The Centrality of User Modeling to High Recall with High Precision Search

Dan Brassil
H5
San Francisco, USA
dbrassil@h5.com

Christopher Hogan
H5
San Francisco, USA
chogan@h5.com

Simon Attfield
UCLIC
London, UK
s.attfield@cs.ucl.ac.uk

***Abstract*—The objective of Search is to find documents relevant to a particular user's notion of relevance. However, relevance is often a moving target: imperfectly defined and subject to change as more documents are seen. In this paper, we report on systematic User Modeling (UM) and the use of a system-internal agent (proxy) to produce a hybrid human-computer system that achieves extraordinarily high performance on mediated Search tasks.**

We present details of our UM-approach and its four main components: (i) use case (ii) scope (iii) nuance and (iv) linguistic variability. We illustrate how these components provide a framework with which a user and a proxy co-construct a shared representation of information needs and mutual knowledge. This representation serves as the common ground through which external knowledge is shared, mediated, negotiated, synthesized and made accessible to the system.

We evaluated the performance of our system on the Legacy Tobacco Documents Library, a corpus of advertising, manufacturing, marketing, sales and scientific research activities of major US tobacco companies. Independently adjudicated results from NIST's 2008 TREC legal track demonstrate that our approach to UM yields high performance on Search tasks.

Keywords– Accuracy; Complex Litigation; Human-Computer Interaction; Information Retrieval, Sensemaking

I. INTRODUCTION

The effectiveness of a Search system or protocol is measured by how well it retrieves relevant documents from a corpus. Relevance is a derived property that entails a user and an information need: a document is deemed relevant by a user if it satisfies that user's information need (cf. [1]).

A standard assumption for information retrieval systems is that the definition of what makes a document relevant or not exists independently of the system. It assumes that the user of the system has a preexisting, well-defined and unchanging notion of relevance, and that it is the purpose of the system to identify any documents that are relevant according to that notion of relevance.

For certain types of information needs, the assumption of fixed relevance can be reasonable. For example, in known-item search[1], the user is attempting to find an item that he or she knows to exist (such as querying a library's search engine with a specific book's title to locate that book within the library).

For more complex types of information needs, the assumption of fixed relevance breaks down. A user often approaches a Search task seeking to resolve an "anomalous state of knowledge" [3]. Moreover the user often cannot precisely specify what information is needed to resolve his or her anomalous knowledge-state [3]. This type of information-seeking uncertainty has been a particularly enduring finding in information-seeking research. For example, Kuhlthau [4] found that during the information-search process, information needs begin as vague and unclear, but as the user develops greater focus, searching becomes more targeted and precise. Also Vakkari [5] observed a similar impact of uncertainty resolution on the refinement of search tactics, search terms, relevance judgments and sources used during a longitudinal study of an extended information-seeking task. And Bystöm and Järvelin [6] note that information-seeking uncertainty is a particular characteristic of knowledge work which typically features *a priori* indeterminability concerning task outcomes and information requirements.

In these situations, an exploratory information need [7] exists with the assumption that certain aspects of the information need are as yet undefined, and will be further refined through interaction with an information retrieval system. Even if the user does have a well-defined notion of relevance at the outset, that notion may not remain unchanged by the experience of interacting with the system and the information it returns to the user: documents returned by the system may contain information that the user did not previously know about, which may, in turn, refine or change the user's notion of relevance (as observed, for example, by Yang [8] and Tang and Solomon [9]).

Another common assumption is that the user is only interested in a subset of highly relevant documents. For certain information needs, such as the above-mentioned known-item search, a precision-oriented approach is appropriate. In this

[1] See [2] for discussion of known-item search and the assumptions inherent in various definitions of the concept within Library and Information Science research/literature.

case, the relevant set usually consists of one document, so the precision measure effectively captures the performance of the system. For other, more complex, information needs, recall is as important as precision. In these cases, the system needs to retrieve most, if not all, relevant documents in the corpus. In this case, both precision and recall are the measures by which the system's performance is evaluated.

