

Els tesaurus

Damián Martínez Ferreras

PID_00143943



Universitat Oberta
de Catalunya

www.uoc.edu



Els textos i imatges publicats en aquesta obra estan subjectes –llevat que s'indiqui el contrari– a una llicència de Reconeixement-NoComercial-SenseObraDerivada (BY-NC-ND) v.3.0 Espanya de Creative Commons. Podeu copiar-los, distribuir-los i transmetre'ls públicament sempre que en citeu l'autor i la font (FUOC. Fundació per a la Universitat Oberta de Catalunya), no en feu un ús comercial i no en feu obra derivada. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.ca>

Índex

Objectius	5
1. El tesaurus	7
2. Estructura del tesaurus	9
2.1. Unitats lèxiques	9
2.2. Relacions semàntiques	10
2.3. Notes i aclariments	13
3. Formes de presentació del tesaurus	14
3.1. Presentació alfabètica	14
3.2. Presentació sistemàtica	15
3.3. Presentació gràfica	16
3.4. Presentació amb índexs permutats	18
4. Creació i utilització d'un tesaurus	21
4.1. Creació d'un tesaurus	21
4.2. Utilització d'un tesaurus	26
4.2.1. Ús del tesaurus en els processos d'indexació i recuperació de la informació	27
4.2.2. Revisió i actualització del tesaurus	28
5. Sistemes automatitzats de gestió de tesaurus. Tesaurus en línia	29
5.1. Sistemes automatitzats de gestió de tesaurus	29
5.2. Els tesaurus al món digital. Tesaurus en línia	30
6. Tesaurus i ontologies	32
6.1. Els tesaurus	32
6.2. Les ontologies	33
6.3. Similituds i diferències entre tesaurus i ontologies	35
Bibliografia	37

Objectius

Amb l'estudi dels materials associats a aquest mòdul didàctic assolireu els objectius següents:

- 1.** Saber què és un tesaurus i conèixer-ne l'estructura i els principals elements que el componen i el defineixen: construcció, manteniment, tipologies, formes de representació i usos del tesaurus.
- 2.** Analitzar el paper dels tesaurus en el procés d'indexació i recuperació de la informació i la seva importància en els sistemes documentals especialitzats.
- 3.** Conèixer el procés de creació d'un tesaurus i elaborar un microtesaurus.
- 4.** Entendre la funció del tesaurus com una eina específica que fa possible l'objectiu final de la cadena documental, això és, la recuperació i difusió de la informació.

1. El tesaurus

Etimològicament, el terme prové del llatí *tesaurus*, que significa ‘tresor’ i té poca relació amb el concepte que actualment fa servir el professional de la informació.

La norma ISO 2788-1986 defineix un tesaurus com:

Un vocabulari controlat i dinàmic, compost per termes que tenen entre ells relacions semàntiques i genèriques i que s’aplica a un domini particular del coneixement.

Per la seva banda, van Slype (1991) defineix un tesaurus com:

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.

Podríem definir un tesaurus com una llista de termes emprats per a representar els conceptes, temes o continguts dels documents i que guarden entre ells relacions semàntiques i genèriques. Es tracta d’un instrument de normalització i control del vocabulari l’objectiu del qual rau a eliminar l’ambigüitat del llenguatge, convertint el llenguatge natural dels documents en un llenguatge controlat.

Estem parlant del tesaurus com un **vocabulari controlat de descriptors** amb un significat prèviament establert la finalitat última del qual és definir de manera unívoca el contingut d’un document. L’objectiu final d’aquest llenguatge és ser un pont de connexió entre els usuaris i les unitats d’informació, això és, servir tant per a la indexació com per a la recuperació documental.

A diferència d’altres llenguatges documentals, com les classificacions i les llistes d’encapçalaments de matèria, els tesaurus solen ser llenguatges especialitzats, fan referència a una temàtica concreta i estan construïts per servir a una institució o organització determinada.

Davant els llenguatges classificadors, la funció dels quals és descriure el tema d’un document, els termes continguts en un tesaurus responen a l’anàlisi del text o a la seva matèria. Un tesaurus recull tots els conceptes, i cada un d’aquests conceptes constitueix un punt d’accés per a la recuperació del document.

La **potència d'un tesaurus** es basa, a més, en la possibilitat de combinar tots aquests conceptes, anomenats *descriptors*, fet que el converteix en un llenguatge combinatori molt més ric que els encapçalaments de matèries tradicionals a l'hora de fer els processos d'anàlisi i recuperació de la informació.

Les **principals característiques** d'un tesaurus són les següents:

- a) Neix per la necessitat de crear nous sistemes d'indexació més flexibles que els tradicionals.
- b) Presenta una especificitat més gran dels termes relatius a camps disciplinars concrets, cosa que permet una anàlisi més precisa dels documents.
- c) A partir de l'ordenació alfabètica i com a complement a l'ordenació sistemàtica dels descriptors (per temes i significat) el tesaurus facilita la consulta de l'analista i de l'usuari. És precisament aquesta xarxa de relacions semàntiques que s'estableixen entre els descriptors (relacions d'equivalència, de jerarquia i d'associació) les que defineixen un tesaurus i proporcionen l'ajuda necessària en els processos d'indexació i recuperació d'informació.

Els tesaurus poden ser dels **tipus** següents:

- **Generals o especialitzats**, segons el seu camp d'aplicació.
- **Multidisciplinaris o monodisciplinaris**, segons el nombre de disciplines o temes que inclouen.
- **Macrotesaurus o microtesaurus**, segons la naturalesa i compatibilitat amb altres tesaurus. Els macrotesaurus, tot i que tenen les característiques d'un tesaurus quant a estructura i funció, tenen camps semàntics o microdisciplines diferents que poden funcionar com si fossin tesaurus independents, amb total autonomia. En aquest sentit, es podria parlar d'un conjunt de tesaurus més petits o microtesaurus.
- **Monolingües o multilingües**, segons les llengües utilitzades.
- **Alfabètics o sistemàtics**, segons la presentació utilitzada per a mostrar els termes i les seves relacions semàntiques.

Vegeu també

Tractarem els tesaurus alfabètics o sistemàtics en l'apartat 3 d'aquest mòdul didàctic.

2. Estructura del tesaurus

L'estructura conceptual dels tesaurus no és arbitrària en cap sentit, ja que existeixen tant manuals d'elaboració de tesaurus com normes internacionals que recomanen els elements estructurals a tenir en compte. La consulta tant de manuals com de normes amb relació a la construcció de tesaurus ens porta a una estructura conceptual que implica relacions semàntiques i on es tenen en compte els termes com a representacions lèxiques dels conceptes.

En conseqüència, podem basar l'estructura d'un tesaurus en els elements següents: unitats lèxiques, relacions semàntiques i notes i aclariments.

2.1. Unitats lèxiques

Les **unitats lèxiques** són grups de descriptors que agrupen els termes d'indexació bé per camps (temes) bé per classes de termes (facetes).

Als tesaurus, els descriptors representen conceptes o nocions que poden ser concrets, abstractes, entitats individuals o classes. Per a expressar els conceptes, els tesaurus contenen unitats lèxiques anomenades *descriptors* i *no descriptors*.

1) **Descriptors**: termes permesos. Un *descriptor* és una paraula o un conjunt de paraules del llenguatge natural que responen a un concepte, expressat amb un únic terme i aquest terme respon a un únic concepte. Aquesta univocitat s'aconsegueix seleccionant el descriptor entre els seus sinònims i establint una xarxa de relacions semàntiques amb la resta de descriptors del tesaurus. El terme serà utilitzat per a l'anàlisi i la recuperació del document. Poden ser:

a) **Termes simples**: s'empren quan el concepte és clar en si mateix, sense necessitat d'afegir cap altra paraula, o quan s'usa com a genèric de termes més concrets o específics, com, per exemple **Cotxes**.

b) **Termes compostos**: la majoria dels termes d'un tesaurus adquireixen el seu veritable sentit quan es converteixen en termes compostos, ja que així s'aconsegueix una major especificitat. L'estructura sintàctica dels termes compostos sol ser la següent:

- Substantiu + adjectiu: **Cotxes esportius**.
- Substantiu + sintagma preposicional: **Cotxes d'època**.
- Substantiu + sintagma aposicional: **Cotxes F1**.

2) **No descriptors:** termes prohibits. Paraules incloses al tesaurus, que pertanyen a una llista de sinònims i termes emparentats units als descriptors per una relació d'equivalència semàntica, que són susceptibles d'aparèixer en els documents o en les preguntes, però que no són utilitzats per formular la interrogació al sistema. La inclusió d'aquests en un tesaurus té com a finalitat millorar la coherència de la representació dels documents o d'una pregunta en enviar-nos al terme d'indexació.

