



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

Survey on Interactive Visualization and Navigation of Web Search Results

S.Sandhiya, P.Hemalatha

Dept of Information Technology, IFET College of Engineering, Villupuram, India.

Assistant Professor, Dept of Information Technology, IFET College of Engineering, Villupuram, India

ABSTRACT: It argues for making use of text structure when retrieving from full text documents and presents a visualization paradigm called Tile-Bars. The result lists of popular web search engines represent retrieved documents with a title a brief textual summary and a URL into the conventional search result interface. A set of guidelines for the design of exploratory search interfaces. An thoughtful of the impact of categorized overviews on search tactics will be useful to Web search developers. The image in our model is the solar system with its planets and asteroids revolving around the sun. Position, shade, effort, and three-dimensional distance of objects in the visual space are used to represent the semantic relationships between a query and relevant Web pages.

It deals with a certain session of optimization methods, based on conservativeconvexseparable approximations (CCSA), for solving inequality-constrained nonlinear programming problems.

I. INTRODUCTION

Data mining is that the method of extracting or mining information from great deal of information. clustering is one of the most interesting and important topics in data mining. Data mining consists of five major elements:

- Extract, renovate and load contract records onto the data warehouse system.
- Collection and cope the data in a multidimensional database system.
- Provide data access to business bazaar analyst and information technology authorities.
- Analyze the data by presentation software.
- Current the data in a useful format, such as a grid or stall.

Data mining is that the method of extracting or mining information from great deal of information .It's Associate in analytic method designed to explore giant amounts of information in search of consistent patterns and systematic relationships between variables and to validate the findings by the detected patterns to new subsets of information. It is often viewed as a result of natural evolution in development of clustering is one of the most interesting and important topics in data mining. The aim of clustering is to find intrinsic structures in data, and organize them into meaningful subgroups for further study and analysis. There have been many clustering algorithms published every year. the simple algorithm k-means still remains as one of the top 10 data mining algorithms nowadays. It is the most frequently used partition clustering algorithm in practice. Another recent scientific discussion states that k-means is the favorite algorithm that practitioners in the related fields choose to use. Needless t mention, k-means has more than a few basic drawbacks, such as sensitiveness to initialization and to cluster size, and its performance can be worse than other state-of-the art algorithms in many domains. In spite of that, its simplicity,

II. RELATED WORK

In [1]authorsused the field of information retrieval has traditionally focused on textboxes consisting of titles and abstracts. As a consequence, many fundamental molds must be altered for retrieval from full-length text assemblies. This paper argues for making use of text structure when retrieving from full text documents, and presents a imagining paradigm, called Tile-Bars, that demonstrates the usefulness of explicit term distribution information in Boolean-type queries. Tile Bars simultaneously and compactly indicate relative article length, query term amount, and query term distribution. The patterns in a column of Tile Bars can be quickly scanned and deciphered, aiding users in making judgment DEMERITS:Document content information is difficult to display using existing graphical interface



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

techniques because textual information does not conform to the expectations of sophisticated display paradigms. A problem with applying similarity search to full-length text documents is that the structure of full text is quite different from that of abstracts. Central problem is determining where one set of terms ends and the next begins. The algorithm is domain-independent, and is fully implemented. MERITS: Preliminary results indicate that scores can be improved by taking individual term distribution preferences for individual queries into account. Although the time for retrieval is greater than for a standard Boolean full-text query.

In [2] author present a novel approach that incorporates visualization. The result lists of popular web search engines represent retrieved documents with a title, a brief textual summary and a URL into the conventional search outcome interface. For each resulting document, amounts of the entire query are concisely depicted in the form of a small, document-shaped icon. An participant user study was carried out to compare our design with the traditional search result list in terms of accuracy and commission performance. Though statistically major differences in performance were not observed, the participants' particular ratings and views of the visualization's utility were optimistic. Despite the fact that the visualization introduces a new and somewhat complicated variable to consider when evaluating search result lists, the participant's performance did not fall below their performance level with the traditional interface. DEMERITS: The majority of existing Web search engines present relevant Web pages in a list with titles and short paragraph descriptions extracted from the texts. The existing Web search framework, users' topics of interest or contextual information can only be introduced as a different query. Existing Web search visualization tools do not show the degree of relevance between a resulting page or a group of resulting pages and the query in the context of users' subjects of interest. MERITS: The structural information in Web pages in determining the importance of a keyword, we also implement an adjustment mechanism. Primary advantage of the manual mode (i.e. a subject icon is dragged around via mouse) is that users can view the impact of a moving subject on a local area in the visual space by changing both the moving speed and the direction of a subject icon [3] author Categorized overviews of web search results are a promising way to support user examination, sympathetic and detection. These search interfaces combine a metadata-based overview with the list of search results to enable a rich form of interaction. A study of 24 urbane users carrying out complex tasks suggests how searchers may adapt their search tactics when using categorized overviews. This mixed methods study evaluated branded overviews of web search results organized into thematic, physical, and administration categories. Participants conducted four exploratory searches during a 2-hour session to generate ideas for newspaper articles about specified topics such as "human rustling." Results showed that subjects explored deeper while feeling more organized, and that the considered overview helped subjects better

