A Framework for FCA-based Exploratory Web Search

Peter Butka\textsuperscript{1}, Thomas Low\textsuperscript{2}, Michael Kotzyba\textsuperscript{2}, Stefan Haun\textsuperscript{2}, and Andreas Nürnberg\textsuperscript{2}

\textsuperscript{1} Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Slovakia
\texttt{<firstname>.<lastname>@tuke.sk}

\textsuperscript{2} Data and Knowledge Engineering Group, Faculty of Computer Science, Otto von Guericke University Magdeburg, Germany
\texttt{<firstname>.<lastname>@ovgu.de}

Abstract. In this paper we present a conceptual framework for exploratory search by structuring web results based on concept lattices, which are created using methods from Formal Concept Analysis (FCA). The approach is to organize query based search engine results (web documents) by a hierarchy of clusters that are composed of documents with similar attributes. The resulting concept lattice provides a structured view on query-related domains and hence can improve the user’s understanding of document attributes and shared features. Instead of using the traditional approaches, we applied a fuzzy extension of FCA in order to support the usage of different types of attributes within the analyzed result set. Afterward the result set is presented using the graph visualization toolkit CET.

1 Introduction

In Information Retrieval it is still challenging to provide a structured view of retrieved results in order to provide a better navigation within the documents and hence easier exploration of the domain. While standard approaches usually follow simple, linear visualization of result sets, a well organized representation can help the users to understand the current domain of interest and to discover the relationship between the retrieved documents. To reach this goal, methods of exploratory data mining known as Formal Concept Analysis (FCA) [1] are promising. Users can be supported by a hierarchical structure of clusters of documents, which represent document subsets with shared attributes defined within the domain. This approach can be used to support a user during an exploratory search scenario [2, 3], where the user is not familiar with the topic and needs to get an overview. The framework is especially useful in a companion technology to support interactive information access and organization.

To analyze sets of web documents using FCA-based methods, it is necessary to prepare and map the input into an object-attribute model. That is, objects are documents from the query’s result set and attributes are describing elements
of these documents, e.g. the existence of a term within the document, its weight (TF-IDF), meta-data of the document, etc. Usually FCA-based methods are applied only on binary object-attribute models (so-called “crisp” case), where all complex attributes are transformed into binary attributes using some pre-processing methods like conceptual scaling. Due to the huge variety of web results and their content, a binary approach is not fully sufficient. The fuzzy approach used in our framework is able to work with different types of attributes, c.f. [4].

While it is common practice to just visualize the search results, such as concept lattices, it is beneficial to provide an interactive exploration to support the understanding of the search domain [5]. Regarding the visualization, our suggestion is to integrate FCA-based methods into existing graph visualization tools. In this work we present a demonstrator prototype of the proposed framework that is integrated into the Creative Exploration Toolkit (CET), c.f. [6]. The demonstrator allows to visualize concept lattices of query results and is designed to navigate them.

The paper is organized as follows: Section 2 provides information on related work in usage of FCA methods in information retrieval tasks and motivation to use fuzzy FCA approaches. Section 3 describes details regarding our proposed approach. At the end of the paper we add some conclusions and future work ideas.

2 Related Work and Motivation

A survey of FCA-based methods within the scope of information retrieval systems and tasks can be found in [7]. Usually the approaches are divided into three categories:

- **Direct usage of concept lattice structure for exploration:** The concept lattice is created and directly used for search and navigation in the space of hierarchically organized clusters of documents acquired from the query result set.
- **Use of reduction of concept lattice:** The original concept lattice is first reduced in order to simplify result sets for a more understandable visualization and is then used for search and navigation.
- **Use of methods from FCA as analytical function:** For supporting the process of linear (re)ordering of the result sets according to the structure of the concept lattice and its features.

Examples are the CREDO system [8] and its extension called CreChainDo [9]. The systems provide a hierarchy of document clusters returned to the users and their feedback on relevance of concepts (clusters) is used for adaption of the hierarchy. That is, the systems provide an interactive and adaptive exploration in documents from query result sets. The result presented to the user is actually a tree of the simplified, reduced subpart of the concept lattice built on such data. The extended version is able to support and manage chains of the interactive steps. A similar approach, that is using FCA lattices for navigation
and attributes for query tuning has been described in [10]. In [11] an approach for enhancement of keyword-based search has been proposed, where local FCA models with agglomerative clustering were used in order to obtain linear ranking of documents based on combination of full-text and conceptual search. More works on this topic can be found in the survey mentioned before.

