Conceptual Knowledge Processing with Google

Bjoern Koester
Webstrategy GmbH
D-64293, Darmstadt, Germany
http://www.webstrategy.de
bjoern.koester@webstrategy.de

Abstract

This paper introduces a tool for Conceptual Knowledge Processing with Google. The featured prototype, called FooCA, tries to combine the advantages of two research disciplines, Web Mining and Formal Concept Analysis (FCA). Web Mining techniques are used to preprocess search results retrieved via Google, presenting the formal context in an interactive cross table. A new formal context can be iteratively explored by user-configured query refinement. Depending on his personal information need, the user can make decisions about the way he wants to navigate through Google Web search results. Moreover, the user may enable or disable Web Mining or FCA options to help him obtain an overview of the context and to give him more influence than he would have when working with a traditional search engine interface.

1 Introduction

This paper presents a Web-based Internet application for Formal Concept Analysis (FCA) [Ganter and Wille, 1999]. The principal idea behind the application is to enhance information access via a standard search engine. The featured prototype, called FooCA, tries to combine the advantages of two research disciplines, Web Mining and FCA. FooCA stands for a word combination of FCA and Google, since FCA mechanisms are applied to a standard Google search.

Access to Information through Web data plays an important role today [Fürnkranz, 2005]. While facing a rapidly growing information flood 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. Even 'simple searches' usually end up with over one million results.

This paper discusses a different approach to accessing information on the Web using Conceptual Knowledge Processing. We present FooCA as a hybrid system consisting of an exclusively automatic retrieval process on one side, and a highly interactive user-oriented interface based on the idea and methods of FCA on the other side.

2 Conceptual Knowledge Processing

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 [Wille, 2005]. 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.

Concepts as basic units of thought containing both experiences and knowledge of the world build the basis of this kind of understanding of human thinking. Since human beings use concepts to act and communicate with other people, a variety of structures across many disciplines exist about conceptual thinking [Wille, 2000]. A mathematical approach has evolved out of this understanding and will be briefly introduced in the following section.

3 Formal Concept Analysis

Formal Concept Analysis deals with gaining of concepts and its hierarchical implications out of data. FCA studies how objects can be hierarchically grouped together according to their common attributes. FCA is based on a mathematization of the philosophical understanding of concept. For the formalization of concept, an algebraic theory of binary relations and complete lattices is used.

However, in difference to pure mathematical logic, the logic we are interested in is rather contextual, which means that we primarily look at the logical structure of concrete data. The latter represents our understanding of context.

3.1 Context and Formal Context

Informally, when we talk about a context, it has different meanings. When we use the notion "context", we mean the formal context, which we define in a mathematical sense as follows:

Definition 1 (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* $I \subseteq G \times M$ *assigning the appropriate attributes to each object. We call* I *the "incidence relation" and read* $(g, m) \in I$ *as "the object g has the attribute m".*

A context $\mathbb{K}$ can be visualized by a two dimensional table consisting of crosses as elements which we call a cross table as shown in Table 1.

$\mathbb{K}$	**Attr 1**	**Attr 2**	**Attr 3**
Obj 1		×	
Obj 2	×		×
Obj 3	×	×	

Table 1: A cross table for a given context

3.2 Concept and Formal Concept

As we have dealt with the context above, we also have to distinguish between two notions of "concept". The informal understanding of concept and the formal understanding.

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 [Wolff, 1994].

Wille brought both philosophy and mathematics together by defining a formal context. Since objects, attributes and a relation form a context, we can say that a context forms a concept if and only if a context exists and an extension as well as an intension are given.

Before we can define a formal concept for a distinct formal context, we need to introduce the derivation operators.

Definition 2 (Derivation Operators). *For a subset $A \subseteq G$ of the objects we define the set of attributes common to the objects in A as*

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

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

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

The pair of the derivation operators form a *Galois connection*. Thus the following equations 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 allow us to define a formal concept as follows.

Definition 3 (Formal Concept). *A formal concept of the context $\mathbb{K} := (G, M, I)$ is composed of a pair (A, B) consisting of an extension $A \subseteq G$ and an intension $B \subseteq M$ for which apply*

$$A' = B \text{ and } B' = A.$$

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

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 can be visualized by line diagrams. However, a further introduction into these structures would lead far beyond the scope of this paper. Readers interested in a deeper insight into FCA are referred to [Ganter and Wille, 1999].

