

# Topics in Software Industry Transformation Research: A Topic Analysis of Major IS Conferences

Anton Pussep, Markus Schief, Benedikt Schmidt,  
Florian Friedrichs and Peter Buxmann

Technische Universität Darmstadt, Chair of Information Systems,  
Hochschulstraße 1, 62849 Darmstadt, Germany  
{pussep, schief, buxmann}@is.tu-darmstadt.de, florian.friedrichs@stud.tu-darmstadt.de

SAP Research, Bleichstraße 8, 64283 Darmstadt, Germany  
benedikt.schmidt@sap.com

**Abstract.** As the information load grows, it becomes increasingly difficult to follow-up new trends in business and management. However, new developments in technologies and markets pose threats and open up opportunities to firms. Especially the software business changes continuously and profoundly. It is therefore necessary for researchers and practitioners to follow up recent developments and to cope with the information overload. We suggest the application of a data mining technique in order to automatically identify topics: Latent Dirichlet Allocation (LDA). Using a sample of 13,799 publications from ICSOB and major conferences on Information Systems, we identify topics relevant to industry transformation research and review their development on a timescale. As proof of concept, we conduct a short case study using Green IT in order to demonstrate that topic analysis can yield relevant results for literature search beyond the results that can be obtained through a simple keyword search.

**Keywords:** topic analysis, latent dirichlet allocation, LDA, industry transformation, software industry.

## 1 Introduction

IT and software are fast changing businesses. The last decade has seen the rise of Google and Facebook from small startups to the most valuable firms on the planet. While Google made its business with internet search engines, Facebook became the biggest social network. Both businesses engaged in early markets and current technologies, profiting by the first-mover advantage. In presence of network effects, it can be argued that the first-mover advantage could be even more beneficial in the software business [1], [2].

As information grows, it becomes increasingly hard to follow-up important trends and to hence identify sweet spots yielding great business potential. For instance, in our sample the number of sources doubled in the period 2000-2011 and even fourteen-fold in 1993-2011. In research, meta-analysis studies are becoming

increasingly popular in order to cope with the information load: A search in titles in the Business Source Premier database indicates a continuous growth in the period 1990-2010 from 23 to 96 sources per year. State of the art works solely attempt to sum up the current body of knowledge on a particular topic. However, meta-analysis and state of the art analysis is costly, as all work needs to be reviewed manually. A solution can be found in assisting methods from Data Mining, such as the Latent Dirichlet Allocation (LDA). LDA is an unsupervised learning algorithm which can automatically extract topics from a large dataset. In this paper, we suggest the application of LDA to: (1) overview the topics in IS in order to identify relevant topics for software business and management; (2) assess the development of topic relevance over time; and (3) support literature search and review through the unsupervised identification of relevant sources.

Consequently, the research question that we address in this paper is: Which recent topics in IS and software business research are relevant to the field of software industry transformation? Our main contribution to software business and management research is the identification of relevant topics and their development over time. We further demonstrate how the method can support literature search on large datasets beyond a keyword search.

Researchers will find the topics presented here useful to overview their respected fields and assess the current relevance of the topics. Whereas the identified topics are inherent to IS research, they either deal with software directly or can be used as a starting point for software-specific research. Furthermore, the method used in this paper proved to be useful for the identification of relevant sources. We therefore suggest it as an assisting tool for literature search and review. Practitioners will find the results and methods useful for tracking developments in their respective industries, thereby supporting the identification of threats and opportunities to their businesses.

The remainder of the paper proceeds as follows: Section 2 provides an overview on applied topic analysis demonstrating its potential. We proceed with the presentation of the data and method used in this paper in section 3. Next, an overview of relevant topics is presented in section 4, followed by an analysis of how the topic relevance evolved over time in section 5. Finally, section 6 concludes the paper and provides avenues for further research.

## **2 Literature Review**

Topic analysis is a useful tool for the identification of structures in large textual datasets. A set of approaches exists that is based on Bayes methods and enables topic analysis. In the following, we present different applications of topic analysis to datasets and motivate why an approach to software business and management research appears to be promising.

Steyver and Griffiths analyze the proceedings of the national academy of sciences (PNAS) to get insight into the content [3]. They identify topics that show meaningful aspects of the structure of science and show relationships between different science disciplines. By analyzing documents separated by the year of their publication they

demonstrate research trends. A similar study with a focus on the time evolution of documents is performed by Blei and Lafferty on the archives of the journal Science for the period 1880-2000 [4]. Another specialization of topic extraction for time series is topics over time. The method has performed well on 17 years of NIPS conferences, 21 decades of presidential state-union-addresses, and 9 months of email [5].

