

# Llenguatges documentals

Maribel Cuadrado Benítez

P06/09047/02451



Universitat Oberta  
de Catalunya

[www.uoc.edu](http://www.uoc.edu)



# Índex

<b>Objectius</b> .....	5
<b>1. Què són els llenguatges documentals?</b> .....	7
<b>2. El llenguatge natural en la comunicació documental</b> .....	11
2.1. La sinonímia .....	11
2.2. La polisèmia .....	13
<b>3. Per què llenguatges artificials?</b> .....	16
3.1. Objectiu del control del vocabulari .....	17
<b>4. Una polèmica: llenguatges documentals o d'indexació?</b> .....	19
<b>5. Les tipologies dels llenguatges documentals</b> .....	23
5.1. Tipologies .....	23
<b>6. Llenguatge codificat o natural?</b> .....	26
6.1. Llenguatges codificats .....	26
6.2. Llenguatges naturals .....	29
<b>7. Llenguatges documentals i nivell d'indexació</b> .....	32
7.1. Sistemes de classificació i encapçalaments de matèries .....	32
7.2. Llenguatges o vocabularis de descriptors .....	33
7.3. Llenguatges de paraules clau .....	34
<b>8. Llenguatges lliures o controlats?</b> .....	36
8.1. Llenguatges lliures .....	36
8.2. Llenguatges controlats .....	37
<b>9. Precoordinats o postcoordinats?</b> .....	39
9.1. Llenguatges precoordinats .....	39
9.2. Llenguatges postcoordinats .....	40
<b>10. Estructures dels llenguatges documentals</b> .....	44
10.1. Llenguatges d'estructura jeràrquica o arborescent .....	44
10.2. Llenguatges d'estructura combinatòria o associativa .....	45
10.3. Llenguatges d'estructura sintàctica .....	47
<b>11. Quin llenguatge documental?</b> .....	50
11.1. Passat i futur dels llenguatges documentals .....	53

<b>12. Llistats de paraules clau</b> .....	57
12.1. L'ús dels llistats de paraules clau en la indexació i la recuperació .....	58
<b>13. Llistats de descriptors lliures</b> .....	62
13.1. L'ús dels llistats de descriptors lliures en la indexació i la recuperació .....	63
<b>14. Llistats d'autoritats o descriptors controlats</b> .....	65
14.1. L'ús dels llistats d'autoritats en la indexació i la recuperació ....	66
14.2. Els llistats d'autoritats i els identificadors .....	67
<b>Autoavaluació</b> .....	70
<b>Solucionari</b> .....	71
<b>Bibliografia</b> .....	72

## **Objectius**

- 1.** Analitzar els processos i factors necessaris perquè hi hagi una bona comunicació documental.
- 2.** Conèixer la funció dels llenguatges documentals dins de la comunicació documental.
- 3.** Analitzar i comparar les diferents tipologies de llenguatges documentals per a entendre que no hi ha cap llenguatge documental pur.
- 4.** Saber-ne sistematitzar les característiques en la relació a la seva funció.
- 5.** Conèixer les característiques dels principals llenguatges documentals.
- 6.** Conèixer la naturalesa i les tipologies dels diferents llenguatges documentals.



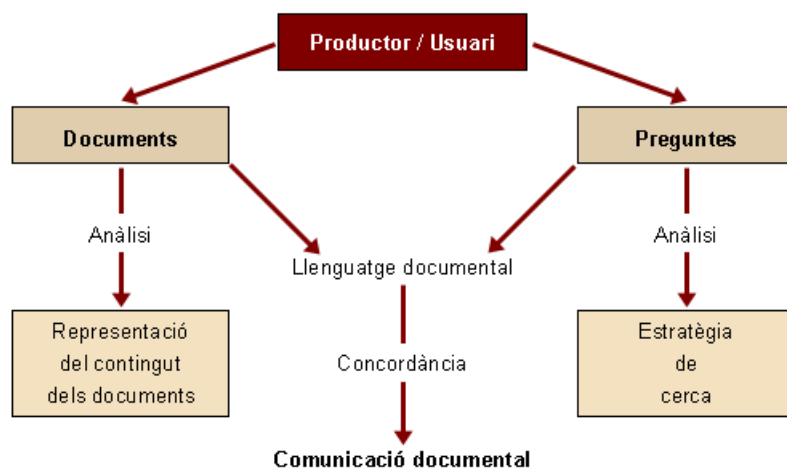
## 1. Què són els llenguatges documentals?

El principal objectiu d'un sistema d'informació és la recuperació de manera ràpida i precisa de la informació continguda en els documents.

L'usuari que s'adreça a un sistema documental amb una necessitat d'informació precisa (quines obres hi ha sobre Eduardo Mendoza?, quins títols s'han publicat a la col·lecció Episodis de Catalunya?), interroga el sistema coneixent l'element o punt d'accés que el portarà a trobar la resposta (Eduardo Mendoza, Episodis de Catalunya).

Però quan l'usuari vol informació sobre un tema o matèria determinat (quina informació hi ha sobre malalties infantils?) interroga el sistema desconeixent l'element o punts d'accés que ha de fer servir per a recuperar de manera exhaustiva la informació (malalties, infants, xarampió, varicel·la?).

Per a poder recuperar la informació continguda en els documents necessitem un sistema que ens permeti d'establir una comunicació documental entre els emissors de la informació (autors dels documents) i els destinataris (usuaris) que necessiten la informació. Com en tot sistema de comunicació, cal que ambdós protagonistes facin servir un mateix codi per a comunicar-se, que parlïn un mateix llenguatge.



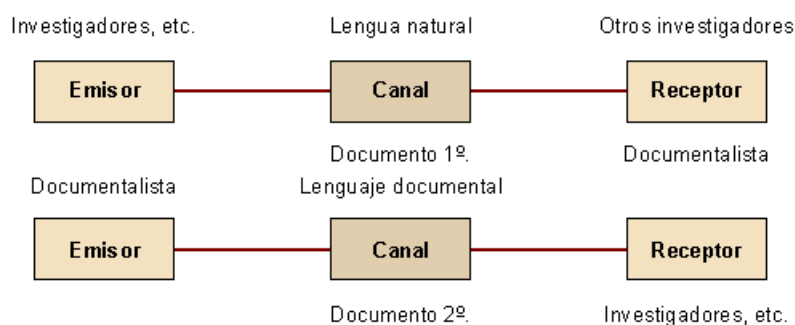
Un usuari busca en una base de dades bibliogràfica informació sobre el període prehistòric calcolític. Formula la consulta amb aquest terme i no troba cap referència documental perquè els autors del document han fet servir en el títol i text del document el terme sinònim *eneolític* per a designar el mateix període.

En aquest cas, la comunicació documental fracassa, ja que un mateix concepte no està codificat de la mateixa manera per l'autor i l'usuari.

A aquests protagonistes de la comunicació documental es pot afegir un tercer, que ve a generar un segon nivell de comunicació: l'analista d'un sistema documental que tracta i analitza els documents.

L'usuari busca en la base de dades informació sobre instal·lacions esportives i no troba cap referència documental perquè l'analista del sistema va indexar els documents que l'interessen amb el terme equipaments esportius. També en aquest cas no es dona la comunicació documental entre el document i l'usuari.

El següent gràfic presenta l'esquema d'aquesta comunicació:



Font: García Gutiérrez, A. (1989) "Los lenguajes documentales". A: López Yepes, J. (comp). *Fundamentos de información y documentación* (pág.318). Madrid: Eudema (Manuales).

Es presenten dos nivells de comunicació. En el pla superior, l'emissor (autor del document) fa ús d'una llengua natural d'expressió (català, castellà, anglès, etc.) per a l'elaboració del document original. El document arriba als receptors, entre els quals hi ha els analistes dels sistemes d'informació que es converteixen d'aquesta manera en emissors de segon nivell, ja que tracten el document original (anàlisi descriptiva i del contingut) i el transformen en un document secundari; una referència documental.

Per tal que l'usuari pugui recuperar la informació continguda en els documents ha de fer ús del mateix llenguatge, després de la traducció de la llengua natural, que han fet servir els analistes.

Els sistemes que permeten d'establir una comunicació entre la informació continguda en els documents i l'usuari d'un sistema documental són els llenguatges documentals.

Els llenguatges documentals són vocabularis de termes o signes emprats en les operacions d'indexació i recuperació dels documents d'un determinat sistema documental. La seva funció bàsica és la representació del contingut del document per a recuperar-los segons la seva temàtica.

Vegem una definició:



"Un lenguaje documental será entonces todo sistema de signos que permita representar el contenido de los documentos con el fin de recuperar los documentos pertinentes en respuesta a consultas que tratan sobre ese contenido. El lenguaje documental no se refiere, pues, a otros criterios utilizados en la búsqueda documental: autor del documento, lengua del texto, fecha de publicación..."

Van Slype, G., pàg. 21

La representació del contingut dels documents mitjançant un llenguatge documental es pot establir a diferents nivells:

- Matèria o tema del document.
- Conceptes tractats en el document.
- Paraules significatives (no buides) contingudes en el títol, resum i/o text complet del document.

Per a portar a terme les dues primeres operacions són necessàries diferents operacions intel·lectuals, realitzades pels analistes del sistema: examen del document, identificació del tema i dels conceptes tractats i traducció d'aquests al llenguatge artificial o codi emprat pel sistema. Per contra, l'extracció automàtica de les paraules significatives del document no requereix la intervenció de cap operació intel·lectual per part de l'analista, fa servir la llengua natural (paraules) que es troba en el document.

Així, doncs, podem afirmar que la comunicació documental es pot establir per mitjà de:

- Llenguatges artificials: representen els conceptes i la matèria del document, en la indexació i la recuperació, mitjançant un llenguatge normalitzat i controlat.
- Llenguatge natural:<sup>1</sup> paraules i/o expressions lliures que pertanyen al llenguatge comú i que es troben en els documents (títol, resum, text...) i en els conceptes de la consulta de l'usuari.

<sup>(1)</sup>La decisió d'incloure el llenguatge natural dins la categoria de llenguatges documentals s'aparta de molts autors per als quals un llenguatge documental és necessàriament un llenguatge artificial; això és cert en els casos dels sistemes de classificació i els tesaurus, però -tal com afirma G. Van Slype- no ho és per a altres llenguatges com els llistats de paraules clau extretes del títol, resum o text dels documents, fruit de l'aplicació d'una indexació automàtica. Les paraules clau es fan servir en la cerca documental d'acord amb els mateixos principis (postcoordinació, cerca booleana) que els descriptors d'un tesaurus. Es presenten com el primer graó, l'estadi menys controlat, en una evolució cap a llenguatges i sistemes més controlats, com els tesaurus. Els dos tipus de llenguatge tenen una mateixa funció: representar el contingut del document.

Tots dos sistemes de comunicació documental presenten avantatges i inconvenients en el moment de la indexació o la recuperació dels documents.

## Els llenguatges documentals o d'indexació

...són sistemes de signes o vocabularis de termes que permeten de representar el contingut dels documents per a recuperar-los depenent de la seva temàtica.

...permeten d'establir una comunicació entre la informació continguda en els documents i l'usuari que formula una necessitat d'informació en un sistema documental.

La comunicació documental es pot establir mitjançant llenguatges artificials, normalitzats i controlats, o bé mitjançant el llenguatge natural contingut en el text del document. Tots dos sistemes presenten avantatges i inconvenients en la indexació o en la recuperació dels documents d'un sistema documental.

### Vegeu també

"La indexació o descripció característica" del mòdul "Anàlisi i recuperació del contingut".

"El llenguatge natural i la comunicació documental", "Per què llenguatges artificials?" i "Una polèmica: llenguatges 'documentals' o 'd'indexació'?", d'aquest mateix mòdul.

### Bibliografia

**García Gutiérrez, A.** (1989). "Los lenguajes documentales". En: J. López Yepes (comp.). *Fundamentos de información y documentación* (págs. 313-342). Madrid: Eudema (Manuales).

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

**Maniez, J.** (1993). *Los lenguajes documentales y de clasificación. Concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Sánchez Ruipérez (Biblioteca del Libro, núm. 56).

**Perpinyà Morera, R.** (1995). "Los lenguajes documentales". En: E. Fuentes (ed.). *Manual de documentación periodística* (págs. 112-132). Madrid: Síntesis (Biblioteconomía y Documentación, núm. 6).

**Pinto Molina, M.** (1989). "El análisis de contenido: la indización de documentos". En: J. López Yepes (comp.). *Fundamentos de información y documentación* (págs. 347-374). Madrid: Eudema (Manuales).

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

## 2. El llenguatge natural en la comunicació documental

Per "llenguatge natural" entenem el llenguatge del discurs comú (català, castellà, anglès...). És, per tant, l'emprat pels autors dels documents i el que fa servir l'usuari en el moment de plantejar una necessitat d'informació. És un dels sistemes emprats pels sistemes documentals per a establir la comunicació documental entre la informació continguda en el document i l'usuari que fa una consulta.

Per tal que aquesta comunicació documental entre document i usuari sigui un èxit, és necessari que el codi aplicat al document en la indexació coincideixi amb l'utilitzat en la consulta: un únic terme per a denominar un mateix objecte o concepte. És a dir, ha d'existir univocitat.

Però com podem aconseguir representar el mateix concepte per mitjà d'un únic terme en els dos processos (indexació/consulta) si s'utilitza un codi tan ambigu com és el llenguatge natural?

### Exemple

Qui no ha tingut l'experiència de buscar en una guia telefònica informació, per exemple, sobre restaurants per a celebrar un banquet de casament, i no ha hagut de consultar la classificació d'activitats en diferents ocasions perquè el terme inicial expressat en la cerca (restaurants) no es corresponia amb l'emprat en la guia (salons per a casaments i convencions, servei d'àpats, *catering*, etc.).

L'ús del llenguatge natural en el procés d'indexació i recuperació dels documents planteja tota una sèrie de problemes derivats de l'ambigüitat d'aquest tipus de llenguatge, és a dir, els generats per la sinonímia, la polisèmia i l'homonímia.

Analitzarem aquests diferents fenòmens i veurem en quina mesura obstaculitzen la comunicació documental necessària entre els documents i l'usuari final.

### 2.1. La sinonímia

Es diu que dues paraules són sinònimes quan tenen un mateix significat i diferents significants. La verificació pràctica de la sinonímia és, doncs, fàcil d'enunciar: dos sinònims són substituïbles l'un per l'altre en qualsevol context.

#### Contingut complementari

J. Maniez fa servir el terme *biunivocitat* per a indicar els dos sentits de la univocitat: un únic terme per concepte i un únic concepte per terme.

## Què és aquesta persona?



Reflexió

un clergue?

un sacerdot?

un religiós?

un mossèn?

El llenguatge natural es caracteritza per la sinonímia: l'existència de més d'un terme o paraula per a designar un mateix objecte o concepte.

El document "Els sacerdots a Barcelona durant el segle XIX" arriba a un sistema documental. El document s'indexa automàticament, extraient les paraules del títol i del resum. Si un usuari vol fer una consulta per a trobar informació sobre els clergues durant la Guerra del Francès i formula la consulta emprant el terme *clergues* no recuperarà aquest document, que és totalment pertinent. S'haurà produït una manca d'informació, un silenci documental. A fi d'evitar-ho, ha de preveure i indicar tots els possibles sinònims de la paraula clergues en la seva estratègia de cerca. També desconeix si el terme estarà en singular o plural i, per tant, cal que faci un truncament:

(Clergue\$ OR sacerdot\$ OR religios\$ OR mossen\$ OR prevere\$) AND guerra AND (francès OR independència)

## 2.2. La polisèmia

La figura semàntica contrària a la sinonímia que també trobem en el llenguatge natural és la polisèmia.

La polisèmia es dona quan una mateixa paraula o signe lingüístic pot expressar més d'un concepte: un únic significat i diversos significats, però amb trets semàntics comuns.

Per exemple:

### **Exemple**

llis (sense aspreses)

llis (color homogeni, sense estampat)

Exemple extret de G. Van Slype, pàg. 28.

La polisèmia també es pot manifestar per mitjà de l'homonímia.

### **L'homonímia**

És la similitud formal de paraules diferents.

És una varietat de la polisèmia: un únic significat i diversos significats, sense trets semàntics comuns.

