

7th NKOS workshop, Aarhus, 19 September 2008

Potential of freely faceted classification for knowledge retrieval and browsing

Claudio Gnoli

University of Pavia, Italy

Classification



Mr. Dewey

has a reputation of old KOS,
aimed at pre-digital applications...

...but can be applied to
indexing and extracting
any information
from a database !

Navigatore Dewey: Cliccare sulla descrizione o sul numero

CLASSI PRINCIPALI

- ↳ Classe: [LINGUAGGIO](#) (14108 notizie collegate notizie co
- ↳ Divisione: [LINGUE GERMANICHE \(TEUTONICHE\)](#)
- ↳ Sezione: [ALTRE LINGUE GERMANICHE](#) (

CDD	Descrizione
439.7341	LINGUA SVEDESE. Dizionari francesi
439.7351	LINGUA SVEDESE. Dizionari in italiano
439.7361	LINGUA SVEDESE. Dizionari in spagnolo
439.78	LINGUA SVEDESE. Uso standard (Linguistica)
439.81321	LINGUA DANESE. Dizionari in inglese
439.81351	LINGUA DANESE. Dizionari in italiano
439.82321	LINGUA NORVEGESE (BOKMAL, RIKSMAL). D
439.82351	LINGUA NORVEGESE (BOKMAL, RIKSMAL). D
439.9	LINGUE GERMANICHE ORIENTALI
439.91	LINGUE GERMANICHE ORIENTALI

Classification

unlike other KOSs,
allows for **systematic browsing**
("helpful sequences").

Ontology researchers
don't look much interested
in **sorting**,

though this is essential to
find **relevant** information
(e.g. Google Page Rank)

590	Religion. Theology
590A	Theory and philosophy of
590A1	Schools of theology of
590A2	Concepts in religion.
590A3	The Holy. The sacred
590A32	Definition
590A33	Origin
590A34	Nature. Taboo, he
590A35	Representation. S
590A352	Individual sym
590A353	Letters and nu
590A36	Manifestation of th
590A362	Form of the su
590A363	Dynamism. Pr
590A364	Fetishism. Wo
590A366	Manism. Mana
590A367	Animism
590A368	Totemism
590A37	Objects of worshi
590A38	Humans and parts

Classified browsing

Systematic sorting is functional...

...as compared to alphabetical

A engagement

B marriage

C separation

D divorce

divorce

engagement

marriage

separation

(from a real case while designing
a government website)

Frequent limitations to digital exploitation. 1

Notation not expressive of hierarchy:

A animals
B birds
C mammals
CA carnivores
CB rodents
...

⇒ truncated queries are problematic:
how to search for “any animal” ?

Frequent limitations to digital exploitation. 2

Single concept = different notations depending on discipline

546	inorganic chemistry
546.72	iron
553	economic geology
553.3	iron and manganese ores
669	metallurgy
669.1	ferrous metallurgy
67	industries, trades and crafts
672	articles of iron and steel

⇒ multidisciplinary search is difficult:
how to search for “anything related to iron” ?

Current efforts in UDC to improve this

Integrative Level Classification

Developing and testing a non-disciplinary, freely faceted classification scheme

www.iskoi.org/ilc

main classes: *(select + to expand)*

a	forms	+
b	branes	+
c	spaces	+
d	particles	+
e	atoms	+
f	molecules	+
g	bodies	+
h	celestial objects	+
i	rocks	+
j	landforms	+
k	genes	+
l	cells	+
m	organisms	+
n	populations	+
o	instincts	+
p	ideas	+
q	signs	+
r	languages	+
s	communities	+

search by verbal caption:

browse by notation:

browse facets of category:

0	under aspect	+
1	at time	+
2	in place	+
3	through process	+
4	made of element	+
5	with organ	+
6	from origin	+
7	to destination	+
8	like pattern	+
9	of kind	+

browse special classes:

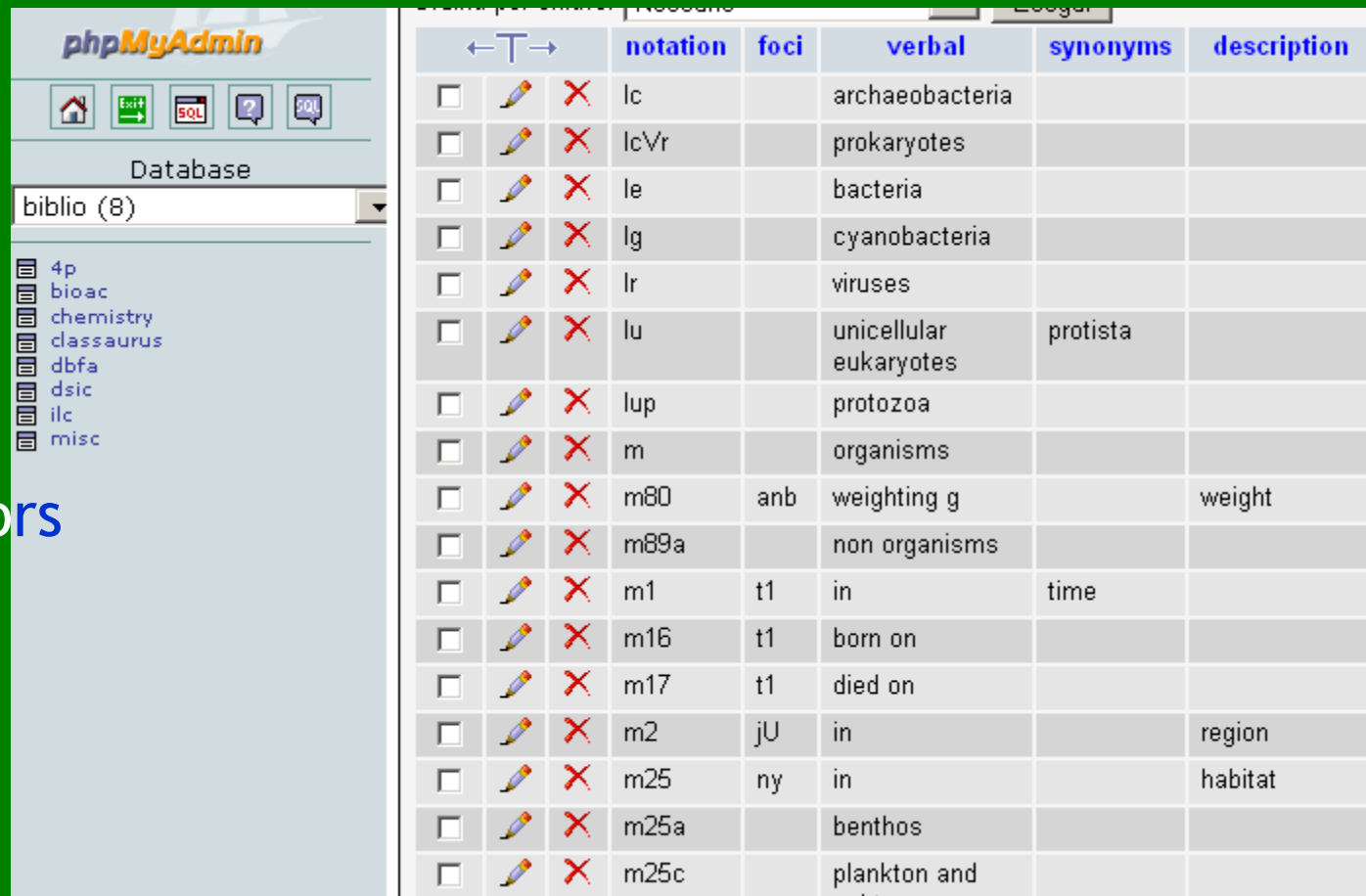
A	those	+
B	those first	+
C	those second	+
DVT	(etc.)	
U	the typical	+
..		