Wang et al. [6] use the group topic model to analyze voting data to the senate dataset (voting records of Senators in the 101<sup>st</sup>-109<sup>th</sup> US Senate) and the UN dataset (voting records of the UN general assembly). They identify topic groups that show the voting attitude of groups with respect to different topics.

Topic models are applied by Schmidt et al. [7] to analyze the text which knowledge workers interact with while executing a set of tasks. The emerging topics give hints about the task types and are able to be used to identify structures in the work process.

Despite the work on particular topic modeling methods for specific applications (e.g. time series data), accompanied by a dedicated example, few work exists that shows a general acceptance of these techniques in the scientific communities. Ramage suggests that topic analysis provides valuable support for social scientists, as textual datasets grow in size and scope [8]. The adaptation is complex: on the one hand, the corpora for analysis need to be transformed to specific data formats which can be tedious task for large collections of data. On the other hand, the application of topic modeling techniques requires theoretical background, as long as no “out of the box tool” exists. Recently, libraries like the Stanford Topic Modeling Toolbox (TMT) emerged that support the general application of topic modeling.

Talking about industry transformation indicates a very broad scope that implies large sets of text documents that may be considered relevant. Existing work focuses on specific industries [9, 10] or takes a system perspective [11]. An application of topic modeling approach to literature in the domain of software industry transformation does not exist to our best knowledge.

### **3 Data and Method**

This paper aims at identifying scientific topics relevant to industry transformation research in software business and management. As this is a very broad theme, we used the description provided by the ICSOB 2012 call for papers description in order to identify issues relevant to the overall theme. By this, we obtained nine issues relevant to industry transformation: (1) increasing verticalization of software; (2) software industry evolution and technological change; (3) empirical research issues; (4) security; (5) user experience; (6) green software; (7) internationalization; (8) outsourcing; (9) open source. However, we acknowledge that further issues can be relevant to industry transformation research.

The method applied in this paper relies on unsupervised learning to identify topics from a large dataset. As there is only one relevant conference directly targeted at software business and management, we extended the dataset by major IS conferences. The IS field is relevant as software is part of IS, while having distinguishing characteristics as well [12]. Furthermore, the IS field inherently deals with topics

issues on the intersection between IT and business. Arguably, conferences are more suitable to track current topics as there are large IS conferences with up to twenty different tracks per conference. Thus, in addition to the International Conference on Software Business, our sample comprises six conferences from the Association for Information Systems (AIS): AMCIS, Conf-IRM, ECIS, ICIS, MCIS, PACIS. From those, we collected all publications dating back to 1993, resulting in a total of 14,077 documents for the period 1993-2011.

We downloaded all documents as PDF files using a crawler which is based on the Apache httpclient library. Before the PDF files could be converted to text files readable by the topic analysis tool, we had to apply Optical Character Recognition (OCR) to the PDF files that were saved as pictures and not as text. For that, we used the Adobe Acrobat Professional 9. For conversion to text we used the Apache pdfbox library. As three files could not be processed with this library, we manually copied the text from those files. 278 documents could not be converted to text files due to bad image quality or were recognized as documents not written in English. This left us with 13,799 text files usable for topic analysis.

To identify topics in the document corpus, we use Latent Dirichlet Allocation (LDA) [13]. Deerwester [14] has empirically shown that the latent structure of topics can be recovered based on the co-occurrence structure of terms in text documents. For this, bags of words for document sets are created as input. LDA creates a generative probabilistic model of a text corpus using Bayesian methods. For LDA, documents are random mixtures over topics and each topic is a distribution over words. In the following, we specify this idea and describe the method application. A good introduction is also given in [15] and [16].