The intersection of exploratory information needs and high-recall-with-high-precision needs is often seen in complex litigation. In complex litigation a senior litigator must oversee a large document production to opposing counsel in order to comply with his or her legal obligations as outlined in US Federal Rule 26(b). The problem arises because large, heterogeneous, and unstructured document populations [10] are often the corpora senior litigators must evaluate to produce the relevant document productions to opposing counsel. While there are a number of methods that assist a senior litigator in complying with rule 26(b), such as manual document review, key-word filtering, Boolean queries, latent semantic indexing, *et cetera*, the fact remains that the senior litigator must first determine what he or she needs to find. Without such a determination of information need, any resultant search or classification task will fail (cf. [11]). This type of information-need determination characteristically evolves through interaction with document exemplars and ongoing issue refinement [12]. As such, complex litigation can be seen as a class of *sensemaking* activity, where sensemaking is, "the reciprocal interaction of information seeking, meaning ascription and action" [13, p240].

The key questions to be answered, then, are these: How can we most effectively help the user collect, organize and create representations of the complex information they need to understand? Given limitations on the user's time and attention, what is the best way to structure the conversation with the user so as to acquire the most information with the least effort? And finally, given a certain amount of information, how best to go about representing it in a way that it is consumable by an automated system? These are the questions with which this paper is concerned.

II. USER MODELING

We have already noted that the efficacy of Search is determined by how well it returns relevant documents from a corpus; this entails that, at some level of the information retrieval system, there must exist a representation of a user and his or her information need. In other words, some amount of User Modeling (UM) must occur for a search task to be successful.

We understand UM to be a two-fold endeavor: (i) constructing a definition of relevance and (ii) iteratively interacting with a user to increase the likelihood of relevance in the output. We follow Saracevic et al [14] in positing that mediated interaction, that is interaction of a user, a human intermediary and an information retrieval system, is the most effective form of UM for Search. Within such a model, an intermediary is an "intelligent agent constructing, implementing and modifying user models in all their complexity with considerable feedback" [14][2].

A. UM as co-construction

There are two central tenets of our approach to UM: (i) a user is seeking to resolve an "anomalous state of knowledge" and (ii) the user is unable to precisely specify what information is needed to resolve the anomalous knowledge-state [3]. These tenets underlie our own endeavors as intermediaries: we are seeking to resolve an anomalous state of knowledge as it pertains to satisfying the user's information need and we are unable to precisely define what information will satisfy the user's information need. Moreover, we recognize that users and intermediaries have access to external knowledge sources (personal knowledge, reference guides, the target corpus, etc.) that can be leveraged to inform and refine the model. Thus, the act of UM is a co-construction of information needs and mutual knowledge[3] in a shared representation.

Underpinning our approach to UM is sensemaking which proceeds through an ongoing interaction between *top-down* processes (i.e. guided by existing knowledge and interpretations), and *bottom-up* processes (i.e. guided by new data). This dynamic lies at the heart of several prominent accounts of sensemaking, including Klein et al's data frame theory [18]; Weick's account of sensemaking in organizational settings [19]; Russell et al's learning loop complex model [10]; and Pirolli and Card's account of sensemaking by intelligence analysts [20]. In information-rich environments, this leads to iteration between two complementary focusing strategies: *issue focusing* involves refining issues and questions in the light of new information; *data focusing* involves selecting (e.g. searching for) information to address the current formulation of issues and questions

In this context, User Modeling (UM) serves as a powerful shared source of information by providing a mechanism by which external knowledge can be formalized into the system via vocabularies, query development, and various forms of training data.

We assume a model, based on [21] and depicted in Figure 1, in which the representation serves as the common ground through which external knowledge is shared, mediated, negotiated, synthesized and made accessible to the system. It is this aspect of our approach to UM that allows the intermediary to become a *proxy* for the user thereby permitting the proxy to arbitrate whether information is assessed as relevant or not relevant.

Our framework for UM comprises four component areas: (i) use case, (ii) scope, (iii) nuance and (iv) linguistic

---

[2] The relationship of the intermediary to the user and the information retrieval system is one of systems boundaries. Buckland and Plaunt [15] write that "systems boundaries define what is considered the 'system' and what is considered the 'environment'". On this definition , whether or not the intermediary is within the system is determined by how integrated the intermediary is into the design of the overall system.

