Text Mining Scientific Papers: A Survey on FCA-Based Information Retrieval Research

Jonas Poelmans[1,4], Dmitry I. Ignatov[4], Stijn Viaene[1,2], Guido Dedene[1,3], and Sergei O. Kuznetsov[4]

[1] K.U. Leuven, Faculty of Business and Economics, Naamsestraat 69, 3000 Leuven, Belgium
[2] Vlerick Leuven Gent Management School, Vlamingenstraat 83, 3000 Leuven, Belgium
[3] Universiteit van Amsterdam Business School, Roetersstraat 11 1018 WB Amsterdam, The Netherlands
[4] National Research University Higher School of Economics (HSE), Pokrovsky boulvard 11, 101000 Moscow, Russia
{Jonas.Poelmans,Stijn.Viaene,Guido.Dedene}@econ.kuleuven.be, {dignatov,skuznetsov}@hse.ru

Abstract. Formal Concept Analysis (FCA) is an unsupervised clustering technique and many scientific papers are devoted to applying FCA in Information Retrieval (IR) research. We collected 103 papers published between 2003-2009 which mention FCA and information retrieval in the abstract, title or keywords. Using a prototype of our FCA-based toolset CORDIET, we converted the pdf-files containing the papers to plain text, indexed them with Lucene using a thesaurus containing terms related to FCA research and then created the concept lattice shown in this paper. We visualized, analyzed and explored the literature with concept lattices and discovered multiple interesting research streams in IR of which we give an extensive overview. The core contributions of this paper are the innovative application of FCA to the text mining of scientific papers and the survey of the FCA-based IR research.

1 Introduction

According to Manning et al. (2008), "information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)." In the past, only specialized professions such as librarians had to retrieve information on a regular basis. These days, massive amounts of information are available on the www and hundreds of millions of people make use of information retrieval systems such as web or email search engines on a daily basis. Formal Concept Analysis (FCA) was introduced in the early 1980s by Rudolf Wille as a mathematical theory (Wille 1982) and is a popular technique within the IR field. FCA is concerned with the formalization of concepts and conceptual thinking and has been applied in many disciplines such as software engineering, knowledge discovery and ontology construction during the last 15 years. The core contributions of this paper are as follows. We visually represent

P. Perner (Ed.): ICDM 2012, LNAI 7377, pp. 273–287, 2012.

the literature on FCA and IR using concept lattices, in which the objects are the scientific papers and the attributes are the relevant terms available in the title, keywords and abstract of the papers. We developed a toolset with a central FCA component that we use to index the papers with a thesaurus containing terms related to FCA research and to generate the lattices. We zoom in and give an extensive overview of the papers published between 2003 and 2009 on using FCA in information retrieval.

The remainder of this paper is composed as follows. In section 2 we introduce the essentials of FCA theory and the knowledge browsing environment we developed to support this literature analysis. In section 3 we describe the dataset used. In section 4 we visualize the FCA literature on information retrieval using FCA lattices and we summarize the papers published in this field. Section 5 concludes the paper.

2 Formal Concept Analysis

FCA (Ganter et al. 1999, Wille 1982) is a well established technique in mathematics and computer science and multiple partial surveys were published during the past years. A textual overview of part of the literature published until the year 2004 on FCA is given by Priss (2006). An overview of available FCA software is provided by Tilley (2004) and in Tilley et al. (2007), an overview of 47 FCA-based software engineering papers is given. The authors categorized these papers according to the 10 categories as defined in the ISO 12207 software engineering standard and visualized them in a concept lattice. In Lakhal et al. (2005), a survey on FCA-based association rule mining techniques is given. Poelmans et al. (2010) give an extensive overview of KDD applications of FCA. FCA groups scientific papers containing terms from the same term-clusters in concepts. The starting point of the analysis is a formal context (G, M, I) consisting of rows G (i.e. objects), columns M (i.e. attributes) and crosses $I \subseteq G \times M$ (i.e. relationships between objects and attributes).

Table 1. Example of a formal context

	browsing	mining	Software	web services	FCA	information retrieval
Paper 1	X	X	X		X	
Paper 2			X		X	X
Paper 3		X		X	X	
Paper 4	X		X		X	
Paper 5				X	X	X

An example of a cross table is displayed in Table 1. In the latter, scientific papers (i.e. the objects) are related (i.e. the crosses) to a number of terms (i.e. the attributes); here a paper is related to a term if the title or abstract of the paper contains this term. Given a formal context, FCA then derives all concepts from this context and orders them according to a subconcept-superconcept relation, resulting in a lattice.

The notion of concept is central to FCA. The way FCA looks at concepts is in line with the international standard ISO 704, that formulates the following definition: “A concept is considered to be a unit of thought constituted of two parts: its extent and its intent." The extent consists of all objects belonging to the concept, while the intent comprises all attributes shared by those objects. Let us illustrate the notion of concept of a formal context using the data in Table 1. For a set of objects $O \subseteq G$, the common features can be identified, written O', via:

$$A = O' = \{m \in M \mid \forall o \in O : (o, m) \in I\}$$

Take the attributes that describe paper 4 in Table 1, for example. By collecting all papers of this context that share these attributes, we get to a set $O \subseteq G$ consisting of papers 1 and 4. This set O of objects is closely connected to set A consisting of the attributes "browsing", "software" and "FCA": $O = A' = \{o \in G \mid \forall a \in A : (o, a) \in I\}$

That is, O is the set of all objects sharing all attributes of A, and A is the set of all attributes that are valid descriptions for all the objects contained in O. Each such pair (O, A) is called a formal concept (or concept) of the given context. The set $A = O'$ is called the intent, while $O = A'$ is called the extent of the concept (O, A).

