

# Sorting documents by base theme with synthetic classification: the double query method

*Claudio Gnoli & Alberto Cheti*



# Knowledge organization A to Z ?...



**Friday**  
**Monday**  
**Sathurday**  
**Sunday**  
**Thursday**  
**Tuesday**  
**Wednesday**

# Knowledge organization A to Z ?...

**A solution:**

**good old  
classification :-)**

- 1 Sunday**
- 2 Monday**
- 3 Tuesday**
- 4 Wednesday**
- 5 Thursday**
- 6 Friday**
- 7 Sathurday**

# Knowledge organization A to Z ?...

**Systematic presentation can act as an intellectual guide to contents**

expand all | collapse all

TOP

- AUXILIARY TABLES
  - + COMMON AUXILIARY SIGNS
  - + COMMON AUXILIARY NUMBERS
- MAIN TABLES
  - 0 SCIENCE AND KNOWLEDGE. ORGANIZATION. COMPUTERS
  - 1 PHILOSOPHY. PSYCHOLOGY
  - 2 RELIGION. THEOLOGY
  - 3 SOCIAL SCIENCES
  - 5 MATHEMATICS. NATURAL SCIENCES
  - 6 APPLIED SCIENCES. MEDICINE. TECHNOLOGY
  - 7 THE ARTS. RECREATION. ENTERTAINMENT. SPORT
  - 8 LANGUAGE. LINGUISTICS. LITERATURE
  - 9 GEOGRAPHY. BIOGRAPHY. HISTORY

# Knowledge organization A to Z ?...

Original German term:

*Wissensordnung*

= “ordering of knowledge”



# Classification

Often poorly applied  
in online resources...

Lack of integration  
between cataloguers'  
and OPACmasters' work

[Bland & Stoffan, 2008; Rozman, 2009;  
Casson et al. 2011]

The screenshot shows the SBN website interface. At the top, there is a dark red header with the text "SERVIZIO BIBLIOTECARIO NAZIONALE". Below the header, there is a navigation bar with links for "Glossario", "Contacts", and "Italiano". A search bar is visible on the left, and a dropdown menu is open, listing various search criteria: "Author", "Title", "Series", "Uniform title", "Publication place", "Publisher", "Subject", "Dewey classification" (highlighted in blue), "Dewey description", "Keywords", "ISBN", "ISSN", and "BID". Below the dropdown, there are input fields for search terms, with "720\_" entered in the first field. A "Perform search" button is visible on the right. The bottom of the page shows a "Filters" section.

# Compound subjects

**Most real documents are about combinations of concepts, e.g.:**

**«the corrosion of tinplate  
by acid fruit products»...**

[Foskett 1958]

→ **Synthetic classmarks needed  
(subdivisions, auxiliaries, facets, roles, links...)**

# Citation order matters

**1 : 34 «philosophy – law»**

**34 : 1 «law – philosophy»**

No. of Recs	Entry
1	1:5 WEI
4	1:5 WEY
3	1:5 WHI
1	1:5 WIL
1	1:5 ZAH
2	1:5 ZIM
1	1:5 ZUC
1	1:51 ASC
1	1:51 BOR
1	1:51 BOU



# The PRECIS-GRIS tradition

(Verbal) subject strings  
should be ordered combinations  
of terms (concepts)

Law - influence of philosophy - U.K. - dictionaries

# Base vs. particular theme

## Notions coming from text linguistics

[Beaugrande & Dressler 1981]

«Influence of the abundance of  
wild ungulates on wolf diet  
in Northern Apennines»

Wolf - diet - effect of ungulate abundance - N Apennines

# Two-step search [GRIS]

Interfaces should allow to:

-- identify **a concept**  
(finding the right term,  
discarding homographs etc.)

-- examine all **combinations**  
of it with other concepts

...starting with those  
where it is the base theme!

Bose algebras

Bose-Einstein condensation

Bose-Einstein condensation Congress

Bose-Einstein condensation -- Mathem

Bose-Einstein gas

Bose-Einstein gas -- Congresses

Bose-Einstein gas -- Mathematics

Bose, J.C. (Jagadis Chandra), Sir, 185

Bose, S. (Satyendranath), 1894-1974

# Double query method

Let's give the user  
what (s)he's asked for:

- (1) all combinations where  
the search term is the base theme
- (2) all combinations where  
the search term is a particular theme

