# Conceptual Knowledge Retrieval with FooCA: Improving Web Search Engine Results with Contexts and Concept Hierarchies

Bjoern Koester

Webstrategy GmbH, Darmstadt, Germany
bjoern.koester@webstrategy.de
http://www.webstrategy.de

**Abstract.** This paper presents a new approach to accessing information on the Web. FooCA, an application in the field of Conceptual Knowledge Processing, is introduced to support a holistic representation of today's standard sequential Web search engine retrieval results. FooCA uses the itemset consisting of the title, a short description, and the URL to build a context and the appropriate concept hierarchy. In order to generate a nicely arranged concept hierarchy using line diagrams to retrieve and analyze the data, the prior context can be iteratively explored and enhanced. The combination of Web Mining techniques and Formal Concept Analysis (FCA) with contextual attribute elicitation gives the user more insight and more options than a traditional search engine interface. Besides serving as a tool for holistic data exploration, FooCA also enables the regular user to learn step by step how to run new, optimized search queries for his personal information need on the Web.

## 1 Introduction

Access to information through Web data plays an important role today [13]. While facing a rapidly growing flood of information on the World Wide Web, we see an increasing need for advanced tools that guide us to the kind of information we are looking for. Retrieval results of major search engines are growing every day. Especially searches for general terms usually end up with over one million results.

Judging over a larger, sequential list of results is almost impossible with the current user interfaces provided by major Web search engines today. Inspired by a discussion with Prof. Rudolf Wille who criticized the way society has been 'googlelized' by limiting searches to the first few pages retrieved via a regular Web search, this work aims at introducing an alternative way of retrieving Web search results.

The challenge for this explorative work was to see whether Google's three-row result itemset consisting of the title, a short description (a so-called snippet), and the URL could be used to build a meaningful context and an appropriate concept hierarchy.

In earlier stages of the project ([17], [18]), it had turned out that the presentation of the retrieved search engine results within a context already gives a valuable and holistic overview of information units and their relations. Now, the additional perspective of the corresponding concept hierarchy further improved and elaborated this effort.

This paper presents a new approach to accessing information on the Web. Conceptual Knowledge Processing is derived from a pragmatic understanding of knowledge according to which human knowledge is acquired and supported in a process of human thinking, reasoning, and communicating [23]. Methods and tools supporting conscious reflection, judgment and action are proposed that facilitate such a form of information gathering. This understanding of knowledge processing serves as the background of FooCA.

FooCA stands for a word combination of FCA and Google, the described FCA mechanisms having first been applied to a standard Google search. Since all major search engines do not differ significantly in the way they present their search results, FooCA aims at supporting a general approach for search engines.

The first part of FooCA uses a regular Web search engine for the retrieval of potential Web objects[1]. Giving autonomy and thus the decision-making ability back to the user, the second part is an interactive interface that allows the searcher to participate in selecting, filtering and choosing the best methods for customizing the results according to his information need. The user can choose among multiple search strategies and select those preferences best suited for his search interest. Depending on the strategies he has chosen, the search can range between being more accurate but slower, and less accurate but faster. After sending his search request the system assists the user in further refining his search query.

Search engine retrievals will be visualized first in a context to assist in further refining the search and reducing the complexity. Later, when the context has been successfully refined, the data can be visualized by a line diagram to establish a concept hierarchy.

Many expressions used in this paper are based on the terminology of Conceptual Knowledge Processing introduced in [24]. In general, when speaking about a *concept lattice*, we will refer to it by using the more general term *concept hierarchy*.

## 2   Formal Concept Analysis (FCA)

Formal Concept Analysis aims at gaining concepts and hierarchical implications out of data [14]. FCA studies how objects can be hierarchically grouped together according to their common attributes. We give a brief introduction.

---

[1] The term 'Web objects' is used throughout this paper to address various types of media that can be accessed via a URL (Unified Resource Locator). These are mainly Web pages, PDF documents or images.

**Formal Context.** A formal context $\mathbb{K} := (G, M, I)$ is composed of a set of objects $G$, a set of attributes $M$, and a binary relation $I \subseteq G \times M$. We call $I$ the 'incidence relation' and read $(g, m) \in I$ as 'object g has the attribute m'.

A context $\mathbb{K}$ can be visualized by a two dimensional table, a cross table, containing crosses which indicate the incidence relation.

