

TAXONOMIES: QUÈ SÓN, PER A QUÈ SERVEIXEN, I COM ES CONSTRUEIXEN

Miquel Centelles

Departament de Biblioteconomia i Documentació
Universitat de Barcelona
miquel.centelles@ub.edu

Resum

Es defineix el concepte de taxonomia en el marc dels processos de revisió de les normes nacionals i internacionals de construcció de vocabularis controlats, i, fonamentalment, de la norma ANSI/NISO Z39.19-2005. Es presenten els aspectes més rellevants que aquesta darrera norma ha introduït en la construcció i aplicació de les taxonomies, i que tenen a veure amb els mètodes i les fonts per a la selecció de termes i per a l'establiment de relacions entre els termes. Finalment, s'analitzen les oportunitats i les amenaces que es presenten per a les taxonomies com a model de representació i recuperació del coneixement en l'entorn web, tenint en compte els models generats pel paradigma del Web 2.0.

Paraules clau

taxonomies, Web 2.0, sistemes de representació del coneixement

Resumen

Se define el concepto de taxonomia en el marco de los procesos de revisión de las normas nacionales e internacionales de construcción de vocabularios controlados, y, fundamentalmente, de la norma ANSI/NISO Z39.19-2005. Se presentan los aspectos más relevantes que esta última norma ha introducido en la construcción y aplicación de las taxonomías, y que tienen que ver con los métodos y las fuentes para la selección de términos y para el establecimiento de relaciones entre los términos. Finalmente se analizan las oportunidades y las amenazas que se presentan para las taxonomías como modelo de representación y recuperación del conocimiento en el entorno web, teniendo en cuenta los modelos generados por el paradigma de la Web 2.0.

Palabras clave

taxonomías, Web 2.0, sistemas de representación del conocimiento

Abstract

The concept of taxonomies is defined within the framework of revisions of national and international standards for constructing controlled vocabulary, especially the standard ANSI/NISO Z39.19-2005. The most important aspects introduced by this standard for constructing and applying taxonomies deal with the methods and sources for the selection of terms and for establishing relations between terms. Finally an analysis is provided of the opportunities and threats for taxonomies as models for representing and retrieving knowledge on the web, with special consideration of models generated by the Web 2.0 paradigm.

Key words

Taxonomies, Web 2.0, knowledge representation systems

1. Però... què és una taxonomia?

A l'article Centelles (2005) s'identifiquen i s'analitzen 36 definicions diferents del terme taxonomia proposades durant el període 2002-2005 en diferents tipus de fonts d'informació: diccionaris i enciclopèdies especialitzades, articles de consultors en gestió de continguts i arquitectura de la informació, articles d'especialistes vinculats a l'àmbit acadèmic, i documents tècnics i comercials d'aplicacions informàtiques per al desenvolupament de taxonomies. En aquell moment encara no disposàvem d'una definició normativa: la revisió de la norma ANSI/NISO Z39.19-1993, *Guidelines for the construction, format, and management of monolingual thesauri* es trobava en plena fase de discussió, i encara no s'havia generat una versió provisional completa. El 5 de juliol de 2005, s'aprova la norma ANSI/NISO Z39.19-2005, *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Com es pot observar en el títol, un resultat fonamental de la revisió és que no limita l'abast de les seves directrius i recomanacions als tesaurs, sinó que l'amplia a altres tres tipus de vocabularis controlats: les llistes de termes, els anells de sinònims, i les taxonomies. En l'apartat «Definitions, abbreviations, and acronyms», s'inclou la següent definició «normativa» del terme taxonomia:

A collection of controlled vocabulary terms organized into a hierarchical structure. Each term in a taxonomy is in one or more parent/child (broader/narrower) relationships to other terms in the taxonomy.

L'abast de la revisió de la norma Z39.19 iniciada el 2003 és la més profunda de les que s'han portat a terme des de la seva primera versió, publicada el 1974, seguint els cicles regulars de revisió quinquennals. La norma va nàixer per donar resposta a les necessitats d'un entorn de la informació caracteritzat pel predomini de sistemes d'informació creats i, fonamentalment, utilitzats per bibliotecaris, indexadors i altres professionals de la informació que incorporaven i indexaven documents analògics. A finals del segle xx, el nou entorn de la informació havia canviat profundament. El divorci entre el nou entorn de la informació i la norma per a la construcció de tesaurs es posa de relleu en els resultats de l'enquesta sobre l'ús de la norma Z39.19 que va ser portada a terme per NISO el febrer de 2003 entre els membres de diferents comunitats virtuals.

Altres organismes de normalització estan desenvolupant processos de revisió de normes de tesaurs en la mateixa línia que l'ANSI; tot i que encara no disposem de resultats tangibles d'aquests processos.

- Pel que fa a ISO, i d'acord amb la National Information Standards Organization (2005: viii),¹ s'està portant a terme un procés de revisió similar a la norma ISO 2788:1986, *Guidelines for the establishment and development of monolingual thesauri*. No obstant això, al lloc web oficial del Comitè Tècnic d'ISO responsable d'aquesta norma —el TC46/SC9²— no es fa cap referència a aquesta revisió.
- El British Standards Institute (BSI) està portant a terme la revisió de la norma BS 5723:1987, *Guide to establishment and development of monolingual thesauri*, que ha esdevingut BS 8723:2005, *Structured vocabularies for information retrieval: guide* organitzada en cinc

1. «Similar revision work is underway internationally on ISO 2788, Guidelines for the establishment and development of monolingual thesauri.»

2. ISO TC 46/SC 9: *Information and Documentation - Identification and Description* [en línia]. Ottawa: Library and Archives Canada, Last update: 2006-09-09. <<http://www.collectionscanada.ca/iso/tc46sc9/index.htm>>. [Consulta: 15/12/2006].

parts diferents: «Part 1: Definitions, symbols and abbreviations»; «Part 2: Thesauri»; «Part 3: Vocabularies other than thesauri»; «Part 4: Interoperation between multiple vocabularies»; «Part 5: Interoperation between vocabularies and other components of information storage and retrieval systems». De moment, se n'han publicat les dues primeres parts. La tercera part, ha d'incorporar les directrius i recomanacions relatives a les taxonomies.

La norma Z39.19-2005 defineix les taxonomies a partir de tres característiques fonamentals:

- Una taxonomia és un tipus de vocabulari controlat.
- Una taxonomia incorpora les relacions d'equivalència i de jerarquia.
- L'àmbit preferent d'aplicació d'una taxonomia és l'exploració i navegació dels continguts dels llocs web.

A continuació analitzarem cada una d'aquestes tres característiques.

a) Una taxonomia és un tipus de vocabulari controlat

Un vocabulari controlat és una llista de termes seleccionats per una autoritat determinada i que disposen d'una definició no ambigua i no redundant. Per assolir aquestes dues condicions, s'han d'aplicar, com a mínim, dues regles:

1. En el cas que un mateix terme s'utilitzi per designar diferents conceptes, aquest terme ha de ser explícitament qualificat per resoldre l'ambigüitat.
2. En el cas que diferents termes s'utilitzin per designar el mateix concepte, un dels termes ha de ser identificat com a terme preferent en el context del vocabulari controlat i la resta ha de ser incorporats com a sinònims no preferents.

Per controlar el vocabulari cal aplicar tres mètodes: definir l'abast o significat dels termes; utilitzar la relació d'equivalència per enllaçar termes sinònims i quasi-sinònims; distingir entre homògrafs.

b) Una taxonomia incorpora les relacions d'equivalència i de jerarquia

Hi ha tres tipus de relacions semàntiques que poden connectar els termes d'un vocabulari controlat: la relació d'equivalència, la relació de jerarquia i la relació associativa.