# An application

The screenshot shows the ISKO website interface. On the left is a vertical navigation menu with links for home, about ISKO, join ISKO, Knowledge organization journal, ISKO events, ISKO chapters, ISKO people, ISKO publications, KO literature, KO institutions, and contact us. The main content area features a header for 'International Society for Knowledge Organization' and a section for 'Knowledge organization literature' with 'Recent bulletins' from 2009 to 2013. Below this is a search section titled 'Search in the cumulative database (1997-current)'. A search results list is displayed, with '757 Expert Systems in Searching. Search Engines' highlighted in blue. A 'y items' button is visible on the right side of the list.

www.isko.org/lit.html

## ISKO International Society for Knowledge Organization

[home](#)  
[about ISKO](#)  
[join ISKO](#)  
[Knowledge organization journal](#)  
[ISKO events](#)  
[ISKO chapters](#)  
[ISKO people](#)  
[ISKO publications](#)  
[KO literature](#)  
[KO institutions](#)  
[contact us](#)

### Knowledge organization literature

#### Recent bulletins

- [2013, n. 3](#) — [2013, n. 2](#) — [2013, n. 1](#)
- [2012, n. 4](#) — [2012, n. 3](#) — [2012, n. 2](#) — [2012, n. 1](#)
- [2011, n. 4](#) — [2011, n. 3](#) — [2011, n. 2](#) — [2011, n. 1](#)
- [2010, n. 4](#) — [2010, n. 3](#) — [2010, n. 2](#) — [2010, n. 1](#)
- [2009, n. 4](#) — [2009, n. 3](#) — [2009, n. 2](#) — [2009, n. 1](#)

#### Search in the cumulative database (1997-current)

75 Online Retrieval Systems and Technologies: ▼

- 6 On Special Subjects Classification Systems and Thesauri**
  - 65 On CS & T in Medicine, Psychology, Education, Labour, Sports, Household
- 7 Knowledge Representation by Language and Technology**
  - 71 General Problems of Natural Language in Relation to KO
  - 72 Semantics
  - 73 Automatic Language Processing
  - 74 Grammar Problems
  - 75 Online Retrieval Systems and Technologies:
    - 751 General and Theoretical Problems
    - 752.2 Interactive catalogues. On-line catalogues
    - 752.3 Internet gateways. Subject gateways. Access to the Internet
    - 753 On-Line Access, Queries, Navigation, Full/Free Text Searching
    - 757 Expert Systems in Searching. Search Engines**
    - 759 Evaluation of On-Line Information Retrieval Systems and Techniques
  - 77 Problems of Terminology
  - 78 Subject-Oriented Terminology Work
  - 79 Problems of Multilingual and Cross Language Systems and Translation

y items

niel Martínez-Ávila, Ot

# (1) Results as base theme

## Knowledge organization literature. Selected items

Your search for class 757 **Expert Systems in Searching. Search Engines. Intelligent Agents. Routing. SDI. Data Mining. Data Fusion. Collection Fusion. Current Awareness Services** found the following 104 items: [\[new search\]](#)

### Items on this class as the base theme

6996

search engines • 757

Lepsky, Klaus. – (Book review of) Dirk Lewandowski, Christian Maass, eds. *Web-2.0-Dienste als Ergänzung zu algorithmischen Suchmaschinen*. Berlin, Logos, 2008. ix, 189pp. ISBN: 978-3-8325-1907-0 [Web 2.0 services as a complement to algorithmic search engines] (Lang.: ger). – In: *Zeitschrift für Bibliothekswesen und Bibliographie*, 56(2008)2, pp. 134-135.

7074

search engines • 757

Blenkle, Martin, Ellis, Rachel, Haake, Elmar. – *E-LIB Bremen: Automatische Empfehlungsdienste für Fachdatenbanken im Bibliothekskatalog; Metadatenpools als Wissensbasis für bestandsunabhängige Services* [E-LIB Bremen: Automatic recommender services for subject specific databases in the library catalogue: metadata pools as the knowledge base for services unrelated to catalogue resources] (Lang.: eng). – In: *Bibliotheksdienst*, 43(2009)6, pp. 618-625. – Available at [http://www.zlb.de/aktivitaeten/bd\\_neu/heftinhalte2009/Erschliessung010609BD.pdf](http://www.zlb.de/aktivitaeten/bd_neu/heftinhalte2009/Erschliessung010609BD.pdf)

7075

search engines • 757

Summann, Friedrich, Wolf, Sebastian. – *Suchmaschinentechnologie und wissenschaftliche Suchumgebung* [Search engine technology and the environment of scholarly searching] (Lang.: eng). – In: *Online-Mitteilungen*, 86(2006), pp. 3-18.

7235

search engines • 757

Marin, M. et al. – *Sync/Async parallel search for the efficient design and construction of web search engines* (Lang.: eng). – In: *Parallel computing*, 36(2010)4, pp. 153-168.