Un homònim és una cadena de caràcters que té més d'un significat però una única pronunciació (per exemple, *planta*).

Si la similitud apareix en la forma oral s'anomena **homofonia**; si apareix en la forma escrita es parla d'**homografia**:

- L'homòfon fa referència a cadenes de caràcters diferents però que es pronuncien igual. Així, les paraules *vell* i *bell* són homòfones però no homògrafes.
- Un homògraf consisteix, doncs, en una cadena de caràcters amb més d'un significat (per exemple, *escuma*); també pot tenir més d'una pronunciació.

L'homografia es distingeix de la polisèmia en el fet que els termes homògrafs són diferents pel seu origen i la coincidència formal que els confon és merament fortuïta.

Per exemple:

### **Exemple**

L'**escuma** va caure sobre un matalàs d'**escuma**

Exemple extret de J. Maniez, pàg. 199.

Un usuari s'adreça a un sistema documental i vol recuperar documents sobre bancs, en el sentit d'entitat financera. Si formula la consulta amb el terme *bancs*, recuperarà documents que facin referència a bancs (entitats financeres), bancs (mobiliari urbà), bancs (conjunt de peixos), bancs (de dades), bancs (de sang), etc.

Ha recuperat tota una sèrie de documents amb temàtiques no pertinents a la seva consulta. Si la sinonímia provoca manca d'informació, la polisèmia i l'homonímia comporten soroll documental.

Hem pogut constatar que l'ambigüitat del llenguatge natural és un obstacle per a la recuperació documental. Si un terme pot designar indiferentment diversos objectes o conceptes i aquests objectes o conceptes poden estar representats per diversos termes, l'usuari:

- trobarà documents diferents als que cercava (soroll documental).
- no trobarà els documents que cercava (silenci documental).

Per a eliminar el soroll i el silenci documentals és necessari que els termes emprats en la indexació i recuperació dels documents siguin tan unívocs com sigui possible.

L'ús del llenguatge natural en la indexació i en la recuperació dels documents

...planteja una sèrie de problemes derivats de la seva ambigüitat: sinonímia, polisèmia i homonímia.

...provoca soroll i silenci documentals en el moment de la recuperació.

...no permet la univocitat (un únic terme, un únic concepte) necessària per tal que hi hagi una bona comunicació documental.

### **Vegeu també**

Mòdul 2, nucli de coneixement "Els llenguatges d'interrogació"

Mòdul 3, nuclis de coneixement "Llenguatges lliures o controlats?" i "Llistas de paraules claus"

Mòdul 6, nuclis "Els no descriptors i termes preferents" i "Les relacions d'equivalència o sinonímia"

**Bibliografia**

**García Gutiérrez, A.** (1989). "Los lenguajes documentales". En: J. López Yepes (comp.). *Fundamentos de información y documentación* (págs. 313-342). Madrid: Eudema (Manuales).

**Maniez, J.** (1993). *Los lenguajes documentales y de clasificación. Concepción, construcción y utilización en los sistemas documentales* (págs. 192-202). Madrid: Pirámide / Fundación Sánchez Ruipérez (Biblioteca del Libro, núm. 56).

### 3. Per què llenguatges artificials?

Per tal que la comunicació documental entre document i usuari sigui un èxit és necessari que hi hagi univocitat: un únic terme per concepte i un únic concepte per terme.

Es fa necessari comptar amb un llenguatge que controli l'ús i els significats dels termes emprats en la indexació i en la recuperació dels documents. Aquests llenguatges són les eines que faran de pont entre el document, l'analista i l'usuari final.

La primera justificació de l'existència dels llenguatges documentals artificials o controlats és la necessitat d'univocitat (un únic terme, un únic concepte): com més propers siguin al llenguatge natural, més grans seran els riscos de trobar silenci i soroll documentals en la recuperació de la informació.

L'ús d'un únic codi per a designar un únic concepte -conegut per l'analista (que tradueix el concepte del document, en llenguatge natural, a un llenguatge controlat) i per l'usuari (que a la vegada tradueix el concepte de la seva consulta, en llenguatge natural, al mateix llenguatge controlat)- permet la comunicació documental entre tots dos.

#### Exemple

El document "Els sacerdots a Barcelona durant la Guerra del Francès" arriba a un sistema documental. L'analista que indexa el contingut del document ha de traduir, entre d'altres, els conceptes "sacerdots" i "Guerra del Francès" al llenguatge documental emprat en el sistema documental. Suposem que es tracta d'un vocabulari controlat de descriptors tipus tesaurus. L'analista consulta el terme *sacerdots* i veu que el llenguatge documental no li permet de fer ús del terme *sacerdots*, sinó que ha d'emprar el seu sinònim *clergues*:

Sacerdots

Empreu *clergues*

De la mateixa manera, consulta el terme *Guerra del Francès* i comprova que està reconegut com a descriptor preferent sobre altres possibles sinònims com *Guerra de la Independència*:

Guerra del Francès

Emprat per *Guerra de la Independència*

L'usuari que s'adreça a aquest sistema documental per a recuperar documents sobre la Guerra de la Independència a Catalunya en primer lloc ha de consultar el llenguatge documental emprat pel sistema en el qual s'indica quin és el descriptor que ha d'utilitzar per a poder recuperar els documents sobre aquest tema:

Guerra de la Independència

Empreu *Guerra del Francès*



Analista i usuari han d'emprar els mateixos termes per a indexar i recuperar els documents. S'ha produït una comunicació documental que ha evitat els fenòmens de soroll i silenci documentals associats a l'ambigüïtat del llenguatge natural.

### 3.1. Objectiu del control del vocabulari

En resum, l'objectiu del control del vocabulari és facilitar la representació de les matèries per part dels analistes i dels usuaris que les recuperen. Això s'aconsegueix:

- controlant la sinonímia i la polisèmia dels termes d'indexació.
- fent possible la univocitat del llenguatge (un únic terme, un únic concepte).
- facilitant la realització de cerques àmplies sobre una matèria enllaçant els termes amb relacions d'associació semàntiques.

La indización tiende a ser más consistente cuando el vocabulario utilizado está controlado. Será más probable que los indizadores estén de acuerdo sobre los términos necesarios para representar una determinada materia, si éstos son seleccionados de una lista previa que si pueden utilizar cualquier término. Lo mismo ocurre en el proceso de búsqueda: será más fácil identificar los términos apropiados a una necesidad de información si son seleccionados de una lista determinada. Por lo tanto, el vocabulario controlado facilita la coincidencia entre el lenguaje de los indizadores y el de los usuarios que hacen la búsqueda.

F. Lancaster, pàg. 22

Un vocabulari controlat o llenguatge documental artificial pot ser simplement un conjunt limitat de termes que puguin ser emprats per l'analista i l'usuari. No obstant això,

normalment els termes d'indexació estan sotmesos a algun tipus d'estructura per tal que els que tenen significats propers apareguin, d'alguna manera, junts o bé relacionats.

Aquest fet ajuda tant l'analista com l'usuari que fa la consulta en el moment de seleccionar els termes més adequats per a representar una determinada matèria. I també ajuda l'usuari a identificar tots els termes necessaris per a realitzar una cerca àmplia sobre una matèria.

Les diferents maneres d'agrupar o estructurar els termes, així com el sistema d'indexació seleccionat, el nivell de control exercit sobre el vocabulari i els signes lingüístics emprats pel vocabulari dóna lloc a una tipologia diversa de llenguatges documentals.

## Els llenguatges documentals artificials

...són vocabularis controlats dels termes emprats en les operacions d'indexació i recuperació dels documents d'un determinat sistema documental.

...tenen com a principal objectiu facilitar la representació del contingut dels documents a analistes i usuaris: controlant la sinonímia i la polisèmia dels termes d'indexació i fent possible la univocitat (un únic terme, un únic concepte)

...presenten generalment una estructura que permet d'agrupar o relacionar els termes amb afinitats de significat, amb l'objectiu d'ajudar a analistes i usuaris en la selecció del termes d'indexació més idonis.

### Vegeu també

"La indexació o descripció característica", "La recuperació de la informació" del mòdul "Anàlisi i recuperació del contingut".

"El llenguatge natural en la comunicació documental", "Les tipologies dels llenguatges documentals" d'aquest mateix mòdul.

### Bibliografia

**García Gutiérrez, A.** (1989). "Los lenguajes documentales". En: J. López Yepes (comp.). *Fundamentos de información y documentación* (págs. 313-342). Madrid: Eudema (Manuales).

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

**Perpinyà Morera, R.** (1995). "Los lenguajes documentales". En: E. Fuentes (ed.). *Manual de documentación periodística* (págs. 112-132). Madrid: Síntesis (Biblioteconomía y Documentación, núm. 6).

**Pinto Molina, M.** (1989). "El análisis de contenido: la indización de documentos". En: J. López Yepes (comp.). *Fundamentos de información y documentación* (págs. 347-374). Madrid: Eudema (Manuales).

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

## 4. Una polèmica: llenguatges documentals o d'indexació?

Entre els estudiosos de la disciplina hi ha una polèmica pel que fa a la terminologia que s'ha d'emprar per a anomenar els sistemes de signes que ens permeten de representar el contingut dels documents amb la finalitat de recuperar-los per la seva temàtica: llenguatges documentals o llenguatges d'indexació?

Generalment es fan servir indistintament per a designar el mateix concepte: el conjunt de termes d'indexació emprats en un sistema documental per a la recuperació, independentment que la llengua d'aquests sigui natural (paraula clau), controlada (descriptor) o codificada (símbol de classe). Segons Maria Pinto, l'ús d'una terminologia o altra depèn dels àmbits professionals i escoles dels autors:

"...las expresiones más empleadas por profesionales de la documentación e información científica son 'lenguajes de indización' y 'lenguajes documentales'; la primera, localizada en el ámbito anglosajón, donde el término documentación había cedido su lugar al de *information science*, lo que hacía suponer que no se escogiera el calificativo documental. Sin embargo, no ocurre lo mismo en los países mediterráneos, que han prodigado el uso del vocablo *documentación* y sus derivados. Además se da el caso de que un mismo teórico emplea más de una expresión, confundiendo al usuario."

M. Pinto Molina, 1989, pàg. 382

No obstant això, alguns autors (G. Van Slype, J. Maniez) fan ús dels dos termes deliberadament amb l'objectiu de fer una distinció entre els anomenats *llenguatges de classificació* (sistemes de classificació) i els *llenguatges d'indexació* (tesaurus i lèxics controlats) que es presenten com les dues tipologies possibles dins dels llenguatges documentals.

El fons de la qüestió resideix en la polèmica existent sobre la consideració de l'operació de la *classificació* com a diferent i, per tant, independent de la d'*indexació*. Per a agrupar aquestes dues operacions que actuen sobre el contingut del document, fan servir el terme *caracterització*. Això els permet diferenciar-les de la tècnica de condensació de la informació del contingut del document, que es porta a terme amb el resum.

Aquest plantejament els fa emprar vocables diferents per a referir-se als llenguatges aplicats en el procés de classificació i d'indexació:

"Existen dos tipos principales de lenguajes documentales: los lenguajes de indización, denominados también lenguajes combinatorios, que permiten representar el contenido de los documentos y de las consultas de forma analítica; los lenguajes de clasificación utilizados más generalmente para representar este contenido de forma sintética."

G. Van Slype, pàg. 21

La Norma UNE 50-121, que determina els mètodes per a l'anàlisi dels documents i selecció dels termes d'indexació, no recull aquesta distinció i defineix la *indexació* com:

"Acción de describir o identificar un documento en relación con su contenido." (UNE 50-121: 254)

i els *termes d'indexació* com:

"la representación de un concepto en forma de: un término derivado del lenguaje natural, preferiblemente sustantivo simple o compuesto; un código de clasificación." (UNE 50-121: 254)

Davant d'aquesta polèmica, cal aclarir que:

1) S'ha optat per seguir la normativa i, per tant,

entendrem com a indexació un únic procés que es pot portar a terme a diferents nivells segons la naturalesa i necessitats del sistema documental i que implica, per tant, l'ús de llenguatges documentals amb característiques i tipologies diferents:

- **Indexació per matèries o classificació**

És un sistema d'indexació que té per objectiu sintetitzar el contingut dels documents generalment en un únic terme d'indexació. Aquest terme es pot traduir a un sistema de classificació o a un llistat d'encapçalament de matèries. Si es tradueix a un sistema de classificació (per exemple, CDU) podem emprar el terme *classificació* per a designar aquest sistema d'indexació.

- **Indexació per conceptes o descriptors**

És un sistema d'indexació que té per objectiu identificar els conceptes que conformen el contingut del document i representar-los mitjançant una sèrie de descriptors, multiplicant així els punts d'accés o entrades del document en el catàleg.

- **Indexació per paraules clau**

És un sistema d'indexació que té per objectiu analitzar el contingut dels documents mitjançant les paraules significatives contingudes en el seu títol, resum o text complet. Generalment és un procés realitzat automàticament.

2) Es fa ús del terme *llenguatges documentals* en lloc del de *llenguatges d'indexació* perquè és l'aconsellat per la normativa que el defineix com:

"Lenguaje formalizado utilizado para caracterizar los datos o el contenido de los documentos y permitir su almacenamiento o recuperación." (UNE 50-113, pàg. 421)

### Vegeu també

"Llenguatges documentals i nivell d'indexació", "Precoordinats o postcoordinats?" d'aquest mòdul.

"La indexació o descripció característica" del mòdul "Anàlisi i recuperació del contingut".

"Els sistemes de classificació" del mòdul "Sistemes de classificació".

No obstant aquesta polèmica terminològica i de procediment, tots els autors coincideixen en el fons en el fet que hi ha diferents maneres d'identificar, extreure i traduir el contingut dels documents. Les etiquetes varien segons els autors, però les operacions són les mateixes.

**Maniez, J.** (1993). *Los lenguajes documentales y de clasificación. Concepción, construcción y utilización en los sistemas documentales* (pàg. 219-220). Madrid: Pirámide / Fundación Sánchez Ruipérez (Biblioteca del Libro, 56). La Fundación Sánchez Ruipérez ha autoritzat la reproducció gratuïta d'aquest capítol.

J. Maniez és un dels autors que sosté la tesi que l'anàlisi del contingut del document comprèn dues operacions: caracterització (classificació i indexació) i condensació (resum). Els llenguatges documentals, segons aquest autor, es divideixen en llenguatges de classificació (sintètics) i llenguatges d'indexació (analítics) depenent de si es fan servir en l'operació de classificació o en la d'indexació. Per a aquest autor, *caracterització* és el terme que agrupa les operacions que la Norma UNE 50-121 denomina *indexació*.

Us proposem de fer una lectura de les conclusions de J. Maniez que ens serviran per a introduir-nos en l'anàlisi dels llenguatges documentals i de les seves característiques.

### Recapitulación

Recapitemos los caracteres funcionales de los lenguajes documentales y, según cada carácter, ordenemos los diferentes lenguajes según su grado de adecuación funcional:

1) Todos los lenguajes documentales son instrumentos de ordenación destinados a encontrar rápidamente los documentos solicitados en caso de necesidad. Son, pues, *repertorios de signos distintivos*, susceptibles de ser asignados a un documento, ordenados para estar disponibles en cualquier momento.

La asignación y la ordenación pueden hacerse directa y concretamente mediante una etiqueta pegada al documento (caso de los índices de clasificación), pero lo más frecuente es que se hagan indirectamente, por ejemplo, con una inscripción en la ficha, una lista o un soporte electrónico, a partir de los cuales puede encontrarse la referencia de la obra.

La propia naturaleza de estos signos es en sí misma secundaria: alfanumérica, numérica, gráfica; y secundaria también su *naturalidad*. Lo esencial es que respondan a las exigencias de la *biunivocidad*: un solo signo distintivo por concepto y un solo concepto para un signo distintivo.

Esta exigencia lleva consigo la necesidad de fijar reglas precisas para el empleo de estos repertorios, de *controlarlos* en su aspecto formal, en la relación forma/contenido y, eventualmente, en su sintaxis.

Por orden de control decreciente, figuran las clasificaciones jerárquicas, las clasificaciones facetadas, los tesauros con reglas sintácticas, los tesauros y listas de autoridades y las listas de términos en lenguaje natural. Los unitérminos, extraídos por indexación automática, no responden a las exigencias de la biunivocidad.

