Universität Ulm
Fakultät für Mathematik und
Wirtschaftswissenschaften

Research in Business Process Management: A bibliometric analysis

Diplomarbeit

in Wirtschaftswissenschaften

vorgelegt von
Wohlhaupter, Peter
am 7. März 2012

Gutachter

Prof. Dr. Manfred Reichert
Dr. Edgar Schiebel

Acknowledgement

I would like to thank Dr. Edgar Schiebel from the Austrian Institute of Technology for his permission to use the BibTechMon software in this work, for his valuable input during the making of this thesis and for his willingness to be one of the two supervisors of my thesis. I would also like to thank Rüdiger Pryss, as the mentor of this work, for his always encouraging and helpful support. Last but not least, I would like to thank Professor Manfred Reichert for his willingness to be one of the supervisors of my thesis and for giving me a very insightful interview on the field of business process management.

Table of Contents

Acknowledgement i
List of Figures iv
List of Tables v
List of Abbreviations vii

1 Introduction 1
1.1 Business Process Management as an Important Field in Theory and Practice 1
1.2 Aims of this Work 2
1.3 Structure of this Work 3
2 Business Process Management 5
2.1 History 5
2.2 Definitions 6
2.3 Topics within Business Process Management 7
2.3.1 BPMN, BPEL 7
2.3.2 Data-driven Workflows 7
2.3.3 Metrics 8
2.3.4 Compliance 8
2.3.5 Mobile Processes 9
2.4 Topics related to Business Process Management 9
2.4.1 Business Intelligence 9
2.4.2 ERP 10
2.4.3 Knowledge Management 11
2.4.4 SOA 11
3 Bibliometrics 13
3.1 History 13
3.2 Definitions 14
3.3 Bibliometric Methods 14
3.3.1 One-dimensional Methods 15
3.3.2 Indexes 15
3.3.3 Two-dimensional or Relational Methods 16
3.4 Bibliometric Terms 17
3.5 BibTechMon Software 17
3.6 Scientific Databases 24
4 Work with Google Scholar 26
4.1 Steps of the Analysis 26
4.2 Search in Google Scholar 27
4.3 Options in Google Scholar 28
4.4 Possible Networks from the Google Scholar Data 29
4.4.1 Author Network 29
4.4.2 Cited by Network 30
4.4.3 Inverted Cited by Network 30
4.4.4 Co-Citation Network 31
4.4.5 BibCoup Network 31

4.5 Comparison of Networks 32
4.6 Time Constraints when Working with Forward Citations 32
4.7 Metrics.. 32
5 Bibliometric Analysis 35
5.1 Analysis of BPM Data in General 37
5.1.1 BPM Search Term 38
5.1.2 Workflow Search Term 44
5.1.3 Combined Business Process/Workflow Management Search Term 48
5.1.4 Author Network 53
5.2 Analysis of Specific Fields of BPM 55
5.2.1 BPMN and BPEL 56
5.2.2 Data-driven Workflows 61
5.2.3 Metrics 66
5.2.4 Compliance 71
5.2.5 Mobile Processes 75
5.3 Analysis of Fields Related to BPM 81
5.3.1 Business Intelligence 82
5.3.2 ERP 86
5.3.3 Knowledge Management 89
5.3.4 SOA 93
5.4 Analysis of one Journal and two Conferences 97
5.4.1 Data & Knowledge Engineering Journal 98
5.4.2 International Conference on Business Process Management 103
5.4.3 CAiSE 107
5.4.4 CAiSE Inverted Cited by Network 112
6 Conclusions 114
6.1 Interview with Professor Reichert 114
6.2 Results of the Bibliometric Analysis 114
6.3 Comparison with the Results from the Interview 115
6.4 Further Analysis of the Results 115
6.5 Future Prospects of BPM and Bibliometrics 116

Appendix 117
I. Source Code of the Google Script 117
II. Changes of the Script after Google Scholar Altered its Format 127
III. Functionalities of the Script and Further Notes to the Usage of the Script 129
IV. Transformation of the CSV File into a MDB File 130
V. Complete Text of the Interview with Professor Reichert 131

References 141

Ehrenwörtliche Erklärung 147

List of Figures

Figure 1: Illustration of the structure of this work 4
Figure 2: Extract Keywords screen from BibTechMon 18
Figure 3: Random distribution of elements in BibTechMon before the iteration 19
Figure 4: Screenshot of the options possible to regulate related to the iteration in BibTechMon 20
Figure 5: Distribution of elements after the iteration is complete 21
Figure 6: Density map placed over the distribution of elements 22
Figure 7: Names of marked elements from the network in BibTechMon 23
Figure 8: Articles containing selected elements in BibTechMon 23
Figure 9: Advanced search options in Google Scholar. 27
Figure 10: Cited by network of the business process management search term 38
Figure 11: Cited by network of the workflow management search term 45
Figure 12: Cited by network of the combined business process/workflow management search term 48
Figure 13: Author network on basis of the BPM search term 53
Figure 14: Cited by network of the BMPN/BPEL search term 56
Figure 15: Cited by network of the data-driven workflows search term 61
Figure 16: Cited by network of the business process metric search term 66
Figure 17: Cited by network of metrics, different elasticity threshold 67
Figure 18: Cited by network of the compliance search term 71
Figure 19: Cited by network of the mobile processes search term 75
Figure 20: Cited by network of the business intelligence search term 82
Figure 21: Cited by network of the ERP search term. 86
Figure 22: Cited by network of the knowledge management/BPM search term 89
Figure 23: Cited by network of the SOA search term 93
Figure 24: Cited by network of the Data & Knowledge Engineering Journal search term 98
Figure 25: Cited by network of the BPM Conference search term 103
Figure 26: Cited by network of the CAiSE search term 107
Figure 27: Inverted cited by network of the CAiSE search term 112

List of Tables

Table 1: Comparison of the different networks 32
Table 2: Numbers of the business process management search and of the corresponding network.......... 39
Table 3: Clusters in the cited by network for the business process management search term 39
Table 4: Numbers of the workflow management search and of the corresponding network 45
Table 5: Clusters in the cited by network for the workflow management search term 45
Table 6: Numbers of the combined business process/workflow management search and of the corresponding network.......... 49
Table 7: Clusters in the cited by network for the combined business process/workflow management search term.......... 49
Table 8: Numbers of publications.......... 54
Table 9: Numbers of the BPMN/BPEL search and of the corresponding network 57
Table 10: Clusters in the cited by network for the BPMN/BPEL search term 57
Table 11: Numbers of the data-driven search and of the corresponding network 62
Table 12: Clusters in the cited by network for the data-driven workflows search term 62
Table 13: Numbers of the metrics search and of the corresponding network 67
Table 14: Clusters in the cited by network for the business process metrics search term 68
Table 15: Numbers of the metrics search and of the corresponding network 72
Table 16: Clusters in the cited by network for the compliance search term.......... 72
Table 17: Numbers of the metrics search and of the corresponding network 76
Table 18: Clusters in the cited by network for the mobile processes search term...... 76
Table 19: Numbers of the business intelligence search and of the corresponding network 83
Table 20: Clusters in the cited by network for the business intelligence search term 83
Table 21: Numbers of the ERP search and of the corresponding network.......... 87
Table 22: Clusters in the cited by network for the ERP search term.......... 87
Table 23: Numbers of the knowledge management search and of the corresponding network 90
Table 24: Clusters in the cited by network for the knowledge management search term 90
Table 25: Numbers of the SOA search and of the corresponding network 94
Table 26: Clusters in the cited by network for the SOA search term 94
Table 27: Numbers of the Data & Knowledge Engineering Journal search and of the corresponding network.......... 99
Table 28: Clusters in the cited by network for Data & Knowledge Engineering Journal search term 99
Table 29: Numbers of the BPM Conference search and of the corresponding network 104
Table 30: Clusters in the cited by network for the BPM Conference search term ... 104
Table 31: Numbers of the CAiSE search and of the corresponding network........... 108

Table 32: Clusters in the cited by network for the CAiSE search term 108
Table 33: Clusters in the CAiSE inverted cited by network 113

List of Abbreviations

ACM	Association for Computing Machinery
BAM	Business Activity Monitoring
BI	Business Intelligence
BPEL	Business Process Execution Language
BPM	Business Process Management
BPMS	Business Process Management System
BPMN	Business Process Modeling Notation
CoopIS	International Conference on Cooperative Information Systems
CRM	Customer Relationship Management
CSV	Comma-Separated Values
DBIS	Institute of Databases and Information Systems
ebXML	Electronic Business using eXtensible Markup Language
EPC	Event-driven process chain
ERP	Enterprise Resource Planning
ICWS	International Conference on Web Services
IEEE	Institute of Electrical and Electronics Engineers
IT	Information Technology
MIS	Management Information Systems
PAIS	Process-Aware Information System
PKM	Process-oriented Knowledge Management
SCC	International Conference on Services Computing
QoS	Quality of Service
SCM	Supply Chain Management
SME	Small and Medium Enterprises
SOA	Service-Oriented Architecture
SOAP	Simple Object Access Protocol
SOC	Service-Oriented Computing
UML	Unified Modeling Language
WFM	Workflow Management
WfMC	Workflow Management Coalition
WFMS	Workflow Management System
WS-CDL	Web Services Choreography Description Language
WSDL	Web Service Description Language
WSFL	Web Service Flow Language
XML	eXtensible Markup Language

1 Introduction

1.1 Business Process Management as an Important Field in Theory and Practice

The work life of today is pretty much unthinkable without division of labor. In every business, labor is divided in different processes, which in turn consist of different activities. With successful management of those business processes, one can increase efficiency, lower costs and expand flexibility of the enterprise. In business process management, or BPM for short, business processes can be modeled and viewed from two main perspectives: From the perspective of, for example, a manager without a lot of technical knowledge, and from a more technical, more IT-oriented perspective which is then needed, when BPM systems shall be used. These business process management systems support enterprises with every step related to the business processes. Typically, they include components to model, edit, run and change processes, to evaluate and monitor processes and also to integrate the processes with the organizational structure of the enterprise.

Business process management is a popular topic both in business practice and in science. The importance in business practice can be seen in the growing revenues in the BPM market. The market researcher IDC predicts for 2013 a market volume for BPM software of 3 billion US dollars in comparison to a market volume of 1.7 billion US dollars in 2009.[1] Big companies such as IBM invest heavily in the BPM market.[2] The scientific importance follows from the big and growing number of publications in the topic of BPM. While only 422 publications can be found in Google Scholar for the term “business process management” that have been published in the year 2000, already 2570 articles were published with the same topic in 2005 and in 2010 the number of BPM publications from that year has risen to 5810.[3]

There are several relevant subtopics of BPM, for example data-driven workflows, process compliance or process metrics. As well, there are several relevant topics that are related to BPM, since they also are connected to business processes. Among these are topics like business intelligence, enterprise resource planning and service-oriented architecture. Since there are so many topics and even more publications, it is difficult to keep track on all the new developments in BPM. This makes it useful to use bibliometric methods. With bibliometrics one can, in general, analyze scientific publications quantitatively. With advanced bibliometric methods one can also analyze networks among researchers or try to detect thematic clusters in scientific fields.

---

[1] IDC: Worldwide Business Process Management Software 2009-2013 Forecast, as cited by IBM: IBM to Acquire Lombardi, URL: http://www-03.ibm.com/press/us/en/pressrelease/28890.wss, accessed January 15, 2012

[2] See for example the purchase of Lombardi by IBM, ZDNet: IBM kauft Prozessmanagement-Software-Anbieter Lombardi, URL: http://www.zdnet.de/news/41524612/ibm-kauft-prozessmanagement-software-anbieter-lombardi.htm, accessed February 2, 2012

[3] All results from Google Scholar, URL: http://scholar.google.com, accessed December 15, 2011

In this work, I will use the bibliometric software BibTechMon[4] of the Austrian Institute of Technology to perform different bibliometric analyses in the field of business process management. For the first time, the free database Google Scholar will be used in a BibTechMon analysis as the source of the scientific articles instead of paid sources like Web of Science or Scopus. In the course of this work I have written a script to extract the data from Google Scholar pages and to transform that data into a format that can be used with the BibTechMon software. Also, the work with Google Scholar data somewhat changes the way of how the bibliometric analysis works in comparison to using data from Web of Science or Scopus. I want to point out the new possibilities but also the restrictions that come along when working with Google Scholar data.

Based on these experiences, I will then acquire data sets from Google Scholar about several different topics in business process management or related to it and I will analyze this data in order to discover thematic clusters in those fields. At the same time, I want to get information about the positioning of one particular institute in the field of BPM. This institute is the Institute of Databases and Information Systems, or DBIS, which is part of the Faculty of Engineering and Computer Science at University of Ulm and very active in the field of BPM. For this purpose, I will interview Professor Reichert of DBIS about his opinions about the BPM field in general and the positioning of DBIS in particular and then I will compare his statements to my findings in BibTechMon.

1.2 Aims of this Work

This work wants to apply bibliometric methods on the field of business process management, something that has – to the knowledge of the author – so far only be done once, in relation to collaboration networks of authors in the BPM conferences[5], but never on the BPM field as a whole.

The first goal of this work is, to make it possible to use Google Scholar data in the bibliometric software BibTechMon. For this reason I have written a Perl script, which I will call Google script, that facilitates the extraction of data from Google Scholar and generates CSV files that in turn can be converted to files usable in BibTechMon.

The next goal is it to compare Google Scholar to other scientific databases such as Web of Science and Scopus and to discuss the features of Google Scholar data as well as the implications this data has on the work in BibTechMon and bibliometrics in general. The most pressing topics will be the differences between forward citations and backward citations, as well as the dimension of time.

The third goal is to use the Google script to extract Google Scholar data for various search terms related to business process management and to create different network graphs in BibTechMon based on these results. Then I want to use those network graphs to identify research clusters in the various fields of business process management.

[4] See Noll, Fröhlich, Schiebel (2002): Knowledge Maps of Knowledge Management Tools – Information Visualization with BibTechMon

[5] Reijers et al. (2009): A Collaboration and Productiveness Analysis of the BPM Community

As a fourth goal, I will look into the positions that the DBIS institute has in certain BPM networks. In addition to that, an interview with Professor Reichert will be conducted and I will compare the statements of Professor Reichert with the results found with the bibliometric analyses.

1.3 Structure of this Work

This work starts with this introduction chapter which contains three subchapters. The first subchapter contains a motivation for this work and tries to show the importance of the field of business process management and why it is necessary to analyze it with bibliometric methods. The second subchapter states the concrete goals of this work. Then, the current subchapter with the structure of this work follows.

Chapter 2 sums up the history of business process management, gives definitions for the term and gives definitions for topics both within business process management and related to business process management.

In Chapter 3, first the history of bibliometrics is recapitulated. After that, definitions for the term bibliometrics and related terms are given, followed by a subchapter on bibliometric methods and a subchapter on the definition of further bibliometric terms. This is followed by a description of the functionality of the BibTechMon software that I will use for the bibliometric analyses in the course of this work.

Chapter 4 goes into the specifics of how the analysis on basis of Google Scholar data will be performed. First it will describe what steps need to be taken to perform the bibliometric analyses. Then it will describe how the search in Google Scholar works and what options can be chosen. After that, I will show which networks can be created in BibTechMon on basis of the data from Google Scholar. This is followed by a comparison of the networks that are possible. I will then look into possible time restrictions when performing bibliometric analyses and in the last subchapter I present a quick outline of how the quality of a search time might be assessed.

Chapter 5 is the main part of this work where I will perform several bibliometric analyses with BibTechMon on basis of data from Google Scholar. The analyses are divided into four groups: First, analyses of the BPM field in general will be performed. Second, specific subtopics of BPM will be analyzed. Third, topics related will be analyzed and last I will analyze articles related to one publication and two conferences from the BPM field.

In Chapter 6, the conclusion of the analyses will be presented. I will sum up the core results of the interview that I conducted with Professor Reichert and the results of the bibliometric analyses in Chapter 5. Thereafter, I will compare the statements from Professor Reichert with the results of the analyses. Then I will try to evaluate the results of this work and give an outlook for both business process management and bibliometrics.

The following graphic illustrates the structure of this work. This work starts with an introduction to the topic, followed by three chapters that form the basis of the bibliometric analysis, which is the core of this work. The work ends with the conclusion.

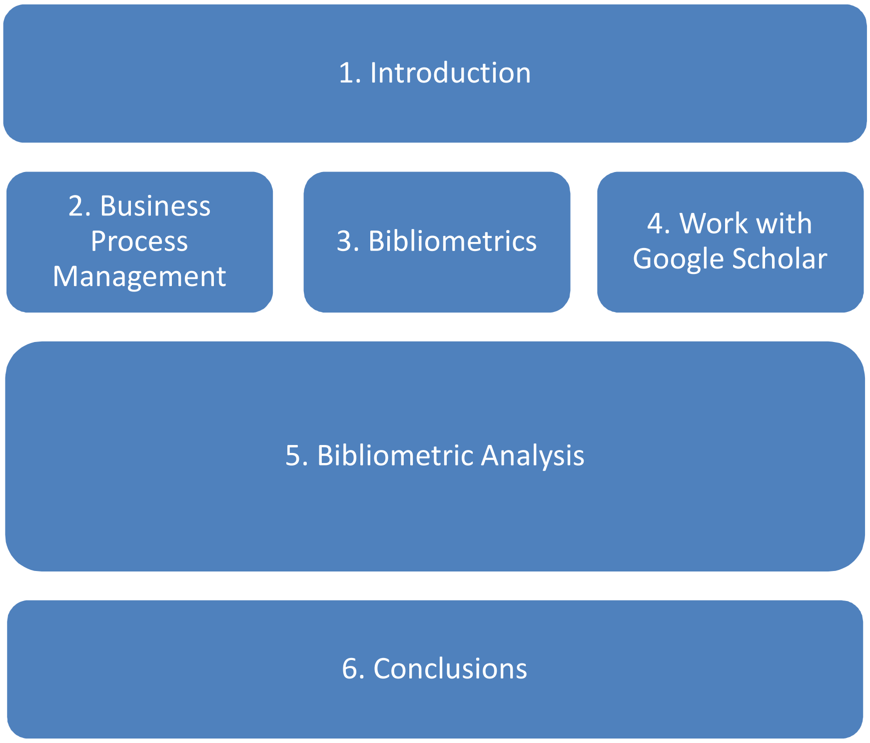


Figure 1: Illustration of the structure of this work

2 Business Process Management

2.1 History

The following chapter is mostly based on Mendling (2008).[6] For additional bibliographic data the Encyclopedia Britannica[7] has been used.
The first notable scientist that has considered business process management in an early form is the Scottish economist Adam Smith (1723-1790). Adam Smith saw the potential of subdivision of labor, which is a precondition for business processes, and he described that potential at the example of the production of pins. This idea was then picked up by the French mining engineer Henri Fayol (1841-1925). Fayol realized that a subdivision of labor can lead to increased productivity.
Following on Fayol, the American engineer Frederick Taylor (1856-1915) became the next important figure in the early history of processes-related thoughts. His "Taylorism" focused on the optimization of process steps.
The next important figure was Henry Ford (1863-1947) who greatly popularized the assembly line idea, thus optimizing the processes of his business, the Ford Company.
The first author in the field of business organization, who then proposed a distinction between structural organization and process organization, was Fritz Nordsieck (1906-1984).
After World War 2 discussions about the automation of office work began. In the early 1970s the first information systems have been created. Their focus was mainly on structural aspects, however, and not on process aspects. In the late 70s the idea of flow control has been introduced to office automation. After that, Michael Zisman was the first to introduce petri nets – a notation originally invented by Carl Petri to describe chemical processes – in order to model business processes in 1977. Similarly, Skip Ellis used information control nets to do the same.
In the early 1990s workflow management was presented as a new technology. At roughly the same time, new business administration concepts such as process innovation and business process redesign have been introduced. As well, in the 1990s the application of workflow systems has become more widespread.
This was followed by an increase in scientific publications on workflow technology.
On the technical side languages for the execution and choreography of processes, like BPEL[8], WS-CDL[9] and ebXML[10] have been created.
What we can conclude from this short outline about the history of business process management is that the origins of business process management lie in economics and management science, but with the development of computer science and information technology, it is a topic that has become more and more dominated by IT-related

[6] Mendling (2008): Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness, p. 2-4
[7] Encyclopedia Britannica, URL: http://www.britannica.com, accessed December 12, 2011
[8] See OASIS: OASIS Web Services Business Process Execution Language (WSBPEL) TC, URL: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsbpel, accessed February 25, 2012
[9] See W3C: Web Services Choreography Description Language Version 1.0, http://www.w3.org/TR/2004/WD-ws-cdl-10-20041217/, accessed February 25, 2012
[10] See OASIS: About ebXML, URL: http://www.ebxml.org/geninfo.htm, accessed February 25, 2012

topics. This work will focus on the IT aspects of BPM, but it will also take the management-related aspects into account.

2.2 Definitions

The following chapter is mostly based on van der Aalst, ter Hofstede and Weske (2003): Business Process Management: A Survey, one of the fundamental articles on business process management.[11]

The basis of business process management is business processes. Starting with a definition of the term process, business processes can be defined as follows:

"A process is a completely closed, timely and logical sequence of activities which are required to work on a process-oriented business object. Such a process-oriented object can be, for example, an invoice, a purchase order or a specimen. A business process is a special process that is directed by the business objectives of a company and by the business environment. Essential features of a business process are interfaces to the business partners of the company (e.g. customers, suppliers)."[12]

A first description of BPM is given by van der Aalst et al in the following way: "Business Process Management (BPM) includes methods, techniques, and tools to support the design, enactment, management, and analysis of operational business processes. It can be considered as an extension of classical Workflow Management (WFM) systems and approaches."

In order to understand the reference to workflow management I will have a look at the definition of workflow and workflow management systems.

The Workflow Handbook defines workflow as follows: "The automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules."[13] Based on that the Workflow Handbook defines a workflow management system: "A system that defines, creates and manages the execution of workflows through the use of software, running on one or more workflow engines, which is able to interpret the process definition, interact with workflow participants and, where required, invoke the use of IT tools and applications"[14].

Later van der Aalst, ter Hofstede and Weske go on to define BPM as: "Supporting business processes using methods, techniques, and software to design, enact, control, and analyze operational processes involving humans, organizations, applications, documents and other sources of information."

They then define a business process management system as: "A generic software system that is driven by explicit process designs to enact and manage operational business processes".

[11] The article is cited 710 times in Google Scholar (excluding patents), URL: http://scholar.google.com, accessed December 13, 2011

[12] Becker, Kahn (2003): The Process in Focus

[13] Lawrence (Editor) (1996): Workflow Handbook 1997

[14] Lawrence (Editor) (1996): Workflow Handbook 1997

2.3 Topics within Business Process Management

In this chapter, I will define important terms within the field of business process management.

2.3.1 BPMN, BPEL

BPMN and BPEL are a notation and a language to describe business processes. BPMN is targeted at analysts of processes that not necessarily have a lot of IT knowledge, while BPEL is targeted at IT developers.[15]
BPMN stands for "business process modeling notation". It is a graphical notation to model business processes within one organization or between several organizations. It has been designed by the Object Management Group and is currently available in its second version.[16] It is a notation widely accepted in the industry and is supposed to replace the Event-driven Process Chains or EPCs.
BPEL on the other hand stands for "business process execution language"[17]. It is the result of the combination of two process execution languages, XLANG developed by Microsoft and the Web Service Flow Language, or WSFL, by IBM. The two different approaches in the modeling of the processes, a block-based approach from XLANG and a graph based approach from WSFL, have both been incorporated into BPEL. While BPMN is targeted at business users, BPEL is mostly used for the modeling of the technical side of the process. BPEL can be run in process engines such as Intalio, Apache ODE, Microsoft Biztalk or IBM Websphere.
Due to the different natures and fields of application of BPMN and BPEL there have also been several publications about transforming BPMN to BPEL and vice versa.

2.3.2 Data-driven Workflows

Workflow structures are typically categorized into two major groups: Control-driven workflows and data-driven workflows. While in control-driven workflows the focus lies on sequences, conditions and iterations, data-driven workflows focus on the product or the data on which the processes are centered. Or in other words: In data-driven processes "the product structure defines the sequence of process executions".[18]
Related terms are product-based workflows, object-aware/object-centric workflows as well as artifact-based workflows.
In artifact-based workflows, the processes are centered on "business artifacts" which are supposed to represent "key business entities".[19] The terms object-aware and object-centric are used in a similar sense as the data-driven in data-driven workflows.

[15] Wohed et al. (2006): On the Suitability of BPMN for Business Process Modelling
[16] Object Management Group: Business Process Model and Notation, URL: http://www.bpmn.org, accessed February 20, 2012
[17] See OASIS: OASIS Web Services Business Process Execution Language (WSBPEL) TC, URL: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsbpel, accessed February 25, 2012
[18] Müller, Reichert, Herbst (2006): Flexibility of Data-driven Process Structures
[19] Fritz, Hull, Su (2009): Automatic construction of simple artifact-based business processes

In this work, I will use the term data-driven workflow to encompass the whole field related to these terms. Examples for data-driven workflow can be found in development processes or production processes, where each sub-component of the product has several processes related to it.[20] Another example can be an application process for a job vacancy that is centered on the applications sent by the applicants.[21]

2.3.3 Metrics

According to the Merriam-Webster dictionary, a metric is "a standard of measurement".[22] In software engineering, it has proven to be useful, to use metrics to measure the understandability and the quality of programming practices and software design. Along these lines, there are considerations in the field of business process management, to use metrics, as well. There are mainly two groups of metrics used in BPM: quality metrics and similarity metrics. The quality metrics intend to measure such qualities of business process models as whether their size is appropriate, whether the process models are easy to understand and whether they are clearly structured.[23] The similarity metrics are used to measure whether two given process models are similar to each other. This can be useful when there are already large process repositories and new processes should be added or when the processes of merging companies are analyzed in order to see, whether processes are similar to each other and where they have to be changed. [24]

2.3.4 Compliance

Compliance is "the act or process of complying to a desire, demand, proposal, or regimen or to coercion"[25], whereas to comply is defined as "to conform, submit, or adapt (as to a regulation or to another's wishes) as required or requested".
In business life, it is necessary for companies to comply with different sorts of regulations. These standards can relate for example to quality standards or internal controls.[26]
It is an import part of business process management, to ensure that the business processes used comply with such standards. When checking the compliance of business processes, there are two main categories:

- Forward compliance checking, i.e. the attempt to assure compliance when designing a process, before the process will be performed

[20] Müller, Reichert, Herbst (2006): Flexibility of Data-driven Process Structures
[21] Künzle, Reichert (2009): Towards Object-aware Process Management Systems: Issues, Challenges, Benefits
[22] Merriam-Webster: Definition of metric, URL: http://www.merriam-webster.com/dictionary/metric, accessed February 17, 2012
[23] Vanderfeesten et al. (2009): Quality Metrics for Business Process Models
[24] Dijkman et al. (2011): Similarity of Business Process Models: Metrics and Evaluation
[25] Merriam-Webster, Definition of compliance, URL: http://www.merriam-webster.com/dictionary/compliance, accessed February 17, 2012
[26] El Kharbili et al. (2008): Business Process Compliance Checking: Current State and Future Challenges

- Backward compliance checking, i.e. checking whether a process is compliant after it has been performed

Since compliance is getting more and more important in basically all industries[27], a growing importance of compliance within business process management is to be expected as well.

2.3.5 Mobile Processes

In many fields, such as health care, logistics and sales,[28] it is necessary to include mobile users into processes. Examples can be an absent manager that still has to take part in business decisions[29] or a chronically ill patient that needs assistance. In both cases, mobile devices can be used to assist the users and this assistance will usually occur in a process-oriented context. So far, no comprehensive systems in the field of mobile processes exist. Several requirements have been identified by Pryss et al. when it comes to running mobile processes.[30] These requirements can be split into three different categories:

- Process implementation requirements, this includes for example the partitioning of processes
- Supporting infrastructure requirements, e.g. the handling of broken connections
- Runtime requirements, for example, the synchronization of the process on different devices

With mobile devices becoming more and more a part of our everyday life, it is to be expected that the management of mobile processes becomes more widespread, as well.

2.4 Topics related to Business Process Management

In this chapter, I will define important topics related to the field of business process management.

2.4.1 Business Intelligence

The term "business intelligence" was coined by Hans Peter Luhn at IBM in 1958 in his article "A Business Intelligence System". Herein, he describes business intelligence as "[t]he ability to apprehend the interrelationships of presented facts in

[27] Lu, Sadiq, Governatori (2008): Compliance Aware Business Process Design
[28] Pryss et al. (2011): Towards Flexible Process Support on Mobile Devices
[29] Pousttchi, Thurnher (2006): Usage of mobile technologies to support business processes
[30] Pryss et al. (2011): Towards Flexible Process Support on Mobile Devices

such a way as to guide action towards a desired goal".[31] A slightly newer definition stems from Negash and Grey. They define business intelligence as "a data-driven DSS [decision support system] that combines data gathering, data storage, and knowledge management with analysis to provide input to the decision process".[32]
Depending on whether broader or narrower definitions are used[33] it encompasses functions such as the following:

- Data/Text Mining
- Data Warehousing
- Online Analytical Processing
- Knowledge Management

There is a significant connection between business intelligence, or shortened BI, and BPM: When using BPM systems, a lot of data about processes will be provided by these systems. This data can be analyzed with business intelligence methods. The "application of business intelligence techniques to business processes"[34] is then called business process intelligence.

2.4.2 ERP

Jarrar, Al-Mudimigh and Zairi define enterprise resource planning systems or ERP systems as "comprehensive package software solutions that seek to integrate the complete range of business's processes and functions in order to present a holistic view of the business from a single information and IT architecture".[35]
What distinguishes ERP systems from former stand-alone business information systems, is that ERP systems try to integrate the complete business process into one system. The triumph of ERP began in the 1990s and 2009 the market for ERP software comprised more than 20 billion dollars a year.[36] What is typical for ERP systems is that their size and complexity requires careful planning and implementation. The critical success factors of implementing ERP systems are "top management support, a clear business vision, and issues specific to ERP such as ERP strategy and software configuration"[37]. However, processes are also relevant, what can be seen in the following quote: "some of the more important factors are the issues related to re-engineering business processes and the integration of various core

---

[31] Luhn (1958): A Business Intelligence System
[32] Negash, Gray (2008): Business Intelligence
[33] More information about definitions of business intelligence can be found in: Gluchowski (2011): Business Intelligence - Konzepte, Technologien und Einsatzbereiche and Golfarelli, Rizzi, Cella (2004): Beyond data warehousing: what's next in business intelligence?
[34] Grigori et al. (2004): Business Process Intelligence
[35] Jarrar, Al-Mudimigh, Zairi (2000): ERP implementation critical success factors-the role and impact of business process management
[36] Gartner (2010): Magic Quadrant for ERP for Product-Centric Midmarket Companies
[37] Jarrar, Al-Mudimigh, Zairi (2000): ERP implementation critical success factors-the role and impact of business process management

processes to the ERP system"[38]. Thus, we can see that processes play a vital role in ERP systems, thereby relating ERP to BPM.

2.4.3 Knowledge Management

Knowledge is defined by Davenport and Prusak as follows: "Knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information."[39]

The management of said knowledge is an important part of every business. For this purpose knowledge management systems have been developed. Their objective is "to support construction, sharing and application of knowledge in organizations."[40] As a more recent development, the notion of Process-oriented Knowledge Management, also called PKM, has been introduced. PKM thrives to integrate knowledge management and business process management. Topics in this field are the integration of the knowledge lifecycle with the process lifecycle and process knowledge. Process knowledge can be divided into three different groups: Process template knowledge, which is the knowledge about the process models, process instance knowledge, which is the knowledge gathered during the execution of the process, and process-related knowledge, which is the knowledge process-activity performers can use during the process. The core difference of the process-related knowledge approach to the normal knowledge approach is that in a PKM the knowledge will be presented at the right time and the right place.[41]

2.4.4 SOA

Service-oriented architecture or SOA is an architectural paradigm for computer systems which supports the thinking in processes. It requires an alignment of IT processes to the business processes. It also requires a unified IT infrastructure and an enterprise service bus. Additionally, in order to build a working SOA, one needs to follow certain principles when creating the services. Commonly eight principles are mentioned[42] and they include:

- Loose coupling: Reducing dependencies between services
- Reusability: Services can be re-used, possibly in different contexts
- Standardized contracts: Services adhere to standardized service contracts
- Composability: Larger services can be constructed by using other services

---

[38] Jarrar, Al-Mudimigh, Zairi (2000): ERP implementation critical success factors-the role and impact of business process management

[39] Davenport, Prusak (1998): Working Knowledge– How organizations manage what they know

[40] Alavi, Leidner (2001): Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues

[41] Jung, Choi, Song (2007): An integration architecture for knowledge management systems and business process management systems

[42] Erl: The Service-Orientation Design Paradigm, URL: http://www.soaprinciples.com/p3.php, accessed February 27, 2012

Another important component is the process-orientation of SOA, and this is what relates it to BPM. The services must be aligned to the processes in the company. Furthermore, the processes should be supported by an IT system and the orchestration of services – i.e. the creation of a larger service making use of smaller sub-services – can be done using process engines. In SOA, the business process execution language BPEL mentioned before is also commonly used. Other relevant standards include WSDL and SOAP.[43]

[43] Pasley (2005): How BPEL and SOA are changing Web services development

3 Bibliometrics

In this chapter, I will first present a short history of bibliometrics and then show different definitions of the term. After that, I will have a look at the different kinds of bibliometric methods and define some additional bibliometric terms. Then, I will give a short overview over the functionalities of the BibTechMon software that I will use in the course of this work. In the last chapter, I will present several scientific databases. From these databases, later the Google Scholar database will be used.

