Identifying Innovative Documents: Quo vadis?

Ivonne Schröter[1,2], Jacob Krüger[1,3], Philipp Ludwig[1], Marcus Thiel[1], Andreas Nürnberger[1] and Thomas Leich[2,3]
[1]*Otto-von-Guericke-University, Universitätsplatz 2, 39106, Magdeburg, Germany*
[2]*METOP GmbH, Sandtorstraße 23, 39106, Magdeburg, Germany*
[3]*Harz University of Applied Sciences, Friedrichstraße 57-59, 38855, Wernigerode, Germany*

Keywords: Enterprise Information System, Empirical Study, Survey, User Questionnaire, Requirements Engineering.

Abstract: The number of new research documents and patents published each year is steadily increasing. Despite this development, identifying innovative documents in a timely manner has received only little attention in research. Nevertheless, this use case is important for companies that strive to keep up with current innovations in their field. However, since existing solutions do not take context and background of the particular firm or researcher into account, they fall short in supporting the user in his search for suitable documents. In this paper, we describe an industrial case study we conducted within sheet-metal working companies and related research institutes in Germany. We *i)* report a qualitative study on innovation research, *ii)* provide a list of features that industrial researchers demanded, and *iii)* discuss implementation challenges for systems that support interactive retrieval of innovative documents. Based on the initial results, we argue that existing systems fall short to provide an integrated workflow. Overall, we discuss how to implement such a system and the corresponding problems.

1 INTRODUCTION

Monitoring patents and academic publications is an essential task for scientists and companies alike to not miss opportunities for innovation, research, or workflow improvement. However, this usage scenario differs greatly from conventional research processes: Evaluating how innovative a document is requires additional effort and thorough reviews. Hence, assessing innovative documents provides new challenges to search engines, user guidance, and the presentation of results (Kuhlthau, 1991; Marchionini, 2006).

A lot of digital libraries are available (Meyyappan et al., 2000; Aghaei Chadegani et al., 2013) and several approaches to browse and explore such collections have been presented in the past (Lehmann et al., 2010; Lashkari et al., 2009). Still, only little attention has been brought to the task of helping users in identifying innovative documents (Xie, 2006). While topic-oriented search engines such as *dblp*[1] (in the domain of computer science) are available, these systems are hardly sufficient to consider individual demands (Frias-Martinez et al., 2006; Jayawardana et al., 2001). For example, experts interested in new development in their field are most likely not interested in finding patents published by the company they work for or which they published themselves. However, these documents might be considered new by other people.

Furthermore, experts have to sight a great variety of resources, for instance, papers, patents, or norms. This is especially a problem in smaller firms, where sufficient resources for technological monitoring may not be available. To this point, these users could benefit from systems that support them in retrieving documents and patents relevant to their current problems.

Hence, we argue that innovation researchers require an tool that integrates all steps of search processes and considers their context. In order to discuss this claim, we conducted a case study during which we interviewed experts in the field of a particular industrial domain, sheet-metal working. We further analyzed our findings, derived opportunities to support the search process, and possible problems with two experts. Overall, we provide a detailed discussion on important features for innovation research. In particular, we describe the following:

- A case study we conducted in the sheet-metal

[1]http://dblp.uni-trier.de/, 06.09.2016

Schröter, I., Krüger, J., Ludwig, P., Thiel, M., Nürnberger, A. and Leich, T.
Identifying Innovative Documents: Quo vadis?.
DOI: 10.5220/0006368706530658
In *Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017) - Volume 1*, pages 653-658
ISBN: 978-989-758-247-9

working domain to identify characteristics of the corresponding search process for innovative documents. With this, we analyze search and evaluation behavior of users and direct further research.

- A discussion on potential features and challenges to implement systems to support search processes. We argue for our position that such systems are required for users and researchers but also result in new problems.

In this paper, we first introduce design and results of our empirical study in Section 2. We discuss the results and their implications in Section 3. Finally, we summarize our contributions in Section 4.

2 STUDY DESIGN

The goal of our study was to identify experts' workflows while they investigate innovations. Therefore, we designed a four-step study with close cooperation to industry and experts. We illustrate our workflow in Figure 1.

