Afrika-Studiecentrum
FROM CLASSIFICATION TO THESAURUS ... AND BACK?

SUBJECT INDEXING TOOLS AT THE LIBRARY OF THE AFRIKA-STUDIECENTRUM, LEIDEN, THE NETHERLANDS

Marlene van Doorn & Katrien Polman
Presentation at International UDC Seminar 2009
The Hague, 29-30 October 2009
Library collection

Geographical focus: Africa (North & Sub-Saharan) & associated islands

Subject focus: social sciences & humanities
socioeconomic & political developments;
government, law & constitutional development;
history; religion; anthropology; women’s studies;
education; literature

Languages: English, French, German, Dutch, Afrikaans, Portuguese, Spanish

Size: 75,000 books, 2,000 periodicals, 25,000 microfiches, 1,000 documentaries & feature films
Alphabetical card index to UDC codes
Searching with a UDC code in the ASC catalogue

“\(6\-15\:44\text{Sahel}\)” → Sahel
A new system for subject access

- user friendly
- word based
- multilingual, English and Dutch
- scope and coverage suited to library collection
- available electronically
No obvious ready-made system
Using the UDC to develop a word-based system

- Making an index
  
  *Example:* University Library of Kiel

- Extracting a thesaurus
  
  *Example:* ETH-Bibliothek Zürich
Digitizing the alphabetical card index

#001 06699
#100 markten (abstracte -)
#200 380.12

#001 06699
#100 markten, markt v.e. bepaald product
#200 380.12... (code product)

#001 06699
#100 markten, afzetmarkten
#200 380.123

#001 06699
#100 markten, marktanalyse
#200 380.13

#001 06699
#100 markten, marketing
#200 381.743
#201 (ook: 380.12)

#001 06699
#100 markten, co"peratieve marketing
#200 334.6
#201 (ook: 334.54)

#001 06699
#100 markten, marketing boards
#200 061.5:381.743
#201 (ook: 061.5:380.12)
Which UDC option?

Making an index to UDC codes
+ Potentially multilingual, partially automated construction process
- Considerable editing, incomplete overlap
  ASC UDC and MRF, indexing still classification based

✓ Extracting a thesaurus (decision mid 2000)
Importing digitized card file into MultiTes

<table>
<thead>
<tr>
<th>Ingressus file</th>
<th>MultiTes</th>
</tr>
</thead>
<tbody>
<tr>
<td>#001 00020</td>
<td>aardolie - voorkomen</td>
</tr>
<tr>
<td>#100 aardolie - voorkomen</td>
<td>SC: 553.98*</td>
</tr>
<tr>
<td>#200 553.98*</td>
<td>SN: *) is code voor het mineraal aardolie</td>
</tr>
<tr>
<td>#500 *) is code voor het mineraal aardolie</td>
<td>#001 00021</td>
</tr>
<tr>
<td>#100 aardrijkskunde</td>
<td>aardrijkskunde</td>
</tr>
<tr>
<td>#400 zie: geografie</td>
<td>RT: zie: geografie</td>
</tr>
<tr>
<td>#001 00023</td>
<td>Aha riots</td>
</tr>
<tr>
<td>#100 Aha riots</td>
<td>SC: (669) 323.274*1929</td>
</tr>
<tr>
<td>#200 (669) 323.274*1929</td>
<td>UF: women's war (opstand van Ibo-vrouwen in 1929)</td>
</tr>
<tr>
<td>#300 women's war (opstand van Ibo-vrouwen in 1929)</td>
<td>#001 kaartnummer</td>
</tr>
<tr>
<td>#100 UDC-omschrijving</td>
<td>Verwijderd</td>
</tr>
<tr>
<td>#200 UDC-code¹</td>
<td>Descriptor</td>
</tr>
<tr>
<td>#300 omschrijvingen/synoniemen²</td>
<td>SC:</td>
</tr>
<tr>
<td>#400 zie-verwijzingen (zie ook, zie verder)³</td>
<td>UF:</td>
</tr>
<tr>
<td>#500 opmerkingen⁴</td>
<td>RT:</td>
</tr>
<tr>
<td></td>
<td>SN:</td>
</tr>
</tbody>
</table>

- Tekst in #200 wordt verwijderd (zie op, hulgetal…)
- Blanks in #200 worden door Underscores vervangen

**Notes**
- DAT: Also covers publications on the petroleum industry indexed before 2006.
- IND: 1
- PPN: 294931708
- UCD: 553.98
- 662.753
- 665.6/.7
- 665.61
- 665.61.003.13

**Relationships**
- UF: crude oil
- UF: oil
- BT: fuels
- RT: hydrocarbon policy
- RT: natural gas
- RT: petroleum industry
- RT: petroleum products
- RT: petroleum refineries
The African Studies Thesaurus

• Completed June 2006

• Contains 12,319 terms
  ➢ 5,257 preferred terms
  ➢ 7,062 non-preferred terms

• 1,600 fewer descriptors than UDC ‘source’ codes (upward posting, compound codes split, codes from different classes combined)
Challenges

- Conceptual
- Thesaurus construction issues
Conceptual challenges

• Language (Dutch/English)
• Differences between a numerical code in a classification and words in a natural language
• Multitasking: constructing thesaurus, mapping codes and descriptors, reviewing classifying practice
Thesaurus construction issues
Specificity of vocabulary

• Splitting compound codes into component concepts

  *Example*: literature

  User tips: For publications about the literature of a particular ethnic group, combine 'literature' with the name of the ethnic group, e.g. Hausa literature = literature + Hausa.

• Upward posting

  *Example*: felines UF cats, leopards, lions
Thesaurus construction issues

Vocabulary control

Polysemy

• Qualifiers

  *Example*: Ndebele language (South Africa), Ndebele language (Zimbabwe)

• Alternative terms

  *Example*: drugs, medicinal drugs
Thesaurus construction issues
Vocabulary control

Names of languages and ethnic groups

*Example:* Fulfulde language

- UF Ful language
- UF Fula language
- UF Fulah language
- UF Fulani language
- UF Fulbe language
- UF Fulde language
- UF Peul language
- UF Poul language
Thesaurus construction issues
The organization of country related descriptors

Example: Ghana
RT Mossi polities
RT Nankanse language
RT Nanumba
RT Nawuri
RT Nawuri
RT New Patriotic Party
RT Sefwi polities
RT Sisala
RT Tem language
Thesaurus construction issues
The organization of country related descriptors

Ghana
RT Ghanaian languages
RT Ghanaian peoples
RT Ghanaian political parties
RT Ghanaian polities

Ghanaian peoples
BT African peoples
NT Adangme
NT Akan
NT Anufo
....
Broad Subject Ordering

• **Why?** subject browsing, mapping terms from multiple databases
• **How?** using UDC codes assigned to descriptors
• **Challenges?** right truncation of UDC codes, level of division, notation
Thesaurus and Classification

- **Thesaurus**: user friendly, multihierarchichal and associative relations
- **Classification**: subject categories, subject browsing, monitoring collection profile

Complementary information languages for subject indexing and retrieval
The African Studies Thesaurus

Comments?
Questions?
polman@ascleiden.nl
doorn@ascleiden.nl

http://www.ascleiden.nl/Library/Thesaurus/