Analyzing the Information Search Behavior and Intentions in Visual Information Systems

Kawa Nazemi[1,*], Dirk Burkhardt[1,2] and Arjan Kuijper[2]

[1]*Darmstadt University of Applied Sciences, Darmstadt, Germany*

[2]*TU Darmstadt, Darmstadt, Germany*

Abstract: Visual information search systems support different search approaches such as targeted, exploratory or analytical search. Those visual systems deal with the challenge of composing optimal initial result visualization sets that face the search intention and respond to the search behavior of users. The diversity of these kinds of search tasks require different sets of visual layouts and functionalities, *e.g.* to filter, thrill-down or even analyze concrete data properties. This paper describes a new approach to calculate the probability towards the three mentioned search intentions, derived from users' behavior. The implementation is realized as a web-service, which is included in a visual environment that is designed to enable various search strategies based on heterogeneous data sources. In fact, based on an entered search query our developed search intention analysis web-service calculates the most probable search task, and our visualization system initially shows the optimal result set of visualizations to solve the task. The main contribution of this paper is a probability-based approach to derive the users' search intentions based on the search behavior enhanced by the application to a visual system.

Keywords: Information search behavior, Human-computer-interaction, Information retrieval, Predictive analysis, Interactive search, User-centered-design, Adaptive visualization, Distant supervision.

1. INTRODUCTION

Modern information systems deal with a variety of large data sources. In particular semantically annotated data sources, such as Linked Open Data sources like DBpedia or Freebase, can become very complex and hence consist of multidimensional data and high volume of information density. This mass of data and the high information density enables combined with todays' high performance computer systems a broad band of use cases. However, the more functionalities users get in using information systems, the more difficult and confusing the interaction and search may become.

To support users in their work with information systems a variety of approaches has been investigated. Successfully proofed approaches are among others guidance or assistance of users in information systems. Recommender systems or adaptive systems are able to adapt the user-interface in dependence of concrete retrieved data, towards the user's behavior or the task a user is performing. This comes along that the user will see only those data visualizations that are really useful. All of these approaches have in common that they need some pre-knowledge, either based on a couple of user's interactions or in general about the search context such as the concrete task and goal, until they are able to support a certain user. This lack of current information systems increased the trend of providing only a simplified search interface. Search interfaces do often only provide a single search input field, where user can enter a key or search phrase to query one or multiple data sources. In particular for information systems that can handle different types of data sources, such as knowledge bases like DBpedia or digital library sources such as the digital library of the Eurographics Association or statistical data-sources such as Eurostat, it is difficult to provide the right choice of result visualizations. To support users in working with computational information systems there is merely the query phrase that can be analyzed.

Search intention analysis is the idea of identifying the search purpose of certain users almost on the basis of the entered key phrase. The current existing approaches focus often on basic context information, such as the GPS position or the device type itself. More advanced approaches using a semantic-based analysis, for instance based on the identification of interrogative pronouns or via mapping single words of the key phrase to semantical concept, *e.g.* a geographic region requires showing the results on a map. The degree of user support is often limited, either they ignore the purpose (based on the semantic meaning of the query) of an entered key phrase or they depend strongly on the query language and makes them hard to use in multi-lingual environments.

In this paper we describe a new approach to classify search phrases into three groups of search

[*]Address correspondence to this author at the Darmstadt University of Applied Sciences, Darmstadt, Germany; Tel: +49.6151.16-39393; Fax: +49.6151.16-39413; E-mail: kawa.nazemi@h-da.de

E-ISSN: 2410-2938/17

intentions. As search intention we defined the three search types: (1) targeted, (2) exploratory and (3) analysis search. The paper explains the major advantage of this classification and provides a brief study on existing approaches and a comparison. To classify search phrases, our approach uses the parameters: word frequency, query length and the distant supervision method. The approach focuses mainly on word frequency analysis, which is realized with the help of two web-services, namely the *Wortschatz* frequency service by the University of Leipzig and the *Microsoft Ngram* service (now part of the Microsoft Cognitive Services). It should be mentioned that our approach is not restricted to these services, since these services just calculate a frequency value on the basis of a large document collections. Majorly, on the basis of the retrieved word frequency for each word of the search query, we created an algorithm that calculates if the search intention is more likely a targeted, an explorative or an analysis search. In a first prototype we enhanced our information search system that supports a variety of different data sources. On the basis of the defined query classification, the system is now able to present an appropriate initial set of visualizations and therewith support the users directly in their work. To proof our concept, we have evaluated the classification algorithm with hundreds of real user search queries and their relation to the main users' intentions. Therefore, we created a survey tool on which users could mention their performed search queries and their intentions of search (based on their real performed searches on google, idealo etc.).

