

# Scientometrics: How to perform a Big Data Trend Analysis with ScienceMiner

Volker Frehe, Vilius Rugaitis, Frank Teuteberg

Osnabrück University  
Accounting and Information Systems  
Katharinenstr. 1  
49074 Osnabrück  
volker.frehe@uni-osnabrueck.de  
vrugaiti@uni-osnabrueck.de  
frank.teuteberg@uni-osnabrueck.de

**Abstract:** This paper describes the results of the implementation of an application that was designed under the design science principles. The purpose of this application is to identify trends in science. First, the status quo of similar applications as well as the knowledge base about data mining in the field of scientometrics is analyzed. Afterwards, the implementation as well as the evaluation of our application is described. Our web-based application allows to search for contributions (literature and internet, e.g., twitter, news), executes several data mining methods and visualizes the results in seven different ways. Each visualization has some filters and further control elements. It is the first application to provide the complete process from data acquisition to data visualization in an automated way.

## 1 Motivation

Independent of the research field, the literature review is an important and essential yet time-consuming method to gather the status quo in science. There are several indices, like the h-index [Hi05], by means of which authors can be rated and distinguished authors and literature can be identified. It is a broadly accepted method to separate relevant from irrelevant literature by means of the various variants of the h-index, e.g., the one for institutions [Ki07] or else completely new variants like the g-index [Eg06] that is also based on the h-index. But scientific knowledge is not only distributed in literature, it can also be found in the internet, e.g., in social networks like Twitter, Facebook, etc. As the knowledge base continues to increase, new methods need to be developed to capture it. There are already automated methods from the field of information retrieval (IR), that are used in scientific knowledge capturing, like co-classification [AG10] and co-word analysis [DCF01, Le08]. Moreover, it has been proven that automated citation analysis is able to reduce the workload of the scientists [Co06, Ma10].

Therefore, automation and the use of IR and data mining methods seem to be indispensable in the field of scientific research, especially for knowledge discovery.

This contribution describes the process of the development of an application for big data trend analysis in the area of science. As the process follows the design science principles [He04], for the implementation, we use a great knowledge base deriving from a literature review as well as information from the environment (e.g., from existing applications, round tables and surveys). There are already several similar applications, but our application is the first to provide the complete process from data acquisition to data visualization in an automated way.

The paper is structured as follows. At first, the literature search and similar applications are described. Afterwards, in section 3, the research methodology is explained. Section 4 provides an extensive literature review, and section 5 contains information about the implementation of the application. Before closing the paper with a conclusion, section 6 shows the results of the evaluation among scientists.

## 2 Status Quo and Related Work

A systematic literature review based on [WW02] was performed to gain information about the topic's trend analysis and scientometrics. We used several databases with specific search terms to get a broad body of literature. Following the guideline of [Br09] for transparency of the search process, the complete list of databases, search words and results can be downloaded<sup>1</sup>. Overall, our search resulted in 2,674 contributions. As the journal *Scientometrics* is solely about the "science of science" and thus is very important for our research, we also investigated 1,190 additional contributions from this journal. In order to consolidate the huge amount of 3,864 contributions, we first eliminated all non-academic contributions. Further, we sorted out the irrelevant papers by reading the titles, which resulted in a list of 594 contributions. In addition, we accepted papers exclusively in English and removed contributions without relevance to the topic by reading the abstracts. Ultimately, we used 289 contributions for the analysis.

Our literature search revealed several applications with some kind of scientometric analysis. The list of identified applications is shown in table 1, which also provides information about the type of the data that the application can analyze. We divided it into bibliometric data, altmetric data or other scientometric data. Most applications access bibliometric data, however, only few use altmetrics. The most common visualization methods are tables and diagrams, whereas methods like tag clouds, world maps or heat maps are under-represented. A detailed analysis of the applications can be downloaded<sup>2</sup>.

Our application differs from these applications as, to the best of our knowledge, it is the first to comprise all functions, from data acquisition over data preparation and analysis to data visualization in form of a user-friendly web application.

