

El portal CiT (Terminologia de Ciències i Tecnologia)

Salvador Alegret

Secció de Ciències i Tecnologia (IEC)

Joana Torres

Tècnica lingüística del CiT

Introducció

El portal CiT, Terminologia de Ciències i Tecnologia (<http://cit.iec.cat>), és fruit d'un projecte de recerca de la Secció de Ciències i Tecnologia de l'Institut d'Estudis Catalans nascut l'any 2004 arran de la Declaració de Berlín sobre l'accés lliure al coneixement de les ciències i de les humanitats (2003). L'objectiu d'aquesta declaració era promoure el coneixement científic i el pensament humà, incloent-hi el patrimoni cultural, i garantir-ne l'accés mundial. Tant la Generalitat de Catalunya com l'Institut d'Estudis Catalans s'hi adheriren, i la Secció de Ciències i Tecnologia (SECCT) de l'IEC va decidir treballar en aquesta direcció. A partir de la idea que determinades dades de les institucions públiques o privades havien de ser d'accés lliure per a ser reutilitzades sense restriccions d'ús ni mecanismes de control, d'acord amb el moviment internacional conegut com a *open data* (o 'dades obertes'), s'emprenquè la creació d'aquesta plataforma informàtica amb la finalitat de posar a l'abast de la comunitat científica la terminologia de ciències i tecnologia generada per l'IEC i per institucions vinculades, la qual s'ha fet pública recentment per mitjà del portal CiT.

És conegut que actualment molts dels científics experimentals, els que treballen en recerca pura i aplicada, i que bonament generen coneixement —com ara els nanotecnòlegs o els químics analítics—, utilitzen la nostra llengua als laboratoris, però difonen internacionalment el coneixement generat en anglès. Malgrat això, cal disposar de terminologia en català per a la difusió científica dins de la nostra comunitat lingüística. Necessitem eines que ens facilitin traslladar de l'anglès al català quotidià els continguts de les revistes científiques internacionals (en llengua

anglesa). L'editorial catalana més important en aquest àmbit és l'IEC. Per això, vam dissenyar un portal amb unes característiques determinades, amb la intenció que fos útil per a dur a terme la difusió de les ciències i la tecnologia en llengua catalana; un portal modular, format per diverses eines i projectes, moltes de les quals encara estan en desenvolupament (figura 1).



FIGURA 1. Salvador Alegret i Joana Torres presentant el portal CiT.

La primera eina que vam dissenyar fou una biblioteca d'obres terminològiques (BiblioCiT), fruit de la limitació de les edicions existents que, en molts casos, només tenien difusió dins la càtedra universitària corresponent. Algunes editorials van apostar per aquestes publicacions especialitzades, com és l'obra del doctor Oriol Riba, geòleg de gran experiència que es va dedicar a la confecció del *Diccionari de geologia* a la darrereria de la seva vida professional. Però també és sabut que la ciència «neix, creix i mor», i, en aquell moment, la major part de les edicions eren molt limitades, amb tiratges d'uns dos-cents cinquanta exemplars que l'autor distribuïa directament entre els col·legues d'especialitat.

Així, la biblioteca en línia fou ideada per a recuperar aquestes obres tan interessants i, actualment, exhaurides. Superats els tràmits legals de recuperació dels drets d'edició —cal dir que sempre hem trobat una bona complicitat amb les grans editorials catalanes en aquest aspecte—, vam poder començar a presentar les obres de

manera actualitzada i a posar-les a l'abast dels usuaris. Aquesta iniciativa es descabellà pensant, sobretot, en la comunitat científica i universitària, ja que la major part de les obres, que van fixar la terminologia al seu moment, eren rareses difícils de trobar a les biblioteques. També es va procurar de protegir-les amb les llicències Creative Commons, per tal de resguardar-ne l'autoria i evitar que se'n facin obres derivades.

A continuació, presentarem les característiques principals del portal CiT i mostrem totes les eines de què disposa —entre les quals hi ha la biblioteca— mitjançant exemples il·lustratius.

1. El portal CiT

El portal CiT (Terminologia de Ciències i Tecnologia) és una plataforma web que aplega obres de terminologia de ciències —en un sentit ampli— i de tecnologia. Està concebut com una estructura web amb diversos mòduls, cadascun dels quals correspon a una eina (figura 2).



FIGURA 2. Aparença de la pantalla d'inici del portal CiT a l'inici del mes d'abril del 2015.