3.1 History

The following subchapter about the history of bibliometrics is based on Hood/Wilson (2001): The literature of bibliometrics, scientometrics and informetrics.

The first early version of bibliometrics can be found in Hebrew literature from about the 12th century which used citation indexes for the first time. Later citation indexes can be found in legal literature from 1743. Publication counts have been used at least since 1817. The possibly first bibliometric study has been published in 1896 by Campbell[44]. Campbell used statistical methods to analyze how subjects are scattered in publications. Other early works include a bibliometric study on anatomy literature by Cole and Eales from 1917.[45]

The term "bibliometrics" itself was probably first used – in his French equivalent "bibliometrie" – by Otlet in 1934 in his work "Traitée de Documentation. Le livre sur le Livre. Théorie et Pratique". However, usually Pritchard is attributed with coining the term in his publication from 1969. He suggested replacing the term "statistical bibliography" with the term "bibliometrics".[46]

Several alternative and related terms have been proposed, but only two got at least some recognition: The same year as Pritchard's coining of the term "bibliometric", Nalimov and Mulchenko proposed the term "scientometrics".[47] Scientometrics is supposed to "study all aspects of the literature of science and technology".[48] Scientometrics gained some recognition by the founding of a journal with the same name. However, much of scientometrics and bibliometrics overlap and much of bibliometric research is also published in the Scientometrics journal.

Another related term is "Informetrie" or "informetrics", a term that has been proposed by Nacke in 1979.[49] It covers "the measurement of information phenomena"[50] and is supposed to encompass both bibliometrics and scientometrics.

---

[44] Campbell (1896): The Theory of the National and International Bibliography: with Special Reference to the Introduction of System in the Record of Modern Literature

[45] Cole, Eales (1917): The history of comparative anatomy. Part I: A statistical analysis of the literature

[46] Pritchard (1969): Statistical bibliography or bibliometrics?

[47] Nalimov, Mulchenko (1969): Scientometrics. Study of the Development of Science as an Information Process (English translation of the title), as cited in Hood, Wilson (2001): The literature of bibliometrics, scientometrics, and informetrics

[48] Hood, Wilson (2001): The literature of bibliometrics, scientometrics, and informetrics

[49] Nacke (1979): Informetrie: Ein neuer Name für eine neue Disziplin

[50] Hood, Wilson (2001): The literature of bibliometrics, scientometrics, and informetrics

An analysis of the distribution of these and some other related term performed by Hood/Wilson in the Dialog database[51], show that "bibliometrics" is by far the most commonly used term,[52] and because of that, it is the term used in this work.

3.2 Definitions

There are several definitions of the term bibliometrics.
Pritchard, who coined the term, defined the goal of bibliometrics as: "to shed light on the processes of written communication and of the nature and course of development of a discipline (in so far as this is displayed through written communication), by means of counting and analyzing the various facets of written communication […] the application of mathematics and statistical methods to books and other media of communication"[53].
Another slightly broader definition is given by Fairthorne, who defines bibliometrics as "quantitative treatment of the properties of recorded discourse and behavior appertaining to it"[54].
Based on a review of earlier definitions Broadus gives the following definition: According to him, bibliometrics is "the quantitative study of physical published units, or of bibliographic units, or of surrogates of either"[55].
This is the definition I will use in this work.

3.3 Bibliometric Methods

Bibliometric methods are methods to measure properties of publications. Bibliometric methods can be divided into three groups:
The first group contains the one-dimensional methods. The one-dimensional methods are about counting occurrences of certain elements, for example the count of citations of an article.
The second group consists of the so-called indexes that have been created on basis of the counts of such occurrences. These indexes use certain formulae in order to integrate several counts into one single number.
The third group contains of the two-dimensional or relational methods, where the co-occurrence of elements is used as the basis of analysis.

[51] An information service, which can be found on http://www.dialog.com
[52] In their search the term "bibliometrics" was found 5097 times, the term "scientometrics" 1326 times and the term "informetrics" 418 times.
[53] Pritchard (1969): Statistical bibliography or bibliometrics?
[54] Fairthorne (1969): Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction.
[55] Broadus (1987): Toward a definition of 'bibliometrics'

3.3.1 One-dimensional Methods

The one dimensional methods are about counts of occurrences. They include:[56]

- Publication counts: One of the most simply measures is to count the number of publications a certain author has published.
- Weighted publication counts: A slightly more sophisticated measure is to weigh the publications with a value related to the importance of the journal where the publication has been published in.
- Citation counts: To count how many times a certain publication has been cited by other publications.

3.3.2 Indexes

Based on the citation counts of the publications of an author, two indexes have been established that are commonly used, in order to evaluate the importance of an author in a certain field, these are the h-index and the g-index. In order to measure the importance of scientific journals, the Journal Impact Factor is used.

h-index

A commonly used index to measure the importance of a researcher in a specific field, is the h-index or Hirsch-index. The h-index is defined by its creator, Jorge Hirsch, as follows:

"A scientist has index h if h of his or her N_p papers have at least h citations each and the other (N_p – h) papers have <=h citations each."[57]

So, if a scientist has in total published five publications, of which three have been cited at least three times each, and the other two publications have been cited at most three times each, then this scientist has an h-index of three.[58]

g-index

In 2006, after the h-index has gained popularity, an improvement of the h-index has been proposed by Egghe. This improved index is called g-index. Based on a set of articles, it is defined as follows:

"If this set is ranked in decreasing order of the number of citations that they received, the g-index is the (unique) largest number such that the top g articles received (together) at least g^2 citations."[59]

Given a scientist that has published five articles that have been cited 7, 6, 6, 3 and 1 time, respectively, he receives a g-index of 4.

It should be noted, that neither of these indexes are suitable to compare authors from different scientific fields. This is caused by the fact that in different disciplines the way of publishing and citing differs greatly.

[56] Meyer et al. (2009): Research Evaluation for Computer Science
[57] Hirsch (2005): An index to quantify an individual's scientific research output
[58] Further examples can be found in Robecke (2011): Development of an iPhone business application
[59] Egghe (2006): Theory and practise of the g-index

Journal Impact Factor

While the h-index and the g-index are used for single authors or groups of authors, there is also an index to evaluate the impact of journals. That index is the Journal Impact Factor and it can be defined as follows[60]:

A = citations from the given year to articles published in the two years before the given year
B = number of articles published in the two years before the given year
Journal Impact Factor for the given year = A/B

For example, a journal that published 60 articles in the years 2009 and 2010 and received 200 citations to these articles in the year 2011, has an Journal Impact Factor for the year 2011 of 200/60=3.33.

3.3.3 Two-dimensional or Relational Methods

The following subchapter is mostly based on van Raan and Tijssen (1993): The neural net of neural network research: An exercise in bibliometric mapping.

While one-dimensional methods work on counts or simple occurrences of elements, such as publications or citations, two-dimensional or relational methods work on the co-occurrence, i.e. the occurrence at the same time of different elements.

Every publication contains certain elements, such as the authors, the text of the article and keywords. Some of these elements can consist of a list of entries, for example a list of authors, when the article is written by more than one author, or a list of keywords, since commonly publications are described by more than one keyword, or a list of citations, with the list of articles the publication cites. For each value of such elements one can count how often they occur together with the other values of those elements (this is the so-called co-occurrence). For example if there is an article published together by Miller and Meyer, then this means the author names Miller and Meyer co-occur at least once. These co-occurrences can be counted for different elements.

Once all the co-occurrences have been counted, a matrix can be compiled of pair-wise relations between those values.

These matrices are typically called co-word-matrix for a matrix of the co-occurrences of keywords, co-citation-matrix for the co-occurrence of citations, co-author-matrix for the co-occurrence of authors etc.

The information of these matrixes can then be converted via clustering technologies into 2-dimensional representations or “maps”.

In such maps, the elements that have stronger co-occurrences than others will have a stronger connection and will be closer to each other, as well. Because of this, clusters of elements can then be identified.

[60] Thomson Reuters: Impact Factor, URL http://thomsonreuters.com/products_services/science/free/essays/impact_factor/, accessed February 1, 2012

3.4 Bibliometric Terms

In this subchapter I will define some terms that will be used later in this work.

Co-citations and Knowledge Bases
Two articles that are both cited by the same other article are called co-cited. Co-cited articles are connected and the more often two articles are co-cited the stronger that connection is. On basis of these connections clusters of co-cited articles will emerge. These clusters are called knowledge bases. They define research topics within the scientific field and serve as the basis of the articles that cite them.[61]

Bibliographic Coupling and Research Fronts
Two articles are bibliographically coupled if they have at least one reference in common. Bibliographically coupled articles are connected and the more references they have in common, the stronger that connection is. As with the co-citation, bibliographic coupling leads to clusters, this time to clusters of bibliographically coupled articles. These clusters are called research fronts. They define a research topic within the scientific field and can be seen as the outcome of the articles they cite.[62]

Backward Citation and Forward Citation
A Backward citation is a reference of an article, i.e. another article that is cited by the article. A forward citation is an article a specific other article is cited by.[63]

Google Scholar related terms
Additionally, I want to define two terms that I will frequently use when working with Google Scholar later in this work:

- Result articles: The set of articles that is the result of a specific search in Google Scholar
- Cited bys: The set of articles that are citing one specific article out of the result articles.

3.5 BibTechMon Software

BibTechMon is a program developed by the Austrian Institute of Technology.
With the BTM software one can create networks on basis of data from scientific databases, similarly like the maps mentioned by van Raan.[64] In the following I will

---

[61] See Schiebel (2011): Lecture notes in Technologie- und Innovationsmanagement III
[62] See Schiebel (2011): Lecture notes in Technologie- und Innovationsmanagement III
[63] See the equivalent definition for patent citations in Duguet; MacGarvie: How Well Do Patent Citations Measure Flows of Technology? Evidence from French Innovation Surveys
[64] For further information about BibTechMon see: Kopcsa, Schiebel (1998): Science and Technology Mapping: A New Iteration Model for Representing Multidimensional Relationships and Noll, Fröhlich,

show the basic of steps how to create networks and detect clusters within the networks in BibTechMon.

At first, one has to create a project, in which the relevant data will be saved. Then, one adds the database with the information gathered from the scientific database to the project. From that database one can then extract the keywords or the elements from the database on which the network will be based on. A screenshot can be seen in Figure 2:

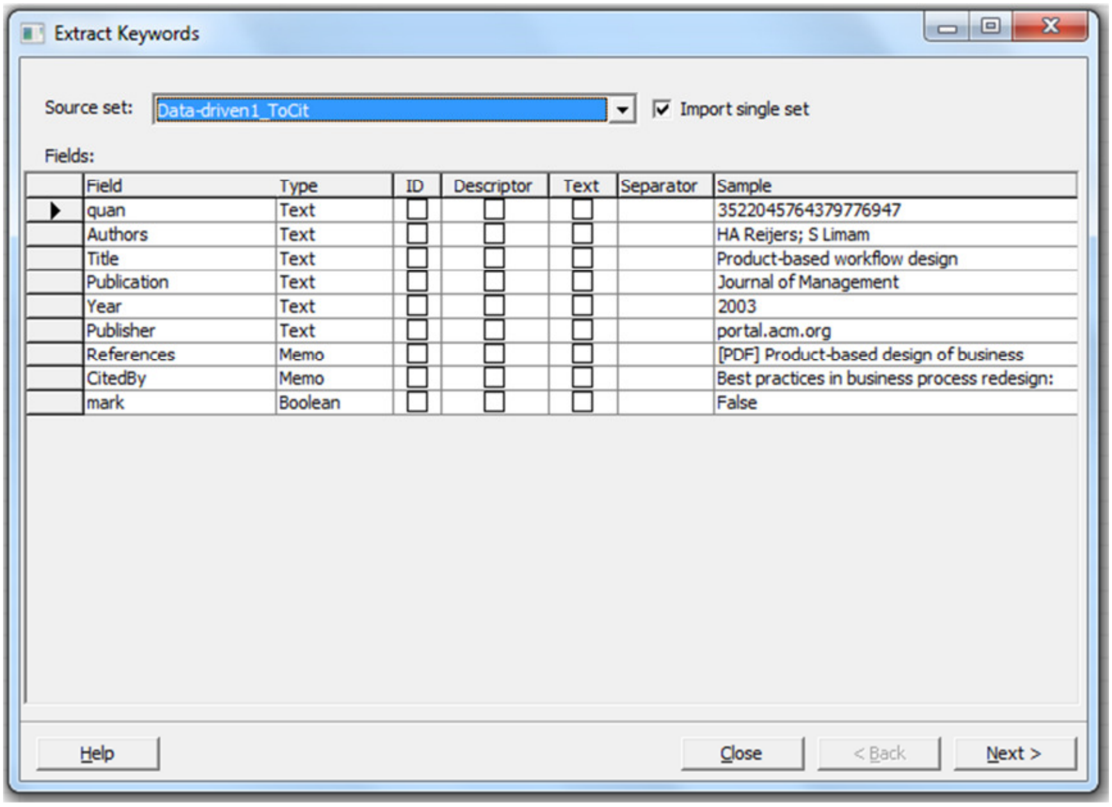


Figure 2: Extract Keywords screen from BibTechMon

In order to work with these elements, an ID and a descriptor have to be chosen. As the ID the *quan* field will be chosen. The *quan* contains a unique number for each article, so that each article can be identified. As for the descriptor, it is necessary to choose a field that contains a list of elements. In this case, one of the fields Authors, References and CitedBy could be chosen. In each case, the separator has to be given as *;* (semicolon), in order to split the list of authors, references and cited bys into their single elements. Subsequently, the list of elements found in the fields will be displayed. Usually, the number of elements will be limited to approximately 1,000 because with higher numbers, the iteration that will be performed later would take too long.

Before the iteration, the terms will be placed randomly as can be seen in Figure 3.

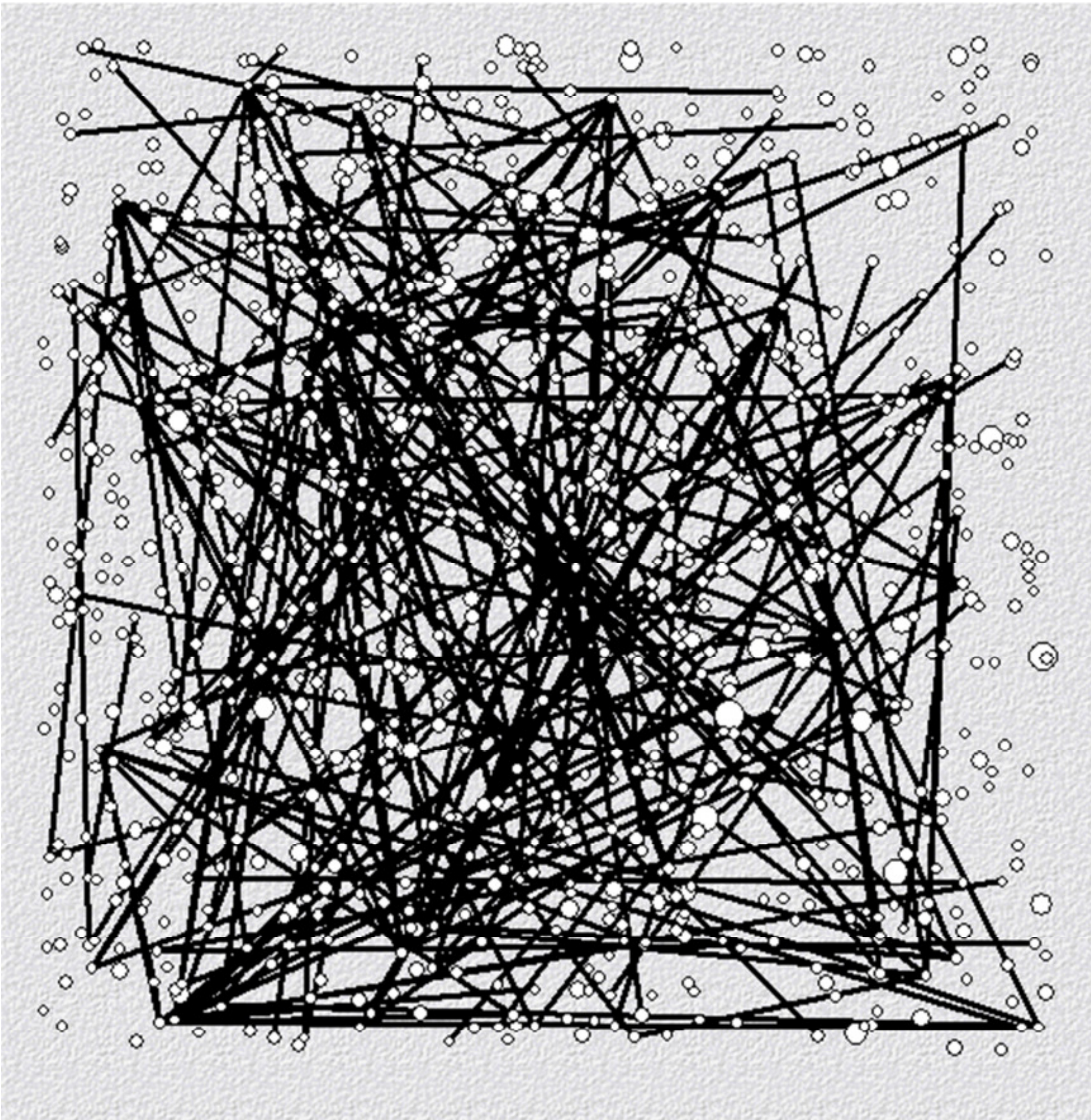


Figure 3: Random distribution of elements in BibTechMon before the iteration

Before one starts the iteration, one can regulate several options. The most important options are the step size and the Sonstwert ("other value"). The step size defines how much the position of one element can change from one iteration step to the next. The Sonstwert defines the repelling force between the elements. Commonly the step size will be placed close to the minimum of the possible values. The Sonstwert will be placed close to the maximum of the possible values. Both can be seen in the screenshot in Figure 4.
Also, the elasticity threshold can be changed, but usually it will remain in its original position. With the elasticity threshold one can determine that only the stronger connections between elements will be relevant for the iteration.

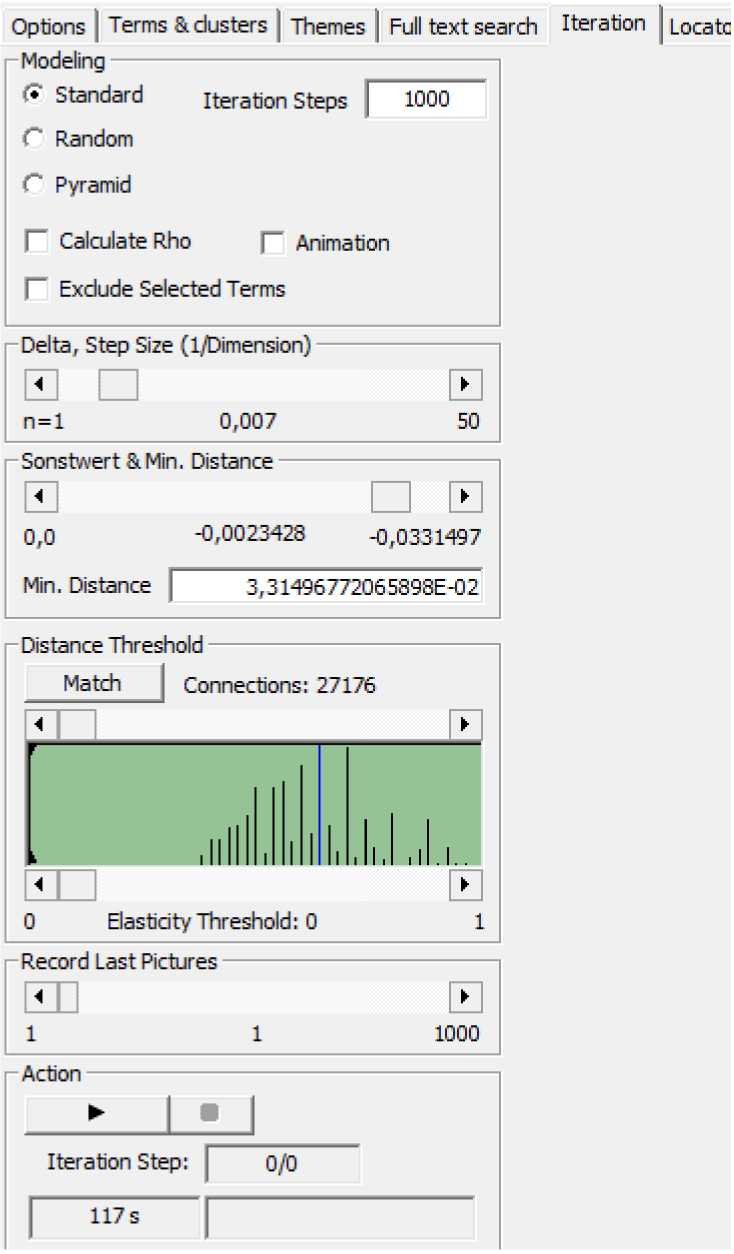


Figure 4: Screenshot of the options possible to regulate related to the iteration in BibTechMon

Usually, with 1,000 iteration steps one receives a network where clusters can already be identified. In the graph in Figure 5 we see the result of such an iteration of 1,000 iteration steps.

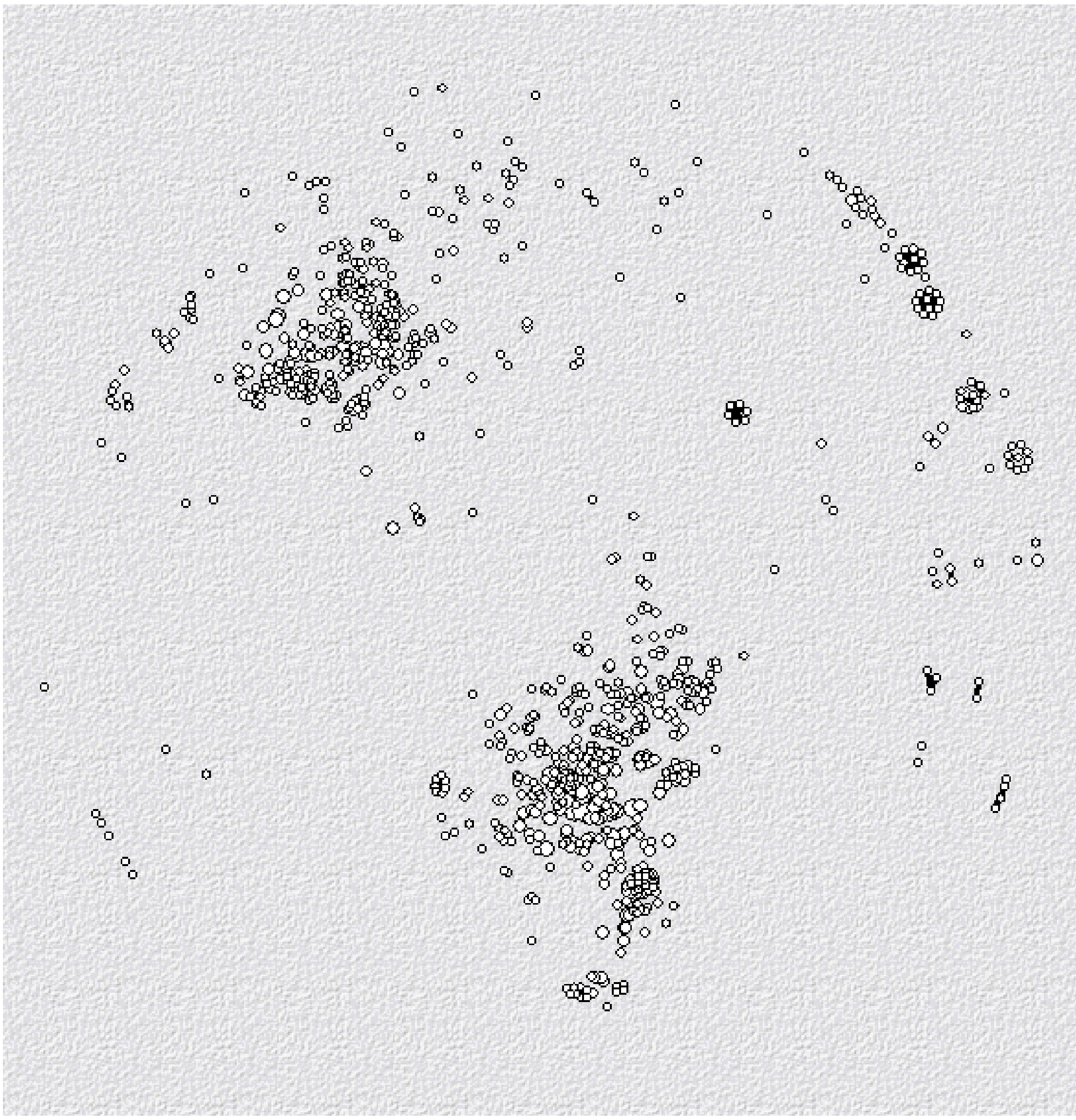


Figure 5: Distribution of elements after the iteration is complete

The density map that can be placed over the network can be of additional help for the identification of clusters. This can be seen in the screenshot in Figure 6:

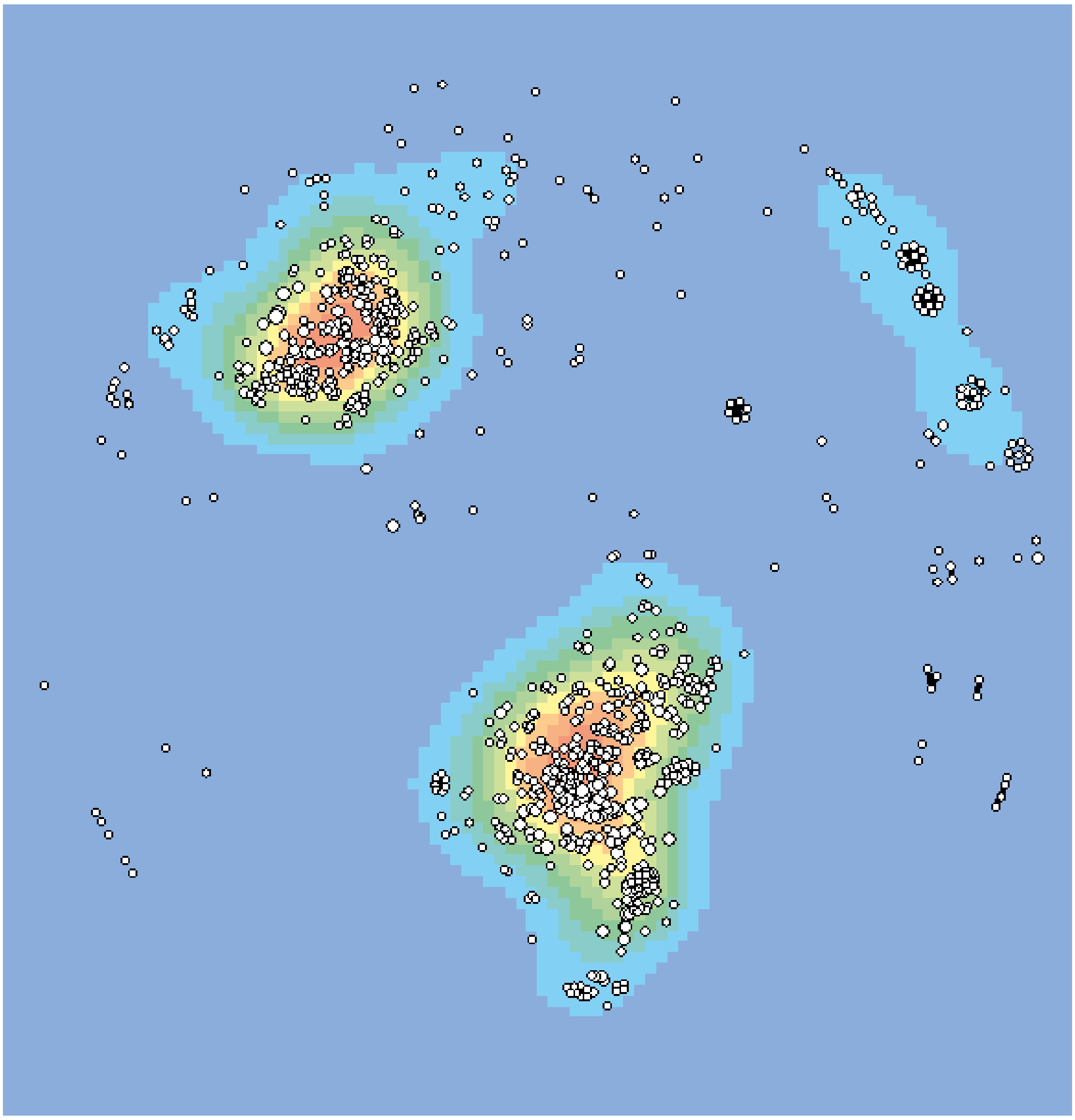


Figure 6: Density map placed over the distribution of elements

After the iteration is done and the density map is placed over the network, one can adjust the presentation of the network. For example, one can adjust the size of the circles that represent the elements in the network. As well, one can limit the amount of connections between the elements that are displayed in order to increase the visibility of the elements.