Quan un mateix concepte es pot expressar mitjançant dos o més sinònims, un d'ells ha de seleccionar-se com a terme preferent o descriptor. Els altres passen a ser no descriptors.

Exemple

```
"Cotxe" (no descriptor)  
EMPREU "Automòbil" (descriptor)
```

Si busquem informació de vendes de cotxes ens trobarem amb el que es denomina *silenci documental*, ja que hi haurà documents que parlin sobre aquest tema però estiguin indexats utilitzant la paraula "Automòbil". En canvi, si busquem només per la paraula "Automòbil" només ens apareixeran documents que continguin aquesta paraula i no la paraula "Cotxe".

Aquest problema s'hagués evitat si al tesaurus hagués figurat la relació a dalt expressada i, en conseqüència, tant a l'hora d'indicar un document sobre vendes de cotxes com a l'hora de buscar informació sobre aquest tema, no haguéssim tingut cap altra opció que utilitzar el terme pertinent o descriptor, això és, el terme "Automòbil".

Així doncs, podem dir que els no descriptors formen part del tesaurus establint una relació d'equivalència amb el descriptor pertinent, però no compten ni per a la indexació ni per a la recuperació documental. Gràcies als no descriptors, descartem tot allò que podem denominar *accidents del llenguatge natural*, a fi de no retenir més que un terme referencial.

2.2. Relacions semàntiques

Els camps semàntics agrupen aquells descriptors adscrits al mateix àmbit de coneixement del tesaurus i aquesta agrupació es realitza per mitjà d'una relació de pertinença als esmentats camps o microdisciplines. Aquests no es poden utilitzar com a descriptors, això és, no intervenen en la indexació i recuperació dels documents, i només serveixen per a agrupar els descriptors afins a la seva temàtica.

Quant a les **relacions semàntiques** pròpiament dites, són totes aquelles relacions que s'estableixen entre els termes d'un tesaurus i tenen com a objectiu principal especificar el significat d'un descriptor assignant-lo a un camp semàntic o microdisciplina determinada.

Les relacions semàntiques establertes als tesaurus són de quatre tipus: de pertinença, d'equivalència, de jerarquia i d'associació. A continuació en veurem una descripció de cada una i recomanem recórrer als manuals i a les normes citades en la bibliografia per a una definició més detallada:

1) **Relacions de pertinença:** s'estableixen entre un descriptor i el camp semàntic a què pertany.

Notació

No existeixen sigles o abreviatures normalitzades per a expressar la relació de pertinença.

2) **Relacions d'equivalència:** l'equivalència semàntica entre descriptors s'aplica als sinònims, termes que tenen significats iguals en diferents contextos, i als quasisinònims, termes considerats de significat diferent en l'ús comú però tractats com si fossin sinònims en matèria d'indexació.

Notació

Les relacions d'equivalència se solen representar mitjançant el terme EMPREU (relació entre el terme no preferent i el terme preferent, o terme a usar més acceptat) i les sigles UP (Emprat per), que en anglès són USE i UF.

Notació usada en les relacions d'equivalència

	Català	Castellà	Anglès
Empreu	EM	USE	USE
Emprat per	EP	UP	UF

Descriptor	No descriptor
Automòbil	cotxe
EP cotxe	EMPREU Automòbil

S'aconsegueix així evitar l'ambigüitat terminològica de la sinonímia del llenguatge natural expressada en diversos termes per a referir-se a un mateix concepte.

3) Relacions de jerarquia: són les relacions que s'estableixen entre els descriptors. En una relació de jerarquia, un terme és superior o genèric d'un altre i aquest altre terme és inferior o específic de l'anterior. Aquesta estructura classificadora distingeix el tesaurus de les llistes alfabètiques de matèries o paraules clau, així com de diccionaris i lèxics. D'aquesta manera es creen camps conceptuals que classifiquen els termes i els agrupen en un camp delimitat de significats o facetes que es van agrupant en disciplines i subdisciplines.

Notació

Per a la presentació de les relacions jeràrquiques se solen utilitzar determinades sigles com ara TA (*terme ampli*) i TE (*terme específic*); o bé sigles en anglès, BT (*broad term* o *terme més ampli*), NT (*narrow term* o *terme més específic*).

Notació usada en les relacions de jerarquia

	Català	Castellà	Anglès
Terme genèric	TA (<i>terme ampli</i>)	TG	BT
Terme específic	TE	TE	NT

Descriptor	Descriptor
Productes lactis	logurt
TE iogurt	TA Productes lactis

Les relacions de jerarquia constitueixen el tret més característic dels tesaurus i proporcionen una ajuda importantíssima a l'usuari en els processos d'indexació i recuperació documental.

4) Relacions d'associació: són les relacions asimètriques establertes entre descriptors de famílies diverses, diferents a les jeràrquiques o les relacions d'equivalència, i que designen conceptes capaços d'evocar mútuament, per associació d'idees, relacions entre ells.

Notació

Aquest tipus de relació se sol representar mitjançant les sigles TR (*terme relacionat*) o en anglès RT (*related term*).

Notació usada en les relacions d'associació

	Català	Castellà	Anglès
Terme relacionat	TR	TR	RT

Descriptor	Descriptor
Càrrecs municipals	Administració Central
TR Administració Central	TR Càrrecs municipals

La majoria de les relacions associatives s'estableixen entre termes que comparteixen alguna característica comuna però que pertanyen a camps semàntics diferents.

Ajuden l'usuari en les tasques d'indexació i recuperació documental ja que proporcionen nous descriptors relacionats que faciliten la formulació de les tasques esmentades.

2.3. Notes i aclariments

Quan el significat d'un descriptor és difús o no queda ben precisat pel seu context semàntic o lèxic es pot afegir una explicació addicional relacionada amb el descriptor, de manera que no es produeixi ambigüitat dins del tesaurus.

Aquestes explicacions poden ser notes (explicatives, històriques o d'aplicació) o modificadors:

1) **Notes.** Utilitzen com a notació NA:

Exemple

INSTAL·LACIONS ESPORTIVES

NA Empreu aquest terme per a fer referència a instal·lacions dels diversos esports des del punt de vista de les seves característiques i descripció. Per a fer referència a aspectes econòmics i de gestió, empreu EQUIPAMENTS ESPORTIUS

2) **Modificadors.** Afegixen entre parèntesi l'explicació per evitar l'ambigüitat.

Exemple

Jocs Olímpics (Hivern).

3. Formes de presentació del tesaurus

Existeixen diferents formes de presentar els termes que componen un tesaurus i les relacions que s'estableixen entre aquests termes dins del tesaurus. Els principals tipus de presentacions són les següents: alfabètica, sistemàtica, gràfica i amb índexs permutats.

3.1. Presentació alfabètica

En la presentació alfabètica tots els termes del tesaurus, descriptors i no descriptors, es troben agrupats en una sola seqüència alfabètica, acompanyats de les seves relacions. S'utilitza principalment en el moment de la indexació dels documents ja que un simple cop d'ull al llistat alfabètic ens permet identificar els descriptors pertinents.

En la presentació alfabètica, les informacions associades a cada descriptor (informacions que especifiquen les relacions semàntiques del tesaurus) van precedides de les seves abreviatures o notacions respectives de la manera següent:

Informació associada i descriptors usats en la presentació alfabètica

Descriptor	Informació associada
DOM	Camp semàntic o microdisciplina a què pertany el descriptor (relació de pertinença)
NA	Nota d'aclariment per a evitar l'ambigüitat del terme
EP	<i>Emprat per.</i> El terme preferent o descriptor fa una referència al no descriptor o terme no preferent (relació d'equivalència)
EMPREU	El terme no preferent o no descriptor fa referència al descriptor (relació d'equivalència)
TC	Terme capçalera no és d'ús obligatori (relació de jerarquia)
TA	Terme genèric que fa referència al terme al qual pertany el descriptor en qüestió (relació de jerarquia)
TE	Terme específic que fa referència al terme que s'inclou dins del terme en qüestió (relació de jerarquia)
TR	Terme relacionat que fa referència al terme que té relació amb el descriptor que ens ocupa, però que no pertany al mateix grup del descriptor analitzat (relació d'associació)

Exemple

Begudes espirituoses

DOM Indústries

NA Es consideren begudes espirituoses aquelles begudes, exceptuant-ne els vins tradicionals i la cervesa, amb contingut alcohòlic procedents de la destil·lació de matèries primeres agrícoles.