2) Todos los lenguajes documentales se construyen para seleccionar los documentos *en función de su contenido*. No son repertorios de objetos, sino de *temas*, una especie de casilleros de análisis semántico que cubren un dominio del saber o el conjunto del saber.

Desde este punto de vista, los lenguajes documentales son forzosamente tributarios del lenguaje natural, ya que los conceptos se forjan y se expresan por medio del lenguaje.

Hay dos grandes categorías de casilleros semánticos: unos, con casillas amplias, previstas para que cada documento pueda encontrar en ellas un lugar y sólo uno. Son los lenguajes de tipo sintético (se les llama también *precoordinados*): clasificaciones enumerativas, listas

de encabezamientos de materias. En las otras, con casillas mucho más pequeñas, cada casilla representa un concepto: en general varias casillas afectan al mismo documento y es la combinación de estas casillas la que indica el contenido. Son los lenguajes de tipo analítico (se les llama también *postcoordinados*): tesauros, listas de descriptores. Entre los dos se sitúan las clasificaciones por facetas y los lenguajes de indización provistos de una sintaxis obligatoria.

Esta particularidad de los lenguajes documentales es lo que da lugar a que ciertos autores digan que son metalenguajes, lenguajes que describen otro lenguaje. De hecho, no pretenden describir el lenguaje de los documentos sino expresar los conceptos o las materias, sea cual sea el lenguaje de los propios documentos (natural, visual o sonoro).

3) Una tercera característica de los lenguajes documentales es importante, aunque no esencial: la estructuración semántica. Esta estructuración es inherente al sistema de creación de las clasificaciones y de los tesauros, existiendo también una tendencia a alinear estas dos categorías diferentes entre las clasificaciones en el sentido amplio del término (conjunto de clases unidas explícitamente por relaciones semánticas).

En la medida en que todo análisis documental es un casillero de análisis semántico a disposición de los usuarios, es natural que proporcione vías de acceso al contenido implícito de los documentos, ya sea por su organización sistemática o por medio de un instrumento auxiliar como un índice.

4) Otro aspecto facultativo es la presencia de una sintaxis. Hemos encontrado casos de sintaxis documental en los lenguajes de indización y en las clasificaciones. Pero un lenguaje documental puede muy bien limitarse a un simple repertorio de términos: es el caso de las clasificaciones enumerativas y de los numerosos lenguajes de indización.

## Bibliografía

Asociación Española de Normalización y Certificación (AENOR) (1997). "UNE-50-113/6 (ISO 5127/6). Documentación e información. Vocabulario. Parte 6: lenguajes documentales". *Revista Española de Documentación Científica* (vol. 20, núm. 4, págs. 417-436).

Asociación Española de Normalización y Certificación (AENOR) (1991). *UNE-50-121/91 (ISO 5963-1985). Documentación: métodos para el análisis de documentos, determinación de su contenido y selección de los términos de indización*.

Maniez, J. (1993). *Los lenguajes documentales y de clasificación. Concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Sánchez Ruipérez (Biblioteca del Libro, núm. 56).

Pinto Molina, M. (1989). "El análisis de contenido: la indización de documentos". En: J. López Yepes (comp.). *Fundamentos de información y documentación* (págs. 347-374). Madrid: Eudema (Manuales).

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

## 5. Les tipologies dels llenguatges documentals

Els llenguatges documentals han estat construïts per a una finalitat precisa, generalment per a satisfer les necessitats de la indexació i recuperació d'un determinat sistema documental: facilitar l'accés a la informació d'acord amb les demandes dels usuaris. Per tant, l'objectiu per al qual s'elaboren, és a dir, les característiques de cada sistema documental, condicionen la naturalesa i estructura del llenguatge documental que s'ha d'aplicar.

Els principals llenguatges documentals són els següents:

- Els sistemes de classificació
  - Classificació Decimal Dewey (DDC)
  - Classificació decimal universal (CDU)
  - Classificació de la Library of Congr s (LCC)
  - Classificació Colon (CC)
- Els llistats d'encapçalament de mat ries
- Els tesaurus
- Els llistats de paraules clau
- Els llistats de descriptors lliures
- Els llistats d'autoritats o descriptors controlats

### 5.1. Tipologies

L'estudi de la naturalesa dels llenguatges documentals planteja la necessitat d'analitzar-los des de m ltiples variables, depenent dels factors i funcions que els condicionen. La tipologia dels llenguatges documentals pot variar, doncs, depenent del punt de vista del seu estudi, generalment segons:

- la naturalesa dels termes d'indexaci  (s mbols codificats o paraules del llenguatge natural), parlem de:
  - Llenguatges codificats
  - Llenguatges naturals

- el nivell d'anàlisi en el moment de la indexació dels documents: representació del tema, conceptes o paraules del document, parlem de:  
Sistemes de classificació  
Llenguatges de descriptors  
Llenguatges de paraules clau
- el nivell de control exercit sobre els termes d'indexació, parlem de:  
Llenguatges lliures  
Llenguatges controlats
- el nivell de coordinació i combinació dels termes en el moment de la indexació o la recuperació, parlem de:  
Llenguatges precoordinats  
Llenguatges postcoordinats
- l'estructura o manera d'agrupar i presentar els termes d'indexació, parlem de:  
Llenguatges d'estructura jeràrquica  
Llenguatges d'estructura combinatòria  
Llenguatges d'estructura sintàctica

Cap llenguatge documental és totalment pur, els diferents aspectes es poden barrejar en un mateix llenguatge documental i presentar avantatges i inconvenients en el procés d'indexació o recuperació.

La tipologia dels llenguatges documentals és diversa depenent de...

...les característiques, naturalesa i funció estudiades.

Cap llenguatge documental és totalment pur...

...tots combinen diferents aspectes i presenten avantatges i inconvenients en la indexació i recuperació del contingut dels documents.

### **Vegeu també**

Els nuclis d'aquest mateix mòdul:

"Llenguatge codificat o natural?"

"Llenguatges documentals i nivell d'indexació"

"Llenguatges controlats o lliures?"

"Precoordinats o postcoordinats?"

"Estructures dels llenguatges documentals"

### **Bibliografia**

**García Gutiérrez, A.** (1989). "Los lenguajes documentales". En: J. López Yepes (comp.). *Fundamentos de información y documentación* (págs. 313-342). Madrid: Eudema (Manuales).



**Perpinyà Morera, R.** (1995). "Los lenguajes documentales". En: E. Fuentes (ed.). *Manual de documentación periodística* (págs. 112-132). Madrid: Síntesis (Biblioteconomía y Documentación, núm. 6).

**Pinto Molina, M.** (1989). "El análisis de contenido: la indización de documentos". En: J. López Yepes (comp.). *Fundamentos de información y documentación* (págs. 347-374). Madrid: Eudema (Manuales).

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

## 6. Llenguatge codificat o natural?

El contingut dels documents en un llenguatge documental es representa mitjançant els termes d'indexació. Aquests poden ser:

- Una paraula o expressió del llenguatge natural o comú (paraules clau i descriptors)
- Un codi de classificació numèric o alfanumèric (símbol de classe)

L'adopció d'un o altre tipus de termes d'indexació ens permet de fer una distinció entre llenguatges codificats i llenguatges naturals.

L'ús d'un determinat tipus de terme d'indexació no és un caprici, sinó que respon a les necessitats, naturalesa i objectius de cada tipus de llenguatge documental.

### 6.1. Llenguatges codificats

Són llenguatges que fan servir un sistema de notació per a la indexació dels documents i el plantejament de les consultes.

La notació és un codi artificial compost de símbols (numèrics, alfabètics...) que tradueix les matèries d'una classificació bibliogràfica en índexs ordenats.

Aquests llenguatges codificats es fan servir tradicionalment en els sistemes de classificació on estan plenament justificats:

- La necessitat d'extreure de manera sintètica la matèria principal del document i representar-la en una determinada posició d'un índex preestablert representa la traducció a un únic terme o signe precoordinat i codificat.
- A la vegada aquests codis són emprats per a facilitar l'ordenació física dels documents. La notació permet de conciliar l'ordenació i la classificació.

No obstant això, a vegades és necessari de fer la traducció d'una notació codificada d'un índex classificatori al llenguatge natural per a facilitar la consulta a l'usuari no expert.

És el cas de productes elaborats pels sistemes documentals i adreçats a satisfer les demandes d'informació dels seus usuaris com, per exemple, els butlletins d'índexs per matèries. Aquests són butlletins bibliogràfics editats periòdicament que poden contenir referències de:

- documents registrats per un sistema documental durant un temps determinat
- articles apareguts en una sèrie de publicacions periòdiques durant un temps determinat (butlletí de sumaris)
- novetats bibliogràfiques en un camp o disciplina, etc.

Cada entrada de l'índex conté:

- un o més termes d'indexació posats com a encapçalaments i ordenats alfabèticament
- una o més referències de documents. Aquesta referència pot ser completa o parcial (autor i títol, etc.).

Generalment a fi de no multiplicar les entrades d'un mateix document sota diferents termes d'indexació es fa servir un sistema de classificació per a sintetitzar la matèria principal del document.

Però, per tal que la consulta resulti més entenedora per a l'usuari, aquests símbols de classe es tradueixen al llenguatge natural, són les anomenades rúbriques.

"RÚBRICA: es la traducción en lenguaje natural de una materia correspondiente a un índice en la presentación de una clasificación documental. Ejemplo: 592 (índice) = invertebrados (rúbrica). Sinónimo: encabezamiento de materia."

J. Maniez, pàg. 23

*CINDOC. Revista Española de Documentación Científica. Índice acumulativo 1986-1996. Repertorio de títulos (pág. 9).*

## Repertorio de títulos

### 11. Ciencias de la Información

#### 1101. Generalidades

1

**¿Es posible planificar la ciencia?**

Moravcsik M.J.

REDC 1989, 12,(2): 181-189.

2

**¿Cómo evaluar la ciencia y a los científicos?**

Moravcsik M.J.

REDC 1989, 12, (3): 313-325.

3

**The knowledge base of a profession: the case of social work**

Berman Y.  
REDC 1994, 17, (2): 196-204.

4

**Enfoques objetivo y subjetivo del concepto de información**  
Fernández-Molina J.C.  
REDC 1994, 17, (3): 320-331.

5

**Modelo conceptual de un sistema de información documental**  
Codina L.  
REDC 1994, 17, (4): 440-449.

6

**Factores condicionantes de las operaciones documentales en medios de comunicación**  
Amat C.B., Castillo Blasco L.  
REDC 1996, 19,(2): 188-201.

7

**Información y relevancia. Una hipótesis acerca de cómo procesamos los seres humanos la información lingüística**  
Bonilla Álvarez S.  
REDC 1996, 19, (4): 392-410.

#### 1102. Organismos internacionales

8

**La historia de la FID**  
Gietz R.A.  
REDC 1986, 9, (3): 237-247.

9

**La FID en los años noventa. Información sobre los distintos comités**  
Keenan S., de Regt W.F. Dahlberg I., Gorkova V.I., Dosa M., Rajan T.N., Karlgren H., Mikhailov A., Pérez Álvarez-Ossorio J.R., Földi T.  
REDC 1986. 09. (3): 251-285.

10

**Futuro de la FID**  
Hill M.W.  
REDC 1986, 9, (3): 289-296.

11

**Pasado cercano y futuro inmediato de la FID/CLA. Una versión a mitad del camino**  
Arango H  
REDC M2, 15, (4): 331-339.

12

**Las políticas de información de la UNESCO y su repercusión en América Latina**  
Cruz-Ramos R Zetter-Leal J.  
REDC 1992, 15, (4): 340-358.

13

**La documentación en las instituciones de la Comunidad Europea. Una aproximación bajo la perspectiva del tratado de Maastricht**  
Rodríguez Bouyssi J.  
REDC 1993, 16, (1): 42-51.

#### 1104. Normalización

14

**Documentación. Índice de una publicación**  
UNE 50-111. ISO 999  
REDC 1989, 12, (1): 75-77.

15

**Documentación. Número internacional normalizado de publicaciones en serie -ISSN-**  
UNE 50-107. M 3297  
REDC 1989, 12, (2): 191-195.

16

**Información y documentación. Vocabulario. Parte 3.<sup>a</sup>. Adquisición, identificación y análisis de documentos y datos**

UNE 50-113/3.<sup>a</sup>. ISO 5127-3a)

REDC 1989, 12, (3): 327-348.

17

**Documentación. Directrices para el establecimiento y desarrollo de tesauros monolingües. Parte I**

UNE 50-106. ISO 2788

REDC 1989, 12, (4): 463-483.

18

**Documentación. Directrices para el establecimiento y desarrollo de tesauros monolingües. Parte II**

UNE 50-106. W 2788

REDC 1990, 13, (1): 601-629.

19

**Documentación. Sumario de las publicaciones periódicas**

UNE 50-110. ISO 18

REDC 1990, 13, (2): 757-758.

*Índice Acumulativo, 1986/1996.*

Els principals tipus de llenguatges documentals codificats són, doncs, els sistemes de classificació:

- Classificació Dewey (DDC)
- Classificació decimal universal (CDU)
- Classificació Colon (CC)
- Classificació de la Library of Congress (LCC)

## 6.2. Llenguatges naturals

Són llenguatges documentals que estan constituïts per elements lèxics del llenguatge usual: paraules i expressions del llenguatge natural no codificats.

El sistema analític d'indexació per conceptes, a diferència de la indexació per matèries o classificació, més sintètica, exigeix uns llenguatges més especialitzats i que es puguin actualitzar fàcilment per a adaptar-se al desenvolupament de la terminologia científica i tècnica. No interessa crear un sistema de notació amb codis universals, sinó fer ús del mateix llenguatge de la disciplina.

En un sistema de recuperació amb llenguatge natural, la matèria o conceptes dels documents i de les necessitats d'informació està representada per un vocabulari limitat de paraules emprades habitualment en el camp temàtic o disciplina i, per tant, més properes a l'usuari.

Els termes d'indexació que s'expressen mitjançant el llenguatge natural són:

- Les paraules clau
- Els unitermes

- Els descriptors

Els llenguatges documentals que fan servir termes d'indexació basats en el llenguatge natural o comú són:

- Encapçalaments de matèries
- Llistats de descriptors lliures
- Llistats d'autoritats o descriptors controlats
- Tesaurus
- Llistats de paraules clau

Els llenguatges documentals segons la naturalesa dels seus termes d'indexació poden ser...

Llenguatges codificats

- Representen la matèria del document amb un únic signe lingüístic: un codi o notació de símbols de classe (numèrics, alfanumèrics...).
- Faciliten la classificació i ordenació física dels documents.

Llenguatges naturals

- Representen els conceptes del document amb paraules i expressions del llenguatge natural no codificades.
- Els seus termes d'indexació poden ser paraules clau, unitermes o descriptors.

### **Vegeu també**

"Llistats de paraules clau", "Llistats de descriptors lliures", "Llistats d'autoritats o descriptors controlats" d'aquest mòdul.

"La indexació o descripció característica" del mòdul "Anàlisi i recuperació del contingut".

"Els sistemes de classificació" del mòdul "Sistemes de classificació".

"Els llistats d'encapçalaments de matèria" del mòdul "Llistes d'encapçalament de matèries".

"Els tesaurus" del mòdul "El tesaurus".

### **Bibliografia**

**Maniez, J.** (1993). *Los lenguajes documentales y de clasificación. Concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Sánchez Ruipérez (Biblioteca del Libro, 56).

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

## 7. Llenguatges documentals i nivell d'indexació

Atès que l'objectiu de la tasca documental és transmetre la informació continguda en els documents, és important conèixer quina quantitat d'informació d'un document original permet de transmetre un llenguatge documental determinat.

Partint dels tres sistemes d'indexació i anàlisi dels documents (per matèries, per conceptes i per paraules del document) podem establir una primera tipologia dels llenguatges documentals.

### 7.1. Sistemes de classificació i encapçalaments de matèries

Aquests tipus de llenguatges documentals permeten de representar de manera sintètica el contingut d'un document i ordenar físicament les referències i els documents. Són els tipus de llenguatges més idonis per a traduir la matèria del document: responen a la qüestió "quin és el tema del document?"