Integrative Level Classification

Each concept has constant notation, combinable with any other
⇒ Notation acts as **unique concept identifier**

recorded in
MySQL with:

- verbal caption
- synonyms
- facets
- semantic factors



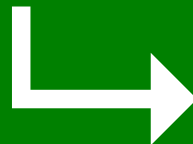
The screenshot shows the phpMyAdmin interface. On the left, the 'Database' dropdown is set to 'biblio (8)'. Below it, a list of tables is visible: 4p, bioac, chemistry, classaurus, dbfa, dsic, ilc, and misc. The main area displays a table with the following columns: notation, foci, verbal, synonyms, and description. The table contains 16 rows of data, each with a checkbox, an edit icon, and a delete icon in the first three columns.

			notation	foci	verbal	synonyms	description
<input type="checkbox"/>			lc		archaeobacteria		
<input type="checkbox"/>			lcVr		prokaryotes		
<input type="checkbox"/>			le		bacteria		
<input type="checkbox"/>			lg		cyanobacteria		
<input type="checkbox"/>			lr		viruses		
<input type="checkbox"/>			lu		unicellular eukaryotes	protista	
<input type="checkbox"/>			lup		protozoa		
<input type="checkbox"/>			m		organisms		
<input type="checkbox"/>			m80	anb	weighting g		weight
<input type="checkbox"/>			m89a		non organisms		
<input type="checkbox"/>			m1	t1	in	time	
<input type="checkbox"/>			m16	t1	born on		
<input type="checkbox"/>			m17	t1	died on		
<input type="checkbox"/>			m2	jU	in		region
<input type="checkbox"/>			m25	ny	in		habitat
<input type="checkbox"/>			m25a		benthos		
<input type="checkbox"/>			m25c		plankton and		

Integrative Level Classification

Retrieval and sorting are made through notation,
although users can interact through verbal captions

search by verbal caption:



▽ Q	nynVy	terrestrial ecosystems
▽ Q	o43n	nervous system [neuroethology] < mq5n
▽ Q	oy7n	sexual system
▽ Q	r3w	writing system, symbol
▽ Q	t8s	welfare, welfare system
▽ Q	tn5o	court, judiciary, judicial system [law]
▽ Q	u	economies, economical systems
▽ Q	u6y	pension systems
▽ Q	ukf	feudal systems
▽ Q	um	mixed economies, compromise systems
▽ Q	umn	mercantilism, mercantile systems
▽ Q	up	planned economies, hands-on systems
▽ Q	v	cultures, technological systems [applied sciences] < u sv
▽ Q	wtqv	VHS, Video home system
▽ Q	wtsh	public address system
▽ Q	y49ma	essay, non systematic exposition
▽ Q	y49mr	treatise, systematic exposition
▽ Q	y49xs	subject index, knowledge organization system [

Applications

different domains

- chemistry
- bioacoustics
- government
- traditional culture
- facet analysis

different resource types

- bibliographies
- web directory
- website architecture

The screenshot shows a web browser window with the URL <http://www.iskoi.org/ilc/4p/index.en.php>. The page title is "Where the Apennine begins" and the main heading is "Books and articles".

Below the heading, there are two search options:

- expand a class:* A list of classes with expandable buttons (red circles with a plus sign):
 - A Polcevera valley
 - B Bisagno valley
 - C Scrivia valley
 - D Spinti valley
 - E Borbera valley
 - F Ossonova valley
 - G Grue valley
 - H Curone valley
 - J Staffora valley and Oltrepò
 - K Tidone valley
 - o Trebbia valley
 - Q Nure valley
 - S Aveto valley
 - T Fontanabuona valley
 -
 - g bulk matter
 - h rocks
 - j territory
 - m organisms
- or search the classes corresponding to the subject:* A search box and an "enter" button.
- or search directly by name of an author:* A search box and an "enter" button.
- or search directly by words in the title:* A search box and an "enter" button, with a note: "(up to 5 words separated by blanks)".

At the bottom right, there is a note: "The bibliographic database currently consists of 580 records. The scheme is arranged according to the Integrative level classification."