**Derivation Operators.** For a subset $A \subseteq G$ of the objects we define the set of attributes common to the objects in $A$ as

$$A' := \{m \in M \mid gIm \ for \ all \ g \in A\} \tag{1}$$

respectively, for a subset $B \subseteq M$ of the attributes we define a set of objects which have all attributes in $B$ as

$$B' := \{g \in G \mid gIm \ for \ all \ m \in B\} \tag{2}$$

**From a Galois Connection to a Formal Concept.** The pair of the derivation operators form a *Galois connection*. Thus the following statements are true for a given context $(G, M, I)$, its subsets $A, A_1, A_2 \subseteq G$ of objects as well as its subsets $B, B_1, B_2 \subseteq M$ of attributes:

$$A_1 \subseteq A_2 \Rightarrow A_2' \subseteq A_1' \ \text{and} \ B_1 \subseteq B_2 \Rightarrow B_2' \subseteq B_1' \tag{3}$$

$$A \subseteq A'' \ \text{and} \ B \subseteq B'' \tag{4}$$

$$A' = A''' \ \text{and} \ B' = B''' \tag{5}$$

$$A \subseteq B' \Leftrightarrow B \subseteq A' \Leftrightarrow A \times B \subseteq I \tag{6}$$

Both, the derivation operators and the formed Galois connection now allow us to define a formal concept as follows.

**Formal Concept.** In a philosophical sense a concept consists of two parts: the extension and the intension. The extension covers all objects belonging to this concept and the intension comprises all attributes valid for all those objects.

A formal concept of the corresponding context $\mathbb{K} := (G, M, I)$ is composed of a pair $(A, B)$ consisting of an extension $A \subseteq G$ and an intension $B \subseteq M$, to which the following apply:

$$A' = B \ \text{and} \ B' = A \tag{7}$$

We denote by $\mathfrak{B}(G, M, I)$ the set of all concepts of the context $(G, M, I)$ and write $\mathfrak{B}(\mathbb{K})$ for short.

**Concept Hierarchy.** An important structure can be obtained by defining a subconcept-superconcept relation building a formal order relation on $\mathfrak{B}(G, M, I)$ which then enables us to form a mathematical lattice denoted by $\underline{\mathfrak{B}}(G, M, I)$. Such a lattice structure[2] can be visualized by line diagrams, as shown in Figure 2. A further mathematical introduction into these structures would lead far beyond the scope of this paper. Readers interested in a deeper insight into FCA are referred to [14].

## 3  Retrieval Via a Standard Web Search Engine

Our approach[3] is to launch a search request using the official programming interfaces provided by the search engines. The returned set of ranked items is then analyzed by means of Web Mining and FCA.

We first obtain ranked items that match the criteria of the internal and unrevealed ranking algorithms of the search engine. Since the search engine market has become competitive, the exact algorithms used for the ranking of Web results are classified.

### 3.1  Query Evaluation

For our frequent use of the term 'query' and its fundamental operations in the context of FooCA, we would like to propose a short introduction first.

**Query.** A query $Q_i$ is a list of terms forming a request for information from a database. In almost all Web search engines, the entered query is parsed and segmented into terms. Adding additional terms to the query means that each term should appear in a document. We would like to simplify our understanding of a query as a list of terms with two operations only, the concatenation and explicit exclusion of terms:

**Implicit AND Operator.** The query 'data mining' results in a search for pages containing the words 'data' and 'mining'. Since there is no 'and' operator between the terms, this notation is called implicit AND. The focus of an AND query is to add more terms to establish a 'list of terms' that specializes or characterizes the search concept.

---

[2] We prefer using the more general term *concept hierarchy* instead, as this paper would like to address and inspire a broader group of people outside of abstract algebra as well.

[3] One other way of retrieving Web objects is to implement just another Web crawling agent that spiders and indexes all Web pages it can possibly find. But this approach would involve not only considerable cost through having to run the hardware and storage systems but also intensified internet traffic. In addition, it would take some time to make a crawler robust enough and the indexing fast enough. And then, we would lose a lot of time implementing a general retrieval task as it has already been done by so many institutions and companies worldwide.

**Query Term Negation.** Instead of adding terms to a query, it sometimes seems necessary to avoid the appearance of specific terms. As an example, a query for the term 'data mining' will return a lot of Web objects about data and mining in general (also mineral mining). Assuming that we are primarily interested in data mining, we could use the negation operator to exclude a term from the search by simply adding '-minerals' to our search query resulting in the following query that better suits our information need: 'data mining -minerals'. All pages returned contain 'data' and 'mining' and no instance of the term 'minerals'.

### 3.2   From Ranked Lists to Conceptual Structures

Considering the fact that we obtain a ranked set of documents which ran through a completely automated machinery, how would the automated system know exactly what the user really wants? A user might search for a term and get back more than one million documents matching his query. Now he has the option of clicking sequentially through all the results lists, checking each item and hopefully finding the relevant pages he was looking for. An alternative would be to further specify the query to limit results. This, however, requires a degree of familiarity with techniques for narrowing down a search query.