[3] For more on co-construction of knowledge and mutual understanding see [16] and [17].

variability. The resultant representation is a description of subject matter, that, if found in a document, would make that document relevant.

1) Use Case

When a user approaches a search task, aspects of the task other than what needs to be found, are often left implicit. While defining and understanding what it is a user is seeking, equally important to understand are:

a) User Objective(s) – What goal is the user attempting to achieve? How does his or her information-seeking behavior [22] assist in achieving this goal?

b) *Performance Requirements* – What degree of performance is required for effective retrieval? Is a low recall but highly precise output accurate enough for the task at hand? Or is high recall with high precision needed for the retrieval to be deemed a success?

Answers for these questions provide valuable information for overall system design. Moreover, the proxy tests where and how use case considerations might manifest themselves in the other components of UM: scope, nuance and linguistic variability.

2) Scope

Some information needs are more difficult than others. One way in which difficult information needs differ from easier ones is in how challenging it is to evaluate whether a particular document is relevant or not. In the case of known-item retrieval [23], for example, relevance can be evaluated by comparing equality with a known result. In more difficult cases, determining whether a single document is relevant or not can be challenging. To wit, we engage in an iterative process of scope definition with the user.

We define scope as the breadth of concepts considered relevant by the user. When engaging in scope discussions, we seek to define the boundaries of relevance for a given conceptual domain. Potentially relevant subject matter is divided into *core* concepts and *peripheral* concepts in order to provide a semantic framework for user-proxy interactions. During UM, a combination of issue-focusing and data-focusing techniques are used such as employing questionnaires, exemplar documents drawn from the target corpus, and user interviews to draw the user's attention to the (potential) complexity of his or her information need. The user's responses are analyzed and the ramifications and logical extensions of the responses are discussed with the user. This process is iterated until a shared definition of scope is agreed to.

3) Nuance

While scope discussions allow the user's information need to be modeled on a conceptual-semantic level, a mechanism is needed for identifying and assessing how relevant concepts are linguistically manifested within documents that comprise the corpus. We propose that an understanding of nuance, which we define as the degree of specificity required to be relevant, provides such a mechanism.

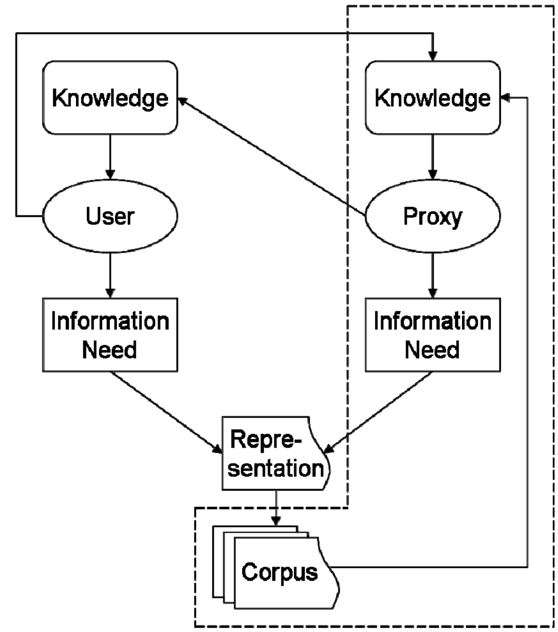

Figure 1: Representation of proposed User Modeling framework. The portion within the dashed line is internal to the system.

In UM, nuance discussions revolve around lexical relationship between words such as meronymy (part-whole relationship), hypernymy-hyponymy (class-member relationship), metonymy (close-association substitution), *et cetera*. Combinations of lexical relationships are tested and explored to further refine the model's ability to inform the identification of linguistic manifestations of relevance. Again, issue focusing and data focusing techniques similar to the ones using in Scoping discussions (i.e questionnaires, exemplar passages drawn from the corpus and user-interviews) are used to facilitate the process which is iterated until a shared definition is achieved.

