

L'Arbre del Coneixement: un projecte de classificació i tesaurització de la terminologia en català

Marina SALSE I ROVIRA, Marga PRADAS I GARIJO

RESUM: *El projecte Arbre del Coneixement, dirigit i promogut pel Grup Enciclopèdia Catalana, neix el 1995 amb un doble propòsit: proporcionar una eina de revisió de la Gran enciclopèdia catalana (obra emblemàtica de l'editorial) i experimentar noves formes d'accés a la informació en el marc d'una societat cada vegada més oberta a les noves tecnologies i als sistemes hipertextuals. El projecte, que es considera finalitzat en el 80 %, organitza, seguint una classificació polijeràrquica, les gairebé 225.000 accepcions que constitueixen el corpus enciclopèdic de la Gran enciclopèdia catalana. En l'actualitat l'Arbre del Coneixement ha estat el suport d'una sèrie de realitzacions editorials encarades tant a l'interior de l'editorial (revisió de l'enciclopèdia, generació de diccionaris temàtics), com orientades als usuaris externs (Hiperenciclopèdia, Fototeca, creació de CD, etc.).*

1. NAIXEMENT DEL PROJECTE “ARBRE DEL CONEIXEMENT”

El projecte “Arbre del Coneixement” (d'ara endavant AC) va néixer a finals de 1994 i es va començar a desenvolupar a principis de 1995. Sorgia tant d'unes necessitats internes de l'editorial mateixa, com d'uns condicionaments externs.

Marina Salse i Rovira, Facultat de Biblioteconomia i Documentació de la Universitat de Barcelona, salse@fbd.ub.es.

Marga Pradas i Garijo, Enciclopèdia Catalana, mpradas@grec.com.

Article rebut el desembre de 2001.

Article basat en la comunicació presentada per les autores a les 8es Jornades Catalanes de Documentació, 15 i 16 de novembre de 2001, l'Hospitalet de Llobregat.

1.1 Els condicionaments interns

Les necessitats internes venien donades per la mateixa natura de les publicacions editorials de l'editorial Enciclopèdia Catalana, una part important de la producció de la qual era representada per les àrees de *diccionaris* i de *grans obres*. Dins d'aquestes hi ha la publicació emblemàtica que ha donat nom a la mateixa editorial: la *Gran enciclopèdia catalana* (d'ara endavant GEC).

La GEC plantejava ja en aquells moments una problemàtica d'actualització molt concreta a causa de la seva envergadura (24 volums i suplement) i al pas del temps. Aquests dos elements, el temps i el volum, provocaven en els col·laboradors que revisaven l'obra (especialment en matèries grans com la història, la física, la química o la literatura...):

- Dificultat d'obtenir una visió global de la forma en què la GEC tractava una matèria. Si pensem que, a més, l'obra havia estat reactualitzada sovint per persones diferents, trobarem que moltes vegades la visió global s'havia fragmentat en una disparitat d'enfocaments que calia tornar a reagrupar.
- Dificultat en detectar mancances. O sigui manca de nous conceptes o noves branques dins d'una àrea concreta de coneixement.
- Dificultat en detectar obsolescències: visions antiquades, articles desfats temporalment.
- Dificultat per coordinar les diferents matèries, tractades, com és lògic, per diferents col·laboradors.

Contra aquesta situació ja s'havien efectuat diverses accions a principis dels 90. D'entre les més importants destaca la creació de la base de dades de la GEC, la unitat documental de la qual es feia respecte a l'accepció. El sistema utilitzat per aquesta base de dades fou l'AS-400, d'IBM. Ultra diversos camps com ara la categoria gramatical, data de naixement, data de mort, etc., s'havia creat un petit thesaurus de matèries que permetia assignar cadascuna de les accepcions de la GEC a una matèria (o submatèria) concreta.

Tanmateix, amb això no n'hi havia prou ja que l'aproximació a l'obra per part dels col·laboradors continuava sent molt lineal, i l'agrupació per matèries, insuficient. Calia esquematitzar l'obra, proporcionar als col·laboradors un punt de vista jeràrquic que els permetés visualitzar en conjunt el tractament de la seva matèria de forma esquemàtica i senzilla i ajudés a un replantejament global de l'obra.

Abonava la generació d'un projecte similar l'existència dins de l'editorial d'una important àrea de diccionaris que es presentava com a potencial "con-

sumidora” interna de la feina desenvolupada per a la GEC. Els diccionaris treballen a partir d’arbres de camp que els permeten generar una primera estructuració de les entrades que han de contenir. La idea era, doncs, usar l’AC com a base dels arbres de camp dels nous diccionaris de l’Editorial.