Hence, new ways of controlling and obtaining an overview of information need to be established to guide and assist the user instead of ignoring human skills, such as the intuitive understanding of a concept.

This paper proposes a vision as well as a working prototype that shows how traditional, sequential lists can indeed be replaced by conceptual structures.

## 4   Conceptual Knowledge Retrieval with **FooCA**

As introduced above, FooCA uses the search facilities provided by a standard Web search engine and enriches them by assisting the user and enabling him to control the search in more advanced ways.

### 4.1   Architecture

FooCA runs on a standard Linux system and is written in Perl. It communicates with the search engines using their official APIs.[4]

In a query result, we are interested in the general items returned by the Web search engine to a normal user: the title of the Web document, its URL, and the snippet that represents the semantic context in which the search query (or parts of it) occurred.

As shown in Figure 1, FooCA enables the user to interact with the search engine. The user has a specific but informal concept in mind that he wants to

---

[4] More information on the API for Google can be obtained at `http://www.google.com/apis/reference.html`, the API for Yahoo can be downloaded at `http://developer.yahoo.net`.
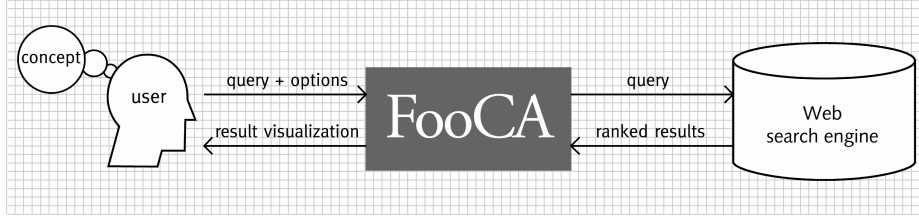
**Fig. 1.** FooCA interacts between the user and the search engine. Queries are passed to the search engine along with additional options. The results are then post-processed for visualization to the user.

search for. He then enters an approximate query that represents his concept, along with chosen options for later FooCA processing. FooCA then receives that information, evaluates the personal options for that user and re-submits his query to the search engine which interprets the query as if it were a normal search and processes it, returning the results in a ranked order to FooCA. Using the personal options and the search results retrieved, FooCA now generates its internal representation of the context and presents it to the user in a visualized form. From this point on, the user can refine his search with the FooCA interface.

### 4.2   Basic Feature Extraction Operations

In order to process attributes within FooCA, we need to identify word tokens. Therefore, some basic feature extraction operations are applied to the standard Web search engine retrieval results: Identification of tokens divided by space or punctuation, stripping of all HTML format tags, transforming all characters from upper- to lower-case and finally removing all special characters except for '-'. Using these basic operations,[5] a list of useful word tokens originating from the snippet can easily be generated.

### 4.3   From a Web Search Engine Retrieval to a Context

FooCA lets the user enter a query $Q_i$ which is directly passed on to the Web search engine without modification. Web search engines usually offer the user a short excerpt of words before and after the occurrence of a term that appears in the query. This short excerpt is called a 'snippet'. The idea is to use that snippet as a starting point since it provides us with a short, non-formal context in which the search query (or parts of it) are embedded. In cases where no snippet is retrieved, the page title is used instead. After extracting feature terms from the retrieved snippets, we gain a context $\mathbb{K}$ considering the URLs as objects $G$ and the extracted feature terms as attributes $M$ as follows: $\mathbb{K}(Q_i) := (G, M, I)$.

---

[5] When using FooCA in German, an additional processing step to rewrite the German umlauts such as 'ö' or 'ä' to 'oe' and 'ae' is provided to map different writings of the same tokens onto one single attribute.

### 4.4  Representing the Context in a Cross Table

Once the user has enabled or disabled specific search strategies and entered the search query, FooCA presents the retrieved results in an interactive two-dimensional cross table. The row headers are object names, which are clickable numbers in our case, representing the ranked Web object search results. Columns are headed by attributes which are the extracted feature terms of the snippets. The incidence relation $I$ of the context $\mathbb{K}$ between the Web objects $G$ and its attributes $M$ is marked by a cross '$\times$' in the table. The $i$th Web object possesses the $j$th attribute indicated by a cross in the $(i, j)$ position.

The cross table can be navigated using the mouse. As the mouse cursor moves along the table, the underlying row is highlighted indicating the Web object. The user can click anywhere inside the table and is promptly directed to the related Web object.

Apart from navigating inside the table by way of the incidence relation of the context, another navigation method using query refinement is offered. The user can click on any listed attribute name in order to either search for that name directly, launching a new query with that attribute only, or he can include or exclude an attribute by further qualifying the previous query.

