

Vocabulària: un multicercador temàtic

XAVIER ALBONS, PEP CARA, ÀNGELS EGEA, MONTSERRAT LLEOPART
Serveis Lingüístics
Universitat de Barcelona

Resum

Aquesta comunicació presenta el projecte Vocabulària, desenvolupat per la Universitat de Barcelona, que pretén la difusió de la terminologia catalana correcta. S'exposen els antecedents del projecte i els multicercadors que utilitza, i s'explica la gestió que fa dels continguts i la presentació dels resultats.

PARAULES CLAU: bloc, multicercador, terminologia, Vocabulària.

Abstract: *Vocabulària: a theme-based multi-search engine*

This communication presents the Vocabulària project developed by the University of Barcelona, which seeks to disseminate correct Catalan terminology. The background of the project is outlined, and the multi-search engines it uses, and the way it manages contents and presents results is explained.

KEY WORDS: blog, multi-search engine, terminology, Vocabulària.

1. INTRODUCCIÓ

Vocabulària és un projecte desenvolupat pels Serveis Lingüístics de la Universitat de Barcelona (UB) (<http://www.ub.edu/sl>) en què han participat el professorat i l'alumnat de la Xarxa de Dinamització Lingüística. Té l'objectiu de facilitar l'accés a la terminologia correcta usada en cada àmbit de coneixement mitjançant l'ús de multicercadors i la difusió d'obres en línia.

2. ANTECEDENTS DEL VOCABULÀRIA

Fa una vintena d'anys, alumnes de la Facultat de Química van fer un recull inicial de termes de química a partir del buidatge dels seus apunts de classe. Aquest

material de partida, amb l'aportació conceptual del professorat i metodològica dels Serveis Lingüístics (aleshores Servei de Llengua Catalana), va ser l'embrió de la col·lecció de «Vocabularis Bàsics per a l'Alumnat» (www.ub.cat/enllaca/directori.php?branca=498). Els vocabularis d'aquesta col·lecció, uns llibrets que recullen la terminologia bàsica de les matèries tractades a la UB amb equivalències en castellà i en anglès principalment, han acomplert durant gairebé dues dècades l'objectiu inicial de difusió de la terminologia catalana correcta, i més endavant han esdevingut un recurs de suport per a alumnat nouvingut i professorat visitant.

Els avenços tecnològics i la implantació de les noves tecnologies de la informació i la comunicació dins i fora de la comunitat universitària han permès, d'una banda, passar del format en paper al format electrònic —més barat de produir i més fàcil de difondre i d'actualitzar— i, de l'altra, han millorat les possibilitats de cerca d'informació i, per tant, d'aprofitament d'altres recursos en línia. Això és especialment important si tenim en compte la magnitud de la comunitat universitària: uns 81.000 estudiants (comptant-hi els de formació continuada), 4.700 professors i 2.200 treballadors d'administració i serveis, i la renovació periòdica d'una bona part d'aquest col·lectiu.

En aquest marc neix el projecte Vocabulària, que manté els mateixos objectius que la col·lecció de «Vocabularis Bàsics per a l'Alumnat» (difusió de la terminologia catalana correcta), però que, aprofitant les possibilitats que ofereixen les noves tecnologies, vol difondre la terminologia d'elaboració pròpia i també la de totes les obres consultables en línia en un sol corpus de consulta global o temàtica. La tecnologia de base per al desenvolupament del Vocabulària són els multicercadors.

3. ELS MULTICERCADORS

Els multicercadors dels Serveis Lingüístics de la UB estan inspirats en el OneLook (www.onelook.com), un motor de cerca que conté, indexats i classificats temàticament, més de mil diccionaris en anglès, en els quals es poden fer cerques simultànies amb una resposta immediata i exhaustiva en què es detallen totes les fonts que contenen la cadena cercada i l'àrea temàtica a què pertanyen.

És una eina de cerca d'informació ràpida, perquè es fa una sola cerca a tot el conjunt i la resposta és immediata perquè la informació està indexada; i és una eina exhaustiva perquè inclou un nombre elevat de recursos d'una tipologia determinada. A més, presenta l'avantatge que l'usuari no necessita conèixer els recursos que li poden ser útils per trobar una determinada informació, sinó que el mateix motor de cerca mostra en la resposta els recursos que contenen la informació cercada. Així, un multicercador d'aquestes característiques és útil també com a eina de difusió de recursos.