When the iteration is done, one can select a group of elements, for example, the articles of one cluster, and display the information of the elements selected. The information of such selected elements might look like in the screenshot in Figure 7:

(i)	name	freq.	sel.
570	A taxonomic description of computer-based clinical decision	2	▶
571	A three-layer architecture for e-contract enforcement in an e-service	2	▶
221	A three-tier view-based methodology for M-services adaptation	3	▶
106	A web services implementation framework for financial enterprise content	4	▶
575	Achieving interoperation of grid data resources via workflow level integration ()	2	▶
580	Alert-driven e-service management (290734198379412141)	2	▶
581	Alerts for healthcare process and data integration (5672858102209725896)	2	▶
586	An EREC framework for e-contract modeling, enactment and monitoring	2	▶
587	An event driven approach to customer relationship management in e-brokerage	2	▶
107	An Integration of Web Service and Workflow to a Wealth Management Order	4	▶
594	Applying recommender system based mashup to web-telecom hybrid service	2	▶
615	Building enterprise mashups (6063389704148447000)	2	▶
239	Constraint-based Negotiation in a Multi-Agent Information System with Multiple	3	▶
636	Context-aware adaptive service mashups (6813751578305842506)	2	▶
647	Data-driven Methodology to Extending Workflows to E-services over the	2	▶
661	DMIS: design and prototype of a future clinical eHealth system	2	▶
694	FAST ACCESS: A system architecture for RESTful Business Data	2	▶
256	Flows and views for scalable scientific process integration	3	▶
717	Identifying and Understanding Clinical Processes ()	2	▶
730	Knowledge web-based system architecture for collaborative product	2	▶
777	Process modelling in process-oriented enterprise mashups	2	▶
780	Process-oriented enterprise mashups (4784947803639458259)	2	▶
806	SeDeLo: Using Semantics and Description Logics to Support Aided Clinical	2	▶
812	Situational Enterprise Services ()	2	▶
849	Towards a Context-Oriented Dynamic Medical Information System supporting	2	▶
883	Using three AOSE toolkits to develop a sample design	2	▶

Figure 7: Names of marked elements from the network in BibTechMon

As well, one can display the articles containing the selected elements. This leads to the screenshot in Figure 8:

hits	Authors	Title	Publication	Year	Publisher	References	CitedBy
11	DKW Chiu; SC	An architecture for	36th Hawaii	2003	Citeseer		An EREC framework for e-contract
9	SC Cheung; DKW	A data-driven	36th Hawaii	2003	Citeseer		A three-layer architecture for e-co
9	F Rosenberg; F	Composing restful	IEEE Internet	2008	computer.or		Composing restful services with jo
9	D Benslimane; S	Services mashups: The	Internet Computing,	2008	ieeexplore.i		[PDF] Outsourcing business to clou
8	DKW Chiu; WCW	An event driven	Conference on	2003	Citeseer		A three-layer architecture for e-co
7	PCK Hung	Developing		2004	computer.or	A data-driven	An EREC framework for e-contract
3	DKW Chiu; Z Shan;	Designing workflow	Technologies for	2005	Springer	A data-driven	Flows and views for scalable scien
3	D Domazet; MC	An infrastructure for	. Proceedings of	2000	ieeexplore.i	Active	Knowledge web-based system arc
2	D Sluis; KP Lee	[PDF] DICOM	Medicamundi	2002	healthcare.p	[PDF]	[PDF] Opportunities and Challenge
2	T Wendler	[PDF] Workflow	Medicamundi	2001	healthcare.p		[PDF] DICOM SR-integrating struc
1	SH Kuk; HS Kim; JK	An e-Engineering	Computers in	2008	Elsevier		Design of service-oriented archite
1	P Ciancarini; O	[PDF] A case study in		1998	Citeseer		Agent-oriented software engineer
1	P Ciancarini; A	Multiagent system	Intelligent Agents	2000	Springer	[PDF] A case	Developing multiagent systems: Th
1	F Biennier; J Favrel	[CITATION]	Proceedings of the				Knowledge web-based system arc
1	DS Domazet; FN	Active data-driven	CIRP Annals-	1995	Elsevier		Integrative design environment to
1	T Kiss; A Tudose; P	[PDF] SRB data	of the UK	2007	allhands.org		Achieving interoperation of grid da
1	DF Sittig; A Wright;	Grand challenges in	Journal of	2008	Elsevier		Creating and sharing clinical decisi
1	A Harrison; I Kelley;	[PDF] Workflows	of the UK	2007	omii.ac.uk	[BOOK]	Solving the grid interoperability pr
1	D Kirsh	[PDF] A few thoughts	Intellectica	2000	Citeseer		Personal information management

Figure 8: Articles containing selected elements in BibTechMon

Both of these options are very useful in order to acquire information about the topics of the clusters.

3.6 Scientific Databases

Scientific databases in this context are web sites that allow the user to search in a large collection of scientific publications, to access these publications and to get further information about them. Three of the biggest and most relevant databases for scientific articles are Web of Science, Scopus and Google Scholar. These three are commonly used as sources and are often compared with each other.[65] Web of Science is published by Thomson Reuters. It encompasses, via the Web of Knowledge, also by Thomson Reuters, over 40 million source items according to the company and covers 23,000 journals[66]. Scopus is run by Elsevier and encompasses 46 million records and 18,500 journals.[67]

Both offers contain numerous options for the search of scientific articles. As well, they offer the possibility to download sets of articles including information suitable for bibliometric analysis, such as the references, the organizations the authors belong to etc. Both products are subscription-based and hence only open to certain users.

Google Scholar, on the other hand, is free and open for all users. There are, however, significant differences between Google Scholar, Web of Science and Scopus. While the latter two concentrate on articles published from selected journals and other strictly scientific sources, Google Scholar indexes articles directly from the web.[68] This leads to the situation that also publications that are not strictly scientific, such as diploma theses might get indexed by Google Scholar. Also, Google Scholar does not provide additional bibliometric data, such as the organizations of the authors or lists of references. Another disadvantage of Google Scholar is its higher error rate when it comes to titles and authors. This is caused by the automatic indexing of articles, while Web of Science and Scopus offer more consistent data. Google Scholar does not publish how many articles it covers. Google Scholar also does not allow the download of sets of articles. Therefore, I will use a self-written script to access the data.

There are, however, two important reasons for using Google Scholar instead of the other databases:

- As previously mentioned, Google Scholar is the only one of these databases that is free of charge and accessible to all. This makes it much easier for other researchers to reproduce one's results, because they don't need to have a subscription to the paid databases.
- It has been shown, that the coverage by Google Scholar of certain fields is significantly better than the coverage by Web of Science and Scopus. These fields are social sciences, business and most relevant for this topic: computer science.[69] As for the Web of Science, Meyer at al. even state: "In assessing

[65] For example, see Meho, Yang (2007): Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar

[66] Thomson Reuters: Web of Knowledge, URL: http://thomsonreuters.com/content/science/pdf/Web_of_Knowledge_factsheet.pdf, accessed February 17, 2012

[67] Scopus: Content Coverage Guide, URL: http://www.info.sciverse.com/scopus/scopus-in-detail/facts, accessed February 17, 2012

[68] Google Scholar: Help page, URL: http://scholar.google.com/intl/en/scholar/help.html, accessed February 20, 2012

[69] Harzing, van der Wal (2008): Google Scholar as a new source for citation analysis

> publications and citations, the ISI Web of Science is inadequate for most areas of computer science and must not be used."[70]

These are the reasons why in this work I will use Google Scholar as the source of the data.

[70] Meyer et al. (2009): Research Evaluation for Computer Science

4 Work with Google Scholar

In this chapter, I will describe the procedure of the analysis of different BPM-related search terms, on basis of Google Scholar data. In the first subchapter, I will describe the raw procedure involving the programs I used. In the following subchapters, I will look more closely into the specifics of the work with Google Scholar and the data that can be acquired from it.

4.1 Steps of the Analysis

As mentioned in the chapter about the structure of this work, several different search terms and fields of BPM will be analyzed. For each of those search terms, the steps that will be done are essentially the same:

Preparation and execution of the search in Google Scholar
First, a search term has to be chosen. This term needs to be based on the terminology that is used in the specific field. Then, it is suitable to experiment with different versions of the search term and to compare the results. First of all, it needs to be checked, if the results fit the topics, if they don't fit the topic, other search terms should be used or certain restrictions can be applied. These restrictions are further described in Chapter 4.3. In general, search terms with a high number of results are desirable. Additionally, a high average number of forward citations is desirable, as well, in order to yield useable cited by networks (this is also further described in 4.7). After a search term has been chosen, the search term and the chosen restrictions can be entered into the Google script. The Google script will then download the search results from Google Scholar and turn them into a "comma-separated values" file or CSV file. In the Google script, the number of search results and the number of cited bys can be chosen. In this work, I will always choose the maximum number of articles of 1,000 (which is also the maximum possible in Google Scholar). As the maximum number of citing articles for each result article, I will always choose 100. It should be noted that Google Scholar does not allow automated download of its search results. Instead the results should be manually saved and the Google script can then be used on this saved data. The source code of the Google script can be found in Appendix I and the changes of the source code that were necessary after Google slightly altered the Google Scholar format can be found in Appendix II. A short overview of the functionality of the script is given in Appendix IV.

Conversion of CSV file and work in BibTechMon
After the CSV file has been created, it must be converted into a MDB data base. The precise steps on how to do this can be found in Appendix IV. This MDB file must then be loaded into BibTechMon, where the different networks can be created. The networks will be created as described in Chapter 3.5. In the networks, I will then identify cluster and have a look at frequent keywords in the titles and will also look at the most frequently cited articles.

4.2 Search in Google Scholar

At first, I will describe the kind of search terms that can be entered in Google Scholar. The same search terms can also be entered into the corresponding field of the Google script.
In the search field the following constructs can be used:

- Quotation marks in order to search for a complete phrase (e.g. *"business process management"*).
- *OR* and *AND* constructs, in order to search for alternative terms (e.g. *process OR workflow*) or lists of terms (e.g. *business AND process AND flexibility*; the *AND* is actually optional, thus this search is equivalent to *business process flexibility*).

It should be noted, that the *AND*/*OR* constructs cannot be nested, i.e. more complicated constructs like *(workflow flexibility OR adaptivity) OR (process flexibility OR adaptivity)* are not possible. The example term would be interpreted as *workflow flexibility OR adaptivity OR process OR flexibility OR adaptivity*.
It is possible to restrict the search of these terms to the titles of the articles; however, since for us the whole article is relevant, I ignored this option. In the following screenshot the advanced search options of Google Scholar can be seen.

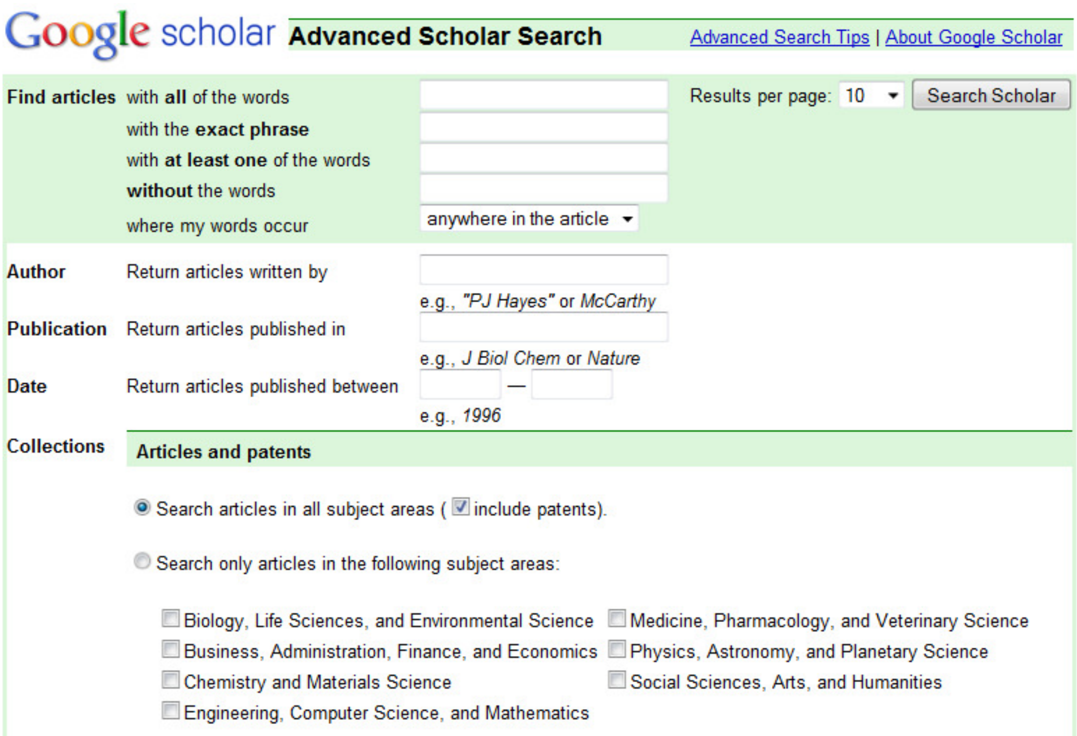


Figure 9: Advanced search options in Google Scholar.
Note: Further options regarding legal publications have been cut out.

4.3 Options in Google Scholar

There are several further options, on how the search with Google Scholar can be refined. I will now describe the most relevant options and their implementation in the Google script.

First of all, the language in which Google Scholar will be used can be chosen. In this work, I will always use the English version, because some options are only available in that version (for example the exclusion of patents). The script will as well automatically access the English version.
Other options include: Limitation of the search to one or several scientific fields. The available fields are:

- Biology, Life Sciences, and Environmental Science
- Medicine, Pharmacology, and Veterinary Science
- Business, Administration, Finance and Economics
- Physics, Astronomy, and Planetary Science
- Chemistry and Materials Science
- Social Sciences, Arts, and Humanities
- Engineering, Computer Science, and Mathematics

The limitation to certain scientific fields can also be done in the script, by entering the abbreviations of these scientific fields in the corresponding field of the script. However, most of the time, I will not use this option, because sometimes articles do not get the correct classification and might hence be excluded if we limit our search to certain fields.
It is also possible to enter the year or a time span in order to limit the result articles to a certain time period.
It can also be set whether patents should be included or not. In this work, I will always exclude patents, since I am limiting my interest to scientific publications and also because patents do not play a very important role in this field.
For the cited articles, as well, we will exclude patents.
The script excludes patents both from the result articles and for the forward citations automatically.
Furthermore, Google Scholar allows us to restrict the search to certain publications, e.g. certain journals. In this work I will do this exemplary for the Data & Knowledge Engineering Journal in Chapter 5.4.1. For all the other searches I will not use this option. However, the script generally gives the option of limiting the search to any kind of publication or journal, if needed.
As well, it is possible to limit the result articles to a certain author. The author search can also be done in the script by simply using the construct *author:* in the search term, e.g. *author:"jorge hirsch",* in order to search for articles by Jorge Hirsch. The *author:* construct can be combined with other search terms.
Additionally, it is possible to search legal opinions and journals. Since this is not relevant to our search field, I ignored these options and I also did not include this possibility in the script.

4.4 Possible Networks from the Google Scholar Data

The networks are created on basis of the keywords that can be extracted from the Google Scholar data. In order to create a meaningful network, it is necessary that the keywords can “connect” different articles. That means, that each article has one or more of said keywords and other articles might have keywords in common with that article. In the case of the Google Scholar data, these keywords are:

- Authors
- Cited bys
- References

In case of the cited bys and the references it is also possible to invert the relation between connecting terms and presented terms. Thus, in total five network types are possible:

- The author network
- The cited by network
- The inverted cited by network
- The references network called co-citation network
- The inverted references network called BibCoup network

In this chapter I will discuss the possibilities and limitations of each of these network types.

4.4.1 Author Network

This is the network of the authors. Two authors will be connected when they have published an article together. The more articles they have published together the stronger the connection will be. The more articles a single author has published or co-published, the bigger his circle in the network will be. With this type of network, one can analyze which authors frequently publish together, which authors are important in general and if there are connections between different groups of authors that publish together.
When working with Google Scholar there are two challenges regarding the author data and the resulting author network.

Challenge 1
The name of one author might be written in different forms for different articles, especially if the name consists of more than the usual two parts (first name and last name). One prominent example here is WMP van der Aalst. His name is sometimes found as W van der Aalst, sometimes as WMP van der Aalst and sometimes even as WMP Aalst. BibTechMon, however, will consider each variant of this name as a different author, hence in the author network there would be all kinds of variants of

that name instead of treating them all as the same name. In order to address this challenge, I added a function to the script that tries to standardize these more complicated names. This function will convert any name into the following form: First letter of the first part of the name + last part of the name. Everything that is between those parts will be ignored. This leads to transformations of the following kind: All kinds of variants of the name WMP van der Alst will be transformed to the name W Aalst. Other examples include the variants of the name of Arthur ter Hofstede that will be transformed to A Hofstede. All names that are already in the form of one first letter of the first name + one last name will not be affected.

Challenge 2
Google Scholar does often not return the full set of authors of a given article but only the first two or three. This can lead to an incomplete list of authors and also to incomplete author networks. However, usually the most important authors should be mentioned first and only the less important ones will be ignored. There is the possibility of downloading additional information for each article via the "Cite" function which should give the complete set of authors. However, this has been ignored so far, because it would significantly increase the number of requests that need to be made to Google Scholar.

4.4.2 Cited by Network

This is the network that will be used most when working with Google Scholar data, since it yields the most promising results. In this network the method of bibliographic coupling will be used on the "cited bys" or the forward citations of the result articles. Two forward citations will be connected if they are both citing the same article. The more articles they are citing together, the stronger the connection between the two forward citations will be. The more articles have the same forward citation (i.e. are cited by the same other article), the bigger the circle of the forward citation will be in the cited by network. This network can be used to determine clusters or research fronts among the citing articles.

The cited bys are accessible for each article. Up to 1,000 cited bys of a single article can be accessed in Google Scholar, given that the article is cited that many times which is quite rare, at least in our field of BPM. However, to limit the necessary queries to Google Scholar, in this work we only take the first 100 cited bys for each article. In the vast majority of cases, this is sufficient, since most articles are not cited by more than 100 other articles, anyway.

It is important to note that the number of cited bys is crucial for the quality of the network. If the result articles are not cited by a significant number of articles, no useful cited by network can be drawn.

4.4.3 Inverted Cited by Network

This network is based on the same data as the cited by network. However, it exchanges the object on which will be worked on. While the cited by network works on the cited bys and the connections are made via the result articles, the inverted cited by network works on the result articles and makes the connections via the cited bys. The method used is hence the method of co-citation: Two result articles that are cited by the same article are connected. The more articles they are both cited by, the stronger the connection will be. And the more often the result article is cited, the bigger its circle in the network will be.

This network can be used in order to determine thematic clusters among cited articles or knowledge bases. They can be particularly interesting, when comparing them with the clusters found in the cited by network. This way it can be explored which thematic clusters developed in the research front on the basis of the clusters in the knowledge base.

4.4.4 Co-Citation Network

The co-citation network works on the references or citations of the result articles. As the name implies, the method used is the co-citation method: Two articles are connected if they are both cited by the same article. The more articles they are both cited by, the stronger the connection will be. And the more often a reference appears in the result articles, the bigger its circle will be in the graph. As in the inverted cited by network this allows for the detection of knowledge bases or thematic clusters among cited articles. The difference is that in the co-citation network the cited articles are the references of the result articles, while in the inverted cited by network the result articles themselves are the cited articles.

Google Scholar does not explicitly return the references of an article. Because of that, they can only be partly recovered in an indirect way. If one article *X* of the result articles is also in the list of cited bys of article *Y*, then article *Y* is a reference of article *X*. I.e. references can only be detected if there is an overlap between result articles and cited bys. If, however, the result articles are, for example, limited to articles from the year 2000 and the cited bys are limited to articles from the year 2005, then no references will be found, because an overlap of result articles and cited bys has been excluded via the year.

Due to the highly incomplete nature of the references that can be reconstructed on basis of the Google Scholar data, co-citation networks will not be used in this work.

4.4.5 BibCoup Network

The BibCoup network works on the result articles with the method of bibliographic coupling. Two of these articles will be connected if they have an identical reference. The more references they have in common, the stronger the connection between the citing articles will be. The more references a citing article has, the bigger its circle will be.

This network requires the references of the result articles. Since the reference data that can be acquired from Google Scholar is highly incomplete, also the BibCoup networks will not be used in this work.

4.5 Comparison of Networks

In the following table I sum up the differences between the cited by network, the inverted cited by network, the co-citation network and the BibCoup network.

Table 1: Comparison of the different networks

<table>
<tr><th></th><th>Cited by network</th><th>Inverted cited by network</th><th>Co-citation network</th><th>BibCoup network</th></tr>
<tr><td>Method:</td><td>Bibliographic coupling</td><td>Co-citation</td><td>Co-citation</td><td>Bibliographic coupling</td></tr>
<tr><td>Result:</td><td>Research fronts</td><td>Knowledge bases</td><td>Knowledge bases</td><td>Research fronts</td></tr>
<tr><td>Citing articles are acquired from:</td><td>Forward citations of result articles</td><td rowspan="4">See Cited by network</td><td>Result articles</td><td rowspan="4">See References network</td></tr>
<tr><td>Implication:</td><td>Forward citations are variable and only indirectly determined by the search terms</td><td>Result articles directly determined by search terms</td></tr>
<tr><td>Cited articles are acquired from:</td><td>Result articles</td><td>References of result articles</td></tr>
<tr><td>Implication:</td><td>Result articles are directly determined by the search terms</td><td>References are fixed</td></tr>
</table>

4.6 Time Constraints when Working with Forward Citations

The time constraints are different when working with forward citations (as it is the case when working with the cited bys in Google Scholar) or backward citations (as it is the case when working on the reconstructed references in Google Scholar or the normal references in the Web of Science or Scopus). When working with the forward citations, the result articles are the cited articles and the forward citations or the "cited bys" are the citing articles.

With the options in Google Scholar the year of the publication of the result articles can either be unfiltered (i.e. all years are fine), limited to the publications from a certain year and after that year, or limited to a certain span of years which can also be just one year.

If the year of one result article is y_i and n is the number of articles in the result list, then the years of the articles are $y_0 ... y_n$.

If the search was limited to one specific year s then $y_0=...=y_n=s$.
If $c_{i,j}$ are the years of the citing articles of article y_i then the relation between y_i and $c_{i,j}$ is $c_{i,j} >= y_i$.
If $y_0=...=y_n=s$, then $c_{i,j}>=s$.
Additionally, the years of the citing articles, i.e. the forward citations, can be limited in the same fashion.
In this way, the articles that form the knowledge bases can be limited to publications from a certain year (or time span) and at the same time the articles that form the research fronts can also be limited to publications from another year (or another time span). In this work, however, we will not use this option, in order to get as many result articles and cited bys as possible.
Now, we will compare that to the time constraints when using the normal references data, either in its reconstructed and incomplete form from Google Scholar or in its fairly complete form from sources like Web of Science or Scopus.
If we work with the references, the result articles are the citing articles and the references are cited articles.
Again, if the year of one result article is y_i and n is the number of articles in the result list, then the years of the articles are $y_0...y_n$.
If $r_{i,j}$ are the years of the references of article y_i, then the year of the publication of the references will be $r_{i,j} <= y_i$.
If the references are provided by the database itself, as it is the case with Web of Science and Scopus, then usually it is not possible to automatically restrict the references to a certain time span. If the references are reconstructed via the cited bys as mentioned in Chapter 4.4.4, they can again be limited to a certain time span. However, it must be taken into account, that the time span of the citing articles and the time span of the result articles must overlap each other, since otherwise no references will be able to be reconstructed. Also, it must be noted, that usually only few references can be reconstructed in the way described.
So far, the restriction of the results to certain years has not yet been implemented into the Google script.

4.7 Metrics

In order to assess the quality of a search term (including restrictions to certain publications or research areas), at first, of course the number of search results should be taken into account. If this number is low – especially, if it is lower than the 1,000 articles that only can be accessed anyway – the search term might have been too restrictive.
Another metric for the quality of the search results is the count of the number of citing articles, i.e. the sum of the number of "cited bys".
The more forward citation the result articles have, the more connections will usually be between the cited bys in the cited by networks. In the course of this work I have discovered that the more cited bys the result articles had, the better cited by networks could be created and the clusters could be identified more easily. Based on these experiences, I will use the number of cited bys of the result articles as an indicator for the quality of the results. The script gives the number of cited bys for each article in

the CSV file.[71] Of, course one still has to check whether the articles fit the search terms topic-wise, this indicator is merely about the quality of the cited by network that can be created from that data and not about the quality of the content per se.

[71] Note: The script gives the number of the total number of cited bys, i.e. including patents. That means, when patents are excluded, the real number of cited bys might be lower. However, in our field, citations through patents are rather uncommon, so the citation count should not be altered greatly.

5 Bibliometric Analysis

In this chapter, I will perform bibliometric analyses with BibTechMon on the basis of Google Scholar data. In total, the results of 15 different search terms will be presented. Those search terms are grouped into four categories:

- Search terms related to BPM in general
- Search terms related to specific fields within BPM
- Search terms related to fields that are linked to BPM
- Search terms related to one journal and two conferences in the BPM field

For each of these categories a number of different sub topics have been chosen. Each sub topic will receive its own subchapter.
Each subchapter will have the same structure that is given as follows:
At first, the topic of the chapter and the exact search term that has been used will be given.
Then, the network that has been created in BibTechMon on basis of the data from Google Scholar acquired with that search term will be shown.
In those networks, I identified clusters, after executing the iteration in BibTechMon and placing the density map over the network as described in Chapter 3.5. I looked at the titles of the elements in the cluster, again as described in the mentioned chapter. From the titles of the elements I assessed the contents of the cluster and I named each cluster according to its content. The names of the identified clusters will be added to the graph of the network.
In a table below the network, I will give some numbers related to the search and to the network. First, I will give the number of search results in Google Scholar for that particular search term. This is an indicator for the size of the search field in total. It should be noted again, that only the first 1,000 results can actually be accessed in Google Scholar. Then, I will give the number of cited bys for these first 1,000 articles in Google Scholar, in the sense of the metrics I mentioned in Chapter 4.7. This number shows us how often the first thousand articles in that search field are cited, which can be seen as an indicator for the maturity of the field. Then I mention the number of terms shown in the network. This is necessary in order to compare the size of the clusters in the network to the size of the whole network.
Below that, I will give the number of elements in each identified cluster in another table.
After these numbers, I will describe each cluster. From the titles of the elements in the cluster I chose terms that appear frequently and describe the contents of the cluster. For each cluster I will give a number of those terms, which I named "keywords", in order to give an impression of the content of the cluster. As well, for each cluster I will present the three most cited articles from the cluster. These articles I will receive with the option in BibTechMon to display the articles containing the selected elements, also described in Chapter 3.5. I will also give the authors of these articles as they are given by Google Scholar. As mentioned in Chapter 4.4.1 the list of authors given by Google Scholar is not always complete.

The aim of this chapter is to identify and describe the clusters in the field of BPM and in related fields, as well as to note which researchers play an important role in the field of BPM.

5.1 Analysis of BPM Data in General

In order to cover as much of the field of business process management as possible, the data from three different search terms will be used in this chapter, to perform analyses about BPM in general. The search terms used are as follows:

- *business process management* (without quotation marks and not limited to any particular scientific field)
- *workflow management* (without quotation marks and not limited to any particular scientific field)
- *"business process" OR workflow management* (here, the search was limited to the business and engineering related fields, in order to avoid too many articles not belonging to our field. The quotation marks around business process are necessary to maintain the OR-structure.)

For each of these search terms I will create a cited by network. For the first search term, I will additionally create an author network, in order to gather information about the most important authors in the field of BPM.

5.1.1 BPM Search Term

The first network I will analyze, is the cited by network created on basis of the business process management search term. The network with the clusters can be seen in the following figure.

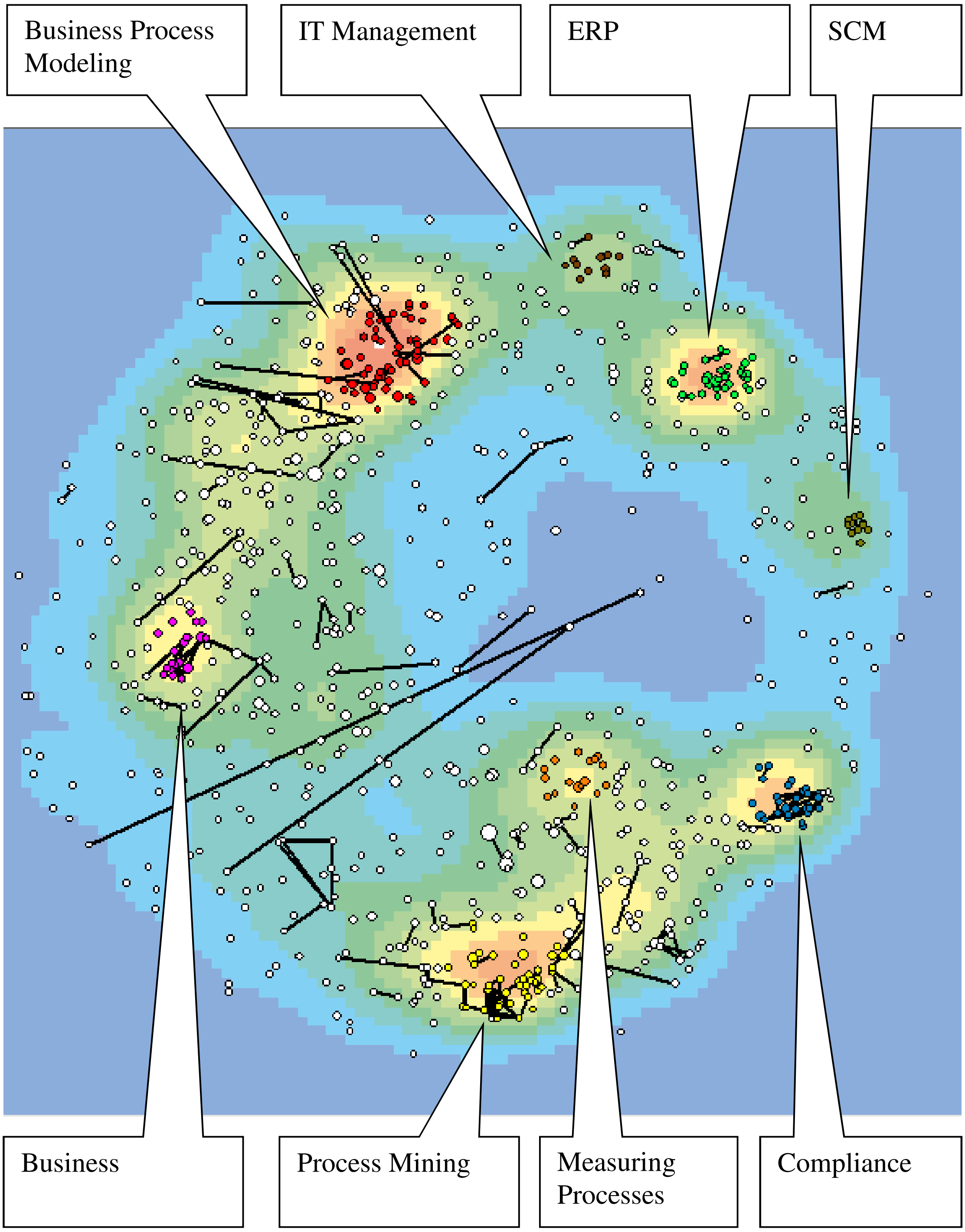


Figure 10: Cited by network of the business process management search term

Table 2: Numbers of the business process management search and of the corresponding network

Number of search results:	1,730,000
Number of cited bys (includes patents) of the first thousand articles:	74,624
Number of terms in the graph:	886

In this network, the following clusters have been identified:

Table 3: Clusters in the cited by network for the business process management search term

Name of the cluster	Number of articles in the cluster
Business Process Modeling	57
Process Mining	45
ERP	35
Compliance	29
Business	20
Measuring Processes	17
IT Management	12
SCM	11

Now, we will look at the articles within those clusters and the articles by which they are connected.

Business Process Modeling Cluster
The biggest cluster covers the topics of business process modeling and business process reengineering. Frequent terms in the article of the cluster include:

- Redesigning Processes
- Process Redesign
- Business Process Reengineering/BPR
- Process Modeling
- Process Change

The most frequently cited articles are:

- Reengineering: business change of mythic proportions? (Davenport)
- The new and the old of business process redesign (Earl)
- A methodology for business process redesign: experiences and issues (Wastell; White)

Many of the cited articles are from business-related journals such as the Business Process Management Journal, Sloan Management Review and Harvard Business Review. Additionally, articles from journals from the computer science field and the information sciences are cited. Those journals include the Information Systems Journal, the MIS quarterly and the Journal of Strategic Information Systems.

Process Mining Cluster
The next bigger cluster is mainly about process mining. It includes terms like:

- Process Mining
- Process Discovery
- Workflow Mining
- Event Logs
- Conformance Checking
- Conformance Testing
- Business Process Analysis
- Petri Nets

The most frequently cited articles are:

- Workflow mining: Discovering process models from event logs (van der Aalst; Weijters)
- Mining process models from workflow logs (Agrawal; Gunopulous)
- Conformance testing: Measuring the fit and appropriateness of event logs and process models (Rozinat)

Apart from articles from van der Aalst, also articles from Casati and Leymann, two other well-known authors in the BPM field, are cited. The cluster can hence be considered a cluster of typical BPM articles.

ERP Cluster
The next biggest cluster is the ERP cluster. It almost exclusively contains articles about enterprise resource planning. Many of these articles cite articles from the Business Process Management Journal, which has a stronger business-oriented focus. However, those articles might still be relevant to our field due to their process-related nature.
The most common keywords in the titles of the articles are:

- ERP
- Small and Medium Enterprises
- ERP Implementation
- Critical Success Factors

The most frequently cited articles are:

- Enterprise resource planning: a taxonomy of critical factors (Al-Mashari; Al-Mudimigh)
- Planning for ERP systems: analysis and future trends (Chen)
- Change management strategies for successful ERP implementation (Aladwani)

Compliance Cluster
The fourth biggest cluster is about compliance and checking of business processes. Keywords include:

- Compliance
- Checking
- Rules
- Process Analysis
- Semantic

The most frequently cited articles are:

- Modeling control objectives for business process compliance (Sadiq; Governatori)
- Auditing business process compliance (Ghose)
- A static compliance-checking framework for business process models (Liu; Muller)

The articles that are cited are often from the Business Process Management Conference and from the publications by Springer about the Business process management workshops.