Aquestes eren les bases teòriques que justificaven, doncs, l’existència d’un projecte d’aquestes característiques dins de l’editorial. Tanmateix, la inversió en temps i diners que calia era molt gran; s’ha de tenir present que calia mobilitzar centenars de milers d’entrades i destinar personal especialitzat tant a la confecció de l’arbre com a la seva revisió i “encaixament”

Els condicionaments externs varen decidir finalment inclinar la balança cap a la realització del projecte.

1.2 Els condicionaments externs

Els primers anys 90, època de gestació del projecte AC, són l’inici d’un període de forta transformació editorial. La informàtica està irrompent amb força a les llars, on comencen a proliferar els PC. Aquests PC comencen a incorporar en aquesta època lectors de cd-rom i en aquests lectors s’hi insereixen cd-roms que usen noves formes d’aproximar-se a la informació i permeten superar l’estricta linealitat en paper.

Molts lectors (i, en conseqüència, molts usuaris d’enciclopèdies) s’aproximen per primera vegada a un tipus d’accés a la informació que ja no està guiat per les lletres marcades al lloc dels volums (o sigui, per una aproximació alfabètica) i per les remissions entre entrades dins de cada volum. El concepte d’hipertext, tot i que ja ve d’antic s’aproxima als usuaris i es popularitza. Apareixen les primeres enciclopèdies en cd-rom.

Els usuaris comencen a demanar que se’ls proporcionin accessos alternatius a les enciclopèdies tradicionals, mentre que les editorials comencen a preguntar-se on quedarà el paper de les enciclopèdies en paper (valgui la redundància). Concepcions com les de Nicholas Negroponte afirmant per aquestes dates que hem de començar a moure bits i no àtoms¹ estan sacsejant els fonaments editorials tradicionals.

A més, i també en aquestes dates, Tim Berners-Lee ja ha creat el WWW i aquest comença a popularitzar-se. Internet comença a estendre els seus tentacles i s’endevina com un nou revulsiu contra les concepcions editorials tradicionals.

Les editorials no deixen de reaccionar contra la nova situació i la nova demanda. Dues grans editorials, Larousse i Encyclopaedia Britannica, llencen

1. Negroponte, Nicholas. *El mundo digital*. Barcelona: Ediciones B, 1995.

aproximacions jeràrquiques en paper a les seves enciclopèdies, tot ajustant-se a les demandes del mercat. El gruix de les editorials inaugura línies de treball per a produir cd-roms i per a resseguir el nou fenomen d'Internet.

Aquest món canviant, bàsicament per raons tecnològiques, va fer prendre al projecte AC una importància que depassava a bastament l'àmbit editorial intern, el qual, d'altra banda, sempre va justificar una obra d'aquesta mena. Havia de servir per a la millora de la qualitat del treball intern, però també havia de ser la base sobre la qual muntar una sèrie de productes externs adequats a les noves aproximacions a la informació que anaven (i van) sorgint sense aturar-se.

1.3 Objectius de l'Arbre del Coneixement

Després de fer aquesta panoràmica de la situació podem compendiar els objectius / finalitats de l'arbre en els següents punts:

- Actuar com a base de revisió de la GEC: per les reimpressions i futures edicions.
- Servir de base per a la confecció de diccionaris i altres productes enciclopèdics en paper: diccionaris temàtics i Enciclopèdia Temàtica.
- Servir de base per a la confecció de productes en línia i multimèdia: *Hiperenciclopèdia* i *Índex electrònic* de la GEC.
- Servir com a base per a la superació de la linealitat d'edicions impreses d'enciclopèdies: ofereix vies alternatives de consulta.
- Col·laborar a la fixació terminològica catalana dels diversos camps del coneixement.
- Aconseguir la tesaurització del coneixement en català.

Aquest darrer objectiu, tot i que apareix de bon principi, i que vindria a omplir un buit considerable dins de la cultura catalana, és encara el darrer i més poc desenvolupat dels que es va plantejar l'Editorial en un principi. No es plantejarà en la seva totalitat fins a la finalització definitiva del projecte.

2. DESCRIPCIÓ DEL PROJECTE ARBRE DEL CONEIXEMENT

El projecte, en línies generals, consisteix en una classificació polijeràrquica de les accepcions de la GEC (en l'actualitat unes 225.000).

2.1 Principis constructius

Els principis constructius de l'arbre, enunciats des dels seus inicis, foren els següents.