---

<sup>1</sup> Literature Search Details & Results, <http://uwi.uos.de/att/SM-LiteratureSearch.pdf> [last access 27.06.2014]

<sup>2</sup> Investigated Tools, <http://uwi.uos.de/att/SM-Tools.pdf> [last access 27.06.2014]

Tool	Type			Visualization								
	Bibliometrics	Scientometrics	Altmetrics	Table	Diagram	Landscap Map	Heat Map	Tag Cloud	World Map	Stream Graph	Strategy/ strategic map	Network Map
Bibexcel	x											
CiteSpace	x				x				x			x
CoPalRed				x	x						x	x
IN-SPIRE				x	x	x				x		x
Leydesdorff's Software	x											
Network Workbench Tool		x			x							x
Science of Science (Sci <sup>2</sup> ) Tool	x			x	x				x			x
VantagePoint	x			x	x							x
VOSViewer						x						x
Sitkis	x											
BiblioTools	x											
SAINT	x		x									
SciMAT	x			x							x	x
CATAR	x											
TEXTREND	x				x							x
ImpactStory			x	x								
Altmetric			x	x	x		x		x			
SciCombinator			x	x				x				
PlumX			x	x	x							

Table 1: Investigated Applications and Visualization Methods

### 3 Research Method

Our research is following the design science [He04] method, as we want to include scientific knowledge as well as information from the environment in the development, as shown in figure 1.

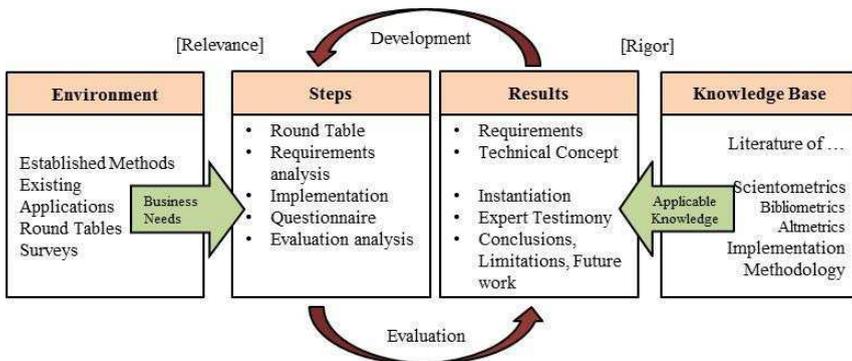


Figure 1: Design science Method by [He04]

The seven design science principles [He04] are met, as shown in table 2.

Guideline	Description
Design as an Artifact	The application, which is our developed artifact, follows the definition of [MS95].
Problem Relevance	The relevance of the topic is already mentioned in section 1 and further worked out in section 4.
Design Evaluation	At the moment, there is already a first evaluation (cf. section 6); further evaluations will follow.
Research Contributions	Our application is the first of this kind, providing all functions from data acquisition over data preparation and analysis to data visualization in form of a user-friendly web application.
Research Rigor	In our research, we follow the methodological requirements for literature research [WW02], design science [He04, MS95] and the evaluation by survey [My09]
Design as a Search Process	The iterative search process is shown in figure 1.
Communication of Research	The web application itself is free for everyone <sup>3</sup> . The results will be presented to the scientific community inter alia via this contribution.

Table 2: Design science Guidelines [He04]

## 4 Literature Review

Our literature search resulted in 289 contributions. We also conducted a cluster analysis via Rapidminer<sup>4</sup> and used the title, abstract and keywords of each contribution for clustering. Common stop words, as well as an extra list of 71 stop words, were eliminated from the list of words. We equally eliminated words with less than 4 and more than 25 characters. After the selection of the valid words, a word stemming has been performed. The cluster analysis was performed by a k-means algorithm and resulted in 9 clusters as displayed in table 3.

Most papers belong to clusters 6, 7 and 3, which represent some kind of bibliometric citation analysis (cluster 6), trend detection (cluster 7) and co-word/co-citation analysis. In the following, we want to get a deeper insight in each cluster.

**Cluster 0** is about indices. Radicchi and Castellano (2013) investigate the relationship of the h-index and the amount of publications and citations of a scientist [FC13]. They detect a weak connection between amount of publications and h-index and a strong correlation between the h-index and the amount of citations. A study about the influence of age, field and uncitedness on author ranking is performed by [Am12]. It shows that the age and the field of the scientist have great influence on the ranking. **Cluster 1** is about social network analysis (SNA), like the contributions of [No12], who develops a method to detect trends in social networks (SN). Therefore, several self-organizing maps

<sup>3</sup> ScienceMiner, <http://scienceminer.uwi.uos.de> [last access 27.06.2014]

<sup>4</sup> Predictive Analytics, Data Mining, Self-service, Open source – RapidMiner, <http://rapidminer.com> [last access 27.06.2014]

(SOM) are created at various times who visualize the results of a cluster analysis. These SOMs can be used to identify trends in social networks.