Actualment, hi ha tres recursos operatius. El primer és la BiblioCiT, una biblioteca en línia d'obres terminològiques que poden ser consultades de manera individual fent servir opcions de cerca avançada. Aquest recurs aplega una trentena d'obres especialitzades de diferents àmbits del saber. En segon lloc, trobem el CercaCiT, un motor de cerca avançat que permet consultar totes les obres de la BiblioCiT simultàniament o per àrees temàtiques. Actualment, s'hi poden consultar un total de més de 120.000 unitats terminològiques catalanes.¹ I, en tercer lloc, disposem de l'eina ContextCiT, que permet trobar contextos d'ús dels termes apareguts en les revistes especialitzades recollides en l'Hemeroteca Científica Catalana (HCC). La plataforma tecnològica està construïda sobre una base de dades relacional i un entorn de consulta web basat en pàgines dinàmiques.

2. La biblioteca en línia d'obres terminològiques BiblioCiT

Actualment, la biblioteca d'obres terminològiques comença a fer goig: les obres es presenten agrupades en «prestatges» segons l'àmbit d'especificitat a què pertanyen: ciències de la terra i de l'espai, ciències de la vida i la salut, ciències exactes i experimentals, ciències socials i humanes, i enginyeria i tecnologia (figura 3). La major part de les obres que s'hi apleguen han estat publicades originàriament en paper per l'IEC, per altres institucions —com la Universitat de les Illes Balears— i per editorials de prestigi reconegut —com ara Enciclopèdia Catalana, Barcino o Reverté. Tot i això, la biblioteca virtual també conté obres inèdites, com ara el *Diccionari de les ciències ambientals*, publicat fins ara únicament en línia dins la plataforma CiT.

Certament, es tracta d'un projecte ambiciós, ja que la BiblioCiT inclou obres de naturalesa molt diversa: diccionaris (*Diccionari de geologia*, *Diccionari de l'utilitatge químic*); diccionaris enciclopèdics (*Diccionari jurídic*, *Diccionari enciclopèdic de medicina*); vocabularis (*Vocabulari de luminotècnia*, *Vocabulari de biologia de la reproducció*, etc.); nomenclatures (*Codi Internacional de Nomenclatura Zoològica*; *Magnituds, unitats i símbols en química física*); terminologies de manuals científics (*Fonaments d'economia*, de Krugman); glossaris i diccionaris multilingües (*Vocabulari multilingüe de la ciència del sol*, *Diccionario multilingüe de la ciencia del suelo*); terminologies d'obres d'autor (la terminologia continguda en l'obra *Els basti-*

1. En el moment de tancar l'edició d'aquestes actes (abril del 2015), la BiblioCiT aplega un total de 156.058 unitats terminològiques procedents de cinquanta obres especialitzades.

BiblioGT

Aquesta biblioteca en línia conté bàsicament publicacions terminològiques de ciència i tecnologia de diferent tipologia (diccionaris, vocabularis, nomenclatures, manuals i altres) editades per l'Institut d'Estudis Catalans o vinculades a aquesta institució.

El total d'unitats terminològiques catalanes aplegades fins ara és aproximadament 170.000.

Ciències de la Terra i de l'espai

| | | | | |
|---|--|---|--|-------------------------------------|
| | | | | |
| Aportació a la terminologia geogràfica catalana (1998) | Diccionari de geologia (1997) | Diccionari de les ciències ambientals (2008) | Diccionari multilingüe de la ciència del sòl (2014) | Ecologia I (2011) |
| | | | | |
| Ecologia II (2011) | Guia estratigràfica internacional (versió abreujada) (2001) | Sistemes d'informació geogràfica (2011) | Una taula dels temps geològics (1986) | Vocabulari de mineralogia (2001) |

FIGURA 3. Secció de la biblioteca que aplega les obres de ciències de la terra i de l'espai.

ments menors catalans de construcció artesanal del segle XX, d'Alexandre Ribó Golovart), i altres obres de terminologia (*Novetat i llenguatge*, de Lluís Marquet).

Hem volgut preservar la identitat de les obres, que els autors hi siguin reconeguts. Però la nostra voluntat no solament és que les obres guardin l'autoria, sinó que també mantinguin la integritat, és a dir, que apleguin tots els continguts —encara que no es presentin en el format tradicional de llibre—, tant lexicogràfics com supraestructurals. L'estructura, evidentment, com que forma part de la història de la terminologia reglada de les ciències i la tecnologia, és molt diversa i complexa.

