Automated Construction and Analysis of Scientific and Educational Profiles of the University Department

Authors:
Mokhov Andrey
Tolcheev Vladimir
National Research University “MPEI”

Andrey Mokhov
NRU “MPEI”
Aim:
Determine similarity between the department’s educational courses and the subjects of the department’s scientific work.

Reason:
Possibility of replacing department courses with online-disciplines of leading Russian universities in case of teacher (lecturer) of the particular course doesn’t have publications on the topic of the lecture.

How to identify the courses that need to be replaced?
Terminological profile – a vector of words (descriptors) ordered by frequency of occurrence in the documents.

Samples:
Automated Construction and Analysis of Scientific and Educational Profiles of the University Department

Scientific profile of department

Profiles of educational courses

\[ X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \]

\[ Y_t = \begin{bmatrix} y_1^{(t)} \\ \vdots \\ y_n^{(t)} \end{bmatrix} \]

n = the size of the vocabulary (terms)

t = 1, ..., T where T – the number of educational courses

\( x_i \) – the frequency of the i-th term in the department’s profile,

\( y_i^{(t)} \) – the frequency of the i-th term in the profile of the t-th course

Cosine similarity - a degree of proximity of two profiles:

\[ \cos(X, Y_t) = \frac{\sum_{i=1}^{n} x_i \cdot y_i^{(t)}}{\sqrt{\sum_{i=1}^{n} (x_i)^2} \cdot \sqrt{\sum_{i=1}^{n} (y_i^{(t)})^2}} \]
Profiles of the *department of Control and Informatics (C&I)* of the NRU "MPEI":

\[ N = 387 \text{ articles for scientific profile. Vocabulary consists of } n = 651 \text{ words.} \]

*Profiles of 9 bachelor courses + 3 master courses:*

- Theory of automatic control (TAC).
- Methods of optimization (MO).
- Data processing methods (DPM).
- Neurocomputers and their applications (NC).
- Identification of dynamic objects (IDO).
- Simulation of control systems (SCS).
- Database management systems (DBMS).
- Microcontrollers and microprocessors in control systems (MMCS).
- Information technology (IT).
- Neuro-fuzzy diagnostic systems (NFDS, master course).
- Fuzzy control algorithms (FCA, master course).
- Decision support systems (DSS, master course).

Moscow, Russia
14-17 April, 2020
Automated Construction and Analysis of Scientific and Educational Profiles of the University Department

• Cosine similarity between C&I profile and bachelor courses

<table>
<thead>
<tr>
<th>Cosine similarity</th>
<th>Bachelor courses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TAC</td>
</tr>
<tr>
<td>C&amp;I</td>
<td>0.522</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cosine similarity</th>
<th>Undergraduate courses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SCS</td>
</tr>
<tr>
<td>C&amp;I</td>
<td>0.523</td>
</tr>
</tbody>
</table>

• Cosine similarity between C&I profile and master courses

<table>
<thead>
<tr>
<th>Cosine similarity</th>
<th>Master courses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NFDS</td>
</tr>
<tr>
<td>C&amp;I</td>
<td>0.465</td>
</tr>
</tbody>
</table>
PROXIMITY DETERMINATION BETWEEN SCIENTIFIC PROFILES OF DIFFERENT DEPARTMENTS

Scientific profile of Applied Mathematics (AM) department of the NRU "MPEI“ – consists of 336 articles.
Vocabulary consists of 884 terms.

First 20 terms of the AM and C&I profiles ordered descending by their weight in the profile of the AM:

<table>
<thead>
<tr>
<th>Term</th>
<th>AM</th>
<th>C&amp;I</th>
</tr>
</thead>
<tbody>
<tr>
<td>System</td>
<td>89</td>
<td>88</td>
</tr>
<tr>
<td>Decision</td>
<td>51</td>
<td>5</td>
</tr>
<tr>
<td>Artificial</td>
<td>41</td>
<td>9</td>
</tr>
<tr>
<td>Intelligence</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>intellectual</td>
<td>38</td>
<td>7</td>
</tr>
<tr>
<td>Support</td>
<td>35</td>
<td>6</td>
</tr>
<tr>
<td>computing</td>
<td>33</td>
<td>3</td>
</tr>
<tr>
<td>Informational</td>
<td>29</td>
<td>34</td>
</tr>
<tr>
<td>Logic</td>
<td>28</td>
<td>11</td>
</tr>
<tr>
<td>inference</td>
<td>26</td>
<td>15</td>
</tr>
<tr>
<td>modeling</td>
<td>22</td>
<td>13</td>
</tr>
<tr>
<td>programming</td>
<td>22</td>
<td>3</td>
</tr>
<tr>
<td>argument</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>Time</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>Network</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>Model</td>
<td>19</td>
<td>39</td>
</tr>
<tr>
<td>mathematical</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>Program</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>Neural</td>
<td>15</td>
<td>14</td>
</tr>
</tbody>
</table>

Cos: 0.523
Automated Construction and Analysis of Scientific and Educational Profiles of the University Department

Word clouds of department profiles:

**AM department**
- System
- Decision
- Artificial
- computing
- Model
- Argumnet
- Network
- Mathematical
- Inference
- Intelligence

**C&I department**
- System
- Model
- Control
- Function
- Network
- Characteristic
- Pathology
- Document
- Mathematical
- Dynamic
- Budget
Key topics of departments

- AM department: “Artificial Intelligence System”, “Decision Support System” and “Computing systems”

Can be found shared publications, prepared by the scientists of two departments: "Replenishment of base of signs of pathologies of a retina of an eye by processing of rhythmic electroretinograms for diagnostics by methods of artificial intelligence"
Limitations

The present analysis does not include:

• scientific works, published in foreign journals indexed in the Scopus and Web of Science databases,
• patents,
• certificates of registration of computer programs,
• acts of implementation in practice,
• research reports,
• incomplete data in eLibrary.ru – some articles and even scientists were not found in the digital library.
Related problem solved by presented technique - constructing profiles of each teacher of the department for scientific groups analysis.

By clustering the obtained profiles we found three groups of teachers in C&I conducting researches: "Fuzzy control systems", "Neural networks", "Information analysis".

**Word cloud of the cluster "Neural networks":**
**Conclusions:**

Presented technique allows:

- identify courses taught by teachers who have (or don’t have) scientific publications in course field
- Identify the best qualified teacher for certain course
- Identify the similarity of scientific fields of university departments
- Identify the scientific groups of certain department

Presented technique was approbated on the example of two departments of NRU “MPEI” and the results are well interpreted and close to experts’ opinions.

**Future plans:**

- Improve the quality of results by increasing the number of terms and using informative words from titles and annotations.
- Analyze Web of science and Scopus articles
Thank you for attention!

Speaker’s contacts:

Andrey Mokhov
asmokhov@mail.ru