INDEXED SEARCH USING SEMANTIC ASSOCIATION GRAPH

Kiran P[1], Shreyasi A N[2]
[1] *Associate Professor, Department of CSE, RNS Institute of Technology, Bengaluru, Karnataka, India.*
[2] *PG Scholar, Department of CSE, RNS Institute of Technology, Bengaluru, Karnataka, India.*

ABSTRACT

This project would help users when they try to explore unfamiliar fields or when they want to learn more about a particular subject area or to increase their knowledge rather than to restrict themselves to a specific problem. Indexed search using semantic association graph method will provide browse guidance. It allows users to discover new associations and new knowledge, and helps them to find the information and or data they are interested in. This method defines a new set of criteria containing clarity, relevance, novelty and diversity to analyse the effectiveness of an exploratory search. It is more suitable for learning new associations and discovering new knowledge with high likely relevance to a query using Query topic and Semantic association graph. This work is more suitable than many other search engines for an exploratory search.

Keywords: *Exploratory search, indexing network, relevance information search, search expansion*

1. INTRODUCTION

INTERNET data are heterogeneous, redundant, disordered, and exponentially growing. Since the 1990s, Internet search has focused on finding specific webpages that best meet user information requirements. Most search engines crawl Web and index webpage, matching worldwide "best" candidate webpage with excellent precision. Such ability has been proven adequate to full fill the majority of user needs. However, a matching query style has significant limitations. Search results are satisfactory only when users give the right search words. When users cannot specify accurate search words, the search results are often useless. Further, as information becomes more ubiquitous and demands for various searches grow, there is an increasing need to support search behaviours beyond simple lookup. A matching query style is not suitable to investigate information in unfamiliar field and to learn new associations and knowledge with relevance to a query.

Exploratory search describes an information-seeking problem context that is open-ended, persistent, multi-faceted and an information-seeking process that is opportunistic, iterative, and multi tactical. The existing studies about it mainly focus on its implementation modes, evaluation methods, and how to analyse the information about a topic. By using topics as a basic unit, Project analyses the semantic associations between a topic and other topics on the Internet. Based on those semantic associations, this project proposes a topic oriented exploratory search, which can provide effective browse guidance for users and display new information that is likely relevant to their queries. It allows users to discover new associations and knowledge, and help them find their interested information.

To provide the desired browse guidance, this project designs an interactive search expansion mode, which expands searches in two directions: 1) other keywords in the same topic and 2) other topics related to the browsing topic. For example, users enter "Ming Yao" (a former basketball player's name). By interacting with users, the system knows that they want to browse information belonging to topic "basketball." Figure 1.1 shows navigation mode. The browsing topic basketball is represented by a grid circle. Keywords in the browsing topic, e.g., NBA, basketball match, sports, Ming Yao, and the rockets, are all shown at the bottom. Topics related to the browsing topic, e.g., new media and sport image, are shown with different-size circles where a larger circle implies that the topic contains more webpages. Thicker lines between two circles mean a closer relation between their corresponding topics.

To achieve the proposed search expansion, our system must understand semantic associations among topics. Hence, some kinds of topic association structures are needed. This project presents a method to generate such structures based on hyperlinks on the Internet. As we all know, webpages on the Internet are linked by hyperlinks. A hyperlink between two webpages usually implies some real-life association. For example, there are some hyperlinks between cooking webpages and ingredient sale webpages. According to them, one can identify some associations between cooking and ingredient sale. We target at dealing with this kind of associations and present a semantic association graph of topics, there by greatly helping users' topic-oriented exploratory search.

