

# La importància del llenguatge en la recuperació de la informació

Ana María Checa Rubio  
Pere Masip Masip

P06/09046/00262



# Índex

|  |    |
|--|----|
| <b>Introducció</b> .....   | 5  |
| <b>1. Els llenguatges documentals: una definició</b> .....                                   | 7  |
| <b>2. Els llenguatges documentals: una classificació</b> .....                               | 8  |
| 2.1. Llenguatges documentals segons la funció .....  | 8  |
| 2.2. Els llenguatges documentals segons el control sobre el<br>vocabulari .....              | 9  |
| 2.3. Llenguatges documentals segons el grau de coordinació .....                             | 9  |
| <b>3. L'ús dels llenguatges documentals: el procés d'indexació</b> ..                        | 11 |
| 3.1. Anàlisi documental de forma .....   | 12 |
| 3.2. Anàlisi documental de contingut .....   | 14 |
| <b>4. L'ús del llenguatge lliure com a alternativa al<br/>    llenguatge controlat</b> ..... | 19 |
| <b>5. Els llenguatges documentals: una descripció</b> .....                                  | 22 |
| 5.1. Llistes de paraules clau .....  | 22 |
| 5.2. Llistes de descriptors lliures .....  | 23 |
| 5.3. Tesaurus .....  | 24 |
| 5.4. Llistes d'encapçalaments .....  | 27 |
| 5.5. Classificacions .....   | 29 |
| 5.6. Llistes d'autoritats .....  | 31 |
| <b>Bibliografia</b> .....  | 33 |



## Introducció

La cerca d'informació és un procés que té el seu punt de partida en el moment en què es reconeix l'existència d'una necessitat d'informació. Les necessitats informatives no són totes iguals; per tant, les cerques d'informació tampoc no tindran totes les mateixes característiques.

S'ha vist en el mòdul "Introducció a la cerca i recuperació de la informació" que tradicionalment les cerques d'informació es divideixen en dues grans categories:

- Cerca d'un ítem conegut.
- Cerca temàtica.

Les primeres no solen presentar dificultats excessives. Es produeixen quan es fa necessari localitzar un document concret del qual es coneix, per exemple, el títol, l'autor, l'editor...

La cerca temàtica, en canvi, ofereix habitualment més problemes. Ens enfrontem a una cerca d'aquest tipus quan necessitem, per exemple, conèixer la producció científica en una disciplina determinada, una dada actual específica, informació d'actualitat, etc.

Una de les principals dificultats que caracteritzen les cerques temàtiques està determinada per la participació del llenguatge natural. El **llenguatge natural** és el llenguatge que usem quan parlem i escrivim. És un llenguatge extremadament ric, que s'ha mostrat especialment eficaç a l'hora de comunicar-nos, però que presenta alguns inconvenients quan l'hem d'utilitzar per a intentar recuperar informació.

Aquests inconvenients estan relacionats amb les pròpies característiques del llenguatge natural, que és un llenguatge equívoc i ambigu. Així, el mot *Mercuri*, tant pot fer referència al planeta més proper al Sol, al metall, com al déu de la mitologia clàssica. De la mateixa manera, per a referir-nos a la malària, podem usar també el terme sinònim *paludisme*.

En contraposició al llenguatge natural, i precisament per a facilitar la recuperació de la informació, si més no *a priori*, disposem dels **llenguatges documentals**, que són, com es veurà, llenguatges unívocs. És a dir, cada terme representa un únic concepte.

Els llenguatges documentals exerceixen un control sobre el vocabulari i faciliten, en principi, el procés de cerca de la informació i l'obtenció de la informació pertinent.

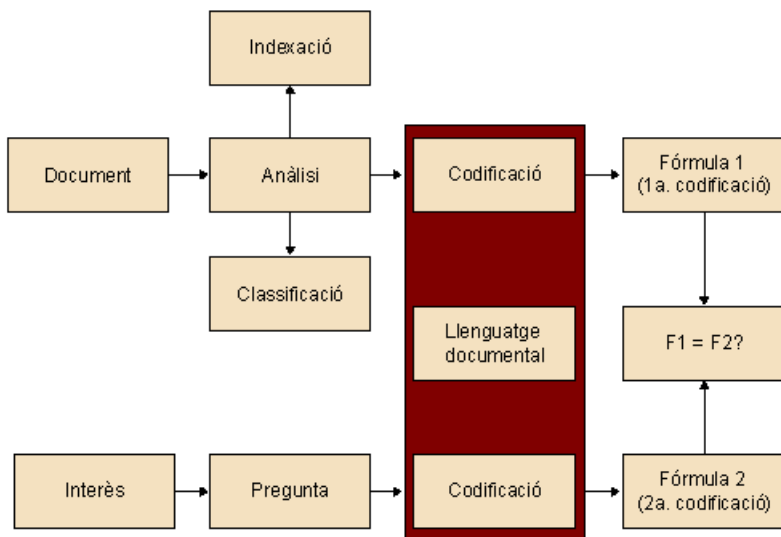
Però, per què és important el control del vocabulari en la recuperació de la informació? Vegem-ho en l'exemple següent.

### Paludisme o malària

Un centre de documentació incorpora a la seva col·lecció un document sobre el paludisme. Els documentalistes en fan l'anàlisi documental, descriuen el document i n'introdueixen els detalls (autor, títol, editorial, paraules clau...) al catàleg. Dies més tard un usuari interessat en el tema acudeix al catàleg i busca la informació sobre aquesta malaltia, però interroga el catàleg a partir del mot *malària*, un sinònim de *paludisme*. Com que els documentalistes havien usat el mot *paludisme* per a descriure el contingut del document, l'usuari no recupera el document que uns dies abans havia estat adquirit pel centre de documentació.

El que ha succeït en aquest exemple és que no ha coincidit la manera en què els documentalistes han descrit el contingut del document amb com ho ha fet l'usuari, ha fracassat el que Jacques Maniez anomena "doble codificació".

Maniez denomina la comunicació entre el que demana i el que subministra la informació, *mediació documental*, que es realitza per mitjà d'un procés més complex: la **doble codificació**. La doble codificació es pot representar de forma gràfica de la manera següent:



El procés de doble codificació és, per tant, una comparació entre els resultats del procés d'indexació dut a terme per l'analista i els de l'usuari final. L'èxit es produeix quan coincideixen els resultats de l'un i de l'altre.

### Utilització dels llenguatges documentals

En l'exemple, els termes usats en la codificació no han coincidit i això ha conduït al fracàs la cerca documental. Per a facilitar la coincidència en el procés de codificació, els documentalistes usen els llenguatges documentals.

## 1. Els llenguatges documentals: una definició

Una definició clàssica de llenguatge documental és la que proporciona Georges Van Slype:

"Tot sistema de signes que permeti representar els continguts dels documents amb la finalitat de recuperar els documents pertinents en resposta a consultes que tractin d'aquell contingut."

G. Van Slype (1991). *Los Lenguajes de indización: concepción, construcción y utilización en los sistemas documentales* (pàg. 21). Madrid: Fundación Germán Sánchez Ruipérez / Pirámide.

És a dir, els llenguatges documentals són eines que ens permeten descriure el contingut dels documents d'una manera clara i inequívoca per a la seva recuperació posterior, tot evitant els problemes del llenguatge natural.

El control del vocabulari que s'exerceix en els llenguatges documentals permet que un únic terme denomini un únic concepte, de manera que:

- Un terme no pot definir més d'un concepte.
- Un concepte no pot ser definit per més d'un terme.

Perquè això sigui possible, en els llenguatges documentals:

- Els termes estan relacionats semànticament,
  - A nivell de sinònims i quasisinònims: paraules amb el mateix significat o que volen dir pràcticament el mateix.
  - A nivell d'antònims: paraules que signifiquen el contrari.
- S'estableix una diferenciació entre termes homònims.
- S'estableixen relacions entre els termes que conformen el llenguatge documental.