2. OVERVIEW AND RELATED WORKS

User interfaces can be adapted in general based on the visualized data, the user that is using the system and therewith based on his preferences and user's behavior, and the task the user solves with the system [25]. In perspective of big data and the increasing funtionalities to perform analyses, the task-based adaption is an adequate way to support users in their data analysis work. Adapting the user-interfaces based on each single task requires huge effort in investigating on all necessary steps that needs to be done. However, most often it is not necessary to support the user on the basis of individual tasks. Instead, it is more performant and even similar effective to group the tasks and provide support on the basis of the identified task groups. In particular, graphical information system that include various data sources enabling the performance of different tasks with the same tools and functions but on different data sources. There is commonly no need to consider each possible analysis task for each single data source. The challenge is the definition of optimal task groups.

A number of visual task classifications already exist, such as the classification of Shneiderman [7], Buja *et al*. [6] or Zhouh and Feiner [4]. Zhouh and Feiner investigate the way from information to user [4]. They elaborated the enabling and informing users. Informing users by elaboration and summarization, premises the information *(targeted) searching task*. They further propose that search is a sub-task of exploration in their enabling category [4]. Fluit *et al*. propose in their simplified task classification search as one of the three high-level tasks and call it query [5, 6]. As the most of the presented classifications consider search as an important and fundamental task in visualizations.

Beside search, the high-level visual task *explore* plays an essential role. The task classifications show that exploring information plays a key-role for each classification. Shneiderman's model proposes a top-down seeking model with the characteristics of exploration [7]. From overview to detailed information can be assigned as an exploration tasks [7]. Zhou and Feiner explicitly name the task explore is a higher-level task of search and verify [4]. Yi *et al*. have their own categorization for explore, although their classification is not considering the high-level tasks [8]. The classification of Pike *et al*. [9] assigns the task explore as the high-level task on the user-goal and tasks level.

In particular the classification by Keller [10] proposes a different view on solving visual tasks. They propose that the main visual interactions are solving more *analytical (search) tasks*. Their task classification can be abstracted to analyze. Zhouh and Feiner differentiate in their model different aspects of the task "analyze". In particular, the task category enables and verifies leads to the higher-level task analyze, whereas some aspects of inform and summarize are addressing the analysis task [4]. Pike *et al*. identified the task analyze already as a high-level task in their model [9]. Further they assigned the task "compare" as a high-level task too, whereas other works (*e.g*., [10]) assigns compare as a sub-task of the analytical visual problem solving process. As the analytical tasks plays an important role in all presented classification, analyze should be assigned as a high-level visual task.

In perspective of user-interface adaption based on search queries and the search intention they are very detailed. So, to classify search tasks, we oriented on the given task classification and reduced them on the three higher level tasks or task categories (see also Figure **1**) [1, p. 40]: (1) *targeted search*, (2) *exploratory search*, and (3) *analysis search*. The reduction of search tasks on these three groups is the result of the experience we collected over the past years. The search intention of almost any web search can be classified in one of these three categories. In dependence of the identified search category, the initial visualization orchestration can be defined. For instance, a targeted search requires visualizations that enable an effective filter mechanism to a concrete result entity. In contrast to a targeted search, an exploratory search requires overview visualization followed by thrill-down functionalities. And analysis search requires visualizations that allow an analysis *e.g*. on quantitative data.

Recent search intention analysis approaches make use of additional information (most often context information). This additional information investigate focus more on the behavior, feelings and process during the search process [11]. An overview of features that can be used as the above-mentioned user's searching behavior are listed in Figure **2**. However, most of the approaches are dedicatedly designed for concrete scenarios, such as filtering data on geographical information such as the GPS position of user, who uses his smartphone in a city in front of a shop, and are less beneficial in generic application use-cases. In contrast to these specialized search scenarios we focus more on classical search scenarios that also work on desktop computers.

3. DESIGN OF AN VISUAL INFORMATION SYSTEMS THAT INCLUDES SEARCH INTENTION ANALYSIS

3.1. Overview and Architecture

The general approach aims on the provision of a micro-service to enable the inclusion in different graphical information search systems. This enables to enrich any kind of systems with the feature of search intention analysis (see also Figure **2**). Therefore, the API consists of a simple design. As argument the

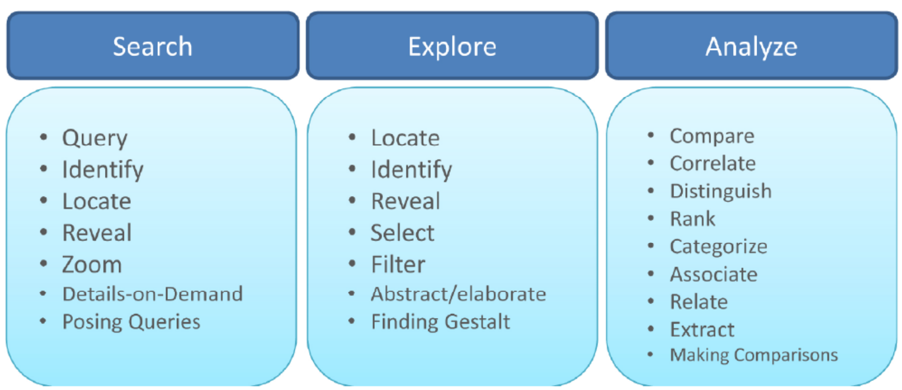

Figure 1: High-Level Tasks with their Sub-Tasks [1, p. 40].

