Search Intention Analysis for User-Centered Adaptive Visualizations

Dirk Burkhardt, Matthias Breyer, Kawa Nazemi, and Arjan Kuijper

Fraunhofer Institute for Computer Graphics Research, Fraunhoferstr. 5
64283 Darmstadt, Germany
{dirk.burkhardt,matthias.breyer,kawa.nazemi,
arjan.kuijper}@igd.fraunhofer.de

Abstract. Searching information on web turned to a matter of course in the last years. The visualization and filtering of the results of such search queries plays a key-role in different disciplines and is still today under research.

In this paper a new approach for classifying the search intention of users' is presented. The approach uses existing and easy parameters for a differentiation between explorative and targeted search. The results of the classification are used for a differentiated presentation based on graphical visualization techniques.

Keywords: User Intention Analysis, Search result visualization, Semantic Visualization, User-Centered Interaction, Adaptive Visualization, Semantic Web.

1 Introduction

Acquiring information and knowledge from web-based knowledge repositories is already part of our daily life. In our daily routine we are confronted with many expressions that we either do not know or have superficial knowledge about. These situations often force us to use web-repositories for gathering mostly rudimentary knowledge about a given topic. This behavior makes the most of us knowing about many issues very less and leads to elementary changes in our society. For offering an adequate level of information quantity, granularity and depth, it is more and more necessary to possess knowledge about the users' pre-knowledge and intentions. Today many systems collect information about the users and model them in different ways for offering or recommending different information artifacts and different types of information, whereas the way of their presentation is not investigated in depth.

Beside the changes in the information acquisition, the already existing trend of alternative interaction paradigms with computer systems leads to a need for investigation and research alternative and optimized ways of presenting information and search results. Thus natural gesture-based interaction ways are a promising alternative to the outdated mouse and keyboard interaction. Especially graphical user interfaces will play a key-role in future information and search systems. For the usual user a natural interaction with graphical objects is easier to proceed than with text-based systems.

C. Stephanidis (Ed.): Universal Access in HCI, Part I, HCII 2011, LNCS 6765, pp. 317–326, 2011.

Graphical objects are more similar to our perception of and interaction with the real world and the objects within.

Although the usage of graphical information representation is obvious, today's search results are almost entirely presented as text-based lists. A widespread deployment of information visualizations has not taken place yet. The usage of information visualization systems are principally reserved for scientific and economic areas. Users with no or average skills in information visualization perceive the interaction with and the information acquisition from these system not as intuitive and are overstrained. Other users with a higher graphical comprehension experience the graphical representation as infantile and not gainful.

To bridge the gap between the existing heterogeneous users, it is necessary to develop adaptive systems, which are able to adapt automatically to the requirements, pre-knowledge and intentions of users. The challenges in this area are the identification of relevant information about the user, the implicit extraction of this information and the problems of "new user", "new content" and "new context", based on the adaptation method.

To face these challenges an adaptive visualization for semantically annotated data will be presented in this paper. The adaptation will especially face the new user and new content problem by analyzing the first user interaction in a searching process: the entered search query terms. Therefore a new method for analyzing the user's intention based on the entered search query input is presented. This method classifies the search intentions of users as "explorative" and "targeted", thus these are the two most common and most clear searching behaviors. For differentiating between the search-classes, word frequency algorithms are used. Each of the search terms are analyzed according their Frequency List class, in which the words of a language are classified, based on their occurrence. The targeted search behavior on the one hand is a specific and very clear search and requires pre-knowledge in the searched topic. A targeted search consists always of multiple search terms and indicates precise results. An example for this class of search could be the input of the terms "Workflow WFMC XPDL". A targeted search is not only defined by the number of the search terms, but also by the frequency of each term compared to the occurrence of the given language. An explorative search on the contrary consists of one or more words which are high frequently used. These words indicate that the user wants general information about a topic. In most cases the searched topic is unknown or the user's pre-knowledge is rudimentary. Examples for this class of search could be the input of the search terms "Workflow" or "what is a workflow". The explorative search contains only one term with a high Frequency Class and hence with a low occurrence compared to the given language.