### 4.5  Search Preferences and Strategies

To cope with the rising flood of information in almost every field, the user has surrendered a part of his own authority of judgment to an automated evaluation process that makes decisions and acts based on certain rules.

FooCA gives the searcher the possibility to gain more control over the decision-making process by using methods and operations that are usually processed automatically in general search engines without human intervention. Those methods and operations provided by the FooCA application are described in the following subsections:

**Choice of Search Engine.** The user has the choice between a set of pre-integrated Web search engines. In fact, any search engine providing a sequential listing of search retrievals and a snippet can be integrated into FooCA by building a specific interface. As for the moment, interfaces for the two most popular search engines, Google and Yahoo, have been deployed successfully and tested thoroughly. More interfaces will be provided in the short term.

By using different search engines as the input source of the FooCA system, various different contexts can be generated and explored.

**Language Restriction.** Both, Google and Yahoo offer the possibility to restrict the search to a predefined subset of its Web object index. FooCA supports the search in English and German. However, depending on the facilities of the underlying search engine, other languages could easily be integrated as well.

**Removal of Stop Words.** Common words in a language that occur frequently but are insignificant for searching purposes can simply be removed. English stop words are, for instance, 'I', 'you', 'are', 'it', 'the', 'and', 'on' and 'of'. FooCA provides a list of general stop words for English and German.

**Stemming.** Stemming means reducing words to their respective grammatical root. This is accomplished either by using a rule-based stemming algorithm (such as the Porter stemmer [20]) or a knowledge base consisting of words and their related root forms. FooCA uses Porter stemming.

**Clarification of the Context.** The context $\mathbb{K} := (G, M, I)$ can be attribute-clarified to a context $(G, M/\sim, \widetilde{I})$, where $\sim$ is the equivalance relation with $m \sim n :\Longleftrightarrow m' \sim n'$. A corresponding clarification of objects cannot be implemented due to the line-based navigation within the cross table which associates each line with a distinct Web object.

**User-Based Query Refinement.** In FooCA, we understand a query $Q_i$ to be a set of attributes of our attribute set $M$, $Q_i \subseteq M$. Although this is a very simplified view on queries – we are not considering any advanced operations here – we gain a lot of power by simply adding and removing attributes from an attribute set.

By letting the user decide about the importance or unimportance of the attributes presented, the system enables him to refine the search space accordingly and trigger a new information retrieval process. Just as in the process of adding new query terms into the Web search engine form, the user can simply click an attribute representing a query term and decide to either include or exclude that term in a new search process. The main difference between the FooCA-based refinement process for a query and the manual refinement using the original Web search interfaces consists in the set of given attributes. Typically, the user is not entirely clear from the beginning which term(s) are needed to narrow a specific search. With FooCA presenting the attributes in an attribute-object relation, the user is able to inspect the context and make his decision based on that specific contextual knowledge. Furthermore, new relationships or erroneous relationships become apparent in a holistic way.

In our approach, removed attributes are not actually removed from the retrieved document collection; instead a new query is defined explicitly excluding that removed attribute. A new context for the new, refined query is then generated and displayed.

For easy navigation and intuitive handling of the refinement process when the query refinement option is enabled, each attribute column is headed by a big green checkmark followed by a red cross. When the user wants to refine his search by accepting the whole concept formed by the attribute set contained in a column, he can simply click on the green checkmark. The next iteration of the search is then refined including all search attributes shown underneath the previously clicked checkmark. Correspondingly, clicking on the big red cross, the search is refined by excluding all listed attributes within that column. For a more subtle refinement, a smaller green plus and a red minus symbol are placed behind each attribute to initiate a single-attribute refinement.

**Limiting by an Object Count for Attributes.** A reduction of the corresponding context seems appropriate. The user can limit his view of the context

by increasing the object count for attributes. The default value for the object count corresponds to the full context. Increasing the object count decreases the number of attributes shown in the context and hence makes visualization easier when concentrating on the more prominent attributes only.

**Minimum Attribute Length.** In most cases it seems necessary to filter out tokens that are small in size, especially words with less than three characters. However, in some search scenarios even small words are relevant. The user decides whether he wants to use this option to remove all words with less than $n$ characters.

**Attribute Ranking.** FooCA retrieves Web object references in a pre-ranked order based on the search engine's own ranking criteria. A hybrid approach for ranking seems to be a pragmatic solution, since we can maintain the ranking for the objects and combine them with a new ranking of the related attributes. This approach seems natural as we read from top left to bottom right. The goal is to achieve a diagonal area of crosses in the cross table starting from the upper left to the lower right. To that end, attributes are ranked first by the number of objects they are related to and second by the sum of the ranking positions of the related objects. The latter corresponds to, and respects, the prior ranking of the search engine.

