

The Effects of Topic Map Components on Serendipitous Information Retrieval

Sunmin Won¹ and Sam Gyun Oh²

¹Sungkyunkwan University, Department of Library & Information Science
won.sunmin@gmail.com

²Sungkyunkwan University, Professor, Department of Library & Information Science.
samoh@skku.edu

Abstract. Some recent studies on information retrieval and information seeking have examined the utility of serendipitous discovery. This research argues that serendipitous discovery has a positive impact on information retrieval and can happen most frequently in semantic web built on the framework of topic maps. This paper discusses the components of topic map that influence serendipitous discovery as well as the elements of topic map designs that may enhance serendipitous discovery. To that end the results of a study on the effects of serendipitous discovery in topic-map-based ontology systems are discussed in the context of information seeking.

1 Introduction

Fast and easy access to massive and various resources are now available to information users. Yet finding appropriate information that meets users' needs has become an increasingly difficult task. While the challenges multiply with polysemous words and hard-to-define needs of information users, current web technologies primarily support keyword-based searches.

The semantic web environment seeks to enable machines to understand and process information, allowing users to better obtain relevant information. One of the latest concerns among researchers is the utility of serendipity in information retrieval (IR) and information seeking. Russell(2007) defines serendipity as fortunate discoveries made by accidents and as a scientific procedure[1]. With

well-organized data the potential for web encounters of a vast number of interesting materials increases.

The topic map, which is one of the frameworks for building the semantic web, provides capability to connect all semantically-related information. Therefore it enables human judgment and decision in relation to relevant information, and enhances serendipity through systematic facilitation of users' chance encounters with information.

The current paper lists suggestions for the design of the topic map that facilitates serendipitous information retrieval. To that end, data from a qualitative study in the context of serendipity which examined the effects of topic map-based ontology system on the information needs of information seekers was analyzed. The study consisted of observations and interviews of 10 college students.

2 Serendipity and Information Retrieval

2.1 The Concept of Serendipity in IR

Serendipity is defined as a situation in which one stumbles upon appropriate information. While information acquisition commonly involves the use of a search/query mechanism or browsing/scanning, it is believed that the latter is more inductive to serendipity.

There are a number of conceptual studies on serendipity. Erdelez (1997), for instance, identified a "serendipity-prone" group among some academic library users who reported "the pressure of the abundance of information waiting to be encountered". Ross(1999) interviewed 194 committed readers and found that these readers were "finding without seeking", gaining valuable insights from their pleasure reading of materials that interested them. Williamson(1998) also discovered incidental information acquisitions among 202 Australian senior citizens from their phone conversations. These studies support that there exists certain dynamics between serendipitous discovery and individuals [3].

2.2 The Usefulness of Serendipity in IR

What makes serendipity useful in IR is its implications for opportunities to find information that is either potentially valuable or capable of generating new knowledge. Some research has found serendipitous discovery as a common

experience among information users and as one that assists their information seeking tasks.

Toms(2000) reported an experiment in which the effects of unintended information retrieval were studied using a specially developed information retrieval application. Results showed that the information obtained through serendipitous discovery was thought to be more meaningful than the information obtained through keyword searching method. Furthermore, information seekers were more involved with retrieval that was inductive to serendipitous discovery. Allen E. Foster and Nigel Ford(2003)'s qualitative study of serendipitous discovery among inter-disciplinary researchers revealed that such discovery was broadly experienced and was likely to provide information relevant to their research process. They also found that certain attitudes and strategic decisions were perceived to be effective in exploiting serendipity when it occurred.

3 Serendipity in Topic Map

3.1 Overview of Current Research

The initial purpose of this study was to analyze the impact that serendipity, which is caused by a topic-map-based ontology system, brings to the clarification of users' information needs. For this purpose a topic-map-based information retrieval system was developed using the ontopoly editor and was given the name TIRS(Topic Map Information Retrieval System). TIRS provided information on every nation of the world as well as on the top 100 global corporations. It had 1,372 topics, 3,924 associations and 4,865 occurrences. Ten participants searched information with TIRS in a largely uncontrolled environment. The only constraint the participants had to follow was that they employ only browsing and scanning as their search strategy. Data was collected by qualitative methods of observations, interviews and a 'Think Aloud' technique which recorded the participants' thoughts during the entire search process. Then unitization and categorization of the data was done, followed by data analyses.

3.2. Findings

The data from the study revealed the following:

- a) The participants experienced serendipitous discovery during information retrieval and were able to locate relevant information with ease using the TIRS.

- b) The participants either adopted the serendipitous discovery of information or utilized it to modify their information needs and retrieval strategies. Modifications included more specific information needs and wording as well as the addition of words and narrowing down of retrieval ranges.
- c) Topic-map-induced serendipity was an element that influenced the changes in the types of information that the participants felt that they needed.

4. TAO and Serendipity

The present study examined how the information seekers utilized the components of the topic map. It also identified the components that induced most serendipitous discoveries.