Cluster	Description	# of Papers
0	h; h_index; index; indic; type; citat; number; public; individu; evalu; measur; scientist	23
1	network; social; social_network; mine; data; futur; analysi + analyz; domain; user; relationship; knowledg; techniqu; pattern	17
2	technolog; literatur; trend; bibliometr; emerg; analysi + analyz; model; bibliometr_analysi; appli + applic; system; network; develop; research; case; studi; citat	29
3	co; cluster; map; network; analysi + analyz; knowledg; structur; document; citat; tool; china; scienc; research; keyword; field	43
4	collabor; intern; scientif; author; institut; countri; level; network; china; nation; paper; co; field; scienc; analysi + analyz; public; bibliometr; output; impact; research; pattern; year	31
5	factor; journal; impact; citat; publish; cite; articl; paper; individu; author; web; discuss; effect	12
6	citat; public; countri; research; cite; scienc; bibliometr; output; evalu; indic; articl; product; analysi + analyz; number; scientif; scientist; area; disciplin; knowledg; journal; paper; assess; publish	68
7	detect; topic; user; trend; emerg; model; collect; data; social; inform; search; propos; approach; cluster; method; interest; dynam; system; appli + applic; retriev; analysi + analyz	44
8	author; rank; univers; citat; co; approach; research; measur; base; bibliometr; assess; compar; index; topic; propos	22

Table 3: Cluster description

**Cluster 2** is also about trend analysis in networks, but from a more bibliometric point of view. Guille (2013) indicate that the mentioning frequencies (e.g., re-tweets) are a better indicator for the popularity of a topic than the global frequency of a topic [Gu13]. These indicators can be used to measure information diffusion in SN. Khan et al. (2011) created a concept (network of core, based on the mathematical graph theory), to discover hidden structures in scientific networks by the visualization of theoretical constructs, models and concepts of a specific scientific domain through a network [KMP11]. **Cluster 3** is mainly about co-word analysis und co-citation analysis. An analysis of co-citation performance of six retrieval methods has been conducted by [Et12]. A positive effect on performance could be found by using the co-citation context and the normalization technique of cited frequency. Yang et al. (2012) have combined several visualization techniques (cluster tree, strategy diagram and social network maps) of the co-word analysis to use the advantages of each technique and to weaken the disadvantages [YWL12]. A problem in the field of co-word analysis is the use of keywords as a weak point of literature search [NPS13, Wa12]. Solutions are to use the Knowledge Discovery Process (KDP) for cluster analysis of all available contribution data [NPS13] or to integrate expert knowledge into the co-word analysis in form of a new method, the semantic based co-word analysis [Wa12]. **Cluster 4** deals with the collaboration of scientists. Gazni et al. (2012) have investigated that collaborations between authors, institutions and countries have gradually increased in the past years