La relació d'equivalència connecta els termes que, en el marc del vocabulari controlat, designen un mateix concepte. Els termes equivalents poden ser sinònims, variants lèxiques, etc. D'acord amb la norma Z39.19-2005, la relació d'equivalència permet assolir dos dels objectius del control del vocabulari: l'eliminació de l'ambigüitat i el control de la sinonímia. És la relació que, en el context d'un vocabulari controlat determinat, pot connectar *Logística industrial*, *Logística de mercaderies* i *Logística empresarial*.

La relació de jerarquia, que només es dona entre termes preferents, es basa en graus o nivells de superordinació o subordinació, on el terme superordinat representa una classe o un tot, i els termes subordinats fan referència als seus membres o parts. Aquesta és l'única relació semàntica que afecta tots els termes preferents del vocabulari controlat i, per tant, és la relació que defineix la seva macroestructura.

Concebuda d'aquesta forma, la relació de jerarquia presenta tres situacions diferents des del punt de vista lògic i mútuament excloents:

- La relació genericoespecífica s'anomena també relació *és-un* i s'estableix entre una classe i els seus membres o espècies. És la relació que connecta *Electrodomèstics* (classe) amb *Nevera*, *Congelador*, *Rentadora*, *Màquina de rentar plats*, *Assecadora*, *Cuina*, *Forn*, etc. (membres de la classe).
- La relació partitiva s'anomena també relació *part-tot* o *d'inclusió*, i s'estableix quan la classe correspon al tot i els seus membres en són les parts. És la relació que connecta *Automòbil* (tot) amb *Motor*, *Suspensió*, *Direcció*, *Transmissió*, *Carrosseria*, *Circuit elèctric*, *Frens*, *Accessoris* (parts).
- La relació d'ocurrència s'estableix entre una classe general expressada amb nom comú i una ocurrència individual d'aquesta classe expressada amb un nom propi. És la relació que connecta *Mar* amb *Mediterrània*, *Egeu*, etc.

Cal dir que, estrictament parlant, la relació partitiva no és una relació jeràrquica. Entre altres requeriments, la relació jeràrquica exigeix que un terme subordinat hereti els atributs del terme superordinat, i que incorpori, com a mínim, un nou atribut. Així, totes característiques de *Mamífers*, ho han de ser també de *Gossos*; a més, *Gossos* ha d'afegir característiques que no tenia *Mamífers*. En canvi, el terme *Seient* no hereta totes les característiques de *Bicicletes*.

La relació d'associació, que només es dona entre termes preferents, connecta termes que designen conceptes que, encara que no estan lligats entre si per una equivalència semàntica o una jerarquia, són susceptibles d'evocar-se mútuament, per associació d'idees. Hi ha diferents tipus de relacions associatives:

- disciplina i objecte d'estudi, per exemple: *dermatologia* ? *pell*
- entitat i material, per exemple: *espelmes* ? *cera*
- causa efecte, per exemple: *pesticides* ? *contaminació agrícola*
- etc.

Els quatre tipus de vocabularis controlats que identifica la norma Z39.19-2005, es diferencien, fonamentalment, pels tipus de relacions semàntiques que s'estableixen entre els termes: les llistes de termes i els anells de sinònims incorporen únicament la relació d'equivalència; els tesaurus incorporen les relacions d'equivalència, de jerarquia i d'associació. Les taxonomies incorporen les relacions d'equivalència i de jerarquia.

c) L'àmbit preferent d'aplicació d'una taxonomia és l'exploració (browsing) i la navegació dels continguts dels llocs web

La definició del terme *taxonomia* que inclou la norma Z39.19-2005 no vincula aquest tipus de vocabulari controlat amb una aplicació preferent. No obstant això, l'apartat «B.3. Typical applications of Lists, Synonym Rings, Taxonomies, and Thesauri» indica que:

Taxonomies are often created and used in indexing applications and for Web navigation. Because of their (usually simple) hierarchical structure, they are effective at leading users to the most specific terms available in a particular domain.

Aquest tret definitori de les taxonomies és corroborat per les definicions que presenten altres fonts no normatives, com s'indica a Centelles (2005):

Las definiciones que vinculan las taxonomías al entorno digital destacan, como finalidades prioritarias, la mejora de la navegación y el desarrollo de sistemas de búsqueda basados en la exploración («browsing») y en la recuperación («searching»).

D'acord amb Garrett (2002), l'arquitectura de la informació ocupa una posició central en el disseny web centrat en l'usuari, i la defineix com el disseny estructural de l'espai d'informació que facilita l'accés intuïtiu al contingut. Segons Montero i Peña (2005) les taxonomies desenvolupen un paper clau en aquest procés, ja que permeten l'organització jeràrquica del conjunt de categories sota les quals es poden classificar les unitats de contingut.

2. Aspectes rellevants en la construcció i de l'aplicació de les taxonomies

2.1. Introducció

En l'apartat 3, presentarem les oportunitats i amenaces amb què es troben les taxonomies, com a sistemes d'organització del coneixement, en un moment on coincideixen els beneficis que aporta la revisió de la norma de vocabularis controlats, i el nou paradigma general d'Internet (Web 2.0). La lliçó que aprendrem és que hi ha entorns on resulta adient la construcció i aplicació de vocabularis controlats —i, més concretament, taxonomies—, i altres entorns on és més adient optar per sistemes d'organització del coneixement no jeràrquics, descentralitzats i estàtics. Els entorns web on és recomanable l'ús de vocabularis controlats són diversos pel que fa a destinataris (llocs públics, intranets, extranets), models de negoci, característiques dels equips que els gestionen, etc. Aquesta diversitat ha motivat l'existència de diferents models de construcció dels vocabularis controlats, que divergeixen en les fonts utilitzades per a la recopilació de termes (representació dels conceptes, selecció de termes, determinació de les formes preferents, etc.) i l'establiment de relacions entre termes. La revisió de la norma ANSI/NISO 39.19 del 2005 es fa ressò, parcialment, d'aquesta diversitat.

Segui quin sigui l'entorn web a què s'aplicarà, el procés de construcció d'una taxonomia s'inicia amb la planificació estratègica, un procés que es basa en l'anàlisi de tres dimensions de la corporació que crea la taxonomia, i del lloc web on s'aplicarà: *el context*, *l'audiència*, i *els continguts*.

- L'objectiu de l'anàlisi del context és definir l'entorn corporatiu i el lloc web. De la corporació ens interessa conèixer el tipus, l'àrea d'activitat, la posició estratègica, les dimensions, l'organigrama i l'estructura de funcionament, etc. Del lloc web, ens interessa conèixer la tipologia, la posició estratègica, les dimensions, aspectes organitzatius i tecnològics de la seva creació i manteniment, etc.
- L'objectiu de l'anàlisi de l'audiència és identificar els grups d'usuaris del lloc web (interns i externs), i la caracterització dels grups pel que fa a necessitats, comportaments, i habilitats informatives.
- Finalment, l'objectiu de l'anàlisi dels continguts és identificar els tipus de continguts del lloc web, inventariar-los i descriure'ls a partir d'aspectes com el context de creació i manteniment, funcions, audiència, formats, accessibilitat, etc.

La planificació estratègica es concreta en un informe de síntesi on conflueixen aquestes tres anàlisis, i on es defineix la taxonomia que cal dissenyar en aspectes com l'òrgan responsable de la creació i el manteniment; l'abast i les dimensions; els usos i les funcions; etc. Aquest informe s'ha de completar amb un estudi de la situació de partida; és a dir, investigar si hi ha vocabularis controlats en la corporació i, en cas afirmatiu, quants n'hi ha, de quin tipus són, amb quina finalitat s'utilitzen i qui els utilitza, etc.

Cada una de les tres anàlisis exigeix la recopilació i estudi de fonts d'informació apropiades.