Approach	Advantages	Disadvantages
Inferring user Search Intention Based on Transaction Logs of users (*e.g.* [17])	Does not depend on external content, Implemented in real time, Uses large data-set	Assigned each query to one and only one category (can have multiple intents)
Characterizing Queries using The Variables based on Text (*e.g.* [18-20])	Tackles the problem of building query classifiers based on different sources of information	Requires a method for incorporating the influence of new words
Classifying User Queries based on Query Phrase (*e.g.* [21])	Do not require any external or additional information	Not enough to classify all the user intents by using only the content of words in query
User Search Intention based on Situation Analysis of the Physical World (*e.g.* [22])	Uses a consistent Framework for both the extraction of a user's search intentions and low-level situation inference.	Requires a hardware setup
Discovering User's Specific Geo Intention analysis from Web search logs (*e.g.* [23])	Works even when the explicit geo information is missing	Only analyzing the geo intent is not enough

Figure 2: Table of features and approaches that can be used as user's searching behavior to enable prediction of search intention.

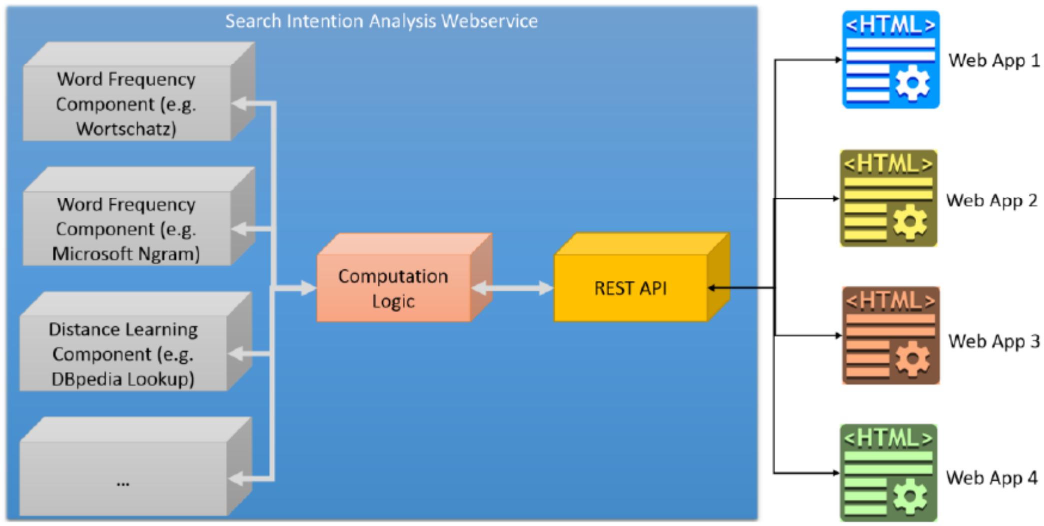

Figure 3: Architectural Overview of the Search Intention Analysis Web-service.

REST API expect the query, the user has formulated as search phrase in the search field of the information system. The computational logic generates now the scores for the three major search tasks (1) targeted search, (2) exploratory search and (3) analysis search [27]. The calculated scores are then sent back to the information search system that now use it to adapt the user interface in dependence of the most probably performing task.

3.2. Search Intention Analysis Calculation Model

The search intention is identified based on the highest score for the given search task. Here, we differentiate between *Exploratory Score* (S_E), *Targeted Score* (S_T) and *Analytical Score* (S_A). Each score is calculated in perspective of the above identified patterns. The calculation is based on the *Query Length Factor* (F_{QL}), *Least Frequent Factor* (F_{LF}), *DBpedia Matching Factor* (F_{DBp}) and *Analytical Factor* (F_A). Furthermore, we define that a *Query* (Q) represents a *set* of *terms or words* $Q = \{t_1, t_2, ..., t_n\}$. We could identify the following score calculations for the mentioned search tasks:

$$S_{\mathrm{E}} = \sqrt{(F_{\mathrm{QL}})^2 + (1 - F_{\mathrm{LF}})^2 + (1 - F_{\mathrm{DBp}})^2 + (1 - F_{\mathrm{A}})^2} \tag{1}$$

$$S_{\mathrm{T}} = \sqrt{(1 - F_{\mathrm{QL}})^2 + (F_{\mathrm{LF}})^2 + (F_{\mathrm{DBp}})^2 + (1 - F_{\mathrm{A}})^2} \tag{2}$$

$$S_{\mathrm{A}} = \sqrt{(1 - F_{\mathrm{QL}})^2 + (F_{\mathrm{LF}})^2 + (F_{\mathrm{DBp}})^2 + (F_{\mathrm{A}})^2} \tag{3}$$