[GSD12], which indicates the importance of this topic. He et al. (2011) explore co-author networks via a context subgraph [HDN11]. Through this subgraph quantitative factors should be obtained by the integration of the author's background in the analysis. **Cluster 5** is concerned with journal impact factors. Vanclay (2012) critically study the Thomson Reuters Impact Factor (TRIF) and demonstrates the influence of wrong links, misspelling, missing cites and advocate a complete overhaul of the TRIF [Va12]. Thelwall (2012) additionally request for adding new indicators (altmetrics), like online readership indicators, social bookmarking indicators, link analysis, web citations and Twitter (tweets) in order to enhance the bibliometric indicators [Th12]. To avoid manipulation, a mixture of several indicators should be used. **Cluster 6** is about citation analysis. Franceschet (2009) conducted a correlation analysis to reduce quantitative, bibliometric indicators for scientist assessment [Fr09]. The analysis includes 13 indicators. The amount of papers (for productivity assessment), the amount of citations (for impact assessment), the average citation amount per paper (for relative impact assessment) and the m-quotient (for long-term impact assessment) are identified as the most important indicators. **Cluster 7** deals with trend analysis. Tseng et al. (2009) investigate several trend indices [Ts09]. It was figured out that the linear regression is best for timeline analysis, which supports the extensive usage of this method. Guo et al. (2011) use several indicators (increase of specific word usage, amount of new authors in research field and amount of interdisciplinary citations) in a mixed model [GWB11]. Their research indicates that new authors explore a new research field first, then, they reference interdisciplinary literature before they use some specific words more often. Through this information, new trends and hot topics can be identified. **Cluster 8** covers author rankings. Wang et al. (2012) identify that the co-citation analysis can also be used to recognize research patterns, find research communities and is in a position to identify hot topics in science [WQY12]. Ding (2011) criticizes that author rankings are field independent [Di11]. He proposes a new ranking which includes the authors' fields (topic-based PageRank for authors). The author-conference-topic model (ACT) is used to gain information about the authors' fields and it is integrated with the PageRank algorithm to enable a field dependent author ranking. The results of the literature review have been used in the conceptual phase of the implementation of our web application.

## 5 Implementation of the Prototype

The framework of the application and the interaction of the several modules are displayed in figure 2. As our framework is built on a modular basis, enhancements are possible in every step (e.g., adding new data sources or mining/visualization methods).

The developed artifact is a web application for automated trend detection via bibliometric and altmetric analysis. We follow the Knowledge Discovery in Databases (KDD) process of [FPS96], which consists of the steps selection, preprocessing, transformation, data mining and interpretation. Therefore, the web application is designed as user-friendly as possible. At first, the user states a keyword for a topic to search for. The application will execute the next steps in the background so that the user gets a result of the process in form of some visualizations.

The first step is the selection of data. Therefore, we integrated several data sources from the internet which are accessed through application programming interfaces (API). Because of usage and technical restrictions or bad quality of data, our prototype has access to Microsoft Academic Search<sup>5</sup> as source for bibliometric data and the service altmetric<sup>6</sup> for altmetric data. The service altmetric combines access to several sources like Facebook, Google+, Twitter, Reddit and several blogs and news sites. The data selection is performed via a batch process on the server side. This allows the user to state a query and leave the web application while the search query is executed in the background. This approach provides flexibility for the end-user, since most of the data sources suffer of technical and legal restrictions, which lead to a long execution time. This way the time-consuming queries can be initiated and then executed in the background without the need of permanent user presence. When logging in again, the user has access to all his executed queries. The batch process also enables multi-threading and parallel processing of various queries, which enhances the performance. The batch process also performs the second step of the KDD process (preprocessing). Irrelevant words (stop words) are eliminated, a word stemming is executed and synonyms are combined through the use of a dictionary. The mainly utilized entities like users, administrators, contributions and dictionaries in form of a Unified Modeling Language class diagram can be downloaded<sup>7</sup>.

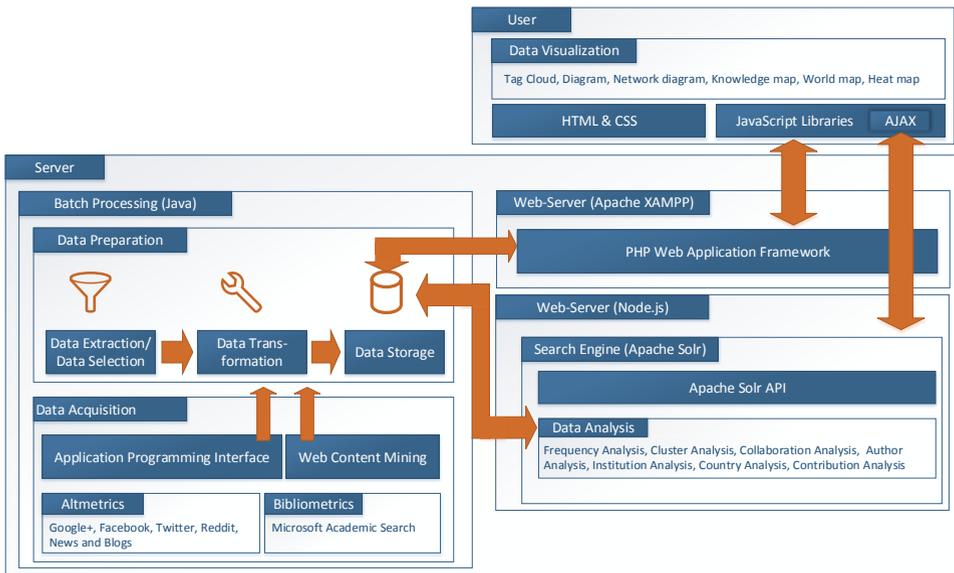


Figure 2: Architecture of application

After the data selection, the next step is the transformation which is also performed on the server. The contributions' data enriched with altmetric data are converted to the