Search by class

User browses classes
and selects one:



Where the Apennine begins

Books and articles

Expanded class

On class **m** depend these classes. *Please choose the code of a class in order to display the corresponding documents.*

<u>m</u>	organisms
<u>m81y</u>	fossils
<u>mf</u>	mushrooms
<u>mp</u>	plants
<u>mpe</u>	conifers
<u>mpf</u>	flowering plants
<u>mpfag</u>	orchids
<u>mpfbl</u>	olives
<u>mpfdg</u>	vines
<u>mpfmyc</u>	chestnuts
<u>mpfmyf</u>	beeches
<u>mq</u>	animals
<u>mqm</u>	mollusca
	arthropoda

Classified retrieval

Occurrences of the class combined with any other are displayed

and sorted by class

Results of the search by class

To the class **flowering plants** correspond these 10 documents. *In order to extract only those a click on the corresponding term.*

mpf	flowering plants	<i>I nostri fiori / G De Maria, G Meriana – SAGEP : Genova</i>
mpf	flowering plants	<i>Il mondo dei fiori / G De Maria – SAGEP : Genova : 1988</i>
mpf hs h1e JA	flowering plants : sedimentary rocks : Staffora valley	<i>Un'angiosperma fossile nel cretaceo dell'Appennino pavese / Vittadini = (Atti della Società italiana di scienze naturali)</i>
mpfag E H	orchids : Borbera valley : Curone valley	<i>Le orchidee dell'Appennino alessandrino / M Carrega – Gr</i>
mpfag H JA	orchids : Curone valley : Staffora valley	<i>Nota preliminare sulle specie di Orchidaceae delle valli Stafforine / Abeli = Quaderni del Civico museo di scienze naturali di V</i>
mpfag O	orchids : Trebba valley	<i>Le orchidee spontanee del Piacentino / L Bongioni – Amn Piacenza : 1989</i>
nyr mpfmyf jm S	woods : beeches : mountains : Aveto valley	<i>La foresta del monte Penna / A Cotta = (Il legno. 1938.06</i>
rD r5 vm mpfdg J	Lombardian dialects : words : agriculture : vines : Staffora valley and Oltrepò	<i>La parlata di Portàlbera e la terminologia vinicola nell'Oltrepò / dell'Istituto di Glottologia) = Studi e ricerche. 5 – Zuffi : Bo</i>
vm mpfbl mpfmyc	agriculture : olives : chestnuts	<i>Il castagno e l'ulivo / Giovanni Meriana – SAGEP : Genova</i>
vm mpfmyc J	agriculture : chestnuts : Staffora valley and Oltrepò	<i>Il castagno nell'Oltrepo pavese / Gianluca Gaiani, Silvano Lombardia. Azienda regionale delle foreste : 1999</i>

Global vs. local

Universal KOSs help interoperability,
as each concept is uniquely identified:

mpf **flowering plants**

But, within a domain,
specialized concepts with long notations
can be frequent:

jUe	Europe
jUei	Italian peninsula
jUeip	Apennine
jUeipg	Ligurian Apennine
jUeipg	Antola chain
jUeipgh	Curone valley

Global vs. local

To conceal the two,
we need to specify both global and local meanings
by an **AR-complex** [Wåhlin 1971]

We state:

H = jUeipgh

Curone valley

⇒ **mpf H**

flowering plants : Curone valley (free cl.)

mpf2H

flowering plants, in Curone valley
(freely faceted cl.)