4 Google

One way of retrieving Web documents 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 of running the hardware and storage systems, but it would also cause 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 that as it has been done by so many institutions and companies worldwide.

Our approach is to launch a search request using the Google API[1] and to analyze the returned set of ranked items by means of Web Mining and FCA.

4.1 Google Ranking

We first obtain ranked items that match the criteria of the Google algorithms. Since the search engine market has become competitive, the exact algorithms used by Google for the ranking part of Web results are classified. However, the heart of Google is the PageRank algorithm[2], which has been documented and published [Brin and Page, 1998]. The PageRank algorithm continues to be Google's core technology, with minor adjustments being made "on a daily basis", as Google states on their website.

PageRank

The PageRank PR builds a probability distribution across Web documents with $\sum_{i=1}^{n} PR(A_i) = 1$, where A_i is a Web document. Then, the PageRank of a distinct document A can be formalized as

$$PR(A) = (1 - d) + d \cdot \sum_{i=1}^{n} \frac{PR(T_i)}{C(T_i)} \tag{7}$$

where d with $0 \leq d \leq 1$ is a dampening factor, T_i are all Web documents linking to document A. $C(T_i)$ is the number of overall links in a document.

According to the PageRank introduced, any Web document gains importance if either many other documents or a high ranked document link to it. A link from another document is of high value if the document itself has a high PageRank or there are only a few outgoing links to other documents.

4.2 Query Evaluation

For later 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 set of terms forming a request for information from a database. In Google, the entered query is parsed and segmented into terms. Adding additional terms to the query means that each term should appear in a document.

Since a query is embedded into a complex query language, we would like to simplify our understanding of a query as a set of terms with two operations, the concatenation and exclusion of terms:

Implicit AND Operator

The query 'formal concept analysis' results in a search for pages containing the words 'formal' and 'concept' and 'analysis'. 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.

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 'formal concept analysis' will return a lot of introductions. Assuming that we are

[1] `http://www.google.com/apis/`

[2] `http://www.google.com/intl/en/technology/`

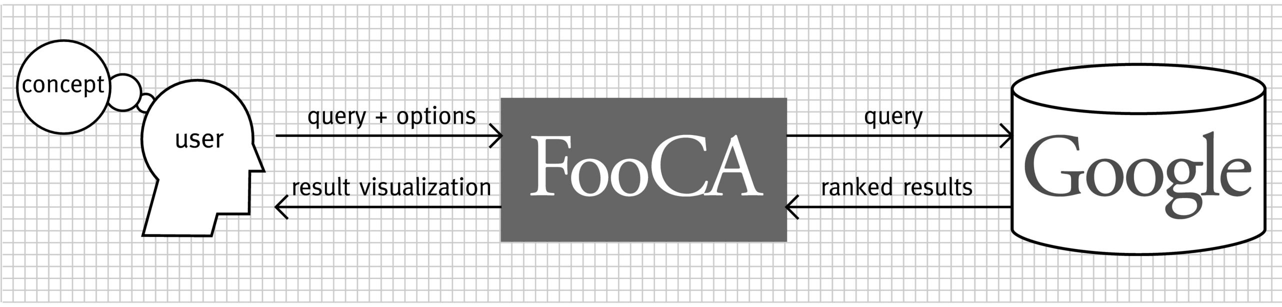


Figure 1: FooCA interacts between the user and the search engine.

experts in this field, we are not really interested in introductions to FCA. Therefore, we can use the negation operator to exclude a term from the search by simply adding '-introduction' to our search query resulting in the following query that better suits our information need: 'formal concept analysis -introduction'. All pages returned contain 'formal' and 'concept' and 'analysis' and no instance of the term 'introduction'.

Google supports a number of more advanced operators[3], query operators that have a special meaning and enable different types of searches. However, for the moment the two query term operations presented shall suffice.

4.3 Quality of Results

First of all, the ranking mechanism relies primarily on the PageRank. Its main focus is link structure, rather than semantics of a Web document.