**Exporting the Context.** In addition to FooCA, the corresponding lattice visualization of a context, the explored formal context can easily be exported into your favorite visualization program for further operations. FooCA offers an export interface using the Burmeister Format (CXT). We have tested importing contexts generated by FooCA in ToscanaJ/Siena[6] [2].

### 4.6   Visualization of the Concept Hierarchy Using FlashLattice

Finally, after successfully refining the context $\mathbb{K}$, its concept hierarchy $\underline{\mathfrak{B}}(\mathbb{K})$ can be visualized by a line diagram using FlashLattice as used in SurfMachine[10] and D-SIFT[11], which has been contributed as a plugin for FooCA by Jon Ducrou from the University of Wollongong, Australia. FlashLattice provides the ability to visualize the prior refined context in a corresponding concept hierarchy, where each circle represents a concept, and lines between the concepts represent their subconcept-superconcept relation as shown in [8]. When clicking a concept button, the URLs attached to it will be opened in separate browser windows.

## 5   Example of a FooCA Web Search Analysis

### 5.1   Stepwise Context Size Reduction Via a Cross Table

Figure 2 shows the results for the query 'ICDM 2006' in FooCA. The context has already been refined by allowing only a minimum of four objects per attribute
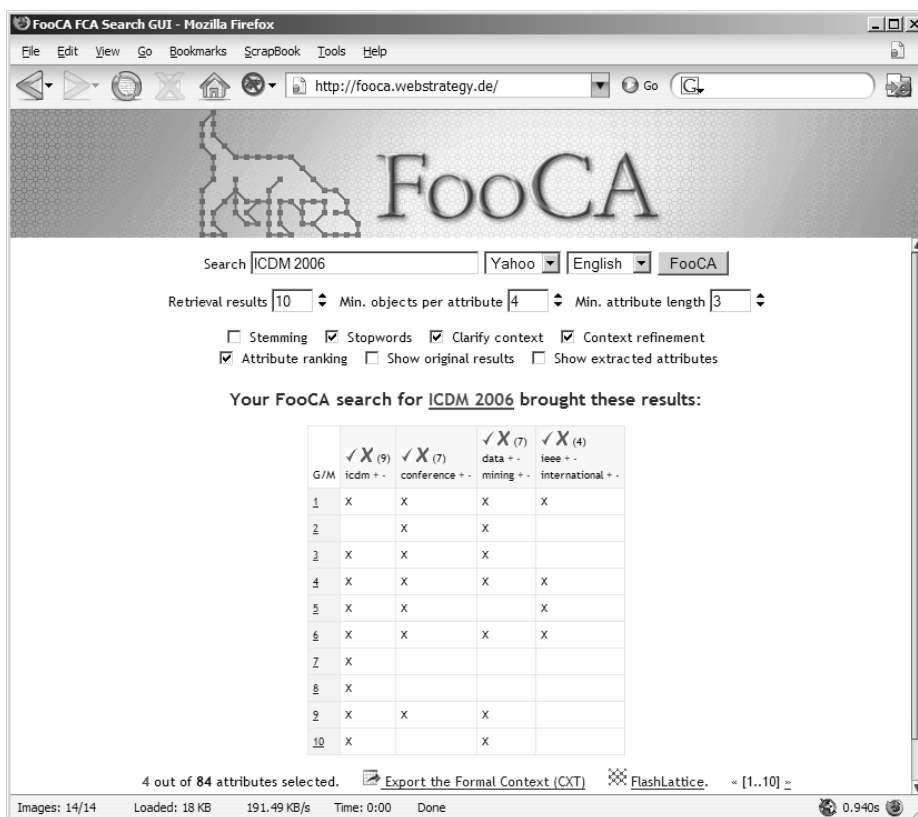
---

[6] http://toscanaj.sourceforge.net/

**Fig. 2.** FooCA result for the query 'ICDM 2006' on ten Yahoo results after a human-oriented context refinement process

and at least three characters of size for the attributes. Further optional preferences such as enabling stopword filtering, clarification of the context, query refinement and attribute ranking have been enabled in this example.

