

A Topic-Based Browser for Large Online Resources

Heiner Stuckenschmidt¹, Anita de Waard², Ravinder Bhogal³, Christiaan Fluit⁴, Arjohn Kampman⁴, Jan van Buel⁵, Erik van Mulligen^{6,7}, Jeen Broekstra⁴, Ian Crowlesmith⁵, Frank van Harmelen¹, Tony Scerri²

¹Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081HV Amsterdam,

²Advanced Technology Group, Elsevier, Amsterdam, NL / London, UK,

³User-Centered Design Group, Elsevier, London, UK,

⁴Aduna, Amersfoort, NL,

⁵Bibliographic Databases, Elsevier, Amsterdam, NL,

⁶Collexis B.V., Geldermalsen, NL,

⁷Department of Medical Informatics, Erasmus University, Rotterdam, NL

Abstract. The exploration of large information spaces is a difficult task, especially if the user is not familiar with the terminology used to describe information. Conceptual models of a domain in terms of thesauri or ontologies can leverage this problem to some extent. In order to be useful, there is a need for interactive tools for exploring large information sets based on conceptual knowledge. We present a thesaurus based browser that supports a mixed-initiative exploration of large online resources that provides support for thesaurus-based search and topic-based exploration of query results. We motivate the chosen exploration strategy the browser functionality, present the results of user studies and discuss future improvements of the browser.

1 Introduction

The exploration of large information spaces is a difficult task. Users need to be supported by appropriate tools to access and present available information. Looking at the World Wide Web as the largest information space that is currently available to a large group of users, we recognize two main strategies for accessing information:

- Search engines provide keyword based methods to retrieve sets of relevant documents based on string matching techniques.
- Hyperlinks can be used to navigate between web pages in a way predefined by the developer of the particular web site.

Users typically employ an exploration strategy that combines these two techniques. Search engines are used to jump to a particular page that seems to be relevant to a query. From there on the information space connected to the page is explored by browsing existing hyperlinks. Both of these paradigms have

serious drawbacks with respect to efficiency and precision of information access. Keyword based search engines suffer from the problem of ambiguity of terms used in the queries. Often the same keyword is used to describe different things leading to a low precision of the result or different words are used for the same thing lowering the recall. When using the browsing paradigm, exploration is limited to predefined paths that do not reflect the users view on relevance, but the one predefined by the information provider.

The problems mentioned above turn out to be the main bottleneck for efficient exploration of large information spaces, especially if the user is not familiar with the dominant terminology used. Current developments in the area of the semantic web aim at using explicit models of the terminology of a domain to improve information access (see e.g. [5]). Using conceptual structures has benefits with respect to searching as well as for browsing:

- The precision and recall of querying can be improved by normalizing the vocabulary used for indexing and for querying documents to a common set of terms.
- Navigation in the information space can be based on conceptual relations rather than existing links and the user can be given the choice of which relations to use.

We have developed a browser that uses conceptual knowledge in terms of a thesaurus to support the exploration of large information spaces. The browser is designed in such a way that it directly supports the mixed-initiative strategy employed by users and makes use of the conceptual knowledge to support searching as well as browsing. We applied the browser to large repositories of documents in the life science domain and performed user studies for assessing the benefits of our approach.

In the following, we briefly describe the DOPE¹ project in which the browser was developed and present the specific requirements for supporting the user in the exploration tasks we identified in the project. In section 2 we review the design rationales of the system we have developed in the DOPE project referring to results from empirical studies in the area of Information retrieval. Section 3 presents the browser we developed in order to meet the requirements. The results of a user study with real end users that confirm our claims about the adequacy of the approach are presented in section 4. We conclude with a summary of our results and some directions for future work.

2 The DOPE Project

Innovative science and technology institutes rely on the availability of complete and accurate information about new research and development. It is the business

¹ Drug Ontology Project for Elsevier

of information providers such as Elsevier to provide the required information in a cost-effective way. The semantic web will very likely provide an important contribution to future scientific information solutions, since it facilitates easy and effective access to a large variety of data. With the unremitting growth of scientific information sources, the need for integrated access to such sources becomes even more urgent.

Thesauri have proven to be a key technology to effective information access as they provide a controlled vocabulary for indexing information, and thereby help to overcome some of the problems of free-text search by relating and grouping relevant terms in a specific domain. A number of thesauri have been developed in different domains of expertise. Examples from the area of medical information include MeSH² and Elsevier's life science thesaurus EMTREE³

These thesauri are used to access information sources (in particular document repositories) like EMBASE.com⁴. But currently there is no open architecture to allow access to these thesauri for querying other data sources. When we move from the centralized and controlled use of EMTREE inside Elsevier to a distributed setting, we need to improve the accessibility of the content by delivering it in a standardized representation using open data standards that allow for semantic qualifications, such as RDF.