In this paper we first present related works to behavior-driven adaptive visualizations. We will show that the number of these systems is very limited, whereas the added value is higher than expected. After that the analysis method will be describe, with a differentiation between targeted and explorative search behavior. In the following section the visualization of the search results based on the search behavior analysis will be described. Here the aspect of the adaption to the users' requirements is the focus. We will conclude this paper with an evaluation of the system, which will argue the deployment of our method in future applications.

2 Related Works

The search process is in focus since larger databases can be created and conducted with technical systems, but they becoming a central information source since a large number of users are able to use or enrich such data sources in the internet. For an economical success of a search engine the quality of a search result is very important, because an information database can only be useful contact point, if the results are relevant to the searched query. Thus the search engine plays an important role with its algorithm generating a list of results, basing on a issued query. Many algorithms were developed in the past and some of them are also published in journals and on conferences. But often these basic algorithms like the Keyword search over relational tables and streams [3] or by using a partially observable markov model [4] are only a few of different used algorithms that often will be combined to determine well matches. Often these combinations of different algorithms are issued confidential by the companies which provide a search engine. But this circumstance makes it hard to compare the different existing search engines on the market.

The evaluation of existing search engines and also their usability is in focus not only by the providers themselves. Also for partners or researchers it is important to know, which search engine is adequate for testing a new approach or algorithm. Often the evaluation primary concerns to the quality of the generated result list. The challenge for this goal is to find an automatic method for determining the quality, for instance “how can the quality of a generated result set be measured?” Two examples for realizing published by Hochstotter et. al. [1] and Zaragoza et. al. [2], but of course there exist more strategies.

With the up growing of the Semantic Web also further search strategies were possible, by regarding its semantic structured background [5][6]. In the past, if a user was searching for e.g. the term ‘Golf’, the search engines were not able to identify if the automotive or the sport was meant by the user. By searching through a structured e.g. semantics database it is possible to let the user choose the relevant concept and provide him so a better result set with more helpful information. Another advantage is that further kinds of interacting through the result set can be provided, like an interactive and dynamic search request, if the user is clicking on a concept or instance, the search request can be extended to improve the quality of the result set. By supporting this interactivity, the user is not forced to improve the results only by adding further key words, he also can improve it by clicking on a concept e.g. to filter the result set.

Increasing the quality of search results can also be achieved by adapting the search strategy of the search engine in dependence to the user, so the algorithm adapts the results to the behavior of the user. The adaption of the search results will be made by generating a user model, which regards the goals, experiences etc. of a user to optimize the result set. This generation of a user model can be achieved with different approaches. A common approach is classifying into a limited number of groups. An example for this procedure is presented in Teevan et. al. [7]. A group is characterized by different user attributes like the age, gender or location. So with these characters a user can be allocated in a group, which goes along with an influence and an optimization of the search results. Similar to this approach, the search results can be optimized by the social context of a user [8]. The idea based on the existing

bookmarks of the users on which a profile will be generated and in fact of this profile, it can be matched and compared with done searches by users with a similar profile. It is also possible to regard the contents of social networks. Another aspect is the optimization basing on the interest, like it is described in Ma et. al. [9]. During the entire search process of the user, the search request and results will be classified to an interest field and stored in a database. By every other following search request, these interests will be used, to optimize the search results. This can be done for a single user and also for all or also a group of users.

The chapter described approaches basing on the idea of generating adequate results to a user query. But not only the generated results are important for a user, the type how information will be presented for the user is a main topic too. So in difference to result generation, aspects for usability and user experiences are main factors that have to be investigated. For providing an adequate data visualization an adaptive visualization system can be used to adopt the results for the presentation on the users' behavior [10]. But also non-automatic adaption systems can support the user by the information visualization, like the possibility to create a so called "Knowledge Cockpit", where a user can couple preferred visualization to work and interact with the results of a search request [11].