AR-complex

mpf H flowering plants : Curone valley

- lower-case letters are from the Reference system
- upper-case letters are from the Adapted system
(*deictics*: meaning changes according to context)

In ASCII, upper-case letters file before,
thus local concepts will be listed before
(*favoured host class* [Ranganathan 1967])

A	Polcevera valley
B	Bisagno valley
C	Scrivia valley
D	Spinti valley
E	Borbera valley
F	Ossona valley
G	Grue valley
H	Curone valley
J	Staffora valley and Ot...
K	Tidone valley
O	Trebbia valley
Q	Nure valley
S	Aveto valley
T	Fontanabuona valley
-	
g	bulk matter
h	rocks
j	territory
n	organisms

AR-complex

Local KOSs can interoperate with a global KOS by mapping their deictics on it:

H = jUeipgh Curone valley

Mapping can be more complex if the two KOSs have different hierarchies:

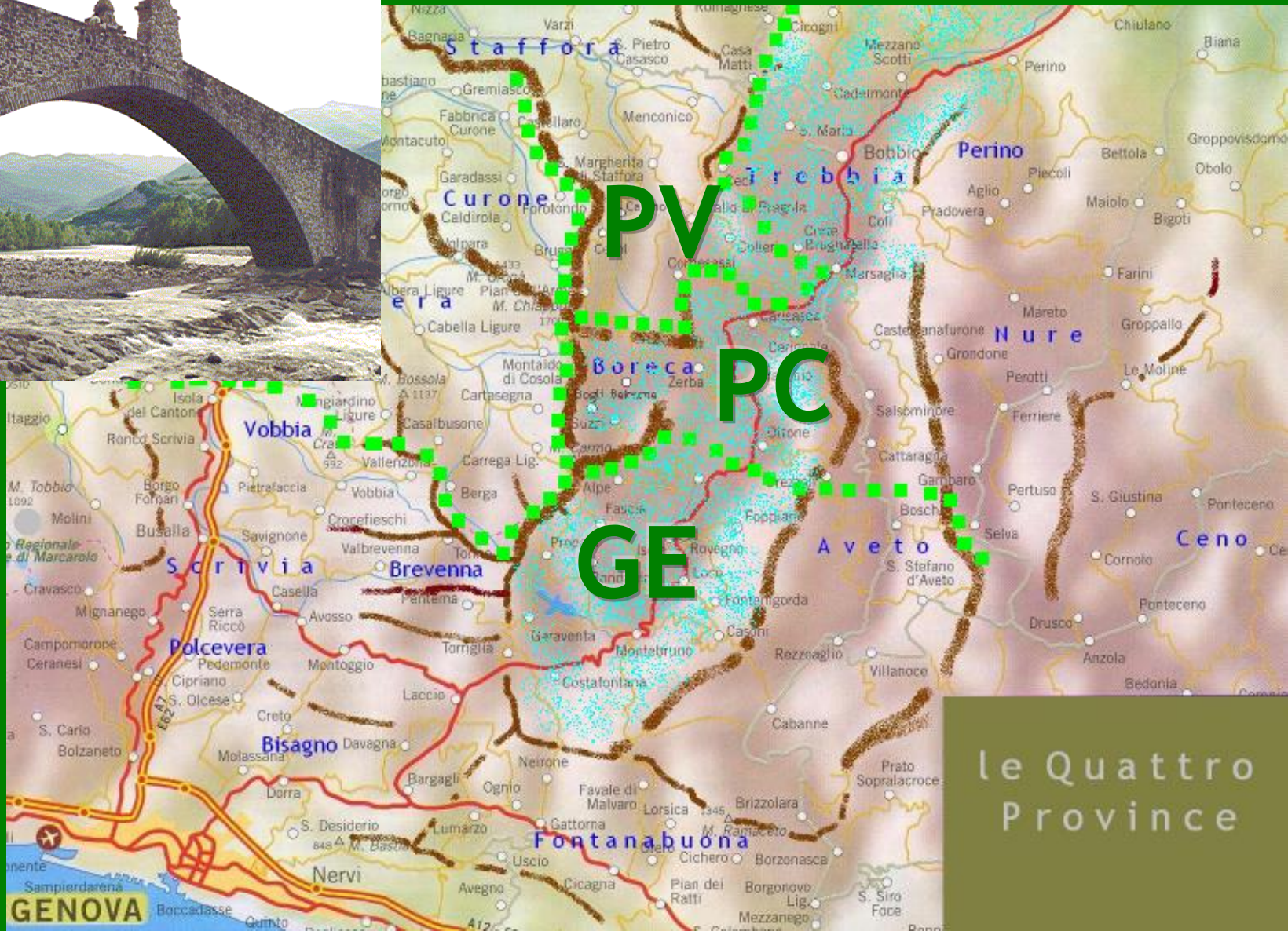
O = jUeipgo + jUeilgau + ...

