

Introducció

Funcions d'un thesaurus:

Un thesaurus és
Un thesaurus no és
Per qui és un thesaurus?

El thesaurus SINERA:

Origen
Abast
Seccions
Relacions
Estructura d'un enregistrament

Utilització del thesaurus:

Presentacions
Utilització per si mateix
Connexió amb la base SINERA
Mètodes de cerca

Un thesaurus és

un vocabulari d'una àrea de coneixement o d'una àrea d'activitat, que es presenta estructurat semànticament i que és emprat amb caràcter normatiu per un servei d'informació documental en la indexació de determinat tipus de materials.

estructurat semànticament: els termes són agrupats en funció del seu significat, i no alfabèticament ni seguint una filiació conceptual.

indexació: atribució a cada document d'un conjunt de termes que en descriuen el contingut informatiu

Un thesaurus no és

- un inventari exhaustiu de les expressions i els termes emprats en el seu camp, sinó una selecció dels termes més rellevants i susceptibles de ser distingits en les descripcions de contingut dels documents que es volen indexar.

- un diccionari, que acompanya cada mot amb la seva definició, les peculiaritats gramaticals i els exemples d'ús, sinó un vocabulari que només conté termes i pressuposa conegut el seu significat.

- una llista alfabètica amb equivalències lingüístiques, sinó que els termes que hi conté vénen agrupats d'acord amb la seva proximitat conceptual i és aquesta estructura de sentits la que li dóna valor.

- una normativa de tipus terminològic que fixa la forma correcta de les paraules que hi figuren, sinó que reflecteix l'ús comú entre els qui les empren, i hi afegeix i declara com a tals les variants significatives que són d'interès.

Per qui és un thesaurus?

Un thesaurus és un instrument de treball per a un sistema o un servei d'informació i documentació, i dintre d'ell per a les dues grans categories de persones que hi participen:

- Els indexadors o responsables d'atribuir i consignar en cada document els descriptors que n'expressen el contingut informatiu. Per a ells el thesaurus és la norma que dicta els termes que poden emprar per cada document.
- Els usuaris del sistema, que l'exploren a la recerca d'informació, i per als quals el thesaurus és la guia que els permet conèixer els criteris amb què els indexadors han descrit els materials, tot oferint-los camins per recórrer-lo.

Origen del thesaurus SINERA

L'elaboració d'aquest thesaurus va sorgir de la necessitat dels serveis de documentació del Programa d'Informàtica Educativa de disposar d'una eina terminològica per a la compilació i la sistematització de la base de dades SINERA.

La manca d'un thesaurus general en llengua catalana comportà el desenvolupament d'un thesaurus propi de SINERA, inicialment basat en fonts d'altres procedències, principalment el Thesaurus General de la UNESCO (Lausanne, 1984), i d'altres més específics com ara els thesaurus EUDISED, ERIC i THELYCE.

Les grans àrees temàtiques que abasta el thesaurus de SINERA són ciència i tecnologia, ciències socials, humanitats, informàtica i educació. L'àrea d'educació coincideix amb el [Thesaurus Català d'Educació](#), editat pel Departament d'Ensenyament l'any 1992, llevat de les modificacions imprescindibles per a imbricar-lo de forma natural en un context més ampli.

L'elaboració del thesaurus de SINERA ha estat un procés dinàmic i viu, que s'inicià amb la selecció, fusió i traducció de diverses fonts heterogènies i s'enriquí amb un treball terminològic i educatiu en el qual participaren molts membres i col.laboradors del Programa d'Informàtica Educativa, donant lloc a una primera versió provisional i interna del thesaurus impresa l'any 1990 per a ús dels redactors de la base SINERA. A partir d'aquesta versió es realitzà una profunda revisió global, acompanyada d'una incorporació de termes nous que s'havien revelat útils i significatius en el treball dels proveïdors d'informació per a la base SINERA.

El resultat és aquest thesaurus: una realitat dinàmica, no exempta de mancances i d'alguna incongruència, que millorarà i evolucionarà amb el temps i amb el propi creixement de la base SINERA, en un procés d'adaptació a les necessitats reals dels seus usuaris del món de l'educació.

Abast del thesaurus SINERA

El thesaurus SINERA conté **12.276** termes, dels quals **1.414** no són descriptors (sinònims no preferencials) i **10.752** són descriptors de ple dret.

S'hi troben termes i expressions d'ús comú en les diverses àrees del coneixement, procurant que cobreixin la terminologia utilitzada a l'ensenyament, des del nivell primari a un nivell postsecundari aproximadament coincident amb un primer curs universitari i incloent-hi, en alguns casos, termes de nivell més alt, però que sovint es troben en contextos divulgatius o que són necessaris per descriure treballs didàctics especialitzats.