3 Search Intention Analysis

3.1 Users' Searching Behavior

Information and Communication Technologies (ICT) do not only provide new opportunities for searching and gathering information, they have already changed the social behavior of human in different aspects of life.

An example for the "new behavior" could be the research homework of student in school or university: A student gets a topic for research as homework and has to present it. Just few years ago the student had to go to libraries and search in books for the given topic. The information acquisition process was longer and not that easy that easy to handle, but the gathered knowledge was deeper. The students were confronted to process the whole information acquisition process manually, finding the right books and further literature, searching in the books and notice the relevant points for using them in their presentation.

In contrast to the past, the information acquisition process has changed totally. Today's students are searching primary in the web, using search machines or knowledge data bases. The knowledge and information acquisition process is faster, but not deep. The acquired information are mostly rudimentary and thus the process itself s fast, the general knowledge of users are more but not deep.

With the changed information search behavior, different questions overcame in the last years, which need to be investigated. One main question is: what is the search intention of the user searching for a term and how the information can be adequately presented to the users.

Silverstein et al. presented in [12] statistical values of search behavior of 285 million search sessions and about one billion search queries. Our concept for analyzing search intentions considers two main search parameter, on the one hand the

number of terms used in a search query and on the other hand the Frequency List Class of each term. For demonstrating the distribution of the term quantity the work of Silverstein [12] is essential and gives very good overview.

The following table shows the statistics concerning the number of terms per query [12]:

Table 1. Statistics concerning the number of terms per query [12]

0 terms in query:	20.6%	max terms in query:	393
1 term in query:	25.8%	avg terms in query:	2.35
2 terms in query:	26.0%	stddev of terms in query:	1.74
3 terms in query:	15.0%		
> 3 terms in query:	12.6%		

The distribution of the term quantity shows that most valid search queries (not included are search queries with zero terms) contain one term or two terms, with each about 26 percent. The further distribution shows that three terms in query and more than three terms are rough equal with 15 and about 13 percent.

3.2 Targeted and Explorative Search

A very simple classification of search behavior is the differentiation between targeted and explorative search. A targeted search can be defined as an explicit search for a fact or explicit information. This kind of search anticipates Declarative Knowledge or information as result. The explorative search does not require explicit knowledge about the domain by the user. The user may browse in the domain to gain an overview or to get a picture about something he is not familiar with. If the user has some pre-knowledge he may want to perceive some additional facts or he may want to reflect the knowledge by reading, thinking and thus cognitive structuring the facts again.

Adaptive visualizations facilitate users in perceiving the information in search results. These visualization techniques preprocess the data using additional information about the search domain or user specific data to present the information adequately and user adapted. Thereby an important criteria is the differentiation of the users' search intention, thus the differentiation in targeted and explorative search.

The classification of search queries to the class of targeted search intention can be performed by the measurement of two criteria: a) quantity and b) frequency of search terms in the query.

These two criteria quantity and frequency will be discussed in detail in the subsequent sections.

3.3 Term Quantity in Search Queries

A very important parameter for identifying the targeted search is the quantity of the search term. The quantity of terms in search queries is an indication for existing preknowledge by the user. A targeted search must contain more than one term. Only if a search query is really specified, it can be categorized as targeted.

An example of a search with high quantity of search terms can be the typing of the question following question as search terms: “When was president Obama born?”. The anticipated information is very specific and estimates a date. The same targeted search can be formulated as “President Obama birthday” or “day of birth Barack Obama”.

If we take the above example and the user search would be formulated just by one term, e.g. “Obama”, it is not really clear what is really searched by the user or which information is expected. Maybe the user searches for the general information about president Obama or someone else with the same name; further the user may wants to know “what Obama is” and he is searching for example for the Japanese city “Obama”. It can be specified really clear that a targeted search contains more than one search term. But how is it vice versa? Are all search queries which contain more than one term, targeted searches?