Business Cluster
The next cluster is again a more business-related cluster. Terms included are:

- Collaborative Business Process Management
- Open Innovation
- Organization
- Business Process Management
- Process Management

The most frequently cited articles are:

- Implications of business process management for operations management (Armistead)
- Business process management-lessons from European business (Pritchard)
- Business process management as competitive advantage: a review and empirical study (Hung)

Frequently cited publications are the Business Process Management Journal, Harvard Business Review and the Journal of Management Information.

Measuring Processes Cluster

The next cluster contains the topics quality, complexity of processes and process evaluation.

The most common keywords are:

- Complexity
- Modularity
- Evaluation
- Metrics
- Quality

The most frequently cited articles are:

- What makes process models understandable? (Mendling; Reijers)
- Guidelines of business process modeling (Becker; Rosemann)
- Complexity metrics for business process models (Gruhn; Laue)

IT Management Cluster
The 7th cluster is the IT management cluster, which focuses on business aspects of information technology. Often mentioned keywords are:

- IT Capability.
- Business Value
- Resource-based Analysis (of IT)
- Business Process

The most frequently cited articles are:

- Develop long-term competitiveness through IT assets (Ross; Beath)
- The implications of information technology infrastructure for business process redesign (Broadbent; Weill)
- Information technology as competitive advantage: The role of human, business and technology resources (Powell)

Publications cited are mostly journals related to management and related to management information systems.

SCM Cluster
The smallest cluster is the supply chain management or SCM cluster. The most common keywords here are:

- SCM Frameworks
- SCM Concepts

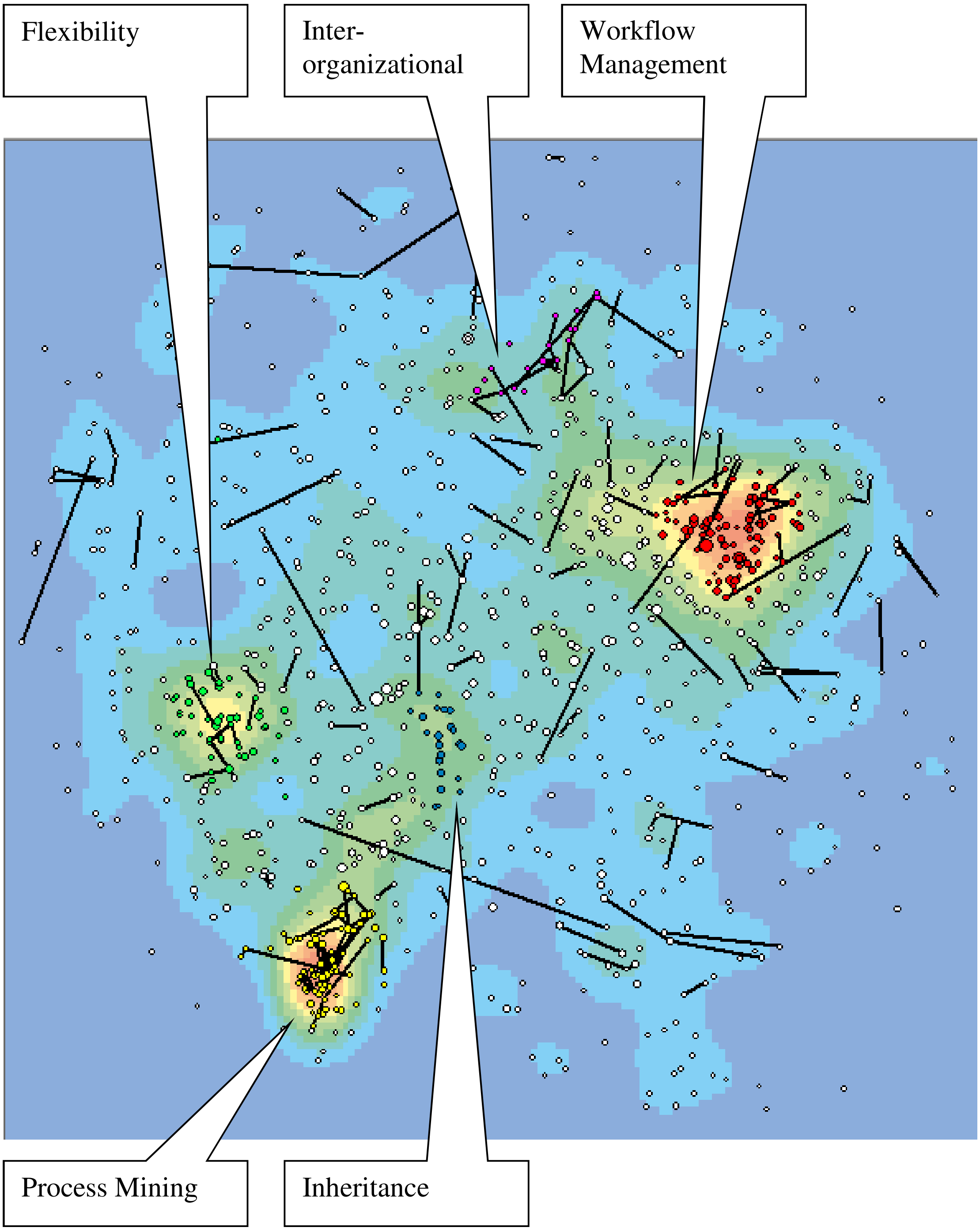
The most frequently cited articles are:

- Supply chain management: more than a new name for logistics (Cooper; Lambert)
- Issues in supply chain management (Lambert)
- Supply chain management: an analytical framework for critical literature review (Croom; Romano)

The cited publications include journals from the areas of logistics, marketing and business.
An analysis of the result articles acquired from Google Scholar shows that the reason why SCM-related articles are found in our search result is that many supply chain management articles contain references to business processes, BPR and similar topics.

5.1.2 Workflow Search Term

In order to possibly receive additional clusters in the BPM field, we analyze the cited by network on basis of the workflow network search term, which is shown in the following figure. The term workflow management is strongly related to business process management[72] and in fact has been used before the term BPM became widespread.



[72] van der Aalst, ter Hofstede, Weske: Business Process Management: A Survey

Figure 11: Cited by network of the workflow management search term

Table 4: Numbers of the workflow management search and of the corresponding network

Number of search results:	266,000
Number of cited bys (includes patents) of the first thousand articles:	39,505
Number of terms in the graph:	1,010

Table 5: Clusters in the cited by network for the workflow management search term

Name of the cluster	**Number of articles within the cluster**
Workflow Management	91
Process Mining	71
Flexibility	48
Inheritance	22
Inter-organizational	18

Below we will have a closer look at those clusters:

Workflow Management Cluster

The biggest cluster in this network is the workflow management cluster with various workflow management topics and a special focus on flexibility and distributed workflows. It includes keywords like:

- Adaptive Workflow Management Systems
- Distributed Workflow Management Systems
- Flexibility
- Cross-organizational Workflows
- Enterprise-wide Workflows

The most cited articles are:

- Failure handling in large scale workflow management systems (Alonso; Kamath; Agrawal)
- INCAs: Managing dynamic workflows in distributed environments (Barbara; Mehrotra)
- Providing high availability in very large workflow management systems (Kamath; Alonso; Günthör)

Other topics among the cited articles include distributed environments, large workflow management systems and collaboration.

Process Mining Cluster

The second biggest cluster in the Workflow management network is the process mining cluster. Keywords include:

- Process Mining

- Workflow Mining
- Discovering (in combination with the following terms: Petri Nets, Expressive Process Models, Models of Behavior, Simulation Models, Social Networks)
- Genetic Process Mining
- Interactive Workflow Mining

The most frequently cited articles are:

- Mining process models from workflow logs (Agrawal)
- Rediscovering workflow models from event-based data using little thumb (Weijters)
- A machine learning approach to workflow management (Herbst)

Another author whose articles are cited in this cluster is van der Aalst.

Flexibility Cluster
The next cluster is about flexibility and case handling. Keywords include:

- Flexibility
- Flexibility Schemes
- Adaptive
- Dynamic
- Change Patterns
- Change Support
- Case handling

The most cited articles are:

- Correctness criteria for dynamic changes in workflow system--a survey (Reichert; Rinderle)
- Formal foundation and conceptual design of dynamic adaptations in a workflow management system (Weske)
- Worklets: A service-oriented implementation of dynamic flexibility in workflows (Adams; ter Hofstede; Edmond)

Authors of other frequently cited articles include van der Aalst and again Weske.

Inheritance Cluster
The next cluster is the inheritance/inter-organizational cluster. Keywords included in the titles of the articles are:

- Inheritance Patterns
- Inter-organizational
- Cross-organizational

The most frequently cited articles are:

- The application of workflow nets to workflow management (van der Aalst)
- Workflow management: modeling concepts, architecture and implementation (Jablonski)
- Production workflow: concepts and techniques (Leymann)

Among the other cited articles general workflow management topics are dominating, as well.

Inter-organizational Cluster
The last cluster is about inter-organizational workflows. Frequent terms include:

- Cross-organizational
- Cooperative (also in the German spelling kooperativ)
- Peer-to-peer
- E-services

The most frequently cited articles are:

- CrossFlow-cross-organizational workflow for virtual organizations (Grefen)
- WW-Flow: Web based workflow management with runtime encapsulation (Y Kim; Khang; D Kim; Bae)
- DartFlow: A workflow management system on the web using transportable agents (Cai; Gloor)

5.1.3 Combined Business Process/Workflow Management Search Term

Now we will analyze the cited by network that has been created on basis of the search term that includes both business process management and workflow management and that was limited to the business field and the computer science field. The network and the identified clusters can be seen in the following figure:

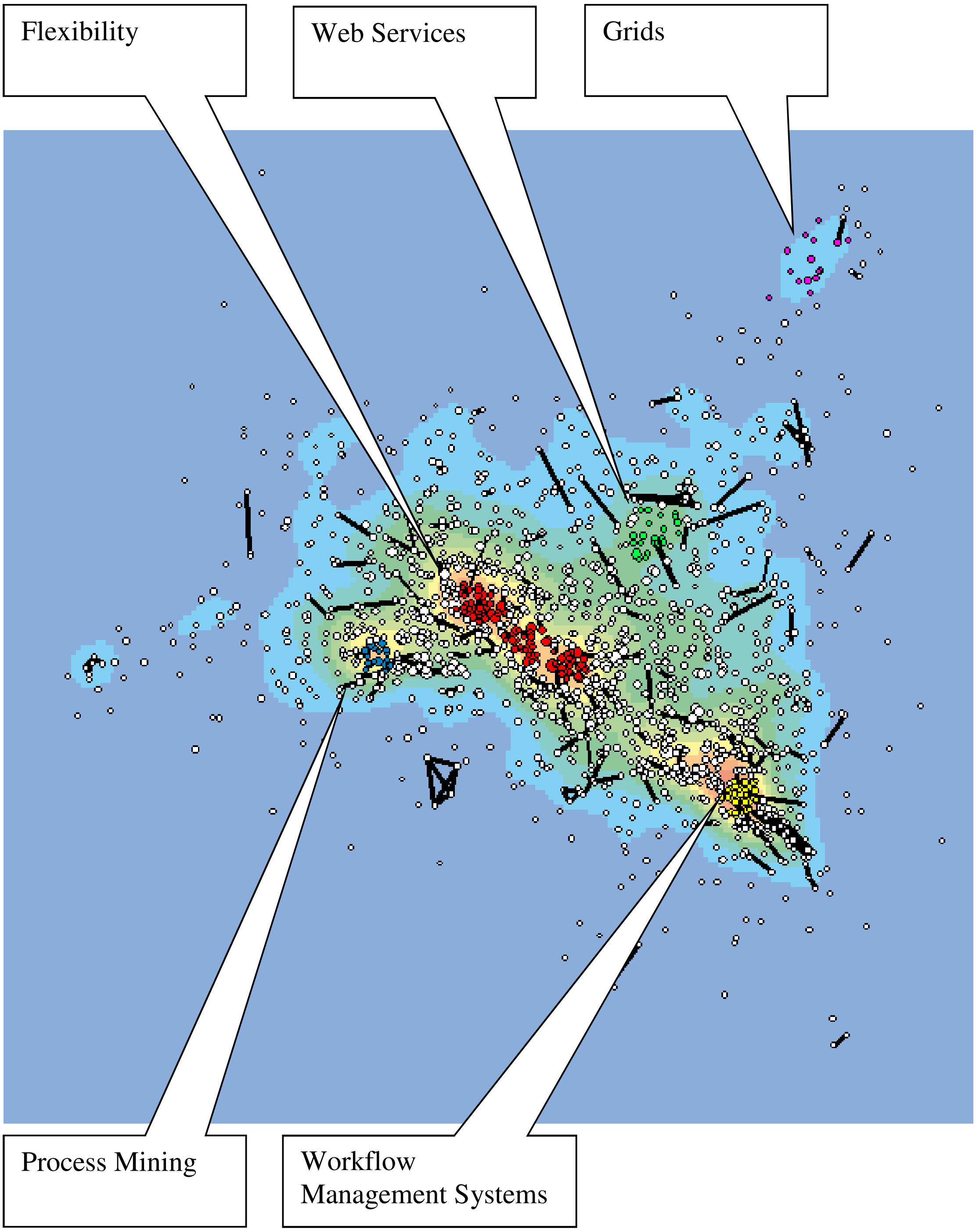


Figure 12: Cited by network of the combined business process/workflow management search term

Table 6: Numbers of the combined business process/workflow management search and of the corresponding network

Number of search results:	28,700
Number of cited bys (includes patents) of the first thousand articles:	59,465
Number of terms in the graph:	1,549

The names of the clusters and the numbers of the articles in those clusters can be seen in the following table:

Table 7: Clusters in the cited by network for the combined business process/workflow management search term

Name of the cluster	Number of articles within the cluster
Flexibility	96
Workflow Management Systems	36
Web Services	24
Process Mining	16
Grids	15

Now we will have a look at the contents of those clusters:

Flexibility Cluster

There is one big cluster with the topic of process flexibility/adaptivity/dynamic which can be divided into three smaller clusters. In the first sub-cluster important authors that are cited include Reichert, Rinderle and van der Aalst. Frequent keywords are:

- Flexible
- Change
- Dynamic
- Adaptive
- Process-Awareness
- Patterns
- Verification
- Modeling

The next sub-cluster still cites Reichert and Rinderle, however the articles from van der Aalst are dominant among the cited publications. Terms in this sub-cluster include:

- Dynamic Flexibility
- Flexible
- Adaptive
- Petri Nets
- Pi Calculus
- Various projects systems, such as ADEPT, YAWL and MANET

In the last sub-cluster, again keywords like the following occur:

- Flexibility
- Process Evolution
- Change
- Patterns (Flexibility Patterns, Exception Handling Patterns)
- Process Evolution
- Verification

The most cited articles of the whole cluster are:

- Application of Petri nets to workflow (van der Aalst)
- Three good reasons for using a Petri-net-based workflow management system (van der Aalst)
- Workflow management, modeling concepts, architecture and implementation (Jablonski)

Other cited articles contain topics like "workflow evolution", "inheritance of workflows" and formal topics.

It should be noted, that all of these three sub-clusters do contain a number of articles that do not directly belong to one single topic, hence the topics "drifts" among various fields like "process flexibility" and "process models".

Workflow Management Systems Cluster
The next cluster concentrates on WFMS and distributed and cross-organizational workflows. Particularly, several specific systems are mentioned such as:

- OPERA
- EVE (Event Engine)
- MARIFlow
- METEOR

The most frequently cited articles are:

- WIDE-a distributed architecture for workflow management (Ceri; Grefen)
- Functionality and limitations of current workflow management systems (Alonso; Agrawal; Abbadi; Mohan)
- Providing high availability in very large workflow management systems (Kamath; Alonso; Günthör)

Web Services Cluster
The next cluster focuses on web services, choreography and cross-organizational workflows. Within the cited articles, the ones with cross-organizational topics are dominant. Another topic in the cluster itself is modeling and constraints.

Common keywords in the titles of the articles are:

- Cross-organizational
- Inter-organizational
- Workflow
- Web Service
- Choreography

The most frequently cited articles are:

- Facilitating cross-organizational workflows with a workflow view approach (Schulz)
- Crossflow: Cross-organizational workflow management for service outsourcing in dynamic virtual enterprises (Grefen; Aberer; Ludwig)
- The view-based approach to dynamic inter-organizational workflow cooperation (Chebbi; Dustdar)

Process Mining Cluster

The 4th cluster is again a process mining cluster. Keywords include:

- Genetic Process Mining
- Fuzzy Mining
- Specific mining tools and frameworks, such as EMiT and the ProM framework

The most frequently cited articles are:

- Workflow mining: Discovering business process models from event logs (Aalst; Weijters)
- Rediscovering workflow models from event-based data using little thumb (Weijters)
- Workflow mining: a survey of issues and approaches (Aalst; Dongen; Herbst)

Grids Cluster

The last cluster is about grid workflows and scientific workflows. In the cited articles, as well, grid workflows and scientific workflow are the dominant topic. Keywords include:

- Grid Workflows
- Scientific Workflows
- Grid Computing

The most frequently cited articles are:

- A taxonomy of workflow management systems for grid computing (Yu)
- Programming scientific and distributed workflow with Triana services (Churches; Gombas; Harrison)
- Pegasus: A framework for mapping complex scientific workflows onto distributed systems (Deelman; Singh; Su; Blythe)

5.1.4 Author Network

In addition to the cited by networks, I also looked at the author network created on basis of the BPM search term. The author network contains a total of 1,173 authors and can be seen in the following figure:

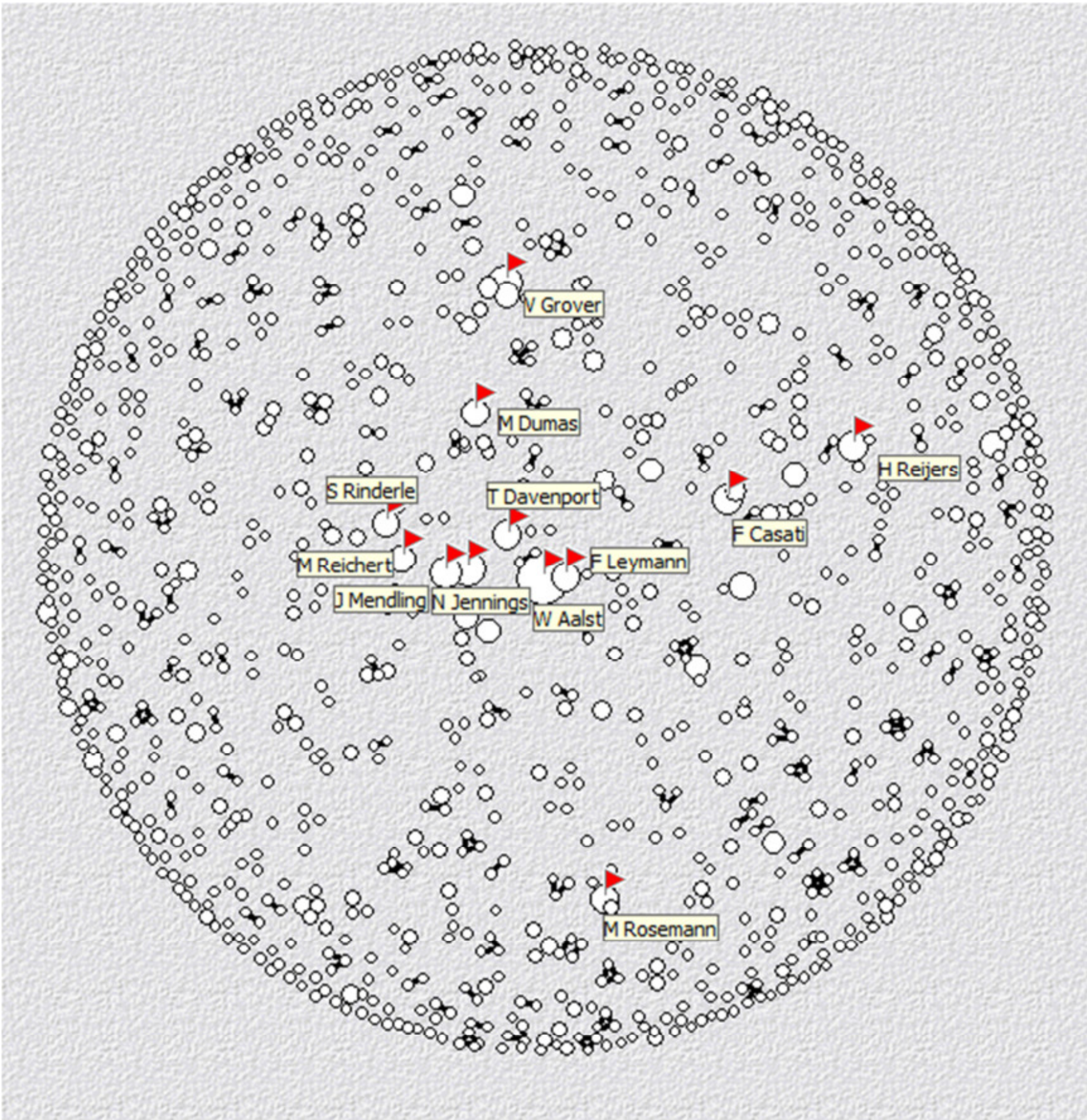


Figure 13: Author network on basis of the BPM search term

The marked authors with the labels on them each published seven or more articles that were covered by that search term. The bigger the circles in the graph are, the more articles have been published by the author, the biggest circle representing van der Aalst who published 26 articles.

The following numbers of publications have been found in the data for each author:

Table 8: Authors with the highest numbers of publications in the BPM search field

Author	Number of publications found
W Aalst	26
V Grover	12
N Jennings	12
J Mendling	9
F Casati	8
T Davenport	8
F Leymann	8
H Reijers	8
M Dumas	7
M Reichert	7
S Rinderle	7
M Rosemann	7

5.2 Analysis of Specific Fields of BPM

In the last chapter the BPM topic in general has been analyzed. In this this chapter I will now analyze certain subtopics of BPM.
The analyzed fields are as follows:

- BPMN and BPEL
- Data-driven Workflows
- Metrics
- Compliance
- Mobile Processes

For a description of these fields, see Chapter 2.3. For each of these fields the cited by network will be created.

5.2.1 BPMN and BPEL

In order to determine clusters in the search field of BPMN and BPEL I used the following search term: *BPMN OR BPEL*.
The cited by network of the data gathered with this search term can be seen in the following figure:

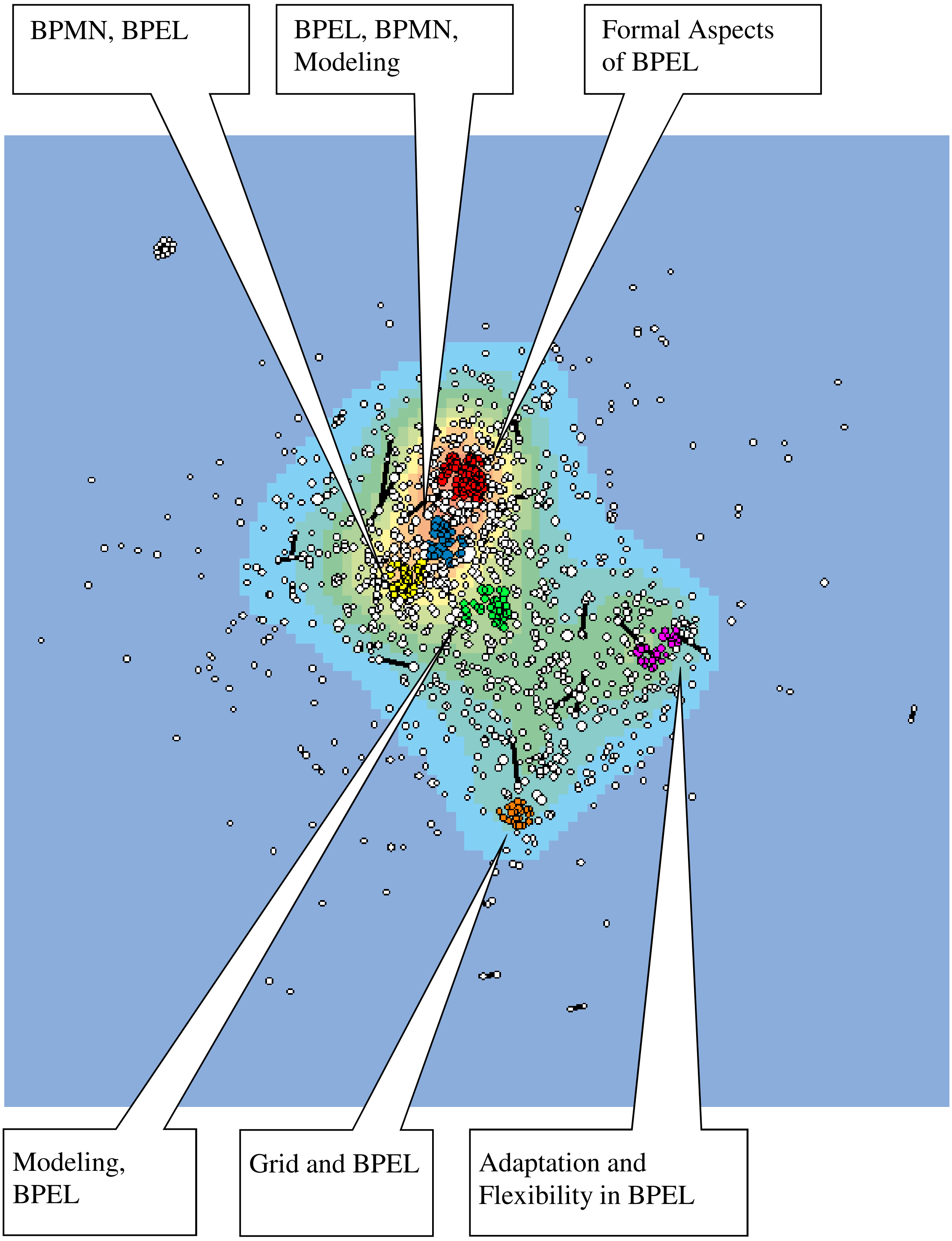


Figure 14: Cited by network of the BMPN/BPEL search term

Table 9: Numbers of the BPMN/BPEL search and of the corresponding network

Number of search results:	27,100
Number of cited bys (includes patents) of the first thousand articles:	14,071
Number of terms in the graph:	1,300

The following clusters have been identified:

Table 10: Clusters in the cited by network for the BPMN/BPEL search term

Name of the cluster	Number of articles within the cluster
Formal Aspects of BPEL	94
BPMN, BPEL	39
Modeling, BPEL	38
BPEL, BPMN, Modeling	32
Adaptation and Flexibility in BPEL	29
Grid and BPEL	23

Formal Aspects of BPEL Cluster
This cluster is mostly about formal aspects of BPEL. Frequent keywords include:

- Petri Nets
- Analysis
- Compliance
- Verifying
- Web Service (Composition)
- Modeling/Formal Modeling
- Semantics

All of the most frequently cited articles are connected to BPEL. The three most frequently cited articles are:

- Transforming BPEL to Petri nets (Hinz; Schmidt)
- Formal semantics and analysis of control flows in WS-BPEL (Ouyang; Verbeeck; van der Aalst)
- ASM-based semantics for BPEL: The negative control flow (Fahland)

Among the other cited articles BPEL also is the dominant topic.

BPMN, BPEL Cluster
This cluster covers both BPMN and BPEL related topics. Frequent keywords are:

- Metrics
- Semantics
- Modeling

Also, the transformation of BPMN into BPEL and vice versa is mentioned.

The most frequently cited articles are:

- On the translation between BPMN and BPEL: Conceptual mismatch between process modeling languages (Recker)
- Translating bpmn to bpel (Ouyang, van der Aalst, Dumas)
- Using BPMN to model a BPEL process (White)

Also, in most of the other cited articles BPEL and/or BPMN are mentioned. As well, a couple of other process-related languages are mentioned.

Modeling, BPEL Cluster

This cluster focuses more on BPEL and on topics like collaboration between organizations, as well as the topic of web services. Common keywords other than BPEL include:

- Collaboration
- Coordination
- Inter-organizational
- Web Services
- Composition
- Choreography

The three most frequently cited articles are:

- BPEL4Chor: Extending BPEL for modeling choreographies (Decker)
- From RosettaNet PIPs to BPEL processes: A three level approach for business protocols (Khalaf)
- From inter-organizational workflows to process execution: Generating BPEL from WS-CDL (Mendling; Hafner)

BPEL, BPMN, Modeling Cluster

In this cluster, the modeling of processes in BPMN and BPEL is a topic, as well as the composition and orchestration of web services using BPEL. Keywords include:

- (Business) Modeling/Models
- Life Cycle Modeling
- Orchestration
- Composition

The most frequently cited articles are:

- From BPMN process models to BPEL web services (Ouyang; Dumas; ter Hofstede)
- Using BPMN to model a BPEL process (White)

- On the translation between BPMN and BPEL: Conceptual mismatch between process modeling languages (Recker)

As we can see from the cited articles, the transformation from BPMN process models to BPEL are important again.