The scores for each kind of task is calculated on a couple of identified factors. These factors will be calculated as follow:

$$F_{\mathrm{QL}} = \frac{2}{|Q|} \qquad \text{for } Q \neq \{\ \} \tag{4.1}$$

$$F_{\mathrm{DBp}} = \max_{0<n<|P(Q)/\{\}|} (f_{DBp}(Q, Q'_n)) \quad \text{for } Q'_n \in P(Q)/\{\} \quad (P() \text{ is power set}) \tag{4.2}$$

$$f_{\mathrm{DBp}}(Q, Q') = \begin{cases} 2 \cdot \dfrac{|Q'|}{|Q|} & \text{If whole or part of query } Q \text{ matches to a DBpedia resource and has class references, } \textit{for } Q' \in P(Q)/\{\} \\ 0.5 & \text{If whole or part of query } Q \text{ matches to a DBpedia resource, but has no class reference} \\ 0 & \text{Else (in particular if there is no match to a DBpedia resource)} \end{cases} \tag{4.2.1}$$

$$F_{\mathrm{A}} = F_{\mathrm{A}_1} + F_{\mathrm{A}_2} + F_{\mathrm{A}_3} \tag{4.3}$$

$$F_{\mathrm{A}_1} = \begin{cases} 0.3 & \text{If } |Q| > 6 \\ 0 & \text{else} \end{cases} \tag{4.3.1}$$

$$F_{\mathrm{A}_2} = \sum_{n=1}^{|Q|} \begin{cases} 0.2 & \text{If } t_n \in \mathrm{Q} \text{ is a number: } t_n \in R \\ 0 & \text{else} \end{cases} \tag{4.3.2}$$

$$F_{\mathrm{A}_3} = \sum_{n=1}^{|Q|} \begin{cases} 0.1 & \text{If } F_{\mathrm{A}_1} > 0 \text{ and } f_{\mathrm{WC}}(t_n) > 10 \text{ for } t_n \in Q \\ 0 & \text{else} \end{cases} \tag{4.3.3}$$

Based on APIs such as *WordSchatz* (for German) or *Microsoft Web N-gram*, (for English) we can define a general function that calculates the word frequency class for each term of the searched query. Our model consists about sixteen word-frequency-classes, where class 0 represents a word or words with the highest occurrence in a particular language, *e.g*. the word 'the' in the English language.

$$f_{\mathrm{WC}}(t) = c \qquad \text{for} \qquad c \in \mathrm{N}_0,\ 0 \le c \le 15 \qquad (4.4.1)$$

- c *… represents the word frequency class of a word, the lower the class the higher the word frequency in a language. The calculation is performed through word-frequency APIs*

This function is underspecified since it depends strongly on the used calculation model or API. In reference to this general function, we can now calculate the *Least Frequent Factor* as follow:

$$F_{\mathrm{LF}} = \frac{(10 - \max_{0<n\le|Q|}(f_{\mathrm{WC}}(t_n))}{5} \qquad \text{for } t_n \in Q \text{ and } n \in N \qquad (4.4)$$

Further details to the calculation model can be found in our previous technical paper [27], where also the identified features and the final evaluation of the model is described.

3.3. Web-Service Design of the Search Intention Analysis

The model follows the distributed computing approach in which the instances of the model are distributed over the cluster of servers which are implemented as RESTful web services. Figure **3** depicts the system architecture of the distributed information system. Each web-service can be accessed over an HTTP call by passing the query to be categorized as a query parameter. The web-service is implemented in Java and it further uses several other third-party services. Every instance is installed with the local MySQL Databases (one for each Non-German language supported by the system) containing the *Wortschatz* frequency information. As of now, the model uses the *Wortschatz* only for English (as a local database) and German languages (as a SOAP service). The architecture shown in Figure **3** also indicates the steps followed by the System.

The implementation is realized as a REST web-service, which enables the integration in a couple of external applications. The request needs to have the format: *http://server.tld/WordFrequencyService/rest/frequency/[search phrase]*

The response will be given in XML (example result output as shown in Figure **5**). As output the most important calculations of the scores for targeted, explorative and analytical search will be given. The individual evaluation of the words which consists of the “frequency information”, “language”, “Ngram value” along with the “Ngram offset”, “*WortSchatz* frequency” and the corresponding “frequency class” will be given as separate “word” tags and all the word tags are grouped under “word-details" tag.

The output of the model is informative enough for any system which depends on it for the query categorization. Therewith the inclusion in existing information search application as REST service makes it easy to advance existing application with the ability to initially show a more appropriating composition of visualization that show the search results.