There is a natural hierarchical ordering relation between the concepts of a given context that is called the subconcept-superconcept relation.

$$(O_1, A_1) \leq (O_2, A_2) \Leftrightarrow (O_1 \subseteq O_2 \Leftrightarrow A_2 \subseteq A_1)$$

A concept $d = (O_1, A_1)$ is called a subconcept of a concept $e = (O_2, A_2)$ (or equivalently, e is called a superconcept of a concept d) if the extent of d is a subset of the extent of e (or equivalently, if the intent of d is a superset of the intent of e). For example, the concept with intent "browsing", "software", "mining" and "FCA" is a subconcept of a concept with intent "browsing", "software" and "FCA." With reference to Table 1, the extent of the latter is composed of papers 1 and 4, while the extent of the former is composed of paper 1.

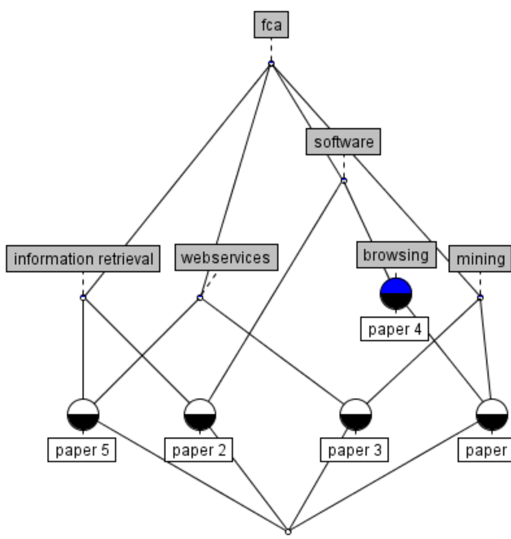

Fig. 1. Line diagram corresponding to the context from Table 1

The set of all concepts of a formal context combined with the subconcept-superconcept relation defined for these concepts gives rise to the mathematical structure of a complete lattice, called the concept lattice of the context. The line diagram in Figure 1 is a representation of the concept lattice of the formal context abstracted from Table 1. The circles or nodes in this line diagram represent the formal concepts. The shaded boxes (upward) linked to a node represent the attributes used to name the concept. The non-shaded boxes (downward) linked to the node represent the objects used to name the concept. The information contained in the formal context of Table 1 can be distilled from the line diagram in Figure 1 by applying the following reading rule: An object “g” is described by an attribute “m” if and only if there is an ascending path from the node named by “g” to the node named by “m.” For example, paper 1 is described by the attributes “browsing”, “software”, “mining” and “FCA.”

We developed a knowledge browsing environment CORDIET to support our literature analysis process (Poelmans et al. 2010b, Poelmans et al. 2010c). One of the central components of our text analysis environment is the thesaurus containing the collection of terms describing the different research topics. The initial thesaurus was constructed based on expert prior knowledge and was incrementally improved by analyzing the concept gaps and anomalies in the resulting lattices. The thesaurus is a layered thesaurus containing multiple abstraction levels. The first and finest level of granularity contains the search terms of which most are grouped together based on their semantical meaning to form the term clusters at the second level of granularity.

The papers that were downloaded from the World Wide Web (WWW) were all formatted in pdf. These pdf-files were converted to ordinary text and the abstract, title and keywords were extracted. The open source tool Lucene was used to index the extracted parts of the papers using the thesaurus. The result was a cross table describing the relationships between the papers and the term clusters or research topics from the thesaurus. This cross table was used as a basis to generate the lattices.

3 Dataset

This Systematic Literature Review (SLR) has been carried out by considering a total of 103 papers related to FCA and IR published between 2003 and 2009 in the literature and extracted from the most relevant scientific sources. The sources that were used in the search for primary studies contain the work published in those journals, conferences and workshops which are of recognized quality within the research community. These sources are: *IEEE Computer Society, ACM Digital Library, Sciencedirect, Springerlink, EBSCOhost, Google Scholar, Conference repositories: ICFCA, ICCS and CLA conference.* Other important sources such as DBLP or CiteSeer were not explicitly included since they were indexed by some of the mentioned sources (e.g. Google Scholar). In the selected sources we used various search strings including “Formal Concept Analysis”, “FCA”, “concept lattices”, “Information Retrieval”. To identify the major categories for the literature survey we also took into account the number of citations of the FCA papers at CiteseerX.

4 FCA-Based Information Retrieval Research

In Conceptual Knowledge Processing (CKP) the focus lies on developing methods for processing information and knowledge which stimulate conscious reflection, discursive argumentation and human communication (Wille 2006, Eklund et al. 2007). The word "conceptual" underlines the constitutive role of the thinking, arguing and communicating human being and the term "processing" refers to the process in which something is gained which may be knowledge. FCA can be particularly suited for IR because of its human-centeredness. The efficient retrieval of relevant information is promoted by the FCA representation that makes the inherent logical structure of the information transparent. FCA can be used for multiple purposes in IR (Priss 2006). First, FCA is an interesting instrument for browsing through large document collections. FCA can also support query refinement. Because a document-term lattice structures the available information as clusters of related documents which are partially ordered, lattices can be used to make suggestions for query enlargement in cases where too few documents are retrieved and for query refinement in cases where too many documents are retrieved. Third, lattices can be used for querying and navigation. An initial query corresponds to a start node in a document-term lattice. Users can then navigate to related nodes. Further, queries are used to "prune" a document-term lattice to help users focus their search (Carpineto et al. 1996b). For many purposes, some extra facilities are needed such as processing large document collections quickly, allowing more flexible matching operations, allowing ranked retrieval and give contextual answers to user queries. The past years many FCA researchers have also devoted attention to these issues.