In LDA, a word is generated by a convex combination of topics  $z$ . The probability that a word is chosen from a sampled topic-word distribution, i.e. the sum of the probabilities that a word belongs to a topic and that a topic belongs to a document:

$$P(w_i) = \sum_{j=1}^T P(w_i | z_i = j) P(z_i = j)$$

By setting  $P(w|z=j) = \Phi^{(j)}$  and  $P(z) = \Theta^{(j)}$  we get the following equation:

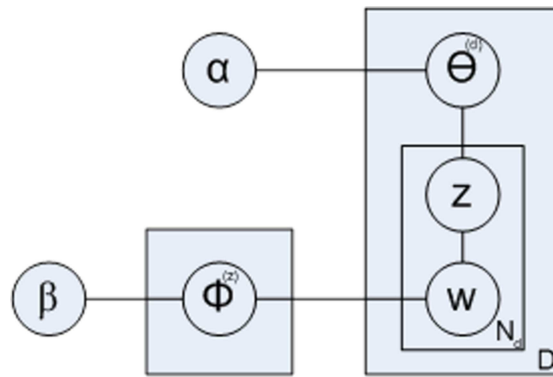
$$P(w_i) = \sum_{j=1}^T \Phi^{(j)} \Theta^{(j)}$$

The distribution referring to  $\Phi$  is the multinomial distribution over words for topic  $j$  and indicates which words are important for which topic.  $\Theta$  says which documents are important for a specific topic.

The following definitions hold for LDA:

- $D$  is the document corpus
- $N$  is the total amount of word tokens
- $d$  is a document
- $d$  consists of  $N_d$  word tokens
- the distributions in the equation above  $\Theta$ ,  $\Phi$  and the topics  $z$  are latent variables

A good overview of the LDA algorithm can be given based on the plate notation (see Figure 1). The outer plate represents documents. The inner plate stands for repeated choices of topics and words within a document. Blei uses a Dirichlet prior on  $\Theta$  with a single hyperparameter  $\alpha$ . As described by Heinrich [15] the process can be seen as follows: LDA generates observable words  $w$ , partitioned by documents  $D$ . For each word, a topic proportion  $\Theta$  is drawn. This is used to sample  $z$ . By this, a relation between words and documents can be identified for  $z$ . Then the corresponding topic-specific term distribution  $\Phi(z)$  is used to draw a word. The topics  $\Phi$  are sampled once for the entire corpus. After finalizing the process, the relevance of a word to a topic and the relevance of a topic to a document have been quantified.



**Figure 1.** Latent Dirichlet Allocation in plate notation.

An important aspect is the need of the definition of the amount of topics to be sampled by LDA. It assumes that the number of topics is given. As the actual number is unknown, an optimal number of topics can be estimated using an objective function, such as the perplexity. The perplexity “is monotonically decreasing in the likelihood of the test data, and is algebraically equivalent to the inverse of the geometric mean per-word likelihood” [13]. The lower the perplexity score, the better the generalization performance for the given number of topics. The perplexity curve is U-shaped, where the minimum indicates the optimal number of topics.

To improve the results, different operations can be performed on the text data, before the bags of words per documents that are input to LDA are created. A language detection based on language specific n-grams helps to focus on one language, if the corpus contains documents in several languages. Stemming and part-of-speech tagging help to standardize the used words and filter those word types considered meaningful, e.g. verbs and nouns. Stopword lists help to filter words that are considered as not meaningful.

**Table 1.** Sample distribution over the sample period.

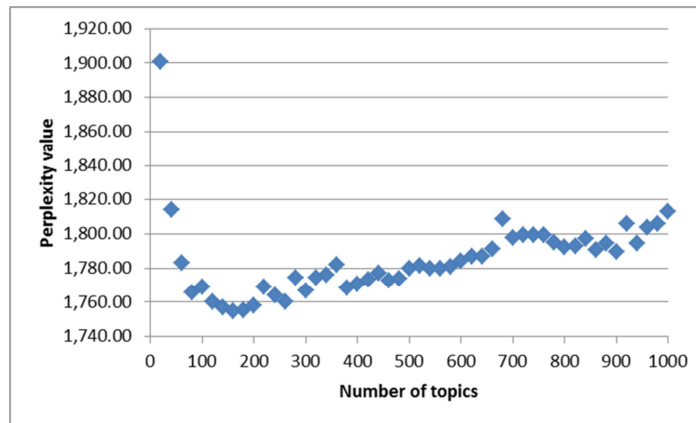
Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Documents	2	3	4	2	84	82	34	11	19	23	42	65	60	69	63	68	92	78	90

As we focus on topics relevant to industry transformation research and the calculations are very intense, we reduced the sample by papers that appeared not to be relevant. For that, we computed 120 LDA topics. The value of 120 was selected using an optimal perplexity value. Out of these topics, 30 topics could be mapped to the issues in industry transformation research. The final document sample comprises documents that are at least 50% relevant to these 30 relevant topics, leaving us with a sample of 891 documents. The sample distribution is shown in Table 1.

