ADAPTIVE MULTIMEDIA INFORMATION ACCESS

Ask Questions, Get Answers

Mark T. Maybury
Information Technology Division
The MITRE Corporation
202 Burlington Road
Bedford, MA 01730, USA
maybury@mitre.org
http://www.mitre.org/resources/centers/it

QUESTION ANSWERING

MITRE is pursuing a line of research aimed at creating systems that transform users from the traditional web search approach in which the user poses queries and gets (typically an overwhelming amount of purportedly relevant) pages, to one in which the user asks questions and gets answers. As implied by Figure 1, this suggests moving away from a conventional information retrieval strategy of document/web page retrieval to one requiring both multilingual and multimedia information extraction coupled with personalized presentation planning.

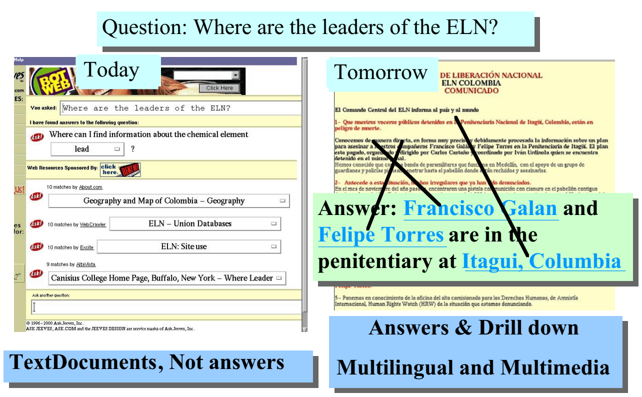


Figure 1. Ask Questions, Get Answers

The inadequacy of the current document retrieval strategy is underscored by Figure 2 that illustrates the (normalized into common units) growth in computing, storage, and high speed networks, suggesting that growth of distributed data is actually accelerating. Even more intense, .com organizations typically experience storage requirements that double every 90 days.

This paper outlines some of the key initiatives MITRE is pursuing to achieve the "ask questions, get answers" vision including spoken language information access, data mining of user information seeking activities and tool use, automated discovery of experts and their knowledge, collaborative searching, and multilingual and multimedia search. The paper concludes by articulating lessons learned from pursuing these areas.

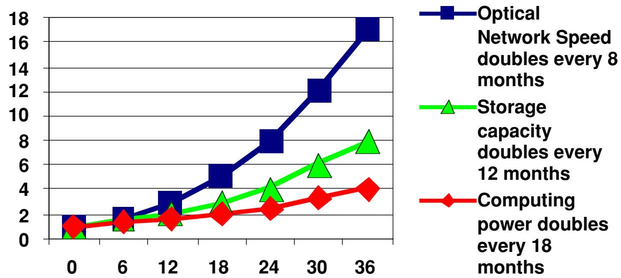


Figure 2. Acceleration of Infrastructure Growth

COMMUNICATOR

One means for enhancing access to the structured elements of the massive data stream of Figure 2 is to provide direct access to it in a manner that does not require the user to learn any special query language. The DARPA Communicator initiative aims to support natural conversational interaction to distributed on-line resources. This includes spoken access to web content, navigation, and summarization. Research foci of this initiative include dialogue management, multimodal input (speech, gesture), and output (synthesis, generation). One of the objectives of Communicator is to create a market and facilitate development via a component based, distributed architecture (see Figure 3) that is available via open source (*fofoca.mitre.org*).