1. *Base constructiva*: entrades en vigència a la base de dades de la GEC amb matèria assignada. Per entendre la gran varietat d'entrades que inclou l'AC cal fer una referència a les quatre grans subbases que componen la base de dades automatitzada de la GEC:

Caixa baixa

- Entrades que pertanyen a la base de dades LÈXIC que inclou tot el corpus lexicogràfic de la GEC en qualsevol de les diferents categories gramaticals. No totes les accepcions d'aquest grup tenen assignada matèria, ja que es considera que molts termes són d'un lèxic massa comú com per assignar-los a una determinada matèria. Dins de l'AC només es va treballar amb els termes que duïen matèria assignada i, dins d'aquests, es va donar prioritat absoluta a la inclusió de substantius.

Tot i que en qualsevol tesaurus o llista de matèries el substantiu és la categoria gramatical emprada per excel·lència, a l'AC en molts casos s'han utilitzat altres categories gramaticals (adjectius, locucions, etc.) quan l'accepció era bàsica en la construcció de l'arbre i no existia un substantiu dins del corpus de l'obra que el pogués substituir. Per exemple, en el cas dels tipus de motors podem trobar alguns motors entrats dins la GEC pel concepte “motor” (motor pneumàtic), però altres són entrats pel seu adjectiu (com el “motor anaeròbic”, entrat pel concepte “anaeròbic”).

Caixa alta

- Entrades que pertanyen a la base de dades TOPONÍMIA: inclou llocs geogràfics d'arreu del món, amb especial atenció a tots aquells que pertanyen als Països Catalans. Constitueixen fins a l'actualitat un arbre separat de l'AC tot i que s'ha plantejat en alguns casos la seva fusió. Aquesta darrera tasca està pendent d'estudi i disseny per a la seva realització.

- Entrades que pertanyen a la base de dades BIOGRAFIES: inclou biografies de persones reals, ja que déus i éssers mitològics formen part de la base de dades GENERAL.
 - Entrades que pertanyen a la base de dades GENERAL: s'inclouen entrades de caixa alta que no es localitzen a les bases de toponímia i biografia. Inclou, per exemple, empreses, entitats, déus, periòdics, obres d'art o llibres.
2. *Format*: arbre polijeràrquic. Es parteix del fet que una mateixa entrada pot assignar-se a més d'una matèria (així, la dansa pertany a la vegada a les matèries de música i d'espectacles). Tanmateix, per qüestions de coherència en l'estructura només un dels col·laboradors treballa la matèria.
3. *Metodologia constructiva*: termes genèrics que inclouen termes específics, tot intentant reproduir l'estructura interna i subjacent de la GEC. No importa que aquesta estructura reveli errors, mancances o concepcions obsoletes. Ans al contrari, es tracta d'aconseguir que les detecti per tal de poder-les solucionar. La base de la metodologia constructiva són les definicions de cadascuna de les entrades de la GEC. Secundàriament es pot acudir als desenvolupaments enciclopèdics per esclarir dubtes.
4. *Metodologia de revisió*: Una de les finalitats constructives de l'arbre és anar detectant mancances i obsolescències dins de cadascuna de les grans àrees de coneixement. Aquesta ha estat una feina principal dels col·laboradors de cadascuna de les branques de l'arbre. Es volien detectar, entre altres:
- Mancances d'entrades.
 - Mancances de grups d'entrades pertanyents a algun nou àmbit científicotècnic.
 - Entrades obsoletes.
 - Definicions obsoletes.
 - Estructuracions anacròniques de determinades àrees del coneixement.
5. *Nombre desitjat de fills per pare*: 10/15. Un dels defectes més importants detectats en productes similars fou que molts d'ells es limitaven a fer aproximacions de molt pocs nivells al corpus enciclopèdic, de forma que per sota dels pares quedaven sovint centenars d'entrades. Això documentalment és un error perquè atabala l'usuari i fa inútil la cerca per aquella matèria en concret. Dins de l'AC es varen voler establir nivells

amb prou pocs fills com per no produir soroll en els possibles usuaris. Això no vol dir, evidentment, que no hi hagi nivells amb més de quinze fills (no es poden definir divisions coherents per les 24 lletres de l'al·fabet) però l'intent d'evitar-ho ha estat una constant de tota l'obra.

2.2 L'estructura de capçalera (els nivells superiors de l'arbre)

Establerts els principis constructius de l'arbre, el següent pas va ser generar l'estructura de capçalera de l'arbre. És a dir, establir els nivells superiors a partir dels quals començar la feina d'estructuració. Per poder establir aquests nivells es van tenir en compte diverses fonts:

- El petit tesaurus de matèries de la GEC. Molt important, ja que era la base de la distribució d'entrades entre els col·laboradors.
- La importància donada per la pròpia GEC a les branques principals del coneixement. Es va fer una anàlisi exhaustiva del que es consideraven branques principals.
- Productes documentals de classificació i indexació: el tesaurus de la Unesco, la CDU o les llistes d'encapçalaments de matèria en català, entre molts d'altres.
- Productes similars de l'àmbit editorial.
- Propostes d'especialistes o de grups d'especialistes dins de determinades àrees del coneixement.

