

Universitat Oberta de Catalunya  
Programa de Doctorat en Societat de la Informació i el Coneixement

TESI DOCTORAL

CAPTURA I ANÀLISI DEL COMPORTAMENT DELS  
ESTUDIANTS EN ENTORNS VIRTUALS  
D'APRENENTATGE:  
EL CAMPUS VIRTUAL DE LA UOC

Autor:

Enric Mor Pera

Director:

Dr. Julià Minguillón Alfonso

Barcelona, febrer de 2008



## Taula de continguts

Taula de continguts .....	I
Índex de figures .....	V
Índex de taules .....	VII
Agraïments .....	IX
Capítol 1 Introducció .....	1
1.1 Plantejament del problema .....	1
1.2 Objectius i aportacions originals .....	5
1.3 Mètode de treball .....	6
1.3.1 Planificació general .....	6
1.3.2 Aproximació conceptual .....	8
1.4 Estructura del document .....	9
Capítol 2 Estat de la qüestió .....	11
2.1 Interacció Persona - Ordinador .....	11
2.1.1 Disseny Centrat en l'Usuari .....	19
2.1.2 Mètodes de disseny centrat en l'usuari i d'avaluació de la usabilitat .....	24
A. Mètodes d'indagació .....	25
B. Prototipatge, card sortint, personatges, escenaris i test d'usuaris .....	28
C. Mètodes d'inspecció .....	32
2.1.3 Model mental .....	35
2.1.4 Navegació i comportament d'usuari .....	39
2.2 E-Learning i Technology Enhanced Learning .....	42
2.2.1 Models pedagògics, tecnològics i organitzatius .....	46
2.2.2 Sistemes d'e-learning .....	51
2.2.3 Estàndards d'e-learning .....	54
2.2.4 E-Learning i Disseny Centrat en l'Usuari .....	59
2.3 Anàlisi d'ús de llocs web .....	62
2.3.1 Fitxers de log .....	63
2.3.2 Estructura dels fitxers de log .....	64
2.3.3 Anàlisi de fitxers de log .....	67
2.3.4 Mètriques remotes .....	71
2.3.5 Consideracions .....	73
2.4 Tecnologies de modelització d'usuaris i personalització .....	75
2.4.1 Tecnologies de la personalització .....	78
2.4.2 Personalització i e-learning .....	81
2.5 Resum .....	83
Capítol 3 Anàlisi de l'ús i dels usuaris de sistemes d'e-learning .....	85
3.1 Introducció .....	85
3.2 Proposta d'anàlisi multinivell .....	87
3.3 Validació de la proposta multinivell .....	94
3.3.1 Establiment dels nivells d'anàlisi: personatges i escenaris als sistemes d'e-learning .....	94
3.3.2 Avantatges de l'anàlisi a tres nivells .....	97
3.3.3 Nivell 1: Curt termini .....	99
Sessions d'un usuari i patró de navegació .....	101
Agrupació de patrons dels usuaris .....	102
Distribució dels patrons al llarg del temps .....	102
3.3.4 Nivell 2: Mig termini .....	104
3.3.5 Nivell 3: Llarg termini .....	107

3.4 Processament i anàlisi de logs en entorns virtuals d'aprenentatge .....	108
3.5 Proposta de marcatge .....	112
3.5.1 Metodologia de marcatge .....	113
3.5.2 Estratègia basada en l'acció .....	116
3.5.3 Estratègia basada en el contingut .....	118
3.5.4 Consideracions .....	120
3.6 Aspectes de privacitat i seguretat .....	121
3.7 Personalització .....	123
3.7.1 Elements de personalització del campus virtual de la UOC .....	126
3.8 Resum .....	129
Capítol 4 Resultats experimentals .....	131
4.1 Introducció .....	131
4.2 La UOC com a escenari únic .....	132
4.2.1 Disseny del sistema i arquitectura .....	133
4.2.2 Serveis de campus .....	136
4.3 Experiments inicials amb les dades del portal de la UOC .....	139
4.3.1 Navegació dels usuaris i arbres de decisió .....	140
4.3.2 Conjunt de dades .....	140
4.3.3 Extracció de les variables de classificació .....	143
4.3.4 Arbres de decisió de profunditat limitada .....	143
4.3.5 Anàlisi de les dades .....	144
4.3.6 Anàlisi estàtic .....	145
4.3.7 Anàlisi dinàmic .....	146
4.3.8 Conclusions .....	148
4.4 Experiments inicials amb fitxers de log del campus virtual .....	149
4.4.1 Navegació dels usuaris .....	149
4.4.2 Conjunt de dades .....	150
4.4.3 Identificació d'usuaris .....	153
4.4.4 Preprocessat de les dades .....	154
4.4.5 Obtenció dels camins de navegació .....	157
4.5 Gestió i preprocessat dels fitxers de log .....	160
Etapa 1: Filtratge .....	164
Etapa 2: Modificació del format de les línies de log .....	165
Etapa 3: Creació del fitxer "CampusX" .....	166
4.6 Experiments al nivell de sessió .....	167
4.6.1 Experiment: plana d'inici del campus virtual .....	168
4.6.2 Experiment: procés d'obtenció de camins de navegació .....	171
Etapa 1: Fitxers "CampusX" necessaris i línies vàlides .....	173
Etapa 2: Obtenció dels números de sessió i els "uid" .....	174
Etapa 3: Informació de navegació de l'usuari .....	174
Etapa 4: Generació de camins .....	175
Etapa 5: Obtenció del camp "idp" .....	176
4.6.3 Experiment: semestre de primavera del curs 2006-2007 .....	177
Conjunt de dades i preprocessat .....	177
Sessions de navegació .....	180
Caracterització d'usuaris per la seva navegació .....	195
4.6.4 Conclusions dels experiments de nivell 1 .....	199
4.7 Experiments al nivell de curs o semestre acadèmic .....	200
4.7.1 Experiment: estudi del lliurament d'activitats en assignatures .....	200
Dades dels fitxers de log i creuament amb el perfil de navegació .....	202
Preprocessat de les dades i obtenció de variables .....	205
Aplicació de tècniques de mineria de dades .....	210
<i>Data fusion</i> .....	213

4.7.2	Conclusions dels experiments de nivell 2 .....	214
4.8	Experiments al nivell de lifelong learning.....	215
4.8.1	Experiment: segon cicle d'Enginyeria en Informàtica.....	216
	Anàlisi de la matrícula per semestre .....	218
	Combinacions d'assignatures.....	226
	Titulacions progressives.....	228
	Anàlisi de l'abandonament.....	230
	Navegació dels estudiants d'Enginyeria en Informàtica.....	233
4.8.2	Conclusions dels experiments al nivell 3 .....	238
4.9	Resum .....	238
Capítol 5	Conclusions .....	241
5.1	Assoliment dels objectius plantejats.....	241
5.1.1	Resum de l'objectiu general.....	242
5.1.2	Objectius específics i assoliment .....	243
5.2	Revisió de les preguntes de recerca .....	244
5.3	Conclusions principals.....	246
5.4	Línies de treball futures .....	248
5.4.1	Extensions de la proposta.....	248
5.4.2	Aplicació a d'altres entorns interactius.....	249
5.4.3	Perspectives de treball relacionades.....	249
5.5	Consideracions finals.....	250
Annex A	Dades quantitatives de la UOC i dels seus estudiants .....	251
A1	Dades quantitatives del semestre de primavera del curs 2006-2007 .....	251
Annex B	Llistat de tasques dels estudiants de la UOC .....	257
	Període de matrícula .....	257
	Període lectiu .....	258
	Període d'avaluació final .....	258
	Altres tasques que poden dur a terme durant el semestre.....	258
Annex C	Personatges i escenaris a la UOC.....	259
C.1	Introducció .....	259
C.2	Dades de la UOC i dels seus estudiants .....	260
C.3	Tipologies i personatges.....	262
	Hobby.....	263
	Executius.....	264
	Jaume .....	266
	Olga.....	267
C.4	Escenaris, situacions i tasques que realitzen aquests personatges .....	268
Annex D	Mapa de marques del campus virtual de la UOC .....	273
D.1	Introducció .....	273
D.2	Mapa de marques .....	274
D.3	Entrades generades per les marques als fitxers de log.....	279
D.3.1	Login .....	279
D.3.2	Menú Lateral .....	280
D.3.3	Menú superior .....	282
D.3.4	Graella d'aules .....	286
Annex E	Resum d' <i>scripts</i> i procediments de captura i processament de fitxers de log.....	289
E.1	Introducció .....	289
E.2	Servidor: einlinux3.uoc.edu .....	290
E.3	Servidor: personal.uoc.edu.....	291
Annex F	Aportacions originals .....	293
Bibliografia	.....	299



## Índex de figures

Figura 2.1: Esquema de la IPO (Extret de [159]).	13
Figura 2.2: Retroalimentació entre els usuaris i la tecnologia.	16
Figura 2.3: Adaptació (i personalització) de la tecnologia a les persones.	17
Figura 2.4: Cicle d'activitats de la ISO13407.	22
Figura 2.5: Principals mètodes de disseny centrat en l'usuari i d'avaluació de la usabilitat.	24
Figura 2.6: Relació entre el model mental del dissenyador, la imatge del sistema i el model mental de l'usuari.	37
Figura 2.7: Elements del model de la UOC (Extret de [22]).	48
Figura 2.8: Dimensions dels sistemes d'e-learning.	52
Figura 3.1: Interrelació entre els nivells.	90
Figura 3.2: Procés de millora basat en la captura i anàlisi de dades d'ús.	93
Figura 3.3: Diagrama d'espais i enllaços principals del campus virtual.	115
Figura 3.4: Marques del menú lateral.	117
Figura 3.5: Marques del menú superior.	117
Figura 3.6: Cicle de vida de l'estudiant de la UOC.	125
Figura 4.1: Estructura i traspàs de dades en un sistema de tres capes.	135
Figura 4.2: Plana d'Inici del Campus Virtual de la UOC.	138
Figura 4.3: Estructura de la plana principal del lloc web.	141
Figura 4.4: Esquema de tall i fusió de fitxers de log.	155
Figura 4.5: Resultat generat per Pathalizer.	158
Figura 4.6: Esquema de tractament dels fitxers de log.	160
Figura 4.7: Procés de generació del fitxer "CampusX".	164
Figura 4.8: Camins de navegació més freqüents.	170
Figura 4.9: Procés d'obtenció dels camins de navegació.	173
Figura 4.10: Connexions al campus virtual a les diferents hores del dia, d'un dia seleccionat a l'atzar.	179
Figura 4.11: Nombre de sessions distribuïdes per la seva durada (sessions amb durada inferior a 60 segons).	182
Figura 4.12: Distribució del promig de la durada (en segons) al llarg dels dies del semestre.	183
Figura 4.13: Distribució de les sessions al llarg de les hores del dia.	184
Figura 4.14: Distribució de les sessions al llarg dels dies de la setmana.	184
Figura 4.15: Distribució de les sessions al llarg del semestre.	186
Figura 4.16: Nombre d'accions per sessió.	187
Figura 4.17: Nombre d'accions per sessió, de les sessions on apareix la marca "B_SURT".	187
Figura 4.18: Distribució al llarg del curs de la primera acció en la navegació pel campus virtual.	192
Figura 4.19: Distribució al llarg del curs de l'accés als diferents espais en la navegació pel campus virtual.	194
Figura 4.20: Etapes del procés d'anàlisi.	204
Figura 4.21: Nombre d'assignatures matriculades per semestre.	222
Figura 4.22: Nombre d'assignatures aprovades per semestre.	223
Figura 4.23: Nombre d'estudiants (matrícules) que ho aproven tot a final de curs.	224
Figura 4.24: Relació entre assignatures matriculades i assignatures aprovades per cada matrícula.	225
Figura 4.25: Navegació entre semestres amb assoliment del títol.	231
Figura 4.26: Arbre de decisió generat per l'algorisme C5.0 de SPSS Clementine.	235

Figura A.1: Distribució de l'edat.....	253
Figura D.1: Plana d'inici del campus virtual tal i com la veu l'estudiant. ....	275
Figura D.2: Diagrama d'espais i enllaços principals del campus virtual. ....	275
Figura D.3: Marques del menú del campus virtual. ....	277
Figura D.4: Marques dels altres accessos a la bústia. ....	278



## Índex de taules

Taula 2.1: Exemple d'entrada de fitxer de log .....	66
Taula 2.2: Semàntica de les entrades dels fitxers de log .....	66
Taula 3.1: Noms que reben els diferents nivells. ....	98
Taula 3.2: Marques de la plana d'inici. ....	118
Taula 3.3: Relació dels nivells d'anàlisi amb les etapes del cicle de vida de l'estudiant.....	127
Taula 4.1: Resultats obtinguts amb el sistema de classificació estàtic.....	145
Taula 4.2: Partició del conjunt de dades d'acord a una temporalització mensual.....	147
Taula 4.3: Resultats obtinguts amb el sistema de classificació dinàmic. ....	147
Taula 4.4: Resultats obtinguts amb el sistema de classificació dinàmic acumulat. ....	148
Taula 4.5: Primera acció dels estudiants en el campus virtual. ....	189
Taula 4.6: Primera acció dels estudiants en el campus virtual el primer dia de curs. ....	190
Taula 4.7: Resultat del PCA Factorial sobre el conjunt de dades. ....	197
Taula 4.8: Correlació de cada component amb les variables. ....	197
Taula 4.9: Qualificacions obtingudes en la primera activitat. ....	202
Taula 4.10: Distribució de les qualificacions d'acord als clústers obtinguts a "Compiladors I". .....	212
Taula 4.11: Distribució de les qualificacions d'acord als clústers obtinguts a "Fonaments de programació". ....	213
Taula 4.12: Assignatures recomanades amb les troncal en negreta per als dos primers semestres de l'Enginyeria en Informàtica. ....	218
Taula 4.13: Nombre d'estudiants matriculats a cada assignatura, d'acord a cada semestre acadèmic. ....	220
Taula 4.14: Combinacions d'assignatures matriculades i aprovades amb el marge entre suspendre alguna assignatura i aprovar-ho tot. ....	226
Taula 4.15: Combinacions d'assignatures i el nombre d'estudiants que les aproven totes o en suspensen alguna. ....	227
Taula 4.16: Nombre d'estudiants que assoleix cada titulació progressiva.....	229
Taula 4.17: Abandonament a cada semestre. ....	232
Taula A.1: Dades de matrícula de titulacions homologades en català en el semestre de primavera del curs 2006-2007. ....	252
Taula A.2: Sexe. ....	252
Taula A.3: Edat. ....	253
Taula A.4: Motivació per estudiar a la UOC.....	254
Taula A.5: Motivació per continuar estudiant a la UOC.....	254
Taula A.6: Situació laboral.....	254
Taula A.7: Estudis previs. ....	255
Taula A.8: Comarca de residència. ....	255
Taula D.1: Marques de la botonera superior. ....	276
Taula D.2: Marques de la botonera lateral. ....	276
Taula D.3: Marques dels altres accessos a la bústia.....	277



## Agraïments

Des del dia que vaig començar a treballar en aquesta tesi he rebut molta comprensió i suport de les persones del meu voltant i aquestes línies volen mostrar el meu agraïment. A nivell professional, el meu director i amic, el Julià Minguillón, ha sabut estar pendent en tot moment. Moltes gràcies, sense el teu suport, ajut i correccions no hauria portat a terme aquesta tesi. Dono gràcies als meus col·legues i amics del grup de recerca i dels Estudis d'Informàtica, Multimèdia i Telecomunicacions, on he trobat un entorn favorable i col·laboratiu per a realitzar aquest treball. Gràcies també a tots els amics de Tecnologia Educativa amb qui he compartit treball, projectes i aspiracions al voltant de molts dels aspectes que es presenten en aquest treball.

A nivell personal, em sento molt afortunat d'haver tingut a la Maria al meu costat, a qui dedico aquest treball. Moltes gràcies! També vull expressar el meu agraïment als meus pares, Ramon i Teresa, per la seva paciència, comprensió i suport durant aquest temps, a la Maria i la Teresa, i al Joan Ramon i la Maria. Moltes gràcies també al Joan i la Laieta, a l'Anna i l'Eulàlia i als dos Àngels. Dono especialment gràcies al meu amic i germà Vini, que des del primer moment m'ha recolzat incondicionalment i m'ha sabut donar consells i suport sense els quals ara no estaria escrivint aquestes línies.

A tots els amics i amigues que han passat i han estat al meu costat durant la realització d'aquest treball. Tot i que no escrigui el vostre nom sou al meu cap i al meu cor ara mateix. Moltes gràcies!

Barcelona, 15 de febrer 2008

Enric Mor Pera



# Capítol 1

## Introducció

L'objectiu d'aquest capítol és oferir una breu explicació del context en el que es desenvolupa el treball realitzat en aquesta tesi doctoral i una síntesi de la qüestió que es vol resoldre, descrivint tant els objectius com el mètode de treball utilitzat en la recerca. Les principals aportacions també s'introdueixen i resumeixen aquí, tot i que s'aniran detallant al llarg de tot el document.

La idea bàsica que va motivar aquesta tesi doctoral és l'observació, com a professor, que en un entorn virtual d'aprenentatge hi ha una gran quantitat d'informació implícita en l'ús que en fan els seus usuaris que pot ser aprofitada per a millorar el propi entorn virtual i, de retruc, millorar l'experiència dels estudiants.

### 1.1 Plantejament del problema

L'educació en línia o e-learning representa, cada cop més, una alternativa real a l'ensenyament tradicional, més enllà d'una nova forma d'accedir a la formació per a aquelles persones que, per diferents motius, no ho poden fer de cap altra manera [90]. A més del consum de recursos digitals, la formació mitjançant entorns virtuals d'aprenentatge dona a l'estudiant la possibilitat de ser el centre del procés d'aprenentatge, amb un grau elevat d'interacció amb la resta d'elements: altres estudiants, professors, recursos docents i el propi entorn virtual.

El fet que cada cop més persones accedeixin a l'educació en línia fa que els diferents aspectes d'ús i interacció amb els ordinadors, els entorns virtuals d'aprenentatge i els materials didàctics siguin un aspecte essencial que determina el seu èxit o fracàs, així com la satisfacció o frustració dels seus usuaris [32]. Al dissenyar un entorn virtual d'aprenentatge és fàcil caure en el parany de pensar que s'està dissenyant per a un usuari ideal, per a un estudiant arquetípic, i que tots els usuaris es comporten de forma similar. Aquest usuari ideal no existeix i és necessari tenir en compte que el sistema serà utilitzat per persones reals i no per estudiants arquetípics, i és necessari saber qui són i què necessiten. Investigant als usuaris es trenca la perspectiva limitada que es pot tenir al construir un sistema i es pot dissenyar des del punt de vista de les persones que l'utilitzaran. Així, a l'hora de dissenyar un sistema o entorn virtual, és important tenir en compte aquests aspectes bàsics del camp de la interacció persona-ordinador: el dissenyador no és l'usuari, i malgrat el sentit comú és important, no és suficient. Cal observar i estudiar què fan i com es comporten els usuaris i, només llavors, aplicar el sentit comú [63].

L'objectiu que es persegueix al dissenyar sistemes interactius tenint en compte els seus usuaris és proporcionar una experiència d'ús satisfactòria [73], és a dir, que les persones que utilitzen el sistema puguin fer allò que volen fer, de manera senzilla, eficient i satisfactòria. És per tant important disposar d'informació sobre els usuaris, qui són, les seves característiques, necessitats i objectius, conèixer les interaccions que realitzen amb un sistema, com l'utilitzen per assolir els seus objectius, resumint, què fan. Aquesta informació recollida en un model d'usuari, ja sigui formal o informal, permet saber com s'està utilitzant un sistema i millorar el seu disseny, tant a nivell intern i de rendiment, com a nivell del disseny d'interacció [155]. L'objectiu primordial d'aquesta tesi doctoral és estudiar com identificar i analitzar aquesta informació d'ús per a millorar el propi sistema.

Conèixer què fan realment els usuaris no és una tasca senzilla i requereix investigar en profunditat. A l'hora de dissenyar les tasques i processos que els usuaris portaran a terme en un sistema, molts cops es parteix d'hipòtesis d'ús que tenen en compte els perfils tant demogràfics i socio-culturals com psicològics dels usuaris. Tot i que

aquestes hipòtesis constitueixen un bon punt de partida, es fa necessari conèixer què fan realment els usuaris per a poder contrastar-les. Saber quines són les seves accions i el seu comportament dins d'un entorn virtual proporciona informació al seu dissenyador per a millorar-lo, i així millorar l'experiència dels usuaris. D'aquesta manera, s'estableix un cercle de millora contínua, fent evolucionar el sistema amb els seus usuaris.

En els entorns i plataformes d'e-learning, és encara més important conèixer com actuen i es comporten els estudiants dins del sistema. Conèixer el comportament de navegació dels diferents estudiants pot proporcionar una informació que serà de gran ajuda per a adequar l'entorn als estudiants i millorar la usabilitat i l'experiència d'ús del sistema. A més, aquesta informació proporciona un coneixement inestimable als professors i docents que dissenyen accions formatives, ja que el podran utilitzar per a la millora del procés d'ensenyament i aprenentatge, i per adaptar i personalitzar itineraris formatius a diferents nivells. D'aquesta manera, el disseny de sistemes d'e-learning planteja un doble repte, per una banda ha de proporcionar funcionalitats avançades mitjançant una interfície senzilla i, per altra banda, ha de ser pedagògicament adequat de manera que els estudiants assoleixin els objectius d'aprenentatge. Aquesta dualitat necessita ser contrastada amb dades reals d'ús.

El coneixement sobre els estudiants, sobre les seves accions i l'ús que fan dels entorns virtuals es pot usar per a dissenyar un sistema d'e-learning basat en estàndards que permeti, per exemple, adaptar els itineraris formatius a les diferents necessitats de cada estudiant. La idea d'aquests itineraris és convertir el pla docent o guia d'estudi, un element que normalment és lineal, estàtic i aïllat, en el mitjà articulador d'un procés dinàmic i variable que, tenint en compte els elements claus de la usabilitat i la experiència d'usuari, involucri aspectes de disseny instruccional, de personalització centrada en l'estudiant i es relacioni amb tots els objectes i activitats d'aprenentatge que apareixen al llarg d'un període acadèmic. La utilització d'estàndards d'e-learning per a la representació i construcció d'aquests itineraris proporcionarà una presentació dels continguts adequada a cada estudiant i la llibertat per avançar al seu ritme, tenint en compte un marc prèviament definit. Aquest marc està constituït per un equip docent tenint en compte els aspectes de disseny instruccional i d'experiència d'ús. Per la seva

banda, es vol que l'entorn o plataforma faci un seguiment dels usuaris i de totes les accions que realitzen amb dos objectius: primer, recollir informació que pugui ser posteriorment analitzada per a extreure coneixement útil per l'avaluació de la usabilitat i la millora de l'experiència d'ús; segon, adaptar el procés d'aprenentatge segons les necessitats, el ritme i les accions de cada usuari, així com dels resultats acadèmics que obté, revisant totes les hipòtesis implícites en el seu disseny.

Adaptar i adequar el procés d'aprenentatge a les necessitats específiques dels estudiants serveix com a excusa per a dissenyar i construir un entorn virtual d'aprenentatge que, a partir del coneixement dels usuaris, fa que el procés d'ensenyament i aprenentatge i les diferents activitats relacionades constitueixin una experiència educativa satisfactòria per a l'estudiant. Aquesta millora té lloc no només perquè s'identifiquen i s'eliminen els possibles elements de frustració, sinó perquè els estudiants realment assoleixen els seus objectius d'aprenentatge i ho fan d'una manera emocional [62, 130], és a dir, que té una influència positiva sobre les emocions que es produeixen a l'utilitzar el sistema. A més, una situació negativa que es produeix especialment en els ensenyaments no presencials, facin ús intensiu de la tecnologia o no, és l'elevat grau d'abandonament dels estudis per part dels estudiants [106], element clau que cal tenir en compte i que mereix ser estudiat en profunditat, més enllà dels resultats que es presenten en aquesta tesi doctoral.

Aquest treball té la característica que el seu camp d'estudi és el de la Universitat Oberta de Catalunya i el seu campus virtual, disposant així d'una quantitat important d'usuaris i de dades que permetran presentar resultats significatius. La UOC i el seu campus virtual constitueixen un escenari únic, i el fet d'utilitzar-los com a banc de proves i validació permetrà generar resultats directament relacionats amb les seves particularitats, així com resultats més generals que tindran com a focus les mètriques d'ús, la navegació, la personalització i les seves relacions amb l'e-learning. La difusió d'aquests resultats és, per tant, una difusió de la realitat de la pròpia UOC en primer lloc i del seu campus virtual en segon.

Així doncs, aquest treball de tesi engloba diversos punts de vista com són els de la interacció persona-ordinador, la usabilitat, l'e-learning i l'anàlisi d'ús de llocs web, i el campus virtual és l'escenari on s'utilitzen tots ells.



## 1.2 Objectius i aportacions originals

Metafòricament, aquest treball es planteja fer una fotografia detallada del campus virtual de la UOC i dels seus estudiants, i proposa una manera específica de fer-la. L'objectiu d'aquesta fotografia és mostrar el comportament dels usuaris del sistema en diferents situacions a diferent resolució, segons el que s'estigui analitzant. Aquesta informació permetrà validar o refutar hipòtesis de disseny del propi sistema i introduir-hi millores. La fotografia mostrarà l'entorn virtual d'aprenentatge, els estudiants i les seves interaccions en un moment específic i, per tant, no serà exactament reproduïble. No obstant, la metodologia es podrà utilitzar de nou per realitzar nous estudis que millorin l'entorn virtual d'aprenentatge i l'experiència dels seus estudiants, en un cicle continu.

L'extracció d'informació sobre els estudiants i la seva navegació es vol utilitzar per a fer propostes d'adaptació i personalització del procés d'aprenentatge adequant-lo a les seves necessitats. A partir d'aquests objectius es plantegen les preguntes de recerca que es presenten a continuació.

- És possible capturar la navegació dels usuaris d'un campus virtual? Com es poden obtenir i quins són els patrons d'ús i de navegació d'aquests usuaris? Les dades disponibles actualment proporcionen informació suficient o es necessiten altres fonts de dades?
- És adequat analitzar les dades i estudiar els usuaris establint diferents nivells d'anàlisi? Quins són aquests nivells? Quina és la informació rellevant a cada nivell? Com es pot conceptualitzar aquest anàlisi multinivell?
- Poden obtenir-se noves evidències, o millorar les existents, sobre el campus virtual i els seus usuaris? Quines mètriques i indicadors d'ús s'obtenen? Aquesta informació permet avaluar el sistema i proposar millores?

- Quines tendències de navegació s'observen en un campus virtual? Quines tipologies d'usuaris es desprenen de la navegació i quines característiques presenten? Es pot millorar l'experiència dels estudiants amb la informació obtinguda?
- Com es pot adequar l'entorn virtual als usuaris i com es poden adaptar els itineraris i les accions formatives als diferents estudiants en funció de les seves necessitats, comportament i hàbits d'ús?

Els dos primers blocs de qüestions s'adrecen al capítol 3, amb un enfocament més metodològic, mentre que la resta s'adrecen entre el capítol 3 i el capítol 4, aquest darrer amb una aproximació de caire més experimental.

### **1.3 Mètode de treball**

En aquesta secció es descriu la planificació general del treball realitzat, així com l'aproximació conceptual del model seguit.

#### **1.3.1 Planificació general**

La metodologia que s'ha seguit en aquest treball de recerca comprèn les etapes següents:

1. Estudi de l'estat de la qüestió
2. Formulació de la proposta (definició del model).
3. Descripció de l'escenari objecte d'estudi.
4. Obtenció de resultats experimentals (validació del model).
5. Formulació de conclusions.

En una primera etapa es realitza l'estudi de l'estat de la qüestió de les diferents àrees de coneixement entre les quals es situa aquest treball. Es presenta el camp de la interacció persona-ordinador i la importància del disseny centrat en l'usuari en els sistemes

interactius en general i en els sistemes d'e-learning en particular. El disseny centrat en l'usuari orienta la recollida i anàlisi de les dades d'ús de l'entorn, així com mostra la idoneïtat dels mecanismes d'adaptació i personalització proposats com a via per a assolir un entorn fàcil d'utilitzar per totes les persones. També es presenta com la tecnologia ha canviat els processos d'ensenyament i aprenentatge, es defineixen els sistemes d'e-learning i es descriuen els seus components, així com la importància dels estàndards d'e-learning per a la creació d'entorns virtuals d'aprenentatge més usables. Per finalitzar, es presenten les aproximacions d'anàlisi d'ús de llocs web així com els avantatges i inconvenients d'utilitzar fitxers de log. Es descriu el context en el qual es desenvoluparà el model d'anàlisi plantejat.

En la següent etapa s'aborda l'estudi del comportament i navegació dels usuaris mitjançant tres nivells d'anàlisi. Aquests nivells són els de sessió, curs i *lifelong learning*, necessaris per la proposta d'anàlisi que es presenta en aquesta tesi. A cada nivell s'identifiquen les fonts de dades necessàries i s'estableixen diferents objectius, relacionats amb el procés d'aprenentatge. Aquesta proposta multinivell permet descriure el model plantejat seguint diferents objectius en cada cas.

La UOC i el seu el campus virtual constitueixen l'escenari d'estudi d'aquest treball i és on s'aplica la metodologia d'anàlisi que es proposa i on es realitzen experiments a cadascun dels nivells definits. Les dades ja existents a aquesta universitat, moltes de manera implícita, són les que s'utilitzen en els experiments que es realitzen, especialment els fitxers de log generats pel campus virtual i que queden recollits a diari en els servidors. Les dades de log presenten mancances i, per tant, és necessari enriquir-les amb dades de marques de navegació integrades en el campus virtual, així com altres dades addicionals de les bases de dades institucionals.

L'obtenció de resultats i evidències rellevants es porta a terme a cada nivell d'anàlisi mitjançant diferents tècniques i mètodes. Així, al primer nivell, l'obtenció de camins de navegació requereix un complex sistema de gestió i processat de fitxers de log, mentre que l'obtenció de patrons i de resultats a nivell de curs i *lifelong learning* es realitza seguint mètodes propis de la mineria de dades [65].

Els resultats que s'obtenen han de permetre identificar si existeix una divergència entre els diferents aspectes del disseny, tant del sistema com del procés d'aprenentatge, i de l'ús que se'n fa. Aquests resultats són interessants per una banda per a la millora del campus virtual de la UOC i, per altra banda, pel disseny de sistemes d'e-learning en general, que siguin propers a l'usuari. A partir dels resultats obtinguts es posa en evidència certs fets propis del disseny que poden millorar-se, a partir d'un conjunt de propostes de personalització sobre el procés d'aprenentatge.

### **1.3.2 Aproximació conceptual**

A la vegada, aquesta tesi doctoral fa una aproximació empírica i experimental per a mostrar la importància de l'anàlisi de l'ús i la navegació en els entorns i plataformes d'e-learning, i les estratègies per a augmentar i millorar l'experiència dels estudiants, utilitzant tota la informació disponible, ja sigui provinent de fitxers de log o d'altres fonts d'informació. El treball que es presenta realitza una exploració sobre les diferents fonts de dades de la UOC, per així obtenir informació rellevant que permeti descobrir evidències i descriure els fets observats en relació a la navegació dels usuaris i el seu comportament dintre d'un entorn virtual d'aprenentatge.

L'aproximació que es porta a terme es basa, per una banda en el *design research* [79] i el *desing and creation research* [135] utilitzat en l'àrea de sistemes d'informació. Per altra banda, també s'utilitza la pràctica basada en la evidència, mètode utilitzat habitualment en la tecnologia en general i en la informàtica en particular, on es dissenya, construeix i refina un prototipus fins a assolir els resultats desitjats.

Per tant, les principals qüestions que s'aborden en aquest treball són la captura i l'anàlisi de la navegació i comportament dels usuaris del campus virtual de la UOC en diferents períodes de temps. Així, es porta a terme una proposta sistemàtica que té en compte l'estat de la qüestió, la recollida de dades, el seu processament i anàlisi apropiats i la interpretació adequada dels resultats obtinguts, usant una estructura multinivell. A partir d'aquests resultats es fan propostes de millora per a cadascun dels nivells estudiats.

## **1.4 Estructura del document**

Aquesta memòria de tesi doctoral s'estructura com es descriu a continuació. En aquest primer capítol es presenten els objectius i aportacions del treball de tesi, així com la metodologia del treball de recerca. En el capítol 2 es descriu l'estat de la qüestió en relació a la interacció persona-ordinador, l'e-learning, l'anàlisi d'ús de llocs web i el modelatge d'usuaris i personalització. En el capítol 3 es presenta la proposta de recollida de dades i d'anàlisi dels usuaris d'un entorn d'e-learning, establint tres nivells que delimiten diferents intervals temporals, determinats per la naturalesa de l'escenari que s'estudia i els objectius dels seus usuaris. En el capítol 4 es descriu la captura i recollida de dades, així com el processament necessari per a obtenir la navegació dels usuaris. També es presenten els resultats dels experiments realitzats, tant els preliminars com a cadascun dels tres nivells proposats. En el capítol 5 es presenten les conclusions d'aquest treball i les futures línies de recerca que es poden desenvolupar a partir del mateix.



## Capítol 2

### Estat de la qüestió

La tesi que es proposa versa sobre l'estudi d'entorns d'e-learning i l'obtenció de camins i patrons de comportament dels usuaris que puguin servir com a mesures i indicadors per a avaluar la usabilitat i millorar l'experiència d'ús d'aquests entorns, i que permetin proporcionar adaptació i personalització del procés d'aprenentatge. Els àmbits de coneixement que tenen relació amb aquest objectiu són la interacció persona-ordinador, l'e-learning (o *technology enhanced learning*), l'anàlisi d'ús de llocs web i la personalització. A continuació es mostra l'estat de l'art d'aquests àmbits de coneixement i la seva vinculació amb el treball que es planteja. Tot i que no es tracta d'una tesi exclusivament d'interacció persona-ordinador, es considera important començar per una visió més centrada en l'usuari, l'element clau del procés d'aprenentatge en entorns virtuals.

#### 2.1 Interacció Persona - Ordinador

Tot i que és difícil proporcionar una definició àmpliament acceptada, la Interacció Persona-Ordinador (IPO) es pot definir breument com l'àrea de coneixement que estudia la interacció entre les persones i els ordinadors. Una de les definicions més acceptades la va proposar l'any 1996 el grup d'especial interès en Computer-Human Interaction de la Association for Computing Machinery<sup>1</sup>, i defineix la IPO com la

---

<sup>1</sup> ACM SIGCHI, <http://www.acm.org/sigchi/>

disciplina relacionada amb el disseny, avaluació i implementació de sistemes interactius per a l'ús per part d'humans, i amb l'estudi dels aspectes i fenòmens amb què està relacionada. Altres autors, com Terry Winograd al llibre de Preece [143], defineixen la IPO com la disciplina que s'ocupa no de l'estudi dels humans, ni tampoc de l'estudi de la tecnologia, sinó de l'espai que hi ha entre els dos.

La interacció persona-ordinador tracta sobre el disseny de sistemes informàtics que tenen en compte a les persones per tal de facilitar que aquestes realitzin les seves activitats de manera productiva i segura. La IPO juga un paper important en el disseny i el desenvolupament de tot tipus de sistemes, des d'aquells on la seguretat és extremadament important, com poden ser els sistemes de control aeri, fins als jocs d'ordinador, orientats a l'entreteniment, passant per les aplicacions ofimàtiques, on la productivitat i satisfacció són elements clau [143].

La IPO persegueix que les persones no s'hagin d'adaptar a les màquines sinó que aquestes estiguin dissenyades per a satisfer les necessitats de les persones. Per aquest motiu, la disciplina aporta un conjunt de tècniques centrades en l'usuari, tant per conèixer els seus objectius i necessitats, com per dissenyar tenint-los en compte [63].

El principal objectiu de la Interacció Persona-Ordinador és doncs millorar la interacció entre les persones i els ordinadors. Aquest objectiu es concreta en els següents: entendre els factors que determinen com les persones utilitzen la tecnologia, desenvolupar eines i tècniques que facilitin el disseny i construcció de sistemes adequats a les persones i assolir una interacció eficaç, eficient i segura [143].

La Interacció Persona-Ordinador inclou, per tant, diferents àrees de coneixement, que estudien tant els aspectes humans com els aspectes tecnològics, passant per un ventall de camps com són la psicologia, sociologia, lingüística, disseny, enginyeria i informàtica. El que siguin moltes les disciplines que han de treballar conjuntament per aconseguir els objectius fa que la IPO sigui eminentment interdisciplinària. La tecnologia juga un paper molt important a la IPO, però com és pot veure és una disciplina que no pot treballar-se únicament des del punt de vista tecnològic. A la Figura 2.1 es mostra el esquema de la disciplina de la Interacció Persona-Ordinador tal com



apareix al *ACM SIGCHI Curricula for Human-computer Interaction* [159] amb tots els elements que intervenen.

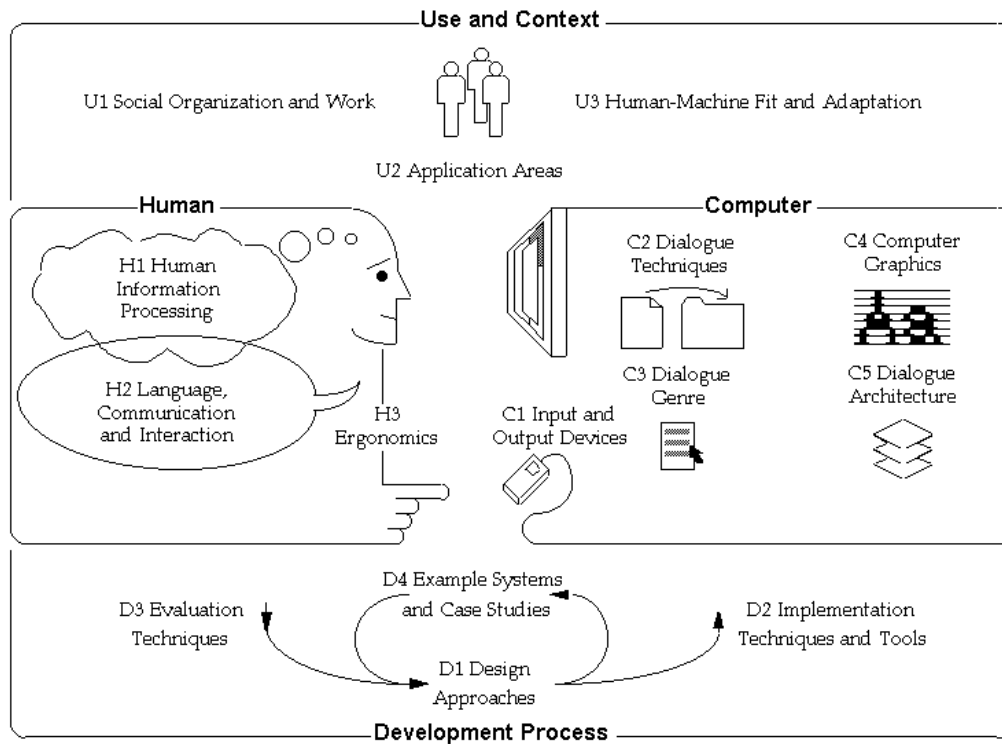


Figura 2.1: Esquema de la IPO (Extret de [159]).

La interacció persona-ordinador esdevé d'especial importància en l'e-learning donat que actualment l'ordinador és el mitjà d'accés al campus virtual i és un dels recursos més utilitzats per la distribució de cursos i continguts educatius. A més, els ordinadors són adequats no només per presentar entorns rics en els quals els estudiants poden treballar col·laborativament, sinó també perquè proporcionen capacitats multimèdia que fan possible distribuir i presentar materials educatius d'una manera difícil d'imaginar fa uns anys. Aquests aspectes representen un gran repte tant pels docents com pels dissenyadors instruccionals. Tot i això, aquest escenari que es presenta magnífic i prometedor, molts cops no s'acaba d'aprofitar o quan s'intenta s'obtenen resultats no gaire bons. Això ve motivat perquè es necessiten uns coneixements i unes guies clares sobre com dissenyar i construir sistemes interactius que realment lliguin amb les necessitats dels estudiants.

Directament relacionats amb la interacció persona-ordinador, hi ha dos conceptes que resulten especialment interessants, el disseny centrat en l'usuari i la usabilitat. Ambdós presenten aspectes concrets orientats al disseny, construcció i avaluació de sistemes interactius que tenen en compte als usuaris que els utilitzaran.

La usabilitat fa referència a la facilitat d'ús dels productes i dels sistemes interactius, acostant-los a les necessitats de les persones. Els estàndards relacionats amb la usabilitat definits i acreditats per la ISO (*International Organization for Standardization*) [14] proposen algunes definicions. En relació a la qualitat, la norma ISO9241-11 [86] defineix la usabilitat com la eficàcia, la eficiència i la satisfacció amb la que un producte pot ser usat per usuaris específics per assolir objectius específics en un context determinat.

Així doncs, la usabilitat té en compte diversos aspectes:

- a) l'eficàcia, en tant que les persones poden portar a terme i completar tasques, assolir els seus propis objectius amb el producte o sistema, és a dir, poden fer el que necessiten fer;
- b) l'eficiència, entesa com la possibilitat que les persones completin tasques i assolixin objectius de la millor manera possible, sense que els suposi un esforç ni quantitat de temps excessius o desmesurats;
- c) la satisfacció, mesurable a partir del què opinen els usuaris de la facilitat d'ús dels productes o sistemes i de si poden o no portar a terme i completar el que necessiten fer, de la millor manera possible i de forma satisfactòria, agradable i fins i tot emocional.

Al mateix temps, tal com la pròpia definició mostra, la usabilitat es veu afectada per:

- a) Els usuaris: Qui utilitza o utilitzarà el producte o sistema? Són usuaris novells o experts? Han rebut algun tipus de formació o entrenament per a utilitzar-lo?
- b) Els seus objectius: Que volen fer els usuaris amb el producte o sistema? El producte o sistema serveix per fer el que els usuaris volen fer?
- c) El context o situació d'ús: On i com els usuaris utilitzaran el sistema o producte? Aquest aspecte és tant important com els anteriors, malgrat sovint es deixa de

banda, cosa que pot portar al fracàs dels sistemes o productes que es comercialitzen.

La usabilitat i els seus atributs són aplicables tant a productes físics com a productes digitals o sistemes *software* interactius. La enginyeria del programari és l'àrea de la informàtica que compren els mètodes, tècniques i eines necessàries per dur a terme els projectes de desenvolupament de programari i el seu manteniment posterior. En l'enginyeria del programari, la usabilitat ha estat sempre fortament lligada amb el disseny i avaluació de les interfícies d'usuari. La norma ISO/IEC9126 [84, 87] fa referència a la qualitat del programari i la seva avaluació, i defineix la usabilitat com el conjunt d'atributs que dirigeixen sobre l'esforç necessari per a l'ús, i sobre l'avaluació individual de tal ús, per a un grup concret d'usuaris. Aquesta definició de l'any 1991 va ser actualitzada l'any 2000, per passar a parlar de la usabilitat com la capacitat d'un sistema software de ser copsat, après, utilitzat i ser atractiu per l'usuari en unes condicions específiques. Al parlar de condicions específiques es fa referència al context d'ús, que sempre es té en compte en la IPO, i fa referència a que un sistema o producte no té usabilitat intrínseca sinó que pot tenir la capacitat de ser usat en un context particular.

Altres organitzacions relacionades amb la tecnologia, la innovació, el disseny i el desenvolupament de sistemes tecnològics i que defineixen estàndards també proposen definicions per la usabilitat. El IEEE (*Institute of Electrical and Electronics Engineers*) [11] l'any 1990 va definir la usabilitat com la facilitat amb la que un usuari pot aprendre a operar, preparar entrades i interpretar les sortides d'un sistema o component [2].

En aquestes definicions s'observen aspectes i elements comuns. Aquests elements són les persones, els sistemes i com els sistemes han de poder ser utilitzats per les persones. Són definicions recents ja que fan referència a sistemes tecnològics i informàtics que fins que no es van a començar a construir, implantar i estendre no va sorgir la necessitat d'estudiar les seves implicacions relacionades amb l'ús que les persones en fan. És per això que, a l'igual que l'ergonomia és un element fonamental en la vida de les persones ja que constantment interaccionen amb objectes físics, la usabilitat està esdevenint un altre element clau en el dia a dia de les persones ja que fa referència a la ergonomia de

la tecnologia [43], i cada cop més la tecnologia està present en els diferents objectes quotidians.

Per tant, la interacció persona-ordinador i la usabilitat pretenen fer la vida més fàcil a les persones promovent que la interacció amb la tecnologia sigui una experiència satisfactòria. Això és tot un repte que implica dissenyar i construir sistemes que tinguin en compte a l'usuari i que s'adaptin a les seves necessitats i no a l'inrevés. L'objectiu és, per tant, que la interacció de les persones amb la tecnologia no hagi d'anar acompanyada d'un complicat manual d'instruccions. En relació a l'e-learning, aquest objectiu es tradueix en dissenyar i construir entorns virtuals d'aprenentatge que tenen en compte les característiques i necessitats dels estudiants, i que converteixen els diferents processos i tasques en una experiència positiva tant pel que fa referència a la navegació per l'entorn com al propi aprenentatge.

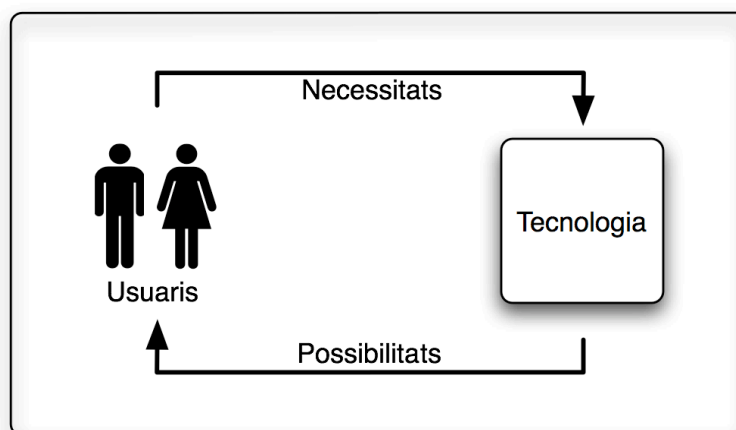


Figura 2.2: Retroalimentació entre els usuaris i la tecnologia.

La relació de les persones amb la tecnologia és un procés dinàmic que es retroalimenta i es mostra a la Figura 2.2. A mesura que les persones demanen a la tecnologia que els proporcioni la possibilitat de fer coses noves, la tecnologia no només ofereix allò que se li demana, sinó que amplia les possibilitats del que es pot fer, oferint noves oportunitats. Aprofitar aquestes oportunitats fa que es puguin fer coses noves i això provoca que se li tornin a plantejar noves necessitats [143]. Es pot dir que la tecnologia augmenta a les persones. Per això es prefereix parlar de *Technology Enhanced Learning* més que no pas e-learning.

En aquest cercle d'oportunitats que ofereix la tecnologia no s'ha d'oblidar que l'objectiu de la interacció persona-ordinador és fer que la tecnologia s'adapti a les persones i no a l'inrevés (Figura 2.3). Així doncs, a l'hora de dissenyar sistemes interactius s'ha de tenir en compte no només la tecnologia que hi haurà d'haver al darrere sinó també les persones que l'hauran d'utilitzar, les seves necessitats i desitjos, les emocions que suscitarà, la navegació i interacció, etc. Per tant, l'objectiu és que els ordinadors, rígids i digitals, s'adaptin a les persones, flexibles i analògiques. Només així s'aconseguirà que els usuaris experimentin una veritable experiència d'ús satisfactòria.

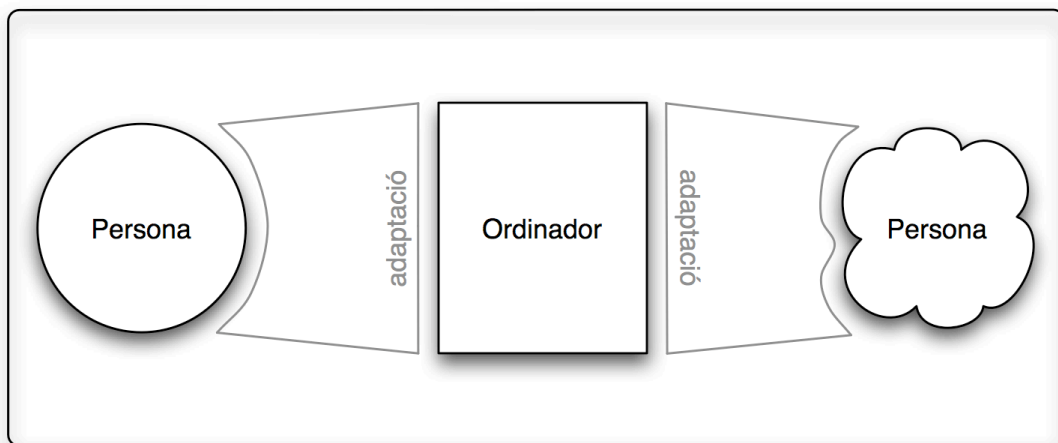


Figura 2.3: Adaptació (i personalització) de la tecnologia a les persones.

La interfície d'un sistema és la part física d'aquest que l'usuari veu i amb la que interactua. La interfície d'usuari és tot allò dissenyat i incorporat en un ordinador o dispositiu, amb el que una persona pot interactuar [63]. Les interfícies gràfiques d'usuari aprofiten la capacitat gràfica dels ordinadors per presentar en la pantalla un conjunt d'elements visuals, gràfics i textuais amb l'objectiu de presentar una interfície intuïtiva, amigable i fàcil d'utilitzar. El disseny d'aquestes interfícies gràfiques es recolza en metàfores i analogies del món real.

La interacció es pot veure com un diàleg entre el sistema i l'usuari [63]. La manera en com els usuaris es comuniquen amb un sistema interactiu s'anomena estil d'interacció. Els estils d'interacció més comuns són la línia de comandes, els menús i formularis, i la manipulació directa. L'estil d'interacció que més s'utilitza actualment és la manipulació

directa. La elecció de l'estil d'interacció té un efecte important en el diàleg dels usuaris amb el sistema.

La manipulació directa [157] és un estil d'interacció que es basa en una interfície que mostra a l'usuari totes les seves possibilitats en forma de menús representant accions, i icones representant objectes. Respecte als paradigmes previs, la manipulació directa representa el canvi des d'una complexa sintaxi de comandes que calia memoritzar, a la manipulació d'objectes mitjançant accions ràpides, incrementals i reversibles que provoquen un efecte visible immediat en l'objecte seleccionat. La manipulació directa permet als usuaris novells explorar la interfície i accedir, sense la necessitat d'un estudi en profunditat, a les funcionalitats del sistema, proporcionant una sensació de control. És molt important la realització d'un bon disseny visual i així evitar analogies incorrectes i conclusions errònies dels usuaris.

Les interfícies gràfiques d'usuari de manipulació directa es basen en una metàfora, una representació familiar del món real per a que així els usuaris entenguin fàcilment el significat dels elements gràfics i les accions que poden realitzar. Les metàfores juguen un paper molt important en el disseny de les interfícies [58]. El disseny i disposició dels elements a la pantalla, el tipus d'interacció que poden realitzar els usuaris, els noms de les accions i components tot està condicionat per la metàfora que s'utilitza. Per exemple, molts dels sistemes operatius actuals fan servir una interfície gràfica que segueix la metàfora de l'escriptori. D'aquesta manera, els usuaris poden veure i reconèixer entre d'altres elements els directoris representats amb el dibuix d'una carpeta, i el lloc on esborrar arxius i carpetes es representa amb una paperera. Les metàfores i la manipulació directa en el disseny d'interfícies gràfiques d'usuari està molt relacionat amb el model mental dels usuaris, que es presenta amb més detall a la secció 2.1.3.

Tot sovint, al parlar d'interfícies d'usuari i de la seva facilitat d'ús, es pot caure en el parany d'identificar la usabilitat d'un sistema amb el disseny de les interfícies d'usuari. Tenint en compte això, és interessant fer especial èmfasi no únicament en el disseny d'interfícies sinó en les metodologies que, ben utilitzades i aplicades, redunden i desemboquen en una interfície entenedora, intuïtiva i fàcil d'utilitzar, i que fan que la

interacció amb el sistema o entorn sigui una experiència satisfactòria. Aquestes metodologies formen part del que s'anomena Disseny Centrat en l'Usuari (DCU) [25] que és el procés de disseny que involucra a l'usuari en totes les fases del procés, des de la conceptualització fins a l'avaluació, passant per la seva implementació. L'objectiu del disseny centrat en l'usuari és la creació de productes que els usuaris trobin útils i usables.

### **2.1.1 Disseny Centrat en l'Usuari**

El Disseny Centrat en l'Usuari (DCU) és el terme general que s'utilitza per descriure el disseny on l'usuari influeix en el resultat final. És, al mateix temps, una filosofia i un procés. Una filosofia, una orientació estratègica, que situa a la persona en el centre per tal de desenvolupar un producte adequat a les seves necessitats i requeriments, i un procés de disseny que es focalitza en els factors cognitius de les persones i com aquests intervenen en les seves interaccions amb els productes [134].

El terme s'originà al laboratori de recerca del Donald A. Norman a la *University of California San Diego* (UCSD). La seva àmplia difusió es va produir després de la publicació del llibre *User Centered System Design: New Perspectives on Human-Computer Interaction* [134], i es va consolidar amb *The Design of Everyday Things* [131] originalment publicat com *The Psychology of Everyday Things* [133], que constitueix una fita en l'àmbit de coneixement.

Tot i que el paradigma del disseny centrat en l'usuari és aplicable al desenvolupament de qualsevol tipus de producte, és en els productes amb una forta component tecnològica (tant de maquinari com de programari) on té una especial importància, donat que és molt habitual que en els desenvolupaments d'aquest tipus de productes es faci molt èmfasi en les seves prestacions tècniques, en detriment molts cops d'aspectes que poden facilitar el seu ús per part de les persones que els utilitzaran [131].

El disseny centrat en l'usuari persegueix obtenir informació sobre els usuaris, les seves tasques i els seus objectius, i utilitzar la informació obtinguda per orientar el disseny i

desenvolupament dels productes [151]. D'entre les diferents informacions que es consideren, a tall d'exemple es destaquen les següents:

- Qui són els usuaris del producte?
- Quines són les tasques i objectius dels usuaris?
- Quin és el nivell de coneixement i la experiència prèvia dels usuaris amb la tecnologia?
- Quina és l'experiència dels usuaris amb el producte o amb productes similars?
- Quines són les funcionalitats que els usuaris esperen del producte?
- Quina informació del producte poden necessitar els usuaris? De quina manera la necessitaran?
- Com pensen els usuaris que funciona el producte?
- Com pot el disseny del producte facilitar els processos cognitius dels usuaris?

En el context de l'e-learning i del disseny d'entorns virtuals d'aprenentatge, aquestes preguntes s'han d'utilitzar per orientar i enfocar el disseny i desenvolupament dels entorns i continguts d'e-learning. Així doncs, caldrà saber qui són els estudiants, les seves característiques, les tasques i objectius d'aprenentatge que hauran de realitzar, el seu coneixement i experiència prèvia en educació en línia, etc.

El disseny centrat en l'usuari pot millorar la utilitat i la usabilitat dels productes i objectes quotidians, des d'aparells de telèfon mòbil a sofisticats sistemes de programari, i en general de qualsevol objecte amb el qual les persones interactuen. Un exemple de sistemes on s'han aplicat els principis del disseny centrat en l'usuari són algunes màquines de pagament d'aparcament (tot i que es poden trobar màquines que serveixen d'exemple de mal disseny). En alguns aparcaments es poden trobar màquines on efectuar molt fàcilment el pagament de la estada del vehicle, la màquina guia totes les passes que s'han de portar a terme i gairebé no hi ha possibilitat d'error; indica que s'ha d'introduir el tiquet i assenyala clarament on i com s'ha de fer, indica la quantitat a pagar, mostra les modalitats de pagament i indica on i com s'ha de fer, retorna el tiquet validat per la sortida i, per finalitzar, ofereix la possibilitat de demanar un comprovant de pagament. Un contraexemple de sistema on no s'ha aplicat el disseny centrat en l'usuari poden ser les portes d'entrada al metro de la ciutat de Barcelona. En algunes estacions els usuaris han de cancel·lar el bitllet introduint-lo a la seva dreta i a d'altres



estacions ho han de fer per la seva esquerra; aquest comportament diferent i aparentment sense raó justificada, provoca desconcert i confusió diàriament, a molts usuaris del transport públic.

En tant que procés, el disseny centrat en l'usuari involucra a l'usuari en totes les fases a través de les quals es desenvolupa un producte, des de la seva conceptualització fins a l'avaluació, passant, en molts casos, pel seu desenvolupament. L'objectiu del disseny centrat en l'usuari és la creació de productes que els usuaris trobin útils i usables [25].

El disseny centrat en l'usuari es relaciona estretament amb la utilitat dels productes i sistemes i la seva usabilitat, i es basa en la informació sobre les persones que utilitzaran el producte [75]. Els processos de disseny centrats en els usuaris es focalitzen en els usuaris durant la planificació, el disseny i el desenvolupament d'un producte o sistema.

Hi ha un estàndard internacional que estableix una base comú per als mètodes del disseny centrat en l'usuari. Aquest estàndard és la norma ISO 13407: *Human-centred design process for interactive systems* [85], i proposa un procés genèric per incloure les activitats centrades en les persones mitjançant un cicle de vida de desenvolupament. Tot i que no recull tècniques o mètodes concrets, la ISO 13407 és una guia per incorporar disseny centrat en l'usuari en el cicle de vida del desenvolupament d'aplicacions interactives i així obtenir productes millors i més usables.

L'estàndard proposa quatre principis bàsics pel disseny centrat en les persones:

- Involucrar activament als usuaris i entendre els requisits dels usuaris i de les tasques a realitzar.
- Una correspondència adequada entre les funcionalitats que proporciona la tecnologia i els usuaris
- Iteració de solucions de disseny
- Disseny pluridisciplinar

El principal cicle de treball del procés està constituït per quatre activitats a realitzar de manera cíclica, tal i com es mostra a la Figura 2.4. Aquestes activitats són:

1. Especificació del context d'ús. Identificació de les persones que utilitzaran el producte, per a què el faran anar i sota quines condicions l'utilitzaran, és a dir, en quin context i on el faran servir.
2. Especificació de requeriments. Identificació de les necessitats i objectius dels usuaris, així com els requeriments organitzacionals i d'ús del producte.
3. Creació i desenvolupament de solucions de disseny. A partir de la informació recollida a les dues etapes anteriors es porten a terme els dissenys. Depenent de cada cas, aquesta activitat es pot descompondre en diferents subetapes.
4. Avaluació dels dissenys. Aquesta és una de les parts més importants del procés, on els dissenys realitzats s'avaluen tenint en compte les persones que els hauran d'utilitzar, els requeriments i el context d'ús. Si l'avaluació satisfà els requeriments el procés acaba, si no, torna a repetir el procés des de la primera etapa, refinant els resultats obtinguts.

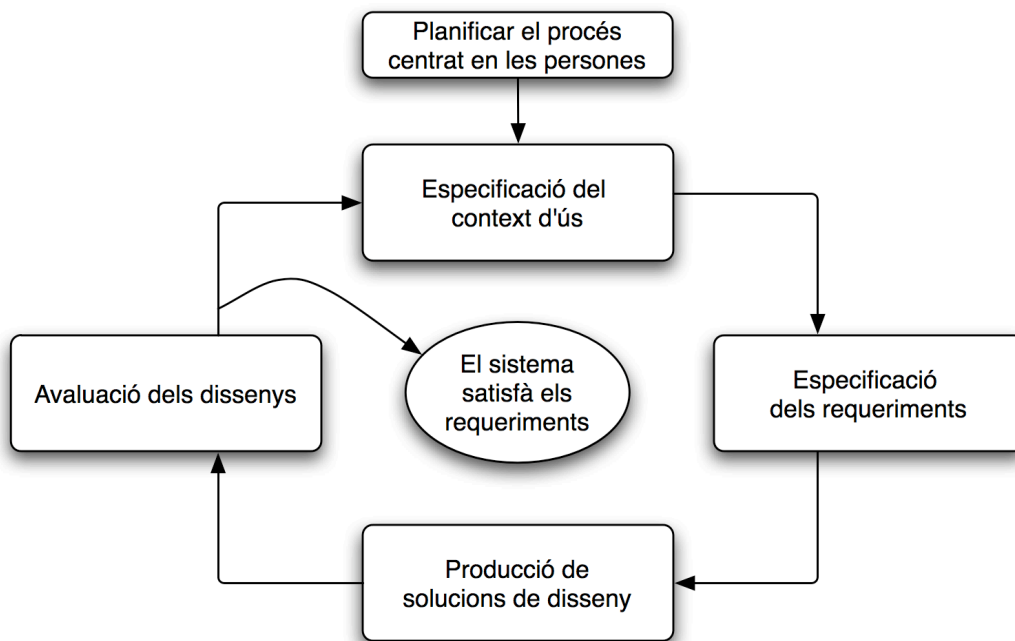


Figura 2.4: Cicle d'activitats de la ISO13407.

A cadascuna de les etapes de treball que proposa el disseny centrat en l'usuari es pot fer ús de diferents mètodes i tècniques per a aconseguir els objectius i generar els resultats

esperats. Així doncs, segons es defineixin i concretin els mètodes a emprar en cada una de les etapes del cicle de treball centrat en l'usuari, es poden tenir diferents models de procés. En la literatura relacionada amb el disseny centrat en l'usuari i la usabilitat es poden trobar diferents propostes de cicle o procés i les recomanacions per aplicar-lo segons siguin les especificitats de cada projecte. A més, diferents empreses, sobretot les relacionades amb el desenvolupament de programari, també suggereixen i posen a disposició el model de procés que utilitzen per executar projectes que desenvolupen productes seguint el paradigma del disseny centrat en l'usuari [10, 35, 153].

És important mesurar els beneficis que s'obtenen gràcies a la incorporació del disseny centrat en l'usuari en el desenvolupament d'un projecte. El disseny centrat en l'usuari aporta beneficis importants com són el desenvolupament més eficient del producte, donat que la usabilitat és més barata i fàcil d'incorporar si es fa a partir de les primeres etapes del desenvolupament; costos de suport i de documentació baixos, donat que els productes usables són més senzills de documentar i requereixen un nivell baix de suport o ajuda; equips de desenvolupament més productius i satisfets, donat que les tècniques del disseny centrat en l'usuari ofereixen solucions pels dilemes de disseny; i, el que és més important, usuaris més productius i satisfets [75]. Els productes usables permeten als usuaris portar a terme les seves tasques de manera més eficaç i eficient.

La experiència d'ús, de vegades també anomenada experiència d'usuari, és el terme que s'utilitza per descriure tot allò que un usuari percep, intueix, sent i experimenta quan interacciona amb un sistema. És un factor clau a l'hora de dissenyar i construir sistemes interactius i està directament relacionat amb la usabilitat, la facilitat d'ús i especialment amb el disseny centrat en l'usuari. El terme d'experiència d'usuari s'ha anat consolidant al llarg del temps i on ha tingut una especial repercussió ha estat en la indústria del programari. Constitueix un marc per parlar de totes aquelles àrees de coneixement que s'orienten a fer la vida més fàcil a les persones i, per tant, té una vinculació especial amb la usabilitat i el disseny centrat en l'usuari.

L'aproximació de Morville [123] a la experiència d'ús és interessant, i es realitza no tant des del desenvolupament d'aplicacions sinó des de l'organització i accés a la informació. Els principals elements de qualitat en la experiència d'usuari que l'autor

proposa són: utilitat, usabilitat, accessibilitat, credibilitat, que sigui desitjable, que sigui trobable (*findability*) i que aportí valor. Un element clau en la experiència d'usuari és l'aspecte afectiu i emocional, és a dir, tenir en compte les emocions que experimenten els usuaris a l'utilitzar un sistema i fer que els sistemes reconeixin, entenguin i expressin opinions [139]. L'experiència d'usuari pot treure profit de les emocions per millorar el disseny [130] i la interacció. El sistema "coneix" a l'usuari, en definitiva.

### 2.1.2 Mètodes de disseny centrat en l'usuari i d'avaluació de la usabilitat

Per cadascuna de les etapes del desenvolupament de sistemes i aplicacions, el disseny centrat en l'usuari proposa la utilització d'un conjunt de mètodes o metodologies per així assolir els objectius proposats. Els principals mètodes utilitzats a les diferents etapes es poden veure en forma de sumari a la Figura 2.5.

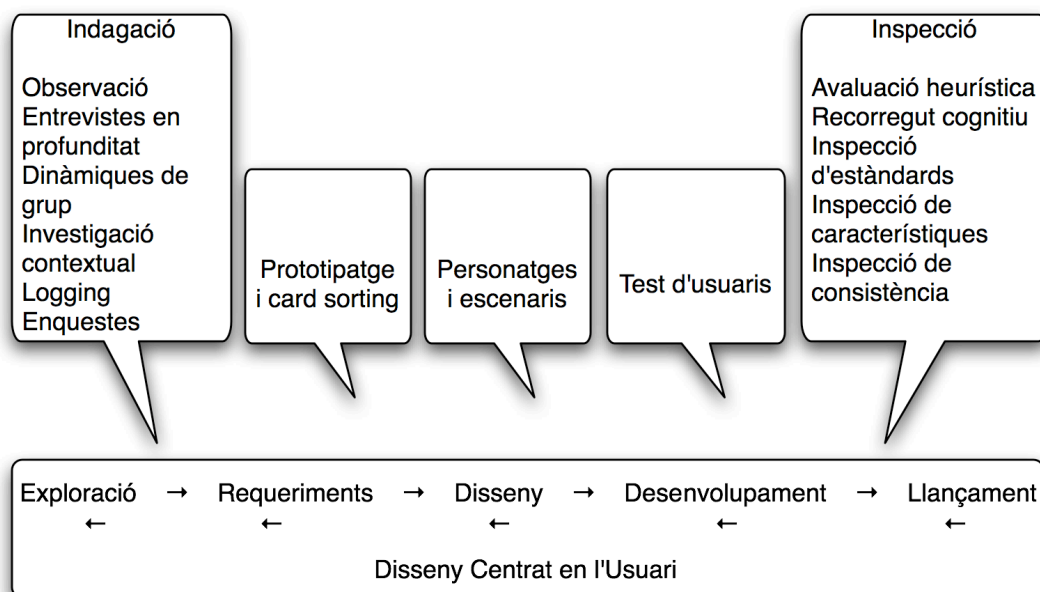


Figura 2.5: Principals mètodes de disseny centrat en l'usuari i d'avaluació de la usabilitat.

Com es pot veure, s'han classificat els diferents mètodes que s'utilitzen en el disseny centrat en l'usuari i la usabilitat en quatre grups. És important destacar que la

classificació de mètodes no és estricta i que hi ha mètodes que poden pertànyer a diferents grups o, dit d'una altra manera, hi ha mètodes que seguint objectius i resultats diferents pot ser útil de portar-los a terme en diferents etapes del procés de disseny centrat en l'usuari.

A continuació es presenten breument els principals mètodes, fent un especial èmfasi en aquells que presenten una rellevància especial amb el treball de tesi que es presenta. Les fonts d'informació consultades per descriure aquest mètodes són [60, 75, 102, 112, 128, 129].

## **A. Mètodes d'indagació**

Els mètodes d'usabilitat per indagació es porten a terme en les etapes d'especificació del context d'ús i dels requeriments. Es basen en involucrar als usuaris en les diferents activitats que es porten a terme i estan orientades a la obtenció d'informació. Aquest coneixement sobre els usuaris, les seves necessitats i objectius és imprescindible per un disseny centrat en l'usuari i per desenvolupar entorns usables. Les tècniques més utilitzades es presenten a continuació.

### **a) Observació**

Tal com indica el seu nom, aquesta tècnica consisteix en la observació dels usuaris en el seu entorn habitual o en un laboratori. És una de les tècniques més útils per conèixer de manera objectiva què fan els usuaris, cosa que pot servir tant per determinar els seus comportaments com per avaluar la usabilitat. L'estudi de camp és de gran utilitat abans de començar el desenvolupament i en les seves primeres etapes. Un lloc habitual on es porten a terme observacions d'usuaris és un Internet Cafè, per exemple. L'observació en aquest lloc pot proporcionar informació interessant en relació a com les persones utilitzen els ordinadors, com es connecten i naveguen per Internet i quins són els llocs web més visitats. A l'hora d'utilitzar la informació obtinguda cal tenir en compte l'entorn i el context d'ús on s'ha portat a terme la observació, i que molts cops s'observa sense conèixer els objectius reals dels usuaris. Aquest mètode permet conèixer el comportament i la interacció dels usuaris amb el sistema, en el seu entorn i

context d'ús habituals. Per altra banda, el seu principal inconvenient es deu al fet que no es produeix cap comunicació amb l'usuari, cosa que dificulta esbrinar les seves actituds i intencions.

#### b) Enquestes

Les enquestes es duen a terme entre una mostra representativa d'usuaris. Per a obtenir les dades es fan servir formularis estructurats amb preguntes que recullen totes les alternatives possibles de resposta. Aquest mètode proporciona informació quantitativa i permet obtenir informació de manera remota, sense que els usuaris hagin d'estar presents. Presenta l'avantatge que la informació que s'obté té significació estadística. Per altra banda, a l'igual que pot passar amb altres tècniques d'indagació com les entrevistes i els *focus groups*, les enquestes presenten l'inconvenient que el que els usuaris diuen que fan no sempre coincideix amb el que realment fan.

#### c) Entrevistes i *focus groups*

Les entrevistes i els *focus groups* permeten obtenir informació de tipus qualitatiu. Es duen a terme entre mostres petites, i a vegades també utilitzen formularis poc estructurats i oberts, i no es busca representativitat sinó processos i significat. No persegueixen la quantificació de les respostes sinó una comprensió en profunditat sobre les necessitats, preferències i experiències dels usuaris amb un producte o sistema. Al basar-se en la comunicació amb l'usuari, permeten recollir tot tipus d'informació sobre comportaments passats, actituds i intuïcions. No obstant, la informació recollida pot estar afectada per diferents biaixos i influències. Per tant, cal estar molt entrenat per portar a terme mètodes com aquest i poder obtenir coneixement rigorós.

Hi ha una modalitat concreta d'entrevista que s'anomena investigació contextual (en anglès, *contextual inquiry* [82] ) i que es recolza fortament en el concepte de context d'ús. Consisteix en realitzar la entrevista en el lloc on s'utilitzarà el sistema o lloc web i mentre s'està navegant i interaccionant per ell i, per tant, és molt similar a l'observació participant. Aquesta interacció de l'usuari proporciona la oportunitat d'adreçar qüestions de caire més específic que permetran entendre les seves dificultats i motivacions. La investigació contextual és un dels millors mètodes per descobrir i entendre el context dels usuaris i com el seu entorn influeix en les seves interaccions.

No obstant, presenta l'inconvenient que cal desplaçar-se al lloc d'interacció dels usuaris i que requereix una quantitat de temps considerable.

#### d) *Logging*

Les tècniques de *logging* es basen en la monitorització de l'activitat dels usuaris i consisteixen en recollir i analitzar les dades del registre de la seva activitat en un sistema o lloc web. Normalment, la recollida i processat de les dades es porta a terme mitjançant un sistema automatitzat, que passa desapercebut a l'usuari. Molts cops s'aprofiten els propis fitxers de registre d'activitat del sistema. Tradicionalment, aquests fitxers tenen l'extensió log, i d'aquí agafen el nom aquestes tècniques.

En els llocs web, el processament i anàlisi dels fitxers de log o de registre d'activitat permet conèixer el comportament dels usuaris, les funcionalitats que utilitzen, quant temps destinen a visualitzar pàgines, què dura la visita al lloc web i quin és el seu itinerari de navegació. Aquesta informació és molt útil i pot ser utilitzada per modificar i millorar els dissenys i la estructura del lloc web. Per altra banda, tot i conèixer les pàgines que visiten i les navegacions dels usuaris, des del punt de vista de l'avaluació de la usabilitat aquest mètode presenta l'inconvenient que es desconeix quins són els objectius dels usuaris. És, de fet, un mètode observacional. Sense conèixer els seus objectius, no es pot saber si accedeixen a una pàgina per error o perquè així ho desitgen, així com també es desconeix si un determinat itinerari de navegació s'ha realitzat volgutament o perquè l'usuari s'ha perdut. Tot i així, la informació que s'obté amb el *logging* és de gran utilitat ja que mostra l'activitat real dels usuaris en el lloc web.

El principal avantatge que ofereix el *logging* és que amb un cost relativament baix, proporciona informació detallada de les accions i navegacions que realment fan els usuaris. És una tècnica molt indicada per conèixer el què, però no proporciona informació del perquè. Així doncs, s'ha de combinar amb d'altres tècniques d'indagació que proporcionin informació dels objectius dels usuaris i les seves motivacions.

## **B. Prototipatge, card sorting, personatges, escenaris i test d'usuaris**

El prototipatge i el *card sorting* són tècniques indicades per projectes que presenten certa complexitat tant pel disseny de la interfície d'usuari del lloc web com per determinar la seva estructura i terminologia. El *card sorting* i algunes modalitats de prototipatge també s'anomenen mètodes de disseny participatiu. La creació de personatges i escenaris constitueix una tècnica molt útil per situar a l'usuari en el centre del procés de disseny. Per la seva banda, els anomenats tests d'usuaris són proves que es realitzen amb els dissenys, ja siguin prototipus o interfícies d'usuari acabades.

### a) Prototipatge

El prototipatge consisteix en la construcció d'un o diferents models del sistema o lloc web que s'està dissenyant. Aquests models simulen o ja tenen construïdes parts del sistema final i s'utilitzen per portar a terme proves que no es podrien realitzar fins que el projecte estigués completament finalitzat. Aquests models no pretenen reproduir el lloc web o el sistema en la seva totalitat sinó que normalment corresponen a parts concretes que presenten una especial complexitat. D'aquesta manera es pot involucrar directament als usuaris en el disseny de determinats elements del sistema com, per exemple, l'aula virtual d'un entorn virtual d'aprenentatge. Normalment, les tècniques de prototipatge són portades a terme per experts en usabilitat amb la col·laboració de l'equip de desenvolupament. Les modalitats d'aquesta tècnica que s'utilitzen amb més freqüència són el prototipatge de baixa fidelitat, alta fidelitat, vertical i horitzontal [152].

Els prototipus de baixa fidelitat modelen elements generals del sistema, sense baixar al detall. Es poden construir utilitzant paper i llapis i consisteixen en una representació esquemàtica del sistema o del disseny de la interfície. Així doncs, no inclouen els aspectes de disseny gràfic de la interfície ni els aspectes funcionals de l'aplicació [161]. El seu objectiu és proporcionar una primera idea de com haurà de ser la interfície, de la disposició dels seus elements i de la visibilitat que han de tenir. Molts cops es demana que siguin els propis usuaris qui portin a terme els prototipus de baixa fidelitat i així obtenir informació de les seves preferències i necessitats. Per altra banda, el prototipatge d'alta fidelitat tracta de construir un model el més proper possible al sistema que s'està dissenyant i desenvolupant. Aquest prototipus s'utilitza per avaluar, de manera més



precisa, aspectes funcionals i d'usabilitat, tant per un expert en usabilitat com mitjançant un test d'usuaris. De la mateixa manera, es pot parlar de prototipatge vertical i horitzontal. El prototipatge vertical consisteix en l'elaboració d'un model d'una part concreta del lloc web amb l'aparença i funcionalitat que tindria si aquest estigués completament acabat. Això permet portar a terme mètodes d'avaluació de la usabilitat sobre una part específica o unes funcions determinades. El prototipatge horitzontal es basa en la construcció d'un model que abasta totes o casi totes les funcionalitats, espais i menús del lloc web, sense que estiguin implementades i funcionin realment. Serveix per avaluar no tant l'ús del sistema sinó el seu abast, la navegació i l'arquitectura de la informació.

El prototipatge, en les seves diferents modalitats, té l'avantatge que permet modificar i corregir els dissenys abans que es desenvolupin completament. És un mètode que permet involucrar fàcilment als usuaris i fer-los participar en el procés de disseny. No obstant, té el desavantatge que al treballar amb models o amb parts incompletes hi ha aspectes del sistema que no es poden avaluar, i que el desenvolupament d'alguns tipus de prototipus pot tenir un alt cost.

#### b) *Card sorting*

La tècnica del *card sorting* consisteix en la realització d'un exercici d'ordenació i classificació de conceptes. Es demana als usuaris que ordenin i categoritzin un seguit de cartes o targetes que representen conceptes. De la ordenació d'aquestes cartes s'obté una categorització que és molt adequada per a organitzar i estructurar la informació de manera natural pels usuaris.

La tècnica del *card sorting* es pot realitzar molt fàcilment amb cartes o cartolines de paper o mitjançant un programari específic. L'avantatge d'utilitzar un programari és que ell mateix pren nota de la ordenació de cada usuari i realitza els càlculs i el sumari dels resultats. El *card sorting* és una tècnica de gran utilitat pel disseny i construcció de llocs web donat que permet descobrir la organització i l'arquitectura de la informació més adient, tant per estructurar un menú de navegació com per entendre on els usuaris anirien a buscar la informació en un catàleg de productes. El principal avantatge del *card sorting* és que és una tècnica senzilla de portar a terme i té un cost baix. Per altra

banda, presenta el desavantatge que per a obtenir resultats representatius, es requereix un cert volum d'usuaris que realitzin l'exercici.

### c) Personatges i escenaris

Un personatge és la descripció d'un usuari arquetípic que pot servir com a guia en el procés de disseny. Es va desenvolupar com originalment com una tècnica d'ajuda al disseny, proposada per Alan Cooper [59] en la seva aproximació al desenvolupament de sistemes que tenen en compte a l'usuari, anomenada *Goal-directed design* [58]. Un personatge es construeix amb precisió i rigor a partir de la informació quantitativa i qualitativa de la investigació d'usuaris: segmentació, perfilatge d'usuaris, observació, enquestes, entrevistes i *focus groups*. Per tant primer cal investigar els usuaris, analitzar les dades obtingudes i finalment modelar els usuaris en personatges. Es considera, d'aquesta manera, que un personatge és un model de l'usuari.

El terme original en anglès és *persona*, i en l'actualitat s'utilitza conjuntament amb personatge, perfil o usuari representatiu. La tècnica s'inspira en el mètode que utilitzen els actors per desenvolupar un personatge per així realitzar una actuació més realista, d'acord al paper que han de representar. De la mateixa manera, els personatges s'utilitzen per a que els dissenyadors i els desenvolupadors tinguin en compte als usuaris en les diferents etapes del procés de disseny i desenvolupament i, per tant, serveixen de recordatori de per a qui s'està dissenyant i ajuden a tenir en compte als usuaris en tot el procés de disseny i desenvolupament.

Per la seva banda, un escenari és la descripció d'un personatge en una situació d'ús del sistema, amb uns objectius concrets [150]. Aquesta descripció inclou el context en el que té lloc l'acció i la seqüència d'accions que es realitzen. Les situacions concretes que es generen a partir d'aquests elements són de gran ajuda per explorar idees i considerar decisions de disseny.

Els escenaris són una eina que facilita fer hipòtesis sobre les situacions en les que trobaran els usuaris i les necessitats que tindran per portar a terme els seus objectius. No són per tant una forma de documentar les interaccions concretes d'un usuari sinó de posar de manifest el context d'ús i els objectius dels usuaris amb les seves motivacions.

Per construir els escenaris cal també un seguit d'informació que, en part, és diferent a la que s'utilitza per la creació de personatges. Aquesta informació inclou els objectius i motivacions dels usuaris, les seves necessitats d'informació, les accions que poden portar a terme i les funcionalitats necessàries per adreçar aquestes qüestions [47].

Els escenaris també s'utilitzen a l'enginyeria del programari, de manera més formal i específica, per modelar els requeriments i descriure les funcionalitats del sistema [167]. Des de la perspectiva de la interacció persona-ordinador i el disseny centrat en l'usuari, es tracta d'un recurs per ajudar a entendre a l'usuari i l'ús que farà del sistema [150] i no pas d'una descripció de les seves funcionalitats.

Els personatges i els escenaris són tècniques que apropen als usuaris i a les seves motivacions, objectius i situacions d'ús. Són tècniques que s'utilitzen principalment per orientar el disseny, tot i que també s'utilitzen per entendre i analitzar els usuaris i l'ús que fan dels sistemes interactius.

#### d) *Test d'usuaris*

Els mètodes d'avaluació de la usabilitat per comprovació o test permeten realitzar experiments amb els dissenys i portar terme proves i comprovacions sobre ells. A aquests mètodes i les seves diferents modalitats se'ls anomena test d'usuaris. Un test d'usuaris proporciona informació de caire més aviat informatiu que no pas quantitatiu o estadístic, així doncs cal donar més importància a la interpretació dels resultats que no pas a la elaboració de mètriques o estadístiques. Mitjançant el test d'usuaris es comprova si les interfícies i els processos estan ben dissenyats i si els usuaris hi troben obstacles i dificultats per assolir els seus objectius [128].

Els test d'usuaris poden variar des del senzill test amb una sola tasca i pocs usuaris a tests més complexos on s'avaluen un conjunt de tasques amb un volum important d'usuaris repartits en diferents perfils. A més, els tests poden ser molt diferents depenent de la seva tipologia i modalitat utilitzada. El protocol *thinking aloud* estableix el demanar l'usuari participant en el test que vagi dient en veu alta el que pensa i fa, d'aquesta forma el conductor del test obté més informació a banda de la observació de l'activitat del usuari. Hi ha altres protocols on en lloc de demanar que els usuaris

s'expressin en veu alta, el conductor del test els hi va formulant preguntes relacionades amb les tasques o la interfície. Normalment els test d'usuaris es porten a terme en una sala preparada per test que s'anomena laboratori d'usabilitat, on s'enregistra la pantalla on l'usuari interacciona amb la interfície i, de forma simultània, es filmen les seves expressions. D'aquesta manera experts en usabilitat poden analitzar el test sense necessitat d'estar-hi presents. Un dels principals desavantatges del test d'usuaris és el cost de portar als usuaris al laboratori, i que molts cops, els usuaris canvien el seu comportament només pel fet d'estar en un laboratori i realitzar un test. Una tècnica que evita aquesta situació és el test remot, que permet realitzar test d'usuaris sense haver de portar els usuaris al laboratori. L'avantatge del test remot és que la captació d'usuaris és més senzilla, l'usuari no perd el context d'ús habitual i per tant el test no s'esbiaixa en aquest sentit. L'inconvenient principal és que requereix d'una infraestructura tecnològica tant en el lloc web com a casa de l'usuari que pot ser complexa i que, depenent de com sigui, no permet portar a terme alguns protocols de test com el *thinking aloud*. El principal avantatge del test d'usuaris és que permet avaluar la usabilitat d'una interfície de manera força fiable i permet descobrir on els usuaris trobaran obstacles o dificultats.

