View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UCL Discovery

Automated Legal Sensemaking: The Centrality of Relevance and Intentionality

Robert S. Bauer, Teresa Jade, Bruce Hedin, Chris Hogan
H5
71 Stevenson St.
San Francisco, CA 94105
{rbauer, tjade, bhedin, chogan}@H5.com

1. Introduction

In a perfect world, discovery would ideally be conducted by the senior litigator who is responsible for developing and fully understanding all nuances of their client's legal strategy. Of course today we must deal with the explosion of electronically stored information (ESI) that never is less than tens-of-thousands of documents in small cases and now increasingly involves multi-million-document populations for internal corporate investigations and litigations. Therefore scalable processes and technologies are required as a substitute for the authority's judgment. The approaches taken have typically either substituted large teams of surrogate human reviewers using vastly simplified issue coding reference materials or employed increasingly sophisticated computational resources with little focus on quality metrics to insure retrieval consistent with the legal goal. What is required is a system (people, process, and technology) that replicates and automates the senior litigator's human judgment.

In this paper we utilize 15 years of sensemaking research to establish the minimum acceptable basis for conducting a document review that meets the needs of a legal proceeding. There is no substitute for a rigorous characterization of the explicit and tacit goals of the senior litigator. Once a process has been established for capturing the authority's ***relevance*** criteria, we argue that literal translation of requirements into technical specifications does not properly account for the activities or states-of-affairs of interest. Having only a data warehouse of written records, it is also necessary to discover the ***intentions*** of actors involved in textual communications. We present quantitative results for a process and technology approach that automates effective legal sensemaking.

2. Sensemaking and Relevance

We look to cognitive-task-analysis research to characterize the sensemaking behaviors ("making sense" of it all) of a senior litigator conducting a document review. "Sensemaking" is necessary for any decision-making; however, as today's information environments have become increasingly complex, decision-making has become much more difficult and time-consuming. So understanding massive and diverse content is not just a simple matter of consuming information or finding it faster. More advanced approaches for interacting with information are needed and current keyword-search and data-mining methods simply cannot meet these needs. [1] Indeed, the courts have recently ruled that the drastic oversimplification of the e-Discovery task as a simple search exercise is wholly inadequate:

> "Whether search terms or 'keywords' will yield the information sought is a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics ... Given this complexity, for lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information than the terms that were used is truly to go where angels fear to tread." [2]

COPYRIGHT © 2008 – H5

> "Use of search and information retrieval (IR) methodology, for the purpose of identifying and withholding privileged or work-product protected information from production, requires the utmost care in selecting methodology that is appropriate for the task because the consequence of failing to do so, as in this case, may be the disclosure of privileged/protected information to an adverse party, resulting in a determination by the court that the privilege/protection has been waived." [3]

Indeed, the effective modeling of the e-Discovery task requires 'making sense' of the salient aspects of the senior litigator's sensemaking efforts. This task description holds whether a computer system is used or a team of human surrogate reviews conducts the review (with or without technology support.) Therefore, the proper framework for considering how to tag relevant documents within a discoverable document population must incorporate two iterative sensemaking loops, shown schematically in Figure 1.

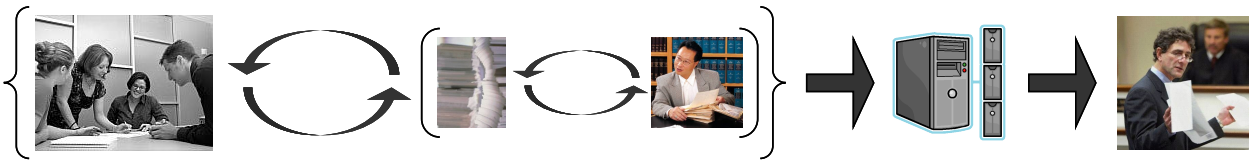

FIGURE 1: Two sensemaking efforts are involved in providing a legal team with an effective document set from a large document population of electronically stored information (ESI:) 'Making sense' {outer braces} of the sensemaking activities of the senior litigator [inner brackets.] Machine automation of e-Discovery is shown as the computation over the ESI utilizing the knowledge-based system (KBS) developed by an interdisciplinary team.