Com a resultat de nombroses discussions s'establiren els següents nivells de treball:²

BELLES ARTS

Arts menors
Arts plàstiques
Espectacles
Música
Restauració

CIÈNCIES SOCIALS

Antropologia
Biblioteconomia, Documentació i Arxivística

2. En rodona són accepcions de la GEC, en cursiva apareixen les facetes.

Ciències de l'educació

Dret

Economia

Política

Psicologia

Sexologia

Sociologia

CIÈNCIA

Ciències de la terra

Geografia

Geologia

Climatologia i meteorologia

Topografia i cartografia

Ciències de l'espai

Astronomia

Astronàutica

Ciències de la vida

Biologia

Botànica

Farmàcia

Medicina

Zoologia

Física

Matemàtiques

Química

FILOSOFIA

HISTÒRIA

Ciències auxiliars de la història

Arqueologia

Heràldica

Genealogia

Numismàtica

Escriptures, textos i documents antics

LINGÜÍSTICA I LITERATURA

Lingüística

Literatura

LLEURE

Caça

Muntanyisme

Pesca (esportiva)

Tauromàquia

Joc

Esport

Turisme

RELIGIÓ

TECNOLOGIA

Informàtica

Mineria

Ciència dels materials

Tecnologia militar

Tecnologia paperera i arts gràfiques

Electrònica i electrotècnia

Tecnologia de materials

Mecànica, màquines i fabricació

Automatització i control

Enginyeria civil

Arquitectura, urbanisme i construcció

Transport i tecnologia marítima i aeroespacial

Tecnologies de la informació i de les comunicacions

Agricultura, ramaderia, pesca i alimentació

Tecnologia tèxtil i de la indumentària

Tecnologia energètica

2.3 L'estructuració interna de l'Arbre del Coneixement

A diferència de projectes similars, l'AC té un nivell molt elevat d'estructuració. En l'apartat de bases constructives ja s'ha especificat quina és la filosofia de creació de l'arbre. En aquest apartat definirem els elements constructius que fan possible la seva creació i el seu alt nivell d'estructuració interna.

a) Els nodes

Un dels conceptes estructurals importants és el de *node*, que podem definir com qualsevol embrancament de l'arbre. El concepte de node és, doncs, molt ampli de forma que poden ser nodes:

— Les *entrades de la GEC*, amb preferència sobre qualsevol altre tipus de node. Sempre que hi hagi una entrada de la GEC s'ha evitat posar qualsevol altre tipus de node. Dins de les entrades de la GEC es poden distingir tres categories:

1. Entrades amb una única ubicació
2. Entrades amb múltiples ubicacions, les quals es consideren adequades per a diferents àmbits del coneixement o bé ubicables diverses vegades en un únic àmbit del coneixement. En aquests casos es poden donar tres tipus de nodes
 - a. *Nodes de salt*: quan un concepte apareix en dos arbres que han de ser desenvolupats per col·laboradors diferents. El coordinador estableix abans de lliurar els arbres als especialistes si hi ha conceptes que es trobin en aquest cas i decideix en quin dels arbres es farà el desenvolupament (un cas molt clar és el de la branca del coneixement anomenada *química física*). En l'arbre on el concepte no es desenvoluparà es col·loca un node de salt que indica al col·laborador que aquell concepte es desenvolupa en un altre arbre. Dels nodes de salt no hi poden penjar fills.
 - b. *Nodes principals* i *nodes secundaris*: quan un concepte pot ser posat més d'una vegada en un mateix arbre. Per garantir la coherència en el desenvolupament només es podrà penjar fills del node principal. Cal notar que, en l'arbre visualitzat en el seu conjunt els nodes de salt abans esmentats desapareixen per convertir-se en entrades principals i secundàries.
3. Les remissions: dins de qualsevol enciclopèdia en paper s'inclouen referències de “vegeu” (les de “vegeu també” existeixen, però no s'han considerat en l'arbre). El tractament d'aquestes remissions / referències ha generat un tipus concret de nodes sense fills que envia directament a l'entrada remesa / referenciada. Una de les funcions principals dels col·laboradors ha estat identificar remissions i enllaçar-les amb l'entrada principal.