### **C. Mètodes d'inspecció**

Els mètodes d'inspecció, també anomenats simplement mètodes d'avaluació de la usabilitat, consisteixen fonamentalment en uns avaluadors experts en usabilitat que estudien i examinen el sistema a avaluar [129]. A continuació es presenten els mètodes d'inspecció que es consideren més interessants per aquest treball.

#### **a) Avaluació heurística**

L'avaluació heurística consisteix en l'estudi i avaluació de la interfície per un o més experts en usabilitat d'acord a un conjunt de regles i principis d'usabilitat prèviament establerts. Aquests principis d'usabilitat en base als que s'avalua, s'anomenen heurístics.

Aquest mètode va ser desenvolupat originalment per Jakob Nielsen [126] i diferents autors han proposat conjunts d'heurístics per diferents tipus d'interfícies i sistemes interactius, com per exemple [112] i [156]. L'avaluació heurística pot portar-se a terme en diferents moments del procés de disseny centrat en l'usuari. Es pot realitzar durant el procés de disseny, un cop s'han elaborat les primeres propostes de disseny i un cop s'han construït els primers prototipus del sistema, ja que això permetrà corregir aspectes d'usabilitat abans de desenvolupar-lo completament. També es pot portar a terme en la etapa de llançament, un cop desenvolupat i així acabar de diagnosticar diferents elements d'usabilitat que no eren visibles en etapes prèvies.

L'avaluació heurística té com a avantatge que és un mètode raonablement ràpid de portar a terme i que, quan hi participa més d'un expert, proporciona diferents opinions tot seguint un mateix conjunt de regles. Per altra banda, com a desavantatges cal destacar que l'avaluació està especialment condicionada tant pel conjunt d'heurístics escollits com pels experts.

#### b) Recorregut cognitiu

En el recorregut cognitiu, en anglès *cognitive walkthrough* [[176], un expert construeix escenaris amb les tasques que l'usuari haurà de realitzar i les executa assumint el rol del propi usuari. Durant l'execució d'aquesta tècnica, s'analitzen les tasques que l'usuari ha de portar a terme i es simula el procés de resolució de problemes amb que es troba en cada etapa del procés d'interacció. D'aquesta manera es comprova si els objectius simulats i la càrrega cognitiva podran ser assumits de cara a que l'usuari realitzi una nova acció correcta. Aquest mètode és va desenvolupar per a aquelles interfícies i sistemes que es poden aprendre a utilitzar per exploració, tot i que també és útil en qualsevol tipus d'interfície.

D'entre els avantatges que presenta aquest mètode cal destacar que es pot portar a terme un cop dissenyat el sistema, inclús sense necessitat de prototipus. És relativament ràpid de portar a terme i es focalitza en l'usuari i en els aspectes que se sap que poden ser problemàtics. No obstant, pot ser un procés llarg i tediós que pot resultar molt esbiaixat segons l'expert que el porta a terme, els escenaris i les tasques escollides. També presenta l'inconvenient que no mostra tots els problemes d'usabilitat, i que posa a

l'expert o al dissenyador en el lloc de l'usuari, essent aquest darrer aspecte una de les errades més freqüents a l'hora de dissenyar [63].

#### c) Inspecció d'estàndards

En la inspecció d'estàndards, un expert en usabilitat examina si la interfície segueix les especificacions i estàndards definits per la indústria o per organitzacions internacionals. Les especificacions elaborades per la indústria poden ser, per exemple, els documents que defineixen la interfície gràfica i la interacció del sistema operatiu Microsoft Windows o Mac OS X. Els estàndards que defineix el W3C<sup>2</sup> afecten especialment a tots els desenvolupaments que es fan a la web, especialment els relacionats amb interactivitat i accessibilitat. Aquesta tècnica té com a avantatge que és relativament senzilla i ràpida de portar a terme, però, en canvi, requereix d'experts amb un ampli coneixement dels estàndards i de la seva aplicació. Tot i que aquest mètode és especialment important donada la rellevància que tenen els estàndards a la web, és un mètode que no posa directament de manifest els possibles problemes d'usabilitat.

#### d) Inspecció de característiques

A la inspecció de característiques, en anglès *feature inspection*, s'analitza un conjunt de propietats i característiques extretes a partir de la definició d'un escenari i de les seves tasques associades. Cada conjunt de característiques s'analitza i avalua tenint en compte la seva utilitat, disponibilitat i comprensibilitat. La inspecció de característiques fa especial èmfasi en aquells aspectes de la interfície que no resulten del tot naturals pels usuaris ja sigui perquè no hi estan acostumats o perquè requereixen coneixements més avançats. Presenta l'avantatge que l'anàlisi es focalitza en els aspectes que poden representar una dificultat per l'usuari. No obstant, l'anàlisi pot sortir esbiaixat segons l'escenari i tasques seleccionats i a més, l'obtenció del conjunt de característiques a analitzar pot resultar difícil.

#### e) Inspecció de consistència

A la inspecció de consistència, en anglès *consistency inspection*, s'avalua si un disseny està en concordança amb altres dissenys que també s'han de presentar a l'usuari. Un

---

<sup>2</sup> W3C és el World Wide Web Consortium ([www.w3c.org](http://www.w3c.org)) que és una organització internacional que té l'objectiu de definir els estàndards per la Web

expert en usabilitat analitza els diferents dissenys i interfícies d'usuari per comprovar que les interaccions i processos es realitzen de forma coherent i similar entre ells. És recomanable portar a terme aquesta avaluació en les primeres etapes del projecte i així evitar desenvolupaments que més tard s'hagin de corregir. Aquest mètode és molt recomanable per aquelles aplicacions o llocs web amb una mida mig i gran, a més l'anàlisi és ràpid i fiable, especialment si es fa seguint un conjunt de guies o pautes prèviament establerts. Entre els seus desavantatges cal destacar que si l'anàlisi revela inconsistències entre dissenys, no és senzill posar d'acord a diferents equips de disseny i desenvolupament i pactar un disseny i interacció unificats i coherents.

Els mètodes d'avaluació de la usabilitat per inspecció requereixen d'un o més experts en usabilitat que portin a terme les avaluacions i tenen l'avantatge que es poden portar a terme un cop el disseny ja està fet, abans que el sistema estigui completament implementat. Per altra banda, presenten l'inconvenient que no tots ells involucren directament als usuaris i això molts cops resulta contraproductiu.

Tots aquests mètodes de disseny centrat en l'usuari resulten interessants i adequats en sistemes d'e-learning. A l'hora d'aplicar-los, caldrà tenir en compte les característiques i singularitats pròpies de l'e-learning i dels seus usuaris, els estudiants. En aquest treball de tesi tenen especial importància els mètodes d'indagació, especialment el *logging* i la seva capacitat de combinació amb mètodes de caire més qualitatiu com són les entrevistes, els focus groups, així com els test d'usuaris. Al capítol 3 es presenta amb més profunditat la proposta d'utilització d'aquests mètodes a l'àmbit del e-learning.

### **2.1.3 Model mental**

El model mental és un dels conceptes fonamentals de la interacció persona-ordinador. Es va introduir per primer cop l'any 1943 pel filòsof i psicòleg Kenneth Craik [61] que el va definir com petits models de la realitat construïts a la ment que s'utilitzen per raonar, desenvolupar explicacions i anticipar esdeveniments. Els models mentals, per tant, determinen la interacció de les persones amb els objectes quotidians que els

envolten. Així doncs, el concepte de model mental és de gran rellevància a la IPO donat que del model mental que cada persona es fa d'un sistema interactiu depèn que el pugui utilitzar amb eficàcia, eficiència i satisfacció.

Donald Norman va consolidar la importància del concepte de model mental a l'àmbit de la interacció persona-ordinador amb la publicació de *The Psychology of Everyday Things* [133] i una mica abans amb *Some observations on Mental Models* a [74]. A l'interaccionar amb l'entorn, amb altres persones i amb els objectes tecnològics, les persones construeixen models mentals d'ells mateixos i de les coses amb les que interaccionen. Aquests models els hi proporcionen ajuda predictiva i explicativa per entendre la interacció. Així, el model mental que un usuari es fa d'un sistema l'ajuda a comprendre'n el funcionament, predir els efectes de les seves accions i entendre els resultats.

Els models mentals es formen a partir de l'experiència, l'entrenament i l'aprenentatge. El model mental d'un dispositiu o sistema es forma, a més, a partir de la interpretació de la seva estructura visible. Norman anomena imatge del sistema a la part visible del sistema. Quan la imatge del sistema és incoherent o inapropiada, l'usuari no pot usar fàcilment el dispositiu, i si a més és incompleta o contradictòria, tindrà dificultats per utilitzar-lo.

Al desenvolupar un sistema, el seu dissenyador es basa en un model de com ha de ser aquell sistema i de com s'ha d'utilitzar. Aquest model s'anomena model conceptual i és un model o representació del sistema tant a nivell de estructura, com de funcionalitat i interfície. El model conceptual hauria de proporcionar una representació adequada del sistema, una representació que sigui exacta, consistent i completa. Conèixer el model conceptual d'un sistema constitueix el pas previ i necessari abans d'intentar entendre com les persones l'utilitzen i interaccionen amb ell, i també és un element fonamental per ensenyar a utilitzar els sistemes.



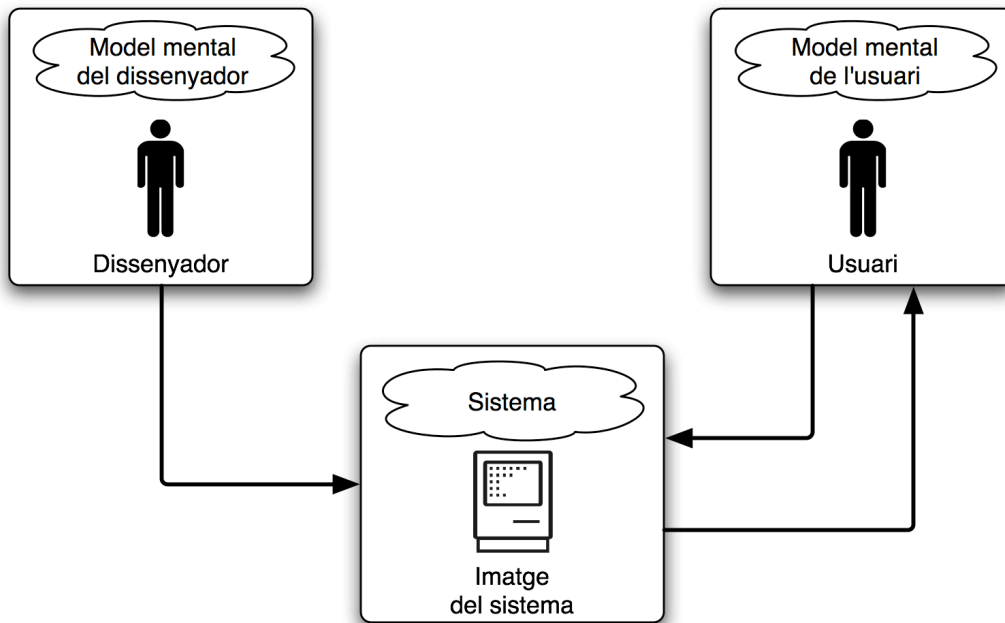


Figura 2.6: Relació entre el model mental del dissenyador, la imatge del sistema i el model mental de l'usuari.

Per tant, qui dissenya i construeix un sistema, intenta plasmar el model conceptual en el seu desenvolupament. L'usuari, a partir de la part visible o imatge del sistema, de la seva experiència prèvia i d'altres factors forma el seu model mental. La situació ideal es produiria en el cas d'existir una correspondència completa entre model mental de l'usuari i el model conceptual del dissenyador tot i que, malauradament, no sempre és el cas. A la Figura 2.6, adaptada de [133] es mostra aquesta situació i la relació entre el model conceptual, la imatge del sistema i el model mental. El problema per a que el model mental de l'usuari es correspongui amb el del dissenyador es posa de manifest quan el dissenyador no parla directament o no involucra a l'usuari. D'aquesta manera, el dissenyador només s'adreça a l'usuari mitjançant la imatge del sistema, de la que l'usuari en pot fer moltes i diferents interpretacions.

Relacionat amb el concepte de model mental, el mapatge, els permisos i les restriccions proporcionen ajuda per entendre com funcionen els sistemes i, per tant, per formar el model mental.

Mapatge (*mapping*) és el terme utilitzat per a definir la correspondència entre l'acció sobre els controls d'un sistema i els resultats que es produeixen. El mapatge és natural quan gràcies a les analogies amb el món físic o als estàndards culturals el resultat de l'acció sobre un control és predictable. Un dispositiu o sistema interactiu és fàcil d'utilitzar quan proporciona una bona visibilitat de les possibles accions i s'utilitzen mapatges naturals pels controls i els resultats.

Els permisos (*affordances*) són les propietats percebudes d'un objecte que determinen com pot ser utilitzat. Per exemple, una cadira permet seure, una porta permet obrir-la i tancar-la i un botó permet ser premut. Gràcies als permisos percebuts dels objectes, amb un cop d'ull les persones saben què han de fer, sense necessitar instruccions o informació addicional. Els objectes complexos poden necessitar explicacions però els simples no, i quan els objectes simples necessiten instruccions vol dir que el disseny ha fallat. Els permisos també juguen un paper molt important en el disseny de sistemes interactius donat que proporcionen informació als usuaris del que poden fer amb aquell sistema.

Com a complement als permisos, les restriccions (*constraints*) són tots aquells factors tant físics, semàntics, culturals com lògics que propicien accions adequades i n'eviten d'errònies. Es diu que les restriccions complementen els permisos ja que de la mateixa manera que els permisos permeten percebre com s'ha d'utilitzar, les restriccions permeten percebre com no s'ha d'utilitzar.

El model mental és de gran importància pels sistemes d'e-learning i per la personalització. Cal establir mecanismes per comprovar si el model mental dels estudiants en relació als entorns virtuals d'aprenentatge i els continguts educatius es correspon amb el model conceptual dels qui els han desenvolupat, és a dir, professors, dissenyadors instruccionals i tecnòlegs. Els sistemes de personalització són normalment complexos i per tant encara té més importància que el model mental dels usuaris del sistema s'ajusti al model conceptual de qui l'ha construït, entenent clarament que allò que se'ls hi presenta respon a les seves necessitats i especificitats.

### 2.1.4 Navegació i comportament d'usuaris

La navegació és un element principal de la interacció i experiència dels usuaris a la web. La major part del temps que els usuaris dediquen a la web el passen navegant, és a dir, buscant, localitzant i accedint continguts i serveis. La navegació presenta certs inconvenients per a les persones a l'hora de moure's per la gran quantitat d'informació disponible a la web. Aquests inconvenients són, al mateix temps, reptes pels dissenyadors i se'n destaquen dos, la sobrecàrrega cognitiva [103, 108] que es produeix quan els usuaris reben més informació de la que necessiten o poden gestionar, i la desorientació [56, 105] que es produeix quan els usuaris s'han de moure per espais amb molta informació, i que generalment són complexos i dinàmics.

La navegació és, per tant, aspecte essencial del coneixement dels usuaris. És important entendre com utilitzen els sistemes i llocs web, quines opcions accedeixen i utilitzen, com naveguen per accedir a funcionalitats i continguts [52]. Aquesta informació és especialment important en el context de l'e-learning, com es veurà més endavant. És necessari, doncs, dedicar temps i recursos a observar a les persones navegant i interaccionant amb els llocs web, en el seu context i entorn habitual. En aquest sentit, es poden tenir en compte diferents aspectes relacionats amb el comportament dels usuaris, com les estratègies de navegació, el seu grau d'implicació en la lectura de la informació que se'ls presenta o la manera en què prenen les decisions per portar a terme una acció.

Cal distingir la navegació que es realitza fins a arribar a un lloc web concret i la que es realitza dins d'un lloc web en particular. La primera es realitza a partir de buscadors i d'enllaços que es troben en planes web i la majoria de cops no hi ha manera de proporcionar un guiatge a l'usuari. La navegació que es realitza en un lloc web es més focalitzada i es pot proveir a l'usuari de guies i ajudes de navegació [125]. En ambdues situacions de navegació l'usuari es pot sentir frustrat, es pot desorientar i perdre quan tracta de trobar un lloc web i també a l'hora de trobar un contingut o servei específic en un lloc web concret.

Els usuaris orienten les seves decisions de navegació segons una o més estratègies concretes que, a més, poden anar canviant en el temps [46]. Molts cops els propis

usuaris no són conscients de que estan navegant i interaccionant tot seguint una estratègia relacionada amb els seus objectius i el model mental del lloc web. Tot i així, aquest és un aspecte important a tenir en compte. Habitualment, les diferents estratègies de navegació es classifiquen en dos grans grups, com estratègies de cerca o *searching*, i com estratègies d'exploració o *browsing* [109]. L'estratègia de *searching* és la utilitzada quan els usuaris busquen un determinat producte o informació en un lloc web. En aquest cas, els usuaris tenen un objectiu i saben bé el què volen i orienten la seva navegació a aquest objectiu. L'estratègia de *browsing* és la que utilitzen els usuaris quan simplement naveguen a veure què troben. Tot i que poden tenir un objectiu final, aquest objectiu no orienta necessàriament les seves interaccions i decisions de navegació.

L'aplicació o el lloc web hauria d'incorporar un conjunt de recursos i elements [125, 169] per tal d'assegurar que, sigui quina sigui l'estratègia que segueix cada persona en cada moment, no trobi dificultats en la navegació i la seva visita sigui òptima i satisfactòria. Això finalment es tradueix en elements de cerca i accés ràpid als continguts i funcionalitats per a aquelles persones que volen realitzar una acció o consultar una informació determinada, així com elements de navegació i exploració per a aquells usuaris que simplement naveguen comprovant si troben alguna cosa del seu interès.

El model mental que els usuaris es formen d'un lloc web és l'element principal de guia en la seva navegació. És per aquest motiu que és molt important que el lloc web ofereixi una molt bona visibilitat del model o imatge del sistema i que aquesta sigui clara i propera al model conceptual amb la que es va dissenyar el lloc web.

Són coneguts alguns elements claus en el comportament dels usuaris [101, 125], com per exemple que gairebé no llegeixen els textos, simplement els hi fan una ullada (també es diu que *escanegen* els textos); tot i que aquesta afirmació és una generalització i pot variar depenent dels usuaris, de la tipologia dels llocs web que visiten i de l'interès que els hi desperta. Molts usuaris accedeixen a Internet amb pressa ja que tenen la percepció que els ajuda a estalviar temps. També saben que normalment no cal llegir-ho tot, perquè molts cops hi ha molt de text complementari revestint les idees principals. A

més, cal tenir en compte que la qualitat i situació de les pantalles dels ordinadors no faciliten una lectura còmoda i relaxada.

Un altre aspecte relacionat amb el comportament dels usuaris és com prenen les diferents decisions necessàries per portar a terme una acció concreta. Molts cops els usuaris no exploren i avaluen amb deteniment totes les opcions possibles a l'hora d'interactuar sinó que seleccionen l'opció que consideren més raonable tot assumint un cert risc d'equivocar-se. Aquest fet ha de tenir-se en compte al dissenyar un lloc web que faciliti als usuaris desfer els possibles errors, fent passes enrere tant en la seva navegació com en els processos que hi portin a terme. En aquest sentit, l'ús del botó *back* o enrere en el navegador es converteix en una acció habitual a l'hora de visitar llocs web i que provoca que l'usuari tingui la percepció que sempre pot tornar enrere i desfer aquelles passes i accions que ha realitzat [101].

Un altre comportament interessant és que quan un usuari descobreix com portar a terme una tasca o acció determinada, habitualment ja no busca una manera millor de fer-la. Simplement, considera que allò funciona i li serveix per assolir els seus objectius i es limita a repetir-ho en ocasions posteriors. Aquest comportament es degut, en gran mesura, a no tenir un bon model del lloc web o sistema i que, per tant, el cost per a l'usuari de optimitzar les tasques que hi porta a terme és aparentment més gran que el portar-les a terme de la manera com se sent segur. Aquest fet és de gran importància en els entorns virtuals d'aprenentatge on els usuaris poden acabar dedicant més temps del necessari a realitzar les tasques de navegació per l'entorn que no pas les directament relacionades amb el seu procés d'aprenentatge [101, 125].

El concepte de model mental i de model conceptual són de gran importància en la personalització. Quan un sistema ofereix personalització, augmenta en complexitat donat que hi ha moltes més opcions disponibles que només es mostren i són visibles depenent de cada usuari i situació. És, per tant, important proporcionar una bona imatge del sistema per a que l'usuari es pugui formar un model mental consistent i adequat, que li sigui útil a l'hora de desenvolupar les tasques que ha de portar a terme i, d'aquesta manera, treure el major profit del sistema i les opcions de personalització. El disseny del

sistema ha de tenir en compte el perfil dels usuaris que l'utilitzaran i el comportament i estratègies que utilitzaran aquests usuaris.

En aquesta secció s'han presentat els aspectes de la interacció persona-ordinador més rellevants pel treball de tesi doctoral que es presenta. A continuació es presenten els conceptes d'e-learning i *technology enhanced learning* que, en gran part, també es fonamenta en la interacció de les persones amb la tecnologia.

## **2.2 E-Learning i *Technology Enhanced Learning***

En aquesta secció es presenta el concepte d'e-learning, fent un especial èmfasi en la tecnologia, que aplicada a l'aprenentatge (*Technology Enhanced Learning*) no només el millora sinó que produeix un canvi de paradigma.

L'e-learning és una nova tendència de l'educació que emergeix de la societat de la informació i fa referència a la interrelació de la tecnologia i l'aprenentatge, suposant una transformació cultural del que és i significa l'experiència en un entorn educatiu [148]. Es fa difícil trobar una definició d'e-learning àmpliament acceptada i consensuada; a més, moltes de les definicions es presenten esbiaixades o bé cap a la tecnologia o bé cap a l'educació. Per altra banda, es troben altres termes que s'utilitzen en lloc de e-learning. Aquests termes són, entre d'altres, *Web-Based Learning*, *Technology Enhanced Learning*, o *Distributed Learning*.

Així per exemple, la *American Society for Training & Development* [6] defineix e-learning com tot allò distribuït, presentat i facilitat mitjançant la tecnologia, amb la intenció de l'aprenentatge. Per la seva banda, Cisco Systems [7] defineix l'e-learning com l'aprenentatge basat en Internet que inclou la distribució i lliurament de continguts, el desenvolupament de cursos, i la gestió de l'aprenentatge i de la xarxa formada per la comunitat educativa. Engelbrecht [68] defineix l'e-learning com la instrucció que es distribueix mitjançant els mitjans digitals, ja sigui Internet o altres mitjans. Rosenberg [149] basa la seva definició en tres elements: la xarxa, la distribució i una nova mirada a la educació. L'e-learning és en xarxa i es distribueix, emmagatzema, comparteix i

actualitza gràcies a la xarxa; es distribueix als estudiants mitjançant un ordinador connectat a la xarxa i mira més enllà dels paradigmes tradicionals en la educació. Defineix l'e-learning com l'ús de tecnologies basades en Internet per proporcionar un ampli ventall de solucions que permetin l'adquisició de coneixements, habilitats i capacitats. Altres definicions situen la tecnologia en un segon terme i destaquen la idea d'aprenentatge flexible en contraposició a l'aprenentatge fixat. Els seus autors [55] destaquen les noves oportunitats i opcions que l'e-learning presenta als estudiants en quant a la flexibilitat en relació a la localització, el temps, els continguts, els requisits d'accés, l'aproximació instruccional o la distribució i accés als recursos d'aprenentatge. Altres autors [160] posen l'èmfasi en la connectivitat, definint l'e-learning com la tecnologia aplicada a la educació que aprofita els avantatges de la connectivitat que ofereix Internet i la web, així com altres xarxes com intranets o extranets [158]. Sigui com sigui, qualsevol iniciativa educativa etiquetada com e-learning hauria de contemplar el model pedagògic, el model tecnològic i el model organitzatiu, i amb aquesta visió global es quan es poden incloure aspectes que millorin la experiència dels estudiants, com és la personalització.

La utilització de Internet en la educació s'està convertint en un fet habitual i normal i per aquest motiu l'e-learning s'està convertint en sinònim de la utilització d'Internet amb tots els seus recursos, oberts o tancats, per a la distribució de continguts i per l'establiment d'entorns virtuals d'aprenentatge. La tecnologia aplicada a la educació proporciona la possibilitat de fer les mateixes coses amb nous mitjans, però el més important és que proporciona l'oportunitat de fer noves coses amb nous mitjans. Aquest és el motiu principal pel que la tecnologia aplicada a la educació produeix un canvi de paradigma [71] i es basa principalment en aprofitar les oportunitats que ofereix la tecnologia. La retroalimentació de la tecnologia que s'ha presentat a la secció anterior succeeix de forma important en l'e-learning. L'ús de entorns virtuals d'aprenentatge i de materials en suport digital fa que apareguin noves oportunitats d'interacció i aprenentatge, i aquestes possibilitats deriven en nous requeriments cap a la tecnologia que, al mateix temps, oferirà nous avanços tecnològics per l'aprenentatge que representaran noves oportunitats. El que és important no és la tecnologia de *per se*, sinó l'ús que fan les persones d'aquesta tecnologia.

L'e-learning constitueix un canvi de paradigma en la educació, proporcionant nous reptes a institucions, educadors i estudiants. Hi ha un canvi en com es veu l'educació i l'aprenentatge, en com té lloc i en com es distribueix. Gràcies a l'ús de la tecnologia es trenquen els requisits de lloc i temps, doncs l'aprenentatge no ha de seguir un ritme donat i ja no està lligat a un lloc establert en un moment fixat. Per a algunes persones també fa que l'accés a la educació sigui més econòmic donat que s'estalvien molts dels costos col·laterals com pot ser el temps i les despeses dels desplaçaments. Un altre canvi és que gràcies a la tecnologia es poden utilitzar nous i diferents recursos per a l'aprenentatge i es poden tenir en compte les necessitats individuals dels estudiants i adequar la seva formació segons diferents estils d'aprenentatge proporcionant més flexibilitat en el procés d'ensenyament. Un dels canvis més importants que suposa l'aplicació de la tecnologia a la educació és que canvia qui porta el control o condueix el procés, passant del professor a l'estudiant, que és qui decideix quan, com i on estudia. Això suposa un canvi de perspectiva del que era el procés d'ensenyament i aprenentatge fins ara. També suposa un recolzament definitiu a l'educació a distància, que deixa de ser un model educatiu secundari i esdevé un nou paradigma educatiu adaptat a les necessitats formatives de les persones, al llarg de la vida.

Des del punt de vista del món empresarial, l'e-learning és un bon exemple del canvi cap a la economia del coneixement [172] on els canvis tecnològics fan que els entorns professionals i els llocs de treball canviïn ràpidament i les competències dels treballadors necessitin una actualització contínua. Es parla del canvi de la era industrial a la era del coneixement [49] on la tecnologia canvia ràpidament, es redueixen els temps de desenvolupament de productes i per tant apareix una necessitat de formació contínua de les persones per a que puguin adaptar-se a aquests canvis. La innovació tecnològica està canviant contínuament, i a vegades de forma imperceptible, la manera en que el treball es realitza. Això fa que la formació i l'aprenentatge en el lloc de treball sigui necessari també de manera contínua, en el instant i lloc que es necessiti. Les organitzacions han de tenir en compte el canvi cultural que comporta la tecnologia i la economia del coneixement i utilitzar l'e-learning com un avantatge competitiu superant el parany de la resistència al canvi i afavorint un canvi cultural d'aprenentatge al llarg de la vida. Així, l'e-learning es presenta com una alternativa a la educació tradicional que presenta un gran valor per a les empreses i per a les persones.



Les institucions educatives estan especialment afectades pel canvi de paradigma de l'e-learning i han de saber aprofitar-lo i adaptar les seves estratègies a la nova societat del coneixement i a les oportunitats que presenta la tecnologia[34]. Per altra banda, amb la capacitat d'accés a la xarxa i les eines tecnològiques actuals és molt fàcil per a qualsevol agent distribuir una oferta formativa. Les institucions educatives, ara més que mai, esdevenen elements clau com autoritats acreditatives de la qualitat de les ofertes formatives en línia. És important tenir en compte que el simple ús d'Internet i les tecnologies no garanteix una oferta formativa de qualitat. Per tant, cada institució educativa ha d'establir la seva estratègia d'ensenyament i aprenentatge que fa ús de la tecnologia. El model pedagògic, la infraestructura tecnològica, l'organització interna de la institució, les noves capacitacions i innovacions que afavoreix la tecnologia són elements clau de la estratègia.

L'e-learning també representa un canvi i proposa nous reptes als educadors, és a dir, per a totes aquelles persones que treballen en la educació: professors, tutors, dissenyadors instruccionals, etc. És important que tots els actors implicats tinguin una actitud positiva i oberta envers la tecnologia. Els educadors han de tenir en compte que l'aprenentatge amb la tecnologia dóna el control als estudiants, i han de saber cedir aquest control i no caure en el parany d'utilitzar la tecnologia per continuar fent el mateix que venien fent fins ara [71]. Els estudiants decideixen com i on utilitzen els continguts educatius i quin ús fan de l'entorn virtual d'aprenentatge. A més tenen a la seva disposició tots els continguts i recursos disponibles i accessibles a Internet. És per això que els formadors esdevenen facilitadors [165] i, tot i que l'estudiant tingui el control del seu procés d'aprenentatge, l'acció docent del professor és fonamental donat que guiarà, proporcionarà criteri per a discriminar els recursos que li poden ser útils, donarà consell per treure el millor partit de l'entorn virtual d'aprenentatge i sobretot aportarà el necessari per a que cada estudiant pugui desenvolupar el seu procés d'aprenentatge de manera satisfactòria. Tony Bates [34] afirma que si la educació tradicionalment es fonamentava en l'adquisició i memorització de continguts, amb l'aparició d'Internet això canvia i altres habilitats com la cerca i selecció de la informació passen a un primer nivell. De la mateixa manera, el rol del professor esdevé obsolet i, al mateix temps, es multiplica, és el que A. Naeve anomena com a *Knowledge Manifold* [124].

Un factor clau de la educació en línia és la flexibilitat que ofereix als seus estudiants, donat que gràcies a la tecnologia es trenquen les barreres d'espai i temps. Tant la flexibilitat com la personalització són, cada cop més, factors clau en el món *online* en general i en l'àmbit de l'e-learning en particular. Més endavant en aquest mateix capítol es presenta què és i en què consisteix la personalització. Pel que fa referència a l'e-learning, la personalització en les seves diferents vessants és un tema d'estudi que s'ha treballat extensament al camp anomenat *Adaptive Hypermedia* [40, 42]. L'hipermèdia adaptatiu és l'àrea de recerca que s'emmarca entre l'hipertext i el *user modeling* i representa una alternativa als tradicionals sistemes web de "talla única" donat que tracten d'adaptar tant la interacció com els continguts als objectius, interessos i coneixement dels usuaris concrets, representats en els models d'usuari individuals. L'avaluació del sistema d'adaptació així com l'avaluació de la usabilitat i l'experiència d'ús dels sistemes adaptatius és un dels factors claus del seu èxit [67]. Un dels objectius d'aquesta tesi és establir el lligam entre els resultats obtinguts de l'anàlisi del comportament dels usuaris i les possibilitats d'adaptació de diferents escenaris educatius que es generen a un entorn virtual d'aprenentatge, identificant les opcions de personalització.

### **2.2.1 Models pedagògics, tecnològics i organitzatius**

La tecnologia aplicada a la educació té un gran nombre d'implicacions, factors i elements que la influeixen i que cal tenir en compte. Això fa que hi hagi una gran varietat d'aproximacions a l'e-learning que venen definides per diferents elements, que acostumen a estar establerts per la pròpia institució educativa com per exemple: la forma de distribució de continguts, les característiques de l'entorn virtual d'aprenentatge, la necessitat o no de sincronia, l'estil d'aprenentatge que es segueix, etc. Un element clau és el grau de virtualitat que assumeix la institució, que dependrà de la seva estratègia i de la tecnologia que utilitzi, podent anar des de la virtualitat com a complement o semi-virtualitat fins a la virtualitat completa. No forma part de l'abast d'aquest treball el fer un repàs exhaustiu de les diferents aproximacions a l'e-learning,

no obstant es descriurà el model de la UOC donat que constitueix l'àmbit d'estudi d'aquesta tesi doctoral.

La Universitat Oberta de Catalunya segueix un model de formació no presencial que ha optat per l'ús d'un entorn virtual d'aprenentatge, el campus virtual. La UOC s'ha decantat per desenvolupar un model centrat en l'estudiant, i presenta totes les característiques associades a la formació no presencial desenvolupada en entorns virtuals d'aprenentatge [165].

El model d'aprenentatge de la UOC i el funcionament general de la universitat s'han dissenyat tenint en compte la tipologia i les necessitats dels estudiants, persones que tenen responsabilitats professionals i familiars, amb poca disponibilitat de temps per a estudiar i amb una alta motivació per a estudiar. Aquestes i altres dades sobre els estudiants de la UOC es descriuen a l'annex A i l'annex C.

L'objectiu principal del model de la UOC és permetre a l'estudiant organitzar el seu propi procés d'aprenentatge d'acord amb les seves necessitats i interessos, adaptant el seu ritme d'estudi a la seva disponibilitat. D'aquesta manera es volen reforçar els aspectes positius de l'ensenyament a distància, és a dir, la no necessitat de coincidir en l'espai i el temps, i suavitzar els possibles aspectes negatius com poden ser estudiar sol i la necessitat de més autodisciplina i autoorganització. Per aconseguir-ho, el model educatiu de la universitat proporciona materials didàctics i continguts motivadors, suport a la resolució de dubtes, propostes de treball dels continguts, suport a la planificació personal, facilitant la interacció amb altres persones, especialment amb el personal docent i els altres estudiants.

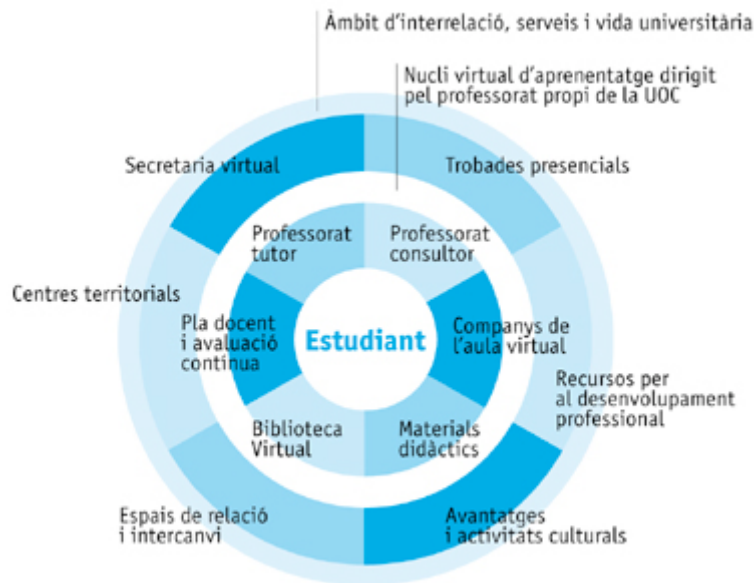


Figura 2.7: Elements del model de la UOC (Extret de [22]).

Els elements del model de la UOC s'estructuren al voltant de l'estudiant i s'organitzen en el que la pròpia universitat anomena “nucli virtual d'aprenentatge” [22] i es mostra a la Figura 2.7. Els elements principals del model són els següents [22, 171]:

- L'estudiant. És la persona que aprèn i, per aquest motiu, el centre del model que s'orienta a proporcionar atenció personalitzada i un estímul i atenció constants en el temps per així facilitar els processos d'ensenyament i aprenentatge.
- El personal docent. L'equip docent de la UOC està estructurat en professors, consultors i tutors. Els professors tenen dedicació exclusiva a la universitat i són els responsables de la oferta formativa i vetllen per la qualitat dels cursos i assignatures. Els consultors i tutors són col·laboradors docents que proporcionen als estudiants atenció docent i atenció acadèmica, són experts en el seu àmbit de coneixement i procedeixen tant del món professional com de l'universitari. Els consultors proporcionen atenció docent i actuen com a facilitadors de l'aprenentatge en el marc d'una assignatura. La seva actuació es desenvolupa en l'aula virtual, a partir del pla docent i dels materials didàctics i la seva tasca serveix d'estímul i guia als estudiants, fent que el procés d'ensenyament s'ajusti

als diferents ritmes i necessitats dels estudiants. Els tutors proporcionen atenció acadèmica i, d'aquesta manera, s'encarreguen d'orientar i assessorar als estudiants en relació al seu itinerari curricular, el coneixement de la universitat amb la seva normativa, els drets i deures dels estudiants i el calendari acadèmic, entre d'altres aspectes.

- El pla docent. És l'instrument que estructura i ordena els diferents elements de cada assignatura. En el pla docent es defineixen els objectius d'aprenentatge i les competències a assolir, es presenten els materials didàctics que cal estudiar i treballar per assolir els objectius, i es descriuen les diferents activitats a realitzar, així com model l'avaluació. El pla docent també inclou el calendari o planificació dissenyat per a cada assignatura que proposa una distribució temporal del treball, situant en cada moment del semestre els materials didàctics a estudiar i les activitats a realitzar.
- L'avaluació contínua. La UOC ofereix la possibilitat de seguir un sistema d'avaluació contínua que consisteix en realitzar durant el curs un conjunt d'activitats que són guiades i valorades pels consultors de cada assignatura. Aquest sistema facilita als estudiants aprofitar el curs i assolir els objectius proposats de forma gradual i progressiva, alhora que els permet fer un seguiment actiu del seu procés d'aprenentatge i mesurar el seu progrés.
- Els materials didàctics. Per a cada assignatura, l'estudiant disposa d'uns materials didàctics, en format paper o digital, segons els objectius d'aprenentatge de cada assignatura. Els materials didàctics i els recursos d'aprenentatge de les assignatures proporcionen els continguts necessaris per a la obtenció dels coneixements i les competències previstes.
- La biblioteca digital. La biblioteca digital de la universitat posa al servei dels estudiants les eines, serveis bibliogràfics i els recursos d'informació que els ajudaran en el procés d'aprenentatge i l'assoliment dels seus objectius.

- Els serveis de comunitat universitària. L'entorn virtual d'aprenentatge de la UOC permet, a més de desenvolupar les activitats relacionades amb l'aprenentatge, crear una comunitat universitària virtual on és possible l'intercanvi personal i el treball cooperatiu. Posa al servei de l'estudiant l'espai que permet establir relacions personals o professionals, crear grups d'interès o accedir a informacions i avantatges diversos, tant a nivell d'oci com a nivell professional.
- El campus virtual. És l'entorn virtual d'aprenentatge propi de la UOC. Està construït sobre diferents tecnologies que utilitzen una única interfície web comú, d'aquesta manera, és accessible des de qualsevol lloc amb connexió a Internet, amb un navegador web. El campus virtual proporciona accés als elements i eines d'ensenyament i aprenentatge i, al mateix temps, fa possible la interacció entre estudiants, l'equip docent i l'equip de gestió de la universitat, sense necessitat de coincidir ni en l'espai ni en el temps.

Tots aquests elements estan interrelacionats, de manera que constitueixen un sistema integral que proporciona als estudiants el suport necessari en un entorn no presencial. Aquests elements, al mateix temps, reflecteixen el model educatiu, tecnològic i organitzatiu establert per la universitat. El model de la UOC constitueix un exemple d'aproximació l'e-learning molt interessant on el model educatiu completament virtual, l'estratègia de la institució, la organització de l'equip docent i de gestió, i la tecnologia utilitzada en el campus virtual estan orientats a un mateix objectiu, proporcionar als estudiants una experiència educativa satisfactòria.

A l'hora d'establir el model tecnològic d'una iniciativa d'e-learning és important recordar que la tecnologia es fonamenta en els ordinadors i en les xarxes d'ordinadors tant per posar en contacte estudiants i professors com per presentar continguts educatius. D'aquesta manera, l'interacció dels estudiants amb els seus companys i amb els professors es veurà influenciada per la interfície i el disseny de l'entorn virtual d'aprenentatge. Per la seva banda, els continguts educatius desenvolupats sobre mitjans digitals poden presentar diferents graus d'interactivitat que també es veurà afectada pel seu disseny. És per això que la interacció persona-ordinador és clau en l'e-learning

donat que estudia els diferents aspectes de la interacció de les persones amb la tecnologia. Com s'ha descrit a la secció 2.1, l'àmbit de treball de la IPO són per una banda els usuaris i per l'altra el disseny, implementació i avaluació de sistemes interactius, i això és el que són els continguts educatius i els entorns virtuals d'aprenentatge, sistemes interactius utilitzats per estudiants i professors, amb una finalitat formativa. El fet que hi hagin diferents models o aproximacions a l'e-learning provoca que hi hagi diferents maneres d'interacció. Un cop més, aquestes interaccions han d'estudiar-se per així dissenyar adequadament els sistemes d'e-learning. El disseny centrat en l'usuari aporta el coneixement necessari per a que els sistemes d'e-learning, entorns i continguts es dissenyin tenint en compte les necessitats dels seus usuaris, els estudiants.

### **2.2.2 Sistemes d'e-learning**

Amb el terme "sistemes d'e-learning" s'inclou en aquest treball tant els continguts i recursos educatius basats en mitjans digitals com els entorns virtuals d'aprenentatge. És a dir, s'utilitza com a terme general per fer referència tant a continguts educatius com a plataformes.

Des d'un punt de vista més formal, els sistemes d'e-learning es poden conceptualitzar en un model format per tres components bàsics: els usuaris, el sistema de gestió i interfície i el curs o cursos formats a partir dels recursos docents disponibles [48, 81]. Els usuaris són principalment els estudiants, tot i que també s'inclou com a agents clau als professors i als administradors del sistema. El sistema de gestió es recolza en bases de dades amb informació sobre els estudiants i els cursos disponibles. La interfície permet la interacció dels usuaris amb el sistema gestor i presenta la informació adequada depenent de cada tipus d'usuari. Els cursos estan formats principalment per continguts i recursos educatius que, al mateix temps, estan estructurats en objectes d'aprenentatge. El diagrama amb aquests tres components o dimensions es presenta a continuació, a la Figura 2.8.

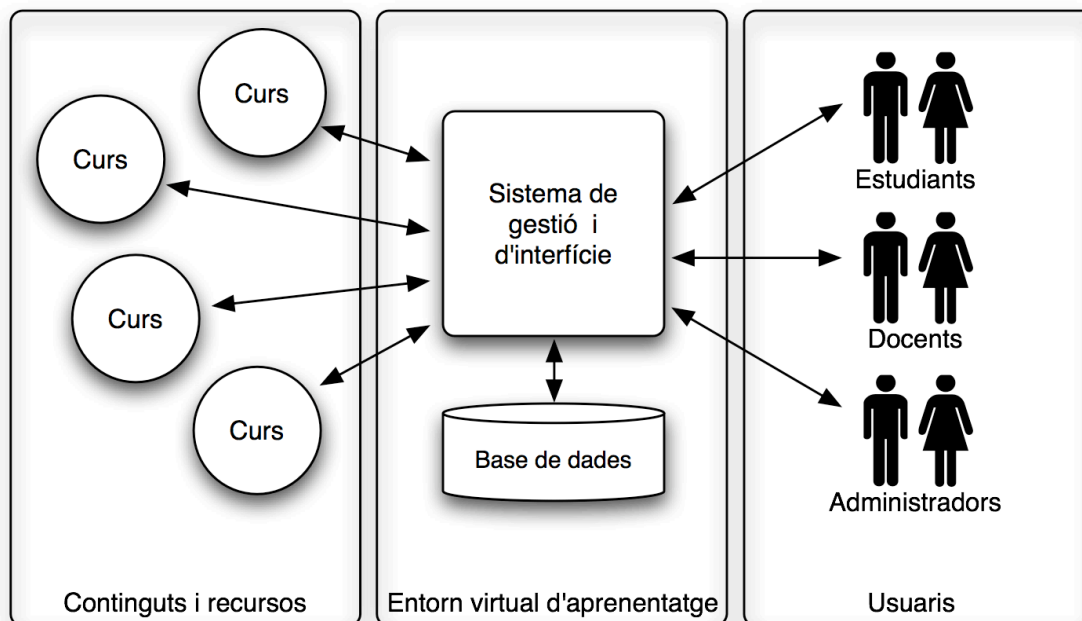


Figura 2.8: Dimensions dels sistemes d'e-learning.

Tenint en compte aquestes tres dimensions, el funcionament és el següent: els estudiants, professors i administradors accedeixen al sistema de gestió mitjançant la interfície. El sistema de gestió accedeix als cursos i consulta la base de dades per proporcionar i visualitzar informació pels diferents usuaris, segons el seu tipus i estat. El sistema de gestió recull les dades generades pels diferents usuaris i les envia a la base de dades on queden emmagatzemades. La recollida de dades generades pels estudiants en cada curs conforma el que s'anomena monitorització o seguiment de l'estudiant i registra dades sobre els continguts accedits, exercicis realitzats, temps utilitzat, etc.

Les tres dimensions són els elements bàsics i necessaris d'un sistema d'e-learning i estan afectades directament, tot i que de diferents maneres, per la tecnologia. Els components de gestió, interfície i base de dades constitueixen els anomenats Entorns Virtuals d'Aprenentatge (EVA) tot i que també reben altres noms com Plataforma Educativa i també *Learning Management System* (LMS) [81]. Els EVA inclouen també eines relacionades amb la comunicació entre els usuaris de l'entorn, així com eines de col·laboració. Cada cop més els LMS es basen en tecnologies web, d'aquesta manera es facilita que els estudiants puguin accedir a l'entorn independentment del lloc, del moment i del seu punt de treball. Depenent de la institució educativa i del seu grau de



virtualització, l'entorn virtual d'aprenentatge pot arribar a constituir un campus virtual complet, oferint altres serveis com són secretaria, biblioteca o serveis de comunitat educativa. Altres objectius dels LMS són centralitzar i simplificar la gestió i administració de la oferta formativa i de les persones.

Molts cops els EVA es focalitzen en el sistema de gestió i la base de dades, deixant en un segon terme la interfície. Això és un error ja que mitjançant la interfície és com estudiants i professors interaccionen i es comuniquen entre ells, i és com accedeixen als continguts. El disseny de la interfície influeix en la experiència educativa i juga un paper important en la personalització del procés d'aprenentatge.

El curs o cursos constitueixen l'altra dimensió tecnològica dels sistemes d'e-learning, i el seu element fonamental són els continguts educatius. El sistema que gestiona els continguts rep del nom de *Content Management System* (CMS) i a vegades està integrat en el mateix LMS, donant lloc al que es coneix com LCMS. El seu objectiu és permetre la creació, emmagatzemament, gestió i manteniment dels continguts educatius. Els LCMS involucren a autors, dissenyadors (gràfics, instruccionals, d'interfícies i d'interacció), tecnòlegs i administradors. Diferents autors, estàndards i especificacions recomanen que els continguts s'estructurin mitjançant *Learning Objects* o Objectes d'Aprenentatge [177]. Els objectes d'aprenentatge són continguts digitals marcats i etiquetats de manera sistemàtica i estàndard, per exemple d'acord amb un estàndard com pot ser el LOM [83]. La interoperabilitat i la reutilització són dos dels objectius principals que es persegueix amb l'ús d'objectes d'aprenentatge. Cada objecte d'aprenentatge cobreix un o més objectius concrets d'aprenentatge, de forma que la construcció de cursos més o menys complexos es pot realitzar a partir de la combinació de diferents objectes d'aprenentatge. L'ús d'estàndards per a l'estructuració del contingut dels objectes d'aprenentatge és, per tant, una necessitat per a assegurar una correcta connexió i compatibilitat entre diferents objectes i per poder-los utilitzar en diferents plataformes o entorns educatius. Així doncs, si es fragmenta el procés d'aprenentatge i s'estructura amb objectes d'aprenentatge, aquest permetrà una gran riquesa en forma d'un elevat nombre de combinacions que poden generar-se a partir de tots els objectes d'aprenentatge que constitueixen un curs complet. Això permet generar itineraris formatius de manera que l'aprenentatge es pot flexibilitzar i adaptar a les

necessitats dels estudiants [116, 121]. Tot i així, la reutilització dels objectes d'aprenentatge presenta un seguit de qüestions i aspectes a treballar [141, 174].

Els usuaris constitueixen la tercera dimensió a considerar en els sistemes d'e-learning: estudiants, professors i gestors. A l'hora de dissenyar el sistema d'e-learning, les persones que l'utilitzaran són un element clau. El coneixement sobre els usuaris, saber qui són, des d'on i com es connecten, el seu coneixement previ, les seves competències tecnològiques, etc. és una informació clau per dissenyar un entorn virtual d'aprenentatge i uns continguts educatius que siguin útils i afavoreixin l'aprenentatge. El disseny centrat en l'usuari proporciona mètodes per conèixer als usuaris i el seu context, i també ofereix una perspectiva de disseny que els té en compte en les diferents etapes. Per la seva banda, la usabilitat proporciona el coneixement per avaluar la eficàcia, eficiència i satisfacció dels estudiants, professors i gestors utilitzant els entorns i continguts d'e-learning.

Per tant, en un sistema ideal d'e-learning, és important poder disposar de descripcions formals dels seus components, per assegurar una major interoperabilitat i riquesa semàntica que permeti adaptar el sistema a les necessitats reals dels usuaris. D'aquesta manera, la utilització d'estàndards i especificacions esdevé clau.

### **2.2.3 Estàndards d'e-learning**

Els estàndards tenen una gran importància en la tecnologia i per aquest motiu han esdevingut de gran importància a l'e-learning. Tant els continguts com les plataformes educatives, al tractar-se de sistemes tecnològics, es beneficien fortament de l'ús d'estàndards. Els estàndards són, segons la ISO [14], acords documentats que contenen les especificacions tècniques o altres criteris precisos per a ser utilitzats com a regles, pautes, o definicions de característiques per a assegurar que les matèries, els productes, els processos i els serveis siguin convenients per al seu propòsit. Es poden diferenciar dues tipologies d'estàndards, els estàndards acreditats o *de iure* i els que s'anomenen estàndards *de facto*. Els estàndards *de iure* són aquells que han estat aprovats i acreditats per una organització d'estàndards com poden ser l'IEEE [11] o la ISO. Per altra banda,

els estàndards *de facto*, són els que tot i que no han rebut l'aprovació i acreditació de cap organització, són utilitzats extensivament en el mercat. L'adopció d'estàndards afavoreix el desplegament i progrés de les àrees on la tecnologia juga un paper important, com per exemple ha passat amb Internet i els estàndards definits pel W3C. És interessant destacar la diferència entre estàndard i especificació donat que es tracta d'un aspecte que genera confusió, ja que molts estàndards, abans de ser-ho, han estat especificacions. Una especificació és també un acord documentat però que encara no han estat aprovats o acreditats per una organització d'estàndards.

Per portar a terme una proposta d'e-learning cal partir d'una aproximació interdisciplinària donat que es requereixen coneixements i tècniques de diferents caires: tecnològics, educatius i d'administració i gestió. Així doncs, per desenvolupar un sistema d'e-learning es necessària la intervenció de diferents equips que principalment són: docent, producció tecnològica, producció de continguts, administració tecnològica i gestió docent, sense oblidar als estudiants. En un entorn interdisciplinari, l'existència i ús d'estàndards aporta llum, claredat i perspectiva en els diferents aspectes a treballar i consensuar. Molts cops, l'estàndard esdevé un document de consens i el resultat de la reflexió d'un conjunt d'experts en les diferents disciplines implicades. Els estàndards tecnològics afavoreixen la compatibilitat i la usabilitat dels productes d'e-learning.

Els estàndards educatius tenen els seus orígens allà cap als anys vuitanta amb les primeres aplicacions de la tecnologia a l'aprenentatge, el que es va anomenar *Computer Based Instruction* [30], però és a finals de la dècada dels anys noranta quan es comencen a definir els estàndards que actualment estan en ús i que aporten valor a les plataformes d'e-learning. La primera especificació per a un sistema d'e-learning va ser la CMI (*Computer Managed Instruction*) desenvolupada per la AICC (*Aviation Industry CBT Committee*). Les especificacions i estàndards educatius parteixen de diverses iniciatives i institucions i els que han obtingut més rellevància es presenten a continuació:

- El *World Wide Web Consortium* o W3C [24] és el principal organisme internacional pels estàndards a la web. Es va crear per assegurar que hi hagi compatibilitat i acord en el desenvolupament i adopció d'estàndards per la web. Tot i que no estableix

estàndards pròpiament educatius, el W3C ha jugat i juga un paper clau en la definició d'especificacions per a els continguts i la interactivitat a la web i per aquest motiu l'e-learning se'n veu afectat i beneficiat. Moltes tecnologies aplicades a l'aprenentatge es basen al mateix temps en tecnologies basades en les especificacions del W3C.

- *IMS (Instructional Management System) Global Learning Consortium* [12] és un consorci format per més de 200 institucions educatives, fabricants i organitzacions que proposa una arquitectura completa per l'aprenentatge en línia basada en l'interoperabilitat. Aquesta arquitectura incorpora interfícies per metadades independents de la plataforma, continguts agregats, serveis de gestió, perfils d'usuari i serveis externs. Les especificacions proposades per IMS afavoreixen la interoperabilitat tant a nivell de continguts com de plataformes i fan possible que sistemes diferents puguin funcionar plegats. També proposen especificacions per a fer seguiment del progrés dels estudiants, generar informes de la seva activitat i poder intercanviar la seva informació entre diferents sistemes administratius. Cal tenir en compte que les especificacions proposades pel IMS encara no han esdevingut estàndards però han tingut una gran acceptació a la indústria. És destacable *Learning Design* [98] per ser el primer estàndard orientat a la descripció del procés d'aprenentatge, incloent-hi aspectes de personalització.
- El *Learning Technology Standards Committee* del IEEE (IEEE LTSC) [16] és el comitè del IEEE que s'encarrega de desenvolupar els estàndards educatius acreditats. El IEEE LTSC és qui finalment transforma en estàndards *de iure* les especificacions proposades per altres consorcis com IMS o ADL. Els àmbits que treballa inclou metadades per objectes d'aprenentatge, perfils d'estudiants, seqüenciació de continguts, definició de competències, i localització i empaquetament de continguts. Un dels estàndards més importants definit pel IEEE LTSC és el *Learning Object Metadata (LOM)* [83]. LOM és un model de dades per descriure objectes d'aprenentatge i altres recursos digitals per l'aprenentatge amb l'objectiu d'afavorir la seva reusabilitat i interoperabilitat.

- *Advanced Distributed Learning Initiative* (ADL) [3] és una organització que estableix col·laboracions entre les institucions educatives, el govern i la indústria per així avançar en l'aprenentatge basat en la tecnologia. ADL va proposar *Sharable Content Object Reference Model* (SCORM) [26], un model de referència que defineix un model de continguts per l'aprenentatge basat en la web. SCORM incorpora les especificacions de IMS i permet la interoperabilitat entre sistemes de diferents fabricants. SCORM també defineix com s'han d'empaquetar continguts i defineix el concepte de seqüenciació, que fa referència a les regles que especifiquen l'ordre en com els continguts es presenten als estudiants.
- *Open Knowledge Initiative* (OKI) [17] proposa especificacions que descriuen com els components d'un entorn es poden comunicar entre ells i amb altres sistemes. Les especificacions OKI afavoreixen l'interoperabilitat i integració de sistemes i defineixen les bases per una arquitectura orientada al servei (*Service Oriented Architecture*, SOA). El model d'arquitectura proposat per OKI es basa en la separació, jerarquització i ocultació per capes i així aconseguir una integració senzilla i una major interoperabilitat. Un dels aspectes claus dels objectes d'aprenentatge és la interoperabilitat i per això les especificacions OKI són de gran importància a l'e-learning. OKI, juntament amb altres institucions ha promogut el projecte Sakai [20] que consisteix en el desenvolupament d'un entorn col·laboratiu i d'aprenentatge que fa servir les especificacions OKI i, per tant, ofereix un alt grau d'integració i interoperabilitat que fa possible utilitzar-lo i adaptar-lo per necessitats grans o petites i permet afegir o canviar components amb molta facilitat. Un aspecte important de Sakai és que utilitza una llicència de codi obert anomenada *Educational Community License.*, el que assegura les quatre llibertats del programari lliure: llibertat d'execució dels programes sense limitacions, llibertat d'adaptar els programes a les pròpies necessitats, llibertat de distribuir els programes sense restriccions, llibertat per a fer millores dels programes i publicar-les.

Hi ha altres organismes i institucions que promouen els estàndards en general i, per tant, l'e-learning també se'n veu beneficiat. La ISO, per exemple, promou estàndards com el MPEG-7 i el MPEG-21 amb grans implicacions al món de l'e-learning donat que

permeten un salt qualitatiu en els continguts basats en vídeo i multimèdia ja que permeten la descripció formal del seu contingut. També es interessant mencionar que hi ha diferents projectes tant a nivell europeu com mundial que intenten suplir la mancança d'estàndards entre els que es destaca el projecte ARIADNE. La *Alliance of Remote Instructional Authoring and Distribution Networks for Europe* (ARIADNE) [5] es va finançar mitjançant el IV Programa Marc de la Unió Europea i es focalitza en el desenvolupament d'eines i metodologies per produir, gestionar i reutilitzar objectes digitals d'aprenentatge.

Els estàndards d'e-learning permeten i ofereixen un conjunt d'avantatges i beneficis importants: llibertat d'elecció, escalabilitat del sistema, portabilitat dels materials didàctics, reutilització dels materials i la no dependència d'un únic proveïdor. És important reflexionar sobre la diferència dels sistemes d'e-learning basats en estàndards i els sistemes anomenats propietaris. Un sistema basat en estàndards permet la modularitat de les seves parts i que un entorn incorpori mòduls de diferents sistemes amb completa compatibilitat en l'intercanvi d'informació i protocols entre ells. Per tant, els estàndards garanteixen la escalabilitat del sistema, és a dir, que el sistema pugui estar compost de diferents mòduls independents que poden créixer amb les necessitats del sistema. Això permet modificar o actualitzar diferents parts del sistema sense haver de canviar-lo en la seva totalitat. En relació als continguts educatius i als materials didàctics, els estàndards permeten la portabilitat dels mateixos continguts entre diferents plataformes que utilitzin el mateix estàndard, i per tant la seva reutilització. A més, utilitzar estàndards en els sistemes d'e-learning significa que les seves tres dimensions estan ben descrites. La escalabilitat i la descripció semàntica d'aquests sistemes afavoreix la incorporació d'aspectes de personalització i adaptació del procés d'aprenentatge a les particularitats dels estudiants.

La tecnologia aplicada a l'aprenentatge es veu directament afectada per les interaccions i les interfícies. La interfície dels sistemes d'e-learning no és només una part instrumental de la plataforma sinó que cada cop més es revela com una part essencial del seu bon funcionament i de que els estudiants assoleixin els seus objectius d'aprenentatge. És per aquest motiu, i tenint en compte que també és un factor clau per a la personalització de continguts, serveis i itineraris formatius, que en aquest treball la

interacció persona-ordinador i especialment el disseny centrat en l'usuari i la usabilitat tenen una singular importància.

#### **2.2.4 E-Learning i Disseny Centrat en l'Usuari**

Els sistemes d'e-learning, entorns i continguts, al tractar-se de sistemes interactius, es important que es dissenyin seguint un disseny centrat en els seus usuaris. Cal tenir en compte les característiques dels estudiants i les seves habilitats interaccionant amb l'entorn virtual i els materials d'aprenentatge, i considerar els diferents atributs de la usabilitat per així proporcionar un aprenentatge eficaç, eficient i satisfactori, ajustat a les seves necessitats formatives.

Des del desenvolupament dels primers ordinadors personals, aquests s'han utilitzat en l'àmbit de la educació i la formació. Així, els primers usos de la tecnologia en la educació amb persones no expertes i habituades al seu ús va posar de manifest la necessitat de sistemes senzills, efectius i a l'abast de tothom. Això, com en moltes altres àrees on per primer cop es va utilitzar la tecnologia, coincideix amb el moment en que l'àrea de coneixement de la interacció persona-ordinador es desenvolupa com a tal i apareix la noció de disseny centrat en l'usuari [134].

Quan s'estudia la relació i implicacions de la usabilitat amb els sistemes d'e-learning és interessant partir de les tres dimensions dels sistemes d'e-learning descrites a la secció anterior, usuaris, entorns i continguts. La dimensió dels usuaris o estudiants inclou la identificació dels usuaris, les seves necessitats, preferències i característiques. La dimensió dels continguts i recursos inclou la identificació de guies, tècniques i requeriments de disseny i avaluació que s'han de seguir, i els aspectes relacionats amb la separació del contingut i la visualització dels materials didàctics. La dimensió de l'entorn virtual d'aprenentatge té en compte la identificació dels requeriments i funcionalitats de l'entorn educatiu, l'anàlisi de tasques i el disseny de la interacció. Alguns treballs mostren la importància de la usabilitat en el camp de l'e-learning [32, 166, 170].

Directament relacionat amb el concepte d'eficiència com a atribut de la usabilitat, és interessant prestar atenció al concepte de “facilitat d'esser après”, en anglès *learnability*. Aquest és un concepte que normalment fa referència a la rapidesa amb la que algú aprèn a utilitzar una aplicació o sistema. Així, *learnability* és el grau amb el que una interfície d'usuari es pot aprendre de manera eficaç i ràpida, però també fa referència a la eficàcia amb la que un objectiu d'e-learning és assolit [69]. La usabilitat fa referència a la funcionalitat i *learnability* fa referència al procés cognitiu, per tant quan menor esforç dedicat a la funcionalitat més gran podrà ser l'esforç dedicat a l'aprenentatge. Per tant, en el disseny i construcció de sistemes d'e-learning és important assolir tant els objectius d'usabilitat com els educatius. Els estudiants haurien de poder interactuar amb naturalitat amb els continguts i els entorns d'aprenentatge, i concentrar-se en l'assoliment de les seves competències i objectius d'aprenentatge. Així doncs, es necessari identificar solucions de disseny de sistemes d'e-learning que redueixin la frustració dels usuaris, incrementin la usabilitat, *learnability* i satisfacció. Tot sovint es presenten experiències d'e-learning on es pot observar que la educació en línia es una simple replicació electrònica dels continguts i mètodes tradicionals amb l'afegit de fer-ho amb interfícies dolentes i uns esquemes d'interacció rígids. No es pot oblidar que una interfície mal dissenyada constitueix un obstacle efectiu per a l'aprenentatge. Un dels reptes principals del disseny de l'e-learning és assegurar la seva usabilitat i proporcionar interfícies atractives i engrescadores que, a més, siguin pedagògicament adequades.

Quan els estudiants dediquen més temps entenent interfícies poc usables que treballant activitats o aprenent els continguts dels materials didàctics això distreu i retarda l'assoliment dels objectius d'aprenentatge. Les interfícies han de centrar-se en les necessitats dels estudiants i els seus objectius d'aprenentatge, proporcionant una idea clara de l'organització dels continguts i de les funcionalitats de l'entorn, una navegació senzilla i la possibilitat de personalitzar camins, processos i itineraris formatius. S'ha d'involucrar, sense que això resulti en un atabalament, a l'estudiant en les diferents parts del procés d'aprenentatge.

El disseny centrat en l'usuari proporciona mètodes per desenvolupar un sistema que respongui a les necessitats dels usuaris, que tinguin en compte els seus objectius i que,



per tant, sigui usable. La concreció del disseny centrat en l'usuari en el disseny centrat en l'estudiant [162], en anglès *learner-centered design* (LCD) és interessant donat que no només s'ha d'utilitzar un entorn i uns continguts per assolir uns objectius, sinó que cal tenir en compte que cada estudiant té unes estratègies d'aprenentatge diferents, una experiència prèvia diferent i també té motivacions per aprendre diferents als altres estudiants. Així, en el context de l'e-learning, a més dels mètodes i recomanacions del disseny centrat en l'usuari, descrites a la secció 2.1.1, es necessiten mètodes específics de disseny centrat en l'estudiant [144] per fer que els nous entorns virtuals d'aprenentatge i els continguts i recursos educatius siguin accessibles i utilitzables d'una manera educativament productiva.

El disseny centrat en l'estudiant persegueix l'objectiu de desenvolupar eines i sistemes que serveixin als estudiants per a adquirir nous coneixements, habilitats i competències. D'aquesta manera, es poden distingir quatre etapes d'alt nivell, com a qualsevol altre sistema interactiu: anàlisi d'usuaris i tasques, especificació de requeriments, disseny i avaluació. La clau està en utilitzar els mètodes adequats en cada etapa, tenint en compte la natura específica dels sistemes a desenvolupar. Diferents autors adrecen aquest aspecte [31, 114].

L'objectiu final dels sistemes d'e-learning és, per tant, proporcionar una experiència d'aprenentatge satisfactòria [62]. Donald Norman, en el seu llibre "*Things that make us smart*" [132] afirma que per a que un producte educatiu representi una experiència positiva pels estudiants, ha de:

- ser interactiu i proporcionar feedback,
- tenir objectius específics,
- motivar, comunicant una sensació continua de repte,
- proporcionar eines adequades,
- i evitar qualsevol factor de molèstia que interrompi l'aprenentatge

Els sistemes d'e-learning proporcionen nous reptes des del punt de vista pedagògic i de les estratègies d'aprenentatge, per així treure el màxim profit de les tecnologies, adaptant-les al perfil de l'estudiant. Així doncs, el repte és doble, els sistemes d'e-learning efectius han d'incloure, per una banda, funcionalitats educatives avançades, i

per altra banda, ho han de fer amb una interfície que no presenti complexitat als estudiants, proporcionant una interacció senzilla i estimulants l'interès dels estudiants.

### 2.3 Anàlisi d'ús de llocs web

L'anàlisi de l'ús de llocs web estudia com els usuaris visiten el web i com utilitzen els seus serveis. Es pot enfocar des de diferents punts de vista tenint en compte el mètode i els objectius que es persegueixin [163, 164]. L'anàlisi d'ús de llocs web es pot realitzar amb sistemes d'analítiques web, en anglès *web analytics*, també reben el nom de mètriques web, *e-metrics*, estadístiques web i *clickstream*, entre d'altres. Generalment inclou estadístics com poden ser la freqüència d'accessos a les pàgines, els camins de navegació pel web, la longitud dels camins, la durada de cada sessió de navegació, les pàgines d'entrada, les pàgines de sortida, etc.

Les dades d'ús de la web es poden obtenir de diferents fonts: servidors web, *cookies*, clients web, servidors *proxy*, bases de dades d'ús i també de bases de dades corporatives [164]. Això proporciona diferents aproximacions a les mètriques i analítiques web. La primera diferenciació té en compte on es produeix la generació i recollida de les dades. Principalment les dades es poden generar i recollir a la part de l'usuari o client o bé a la banda del servidor. La recollida de dades a la banda del client és la que pot proporcionar dades més acurades donat que es recull el que l'usuari realment fa quan interacciona amb el seu ordinador tot i que presenta el gran inconvenient que cal instal·lar un programari específic al seu ordinador. Això requereix un esforç de gestió del programari i de les dades que recull i una molèstia per a l'usuari que, a més, en moltes ocasions no està d'acord en la instal·lació de programari de tercers que directament monitoritzin la seva interacció. La recollida i anàlisi de dades d'ús a la vessant del servidor pot ser principalment de dos tipus: sistemes basats en l'anàlisi dels fitxers de log dels servidors web i els sistemes de mètriques remotes que es basen en un codi que s'introdueix a les pàgines del lloc web que es vol mesurar i que per a cada accés genera una entrada en una base de dades remota. Aquest treball s'orienta a la recollida i anàlisi de dades de navegació de manera transparent i no intrusiva per a l'usuari i per tant es focalitza en les

dades generades en els servidors. Tot i així es veurà en que consisteixen les mètriques remotes i els avantatges i inconvenients que presenten.

### 2.3.1 Fitxers de log

Els sistemes basats en l'anàlisi de logs consisteixen, a grans trets, en un programari que processa els fitxers de log i genera resultats intel·ligibles i interessants. Aquests sistemes reben com a entrada els fitxers de log generats i emmagatzemats pels servidors web, com per exemple Apache [4]. La sortida que generen és un informe d'ús del lloc web.

Els logs són una font d'informació rica i fiable de l'activitat d'un lloc web, del seu ús i del que fan els usuaris. La informació que proporcionen es bàsica, en el sentit que és informació sense processar i apareix tal com l'emmagatzema el servidor web. Aquest fet cal tenir-lo en compte a l'hora d'interpretar el resultat del seu processament i anàlisi. Els fitxers de log proporcionen informació sobre què fa l'usuari quan visita el lloc web, la navegació i les pàgines sol·licitades, però no donen informació de les seves intencions, objectius i necessitats. I el que és més important, no donen informació sobre si allò que han fet o visitat era el que volien fer o visitar. També proporcionen informació sobre el propi lloc web, les pàgines més i menys visitades i la seva estructura i navegabilitat.

Un dels handicaps que presenten els logs és que la web ha evolucionat i el format i la informació que recullen els fitxers de log no ho han fet. Per entendre bé les característiques dels fitxers de log cal tenir en compte el seu ús original als inicis de la web. En aquell moment, els llocs web eren principalment estàtics presentant continguts fixos organitzats jeràrquicament. Ara, el dinamisme és present a la web i els llocs web poden presentar tant contingut estàtic com funcionar com una sofisticada interfície d'un sistema informàtic [73]. Un senzill exemple de que els fitxers de log no han canviat és que en molts servidors el temps es continua mesurant amb resolució de segons mentre que gràcies a la capacitat del servidor, està resolent milers de peticions al mateix temps i queden totes registrades al mateix *timestamp*, a la mateixa hora, minut i segon. Un altre aspecte important relacionat amb el disseny original dels logs i el seu format és que aquest en gran part respon a les eines amb les quals s'havia de tractar. Els fitxers de log

són textuals i seqüencials i havien de ser processats amb les eines bàsiques de tractament de fitxers del sistema operatiu *Unix* com son les comandes *grep*, *count*, *awk*, etc. Avui en dia, els fitxers de log poden adquirir una grandària i volum que fa que aquestes eines no els puguin processar i que s'hagin de buscar mètodes alternatius per fer-ho.

### 2.3.2 Estructura dels fitxers de log

Un log és bàsicament el registre d'activitat que genera un servidor. Consisteix en un fitxer o un conjunt de fitxers que son generats i mantinguts pel propi servidor. D'aquesta manera, els logs de servidors web recullen l'activitat d'un servidor web, emmagatzemen l'històric de peticions i el resultat de cada petició. Les entrades s'emmagatzemen cronològicament i en fitxers independents per a cada dia.

Les entrades dels arxius de log es registren en un format estàndard que es coneix amb el nom de *Common Log Format* (CLF) [15] proposat pel World Wide Web Consortium (W3C) [23]. El W3C proposa un estàndard pel format i estructura dels fitxers de log amb l'objectiu d'afavorir la captura de dades i poder realitzar un anàlisi ric amb eines desenvolupades independentment que entenen l'estructura dels fitxers. Així, el principal avantatge de recollir els logs en format estàndard és que aquests poden ser processats per moltes i diverses eines d'anàlisi. Es pot donar el cas que es vulgui o es necessiti recollir més informació que la que defineix el CLF i així enriquir els resultats de l'anàlisi, llavors es pot seguir la especificació del *Extendend Log File Format* [76] també definit pel W3C. Un dels servidors web més utilitzats és el Apache HTTP Server [4] que és un programari de codi obert. Apache genera fitxers de log que segueixen el CLF i al mateix temps permet configurar les dades concretes que es registren a les entrades. Això permet recollir dades més específiques de cada lloc web en concret i també permet no registrar d'aquelles informacions confidencials o aquelles que embruten els fitxers de log i dificulten el seu processat i anàlisi. Per tant, tot i seguir el format estàndard, el format final del arxiu de log depèn del servidor que s'utilitzi i de la seva configuració, cosa que en moltes ocasions dificulta el seu posterior processament.

Els servidors web generen diferents arxius de log que registren diferents tipus d'activitat. Els més freqüents són el *Error Log* o log d'error i el *Access Log* o log d'accés. El log d'error recull les errades que s'han produït en el servidor en el seu funcionament normal de procés de peticions. Les entrades d'aquest fitxer constitueixen una font d'informació fonamental per solucionar els possibles problemes de funcionament del servidor. El log d'accés recull totes les transaccions o peticions processades pel servidor. Cada cop que un usuari, navegant pel lloc web, selecciona un enllaç i demana carregar una plana, es genera un conjunt de peticions que hauran de ser processades pel servidor. Aquestes peticions corresponen a tots els elements que s'han de carregar per mostrar una plana: arxius de text o html, fulls d'estil, gràfics i imatges. El servidor web genera una entrada en el fitxer de log d'accés per cada transacció. Per tant, cada cop que un usuari visita un lloc web, es genera un nombre limitat de transaccions que queden registrades en els arxius de log. A més de guardar un registre de l'activitat, l'objectiu principal del log d'accés és poder generar anàlisis de l'activitat i l'ús del lloc web, tot i que no ha estat usat massivament.

La informació que queda recollida als fitxers de log inclou principalment l'adreça IP del client, la data i hora de la petició, el tipus d'operació, la plana (o component) sol·licitat, el codi de retorn, el nombre de bytes servits, el navegador i el sistema operatiu del client i el *referer*, que és la plana o lloc web des de on s'ha originat la petició que s'està servint.. Una entrada del fitxer de log té el següent aspecte:

```
213.73.39.193 - - [23/May/2007:01:10:08 -0700] "GET /folder/page.html
HTTP/1.1" 304 - "http://www.uoc.edu/" "Mozilla/5.0 (Macintosh; U; Intel Mac OS
X; es-ES; rv:1.8.1.3) Gecko/20070309 Firefox/2.0.0.3"
```

Els camps que presenta són: adreça IP, identificador, hora i data, recurs i protocol, estat, grandària, *referer* i agent. Si es col·loquen tots els camps en una taula tal com es presenta a la Taula 2.1 es veu que té una estructura molt simple, donat que el disseny original preveia processar i analitzar els fitxers de log amb les eines pròpies del sistema operatiu.

IP	Id	Id	Data i hora	Recurs	Estat	Bytes	Referer	Agent
213.73.39.193	-	-	[23/May/2007:01:10:08 -0700]	"GET /folder/page.html HTTP/1.1"	304	-	"http://www.uoc.edu/"	"Mozilla/5.0 (Macintosh; U; Intel Mac OS X; es-ES; rv:1.8.1.3) Gecko/20070309 Firefox/2.0.0.3"

Taula 2.1: Exemple d'entrada de fitxer de log.

Entendre la semàntica dels diferents camps és indispensable per poder processar i analitzar els logs. A la Taula 2.2 es presenta una breu descripció del significat de cada camp.

213.73.39.193	Adreça IP del client des de la que es realitza la petició al servidor.
-	Usuari, utilitzat en els llocs web que requereixen autenticació estàndard. El fet que a l'exemple aparegui un guió significa que la informació no està disponible.
-	Password, utilitzat en els llocs web que requereixen autenticació estàndard.
[23/May/2007:01:10:08 -0700]	Dia i hora en que el servidor ha processat la petició. A més de la data i l'hora també s'indica la zona horària. La configuració dels servidors permet mostrar la data i hora en diferents formats.
"GET /folder/page.html HTTP/1.1"	Aquesta és la petició pròpiament dita, té els següents camps: Comanda o tipus d'operació: GET Recurs sol·licitat pel client: /folder/page.html Protocol utilitzat pel client: HTTP/1.1
304	Codi d'estat que el servidor envia al client quan processa la petició. Els codis mostren si la petició s'ha processat correctament o ha resultat en un error.
-	Mida en bytes de l'objecte servit al client. Si no s'ha servit cap contingut aquest camp pren el valor "-".
"http://www.uoc.edu/"	<i>Referer</i> . És el lloc web que el client reporta que ha originat la petició que s'ha servit.
"Mozilla/5.0 (Macintosh; U; Intel Mac OS X; es-ES; rv:1.8.1.3) Gecko/20070309 Firefox/2.0.0.3"	Agent d'usuari. És el sistema operatiu i el navegador que el client reporta d'ell mateix. Navegador Mozilla-Gecko-Firefox Sistema Operatiu: Mac OSX

Taula 2.2: Semàntica de les entrades dels fitxers de log.

### 2.3.3 Anàlisi de fitxers de log

Els fitxers de log són molt rics en dades i per tant constitueixen una font d'informació important. Poden ser estudiats i analitzats, i per tant obtenir informació rellevant sobre l'ús del lloc web i el seu tràfic segons l'hora del dia, el dia de la setmana o l'origen, per exemple. Tot i que es pot pensar que quantes més dades d'ús del lloc web es tinguin millor es podrà modelitzar l'ús del lloc web, el gran volum de dades que emmagatzema un fitxer de log fa que aquests arxius siguin impossibles de processar per humans i que no sempre sigui senzill processar-los usant programes.

Els fitxers de log normalment tenen un gran volum i són de difícil procés, i això molts cops representa un problema. Tot i la simplicitat de les dades de les quals es compon una entrada individual, l'estudi del conjunt de totes les entrades i el seu anàlisi i agregació proporciona un gran conjunt d'informació útil. De manera bastant directa i senzilla es pot saber el nombre de pàgines servides, la seva distribució tenint en compte diferents períodes de temps, la localització geogràfica dels visitants, els llocs web que enllacen el lloc web que estem analitzant i li dirigeixen tràfic, i altres.

L'anàlisi de fitxers de log de llocs web presenta unes avantatges molt interessants tot i que també presenta alguns problemes i dificultats. Aquestes dificultats han estat revisades i presentades en diferents treballs i estudis [77, 140]. A continuació es descriuen els avantatges i les dificultats més importants i aquelles amb una especial relació amb aquest treball.

El principal avantatge és que la informació que s'extreu dels logs es pot utilitzar per descriure l'ús real del lloc web. L'anàlisi de logs d'un lloc web senzill presenta els camins de navegació dels seus usuaris, les paraules clau utilitzades als cercadors per arribar al lloc web, les que s'han utilitzat en el propi lloc web per cercar informació, i els llocs web o pàgines que han servit de *referer*, és a dir, aquelles pàgines amb enllaços que referencien al lloc web i que han enviat usuaris. A més, tota la recollida de dades es realitza de manera automàtica, sense necessitat d'intervenció humana.

Si es processen els logs amb tècniques de *web mining* es poden descobrir informacions i patrons molt més interessants que els estadístics [66, 99, 164]. Per exemple, es poden construir grups o perfils de visitants i tenir-los en compte a l'hora de redissenyar el lloc web, oferint itineraris adequats per a cada perfil. També es pot utilitzar el *web mining* per oferir continguts i serveis personalitzats.

Els inconvenients i dificultats de l'anàlisi de fitxers de log fan referència principalment a la grandària dels fitxers, el processat de les dades i la identificació dels usuaris, sent aquest darrer el més important per a poder extreure informació relacionada amb el comportament dels usuaris.

La identificació dels usuaris és un dels elements claus per a que l'anàlisi de logs pugui generar resultats interessants, altrament els resultats es poden veure alterats i fins i tot falsejats. La majoria de llocs web no requereixen identificació dels seus usuaris i el segon i tercer camp de les línies dels seus fitxers de log queden buits.. En els fitxers de log, cada ordinador client que accedeix al lloc web queda identificat mitjançant l'adreça IP de l'ordinador. Amb l'adreça IP no sempre es pot identificar als usuaris individuals i, per tant, no es pot saber el comportament i la navegació d'un mateix usuari. Un mateix usuari pot visitar un lloc web des de diferents adreces IP ja que depenent del seu proveïdor d'Internet o ISP (*Internet Server Provider*), cada cop que es connecta se li pot assignar una IP diferent. Per altra banda, diferents usuaris poden compartir una mateixa adreça IP, per exemple, un mateix ordinador en un punt d'accés públic pot ser utilitzat per diferents persones per visitar el mateix lloc web. Algunes xarxes d'ordinadors petites comparteixen una mateixa adreça IP i és la que presenten externament. Es pot donar el cas que un proveïdor d'accés a Internet hagi instal·lat un proxy, llavors, tots els ordinadors que accedeixin a Internet mitjançant aquest proveïdor i que quedin afectats pel *proxy*, tindran la mateixa adreça IP. Un *proxy* és un servidor, que pot ser un programari o un maquinari, que s'encarrega de realitzar les peticions d'un conjunt d'ordinadors (es diu que queden "al darrere" del *proxy*). D'aquesta manera, l'ordinador que vol accedir a un lloc web contacta al *proxy*, i és el *proxy* qui realitza la petició al lloc web. Els servidors *proxy* també poden tenir memòria cau per optimitzar el tràfic i l'accés a llocs web, de manera que si hi ha peticions a planes que estan



emmagatzemades en la memòria cau, aquestes no generen una petició al lloc web i per tant no queden registrades als fitxers de log dels servidors.

Es pot utilitzar l'adreça IP per a obtenir el nom del host (*hostname*) associat a la IP, tot i que no sempre revela informació útil i adequada per a identificar usuaris. Hi ha programaris d'anàlisi que utilitzen el *hostname* per a obtenir informació demogràfica dels usuaris, així com el país d'origen i la localitat, l'afiliació institucional de l'adreça IP i del domini vinculat, es a dir, si és tracta d'una institució educativa, comercial o de comunicacions, per exemple. Aquesta informació no sempre és del tot fiable. Per exemple, es pot identificar els usuaris de moltes universitats segons el sufix *.edu* de l'adreça o bé pel propi nom de la universitat, tot i que només s'identificaran les connexions que estudiants i treballadors de la universitat realitzin des de la xarxa del campus. En canvi, molts dels estudiants, quan es connectin des de casa seva ho faran mitjançant un proveïdor d'Internet. Amb l'adreça IP i la seva traducció a nom de domini es pot obtenir la localització geogràfica dels usuaris. Hi pot haver dificultats quan s'utilitzen dominis genèrics com ara *.com .net o .org*, que tot i que hi ha el registre de la seva localització geogràfica, a la pràctica no sempre identifiquen geogràficament les adreces d'origen.

Les *cookies* o galetes es poden utilitzar per solucionar alguns dels inconvenients relacionats amb les adreces IP dels fitxers de log i la identificació d'usuaris. Les *cookies*, també anomenades *web cookies* o *HTTP cookies*, són petites porcions de text que el servidor web envia al navegador i que cada cop que es torna a accedir al servidor el navegador li envia el mateix text al servidor. Les cookies s'utilitzen per identificar usuaris, conservar informació sobre ells o per monitoritzar-los. Assignant una *cookie* a un usuari el primer cop que visita el lloc web pot servir per monitoritzar i fer seguiment de les seves visites. Hi ha l'inconvenient que si els usuaris de la web no estan identificats, assignar una identificació als usuaris únics fent servir *cookies* pot no tenir sentit en el cas que l'usuari canviï d'ordinador, ja que les *cookies* queden guardades en local, en el disc dur de l'ordinador des d'on es fa la connexió. Al utilitzar galetes també pot aparèixer l'inconvenient que molts usuaris o bé tenen bloquejades les galetes dels seus navegadors o bé periòdicament les esborren, donat que el control el tenen els usuaris i no els servidors.

