

# Using Contextual Factors to Match Intent

Luanne Freund  
Faculty of Information Studies  
University of Toronto  
Toronto, Ontario, Canada  
freund@fis.utoronto.ca

Elaine G. Toms  
Faculty of Management  
Dalhousie University  
Halifax, Nova Scotia, Canada  
etoms@dal.ca

## 1. INTRODUCTION

*We need to free ourselves from the assumption that we can describe information behavior by starting from the system, the service, the knowledge base, or the information carrier. To use these as definers of useful information is misleading: only the recipient, the user, can define information in his or her context. This is not to say that this interface between information system, information artifact and user is not complex and intellectually interesting, but the final determiner of information value is the user who sits in a particular context and develops criteria of information value from that context.*  
Robert Taylor [1]

Contextual information retrieval (IR) is an open and promising field of research. *Open*, because context can refer to virtually everything and anything, and *promising*, because it has the potential to push IR research towards a more complex and naturalistic approach to human information interaction. There are some associated challenges. Firstly, to define which context we are concerned with, and secondly, to deal with the “messiness” of human behaviour, which is far from algorithmic. In this brief position paper, we will outline our general approach to contextual IR, argue for the importance of a number of key inter-related contextual factors, and review some of our related research projects in this area.

Context is important to human information behaviour for two key reasons: context disambiguates meaning in language and communication, and context shapes and delineates human needs and behaviour [2]. In the first case, context can be useful in determining aboutness and topical relevance. In the second case, it can be useful in determining how people search and on what basis they select documents, i.e. situational, cognitive and affective relevance [3, 4]. Our interest in contextual IR is less in the first sense, as a means of constraining the semantic space, and more in the latter sense of trying to make use of information about search behaviour and document preferences in different situations to achieve “higher-order relevance”[5]. We argue, that to move beyond topical IR, we need to use context to match documents with users/information needs on the basis of function, i.e., the intended use.

A document is more than a ‘bag of words’: it is the product of the author’s intent to communicate some idea for some purpose. This purpose is associated with an information goal, i.e., explaining, teaching, disproving, etc., and may also be associated with a work

task, i.e., reviewing a film, designing yachts, or calculating highway capacity. This purpose is expressed in the document in the decisions that the author makes with respect to use of language, structure and design. Recurrent situations and intents within particular domains engender recurring patterns of these document features, which can be interpreted as document *genres*. So, from the document side, we view genre as an expression of the authors’ intent with respect to the function of the document.

On the other side of the matching algorithm, the searcher also has intent, which likewise is associated with information goals and possibly motivated by work tasks. Although the searcher’s intent is seldom expressed directly in search queries, which tend to be primarily topical in nature, the underlying intent plays an important role in the document evaluation and selection process. It stands to reason, then, that identifying correlations between the intent of the document’s author as expressed in the genre, and the intent of a document searcher would offer some benefit in calculating relevance. We are not suggesting that these correlations are simple, absolute or one-to-one. On the contrary, the relationship is likely to be far more complex. However, we suggest that within specific domains, there are patterns of relationships sufficiently strong and consistent to be identifiable and to provide some benefit.

## 2. CONTEXTUAL FACTORS

Following from the discussion above, the key contextual factors that we are currently exploring in our research are: **document genre, domain, work task and information goal.**

### 2.1 Genre

We take a functional rather than strictly classificatory approach to genre, based on an approach to genre theory that has been developing over the past 25 years in a range of disciplines: linguistics, communications, organizational behaviour, etc. From this perspective, genre can be defined as “*socially recognized types of communicative actions [...] that are habitually enacted by members of a community to realize particular purposes*” [6]. This approach suggests that genre provides a means of facilitating a connection between the communication purposes of the document creator and the tasks and goals of the reader/user. For example, a *manual* that communicates instructions on how to carry out some process, is of use to a reader who needs to complete that process, but may be of little use to a reader who needs to decide if the process is worth doing. Genre is typically described as a function of the *form, content and purpose* dimensions of documents. Form can be further deconstructed to medium, structure and linguistic features [7]. More recent works that have focused specifically on digital genres have added a further dimension relating to the *mode of interaction* [4, 8]. In

practice, most implementations of genre in information retrieval have been limited to providing the searcher with the option to either limit or cluster results according to genre [9, 10] rather than weighting genre in the matching / ranking algorithms.