3.4. SemaVis a System for Visual Information Search, Exploration and Analysis

We named our technology that integrates the conceptual design for adaptive semantics visualization SemaVis. SemaVis was originally designed as a modular framework for semantic visualization, editing and annotation of semantic content [1, 13, 24]. We enhanced SemaVis with respect to an adaptive behavior and focused in particular on visualization and

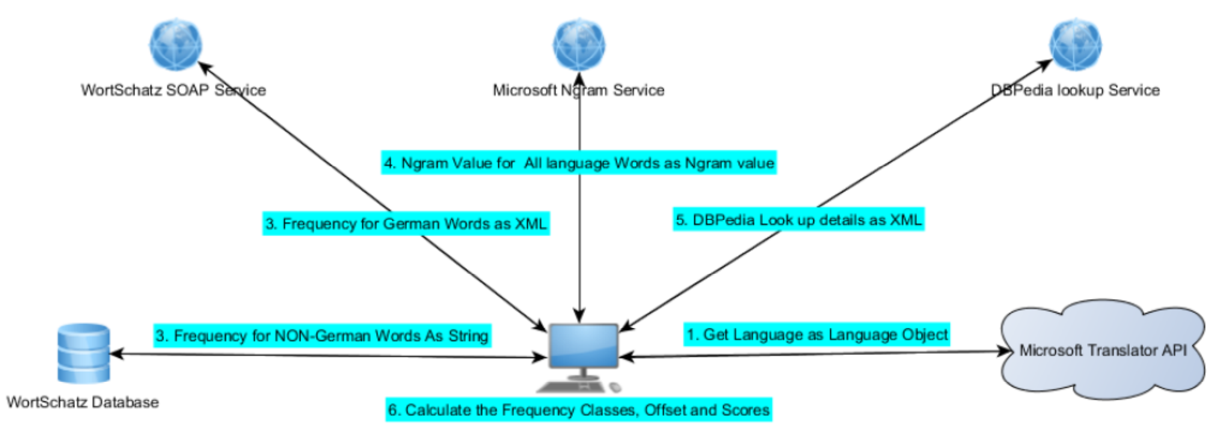

Figure 4: The System Architecture with Steps of Calculation.

```
▼<frequencyDetails>
   <query>mobile phone</query>
   <language>ENGLISH</language>
   <queryType>Explorative</queryType>
 ▼<calculationDetails>
     <minimumNormalizedClass>9</minimumNormalizedClass>
     <explorativeScore>1.7</explorativeScore>
     <targetedScore>1.1357816691600546</targetedScore>
     <analyticalScore>0.5385164807134505</analyticalScore>
   </calculationDetails>
 ▼<word-details>
   ▼<words>
       <text>mobile</text>
       <wordLanguage>ENGLISH</wordLanguage>
       <frequency>144</frequency>
       <wortSchatzFrequencyClass>11</wortSchatzFrequencyClass>
       <ngramValue>-3.486</ngramValue>
       <ngramClass>7</ngramClass>
       <ngramOffset>0</ngramOffset>
       <normalizedClass>9</normalizedClass>
     </words>
   ▼<words>
       <text>phone</text>
       <wordLanguage>ENGLISH</wordLanguage>
       <frequency>481</frequency>
       <wortSchatzFrequencyClass>10</wortSchatzFrequencyClass>
       <ngramValue>-3.514</ngramValue>
       <ngramClass>7</ngramClass>
       <ngramOffset>0</ngramOffset>
       <normalizedClass>9</normalizedClass>
     </words>
   </word-details>
</frequencyDetails>
```


Figure 5: API result in XML Notation of a Performed Query Phrase Analysis for the Query 'mobile phone'.

visual adaptation. SemaVis as a visualization technology enables visualizing various datatypes, adapting to various influencing factors, and provides more functionalities than described in our conceptual model. SemaVis is implemented as a client-server technology, but it can also be used as a client application or compiled as desktop application with limited functionalities.

In different projects we even extended SemaVis to support a number of different data sources, formats and data types, which enable the system to support all of the mentioned search tasks. In particular when statistical data sources such as EuroStat or statistic simulators [28], semantically enriched data sources such as DBpedia, Freebase or Eur-Lex [24], and bibliographic data sources such as EuroGraphics [29] are available for analysis purpose, it is difficult to show an initial set of visualizations that fits to all possible visualization task the user has in mind. But exactly this situation can easily be faced with our search intention analysis service that enables to predict the main intentions based on the first interaction with the system, namely querying a term or a set of terms. Based on the initial analysis the system is able to provide a set of adequate visual layouts [1] and solve the new user problem in adaptive systems.

For the inclusion of the search intention analysis web-service into our SemaVis visualization technology, we defined optimal sets of visualizations for each of the three search task categories. If the user is now performing a search, the optimal set of visualization is shown in dependence of what search task category has the highest score (schematic sketch in Figure **6**). The set of the visualizations are not predefined. We trained them with a statistical probability algorithm, which is described detailed in our previous work [1].