Also, since the mental models and common terms for accessing data diverge between subject areas and communities, different ontologies will be developed and used, that need to coexist. An ideal architecture would therefore allow for the disclosure of distributed and heterogeneous data sources through different ontologies. The aim of the DOPE project (Drug Ontology Project for Elsevier) is to investigate the possibility of providing access to multiple information sources in the area of life science through a single interface

2.1 The current situation

Besides asking an information specialist at a company or academic institute to do a full literature or product review there is a tendency that researchers perform their own searches for relevant information. The two ways of searching used by the overwhelming majority of life science researchers (including Drug Discovery) are specialized Bibliographic resources and internet search engines.

Bibliographic Resources PubMed⁵ is an online repository produced by the National Centre for Biotechnology Information. The data searched through

² <http://www.nlm.nih.gov/mesh/meshhome.html>

³ <http://www.elsevier.com/homepage/sah/spd/site/>

⁴ <http://embase.com/>

⁵ <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>

PubMed are the abstracts contained in MEDLINE, which contains over 12 million bibliographic citations and author abstracts from more than 4,600 biomedical journals. PubMed allows for fielded searches (e.g. search for Author name = Rat vs. Species = Rat)

Internet Search Engines Scientists also use Google or other full-text web search engines such as Scirus (www.scirus.com) to search the web for information about their specific topic. In this case, queries are usually one- or two keyword phrases that are so specific the records returned can be oversene by the searcher. A Google query of "MHC, peptide epitope" (one of the starting terms for one of our test users) retrieved 29,600 records.

Additionally, researchers search specific databases of either chemical compounds (one of the many drug databases available via the internet or the local intranet), genes (e.g. GenBank, <http://www.ncbi.nlm.nih.gov/Genbank/>, also published by NCBI) and proteins (e.g. SwissProt, <http://www.ebi.ac.uk/swissprot/>, produced by the European Bioinformatics Institute) as well as many proprietary, subject-specific or other data bases of compounds, nucleotides, proteins, genes, and other biochemical entities.

Integrating and sorting through the data of these user-driven searches is one of the main issues driving drug-discovery (there are entire conferences devoted to the subject of information integration in drug discovery,⁶. There is currently no one product that can index independent data sources to one or more thesauri and integrate the search over all of them.

2.2 Requirements for Tool Support

The quality of a system supporting the user in exploring large information spaces depends on many factors. Some of them are linked to human cognition, our way of thinking and our mental limitations. We will not discuss these general requirements but rather concentrate on the specific requirements that can be derived from the application in the context of the DOPE project. These requirements are mainly concerned with assisting the user in exploring the different document repositories linked to the browser based on terms from the thesaurus.

Some of the requirements directly arise from the nature of the application. First of all, the visualization has to support the use of conceptual knowledge, but not necessarily its creation or maintenance. A requirement the application area shares with the Web in general is the ability to deal with a high number of instances compared to a smaller conceptual model. At the current state of the DOPE project, the accessible information space contains about 10.5 million

⁶ <http://www.wbresearch.com/DDDII/agenda.html>

documents that are indexed using a thesaurus that contains about 50000 terms. It is quite obvious that we have to limit the the number of documents and terms presented to the user at the same time. Another requirement shared with the web is the fact that the assignment of documents to terms is not unique, because every document will be linked to many different terms in the thesaurus. Any system supporting the exploration of the documents will have to deal with a large amount of objects that are assigned to many different classes in the conceptual model.

From a functional point of view, the system has to support the user in navigating in the information space providing the following functions:

Query formulation query interfaces that contain a graphical presentation of the available knowledge (ontology), make the query formulation task much easier as the user is already acquainted with the used terminology. The ideal situation would be that queries can be formulated visually.

Query Result presentation inadequate presentation of query answers shadows the performance of many search engines as the user is overwhelmed with results without being able to pick the part that is of interest for him. Graphical presentations can explain the relation of the answer to the original terms of the query.

Query reformulation Very often the user is not satisfied by the returned answer: there are either too many items or no items at all. Graphical presentation of results allows a user to perform:

- Query relaxation: in case of empty sets, by giving up some of the search terms and choosing the results that come closer to the original query.
- Query broadening: in the case of an empty answer set some terms can be replaced by their super-classes. thereby broadening the scope of the query.
- Query narrowing: if there are too many items, the query can be repeated with more specialized sub-classes of certain query terms, narrowing the scope of the query.

Visualizations should help the user in finding relevant information and in addition should make the formulation of queries easier by giving visual clues.

3 Design Rationals

Experimental research has a long tradition in the field of information retrieval and valuable insight has been gained in user behavior and the pros and cons of different retrieval techniques and visualizations. In the following we review existing results in information retrieval focussing on the role of interactive strategies for query refinement including the use of conceptual knowledge and techniques for visualizing search results in a structured way. We will further link previous results to design decisions we made when designing the DOPE browser that will be described in the next section.