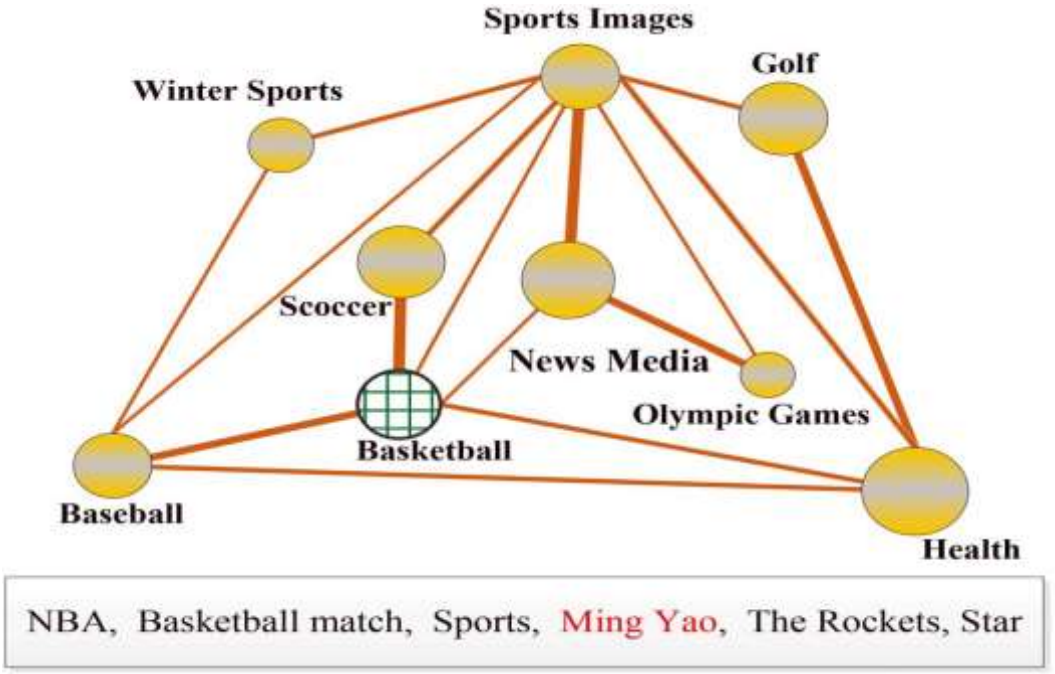

Fig-1: Illustrative result of exploratory search

The contributions of this project can be summarized as follows.
1) Proposing a method to build a semantic association graph based on hyperlinks on the Internet.
2) Giving a method to suggest several candidate topics based on search keywords. By interacting with users, the system determines their browsing topics.
3) Designing an interactive search expansion mode and expanding searches in two directions to provide good browse guidance for users. Different from common search expansions used for finding accurate webpages to queries, the proposed one helps users find their interested information, thereby giving them some closely related but disjoint keywords. It can better support the exploratory search.
4) Defining some new metrics that can be well used to evaluate whether a kind of search expansion is helpful for an exploratory search.
5) Conducting several experiments to compare the proposed method with many search expansion and a co-occurrence frequency method. The experimental results validate the advantage of the proposed method in performing high-quality exploratory search.

2. RELATED WORK

[1] A. Bozzon, M. Brambilla, S. Ceri, and P. Fraternali et al., proposed Liquid Query, a paradigm that exploits the power of underlying search services and provides the user with a multi-domain exploratory search environment. Liquid Query combines search interfaces and data visualization facilities to improve the user experience and the productivity in information seeking. The approach helps users in performing exploratory search tasks, without the need of different online tools for composing heterogeneous information. Future work the integration with the Search Computing engine, currently under development, will enable the Liquid Query interface to exploit sophisticated top-k query optimization and execution algorithms, currently under implementation in the platform.

[2] R. W. White, B. Kules, S. M. Drucker, and M. C. Schraefel et al., Proposed a Supporting exploratory search is an exciting multidisciplinary area that will have a profound effect on how information is gathered, used, and shared. Rather than just providing search results, search systems should help users explore, overcome uncertainty, and learn. To accomplish this, researchers and practitioners must leverage their skills and experience to develop search systems that actively engage searchers by using semantics, inherent structure, and meaningful categorization to organize intuitive visual workspaces.