The result is a better search experience since the users will be less confused or frustrated with a more appropriate result of visualized information.

4. EVALUATION

In this section we describe the evaluation of the calculation model. To perform the evaluation, we invited users to commit real performed searches on google, Bing etc. and let them choose if these searches intended to be most likely targeted, exploratory or analytics (the evaluation is described more detailed in [27]). The evaluation was performed separately for each category and can be easily extended to accommodate a new category with few changes. The data from all the separate evaluations was gathered and analyzed to investigate the overall performance and efficiency of the model.

As the data collected from the survey are labeled data, the supervised machine learning evaluation techniques [13] of precision, recall and accuracy were applied. These techniques in turn depend on the True/False Positive and True/False Negative predictions [14] of the model. As these methods are highly recommended for any classification system, they were chosen as the best suited for our model. The precision in the field of information retrieval is given as the ratio of the number of records retrieved, which are relevant to the total number of records retrieved (irrelevant and relevant) [14]. Recall in the field of information retrieval is given as the ratio of the number of records retrieved, which are relevant to the number of relevant records in the repository [14]. In information retrieval, precision is considered as a result relevancy measure and recall as a measure of how many truly relevant results are returned. It is also believed that, with high precision and recall, the system will return many correctly labeled results [15]. With the same set of parameters, another measure accuracy can be derived. In the field of information retrieval accuracy is

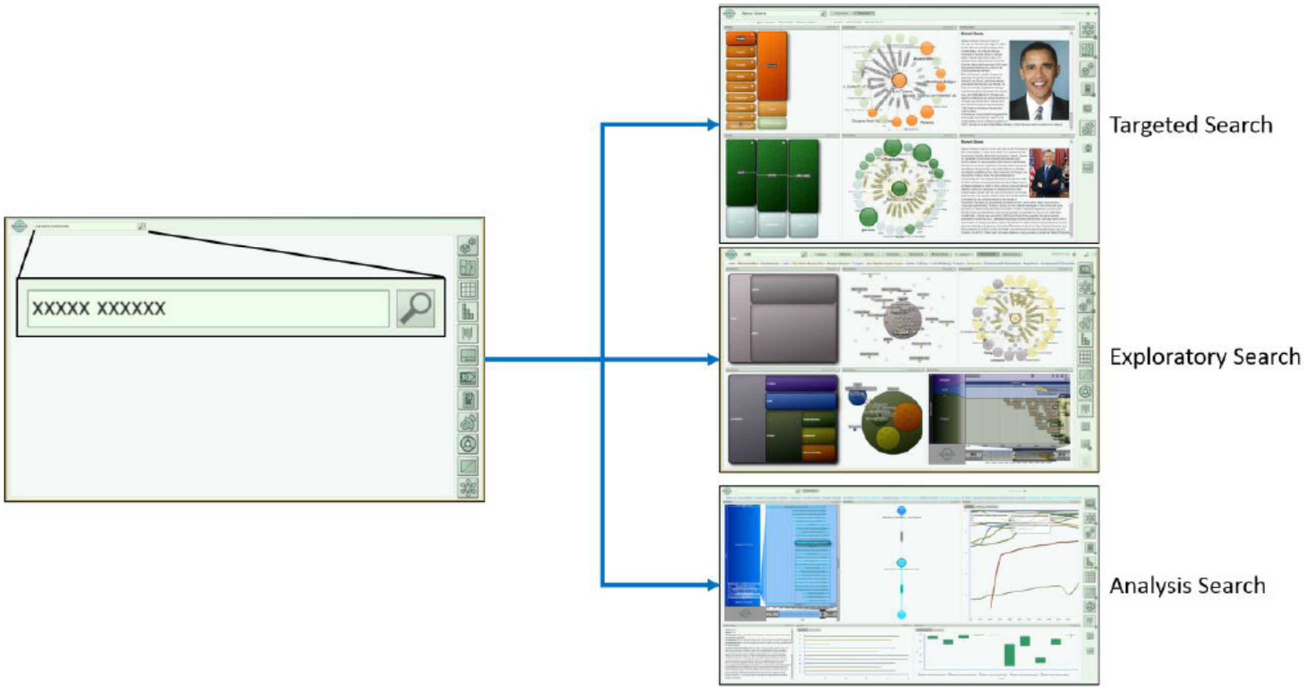

Figure 6: Scenario image that shows how different visualization compositions where arranged in dependence of the identified search task category.

the ratio of true correct retrieval values (both true positives and true negatives) to the total number of cases examined [16].

The table and the barchart representation in Figure **7** shows the precision, recall and accuracy values of the model for the categories "exploratory search", "targeted search" and "analysis search", based on the True/False Positive and True/False Negative values. The results are obtained by providing the model with 100 queries from the survey that was conducted (testing data).