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

5 FooCA

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

5.1 Architecture

The current prototype runs on a Linux system and is written in Perl. It communicates with Google using the official Google API[4] via SOAP[5].

In a query result, we are interested in the general items returned by Google 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 search for. He then enters an adequate 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 Google. Google receives the query as if it were a normal Google search and processes it, returning the results in a ranked order to FooCA. Using the personal options and the search results retrieved from Google, FooCA now generates its internal representation of the formal context and presents it to the user in a visualized form. From this point on, the user can refine his search, practically submitting a new query.

5.2 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 Google retrieval results:

- Identification of tokens
- Stripping of all HTML format tags
- Rewriting of German umlauts (e.g. 'ö' to 'oe')
- Transforming all characters from upper- to lower-case
- Removing all special characters except for '-'
- Removing all words ≤ 3 characters in length

Using these basic operations, a list of useful word tokens originating from the Google snippet can easily be generated.

5.3 From a Google Retrieval to a Formal Context

FooCA lets the user enter a query Q_i which is directly passed on to Google without modification.

To limit the number of words, we extract the words surrounding the search terms of each result. Google offers the user a short excerpt of words before and after an occuring term that appears in the query called 'snippet'. The idea is to use that snippet as a basis 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 formal context $\mathbb{K}$ considering the URLs as objects G and the extracted feature terms as attributes M such that $\mathbb{K}(Q_i) := (G, M, I)$.

5.4 Interactive Cross Table

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

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 document search results. Columns are headed by attributes

[3] http://www.google.com/help/operators.html

[4] http://www.google.com/apis/reference.html

[5] SOAP stands for Simple Object Access Protocol and is a standard for exchanging XML-based messages over the internet using HTTP.

which are the extracted feature terms of the Google snippets. The incidence relation I of the formal context $\mathbb{K}$ between the Web document objects G and its attributes M is marked by a cross 'x' in the table. The ith Web document possesses the jth 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 each active row is highlighted indicating the Web object. The user can click anywhere inside the table and is promptly directed to the related Web document.

Apart from navigating inside the table by way of the incidence relation of the formal 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 the user can include oder exclude an attribute by defining a specialized version of the previous query.

5.5 Search Preferences and Strategies

FooCA gives the advanced searcher the possibility to gain more control over the making of decisions by using methods and operations that are usually processed automatically in general search engines without human intervention. Those methods and operations provided by the current FooCA prototype are described in the following subsections:

Removing stop words

Common words in a language that occur frequently but are insignificant can simply be removed. In 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

Since English is a fairly simple language, stemming algorithms perfom quite well. One of the most popular and fastest algorithms introduced by Porter is called the Porter Stemmer [Porter, 1997] which is used in our application.

Clarification of the Formal Context

The formal context $\mathbb{K} := (G, M, I)$ can be attribute-clarified to a context $(G, M/\sim, \tilde{I})$, where $\sim$ is the equivalance relation with $m \sim n :\Longleftrightarrow m' \sim n'$.

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 a set of attributes.

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 retrievel process. Just as in the process of adding new query terms into the Google search 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 Google interface consists in the set of given attributes. Typically, the user is not entirely clear right from the beginning just which term(s) he needs to narrow down 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 context knowledge. Furthermore, new relationships or erroneous relationships become apparent instantly.

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 big red cross. When the user would like to refine his search by accepting the whole concept formed by the attribute set contained in 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 small green plus and a small red minus symbol are placed behind each attribute to initiate a single-attribute refinement.

Exporting the Formal Context

As FooCA does not yet offer any other form of visualization of the corresponding lattice to a formal context, the explored formal context can easily be imported by your favorite visualization program. FooCA offers an export interface using the Burmeister Format (CXT). We have successfully tested importing FooCA-generated contexts into ToscanaJ/Siena[6] and ConExp[7].

Limiting by an Object Count for Attributes

As experiments have shown, the FooCA-generated formal context grows linearly with the number of document results. A limitation of the focus on the corresponding context seems appropriate. The user can limit his view of the formal context by increasing the object count for attributes. The default value for the object corresponds to a 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.