3.1 Search Process

Different kinds of search processes can be identified that differ in their characteristics. These include monitoring, performing pre-defined search patterns and exploration of an largely unknown area [23]. Search tasks in the area of drug research largely fall in the last category which is the least structured one and requires special strategies [26]. Traditional information retrieval assumes a process model where the user poses a query to the system based on an existing information need, has the result retrieved and based on the result decides to stop or to reformulate the query in a way that promises to improve the result [25]. This is repeated until the 'perfect' answer is found (compare figure 1). It has been argued that this simple model is not adequate for complex search tasks [8]. This holds in particular for exploration tasks where the information need is not always well defined. As an alternative to the traditional model, the so-called berry-picking model has been proposed [2]. This model assumes that the information need changes during the search process as a result of partial results. In analogue to berry-picking the user takes some of the search results and moves on to a different topic area. This model is more adequate for the exploration task and has been adopted in our work. Figure 1 shows the corresponding process model underlying our work. It assumes that the user starts off with a query on a particular topic. Based on the result the user can either further explore the result set in order to get a better understanding of the results or rescope the search by re-defining the information need and posing a new query to the system. This two level process provides the combination of search (retrieving potential results) and browsing (exploring the retrieved results).

It has been argued that people differ in their search behavior [10] which makes it difficult to provide adequate tool support for the exploration process. Results of a user study indicate however, that users normally prefer interactive methods where the system only performs parts of the search task provides feedback and lets the user decide about next steps [17]. This is a strong argument for an interactive exploration process, despite the fact that it has been shown that interactive methods fail to outperform completely automatic ones, mainly because users often have difficulties in choosing appropriate search terms [20]. It is therefore important to implement a method that leaves the user in control and supports him or her to combine different search strategies (querying and exploration) and to chose appropriate search terms [4]. Controlled vocabularies such as thesauri or ontologies play a key role in in this context [9].

3.2 User Support

As motivated above, the main task of an exploration tool is to assist the user in the different steps of the search process. In the process underlying our work, this mainly means supporting the user in formulating queries that scope the search and in exploring the result of this query.

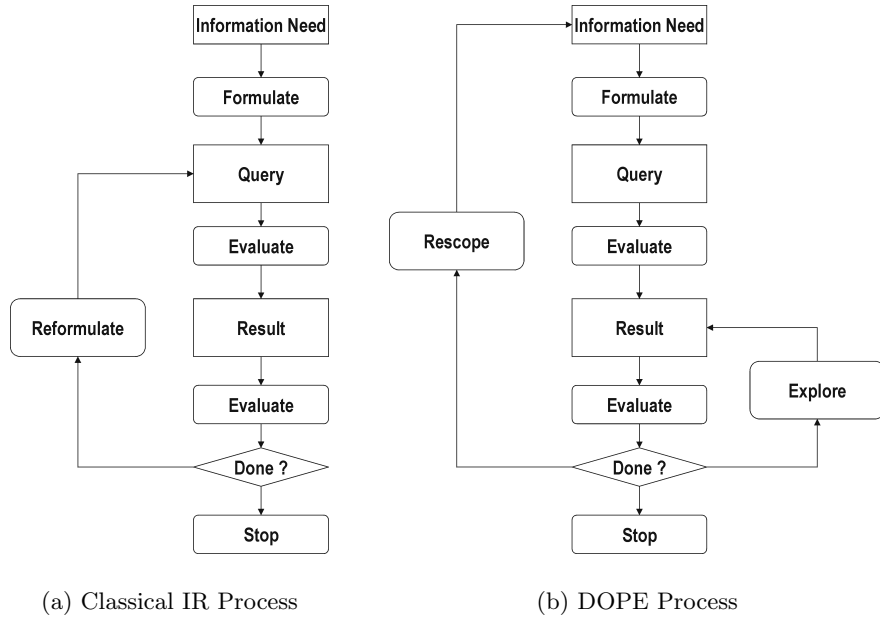


Fig. 1. Process model for Information Search

Query Formulation Considering the task of formulating the initial query, studies have shown that users normally start with quite short and simple queries [1]. It has also been shown that the use of Boolean operators in queries is problematic, because users tend to have an intuition about the meaning of these operators that is different from their logical interpretation used by the system [12]. We therefore abstain from providing a complex query formulation interface for this initial step and rather restrict it to the selection of a search term. As a consequence, supporting the user in the selection of this term is the main task in this first step. In a comparative study Chen and others have shown that term hierarchies provide better support for the user than methods that try to provide an overview of the existing information using term distribution maps and similar techniques [6]. In the DOPE context, we use the EMTREE thesaurus to suggest query terms: The user can browse the hierarchy and select a term or type in a free term that is matched against EMTREE and a corresponding term is proposed. EMTREE is also used to provide support for the re-scoping step. Results of Efthimiadis show that the most useful terms are the narrower terms from the hierarchy [7]. For re-scoping we therefore support the user in navigating the EMTREE hierarchy starting from the previous search term.