El seu objectiu és classificar la matèria específica del document, la síntesi total dels continguts tractats. Per a portar-lo a terme, hi ha d'haver una operació intel·lectual per part de l'analista del sistema.

#### Sistemes de classificació

Atesa la dificultat que representa extreure la matèria específica d'un document amb un enunciat simple, s'ha de recórrer a un sistema de classificació que sigui capaç de representar la complexa estructura del contingut informatiu.

Tradicionalment aquests sistemes han estat eines de treball vinculades a les necessitats de funcionament d'una biblioteca.

Entre els principals sistemes de classificació hi ha:

- Les classificacions enciclopèdiques, per facetes i especialitzades: classificació Dewey (DC), classificació decimal universal (CDU), classificació Colon (CC), classificació de la Biblioteca del Congrés (LC), classificació bibliogràfica de Bliss (BC).

#### Llistats d'encapçalament de matèries

Els encapçalaments de matèries<sup>2</sup> (*subject heading*) van aparèixer com la solució més idònia per a elaborar els catàlegs manuals a les biblioteques. La seva naturalesa (llenguatge precoordinat i sintètic) permetia:



<sup>(2)</sup>Malgrat l'origen i funció inicial dels llistats d'encapçalaments de matèries, s'ha de fer constar que l'evolució i automatització dels sistemes d'informació, que ha possibilitat l'ampliació del nombre d'entrades de matèries dins del catàleg, ha comportat l'evolució i desenvolupament d'aquest llenguatge fins al punt de convertir-se en un llenguatge combinatori, més pròxim al tesaurus que no pas als sistemes de classificació tradicionals. Podríem afirmar que les llistes són instruments per a la indexació analítica i sintètica, que combinen aspectes de tots dos sistemes.

- sintetitzar la matèria del document en un únic terme d'indexació per a no multiplicar les entrades al fitxer manual.
- emprar un llenguatge natural (no codificat amb signes i números) més pròxim als usuaris d'aquests sistemes documentals.

Los primeros avances en la descripción característica de los documentos fueron registrados en las bibliotecas a finales del siglo xix, con la aparición de las clasificaciones enciclopédicas y facetadas, basadas en la sistematización de materias. Fueron las bibliotecas -con esas clasificaciones- las creadoras del concepto de indización por matèries, pero no del término, al utilizar la denominación encabezamiento de materia (subject heading). Consiste en la correlación sucesiva de diferentes encabezamientos que expresan el tema o temas de un documento. Para ello, serán requisitos necesarios la uniformidad y el haber establecido previamente una lista a la que deberá ajustarse la redacción de los términos de indización.

M. Pinto Molina, 1989, pàg. 384

## 7.2. Llenguatges o vocabularis de descriptors

Aquests tipus de llenguatges permeten de representar de manera analítica el contingut d'un document. Són els llenguatges més idonis per a traduir els conceptes del document: responen a la qüestió "quins són els conceptes del document?"

Impliquen sempre una operació intel·lectual per part de l'analista del sistema, que pot ser:

- Només en les dues primeres operacions del procés d'indexació (examen del document i identificació dels conceptes principals, per exemple, llistats de descriptors lliures)
- En el procés sencer (examen, identificació i traducció dels conceptes al llenguatge documental emprat pel sistema, per exemple, llistats de descriptors controlats o llistats d'autoritats i els tesaurus)

La seva aplicació va lligada a l'existència de sistemes automatitzats que facin possible la recuperació de la informació mitjançant la combinació dels descriptors.

Aquests llenguatges documentals van representar un gran avenç en la descripció del contingut dels documents i, a més, una aproximació coherent envers els usuaris dels sistemes d'informació, que veien l'oportunitat de sortir dels

laberints de la CDU i les seves notacions numèriques i d'emprar la seva pròpia terminologia (llenguatge natural o comú) en el moment de recuperar la informació.

Són els llenguatges documentals més emprats als centres de documentació especialitzats.

Els principals llenguatges de descriptors són:

- Els llistats de descriptors lliures
- Els llistats d'autoritats o descriptors controlats
- Els tesaurus

### **7.3. Llenguatges de paraules clau**

Aquests tipus de llenguatges també permeten de representar de manera analítica el contingut d'un document. Però, a diferència dels llenguatges de descriptors, no necessiten la intervenció de cap operació intel·lectual per part de l'analista.

Són els llenguatges documentals que extreuen les paraules significatives contingudes en el títol, resum i a vegades el text complet del document. Generalment són el resultat de l'aplicació d'un sistema d'indexació automàtica.

Actualment són llenguatges emprats per tots els sistemes d'informació que disposen d'una base de dades o un catàleg automatitzat. Normalment el seu ús dins del sistema documental es complementa, gairebé sempre, amb altres llenguatges documentals que presenten una estructura més controlada i unívoca en el moment de la indexació i la recuperació dels documents, tant si són tesaurus de descriptors com si són encapçalaments de matèries.

Tenen l'avantatge que són simples d'usar i que registren els majors nivells de coherència en la indexació, ja que extreuen els conceptes significatius tal com es troben en el document original. Però són ambigus en la pràctica, ja que en emprar el llenguatge natural dels documents, els seus termes presenten multiplicitat de sentits i manca d'univocitat.

Els principals llenguatges de paraules clau són els:

- Llistats de paraules clau

Els diferents nivells d'indexació no són excloents, ja que poden ser perfectament complementaris en un determinat sistema documental. Així, un document pot estar representat per un codi de classificació, un nombre determinat de descriptors i a més pot ser recuperable per mitjà de les paraules clau del seu títol, resum o text.

La quantitat d'informació que un sistema documental extreu d'un document (segons les necessitats dels seus usuaris) determina el nivell d'indexació i el llenguatge documental que cal emprar.

Si és necessari sintetitzar i extreure la matèria del document, s'han de fer servir els sistemes de classificació i els encapçalaments de matèries.

Si interessa recuperar de manera més precisa els conceptes continguts en els documents, s'han d'emprar els llenguatges de descriptors i paraules clau.

#### **Vegeu també**

"Els llistats de paraules clau", "Els llistats de descriptors lliures" i "Els llistats d'autoritats o descriptors controlats" d'aquest mateix mòdul.

"La indexació o descripció característica" i "El procés d'indexació" del mòdul "Anàlisi i recuperació del contingut".

"Els sistemes de classificació" del mòdul "Sistemes de classificació".

"Els encapçalaments de matèries" del mòdul "Llistes d'encapçalaments de matèries".

"Els tesaurus" del mòdul "El tesaurus".

#### **Bibliografia**

Pinto Molina, M. (1989). "El análisis de contenido: la indicación de documentos". En: J. López Yepes (comp.). *Fundamentos de información y documentación* (págs. 347-374). Madrid: Eudema (Manuales).

## 8. Llenguatges lliures o controlats?

La comunicació documental entre la informació continguda en el document i l'usuari final es pot establir mitjançant el llenguatge natural o comú o bé per mitjà de llenguatges artificials construïts per a garantir una millor recuperació de la informació.

L'ús d'un sistema o l'altre representa l'existència o no d'un control del vocabulari emprat per a indexar i recuperar els documents. Així, segons el nivell de normalització de la seva terminologia, els llenguatges documentals poden ser els següents.

### 8.1. Llenguatges lliures

Els llenguatges lliures són vocabularis formats per termes d'indexació extrets del llenguatge natural emprat en els documents (títol, resum i text complet), que es fan servir per a representar-ne i recuperar-ne el contingut. Normalment són el resultat de l'aplicació d'un sistema d'indexació automàtica.

Són llenguatges construïts *a posteriori*, és a dir, que es nodreixen dels termes d'indexació que extreuen dels documents que arriben al sistema.

El llenguatge lliure més comú en tots els sistemes documentals és el fitxer invers o diccionari d'una base de dades documental. Aquest fitxer recull totes les paraules significatives del títol, resum i altres camps definits a la base de dades ordenades en una mateixa seqüència alfabètica. L'usuari pot recuperar la informació per mitjà de la combinació d'aquests termes fent ús dels operadors de cerca (booleans, proximitat, truncaments...).

Els llenguatges lliures reporten molts avantatges en el moment de la indexació (costos molt baixos, actualització immediata, màxima coherència), però presenten molts problemes en el moment de la recuperació, derivats de l'ambigüitat del llenguatge natural (sinonímia, polisèmia i homonímia).

"Los lenguajes libres adolecen de todas las dificultades y ambigüedades de la indización por lenguaje natural (polisemia y sinonimia). Aunque debe reconocérseles las ventajas derivadas de la riqueza terminológica del lenguaje, del grado de coherencia (puesto que un mismo documento siempre estará indizado con los mismos términos) y de que los costes de fabricación son prácticamente nulos".

R. Perpinyà, 1995, pàg. 117

Els principals llenguatges lliures són:

- Els llistats de paraules clau

- Els llistats de descriptors lliures

## 8.2. Llenguatges controlats

Els llenguatges controlats són llenguatges artificials que tenen com a objectiu la representació unívoca del contingut dels documents i de les consultes: un únic concepte per terme i un únic terme per concepte.

Constitueixen un vocabulari controlat i finit, és a dir, un limitat conjunt de termes que han de fer servir els analistes i els usuaris que recuperen els documents.

Només els termes d'indexació inclosos en el vocabulari es poden emprar en el moment de la indexació i la recuperació.

Eliminen l'ambigüitat del llenguatge natural perquè:

- seleccionen un únic terme d'indexació entre els seus possibles sinònims.
- atorguen al terme d'indexació un únic significat entre els que pugui tenir en el llenguatge natural i col·loquial.

Són llenguatges construïts *a priori*, és a dir, abans de començar la indexació dels documents. Generalment requereixen uns costos alts de construcció (personal i temps), ja que s'han d'estudiar amb aprofundiment les característiques del sistema documental on s'ha d'aplicar.

Els responsables de la seva construcció han de conèixer les necessitats d'informació dels usuaris del sistema (nivell de precisió i exhaustivitat) i els dominis del camp temàtic o disciplina que ha de cobrir el vocabulari. Generalment s'estableix un llistat de termes d'indexació del qual s'eliminen els sinònims i es precisa el significat atorgat a cada terme. Una vegada elaborat, s'aplica al sistema documental on se'n pot comprovar el grau d'eficàcia en la indexació i recuperació dels documents.

Són els llenguatges estrictament documentals.

"Un vocabulario controlado es un subconjunto del lenguaje natural de un campo temático (...) puede ser tan simple como una lista de términos autorizados. Sin embargo, lo más frecuente es que tenga algún tipo de estructura, como es el caso de los esquemas de clasificación, los tesauros y las listas de encabezamiento de materias".

F. Lancaster, 1995, pàg. 177

Els principals tipus de llenguatges controlats són:

- Els tesaurus
- Els llistats d'autoritats o descriptors controlats
- Els llistats d'encapçalaments de matèries
- Els sistemes de classificació

Segons el nivell de control del termes d'indexació, els llenguatges documentals poden ser...

#### Llenguatges lliures

- Extreuen els termes del llenguatge natural dels documents (títol, resum, text)
- Es construeixen a posteriori.
- Presenten avantatges en la indexació, però inconvenients en la recuperació (sinonímia i polisèmia)

#### Llenguatges controlats

- Tenen com a objectiu la univocitat dels termes d'indexació, ja que eviten la sinonímia i la polisèmia.
- Només els termes recollits en el llenguatge es poden emprar en la indexació i recuperació.
- Són construïts a *priori* pels responsables del sistema documental.

#### **Vegeu també**

"El llenguatge natural en la comunicació documental", "Per què llenguatges artificials?" i "Els llistats d'autoritats" d'aquest mateix mòdul.

"Els llistats d'encapçalaments de matèries" del mòdul "Llistes d'encapçalaments de matèries".

"Els tesaurus" del mòdul "El tesaurus".

#### **Bibliografia**

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

Perpinyà Morera, R. (1995). "Los lenguajes documentales". En: E. Fuentes (ed.), *Manual de documentación periodística* (págs. 112-132). Madrid: Síntesis (Biblioteconomía y Documentación, núm. 6).

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

## 9. Precoordinats o postcoordinats?

Els llenguatges documentals segons la seva naturalesa (sintètica o analítica) presenten necessitats diferents en el moment d'assignar i recuperar els termes d'indexació:

- En els llenguatges sintètics s'ha de traduir la matèria del document a un únic terme d'indexació. Aquesta operació significa que l'analista ha de fer l'operació intel·lectual de síntesi i coordinació dels termes en el moment de la indexació.
- En els llenguatges analítics els conceptes del document es tradueixen en diferents termes d'indexació. La recuperació de la informació la porta a terme l'usuari combinant i coordinant els diferents termes mitjançant una estratègia de cerca.

Podem afirmar, doncs, que els llenguatges documentals, depenent del sistema emprat per a combinar els termes en el moment de la indexació i la recuperació, es poden dividir en dos grans tipus.

### 9.1. Llenguatges precoordinats

Són els llenguatges documentals que coordinen els diferents conceptes que conformen una matèria en el moment de la indexació, és a dir, abans d'emmagatzemar-lo.

Els sistemes precoordinats preveuen generalment una única matèria per document i, per tant, són eficaços per a ordenar les obres en prestatgeries i fitxers manuals.

La precoordinació generalment permet només un únic terme d'indexació, però proporciona en una única cerca els elements essencials de la consulta, i possibilita l'ordenació física dels documents i de les referències (fitxers manuals).

La funció d'aquest tipus de llenguatge tradicionalment està lligada a les necessitats dels catàlegs manuals en els quals no es poden combinar els termes.

#### **Exemple**

Així, per exemple, en un llenguatge precoordinat com el llistat d'encapçalaments de matèries, el concepte "psicologia infantil" té una única entrada ("infants-psicologia"), mentre que en un llenguatge postcoordinat es mantenen els dos termes per separat, sense

necessitat de crear cap concepte coordinat, ja que el significat s'estableix amb la combinació de tots dos "psicologia" i "infantil".

En el primer cas, la referència documental es pot trobar al catàleg al costat d'altres entrades encapçalades pel tema infants (infants-biografia, infants-cirurgia, infants-creixement, etc.).

Si l'usuari consulta el catàleg per les entrades de psicologia trobarà una referència que el remetrà a l'entrada o terme acceptat:

psicologia infantil

v. (vegeu) **infants-psicologia**

"Los catálogos manuales implican que un documento sólo se puede buscar por un concepto pero no por la combinación de dos o más, de ahí la necesidad de la precoordinación (cuyo instrumento básico es la lista de encabezamientos de materia), mientras que las prestaciones de los sistemas informáticos aplicados a la automatización de la descripción de documentos son múltiples y permiten que la yuxtaposición de términos para llegar a formular un concepto no se haga en el momento del análisis documental sino en la operación de la consulta."

R. Perpinya, 1995, pàg. 117

Els principals llenguatges precoordinats són:

- Els sistemes de classificació
- Els llistats d'encapçalaments de matèries

## 9.2. Llenguatges postcoordinats

Són els llenguatges documentals que permeten la coordinació de conceptes en el moment de la recuperació.

Juxtaposen els termes d'indexació per a coordinar-los després de l'emmagatzematge. Aquest sistema permet de fer servir múltiples termes d'indexació i, per tant, d'aprofundir en l'anàlisi del contingut del document.

Cada terme d'indexació representa un punt d'accés als documents: com més termes, més possibilitats de recuperar el document des de diferents conceptes i temes. En el moment de la cerca, s'estableix el significat o concepte desitjat per la combinació d'un o més termes d'indexació.

Els termes es poden fer servir lliurement, sense referència a un altre amb el qual puguin tenir relació.