Attribute Ranking

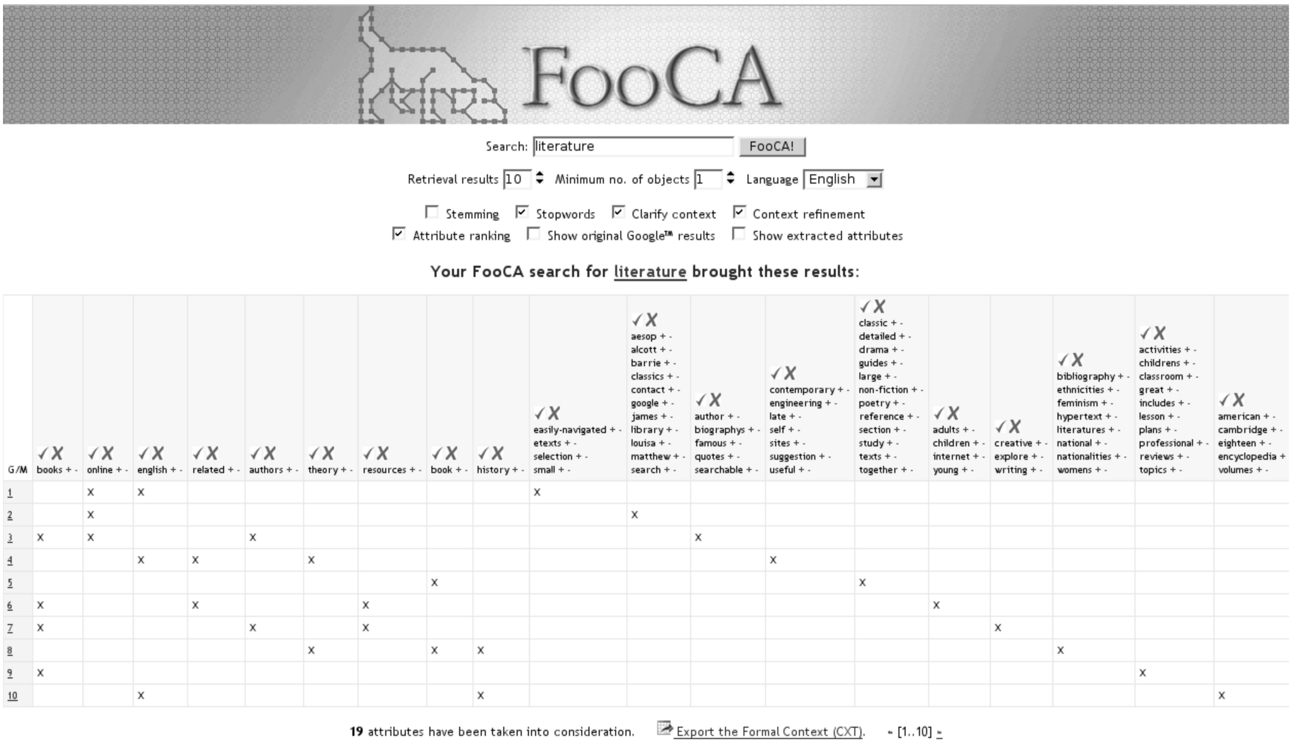
FooCA retrieves Web document references in a pre-ranked order based on Google's criteria. A hybrid approach for ranking seems to be a pragmatic solution, since we can maintain the ranking for the objects and combine them with a new ranking of the related attributes. This approach seems natural as we 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 Google positions of the related objects. The latter corresponds to and respects the prior Google ranking.

Language Restriction

Google offers the possibility to restrict the search in a predefined subset of its Web document index. To allow both a country- and a language-based limitation of a search, Google offers two optional restriction parameters. The first

