Serendipitous Information Retrieval

Elaine G. Toms
Faculty of Information Studies
University of Toronto
140 St. George St.
Toronto, Ontario, Canada M5S 1A1
Voice: (416) 978-7802
Fax: (416) 971-1399
E-Mail: toms@fis.utoronto.ca

Abstract

The acquisition of information is generally thought to be deliberately sought using a search or query mechanism or by browsing or scanning an information space. People, however, find information without seeking it through accidental, incidental or serendipitous discoveries, often in combination with other information acquisition episodes. The value of this phenomenon to an individual or an organization can be equated with the impact of serendipitous breakthroughs in science and medicine. Although largely ignored in information systems development and research, serendipitous retrieval complements querying and browsing, and together they provide a holistic, ecological approach to information acquisition and define the key approaches to a digital library. In this paper, the concept of serendipitous information retrieval is introduced and validated with data from a study of news readers, along with some approaches for how to facilitate it.

Introduction

Information retrieval in both traditional information systems and in digital libraries is based on the premise that users know or partially know something about the information they seek. Yet significant evidence exists to support the concept that people also acquire information that was never sought and about which the individual may have had no predisposition. In essence, there are three ways in which people acquire information, each of which should be supported by a digital library:

1. from the search for information about a well-defined and known object(s);
2. from the search for information about an object that cannot be fully described, but will be recognized on sight; and,
3. from the accidental, incidental, or serendipitous discovery of an object.

Systems that support the first prompt users for search terms and keywords, and provide options for manipulating the query and the result. Hypermedia, visualization and menu-driven systems that provide views and overviews of the data facilitate the second. In both cases, the next step taken, e.g., next index term selected, next page examined or information node accessed has a conceptual and/or a semantic relationship with the previous one. In the first, that relationship is predetermined by the system's response to the query; while in the second, the relationship is enabled for the most part by directed pathways through the system. While the process of information seeking is more complex than this, the outcome is generally the same: to locate the information object; this is a highly purposive task. The third type, the serendipitous approach, is a type of information seeking that is not traditionally examined in information retrieval research and has received little attention by both developers and researchers. It is this latter type that is the focus of this paper.

Examining Serendipity in the Context of Information Retrieval

The importance of stimulating curiosity and encouraging exploration in the pursuit of knowledge has been around for quite some time. Grose and Line (1968) suggested that books be randomly shelved to facilitate novel browsing while the *Oceanographic Literature Review* contains much more than just oceanographic literature in an attempt to induce serendipity (Shepherd, 1983). In a speculative paper, Ford (1999) discusses how information retrieval systems might support the creative process.

While the notion of serendipitous interactions with information systems is generally acknowledged, a growing body of evidence in a diverse series of research studies is illuminating the concept. Erdelez (1997) interviewed

academic library users and discovered that some were „serendipity-prone," tending to be super-encounterers who actually „felt the pressure of the abundance of information waiting to be encountered." Ross (1999) found from interviewing 194 committed readers that many found insights from pleasure reading that related to themselves and their own problems, concluding that they were „finding without seeking." Williamson (1998) studied the information-seeking behavior of 202 senior citizens in Australia by taping their phone conversations. She discovered that they often experienced „incidental information acquisition." The common thread throughout these studies was the relationship between the serendipitous discoveries and the individual: each matched a current information need, either implicit or explicit, or to some aspect of the person's life experiences. While none of these studies were completed within digital information environments, they do provide solid evidence of the role that serendipity plays in information acquisition. This must not be ignored in the design of digital libraries.

Serendipitous information retrieval occurs when a user with no a priori intentions interacts with a node of information and acquires useful information. We can say that the retrieval was triggered 'by chance' or a chance encounter (Toms, 1998b). A chance encounter occurs at the point in human interaction with an information system when a human makes an accidental and often sagacious discovery. The encounter is generally influenced by the person's prior knowledge, although not necessarily, and by the person's recognition of the 'affordances' (Gibson, 1977) in the text (Toms, 2000). Like the role of serendipity in scientific breakthroughs and discoveries, these chance encounters may lead to an outcome not anticipated at the outset.

Inducing Serendipity

In her study of 47 readers of a daily digital newspaper, Toms (2000) experimentally manipulated the purpose with which users approached the paper and the tools used to facilitate use. She compared:

1. two purposes: one with a specified goal and one with no pre-defined goal. The first was operationalized as 'find the answer to a set of questions', while the second was 'read/browse the newspaper for the next 20 minutes'; and,
2. two methods of information access: one represented by a search tool which prompted for keywords, and one by a dynamic list of ten suggested news articles that were somewhat similar to the article currently displayed.