Altres cops s'utilitza la tècnica de simular sessions. Quan un lloc web detecta que rep una visita, li assigna un identificador de sessió tot i que aquell usuari no tingui usuari i paraula clau, i es simula una sessió identificada. D'aquesta manera, en els logs queda registrat el numero de sessió i per tant se sap quines entrades pertanyen a cada usuari, independentment de la adreça IP. Aquesta tècnica presenta l'avantatge que es pot fer un seguiment acurat de cada usuari únic en les seves sessions de navegació i presenta els inconvenients propis de les *cookies*.

També cal considerar que els fitxers de log recullen moltes més dades que les pròpies d'ús, navegació i activitat dels usuaris. Aquestes dades embruten la informació dels logs i cal netejar-les aplicant els filtres corresponents abans de realitzar l'anàlisi. Aquestes dades són originades tant per les transaccions que sol·liciten elements relacionats amb el disseny o *layout* de les planes com per robots de buscadors, servidors *proxy* o errades del servidor. El resultat de recollir tota aquesta informació als fitxers de log implica que els fitxers tinguin una gran mida, que és un altre dels principals problemes dels fitxers de log. La gran mida dels fitxers provoca dificultats tant per processar els fitxers de log com per filtrar-los i seleccionar la informació rellevant. Si es vol estudiar la navegació a partir de les línies d'un fitxer de log, la relació entre línies útils i no útils és a partir de 1 sobre 10, una línia que aporta informació rellevant per cada 10 que són irrellevants.

Finalment, una altra qüestió important que cal adreçar al processar i analitzar fitxers de log és que no tot el que passa al navegador de l'ordinador client queda registrat en els fitxers de log. Per exemple si una plana està emmagatzemada en la memòria cau del navegador, quan l'usuari fa la petició per veure-la, el navegador la recupera de la memòria cau i no fa cap petició al lloc web, de forma que l'accés a la plana no queda registrat als fitxers de log. El temps dedicat a cada plana tampoc es registra als fitxers de log. Hi ha mètodes d'anàlisi de logs que consideren el temps entre peticions com el temps dedicat a la lectura de la plana. Això en alguns casos pot ser correcte, però en molts d'altres no. En cap moment es té control ni queda registrat què està fent l'usuari en l'ordinador, pel que després de fer una petició pot passar a treballar amb una altra aplicació, pot tenir més d'una finestra oberta de manera simultània o simplement pot abandonar l'ordinador.

Els fitxers de log capturen la realitat del comportament dels usuaris i l'ús que fan del lloc web. L'anàlisi de fitxers de log, tot i les seves limitacions, representa una eina destacable per entendre un lloc web i l'ús que en fan els usuaris i proporciona dades per la millora del seu rendiment, organització i usabilitat. Si es compara la informació que s'obté amb l'anàlisi de fitxers de log amb altres tècniques de recollida de dades com les enquestes o entrevistes, l'anàlisi de logs presenta l'avantatge que les dades que es recullen són reals i no pertanyen a una mostra sinó a tots els usuaris, i a més es captura tot l'espectre de usos del lloc web.

En relació a la privacitat i a la seguretat i la recollida i anàlisi dels fitxers de log cal fer algunes consideracions. Els fitxers de log no estan disponibles per als usuaris o visitants dels llocs web. Només hi tenen accés els webmasters o els administradors del sistema. De fet, aquests fitxers queden protegits per les legislacions i cal ser curós a l'hora d'emmagatzemar-los, manipular-los i estudiar-los.

### **2.3.4 Mètriques remotes**

Els sistemes de mètriques remotes es basen en un codi en les planes del lloc web que genera una entrada en una base de dades cada cop que s'accedeix i d'aquesta manera la base de dades conté tota la informació d'ús del lloc web. Aquests sistemes requereixen que la web que es vol estudiar incorpori en totes les seves planes un codi escrit en *Javascript*. Cada cop que un usuari accedeix al lloc web, el seu navegador al accedir al codi corresponent a una plana genera una entrada en una base de dades remota. Els camps d'aquesta entrada poden variar en funció de cada sistema concret, tot i que els camps bàsics són la data i hora, la plana accedida, el *referer* i l'agent. El sistema utilitza les entrades de la base de dades per a generar els informes d'accés i ús del lloc web. De fet, es pot considerar la base de dades que es construeix com un fitxer de log orientat exclusivament a les analítiques web. D'aquesta manera, el processament dels registres i el seu anàlisi és molt més senzill. Les aproximacions basades en mètriques remotes o que utilitzen *Javascript* normalment són de subscripció. La empresa que ofereix el producte també ofereix el hosting i manteniment de la base de dades que emmagatzema

les mètriques, tot i que també hi ha solucions més a mida on la empresa de mètriques instal·la les bases de dades en la pròpia localització del client. Hi ha moltes empreses que ofereixen serveis de mètriques remotes, aquestes són algunes: Clicktracks [8], Google Analytics [9] o IndexTools [13].

El principal avantatge de les mètriques remotes és que el processat de les dades es redueix dràsticament si es compara amb l'anàlisi de logs clàssic. El fet de tenir les dades d'ús en una base de dades simplifica l'anàlisi i fa que la generació de reports i informes sigui relativament senzilla. Un altre avantatge és que el propietari d'un lloc web pot tenir les dades d'ús directament, sense haver de manegar i processar les dades ell mateix. Les dades que es recullen són les que realment són útils per obtenir resultats i informes amb significat. El principal inconvenient d'aquesta aproximació és que requereix una infraestructura complexa si es compara amb l'anàlisi de logs clàssic, donat que cal disposar d'una base de dades que, depenent del tràfic del web que s'està estudiant, ha de tenir una versatilitat i robustesa considerables. Al mateix temps, depenent del tràfic del lloc web, la base de dades haurà de poder suportar un gran volum d'entrades i oferir un bon rendiment a l'hora de gestionar-les. Els sistemes de mètriques remotes normalment no requereixen descarregar i instal·lar un programari específic en l'ordinador del client. Tant els logs que fa servir, com la eina d'anàlisi estan hostatjats remotament. El client accedeix per consultar els sumaris i informes de l'anàlisi. Pel client, això representa un avantatge ja que no ha de conèixer un paquet informàtic específic i no ha de manegar els arxius de log, que sempre ocupen grans volums de disc i el seu processament requereix alts requeriments de maquinari. Per altra banda, aquesta aproximació requereix definir una estratègia de seguiment dels usuaris que servirà per definir una estratègia de marcatge o d'introducció del codi *javascript* en el lloc web. Això vol dir que s'ha de conèixer molt bé el lloc web, la seva arquitectura i navegació, i no només saber on cal posar els codis per fer el seguiment sinó també cal poder interpretar els resultats de la base de dades reconeixent les marques de cada àrea del lloc web.

Un dels principals aspectes a tenir en compte de les mètriques remotes és el relacionat amb la confidencialitat i privacitat de les dades. El sistema es basa en recollir les dades d'ús i emmagatzemar-les en una base de dades remota, en un servidor que normalment

es troba ubicat a la seu de l'empresa a qui es contracta el servei de mètriques. Per tant, aquesta empresa és dipositaria de tota la informació de navegació i ús del lloc web i dels usuaris que l'han utilitzat i això s'ha de tenir en compte a l'hora de valorar si fer ús dels seus serveis i de exigir unes garanties de seguretat i confidencialitat de les dades.

Hi ha sistemes d'anàlisi d'ús de llocs web que es basen en la combinació de les dues aproximacions, els fitxers de log i les mètriques remotes. La principal diferència entre les dues aproximacions fa que l'anàlisi de logs el realitzi el client en una màquina en local i que les mètriques de servidor presentin els informes de manera remota.

### **2.3.5 Consideracions**

Normalment, l'anàlisi tant de logs com de mètriques remotes generen resultats que tenen a veure amb el nombre de visitants, els visitants únics, el nombre de pàgines visitades i els hits de les pàgines. Aquestes figures és el que s'anomena *traffic data* o *traffic numbers*. Altres dades que es consideren tenen en compte la connexió (país, proveïdor, tipus de connexió) i l'origen o *referrer*.

A partir d'aquestes dades de tràfic, es generen una gran varietat d'informes que presenten les dades agregades o segmentades segons sigui el cas. Així doncs es pot consultar les visites segmentades pel buscador o enllaç d'origen o agrupades per origen geogràfic. Els elements diferenciadors bàsics entre els diferents productes, comercials o lliures, es la manera en com presenten aquests sumaris i informes. El disseny i estructura dels informes condicionen una adequada interpretació de les dades, per tant cal fer èmfasi en el significat de les dades, més que en els informes que es generen. La visualització de les dades és, cada dia més, un aspecte essencial de l'anàlisi de dades i ha donat lloc a publicacions i conferències especialitzades [45].

A l'hora d'interpretar les dades cal tenir en compte alguns aspectes, de vegades clau, però no caure en paranys i malinterpretacions de les dades. Alguns d'aquests aspectes són el nombre de visitats i visitants únics, la durada de la visita o sessió, o la manera en com es calculen alguns resultats. Per exemple, la durada de la visita no es pot mesurar ja

que no se sap mai amb certesa quan un usuari abandona el lloc web, però si es pot tenir una aproximació si la visita ha realitzat més d'una petició. Hi ha molts usuaris que accedeixen al lloc web però no realitzen cap clic, això pot voler dir que arriben al lloc web, li fan una ullada, i marxen [101, 125].

Com s'ha introduït, és important distingir entre visitants i visitants únics. Si la discriminació es realitza mitjançant l'adreça IP es corre el risc de comptar malament tant els hits com els visitants. Hi ha usuaris que comparteixen la mateixa adreça IP tot i accedir des de diferents ordinadors. Això es produeix quan s'accedeix a través d'un proxy. Els proxy són molt utilitzats en xarxes corporatives o institucionals i pels ISP. No se sap quans usuaris diferents hi ha al darrere d'un proxy, i totes les entrades que generen en els fitxers de log i en les eines de mètriques porten la mateixa adreça IP. Per altra banda, es pot donar el cas que un mateix usuari accedeixi amb diferents adreces IP, ja sigui perquè canvia d'ordinador o perquè el seu proveïdor d'accés a Internet dinàmicament li va canviant l'adreça. Per comptar visites acuradament cal tenir als usuaris identificats. Si el lloc web no requereix usuari i password, la identificació es pot portar a terme simulant sessions o be utilitzant cookies. Per altra banda, el fet d'utilitzar el concepte de sessió pot fer que si un usuari deixa d'utilitzar el lloc web durant un període de temps determinat (superior al que s'hagi establert com a durada de la sessió), la seva navegació es comptabilitzarà com la d'un usuari diferent. Normalment el temps d'inactivitat que s'estableix per a una sessió abandonada és de 30 minuts [140].

Les eines d'anàlisi han de generar uns resultats agregats que tinguin significat. Per exemple, les planes vistes en promig no s'haurien de comptabilitzar com la divisió de totes les planes visitades dividit pel nombre d'usuaris únics. La solució adequada seria fer el seguiment de cada usuari individual i monitoritzar i comptar les planes que visita. A partir de dades més treballades, com són aquestes, es poden calcular sumaris de manera més acurada. A l'hora d'interpretar les dades cal tenir en compte que les dades són agregades i que la realitat pot ser molt diferent. Les dades promig normalment no corresponen al comportament real de cap usuari individual tot i que proporcionen una visió general que pot resultar útil en una primera aproximació.

Les dades corresponents al nombre total de visites i de pàgines vistes no proporcionen informació sobre el comportament i navegació dels usuaris però sí que proporcionen un indicador molt interessant de la activitat general i ús del lloc web. Per tant, resulta interessant fer un seguiment individualitzat a usuaris concrets i obtenir informació detallada de la seva navegació. Aquesta informació amb més detall es pot utilitzar per dotar de significat a les dades que s'obtinguin amb els anàlisis de logs i les mètriques, i es podrà combinar amb informacions obtingudes amb altres tècniques, ja siguin qualitatives o quantitatives, provinents del disseny centrat en l'usuari.

Per interpretar correctament els resultats de les analítiques web cal conèixer la estructura de lloc web, les funcionalitats que ofereix i definir escenaris d'ús realistes. A més, és interessant combinar les analítiques web amb mètodes de disseny centrat en l'usuari presentats a la secció 2.1.2 i els trets de navegació i comportament dels usuaris descrits a la secció 2.1.4. Així, coneixent tant el lloc web com els usuaris i les accions que hi poden realitzar, es pot avaluar si té sentit que una determinada plana rebí un determinat nombre de visites tant a nivell global com d'un mateix usuari individual i esbrinar si es deu a un problema de disseny.

L'anàlisi d'ús de llocs web és una eina molt útil per avaluar la usabilitat [80, 88] i així millorar el seu disseny. Conèixer com és comporten els usuaris i com utilitzen un sistema interactiu és un factor clau en el seu disseny i, per tant, del seu èxit.

## **2.4 Tecnologies de modelització d'usuaris i personalització**

El concepte de personalització és un aspecte clau avui en dia. El fet bàsic que fonamenta la personalització és recollir informació de l'usuari i emprar-la d'una manera adequada; així, la personalització té en compte com les institucions recullen informació sobre els seus usuaris o clients, i com la utilitzen per canviar la manera en com tracten a cada persona individualment. Amb el fet de recollir dades sobre els usuaris no n'hi ha prou per personalitzar, és necessari una anàlisi adequada d'aquestes dades per a convertir-les en informació i aquesta informació en coneixement que pugui ser útil tant per a la institució com per a l'usuari, és a dir, una situació *win-to-win* on tots dos guanyen [92].

Per tant, per a poder oferir personalització es requereix un conjunt d'informació sobre les persones i això produeix una dualitat entre personalització i privadesa. El fet d'haver de proporcionar informació personal, ja sigui implícitament o explícitament, per part de l'usuari per així obtenir un millor servei provoca contradiccions tant en les persones a l'hora de proporcionar les dades, com en les institucions a l'hora de fer-ne un bon ús. Per a que la personalització arribi a bon port, en primer lloc és important proporcionar un bon servei als usuaris i no caure en el parany d'utilitzar la informació que es té per intentar obtenir un rendiment econòmic ràpid i fàcil [92, 96].

La personalització és útil per a tenir en compte les necessitats específiques de cada usuari, que tal i com s'ha descrit a la secció 2.1.1 és un aspecte clau del disseny de sistemes interactius centrat en l'usuari. Així, a l'hora de dissenyar i construir sistemes que proporcionen personalització cal preguntar-se si les funcionalitats personalitzables que s'ofereixen constitueixen un valor real als usuaris, i cal veure també com la personalització encaixa en el disseny general del sistema [127]. Per tant, és important analitzar el rol de la personalització en el disseny de l'experiència dels usuaris amb el sistema i integrar-lo en el procés de disseny centrat en l'usuari [100]. Al mateix temps, es produeix la paradoxa que, en moltes ocasions, els sistemes amb personalització produeixen desconcert en l'usuari donat que a l'adaptar opcions i funcionalitats, l'usuari té uns millors serveis però al mateix temps té la percepció de pèrdua de control de la interacció [91]. Finalment, els mecanismes de personalització poden afegir una complexitat important al sistema [89], afectant-ne el seu rendiment, el que pot ser percebut negativament pels usuaris.

L'estudi i la recerca en personalització es realitza des de diferents aproximacions que tenen en compte des del grau de personalització que es vol aconseguir fins l'àmbit específic d'aplicació, passant totes elles per l'estudi del conjunt de tecnologies que ho fan possible. Aquestes tecnologies es basen principalment en la recollida de dades sobre els usuaris i el seu adequat anàlisi per a així construir un model, anomenat model d'usuari [97], que permeti saber en cada moment les necessitats que pot tenir en funció del context. Així, la personalització ha donat lloc a diferents àrees de recerca com són el *Adaptive Hypermedia* [40], *User Modeling* [97] o *Adaptive User Interfaces* [104]. En



aquesta secció es presenten aquelles àrees i tecnologies que tenen un especial interès i rellevància per a aquest treball de tesi doctoral. És interessant destacar que en aquest treball no es diferencia entre personalització i adaptació, mentre que aquests dos termes difereixen del concepte de customització, que fa referència a les opcions d'un sistema que l'usuari configura i estableix segons les seves preferències, com ara per exemple l'aspecte de la interfície.

La personalització té moltes dimensions i, per tant, els sistemes que la incorporen ho poden fer de moltes maneres i estils. Independentment del domini d'aplicació del sistema de personalització, la personalització pot adaptar enllaços, continguts, el context, etc. La personalització no és un objectiu senzill donat que tracta amb un gran conjunt de dades i vol aconseguir uns resultats específics i acurats. Involucra, per tant, tot un conjunt de diferents tecnologies que inclouen des del senzill accés a una base de dades fins a algorismes de mineria de dades, passant per complexos sistemes de filtratge col·laboratiu i sistemes de presa de decisions basats en regles.

Els sistemes hipermèdia adaptatius es basen en el *Adaptive Hypermedia* (AH) [40, 41]. Van començar a aparèixer als anys 90 quan es va combinar la recerca de conceptes com *user modeling* i *user adaptation* amb l'hipertext. La primera aplicació de l'hipermèdia adaptatiu va ser a l'àmbit de l'educació on la llibertat de navegació que ofereix el hipertext es va introduir a l'àrea dels sistemes tutors intel·ligents. A partir d'aquell moment, el desenvolupament dels sistemes hipermèdia adaptatius va coincidir amb el desenvolupament de la web, amb una gran interrelació entre tots dos camps. Aquests sistemes ofereixen diferents graus i modalitats d'adaptació i depenent de l'àrea d'aplicació es poden classificar en sistemes adaptatius educatius, sistemes d'informació adaptatius i sistemes recomanadors adaptatius.

L'*adaptive hypermedia* vol donar solució als problemes amb els que es troben els usuaris a l'interaccionar amb sistemes web dinàmics. Així, es pot guiar la navegació, adaptar els enllaços que es presenten i també el contingut. Brusilovsky [40] distingeix dos tipus d'adaptació: *adaptive presentation* i *adaptive navigation support*. L'adaptació de la presentació adequa la informació que es presenta a cada usuari, focalitzant-se en la presentació del contingut. Això pot implicar, per exemple, presentar un text més o

menys detallat, o el mateix contingut en format àudio o textual. Per altra banda, la navegació adaptativa estableix la seqüència o ordre que considera més adequada per presentar els continguts. En el seu treball, l'autor fa un anàlisi en profunditat de cadascun dels dos tipus amb els diferents elements que hi intervenen i els aspectes a tenir en compte.

### **2.4.1 Tecnologies de la personalització**

Les anomenades tecnologies de personalització proporcionen mètodes i tècniques per a construir sistemes que aporten personalització als entorns, continguts o serveis, i que en molts casos fan servir les diferents tecnologies de modelatge d'usuaris, anàlisi d'ús de llocs web i de mineria de dades. Aquestes tecnologies s'utilitzen per a recollir informació dels usuaris, emmagatzemar-la i així inferir els seus interessos i preferències, oferir recomanacions i, fins i tot, predir la seva navegació. Les principals tecnologies de personalització són les anomenades tècniques de filtratge: *Collaborative Filtering*, *Rule-Based Filtering* i el *Content-Based Filtering* [78, 145, 146].

Per a poder personalitzar és necessari un seguit d'informacions sobre les preferències, necessitats, objectius i expectatives dels usuaris. El conjunt d'informació que descriu a un usuari s'anomena perfil o model d'usuari [97]. El perfil d'usuari està format per dos components principals, el component factual i el component conductual [28]. El component factual inclou informació demogràfica i transaccional, com pot ser l'edat, ingressos, nivell educatiu, marques preferides, etc. El component conductual inclou informació sobre l'activitat que porta a terme l'usuari com pot ser la seva navegació i ús del sistema. Aquesta informació normalment s'emmagatzema en forma de descripcions lògiques, regles de classificació, parells atribut-valor, etc. El sistema de representació més utilitzat són les regles d'associació [29]. Aquestes regles es poden obtenir de dues formes, definides per un expert en el domini o obtingudes a partir de l'aplicació de tècniques de mineria de dades sobre conjunts de dades de registre de l'activitat dels usuaris.

L'anàlisi d'ús de llocs web i la informació que recull i processa, i que s'ha descrit a la secció 2.3, resulta de gran utilitat per a la personalització i constitueix la primera font de dades sobre els usuaris, especialment en aquells casos on es monitoriza i guarda la informació relativa a la navegació individual de cada usuari. La informació d'ús del sistema o del lloc web emmagatzema tant els perfils individuals com agregats per tal d'anar completant i perfeccionant el model d'usuari de manera gradual. Aquesta anàlisi es realitza en una primera etapa utilitzant mètodes estadístics, tot i que els sistemes més avançats fan ús de tècniques de mineria de dades en general i de *web mining* [163] i de regles d'associació [27] en particular.

El procés de construcció de models d'usuari mitjançant regles consta normalment de dues fases: descobriment de regles i validació de regles. Per descobrir les regles es poden utilitzar diversos algorismes de mineria de dades. Hi ha un tipus especial de regles d'associació anomenades *profile association rules* [28] on la part esquerra de la regla té la informació del perfil d'usuari (edat, ingressos, educació, etc.) i la part dreta de la regla conté informació sobre el comportament de l'usuari (compra cervesa, utilitza vals de descompte, etc.). El principal problema amb els diferents mètodes d'obtenció de regles és el gran nombre de regles que s'obtenen, normalment moltes, que tot i ser estadísticament acceptables, poden ser il·legítimes, irrellevants o trivials. Normalment es realitza una anàlisi posterior manualment per filtrar i eliminar les regles il·legítimes i irrellevants. Molts sistemes comercials de mineria de dades porten a terme el filtratge de regles fent que un expert humà realitzi el filtratge inspeccionant una a una les regles generades, rebutjant les inacceptables. Aquesta aproximació al filtratge de regles, però, no és factible quan tenim un gran nombre de regles i de perfils d'usuari.

Les dades per construir perfils d'usuari poden ser recollides de forma implícita o explícita. La recollida explícita de dades requereix la participació activa de l'usuari; d'aquesta manera es permet a l'usuari controlar la informació emmagatzemada en el seu perfil. El perfilatge explícit pot ser de diferents tipus. L'usuari pot omplir un formulari, respondre una enquesta, contestar un qüestionari o emplenar informació personal a l'hora de registrar-se en un servei. Aquest mètode de recollir informació té l'avantatge d'obtenir la informació directament de l'usuari i que digui què vol i com ho vol.

Per altra banda, el perfilatge implícit no requereix la participació directa de l'usuari i normalment es realitza "per sota". El perfilatge implícit significa realitzar una monitorització i una traça de l'activitat i interaccions dels usuaris, per tal d'identificar patrons de comportament, activitat i ús. En la majoria dels casos, el seguiment es realitza sense el consentiment explícit dels usuaris i es manté la transparència del sistema, és a dir, l'usuari no se'n adona de que se li està fent un seguiment i monitorització, ja que aquest es porta a terme en el servidor. En el cas de la web, la informació implícita pot ser recollida tant a la banda del servidor com a la del navegador o client. La informació del servidor pot ser els logs d'accés generats automàticament pel servidor, logs de referència o logs d'agents. La informació de la part del client pot incloure les cookies o galetes, la traça del teclat o del ratolí o la traça del DOM (*Document Object Model*). Una altra font d'informació i de dades sobre els usuaris o clients poden ser les bases de dades que recullen les diferents transaccions realitzades, tan siguin de venda, d'atenció al client, de matrícula o dades demogràfiques i de satisfacció recollides al fer enquestes. En molts casos, les dades estan recollides i emmagatzemades en diferents formats i en bases de dades diverses i, per tant, serà necessària una feina de preprocessament i integració de les dades per tal d'evitar inconsistències i redundàncies, amb l'objectiu d'obtenir una única font fiable de dades. El perfilatge implícit elimina la dificultat associada amb el fet d'obtenir informació personal i, per tant, sensible de l'usuari. En lloc d'obtenir la informació directament de l'usuari, el sistema recull les dades considerades rellevants i tracta d'inferir informació específica de l'usuari. Tot i que el perfilatge implícit és menys intrusiu, cal tenir en compte els diferents aspectes de privacitat, com s'ha esmentat amb anterioritat.

Els perfils d'usuari i els seus components es poden classificar en estàtics o dinàmics, individuals o agregats (perfils de grup). Un perfil és estàtic quan rarament o mai canvia (per exemple, la informació demogràfica). Un perfil és dinàmic quan la seva informació canvia en el temps, com per exemple les preferències de compra de l'usuari. Aquests perfils s'hauran d'anar actualitzant per tal de que reflecteixin els canvis, ja que aquests canvis aporten també informació sobre l'usuari.

La modelització d'usuaris i les tecnologies de personalització poden ser utilitzades de manera òptima per a oferir serveis web personalitzats, adequats a les necessitats de cada

persona o grup de persones. Per altra banda, es pot caure en el parany d'utilitzar-les de manera intrusiva presentant continguts no sol·licitats per l'usuari, per exemple. És important fer notar que en general hi ha un compromís entre la privacitat i la personalització, quanta més informació els usuaris mostren, més serveis personalitzats se li poden oferir i més enfocada serà la personalització.

#### **2.4.2 Personalització i e-learning**

Els avantatges que proporciona la personalització a l'hora de adequar la experiència d'ús a cada usuari o grup d'usuaris, resulta de gran interès en els sistemes d'e-learning. Les tres dimensions dels sistemes d'e-learning es veuen afectades per la personalització, és a dir, els continguts, els entorns virtuals d'aprenentatge i els usuaris. Els continguts es poden adaptar als estudiants tant a nivell de presentació, com de format, de profunditat i de seqüenciació. Els entorns poden proporcionar suport a la navegació i ajuda per realitzar les diferents tasques relacionades amb el procés d'aprenentatge. Finalment, els usuaris, principalment els estudiants, es beneficien de sistemes educatius que tenen en compte les seves preferències i necessitats, tant a l'hora de navegar per l'entorn virtual d'aprenentatge com a l'accedir als continguts educatius.

L'objectiu d'incorporar elements de personalització a l'e-learning ha produït diferents aproximacions. Així, des de la intel·ligència artificial s'ha abordat amb l'àrea *Intelligent Computer-Aided Instruction (ICAI)*, també anomenada *Artificial Intelligence in Education (AIED)* [154]. L'AIED vol aplicar els mètodes i tècniques pròpies de la intel·ligència artificial al desenvolupament de sistemes tutors intel·ligents [175], és a dir, sistemes educatius que proporcionen funcionalitats avançades i que adapten la instrucció als seus usuaris. Així, per exemple, s'apliquen models predictius basats en tècniques estadístiques [179] com per exemple les xarxes bayesianes [137] que resulten flexibles i proporcionen resultats acurats per a poder construir models predictius i poder, d'aquesta manera, realitzar recomanacions als estudiants. Millorar el feedback que reben els usuaris per part del sistema és un objectiu d'aquests tipus de models.

Per altra banda, hi ha aproximacions que tenen en compte el gran volum de dades que generen els sistemes d'e-learning, tant acadèmiques com d'interacció dels estudiants amb els continguts educatius, com amb el propi entorn virtual d'aprenentatge. Aquest gran volum de dades pot ser analitzat i estudiat per a obtenir informació rellevant, per tal de conèixer millor als estudiants i per millorar les seves interaccions amb el sistema, així com el procés d'ensenyament i aprenentatge. Aquesta tendència ha donat lloc a una àrea de recerca que rep el nom de *Educational Data Mining* [147] i persegueix l'objectiu d'adaptar les tècniques i mètodes de la mineria de dades a la especificitat de les dades i del context dels sistemes d'e-learning.

Des de la comunitat de l'hipermèdia adaptatiu, ja presentat en aquesta secció, es proposen els *adaptive hypermedia educational systems* [39]. Aquests sistemes tenen el seu punt de partida en els sistemes tutors intel·ligents que proposen adaptar els continguts del curs als estudiants, tenint en compte les instruccions del tutor. A diferència d'aquests, els sistemes educatius basats en l'hipermèdia adaptatiu posen el seu focus en els estudiants i tracten d'adaptar els sistemes a les necessitats específiques d'aquests, proporcionant una major llibertat de navegació, de manera que diferents estudiants puguin estudiar els continguts en diferent ordre i que, a més, hi hagi una adaptació dels continguts en forma d'explicacions addicionals per a aquells estudiants que ho necessitin per a entendre millor els continguts, per exemple.

En aquests sistemes s'adrecen dues qüestions clau, què es pot adaptar i quins paràmetres es poden adaptar al sistema. Brusilovsky [40] proposa una taxonomia d'adaptació que parteix de l'adaptació de la presentació i de l'adaptació de la navegació. En els sistemes educatius, l'objectiu d'adaptar la presentació es tradueix en adequar els continguts tenint en compte el model d'estudiant. Per la seva banda, la finalitat de la navegació adaptativa és donar suport als estudiants a l'hora de navegar, tant pels continguts com per l'entorn virtual d'aprenentatge.

Els sistemes adaptatius s'estructuren al voltant del model d'usuari i, de la mateixa manera, els sistemes d'e-learning adaptatius es construeixen al voltant del model d'estudiant, que emmagatzema la informació del mateix. Aquest model normalment conté informació personal en relació als seus estudis, els seus objectius d'aprenentatge,

les qualificacions obtingudes, les seves preferències i estil d'aprenentatge, i la informació sobre navegació per l'entorn virtual d'aprenentatge. Tota aquesta informació ha de servir per millorar les seves interaccions amb el sistema d'e-learning.

## **2.5 Resum**

En aquest capítol s'ha presentat l'estat de la qüestió de les àrees de coneixement que són interessants per aquest treball de tesi doctoral. S'ha presentat la interacció persona-ordinador i el disseny centrat en l'usuari, on conèixer els usuaris constitueix l'aspecte clau a l'hora de dissenyar i construir sistemes interactius. A continuació s'ha presentat l'e-learning, amb els seus sistemes i estàndards. Els entorns virtuals d'aprenentatge són sistemes interactius, en ocasions molt complexos, que promouen la interacció entre persones, continguts i serveis. Una manera d'ajudar als usuaris d'aquests entorns és analitzar la seva interacció per a que el sistema "els conegui". Per aquest motiu també s'ha presentat com analitzar l'ús de sistemes web interactius, per així obtenir informació sobre els usuaris i poder adaptar el sistema a les seves necessitats. La personalització pretén portar a terme aquesta adaptació del sistema a les necessitats de cada usuari. Per aconseguir-ho es basa en models d'usuari que es construeixen amb la informació disponible dels usuaris i la que es recull de l'ús que fan del sistema. A partir d'aquest anàlisi de la qüestió, en el següent capítol es presenta la proposta d'anàlisi de l'ús i dels usuaris d'un sistema d'e-learning complex i canviant, com és el de la Universitat Oberta de Catalunya.





## Capítol 3

### Anàlisi de l'ús i dels usuaris de sistemes d'e-learning

En aquest tercer capítol es presenta l'aportació principal d'aquest treball de tesi a l'àmbit de coneixement. L'objectiu principal és desenvolupar una metodologia de recollida de dades i d'anàlisi d'usuaris per a sistemes web amb un gran volum d'usuaris, especialment per als entorns d'e-learning i que, a més, també sigui adaptable a d'altres entorns web complexos.

El resultat que s'obtenen amb la investigació d'usuaris són útils per conèixer els usuaris d'un sistema web i així adequar el disseny i la interacció a les seves necessitats. Més concretament, la informació que s'obté s'utilitzarà per l'avaluació i millora de la usabilitat del sistema [72], per dissenyar itineraris formatius que tenen en compte les característiques dels estudiants i les seves necessitats, i per introduir elements d'adaptació i personalització al sistema [116, 121].

#### 3.1 Introducció

A Internet i als entorns d'e-learning encara no és gens clar quins indicadors i quines mètriques són les adequades i rellevants a l'hora de dissenyar entorns, processos i interaccions. Moltes decisions de disseny encara es recolzen en la intuïció i en hipòtesis sobre els usuaris i l'ús que fan dels sistemes, que no estan contrastades amb cap tipus de dada real u objectiva. Aquest treball afronta aquesta qüestió i proposa una metodologia

que serveix per a obtenir dades i resultats que proporcionen coneixement dels usuaris, del seu comportament i de l'ús que fan dels entorns i dels sistemes web, més enllà de les mètriques habituals com ara els *hits* o el nombre d'accessos a una plana. Aquesta metodologia segueix un procés cíclic i un refinament progressiu seguint, d'aquesta manera, la perspectiva del *design research* [79], on el disseny és constantment revisat fins a assolir els objectius plantejats. Es pot obtenir informació sobre els usuaris en forma de mètriques, camins i patrons de navegació, a partir de les dades o *logs* que automàticament genera i recull un entorn web o una plataforma d'e-learning i de la informació que emmagatzema dels usuaris i les seves interaccions. A més d'aquestes dades, per tal de construir camins i patrons acurats i útils, la informació dels fitxers de log no és suficient i s'ha de complementar amb altres tipus de dades com són les marques d'ús, situades estratègicament en diferents punts del sistema o de l'entorn virtual. La informació que s'obté d'aquesta manera és de gran rellevància i utilitat ja que permet conèixer als estudiants i l'ús que fan del sistema d'e-learning i, a partir d'aquí, dissenyar millores tant del propi entorn educatiu com del procés d'aprenentatge.

L'aproximació a la captura, anàlisi i modelatge del comportament dels usuaris que es realitza en aquest treball s'orienta al coneixement dels usuaris a partir dels seus camins de navegació i de la construcció de patrons de navegació i comportament. És per aquest motiu que es proposa un *framework* de recollida de dades i d'anàlisi estructurat a diferents nivells per a l'estudi dels usuaris dels entorns d'e-learning. La metodologia que es proposa no només pot ser aplicada l'àmbit de l'e-learning i als entorns virtuals d'aprenentatge, sinó que tenint en compte els diferents contextos, també pot ser utilitzada per a l'estudi d'usuaris d'entorns o sistemes web complexos i amb usuaris recurrents, com ara els usuaris d'establiments de comerç electrònic o de banca *online*. Els nivells d'anàlisi que es proposen es corresponen amb els terminis o intervals temporals de la interacció, tenint en compte la idiosincràsia de l'entorn i les diferents dades recollides en diferents períodes de temps i distingeixen, per tant, diferents patrons de navegació i comportament dels usuaris. El mateix concepte de navegació té un significat diferent en funció del nivell d'anàlisi.

El treball que es presenta té la característica que el seu camp d'estudi i anàlisi és el de la Universitat Oberta de Catalunya (UOC) i el seu campus virtual, disposant per tant

d'un entorn d'e-learning amb gran nombre d'usuaris i una quantitat important de dades que permeten presentar resultats interessants i significatius. La UOC i el seu campus virtual constitueixen un escenari únic, i el fet d'utilitzar-los com a banc de proves permet validar l'anàlisi a diferents nivells i generar resultats directament relacionats amb les particularitats de la pròpia universitat, així com resultats més generals que tenen com a focus les mètriques d'usuari, la usabilitat i les seves relacions i concrecions amb l'e-learning.

### **3.2 Proposta d'anàlisi multinivell**

L'estudi d'usuaris des de la perspectiva de l'e-learning es pot descompondre i estructurar en nivells diferents. La definició d'aquests nivells s'ha portat a terme a partir de la informació que disposa la UOC sobre els seus estudiants, la naturalesa pròpia de la universitat, dels ensenyaments a distància i dels sistemes d'e-learning. Aquestes anàlisis de dades preliminars que es presenten a la secció 4.3 i a la secció 4.4. Totes aquestes fonts d'informació s'han refinat i validat amb la informació obtinguda de l'aplicació de la tècnica de construcció de personatges i escenaris, que es presenta a la secció 3.3 d'aquest capítol.

De forma intuïtiva, es pot dir que els nivells d'anàlisi es corresponen amb el curt, mig i llarg termini dels estudis que les persones realitzen en entorns virtuals d'aprenentatge. El primer nivell és el nivell de sessió o de connexió individual al campus virtual, i és el que té en compte el període més curt de temps. El segon nivell és el nivell de curs o d'assignatura i la seva durada correspon a la durada d'un curs acadèmic. El tercer nivell és el nivell d'aprenentatge al llarg de la vida o *lifelong learning*, i té una durada llarga i molts cops indeterminada, en funció dels objectius de l'estudiant. Com es pot veure, aquests nivells venen donats de manera natural per l'ús que els estudiants fan del campus virtual i la dinàmica pròpia de la seva formació al realitzar estudis no presencials de nivell universitari. Cadascun d'aquests nivells de comportament, navegació i ús proporciona informació rellevant per construir un model d'usuari que es pot utilitzar per personalitzar el procés d'aprenentatge i per assolir diferents objectius relacionats amb el disseny de l'entorn d'aprenentatge, la usabilitat i la experiència dels

estudiants. Considerar conjuntament els tres nivells permet construir una imatge global dels usuaris i el mapa de l'ús que fan del sistema, amb l'objectiu de millorar-lo. Com es veurà més endavant en aquest capítol, només un o dos nivells d'anàlisi no són suficients per analitzar l'escenari proposat, mentre que quatre o més seria inconsistent.

El nivell de sessió estudia el comportament de navegació a curt termini, és a dir, què fa cada usuari individual cada cop que es connecta al campus virtual. Un cop s'obté la informació de cada usuari es fan agrupacions per obtenir comportaments generals d'usuari i entendre el seu comportament. També es pot realitzar un seguiment de les diferents sessions d'un mateix usuari per veure si el mateix usuari té patrons de navegació similars al llarg del temps. Molts cops, al dissenyar accions formatives en entorns virtuals d'aprenentatge es parteix de diferents hipòtesis de treball. Per exemple, a la UOC es considera que els estudiants es connecten al campus virtual en sessions de vint minuts on en primer lloc consulten la seva bústia personal i després accedeixen als cursos que han matriculat. Per altra banda, l'entorn virtual d'aprenentatge pot no haver estat dissenyat tenint en compte hipòtesis com aquesta, el que planteja contradiccions. La informació obtinguda en aquest nivell permetrà contrastar aquestes i d'altres hipòtesis de treball utilitzades per dissenyar accions formatives.

El comportament de navegació a mig termini s'estudia al nivell de curs o, en el cas de la UOC, a nivell de semestre acadèmic. En aquest nivell es poden construir els patrons de comportament de navegació, tenint en compte la informació obtinguda en el nivell de sessió, en el període de temps d'un curs acadèmic, un semestre. Els resultats obtinguts en aquest nivell de mig termini també serveixen per a fer agrupacions d'usuaris i obtenir perfils. La informació que es treballa en aquest nivell permet conèixer els patrons de comportament dels estudiants durant el curs acadèmic i permetrà esbrinar si hi ha patrons d'actuació correlacionats amb l'assoliment dels objectius d'aprenentatge i la superació del curs. Amb les dades que s'estudien en aquest nivell també es poden descobrir els patrons d'ús dels diferents espais i serveis del campus virtual, informació que és d'especial importància, tant per tenir mètriques concretes dels espais i serveis, com per identificar les accions que s'haurien de portar a terme tant d'utilitat com d'usabilitat.

En el tercer nivell, el de *lifelong learning*, es troben diferents etapes que es corresponen amb les etapes acadèmiques o de vida acadèmica de l'estudiant. Poden ser entrada o accés, primera matrícula, altres matrícules, matrícules després l'obtenció del títol, etc. L'estudi del comportament de navegació dels usuaris a llarg termini serveix per identificar els aspectes i variables que segmenten les fases de la vida acadèmica dels estudiants. Per exemple, en aquest nivell es pretén descobrir les combinacions d'assignatures o cursos que afavoreixen l'assoliment d'una titulació, així com les necessitats de formació de les persones al llarg de la vida o les causes de l'abandonament dels estudis.

Pel que fa al concepte de navegació, aquest canvia de significat en funció del nivell d'anàlisi. Així, pel nivell de sessió, per navegació s'entén com l'usuari accedeix als serveis disponibles, tenint en compte en quin ordre i a través de quins mecanismes ho fa. En el nivell de semestre, per navegació s'entén com l'usuari visita o no certs espais de l'entorn virtual d'aprenentatge, i com realitza les activitats predefinides a cada assignatura com, per exemple, descarregar-se un enunciat. A nivell de *lifelong learning* per navegació s'entén com l'estudiant selecciona les assignatures i com va progressant en l'obtenció del títol.

Els tres nivells d'estudi d'usuaris estan fortament interrelacionats entre ells, tal i com es mostra a la Figura 3.1, de dalt cap a baix, és a dir, les dades del nivell de sessió complementen les dades dels altres dos nivells, el de curs i el d'aprenentatge al llarg de la vida. També hi ha una interrelació de baix cap a dalt, on les dades del tercer i segon nivell poden explicar la navegació dels usuaris pel campus virtual en el nivell de sessió. Així, per exemple, se sap que a l'inici de semestre els estudiants naveguen principalment per les aules i que els estudiants que porten més temps a la universitat naveguen menys pel campus virtual. També hi ha una forta interrelació entre els nivells més propers, el primer amb el segon i el segon amb el tercer, donat que, a cada nivell, les dades i els fets que se'n desprenen afecten més directament als nivells més propers. Així, els resultats d'un nivell poden ser usats com a variables que complementen l'anàlisi del següent. D'aquesta manera hi ha un fil conductor entre els nivells que dona coherència i unitat a l'anàlisi.

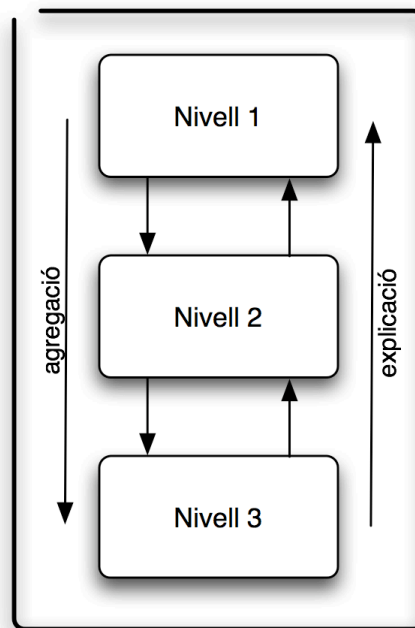


Figura 3.1: Interrelació entre els nivells.

Per altra banda, l'estudi d'usuaris d'entorns d'e-learning a tres nivells permet, des del punt de vista de la interacció persona-ordinador, analitzar l'ús real dels sistemes interactius i així estudiar la correlació entre el model conceptual del disseny i el model mental dels usuaris que utilitzen el sistema. També permet fer recerca en el que s'anomena *Long Term HCI* [94], també anomenat *ongoing relationships* [117] que consisteix en estudiar la evolució dels usuaris d'un entorn interactiu i recurrent al llarg del temps, tenint en compte que els usuaris aprenen a utilitzar el sistema i poden passar de usuaris novells a usuaris experts i, per tant, la seva interacció canvia.

La interacció persona-ordinador recorda que al dissenyar un sistema interactiu cal tenir sempre en compte els factors humans, és a dir, com les persones interaccionen i utilitzen els sistemes [63, 131]. Des d'aquest punt de vista, els camins de navegació obtinguts a partir de la recollida de dades i el seu processament i anàlisi, i sense involucrar directament a l'usuari, proporcionen una informació d'ús concreta que en determinades ocasions cal completar amb altres dades que s'obtenen directament de l'usuari. La informació que s'obté a partir de les dades d'ús dels sistemes interactius és, en alguns casos, una informació incompleta donat que no es coneix l'objectiu de l'usuari i, per

tant, es desconeix la intenció amb que ha realitzat certes accions i navegacions. En aquest treball, aquesta qüestió apareix suavitzada donat que s'estudia un sistema tancat i amb usuaris registrats i identificats que poden portar a terme un conjunt específic de tasques. A més, es disposa d'un gran volum de dades d'ús amb les quals, correctament processades i analitzades, es pot obtenir coneixement rellevant i altament significatiu. Per altra banda, tal i com es presenta a la secció 3.5 d'aquest mateix capítol, a l'introduir un conjunt de marques als fitxers de log, això els hi proporciona una riquesa superior i una semàntica que elimina moltes de les ambigüitats que presenten, al mateix temps que facilita el seu processat i anàlisi posteriors.

Un aspecte important en el estudi del comportament dels usuaris és com es tenen en compte els seus objectius i necessitats. Tant important com descobrir el camí de navegació d'un usuari és conèixer si la navegació obtinguda respon a una estratègia i si és així, quin és l'objectiu que es vol assolir. Les dues estratègies de navegació per la web utilitzades més freqüentment són la cerca i la exploració, tal i com s'ha presentat a la secció 2.1.4. En un entorn virtual d'aprenentatge hi poden haver moltes més estratègies, donat que es tracta d'un entorn ben definit i els usuaris tracten d'assolir uns objectius d'aprenentatge. L'avantatge és que els possibles objectius dels usuaris són coneguts a priori i el seu nombre, tot i ser elevat, és limitat. Per tant, el fet de contemplar l'estudi de dades a tres nivells permet, tot i desconèixer l'objectiu de l'usuari a curt termini, fer hipòtesis fiables dels seus objectius a mig i llarg termini. Entre els objectius dels usuaris a mig termini es consideren el seguiment d'un curs, l'estudi d'uns continguts educatius, la realització de les activitats previstes al pla docent, l'assoliment d'objectius d'aprenentatge o la superació d'un examen, per exemple. A llarg termini els usuaris persegueixen assolir una titulació i obtenir una formació que s'adapti a les seves necessitats professionals i d'acord amb el progrés del seu coneixement.

A més, la utilització i aplicació de mètodes de disseny centrat en l'usuari proporciona informació sobre els estudiants, les seves necessitats, preferències i hàbits d'ús. Els mètodes aplicats fins ara han proporcionat informació sobre objectius, motivacions i actituds dels estudiants, recollida principalment en la documentació del projecte Campus en Programari Lliure[19]. L'estudi que es planteja d'anàlisi d'interaccions a

diferents nivells i a partir de grans volums de dades és molt interessant que sigui complementada i al mateix temps serveix de complement a d'altres metodologies que parteixen de conèixer què vol fer l'usuari, però que en molts casos escassament tenen dades del què realment fa i com utilitza un sistema. Per tant, l'obtenció de camins i patrons de navegació a partir de la informació que el sistema emmagatzema de l'usuari, complementa a la perfecció, i en molts cassos pot suplir les carències pròpies, de mètodes de disseny centrat en l'usuari com són el test observacional, qüestionaris, entrevistes o *focus groups*, entre d'altres.

En conseqüència, tenint en compte aquestes dues aproximacions, la de l'e-learning i la del disseny centrat en l'usuari, a cada nivell es poden conèixer uns objectius i preferències dels estudiants i es poden assumir uns altres, i d'aquesta manera orientar el processament i anàlisi de les dades cap a objectius concrets.

A cadascun dels nivells, a més dels estadístics de caire més clàssic, s'utilitzen tècniques i mètodes que formen part de la mineria de dades i del *web mining* [164], que és l'aplicació de les tècniques pròpies de la mineria de dades a la web. La mineria de dades és el procés de descobriment de coneixement dins de grans volums de dades, i l'aplicació de les seves tècniques i algorismes a les dades de cada nivell ajuda a descobrir quines són les variables i els indicadors que proporcionen informació rellevant per caracteritzar el coneixement dels usuaris de l'ús que fan del sistema.

En relació a la personalització dels itineraris formatius i l'adequació de continguts i serveis a les necessitats dels usuaris, per a cada objectiu de personalització la rellevància de les variables i els indicadors pot ser diferent i així els que proporcionen menys informació per segons quines recomanacions, es mostraran més rellevants per segons quines altres. A cadascun dels tres nivells, personalitzar també té un significat diferent i es basa en diferents dades sobre el sistema i els usuaris. La proposta que es realitza en aquesta tesi es presenta a la secció 3.7.

La Figura 3.2, que es mostra a continuació, sumaritza en forma d'esquema els objectius exposats, la captura, processament i anàlisi de les dades d'ús i les interaccions dels usuaris sota el prisma del disseny centrat en l'usuari per obtenir patrons de navegació i



mètriques d'ús, i que aquest coneixement retorni al campus virtual millorant el seu disseny, usabilitat i utilitat.

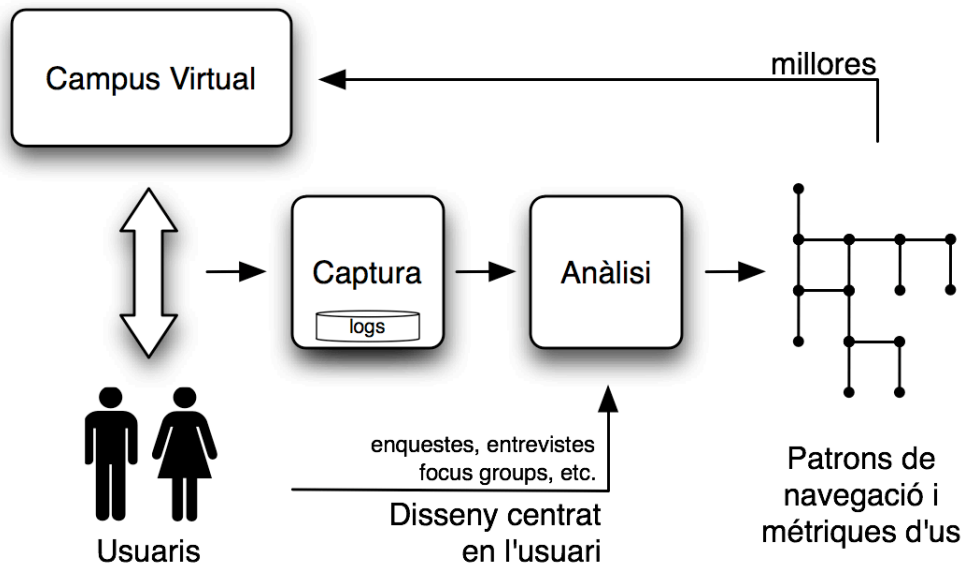


Figura 3.2: Procés de millora basat en la captura i anàlisi de dades d'ús.

A l'obtenir els camins i patrons del que fan els usuaris en el sistema el que s'està fent és construir mapes de navegació per a cada usuari individual i per a cada grup d'usuaris. I no només un mapa de navegació individual i per perfil, sinó que tenint en compte els tres nivells d'estudi (de sessió, de semestre i de *lifelong learning*) el mapa que s'està construint és el mapa global del sistema al llarg del temps. Així doncs, els resultats de l'estudi a tres nivells del comportament dels usuaris serà d'utilitat per discernir si a partir d'aquests mapes es pot inferir o construir un model mental dels usuaris del sistema i si és necessària la combinació dels mapes obtinguts a cada nivell per a inferir el model mental. Tal i com s'ha descrit a la secció 2.1.3, la adequació entre el model mental de l'usuari i el model mental del sistema és clau per a que el usuaris tinguin una bona experiència d'ús, a qualsevol nivell.

### **3.3 Validació de la proposta multinivell**

En aquest treball es proposa analitzar l'ús i els usuaris dels sistemes d'e-learning amb una aproximació a tres nivells. La descomposició en nivells afavoreix poder analitzar diferents aspectes amb més detall i el combinar les diferents fonts d'informació disponibles a cada nivell. Els tres nivells, introduïts a la secció 3.2 d'aquest capítol, es defineixen tenint en compte les característiques pròpies del sistema que es vol estudiar, en aquest cas el sistema d'e-learning de la UOC amb els seus tres components principals, els estudiants, els continguts educatius i l'entorn virtual que dóna suport al procés d'aprenentatge.

La interacció dels usuaris amb el sistema d'e-learning es realitza repetida i recurrentment al llarg del temps i sembla adequat establir franges de temps per abordar més acuradament l'estudi i anàlisi. Malgrat el fet que tenir tres nivells sigui intuïtivament raonable, s'ha volgut estudiar més a fons la interacció dels usuaris amb l'entorn virtual d'aprenentatge i esbrinar si l'aproximació multinivell és l'adequada. Per validar-ho s'ha utilitzat la informació disponible sobre els estudiants i la construcció de personatges i escenaris, considerant sempre la naturalesa pròpia dels ensenyaments en línia.

#### **3.3.1 Establiment dels nivells d'anàlisi: personatges i escenaris als sistemes d'e-learning**

Per estudiar i entendre la navegació i comportament dels usuaris i l'ús que fan del sistema, en primer lloc cal estudiar i entendre completament l'entorn en el que interaccionen i naveguen. És interessant realitzar una aproximació holística a l'entorn i situació en que els usuaris interaccionen i, d'aquesta manera, conèixer en profunditat el context i necessitats. La tècnica dels personatges i escenaris, provinent de l'àmbit de la interacció persona-ordinador, permet portar a terme aquesta aproximació i obtenir la informació per construir els nivells d'anàlisi que necessiti cada domini i entorn. Com s'ha dit, en els sistemes d'e-learning és necessari tenir en compte les seves tres dimensions, els usuaris, l'entorn virtual d'aprenentatge i els continguts educatius.

La tècnica dels personatges, presentada a la secció 2.1.2, consisteix en la creació de personatges ficticis basats en les característiques dels usuaris reals. Aquests personatges són una modelització dels usuaris reals del sistema i ajuden a conduir la construcció dels escenaris. Un escenari és bàsicament la descripció d'una persona o personatge interaccionant amb un sistema. Més concretament, un escenari està format per una història personalitzada i fictícia, i els personatges, esdeveniments, productes i entorns que hi intervenen [143]. Les situacions concretes que es generen a partir dels seus elements ajuden al dissenyador a explorar idees i les ramificacions de les decisions de disseny. Els escenaris també s'utilitzen de manera més formal i específica a l'àrea de l'enginyeria del programari i també reben el nom de casos d'ús [167].

Tot i que es tracta d'una tècnica que es va desenvolupar i s'utilitza principalment per enfocar el disseny, els escenaris resulten de gran utilitat per entendre com els usuaris d'un sistema interactiu l'utilitzen i porten a terme les seves tasques, donat que constitueixen una eina d'anàlisi d'experiències. Així doncs, la tècnica dels escenaris aplicada als sistemes d'e-learning, no només considerant la seva vessant tecnològica sinó tenint en compte les seves tres dimensions, ajuda a entendre el sistema i constitueix una potent eina per enfocar i abordar propostes de millora. De fet, els escenaris són útils per desviar el focus de la tecnologia i dirigir-lo als usuaris i a les seves interaccions, millorant d'aquesta manera el disseny de la tecnologia [150]. En el cas dels estudiants cal doncs focalitzar-se en les seves activitats i interaccions amb els continguts i l'entorn educatiu. Per construir escenaris de sistemes d'e-learning és necessari un coneixement sòlid de les tasques que es poden portar a terme a l'entorn virtual d'aprenentatge i del accés i usos dels materials didàctics. Cal descriure amb llenguatge planer la interacció que s'ha de portar a terme i fer menció a la tecnologia quan aquesta representi un aspecte que condiciona la interacció. S'han d'incloure els aspectes rellevants de la interacció incloent el context i, fins i tot, l'actitud de l'usuari.

En el cas concret de la utilització de la tècnica d'escenaris a la UOC, el seu desenvolupament ha estat relativament senzill si es compara amb el desenvolupament de nous sistemes, ja que no ha calgut definir el sistema ni investigar des de zero els seus usuaris, donat que es tracta d'un sistema d'e-learning en funcionament amb un gran

nombre d'usuaris que l'utilitzen a diari i dels quals es disposa informació. Cal tenir en compte que s'ha utilitzat la tècnica no tant per desenvolupar solucions de disseny sinó per entendre les singularitats específiques de l'entorn i dels seus usuaris. Per tant no hi ha hagut necessitat de fer hipòtesis sobre les possibles tasques que els usuaris podrien portar a terme, les situacions en les que es podrien trobar i les necessitats que tindrien per assolir els seus objectius. D'aquesta manera, s'han definit dos personatges que són representatius dels perfils d'usuaris matriculats a la UOC. A partir d'aquesta informació s'han definit els escenaris més representatius i característics, així com d'altres de caire més particular o singular, recollint les situacions, accions i activitats que els usuaris del campus virtual porten a terme de manera habitual, al llarg del temps. Un cop definits personatges i escenaris, s'han agrupat les tasques i els seus objectius tenint en compte l'àmbit i abast temporal i s'ha obtingut una agrupació en tres grans conjunts. La característica principal que propicia aquesta agrupació lligada al concepte de tasca és el període de temps on es situen els objectius, tasques i fites dels estudiants a cada escenari plantejat. D'aquesta manera, s'identifiquen tres intervals de temps: la durada de la sessió d'estudi i navegació a l'entorn virtual d'aprenentatge, la durada del curs o assignatures que ha matriculat l'estudiant i la durada dels seus estudis, no només fins a l'assoliment d'un grau o un títol, sinó al llarg de la vida. Les dades de la UOC i els seus estudiants que s'han utilitzat es poden consultar a l'annex A, el recull o *checklist* de tasques que poden realitzar els estudiants al llarg del semestre acadèmic es presenta a l'annex B i, finalment, el detall de la modelització de personatges i la construcció d'escenaris es troba a l'annex C.

Des d'un punt de vista formal, és important poder validar el que intuïtivament es pensa amb una metodologia com la presentada. És possible proposar una estructuració alternativa a l'anàlisi en tres nivells per a estudiar un entorn virtual d'aprenentatge com el de la UOC. Per exemple, es podria proposar un nou nivell, entre el nivell de sessió i el nivell de semestre, que consideri el mòdul didàctic o la unitat d'aprenentatge com a període de temps, amb uns objectius clarament definits. No obstant, presenta l'inconvenient que al no ser aquest un element estàndard del model educatiu de la universitat, en cada assignatura la unitat d'aprenentatge pot tenir una durada i estructura diferents i, per tant, podria desvirtuar o fins i tot impossibilitar un anàlisi correcte. A més, el nou nivell no es correspondria amb els perfils d'estudiants detectats i descrits a

l'annex C. Cal tenir en compte que la organització del calendari acadèmic semestral imposa una estructura força rígida que no es pot obviar. Fixar més elements en el procés d'aprenentatge podria proporcionar més elements d'anàlisi però restaria la flexibilitat actual del procés d'aprenentatge, molt valorada pels estudiants.

Per altra banda, a cada context i àmbit d'aplicació cal establir el nombre de nivells d'anàlisi a partir del sistema interactiu o entorn virtual, els usuaris, els seus objectius i les tasques que hi realitzaran. Així, en altres àmbits d'aplicació, com per exemple la banca *online*, seria necessari un anàlisi en profunditat del nombre de nivells. Altres entorns virtuals d'aprenentatge d'institucions que ofereixen cursos puntuals i no ofereixen titulacions potser tindrien dos nivells, el de sessió i el de curs, per exemple, o tres si el concepte d'unitat d'aprenentatge està ben definit.

### **3.3.2 Avantatges de l'anàlisi a tres nivells**

L'estudi d'entorns recurrents, com per exemple els sistemes d'e-learning amb un gran nombre d'usuaris, que generen grans volums de dades d'ús al llarg del temps, és un repte que presenta una problemàtica pròpia. Els usuaris interaccionen amb el sistema al llarg del temps generant dades d'ús. Si l'estudi d'aquests sistemes es fa en nivells, tot mantenint la visió general del sistema i de les seves dimensions, es pot portar a terme i permet obtenir resultats interessants, tant pel que fa a la validació o refutació d'hipòtesis acceptades com a vàlides, com pel que fa a proposar-ne de noves.

El principal avantatge de descompondre l'anàlisi dels usuaris i l'ús del campus virtual de la UOC és, tot i que pot semblar evident, que al tractar-se d'un sistema complex i de gran mida, amb un gran volum d'usuaris, poder atacar el problema en parts més manejables, facilita la seva resolució i l'obtenció de resultats.

Des de la perspectiva de l'e-learning, l'anàlisi a tres nivells proporciona informació sobre què fan els estudiants d'entorns virtuals que d'altra manera seria difícil d'obtenir. També permet complementar-les amb altres fonts d'informació sobre els estudiants, com per exemple l'expedient acadèmic, les dades d'accés, la matrícula o les bases de

dades de la universitat. Així doncs, al dividir l'anàlisi en períodes de temps que es corresponen amb els àmbits temporals propis dels processos d'ensenyament i aprenentatge, cada nivell es pot combinar amb la informació acadèmica adequada, amb l'avantatge que aquesta es troba disponible a les bases de dades de la pròpia universitat. Per aquesta raó, el nom que s'ha donat a cada nivell és nivell de sessió al nivell 1 o curt termini, nivell de curs al nivell 2 o mig termini i nivell de *lifelong learning* al nivell 3 o de llarg termini, en funció del període que contempla.

Des de la perspectiva de la interacció persona-ordinador, l'aproximació a tres nivells permet descompondre l'anàlisi d'usuaris en períodes de temps on es poden assumir els seus objectius, complementant les dades amb altres fonts d'informació i mètodes del disseny centrat en l'usuari. El *long term HCI* [142] persegueix observar i analitzar l'evolució dels usuaris al llarg del temps i proporcionar els indicadors a tenir en compte per poder gestionar el canvi d'usuaris novells a usuaris experts. Entendre la transició d'usuaris novells a usuaris experts és important donat que molts cops no es produeix el canvi a expert, i els usuaris experimenten frustració i són més propensos a abandonar l'ús del sistema. A [117] s'ha presentat una aproximació a l'anàlisi del *long term HCI* en sistemes d'e-learning utilitzant els nivells d'anàlisi. Així, el nom que s'ha donat a cada nivell és *session* al nivell de curt termini, *activity* al nivell de mig termini i *wish* al nivell de llarg termini, seguint en aquest cas una terminologia més propera a la interacció persona-ordinador.

Així doncs, depenent de la perspectiva de l'e-learning o de la IPO els mateixos nivells reben diferents noms. A la Taula 3.1 hi ha el sumari dels noms donats a cada nivell.

Nivell	Període de Temps	E-Learning	IPO
1	Curt termini	Sessió	<i>Session</i>
2	Mig termini	Curs	<i>Activity</i>
3	Llarg termini	<i>Lifelong learning</i>	<i>Wish</i>

Taula 3.1: Noms que reben els diferents nivells.

A cada nivell es coneixen, de manera no individualitzada i depenent del cas amb major o menor detall, els objectius dels usuaris. Es tracta d'objectius assumits, donat que no s'està considerant individualment a cada usuari, però sí es coneixen a cada nivell, tal i com es mostra a l'annex B i a l'annex C, quins objectius i fites persegueixen i quines necessitats tenen. Això és de gran ajut per conduir l'anàlisi, sobre tot quan es basa en el processament semiautomatitzat de grans volums de dades.

El campus virtual de la UOC és un entorn virtual d'aprenentatge que evoluciona i, per tant, s'hi introdueixen canvis i millores. Alguns d'aquests canvis tenen una repercussió directa en els elements externs i d'interfície d'usuari i, d'aquesta manera, al llarg del temps poden aparèixer i desaparèixer botons i enllaços de primer nivell. Normalment, aquests canvis es produeixen en el període entre dos semestres acadèmics, tot i que aquest fet desitjable no sempre és així per altres criteris de producció. Per tant, a l'hora de dissenyar experiments amb les dades, és recomanable considerar el semestre com a unitat de mesura principal. Si es repeteix un mateix anàlisi en un altre semestre és important revisar si s'han produït canvis i adaptar el processat de les dades conseqüentment.

A la Figura 3.1 s'ha presentat la relació entre els nivells i el fil conductor que dona coherència a l'anàlisi. Ara que s'han introduït amb més detall els nivells, és interessant visitar la figura i fer notar els elements de cada nivell que afavoreixen la relació entre els mateixos. Així, al primer nivell s'obtenen les sessions de navegació dels usuaris i del processament d'aquestes sessions s'obtenen patrons. Al segon nivell es fixa l'àmbit temporal a la durada d'un semestre acadèmic i les sessions i patrons obtinguts al primer nivell s'agrupen en cursos. Al tercer nivell hi ha agrupacions de cursos i agrupacions d'aquests cursos en graus. A continuació es presenta de forma detallada cadascun dels nivells identificats.

### **3.3.3 Nivell 1: Curt termini**

En el primer nivell es recull i analitza l'activitat dels usuaris cada cop que es connecten al sistema. Aquesta connexió i navegació pel campus virtual rep en la literatura el nom

de sessió [140]. Tractant-se d'un entorn virtual d'aprenentatge al qual s'accedeix mitjançant Internet, les dades que es recullen i s'utilitzen en aquest nivell es basen en el format comú de fitxers de log d'Apache i principalment proporcionen informació sobre l'adreça IP, la data i hora, l'ítem sol·licitat, l'ítem accedit i d'altres menys rellevants com el navegador i el sistema operatiu. Les entrades d'aquests fitxers tenen un aspecte com aquest:

```
3.33.216.136 - - [16/Oct/2007:09:19:31 +0200] "GET /cgi-bin/avis?s=c8edf7f1949e7bc728e5db2e7aaadbcaf4e753a5003c9db622870a9f56d1e66e220c92112505b81d1a640d0dd772ddfc79c99f861c33d2f79eb415d39b857625&appid=UOC2000&lb=b HTTP/1.1" 200 898 "http://cv.uoc.edu/cgi-bin/uocapp?s=c8edf7f1949e7bc728e5db2e747fdbcaf4e753aaaa3c9db622870a9f56d1e66e220c92112505b81d1a640d0dd772ddfc79c99f861c33d2f79eb415d39b857625" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)"
```

```
83.49.176.17 - - [16/Oct/2007:09:19:32 +0200] "GET /UOC/a/extcgi_0.html?s=dbe6b0d4f8baf435975a80d96e2ecc81a790db9b734773337d52588bc6d129dda3bbe938732d4bee62a7ef2a13290c84c908dd28f0defffe79e7f20b157dbe81&img=hola&cgi=/cgi-bin/hola HTTP/1.1" 200 1740 "http://cv.uoc.edu/cgi-bin/uocapp?s=dbe6b0d4f8baf435975a80d96e2ecc81a790db9b734773337d52588bc6d129dda3bbe938732d4bee62a7ef2a13290c84c908dd28f0defffe79e7f20b157dbe81" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)"
```

```
83.49.176.17 - - [16/Oct/2007:09:19:32 +0200] "GET /UOC/a/mc-icons/labels/banner_main.gif HTTP/1.1" 304 - "http://cv.uoc.edu/UOC/a/menu.htm" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)"
```

Els resultats que s'espera obtenir amb l'anàlisi es situen principalment en dos àmbits. En primer lloc es vol obtenir un resultat d'anàlisi d'ús, obtenint els camins de navegació dels usuaris i proposant agrupacions que proporcionin patrons. En segon lloc es vol relacionar aquest primer resultat de patrons amb dades de caire més qualitatiu que provenen del coneixement de l'entorn i el context. Evidentment això es complementa



amb uns descriptors estadístics que permeten conèixer millor la realitat sobre el concepte de sessió.

### **Sessions d'un usuari i patró de navegació**

L'anàlisi d'ús i de comportament de navegació es realitza mitjançant el processat de les dades recollides i calculant el vector corresponent a la navegació d'un usuari durant una sessió. De fet, el resultat d'aquest primer filtratge i anàlisi són les diferents sessions de navegació de cada usuari. D'aquesta manera, amb les entrades dels fitxers de log es pretén obtenir camins de navegació d'aquest tipus:

```

Usuari01 2007-10-17 09:09:01 09:09:27 Login,Bustia,Sortir
Usuari02 2007-10-17 09:06:32 09:17:30 Login,Bustia,Aules
Usuari03 2007-10-17 09:02:18 09:04:55 Login,Secretaria,Agenda,Sortir
Usuari04 2007-10-17 09:04:39 09:25:55 Login,Bustia,Aula-Planificacio,
Biblioteca
Usuari05 2007-10-17 09:06:26 09:06:45 Login,Bustia
Usuari06 2007-10-17 09:07:40 09:10:32 Login,Aula-Planificacio,Aula-
Comunicacio,Debat,Sortir
Usuari07 2007-10-17 09:08:29 09:08:33 Login,Bustia,Sortir

```

Aquests camins i les marques de navegació són simulats. No obstant, serveixen d'exemple per mostrar l'objectiu que es persegueix quan es parla de capturar la navegació dels usuaris a partir dels fitxers de log.

El procediment per a obtenir els camins de navegació de cada usuari del campus virtual és llarg i presenta certa complexitat. Per una banda requereix gestionar i preprocessar els fitxers de log, tal i com es descriu a la secció 4.5 i per altra banda inclou el processament d'aquests fitxers, tal i com es presenta a la secció 4.6.2. Un cop obtingudes les sessions de navegació, es vol estudiar com, utilitzant tècniques de mineria de dades, es poden obtenir agrupacions, patrons i altres informacions útils. Al capítol 4 es presenta l'anàlisi realitzada amb la informació de les sessions i camins de navegació.

## **Agrupació de patrons dels usuaris**

Un altre resultat interessant és el que s'obté agrupant comportaments similars d'usuaris en base a les característiques de les sessions de navegació que realitzen. Les característiques, és a dir, els elements de cada sessió de navegació que s'han considerat caracteritzadors, són molt diversos. Entre d'altres es poden destacar la durada, el dia de la setmana en que té lloc, la franja horària, el moment del dia, la primera acció (o primer clic en el campus virtual) i el recorregut realitzat.

Així doncs, com a resultat de l'anàlisi en aquest primer nivell es poden obtenir patrons que poden ser de dos tipus:

- Patrons de navegació d'un usuari (en un interval de temps).
- Patrons de navegació d'un grup d'usuaris amb sessions de navegació similars (en un interval de temps).

## **Distribució dels patrons al llarg del temps**

La informació que proporcionen els camins de navegació dels usuaris i les seves possibles agrupacions és interessant que es complementin amb altres fonts d'informació i també amb dades de caire més qualitatiu. És interessant veure, per exemple, com és la distribució dels patrons al llarg del temps tenint en compte el coneixement que es té del moment del curs acadèmic en el que es produeixen les sessions i tenint en compte també l'objectiu assumit d'aquell moment.

Durant el curs es distingeixen tres moments clau: els dies d'inici de curs, les setmanes durant el semestre o de desenvolupament del curs i els dies que van de la fi de semestre fins a la publicació de les notes. En un context educatiu com el de la UOC, aquests tres moments a tenir en compte venen definits pel calendari acadèmic de la universitat. El seu impacte sobre els estudiants s'ha identificat a partir de les dades d'activitat del campus virtual i de la investigació d'usuaris del projecte Campus en Programari Lliure [19]. Aquests tres moments clarament propicien actituds i comportaments diferents en els estudiants, especialment els d'inici i fi de curs. El període de desenvolupament del

curs és el més variable en funció del conjunt d'estudiants que s'analitza i, especialment, de cada assignatura, la seva temàtica, de les activitats previstes i de la seva programació i planificació.

Els primers dies de curs, a l'inici del semestre, els estudiants es connecten al campus virtual per comprovar que tenen accés a les aules de les assignatures de les quals s'han matriculat. Les entrevistes realitzades amb estudiants mostren que un cop han accedit a les aules, els dos espais principals de consulta són l'espai de planificació i l'espai de comunicació de l'aula, que es descriuen a la secció 4.2. A l'espai de planificació els estudiants hi accedeixen per imprimir el pla docent i consultar el calendari de l'aula, per obtenir així la planificació d'activitats durant les setmanes que dura el curs. A l'espai de comunicació hi accedeixen per llegir els missatges de benvinguda del consultor i dels companys de l'aula i per escriure el seu propi missatge de presentació, tot i que això últim no ho fan tots els estudiants.

Durant el període de desenvolupament del curs, els estudiants accedeixen al campus per fer el seguiment de les assignatures de les que estan matriculats. Aquest seguiment implica, principalment, estar al dia dels missatges de les bústies compartides de les aules i la realització de les activitats previstes al pla docent. Algunes de les activitats previstes són avaluable i reben el nom de PAC (Prova d'Avaluació Continuada). Les setmanes on hi ha lliurament de PAC d'una assignatura, augmenta l'activitat en la seva aula, tant a nivell de connexions com de missatges al fòrum. Així es pot observar en les dades d'ús de l'aula i també ho afirmen els propis estudiants quan se'ls hi pregunta al respecte. Els estudiants molts cops reconeixen que fan un seguiment de les assignatures segons el lliurament de les PAC previstes i, d'aquesta manera, per cada PAC dediquen aquella setmana a la seva resolució, deixant de banda les altres assignatures. Els estudiants mateixos posen de manifest que el problema apareix quan en una setmana coincideixen PAC de diferents assignatures.

Els darrers dies de curs poden ser claus en diferents vessants. Tots els estudiants estan pendents de les notes finals d'avaluació continuada que és la que determina si s'han de realitzar un examen final, una prova de validació o no rés. Així, hi ha estudiants que han de realitzar proves finals i les han de preparar. Això augmenta la interacció a l'espai de

comunicació de l'aula i els missatges directes als consultors. Un cop realitzades les proves finals, l'activitat a l'aula baixa fins casi desaparèixer. Els dies de publicació de notes finals torna a haver una alta taxa de connexió, tot i haver molts estudiants que sol·liciten rebre les notes amb un avís al telèfon mòbil, el que fa que no els calgui connectar-se al campus virtual.

L'anàlisi dels usuaris a tres nivells fa que a cada nivell es pugui realitzar un processament i anàlisi de dades independent, alhora que no es pot perdre de vista que es tracta d'un procés global. D'aquesta manera, el processament i anàlisi de dades en el primer nivell es veu afectat pels objectius del segon i tercer nivell, en el sentit que el conjunt d'assignatures que es decideixi analitzar en el tercer nivell, per exemple, filtra el conjunt de dades a utilitzar en el segon i primer nivells. Al mateix temps, les dades del segon i tercer nivell ajuden a caracteritzar i explicar les navegacions observades en el nivell de sessió. Una de les utilitats principals dels resultats obtinguts en aquest primer nivell d'anàlisi és l'avaluació de la usabilitat i la millora del disseny del sistema. Així, la navegació real que porten a terme els usuaris del campus virtual es pot comparar amb la navegació teòrica que haurien de portar a terme per a realitzar les tasques previstes i, d'aquesta manera, detectar els aspectes a millorar. Conèixer la navegació dels usuaris a nivell de sessió també permet adaptar alguns aspectes del procés d'aprenentatge, tal com es proposa a la secció 3.7 d'aquest capítol.

### **3.3.4 Nivell 2: Mig termini**

En el segon nivell es recull i analitza l'activitat dels estudiants durant la realització d'una assignatura. A la UOC, aquest període de temps correspon a un semestre acadèmic i té una durada aproximada d'unes catorze setmanes, tot depenent del calendari establert en cada curs acadèmic.

Les dades que s'utilitzen en aquest nivell també provenen dels fitxers de log del campus virtual, així com de les bases de dades amb la informació acadèmica dels estudiants. Aquestes bases de dades contenen el registre de les assignatures matriculades i de les

activitats acadèmiques realitzades i les seves qualificacions. Una altra font de dades són les enquestes de satisfacció que es realitzen semestralment.

Des de la perspectiva de la interacció persona-ordinador, aquest nivell és el d'activitat i, per tant, es poden assumir alguns dels objectius dels usuaris. Des de la perspectiva de l'e-learning es pot saber alguns d'aquests objectius; principalment, l'objectiu que persegueixen els estudiants en aquest nivell és l'assoliment de les competències i objectius de les assignatures que realitzen, dit d'una altra manera, aprovar o superar les assignatures de les quals estan matriculats. En el cas de l'entorn virtual d'aprenentatge de la UOC, aquest objectiu s'assoleix portant a terme les tasques recollides a l'annex B.

En aquest nivell es complementa la informació obtinguda en el primer nivell, ja que es disposa d'informació de cada assignatura, concretament de la seva estructura, planificació i les seves tasques previstes. A més, aquest nivell aporta una informació qualitativa molt important com són les qualificacions d'avaluació continuada que cada estudiant obté durant el semestre, així com la qualificació de l'avaluació final de l'assignatura. Així doncs, l'objectiu assumit d'aquest segon nivell està directament vinculat amb les qualificacions.

L'anàlisi que es realitza en aquest nivell ha de portar a descobrir si existeix una correlació entre els resultats acadèmics i els camins de navegació dels estudiants. Per una mateixa assignatura s'haurien de considerar per una banda els camins dels estudiants i, per l'altra, les qualificacions obtingudes. Si es descobreix que hi ha una correlació, es poden prendre mesures proactives per potenciar els comportaments que afavoreixen l'èxit acadèmic, així com establir mesures que permetin evitar o disminuir el fracàs.

La distribució de connexions al campus virtual al llarg del semestre acadèmic és una informació interessant d'aquest nivell que es pot calcular a partir de les dades de navegació obtingudes en el nivell anterior. Aquesta informació serà útil tant des del punt de vista de l'estudi de l'ús del sistema com del disseny instruccional de les assignatures. L'anàlisi de la distribució de connexions és interessant realitzar-lo tant a nivell d'un usuari individual com de tots els estudiants matriculats. Aquesta distribució

de connexions pot presentar un patró que a més cal veure si està relacionat amb les qualificacions obtingudes i amb la superació de l'assignatura.

Una altra anàlisi que es pot realitzar en aquest segon nivell consisteix en comparar, per cada estudiant, els seus patrons de navegació en les diferents assignatures que ha matriculat aquell mateix semestre. Pot ser interessant veure si hi ha consistència de patrons entre diferents assignatures o si hi ha diferències significatives. Si, a més, es té en compte el rendiment acadèmic, es pot obtenir un coneixement important sobre els estudiants. En funció dels objectius plantejats, es poden dissenyar experiments ad-hoc per així veure si les hipòtesis plantejades són certes o no.

L'estudi de les dades en aquest nivell de semestre acadèmic també permet comparar diferents semestres entre ells i així poder explicar-los. Els patrons de navegació i comportament dels usuaris pot ajudar, per exemple, a explicar diferències del rendiment acadèmic d'una mateixa assignatura entre diferents semestres acadèmics.

En aquest nivell de mig termini s'estudia i analitza tot allò que succeeix en una assignatura o al voltant d'ella. Es podria dir que l'estudi que es porta a terme és *intra-assignatura* o a dins de cada assignatura, i no es tenen en compte les altres informacions. L'estudi de tot allò que passa entre les diferents assignatures es realitza en el tercer nivell, per exemple, obtenir les agrupacions d'assignatures que es matriculen cada semestre acadèmic i veure si es segueix un patró al llarg del temps. L'activitat entre assignatures també té una repercussió al que passa en cada assignatura, per aquest motiu és important mantenir en tots els nivells una visió completa de l'anàlisi que es realitza.

Els resultats que s'obtenen en aquest segon nivell d'anàlisi tenen una aplicació molt directa en el procés d'aprenentatge de l'estudiant, ja que permet detectar si el comportament d'estudiants concrets s'allunya del disseny general de l'assignatura i, d'aquesta manera, prendre les decisions adequades. Per exemple, tal i com es presenta a la secció 4.7, es pot estimar la probabilitat que un estudiant no lliuri la resolució d'una PAC i avisar, si s'escau, al seu professor per a que realitzi les accions que consideri convenients.

### 3.3.5 Nivell 3: Llarg termini

En el tercer nivell es recull i analitza l'activitat dels usuaris a llarg termini i té en compte tot el temps que dura la seva relació amb la institució i amb l'entorn virtual. En un entorn d'e-learning això significa que es té en compte tot el procés d'aprenentatge que es realitza al llarg de la vida, és a dir, les diferents assignatures, cursos i accions formatives que es realitzen en cada moment.

Les dades que s'utilitzen en aquest nivell provenen principalment de les bases de dades de la universitat i del campus virtual, tot i que també s'utilitzen dades dels nivells 1 i 2. Pròpiament, les dades amb les que es treballa no són tant dades d'ús sinó dades que més aviat es podrien anomenar de resultat.

El principal processament i anàlisi que es pot portar a terme en aquest nivell consisteix en analitzar les agrupacions d'assignatures que es matriculen en cada semestre i, a partir d'aquestes dades, buscar les relacions amb el rendiment acadèmic i els canvis que es produeixen al llarg del temps. D'aquesta manera per cada estudiant o grup d'estudiants es pot calcular el seu camí de matriculació o de navegació entre assignatures. Cadascun dels camins obtinguts es pot relacionar amb el rendiment acadèmic i estudiar els camins que porten a l'èxit o fracàs des del punt de vista acadèmic. Així, es poden obtenir les combinacions d'assignatures que fan que els estudiants es formin al llarg de la vida, així com aquelles combinacions que afavoreixen el *dropout* o abandonament dels estudis.

Els resultats que s'obtenen en aquest nivell són importants per conèixer l'adequació tant dels plans d'estudis com de les recomanacions de matrícula a les necessitats de formació dels estudiants que queden reflectides en cada matrícula que realitzen. També s'obté informació rellevant sobre aquelles assignatures que, realitzades conjuntament, dificulten el procés d'aprenentatge dels estudiants durant un o més semestres acadèmics. A la secció 4.8 es presenten els experiments realitzats i els resultats obtinguts en aquest nivell d'anàlisi a llarg termini.

### 3.4 Processament i anàlisi de logs en entorns virtuals d'aprenentatge

Una de les qüestions relacionades amb el disseny de sistemes interactius en general i amb el disseny d'entorns d'e-learning en particular és l'obtenció d'informació sobre l'ús real que en fan els usuaris. No sempre és senzill monitoritzar què fan els usuaris i les tasques i navegació que porten a terme. El disseny instruccional, la pedagogia i la usabilitat són elements que es tenen en compte al dissenyar entorns virtuals d'aprenentatge, però és capturant i analitzant l'ús que es fa del sistema com es pot proporcionar informació per determinar si els usuaris se senten còmodes amb l'entorn i si són capaços de dur a terme les tasques relacionades amb el procés d'aprenentatge. A més, la informació obtinguda es pot utilitzar per introduir elements de personalització al sistema.

Els fitxers de log poden ser de gran utilitat per capturar el comportament i activitat dels usuaris [93]. Malgrat els inconvenients que presenten i que s'han descrit en detall a la secció 2.3, poden resultar de gran interès en entorns virtuals d'aprenentatge i més encara si aquests donen servei a un gran volum d'estudiants [122].

La qüestió principal en relació a recollir i analitzar el comportament dels usuaris en entorns virtuals d'aprenentatge amb un nombre elevat d'estudiants és el gran volum de dades disponibles de diferents fonts i la manca de dades específiques per portar a terme una anàlisi concreta. En un entorn interactiu web, els usuaris interaccionen amb el sistema mitjançant un navegador web. Examinant els fitxers de log generats pels servidors web és possible obtenir una certa traça de la interacció dels usuaris amb l'entorn. Hi ha diferents aproximacions a obtenir aquesta traça, essent el *web mining* la que proporciona resultats més interessants [53]. Tanmateix, poden aparèixer inconvenients quan el lloc web a estudiar és complex, tant a nivell de mida com d'estructura interna. A més dels logs d'accés generats pels servidors web, en un entorn virtual d'aprenentatge com el de la UOC es generen altres tipus de logs. Aquests logs es corresponen amb els principals elements o capes en que s'estructura el campus virtual; aquestes capes són els servidors web, la lògica de negoci i el model de dades i es descriuen a la secció 4.2.1.