Trebbia valley, being partly in Genova province (Liguria),
partly in Piacenza province (Emilia),
and a little section in Pavia province (Lombardia)

Mapping different hierarchies



H
Trebbia
valley



le Quattro
Province

Retrieval of a facet

IR systems can retrieve a facet
from any position within a string:

mpf H flowering plants : **Curone valley**

Still, citation order is crucial for browsing:

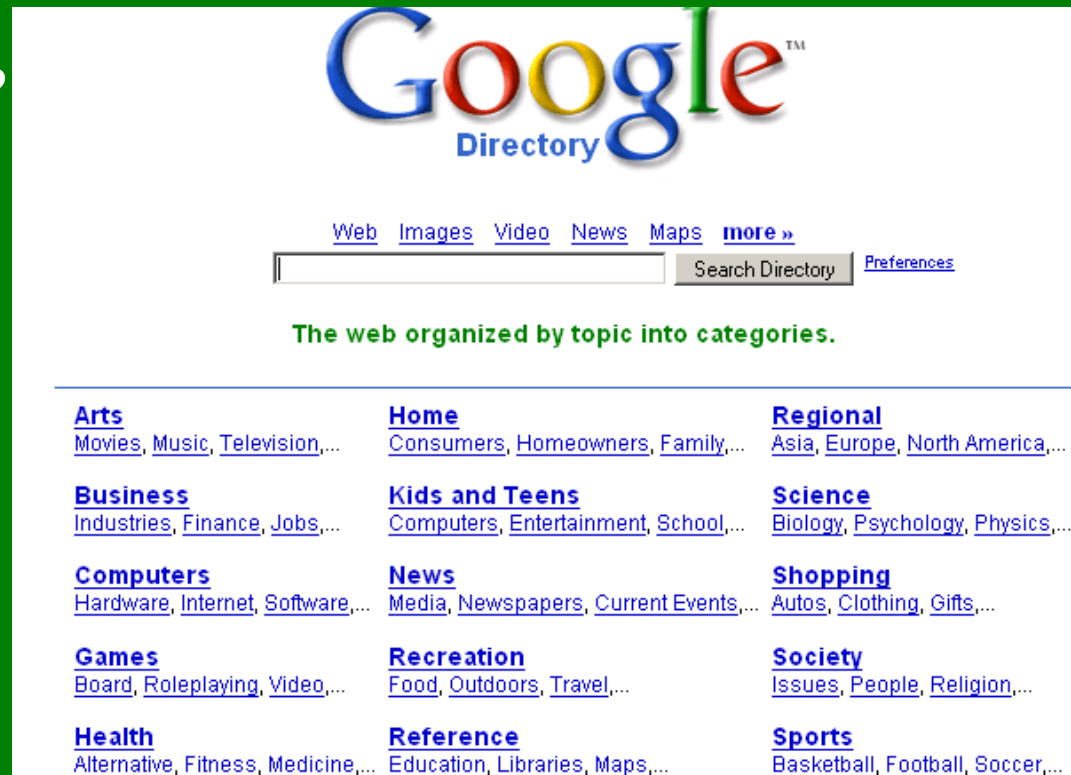
H Curone valley
mpf H flowering plants : Curone valley
nyr H woods : Curone valley
nyr mpf H woods : flowering plants : Curone valley

Freely faceted classification

is powerful for retrieval, sorting, browsing...

...so why is it poorly used?