Per aquest motiu, els camps en els quals es poden fer consultes no són idèntics en totes les obres. Per exemple, el *Vocabulari multilingüe de la ciència del sol* permet la cerca per àrea temàtica (tots els termes en tenen una o més d'una), però no la cerca per definició, ja que no n'hi ha; mentre que l'*Aportació a la terminologia geogràfica catalana* no disposa de cerca per àrees temàtiques, però sí que s'hi poden fer cerques en la definició.

Entrem a veure quina aparença té aquesta última obra terminològica de geografia catalana. D'entrada, es pot observar que la part situada a l'esquerra de la pantalla s'adapta específicament a l'estructura lexicogràfica adoptada per l'autor i que la part de la dreta conté el motor avançat de cerca (figura 4).

La informació mínima que té cada terme del recull és l'entrada i la categoria gramatical. D'altra banda, hi ha termes que tenen informació addicional, com són la definició, exemples, sinònims, notes i termes relacionats. Aquest recull terminològic en línia és una base de dades accessible a través d'una interfície. La manera més similar al que seria «fullejar» el recull a la manera tradicional es pot aconseguir fixant una lletra —per exemple, la *a*— en la casella «Mot, fragment de mot o sintagma a cercar» i seleccionant l'opció «Que comenci per»; d'aquesta manera, el cercador farà que es despleguin totes les entrades que comencen per la lletra escollida.



FIGURA 4. Aparença en pantalla de l'obra *Aportació a la terminologia geogràfica catalana*, de Marc-Aureli Vila i Comaposada.

A continuació mostrarem una prova de cerca en el *Diccionari jurídic*. Es tracta, principalment, d'un diccionari enciclopèdic especialitzat que vol recollir els termes que formen part de les diverses disciplines del dret, com ara el dret públic, el dret civil, el dret penal, el dret processal, el dret canònic, el dret ambiental i el dret mercantil, entre d'altres. També inclou termes que pertanyen a la història del dret, ressenyes biogràfiques de juristes eminents ja traspassats que han tingut una producció científica destacable i conceptes jurídics de les diverses branques. Igualment, el diccionari aplega fraseologia jurídica, molt necessària per a la redacció, la correcció i la traducció de llenguatge jurídic (figura 5).

Els articles es divideixen en dues parts ben diferenciades: en primer lloc, hi ha la definició lexicogràfica del terme (llevat del cas de les biografies), i, en segon lloc, l'ampliació enciclopèdica corresponent, que pot arribar a ser molt extensa i que comprèn informacions des del punt de vista de la doctrina, la legislació, la jurisprudència, i, pel que fa als termes històrics, l'explicació de les circumstàncies històriques i socials que han envoltat i han permès el desenvolupament del dret dels diferents territoris de llengua catalana. Hi ha entrades, però, que no inclouen aquesta ampliació, ja que la redacció del conjunt enciclopèdic està en curs.



FIGURA 5. Exemple de cerca per camps del terme *legítima* dins el *Diccionari jurídic*.

Altres informacions que contenen els articles i que poden ser útils són les equivalències en espanyol, els termes relacionats amb l'entrada consultada —i que permeten fer-se una idea global de l'abast del terme— i els sinònims. A tall d'exemple podem fer una cerca avançada del terme *legítima* i veure quants articles contenen informació enciclopèdica d'aquest mot dins l'àmbit territorial de Catalunya; el resultat són noranta registres amb aquests paràmetres de cerca.

3. CercaCiT, el cercador transversal dels termes continguts a BiblioCiT

El motor de cerca CercaCiT ofereix la possibilitat de fer una cerca transversal en totes les obres recollides a la BiblioCiT. Així, es poden consultar totes les publicacions de manera simultània, però també de manera selectiva i amb criteris de cerca diversos.

A diferència del que passa a la BiblioCiT, al CercaCiT s'han homogeneïtzat les fitxes de les unitats terminològiques, de manera que totes tenen la mateixa estructura i es mostren els mateixos camps d'informació: entrada, categoria gramatical, títol de l'obra de la qual prové l'entrada i àrea temàtica. A més, també apareixen, si n'hi ha, les abreviacions, la definició, les equivalències, els exemples d'ús, els sinònims i els termes relacionats. El motor de cerca permet fer dos tipus generals de selecció: d'una banda, indicar l'obra o obres en què es vol fer la consulta i, de l'altra, indicar en quin camp o camps volem que es faci la cerca i la manera de dur-la a terme. Pel que fa a la tria d'obres, es pot fer una cerca en totes, especificar-ne una de concreta o triar un dels blocs temàtics d'obres de la BiblioCiT. En relació amb la segona selecció, es pot especificar en quins camps es fa la consulta (entrada, definició, equivalències, sinònims, exemples i termes relacionats, o en tots els camps alhora) i quin grau de coincidència hi ha d'haver amb la cadena de caràcters introduïda («Coincident», «Començada per», «Acabada per» o «Que conté»). D'altra banda, mitjançant un menú desplegable, es pot triar entre les diverses àrees de coneixement per a delimitar més la cerca.

