

Taxonomies al web: planificació i posada en marxa d'un sistema d'organització i recuperació de continguts en entorns corporatius

9es Jornades Catalanes d'Informació i Documentació
Centre de Convencions Internacional
Barcelona, 26 i 27 de novembre

Miquel Centelles Velilla
Facultat de Biblioteconomia i Documentació

Objectius del taller

- Conèixer el concepte de taxonomia corporativa i les especificitats que presenta en el context dels llenguatges documentals.
- Conèixer els processos i les eines que cal aplicar en la construcció d'una taxonomia corporativa.
- Aplicar els processos i les eines a supòsits reals de construcció de taxonomies corporatives.

Organització del taller

Presentació de conceptes introductoris.

Presentació de la metodologia per al disseny i implementació de taxonomies.

Es diferencien cinc fases:

- Planificació estratègica.
- Construcció de la taxonomia corporativa.
- Categorització.
- Aplicació i presentació de la taxonomia.
- Cicle d'avaluació i millora contínues. Difusió.

Per a cada fase, es presenten:

- Conceptes específics.
- Resultats que se n'esperen.
- Tècniques, recursos i eines informàtiques.
- Normes, recomanacions i directrius.
- Processos.

Realització d'una activitat relativa a cada fase.

Conceptes introductoris

Què és una taxonomia corporativa?

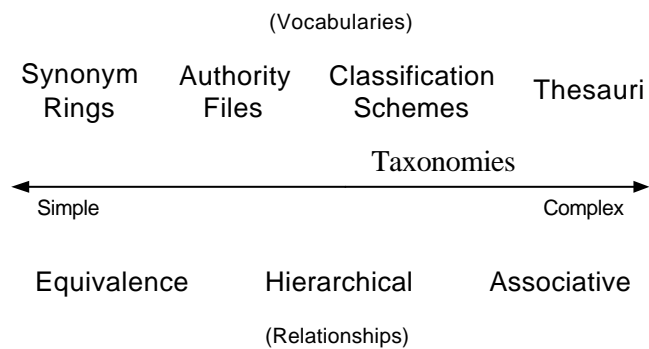
- Des del punt de vista de l'**estructura**, no presenta diferències respecte a altres eines d'organització del coneixement i llenguatges documentals controlats.
- Les diferències fonamentals es troben en l'àmbit de les **funcions i àmbits d'aplicació**.

Conceptes introductoris

Estructura

- L'estructura consisteix en un conjunt de termes (o categories) i en relacions entre aquests termes.
- Elements típics de l'estructura:
 - Relacions d'equivalència entre els termes acceptats i els termes no acceptats (sinònims, les variants lèxiques, els quasi-sinònims, els termes compostos).
 - Relacions jeràrquiques.
 - Relacions associatives.

Conceptes introductoris



Font: Amy J. Warner. *Metadata & taxonomies for a more flexible information architecture* (2002)

Conceptes introductoris

Funcions i àmbits d'aplicació

- Reflecteix l'organització del coneixement d'una corporació determinada, és a dir, la forma com es crea, capta, cerca i utilitza la informació en aquesta entitat, la seva cultura organitzativa, les seves terminologies, etc.
- Permet la representació de totes les formes d'informació generada i utilitzada per una organització determinada:
 - Informació interna i externa.
 - Informació formal i informal / estructurada i desestructurada.
 - Informació sobre totes les entitats cognitives: matèries, persones, entitats, productes i serveis, funcions i activitats, processos ...).

Conceptes introductoris

Funcions i àmbits d'aplicació

- Llocs web corporatius.
- Sistemes d'organització i exploració de la informació ("browsing").
- Sistemes de recuperació de la informació (cercadors).
- Sistemes de filtratge de la informació (DSI).
- Sistemes de navegació de la informació.
- Posicionament en cercadors.
- Minería de dades i avantatge competitiva.
- Col·laboració entre experts.

Conceptes introductoris

Què NO és una taxonomia corporativa?

- La resposta d'un comitè de savis.
- El resultat d'una eina informàtica màgica.
- Una eina acabada.
- Una oportunitat per fer feina pendent.

Conceptes introductoris

Punts d'atenció abans de continuar

- Disseny centrat en les persones.
- Disseny en processos diferenciats:
 - La construcció que comporta l'establiment de l'esquema de classificació i el model d'estructura.
 - La categorització
 - L'aplicació i presentació.
- Disseny conduït per l'avaluació contínua

Planificació estratègica

Resultat

Document en forma d'informe on es defineixi el model de taxonomia que cal dissenyar i implementar, a partir de l'anàlisi del context organitzatiu, de l'audiència i dels continguts del lloc web.