4) Linguistic Variability

Related to, but distinct from, nuance is linguistic variability. Linguistic variability is the variety of ways a concept can be lexically (e.g. polysemy and synonymy (cf. [24]) or syntactically (e.g. agentless passive constructions) expressed. The problem that linguistic variability poses for Search tasks have been well-studied and a number of proposals to deal with it have been offered ([25] *inter alia*). We propose that the method by which linguistic variability is handled within a document retrieval system is intrinsically related to the user's use case, scope and nuance.

Within our UM framework, two approaches to addressing linguistic variability are evaluated: defining each concept as a closed set of words/phrases or defining each concept in terms of pertinent characteristics (which allows the user/proxy to recognize unforeseen instantiations of relevance).

The appropriateness of employing one approach over the other is determined by analyzing the interplay between the user's objectives and performance requirements, the complexity of the information need and how the concept in question manifests within the target corpus.

B. Modification

Belkin [3] notes that “a change in one’s state of knowledge, by virtue of having engaged with text, will be reflected in some change in the anomalous state of knowledge”. Because our approach to UM assumes anomalous states of knowledge on the part of both the user and proxy, we build into the UM process a data focusing “check-in” procedure which allows the system to adapt to new information as do the user and the proxy so as to converge on a definition of relevance that meets the user’s information need.

III. SYSTEM OVERVIEW

In this section we situate UM into the overall design of our system. Our system comprises one main agent, the proxy, and four separate, yet interconnected, processes: User Modeling, Assessment, Classification and Measurement. A diagram is shown in Figure 2.

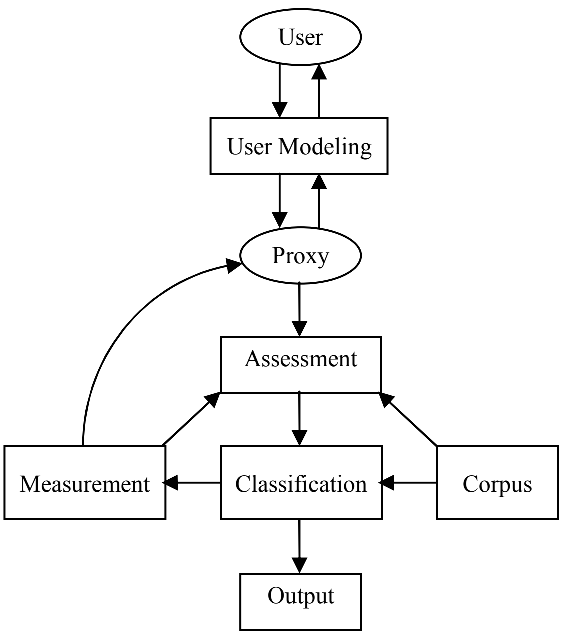

Figure 2: System Architecture

A. Proxy

The proxy is an internal agent who co-constructs a theory of relevance with the user via User Modeling. The proxy provides guidance to document assessors and resolves intra-and inter-assessor discrepancies to ensure that errors are re-solved in favor of the proper interpretation of relevance.

B. User Modeling

As discussed in greater detail in section II, User Modeling is the process by which the proxy co-determines a theory of relevance with the user, iterating the process to increase the likelihood of relevance within the system’s output.

C. Assessment

The assessment process is designed to (i) generate a large amount of training data (ii) of the appropriate kind (iii) with minimal error. The assessment process consists of an initial assessment of all documents of interest and subsequent error correction procedures.

By bringing assessment into the system, several improvements are actualized. If relevance is to be modified in response to further knowledge, it is more cost effective to modify an internal notion of relevance and verify that it remains consistent with the external notion than to attempt to modify the external notion.

D. Classification

Document-assessment pairs generated during assessment are used as training data for a supervised classification system. The classifier is trained over available assessments and the resulting model used to perform a binary classification of all documents.

E. Measurement

The performance of the classification system is regularly evaluated in order to test its efficacy. The classification system is run over all documents in the corpus. Following classification, a random sample is drawn and reviewed by document assessors. Data generated by the evaluation process are used to tune the system and may result in the proxy and user modifying the theory of relevance.