<sup>5</sup> Microsoft Academic Search, <http://academic.research.microsoft.com> [last access 27.06.2014]

<sup>6</sup> Altmetric API documentation, <http://api.altmetric.com> [last access 27.06.2014]

<sup>7</sup> ScienceMiner UML diagram, <http://uwi.uos.de/att/SM-UML.pdf> [last access 27.06.2014]

needed format, if necessary merged and stored in a relational database. This is the last step of the batch process.

The next KDD process step (data mining) is done by Apache Solr<sup>10</sup> on the server. This product is suitable due to the provision of advanced text analysis methods, fast response times, import and export functionalities and enhancement possibilities. Because of performance reasons, the data is imported to Solr and is not analyzed in the relational database. This provides the flexibility which is needed for the interactive visualization of the results. However, Solr does not provide any security mechanisms for the data exchange. This is why we decided to use Node.js<sup>11</sup> as proxy server for Solr to handle the access. At the moment, only clustering and frequency analysis is used for data mining. The last step of the KDD process (interpretation) involves the user again. The web application provides HTML and JavaScript functionalities that communicate with the web server via Asynchronous JavaScript and XML (AJAX). Several visualization possibilities are given, which aid the user interpreting the results.

The most important part for the user is the visualization of the results. There are several methods provided to display the mining results. An example of the user interface with a result of the query “Scientometrics” is shown in figure 3. At first, there is general information providing an overview of the data gathered by the query (e.g., how many publications and altmetric data have been found, the date of the first and last publication, etc.).

As our literature review reveals numerous visualization techniques, our application implements several of them. The tag cloud provides an overview about the most-relevant terms, keywords, authors, etc. The diagram allows to show a timeline of the publication dates and to also visually view authors, affiliations, etc. as well as the amount of their publications. The network map is a construct to visualize the connections between entities like authors, countries and affiliations. The topic map enables to cluster the contributions and show main topics and the associated keywords. The world map is a construct by which the origin (and amount) of the contributions is displayed on a world map. The heat map (cf. figure 3) shows the diffusion of several topics over time. Each visualization element has some controls. There are controls to specify the timeframe, choose the element to be analyzed (e.g. author vs. affiliation), specify the amount of elements to be shown, etc. Depending on the visualization element, the respective controls are depicted. A complete overview of all visualization elements can be downloaded<sup>12</sup>.

Every method can be displayed or hidden and also the order of the methods can be changed. The left navigation panel can also be hidden in order to use the available space for the visualization elements.

---

<sup>10</sup> Apache Lucene - Apache Solr, <https://lucene.apache.org/solr> [last access 27.06.2014]

<sup>11</sup> node.js, <http://nodejs.org> [last access 27.06.2014]

<sup>12</sup> ScienceMiner Visualization, <http://uwi.uos.de/att/SM-Visualization.pdf> [last access 27.06.2014]