Result Exploration The result of the search step is a collection of documents from different sources. As the number of these documents can be quite high, we

have to support the user in exploring these documents. Research studies indicate that other terms occurring in the result documents provide valuable support for further structuring the result set [16]. We therefore extract relevant terms from these documents and offer them to the user as an exploration tool. Major problems have been reported with respect to linking individual results to the different terms used in querying. We address this problem using the principle of faceted queries where the query is split up according to different aspects and documents are explicitly linked to aspects rather than the query as a whole [21]. In our case, documents are explicitly linked to individual terms used for exploration. Different ways of representing these links have been proposed: The keyword in context method [19] extracts a relevant text passage from the document that contains the respective term and presents it to the user. TileBars [13] display the occurrence of different terms in different parts of the document and therefore also provide information about direct co-occurrences. The solution we have chosen is based on the visual representation of terms and documents in a network where documents are connected to terms they contain (compare [14, 18]). The advantage of this representation is that it provides an overview over the result set rather than individual information for each document. We argue that such an overview better supports the exploration task. The specific representation chosen here is based on earlier ideas on providing graphical representations of Boolean Queries [22]. We use the clustermap representation [11] developed by Aduna. The underlying concept of cluster maps is similar to the use of Venn diagrams for query formulation and result representation [15]. In our system, the use of is restricted to the representation of conjunctions of search terms.

3.3 System and Interface Design

When developing a system that provides the user support mentioned above, we are facing several practical design decisions. From the point of view of this paper, the most relevant of these decisions are the ones concerned with the design of the user interface. It has been argued that monolithical interfaces that combine all the information in a single screen are preferably from a cognitive point of view, because otherwise the user quickly loses track of current state of the search process [3, 24]. We therefore chose to provide such an interface. The main disadvantage of this approach lies in the scalability. A single-window interface is naturally limited by the size of the screen and therefore does not scale arbitrarily. We approach this problem by introducing a number of thresholds that limit the amount of information shown at the same time:

- When matching a user-specified keyword with EMTREE we limit ourselves to showing maximally 100 potentially matching concepts.
- The result set consists of all matching documents with a minimum relevance (above 0.02), to a maximum of 500 documents.
- For supporting the exploration of the result set the 10 most relevant concepts from each document are chosen, with a minimum relevance of 0.5.

The first threshold is not a real limitation, because matches not in the first 100 are almost never relevant. The other thresholds are more critical. We rarely encountered less than 500 results for a query. This has been chosen based on technical considerations. As the amount of data influences the response time of the system, the challenge was to find an appropriate trade-off between amount of information and system performance. The current implementation is limited to 500 documents mainly because of high communication costs. Using improved communication protocols will enable us to retrieve more documents in future implementations. The relevance thresholds were determined empirically. They will have to be adjusted based on user feedback.

4 The DOPE Browser

In the presence of an information space of the size of the one used in DOPE, it is not enough to provide a technical infrastructure to query available sources. The user will regularly be overwhelmed by the amount of results and will often not even know what to ask for. In order to address these common problems of information disclosure, we have to provide an intelligent interface that guides the user when exploring the information space and presents the results of a query in a structured way. In this section we describe the user interface of the DOPE system that is based on Aduna visualization technology and explain how to use it when searching for information. Further, we report the results of user studies carried out with potential end users.

4.1 The Browser Interface

We have designed and created a prototype of a user interface client called the "DOPE Browser". It gives users the ability to query and navigate a collection of documents using thesaurus-based techniques, while hiding much of the complexity of the back-end, such as the existence of multiple data sources, any thesaurus or ontology mapping that may take place, etc. In this system, the user sees a single virtual document collection made navigable using a single thesaurus (EMTREE). Typical document metadata such as e.g. title, authors and journal information is associated with each document. Due to this simplified view on the data, the user interface will be easily reusable on other data sets and thesauri.

Suppose a user wants to browse through the existing literature on aspirin. She can enter the string "aspirin" in the text field at the upper left part of the window. The system then consults Sesame for all concepts than can be related to this string. It responds with a dialog showing four possible EMTREE terms, asking the user to pick one. (This dialog is omitted when there is only one exact match with an EMTREE term.) We assume the user chooses the first term, "acetylsalicylic acid", which is the chemical name corresponding with the brand name, and is now the focus term.