The main disadvantage of these approaches is that only the classical framework for FCA is used, i.e., every description of objects and their sets is based on binary attributes. In our framework we implemented an one-sided fuzzy approach, where objects are considered binary as well\(^1\), but the attributes can obtain fuzzy values. In this case particular documents are described by the set of different types of attributes. That is, not only binary attributes are used, but also the documents within the query result set can be described by some nominal, ordinal, numeric, or general lattice attributes. This allows a more natural result description. Furthermore, the user does not need to specify the preprocessing and transformation (called conceptual scaling) into a binary input data table (which can be less interpretable for users) for the FCA building process. Here a fuzzy approach, like our generalized one-sided concept lattice approach, which is able to combine different types of attributes within one heterogeneous data table, can be very useful and provides a multitude of applications for interactive exploration.

### 3 FCA-supported Exploratory Web Search

To demonstrate the proposed framework, we use the already mentioned CET and integrate the fuzzy FCA approach into it. CET allows a graph visualization of created concept lattices from query result sets and/or its reduction or some projection. The user can interact with the graph model of query result set and explore the search domain. The CET provides an input for a search query, calls the search API, e.g. the Bing API by Microsoft, analyzes the results and generates the graph of nodes and edges. Usually, nodes in this tool are particular pages (documents) and edges are relationships found between some pages. In our extended version, first the concept lattice is generated from the retrieved web results and afterwards the lattice, which is a graph itself, is visualized. The CET tool is using the XMPP [12] protocol for communication with the backend functions in order to provide changes in the visualized graph. The provided graph is interactive and it is possible to change the visualization views, obtain information on nodes and edges, as well as change the global visual appearance of the graph, e.g. to reduce the provided level of detail.

We have integrated the FCA-based methods, especially the fuzzy version, into the CET. Furthermore, we have adapted and extended the FCA-based exploration to navigate through the retrieved web results. The demonstrator implements the following functions:

---

1 Defined clusters are subsets, i.e., an object is or is not part of a concept and concepts represent clusters
– **Modeling nodes as concepts within the concept lattice**: The lattice is generated according to a query result set. Nodes describe the content of the concept. Furthermore, the adapted CET tool can be used for visualization of more detailed information on particular concepts (clusters).

– **Modeling edges as a relation of generalization or specification within the concept lattice**: The lattice can be used in the traditional way using the extracted hierarchy (hierarchy of concepts = clusters).

– **Supporting interactive search and navigation within the subsets of results and their hierarchy**: The function includes highlighting of most important attributes (and their respective values) responsible for particular step in generalization/specialization (using edge description).

– **Using fuzzy extension of FCA in order to analyze the lattice**: Here different features like weighting of terms in results (vector model from extracted words) are used. Additionally, meta-data descriptions of different types, external evaluation of results by users, or any other more complex ordering of values within specific attributes can be integrated.

The result is an interactive exploration tool for an arbitrary search domain. All obtained search results are analyzed by an one-sided fuzzy concept lattice that allows the user to interact within. The results – clusters of documents from the query – are hierarchically organized according to different specific combinations.

![Illustration of the CET tool visualizing the first ten web results for the input query “snow leopard” organized as a concept lattice.](image)

**Fig. 1.** Illustration of the CET tool visualizing the first ten web results for the input query “snow leopard” organized as a concept lattice.
to different types of attributes. Figure 1 illustrates an example for the ambiguous query “snow leopard” (animal vs. operating system) generated from the first ten search engine results. The concept at the top consists of all ten web results, because each result contains both terms “snow” and “leopard”. Furthermore, below the upper concept there are two concepts on the right side with two web results named “apple” and “animal”. The web documents in these concepts share attributes (terms) related to the operating system from Apple or the animal respectively. All other concepts only contain one web result or less, because their attributes are specific and no other result shares the same attributes. At the bottom the concept with all attributes is placed. Since no web result contains all attribute terms like “snow”, “leopard”, “store”, “support” and “animal” together, this concept is empty.

4 Conclusions and Future Work

In this paper we presented the framework for a FCA-based exploratory search that generates a structured view on search engine results. A structured view can provide a better navigation within the documents and hence easier exploration of the domain. The framework was implemented as a demonstrator that allows an interactive visualization of concept lattices created by a fuzzy FCA approach, which is able to take different types of attributes into account. The demonstrator is based on the integration with the graph visualization tool CET to explore the query results.

For future work the framework will be extended and the demonstrator will be finalized. This includes the extraction of more complex attributes from query result sets, the integration of such extracted data table with fuzzy FCA algorithm, a detailed description of elements and edges in the interface, improved design of visualized elements, as well as more processes specific interactions with the concept lattices for exploration. After finalization, the tool will be tested in user studies to investigate the users experience and to investigate more ways to support the users by navigating the concept lattice and hence understanding the search domain.

Acknowledgments

This work was supported by the Transregional Collaborative Research Centre SFB/TRR 62 “A Companion-Technology for Cognitive Technical Systems” funded by the German Research Foundation (DFG).

References