El thesaurus no és, evidentment, exhaustiu. S'ha procurat delimitar els termes atenent als objectius d'indexació que ens proposàvem i, per tant, han estat seleccionats prioritàriament els termes susceptibles de figurar en la indexació (temàtica, tipològica o topogràfica) de recursos educatius de qualsevol tipus. Han quedat fora els termes molt particulars i que, des d'aquest punt de vista, no han semblat significatius.

Seccions del thesaurus SINERA

Els termes del thesaurus estan dividits en cinc seccions o grans àrees temàtiques, d'extensió molt desigual, que són:

Ciència i tecnologia: ciències ambientals, ciències de l'espai, ciències de la Terra, ciències de la vida, física, matemàtiques, química, tecnologia...

Ciències socials: administració, comunicació, dret, economia, política, psicologia, sociologia...

Educació

Humanitats: antropologia, arts, cultura, documentació, ètica, filologia, filosofia, història, lleure, religió...

Informàtica

ciències de la Terra:

Aquesta secció conté la terminologia específica de les ciències de la Terra, i una extensa enumeració jeràrquica d'entitats pròpies de la geografia política: els estats del món amb les seves capitals, les comunitats autònomes d'Espanya i, en el cas de Catalunya, les comarques i els municipis que les integren.

Els termes d'aquesta secció són, pràcticament, els únics noms propis que figuren al thesaurus.

La distribució territorial que reflecteixen és la vigent en el moment de la seva compilació.

Educació:

L'àrea d'educació del thesaurus SINERA coincideix amb el [Thesaurus Català d'Educació](#), editat pel Departament d'Ensenyament l'any 1992, llevat de les modificacions imprescindibles per a imbricar-lo de forma natural en un context més general. És el cas, per exemple, dels termes relatius a la salut o a les institucions socials.

Informàtica:

Encara que és indiscutible la pertinença de la informàtica a la tecnologia, se l'ha situat en una secció específica en base a l'origen i finalitat d'aquest thesaurus, nascut en el marc de les actuacions de documentació sobre aplicacions educatives de la informàtica realitzades pel Programa d'Informàtica Educativa. Aquesta secció és per tant especialment adequada a la sistematització de la terminologia usual en les publicacions especialitzades en aquest camp.

Relacions emprades al thesaurus SINERA

Sinonímia

Jerarquia

Què és una relació jeràrquica

Jerarquies acceptades

Jerarquies no acceptades

Termes polijeràrquics

Sinonímia

S'han declarat sinònims els termes que, donat l'àmbit d'aplicació del thesaurus, poden considerar-se com a designacions d'un mateix concepte. És a dir, dos termes es consideren sinònims quan es pot esperar raonablement que les persones interessades per la informació que aporta un d'ells també ho estan en la informació que aporta l'altre, i recíprocament.

Per exemple, *diputats* i *senadors* no són el mateix, però han estat declarats sinònims i són formes no preferents d'un tercer sinònim que és *parlamentaris*.

En alguns casos, els termes declarats com a sinònims ho són realment; en d'altres, són variants d'ús corrent d'un terme que algun usuari pot esperar de trobar o indiquen conceptes estretament lligats i distingibles només per consideracions més precises; finalment, hi ha formes fluctuants ortogràficament.

La relació de sinonímia ve indicada per les sigles EP .

El thesaurus conté **1.414** sinònims, però és clar que no hi són tots els possibles.

Què és una relació jeràrquica

El thesaurus presenta el parell de relacions jeràrquiques recíproques terme ampli i terme específic, per mitjà de la indentació dels termes: els termes són més específics quant més a la dreta es troben. En la literatura més habitual sobre thesaurus, aquestes relacions són d'abast molt restringit i només s'estableixen entre parells de termes on un designa una part o un cas particular de l'altre.

Aquesta interpretació "taxonòmica" és massa estricta fora d'algunes àrees de coneixement molt determinades. En aquest thesaurus s'ha optat per un criteri més ampli: un terme es considera jeràrquicament superior a un altre terme quan es pot suposar raonablement que la persona interessada per la informació que aporta el terme superior, també considerarà interessant la informació que aporta el terme inferior, però no recíprocament.

Dins de cada nivell jeràrquic els termes no es troben en cap ordre genètic o conceptual, sinó purament alfabètic.