#### 4 Topics in Industry Transformation Research

Within this section, we present and discuss the results of the topic analysis. As described above, the topics have been extracted from the 891 documents that appeared to be relevant to the ICSOB 2012 track “Industry Transformation”.

In order to obtain the optimal number of topics, we calculate the perplexity value for 20 to 1,000 topics (using steps of 20). The results are depicted in Figure 2. From this graph it can be seen that the optimal number of topics is around 160. Thus, we run LDA with 160 topics to retrieve the actual topics.



**Figure 2.** Perplexity value for different numbers of topics.

As the obtained number of topics is too large to be reviewed here individually, we map those to the industry transformation issues first and analyze them on this aggregated level further on. The mapping was done by two authors in three steps:

(1) exclusion of “noisy” topics; (2) individual expert mapping of topics to track issues; and (3) detailed analysis of mismatches. The main result is shown in Table 2.

The noisy topics could be easily identified by both authors individually as they contained nonsense words or conference names and locations. Further, some automatic rules have been applied to identify irrelevant topics: (1) at least three documents must have a relevance of at least 1%; (2) at least one source must have a relevance of at least 50%; (3) total topic relevance must be at least 100 (this is a low value, as the average relevance is 11,000); (4) the term with the maximal relevance must attribute for at least 1% of the total topic relevance. This left us with 101 topics.

Within the separate mapping both authors assigned each topic to exactly one track issue. 70 mappings have been assigned consistently by both. For the remainder, both authors reviewed them together and either excluded the topics or assigned them to an issue. Where no decision was possible, the relevant documents were briefly reviewed in order to make a decision. Finally, this process resulted in 74 relevant topics.

Given the association of LDA topics to conference issues, we identified the documents relevant to the conference issues. For that, we assigned each document to the most relevant issue, where the relevance was bigger than 1%. As a result, we obtained an association of 701 documents to the given conference issues.

The results in Table 2 show the topics per industry transformation issue and the document numbers associated with them. In the following paragraphs, all issues are briefly reviewed. Further, as a detailed example, we provide a profound review of the topics related to industry transformation issue Green IT.

It becomes apparent that the issue of verticalization in software markets is not associated with any topics or sources. In general, this result confirms our feeling that verticalization is not a major issue in software business and IS research yet.

The issue of empirical research in software business has a cross-cutting role as it focuses on a method that is applicable to all other issues. Naturally, we find topics there, which explicitly deal with measurement and data collection. Another stream is related to stock market studies, as those are empirical in nature and not associated with the other issues. Thus, in order to identify sources applying particular methods, both relevance measures should be analyzed: the relevance to the method and the particular field.

The largest issue deals with industry evolution and technical change as it spans multiple streams. The main streams can be divided in technology adoption and diffusion, social networks and web 2.0, and ICT development of countries. Furthermore, we find four topics associated with cloud computing representing 29 documents: application service providing (ASP) and software as a service, cloud computing and pricing, as well as service level agreements (SLA). Two topics with 26 documents represent the mobile technology.