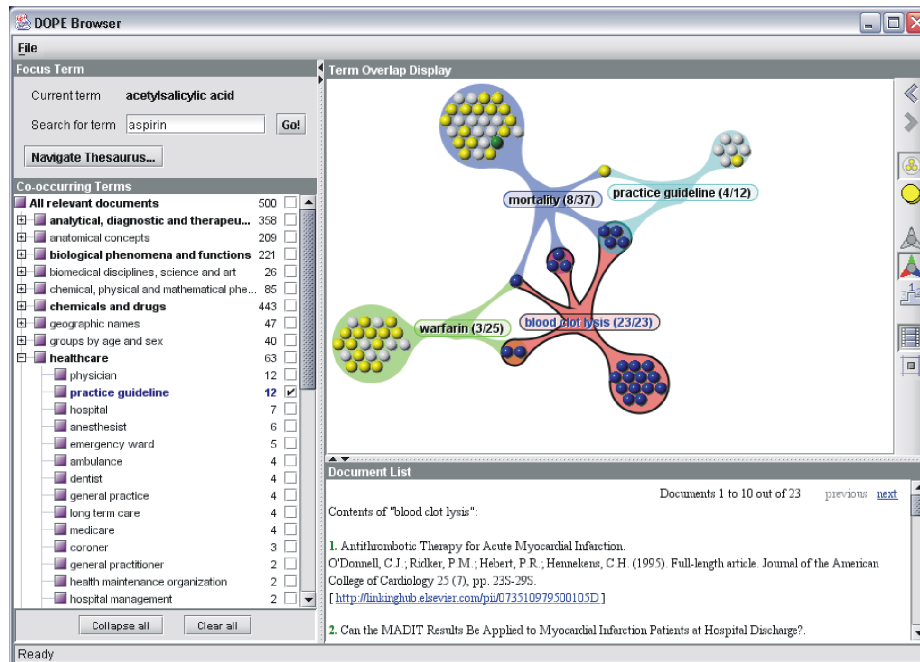


Fig. 2. The Browser Interface

The system consults Sesame again and retrieves the maximally 500 most relevant documents about "acetylsalicylic acid", their metadata fields (such as titles and authors) and the other most important terms with which these documents are indexed. The co-occurring terms are presented in the tree at the left hand side, grouped by their facet term (the most generic broader term, i.e. the root of the tree they belong to).

The user can now browse through the tree and check one or more checkboxes that appear behind the term names, to see their contents visualized at the right hand side.

Figure 2 shows the state of the interface after the user has checked the terms "mortality", "practice guideline", "blood clot lysis" and "warfarin". The visualization graph shows if and how their document sets overlap. Each sphere in the graph represents an individual document, with its color reflecting the document type, e.g. full article, review article or abstract. The colored edges between terms and clusters of spheres reveal that those documents are indexed with that term. For example, there are 25 documents about warfarin, 22 of them are only labelled with this term, two have also been labelled with blood clot lysis, and one is about warfarin, blood clot lysis and mortality. This visualization shows that within the set of documents about aspirin there

is some significant overlap between the terms blood clot lysis and mortality, and that 4 of the practice guidelines documents relate to these two topics as well.

Various ways exist to further explore this graph. The user can click on a term or a cluster of articles to highlight their spheres and list the document metadata in the panel at the lower right. Moving the mouse over the spheres reveals the same metadata in a tool tip. The visualizations can also be exported to a clickable image map that can be opened in a web browser.

The user can start with a new query by typing in a new search string. This will empty the rest of the interface and load a new set of documents and co-occurring terms. The Thesaurus Browser provides an alternate starting point for a next query. When a focus term has been selected, the user can click the "Navigate Thesaurus..." button at the upper left. He is then confronted with a dialog that lets her select a new focus term, by browsing through the thesaurus, starting from the focus term. The user can iteratively select a broader, narrower or alternative term until he encounters a term that she wants to make the new focus term.

The visualization conveys several types of information. First, the user obviously sees document characteristics such as index terms and article types. Visualizing a set of terms shows all Boolean combinations, without the need for the user to express them all separately. Furthermore, the graph also shows within the scope of the selected set of documents how these terms relate, i.e. if they have some overlap and if so, which documents constitute that overlap. Consequently, the geometric distance between terms or between documents is an indication of their semantic distance: terms that share documents are located near one another, and so are documents with the same or similar term memberships.

5 User Studies

At a Drug Discovery conference held in the US in 2003, ten potential end users were given the prototype of DOPE and asked to conduct various tasks. The user group presented a mixture of academic users (6 users from 4th year student to professor) and industrial users, (4 users mainly in leading functions). Users were first given a brief overview of the prototype and were then asked to conduct tasks in the domain of their expertise. To make the results comparative, all users were given similar information and asked to follow the following steps:

1. Identify a search term
2. Explore the co-occurring terms
3. Explore the results with the use of visualization graph
4. Discussion to identify the potential benefits and issues of the prototype.

We will discuss the feedback we got from the users in the individual steps in the following.

5.1 Identifying a search term

The first step the users were asked to perform is the selection of an appropriate term to focus the search and to limit the results shown in the DOPE Browser. It turned out that in many cases the decision to enforce a focus term is too restrictive, as users often wanted to use more than one term for focussing. Another observation was, that users do not always want to start their search using a domain term from the thesaurus, but rather the name of an author or a journal. We conclude that the current prototype is rather suited for exploring the space of available information than for searching for a specific article which would normally require the possibility to define more search criteria. The following terms are examples of typical topics users looked for in the system:

- Genomic
- Glucagons
- Kinase
- Diabetes
- Serotonin
- MHC, peptide epitope
- COX-2 cyclo oxygenase
- Hypertension

5.2 Exploring Co-occurring Terms

The general reaction of users to the use of co-occurring terms for narrowing down the search result was quite positive. Users indicated that the additional terms provide useful context information to refine their initial query. Most user liked the way of organizing the terms in a two level hierarchy, though there is a lack of support for finding a specific co-occurring term. Users mainly used terms from the second level. It turned out that users feel the need for using quite specific terms for narrowing down the query and there were complaints that often the co-occurring terms were too general given the quite specific focus term. We conclude that the use of co-occurring terms as a mechanism for narrowing down the search is useful approach, but that the selection and ordering of the terms for the interface needs more sophisticated solutions.