## 2. Els llenguatges documentals: una classificació

Els llenguatges documentals es poden classificar d'acord amb diversos criteris:

- Segons la funció que desenvolupin:
  - Llenguatges d'indexació.
  - Llenguatges de classificació.
- D'acord amb el nivell de control:
  - Llenguatges lliures.
  - Llenguatges controlats.
- Pel grau de coordinació
  - Llenguatges precoordinats.
  - Llenguatges postcoordinats.
- Segons l'estructura, principalment:
  - Llenguatges d'estructura jeràrquica.
  - Llenguatges d'estructura combinatòria.

### 2.1. Llenguatges documentals segons la funció

Els **llenguatges d'indexació** són els emprats en la indexació i permeten la representació dels continguts dels documents mitjançant paraules obtingudes del llenguatge natural.

#### **Llenguatges d'indexació**

Són llenguatges d'indexació els tesaurus, les llistes d'encapçalaments, les llistes de paraules clau i de descriptors lliures.

#### **Indexació**

La indexació consisteix a descriure el contingut dels documents a partir de l'assignació d'uns termes (paraules clau, descriptors, encapçalaments de matèria...).

Els **llenguatges de classificació**, com els d'indexació, permeten descriure els continguts dels documents, però en aquest cas mitjançant codis numèrics o alfanumèrics que corresponen a categories en què prèviament s'ha dividit el coneixement.

#### **Llenguatges de classificació**

Són llenguatges de classificació les classificacions.



## 2.2. Els llenguatges documentals segons el control sobre el vocabulari

Els **llenguatges lliures**, formats *a posteriori*, estan constituïts per termes obtinguts del mateix document, per tant del llenguatge natural. Per alguns autors no es poden considerar llenguatges documentals.

### Llenguatges lliures

Hi ha dos tipus fonamentals de llenguatges lliures, les llistes de paraules clau i les llistes de descriptors lliures.

Els **llenguatges controlats**. Construïts *a priori* del moment de la indexació, en aquests llenguatges s'exerceix un rigorós control terminològic, de manera que un concepte únicament pot estar representat per un sol terme. Així, desapareixen, o disminueixen, els problemes identificats en l'ús del llenguatge natural com la polisèmia, la sinonímia, l'homonímia...

### Llenguatges controlats

Són llenguatges controlats: els tesaurus, les llistes d'encapçalaments i les classificacions.

## 2.3. Llenguatges documentals segons el grau de coordinació

El grau de coordinació fa referència al moment en què es combinen els elements que componen els llenguatges documentals. Si els elements es combinen en el moment de la descripció –operació que realitza el documentalista–, parlem de **llenguatges precoordinats**; si la combinació es fa en el moment de la cerca documental –en mans de l'usuari–, estem davant de **llenguatges postcoordinats**.

### Llenguatges precoordinats i postcoordinats

Les llistes d'encapçalaments i les classificacions s'englobarien dins de la categoria de llenguatges precoordinats; per contra, els tesaurus, les llistes de descriptors lliures i les llistes de paraules clau són postcoordinats.

### Vegeu

Més endavant es veuran amb més detall les principals característiques dels llenguatges lliures i dels controlats. Ara és el moment de veure el paper que el llenguatge desenvolupa en la recuperació de la informació.

### Tipologies de llenguatges documentals

| Tipologies de llenguatges documentals |  |  |
|---------------------------------------|--|--|
| Segons la funció que desenvolupen     |  |  |

| <b>Tipologies de llenguatges documentals</b>                |                              |  |
|---|------------------------------|--|
|   | Llenguatges d'indexació      | <ul style="list-style-type: none"> <li>• Llistes de descriptors lliures</li> <li>• Llistes de paraules clau</li> <li>• Tesaurus</li> <li>• Llistes d'encapçalaments</li> </ul> |
|   | Llenguatges de classificació | <ul style="list-style-type: none"> <li>• Classificacions</li> </ul>  |
| <b>Segons el control que s'exerceix sobre el vocabulari</b> | Llenguatges lliures          | <ul style="list-style-type: none"> <li>• Llistes de descriptors lliures</li> <li>• Llistes de paraules clau</li> </ul>   |
|   | Llenguatges controlats       | <ul style="list-style-type: none"> <li>• Tesaurus</li> <li>• Llistes d'encapçalaments</li> <li>• Classificacions</li> </ul>  |
| <b>Segons quan es coordinin els termes</b>                  | Llenguatges precoordinats    | <ul style="list-style-type: none"> <li>• Llistes d'encapçalaments</li> <li>• Classificacions</li> </ul>  |
|   | Llenguatges postcoordinats   | <ul style="list-style-type: none"> <li>• Llistes de descriptors lliures</li> <li>• Llistes de paraules clau</li> <li>• Tesaurus</li> </ul>                                     |
| <b>Segons l'estructura interna</b>                          | Estructura jeràrquica        | <ul style="list-style-type: none"> <li>• Classificacions</li> </ul>  |
|   | Estructura combinatòria      | <ul style="list-style-type: none"> <li>• Llistes d'encapçalaments</li> <li>• Tesaurus</li> </ul>   |

### 3. L'ús dels llenguatges documentals: el procés d'indexació

Quan un document arriba a una biblioteca o centre de documentació experimenta una sèrie de processos abans de quedar a disposició de l'usuari. Aquests processos s'engloben en el que s'anomena *anàlisi o tractament documental*.

Pinto Molina defineix l'anàlisi documental com:

"un conjunt d'operacions, unes d'ordre intel·lectual i altres mecàniques i repetitives que afecten la forma i el contingut dels documents originals, reelaborant-los i transformant-los en altres de caràcter instrumental o secundari que faciliten a l'usuari la identificació precisa, la recuperació i la difusió [dels documents originals]."

M. Pinto Molina (1989). "El análisis documental". A: J. López Yepes (comp.). *Fundamentos de información y documentación* (pàg. 267). Madrid: Eudema.

Si es dissectiona aquesta definició, s'observa que l'**anàlisi documental** consisteix en l'aplicació d'un seguit de tècniques que tenen com a objectiu facilitar la identificació i la recuperació dels documents, i també la creació d'una representació dels documents que permeti identificar-los i que desenvolupi un rol d'intermediària entre aquests i els usuaris.

#### El registre bibliogràfic d'un catàleg

El registre desenvolupa un paper d'intermediari entre el document original i els usuaris.

The screenshot shows a web browser window displaying the 'Catàleg Col·lectiu de les Universitats de Catalunya' (CCUC) interface. The page title is 'Catàleg Col·lectiu de les Universitats de Catalunya'. Below the title, there is a section for 'INFORMACIÓ BIBLIOGRÀFICA'. The main content is a table with the following data:

|              |  |            |
|--------------|--|------------|
| LOCALITZACIÓ | UAB URV  | Demanar-lo |
| AUTOR        | Pavlik, John V.  |            |
| TÍTOL        | El Periodismo y los nuevos medios de comunicación / John V. Pavlik |            |
| PUBLICACIÓ   | Barcelona [etc.] : Paidós, cop. 2005                               |            |
| DESCRIPCIÓ   | 351 p.   |            |
| COL·LECCIÓ   | Paidós comunicació . 160   |            |
| NOTA         | Bibliografía. Index  |            |
| MATÈRIA      | Periodisme -- Processament de dades                                |            |
| MATÈRIA      | Periodisme -- Innovacions tecnològiques                            |            |
| ISBN         | 84-493-1730-4  |            |

Below the table, there are three buttons: 'Guardar', 'Registre MARC', and 'Nova cerca'. At the bottom of the page, there is a footer with the text: 'Consortori de Biblioteques Universitàries de Catalunya' and 'Comentaris i suggeriments: info@cuc.es'. The copyright notice reads 'Copyright (c) 1998, Y.T.S. Inc.'.

Registre del CCUC del llibre de John V. Pavlik. El registre ofereix informació detallada del document original.

En aquest cas el document original seria el llibre de John V. Pavlik *El periodismo y los nuevos medios de comunicación*.