Adaptation and Flexibility in BPEL Cluster
The articles in this cluster focus on topics such as flexibility and adaptivity of BPEL processes. Another topic mentioned is the field of self-healing processes.
The most common keywords among the articles include:

- Flexibility
- Dynamic
- Self-healing
- Self-adaptive
- Composition
- Adaptation

The three most frequently cited articles are:

- Non-intrusive monitoring and service adaptation for WS-BPEL (Moser/Rosenberg)
- Ao4bpel: An aspect-oriented extension to bpel (Charfi)
- Self-healing BPEL processes with Dynamo and the JBoss rule engine (Baresi; Guinea)

Most of the other cited articles are also BPEL-related

Grid and BPEL Cluster
This cluster is about the usage of BPEL in grid environments, particularly in the context of scientific workflows and grid services.
The most frequent keywords in this cluster are:

- Workflow
- Grid
- Services
- Composition
- Scientific Workflow

The most frequently cited articles are:

- Grid service orchestration using the business process execution language (BPEL) (Emmerich; Butchart; Chen)

- Choreography for the Grid: towards fitting BPEL to the resource framework (Leymann)
- Evaluation of BPEL to scientific workflows (Akram; Meredith)

5.2.2 Data-driven Workflows

In order to catch the relevant articles for data-driven workflows, the following search term has been used: *workflow data-driven OR object-aware OR object-centric OR artifact-based OR product-based*. The cited by network created on basis of this search term can be seen in the following figure:

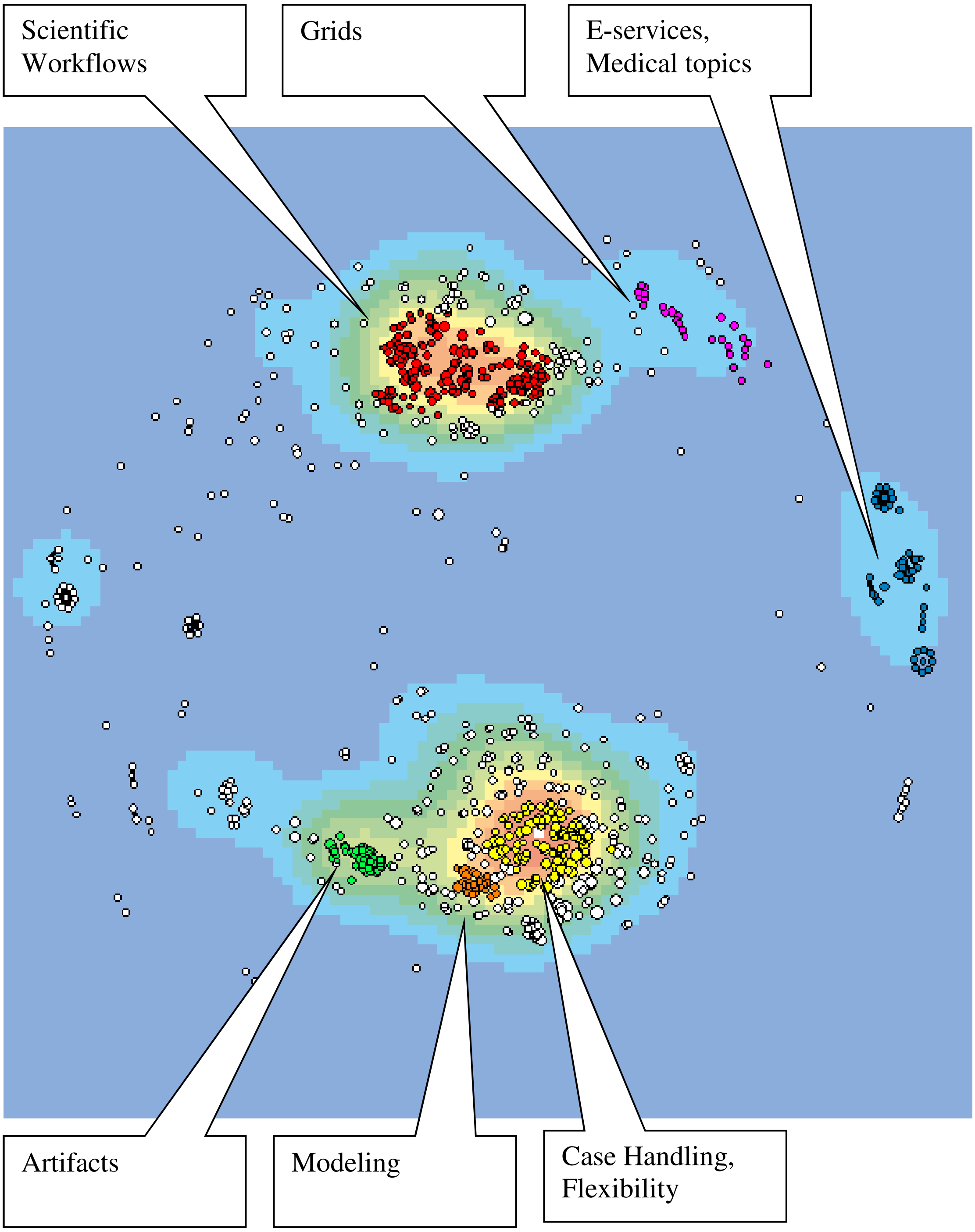


Figure 15: Cited by network of the data-driven workflows search term

Table 11: Numbers of the data-driven search and of the corresponding network

Number of search results:	9,179
Number of cited bys (includes patents) of the first thousand articles:	8116
Number of terms in the graph:	910

In this network, the following clusters have been identified:

Table 12: Clusters in the cited by network for the data-driven workflows search term

Name of the cluster	Number of articles in the cluster
Scientific Workflows	175
Case Handling, Flexibility	108
Artifacts	49
E-services, Medical Topics	43
Grids	28
Modeling	26

Scientific Workflows Cluster

This cluster focuses on scientific workflows and services in a scientific environment. Important keywords used include:

- Web Service
- Workflows
- Scientific
- Neuro-Imaging
- Biomedical
- Scientific Workflow/Process
- Grid (Workflows)
- Data-intensive
- Service-oriented
- Cloud

The most frequently cited articles are:

- Workflows and e-Science: An overview of workflow system features and capabilities (Deelman; Gannon; Shields)
- Provenance collection support in the kepler scientific workflow system (Altintas; Barney)
- Workflows for e-Science: Scientific Workflows for Grids (Taylor; Deelman)

Case Handling, Flexibility Cluster

This cluster covers various topics such as case handling, exception handling, flexibility and data-driven/product driven workflows.

The most relevant keywords used in the titles are:

- Product-based
- Case Handling
- Flexible/Flexibility
- Change Support
- Exception Handling
- Dynamic
- Schema Evolution
- Business Process Redesign
- Constraint-based
- Data-driven

The most frequently cited articles are:

- Case handling: a new paradigm for business process support (van der Aalst; Weske)
- Beyond workflow management: product-driven case handling (van der Aalst)
- Correctness criteria for dynamic changes in workflow systems--a survey (Reichert; Rinderle)

Artifact Cluster
This cluster is strongly about artifact-based workflows. The most important keywords used are:

- Business Artifacts
- Artifact-centric
- Data-centric
- Conformance
- Cross-organizational
- Nested Dynamic Condition

The most frequently cited articles are:

- Towards formal analysis of artifact-centric business process models (Bhattacharya; Gerede; Hull; Liu)
- Automatic construction of simple artifact-based business processes (Fritz; Hull)
- Artifact-centric business process models: Brief survey of research results and challenges (Hull)

E-services, Medical Topics Cluster
This smaller cluster is about e-services and e-health. Important keywords include:

- E-service
- Collaboration
- Inter-enterprise
- Decision Support Systems
- Medical (Information System)
- Clinical Processes

The most frequently cited articles are:

- Standards for clinical decision support systems (Broverman)
- An architecture for e-contract enforement in an e-service environment (Chiu; Cheung)
- A pragmatic framework for understanding clinical decision support (Perreault; Metzger)

Modeling Cluster
This cluster is about the modeling and remodeling of processes. Frequent keywords include:

- Models
- Data
- Modeling
- Best Practices
- Business Process Reengineering

The most frequently cited articles are:

- Design and control of workflow processes: business process management for the service industry (Reijers)
- Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics (Reijers)
- Product-based workflow design (Reijers)

Grids Cluster
The last cluster is about grid workflows and particularly about grids in the context of data-intensive applications. Frequently mentioned terms are:

- Workflow (Patterns)
- Parallel Computing
- Grid

- Data-intensive (Applications)

The most frequently cited articles are:

- Distributed computing with Triana on the Grid (Taylor; Wang; Shields)
- Grid-enabled workflows for data intensive medical applications (Glatard; Montagnat)
- An optimized workflow enactor for data intensive grid applications (Glatard; Montagnat)

5.2.3 Metrics

In order to cover the field of metrics in BPM, we used the following search term: *"business process" OR workflow metrics.* On the basis of the results of this search term, the following cited by network has been derived:

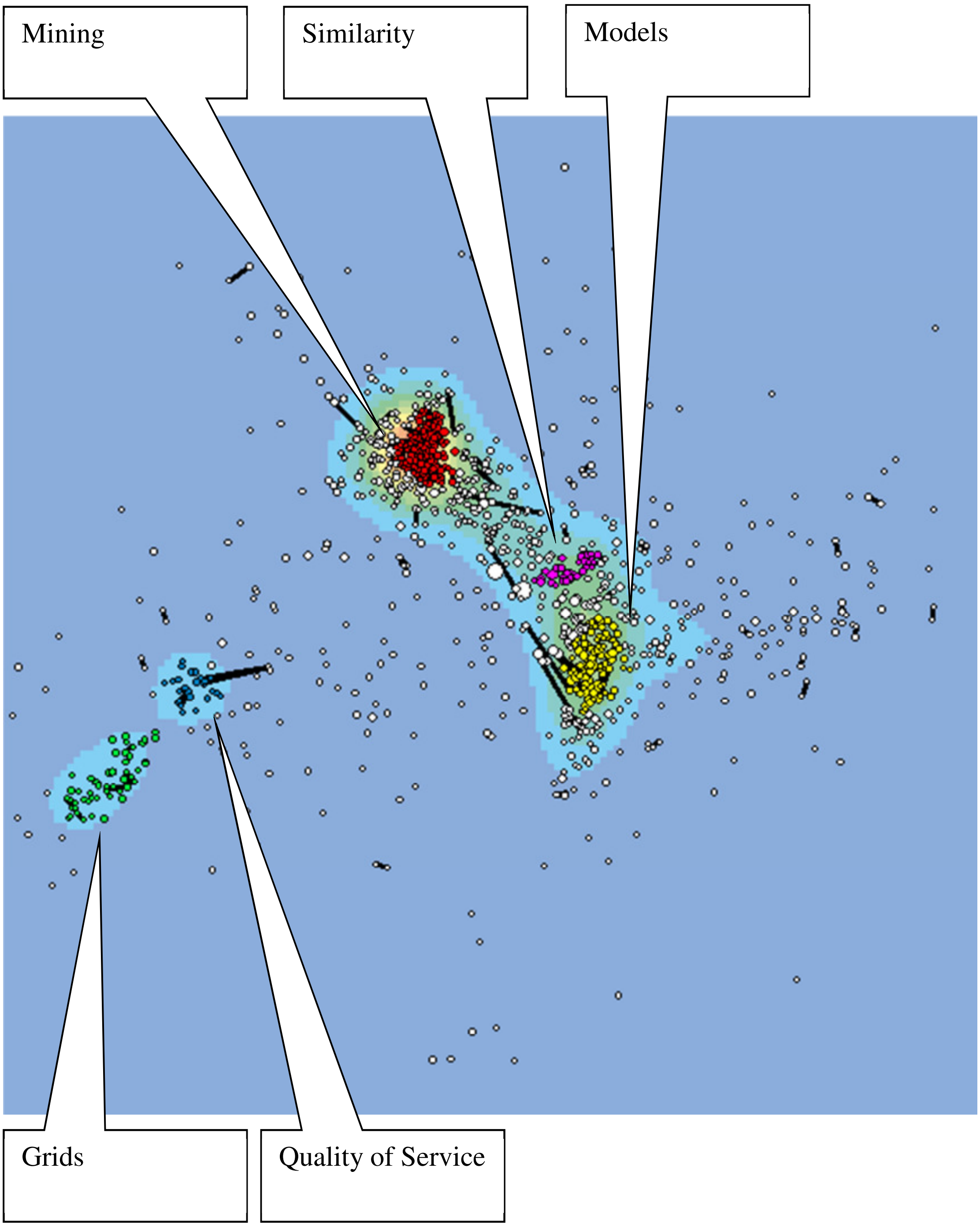


Figure 16: Cited by network of the business process metric search term

In order to demonstrate the effect of a different elasticity threshold in the iteration – see Chapter 3.5 – I created two different network graphs with the Metrics data. One network graph with the elasticity threshold of 0, this graph can be seen in Figure 16, and one graph with the elasticity threshold of approximately 0.40. With the higher elasticity threshold the clusters are more dispersed. This network can be seen in the following graph:

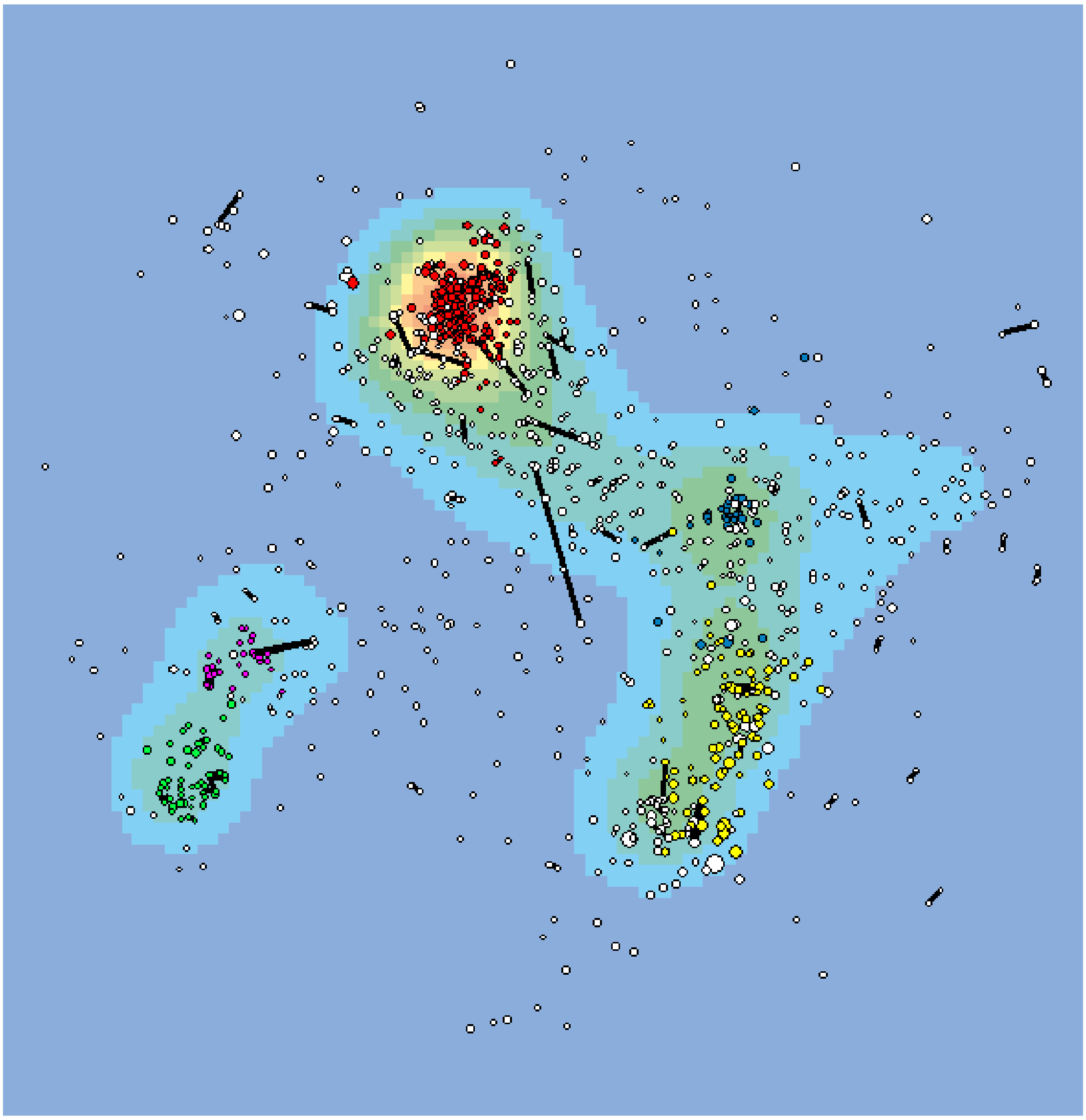


Figure 17: Cited by network of metrics, different elasticity threshold

Table 13: Numbers of the metrics search and of the corresponding network

Number of search results:	41,600
Number of cited bys (includes patents) of the first thousand articles:	17,548
Number of terms in the graph:	1,006

Table 14: Clusters in the cited by network for the business process metrics search term

Name of the cluster	Number of articles within the cluster
Mining	154
Models	89
Grids	52
Similarity	32
Quality of Service	26

Mining Cluster

The first cluster is again a process mining cluster. In the titles of the articles keywords such as the following can be found:

- Process Mining
- Workflow Mining
- Discovering (in combination with several other terms such as: Process Models, Social Networks, Colored Petri Nets, Workflow Models, Reference Models, Simulation Models)
- Machine Learning
- Event Logs
- Business Intelligence/Business Process Intelligence

The most frequently cited articles are:

- A machine learning approach to workflow management (Herbst)
- Rediscovering workflow models from event-based data using little thumb (Weijters)
- Discovering workflow performance models from timed logs (Aalst, Dongen)

Models Cluster

This cluster is about process modeling. The most common keywords are:

- Modeling Grammar
- Semantic
- Framework
- Collaborative Process Modeling

The most frequently cited articles are:

- What makes process models understandable? (Mendling; Reijers)
- On a quest for good process models: the cross-connectivity metric (Vanderfeesten; Reijers; Mendling)
- Influence factors of understanding business process models (Mendling)

Among the other cited articles, Mendling is also one of the dominant authors.

Grids Cluster
This cluster is about grids. The most frequent keywords are:

- Scientific Grid
- Grid Workflow

The most frequently cited articles are:

- Cost-based scheduling of scientific workflow application on utility grids (Yu; Buyya)
- A taxonomy of workflow management systems for grid computing (Yu)
- Workflow enactment in ICENI (McGough; Young; Afzal)

Similarity Cluster
The fourth cluster is about similarity of process models. Common keywords in the titles of the articles are:

- Similarity
- Merging
- Behavior
- Semantic

The most frequently cited articles are:

- Measuring similarity between business process models (Dongen; Dijkman)
- Graph matching algorithms for business process model similarity search (Dijkman; Dumas)
- Measuring similarity between semantic business process models (Ehrig; Koschmider)

Quality of Service Cluster
The fifth cluster is about quality of service (QoS) in a workflow environment. The most common keywords are:

- Time Management
- Quality of Service
- Quality of Service Management
- Web Services
- Evaluation
- Composition

The most frequently cited articles are:

- Workflow management with service quality guarantees (Gillmann; Weikum)
- Workflow quality of service (Cardoso; Sheth)
- Quality of service for workflows and web service processes (Cardoso; Sheth; Miller; Arnold)

5.2.4 Compliance

In this chapter we will look at the results generated on basis of the compliance search term. The search term was: *"business process" OR workflow compliance*. The cited by network can be seen in the following figure:

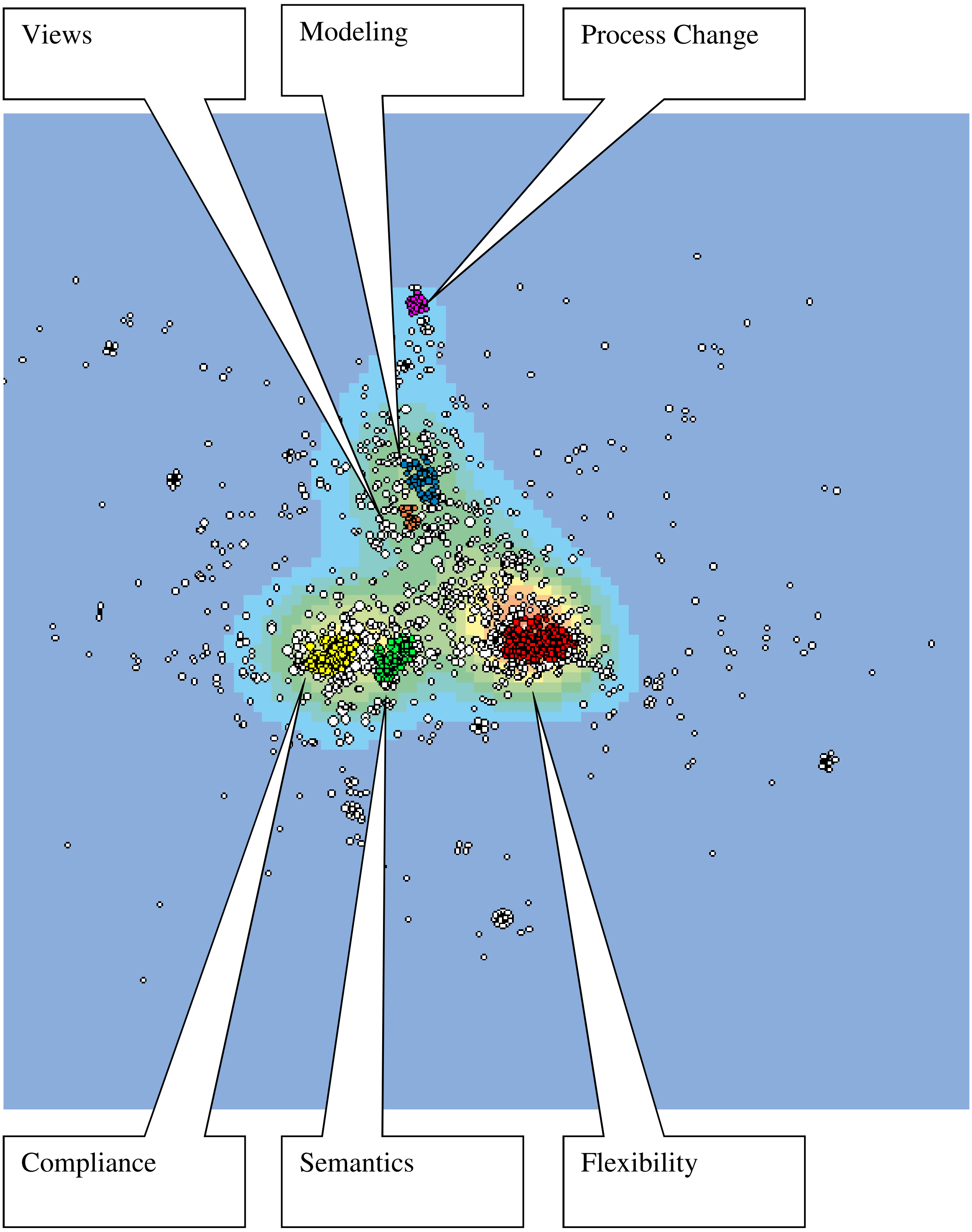


Figure 18: Cited by network of the compliance search term

Table 15: Numbers of the metrics search and of the corresponding network

Number of search results:	38,600
Number of cited bys (includes patents) of the first thousand articles:	12,139
Number of terms in the graph:	1,618

The following clusters have been identified:

Table 16: Clusters in the cited by network for the compliance search term

Name of the cluster	**Number of articles within the cluster**
Flexibility	258
Compliance	126
Semantics	78
Modeling	44
Process Change	25
Views	24

Flexibility Cluster
The first cluster is about flexibility and adaptivity of processes. The most common keywords are:

- Flexibility
- Adaptivity
- Evolution
- Dynamics
- Exception
- Process-Awareness

The most frequently cited articles are:

- Flexible support of team processes by adaptive workflow systems (Reichert; Rinderle)
- Correctness criteria for dynamic changes in workflow-systems--a survey (Reichert; Rinderle)
- Workflow evolution (Casati; Ceri; Pernici)

Among the other cited articles there are a significant number of further articles by Reichert/Rinderle.

Compliance Cluster
The second cluster is about the intended topic of the whole network: Compliance of business processes. The most common keywords in the titles of the articles are:

- Compliance
- Compliance Management
- Business Process Compliance

- Compliance Governance
- Compliance Verification
- Framework
- Implementation

The most frequently cited articles are:

- A static compliance-checking framework for business process models (Liu; Muller)
- Modeling control objectives for business process compliance (Sadiq; Governatori)
- Compliance checking between business processes and business contracts (Governatori; Milosevic)

Semantics Cluster
This cluster is about the semantics of business processes models. Its main keywords are:

- Semantic (in combination with other keywords such as Business Process Management, Event-driven Process Chains, Process Modeling and Benchmarking of Process Models)
- Ontology
- Integration
- Composition
- Design/Redesign
- Web Services

The most frequently cited articles are:

- Semantic business process management: A vision towards using semantic web services for business process management (Hepp; Leymann; Domingue)
- An ontology framework for semantic business process management (Hepp)
- Generation of business process models for object life cycle compliance (Küster; Ryndina)

Modeling Cluster
This cluster is about the modeling of business processes. Main keywords are:

- Process Modeling Grammars
- Process Patterns
- Business Process Documentation
- Process Modeling Methodology
- Collaborative Business Process Modeling

The most frequently cited articles are:

- Business process modeling-a comparative analysis (Recker; Rosemann)
- Factors and measurs of business process modelling: model building through a multiple case study (Bandara; Gable)
- Business process modeling: Perceived benefits (Indulska; Green; Recker)

Process Change Cluster
In this cluster the articles are focused on process change topics. The most common keywords are:

- (Business) Process Change
- (Business) Process Redesign
- e-Government
- ERP

The most frequently cited articles are:

- Special section: toward a a theory of business process change management (Kettinger)
- Business process change and organizational performance: exploring an antecedent model (Guha; Grover; Kettinger)
- Developing strategic perspectives on business process reengineering: from process reconfiguration to organizational change (Teng; Grover)

Most of the cited articles have been published in management journals. This indicates, that the cluster is more focused on business aspects than on IT aspects.

Views Cluster
The smallest cluster is about workflow views. The relevant keywords are:

- View
- View-based
- Workflows

The most frequently cited articles are:

- Workflow view based e-contracts in cross-organizational e-services environment (Chiu; Karlapalem; Li)
- Workflow view driven cross-organizational interoperability in a web service environment (Chiu; Cheung; Till; Karlapalem)
- WW-Flow: Web based workflow management with runtime encapsulation (Y Kim; Kang; D Kim; Bae)

5.2.5 Mobile Processes

In this chapter we will look at the results generated on basis of the mobile processes search term. The search term was: *mobile "business process" OR workflow*. The cited by network of this search term can be seen in the following figure:

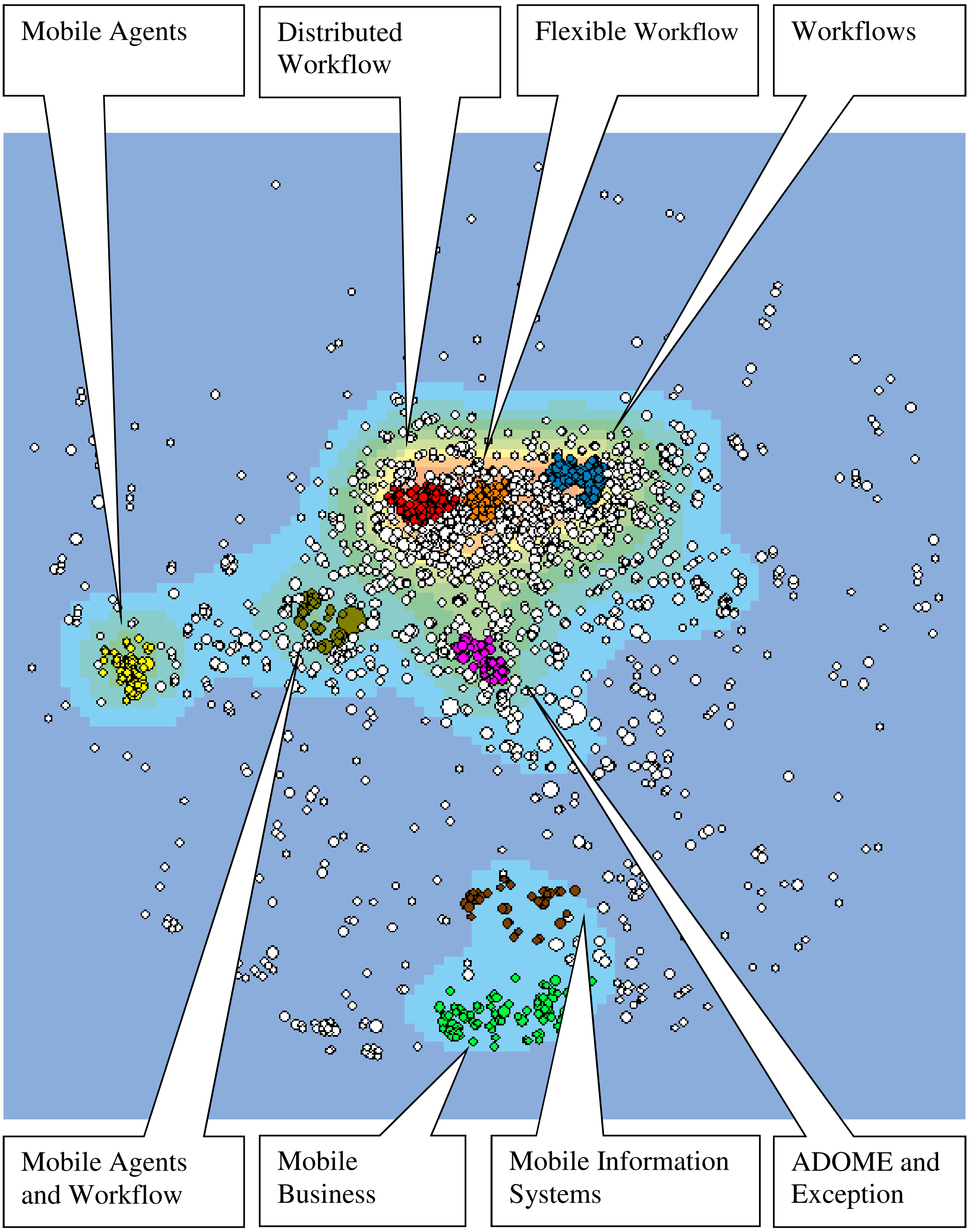


Figure 19: Cited by network of the mobile processes search term

Table 17: Numbers of the mobile processes search and of the corresponding network

Number of search results:	61,100
Number of cited bys (includes patents) of the first thousand articles:	16,240
Number of terms in the graph:	2,052

The following clusters have been identified:

Table 18: Clusters in the cited by network for the mobile processes search term

Name of the cluster	**Number of articles within the cluster**
Distributed Workflow	108
Mobile Agents	87
Mobile Business	86
Workflows	85
ADOME and Exception	52
Flexible Workflow	51
Mobile Information Systems	40
Mobile Agents and Workflow	37

Distributed Workflow Cluster
The biggest cluster is about distributed workflows. The most frequent keywords are:

- Distributed
- Cross-organizational
- Mobile Environments
- Agile
- Workflow Migration
- Knowledge
- ADEPT
- Adaptive

The most frequently cited articles are:

- Functionality and limitations of current workflow management systems (Alonso; Agrawal; Abbadi)
- INCAs: Managing dynamic workflows in distributed environments (Barbara; Mehrotra)
- From centralized workflow specification to distributed workflow execution (Muth; Wodtke; Weissenfels; Dittrich)

Mobile Agents Cluster
This cluster is about the concept of mobile agents. The most common keywords are:

- Mobile Agents
- Mobile Computing
- Intelligent Agent
- Intelligent Systems
- Distributed
- Architecture
- Applications

The most frequently cited articles are:

- Mobile Agents: Are they a good idea? (Chess; Harrison)
- Agent Tcl: A flexible and secure mobile-agent system (Gray)
- Seven good reasons for mobile agents (Lange; Oshima)

Mobile Business Cluster
The topics of this cluster are mobile business and mobile commerce. The most common keywords in the titles of the articles are:

- Mobile Phones
- Mobile Processes
- Mobile Services
- Mobile Work Context
- Mobile User Interface in BPM
- m-Business
- m-Health

The most frequently cited articles are:

- Business models and transactions in mobile electronic commerce: requirements and properties (Tsalgatidou; Pitoura)
- Introduction to the special issue: mobile commerce applications (Liang)
- Development perspectives, firm strategies and applications in mobile commerce (Buellingen)

Workflow Cluster
This cluster is about workflows and business processes in general. Different topics such as process/workflow mining, process flexibility and process verification are mentioned. The corresponding keywords are:

- Mining
- Pattern

- Analysis
- Flexible/flexibility
- Change
- Discovering
- Verifying/verification

The most frequently cited articles are:

- Workflow management: modeling concepts, architecture and implementation (Jablonski)
- Workflow patterns (van der Aalst; ter Hofstede)
- YAWL: yet another workflow language (van der Aalst)

ADOME and Exception Cluster
This cluster is about exception handling in combination with the ADOME workflow management system. Frequent keywords are:

- ADOME
- Exception Handling
- e-Service
- Services
- Workflows

The most frequently cited articles are:

- Web interface-driven cooperative exception handling in adome workflow management system (Chiu; Li)
- A meta modeling approach to workflow management systems supporting exception handling (Chiu; Li)
- Workflow view-based e-contracts in a cross-organizational e-services environment (Chiu; Karlapalem; Li)

Flexibility Cluster
This cluster is about adaptivity and flexibility in workflow management. The most common keywords in the titles of the articles are:

- Workflow
- Flexible
- Changes
- Adaptation
- Exception Handling

The most frequently cited articles are:

- A taxonomy of adaptive workflow management (Han; Sheth)
- A comprehensive approach to flexibility in workflow management systems (Heinl; Horn; Jablonski; Neeb)
- A framework for dynamic changes in workflow management systems (Reichert)

Mobile Information Systems Cluster
This cluster is about mobile information systems, particularly mobile information systems in hospitals. The most common keywords in the titles of the articles are:

- Computer Use by Doctors and Nurses
- Clinical Decisions Support Systems
- Healthcare
- Hospital
- Mobile Information

The most frequently cited articles are:

- Mobile information and communication tools in the hospital (Ammenwerth; Buchauer; Bludau)
- Research areas and challenges for mobile information systems (Krogstie; Lyytinen; Opdahl)
- Requirement engineering for mobile information systems (Krogstie)

Mobile Agents and Workflow Cluster
In this cluster, the topics are workflows in combination with agents and distributed workflows. The main keywords are:

- Inter-organizational Workflows
- Agent-based Workflow
- Distributed Workflow/Distributed Processes
- Agent Technology
- Web Services

The most frequently cited articles:

- Using mobile agents to support interorganizational workflow management (Merz; Liberman)
- Agent-based workflow: TRP support environment (TSE)
- Decentralized and flexible workflow enactment based on task coordination agents (Joeris)

Most of the cited articles in the Mobile Agents and Workflow Cluster are already a bit older, for example, the three articles mentioned are from 1997, 1996 and 2000. This might indicate that the topic of the cluster has not received a lot of attention in the last couple of years.