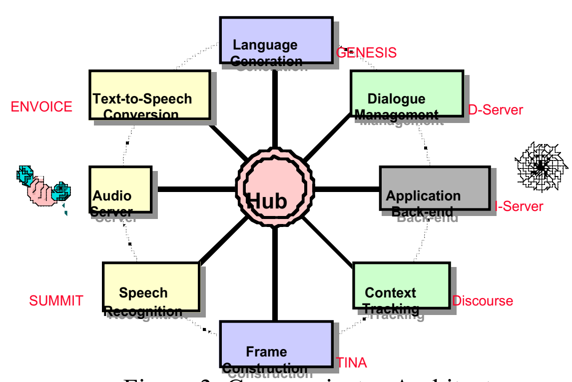
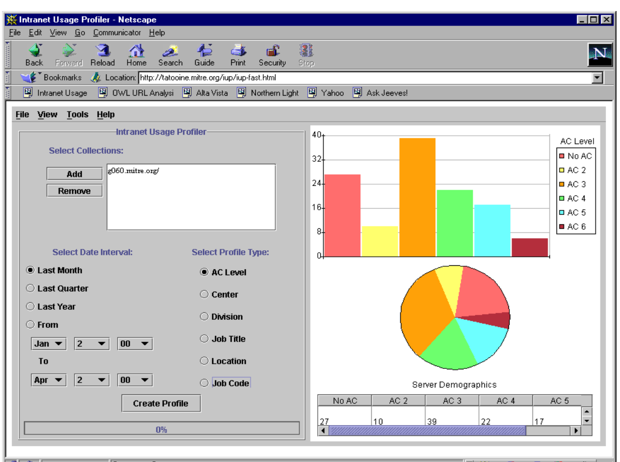


Figure 3. Communicator Architecture

The Communicator initiative is leveraging standards for plug & play and portability to new domains with the intent of lowering entry barrier to system development through componentware. For example, the JUPITER demonstration system created at the MIT LCS provides user with mobile access to weather information via a speaker independent phone interface. Try out Jupiter at www.sls.lcs.mit.edu/sls/whatwedo/applications/jupiter.html.

INTERNET USAGE PROFILER

As illustrated in Figure 2, the data explosion exists both within enterprises and across the web. We can exploit user information seeking behavior to develop interest models that can be used to tailor delivered information and/or leveraged by other users who have similar information needs. For example, MITRE's intranet, called The MITRE Information Infrastructure (MII), contains over 1 million pages from over 90 web servers and 300+ Collections (sets of URLs). Each of these collections has a steward who needs to understand collection usage to quantify benefits. To assist in collection stewards, MITRE's Intranet Usage Profiler (IUP) shown in Figure 4 supports interactive viewer demographic analysis, e.g., "What kinds of folks are looking at my web collection?" For example, Figure 4 depicts the distribution by technical level of individuals that are viewing the Information Technology Division (G60) web site within the MITRE Corporation.



viewer demographics, a sister system, the Intranet Usage Profiler Recommender System (IURS) turns this data around to enable the user to specify one's demographic profile (e.g., Applied Capability level 3 (AC-3) networking engineers living in Washington working in a particular division) and see specific URLs one's peers have visited.

KEAN

One of the problems with usage data is just because people visited a site does not mean that they found it useful. We would like a way we could implicitly (versus costly explicit asking) determine the value of a web site to a user. The Knowledge Editing and Annotation Environment (KEAN) addressed, among other things, the "cold start" problem in recommender systems. In particular, the first users of a recommender system can't benefit from the collective experience/knowledge of the community because no explicit utility ratings of sites exist. Mattox, Maybury, and Morey (1999) reported the study of the use of 295 URLs by 26 individuals to perform a range of information seeking tasks regarding directory services (e.g., "Which standards organization defines the X.500 specification?", "How does LDAP differ from X.500?", "Name some of the data types that can be stored in an LDAP attribute."). Participants were asked to rate the utility of web pages visited (on a Likert scale of 1 (low) to 10 (high) utility).

A statistical analysis of the web logs and associated utility ratings resulted in the discovery of a positive correlation between utility and the delta between individual URL accesses, or "dwell" time. In particular, we discovered explicit utility equals .0113 times "dwell" time. A data mining tool discovered that 66% of all URLs "dwelled upon" for greater than 78 seconds were classified as high utility (6-10) by users which, consistent with an initial hypothesis that short time deltas between web accesses indicated lack of relevancy, moderate deltas indicated reading (e.g., perception, cognition, assessment), and extremely long deltas (e.g., > 10minutes) indicated intermittent activity (e.g., other tasks, coffee breaks). With this finding, for similar tasks we can know monitor web logs and implicitly construct utility ratings to support a recommender system, such as KEAN.