El registre bibliogràfic d'un catàleg és el fruit de l'anàlisi documental aplicada sobre el document original per part del personal de la biblioteca. Gràcies a aquest registre, aquest document instrumental, en paraules de Pinto Molina, disposa d'informació detallada del document original que permeti identificar-lo dins el conjunt del fons bibliogràfic del CCUC, conèixer les seves característiques principals, saber de què tracta i localitzar-lo. A més, el registre esdevé una eina que facilita la recuperació dels documents.

Com indica Pinto Molina, el tractament o l'anàlisi documental es realitza a dos nivells:

- A nivell formal.
- A nivell de contingut.

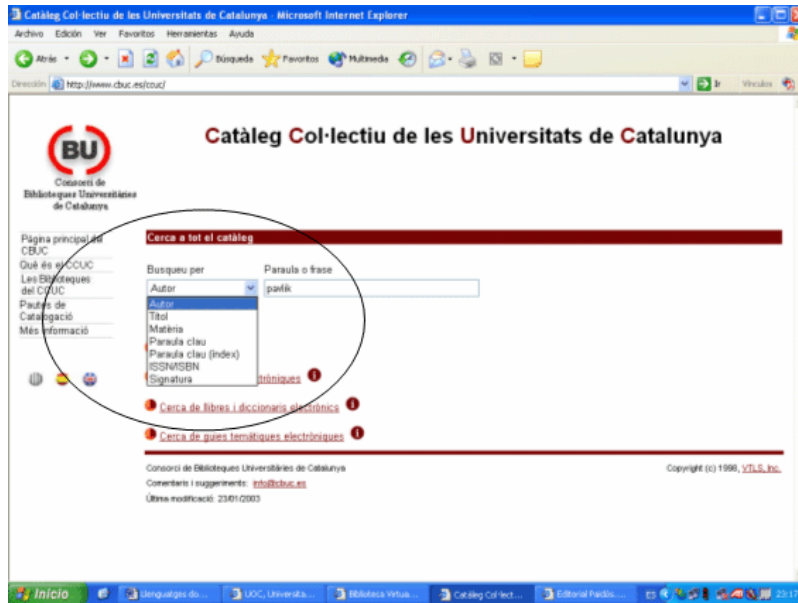
### 3.1. Anàlisi documental de forma

A nivell formal, l'anàlisi dels documents permet la descripció formal amb l'objectiu de facilitar-ne la identificació. Inclou dues operacions principals:

- **La descripció bibliogràfica.** Per mitjà de la descripció bibliogràfica es descriuen tots els elements del document (autor, títol, editorial, ISBN...) que en possibiliten la identificació dins d'un fons concret.
- **La catalogació.** Determina els punts d'accés al document, és a dir, es determinen a partir de quins elements formals el document podrà ser recuperat.

#### El Catàleg Col·lectiu de les Universitats de Catalunya

Observem que el CCUC ofereix sis punts d'accés al document: autor, títol, matèria, paraula clau, número normalitzat (ISSN/ISBN) i signatura.



El CCUC ofereix sis punts d'accés al document.

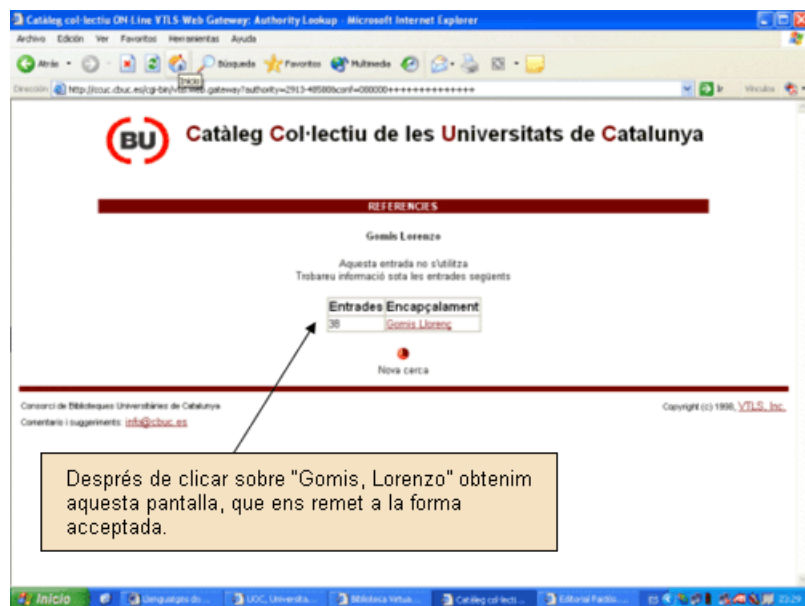
Idealment, els punts d'accés i la manera com es representen han d'estar estandaritzats per tal d'evitar ambigüitats a l'hora de la recuperació.

### Diferents signatures o variants ortogràfiques

Un mateix autor pot signar de diverses maneres. La catalogació ha de permetre recuperar tots els documents d'aquell autor disponibles a la biblioteca, independentment de les variants que utilitzi. Aquest és el cas, per exemple, del periodista i escriptor Llorenç Gomis, que habitualment signa com a *Llorenç Gomis* quan publica en català i com a *Lorenzo Gomis* quan ho fa en castellà. La catalogació permet que tots els documents de Llorenç Gomis puguin ser recuperats independentment de la forma que hagi adoptat l'autor o la llengua de les seves obres.



Entrada no acceptada per l'autoritat Llorenç Gomis.



Forma acceptada per l'autoritat Llorenç Gomis.

A vegades, els problemes deriven de l'existència de diverses variants ortogràfiques d'un mateix nom, la catalogació també les ha d'unificar. Així, ha de permetre recuperar l'obra *L'oncle Vània*, independentment que es busqui per alguna de les múltiples variants en què ens podem trobar el cognom del seu autor: Txèkhov, Chekhov, Cjevov, Cekhov, Tschekhov.

### Vegeu

Aquest aspecte de l'anàlisi documental es recuperarà al final del mòdul quan es tracti breument de les llistes d'autoritats.

## 3.2. Anàlisi documental de contingut

L'anàlisi documental pel que fa al contingut té com a funció proporcionar a l'usuari una idea precisa del contingut d'un document, i inclou tres operacions:

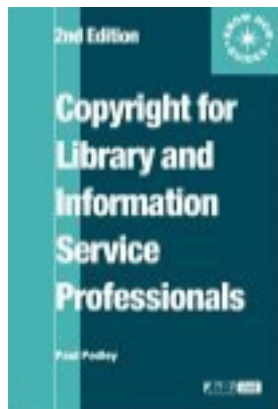
- La **indexació**, per mitjà d'un seguit de paraules clau provinents d'un llenguatge estandaritzat amb la finalitat principal de facilitar-ne la recuperació.
- La **classificació**, mitjançant uns codis que en l'actualitat s'usen fonamentalment per a facilitar-ne la localització física dins d'una col·lecció.
- El **resum**, per mitjà del llenguatge natural, ofereix una representació abreujada del document.

Aturem-nos, ara, un moment, en la primera de les operacions, la indexació. Com s'ha vist, la indexació consisteix a descriure el contingut dels documents a partir de l'assignació d'uns termes (paraules clau, descriptors, encapçalaments de matèria...). El documentalista, tanmateix, no atorga lliurement els termes que considera més oportuns, sinó que s'ha d'ajustar a les possibilitats que li ofereix el llenguatge d'indexació adoptat per la biblioteca o centre de documentació. Vegem-ho amb més deteniment:

- 1) Arriba un document al centre de documentació i el documentalista l'examina per a identificar-ne els conceptes significatius.
- 2) Tradueix aquests conceptes al llenguatge d'indexació adoptat pel centre de documentació.

### Portem-ho a la pràctica

Arriba a una biblioteca, per exemple, la del CSIC, el llibre de Paul Pedley *Copyright for Library and Information Services Professionals*.