5.3 Analysis of Fields Related to BPM

In this chapter we will analyze four fields that are related to business process management. Those fields are as follows:

- Business Intelligence
- ERP
- Service-oriented Architecture
- Knowledge Management

For a description of these fields, see Chapter 2.4.

5.3.1 Business Intelligence

For the analysis of the business intelligence field I chose the search term *business intelligence*, restricted to the two fields “Business, Administration, Finance, and Economics” and “Engineering, Computer Science, and Mathematics”. The cited by network based on that search term can be seen in the following figure:

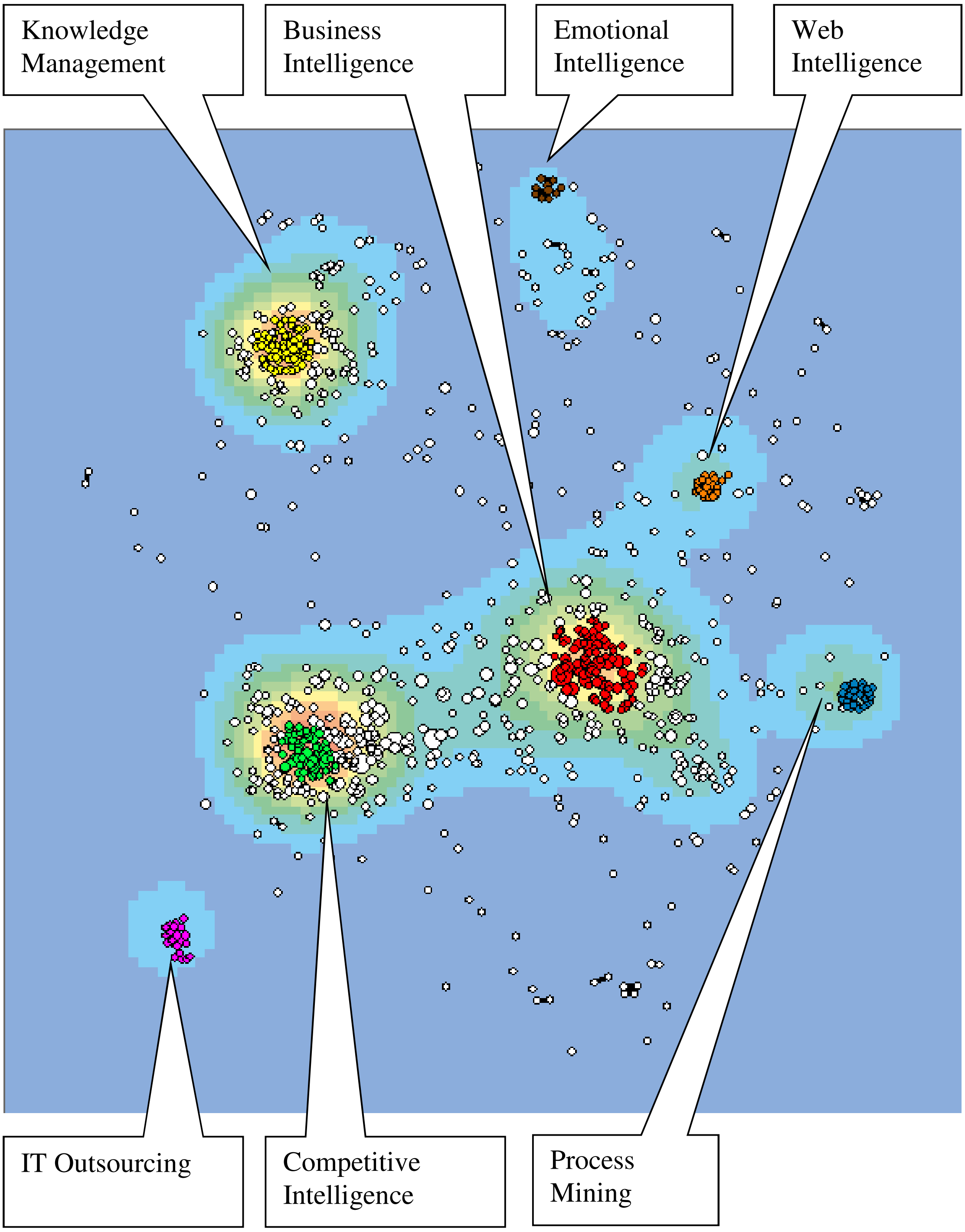


Figure 20: Cited by network of the business intelligence search term

Table 19: Numbers of the business intelligence search and of the corresponding network

Number of search results:	667,000
Number of cited bys (includes patents) of the first thousand articles:	34,023
Number of terms in the graph:	958

In the this network, we can identify a total of seven clusters:

Table 20: Clusters in the cited by network for the business intelligence search term

Name of the cluster	Number of articles within the cluster
Business Intelligence	87
Knowledge Management	70
Competitive Intelligence	58
Process Mining	32
IT Outsourcing	21
Web Intelligence	21
Emotional Intelligence	10

Business Intelligence Cluster
The biggest cluster is about business intelligence itself. The most common keywords in the titles of the articles are:

- Business Intelligence
- Business Analytics
- Data Mining
- Data Warehousing
- Supply Chain/Distribution Chain

The most frequently cited articles are:

- Business Intelligence (Negash)
- Beyond data warehousing: what's next in business intelligence? (Golfarelli; Rizzi)
- Business intelligence roadmap: the complete project lifecycle for decision-support applications (Moss)

Knowledge Management Cluster
The next cluster is about knowledge management and particularly the implementation of knowledge management systems. The most common keywords are:

- Knowledge Management
- Knowledge Management Implementation
- Critical Success Factors

The most frequently cited articles are:

- The knowledge agenda (Skyrme)
- Strategies for implementing knowledge management: role of human resources management (Soliman)
- Excellence in knowledge management: an empirical study to identify critical factors and performance measures (Chourides; Longbottom)

It is noticeable, that many of the cited articles are published in the Journals of Knowledge Management.

Competitive Intelligence Cluster
The third cluster is about competitive intelligence. The main keywords are:

- Competitive Intelligence
- Information
- Scanning

The most frequently cited articles are:

- The new competitor intelligence: the complete resource for finding, analyzing, and using information about your competitors (Fuld)
- Key intelligence topics: a process to identify and define intelligence needs (Herring)
- The business intelligence system: A new tool for competitive advantage (Gilad)

It is noticeable that there are quite a lot of Spanish and Portuguese-language articles in the cluster.

Process Mining Cluster
The fourth cluster is about processes and particularly about process mining. The most common keywords are:

- Discovering
- Mining
- Conformance Checking
- Conformance Testing

The most frequently cited articles are:

- Workflow mining: a survey of issues and approaches (Aalst; Dongen; Herbst)
- Business process cockpit (Sayal; Casati; Dayal)
- Business process intelligence (Grigori; Casati; Castellanos; Dayal)

IT Outsourcing Cluster
This cluster that is a bit separate from the others is about IT outsourcing. The most common keywords are:

- IT Outsourcing
- Information Systems Outsourcing

The most frequently cited articles are:

- Information technology outsourcing in Europe and the USA: Assessment issues (Willcocks; Lacity)
- Co-operative partnership and total IT outsourcing: From contractual obligation to strategic alliance? (Willcocks)
- Contracts and partnerships in the outsourcing of IT (Fitzgerald)

Web Intelligence Cluster
The smallest cluster is about web intelligence. The relevant keywords are:

- Web Intelligence
- Wisdom Web

The most frequently cited articles are:

- Web Intelligence (WI) Research Challenges and Trends in the New Information Age (Yao; Zhong)
- Web intelligence (Zhong; Liu)
- Web intelligence (WI): What makes wisdom web? (Liu)

Emotional Intelligence Cluster
The smallest cluster is about the topic of emotional intelligence in the context of business. The relevant keywords are:

- Emotional Intelligence
- Leadership
- Manager/Managerial

The most frequently cited articles are:

- Transformational leadership and emotional intelligence: An exploratory study (Barling; Slater)
- Emotional intelligence and its relationship to workplace performance outcomes of leadership effectiveness (Rosete)
- Linking emotional intelligence abilities and transformational leadership styles (Leban)

5.3.2 ERP

In order to cover the relations between ERP systems and BPM the following search term has been used: *ERP business process management*.
The following figure shows the graph created on the basis of that search term.

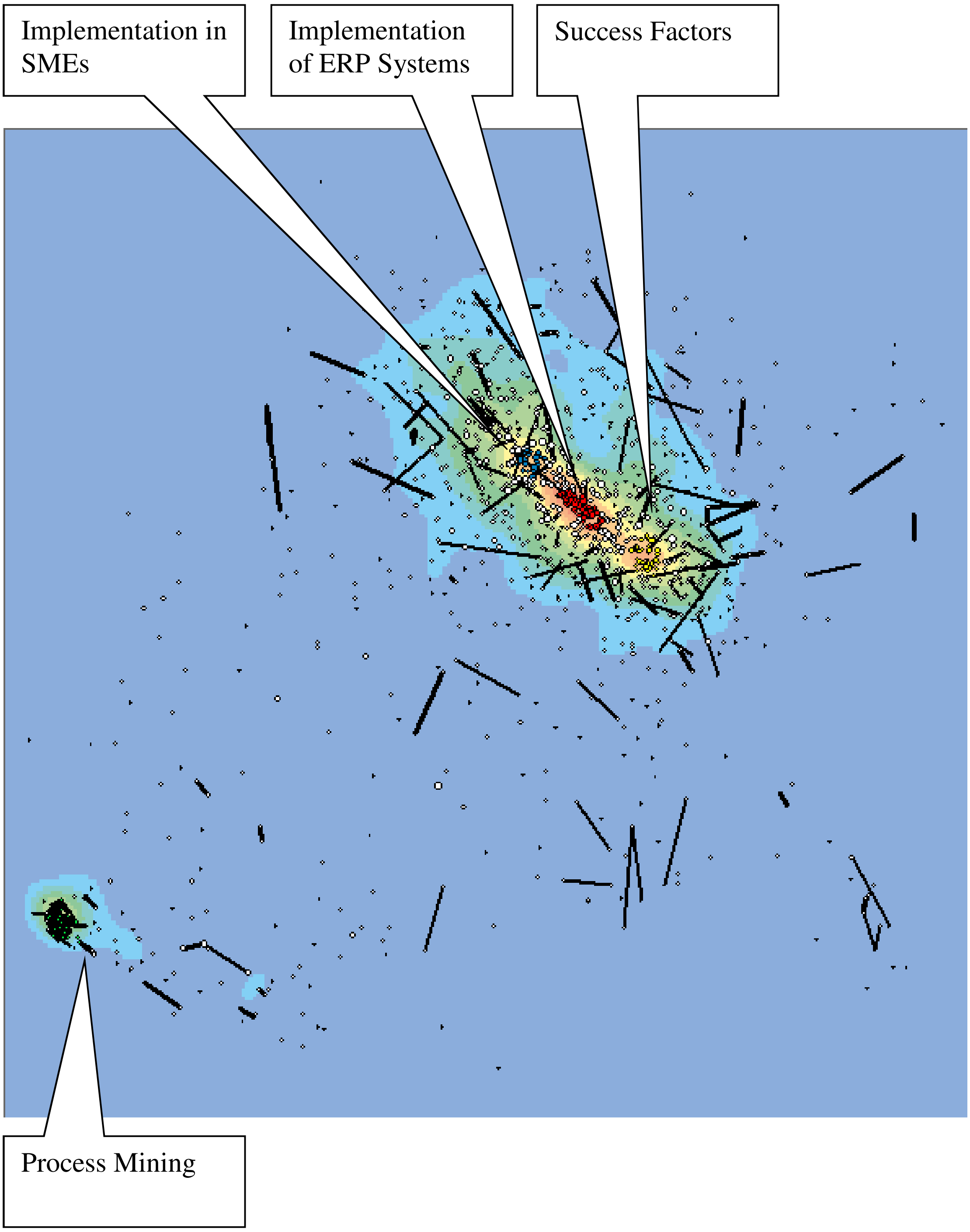


Figure 21: Cited by network of the ERP search term

Table 21: Numbers of the ERP search and of the corresponding network

Number of search results:	64,100
Number of cited bys (includes patents) of the first thousand articles:	48,568
Number of terms in the graph:	1,213

The following clusters have been identified:

Table 22: Clusters in the cited by network for the ERP search term

Name of the cluster	Number of articles within the cluster
Implementation of ERP Systems	44
Success Factors	33
Process Mining	31
Implementation in SMEs	21

There are two clusters of which the first one can be split into three subclusters.

Implementation of ERP Systems Cluster
The biggest cluster is about the implementation of ERP systems in general. The most common keywords are:

- Implementation
- Key Factors
- Critical Success Factors

The most frequently cited articles are:

- Enterprise resource planning: A taxonomy of critical factors (Al-Mashari; Al-Mudimigh)
- Enterprise resource planning: Managing the implementation process (Mabert; Soni)
- ERP selection process in midsize and large organisations (Bernroider)

Success Factors Cluster
This cluster is about success factors of ERP implementation. The most common keywords are:

- Implementation
- Critical Success Factor
- Organizational Factors
- Institutional Forces.

The most frequently cited articles are:

- The impact of critical success factors across the stages of enterprise resource planning implementations (Somers)
- Enterprise resource planning: multisite ERP implementations (Markus; Tanis)

- Critical issues affecting an ERP implementation (Bingi; Sharma)

Process Mining Cluster
This cluster is about the topic of Process Mining. Frequent key words in that cluster include:

- Mining
- Discovering
- Event Logs

The most frequently cited articles are:

- Conformance testing: Measuring the fit and appropriateness of event logs and process models (Rozinat)
- Workflow mining: Discovering process models from event logs (van der Aalst; Weijters)
- Workflow mining: a survey of issues and approaches (van der Aalst; van Dongen; Herbst)

Implementation in SMEs Cluster
The third sub-cluster as well focuses on implementation, particularly in small and medium-sized enterprises (SMEs) and the additional key words include:

- Small and Medium-sized Enterprises
- Different types of organizations (universities and government organizations)
- Different countries (India, China and Thailand).

The most frequently cited articles are:

- A framework of ERP systems implementation success in China: An empirical study (Z Zhang; Lee; Huang; L Zhang)
- Enterprise information systems project implementation: A case study of ERP in Rolls-Royce (Yusuf; Gunasekaran)
- Enterprise resource planning: An integrative review (Shehab; Sharp)

5.3.3 Knowledge Management

In order to analyze the relations between knowledge management and BPM, the following search has been performed: *"business process management" OR "workflow management" "knowledge management"*. The quotation marks are necessary to maintain the OR-structure in the intended way and in order to avoid articles that simply contain the word knowledge at some point in the article. Below, we will analyze the cited by network created on basis of this search term. The cited by network and the identified clusters can be seen in the following figure:

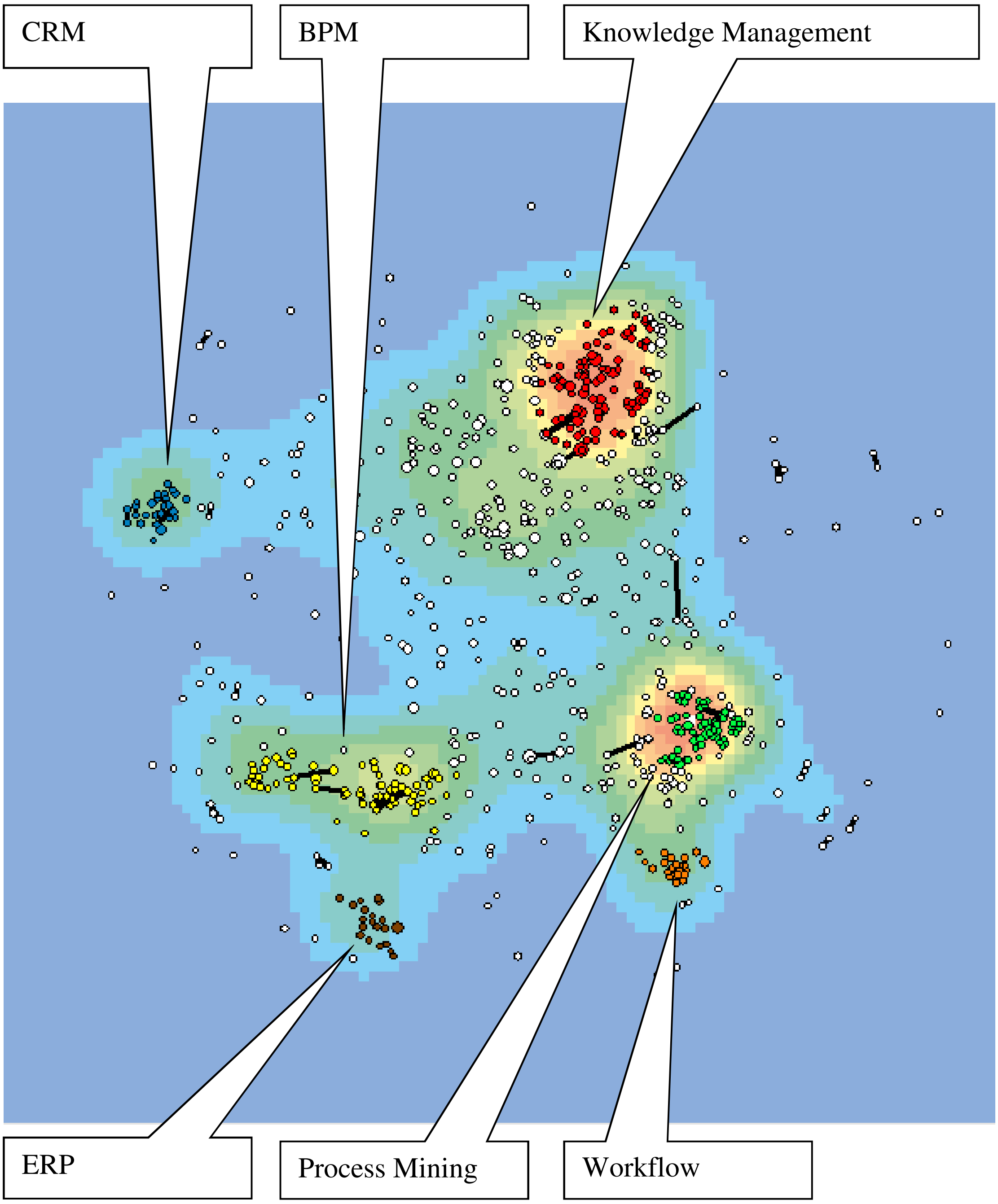


Figure 22: Cited by network of the knowledge management/BPM search term

Table 23: Numbers of the knowledge management search and of the corresponding network

Number of search results:	12,200
Number of cited bys (includes patents) of the first thousand articles:	17,194
Number of terms in the graph:	750

The following clusters have been identified:

Table 24: Clusters in the cited by network for the knowledge management search term

Name of the cluster	Number of articles within the cluster
Knowledge management	89
BPM	71
Process Mining	65
CRM	26
Workflow management	23
ERP	20

Knowledge Management Cluster

The biggest cluster is the knowledge management cluster. The most commom terms are:

- Knowledge Management
- Experience Management
- Business-process-oriented Knowledge Management
- Process-based Knowledge Management
- Business Decision Processes
- Knowledge-intensive Business Processes (or „wissensintensive Geschäftsprozesse“ in German)

The most commonly cited articles are:

- Information supply for business processes: coupling workflow with document analysis and information retrieval (Abecker et al).
- An organizational-memory-based approach for an evolutionary workflow management system-concepts and implementation (Wargitsch, Wewers)
- Context Framework-an Open Approach to Enhance Organisational Memory Systems with Context Modelling Techniques (Klemke)

Among the cited articles other topics are: integration of knowledge and business processes, organizational memories and the support of business processes through knowledge-based systems.

BPM Cluster
The second biggest cluster is the BPM cluster. The most frequent keywords are:

- Process Modeling
- Evaluation
- Maturity

In addition to this several general BPM topics are mentioned. The most frequently cited articles are:

- Towards a business process management maturity model (Rosemann)
- Factors and measures of business process modeling: model building through a multiple case study (Bandara, Gable)
- Potential pitfalls in process modeling: part A (Rosemann).

The most common publications among the cited articles is the Business Process Management Journal.

Process Mining Cluster
The next cluster is again a process mining cluster. Frequent terms include:

- Process Mining
- Workflow Mining
- Event Logs

The most frequently cited articles are:

- Integrating machine learning and workflow management to support acquisition and adaptation of workflow models (Herbst)
- Workflow mining: a survey of issues and approaches (Aalst; Dongen; Herbst)
- Process miner—a tool for mining process schemes from event-based data (Schimm)

CRM Cluster
This cluster focuses on Customer Relationship Management or CRM. Frequent keywords include:

- CRM
- Customer Information
- Customer Knowledge

A process-related term found in the cluster is people-driven processes in CRM. One of the most frequently cited article is also process-related. The most frequently cited articles are:

- Improving performance of customer-processes with knowledge management (Bueren; Schierholz; Kolbe)
- CRM and customer-centric knowledge management: an empirical research (Stefanou; Sarmaniotis)
- Mobilizing customer relationship management: A journey from strategy to system design (Schierholz; Kolbe)

Workflow Management Cluster

This cluster focuses on workflow management topic. In this cluster frequent keywords are:

- Workflow
- Process
- Web Service
- e-Service
- Information

The most frequently cited articles are:

- A meta modeling approach to workflow management systems supporting exception handling (Chiu; Li)
- Web interface-driven cooperative exception handling in adome workflow management system (Chiu; Li)
- DartFlow: A workflow management system on the web using transportable agents (Cai; Gloor)

ERP Cluster

The last cluster is about ERP systems. The most common keywords are:

- ERP
- Implementation

The most frequently cited articles are:

- Enterprise resource planning: A taxonomy of critical factors (Al-Mashari; Al-Mudimigh)
- Enterprise resource planning in engineering business (Sirigindi
- Enterprise resource planning: An integrative review (Shehab; Sharp)

5.3.4 SOA

In order to cover the field of service-oriented architecture (SOA) and service-oriented computing (SOC) we used the search term: *"service oriented" architecture OR computing*. The graph based on the cited by field of those results can be seen in the following figure:

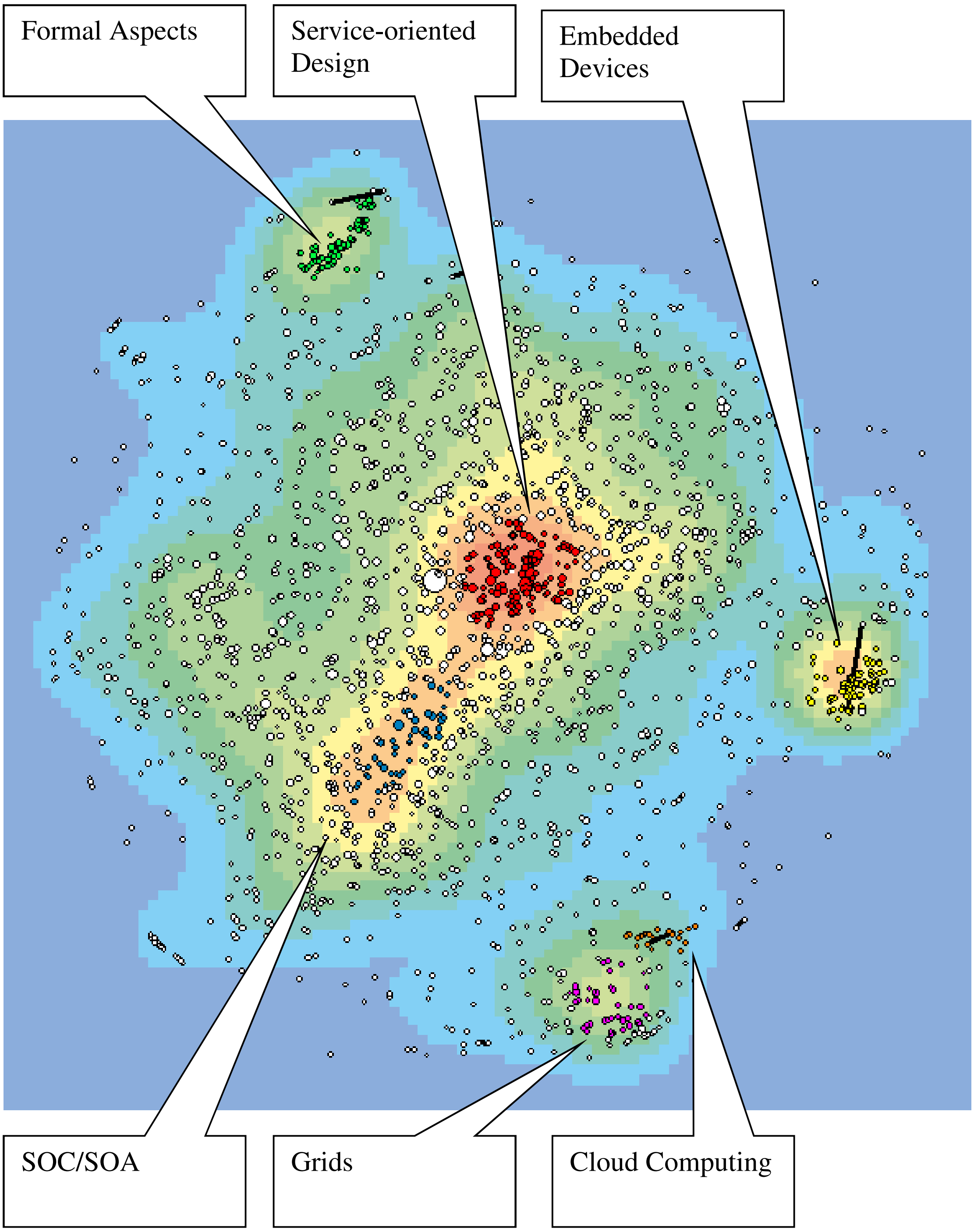


Figure 23: Cited by network of the SOA search term

Table 25: Numbers of the SOA search and of the corresponding network

Number of search results:	113,000
Number of cited bys (includes patents) of the first thousand articles:	32,879
Number of terms in the graph:	2,130

In the following table the identified clusters are presented.

Table 26: Clusters in the cited by network for the SOA search term

Name of the cluster	Number of articles within the cluster
Service-oriented Design	113
Embedded Devices	81
Formal Aspects	64
SOC/SOA	47
Grids	41
Cloud Computing	21

As follows, a description of the identified clusters:

Service-oriented Design Cluster
This cluster is about service-oriented modeling and design. The following keywords are the most relevant:

- Modeling
- SOA
- Service Identification
- Framework
- Development
- Model-driven Service Engineering

The most frequently cited articles are:

- Elements of service-oriented analysis and design (Zimmermann; Krogdahl)
- Service-oriented modeling and architecture (Arsanjani)
- Service-oriented design and development methodology (Papazoglou)

Embedded Devices Cluster
This cluster is about embedded devices in SOA environments. In this cluster the following keywords are common:

- Real-time Control
- Control System
- SOA-ready Device
- Industrial Automation
- SOA-based Devices

- Device Integration
- SOA for Devices
- Embedded Devices
- SOAP4IPC
- SOAP4PLC

The most frequently cited articles are:

- Service-oriented device communications using the devices profile for web services (Jammes; Mensch)
- SIRENA-Service Infrastructure for Real-time Embedded Networked Devices: A service oriented framework for different domains (Hohn; Bobek)
- Integration of soa-ready networked embedded devices in enterprise systems via a cross-layered web service infrastructure (Karnouskos; Baecker)

Formal Aspects Cluster
This cluster is about formal aspects of SOA. The most common keywords are

- Formal Approach
- Formal Framework
- Formal Model
- Specification
- Verification
- Semantics
- Formal Basis

The most frequently cited articles are:

- A formal approach to service component architecture (Fiadeiro; Lopes)
- Disciplining orchestration and conversation in service-oriented computing (Lanese; Vasconcelos)
- SOCK: a calculus for service oriented computing (Guidi; Lucchi; Gorrieri; Busi)

SOC/SOA Cluster
This cluster is about general SOC/SOA topics. The most common keywords are:

- Service Component Integration
- Semantic Web Services
- Quality of Service
- SOA
- Web Services

The most frequently cited articles are:

- Service-oriented computing (Bichier)
- Current solutions for web service composition (Milanovic)
- Service-oriented computing: Concepts, characteristics and directions (Papazoglou)

Grids Cluster
This cluster is about the topic of grids again. The most common keywords are:

- Grid (in combination with the keywords Computing, Economics and Technologies)
- Cloud Computing
- Service/Service-oriented

The most frequently cited articles are:

- The gridbus toolkit for service oriented grid and utility computing: An overview and status report (Buyya)
- Virtual workspaces: Achieving quality of service and quality of life in the grid (Keahey; Foster; Freeman)
- A computational economy for grid computing and its implementation in the Nimrod-G resource broker (Abramson; Buyya)

Cloud Computing Cluster
This cluster is about the area of cloud computing. The most common keywords are:

- Cloud Computing
- Business Collaboration
- Service-oriented
- Grid

The most frequently cited articles are:

- Cloud computing and grid computing 360-degree compared (Foster; Zhao; Raicu)
- Toward a unified ontology of cloud computing (Youseff; Butrico)
- Cloud computing-Issues, research and implementations (Vouk)

5.4 Analysis of one Journal and two Conferences

In this subchapter I will have a look at one journal related to BPM and two conferences related to BPM.

The journal is the Data & Knowledge Engineering Journal from Elsevier. Its topics include "[a]rchitectures of database, expert, or knowledge-based systems" and "[a]pplications, case studies, and management issues"[73]. BPM is also an important topic for this journal, for example it also publishes papers from BPM conferences. For my analysis I searched for all articles in that journal that contain the term process or processes.

The first conference I will look at is the International Conference on Business Process Management. Its topics are, as also the name suggests, "all aspects of BPM"[74]. In 2011 this conference has been held for the ninth time.

The second conference is the Conference on Advanced Information Systems Engineering, also known as CAiSE. In 2011 it was held for the 23rd time. Other than the International Conference on Business Process Management the CAiSE conference is not only about BPM. Its topics include enterprise architectures, services and processes.[75] In my search, I will concentrate on the process-related articles.

The importance of the journal and the two conferences for the BPM field has also been confirmed by Professor Reichert. See the interview in Appendix V for his complete statement.

For the journal and the two conferences I each created the cited by network to identify clusters. In case of the Data & Knowledge Engineering Journal I used a search term limiting the result articles to articles published in that journal. In case of the two conferences I used a search that includes articles that contain the name of the conference in the article, because otherwise the number of result articles would not have been high enough. For each of the three search terms, I created the cited by network in order to identify clusters. In case of the CAiSE search term, I also created the inverted cited by network in order to compare the research fronts from the cited by network with the knowledge bases from the inverted cited by network.

[73] See also Elsevier: Data & Knowledge Engineering, URL: http://www.journals.elsevier.com/data-and-knowledge-engineering/, accessed February 2, 2012, for a full list of their topics.

[74] BPM 2011 9th International Conference on Business Process Management: Welcome to BPM 2011, URL: http://bpm2011.isima.fr/, accessed February 2, 2012

[75] CAiSE '11 - 23rd International Conference on Advanced Information Systems Engineering: Call for Papers, URL: http://www.caise2011.com/callForPapers.php, accessed February 2, 2012

5.4.1 Data & Knowledge Engineering Journal

In order to analyze the process-related articles from the Data & Knowledge Engineering Journal the search term *process OR processes* has been used. The search has been limited to the articles from the Data & Knowledge Engineering Journal.