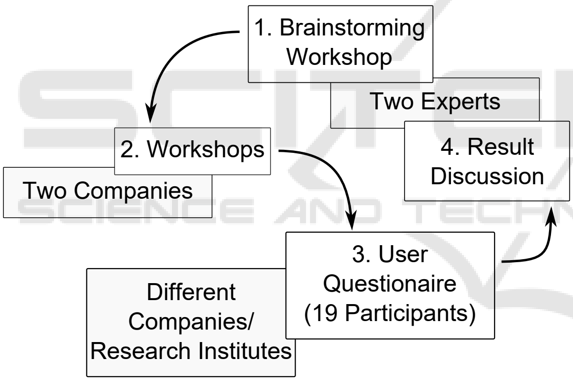

Figure 1: Steps of our conducted study.

Firstly, we conducted a brainstorming workshop with two experts in the field of sheet-metal research. We discussed workflows and tasks in their search and analysis processes. As a result, we have identified possible participants and questions for our next steps.

Secondly, we performed workshops in two sheet-metal working companies. During this on-site study, we asked industrial researchers to specify their work processes. In addition, we monitored their search and analysis behavior to identify steps and tools in their workflow. This allowed us to obtain a first overview about the research focus and potential points of support with tools.

Thirdly, based on the results, we developed a user questionnaire and sent it to several sheet-metal working companies and research institutions. 19 employees in different positions rated ideas that we gathered in the previous steps. One main result is a practice-driven rating of important features in applications in order to support the research for innovation.

Finally, we discussed our observations with the experts that aided us during our initial workshop. With them, we summarized our results and closed remaining gaps. Moreover, we identified and discussed the industrial context in research for innovation and corresponding tools. In the following sections, we provide detailed information on applying our study in the proposed context (i.e., Steps 2 and 3 in Figure 1).

2.1 Workshops

For our on-site case study (Step 2 in Figure 1), we performed workshops in two companies. During these, we used guided interviews to receive information from domain experts. For this, we prepared a catalog of questions to ensure similar treatment. These questions were acquired through the previous brainstorming workshop (Step 1 in Figure 1). We used a fixed process to slowly introduce the interviewees from general aspects of their research behavior to our actual goal. To do this, we performed the following steps:

1. Explaining the procedure of our study.
2. Investigating the reasons for research.
3. Identifying the research process.
4. Detecting tasks that can be supported with tools.
5. Reviewing these points of support with the interviewees.

To address all steps, we developed questions with different aims and scopes. For instance, we wanted to identify the expertise of our workshop groups while investigating their reasons for research. These questions were on a more general level, for example, we asked: "*How often do you research?*", "*To what purpose do you research?*", and "*How much expertise do you have in technological research?*" to slowly introduce the participants into our study. Other questions, for instance, "*Is there an operational workflow for the search/analysis of innovative documents?*" or "*Which resources do you access in your research process?*", helped us to identify workflows and opportunities for tool support.

We combined the results at the end to summarize views on the same questions. The agreement on perceived problems and challenges was quite high, resulting in a unified view. This means, that most participants reported similar on their search behavior and its challenges. Hence, we identified five main issues:

- Research is mainly done with classical search engines for the web (e.g., Google), patents (e.g., DEPATISnet[2] (Jürgens and Herrero-Solana, 2015)) or other resources (e.g., research institutes).
- Direct search for innovations is usually not possible, for instance, due to only partial availability of solutions. Hence, an extensive search with seemingly fitting keywords is done. Also, incremental specification of the research is conducted in an exhaustive manner.
- Exchanges on the research rarely or never happen, especially not across several departments. This proves prior results in the professional search domain (Knight and Spink, 2008; Nürnberger et al., 2015).
- Pre-filtering is usually done by source, institution, or abstract. This may leave out potentially interesting and relevant documents.
- Patents must be carefully read and evaluated, while pre-selection is done via fitting, but also broad, keywords (Lupu et al., 2013).

In order to address these issues, workshop participants proposed solutions. Additionally, we deduced possibilities from the described search processes. To reduce the effort and time necessary for our further study, we limited the number of potential solutions to the ten ideas that the majority of participants emphasized. We illustrate these ideas in Table 1.

Table 1: Proposed ideas for supporting research.