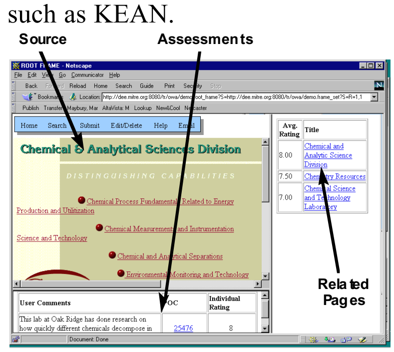


Figure 5. KEAN

Figure 5 illustrates the results of a KEAN query. KEAN couples a standard information retrieval engine with both explicit user qualitative and quantitative annotations (in lower pane) and the relevant and useful pages as determined implicitly by users browsing behaviour (in the right hand pane) in relation to a particular page (the main pane). KEAN provides users with the ability to search by subject, keyword, expert, rating level, and date/time. This enables new classes of user tailored queries and recommendations, such as:

- What information does Alfred (an expert in Adaptive Hypermedia) think is useful for adaptive hypermedia?
- What information do people in the Adaptive Hypermedia community of practice find useful on adaptive hypermedia?
- What information does everyone think is useful on adaptive hypermedia in the past few weeks?
- What information on adaptive hypermedia have I found to be useful in the past?

Users of this new class of recommender system have a distinct advantage both in terms of directly getting to the most "useful" information, but also by being able to explicitly compare their own assessments of utility with those of their community.

OWL

In addition to instrumenting user's information seeking behavior, we can track their utilization of tools, compare their performance to their peers and provide personalized, collaborative help to enhance their learning and performance. For example, Linton et al.'s (1999, 2000) Organizational Wide Learning (OWL) project instrumented Microsoft Word commands for dozens of users over many months of usage. As shown in Figure 6, approximately 10 commands account for nearly 80% of usage. More significantly, an individual user can compare their usage to the group's pooled knowledge to identify opportunities for learning. In the future, this could be extended, for example, to search for experts in particular functionality (e.g., graphics, tables, formatting), a topic of the next described system.

	Command	Percent	Cumulative
1	File Open	13.68	13.68
2	Edit Paste	12.50	26.18
3	File Save	11.03	37.22
4	File DocClose	10.25	47.47
5	Edit Delete	9.50	56.97
6	Edit Copy	7.86	64.83
7	Format Bold	4.22	69.05
8	File Print	4.12	73.16
9	Edit Cut	3.50	76.66
10	File Quit	2.73	79.39

Figure 6. OWL

EXPERT FINDER

Indeed, frequently the best way to answer your questions is to find an expert. "Skill mining automatically identifies the skills of knowledge workers by analyzing their past behavior" (Fenn 1999). Unfortunately, commercial tool and/or web sites that provide expert finding services have typically either depend upon recommender technology that depend upon explicit user ratings or on requiring (typically overcommitted) experts to manually maintain profiles of their expertise. These processes are often errorful, incomplete, expensive to maintain and quickly perishable. In addition to KEAN's instrumentation of user information seeking behavior and OWL's instrumentation of application function usage to discover skills, we can also utilize the intellectual by products of human effort to learn about and help the individual. A key knowledge management challenge is discovering who knows what independent of location or organization.

Expert Finder (Mattox et al. 1999) implicitly determines expertise from multiple sources of evidence including intellectual products (e.g., briefings, papers, web pages, resumes) created by a staff as well as what is publishing about a staff by others (e.g., in corporate newsletters, corporate directory services, project leadership information). Using information extraction (language processing) technology to extract proper names from documents, the system is able to correlate co-occurrences of keywords and terms with individuals. This evidence can be combined using a model of quality of sources. For example, content in a resume can be assumed more accurate than that from a document. Similarly, content published about an expert by a sanctioned corporate source is likely more reliable than an arbitrary source.