5.3 The Visualization tool

In general, most users reacted very positively to the visualization and many referred to the graph as "this is what I'm thinking about" or "neat way of jumping between categories". The main two points that emerged from the study are that the visualization provides:

- Richer contextual information about the articles Many users inferred that using the graph helped them to see information that they would have otherwise missed enhancing serendipitous discovery.

- Simpler scanning of the articles. Most users commented that when they scan a list of articles they are looking for one or two keywords to appear in the article title; if the combination is more complex then it becomes increasingly cumbersome to scan the list effectively. The DOPE visualization tool acts as a reminder and map of their search criteria and alleviates cumbersome scanning.

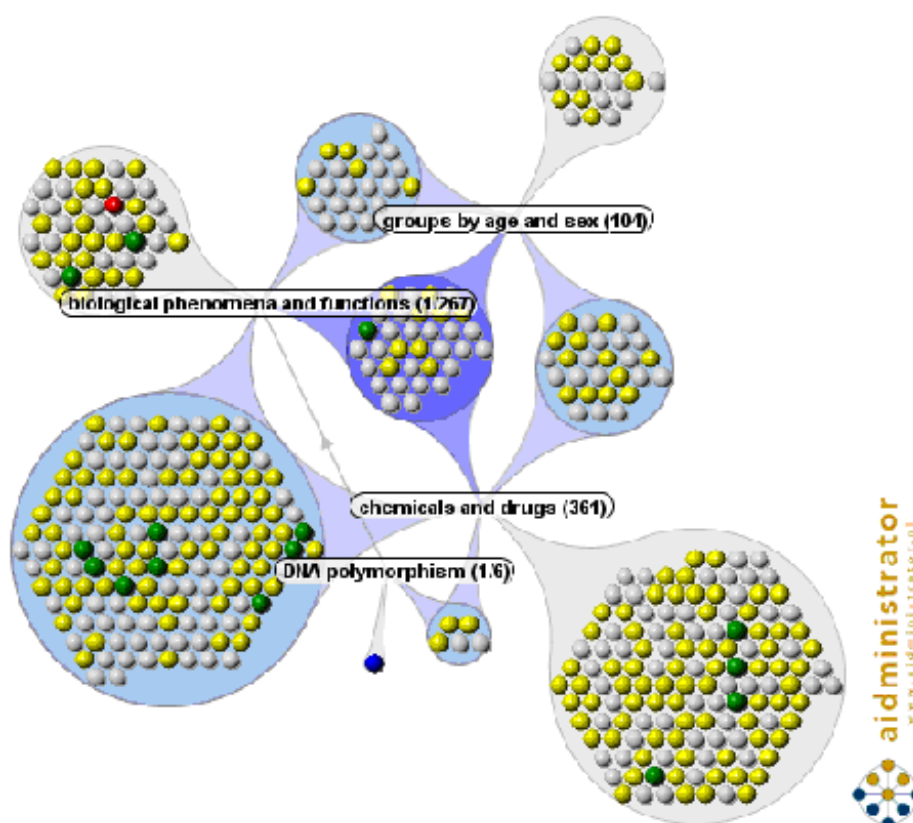


Fig. 3. The visualization where the user has selected 3 terms and a subset term.

There were three issues identified for improving the visualization:

- Interpretation of the subset names In figure 3 above, the user has selected three terms in an area of his research. When users were asked what information they get from this graph, most referred to the terms as labels rather than the unions of the spheres.
- Difficulty to interpret complex term overlaps As with Venn diagrams, the complexity of representation increases rapidly after more than three terms

are applied. Figure 4 above illustrates when the user has selected 4 terms and the visualization gets very complex. There is no immediate way to resolve this in the current visualization but one possibility may be to limit the terms displayed to alleviate from complexity.

- Manipulation of the graph The selection and de-selection of the keywords is available only on the tree at the left hand side of the graph. However, most users were attempting to interact directly with the graph and to manipulate it in the process. This behavior indicates the need for supporting providing direct manipulation of the graph. In addition users were mainly scanning the titles with the rollover feature on the graph and they paid little attention to the list at the bottom frame.