P-1	Marking and extracting parts of a document.
P-2	Creating a research history.
P-3	Comparing different search result lists.
P-4	Development of a topic over time (trend analysis).
P-5	Listing important institutions and authors for a topic.
P-6	Filtering documents by domain, country of origin, etc.
P-7	Comparing different versions of the same document.
P-8	Sorting of documents by their topics.
P-9	Referencing documents to other publications.
P-10	Summarizing parts of a document.

Some of these proposals are already implemented in similar applications. For instance, marking document parts (P-1) is possible in most digital readers and comparing different document versions (P-7) is part of version control systems to some extend. Also, many features for filtering, sorting, or similar tasks are partly supported in reference management systems (Gilmour and Cobus-Kuo, 2011). However, we are not aware of an application which combines all ideas in an integrated system and supports search processes as a whole.

[2]https://depatisnet.dpma.de, 06.09.2016

To this point, we argue that the most important ideas have to be identified and defined as requirements. Further assessing context information and investigating potential challenges is necessary. In order to initiate corresponding research, we conducted user questionnaires to evaluate the ideas.

2.2 User Questionnaire

During the third step (i.e., Figure 1), we determined the importance of previously proposed ideas. To this point, we conducted a user questionnaire with 19 experts from different sheet-metal working companies or related research institutes. We asked them to rank the ideas mentioned in Table 1 from 1 (very important) to 10 (not that important). Each priority number could only be selected once. In addition, the proposals could be rated in high, middle and low relevance (as a complementary ranking) to track whether the participants understood the system. Furthermore, we recorded additional ideas, if a participant proposed an important feature the workshops did not cover.

We did not evaluate five questionnaires, since they were not completely or wrongly filled out. For example, participants described an idea as important but to have low relevance (the complementary ranking), which is contradictory, or used the same rank multiple times. We display the results of the remaining 14 responses in Table 2. While we are aware that this can only be considered an initial sample, we could still identify the most important ideas from Table 1 that were consistent over almost all questionnaires to reason our position. For example, summarizing (P-10) and referencing documents (P-9), as well as listing important authors (P-5) were regularly demanded. In contrast, the participants considered pre-filtering of documents (P-6) as least relevant.

Several additional ideas were named, for example to create a report from result lists, adding information about addresses of institutions and contact persons, or to display the relevance of a document to certain topics. During our evaluation, we mapped these ideas with those we already had. Then, we merged all results and discussed them with the two experts that initially helped us. We identified opportunities but also challenges to support the process of retrieving innovative documents.

3 DISCUSSION

During our study, we found several tasks for innovation research we can support with tools. However,

Table 2: Results of the user questionnaire. Smaller priority numbers indicate a more important proposal. Thus, a low average ($\bar{x}$) shows the ideas most interesting to the participants.

Proposal		Priority										
		1	2	3	4	5	6	7	8	9	10	$\bar{x}$
P-1:	Mark/extract document parts	1	1	2	1	1	0	1	3	0	4	**6.36**
P-2:	Research history	0	0	0	0	2	3	4	2	0	3	**7.29**
P-3:	Compare result lists	0	0	1	0	4	4	1	2	1	1	**6.36**
P-4:	Temporal evolution	0	3	2	3	1	2	0	1	2	0	**4.79**
P-5:	List institutions and authors	3	1	4	1	0	0	2	0	3	0	**4.43**
P-6:	Pre-filter by domain, etc.	0	0	1	1	0	1	2	3	3	3	**7.71**
P-7:	Display version differences	1	1	0	1	2	2	2	0	3	2	**6.43**
P-8:	Sort documents by topic	1	2	1	3	1	2	1	2	0	1	**5.00**
P-9:	Reference documents to another	3	3	2	2	2	0	0	1	1	0	**3.57**
P-10:	Summarize parts of a document	5	3	1	2	1	0	1	0	1	0	**3.07**

not all of them are engineering tasks but require additional research. In further discussions with the two experts (i.e., Step 4 in Figure 1), we identified several challenges. In the following, we analyze the *research context*, *how to support search processes*, and the *evaluation of innovation degree*. For each aspect, we discuss its importance and map it with the proposals we display in Table 2.