## 5.2   Visualization of the Concept Hierarchy Using a Line Diagram

In order to support data analysis, and hence, enable judgment over Web search engine data retrieval, data is presented to the user in the expanded form of a 'landscape structure'. Fortunately, FCA provides us with methods to support such a structured visualization. Every (finite) context can be represented by a concept hierarchy which is best visualized by a line diagram. A concept hierarchy represents the subconcept-superconcept relation. Concepts are represented by circles, whereas the relationships between concepts are symbolized by lines. Object labels are attached below the circles, whereas attribute labels are attached above attributes. These labels allow the diagram to be read and interpreted in two ways. As stated earlier in the FCA introductory section, the
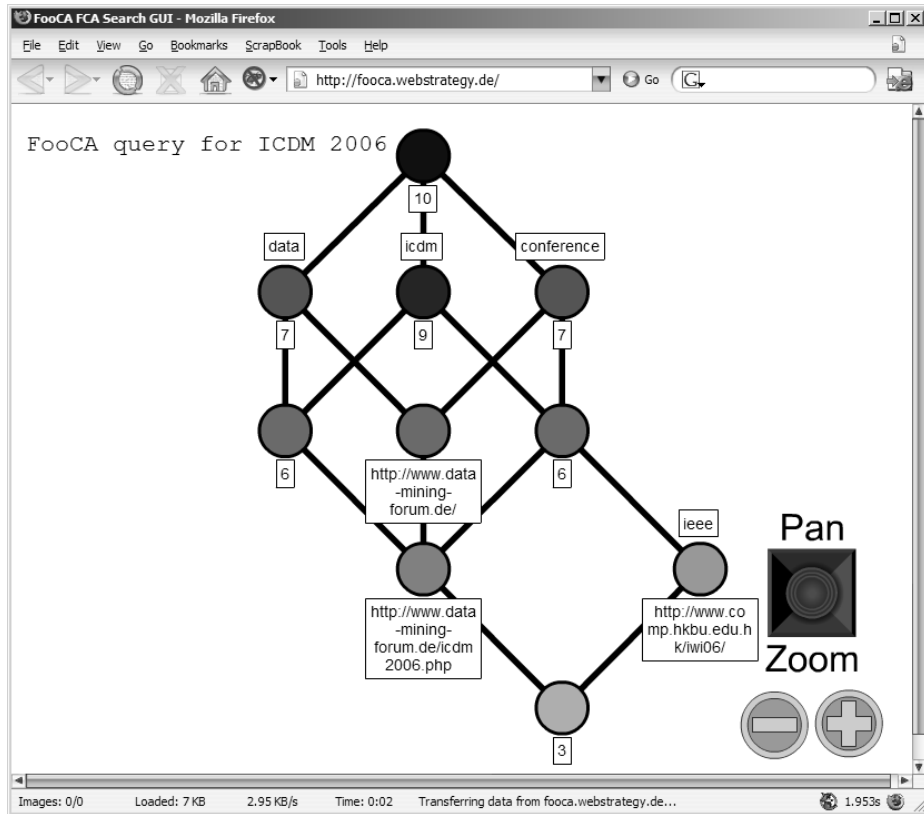
**Fig. 3.** FooCA showing the concept hierarchy for the query 'ICDM 2006' of ten Yahoo results. All Web objects can be opened by clicking on a concept. New or previously unknown relationships within the data set such as that there are two conferences on different continents called 'ICDM 2006' have been made visible.

extension covers all objects belonging to a concept and the intension comprises all attributes valid for all those objects. The extension of a distinct concept can then be read in a line diagram by starting at the very concept, following the objects downward along the path. An equivalent procedure can be done on the intension of a certain concept by following the attributes upward along the path.

   The line diagram in Figure 3 can be read as follows: attributes, which are labeled above the circles, are 'data', 'icdm', 'conference' and 'ieee'. Labeled below the circles are the objects. At first, the objects are hidden behind a white label only indicating the count of objects that are clustered within or below the concept. By clicking at the label, the label window opens showing all objects associated with that specific object. In this example two object labels have been opened to reveal different semantics for the search query 'ICDM 2006': Considering the attributes 'data' and 'conference', the top concept for the object URL `http://www.data-mining-forum.de` is visualized right in the center of the hi-

erarchy. The web page behind that URL is the host of the ICDM 2006 conference in Leipzig, Germany. Right below this concept, the official ICDM 2006 homepage can be found. The difference between them is the attribute 'icdm' which does not belong to the host web page, but instead to the official conference web page at `http://www.data-mining-forum.de/icdm2006.php`. All attributes except for 'ieee' are in its intension. If you look on the right hand side, you will find the two labels 'ieee' and the URL `http://www.comp.hkbu.edu.hk/iwi06` attached to a concept which is just another conference called ICDM hosted in China. By clicking on one of the concepts, a new browser window opens directing right to the URLs represented by the related Web objects.

## 6    Conclusion and Discussion

In this paper, we have presented an approach combining Formal Concept Analysis with Web Mining techniques to build a tool that applies Conceptual Knowledge Processing in information retrieval.