Conference issue	LDA topic	Sources	LDA topic	Sources
Verticalization	<i>no topics and no sources</i>			
Impact of industry evolution and technological change(e.g. mobile, on-demand, cloud)	adopt innov diffus technolog roger	9	network access cost wireless technolog	12
	adopt organ organiz technolog institut	10	phone countri 0 readi network	3
	asp risk applic vendor saa	5	platform open develop applic technic	1
	broadband adopt consum research	13	resist inhibitor adopt non-adopt belief	5
	internet			
	cloud price resourc comput revenu	12	rfid tag technolog reader retail	7
	commun social presenc virtual interact	2	site answer wikipedia portal q&a	2
	countri develop industri econom sector	7	social movement polit media twitter	3
	develop ict countri access africa	15	social network user facebook onlin	14
	ict countri develop fdi corrupt	6	technolog micro-blog social commun http	2
	ict develop countri econom sustain	5	telecommun broadband industri polici	2
			monopoli	
	invest telecom econom economi countri	0	vol attitud smart control meter	3
	mobil servic phone consum devic	23	web 2.0 applic content share	5
	mobil wireless commerc solut applic	3	web workflow execut leas sla	6
	model workload oper instanc sla	6	wireless mobil network user devic	24
	nation countri diffus technolog develop	10		
				215
Open-source software business and management	softwar sourc open oss product	9		
				9
Empirical research issues in software business and management	breach market event return announc	7	model measur construct research item	63
	market futur index predict stock	3	qualiti dimens servic measur user	18
	market invest stock announc reaction	7	web-bas survei respons web mail	3
	measur success model dimens survei	10		
				111
Issues in user experience	aesthet color websit design usabl	3	usabl develop user oss cost	0
	bank risk phone cell perceiv	0	usabl qualiti web design instrument	4
	continu satisfact perceiv user intent	15	usag behavior organiz implement user organ	7
	enjoy motiv intrins perceiv activ	5	user design devic interfac task	9
	internet onlin user usag access	0	user profession percept satisfact congruenc	1
	satisfact expect user factor perform	9	user visual input access error	9
	success user satisfact qualiti model	14	virtual world life second user	0
	technolog perceiv model accept intent	134		
				210
Issues in security	concern secur factor trust dimens	3	protect threat password secur cost	4
	kei secur authent network encrypt	9	secur bluetooth comput respond health	3
	osn privaci user trust provid	6	secur home comput behavior practic	2
	person data concern privaci user	5	secur polici employe comput organ	10
	polici user privaci wap site	3	tag user anonym tor latenc	5
	privaci concern person control trust	26	vulner patch attack vendor time	5
	privaci protect polici iso control	4		
				85
Issues in "green" software	green environment motiv technolog	5	grid cost energi power server	5
	sustain			
				10
Internationalization of software firms	cultur organiz organis technolog	4	cultur valu nation countri dimens	20
	internet			
				24
Outsourcing	offshor distanc outsourc term nearshor	4	outsourc risk bpo deal market	4
	outsourc cost decis relationship develop	7	outsourc success cost benefit organ	13
	outsourc relationship contract manag	9		
	govern			
				37
Number of issues: 9	Number of LDA topics: 74		Number of sources: 701	



The second-largest issue is user experience. By far the largest topic deals with the technology acceptance model. Interestingly, it appears that terms are used especially consistently in this stream, as 134 documents are associated with one topic. This is different for the stream related to the IS success model, which comprises various LDA topics. Another stream in this issue deals with the design and usability of systems and interfaces.

The issue security encompasses 85 documents, covering diverse streams. Those include privacy, trust, anonymity, and policies in organization. Rather technical streams are vulnerability of systems and encryption. Minor streams cover Bluetooth and home security.

37 documents cover the issue outsourcing. The topics appear to be quite indicative with regard to the stream discussed. Those include (1) the definition of terms; (2) the outsourcing decision; (3) the contract management; (4) the risk associated with outsourcing; and (5) the success of outsourcing.

Two rather small issues are open source software and internationalization of software firms. The open source issue includes just one topic. However, a closer look indicates that the included documents are quite diverse. It appears that they have been summarized under this topic due to their high relevance to open source, but they also deal with other topics such as usability and security. The issue internationalization subsumes two cultural topics, whereas the bigger stream deals with cultural differences in different countries. The smaller stream deals with cultural differences on the organizational level.

The issue green software is discussed in more detail here. The comprising topics and associated documents are shown in Table 3. We combine the detailed view with a short case study, in order to evaluate if the topic analysis yields useful results for literature search.

**Table 3.** Documents associated with Green IT.

Terms	Relevance	Title	Year	Conference	"Green IT" in title
green environment motiv technolog sustain	0.35	Testing Multimedia for Ecological Sustainability	1998	AMCIS	no
	0.62	Green IT Adoption: A Motivational Perspective	2011	PACIS	yes
	0.72	Organizational Motivations for Green IT: Exploring Green IT Matrix and Motivation Models	2009	PACIS	yes
	0.68	IT and Eco-sustainability: Developing and Validating a Green IT Readiness Model	2009	ICIS	yes
	0.04	Solving the Traffic Problem by Using A Simulation Model	2008	Conf-IRM	no
grid cost energi power server	0.72	Does Green IT Matter? Analysis of the Relationship between Green IT and Grid Technology from a Resource-Based View Perspective	2009	PACIS	yes
	0.37	Reducing datacenter energy usage through efficient job allocation	2011	ECIS	no
	0.76	A multi-model algorithm for the cost-oriented design of the information technology infrastructure	2003	ECIS	no
	0.69	The Impact of MIS Software on IT Energy Consumption	2010	ECIS	no
	0.38	Grid Technology as Green IT Strategy? Empirical Results from the Financial Services Industry	2010	ECIS	yes

In order to analyze the applicability of LDA to literature identification, we run a benchmark test on the defined document sample. By searching for “green IT” within the titles of all 701 documents, we retrieved five documents, which are also included in the LDA results list. Further, we searched for the term “green” only and identified one further paper “Why are consumers going green? The role of environmental concerns in private green-IS adoption” [17]. In our approach, this paper was not assigned to the green IT topics as its content is more related to an industry evolution topic (smart meters) than to a Green IT topic (only 5%).