[3] T. Ruotsalo*et al.*, Paper has contributed a novel approach for interactive exploratory search. The paper demonstrated that interactive user modelling allows the user to control their exploratory search in an intuitive way and the user studies show that users can readily adopt this interaction to partially replace query typing as the input mechanism. Most importantly, above adoption leads to significantly improved retrieval and task performance. Future improvements, the effect of different types of tasks and user modelling durations, such as long-term modelling that goes beyond individual search sessions should be investigated. Also the role of other personalization dimensions, such as difficulty of the retrieved contents and different levels of user pre-knowledge could be interesting to gain more insight beyond topical customization.

[13] S. Wei *et al.*, presented a multi-aspect oriented query summarization task, which aims to summarize a query from multiple aspects aligned to user intents. Specifically, the paper proposed a composite query based strategy, where a set of component queries are used as data sources for the original query. The strategy differs from traditional search result organization and text summarization, where the set of documents to be deal with is seen as a given system input.

3. IMPLEMENTATION

3.1 QUERY TOPIC

A user browsing topic is called a query topic. A topic represents a collection of similar webpages around the same problem. This project views a webpage category as a topic. We use a topic as a basic unit of semantic association graphs for two reasons. First, information on the Internet is redundant and unordered. Many similar webpage exist there webpage classification we can gather them together to benefit to information analysis. Second, hyperlinks are created by developers or administrators of websites and webpages hyperlink noises exist on the Internet. It uses classification and statistics to lessen their influence. Setting proper threshold can help minimize their unfavourable impact on establishing associations among topics.

3.2 SEMANTIC ASSOCIATION GRAPH

Webpage domain contains all webpages in a semantic association graph. A webpage category (or a topic) is a subset of a webpage domain. A topic is represented by several keywords appearing in the Google's artificial directory. And, feature words of a topic are keywords but they are relatively more representative ones. An indexing network is a webpage organization model. It defines semantic relationships among keywords, webpages and categories, and generates a semantic association graph. The exploratory search method presented depends on a semantic association graph.

3.3 POPULARITY OF TOPICS

When users surf on the Internet, they browse the next webpage via a hyperlink. Their behaviours that switch among topics are random. Association degrees among topics decide the probability that users switch from one topic to another. The idea is inspired by a classic ranking algorithm PageRank. But we use it to look for popular topics, which are publically concerned ones that have semantic associations with many topics. Different from PageRank, use transition probabilities among topic.

3.4 SYSTEM ARCHITECTURE:

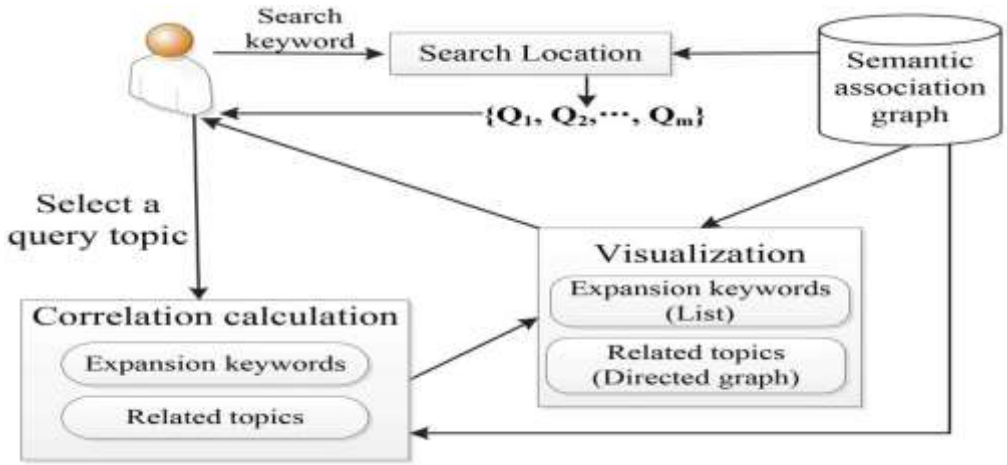

Fig-3.1 System Architecture

User will give a search topic by means of query, it will be processed in search location and search location will suggest m number of queries and also corresponding semantic association graph is drawn for that particular query.