In FooCA as well as in any other information retrieval system, the user enters his own concepts in the form of a query of terms into the system. New with FooCA are the pre- and post-processing steps. A query is submitted along with user-selected options and the results are post-processed in order to obtain conceptual information.

After submitting the query, the user receives feedback in the form of a retrieval result presented in a context as shown in Figure 1 and represented by a line diagram of the concept hierarchy as shown in Figure 2. By his evaluation of, and judgment on, a larger conceptual presentation of terms and their relations to certain objects, the user learns to refine his query, developing similar or closely related queries and exploring their results interactively to define entirely new queries. By personally influencing the refining process using the interactive cross table, the user therefore not only finds new queries but also develops and adapts new concepts of thinking. The graphical representation of a line diagram uncovers conceptual structures that are closely related to cognitive structures of the human mind [21]. FooCA has indeed shown that it not only allows new concepts to be explored by searching for related queries using the query refinement mechanism, but that it also supports a holistic interpretation of, and judgment over, a larger amount of data by visualizing conceptual structures.

Consequently, using FooCA involves a learning process whereby the user understands how to close the systematically-evolving, semantic-pragmatic gap as described by Andelfinger [1].

According to Weber [22], the Kantian capacity to judge [16], whereby symbols involve a double judgment, is extended by a new technological dimension that unfolds in three steps:

> First, the capacity to judge applies a concept of a model to an object of sensible intuition. Then, it applies the mere rule of reflection to a different object of which the first is only a symbol. Finally, it establishes a relationship allowing the model to affect the object.

As a result, FooCA is a highly supportive tool assisting a user in assessing search retrievals. It gives the user more insight and an increased range of active possibilities than he would have with a traditional search engine interface.

Consequently, there is a definite need for establishing the idea of Conceptual Knowledge Processing in a next generation of information retrieval systems, a need for solutions that offer assistance. Further research in this area would appear appropriate.

### 6.1   Related Work

Approaches to enhancing information retrieval results have been studied for some time. Hearst [15] gives a general overview of user interfaces for information retrieval systems. Marchionini et. al. [19] have reported on ongoing efforts to develop and test generalizable user interfaces that provide interactive overviews for large-scale Web sites and portals.

Carpineto and Romano [6,7] introduced an FCA approach in 2004 with CREDO.[7] However, CREDO presents results in the look of a cluster-based search engine such as Clusty.[8] Although hierarchically focussing on specialized concepts, the user is confronted with a standard sequential list view consisting of title, link and a text snippet.

### 6.2   Future Work

FooCA has shown a successful way of giving structure to mostly unstructured or poorly structured data. The quality of the concept hierarchy and its corresponding line diagram can of course be heavily increased by obtaining more background knowledge. A possible approach to obtain background knowledge is inherent in the Web itself. Due to its inherent knowledge, a huge variety of freely-available knowledge bases can either be used or newly created by means of (semantic) Web Mining [3]. In Tim Berners-Lee's vision of the Semantic Web [4], all information on the World Wide Web can be connected. As for today, the Semantic Web is still a vision. Furthermore, doubts have arisen regarding the practicability of the Semantic Web approach [9].

Nevertheless, once we have a knowledge base that is suitable for a specific search domain, interfaces can be implemented. Knowledge bases could consist of domain-specific vocabularies or thesauri that limit the attributes in our example to the desired topic or domain, such as the online thesaurus that contains information about synonyms and hypernyms, WordNet.[9]

We hope we have been able to show new ways of combining fast search results, giving the user full control over fundamental searching strategies, and unlocking data through new forms of representation such as contexts and concept hierarchies with FooCA.

In a nutshell, FooCA is an innovative, practical implementation of Conceptual Knowledge Processing.

---

[7] `http://credo.fub.it`
[8] `http://www.clusty.com`
[9] `http://wordnet.princeton.edu`