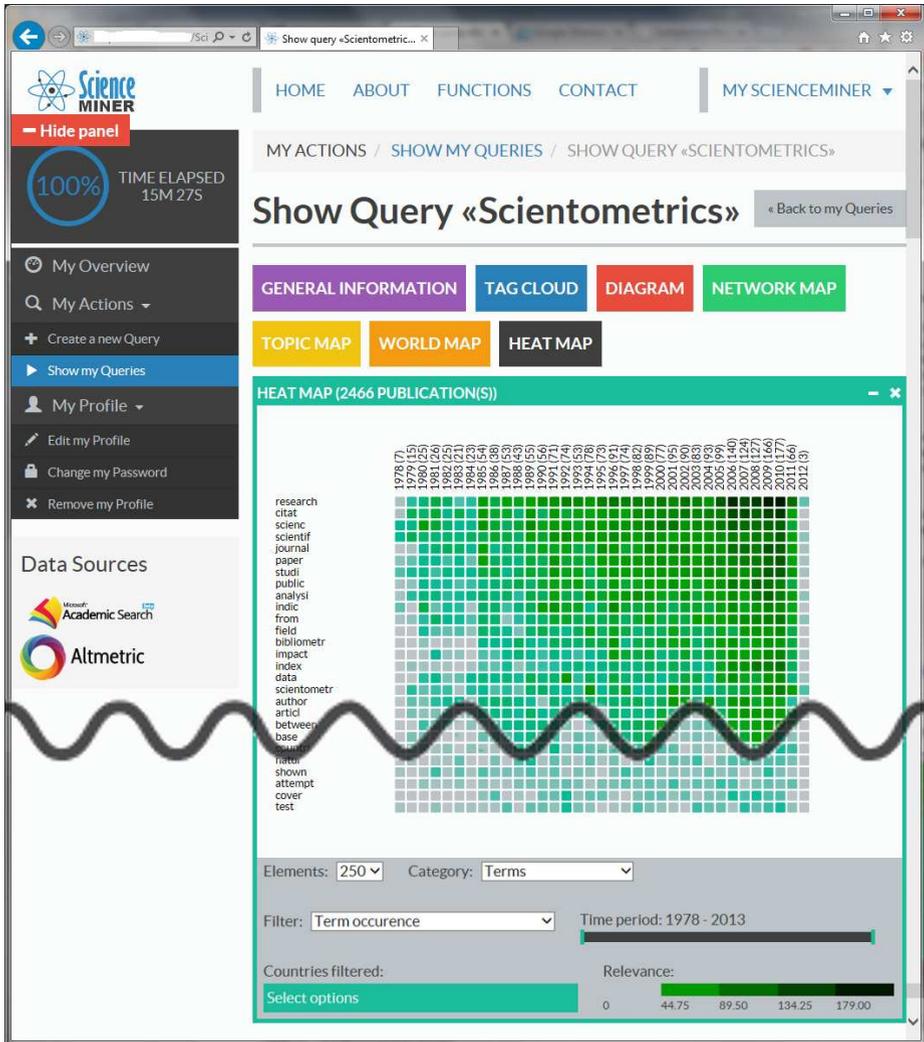


Figure 3: Frontend with heat map of web application<sup>13</sup>

## 6 Evaluation

After the experimentation phase, we invited 40 scientists and young researchers via e-mail to take part in an evaluation of the web application. We asked them to use the application and fill in an online survey. Apart from the integrated online help, no further support was given. Up to now, 14 of the invited scientists and researchers have completed the survey. The average age of the participants is 27.8 and all are male. Four of them are students, two graduates, seven research assistants and one professor. Of the

<sup>13</sup> The curved line indicates that we merged two screenshots into one.

respondents, 71.5% come from the IS field, 21.5% from the field of economics and 7% from other fields. The survey consists of 7-point Likert scale questions as well as free text fields for notes and recommendations. The Likert Scale reaches from “strongly agree” to “strongly disagree”. The questions are grouped into clusters to evaluate the design, the content, the usability and the functionality of the application as well as to raise general questions about bibliometrics and altmetrics. Figure 4 shows one sample question for each category and the associated results. The survey shows good results for the design, the content, the usability and the functionality of the application. The bibliometrics and altmetrics seem to be accepted methods by scientists, but only in addition to other methods (see next paragraph). The complete survey consists of 46 questions; the results are comparable with the ones mentioned here. As the survey has not been concluded yet, the presented results only serve the purpose of giving first insights.