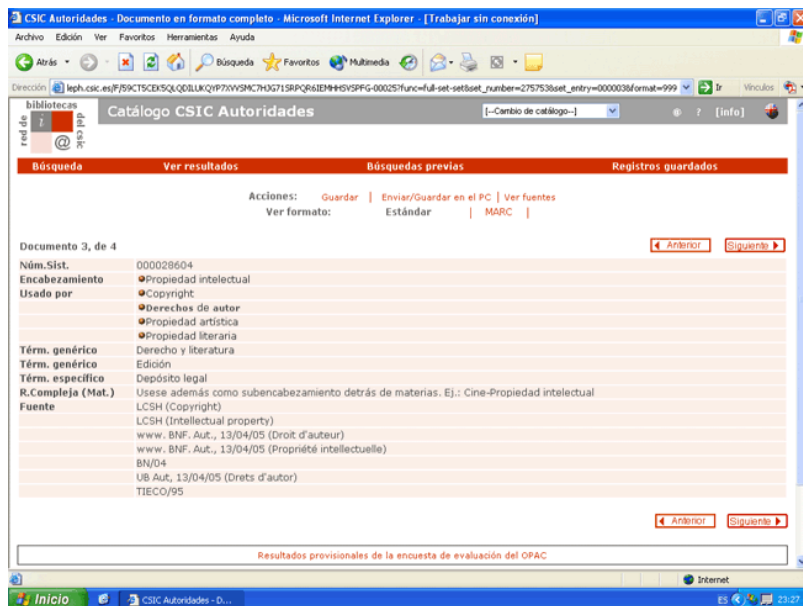
Document original.

El documentalista fa l'anàlisi documental del document, incloent-hi la indexació. Per a indexar el document, òbviament, el documentalista no pot llegir tot el document, sinó que n'analitza parts concretes, habitualment, el títol, el resum, el sumari, la introducció... i tot seguit determina els conceptes essencials explícits i implícits de què tracta el document. En el cas que ens ocupa, aquests conceptes podrien ser *drets d'autor*, *biblioteques*.

A continuació el documentalista haurà de convertir aquests conceptes seleccionats del llenguatge natural en els termes acceptats en el llenguatge d'indexació adoptat per la seva biblioteca. En el cas del CSIC, usen com a llenguatge documental d'indexació la seva pròpia llista d'encapçalaments.

El primer que caldrà fer és confirmar si el primer dels conceptes, *drets d'autor*, és un encapçalament acceptat. En aquest cas, observem que no és un terme acceptat, però sí que ho és la forma sinònima  *propietat intel·lectual*. El documentalista usará el terme  *propietat intel·lectual* per a indexar el document i continuarà el procés amb la resta de termes.

En un futur, aquest document podrà ser recuperat per qualsevol usuari que cerqui per qualsevol dels termes pels quals el document ha estat indexat, però gràcies a l'existència del llenguatge documental, també pels termes sinònims no acceptats.



Consulta de la llista d'encapçalaments de la biblioteca del CSIC.

Així doncs, l'ús del llenguatge d'indexació permet que cada concepte estigui representat per un únic terme i que un terme representi exclusivament un concepte. En el cas de l'existència de termes sinònims o quasisinònims, en aquest cas *drets d'autor*, *copyright*, *propietat artística* i *propietat literària*, el sistema remet a la forma acceptada mitjançant l'epígraf *Usat per*. En definitiva, l'ús del llenguatge documental, controlat en aquest cas, permet reduir els problemes associats al llenguatge natural, descrits anteriorment.

### Llenguatges documentals controlats



Els llenguatges controlats permeten que cada concepte estigui representat per un únic terme.

La indexació, però, no únicament es produeix en el moment de l'anàlisi documental, que realitza el documentalista, també ens trobem davant d'una indexació en el moment que un usuari realitza una cerca documental. Com el cas anterior, la indexació presenta dues fases:

- 1) Identificació en el llenguatge natural dels conceptes que millor descriuen la necessitat d'informació plantejada.
- 2) Conversió dels conceptes en llenguatge natural al llenguatge controlat utilitzat pel recurs d'informació que s'interroga.

L'èxit o fracàs de la cerca dependrà, en gran mesura, de la correcta elecció dels conceptes i de la seva traducció als termes pels quals el contingut dels documents hagin estat identificats.



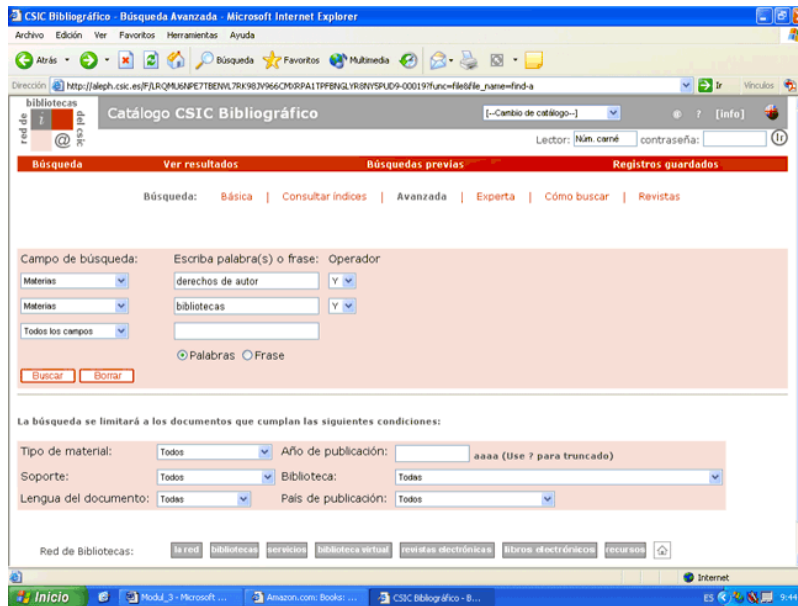
Recuperem, en aquest punt, el concepte de *mediació documental* de Jacques Maniez, que ja s'havia descrit a l'inici del mòdul. Ara, però, gràcies a l'ús del llenguatge documental els resultats del procés de doble codificació poden ser ben diferents dels que s'han vist llavors.

### Recordeu

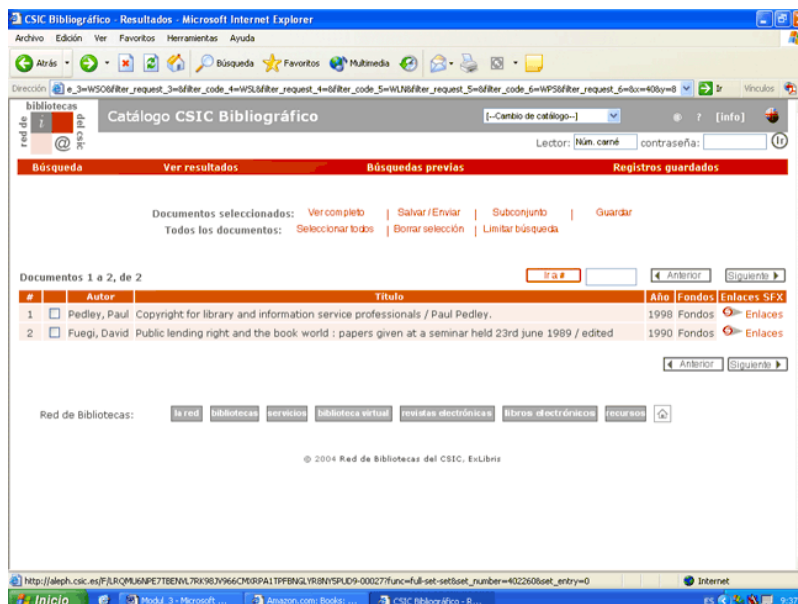
Hem plantejat el concepte de *mediació documental* en la introducció d'aquest mòdul.

## Mediació documental

Tornem al llibre de Paul Pedley, imaginem que arriba un usuari a la biblioteca del CSIC que necessita informació sobre els drets d'autor a les biblioteques. Per a recuperar el que hi ha sobre el tema, acudeix al catàleg i realitza una cerca per matèria, combinant amb l'operador booleà Y (AND) els dos conceptes que, segons ell, millor descriuen la necessitat d'informació plantejada, en aquest cas "Drets d'autor" i "Biblioteques".



Cerca al catàleg del CSIC.



Resultats obtinguts de la cerca al catàleg del CSIC.