— Les *facetes*: agrupaments d'entrades que no disposen ni es preveu que hagin de disposar de cap entrada a la GEC. Per exemple “Tipus de vehicles” (dins el subarbre *transport*). Les facetes han estat la forma d'agrupar entrades que, d'altra manera penjarien en llistes massa llargues

del node principal. També han estat formes d'agrupar àrees de coneixement en els nivells superiors, com es pot veure en l'apartat anterior, en el qual s'han assenyalat en cursiva totes les facetes.

- Els *nodes provisionals* i *neologismes* (que indiquen mancances i propostes de noves entrades en algun punt d'un subarbre o d'un arbre general). Els *nodes provisionals* s'usaven en treballar sobre arbres parcials i eren una eina dels col·laboradors per indicar que allà consideraven que hi havia d'haver una entrada. Els *neologismes* només s'usen en l'arbre complet i solen ser la conseqüència lògica quan hi ha nodes provisionals que s'ha comprovat que no figuren a la GEC. Són una proposta que l'equip de redacció de la GEC ha de valorar i, si s'escau, redactar i incorporar al corpus de l'obra.

Sobre qualsevol dels tipus de nodes (llevat de les facetes) els col·laboradors o la coordinació del sistema poden afegir comentaris que ajudin a comprendre les decisions adoptades o que permetin recollir propostes de millora de la GEC.

b) Polisèmies i homonímies

Dins de qualsevol llengua, i, en conseqüència, dins de la GEC, com a reflex d'un llenguatge, la polisèmia i l'homonímia (igualtat de forma i diversos significats) són elements comuns.

Normalment la pròpia GEC ja intenta solucionar molts d'aquests conflictes per mitjà del concepte d'*accepció*, que permet distingir tipus de conceptes que en principi tindrien igual forma. Per exemple les accepcions s'usarien per distingir diversos tipus de màquines, encara que el terme comú a totes les màquines sigui "màquina." Per exemple:

Màquina (pel concepte general de màquina)

Màquina (màquina hidràulica)

Màquina (màquina pneumàtica)

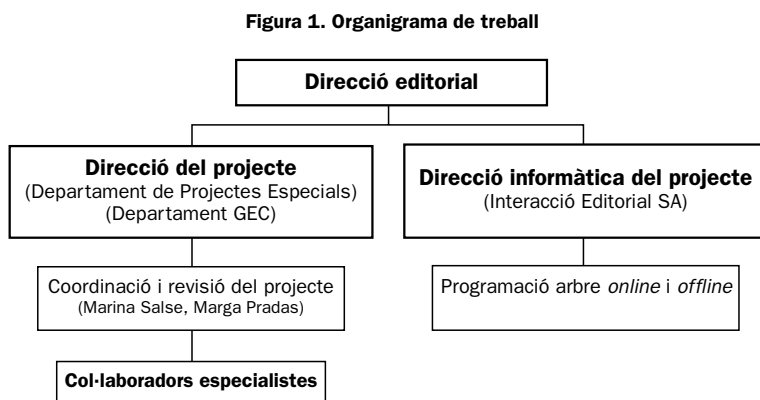
Tot i això, hi ha homonímia que només es diferencia per la definició. Per exemple, hi ha dues entrades "carnívor" diferents, una fa referència a l'animal i l'altra a l'ordre de l'animal, però no s'han posat accepcions per distingir-les. El problema es féu evident, quan, com en aquest cas, en l'arbre alguns d'aquests conceptes queden al mateix nivell (germans) o bé tenen relacions familiars directes (pare-fill, avi-nét). En aquests casos, per esclarir el contingut de cadascun dels nodes es col·loca un comentari entre parèntesis. Aquests comentaris es generen intentant seguir les mateixes regles de construcció d'accepcions de la GEC. Així, en l'exemple anterior:

Carnívor (ordre dels carnívors)

Carnívor (animal carnívor)

3. L'ORGANITZACIÓ DEL TREBALL

En la realització d'aquesta tasca s'han involucrat diferents departaments i àrees de l'editorial. L'organigrama de treball es presenta en el següent esquema:



a) Els especialistes

Des d'un primer moment el plantejament va ser el d'oferir la realització de la feina a un grup d'especialistes per a cadascuna de les àrees del coneixement. Cal assenyalar que fou difícil trobar especialistes amb nocions de documentació i, quan no fou possible localitzar-los, es va dur a terme una tasca de formació en aquest sentit dels especialistes.

A aquests especialistes se'ls proporcionava un llistat d'entrades extretes de la base de dades GEC amb matèries assignades que es consideraven adequades per a aquell àmbit del coneixement, així com un programari de tractament fet a mida (vegeu l'apartat corresponent).