Seguidament, mostrem dues proves realitzades amb el motor de cerca avançada:

a) Posem d'exemple el terme *aerobi*: fent una cerca general en tots els camps per veure quantes entrades hi ha que contenen aquesta paraula, ens apareixen 388 registres. Ara bé, podem restringir-ne els resultats per mitjà de la cerca avançada. Delimitem que es busqui *aerobi* només a l'entrada i que hi coincideixi (7 registres) o bé que contingui *aerobi* a l'entrada, dins l'àrea d'ecologia (10 registres) (figura 6). També

podem provar de cercar les equivalències en llengua anglesa que continguin *aerobi* (46 registres).²

b) També es pot cercar per una seqüència de caràcters per mitjà de les opcions «Que comenci», «Que acabi» o «Que contingui». Posem per cas que ens interessa saber la forma exacta d'un terme: *calibrat*, *calibratge* o *calibració*. Cal que acotem la cerca perquè busqui totes les entrades que acaben en *-atge* o que comencen amb *cali-* (les entrades acabades en *-atge*, per exemple, es corresponen amb 376 registres dins el total de trenta-quatre obres).³

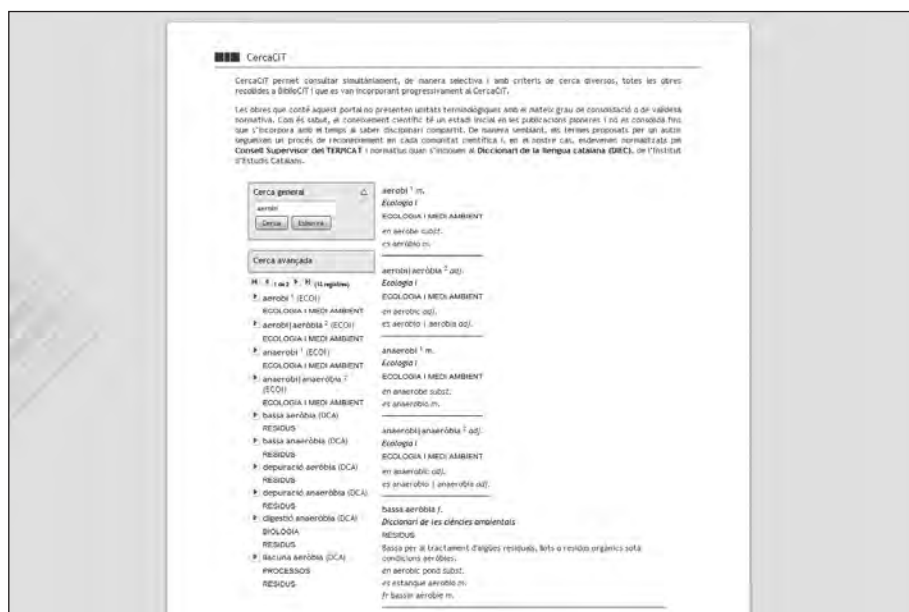


FIGURA 6. Exemple del resultat de cerca avançada que contingui el terme *aerobi* a l'àrea d'ecologia.

2. Fetes de bell nou les cerques d'aquest apartat *a* a començament d'abril del 2015, per raó de la incorporació constant d'obres noves en aquesta plataforma, els resultats han estat els següents: el terme *aerobi* apareix dins 420 registres, figura com a entrada en 9 registres, figura com a part de l'entrada en 13 registres i hi ha 73 registres amb l'equivalència anglesa d'aquest terme.

3. Fetes de bell nou les cerques d'aquest apartat *b* a començament d'abril del 2015, els resultats han estat els següents: hi ha 495 registres que tenen entrades acabades en *-atge*.