## 2.2 Domain

The idea that domain or discipline shapes information behaviour is not new. It is theoretically grounded in work in a range of information science related areas: domain analysis, information seeking studies, and communities of practice, to name a few. Practical implementations of domain context in information retrieval have focused largely on domain-specific search systems, some of which include customized search features, interfaces, and controlled vocabularies. It is less common to use domain as a means of differentiating between individual searches in a multi-domain search system, however some steps are being made in this direction [11, 12]. There is a clear connection between domain and genre, in that, "different disciplines or discourse communities develop special kinds of documents as adaptations to their special needs"[13]. The sets of genres in common usage within particular domains or communities of practice are known as *genre repertoires* [7].

## 2.3 Work Task

The role of tasks in motivating and framing information behaviour is receiving increasing attention in information science research [14-16]. Work tasks are particular cases of tasks, in which the doer is assigned an activity or unit of work that needs to be completed in order to meet a pre-defined goal, usually having some extrinsic value. When there is an information gap that stands in the way of work task completion, it prompts a search for information [17]. Work tasks can be identified at varying levels of granularity, and typically involve some sub-tasks or steps. To date, research on work tasks has shown that there is a significant relationship between various characteristics of work tasks (time constraints, importance, stage of completion, complexity, etc.) and various measures of information seeking behaviour (selection of channels, number and type of sources consulted, cognitive activities, etc.) [14, 15, 18, 19]. However, there has been little practical application of these findings in information retrieval systems to date. Furthermore, the relationship between work task and domain, although implicit in genre research, has been explored to a limited extent in select domains only, such as bio-informatics [20].

## 2.4 Information Goal

Nested within the work task is another layer of context for the search situation: the *information goal*. Information goals refer to the kinds of information that people are seeking and what they intend to do with it. Limberg [21] found that students with different broad information goals (fact-finding, assessing an issue and reaching a decision, understanding a topic) handled and evaluated information differently with respect to a number of parameters, including relevance, bias, information quantity and authority. Studies of searching behaviour have also found that different types of searching and selecting behaviours are observed for different types of information goals, such as known item searches, subject or topical searches, fact finding, and question-answering [22]. But the relationship among information goal, work task and document genre has been largely unexplored.

## 3. RESEARCH PROJECTS

In the past few years we have investigated some of these variables in independent studies. A study of 48 Web searchers searching in four domains (consumer health, research, shopping and travel) found significant differences among the four domains on many variables including genre [11]. Different web page genres were selected as relevant with respect to the questions from each domain. In some cases, genres were used across multiple domains, while other genres were clearly domain specific [23]. In a more recent study that examined the browsing actions of those who view cultural and heritage websites, we again found unique genres within sub domains of culture and heritage [24]. Compare for example, the genres of sports websites with those of art gallery sites. Thus, early work suggests a relationship between genre and domain.

In our most recent work, we are examining a more complex set of relationships: the associations among genre, work task and information goal within a single work domain. We are working with a community of software engineers within a large high tech company to develop and test an approach to contextual search for their internal information resources. In an initial study of these engineers we discovered that work tasks, genres and information goals were inter-related [25], which we later confirmed using a metadata-rich database of user-authored documents [26]. We were able to define sets of common tasks (installation, troubleshooting, etc), information goals (learning, doing, finding facts, etc) and genres (FAQs, product manuals, whitepapers, etc.) and the relationships between them. We are currently implementing these relationships in a search engine for this domain. The system will collect task and goal input from searchers in addition to the query, and use this information to adjust the ranking of documents according to their genre.

## 4. CONCLUSION

Understanding context and isolating selected contextual variables has the potential to enhance and focus retrieval. The keys are in uncovering which of the multiple variables to consider and in understanding the relationship between the variables and retrieval results. We believe that relating domain, work task, genre and information goal can make a contribution, and hope through ongoing work to further define and test the strength of the effect on retrieval results. This approach has the potential to be applied in a range of retrieval environments in which task and genre are prevalent, including digital libraries, enterprise search and Internet retrieval.

## 5. ACKNOWLEDGMENTS

This research was funded by an IBM Centre for Advanced Studies Doctoral Fellowship to the first author, and SSHRC and Canada Research Chairs programs grants to the second author. The authors acknowledge the support of Charlie Clarke, Julie Waterhouse and Gordon Lee.

## 6. REFERENCES

- [1] R. S. Taylor, "Information use environments," *Progress in Communication Sciences*, vol. 10, pp. 217-255, 1991.