El processament i anàlisi de logs resulta de gran interès a l'entorn virtual d'aprenentatge de la UOC. El motiu principal és que amb l'estructura i configuració actuals del campus virtual, el processament i anàlisi de logs és la única manera d'obtenir informació de l'ús real del sistema i del comportament dels usuaris. Una aproximació mitjançant mètriques remotes requeriria, a més de la introducció del codi Javascript corresponent en les planes del campus virtual, la infraestructura i recursos necessaris per emmagatzemar i gestionar les entrades generades pels milers d'usuaris del campus virtual. En qualsevol cas, això no significa que no es pugui construir o contractar un sistema de mètriques remotes per la UOC, i correspon a una decisió estratègica de la institució. Un altre objectiu d'aquest treball ha estat, doncs, proposar un sistema de captura i anàlisi de l'activitat dels usuaris, fent ús de la infraestructura tecnològica existent i dels recursos econòmics disponibles. Altres eines comercials com Google Analytics [9] o Indextools [13] no són adequades per a fer el seguiment individualitzat de milers d'usuaris en un entorn tan complex.

Cada acció que realitza un usuari del campus virtual es veu reflectida i registrada en una o més entrades en els fitxers de log generats pels servidors. A més, depenent del tipus d'acció, diferents servidors poden guardar el registre de la mateixa acció però utilitzant informació diferent. En aquest treball s'han utilitzat principalment els fitxers de log generats pels servidors web Apache que actuen com a *front-ends* del campus virtual. Els fitxers de log generats pels servidors web del campus virtual de la UOC presenten la problemàtica i inconvenients habituals dels fitxers de log que s'han presentat detalladament a la secció 2.3: mida gran, fitxers incomplets o que presenten informació incompleta i, amb menys mesura, la identificació dels usuaris, donat que el campus virtual identifica unívocament cada sessió que es realitza.

Un dels principals inconvenients que sorgeix a l'estudiar els logs generats per servidors web és el de la identificació dels usuaris. Aquest inconvenient apareix parcialment en un entorn virtual d'aprenentatge com el de la UOC, ja que al tractar-se d'un entorn recurrent amb usuaris identificats, cada usuari té un identificador que es manté en el temps. D'aquesta manera, tots els usuaris que accedeixen a l'entorn d'aprenentatge estan registrats i al entrar al campus virtual es genera un número de sessió únic que els identifica durant aquella sessió de navegació. Per tant, és possible no només identificar

els usuaris individuals sinó també cada sessió de navegació concreta. No obstant això, el numero de sessió no s'utilitza sempre, sinó només en aquelles crides que es consideren crítiques de cara a garantir la identitat de l'usuari. D'aquesta manera, es produeixen crides al sistema i, per tant, entrades en els fitxers de log on no apareix el numero de sessió de l'usuari i, conseqüentment, no es pot analitzar la traça de la seva navegació. Per fer seguiment de la navegació en les àrees on no s'utilitza el número de sessió, les adreces IP també s'han de descartar ja que hi ha la possibilitat que diferents usuaris accedeixin al campus virtual amb el mateix servidor *proxy*, la qual cosa emmascara les seves adreces IP reals.

Altres dels inconvenients habituals en l'anàlisi de fitxers de log, de fet la majoria, continuen apareixent: la grandària dels fitxers, les habituals errades de registre d'entrada, la impossibilitat de registrar certes accions dels usuaris quan s'utilitzen altres tecnologies a més de HTML en la mateixa plana, com per exemple Javascript, JSP, etc.

En el campus virtual de la UOC, els fitxers de log es generen a diari pels diferents servidors web. Aquests fitxers es copien en un altre servidor on són fusionats en un únic fitxer i on s'apliquen els filtres que es descriuen a continuació. Un cop s'ha obtingut un únic fitxer amb tots els logs del campus virtual corresponents a un dia, aquest fitxer es preprocessa amb l'objectiu d'eliminar totes aquelles entrades que no són generades directament pels usuaris, com la càrrega de imatges, icones o elements d'estil. Aquest preprocessat redueix la quantitat de línies en un 90%. No obstant, si es considera el període de temps d'una setmana, el nombre de línies que han de ser processades es troba encara al voltant dels 36 milions, aproximadament 12 GB, que en qualsevol cas és un número molt alt. Cal notar que aquestes dades poden variar sensiblement depenent del moment concret del calendari acadèmic. Per aquesta raó és necessari realitzar un segon preprocessat i acostar més les dades a l'anàlisi que es vol realitzar, tal com es descriu a [57]. A més, com es descriu a [95], no és possible aplicar tècniques clàssiques de mineria de dades si la quantitat d'entrades vàlides a analitzar no es redueix considerablement i s'utilitza un sistema distribuït.

Com s'ha esmentat anteriorment, quan els usuaris naveguen per les àrees de l'entorn virtual d'aprenentatge que no requereixen identificació, no es pot seguir la seva traça

adequadament i, per tant, les línies de log corresponents a aquesta navegació no aporten informació útil i s'eliminen, afegint aquest pas al preprocessat de les dades.

La gestió i preprocessat dels fitxers de log del campus virtual de la UOC es presenta detalladament a la secció 4.5. Bàsicament consisteix en rebre els fitxers de log corresponents a cada front-end, aplicar els filtres previstos i fusionar-los en un únic fitxer per cada dia. Aquests fitxers preprocessats constitueixen el punt de partida de l'anàlisi d'activitat i navegació dels usuaris.

Si bé aquests fitxers estan preprocessats i preparats per l'anàlisi, la seva mida resulta encara excessiva i contenen molta informació irrellevant de cara a analitzar-los. A més, encara poden presentar forats en el registre de l'activitat de l'usuari deguts a l'arquitectura de l'aplicació que produeix entrades sense identificació dels usuaris i altres entrades que trenquen la estructura clàssica dels fitxers de log degut a les tecnologies que s'utilitzen.

Tot i que el principal inconvenient del processament i anàlisi de logs d'entorns web desapareix parcialment en el cas de l'entorn d'aprenentatge de la UOC, els experiments realitzats i presentats al capítol 4 mostren que usar únicament els fitxers de log no proporciona informació suficientment rica per descobrir la navegació dels usuaris del campus virtual. En aquests experiments s'han processat els fitxers de log del campus virtual, s'ha reduït la seva mida, s'ha estandarditzat la identificació de l'usuari i s'ha convertit el format del fitxer en un format clàssic. Malgrat això, s'ha comprovat que les habituals eines d'anàlisi de llocs web al processar aquests fitxers de log no generen resultats correctes o útils. El motiu principal és la complexitat pròpia del campus virtual, tant a nivell de d'estructura i mida, com dels serveis que ofereix i les tecnologies que utilitza.

Per aquesta raó, en lloc d'analitzar els fitxers de log en brut tal i com es generen en els servidors, s'ha decidit establir un sistema amb un conjunt mínim de marques que s'introdueixen en llocs estratègics del campus virtual, que es consideren més adients de cara a l'anàlisi que es vol portar a terme. La utilització de marques persegueix un doble

objectiu, capturar l'activitat i navegació reals dels usuaris i reduir el volum de dades necessari per l'anàlisi.

### **3.5 Proposta de marcatge**

Amb la intenció de trobar una solució als inconvenients i dificultats de l'anàlisi de fitxers de log es proposa utilitzar un sistema que proporcioni informació addicional. Tenint en compte la informació que proporcionen els fitxers de log, hi ha una necessitat real d'enriquir-la amb un sistema transparent, que no interfereixi amb el funcionament del campus virtual, que permeti seguir l'activitat dels usuaris i que simplifiqui el processament dels fitxers de log.

El sistema que es proposa utilitzar consisteix en col·locar marques en diferents punts de l'entorn virtual d'aprenentatge i així registrar de manera unívoca la navegació de l'usuari. Aquestes marques s'integren amb el propi sistema web de manera que quan un usuari accedeix a un enllaç o carrega una plana, el propi sistema deixa registrada la marca corresponent en les entrades dels fitxers de log, complementant-les. Per exemple, d'aquesta manera és possible saber per on accedeix a la bústia l'usuari un cop ha accedit a l'entorn, donat que la plana d'inici del campus virtual li ofereix més d'una possibilitat.

Aquesta aproximació al seguiment de l'activitat dels usuaris i l'anàlisi d'ús es basa en els fitxers de log i s'inspira en les mètriques remotes, en el fet que s'introdueix un senyal en aquells llocs que es volen monitoritzar i es genera una entrada quan l'usuari hi accedeix. Tanmateix, no es pot considerar com una aproximació híbrida de fitxers de log i mètriques remotes, ja que es recolza completament en els fitxers de log i no requereix nova infraestructura per emmagatzemar les entrades generades.

Per introduir marques per capturar i analitzar el comportament dels usuaris apareixen dos qüestions importants: el coneixement del lloc web i la metodologia de marcatge. Conèixer en profunditat el lloc web o l'entorn virtual que es vol marcar i analitzar és fonamental, per una banda per saber quines àrees i serveis són les importants i de les quals es vol conèixer l'ús que se'n fa i, per altra banda, per saber interpretar

adequadament la informació que s'obté mitjançant les marques. La metodologia de marcatge estableix en primer terme obtenir una estratègia de marcatge i així definir què es marca i què no, què es vol obtenir amb aquella marca i la manera en que es porta a terme.

### 3.5.1 Metodologia de marcatge

Un dels resultats més interessants que s'obtenen de la proposta de marques integrades en el lloc web és que proporciona un mapa senzill però realista dels serveis disponibles en l'entorn virtual d'aprenentatge i com aquests serveis són utilitzats pels estudiants o pels usuaris de l'entorn web, en general. Sense aquest mapa, qualsevol esforç per entendre el comportament dels usuaris representa un gran cost degut a la necessitat d'identificar les entrades específiques, relacionades amb l'activitat a monitoritzar, en els fitxers de log. Fins i tot reproduir experiments ja realitzats amb èxit [44] amb noves dades i nous fitxers de log, esdevé una tasca molt complexa si l'entorn virtual d'aprenentatge ha canviat encara que sigui mínimament, cosa que és freqüent en entorns web complexos, com és el cas del campus virtual de la UOC.

Com a conseqüència d'això, la proposta de marques que es porti a terme s'ha de fer d'acord als següents requeriments: independència de l'arquitectura, independència tecnològica, jerarquia, automatització, escalabilitat i consistència.

- Independència de l'arquitectura: tots els canvis en el sistema web s'han de fer únicament en la capa de presentació, sense afectar la capa de lògica de negoci. Així, la integració de marques no hauria d'interferir amb el correcte funcionament del sistema.
- Independència tecnològica: qualsevol canvi en els aspectes tecnològics de la capa de lògica de negoci no hauria d'interferir amb la interfície web del sistema i, per tant, amb el sistema de marques integrades.
- Jerarquia: per tal de poder focalitzar o generalitzar els anàlisis que es porten a terme, s'haurien de poder activar i desactivar fàcilment arbres i subarbres de

marques. Per exemple, d'aquesta manera es podrien activar tot el conjunt de marques de les aules virtuals en un moment concret.

- Automatització: el sistema de marques integrades en un sistema web complex, amb centenars de planes, s'hauria de poder modificar de manera automatitzada amb un *script* o un panel de control ad-hoc.
- Escalabilitat: la proposta de marques integrades ha de ser independent de la mida del lloc web que es vol monitoritzar.
- Consistència: tota la informació rellevant necessària per portar a terme l'anàlisi es troba en els fitxers de log, que és on queden enregistrades les marques, i no són necessàries altres fonts de dades.

Per a estudiar el comportament dels usuaris d'un lloc web és necessari establir a priori i de manera raonada quins elements i accions seran monitoritzats i posteriorment analitzats. En un entorn web que ofereix als seus usuaris un ampli ventall de continguts i serveis, serà necessari identificar els objectius dels usuaris, les principals tasques que porten a terme i les possibles accions relacionades amb aquestes tasques. La interacció persona-ordinador i el disseny centrat en l'usuari proporcionen mètodes adequats [63, 129] per portar-ho a terme. D'aquesta manera, gràcies per exemple a l'anàlisi de tasques o al recorregut cognitiu (veure secció 2.1) es pot obtenir la seqüència de passes i accions que un usuari ha de realitzar interaccionant amb la interfície del sistema web per tal d'assolir un objectiu o portar a terme una tasca concreta. Prenent aquesta seqüència d'accions com a punt de partida i amb el coneixement de les àrees i serveis que es vol analitzar, es pot establir un primer conjunt de marques per a generar un primer mapa de l'entorn web. A partir d'aquesta primera versió del mapa, es porta a terme un nou estudi del lloc web i de les marques generades, intentant descobrir si s'ha omès la descripció d'alguna tasca. Aquest procés es va repetint fins a aconseguir un mapa de marques adequat. A continuació, a la Figura 3.3, es mostra el mapa amb els espais, serveis i accessos principals del campus virtual de la UOC que es volen monitoritzar i analitzar. La figura també mostra el principal diagrama de navegació donat que indica els *links* o enllaços que es consideren més importants.

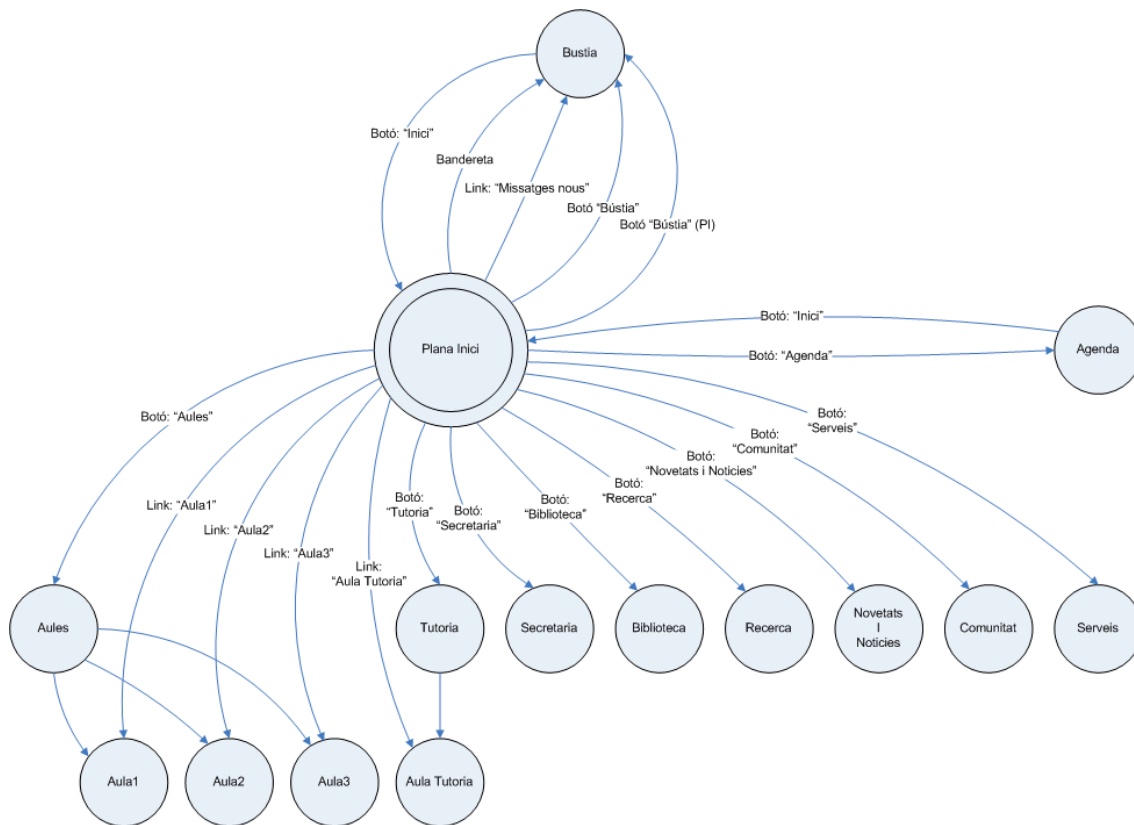


Figura 3.3: Diagrama d'espais i enllaços principals del campus virtual.

En el cas concret del campus virtual de la UOC, això implica que cal integrar marques en els elements del menú lateral i superior i també en alguns dels enllaços o accessos directes que apareixen a la plana d'inici (veure Figura 4.2). Aquestes marques, tal i com s'ha dit, han de deixar registrats a les entrades dels fitxers de log els serveis i espais visitats. Per tant, cal definir els identificadors de cada espai i de cada enllaç a fer seguiment. Tota aquesta informació queda recollida al mapa de marques del lloc web. El mapa de marques detallat del campus virtual de la UOC que s'ha definit i usat en aquesta tesi està disponible a l'annex D.

A partir dels resultats obtinguts en els anàlisis preliminars, el coneixement que es té del campus virtual de la UOC i dels espais, serveis i accessos que es volen monitoritzar, s'han definit dues estratègies de marcatge complementàries i se les ha anomenat estratègia basada en l'acció i estratègia basada en el contingut. Les dues estratègies es combinen per proporcionar la informació necessària per a l'anàlisi posterior.

### 3.5.2 Estratègia basada en l'acció

Per cada clic que un estudiant realitza en un element crític pel procés d'aprenentatge es genera una crida en la capa de lògica de negoci, que genera una o més línies en els fitxers de log dels servidors web. El camí més senzill per capturar aquesta acció és afegir un paràmetre especial d'acord al context on aquest paràmetre apareix.

Una proposta senzilla és afegint una cadena com “&ACCIO=nom” a cada enllaç rellevant de l'entorn virtual, on ‘nom’ és l'identificador de l'acció que realitza l'usuari al seleccionar a l'enllaç. Això provoca que aquesta cadena quedi registrada en les entrades dels fitxers de log quan l'usuari selecciona aquella acció. Aquesta cadena s'introdueix simulant que es tracta d'un paràmetre més en les crides ja existents en els enllaços, per aquest motiu va precedida del símbol “&” i s'utilitza un nom de paràmetre que no interfereixi amb el funcionament normal del sistema. D'aquesta manera, totes les accions on s'ha afegit la cadena apareixen als fitxers de log i resulta relativament senzill fer-ne un seguiment.

D'aquesta manera, per cada acció que es consideri rellevant s'ha d'introduir una cadena que identifiqui aquella acció. La política de noms que s'utilitzi facilitarà el posterior processat, anàlisi i comprensió. Per facilitar les tasques relacionades amb l'anàlisi és important escollir noms d'acció que no siguin prefixes els uns dels altres i, d'aquesta manera, facilitar i simplificar l'ús de la comanda *grep* o d'eines similars a l'hora d'inspeccionar i processar els fitxers de log generats. Això pot no ser així en el cas de voler analitzar tot un bloc de marques a la vegada, per exemple les corresponents a la bústia, on un mateix prefix pot ser usat per tot un grup de marques.

Per altra banda, cal ser curós a l'hora d'introduir aquest tipus de marques en el lloc web donat que podrien interferir amb el correcte funcionament de les aplicacions que reben les crides. Per exemple, si un *servlet* o un *CGI* té un mecanisme de seguretat que controla el nombre de paràmetres que rep, aquest mecanisme s'hauria de modificar per tal que pugui incorporar els nous paràmetres. Això pot suposar un inconvenient en alguns llocs web, depenent de la seva estructura d'enllaços i crides, i en entorns web complexos on pot haver-hi centenars de serveis disponibles mitjançant crides a *servlets*



o *CGI* que ja estan en funcionament. Com a efecte col·lateral, implementar el sistema de marques pot servir per detectar crides i *scripts* mal dissenyats, com ha estat el cas.

En el campus virtual de la UOC s'han definit un conjunt de marques orientades a l'acció. Aquestes marques han de recollir l'activitat dels usuaris en les dues botoneres o menús del campus, així com alguns enllaços directes de la plana d'inici. Les marques definides es descriuen en detall a l'annex D i són les que es presenten a continuació a les figures 3.4 i 3.5 i a la Taula 3.2.



Figura 3.4: Marques del menú lateral.

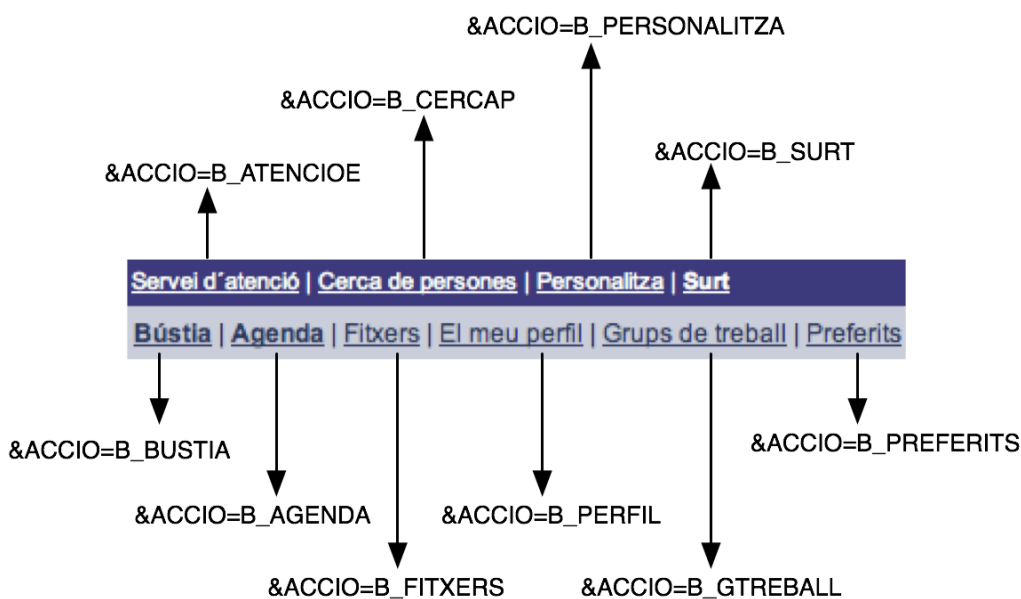


Figura 3.5: Marques del menú superior.

Botó "Bustia" de la plana d'inici	&ACCIO=B_BUSTIA_PI
Link "Missatges nous" de la plana d'inici	&ACCIO=L_MNOUS
Bandereta	&ACCIO=BANDERA

Taula 3.2: Marques de la plana d'inici.

Aquest conjunt de marques genera informació suficient per a analitzar el comportament dels usuaris en la plana d'inici i dels serveis més habituals del campus virtual, però es pot complementar.

### 3.5.3 Estratègia basada en el contingut

La segona estratègia que es proposa es basa en situar recursos invisibles en les planes del campus virtual de les quals es vol fer seguiment, de manera que quan un usuari accedeix i carrega una plana marcada es generi una entrada en els fitxers de log. D'aquesta manera el que s'aconsegueix és un registre de les planes i els continguts que ha visitat cada usuari. En realitat, el que s'està fent és que quan el navegador carrega un element invisible, com ara una petita imatge transparent, s'executi un codi que força una nova entrada en els fitxers de log. Això s'aconsegueix amb una sentència HTML com aquesta:

```

```

Aquesta sentència es col·loca en aquelles planes que es consideren rellevants pel procés d'aprenentatge i així poder capturar si els estudiants visiten aquestes planes. Al mateix temps, aquest mètode també resulta de gran utilitat per generar una entrada específica al fitxer de log quan un enllaç és seleccionat.

El codi que genera una entrada específica en els fitxers de log quan un enllaç és seleccionat és el següent:

```

<script>
Function tag2log(action) {
uocTrack_1_0.src=uocTrack_1.0.src + "?ACCIO=" + action
}
</script>

```

I, per exemple, es pot cridar així:

```

<a href="http://www.uoc.edu/wiki"
onClick="tag2log('WikiAccess');">Wiki</a>

```

Aquesta estratègia i el tipus de marques que requereix consumeix més recursos computacionals que l'anterior, però presenta alguns avantatges: no depèn d'una crida en particular, no interfereix amb els servlets o CGI existents i, el que potser és més important, és un mètode que pot ser automatitzat i així incorporar-lo a totes les planes que continguin un recurs o servei en particular que s'hagi seleccionat per fer-ne seguiment.

Les marques que s'han definit pel campus virtual de la UOC seguint aquesta estratègia marquen la càrrega de la plana d'inici i de la plana d'aules dels estudiants. També s'han utilitzat per marcar el quadre d'accessos directes a les aules que apareix en aquestes dues planes. El llistat d'ítems marcats es presenta a continuació:

- Plana d'inici
- Plana d'inici d'aules

Per cada assignatura que apareix a la plana d'inici:

- Enllaç amb el nom de l'assignatura
- Enllaç al Tauler
- Enllaç al Fòrum
- Enllaç al Lliurament d'activitats

### 3.5.4 Consideracions

Les dues estratègies de marcatge es poden utilitzar al mateix temps, capturant diferents aspectes de l'activitat i comportament de l'usuari. En sistemes web rics en mida i complexitat, on el mateix objectiu es pot assolir seguint diferents camins, és necessari no només conèixer les accions dels usuaris, sinó també tenir constància de les àrees que visiten i així poder construir un model d'usuari més acurat. A continuació es presenta un fragment de les línies de log que mostren les marques introduïdes i així exemplificar el resultat de cada estratègia de marcatge. Amb l'objectiu de facilitar la seva lectura, en aquest document es presenta un format simplificat del clàssic *Apache CLF*, substituint l'adreça IP de l'usuari pel seu identificador.

Estratègia orientada a l'acció:

```
166877 20060920183505 GET
/UOC/a/extcgi_0.html?s=166877&ACCIO=B_INICI&img=hola&cgi=hola
...
166877 20060920183518 GET
/UOC/a/cgi-bin/ma_mainMailFS?s=166877&ACCIO=L_MNOUS...

166877 20060920183555 GET
/webapps/classroom/032_common/plan.jsp?s=166877&ACCIO=plan.jsp&...
...
166877 20060920183625 GET
/webapps/classroom/032_common/docum.jsp?s=166877&ACCIO=docum.jsp&...
```

Estratègia orientada al contingut:

```
166877 20060920180653 GET /trackingTag.gif?PAGE=/UOC/a/cgi-bin/hola

166877 20060920180702 GET
/webapps/classroom/032_common/plan.jsp?s=166877&ACCIO=plan.jsp&...
166877 20060920180703 GET
/trackingTag.gif?PAGE=/webapps/classroom/032_common/plan.jsp

166877 20060920180713 GET
```

```

/webapps/classroom/032_common/comms.jsp?s=166877&ACCIO=comms.jsp&...
166877 20060920180713 GET
/trackingTag.gif?PAGE=/webapps/classroom/032_common/comms.jsp

```

En ambdós exemples es poden veure les entrades generades per l'usuari amb identificador 166877, la marca de càrrega de la plana d'inici (`trackingTag.gif?PAGE=/UOC/a/cgi-bin/hola`), com accedeix a la bústia mitjançant l'enllaç de missatges nous (`&ACCIO=L_MNOUS`), com accedeix a l'àrea de planificació de l'aula (`&ACCIO=plan.jsp`) i la marca de càrrega de l'àrea de planificació de l'aula (`trackingTag.gif?PAGE=/webapps/classroom/032_common/plan.jsp`).

El mapa de marques definit, el conjunt de marques introduïdes i el format de les entrades que es genera en els fitxers de log es troba descrit en profunditat a l'annex D. Les marques integrades en el campus virtual permeten, a partir d'un adequat processament dels fitxers de log, obtenir els camins de navegació dels usuaris. Analitzant aquest camins es pot obtenir informació rellevant dels estudiants i de l'ús que fan de l'entorn virtual d'aprenentatge. El procediment de processat de fitxers de log i obtenció de camins, així com les anàlisis realitzats es descriuen al capítol 4.

### 3.6 Aspectes de privacitat i seguretat

Hi ha dos qüestions importants que s'han de tenir en compte al dissenyar qualsevol proposta de marques integrades, com són la privacitat i la seguretat. A més, es tracta de dues qüestions que cada dia preocupen més als usuaris [92] i que cal gestionar adequadament.

En relació a la privacitat, és important posar de manifest que l'ús d'un sistema de monitorització continu de l'activitat dels usuaris no ha de ser vista com una amenaça a la seva privacitat [1] (per part dels usuaris del sistema), sinó més aviat com una part important d'un entorn virtual d'e-learning, que permet l'avaluació i millora de la usabilitat i la personalització. El primer fet és que els usuaris han de conèixer l'existència del mecanisme de monitorització [92]. Estudis com el que es planteja en aquest treball poden posar sobre la taula diferents qüestions sobre els aspectes de

privacitat. Així, és important remarcar que tota la informació que es recull a l'entorn dels estudiants, s'utilitza únicament amb finalitats acadèmiques i de millora de l'entorn virtual d'aprenentatge, sempre en el context de la Universitat Oberta de Catalunya.

Tota la informació sensible de la navegació dels usuaris, quines parts de l'entorn es visiten i quines no, quant temps dediquen a cada activitat, els resultats obtinguts per cada activitat, etc; només s'utilitza en el procés de redisseny i millora del sistema i per prendre decisions acadèmiques relatives als itineraris disponibles. A més, tota la informació sensible dels usuaris (nom, adreça, telèfon, etc) no s'utilitza en cap cas pel sistema d'anàlisi i de mineria de dades, pel que el processat de les dades es realitza sempre de forma anònima. Per altra banda, si els estudiants coneixen l'existència de l'eina de monitorització, aquesta s'entén com a part habitual de la recollida de dades que qualsevol equip docent realitza sobre el conjunt d'estudiants que porta a terme una assignatura o un curs. De la mateixa manera, la personalització no ha de veure's com quelcom intrusiu, sinó com una eina d'ajuda amb l'objectiu de millorar l'ús del sistema, en aquest cas, la plataforma d'e-learning, de forma que és l'usuari qui en tot moment té el control i pot, en qualsevol moment, deixar de banda les recomanacions obtingudes. Evidentment, el grau de llibertat de l'usuari dins d'una assignatura o curs queda prefixat amb anterioritat, tenint en compte els criteris que el procés de disseny instruccional determini per al curs o assignatura en qüestió. Igualment passa amb tots els aspectes relacionats amb la usabilitat i l'experiència d'ús del sistema de recomanacions, de manera que els estudiants puguin fer un ús satisfactori del mateix, i que aquest s'integri de forma completa amb les altres eines que la plataforma d'e-learning ofereix. La personalització no pretén vigilar a l'usuari, sinó ajudar-lo i oferir-li més possibilitats de les que pot ser conscient.

En relació a la seguretat, les qüestions relacionades poden aparèixer especialment depenent de la implementació del sistema de marques integrades. Un usuari avançat podria interceptar les crides utilitzant un proxy i llavors determinar la informació que s'està recollint del sistema, i intentar utilitzar aquesta informació amb el propòsit de portar a terme algun tipus d'acció. Malgrat que això pot ser un aspecte molt important en un sistema de comerç electrònic, no està clar que ho pugui ser en un entorn d'e-learning i l'ús fraudulent que un estudiant podria fer del coneixement obtingut. No

obstant, si la informació d'ús s'utilitza en temps real pel sistema d'e-learning, per exemple per personalitzar l'entorn i el procés d'aprenentatge, aquesta qüestió s'hauria de tractar com a crítica. En qualsevol cas, cal dir que el sistema de marques que es proposa en aquest treball i que ha estat implementat en el campus virtual de la UOC, ha estat implementat pel propi equip desenvolupador del sistema, passant les diferents proves de seguretat previstes per la institució.

Una opció de treball que s'ha utilitzat en alguns dels experiments que es presenten al capítol 4 consisteix en reemplaçar per un nou identificador totes les referències als números de sessió que és l'element que identifica la navegació de cada usuari. Aquest nou identificador d'usuari es genera seqüencialment i fa impossible identificar cap usuari individual. D'aquesta manera, d'aquesta manera els fitxers de log preprocessats serien del tot inútils davant de possibles atacs.

### **3.7 Personalització**

Una de les possibilitats més interessants en qualsevol sistema interactiu i especialment en els sistemes d'e-learning és monitoritzar la navegació dels usuaris amb el propòsit de realitzar anàlisis, tal i com s'ha descrit al llarg d'aquest capítol. D'aquesta manera, es pot conèixer la navegació i comportament real dels usuaris i, a més, es poden descobrir fets interessants i aspectes insòlits sobre el propi sistema. Els resultats que s'obtenen constitueixen una font d'informació de gran utilitat per a introduir elements d'adaptació i personalització en el sistema d'e-learning, així com per validar el seu disseny.

A partir dels fitxers de log es pot construir un model de navegació de l'usuari [93], que es pot enriquir amb altres fonts d'informació sobre els usuaris. Aquesta informació es pot utilitzar per l'avaluació i millora de la usabilitat [80] detectant si la interfície web del sistema interactiu ha estat dissenyada correctament i si els usuaris troben obstacles i dificultats per portar a terme els seus objectius [111]. El model d'usuari també es pot utilitzar per a millorar l'experiència d'aprenentatge dels estudiants, adequant els itineraris formatius a les necessitats específiques dels estudiants i proporcionant mecanismes de suport a les tasques que han de realitzar [121].

En el cas particular d'un entorn virtual d'aprenentatge, l'anàlisi es pot portar a terme a tres nivells, tal i com s'ha descrit en aquest mateix capítol. A cada nivell es poden respondre diferents preguntes i es poden adreçar diferents aspectes de personalització. Els professors, els dissenyadors instruccionals i, fins i tot, els dissenyadors del sistema, necessiten eines potents per entendre tota la informació recollida a l'entorn virtual d'aprenentatge de cara a millorar el procés d'aprenentatge i, per tant, l'experiència i satisfacció, mitjançant un entorn adaptatiu [136], monitoritzant i interpretant l'activitat dels seus usuaris [168], inferint requeriments i preferències d'usuaris, i prenent decisions a partir del coneixement disponible dels mateixos.

Per tant, l'extracció dels patrons reals de navegació i comportament pot ser de gran utilitat, entre altres coses, per determinar la qualitat de l'entorn virtual d'aprenentatge i avaluar el grau de concordança entre la navegació real dels usuaris, que reflecteix el seu model mental i el model conceptual del propi sistema d'e-learning. A més, resulta especialment interessant per introduir elements de personalització, adequant el procés d'aprenentatge als coneixements previs i necessitats dels estudiants [116, 121]. D'aquesta manera, es pot establir un cicle que es retroalimenta i que constitueix una millora continua, en la línia del que promou el *design research*.

Un dels objectius d'aquest treball de tesi doctoral és identificar com analitzar un entorn complex per millorar-lo i personalitzar-lo. Així, en un entorn d'e-learning complex com el de la UOC es pot realitzar una proposta de personalització que relaciona els tres nivells d'anàlisi presentats amb accions de recolzament que es poden portar a terme a cada nivell. El sistema, mitjançant el model d'usuari, coneix a l'estudiant i li pot proporcionar ajuda dins d'un marc establert. Aquest marc el constitueix el que es pot anomenar cicle de vida de l'estudiant i les tasques que ha de realitzar en cada moment.

El cicle de vida de l'estudiant inclou totes les etapes per les quals passa un estudiant, des de que sol·licita informació i demana l'accés a la universitat, fins que finalitza els seus estudis o els abandona. És interessant remarcar que en un escenari d'aprenentatge al llarg de la vida o de *lifelong learning*, aquest cicle de vida no acaba mai, ja que



l'estudiant es forma al llarg del temps, durant tota la vida, encadenant diferents propostes formatives. Aquest cicle de vida es presenta gràficament a Figura 3.6.

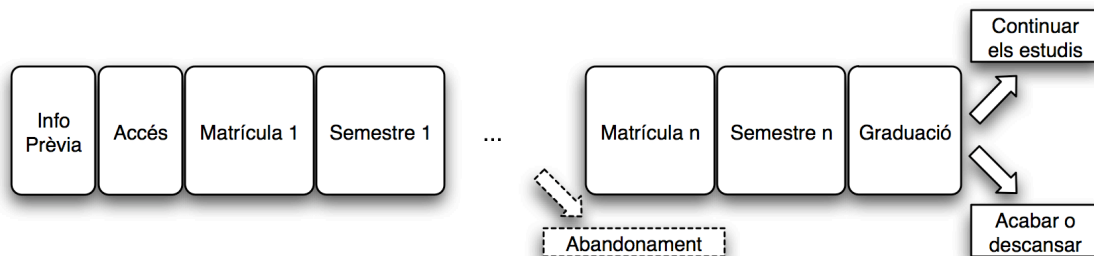


Figura 3.6: Cicle de vida de l'estudiant de la UOC.

El primer que fa un estudiant que vol estudiar a la UOC és sol·licitar informació sobre els estudis i demanar l'accés a la universitat. Cada estudiant arriba a la universitat amb una “motxilla”, és a dir, amb uns estudis previs amb els que, si es tracta d'estudis universitaris, pot demanar adaptació als nous estudis que vol realitzar. Evidentment, en aquesta motxilla hi ha tota la informació socio-demogràfica de l'estudiant, així com les seves preferències i objectius. Un cop ha accedit a la universitat, l'estudiant formalitza la primera matrícula que, per a molts d'ells, es tracta del primer contacte amb el pla d'estudis i la seva estructura d'assignatures. Un cop formalitzada la matrícula, la següent etapa és la del seguiment del semestre. A partir d'aquest punt, el cicle de vida és, com el propi nom implica, cíclic: l'estudiant va matriculant i seguint els diferents semestres, fins que arriba a dos possibles punts clau. Per una banda, l'estudiant pot arribar al punt de la seva graduació o, malauradament, abandonar els seus estudis. Un cop s'ha graduat pot decidir continuar estudiant o bé donar per finalitzada, ja sigui definitivament o temporalment, la seva formació. És interessant mencionar que, amb certa freqüència, els estudiants es prenen un semestre de descans, és a dir, un semestre on no es matriculen de cap assignatura i que, per tant, no estudien. Això és important per definir correctament el què s'entén per abandonament, per exemple.

L'abandonament dels estudis és un fet que pot passar en qualsevol moment i per això a la figura s'indica amb una línia discontinua. L'abandonament és una qüestió de gran

importància en els ensenyaments no presencials, donat que és més propens a produir-se si es compara amb els ensenyaments presencials tradicionals [106]. La proposta que es realitza en aquest treball no inclou l'estudi de l'abandonament, que mereix un estudi en profunditat per ell mateix, però sí adreça aquells aspectes que poden ajudar i facilitar les tasques als estudiants, eliminant obstacles, reduint frustracions i millorant la seva experiència d'aprenentatge i, d'aquesta manera, reduir la taxa d'abandonament. També s'aborden indicadors que poden ser usats per analitzar si la proposta formativa és l'adequada.

### **3.7.1 Elements de personalització del campus virtual de la UOC**

La proposta de personalització que es realitza es deriva de la relació entre tres elements principals: el cicle de vida de l'estudiant, el conjunt de tasques que realitzen o poden realitzar en cadascuna de les etapes, i el coneixement sobre els estudiants i les accions que porten a terme, que s'obté com a resultat de l'anàlisi a tres nivells. Les tasques que realitzen o que poden realitzar els estudiants al llarg del semestre es presenten a l'annex B. Els resultats de l'anàlisi a tres nivells i la informació rellevant pel procés d'aprenentatge es descriuen al llarg de tot el capítol 4. A continuació, a la Taula 3.3 es presenta esquemàticament la interrelació d'aquests tres elements, diferenciant per cada semestre acadèmic els períodes d'inici, desenvolupament i fi de curs. Aquesta taula resumeix les aportacions experimentals realitzades com a mostra del potencial del sistema de personalització.

Amb aquesta interrelació el que es vol obtenir és la informació necessària per donar opcions de personalització, és a dir, oferir recomanacions a l'estudiant en cada moment del seu cicle de vida, tenint en compte les etapes per les que ha passat. El propòsit de la taula és identificar els aspectes en els que es pot oferir suport als estudiants i l'objectiu d'aquesta secció es identificar i presentar els escenaris de personalització en un entorn com el de la UOC.

Nivell 1	(1)	(1b)	(1b)	(4)	(4)	(4)	...
Nivell 2					(3)		...
Nivell 3			(2)	(5)	(5)	(5)	...
	Info Prèvia	Accés	Matrícula	Semestre <sub>1</sub> Inici	Semestre <sub>1</sub> Durant	Semestre <sub>1</sub> Fi	...

Taula 3.3: Relació dels nivells d'anàlisi amb les etapes del cicle de vida de l'estudiant.

(1) En aquest punt es considera la navegació dels usuaris per l'espai d'informació dels estudis, disponible al portal web de la UOC. L'objectiu que es persegueix consisteix en proporcionar informació adequada als interessos dels futurs estudiants. Amb un estudi com el que es presenta a la secció 4.3, es pot saber quins estudiants sol·licitaran informació per l'accés i, per tant, es poden portar a terme accions de suport orientades a aquells estudiants que es preveu que no sol·licitaran més informació.

(1b) Experiments preliminars mostren que les persones que en (1) només sol·liciten informació d'una titulació demanen accés i es matriculen amb més probabilitat que aquells que sol·liciten informació de 2 o 3 estudis diferents. Això pot usar-se per a orientar millor als futurs estudiants durant el procés de captació.

(2) A l'hora de matricular-se, els estudiants accedeixen a les recomanacions de matrícula disponibles a la secretaria del campus virtual i també demanen consell als seus tutors. En el nivell 3, el nivell d'aprenentatge al llarg de la vida, s'estudien, entre d'altres elements, les assignatures matriculades pels estudiants. Els experiments realitzats i que es presenten a la secció 4.8 mostren que els estudiants realitzen la matrícula sense tenir en compte les recomanacions proporcionades per la universitat. Això molts cops repercuteix negativament en el seu rendiment acadèmic, per lo que a l'hora de matricular-se el sistema de personalització podria aconsellar novament als estudiants sobre la matrícula d'assignatures que, segons el seu perfil, han de realitzar, o com a mínim, avisar als estudiants i els seus tutors de potencials perills.

(3) En el nivell de curs acadèmic s'estudia el comportament dels estudiants en relació al pla docent i les activitats previstes. Els resultats obtinguts a la secció 4.7 mostren que aquells estudiants que no descarreguen l'enunciat de l'activitat proposada els primers dies després d'haver-se publicat l'enunciat, no la lliuren o no obtenen una bona qualificació. El sistema podria proporcionar recordatoris a aquells estudiants que no descarreguen l'activitat, o bé podria advertir al consultor sobre els estudiants que és probable que no la lliurin, i així poder prendre les mesures adequades, readaptant el seu calendari en funció d'aquest fet.

(4) En el nivell de sessió, independentment del moment del semestre acadèmic, s'estudia la navegació dels estudiants. Coneixent aquesta navegació es poden descobrir fets no coneguts prèviament en relació a la seva navegació i comportament, tal i com es presenta a la secció 4.6. Aquesta informació pot servir per millorar aspectes d'usabilitat de l'entorn virtual d'aprenentatge i, si es relacionen amb les dades dels altres dos nivells, permet construir perfils d'usuaris relacionant la seva navegació amb el seu resultat acadèmic.

(5) Utilitzant dades agregades de (3) i (4) combinades amb l'historial de matrícula de l'estudiant és possible detectar situacions de risc que indiquin un possible abandonament dels estudis.

És evident que la relació entre els tres elements expressada a la Taula 3.3 pot arribar a generar connexions virtualment infinites i no és l'objectiu d'aquest treball recollir-les totes, sinó mostrar aquelles més representatives. L'objectiu que es persegueix no és individualitzar el procés d'aprenentatge ni tampoc construir un sistema d'autoaprenentatge sinó recomanar la millor opció pels estudiants en cada moment del procés d'aprenentatge, tal i com es descriu a [116], tenint en compte la seva navegació i actuació prèvia.

### 3.8 Resum

En aquest capítol s'ha presentat la proposta d'anàlisi a tres nivells per a sistemes d'e-learning complexos i canviants. Aquesta proposta es desenvolupa a partir de l'anàlisi del context i dels estudiants del sistema, fent ús de la tècnica dels personatges i escenaris. La proposta d'anàlisi permet recollir i analitzar dades sobre els estudiants i l'ús de l'entorn virtual d'aprenentatge gràcies a un sistema de marques integrades en el sistema. Els resultats que s'obtenen de l'anàlisi a tres nivells són rellevants per així poder eliminar incoherències entre el disseny del sistema i l'ús que se'n fa, així com per proposar aspectes de personalització de l'entorn virtual d'aprenentatge que millorin l'experiència educativa. En el següent capítol es presenta la implementació del sistema de marques que permet capturar i analitzar la navegació dels usuaris i es descriuen els experiments i anàlisis portats a terme a cada nivell, així com els resultats obtinguts que donen suport a aquesta proposta.



## Capítol 4

### Resultats experimentals

Un cop definida l'aproximació a l'anàlisi de l'ús i dels usuaris dels sistemes d'e-learning mitjançant tres nivells i una proposta de marques, i presentats els principals reptes que això implica, s'aborda ara la validació experimental del model presentat. Per tant, l'objectiu principal d'aquest capítol és presentar la implementació de la proposta de marques de seguiment i del sistema de recollida i gestió de fitxers de log, així com les dades de navegació que s'obtenen, i la experimentació portada a terme a cadascun dels nivells i com el resultat dels experiments ha influït en el treball de tesi doctoral que es presenta.

#### 4.1 Introducció

En el capítol anterior, a la secció 3.7 s'han descrit les diferents possibilitats d'actuació en el campus virtual, a partir de la informació obtinguda a cada nivell. A cadascun dels nivells descrits al capítol 3 d'aquest treball, s'han portat a terme anàlisis sobre les dades disponibles. Una part important del treball realitzat ha consistit en la implementació i posada en marxa del sistema de marques de seguiment, així com la recollida, gestió i preprocessat de les dades dels fitxers de log. A cada nivell s'han realitzat diferents processats de les dades, d'acord al format i semàntica d'aquestes. Els resultats obtinguts, la major part publicats i d'altres encara inèdits, es presenten en aquest capítol.

En primer lloc es presenta i es descriu el camp d'experimentació utilitzat, la UOC i el seu campus virtual, que constitueixen un escenari únic i de gran interès donat les seves característiques i el volum d'usuaris als quals dona servei. A continuació es presenten, de manera cronològica en el temps, els treballs inicials realitzats amb les dades de la UOC. Després es mostren els experiments que recolzen la proposta de nivells i les possibilitats de personalització.

No entra en l'abast d'aquesta tesi doctoral fer un recorregut exhaustiu de tots els estudis i experimentacions que es poden realitzar en cada nivell, sinó mostrar que amb l'anàlisi realitzat a cada nivell s'obté informació rellevant i es té més coneixement dels usuaris i de l'ús que fan del sistema i, per tant, es disposa de la informació adequada per millorar el disseny del sistema d'e-learning i proposar aspectes de personalització a cada nivell. Així, l'objectiu és fer una fotografia de la UOC i dels seus estudiants, mitjançant una metodologia concreta, i que aquesta fotografia proporcioni informació rellevant i interessant per millorar la pròpia universitat i la experiència d'aprenentatge dels seus estudiants. L'escenari a estudiar és enorme i canviant, el campus virtual és molt complex i l'estudi d'usuaris també presenta complexitat i, per tant, el mètode que s'utilitza és mou a diferents nivells, utilitzant la informació disponible en cada moment.

## **4.2 La UOC com a escenari únic**

La Universitat Oberta de Catalunya (UOC) [22] és una universitat no presencial, completament virtualitzada que ofereix 20 titulacions oficials i diversos postgraus i estudis de màster, a més d'un programa de doctorat, amb gairebé 40.000 estudiants i uns 2.500 treballadors, incloent professorat, personal acadèmic i de gestió, i col·laboradors externs.

El portal o lloc web de la universitat ofereix diferents serveis tant per a visitants com per a estudiants o usuaris registrats. Principalment, el web ofereix informació sobre la universitat i la seva oferta formativa, els projectes que s'hi desenvolupen i la recerca que es realitza. A més, ofereix informació cultural, social, artística, tecnològica i sobre altres activitats relacionades. En una setmana del semestre de tardor del curs 2007-2008



el portal ha servit gairebé 4 milions de planes, unes 550.000 al dia. El nombre de visitants de la mateixa setmana supera els 350.000, dels quals els visitants únics són gairebé 180.000, i la meitat d'aquests són visitants nous, i tots aquests visitants realitzen més de 800.000 sessions de navegació.

El campus virtual és l'entorn virtual d'aprenentatge propi de la UOC. Els usuaris registrats hi poden accedir a través del portal. Amb el seu nom d'usuari i paraula clau són autenticats i se'ls proporciona accés a distintes àrees, tot depenent del seu perfil i característiques. Hi ha diferents perfils d'usuari, principalment estudiants, professors, consultors, tutors, personal de gestió i administradors del sistema. Tots els perfils comparteixen el mateix entorn web amb un disseny i arquitectura similars, tot i que el seu contingut, navegació i serveis varien en funció del seu perfil i personalització.

#### **4.2.1 Disseny del sistema i arquitectura**

El campus virtual de la UOC ha estat dissenyat tenint en compte alts requeriments instruccionals i d'usabilitat, perseguint el següent objectiu: crear un entorn simple i intuïtiu que permeti a un gran nombre d'usuaris connectar-se simultàniament, tot mantenint una concurrència i rendiment raonable. En el disseny i avaluació de la interfície del campus virtual es van portar a terme un gran nombre de mètodes de disseny centrat en l'usuari com són: entrevistes, *focus groups*, *card sorting*, avaluació heurística i test d'usuaris, descrits a la secció 2.1.2.

Des del punt de vista tecnològic, el campus virtual està construït sobre un complex sistema servidor de bases de dades i utilitza una estructura jeràrquica de servidors que fan front als diferents tipus de peticions dels usuaris. El campus virtual treballa a partir de 43 servidors que actuen com a *front-ends* i que processen les peticions dels usuaris que volen accedir a l'entorn, constituït per 39 servidors que donen servei al campus virtual. Aquests servidors no inclouen aquells usats en l'entorn de reproducció, desenvolupament i proves. El nombre de *front-ends* actius depèn de la càrrega del sistema en cada moment. Un sistema balancejador de càrrega distribueix als usuaris de manera automàtica entre tots els *front-ends* disponibles. En les hores punta, com per

exemple les 18 hores, es poden trobar fins a 3.000 usuaris connectats concurrentment i navegant simultàniament pel campus virtual, i tots els *front-ends* estan actius. Quan el nombre d'usuaris decreix en les hores de menor ús, com per exemple a les 4 hores de la matinada on només hi ha al voltant de 70 usuaris connectats, alguns dels *front-ends* es desactiven de manera automàtica, reduint d'aquesta manera la càrrega de tot el sistema. Considerant un dia del semestre acadèmic agafat a l'atzar, el nombre mig d'usuaris connectats al campus virtual és de 1.700, i el nombre màxim al que s'ha arribat en un dia concret, el dia d'inici de semestre, supera els 5.000 usuaris concurrents. També hi ha altres sistemes i aplicacions com servidors de bases de dades que s'utilitzen per processar serveis requerits pels usuaris per a accedir a àrees específiques del campus virtual, com per exemple la biblioteca digital o la consulta de l'expedient acadèmic.

El campus virtual de la UOC utilitza tecnologia web basada en el model client-servidor amb una interfície comú per així integrar els diferents serveis i funcionalitats. Aquestes funcionalitats són, principalment, l'accés als materials didàctics i altres recursos educatius de la biblioteca digital, informació i tutorització acadèmica, i la interacció amb els docents i altres estudiants mitjançant les eines de comunicació asíncrona basades en el correu electrònic. Els serveis oferts es descriuen a la secció 4.2.2 i són, principalment, correu electrònic, aules virtuals, agenda, secretaria, biblioteca i novetats i notícies. Quan els usuaris utilitzen aquests serveis i funcionalitats, deixen una traça que, posteriorment, pot ser estudiada amb l'objectiu de modelar la seva activitat. Principalment, la informació es recollida en els fitxers de log dels servidors web del campus, seguint el format Apache CLF (*Common Log Format*) [4].

El campus virtual de la UOC segueix l'arquitectura clàssica de molts sistemes d'informació implementats en web i s'estructura en tres capes que donen resposta a les peticions dels usuaris. La primera capa és la de presentació i se'n ocupen els servidors web, que són els que generen logs d'accés amb el registre de les interaccions dels usuaris. Moltes d'aquestes interaccions no només són navegacionals sinó que inclouen l'ús de serveis i components a càrrec de la segona capa, de lògica de negoci. Aquesta capa també genera els seus propis logs d'execució i el seu format depèn de cada programa o component específic. Per exemple, el motor de SCORM usat en alguns materials docents, enregistra totes les peticions d'accés a *assets* o recursos en una base

de dades pròpia. La tercera capa és la del model de dades i se'n fa càrrec el sistema de base de dades que actua com a *back-end* del sistema. Aquest sistema emmagatzema totes les dades generades i utilitzades per les aplicacions de la capa prèvia. Les dades emmagatzemades en aquesta capa també poden ser utilitzades per fer seguiment de l'activitat en el sistema i portar control, per exemple, del darrer canvi en un recurs específic. La Figura 4.1 presenta gràficament l'estructura i traspàs de dades en aquest model de tres capes.

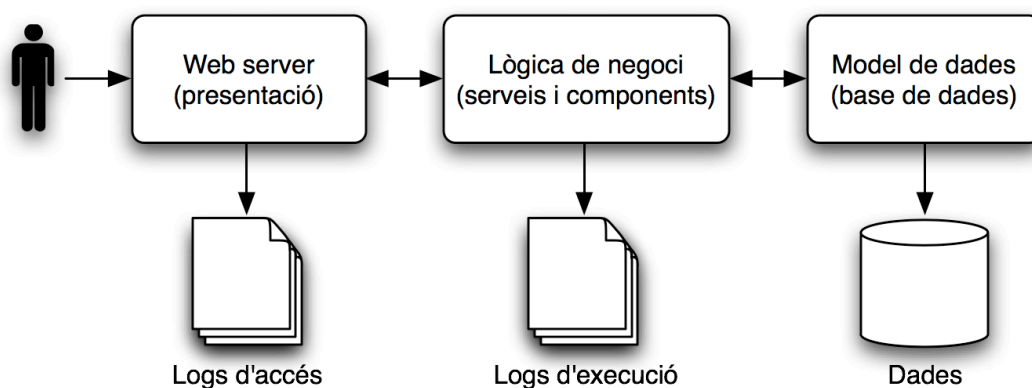


Figura 4.1: Estructura i traspàs de dades en un sistema de tres capes.

En aquest treball es tenen en compte principalment les dades generades per la capa de presentació, que són les que es generen en els servidors web del campus virtual i queden enregistrades en els fitxers de log d'accés. El fet d'utilitzar els fitxers de log afavoreix que la metodologia d'anàlisi es pugui reproduir en altres sistemes web que utilitzin una estructura similar. Tanmateix, només quan es disposa de les dades generades en la capa de lògica de negoci és quan es poden realitzar anàlisis en el segon i tercer nivell dels presentats en la secció 3.2, i l'anàlisi és més orientat i enfocat a una o unes activitats determinades [113]. En una situació ideal, les diferents fonts de dades es fusionarien per així construir un model d'usuari més acurat. Això passa per dissenyar una arquitectura de la informació que tingui en compte aquest objectiu d'anàlisi.

### 4.2.2 Serveis de campus

El campus virtual de la UOC és un entorn d'e-learning integrat que permet als seus usuaris comunicar-se entre ells mitjançant un sistema de correu electrònic; també integra una agenda, un sistema de notícies i novetats, aules virtuals, una biblioteca digital i d'altres eines relacionades amb el procés d'aprenentatge. Els elements principals del campus virtual es presenten a continuació.

- **Plana d'inici.** Quan els usuaris accedeixen al campus virtual, se'ls presenta una plana d'inici, des de la qual poder accedir a les diferents àrees i serveis de l'entorn virtual, d'entre els que es destaquen els següents: correu electrònic, aula virtual, agenda, novetats i notícies. A la plana d'inici se li presenta a l'usuari informació personalitzada i la que ha canviat des del seu darrer accés al campus virtual: missatges nous a la bústia, els missatges nous a les diferents aules, enllaços directes a les novetats i notícies més actuals, així com les cites de l'agenda per a aquell dia.
- **Correu electrònic.** Consisteix en un sistema personal de correu electrònic integrat en l'entorn virtual d'aprenentatge que permet als usuaris comunicar-se entre ells, així com llegir i escriure en les bústies compartides de les aules. El sistema de correu electrònic utilitza adreces estàndards d'Internet i, a més, garanteix la identitat de l'usuari, essent aquest darrer aspecte molt important en un entorn virtual d'aprenentatge. El sistema de correu també ofereix altres funcionalitats com la consulta de la història i estat dels missatges.
- **Aula virtual.** És el lloc on els estudiants tenen accés a tota la informació relacionada amb el procés d'aprenentatge. A l'aula hi troben les eines d'interacció i compartició de recursos amb el professor i els altres estudiants. El sistema informa a l'estudiant dels canvis que s'han produït des de la seva darrera visita. L'aula virtual s'estructura en quatre espais principals: planificació, comunicació, recursos i avaluació. L'espai de planificació conté la descripció de tots els elements de l'assignatura, els objectius d'aprenentatge i el calendari d'activitats que es seguirà durant el curs. L'espai de comunicació proporciona

les eines per a que els estudiants es comuniquin entre ells i amb el seu consultor. És aquí on es troben les bústies compartides que reben el nom de “Tauler” i “Fòrum”. En aquest espai també es troben la llista de companys de l’aula i la informació sobre si estan connectats al campus virtual en aquell moment. L’espai de recursos conté els enllaços a tots els recursos digitals necessaris, així com el llistat de recursos no digitals que també són un requisit de l’assignatura. Finalment, l’espai d’avaluació conté les bústies de lliurament d’activitats d’avaluació continuada, així com els accessos directes a les aplicacions relacionades amb el procés d’avaluació, com per exemple la consulta de qualificacions.

- Agenda. Cada usuari disposa d’una agenda on automàticament apareixen les cites relacionades amb les assignatures o cursos que ha matriculat i on pot afegir noves cites, així com fer anotacions personals.
- Novetats i notícies. A cada usuari se li presenten les novetats i notícies que són adequades i significatives segons el seu perfil.
- Biblioteca. Integra tots els recursos i continguts digitals en el campus virtual i permet cercar i sol·licitar en préstec els recursos en suports no digitals.
- Altres. El campus virtual ofereix diferents informacions i utilitza certs recursos per així a generar la sensació d’entorn proper i potenciar la percepció de comunitat virtual. Així, per exemple, el correu electrònic mostra sempre, juntament amb el nom, la fotografia del remitent dels missatges; a la plana d’inici es mostra el nombre d’usuaris connectats al campus virtual i també la “foto del dia” que està relacionada amb algun esdeveniment important de la universitat, etc.

Un cop els usuaris accedeixen al campus virtual, sempre troben una plana d’inici com la que es mostra a la Figura 4.2. Depenent del perfil d’usuari, aquesta plana inicial pot ser significativament diferent. La aproximació d’aquest treball es focalitza en els estudiants de titulacions homologades en català com a subjectes principals d’estudi. S’ha escollit

aquest grup d'usuaris perquè és el grup més nombrós que presenta característiques comuns, a banda de tenir un interès rellevant de *per se*. Aquestes característiques són el mateix entorn i configuració del campus virtual, amb la mateixa llengua i botons al menú lateral i superior, així com la informació i continguts que se'ls hi presenten. També comparteixen el mateix calendari acadèmic i, per tant, les dates clau durant el curs coincideixen, un aspecte que afavoreix escollir aquestes dates per realitzar estudis de comportament i navegació. El fet que un mateix i nombrós grup d'usuaris comparteixi un mateix entorn, en un mateix període de temps i amb uns objectius comuns que es concreten, principalment, en participar en el procés d'aprenentatge, afavoreix una recollida de dades homogènia i que es pugui realitzar un processat i anàlisi amb garanties d'obtenir resultats representatius.

The screenshot shows the UOC Campus Virtual homepage. At the top, there is a navigation bar with the UOC logo and links for 'Servei d'atenció', 'Cerca de persones', 'Personalitza', and 'Surt'. Below this is a secondary navigation bar with 'Bústia', 'Agenda', 'El meu perfil', 'Grups de treball', and 'Preferits', along with a user profile dropdown showing 'UOC - Estudiant - Humanitats'. The main content area starts with a 'Hola' greeting and 'Benvinguda al Campus Virtual.' It includes a 'Qui hi ha' section showing '1622 usuaris connectats' and a 'Nulla dies sine linea' section with a link to 'faca\_regne\_pettit [...]'. There are also buttons for 'BÚSTIA' (with 'Missatges nous: (23)') and 'AGENDA' (with 'Cites de la setmana: (3)'). The 'AULES' section is divided into three columns: 'Anglès I aula 18' (with links for Notice board, Forum, Exercises, Assessment, and News), 'Literatura i civilització gregues aula 3' (with links for Tauler, Fòrum, and Debat), and 'Tutoria inici maig06' (with links for Tauler and Fòrum). Below this is a 'NOVETATS I NOTÍCIES' section with two main news items: 'Modificació full personal d'exàmens: termini obert' and 'Nou apartat de centres de suport'. The first news item includes a small image and text about exam schedules. The second news item includes a small image and text about support centers. At the bottom right, there is a '+ Notícies' button.

Figura 4.2: Plana d'Inici del Campus Virtual de la UOC.

La plana d'inici del campus virtual es distribueix en un sistema de *frames* o marcs amb una disposició clàssica, que ha estat dissenyada tenint en compte estudis d'usuari i els estàndards establerts. La banda esquerra presenta un menú dinàmic que proporciona accés a totes les opcions disponibles i als serveis, personalitzats per a cada perfil d'usuari. La banda superior presenta els botons que donen accés ràpid i directe als

serveis i funcionalitats principals del campus virtual. La part central de la plana d'inici és l'àrea principal on es presenta la informació i on l'usuari realitza la major part de tasques. Aquesta àrea va canviant a mesura que l'usuari navega amb el campus virtual, utilitzant els serveis oferts i accedint als continguts disponibles. La plana d'inici és el centre d'operacions i seguiment que té l'estudiant respecte al seu procés d'aprenentatge.

Un cop presentada la UOC i el seu campus virtual amb els seus diferents aspectes i elements principals, a continuació es descriuen els experiments i anàlisis que s'han portat a terme amb les dades recollides a la universitat i a la seva plataforma educativa. En primer lloc es presenten els estudis i experiments inicials que van ser la llavor de la idea presentada en aquesta tesi, és a dir, la informació implícita del comportament dels usuaris pot ser usada per millorar el sistema. A continuació es descriu el sistema de gestió i preprocessat dels fitxers de log, necessari per a la captura de la navegació dels usuaris i per portar a terme els experiments. Per finalitzar es descriuen, per cadascun dels tres nivells, els experiments i anàlisis realitzats.

### **4.3 Experiments inicials amb les dades del portal de la UOC**

En aquesta secció es presenta el primer treball d'anàlisi de dades d'ús per modelar l'activitat dels usuaris que es va portar a terme en el context d'aquest treball de tesi doctoral. Al tractar-se d'un experiment preliminar, es va decidir estudiar un lloc web d'estructura jeràrquica i de caire més estàtic que el campus virtual de la UOC. Aquesta decisió va tenir dues implicacions principals: es facilita la comprensió del lloc web i el seguiment de la navegació dels seus usuaris, i va afavorir el treballar amb un format de logs més senzill que el que proporciona el campus virtual. El lloc web que es va escollir és el propi portal de la UOC, seleccionant específicament l'àrea dedicada a l'oferta formativa, que presenta la informació sobre els estudis que es poden cursar a la universitat. Els resultats d'aquest treball es van publicar a [118]. Aquest estudi es va realitzar amb dades d'accés i ús del lloc web de la UOC de l'any 2002.

La motivació principal del treball va ser l'estudi dels diferents perfils de navegació dels usuaris del lloc web i intentar obtenir dos tipus d'informació a l'entorn dels usuaris:

quins espais del lloc web visiten i de quina manera ho fan. Aquesta informació es pot utilitzar en dues línies, primera per millorar la usabilitat del lloc web, millorant l'arquitectura de la informació del lloc web [80, 163] i segona, per fer una proposta de navegació adaptativa [53, 138].

### 4.3.1 Navegació dels usuaris i arbres de decisió

Els usuaris accedeixen al lloc web de la UOC seguint, entre d'altres objectius, buscar informació sobre la seva oferta formativa, especialment durant el període d'accés. Molts usuaris que naveguen pel lloc web visiten un petit nombre de pàgines, entre sis i vuit, i aquest fet proporciona un interès en poder fer prediccions de manera ràpida i raonablement aproximades utilitzant classificadors simples, amb l'objectiu d'adaptar l'estructura de la plana i reforçar, per exemple visualment, els elements en els que els usuaris poden estar més interessats. Els arbres de decisió [38] permeten construir sistemes de classificació que parcialment aconsegueixen l'objectiu de la simplicitat i l'exactitud. Per a la construcció dels arbres de decisió s'utilitzen únicament hiperplans ortogonals que permeten obtenir una interpretació molt valuosa dels classificadors construïts, per a un anàlisi posterior, donat que només una variable és usada en cada decisió.

### 4.3.2 Conjunt de dades

Com s'ha descrit anteriorment, el conjunt de dades utilitzat es va obtenir d'una part del lloc web de la UOC<sup>3</sup> que ofereix informació sobre els estudis que es poden cursar a la universitat. Les persones interessades en estudiar tenen a la seva disposició, en aquesta àrea del lloc web, el llistat de titulacions organitzat per àrees de coneixement. El lloc web segueix una estructura clàssica amb un disseny basat en *frames* o marcs, amb un gran *frame* a la dreta i centre de la pantalla on es presenta la informació i un *frame* més petit a l'esquerra que proporciona accés a informació addicional relacionada amb el model educatiu i pedagògic de la universitat. Aquesta disposició de la informació es mostra a la Figura 4.3. El marc de l'esquerra és l'àrea considerada objectiu en aquest

---

<sup>3</sup> <http://www.uoc.es/web/cat/launiversitat/estudis>



estudi, donat que conté els enllaços d'informació addicional i de sol·licitud d'accés als estudis.



Figura 4.3: Estructura de la plana principal del lloc web.

Durant sis mesos s'han recollit les dades d'accés a les planes del lloc web i que corresponen a la navegació dels usuaris. S'ha creat una gran base de dades de més de 167.000 entrades corresponents a totes les visites al web. Aquest conjunt de dades no és un fitxer de log clàssic, sinó un seguit de entrades generades per codi introduït a les planes web que es volen monitoritzar. El conjunt d'entrades que generen aquests codis amagats s'utilitza a mode d'un fitxer de log reduït. Aquesta aproximació també es utilitzada per [36] i és la que a l'apartat 2.3.4 s'ha presentat com la recollida de dades de les mètriques remotes. Utilitzar les entrades generades per les mètriques remotes en lloc dels fitxers de log clàssics ofereix diferents avantatges per l'estudi que es realitza. En primer lloc, es simplifica notablement el preprocessat de les dades donat que només s'emmagatzema la informació rellevant i, d'aquesta manera, es redueix el temps de processador i l'espai de disc necessaris. El segon avantatge que ofereix aquesta aproximació és que facilita la traça de la navegació dels usuaris i, per tant, l'estudi dels comportaments de navegació. Per altra banda, depenent del nombre d'usuaris que visitin el lloc web, aquesta aproximació presentaria l'inconvenient que necessitaria gestionar un gran volum d'entrades generades pels codis insertats a les planes web.

Cada entrada generada enregistra l'identificador d'usuari, la data i l'hora de la visita. Aquest identificador és únic per cada usuari i es crea a partir de l'adreça IP de la seva primera visita i un "numero màgic" que s'assigna aleatòriament a cada usuari el primer cop que visita el lloc web. S'utilitza un mecanisme basat en *cookies* per monitoritzar les futures visites de l'usuari al lloc web. Com s'ha vist a la secció 2.3, aquest sistema presenta l'inconvenient que és impossible monitoritzar les futures visites d'un usuari si es connecta des de diferents ordinadors o si esborra les galetes del seu navegador habitual.

És necessari preprocessar les dades per així eliminar aquelles entrades que no s'han registrat correctament, ja sigui perquè ha fallat l'enregistrament de l'entrada o bé perquè l'usuari esborra o no accepta *cookies* i, per tant, no ha funcionat el mecanisme d'identificació i monitorització dels visitants. En aquest preprocessat també s'esborren les entrades duplicades, si en hi ha. Un total de 17 enllaços s'utilitzen com a informació bàsica extreta de cada usuari: els enllaços que ha visitat i el nombre de vegades que ho ha fet. Dels 17 enllaços, 13 fan referència a les opcions disponibles i 4 fan referència a la informació addicional. El total d'entrades després d'aquesta fase és de 160.000, mostrant que més del 96% del conjunt de dades original s'està utilitzant i, per tant, només un petit percentatge d'usuaris no són seleccionats per l'estudi.

Per cada usuari diferent en el conjunt de dades, més de 61.000, es construeix un vector amb 13 variables, una per cada enllaç del lloc web. Aquestes variables indiquen si l'usuari ha visitat o no un enllaç, i en quin ordre. L'etiqueta assignada a cada usuari és també un valor binari que mostra si aquell usuari ha visitat al menys un dels enllaços de l'àrea d'informació relacionada (denominada classe 1) o no l'ha visitat (denominada classe 0). Aquesta és la variable objectiu sobre la qual s'intenta realitzar una predicció. Concretament es vol predir els elements de la classe 1. Aquest procés genera un conjunt de dades amb 61.742 entrades on cada entrada és un valor 13-dimensional amb una etiqueta binària com a resultat d'aquell vector. La probabilitat estimada *a priori* de que un usuari accedeix a la informació addicional és troba al voltant del 33%.

### 4.3.3 Extracció de les variables de classificació

En aquest punt, cada mostra del conjunt de dades és un vector binari on cada element indica si l'usuari ha visitat o no un cert enllaç durant la seva navegació. L'ús de hiperplans ortogonals amb variables binàries pot produir arbres de decisió massa grans per poder capturar la complexitat del problema, especialment quan es requereix la unió de diferents subproblemes per ser resolt [107]. A més, un simple anàlisi del conjunt de dades mostra que els usuaris que visiten un enllaç d'una temàtica són molt propensos a visitar la resta d'enllaços de la mateixa temàtica, per lo que resulta interessant generar cinc variables noves, una per cada temàtica, resultat de combinar mitjançant la funció OR totes les variables relatives, es a dir, els enllaços relacionats. També es calcula una variable nova que indica el número d'enllaços visitats abans de visitar un dels enllaços considerats objectiu. D'aquesta manera, s'utilitzen un total de 19 variables per construir els arbres de decisió.

L'ús de variables binàries és molt interessant donat que es pot optimitzar l'algorisme de creixement dels arbres de decisió al només tenir que provar dos valors, cert i fals. La complexitat de l'algorisme d'ordenació necessari en cada etapa del procés de creixement es veu reduïda de  $O(N \log N)$  a  $O(N)$ . Anàlogament, la classificació de noves mostres també es veu afavorida al tenir que avaluar únicament una expressió binària. Finalment, els arbres de decisió construïts donen informació relativa a la importància de cada variable depenent de lo alt (en el sentit de la posició a l'arbre de decisió) que aquesta sigui escollida per l'algorisme en l'etapa de creixement.

### 4.3.4 Arbres de decisió de profunditat limitada

Per a poder realitzar prediccions de manera ràpida i que siguin realistes una opció és utilitzar arbres de decisió de profunditat limitada [115] amb possibilitat de combinar-los en un esquema d'encadenament. No tindria sentit construir sistemes de classificació que necessiten més preguntes que el número promig d'enllaços que visiten els usuaris. Els arbres de decisió es construeixen utilitzant un algorisme modificat del que es defineix a [38]. Es construeix un arbre de decisió complet de profunditat màxima  $d$ , i es poda amb

l'algorisme BFOS [38]. Degut a que  $d$  es limitada, en aquest cas  $d=8$ , és molt possible que només algunes de les branques de l'arbre de decisió complet siguin podades, i que només unes poques fulles de l'arbre estiguin etiquetades com a classe 1, la classe objectiu que es vol predir. L'estructura de l'arbre resultant pot ser utilitzada per extreure informació útil relativa al problema que es planteja. Els paràmetres dels algorismes de creixement i poda són els següents: s'utilitzen hiperplans ortogonals, la maximització de la entropia és el criteri de partició, la regla d'etiquetatge és per majoria (els costos de classificació són simètrics), i la poda es realitza d'acord a l'arbre de decisió més petit en número de fulles respecte a l'error de classificació mínim.

En relació a l'esquema d'encadenament, la idea bàsica és observar si la informació extreta en una etapa pot ser utilitzada en les següents, amb l'objectiu de millorar la precisió del sistema de classificació, tal com es descriu a continuació.

#### **4.3.5 Anàlisi de les dades**

A partir del conjunt de dades descrit es presenten dues alternatives: la primera és utilitzar tot el conjunt de dades, abastant tot el període de recollida de dades, per construir classificadors que proporcionin informació útil sobre la navegació que realitzen els usuaris. Per portar-ho a terme s'utilitzen tècniques específiques d'anàlisi del biaix i la variància [64] i de generació de conjunts d'entrenament i avaluació. Aquest procés rep el nom d'estàtic. La segona opció és dissenyar sistemes que únicament utilitzin una part de les dades seguint una partició temporal per l'entrenament, utilitzant la resta de dades per l'avaluació. És a dir, les dades disponibles fins a un cert moment són utilitzades per a entrenar del sistema de classificació, i la resta, que representen el futur, s'utilitzen per avaluar d'una manera realista el sistema construït. Aquest procés rep el nom de dinàmic. L'objectiu és veure si el coneixement extret amb el primer mètode, l'estàtic, es vàlid per a un escenari real o dinàmic.

### 4.3.6 Anàlisi estàtic

En aquest primer experiment s'utilitza tot el conjunt de dades per construir els conjunts d'entrenament i d'avaluació dels arbres de decisió construïts. Equival a realitzar un anàlisi a posteriori de l'ús del lloc web amb l'objectiu d'estudiar si la navegació dels usuaris pot ser predita en alguns casos. És per això que s'utilitza la següent configuració: el conjunt de dades amb 61.742 mostres es particiona utilitzant *N-fold cross-validation* amb  $N=3$ , i un total de 25 conjunts d'entrenament són generats utilitzant *bootstrapping* per a cada partició, amb l'objectiu d'estudiar la descomposició de l'error de classificació en biaix i variància [64]. Aquest procés es repeteix 5 cops i es fa el promig dels resultats. D'aquesta manera es construeixen un total de 125 arbres de decisió. La Taula 4.1 mostra les resultats obtinguts.

Classe $i / j$	0	1	Total	Valor $\alpha$
0	1720891	7539	1728250	99.6%
1	476677	367698	844375	43.6%
Total	2197568	375037	2572625	---
Valor $\beta$	78.3%	98.0%	---	81.2%

Taula 4.1: Resultats obtinguts amb el sistema de classificació estàtic.

El valor  $\alpha$  és la probabilitat de que un element d'una classe sigui classificat com a tal, mentre que el valor  $\beta$  és la probabilitat de que la classe assignada a un element sigui realment la classe en qüestió. En aquest cas, l'objectiu es predir satisfactòriament els elements de la classe 1, lo que es realitza correctament (un 98.0%) tot i que només per a una part d'ells (un 43.6%), tal i com es pot veure a la Taula 4.1. La precisió global del sistema és del 81.2%, que és bastant millor que el 67% que es podria esperar en el pitjor cas, i indica que un percentatge dels usuaris sí segueixen pautes de comportament en quant a la seva navegació pel lloc web.

La descomposició de l'error de classificació  $E=0.188$  en els components de biaix i variància ( $E=B+V$ ) és la següent;  $B=0.186$  i  $V=0.002$ , indicant que la majoria d'errors es cometten degut a una representació incorrecta del problema, possiblement a causa de

l'ús d'arbres de decisió de profunditat limitada amb hiperplans ortogonals, i no pas a la dependència dels conjunts d'entrenament, el que assegura una correcta configuració de l'experiment.

En relació a les característiques dels arbres de decisió construïts, la profunditat màxima mitja és 7.832, molt a prop del màxim preestablert ( $d=8$ ) però inferior, lo que indica que el màxim triat resulta adequat. La profunditat mitja, és a dir, el número de preguntes que es fan en promig per arribar a una resposta, és  $R=3.7$ , lo que indica que és impossible fer prediccions raonables amb menys de 4 preguntes aproximadament. La major part d'arbres de decisió comparteixen la mateixa estructura interna, de fet més de 100 dels 125 arbres de decisió construïts utilitzen les mateixes variables als primers nivells, lo que indica una clara agrupació de les mostres al voltant d'aquestes variables. Concretament, els visitants del lloc web que naveguen per les planes relatives a la temàtica d'economia són els més propensos a visitar alguns dels enllaços de l'àrea considerada objectiu, mentre que un gran nombre de visitants presenta un comportament qualificable de erràtic o impredecible. El valor elevat del paràmetre  $\alpha$  assegura que només aquells usuaris realment interessats en visitar els enllaços de l'àrea d'informació addicional són detectats, de manera que potser es podrien suggerir aquests enllaços utilitzant altres metàfores o disposicions visuals, com botons o finestres emergents en lloc dels enllaços, sense haver de tenir por en cap moment que es produeixin molèsties per a la resta d'usuaris. Aquest és un aspecte clau de tots sistema de personalització basat en recomanacions.

#### **4.3.7 Anàlisi dinàmic**

Per l'anàlisi dinàmic es suposa que es realitza una partició d'acord a una temporalització. Si el sistema recull dades que són processades mensualment, per exemple, s'obtenen els conjunts d'entrenament i avaluació tal com es mostra a la Taula 4.2. Evidentment es podria haver triat una altra unitat de partició, per exemple, setmanalment.

Mes	Entrenament / Acumulat	Avaluació
1	14086 / 14086	47656
2	12890 / 26976	34766
3	8008 / 34984	26758
4	6134 / 41118	20624
5	8561 / 49769	12063

Taula 4.2: Partició del conjunt de dades d'acord a una temporalització mensual.

La pregunta que ara es planteja és la següent: és possible obtenir resultats semblants als que s'han obtingut amb l'anàlisi estàtic? És a dir, en un escenari real, és possible anar construint sistemes de classificació basats en arbres de decisió de manera que es pugui obtenir informació similar? I, és possible, també, utilitzar combinacions de sistemes de classificació per millorar el resultat, tenint en compte tant les dades recents (per exemple les del darrer mes) com les acumulades?

És factible pensar que potser els usuaris segueixin el mateix comportament al llarg de tot el període de recollida de dades, per lo que tota la història acumulada pot ser útil per construir l'arbre de decisió. No obstant, també sembla plausible que els usuaris vagin canviant lleugerament el seu comportament, pel que l'ús de la història recent exclusivament hauria de ser més precís.

La Taula 4.3 mostra els resultats obtinguts per a un arbre de decisió construït utilitzant els conjunts d'entrenament definits per la partició temporal mensual, utilitzant només les dades més recents. Els valors  $\alpha$  i  $\beta$  mostrats són els corresponents a la classe 1.

Mes	$R$	$d$	$1-E$	$\alpha$	$\beta$
1	3.14	6	81.7%	44.3%	100.0%
2	3.71	8	82.1%	48.3%	97.0%
3	3.32	6	82.0%	43.8%	100.0%
4	3.12	6	79.0%	42.0%	100.0%
5	3.63	7	79.3%	41.5%	97.3%

Taula 4.3: Resultats obtinguts amb el sistema de classificació dinàmic.

La idea bàsica d'aquest experiment és veure com es pot capturar la variabilitat del comportament dels usuaris al llarg del període complet de recollida de dades. S'observa que per a tres períodes de temps, els mesos 1, 3 i 4, és possible construir sistemes de classificació que aïllin perfectament a potencials grups d'usuaris. En canvi, pels altres mesos els resultats són lleugerament pitjors que amb el classificador estàtic, tot i que es podrien usar-se donat que segueixen sent millors que un classificador aleatori ( $I-E=67\%$ ).

Si en lloc d'utilitzar únicament les dades corresponents al període de recollida definit (un mes) s'utilitzen totes les dades disponibles, s'obtenen els resultats mostrats a la Taula 4.4

Mes	$R$	$d$	$I-E$	$\alpha$	$B$
1	3.14	6	81.7%	44.3%	100.0%
2	3.93	8	82.1%	47.9%	98.1%
3	3.39	8	82.1%	44.3%	99.2%
4	3.51	7	81.8%	42.6%	98.9%
5	3.80	8	81.8%	42.0%	97.3%

Taula 4.4: Resultats obtinguts amb el sistema de classificació dinàmic acumulat.

El primer arbre de decisió és, evidentment, el mateix que en el cas anterior, ja que no hi ha història prèvia. És interessant observar que en aquest cas la probabilitat d'error és sempre menor que en el cas del classificador estàtic i, que excepte en el darrer període mensual, el valor  $\beta$  també es millor. Per altra banda, els arbres de decisió construïts són lleugerament majors, degut a la major mida dels conjunts d'entrenament.

#### 4.3.8 Conclusions

Els resultats obtinguts i presentats mostren que utilitzant el coneixement extret de les dades recollides durant un semestre, és possible predir el comportament dels usuaris, tot



i que una gran part d'ells mostren un comportament difícil de predir. Hi ha un petit grup d'usuaris que, clarament, es regeixen per uns mateixos principis o estratègies a l'hora de navegar pel lloc web: després de navegar per gairebé tots els enllaços de l'àrea temàtica relacionats entre sí, sol·liciten informació addicional sobre la universitat i la metodologia d'estudi. El sistema de classificació que aquí es presenta, pot permetre l'anticipació del sistema als objectius de navegació de l'usuari, mostrant-li la informació addicional, o bé ressaltant els enllaços relacionats, després que l'usuari hagi visitat un nombre reduït però suficient d'enllaços. Un dels punts crítics del tot sistema de classificació és la seva adaptabilitat als canvis lògics que es produiran al llarg del temps. La utilització de diferents esquemes pot fer front a aquests canvis de manera senzilla, mentre que la utilització d'esquemes de combinació basats en els paradigmes de votació i encadenament també es presenta com una possible alternativa per a la millora dels resultats obtinguts. El sistema presentat constitueix un exemple senzill que mostra les possibles aplicacions de l'anàlisi de dades com a eina de disseny centrat en l'usuari, que pot millorar la experiència d'usuari, i mostra com la informació implícita en la navegació pot ser usada per a millorar el sistema, en el que seria una aplicació concreta del nivell de sessió.

#### **4.4 Experiments inicials amb fitxers de log del campus virtual**

En aquesta secció es presenta el primer anàlisi de dades de log dels servidors del campus virtual de la universitat. Aquest estudi va tenir dues motivacions principals, per una banda conèixer tant la estructura dels fitxers de log generats pel campus virtual com entendre la seva semàntica per poder generar un mapa dels serveis oferts i, per altra banda, comprovar si a partir d'aquests fitxers es podia obtenir la navegació dels usuaris i informació sobre l'ús que fan del sistema.