The primary sensemaking loop of e-Discovery involves the senior litigator who inherently evaluates document relevance by assessing the intentions of the document's author relative to the legal strategy of the case [shown in Fig. 1 within the inner brackets.] Any effort that scales beyond a single litigator must simulate this sensemaking activity as closely as possible. Consistent with many other findings [4,] any method that depends primarily on human review fails to transfer properly the requisite knowledge of the senior litigator's sensemaking into a consistent, reproducible document review. We overcome this inherently human limitation by utilizing a multi-disciplinary team of linguists, lawyers, and subject matter experts to codify their meta-sensemaking model {shown in Fig. 1 within the outer braces,} into a knowledge-based system that replicates the litigator's primary sensemaking [inner brackets]. Rigorous relevance criteria and in-process measurements of statistically-valid ESI samples are required to assess and ensure accuracy. The document set needed by the legal team can then be produced without further human participation by applying the KBS to the entire ESI document population; this automation often employs a massively distributed computational infrastructure because the ESI scale is typically massive.bb

As Russell, et. al. describe: "Sensemaking is simple—it's the way people go about their process of collecting, organizing and creating representations of complex information sets, all centered around some problem they need to understand." [5] Clearly, if the "problem" requiring "understanding" is not fully characterized, then the resulting document review will fail. As Russell, et. al put it in their seminal 1993 paper on "making sense" of large, heterogeneous, and often unstructured document content populations: This is a "general phenomena in which part of the job of sensemaking is to establish the goals of the task." [6] As noted above, the e-Discovery goal must be established by the senior litigator whose authoritative judgment serves as the only true criteria for a successful review.

COPYRIGHT © 2008 – H5

There are any numbers of ways to establish what we call 'Relevance Criteria' (RC) to guide document review. Engineering frameworks abound for capturing requirements; for example, the Institute of Electrical and Electronics Engineers (IEEE) defines a requirement as "a condition or capability needed by a user to solve a problem or achieve an objective." While this may seem obvious, it is well established that requirements engineering is the hardest single part of building a software system. [7] In recognition of the central role of authoritatively establishing relevance for document review, the 2008 TREC Legal Track has instituted an interactive task that incorporates a 'Topic Authority' to "represent the senior litigator who engages the services of an e-discovery firm." [8] This is the only way to ensure that legal sensemaking is relevant to the proceedings.

3. Sensemaking and Intentionality

The means of establishing the e-Discovery goal (i.e., RC development) must be coupled with an understanding of what the senior litigator would consider to be important evidence that meets the objectives. In addressing this aspect of sensemaking, it is critically important to recognize that the reasoning of senior and junior members of a litigation team is usually quite different. A classical method for understanding the structure of a cognitive activity is to study occupational experts of that activity; for the case of sensemaking, cognitive task research of intelligence analysts has been conducted for at least the past 10 years. Takayama and Card report that senior and junior analyst behaviors are nearly the opposite of each other: "A general trend of more top-down behavior in seniors and more bottom-up behaviors in juniors has become apparent. Senior analysts begin with their own hypotheses and large personal repositories of information before reaching out to more distant sources to fill in gaps or get updates." [9] The "hypotheses" for a senior litigator is the context for and intentionality of the author of documents comprising the ESI; the "large personal repositories of information" are accumulated by the senior litigator over his/her years of legal experience. This distinction between senior and junior analyses has been characterized by Pirolli as information processing "driven by bottom-up processes (from data to theory) or top-down (from theory to data)." [10] This is central to efforts to automate the sensemaking task because it dictates that replicating senior litigator sensemaking must be rooted not in 'data mining' approaches but in systems that reason from a set of "hypotheses." As noted for the legal domain, these hypothetical constructs characterize the expected intentions of individuals that are involved in the topics of interest in the case.