Terme ampli (TA):

Referència hiperònims o conceptes jeràrquicament més generals. També es fa servir per a la relació partitiva d'un conjunt amb els seus components

Terme específic (TE):

Referencia hipònims o conceptes jeràrquicament més específics. També es fa servir per a la relació partitiva d'un component i el conjunt que l'inclou.

Jerarquies acceptades

- Les relacions jeràrquiques general/particular formalment correctes, com la que hi ha entre *transport* i *transport aeri*: el segon concepte és un cas particular del primer.
- Les relacions d'abast: la *sociologia* s'ocupa de la *família*, però això no vol dir que *família* sigui una part o una branca de la sociologia.
- Les relacions genètiques: l'estudi de la *música* porta a considerar el tema dels *instruments musicals*, però no són un cas particular ni un objecte d'estudi.

Jerarquies no acceptades

Els criteris per establir o rebutjar una relació jeràrquica determinada són molt difícils de formular i, en un thesaurus d'aquesta grandària, també són difícils d'aplicar de forma consistent. La realitat és molt rica i multiforme, i desborda qualsevol estructura de significats que se li vulgui superposar. Molt sovint ha estat necessari decidir l'adscripció d'un terme en una cadena jeràrquica concreta, tot i que moltes cadenes podien admetre'l raonablement.

Per tal de resoldre aquests conflictes, els thesaurus disposen de les relacions associatives que connecten "informalment" cada terme amb tots els termes que denoten conceptes afins. En la versió que esteu consultant i en el seu estat present, aquest thesaurus no implementa les relacions associatives.

No figuren, doncs, com a relacions jeràrquiques:

- Moltes relacions perfectament correctes, però que han estat sacrificades a una alternativa que s'ha cregut preferible: per exemple, *òpera* és un terme específic d'*espectacles* i no de *música vocal*.

- Relacions del tipus general/particular que resulten artificioses o multipliquen en excés els nivells jeràrquics: per exemple, *processador de textos* i *verificador ortogràfic* estan al mateix nivell, tot i que el segon opera habitualment dins l'àmbit del primer. Anàlogament, no s'ha tingut en compte que un *diccionari* és un cas particular de *llibre*, i els termes s'han ubicat en famílies semàntiques que no estan directament connectades.

- Relacions del tipus exemple, que s'estableixen entre un terme general i les seves realitzacions concretes: tot i que la *revolució francesa* sigui una *revolució*, no s'ha considerat com un dels seus termes específics: el primer terme pertany a la branca **Història** del thesaurus i el segon, a la branca **Política**.

- Relacions creuades entre diversos termes que indiquen aspectes diferents d'un tema, i que s'ha preferit deixar al mateix nivell. Per exemple, *transport internacional*, *transport marítim* i *transport de mercaderies* figuren al mateix nivell, tot i que cadascun d'aquests termes admet els altres dos com a particularitzacions.

Termes polijeràrquics

Un thesaurus ideal ha de contenir cada terme en una sola posició d'una sola cadena jeràrquica. Però això, és pràcticament impossible i ha estat necessari, i sovint difícil, prendre decisions d'aquest tipus.

En alguns casos, però, un terme ha de considerar-se formant part simultàniament de dues cadenes jeràrquiques. En aquest thesaurus hi ha 279 termes polijeràrquics, que normalment són terminals (sense termes específics) o tenen molt pocs termes connectats.

Els termes principals són:

- *desenvolupament infantil*, que depèn de *desenvolupament individual* i de *psicologia infantil*,
- *filosofia de la ciència*, que depèn de *ciència (conceptes generals)* i de *filosofia*,
- *justícia*, que depèn de *teoria jurídica* i de *virtuts*,
- *literatura folklòrica*, que depèn de *cultura popular* i de *formes i gèneres literaris*,
- *marxisme*, que depèn de *doctrines econòmiques* i de *filosofia política*,
- *sindicats*, que depèn de *relacions laborals* i d'*associacions*.

Presentacions del thesaurus SINERA

Actualment el thesaurus general de la base SINERA està disponible en tres presentacions diferents:

- **Telemàtica:** com a part de la consulta en línia a la base SINERA per mitjà de la xarxa XTEC. El thesaurus es presenta complet, amb les relacions associatives incloses, i està pensat perquè s'utilitzi com una eina auxiliar per a la consulta de la base, sense que sigui possible recórrer-lo lliurement. Està gestionat pels programes Mistral i Atena.
- **En suport magnètic:** desvinculat de la consulta de la base, constitueix una eina de treball per als redactors de SINERA. Resideix normalment al disc dur d'un microordinador i està gestionat pel programa Beat, que explota al màxim les seves possibilitats de navegació. Podeu consultar el thesaurus en format Beat des de la pantalla Miscel·lània del Programari o cercant la referència de Beat a la base SINERA.
- **En suport CD-ROM:** és la versió més assequible a l'usuari que estigui llegint aquesta informació. Es presenta sense relacions associatives i pot ser recorregut jeràrquicament. **SINERA en Disc** permet fer una cerca directa dels termes del thesaurus en conjunció amb la cerca a la base SINERA.