3.1 Research Context

We argue that one of the main aspects of innovation research is its context. During our studies, we found that most researchers demand for features that improve search speed (e.g., P-1, P-6, P-10) or help them to refine results to their background (e.g., P-6, P-8, P-9). Overall, during our study we identified four main types of context information that we display in Figure 2 and which influence searches for innovative documents:

- *User*: The users' background influences which innovations are important to them. For example, an employee who works on aluminum research may not be interested in copper, but it could be important.
- *Company*: A company's research agenda or market segment further scope which innovations are interesting for its researchers. For instance, if a company develops metal it might be unnecessary to search documents in regard to plastic research.
- *Documents*: Researchers might be interested in specific documents, such as, its type, innovation degree, or topic, which define the basis of further research. For example, when informing about a new alloy, research articles are interesting while patents are important when development shall start.
- *Trends*: Trends help researchers to assess in which direction a specific field evolves. For instance, they can assess which production methods are out of date or will become standards.

So, a suitable tool for innovation research should track and utilize these context information. Thus, we can influence the search process and evaluation results without direct intervention of the user. For instance, a tool could automatically filter documents to context specifics, such as a domain or country (P-6). This leads to improved results and less time consuming search processes. Still, users may want to search documents outside of their context, which must also be supported.

In the following, we propose features for a system that integrates the context in a search process. Furthermore, we discuss a main challenge in this regard: Determining the innovation degree of documents.

3.2 Supporting Search Processes

To support the development of a corresponding tool, we discussed approaches to support searches for innovative documents. In particular, the proposed features allow users to easier retrieve documents related to their context. The following examples provide an initial overview:

- *Extracting meta data and abstracts* helps to summarize, filter, and sort documents (P-6, P-8, P-10), to identify leading names in a topic (P-5), or to reference literature (P-9).
- *Filtering and categorizing found documents* enables researchers to map their results to specific topics (P-8) or separate the works of specific authors and institutions (P-5).
- *Automated text generation* helps developers to

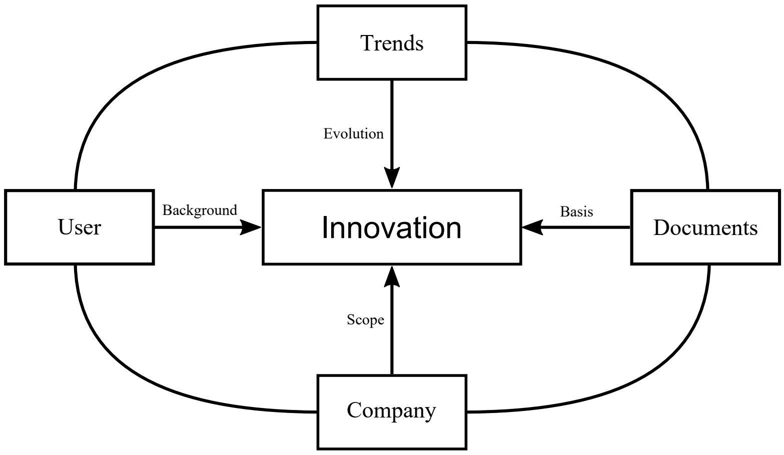

Figure 2: Context of innovation research.

faster summarize gained information (P-10) or extract marked text (P-1).

- *Analyzing and visualizing the publication date* can help researchers to overview temporal developments of a specific topic (P-4).
- *Storing search processes separately* provides the possibility to create a research history and reuse old results (P-2).
- *Illustrating results in a list* provides an overview of found documents and allows to compare literature (P-3) or different versions (P-7).
- *Supporting an ontology tree* allows researchers to store and connect keywords, allowing them to filter (P-6), sort (P-8), and assess documents.

These are only some examples of features that can help users. As we stated before, several applications have been introduced to partly support such functions. However, we emphasize the importance of the features and of a tool that integrates these into a defined process. Many of these features are part of more general tasks in exploratory search as discussed by Marchionini (Marchionini, 2006) and, thus, elements of corresponding retrieval tools could be applied here.

3.3 Evaluating Innovation Degree

In addition to the interactive support of the users' search workflow, we propose to use semi-automatic evaluation of a documents' innovation degree. For instance, we could base this analysis on previously assessments of users and provide them to a recommender system (Felfernig et al., 2007). Also, simpler mechanisms may provide assistance, for example evaluating the development of a topic over time.

Increasing numbers of citations or publications can be an indicator for a research field that is significant for practice. Other interesting factors are the topics addressed or combined, the authors and their networks, or where the document is published. For example, a new topic introduced by a reputable author at an innovative conference is likely to have high innovation degree. Still, automating this analysis requires further investigations. We argue that this is the most challenging step towards automation and considering context information is essential.