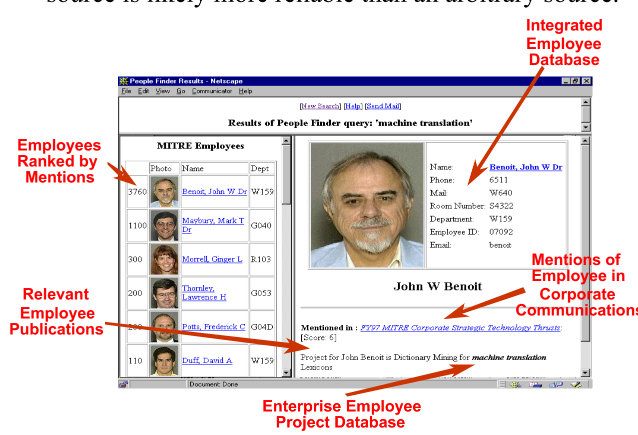


Figure 7. Expert Finder "machine translation" Experts

As shown in Figure 7, the Expert Finder system responds to the user keyword request (in this case "machine translation") by combining evidence and rank ordering the experts in the corporation and presenting these to the user. Empirical evaluation of Expert Finder and human performance (by 20 human resource managers) found 63% of the time humans identify the same individual when giving a list expert of the top 5 experts. Presented with the same task, expert finder identified approximately 30% of the experts the humans did in its top 5 (recall) and approximately 40% of the experts it listed in its top 5 were considered experts by the human judges (precision).

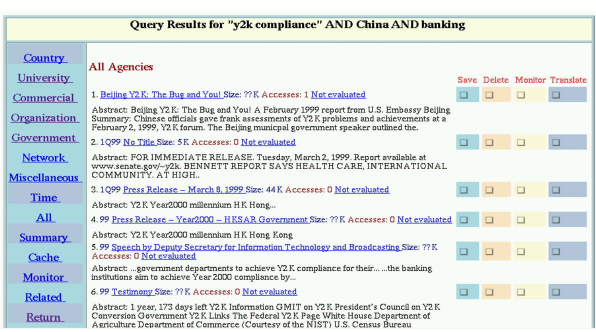
PRIVACY AND SECURITY

It is important to note that the ability to instrument individual information seeks behavior and applications implies both an opportunity and responsibility for protection of private information. While a full treatment of the subject of privacy and security is outside the scope of this article and US and European laws continue to evolve in this area, it is generally prudent to work only with public and/or explicitly provided information (as does Expert Finder), to ensure informed consent, to collect the minimal amount of personal information needed, to use it only for what you collected it for, to protect individual identity and so on (Kobsa 2000).

SCOUT

In addition to discovering , groups of experts working collaboratively may wish to perform their information seeking behaviors jointly. Doing this across time and space can be a challenge. Scout (D'Amore and Konchady 1999) is a multi-user collaborative retrieval tool that addresses multi-user, coordinated searching, shared analysis, using a built-in recommender system. It tracks topics, users, and provides a persistent knowledge store. As shown in Figure 8, a user of Scout generates task folders to organize the results from their searches (e.g. a list of monitored topics) that are shared among a group of searchers and/or search agents or web-page monitors.