4. El cercador d'exemples d'ús de la terminologia científica ContextCiT

A diferència de les dues eines descrites fins ara, que se centren en el significat i les relacions lingüístiques dels termes (sinonímia, equivalències, etc.), el ContextCiT és un cercador avançat que troba termes en el seu context mitjançant un corpus textual. El conjunt de textos sobre el qual es realitza el buidatge el formen les revistes recollides en l'Hermetoteca Científica Catalana de l'IEC (HCC) (<http://revistes.iec.cat>), que recull més de disset mil articles. Els textos que apareixen en les revistes especialitzades són una font molt valuosa d'unitats terminològiques, ja que les difonen i, fins i tot, les fixen. Per aquest motiu i aprofitant que bona part de les revistes de la HCC estan digitalitzades en format PDF, amb l'ajuda d'un cercador de Google personalitzat, es fan consultes en els documents, els resums i les metadades de la HCC. Actualment, la cerca es limita als documents que incorporen sistemes de reconeixement òptic de caràcters (OCR).

D'aquesta manera disposem d'una eina que ens ofereix la possibilitat de documentar-nos a l'hora de prendre decisions davant d'un dubte sobre qualsevol terme científic. Per exemple, fa temps ens vam topar amb la paraula *serendipity* en un text, marcada amb cursiva com un manlleu no adaptat gràficament. Fent una cerca al ContextCiT, a fi d'analitzar-ne les freqüències d'ús i saber si es fa servir en algun context el mot adaptat en català, obtenim cinc resultats de contextos amb el terme *serendipity* (figura 7): dos d'aquests es troben en anglès; dos en un article en català a la revista *Ciència. Revista Catalana de Ciència i Tecnologia* (1983) titulat «LSD, de la química a la reverència», de Montserrat Majenc, i el cinquè en el context d'*Omnis Celulla* (desembre del 2004), en la secció «Calaix de sastre», on es reproduïx un fragment d'un llibre de Mercè Piqueras titulat *Cròniques de l'altra veritat* en què precisament dos personatges tracten, a propòsit d'una conversa, de l'adaptació d'aquest terme al català i decideixen que se n'ha de dir *serendipitat*.



FIGURA 7. Resultats de la cerca al ContextCiT del terme *serendipity*.

5. Una nova eina en construcció: ExtractCiT

El programa CiT és un projecte viu que participa en la creació de noves eines amb vista al futur. ExtractCiT serà la plataforma de presentació de la terminologia extra automàticament d'un corpus de textos, per mitjà de l'estació de treball Terminus 2.0, desenvolupada pel grup IULATERM, de la Universitat Pompeu Fabra, amb el qual el CiT ha establert un conveni de col·laboració. Les tasques que s'han dut a terme fins ara són la constitució d'un corpus i l'extracció de la terminologia de les revistes de la HCC. El resultat és una llista de «possibles candidats» a terme amb el grau corresponent de «terminologicitat» tenint en compte la freqüència d'aparició, els contextos d'ús i la referència (obra i localització dins l'obra).

A partir del buidatge, els termes es presentaran en format de base de dades d'accés lliure amb el propòsit que la comunitat científica tingui a l'abast una font terminològica actualitzada, útil per a fer estudis transversals o analitzar la vida dels termes, entre altres possibilitats. Tal com hem vist en la consulta del CercaCiT, aquesta

nova eina també permetrà seleccionar la revista que conté la terminologia objecte de cerca, així com buscar termes en totes les revistes de manera simultània o bé acotant la cerca només a unes quantes.

6. Conclusions

La terminologia és l'element essencial de vehiculació i difusió del progrés científic i tecnològic. Per aquest motiu, la Secció de Ciències i Tecnologia de l'IEC té la tasca de vetllar perquè les unitats lèxiques dels àmbits científics siguin accessibles al públic de tota mena —des de científics fins a estudiants, passant pels professionals de la llengua. Tal com hem apuntat en la presentació, la naturalesa de les obres terminològiques catalanes, destinades a un públic especialitzat reduït, fa que les edicions en paper tinguin tiratges també reduïts i quedin descatalogades en un període de temps breu. Això mateix també passa amb algunes obres científiques en català que no són estrictament terminològiques.

El portal CiT té per objectiu donar una nova vida a aquestes obres i a la terminologia que contenen. Així, les trasllada a un entorn informàtic perquè tinguin més accessibilitat, visibilitat i funcionalitat —respectant sempre l'autoria, la integritat i la identitat dels continguts originals. D'aquesta manera, a més, passen a formar part d'una xarxa terminològica que s'enriqueix incorporant-hi noves obres. Per aquesta raó, l'estructuració de les bases de dades que es poden consultar al portal CiT, procedents d'obres molt diverses i de concepció lexicogràfica molt variada, amb el suport de cercadors avançats i criteris de cerca específics, està orientada a trobar noves correlacions entre els significats dels termes i les relacions lingüístiques i terminològiques que mantenen.