While the benchmark validation by standard term search confirms that our topics do not omit any documents, our Green IT topics, notably, cover five further documents that do not contain “green IT” within their title. Thus, content-wise related papers appear in our search result. These documents would be hard to find with predefined keywords as their titles are very heterogeneous, whereas a search on the full text of all articles would most likely yield too many unrelated results.

We further examined the content of the retrieved documents in order to evaluate if the selected documents fit to the proposed topics. In general, our findings confirm the logical soundness of the topics. While the first topic mainly comprises documents dealing with an organizational perspective (motivation and readiness for green IT as well as adoption) of green IT, the second topic primarily consists of documents

focusing on technical matters (energy consumption and costs of IT infrastructure such as grid technology). However, the relevance of the assigned documents may not be neglected. For instance, the paper “Solving the Traffic Problem by Using a Simulation Model” [18] has only a 4% relevance and is hence only partially related to the other papers’ content. The proposed optimization algorithms of the traffic domain do relate in certain aspects to green IT (e.g. reduce carbon emissions or transfer optimization approach to the ones in datacenters), but not to the full extent. Consequently, the relevance values for each document need to be considered.

## 5 Topic Evolution

Figure 3 shows the development of the conference issues over the sample period. The values reflect the relevance of the issues. Whereas the absolute altitude is not indicative, the changes in time and relative differences between the values reflect relative differences in relevance. We excluded the issues empirical research and internationalization of firms in order to make the figure more readable. Whereas the issue internationalization was too unimportant, the empirical research issue deals with a particular method. Thus, we decided to focus on the other issues.



**Figure 3.** Evolution of the industry transformation issues over time.

For each year and each issue, the relevance is calculated using the associated LDA topics. For each topic, each of the 891 documents has a relevance value for the topic. The sum of all these relevancies is the topic relevance. The issue relevance is the sum of all relevancies from associated topics. Interestingly, nearly the same results were obtained by looking at the absolute number of documents and associating each document with just one topic, as we did in section 4. As working with these fragment-relevancies is the common procedure [13] and is provided directly through Stanford TMT, we use this representation form in this section.

The results indicate that the largest issues being industry evolution and user experience continue to increase in relevance. However, as those are quite broad issues, comparisons with more specific topics such as open source and green software are less meaningful.

Security appears to be of continuous interest. Whereas it only came up around the year 2000, it soon became more important than outsourcing and continuously grew in relevance faster than all the smaller issues.

Outsourcing shows a similar behavior to security until the year 2003. The following five years, outsourcing reaches its highest attention, before the interest even declines. It appears that the relevance starts to stabilize again, even though it is hard to conclude on a change from the last two years.

Open source displays a similar development as outsourcing, but with a lower relevance over the entire period. Also, there is no indication of a change in the trend downwards.

Apart from two small occurrences, one publication in 1998 and one in 2003, Green IT is a very recent topic. Eight out of ten relevant documents have been published since 2008. These figures go along with the observation by Molla et al. [19], concluding that despite increased attention from business and government, the interest from IS researchers is a more recent phenomenon. It hence can be assumed that more publications are yet to come with increasing scientific interest.

## 6 Conclusion

Topic analysis can be useful for overviewing the topics in a particular field and their relevance in time. Furthermore, it helps to identify sources relevant to particular topics and can therefore support literature search.

In this paper, we identified topics relevant to industry transformation research in software business and management. For that, we performed a topic analysis of scientific publications in IS research using the LDA approach. Our sample comprised 13,799 documents from seven conferences: ICSOB, AMCIS, Conf-IRM, ECIS, ICIS, MCIS, and PACIS. We assigned the topics identified by LDA to the various issues in software industry transformation research. Thus, indicating the particular topics within the broad issues and their development in time. A short case study on the issue Green IT confirmed that topic analysis can be used to identify relevant sources beyond the possibilities of mere search engines. An overview of the development of the topics in the sample period in general confirmed our impression, e.g. the diminishing interest in outsourcing, steady increase of security topics, and the recent rise of Green IT.