UP Licors i aiguardents

TC Indústria alimentària

TA Indústria vinícola

TE Brand

Ginebra

Ron

Vodka

Whisky

TR Alcoholisme

Quant a les altres llengües, a continuació es presenta una taula amb les equivalències de les notacions:

Descriptor	Català	Castellà	Anglès
DOM	No és obligatori	No és obligatori	No és obligatori
NA	NA/NE	NA	SC
EP/EMPREDU	EP/EM	UP/USE	UF/USE
TC	TC	TC	TT
TG	TA	TA	BT
TE	TE	TE	NT
TR	TR	TR	RT

3.2. Presentació sistemàtica

La presentació sistemàtica, que també s'anomena *jeràrquica*, mostra els descriptors agrupats per camps semàntics o microdisciplines i, dins d'aquests, per cadenes jeràrquiques, en funció de les relacions jeràrquiques dels termes (terme genèric i/o específic). En la presentació sistemàtica els títols dels camps semàntics encapçalen la jerarquia i, a continuació, apareixen per ordre jeràrquic el conjunt de termes genèrics i específics.

L'ordenació dels descriptors que pertanyen al mateix camp semàntic es realitza per sagnats (un sagnat a la dreta per cada nivell jeràrquic) i per una ordenació alfabètica dels descriptors que pertanyen al mateix nivell jeràrquic i que, per tant, pertanyen al mateix descriptor genèric.

Exemple

Posem en negreta els termes emprats en tractar la representació alfabètica:

Indústries
Indústria alimentària
Indústria càrnica
Indústria conservera
Aliments congelats
Indústria de l'oli
Oli de gira-sol
Oli d'oliva
Indústria vinícola
Begudes alcohòliques
Begudes espirituoses
Brandi
Ginebra
Rom
Vodka
Whisky
Licors i aiguardents
Cava
Vi

3.3. Presentació gràfica

La representació gràfica permet visualitzar les relacions entre els termes i la seva estructura jeràrquica de manera molt efectiva. Les relacions es disposen com una figura que permet a l'usuari associar els termes que es troben relacionats.

Tanmateix, necessiten anar acompanyats d'un índex alfabètic, ja que és molt difícil representar aquelles relacions que no apareixen al gràfic, com poden ser els no descriptors o les notes explicatives. Es dona així la circumstància que la secció auxiliar de l'índex conté més informació que la part en teoria principal, la qual cosa converteix aquesta part gràfica en una presentació de suport.

Hi ha dos tipus de representacions gràfiques: l'estructura arborescent i el diagrama de fletxes.

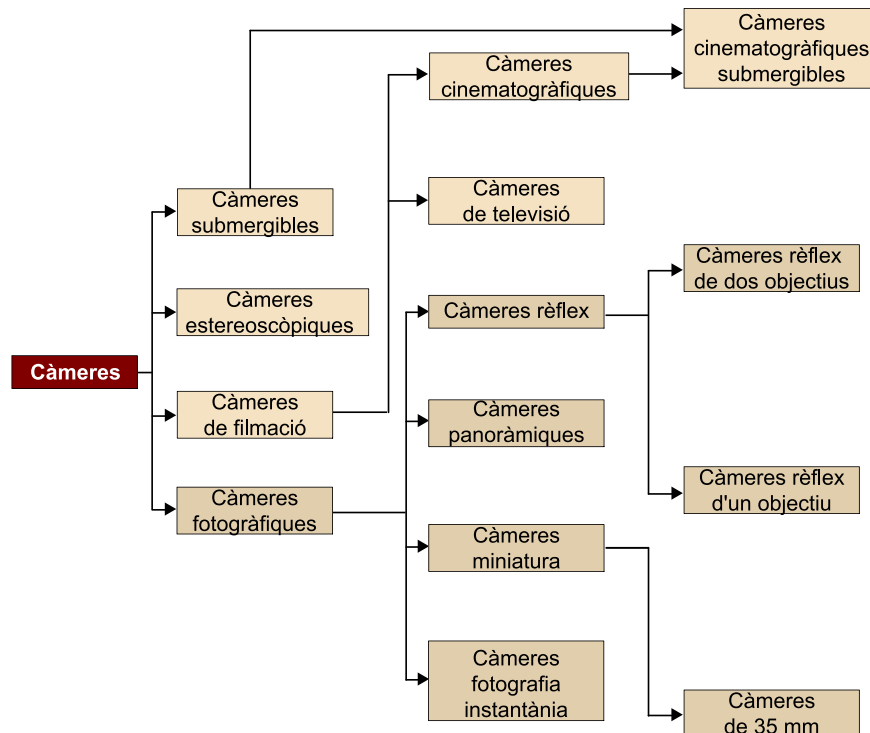
1) Estructura arborescent

Aquesta representació es caracteritza per la forma d'arbre, en què el terme principal se situa en una posició destacada del gràfic. A continuació els termes específics se situen en funció dels diferents graus de jerarquia que tenen.

Aquest tipus de presentació només permet visualitzar les relacions jeràrquiques i és necessari l'índex alfabètic per a entendre la resta de relacions que es donen entre els termes (d'equivalència, associació i notes d'aplicació).

Exemple

Figura 1. Exemple d'estructura arborescent



Font: Norma UNE 50-UNE-50-106-90 (pàg. 44).

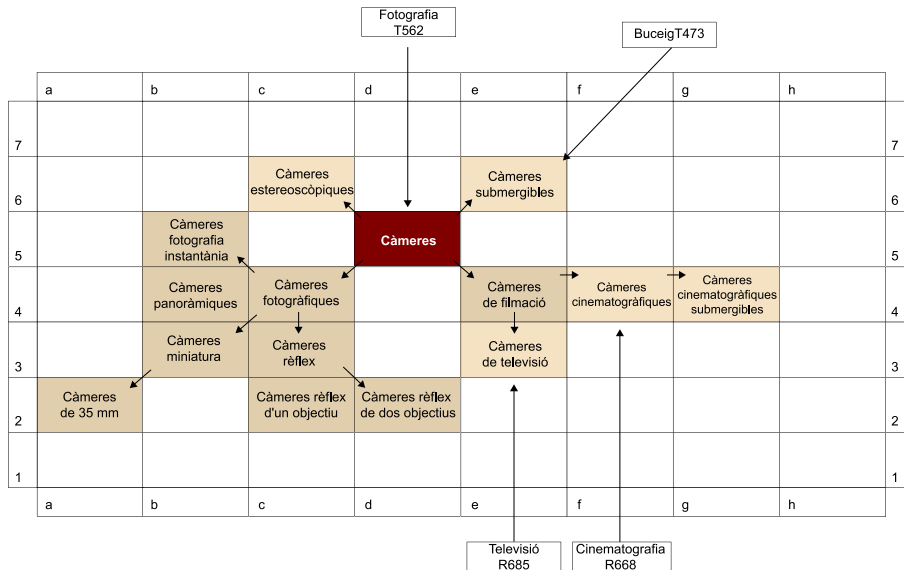
2) Diagrama de fletxes

En aquesta presentació cada camp semàntic apareix separatament i té una forma de presentació rectangular:

- En el centre del rectangle es troba el camp semàntic o terme principal i al seu voltant graviten els descriptors que pertanyen a l'esmentat camp semàntic, i també les relacions jeràrquiques i associatives que s'estableixen entre ells. Les relacions jeràrquiques s'indiquen per mitjà de fletxes que van dels termes més genèrics cap als més específics i les relacions associatives es fixen a través de línies rectes que connecten amb els descriptors relacionats.
- Fora del rectangle es troben els descriptors que pertanyen a altres camps semàntics però que estan relacionats amb els descriptors situats en l'interior del rectangle. Igual que en el cas anterior, s'utilitzen fletxes o línies rectes per a enllaçar amb els descriptors de dins del rectangle.

De la mateixa manera com ocorre amb la representació arborescent, la representació mitjançant el diagrama de fletxes necessita l'índex alfabètic per a entendre la resta de relacions (d'equivalència, associació i notes d'aplicació) que es donen entre els termes.

Figura 2. Exemple de diagrama de fletxes



Font: Norma UNE 50-UNE-50-106-90 (pàg. 46).

3.4. Presentació amb índexs permutats

Un índex és una sèrie ordenada de punts d'accés que ens remeten d'un terme conegut per l'usuari als conceptes continguts als documents. El seu objectiu se centra, per tant, en la recuperació documental. L'automatització dels sistemes documentals ha possibilitat la realització d'índexs més complexos, amb permutacions dels termes.