##### **4.4.1 Navegació dels usuaris**

L'estudi dels patrons de navegació dels usuaris és important per validar el disseny que es va decidir al planejar i construir el sistema web i per avaluar la seva usabilitat. Molts cops, els aspectes de la interacció de les persones amb el sistema es treballen abans de la

posada en marxa del sistema, i el sistema es construeix en base a aquest treball [63, 155]. Un cop posat en marxa el sistema, s'espera que els usuaris, a partir del disseny original, tinguin una experiència d'ús satisfactòria. No obstant, és important realitzar un estudi de l'ús real del sistema un cop està en funcionament, així com periòdicament en el temps. Mitjançant l'obtenció dels patrons de navegació a partir dels fitxers de log, és possible fer una aproximació als patrons de navegació i d'interacció dels usuaris amb el sistema. En el cas particular d'un entorn d'e-learning com és un campus virtual, els usuaris porten a terme un conjunt de tasques que, tot i ser conegudes a priori, les combinacions d'accions que l'usuari pot realitzar resulten molt nombroses i, per tant, no es poden fer hipòtesis generalistes sobre l'ús que se'n fa dels serveis accessibles.

Per conèixer l'activitat dels usuaris en un lloc web, una bona alternativa és utilitzar els fitxers de log que es generen en els servidors. Això presenta alguns inconvenients, que s'han presentat en detall a la secció 2.3, inherents a les característiques dels fitxers de log: la mida dels fitxers, els habituals errors de registre d'entrada, la impossibilitat de registrar certes accions dels usuaris quan s'utilitzen altres tecnologies a més de HTML en la mateixa plana com per exemple Javascript, JSP o CGI, i sobre tot, disposar d'una identificació adequada i unívoca per a cada usuari. Per tant, si s'utilitzen les eines i recursos habituals i disponibles, no sempre s'aconsegueixen resultats útils i complets degut a les limitacions esmentades.

El preprocessat i anàlisi de dades que es presenten a continuació pretenen il·lustrar aquestes limitacions dels fitxers de log i, al mateix temps, explorar les alternatives existents per així poder aconseguir resultats interessants sobre la navegació dels usuaris. Coneixent amb profunditat l'estructura, la informació que aporten i les mancances que presenten els fitxers de log generats pel campus virtual, es podran proposar solucions adequades.

#### **4.4.2 Conjunt de dades**

A continuació es presenta l'anàlisi dels fitxers de log d'un dia escollit a l'atzar del semestre de primavera del curs 2003-2004. L'estudi persegueix, per una banda,

descobrir la informació sobre navegació dels usuaris que proporcionen els logs del campus virtual, i per l'altra, posar de manifest les seves limitacions. L'estructura i navegació de l'entorn virtual d'aprenentatge de la UOC que es vol analitzar és el que s'ha presentat en la secció 4.2 d'aquest capítol. Es tracta d'un sistema del que es coneix la seva estructura externa (interfície) i, amb menys profunditat, la seva estructura interna (arquitectura), donada la seva complexitat. Tot i així, s'ha hagut de fer un gran esforç per entendre el conjunt de dades, és a dir per desxifrar i entendre el significat de moltes de les línies que apareixen als arxius de log.

Amb un primer preprocés dels arxius de log s'obtenen indicadors suficients per aproximar, a grans trets, l'arquitectura del sistema. De fet, el preprocés de les dades no només està orientat a eliminar les entrades incorrectes i a reduir el volum d'informació a processar, sinó que depèn en gran mesura de la pròpia estructura del sistema. El preprocessat és una etapa activa i cíclica on es descobreix parcialment l'estructura interna de l'entorn que es vol analitzar. S'ha realitzat un estudi a curt termini amb les dades dels fitxers de log corresponents a un dia, amb l'objectiu d'identificar i presentar els problemes i dificultats que apareixen, i presentar una proposta basada en un sistema de captura de dades addicional que permeti obtenir la informació desitjada per a un anàlisi més complet.

Els arxius de log generats pels servidors web de cada *front-end* del campus virtual es generen a diari a les 01:00 hores seguint una rotació diària. L'arxiu de log, en el moment de realitzar aquesta anàlisi, s'obté de cada *front-end* actiu té una mida de l'ordre de 600 MB diaris. El fitxer resultat de fusionar els diferents fitxers de log de cada *front-end* presenta alguns dels problemes inherents dels fitxers de log i que el fan difícil de tractar: mida, conflictes d'adreces IP, conflictes d'hora i *web caching*.

- Mida dels fitxers. La gran mida dels fitxers de log fa que aquests siguin difícils de gestionar i analitzar. Un típic fitxer de log d'un servidor Apache corresponent a un dia laborable com és un dilluns, per exemple, pot contenir més de 19 milions d'entrades. Un cop processat aquest fitxer per a eliminar totes aquelles entrades que no identifiquen als usuaris, el fitxer resultant conté més de 12 milions d'entrades vàlides. Aquestes entrades són generades per més de 21.000

usuaris diferents. Així, la capacitat d'emmagatzemament i el temps de processador necessari per tractar aquests fitxers és, en qualsevol cas, enorme.

- **Conflictes d'adreça IP.** Els usuaris accedeixen al campus virtual mitjançant diferents proveïdors d'Internet. Alguns d'ells estableixen polítiques de compartició de la mateixa adreça IP, actuant com un *proxy* i, per tant, l'adreça IP no es pot utilitzar com identificador dels usuaris individuals, sinó únicament com identificador de proveïdors d'Internet. Per tant, totes aquelles entrades dels fitxers de log que no porten el número de sessió propi del campus virtual no poden ser, normalment, identificades de forma unívoca.
- **Conflictes d'hora.** A cadascuna de les entrades dels fitxers de log es pot trobar el dia i l'hora, amb una resolució de segons com a unitat mínima de temps. Degut a l'estructura multiprocés del servei de registre dels accessos al sistema, l'ordre en que les entrades apareixen en el fitxer de log pot no ser exactament el mateix en el que es van produir. La combinació d'aquests dos aspectes provoca que les entrades del fitxer de log puguin estar desordenades.
- **Web caching.** Cada petició HTTP que apareix al fitxer de log té un codi de retorn que pot ser utilitzat per detectar les possibles errades HTML. Aquest codi de retorn també es pot utilitzar, per exemple, per saber quan un usuari visita de nou alguna part del lloc web ja visitada anteriorment. Tanmateix, i tal com s'ha fet esment, alguns proveïdors d'Internet actuen com a *proxy*, per lo que el servidor web no rep cap petició i, per tant, aquesta informació no es registra al log.

Com a conseqüència d'això, a l'hora d'utilitzar els fitxers de log clàssics per a descobrir els camins i patrons de navegació dels usuaris, cal tenir en compte que es poden obtenir resultats que en alguns casos poden ser erronis o considerablement inexactes i, per tant, han de ser utilitzats amb prudència [57]. Tot i així, els fitxers de log són de gran utilitat per a l'obtenció de mètriques d'ús del lloc web, i en cap cas poden ser descartats a l'hora de portar a terme anàlisis d'ús.

#### 4.4.3 Identificació d'usuaris

La interacció dels usuaris amb campus virtual, la seva navegació, càrrega de continguts i selecció de botons i enllaços, deixen una traça que es pot utilitzar per estudiar el seu comportament de navegació en el campus virtual. Aquestes dades d'ús es generen pels *front-ends* o servidors web i es registren en fitxers de log clàssics.

En el cas concret del campus virtual que s'analitza, és possible identificar als usuaris atès que accedeixen a l'entorn d'aprenentatge amb un nom d'usuari i una paraula clau. Cada cop que un usuari accedeix al campus virtual, es genera un número de sessió únic que identifica a l'usuari segons el seu perfil i li dona accés a espais i continguts protegits. Aquest identificador de sessió és una cadena hexadecimal de 128 xifres de longitud i es calcula utilitzant una funció de hash que combina el nom d'usuari, l'hora d'accés i altres elements d'ús intern. Així doncs, tenint en compte els components de l'identificador de sessió, es pot obtenir una traça d'un mateix usuari, independentment que accedeixi al sistema en diverses ocasions i generi més d'un identificador de sessió. En relació a la privadesa i a la seguretat, és important destacar que és impossible obtenir el nom d'usuari a partir de la informació que queda registrada als fitxers de log, així com calcular nous identificadors de sessió vàlids. El número o identificador de sessió s'utilitza en totes les operacions que es consideren crítiques que l'usuari porta a terme en el sistema. Aquestes operacions poden ser, per exemple, accedir a la bústia personal o al seu expedient acadèmic. Per tant les successives peticions HTTP porten el número de sessió com a informació valuosa que s'utilitza per determinar si l'usuari té accés o no a les aplicacions.

Hi ha àrees del campus virtual on no és necessari l'identificador de sessió i, per tant, les peticions HTTP no el porten. Tampoc s'envia el número de sessió per a aquelles peticions relacionades amb els elements gràfics del sistema com són, per exemple, els fulls d'estil, les icones o les imatges. Aquest fet té grans implicacions de cara a seguir la traça de navegació de l'usuari, ja que hi haurà logs d'accés a àrees del campus virtual en les quals no es podrà identificar de quin usuari es tracta. Quan l'usuari abandona el sistema o bé resta inactiu durant un període de temps, l'identificador de sessió es

descarta i deixa de ser vàlid, i no es pot utilitzar de nou per tornar a entrar al campus virtual.

És important mencionar que a l'hora de realitzar aquest estudi no es va tenir accés a cap altra font de dades que no fossin els fitxers de log. D'aquesta manera, es podia obtenir el numero de sessió, que identifica a l'usuari durant una sessió de navegació, però no es va tenir accés a altres fonts de dades que permeten identificar a l'usuari d'aquella sessió, ja que l'objectiu era estudiar la navegació i no pas fer un creuament amb altres elements, com ara l'expedient acadèmic, per exemple.

#### **4.4.4 Preprocessat de les dades**

Com s'ha comentat, els fitxers de log es generen cada dia a les 01:00 hores, per lo que es disposa d'un fitxer per cada dia. Cada *front-end* genera el seu propi fitxer de log en format Apache combinat [4], per lo que tots han de ser fusionats per a poder disposar de totes les dades de forma conjunta. Les estadístiques d'accés al campus, disponibles per a l'ús intern de la universitat, mostren que l'hora de menys activitat són les 04:00 hores, per lo que per realitzar l'estudi d'un dia concret es fusionen els arxius de dos dies i es seleccionen les entrades generades a partir de les 04:00 del primer dia fins a les 04:00 del segon, tal com es mostra a la Figura 4.4. No obstant, aquest pas es realitzarà posteriorment en una altra etapa del preprocessat per així poder assegurar no perdre les entrades relatives als usuaris que estan navegant durant les hores en les que es realitza la selecció. La fusió dels fitxers de log dels dos dies genera un fitxer de més de 10 GB amb més de 2.700.000 entrades, lo que mostra els problemes esmentats sobre la grandària dels arxius i el temps necessari per al seu processat.

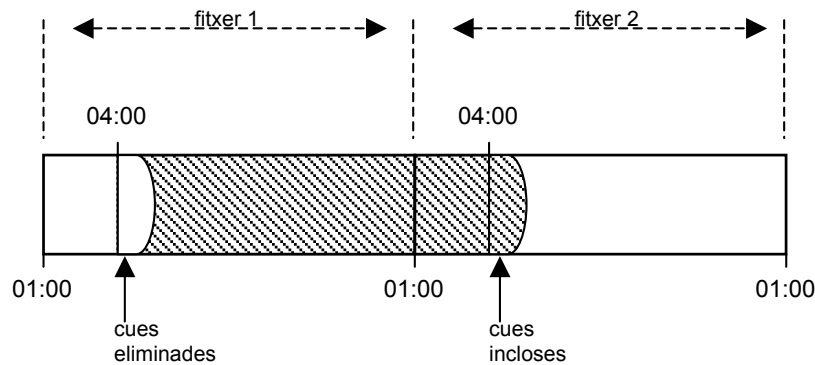


Figura 4.4: Esquema de tall i fusió de fitxers de log.

La primera etapa del preprocessat és la eliminació de totes aquelles entrades que són generades des de la xarxa interna pròpia de la UOC. També s'eliminen aquelles entrades corresponents als robots que naveguen pel campus de manera automàtica. Aquests robots s'encarreguen d'assegurar que els diferents subsistemes del campus virtual funcionen correctament, així com comprovar les connexions des de diferents tipologies de proveïdors d'accés. Aquest filtrat es realitza utilitzant l'adreça IP com a element de filtrat. Amb aquesta etapa, les 2.700.000 entrades es converteixen en 2.300.000.

La següent etapa del preprocessat és la corresponent a l'eliminació de totes aquelles entrades que fan referència a la càrrega i presentació d'elements gràfics i de format, com són imatges, icones o fulls d'estil. Aquests elements es carreguen de manera automàtica i no resulten significatius per a descobrir els camins i patrons de navegació, ja que la seva càrrega no sempre està directament vinculada a les decisions de navegació de l'usuari. En acabar aquesta etapa, el fitxer amb el que s'està treballant queda reduït a 600.000 entrades, amb una grandària de 2.5 GB.

A continuació, en la següent etapa, es realitza un filtratge per eliminar totes aquelles entrades corresponents a peticions que obtenen un error com a codi de retorn. Només les entrades amb codis de retorn del protocol HTTP 200 i 304 [4] són acceptades. Aquest pas elimina poques entrades tot i que evita tenir entrades incorrectes en els fitxers de log i que no aporten cap informació.

El preprocessat realitzat fins a aquest punt es genèric i no depèn especialment de les característiques del sistema web que s'està estudiant. Les etapes i filtres que es descriuen a partir d'aquest punt són específics del campus virtual de la UOC, tot i que en altres entorns web tindrien el seu equivalent. El campus virtual és compartit per tots els estudiants que estudien a la UOC, tot i que els estudis que segueixen o l'idioma sigui diferent. El filtre a aplicar consisteix en seleccionar a aquells estudiants que utilitzen l'entorn virtual en un mateix idioma. Un altre filtre s'utilitza per normalitzar els camins que poden ser seguits des de diferents accessos, és a dir, eliminar aquelles referències absolutes i deixar només les referències relatives, de manera que totes les crides a un mateix recurs apareixen de la mateixa manera en els fitxers de log. Els paràmetres que apareixen en els elements que formen les planes que es carreguen mentre l'usuari les visita també s'han de processar per a eliminar tota la informació que no aportí significat al camí recorregut.

En la darrera etapa del preprocessat, les entrades del fitxer log es manipulen per a evitar el problema d'adreces IP repetides que es produeix quan diferents usuaris que accedeixen al campus virtual i l'adreça IP que queda registrada en les entrades dels fitxers de log és la mateixa. Per a aquesta manipulació concreta es fa ús del fet de tenir als usuaris identificats amb el número de sessió. En els fitxers de log en format Apache combinat, el primer camp és l'adreça IP del visitant i el segon camp correspon a la seva identificació. Aquest segon camp acostuma aparèixer en els fitxers de log amb valor "-", donat que normalment no es disposa d'aquesta informació. La proposta que es va utilitzar consisteix en assignar un nou número d'adreça IP a cada usuari, que es genera de manera consecutiva (i a partir del 192.168.0.0). El número o identificador de sessió s'elimina dels paràmetres de les crides que hi ha a les peticions i s'utilitza en el segon camp, com a identificador. Amb aquesta manipulació el que s'aconsegueix és un fitxer de log que segueix el format estàndard de Apache combinat i, el que és més important, no presenta els problemes de col·lisió d'adreces IP i, a més, permet la identificació de cada usuari sense cap mena de confusió.

Per tant, el que s'ha aconseguit fins a aquest punt és no només un filtratge del fitxer de logs sinó un acondicionament d'aquest fitxer, de manera que s'aproxima a un format i



contingut ideals per al seu posterior anàlisi, on cada usuari s'identifica de manera unívoca i no hi ha col·lisions d'adreces IP.

La darrera etapa abans de l'anàlisi selecciona les entrades corresponents a un període de temps concret. Un dels problemes o dificultats que pot aparèixer al seleccionar les entrades entre dues dates concretes és la possibilitat de tallar una o més sessions d'usuari pel mig. Per poder minimitzar aquest efecte, és millor fer el tall per la hora de menys activitat; tal i com s'ha dit aquesta hora és a les 04:00. Al tenir les entrades identificades de manera unívoca pel número de sessió, el tall que es realitza no és estricte, sinó que es mantenen les entrades d'aquells usuaris que entren al sistema abans de l'hora de tancament. De la mateixa manera, s'eliminen totes les entrades corresponents a usuaris que apareixen a partir de la hora d'inici, és a dir, aquells que havien iniciat sessió abans d'aquesta hora. La Figura 4.4 mostra gràficament aquest procés per la selecció de les entrades corresponents a un dia, a partir dels fitxers de log de dos dies consecutius.

#### **4.4.5 Obtenció dels camins de navegació**

Un dels objectius d'aquest estudi va ser constatar si es podien utilitzar eines de mercat o de programari lliure per analitzar el fitxer de log corresponent a les entrades d'un dia. Una de les eines de lliure distribució disponible per l'anàlisi de camins de navegació que en el moment de portar a terme l'estudi oferia avantatges i característiques interessants és Pathalizer [18]. Pathalizer és una eina que a partir de fitxers de log genera una visualització dels camins seguits per la majoria dels visitants del lloc web. D'aquesta manera, el resultat que genera aquesta eina pot ser utilitzat per a prendre decisions de disseny del lloc web i millorar la seva estructura i navegació.

A partir d'un fitxer de logs amb format Apache, o altres formats que puguin ser descrits de manera estàndard, Pathalizer genera un graf dirigit i amb pesos. Per fer això, en primer lloc genera una llista d'events ordenats per l'adreça d'origen i el *timestamp* (la data i hora). Aquesta llista es converteix en una llista de grafs on els nodes són adreces web i els arcs són transicions o clics de l'usuari. A continuació, aquests grafs

s'unifiquen en un únic gran graf. En aquest graf generat els nodes que no són visitats amb certa freqüència, segons un valor fixat amb anterioritat, es filtren i es descarten. D'aquesta manera, a més d'eliminar possibles errades en les entrades dels fitxers de log, s'obtenen els camins recorreguts amb més freqüència. En acabar aquest procés, l'eina genera una representació gràfica del graf construït. Pthalizer funciona raonablement bé per l'anàlisi de fitxers de log generats per llocs web de caire estàtic i que no tinguin una excessiva programació en la part del servidor. Amb les proves realitzades, també s'ha comprovat que la eina presenta certs problemes amb fitxers de log de mida gran.

A l'utilitzar una eina de propòsit general amb els fitxers de log generats pel campus virtual de la UOC no sempre s'obtenen resultats significatius, tot i utilitzar fitxers de log preprocessats i adaptats al format requerit per l'eina. Els resultats de l'anàlisi que realitza Pthalizer a partir dels fitxers de log del campus virtual no tenen significat atès que el sistema perd el rastre de la navegació de l'usuari, principalment a causa de les tecnologies utilitzades en les planes del campus virtual i de la manera com es realitzen les crides i peticions. A mode d'exemple, a la Figura 4.5 es pot veure el resultat que produeix Pthalizer sobre el fitxer de logs processat tal i com s'ha descrit.

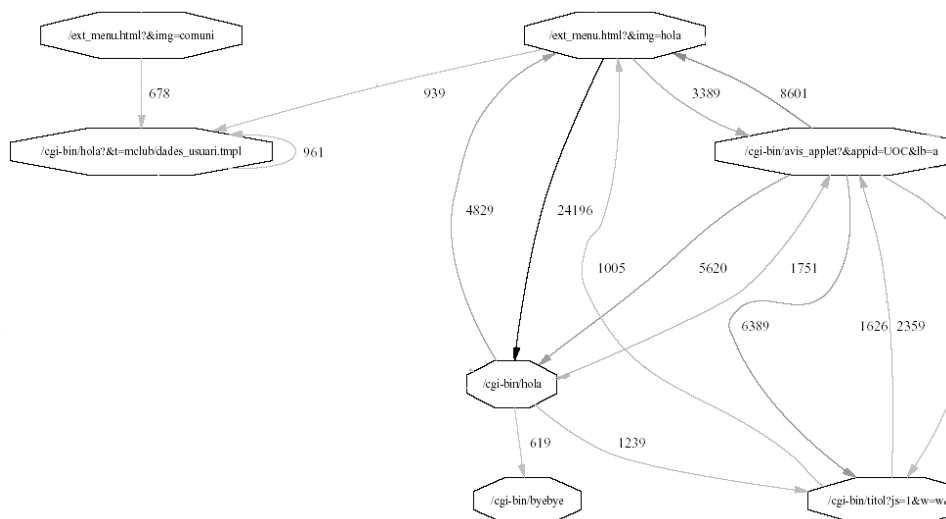


Figura 4.5: Resultat generat per Pthalizer.

A continuació es destaquen alguns dels inconvenients que apareixent a l'utilitzar Pthalizer. Els paràmetres que apareixen entre crides consecutives d'elements de la

plana web s'interpreten com si es tractés de planes web diferents, i això provoca que un mateix enllaç aparegui de manera repetida en els resultats. Això es pot resoldre parcialment complicant el preprocessat. Un altre inconvenient és que els camins de retorn es perden quan l'usuari accedeix a objectes de tipus JSP o a aplicacions específiques del campus virtual i això provoca que no es pugui construir un graf complet. Per altra banda, les entrades en els fitxers de log apareixen amb una resolució de segons i això provoca que les entrades puguin no estar ben ordenades i, per tant, es construeixen subgrafs que en realitat no són camins de navegació, sinó combinacions de possibles camins. Cal tenir en compte que això pot voler dir que segurament es fa un ús incorrecte del camp *referer* de les entrades del fitxer de logs. Per altra banda, aquest comportament resulta de gran utilitat per descobrir la estructura interna del sistema ja que mostra les relacions entre planes i elements que no crida i carrega directament l'usuari, sinó que es carreguen des d'altres planes. Aquest fet és clau per a establir una metodologia d'anàlisi de fitxers de log. El preprocessat és la etapa més important i s'ha de portar a terme combinant informació obtinguda amb eines com Pathalizer juntament amb la informació que es té del lloc web que s'està analitzant. Aquesta estratègia es descriu de manera similar a [54], on la estructura del lloc web es fa servir a priori per a establir els camins que s'estudiaran.

La conclusió clara que s'extreu d'aquest treball és que, tot i realitzar un processat dels arxius de log per a que puguin ser tractats i interpretats per un algorisme constructor de camins, és necessari dissenyar un mecanisme que permeti processar i analitzar els arxius de logs tot suplint les deficiències que presenten els arxius de log clàssics. El treball descrit en aquesta secció es va publicar a l'any 2004 a [120].

La proposta que es presenta en aquesta tesi per superar les limitacions que ofereixen els fitxers de log clàssics i, al mateix temps, fer front a la complexitat particular del campus virtual de la UOC, és un sistema de marques integrades en les entrades dels propis fitxers de log i la descomposició en nivells del procés d'anàlisi, per així poder abastar en la seva totalitat l'activitat i el comportament dels usuaris de l'entorn virtual d'aprenentatge de la UOC.

La proposta de marques que s'ha desenvolupat, així com la proposta d'anàlisi per nivells, s'ha descrit amb profunditat al capítol 3 d'aquesta memòria de tesi doctoral. La implementació d'aquesta proposta de marques ha comportat un laboriós procés de recollida, preprocessat i filtratge dels fitxers de log, que es descriu en la següent secció.

#### 4.5 Gestió i preprocessat dels fitxers de log

La recollida i gestió de les dades dels logs no és una tasca senzilla. En aquest treball de tesi s'ha fet un èmfasi especial en aquesta tasca, per així establir un procediment de recollida, emmagatzemament i preprocessat dels fitxers de log de manera que permeti realitzar un anàlisi adequat i que faciliti la tasca d'anàlisi de futurs investigadors. El resultat d'aquest procés són els camins de navegació dels diferents usuaris del campus virtual que, al mateix temps, constitueixen el punt de partida per altres estudis amb més profunditat.

El procediment dissenyat i implementat pel processament dels fitxers de log es descompon en dues fases; la primera consisteix en la recollida, gestió i preprocessat dels fitxers de log, mentre que la segona consisteix en el processat dels logs un cop filtrats per a obtenir els camins de navegació dels usuaris. La Figura 4.6 mostra esquemàticament tot el procés a partir de la recepció dels logs d'accés del campus virtual fins a la obtenció dels camins de navegació. Aquest procediment s'ha automatitzat i deixat en funcionament en els servidors del grup de recerca en el que s'ha realitzat aquest treball de tesi doctoral, tal i com es descriu en l'annex E.

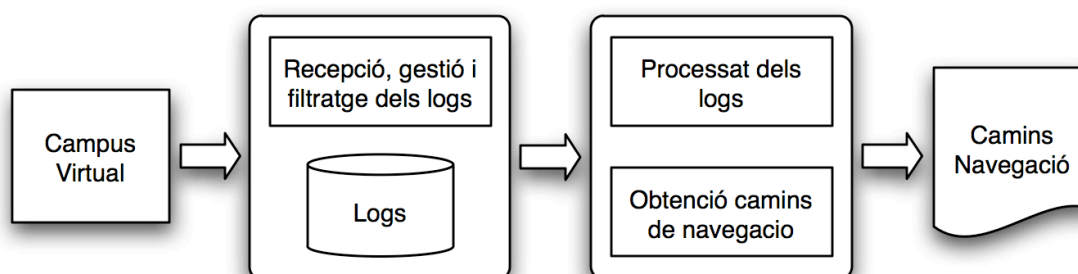


Figura 4.6: Esquema de tractament dels fitxers de log.

En aquest procediment hi intervenen principalment tres servidors i nombrosos algorismes i *scripts*. Els tres servidors intervenen en el processat dels fitxers de log són: haiti2.uoc.edu, einflinux3.uoc.edu i personal.uoc.edu. Els dos darrers formen part de la infraestructura del grup de recerca, mentre que el primer, haiti2, forma part de la infraestructura de sistemes d'informació de la universitat. Els servidor einflinux3 s'encarrega de realitzar les tasques corresponents a la primera fase del procés: la recepció, gestió, filtratge i emmagatzemament dels fitxers de log. L'altre servidor, personal, s'encarrega de realitzar tots els processos de la segona fase, el processat dels logs fins a l'obtenció dels camins de navegació. Les tasques corresponents la primera fase, la de recepció, filtratge i emmagatzemament, estan a hores d'ara completament automatitzades. Un cop s'ha comprovat que funcionen correctament, s'ha programat que s'executin cada dia a l'hora establerta. D'aquesta manera es poden dedicar més esforços en la fase d'obtenció dels camins de navegació i el seu anàlisi. Per altra banda, les tasques incloses en la segona etapa, la de processat i generació dels camins, al ser computacionalment més costoses, s'ha previst que es realitzin sota demanda. Aquests processos es poden executar directament a la línia de comandes del sistema operatiu del servidor o mitjançant un formulari web que s'ha desenvolupat *ad-hoc* i que de manera senzilla permet que els investigadors que vulguin tenir accés a les dades d'ús i navegació ho puguin fer sense necessitat de conèixer específicament els *scripts*, els processos implementats i la seva semàntica. A continuació, en aquesta mateixa secció es descriu la primera fase d'aquest procediment, la recepció i preprocessat dels fitxers de log. La segona fase, la que s'encarrega de l'obtenció dels camins de navegació, es descriu a la secció 4.6.2, donat que, pròpiament, ja forma part dels experiments del primer nivell d'anàlisi, el nivell de sessió.

El servidor haiti2 envia, a diari, els fitxers de log generats en els diferents servidors que actuen com a *front-end* del campus virtual. Aquests fitxers porten el nom del servidor i la data, i estan comprimits en format *gz*. També s'encarrega d'enviar el fitxer "loginhistory" del mateix dia que, com es descriu més endavant, conté la relació entre els números de sessió utilitzats aquell dia i l'identificador de l'usuari. Aquest servidor guarda els fitxers de log durant, aproximadament, 30 dies. Això vol dir que si hi ha algun problema en el tractament dels logs dels darrers dies, es poden tornar a recuperar

fàcilment. Un cop transcorregut aquest temps els fitxers es desen en cintes d'emmagatzemament i s'arxiven. La experiència fins al moment actual ha mostrat la dificultat en recuperar fitxers de les cintes d'emmagatzemament, fins al punt que alguns dels anàlisis que es presenten en les següents seccions s'han portat a terme mancants dades d'alguns dies concrets.

El servidor *einlinux3* és la màquina que rep els fitxers de log des del servidor *haiti2*. Resumidament, aquest servidor realitza el primer procés i filtratge dels fitxers que rep diàriament i genera un resultat que és enviat a una altra màquina. D'aquesta manera, aquest servidor es pot considerar bàsicament intermediari, rep els fitxers de log, els neteja, processa i filtra i, finalment, envia les dades llestes per ser analitzades amb el servidor personal.

Els fitxers que a diari es reben són els fitxers de log dels diferents *front-ends* del campus virtual i el fitxer "loginhistory". Aquests fitxers es reben al voltant de les 3 hores i es desen a la carpeta "/extra/webminer/incoming". El nombre de fitxers que es reben pot variar cada dia, i depèn principalment del nombre d'usuaris que s'han connectat al campus virtual i dels servidors *front-end* que s'han utilitzat.

El fitxer "loginhistory" conté la relació entre cada número de sessió utilitzat aquell dia i el "uid", que és l'identificador intern de campus de cada usuari, mentre que el número de sessió és un número de 128 xifres que identifica a un usuari en una sessió de navegació, des de que accedeix fins que surt del campus virtual. Si un mateix usuari accedeix al campus virtual més d'una vegada al dia, ho fa amb un número de sessió diferent, que es genera a partir del seu "uid" propi i altra informació.

Els logs del campus virtual d'un únic dia estan formats aproximadament per més de 40 arxius. Hi ha un arxiu per cada *front-end* i el nombre de *front-ends* pot variar depenent del dia. Cada fitxer de log generat d'un *front-end* recull les línies de log generades entre les 01 hores del dia anterior i les 00:59:59 del dia que porten en el nom del fitxer. El nom dels arxius està format per "campus" seguit del nom del *front-end* (que tenen nom de països), la data en que s'ha generat i el format de compressió, tots ells separats per

punts. D'aquesta manera, un arxiu de log d'un *front-end* té un aspecte com aquest: “campus.colombia.071017.gz” on:

- “colombia” es el *front-end* que ha generat el fitxer
- “071017” és la data en que s’ha generat, el 17/10/2007 i conté les línies de log del dia anterior
- “gz” indica que el fitxer està comprimit amb format *gzip*

Per exemple, el dia 17 d’octubre de 2007 s’han rebut els logs del dia 16 i s’han desat a una carpeta que porta per nom 071017. Aquesta carpeta conté 43 arxius de log dels diferents servidors, tots ells de diferents mides, destacant dos blocs molt diferenciats de fitxers, els que estan gairebé buits i els que tenen una mida al voltant dels 100 MB o superior. D’aquesta manera, un fitxer de log comprimit, tal i com es rep, corresponent a un únic *front-end*, té una mida aproximada de 100 MB. Aquest mateix fitxer, sense comprimir té una mida de 900MB i conté aproximadament 2 milions de línies. Si es consideren tots els fitxers comprimits rebuts en un dia, per exemple el 17 d’octubre de 2007, aquests ocupen 1,5 GB. Aquests mateixos arxius, un cop descomprimits ocupen 14 GB i contenen més de 34 milions de línies, el que els fa intractables sense un preprocessat.

Els diferents fitxers de log del campus virtual i el corresponent fitxer “loginhistory” es reben cada dia al directori “/extra/webminer/incoming/”. Un cop han estat processats i s’ha enviat el resultat al servidor personal, es desen en una carpeta que porta per nom el dia que s’han generat. Aquesta carpeta, juntament amb totes les dels altres dies es troba a “/home/webminer/bin/script\_agafar\_logs/comprimits/”. Així, per exemple, els logs del dia 17 d’octubre de 2007 es guarden en el directori “/home/webminer/bin/script\_agafar\_logs/comprimits/071017/”. El detall i el codi dels *scripts* que porten a terme el processat i filtratge dels fitxers de log es troba documentat a l’annex E. Per a evitar la saturació del disc del servidor, s’ha preparat un procés que cada primer dia de mes s’encarrega d’esborrar tots els fitxers de log ubicats en el directori d’emmagatzemament “comprimits”.

El procés de filtratge dels fitxers de log el du a terme un *script* anomenat “Generador.sh” ubicat en la carpeta “/etc/cron.logsDia/” i que s’executa de manera

automàtica cada dia a les 06:00 del matí. Aquest procés està estructurat en tres etapes principals, que es mostren gràficament a la Figura 4.7. El resultat és un únic fitxer de log filtrat, que s'envia al servidor personal.

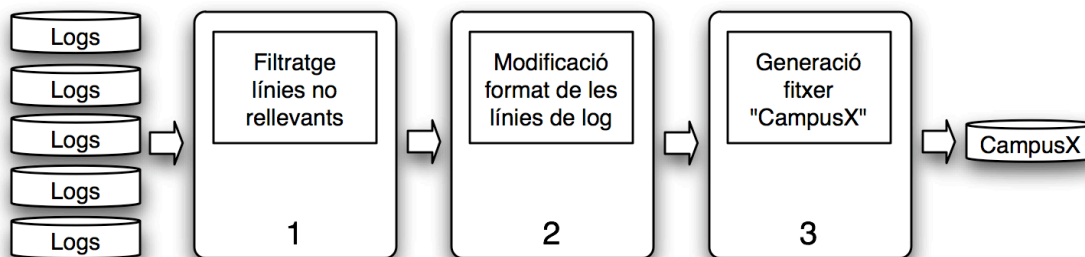


Figura 4.7: Procés de generació del fitxer “CampusX”.

### **Etapa 1: Filtratge**

En aquesta etapa es seleccionen les línies que són interessants a nivell de navegació dels usuaris i s'eliminen les que no interessin. Així, les línies que s'eliminen són aquelles que no aporten informació rellevant sobre la navegació i comportament dels usuaris del campus virtual. Les línies útils són aquelles que contenen la cadena “uocapp” i les diferents marques de seguiment (“ACCIO” o “TrackingTag”), tal com s'ha presentat al capítol 3 i amb més detall a l'annex D. Les línies que s'eliminen són, principalment:

- les que es generen en els diferents punts de treball de la UOC, ja que recullen les peticions de la navegació dels treballadors de la universitat i, per tant, queden fora de l'abast de l'anàlisi de navegació que es porta a terme. Aquestes línies es filtren per l'adreça IP i les que s'eliminen són “192\168\233\69”, “213\73\3[2,5,6,9]” i “213\73\41\.”,
- les generades per els robots o processos de control del campus virtual, com per exemple, els que obtenen estadístiques de temps de resposta de la plataforma,
- i aquelles que contenen peticions que no aporten informació d'ús, com per exemple, la càrrega de l'avís de missatge nou, la càrrega d'imatges i totes aquelles relacionades amb el format i la presentació.



Per tant, les línies que s'emmagatzemen són les que hi apareix alguna de les marques de seguiment, les quals apareixen precedides de la cadena "ACCIO=", així com "trackingTag" i "uocapp". La cadena "uocapp" és la que genera el campus virtual quan un usuari hi entra amb el seu usuari i clau personal, i és aleshores quan es genera un número de sessió.

## **Etapa 2: Modificació del format de les línies de log**

En aquesta etapa es modifica el format de les línies amb el propòsit de fer-les més llegibles i així facilitar el processat posterior. Aquesta etapa es porta a terme executant el *script* "camps\_ok.awk" ubicat en el directori "/home/webminer/bin/". Aquest *script* modifica el format de la data i hora i també s'encarrega d'eliminar de cada línia del log la informació que s'ha considerat poc rellevant de cara a l'anàlisi de la navegació, com per exemple, els camps corresponents al sistema operatiu i el navegador.

Així, per exemple, en aquesta etapa es reben les línies en un format com aquest:

```
89.129.180.225 - - [23/Oct/2007:01:00:53 +0200] "GET /cgi-
bin/byebye?s=640018d7b66be51753c5ca7d1f8912db5e7357f7e932d8936ba5bceb7
99986a09b7076217e7acffc00598bbaa7741627bd33e843eab4d0a492fa018e7305995
1&ACCIO=B_SURT&alert=false&text=Segur%20que%20voleu%20sortir%20del%20C
ampus? HTTP/1.1" 200 327 "http://cv.uoc.edu/cgi-
bin/uocapp?s=640018d7b66be51753c5ca7d1f8912db5e7357f7e932d8936ba5bceb7
99986a09b7076217e7acffc00598bbaa7741627bd33e843eab4d0a492fa018e7305995
1" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR
1.1.4322) "
```

I un cop processada i transformada, presenta un aspecte com aquest:

```
89.129.180.225 2007_10_23_01_00_53 GET /cgi-
bin/byebye?s=640018d7b66be51753c5ca7d1f8912db5e7357f7e932d8936ba5bceb7
99986a09b7076217e7acffc00598bbaa7741627bd33e843eab4d0a492fa018e7305995
1&ACCIO=B_SURT&alert=false&text=Segur%20que%20voleu%20sortir%20del%20C
ampus? http://cv.uoc.edu/cgi-
bin/uocapp?s=640018d7b66be51753c5ca7d1f8912db5e7357f7e932d8936ba5bceb7
```

99986a09b7076217e7acffc00598bbaa7741627bd33e843eab4d0a492fa018e7305995

1

### **Etapa 3: Creació del fitxer “CampusX”**

Un cop s’han realitzat les dues etapes prèvies per cadascun dels fitxers de log dels diferents *front-ends* que es reben a diari, el resultat es guarda en un únic fitxer i s’ordenen les seves línies d’acord a la data i hora. El fitxer generat s’anomena “CampusX” seguit de la data. Així, per exemple, el fitxer “CampusX” que conté els logs filtrats del dia 16 d’octubre s’anomena “CampusX.071017”. Un cop creat el fitxer resultat, es guarda al servidor personal a la carpeta “extra/webminer/incoming” i els fitxers de log que s’han processat es mouen a la carpeta d’emmagatzemament “/home/webminer/bin/script\_agafar\_logs/comprimits/”.

Amb l’objectiu de portar un control del procés de filtratge dels fitxers de log rebuts, aquest *script* genera un informe d’activitat. Quan finalitza el procés, es genera un fitxer de log que conté la descripció dels possibles errors que es puguin haver produït. Aquest fitxer de log es desa al directori “/var/log/Generador/”. Aquest log d’activitat també es pot rebre per correu electrònic i, així, conèixer amb immediatesa si hi ha hagut problemes amb la recepció i processat dels logs i poder prendre les accions adients. En el temps que es porta realitzant aquest processat de filtratge dels fitxers de log, els dos únics problemes que han aparegut han estat deguts al fet de que no s’han rebut els fitxers de log o perquè el servidor a esgotat l’espai de disc.

Fins a aquest punt s’ha descrit el procés de gestió i preprocessat dels fitxers de log. Els fitxers que s’obtenen ja estan filtrats i preparats per ser analitzats, donat que només contenen informació rellevant en relació a la interacció i navegació dels usuaris. En la següent secció es descriuen els experiments realitzats en el nivell de sessió a partir d’aquests fitxers filtrats i es descriu el procés d’obtenció de camins de navegació.

## 4.6 Experiments al nivell de sessió

Aquest primer nivell d'anàlisi, també anomenat nivell de curt termini, persegueix estudiar la navegació dels usuaris cada cop que es connecten a l'entorn virtual d'aprenentatge de la UOC. Així, el principal objectiu en aquest nivell és capturar la navegació real de cada estudiant pel campus virtual de la universitat, posant el focus especialment en aquelles accions i tasques relacionades amb el procés d'aprenentatge (accessos a la bústia i a les aules virtuals, utilització de recursos d'aprenentatge, etc.). La naturalesa dels logs generats pels servidors web del campus virtual, tal i com s'ha vist a l'estudi presentat a la secció 4.4, no permet portar a terme un seguiment adequat de l'activitat dels estudiants. Per aquest motiu s'han introduït un conjunt de marques que serveixen per realitzar un seguiment de les accions que realitzen els estudiants. La estratègia de marcatge així com el conjunt de marques proposades es descriu amb detall a la secció 3.5 i el mapa de marques es presenta a l'annex D. Aquestes marques han de permetre, per una banda, capturar la navegació dels usuaris i, per altra banda, el processat de les dades recollides i, descobrint així fets rellevants en relació al comportament dels usuaris i, d'aquesta manera, millorar el disseny tant dels itineraris formatius com de l'entorn virtual d'aprenentatge.

Una de les característiques principals de les marques que s'introdueixin és que no han d'interferir amb el sistema original i han de ser transparents per altres anàlisis i, sobre tot, per l'usuari final del sistema. Per exemple, seria inacceptable que una marca introduïda en una plana o en un enllaç provoqués un retard en el seu temps de càrrega. Aquest fet seria interpretat pels usuaris com un inconvenient important del sistema interactiu. A més, aquest és un element especialment important i crític per un entorn web interactiu amb milers d'usuaris, on un petit retard en el temps d'accés a determinats recursos pot provocar tal augment de la demanda que pot causar que el sistema sencera caigui. Aquest, tal i com s'ha presentat a la secció 2.3, és un dels motius principals pels que es proposa seguir una aproximació a l'anàlisi d'ús basada en la combinació de fitxers de log i marques, i no pas una aproximació de mètriques remotes. L'altre motiu principal és la infraestructura i recursos necessaris per registrar l'activitat i interacció de milers d'usuaris mitjançant servidors i bases de dades remotes.

Així, en aquest nivell, inicialment es van proposar dos experiments: el primer consisteix en descobrir quina és la primera acció que porten a terme els estudiants i el segon busca recollir la seqüència d'accions que realitzen els estudiants per així, posteriorment, analitzar els seus patrons de comportament.

#### **4.6.1 Experiment: plana d'inici del campus virtual**

Amb l'objectiu de validar la proposta de marques integrades descrita a la secció 3.5, a continuació es presenta un experiment que mostra el comportament dels usuaris un cop han entrat al campus virtual. L'objectiu de l'experiment és analitzar els fitxers de log i, mitjançant les marques introduïdes, determinar els camins seguits pels estudiants a partir del moment que es registren i entren en el campus virtual. Aquest objectiu presenta alguns aspectes interessants ja que, en el campus virtual, es pot arribar al mateix objectiu de diferents maneres. D'aquesta manera es pot avaluar el disseny de la plana d'inici com a àrea integradora dels accessos directes que es consideren més importants per als usuaris i, al mateix temps, valorar la importància dels elements que hi apareixen. Els resultats obtinguts en aquest experiment s'han publicat a [122].

El conjunt de dades que es va escollir correspon al semestre acadèmic actual en el moment de la realització de l'experiment, el semestre de tardor del curs 2006-2007. El semestre acadèmic es va iniciar el dia 20 de setembre de 2006. És aquest un dia especial del semestre on es registra una major afluència d'usuaris, ja que es tracta del primer dia de curs. La major part dels estudiants del curs que comença entren al campus amb la motivació d'accedir a les aules i consultar el pla docent, la planificació i els recursos docents de cadascuna de les assignatures que han matriculat. A més, també accedeixen a l'espai de comunicació de les aules per llegir el missatge de presentació del consultor i també per escriure i publicar el seu propi. El fitxer de log preprocessat corresponent a aquest dia té 7,8 milions de línies i ocupa al voltant de 1,5 GB. Conté informació de 105.000 sessions de les quals 42.000 són d'estudiants, d'un total de 20.000 estudiants diferents. Per a realitzar l'experiment s'han seleccionat 30.000 sessions d'un subconjunt de 14.600 estudiants, concretament aquells que estan matriculats d'una titulació homologada en llengua catalana. Una primera inspecció de les sessions revela que hi ha

algunes sessions que presenten un comportament estrany, probablement provocat per un malfuncionament del procés de *login*. És important destacar que cada semestre, en el primer dia de curs, el campus virtual arriba al seu límit d'usuaris concurrents i això provoca, en ocasions, que processos de *login* i sessions es quedin penjats i l'usuari hagi de tornar a repetir el procediment per entrar al campus virtual. Per exemple, hi ha sessions curtes que no presenten senyals d'interacció de l'usuari; per a aquest tipus de sessions s'ha decidit eliminar aquelles amb una durada menor de 3 segons, que en un dia com el que s'està considerant és el que tarda en promig la càrrega dels menús i la plana d'inici.

Des de la plana d'inici, tal i com es pot veure a les figures D.1 i D.2, els estudiants tenen accés principalment a:

- La bústia, mitjançant tres accessos diferents: la icona animada en forma bandera que apareix quan hi ha missatges nous, el missatge de notificació del nombre de missatges nous i el botó principal d'accés a la bústia situat al menú recursiu.
- El llistat d'assignatures que han matriculat.
- El llistat de les diferents bústies compartides d'aquestes assignatures: tauler, fòrum, debat, etc.
- L'aula virtual per cadascuna d'aquestes assignatures.
- Altres serveis del campus virtual: biblioteca, secretaria, comunitat, etc.

El resultat de l'anàlisi de les marques de navegació recollides mostra que la primera acció del 49,5% dels estudiants consisteix en accedir a les bústies compartides de les assignatures. Aquest és un comportament normal per un primer dia de curs, ja que en aquestes bústies es troben els missatges de benvinguda dels seus consultors, així com els missatges de presentació dels seus companys. A més, cada estudiant té el desig d'accedir a aquestes bústies per enviar el seu propi missatge de presentació.

Per a un 26,4% dels estudiants la seva primera acció és accedir a les aules virtuals. Un altre 12,0% dels estudiants selecciona anar directament a la bústia per l'enllaç de la plana d'inici que indica el nombre de missatges nous i només un 1,7% i un 0,6% seleccionen anar a la bústia pel botó principal i per la icona de la bandera, respectivament. El 8,1% dels estudiants accedeix a les aules mitjançant el llistat

d'assignatures que troba a la plana d'inici i l'1,7% restant realitzen altres accions. És interessant destacar la baixa que és la xifra dels estudiants que la seva primera acció no té a veure amb la bústia o les assignatures matriculades, malgrat que és un valor que presenta coherència si es té en compte que és el primer dia de curs i l'objectiu dels estudiants es conèixer les assignatures que cursaran.

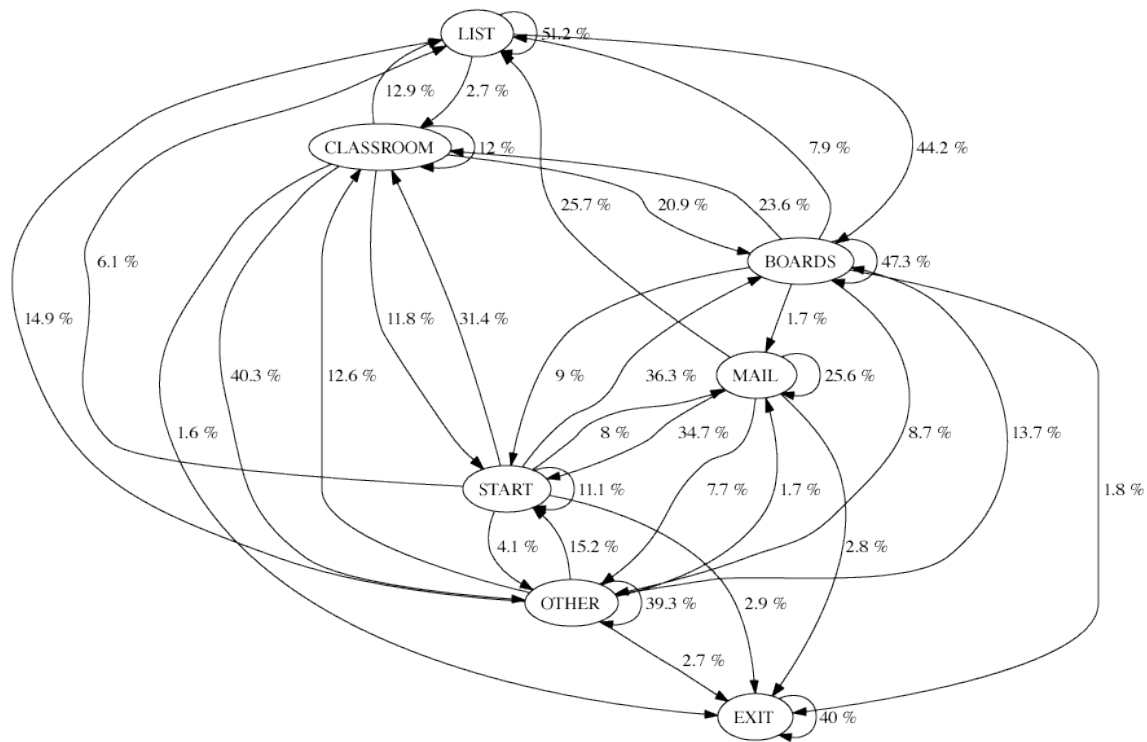


Figura 4.8: Camins de navegació més freqüents.

La Figura 4.8 mostra els camins més freqüents, seguits pels estudiants, entre els diferents espais del campus virtual. Només s'han dibuixat els enllaços rellevants entre els nodes. S'ha considerat que un enllaç és rellevant quan el percentatge d'estudiants que l'ha seguit és al menys del 5% per ambdós nodes. Cal destacar que el graf inclou tots els possibles camins i no només la primera acció realitzada des de la plana d'inici, d'aquí les diferències en els percentatges.

Els resultats d'aquest experiment es van publicar a [122]. La conclusió a la que es va arribar és que aquests experiments preliminars mostren que la aproximació als dos problemes coneguts de l'anàlisi de sistemes web de gran mida i complexitat mitjançant un sistema de marques integrades, és útil i adequat. En primer lloc, l'activitat i comportament dels usuaris es captura de manera més adequada des del moment que només es consideren les marques que són rellevants per l'anàlisi. En segon lloc, les dades que es generen només representen una fracció considerablement petita de totes les dades que es registren en els fitxers de log, de manera que es redueixen considerablement els requeriments d'espai i de temps necessaris per al seu procés. A més, els resultats de l'experiment realitzat són de gran interès per conèixer a l'estudiant i per millorar la plana d'inici, que és el que s'estava analitzant. Ara, finalment, es coneix amb dades reals que el primer dia de curs el principal interès dels estudiants és accedir a la informació de les assignatures, llegir els missatges de benvinguda i enviar el seu, i a partir d'aquí es poden prendre les decisions de disseny que es considerin adequades per adreçar els problemes generats per la gran concurrència del primer dia de curs. Els resultats d'aquest experiment mostren que es poden portar a terme experiments similars a d'altres àrees del campus virtual i obtenir diferents informacions valuoses per la millora de l'entorn i de la experiència dels seus usuaris.

A partir d'aquest primer experiment al nivell de sessió, s'ha definit i refinat el procés de processat dels fitxers de log, tal i com es descriu en la següent secció, per a l'obtenció dels camins de navegació dels estudiants del campus virtual.

#### **4.6.2 Experiment: procés d'obtenció de camins de navegació**

Amb el filtratge dels fitxers de log i la generació del fitxer "CampusX", tal com s'ha descrit a la secció 4.5, finalitza la fase de preprocessat dels fitxers de log i comença la fase d'obtenció dels camins de navegació dels usuaris. És aquesta, per tant, la segona fase del procés principal i que permet obtenir un resultat que constitueix un conjunt de dades molt ric per realitzar estudis amb profunditat.

Els processos que es realitzen en aquesta fase s'executen en el servidor personal. En aquest servidor s'emmagatzemen els fitxers de log preprocessats, s'executen els processos que generen els camins de navegació dels usuaris, i també és on està ubicada la plana web<sup>4</sup> que permet executar aquests processos amb una interfície senzilla i s'hi emmagatzemen els resultats. Aquest és el servidor principal del grup de recerca que, a més, s'encarrega de realitzar tasques relacionades amb altres projectes afins.

El procés que genera els camins de navegació dels usuaris del campus virtual s'ha programat en un *script* anomenat "ProcessarLogs.sh". Està ubicat a "/home/webminer/bin" i es pot executar des de la línia de comandes o des de la plana web construïda amb aquesta finalitat. Quan s'executa el procés de generació dels camins de navegació, aquest calcula tots els camins de navegació dels usuaris del campus virtual amb perfil estudiant, durant un període de temps especificat. Per tant, a l'hora de cridar al procés, cal especificar aquest període. L'*script* "ProcessarLogs.sh" crida a d'altres *scripts*, tots ells documentats a l'annex D. El procés de generació dels camins de navegació està format per cinc etapes principals, tal com es presenta a la Figura 4.9. Aquestes són la obtenció dels fitxers "CampusX" necessaris i la extracció de les línies vàlides, la obtenció dels números de sessió i el "uid" dels usuaris, la obtenció de les dades de navegació, la generació pròpiament dels camins i la obtenció del "idp" per futurs anàlisis i experiments.

---

<sup>4</sup> <http://personal.uoc.edu/logs/scriptlogs.php>



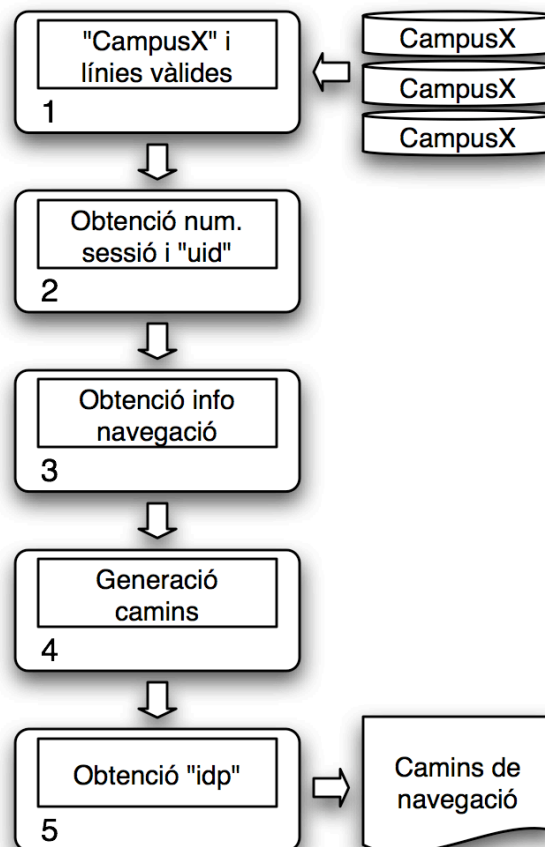


Figura 4.9: Procés d'obtenció dels camins de navegació.

### Etapa 1: Fitxers "CampusX" necessaris i línies vàlides

La generació dels camins de navegació es fa en un període de temps donat. Per tant, el primer pas del procés consisteix en crear una llista amb el nom de tots els fitxers "CampusX" inclosos en el període de temps establert, necessaris per a realitzar el procés. Per exemple, si es volen obtenir els camins del 17 d'octubre a partir de les 9 del matí fins a les 5 de la tarda del 18 d'octubre els fitxers "CampusX" necessaris serien CampusX.071018 i CampusX.071019. Cal tenir en compte que els fitxers "CampusX", tal com passa amb els fitxers de log, contenen les línies de log del dia anterior al nom que porten, a partir de les 01:00 fins a les 00:59 del dia següent. Un cop localitzats els fitxers "CampusX" necessaris es comprova que aquests fitxers estan disponibles en el disc del servidor.

En aquesta etapa es processen els fitxers “CampusX” necessaris, seleccionant les línies que tenen la hora compresa entre la hora d’inici i de fi que s’ha especificat al llançar procés. Totes les línies que es seleccionen es guarden en un fitxer temporal anomenat “linieslog”.

### **Etapa 2: Obtenció dels números de sessió i els “uid”**

Els números de sessió utilitzats s’obtenen a partir del fitxer “linieslog”. Aquest fitxer es filtra seleccionant els números de sessió i es guarden en el fitxer “sessions”. Per a obtenir els “uid” dels usuaris s’utilitzen els fitxers “loginhistory” corresponents als diferents “CampusX”. Cada fitxer “loginhistory” conté la relació dels diferents números de sessió de cada usuari. Donat que només cal obtenir un sol cop el “uid” de l’usuari, els fitxers “loginhistory” necessaris per l’experiment es filtren i s’eliminen les línies repetides. El resultat es guarda en el fitxer “Floginhistory”. Aquest pas és costós a nivell de temps de procés per la mida dels fitxers involucrats.

Finalment es genera un fitxer amb la relació de tots els números de sessió utilitzats en el període de temps establert i el seu “uid” corresponent. Aquest fitxer s’anomena “Fuidsessions” i cada línia està formada per un “uid” i un número de sessió. Per generar aquest fitxer es busca per cada numero de sessió de “sessions” el seu “uid” a “Floginhistory”. Cada cop que s’entra al campus es genera un número de sessió diferent, per tant aquest fitxer pot tenir varies línies amb el mateix “uid” i amb diferents números de sessió.

### **Etapa 3: Informació de navegació de l’usuari**

En aquesta etapa es processa el fitxer temporal “linieslog” per a calcular la informació següent: el número de vegades que un usuari inicia sessió en el campus (per l’interval de temps establert), l’hora d’inici de la sessió i l’hora de fi i, per acabar, el tipus d’usuari (si està especificat). El filtratge d’aquesta etapa es porta a terme amb l’algorisme

programat mitjançant el *script* “proces.awk” ubicat a “/home/webminer/bin/”, que revisa el llistat de “uid” i números de sessió de “Fuidsessions”.

#### **Etapa 4: Generació de camins**

En aquesta etapa s’obtenen totes les marques de navegació d’un usuari, és a dir, d’un mateix “uid” en cada sessió. El procés també el porta a terme l’algorisme programat al *script* “proces.awk”. Aquesta cadena de marques representa el camí de navegació de cada usuari en cada sessió.

El resultat d’aquestes dues darreres etapes es combina en un únic fitxer de resultat. Aquest fitxer presenta el següent format:

```
247798 1 20070531083252 20070531083320 ;LOGIN;B_SURT - -
175759 1 20070531103034 20070531115535
;LOGIN;Tauler;B_AULES;B_AULES;B_AULES;B_AULES;B_AULES;B_AULES;B_AULES;B_AULES;
B_AULES;B_AULES;B_AULES;B_AULES;plan.jsp;eval.jsp;B_AULES;B_AULES;plan.jsp;doc
um.jsp;docum.jsp;B_BUSTIA ESTUDIANT UOC
280299 3 20070531144015 20070531144931
;LOGIN;L_MNOUS;B_BUSTIA;B_BUSTIA;B_BUSTIA;B_INICI;B_BUSTIA;B_INICI;B_INICI;B_S
URT - -
47600 1 20070531194202 20070531194301
;LOGIN;L_MNOUS;B_INICI;plan.jsp;eval.jsp;plan.jsp ESTUDIANT UOC
213106 21 20070531175558 20070531175558 ;LOGIN - -
275220 2 20070531110419 20070531110805 ;LOGIN;plan.jsp;eval.jsp;B_INICI;B_SURT
ESTUDIANT UOC
247798 2 20070531090222 20070531090230 ;LOGIN;B_SURT - -
175759 2 20070531121702 20070531121711 ;LOGIN;plan.jsp;docum.jsp ESTUDIANT UOC
280299 4 20070531145019 20070531145120 ;LOGIN;B_INICI;B_BUSTIA;B_INICI;B_SURT
- -
85855 1 20070531154621 20070531154755 ;LOGIN;plan.jsp;eval.jsp;B_SECRETARIA
ESTUDIANT UOC
47600 2 20070531202528 20070531202555
;LOGIN;eval.jsp;plan.jsp;B_INICI;Debat;Debat ESTUDIANT UOC
213106 22 20070531181013 20070531181052 ;LOGIN;B_INICI - -
85855 2 20070531174058 20070531174116 ;LOGIN;plan.jsp;eval.jsp ESTUDIANT UOC
295808 1 20070531173125 20070531173127 ;LOGIN;B_BUSTIA - -
```

## Etapa 5: Obtenció del camp “idp”

Aquesta és la darrera etapa del procés i la porta a terme l'*script* “ProcessarLogs.sh”. En aquesta etapa s’afegeix al resultat de la etapa anterior el camp “idp” de cada usuari. El camp “idp” es l’identificador que permet accedir a les dades de cada estudiant a nivell d’aplicacions acadèmiques (matrícula, expedient, notes, etc). Afegir el camp “idp” als resultats permetrà estudis en profunditat sobre com afecta la navegació dels estudiants al seu rendiment acadèmic.

El resultat d’aquesta etapa és el resultat final del procés, i presenta el següent aspecte:

```
247798 485837 1 20070531083252 20070531083320 ;LOGIN;B_SURT - -
175759 435843 1 20070531103034 20070531115535
;LOGIN;Tauler;B_AULES;B_AULES;B_AULES;B_AULES;B_AULES;B_AULES;B_AULES;B_AULES;
B_AULES;B_AULES;B_AULES;B_AULES;plan.jsp;eval.jsp;B_AULES;B_AULES;plan.jsp;doc
um.jsp;docum.jsp;B_BUSTIA ESTUDIANT UOC
280299 511892 3 20070531144015 20070531144931
;LOGIN;L_MNOUS;B_BUSTIA;B_BUSTIA;B_BUSTIA;B_INICI;B_BUSTIA;B_INICI;B_INICI;B_S
URT - -
47600 258552 1 20070531194202 20070531194301
;LOGIN;L_MNOUS;B_INICI;plan.jsp;eval.jsp;plan.jsp ESTUDIANT UOC
213106 455890 21 20070531175558 20070531175558 ;LOGIN - -
275220 507878 2 20070531110419 20070531110805
;LOGIN;plan.jsp;eval.jsp;B_INICI;B_SURT ESTUDIANT UOC
247798 485837 2 20070531090222 20070531090230 ;LOGIN;B_SURT - -
175759 435843 2 20070531121702 20070531121711 ;LOGIN;plan.jsp;docum.jsp
ESTUDIANT UOC
280299 511892 4 20070531145019 20070531145120
;LOGIN;B_INICI;B_BUSTIA;B_INICI;B_SURT - -
85855 397654 1 20070531154621 20070531154755
;LOGIN;plan.jsp;eval.jsp;B_SECRETARIA ESTUDIANT UOC
47600 258552 2 20070531202528 20070531202555
;LOGIN;eval.jsp;plan.jsp;B_INICI;Debat;Debat ESTUDIANT UOC
213106 455890 22 20070531181013 20070531181052 ;LOGIN;B_INICI - -
85855 397654 2 20070531174058 20070531174116 ;LOGIN;plan.jsp;eval.jsp
ESTUDIANT UOC
295808 525131 1 20070531173125 20070531173127 ;LOGIN;B_BUSTIA - -
```

Com es pot veure, els camins de navegació obtinguts ofereixen informació rellevant sobre la navegació dels usuaris del campus virtual. Aquest resultat és important perquè presenta una informació que no s'havia obtingut fins a aquest moment i proporciona un coneixement valuós sobre el comportament real dels estudiants d'un entorn virtual d'aprenentatge. Al mateix temps, aquestes dades constitueixen un interessant punt de partida per a posteriors estudis, donat que poden ser analitzats en més profunditat i, d'aquesta manera, descobrir fets i patrons que afecten el procés d'aprenentatge i que permetran millorar la experiència educativa dels estudiants.

#### **4.6.3 Experiment: semestre de primavera del curs 2006-2007**

En aquesta secció es presenten els experiments realitzats amb les dades del semestre de primavera del curs 2006-2007, també anomenat 20062. Al llarg del semestre s'han recollit els fitxers de log, s'han processat i emmagatzemat per poder ser processats i obtenir els camins de navegació. Aquest ha estat el conjunt de dades principal de l'estudi realitzat. L'objectiu de l'experiment consisteix en obtenir les sessions de navegació dels estudiants, analitzar-les per a obtenir fets rellevants, i estudiar els resultats que s'obtenen.

#### **Conjunt de dades i preprocessat**

El semestre acadèmic que s'estudia va tenir una durada de 136 dies, des del dia d'inici de curs, el 28 de febrer, fins als dies de publicació de les notes finals, el 12 i 13 de juliol. Per aquest conjunt de dies es vol obtenir els camins de navegació dels estudiants i, a partir dels camins, descobrir fets interessants i rellevants sobre l'ús que se'n fa del campus virtual i les activitats que s'hi porten a terme. El conjunt d'estudiants que s'estudia, tal i com s'ha descrit a la secció 4.1, és el conjunt d'estudiants de les titulacions en català. Tots aquests estudiants comparteixen el mateix entorn virtual d'aprenentatge, així com les dates clau del semestre acadèmic. El nombre d'estudiants matriculats durant aquest semestre és 29.531, i les seves característiques principals així com la informació de caire més quantitatiu es presenten amb detall a l'annex A.

Durant el desenvolupament del semestre es van rebre a diari, llevat d'alguna excepció, els fitxers de log dels *front-ends* del campus virtual. Alguns dies puntuals, per motius aliens al procés de captura, els fitxers de log no es van rebre i, per tant, no es disposa de les dades corresponents. Hi ha també altres dies on es van rebre els fitxers de log però s'ha detectat que no són complets. Concretament, manquen les dades de 5 dies, que corresponen a les dates 24/03, 13/04, 14/04, 25/06 i 05/07. Així, el conjunt de dies dels quals es disposa dades és de 131. Els dies amb dades possiblement incompletes són, per la seva banda, 12/04, 30/05 i 24/06. Malgrat això, es considera que donat el nombre de dies i d'estudiants dels quals es disposa de dades, els resultats que es presenten en aquesta secció no es veuen afectats.

Els logs rebuts a diari s'han anat preprocessant seguint el procediment descrit a la secció 4.5. D'aquesta manera, s'han obtingut 131 fitxers de log filtrats, del tipus "CampusX". Aquests fitxers preprocessats ocupen un espai de disc de més de 150 GB. Per donar suport a la gestió i emmagatzemament d'aquest gran volum d'informació s'han necessitat dos discs externs de suport, cadascun d'ells de 500 GB. D'aquesta manera, un disc s'ha utilitzat de suport de totes les dades de l'anàlisi i, l'altre, ha realitzat principalment funcions de *backup*.

El procés amb l'algorisme de generació de camins que s'ha descrit a la secció 4.6.2 genera, per així poder calcular els camins de navegació dels usuaris, uns fitxers temporals d'una mida considerable. Així per exemple, els fitxers temporals corresponents a la generació dels camins de navegació d'un dia del semestre escollit a l'atzar ocupen més de 3 GB. Això provoca que no es puguin processar a l'hora totes les dades recollides durant el semestre, ja que provocaria el col·lapse dels discs del servidor i, per tant, la caiguda d'aquest. Per solucionar aquest inconvenient s'ha decidit dividir en parts el procés de generació de camins de navegació, escollint, com a unitat de procés, un dia. Tot i així, no s'ha agafat el dia natural (de 0 a 24 hores), sinó que s'ha estudiat quin és el moment del dia amb menys usuaris connectats, per a establir en aquell moment l'hora de tall. S'han estudiat les connexions de diferents dies del semestre agafats a l'atzar, i l'hora amb menys usuaris connectats és a les 04 hores, tal com es mostra a la Figura 4.10. Per tant, s'ha establert les 04 hores com a hora de tall

per processar i obtenir els camins de navegació del semestre, dia a dia, perdent el mínim d'informació possible (inferior al 0,1%).

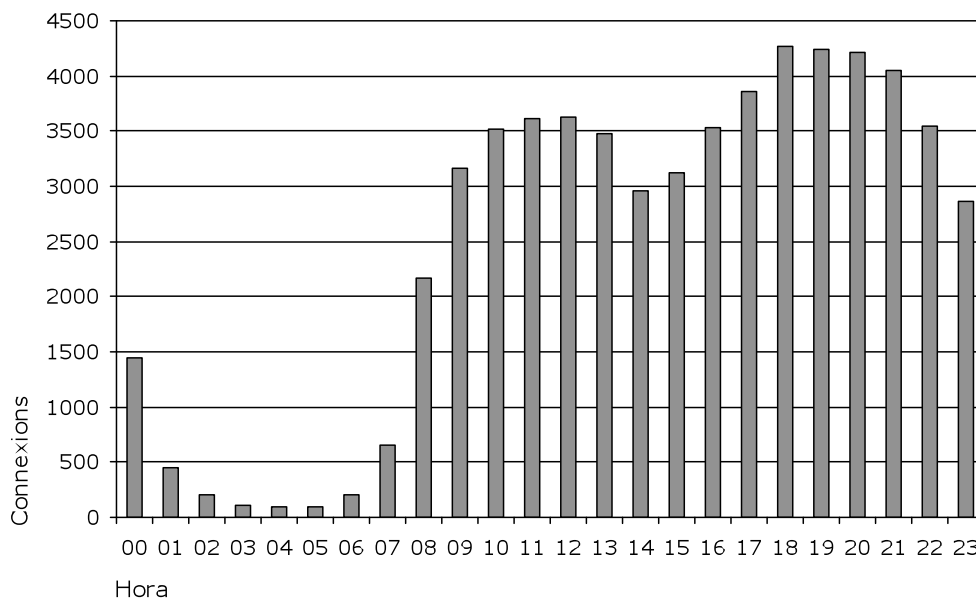


Figura 4.10: Connexions al campus virtual a les diferents hores del dia, d'un dia seleccionat a l'atzar.

Per processar les dades de tot el semestre es va programar un nou *script* que de manera automàtica generés tots els camins de navegació de tots els estudiants durant cada dia del semestre. Aquest procés va generar 131 fitxers de resultat, i va necessitar 11 dies per realitzar el procés, el que mostra la capacitat de càlcul necessària. Els resultats obtinguts s'han guardat en un únic fitxer que té una mida de 797MB i els camins de navegació que conté tenen un aspecte com aquest:

```
22163 180723 3 20070304173832 20070304173906 ;LOGIN;L_MNOUS
3479 110390 2 20070304230653 20070304235014 ;LOGIN;L_MNOUS;B_SURT
248293 486235 1 20070304212548 20070304213142
;LOGIN;B_BUSTIA_PI;B_SECRETARIA;B_INICI;B_SURT
22163 180723 4 20070304202756 20070304214027
;LOGIN;L_MNOUS;B_INICI;Tauler;plan.jsp;B_INICI;B_BUSTIA_PI;B_INICI;B_B
USTIA_PI
```

```

202068 450425 1 20070304165849 20070304170445
;LOGIN;Forum;Tauler;Tauler;plan.jsp;comms.jsp;B_AULES;B_AULES;Forum
248293 486235 2 20070304233042 20070304233132 ;LOGIN;B_INICI;B_SURT
202068 450425 2 20070304230411 20070304230834
;LOGIN;L_MNOUS;B_AULES;Forum;B_BUSTIA
22163 180723 5 20070304222549 20070304222609 ;LOGIN;B_BUSTIA_PI
277534 509787 1 20070304181851 20070304182647
;LOGIN;plan.jsp;plan.jsp;B_AULES;B_AULES;plan.jsp
277534 509787 2 20070304230616 20070304232923
;LOGIN;plan.jsp;docum.jsp;B_AULES;B_AULES;plan.jsp;docum.jsp;plan.jsp
212659 455509 1 20070304210453 20070304212831
;LOGIN;Tauler;Tauler;Tauler;
62538 344036 2 20070304204238 20070304204302 ;LOGIN;
292189 521895 1 20070304183500 20070304183534 ;LOGIN;plan.jsp

```

Aquests camins representen una font d'informació molt valuosa que fins ara no s'havia obtingut mai. L'anàlisi d'aquests resultats ja s'ha començat a aplicar a diversos projectes interns de la universitat, orientats a la millora de les aules i de la plana d'inici del campus virtual. Les dades de navegació, amb el format que s'obtenen, ja mostren molta informació sobre els usuaris i l'ús que fan del campus virtual. Tot i així, és interessant continuar processant i analitzant aquestes dades. De fet, aquests camins de navegació representen el punt de partida de nous anàlisis i estudis sobre el campus virtual i els seus usuaris i, com a mostra, a continuació es presenten alguns dels estudis realitzats.

### **Sessions de navegació**

Com s'ha comentat, tots els camins obtinguts del processat del semestre 20062 s'han guardat en un únic fitxer on cada línia representa una sessió de navegació pel campus virtual. Aquest fitxer conté 5.326.697 línies i, per tant, corresponen al nombre total de sessions del semestre. Entre aquestes sessions s'observa que n'hi ha algunes amb una durada de zero segons i es decideix eliminar-les ja que no aporten informació de navegació dels usuaris i segurament es corresponen a sessions fallides. Un cop eliminades aquestes sessions s'obté un fitxer amb 5.324.395 línies. D'aquestes sessions també es descarten aquelles que no comencen amb la marca "LOGIN", donat que es pot



tractar de sessions que, per diversos motius, no s'han capturat correctament i estan tallades. Un cop eliminades, s'obté un conjunt de 5.293.237 sessions, més del 99% de sessions del fitxer original.

El nombre mig de sessions per cada dia és lleugerament superior a 40.000, i la seva durada mitja és de 826 segons, uns 14 minuts. No obstant això, la moda de la durada és 7 segons, que és el que un usuari necessita per carregar la plana d'inici del campus virtual i fer una ullada per veure si hi ha missatges nous. Aquest valor és important que es tingui en compte en la implementació d'un sistema de personalització, ja que aquest hauria de ser capaç de generar i presentar informació útil en menys de 7 segons si es vol capturar l'atenció de l'usuari en la plana d'inici. Cal comentar que s'han capturat sessions amb una durada extremadament llarga, el que significa que el sistema de desconnexió per temps del campus virtual no sempre funciona bé.

A la Figura 4.11 es mostra el nombre de sessions segons la seva durada. Per facilitar la seva llegibilitat es mostren les sessions amb una durada inferior a 60 segons i, com es pot observar, un cop superat el valor més freqüent, 7 segons, la durada de les sessions decreix progressivament però suament. Aquesta distribució no es correspon amb cap distribució coneguda, sinó que segurament és la suma de dues distribucions, resultat de combinar dues operacions diferents: una normal per entrar a veure si ha novetats i sortir, centrada en els 7 segons, i una segona per a fer accions dins del campus virtual.

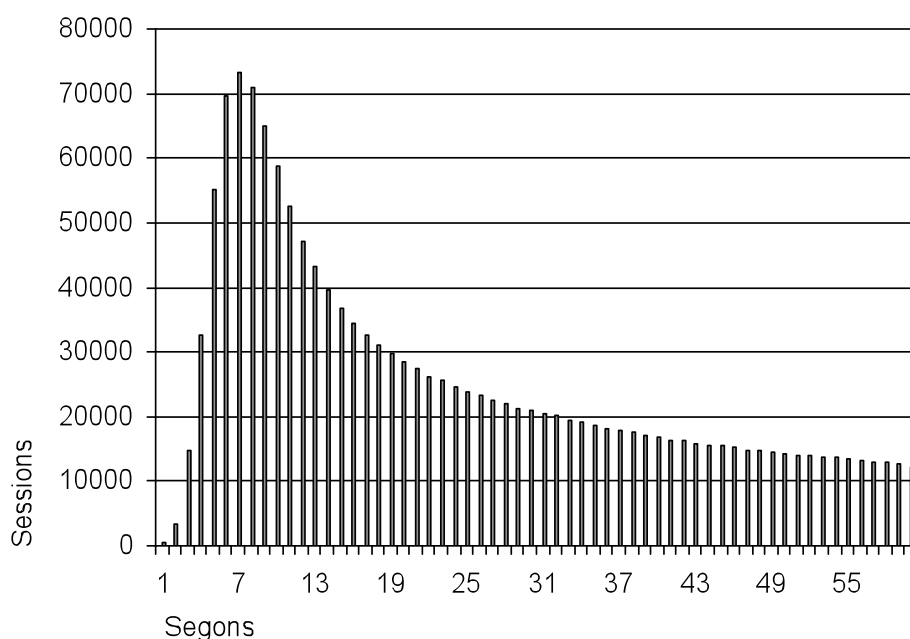


Figura 4.11: Nombre de sessions distribuïdes per la seva durada (sessions amb durada inferior a 60 segons).

Per altra banda, les sessions que tenen la marca “B\_SURT” com a darrera acció són les que permeten calcular la durada real de la sessió. Només 1.571.486 sessions de les 5.293.237 del conjunt de dades porten la marca, menys d’una tercera part del total de sessions. Això indica un comportament clar dels estudiants, on dues terceres parts no utilitzen el botó “Surt” per tancar la seva sessió del campus virtual. Si s’estudia la durada d’aquest conjunt de sessions s’obté que la moda ha canviat lleugerament respecte el total de sessions, i pren un valor de 8 segons, segurament pel fet de localitzar el botó “Surt” i fer-ne clic. La durada mitjana també canvia i baixa a 753 segons, gairebé 13 minuts.

Una informació interessant és la distribució de la durada de les sessions al llarg del temps, durant el desenvolupament del curs o semestre acadèmic. Per obtenir aquesta distribució s’ha utilitzat el conjunt de sessions de les quals es coneix amb certesa la seva durada, i s’ha calculat la mitjana de la seva durada per cada dia del semestre. El resultat es presenta a la Figura 4.12. Com es pot veure, la durada mitjana de les sessions varia dependent del dia i decreix conforme va avançant el curs. Aquesta dada té sentit si es té

en compte que normalment el contingut de les aules és estàtic i un cop, a l'inici de curs, els estudiants han descarregat el pla docent, el calendari i els recursos de l'assignatura. Per tant, no necessiten dedicar temps a navegar per l'aula i realitzen sessions més curtes, dedicades principalment als taulers de la plana d'inici i a la bústia personal.

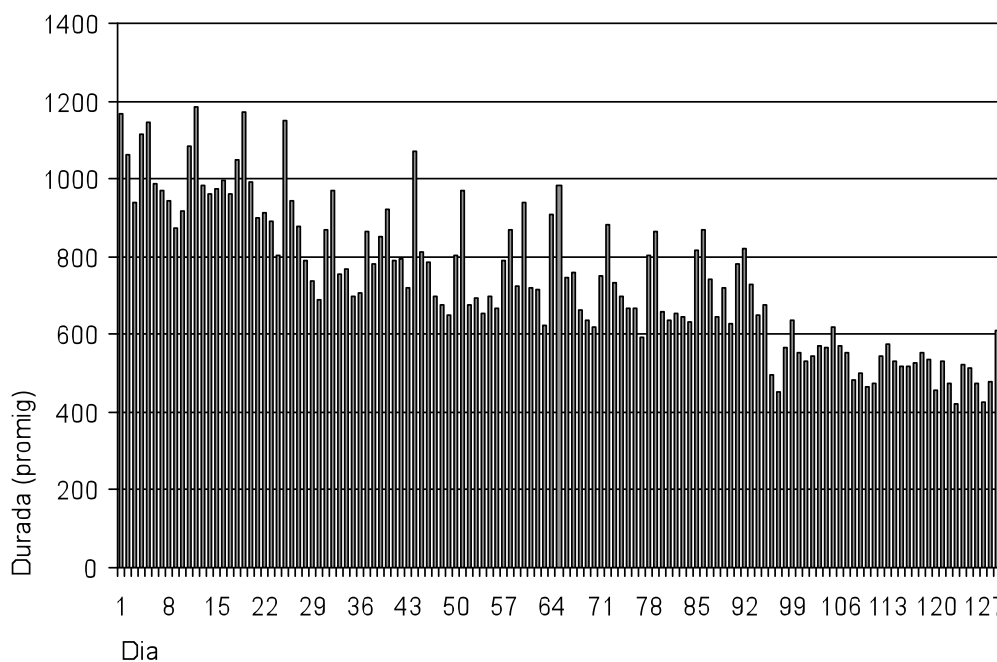


Figura 4.12: Distribució del promig de la durada (en segons) al llarg dels dies del semestre.

Com s'ha dit, cada dia del semestre es produeixen una mitjana de 40.000 sessions de navegació d'estudiants. Una altra informació rellevant al voltant d'aquestes sessions de navegació té a veure amb el moment del temps en que es produeixen, és a dir, l'hora del dia i el dia de la setmana. El nombre de sessions que es produeixen a cada franja horària del dia es mostra a la Figura 4.13, on es presenta la distribució de les sessions al llarg de les hores del dia. Per calcular-ho s'han tingut en compte totes les sessions obtingudes durant el semestre que s'estudia, el 20062, i s'han comptat el nombre de sessions en cada franja horària. Es pot observar que aquesta distribució general coincideix plenament amb la mostrada a la Figura 4.10, com era d'esperar.

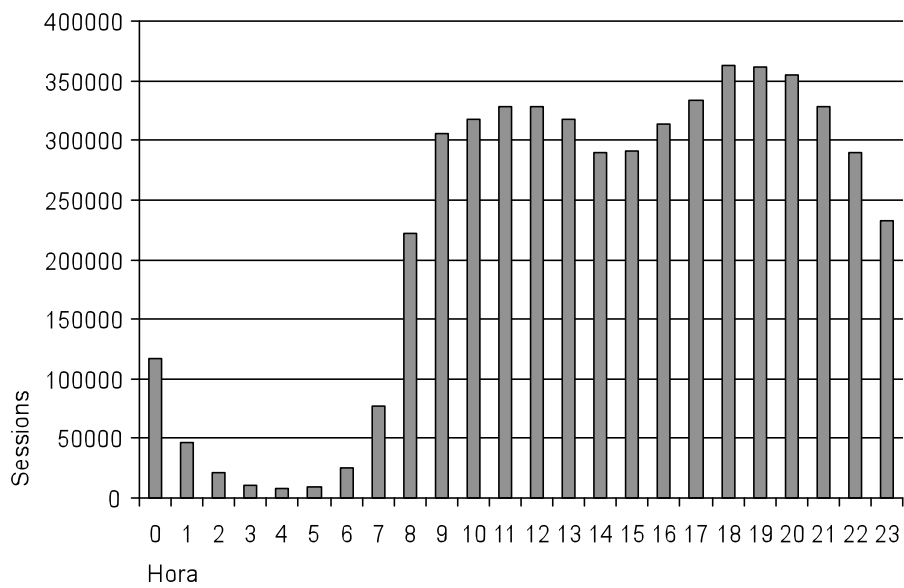


Figura 4.13: Distribució de les sessions al llarg de les hores del dia.

Per altra banda, a la Figura 4.14 es presenta la distribució de les sessions al llarg del dia de la setmana. Per calcular-ho també s'han tingut en compte totes les sessions del conjunt de dades.

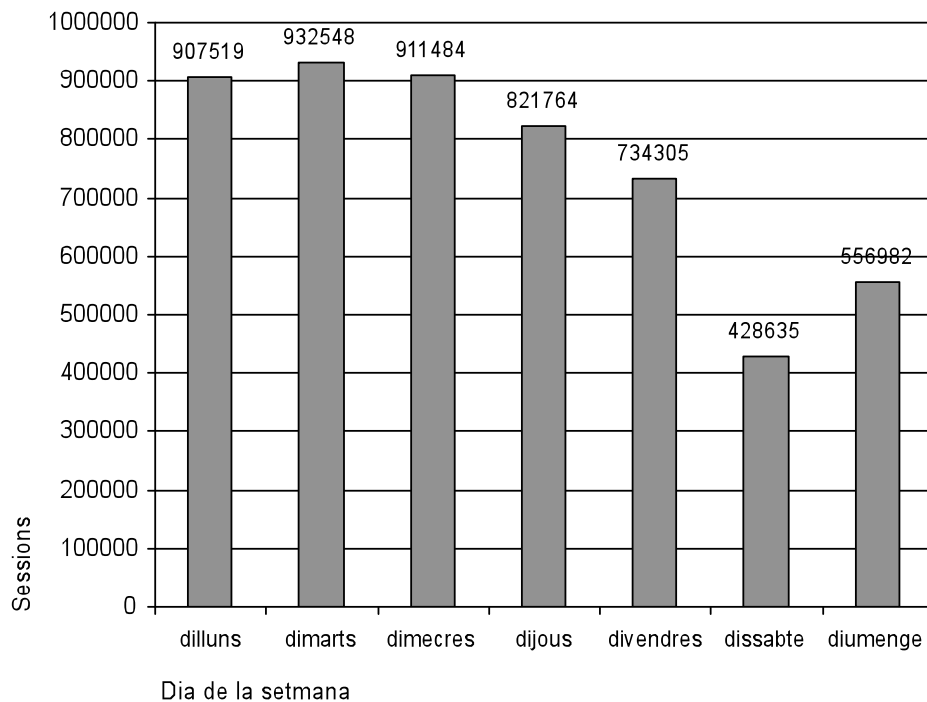


Figura 4.14: Distribució de les sessions al llarg dels dies de la setmana.