[6] `http://freshmeat.net/projects/toscanaj`

[7] `http://sourceforge.net/projects/conexp`



G/M	books + -	online + -	english + -	related + -	authors + -	theory + -	resources + -	book + -	history + -	easily-navigated + - etexts + - selection + - small + -	aesop + - alcott + - barrie + - classics + - contact + - google + - james + - library + - louisa + - matthew + - search + -	author + - biographys + - famous + - quotes + - searchable + -	contemporary + - engineering + - late + - self + - sites + - suggestion + - useful + -	classic + - detailed + - drama + - guides + - large + - non-fiction + - poetry + - reference + - section + - study + - texts + - together + -	adults + - children + - internet + - young + -	creative + - explore + - writing + -	bibliography + - ethnicities + - feminism + - hypertext + - literatures + - national + - nationalities + - womens + -	activities + - childrens + - classroom + - great + - includes + - lesson + - plans + - professional + - reviews + - topics + -	american + - cambridge + - eighteen + - encyclopedia + volumes + -
1		x	x							x									
2		x									x								
3	x	x			x							x							
4			x	x		x							x						
5								x						x					
6	x			x			x								x				
7	x				x		x									x			
8						x		x	x								x		
9	x																	x	
10			x						x										x

Figure 2: The context in FooCA for the query 'literature' with attribute ranking enabled. Ranking of the attributes obviously improves readability and gives a faster overview over the more relevant (not necessarily more important) attributes.

parameter allows a limitation to one of 28 specific languages. The second parameter allows to distinguish between more than 200 countries. FooCA supports the search in English and German with other languages to be easily integrated as well.

6 Example

Figure 2 shows a context for the search query 'literature' limited by ten results. Stopword filtering, clarification of the context, attribute ranking and query refinement as optional preferences have been enabled. This example shows that ranking the attributes can indeed improve readability, with the attributes related to most frequently appearing in an ordered fashion from left to right. It therefore gives the user a faster overview of the most relevant attributes. It should be made clear, however, that these attributes are not necessarily more important than any other attributes in any given context. In some cases the most important attribute has only a small object count, although being distinctive and characterizing.

Looking at the distribution of the crosses in the table, it is very nice to see that all combined attributes in this example relate to only one object, forming a descending line from top left to the bottom right. In most explorative searches it proved useful to ignore those attribute columns. This requirement led to the creation of an option of a minimum object count per attribute.

If the user is interested in literature, but only in English literature, he can refine his search by clicking at the green checkmark above or at the plus sign behind the attribute 'english' in the cross table. As a feedback, the user is presented with a narrowed down formal context. If we suppose the user is interested in British but not in American literature, he might refine the search by excluding the attribute 'american'. After his two-step refinement process, the user is presented with the context featuring the following attributes (limited by a minimum of four objects per attribute): 'authors', 'resources', 'language', 'university', 'links', 'resource', 'results', 'century', 'subjects'. As we see, an additional stemming process could save us one column by combining 'resources' with 'resource'. However, there might be a semantic distinction between plural and singular. A 'resource' could stand for a single resource whereas 'resources' might lead to multiple resources. The latter could also be a synonym for 'links'. Accordingly, interpretation of the given context is completely up to the user. The context is what the user personally sees in it.

7 Related and Future Work

Approaches to enhancing information retrieval results have long been studied. Hearst [Hearst, 1999] gives a general overview of user interfaces for information retrieval systems. Marchionini et. al. [Marchionini and Brunk, 2003] have reported on ongoing efforts to develop and test generalizable user interfaces that provide interactive overviews for large-scale Web sites and portals.

A very similar and current FCA-based approach was introduced in 2004 by Carpineto and Romano [Carpineto and Romano, 2004a; 2004b] with CREDO[8]. However, the fundamental difference between the CREDO approach and FooCA is the way CREDO simulates the look of cluster-based search engines such as Clusty[9]. The complete formal context is hidden to the user. One other aspect concerns the influence of the user himself. With CREDO the user cannot control the underlying strategies.

A next step will be to extend FooCA's visual presentation from the cross table to an optional embedded hierarchical or graphically visualized concept lattice view without the need to export the formal context from the FooCA system. Further simplification of the formal context for a larger amount of Web objects will then be necessary.

[8]http://credo.fub.it

[9]http://www.clusty.com

A possible approach to reducing the number of attributes 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 [Berendt *et al.*, 2002]. In Tim Berners-Lee's vision of the Semantic Web [Berners-Lee, 1999], all information on the World Wide Web can be connected. As for today, the Semantic Web is still a vision and furthermore, doubts arise regarding the practicability of the Semantic Web approach [de Moor, 2005].

8 Conclusion and Discussion

In this paper, we have presented an approach and its prototype combining Formal Concept Analysis with Web Mining techniques to serve as a tool for Conceptual Knowledge Processing in the field of information retrieval.