The first attempts to use lattices for information retrieval are summarized in Priss (2000), but none of them resulted in practical implementations. Godin et al. (1989) developed a textual information retrieval system based on document-term lattices but without graphical representations of the lattices. The authors also compared the system's performance to that of Boolean queries and found that it was similar to and even better than hierarchical classification (Godin et al. 1993). They also worked on software component retrieval (Mili et al. 1997). In Carpineto et al. (2004a), their extensive work on information retrieval was summarized.

86 % of the papers on FCA and information retrieval are covered by the research topics in Figure 2. In section 4.1 and 4.2 we intuitively introduce the process of transforming data repositories into browsable FCA representations and performing query expansion and refinement operations. In section 4.3 and 4.4, the 28 % of papers on using FCA for representation of and navigation in image, service, web, etc. document collections are described. Defining and processing complex queries covers 6% of the papers and is described in section 4.5. Section 4.6 summarizes the papers on contextual answers (6% of papers) and ranking of query results (6% of papers).

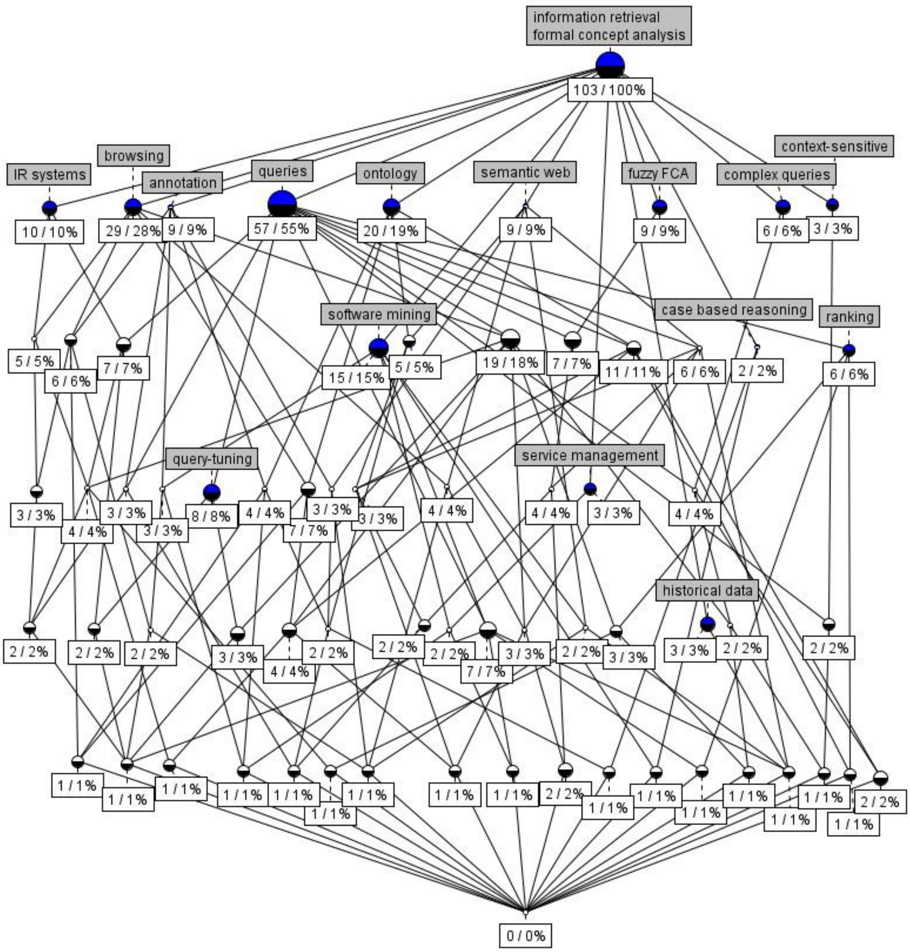

Fig. 2. Lattice containing 103 papers on using FCA in IR

4.1 Knowledge Representation and Browsing with FCA

In 28 % of the 103 selected papers, FCA is used for browsing and navigation through document collections. In more than half of these papers (18% of total number of papers), a combination of navigation and querying based on the FCA lattices is proposed. Annotation of documents and finding optimal document descriptors play an important role in effective information retrieval (9% of papers). All FCA-based approaches for information retrieval and browsing through large data repositories are based on the same underlying model. We first have the set G containing objects such as web pages, web services, images or other digitally available items. The set A of attributes can consist of terms, tags, descriptions, etc. These attributes can be related to certain objects through a relation $I \subseteq G \times M$ which indicates the terms, tags, etc.

can be used to describe the data elements in G. This triple (G, M, I) is a formal context from which the concept lattice can be created. The mathematical details of such a concept lattice are described in section 2.1. The process of obtaining a browsable FCA representation from such data is displayed in Fig. 3.