- En el cas de l'anàlisi del context, són especialment útils fonts documentals com les memòries o informes d'activitat, els organigrames i catàlegs de llocs de treball amb descripció de funcions, els protocols d'actuació i els manuals de procediments; i fonts personals com els mateixos membres de l'organització, amb qui es faran reunions i entrevistes que permetran obtenir una excel·lent informació complementària que no sol estar reflectida documentalment.
- En el cas de l'anàlisi de l'audiència, són especialment útils les entrevistes, els qüestionaris i enquestes, les discussions de grup, l'observació i les anàlisis de protocols en entorns reals o experimentals de recuperació, les anàlisis transaccionals, els tests d'usabilitat, etc.
- En el cas de l'anàlisi dels continguts és especialment útil elaborar un inventari o mapa de continguts, que identifiqui tots els tipus que s'incorporen al lloc web i que inclogui una mostra representativa de cada un dels tipus.

La planificació estratègica és un procés necessari per a la correcta construcció de totes les taxonomies. Per tant, independentment de les condicions de l'entorn web, és necessari portar a terme les tres anàlisis i, per tant, utilitzar totes les fonts d'informació apropiades a totes tres.

En canvi, les condicions de l'entorn web sí determinaran importants decisions metodològiques en fases posteriors de la construcció de la taxonomia: la recopilació dels termes (o categories) i l'establiment de relacions semàntiques entre ells. Vegem-ho a continuació.

2.2. Mètodes de selecció dels termes i de l'establiment de relacions entre termes

La norma Z39.19-2005, i totes les versions anteriors, diferencien tres mètodes diferents per a la construcció de taxonomies en funció dels mecanismes que apliquen per a la recopilació dels termes.

- Perspectiva d'un comitè o consell editorial: els termes són proposats per un equip d'especialistes en el domini de la taxonomia, amb l'ajut d'especialistes en gestió de la informació.
- Perspectiva empírica: els termes són extrets dels propis objectes de contingut.
- Combinació de mètodes: els termes són recopilats a partir d'una combinació de les dues opcions anteriors.

Els dos primers mètodes presenten variacions segons com s'estableixin les relacions entre els termes i, fonamentalment, les relacions jeràrquiques. En el cas de la perspectiva d'un comitè o consell editorial, les dues variacions són el *mètode de dalt a baix* i el *mètode de baix a dalt*:

- Mètode de dalt a baix: en primer lloc s'identifica un nombre reduït de classes principals. Per a cada classe principal, s'identifiquen els termes específics, i se situen en els diferents nivells d'especificitat fins arribar al nivell desitjat. Aquest mètode és especialment recomanable quan s'ha de crear una taxonomia ex novo.
- Mètode de baix a dalt: en primer lloc s'identifiquen els termes que tenen un abast conceptual més específic. Els termes es van agrupant en classes més generals, fins assolir el nivell màxim de generalitat. Aquest mètode és especialment recomanable quan ja es disposa de la taxonomia, i l'objectiu és ampliar la seva base terminològica i/o reestructurar-la.

En el cas de la perspectiva empírica, les dues variacions són el *mètode deductiu* i el *mètode inductiu*:

- Mètode deductiu: un cop s'ha extret un nombre suficientment representatiu de termes dels objectes de continguts, s'estableixen les relacions estructurals (principalment, les relacions jeràrquiques) entre els termes. La progressió en l'establiment de les relacions és de dalt a baix. Un cop s'han establert les relacions entre els termes, aquests són sotmesos al control del vocabulari.
- Mètode inductiu: conforme els termes són extrets dels objectes de continguts, es sotmeten al control del vocabulari i se situen en una o més classes principals. A continuació es desenvolupa l'estructura jeràrquica seguint una progressió de baix a dalt.

2.3. Fonts dels termes i de les relacions entre els termes

Des del punt de vista de les directrius i recomanacions relatives a la construcció de taxonomies, la principal innovació que incorpora la revisió de la norma Z39.19 publicada el 2005 és la diversificació de les fonts aplicables als processos de recopilació de termes (que inclou la identificació dels conceptes que han de ser representats, la selecció de termes, i la determinació dels termes preferents) i l'establiment de relacions entre termes.

En les versions anteriors de la norma, les fonts aplicables eren fonamentalment bibliogràfiques (justificació bibliogràfica), i, de manera limitada, basades en les preferències i el comportament dels usuaris de la taxonomia (justificació basada en la corporació). En canvi, la revisió de 2005 incorpora un tercer tipus de fonts, basades en les necessitats i prioritats de la corporació (justificació basada en la corporació). A més, cap d'aquests tres tipus de fonts no és considerada, d'antuvi, prioritària; cal seleccionar-ne un tipus tenint en compte les característiques del lloc web on s'aplicarà.

Aquests canvis estan profundament connectats amb els fonaments de la revisió de la norma. Efectivament: la finalitat de la revisió és ampliar i diversificar els entorns informatius i l'audiència. Ja no són prioritaris els entorns bibliotecaris orientats a la recuperació de documents en suports analògics (catàlegs bibliogràfics i bases de dades documentals), ni l'audiència formada per documentalistes que creen i mantenen aquests instruments. Ara és necessari donar resposta a les necessitats d'organització i recuperació de continguts dels sistemes d'informació situats fora dels entorns bibliotecaris. En aquests nous contextos, la justificació bibliogràfica no és l'única font vàlida dels termes; la captació i la fidelització d'usuaris exigeix la incorporació dels termes i formes de termes realment utilitzats pels usuaris, i la transferència als llocs web de les prioritats estratègiques de la corporació exigeix la incorporació dels termes preferits pels seus membres.

a) Justificació basada en la bibliografia

El glossari de la norma Z39.19-2005 defineix la justificació bibliogràfica («literary warrant») de la manera següent:

Justification for the representation of a concept in an indexing language or for the selection of a preferred term because of its frequent occurrence in the literature. (Z39.19-2005: 6).

Són exemples representatius de fonts bibliogràfiques els següents:

- Documents i altres objectes de contingut especialitzats en el domini de la taxonomia.
- Fonts de referència especialitzats en el domini de la taxonomia com diccionaris, glossaris, enciclopèdies, tractats, manuals, etc.
- Taxonomies i vocabularis controlats ja existents, serveis de resum i índexs, i llistes d'encapçalaments de matèries especialitzats en el domini de la taxonomia.
- Opinions d'experts en el domini de la taxonomia.

Les fonts bibliogràfiques poden ser internes —és a dir, creades o utilitzades per la corporació que ha de desenvolupar la taxonomia— o externes —és a dir, creades per altres instàncies. Les fonts internes poden haver estat recopilades en les fases d'anàlisi de la situació de partida i d'anàlisi dels continguts del lloc web. Alguns exemples clau d'aquestes fonts són les taxonomies i altres vocabularis controlats ja utilitzats en el lloc web o en altres sistemes d'informació de la corporació; les mostres representatives dels diferents tipus de continguts que han estat recopilats en la fase d'anàlisi dels continguts; els consultors de les diferents àrees d'activitat de la corporació.

Per a la identificació dels termes en els documents representatius i en altres fonts bibliogràfiques es poden utilitzar eines d'extracció de termes desenvolupats, principalment, en les àrees de la terminologia i de la lingüística computacional. Per exemple, AntConc <<http://www.antlab.sci.waseda.ac.jp/software.html>> o WordSmith Tools <<http://www.lexically.net/wordsmith/index.html>>.

b) Justificació basada en els usuaris

El glossari de la norma Z39.19-2005 defineix la justificació basada en els usuaris («user warrant») de la manera següent:

Justification for the representation of a concept in an indexing language or for the selection of a preferred term because of frequent requests for information on the concept or free-text searches on the term by users of an information storage and retrieval system. (Z39.19-2005: 10).

En les versions anteriors de la norma, ja havia estat introduïda com a criteri alternatiu a la justificació bibliogràfica en un sol àmbit: la determinació de l'ús de la forma singular o plural dels termes preferents.