Most Internet directories prefer alphabetical sorting:



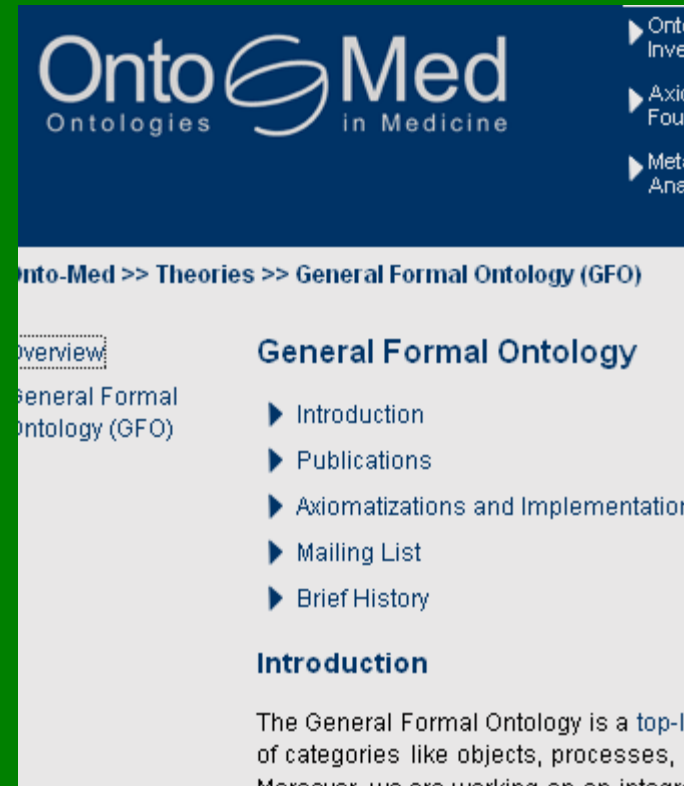
The screenshot shows the Google Directory homepage. At the top is the Google logo with 'Directory' underneath. Below the logo are navigation links for 'Web', 'Images', 'Video', 'News', 'Maps', and 'more »'. There is a search bar with a 'Search Directory' button and a 'Preferences' link. Below the search bar is the tagline 'The web organized by topic into categories.' The main content is a grid of 12 topic categories, each with a bold title and a list of sub-topics.

Arts Movies , Music , Television ,...	Home Consumers , Homeowners , Family ,...	Regional Asia , Europe , North America ,...
Business Industries , Finance , Jobs ,...	Kids and Teens Computers , Entertainment , School ,...	Science Biology , Psychology , Physics ,...
Computers Hardware , Internet , Software ,...	News Media , Newspapers , Current Events ,...	Shopping Autos , Clothing , Gifts ,...
Games Board , Roleplaying , Video ,...	Recreation Food , Outdoors , Travel ,...	Society Issues , People , Religion ,...
Health Alternative , Fitness , Medicine ,...	Reference Education , Libraries , Maps ,...	Sports Basketball , Football , Soccer ,...

Faceted ontologies?

Integration between
ontology and FC
is rare
(ongoing research in DRTC, GFO)

implying a high risk of
reinventing the wheel.



The screenshot shows the OntoMed website interface. The header features the logo "OntoMed" with the tagline "Ontologies in Medicine". A navigation menu on the right includes links for "Onto", "Inve", "Axi", "Fou", "Met", and "Ane". The main content area displays the breadcrumb "Onto-Med >> Theories >> General Formal Ontology (GFO)". Below this, there is a section titled "General Formal Ontology" with a list of links: "Introduction", "Publications", "Axiomatizations and Implementation", "Mailing List", and "Brief History". A sub-section titled "Introduction" begins with the text: "The General Formal Ontology is a top-1 of categories like objects, processes, Moreover, we are working on an inter".

Reasons for under-exploitation of FC

- poorly known, nor taken seriously enough in LIS schools;
- not considered by designers of interfaces, even if known by librarians cooperating in the same projects;
- more complex than other KOSs, thus requiring initial investment;
- convincing examples hardly available on the Internet...


Complex KOSs

need several layers working together:

- conceptual structure
- local schemes (deictics)
- notation
- verbal captions in natural languages
- database management
- indexing interface ← *indexer-friendly!*
- user interface
- ...