4.1. Use of the Topic Map Components

As is summarized in Table 1., the component that yielded the highest percentage of usage(40.8%) was the "topic type". The participants employed this component primarily to access individual information (29.5% of all cited purposes), but relied on it as well when checking and searching the information that belonged to super-class categories.

Next in usage was the "association type", which yielded 32.0% of the frequencies. The participants employed this component primarily to check information by its associative relationships (16.3% of all cited purposes), and to identify information relevant to their known needs rather than to search for one that might generate needs.

Lastly the "occurrence type", which accounted for 27.3% of the usage, was most applicable when the participants wanted to access detailed information; 11.8% of all cited purposes fell into this category. The "occurrence type" was also useful when searching for the value of a specific item or checking the overall content of a specific item. Incidentally this component was cited as particularly helpful when checking ranks.

Nevertheless it is important to note that during the face-to-face interviews administered at completion of retrieval sessions six participants cited the "association type" as the most frequently used component, while only one participant mentioned the "topic type" as such. An explanation for the discrepancies may lie in the degrees of actual satisfaction: The information

obtained or discovered by association type may have been far more effective in meeting the information needs of these participants.

Component name	Purpose of using	The Number of times	Rate	Total	Rate
Topic Type	- Accessing to individual information	118	29.5%	163	40.8%
	- Holistic searching in super ordinate category	45	11.3%		
Association Type	- Related information encountering	61	15.3%	128	32.0%
	- Encountering information what generate new information need	2	0.5%		
	- Browsing organized information what relation based	65	16.3%		
Occurrence Type	- Holistic searching in super ordinate category	37	9.3%	109	27.3%
	- Related information encountering	19	48.0%		
	- Rank checking	6	1.5%		
	- Specific information checking	47	11.8%		
Total		400	100%	400	100%

Table 1. Rate of using of Topic Map component during information seeking behavior

4.2. Topic Map Components Inducive to Serendipitous Discoveries

Valuable insight into what components may best enhance serendipitous discoveries can be gained from the data from face-to-face interviews. In order to identify these components, cases where the participants recognized serendipitous discoveries either instantly or after they had initiated modifications on their information needs were carefully followed up. Results confirmed that the "association type", which produced 17 cases of serendipitous discoveries, had been the most effective component. Interestingly the component "topic type"

produced no more than a single reported case, while "occurrence type" was cited in 12 cases. Related statistics are summarized in Table 2.

Component name	The number of time what inducing serendipitous situation	Rate
Topic Type	1	3.3%
Association Type	17	56.7%
Occurrence Type	12	40.0%
Total	30	100%

Table 2. Rate of Topic Map component what inducing serendipitous discovery situation

Despite the fact that it accounted for 40.8% of the usage, only 3.3% of the serendipitous discoveries was linked to the "topic type". In comparison, the components "association type" and "occurrence type" corresponded to 56.7% and 40.0% of the serendipitous discoveries, respectively. In all, serendipitous discoveries were largely unpredicted by the mere frequencies of the usage of the components; what seems more crucial is the degrees in which the participants' information needs had been met by these components.

As indicated in Table 1, the participants used the components mostly to browse information of relationship-based presentations in the "association type", and to check specific information in the "occurrence type". That is, in terms of serendipitous discoveries, information organized by relationships and the discovery of specific information seemed to have been the effective inducer for each type. Results from further interviews with the participants confirmed this speculation. They noted that the strength of the "association type" lies in its grouping of information by relationships and its exhaustive displays of related information for any given selections. In the "occurrence type" they valued the presentation of information in descending/ascending orders as well as the inclusion of specific values and substance as these features enabled them to compare different pieces of information.

5 Suggestions from Participants

During the course of the in-depth interviews the participants discussed the strengths of the topic map system and expressed ideas that might enhance serendipity.

5.1 Association Type as Component: Strengths and Desired Improvements

A. Utility of Relationships

The participants reported that the association type proved convenient in that it allowed searches, in a chain fashion, of directly related information. High level of relatedness among the pieces of information also meant fewer efforts to filter out unnecessary information on their part.

B. Reduction in Comparisons between Information

Information was arranged using diverse relationships among different topic types, making it possible for the participants to peruse it at one glance. The participants were of the opinion that their searches had been made more economical and less cumbersome by these arrangements. However, they also requested that the singular criterion of descending/ascending order of arrangements be expanded upon. They said, for instance, that different pieces of information could be displayed according to their relatedness to another piece of information.

C. Semantic Recognition

The participants judged that the association type facilitated their searches as it was expressive enough of the contents and the types of relationships in the information. Such semantic representations gave them the necessary first clues. With the relatively easy guessing work on what each piece of the information was about, the participants successfully managed to narrow down their searches.

On the other hand, the participants also noted that ambiguous expressions only served as deterrents since these expressions caused them to guess wrong. Consequently, they suggested that the association type names be more explanatory and analogical in relation to the contents and relationships of the information, .