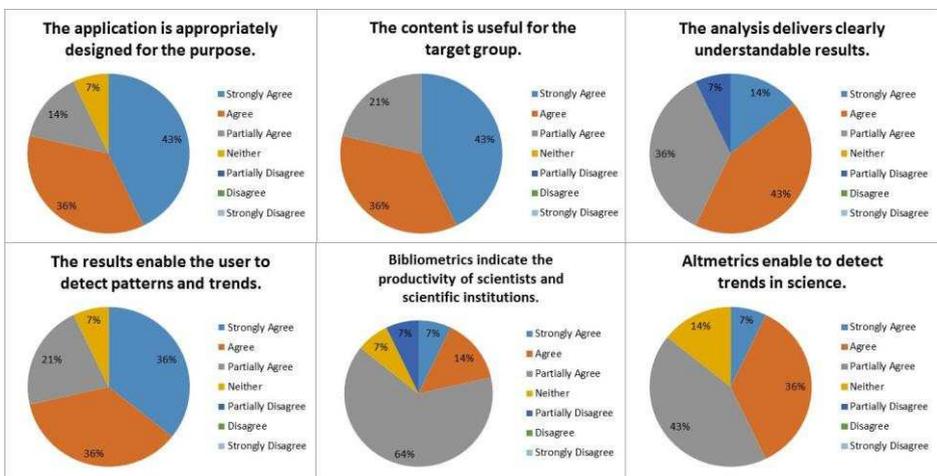


Figure 4: Survey results

However, already the annotations received so far provide some valuable elements of improvement for the application. Most people still perceive the qualitative review to be indispensable. According to them, the bibliometric and altmetric analyses can only be used in a subsidiary manner or just to identify relevant literature. Although our application is deemed useful, there are also some improvement suggestions, for instance, to integrate a spellchecker in the research as well as the inclusion of acronyms in the search. As the search is time-consuming, apparently there is a need for some kind of fast pre-search. Furthermore, some comments referred to the wish, that not only the abstracts should be investigated, but the entire contributions. Additionally, more search engines (like Google Scholar) should be integrated to obtain more results. One person asked for a list of all identified publications. However, this feature can due to legal restrictions not be integrated as it would be an imitation of the search engine’s functionality. Two people asked for a comparison of two search results. Also, more visualization methods were wanted as well as the possibility to export the results. If procurable, all these recommendations will be implemented to further improve the application in the aim of design science (cf. section 3).

## 7 Conclusion and Future Work

Following the design science principles, the developed application proves how theoretical knowledge from scientometrics and data mining theories can be used in a practical way. The application can be used by scientists to get new insights into several fields of their research. The evaluation indicates that the application is practicable and useful. However, the automated data mining should only be used in addition to traditional literature research methods. Nevertheless, the developed application can be seen as an enhancement to the traditional methods and although it prods to new trends and discovers undetected contributions by the use of not only scientific contributions, but also information from the web (like Facebook, Twitter, etc). We are well aware of the fact, that our application has only been evaluated by 14 people so far, which represents a limitation. However, with this contribution we pursue the goal of stimulating a broad use of our prototype. Thereby, more scientists might work with it and we might obtain further meaningful recommendations from the science community.

## Acknowledgments

The authors would like to thank the anonymous reviewers and Ms. Marita Imhorst, who provided valuable insights, help and substantive feedback during the research process.

## References

- [AG10] M.-R. Amini and C. Goutte: A co-classification approach to learning from multilingual corpora. *Machine Learning*, 79(1–2):105–121, 2010.
- [Am12] L. Améz: Citation Measures at the Micro Level: Influence of Publication Age , Field , and Uncitedness. *Journal of the American Society for Information Science and Technology*, 63(7):1459–1465, 2012.
- [Br09] J. vom Brocke, A. Simons, B. Niehaves, K. Riemer, R. Plattlauf and A. Cleven: Reconstructing the Giant: On the Importance of Rigour in Documenting. *ECIS 2009 Proceedings*, 2009.
- [Co06] A. M. Cohen, W. R. Hersh, K. Peterson and P.-Y. Yen: Reducing workload in systematic review preparation using automated citation classification. *Journal of the American Medical Informatics Association*, 13(2):206–19, 2006.
- [DCF01] Y. Ding, G. G. Chowdhury and S. Foo: Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management*, 37(6):817–842, 2001.
- [Di11] Y. Ding: Topic-Based PageRank on Author Cocitation Networks. *Journal of the American Society for Information Science and Technology*, 62(3):449–466, 2011.
- [Eg06] L. Egghe: Theory and practise of the g-index. *Scientometrics*, 69(1):131–152, 2006.
- [Et12] M. Eto: Evaluations of context-based co-citation searching. *Scientometrics*, 94(2):651–673, 2012.
- [FC13] F. Radicchi and C. Castellano: Analysis of bibliometric indicators for individual scholars in a large data set. *Scientometrics*, 97(3):627–637, 2013.
- [FPS96] U. M. Fayyad, G. Piatetsky-Shapiro and P. Smyth: From data mining to knowledge discovery: An overview. (U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R.