Indexing interface

The indexer
can edit the
classmark and
dynamically
see the caption
she is producing



BioAcoustic Reference Data

Record editing

Lammers M.O., Au W.W.L. 2003

Directionality in the whistles of Hawaiian spinner dolphins (*Stenella longirostris*)

Marine Mammal Science, 19 (2), 249-264
[74.1001 PDF, id 2]

[look in ILC](#) [update ILC](#)

signs, by spinner dolphin, localisation, acoustic, some direction, possi

Indexing interface

She can be helped by automatic suggestions generated by matching title with DB thesaurus

BioAcoustic Reference Database

Suggested classes

Lammers M.O., Au W.W.L. 2003

Directionality in the whistles of Hawaiian spinner dolphins (*Stenella longirostris*): a signal feature

Marine Mammal Science, 19 (2), 249-264
[74.1001 PDF, id 2]

signs, by spinner dolphin, localisation, acoustic, some direction, possibility

mqrifze	whospinners	embioptera
mqvtnisl	spinner dolphin	<i>Stenella longirostris</i>
mqvtni	delphinidae	Dolphin
mqvtnicc	Commerson's dolphin	<i>Cephalorhyncus commersonii</i>
mqvtnice	Chilean dolphin	<i>Cephalorhyncus eutropia</i>

document title →

edited notation →

automatic caption →


suggested classes →

Work in progress

to produce the convincing examples
(FACET docet).

Strategies:

- produce user-friendly interfaces
- through open source software (MySQL, PHP)
- make results freely available on the Internet
- provide links to very popular resources (eg Wikipedia) ?
- join FFC with some Web 2.0 resource?



The screenshot shows a web page for the Hypermedia Research Unit. The page title is "The FACET Project". Below the title is a link: "[Back to Knowledge Organisation S...". A list of nine items is displayed, each with a number and a blue underlined link:

1. [Background](#)
2. [Objectives](#)
3. [Beneficiaries](#)
4. [Facet Architecture and Inter](#)
5. [Multi Concept Matching Fun](#)
6. [Faceted Query Editor and S](#)
7. [Semantic Browsing](#)
8. [Semantic Closeness](#)
9. [Web demonstrator](#)

Below the list, the following information is provided:

Dates: April 2000 - March 2003
Principal Investigator: Douglas T
Co-Investigator: Daniel Cunliffe
Research Associates:

On the left side of the screenshot, there is a book cover titled "Hypermedia and Multimedia" with a blue background and white text.

ILC people: Claudio Gnoli, Mela Bosch, Enzo Cesanelli, Philippe Cousson, Viviana Doldi, Hong Mei, Gabriele Merli, Marcella Patania, Roberto Poli, Rick Szostak, Lorena Zuccolo

Published reports:

Gnoli & Poli 2004, Levels of reality and levels of representation, Knowl org 31, 3, 151-160

Gnoli & Merli 2005, Notazione e interfaccia di ricerca per una classificazione a livelli, AIDA informazioni, 23, 1-2, 57-72

Hong 2005, A phenomenon approach to faceted classification, 53th conf Japan Soc LIS

Gnoli 2006, The meaning of facets in nondisciplinary classifications, proc 9th ISKO conf, Vienna, 11-18

Gnoli & Hong 2006, Freely faceted classification for Web-based information retrieval, New rev hypermedia & multimedia, 12, 1, 63-81

Gnoli, Bosch & Mazzocchi 2007, A new relationship for multidisciplinary knowledge organization systems: dependence, proc 8th ISKO Spain conf, León, 399-409

Gnoli 2007, “Classic” vs. “freely” faceted classification, ISKO UK meeting Ranganathan revisited, London

Gnoli 2008, Freely faceted classification for a Web-based bibliographic archive: the BioAcoustic Reference Database, proc ISKO D conf, Konstanz

Szostak & Gnoli 2008, Classifying by phenomena, theories, and methods, proc 10th ISKO conf, Montréal

Thank you!



Comments welcome:

gnoli@aib.it

www.iskoi.org/ilc