El model és definit pels següents elements:

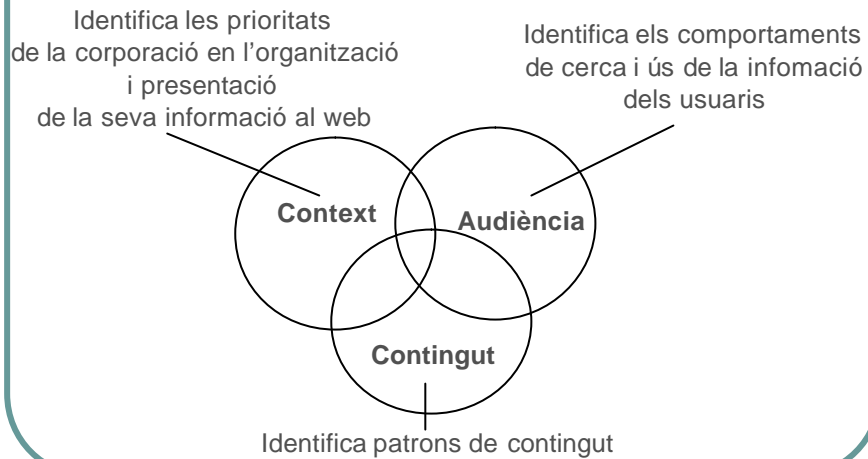
- Tipus de taxonomia
- Abast i dimensions : temàtica, usuaris , continguts i llengües .
- Usos i funcions .
- Dimensions des del punt de vista dels termes i les relacions entre els termes
- Sistema de gestió i d'actualització
- Recursos necessaris : costos econòmics i materials , humans i de temps .

Planificació estratègica

Processos

- Anàlisi de la situació de partida.
- Anàlisi del context.
- Anàlisi de l'audiència.
- Anàlisi dels continguts.
- Existeix alguna alternativa a la construcció de la taxonomia ex novo?

Planificació estratègica



Planificació estratègica

Anàlisi de la situació de partida

- Descobriment de l'escenari en què es desenvolupa el projecte de taxonomia.
- Motivacions del projecte:
 - Ja existeix un o diversos llenguatges controlats, però no acaben de funcionar o no s'adapten bé a noves aplicacions.
 - No existeix cap llenguatge controlat: s'ha de crear de nou.
- Identificació de les persones i unitats implicades.
- Resultats:
 - Contacte amb les persones i unitats implicades per a crear un marc de confiança i col·laboració entre les diferents unitats implicades.
 - Inventari de llenguatges controlats existents.

Planificació estratègica

Anàlisi del context

- Definició de la corporació:
 - Tipologia.
 - Àrea d'activitat.
 - Missió, metes i objectius.
 - Dimensions: membres, clients/usuaris, facturació...
 - Estructura organitzativa.
- Definició del lloc web:
 - Tipologia.
 - Missió, metes i objectius.
 - Dimensions: visites, transaccions, continguts...
 - Estructura organitzativa.
- Resultat: coneixement de la corporació i del lloc web.

Planificació estratègica

Anàlisi del context

- Tècniques, recursos i eines informàtiques:
 - Comparació.
 - Entrevistes amb les unitats implicades.

Planificació estratègica

Tipologia de llocs web

- Llocs web accessibles a través de xarxes privades: intranets i extranets.
 - Les intranets han evolucionat en els portals corporatius que poden ser: portals orientats a la presa de decisions, portals orientats al treball col·laboratiu o portals orientats a ambdues funcions.
 - Extranets empresa-empresa, extranets empresa-consumidor...
- Llocs web accessibles a través d'Internet.
 - Llocs principalment informatius.
 - Llocs que incorporen elements d'interacció amb l'usuari.
 - Llocs principalment transaccionals.
 - Portals de model transversal: portals públics horitzontals (*Yahoo!* <http://www.yahoo.com>) o portals verticals o vortals (*vLex.com* <http://www.vlex.com>).

Planificació estratègica

Anàlisi de l'audiència

- Identificació de la comunitat d'usuaris i segmentació en grups.
- Identificació de necessitats, comportaments, expectatives i... frustracions dels segments d'usuaris.
- Priorització de grups d'usuaris.
- Tècniques, recursos i eines informàtiques:
 - Existeix una gran diversitat de mètodes d'anàlisi de l'audiència:
 - Indirectes:
 - Directes:
 - En aquesta fase, però, l'entrevista amb representants dels segments identificats resulta molt eficaç.

Planificació estratègica

Termes més cercats pels usuaris al cercador d'un lloc web d'una administració pública

10124 hipermapa	2554 dogc	1835 itv	1369 sanitat
4524 mossos	2297 epoca	1816 subvencions	131 mapes
3865 oposicions	2229 transit	1441 palau robert	1285 joventut
2783 mapa	2209 ensenyament	1440 camps de treball	1222 ogu
2705 calendari laboral	2034 ics	1370 bombers	1180 benestar social

Dades corresponents a novembre de 2004

Planificació estratègica

Anàlisi del contingut

- **Concepte ampli de contingut, que inclou:**
 - Dades i informació amb diferents graus d'estructuració, en diferents formats i amb diferents nivells d'accessibilitat.
 - Aplicacions o funcionalitats.
- **Eines per al coneixement dels continguts:**
 - **Si gran part del contingut ja existeix, cal utilitzar eines com:**
 - Auditoria dels continguts.
 - Inventari de continguts.
 - ...
 - Mapa del lloc web.
 - **Si gran part del contingut encara NO existeix:**
 - Llistes de desiderata.
 - No obstant, cal diferenciar la identificació de nous continguts del desenvolupament de la taxonomia.

Planificació estratègica

Anàlisi del contingut

- Cal identificar patrons, similituds:
 - Lògics: temes, confidencialitat, especialització...
 - Físics: formats, dinàmics o estàtics, estructurats o desestructurats...
- Cal identificar relacions i connexions: dependència, jerarquia, associació...
- Cal començar a crear agrupacions.

Planificació estratègica

- | | |
|--------------------------|---|
| - context de creació | - actualització i conservació |
| - definició | - responsable de la creació i manteniment |
| - contingut | - audiència |
| - condicions de validesa | - accessibilitat |
| - funcions | - format/s |
| | - exemple |

Planificació estratègica

Activitat 1: quin és el nostre escenari?