5.1.1 Exception to Plural Count Nouns. If in the domain of the thesaurus there is literary or user warrant for the expression of count nouns in the singular, establishment of descriptors in that form is acceptable. For example, in the field of biomedicine, the names of parts of the body are generally formulated in the singular. (Z39.19-2003: 13).

De tots els mecanismes possibles per obtenir informació sobre les preferències dels usuaris, la norma Z39.19-2005 únicament en destaca un: la identificació dels termes i formes dels termes que els usuaris utilitzen de manera més freqüents en les consultes plantejades als sistemes de recuperació de la informació (especialment, als cercadors). Hi ha diferents tipus d'aplicacions informàtiques que ens permeten identificar aquests termes, que presentem a continuació ordenats de més estàndard a més específics:

- Aplicacions d'anàlisi de webs. Per exemple, Web Log Storming <<http://www.datalandsoftware.com/weblog/index.html>>, WebSTAT <<http://www.webstat.com>> i WebTrends <<http://www.webtrends.com>>.
- Aplicacions i mòduls d'extracció de registres de consultes dels cercadors. En general, tots els cercadors generen registres de transaccions de cerca. SearchTools.com, per exemple, <<http://www.searchtools.com/tools/tools.html>> relaciona aproximadament 200 productes que es poden implementar com a enginy de cerca intern en un lloc web, i possiblement cada un d'ells té el seu format propietari d'elaboració de registres de consultes. Específicament, HBX On-Demand de WebSideStory <<http://www.hbxondemand.com>> ha desenvolupat una aplicació d'anàlisi de transaccions per als usuaris d'AtomZ.
- Aplicacions d'anàlisi de la recuperació. Per exemple, BehaviorTracking de Mondosoft <<http://www.behaviortracking.com/reports.asp>>.

Més enllà de la norma Z39.19-2005, no podem oblidar altres mecanismes directes per a la identificació del llenguatge utilitzat pels usuaris, com són les entrevistes de representants dels grups identificats a la fase d'anàlisi dels usuaris, i els grups de seguiment. Cal destacar, d'altra banda, que els mecanismes directes i indirectes d'identificació del llenguatge dels usuaris no ens han de permetre únicament identificar termes i formes dels termes preferits, sinó també relacions semàntiques entre els termes: és a dir, quines equivalències, quines relacions jeràrquiques i quines associacions fan els usuaris amb els termes que utilitzen. En aquest sentit, resulta especialment interessant l'anàlisi de coocurrència de termes en les consultes als sistemes de recuperació de la informació; l'anàlisi de les rutes de navegació, i els tests d'ordenació de targes amb aplicacions com CardCluster <<http://condor.depaul.edu/~jtoro/cardzort/cardzort.htm>>.

c) Justificació basada en la corporació

La justificació basada en la corporació («organizational warrant») és una novetat de la revisió de 2005 de la norma Z39.19.c El glossari la defineix de la manera següent:

Justification for the representation of a concept in an indexing language or for the selection of a preferred term due to characteristics and context of the organization. (Z39.19-2005: 7).

La seva finalitat és traslladar a la taxonomia les necessitats i prioritats de la corporació, a través de la identificació dels termes i formes dels termes que són preferits i utilitzats pels membres de la corporació.

La norma no proposa mecanismes per a la identificació de termes procedents de la justificació basada en la corporació. Podem citar-ne les següents:

- Fonts directes: entrevistes i grups de seguiment amb representants de les diferents unitats de la corporació.

- Fonts indirectes: extracció de termes a partir d'eines d'organització i de gestió, com poden ser els documents fundacionals, els plans estratègics, els organigrames, els manuals de procediments...

Cada tipus de justificació exigeix l'ús de fonts específiques per a la identificació de termes i formes dels termes. Aquestes fonts coincideixen, en gran mesura, amb les que es van utilitzar per a les tres anàlisis que fonamenten la planificació estratègica de la taxonomia: context, audiència i continguts. En aquest sentit, podem establir les següents correspondències:

- Les fonts d'informació de l'anàlisi del context resulten especialment útils per a la justificació de la representació de conceptes i la selecció de termes preferents sobre la base de les característiques de la corporació (justificació basada en la corporació).
- Les fonts d'informació de l'anàlisi de l'audiència resulten especialment útils per a la justificació de la representació de conceptes i la selecció de termes preferents sobre la base del comportament dels usuaris (justificació basada en els usuaris).
- Les fonts d'informació de l'anàlisi del contingut resulten especialment útils per a la justificació de la representació de conceptes i la selecció de termes preferents sobre la base dels continguts (justificació basada en la bibliografia).

3. Oportunitats i amenaces en el desenvolupament de taxonomies

L'existència de normes i documents similars en un sector d'activitat i de producció reporten beneficis a tots els seus agents: empreses, usuaris i clients, administració... la societat en general. Diferents instàncies han identificat i mesurat aquests beneficis, tant de les normes d'àmbit nacional —*Economic benefits of standardization: summary of results*, elaborat pel Deutsches Institut für Normung e. V. (DIN)—, com internacional —*How ISO standards benefit society*. D'acord amb el World Standards Services Network (WSSN) <<http://www.wssn.net/WSSN/index.html>>, una norma és, per als agents econòmics, un factor de racionalització de la producció; un factor de clarificació de les transaccions; un factor d'innovació i desenvolupament de productes; un factor de transferència de noves tecnologies; un factor d'elecció estratègica de les empreses. En el sector de la gestió de la informació, la revisió de la norma Z39.19 publicada el 2005 tenia la possibilitat d'ampliar els beneficis i beneficiats. Pel que fa als beneficis, i a través del posicionament dels vocabularis controlats en l'entorn web, la revisió de la norma ofereix mecanismes de racionalització dels processos de disseny i ús de sistemes d'organització i cerca d'informació, així com la reutilització i interoperabilitat dels sistemes. Pel que fa als beneficiats, als destinataris tradicionals de la norma, s'incorporen ara totes les organitzacions que han de dissenyar llocs web públics, intranets i extranets; les empreses de consultoria en arquitectura de la informació; les empreses productores d'aplicacions per a les diferents funcionalitats de la gestió dels continguts digitals; i els usuaris i clients de tots aquests agents.

Més concretament, la norma Z39.19-2005 podia significar un impuls definitiu a les taxonomies com a eines d'organització i recuperació de continguts en els llocs web. No obstant, gairebé coincidint amb el període de revisió de la norma (2002-2005), Internet va experimentar un procés de redefinició que es va concretar, el 2004, amb l'encunyament de la marca Web 2.0.³ Aquest terme identifica un nou model de negoci a Internet, diferent del model de les empreses

3. Aquest terme va ser encunyat per O'Reilly Media el 2004.

puntcom —el Web 1.0—, que van fracassar amb l'esclat de la bombolla tecnològica la tardor del 2001. D'acord amb la *Wikipedia*:⁴

Web 2.0 [...] refers to a perceived or proposed second generation of Internet-based services —such as social networking sites, wikis, communication tools, and folksonomies— that emphasize online collaboration and sharing among users.

En el context de l'organització i cerca d'informació, les preferències del Web 2.0 semblen posar en qüestió els beneficis de les taxonomies. De fet, a la primera conferència sobre Web 2.0. (Web 2.0 Summit <<http://www.web2summit.com>>), a la pluja d'idees que confrontava aquesta segona generació d'Internet amb la primera, es va incorporar la dicotomia:

Web 1.0	Web 2.0
directories (taxonomy)	tagging («folksonomy»)

Els models del Web 2.0 constitueixen amenaces per a la continuïtat de les taxonomies com a sistemes òptims d'organització i cerca de continguts al web. És possible identificar tres punts febles de les taxonomies des del punt de vista del paradigma del Web 2.0: el model estructural jeràrquic; el model centralitzat; i el model estàtic.