Fruit de la recerca realitzada, s'obtenen les referències de dos documents, entre els quals apareix el llibre de Paul Pedley. S'han pogut recuperar aquests dos documents perquè han coincidit els termes seleccionats en el procés d'indexació dut a terme per l'analista i els de l'usuari final.

The screenshot shows a web browser window displaying the CSIC Bibliográfico catalog. The page title is "Catálogo CSIC Bibliográfico". The search results section shows "Documento 1, de 2". The document details are as follows:

|                 |  |
|-----------------|--|
| Núm.Sistema     | 000795432  |
| Autor principal | ●Pedley, Paul  |
| Título          | ●Copyright for library and information service professionals / Paul Pedley.      |
| Pie imprenta    | London : Aslib, cop. 1998.   |
| Descr. física   | 64 p. : 24 cm.   |
| Serie           | ●ASLB know how series  |
| Bibliogr.       | Bibliografía   |
| ISBN            | ●0851424120  |
| Materia         | ●Bibliotecas -- Servicios de información bibliográfica<br>●Propiedad intelectual |
| CDU             | ●025.5:347.78  |
| Ejemplares      | Todos los ejemplares   |
| Biblioteca      | M-BG.Hum   |

At the bottom of the page, there is a navigation bar for the "Red de Bibliotecas" with links to "la red", "bibliotecas", "servicios", "biblioteca virtual", "revistas electrónicas", "libros electrónicos", and "recursos".

Registre del catàleg del CSIC del document buscat.

En aquest cas concret, tot i que el terme usat en la indexació per part del documentalista fou *propiedad intelectual (propiedad intelectual)* i que l'usuari busca a partir de drets d'autor (*derechos de autor*), el document ha estat recuperat perquè la presència del llenguatge documental ha fet desaparèixer el problema de la sinonímia. Internament, el sistema, davant una cerca a partir d'un terme no acceptat, remet a la forma acceptada pel llenguatge documental.

## 4. L'ús del llenguatge lliure com a alternativa al llenguatge controlat

Malgrat els avantatges que ofereix l'ús de llenguatges documentals controlat per a la recuperació de la informació, l'ús del llenguatge lliure no s'ha de descartar, ja que cada cop és més present en els sistemes de recuperació de la informació. Usem el llenguatge lliure, per exemple, en les cerques a cercadors d'Internet i a la gran part de bases de dades de text complet.

Com s'ha descrit en l'apartat corresponent, l'ús del llenguatge lliure per a la recuperació de la informació consisteix a usar els termes que apareixen en el mateix document i en la mateixa forma en què ho fa. Els termes poden procedir del títol, del resum o del mateix text del document.

Mentre que els llenguatges documentals controlats són:

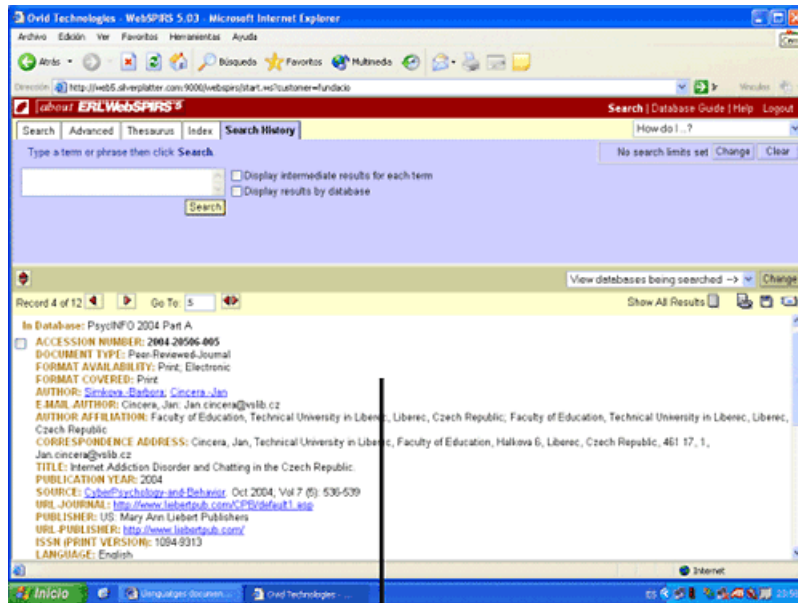
- un codi unívoc i estereotipat (fixat de manera permanent),
- un codi controlat i no lliure,
- normalitzat i no arbitrari.

El llenguatge lliure és equívoc, en què coexisteixen diversos significats per a un sol significat o diversos significats sinònims, característiques que, com s'ha vist, no ajuden a recuperar la informació. Amb tot, ofereix almenys un avantatge sobre els llenguatges controlats: a causa del seu caràcter il·limitat permet una especificitat més gran a l'hora de realitzar la cerca documental.

Efectivament, si la necessitat informativa és molt específica, habitualment s'obtenen millors resultats usant el llenguatge lliure en lloc del controlat. Vegem-ne un exemple.

### Llenguatge lliure enfront de llenguatge controlat

Si necessitem identificar la producció científica recent sobre l'addicció als xats en una base de dades que usi un llenguatge documental controlat, posem per cas a la base de dades PsycInfo, és molt probable que hàgim de fer la cerca a partir dels termes *addicció* i *Internet*, en lloc d'*addicció* i *xat*. El motiu és que el terme *xat*, a causa de la seva especificitat i actualitat, possiblement no estarà recollit en el llenguatge documental de la base de dades. Fet que ens obligarà a haver d'usar el terme genèric *Internet*.



ACCESSION NUMBER: 2004-20506-005  
 DOCUMENT TYPE: Peer-Reviewed-Journal  
 FORMAT AVAILABILITY: Print, Electronic  
 FORMAT COVERED: Print  
 AUTHOR: [Srnkova, Barbora](#), [Cincera, Jan](#)  
 E-MAIL AUTHOR: Cincera, Jan, Jan.cincera@vslib.cz  
 AUTHOR AFFILIATION: Faculty of Education, Technical University in Liberec, Liberec, Czech Republic; Faculty of Education, Technical University in Liberec, Czech Republic  
 CORRESPONDENCE ADDRESS: Cincera, Jan, Technical University in Liberec, Faculty of Education, Halkova 6, Liberec, Czech Republic, 461 17, 1, Jan.cincera@vslib.cz  
 TITLE: Internet Addiction Disorder and Chatting in the Czech Republic  
 PUBLICATION YEAR: 2004  
 SOURCE: [CyberPsychology and Behavior](#), Oct 2004, Vol 7 (5): 536-539  
 URL JOURNAL: <http://www.liebertpub.com/CPB/default1.asp>  
 PUBLISHER: US: Mary Ann Liebert Publishers  
 URL PUBLISHER: <http://www.liebertpub.com/>  
 ISSN (PRINT VERSION): 1094-9313  
 LANGUAGE: English  
 ABSTRACT: (from the journal abstract) This study analyzes connections between online chatting and the Internet Addiction Disorder (IAD). In the research a compared two groups of respondents—Czech online chat users and Czech university students. The set of questions designed by Kimberly Young, of the Unive Pittsburgh, was the common base for research. The research showed that IAD is known among university students, that some of them spend an enormous t Internet, and that many of them feel it is a problem. The research also clearly showed that the most serious chat users more often have problems with IAD th university students. (PsycINFO Database Record (c) 2004 APA, all rights reserved)  
 KEY WORDS: [Internet Addiction Disorder](#), [online chat](#), [university students](#), [chatting](#), [Internet](#)  
 MAJOR DESCRIPTORS: [Addiction](#), [College Students](#), [Internet](#)  
 CLASSIFICATION CODE: Behavior-Disorders and Antisocial Behavior, 3230, 3  
 AGE GROUP: [Adolescence \(13-17 yrs\)](#), [Adulthood \(18 yrs and older\)](#), [Young Adulthood \(18-29 yrs\)](#), [Thirties \(30-39 yrs\)](#), [Middle Age \(40-64 yrs\)](#)  
 POPULATION: [Human](#), [Male](#), [Female](#)  
 POPULATION LOCATION: Czech-Republic

A

B

C

Malgrat que el document parla de l'addicció al xat, el mot *xat* no apareix entre els descriptors usats en la indexació (*Major descriptors*) (A). Tot i així, en aquest cas el podem recuperar gràcies al fet que a apareix en el resum (llenguatge natural) (B). En aquest cas, a més, també es pot recuperar perquè PsycInfo també indexa els documents en llenguatge lliure (descriptors lliures) (C).