The crucial insight based on sensemaking research is that in for e-Discovery, senior litigators are NOT reviewing the literal content of text (i.e., bottom-up), but rather the overarching aspects of the situation and the author's intent (i.e., top-down.) Figure 2 depicts five essential elements required to characterize the intentions that underlay the relevance goals of an e-Discovery effort. [11]

COPYRIGHT © 2008 – H5

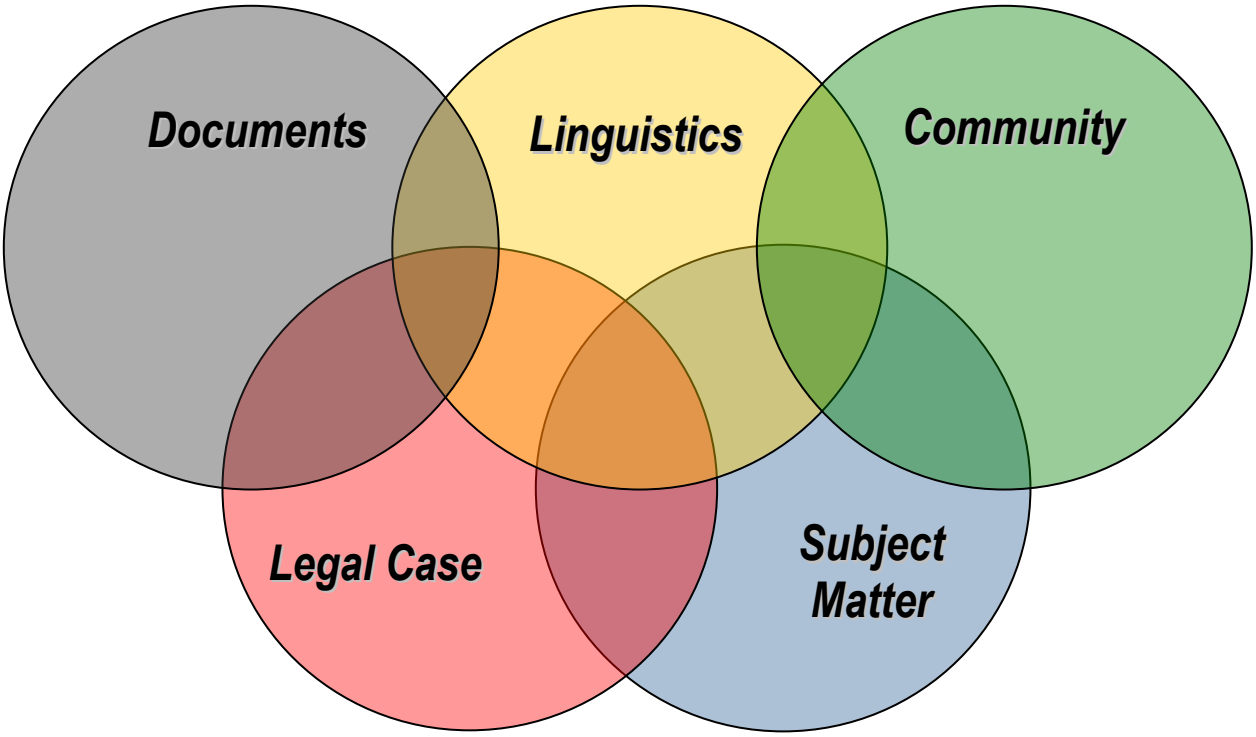

FIGURE 2: Five inter-related aspects that must be considered in characterizing the intention behind the written record. A system or approach that does not at least partially take all of these elements into account is unlikely to consistently achieve the kinds of retrieval results that are acceptable in the legal context.