5.2 Occurrence Type as Component: Strengths and Desired Improvements

A. Presentation of Details

The participants agreed that the at-a-glance display of numerical values and specifics on contents is the strength of the occurrence type. They also appreciated the fact that direct links to the topic types were a part of this display. These features of the occurrence type facilitated searches for information with concrete

data, and at times provided clues for narrowing down the categories of searches. Yet the participants mentioned that grouping numerical values and specifics on contents according to the relatedness among different pieces of information would be a welcome option to improve the functionality of the occurrence type.

6 Suggestions for Topic Map Designs to Enhance Serendipity

On the basis of the quantitative data and the results from the interviews some considerations for topic map designs for better serendipity follow.

6.1. Organization of Information

Classification by relationships, which renders relative ease in locating relevant information, was the element that made topic map based IRS convenient. Maximizing the use of the filtering function of the topic map is crucial in the expression of the distinct relationships among different pieces of information. Presenting these pieces of information under similar or related categories is also necessary.

"Association" and "association type" provide the most basic filtering function of the topic map, while "hierarchy", which represents up and down relationships, is most suitable for displays of information by similar or related categories. "Scope" and "role type", which are not directly shown to users yet serve as the cornerstone for limiting ranges and clarifying relationships among information, are two other components that must be fully utilized.

6.2. Semantic Labeling

Explicitness in the expressions of the relationships between different pieces of information is a key element. That is, labeling of the association should be made sufficiently explanatory of the relationships between topics. If the related topics are represented in the label itself, it would increase the efficiency and effectiveness of the association type.

The data from the present study were drawn from undergraduate students in Korea using the Korean language as the medium of communication. Consequently the implications of the study may not be applicable to all populations and languages. Much more linguistic research is required to clarify issues in labeling.

Nevertheless certain aspects of labeling that emerged from the present study do deserve a close attention. In general, when the understanding of the topic map is less than complete, labeling in and by itself does not sufficiently express the relationships between topics. Furthermore, one label cannot represent all relationships: One must employ different grammar if, for instance, the position of the subject and the predicate is reversed.

6.3. Organization of Displays

The participants agreed that the most prominent strength of the topic map based IRS is its capacity to provide information that has been arranged by relationships. Organized information meant better access to relevant information. Yet the participants also noted that the IRS does not go beyond the level of showing the relationships between two directly involved topics. A desirable enhancement, they said, would be the addition of a second-level organization by which other related information is displayed as well. Even though such expressions exceed the basic functions of the topic map, they are still feasible in its construction with query languages such as TOLOG or TMQL.

7. Conclusion

The participants of the present study highly valued the convenience and utility of the topic map. The majority of these participants experienced serendipitous discoveries and judged them to be an integral of information retrieval.

The topic type was the most-used, yet the least-productive of all topic map components as it was connected to but a single case of serendipity. The participants used this component primarily to access individual information. There may be various features that trigger serendipitous discoveries, but the data from the present study suggest that relatedness among information is an extremely important feature for such discoveries. This premise was supported by the fact that most participants reported serendipitous discoveries with the association type, and that their use of the association type increased more toward the end of their information retrieval sessions.

Effective and efficient serendipitous discoveries require maximum utilization of relatedness. This necessitates better organizations of information and enhanced semantic expressions as well as technical applications of TMQL or TOLOG. In all, what is fundamental to the design of topic map based IRS seems to be the crucial role that the association type may play for serendipity. The association

type and the association name should be explanatory and analogical if one is to reflect the well-defined and various relationships between the pieces of information (i.e. the topic types in some respect).

The present study produced a small quantity of data involving a small number of subjects. However, within the context of this experiment it revealed that topic map is inductive to serendipitous discoveries, and that these discoveries facilitated the participants' information retrieval, especially in relation to the clarification of their information needs.

References

- Allen Foster & Nigel Ford. 2003. Serendipity And Information Seeking: An Empirical Study. *Journal of Documentation* 59(3) : 321-340.
- A. Green. 1990. What do we mean by User Needs. *British Journal of Academic Librarianship* 5 : 65-78.
- Brenda Dervin. December 1998. Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of Knowledge Management* 2(2) : 36-46
- Catherine C. Marshall & William Jones. January 2006. Keeping Encountered Information. *Communications of the ACM* 49(1) : 66-67.
- E. Toms. 2000. Serendipitous Information Retrieval. Proceedings of the First DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries, Zurich, Switzerland European Research Consortium for Informatics and Mathematics. [online] http://www.ercim.org/publication/ws-proceedings/DelNoe01/3_Toms.pdf
- Graham Moore. December 2000. Topic Map technology - the state of the art. XML 2000 Conference & Exposition, Washington, USA.
- Jürgen Beier. 2001. Navigation and interaction in medical knowledge spaces using topic maps. *International Congress Series* 1230 : 384-388.
- S. Erdelez. 1999. Information encountering: it's more than just bumping into information. *Bulletin of the American Society for Information Science* 25(3): 25-29.
- S. Pepper. 2002. The tao of topic maps : Finding the Way in the Age of Infoglut. *Ontopia*. [online]. <http://www.ontopia.net/topicmaps/materials/tao.html>
- S. Pepper, G. Moore. 2001. XML topic maps (XTM) 1.0. [online] <http://www.topicmaps.org/xtm/1.0/>