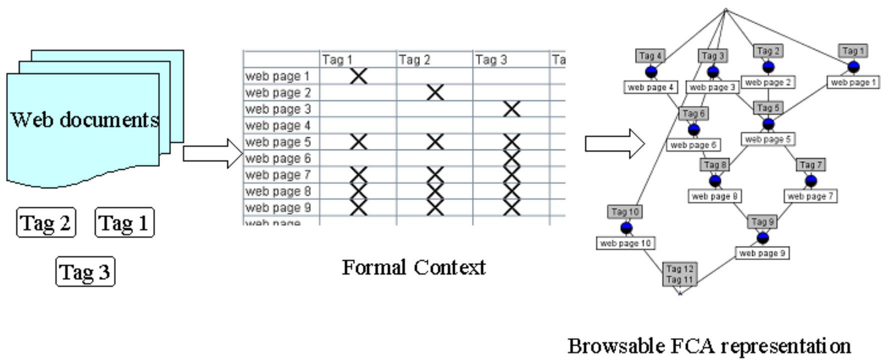

Fig. 3. Data transformation process

4.2 Query Result Improvement with FCA

Search engines are increasingly being used by amongst others web users who have an information need. The intent of a concept in an FCA lattice corresponds to a query and the extent contains the search results. A query $\wedge A$ uses a set of terms A and the system returns the answer by evaluating A'. Upon evaluating a query $\wedge A$ the system places itself on the concept (A', A'') which becomes the current concept c. For example in Fig. 4, the intent of the current concept A_c = {Tag 1, Tag 2, Tag 3, Tag 4, Tag 5, Tag 6, Tag 8} and the extent of the current concept O_c = {web page 8, web page 9}. Since a query provided by a user only approximates a user's need, many techniques have been developed to expand and refine query terms and search results. Query tuning is the process of searching for the query that best approximates the information need of the user. Query refinements can help the user express his original need more clearly. Query refinement can be done by going to a lower neighbor of the current concept in the lattice by adding a new term to the query items. The user can navigate for example to a subconcept $((A_c \cup \{t\})', (A_c \cup \{t\})'')$ by adding term t.

Query enlargement, i.e. retrieving additional relevant web pages, can be performed by navigating to an upper neighbor of the current concept in the lattice by removing a term from the query items. The user can navigate for example to a superconcept $((O_c \cup \{o\})'', (O_c \cup \{o\})')$ by adding object o. The combination of subsequent refine and expand operations can be seen as navigation through the query space. Typically, navigation and querying are two completely separate processes, and the combination of both results in a more flexible and user-friendly method. These topics are investigated in 8 % of the IR papers.

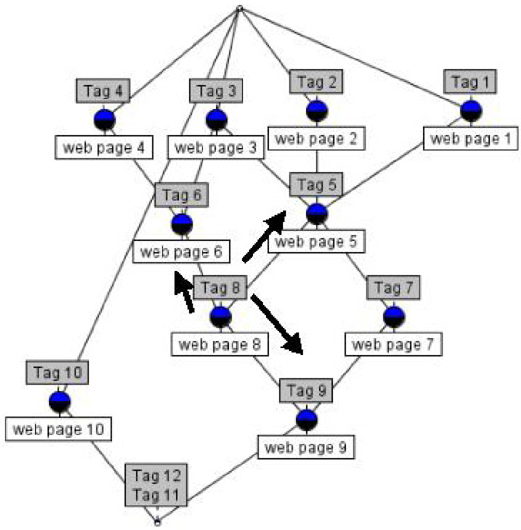

Fig. 4. Query tuning: from the current concept: an upward arrow for query expansion and a downward arrow for query refinement

4.3 Web and Email Retrieval

FCA has been used as the basis for many web-based knowledge browsing systems developed during the past years. Especially its comprehensible visualization capabilities seem to be of interest to the authors of these papers.

The results returned by web search engines for a given query are typically formatted as a list of URLs accompanied by a document title and a short summary of the document. Several FCA based systems were developed for analyzing and exploring these search results. CREDO (Carpineto et al. 2004), FooCA (Koester 2005, Koester 2006) and SearchSleuth (Ducrou et al. 2007, Dau et al. 2008) build a context for each individual query which contains the result of the query as objects and the terms found in the title and summary of each result as attributes. The CREDO system then builds an iceberg lattice which is represented as a tree and can be interactively explored by the user. FooCA shows the entire formal context to the user and offers a great degree of flexibility in exploring this table using the ranking of attributes, selecting the number of objects and attributes, applying stemming and stop word removal etc. SearchSleuth does not display the entire lattice but focuses on the search concept, i.e. the concept derived from the query terms. The user can easily navigate to its upper and lower neighbors and siblings. Nauer et al. (2009) also propose to use FCA for iteratively and interactively analyzing web search results. The user can indicate which concepts are relevant and which ones are not for the retrieval task. Based on this information the concept lattice is dynamically modified. Their research resulted in the CreChainDo system. Kim et al. (2004) presented the FCA-based document navigation system KAnavigator for small web communities in specialized domains. Relevant documents can be annotated with keywords by the users. Kim et al. (2006) extended the search functionality by combining lattice-based browsing with conceptual scales to reduce the complexity of the visualization. Cigarran et al. (2004) present the JBraindead IR System which combines free-text search with FCA to organize the results of a query.

Cole et al. (2003) discuss a document discovery tool named Conceptual Email Manager (CEM) which is based on FCA. The program allows users to navigate through emails using a visual lattice. The paper also discusses how conceptual ontologies can support traditional document retrieval systems and aid knowledge discovery in document collections. The development of this software is based on earlier research on retrieval of information from semi-structured texts (Cole et al. 2001, Cole et al. 2000). Building further on this work is the Mail-Sleuth software (Eklund et al. 2004) which can be used to mine large email archives. Eklund et al. (2005) use FCA for displaying, searching and navigating through help content in a help system.

Stojanovic (2005) present an FCA-based method for query refinement that provides a user with the queries that are "nearby" the given query. Their approach for query space navigation was validated in the context of searching medical abstracts. Stojanovic (2004) presents the SMART system for navigation through an on-line product catalog. The products in the database are described by elements of an ontology and visualized with a lattice, in which users can navigate from a very general product-attribute cluster containing a lot of products to very specific clusters that seem to contain a few, but for the user highly relevant products. Spyratos et al. (2006) describe an approach for query tuning that integrates navigation and querying into a single process. The FCA lattice serves for navigation and the attributes for query formulation. Le Grand et al. (2006) present an IR method based on FCA in conjunction with semantics to provide contextual answers to web queries. An overall lattice is built from tourism web pages. Then, users formulate their query and the best-matching concepts are returned, users may then navigate within the lattice by generalizing or refining their query. Eklund et al. (2008) present AnnotationSleuth to extend a standard search and browsing interface to feature a conceptual neighborhood centered on a formal concept derived from curatorial tags in a museum management system.