We find that Relevance Criteria for topics in a review can vary considerably. Some topics are rather simple and readily map to the documents; straight-forward information retrieval techniques can do quite well in finding material that captures such intent. However, other topics are frequently quite complex and require modeling of specific community practices in order to adequately characterize the often subtle, but critical distinctions between relevant and irrelevant documents. Indeed, combining these various dimensions of e-Discovery is a good way to distinguish different approaches and understand their limitations.

a. Documents + Legal Case
   In any kind of text-based classification or retrieval system, the documents may simply serve as the target or may also inform the query. In a legal context, the documents contain content and are often associated with metadata (for example, the author of the document, sent date, etc.) The most basic approach to finding relevant documents in the context of litigation involves querying the documents for the topics of interest in the legal case. For example, an attorney might search for "evidence to support a damages claim." However, the complexity of the legal topics and the fact that the documents were created for other purposes makes a direct mapping of the documents to the legal topics very difficult.
   Examples of systems that only consider these 2 dimensions include:
   - Manual (human) review conducted by attorneys
   - Basic keyword searches targeted to legal issues
   - Supervised learning with relevance feedback

b. Documents + Legal Case + Subject Matter
   Some legal teams hire subject matter experts to assist them in reviewing particular sets of results. For example, experts in accounting are frequently consulted in cases where calculations regarding damages are required. A key requirement for a

COPYRIGHT © 2008 – H5

successful system in such a case is a clear and concrete characterization of the target documents. We have found that subject matter can form a bridge from the documents to the legal case. In other words translating the legal case into key subject matter areas creates clarity around which documents will be of interest. For example, when searching for evidence of anti-competitive activities, defining the search in terms of the sales and marketing practices can illuminate where opportunities in the market are unfairly blocked by a competitor.
Examples of systems that consider only these 3 dimensions include:
Subject matter experts review results under legal team direction
Use of Domain-specific lexicons

c. Documents + Legal Case + Subject Matter + Linguistics
Meaning is encoded in language. Linguists are trained to model the salient morphological, syntactic, semantic, sociolinguistic and discourse aspects of the way meaning is encoded in language. The same concepts expressed in different contexts will involve different phraseology. For example, a linguist who analyzes the content of a PowerPoint slide deck can model both the language of that presentation as well as how the presenter might express the same concepts when communicating through email with the boss.
Examples of systems that consider these 4 dimensions include:
Supervised learning (including both relevance feedback and semantic analysis)
Semantic search

d. Documents + Legal Case + Subject Matter + Linguistics + Community
The final key element required to achieve consistently high Recall (R) with high Precision (P) is the characterization of the community in which the documents are created and used. Legal teams often consider various individuals as they prepare for depositions and may compile organization charts and the like to understand roles and responsibilities of key personnel; however, review teams rarely if ever model the processes, states-of-affairs, or idiosyncratic terminology of the communities in which these players participate. In order to understand the document population indicative of the activities of document authors, such rich operating characterizations of communities are vital.
An example system that combines all 5 dimensions is:
Socio-Technical Information Retrieval (STIR) [11]

It is inherent to the top-down sensemaking approach of senior litigators that they review documents with an implicit model of the communities in which actors participate. Individuals create documents in the course of their activities as members of "communities of practice." [12] We all work in communities over significant period of time wherein specific tools, processes, resources, and language develop; just recall the idiosyncratic acronyms that must be decoded to understand what is transpiring in a new organization you have joined. The language and linguistic forms used to encode meaning are informed by an author's memberships and roles in their multiplicity of communities. Text cannot be analyzed for the intention of the writer without accounting for the community in which it was created and used.

While the specific framework (such as [13] for 3.d. Communities) may vary, and not every topic requires the same amount of depth, any system that does not take into account all five elements in Figure 2 is unlikely to consistently achieve the kinds of retrieval results that are desirable in the legal context. The nearly universally overlooked sensemaking requirement for e-Discovery that accounts for author intent is the characterization of the community context.