When user selects one query among m queries correlation calculation is done by giving expansion keyword and related topics, with the help of visualization in which list of expansion keywords and related topics are present. For each selection of keyword or related topic semantic association graph will be extended till user gets the needed information.

4. CONCLUSION

This project will develop the "search engine" with added advantages of semantic association graph based on hyperlinks present on the Internet. Indexed search using semantic association graph will suggest several candidate topics based on search keywords and by interacting with users the system determines their browsing topics. Designing an interactive search expansion mode and expanding searches in two directions to provide good browse guidance for users. Different from common search expansions used for finding accurate webpages to queries, the proposed one helps users to find their interested information, thereby giving them some closely related but disjoint keywords. It can better support the exploratory search to discover new associations and new knowledge, and helps users to find the information and data they are interested in.

5. ACKNOWLEDGEMENT

It is my privilege to acknowledge with deep sense of gratitude towards project guide, Dr. Kiran P, for his valuable suggestions and guidance throughout course of study and timely help in completion of the preliminary project work on "Detection of unknown and identical users across social media network". I would also like to thank project co-ordinator Dr. S Satish Kumar, and all other faculty members of Computer Engineering department who directly or indirectly kept the enthusiasm and momentum required to keep the work done. I hereby extend my thanks to all concerned person who co-operated with me in this regard.

6. REFERENCES

[1] A. Bozzon, M. Brambilla, S. Ceri, and P. Fraternali, "Liquid query: Multi-domain exploratory search on the Web," in *Proc. 19th Int. WorldWide Web Conf.*, Raleigh, NC, USA, 2010, pp. 161–170.
[2] R. W. White, B. Kules, S. M. Drucker, and M. C. Schraefel, "Supporting exploratory search," Commun. ACM, vol. 49, no. 4, pp. 37–39, Apr. 2006.
[3] T. Ruotsalo et al., "Supporting exploratory search tasks with interactive user modeling," in Proc. Amer. Soc. Inf. Sci. Technol., Montreal, QC, USA, 2013, pp. 1–10.
[4] Blucora, Inc. (2010, Jul.). *Dogpile*[Online]. Available: http://www.dogpile.com/

[5] InfoSpace LLC. (2014, Nov.).*Infospace*[Online]. Available: http://infospace.com/
[6] CUIL. (2008, Jul.).*CUIL* [Online]. Available: http://www.cuil.pt/
[7] M. T. Mills and N. G. Bourbakis, "Graph-based methods for natural language processing and understanding—A survey and analysis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 1, pp. 59–71.
[8] M. F. Costabile, D. Fogli, P. Mussio, and A. Piccinno, "Visual interactive systems for end-user development: A model-based design methodology," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 37, no. 6, pp. 1029–1046, Nov. 2007.
[9] F. Mohamethet al., "User centered and ontology based information retrieval system for life sciences," BMC Bioinformatics, vol. 13(Suppl1):S4, pp. 1–12, Jan. 2012.
[10] X. G. Qi and B. D. Davison, "Web page classification: Features and algorithms," ACM Comput. Surv., vol. 41, no. 2, pp. 1–31, Feb. 2009.
[11] L. Y. Lin. (2012, Dec.). IK Analyzer 2012FF_hf1_source.rar [Online].Available: https://code.google.com/p/ik-analyzer/downloads/list
[12] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz, "Analysis of a very large Web search engine query log," in Proc. 22nd Conf. ACMSpec. Interest Group Inf. Retrieval, Berkeley, CA, USA, 1999, pp. 6–12.
[13] S. Wei et al., "Multi-aspect query summarization by composite query," in Proc. 35th Conf. ACM Spec. Interest Group Inf. Retrieval, Portland, OR, USA, 2012, pp. 325–334.