Mapping science based on keywords of articles antecedences, presences, and consequences:
An application of CEON/CEES model of multi-perspective description of articles

Tanja Jevremov¹, Dejan Pajić¹, Miša Sotirović² & Pero Šipka²

¹University of Novi Sad, Serbia, Faculty of Philosophy
²Centre for Evaluation in Education and Science, Belgrade
Centre for Evaluation in Education and Science projects
Description of articles

- Main purposes:
  - Information retrieval
    - Classical
    - Visual
  - Author profiles

- New SEESAmE functionalities / modules:
  - AKwA: Automatic Keywords Assignment
  - KwASS: Keywords Assignment Support System - online verification and editing by the authors
Automatically generated descriptors: three types

citing direction

TI of cited references
AKw-R
intellectual base
antecedences

TI & AB
AKw-TA
research front
presences

TI & AB of citing papers
AKw-C
attracted area
consequences

time
Research questions

1. Aggregation of different types of keywords into a joint description of
   - papers
   - authors’ expertise

2. Explanation of potential differences among three types of keywords
   - knowledge transfer?
Method: MISH indexing

Multistage Indexing Subject Heading method: MISH

phases:
- term extraction
- term standardization - controlled vocabulary

method:
- Keyphrase Extraction Algorithm KEA

KEA problems:
- large number of
  1. unique terms
  2. general terms
Method: KEA / MISH comparison

prop. of different terms

prop. of 20 most frequent terms
Method: Data sample

- Sample of papers
  - 13032 SCIndeks papers, indexed with all three types of automatically generated descriptors
  - Majority of chemistry, agriculture and medicine

- Sample of words

<table>
<thead>
<tr>
<th></th>
<th>avr. per paper</th>
<th>lower threshold frequency</th>
<th>selected for analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKw-R</td>
<td>7.4</td>
<td>14</td>
<td>1279</td>
</tr>
<tr>
<td>AKw-TA</td>
<td>7.4</td>
<td>17</td>
<td>1006</td>
</tr>
<tr>
<td>AKw-C</td>
<td>11.8</td>
<td>27</td>
<td>1030</td>
</tr>
</tbody>
</table>
Method: Data sample

- Sample of papers
  - 13032 SCIndeks papers, indexed with all three types of automatically generated descriptors
  - Majority of chemistry, agriculture and medicine

- Sample of words

<table>
<thead>
<tr>
<th></th>
<th>avr. per paper</th>
<th>lower threshold frequency</th>
<th>selected for analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKw-R</td>
<td>7.4</td>
<td>14</td>
<td>1279</td>
</tr>
<tr>
<td>AKw-TA</td>
<td>7.4</td>
<td>17</td>
<td>1006</td>
</tr>
<tr>
<td>AKw-C</td>
<td>11.8</td>
<td>27</td>
<td>1030</td>
</tr>
</tbody>
</table>
Method: Data analysis

- Distance matrices: Jaccard index
- Cluster analysis: Partitioning Around Medoids method (PAM)
  - number of clusters: silhouette and connectivity measures
- Multidimensional scaling
  - distances among clusters = average distances of best represents of clusters
  - sizes of clusters = number of words in clusters
Results: AKw-TA map

- Agriculture: cattle breeding
- Agriculture: growing crops
- Microbiology: fermentation
- Soil characteristics
- Chemistry & technology
- Physics & technique
- Medicine: risk factors
- Psychology: education
- Society: law
Results: AKw-TA map

- Agriculture: cattle breeding
- Agriculture: growing crops
- Medicine: risk factors
- Microbiology: fermentation
- Soil characteristics
- Chemistry & technology
- Physics & technique
- Psychology: education
- Society: law
Results: AKw-TA map

- Agriculture: cattle breeding
- Agriculture: growing crops
- Soil characteristics
- Microbiology: fermentation
- Physics & technique
- Chemistry & technology
- Society: law
- Medicine: risk factors
- Psychology: education
Results: AKw-TA map

- Agriculture: cattle breeding
- Agriculture: growing crops
- Microbiology: fermentation
- Medicine: risk factors
- Soil characteristics
- Chemistry & technology
- Physics & technique
- Psychology: education
- Society: law
Results: AKw-TA map

- Agriculture: cattle breeding
- Agriculture: growing crops
- Soil characteristics
- Microbiology: fermentation
- Chemistry & technology
- Physics & technique
- Society: law
- Medicine: risk factors
- Psychology: education
Results: AKw-R map

- agriculture: growing crops
- biochemistry
- biology
- fermentation
- microbiology
- mathematical chemistry
- medicine: cardiology
- psychology
- society
- geology: earthquake estimation
Results: AKw-R map
Results: AKw-R map
Results: AKw-R map
Results: AKw-R map

- Agriculture: growing crops
- Biochemistry
- Biology
- Microbiology: fermentation
- Applied chemistry
- Geology: earthquake estimation
- Mathematical chemistry
- Medicine: cardiology
- Psychology
- Society
Results: AKw-C map
Results: AKw-C map

- Agriculture: growing crops
- Agriculture: food production
- Agriculture: biotechnology
- Soil characteristics
- Mathematics & physics
- Mathematical chemistry
- Society: law
- Psychology: education
- Medicine: cardiology
Comparison of maps in 2D space

Similarity of dimensions:
- AKw-R (from foreign publications)
- AKw-TA (from foreign publications)
- AKw-C (from national journals)

Similarity of clusters:

Comparison of maps in 2D space

- AKw-TA
- AKw-R
- AKw-C

From publication linked with citation relations
Conclusions

1. Maps are generally similar: acceptability of the idea of multi-perspective description

2. Differences among maps are mostly determined by:
   - Characteristics of analysis: descriptors from foreign or national papers
   - Characteristics of peripheral science in Serbia:
     - isolated fields
     - orientation on national science