3.1. Escac al model estructural jeràrquic

El model estructural es defineix pel tipus de relacions que connecten totes les categories del vocabulari controlat i que constitueixen, per tant, la columna vertebral del sistema. Les normes internacionals i nacionals de tesaurus han reduït els models estructurals possibles al jeràrquic, i han incorporat dins d'aquest el model arbori basat en la relació partitiva. L'escac al model jeràrquic té dos orígens:

- Un model vinculat al Web 2.0, que reivindica el valor dels enllaços com a base de l'organització dels continguts al web —«the link, which can point to anything» en paraules de Shirky (2005).
- El model de l'anàlisi de facetes, ja existent en la tradició dels SOC, però que havia estat desplaçat de les normes dels tesaurus fins que ha estat incorporat, tímidament, a la darrera revisió de la Z39.19 publicada el 2005.

Com hem indicat anteriorment, el model estructural jeràrquic és un dels trets definitoris de la taxonomia. Els punts forts de la jerarquia són:

- Aporten molta informació sobre els atributs dels membres. L'herència (transitivitat) al llarg de la cadena jeràrquica permet inferir moltes propietats. Permeten definir un concepte a partir de la inclusió en una classe, les propietats compartides i les diferències respecte als membres d'una mateixa sèrie.
- Aporten economia de notació: amb pocs símbols podem representar molta informació.
- Aporten una visió de conjunt dels fenòmens que caracteritzen un àmbit temàtic, sempre que es respectin les regles de divisió dels conceptes en classes i subclasses, i les característiques

4. «Web 2.0.» En: *Wikipedia: the free encyclopedia*. Wikimedia Foundation, Inc., last modified 12:54, 17 January 2007. <http://en.wikipedia.org/wiki/Web_2.0>. [Consulta: 17/1/2007].

escollides reflecteixen diferències clares, perdurables i adequades a l'àmbit d'aplicació del sistema.

La jerarquia també presenta punts febles. Shirky (2005) revisa els sistemes de categorització tradicionals procedents de l'entorn bibliotecari que s'han aplicat als llocs web. Denomina aquests sistemes de classificació «classificacions o categoritzacions ontològiques», i les defineix com el resultat del procés d'organitzar un conjunt d'entitats en grups, sobre la base de les seves essències i relacions *a priori*. Aquestes relacions són, fonamentalment, jeràrquiques, i imposen límits a les relacions que es poden establir entre els objectes categoritzats mitjançant una classificació ontològica (límits a la policlassificació), i entre les pròpies categories (límits a la polijerarquia). En el context web, la multiplicitat de relacions entre els objectes es poden definir a través dels vincles, sense imposar límits *a priori*:

[...] if you've got enough links, you don't need the hierarchy anymore. There is no shelf. There is no file system. The links alone are enough.

Shirky (2005) no rebutja absolutament el model jeràrquic en l'organització dels continguts web ni, per tant, les taxonomies, sinó que caracteritza contextos en què aquest model resulta adient, i altres on no resulta adient.

CONTEXTOS FAVORABLES A LES CLASSIFICACIONS ONTOLÒGIQUES	CONTEXTOS NO FAVORABLES A LES CLASSIFICACIONS ONTOLÒGIQUES
DOMINI QUE CAL ORGANITZAR	DOMINI QUE CAL ORGANITZAR
Corpus reduïts	Corpus extensos
Categories formals	Categories no formals
Entitats estables	Entitats inestables
Entitats restringides	Entitats no restringides
Límits definits	Límits no definits
PARTICIPANTS	PARTICIPANTS
Catalogadors experts	Usuaris no coordinats
Font autoritzada de judici	Usuaris no professionals
Usuaris coordinats	Catalogadors no experts
Usuaris experts	Sense fonts autoritzades

L'anàlisi de facetes consisteix a identificar un conjunt limitat de característiques o dimensions que permeten representar de forma completa qualsevol objecte en el context d'un sistema de representació determinat. Cada una d'aquestes característiques constitueix una categoria fonamental, format per un conjunt limitat de conceptes o facetes. L'organització dels conceptes dins d'una categoria es pot basar en la relació que es consideri més adient: jeràrquiques, arbres o, de nou, facetes. Tots els objectes que s'incorporen al lloc web són descrits per, com a mínim, una faceta de cada una de les categories fonamentals. La suma d'aquestes facetes constitueix la representació completa de l'objecte. Per exemple: nines japoneses de fusta tallada del segle xv.

Enfront del model jeràrquic, l'anàlisi de faceta presenta punts forts i punts febles. Entre els primers, cal destacar els següents:

- Només cal identificar les categories fonamentals que hi intervenen, les facetes s'organitzen a dins segons diferents característiques, no cal definir una característica essencial i les subordinades a ella. És especialment adient per àrees noves i en constant evolució.
- Hospitalitat: l'anàlisi dels conceptes compostos en elements més simples permet ubicar nous conceptes a partir de les categories fonamentals.
- Flexibilitat: permet fer diverses combinacions de característiques.
- Expressivitat: permet incorporar l'estructura més adient dins les categories, d'aquesta manera pot integrar estructures i models teòrics diferents.

En canvi, hi ha dos punts febles a tenir en compte:

- Rigidesa en el nivell de les classes principals. Un cop establertes les categories fonamentals adients per a l'àmbit d'especialitat, no es poden integrar noves categories amb facilitat.
- Dificultat de mostrar la connexió entre facetes dins categories diferents, tant pel que fa a la visualització de l'estructura global, com a aspectes més puntuals.

Com hem indicat anteriorment, la norma Z39.19-2005 incorpora l'anàlisi de facetes com un model estructural alternatiu al jeràrquic en els vocabularis controlats.

Controlled vocabularies —especially large ones consisting of thousands of terms— may be easier to use if they are organized in some way other than hierarchically. Faceted analysis, based on the work of Ranganathan and refinements by the Classification Research Group, is another way of organizing knowledge. (Z39.19-2005: 14).

La part dispositiva de la norma dedica a l'anàlisi de facetes una secció força completa que inclou, fins i tot, procediments d'elaboració. No obstant això, aquesta secció no s'integra a la secció de relacions semàntiques («5.3.3 Semantic Relationships»). En canvi, a l'apèndix «B.1. Properties of Lists, Synonym rings, Taxonomies, and Thesauri», l'anàlisi de facetes sí és tractada com un tipus de relació semàntica entre termes, disponible per als tesaurus, però, en canvi, no apropiada per a les taxonomies.

A la pràctica, però, trobem exemples de taxonomies que apliquen, com a model estructural, l'anàlisi de facetes. Adkisson (2005) va analitzar l'octubre de 2003 setanta-cinc llocs web líders en comerç electrònic i va descobrir que el 69 % d'aquests llocs feien algun ús de classificació facetada. Aquest ús és especialment intens en entorns on s'ha individualitzat el tipus d'objecte que ha de ser representat mitjançant la taxonomia; per exemple, la recepta culinària en el cas d'Epicurious <<http://www.epicurious.com>> o l'ampolla de vi en el cas de Wine.com <<http://www.wine.com>>.

Cal destacar, a més, que els llocs web mostren diferents possibilitats pel que fa a la presentació de les taxonomies facetades. En concret, Adkisson (2005) va identificar tres opcions: aplicació en el sistema d'exploració (*browsing*); aplicació en l'opció avançada del sistema de recuperació (cerca paramètrica); o aplicació en ambdós sistemes.

El desenvolupament del model facetat de les taxonomies està profundament lligat a l'evolució d'XFML (*eXchangeable Faceted Metadata Language*), un estàndard basat en XML que permet la publicació i intercanvi de metadades basades en estructures facetades. Les especificacions d'aquest format obert es poden consultar a XFML <<http://www.xfml.org>>. Facetmap <<http://facetmap.com>> és una aplicació que permet la construcció i gestió de taxonomies facetades basades en l'estàndard XFML.