Els índexs permutats mostren la relació alfabètica de tots els termes significatius d'un tesaurus mitjançant l'agrupació dels descriptors afins des del punt de vista semàntic, facilitant d'aquesta manera la recuperació dels descriptors compostos independentment de la seva situació dins del descriptor. En definitiva, es recorre a una agrupació de descriptors afins des del punt de vista semàntic que en un altre tipus de representació apareixerien més distanciats.

L'aplicació d'aquests tipus d'índexs als descriptors d'un tesaurus ha generat nous tipus de presentacions dels seus termes. Entre aquests tipus d'índexs produïts per l'ordinador, destaquen els índexs KWIC, els KWOC i els KWAC.

1) Índexs KWIC

Els índexs KWIC (*key words in context*) mostren una relació de descriptors en la qual les paraules significatives s'ordenen alfabèticament en una columna central. La resta del text se cita abans i després de les esmentades paraules.

El KWIC crea l'índex de tal manera que una columna central ordenada de forma alfabètica mostra la paraula clau amb una tipografia destacada (aquesta representació es fa per a totes la paraules claus del document, és a dir, es produeix una permutació de termes per tal de deixar sempre a la columna central

el terme significatiu). Un mateix document apareixerà representat d'aquesta manera en el sistema tantes vegades com termes indexats posseeixi a l'índex KWIC. La notació de l'índex és numèrica.

Exemple d'índex KWIC

reajuste del funcionamiento de **COMPRESORES** centrífugos. Posibilidades de químicos - La selección de **COMPRESORES** para las industrias de procesos velocidad. Particularidades de **COMPRESORES** de procesos con rotor de gran de pirolisis - Obtención de **COQUE** electrodico a partir de productos líquidos Consideraciones críticas sobre **COSMÉTICOS** para la piel

2) Índexs KWOC

Als índexs KWOC (*key word out context*), els descriptors s'enumeren posant totes les paraules importants en ordre alfabètic, seguides pel nom sencer del descriptor.

Exemple d'índex KWOC

COMPRESORES

Posibilidades de reajuste del funcionamiento de **compresores** centrífugos
La selección de **compresores** para las industrias de procesos químicos
Particularidades de **compresores** de procesos con rotor de gran velocidad

COQUE

Obtención de **coque** electrodico a partir de productos líquidos de pirolisis

COSMÉTICOS

Consideraciones críticas sobre **cosméticos** para la piel

3) Índexs KWAC

L'índex KWAC (*key word and context*) és una barreja del KWIC i el KWOC.

Exemple d'índex KWAC**COMPRESORES**

reajuste del funcionamiento de **COMPRESORES** centrífugos. Posibilidades de químicos - La selección de **COMPRESORES** para las industrias de procesos velocidad. Particularidades de **COMPRESORES** de procesos con rotor de gran

COQUE

de pirolisis - Obtención de **COQUE** electródico a partir de productos líquidos

COSMÉTICOS

Consideraciones críticas sobre **COSMÉTICOS** para la piel

4. Creació i utilització d'un tesaurus

Abans d'iniciar la construcció d'un nou tesaurus hem de considerar les qüestions següents:

a) Comprovar si ja existeix algun tesaurus sobre la temàtica que ens interessa i si aquest pot servir per al nostre sistema documental.

b) Si decidim iniciar la construcció d'un de nou, haurem d'establir-ne les característiques principals, és a dir: temes que ha de tractar, mida, si serà monolingüe o multilingüe, tipus de relacions entre els descriptors i una altra sèrie de qüestions formals relatives a la tipografia dels termes, longitud dels mateixos, qüestions relatives al seu gènere, forma, etc.

4.1. Creació d'un tesaurus

La construcció del tesaurus inclou vuit o nou etapes, segons si és monolingüe o multilingüe (Normas UNE-50-106 y UNE 50-125):

- Recollida del vocabulari en llenguatge natural dins dels dominis que cobreixi el tesaurus.
- Subdivisió del conjunt dels dominis que es tindran en compte en una sèrie de microdisciplines.
- Transformació del vocabulari lliure en un llenguatge controlat, establint les relacions de pertinença, d'equivalència semàntica, de jerarquia, i redactant les notes explicatives.
- Recerca de les equivalències interlingüístiques (si es tracta d'un tesaurus multilingüe).
- Enriquiment del tesaurus per mitjà de relacions associatives.
- Elaboració de l'esborrany del tesaurus.
- Formació dels indicadors.
- Test del tesaurus.
- Revisió final i primera edició.

1) Recollida dels termes

Aquesta fase consisteix a buscar i registrar totes les paraules i expressions significatives que intervenen en la disciplina o disciplines incloses en el tesaurus.

Poden utilitzar-se algunes **fonts terminològiques alternatives** o complementàries:

- Un llenguatge documental ja existent en el servei de documentació: sistemes de classificació, llistes de paraules clau o de descriptors lliures, llistes d'autoritats.
- Tesaurus aliens al servei de documentació, però dedicats als mateixos dominis o a dominis pròxims, descoberts gràcies a algun dels repertoris de tesaurus disponibles al mercat. Un tesaurus ja existent resulta útil perquè aporta una col·lecció organitzada de conceptes de la qual es podrà extreure el que convingui, però poques vegades s'adoptarà en la seva totalitat per a organitzar el propi patrimoni informatiu. A Internet es poden localitzar molts tesaurus disponibles per a consultar-los.
- Terminologia utilitzada als tractats, manuals, lèxics especialitzats, reglamentacions tècniques publicades o específiques de l'organisme. Aquestes fonts tenen el gran avantatge d'oferir una terminologia generalment admesa i estructurada pels seus autors.
- Bases de dades terminològiques.
- Indexació intel·lectual, en llenguatge natural, per mitjà de descriptors lliures, d'una mostra de centenars de documents representatius per a cada un dels grans àmbits que inclourà el tesaurus. S'obté d'aquesta manera un inventari no ordenat de la terminologia efectivament utilitzada pels autors per a designar els conceptes que manegen habitualment; pot ser que sigui terminologia en diverses llengües, si es tracta d'una mostra multilingüe per a crear un tesaurus multilingüe.
- Indexació automàtica, en llenguatge natural, per mitjà de paraules clau, d'aquests mateixos documents. Aquest mètode és poc eficaç perquè la llista obtinguda només inclou paraules aïllades (unitermes), que només en alguns casos, quan estan aclarides per context, serveixen per a trobar els conceptes. A més, encara que s'utilitzi una gran llista de paraules buides, l'ordinador proporciona un percentatge extremadament elevat de termes sense cap interès discriminant.
- Consulta a especialistes vinculats amb l'organisme a què pertany el sistema documental. Aquesta font és indispensable per a obtenir la terminologia pròpia de les diverses pràctiques professionals dins de l'organisme, i també una explicació de l'accepció específica de certs termes per a cada un d'aquests especialistes.

2) Creació dels camps semàntics o microdisciplines

Amb tot el conjunt de termes recollits durant la fase anterior podem establir una llista provisional de microdisciplines que completarà el cos del tesaurus.

La **llista de les microdisciplines** té com a única finalitat preparar un sistema d'agrupació dels descriptors que sigui acceptable per a l'usuari i còmode de consultar. D'aquesta manera la llista així disposada serà utilitzada directament durant l'etapa següent i es revisarà contínuament al llarg de tot el procés de construcció del tesaurus.

3) Reducció a un llenguatge documental. En aquesta fase s'estableix les relacions entre els descriptors definits i s'introdueixen les notes d'aclariment escaients:

a) Relacions de pertinença: el primer que hem d'aconseguir és la distribució de tots els termes del llenguatge natural recopilats durant la primera fase en les diferents microdisciplines que hem establert en la segona. D'aquesta manera, tots els termes queden adscrits provisionalment a un camp semàntic o una microdisciplina.

b) Relacions d'equivalència: una vegada tenim tots els descriptors distribuïts entre les diferents microdisciplines, les hem d'examinar una per una a fi d'establir entre els seus conceptes el llistat de termes sinònims o quasisinònims. D'aquesta manera, comencem a seleccionar els termes que passaran a ser descriptors i els que seran no descriptors, i que haurem d'enllaçar per mitjà d'una relació d'equivalència semàntica.

Els criteris per a elegir el terme descriptor entre els seus sinònims s'han de fonamentar en la selecció del menys ambigu o del més utilitzat en la professió o pels usuaris del sistema documental. Probablement, haurem de consultar obres terminològiques per tal d'assegurar la correcció de l'ortografia d'aquestes paraules.

c) Notes explicatives: si el terme seleccionat com a descriptor encara pot provocar alguna ambigüitat dins del tesaurus, haurem de redactar la nota explicativa corresponent (de definició, històrica o d'aplicació).

d) Relacions jeràrquiques: el pas següent consisteix a establir les cadenes jeràrquiques entre descriptors genèrics i específics dins de cada camp semàntic o microdisciplina.