Figure 8. Collaborative Document Search

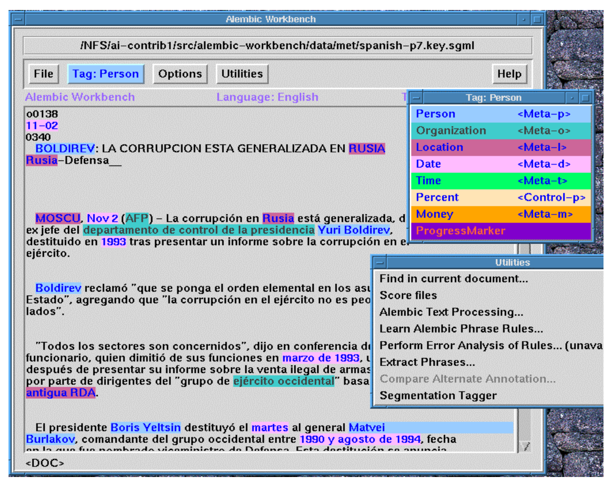


Relevance is inferred from user actions (e.g., save, delete, monitor, translate). This enables users to efficiently have a shared vision of the information space and to leverage one another's knowledge and information seeking activities, synchronously or asynchronously. Finally, user variability in information seeking tasks (e.g., querying, browsing, relevance assessments) can be exploited by the group. We are presently evaluating the ability of Scout to increase to accuracy and coverage of information seeking performance.

MULTILINGUAL ALEMBIC

In addition to leveraging and learning from multiparty behavior to enhance our ability to answer questions, in the future increasing amounts of information will be available on the web in language other than English. Researchers at Carnegie Mellon University estimated that in the Fall of 1999 web content globally became less than 50% English, underscoring the need for multilingual information access. MITRE's Alembic information extraction system, shown in Figure 9 and available on the web at www.mitre.org/technology/alembic-workbench. Alembic utilizes a mixed-initiative approach sometimes called "tag-a-little, learn-a-little, tag-a-little" in which the computer utilizes human judgement about lexical, syntactic and semantic classifications to learn linguistic rules. Figure 9 shows its operation in the task of being trained to performing Spanish named entity extraction. Alembic supports multilingual annotation (via UNICODE characters and fonts) and has builtin services for machine learning and content extraction evaluation.

Figure 9. Spanish Markup in the Alembic Workbench



MULTIMEDIA ACCESS: BNN

Just as multilingual content is increasing, so too individuals are faced with vast quantities of non-text multimedia (imagery, audio, video). Applications that promises on-demand access to multimedia information such as radio and broadcast news on a broad range of computing platforms (e.g. kiosk, mobile phone, PDA) offer new engineering challenges. Synergestic processing of speech, language and image/gesture promise both enhanced interaction at the interface and enhanced understanding of artifacts such as web, radio, and television sources (Maybury 2000). Coupled with user and discourse modeling, new services such as delivery of intelligent instruction and individually tailored personalcasts become possible.

Figure 10. Tailored Multimedia News

Figure 10 illustrates one such system, the Broadcast News Navigator (BNN) (Maybury et al. 1997). The web-based BNN gives the user the ability to browse, query (using free text or named entities), and view stories or their multimedia summaries (Figure 10 displays all stories about the Russian nuclear submarine disaster from multiple North American broadcasts from 14-18 August 2000). For each story, the user is given the ability to view its closed caption text, named entities (i.e., people, places, organizations, time, money), a generated multimedia summary, or view the full original video of a story.

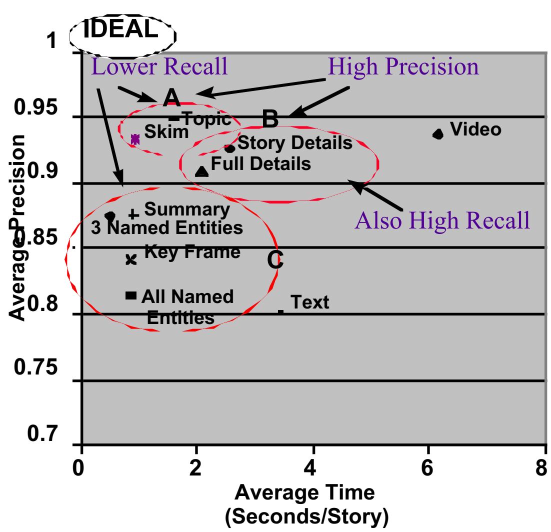


Figure 11. Relevant Judgement Performance with Varying Multimedia Displays

In empirical studies, Merlino and Maybury (1999) demonstrated (See Figure 11) that users could enhance their retrieval performance (a weighted combination of precision and recall) by utilizing BNN's StorySkim and Details presentations. In addition to task performance, users reported user satisfaction (1 dislike, 10 like) of 7.8% (for retrieval) and 8.2% for mixed media display (e.g., story skim, story details).