3.2. Escac al model centralitzat de les taxonomies

La construcció i manteniment dels vocabularis controlats i, per tant, de les taxonomies, està en mans d'un ens central; d'acord amb la norma ANSI/NISO Z39.19-2005, és responsabilitat d'un consell editorial formal. En canvi, un dels principis en què es fonamenta el Web 2.0 és l'aprofitament de la intel·ligència col·lectiva en els processos de creació, organització/categorització, filtratge, etc., dels continguts. És a dir: en la descentralització de la responsabilitat sobre aquests processos. En el marc del Web 2.0, s'ha desenvolupat un nou model de categorització dels continguts coherent amb la descentralització: l'etiquetatge (*tagging*). Consisteix en l'assignació de paraules claus o etiquetes (*tags*) als recursos que es publiquen o consulten en els llocs web. En l'assignació de les etiquetes, els usuaris no estan sotmesos (encara) a cap control terminològic previ. Si el model basat en el control terminològic pot donar lloc a les taxonomies, l'etiquetatge col·laboratiu pot donar lloc a les *folksonomies*.⁵ Una *folksonomia* és un llenguatge documental per a la representació i de recuperació de la informació en un entorn col·laboratiu, creat a partir de l'etiquetatge descentralitzat, on els termes —millor dit, les etiquetes— no estan connectats mitjançant relacions semàntiques, sinó mitjançant altres tipus de relacions que podríem denominar pragmàtiques. Així, per exemple, Connotea <<http://www.connotea.com>> ens ofereix la possibilitat de consultar les etiquetes relacionades («related tags»):

Related tags are those which have been used by other people to categorise links you have in common. For example, if you are saving papers under AvianFlu, and another user is saving some of the same papers, but using the tag Bird Flu, you will see Bird Flu as a related tag. This opens the door for you to discover interesting content which is related to your own but has been tagged differently.

Una folksonomia es presenta, per tant, com «un espai de noms pla, sense jerarquies ni relacions de parentiu predeterminades».

Tenint en compte la comparació entre els models centralitzat i descentralitzat de categorització, és fàcil endevinar els punts forts i febles de les folksonomies. Pel que fa als primers, cal destacar el fet que són sistemes sotmesos a la lògica del mercat, que maximitzen el valor de la col·lectivitat. Pel que fa als punts febles, podem destacar-ne dos:

- Quant a l'assignació d'etiquetes, la folksonomia presenta un inconvenient comú a tots els llenguatges documentals no controlats: la presència d'ambigüitats i errors. Els usuaris utilitzen diferents termes per a designar els mateixos conceptes i un mateix terme per a diferents conceptes; de vegades, simplement els termes i conceptes utilitzats són erronis. A més, la participació de diversos usuaris comporta inconsistències en el nombre d'etiquetes assignades, i en els principis d'especificitat i exhaustivitat.
- Quant a la consulta de la folksonomia, el treball col·laboratiu dona lloc a dimensions sovint difícilment abastables pels usuaris. Els llocs web que han aplicat folksonomies, han desenvolupat diferents mecanismes per superar aquest inconvenient: la diversificació de les presentacions de la folksonomia: llistes alfabètiques, llistes ordenades de més a menys usades, etc.; i la fragmentació de les folksonomies a partir de criteris pragmàtics com grups d'usuaris, etiquetes aplicades a un mateix recurs, etc.

5. El terme *folksonomia* és un calc de l'anglès *folksonomy*, creat mitjançant la combinació de *folk* i (*taxo*)*nomy*. Altres termes utilitzats a la bibliografia per a designar aquest concepte són la variant ortogràfica *folcsonomia*, i les formes *gentonomia* o *etnoclassificació*. Termcat no s'han pronunciat sobre aquesta qüestió.

La confrontació entre el model taxonòmic i folksonòmic de categorització i organització dels continguts ha donat lloc a interessants debats en els fòrums especialitzats en arquitectura de la informació, com és el cas del que amb el títol «The resilience of information architecture» va ser dinamitzat per Morville entre els mesos d'abril i maig de 2006 a *Findability.org* <<http://findability.org>>. Morville (2006) considera que existeixen contextos on les folksonomies són més adients per a l'organització dels continguts, i contextos on les taxonomies resulten més oportunes.

Tagging has flourished in the free-wheeling blogosphere, but has had little traction in the realm of corporate and government web sites where authority is an equal partner to findability. It remains unclear when and whether tags belong in the information ecologies in which IAs typically practice.

Aquesta opinió se situa en la línia de l'expressada per Shirky (2005) quan defensa l'existència d'entorns on les classificacions ontològiques funcionen correctament, i d'altres on no són recomanables.

En tot cas, són múltiples els entorns on les folksonomies s'han desenvolupat amb èxit:

- Serveis que permeten publicar i compartir favorits: generalistes —com *del.icio.us* <<http://del.icio.us>>— o especialitzats —com Connotea <<http://www.connotea.com>>.
- Serveis que permeten publicar i compartir fotografies: Flickr <<http://www.flickr.com>>.
- Serveis basats en programari social per a la comunicació d'un a diversos (blocs com *Taxonomy Watch* <<http://taxonomy2watch.blogspot.com>>), i de diversos a diversos (sistemes wiki com Taxonomy Community of Practice Wikispace <<http://taxocop.wikispaces.com>>).

Una línia de treball que pot obrir-se en el futur més immediat és la combinació i, fins i tot, integració d'ambdós models d'organització del coneixement en llocs web.

3.3. Escac al model estàtic de les taxonomies

Les taxonomies són, com la resta de vocabularis controlats, llenguatges documentals fonamentalment estàtics, tant des del punt de vista de la construcció com de la seva aplicació en la categorització d'objectes informatius. Des del punt de vista de la construcció, una taxonomia és estàtica perquè únicament es pot utilitzar quan ja han finalitzat completament els processos d'identificació de conceptes, de selecció dels termes que representen els conceptes, i d'establiment de relacions semàntiques entre els termes. Evidentment, una taxonomia és un instrument viu, i és possible (i necessari) introduir-hi canvis tant en els termes com en les relacions entre els termes. Però aquests canvis formen part dels processos d'actualització i manteniment de la taxonomia, i segueixen uns protocols preestablerts. Des del punt de vista de la categorització, una taxonomia és estàtica perquè els termes s'assignen als objectes informatius abans del desenvolupament de processos de cerca en els sistemes d'exploració (*browsing*) i de recuperació, i no de manera simultània a aquests processos.

Les taxonomies construïdes en el marc de la norma la norma ANSI/NISO Z39.19-2005 són llenguatges documentals estàtics que s'apliquen, fonamentalment, a processos de classificació i de recuperació estàtics.

Fora d'aquest marc, és possible trobar aproximacions a les taxonomies com a instruments dinàmics de categorització d'objectes d'informació. Una cerca al web del terme «taxonomia dinàmica» ens ofereix definicions com les següents:

- Per designar el model de desenvolupament de taxonomies que incorporen mecanismes molt àgils d'incorporació de nous termes i noves relacions entre termes en vocabularis ja existents. És el cas del mòdul Taxonomy del sistema de gestió de continguts Drupal <<http://digitalsolutions.ph/couchkamotereviews/book/view/47>>.
- Per designar un model de desenvolupament de taxonomies que aplica la policlassificació en la categorització dels recurs. És el cas del model proposat per Sacco (2002 i 2003):

[...] conventional taxonomies are monodimensional (i.e. an item is classified under one and only one concept), whereas dynamic taxonomies are multidimensional (i.e. an item can be classified under several different concepts at any level of abstraction).