Construcció

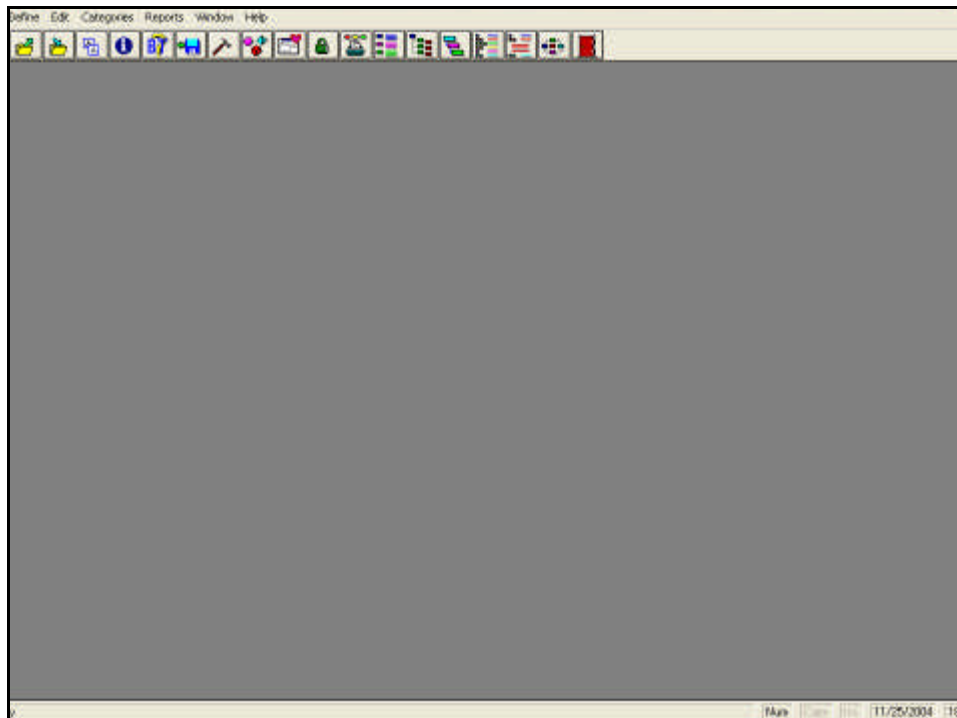
Resultat

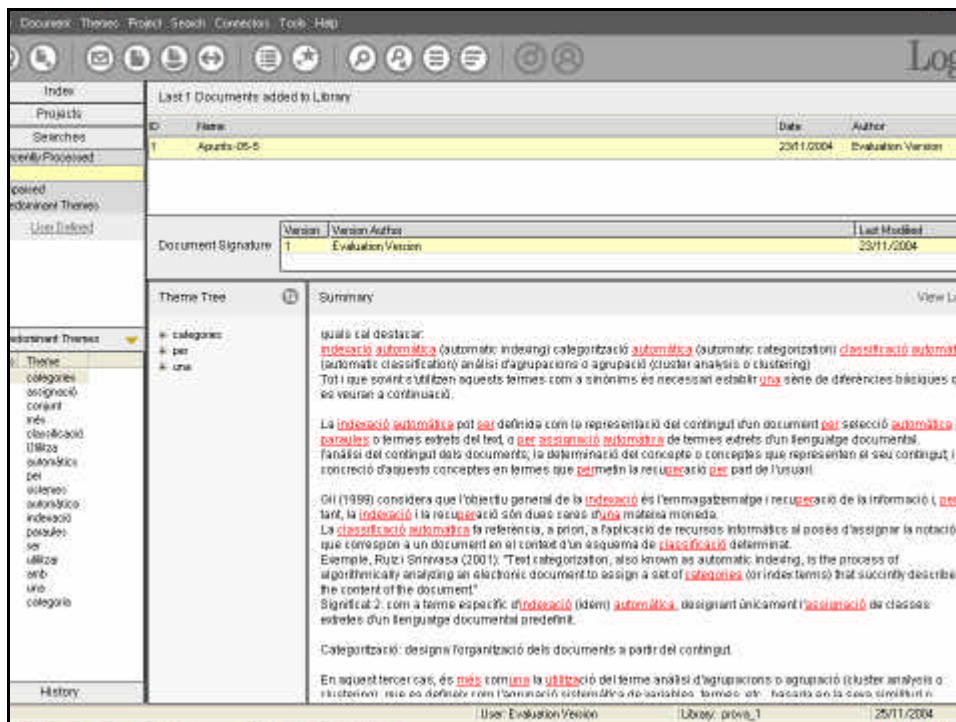
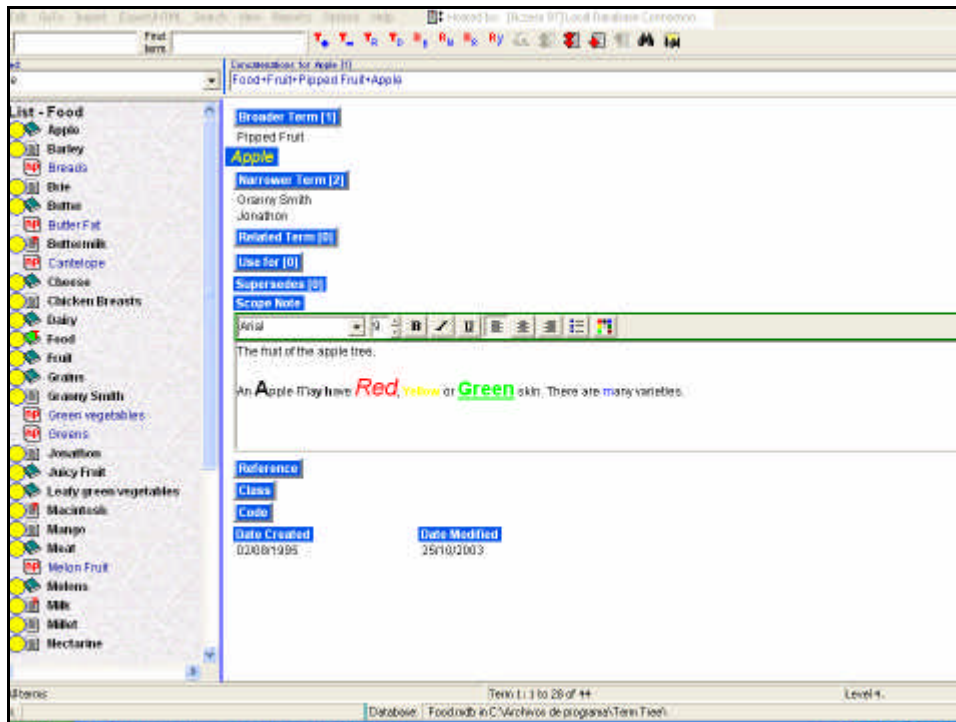
Col·lecció de termes connectats entre ells amb els tipus de relacions establerts a la definició del model, i integrada en una eina informàtica que permeti els processos posteriors de categorització, aplicació i presentació.