Partim del descriptor més genèric (terme capçalera) i desenvolupem en forma arborescent les diferents ramificacions fins a arribar als descriptors més específics. L'ordenació dels termes dins d'un mateix nivell jeràrquic pot fer-se per facetes o temes.

4) Recerca d'equivalències interlingüístiques

En el cas dels tesaurus multilingües, una vegada que s'ha fixat el contingut semàntic dels descriptors (gràcies a les relacions de pertinença, equivalència i jerarquia i a les notes explicatives), podem començar a establir les equivalències lingüístiques del terme en els diferents idiomes definits als tesaurus.

5) Establiment de relacions d'associació

El pas següent en el procés de construcció del tesaurus és l'establiment de les relacions d'associació entre els descriptors dels seus diferents camps semàntics. És, sens dubte, la fase més laboriosa, ja que s'ha d'examinar cada un dels descriptors del tesaurus a fi que trobi termes temàticament afins entre la resta dels descriptors.

També s'hauran de revisar els descriptors localitzats en altres camps semàntics amb què el descriptor que s'està examinant pugui tenir una relació d'associació. Si això es produeix, les relacions d'associació podran ser creades entre descriptors que pertanyin a cadenes jeràrquiques diferents dins de la mateixa disciplina.

En cap cas no es crearan relacions associatives:

- Descriptors de la mateixa cadena jeràrquica.
- Descriptors i no descriptors.

6) Elaboració de l'esborrany del tesaurus

L'elaboració de tot el procés anterior donarà lloc a una edició pilot o experimental del tesaurus. S'ha de procurar fer, almenys, una presentació alfabètica i una altra de jeràrquica.

Aquesta edició, o edició 0, està destinada a la formació dels indexadors o analistes i a la prova de funcionament del tesaurus.

L'exercici dura diversos dies i consisteix a:

- Presentar l'esborrany del tesaurus.
- Fer que els indicadors indexin els mateixos documents.

Després de la indexació de cada document:

- Es dialoga sobre els punts de vista adoptats per cada indexador, per fer que coincideixin progressivament.
- Es calcula la taxa de coherència (proporció entre el nombre de descriptors comuns i el nombre total de descriptors diferents utilitzats per dues persones o dos grups per a indexar el mateix document); al principi de l'exercici

la taxa serà escassa (entre el 20 i el 30% segons van Slype), però en acabar la formació tendirà cap al seu valor ideal (entre el 50% i el 80%).

7) Test del tesaurus

El test del tesaurus té com a finalitat la identificació i resolució dels problemes que presenta el tesaurus com a eina per a la indexació d'un sistema documental determinat. Totes les anomalies detectades pels indexadors durant la indexació d'una mostra significativa de documents s'assenyalen en un informe en forma de propostes per a:

- Afegir descriptors i no descriptors oblidats durant la construcció.
- Modificar les estructures jeràrquica i associativa.
- Explicitar per mitjà de notes explicatives descriptors encara ambigus.

Per a organitzar aquest test:

- a) Se seleccionen diversos centenars de documents i, si és possible, de consultes.
- b) Es fa que els documentalistes els indexin usant el nou tesaurus.
- c) Es demana als documentalistes que preparin un informe especial en el qual apareguin identificats els tipus de propostes descrites més amunt.
- d) Es realitza, finalment, una estadística d'aquestes propostes, que seran valorades pels responsables d'elaborar el tesaurus.

8) Revisió final i primera edició

Una vegada que s'han afegit i han estat revisades les propostes dels indexadors, estem en condicions de preparar l'edició definitiva del tesaurus.

L'edició d'un tesaurus ha d'incloure, en forma de prefaci o introducció, les informacions:

- Noms dels autors i del servei o centre responsable de l'edició.
- L'objectiu que es vol assolir amb la creació del tesaurus.
- El llistat de les microdisciplines que conté.
- Les fonts terminològiques utilitzades.
- El procés de construcció del tesaurus.
- Les característiques del tesaurus: llengües, nombre de descriptors, de no descriptors, de microdisciplines, de termes específics, de termes relacionats, etc.
- Alguns exemples de les presentacions del tesaurus: relacions incloses, abreviatures o signes utilitzats.
- Instruccions per a la indexació i la formulació de consultes.

- Informació sobre el manteniment del tesaurus: futures, perspectives, ús de sistemes automatitzats, etc.

Una vegada redactat el prefaci amb aquesta informació, podem iniciar la impressió o edició del tesaurus.

4.2. Utilització d'un tesaurus

El tesaurus és una eina viva i dinàmica que es nodreix de nous descriptors, gràcies a la indexació de nous documents i als canvis qualitius que el mateix procés de construcció genera. Per tant, el manteniment i l'actualització periòdica del tesaurus es converteix en una necessitat.

S'ha d'actualitzar periòdicament, tant per corregir errors i omissions com per adaptar-lo a l'evolució científica de la disciplina que inclou. El procés de control ha de seguir els passos següents:

1) Control de la freqüència d'ús dels descriptors

Per a dur això a terme, s'ha de confeccionar periòdicament un llistat dels descriptors amb el nombre d'ocurrències, és a dir, la quantitat de vegades que han estat utilitzats en la indexació dels documents.

L'elaboració d'aquests llistats, normalment facilitada pel *programari* de gestió documental, permet detectar:

- Descriptors que s'han utilitzat molt poc, la qual cosa indica que el descriptor és problemàtic.
- Descriptors que s'han utilitzat molt, la qual cosa també pot indicar problemes.

2) Detecció d'absència de descriptors

Amb l'ús del tesaurus en el procés d'indexació es poden detectar nous termes que apareixen als documents, però que no estan inclosos al tesaurus.

Generalment, la majoria dels sistemes documentals, especialment les bases de dades documentals, ja han previst aquesta incidència amb la inclusió d'un camp denominat *descriptors candidats* en la base de dades. Tots els conceptes que no són al tesaurus s'inclouen en aquest camp.

Periòdicament, els responsables del tesaurus elaboren una llista d'aquests descriptors candidats i estudien la seva possible inclusió en el tesaurus.

3) Problemes d'ús del tesaurus

Descriptors poc utilitzats

Això pot succeir, per exemple, perquè el descriptor és massa específic, no s'entén el seu significat, etc.

Descriptors massa utilitzats

Això pot succeir perquè el descriptor és massa genèric i, en conseqüència, cal crear descriptors més específics, etc.

Amb l'ús del tesaurus també es comencen a detectar diferents problemàtiques:

- Descriptors amb un significat molt pròxim i que, per tant, provoquen el dubte de l'indexador en el moment de seleccionar-los.
- Relacions semàntiques l'absència de les quals es nota. Per exemple, la d'un no descriptor que ajudaria a localitzar el descriptor admès.
- Divergències entre les diferents presentacions del tesaurus: un terme apareix adscrit a un camp semàntic determinat en la presentació jeràrquica, i a un altre en l'alfabètica.

La detecció d'aquests problemes s'ha de dur a terme amb la col·laboració estreta dels analistes, que són les persones que els troben durant l'anàlisi del document.

4.2.1. Ús del tesaurus en els processos d'indexació i recuperació de la informació

L'ús del tesaurus en el procés d'indexació d'un document té relació fonamentalment amb la traducció dels conceptes extrets del llenguatge natural a un llenguatge controlat, és a dir, amb la conversió dels termes del llenguatge natural en descriptors del tesaurus.

Ara bé, en tractar-se d'un llenguatge documental molt flexible, a l'hora d'indexar amb un tesaurus hem de seleccionar els conceptes rellevants del document de la manera més precisa possible, ja que l'estructura semàntica del tesaurus garanteix la coherència del significat de tots els descriptors. Aquesta circumstància influirà de manera definitiva en un procés de recuperació de la informació eficaç.

En aquest últim procés, això és, en el moment de procedir a la consulta i la recuperació pertinents de la informació, el tesaurus dirigeix l'usuari a l'hora d'interrogar la base dades o sistema documental concret ja que tradueix el llenguatge natural de l'usuari i el dirigeix dels conceptes o descriptors no acceptats als descriptors aptes per a procedir a una interrogació efectiva del sistema.