En la transició del Web 1.0 al Web 2.0, la confrontació entre taxonomies estàtiques i dinàmiques s'associa a una confrontació entre dos models de cerca d'informació: els sistemes d'exploració (*browsing*), que es vinculen al Web 1.0, i els sistemes de recuperació (*searching*), que es vinculen al Web 2.0.⁶ Els sistemes d'exploració, dissenyats sobre la base de vocabularis controlats com les taxonomies, organitzen els recursos *a priori* (*just-in-case*) i de manera estàtica: *a priori* perquè la categorització dels recursos és anterior a les necessitats d'informació i a la visió del món de l'usuari que utilitza el sistema; estàtica, perquè no té la capacitat d'adaptar-se al context de cerca específic de l'usuari. En canvi, els sistemes de recuperació, que poden oferir als usuaris vocabularis controlats per a la definició de les equacions de cerca, organitzen els recursos *a posteriori* (*just-in-time*) i de manera dinàmica: *a posteriori* perquè la categorització dels recursos és posterior a les necessitats d'informació i a la visió del món de l'usuari que utilitza el sistema; dinàmica, perquè té la capacitat d'adaptar-se al context de cerca específic de l'usuari.

En el context dels sistemes de recuperació d'informació, el terme *taxonomia dinàmica* s'ha aplicat per designar mecanismes d'organització temàtica dels resultats de les cerques dels usuaris, com alternativa a l'ordenació tradicional dels resultats basada en la rellevància, la data de creació, el creador, etc. Estrictament parlant, però, la majoria d'aquestes taxonomies dinàmiques són realment mecanismes d'agrupació temàtica (*topic clustering*) dels continguts. Aquests mecanismes actuen de manera simultània al càlcul dels resultats de la cerca de l'usuari, i fan anàlisis estadístiques de freqüència, coocurrència, posició, etc., del contingut dels resultats per identificar els termes que són comuns a grups significatius de resultats. Els mecanismes d'agrupació temàtica presenten els termes que identifiquen els grups significatius de resultats per tal que l'usuari en seleccioni els més rellevants i pugui accedir als resultats que contenen. En la presentació dels resultats, poden oferir un sol nivell d'agrupació de resultats, o diferents nivells de subagrupació que simulen falses jerarquies. L'agrupació temàtica s'ha incorporat tant en cercadors disponibles a Internet (*search engines*) com en aplicacions orientades al desenvolupament de cercadors en intranets corporatives. Són exemples de la primera opció Ask.com <<http://search.ask.com>>; Query Server <<http://www.queryserver.com>>; Turbo10 <<http://www.turbo10.com>>; Vivísimo <<http://vivisimo.com>>; i WiseNut <<http://www.wisenut.com>>. Pel que fa a les aplicacions per al desenvolupament d'intranets corporatives, cal destacar l'IDOL Server d'Autonomy, amb el mòdul Automatic Taxonomy Generation <<http://www.autonomy.com/content/Products/IDOL/f/Taxonomies/index.en.html>>.

Estrictament parlant, els mecanismes d'agrupació temàtica no satisfan els principis per a la construcció de taxonomies o, en general, de vocabularis controlats. Hi ha, però, models de

6. En opinió de Shirky (2005), Yahoo! es el principal exponent dels sistemes d'exploració, i Google dels sistemes de recuperació.

taxonomies dinàmiques que sí compleixen aquests principis. En aquests models, els documents són categoritzats a partir de vocabularis controlats predefinitos i, més específicament, de taxonomies. En el moment de presentar els resultats de la cerca, el sistema mostra les categories dels documents recuperats i reconstrueix les relacions semàntiques entre les categories dels documents recuperats. D'aquesta manera, és possible l'exploració de l'organització temàtica dels resultats, amb els avantatges que ofereixen les relacions d'equivalència i de jerarquia pre-determinades. A diferència de l'agrupació (*clustering*) temàtica, les categories i les relacions entre les categories no es creen en el moment de la cerca, sinó que ja estan preestablertes. Són mostres d'aquest model de taxonomia dinàmica la proposta de Wollersheim i Rahayu (2002) o la funcionalitat Dynamic Classification que incorpora la versió 8 de RetrievalWare <<http://www.convera.com/solutions/retrievalware/Default.aspx>>.

El panorama actual dels sistemes d'organització i cerca de continguts al web no sembla establir una oposició entre els models representats per les taxonomies i les folksonomies. En aquest sentit, resulten molt aclaridores les paraules de Morville (2005, p. 139):

Ontologies, taxonomies, and folksonomies are not mutually exclusive. In many contexts, such as corporative web sites, the formal structure of ontologies and taxonomies is worth the investment. In others, like the blogosphere, the casual serendipity of folksonomies is certainly better than nothing. And in some contexts, such as intranets and knowledge networks, a hybrid metadata ecology that combines elements of each may be ideal.

En el marc de les solucions híbrides, l'evolució de Youtube <<http://www.youtube.com>> resulta molt reveladora. Youtube va iniciar la seva singladura el febrer del 2005 com «the first online community site that allows members to post and share personal videos». Constituïa una de les principals iniciatives vinculades al paradigma del Web 2.0. En el seu origen, la categorització dels enregistraments vídeo es basava en l'etiquetatge distribuït. El 12 de setembre de 2005, però, s'anuncia la novetat següent:

While the video uploads have been pouring in, we realized it grew ever more difficult for viewers to find relevant and interesting videos. Though tags are very efficient at identifying specific details about content, we wanted to further simplify the browsing experience. With the release of Channels, similar content will be categorized and grouped into common channels.

El 19 d'abril de 2006, els *Channels* esdevenen *Categories*: una llista plana de categories, creades de forma centralitzada, i aplicades de manera distribuïda pels col·laboradors de Youtube quan hi publiquen vídeos.

4. Conclusions

La revisió de les normes nacionals i internacionals de tesaurus en la línia desenvolupada per la norma Z39.19-2005 pot significar una millora del posicionament de les taxonomies en el marc de l'arquitectura de la informació dels llocs web. L'impuls que, en el marc del nou paradigma d'Internet, el Web 2.0, s'està donant als sistemes d'organització i recuperació dels continguts no jeràrquics, descentralitzats i dinàmics, pot representar una amenaça per a la supervivència de les taxonomies. No obstant això, els entorns de creació i desenvolupament de llocs web són molt diversos i, per tant, també són diverses les seves necessitats i exigències pel que fa als sistemes d'organització i recuperació d'informació. Cal preveure, per tant, la supervivència dels diferents models i la seva aplicació a contextos adjacents.