L'ús de llenguatge lliure representa un esforç més gran per part de l'usuari, que haurà de pal·liar el seu caràcter equívoc. Serà per mitjà de l'estratègia de cerca que ho podrà fer, per exemple, considerant els possibles sinònims per a descriure el concepte buscat.

### Exemple dels drets d'autor

Recuperant l'exemple dels drets d'autor vist anteriorment, mentre que per mitjà de l'ús del llenguatge controlat el problema de la sinonímia desapareix en la mesura que és el mateix llenguatge qui regula l'ús de possibles equivalències; si la cerca es fa per mitjà del llenguatge lliure en el moment d'establir l'estratègia de cerca caldrà tenir en compte els diversos sinònims, per exemple, per mitjà de la construcció d'una equació de cerca que les aplegui:

drets d'autor OR propietat intel·lectual OR copyright

El llenguatge lliure s'utilitza molt en cerques fetes en cercadors d'Internet i en algunes bases de dades de text complet. Per contra, els llenguatges documentals controlats els trobem especialment en catàlegs i bases de dades.

### Llenguatges controlats i lliures

| Llenguatges controlats  | Llenguatges lliures   |
|---|---|
| <b>Avantatges</b>   | <b>Avantatges</b>   |
| <ul style="list-style-type: none"> <li>• Un mateix terme per a denominar un mateix objecte.</li> </ul>  | <ul style="list-style-type: none"> <li>• Més ric que el llenguatge controlat.</li> <li>• No hi ha problemes a l'hora de representar nova terminologia.</li> <li>• Més fàcils d'usar. L'usuari no necessita aprenentatge.</li> </ul> |
| <b>Inconvenients</b>  | <b>Inconvenients</b>  |
| <ul style="list-style-type: none"> <li>• Es perd l'especificitat del llenguatge lliure.</li> <li>• El nombre de termes pels quals s'indexa un document és menor.</li> <li>• Resulta difícil determinar el tema del document i reduir-lo a pocs termes.</li> <li>• Problemes a l'hora de representar matèries actuals.</li> <li>• La creació i manteniment d'un llenguatge controlat, en especial els tesaurus, és car.</li> </ul> | <ul style="list-style-type: none"> <li>• Implica un esforç més gran per part de la persona que ha de buscar.</li> <li>• Els resultats que s'obtenen són més amplis i és necessària una revisió aprofundida.</li> </ul>              |

## 5. Els llenguatges documentals: una descripció

Per a abordar amb una mica més de detall la descripció dels llenguatges documentals seguirem la classificació basada en l'anàlisi dels llenguatges segons el control exercit sobre el vocabulari.

Com hem vist, la majoria dels autors coincideixen a dividir els llenguatges documentals segons el control del vocabulari en les tipologies següents:

- **Llenguatges lliures.** Estan formats per llistes de paraules construïdes *a posteriori* a partir de la indexació dels documents.

### Llenguatges lliures

Presenten dues tipologies: les principals llistes de paraules clau i les llistes de descriptors lliures.

- **Llenguatges controlats.** Són llistes de termes, tancades i normatives, construïdes *a priori*. Defineixen tots els termes, estableixen aquells termes que es poden usar per a la descripció dels documents i disposen les relacions entre termes.

### Llenguatges controlats

Són llenguatges controlats els tesaurus, llistes d'encapçalaments i classificacions.

En el cas de les classificacions, no estem davant de llistes de termes, sinó de codis numèrics o alfanumèrics que identifiquen conceptes.

### 5.1. Llistes de paraules clau

Les llistes de paraules clau són llistes no estructurades de termes significatius seleccionats automàticament per sistemes informàtics per a descriure el contingut dels documents. Habitualment els termes s'obtenen del títol, el resum... i, cada cop més, del text complet. En aquest cas, qualsevol paraula del document és susceptible de ser usada en la recuperació de la informació.

- Estan constituïts per paraules soltes (biblioteca, biblioteques, centre, documentació...).
- De totes les categories gramaticals (noms comuns, noms propis, adjectius, verbs...).

- En totes les formes gramaticals: masculí, femení, singular i plural, conjugacions dels verbs...
- Totes les variants ortogràfiques.
- Paraules en totes les llengües dels documents introduïts en el sistema documental.

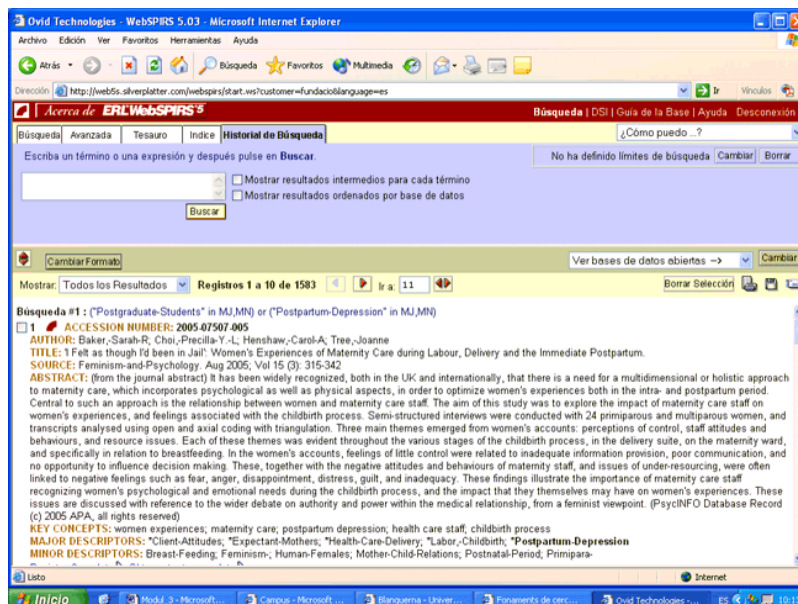
Presenten l'avantatge que apleguen termes molt específics i actuals, però la manca de control els fa estar subjectes a l'ambigüitat del llenguatge natural.

## 5.2. Llistes de descriptors lliures

Les llistes de descriptors lliures són llistes no estructurades de conceptes representats per paraules o expressions que han estat assignades mitjançant un procés intel·lectual pel documentalista. Aquest selecciona del llenguatge natural les paraules que considera més significatives i exclou els casos més evidents de polisèmia i sinonímia.

### Psycinfo

A la base de dades Psycinfo, els documents són indexats usant un tesaurus (camps MAJOR i MINOR DESCRIPTORS), però també usen descriptors lliures (camp KEY CONCEPTS) per a completar la indexació realitzada mitjançant el llenguatge controlat.



Malgrat no estar davant d'un llenguatge controlat, les llistes de descriptors lliures són una primera aproximació al control dels vocabularis.

Després de l'anàlisi del document, els descriptors són atribuïts per documentalistes –i no per un sistema informàtic com en el cas de les paraules clau. Els descriptors poden ser unitermes (*biblioteca*) o expressions (*centre de documentació*), generalment en una única llengua, independentment de l'idioma dels documents.

Habitualment els termes seleccionats són noms i s'intenta evitar formes adjetivades o verbals, i en singular. Malgrat que en els casos més evidents es pot intentar evitar la sinonímia i la polisèmia, a causa de la inexistència de control sobre els termes, l'homogeneïtat entre ells no existeix.

### 5.3. Tesauros

Un tesauro és una llista estructurada de conceptes, destinats a representar de manera unívoca el contingut dels documents i de les consultes dins d'un sistema documental determinat, i a ajudar l'usuari en la indexació dels documents i de les consultes (Van Slype, 1991, pàg. 24).