### Exemple

Així, per exemple, un document que tracta sobre les troballes i estructures de jaciments arqueològics del neolític final a Catalunya s'ha indexat amb els conceptes següents:

Jaciments arqueològics



Objectes arqueològics

Estructures funeràries

Ritus funeraris

Restes humanes

Datació arqueològica

Indústria lítica

Neolític final

Catalunya

I podrà ser recuperat en resposta a consultes realitzades sobre una combinació qualsevol dels conceptes:

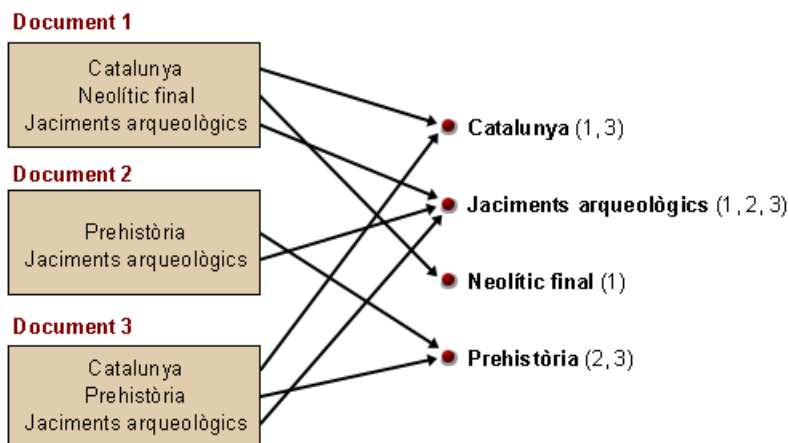
- Informació sobre la indústria lítica en el neolític final  
(*indústria lítica i neolític final*)
- Informació sobre ritus, objectes i costums funeraris en qualsevol període  
(*objectes arqueològics, ritus funeraris i estructures funeràries*)
- Informació sobre els problemes de datació de les restes humanes en el neolític final  
(*restes humanes, neolític final i datació arqueològica*)
- Informació sobre els jaciments arqueològics existents a Catalunya al període del neolític final  
(*jaciments arqueològics, neolític final i Catalunya*)

La postcoordinació permet d'emprar un gran nombre de termes d'indexació o punts d'accés al document, però s'han d'aplicar en sistemes documentals automatitzats que disposin d'un fitxer invers que faci les operacions d'identificació i localització dels documents pertinents.

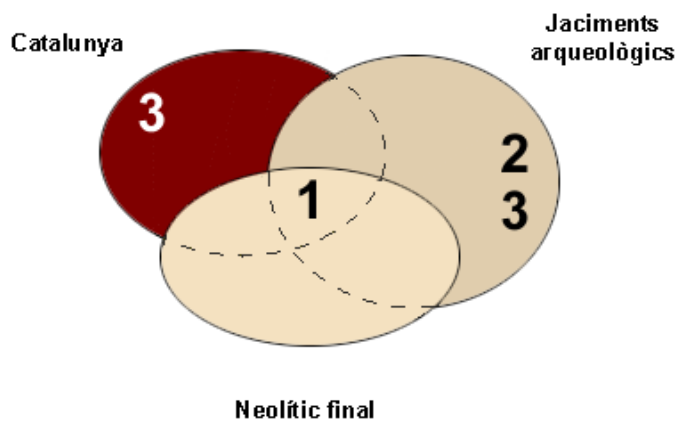
L'existència d'un sistema automatitzat, mitjançant el fitxer invers<sup>3</sup>, permet que en el moment d'indexar els conceptes se situïn un darrere l'altre (segons el sistema i llenguatge emprat), i que en el moment de la recuperació s'expressi una necessitat d'informació concreta mitjançant una estratègia de cerca.

<sup>(3)</sup> **Com actua el fitxer invers d'una base de dades**

Al nostre sistema documental arriben tres documents que s'analitzen, s'indexen amb diferents descriptors i s'emmagatzemen. Per tal que la recuperació de la informació sigui possible, el sistema de gestió documental ha guardat en el seu fitxer invers els diferents descriptors amb la informació associada del número de document en el qual estan presents. És a dir, elabora un índex dels conceptes que emmagatzema.



D'aquesta manera, si un usuari planteja una cerca sobre jaciments arqueològics a Catalunya durant el neolític final, el sistema buscarà en el seu fitxer invers els conceptes i localitzarà els documents que compleixen la triple condició (que tractin de jaciments arqueològics, que el període sigui el neolític final i que estiguin a Catalunya) i n'exclourà la resta. En aquest cas el resultat seria el document número 1.



Els principals llenguatges documentals postcoordinats són:

- Els llistats de paraules clau
- Els llistats de descriptors lliures
- Els llistats d'autoritats o descriptors controlats
- Els tesaurus

Els llenguatges documentals, segons el sistema emprat per a combinar els termes d'indexació (en el moment de la indexació o en el moment de la recuperació), poden ser...

#### Precoordinats

- Coordinen i sintetitzen els diferents conceptes que conformen una matèria en un únic terme d'indexació assignat per l'analista en el moment de la indexació.
- Permeten l'ordenació física dels documents i de les referències (fitxers manuals).

#### Postcoordinats

- Coordinen els diferents termes d'indexació en el moment de la recuperació o consulta.
- Permeten d'emprar un gran nombre de termes d'indexació en l'anàlisi del contingut.
- Només es poden aplicar a sistemes automatitzats.

#### **Vegeu també**

"Llistats de paraules clau", "Llistats de descriptors lliures" i "Llistats de descriptors controlats" en aquest mateix mòdul.

"Els sistemes de classificació" del mòdul "Sistemes de classificació".

"Els llistats d'encapçalaments de matèries" del mòdul "Llistes d'encapçalaments de matèries".

"Els tesaurus" del mòdul "El tesaurus".

#### **Bibliografia**

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

**Perpinyà Morera, R.** (1995). "Los lenguajes documentales". En: E. Fuentes (ed.). *Manual de documentación periodística* (págs. 112-132). Madrid: Síntesis (Biblioteconomía y Documentación, núm. 6).

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

## 10. Estructures dels llenguatges documentals

Els llenguatges documentals, segons la manera d'agrupar i presentar els termes d'indexació, poden presentar diferents tipus d'estructures:

### 10.1. Llenguatges d'estructura jeràrquica o arborescent

Són llenguatges que segueixen un ordre lògic que aproxima els conceptes més específics i els agrupa dins dels més generals. Tots els termes depenen d'un altre superior, de significat més genèric, que a la vegada s'integra en un altre superior i així successivament.

Aquesta estructura comporta una classificació sistemàtica lineal, en la qual els conceptes estan ordenats seguint una jerarquia. Cada terme està representat per un símbol numèric, alfabètic o alfanumèric que n'indica la situació dins la cadena jeràrquica.

Els llenguatges documentals que presenten aquesta estructura jeràrquica són les tradicionals classificacions enciclopèdiques i les especialitzades.

Les categories que estableixen aquests tipus de llenguatges estan definides *a priori*, és a dir, són sistemes tancats que no admeten noves inclusions. Totes les possibles matèries que s'han de tractar estan ja representades en el llenguatge.

Tenen una estructura fixa que es repeteix en la majoria dels sistemes de classificació i que està formada pels tres components bàsics detallats a continuació:

- Les taules en les quals es mostren sistemàticament les matèries i les seves relacions. Generalment unes taules de classificació que es componen d'uns codis organitzats en un determinat nombre de classes i subclasses.
- La notació o codi numèric o alfanumèric que és emprat per a representar la matèria dels documents en índexs o catàlegs i també en la seva ordenació física.
- Índex alfabètic que localitza les diferents matèries dins les taules.

La rigidesa d'aquesta estructura té com a conseqüència una manca de flexibilitat en la indexació o representació de conceptes, de manera que només són útils per a la descripció general de continguts, per a l'assignació de la matèria del document, però no permeten una anàlisi aprofundida en la indexació.

#### Exemple

Exemple d'un llenguatge jeràrquic, la CDU:  
5: Ciències pures  
53: Física  
535: Òptica  
535.2: Propagació de la llum  
535.21: Efectes de la propagació de la llum  
[Exemple extret de J. Rubió, pàg. 16]

No obstant aquests inconvenients, la naturalesa de l'estructura jeràrquica té aspectes positius, ja que permet la reunió dels termes d'indexació per temes. Aquest avantatge és recollit pels tesaurus, en els quals, combinat amb altres tipus de presentacions, ens apareix renovat i en un entorn d'un llenguatge de descriptors postcoordinat i analític.

## 10.2. Llenguatges d'estructura combinatòria o associativa

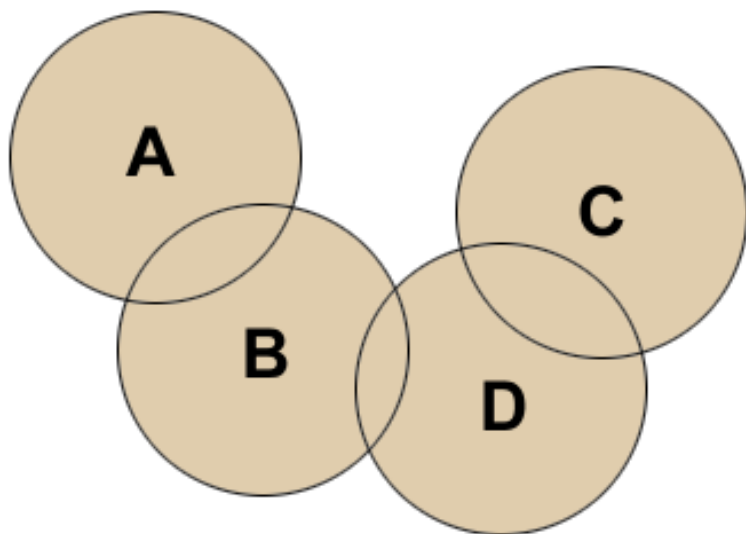
Són llenguatges en els quals els termes d'indexació s'organitzen alfabèticament i es combinen entre si, en el moment de la indexació i la recuperació, sense quedar sotmesos a una posició determinada en el llenguatge.

Aquest tipus d'estructura proporciona un accés més immediat a la informació, però tendeix a dispersar els termes de significat afí.

Els conceptes s'organitzen de manera independent i només se'n produeix la intersecció en les operacions d'indexació i recuperació.

Estructura associativa

Intersecció de termes d'indexació



Sorgeixen com a reacció a la rigidesa dels sistemes tradicionals de tractament de la informació (taules de classificació) i a la necessitat de recollir el nou vocabulari que anava apareixent en les diferents disciplines científiques.

"Para poder recuperar la documentación más reciente que utiliza un léxico científico modernizado, los lenguajes documentales que han de aplicarse en el proceso de búsqueda deben ser reformados mediante la inclusión de nuevos términos o el rechazo de los menos usados, según la línea impuesta por la indización que no es más que la traducción documental de términos representativos de la lengua natural a descriptores."

Aquesta ordenació alfabètica, des del punt de vista lògic, és totalment aleatòria, ja que tendeix a dispersar termes amb significats propers o afins.

No obstant això, hi ha llenguatges combinatoris que incorporen a aquesta estructura alfabètica primària tota una xarxa de relacions d'associació<sup>4</sup>. Aquestes proporcionen a l'usuari una considerable ajuda en la indexació i recuperació, gràcies a l'establiment d'associacions d'idees i de punts de vista entre diferents termes d'indexació.

<sup>(4)</sup>Les relacions d'associació o afinitat entre termes es fan explícites:

- **En un llistat alfabètic de matèries**

Mitjançant l'ús de referències entre termes afins semànticament.

**Vegeu també (vt):** Poden enviar d'un encapçalament a uns altres de més específics o relacionats i viceversa, o bé a un altre grup d'encapçalaments que pertanyen a una mateixa classe.

<b>Avortament</b> vt Educació sexual	<b>Embaràs</b> vt Avortament Cura prenatal ...etc.
---	---

#### Malalties

vt Salut i els noms de les malalties, per exemple:

Tuberculosi

- En un tesaurus

Mitjançant una relació d'associació entre termes afins semànticament, expressada per l'abreviatura TR.

**TR:** Terme relacionat o *related term* (RT). S'estableix de manera recíproca entre els descriptors del tesaurus amb afinitats de significat.

#### Avortament

TR Embaràs

#### Embaràs

TR Avortament

A la pràctica, en els llenguatges documentals més elaborats, tipus tesaurus, es prova d'establir una estructura que integri la flexibilitat dels llenguatges combinatoris i la rigidesa dels jeràrquics o classificatoris.

Les principals característiques dels llenguatges combinatoris o associatius són les següents:

- La seva flexibilitat permet la incorporació de nous termes, i l'eliminació dels termes obsolets, sense afectar la resta de l'estructura. Aquest és un aspecte que facilita la revisió i actualització del llenguatge documental.
- L'especialització, en major o menor grau, a l'àmbit científic o tècnic en el qual s'apliquen.

- La multiplicitat de combinacions que es poden fer entre els termes del seu vocabulari.

Entre els llenguatges d'estructura combinatòria o associativa hi ha:

- Llistats de paraules clau
- Llistats de descriptors lliures
- Llistats d'autoritats o de descriptors controlats
- Llistats d'encapçalaments de matèries
- Tesaurus

### 10.3. Llenguatges d'estructura sintàctica

Són llenguatges més elaborats que els anteriors perquè estan formats per un conjunt de termes d'indexació i una gramàtica que exclou la possible ambigüïtat en relacionar sintàcticament els conceptes.

Un conjunt de descriptors formen una frase documental.

Si assignem a un document titulat "L'ensenyament de la història de la ciència" els termes d'indexació següents:

història / ensenyament / ciència

El document es podrà recuperar també en consultes sobre:

- història de la ciència de l'ensenyament
- la ciència de l'ensenyament de la història

Es produeix un soroll documental en la recuperació amb la presència de documents no pertinents per al tema de la consulta.

Una manera d'evitar el soroll és eliminar l'ambigüïtat dels termes d'indexació amb l'ús de termes compostos o modificadors:

Modificador	Terme compost
ciència (història)	història de la ciència

Aquesta és la tendència que segueixen la majoria dels tesaurus.

Una altra opció per a evitar l'ambigüitat i el soroll documental en la recuperació és l'establiment d'una sintaxi obligatòria en la posició dels termes en el moment de la indexació:

1/ensenyament, 2/història, 3/ciència

D'aquesta manera s'expressa una relació sintàctica entre els termes d'indexació que conformen una frase.

En el moment de la recuperació, el sistema ha de permetre seleccionar els termes en una determinada posició.

Aquests llenguatges més elaborats, segons els experts, en un futur modificaran els tesaurus, encara que hagin partit d'aquests.

S'aconseguirà eliminar la possibilitat de falses relacions mitjançant la intervenció d'operadors simples o normes bàsiques de disposició o correlació dels termes d'indexació.

"Los lenguajes clásicos, jerárquicos o combinatorios tienen a lo sumo un desarrollo léxico y una red de relaciones semánticas, pero no sintácticas, pues son lenguajes sin gramática. Esta falta de relación sintáctica entre los descriptores puede producir una cierta distorsión en la interpretación, bien por el hombre, bien por el sistema, de la frase documental. Ejemplo: /documentación/, /técnica/, /historia/ puede significar "documentación técnica de la historia", o al menos así lo entenderá, por ejemplo, el ordenador si programamos un código. El hecho de establecer un orden fijo en la aparición de los descriptores ya supone el primer indicio de unas reglas gramaticales: 1/técnica, 2/documentación, 3/historia será técnicas de la documentación histórica. Este método puede ser aplicado a los tesoros convencionales."

A. García Gutiérrez, 1989, pàg. 340-341



Els llenguatges documentals segons el sistema d'agrupar i estructurar els termes d'indexació poden ser...

#### Llenguatges jeràrquics

Presenten una estructura en la qual els termes estan ordenats seguint una jerarquia (els conceptes més específics dins els més generals)

#### Llenguatges combinatoris o associatius

Presenten una estructura alfabètica dels termes que es combinen en el moment de la indexació i la recuperació. Alguns incorporen una relació associativa entre termes amb afinitats de significats.

#### Llenguatges sintàctics

Presenten una estructura alfabètica de termes que es combinen en el moment de la indexació segons un ordre fix i obligatori d'aparició, i així estableixen una gramàtica entre els diferents termes.

### Bibliografia

A. García Gutiérrez (1989). "Los lenguajes documentales". En: J. López Yepes (comp.). *Fundamentos de información y documentación* (págs. 313-342). Madrid: Eudema (Manuales).

#### Vegeu també

"Llistats de paraules clau", "Llistats de descriptors lliures" i "Llistats d'autoritats" d'aquest mòdul.