Des de la generalització d'Internet com a mitjà d'informació i comunicació, la presència de recursos lexicogràfics i terminològics en català ha estat molt important. La major part de les institucions productores d'aquests recursos han optat per difondre'ls en línia: el *Diccionari de la llengua catalana* (DIEC) de l'Institut d'Estudis Catalans (IEC); *L'enciclopèdia*, el *Gran diccionari de la llengua catalana* (GDLC) i el *Diccionari enciclopèdic de medicina* (DEM) (Grup Enciclopèdia Catalana); Cercaterm (TERMCAT), etcètera.

La importància d'Internet per a la difusió en línia ha estat decisiva en el cas dels recursos lexicogràfics més petits, com ara els vocabularis de la col·lecció «Vocabularis Bàsics per a l'Alumnat» de la UB, i moltes altres col·leccions similars elaborades en altres universitats catalanes o les obres terminològiques elaborades des de molts altres organismes, com ara la Generalitat de Catalunya o el Consorci per a la Normalització Lingüística (CPNL).

Hi ha hagut iniciatives, fins i tot institucionals —com ara la de Llengua.org (www.llengua.org)—, per a fer directoris exhaustius de recursos en línia sobre diversos aspectes de la llengua catalana. Només cal fer-hi un cop d'ull per a adonar-nos de la gran quantitat de recursos existents i, per tant, de la dificultat de conèixer-los tots i poder-los consultar.

Així, doncs, la gran producció d'obres terminològiques disponibles en línia s'ha acabat convertint en un inconvenient per a arribar als possibles usuaris. En la pràctica, els professionals de la llengua i altres usuaris tendeixen a consultar únicament els grans recursos en línia: DIEC (dlc.iec.cat), GDLC (www.enciclopedia.cat), *L'enciclopèdia* (www.enciclopedia.cat), *Cercaterm* (www.termcat.cat), *Diccionari enciclopèdic de medicina* (www.grec.net/home/cel/mdicc.htm), etc. I la resta de recursos, de dimensions menors, són ignorats i, per tant, desaprofitats, tot i que la suma de tots plegats pot arribar a representar un volum de dades igual o superior a alguns dels grans recursos.

En aquest context, els Serveis Lingüístics de la UB es van plantejar la necessitat de desenvolupar un multicercador terminològic que tingués unes prestacions similars al OneLook, és a dir, una eina que permetés tenir indexats tots els recursos existents d'una determinada tipologia per a poder-hi fer cerques simultàniament.

Es va observar que pràcticament tots els recursos susceptibles de ser integrats en el multicercador terminològic estaven indexats pels motors del Google, és a dir, eren consultables des del cercador Google. Per tant, si podíem aprofitar aquesta plataforma per a construir el nostre multicercador, no hi havia necessitat de construir un corpus de recursos indexats *ad hoc* —com es fa en l'Optimot (optimot.gencat.cat)—, amb totes les dificultats tècniques que això podia comportar i les dificultats d'actualització dels continguts quan la font originària s'ha modificat. A més, la indexació garanteix la rapidesa en la resposta, que no està garantida si no

hi ha indexació prèvia, com passa amb el Metacercador de la UPC (www.upc.es/slt/metacercador).

Només calia trobar la manera de restringir l'univers de cerca al conjunt de recursos que es volia incloure en el multicercador. El mateix Google ens va oferir la manera de construir el multicercador terminològic gràcies a la possibilitat de construir motors de cerca personalitzats (www.google.com/coop/cse). La construcció d'un motor de cerca personalitzat consisteix, simplement, en la introducció dels localitzadors universals de recursos (URL) dels recursos seleccionats en el motor de cerca.

La part més laboriosa de l'elaboració del multicercador va ser la confecció de la llista exhaustiva de tots els recursos que havia de contenir i el manteniment posterior dels URL d'aquests recursos, que es fa trimestralment.

Atesa la naturalesa dels recursos seleccionats, i amb l'objectiu de millorar els resultats de la cerca, es va considerar preferible de construir dos motors de cerca o multicercadors: un de terminològic i un de lingüístic.

El projecte Vocabulària parteix del multicercador terminològic i en fa un desenvolupament pensat per oferir a cada perfil d'usuari els recursos que poden ser més útils.