COPYRIGHT © 2008 – H5

4. Automating Sensemaking: Information Retrieval Processes & Technologies

In order to create an automated legal sensemaking system (shown by computers in Fig.1,) we must consider the meta-sensemaking framework that an expert multi-disciplinary team can use to capture the salient characteristics {shown in braces in Fig.1,} of the senior litigator's sensemaking [shown in inner brackets in Fig.1.] While sensemaking might seem like a vague concept, cognitive task analysis of sensemaking suggests that there is a relatively well-defined structure to the phenomenon. [6, 9] The basic model (called "a learning loop complex") can be summarized in terms of two processes: (1) searching for a representation or framework scheme and (2) actually filling in the framework with the data collected. Attempting to fill in the framework will end up with some data that doesn't fit (called 'residue';) this requires a shift in the representation and then another attempt to fill it in with the data. For our purposes, the knowledge representation used for the multi-disciplinary team's sensemaking is the computational constructs into which the linguistic variations can be expressed and the operations (e.g., Boolean expressions) that enable phrasing of query alternations which capture the meaning being sought in IR computation. The Relevance Criteria developed with the senior litigator (or topic authority in TREC) is the determinant of what constitutes data 'residue' in documents mistakenly tagged as either relevant or irrelevant; this in turn leads to specification of the representational iteration and expressive richness required of the knowledge framework. When the framework accounts for all 5 elements necessary to characterize the intentionality of document authors shown in Figure 2, the information retrieval result is dramatically better than achieved by conventional search technologies plus/or by armies of junior (bottom-up) reviewers.

Results are shown in Figure 3 for iterative development of a litigation review utilizing a hybrid, automated e-Discovery approach addressing all 5 dimensions in Figure 2. Called 'Socio-Technical Information Retrieval' or 'STIR,' [11] it is a knowledge-based system in the classic AI sense of replicating the cognitive sensemaking task of a senior litigator with an automated, computational platform.