Again we will analyze the cited by network and this network and the identified clusters can be seen in the following figure:

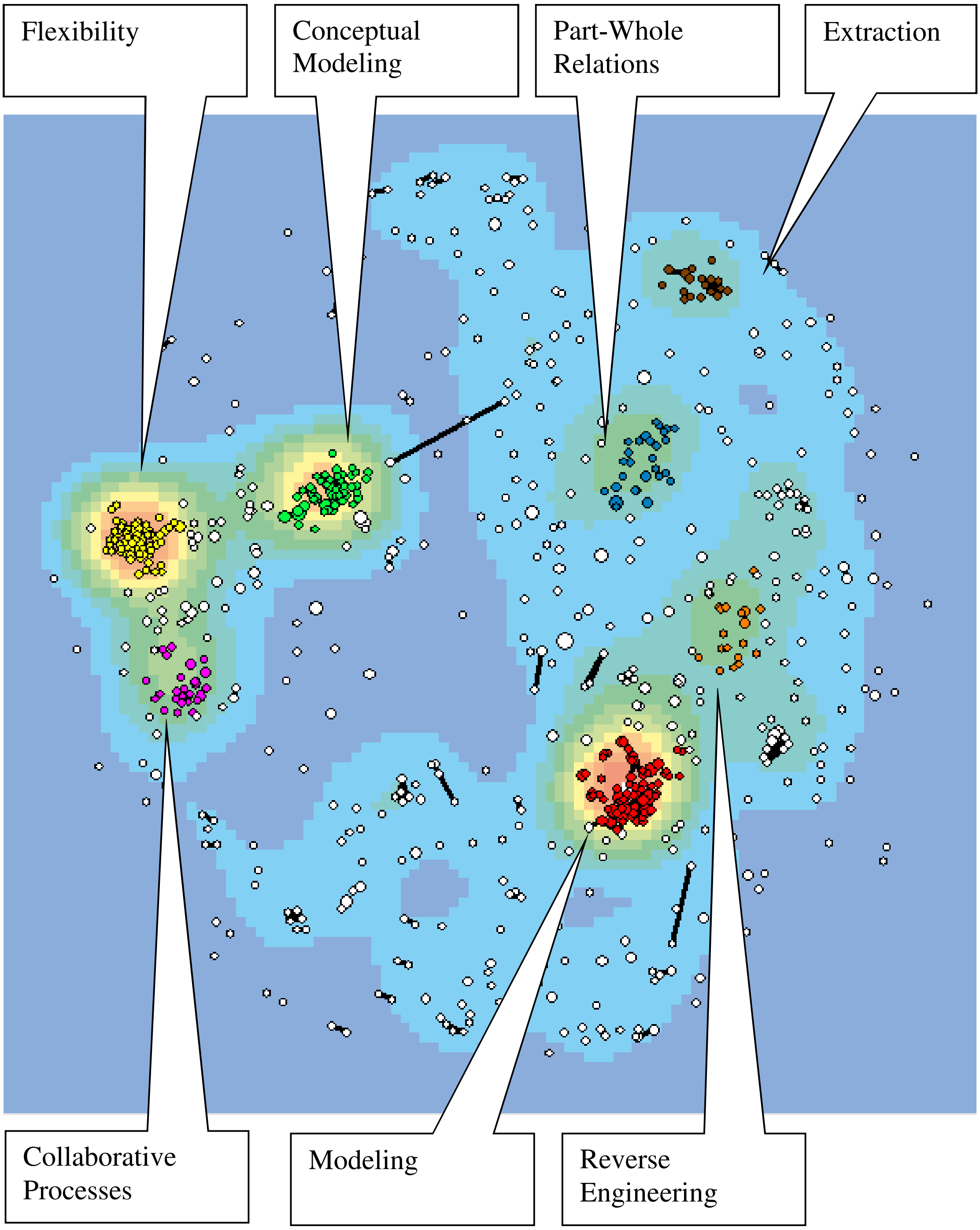


Figure 24: Cited by network of the Data & Knowledge Engineering Journal search term

Table 27: Numbers of the Data & Knowledge Engineering Journal search and of the corresponding network

Number of search results:	3,510
Number of cited bys (includes patents) of the first thousand articles:	23,566
Number of terms in the graph:	730

In the following table we can see the names of the clusters and the numbers of articles within the clusters:

Table 28: Clusters in the cited by network for Data & Knowledge Engineering Journal search term

Name of the cluster	**Number of articles within the cluster**
Modeling	74
Flexibility	63
Conceptual Modeling	53
Part-whole Relations	28
Collaborative Processes	25
Reverse Engineering	18
Extraction	18

Now, we will have a closer look at the articles in the identified clusters:

Modeling Cluster
The biggest cluster focuses on various forms of modeling. Frequent keywords are:

- Data Modeling/Data Models
- Application Models
- Conceptual Modeling
- Workflow Modeling
- Object Role Modeling
- Unified Modeling Language
- Object Relationship Mapping
- Rule Modeling
- Information Modeling
- Databases

Three most frequently cited articles are:

- Expressiveness in conceptual data modeling (ter Hofstede)
- Subtyping and polymorphism in object-role modeling (Halpin)
- Conceptual modelling of database applications using an extended ER model (Engels; Gogolla; Hohenstein)

Flexibility Cluster
The second biggest cluster is about the topics of process flexibility and process change. Frequent keywords are:

- Flexible/Flexibility
- Change Support
- Constraints
- Dynamic
- Process-aware Information Systems
- Variability
- Change Patterns
- ADEPT
- Process Mining

The most frequently cited articles in this cluster are:

- Correctness criteria for dynamic changes in workflow systems--a survey (Reichert; Rinderle)
- Case handling a new paradigm for business process support (van der Aalst; Weske)
- IT support for healthcare processes-premises, challenges perspectives (Lenz)

Conceptual Modeling Cluster
This cluster covers topics about modeling and particularly conceptual modeling. Frequent keywords are:

- Conceptual Modeling
- Process Modeling
- Process Models

Also mentioned are BPMN and BPEL.

The most frequently cited articles are:

- Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions (Moody)
- How do practitioners use conceptual modeling in practice? (Davies; Green; Rosemann; Indulska)
- Complexity and clarity in conceptual modeling: comparison of mandatory and optional properties (Gemino)

Part-Whole Relations Cluster
This cluster is about so-called part-whole relations. Frequent keywords are:

- Data Darehouse
- Part-Whole Relations
- Parthood
- Ontology

The articles that are cited most frequently are:

- Part-whole relations in object-centered systems: An overview (Artale)
- Parts, wholes and part-whole relations: The prospects of mereotopology (Varzi)
- A conceptual theory of part-whole relations and its applications (Gerstl)

Collaborative Processes Cluster
This cluster is about collaborative and cross-organizational workflows. Frequent keywords are:

- Cross-organizational workflow
- Collaborative BPM
- Service outsourcing
- Collaboration between business processes

The most commonly cited articles within the clusters are:

- Facilitating cross-organisational workflows with a workflow view approach (Schulz)
- The view-based approach to dynamic inter-organizational workflow cooperation (Chebbi; Dustdar)
- Constructing customized process views (Eshuis)

Reverse Engineering Cluster
This cluster is about reverse engineering in connection with databases. The most common keywords are:

- Databases
- (Database) Reverse-Engineering.
- Extraction/Extracting

The articles that are cited most frequently are:

- Reverse engineering of relational databases: Extraction of an EER model from a relational database (Chiang; Barron)

- Database reverse engineering: from the relational to the binary relationship model (Shoval)
- A survey of database design transformations based on the Entity-Relationship model (Fahrner)

Extraction Cluster
The smallest cluster is about the topic of extracting data from the web. Common keywords are:

- Extraction
- Web of Knowledge
- Ontology

The most frequently cited articles are:

- Conceptual-model-based data extraction from multiple-record Web pages (Embley; Campbell; Jiang)
- DEByE-data extraction by example (Laender; Ribeiro-Neto)
- Building intelligent web applications using lightweight wrappers (Sahuguet)

5.4.2 International Conference on Business Process Management

For the analysis of articles related to the International Conference on Business Process Management (ICOBPM) the following search term has been used: *"international conference on business process management""* The following cited by network was created on basis of the data acquired with that search term.

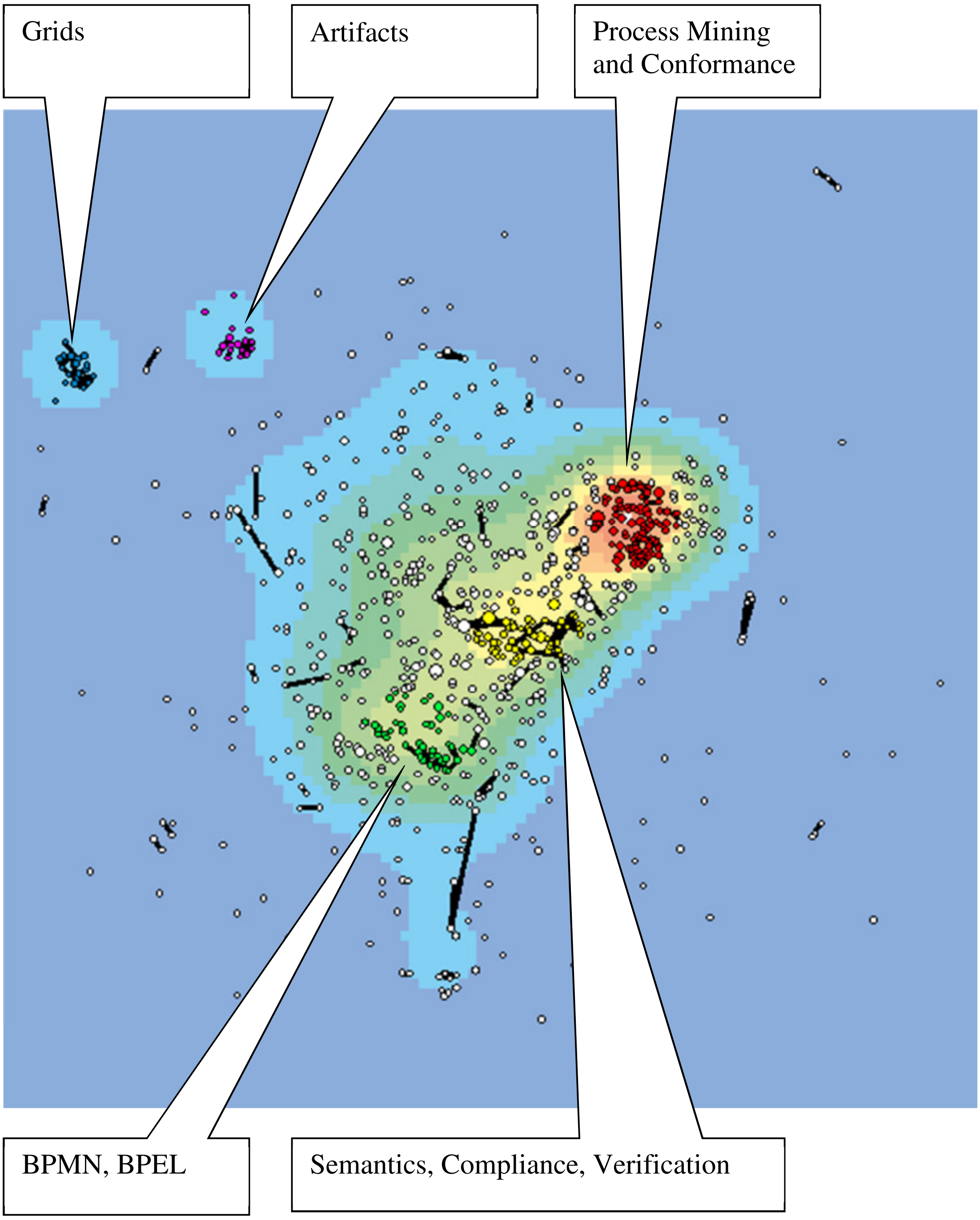


Figure 25: Cited by network of the BPM Conference search term

Table 29: Numbers of the BPM Conference search and of the corresponding network

Number of search results:	1,340
Number of cited bys (includes patents) of the first thousand articles:	9,683
Number of terms in the graph:	823

The following clusters have been identified:

Table 30: Clusters in the cited by network for the BPM Conference search term

Name of the cluster	**Number of articles within the cluster**
Process Mining and Conformance	97
Semantics, Compliance, Verification	54
BPMN, BPEL	44
Grids	19
Artifacts	17

Process Mining and Conformance Cluster

The biggest cluster is about process mining and conformance. The most common keywords are:

- Process Mining
- Discovery
- Conformance
- Compliance Checking
- Conformance Analysis
- Process Conformance
- Change Mining

The most frequently cited articles are:

- Conformance testing: Measuring the fit and appropriateness of event logs and process models (Rozinat)
- Discovering social networks from event logs (van der Aalst; Reijers)
- Conformance checking of processes based on monitoring real behavior (Rozinat)

Semantics, Compliance, Verification Cluster

This cluster is about the semantics, compliance and verification of processes. Main keywords include:

- Metrics
- Semantic Methods
- Verification
- Compliance
- Validation

- Formal Semantics

The most frequently cited articles are:

- Efficient compliance checking using bpmn-q and temporal logic (Awad; Decker)
- Integration and verification of semantic constraints in adaptive process management systems (Ly; Rinderle; Dadam)
- A static compliance-checking framework for business process models (Liu; Muller)

BPMN, BPEL cluster
This cluster is about BPMN and BPEL. The most common keywords are:

- BPMN
- BPEL
- Service Choreography
- Transform/Transformation
- Translating/Translation
- Models
- Modeling

The most frequently cited articles are:

- Translating standard process models to BPEL (Ouyang; Dumas; Breutel)
- Pattern-based translation of BPMN process models to BPEL web services (Ouyang; Dumas; ter Hofstede)
- From BPMN process models to BPEL web sercives (Ouyang; Dumas; ter Hofstede)

Other cited articles cover many further BPMN- and BPEL-related topics.

Grids Cluster
This cluster is about grids and scientific workflows. The most common keywords are:

- Grid Workflows
- Scientific Workflows
- Cloud Workflow
- Grid Environment

The most frequently cited articles are:

- Multiple states based temporal consistency for dynamic verification of fixed-time constraints in grid workflow systems (Chen)
- Adaptive selection of necessary and sufficient checkpoints for dynamic verification of temporal constraints in grid workflow systems (Chen)
- A taxnomy of grid workflow verification and validation (Chen)

Artifact Cluster

The smallest cluster is about artifacts and artifact-centric processes. The main keywords are:

- Artifacts
- Artifact-based
- Artifact-centric
- Data-centric
- XML/AXML

The most frequently cited articles:

- Artifact-centric business process models: Brief survey of reasearch results and challenges (Hull)
- Automatic verification of data-centric business processes (Deutsch; Hull; Patrizi)
- Automatic construction of simple artifact-based business processes (Fritz; Hull)

5.4.3 CAiSE

In order to analyze the process-related publications of the CAiSE conference, the following search term has been used: *CAiSE (process OR processes)*. In the following cited by network, the articles that cite the results from that search term are shown.

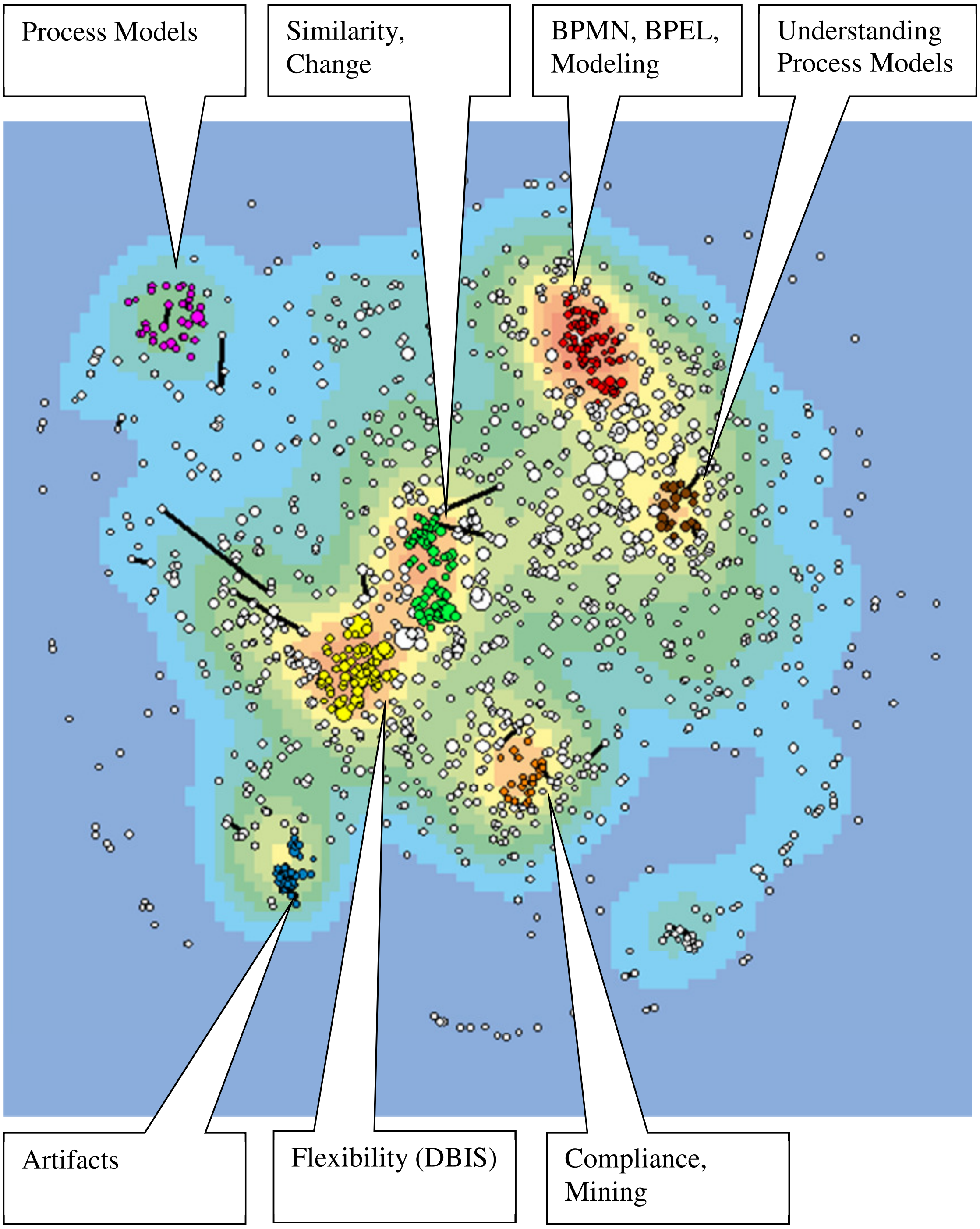


Figure 26: Cited by network of the CAiSE search term

Table 31: Numbers of the CAiSE search and of the corresponding network

Number of search results:	15,200
Number of cited bys (includes patents) of the first thousand articles:	17,906
Number of terms in the graph:	1,416

The following clusters have been identified:

Table 32: Clusters in the cited by network for the CAiSE search term

Name of the cluster	Number of articles within the cluster
BPMN, BPEL, Modeling	68
Flexibility	60
Similarity, Change	56
Artifacts	44
Process Models	33
Compliance, Mining	23
Understanding Process Models	20

BPMN, BPEL, Modeling Cluster

This cluster is mostly about BPMN, BPEL and process modeling. Frequent keywords include:

- BPEL
- UML
- BPMN
- Model/Modeling
- Mapping
- Transformation/Translation
- Choreographies
- Service-oriented/SOA
- Event-driven Process Chains

The most frequently cited articles are:

- On the translation between BPMN and BPEL: Conceptual mismatch between process modeling languages (Recker)
- Translating standard process models to BPEL (Ouyang; Dumas; Breutel)
- Transformation strategies between block-oriented and graph-oriented process modelling languages (Mendling; Lassen; Zdun)

Flexibility (DBIS) Cluster

This cluster is about flexible and dynamic processes. It is notable that most of the cited articles are written by members of the DBIS institute. The most common keywords are:

- Processes

- Flexible/Flexibility
- Dynamic
- Evolution
- Web Services
- Process Variants
- Mining
- Process-Aware Information System
- ADEPT2
- Adaptation
- Patterns

The most frequently cited articles are:

- Change patterns and change support features-enhancing flexibility in process aware information systems (Reichert; Weber)
- Change patterns and change support features in process aware information systems (Weber; Rinderle)
- Unleashing the effectiveness of process-oriented information systems: Problem analysis, critical success factors and implications (Reichert; Mutschler)

Among the other authors Reichert and Rinderle are also dominant.

Artifacts Cluster

This cluster is about business artifacts. The most common keywords are:

- Artifact
- Artifact-centric
- Artifact-based
- Data-aware
- Case Management

The most frequently cited articles are:

- Towards formal analysis of artifact-centric business process models (Bhattacharya; Gerede; Hull; Liu)
- Automatic verification of data-centric business processes (Deutsch; Hull; Patrizi)
- Artifact-centric business process models: Brief survey of research results and challenges (Hull)

Process Models Cluster
This cluster is about process models and process modeling. The most common keywords are:

- Business Process Change
- Modeling
- Metamodeling
- Service
- Business Process Models
- Business Process Engineering
- Method Engineering

The most frequently cited articles are:

- A multi-model view of process modeling (Rolland; Prakash)
- Modelling and Engineering the Requirements Engineering Process: an overwiew of the NATURE approach (Grosz; Rolland; Schwer; Souveyet)
- An assembly process model for method engineering (Ralyté)

Similarity Cluster
This cluster is about the similarity of process models. The most common keywords are:

- Similarity
- Patterns
- Process Metric
- Reference Models
- Lifecycle Model

The most frequently cited articles are:

- Measuring similarity between business process models (van Dongen; Dijkman)
- Measuring similarity between semantic business process models (Ehrig; Koschmider)
- On measuring process model similarity based on high-level change operations (Reichert; Li)

Compliance, Mining Cluster
The next cluster is about the topics compliance and process mining. The most common keywords are:

- Compliance
- Mining

- Monitoring

The most frequently cited articles are:

- Process mining and verification of properties: An approach based on temporal logic (van der Aalst; de Beer)
- Conformance checking of processes based on monitoring real behavior (Rozinat)
- Business process mining: An industrial application (van der Aalst; Reijers; Weiters)

Understanding Process Models Cluster
The last cluster is about understanding process models. The main keywords are:

- Understanding
- Usability
- Cognitive Effectiveness

The most frequently cited articles are:

- What makes process models understandable? (Mendling; Reijers)
- Influence factors of understanding business process models (Mendling)
- Does it matter which modelling language we teach or use? an experimental study on understanding process modelling languages without formal education (Recker)

5.4.4 CAiSE Inverted Cited by Network

In the cited by networks I determined research fronts. In order to also include the knowledge bases, I created the inverted cited by network on basis of the same data as the CAiSE cited by network. The inverted cited by network works on the result articles of the search instead of the forward citations of the result articles. The inverted cited by network can be seen in the following figure. The network contains 647 elements.

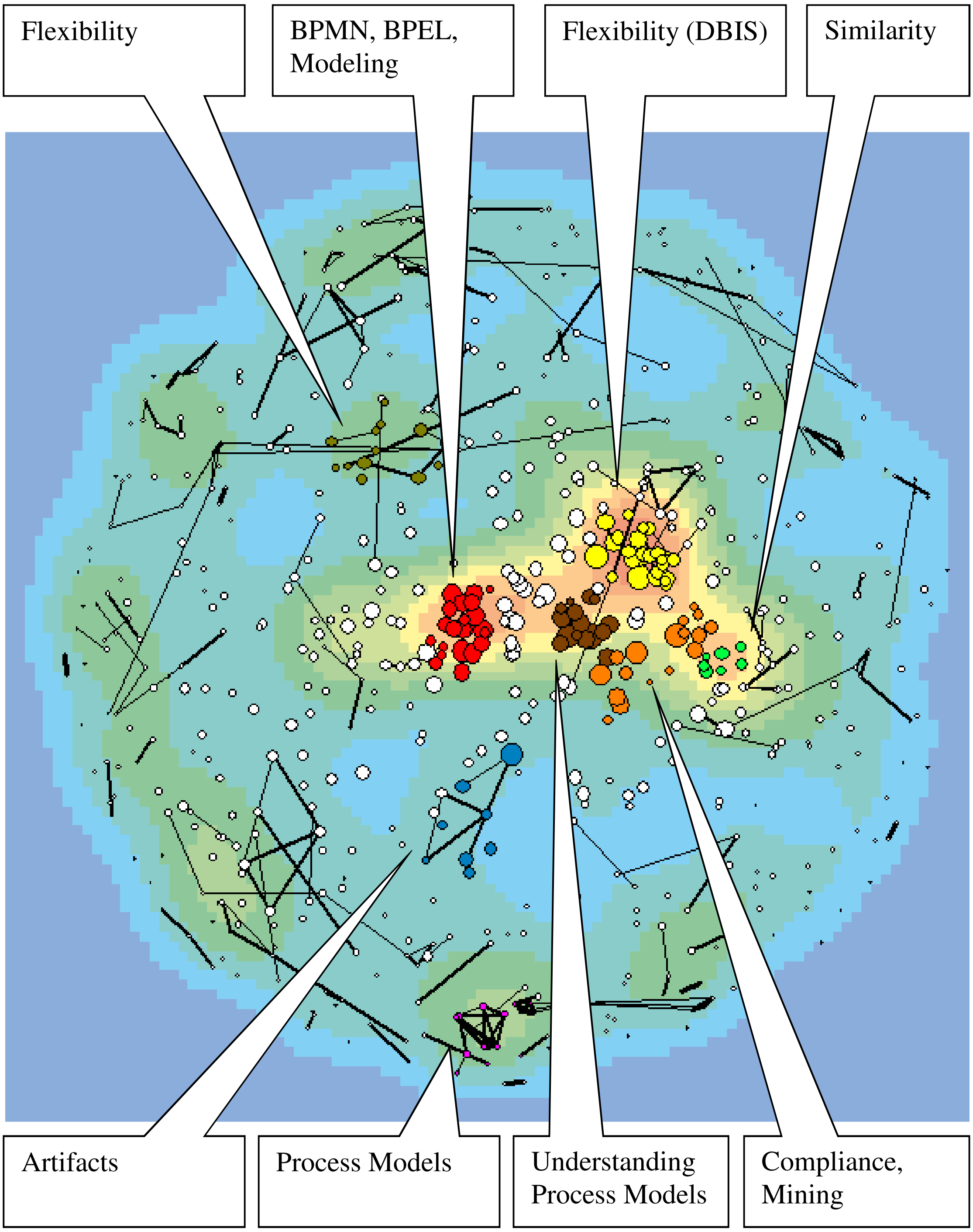


Figure 27: Inverted cited by network of the CAiSE search term

The same clusters were identified as in the cited by network, with the difference that there are two flexibility clusters in this network, instead of just one cluster in the cited by network. In the cited by network, these two cluster merged, with the articles from the DBIS flexibility clusters dominating among the cited articles.

Table 33: Clusters in the CAiSE inverted cited by network

Name of the cluster	Number of articles within the cluster
Flexibility (DBIS)	27
BPMN, BPEL, Modeling	22
Compliance, Mining	17
Process Models	15
Understanding Process Models	15
Flexibility	13
Artifacts	8
Similarity	7

The clusters mostly contain the articles that are mentioned as the cited articles of the clusters of the CAiSE cited by network and articles with similar topics. Also, the keywords the titles are similar to the ones of the clusters in cited by network. I will look now at the two flexibility clusters.

Flexibility Cluster

This cluster contains articles by several different authors such as Regev, Soffer or Sadiq. The keywords include:

- Flexibility
- Rigidity
- Exceptions

Flexibility (DBIS) Cluster

This cluster is dominated by authors like Reichert, Rinderle-Ma and Weber. Common keywords in this cluster are:

- Flexible
- Dynamic
- Adaptive

The articles of the Flexibility (DBIS) cluster were quite important as cited articles in the cited by network, while the articles from the other Flexibility cluster were not as important. So, it is interesting to note that there are two knowledge bases about the same topic, while there is only one research front about the topic, which is dominated by the DBIS flexibility articles.

6 Conclusions

6.1 Interview with Professor Reichert

On December 16, 2011, I conducted an interview with Professor Reichert of the Institute of Databases and Information Systems (DBIS) at the University of Ulm, in order to get the opinion of an expert about the BPM field and to receive some information about where he sees the position of the DBIS institute in comparison to other researchers in the BPM field. The complete text of the interview can be found in Appendix V. Later, I will compare his statements with my findings from the bibliometric analysis.

In the interview, Professor Reichert mentioned the focus of the DBIS institute. The focus of the DBIS is on several topics: BPM, special aspects of SOA, workflow management and e-health.

Within BPM the DBIS focuses on process flexibility and adaptivity, workflow management systems, robustness and correctness of those systems, process variants and change mining as a sub discipline of process mining. A new topic the DBIS follows is mobile processes.

He then mentioned important researchers in the field of BPM. Among others, he mentioned van der Aalst, Dumas, Leymann and Reijers.

When asked about topics within BPM, Professor Reichert mentioned process flexibility and process mining as the two biggest sub topics. Among the other subtopics he considers important are artifact-based processes, workflow management systems, semantics, compliance and verification. As for future prospects of BPM, Professor Reichert stated that BPM might become more of a "background technology" and that "ubiquitous processes" could become important.

6.2 Results of the Bibliometric Analysis

In the last chapters I have analyzed the data of several search fields within the field of business process management or related to business process management.

At first, I analyzed the data gathered from three different search terms, all with the objective of covering as much of the business process management topic in general as possible. In these general BPM networks, several clusters could be identified. Among them are the topics of process mining, process flexibility, process modeling and process compliance. I also created a network of the authors in the BPM field and highlighted the most import ones.

Then, I proceeded to analyze topics within BPM, such as process metrics, process compliance and mobile processes. All the search terms yielded fairly high search results, which suggests, that the chosen topics receive significant attention of the research community. The clusters that I identified in these search fields were, among others, process modeling, process flexibility, artifacts and process mining. A smaller cluster related to healthcare topics has also been found.

After these specific topics within BPM, I looked at topics related to business process management, such as business intelligence, ERP and SOA. In those fields, the clusters were about topics such as IT outsourcing, process management or ERP

implementation. Also, cross relationships between those topics have been found. For example, in the business intelligence field also a knowledge management cluster has been found and in the knowledge management topic an ERP cluster was identified.
In the last chapter, I had a look at articles related to the Data & Knowledge Engineering Journal, to the CAiSE conference and to the International Conference of Business Process Management.
In the articles citing the process-related articles from the Data & Knowledge Engineering Journal, the modeling and flexibility clusters are the biggest ones. Among the articles related to the International Conference on Business Process Management, the biggest clusters are about process mining, the modeling languages BPMN and BPEL and the topic of semantics, compliance and verification. Among the articles related to the CAiSE conference, clusters were about BPEL/BPMN and modeling, about flexibility and about similarity and artifacts. The cluster about flexibility is strongly based on articles by Reichert and other members of the DBIS institute.

6.3 Comparison with the Results from the Interview

I will now compare these findings with the core statements from Professor Reichert. He mentioned process flexibility and process mining as two of the most import subtopics of BPM. This can be confirmed, since process mining and process flexibility appear frequently as clusters in several different search fields. He also mentioned artifact-based processes, compliance and verification. These topics have also been found as clusters. Professor Reichert also mentioned important researchers in the field of BPM. Of the twelve authors I identified as the authors with most publications in the BPM search field (see Chapter 5.1.4), Professor Reichert mentioned eight. With him being among those twelve authors as well, only three authors were not covered. These three authors, Davenport, Grover and Jennings, do arguably not belong to the core BPM field, but to related fields, such as knowledge management and information systems in the case of Davenport and Grover, and to agent theory, in the case of Jennings.[76] As for the positioning of the DBIS institute Professor Reichert mentioned that the main focus of DBIS within BPM is the topic process flexibility and that DBIS is one of the leaders in this field. This can also be confirmed by the fact that the flexibility clusters strongly cite publications from Reichert and other DBIS members.

6.4 Further Analysis of the Results

Judging from the amount of articles found with the several search terms, BPM is a highly active research field. It has various important subtopics such as process

[76] See for example one of Davenport´s main works: Davenport, Prusak (1998): Working Knowledge – How organizations manage what they know, and the websites of Grover and Jennings: Grover: Profile, URL: http://www.clemson.edu/cbbs/faculty-staff/profiles/profile.html?userid=VGROVER, accessed February 20, 2012, and Jennings: Welcome, URL: http://users.ecs.soton.ac.uk/nrj/, accessed February 20, 2012

compliance or mobile processes, the importance of which can also be seen in the high number of articles found in that field. There is also a certain overlap between process topics and other IT topics such as business intelligence or ERP systems. Some of the clusters these other IT topics have in common with the BPM topics are process mining and workflow themes in general.

Clusters that occur several times in the BPM fields and the sub topics of BPM are process modeling, process mining and process flexibility. Other important clusters are inter-organizational and distributed processes as well as artifact-based processes.