Com es pot veure, el dia de la setmana amb més sessions és el dimarts, seguit pel dimecres i el dilluns, és a dir, al principi de la setmana laboral. Curiosament, i a diferència del que es pot pensar dels estudiants en línia, els dies amb menys connexions corresponen al cap de setmana. Aquesta informació, per tant, ha de fer reconsiderar afirmacions del tipus “els estudiants de la UOC es connecten principalment els vespres i els caps de setmana”, freqüents a la universitat al parlar de la tipologia i perfil dels estudiants. Així, el patró de connexió real, com s’ha vist, és principalment entre setmana, durant tota la jornada laborable i els vespres.

També és interessant veure com es distribueix al llarg del semestre el nombre de sessions per dia. Per a facilitar la interpretació de les dades i poder mostrar els resultats afavorint la seva llegibilitat, s’ha aplicat una tècnica de suavització. La tècnica aplicada ha estat la mitjana mòbil amb una amplada igual a 7. El resultat d’aplicar la suavització és una corba més suau, tot i que provoca perdre les dades dels primers i dels darrers dies. No obstant això, l’objectiu no és calcular el resultat exacte per cada dia, sinó mostrar la tendència al llarg del semestre de la primera acció de cada sessió de navegació dels estudiants. El resultat es presenta a la Figura 4.15 i, com es pot observar, la tendència en el nombre de sessions per dia creix les primeres setmanes de curs, va decreixent al llarg del semestre i torna a créixer els darrers dies, quan es publiquen les notes. El primer dia del semestre és el que té més sessions, però a causa de la suavització de les dades que s’ha realitzat, no queda reflectit a la gràfica. El patró observat en relació al baix nombre de sessions en els caps de setmana s’observa clarament en aquesta figura, on cada pic inferior en el nombre de connexions es correspon amb un cap de setmana.

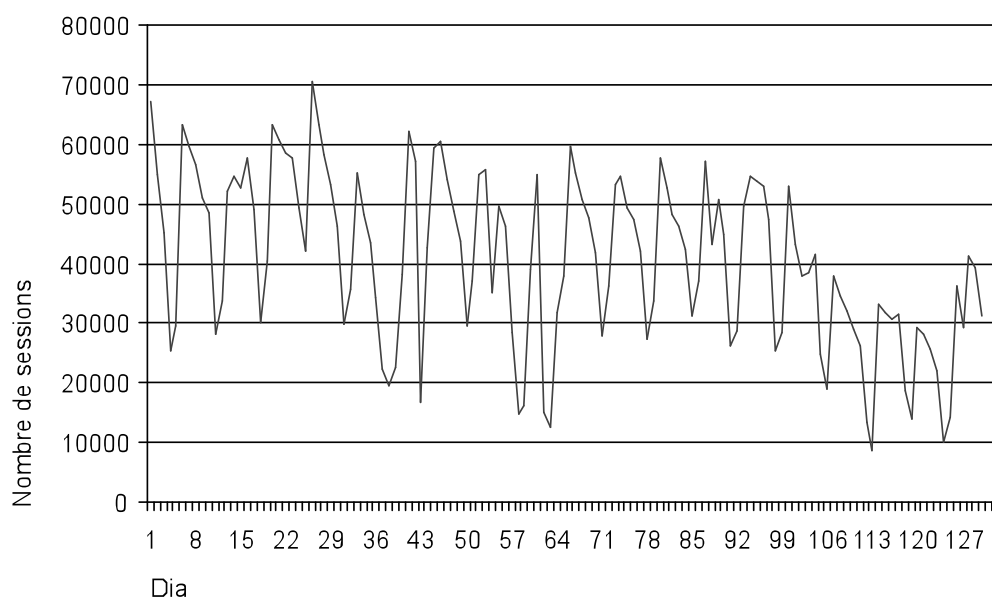


Figura 4.15: Distribució de les sessions al llarg del semestre.

Un altre àmbit que es pot estudiar a partir de les sessions de navegació dels estudiants és el relacionat amb les accions o clics que realitzen. És important tenir en compte que en aquest treball els clics o accions es corresponen necessàriament amb les marques introduïdes al campus i que queden integrades als fitxers de log. Així, quan es parla tan d'accions com de clics dels usuaris no es parla de clics reals, sinó dels corresponents a seleccionar cadascuna de les accions de les marques. Analitzant totes les sessions del conjunt de dades i el nombre d'accions que hi apareixen, el nombre mig d'accions per sessió és de 5,9 i el valor més repetit és de 2 accions per sessió.

A l'hora d'estudiar el nombre d'accions realitzades a cada sessió cal tenir en compte que totes les sessions porten la marca "LOGIN" corresponent a entrar al campus virtual amb el nom d'usuari i contrasenya i, per tant, això estableix un mínim que és que totes les sessions tenen com a mínim una marca o acció. Per altra banda, també cal diferenciar aquelles sessions on apareix la marca "B\_SURT" perquè, de la mateixa manera que abans, totes aquestes sessions tenen com a mínim dues accions. A continuació, es presenta la distribució de nombre d'accions per sessió. A la Figura 4.16 es mostra el nombre de accions que es realitzen a cada sessió, tenint en compte totes les sessions del conjunt de dades. Per la seva banda, la Figura 4.17 es presenta la mateixa

distribució però tenint en compte només aquelles sessions on apareix la marca “B\_SURT”. Per facilitar la llegibilitat de les dues gràfiques, s’han inclòs aquelles sessions que tenen 25 accions o menys.

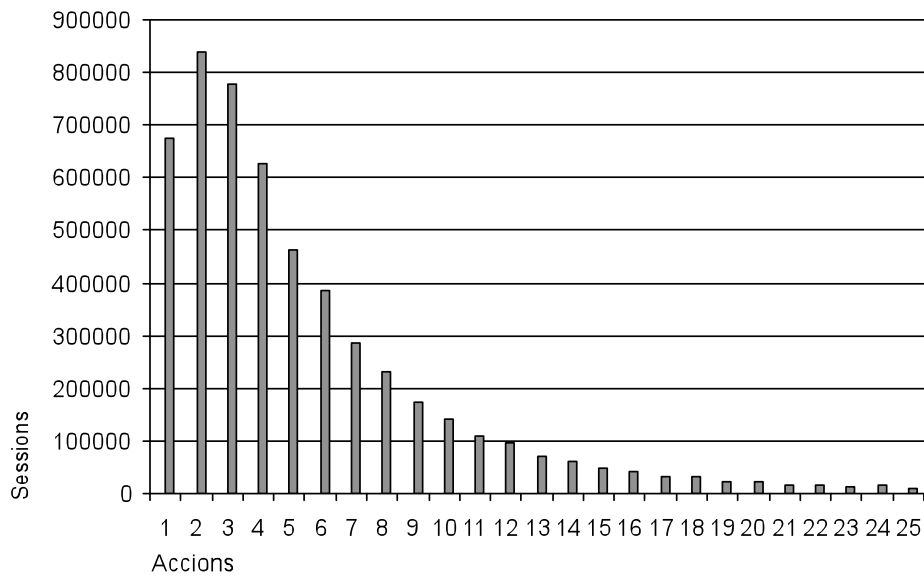


Figura 4.16: Nombre d’accions per sessió.

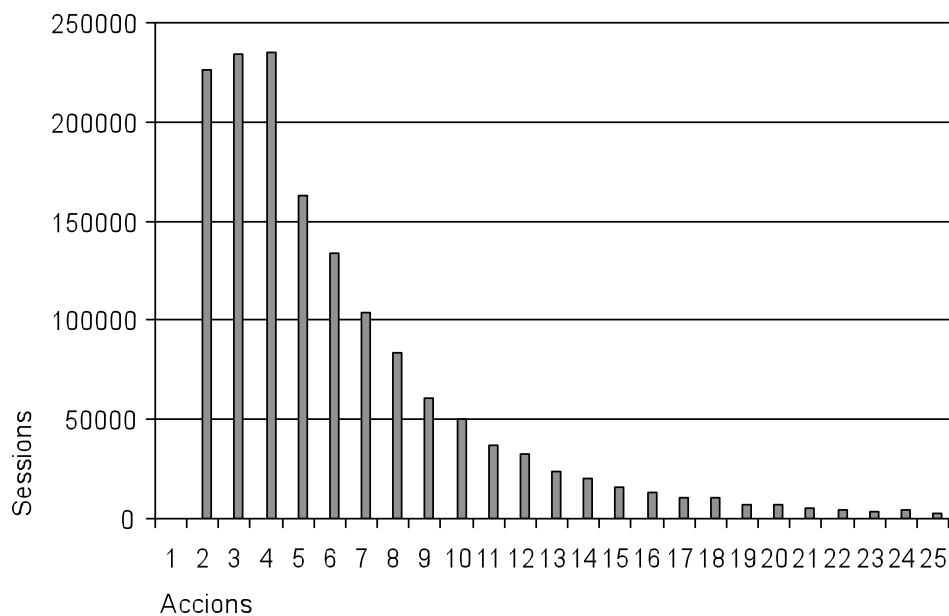


Figura 4.17: Nombre d’accions per sessió, de les sessions on apareix la marca “B\_SURT”.

És interessant veure que les gràfiques són molt similars i, òbviament, la que contempla les sessions amb “B\_SURT” està desplaçada cap a la dreta, ja que té en compte, com a mínim, una marca més, la corresponent al botó de sortir. Analitzant aquest subconjunt de sessions, el promig d’accions per sessió que s’obté és de 6,75 i el valor més repetit és 4, és a dir, la majoria de sessions tenen 4 accions, curiosament 2 més que si es considera tot el conjunt de sessions de navegació.

Un altre aspecte important, tal i com s’ha presentat als experiments de la secció 4.6.1, és l’estudi del primer clic o acció que realitzen els usuaris. Si es realitza una anàlisi similar amb el conjunt de sessions del semestre 20062, s’obtenen els resultats que es presenten a la Taula 4.5.

<b>Espai accedit</b>	<b>%</b>
Enllaç a les aules des de la plana d’inici	20,16%
Botó AULES	4,89%
Agrupació accessos aules	25,05%
Enllaç Missatges Nous des de la plana d’inici	16,06%
Botó BUSTIA de la plana d’inici	3,80%
Botó BUSTIA	2,42%
BANDERA	0,65%
Agrupació accessos bústia	22,95%
Tauler	11,99%
Fòrum	6,43%
Debat	3,39%
Agrupació accessos taulers des de la plana d’inici	21,81%
Botó SURT	4,62%
Botó INICI	4,39%
Botó SECRETARIA	2,28%



Botó COMUNITAT	1,82%
Botó GRUPS DE TREBALL	0,56%
Botó BIBLIOTECA	0,53%
Botó SERVEIS	0,26%
Botó PERFIL	0,21%
Botó NOVETATS I NOTICIES	0,06%
Botó SERVEI D'ATENCIO	0,06%

Taula 4.5: Primera acció dels estudiants en el campus virtual.

Així, com es pot observar, la primera acció més realitzada és accedir a les aules des de la graella d'aules que els estudiants troben a la plana d'inici del campus virtual. En segon terme, la primera acció més realitzada és accedir a la bústia mitjançant l'enllaç de la plana d'inici que indica el nombre de missatges nous. La tercera posició la ocupa l'enllaç al tauler del professor que apareix a la graella d'aules de la plana d'inici. Així, els accessos directes als elements relacionats amb el procés d'aprenentatge i la bústia que apareixen a la plana d'inici són els més utilitzats atorgant, d'aquesta manera, un gran valor a la plana d'inici del campus virtual per part dels estudiants, donat que prefereixen utilitzar els accessos directes de la plana d'inici abans que utilitzar els botons propis del menú per a accedir a aquests espais. Aquesta informació s'ha de tenir en compte a l'hora de modificar el disseny i arquitectura de la informació de la plana d'inici i dissenyar-la amb l'objectiu de proporcionar un millor servei i utilitat als estudiants, sense caure en el parany d'utilitzar-la per donar notorietat a informacions que no són prioritàries pels estudiants, és a dir, que no tenen a veure amb el seu procés d'aprenentatge.

Si s'agrupen les accions segons la seva temàtica, es pot comprovar que el 46,86 %, gairebé la meitat de tots els clics, fan referència als elements relacionats amb les aules virtuals, el 22,95 % inclou les diferents vies d'accedir a la bústia i el 30,19 % restant se'l reparteixen la resta d'accions possibles confirmant, d'aquesta manera, que el principal interès dels estudiants gira al voltant del seu procés d'aprenentatge. No obstant, si es té en compte que les accessos als taulers de les aules a la plana d'inici és una interfície concreta de la bústia, s'obté que un 44,76% dels clics fan referència a la

gestió dels missatges pendents, podent concloure que el principal interès dels estudiants quan entren al campus virtual és la gestió del correu electrònic.

També és interessant obtenir l'indicador de la primera acció, però del primer dia del semestre, i així poder-lo comparar amb els resultats obtinguts tant als experiments de la secció la secció 4.6.1 com amb els resultats de la primera acció de tot el semestre acadèmic. Els resultats es presenten a la Taula 4.5.

<b>Espai accedit</b>	<b>%</b>
Tauler	23,75%
Fòrum	1,16%
Debat	0,44%
Agrupació accessos taulers des de la plana d'inici	25,37%
Enllaç a les aules des de la plana d'inici	14,57%
Botó AULES	8,35%
Agrupació accessos aules	22,92%
Enllaç Missatges Nous des de la plana d'inici	13,34%
Botó BUSTIA	2,23%
Botó BUSTIA de la plana d'inici	2,07
Agrupació accessos bústia	17,66%
Botó INICI	9,43%
Botó SURT	3,11%
Botó SECRETARIA	1,99%
Botó PERFIL	0,37%
Botó BIBLIOTECA	0,28%
Botó AGENDA	0,26%
Botó GRUPS DE TREBALL	0,24%
Botó CERCA DE PERSONES	0,19%
Botó PERSONALITZA	0,15%
Botó SERVEI D'ATENCIO	0,10%

Taula 4.6: Primera acció dels estudiants en el campus virtual el primer dia de curs.

Com es pot veure, els espais més accedits són les bústies de les aules de la plana d'inici, especialment el tauler del professor. El segon espai més visitat com a primera acció són les aules i el tercer és la bústia. És interessant comprovar que els resultats que s'obtenen amb les dades que s'estan estudiant, les del semestre de primavera del curs 2006-2007, coincideix amb els resultats de l'experiment presentat a la secció 4.6.1 on s'han analitzat les sessions del primer dia de curs del semestre de tardor del curs 2006-2007, el semestre previ. També és interessant veure la diferencia de comportament dels usuaris si es compara amb la primera acció que realitzen durant el curs, presentada a la Taula 4.5, on la primera acció més realitzada és la de accedir a les aules.

Al tractar-se del primer dia del curs, és un dels dies del semestre acadèmic on hi ha més connexions al campus virtual, tal com es pot veure a la Figura 4.15. Això provoca que pugui haver moments en els quals el sistema no pot atendre correctament les peticions dels usuaris, especialment la càrrega de la plana d'inici del campus. Això podria explicar el gran nombre de sessions, un 9,43%, que la seva primera acció és seleccionar el botó d'inici, que carrega de nou la plana d'inici del campus virtual. El fet que els estudiants prefereixin accedir al tauler, amb un 23,75%, abans que a les aules o la bústia es pot explicar perquè en el primer dia de curs el consultor envia el missatge de benvinguda i presentació de l'assignatura al tauler de l'aula i, quan els estudiants entren al campus virtual, ho troben indicat amb una bandera a la plana d'inici. De igual manera que s'ha constatat amb l'estudi del primer clic al llarg del semestre, a la Taula 4.5, el principal interès dels estudiants el primer dia de curs té a veure amb el seu procés d'aprenentatge, i per això accedeixen principalment als taulers i les aules de les assignatures que han matriculat, prescindint d'accedir als altres espais del campus virtual. Al mateix temps, aquest interès pel procés d'aprenentatge es concreta, amb el 43,03% dels clics, en la gestió del correu electrònic.

Una altra informació rellevant que es pot extreure analitzant les sessions de navegació dels estudiants consisteix en conèixer com el primer clic o acció que els estudiants realitzen en el campus virtual en cada sessió de navegació es distribueix i va canviant al llarg del semestre acadèmic. Per obtenir aquest resultat s'ha calculat per cada dia la

distribució del primer clic i s'ha aplicat una suavització de les dades amb una finestra de mida 7. Així, a la Figura 4.18 es presenta la distribució al llarg del temps de la primera acció més freqüent en la navegació pel campus virtual. Com s'ha mostrat a la Taula 4.5, les accions més freqüents son: l'agrupació d'accessos a les aules (enllaç a les aules des de la plana d'inici i el botó "Aules"), l'agrupació d'accessos a la bústia (enllaç als missatges nous des de la plana d'inici, botó bústia de la plana d'inici, el botó a la bústia i la icona de la bandera) i l'agrupació d'enllaços als taulers de les aules que apareixen a la plana d'inici (Fòrum, Debat, Tauler). Per comparar-ho amb serveis menys utilitzats, a la figura s'ha afegit l'ús del botó "Secretaria". Altres serveis docents com "Biblioteca" no es mostren en aquesta gràfica donat que estan un ordre de magnitud per sota, i no són representatius.

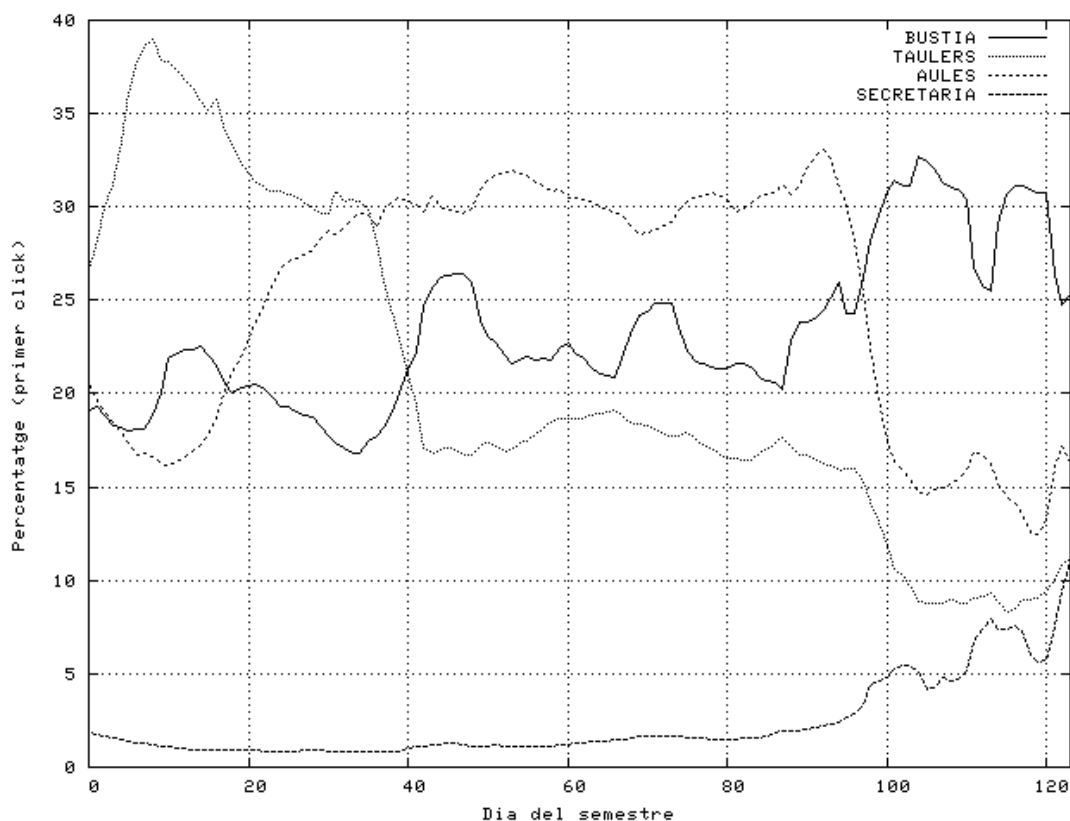


Figura 4.18: Distribució al llarg del curs de la primera acció en la navegació pel campus virtual.

Així, l'ús de la bústia es manté mes o menys constant al llarg del curs, pujant al final, just en el moment que finalitza l'acció docent a les aules i baixa l'activitat dels taulers. L'ús de la secretaria també es manté relativament constant durant el semestre i experimenta una gran pujada a final de curs, quan els estudiants han de gestionar les proves finals, les notes i les possibles incidències relacionades. L'ús dels taulers de les aules és molt alt a l'inici de curs, ja que és l'espai utilitzat per presentar l'assignatura i per conèixer als companys de l'aula. La selecció dels taulers de la plana d'inici com a primera acció va baixant, de manera esglaonada, durant el curs. Quan els taulers deixen de presentar interès pels estudiants, aquests seleccionen accedir a les aules com a primera acció, ja que és l'espai on troben el calendari d'activitats que han de realitzar a cada assignatura.

Una altra tendència que es vol observar en la navegació dels estudiants és l'accés a les diferents àrees del campus virtual i com aquest accés varia al llarg del curs. Concretament, sabent que els espais més utilitzats són les aules i la bústia, és vol veure l'ús dels altres espais principals del campus virtual: Comunitat, Serveis, Secretaria, Recerca, Biblioteca i Notícies. Per obtenir els indicadors s'han comptabilitzat el nombre de sessions que inclouen al menys una marca que indica l'accés a un espai concret. Un cop calculat el nombre d'accessos a cada espai, s'ha aplicat la mateixa tècnica de suavització de les dades. Els resultats obtinguts es mostren a la Figura 4.19.

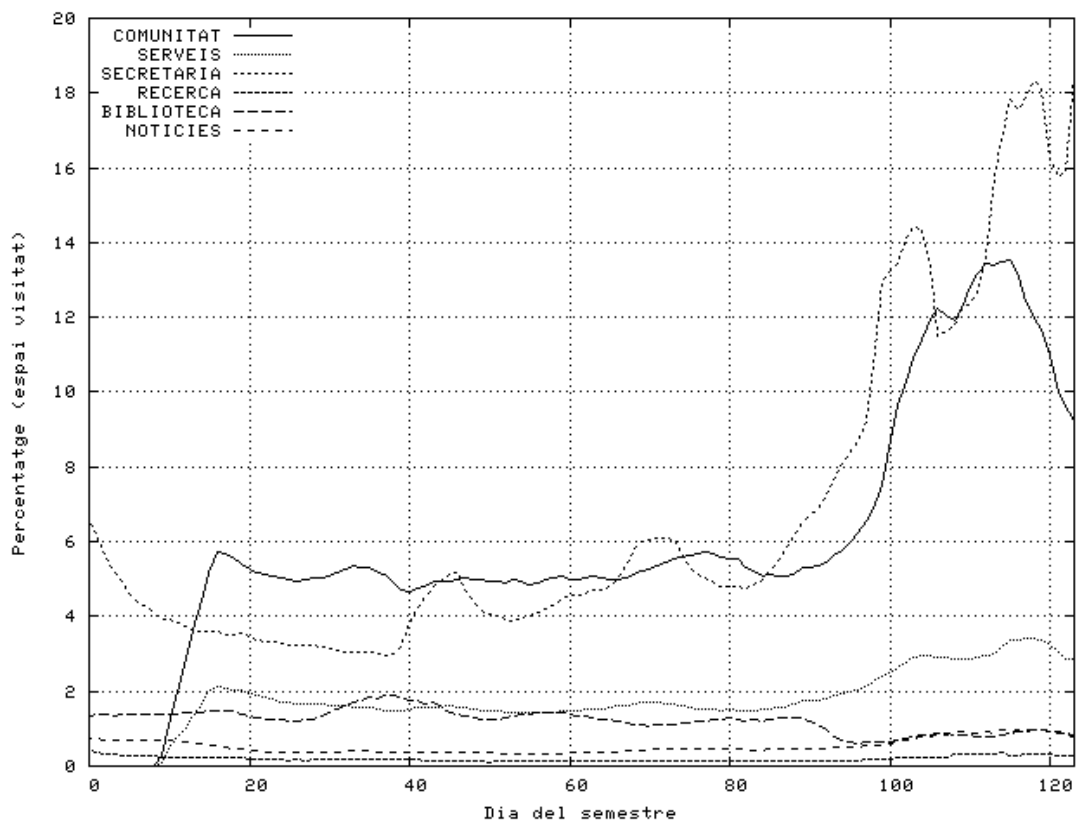


Figura 4.19: Distribució al llarg del curs de l'accés als diferents espais en la navegació pel campus virtual.

Les marques corresponents a “SERVEIS” i “COMUNITAT” no van estar actives els primers dies del semestre i per això apareixen a zero a la Figura 4.19. És interessant observar que tots els espais tenen un ús més o menys constant durant el curs i és cap a final de curs, quan la docència de les assignatures finalitza, que l'ús de la Biblioteca decreix i al mateix temps creix l'ús dels espais Secretaria, Comunitat i, en menor mesura, Serveis. Això és degut a diferents motius, per una banda el nombre de gestions a secretaria augmenta; per exemple, cal actualitzar el full personal d'exàmens, consultar les notes o consultar el pla d'estudis per preparar la matrícula del proper semestre. Per altra banda, hi ha menys obligacions relacionades amb el seguiment de les assignatures i els estudiants tenen més temps per accedir a comunitat, especialment als fòrums d'estudiants. A més, en aquests fòrums troben informació d'altres companys relativa a la preparació de les proves finals.

## Caracterització d'usuaris per la seva navegació

A partir dels camins de navegació disponibles en el conjunt de dades s'ha dissenyat un nou experiment que persegueix caracteritzar els usuaris segons la seva navegació pel campus virtual. A la UOC sempre s'ha pensat que existeixen tipologies d'usuaris en relació als comportaments de connexió al campus virtual; així es considera que hi ha usuaris que es connecten principalment entre setmana, mentre que n'hi ha d'altres que ho fan només als vespres o als caps de setmana. L'objectiu de l'anàlisi que es presenta en aquesta secció és descobrir si existeixen aquests o d'altres patrons, és a dir, si es pot caracteritzar als estudiants per la seva navegació.

El conjunt de dades que s'utilitza és el mateix dels experiments anteriors, que correspon al semestre 20062 i que conté les sessions de navegació al llarg de 131 dies del semestre dels 29.531 estudiants matriculats en els estudis en català. Totes aquestes sessions s'han processat per així calcular un seguit de variables per cada estudiant. Així, s'ha construït un vector de característiques per a cada estudiant que conté les variables següents:

- numsess: Promig de sessions de navegació per dia.
- numsess\_fest: Promig de sessions de navegació per dia festiu. S'han considerat festius els caps de setmana, els dies festius propis del semestre que s'estudia i els ponts.
- ratio\_fest: Ratio de sessions en dia festiu del nombre total de sessions.
- pct\_mati: Percentatge de sessions durant les hores del matí (07h-16h).
- pct\_tarda: Percentatge de sessions durant les hores de la tarda (16h-22h).
- pct\_nit: Percentatge de sessions durant les hores de la nit (22h-02h).
- durada: Promig en segons de la durada de totes les sessions.
- nclicks: Promig del nombre d'accions (marques) de totes les sessions.
- dies\_tard: Nombre de dies que transcorren des de l'inici de curs fins a la primera sessió de navegació de l'usuari.
- diesdiff: Promig de dies de diferència entre dues sessions de navegació consecutives.
- diesdiff\_max: Diferència màxima, en dies, de dues sessions de navegació consecutives.

- pct\_espais: Percentatge de sessions en les quals l'estudiant accedeix a les bústies de l'aula.
- pct\_aula: Percentatge de sessions en les quals l'estudiant accedeix a l'aula.
- pct\_biblio: Percentatge de sessions en les quals l'estudiant accedeix a la biblioteca.

Les tres variables que indiquen la navegació en diferents intervals horaris s'han establert tenint en compte les franges horàries observades a la Figura 4.13 i que es consideren més habituals. Així, els matins s'han establert fins a les 16 pensant en els horaris intensius i els horaris dels funcionaris, i les tardes fins a les 22 pensant en els estudiants amb fills que han d'anar a dormir d'hora. També s'han calculat dues variables més que finalment no s'han fet servir, aquestes variables són el percentatge de sessions de l'usuari que acaba amb el botó "surt" i el nombre total de dies actius en el campus, durant el semestre. La primera perquè no es un element propi de l'ús docent del campus virtual i la segona perquè conté informació que podria distorsionar la predicció realitzada.

Un cop calculades les variables per a cada estudiant, s'han desat totes en un únic arxiu que conté 28.044 línies. El nombre d'usuaris que s'analitzen s'ha vist reduït com a resultat d'eliminar aquelles línies amb resultats invàlids o irrellevants, que es corresponen a estudiants que no s'han connectat o bé a estudiants que tot i estar matriculats a l'inici de semestre s'han donat de baixa abans d'acabar el curs.

Amb aquest conjunt de dades s'ha portat a terme un anàlisi PCA Factorial utilitzant SPSS Clementine 11.1 [21], obtenint 6 components principals o factors amb valor propi més gran que 1. El resultat detallat es mostra a Taula 4.7.



Component	Autovalors inicials		
	Total	% de la variància	% acumulat
1	3.774	26.958	26.958
2	1.931	13.791	40.749
3	1.390	9.927	50.676
4	1.249	8.919	59.594
5	1.059	7.562	67.157
6	1.017	7.261	74.418

Taula 4.7: Resultat del PCA Factorial sobre el conjunt de dades.

Aquests 6 components permeten explicar el 74,42% de la variància. La Taula 4.8 mostra els components segons el pes atorgat a cadascuna de les variables d'entrada originals.

Variable	Factor-1	Factor-2	Factor-3	Factor-4	Factor-5	Factor-6
numsess	<b>0,1914</b>	-0,04647	0,1226	0,0656	-0,008909	-0,003851
numsess_fest	<b>0,229</b>	0,08109	0,178	0,0339	-0,02538	-0,00821
ratio_fest	0,3535	<b>2,994</b>	0,607	-1,746	-0,6364	0,194
pct_mati	0,01145	<b>-2,509</b>	0,2708	-0,7172	0,342	-0,04712
pct_tarda	-0,221	1,196	-1,449	<b>4,771</b>	0,07188	0,003593
pct_nit	-0,08035	<b>2,621</b>	0,8183	<b>-3,899</b>	-0,4506	-0,1029
durada	<b>0,0001129</b>	3,178E-06	0,00009877	0,00005892	0,00004836	-0,00004493
nclicks	<b>0,02239</b>	-0,001356	0,0161	0,01003	0,01151	0,001897
dies_tard	-0,01847	0,01258	0,06173	0,01417	-0,04054	-0,006558
diesdiff	<b>-0,03487</b>	0,007572	<b>0,1215</b>	0,03512	0,05651	0,003876
diesdiff_max	<b>-0,01631</b>	0,001806	<b>0,04063</b>	0,01184	0,02208	-0,0003266
pct_espais	0,05033	0,4279	-0,888	-0,1704	<b>1,922</b>	<b>-2,86</b>
pct_aula	0,1018	0,2866	-0,6038	-0,1119	1,519	<b>3,045</b>
pct_biblio	-0,08238	-0,2545	0,7836	2,167	<b>-13,4</b>	0,3703
Factor constant	-0,6234	-0,8999	-0,3672	-0,8445	-2,081	-0,3891

Taula 4.8: Pes de cada variable a cada component.

En negreta es mostren les variables rellevants per cada factor, usant un criteri de 0,5 per a la correlació entre el factor i les variables que el componen. Així, la component 1 té un criteri de correlació de 0,5 en valor absolut i les variables *numsess*, *numsess\_fest*, *durada* i *nclicks* amb valor positiu; i *diesdiff* i *diesdiff\_max* amb valor negatiu. Aquesta component és molt interessant donat que mesura principalment la freqüència i regularitat de les sessions. La component 2 tria les variables *ratio\_fest*, *pct\_nit* i *pct\_mati*. Així, descriu als estudiants que es connecten en festius o nits i, clarament, representa a aquells estudiants que no es connecten en horari laboral. La component 3 tria les variables *diesdiff* i *diesdiff\_max*, el que vol dir que mesura les sessions esporàdiques, estudiants que es connecten poc i sense regularitat. La component 4 selecciona *pct\_tarda* amb valor positiu i *pct\_nit*, és a dir, sessions de tarda i no de nit, sense ser rellevant el matí. Aquest fet potser indica que la selecció de les franges horàries per calcular les variables relatives a la connexió a la tarda i nit no és del tot òptima i hi ha solapaments. La component 5 tria *pct\_biblio* amb valor negatiu i *pct\_espais* amb valor positiu, és a dir, aquells estudiants que en la seva navegació habitual no utilitzen la biblioteca i sí utilitzen els accessos directes a les bústies de l'aula. La component 6 selecciona *pct\_espais* amb valor negatiu i *pct\_aula* amb valor positiu, és a dir, els estudiants que prefereixen accedir a l'aula virtual abans que a les seves bústies compartides. És interessant comprovar que hi ha una variable que no s'ha fet servir a cap component, *dies\_tard*, el que vol dir que el nombre de dies tard en que es connecta al campus un estudiant, respecte l'inici de curs, no condiciona el desenvolupament adequat del semestre acadèmic.

Aquests sis components o factors són combinacions lineals de les variables amb més força d'entre totes les del conjunt definit, i expliquen 6 comportament diferents dels estudiants del campus virtual, tenint en compte la seva navegació. El següent pas de l'anàlisi realitzada consisteix en aplicar un algorisme de *clustering* sobre els 6 components obtinguts. L'algorisme utilitzat és el bi-etàpic disponible al paquet informàtic SPSS Clementine. Aquest algorisme selecciona automàticament el nombre de clústers, i s'ha utilitzat sense estandarditzar els camps numèrics ja que provenen d'un PCA.

El resultat d'utilitzar l'algorisme bi-etàpic sobre els 6 components són 3 clústers. El clúster 1 conté 2.245 estudiants, el 8% del conjunt de dades, i són estudiants amb sessions de navegació esporàdiques, és a dir, poques connexions al campus virtual, separades en el temps, principalment en festius i nits tot i que també a les tardes. Curiosament són usuaris que accedeixen a la biblioteca, utilitzen els accessos directes a les bústies de l'aula des de la plana d'inici i accedeixen poc a l'aula. El clúster 2 inclou a 4.441 estudiants, el 15,8%, i són estudiants que es connecten molt sovint i regularment, amb una lleugera preferència per fer-ho a les tardes, no utilitzen la biblioteca i la informació d'accés als taulers i a les aules no és significativa. El clúster 3 inclou a la resta d'estudiants, 21.358, el 76,2% i presenta valors baixos per tots els components, el que indica que inclou una gran diversitat d'estudiants amb comportaments de navegació variats.

Es pot repetir el procés d'anàlisi per a avaluar si el clúster 3 proporciona informació rellevant sobre els estudiants que inclou. Si es repeteix tot el procés només pels elements del tercer clúster, el resultat del PCA és similar, excepte pel component 3 que desapareix i el 6 que canvia de signe. Per la seva banda, el resultat d'aplicar l'algorisme bi-etàpic sobre aquests 5 components genera 8 clústers, amb un nombre d'estudiants que varia entre 1850 i 3250 estudiants per clúster, el que mostra una gran diversitat de perfils, però equilibrats en nombre d'individus. Cap d'aquests clústers mostra un comportament navegacional clar.

#### **4.6.4 Conclusions dels experiments de nivell 1**

L'objectiu principal en el nivell de sessió és recollir i analitzar l'activitat dels usuaris quan es connecten al campus virtual. Amb els diferents experiments i anàlisis presentats s'ha vist clarament com s'aconsegueix aquest objectiu. Així, s'ha establert amb èxit un procediment de gestió, processat i emmagatzemament de fitxers de log. A partir d'aquí, s'han obtingut els camins de navegació, una informació nova i rellevant a la que no s'havia tingut accés abans. La navegació constitueix una font d'informació important per conèixer l'ús real del campus virtual, així com les necessitats principals dels usuaris, com així ho han mostrat els resultats obtinguts en els diferents experiments. Així, els

usuaris utilitzen principalment el campus virtual per assolir els seus objectius d'aprenentatge i el servei que més utilitzen és la bústia, en les seves diferents interfícies. Per altra banda, s'ha observat que els estudiants del campus virtual es poden caracteritzar segons la seva navegació en tres grups: estudiants amb patrons de navegació i connexió poc definits, estudiants que es connecten esporàdicament i un tercer grup d'estudiants que ho fan amb molta freqüència i regularitat. Tots aquests resultats constitueixen una informació que fins ara no havia estat possible obtenir i que té una repercussió i aplicació directa en la millora de l'experiència educativa dels estudiants.

#### **4.7 Experiments al nivell de curs o semestre acadèmic**

En aquest segon nivell es va planificar un experiment amb l'objectiu de provar la validesa de les assumpcions fetes sobre el descobriment del comportament de navegació dels usuaris en un entorn virtual d'aprenentatge. L'experiment planificat es situa a l'àmbit del segon nivell o nivell de semestre acadèmic, tot i que, degut a la interrelació entre els nivells que s'ha descrit al capítol 3, requereix que es tinguin en consideració dades corresponents al nivell de sessió. Les dades de nivell de sessió que s'inclouen tenen a veure amb la connexió a l'aula virtual o al període de temps entre dues sessions consecutives d'un mateix usuari, per exemple.

##### **4.7.1 Experiment: estudi del lliurament d'activitats en assignatures**

En aquest experiment s'estudia el comportament dels estudiants a l'hora de portar a terme tasques concretes relacionades amb el procés d'aprenentatge. Per aquest estudi s'han seleccionat dos assignatures dels Estudis d'Informàtica, Multimèdia i Telecomunicació de la UOC. Aquestes assignatures són "Fonaments de Programació" i "Compiladors I", amb 596 i 111 estudiants matriculats respectivament. "Fonaments de Programació" és una assignatura introductòria a la programació que la matriculen principalment els estudiants de primer curs de l'Enginyeria Tècnica en Informàtica i, per altra banda, "Compiladors I" és una assignatura més avançada que es cursada principalment per estudiants de segon cicle en Enginyeria Informàtica. L'estudi que es

realitza es focalitza en les accions i la navegació dels estudiants d'aquestes dues assignatures i no té en compte si el grup d'estudiants matriculats d'aquestes assignatures n'ha matriculat d'altres. No obstant, en una situació real de personalització del procés d'aprenentatge s'hauria de tenir en compte, en el model o perfil de cada estudiant, el seu comportament i navegació en totes les assignatures matriculades, tractant d'assolir una solució de compromís entre la precisió i la complexitat del model. L'objectiu d'aquest experiment és mostrar les diferències entre els patrons de navegació entre els dos col·lectius seleccionats i les possibles aplicacions en cada cas. Els resultats obtinguts han estat publicats a [44, 119].

Bàsicament, els estudiants es connecten l'entorn virtual d'aprenentatge i accedeixen a les aules virtuals per així seguir les activitats d'aprenentatge previstes per a cada assignatura. Aquestes activitats han estat dissenyades d'acord a una planificació pel calendari del curs. En el cas de les assignatures que s'analitzen en aquest experiment, es demana als estudiants que resolguin una activitat optativa que es publica durant la primera setmana del semestre, un cop el curs ja ha començat. El temps previst per a que els estudiants enviïn la solució de l'activitat és es diferent a cada assignatura, variant entre dues i tres setmanes. Els estudiants disposen d'espais ben diferenciats a l'aula virtual per accedir a la publicació de l'enunciat de l'activitat i per a fer el lliurament de la seva solució. D'aquesta manera, aquests espais són clarament identificables en les entrades dels fitxers de log i, per tant, es coneix el moment exacte en que els estudiants descarreguen l'enunciat o en que envien la seva solució. És interessant recordar que l'activitat que es proposa no es obligatòria pel seguiment i superació de l'assignatura, tot i que es recomana fortament la seva resolució tant per iniciar l'estudi de l'assignatura com pel seguiment de l'avaluació continuada.

Així, es pressuposa que tots els estudiants seguiran les activitats proposades, donat que aquells estudiants que no ho facin al final del curs es veuran en la necessitat de realitzar l'examen final que, normalment, ofereix un grau de dificultat superior. En l'anàlisi que es planteja es vol estudiar a aquells estudiants que decideixen no realitzar la primera activitat del curs, i també aquells que tot i realitzar-la obtenen una qualificació baixa, i observar si hi ha alguna relació amb la seva navegació per l'entorn virtual d'aprenentatge.

	A	B	C+	C-	D	N	Total
Fonaments de Programació	216	26	63	0	49	215	569
Compiladors I	22	65	12	1	0	11	111

Taula 4.9: Qualificacions obtingudes en la primera activitat.

La Taula 4.9 mostra les qualificacions obtingudes pels estudiants en la primera activitat. Cal comentar que al tractar-se de la primera activitat del curs el seu objectiu és introduir la matèria d'estudi, amb un grau mig de dificultat i, per tant, no s'espera que hi hagi molts estudiants amb notes baixes, és a dir, amb C- o D. Per altra banda, hi ha un nombre considerable d'estudiants que no la realitzen i per tant la seva qualificació és N. També cal destacar que, com era d'esperar, els valors són sensiblement diferents entre les dues assignatures. A "Fonaments de Programació", 216 dels 569 estudiants, el 37%, no realitzen l'activitat proposada i 49, el 8.6%, realitzen un lliurament que mereix la qualificació més baixa. Aquestes dades són les habituals per a aquesta assignatura i, per tant, qualsevol informació sobre el perfil dels estudiants que obtenen aquestes qualificacions serà extremadament útil. Per altra banda, a l'assignatura de "Compiladors I", només onze estudiants decideixen no realitzar l'activitat proposada i només un obté la qualificació més baixa. Les notes no són tan bones perquè el grau de dificultat d'aquesta prova és superior a la de l'altra assignatura.

### **Dades dels fitxers de log i creuament amb el perfil de navegació**

En l'estudi que es planteja, la principal font de dades bàsiques que s'utilitza és els fitxers de log del campus virtual. Aquets fitxers es preprocessen per així eliminar totes aquelles línies que no són produïdes directament per les accions dels usuaris com la càrrega de fulls d'estil, elements de format, icones, imatges, etc. Aquest preprocés redueix la quantitat de línies en gairebé un 90%. No obstant, durant una setmana típica en mig del semestre acadèmic, el nombre de línies que s'han de processar es troba al voltant dels 24 milions, aproximadament uns 12 GB i és, per tant, un volum

considerable. Per tant, és necessari un segon filtratge més orientat a focalitzar l'experiment, tal com es descriu a [57]

Els usuaris es poden identificar ja que hi ha un identificador de sessió únic que es genera cada cop que un usuari entra en el campus virtual utilitzant el seu nom d'usuari i la seva clau personal d'accés. Les adreces IP que apareixen a les línies de log no són adequades per a identificar els usuaris ja que hi ha la possibilitat que més d'un usuari accedeixi al campus virtual mitjançant el mateix servidor *proxy* i, per tant, la adreça IP real queda emmascarada. Gràcies al número de sessió assignat a cada usuari és possible identificar no només a cada usuari individual sinó també les diferents sessions, cosa que és especialment útil per estudiar el seu comportament de navegació. Quan un estudiant navega per àrees del campus virtual on no es requereix el número de sessió no es pot registrar i fer seguiment de la seva navegació; per tant, les línies dels fitxers de log que no identifiquen als usuaris s'eliminen ja que no aporten informació rellevant.

Per assegurar la privacitat de les dades dels estudiants, el número de sessió s'elimina de la URL de cada crida o línia del fitxer de log, i es substitueix per un nou identificador d'usuari i de sessió, de manera que no es pugui determinar l'usuari original. El nou identificador d'usuari i de sessió tenen un mapatge un a un amb el número de sessió original. Aquest procés afavoreix la privacitat alhora que permet mantenir la possibilitat de creuar la informació de la interacció dels usuaris amb dades socio-demogràfiques o acadèmiques provinents de les bases de dades administratives de la universitat. Tot i les mesures preses, les dades que s'utilitzen en aquest estudi poden continuar representant una preocupació a nivell de privacitat [1], tot i que totes les dades que es recullen i manipulen s'utilitzen únicament amb propòsits acadèmics, tal com pot fer un professor en una classe presencial al realitzar seguiment dels seus alumnes. Els estudiants són conscients que totes les seves interaccions en el campus virtual són registrades i, el que és més important, que les dades recollides només s'utilitzen en la pròpia universitat.

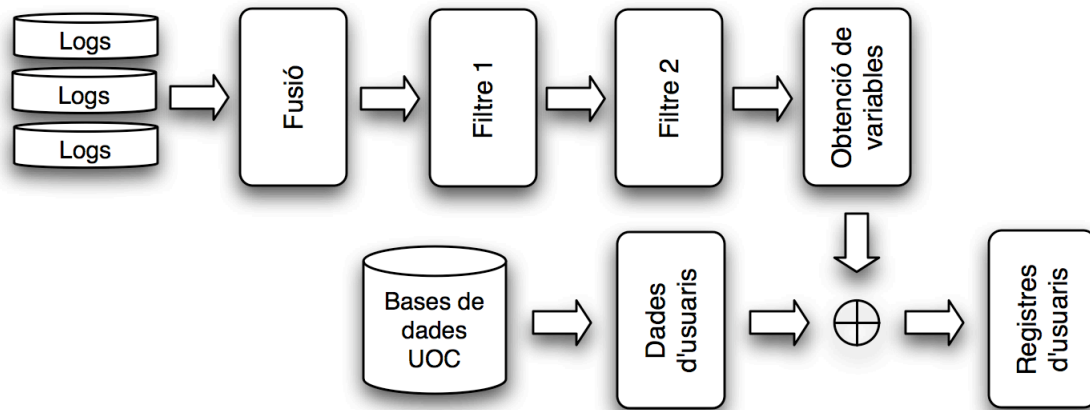


Figura 4.20: Etapes del procés d'anàlisi

El procés d'anàlisi que s'ha establert consta de diferents etapes, tal com es mostra a la Figura 4.20. Aquest procés només està automatitzat parcialment i cal adaptar-lo per cada experiment que es vol resoldre.

Depenent de l'anàlisi que es vulgui realitzar, aquest fitxer de log encara es pot reduir molt més tot seleccionant les línies que es corresponguin amb la interacció a estudiar. Per tal de portar a terme els experiments descrits, es disposa de dos conjunts de dades d'estudiants. El primer conjunt emmagatzema totes les interaccions del grup de 569 estudiants matriculats de l'assignatura "Fonaments de Programació", des del 23 de febrer de 2005 fins al 11 de març del mateix any. L'arxiu de log filtrat corresponent a aquests dies conté unes 1.300.000 línies. El segon conjunt de dades conté la mateixa informació que el primer grup, corresponent a 111 estudiants matriculats de "Compiladors I" entre la mateixa data inicial i el 17 de març de 2005. Per a aquest segon conjunt de dades el fitxer de log filtrat conté als voltats 2.200.000 línies. La intersecció entre els dos conjunts és buida donat que tot i ser dels mateixos estudis, es tracta de dues titulacions diferents.

Els principals motius per a escollir aquests conjunts de dades i els períodes de temps són els que es presenten a continuació. En primer lloc, els conjunts de dades seleccionats tenen la mida adequada per poder ser manegats i per poder experimentar amb diferents metodologies d'anàlisi. En segon lloc, el període de temps establert comença el 23 de



febrer, que correspon al primer dia del curs i transcorre fins que els estudiants han de lliurar la primera activitat prevista en el curs, el 11 o el 17 de març, depenent de l'assignatura. Aquesta primera activitat rep el nom de Prova d'Avaluació Continuada (PAC) i en el cas de les dues assignatures d'aquest estudi es tracta d'una prova optativa, de dificultat mitja. Tot i tractar-se d'una activitat optativa, la seva realització i la obtenció d'una qualificació alta es té en compte en el còmput de la nota final, mentre que si no es realitza o bé si s'obté una qualificació baixa, no es veu reflectit en la nota final. Com a darrera consideració sobre el conjunt de dades, cal esmentar que totes les dades provenen d'un entorn d'e-learning tancat, el campus virtual de la UOC, on totes les accions dels usuaris tenen relació amb els objectius d'aprenentatge. A més, hi ha la possibilitat d'accedir a altres dades com són, per exemple, el nombre d'assignatures matriculades en aquell curs, el nombre de préstecs sol·licitats a la biblioteca o el nombre de semestres que porta cada estudiant a la universitat.

### **Preprocessat de les dades i obtenció de variables**

Els experiments es porten a terme utilitzant un fitxer de log reduït que elimina tota la informació no rellevant dels fitxers estàndards generats pels servidors web Apache. Com s'ha comentat a la secció anterior, dels fitxers de log generats s'eliminen totes les línies que no recullen la interacció dels estudiants i que tenen a veure amb les dues assignatures seleccionades per l'estudi. El resultat són dos fitxers de aproximadament 1.300.000 i 2.200.000 línies, que representa un conjunt de dades raonable per a estudiar el comportament de navegació dels usuaris en un entorn virtual d'aprenentatge. La Figura 4.20 mostra les etapes seguides pel preprocessat de les dades, començant per les dues fonts de dades principals, els fitxers de log i les bases de dades de la universitat, fins a acabar generant el conjunt de dades final.

El principal avantatge de treballar en un entorn virtual d'aprenentatge és que es tracta d'un entorn tancat on hi ha altres fonts de dades que poden ser rellevants per l'anàlisi. Per exemple, a partir de la informació guardada a l'expedient acadèmic de cada estudiant es pot extreure la quantitat total d'assignatures que ha matriculat o el nombre de cursos que porta estudiant a la UOC. Informació d'un altre caire com és el nombre de

llibres demanats en préstec o els resultats de la enquesta de satisfacció, també es troben disponibles. Tal i com es mencionava a la secció anterior, la possibilitat d'accedir a nombroses fonts d'informació sobre els estudiants pot plantejar nombroses qüestions sobre la privacitat de les dades i de la informació personal i cal dir que en tot moment s'és conscient de la seva importància tal com apareix a [1] i que tota la informació que s'utilitza té una finalitat acadèmica. A més, els estudiants estan al corrent que les seves interaccions en el campus virtual són registrades i que tota la informació privada s'utilitza només en l'àmbit de la universitat.

Després del preprocessat inicial, s'apliquen dos filtres, un pel nivell de sessió i un pel nivell de semestre. El primer filtre que s'aplica al fitxer de log és el filtre de sessió. L'objectiu d'aquest filtre és agregar totes les peticions d'una sessió en un únic conjunt de variables. El primer problema que apareix al voler realitzar un anàlisi a nivell de sessió és identificar l'inici i el final d'una sessió en particular. S'han identificat tres tipus de sessió d'acord amb el seu patró d'inici-fi: normal, abandonada i difusa. Les sessions normals són aquelles on els usuaris inicien la seva interacció amb l'entorn virtual d'aprenentatge i se'ls hi assigna un nou número de sessió. Quan l'usuari decideix acabar la sessió ho fa utilitzant el botó de sortir del campus virtual. Les sessions abandonades són com les sessions normals però no acaben amb el botó de sortir, per lo que no se sap del cert quan acaben. Les sessions abandonades s'han de sumaritzar al final de l'anàlisi seqüencial del fitxer de log. Finalment, les sessions difuses són aquelles que porten a terme agents propis del campus virtual. Aquests agents realitzen comprovacions de la configuració, prestacions i rendiment de la plataforma del campus virtual. Degut al funcionament balancejador de càrrega, a la configuració dels *front-ends* del campus virtual i al comportament multiprocés del sistema, pot succeir que en ocasions no s'assigni un número de sessió nou a aquells usuaris que havien abandonat la sessió. D'aquesta manera, hi ha usuaris que mantenen el mateix número de sessió tot i que pròpiament estan realitzant una nova sessió de navegació. Per tant, en aquest context, la única forma de fer seguiment de l'inici i fi de la sessió és establir un interval d'inactivitat. Quan el temps entre dues entrades consecutives d'un mateix usuari és superior a 20 minuts es considera que ha començat una nova sessió, independentment del seu identificador o número de sessió.

El primer algorisme de filtratge recorre seqüencialment el fitxer de log registrant les variables primàries i derivades de la sessió: identificador d'usuari, identificador de sessió, data i hora d'inici de la sessió, la data i hora de fi de sessió, el comptador d'accions de l'usuari, la durada de la sessió en minuts, dia de la setmana a l'inici de la sessió i hora a l'inici de la sessió. A aquestes variables se'ls hi dóna el nom de variables genèriques. També es tenen en compte altres variables que només són rellevants per determinats experiments. L'anàlisi que aquí es presenta té en compte diverses d'aquestes variables addicionals: el nombre de missatges enviats al fòrum de l'assignatura i al fòrum del laboratori virtual durant la sessió, la data i hora en que s'ha descarregat l'enunciat de l'activitat i la data i hora del lliurament de l'activitat. En la següent secció es mostra perquè aquestes variables addicionals són rellevants pel context i l'estudi presentats. La identificació de les variables addicionals normalment implica escriure un o més algorismes per descobrir els esdeveniments que s'han de tenir en compte. En aquest experiment s'ha utilitzat el llenguatge de programació Perl i els recursos que proporciona per treballar amb expressions regulars i així extreure la informació necessària directament de les URL de les línies dels fitxers de log. El cost asimptòtic en temps de l'algorisme d'aquesta primera etapa és lineal en relació a la quantitat de línies de log i el cost asimptòtic de l'espai que utilitza depèn casi exclusivament del nombre de sessions abandonades, ja que aquestes necessiten ser processades al final del procés seqüencial.

A continuació del primer filtre més focalitzat en les dades relatives a la sessió es realitza el segon filtre. El segon filtre té per objectiu agrupar totes les tuples de sessions en una única tupla representant, per cada usuari, les dades relatives al curs. D'aquesta manera, s'utilitzaran les següents variables com a entrada pels algorismes de classificació i *clustering*:

- Sexe (GENDER). Malgrat que és habitual que en els ensenyaments tècnics com la enginyeria en informàtica hi hagi molts més homes que dones, és interessant incloure aquesta variable en l'estudi, i així confirmar la idea intuïtiva que el sexe no afecta el resultat.
- Edat (AGE). Aquesta variable podria proporcionar informació important a l'entorn del context socio-demogràfic dels estudiants. Per exemple, els estudiants de més edat tenen més obligacions familiars que els més joves.

- Nou (NEW). Variable que indica si un estudiant és nou a la UOC o no. Els estudiants nous poden experimentar dificultats amb l'ús del campus virtual i, per tant, és interessant validar aquesta hipòtesi.
- Repetidor (FIRST). Indica si un estudiant ha matriculat per primer cop una assignatura o no. Els estudiants que han matriculat l'assignatura en semestres anteriors són més proclius a comportar-se de manera diferent, donat que ja tenen informació prèvia provinent de la seva pròpia experiència.
- Crèdits (TOTALCREDS i ADAPTEDCREDS). El nombre total de crèdits que l'estudiant ha matriculat i el nombre de crèdits d'estudis anteriors que adapta. TOTALCREDS està directament relacionada amb la quantitat de temps que l'estudiant necessita dedicar a les activitats proposades. Així, aquells estudiants amb un gran nombre de crèdits matriculats és més probable que abandonin una o més assignatures o, el que és el mateix, que no lliurin les activitats proposades durant el curs. Per altra banda, els estudiants amb més crèdits adaptats són més propensos a continuar.

La informació que proporcionen aquestes variables es combinarà amb el comportament de navegació extret de l'anàlisi més bàsic dels seus patrons de navegació, durant el període de temps establert per la data d'inici de curs i el dia previst pel lliurament de la primera activitat del curs.

El conjunt d'informació relacionada amb el nivell de sessió és el següent: el nombre total de sessions en el campus virtual (TOTALSESS), la mitjana de temps entre dos sessions consecutives (MEANINTDUR), la mitjana de la durada de cada sessió (MEANDUR), i la mitjana de *hits* o d'accions de l'usuari en cada sessió (MEANHITS). Malgrat que no es coneix directament la intenció de cada estudiant quan porta a terme accions en el campus virtual, aquestes variables descriuen el patró de navegació bàsic en el període de temps que s'està analitzant. Amb l'objectiu d'estudiar els hàbits dels estudiants, el nombre total de sessions també es calcula per cada dia de la setmana, creant 7 variables noves  $WD_i$  ( $i=1$  per dilluns i  $i=7$  per diumenge). D'aquesta manera, un senzill índex que mesura si un estudiant es connecta preferentment en caps de setmana o no (WEEKENDPCT), es calcula així:

$$WEEKENDPCT = (WD_6 + WD_7) / \sum_{i=1}^{i=7} WDi.$$

El conjunt d'informació relacionada amb el nivell de semestre és el següent: el nombre de missatges enviats al fòrum de l'aula virtual (FORUM), el nombre de missatges enviats al fòrum del laboratori virtual (LAB), i el temps transcorregut entre el moment de publicació de l'enunciat de l'activitat i el moment que l'estudiant accedeix al seu contingut (DELAY).

Així, es consideren un total de 21 variables amb la finalitat de realitzar *clustering* i classificació. L'objectiu es predir, d'alguna manera, la qualificació que obtindrà un estudiant o, al menys, si superarà o no l'activitat proposada. A l'assignatura "Compiladors I" només un estudiant ha presentat una activitat amb una nota baixa, obtenint una "C-", i això és gairebé equivalent a estudiar si l'estudiant lliura l'activitat o no. Per altra banda, a l'assignatura de "Fonaments de Programació" hi ha un grup important d'estudiants que han obtingut la nota més baixa, una "D", mostrant, d'aquesta manera, un comportament molt diferent respecte els estudiants de l'altra assignatura. És interessant mencionar que la intenció d'aquest experiment és entendre el comportament de navegació dels usuaris i explicar, més enllà de la intuïció, aquells fets coneguts per la experiència del professor. En qualsevol cas, no forma part de l'objectiu la construcció d'un acurat sistema de predicció per a un escenari com el que s'ha descrit. D'aquesta manera, qualsevol esforç per incorporar personalització ha de tenir en compte no només el comportament dels estudiants sinó tot el procés d'aprenentatge.

És important comentar que hi ha dades de navegació per a tot el conjunt d'estudiants de l'assignatura "Compiladors I", mentre que la situació és molt diferent per "Fonaments de Programació" on 25 estudiants, el 4,4%, mai s'han connectat al campus virtual i, òbviament, no han realitzat l'activitat proposada. Per tant, els experiments descrits en la següent secció s'han realitzat amb les dades disponibles dels altres 544 estudiants. Durant l'etapa del preprocés de les dades, es va detectar que només 24 dels 25 estudiants que no navegaven van obtenir la nota "N" a l'activitat proposada, i un d'ells tenia una "A", mostrant que el professor va cometre una errada al moment d'introduir les qualificacions en el sistema, generant el que es coneix com un "outlier" mentider.

Això il·lustra la necessitat d'un nivell de control més elevat amb l'objectiu d'evitar errades humanes. Per exemple, el sistema podria posar automàticament una "N" a aquells estudiants que no han lliurat l'activitat i així, per una banda, estalviar aquesta feina al professor i, per l'altra banda, evitar que cometi errades a l'hora d'introduir les notes. El cas que aquí es presenta constitueix un bon exemple de com la usabilitat, la utilitat i la personalització es poden millorar gràcies a l'anàlisi de dades obtingudes de la interacció dels usuaris amb el campus virtual.

### **Aplicació de tècniques de mineria de dades**

Un cop s'han tabulat les dades descrites i s'ha construït un únic registre amb les dades recollides per cada estudiant, es poden aplicar diferents tècniques de mineria de dades. El *clustering* no supervisat amb l'algorisme TwoStep [178] i la classificació supervisada amb arbres de classificació i regressió [38] són, entre d'altres, les tècniques que resulten més útils gràcies a la explicabilitat dels seus resultats, malgrat el fet que les dues tècniques poden no arribar a la precisió òptima de classificació.

#### a) Rellevància de les variables

Els arbres de decisió presenten la característica que també es poden utilitzar per mesurar la importància de les variables, tal i com es suggereix a [38]. Tot i que a la literatura relacionada s'han proposat diversos mètodes [65], en aquest experiment s'utilitzarà un, desenvolupat pels mateixos autors [115, 118], que tracta de treure profit de la diversitat dels arbres de decisió quan aquests es construeixen a partir de conjunts similars, creats utilitzant l'aproximació *bagging* [37]. Bàsicament, aquest mètode construeix un gran nombre de d'arbres de decisió similars, un per cada grup d'entrenament possible i, llavors, la rellevància de les variables es calcula mesurant el nombre de vegades que es selecciona cada possible característica de classificació, tenint en compte la seva posició a l'arbre de classificació, donant més importància a aquelles variables que es troben més properes a l'arrel de l'arbre.

Per l'assignatura de "Fonaments de Programació", les variables més importants són DELAY, TOTALSESS i, sorprenentment, WD<sub>5</sub> (l'activitat s'ha de lliurar en divendres)

i WD<sub>2</sub>. La variable MEANDUR també mereix una atenció especial. A l'assignatura de "Compiladors I", les variables més importants són MEANINTUR, TOTALCREDS, TOTALSESS, MEANDUR i DELAY. En aquest cas, donat que la majoria d'estudiants realitzen l'activitat proposada, DELAY no esdevé tant rellevant com a l'altra assignatura. És sorprenent constatar que variables que reflecteixen la interacció, com són, FORUM i LAB, no s'han considerat importants pel propòsit de classificar. Per altra banda, GENDER i ADAPTEDCREDS són les variables menys rellevants, tal com era d'esperar. Tant WD<sub>6</sub> com WD<sub>7</sub>, també es consideren no rellevants, la qual cosa també és sorprenent donat que intuïtivament es pressuposa que els estudiants tenen menys temps per estudiar durant la setmana ja que la majoria d'ells treballa. Així, la hipòtesi sobre el comportament i interacció dels usuaris durant el cap de setmana ha de ser revisada.

Aquest experiment senzill mostra que les mateixes variables de classificació no són rellevants pels diferents conjunts de dades, tot i que algunes variables com són TOTALSESS, MEANDUR i DELAY semblen més robustes per descriure el comportament dels usuaris, inclús per diferents contextos d'aprenentatge. DELAY és la variable que proporciona informació més rellevant i que es pot utilitzar per personalitzar a nivell de curs. Per exemple, si un estudiant espera un cert temps en descarregar l'enunciat de l'activitat proposada, probablement no la realitzarà o no obtindrà una bona qualificació. Així, es podria dissenyar un sistema automatitzat que adverteixi a l'estudiant, o al seu professor, conforme s'aproxima la fita establerta com a límit raonable per descarregar l'enunciat de l'activitat. Aquesta fita no ha de ser el termini màxim per realitzar l'activitat sinó que es pot establir combinant els resultats dels anàlisis amb l'expertesa del professor. Un avís en pantalla o un correu electrònic personalitzat podrien ser d'utilitat per advertir a l'estudiant sobre l'activitat proposada i animant-lo a la seva descàrrega i realització.

#### b) *Clustering* no supervisat

El segon experiment realitzat tracta d'establir agrupacions d'estudiants en base al seu comportament de navegació, sense tenir en compte les dades socio-demogràfiques dels estudiants, ni les qualificacions que obtenen. L'algorisme *TwoStep* s'utilitza per descobrir patrons en el conjunt de dades. Els registres de dades s'agrupen de manera

que els registres en un grup o clúster tendeixen a ser similars els uns amb els altres, mentre que els registres en diferents clústers tenen diferències. El nombre de clústers s'estableix automàticament. Aquest estudi tracta d'identificar quines variables són rellevants per classificar però amb una aproximació diferent. Posteriorment es pot idear un anàlisi basat en una classificació supervisada i, d'aquesta manera, dissenyar un sistema tutor adaptatiu o de recomanació que combini les dues aproximacions.

Després de l'aplicació de l'algorisme, pels estudiants de l'assignatura de “Compiladors I” es generen dos clústers amb 27 i 84 registres respectivament. Les variables més rellevants pel *clustering* són TOTALSESS, MEANINTDUR (que, òbviament, estan correlacionades), FORUM i LAB, que són significants amb  $p < 0,001$ , i DELAY que és significant amb  $p < 0,05$ . MEANHITS, MEANDUR (que estan fortament correlacionades) i WEEKENDPCT no són significants. Els dos clústers obtinguts classifiquen molt bé les interaccions dels estudiants: els estudiants que es connecten amb més regularitat al campus virtual no envien missatges a les bústies de les aules i la variable DELAY també adquireix valors més elevats per a aquests estudiants que per la resta. Per altra banda, els estudiants de “Fonaments de Programació” s'agrupen en tres clústers, amb 91, 318 i 135 registres respectivament. En aquest cas, el comportament dels usuaris és molt diferent i totes les variables són significants amb  $p < 0,001$ , amb excepció de WEEKENDPCT, que no és rellevant. Un cop més, aquells estudiants amb un patró d'interacció més elevat obtenen millors resultats que la resta (el clúster 1 a la Taula 4.11), mentre que els altres dos clústers mostren diferents valors per la variable DELAY, per exemple.

	A	B	C+	C-	D	N
Clúster 1	11	13	2	0	0	0
Clúster 2	11	52	10	1	0	11

Taula 4.10: Distribució de les qualificacions d'acord als clústers obtinguts a “Compiladors I”.



	A	B	C+	C-	D	N
Clúster 1	69	4	9	0	4	5
Clúster 2	127	16	44	0	38	93
Clúster 3	19	6	10	0	7	93

Taula 4.11: Distribució de les qualificacions d'acord als clústers obtinguts a “Fonaments de programació”.

La Taula 4.10 i la Taula 4.11 mostren les qualificacions obtingudes pels estudiants als clústers obtinguts a cadascuna de les assignatures estudiades. Resulta interessant destacar que a “Compiladors I” aquells estudiants que realitzen l’activitat i no obtenen una nota alta sinó un aprovat amb una “C+” tenen un comportament de navegació més similar a aquells estudiants que decideixen no realitzar l’activitat i, per tant, allunyat dels que realitzen satisfactòriament l’activitat. De fet, el clúster 1 identifica un subconjunt dels bons estudiants, que tenen una “A” o una “B”. En canvi, a “Fonaments de Programació”, aquesta separació no apareix tan clara i observar patrons d’aquest tipus probablement requeriria una altra anàlisi més complexa. Usant els mateixos conjunts de dades però amb aproximacions diferents, s’han obtingut uns altres resultats que també són d’interès pel que fa a la detecció d’estudiants amb patrons de navegació incoherents (*outliers*) el que pot indicar algun tipus de problema o estratègia d’aprenentatge errònia [50, 51].

### ***Data fusion***

En l’actualitat, la UOC, mitjançant el Projecte Campus en Programari Lliure [19], està introduint l’ús de l’estàndard SCORM [26] tant amb la finalitat de desplegar cursos com per realitzar seguiment i anàlisi dels estudiants, tal i com es descriu a [121]. No obstant, altres plataformes educatives com per exemple WebCT ja incorporen la capacitat de fer seguiment dels usuaris [110, 113], ampliant, d’aquesta manera, la capacitat d’obtenir dades dels estudiants. A més, quan els estudiants utilitzen altres serveis del campus virtual, com per exemple la biblioteca digital, també deixen una traça que es pot utilitzar amb finalitats de personalització [70]. Per tant, es fa necessari disposar d’un procés

d'estandardització de les fonts de dades per així garantir que totes les dades es poden descriure adequadament [33]. Aquest procés, que es podria anomenar *data fusion*, seguint la nomenclatura utilitzada en altres àmbits, podria ser molt interessant per superar totes les dificultats que ofereixen els fitxers de log clàssics: identificació d'usuaris, mida gran, manca d'informació dels objectius dels usuaris, etc.

La combinació de diferents estratègies de navegació per perseguir el mateix objectiu com per exemple l'aprenentatge en línia utilitzant un entorn virtual d'aprenentatge combinat amb activitats presencials, canviarà la manera en que els estudiants interaccionen amb el campus virtual i, conseqüentment, el context d'aprenentatge. A més, amb els cursos compatibles amb SCORM es generaran suficients dades tant d'ús com acadèmiques que degudament analitzades serviran per adaptar i personalitzar el procés d'aprenentatge. Aquests nous cursos personalitzats poden derivar a nous paradigmes d'interacció i de contextos d'aprenentatge i, per tant, les dades d'ús generades i emmagatzemades poden ser diferents i, conseqüentment, es necessitaran nous mètodes de processat i anàlisi. D'aquesta manera, serà necessari un sistema adaptatiu capaç de processar, analitzar i personalitzar el procés d'aprenentatge en funció de l'evolució del comportament dels estudiants.

#### **4.7.2 Conclusions dels experiments de nivell 2**

En aquesta secció s'han descrit els experiments portats a terme amb dades de l'entorn virtual d'aprenentatge de la UOC, amb la finalitat d'estudiar les relacions entre els patrons de navegació dels usuaris i els resultats acadèmics obtinguts pels estudiants matriculats en dues assignatures de les enginyeries informàtiques. Els experiments realitzats mostren les possibilitats que sorgeixen a partir de la recollida estructurada de dades d'ús en entorns virtuals i gràcies a la utilització d'eines de mineria de dades, amb la finalitat d'introduir elements de personalització en el procés d'aprenentatge. Els resultats presentats constitueixen un bon punt de partida per futurs estudis i anàlisis de dades i és important comentar que algunes de les idees intuïtives que tenen tant professors com dissenyadors instruccionals es poden validar o refutar amb una recollida de dades ordenada i amb un algorisme de *clustering*. Òbviament, entendre la

complexitat de les accions portades a terme pels estudiants en la seva totalitat requereix una anàlisi amb més profunditat. Els resultats dels experiments mostren, però, que una anàlisi senzilla pot ser útil per determinar quines variables són rellevants per classificar i agrupar (per exemple totes les variables relacionades amb els patrons d'interacció) i quines variables poden no ser-ho, per exemple el sexe i la edat.

#### **4.8 Experiments al nivell de *lifelong learning***

En aquesta secció es descriuen els experiments portats a terme en el nivell de llarg termini o nivell d'aprenentatge al llarg de la vida. L'estudi principal en aquest nivell s'ha portat a terme al voltant de la titulació del segon cicle de la Enginyeria en Informàtica i dels seus estudiants. A continuació es descriuen el context de l'estudi i l'anàlisi portat a terme. Els resultats obtinguts en un primer estudi han estat publicats a [117], tot i que els resultats presentats en aquesta tesi han estat actualitzats amb dades provinents de dos semestres acadèmics més.

Els resultats que s'obtenen de l'anàlisi realitzat en aquest nivell es poden utilitzar per millorar la experiència d'aprenentatge dels estudiants, concretament, del seu recorregut d'aprenentatge al llarg de la vida, el disseny de les competències que han d'adquirir, millorant les titulacions progressives que ofereix la universitat i descobrint si els estudiants segueixen les recomanacions dels seus tutors a l'hora de matricular les assignatures. La informació que s'obté contribueix a la construcció del perfil d'usuari, juntament amb altra informació i característiques dels estudiants.

La UOC constitueix un bon escenari d'estudi per a aquest tercer nivell d'aprenentatge al llarg de la vida, gràcies les múltiples possibilitats de la seva oferta formativa. A més dels escenaris formatius al llarg de la vida que habitualment proporcionen les universitats, les titulacions homologades i els programes de màster i postgrau, la UOC ofereix un seguit de cursos en un escenari d'aprenentatge al llarg de la vida que s'anomena Ateneu Universitari. La particularitat d'aquests cursos és que els pot realitzar qualsevol persona que ho desitgi, sense uns requisits d'accés determinats. D'aquesta

manera, una gran varietat d'estudiants amb diferents coneixements previs es troben en una mateixa aula virtual, el que assegura una gran diversitat de perfils diferents.

#### **4.8.1 Experiment: segon cicle d'Enginyeria en Informàtica**

El segon cicle de l'Enginyeria Informàtica és l'escenari que s'ha decidit analitzar en aquest nivell, donat que es té accés a la totalitat de les seves dades i es coneix en profunditat el seu desenvolupament i estructura. En aquest segon cicle s'ofereixen unes titulacions progressives que volen motivar als estudiants a especialitzar-se en una o més àrees de coneixement determinades, al mateix temps que ofereix algunes assignatures mitjançant l'Ateneu Universitari, orientades a qualsevol persona que les vulgui cursar. El segon cicle de l'Enginyeria en Informàtica constitueix un escenari interessant d'aprenentatge al llarg de la vida ja que els estudiants que hi accedeixen ho fan amb una titulació prèvia, l'Enginyeria Tècnica en Informàtica (de Gestió o de Sistemes), i amb l'objectiu de millorar les seves competències i coneixements en l'àmbit de la informàtica. Per tant, resulta interessant estudiar si els estudiants poden desenvolupar les seves competències i coneixements mitjançant les titulacions progressives dissenyades pel seu escenari d'aprenentatge al llarg de la vida.

L'objectiu del segon cicle de l'Enginyeria en Informàtica és formar professionals i preparar-los per a adquirir un conjunt de competències cobrint diferents àrees del mercat de treball. D'acord amb la informació de la que disposa la universitat, els estudiants d'aquesta titulació tenen una edat mitja de 35 anys i la seva motivació per matricular-se és millorar de les seves competències i habilitats, així com fer un pas endavant en la seva carrera professional, i es correspon amb el perfil "executiu" presentat a l'annex C.

El segon cicle de l'Enginyeria en Informàtica ha estat dissenyat de manera que un estudiant amb dedicació completa el pugui finalitzar en una durada de quatre semestres acadèmics, tot i que la flexibilitat que ofereix la universitat a l'hora de matricular les assignatures permet a cada persona ajustar-ho segons les seves necessitats personals i la seva disponibilitat. D'aquesta manera, cada estudiant pot decidir cada semestre quines assignatures vol cursar ja que, amb l'excepció d'algunes assignatures optatives amb

pocs estudiants, totes les assignatures s'ofereixen cada semestre, i només el projecte final de carrera té prerrequisits.

La titulació ofereix un conjunt de titulacions progressives que tenen per objectiu oferir àrees d'especialització concretes formades per diferents assignatures d'un mateix àmbit temàtic o bloc d'interès. Un cop s'han cursat i superat les assignatures d'una titulació progressiva es pot sol·licitar una acreditació, que és un certificat propi de la universitat. Les titulacions progressives volen motivar l'aprenentatge dels estudiants, així com el desenvolupament d'un conjunt de competències i l'adquisició de coneixements concrets. S'han dissenyat de manera que els estudiants puguin organitzar el seu procés d'aprenentatge d'acord als seus interessos professionals adequant-se a la demanda del mercat laboral.

Actualment, al segon cicle de l'Enginyeria en Informàtica, s'ofereixen set itineraris d'especialització que constitueixen titulacions progressives:

- Construcció d'aplicacions i sistemes distribuïts (34,5 cdt)
- Gestió avançada de les dades i del coneixement (30 cdt)
- Direcció de projectes i de sistemes d'informació (36 cdt)
- Enginyeria del programari (30 cdt)
- Disseny, gestió i seguretat de xarxes (33 cdt)
- Gràfics per computador (18 cdt)
- Interacció persona-ordinador (18 cdt)

Cada titulació progressiva es pot obtenir un cop superades les assignatures que constitueixen l'itinerari d'especialització. La universitat proporciona als estudiants una guia amb una recomanació de les assignatures a matricular en cada semestre. Per exemple, en el primer i segon semestre acadèmic es recomana seguir les assignatures tal com es presenten a la Taula 4.12. Les assignatures que apareixen en negreta són les troncales, és a dir, aquelles que els estudiants estan obligats a cursar i superar per a obtenir el títol, objectiu a llarg termini.

Primer Semestre	Segon Semestre
<ul style="list-style-type: none"> <li>• <b>Enginyeria del programari orientat a l'objecte</b> (11.058)</li> </ul> Informàtica gràfica I (05.020) Bases de dades II (05.020) <ul style="list-style-type: none"> <li>• <b>Disseny de xarxes de computadors</b> (11.061)</li> <li>• <b>Intel·ligència artificial I</b> (11.006)</li> </ul> Finances per a informàtics (11.053)	<ul style="list-style-type: none"> <li>• <b>Enginyeria del programari de components i sistemes distribuïts</b> (11.059)</li> </ul> Interacció humana amb els ordinadors (05.045) <ul style="list-style-type: none"> <li>• <b>Arquitectura de computadors</b> (11.067)</li> </ul> Sistemes de gestió de bases de dades (11.031) <ul style="list-style-type: none"> <li>• <b>Comunicacions sense fils</b> (11.062)</li> </ul> Arquitectura de sistemes distribuïts (11.068) <ul style="list-style-type: none"> <li>• <b>Intel·ligència artificial II</b> (11.007)</li> </ul>

Taula 4.12: Assignatures recomanades amb les troncal en negreta per als dos primers semestres de l'Enginyeria en Informàtica.

Generalment, els estudiants es matriculen cada semestre de dues o tres assignatures que escullen entre aquestes. En l'estudi que aquí es planteja resulta d'especial interès saber si els estudiants seleccionen les assignatures a matricular d'acord a les recomanacions que reben. Els estudiants poden escollir les assignatures que consideren més interessants per les seves necessitats, en lloc de seguir les recomanacions de la universitat, fent que els itineraris d'especialització siguin una oferta més interessant que l'itinerari més clàssic previst a la titulació. En qualsevol cas, es vol estudiar el comportament dels estudiants durant el seu procés d'aprenentatge a l'entorn virtual i les assignatures escollides, per així millorar el disseny dels itineraris d'especialització de les titulacions progressives.

### **Anàlisi de la matrícula per semestre**

Amb l'objectiu d'analitzar el comportament dels estudiants al nivell de llarg termini o d'aprenentatge al llarg de la vida, s'ha portat a terme un experiment que ha consistit en

un conjunt d'anàlisis dels patrons reals de selecció i matrícula de l'oferta d'assignatures per part dels estudiants.

Del segon cicle de l'Enginyeria en Informàtica, es disposa de les dades de 12 semestres acadèmics, des del seu inici el semestre de tardor del curs 2001-2002 (anomenat 20011) fins al semestre de primavera del curs 2006-2007 (anomenat 20062), ambdós inclosos. D'entre les dades disponibles, s'ha decidit seleccionar aquelles cohorts d'estudiants amb una antiguitat a la universitat de 8 semestres. Aquests 8 semestres s'han determinat per dos motius principals. Per una banda, la durada teòrica dels estudis és de 4 semestres si es matriculen 4 o més assignatures a cada semestre. Les dades mostren que el promig d'assignatures matriculades a cada semestre és lleugerament superior a 2, la meitat del previst al pla d'estudis. Per tant, la majoria d'estudiants tardaran 8 semestres, com a mínim, en finalitzar els estudis, el doble del temps previst. Per l'altra banda, s'han analitzat les dades dels graduats disponibles en els 12 semestres de dades del segon cicle de la Enginyeria en Informàtica. El resultat que s'ha obtingut és que el nombre de semestres actius dels graduats és de 8,3. D'aquesta manera, pels experiments que es volen realitzar s'ha decidit establir en 8 el nombre de semestres d'antiguitat dels estudiants.

El nombre total d'estudiants del conjunt de dades és 1.541, que són els estudiants que han accedit a la titulació entre el semestre 20011 i el 20062. D'aquests estudiants 64 ja s'han graduat. De les cohorts disponibles en el conjunt de dades, només 5 compleixen la condició de portar, al menys, 8 semestres a la UOC, concretament del semestre 20011 al 20031. En aquestes 5 cohorts van accedir a la titulació 687 estudiants i s'han graduat 36.

En els experiments que es presenten a continuació, el que es té en compte és cada matrícula formalitzada per cada estudiant. Així, si durant el temps que un estudiant porta a la universitat ha formalitzat 6 matrícules, se'n comptabilitzen 6. No obstant, que cada cohort tingui una antiguitat de 8 semestres no significa que cada estudiant d'aquella cohort hagi formalitzat 8 matrícules, perquè pot haver descansat un o més semestres.

El primer estudi que es realitza consisteix en analitzar les assignatures matriculades a cada semestre, tenint en compte les troncal<sup>5</sup> recomanades i una selecció d'optatives. Aquestes últimes són un total de 8 ja que és el nombre d'assignatures optatives o de lliure elecció que un estudiant ha de matricular per a obtenir el títol. El resultat d'aquesta anàlisi es mostra a la Taula 4.13, que presenta el nombre d'estudiants matriculats cada semestre a cadascuna de les assignatures ofertes, d'acord amb el semestre acadèmic en el que es suposa que s'ha de realitzar cada assignatura, utilitzant l'acumulat de tots els semestres acadèmics de les cinc cohorts seleccionades. Entre parèntesis s'indiquen els estudiants repetidors.

sem assig	1	2	3	4	5	6	7	8
<b>11.058</b>	<b>424</b>	<b>92(11)</b>	47(6)	18(2)	12(1)	4(2)	6(3)	4(0)
<b>11.006</b>	<b>353</b>	<b>101(40)</b>	49(16)	34(8)	13(4)	20(9)	10(2)	6(4)
<b>11.061</b>	<b>133</b>	111(29)	<b>156(31)</b>	58(23)	65(20)	39(18)	21(8)	17(13)
<b>11.059</b>	3	<b>112(1)</b>	<b>195(26)</b>	102(47)	67(39)	54(29)	34(22)	24(15)
<b>11.007</b>	4	63(2)	63(11)	<b>117(18)</b>	<b>97(21)</b>	58(23)	51(17)	32(16)
<b>11.062</b>	13	77(4)	<b>100(21)</b>	<b>102(11)</b>	52(15)	56(10)	37(3)	23(9)
<b>11.067</b>	69	<b>148(5)</b>	<b>76(14)</b>	42(12)	29(11)	26(8)	26(8)	13(8)
<b>11.060</b>	0	9(0)	40(0)	<b>108(4)</b>	<b>69(6)</b>	49(5)	32(4)	28(7)
<b>11.008</b>	48	72(11)	<b>120(16)</b>	<b>81(27)</b>	69(23)	54(23)	41(17)	20(10)
<b>11.012</b>	2	5(0)	20(1)	51(3)	<b>80(8)</b>	<b>69(5)</b>	64(9)	45(12)
<b>11.068</b>	2	29(0)	<b>100(2)</b>	<b>99(7)</b>	62(9)	45(11)	36(7)	25(6)
<b>11.009</b>	0	16(0)	26(1)	51(5)	<b>95(5)</b>	<b>69(13)</b>	64(15)	59(22)
<b>11.013</b>	0	0(0)	3(0)	9(0)	21(1)	48(4)	<b>65(6)</b>	<b>53(8)</b>
5.053	<b>171</b>	<b>63(23)</b>	29(8)	17(7)	11(1)	16(8)	8(1)	4(2)
11.034	<b>48</b>	<b>139(5)</b>	37(3)	14(1)	17(1)	25(3)	11(0)	7(1)
11.031	<b>103</b>	<b>86(14)</b>	30(14)	20(9)	16(2)	17(3)	16(3)	8(2)
5.045	<b>36</b>	<b>49(3)</b>	22(1)	30(1)	37(3)	21(1)	16(2)	13(1)
5.069	6	16(0)	27(4)	<b>37(7)</b>	<b>46(5)</b>	31(5)	22(4)	12(4)
5.070	0	2(0)	16(1)	26(0)	<b>27(0)</b>	<b>42(3)</b>	25(1)	20(2)
5.024	22	<b>32(2)</b>	<b>28(4)</b>	29(4)	15(3)	20(4)	5(3)	6(1)
11.029	5	15(1)	7(1)	<b>26(0)</b>	<b>41(0)</b>	20(0)	16(0)	15(0)

Taula 4.13: Nombre d'estudiants matriculats a cada assignatura, d'acord a cada semestre acadèmic.