Overall, there are several features that can support innovation research. While these are partly supported in existing tools, in our study users asked for further improvements and integration into a workflow. To describe a scope for corresponding research, we argue that context information are the most important aspect.

4 CONCLUSIONS

The number of new research documents increases continuously. Hence, for researchers in academia and companies it becomes more complicated to identify and assess innovative documents. Existing search engines cannot adapt to consider a users background for a given search task. Also, there are different information needs that require the usage of adopted search engines, e.g. for patent retrieval. This results in a great number of potentially interesting documents that must be assessed and aggregated manually.

We argue that considering context information and applying an integrated workflow are essential. To support this position, we reported the results of a qualitative study on industrial research processes. We found strong points on which tasks are costly but could be (partially) automated. Furthermore, we identified several features this user group asks for. Finally, we discussed approaches and challenges that must be addressed while implementing a corresponding system and in further research.

Based on the results, we aim to develop a suitable tool for the search and assessment of innovative documents. While we already identified some features and challenges, this will require additional discussions. Thus, we aim to conduct further qualitative and quantitative user studies and discussions with experts.

ACKNOWLEDGEMENTS

This research is supported by BMWi grants KF3358702KM4, KF3358803KM4, KF2885203KM4, DFG grant LE 3382/2-1, and Volkswagen Financial Services AG.

REFERENCES

Aghaei Chadegani, A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., and Ale Ebrahim, N. (2013). A Comparison Between Two Main Academic Literature Collections: Web of Science and Scopus Databases. *Asian Social Science*, 9(5):18–26.

Felfernig, A., Friedrich, G., and Schmidt-Thieme, L. (2007). Introduction to the IEEE Intelligent Systems Special Issue: Recommender Systems. *IEEE Intelligent Systems*, 22(3):18–21.

Frias-Martinez, E., Magoulas, G., Chen, S., and Macredie, R. (2006). Automated User Modeling for Personalized Digital Libraries. *International Journal of Information Management*, 26(3):234–248.

Gilmour, R. and Cobus-Kuo, L. (2011). Reference Management Software: A Comparative Analysis of Four Products. *Issues in Science and Technology Librarianship*, 66(66):63–75.

Jayawardana, C., Hewagamage, K. P., and Hirakawa, M. (2001). A Personalized Information Environment for Digital Libraries. *Information Technology and Libraries*, 20(4):185–196.

Jürgens, B. and Herrero-Solana, V. (2015). Espacenet, Patentscope and Depatisnet: A Comparison Approach. *World Patent Information*, 42:4–12.

Knight, S. A. and Spink, A. (2008). Toward a Web Search Information Behavior Model. In *Web search*, pages 209–234. Springer.

Kuhlthau, C. C. (1991). Inside the Search Process: Information Seeking from the User's Perspective. *Journal of the American Society for Information Science*, 42(5):361.

Lashkari, A. H., Mahdavi, F., and Ghomi, V. (2009). A Boolean Model in Information Retrieval for Search Engines. In *International Conference on Information Management and Engineering*, ICIME, pages 385–389. IEEE.

Lehmann, S., Schwanecke, U., and Dörner, R. (2010). Interactive Visualization for Opportunistic Exploration of Large Document Collections. *Information Systems*, 35(2):260–269.

Lupu, M., Hanbury, A., et al. (2013). Patent Retrieval. *Foundations and Trends in Information Retrieval*, 7(1):1–97.

Marchionini, G. (2006). Exploratory Search: From Finding to Understanding. *Communications of the ACM*, 49(4):41–46.

Meyyappan, N., Chowdhury, G. G., and Foo, S. (2000). A Review of the Status of 20 Digital Libraries. *Journal of Information Science*, 26(5):337–355.

Nürnberger, A., Stange, D., and Kotzyba, M. (2015). Professional Collaborative Information Seeking: On Traceability and Creative Sensemaking. In *Semanitic Keyword-based Search on Structured Data Sources*, pages 1–16. Springer.

Xie, H. I. (2006). Evaluation of Digital Libraries: Criteria and Problems from Users' Perspectives. *Library & Information Science Research*, 28(3):433–452.