Text and Data Mining for Systematic Reviews

Investigating Trends to Update Collaboration Services

Virginia Pannabecker
Virginia Tech, University Libraries
Background:
Researchers with limited time
Research Questions

In Systematic Reviews and Meta-Analyses:

● To what extent are text and data mining used?
● What theoretical approaches or protocols are described to guide text or data mining?
● What tools, techniques, and specific methods are used to accomplish text and data mining?
● What potential do these methods show for increased efficiency?
Methods

Web Searches:

Text and data mining for systematic reviews - first 5 pages
Text data mining research papers - first 2 pages
Followed related links
Total: 79

Advanced search: ‘any’
Skimmed list of 111 tools for ‘mine’
Total: 5 - Pimiento, PubReminer, SWIFT-Review, Systematic Review Assistant (prototype), and Weka
Methods

Literature Searches: keywords and index terms for text mining, data mining, systematic reviews, meta-analyses, evidence synthesis

Example PubMed Search: (mining[tiab] OR "data mining"[mesh] OR "information extraction") AND ("systematic review" OR "systematic reviews" OR meta-analysis OR meta-analyses OR meta analysis OR meta analyses OR metaanalysis OR metaanalyses OR "evidence synthesis" OR "knowledge synthesis" OR meta synthesis OR metasynthesis OR meta-synthesis OR sysrev_methods [sb])

<table>
<thead>
<tr>
<th>Database</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>CINAHL</td>
<td>48</td>
</tr>
<tr>
<td>Cochrane</td>
<td>45</td>
</tr>
<tr>
<td>Engineering Village</td>
<td>1761</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>200</td>
</tr>
<tr>
<td>PsycINFO</td>
<td>12</td>
</tr>
<tr>
<td>PubMed</td>
<td>426</td>
</tr>
<tr>
<td>Web of Science</td>
<td>854</td>
</tr>
<tr>
<td>Total w/Web</td>
<td>3425</td>
</tr>
</tbody>
</table>
Results

3425 Records
- 880 Duplicates

2545 Records

Categorized via keyword searching of records:

58+ of publications discussing use of TDM techniques for SRs or MAs

24+ SR or MA studies using TDM techniques
## Results - Journals Publishing More than 1 Methods Articles

<table>
<thead>
<tr>
<th>Journal</th>
<th>Number of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic reviews</td>
<td>7</td>
</tr>
<tr>
<td>Research synthesis methods</td>
<td>5</td>
</tr>
<tr>
<td>Journal of Biomedical Informatics</td>
<td>3</td>
</tr>
<tr>
<td>Journal of clinical epidemiology</td>
<td>3</td>
</tr>
<tr>
<td>Journal of the American Medical Informatics Association</td>
<td>3</td>
</tr>
<tr>
<td>Artificial intelligence in medicine</td>
<td>2</td>
</tr>
<tr>
<td>Information and Software Technology</td>
<td>2</td>
</tr>
</tbody>
</table>
Results - Methods for Using TDM in SRs and MAs

Number of Studies per Year

Year


Number of Studies

0 4 8 12 16
Results - SRs or MAs Using TDM

Disciplines

- Business - 2
- Computer Science - 1
- Education - 1
- Health and Medicine - 19
- Life Sciences - 1

TDM methods are used for: developing a search strategy, reviewing records for inclusion, identifying themes, extracting data
Key Readings


Text Mining Practices


**Document classification** - grouping and categorizing info based on trained models

**Document clustering** - grouping and categorizing info with data mining clustering

**Information retrieval** - retrieval and storage of docs, including searches

**Concept extraction** - grouping of words and phrases into similar groups

**Information extraction** - identification and extraction of facts and relationships from text

**Natural language processing** - (also: computational linguistics) tasks such as tagging parts of speech

**Web mining** - data and text mining on the Internet
Text Mining Uses for SRs

Paynter, et al reference: note that TDM is being used, or tested to support systematic review studies at each step:

Literature Search
Screening Records
Abstracting Data
Quality Appraisal
Review Updating
Results - Concerns regarding the use of TDM in SRs

- Maximizing recall to ensure comprehensive study
- Imbalanced datasets
  - Systematic Reviews often select only a very small number of studies (or sometimes no studies) out of hundreds or thousands reviewed
- Understanding how text and data mining tools work - by those using the tools for SRs


[This slide was updated on September 28, 2016, following the September 27, 2016 presentation.]
Results - TDM Tools, Techniques, Methods Used with SRs and MAs

Tools to Consider

- Table 5 (p22-23) and Appendix E (E1-E16) in Paynter, et al. 2016 - selected low-training, free, recommended tools include:
  - Abstrakr
  - PubReminer

More Tools

- Systematic Review Toolbox
- Additional Tools: ContentMine
● Licensing agreements for databases, lack of an API, or unavailability of papers may limit some applications of TDM for SRs or otherwise.

See the Association for Research Libraries’ Issue Brief on Text and Data Mining and Fair Use in the United States.

● A good reason to encourage faculty to continue depositing manuscripts and datasets in institutional or other repositories.
Next Steps

Explore social sciences and policy text and data mining use and tools

Try one or more of these tools with researchers to investigate their ease of use and benefit for various contexts:

- Quick Review to inform longer review protocol and process; to decide the value in conducting a longer, more thorough review
- General Literature Review
- Scoping Review
- SR / MA
Text and Data Mining for Systematic Reviews: Investigating Trends to Update Collaboration Services

When discussing project planning for systematic reviews and meta-analyses with faculty and graduate students, librarians sometimes hear wistful inquiries about automated approaches. Systematic reviews (SRs) require management, analysis, and synthesis of large amounts of data, perhaps particularly those including numerous studies with qualitative text-based data. To investigate how text and data mining approaches might be used in SRs to increase project efficiency, the author conducted, and reports on the results of a literature review.
Key Readings


Recommended Reading for Overview
[1-5]

Studies on Text and Data Mining Methods for Systematic Reviews
[1-57]

Systematic Review / Meta-Analysis Studies Using Text or Data Mining Methods
[58-81]

Full Reference list in VTechWorks Institutional archive record:
http://hdl.handle.net/10919/73042