

Balanç i conclusions de la VII Jornada de la SCATERM

MARINA SALSE, JAUME MARTÍ
Coordinadors de la Jornada

Les ponències i les comunicacions que s'han presentat en aquesta Jornada s'orienten en dos sentits, que reflecteixen en bona part les tendències actuals del treball investigador en els camps en què s'ha centrat.

D'una banda, reflecteixen una preocupació considerable per la gran acumulació d'informació que significa Internet. La Xarxa esdevé un camp de treball, experimentació i investigació molt important que fa prendre formes noves a tècniques de treball usades d'antic:

— Indexació per llenguatges controlats. Usats tradicionalment, continuen ara estant presents a la xarxa, però actualment prenen la forma de tendències noves i de vegades parteixen d'un tractament automàtic de base de la informació. Així, com ens afirmava el doctor José Luis Alonso Berrocal, els tesaurus tradicionals apareixen ara no solament com una eina d'organització de la informació en un entorn no automatitzat, sinó també com a expansions de la consulta en entorns informàtics. A part d'això, el doctor Lluís Codina ens ha parlat de les noves formes de llenguatge controlat aplicades a Internet, com ara les taxonomies, les ontologies o les folksonomies. També podem considerar sistemes de classificació o llenguatges controlats els etiquetatges que es fan de determinats termes per facilitar l'estructuració i la consulta dels recursos electrònics. Un exemple d'això, l'han posat els membres dels Serveis Lingüístics de la Universitat de Barcelona amb llur Vocabulària. Tanmateix, encara les iniciatives són molt aïllades. Els grans projectes, com el Web semàntic, del qual ens ha parlat el doctor Codina, estan de moment poc madurs i no sabem si reeixiran.

— Indexació automàtica. Els estudis en aquest camp es van iniciar els anys seixanta i es van vincular a l'explotació de bases de dades. L'aparició d'Internet i dels seus milions i milions de pàgines ha fet que s'hagin hagut de buscar sistemes per a aconseguir cada vegada més bons resultats, atès que la indexació humana

hauria necessitat una gran inversió de temps i diners. Així, doncs, s'han fet i es fan nombrosos estudis per millorar la indexació automàtica i per crear resums automàtics. Les tendències en investigació han portat a crear dos grans tipus de mètodes: els no lingüístics, especialment centrats en l'estadística i el càlcul de probabilitats, i els lingüístics, que poden anar des de la simple lematització fins a l'ús de tesaurus i/o corpus documentals per a millorar el tractament terminològic de la indexació controlada i, en conseqüència, també la recuperació de la informació. Tot i que normalment els sistemes d'indexació solen usar una combinació d'ambdós tipus de mètodes, en la Jornada d'avui se'ns han presentat essencialment mètodes de tipus estadístic, com el model vectorial, mostrat pel doctor Alonso Berrocal o bé els estudis de Rogelio Nazar i dels professors de la Universitat Oberta de Catalunya, Mercè Vázquez i Antoni Oliver.

Ara hi ha, doncs, un important maridatge entre estadística, lingüística i informàtica. Lletres i ciències, abans tan separades en els plans d'estudis, s'han unit gràcies a la Xarxa i a les noves tendències d'investigació en indexació i recuperació de la informació.

D'altra banda, reflecteixen la vigència dels treballs clàssics de terminologia. El «Vocabulari de preservació i conservació del patrimoni documental» de la profesora Maria Elvira de la UB és una d'aquestes tasques en què la informàtica només intervé com a eina auxiliar. És necessari que es continuïn fent treballs d'aquesta mena per a consolidar la terminologia en els diferents camps del coneixement i incorporar-hi els neologismes; treballs orientats, però, segons els objectius específics, entre els quals destaca el de constituir la base per a un millor tractament automatitzat de la informació i les bases de dades.

Pel balanç fet fins aquí, podem afirmar que les línies més evidents de continuïtat en la recerca pel que fa als temes tractats en aquesta Jornada són en el camp de la recuperació d'informació i en el desenvolupament i l'ús de les xarxes d'Internet, amb el maridatge que esmentàvem entre estadística, lingüística i informàtica.

I, quant a la terminologia, ens ha aparegut en aquesta Jornada com un fi en si mateix, pel valor que té a l'hora de facilitar la comunicació i de resoldre els problemes d'ús lingüístic que en deriven, en l'elaboració de diccionaris especialitzats i en el disseny i la utilització de multicercadors. Però, la presència de la terminologia, l'hem tinguda sobretot com a element auxiliar imprescindible en les operacions de cerca i recuperació d'informació, pròpies dels documentalistes i també dels tractadors.

Per a totes aquestes funcions, la terminologia ha de consolidar conceptes i modes de tractament, per a afinar la resposta a qüestions que directament o indirecta han sorgit en aquesta Jornada: el tractament dels termes que també són lèxic comú i les decisions consegüents sobre inclusió i contingut de les definicions; els

critèris per a la constitució de corpus especialitzats per a contrastar resultats estadístics amb els dels corpus de llengua comuna o general, amb fronteres que cal distingir de les fronteres temàtiques.

No falten, doncs, temes per a futures reflexions i debats en marcs acadèmics com el d'aquesta Jornada de «Terminologia i documentació».