As we can see on the table (Figure **7** left table), the model performs pretty well for "targeted search" and "exploratory search" queries in perspective of precision, recall and accuracy values. As there was less data for the analysis search queries for training and as well as for testing, the results cannot be taken seriously for this category as the model could accommodate only few patterns (hypotheses) for analysis search queries. It is not clear how the model would perform with data sets consisting of many analysis search queries. The major concern being the separation of targeted and exploratory search queries, the results look good, however they can be further improved.

	Precision	Recall	Accuracy
Targeted Query	83.6%	86.44%	82%
Exploratory Query	80.64%	78.12%	87%
Analytical Query	75%	66.66%	95%

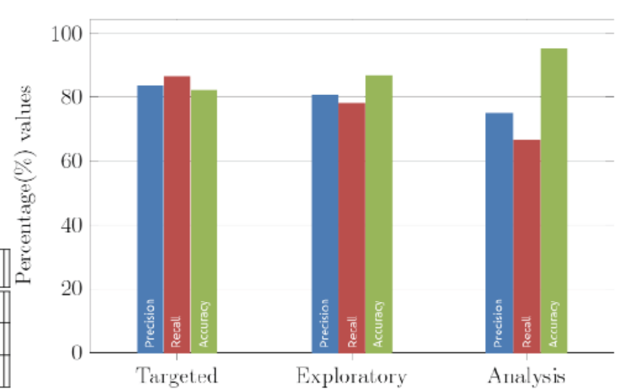

Figure 7: The evaluation shows that the developed system has overall good results to identify the search intention with a high satisfaction quote.

CONCLUSION

In this paper we described a new approach to classify search phrases into three groups of search intentions. As search intention we defined the three search types: (1) targeted, (2) exploratory and (3) analysis search. On the basis of the defined query classification, the system is now able to present an appropriate initial set of visualizations to entered search phrases and therewith support the users directly in his work. We defined the set of visualizations for each of the three visualization task categories manually in respect of the requirements of the corresponding search task and behavior. To proof our concept, we evaluated the classification algorithm with hundreds of real user search queries and their main intentions. The evaluation shows an accuracy of approx. 88%.

In a next step we aim to include the search intention analysis feature to consider in visual trend analysis solution [30] to provide a better support of users in the challenging domain of trend detection in technology and innovation management.

REFERENCES

[1] Nazemi K. Adaptive Semantics Visualization. Springer International Publishing Switzerland 2014.

[2] White RW, Roth RA. Exploratory search: Beyond the Query-response Paradigm. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers 2009.

[3] Marchionini G. Information Seeking in Electronic Environments. New York: Cambridge University Press 1995. https://doi.org/10.1017/CBO9780511626388

[4] Zhou MX, Feiner SK. Visual task characterization for automated visual discourse synthesis. In: CHI '98: Proceedings of the SIGCHI Conference on Human factors in Computing Cystems. ACM Press/Addison-Wesley Publishing Co 1998; p. 392-399. https://doi.org/10.1145/274644.274698

[5] Fluit C, Sabou M, van Harmelen F. Supporting user tasks through visualisation of light-weight ontologies. Handbook on Ontologies. International Handbooks on Information Systems. Springer 2004; p. 415-434.

[6] Fluit C, Sabou M, van Harmelen F. Ontology-based information visualization: Towards semantic web applications. In Visualizing the Semantic Web. XML-Based Internet and Information Visualization. Second Edition. London: Springer-Verlag Limited 2006. https://doi.org/10.1007/1-84628-290-X_3

[7] Shneiderman B. The Eyes Have it: A Task by Data Type Taxonomy for Information Visualizations. In: VL 1996; 1996; p. 336-343. https://doi.org/10.1109/VL.1996.545307

[8] Yi JS, ah Kang Y, Stasko J, Jacko J. Toward a deeper understanding of the role of interaction in information visualization. IEEE Transactions on Visualization and Computer Graphics 2007; p. 1224-1231. https://doi.org/10.1109/TVCG.2007.70515

[9] Pike WA, Stasko J, Chang R, O'Connell TA. The science of interaction. Information Visualization 2009; p. 263-274. https://doi.org/10.1057/ivs.2009.22

[10] Keller PR, Keller MM. Visual cues: Practical data visualization. Los Alamitos: IEEE Computer Society Press; 1994.