La missió dels especialistes era:

- Generar l'arbre de la seva matèria segons els principis constructius enunciats i amb l'ajut dels elements constructius proporcionats.
- Eliminar de la llista aquelles entrades que no fossin pertinents (per exemple adjectius que ja tinguessin un substantiu).
- Rebutjar de la llista aquelles entrades que no fossin de la seva àrea del coneixement.

- Detectar mancances i obsolescències, tot proposant d'afegir entrades, treure entrades o modificar llur definició.
- Enllaçar les remissions amb les entrades remeses.
- Proposar noves interpretacions del gruix de la matèria (propostes de reclassificació).

Els especialistes lliuraven la feina en dues fases. En la primera fase elaboraven una proposta de classificació que lliuraven a l'equip de coordinació. Aquest equip revisava el que s'havia fet i ho retornava al col·laborador amb els comentaris i mancances documentals que es detectaven. En la segona fase el col·laborador lliurava l'arbre estructurat, i la coordinació s'encarregava de fer una validació i incorporació d'aquest arbre parcial a l'arbre principal.

La realització de qualsevol branca de l'AC sovint ha significat diverses interaccions entre els col·laboradors i la coordinació.

b) La coordinació

La coordinació ha estat durant aquests sis anys portada per dues persones: Marina Salse i Marga Pradas, autores d'aquest article. Les feines realitzades per l'equip de coordinació es poden resumir com segueix:

- Formar els col·laboradors i resoldre els seus dubtes.
- Revisar els arbres parcials en les seves diverses fases.
- Enllaçar els arbres parcials amb l'arbre principal.
- Reencaminar entrades rebutjades cap a d'altres matèries.
- Vetllar pel compliment dels principis constructius.
- Redirigir obsolescències i mancances cap als departaments corresponents.
- Assessorar altres departaments en la possible utilitat de l'AC per als seus fins.

c) La direcció

En un origen, el projecte AC va ser atribuït a un nou departament que estava naixent en aquells moments: el departament de Projectes Especials dirigit per Joan Antoni Afonso que es dedicava en certa forma a explorar com les noves tecnologies podien generar nous productes editorials. Aquest departament ha dut el projecte AC fins l'any 2001, any en què, pràcticament acabat, ha passat al departament que realitza la GEC, dirigit per Josep Maria

Vinyes, on es continuarà amb el manteniment i adequació de l'AC a les necessitats editorials.

Les funcions dels caps de departament han estat, entre altres:

- Definir les línies generals del projecte (en col·laboració amb l'equip de coordinació).
- Posar en marxa el projecte, reunint tots els departaments i persones implicades dins de l'editorial i coordinant les seves actuacions. Especialment important va ser en aquest sentit la tasca de coordinació amb l'Àrea d'Informàtica de l'editorial per a la producció del programari.
- Elaborar calendaris i controlar la realització de les diverses fases del projecte.
- Controlar tot l'aspecte econòmic i pressupostari.
- Fomentar la continuïtat del projecte i la seva utilització pels diversos sectors editorials.

Per damunt d'aquests dos departaments la Direcció Editorial, representada per Jesús Giralt, s'ha encarregat de marcar les línies generals de desenvolupament del projecte, de controlar la seva elaboració i de proposar nous camps d'actuació.

d) La direcció informàtica

La direcció informàtica ha estat portada a terme per l'empresa Interacció Editorial SL, pertanyent al Grup Enciclopèdia Catalana. Aquesta empresa ha desenvolupat els dos programes principals que han permès la creació de l'arbre i tot un seguit d'aplicacions adjacents que han permès la comunicació entre ambdues aplicacions principals.

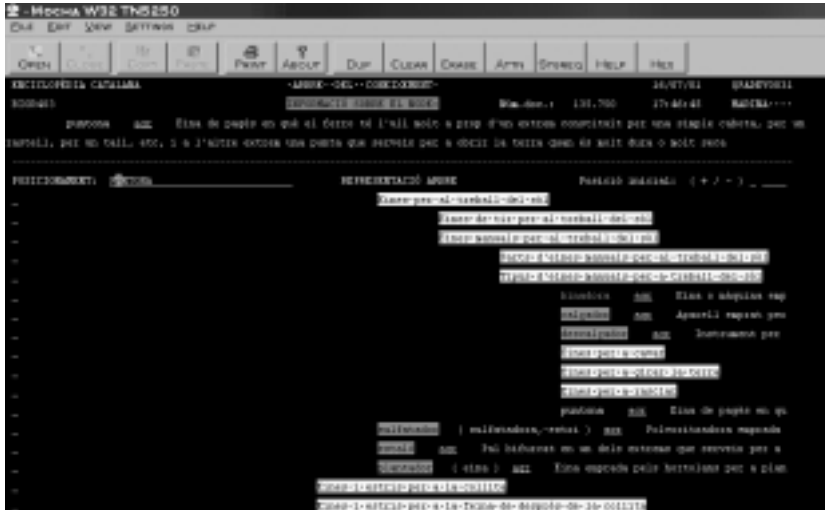
4. EL SUPORT INFORMÀTIC

S'ha treballat tot usant dos programes diferenciats: l'arbre *online* i l'arbre *offline*