Cigarran et al. (2005) focus on the automatic selection of noun phrases as documents descriptors to build an FCA based IR system. Automatic attribute selection is important when using FCA in a free text document retrieval framework. Optimal attributes as document descriptors should produce smaller, clearer and more browsable concept lattices with better clustering features. Garcia et al. (2006) use FCA to perform semantic annotation of web pages with domain ontologies. Similarity matching techniques from Case Based Reasoning can be applied to retrieve these annotated pages as cases. Liu et al. (2007) use FCA to optimize a personal news search engine to help users obtain the news content they need rapidly. The proposed technique combines the construction of user background using FCA, the optimization of query keywords based on the user's background and a new layout strategy of search results based on a "Concept Tree". Lungley et al. (2009) use implicit user feedback for adapting the underlying domain model of an intranet search system. FCA is used as an interactive interface to identify query refinement terms which help achieve better document descriptions and more browsable lattices.

4.4 Image, Software and Knowledge Base Retrieval

Another domain in which FCA has been applied as an information retrieval technique is software engineering. Efficient service management including the classification, semantic annotation and retrieval of web services are important challenges in service-centric software engineering. Peng et al. (2005) present a method for generating a concept lattice depicting conceptual relationships between web services and to accurately retrieve web services from these lattices. Bruno et al. (2005) propose an approach based on FCA and Support Vector Machines to automatically identify key concepts inside service textual documentation, build a lattice based on these service annotations and classify services to specific domains. Poshyvanyk et al. (2007) use a combination of FCA and Latent Semantic Indexing (LSI) for concept location in the source code of Eclipse. LSI is used to map the concepts expressed in queries to relevant parts of the source code. The result is a ranked list of source code elements, organized in an FCA lattice. Muangon et al. (2009) combine FCA with Case Based Reasoning (CBR) for choosing appropriate design patterns for a specific design problem. This approach solves some of the problems of existing design pattern search methods using keyword-search. Design patterns are applied to solve recurring software design problems. Peng et al. (2007) propose a method for the incremental construction of a component retrieval ontology based on FCA. The ontology contains the characterizations of the components stored in the repository.

Ahmad et al. (2003) build concept lattices from descriptions associated to images for searching and retrieving relevant images from a database. In the ImageSleuth project (Ducrou et al. 2006), FCA was also used for clustering of and navigation through annotated collections of images. The lattice diagram is not directly shown to the user. Only the extent of the present concept containing thumbnails, the intent containing image descriptions and a list of upper and lower neighbors is shown. In Ducrou (2007), the author built an information space from the Amazon.com online store and used FCA to discover conceptually similar DVDs and explore their conceptual neighborhood. The system was called DVDSleuth. Amato et al. (2008) start from an initial image given by the user and use a concept lattice for retrieving similar images. The attributes in this lattice are facets, i.e. an image similarity criterion based on e.g. texture, color or shape. The values in the context indicate for each facet how similar an image in the database is with respect to the user provided initial image. By querying, the user can jump to any cluster of the lattice by specifying the criteria that the sought cluster must satisfy. By navigation from any cluster, the user can move to a neighbor cluster, thus exploiting the ordering amongst clusters.

Ducrou et al. (2005b) presented an FCA-based application, D-SIFT, for exploring relational database schema. Tane et al. (2005) introduced the query-based multi context theory, which allows defining a virtual space of FCA-based views on ontological data. Tane et al. (2006) discuss the benefits of the browsing framework for knowledge bases based on supporting the user in defining pertinent views. Hachani et al. (2009) use fuzzy FCA to explain the reasons of a failed database query and generate the nearest subqueries with non-empty answers.

4.5 Defining and Processing Complex Queries with FCA

Multiple techniques have been developed to define and process complex queries and to integrate data coming from heterogeneous sources. This topic is discussed in 6% of the IR papers. De Souza et al. (2004) use FCA for processing user queries over a set of overlapping ontologies, which have been created by independent groups adopting different configurations for ontology concepts. For example in bioinformatics, it is often difficult to relate the resources with a user query since the query needs to be processed and distributed over several heterogeneous data sources. Messai et al. (2005) present an approach based on FCA to search relevant bioinformatics data sources for a given user query. Nafkha et al. (2005b) investigate the possibilities of using FCA for searching similar objects in heterogeneous information sources. Pollailon et al. (2007) present a method based on FCA to provide contextual answers to user's queries from and to help their navigation in heterogeneous data sources. Ceravolo et al. (2007) use FCA for matching and mapping elements from heterogeneous data sources to a common ontology. Hitzler et al. (2006) present a new query language which allows querying formal contexts by means of logic programs written over attributes and objects.

4.6 Domain Knowledge in Search Results: Contextual Answers and Ranking

In this section, we discuss some of the techniques devised to provide contextual answers to user's queries and to incorporate domain knowledge into the organization of search results. Amongst others Carpineto et al. (2005) state that the main advantage of FCA for IR is the possibility of eliciting context and giving contextual answers to user’s queries. This topic is discussed in 9% of the IR papers. Please note that several papers in section 4.3 and 4.4 present methods which provide the user contextual answers, here we focus on ranking of query search results. Several researchers use FCA lattices for measuring query-document relevance, i.e. concept lattice-based ranking (CLR). Messai et al. (2008) partially order the set of attributes with respect to their importance. This hierarchy represents domain knowledge used to improve lattice-based querying and navigation. Hierarchies of attributes are used to define complex queries containing attributes with different levels of importance. Zhang et al. (2008) propose a method based on FCA to build a two-level hierarchy for retrieved search results of a query to facilitate browsing the collection. After formal concepts are extracted using FCA, the concepts most relevant to the query are further extracted. Finally, Ignatov et al. (2009) use FCA for near-duplicate detection in web search results.