This question is already answered in the section above: If a user does not know anything about a term and just wants to have general information or browse through the knowledge related to the term, the queries may be formulated by more than one term, e.g. “What is Obama?”.

So, it is more than just the quantity of the terms is needed to separate between the targeted and explorative search and provide an adequate presentation based on the users’ intentions.

This second criterion and measurement can be solved with a really easy and existing algorithm, the Frequency Lists.

In the following section the classification using the Frequency Lists is described.

3.4 Term Frequency in Search Queries

The term frequency describes the number of occurrences of a term in a specific language. The more a term is used in a language the more common it is. But the more a term is used in a language the more general and unspecific it is. Therefore the term frequency has to be interpreted inverse. The more specific a term is the more preknowledge has to be given by the user and the more targeted is the search query.

A well known inverse interpretation of term frequency is the TF-IDF algorithm, which calculates relevancies of terms for specific documents by examining the term occurrence in the specific document compared to the term occurrence in the whole document space. For the analysis of the relevancies of terms in search queries the likewise examination of occurrences of terms in the query space does not yield reliable results because these queries contain too little information. Nevertheless TF-IDF values can be used to identify the relevance of search query terms in the context of the search space respectively the document space. But this requires an expensive preprocessing of the search space. Even worse if the search space documents change, like it is given in the internet, the preprocessing has to be repeated time after time yield reliable term relevancies.

To overcome this problem a new approach for search term relevance measurement is described in this paper which does depend on the search space but only on the language specific term occurrences.

The search term relevance can be measured using language specific frequency lists. These frequency lists count the number of occurrences of a term in the natural

language and arranges them according the occurrences. Afterwards the terms are chunked and assigned to frequency classes which are valued. The higher the value of a frequency class the more specific is the term. Even more the values are logarithmic, so the assignment of a term to the next higher frequency class indicates that this term is used half as often than terms of the lower frequency class and thus twice as specific.

For example the term 'the' is used most often in the English language. Thus it is in the first frequency class values '0'. Thus a term assigned to the frequency class valued '1', let's say the term 'that', is used half as often as the term 'the'. It follows that the term 'that' is twice as important or specific for the search query in comparison to 'the'.

The main advantage of this approach is that no entire preprocessing of the whole search space is required to gain reliable term relevance values. Adding new documents to the frequency classes only results in increasing numbers of occurrences of used terms. Afterwards the chunks are recalculated which determine the frequency class values. Thus it is not necessary to reprocess the whole search space.

The frequency classes can be used to measure the specifically of a term in the search query. The higher the frequency class or classes are the more a targeted search is given. This holds because the more specific a term is the more accurate is the users' intention about the search domain.

3.5 Targeted search intention deduced by Quantity Measurement and Frequency Lists

A targeted search is given when the user states a search query using multiple terms (quantity) and specific terms (frequency). This can algorithmically be determined by summarizing the frequency class value of the search terms. If the calculated value exceeds a threshold value a targeted search is given.

The quantity of terms in search queries is an indication for existing preknowledge by the user. The more terms are stated in a query a user send the more specific his intended question formulated. This assumption holds if at least one of the two following assumptions are fulfilled: Either the terms localize the result space in a semantically way or the terms localize the result space by their frequency class, thus their specificity. Due to the fact algorithms for semantic interpretation are expensive and not reliable enough by now the presented approach of using frequency classes to determine the assignment of search queries to the classes targeted and explorative search yields to the most reliable results while only a few preprocessing is required.

The frequency class of a term in search queries is an indication for the accuracy with which the search domain is localized. The more the search domain can be enclosed by specific query terms the more awareness about the targeted search results is given for the user.

The combination of these two criteria quantity and frequency determines if a search query represents a targeted search. Furthermore this approach is easily applicable and yields to reliable results.