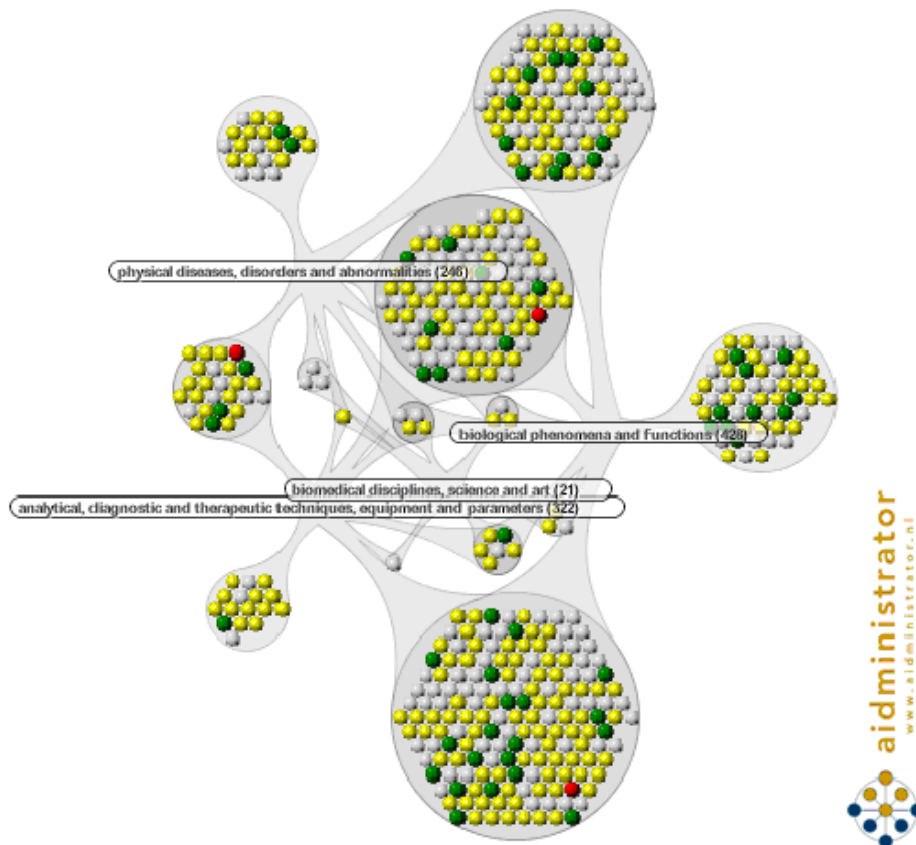


Fig. 4. The visualization where the user has selected 4 terms and a subset term.

5.4 Discussion of potential benefits

Discussions with users about the potential benefits supported the observations made earlier, that the main benefit of the tool is the exploration of a large, mostly unknown information space rather than support for searching for concrete articles. Examples of beneficial applications that were mentioned by potential end users included filtering material for preparing lectures about a certain topic, and supporting graduate students in doing literature surveys. In this context the need for a "shopping basket" for collecting search results was mentioned. Monitoring focus changes in the research community was mentioned as an example of a more advanced potential application. This however would require an extension of the current system with mechanisms for filtering documents based on date of publication as well as advanced visualization strategies for changes that happen over time.

6 Conclusions

We discussed the benefits of using conceptual knowledge for supporting the exploration of large information spaces. We discussed requirements for an exploration tools. The requirements were defined based on user needs in the DOPE project on providing users thesaurus-based access to document repositories in the area of drug development. We reviewed existing tools for exploring RDF-based conceptual knowledge and concluded that these tools do not scale to large instance sets and do not provide sufficient support for formulating and manipulating queries. As our main contribution, we presented the DOPE browser, a topic based browser for large document sets that overcomes of the problems of existing tools.

The main benefit of the DOPE browser is the fact that it supports a mixed-initiative exploration strategy. It combines search for relevant documents based on a keyword with browsing of the result set. Both functionalities are improved using conceptual knowledge provided by a domain specific thesaurus. The thesaurus is used to normalize query terms and extend the query term with related terms. Further, terms from the thesaurus are used to support topic-based instead of link-based browsing. We also reported feedback from user studies. The result of the user studies confirmed our claim that the DOPE browser is a useful tool for exploration.

There are some directions of future work for improving and extending the functionality of the browser. The user studies indicated limitation of the search component of browser. Planned improvements include the ability to use a combination of terms in the initial search and to provide possibilities to use additional metadata such as author names and year of publication to search for documents. We currently extend the browser to the use of more than one thesaurus. In a case study carried out in the tourism domain, we are developing an extended

version of the browser that combines a shared tourism ontology with local classifications used by different users. The connection between the different models is established by automatic classification of documents according to multiple term hierarchies.

Acknowledgments

This work was funded by the Elsevier Advanced Technology Group. The cluster map technology shown in this paper is owned and copyrighted by Aduna (formerly known as Administrator Nederland B.V.).