- Uthurusamy Eds.): *Advances in knowledge discovery and data mining*, pp. 1–34, AAAI Press, Menlo Park, 1996.
- [Fr09] M. Franceschet: A Cluster Analysis of Scholar and Journal Bibliometric Indicators. *Journal of the American Society for Information Science and Technology*, 60(10):1950–1964, 2009.
- [GSD12] A. Gazni, C. R. Sugimoto and F. Didegah: Mapping World Scientific Collaboration: Authors, Institutions, and Countries. *Journal of the American Society for Information Science and Technology*, 63(2):323–335, 2012.
- [Gu13] A. Guille: Information Diffusion in Online Social Networks. *SIGMOD Records*, 42(2):17-28, 2013
- [GWB11] H. Guo, S. Weingart and K. Börner: Mixed-indicators model for identifying emerging research areas. *Scientometrics*, 89(1):421–435, 2011.
- [HDN11] B. He, Y. Ding and C. Ni: Mining Enriched Contextual Information of Scientific Collaboration?: A Meso Perspective. *Journal of the American Society for Information Science and Technology*, 62(5):831–845, 2011.
- [He04] A. R. Hevner, S. T. March, J. Park and S. Ram: Design science in information systems research. *MIS Quarterly*, 28(1):75–105, 2004.
- [Hi05] J. E. Hirsch: An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46):16569–16572, 2005.
- [Ki07] A. L. Kinney: National scientific facilities and their science impact on nonbiomedical research. *Proceedings of the National Academy of Sciences of the United States of America*, 104(46):17943–17947, 2007.
- [KMP11] G. F. Khan, J. Moon and H. W. Park: Network of the core: mapping and visualizing the core of scientific domains. *Scientometrics*, 89(3):759–779, 2011.
- [Le08] W. H. Lee: How to identify emerging research fields using scientometrics: An example in the field of Information Security. *Scientometrics*, 76(3): 503–525, 2008.
- [Ma10] S. Matwin, A. Kouznetsov, D. Inkpen, O. Frunza and P. O’Blenis: A new algorithm for reducing the workload of experts in performing systematic reviews. *Journal of the American Medical Informatics*, 17:446–453, 2010.
- [MS95] S. T. March and G. F. Smith: Design and natural science research on information technology. *Decision Support Systems*, 15(4): 251–266, 1995.
- [My09] M. D. Myers: *Qualitative Research in Business & Management*. London, 2009.
- [No12] P. N. E. Nohuddin, F. Coenen, R. Christley, C. Setzkorn, Y. Patel and S. Williams: Finding “interesting” trends in social networks using frequent pattern mining and self organizing maps. *Knowledge-Based Systems*, 29:104–113, 2012.
- [NPS13] P. Nieminen, I. Pölonen and T. Sipola: Research literature clustering using diffusion maps. *Journal of Informetrics*, 7(4):874–886, 2013.
- [Th12] M. Thelwall: Journal impact evaluation: a webometric perspective. *Scientometrics*, 92(2): 429–441, 2012.
- [Ts09] Y.-H. Tseng, Y.-I. Lin, Y.-Y. Lee, W.-C. Hung and C.-H. Lee: A comparison of methods for detecting hot topics. , 81(1):73–90, 2009.
- [Va12] J. K. Vanclay: Impact factor: outdated artefact or stepping-stone to journal certification? *Scientometrics*, 92(2):211–238, 2012.
- [Wa12] Z.-Y. Wang, G. Li, C.-Y. Li and A. Li: Research on the semantic-based co-word analysis. *Scientometrics*, 90(3): 855–875, 2012.
- [WQY12] F. Wang, J. Qiu and H. Yu: Research on the cross-citation relationship of core authors in scientometrics. *Scientometrics*, 91(3):1011–1033, 2012.
- [WW02] J. Webster and R. T. Watson: Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2):xiii–xxiii, 2002.
- [YWL12] Y. Yang, M. Wu and L. Cui: Integration of three visualization methods based on co-word analysis. *Scientometrics*, 90(2):659–673, 2012.