"Els sistemes de classificació" del mòdul "Sistemes de classificació".

"Els encapçalaments de matèries" del mòdul "Llistats d'encapçalaments de matèries".

"Els tesaurus" del mòdul "El tesaurus".

## 11. Quin llenguatge documental?

Els diferents criteris emprats per a definir els llenguatges documentals (nivell d'anàlisi, control, coordinació, signes lingüístics, estructura) es poden combinar en la pràctica i, per tant, cap llenguatge és totalment pur, sinó que dona lloc a múltiples tipus:

- Llenguatges postcoordinats amb vocabulari lliure (llistats de paraules clau o uniterms)
- Llenguatges precoordinats amb vocabulari controlat (classificacions jeràrquiques)
- Llenguatges postcoordinats amb vocabulari controlat (tesaurus)
- Etc.

Cada tipus de llenguatge té, doncs, els seus avantatges i inconvenients en destaquen els següents:

- Els llenguatges precoordinats permeten pocs termes d'indexació per document, però faciliten en una única cerca els elements essencials d'informació.
- Els llenguatges postcoordinats permeten l'ús de múltiples termes d'indexació, però tenen com a intermediaris fitxers especials que necessiten una cerca en dos temps: primer la identificació dels documents pertinents i després la seva localització.
- Els llenguatges controlats impliquen una gran inversió en temps i personal per a la seva construcció i manteniment, però garanteixen una recuperació ràpida i eficaç dels documents.
- Els llenguatges lliures tenen uns costos molt baixos, però presenten més problemes en la recuperació (soroll i silenci documental).

Aquest quadre<sup>5</sup> ordena els llenguatges documentals a partir dels dos criteris fonamentals següents:

- De dalt a baix, depenent del grau de control (de més a menys)
- D'esquerra a dreta, depenent del grau de precoordinació (de més a menys)

<sup>(5)</sup>Basat en el quadre general dels llenguatges documentals elaborat per J. Maniez, pàg. 221.

	<b>+ sintètics (matèries)</b>		<b>analítics - (conceptes)</b>
<b>+ GRAU</b>	Classificació jeràrquica Llistat d'encapçalaments de matèries	Classificació per facetes	Tesaurus Llistat d'autoritats
<b>DE</b>			Índex de les classificacions
<b>CONTROL -</b>		Índex de títols	Llistat de descriptors lliures Llistat de paraules clau Uniternes

Els llenguatges documentals s'han de seleccionar d'acord amb les necessitats dels usuaris del sistema documental. D'altra banda, l'adopció d'un llenguatge documental determinat no exclou l'ús complementari d'un altre: un document pot estar representat per un codi de classificació, una sèrie de descriptors seleccionats d'un tesaurus o un resum i a més pot ser recuperat per mitjà d'un llistat de paraules clau.

Aquesta complementarietat dels diferents sistemes i llenguatges és el resultat de l'evolució i convergència de les necessitats de recuperació de la informació mitjançant la curta història dels llenguatges documentals.

## Activitat

### Activitat 1: Completeu el gràfic

#### Descripció

En el gràfic següent teniu, d'una banda, totes les tipologies de llenguatges documentals i, de l'altra, en la part superior, alguns dels principals llenguatges documentals.

Assenyaleu, posant una X a la casella corresponent, les tipologies i característiques que afecten cada llenguatge.

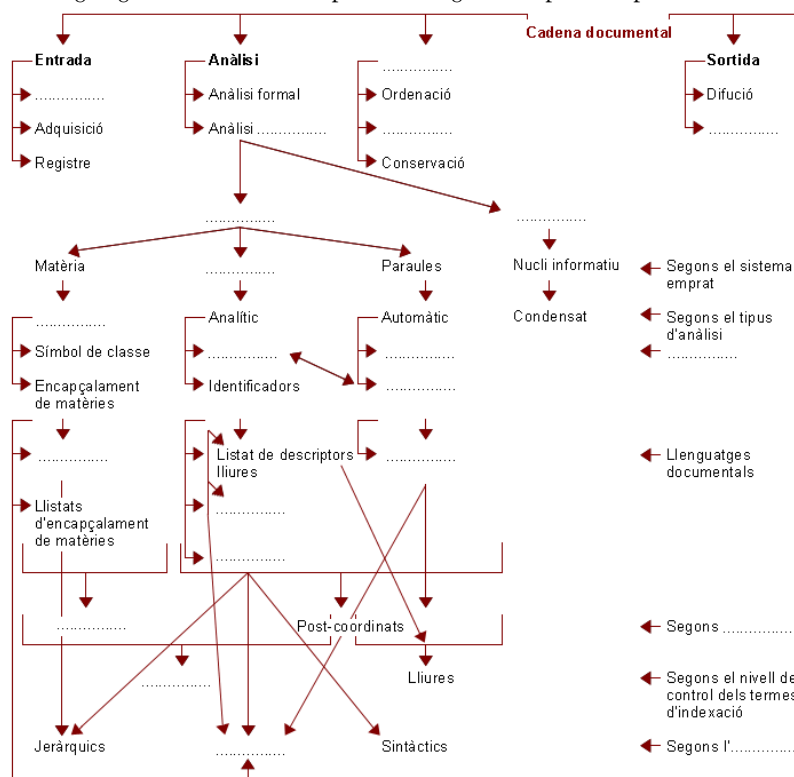
<b>Tipologia de llenguatges documentals</b>	<b>Classificació decimal universal</b>	<b>Llistats d'encapçalaments de matèries</b>	<b>Llistats de descriptors lliures</b>	<b>Llistats d'autoritats</b>	<b>Tesaurus</b>	<b>Llistats de paraules clau</b>
Codificats						
Naturals						
Sintètics (matèria)						
Analítics (conceptes)						
Analítics (paraules)						
Controlats						
Lliures						
Precoordinats						
Postcoordinats						
Jeràrquics						
Combinatoris						
Sintàctics						

Solució

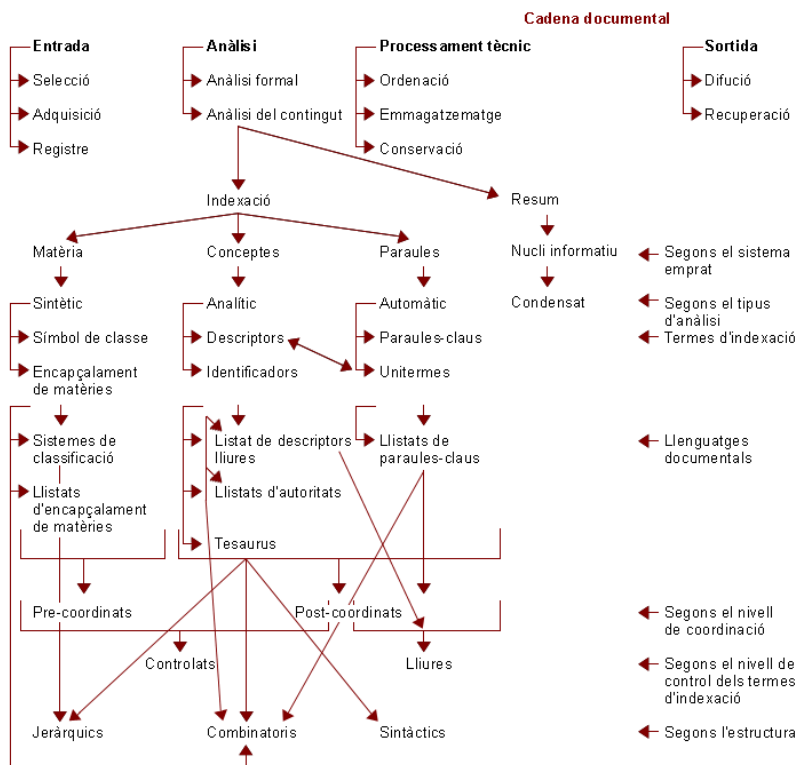
Tipologia de llenguatges documentals	Classificació decimal universal	Llistats d'encapçalaments de matèries	Llistats de descriptors lliures	Llistats d'autoritats	Tesaurus	Llistats de paraules clau
Codificats	X					
Naturals		X	X	X	X	X
Sintètics (matèria)	X	X				
Analítics (conceptes)			X	X	X	
Analítics (paraules)						X
Controlats	X	X		X	X	
Lliures			X			X
Precoordinats	X	X				
Postcoordinats			X	X	X	X
Jeràrquics	X				X	
Combinatoris		X	X	X	X	X
Sintàctics					X	

Activitat 2: Completeu el següent mapa conceptual

Poseu en ordre els vostres coneixements sobre l'anàlisi del contingut dels documents i els llenguatges documentals completant el següent mapa conceptual.



Solució



### 11.1. Passat i futur dels llenguatges documentals

La història dels llenguatges documentals reflecteix la ràpida evolució dels mètodes d'emmagatzematge i cerca de la informació documental.

En un principi hi havia els sistemes de classificació, que es remunten a l'antiguitat però que es van consagrar a la darrereria del segle XIX amb la construcció de les grans classificacions enciclopèdiques: la classificació Dewey, la classificació decimal universal i la de la Biblioteca del Congrés.

Al començament dels anys seixanta van aparèixer els primers tesaurus com a reacció a la rigidesa dels sistemes de classificació: aparició de nova terminologia, com a conseqüència del desenvolupament de les diferents disciplines científiques, que no estava recollida en les classificacions, necessitat creixent d'accedir a la informació mitjançant conceptes més específics i familiars per als usuaris.

La irrupció dels tesaurus i altres llenguatges combinatoris en el panorama documental va provocar una escissió entre els professionals: d'una banda, els bibliotecaris dipositaris d'una tradició d'ús dels sistemes de classificació; de l'altra, un sector de professionals de la documentació defensors dels nous sistemes d'anàlisi i accés a la informació continguda en els documents.

Van fer falta alguns anys per a conciliar les dues posicions i entendre que cada sistema s'adaptava a unes necessitats d'informació diferents.

- A les biblioteques enciclopèdiques, per a la representació sintètica del tema del document dins els catàlegs de matèries, i a vegades per a la seva ordenació física, eren necessaris els sistemes de classificació. També a les biblioteques especialitzades eren una eina útil per a elaborar productes documentals (butlletins bibliogràfics, de sumaris de revistes, etc.).
- Als serveis de documentació especialitzats en un camp del coneixement i especialment gràcies a la presència de les bases de dades eren més útils els llenguatges analítics que permetien de representar el contingut del document mitjançant descriptors. Les possibilitats que donava l'automatització permetien la multiplicació dels punts d'accés al document i una recuperació selectiva per mitjà de la combinació d'aquests.

A mitjan anys setanta el creixement de la indústria de les bases de dades va possibilitar la consulta en línia i va propiciar un nou canvi en el panorama de la documentació: la irrupció del llenguatge natural en la recuperació de la informació. Per què s'han de destinar grans recursos a la construcció de tesaurus si n'hi ha prou d'emmagatzemar a l'ordinador les paraules significatives del títol, resum o text dels documents enregistrats en el sistema? Així, els llistats de paraules clau es va afegir a l'oferta de llenguatges documentals disponibles: sistemes de classificació i vocabularis de descriptors controlats, tipus tesaurus.

La recuperació per mitjà d'aquestes paraules clau contingudes en el text dels documents aviat es va revelar com a insuficient i poc operativa a causa dels problemes d'ambigüitat que presenta el llenguatge natural en la recuperació.

La necessitat de complementar tots dos sistemes -llenguatges controlats (tipus tesaurus) i llenguatges lliures (llistats de paraules clau)- ha estat la tendència durant els anys vuitanta i noranta. D'una banda, els llenguatges controlats, gràcies a la seva manca d'ambigüitat, permeten gestionar cerques documentals amb gran precisió. D'altra banda, la recuperació per mitjà d'un llenguatge lliure extret del text del document permet d'escapar de les restriccions d'un llenguatge controlat i garantir una major exhaustivitat en la cerca.

El desenvolupament de la indexació automàtica amb l'aparició de sistemes experts basats en la intel·ligència artificial i sistemes assistits que extreuen els termes d'indexació a partir d'un tesaurus establert prèviament, fan suposar que el futur de la recuperació de la informació resideix en l'aplicació dels sistemes experts als sistemes documentals.

"Al igual que sucedió con los anteriores cambios de escena en el mundo de los lenguajes documentales, parece que los sistemas expertos aportarán una evolución y no una mutación: los thesaurus constituirán uno de los pilares del sistema de almacenamiento y recuperación documentales: la base de conocimientos, que contendrá la lista de los conceptos evocados en los documentos y en las consultas, bajo una forma normalizada; un segundo pilar, el motor de inferencia, explotará el thesaurus para pasar peticiones, expresadas en lenguaje libre, a una formulación de las ecuaciones de búsqueda en lenguaje controlado, y posteriormente a la extracción de los documentos pertinentes."

G. Van Slype, 1991, pàg. 18

La irrupció i el desenvolupament d'Internet també ha de repercutir en l'evolució dels sistemes de recuperació de la informació. L'explosió informativa present a la xarxa planteja cada vegada més la necessitat de sistematitzar els continguts per a accedir a la informació. Així, hi ha dos tipus de cercadors:

- els orientats a continguts com Yahoo (<http://www.yahoo.com>), que classifiquen els continguts en grans categories.
- els de cerca directa com Altavista (<http://altavista.magallanes.net>), que fan consultes a grans bases de dades després de subministrar-hi algunes paraules clau.

Gairebé tots els cercadors disposen de tècniques per a acotar al màxim les consultes: cerques avançades o expertes mitjançant operadors booleans, de proximitat, truncaments, etc. Alguns assenyalen la rellevància de les pàgines cercades, a partir de la informació subministrada. Això no obstant, tots hem patit l'allau d'informació davant una cerca d'informació a Internet.

No hi ha cap llenguatge documental pur.

Tots combinen diferents característiques i tipologies, i donen lloc a múltiples tipus, cadascun amb els seus avantatges i inconvenients.

La tendència més comuna en els sistemes documentals és l'ús de diferents llenguatges que es complementen en pro d'una millor representació i recuperació de la informació.

Sistemes de classificació, llenguatges controlats i llistats de paraules clau es fan servir dins un mateix sistema documental.

La breu història de la documentació reflecteix aquesta tendència i obre noves perspectives.

Sistemes experts aplicats als sistemes documentals han de permetre la indexació automàtica per mitjà de llenguatges controlats.

### **Vegeu també**

"Les tipologies dels llenguatges documentals", "Llenguatge codificat o natural?", "Llenguatges documentals i nivell d'indexació", "Llenguatges lliures o controlats?", "Precoor-

dinats o postcoordinats?" i "Estructures dels llenguatges documentals" d'aquest mateix mòdul.

### **Bibliografia**

**Maniez, J.** (1993). *Los lenguajes documentales y de clasificación. Concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Sánchez Ruipérez (Biblioteca del Libro, 56).

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



## 12. Llistats de paraules clau

Un llistat de paraules clau és un vocabulari, ordenat alfabèticament, de les paraules significatives, és a dir, de totes les paraules que no són buides (articles, preposicions, conjuncions, pronoms, adverbis, etc.) extretes automàticament amb un programari específic a partir del títol, del resum o bé del text complet dels documents.

### Exemple

#### Referència documental

Estivill, Assumpció; Urbano, Cristobal (1994). "Citacions i referències bibliogràfiques i no bibliogràfiques: una proposta per a la revista *Ítem*". *Ítem. Revista de Biblioteconomia i Documentació* (núm. 15, juliol-desembre, pàg. 4-59).

#### Resum

Proposta per a elaborar referències i citacions de tot tipus de fonts bibliogràfiques i no bibliogràfiques, publicades i inèdites, i de parts d'aquestes. Com a introducció, els autors justifiquen les raons de la proposta i expliquen els models en què es basa. Una vegada aprovades, aquestes directrius van adreçades principalment als autors i correctors de la revista *Ítem*, tanmateix, també s'espera que siguin emprades per a redactar referències i citacions en treballs de biblioteconomia i documentació.