References

1. P.-G. Anick. Adapting a fulltext information retrieval system to the computer troubleshooting domain. In *Proceedings of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 349–358, 1994.
2. M.-J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5):407–431, 1989.
3. Patricia Billingsley. Taking panes: Issues in the design of windowing systems. In Martin Helander, editor, *andbook of HumanComputer Interaction*, pages 413–436. Springer Verlag, 1988.
4. G. Brajnik, S. Mizzaro, and C. Tasso. Evaluating user interfaces to information retrieval systems: a case study on user support. In *Proceedings of the 19th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, pages 128–136, Konstanz, Germany, 1996. Hartung-Gorre Verlag.
5. Leslie Carr, Simon Kampa, Wendy Hall, Sean Bechhofer, and Carole Goble. Ontologies and hypertext. In Steffen Staab and Rudi Studer, editors, *Handbook on Ontologies*, volume XVI of *International Handbooks on Information Systems*. Springer Verlag, 2003.
6. Hsinchun Chen, R.-R. Sewell A.-L. Houston, and B.-R. Schatz. Internet browsing and searching: User evaluations of category map and concept space techniques. *Journal of the American Society for Information Science*, 49(7):582–608, 1998.
7. E.N. Efthimiadis. End-user’s understanding of thesaural knowledge structures in interactive query expansion. In Hanne Albrechtsen and Susanne Oernager, editors, *Advances in Knowledge Organization. Proceedings of the Third International ISKO Conference*, volume 4, pages 295–303, Frankfurt am Main, 1994. Indeks Verlag.
8. D. Ellis. A behavioural model for information retrieval system design. *Journal of Information Science*, 15:237–247, 1989.
9. Raya Fidel. Searchers’ selection of search keys: Ii. controlled vocabulary or free-text searching. *Journal of the American Society for Information Science*, 42(7):501–514, 1991.
10. Raya Fidel. Searchers’ selection of search keys: Iii. searching styles. *Journal of the American Society for Information Science*, 42(7):515–527, 1991.
11. Christiaan Fluit, Marta Sabou, and Frank van Harmelen. Ontology-based information visualisation. In V. Geroimenko and C. Chen, editors, *Visualizing the Semantic Web*. Springer Verlag, 2003.

12. S.L. Greene, S.J. Devlin, P.E. Cannata, and L.M. Gomez. No ifs, ands, or ors: A study of database querying. *International Journal of ManMachine Studies*, 32(3):303–326, 1990.
13. Marti Hearst. Tilebars: Visualization of term distribution information in full text information access. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 59–66, Denver, CO, May 1995.
14. Matthias Hemmje, Clemens Kunkel, and Alexander Willett. Lyberworld – a visualization user interface supporting fulltext retrieval. In *Proceedings of the 17th Annual International ACM/SIGIR Conference*, pages 249–259, Dublin, Ireland, July 1994.
15. Morten Hertzum and Erik Frokjaer. Browsing and querying in online documentation: A study of user interfaces and the interaction process. *ACM Transactions on ComputerHuman Interaction*, 3(2):136–161, 1996.
16. D. I-Iarman. Towards interactive query expansion. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, Grenoble, France, 1988.
17. J. Koenemann and N.J. Belkin. A case for interaction: A study of interactive information retrieval behavior and effectiveness. In *Proceedings of CHI 96 International conference on Human Computer Interaction*, pages 205–212, Vancouver, B.C., Canada, 1996.
18. Robert Korfhage. To see or not to see – is that the query? In *Proceedings of the 14th Annual International ACM/SIGIR Conference*, pages 134–141, Chicago, IL, 1991.
19. Julian Kupiec, Jan Pedersen, and Francine Chen. A trainable document summarizer. In *Proceedings of the 18th Annual International ACM/SIGIR Conference*, pages 68–73, Seattle, WA, 1995.
20. Mark Magennis, J. Cornelis, and K. van Rijsbergen. The potential and actual effectiveness of interactive query expansion. In *ACM SIGIR Forum*, volume 31, pages 324–332, 1997.
21. Charles Meadow, Barbara Cerny, Christine Borgman, and Donald Case. Online access to knowledge: System design. *Journal of the American Society for Information Science*, 40(2):86–98, 1989.
22. A. Michard. Graphical presentation of boolean expressions in a database query language: design notes and an ergonomic evaluation. *Behaviour and Information Technology*, 1(3):279–288, 1982.
23. V.L. O'Day and Robin Jeffries. Orienteering in an information landscape: how information seekers get from here to there. In *Proceedings of the INTERCHI'93*, Amsterdam, Netherlands, April 1993. IOS Press.
24. Ben Shneiderman. *Designing the user interface: strategies for effective humancomputer interaction*. AddisonWesley, Reading, MA, 1997.
25. Ben Shneiderman, Donald Byrda, and W.-B. Croft. Sorting out searching: A userinterface framework for text searches. *Communications of the ACM*, 41(4):95–98, 1998.
26. J.-A. Waterworth and M.-H. Chignell. A model of information exploration. *Hypermedia*, 3(1):35–58, 1991.