5 Conclusions

Since its introduction in 1982 as a mathematical technique, FCA became a well-known instrument in computer science. Over 700 papers have been published over the past 7 years on FCA and 103 of them showed the method’s usefulness for IR. This paper showcased the possibilities of FCA as a meta technique for categorizing the literature on concept analysis with particular focus on the IR field. The intuitive visual interface of the concept lattices allowed for an in-depth exploration of the main research topics. Information retrieval is an important domain in which FCA was found to be an interesting instrument for representation of and navigation in large document collections and multiple IR systems resulted from this research. Also in query tuning and providing contextual answers to user queries, FCA was found to be a useful technique. In the future, we hope that this compendium may serve to guide both practitioners and researchers to new and improved avenues for FCA in the IR field.

Acknowledgements. Jonas Poelmans is aspirant of the "Fonds voor Wetenschappelijk Onderzoek – Vlaanderen" or "Research Foundation Flanders".

References

1. Ahmad, I., Jang, T.S.: Old Fashion Text-Based Image Retrieval Using FCA. In: Proc. IEEE Int. Conf. Image Processing, ICIP-III, vol. 2, pp. 33–36 (2003)
2. Amato, G., Meghini, C.: Faceted Content-based Image Retrieval. In: Proc. 19th IEEE Int. Conf. on Database and Expert Systems Application, DEXA, pp. 402–406. (2008)
3. Bruno, M., Canfora, G., Penta, M.D., Scognamiglio, R.: An Approach to support Web Service Classification and Annotation. In: Proc. IEEE Int. Conf. on e-Technology, e-Commerce and e-Service, pp. 138–143 (2005)
4. Carpineto, C., Romano, G.: A lattice conceptual clustering system and its application to browsing retrieval. Machine Learning 24(2), 1–28 (1996b)
5. Carpineto, C., Romano, G.: Concept data analysis: Theory and applications. John Wiley & Sons (2004a)
6. Carpineto, C., Romano, G.: Exploiting the Potential of Concept Lattices for Information Retrieval with CREDO. J. of Universal Computing 10(8), 985–1013 (2004b)
7. Carpineto, C., Romano, G.: Using Concept Lattices for Text Retrieval and Mining. In: Ganter, B., Stumme, G., Wille, R. (eds.) ICFCA 2005. LNCS (LNAI), vol. 3626, pp. 161–179. Springer, Heidelberg (2005)
8. Ceravolo, P., Gusmini, A., Leida, M., Cui, Z.: An FCA-based mapping generator. In: 12th IEEE int. Conf. on Emerging Technologies and Factory Automation, pp. 796–803 (2007)
9. Cigarrán, J.M., Gonzalo, J., Peñas, A., Verdejo, M.F.: Browsing Search Results via Formal Concept Analysis: Automatic Selection of Attributes. In: Eklund, P. (ed.) ICFCA 2004. LNCS (LNAI), vol. 2961, pp. 74–87. Springer, Heidelberg (2004)
10. Cigarrán, J.M., Peñas, A., Gonzalo, J., Verdejo, M.F.: Automatic Selection of Noun Phrases as Document Descriptors in an FCA-Based Information Retrieval System. In: Ganter, B., Godin, R. (eds.) ICFCA 2005. LNCS (LNAI), vol. 3403, pp. 49–63. Springer, Heidelberg (2005)
11. Cole, R., Eklund, P.: Browsing Semi-structured Web Texts Using Formal Concept Analysis. In: Delugach, H.S., Stumme, G. (eds.) ICCS 2001. LNCS (LNAI), vol. 2120, pp. 319–332. Springer, Heidelberg (2001)
12. Cole, R., Eklund, P., Stumme, G.: Document retrieval for e-mail search and discovery using Formal Concept Analysis. In: Applied Artificial Intelligence, vol. 17, pp. 257–280. Taylor & Francis (2003)
13. Cole, R.J.: The management and visualization of document collections using Formal Concept Analysis. Ph. D. Thesis, Griffith University (2000)