La finalitat de la seva estructura (relacions de pertinença i jerarquia) i de la seva xarxa de relacions (relacions d'equivalència i associació) permet ajudar l'usuari a construir la millor estratègia possible en funció de les seves necessitats d'informació, la qual cosa facilita una precisió i exhaustivitat altes en la recuperació.

4.2.2. Revisió i actualització del tesaurus

Després del seguiment de l'ús del tesaurus i la detecció dels seus principals problemes, és precís iniciar-ne la revisió i actualització. Convé que aquesta revisió sigui duta a terme per la persona o les persones administradores del tesaurus, ja que es tracta de l'única via per garantir el manteniment de la seva coherència.

La periodicitat d'aquesta revisió depèn de les característiques del centre i del tesaurus; pot ser cada sis mesos, anual o bianual.

També és important mantenir una nota històrica en els descriptors que han sofert modificacions, que inclogui la data de la introducció del canvi i el descriptor que s'utilitzava abans de la seva incorporació o modificació. Part del *programari* de gestió de tesaurus inclou entre les seves prestacions la possibilitat de mantenir aquesta nota històrica o fitxer LOG com una informació interna per als administradors.

Durant aquest procés de revisió, s'han de prendre decisions per a solucionar els problemes que pugui presentar el tesaurus:

a) Amb relació als descriptors que s'han utilitzat amb molta freqüència, es poden mantenir com fins al moment o crear-ne d'altres més específics.

b) Amb relació als descriptors que s'han utilitzat amb molt poca freqüència, es poden mantenir com fins al moment o suprimir-los del tesaurus.

c) Amb relació als descriptors candidats, es poden prendre les opcions següents:

- Si només s'han requerit en molt poques ocasions no seria necessari incloure'ls ja que podem utilitzar el descriptor més genèric. Una altra solució és incloure'ls com a no descriptors.
- Si la petició dels descriptors és freqüent, podem incloure'ls, establint prèviament les relacions de pertinença, jerarquia, equivalència i associació corresponents.

d) Un altre problema addicional es deriva de l'eliminació de descriptors que fan referència a termes obsolets que ja no s'usen. El problema d'aquesta operació és que si ja tenim documents indexats amb aquests descriptors, cal que els tornem a indexar a fi de mantenir la coherència entre el tesaurus i el sistema de consulta.

5. Sistemes automatitzats de gestió de tesaurus. Tesaurus en línia

5.1. Sistemes automatitzats de gestió de tesaurus

Els sistemes automatitzats de gestió de tesaurus faciliten la gestió i edició d'un tesaurus de manera automatitzada, i permeten d'aquesta manera reduir el temps i els costos a l'hora d'elaborar-los.

Les prestacions d'els esmentats són variades en funció del *programari* que s'utilitzi però, en línies generals, la seva utilització permet:

a) Incloure termes i registrar automàticament les relacions invertides amb altres termes del tesaurus.

b) Editar diferents presentacions del tesaurus (alfabètiques, sistemàtiques, índexs permutats) i seleccionar el tipus de relació que volem visualitzar a la llista (amb termes relacionats, notes explicatives, termes genèrics, etc. o sense ells).

c) Utilitzar-lo directament en la consulta. Si el *programari* forma part del sistema de gestió de la base de dades documental, pot permetre la selecció dels termes del tesaurus, l'execució immediata de la recerca i la recuperació dels documents pertinents.

d) Disposar de prestacions per als administradors del tesaurus com les següents:

- Possibilitat de mantenir un fitxer històric on es reflecteixin les modificacions i els canvis que es fan en els diferents descriptors.
- Facilitats a l'hora d'eliminar un descriptor del tesaurus (quan aquest desapareix, s'eliminen totes les relacions que tenia amb la resta dels termes).
- Elaboració automàtica d'estadístiques (de freqüència d'ús dels descriptors, del nombre de relacions existents al tesaurus, etc.).

Existeixen dos grans **tipus de sistemes de gestió de tesaurus**:

1) Els més simples estan associats a un sistema de gestió documental concret. Encara que contenen les prestacions bàsiques per a la gestió de tesaurus, presenten també algunes mancances quant a la identificació correcta de les relacions entre els descriptors.

2) Els més complexos són programes específics de gestió de tesaurus. Incorporen funcionalitats diverses, com la de validar automàticament els termes i les relacions entre ells, la qual cosa estalvia temps en la verificació de les esmentades relacions i evita cometre errors.

Exemples de programari de gestió de tesaurus

Recomanem consultar la bibliografia, però per posar algun exemple podem assenyalar els recursos següents segons si usem l'anglès o l'espanyol:

- **Multites** (en anglès): permet la conversió i generació de fitxers en llenguatge HTML i permet incloure el tesaurus a la xarxa Internet.
- **TemaTres** (en espanyol): és una aplicació web per a la gestió de llenguatges documentals. Es troba orientada especialment al desenvolupament de tesaurus jeràrquics, però també pot utilitzar-se per a desenvolupar estructures de navegació web, o com a complement articular amb un gestor de continguts, biblioteques digitals, o en una biblioteca tradicional per a gestionar els llenguatges documentals en ús.

5.2. Els tesaurus al món digital. Tesaurus en línia

En els últims temps, l'àmbit de la representació i recuperació d'informació ha hagut d'assumir l'impacte d'internet i les seves tecnologies associades. Les conseqüències d'aquests canvis tecnològics estan conduint a una digitalització de l'àmbit de la representació i recuperació d'informació que afecta igualment els recursos d'informació, les eines de representació i recuperació, i els requeriments dels usuaris.

En aquest sentit es poden destacar els fets següents:

- 1) Increment del nombre de recursos d'informació que requereixen descripció, identificació i elements de localització, és a dir, una anàlisi formal i de contingut.
- 2) Amb relació a l'anàlisi de contingut es fa necessari un replantejament quant a les eines de gestió i organització d'informació entre les quals es troben els vocabularis controlats.
- 3) És cada vegada més important proporcionar als usuaris unes estructures de coneixement que els permetin un accés ràpid i amigable a la informació. En aquest sentit, el tesaurus, com a vocabulari controlat, pot exercir un paper protagonista: els usuaris requereixen eines conceptuals i semàntiques destinades a una organització de la informació digital efectiva, i els tesaurus són una de les estructures que poden tenir una participació activa en aquest camp.

D'altra banda els tesaurus es veuen significativament beneficiats per l'entorn digital ja que afavoreixen el desenvolupament dels aspectes següents:

- Noves possibilitats quant a les pautes de disseny, elaboració, gestió i ús de les eines utilitzades per a la representació i recuperació d'informació.

- Desenvolupament d'una estructura hipertextual on s'estableixin enllaços entre tots els elements estructurals (descriptors, no descriptors, notes d'abast, etc.), i també entre les diferents parts del tesaurus.
- Simplificació de l'estructura a través de la hipertextualitat, facilitant per tant el seu maneig a partir d'interfícies adequades.
- Reducció dels costos d'actualització.
- Potenciació del desenvolupament de tesaurus multilingües i multidisciplinaris fomentant la cooperació internacional, i, per tant, impulsant la reutilització i interoperabilitat de recursos.
- Accés universal a eines terminològiques que afavoreixen la recuperació d'informació.
- Integració de l'usuari al procés de creació, gestió i optimització de les eines conceptuals documentals.

Enllaç d'interès

Recomanem visitar la pàgina de Willpower, empresa de consultoria independent especialitzada en tots els temes relacionats amb la gestió de la informació i que proporciona exhaustiva i abundant informació relativa a la construcció, gestió, manteniment i utilització de tesaurus, i també nombroses referències a tesaurus de tot el món, normativa, institucions i gestió de tesaurus.

6. Tesaurus i ontologies

Amb relació al procés assenyalat en l'apartat anterior relatiu a la digitalització creixent de l'àmbit de la representació i recuperació d'informació segons Arano (2005), són diverses les eines que han estat objecte d'estudi, igual com ho han estat els camps de coneixement on aquestes eines s'han originat: la lingüística, la intel·ligència artificial, la documentació, l'enginyeria lingüística, etc.

Pel que fa al nostre objecte d'estudi, això és, la representació i recuperació d'informació, aquestes eines són, entre altres, les següents:

- Sistemes de classificació.
- Llistes d'encapçalament.
- Bases de dades lèxiques.
- Bases de coneixement.
- Taxonomies.
- Tesaurus.
- Mapes conceptuals.
- Ontologies.
- Xarxes semàntiques.

Dins d'aquest ampli ventall d'eines de representació i recuperació d'informació, els tesaurus i les ontologies són amb major freqüència objecte de vinculació en la bibliografia encara que provinguin d'àmbits disciplinars diferents.