# ...either alone or combined...

search engines • KM • 757;149

– Data mining for exploring hidden patterns between KM and its performance (Lang.: eng). – In: Knowledge-based pp. 397-401.

search engines • thesauri • 757;214

M., Matheus, C.J. – Uses of ontologies in open source blog mining (Lang.: eng). – In: Frontiers in artificial intelligence 3(2010), pp. 37-56.

search engines • descriptors • 757;226

Google Scholar out-performs many subscription databases when keyword searching (Lang.: eng). – In: Evidence based on practice, 5(2010)3, pp. 39-41.

search engines • facet analysis • 757;325

The use of facets in Web search engines (Lang.: eng). – In: Paradigm and conceptual systems in knowledge 2.23/26), pp. 349-355.

search engines • automatic classification • 757;348

Search engine optimisation and automatic classification (Lang.: eng). – In: Legal information management, 10(2010)1, pp.

search engines • terminology software • 757;773.4

T., Tang, X. – A comparative study of TF\*IDF, LSI and multi-words for text classification (Lang.: eng). – In: Expert systems, 38(2011)3, pp. 2758-2765.

search engines • special terminology • human biology • special terminology • clinical medicine • 757;78-51/4

Discovering hidden connections among biomedical concepts from disjoint biomedical literature sets through semantic-based (Lang.: eng). – In: International journal of intelligent systems, 25(2010)2, pp. 207-223.

## (2) Results as a particular theme

9097

search engines • information use • 757;981

Chaudiron, Stéphane, Ihadjadene, Madjid. – Studying web search engines from a user perspective: key concepts and main approaches (Lang.: eng). – In: Next generation search engines, pp. 411-437 (08.757).

### Further items on this class as a particular theme

8411

concept construction • search engines • facet analysis • 123;757;325

Milonas, Elizabeth. – Wittgenstein and web facets (Lang.: eng). – In: Expanding our horizons, evaluating our parameters, [See 06.11-06-16\17], pp. 33-40.

7341

KO problems in • search engines • 178;757

Lee, J.Y., Kim, H., Kim, P.J. – Domain analysis with text mining: analysis of digital library research trends using profiling methods (Lang.: eng). – In: Journal of information science, 36(2010)2, pp. 144-161.

8607

thesauri • search engines • 214;757

Yufeng, Zhang, Jiaojie, Cai. – [Research on the user interest ontology learning based on web mining technology] (Lang.: chi). – In: Journal of the China Society for Scientific and Technical Information, 30(2011)4, pp. 380-386.

8780

compound descriptors • search engines • 217;757

Baena-Garcia, M., Morales-Bueno, R. – Mining interestingness measures for string pattern mining (Lang.: eng). – In: Knowledge-based systems, 25(2012)1, pp. 45-50.

7365

weights • search engines • citation indexing • 246;757;864

Liu, Xinhai et al. – Weighted hybrid clustering by combining text mining and bibliometrics on a large-scale journal database (Lang.: eng). – In: Journal of the American Society for Information Science and Technology, 61(2010)6, pp. 1105-1119.



# Double query method

```
$queryA =  
"SELECT * FROM `literature`  
WHERE `classmark` REGEXP '^757*'  
ORDER BY classmark";
```

```
$queryB =  
"SELECT * FROM `literature`  
WHERE `classmark` REGEXP ';757*'  
ORDER BY classmark";
```

# Position depends on search

## Knowledge organization literature. Selected items

Your search for class 325 **Facet Analysis** found the following 34 items: [\[new search\]](#)

### Items on this class as the base theme

6937 facet analysis • 325  
Vechtomova, O. – Facet-based opinion retrieval from blogs (Lang.: eng). – In: Information processing & management, 46(2010)1, pp. 71-

### Further items on this class as a particular theme

7933 online systems • facet analysis • 755;325  
Frapp, Dominic. – Using linked data to classify web documents (Lang.: eng). – In: Aslib proceedings, 63(2010)6, pp. 585-595.

7237 search engines • facet analysis • 757;325  
Milonas, Elizabeth. – The use of facets in Web search engines (Lang.: eng). – In: Paradigms and conceptual systems in knowledge organization (06.10.2.23/26), pp. 349-355.

8172 tagging • facet analysis • 835;325  
Wang, Peng, Li, Ying, Wang, Lin. – [Improvement in faceted classification-based folksonomy] (Lang.: chi). – In: Zhonghua Yixue Tushu

# Conclusions

- Principles for combination in verbal indexing (base vs. particular themes) can be extended to classification
- They help users to locate what they are actually searching for among many possible results
- They can be applied to search interfaces by any script (e.g. PHP + MySQL) managing a double query

Thank you!

claudio.gnoli@unipv.it



@scritur