#### Llistat de paraules clau

(Extret de la referència documental anterior)

adreçades	inèdites
aniran	ítem
aprovades	justifiquen
autors	models
basa	parts
bibliogràfiques	principalment
biblioteconomia	proposta
citacions	publicades
correctors	raons
directrius	redactar
documentació	referències
elaborar	revista
emprades	siguin
espera	treballs
expliquen	tipus
fonts	tot
introducció	vegada

Els llistats de paraules clau poden ser monolingües o multilingües depenent de les llengües dels documents que entren en un determinat sistema documental. Es pot donar el cas de tenir recollit dins el llistat un mateix concepte en diferents llengües (*dona/mujer*) sense que hi hagi cap relació d'equivalència entre els dos termes.

Generalment, les paraules clau són paraules simples (uniterms<sup>6</sup>) reconegudes automàticament pel programa informàtic que extreu totes les cadenes de caràcters separades de la resta per un o més espais en blanc o signes de puntuació, sempre que no estiguin incloses en el fitxer de paraules buides dissenyat per al sistema.

<sup>6</sup>Els *uniterms* són l'element significatiu més petit (paraules simples) seleccionat del document i útil per a la seva recuperació. No s'ha de confondre amb les *paraules clau* o *descriptors* que poden ser també paraules compostes.

Poden ser paraules clau:

- Paraules de totes les categories gramaticals (noms comuns, noms propis, adjectius, verbs...)
- Paraules de totes les formes gramaticals (masculí i femení, singular i plural, conjugacions de verbs...)
- Totes les variants ortogràfiques (incloses les faltes d'ortografia o errades que es puguin donar en el text o resum del document)
- Paraules que designen un concepte precís (*radioafccionat*) i paraules amb una significació poc precisa (*ràdio* pot ser el mitjà de comunicació, l'aparell receptor, un element de geometria...)
- Paraules de totes les llengües dels documents registrats en el sistema documental
- Paraules que corresponen al tema dels documents dels quals procedeixen i d'altres que no hi tenen res a veure (comentaris secundaris dels autors, exemples plantejats...)

### 12.1. L'ús dels llistats de paraules clau en la indexació i la recuperació

El fet que caracteritza els llistats de paraules clau és la seva ambigüitat semàntica: cada concepte pot ser designat per diverses paraules (sinonímia) i cada paraula pot designar diferents conceptes (polisèmia i homonímia).

L'usuari, abans de formular una consulta, ha de reunir tots els sinònims possibles dels conceptes que l'interessen. En un llistat de paraules clau, sense cap relació d'equivalència entre sinònims, l'usuari els ha de localitzar llegint-se tot el llistat alfabètic de termes.

La dificultat d'enunciar les consultes de manera completa i el risc de no recuperar els documents pertinents és més gran en aquest tipus de vocabulari lliure per l'existència de:

- Equivalents lingüístics

- Errades ortogràfiques
- Paraules en singular i plural (una solució possible per a aquest problema és el truncament dels termes en el moment de la cerca: "home\$" recuperarà els documents amb el terme *home* en singular i *homes* en plural)
- Paraules amb una mateixa arrel ("inform\$" recuperarà paraules com *informàtica*, *informació*, *informal*, *informatiu*, *informe*, etc.)
- Sigles (ens podem trobar documents que facin referència a *ensenyament secundari obligatori* i d'altres que facin servir la sigla ESO per a designar el mateix concepte)

El fet que les llistes de paraules clau incloguin tots els termes extrets del títol, resum i, a vegades, text complet dels documents implica que siguin el llenguatge documental més ric en terminologia. Tanmateix, això representa un avantatge i un inconvenient a la vegada:

- L'avantatge és que com que recull tots els termes, pot permetre de portar a terme cerques molt precises (el nom d'un lloc, d'una marca, d'un producte, etc.)
- L'inconvenient és que molt freqüentment quan plantegem una cerca sobre una paraula:
- recuperem documents que contenen la paraula però amb un significat diferent (polisèmia) o bé amb un caràcter secundari, sense cap relació amb el tema que ens interessa.
- no trobem els documents en els quals el mateix concepte s'expressa amb una altra paraula.

El llistat de paraules clau és el llenguatge documental que s'actualitza més ràpidament: evoluciona automàticament al mateix temps que la terminologia emprada en els documents.

El principal avantatge que presenta aquest llenguatge, respecte a la resta de llenguatges documentals, és la facilitat d'ús en el moment d'indexar els documents: no necessita que els analistes cerquin els conceptes dels documents, ja que la indexació es fa automàticament.

Això permet:

- Un estalvi en costos de construcció i manteniment del llenguatge (recursos humans).
- Menys demora en el temps transcorregut des de l'arribada dels documents al servei de documentació fins a la seva integració a la base de dades documental.
- Una coherència de la indexació perfecta: un mateix document tractat per un mateix programari informàtic registra sempre els mateixos termes d'indexació. Això no obstant, dos documents que tractin sobre un mateix tema, però en els quals els seus autors hagin emprat termes diferents seran indexats de manera totalment heterogènia.

El principal inconvenient dels llistats de paraules clau es presenta en el moment de la consulta, ja que la manca de rigor d'aquest llenguatge fa que la cerca documental sigui molt complexa:

- Per la necessitat d'agrupar tots els sinònims dels conceptes de la consulta, sense comptar amb cap ajuda (manca de relacions d'equivalència entre els termes).
- Per la necessitat de fer ús del truncament per a no repetir totes les formes gramaticals d'una paraula, o bé dels operadors de proximitat per a evitar la coordinació entre paraules situades en contextos diferents (si les paraules *societat* i *informació* són a la mateixa frase, és molt possible que el document tracti sobre la societat de la informació; si són en paràgrafs diferents, aquesta probabilitat disminueix).
- Per la necessitat d'eliminar una sèrie de documents no pertinents que apareixen a causa de la polisèmia i les falses coordinacions dels termes.
- Per la impossibilitat de recuperar els documents que tracten sobre els conceptes de la consulta si aquests han estat expressats pels autors de maneres diferents.

### Els llistats de paraules clau

...són vocabularis alfabètics dels termes significatius extrets del títol, resum o text complet dels documents de manera automàtica.

...reporten molts avantatges en el moment de la indexació: costos de creació i manteniment molt baixos, coherència d'indexació perfecta i actualització immediata.

...comporten molts inconvenients en el moment de la recuperació: soroll i silenci documental produït per l'ambigüitat semàntica d'aquest llenguatge (sinonímia, polisèmia, homonímia), complexitat en el moment de plantejar l'estratègia de cerca (agrupació de sinònims, ús de truncaments i operadors de proximitat), manca d'una estructura semàntica que ajudi l'usuari a localitzar els termes.

### Vegeu també

El llenguatge natural en la comunicació documental" i "Llenguatges lliures o controlats?" d'aquest mateix mòdul.

"La indexació o descripció característica" i "La indexació automàtica" del mòdul "Anàlisi i recuperació del contingut".

### Bibliografia

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

## 13. Llistats de descriptors lliures

Un llistat de descriptors lliures és un vocabulari, ordenat alfabèticament, de conceptes destacats, per un procés intel·lectual, a partir dels documents registrats dins d'un sistema documental determinat: els conceptes són expressats per paraules o expressions extretes dels documents o proposades pels analistes, sense verificar si existeixen prèviament en un llistat establert a priori. Normalment es tracta d'un vocabulari monolingüe.

### Exemple

#### Referència documental

Estivill, Assumpció; Urbano, Cristobal (1994). "Citacions i referències bibliogràfiques i no bibliogràfiques: una proposta per a la revista *Ítem*". *Ítem. Revista de Biblioteconomia i Documentació* (núm. 15, juliol-desembre, pàg. 4-59).

#### Resum

Proposta per a elaborar referències i citacions de tot tipus de fonts bibliogràfiques i no bibliogràfiques, publicades i inèdites, i de parts d'aquestes. Com a introducció, els autors justifiquen les raons de la proposta i expliquen els models en què es basa. Una vegada aprovades, aquestes directrius van adreçades principalment als autors i correctors de la revista *Ítem*; tanmateix, també s'espera que siguin emprades per a redactar referències i citacions en treballs de biblioteconomia i documentació.

#### Llistat de descriptors lliures

Proposats per l'analista després d'examinar el document.

biblioteconomia

citacions bibliogràfiques

directrius

documentació

fonts bibliogràfiques

fonts inèdites

fonts publicades

Ítem

referències bibliogràfiques

revistes

treballs

Es tracta d'una primera forma de control del vocabulari: els documents són examinats pels analistes que indexen els conceptes principals per mitjà de paraules o expressions, generalment redactades en una única llengua.

Encara que és fàcil la selecció formal dels termes (retenir únicament els noms, deixant de banda adjectius i verbs, i expressar-los només en singular) és molt difícil d'aconseguir una homogeneïtat absoluta, per part dels analistes, en

l'enunciat dels termes emprats per a designar un mateix concepte. Com a conseqüència, el vocabulari registra la presència de molts sinònims, varietats formals en les expressions (ONU i O.N.U.), errades d'ortografia, etc.

En aquests llenguatges, el fenomen de la sinonímia és menys important que en els llistats de paraules clau, gràcies a l'eliminació de les variants formals dels termes (plurals, adjectius, verbs) i de l'ús d'una única llengua per a expressar-los. De totes maneres, la sinonímia natural del llenguatge subsisteix.

La polisèmia no se suprimeix perquè queda en el llenguatge natural, però es redueix en relació amb el llistat de paraules clau gràcies a l'ús d'expressions més significatives.

G. Van Slype posa el següent exemple: el terme *servei militar* és unívoc, mentre que les paraules *servei* i *exèrcit* es poden referir al servei militar, però també a un servei organitzat de l'Exèrcit, a un servei realitzat per a l'Exèrcit, etc.

Juntament amb els llistats de paraules clau, els llistats de descriptors lliures són un dels llenguatges més ràpids i fàcils d'actualitzar. Els analistes tenen la possibilitat d'anar afegint nous termes al llistat a mesura que se'ls van trobant en els documents.

Tenen uns costos de construcció gairebé nuls, només els que es deriven de l'edició *a posteriori* del llistat de descriptors emprats pels analistes.

### **13.1. L'ús dels llistats de descriptors lliures en la indexació i la recuperació**

En el moment de la indexació els llistats de descriptors lliures, com els llistats de paraules clau, presenta l'avantatge que no necessita una traducció dels conceptes trobats en els documents, en llenguatge natural, a un altre llenguatge artificial controlat.

En el moment de la consulta, la manca de rigor del llenguatge lliure, com en el cas dels llistats de paraules clau, fa que sigui més complex de plantejar l'estratègia de cerca:

- Per la necessitat d'agrupar tots els sinònims dels conceptes de la cerca.
- Per la manca d'una estructura semàntica que ajudi l'usuari en la localització dels termes de la cerca.

- Per la impossibilitat de recuperar els documents que tracten sobre els conceptes de la consulta si aquests han estat expressats pels seus autors de maneres diferents.
- Per la impossibilitat de recuperar els documents que tracten sobre els conceptes de la consulta si aquests han estat indexats pels analistes amb formes diferents.

Els llistats de descriptors lliures són els llenguatges que presenten la pitjor taxa de coherència en la indexació:

"La coherencia de la indización es la peor (de un 20 a un 40%) en los sistemas de lista de descriptores libres, ya que al no tener ninguna posibilidad de control automático del vocabulario utilizado, los indizadores son libres de usar los términos que a cada uno de ellos les parezcan más apropiados."

G. Van Slype, 1991, pàg. 33

### Els llistats de descriptors lliures

...són un vocabulari alfabètic de termes seleccionats, per un procés intel·lectual, a partir de l'examen del document.

...recullen termes d'indexació extrets del document o proposats pels analistes sense verificar si existeixen en un llistat establert.

...presenten avantatges en el moment de la indexació (els termes no s'han de traduir a un llenguatge controlat ni s'han de normalitzar), però inconvenients en la recuperació (complexitat de les estratègies de cerca, manca d'una estructura que ajudi a la localització dels termes, soroll documental)

### Vegeu també

"Llenguatges lliures o controlats?" d'aquest mateix mòdul.

### Bibliografia

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



## 14. Llistats d'autoritats o descriptors controlats

Un dels llenguatges documentals controlats més bàsics, per la simplicitat de la seva estructura, és el llistat d'autoritats o llistat de descriptors controlats.

És un vocabulari ordenat alfabèticament de conceptes destinats a representar de manera unívoca el contingut d'un document i de les consultes dins d'un sistema documental; aquests conceptes s'expressen per paraules o expressions extretes d'un llistat finit establert *a priori*; només els termes que figuren en el llistat es poden fer servir per a indexar els documents i les consultes.

Normalment es tracta d'un llistat monolingüe que recull únicament una forma gramatical del terme (singular).

Com el llistat de descriptors lliures, reuneix conceptes expressats per paraules i expressions del llenguatge natural. Però a diferència d'aquell, el llistat d'autoritats és un vocabulari finit i tancat, i només els termes que hi ha en el llistat es poden fer servir per a indexar i recuperar els documents.

Es construeixen *a priori*, és a dir, abans d'iniciar l'anàlisi del fons documental. Els responsables del sistema estableixen els conceptes de la disciplina o camp temàtic del fons que poden tenir interès per a l'usuari del sistema. Cada concepte ha d'estar representat per un únic terme, i per aquesta raó:

- se selecciona d'entre tots els possibles sinònims el que s'emprarà com a descriptor.
- es verifica l'ortografia dels termes que figuren en els documents i en les consultes per a evitar la pèrdua d'informació.

### Exemple

Volem ordenar i sistematitzar la nostra discoteca i petita biblioteca de música amb la construcció d'una base de dades que ens permeti accedir a la informació dels documents mitjançant un llistat de descriptors controlats. Abans d'iniciar l'anàlisi dels documents i com a fruit d'un mostreig dels documents de la discoteca, establim el llistat de descriptors que emprarem:

antologia musical	instrument musical
<i>blues</i>	<i>jazz</i>
cançó	lírica
cançó popular	músic
cançoner	música folklòrica
discografia	música religiosa
ensenyament de la música	música teatral
flamenc	música vocal
història de la música	obra musical
himnes	pastoral
influència musical	tema musical

Malgrat la voluntat d'univocitat que tenen els llistats d'autoritats, a la pràctica la manca d'una estructura semàntica i, especialment, d'equivalències entre descriptors i no descriptors (sinònims o termes amb significats afins) fa difícil l'eliminació de la sinonímia i la polisèmia quan es construeix el llistat (excepte en el cas que es tracti d'un llistat restringit).

#### 14.1. L'ús dels llistats d'autoritats en la indexació i la recuperació

Els llistats d'autoritats, com en tots els llenguatges controlats, en el moment de la **indexació** segueixen el procés següent:

- Examen dels documents (títol, resum, sumari, contingut, etc.).
- Identificació i selecció dels conceptes en la seva expressió en llenguatge natural.
- Traducció d'aquests conceptes per a obtenir-ne una representació en el llistat de descriptors.

A diferència dels llenguatges lliures, amb els quals no és necessari de traduir els termes d'indexació en els llenguatges controlats com el llistat d'autoritats, es fa necessària aquesta última fase de traducció al llenguatge documental per tal de garantir la univocitat.

Aquest fet representa un major grau d'intervenció en el procés de les operacions intel·lectuals necessàries, per part dels analistes, que pot afectar la coherència de la indexació. La manca d'homogeneïtat en els indexadors es pot manifestar en les diferents formes o punts de vista en el moment de percebre el contingut d'un document:

- el contingut real del document
- la part del contingut que serà susceptible de respondre realment a les necessitats dels usuaris
- els conceptes importants que s'han de conservar per a representar el contingut

- els descriptors seleccionats per a representar aquests conceptes

És un fet constatat que davant d'un mateix document la coincidència de descriptors seleccionats per diferents analistes emprant un llistat d'autoritats -segons G. Van Slype- és inferior al 50%: menys de la meitat dels descriptors seleccionats coincideixen.

La subjectivitat dels analistes depèn del grau de formació i coneixement que tingui del llenguatge documental del sistema. També es pot reduir si el llenguatge conté una estructura semàntica que permeti a l'usuari de localitzar els descriptors sinònims o de significat afí -com els tesaurus-, però els llistats d'autoritats generalment no tenen cap estructura que ajudi l'usuari a localitzar els sinònims dels seus termes.