Assess their results, although no significant differences were detected in the quality of the object philosophies. A qualitative analysis of searcher comments identified seven tactics that participants reported adopting when using categorized overviews. DEMERITS: Initially we did not expect to find the Television category under Arts, and they found this troubling. Longitudinal studies have been used to examine changes in tactics and query terms in relation to changes in searchers' information problem stage. The category labels in the overview indicate which categories results are in, but this may be limited to the top few levels because of the limited display space. MERITS: To mitigate these problems, this study adopted a mixed methods approach. It shows how a combination of qualitative and quantitative methods can address research questions related to exploratory search. Benefits of categorized overviews for many search tasks and users seem strong enough to warrant further research, refined designs, and more commercial implementations. Useful for digital library and web search designers, information architects, and web developers because they provide guidance for the appropriate integration of visual overviews with search result lists. [4] It deals with a certain class of optimization methods, based on conservative convex separable approximations (CCSA), for unraveling in equality-constrained nonlinear programming problems. Each generated iteration point is a feasible solution with lower objective value than the previous one, and it is proved that the arrangement of iteration points converges toward the set of Karush-Kuhn-Tucker points. A major advantage of CCSA methods is that they can be applied to problems with a very large number of variables (say 104-105) even if the Sacking matrices of the objective and restraint functions are dense. DEMERITS: There are outer and inner iterations in the methods. An outer iteration starts from the current iterate $x(k)$ and ends up with a new iterate $x(k+1)$. The original objective and constraint functions are replaced by certain convex separable functions which approximate the original functions around $x(k)$. These inner iterations are repeated until the approximating objective and constraint functions become greater than or equal to the original functions at the optimal solution of the subproblem. MERITS: Benefit of CCSA methods is that they can be successfully applied to problems with a very large number of variables, if the Hessian matrices of the



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

objective and constraint functions are dense. This property is to a large extent due to the usage of separable approximations.

V. CONCLUSION AND FUTURE WORK

We introduced Project-Snippet, a novel technique to visualize the collection of textual snippets returned from a web query. The method builds intuitive and meaningful layouts that optimize the placement of snippets by employing an innovative energy functional that considers both overlapping removal and preservation of neighborhood structures. We showed results illustrating how the Project-Snippet layouts convey a global view of the results from a query while allowing for identifying similar content through a clustering mechanism. Since Project-Snippet relies only on information extracted from the textual snippets, it can be plugged into search engines in a straightforward manner, with a modest impact on the computational times. The unique combination of simplicity, low computational cost, and flexibility renders Project-Snippet an attractive alternative for visualizing web queries results. We are currently investigating interactive mechanisms to enable a free navigation in the snippet-based layout as well as on how to modify the energy functional to improve the layout so that it better highlights density information and similarity between neighboring snippets. Visualization techniques to support Queries searching can be split into two major groups: methods for visualizing free text queries, which provide visual tools to assist the querying process, and techniques for visualizing the outcome of a particular query.

We focus our discussion on the latter group to contextualize the technique proposed in this paper. Specifically, we discuss methods tailored to provide visual representations of the results of a textual search process, disregarding approaches aimed at visualizing document collections in general. Albeit potentially applicable in the scenario considered here, they typically neglect the specificities of this kind of application. A comprehensive survey that addresses aspects not covered here may be found in the works by Yao et al. The book by Hearst also surveys contributions on visual interfaces to support general search tasks. Techniques for visualizing textual search results may be organized based on their underlying visualization paradigm. Data mining involves many tasks such as Association, Decision Support system, Pattern Recognition, Classification, Clustering etc in order to obtain useful information from large amount of data Clustering has been used in a number of applications such as engineering, biology, medicine and data mining. One of the most widely used algorithms is K-Means clustering. It partitions the objects into clusters by minimizing the sum of the squared distances between the objects and the centroid of the clusters. The K-Means clustering is simple but it has high time complexity, so it is not suitable for large data set. The algorithm allows early termination of the distance calculation by introducing a premature exit condition in the search process.

REFERENCES

- [1] M.A. Hearst, "TileBars: Visualization of Term Distribution Information in Full Text Information Access," Proc. ACM SIGCHI Conf. Human Factors in Computing Systems, pp. 59-66, 1995.
- [2] T. Heimonen and N. Jhaveri, "Visualizing Query Occurrence in Search Result Lists," Proc. Ninth Int'l Conf. Information Visualisation, pp. 877-882, 2005.
- [3] T. Nguyen and J. Zhang, "A Novel Visualization Model for Web Search Results," IEEE Trans. Visualization and Computer Graphics, vol. 12, no. 5, pp. 981-988, Sept./Oct. 2006.
- [4] K. Svanberg, "A Class of Globally Convergent Optimization Methods Based on Conservative Convex Separable Approximations," SIAM J. Optimization, vol. 12, no. 2, pp. 555-573,
- [5] A. Spoerri, "RankSpiral: Toward Enhancing Search Results Visualization," Proc. IEEE Symp. Information Visualization, pp. 208-214, 2004.
- [6] M. Nizamee and M. Shojib, "Visualizing the Web Search Results with Web Search Visualization Using Scatter Plot," Proc. IEEE Second Symp. Web Soc., pp. 5-10, 2010.
- [7] K. Svanberg, "A Class of Globally Convergent Optimization Methods Based on Conservative Convex Separable Approximations," SIAM J. Optimization, vol. 12, no. 2, pp. 555-573,



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

BIOGRAPHY



S.SANDHIYA currently pursuing INFORMATION TECHNOLOGY in IFET COLLEGE OF ENGINEERING, VILLUPURAM, India. Her area of interest is ooad concepts, web design



AssitantProfesor currently worked in IFET COLLEGE OF ENGINEERING, VILLUPURAM, India. she received her B.E CSE P.S.N.A college of engineering and technology pondicherry and M.E in CSE from IFET COLLEGE OF ENGINEERING, VILLUPURAM .She has published three paper in International Conference level and also Journal paper. Her area of interests include Computer Networks, Wireless Sensor Network and Data Mining.