4. IMPLEMENTACIÓ I DESCRIPCIÓ DEL VOCABULÀRIA

El Vocabulària és una interfície de consultes terminològiques temàtiques i també una plataforma d'informació i difusió feta amb WordPress, un sistema de gestió de continguts de codi obert. Aquest sistema és un dels més usats per a la publicació de blocs a la xarxa. Els gestors de continguts, com ara WordPress, són sistemes que permeten fer i actualitzar webs amb molta facilitat.

Els blocs (figura 1) tenen una pàgina principal o dinàmica on es van publicant notícies o entrades classificades o ordenades per categories o etiquetes temàtiques i, a part, poden tenir diverses pàgines estàtiques. Tant la pàgina dinàmica com les estàtiques comparteixen les barres laterals, la capçalera i el peu de pàgina.

A la pàgina dinàmica hi ha les informacions que són notícia, cadascuna amb un descriptor que les identifica per categories. Aquestes categories es corresponen temàticament amb les facultats —soles o agrupades temàticament— de la Universitat de Barcelona, per exemple, *biologia*, *química*, *ciències de la salut*, *dret*, *economia*, *educació*, *física*, *geologia*, *humanitats*, *matemàtiques*, i, finalment, també una categoria *general* per a les notícies que no corresponen temàticament a cap facultat o grup de facultats. En aquesta part dinàmica, es publicaran les novetats terminològiques fruit d'edicions o de consultes i altres notícies d'actualitat terminològica.

Pel que fa a les barres laterals, es va optar per un esquema estàtic o *tema* que té dues barres laterals. A la barra de l'esquerra hi ha enllaços a una sèrie de recur-

FIGURA 1. Portal del lloc web del projecte Vocabulària (<http://www2.ub.edu/sl/vocabularia>)

sos que estan en estreta relació amb la terminologia. S'ha enllaçat el gestor de consultes lingüístiques i terminològiques Sens Dubte, el diccionari personalitzat per als verificadors ortogràfics, el sistema de traducció automatitzada Internostrum i, finalment, la pàgina de recursos lingüístics de l'Àrea d'Assessorament Lingüístic i Terminologia dels Serveis Lingüístics de la UB, que conté altres eines per a cercar informació lingüística i terminològica o per a la redacció i edició de textos. A la barra de la dreta hi ha una breu descripció de què és el Vocabulària, qui el fa i també els arxius de les notícies de la pàgina principal ordenades per mesos i per categories.

Les pàgines estàtiques tenen una importància bàsica al web i s'hi pot accedir per unes pestanyes, molt visibles, situades sobre la capçalera. De fet, les diferents facultats poden fer un enllaç a la pàgina estàtica que els correspon, de manera que per a elles serà la pàgina principal del Vocabulària. Cada pàgina estàtica correspon a una facultat o un grup de facultats agrupades temàticament. Dins de cada pàgina estàtica hi ha:

- 1) El multicercador de la facultat o les facultats.
- 2) Els vocabularis en PDF elaborats pels Serveis Lingüístics i les comissions

de dinamització lingüística de cada centre, ubicats al dipòsit digital amb el seu identificador (*handle*) corresponent i on també cerca el multicercador.

3) Si es dóna el cas, altres recursos als quals no pot accedir el multicercador, com ara el *Diccionari enciclopèdic de medicina*.

De moment, s'han implementat les pàgines estàtiques corresponents a *biologia, química i ciències de la salut*, i progressivament s'aniran afegint la resta de facultats de la UB.

La participació del professorat i l'alumnat en el Vocabulària consisteix a assessorar sobre les obres terminològiques que formen el corpus de cada branca de cerca. Les comissions de dinamització lingüística de les facultats de la UB, on hi ha un representant de cada departament, donen el vistiplau a la selecció feta pels Serveis Lingüístics i poden fer propostes d'inclusió quan apareguin noves fonts interessants en el futur.

5. DESENVOLUPAMENT FUTUR

Els multicercadors presenten els resultats de cerca en l'ordre que Google té establert per defecte, que no sempre es correspon amb l'ordre de prioritat que es voldria donar a les fonts que constitueixen els multicercadors. En un futur proper està previst d'intervenir en la presentació dels resultats, de manera que apareguin en l'ordre que es consideri més convenient.