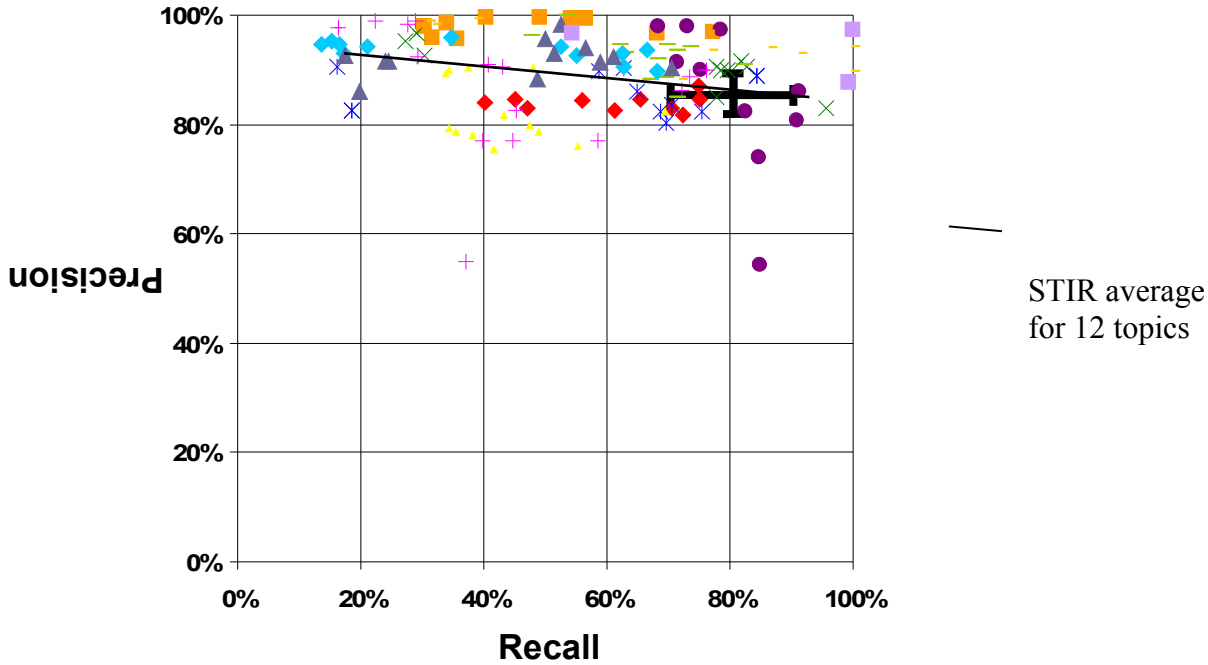

FIGURE 3: Sampled corpus tests for 12 topics during Socio-Technical Information Retrieval (STIR) sensemaking development (i.e., "the process of searching for a representation and encoding data in that representation to answer task-specific

COPYRIGHT © 2008 – H5

questions.") Each topic is a different color and all topics move through time from left to right. The black line (from P~93% @R~18% to P~85%@R~80%) is an average for all the topics measured during the building of the knowledge-based system for the legal case.

When STIR development on a case begins, confidence intervals for statistically-valid Recall estimates are rather large, so particular attention is paid to the measure of lower bound. Recall estimation is usually the hardest part of this sort of evaluation; we estimate Recall by comparing the number of documents machine-model tagged as 'relevant' to the total number of documents determined to be 'relevant' in a double-vetted, quality-controlled, human review of the same, randomly-selected sample population. Over time, retrieval accuracy increases (e.g., increased F-measure [14]) with concomitant narrowing of the confidence intervals (typically +/-6% Recall and +/-3% Precision in our studies.) As would be expected, Precision generally decreases as Recall increases. At the conclusion of the sensemaking development effort utilizing STIR (result shown with confidence intervals of +/-6% R, +/-3% P,) the average quality of the sampled result set for 12 distinct legal issues is dramatically higher than typical TREC interactive task results [15.] The automated STIR sensemaking approach at 80% Recall (i.e., where 4 out of 5 documents of interest to a legal matter would be in the retrieved set) correctly identifies as 'relevant' 4 out of 5 documents in the result set. By comparison, AT LEAST ½ of the documents relevant to a legal case will normally be missed by 'standard' IR approaches when Precision is 80% or greater.

In practice, post-processing of the result sets by junior (bottom-up) reviews would require 4-5 times the resources to complete the e-Discovery effort compared to the automated sensemaking approach in Fig. 1 for the same Precision and Recall. Often the cost/time requirements for achieving acceptable Recall at high Precision with traditional IR review methods is prohibitive; therefore significant simplifying assumptions are made (often unknowingly) in order to reduce the reviewable document population to a manageable size (e.g., further keyword culling.) Such actions dramatically lower Recall at acceptable Precision in most reviews. Meaningful approaches must explicitly deal with fundamental requirements for making sense of vast amounts of information rather than accommodating to the parameters that are easily manipulated by retrieval tools (e.g., keyword specification, Boolean operators, thesauri context/'semantics'.)

Figure 3 shows results obtained by linguists using an appropriate, quantitative methodology that reproducibly employs processes, measures, representations, and technologies to craft queries that increasingly produce retrieval results with simultaneous high P & high R. This hybrid, multi-dimensional approach for conducting high-quality, automated, legal sensemaking (a) captures linguistic expertise, (b) characterizes particular practice communities of subject matter experts, and (c) employs some of the latest advances in AI, Natural Language Processing, and massively parallel processing. The penalty for ignoring intentionality and not using a rigorous quality-controlled development process with measurable goals is always greater than the upfront investment required to craft a case-specific, knowledge-based, IR system.

5. Advances in Automated Legal Sensemaking

Development of a fully automated system for legal sensemaking is a process of evolving representations, wherein people seek increasingly effective representations to support the review task and then use them to mechanistically process massive populations of ESI with any human review. This process of representational change during sensemaking is inherently complex, involving hypothesis and test. [16] As depicted in Figure 1, iterative human participation is necessary during this rich query development (i.e., case-specific KBS building;) such iterative query development by experts in the written discourse of practice communities consistently produces high quality, automated "Socio-Technical Information Retrieval" systems.