## Acknowledgments

## References

1. Andelfinger, U.: Diskursive Anforderungsanalyse. Ein Beitrag zum Reduktionsproblem bei Systementwicklungen in der Informatik. Peter Lang, Frankfurt 1997 (in German).
2. Becker, P., Hereth Correia, J.: The ToscanaJ Suite for Implementing Conceptual Information Systems. In Ganter, B., Stumme, G. and Wille, R. (Eds.): Formal Concept Analysis, Foundations and Applications. Springer-Verlag, Berlin, 2005.
3. Berendt, B., Hotho, A., Stumme, G.: Towards Semantic Web Mining. In International Semantic Web Conference, ISWC02, Springer-Verlag, Berlin, 2002.
4. Berners-Lee, T.: Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor. Texere, London, 1999.
5. Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. In Computer Networks and ISDN Systems, Elsevier, Amsterdam, 1998.
6. Carpineto, C., Romano, G.: Concept Data Analysis: Theory and Applications. John Wiley & Sons, Chichester, 2004.
7. Carpineto, C., Romano, G.: Exploiting the Potential of Concept Lattices for Information Retrieval with CREDO. In Journal of Universal Computer Science, Springer-Verlag, Berlin, 2004.
8. Cole, R., Ducrou, J., and Eklund, P.: Automated Layout of Small Lattices Using Layer Diagrams. In Missaoui, R., Schmid, J. (Eds.): Formal Concept Analysis, 4th International Conference, ICFCA 2006, Dresden, Germany, Proceedings. Springer-Verlag, Berlin, 2006.
9. de Moor, A.: Patterns for the Pragmatic Web. In Dau, F., Mugnier, M., Stumme, G. (Eds.): Conceptual Structures: Common Semantics for Sharing Knowledge, ICCS05, Springer-Verlag, Berlin, 2005.
10. Ducrou, J., Eklund, P.: Combining Spatial and Lattice-based Information Landscapes. In Ganter, B., Godin, R. (Eds.): Proceedings of the 3rd Int. Conference on Formal Concept Analysis, ICFCA 2005, Springer-Verlag, Berlin, 2005.
11. Ducrou, J., Wormuth, B., Eklund, P.: D-SIFT: A Dynamic Simple Intuitive FCA Tool. In Dau, F., Mugnier, M., Stumme, G. (Eds.): Conceptual Structures: Common Semantics for Sharing Knowledge: Proceedings of the 13th International Conference on Conceptual Structures, Springer-Verlag, Berlin 2005.
12. Eklund, P., Ducrou, J., Brawn, P.: Concept Lattices for Information Visualization: Can novices read line diagrams. In Eklund, P. (Ed.): Proceedings of the 2nd International Conference on Formal Concept Analysis, ICFCA 2004, Springer-Verlag, Berlin, 2004.

---

[10] AG Algebra und Logik, Darmstadt University of Technology, Germany
[11] Institut für Algebra, Technische Universität Dresden, Germany
[12] School of Economics and Information Systems, University of Wollongong, Australia

13. Fürnkranz, J.: Web Mining. In Maimon, O. and Rokach, L. (Eds.), Data Mining and Knowledge Discovery Handbook, Springer-Verlag, Berlin, 2005.
14. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer-Verlag, Berlin, 1999.
15. Hearst, M.: User interfaces and visualization. In Modern Information Retrieval, edited by R. Baeza-Yates and B. Ribeiro-Neto. ACM Press, New York, 1999.
16. Kant, I.: Kritik der Urteilskraft. Ed.: Wilhelm Weischedel. Suhrkamp, Frankfurt am Main, 2004 (in German).
17. Koester, B.: Conceptual Knowledge Processing with Google. In Bauer, M., Fürnkranz, J., Kröner, A. et. al. (Eds.): Lernen, Wissensentdeckung und Adaptivität, LWA 2005, GI Workshops, Saarbrücken, October 10th-12th, 2005.
18. Koester, B.: FooCA: Enhancing Google Information Research by Means of Formal Concept Analysis. In Ganter, B., Kwuida, L. (Eds.): Contributions to ICFCA 2006, Verlag Allgemeine Wissenschaft, Mühltal, 2006.
19. Marchionini, G. and Brunk, B.: Towards a General Relation Browser: A GUI for Information Architects. Journal of Digital Information, Volume 4 Issue 1, Texas A&M University Libraries, 2003.
20. Porter, M.F.: An Algorithm for Suffix Stripping. In Readings in information retrieval. Morgan Kaufmann Multimedia Information And Systems Series, San Francisco, 1997.
21. Seiler, Th. B.: Begreifen und Verstehen. Ein Buch über Begriffe und Bedeutungen. Verlag Allgemeine Wissenschaft, Mühltal, 2001 (in German).
22. Weber, D.: Kybernetische Interventionen. Zum Kritischen Verständnis des immanenten Verhältnisses von Multimedia und Pädagogik. Verlag für Sozialwissenschaften, Wiesbaden, 2005 (in German).
23. Wille, R.: Conceptual Knowledge Processing in the Field of Economics. In Ganter, B., Stumme, G., Wille, R.: Formal Concept Analysis. Foundations and Applications. Springer-Verlag, Berlin, 2005.
24. Wille, R.: Methods of Conceptual Knowledge Processing. In In Missaoui, R., Schmid, J. (Eds.): Formal Concept Analysis, 4th International Conference, ICFCA 2006, Dresden, Germany, Proceedings. Springer-Verlag, Berlin, 2006.