En el primer semestre, les recomanacions de matrícula aconsellen que els estudiants es matriculin de tres assignatures troncal<sup>5</sup>, les quals es mostren a la Taula 4.12 i d'algunes assignatures opcionals. L'anàlisi de les dades revela que la majoria dels estudiants

<sup>5</sup> Al segon cicle de l'Enginyeria en Informàtica no hi ha assignatures obligatòries pròpies de la universitat, per tant els estudiants només han de cursar de manera obligatòria les assignatures troncal<sup>5</sup>.



només matriculen una o dues assignatures en el seu primer semestre, concretament matriculen un promig de 2,23 assignatures i, d'aquestes, en superen un promig de 1,42. Es interessant comentar que no s'estudia el nombre de crèdits matriculats donat que la poca diversitat de creditatge de les assignatures del segon cicle de l'Enginyeria en Informàtica, fa que l'estudi per crèdits sigui equivalent al de assignatures matriculades.

D'entre el gran nombre d'assignatures disponibles, les troncal més un considerable nombre d'assignatures optatives disponibles en el pla d'estudis, la majoria d'estudiants trien un subconjunt limitat, les troncal recomanades i unes poques assignatures optatives de diferent naturalesa, per exemple, l'assignatura optativa "Interacció Humana amb els Ordinadors" (05.045) i l'assignatura optativa "Sistemes de Gestió de Bases de Dades" (11.031). Aquest comportament revela que alguns estudiants ja han adquirit les competències bàsiques en la titulació prèvia i que poden avançar més ràpidament en l'àrea d'intensificació en "Gestió avançada de les dades i del coneixement", per exemple. Dues assignatures de les recomanades el primer semestre són les més populars a l'hora de ser triades i matriculades: "Enginyeria del Programari Orientat a l'Objecte" (11.058) i "Intel·ligència Artificial I" (11.006), mentre que altres assignatures que els estudiants saben que són difícils, com ara "Disseny de Xarxes de Computadors" (11.061), tot i estar recomanada, no són triades en el primer semestre.

Un altre resultat interessant és que els estudiants no segueixen amb detall les recomanacions de matrícula a l'hora de triar les assignatures que cursen. Per exemple, "Intel·ligència Artificial II" (11.007) se suposa que s'ha de realitzar després de "Intel·ligència Artificial I" (11.006), però només 63 de 353 possibles estudiants la cursen durant el segon semestre, mentre que "Compiladors I" (11.008) ha estat matriculada per 120 estudiants en el tercer semestre però només 51 han matriculat "Compiladors II" (11.009) en el quart semestre. El motiu principal és que els estudiants prefereixen acabar un semestre complet, és a dir, les recomanacions per a cada semestre, abans que seguir els itineraris d'especialització proposats per la universitat. Una altra raó important és que no tots els estudiants superen les assignatures quan les matriculen i les han de tornar a cursar abans de fer la següent. Així, la Taula 4.13 mostra de manera clara la navegació dels estudiants entre les assignatures del pla d'estudis i permet

observar l'adequació de les recomanacions de matrícula amb la matrícula real dels estudiants a cada semestre acadèmic.

Amb el conjunt de dades disponibles del segon cicle d'Enginyeria en Informàtica s'han realitzat altres anàlisis en relació al comportament de matrícula dels estudiants, per a estudiar la distribució del nombre d'assignatures matriculades i aprovades cada semestre, la correlació entre matrícula i aprovar-ho tot i les combinacions d'assignatures a la matrícula que afavoreixen l'èxit o el fracàs acadèmic.

El resultat de l'estudi del nombre d'assignatures matriculades per semestre es presenta a la Figura 4.21. A l'anàlisi que s'ha portat a terme no s'han inclòs les assignatures adaptades, ja que es vol estudiar el patró que presenten les assignatures que es matriculen i es segueixen, és a dir, que impliquen una dedicació als estudiants. En relació a l'adaptació d'assignatures també és interessant comentar hi ha hagut 84 matrícules d'estudiants que han consistit en només matricular assignatures per adaptar-les.

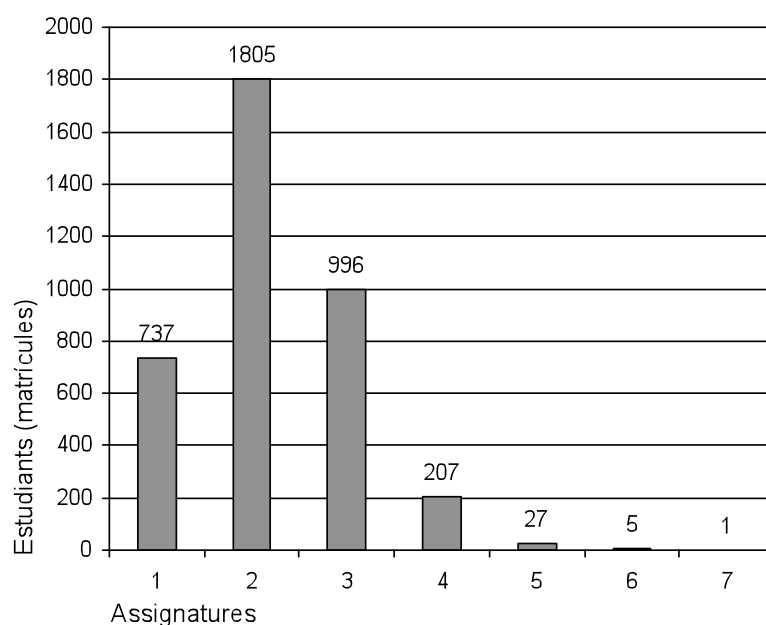


Figura 4.21: Nombre d'assignatures matriculades per semestre.

Com és pot veure a la Figura 4.21, la majoria d'estudiants matriculen dues assignatures cada semestre i, com a curiositat, només un estudiant en tot el conjunt de dades ha matriculat 7 assignatures un cop. La mitjana d'assignatures matriculades per estudiant cada semestre és de 2,16 assignatures.

El nombre d'assignatures aprovades a cada semestre és un altre indicador interessant. Els resultats es presenten a la Figura 4.22. Si es compara amb el nombre d'assignatures matriculades, la gràfica es desplaça sensiblement a l'esquerra. Així, la mitjana d'assignatures aprovades és de 1,48, sensiblement inferior al nombre d'assignatures matriculades.

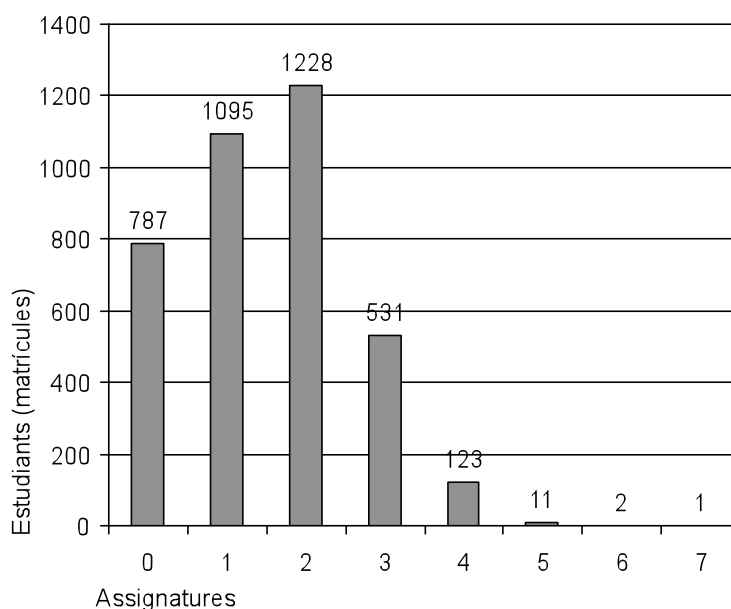


Figura 4.22: Nombre d'assignatures aprovades per semestre.

El estudi de la relació entre la matrícula i el nombre d'assignatures aprovades s'ha dividit en dues parts. Per una banda s'ha analitzat el nombre de matrícules que han aprovat totes les assignatures matriculades. Els resultats obtinguts es mostren a la Figura 4.23. Com es pot veure, el patró més freqüent entre els estudiants que ho aproven tot és matricular i aprovar dues assignatures.

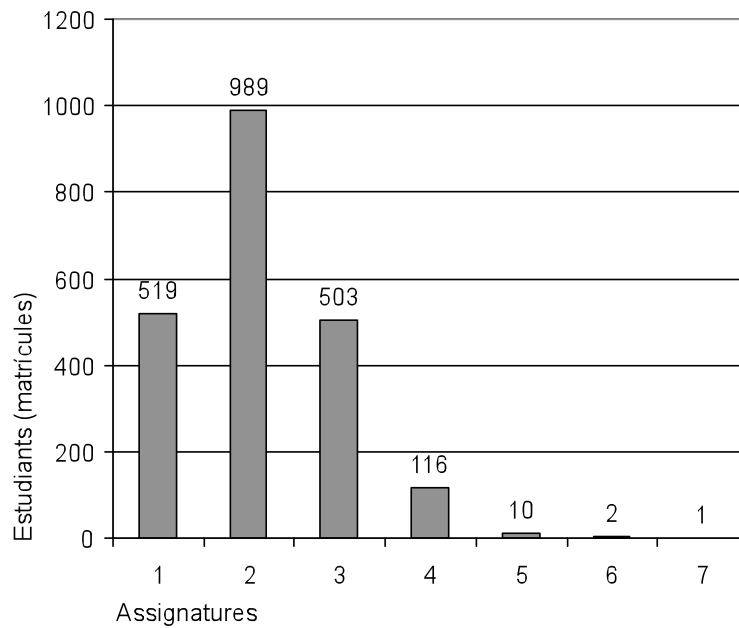


Figura 4.23: Nombre d'estudiants (matrícules) que ho aproven tot a final de curs.

Per altra banda, s'ha estudiat el nombre de matrícules realitzades per estudiants a cada semestre els quals no han aprovat totes les assignatures matriculades. Els resultats es poden veure a la Figura 4.24 on, per facilitar la comparació de dades, també s'ha inclòs la informació de la Figura 4.23. Així, tenint en compte només les noves dades, es pot observar que el patró més freqüent és matricular dues assignatures i aprovar-ne una o cap.

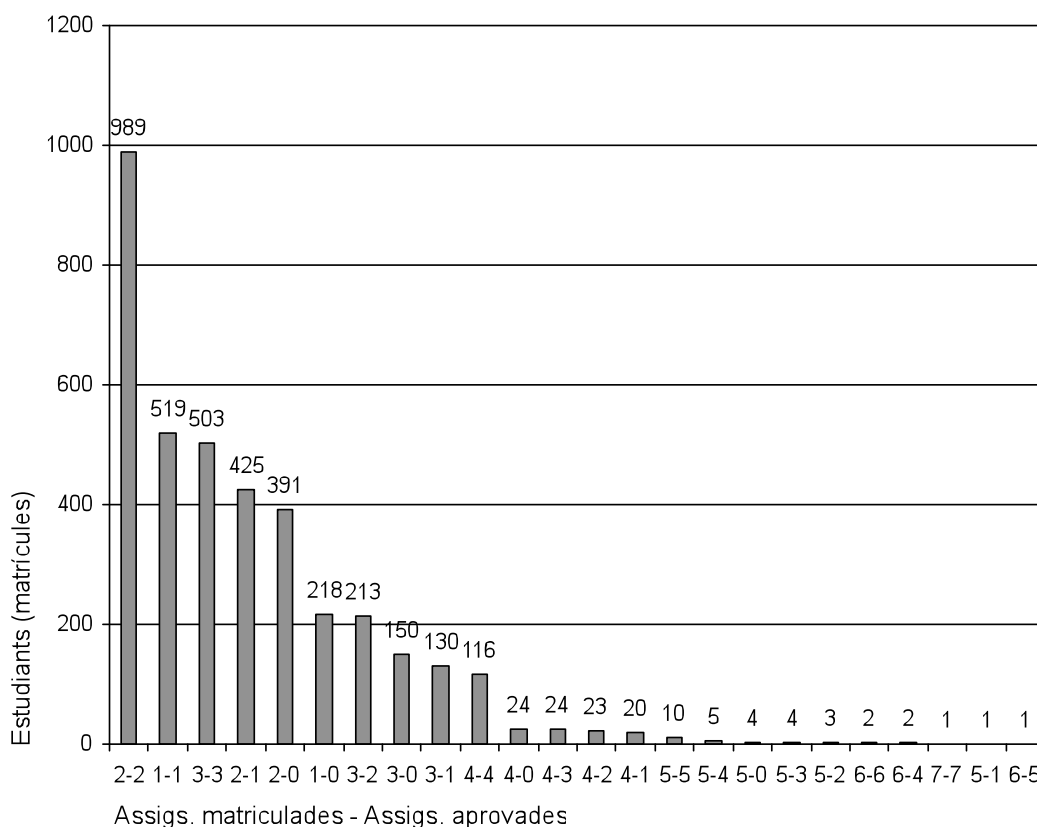


Figura 4.24: Relació entre assignatures matriculades i assignatures aprovades per cada matrícula.

El promig del nombre d'assignatures que es matriculen quan s'aprova tot (Figura 4.23) és de 2,12 i, en canvi, el promig del nombre d'assignatures que es matriculen quan no s'aprova tot és de 2,32. El test de  $\chi^2$  de Pearson realitzat a partir de la taula de contingència de 2x6 (eliminant el cas puntual de 7 assignatures matriculades i aprovades) mostra amb una significació de  $p < 0.001$  que les dues distribucions són diferents. Igualment, el test T de Student de dues cues mostra que les dues mitjanes són diferents i que, per tant, es pot afirmar amb  $p < 0.001$  que els estudiants que suspenen alguna assignatura es matriculen de més assignatures que els que ho aproven tot. Així, tot i que en la pràctica depèn de les assignatures concretes, en cas de dubte es pot aconsellar matricular dues assignatures en lloc de tres.

També és interessant observar les probabilitats per cada grup d'assignatures matriculades mitjançant un índex que és calcula de la forma següent:

$$m = \frac{|n_1 - n_2|}{n_1 + n_2}$$

on  $n_1$  i  $n_2$  són el nombre de casos de cada classe, aprovar-ho tot i no aprovar-ho tot respectivament. Aquest índex mesura el marge entre una possibilitat i l'altra, és a dir, entre aprovar totes les assignatures que es matriculen o suspendre'n alguna. Així, com més gran sigui el marge, es pot establir que és més fàcil aprovar-ho tot en aquella combinació. A la Taula 4.14 es presenta l'índex calculat pels diferents nombres d'assignatures matriculades. Cal comentar que el nombre de mostres que han matriculat més de 4 assignatures no ho fan estadísticament significatiu, tot i que sí que indica una tendència. A partir de la informació obtinguda, un sistema recomanador podria aconsellar matricular una o dues assignatures, però no tres. Curiosament, les matrícules de 4 assignatures també seria recomanable matricular-les, però això es deu principalment al valor petit de  $n_1+n_2$ .

<b>n<sub>1</sub></b>	<b>n<sub>2</sub></b>				<b>Σ n<sub>2</sub></b>	<b>Marge</b>
1-1 = 519	1-0 = 218				218	0,4084
2-2 = 989	2-1 = 425	2-0 = 391			816	0,0958
3-3 = 503	3-2 = 213	3-1 = 130	3-0 = 150		493	0,01
4-4 = 116	4-3 = 24	4-2 = 23	4-1 = 20	4-0 = 24	91	0,1208

Taula 4.14: Combinacions d'assignatures matriculades i aprovades amb el marge entre suspendre alguna assignatura i aprovar-ho tot.

### **Combinacions d'assignatures**

Un altre estudi que s'ha realitzat ha consistit en analitzar les combinacions d'assignatures que els estudiants matriculen conjuntament i observar si hi ha combinacions d'assignatures que afavoreixen un major o menor rendiment acadèmic, és a dir, un major o menor nombre d'assignatures aprovades. Per fer-ho, s'ha agrupat en

una taula totes les combinacions d'assignatures matriculades i s'han comptat el nombre de vegades que s'aproven totes, així com el nombre de vegades que se'n suspèn alguna. D'aquesta manera es té informació d'aquelles combinacions d'assignatures que quan es realitzen juntes s'aproven o es suspenen amb més freqüència. La taula obtinguda aporta informació si s'ordena per les combinacions d'assignatures que afavoreixen no superar-ne alguna. Així, es pot concloure que, amb les dades disponibles actualment, és més fàcil identificar les combinacions d'assignatures que porten al fracàs acadèmic que les que porten a l'èxit. A la Taula 4.15 es presenten les 10 primeres combinacions d'assignatures amb més suspesos, on s'ha inclòs l'índex que mesura el marge entre cadascuna de les dues possibilitats. Un valor elevat del marge indica que una de les dues probabilitats predomina sobre l'altra. En aquest cas, es pot establir que com més gran és el marge més crítica és aquella combinació d'assignatures.

<b>Assignatures</b>	<b>Aprovats</b>	<b>Suspesos</b>	<b>Marge</b>
11.006 - 11.058	58	38	<i>0,2083</i>
11.013	112	31	<i>0,5664</i>
11.059 - 11.061	16	23	<i>0,1794</i>
11.007 - 11.059	5	20	<i>0,6</i>
11.008 – 11.059	9	19	<i>0,3571</i>
05.053 - 11.006 - 11.058	18	19	<i>0,0270</i>
11.059 - 11.068	8	17	<i>0,36</i>
11.006 - 11.031 - 11.058	11	15	<i>0,1538</i>
11.061 - 11.062	6	13	<i>0,3684</i>
05.053 – 11.006	7	13	<i>0,3</i>

Taula 4.15: Combinacions d'assignatures i el nombre d'estudiants que les aproven totes o en suspenen alguna.

Com es pot veure a la Taula 4.15, el codi de les assignatures que apareixen a la primera columna es repeteix amb certa freqüència, això significa que són sempre les mateixes assignatures les que afavoreixen suspendre. La 11.006, 11.059 i 11.058 són les que més

apareixen a la taula i, per tant, això indica una tendència. No obstant, cal dir també que es tracta de les assignatures amb més matrícula. També es pot veure que hi ha combinacions de dues assignatures que quan s'afegeix una tercera, el seu rendiment canvia, com és el cas de combinar qualsevol altra assignatura amb la parella 11.006 - 11.058, que soles s'aproven amb més freqüència que no pas es suspenen, mentre que amb una tercera assignatura això canvia.

Els resultats que es presenten a la Taula 4.15 posen de manifest la necessitat de revisar les recomanacions de matrícula d'algunes assignatures i al mateix temps la necessitat de revisar les pròpies assignatures individualment, per així observar si hi ha qüestions relacionades amb el seu disseny i continguts que dificultin la seva superació. Concretament, el cas de l'assignatura 11.059 resulta exemplar en aquest aspecte. Per altra banda, els resultats de la Taula 4.15 mostren una assignatura que cursada ella sola la suspenen 31 estudiants i per això apareix en segona posició, tot i que al mateix temps la superen 112 estudiants. Aquest comportament és explicable ja que el codi 11.013 es correspon amb el "Projecte Fi de Carrera", que es recomana que es realitzi al final dels estudis i amb dedicació completa, i per aquest motiu els estudiants la matriculen sola. La natura d'aquesta assignatura fa que molts estudiants optin per no presentar-se i repetir-la en el proper semestre.

### **Titulacions progressives**

Un altre estudi realitzat amb el conjunt de dades del segon cicle de l'Enginyeria en Informàtica correspon a l'anàlisi de les titulacions progressives. Un total de 115 dels 687 estudiants de les cohorts seleccionades per l'estudi han obtingut una o més titulacions progressives que ofereix la universitat. Cal tenir en compte que un estudiant pot assolir més d'una titulació progressiva, ja que aquestes s'assoleixen només pel fet d'haver superat el conjunt d'assignatures que les formen. Dels 36 estudiants graduats, 7 ho han fet amb una o més titulacions progressives, per exemple.

S'han estudiat quines són les titulacions progressives més assolides, per així observar l'estratègia que segueixen els estudiants a l'hora de matricular les assignatures. A



continuació, a la Taula 4.16, es presenta per cada area d'especialització o de titulació progressiva quants estudiants l'han assolit. És interessant comentar que el nombre total de titulacions progressives assolides de la taula és més gran que 115, el nombre d'estudiants que han assolit una titulació progressiva. Això vol dir que un mateix estudiant assoleix més d'una titulació progressiva, com s'ha comentat amb anterioritat.

<b>Titulació Progressiva</b>	<b>Estudiants</b>
1. Construcció d'aplicacions i sistemes distribuïts (34,5 crèdits)	33
2. Gestió avançada de les dades i del coneixement (30 crèdits)	4
3. Direcció de projectes i de sistemes d'informació (36 crèdits)	6
4. Enginyeria del programari (30 crèdits)	64
5. Disseny, gestió i seguretat de xarxes (33 crèdits)	31
6. Gràfics per computador (18 crèdits)	0
7. Interacció persona-ordinador (18 crèdits)	0

Taula 4.16: Nombre d'estudiants que assoleix cada titulació progressiva.

Els resultats obtinguts serveixen per conèixer les àrees d'especialització i titulacions progressives per les quals els estudiants mostren una preferència clara. Una anàlisi més en profunditat dels resultats posa de manifest que les titulacions progressives preferides pels estudiants són les que tenen més assignatures troncales i, en canvi, aquelles titulacions progressives amb més assignatures optatives són les que menys estudiants segueixen. Aquest és el cas de les anomenades "Gràfics per computador" i "Interacció persona-ordinador", on totes les assignatures que s'han de cursar són optatives. La conclusió que es pot extreure d'això és que els estudiants no segueixen una estratègia per a assolir una titulació progressiva, sinó que les fan per acumulació d'assignatures, és a dir, van cursant assignatures i assoleixen una titulació progressiva, perquè les assignatures que han cursat coincideixen, com per exemple a l'àrea d'"Enginyeria del programari", on totes les assignatures menys una són troncales. Per tant, els estudiants prioritzen la troncalitat en front de l'optativitat, així que caldria revisar l'oferta de titulacions progressives tenint en compte aquest fet.

## **Anàlisi de l'abandonament**

Un altre estudi que s'ha realitzat amb les dades disponibles de l'Enginyeria en Informàtica correspon a l'anàlisi d'abandonament dels estudiants. Aquest primer estudi de l'abandonament pretén posar sobre la taula uns primers indicadors sobre què significa abandonar a la UOC i com es pot mesurar. Cal tenir en compte que l'anàlisi s'ha realitzat amb les 5 cohorts seleccionades dels estudiants de segon cicle de l'Enginyeria en Informàtica i, per tant, no reflecteix el comportament d'abandonament de tota la universitat.

En aquest estudi s'ha analitzat la navegació dels estudiants entre els semestres de permanència a la UOC, és a dir, si d'un semestre passen al següent, si fan un salt entre diferents semestres, desapareixen o bé es graduen. Aquesta navegació entre semestres permet conèixer com van avançant fins a graduar-se o abandonar, i observar si es produeix algun tipus de patró. La Figura 4.25 mostra el graf amb la navegació per semestres dels estudiants del conjunt de dades. El node que representa el primer semestre inclou tots els estudiants de primer semestre de cadascuna de les cohorts que s'han estudiat. S'ha establert en un 2% el límit per a dibuixar els arcs, és a dir, que només si hi ha més d'un 2% d'estudiants que van d'un node a un altre es dibuixa l'arc. A cada node apareix el número de semestre i, entre parèntesis, el nombre d'estudiants.

Els resultats mostren que els estudiants majoritàriament matriculen un semestre després d'un altre, o bé en descansen un. Es pot comprovar que pocs estudiants que descansin més de dos semestres seguits es reincorporen. Aquests representen menys del 2%, i no resulta significatiu de cara a conèixer els patrons de comportament de matrícula i abandonament dels estudiants. Aquest valor de dos semestres descansant és, per tant, una definició acurada d'abandonament. Així, dels 687 estudiants de les 5 cohorts que componen el conjunt de dades, en el semestre numero 8 en queden 250; la resta, o bé s'han graduat, o bé han abandonat. Més enllà no es pot dir res donat que les cohorts seleccionades només assegura tenir informació de 8 semestres.

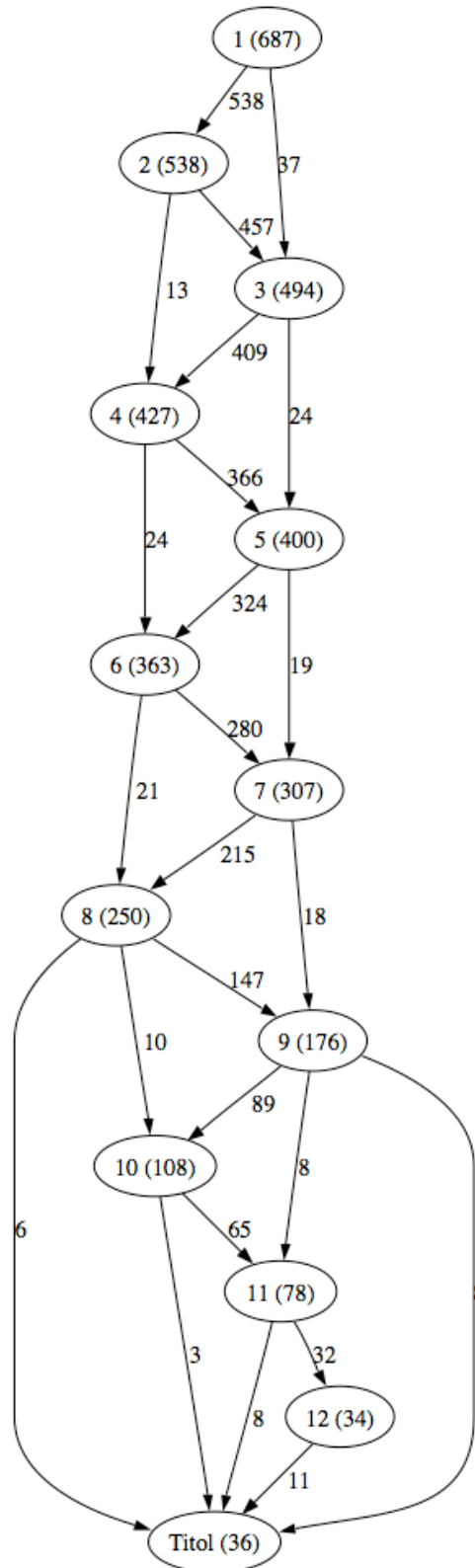


Figura 4.25: Navegació entre semestres amb assoliment del títol.

A continuació s'ha construït una taula on es presenta per cada semestre el nombre d'estudiants que el comencen, el 2% de possibles desapareguts que no apareixen a la Figura 4.25, el nombre total d'estudiants que desapareixen i que, per tant, abandonen, i els percentatge que representa, respecte a aquell semestre i el total acumulat. El càlcul es realitza fins al semestre 6, perquè de les cinc cohorts que s'estudien, només es coneix l'abandonament amb certesa fins al semestre 8, que és el temps que porten a la universitat els estudiants de la darrera cohort. Se'n consideren sis perquè dos semestres és el temps màxim de descans per saber si un estudiant ha abandonat o no, tal i com s'ha definit amb anterioritat.

<b>Semestre</b>	<b>Est.</b>	<b>2%</b>	<b>Aband.</b>	<b>% aband.</b>	<b>% acum.</b>
1	687	14	98	14,3%	14,3%
2	538	11	57	10,6%	22,6%
3	494	10	51	10,3%	30%
4	427	9	28	6,5%	34,1%
5	400	8	49	12,2%	41,2%
6	363	7	55	15,15%	49,2%

Taula 4.17: Abandonament a cada semestre.

Com es pot veure a la Taula 4.17, el percentatge d'estudiants que abandonen va disminuint al llarg dels semestres, però té un lleuger increment a partir del cinquè semestre. Es pot veure que el percentatge acumulat creix uns 8 punts per cada semestre, el que fa que al final del sisè semestre hagin abandonat la meitat dels estudiants que van començar la titulació. Això mostra la importància per la institució d'un estudi detallat de l'abandonament, donat que afecta a un gran percentatge d'estudiants.

## Navegació dels estudiants d'Enginyeria en Informàtica

L'últim experiment realitzat amb les dades de l'Enginyeria en Informàtica recollides durant el semestre 20062 pretén mostrar les possibilitats de l'anàlisi realitzada pel que fa a explorar la relació entre el perfil d'usuari previ a la matrícula d'un semestre, la matrícula concreta d'aquell semestre, el seu perfil de navegació (segons s'ha obtingut a la secció 4.6.3) i els resultats acadèmics obtinguts que serviran com a variable objectiu. En el cas particular dels estudiants involucrats en aquest experiment, matriculen 2,17 assignatures en promig, similar al promig general dels estudiants de les titulacions homologades en català. El primer pas consisteix a generar un conjunt de dades amb tota la informació disponible per cada estudiant, intentant simplificar al màxim el nombre de variables d'entrada fent agrupacions que responen a una qüestió concreta, com es mostra a continuació:

- Sexe.
- Edat en el moment d'iniciar la titulació.
- Número de semestres que porten a la UOC.
- Número de semestres amb matrícula real (sense incloure els semestres que han descansat).
- Nombre de assignatures superades amb anterioritat (aprovades o adaptades). Aquesta variable equivaldria a la "inèrcia" que porta l'estudiant realitzant els seus estudis.
- Ràtio d'assignatures superades amb anterioritat sobre el nombre total d'assignatures matriculades. Aquesta variable mesura com de bo és l'estudiant en realitzar la seva planificació semestral i com va superant les matrícules consecutives, podria dir-se que és la "qualitat" demostrada com a estudiant. Aquesta variable és zero pels estudiants de primer semestre que no tenen un històric de matrícules prèvies al semestre 20062.
- Total d'assignatures i crèdits matriculats en el semestre 20062.
- Total d'assignatures i crèdits cursats en el semestre 20062 (és a dir, no inclou les adaptacions).
- Components del perfil de navegació tal i com s'han obtingut a 4.6.3 (un total de sis).

Com a variable objectiu s'ha definit hoAprovaTot, que permet fer una classificació dels estudiants en dos conjunts, els que ho aproven tot o els que suspensen o no es presenten en alguna de les assignatures que han matriculat. Encara que es podrien haver definit altres variables objectiu (els que ho suspensen tot, el percentatge de crèdits superats, etc.), l'objectiu d'aquest experiment es mostrar si hi ha cap relació entre les variables descrites amb anterioritat, més que no pas construir un sistema classificador acurat que, per altra banda, no tindria massa sentit.

Usant l'algorisme C5.0 de SPSS Clementine s'ha construït un total de 10 arbres de decisió utilitzant *cross-validation* [65], per minimitzar l'impacte que suposa utilitzar un conjunt de dades relativament petit (645 estudiants, un cop eliminats aquells que només han matriculat assignatures adaptades, que es distribueixen en 289+356, és a dir, 289 que suspensen alguna assignatura i 356 que aproven tot allò que matriculen) amb tantes variables d'entrada. No s'han usat cap dels paràmetres d'ajust de l'algorisme, ja que es tracta d'un experiment exploratori: la rellevància de les variables i la seva importància pel que fa a la determinació de la variable objectiu queda explicitada per la profunditat on s'usa com a variable de tall en l'arbre de decisió.

Com ja s'ha dit, el conjunt inicial es divideix en 289 estudiants que suspensen o no presenten alguna assignatura (el 44,8%) i 356 que ho aproven tot (el 55,2%). A cada tall de l'arbre, l'objectiu de l'algorisme C5.0 és maximitzar el marge entre aquests dos valors per un dels dos conjunts generats. La Figura 4.26 mostra els tres primers nivells de l'arbre generat, amb les variables que són més rellevants per a explicar l'objectiu proposat.

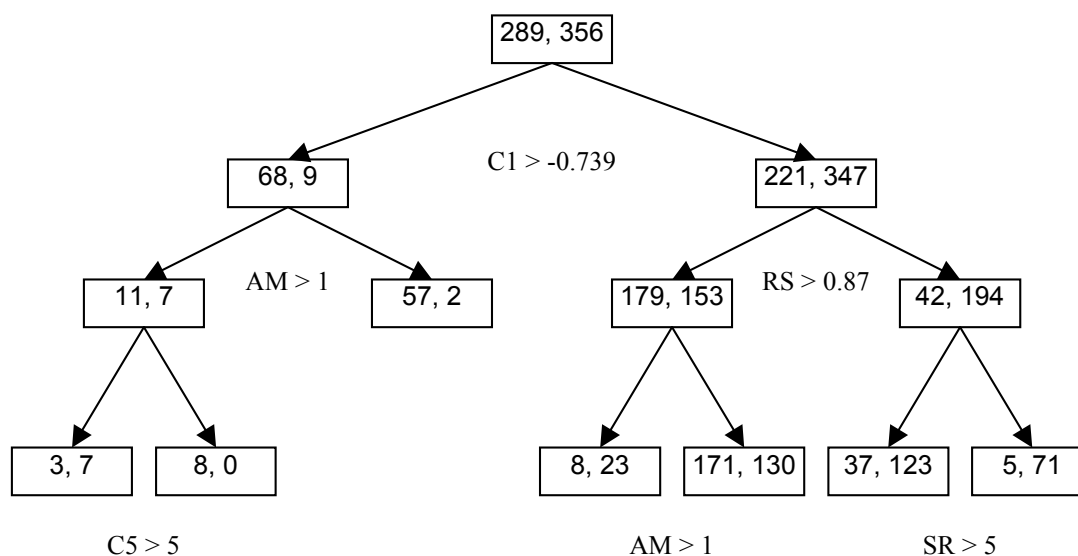


Figura 4.26: Arbre de decisió generat per l'algorisme C5.0 de SPSS Clementine.

És interessant destacar que el primer tall ( $C1 > -0.739$ ) separa 68 estudiants dels 289 que no aproven alguna assignatura (un 23,5%) mitjançant el primer component trobat del perfil de navegació, és a dir, la freqüència del patró de connexions. Aquells estudiants amb un valor inferior a  $-0.739$  (que coincideix aproximadament amb la desviació típica del component, que té una mitjana propera a zero, és a dir, que estan a una  $\sigma$  de distància o més per sota) formen clarament un grup d'interès pel que fa a l'objectiu establert, tot i que també hi ha 9 estudiants que ho aproven tot que cauen en aquest conjunt. És evident que caldria realitzar una anàlisi més detallada del perfil dels estudiants que formen aquest col·lectiu, però no deixa de ser significatiu que la primera variable seleccionada per l'algorisme sigui precisament aquesta, el que relaciona directament el poc ús del campus virtual (els estudiants amb un perfil de connexions esporàdiques) amb l'obtenció de pitjors resultats acadèmics. Aquests  $68+9$  estudiants són de nou separats en dos conjunts mitjançant la variable que mesura el total d'assignatures matriculades ( $AM > 1$ ) en el semestre en curs (incloent-hi les adaptades), però com el tall es realitza pel valor 1 (i les matrícules amb només assignatures adaptades no s'han tingut en compte en aquest estudi) això vol dir que s'estan separant els estudiants que han matriculat i cursat una sola assignatura ( $11+7$  en total) de la resta ( $57+2$ ). Aquest últim conjunt ja no és seleccionat i es considera prou pur per

l'algorisme; de fet explica correctament 57 dels 289 casos (un 19.7%) amb un error de classificació inferior al 3.4%. Per tant, podem concloure que els estudiants que tenen un baix patró de connexió i que matriculen dues o més assignatures són molt propensos a suspendre'n alguna. Els 11+7 estudiants que només matriculen una assignatura són de nou dividits en dos conjunts ( $C5 > -0.027$ ), un de 3+7 estudiants i un altre de 8+0, usant la component 5 amb un valor de tall igual a  $-0.027$ , el que indica que aquells estudiants que fan poc ús de l'espai docent i no accedeixen gairebé mai a la biblioteca són també propensos a suspendre alguna assignatura. La combinació d'aquests dos conjunts explica el 22.5% dels casos amb un error inferior al 3%.

La resta d'estudiants (el gruix de 221+347) són de nou separats en dos conjunts d'acord a la ràtio d'assignatures superades, usant un valor de 0.87 ( $RS > 0.87$ , és a dir, que han aprovat aproximadament 7 de cada 8 assignatures matriculades, el que es pot considerar un nivell de seguiment molt elevat; de fet el promig d'aquesta variable d'aquest conjunt és de 0.981, molt proper al límit d'1 que vol dir aprovat tot allò que s'ha matriculat). De nou seria necessari analitzar quines són les assignatures superades (majoria de troncal o es tracta d'optatives?), però cal interpretar aquesta variable més enllà d'aquest valor. El fet important és que després del patró de connexió, la variable que aporta més informació pel que fa al rendiment acadèmic dels usuaris que ho superen tot és el seu historial "qualitatiu", tal i com s'ha descrit amb anterioritat, és a dir, com ho han fet prèviament, més que no pas la combinació o càrrega docent d'assignatures matriculades en el semestre en qüestió, el que passava amb els que no superaven alguna assignatura. El conjunt d'estudiants que supera gairebé tot el que matricula és de nou dividit en dos conjunts d'acord a una altra variable relacionada, com és el nombre de semestres reals (incloent-hi descansos) que porten matriculats. Els estudiants que porten 6 o més semestres matriculant-se ( $SR > 5$ , incloent-hi el semestre en curs) formen un col·lectiu (71 estudiants, un 19.9%) que ho aprova tot, excepte 5 casos que es poden considerar outliers (un error del 6.6%). Això demostra que l'experiència que els estudiants van obtenint com a estudiants d'un entorn virtual és un factor important: saber triar les matrícules correctament és un element important del procés d'aprenentatge.

Per la seva banda, els 179+153 estudiants que no tenen un historial d'assignatures superades tan bo (un promig de 0.445, és a dir, que només superen una de cada dues



assignatures que matriculen) són de nou dividits en dos conjunts d'acord a la variable que indica el total d'assignatures matriculades en el semestre en curs, pel mateix valor de tall, una única assignatura matriculada i cursada ( $AM > 1$ ). Els estudiants amb una sola assignatura són més propensos a aprovar-la que es que s'han matriculat de dues o més, a l'igual que s'ha descrit anteriorment per l'altre conjunt d'estudiants amb un baix patró de connexió.

L'arbre amb tres nivells no és capaç d'explicar correctament encara la majoria d'estudiants (171+130), per lo que necessita anar dividint aquest conjunt uns quants nivells de profunditat més. Les variables utilitzades en aquests nivells inferiors són de nou el patró de connexió, el nombre d'assignatures superades i el nombre d'assignatures matriculades, el que indica que aquestes tres variables són les de més rellevància a l'hora de determinar el rendiment acadèmic dels estudiants. Una opció seria fusionar de nou tots els conjunts considerats no purs a aquest tercer nivell i començar de nou el procés per detectar noves variables que siguin de rellevància per a l'objectiu proposat, encadenant un seguit d'arbres de decisió petits [115].

En resum, aquest experiment mostra que és possible obtenir predictors fiables del comportament dels usuaris pel que fa al seu rendiment acadèmic, tenint en compte la seva navegació pel campus virtual, el seu historial acadèmic i la quantitat d'assignatures matriculades en el semestre en curs. Els estudiants que cauen en nodes de l'arbre de decisió més propensos a suspendre alguna assignatura podrien veure's afavorits per l'ajuda d'un sistema que detecti aquesta situació el més aviat possible i, via els tutors i els consultors, iniciï un procés de seguiment més personalitzat de cada estudiant per analitzar quines són les causes particulars que poden portar a aquesta situació indesitjable. Aquest procés podria situar-se en el temps al voltant de la publicació dels resultats de la primera prova d'avaluació continuada, que com s'ha vist anteriorment en la secció 4.7.1, pot determinar el rendiment dels estudiants.

### 4.8.2 Conclusions dels experiments al nivell 3

En aquesta secció s'han presentat els experiments realitzats al nivell de *lifelong learning*. L'objectiu que es persegueix en aquest nivell és recollir i analitzar l'activitat dels usuaris durant el temps que dura la seva relació amb la universitat i amb el campus virtual. Els experiments que s'han realitzat han utilitzat dades del segon cicle de l'Enginyeria en Informàtica i els resultats obtinguts mostren que es poden obtenir resultats rellevants. Per tant, és indicat apuntar la idoneïtat de realitzar anàlisis similars amb dades d'altres titulacions, no només per obtenir informació rellevant sobre el propi conjunt de dades sinó també per analitzar les diferències entre titulacions. D'entre les diferents conclusions que s'obtenen dels estudis realitzats, es destaca la necessitat de revisar les recomanacions de matrícula relacionades amb els itineraris d'especialització de les titulacions progressives. Les recomanacions han de ser revisades i redissenyades per així tenir en consideració els itineraris formatius reals dels estudiants de l'entorn virtual d'aprenentatge, que mostren els seus interessos i necessitats. D'aquesta manera es poden adequar i millorar les competències professionals dels estudiants. Per altra banda, és necessari continuar estudiant el comportament d'abandonament dels estudis amb dades d'altres titulacions i, a partir d'aquí, replantejar la definició actual d'abandonament amb la que es treballa.

## 4.9 Resum

En aquest capítol s'ha descrit l'escenari d'estudi d'aquesta tesi, el campus virtual de la UOC i els seus estudiants i s'han presentat els diferents experiments i anàlisis portats a terme. S'han mostrat els experiments inicials que van servir per centrar les bases de la proposta i la investigació realitzada des d'aleshores. S'ha descrit el procediment establert per la recollida, gestió i processat dels fitxers de log del campus virtual i el procés utilitzat per la obtenció dels camins de navegació. També s'han presentat les anàlisis realitzades a cadascun dels nivells definits al capítol 3, que constitueixen l'aproximació a l'anàlisi d'entorns complexos i canviants com és el de la Universitat Oberta de Catalunya. A cada nivell s'han obtingut resultats rellevants que mostren que l'aproximació multinivell a l'anàlisi d'entorns virtuals d'aprenentatge és útil i

proporciona informació útil per a analitzar el disseny del procés d'aprenentatge i prendre decisions al respecte.



## Capítol 5

### Conclusions

Un cop descrita la proposta d'anàlisi d'entorns virtuals d'aprenentatge, el sistema de marcatge utilitzat i els resultats obtinguts en els experiments realitzats, només resta extreure les conclusions, així com fer una síntesi del treball realitzat. En aquest darrer capítol es presenta un resum dels objectius inicials i l'avaluació del seu grau d'assoliment. A continuació s'exposen les conclusions obtingudes del treball d'investigació realitzat i també es descriuen les aportacions originals sorgides del procés d'investigació. Per finalitzar, es presenten les línies de recerca que queden obertes, així com d'altres que poden ser la continuació de l'esforç realitzat dins del marc del grup de recerca on s'inscriu aquesta tesi doctoral.

#### 5.1. Assoliment dels objectius plantejats

En aquesta secció es descriu com s'han assolit, tant l'objectiu general presentat al primer capítol, com els objectius específics, indicant en cada cas la validació realitzada, que permet avaluar el seu grau d'assoliment.

La hipòtesi d'aquesta tesi és que es poden extreure patrons d'ús i navegació a partir del rastre que deixen els usuaris d'un entorn virtual d'aprenentatge, i que aquesta informació pot ser usada per a millorar l'entorn virtual i l'ús que en fan els seus usuaris. A continuació es presenten els resultats que permeten validar aquesta hipòtesi.

### **5.1.1 Resum de l'objectiu general**

Un objectiu d'aquesta tesi és fer una aportació a les àrees de l'e-learning i la interacció persona-ordinador, en relació a obtenir informació dels usuaris i de l'ús que fan d'un sistema. Concretament, s'ha definit una metodologia que permet obtenir informació rellevant de la navegació dels estudiants d'un entorn virtual d'aprenentatge, entenent que ho fan, no només pels espais i serveis del campus virtual, sinó també pels continguts, activitats i recursos educatius de les assignatures, així com per les assignatures d'una o més titulacions i ofertes formatives, és a dir, a diferents nivells. Aquest objectiu es basa en la necessitat de conèixer als estudiants i les seves interaccions amb l'entorn virtual, per així poder millorar la usabilitat del sistema i la seva experiència d'ús, i aportar informació a un model d'usuari que permeti introduir elements de personalització en el procés d'aprenentatge.

Per a assolir aquest objectiu s'ha proposat una metodologia d'anàlisi que es basa en un estudi a tres nivells que permet abastar els objectius i motivacions dels usuaris en diferents períodes de temps. Aquesta proposta d'anàlisi s'ha presentat detalladament al capítol 3 i proposa estudiar el campus virtual i els seus usuaris mitjançant tres nivells que estan interrelacionats entre ells: el nivell de sessió, el nivell de semestre acadèmic i el nivell d'aprenentatge al llarg de la vida. El nombre de nivells d'anàlisi ve donat per les característiques pròpies del campus virtual i la natura específica dels ensenyaments no presencials i en línia. Dos nivells d'anàlisi són insuficients, mentre que quatre no permeten una definició consistent que inclogui les tasques que realitzen les estudiants d'un entorn virtual d'aprenentatge.

Els estudis i experiments preliminars van mostrar que les dades disponibles en el propi campus virtual de la UOC no aportaven informació rellevant per a estudiar la navegació dels usuaris. Per aquest motiu, s'ha fet una proposta de marques integrades que enriqueixen la informació de les entrades dels fitxers de log i que permeten portar a terme un millor seguiment de la navegació dels usuaris. S'han desenvolupat dues estratègies complementàries, una orientada a capturar l'acció i l'altra a capturar el

contingut, per tal de simplificar el posterior procés d'anàlisi. Aquesta proposta s'ha implementat i posat en marxa i, d'aquesta manera, s'ha pogut capturar la navegació dels usuaris del campus virtual i analitzar-la, obtenint resultats rellevants en el nivell de sessió. Per als nivells de curs i de *lifelong learning* s'han portat a terme experiments i anàlisis incloent-hi també dades pròpies de la universitat. Aquests estudis aporten resultats importants per a personalitzar el procés d'aprenentatge i per assessorar als estudiants en el procés de matrícula al llarg dels seus estudis a la universitat.

### **5.1.2 Objectius específics i assoliment**

L'assoliment dels objectius específics d'aquest treball de tesi ha comprès les següents activitats:

1. Proposta d'una metodologia d'anàlisi multinivell dels usuaris d'entorns virtuals d'aprenentatge.
2. Definició de personatges i escenaris a la UOC a partir de la informació disponible a la universitat.
3. Proposta d'anàlisi a nivell de sessió per a estudiar la navegació dels usuaris a partir dels fitxers de log.
4. Definició i implementació d'un sistema de marques integrades en els logs del campus virtual, mitjançant estratègies complementàries.
5. Definició, disseny i implementació d'un mecanisme de recollida de dades de navegació i ús de l'entorn virtual d'aprenentatge.
6. Obtenció dels camins de navegació dels usuaris del campus virtual.
7. Proposta d'anàlisi a nivell de curs que considera la navegació dels estudiants i l'assoliment dels objectius d'aprenentatge.
8. Proposta d'anàlisi a nivell d'aprenentatge al llarg de la vida que té en compte les recomanacions de matrícula, les assignatures que matriculen els estudiants i el seu rendiment acadèmic.
9. Proposta d'aplicació dels resultats de l'anàlisi a la millora del campus virtual i l'experiència d'aprenentatge dels estudiants.

10. Proposta de creuament entre el resultats obtinguts de l'anàlisi del comportament dels usuaris i les possibilitats d'adaptació de diferents escenaris educatius que es generen dins d'un entorn virtual d'aprenentatge, identificant les opcions de personalització plausibles.

Aquests objectius específics han permès respondre les preguntes de recerca, tal i com es descriu a continuació.

## **5.2 Revisió de les preguntes de recerca**

Al capítol 1 s'han plantejat un seguit de preguntes de recerca que, un cop presentada la proposta i els resultats obtinguts, és interessant que siguin revisades. Les diferents preguntes plantejades reflecteixen els objectius proposats; d'aquesta manera cada grup de preguntes té una relació amb els objectius específics i el seu assoliment.

El primer grup de preguntes fa referència a la navegació dels usuaris i la seva captura. Amb les dades disponibles originalment no era possible analitzar la navegació dels usuaris, bàsicament pels següents fets: manca de focus en l'acció dels estudiants i volum intractable de dades disponibles. En aquest sentit s'ha definit i implementat un sistema de marques integrades al campus virtual que genera dades de log suficientment riques, les quals permeten obtenir la navegació dels usuaris. Per fer-ho, s'ha dissenyat i implementat un sistema de recollida i gestió dels fitxers de log i un procediment de processat d'aquests fitxers per a obtenir els camins de navegació. Aquest sistema és absolutament transparent als usuaris i no modifica la sintaxi dels fitxers de log clàssics.

El segon grup de preguntes planteja qüestions relacionades amb l'anàlisi de dades mitjançant una aproximació multinivell. Es realitza una proposta d'una metodologia d'anàlisi dels usuaris d'entorns virtuals d'aprenentatge basada en tres nivells. Aquests nivells s'han establert a partir de les característiques pròpies de la UOC i de la seva oferta formativa, aplicant de la tècnica de personatges i escenaris. Per cadascun dels nivells d'anàlisi establerts s'ha definit una proposta d'anàlisi concreta que té en compte les dades disponibles i la informació rellevant.



El tercer grup de preguntes fa referència a les dades disponibles i els resultats que se'n poden obtenir, així com la utilitat i rellevància dels resultats. A cadascun dels nivells s'han portat a terme diferents experiments, i s'han obtingut resultats interessants que poden ser utilitzats per a millorar el disseny de l'entorn virtual d'aprenentatge i el procés d'aprenentatge dels estudiants. Així, al nivell de sessió s'obté la navegació dels usuaris que, un cop analitzada, proporciona informació sobre l'ús que fan del sistema i les seves necessitats i interessos, tant en un moment concret del semestre com al llarg del temps. Al nivell de curs s'ha estudiat la relació entre la navegació dels usuaris i el seu rendiment acadèmic, mentre que al nivell d'aprenentatge al llarg de la vida s'han obtingut els patrons de matriculació dels estudiants, per exemple.

El quart grup de preguntes fa referència a l'obtenció de dades de navegació dels usuaris i les conclusions que se'n poden extreure. L'obtenció dels camins de navegació dels estudiants i el seu anàlisi s'ha realitzat al nivell de sessió i ha ocupat gran part del treball realitzat. L'anàlisi dels camins obtinguts mostra com naveguen els estudiants, les tendències de navegació que segueixen, així com els espais i serveis del campus virtual més utilitzats i que resulten més interessants. Així, els estudiants naveguen principalment per les àrees del campus virtual estrictament relacionades amb la docència i utilitzen la resta de serveis només quan disminueixen les exigències de les assignatures que realitzen. L'ús reflecteix que alguns espais com ara les aules no són tan usats com caldria esperar, mentre que d'altres, com la bústia, tenen un ús continu. Això pot indicar un cert grau d'anquilosament del campus virtual actual, en el sentit que s'utilitza principalment com a gestor de correu.

Finalment, el darrer grup de preguntes planteja qüestions relacionades amb l'adaptació de l'entorn virtual d'aprenentatge a les necessitats dels estudiants. A partir dels resultats que s'obtenen en les anàlisis realitzades a cadascun dels tres nivells, s'ha proposat un conjunt d'accions per millorar el disseny i utilitat de l'entorn, així com per adaptar l'entorn virtual d'aprenentatge a les necessitats i comportament dels estudiants, usant la informació obtinguda resultat del procés d'anàlisi.

Un cop adreçades les preguntes de recerca, ara ja es possible presentar les conclusions d'aquesta tesi doctoral a partir de la hipòtesi plantejada.

### 5.3 Conclusions principals

La principal conclusió que es desprèn dels resultats obtinguts sobre la UOC i el seu campus virtual és que *els estudiants l'utilitzen principalment pel correu electrònic*, amb un percentatge important (un 20%) de patrons de navegació basats en entrar, mirar si hi ha novetats i sortir, desenvolupant-hi molt poques activitats concretes relacionades amb el procés d'aprenentatge. La separació entre procés d'aprenentatge (estudiar) i interacció (comunicar-se) reforça aquest fet. Tenint en compte aquest resultat, sembla indicat establir els mecanismes adequats per notificar les novetats als estudiants sense necessitat d'entrar a l'entorn virtual, de manera similar a com es fa amb les notes de les proves finals, descongestionant així la càrrega del campus virtual.

Una altra conclusió que es pot extreure a partir dels resultats és que *els estudiants es connecten molt poc durant els caps de setmana* i, en canvi, *es connecten principalment entre setmana, durant tota la jornada laborable i als vespres*. Els resultats mostren que *els estudiants es poden caracteritzar en tres grups* segons la seva navegació: estudiants amb patrons de navegació i connexió poc definits, estudiants que es connecten esporàdicament i un tercer grup d'estudiants que ho fan amb molta freqüència i regularitat. Aquesta informació s'hauria de tenir en compte a l'hora de planificar les activitats docents.

També s'ha observat que *la navegació dels estudiants i la seva participació en les bústies de l'aula té una correlació positiva amb el seu rendiment acadèmic*. Així, es pot conèixer a priori quins estudiants poden tenir necessitats concretes i prendre aquelles mesures que es considerin adequades per al seu procés d'aprenentatge. Amb informació com aquesta es poden adreçar les necessitats específiques dels estudiants d'una aula combinant un sistema de detecció automatitzat i l'acció docent personalitzada del professor. Els factors més rellevants que estan relacionats amb el rendiment acadèmic són el tenir un patró de connexió regular i formalitzar la matrícula de manera ajustada a

les seves possibilitats. Aquesta informació pot utilitzar-se per fer un seguiment més personalitzat del procés d'aprenentatge de cada estudiant.

En relació als espais, servis i navegació pel campus virtual, *la plana d'inici és un bon agregador de serveis* molt utilitzat pels estudiants, mentre que pot ser interessant replantejar-se el disseny del menú lateral i superior actuals, molt menys utilitzats pels estudiants. En aquest sentit, el concepte de Personal Learning Environment [173] proporciona un model interessant a tenir en compte, on els estudiants poden configurar els elements de l'entorn virtual d'aprenentatge segons les seves necessitats.

Finalment, altres conclusions que s'obtenen tenen a veure amb la matrícula dels estudiants i l'abandonament dels estudis i es desprenen de les dades analitzades del segon cicle de l'Enginyeria en Informàtica. Per una banda, els resultats obtinguts mostren que *l'oferta de titulacions progressives actuals no es correspon amb les necessitats dels estudiants* i que aquests no segueixen cap estratègia per assolir-les, sinó que ho fan per acumulació d'assignatures. Cal, per tant, repensar la oferta de titulacions progressives i adaptar-les a les necessitats reals dels estudiants.

Estudiant el comportament de matrícula i el rendiment acadèmic dels estudiants, i com a exemple d'un sistema de suport a la matrícula, s'ha observat que amb l'objectiu de maximitzar el rendiment acadèmic, davant del dubte *és millor matricular una o dues assignatures que fer-ho de tres*. Malgrat que aquest fet també depèn de les assignatures concretes que es matriculen, s'ha comprovat com a cert per a les matrícules i el rendiment acadèmic estudiat en el segon cicle de l'Enginyeria en Informàtica.

Per altra banda, en relació a l'abandonament, se sap amb un alt grau de confiança que *si un estudiant no es matricula durant dos semestres seguits no continuarà els seus estudis*. Establir el criteri d'abandonament és clau per qualsevol estudi que se'n faci al respecte. Donats els resultats d'aquesta tesi, no sembla adequada la definició actual d'abandonament que s'està usant a la institució. El problema de l'abandonament mereix un estudi en profunditat per part de la universitat.

## **5.4 Línies de treball futures**

És important destacar que aquesta tesi s'ha desenvolupat en el marc d'un grup de recerca reconegut de la UOC. Tant durant el desenvolupament del treball d'aquesta tesi com en el moment de la seva finalització s'han observat possibles línies de treball futures. Aquestes línies inclouen, per una banda, extensions al treball presentat i, per altra banda, l'aplicació de la metodologia proposada a d'altres entorns interactius, així com altres perspectives de treball relacionades. En aquesta secció es descriuen les que es consideren més rellevants.

### **5.4.1 Extensions de la proposta**

La proposta d'anàlisi a tres nivells no es una aproximació tancada sinó que afavoreix utilitzar noves fonts de dades, així com l'ampliació dels mètodes utilitzats a cada nivell. Per exemple, es pot voler capturar i analitzar la navegació dels estudiants per les aules i, d'aquesta manera, conèixer el seu comportament de navegació més lligat amb els elements principals del procés d'aprenentatge com és la interacció amb els professors, els altres estudiants i els espais de l'aula. Actualment s'ha realitzat i implementat una proposta de mapa de marques pels espais de l'aula i els seus elements principals. D'aquesta manera es podran obtenir els camins de navegació i analitzar-los per a obtenir resultats que permetin millorar elements concrets del procés d'aprenentatge. És possible dissenyar experiments per analitzar el comportament dels usuaris en qualsevol espai de l'entorn virtual d'aprenentatge sempre i quan es defineixin el mapa d'accions i el conjunt de marques necessaris.

Per altra banda, es vol ampliar l'estudi realitzat al segon cicle de l'Enginyeria en Informàtica i incloure la resta d'estudis i titulacions de la UOC. Aquest estudi ampliat, a més de proporcionar informació rellevant sobre cada titulació, permetrà comparar comportaments de matrícula i d'abandonament entre els diferents grups i tipologies d'estudiants de la universitat. En aquest sentit, conèixer millor els diferents perfils dels estudiants de les diferents titulacions permetrà adaptar el seu procés d'aprenentatge de forma més acurada.

### 5.4.2 Aplicació a d'altres entorns interactius

La metodologia d'anàlisi utilitzada en aquest treball es pot adaptar i utilitzar en altres entorns interactius d'altres àmbits d'activitat. Per fer-ho, caldrà estudiar l'àmbit i els seus usuaris, tal com s'ha fet amb l'estudi de les característiques pròpies de l'e-learning. Així, per exemple, si es vol aplicar a un entorn de banca en línia, és necessari, en primer lloc, establir el nombre de nivells necessaris i, en segon lloc, aplicar la metodologia proposada.

Actualment s'ha iniciat un projecte on s'aplicarà la metodologia presentada a l'anàlisi d'un escenari de mobilitat, usabilitat i accessibilitat a la banca *online*. En aquest projecte s'analitzarà l'ús de sistemes de banca en línia i la seva xarxa de caixers automàtics, per així obtenir informació rellevant per a desenvolupar una proposta d'accés mitjançant dispositius mòbils, que té en compte les persones amb necessitats especials. D'aquesta manera el sistema podrà anar adaptant la interacció amb els caixers a les necessitats i hàbits dels seus usuaris.

### 5.4.3 Perspectives de treball relacionades

El treball de tesi doctoral que s'ha presentat s'emmarca en un grup de recerca on s'estan realitzant altres tesis. Així, s'està iniciant una tesi que analitzarà en profunditat els camins de navegació obtinguts, aplicant tècniques de mineria de dades i adaptant els algorismes a les característiques pròpies de les dades (fitxers de log, bases de dades de la universitat, etc.). Un altre treball en curs està enfocat en realitzar un disseny per competències a partir de l'anàlisi de les necessitats dels estudiants a nivell d'aprenentatge al llarg de la vida, oferint itineraris adaptatius. Per altra banda, un tercer treball de tesi doctoral proposa una metodologia d'anàlisi similar per a estudiar el comportament dels usuaris des del punt de vista informacional, pel que fa referència a la cerca i gestió dels recursos disponibles en un campus virtual. Altres projectes que es realitzen utilitzen les dades de navegació per a conèixer les necessitats dels usuaris per a

realitzar una aproximació a l'accessibilitat dels sistemes d'e-learning des del punt de vista del disseny centrat en l'usuari.

## **5.5 Consideracions finals**

Durant el desenvolupament i realització del treball que s'ha presentat, s'ha tingut la sort d'accedir fàcilment a moltes de les dades de la UOC, dels seus estudiants i del campus virtual. No obstant, també hi ha hagut dificultats associades, especialment a l'accés estable als fitxers de log i a la implementació del sistema de marques de seguiment.

És important que hi hagi una estratègia institucional clara d'accés a dades, implantació d'estàndards educatius i de disseny centrat en l'estudiant, especialment d'usabilitat i accessibilitat, pel que fa a l'entorn virtual d'aprenentatge. Aquesta estratègia també ha d'establir connexions amb altres àrees que, utilitzades adequadament, milloren l'experiència educativa dels estudiants, com per exemple, els continguts oberts, les xarxes socials i el treball col·laboratiu. Per una institució com la UOC, una bona arquitectura de la informació és absolutament necessària per assegurar un creixement sostenible.

Els resultats obtinguts en aquest treball posen sobre la taula que el campus virtual és utilitzat principalment com a eina de correu electrònic. No obstant, els resultats obtinguts també indiquen que si s'estudia el campus virtual i els seus usuaris s'obté informació que pot potenciar l'entorn virtual d'aprenentatge, adequant-lo a les necessitats dels estudiants, millorant la seva experiència d'ús, en un cercle constant d'evolució.

Tots aquests elements configuren un marc excepcional on l'estratègia institucional, la recerca, la innovació i la docència convergeixen i poden posicionar a la UOC com a futur referent en l'e-learning en les àrees de personalització, usabilitat, accessibilitat i mobilitat.

## **Annex A**

### **Dades quantitatives de la UOC i dels seus estudiants**

En aquest annex es presenten les dades quantitatives sobre l'escenari triat com a objecte d'estudi d'aquesta tesi, la UOC i els seus estudiants, a les quals s'ha tingut accés i són interessants per aquest treball.

#### **A1. Dades quantitatives del semestre de primavera del curs 2006-2007**

Les dades consultades i que es tenen en consideració en aquest treball corresponen al semestre de primavera del curs 2006-2007 i als estudiants de les titulacions homologades en català. A continuació es presenta un sumari d'aquestes dades, així com de les que s'han considerat més representatives dels estudiants de la UOC i aquelles que resulten d'un especial interès en aquest treball, un cop finalitzat el semestre en qüestió.

A continuació, a la Taula A1 presenta les dades relatives al nombre d'estudiants matriculats, la matrícula, el nombre d'estudiants graduats i els de nou accés. També es presenta satisfacció dels estudiants, aquest indicador s'obté de la enquesta de satisfacció que cada fi de semestre es realitza entre tots els estudiants

<b>Concepte</b>	<b>Valor</b>
Estudiants matriculats	25.782
Crèdits matriculats	473.747
Graduats	2.079
Estudiants nous	4.850
Satisfacció	80,3%

Taula A.1: Dades de matrícula de titulacions homologades en català en el semestre de primavera del curs 2006-2007.

Si es tenen en compte els estudiants matriculats d'altres ofertes formatives, també en català, com l'Ateneu Universitari, Lliure Elecció Interuniversitària, Intercampus, els Màsters oficials i els convenis amb altres institucions, el nombre d'estudiants matriculats al campus en català és 29.531. Aquests estudiants es van repartir entre les 2.218 aules del campus virtual, de les 1.075 assignatures amb docència activa en aquell semestre.

El perfil bàsic dels estudiants de les titulacions homologades en català és:

- Home o dona, lleugerament superior aquestes últimes.
- Entre 21 i 35 anys.
- Diplomats, llicenciats o amb estudis universitaris no acabats.
- Treballador.
- Estudien per motivacions laborals.
- Residència al Barcelonès.

Aquest perfil bàsic es detalla a continuació. Pel que fa al sexe:

<b>Sexe</b>	<b>%</b>
Home	48,3%
Dona	51,7%

Taula A.2: Sexe.



Pel que fa a l'edat, la Taula A3 mostra que la majoria d'estudiants, el 73,5%, es troba fora del perfil clàssic de la universitat presencial.

Rang d'edat	%
<20	2,8%
21-25	23,4%
26-30	30,5%
31-35	20,3%
36-40	10,9%
41-45	6,6%
46-50	3,5%
51-55	1,1%
>55	0,6%
NS/NC	0,3%

Taula A.3: Edat.

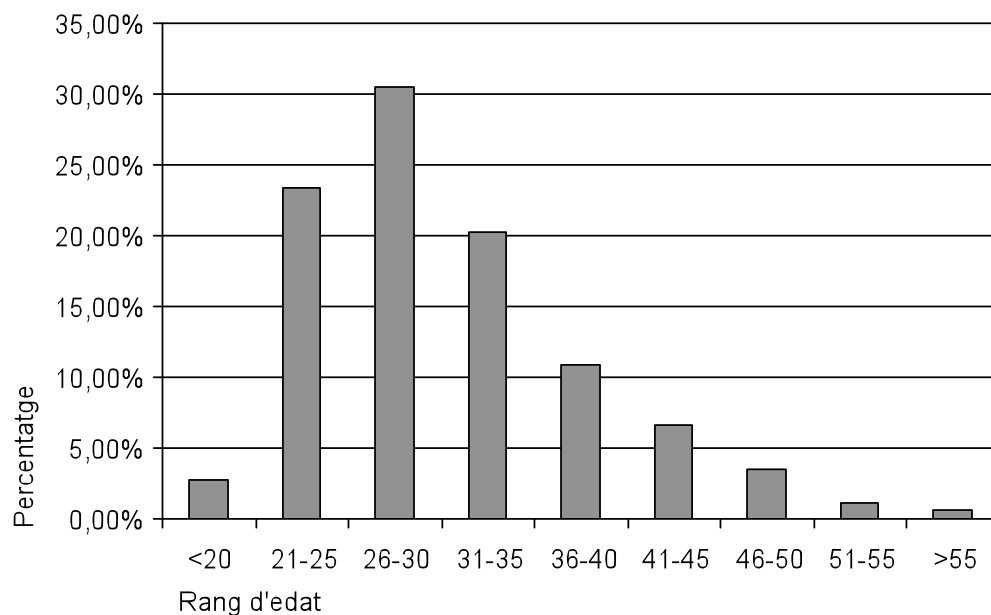


Figura A.1: Distribució de l'edat.

Pel que fa als motius per haver triat la UOC per a estudiar una titulació homologada:

<b>Motivació</b>	<b>%</b>
Millor gestió del temps	66,5%
Universitat moderna i innovadora	8,2%
Sistema d'avaluació	4,0%
Atenció docent personalitzada	3,2%
Prestigi	2,1%
NS/NC	16,0%

Taula A.4: Motivació per estudiar a la UOC.

<b>Motivació</b>	<b>%</b>
Per enriquiment personal	37,1%
Per ampliar coneixements per l'ocupació actual	27,3%
Per canviar de professió o de sector laboral	9,9%
Per promocionar dins de l'empresa	9,9%
NS/NC	15,8%

Taula A.5: Motivació per continuar estudiant a la UOC.

Finalment, altres variables rellevants pel que fa a la situació personal:

<b>Situació laboral</b>	<b>%</b>
Treball estable	53,8%
Treball eventual	8,9%
Per compte propi	4,9%
Aturat	2,8%
Altres	3,8%
NS/NC	25,8%

Taula A.6: Situació laboral.

<b>Estudis previs</b>	<b>%</b>
Llicenciatura o enginyeria	18%
Diplomatura o enginyeria tècnica	26,1%
Estudis universitaris no finalitzats	31,7%
1er cicle d'estudis universitaris	2,5%
COU sense PAAU	1,4%
PAAU	5,6%
Proves d'accés mes grans 25 anys	3,7%
FP 2n grau o Curs formatiu grau superior	10,2%
Doctorat	0,3%
Altres	0,5%

Taula A.7: Estudis previs.

<b>Comarca</b>	<b>%</b>
Barcelonès	30,26%
Baix Llobregat	9,39%
Bages	2,25%
Gironès	2,52%
Maresme	5,44%
Segrià	3,56%
Tarragonès	2,54%
Vallès Occidental	9,78%
Vallès Oriental	3,70%
Fora de Catalunya	14,04%
NSNC	17,52%

Taula A.8: Comarca de residència.



## **Annex B**

### **Llistat de tasques dels estudiants de la UOC**

En aquest annex es presenta el llistat de les diferents tasques que poden portar a terme els estudiants de la UOC. La llista s'ha elaborat tenint en ment als estudiants que porten més de dos semestres a la UOC i, per tant, coneixen l'entorn virtual d'aprenentatge i el funcionament de la universitat.

Les tasques que realitzen els estudiants tenen una temporalització molt clara, sempre vinculada al calendari de desenvolupament del semestre. Així, els períodes clau que s'identifiquen són: període de matrícula, període lectiu i període d'avaluació final. A continuació es presenta el llistat de tasques agrupades en aquests blocs:

#### **Període de matrícula**

- Consultar el pla d'estudis i el contingut de les possibles assignatures.
- Consultar dates d'exàmens de les assignatures a les quals es volen matricular per assegurar-se que les dates no coincideixen.
- Escollir assignatures i fer proposta de matrícula.
- Comunicar-se amb el tutor si aquest no està d'acord amb la proposta.
- Matricular-se.
- Demanar encadenament d'assignatures (per exemple, començar Anglès III i si s'aprofita només cal pagar els crèdits d'Anglès I i II però no cal fer la docència).

- Demanar convalidació de crèdits de lliure elecció d'altres cursos fets a la UOC.

### **Període lectiu**

- Familiaritzar-se amb el calendari del semestre i amb el pla docent de les diferents assignatures (és a dir, amb el que hauran de fer).
- Rebre els materials i, sinó, fer seguiment *online* de l'estat de la tramesa.
- Llegir i estudiar els materials didàctics.
- Fer seguiment de les bústies de l'aula: Taulers, Fòrums (hi ha estudiants que no els llegeixen) i Debats (si decideixen participar-hi o és obligatori).
- Realitzar activitats avaluable i no avaluable (descarregar-se documents, treballar-hi, lliurar documents, treballar en grup, etc).
- Consultar les notes de l'avaluació contínua.

### **Període d'avaluació final**

- Preparar el full personal d'exàmens i proves de validació.
- Anar presencialment a fer els exàmens o proves de validació.
- Consultar les notes.
- Demanar revisió si no s'està d'acord.

### **Altres tasques que poden dur a terme durant el semestre**

- Entrar a la bústia personal.
- Consultar l'expedient acadèmic.
- Consultar la biblioteca.
- Demanar convalidació d'estudis previs (AEPs).
- Fer seguiment de les activitats i ofertes de la comunitat.
- Participar en les activitats de comunitat.
- Accedir a la borsa de treball.

## Annex C

### Personatges i escenaris a la UOC

En aquest annex es presenta la definició de personatges i escenaris que s'ha utilitzat per identificar els nivells d'anàlisi a la UOC. En primer lloc es presenta la informació qualitativa a la que s'ha tingut accés, així com les tipologies d'estudiants que s'han identificat, i a continuació es defineixen els personatges i els seus escenaris.

#### C.1 Introducció

Els personatges i els escenaris són tècniques que apropen als usuaris i a les seves motivacions, objectius i situacions d'ús, i s'han presentat detalladament a la secció 2.x. En aquest treball s'utilitzen per entendre i analitzar els usuaris i l'ús que fan dels sistemes interactius.

Un personatge es construeix a partir de la informació quantitativa i qualitativa de la investigació d'usuaris: segmentació, perfilatge o modelat d'usuaris, observació, enquestes, entrevistes i *focus groups*. Per tant, primer cal investigar els usuaris, analitzar les dades obtingudes i finalment modelar els usuaris en personatges. A partir de les dues tipologies d'estudiants de la UOC que s'obtenen de l'anàlisi realitzada es construeixen dos personatges que constitueixen un model representatiu dels estudiants.

Els escenaris són descripcions d'un personatge en una situació d'ús del sistema, amb uns objectius concrets. A partir del conjunt de tasques dels estudiants de la UOC, que es presenta a l'annex B, es construeixen els escenaris pels dos personatges definits.

## **C.2 Dades de la UOC i dels seus estudiants**

El desenvolupament i modelatge d'usuaris s'ha realitzat a partir de la informació dels estudiants disponible en les bases de dades de la UOC. Es vol destacar especialment que en tot moment s'han consultat de manera anònima i posant el focus no en la identitat i informació personal dels estudiants sinó en aquells trets rellevants per modelar els personatges. Així s'ha considerat la següent informació:

- Dades d'ús del campus virtual, resultat dels primers processos de tractament de fitxer de logs.
- Dades quantitatives i de matrícula dels estudiants.
- Dades socio-demogràfiques de les bases de dades pròpies de la universitat.
- *Checklist* de tasques dels estudiants de la UOC.

Altra informació sobre estudiants prové dels documents de la investigació d'usuaris del projecte Campus en Programari Lliure [19]. Tot i que inclouen més estudiants que els propis de la UOC, presenten informació de gran valor per conèixer la tipologia i característiques dels estudiants en entorns virtuals d'aprenentatge. Aquests documents proporcionen informació sobre els estudis qualitatius portats a terme amb estudiants, concretament entrevistes en profunditat i *focus groups*. També s'han tingut en compte i han resultat de gran utilitat els documents elaborats per la pròpia universitat sobre el perfil i tipologia dels estudiants. Finalment cal destacar que en la elaboració dels personatges i dels escenaris s'ha comptat amb la experiència de l'autor d'aquesta tesi com a professor dels Estudis d'Informàtica, Multimèdia i Telecomunicació, dels quals disposa d'un coneixement en profunditat i en els que ha pogut accedir amb detall suficient a la informació dels seus estudiants.

Així, les característiques principals dels estudiants de la UOC es poden sumaritzar en aquests quatre aspectes:



- Tenen poc temps i van per feina; un cop matriculat, l'objectiu principal de l'estudiant és aprovar.
- La UOC és un element secundari en la seva vida, després de la feina i de la família.
- Estudiar a la UOC representa un cost a tots els nivells (econòmics, de temps d'oci, de dedicació a altres activitats, etc.).
- La mitjana d'assignatures per semestre que fa un estudiant és de 3.

Les dades que es consideren en aquest treball corresponen al semestre de primavera del curs 2006-2007 i als estudiants de les titulacions homologades en català. A l'annex A es presenta el sumari de les dades quantitatives i de matrícula que s'han considerat rellevants per a aquest treball. Els experiments realitzats en el capítol 4 corroboren aquestes premisses.

El *checklist* o llistat de tasques dels estudiants que es presenta a l'annex B constitueix una informació molt important per entendre als estudiants i el seu comportament, sobretot quan es connecten i naveguen pel campus virtual. El llistat de tasques que es presenta mostra de manera exhaustiva i esquemàtica totes les tasques que els estudiants poden portar a terme en els diferents moments en que es desenvolupa el semestre.

Dels estudis qualitatius s'extreu que hi ha un objectiu per estudiar a la UOC i és la millora tant personal com professional. Aquest objectiu que comparteixen la majoria dels estudiants deriva en dos grans blocs de motivacions a l'estudi. Les principals motivacions que es constaten per estudiar a la UOC són motius racionals i motius emocionals. Els motius racionals són aquells que persegueixen la obtenció d'un títol o l'adquisició d'unes habilitats o competències, majoritàriament orientades a l'àmbit professional; aquesta és la tipologia que inclou un major nombre d'estudiants. El segon bloc de motivacions inclou principalment l'estudi com a *hobby*, per ocupar el temps o com a repte personal. Aquests estudiants matriculen menys assignatures cada semestre que no pas els que estudien per motius racionals, que presenten un patró de maximitzar l'aprofitament del temps disponible i obtenir el títol el més aviat possible.

Per altra banda, el perfil psicogràfic obtingut en les entrevistes amb estudiants és que l'estudiant de la UOC es majoritàriament organitzat, disciplinat i constant, ja que el propi fet d'estudiar a la UOC l'obliga a ser-ho. Li agrada el que fa i és, en general, inquiet i autodidacta. Això està directament relacionat amb els objectius que persegueix la personalització, que pretén recolzar a l'estudiant en la presa de decisions durant el procés d'aprenentatge.

D'entre les motivacions per matricular-se a la UOC destaca la flexibilitat que la universitat proporciona en diferents àmbits. Flexibilitat en horaris, no perdre temps en desplaçaments i la facilitat de poder-se connectar des de qualsevol lloc. Flexibilitat també en el nombre d'assignatures a matricular i en el sistema d'avaluació, si es compara amb d'altres universitats i el fet que les assignatures estan més enfocades a aprendre que no pas a superar un examen, gràcies a la seva orientació i plantejament i al model d'avaluació continuada. Els estudiants destaquen com a aspecte positiu i important que permet compatibilitzar l'estudi amb el treball i, sobre tot, amb la família.

### **C.3 Tipologies i personatges**

La definició de personatges i escenaris s'ha portat a terme a partir de les dades presentades a la secció anterior. S'ha comptat amb estudis quantitius i qualitius sobre els estudiants. La naturalesa i complementaritat d'aquestes dades constitueix la matèria primera imprescindible per a la definició de tipologies d'estudiants i per a construcció dels personatges i escenaris.

De la investigació i estudi d'usuaris del Projecte Campus en Programari Lliure [19] s'obté que hi ha dues tipologies principals d'estudiants de la UOC. No és l'objectiu d'aquest annex presentar en profunditat l'estudi realitzat, però sí es destaca a continuació els trets que es consideren interessants. L'estudi es va portar a terme en dues fases, la primera seguint una metodologia qualitativa i, la segona, amb una metodologia quantitativa. En la fase qualitativa s'han realitzat entrevistes en profunditat i *focus groups*. En la fase quantitativa s'ha realitzat una enquesta en línia que s'ha enviat per correu electrònic als estudiants. Amb la informació obtinguda gràcies a la enquesta

es va portar a terme un anàlisi de clústers proporcionant resultats amb significança estadística. Així, les tipologies d'estudiants es van obtenir en base a les seves actituds davant la universitat i els estudis, així com els seus valors vitals. Aquestes tipologies es descriuen i caracteritzen posteriorment en funció de les actituds esmentades i també dels seus comportaments. Es parteix, doncs, de les actituds per explicar els comportaments donat que garanteix una segmentació operativa respecte les variables de comportament.

A partir d'aquest estudi s'identifiquen dues tipologies principals d'estudiants de la UOC. Aquestes tipologies identifiquen als estudiants en base als següents aspectes:

- Els trets en quant a característiques personals, actituds, caracterització acadèmica i caracterització en quant a la relació amb la tecnologia.
- Els comportaments, analitzant els mitjans a través dels quals es relacionen amb la universitat, els dispositius preferits, persones amb les que es relacionen, tasques que realitzen, satisfacció amb els serveis universitaris i millores i demandes que expressen.
- Ús i accés al campus virtual.

Les dues tipologies d'estudiants de la UOC reben el nom de “hobby” i “executius” respectivament. Aquests noms intenten representar la personalitat dels estudiants que componen cada tipologia. A continuació es presenta la fitxa per cada tipologia que inclou un eslògan o lema que els caracteritza, una descripció resumida i els aspectes personal, acadèmic i tecnològic.

## **Hobby**

“Gaudeixo estudiant, és un hobby per a mi”

- Dones.
- Mitjana d'edat al voltant dels 32 anys, caps de família.

### **Perfil Personal**

- Disciplinats i organitzats, persones amb les idees clares que saben què volen a la vida.

- Implicats i actius, els agrada comprometre's amb causes d'interès i aprofitar el temps al màxim.

#### Perfil Acadèmic

- No tenen una ambició professional de tipus materialista, de manera que els estudis no els enfoquen com a mitjà per obtenir un títol que els permeti aconseguir una feina determinada.
- Enfoquen els estudis com quelcom de complementari a la formació que ja tenen.
- Valoren els estudis per l'enriquiment que suposa aprendre.
- Porten els apunts al dia (són disciplinats) i valoren més el fet d'aprendre que no pas la nota mateixa, tot i ser els que menys assignatures repeteixen.

#### Perfil Tecnològic

- Són els que més valoren les noves tecnologies en tots els sentits.
- Malgrat això, no són els que més aparells electrònics tenen.

### **Executius**

“Necessito el títol per impulsar la meva carrera professional”

- Lleugerament més homes que dones.
- Entre 30 i 40 anys, casats.

#### Perfil Personal

- Són disciplinats i organitzats, persones amb les idees clares que saben què volen a la vida.
- No destaquen pel seu nivell d'implicació ni materialisme.

#### Perfil Acadèmic

- No els interessen especialment els estudis en el sentit d'enriquiment personal, sinó més aviat per la vessant utilitarista.
- Volen estudiar per aconseguir un títol, i no els importen les qualificacions sempre que puguin acabar el més aviat possible.

- Perceben els estudis que estan fent ara com a formació complementària d'aquella que ja posseeixen.

#### Perfil Tecnològic

- Reconeixen els avantatges de les tecnologies de la informació i la comunicació tant en l'àmbit professional com en l'àmbit acadèmic.
- Això no obstant, no els resulta especialment natural relacionar-s'hi.

Pel que fa a les actituds, els executius perceben els seus estudis com a complement de la formació que ja posseeixen. En canvi, el grup hobby els veuen com una font d'aprenentatge i enriquiment. En relació als valors vitals, el grup hobby es decanta cap a la implicació i dinamisme davant la vida, així com la disciplina, organització i creativitat. Per altra banda, el grup executiu no es decanta fortament cap a valors concrets, tot i que destaca per la seva necessitat de normes a què atènyer-se en contraposició a la indisciplina. En relació a la tecnologia el grup hobby s'hi relaciona amb més naturalitat que el executiu, i tot i que aquest reconeix més la utilitat, ambdós grups li veuen una utilitat acadèmica. D'entre les àrees de millora que proposen els dos grups es destaca la demanda de senzillesa i agilitat en el campus virtual, així com una relació més àgil amb el personal de la universitat per resoldre incidències. Al mateix temps, posen de manifest la complicació que els hi suposa fer treballs en grup i sol·liciten més materials didàctics en suport físic, per així no haver d'estar lligats a l'ordinador a l'hora d'estudiar-los, valoren poder-los transportar i utilitzar fàcilment en qualsevol lloc, ja sigui al transport públic o al sofà de casa seva.