The examined class of targeted of explorative search can be used to adapt the search result visualization automatically to the users search intent. If the user stated a targeted search visualization techniques should be presented which let the user

perceive the targeted information. In comparison when an explorative search intention is given other visualization techniques should be used to allow an explorative search through the search result space. Even more this explorative search result should provide the user the opportunity to explore through the document space, not limited to the search result space, because it can be assumed the user had no specific intention about what he is looking for.

In the next chapter a case study is given for the automatically adaption of search result visualization for targeted and explorative search intentions.

4 Visualization of Search Results: A Case Study

In the last chapter an approach had been presented to identify if a search query represents a targeted or an explorative search intention. In this chapter an adaptive search result visualization for targeted and explorative search intention is presented.

If the user stated an unspecific search query like 'Obama' the examination of quantity and frequency values yields to the assumption this query represents an explorative search. Therefore visualization techniques are presented which allow an explorative navigation through the search result space and even more through the document space, like presented in Fig. 1.

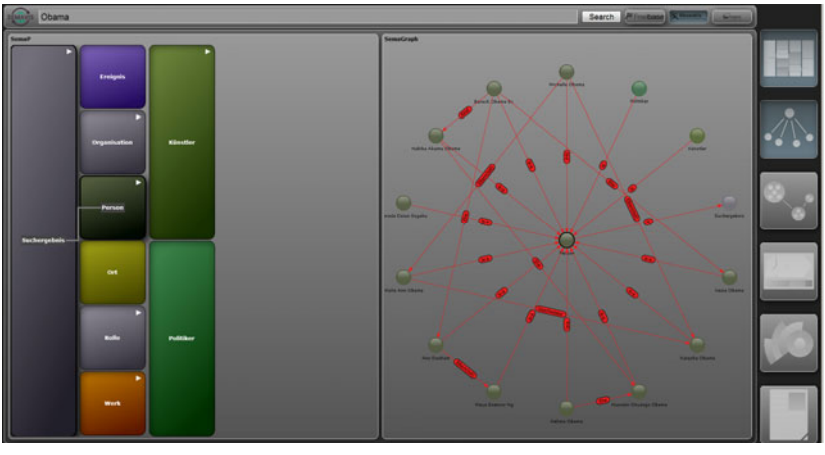

Fig. 1. Search result visualization for an explorative search intention

In contrast, if the user stated a query with a high quantity of terms and these terms are valued high by their frequency classes using frequency lists, a targeted search intention is given. To adapt the search result visualization to the targeted search result the user wants to perceive specific information about his targeted information object. Thus visualization techniques are applied which allow the user to read detailed information about the search result space, like presented in Fig. 2.

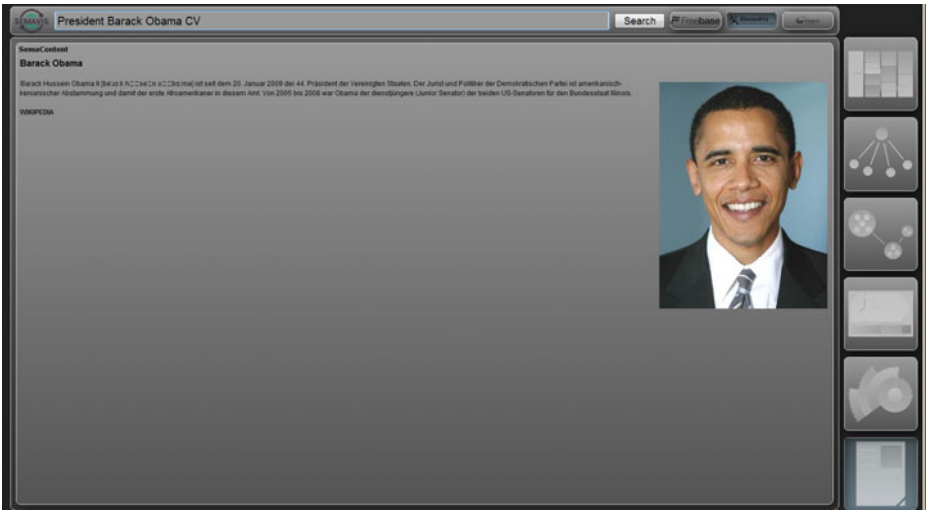

Fig. 2. Search result visualization for an targeted search intention

5 Conclusion

In this paper we presented a new approach for classifying users' search behavior for visualizing the results based on the classification. The approach classifies user search behavior in two categories, targeted and an explorative search. For the classification two different parameters of users' query are used, the quantity of the search terms and the Frequency List class of each searched term.