6.1. Els tesaurus

Com ja hem vist anteriorment, un tesaurus és un tipus de llenguatge documental que representa l'estructuració conceptual d'un determinat camp del coneixement i es converteix en una eina de control terminològic, ja que l'estructuració del tesaurus es fa partint de conceptes.

Els conceptes són representats per termes seleccionats que evidencien aquest control terminològic ja que amb l'esmentat control es busca millorar la precisió de la indexació i recuperació de la informació, dues funcions bàsiques dels tesaurus.

L'ús d'informació conceptual i lingüística emmagatzemada en altres tipus de recursos (per exemple, en una ontologia) permet l'enriquiment a nivell estructural dels tesaurus, i possibilita l'increment de l'amigabilitat d'aquestes eines de cara a la seva utilització per part d'un usuari final no especialitzat.

6.2. Les ontologies

Les ontologies han estat un camp d'interès de la intel·ligència artificial. Una ontologia s'engloba dins de la definició del que és un llenguatge controlat, ja que proporciona una representació formal del coneixement en què els conceptes, les relacions i les restriccions conceptuals són explicitades mitjançant formalismes en un determinat domini.

Són moltes les definicions existents en el camp de la intel·ligència artificial. La més coneguda és la de Gruber (1993), qui defineix *ontologia* com una descripció formal de conceptes en el domini d'un discurs.

Arano (2005), a partir de les definicions existents, proposa la següent.

“Representació explícita i formal d'una conceptualització compartida que implica una perspectiva sobre una certa realitat i que es constitueix en l'estructura conceptual d'una base de coneixement. L'objectiu final és compartir el coneixement que representa.”

Arano (2005)

Les ontologies defineixen conceptes i relacions d'algun domini de manera compartida i consensuada, i aquesta conceptualització ha de ser representada d'una manera formal, llegible i utilitzable pels ordinadors.

Les ontologies consten de termes, les seves definicions i els axiomes que els relacionen amb altres termes que estan organitzats en una taxonomia. La finalitat última d'una taxonomia és presentar un sistema de classificació que agrupi tots els conceptes en diferents nivells jeràrquics.

A continuació, analitzem els components i les principals característiques de les ontologies (Lozano, 2000; Gruber, 1993).

Les ontologies tenen els **components** següents:

a) Conceptes. Són les idees bàsiques que s'intenten formalitzar. Els conceptes poden ser classes d'objectes, mètodes, plans, estratègies, processos de raonament, etc.

b) Relacions. Representen la interacció i l'enllaç entre els conceptes del domini. Solen formar la taxonomia del domini. Per exemple: subclasse-de, part-de, etc.

c) Funcions. Són un tipus concret de relació on s'identifica un element mitjançant el càlcul d'una funció que considera diversos elements de l'ontologia. Per exemple, poden aparèixer funcions com categoritzar-classe, assignar-data, etc.

d) Instàncies. S'utilitzen per a representar objectes determinats d'un concepte.

e) **Axiomes.** Són teoremes que es declaren sobre relacions que han de complir els elements de l'ontologia. Per exemple: "Si *A* i *B* són de la classe *C*, llavors *A* no és subclasse de *B*".

Les **característiques** més representatives de les ontologies són les següents:

- **Hi pot haver ontologies múltiples.** El propòsit d'una ontologia és fer explícit algun punt de vista, per la qual cosa de vegades serà precís combinar dues o més ontologies.
- **Podem identificar nivells d'abstracció de les ontologies.** Aquests nivells de generalització o abstracció ens donen una topologia d'ontologies. La idea és caracteritzar una xarxa d'ontologies usant multiplicitat i abstracció i, ja que no podem aspirar a tenir una descripció completa del món, podem pensar en una estratègia de construcció gradual de baix cap a dalt.
- **Multiplicitat de la representació.** Un concepte pot ser representat de moltes maneres, per la qual cosa poden coexistir múltiples representacions d'un mateix concepte.
- **Mapatge d'ontologies.** Establir relacions entre els elements d'una o més ontologies, per establir connexions, especialitzacions, generalitzacions, etc.

En termes pràctics, el **desenvolupament d'una ontologia** inclou:

- Definir *classes* en l'ontologia, això és, els conceptes i les seves relacions (classes, subclasses).
- Col·locar les classes en una *jerarquia de taxonomies* (subclasse-superclasse).
- Definir atributs i descriure els valors permesos per a aquests atributs.
- Emplenar els valors dels atributs amb exemples.

Alguns exemples de desenvolupaments d'ontologies

Figura 3. Exemple d'ontologia (Mulholland, 1999)

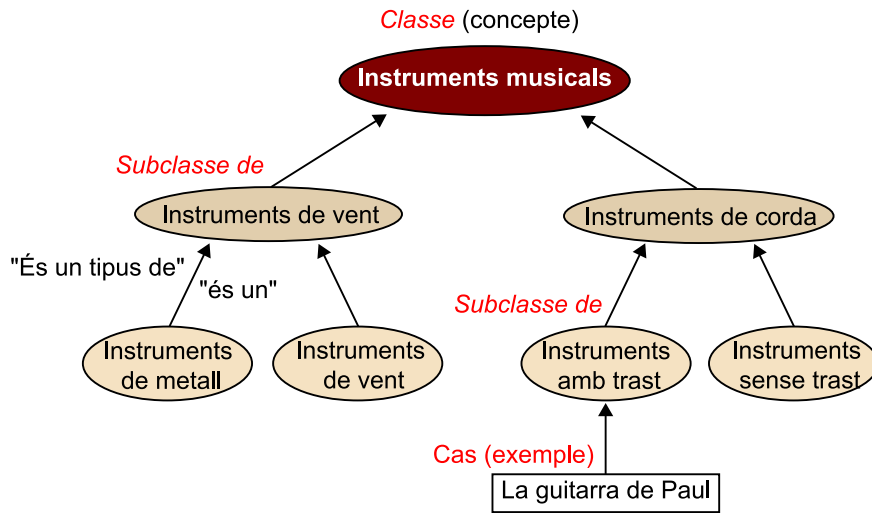
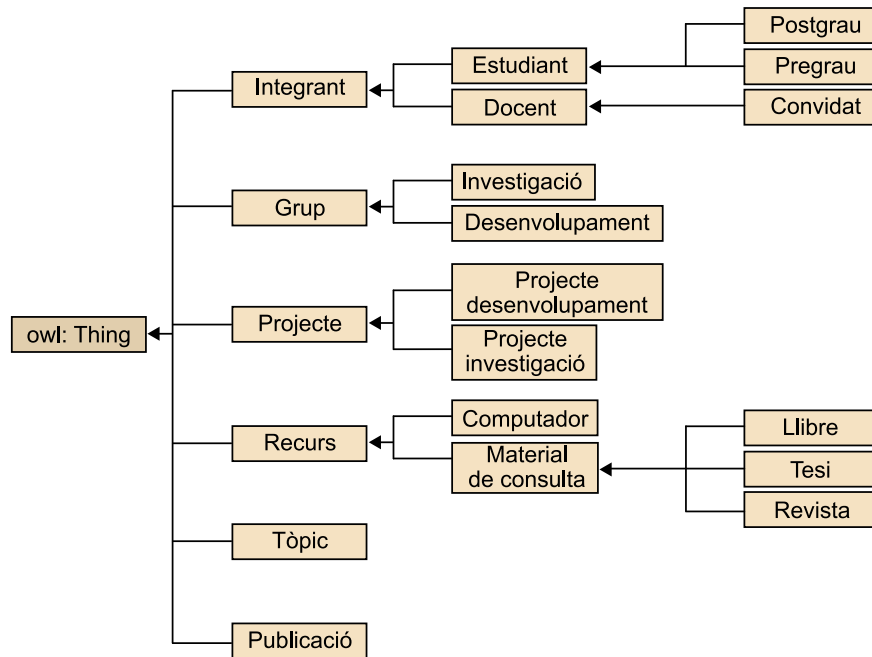


Figura 4. Exemple d'Ontologia per a Laboratoris d'Investigació i Desenvolupament



Font: Hurtado i Sequeda (2006)

6.3. Similituds i diferències entre tesaurus i ontologies

Antonio García Jiménez (2004) esmenta algunes de les diferències i semblances entre tesaus i ontologies. Aquestes es referencien a continuació:

a) Semblances

- S'engloben dins dels llenguatges controlats, ja que utilitzen el vocabulari d'un domini específic a partir d'un nombre de termes bàsics i de les relacions que s'estableixen entre els termes esmentats.