14. Ignatov, D.I., Kuznetsov, S.O.: Frequent Itemset Mining for Clustering Near Duplicate Web Documents. In: Rudolph, S., Dau, F., Kuznetsov, S.O. (eds.) ICCS 2009. LNCS (LNAI), vol. 5662, pp. 185–200. Springer, Heidelberg (2009)
15. Dau, F., Ducrou, J., Eklund, P.: Concept Similarity and Related Categories in SearchSleuth. In: Eklund, P., Haemmerlé, O. (eds.) ICCS 2008. LNCS (LNAI), vol. 5113, pp. 255–268. Springer, Heidelberg (2008)
16. De Souza, K.X.S., Davis, J.: Using an Aligned Ontology to Process User Queries. In: Bussler, C.J., Fensel, D. (eds.) AIMSA 2004. LNCS (LNAI), vol. 3192, pp. 44–53. Springer, Heidelberg (2004)
17. Ducrou, J.: DVDSleuth: A Case Study in Applied Formal Concept Analysis for Navigating Web Catalogs. In: Priss, U., Polovina, S., Hill, R. (eds.) ICCS 2007. LNCS (LNAI), vol. 4604, pp. 496–500. Springer, Heidelberg (2007)
18. Ducrou, J., Vormbrock, B., Eklund, P.: FCA-Based Browsing and Searching of a Collection of Images. In: Schärfe, H., Hitzler, P., Øhrstrøm, P. (eds.) ICCS 2006. LNCS (LNAI), vol. 4068, pp. 203–214. Springer, Heidelberg (2006)
19. Ducrou, J., Eklund, P.W.: SearchSleuth: The Conceptual Neighborhood of an Web Query. In: CLA (2007b)
20. Ducrou, J., Wormuth, B., Eklund, P.: Dynamic Schema Navigation Using Formal Concept Analysis. In: Tjoa, A.M., Trujillo, J. (eds.) DaWaK 2005. LNCS, vol. 3589, pp. 398–407. Springer, Heidelberg (2005b)
21. Eklund, P., Ducrou, J.: Navigation and Annotation with Formal Concept Analysis. In: Richards, D., Kang, B.-H. (eds.) PKAW 2008. LNCS, vol. 5465, pp. 118–121. Springer, Heidelberg (2009)
22. Eklund, P., Ducrou, J., Brawn, P.: Concept Lattices for Information Visualization: Can Novices Read Line-Diagrams? In: Eklund, P. (ed.) ICFCA 2004. LNCS (LNAI), vol. 2961, pp. 57–73. Springer, Heidelberg (2004)
23. Eklund, P., Wille, R.: Semantology as Basis for Conceptual Knowledge Processing. In: Kuznetsov, S.O., Schmidt, S. (eds.) ICFCA 2007. LNCS (LNAI), vol. 4390, pp. 18–38. Springer, Heidelberg (2007)
24. Eklund, P., Wormuth, B.: Restructuring Help Systems Using Formal Concept Analysis. In: Ganter, B., Godin, R. (eds.) ICFCA 2005. LNCS (LNAI), vol. 3403, pp. 129–144. Springer, Heidelberg (2005)
25. Ganter, B., Wille, R.: Formal Concept Analysis. Mathematical foundations. Springer (1999)
26. Recio-García, J.A., Gómez-Martín, M.A., Díaz-Agudo, B., González-Calero, P.A.: Improving Annotation in the Semantic Web and Case Authoring in Textual CBR. In: Roth-Berghofer, T.R., Göker, M.H., Güvenir, H.A. (eds.) ECCBR 2006. LNCS (LNAI), vol. 4106, pp. 226–240. Springer, Heidelberg (2006)
27. Godin, R., Gecsei, J., Pichet, C.: Design of browsing interface for information retrieval. In: Belkin, N.J., et al. (eds.) Proc. GIR, pp. 32–39 (1989)
28. Godin, R., Missaoui, R., April, A.: Experimental comparison of navigation in a Galois lattice with conventional information retrieval methods. Int. J. Man-Machine Studies 38, 747–767 (1993)
29. Hachani, N., Ben Hassine, M.A., Chettaoui, H., et al.: Cooperative answering of fuzzy queries. Journal of Computer Science and Technology 24(4), 675–686 (2009)
30. Hitzler, P., Krötzsch, M.: Querying Formal Contexts with Answer Set Programs. In: Schärfe, H., Hitzler, P., Øhrstrøm, P. (eds.) ICCS 2006. LNCS (LNAI), vol. 4068, pp. 260–273. Springer, Heidelberg (2006)

31. Ignatov, D.I., Kuznetsov, S.O.: Frequent Itemset Mining for Clustering Near Duplicate Web Documents. In: Rudolph, S., Dau, F., Kuznetsov, S.O. (eds.) ICCS 2009. LNCS (LNAI), vol. 5662, pp. 185–200. Springer, Heidelberg (2009)
32. Kim, M., Compton, P.: Evolutionary Document Management and Retrieval for Specialised Domains on the Web. Int. J. of Human Computer Studies 60(2), 201–241 (2004)
33. Kim, M., Compton, P.: A Hybrid Browsing Mechanism Using Conceptual Scales. In: Hoffmann, A., Kang, B.-H., Richards, D., Tsumoto, S. (eds.) PKAW 2006. LNCS (LNAI), vol. 4303, pp. 132–143. Springer, Heidelberg (2006)
34. Koester, B.: Conceptual Knowledge Retrieval with FooCA: Improving Web Search Engine Results with Contexts and Concept Hierarchies. In: Perner, P. (ed.) ICDM 2006. LNCS (LNAI), vol. 4065, pp. 176–190. Springer, Heidelberg (2006)
35. Lakhal, L., Stumme, G.: Efficient Mining of Association Rules Based on Formal Concept Analysis. In: Ganter, B., Stumme, G., Wille, R. (eds.) ICFCA 2005. LNCS (LNAI), vol. 3626, pp. 180–195. Springer, Heidelberg (2005)
36. Le Grand, B., Aufaure, M.A., Soto, M.: Semantic and Conceptual Context-Aware Information Retrieval. In: Damiani, E., Yetongnon, K., Chbeir, R., Dipanda, A. (eds.) SITIS 2006. LNCS, vol. 4879, pp. 247–258. Springer, Heidelberg (2009)
37. Liu, M., Shao, M., Zhang, W., Wu, C.: Reduction method for concept lattices based on rough set theory and its application. Computers and Mathematics with Applications 53, 1390–1410 (2007)
38. Lungley, D., Kruschwitz, U.: Automatically Maintained Domain Knowledge: Initial Findings. In: Boughanem, M., Berrut, C., Mothe, J., Soule-Dupuy, C. (eds.) ECIR 2009. LNCS, vol. 5478, pp. 739–743. Springer, Heidelberg (2009)
39. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press (2008)
40. Messai, N., Devignes, M.D., Napoli, A., Smail-Tabbone, M.: Extending Attribute Dependencies for Lattice-based Querying and Navigation. In: Eklund, P., Haemmerlé, O. (eds.) ICCS 2008. LNCS (LNAI), vol. 5113, pp. 189–202. Springer, Heidelberg (2008)
41. Messai, N., Devignes, M.-D., Napoli, A., Smaïl-Tabbone, M.: Querying a Bioinformatic Data Sources Registry with Concept Lattices. In: Dau, F., Mugnier, M.-L., Stumme, G. (eds.) ICCS 2005. LNCS (LNAI), vol. 3596, pp. 323–336. Springer, Heidelberg (2005)
42. Mili, H., Ah-Ki, E., Godin, R., Mcheick, H.: Another nail to the coffin of faceted controlled-vocabulary component classification and retrieval. VCM SIGSOFT Software Engineering Notes 22(3), 89–98 (1997)
43. Muangon, W., Intakosum, S.: Retrieving Design Patterns by Case-Based Reasoning and Formal Concept Analysis. In: 2nd Int. Conf. Comp. Sc. Inf. Technology, pp. 424–428 (2009)
44. Nafkha, I., Jaoua, A.: Using Formal Concept Analysis for Heterogeneous Information. In: Belohlavek, R.R., et al. (eds.) CLA, pp. 107–122 (2005)
45. Nauer, E., Toussaint, Y.: CreChainDo: An iterative and interactive Web information retrieval system based on lattices. International Journal of General Systems 38(4), 363–378 (2009)
46. Peng, D., Huang, S., Wang, X., Zhou, A.: Concept-Based Retrieval of Alternate Web Services. In: Zhou, L.-z., Ooi, B.-C., Meng, X. (eds.) DASFAA 2005. LNCS, vol. 3453, pp. 359–371. Springer, Heidelberg (2005a)
47. Peng, X., Zhao, W.: An Incremental and FCA-based Ontology Construction Method for Semantics-based Component Retrieval. In: 7th Int. Conf. on Quality Soft, pp. 309–315 (2007)

48. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G.: Formal Concept Analysis in Knowledge Discovery: A Survey. In: Croitoru, M., Ferré, S., Lukose, D. (eds.) ICCS 2010. LNCS, vol. 6208, pp. 139–153. Springer, Heidelberg (2010)
49. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G.: Concept Discovery Innovations in Law Enforcement: a Perspective. In: IEEE CINS Workshop (INCos), Greece (2010b)
50. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G.: Curbing domestic violence: Instantiating C-K theory with Formal Concept Analysis and Emergent Self Organizing Maps. Intelligent Systems in Accounting, Finance and Management 17(3-4), 167–191 (2010c)
51. Polaillon, G., Aufaure, M.A., Le Grand, B., Soto, M.: FCA for contextual semantic navigation and information retrieval in heterogeneous information systems. In: 8th IEEE Int. Workshop on Database and Expert Systems Applications, pp. 534–539 (2007)
52. Poshyvanyk, D., Marcus, A.: Combining Formal Concept Analysis with Information Retrieval for Concept Location in Source Code. In: Proc. IEEE Int. Conf. on Program Comprehension, pp. 37–48 (2007)
53. Priss, U.: Lattice-based Information Retrieval. Knowledge Organization 27(3), 132–142 (2000)
54. Priss, U.: Formal Concept Analysis in Information Science. In: Blaise, C. (ed.) Annual Review of Information Science and Technology, ASIST, vol. 40, pp. 521–543 (2006)
55. Spyratos, N., Meghini, C.: Preference-Based Query Tuning Through Refinement/Enlargement in a Formal Context. In: Dix, J., Hegner, S.J. (eds.) FoIKS 2006. LNCS, vol. 3861, pp. 278–293. Springer, Heidelberg (2006)
56. Stojanovic, N.: On the query refinement in the ontology-based searching for information. Information Systems 30(7), 543–563 (2005)
57. Stojanovic, N.: On Using Query Neighborhood for Better Navigation through a Product Catalog: SMART Approach. In: IEEE Int. Conf. e-Tech., e-Com. and e-Service (2004)
58. Tane, J.: Using a Query-Based Multicontext for Knowledge Base Browsing. In: 3rd Int. Conf., ICFCA - Supplementary, Lens, France, pp. 62–78 (2005)
59. Tane, J., Cimiano, P., Hitzler, P.: Query-Based Multicontexts for Knowledge Base Browsing: An Evaluation. In: Schärfe, H., Hitzler, P., Øhrstrøm, P. (eds.) ICCS 2006. LNCS (LNAI), vol. 4068, pp. 413–426. Springer, Heidelberg (2006)
60. Tilley, T.: Tool Support for FCA. In: Eklund, P. (ed.) ICFCA 2004. LNCS (LNAI), vol. 2961, pp. 104–111. Springer, Heidelberg (2004)
61. Tilley, T., Eklund, P.: Citation analysis using Formal Concept Analysis: A case study in software engineering. In: 18th Int. Conf., DEXA, pp. 545–550 (2007)
62. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In: Rival, I. (ed.) Ordered Sets, pp. 445–470. Reidel, Dordrecht-Boston (1982)
63. Wille, R.: Methods of Conceptual Knowledge Processing. In: Missaoui, R., Schmidt, J. (eds.) ICFCA 2006. LNCS (LNAI), vol. 3874, pp. 1–29. Springer, Heidelberg (2006)
64. Zhang, Y., Feng, B., Xue, Y.: A New Search Results Clustering Algorithm based on Formal Concept Analysis. In: 5th Int. Conf. on FSKD, pp. 356–360 (2008)