IV. EXPERIMENTAL RESULTS

We evaluated our system in the context of the 2008 TREC Legal Track Interactive Task. The Legal Track was established by TREC to evaluate approaches to information retrieval with application to the problems encountered in the legal domain, including especially electronic discovery (e-discovery)[4]. The Interactive Task was formulated in order to provide a more realistic setting for the e-discovery information retrieval task, and includes the following features [26]:

- A Topic Authority (i.e. user) is the sole determiner of relevance
- Teams have 10 hours to interact with the Topic Authority
- Teams must submit a binary classification for every document in the population
- Teams can appeal assessment decisions with Topic Authority making the final decision

The Interactive Task was evaluated on the Legacy Tobacco Documents Library, comprising 6,910,192 documents released under the tobacco “Master Settlement Agreement”.

Given this context, we evaluated on the following topic:

> **Topic 103.** All documents which describe, refer to, report on, or mention any “in-store,” “on-counter,” “point of sale,” or other retail marketing campaigns

[4] Electronic discovery refers to any process in which electronically stored information is located, secured and searched in order to use it as evidence in a civil or criminal legal case.

Of the 600 minutes (10 hours) allotted each team to interact with the Topic Authority, we used 490 minutes to conduct UM. During this time, we determined the following:

Use Case – The Topic Authority had two objectives: to produce to opposing counsel a set of documents deemed responsive to the Request for Production (RFP) [primary objective] and to mitigate the risk of being accused of under-producing (i.e. intentionally withholding responsive documents) or over-producing (i.e. intentionally delivering non-responsive documents) [secondary objective].

The decision to prioritize one risk over the other has far-reaching design decisions: under-production < over-production implies a narrow, more exclusive conception of relevance whereas under-production > over-production implies a broad, more inclusive conception of relevance.

During UM, we learned the user felt that the risk of under-production accusations outweighed the risk of over-production accusations. It was also determined that the user's performance requirements, required high performance in both recall and precision.

Scope – During scope discussions, we sought to understand how the topic authority interpreted *retail marketing campaigns*. We analyzed the phrase, creating questions that tested the scope of each word: types of retail outlets, the activities that constitute marketing, and the characteristics of a campaign.

Based on the Topic Authority's responses, we discovered that she did not consider *retail* a delimiting factor, thereby broadening the scope of the RFP to (cigarette) marketing campaigns in general. Furthermore, it was decided that the breadth, depth and complexity of the concepts implicated by the Topic Authority's interpretation of the RFP was best handled by deconstructing it into 18 sub-topics.

Nuance - In the context of the TREC Interactive Task, discussions of nuance and specificity centered on the semantic relations hyponymy and hypernymy. For instance, it was agreed to that a hyponym of campaign, such as *Marlboro Ranch* (a name of a specific marketing campaign) should be considered, in and of itself, a marker of relevance, whereas the non-specific hypernym *campaign* should not be considered, in and of itself, a marker of relevance.

Linguistic Variability – An analysis of the Topic Authority's use case, the complexity of the topic (and its sub-topics) and the level of specificity required by some concepts, determined that the most appropriate method for handling linguistic variability was definition-by-characteristic.

Modification – As mentioned in section II.B, designed into UM was a "check-in" procedure to ensure convergence of relevancy interpretation. The check-in was implemented as a mechanism by which the proxy could evaluate interpretation discrepancies that might have arisen between the user and proxy, in recognition that interaction with external knowledge sources (such as the corpus) impacts knowledge states and thus might necessitate up-dating the co-defined theory of relevance (cf. [9]). During the check-in, such a discrepancy was discovered: the user presented an alternate interpretation of relevance concerning the degree of specificity required for a

TABLE I. INITIAL DEFINITION OF RELEVANCE – PROMOTIONS AND MEDIA

	Specific Media	Non-specific Media
Specific Brand	R	NR
Non-specific Brand	NR	NR

TABLE II. FINAL DEFINITION OF RELEVANCE – PROMOTIONS AND MEDIA

	Specific Media	Non-specific Media
Specific Brand	R	R
Non-specific Brand	R	NR

determination of relevance for discussions of cigarette brand promotions via media outlets. Prior to the check-in, a discussion of promoting a cigarette brand through a media outlet required a specific brand and specific media outlet be discussed for an assessment of relevant to be valid (e.g. A marketing budget indicating an advertisement for *Lucky Strike* being placed in *Newsweek*). Generality in either domain did not meet the definition of relevance (cf. Table 1)[5]. The user's alternate interpretation allowed non-specificity in one domain but not both (cf. Table 2).