- Estan relacionats amb la terminologia emprada per a representar els conceptes d'un domini específic.
- Tots dos utilitzen jerarquies per a agrupar termes en categories i subcategories i es poden fer servir per a catalogar i organitzar recursos d'informació.

c) Diferències

- Si bé els destinataris dels tesaurus són els humans, les ontologies poden ser usades per humans per a compartir coneixement, i per agents de programari per a processar-lo.
- En els tesaurus, els descriptors s'enllacen mitjançant tres tipus principals de relacions molt simples, relacions enfocades principalment a la recuperació pertinent d'informació. En canvi, les ontologies tenen una varietat més gran de relacions entre conceptes. La descripció dels conceptes i les seves relacions es poden emprar de manera més general que en els tesaurus actuals i, per tant, es pot establir un nombre més gran de relacions diferents, i tenir-les definides.
- En els tesaurus, la descripció dels conceptes és limitada i no formalitzada, mentre que en les ontologies és explícita i formalitzada.

Bibliografia

Bibliografia bàsica

Aitchison, J.; Gilchrist, A.; Bawden, D. (2000). *Thesaurus construction and use: a practical manual* (4a. ed.). Chicago: Fitzroy Dearborn.

Arano, S. (2005). "Los tesauros y las ontologías en la Biblioteconomía y la Documentación" [article en línia]. *Hipertext.net* (núm. 3). [Data de consulta: setembre de 2009].

Chaumier, J. (1986). *Análisis y lenguajes documentales: el tratamiento lingüístico de la información documental*. Barcelona: Mitre ("Ciencias de la Comunicación").

Craven, T. C. (1997). *Introductory tutorial on thesaurus construction*. Londres / Ontario, Canadá. University of Western Ontario. (Actualitzat el 2008). [Data de consulta: setembre de 2009].

Currás, E. (1998). *Ontologías, taxonomía y tesauros: manual de construcción y uso*. Gijón: Ediciones Trea.

García Jiménez, A. (2004). "Instrumentos de representación del conocimiento: tesauros versus ontologías". *Anales de Documentación* (núm. 7, pàg. 79-95).

Gruber, T. R. (1993). "A Translation Approach to Portable Ontology Specifications". *Knowledge Acquisition* (vol. 5, núm. 2, pàg. 199-220). [Data de consulta: setembre de 2009].

Hurtado, D; Sequeda, J. (2006). "Propuesta del Uso de Ontologías para la Búsqueda de Semántica en Laboratorios de Investigación y Desarrollo: OLID"

NISO Z39.19-2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*.

NISO Z39.19-2003. *Guidelines for the Construction, Format, and Management of Monolingual Thesauri*.

Lancaster, F. W. (1995). "El control del vocabulario en la recuperación de la información". *Educación. Materiales* (núm. 12). València: Universitat de València.

Lozano Tello, A. (2000). "Ontologías en la Web semántica". I Jornadas de Ingeniería Web'01. [Data de consulta: setembre de 2009].

Mulholland, P. (1999). *Introduction to Ontologies*. Knowledge Media Institute / The Open University (Walton Hall) Milton Keynes, RU.

Slype, G. van (1991). *Los lenguajes de indización. Concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Sánchez Ruipérez ("Biblioteca del Libro").

Tomás López, M. (2004). "Diseño de un tesaruo conceptual de folklore de la Región de Murcia". *Scire* (vol. 10, núm. 1, pàg. 65-76).

UNE-50-106 (ISO 2788-1986). *Documentación: Directrices para el establecimiento y desarrollo de tesauros monolingües* (1990). Asociación Española de Normalización y Certificación (AENOR).

UNE-50-125 (ISO 5964-1985). *Documentación: Directrices para la creación y desarrollo de tesauros multilingües* (1997). Asociación Española de Normalización y Certificación (AENOR).

Willpower Information (1999). *Thesaurus principles and practice*. [Data de consulta: setembre de 2009].

Recopilació d'enllaços a recursos d'informació sobre tesauros en línia

American Society of Indexers. *Thesauri online*. Wheat Ridge, Colorado: American Society of Indexers. [Data de consulta: setembre de 2009].

Arano, S.; Codina, L. (2004). "La estructura conceptual de los tesauros en el entorno digital: ¿nuevas esperanzas para viejos problemas?" 9s. Jornades Catalanes d'informació i Documentació. Barcelona, 24-26 de novembre de 2004. [Data de consulta: setembre de 2009].

BUBL Link/5:15 Thesauri / BUBL Information Service. Glasgow: Centre for Digital Library Research, University of Strathclyde [Data de consulta: setembre de 2009].

CINDOC. *Tesaurus y Glosarios IEDCYT en línea*. Madrid: Centro de Información y Documentación Científica. [Data de consulta: setembre de 2009].

Koch, T. “Controlled vocabularies, thesauri and classification systems available in the WWW”. Lund, Suècia: Lund University. [Data de consulta: setembre de 2009].

Library of Congress Thesauri. Washington: Library of Congress. [Data de consulta: setembre de 2009].

McKiernan, G. “Controlled vocabularies: Beyond bookmarks: schemes for organizing the web”. Ames: Iowa State University. [Data de consulta: setembre de 2009].

Middleton, M. Thesaurus online sites. Controlled vocabularies: resource guide. Brisbane: School of Information Systems, QUT. [Data de consulta: setembre de 2009].

Recopilació d'enllaços a recursos sobre programari de gestió de tesaurus

Jessica Milstead. Thesaurus management software. Wheat Ridge: American Society of Indexers. [Data de consulta: setembre de 2009].

Michael Middleton. Thesaurus management software: Controlled vocabularies: resource guide. Brisbane: School of Information Systems. [Data de consulta: setembre de 2009].

Willpower Information. Software for building and editing thesauri. [Data de consulta: setembre de 2009].

Recopilació d'enllaços a recursos d'informació generals sobre ontologies

a) Pàgines web

Consorcio World Wide Web (W3C). *Guía Breve de la Web Semántica*. [Data de consulta: setembre de 2009].

“Ontologías” (Juan Carlos Cámara, 2002). *Learning Metadata Standards*. [Data de consulta: setembre de 2009].

“Ontology” (Tom Gruber). [Data de consulta: setembre de 2009].

Semanticweb. [Data de consulta: setembre de 2009].

“Sites Relevant to Ontologies and Knowledge Sharing” (Richard Fikes). [Data de consulta: setembre de 2009].

b) Articles

Arano, Silvia; Codina, Lluís (2004). “La estructura conceptual de los tesaurus en el entorno digital: ¿nuevas esperanzas para viejos problemas?”. 9s. Jornades Catalanes d'Informació i Documentació. Barcelona, 24-26 de novembre de 2004. [Data de consulta: setembre de 2009].

Arano, Silvia (2005). “Los tesaurus y las ontologías en la Biblioteconomía y la Documentación” [article en línia]. *Hipertext.net* (núm. 3). [Data de consulta: setembre de 2009].

Caldera-Serrano, J.; Sanchez-Jiménez, R. (2008, gener-desembre). “Ontología para el control y recuperación de información onomástica en televisión”. *El Profesional de la Información* (vol. 17, núm. 1, pàg. 86-91).

García Jiménez, Antonio (2004). “Instrumentos de representación del conocimiento: tesaurus versus ontologías”. *Anales de Documentación* (núm. 7, pàg. 79-95).

Hurtado Bustamante, Diana Paola; Sequeda Sanclemente, Juan Federico (2006). “Propuesta del Uso de Ontologías para la Búsqueda Semántica en Laboratorios de Investigación y Desarrollo”. OLID. [Data de consulta: setembre de 2009].

Legg, C. (2007). “Ontologies on the Semantic Web”. *Annual Review of Information Science and Technology* (vol. 41, núm. 1).

Lozano Tello, Adolfo (2000). “Ontologías en la Web semántica”. *I jornadas de Ingeniería Web'01*. [Data de consulta: setembre de 2009].

Pedraza-Jiménez, R.; Codina, L.; Rovira, C. (2007, novembre-desembre). "Web semántica y ontologías en el procesamiento de la información documental". *El Profesional de la Información* (vol. 16, núm. 6, pàg. 569-578).

Reiss, M.; Moal, M.; Barnard, Y.; Ramu, J.P.; Froger, A. (2006). "Using Ontologies to Conceptualize the Aeronautical Domain". A: *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics* (pàg. 56-63).

Sanchez-Jiménez, R.; Gil-Urdiciain, B. (2007, novembre-desembre). "Lenguajes Documentales y ontologías". *El Profesional de la Información* (vol. 16, núm. 6, pàg. 551-560).