Construcció

Tècniques, recursos i eines informàtiques

- Mètodes "de dalt a baix" i "de baix a dalt".
- Grau d'automatització:
 - Sistema semiautomàtic. Exemples:
 - Multites <http://www.multites.com>
 - Term Tree 2000 <http://www.termtree.com.au>
 - Sistema automàtic. Exemple:
 - Comuns als sistemes automàtics de categorització.
 - Logik <http://www.coredge.com>
- Manual de gestió de la taxonomia.





Construcció

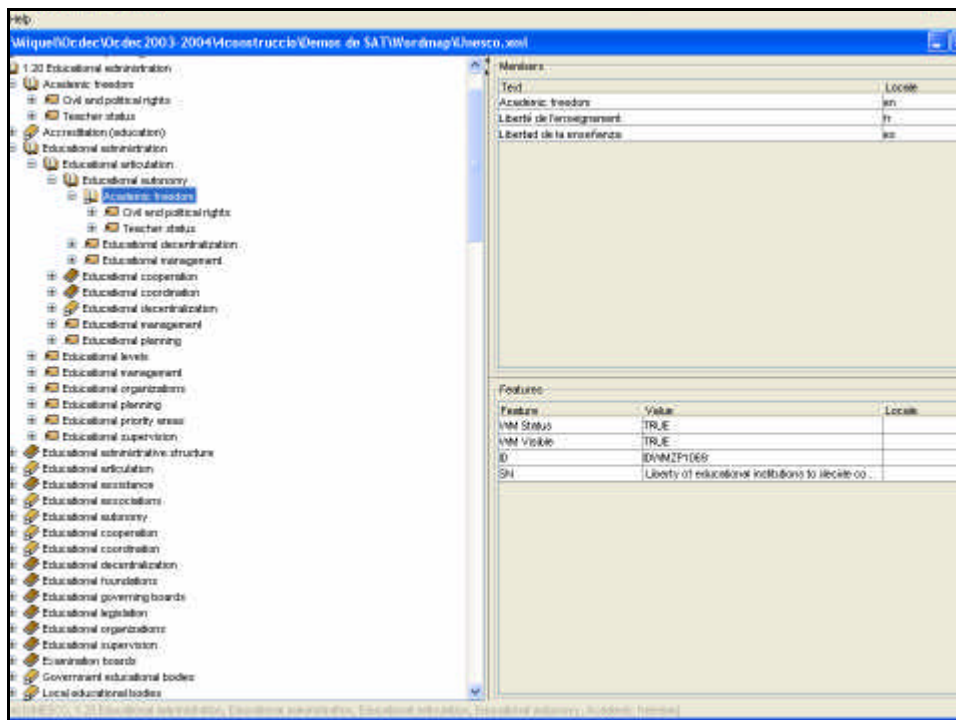
Normes, recomanacions i directrius

Quines són:

- Norma ISO 2788 (1974, 1985, 1986).
- Norma UNE 50-106-90 (1990).
- Norma ANSI/NISO Z39.19 (1994), actualment en procés de revisió
<http://www.niso.org/committees/MT-info.html>

Què regulen:

- Forma dels termes.
- Relacions entre els termes.



```

Edició Ver Favoritos Herramientas Ayuda
U:\Miquel\Odecs\Odecs2003-2004\construccio\Odecs de SAT\Wordmap\Unesco.xml
[google] Search! Send to Log... Mail
?xml version="1.0" encoding="UTF-8" ?
WORDMAP xmlns:ksi="http://www.w3.org/2001/XMLSchema-instance"
ksi:noNamespaceSchemaLocation="http://www.wordmap.com/dtd/wordmap_schema_v401.xsd"
<SOURCE name="Unesco" version="1.0" description="testing of v4 converter" fset="fset">
  <FEATURE_TYPE name="WM Basechstring" data_type="CHAR" locale_specific="0" />
  <FEATURE_TYPE name="WM Original Code" data_type="CHAR" locale_specific="0" />
  <FEATURE_TYPE name="WM Status" data_type="BOOLEAN" locale_specific="0" />
  <FEATURE_TYPE name="WM Visible" data_type="BOOLEAN" locale_specific="0" />
  <FEATURE_TYPE name="WM Start Date" data_type="DATE" locale_specific="0" />
  <FEATURE_TYPE name="WM End Date" data_type="DATE" locale_specific="0" />
  <FEATURE_TYPE name="WM Associated Term" data_type="CHAR" locale_specific="0" />
  <FEATURE_TYPE name="WM TTL" data_type="TTL" locale_specific="0" />
  <FEATURE_TYPE name="ID" data_type="CHAR" locale_specific="0" />
  <FEATURE_TYPE name="HV" data_type="CHAR" locale_specific="0" />
  <FEATURE_TYPE name="INP" data_type="CHAR" locale_specific="0" />
  <FEATURE_TYPE name="SN" data_type="CHAR" locale_specific="0" />
  <FEATURE_TYPE name="TNR" data_type="CHAR" locale_specific="0" />
  <FEATURE_TYPE name="UPD" data_type="CHAR" locale_specific="0" />
  <FEATURE_TYPE_SET name="fset">
    <FEATURE_TYPE_SET_ITEM feature_type="ID" editable="1" required="1" unique="1" key="1" cardinality="8" />
    <FEATURE_TYPE_SET_ITEM feature_type="HV" editable="1" required="0" unique="0" key="0" cardinality="M" />
    <FEATURE_TYPE_SET_ITEM feature_type="INP" editable="1" required="0" unique="0" key="0" cardinality="M" />
    <FEATURE_TYPE_SET_ITEM feature_type="SN" editable="1" required="0" unique="0" key="0" cardinality="M" />
    <FEATURE_TYPE_SET_ITEM feature_type="TNR" editable="1" required="0" unique="0" key="0" cardinality="M" />
    <FEATURE_TYPE_SET_ITEM feature_type="UPD" editable="1" required="0" unique="0" key="0" cardinality="M" />
  </FEATURE_TYPE_SET>
</SOURCE>
<DATA>
  <ROOT id="WM1" status="true" visibility="true">
    <PHRASE text="UNESCO" position="1" locale="en" />
  </ROOT>
  <WORDSET id="WM2" status="true" visibility="true">
    <PRIMARY PHYSICAL PARENT id="WM3" type="ts a" />
  </WORDSET>
</DATA>

```

Construcció

Processos

- Cal seqüenciar-los segons les tècniques aplicades.
- No obstant, l'ordre lògic és:
 - Delimitació de la realitat que cal representar.
 - Extracció del lèxic.
 - Control del lèxic.
 - Establiment del model estructural de la taxonomia

Edició Document: Herramientas Ver Ventana Ayuda

114%

Recomanacions específiques

Àmbit de les recomanacions	Recomanacions	Exemples i aclariments	Apartat de la norma UNE 50-105-00
Formes en diferents categories gramaticals	Els termes han de ser, preferentment, substantius o frases substantivades.		6.2
Forma amb aïlla i acrònim i forma desenvolupada	Es preferirà la forma desenvolupada del nom que representen.	ADS vs. Agrupacions de defensa sanitària (ADS) = Agrupacions de defensa sanitària (ADS); Referència des de les formes en sigles i acrònims.	6.2.5
Forma en singular i forma en plural	Substantius que representen ers comuns: - Quantificables, preferentment en plural. - No quantificables, preferentment en singular. Substantius que representen conceptes abstractes, preferentment en singular.	Referència des de les formes no considerades com a preferents, d'acord amb les indicacions de l'apartat 6.3.	6.3
Diferents formes ortogràfiques	Ortografia més generalment acceptada del terme: - Cal seguir les normes ortogràfiques de la llengua catalana. - Cal seguir la pràctica d'algun diccionari o glossari autoritzat: diccionari normatiu GEC, Termcat.	Eco-eficiència vs. Ecoeficiència = Ecoeficiència Referència des de les formes no considerades com a preferents, si són gramàticament correctes.	6.3.1

Recomanacions per al context de vocabulari

247 x 202,8 mm

Construcció

Delimitació de la realitat que cal representar

- Primera aproximació: identificació de tot el conjunt de components fonamentals (primer nivell) de l'organització.
 - L'organització, les competències, les funcions i els serveis de l'administració X (lloc web governamental).
 - L'organització, els membres i els recursos de docència, d'investigació i de gestió de la universitat Y (lloc web d'una universitat).
 - Els recursos de gestió, d'informació i de comunicació de la biblioteca Z (intranet d'una biblioteca).

Construcció

Delimitació de la realitat que cal representar

- Segona aproximació: de tots els components fonamentals que cal representar, quin concentra l'interès estratègic de l'organització? És el nucli de la representació.
 - Els serveis que s'ofereixen als ciutadans (lloc web governamental).
 - Els grups d'usuaris (lloc web d'una universitat).
 - Els procediments (intranet d'una biblioteca).