Xarxa Telemàtica Educativa de Catalunya

Utilització del thesaurus per si mateix

Tot i que un thesaurus assoleix plenament la seva funció com a auxiliar en la recerca de documents en una base de dades, hi ha dos motius principals que fan que la seva utilització autònoma sigui valuosa:

- Un thesaurus és un recull, altament estructurat, dels termes més importants d'una àrea de coneixement o d'activitat, i fer-ne un recorregut permet obtenir una panoràmica de gran interès, que destaca els conceptes més fonamentals i els agrupa en una estreta i rigorosa xarxa de significacions.
- Un thesaurus és un model normatiu, amb suport institucional, per a les persones que treballen en l'anàlisi del contingut informatiu de documents, i que estableix una comunitat de sentit entre elles i els destinataris de la informació.

Consulta en connexió amb la base SINERA

Les mateixes tècniques de consulta són efectives tan al thesaurus com a la base SINERA a la qual dóna suport. Aquestes tècniques estan recollides a l'opció **Guia del CD-ROM** del menú Ajuda de SINERA en Disc, també es descriuen a la secció mètodes de consulta de la **introducció a la base SINERA**. Un breu resum el podeu trobar a la secció mètodes de cerca d'aquesta introducció.

Si s'activa la recerca simultània en els dos contextos, la investigació d'un terme o d'un concepte tindrà una doble resposta: el sistema informàtic recuperarà els documents de la base on figuri, i també l'enregistrament del thesaurus al qual pertany.


La lectura atenta d'aquesta branca de l'arbre, i d'aquelles que li són properes jeràrquicament, pot ampliar la perspectiva de la recerca original i suggerir aproximacions noves i més adequades a la necessitat inicial d'informació.


EP = emprat per

Estructura d'un enregistrament

Cada enregistrament del thesaurus conté una secció de l'arbre jeràrquic de termes, composta per un terme d'encapçalament i pels termes que hi estan subordinats, fins a cert nivell d'especificitat.


Aquest nivell de desplegament de l'arbre varia d'un enregistrament a un altre, i s'ha fixat per tal de contenir en cada finestra fragments homogenis d'informació d'una extensió que els faci fàcilment consultables.


Els termes que figuren en **verd** i subratllats, i sobre els quals el punter del ratolí adopta la forma d'una mà  són les portes a enregistraments nous i clicant-los es va davallant per l'estructura jeràrquica.


El desplaçament en sentit ascendent es fa clicant successivament el botó  que figura a la dreta de la línia superior de tots els enregistraments.

Mètodes de cerca

Per cercar un terme al thesaurus és recomanable emprar l'opció corresponent al botó

 L'índex del thesaurus conté tots els termes que hi figuren i exactament tal com hi figuren.

El desplaçament per l'índex es pot fer de manera seqüencial, però és preferible inserir el terme buscat a la línia d'edició que figura a la part superior de la finestra **Índex**. Un cop localitzat el terme, el botó  indica quins enregistraments el contenen. Cada terme figura en un sol enregistrament, excepte aquells termes que són nusos de l'arbre, que figuren en dos arbres, o que són polijeràrquics.

El botó  procedeix a mostrar l'enregistrament en qüestió.

Si l'usuari no coneix la forma precisa dels termes que desitja trobar és preferible que activi l'opció Cerca documental al thesaurus.

Cerca documental al thesaurus

Aquesta opció és la més adequada per a la cerca d'informació en enregistraments que contenen un text redactat lliurement, com és el cas de la base SINERA.

Els avantatges que presenta aquest tipus de cerca en la consulta del thesaurus són dos:

- pot recuperar paraules aïllades, que han estat incorporades a l'índex formant part d'una expressió,
- permet consultar simultàniament la base i el thesaurus i, per tant, utilitzar el thesaurus per guiar la cerca a la base.

La cerca documental s'activa amb el botó  i ofereix com a resposta el nombre d'enregistraments de les aplicacions actives on figuri la paraula o la fórmula de cerca introduïda a la línia d'edició. Aquests enregistraments es poden llegir clicant el botó



Per a més detalls sobre les fórmules i els mètodes de cerca consulteu l'epígraf [Mètodes de consulta](#) de la [Introducció a la base SINERA](#).