We contribute to the software business and management research by identifying topics relevant to industry transformation, as well as through the analysis of their development over time. Furthermore, we demonstrate how topic analysis can be used for literature search and evaluation. Researchers in the area of software business and management will find the results useful in order to overview topics relevant to their field. The results on source identification further suggest a powerful method for assisted literature search and review beyond search engines. Practitioners can use our

results in particular and the method in general in order to identify threats and opportunities to their businesses. The sample can be easily switched to other data sources such as social networks or company wikis, as the input documents are mere text files.

There are certain limitations to this paper. Due to the large sample, we could not process the raw data manually. Also, we could not review all final sources in detail. Furthermore, as no sufficient dataset is available on the software business and management field in particular, we focused on the larger set of IS research. Finally, the evaluation of topics can be greatly improved by applying a structured approach such as systematic mapping. In our further research we attempt to extend the sample by major journals, include publications from the software and business field, apply the topic analysis to issues beyond industry transformation, and provide a software tool for literature search that incorporates our method and findings.

## Acknowledgments

This work was supported by the Software-Cluster and partially funded within the Leading-Edge Cluster competition by the German Federal Ministry of Education and Research (BMBF) under grant “01|C10S05”. The authors take the responsibility for the contents.

## References

1. Lieberman, M.B. and Montgomery, D.B.: First-mover advantages. *Strategic Management Journal* 9, 41-58 (1988)
2. Katz, M.L. and Shapiro, C.: Network Externalities, Competition, and Compatibility. *American Economic Review* 75, 424-440 (1985)
3. Griffiths, T.L. and Steyvers, M.: Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America* 101, 5228 (2004)
4. Blei, D.M. and Lafferty, J.D.: Dynamic topic models. pp. 113-120 (2006)
5. Wang, X. and McCallum, A.: Topics over time: a non-Markov continuous-time model of topical trends. pp. 424-433 (2006)
6. Wang, X., Mohanty, N., and McCallum, A.: Group and topic discovery from relations and text. pp. 28-35 (2005)
7. Schmidt, B., Stoitsev, T., and Mühlhäuser, M.: Analysis of Knowledge Work Execution at Computer Workplaces. In: *Proceedings of 12th European Conference on Knowledge Management* (2011)
8. Ramage, D., Rosen, E., Chuang, J., Manning, C.D., and McFarland, D.A.: Topic modeling for the social sciences. (2009)
9. Berger, A.N., Kashyap, A.K., Scalise, J.M., Gertler, M., and Friedman, B.M.: The transformation of the US banking industry: What a long, strange trip it's been. *Brookings papers on economic activity* 1995, 55-218 (1995)
10. De Vany, A. and David Walls, W.: Natural gas industry transformation, competitive institutions and the role of regulation:: Lessons from open access in US natural gas markets. *Energy policy* 22, 755-763 (1994)

11. Dennis, D. and Meredith, J.: An empirical analysis of process industry transformation systems. *Management Science*, 1085-1099 (2000)
12. Messerschmitt, D.G. and Szyperski, C.: *Software Ecosystem*, Cambridge, MIT Press (2003)
13. Blei, D.M., Ng, A.Y., and Jordan, M.I.: Latent dirichlet allocation. *The Journal of Machine Learning Research* 3, 993-1022 (2003)
14. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., and Harshman, R.: Indexing by latent semantic analysis. *Journal of the American society for information science* 41, 391-407 (1990)
15. Heinrich, G.: Parameter estimation for text analysis. *Fraunhofer Technology Report* (2009)
16. Steyvers, M. and Griffiths, T.: Probabilistic topic models. *Handbook of latent semantic analysis* 427, 424-440 (2007)
17. Kranz, J. and Picot, A.: Why Are Consumers Going Green? the Role of Environmental Concerns in Private Green-is Adoption, in *ECIS 2011 Proceedings* (2011)
18. Yang, C.-L. and Wen, W.W.: Solving the Traffic Problem by Using A Simulation Model. In: *CONF-IRM 2008 Proceedings* (2008)
19. Molla, A., Cooper, V.A., and Pittayachawan, S.: IT and Eco-sustainability: Developing and Validating a Green IT Readiness Model. In: *ICIS 2009 Proceedings* (2009)