Construcció

Delimitació de la realitat que cal representar

- Tercera aproximació: l'anàlisi de facetes.
 - Quins aspectes del nucli de la representació permeten descriure'l de forma completa?
 - Els serveis... (lloc web governamental).
 - Els grups d'usuaris... (lloc web d'una universitat).
 - Els procediments... (intranet d'una biblioteca).

Construcció

Extracció del lèxic

- Identificació de fonts del lèxic.
- Establiment d'un ordre de prelación de les fonts.
- Eines d'extracció de lèxic:
 - Extractors lingüístics.
 - Aplicacions d'agrupament ("clustering").
 - Registres de transaccions.
- Registre dels termes.

Construcció

Extracció del lèxic

- Fonts internes:
 - Eines d'indexació i de classificació utilitzades a l'organització. Per exemple: quadre de classificació de l'arxiu.
 - Terminologies, nomenclatures i glossaris.
 - Eines d'organització i de gestió. Per exemple: organigrames i manuals de procediments.
 - Fonts personals: membres de l'organització i usuaris.
 - Directes
 - Registres de cerques realitzades pels usuaris.
 - Fonts documentals.
- Fonts externes:
 - Taxonomies externes.
 - Fonts personals: experts.

Construcció

Control del lèxic

- Diferenciació entre termes d'indexació, termes preferents i termes no preferents.
- Elecció de termes acceptats.
- Establiment de la forma dels termes.
- Establiment de la relació d'equivalència entre termes sinònims i quasisinònims.
- Desambiguació de termes.

Construcció

Establiment del model estructural de la taxonomia

- Estructures jeràrquiques.
 - La relació entre les categories es basa en la relació genèric-específica ("és un").
 - Qüestió problemàtica: la polijerarquia.
- Estructures arbòries.
 - Les categories presenten relacions sistemàtiques, però no la relació és-un.
 - Exemples: relació de successió en el temps, relació partitiva o d'inclusió, etc.

Construcció

Establiment del model estructural de la taxonomia

- Estructures facetades
 - Representació basada en la descomposició en propietats essencials.
 - Aspectes, propietats o característiques d'un ens real o abstracte, clarament definits, mútuament exclusius i, en conjunt, exhaustius.
 - Per exemple, un representant de l'objecte "Document" es pot descriure amb facetes procedents de les següents categories fonamentals: Temes, Autors, Audiències, Nivells intel·lectuals, Formes, Tipus, Llengües, Dates...

Construcció

Establiment del model estructural de la taxonomia

- Estructures facetades
 - Requisits:
 - Cal decidir en primer lloc les categories fonamentals que ens han de permetre estructurar els conceptes d'un camp determinat. (ex: en obres d'art: temps, espai, procés, material i objecte).
 - L'organització dels conceptes dins d'una categoria es pot basar en la relació que es consideri més adient (jerarquies o arbres).
 - La representació dels conceptes combina facetes vinculades a les diferents categories. (ex: nines japoneses de fusta tallada del s.XV).

Construcció

Establiment del model estructural de la taxonomia

- Estructures facetades
 - Facetes de “Servei”: Temes, Unitats responsables, Àmbits territorials, Accions, Destinataris...
 - Facetes de “Procediment”: Tema, Terminis, Unitats responsables, Normativa, Documents, Circuits...

Construcció

Activitat 2: L'ordenament de fitxes (“card sorting”) com a tècnica per a incorporar la lògica dels usuaris a la construcció de la taxonomia

Categorització

Resultat

Assignació als ítems del lloc web de termes procedents de la taxonomia.

El model de categorització està determinat per:

- Quins documents es categoritzen?
- Com es categoritzen?
 - Grau d'automatització
 - Llenguatge de marcatge de pàgines
 - Model de metaetiquetes
- Qui els categoritza?

Categorització

Normes, recomanacions i directrius

- Quines són:
 - Dublin Core
 - RDF (Resource Description Framework)
 - DAML (DARPA Agent Markup Language)
- Què regulen:
 - Metaetiquetes.
 - Marcatge semàntic de les pàgines.

Categorització

Tècniques, recursos i eines informàtiques

- Grau d'automatització.
- Llenguatge de marcatge de pàgines.
- Model de metaetiquetes.

Categorització

- XML és realment un metallenguatge per a la descripció de llenguatges de marques que proporciona un instrument per a la definició d'etiquetes i les relacions estructurals entre les etiquetes.
- Tota la semàntica d'un document XML pot ser definida o bé per l'aplicació que el processa o bé per fulles d'estil.
- On s'aplica XML?
 - En el propi contingut dels documents.
 - En les metadades per al contingut.
 - En els fitxers de la taxonomia.

Categorització

```
<?xml version="1.0"?>
<Libros>
<Libro><Titulo>Juicio universal</Titulo><Tema>Filosofía</Tema>
<Autor>Giovanni Papini</Autor><Editorial>Planeta</Editorial></Libro>
<Libro><Titulo>El crimen del padre Amaro</Titulo><Tema>Novela</Tema>
<Autor>José María Eça Queiroz</Autor><Editorial>Alianza Editorial</Editorial>
</Libro>
<Libro><Titulo>Programacion con Delphi 5</Titulo><Tema>Informatica</Tema>
<Autor>Francisco Charte
Ojeda</Autor><Editorial>Anaya Multimedia</Editorial></Libro>
</Libros>
```

Categorització

Activitat 3: comparació entre dos models de categorització semiautomàtica.