The purpose operationalized as 'no goal' was devised to simulate a type of browsing. However, it also created the potential for a chance encounter or serendipitous interaction, i.e., the user had no pre-determined objective in mind. The dynamic list constantly presented the user with novel articles and served as a trigger, to evoke a new connection or new idea. In the process of conducting this test, Toms attempted to trigger serendipity in user interaction with digital information system

The system used in this study (and fully described in Toms (2000)) had three fixed windows. One, occupying the right half of the screen, contained the text of a news article. Two windows, which split the left half, contained methods for accessing the news articles. The lower of the two windows contained a three-level hierarchical menu; the upper contained either the results from using the search tool or the list of suggestions; only one was accessible in a single experimental session. The system used a nearest neighbor (weighted Boolean) approach. Whenever a new article was selected for display in the right side (the text of only one article was displayed at a time), the list of suggestions was dynamically revised to contain the ten most similar articles. The similarity measure did not control for size of document. The list in actuality could range from most to least similar in topic.

Participants with a goal, even a goal that was only partially understood, searched thoroughly with the tools they were given and examined contents of articles clearly with the intention only of extracting the appropriate piece of information and moving on quickly. Those without a goal, the serendipitous, were less concerned about selecting a priori meaningful content, but were more interested in coverage and exploration. Thus, when the interaction was not guided by an objective, user decisions seemed less definitive and less predictable. Although participants used menus as the chief means of accessing newspaper articles, participants found the most interesting articles from the dynamic list. Some were esoteric and sensational while others evoked a thought or event held in memory. Their motivations were varied (Toms, 1999).

Participants, who came from all age groups and all walks of life (from cooks and unemployed welders to senior citizens), also recognized the need for and value of these chance encounters. Typically, participants in post-session reviews commented, „if you focus on your interests, then your interests are going to stay what they are.“ They expressed a need to see things that they defined as „low interest“ because otherwise „you're living with tunnel vision.“ Their responses suggest that encounters of this type are enriching and rewarding to the individual. Thus, participants had chance encounters, which may not have otherwise occurred had the dynamic list of suggested articles not been present.

Facilitating Serendipity

Serendipitous retrieval takes place in the context of browsing or searching a digital information space; people immerse themselves in the items that interest them, meandering from topic to topic while concurrently recognizing interesting and informative information en route. These activities may range from scanning a non-fiction book, to reading the morning newspaper, or perusing a business report. Each of these acts may result in the acquisition of new information, the rejection or confirmation of an idea, or the genesis or new, perhaps not wholly formed thoughts on a topic – none of which were the original intent of the user. In addition to supporting the typical information retrieval tasks, a digital library must also support these types of activities just as generations of physical libraries have by accident supported serendipitous interactions merely through the juxtapositioning of books on a shelf. A digital library must stimulate curiosity and encourage exploration so that user may make opportune discoveries. As illustrated above, this is as important to those who are just 'surfing' or examining the content of a digital library as it is to those making significant discoveries. As Csikszentmihalyi (1996) states, „Learning for its own sake is rewarding even if it fails to result in a public discovery.“ How might we facilitate such an activity?

O'Connor (1988) claims that such an environment can be simulated simply through the navigation of a collection by maximizing a user's ability to explore. He does not, however, suggest how that navigational trail is to be implemented. In the case of my study, the serendipitous activity was stimulated and triggered by a list of articles that bore some degree of similarity to the one currently displayed. Sometimes that similarity was low and often the titles of newspaper stories were cryptic and sensational, stimulating considerable interest. But, the triggering mechanism is not fully understood, although most people could explain the connection. Potentially, it was the placement of the current article with the most/least similar one. But many other scenarios could be hypothesized.

Just as the creative process must result in new ideas that are intelligible to the creator and understood in terms of prior knowledge (Boden, 1996), so too must the results of a chance encounter. This means that for any node to be of potential value to the user, it must either be related to prior experiences and exposure to a node or similar nodes, or must indirectly trigger a novel event. According to Foskett (1983), it is the concurrence in time or space of different ideas or information nodes that stimulates the cognitive processes. O'Connor (1988) similarly suggests that „…the recognition of a valuable document attribute connection discovered by means outside established access system rules and relying on a user's self-knowledge“ induces a serendipitous and likely sagacious response. The difficulty is in translating these conceptual ideas into operational components. My approach is to look to the creative process.

