UDC Seminar 2013, The Hague

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

Claudio Gnoli & Alberto Cheti
Knowledge organization A to Z ?...

Friday
Monday
Saturday
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  Saturday
Systematic presentation can act as an intellectual guide to contents
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]
Compound subjects

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

«the corrosion of tinplace 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»

<table>
<thead>
<tr>
<th>No. of Recs</th>
<th>Entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1:5 WEI</td>
</tr>
<tr>
<td>4</td>
<td>1:5 WEY</td>
</tr>
<tr>
<td>3</td>
<td>1:5 WHI</td>
</tr>
<tr>
<td>1</td>
<td>1:5 WIL</td>
</tr>
<tr>
<td>1</td>
<td>1:5 ZAH</td>
</tr>
<tr>
<td>2</td>
<td>1:5 ZIM</td>
</tr>
<tr>
<td>1</td>
<td>1:5 ZUC</td>
</tr>
<tr>
<td>1</td>
<td>1:51 ASC</td>
</tr>
<tr>
<td>1</td>
<td>1:51 BOR</td>
</tr>
<tr>
<td>1</td>
<td>1:51 BOU</td>
</tr>
</tbody>
</table>
(Verbal) subject strings should be ordered combinations of terms (concepts)

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!
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
(1) Results as base theme

Knowledge organization literature. Selected items


Items on this class as the base theme

6996

7074

7075

7235
...either alone or combined...


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

(2) Results as a particular theme


Further items on this class as a particular theme

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


$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
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

7237

8172
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

@scrittur