Categorització

Grau d'automatització

- Es pot concebre com un contínuum:
 - En un extrem es troben els mètodes totalment automàtics, en què una aplicació agrupa documents ("clustering") basant-se en determinats patrons del seu contingut, etiqueta les agrupacions ("clusters") i organitza aquestes agrupacions d'alguna forma per tal de crear un sistema de categories a mida.
 - En l'altre extrem es troben els mètodes totalment manuals (millor dit, intel·lectuals) en què un indexador analitza el contingut d'un document i li assigna una o més categories extretes d'un llenguatge documental controlat, o del propi document.

Categorització

Categorització automàtica

- Es fonamenta en algorismes estadístics que permeten agrupar documents que presenten similituds en el comportament de les paraules que contenen.
- La similitud es pot basar en: col·locació de les paraules, ordre de les paraules, proximitat de les paraules, freqüència de les paraules o la combinació de dos o més d'aquests principis.
- Un sistema d'agrupament ha de ser capaç de realitzar els següents processos:
 - Analitzar estadísticament les seqüències de paraules d'un document.
 - Computar el valor que representa numèricament el contingut del document
 - Comparar els valors de dos (sub)documents i determinar el seu grau de similitud.

Categorització

Categorització automàtica

Els mètodes de classificació estadística que empren les freqüències de paraules com a fonament dels algoritmes:

- Mètodes probabilístics: mètode bayesià, mètode de Rocchio...
- Mètodes vectorials: mètode K-Nearest Neighbor, Support Vector Machines...
- Arbres i llistes de decisió.

Categorització

Categorització automàtica

- Punts forts:
 - Eficàcia i rapidesa de processament.
 - Alta escalabilitat.
 - Alt nivell de consistència.
- Punts febles
 - Fragilitat de les regles/algoritmes.
 - Dificultats per a l'entrenament.
 - Alt nivell d'exactitud.
 - La lògica interna pot ser poc clara.

Categorització

Categorització automàtica

- Exemples:
 - Logik – My Edition (CoreEdge Software)
<http://www.coreedge.com>.
 - Autonomy <http://www.autonomy.com>.
 - Es basa en el mètode probabilístic baiesà.
 - Ocupa una quota de mercat del 14%.
 - Lotus Discovery Server <http://www.lotus.com>.
 - Es basa en el clustering estadístic de documents utilitzant mètodes vectorials.
 - Ocupa una quota de mercat del 7%.

Categorització

Categorització semiautomàtica o híbrida

- El fonament d'aquests sistemes és la combinació de la intel·ligència humana i l'eficiència de les màquines.
- Se n'han desenvolupat diferents tipus de sistemes que podem agrupar en quatre grans models.

Categorització

Categorització semiautomàtica o híbrida

- Sistemes que analitzen estadísticament els documents i ofereixen als experts humans opcions de categorització perquè aquests les revisin i aprovin.
- Exemple: Topic Advisor (Inktomi) <http://www.inktomi.com> (adquirit per Yahoo!).
- Categorització basada en regles de cerca.
 - Exemple: K2 (Verity) <http://www.verity.com>.
 - Clustering estadístic i regles de cerca.
 - Ocupa una quota de mercat del 15%.

Categorització

Categorització semiautomàtica o híbrida

La regla bàsica per a la categoria AUSTRALIAN DOLLAR és:

```
(if
test:
  (or [australian-dollar-concept]
and[dollar-concept]
  [australia-concept]
  (not [us-dollar-concept])
  (not [singapore-dollar-concept])))
action: (assign australian-dollar-category))
```

Categorització

Categorització semiautomàtica o híbrida

La regla bàsica per a la categoria GOLD és:

```
(if
test:
  (or (and [gold-concept :scope headline 1]
          [gold-concept :scope body 1])
       [gold-concept :scope body 4])
action: (assign gold-category))
```

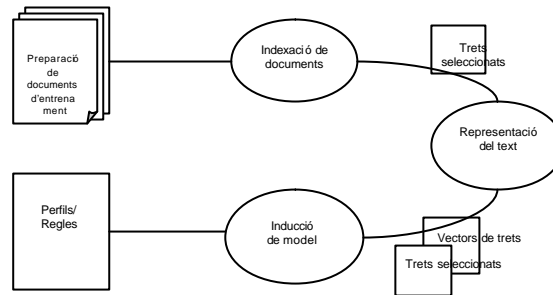
Categorització

Categorització semiautomàtica o híbrida

- Categorització basada en conjunts de documents d'entrenament o exemplars.
 - Exemple: MohoClassifier (Mohomine) <http://www.mohomine.com>.
- Categorització basada en l'anàlisi lingüística.
 - Exemple: Smart Discovery (InXight) <http://www.inxight.com>.
 - Anàlisi lingüística amb afinament estadístic.
 - Ocupa una quota de mercat del 4%.

Categorització

Categorització semiautomàtica o híbrida



Categorització

Categorització semiautomàtica o híbrida

- Punts forts:
 - Bon equilibri entre eficiència i exactitud.
 - Procés conduït pel raonament humà.
 - Capacitat d'acumular i generar aprenentatge.
- Punts febles:
 - Exigència de gestió del sistema.
 - Exigència d'habilitats especials.
 - Exigència d'esforços de manteniment.

Categorització

Grau d'automatització

- Què diuen les empreses?
 - En una enquesta realitzada per Delphi Research el 2004, a la pregunta sobre l'estratègia d'implementació de la taxonomia, 300 grans empreses de tot el món (el 60%, nordamericanes) van donar les següents respostes:
 - el 36%, híbrida;
 - el 26%, automàtica;
 - el 23%, manual;
 - la resta, o altres opcions o no van respondre

Font: *Information intelligence* 2004 (p. 26).