Current and Future Research

This research is currently looking at serendipitous retrieval as the outcome from a creative act and looking to creativity research (for example, Austin, 1978; Beveridge, 1950; Ackoff & Vergara, 1981; Bawden, 1986) for a sound conceptual foundation on how to trigger serendipitous encounters. Some of the approaches we are exploring are simplistic while others are more complex such as:

1. Role of chance or 'blind luck': implemented via a random information node generator.
2. Pasteur principle („chance favours the prepared mind“): implemented via a user profile.
3. Anomalies and exceptions: partially implemented via poor similarity measures.
4. Reasoning by analogy: implementation is unknown at the moment.

These are just a few of the relationships among the information nodes that might be made explicit to induce serendipity or provide an environment for serendipitous experiences to occur. It remains to be seen how human will perform with a variety of tools present at the interface and which of those tools might be the most useful.

Developing appropriate triggers for chance encounters is but one part of the problem, integrating triggers into the digital library interface so that the system is somewhat intuitive to users and passes usability is an added conundrum. Ford (1999), however, cautions that there are inherent limitations in a system that must deliver precision along side diversity to support seemingly conflicting goals.

Conclusion

In this paper, I have argued for the inclusion of the creative process in information retrieval systems and in particular in digital libraries. While significant evidence exist to support the value of serendipitous experiences, few information systems support such an method. Yet, their value is generally assumed. Serendipitous retrieval demands approaching information retrieval in an unorthodox manner, one that does not tightly couple the explicit match of query with result, but instead takes a fuzzy approach to the problem.

References

Austin, JH (1978). *Chase, Chance and Creativity*. Columbia University Press.

Bawden, D (1986). „Information systems and the stimulation of creativity." *Journal of Information Science*, 12, 203-216.

Beveridge, WIB (1980). *Seeds of Discovery*. London: Heinemann.

Boden, MA (1996). „Agents & creativity." In B. Gorayska and JL Mey (eds), *Cognitive Technology: In Search of a Humane Interface* (pp. 119-127). Elsevier Science B.V.

Csikszentmialyi, M (1996). *Creativity: Flow and the Psychology of Discovery and Invention.* HarperPerennial.

Erdelez, S (1997). „Information encountering: a conceptual framework for accidental information discovery." In *Information Seeking in Context: Proceedings of an International Conference on Research in Information Needs, Seeking and Use in Difference Contexts, Tampere, Finland,* (pp.412-421). London: Taylor Graham.

Ford, N (1999). „Information retrieval and creativity: towards support for the original thinker.*" Journal of Documentation*, 55(5): 528-542.

Foskett, DJ (1983). *Pathways to Communication*. London: Bingley.

Gibson, JJ (1977). „The theory of affordances," In R. Shaw & J. Bransford (eds.), *Perceiving, Acting and Knowing: Toward an Ecological Psychology* (pp.67-82). Hillsdale: Lawrence Erlbaum.

Grose, MW & Line MB (1968). „On the construction of white elephants – some fundamental questions concerning the catalogue." *Library Association Record,* 70, 2-5.

O'Connor, B (1988). „Fostering creativity: enhancing the browsing environment." *International journal of information management*, *8*(3), 203-210.

Ross, CS (1999). „Finding without seeking: the information encounter in the context of reading for pleasure." *Information Processing and Management*, 35, 783-799.

Shepherd, FC (1983). „The various roles of secondary publications (some thoughts)." *International Journal of Micrographics and Video Technology* 2, 101-104.

Toms, EG (1998). „What motivates the browser?" In *Exploring the Contexts of Information Behaviour: Proceedings of the Second International Conference on Research in Information Needs, Seeking and Use in Different Contexts* (pp. 191-208). Taylor Graham.

Toms, EG (1998b). „Information exploration of the third kind: the concept of chance encounters." A position paper for the *CHI 98 Workshop on Innovation and Evaluation in Information Exploration Interfaces*. (http://www.fxpal.com/chi98ie/submissions/long/toms/index.htm)

Toms, EG (2000). „Understanding and facilitating the browsing of electronic text." *International Journal of Human Computer Studies*, 52, 423-452.

Williamson, K (1998). „Discovered by chance: the role of incidental information acquisition in an ecological model of information use. *Library and Information Science Research*, 20(1), 23-40.