a) L'arbre online

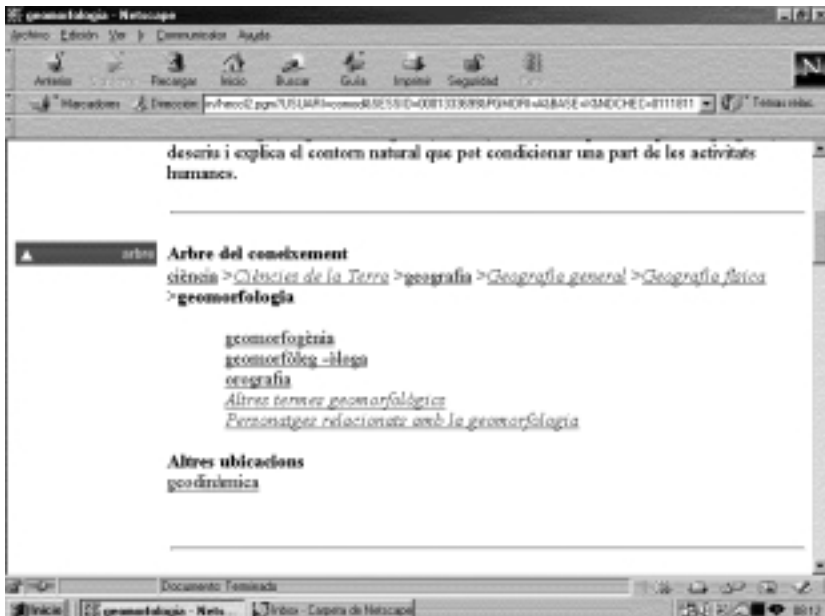
Es coneix amb aquest nom l'aplicació que es va construir tot programant directament l'ordinador central d'Enciclopèdia Catalana SA, i usant com a suport les bases de dades documentals que contenen les entrades de la GEC.

Figura 2. Aspecte de l'arbre *online* en un dels arbres que ja es consideren acabats (secció de l'arbre d'agricultura)



Aquest arbre reuneix el conjunt de l'AC i és aquí on s'han anat bolcant els arbres parcials a mesura que els col·laboradors els lliuraven i a mesura que eren validats per l'equip de coordinació. Aquí es fan també els ajustaments finals i des d'aquí es porta l'actualització de l'arbre.

Figura 3. L'arbre adaptat a la Hiperenciclopèdia. És consultable per Internet.

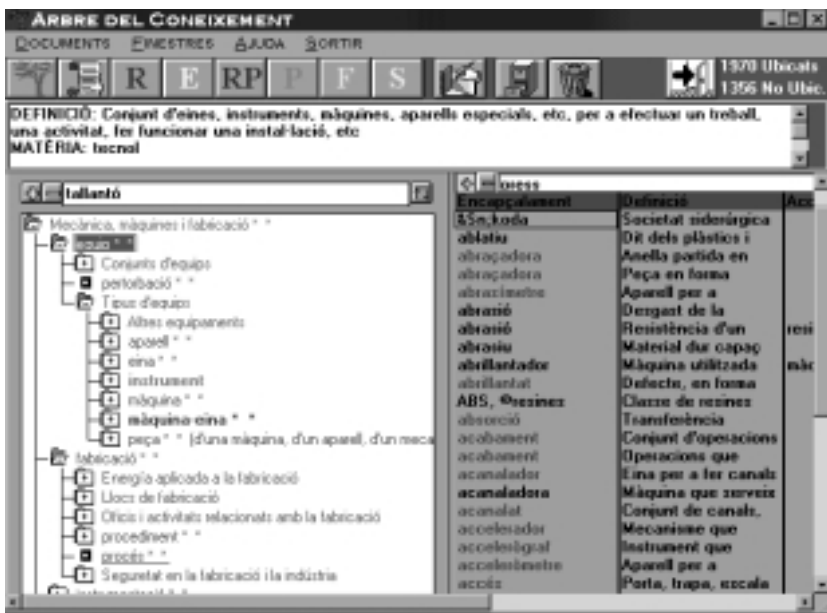


També és aquest arbre la font principal d'on es generen productes informàtics derivats, com l'arbre que apareix a la Hiperenciclopèdia (consultable en línia a <http://www.grec.net>, tot i que no és de consulta lliure), del qual en teniu aquí una mostra.

b) L'arbre offline

L'arbre *offline* és una aplicació creada en Visual Basic sobre Access i va ser generada com l'eina que es proporcionava directament als col·laboradors per a ajudar-los en la confecció de l'arbre.