Aplicació i presentació

Resultats

Assolir el màxim nombre d'aplicacions de ta taxonomia en el lloc web (reutilització) per tal d'incrementar la consistència i millorar la imatge corporativa, i rendibilitzar els recursos invertits en el seu desenvolupament.

Utilitzar criteris i recomanacions d'usabilitat i d'accessibilitat en les diferents aplicacions de la taxonomia.

Aplicació i presentació

Reutilització en diferents àmbits del lloc web

Arquitectura de la informació

Facilita el disseny de sistemes de cerca d'informació i millora la seva eficàcia:

- Sistemes d'exploració o "browsing" (directoris temàtics, directoris alfabètics, etc.)
- Sistemes de recuperació (cercadors)
- Sistemes de filtratge (DSI)

Facilita el disseny de sistemes de navegació de la informació.

Facilita la interrelació dels diferents sistemes de cerca i dels sistemes de navegació.

Posicionament als cercadors

Dos criteris fonamentals en els algorismes de ponderació són la localització i la freqüència de les paraules clau a les pàgines.

El posicionament pot millorar si s'utilitzen termes de la taxonomia no únicament a les metaetiquetes, sinó també als dominis i directoris; als títols, als resums, als primers paràgrafs de les pàgines, etc.

Aplicació i presentació

Reutilització en diferents àmbits del lloc web

Altres àrees d'aplicació

Personalització de serveis.

Creació de comunitats virtuals.

Mineria de dades.

Aplicació i presentació

Presentació

Forma de la presentació:

- Textual: ordenament alfabètic, sistemàtic...
- Gràfica: arbres, xarxes...
- Metafòrica.

Condicions relatives a l'esquema de classificació:

- Esquemes purs: un sol criteri de divisió en categories.
- Esquemes híbrids: diferents criteris de divisió en categories.

Aplicació i presentació

Presentació

Condicions relatives a les estructures jeràrquiques.

- Amplada màxima recomanable:
 - Nombre de categories diferents en un mateix nivell de la jerarquia.
 - Tenint en compte les capacitats de memòria, atenció i visió de l'usuari estàndard, es recomana entre 7 i 12.
- Profunditat màxima recomanable:
 - Nombre de nivells de la jerarquia.
 - Tenint en compte les capacitats de memòria, atenció i visió de l'usuari estàndard, es recomana un màxim de 4.

Cicle d'avaluació i millora contínues. Difusió

Resultat

Sistema que assegurí que el disseny i implementació de la taxonomia s'han desenvolupat tenint en compte criteris de qualitat, que s'adapta als canvis de context, audiència i continguts, i que és coneguda i utilitzada de forma intensiva pel públic objectiu.

Aquest sistema inclou:

- Un pla d'avaluació.
- Un pla de manteniment i actualització.
- Un pla de difusió.

Cicle d'avaluació i millora contínues. Difusió

Pla d'avaluació

Quan:

- Durant el període de disseny.
- Abans del llançament.
- Després de la implementació.

Com:

- Els objectius i funcions de la taxonomia.
- Els components que són objecte d'avaluació (formulari d'avaluació).
- Els mètodes que s'apliquen a l'avaluació.
- Els indicadors o mesures de valoració.

Cicle d'avaluació i millora contínues. Difusió

Els objectius i funcions de la taxonomia

- Objectius: reutilització i consistència en els processos de categorització i recuperació de la informació.
- Funcions:
 - Integració de diferents fonts i repositoris d'informació.
 - Millora del posicionament als cercadors.
 - Desenvolupament de sistemes de navegació i *browsing* (exploració).
 - Desenvolupament de sistemes de recuperació (cercadors).
 - Desenvolupament de sistemes de filtratge: serveis de personalització dels continguts, canals per a comunitats virtuals, etc.

Cicle d'avaluació i millora contínues. Difusió

Els components que són objecte d'avaluació

- Grau d'exhaustivitat (abast) i grau d'especificitat.
- Estandardització.
- Construcció.
 - Control dels termes.
 - Estructura.
- Presentació.
- Aplicacions i ús.
- Organització i gestió.

Cicle d'avaluació i millora contínues. Difusió

Els mètodes que s'apliquen a l'avaluació

- Mètodes indirectes:
 - Anàlisi de registres transaccionals ("transactional logs").
 - Anàlisi del procés de categorització amb la utilització de la taxonomia: temps i facilitat d'ús.
 - Anàlisi del procés de recuperació d'informació amb la utilització de la taxonomia: temps, rellevància i facilitat d'ús.
- Mètodes directes:
 - Observació.
 - Enquestes.
 - Proves d'usabilitat.

Cicle d'avaluació i millora contínues. Difusió

Els indicadors o mesures de valoració

- Nivell d'ús.
- Nivell de satisfacció.
- Retorn sobre la inversió (ROI).

Cicle d'avaluació i millora contínues. Difusió

Pla de millora contínua

- Manteniment de la xarxa d'experts utilitzada en el projecte inicial.
- Sistema de presa de decisions sobre els canvis: qui els proposa i qui els valida.
- Sistema de comunicació de les decisions: qui, com i a qui es comuniquen.

Cicle d'avaluació i millora contínues. Difusió

Activitat 4: establiment d'elements d'un pla d'avaluació.