La relació d'aquests dos grups amb el campus virtual presenta força trets en comú. Ambdós grups utilitzen Internet i el correu electrònic però es mostren reticents a usos avançats de la tecnologia. Necessiten una interfície d'accés clara, senzilla i usable i no mostren interès en accedir al campus mitjançant dispositius mòbils. Tot i així, el grup hobby es mostra més proper a la tecnologia i als dispositius que el grup executiu. El grup hobby mostra una predisposició per aprofitar al màxim les possibilitats educatives que ofereix el campus virtual, mostren interès per la seva carrera i també en els recursos educatius perifèrics als seus estudis. Per la seva banda el grup executiu no mostra una predisposició a fer un ús intensiu del campus sinó que es connecten i "van per feina";

valoren, per tant, una interfície que els hi proporcioni facilitat i senzillesa i no cerquen informació complementària més enllà del que els hi requereix els seus estudis. Ambdós grups valoren molt positivament el tenir accés amb facilitat al professorat i poder resoldre les qüestions acadèmiques, i troben a faltar la mateixa facilitat d'accés per resoldre els aspectes de gestió relacionats amb els seus estudis.

Els personatges que es proposen a continuació responen a les dues tipologies principals dels estudiants de la UOC, “hobby” i “executiu”. D'aquesta manera, s'han definit dos personatges que representen i personalitzen aquests dos perfils, caracteritzant així el conjunt d'estudiants de la UOC. A aquests personatges se'ls hi a donat el nom de Jaume i Olga.

El fet de tenir dos personatges permet considerar els estudiants de la universitat tenint en compte de manera més personalitzada les seves característiques, motivacions i objectius. També permet focalitzar els resultats i simplificar el procés de construcció d'escenaris. Per cada personatge es presenten els seus objectius, motivacions, necessitats, decisions d'estudi, dificultats, expectatives, preferències i relació amb la tecnologia.

## **Jaume**

El Jaume té 40 anys, està casat i té un fill petit. Està matriculat del segon cicle de la enginyeria en informàtica. Treballa d'informàtic en una població diferent de la que viu i cada dia s'ha de desplaçar en cotxe, dedicant-hi molt de temps. Estudia principalment els caps de setmana, tot i que també li dedica al menys dos vespres entre setmana dependent, sobre tot, de les dates de lliurament de les activitats. Li agrada la tecnologia i hi té una relació constant principalment en el seu exercici professional. Voldria aprofitar els desplaçaments en cotxe per avançar feina de la UOC, però la única manera de fer-ho seria disposar dels materials didàctics en format àudio i malauradament la UOC no els hi proporciona. Quan es connecta al campus virtual va molt per feina i només accedeix a les aules de les seves assignatures. Normalment llegeix el tauler del professor i, en canvi, el fòrum el llegeix en diagonal, destriant els missatges que creu que el poden

interessar, i molts dies ni tant sols el llegeix. Cada semestre es matricula del màxim d'assignatures que pot per així avançar a bon ritme i poder obtenir el títol aviat. Això l'ha portat a tenir una comunicació habitual i fluida amb la seva tutora. Està a punt d'obtenir el títol d'enginyer i això li representarà una important millora professional, al mateix temps que constitueix una fita personal important i està emocionat al respecte. Abans de fer el segon cicle ja va fer la enginyeria tècnica a la UOC i de moment no es tornarà a matricular a la UOC, ja que els estudis actuals li han suposat moltes renúncies i sacrificis i vol dedicar més temps a la família, tot i considera que ha estat una bona experiència i no descarta tornar a estudiar en el futur.

## **Olga**

L'Olga és economista i està estudiant humanitats. Es planteja els seus estudis com un complement a la seva formació i com a millora personal. Tot i que creu que també l'ajudaran a millorar les seves competències professionals, la motivació principal per estudiar és per passar-ho bé, ja que sempre havia volgut estudiar humanitats. Té 34 anys, no està casada i viu amb la seva parella. Estudia principalment als vespres i al cap de setmana, tot i que, alguns dies, depenent del moment del curs, dina ràpid per així poder avançar feina connectada al campus durant el temps del dinar. Està contenta de poder estudiar a la UOC i se sent satisfeta amb el que aprèn i amb les activitats que realitza. Quan acabi humanitats voldria continuar estudiant a la UOC i segurament es matricularia d'estudis asiàtics. Estudiar la fa sentir que millora professionalment però, sobre tot, personalment. Li agraden les tecnologies, les utilitza com a eina de treball i a casa de manera més lúdica amb l'ordinador i la càmera digital. També utilitza habitualment un telèfon mòbil amb pda. Li agradaria poder aprofitar per estudiar els desplaçaments amb metro quan va i torna a la feina tot i que normalment no ho fa degut a la mida dels materials que la UOC li envia a casa. És una persona inquieta i li interessen la cultura i l'actualitat.

## C.4 Escenaris, situacions i tasques que realitzen aquests personatges

La construcció d'escenaris donats uns personatges requereix incloure les accions que es porten a terme, tant *online* com *offline*, les frustracions i emocions dels personatges, les necessitats d'informació que tenen i les funcionalitats que utilitzen per resoldre les seves necessitats i objectius. Les tasques que realitzen els dos personatges definits són, principalment, les presentades a l'annex B. Els dos personatges realitzen les mateixes tasques tot i que les afronten de diferents maneres, ritme i profunditat. Tots dos persegueixen la obtenció del títol, tot i que, principalment, el Jaume té pressa per aconseguir-lo i la Olga té menys pressa i més inquietud per aprendre i gaudir del que aprèn.

- El Jaume fa la seva proposta de matrícula i espera a que la seva tutora la validi. És conscient que la elecció de les assignatures és un factor clau per poder avançar ràpidament per aconseguir el títol.
- El Jaume formalitza la matrícula d'aquell semestre.
- El Jaume es connecta el primer dia de curs per veure si té accés a l'aula i al pla docent, amb el calendari i planificació de tot el semestre. Al cap d'una estona es torna a connectar per llegir el missatge de benvinguda a l'aula i enviar el seu missatge de presentació.
- Al Jaume li preocupa el temps que li queda per acabar la titulació i envia una pregunta al tutor en relació al pla d'estudis.
- El Jaume es connecta a l'aula perquè sap que avui es publica l'enunciat d'una PAC, se'l descarrega i es posa a realitzar-la.
- El Jaume dedica cinc dies a resoldre la PAC. Durant aquests dies només es connecta al correu i a l'aula d'aquella assignatura per veure si el consultor ha donat indicacions per la seva realització.
- A l'estar matriculat de tantes assignatures, li agradaria tenir una eina o espai al campus, similar als que utilitza en la seva feina, que sigui com un quadre de comandament de les seves assignatures i activitats pendents.
- El Jaume accedeix al seu full personal d'exàmens per així decidir i comunicar a quin torn i a quina seu anirà a fer les proves finals.



- El Jaume envia una consulta a la secretaria sobre la convalidació d'unes assignatures que va cursar fa un temps.
- El Jaume demana informació als Estudis sobre els TFC disponibles. El cursarà el segon semestre d'aquell curs i per ell representa la darrera fita abans d'aconseguir el grau, i és molt important per a ell.

Pel que fa a l'altre personatge:

- L'Olga s'ha de matricular i dedica força temps a informar-se sobre les diferents assignatures. Durant aquest temps envia algunes consultes al seu tutor i fa la proposta de matrícula.
- L'Olga formalitza la matrícula i realitza el pagament del seu cost.
- Es connecta el primer dia de curs, a l'hora de dinar, des de l'ordinador del despatx. Llegeix els missatges de benvinguda i imprimeix els plans docents de les assignatures que ha matriculat.
- L'Olga està força satisfeta amb el campus virtual ja que li permet accedir a les aules i seguir les assignatures, malgrat això creu que el correu electrònic és poc àgil i flexible.
- Accedeix a la planificació de cadascun dels cursos matriculats per prendre nota de les dates clau, i així organitzar-se la feina que haurà de fer durant el semestre. Li agradaria poder sincronitzar el calendari del campus amb el calendari personal de la seva agenda electrònica.
- L'Olga es connecta al vespre des de casa seva per llegir les bústies de les aules. Després de llegir les activitats a les aules dedica 45 minuts abans d'anar a dormir a estudiar els materials didàctics.
- S'acaba el termini per fer l'aportació a un debat. Al migdia, l'Olga es connecta al campus virtual i llegeix les aportacions dels seus companys i comença a preparar la seva. Al vespre, des de casa accedeix a l'espai de debat per a enviar-la.
- Es connecta a l'aula i després accedeix a la biblioteca per consultar si troba recursos complementaris a la bibliografia recomanada.
- Es connecta al campus per veure si el consultor ha publicat les notes de la darrera PAC.

- L'Olga ha estat fora el cap de setmana i lliura tard una PAC. Prioritza el fer-la be i aprendre que no pas lliurar-la en el termini establert i per tant estudia detalladament els materials i realitza les activitats recomanades abans de lliurar-la.
- L'Olga accedeix a secretaria per sol·licitar un encadenament d'assignatures i també aprofita per consultar l'estat de la seva titulació progressiva.
- L'Olga actualitza el full personal d'exàmens i sol·licita rebre les notes al mòbil.

A partir de la definició d'aquests escenaris s'han analitzat les accions que porten a terme el Jaume i l'Olga i s'han buscat agrupacions en base al tipus d'acció, la connectivitat al campus virtual i el tipus d'objectiu.

L'objectiu de cada tasca té dues vessants depenent de la finalitat, la de la realització immediata i la relativa a si es tracta d'un subobjectiu d'un objectiu d'abast més general. Per exemple, l'objectiu immediat de sol·licitar informació sobre el TFC és enviar un missatge als Estudis sol·licitant la informació. Aquest objectiu forma part d'un més general que és finalitzar els estudis i obtenir el títol. Totes les tasques que requereixen connexió al campus virtual tenen aquesta doble vessant, la relativa la connexió al campus, que és el curt termini, i la relativa a l'objectiu global que pot ser en relació al semestre en curs o respecte la titulació o altres estudis que es volen realitzar.

Per tant, cada tasca té una finalitat que es situa en el nivell de la pròpia realització de la tasca i una altra finalitat que traspasa nivells. Si es realitza una agrupació de les tasques i els seus objectius tenint en compte això s'observen tres grans blocs. Bàsicament: sessió de treball o de connexió al campus virtual, semestre acadèmic i titulació.

Les activitats i objectius relacionats amb la sessió de treball tenen a veure amb connectar-se al campus virtual per portar a terme una acció concreta com pot ser accedir al tauler del professor, enviar un correu al consultor, consultar el pla docent, descarregar o lliurar una PAC, accedir a la biblioteca, llegir les aportacions d'un espai de debat o enviar una consulta a secretaria. En totes aquestes tasques l'element comú es que es realitzen mitjançant una connexió al campus virtual i navegant pels seus diferents espais

i serveis. Es tracta d'accions concretes i de durada determinada que els usuaris porten a terme i que el disseny del campus virtual i la seva navegació condicionen profundament.

Els objectius i tasques a nivell de curs o semestre acadèmic inclouen tots aquells aspectes relacionats amb la bona consecució del curs, des de matricular les assignatures a l'inici de curs fins a la realització de les proves finals i la consulta de les notes a final de curs, passant per totes aquelles que es realitzen durant el semestre. Les tasques concretes relacionades amb el desenvolupament del semestre com és accedir a l'aula, descarregar un enunciat de PAC, lliurar la realització d'una PAC, són tasques que es porten a terme en una sessió però que tenen un objectiu comú de superar l'assignatura i per tant són objectius de semestre o curs acadèmic que es porten a terme un diferents sessions individuals.

L'altre grup d'objectius i activitats està relacionat amb l'aprenentatge al llarg de la vida. Es té en compte les diferents matrícules realitzades en diferents semestres, l'èxit o fracàs en aquestes matrícules, és a dir, si es superen les assignatures, les matrícules futures, i la convalidació d'estudis previs. Per tant, en aquest bloc passa el mateix que en els blocs anteriors: les tasques concretes relacionades amb l'aprenentatge al llarg de la vida es realitzen a nivell de curs acadèmic i es porten a terme en sessions concretes de connexió al campus virtual.

Aquesta estructura jeràrquica dels objectius determina els nivells d'anàlisi que s'han considerat en aquesta tesi. És impossible relacionar les tasques puntuals que du a terme un estudiant quan es connecta al campus virtual amb l'objectiu a llarg termini, excepte si s'usa aquesta estructura multinivell. Tot i que seria possible estendre el nombre de nivells, el propi calendari semestral imposa una certa estructura que no es pot obviar, així com el comportament dels propis usuaris. Per exemple, seria lícit plantejar-se un quart nivell entre el nivell de sessió i el nivell de semestre, per exemple el nivell d'unitat d'aprenentatge o de mòdul, però aquest nivell no es correspon als hàbits i comportament d'un dels perfils detectats, l'anomenat "executiu".



## **Annex D**

### **Mapa de marques del campus virtual de la UOC**

L'objectiu d'aquest annex és, per una banda, presentar el detall de la proposta de mapa de marques integrades pel campus virtual de la UOC i, per l'altra banda, recollir com les marques proposades queden reflectides en les entrades dels fitxers de log. També es presenten els problemes que ha comportat aquesta proposta de marques i els aspectes puntuals que encara cal resoldre.

#### **D.1 Introducció**

En aquest annex es proposa un conjunt de marques per a integrar-lo dins l'entorn virtual d'aprenentatge de la UOC, per així permetre el seguiment de la navegació dels usuaris. El conjunt de marques i la seva integració al campus virtual que es proposa segueixen els principis descrits en la secció 3.5: independència de l'arquitectura, independència tecnològica, jerarquia, automatització, escalabilitat i consistència. La proposta del mapa de marques s'ha fet estudiant l'activitat dels usuaris i les tasques que han de portar a terme en el campus virtual. Aquestes tasques, descrites a l'annex B, es corresponen amb els nivells d'anàlisi i els experiments plantejats i descrits en el capítol 4.

D'aquesta manera, la proposta de marques es realitza per fer seguiment dels serveis i espais principals del campus virtual. La qüestió principal que es vol respondre és quins

espais del campus virtual visiten els estudiants i en quin ordre ho fan. L'altre àmbit que es vol observar és la navegació pels espais de les aules i l'ús de les bústies compartides.

Al fer seguiment dels serveis i espais principals es té oportunitat d'estudiar la plana d'inici i analitzar quina és la primera acció que realitzen els estudiants. Aquesta primera acció és important ja que molts cops constitueix l'indicador de les seves motivacions i objectius principals per a aquella connexió o sessió de navegació.

## **D.2 Mapa de marques**

La plana d'inici és la que els usuaris troben en primer lloc quan entren al campus virtual. Aquesta plana no és pròpiament un servei, sinó més aviat un agregador o portal d'accessos directes als serveis i espais principals del campus virtual. En aquesta plana d'inici apareixen els accessos directes a les aules i les seves bústies compartides, la bústia, agenda i les novetats i notícies que es destaquen. Aquests accessos directes es van considerar importants a partir d'un estudi d'usabilitat i de les necessitats dels usuaris. La informació que es presenta està customitzada per a cada tipus d'usuari i, així, alguns dels accessos directes poden canviar segons l'usuari i el seu tipus. A la figura D.1. es mostra la pantalla que troben els estudiants quan entren al campus virtual.

A partir dels estudis de tasques dels usuaris i tenint en compte els experiments que es volen realitzar, s'ha definit el diagrama de navegació que es presenta a la figura D.2. Aquest diagrama recull la navegació principal dels dos serveis més utilitzats i considerats més importants pels estudiants, la bústia i les aules. Com es pot veure, hi ha diferents maneres d'anar a uns mateixos espais o serveis, sobre tot a la bústia i als diferents elements de les aules. Hi ha altres serveis, com per exemple l'agenda, que també ofereixen diferents vies d'accés. Aquests enllaços s'ha decidit no monitoritzar-los inicialment, per així centrar la recerca en els serveis i espais principals. Per afavorir la llegibilitat, en el diagrama no s'han dibuixat totes les possibilitats de navegació; així per exemple, des de qualsevol punt de navegació es possible anar a la plana d'inici o a la bústia.

Figura D.1: Plana d'inici del campus virtual tal i com la veu l'estudiant.

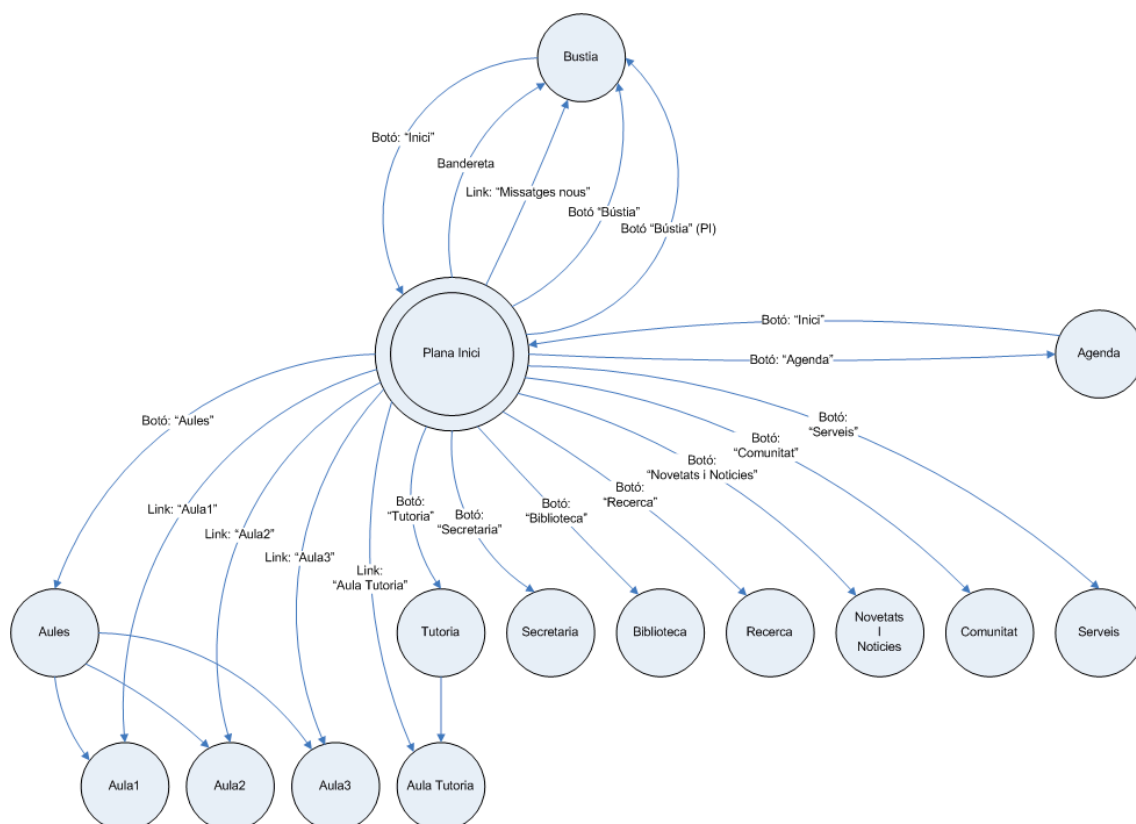


Figura D.2: Diagrama d'espais i enllaços principals del campus virtual.

Per monitoritzar una navegació com la presentada en la figura D.2 és necessari integrar marques en els diferents elements dels menús lateral i superior, així com en alguns dels accessos directes que apareixen a l'àrea de contingut de la plana d'inici. Les marques que es proposen per la botonera superior i lateral es presenten en les taules B.1 i B.2 respectivament. Per facilitar la seva comprensió també es presenten en forma de diagrames en la figura B.3.

Botó "Bústia"	&ACCIO=B_BUSTIA
Botó "Agenda"	&ACCIO=B_AGENDA
Botó "Fitxers"	&ACCIO=B_FITXERS
Botó "El meu perfil"	&ACCIO=B_PERFIL
Botó "Grups de Treball"	&ACCIO=B_GTREBALL
Botó "Preferits"	&ACCIO=B_PREFERITS
Botó "Servei d'atenció"	&ACCIO=B_ATENCIOE
Botó "Cerca de persones"	&ACCIO=B_CERCAP
Botó "Personalitza"	&ACCIO=B_PERSONALITZA
Botó "Surt"	&ACCIO=B_SURT

Taula D.1: Marques de la botonera superior.

Botó "Inici"	&ACCIO=B_INICI
Botó "Comunitat"	&ACCIO=B_COMUNITAT
Botó "Serveis"	&ACCIO=B_SERVEIS
Botó "Aules"	&ACCIO=B_AULES
Botó "Secretaria"	&ACCIO=B_SECRETARIA
Botó "Recerca-IN3"	&ACCIO=B_RECERCA
Botó "Biblioteca"	&ACCIO=B_BIBLIOTECA
Botó "Novetats i Notícies"	&ACCIO=B_NNOTICIES

Taula D.2: Marques de la botonera lateral.



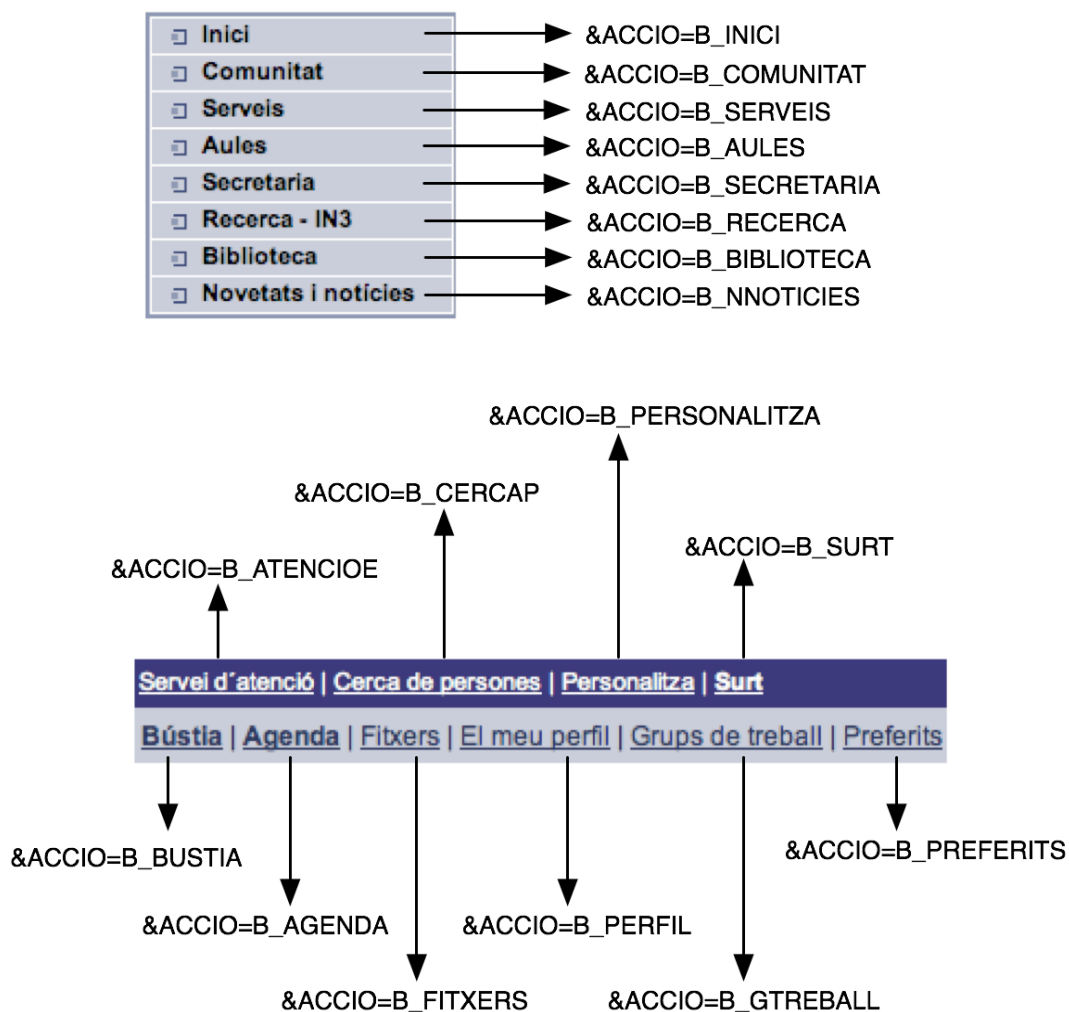


Figura D.3: Marques del menú del camp virtual.

Els altres enllaços a la bústia són els accessos directes de la plana d'inici i la bandera de notificació de missatges nous. Les marques corresponents són les de la Taula D.3.

Icona "Bustia"	&ACCIO=B_BUSTIA_PI
Enllaç "Missatges nous"	&ACCIO=L_MNOUS
Bandereta	&ACCIO=BANDERA

Taula D.3: Marques dels altres accessos a la bústia.

A la figura D.4 es poden veure gràficament els enllaços als que fan referència les marques.

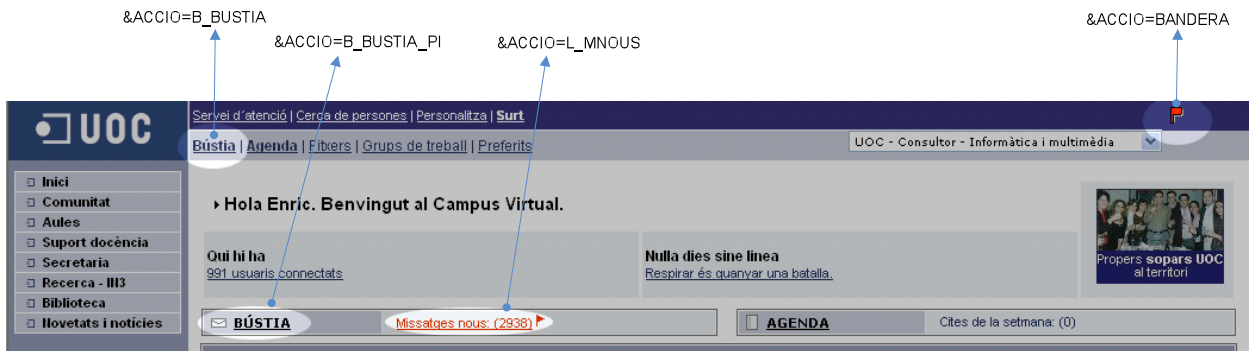


Figura D.4: Marques dels altres accessos a la bústia.

Fins a aquest punt s'han descrit les marques proposades per fer seguiment dels accessos i ús dels serveis i espais principals del campus virtual. Aquestes marques han seguit la estratègia orientada a l'acció descrita a la secció 3.5. A continuació es presenten les marques proposades per les aules i els seus elements i els accessos directes relacionats.

A cada estudiant li apareix en la plana d'inici del campus virtual i en la plana d'inici d'aules, una graella amb les aules de les assignatures matriculades. Aquesta graella presenta, per cada aula, un accés directe a l'aula que porta l'apartat de planificació i els accessos directes a les bústies compartides que hi ha a l'espai de comunicació de l'aula. Aquesta graella es construeix dinàmicament per a cada estudiant cada cop que es presenta per pantalla. Per aquest motiu, per a integrar les marques, s'ha utilitzat l'estratègia orientada al contingut, que també permet construir les marques dinàmicament. Les marques s'han introduït pels següents enllaços: accés directe a l'aula, accés directe al tauler, accés directe al fòrum i accés directe al debat.

Les marques definides s'han integrat al campus virtual pels diferents perfils d'usuari relacionats amb la docència de titulacions homologades en llengua catalana: estudiants, consultors, tutors i professors. Dificultats tècniques derivades de la implementació actual del campus virtual i del gran nombre de tipus d'usuaris diferents, han fet impossible introduir les marques per a tots els perfils. Per portar a terme l'anàlisi

proposat en aquest treball de tesi, el conjunt dels estudiants de les titulacions homologades constitueix el conjunt d'usuaris més gran del campus virtual i, per tant, estudiant aquest conjunt d'usuaris es poden obtenir resultats interessants. No obstant, en el futur seria desitjable poder introduir les marques per tots els perfils d'usuari del campus i així poder realitzar un anàlisi del campus en la seva totalitat.

### **D.3 Entrades generades per les marques als fitxers de log**

A continuació es mostren com apareixen en els fitxers de log les marques definides en la secció anterior. El registre de les marques als fitxers de log és fonamental per a que aquestes siguin útils per l'objectiu d'analitzar l'ús del sistema i la navegació dels usuaris. Per cada marca es mostra com queda registrada i, a més, amb "acció" s'indica l'acció o clic realitzat per l'usuari, i amb "càrrega" s'indica la marca d'haver executat l'acció.

#### **D.3.1 Login**

Acció:

`/cgi-bin/uocapp?s=num_sessió (i sense referer)`

Càrrega:

`trackingTag.gif?PAGE=/UOC/a/cgi-bin/hola`

que ha d'aparèixer sol

També:

`/UOC/a/ext_menu.html?s=num_sessió&img=hola`

`/UOC/a/cgi-bin/hola?s=num_sessió`

El directori /a/ és intercanviable per /b/ i (d'altres similars).

Altres entorns comencen per /EOB/a/ o /UOC2000/b/ i no s'han marcat.

## D.3.2 Menú Lateral

### 1. Botó Inici

Acció: &ACCIO=B\_INICI

```
/UOC/a/extcgi_0.html?s=num_sessió&ACCIO=B_INICI&img=hola&cgi=hola
```

Si busquem per &img=hola&cgi=hola tenim informació d'altres entorns i d'altres perfils.

Càrrega:

```
/trackingTag.gif?PAGE=/UOC/a/cgi-bin/hola
```

que ha d'aparèixer sol

També:

```
/UOC/a/ext_menu.html?s=num_sessió&img=hola
```

```
/UOC/a/cgi-bin/hola?s=num_sessió
```

El directori /a/ és intercanviable per /b/ i (d'altres similars).

Altres entorns comencen per /EOB/a/ o /UOC2000/b/ i no s'han marcat.

### 2. Botó Comunitat

Acció: &ACCIO=B\_COMUNITAT

```
/UOC/a/extern_0.html?s=num_sessió&ACCIO=B_COMUNITAT&img=ser_temp&htm=/club_uoc/comunitat/inici/index.html
```

Càrrega: no hi ha marca de càrrega.

### 3. Botó Serveis

Acció: &ACCIO=B\_SERVEIS

/UOC/a/extern\_0.html?s=num\_sessió&ACCIO=B\_SERVEIS&img=ser\_temp&htm=/club\_uoc/serveis/inici/index.html

Càrrega: no hi ha marca de càrrega.

#### 4. Botó Aules

Acció: &ACCIO=B\_AULES

/UOC/a/ext\_breakcam\_0.htm?s=num\_sessió&ACCIO=B\_AULES&img=aules&t=docencia/estudiant\_aules.tpl

Càrrega:

/trackingTag.gif?PAGE=/UOC/a/cgi-bin/hola

/trackingTag.gif?PAGE=/UOC/a/cgi-bin/hola&ACCIO=&PAGE=P\_INICIAULES

Les marques aquí tornen a estar mal introduïdes. La primera línia amb el trackingTag no hauria d'aparèixer, i en la segona s'ha forçat un element &ACCIO. En qualsevol cas, P\_INICIAULES proporciona la informació necessària.

#### 5. Botó Secretaria

Acció: &ACCIO=B\_SECRETARIA

/UOC/a/extern\_0.html?s=num\_sessió&ACCIO=B\_SECRETARIA&img=at\_secretaria&htm=/secretaria/index.html

Càrrega: no hi ha marca de càrrega

#### 6. Botó Recerca

Acció: &ACCIO=B\_RECERCA

/UOC/a/extern\_0.html?s=num\_sessió&ACCIO=B\_RECERCA&img=in3\_presentacio&htm=/recerca/index\_estudiants.html

Càrrega: no hi ha marca de càrrega.

## 7. Botó Biblioteca

Acció: &ACCIO=B\_BIBLIOTECA

/UOC/a/extern\_0.html?s=num\_sessió&ACCIO=B\_BIBLIOTECA&img=biblio&htm=/biblioteca/index.html

Càrrega: no hi ha marca de càrrega.

La biblioteca obre una finestra nova que es crida així:

[http://biblioteca.uoc.edu/login.html?nom=nom%20cognom1%20cognom2&dni=num\\_dni&mail=adreça\\_mail&entorn=nom\\_entorn&lang=idioma&tipus=tipus\\_usuari&subtipus=subtipus\\_usuari&idp=num\\_idp](http://biblioteca.uoc.edu/login.html?nom=nom%20cognom1%20cognom2&dni=num_dni&mail=adreça_mail&entorn=nom_entorn&lang=idioma&tipus=tipus_usuari&subtipus=subtipus_usuari&idp=num_idp)

## 8 Botó Novetats i Notícies

Acció: &ACCIO=B\_NNOTICIES

/UOC/a/extern\_0.html?s=num\_sessió&ACCIO=B\_NNOTICIES&img=novetats&htm=/novetats/index.html

Càrrega: no hi ha marca de càrrega.

### D.3.3 Menú superior

#### 1. Botó Bústia

Acció: &ACCIO=B\_BUSTIA

```
/UOC/a/cgi-bin/ma_mainMailFS?s=num_sessió&ACCIO=B_BUSTIA
```

Càrrega: no hi ha marca de càrrega.

No hi ha una marca de càrrega de bústia. Els logs de càrrega són més d'un, concretament 4, amb paràmetres. Amb aquests paràmetres s'indica si es tracta de la bústia personal o d'altres bústies del campus. Aquestes són les quatre línies interessants:

```
/UOC/a/cgi-bin/ma_buttons?s=num_sessió&aula=&l=pers&pm=&showTitle=&iforum=&bc_domainId=&bc_preferits=&bc_buscar=&bc_orden=&bc_portada=
```

```
/UOC/a/cgi-bin/ma_filter?s=num_sessió&aula=&l=pers&f=$FOLDER_REBUTS$
```

```
/UOC/a/cgi-bin/ma_folders?s=num_sessió&aula=&l=pers&setrebuts=true
```

```
/UOC/a/cgi-bin/ma_mssgs?s=num_sessió&aula=&l=pers&f=$FOLDER_REBUTS$
```

## 2. Botó Agenda

Acció: &ACCIO=B\_AGENDA

```
/UOC/a/extern_0.html?s=num_sessió&ACCIO=B_AGENDA&img=blanc&htm=http://cv.uoc.edu/webapps/Agenda/NavigationServlet?s=num_sessió
```

Càrrega: no hi ha marca de càrrega.

Es carrega la següent línia i, en qualsevol cas, no significa que l'agenda s'acabi de carregar del tot:

```
/webapps/Agenda/NavigationServlet?s=num_sessió
```

## 3. Botó d'Àrea de Fitxers

Acció: &ACCIO=B\_FITXERS

/webapps/filearea/servlet/iuoc.fileserver.servlets.LoginFilearea?sessionId=num\_sessió&company=/UOC&lang=/a&operation=login&ACCIO=B\_FITXERS

Càrrega: no hi ha marca de càrrega.

#### 4. Botó Perfil

Acció: &ACCIO=B\_PERFIL

/UOC/a/extern\_0.html?s=num\_sessió&ACCIO=B\_PERFIL&img=directori&htm=/club\_uoc/comunitat/el\_meu\_perfil/index.html

Càrrega: no hi ha marca de càrrega.

#### 5. Botó Grups de Treball

Acció: &ACCIO=B\_GTREBALL

/UOC/a/ext\_breakcam\_0.htm?s=num\_sessió&t=grups\_tb/grups.tpl&domainFather=grc&ACCIO=B\_GTREBALL

/UOC/a/cgi-bin/hola?s=num\_sessió&tpl=p://cv.uoc.edu/UOC/a/ext\_breakcam\_0.htm?s=num\_sessió&t=grups\_tb/grups.tpl&domainFather=grc&ACCIO=B\_GTREBALL

Càrrega: no hi ha marca de càrrega.

#### 6. Botó Preferits

Acció: &ACCIO=B\_PREFERITS

/UOC/a/extcgi\_0.html?s=num\_sessió&ACCIO=B\_PREFERITS&img=bookmark&cgi=bookmarkMainFS

Càrrega: no hi ha marca de càrrega.



## 7. Botó Servei d'Atenció

Acció: &ACCIO=B\_ATENCIOE

```
/UOC/a/extern_0.html?s=num_sessió&ACCIO=B_ATENCIOE&img=ajuda&htm=../../  
./CAU/js/cau.html
```

Càrrega: no hi ha marca de càrrega.

## 8. Botó Cerca de persones

Acció: &ACCIO=B\_CERCAP

```
/UOC/a/extern_0.html?s=num_sessió&ACCIO=B_CERCAP&img=directori&htm=htt  
p://cv.uoc.edu/webapps/comuni/cercaContextualServlet?jsp=/jsp/cercaCon  
textual/cercaBasica.jsp&activeView=&l=/a&s=num_sessió
```

Càrrega: no hi ha marca de càrrega.

## 9. Botó Personalitza

Acció: &ACCIO=B\_PERSONALITZA

```
/UOC/a/workarea_0.html?s=num_sessió&ACCIO=B_PERSONALITZA&img=prefer&ht  
m=/prefer/index.html
```

Càrrega: no hi ha marca de càrrega.

## 10. Botó Surt

Acció: &ACCIO=B\_SURT

```
/cgi-
bin/byebye?s=num_sessió&ACCIO=B_SURT&alert=false&text=Segur%20que%20vo
leu%20sortir%20del%20Campus? HTTP/1.1
```

Càrrega: no hi ha marca de càrrega.

### D.3.4 Graella d'aules

La graella d'aules de la plana d'inici funciona amb el sistema trackingTag. Les marques corresponents a clicar sobre un enllaç de tauler, fòrum, debat, etc, agafen el nom del propi recurs així, per exemple, si el recurs es diu "Tauler", la marca serà "&ACCIO=Tauler".

#### 1. Nom Assignatura/Aula

Acció: ACCIO=plan.jsp

```
/webapps/classroom/031_common/index.jsp?s=num_sessió&domainCode=codi_a
ula&img=aules
```

```
/webapps/classroom/032_common/plan.jsp?s=num_sessió&ACCIO=plan.jsp&dom
ainId=144400&domainFatherId=143376&idLang=a&idLangAula=a&userId=ES
TUDIANT&appId=UOC&domainFatherId=143376&domainCode=codi_aula&domainTem
plate=uoc_061_00.002
```

Exemple codi\_aula: 061\_04\_141\_03

Aquesta codificació té el següent significat:

codi semestre: 061

codi assignatura: 04\_141

codi aula: 03

Càrrega:

```
/trackingTag.gif?PAGE=/webapps/classroom/032_common/plan.jsp
```

Aquesta és la marca de càrrega sempre que s'entra a una aula i es carrega directament la pestanya de planificació.

## 2. Enllaç: Tauler

Acció: &ACCIO=Tauler

```
/UOC/a/cgi-bin/ma_mainMailFS?s=num_sessió&e=1&l=codi_tauler
```

Exemple codi\_tauler : 061\_04\_141\_03\_r01

codi semestre: 061

codi assignatura: 04\_141

codi aula: 03

codi tipus bústia: r01, on 01 és tauler, 02 és fòrum i 03 és debat.

Càrrega:

```
/trackingTag.gif?PAGE=/UOC/a/cgi-bin/hola&ACCIO=Tauler
```

Cal anar amb compte amb aquesta marca de càrrega, el sistema del trackingTag concatena les diferents accions, així doncs, si primer s'ha fet clic al fòrum, s'obté una línia com aquesta:

```
/trackingTag.gif?PAGE=/UOC/a/cgi-bin/hola&ACCIO=Fòrum&ACCIO=Tauler
```

Si primer s'accedeix al fòrum i després al debat, s'obté:

```
/trackingTag.gif?PAGE=/UOC/a/cgi-bin/hola&ACCIO=Fòrum&ACCIO=Debat&ACCIO=Tauler
```

El mateix passa si cliquem dos cops sobre el tauler:

```
/trackingTag.gif?PAGE=/UOC/a/cgi-bin/hola&ACCIO=Tauler&ACCIO=Tauler
```

Això cal tenir-ho en compte a l'hora d'analitzar aquestes accions i caldria resoldre-ho per simplificar el procés.

### **3. Enllaç: Fòrum**

Acció: &ACCIO=Fòrum

/UOC/a/cgi-bin/ma\_mainMailFS?s=num\_sessió&e=1&l=061\_04\_141\_03\_r02  
HTTP/1.1

codi semestre: 061

codi assignatura: 04\_041

codi aula: 03

codi fòrum: r02

Càrrega:

/trackingTag.gif?PAGE=/UOC/a/cgi-bin/hola&ACCIO=Fòrum

## **Annex E**

### **Resum d'*scripts* i procediments de captura i processament de fitxers de log**

En aquest annex es presenta el llistat de *scripts* i procediments que s'han implementat per portar a terme la gestió de fitxers de log, el seu processat i l'obtenció de camins de navegació.

#### **E.1 Introducció**

La recollida, gestió i processament dels fitxers de log involucra nombrosos *scripts* i procediments, així com dos servidors del grup de recerca on s'ha realitzat aquesta tesi: einlinux3.uoc.edu i personal.uoc.edu. El detall del tractament dels fitxers de log s'ha descrit a la secció 4.5 i, per la seva banda, el procés d'obtenció de camins de navegació s'ha presentat amb detall a la secció 4.6.2. Per aquest motiu, a continuació es presenta el llistat de procediments i *scripts* a cada servidor i una breu descripció de la seva localització i funcionament. Tenint en compte els objectius d'aquesta memòria de tesi doctoral, i per no allargar innecessàriament l'annex, s'ha considerat interessant no incloure el codi dels procediments i exposar-los mitjançant el resum que es presenta a continuació.

## E.2 Servidor: einflinux3.uoc.edu

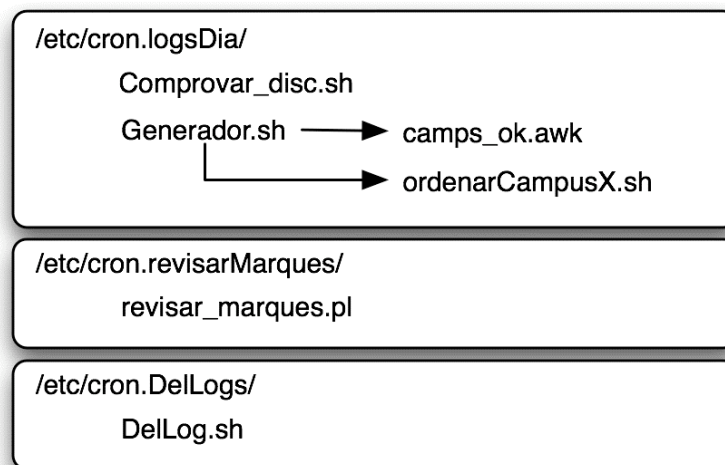


Figura D.1: Localització i dependències dels *scripts* al servidor einflinux3.uoc.edu

### Generador.sh

Ubicació: /etc/cron.logsDia/

Aquest *script* està programat per que s'executi cada dia a les 6:01 de matí (fitxer Crontab). S'encarrega de filtrar els logs que es reben diàriament en format .gz segons els paràmetres especificats en el fitxer “marques” del directori “/etc/cron.logsDia/” i genera un fitxer del tipus “CampusX.dia” amb els resultats, el qual es guarda en el directori “/extra/webminer/incoming/”. Aquest script crida a “ordenarCampusX.sh” i a “camps\_ok.awk”.

### ordenarCampusX.sh

Ubicació: /etc/cron.logsDia/

Ordena per data i hora els fitxers “CampusX” generats per “Generador.sh” a partir dels fitxers de log dels diferents *frond-ends*. Aquest fitxer s'ha d'ordenar donat que es crea unint els diferents fitxers de log sense mantenir la ordenació per data i hora.

### camps\_ok.awk

Ubicació: /etc/cron.logsDia/

Elimina les línies de log que no interessin, com ara, les generades pels elements de format gràfic i que carreguen imatges, fulls d'estil, etc.

**revisar\_marques.pl**

Ubicació: /etc/cron.revisarMarques/

Comprova que als fitxers “CampusX” creats per “Generador.sh” a partir dels fitxers de log dels diferents *front-ends*, no els hi falti cap marca de les introduïdes. Aquestes marques estan registrades en un arxiu en el mateix directori que s’anomena *accio\_marques*. Aquest procediment és necessari donat que, en algunes ocasions, al fer manteniment del campus virtual han eliminat per error marques integrades. Aquest script permet detectar aquest fet de manera automàtica.

**comprovar\_disc.sh**

Ubicació: /etc/cron.logsDia/

Comprova a diari l’espai de disc disponible i en cas que sigui inferior a un llindar preestablert (actualment 2 GB) envia un correu d’avís. Aquesta comprovació és necessària pel gran volum dels fitxers de log que es reben de manera automàtica i la seva variabilitat depenent del dia del semestre.

**DelLog.sh**

Ubicació: /etc/cron.delLogs/

Eliminar els fitxers de log comprimits i emmagatzemats de més de 30 dies d’antiguitat

**E.3 Servidor: personal.uoc.edu****ProcessarLogs.sh**

Ubicació: /home/webminer/bin

Aquest és el procediment principal per l’obtenció dels camins de navegació. Selecciona els fitxers “CampusX” i “loginhistory” necessaris per a portar a terme el procés; primer selecciona les línies que estan entre les hores especificades, elimina les línies duplicades, i extreu el “uid”, el numero de sessió i el “idp” de cada usuari. Aquest procediment crida a “proces.awk”.

**proces.awk**

Ubicació: /home/webminer/bin

Extreu informació sobre l'usuari com ara el número de vegades que es connecta (en el període de temps especificat), l'hora de cada connexió i controla el tipus d'usuari. Aquest procediment crida a "paths.awk".

**paths.awk**

Ubicació: /home/webminer/bin

Obté totes les marques (B\_INICI,B\_BUSTIA,etc...) d'un mateix "uid", les concatena i obté l'entorn de l'usuari si està disponible.

**espaiDisc.sh**

Ubicació: /etc/cron.EspaiDisc/

Comprova l'espai de disc disponible en el servidor. En cas que el percentatge d'espai ocupat en el disc sigui igual o superior a 80%, envia un correu d'avís.

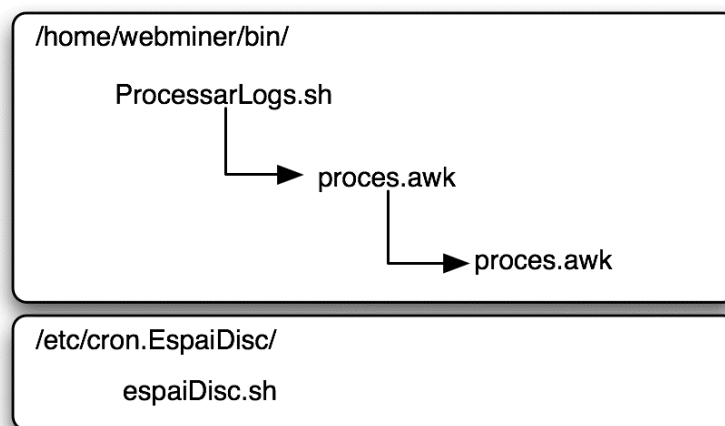


Figura D.2: Localització i dependències dels *scripts* al servidor personal.uoc.edu



## Annex F

### Aportacions originals

L'aportació fonamental d'aquest treball de tesi és una metodologia per a conèixer als usuaris, la seva navegació i l'ús que fan d'un entorn virtual que inclou una proposta d'anàlisi a tres nivells i una proposta de marques de seguiment. Això, ha permès fer una fotografia del campus virtual i dels estudiants de la UOC que proporciona dades i informació de la que mai abans s'havia disposat.

Un dels objectius del treball realitzat ha estat obtenir un seguit de publicacions dels quals es destaquen les següents. Aquestes es poden agrupar en diferents àmbits com, per exemple, anàlisi d'usuaris i navegació, personalització i disseny centrat en l'usuari i usabilitat.

- Mor, Enric, Minguillón, Julià; Garreta-Domingo, Muriel; Lewis, Sheena. "A Three-Level Approach for Analyzing User Behavior in Ongoing Relationships". Lecture Notes in Computer Science (LNCS). Springer-Verlag Editions. Proceedings of the 12th International Conference on Human-Computer Interaction HCII'07, Beijing China, 22-27 July 2007
- Garreta-Domingo, Muriel; Almirall-Hill, Magí; Mor, Enric. "A Game to Promote Understanding about UCD Methods and Process". Lecture Notes in Computer Science (LNCS). Springer-Verlag Editions. Proceedings of the 12th International Conference on Human-Computer Interaction HCII'07, Beijing China, 22-27 July 2007

- Huertas, Maria Antonia; Casado, Carlos; Córcoles, César; Mor, Enric; Guerrero-Roldán, Ana Elena. "Social Networks for learning: wikis, blogs and tagging in education" Proceedings of the EDEN 2007 Annual Conference, Naples Italy, June 2007
- Garreta-Domingo, Muriel; Mor, Enric. "User Centered Design in E-Learning Environments: from Usability to Learner Experience". Proceedings of the EDEN 2007 Annual Conference, Naples Italy, 13-16 June 2007
- Garreta-Domingo, Muriel; Almirall-Hill, Magí; Mor, Enric. "User-Centered Design Gymkhana". Extended Abstracts of the SIGCHI conference on Human Factors in computing systems CHI'07, San Jose USA, May 2007
- Mor, Enric; Minguillón, Julià; Santanach, Francesc. "Capturing user behavior in e-learning environments". Proceedings of the 3rd International Conference on Web Information Systems and Technologies, Barcelona, Spain, 3-6 March 2007
- Mor, Enric; Minguillón, Julià; Carbó, Josep Maria. "Analysis of User Navigational Behavior for E-Learning Personalization", in Romero C. & Ventura S. (Eds.) Data Mining in E-Learning, pp. 227-246, WIT Press 2006
- Casado, Carlos; Mor, Enric. "Accesibilidad en el aprendizaje virtual". Actas del 3r Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables, Oviedo, España, 25-27 Septiembre 2006
- Santanach, Francesc; Garreta, Muriel; Griset, Roger; Llorenç, Laia; Mor, Enric; Minguillón, Julià. "Materiales Interactivos Reutilizables". Actas del 3r Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables, Oviedo, España, 25-27 Septiembre 2006
- Guerrero-Roldán, Ana Elena; Mor, Enric; Minguillón, Julià. "Proceso de adaptación al EEES mediante el uso de competencias e itinerarios personalizados: el caso de Minería de Datos". Actas de las XXI Jornadas de Enseñanza Universitaria de la Informática, Bilbao Spain, July 2006
- Guerrero-Roldán, Ana Elena; Minguillón, Julià; Mor, Enric. "From contents to competences through learning design: the case of data mining". Proceedings of the EDEN 2006 Annual Conference, Vienna, July 2006
- Casado, Carlos; Mor, Enric; Sabaté, Llorenç. "Accesibilidad en la web". MOSAIC: tecnologías y comunicación multimedia. ISSN: 1696-3296

- Ferran, Núria; Mor, Enric; Minguillón, Julià. "Towards personalization in digital libraries through ontologies", a Library Management Journal, vol.25, núm 4/5, pp. 206-217. ISSN 0143-5124.
- Carbó, Josep Maria; Mor, Enric; Minguillón, Julià. "User Navigational Behavior in e-Learning Virtual Environments". Proceedings of the ACM International Conference on Web Intelligence, pp. 243-249, Compiègne, France, 19-22 September 2005.
- Mor, Enric; Minguillón, Julià; Gil, Eva Patricia. "Una aproximación metodológica para el análisis cuantitativo y cualitativo de usuarios de entornos virtuales de aprendizaje". Actas del VI Congreso Internacional de Interacción Persona Ordenador, Granada, Spain, September 2005
- Minguillón Julià, Mor Enric, Santanach Francesc and Guàrdia Lourdes " Personalización del proceso de aprendizaje usando learning objects reutilizables". Actas del 1r Simposio Pluri-disciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables. Guadalajara, España, Octubre 2004
- Mor Enric and Minguillón Julià " E-learning Personalization based on Itineraries and Long-term Navigational Behavior". Proceedings of the Thirteenth World Wide Web Conference, Volume 2, pp 264--265, New York City, NY, May 2004
- Mor Enric and Minguillón Julià " Patrones de navegación de usuarios de un campus virtual". Actas del V Congreso Internacional de Interacción Persona Ordenador, pp 109-112, Lleida, España, Mayo 2004
- Mor Enric and Minguillón Julià " An empirical evaluation of classifier combination schemes for predicting user navigational behavior". Proceedings of the Information Technology: Coding and Computing ITCC'2003, pp 467--471, Publisher IEEE Computer Society.
- Mor Enric and Minguillón Julià " Predicción de la navegación del usuario utilizando árboles de decisión". Actas del IV Congreso Internacional de Interacción Persona Ordenador. Vigo, España, Junio 2003

Entre les aportacions significatives addicionals que sorgeixen com a resultat del procés d'investigació, es pot destacar la participació en comitès així com la l'utilització de

resultats en diferents projectes tant de usabilitat i disseny centrat en l'usuari com d'e-learning. Es destaquen les següents:

Projectes amb finançament extern:

- Projecte A-MEDIDA. Desarrollo de un sistema de transformación de contenidos a medida de los usuarios. FIT-350201-2004-6. Data d'inici:1/7/2004. Data de fi: 31/12/2005
- Projecte Campus en Programari Lliure. Data d'inici: 01/01/2005. Data de fi: 31/12/2007.
- Projecte SUMA: Desarrollo de un sistema de integración de servicios de elearning usando estándares. Data d'inici: 01/01/2007. Data de fi: 31/12/2007.
- Projecte PERSONAL: Personalización del Proceso de Aprendizaje en Entornos Virtuales mediante Itinerarios Formativos Adaptativos basados en Objetos de Aprendizaje Reutilizables y Ontologías. TIN2006-15107-C02-01/02. Data d'inici: 01/10/2006. Data de fi: 30/09/2009.
- Projecte Accessibilitat, Usabilitat i Mobilitat en la banca *online*. Data d'inici: 01/01/2008. Data de fi: 31/12/2008

Participació en comitès científics:

- Comité de Programa de: HCI 2007 Workshop: Design, Use and Experience of E-Learning Systems. Depenent de: The British HCI Group. Setembre de 2007.
- Comité de Programa de: International Workshop on Applying Data Mining in e-Learning (Part of the 2nd European Conference on Technology Enhanced Learning (EC-TEL 2007). Depenent de: International Working Group on Educational Data Mining. Setembre de 2007
- Comité de Programa de: SPDECE 2005, II Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables. Depenent de: REDAOPA. Red Temática de Actividades y Objetos para el Aprendizaje. Octubre de 2005
- Comité de Programa de CHIJOOTE 2005 , I Jornadas de Trabajo sobre Enseñanza de CHI. Depenent de: AIPO. Asociación Interacción Persona-Ordenador. Juliol de 2005.

- Comité de Programa de: Interacción 2004, V Congreso Español de Interacción Persona-Ordenador. Dependent de: AIPO. Asociación Interacción Persona-Ordenador. Maig de 2004.

Projectes i iniciatives internes de la UOC:

- Grup d'experiència d'usuari de la UOC. Aquest grup va néixer de la necessitat de millorar la usabilitat del campus virtual i treballa en el disseny dels diferents aspectes de la experiència d'usuari.
- Projecte de millora materials didàctics (MyWay). Aquest projecte persegueix l'adequació del format dels materials didàctics a les necessitats específiques dels estudiants.
- Projecte aula. Aquest projecte consisteix en el redisseny de l'aula virtual tenint en compte les necessitats d'estudiants i docents en el context de l'Espai Europeu d'Educació Superior.
- Campus en programari lliure. L'objectiu d'aquest projecte és la construcció d'un campus virtual de codi obert seguint un enfocament centrat en l'usuari. Contempla especialment els aspectes d'usabilitat, accessibilitat i la personalització de continguts.
- Projecte de plana d'inici del campus virtual. Com el seu nom indica, en aquest projecte es treballa el redisseny de la plana d'inici dels usuaris del campus virtual, amb l'objectiu de presentar la informació adequada a cada tipus d'usuari i de facilitar elements de personalització.
- Anàlisi d'abandonament. En aquest projecte s'estudien les qüestions relacionades amb l'abandonament dels estudis per part dels estudiants.
- Projecte "*Joy of e-learning*". Aquest projecte persegueix identificar tots aquells elements que afavoreixen l'experiència educativa positiva i eliminar tots els elements de frustració en l'entorn virtual d'aprenentatge.



## Bibliografia

1. CRPITS'14: Proceedings of the IEEE international conference on Privacy, security and data mining. 2002.
2. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. 1990.
3. Advanced Distributed Learning, <http://www.adlnet.gov/> (Consulta: 14/02/2008)
4. Apache Software Foundation, <http://www.apache.org/> (Consulta: 14/02/2008)
5. Ariadne, <http://www.ariadne-eu.org/> (Consulta: 14/02/2008)
6. ASTD (American Society for Training & Development), <http://www.astd.org/> (Consulta: 14/02/2008)
7. Cisco Systems, Inc, <http://www.cisco.com/> (Consulta: 14/02/2008)
8. ClickTracks Analytics, <http://www.clicktracks.com/> (Consulta: 14/02/2008)
9. Google Analytics, <http://www.google.com/analytics/> (Consulta: 14/02/2008)
10. IBM Ease of Use - IBM Design, <http://www.ibm.com/easy> (Consulta: 14/02/2008)
11. IEEE - Institute of Electrical and Electronics Engineers, <http://www.ieee.org/> (Consulta: 14/02/2008)
12. IMS Global Learning Consortium, <http://www.imsglobal.org/> (Consulta: 14/02/2008)
13. IndexTools, <http://www.indextools.com/> (Consulta: 14/02/2008)
14. ISO - International Organization for Standardization, <http://www.iso.org/> (Consulta: 14/02/2008)

15. Logging in W3C httpd: Common Log Format, <http://www.w3.org/Daemon/User/Config/Logging.html#common-logfile-format> (Consulta: 14/02/2008)
16. LTSC - IEEE Learning Technology Standards Committee, <http://ieeeltsc.org/> (Consulta: 14/02/2008)
17. Open Knowledge Initiative, <http://www.okiproject.org/> (Consulta: 14/02/2008)
18. Pathalizer: Visual website usage analysis, <http://pathalizer.sourceforge.net/> (Consulta: 14/02/2008)
19. Projecte Campus en Programari Lliure, <http://www.campus.cat/> (Consulta: 14/02/2008)
20. sakaiproject.org Collaboration and Learning Environment for Education, <http://sakaiproject.org/> (Consulta: 14/02/2008)
21. SPSS Clementine, <http://www.spss.com/clementine/> (Consulta: 14/02/2008)
22. UOC - Universitat Oberta de Catalunya, <http://www.uoc.edu/> (Consulta: 14/02/2008)
23. W3C - World Wide Web Consortium, <http://www.w3c.org/> (Consulta: 14/02/2008)
24. World Wide Web Consortium, <http://www.w3.org/> (Consulta: 14/02/2008)
25. Abras, C., Maloney-Krichmar, D. and Preece, J. 2004. User-Centered Design. In *Encyclopedia of Human Computer Interaction*, W. Bainbridge, Ed. Sage Publications, Thousand Oaks
26. ADL. 2004. Sharable Content Object Reference Model (SCORM) 2004.
27. Adomavicius, G. and Tuzhilin, A. Feb 2001. Using data mining methods to build customer profiles. *Computer* 34, 74-82.
28. Adomavicius, G. and Tuzhilin, A. 2001. Expert-Driven Validation of Rule-Based User Models in Personalization Applications. *Data Mining and Knowledge Discovery* 5, 33-58.
29. Agrawal, R., Imielinski, T. and Swami, A. 1993. Mining association rules between sets of items in large databases. *SIGMOD Record* 22, 207-216.
30. Alessi, S.M. and Trollip, S.R. 1985. Computer-based instruction : methods and development. Prentice-Hall, Englewood Cliffs, NJ.
31. Ardito, C., Costabile, M., Marsico, M., Lanzilotti, R., Levialdi, S., Roselli, T. and Rossano, V. 2006. An approach to usability evaluation of e-learning applications. *Universal Access in the Information Society* 4, 270-283.



32. Ardito, C., Marsico, M.D., Lanzilotti, R., Levialdi, S., Roselli, T., Rossano, V. and Tersigni, M. 2004. Usability of E-learning tools. In *AVI '04: Proceedings of the working conference on Advanced visual interfaces*, Gallipoli, Italy, ACM Press, 80-84.
33. Avgeriou, P., Papasalouros, A., Retalis, S. and Skordalakis, M. 2003. Towards a Pattern Language for Learning Management Systems. *Educational Technology & Society* 6, 11-24.
34. Bates, T. 2000. *Managing Technological Change: Strategies for College and University Leaders*. Jossey-Bass, San Francisco.
35. Bevan, N. 2003. UsabilityNet Methods for User Centred Design. In *Human-Computer Interaction: theory and Practice. Proceedings of HCI International 2003*, Crete, Greece, Lawrence Erlbaum, 434-438.
36. Breese, J.S., Heckerman, D. and Kadie, C. 1998. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, 43-52.
37. Breiman, L. 1996. Bagging predictors. *Machine Learning* 24, 123-140.
38. Breiman, L. 1993. *Classification and regression trees*. Chapman & Hall, New York, N.Y.
39. Brusilovsky, P. 2003. Adaptive and Intelligent Web-based Educational Systems. *International Journal of Artificial Intelligence in Education* 13, 159-172.
40. Brusilovsky, P. 2001. Adaptive Hypermedia. *User Modeling and User-Adapted Interaction* 11, 87-110.
41. Brusilovsky, P. 1996. Methods and Techniques of Adaptive Hypermedia. *User Modeling and User-Adapted Interaction* 6, 87-129.
42. Brusilovsky, P. and Millán, E. 2007. User Models for Adaptive Hypermedia and Adaptive Educational Systems. *The Adaptive Web* 3-53.
43. Cañas, J. and Waern, Y. 2001. *Ergonomía cognitiva: Aspectos psicológicos de la interacción de las personas con la tecnología de la información*. Piramide, Madrid.
44. Carbó, J.M., Mor, E. and Minguillón, J. 2005. User Navigational Behavior in e-Learning Virtual Environments. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*, IEEE Computer Society, Los Alamitos, CA, USA, 243-249.
45. Card, S.K., Mackinlay, J.D. and Shneiderman, B. 1999. *Readings in information visualization: using vision to think*. Morgan Kaufmann Publishers, San Francisco, Calif.

46. Carmel, E., Crawford, S. and Chen, H. 1992. Browsing in hypertext: a cognitive study. *IEEE Transactions on Systems, Man and Cybernetics* 22, 865-884.
47. Carroll, J.M. 2000. Making use : scenario-based design of human-computer interactions. MIT Press, Cambridge, Mass.
48. Casas, J. 2003. Modelos de diseño de las TIC. Universitat Oberta de Catalunya, Barcelona.
49. Castells, M. 2000. The Information Age. The rise of the network society. Blackwell Publishers, Oxford; Malden, MA.
50. Castro, F., Vellido, A. and Nebot, A. 2005. Finding Relevant Features to Characterize Student Behavior on an e-Learning System. In *Proceedings of the International Conference on Frontiers in Education: Computer Science and Computer Engineering, FECS*, 210-216.
51. Castro, F., Vellido, A., Nebot, A. and Minguillón, J. 2005. Detecting Atypical Student Behaviour on an e-Learning System. In *Actas del I Congreso Nacional de Informática Educativa, Simposio Nacional de Tecnologías de la Información y las Comunicaciones en la Educación, SINTICE*, 14-16.
52. Catledge, L.D. and Pitkow, J.E. 1995. Characterizing browsing strategies in the World-Wide Web. In *Proceedings of the Third International World-Wide Web conference on Technology, tools and applications*, Darmstadt, Germany, Elsevier North-Holland, Inc, New York, NY, USA, 1065-1073.
53. Chi, E.H., Pirolli, P. and Pitkow, J. 2000. The scent of a site: a system for analyzing and predicting information scent, usage, and usability of a Web site. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, The Hague, The Netherlands, ACM Press, 161-168.
54. Chi, E.H., Rosien, A. and Heer, J. 2002. LumberJack: Intelligent Discovery and Analysis of Web User. In *WEBKDD 2002 - Proceedings of ACM SIGKDD Workshop on Web Mining for Usage Patterns and User Profiles*, Edmonton, Canada, 1-16.
55. Collis, B. and Moonen, J. 2001. Flexible Learning in a Digital World: Experiences and Expectations. Kogan Page, London.
56. Conklin, J. 1987. Hypertext: an introduction and survey. *IEEE Computer Magazine* 20, 17-41.
57. Cooley, R., Mobasher, B. and Srivastava, J. 1999. Data Preparation for Mining World Wide Web Browsing Patterns. *Knowledge and Information Systems* 1, 5-32.
58. Cooper, A. 2003. About face 2.0 : the essentials of interaction design. Wiley, New York; Chichester.
59. Cooper, A. 1999. The inmates are running the asylum. Sams, Indianapolis, IN.

60. Courage, C. and Baxter, K. 2005. Understanding your users : a practical guide to user requirements : methods, tools, and techniques. Morgan Kaufmann Publishers, San Francisco, CA.
61. Craik, K. 1943. The Nature of Explanation. Cambridge University Press, Cambridge.
62. De Lera, E. and Mor, E. 2007. The joy of e-learning: redesigning the e-learning experience. In *HCI 2007 Workshop: Design, use and experience of e-learning systems*. September 2007, Lancaster.
63. Dix, A. 1998. Human-computer interaction. Prentice Hall Europe, London.
64. Domingos, P. 2000. A Unified Bias-Variance Decomposition and its Applications. In *Proceedings of the Seventeenth International Conference on Machine Learning ICML*, 231-238.
65. Duda, R.O., Hart, P.E. and Stork, D.G. 2001. Pattern classification. Wiley, New York.
66. Eirinaki, M. and Vazirgiannis, M. 2003. Web mining for web personalization. *ACM Transactions on Internet Technology (TOIT)* 3, 1-27.
67. Eklund, J. and Sinclair, K. 2000. An empirical appraisal of the effectiveness of adaptive. *Educational Technology & Society* 3, 165-177.
68. Engelbrecht, E. 2003. A look at e-learning models: investigating their value for developing an e-learning strategy. *Progressio* 25, 38-47.
69. Feldstein, M. 2002. What is "usable" e-learning? *eLearn* 2002, 2.
70. Ferran, N., Mor, E. and Minguillón, J. 2005. Towards personalization in digital libraries through ontologies. *Library Management* 26, 206-217.
71. Galagan, P. 2000. Getting started with e-learning. *Training and Development* 54, 62-64.
72. Garreta-Domingo, M. and Mor, E. 2007. User Centered Design in E-Learning Environments: from Usability to Learner Experience. In *Proceedings of the EDEN 2007 Annual Conference*, June 13-16. Naples Italy.
73. Garrett, J.J. 2003. The elements of user experience : user-centered design for the web. Aiga; New Riders, New York, NY; Indianapolis, Ind.
74. Gentner, D. and Stevens, A.L. 1983. Mental models. L. Erlbaum Associates, Hillsdale, N.J.
75. Hackos, J.T. and Redish, J.C. 1998. User and task analysis for interface design. John Wiley & Sons, Inc, New York, NY, USA.

76. Extended Log File Format, <http://www.w3.org/TR/WD-logfile.html> (Consulta: 24/1/2008)
77. Harley, D. and Henke, J. 2007. Toward an Effective Understanding of Website Users. *D-Lib Magazine* 13, 1082-9873.
78. Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)* 22, 5-53.
79. Hevner, A.R., March, S.T., Park, J. and Ram, S. 2004. Design Science in Information Systems Research. *MIS Quarterly* 28, 75-105.
80. Hilbert, D.M. and Redmiles, D.F. 2000. Extracting usability information from user interface events. *ACM Computing Surveys* 32, 384-421.
81. Holmes, B. and Gardner, J. 2006. *E-learning: Concepts And Practice*. Sage Publications, London.
82. Holtzblatt, K. and Jones, S. 1993. Contextual Inquiry: A Participatory Technique for System Design. In *Participatory Design: Principles and Practices*, D. Schuler and A. Namioka, Eds. Lawrence Erlbaum Associates, Hillsdale, NJ.
83. IEEE Learning Technology Standardization Committee. 2002. IEEE Standard for Learning Object Metadata. 1484.12.1-2002.
84. ISO. 2000. ISO9126-1 ISO/IEC FCD 9126-1: Information technology - Software product quality - Part 1: Quality model, International Organization of Standardization.
85. ISO. 1999. ISO13407: Human-centred Design Processes for Interactive Systems.
86. ISO. 1998. ISO9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability.
87. ISO. 1997. ISO/IEC PDTR 9126-2: Information technology - Software quality characteristics and metrics - Part 2: External metrics, International Organization of Standardization.
88. Ivory, M.Y. and Hearst, M.A. 2001. The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys* 33, 470-516.
89. Jameson, A. 2006. Adaptive interfaces and agents. In *Human-Computer Interaction Handbook*, J.A. Jacko and A. Sears, Eds. Erlbaum, Mahwah, New Jersey.
90. Jonassen, D.H. 2003. *Learning to Solve Problems with Technology: A Constructivist Perspective*. Merrill Prentice Hall, Upper Saddle River, New Jersey.
91. Karat, C., Blom, J.O., Karat, J. and ebrary, I. 2004. Designing personalized user experiences in eCommerce. 5, 348.

92. Kasanoff, B. 2001. Making it personal : how to profit from personalization without invading privacy. Perseus, Cambridge, MA.
93. Kay, J. and Lum, A. 2004. Creating User Models from Web Logs. In *Proceedings of the Intelligent User Interfaces Workshop: Behavior-Based User Interface Customization*, 17-20.
94. Kay, J. and Thomas, R.C. 1995. Studying long-term system use. *Communications of the ACM* 38, 61-69.
95. Kitsuregawa, M., Pramudiono, I., Ohura, Y. and Toyoda, M. 2002. Some Experiences on Large Scale Web Mining. In *Proceedings of the Second International Workshop on Databases in Networked Information Systems*, Springer, 173-178.
96. Kobsa, A. 2007. Privacy-Enhanced Web Personalization. *The Adaptive Web*, Springer, 4321/2007, 628-670.
97. Kobsa, A. 2001. Generic User Modeling Systems. *User Modeling and User-Adapted Interaction* 11, 49-63.
98. Koper, R. and Tattersall, C. 2005. Learning design : a handbook on modelling and delivering networked education and training. Springer, Berlin.
99. Kosala, R. and Blockeel, H. 2000. Web mining research: a survey. *ACM SIGKDD Explorations Newsletter* 2, 1-15.
100. Kramer, J., Noronha, S. and Vergo, J. 2000. A user-centered design approach to personalization. *Communications of the ACM* 43, 44-48.
101. Krug, S. 2006. Don't make me think! : a common sense approach to web usability. New Riders, Berkeley, Calif.
102. Kuniavsky, M. 2003. Observing the user experience. Morgan Kaufmann Publishers, San Francisco, Calif.
103. Landauer, T.K. 1996. The Trouble with Computers: Usefulness, Usability, and Productivity. MIT Press, Cambridge, MA, USA.
104. Langley, P. 1999. User modeling in adaptive interfaces. In *UM '99: Proceedings of the seventh international conference on User modeling*, Banff, Canada, Springer-Verlag New York, Inc, Secaucus, NJ, USA, 357-370.
105. Lazar, J., Bessiere, K., Ceaparu, I., Robinson, J. and Shneiderman, B. 2003. Help! I'm lost: user frustration in web navigation. *IT & Society* 1, 18-26.
106. Levy, Y. 2007. Comparing dropouts and persistence in e-learning courses. *Computers & Education*, 48, 185-204.
107. Liu, H. and Setiono, R. 1998. Feature Transformation and Multivariate Decision Tree Induction. In *Discovery Science*, 279-290.

108. Losee Jr, R.M. 1989. Minimizing information overload: the ranking of electronic messages. *Journal of Information Science* 15, 179.
109. Marchionini, G. 1997. Information Seeking in Electronic Environments. Cambridge University Press, Cambridge.
110. Marquardt, C., Becker, K. and Ruiz, D. 2004. A pre-processing tool for Web usage mining in the distance education domain. In *Proceedings of Database Engineering and Applications Symposium, IDEAS'04*, 78-87.
111. Marsico, M.D. and Levialdi, S. 2004. Evaluating web sites: exploiting user's expectations. *International Journal of Human-Computer Studies* 60, 381-416.
112. Mayhew, D.J. 1999. The usability engineering lifecycle : a practitioner's handbook for user interface design. Morgan Kaufmann Publishers, San Francisco, Calif
113. Mazza, R. and Dimitrova, V. 2004. Visualising student tracking data to support instructors in web-based distance education. In *WWW Alt. '04: Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, New York, NY, USA, ACM, New York, NY, USA, 154-161.
114. Miller, M.J. 2005. Usability in E-Learning. *Learning circuits*.
115. Minguillón, J. 2002. On cascading small decision trees. PhD Thesis.
116. Minguillón, J., Mor, E., Santanach, F. and Guardia, L. 2004. Personalización del proceso de aprendizaje usando learning objects reutilizables. In *Actas del I Simposio Pluri-disciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables*, Guadalajara, España.
117. Mor, E., Garreta-Domingo, M., Minguillón, J. and Lewis, S. 2007. A Three-Level Approach for Analyzing User Behavior in Ongoing Relationships. *Lecture Notes in Computer Science* 4553, 971-980.
118. Mor, E. and Minguillón, J. 2003. An empirical evaluation of classifier combination schemes for predicting user navigational behavior. In *Proceedings International Conference on Information Technology: Coding and Computing [Computers and Communications]*, ITCC 2003, 467-471.
119. Mor, E., Minguillón, J. and Carbó, J.M. 2006. Analysis of User Navigational Behavior for E-Learning Personalization. In *Data Mining in E-Learning*, C. Romero and S. Ventura, Eds. WIT Press, 227-246
120. Mor, E. and Minguillón, J. 2004. Patrones de navegación de usuarios de un campus virtual. In *Actas del V Congreso Internacional de Interacción Persona Ordenador*, Lleida, España.
121. Mor, E. and Minguillón, J. 2004. E-learning personalization based on itineraries and long-term navigational behavior. In *WWW Alt. '04: Proceedings of the 13th*

*international World Wide Web conference on Alternate track papers & posters*, New York, NY, USA, ACM Press, 264-265.

122. Mor, E., Minguillón, J. and Santanach, F. 2007. Capturing user behavior in e-learning environments. In *Proceedings of the 3rd International Conference on Web Information Systems and Technologies*, Barcelona, Spain.
123. Morville, P. 2005. Ambient findability. O'Reilly, Sebastopol, CA.
124. Naeve, A. 2001. The Knowledge Manifold. An educational architecture that Supports Inquiry-Based Customizable Forms of E-learning. In *Proceedings of the 2nd European Web-Based Learning Environment Conference (WBLE 2001)*, Lund, Sweden, 200-212.
125. Nielsen, J. 1999. Designing Web Usability. New Riders Publishing, Thousand Oaks, CA, USA.
126. Nielsen, J. and Molich, R. 1990. Heuristic evaluation of user interfaces. In *CHI '90: Proceedings of the SIGCHI conference on Human factors in computing systems*, Seattle, Washington, United States, ACM, New York, NY, USA, 249-256.
127. Nielsen, J. 1998. Personalization is over-rated. *Jakob Nielsen's Alertbox* .
128. Nielsen, J. 1994. Usability engineering. Morgan Kaufmann.
129. Nielsen, J. and Mack, R.L. 1994. Usability inspection methods. Wiley, New York.
130. Norman, D.A. 2004. Emotional design : why we love (or hate) everyday things. Basic Books, New York.
131. Norman, D.A. 2002. The design of everyday things. Basic Books, New York.
132. Norman, D.A. 1993. Things that make us smart : defending human attributes in the age of the machine. Addison-Wesley, Reading, Mass.
133. Norman, D.A. 1988. The psychology of everyday things. Basic Books, New York.
134. Norman, D.A. and Draper, S.W. 1986. User Centered System Design; New Perspectives on Human-Computer Interaction. Lawrence Erlbaum Associates, Inc, Mahwah, NJ, USA.
135. Oates, B.J. 2006. Researching information systems and computing. Sage, London.
136. Paramythis, A. and Loidl-Reisinger, S. 2003. Adaptive Learning Environments and e-Learning Standards. In *2nd European Conference on E-Learning*, Glasgow Caledonian University, Glasgow.
137. Pearl, J. 1988. Probabilistic reasoning in intelligent systems : networks of plausible inference. Morgan Kaufmann Publishers, San Mateo, Calif.

138. Perkwitz, M. and Etzioni, O. 1998. Adaptive web sites: Automatically synthesizing web pages. In *Proceedings of the 15th National Conference on Artificial Intelligence*, 727-732.
139. Picard, R.W. 1997. *Affective computing*. MIT Press, Cambridge, Mass.
140. Pitkow, J. 1997. In search of reliable usage data on the WWW. *Computer Networks and ISDN Systems* 29, 1343-1355.
141. Polsani, P.R. 2003. Use and Abuse of Reusable Learning Objects. *Journal of Digital Information* 3, 4.
142. Popovic, V. 2000. Expert and Novice User Differences and Implications for Product Design and Useability. In *Proceedings of the IEA 2000 / HFES 2000*, 933-936.
143. Preece, J. 1994. *Human-computer interaction*. Addison-Wesley Pub. Co., Wokingham, England ; Reading, Mass.
144. Quintana, C., Carra, A., Krajcik, J. and Soloway, E. 2001. Learner-Centered Design: Reflections and New Directions. In *Human-Computer Interaction in the New Millennium*, 605-624.
145. Resnik, P. and Varian, H.R. 1997. Recommender systems. *Communications of the ACM, special issue* 40, 56-58.
146. Riecken, D. 2000. Introduction: personalized views of personalization. *Communications of the ACM* 43, 26-28.
147. Romero, C. and Ventura, S. 2007. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications* 33, 135-146.
148. Ros, A. 2004. La verdadera apuesta del aprendizaje virtual: los aspectos sociales del e-learning. *Revista de Universidad y Sociedad del Conocimiento* 1. (Consulta: 12/1/2008).
149. Rosenberg, M.J. 2001. *E-learning : strategies for delivering knowledge in the digital age*. McGraw-Hill, New York.
150. Rosson, M.B. and Carroll, J.M. 2002. *Usability engineering : scenario-based development of human-computer interaction*. Academic Press, San Francisco.
151. Rubin, J. 1994. *Handbook of usability testing*. Wiley & Sons, Inc. New York.
152. Rudd, J., Stern, K. and Isensee, S. 1996. Low vs. high-fidelity prototyping debate. *interactions* 3, 76-85.
153. Schaffer, E.M. 2004. *Institutionalization of usability: a step-by-step guide*. Addison-Wesley, Boston, MA.



154. Self, J. 1988. Artificial intelligence and human learning : intelligent computer-aided instruction. Chapman and Hall, London.
155. Sharp, H., Rogers, Y. and Preece, J. 2007. Interaction design : beyond human-computer interaction. John Wiley & Sons, Chichester, England.
156. Shneiderman, B. 1998. Designing the user interface : strategies for effective human-computer-interaction. Addison Wesley Longman, Reading, Mass.
157. Shneiderman, B. 1981. Direct manipulation: A step beyond programming languages. In *Proceedings of the joint conference on Easier and more productive use of computer systems. (Part - II)*, Ann Arbor, MI, ACM, New York, NY, USA, 143.
158. Siemens, G. Connectivism: A Learning Theory for the Digital Age. *elearnspace* (Consulta: 29/1/2008).
159. SIGCHI Group (Group: US). Curriculum Development. 1992. ACM SIGCHI Curricula for Human-computer Interaction. Association for Computing Machinery.
160. Sloman, M. 2001. The E-Learning Revolution: From Propositions to Action. CIPD, London.
161. Snyder, C.A. 2003. Paper Prototyping: The Fast and Simple Techniques for Designing and Refining the User Interface. Morgan Kaufmann.
162. Soloway, E., Guzdial, M. and Hay, K.E. 1994. Learner-centered design: the challenge for HCI in the 21st century. *interactions* 1, 36-48.
163. Spiliopoulou, M. 2000. Web usage mining for Web site evaluation. *Communications of the ACM* 43, 127-134.
164. Srivastava, J., Cooley, R., Deshpande, M. and Tan, P. 2000. Web usage mining: discovery and applications of usage patterns from Web data. 1, 12-23.
165. Stephenson, J. 2001. Teaching & Learning Online: New Pedagogies for New Technologies. Routledge, New York.
166. Storey, M.A., Phillips, B., Maczewski, M. and Wang, M. 2002. Evaluating the usability of Web-based learning tools. *Educational Technology & Society* 5, 91-100.
167. Sutcliffe, A.G., Maiden, N.A.M., Minocha, S. and Manuel, D. 1998. Supporting scenario-based requirements engineering. *Software Engineering, IEEE Transactions on* 24, 1072-1088.
168. Thomas, R., Kennedy, G.E., Draper, S., Mancy, R., Crease, M., Evans, H. and Gray, P. 2003. Generic usage monitoring of programming students. In *Proceedings of the 20 th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE '03)*, 715-719.

169. Tidwell, J. 2006. *Designing Interfaces: Patterns for Effective Interaction Design*. O'Reilly.
170. Tselios, N., Avouris, N., Dimitracopoulou, A. and Daskalaki, S. 2001. Evaluation of distance-learning environments: impact of usability on student performance. *International Journal of Educational Telecommunications* 7, 355-378.
171. Universitat Oberta de Catalunya. 2007. Document de formació de consultors.
172. Urdan, T.A. and Weggen, C.C. 2000. *Corporate E-learning: Exploring a New Frontier*. WR Hambrecht.
173. Van Harmelen, M. *Personal Learning Environments*, ([http://octette.cs.man.ac.uk/jitt/index.php/Personal\\_Learning\\_Environments](http://octette.cs.man.ac.uk/jitt/index.php/Personal_Learning_Environments) (Consulta: 14/2/2008))
174. Verbert, K. and Duval, E. 2004. Towards a Global Architecture for Learning Objects: A Comparative Analysis of Learning Object Content Models.
175. Wenger, E. 1987. *Artificial intelligence and tutoring systems : computational and cognitive approaches to the communication of knowledge*. Morgan Kaufmann Publishers, Los Altos, Calif.
176. Wharton, C., Rieman, J., Lewis, C. and Polson, P. 1994. The cognitive walkthrough method: a practitioner's guide. In *Usability inspection methods*, Anonymous John Wiley & Sons, Inc. New York, NY, USA, 105-140
177. Wiley, D.A. 2001. *Instructional Use of Learning Objects*. Agency for Instructional Technology.
178. Zhang, T., Ramakrishnan, R. and Livny, M. 1996. BIRCH: an efficient data clustering method for very large databases. In *SIGMOD '96: Proceedings of the 1996 ACM SIGMOD international conference on Management of data*, Montreal, Quebec, Canada, ACM, New York, NY, USA, 103-114.
179. Zukerman, I. and Albrecht, D.W. 2001. Predictive Statistical Models for User Modeling. *User Modeling and User-Adapted Interaction* 11, 5-18.