Figura 4. Aspecte d'un dels arbres que encara estan en construcció.
Els diversos colors assenyalen els diversos tipus de nodes: verd (faceta), blau (entrada de la GEC), vermell (node provisional)



Es va generar una aplicació a mida de l'AC, ja que no hi havia cap aplicació al mercat que reunís les característiques adequades per al tractament de la informació tal com era necessària a la creació de l'arbre.

L'aplicació va ser dissenyada per suportar la creació d'arbres parcials de fins a 10.000 entrades (d'acord amb els principis i elements constructius ja enunciats en apartats anteriors), i en els seus orígens corria (dificultosamente) sota ordinadors 386, encara que es demanava un 486 per no tenir problemes de funcionament (de fet en les primeres fases del projecte vàrem haver de re-

butjar bastants col·laboradors per no tenir equipament informàtic o per tenir-lo massa antic).

c) *Les aplicacions auxiliars*

Complementàries a aquestes dues aplicacions s'han anat generant un bon nombre d'aplicacions auxiliars, encarades a:

- Treure entrades de l'AS/400 per generar arbres parcials.
- Traspasar entrades entre arbres parcials.
- Preparar els arbres parcials per entrar al sistema AS/400 (i, en conseqüència, a l'arbre *online*).
- Assegurar la transmissió sense incoherències entre *offline* i *online*.

5. **ESTAT ACTUAL DE LA QÜESTIÓ I CONCLUSIONS**

Sis anys més tard el projecte AC, amb un 80 % de les entrades ubicades és ja una realitat. També han començat a prendre forma algunes de les utilitats que es varen plantejar el 1995, mentre que d'altres han sorgit de nou encuny vinculades a Internet i a les tecnologies associades.

Entre les realitzacions ja fetes podem assenyalar les següents

1. El projecte AC ha estat el suport de productes hipertextuals/hipermèdia associats a les noves tecnologies. Alguns dels desenvolupats fins ara (que incorporen versions parcials de l'arbre) són la Hiperenciclopèdia o la Fototeca.
2. L'AC és una eina cada vegada més consolidada per a la revisió i actualització de la GEC, així com la base sobre la qual s'han d'elaborar futures edicions completes. Aquest propòsit es basa en el fet que per al col·laborador és més senzill detectar errors i mancances en una estructura arbòria que en un llistat alfabètic de documents afins.
3. L'AC ha estat el suport per a la generació d'arbres de camp per a diversos diccionaris temàtics d'àmbit català.

En l'actualitat, la construcció de l'AC és pràcticament finalitzada, tot i que som conscients que una obra d'aquestes característiques no pot acabar-se mai, ja que el flux d'entrada i sortida de les accepcions és constant.

6. BIBLIOGRAFIA

En tractar-se d'un tipus de tasca inèdita pel seu abast no hem localitzat bibliografia relativa al tema. Per tant, la nostra base i punt de partida han estat els manuals de classificació més habituals en les nostres facultats, així com la consulta a alguns tesaurus generals i especialitzats, com també a diversos diccionaris i enciclopèdies. Per citar-ne alguns:

CDU: Clasificación Decimal Universal. Ed. abreviada española, 6a ed. rev. y actualizada. Madrid: AENOR, 1991. 462 p.

Gran enciclopèdia catalana. 2a ed. Barcelona: Enciclopèdia Catalana, 1986-. 26 v.

Gran Larousse català: consultor temàtic universal. Barcelona: Edicions 62, 1996. 2v.

INSPEC Thesaurus 1989. London: The Institution of Electrical Engineers, 1989. 584 p.

Llista d'encapçalaments de matèria en català. Ed. preliminar actualitzada. Barcelona: Generalitat de Catalunya. Departament de Cultura. Institut Català de Bibliografia, 1991. 512 p.

The new encyclopaedia britannica. 15th ed. Chicago [Ill.] [etc.]: Encyclopaedia Britannica, cop. 1993. 32 v.

Tesouro de la Unesco: lista estructurada de descriptores para la indización y la recuperación bibliográficas en las esferas de la educación, la ciencia, las ciencias sociales, la cultura y la comunicación. Preparado por Jean Aitchison; traducción de Jean Viet. París: Unesco, 1984. 2 v.