COPYRIGHT © 2008 – H5

We refer to such a system as 'Automated' because no human interaction is needed for successful conduct of the subsequent IR task on the full corpus.

Future advances must account for the fundamental characteristics of legal sensemaking. To conduct an effective and efficient document review, a system must replicate the sensemaking of senior litigators as a top-down, automated process of searching for a representation and encoding data in that representation to answer case-topic-specific questions. Therefore, two necessary aspects of any scalable, e-Discovery process or technology are (1) establishing explicit criteria for senior litigator relevance and (2) multi-dimensional coding for author intentionality. Clearly, drastic oversimplification of the review task as a keyword search exercise is not capable of the rich, nuanced queries required for sensemaking. Execution of the sensemaking approach requires rigorous measurement and statistically valid, in-process quality control. Without numerical results that characterize the degree of achievement for Precision and Recall, any claim of 'accuracy' in automating legal sensemaking is unsubstantiated.

References

1. Palo Alto Research Center's description of sensemaking research [see http://www.parc.com/about/pressroom/features/sensemaking_0606.html]
2. John Facciola, *United States v. O'Keefe*, No. 06-CR-249, 2008 WL 44972, at *8 (D.D.C. Feb. 18, 2008)
3. Paul Grimm, *Stanley v. Creative Pipe*, No. MJG-06-2662 p. 25-26 (D.MD. May 29, 2008)
4. See seminal study and its citations: D.C. Blair and M.E. Maron, "An evaluation of retrieval effectiveness," Communications of the ACM, 28 (1985), 289-299.
5. D. M. Russell, R. Jeffries, L. Irani, "Sensemaking for the Rest of Us,"CHI workshop on Sensemaking (2008.) [see http://dmrussell.googlepages.com/sensemakingworkshoppapers]
6. D. M. Russell, M. J. Stefik, P. L. Pirolli, S. K. Card, "The Cost Structure of Sensemaking," Proceeding of InterCHI, ACM. (1993), 269-276.
7. Carnegie Mellon's Software engineering Institute [see: http://www.sei.cmu.edu/productlines/frame_report/req_eng.htm]
8. National Institute of Standards and Technology (NIST) - TREC 2008 Legal Track: Interactive Task. [see: http://trec-legal.umiacs.umd.edu/]
9. L. Takayama, S. K. Card, "Tracing the Microstructure of Sensemaking," CHI workshop on Sensemaking (2008.)
10. P. L. Pirolli, "The Cognitive Structure of User Sense Making." 7th International Workshop of the EU Network of Excellence: DELOS on Audio-Visual Content and Information Visualization in Digital Libraries (AVIVDiLib'05,) Cortona, Italy. [see http://www.parc.com/research/publications/details.php?id=5455]
11. R. S. Bauer, T. Jade, M. P. Marcus, 11th International Conference on AI and Law (DESI I Workshop), Stanford, CA - June 4, 2007 [see http://www.umiacs.umd.edu/~oard/desi-ws/]
12. E. Wenger, "Communities of Practice: Learning, Meaning, and Identity," Cambridge: Cambridge University Press (1998.)
13. A. O. Putman, "Communities," in Advance in Descriptive Psychology, Vol. I, K. E. Davis, ed., JAI Press, Greenwich, CT (1981) 195-209.
14. A single, composite measure of information retrieval accuracy: The weighted harmonic mean of Precision and Recall [see: http://en.wikipedia.org/wiki/Information_retrieval]
15. See for example results in S. T. Dumais & N. J. Belkin, "The TREC Interactive Tracks: Putting the User into Search," Chapter 6 in E. M. Voorhees & D. K. Harman, ed., "TREC: Experiment and Evaluation in Information Retrieval," MIT Press (2005) 123-152.
16. G. W. Furnas, "Representational Change in Sensemaking," CHI Workshop on Sensemaking (2008.)

COPYRIGHT © 2008 – H5