[11] Burkhardt D, Breyer M, Nazemi K, Kuijper A. Search intention analysis for user-centered adaptive visualizations. In: Universal Access in Human-Computer Interaction. Design for All and eInclusion. Berlin, Heidelberg: Springer 2011; p. 317-326. https://doi.org/10.1007/978-3-642-21672-5_35

[12] Nazemi K, Breyer M, Burkhardt D, Stab C, Kohlhammer J. SemaVis - A new approach for visualizing semantic information. In Towards the Internet of Services: The Theseus Program. Berlin, Heidelberg: Springer 2014; p. 191-202. https://doi.org/10.1007/978-3-319-06755-1_15

[13] Manning CD, Raghavan P, Schütze H. Introduction to information retrieval. New York: Cambridge University Press 2008. https://doi.org/10.1017/CBO9780511809071

[14] Wikipedia. Precision and recall [homepage on the internet]. wikipedia, the free encyclopedia, 2015. [Online; accessed 18-September-2015]. Available from: https://en.wikipedia.org/wiki/Precision_and_recall

[15] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O *et al.* Scikit-learn: Machine learning in python. Journal of Machine Learning Research 2011; 12: p. 2825-2830.

[16] Metz CE. Basic principles of ROC analysis. In: Seminars in nuclear medicine 1978; 8(4): p. 283-298. https://doi.org/10.1016/S0001-2998(78)80014-2

[17] Jansen BJ, Booth DL, Spink A. Determining the user intent of web search engine queries. In: Proceedings of the 16th International Conference on World Wide Web, WWW '07. New York: ACM 2007; p. 1149-1150. https://doi.org/10.1145/1242572.1242739

[18] Mendoza M, Baeza-Yates R. A web search analysis considering the intention behind queries. In: Web Conference 2008; LA-WEB '08. Latin American 2008; p. 66-74. https://doi.org/10.1109/LA-WEB.2008.9

[19] Broder A. A taxonomy of web search. In: SIGIR Forum 2002; 36(2): p. 3-10. https://doi.org/10.1145/792550.792552

[20] Rose DE, Levinson D. Understanding user goals in web search. In: Proceedings of the 13th International Conference on World Wide Web, WWW '04. New York: ACM 2004; p. 13-19. https://doi.org/10.1145/988672.988675

[21] Hernandez I, Gupta P, Rosso P, Rocha P. A simple model for classifying web queries by user intent. In: Proceedings 2nd Spanish Conference on Information Retrieval (CERI-2012) 2012; p. 235-240.

[22] Wang Z, Zhou X, Yu Z, He Y, Zhang D. Inferring user search intention based on situation analysis of the physical world. In: Proceedings of International Conference on Ubiquitous Intelligence and Computing. Springer 2010; p. 35-51. https://doi.org/10.1007/978-3-642-16355-5_6

[23] Yi Y, Raghavan H, Leggetter C. Discovering users' specific geo intention in web search. In: Proceedings of the 18th International Conference on World Wide Web, WWW '09. New York: ACM 2009; p. 481-490. https://doi.org/10.1145/1526709.1526774

[24] Nazemi K, Burkhardt D, Breyer M, Stab C, Fellner DW. Semantic visualization cockpit: Adaptable composition of semantics-visualization techniques for knowledge-exploration. In: Proceedings of International Conference Interactive Computer Aided Learning 2010; p. 163-173.

[25] Burkhardt D, Nazemi K, Sonntagbauer P, Kohlhammer J. Interactive visualizations in the process of policy modeling. In: Proceedings of Electronic Government and Electronic

Participation. Joint Proceedings of Ongoing Research of IFIP EGOV and IFIP ePart 2013. Bonn: Köllen 2013; p. 104-115.

[26] Buja A, Cook D, Swayne DF. Interactive high-dimensional data visualization. Journal of Computational and Graphical Statistics 1996; 5: p. 78-99.

[27] Burkhardt D, Pattan S, Nazemi K, Kuijper A. Search intention analysis for task- and user-centered visualization in big data applications. In: Procedia Computer Science 2017; 104: p. 539-547. https://doi.org/10.1016/j.procs.2017.01.170

[28] Burkhardt D, Nazemi K, Stab C, Steiger M, Kuijper A, Kohlhammer J. Visual statistics cockpits for information gathering in the policy-making process. In: International Symposium Advances in Visual Computing. Berlin, Heidelberg, New York: Springer 2013; p. 86-97. https://doi.org/10.1007/978-3-642-41939-3_9

[29] Nazemi K, Retz R, Bernard J, Kohlhammer J, Fellner DW. Adaptive semantic visualization for bibliographic entries. In: Advances in Visual Computing. 9th International Symposium, ISVC 2013: Proceedings, Part II. Berlin, Heidelberg, New York: Springer 2013; p. 13-24. https://doi.org/10.1007/978-3-642-41939-3_2

[30] Nazemi K, Retz R, Burkhardt D, Kuijper A, Kohlhammer J, Fellner DW. Visual trend analysis with digital libraries. In: Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business (i-KNOW '15). New York: ACM 2015. https://doi.org/10.1145/2809563.2809569

Received on 28-11-2017 Accepted on 22-12-2017 Published on 26-12-2017

http://dx.doi.org/10.15379/2410-2938.2017.04.02.02