Based on this change in interpretation, the definition of relevance was modified as was the representation of the user's information needs. The interpretation modification discussed resulted in an increase in overall yield since documents which contain discussions of placing promotional material of specific brands in non-specific media outlets like those found constitute a fair number of the documents changed from NR to R by internal assessors at the direction of the proxy.

In all, document reviewers made assessments on 7992 documents, providing substantial training data for the system's classifier. The output of the final classification determined that, of the 6,910,192 documents in the population, 608,807 (8.8%) were relevant, and the remaining 6,301,385 (91.2%) non-relevant. According to TREC-adjudicated results[6], an estimated 12.7% of the corpus was relevant (yield), giving an estimated Precision measure of 83.0% and Recall measure of 78.6%. These results significantly exceed those achieved by systems taking more traditional information retrieval approaches [27].

V. CONCLUSIONS

We have presented a novel approach to addressing the task of large-scale Search in which high accuracy, that is high recall with simultaneous high precision, is the desired result. We have shown that a multi-faceted approach to User Modeling that defines a user's information need by addressing it as four distinct yet interrelated areas, is central to this endeavor as it provides a framework to construct a shared understanding of relevance, a means for representing that understanding to an automated system, and a mechanism for iterating and correcting such a system so as to converge on a desired result.

[5] R refers to relevant; NR refers to non-relevant.

[6] These results are the TREC coordinators' OCR-adjusted results (§4.5.4 of [17]).

Central to the approach is the idea of capturing the theory of relevance as it evolves through interaction with information. The decomposition of an RFP into sub-topics and the significance of exploring hyponymy relations (e.g. the identification of specific marketing campaigns) reflects a recursive focusing of a broad information need into multiple lower-level needs. Recursive decomposition of this type has been reported as a key success factor in large regulatory investigations (e.g. the investigation of a class of contracts leads to the identification and investigation of specific contracts in that class) [28]. Recursive refinement is an essential strategy for achieving high precision and recall in complex search problems and it is one that is systematized through the approach described here.

The User Modeling approach mediates and translates between the user and the IR system in a way that both responds to and structures complex information problems involving sensemaking as they naturally evolve towards a solution. The problem of how notions of relevance are defined and converted into a computerized representation is deserving of further research, with consequences not only for the Legal community but for all areas of human endeavor with massive, comprehensive Search problems.

REFERENCES

[1] T. Saracevic. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and Manifestations of Relevance. *Journal of the American Society for Information Science and Technology*, 58(3):1915–1933. 2007

[2] J. H. Lee, A. Renear, L. Smith. Known-Item Search: Variations on a Concept. *Proceedings of the American Society for Information Science and Technology*, 43(1). 2007.

[3] N. Belkin. Anomalous States of Knowledge as a Basis for Information Retrieval. *Canadian Journal of Information Science*, 5:133-143. 1980.

[4] C.C. Kuhlthau. A principle of uncertainty for information seeking. *Journal of Documentation*, 49(4):39–55. 1993.

[5] P. Vakkari. A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of Documentation*, 57(1):44–60. 2001.

[6] K. Byström, K. Järvelin. Task complexity affects information seeking and use. *Information Processing and Management*, 31(2):191–213. 1995.

[7] P. Ingwersen and K. Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer. 2005.

[8] S.C. Yang. Information seeking as problem-solving using a qualitative approach to uncover the novice learner's information–seeking processes in a Perseus hypertext system. *Library and Information Science Research*, 19(1):71–92. 1997.