A diferència dels llenguatges lliures, la llista de termes que constitueix el tesauro està establerta *a priori*. En els tesauros trobem dos tipus de termes:

- Els **descriptors** són els termes escollits per a expressar de forma inequívoca un concepte. Com hem vist, poden estar formats per una o més paraules i són els únics que es poden utilitzar per a indexar els documents i les consultes.
- Els **no descriptors** són termes que no es poden usar en la indexació, però que mitjançant les relacions d'equivalència ens remeten a les formes acceptades (descriptors).

#### Exemples de descriptors i no descriptors

Descriptor: *Entrevista*

No descriptor: *Interviu*

Els tesauros presenten una estructura semàntica que estableix relacions entre termes, fonamentalment relacions d'equivalència, de jerarquia i d'associació:

1) Les **relacions d'equivalència**. S'utilitzen entre termes que es consideren equivalents. En el moment de l'elaboració del tesauro es decideix quin terme serà el terme principal, el terme acceptat (descriptor), i quina o quines les formes no acceptades (no descriptors). Les relacions d'equivalència permeten evitar l'ambigüïtat terminològica i la sinonímia. Les relacions d'equivalència estan identificades per l'expressió USEU o USAT PER (o altres equivalents idiomàtics o gràfics):



**Relacions d'equivalència**

ENTREVISTA

USAT PER: INTERVIU

...

INTERVIU

USEU ENTREVISTA

2) **Relacions jeràrquiques.** Estableixen relacions de generalitat o amplitud i d'especificitat/partitiu entre descriptors.

L'establiment de la jerarquia pot facilitar en el moment de la indexació l'elecció dels descriptors més específics.

La relació genèrica estableix que el descriptor A és més específic que el descriptor B, genèric. Per tant, els descriptors genèrics designen un concepte que engloba altres conceptes més específics.

Les sigles TG identifiquen la relació d'amplitud (terme genèric) i TE la relació d'especificitat/partitiva (terme específic).

**Relacions jeràrquiques**

GÈNERE PERIODÍSTIC

TE: ENTREVISTA

.....

ENTREVISTA

TG: GÈNERE PERIODÍSTIC

ÀSIA

TE: INDONÈSIA

...

INDONÈSIA

TG: ÀSIA

## Tesaurus de la UNESCO



Disponible a: <http://databases.unesco.org/thesaurus>.

3) **Relació d'associació.** És una relació simètrica entre dos descriptors que designen conceptes afins, però que no presenten una relació d'equivalència semàntica o jeràrquica. Aquesta relació es representa mitjançant la sigla TR (terme relacionat).

### Relacions d'associació

AVIACIÓ

TR: AVIÓ

GOMA

TR: ELASTICITAT

COTXE

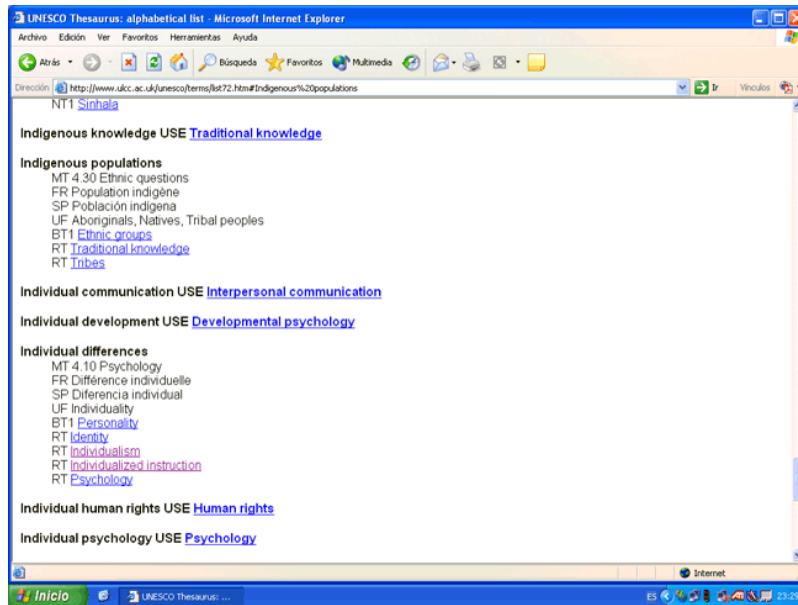
TR: XASSÍS

Els termes que componen un tesaurus i les seves relacions es representen habitualment de tres formes:

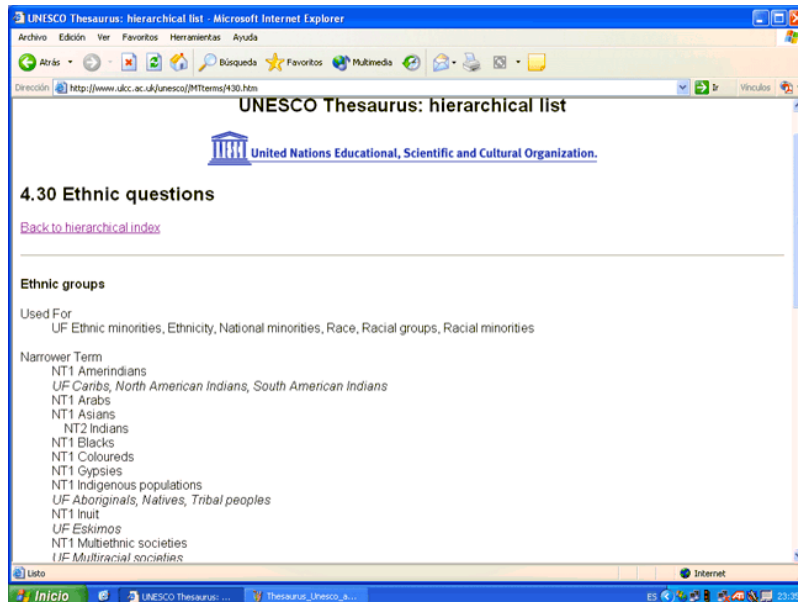
- Alfabètica.
- Sistemàtica.
- Gràfica.

### Presentació alfabètica i sistemàtica de la UNESCO

Tesaurus de la UNESCO. Presentació alfabètica



Tesaurus de la UNESCO. Presentació sistemàtica



## 5.4. Llistes d'encapçalaments

Una llista d'encapçalament és una llista alfabètica de termes, construïda *a priori*, que permet la representació unívoca dels conceptes tractats en els documents i de les consultes. Com en el cas dels tesaurus, únicament els termes inclosos en aquestes llistes es poden usar per a la indexació.

Les llistes estan formades per dos tipus de termes:

1) Els **encapçalaments**, constituïts per una o diverses paraules que representen conceptes. Entre els encapçalaments també s'estableixen relacions, com en els tesaurus, tot i que la xarxa de relacions semàntiques no és tan completa. Les relacions s'expressen mitjançant les referències següents:

- V. (Vegeu): Serveixen per a remetre d'un terme no acceptat a un encapçalament acceptat.
- U. p (Useu per): Serveixen per a identificar els encapçalaments no acceptats.
- V. a (vegeu a més): Remeten a encapçalaments relacionats i a partir dels quals l'usuari pot localitzar informació complementària.

2) Les **subdivisions**, combinades amb els encapçalaments, serveixen per a completar-los i concretar-ne l'abast.

#### **Les subdivisions**

Astronomia – Diccionaris

Ex. *Un diccionari sobre astronomia*

Automòbil – Comerç – Japó

*Un document sobre el comerç de l'automòbil al Japó*

Les llistes d'encapçalaments són llenguatges precontrolats perquè la combinació entre conceptes es produeix abans de la recuperació de la informació. En canvi, en catàlegs o bases de dades que usen tesaurus per a la indexació, la combinació entre descriptors la realitza l'usuari en el moment de la recuperació de la informació mitjançant l'ús dels operadors booleans, per això es diu que els tesaurus són llenguatges postcoordinats.