En el procés de **consulta i recuperació**, el llistat d'autoritats, i en general tots els llenguatges controlats, presenta més avantatges i permet d'eliminar la major part dels inconvenients que presenten els llenguatges lliures:

- No és necessari agrupar sinònims del concepte de cerca.
- L'ús generalitzat del singular i el control que s'exerceix sobre els termes no fa necessaris els truncaments ni els operadors de proximitat.

No obstant això, com en els llenguatges lliures, es poden donar casos de falses coordinacions.

#### **14.2. Els llistats d'autoritats i els identificadors**

L'ús més habitual en els sistemes documentals dels llistats d'autoritats és el control dels identificadors: noms de persones, organismes, entitats, topònims, etc. que apareixen en el procés d'indexació entre els conceptes del document, però que no estan recollits pel llenguatge documental del sistema.

Hi ha pocs sistemes documentals que escapin de la funció estricta dels llenguatges documentals (que es refereixen essencialment al contingut conceptual dels documents i de les consultes). Per tant, hi ha pocs llenguatges documentals controlats, de tipus tesaurus, que incloguin dins dels seus vocabularis els identificadors. Per a assolir la tasca de controlar i normalitzar l'ús dels identificadors en els processos d'indexació i recuperació es fan servir els llistats d'autoritats.

Servei de Documentació d'Història Local de Catalunya. (1999). Llistat d'autoritats. Identificadors personals (pàg. 1).

Abelló i Prat, Joan - 1922- ....  
Aguado y Requejo, Felipe de  
Aguiló i Fuster, Marian - 1825-1897  
Agustí Peypoch, Ignasi - 1913-1974  
Alavedra i Segurañas, Joan - 1896-1981  
Albarranch Blasco, Vicent - 1898-1940

Albó i Corrons, Núria  
Albó i Martí, Ramon - 1871-1955  
Albó, Carlos  
Alcover i Maspons, Joan  
Alejandro, Gerard  
Alemany, Joan  
Alfredi  
Algueró, Josep - 1918-....  
Almató i Ribera, Pere - 1830-1861  
Amadó i Gual, Joan - 1920-...  
Amat de Palou, familia  
Ambrós i Martí, Lluís - 1872-1933  
Aparisi, Joan- 1653-1731  
Aragonés, Francisco  
Arbizu, Andreu  
Archer, Andrés - 1811-1855  
Ardévol Arbués, Àlvar  
Argemir i Mitjà, Miquel - 1591-1625  
Arimany i Nicolau, Josep Lluís - 1923 .....  
Arimon, Josep - 1890-1958  
Armanyà i Font, Francesc - 1718-1803  
Armengol i Castelitèrgol, Agnès  
Arrizabalaga, Josep Maria  
Arumí i Blancafort, Lluís - 1863-1943  
Aumedes Farré, Ramon  
Ayats, Francesc  
Bach de Roda  
Bach i Molas, Josep - 1915 .....  
Bach i Teixidó, Càndid - 1902 .....  
Bach, Johann Sebastian  
Bada i Cazeilles, Berta  
Badia Flaquer, Maria - 1903-1936  
Badia i Durán, Manel  
Badia i Moret, Josep  
Badia i Olivé, Josep Lluís  
Balaguer i Cirera, Víctor - 1824-1901  
Balcelis, Ricard  
Balletbó i Duran, Salvador - 1895-1985  
Balmes i Urpià, Jaume Llucià - 1810-1848  
Baltà i Llopart, Pere  
Balvey i Bas, Tomàs - 1865-1954  
Bannalocha, Berenguer de  
Bannalocha, familia  
Banyuls, Tomás de  
Barbosa, Lluís  
Bardés i Huguet, Josep Maria  
Barnola, Pere  
Barres i Pascual, Maties - 1905-1992  
Barres, Josep Maria - 1861-1923  
Barri i Draper, Pere Nolasc - 1900-19??  
Bassa, Sebastià- 1859-1935  
Batile, Antoni  
Batile, Enric  
Batiles, Joan  
Batiles, Manel  
Beethoven, Ludwig Van - 1770-1827  
Belloch, familia  
Ben Arab, Ridha  
Benet i Morell, Josep - 1920-...  
Berenguer, Guillem de  
Bertran Vila, Antoni - 1870-19??  
Biges, Huguet de  
Bigues, familia  
Bigues, Huguet de  
Bizcarri, Joan Josep  
Blancafort de Rosselló, Manuel -1897-1987  
Blancafort i París, Gabriel - 1929.....  
Blanch, Francesc  
Bofill i Ferro, Jaume  
Bofill i Mates, Jaume - 1878-1933  
Bofill i Portabella, Ramon  
Bofill, Maña Àngela

Bohigas i Guardiola, Oriol  
 Boix, Josep M. - 1887-1918  
 Bonet i Marrugat, Albert - 1894-1974  
 Bonet i Zanuy, Constantí - 1808-1878  
 Bonet, Pep  
 Bordas, Mariano  
 Borrás Brustenga, Josep  
 Bosch de Calderó i Damm, familia  
 Bosch i Cardellach, Antoni - 1758-1829  
 Bosch i Jover, Miquel - 1900-1960  
 Bosch i Pareras, Anna  
 Bosch Planas, A.  
 Botey i Gómez, Josep M.  
 Botey, Antoni  
 Bru i Salilles, Lluís - 1868-1952  
 Brugarolas i Planas, Hilad - 1901 -1996  
 Bruil Orriols, Joaquim - 1914-...  
 Bruy, Francesc - 169?-1713?  
 Busquets i Moliera, Jaume - 1904-1968  
 Butinyà i Hospital, familia  
 Butinyà i Hospital, Francesc Xavier -1834-1899  
 Cabanas i Alibau, Miquel  
 Cabanyes, familia  
 Cabrera i Grinyó, Ramon - 1806-1877  
 Cabrera, Bernat III de - 1350-1368  
 Cadell, Joan  
 Cadevall i Diars, Joan - 1846-1921  
 Calsada, familia  
 Cambó i Batlle, Francesc - 1876-1947  
 Camp Puigdomenech, Joan  
 Camp Puigdomenech, Jordi  
 Camp Puigdomenech, Josep  
 Camp, Maria  
 Companyà, familia  
 Camprodon, Santiago  
 Camprubí i Casas, Jordi - 1948-...  
 Camps i Giró, Joan - 1948-1976  
 Camps i Segalés, Joan - 1845-1???  
 Campsentelles, familia

### Els llistats d'autoritats

...són un vocabulari alfabètic de termes controlats destinats a representar de manera unívoca el contingut d'un document i de les consultes dins d'un sistema documental. Només els termes que figuren en el llistat es poden fer servir per a indexar i recuperar la informació.

...presenten inconvenients en el moment de la indexació (manca d'homogeneïtat per part dels analistes), però avantatges en la recuperació i consulta (estratègies de cerca més simples).

...normalment s'empren per a controlar els identificadors (noms de persones, entitats, marques, topònims, etc.), que són part del contingut del document, però que no solen estar recollits pels llenguatges documentals.

### Vegeu també

"Llenguatges lliures o controlats?" d'aquest mateix mòdul.  
 "El procés d'indexació" del mòdul "Anàlisi i recuperació del contingut".

### Bibliografia

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

## Autoavaluació

Dels enunciats següents, quins són vertaders i quins falsos?

1. La comunicació documental es pot establir mitjançant llenguatges artificials o controlats o bé per mitjà del llenguatge natural contingut en el text del document.
2. L'ús del llenguatge natural en els processos d'indexació i recuperació permet una bona comunicació documental.
3. El llenguatge natural es caracteritza per la seva sinonímia i polisèmia.
4. Els llenguatges documentals artificials o controlats tenen com a principal objectiu garantir la univocitat dels termes d'indexació.
5. Els llenguatges codificats representen la matèria del document per mitjà de múltiples notacions o símbols de classe.
6. Els llenguatges codificats faciliten la classificació i ordenació física dels documents.
7. Si s'ha de sintetitzar i extreure la matèria principal del document, s'ha d'emprar un llenguatge documental amb descriptors.
8. Si interessa recuperar de manera precisa els conceptes continguts en els documents, s'ha d'emprar un sistema de classificació o bé un llistat d'encapçalaments de matèries.
9. Els llenguatges lliures extreuen els termes d'indexació del llenguatge natural dels documents (títol, resum o text complet).
10. Els llenguatges lliures presenten avantatges en la recuperació però inconvenients en la indexació.
11. Els llenguatges controlats tenen com a objectiu la univocitat dels termes d'indexació evitant la sinonímia i polisèmia del llenguatge natural.
12. En un llenguatge controlat només els termes d'indexació recollits en el vocabulari es poden fer servir per a la indexació i la recuperació.
13. Els llenguatges postcoordinats coordinen i sintetitzen els diferents conceptes que conformen una matèria en un únic terme d'indexació.
14. Els llenguatges postcoordinats permeten l'ordenació física dels documents i de les referències.
15. Els llenguatges precoordinats coordinen els diferents termes d'indexació en el moment de la consulta o recuperació de la informació.
16. Els llenguatges combinatoris o associatius presenten una estructura alfabètica de termes d'indexació que es combinen en el moment de la recuperació.
17. Els llenguatges jeràrquics presenten una estructura en la qual els termes estan ordenats seguint el criteri següent: els conceptes més específics dins dels més generals.
18. La construcció d'un llenguatge lliure significa una gran inversió en temps i personal.
19. No hi ha cap llenguatge documental pur.
20. Els llistats de paraules clau són vocabularis alfabètics de termes significatius extrets automàticament del títol, del resum o del text complet dels documents.
21. Els llistats de paraules clau garanteixen la univocitat dels termes d'indexació.
22. Els llistats de descriptors lliures recullen termes d'indexació assignats lliurement per l'analista després l'examen del document.
23. Els llistats de descriptors lliures presenten inconvenients en el moment de la indexació però avantatges en la recuperació de la informació.
24. Els llistats d'autoritats són vocabularis alfabètics de termes controlats destinats a representar de manera unívoca el contingut dels documents i de les consultes dins d'un sistema documental.
25. Els llistats d'autoritats normalment es fan servir per a controlar els identificadors.

## Solucionari

1. La comunicació documental es pot establir mitjançant llenguatges artificials o controlats o bé per mitjà del llenguatge natural contingut en el text del document. **(Vertader)**
2. L'ús del llenguatge natural en els processos d'indexació i recuperació permet una bona comunicació documental. **(Fals)**
3. El llenguatge natural es caracteritza per la seva sinonímia i polisèmia. **(Vertader)**
4. Els llenguatges documentals artificials o controlats tenen com a principal objectiu garantir la univocitat dels termes d'indexació. **(Vertader)**
5. Els llenguatges codificats representen la matèria del document per mitjà de múltiples notacions o símbols de classe. **(Fals)**
6. Els llenguatges codificats faciliten la classificació i ordenació física dels documents. **(Vertader)**
7. Si s'ha de sintetitzar i extreure la matèria principal del document, s'ha d'emprar un llenguatge documental amb descriptors. **(Fals)**
8. Si interessa recuperar de manera precisa els conceptes continguts en els documents, s'ha d'emprar un sistema de classificació o bé un llistat d'encapçalaments de matèries. **(Fals)**
9. Els llenguatges lliures extreuen els termes d'indexació del llenguatge natural dels documents (títol, resum o text complet). **(Vertader)**
10. Els llenguatges lliures presenten avantatges en la recuperació però inconvenients en la indexació. **(Fals)**
11. Els llenguatges controlats tenen com a objectiu la univocitat dels termes d'indexació evitant la sinonímia i polisèmia del llenguatge natural. **(Vertader)**
12. En un llenguatge controlat només els termes d'indexació recollits en el vocabulari es poden fer servir per a la indexació i la recuperació. **(Vertader)**
13. Els llenguatges postcoordinats coordinen i sintetitzen els diferents conceptes que conformen una matèria en un únic terme d'indexació. **(Fals)**
14. Els llenguatges postcoordinats permeten l'ordenació física dels documents i de les referències. **(Fals)**
15. Els llenguatges precoordinats coordinen els diferents termes d'indexació en el moment de la consulta o recuperació de la informació. **(Fals)**
16. Els llenguatges combinatoris o associatius presenten una estructura alfabètica de termes d'indexació que es combinen en el moment de la recuperació. **(Vertader)**
17. Els llenguatges jeràrquics presenten una estructura en la qual els termes estan ordenats seguint el criteri següent: els conceptes més específics dins dels més generals. **(Vertader)**
18. La construcció d'un llenguatge lliure significa una gran inversió en temps i personal. **(Fals)**
19. No hi ha cap llenguatge documental pur. **(Vertader)**
20. Els llistats de paraules clau són vocabularis alfabètics de termes significatius extrets automàticament del títol, del resum o del text complet dels documents. **(Vertader)**
21. Els llistats de paraules clau garanteixen la univocitat dels termes d'indexació. **(Fals)**
22. Els llistats de descriptors lliures recullen termes d'indexació assignats lliurement per l'analista després l'examen del document. **(Vertader)**
23. Els llistats de descriptors lliures presenten inconvenients en el moment de la indexació però avantatges en la recuperació de la informació. **(Fals)**
24. Els llistats d'autoritats són vocabularis alfabètics de termes controlats destinats a representar de manera unívoca el contingut dels documents i de les consultes dins d'un sistema documental. **(Vertader)**
25. Els llistats d'autoritats normalment es fan servir per a controlar els identificadors. **(Vertader)**

## Bibliografía

**Aichison, Jean; Allen, C.G.** (1983). *Bibliografía de vocabularios, tesauros, encabezamientos de materia y esquemas de clasificación de ciencias sociales: mono y plurilingües*. París: UNESCO (Informes y documentos de Ciencias sociales, núm. 54).

**García Gutiérrez, A.** (1989). "Los lenguajes documentales". En: J. López Yepes (comp.). *Fundamentos de información y documentación* (págs. 313-342). Madrid: Eudema (Manuales).

**Gil Urdiciain, B.** (1996). *Manual de lenguajes documentales*. Madrid: Noesis.

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

**Maniez, J.** (1993). *Los lenguajes documentales y de clasificación. Concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Sánchez Ruipérez. (Biblioteca del Libro, núm. 56).

**Perpinyà Morera, R.** (1995). "Los lenguajes documentales". En: E. Fuentes (ed.). *Manual de documentación periodística* (págs. 112-132). Madrid: Síntesis (Biblioteconomía y Documentación, núm. 6).

**Pinto Molina, M.** (1989). "El análisis de contenido: la indización de documentos". En: J. López Yepes (comp.). *Fundamentos de información y documentación* (págs. 347-374). Madrid: Eudema (Manuales).

**Pinto Molina, M.** (1991). *Análisis documental: fundamentos y procedimientos*. Madrid: Eudema.

**Pinto Molina, R.; Gálvez Martínez, C.** (1996). *Análisis documental de contenidos: procesamiento de información*. Madrid: Síntesis (Biblioteconomía y Documentación, núm. 10).

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

### Bibliografía complementaria

**Amat, N.** (1987). *Documentación científica y nuevas tecnologías de la información* (cap. 4 y 5). Madrid: Pirámide.

**Asociación Española de Normalización y Certificación (AENOR)** (1997). "UNE-50-113/6 (ISO 5127/6). Documentación e información. Vocabulario. Parte 6: lenguajes documentales". *Revista Española de Documentación Científica* (vol. 20, núm. 4, págs. 417-436).

**Asociación Española de Normalización y Certificación (AENOR)** (1991). UNE-50-121/91 (ISO 5963-1985). *Documentación: métodos para el análisis de documentos, determinación de su contenido y selección de los términos de indización*.

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

**García Gutiérrez, A.** (1984). *Lingüística documental: aplicación a la documentación de la comunicación social*. Barcelona: Mitre.

**Krippendorff, K.** (1990). *Metodología de análisis de contenido: teoría y práctica*. Barcelona: Paidós.

**Pinto Molina, M.** (1997). *Manual de clasificación documental*. Madrid: Síntesis (Biblioteconomía y Documentación, núm. 15).

**Rubió, J.** (1991) "L' estructura jeràrquica de CDU i els seus recursos de facetització i enumeració de conceptes". *Recull de treballs sobre classificació i indexació* (págs. 15-29). Barcelona: Diputació de Barcelona.