[9] R. Tang, P. Solomon. Toward an understanding of the dynamics of relevance judgment: An analysis of one person's search behaviour. *Information Processing and Management*, 34(2-3):237–256. 1998

[10] D. M. Russell, M. J. Stefik, P. L. Pirolli, S. K. Card. The Cost Structure of Sensemaking. *Proceedings of InterCHI, ACM*, pp. 269-276. 1993.

[11] R. Bauer, T. Jade, B. Hedin, C. Hogan. 2008. Automated Legal Sensemaking: The Centrality of Relevance and Intentionality. *2nd International Workshop on Supporting Search and Sensemaking for Electronically Stored Information in Discovery Proceedings*. 2008. Available online at:
http://www.cs.ucl.ac.uk/staff/S.Attfield/desi/5.%20Bauer.pdf

[12] S. Attfield, A. Blandford. Improving the cost structure of sensemaking tasks: Analyzing user concepts to inform information system design. To appear in *Proceedings of INTERACT 2009*.

[13] J.B. Thomas, S.M. Clark, D.A. Gioia. Strategic sensemaking and organisational performance: linkages among scanning, interpretation, action and outcomes. *Academy of Management Journal*, 36: 239-270. 1993.

[14] T. Saracevic,, A. Spink, and M.-M. Wu. Users and Intermediaries in Information Retrieval: What are they talking about? In *User Modeling. Proceedings of the Sixth International Conference, UM97*: pp. 43–54. 1997.

[15] M. K. Buckland, and C. Plaunt. On the construction of selection systems. *Library Hi Tech*, 12(4):15-28. 1994.

[16] J. S. Brown. *A Symbiotic Theory Formation System*. PhD thesis, University of Michigan. 1972.

[17] J. Roschelle and S. Teasley. The construction of shared knowledge in collaborative problem solving. In C. O'Malley, ed., *Computer-supported collaborative learning*, pp. 69-77. Springer-Verlag, Heidelberg, Germany. 1995.

[18] G. Klein, J.K. Phillips, E.L. Rall, D.A. Peluso. A Data-frame Theory of Sensemaking. In: *Expertise Out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making* 2003. US: Lawrence Erlbaum Assoc. Inc, 2007, pp.113–155.

[19] K. Weick. Sensemaking in organisations. London, England: Sage Publications, Inc. 1995.

[20] Pirolli P. and Card S., The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In: *Proceedings of the International Conference on Intelligence Analysis,* McLean, VA. 2005.

[21] N. Belkin, H., R. Oddy, H. Brooks. ASK for information retrieval: Part I. background and theory. *Journal of Documentation*, 38(2):66–71. 1982.

[22] N. Belkin. Interaction with texts: Information retrieval as information-seeking behavior. In *Information Retrieval '93: Von der Modellierung zur Anwendung*, pp. 55–66. Konstanz. 1993.

[23] A. Broder. A Taxonomy of web search. ACM SIGIR Forum, v.36 n.2, Fall 2002.

[24] H. Uejima, T. Miura, I. Shioya. Improving text categorization by resolving semantic ambiguity. *Systems and Computers in Japan* 36(4):1-8. 2005.

[25] F. W. Kroon. *Linguistic Variation in Information Retrieval Using Query Reformulation*. M.S. thesis. Simon Fraser University. 2002.

[26] J. Baron, B. Hedin, D. Oard, S. Tomlinson. TREC-2008 Legal Track Interactive Task – Guidelines. 2008. Available online at: http://treclegal.umiacs.umd.edu/2008InteractiveGuidelines.pdf

[27] D. Oard, B. Hedin, S. Tomlinson, J. Baron. Overview of the TREC 2008 Legal Track. In *The Seventeenth Text REtrieval Conference (TREC 2008) Proceedings*. 2009.

[28] S. Attfield, A. Blandford, S. De Gabrielle Investigations within investigations: a recursive framework for scalable sensemaking support. *Sensemaking workshop (CHI 2008)* Available online at: http://www.uclic.ucl.ac.uk/people/s.attfield/saabsdgInvestigationsWithinInvestigations.pdf