In FooCA as well as in any other information retrieval system, the user enters into the system his own thought concepts in the form of a query of terms. After submitting the query, he receives a feedback in the form of a retrieval result presented in a formal context as shown in Figure 1. By his evaluation of, and judgment on, a larger contextual 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. In our experiments, FooCA has indeed shown that new concepts can be explored by searching for related queries and then using the query refinement mechanism.

Consequently, using FooCA involves a learning process that helps the user to learn how to close the systematically evolving semantic-pragmatic gap as described by Andelfinger [Andelfinger, 1997].

According to Weber [Weber, 2005], the Kantian capacity to judge [Kant, 2004], 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 influence than he would have when working with a traditional search engine interface.

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

Acknowledgments

This work has been inspired by a personal discussion with Prof. Rudolf Wille at the ICCS05 in Kassel, Germany. We would like to thank Rudolf Wille and Joachim Hereth Correia for their valuable comments and support.

References

[Andelfinger, 1997] Urs Andelfinger. *Diskursive Anforderungsanalyse. Ein Beitrag zum Reduktionsproblem bei Systementwicklungen in der Informatik*. Peter Lang, Frankfurt, 1997.

[Berendt *et al.*, 2002] B. Berendt, A. Hotho, and G. Stumme. Towards semantic web mining, 2002.

[Berners-Lee, 1999] Tim Berners-Lee. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. Harper San Francisco, 1999.

[Brin and Page, 1998] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.

[Carpineto and Romano, 2004a] Claudio Carpineto and Giovanni Romano. *Concept Data Analysis: Theory and Applications*. John Wiley & Sons, 2004.

[Carpineto and Romano, 2004b] Claudio Carpineto and Giovanni Romano. Exploiting the Potential of Concept Lattices for Information Retrieval with CREDO. 10(8):985–1013, August 2004.

[de Moor, 2005] Aldo de Moor. Patterns for the pragmatic web. In Frithjof Dau, Marie-Laure Mugnier, and Gerd Stumme, editors, *ICCS*, volume 3596 of *Lecture Notes in Computer Science*, pages 1–18. Springer, 2005.

[Fürnkranz, 2005] Johannes Fürnkranz. Web mining. In Oded Maimon and Lior Rokach, editors, *The Data Mining and Knowledge Discovery Handbook*, pages 899–920. Springer, 2005.

[Ganter and Wille, 1999] Bernhard Ganter and Rudolf Wille. *Formal concept analysis: mathematic foundations*. Springer, Berlin/Heidelberg, 1999.

[Ganter *et al.*, 2005] Bernhard Ganter, Gerd Stumme, and Rudolf Wille, editors. *Formal Concept Analysis, Foundations and Applications*, volume 3626 of *Lecture Notes in Computer Science*. Springer, 2005.

[Hearst, 1999] M. Hearst. User interfaces and visualization. 1999.

[Kant, 2004] Immanuel Kant. *Kritik der Urteilskraft*. Suhrkamp, 17. aufl. edition, 2004.

[Marchionini and Brunk, 2003] Gary Marchionini and Ben Brunk. Towards a general relation browser: A gui for information architects. *J. Digit. Inf.*, 4(1), 2003.

[Porter, 1997] M. F. Porter. An algorithm for suffix stripping. pages 313–316, 1997.

[Weber, 2005] Dietmar Weber. *Kybernetische Interventionen. Zum Kritischen Verständnis des immanenten Verhältnisses von Multimedia und Pädagogik*. Verlag für Sozialwissenschaften, 2005.

[Wille, 2000] Rudolf Wille. Begriffliche wissensverarbeitung: Theorie und praxis. *Informatik Spektrum*, 23(6):357–369, 2000.

[Wille, 2005] Rudolf Wille. Conceptual knowledge processing in the field of economics. In Ganter et al. [2005], pages 226–249.

[Wolff, 1994] Karl-Erich Wolff. A first course in formal concept analysis – how to understand line diagrams. In F. Faulbaum, editor, *Advances ha Statistical Software, Vol. 4*, pages 429–438. Gustav Fischer Verlag, Stuttgart, 1994.