#### **Operadors booleans**

Automòbil AND Comerç AND Japó

## Lista d'encapçalaments de matèria en català

Index LEMAC -- Biblioteca de Catalunya - Microsoft Internet Explorer

Arxivo Edició Ver Favoritos Herramientas Ayuda

Direcció <http://www6.gencat.net/bc/lemac/indexresults.aspx?FirstRec=5949>

Lista d'encapçalaments de matèria en català (LEMAC)

Resultats de la cerca

| Matèria                                 | Tipus       |
|---|-------------|
| Aviació                                 | Acceptat    |
| Aviació -- Accidents                    | Acceptat    |
| Aviació -- Accidents -- Responsabilitat | Acceptat    |
| Aviació -- Assegurances                 | No acceptat |
| Aviació -- Dret i legislació            | No acceptat |
| Aviació -- Legislació                   | No acceptat |
| Aviació -- Mesures de seguretat         | Acceptat    |
| Aviació -- Pilotatge                    | Acceptat    |
| Aviació -- Psicologia                   | Acceptat    |
| Aviació -- Viatges                      | No acceptat |

Enrera Endavant Nova cerca

? i

Biblioteca de Catalunya, 2002

http://www6.gencat.net/bc/lemac/viewrecord.aspx?strRec=0394-53780&strSubject=Aviació--Accidents&strQualif=1

Registre LEMAC -- Biblioteca de Catalunya - Microsoft Internet Explorer

Arxivo Edició Ver Favoritos Herramientas Ayuda

Direcció <http://www6.gencat.net/bc/lemac/viewrecord.aspx?strRec=0143-65580&strSubject=Aviació&strQualif=1>

Lista d'encapçalaments de matèria en català (LEMAC)

Resultats de la cerca: Aviació

Encapçalament acceptat:

Aviació (Subd. geog.)

Nota d'abast:

Useu-lo per a les obres tècniques i pràctiques. Les obres teòriques s'apleguen sota l'encapçalament Aeronàutica

Termes genèrics:

Comunicacions

Termes relacionats:

Aeronàutica  
Avions

Termes específics:

Aeronaus  
Aerovies  
Aviadors  
Hidroavions  
Línies aèries  
Navegació aèria  
Vol

Detall de la descripció de l'encapçalament Aviació

Disponible a: <http://www6.gencat.net/bc/lemac/>

## 5.5. Classificacions

Els llenguatges de classificació representen el contingut mitjançant codis numèrics o alfanumèrics que corresponen a les categories o classes en què prèviament s'ha dividit l'àmbit del coneixement (Pérez, Adoració, 1998).

Les classificacions permeten, doncs, representar de manera sintètica el contingut d'un document, però també reunir físicament tots els documents que comparteixen una afinitat de continguts.

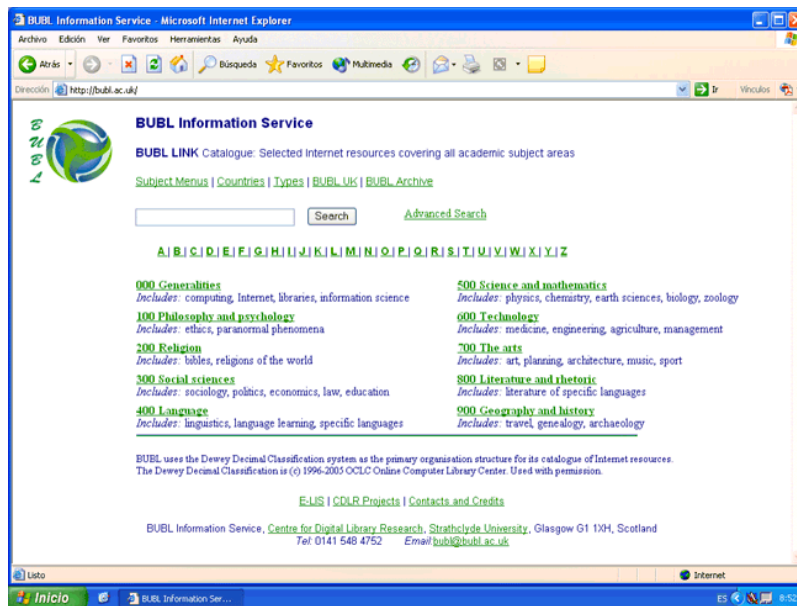
Les classificacions més usades en l'actualitat són la Dewey Decimal Classification (DDC), la Library of Congress Classification (LCC) i la Classificació Decimal Universal (CDU). Aquesta darrera és la més usada a les biblioteques catalanes.

### **Classificació decimal**

- 6 Ciències aplicades, medicina, tècnica
- 65 Gestió i organització de la indústria, el comerç i les comunicacions
- 659 Publicitat
- 659. 1 Publicitat i propaganda
- 659.11 Campanyes de publicitat
- 659.12 Concepció i disseny de la publicitat
- 659.13./16 Mitjans publicitaris /suports de la publicitat
- 659.13 Mitjans visuals
- 659.131 Publicitat en diaris
- 659.132 Publicitat per mitjà de publicacions pròpies
- 659.133 Publicitat en cartells i rètols
- 659.14 Mitjans acústics de publicitat
- 659.144 Publicitat oral
- 659.148 Publicitat audiovisual

Malgrat que les classificacions són un recurs que permet la recuperació dels documents, en els darrers anys, la seva principal funció s'ha limitat a l'ordenació física dels documents. Tanmateix, alguns directoris temàtics d'Internet usen classificacions per a organitzar els seus continguts. Així, per exemple, BUBL usa la classificació decimal de Dewey (DDC).

## Organització de continguts de BUBL



BUBL organitza els seus continguts a partir de la DDC.

### 5.6. Llistes d'autoritats

S'ha vist repetidament que una de les principals preocupacions dels bibliotecaris i documentalistes és la unificació de les formes dels punts d'accés als documents, ja siguin noms de persones, entitats o matèries. Aquest procés d'unificació i normalització es coneix amb el nom de *control d'autoritats*, entenent autoritat com un nom, un títol uniforme, una entitat, una matèria...

La finalitat del control d'autoritats és facilitar la identificació i la recuperació dels documents, evitant les confusions derivades de l'existència de termes sinònims, homònims... o de les diverses variants en què pot ser denominada una persona o una entitat. El fruit del control d'autoritats són les llistes d'autoritats, un conjunt limitat de termes, habitualment ordenat alfabèticament.

Sovint un catàleg es fonamenta en diverses llistes d'autoritats, de matèria (descriptors acceptats pels llenguatges d'indexació), de persones, d'entitats o institucions, noms geogràfics, etc.

#### Consulta d'una llista d'autoritats

Consulta de l'autoritat Vicens Vives a la llista d'autoritats de la Library of Congress:

The screenshot shows the Library of Congress Authorities website in a Microsoft Internet Explorer browser. The page title is "LIBRARY OF CONGRESS AUTHORITIES". The search results are for "Name Authority Headings = vicens vives". The results table is as follows:

| # | Bib<br>Records | select icon in first column to...<br>View Authority Headings/References | Type of Heading |
|---|----------------|---|-----------------|
| 1 | 0              | Vicens-Vives, Editorial   | corporate name  |
| 2 | 0              | Vicens Vives, J. (Jaime)  | personal name   |
| 3 | 41             | Vicens Vives, Jaime   | personal name   |
| 4 | 22             | Vicens Vives, Jaime. [from old catalog]                                 | personal name   |
| 5 | 0              | Vicens Vives, Jaume   | personal name   |
| 6 | 0              | Vicens y Hualde, Ignacio  | personal name   |
| 7 | 1              | Vicens y Marcó, S. P.   | personal name   |
| 8 | 1              | Vicens y Marcó, S. P., [from old catalog]                               | personal name   |
| 9 | 0              | Vicent, Alfredo   | personal name   |

La forma acceptada (*authorized heading*) per Jaume Vicens Vives és Vicens Vives, Jaime.



## Bibliografia

**Pérez, A.** (coord) (1998). *La documentació a l'era de la informació*. Barcelona: Universitat Oberta de Catalunya.

**Pinto Molina, M.** (1989). "El análisis documental". A: J. López Yepes (comp.). *Fundamentos de información y documentación*. Madrid: Eudema.

**Slype, G. Van** (1991). *Los Lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Madrid: Fundación Germán Sánchez Ruipérez / Pirámide.