Bibliografia

- ADKISSON, Heidi P. Use of faceted classification [en línia]. Webdesignpractices, 2005. <<http://www.webdesignpractices.com/navigation/facets.html>>. [Consulta: 12/02/2007].
- ALLEN, Katherine. «Taxonomy and folksonomy». *Information World Review*, núm 212 (2005), p. 7.
- ARGUDO, Sílvia; CENTELLES, Miquel. «Metodología para el diseño de taxonomías corporativas». *Investigación Bibliotecológica*, vol. 19, núm. 39 (jul.- des. 2005), p. 158-177.
- British Standards Institution. *Structured vocabularies for information retrieval: guide. Part 1: Definitions, symbols and abbreviations: BS 8723-1:2005*, Londres: BSI Group, 2005. 14 p.
- CENTELLES, Miquel. «Taxonomías para la categorización y la organización de la información en sitios web». *Hipertext.net*, núm. 3 (2005). <<http://www.hipertext.net>>. [Consulta: 26/09/2005].
- CORNELLA, Alfons. «¿Hacia la gentonomía? (tagging o folksonomy)». *Papeles de infonomia para entender la empresa en red: selección mensual de las mejores ideas publicadas en infonomia.com*, núm. 34 (maig 2005), p. 7-8.
- Economic benefits of standardization: summary of results [en línia]. Berlín: DIN German Institute for Standardization e. V., [2000]. <http://www2.din.de/sixcms/media.php/1344/Studie_englisch.pdf>. [Consulta: 31/12/2006].
- GARRETT, Jesse James. *The elements of user experience: user-centered design for the web*. Nova York: AIGA. Indianapolis (Ind.): New Riders, 2002. XIV, 189. ISBN: 0-7357-1202-6.
- Guide to the Development and Maintenance of Controlled Vocabularies in the Government of Canada [en línia]. Ottawa: Treasury Board of Canada Secretariat. Date reviewed: 2006-01-09. <http://www.tbs-sct.gc.ca/im-gi/mwg-gtm/cvsg-sgvc/docs/2005/vocab/vocabtb_e.asp>. [Consulta: 31/12/2006].
- HAMMOND, Tony; HANNAY, Timo; LUND, Ben; SCOTT, Joanna. «Social bookmarking tools (I): a general overview» [en línia]. *D-lib magazine*, vol. 11, núm. 4 (abril 2005). <<http://www.dlib.org/dlib/april05/hammond/04hammond.html>>. [Consulta: 18/12/2005].
- ISO. *Documentation-Guidelines for the establishment and development of monolingual thesauri: ISO 2788:1986*. Ginebra: International Organization for Standardization; 1986. 32 p.
- . «How ISO standards benefit society» [en línia]. En: ISO. *Overview of the ISO system*. Geneva: ISO, last modified 2006-09-12. <<http://www.iso.org/iso/en/aboutiso/introduction/introdex.html#six>>. [Consulta: 31/12/2006].
- LAMONT, Judith. «Dynamic taxonomies: keeping up with changing content» [en línia]. *KM world magazine*, vol. 12, núm. 5 (maig 2003). <<http://www.kmworld.com/Articles/ReadArticle.aspx?ArticleID=9461>>. [Consulta: 18/12/2005].
- LUND, Ben; HAMMOND, Tony; HANNAY, Timo. «Social bookmarking tools (II): a case study, Coonotea» [en línia]. *D-lib magazine*, vol. 11, núm. 4 (abril 2005). <<http://www.dlib.org/dlib/april05/lund/04lund.html>>. [Consulta: 18/12/2005].
- MERHOLZ, Peter. «Metadata for the masses» [en línia]. *Adaptive Path essay archives*, oct. 19, 2004. <<http://www.adaptivepath.com/publications/essays/archives/000361.php>>. [Consulta: 18/12/2005].
- MONTERO, Yusef Hassan; NÚÑEZ Peña, Ana. «Diseño de arquitecturas de información: descripción y clasificación» [en línia]. *No solo usabilidad*, núm. 4 (14 gen. 2005). <http://www.nosolousabilidad.com/articulos/descripcion_y_clasificacion.htm>. [Consulta: 15/12/2006].
- MORVILLE, Peter. *Ambient findability. Pequin [etc.]*: O'Reilly, cop. 2005. XIV, 188 p. ISBN: 0-596-00765-5.
- . «The resilience of information architecture» [en línia]. En: *Findability.org*. (20 abril 2006). <<http://www.findability.org/archives/000113.php>>. [Consulta: 15/12/2006].

- MOUROUZIS, A. [et al.]. Report on existing VR-related guidelines and standards [en línia]. Atenes: Intuition, 2004. <http://www.intuition-eunetwork.net/documents/deliverables/INTUITION-FORTH-D-WP1_11-R1-V1-Existing%20guidelines%20and%20standards.pdf>. [Consulta: 15/12/2006].
- MYER, Thomas. Aclaración de ideas erróneas sobre la arquitectura de información. Instituto Asilomar para la Arquitectura de Información, Julio 1, 2002. <<http://iainstitute.org/es/translations/000212.html>>. [Consulta: 15/12/2006].
- National Information Standards Organization (Estats Units d'Amèrica). Guidelines for the construction, format, and management of monolingual controlled vocabularies: ANSI/NISO Z39.19-2005. Developed by the National Information Standards Organization; approved July 25, 2005 by the American National Standards Institute. Bethesda, Maryland: NISO, cop. 2005. <<http://www.niso.org/standards/resources/Z39-19-2005.pdf>>. [Consulta: 31/12/2006].
- . «TAG Conference Call, December 15, 2003». En: NISO. Developing the next generation of standards for controlled vocabularies and thesauri. <<http://www.niso.org/committees/TRAG/TAGDec1503.pdf>>. [Consulta: 31/12/2006].
- . Guidelines for the construction, format, and management of monolingual thesauri: ANSI/NISO Z39.19 - 2003 / developed by the National Information Standards Organization: approved August 28, 2003, by the American National Standards Institute. Rev. ed. of: American national standard guidelines for thesaurus structure, construction, and use. New York: ANSI, c1980. Bethesda, Maryland: NISO, cop. 2003. <http://download.www.techstreet.com/cgi-bin/pdf/free/403963/Z39-19_2003.pdf>. [Consulta: 31/12/2006].
- O'REILLY, Tim. What is Web 2.0: design patterns and business models for the next generation of software [en línia]. Sebastopol, CA: O'Reilly Network, 09/30/2005. <<http://www.oreilly-net.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>>. [Consulta: 17/12/2006].
- PLAZA I CERVERA, Enric. Gentonomia (o la Saviesa de la Munió) [en línia]. <<http://www.iiia.csic.es/People/enric/papers/cau/Gentonomia.pdf>>. [Consulta: 18/12/2005].
- SACCO, Giovanni M. Conventional taxonomies vs. dynamic taxonomies [en línia]. University of Torino, Technical report. June 2002. <<http://www.dbworldx.di.unito.it/sacco/papers/sacco-chi03w.pdf>>. [Consulta: 10 febr. 2007].
- . «Dynamic taxonomies: a model for large information bases». IEEE transactions on knowledge and data engineering. Vol. 12, núm 3 (2000), p. 468.
- . Dynamic taxonomies and guided searches [en línia]. CHI 2003: Best Practices and Future Visions for Search UIs: A Workshop. Miami, 2003. <<http://www.dbworldx.di.unito.it/sacco/papers/sacco-chi03w.pdf>>. [Consulta: 10 febr. 2007].
- SHIRKY, Clay. Ontology is overrated: categories, links, and tags [en línia]. 2005. <http://www.shirky.com/writings/ontology_overrated.html>. [Consulta: 18/12/2005].
- «Web 2.0.» En: Wikipedia: the free encyclopedia. Wikimedia Foundation, Inc., last modified 12:54, 17 January 2007. <http://en.wikipedia.org/wiki/Web_2.0>. [Consulta: 17/1/2007].
- WOLLERSHEIM, Dennis; RAHAYU, J. Wenny. «Methodology for creating a sample subset of dynamic taxonomy to use in navigating medical text databases». En: Proceedings of the 2002 International Symposium on Database Engineering & Applications. 2002, p. 276-284.

Agraïments

Agraeixo a Carles Gibernau la informació sobre el treball de Clay Shirky.

Aquest treball ha estat parcialment finançat pel Ministeri d'Educació i Ciència, com a part del projecte HUM2004-03162/FILO.



Miquel Centelles Velilla. És diplomat en Biblioteconomia i Documentació i llicenciat en Filologia Hispànica, i està fent el doctorat interuniversitari Ciència Cognitiva i Llenguatge en la fase de tesi. Les seves àrees d'especialització són els sistemes de representació del coneixement en l'entorn digital i, especialment, els sistemes de recuperació d'informació i les taxonomies.

Des de l'any 2001, és professor titular al Departament de Biblioteconomia i Documentació de la Universitat de Barcelona. A més, col·labora com a professor en el màster de Documentació Digital organitzat per la Universitat Pompeu Fabra.

És membre de dos grups de recerca universitaris: el Grup d'Investigació «Organització i recuperació de continguts digitals», vinculat al Departament de Biblioteconomia i Documentació de la Universitat de Barcelona, i el Grup de Recerca en Ciències de la Documentació (DIGIDOC) de la Universitat Pompeu Fabra.

És autor de diferents articles publicats en revistes especialitzades en informació i documentació com *BiD: textos universitaris de biblioteconomia i documentació*, *Hipertexto*, *Ítem* i *Investigación bibliotecológica*.