As for the work with the forward citations from Google Scholar, the results can be considered fairly successful. It was possible to identify clusters in BibTechMon in the cited by networks and to draw meaningful conclusions from them. However, Google Scholar could facilitate the bibliometric analyses by offering a possibility to download data sets in a CSV file. As a next step it could be possible to implement similar functions like BibTechMon has directly in online services like Google Scholar or Web of Science. This way one could create bibliometric networks without having to download data sets first. Also, it might be possible to create networks with a much larger base of articles and maybe even with the complete set of articles from a certain field, such as computer science, and create bibliometric networks of the whole field. At the same time, some simpler bibliometric functions have already been implemented at sites like authormapper.com[77], a service of Springer, that shows the distribution of authors for a given search term, or at Google Scholar, where additional information of an author can be shown, such as total number of his citations, the number of citations by year and co-authors with which the author has published articles together.

6.5 Future Prospects of BPM and Bibliometrics

Both of the fields I explored in this work, BPM and bibliometrics, still have a lot of potential. BPM might one day, as Professor Reichert pointed out, become a "background technology" that will become as easy to use and as common as databases. In the future, process-support might not just be focused on business processes, but it might be available for any process-related task, at any place and at any time.

Bibliometrics on the other hand will become more and more important because of the growing number of scientific publications and the increasing availability of bibliometric data. Also, the constantly growing computer resources will make it possible to perform complex bibliometric analyses on a much larger scale.

As for bibliometric analyses in the BPM field, I see two areas where further work could be promising: One would be an in-depth analysis of authors, institutions and organizations in the BPM field in general. In order to do this, additional information that is not provided by Google Scholar would be necessary. The other promising area would be the inclusion of the time aspect in the bibliometric analysis, in order to find out how the activity of certain fields, for example process mining or process flexibility, changed over the course of time.

[77] Springer: AuthorMapper, URL: http://authormapper.com/, accessed February 12, 2012

Appendix

I. Source Code of the Google Script

This is the source code of the Google script, which can be used to extract data from Google Scholar and to turn the data into CSV files. The script is written in Perl.
Note: According to its terms of service, it is not allowed to automatically download the results of Google Scholar. Hence the following script may only be used in its "offline mode" on the basis of manually saved Google Scholar files.

```
use LWP::Simple; # module which allows the access of web sites via HTTP
use CGI::Carp; #module for error messages
use URI::Escape; # module for encoding and decoding unsafe characters
require LWP::UserAgent; # implements a user agent

  my $ua = LWP::UserAgent->new( agent => 'Mozilla/5.0 (Windows; U; Windows
NT 5.1; en-US) AppleWebKit/525.13 (KHTML, wie z. B. Gecko) Chrome/0.X.Y.Z
Safari/525.13.'); # a user agent is defined in order to pretend to be a normal web
browser

 $ua->default_header(
   'Accept-Language' => 'en-US',
   'Accept-Charset' => 'utf-8');

$zeit=time; # takes a timestamp for the default file name of the files
$artikel_pro_seite=100; # number of articles on each Google Scholar site
$online=1; # the default value for the mode the program is run in; 0 for the offline
mode; 1 for the online mode; 2 for the hybrid mode
$ordner="data"; # name of the folder in which the Google Scholar data is stored
$datei_prefix="g$zeit"; # default prefix for the file names of the files that will be
saved or opened
$suchstring="business+process+management"; # default search string
$max_artikel=$artikel_pro_seite; # default value for the maximum number of articles
that will be processed (default is 100)
$max_zitate=$artikel_pro_seite; # default value for the  number of cited articles that
will be processed for each article in the article list (default is 100)
$wartezeit=1; # default waiting time between requests to Google Scholar
$einschraenkung_fachbereich=0; # default value for limiting the search to special
field (default is 0; 0 stands for no limitation)
$einschraenkung_publikation=0; # default value for limiting the search to a specific
publication (default is 0; 0 stands for no limitation)
print "\nPeter Wohlhaupters Scholar-Abfrage-Tool\n\n\n";

if (!(-e "data/")) # checks if the folder data exists
```

```
{
mkdir("data"); # if it doesn't exist the folder data will be created
print "Erstelle Ordner data\n";
}

# the mode in which the program will be run is chosen here
print "Fuer Offline-Modus 0 eingeben, fuer Online-Modus 1 eingeben, fuer Hybrid-
Modus 2 eingeben. Enter fuer Standard-Wert.\n";
print "Standardwert: $online\n";
chomp($eingabe=<STDIN>);

if ($eingabe eq "0" || $eingabe eq "1" || $eingabe eq "2") {
$online=$eingabe;
}

# the prefix for the file names of the files written and read is chosen here
print "\nDatei-Prefix fuer Ausgabe- (und ggf. Offline-)Dateien eingeben. Enter fuer
Standard-Prefix.\n";
print "Standard-Prefix: $datei_prefix\n";
chomp($eingabe=<STDIN>);

if ($eingabe ne "") {
$datei_prefix=$eingabe;
}

# the search string is chosen here
print "\nSuch-String eingeben. Enter fuer Standard-Suchstring.\n";
print "Standard-Suchstring: $suchstring\n";
chomp($eingabe=<STDIN>);

if ($eingabe ne "") {
$suchstring=uri_escape($eingabe); # in case the search string is not empty, this will
encode the search string to make it usable in the URI
}

# the maximum number of articles that will be processed is chosen here
print "\nAnzahl Artikel, die maximal abgerufen werden sollen, eingeben (Vielfache
von $artikel_pro_seite). Fuer Standardwert Enter betaetigen.\n";
print "Standardwert: $max_artikel\n";
chomp($eingabe=<STDIN>);

if ($eingabe ne "") {
$max_artikel=$eingabe;
}

# the maximum number of citing articles that will be process for each article can be
chosen here
```

```
print "\nAnzahl Zitate, die je Artikel maximal abgerufen werden sollen, eingeben
(Vielfache von $artikel_pro_seite). Fuer Standardwert Enter betaetigen.\n";
print "Standardwert: $max_zitate\n";
chomp($eingabe=<STDIN>);

if ($eingabe ne "") {
$max_zitate=$eingabe;
}

# here it can be chosen whether the search should be limited to certain specific fields;
this is only relevant if the online or hybrid mode is running

if ($online!=0) {
print "\nKuerzel der Fachbereiche, welche beruecksichtigt werden sollen, mit Komma
getrennt eingeben (ohne Leerzeichen). Kuerzel:  bio, med, bus, phy, chm, soc, eng. 0
fuer alle Fachbereiche eingeben. Fuer Standardwert Enter betaetigen.\n";
print "Standardwert: $einschraenkung_fachbereich\n";
chomp($eingabe=<STDIN>);

if ($eingabe ne "") {
$einschraenkung_fachbereich=$eingabe;
}

# here it can be chosen if the search shall be limited to a certain publication; this is
only relevant if the online or hybrid mode is running
print "\nSoll nur eine bestimmte Publikation durchsucht werden? Wenn ja, dann
Namen der Publikation hier eingeben. 0 fuer keine Einschraenkung eingeben. Fuer
Standardwert Enter betaetigen.\n";
print "Standardwert: $einschraenkung_publikation\n";
chomp($eingabe=<STDIN>);

if ($eingabe ne "") {
$einschraenkung_publikation=$eingabe;
}

# here it can be chosen if the search shall be limited to a certain publication; this is
only relevant if the online or hybrid mode is running
print "\nDurchschnittliche Wartezeit zwischen Suchabfragen eingeben. Fuer
Standardwert Enter betaetigen.\n";
print "Standardwert: $wartezeit\n";
chomp($eingabe=<STDIN>);

if ($eingabe ne "") {
$wartezeit=2*$eingabe;
}

}
```

```
# create subfolder with the file name prefix chosen above, unless the folder already
exists

if (!(-e "data/".$datei_prefix."/"))
{
mkdir("data/".$datei_prefix);
print "Erstelle Ordner data/".$datei_prefix."\n";
}

# define filename for the output csv file
$dateiname_ausgabe=$datei_prefix."_ausgabe_".$zeit.".csv";
# in case it has been chosen that the search shall be limited to certain search field, the
following part creates the necessary string that will be attached to the search URI
$string_einschraenkung="";

if ($einschraenkung_fachbereich ne "0")
{
@subjs=split(/,/, $einschraenkung_fachbereich);
foreach (@subjs) {
$string_einschraenkung.="&as_subj=$_";
}
}

if ($einschraenkung_publikation ne "0")
{
$string_einschraenkung.="&as_publication=".uri_escape($einschraenkung_publikati
on);
}

my@artikel=getAllArticles("http://scholar.google.com/scholar?as_sdt=1,5&hl=en&q
=".$suchstring."&num=100".$string_einschraenkung, $max_artikel); # gets the list of
articles from Google Scholar or from already saved files

foreach (@artikel)
{
($title1, $author1, $journal1, $jahr1, $publisher1, $cit_id1,
$number_of_cited_bys)=getDetails($_); # from each article different kinds of
information is extracted: the title, the author(s), the journal, the year it was published
in, the publisher, an ID that is given to each article by Google Scholar; and the
number of articles the article is cited by

# if the ID is not undefined, the data from the article is added to the relevant lists
if ($cit_id1!=0)
{
push(@liste_ids, $cit_id1);
push(@liste_titles, $title1);
```

```
push(@liste_autoren, $author1);
push(@liste_journals, $journal1);
push(@liste_jahre, $jahr1);
push(@liste_publisher, $publisher1);
push(@liste_number_of_cited_bys, $number_of_cited_bys);
}

}


$k=0;
foreach $artikel_id (@liste_ids) {
@artikel_zitiert_durch=getAllArticles("http://scholar.google.com/scholar?as_sdt=1,5
&sciodt=1,5&hl=en&cites=$artikel_id&num=100", $max_zitate); # for each article
get the articles that cite that article


foreach (@artikel_zitiert_durch)
{
($title_zitiert_durch, $author_zitiert_durch, $journal_zitiert_durch,
$jahr_zitiert_durch, $publisher_zitiert_durch, $cit_id_zitiert_durch,
$number_of_cited_bys_zitiert_durch)=getDetails($_); # for each citing article get the
information that is also extracted from the base articles
$liste_cited_by[$k].=$title_zitiert_durch." (".$cit_id_zitiert_durch."); "; # add the title
and the ID of the citing article to the relevant list entry of the base article
$l=0;
foreach $aid (@liste_ids) {
if ($aid==$cit_id_zitiert_durch) { # if the ID of one of the base articles is identicial
with the ID of the citing article, than the base article gets the cited article as a
reference
$liste_zitate[$l].=$liste_titles[$k]." (".$liste_ids[$k]."); "; # add reference to the base
article
}

$l++;
}

}

$k++;
}

# open a new csv file and save the data gathered for each base article
print "Oeffne Datei $dateiname_ausgabe zum Schreiben\n";
open(DATEI2, ">data/".$datei_prefix."/"."$dateiname_ausgabe") || fehler("Fehler
beim Öffnen der Ausgabedatei $dateiname_ausgabe.");
$h=0;
```

```
print DATEI2 "quan, Authors, Title, Publication, Year, Publisher, References,
CitedBy, NumberOfCitedBys\n";

foreach (@liste_ids)
{
print DATEI2 "$_, \"$liste_autoren[$h]\", \"$liste_titles[$h]\", \"$liste_journals[$h]\",
\"$liste_jahre[$h]\", \"$liste_publisher[$h]\",  \"$liste_zitate[$h]\",
\"$liste_cited_by[$h]\", \"$liste_number_of_cited_bys[$h]\"\n";
$h++;
}

close (DATEI2);

print "\nZum Beenden Enter druecken.";
<STDIN>; # waits for the user to push enter in order to close the program



# coordinates the processing of the articles and the citing articles
sub getAllArticles {
$start=0;
$anzahl=$artikel_pro_seite; # Number of articles espected per site
$max=$_[1]; # maximum number of articles that should be accessed
@a=();
@b=();

while ($start<=($max-$artikel_pro_seite) && $anzahl==$artikel_pro_seite)
{
@a=getArticles($_[0]."\&start=$start");
$anzahl=@a;
$start+=$artikel_pro_seite;
push (@b, @a);

if ($online==1) {
sleep(rand($wartezeit)); # waiting time between requests
}

}

$ges=@b;
$ges2+=@b;
return @b;
}
```

```
sub getArticles {
$_[0] =~m/cites=(.*?)&/; # extracts the ID of the base article
$cite=$1;
 $_[0]=~m/start=(.*)/; # extracts the start value
$startwert=$1;
if ($cite!=0)
{
if ($startwert!=0)
{
$datei="$datei_prefix"."_$cite"."_$startwert.htm";
}
else
{
$datei="$datei_prefix"."_$cite.htm";
}
}
else
{
if ($startwert!=0)
{
$datei="$datei_prefix"."_".$startwert.".htm";
}
else
{
$datei="$datei_prefix"."_0.htm";
}
}

if (($online==0) || ($online==2 && -e "data/".$datei_prefix."/"."$datei")) # in case
the program is run in the offline mode or the hybrid mode, try to open the article as a
saved file
{
 open(DATEI,"<data/".$datei_prefix."/"."$datei") || fehler("Fehler beim Oeffnen der
Datei $datei."); # try to read the file
$result= do {local $/; <DATEI> };
close(DATEI);
print "Oeffne Datei $datei\n";
}

else # if the article is not available offline, try to access it directly from google scholar
{
print "Oeffne Adresse $_[0]\n";
 my $response = $ua->get($_[0]);

 if ($response->is_success) {
   $result=$response->decoded_content;
```

```
if (-e "data"."/".$datei_prefix."/"."$datei")
{
fehler("Datei data"."/".$datei_prefix."/"."$datei bereits vorhanden.");
}

print "Oeffne Datei $datei zum Schreiben\n"; # save file accesses from google scholar
open(DATEI3, ">data"."/".$datei_prefix."/"."$datei") || fehler("Kann $datei nicht zum
Schreiben oeffnen.");
print DATEI3 $result;
close DATEI3;

}

else {
    fehler($response->status_line); # if an error occurs while trying to get the data
from the web, the program is aborted
}

}

@l=split(/<div class=gs_r>/,$result); # splits up the html code into several parts, each
about one article
shift @l; # deletes the first element of the list of raw data, because this element does
not contain article data
return @l;
}

# splits up the raw data of the articles
sub getDetails {

foreach $art (@_) {
($art, $rest) = split(/ <\/div> /,$art);
$art =~ m/<h3>(.*)<\/h3>/; # extract the title which is written between h3-tags
$titel=$1;
$titel=~s/<(.*?)>//g; # eliminate unnecessary tags within the title (such as b-tags
$titel=~s/;//g; # eliminate ;
$art =~ m/<span class=gs_a>(.*?)<\/span>/; # extract the raw data about the author
and the publication and the journal
$autorplus=$1;
$autorplus=~s/<(.*?)>//g; # eliminate html-tags
($autor_raw, $journal_raw, $publisher)=split(/ - /, $autorplus); # split this raw data up
into data about the author, the journal and the publisher
# eliminate unnecessary parts
$autor_raw=~s/<.*?>//g;
$autor_raw=~s/&hellip;//g;
$autor_raw=~s/;//g;
```

```
$autor_raw=~s/,/;/g; # changes delimiter , to delimiter;
# split up different author names and transform them into the form [1. letter first name] [last part of the last name]
@autoren_komplett=split(/; /, $autor_raw);

foreach (@autoren_komplett)
{
@teile=split(/ /, $_);
if ($teile[0] ne "" && $teile[1] ne "") { # in case name only consists out of one element, the name will not be changed
$_=substr($teile[0], 0, 1)." ".$teile[-1]; # if it consists of several elements, the first letter of the first element and the last element will be used (this is done to standardize more complex names that sometimes are given in different forms
}

}

$autor_raw=join("; ", @autoren_komplett);
#end of transformation

#extract journal and year
@werte=split(/, /,$journal_raw);

if ($werte[-1]=~/([0-9]{4}).*/)
{
$jahr=$1;
pop(@werte);
$journal=join(", ", @werte);
}
else {
$journal=join(", ", @werte);
$jahr="";
}

$journal=~s/&hellip;//g;
$journal=~s/;//g;
$publisher=~s/&hellip;//g;
$publisher=~s/;//g;

if ($art=~/cites=(.*?)(&amp|"|>)/)
{
$art =~m/cites=(.*?)(&amp|"|>)/;
$id = $1;
}
else
{
$id="";
```

```
}

if ($art=~/Cited by ([0-9]*?)</) #extracts number of cited bys
{
$zaehler_cited_by=$1;
}

return ($titel, $autor_raw, $journal, $jahr, $publisher, $id, $zaehler_cited_by); #
returns the details of each article

}
}


# sub routine that gives out error messages and waits for the user to push enter before
closing the program
sub fehler {
print "\n\nFehler: $_[0]";
print "\nZum Beenden Enter druecken.";
<STDIN>;
die;
}
```

II. Changes of the Script after Google Scholar Altered its Format

In the end of 2011 or in the beginning of 2012, Google slightly changed the format of the result pages in Google Scholar. This made it necessary to change the source code of the script, as well. All the necessary changes were made in the function getDetails(). The changed function can be seen as follows.

```
sub getDetails {

foreach $art (@_) {
($art, $rest) = split(/ <\/div> /,$art); # eliminates everything after " </div> "; this
prevents that the links to further result pages will be processed
$art =~ m/<h3 class="gs_rt">(.*)<\/h3>/; # extract the title which is written between
h3-tags
$titel=$1;
$titel=~s/<(.*?)>//g; # eliminate unnecessary tags within the title (such as b-tags)
$titel=~s/;//g; # eliminate ;
$art =~ m/<div class=gs_a>(.*?)<\/div>/; # extract the raw data about the author and
the publication and the journal
$autorplus=$1;
$autorplus=~s/<(.*?)>//g; # eliminate html-tags

($autor_raw, $journal_raw, $publisher)=split(/ - /, $autorplus); # split this raw data up
into data about the author, the journal and the publisher
# eliminate unnecessary parts
$autor_raw=~s/<.*?>//g;
$autor_raw=~s/&hellip;//g;
$autor_raw=~s/;//g;
$autor_raw=~s/,/;/g; # changes delimiter , to delimiter;
# split up different author names and transform them into the form [1. letter first
name] [last part of the last name]
@autoren_komplett=split(/; /, $autor_raw);

foreach (@autoren_komplett)
{
@teile=split(/ /, $_);
if ($teile[0] ne "" && $teile[1] ne "") { # in case name only consists out of one
element, the name will not be changed
$_=substr($teile[0], 0, 1)." ".$teile[-1]; # if it consists of several elements, the first
letter of the first element and the last element will be used (this is done to standardize
more complex names that sometimes are given in different forms
}

}
$autor_raw=join("; ", @autoren_komplett);
#end of transformation
```

```
# extract journal and year
@werte=split(/, /,$journal_raw);
if ($werte[-1]=~/([0-9]{4}).*/)
{
$jahr=$1;
pop(@werte);
$journal=join(", ", @werte);
}
else {
$journal=join(", ", @werte);
$jahr="";
}

$journal=~s/&hellip;//g;
$journal=~s/;//g;
$publisher=~s/&hellip;//g;
$publisher=~s/;//g;

if ($art=~/cites=(.*?)(&amp|"|>)/)
{
$art =~m/cites=(.*?)(&amp|"|>)/;
$id = $1;
}
else
{
$id="";
}

if ($art=~/Cited by ([0-9]*?)</)
{
$zaehler_cited_by=$1;
}

return ($titel, $autor_raw, $journal, $jahr, $publisher, $id, $zaehler_cited_by); #
returns the details of each article

}
}
```

III. Functionalities of the Script and Further Notes to the Usage of the Script

The script, which is written in Perl, has the following functionalities:

- Fetch Google Scholar results for a given search term
- For each article in the results fetch the list of articles that cite said article
- Create a CSV file with the following data gathered from the Google Scholar results
    - quan: ID of each article given by Google Scholar
    - Authors
    - Title
    - Publication
    - Year (year the article/book was published)
    - Publisher
    - References: Since there is no way to get a list of articles that is cited by a specific article, a part of the references is recovered using a recursive method: If Article A is in the list of Google Scholar results and Article B is also in that list and in the list of articles citing Article B is Article A again, then Article B is a reference of Article A.
    - CitedBy: List of articles the article was cited by
    - NumberOfCitedBys: Number of other articles the article is cited by

Notes:
Generally speaking, the quality of the data gathered by the script can only be as good as the quality of the data provided by Google Scholar.
In some cases, the data of a specific article is not correct. For example the text given as the title of the article is not actually the title, but another part of the article. Similarly, sometimes the name of the publication is not the real name of the publication.
The list of authors is directly extracted from the results given by Google Scholar. However, the list of authors of one article in the Google Scholar results is not always complete and one or several of the co-authors might be left out. Hence, the data about the authors gathered by this script will also be incomplete.
Additionally, the data about the publication of an article returned by Google Scholar is sometimes incomplete or incorrect, as well. Correspondingly, the data about the publications gathered by this script will also be incomplete or incorrect.

IV. Transformation of the CSV File into a MDB File

In order to transform the CSV file that is created by the Google script into a MDB file, I used Microsoft Access. The following steps are necessary to create the MDB file in the desired form:

- Use comma as separator
- Use quotation marks as text delimiter symbol
- Set decimal delimiter symbol to dot
- Use the first line as names for the columns

Then save the database in the MDB format.

References

Alavi, Maryam; Leidner, Dorothy E.: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. In: MIS Quarterly, Volume 25, Number 1, pages 107-136, March 2001

Becker, Jörg; Kahn, Dieter: The Process in Focus. In: Becker, Jörg; Kugeler, Martin; and Rosemann, Michael (Editors): Process management: a guide for the design of business processes, pages 1-11, Springer, 2003

BPM 2011 9th International Conference on Business Process Management: Welcome to BPM 2011, URL: http://bpm2011.isima.fr/, accessed February 2, 2012

Broadus, R. N.: Toward a definition of 'bibliometrics', Scientometrics, Volume 12, Numbers 5-6, pages 373–379, 1987

CAiSE '11 - 23rd International Conference on Advanced Information Systems Engineering: Call for Papers, URL: http://www.caise2011.com/callForPapers.php, accessed February 2, 2012

Campbell, F.: The Theory of the National and International Bibliography: with Special Reference to the Introduction of System in the Record of Modern Literature, Library Bureau, 1896

Cole, F.J.; Eales, N.B.: The history of comparative anatomy. Part I: A statistical analysis of the literature, In: Science Progress, Volume 11, pages 578-596, 1917

Davenport, Thomas H.; Prusak, Laurence: Working Knowledge – How organizations manage what they know, Harvard Business School Press, 1998

Dijkman, Remco; Dumas, Marlon; van Dongen, Boudewijn; Käärik, Reina; Mendling, Jan: Similarity of Business Process Models: Metrics and Evaluation. In: Information Systems, Volume 36, Issue 2, pages 498–516, April 2011

Duguet, Emmanuel; MacGarvie, Megan: How Well Do Patent Citations Measure Flows of Technology? Evidence from French Innovation Surveys. In: Economics of Innovation and New Technology, Volume 14, Number 5, pages 375-393, July 2005

Egghe, Leo: Theory and practise of the g-index. In: Scientometrics, Volume 69, Number 1, pages 131-152, 2006

El Kharbili, M.; Alves de Medeiros, A.K.M; Stein, S.; van der Aalst, Wil: Business Process Compliance Checking: Current State and Future Challenges. In: Modelling Business Information Systems, November 2008

Elsevier: Data & Knowledge Engineering, URL: http://www.journals.elsevier.com/data-and-knowledge-engineering/, accessed February 2, 2012

Encyclopedia Britannica, URL: http://www.britannica.com, accessed December 12, 2011

Erl, Thomas: The Service-Orientation Design Paradigm, URL: http://www.soaprinciples.com/p3.php, accessed February 27, 2012

Fairthorne, R.A.: Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction. In: Journal of Documentation, Volume 25, Number 4, pages 319-343, 1969

Fritz, Christian; Hull, Richard; Su, Jianwen: Automatic construction of simple artifact-based business processes. In: Proceedings of the 12th International Conference on Database Theory, ACM, 2009

Gartner: Magic Quadrant for ERP for Product-Centric Midmarket Companies, 2010

Gluchowski, Peter: Business Intelligence - Konzepte, Technologien und Einsatzbereiche. In: HMD Praxis der Wirtschaftsinformatik,Volume 222, pages 5-15, 2011

Golfarelli, Matteo; Rizzi, Stefano; Cella, Iuris: Beyond data warehousing: what's next in business intelligence? In: Proceedings of the 7th ACM international workshop on Data warehousing and OLAP, ACM, 2004

Google Scholar: General search, URL: http://scholar.google.com, accessed December 15 2011

Google Scholar: Help page, URL: http://scholar.google.com/intl/en/scholar/help.html, accessed February 20, 2012

Grigori, Daniela; Casati, Fabio; Castellanos, Malu; Dayal, Umeshwar; Sayal,Mehmet; Shan, Ming-Chien: Business Process Intelligence. In: Computers in Industry, Volume 53, Issue 3, pages 321–343, April 2004

Grover, Varun: Profile, URL: http://www.clemson.edu/cbbs/faculty-staff/profiles/profile.html?userid=VGROVER, accessed February 20, 2012

Harzing, Anne-Wil; van der Wal, Ron: Google Scholar as a new source for citation analysis. In: Ethics in Science and Environmental Politics, Volume 8, Number 1, pages 62-71, 2008

Hirsch, Jorge: An index to quantify an individual's scientific research output. In: Proceedings of the National Academy of Sciiences of the United States of America, Volume 102, Number 46, pages 16569-16572, November 2005

Hood, William W.; Wilson, Concepción S.: The literature of bibliometrics, scientometrics, and informetrics. In: Scientometrics, Volume 52, Number 2, pages 291-314, 2001

IBM: IBM to Acquire Lombardi, URL: http://www-03.ibm.com/press/us/en/pressrelease/28890.wss, accessed January 15, 2012

Jarrar, Y.F.; Al-Mudimigh, A.; Zairi, M.: ERP implementation critical success factors-the role and impact of business process management. In: Proceedings of the 2000 IEEE International Conference on Management of Innovation and Technology, Volume 1, pages 122 – 127, 2000

Jennings, Nick: Welcome, URL: http://users.ecs.soton.ac.uk/nrj/, accessed February 20, 2012

Jung, Jisoo; Choi, Injun; Song, Minseok: An integration architecture for knowledge management systems and business process management systems. In: Computers in Industry, Volume 58, Issue 1, January 2007

Kopcsa, Alexander; Schiebel, Edgar: Science and Technology Mapping: A New Iteration Model for Representing Multidimensional Relationships. In: Journal of the American Society for Information Science, Volume 49, Issue 1, pages 7-17, 1998

Künzle, Vera; Reichert, Manfred: Towards Object-aware Process Management Systems: Issues, Challenges, Benefits. In: Enterprise, Business-Process and Information Systems Modeling - Lecture Notes in Business Information Processing, Volume 29, pages 197-210, Springer, 2009

Lawrence, Peter (Editor): Workflow Handbook 1997; John Wiley & Sons, 1996

Lu, Ruopeng; Sadiq, Shazia Wasim; Governatori, Guido: Compliance Aware Business Process Design. In: Proceedings of the 2007 international conference on Business process management, Springer, 2008

Luhn, Hans Peter: A Business Intelligence System. In: IBM Journal of Research and Development, Volume 2, Issue 4, October 1958

Meho, Lokman; Yang, Kiduk: Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar. In: Journal of the American Society for Information Science and Technology, Volume 58, Issue 13, pages 2105–2125, November 2007

Mendling, Jan: Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness, Springer, 2008

Merriam-Webster: Definition of compliance, URL: http://www.merriam-webster.com/dictionary/compliance, accessed February 17, 2012

Merriam-Webster: Definition of metric, URL: http://www.merriam-webster.com/dictionary/metric, accessed February 17, 2012

Meyer, Bertrand; Choppy, Christine; Staunstrup, Jorgen; van Leeuwen, Jan: Research Evaluation for Computer Science. In: Communications of the ACM, Volume 52, Issue 4, pages 31-34, April 2009

Müller, Dominic; Reichert, Manfred; Herbst, Joachim: In: Business Process Management Workshops - Lecture Notes in Computer Science, Volume 4103, page 181-193, Springer, 2006

Nacke, O. : Informetrie: Ein neuer Name für eine neue Disziplin. In: Nachrichten für Dokumentation, Volume 30, pages 212-216, 1979

Negash, Solomon; Gray, Paul: Business Intelligence. In Handbook on decision support systems 2 - International Handbooks on Information Systems, Volume VII, pages 175-193, Springer 2008

Noll, Margit; Fröhlich, Doris; Schiebel, Edgar: Knowledge Maps of Knowledge Management Tools — Information Visualization with BibTechMon. In: Practicacl Aspects of Knowledge Management - Lecture Notes in Computer Science, Volume 2569, pages 14-27, Springer, 2002

OASIS: About ebXML, URL: http://www.ebxml.org/geninfo.htm, accessed February 25, 2012

OASIS: OASIS Web Services Business Process Execution Language (WSBPEL) TC, URL: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsbpel, accessed February 25, 2012

Object Management Group: Business Process Model and Notation, URL: http://www.bpmn.org, accessed February 20, 2012

Otlet, P.: Traite de Documentation. Le Livre sur le Livre. Theorie et Pratique., Van Keerberghen, 1934

Pasley, J.: How BPEL and SOA are changing Web services development. In: IEEE Internet Computing, Volume 9, Issue 3, pages 60-67, May-June 2005

Pousttchi, Key; Thurnher, Bettina: Usage of mobile technologies to support business processes. In: Wireless Communication and Information, 2006

Pritchard, A.: Statistical bibliography or bibliometrics? In: Journal of Documentation, Volume 25, Number 4, pages 348–349, 1969

Pryss, Rüdiger; Tiedeken, Julian; Kreher, Ulrich; Reichert, Manfred: Towards Flexible Process Support on Mobile Devices. In: Information Systems Evolution - Lecture Notes in Business Information Processing, Volume 72, pages 150-165, Springer 2011

Reijers, Hajo A.; Song, Minseok; Romero, Heidi; Dayal, Umeshwar; Eder, Johann: Koehler, Jana: A Collaboration and Productiveness Analysis of the BPM Community. In: Business Process Management - Lecture Notes in Computer Science, Volume 5701, pages 1-14, Springer, 2009

Robecke, Andreas: Development of an iPhone business application. Diploma thesis, 2011

Schiebel, Edgar: Lecture notes in Technologie- und Innovationsmanagement III, University of Ulm, 2011

Scopus: Content Coverage Guide, URL: http://www.info.sciverse.com/scopus/scopus-in-detail/facts, accessed February 17, 2012

Springer: AuthorMapper, URL: http://authormapper.com/, accessed February 12, 2012

Thomson Reuters: Impact Factor, URL: http://thomsonreuters.com/products_services/science/free/essays/impact_factor/, accessed February 1, 2012

Thomson Reuters: Web of Knowledge, URL: http://thomsonreuters.com/content/science/pdf/Web_of_Knowledge_factsheet.pdf, accessed February 17, 2012

Van der Aalst, Wil.; ter Hofstede, Arthur H.M.; Weske, Mathias: Business Process Management: A Survey. In: Business Process Management - Lecture Notes in Computer Science, Volume 2678, Springer, 2003

Vanderfeesten, Irene; Cardoso, Jorge; Reijers, Hajo A.; van der Aalst, Wil: Quality Metrics for Business Process Models. In: Proceedings of the 9th WSEAS international conference on applied computer science, World Scientific and Engineering Academy and Society, 2009

Van Raan, A.F.J.; Tijssen, R.J.W.: The neural net of neural network research: An exercise in bibliometric mapping, In: Scientometrics, Volume 26, Number 1, pages 169-192, 1993

W3C: Web Services Choreography Description Language Version 1.0, URL: http://www.w3.org/TR/2004/WD-ws-cdl-10-20041217/, accessed February 25, 2012

Wohed, P.; van der Aalst, Wil; Dumas, M.; ter Hofstede, A.; Russel, N.: On the Suitability of BPMN for Business Process Modelling. In: Computer Science - Lecture Notes in Computer Science, Volume 4102, pages 161-176, Springer, 2006

ZDNet: IBM kauft Prozessmanagement-Software-Anbieter Lombardi, URL: http://www.zdnet.de/news/41524612/ibm-kauft-prozessmanagement-software-anbieter-lombardi.htm, accessed February 12, 2012

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit selbständig angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Ich bin mir bewusst, dass eine unwahre Erklärung rechtliche Folgen haben wird.

Ulm, den 7. März 2012

____

(Unterschrift)