The paper concluded with a differentiated visualization of the different search classifications. Therefore a targeted search is visualized with a single content. An explorative search presents the information in acockpit of visualizations and enables an exploration of the information.

Acknowledgments. The here described work was developed as a part of the Core-Technology-Cluster for Innovative User Interfaces and Visualizations of the THESEUS Program, a 60-month program partially funded by the German Federal Ministry of Economics and Technology. For more information visit semavis.de or theseu-programm.de.

References

1. Hochstotter, N., Koch, M.: Standard parameters for searching behavior in search engines and their empirical evaluation. Journal of Information Science 35(1), 45–65 (2009)
2. Zaragoza, H., Cambazoglu, B.B., Baeza-Yates, R.: Web search solved? All result rankings the same? In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM 2010), pp. 529–538. ACM, New York (2010)
3. Markowetz, A., Yang, Y., Papadias, D.: Keyword search over relational tables and streams. ACM Trans. Database Syst. 34(3), Article 17, 51 pages (2009)

4. Wang, K., Gloy, N., Li, X.: Inferring search behaviors using partially observable Markov (POM) model. In: Proceedings of the third ACM International Conference on Web Search and Data Mining (WSDM 2010), pp. 211–220. ACM, New York (2010)
5. Chen, Y., Wang, W., Liu, Z., Lin, X.: Keyword search on structured and semi-structured data. In: Binnig, C., Dageville, B. (eds.) Proceedings of the 35th SIGMOD International Conference on Management of Data (SIGMOD 2009), pp. 1005–1010. ACM, New York (2009)
6. Demartini, G.: Leveraging semantic technologies for enterprise search. In: Proceedings of the ACM first Ph.D. Workshop in CIKM (PIKM 2007), pp. 25–32. ACM, New York (2007)
7. Teevan, J., Morris, M.R., Bush, S.: Discovering and using groups to improve personalized search. In: Baeza-Yates, R., Boldi, P., Ribeiro-Neto, B., Barla Cambazoglu, B. (eds.) Proceedings of the Second ACM International Conference on Web Search and Data Mining (WSDM 2009), pp. 15–24. ACM, New York (2009)
8. Carmel, D., Zwerdling, N., Guy, I., Ofek-Koifman, S., Har'el, N., Ronen, I., Uziel, E., Yogev, S., Chernov, S.: Personalized social search based on the user's social network. In: Proceeding of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009), pp. 1227–1236. ACM, New York (2009)
9. Ma, Z., Pant, G., Sheng, O.R.L.: Interest-based personalized search. ACM Trans. Inf. Syst. 25(1), Article 5 (2007)
10. Nazemi, K., et al.: A conceptual model for adaptive visualization. In: Proc. of the HCI International 2011. LNCS, Springer, Heidelberg (2011)
11. Nazemi, K., Burkhardt, D., Breyer, M., Stab, C., Fellner, D.W.: Semantic Visualization Cockpit: Adaptable Composition of Semantics-Visualization Techniques for Knowledge-Exploration. In: International Association of Online Engineering (IAOE): ICL on Interactive Computer Aided Learning (ICL 2010), pp. 163–173. University Press, Kassel (2010)
12. Silverstein, C., Marais, H., Henzinger, M., Moricz, M.: Analysis of a very large web search engine query log. SIGIR Forum 33(1), 6–12 (1999), http://doi.acm.org/10.1145/331403.331405, doi:10.1145/331403.331405