- [2] J. D. Johnson, "On contexts in information seeking," *Journal of the American Society for Information Science*, vol. 39, pp. 735-760, 2003.
- [3] T. Saracevic, "Relevance reconsidered," presented at Second Conference on Conceptions of Library and Information Science (CoLIS 2), Copenhagen, Denmark, 1996.
- [4] I. Askehave and A. E. Nielsen, "What are the characteristics of digital genres? - genre theory from a multi-modal perspective," presented at 38th Hawaii International Conference on System Sciences, 2005.
- [5] J. Kekalainen and K. Jarvelin, "Evaluating information retrieval systems under the challenges of interaction and multidimensional dynamic relevance," presented at 4th CoLIS Conference, 2002.
- [6] J. Yates, W. J. Orlikowski, and K. Okamura, "Explicit and implicit structuring of genres in electronic communication: reinforcement and change of social interaction," *Organization Science*, vol. 10, pp. 83-103, 1999.
- [7] W. J. Orlikowski and J. Yates, "Genre repertoire: the structuring of communicative practices in organizations," *Administrative Science Quarterly*, vol. 39, pp. 541-574, 1994.
- [8] M. Shepherd and C. Watters, "The functionality attribute of cybergenres," presented at 32nd Hawaii International Conference on System Sciences (HICSS '99), 1999.
- [9] I. Bretan, J. Dewe, A. Hallberg, N. Wolkert, and J. Karlgren, "Web-specific genre visualization," presented at WebNet '98, Orlando Florida, 1998.
- [10] E. J. Glover, S. Lawrence, M. D. Gordon, W. P. Birmingham, and C. L. Giles, "Web search -- your way," *Communications of the ACM*, vol. 44, pp. 97-102, 2001.
- [11] E. G. Toms, L. Freund, R. Kopak, and J. C. Bartlett, "The effect of task domain on search," presented at Proceedings of CASCON 2003, Markham, ON, 2003.
- [12] B. M. Wildemuth, "The effects of domain knowledge on search tactic formulation," *Journal of the American Society for Information Science and Technology*, vol. 55, pp. 246 - 258, 2004.
- [13] B. Hjørland, "Domain analysis in information science: eleven approaches," *Journal of Documentation*, vol. 58, pp. 422-462, 2002.
- [14] P. Vakkari, "Task-based information searching," *Annual Review of Information Science and Technology*, vol. 37, pp. 413-463, 2003.
- [15] K. Bystrom and K. Jarvelin, "Task complexity affects information seeking and use," *Information Processing & Management*, vol. 31, pp. 191-213, 1995.
- [16] K. Bystrom and P. Hansen, "Conceptual framework for tasks in information studies," *Journal of the American Society for Information Science and Technology*, 2005.
- [17] N. J. Belkin, R. N. Oddy, and H. M. Brooks, "ASK for information retrieval: Part I: background and theory," *Journal of Documentation*, vol. 38, pp. 61-71, 1982.
- [18] K. Bystrom, "Information and information sources in tasks of varying complexity," *Journal of the American Society for Information Science*, vol. 53, pp. 581-591, 2002.
- [19] J. Algon, "The effect of task on the information-related behaviors of individuals in a work-group environment," in *School of Communication, Information and Library Studies*. Rutgers,N.J.: Rutgers, 1999.
- [20] J. C. Bartlett and E. G. Toms, "Capturing and modelling the research process of bioinformatics experts: an integrated information behaviour and task analysis approach," *Journal of the American Society for Information Science and Technology*, vol. 56, pp. 469-482, 2005.
- [21] L. Limberg, "Experiencing information seeking and learning: a study of the interaction between two phenomena," *Information Research*, vol. 5, 1999.
- [22] I. Hsieh-Yee, "Research on Web search behavior," *Library and Information Science Research*, vol. 23, pp. 167-185, 2001.
- [23] E. G. Toms, L. Freund, and R. Kopak, "A holistic examination of search results: the relationship between relevance, genre and task domain.," (*in preparation*), 2005.
- [24] E. G. Toms, C. Dufour, and E. Dawe, "Cultural and heritage genres," (*draft in progress, available from the authors*), 2005.
- [25] L. Freund, E. G. Toms, and J. Waterhouse, "Modeling the information behaviour of software engineers using a work - task framework," presented at American Society of Information Science & Technology Annual Meeting, (accepted), 2005.
- [26] L. Freund, E. G. Toms, and C. L. A. Clarke, "Modeling task-genre relationships for IR in the workplace," in *SIGIR 2005 (accepted)*, 2005.