Community defined multimedia evaluations will be essential for progress; the key to this progress will be a shared infrastructure of benchmark tasks with training and test sets to support cross-site performance comparisons.

QUESTION ANSWERING

A new Text Retrieval Evaluation Conference (TREC) track focuses on providing answers to queries as opposed to documents. For example Breck et al's (2000) QANDA system aims to find explicitly stated answers in knowledge sources of varying degrees of structure. QANDA's research aims are to:

1. Understand the question well enough to "ask" the knowledge sources by extracting the characteristics of the answer and routing the question to best knowledge source.

2. Understand the knowledge sources well enough to find the answer weather they are in relational databases, semi-structured data, or human language text without structure.

3. Discover how this capability can help fulfill a user's information need, which could include ad hoc questions against a static database, standing questions against a stream of data, a series of related questions or templated questions.

Figure 12 shows the processing flow in QANDA, which uses a hierarchy of answers (e.g., entity answers could be people, organisations, or locations, locations can be countries, cities, etc.) to facilitate answering.

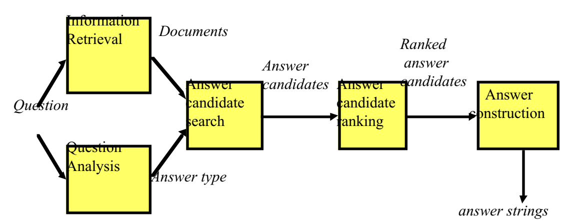


Figure 12. QANDA

In the TREC evaluation series in the new question and answering track involving 25 participating systems, QANDA was given 198 questions which it had to find answers to from an approximately _ million document store (2 Gbytes). Systems in this competition have to return five answers and provide short and long answer (50-byte and 250-byte). The top system in the competition got 56% right in the first place and 66% in the top 5.

LESSONS LEARNED

Our research in building intelligent multimedia information access systems over the past years has yielded the following key learnings. These are listed below are five directives which follow the basic scientific method.

1. *Collect, annotate, and learn from thy corpus*

   Generalizations of data, user, and/or system behavior should be based on large amounts of statistically significant data.

2. *Instrument and mine thy application*

   With users increasingly learning, working and playing in digital environments, fine grained instrumentation of applications can give keen insight into the utility, effectiveness, and enjoyability for system developers as well as users themselves.

3. *Extract thy (multimedia, multilingual, multiparty) content*

   Use powerful, off-the-shelf statistical packages and data mining methods (e.g., for sequence and market basket analysis) and machine learning techniques to discover knowledge from the data. Consider multimedia, multilingual or multiparty sources.

4. *Adapt thy algorithms, content and presentation*

   Leveraging the data previously collected and apply machine learning techniques (e.g., hidden Markov models, neural networks, case based reasoning) to robustly discover models of phenomena within the data in such a manner to enable the automated adaptation of those models to changing and/or new data sources, phenomena, and/or users.

5. *Evaluate thy task ... then do it again.*

   Commit to task based, community wide evaluation that supports the shared collection, annotation and application of data sets. Share system innovations with peers on a regular basis. If possible, share your tools, perhaps via mechanisms such as open source. This in turn will promote repeatable results and more efficient and effective learnings. Repeat the above steps while increasing the scope and complexity of the data and tasks.

REFERENCES

Bayer, S., Maybury, M. T. and Linton, F. 1999. Corpus Based Human Computer Interaction. Proceedings of 8th International Conference on Human-Computer Interaction (HCI International '99), August 22-27, 1999, Munich, Germany. pp. 922-926. http://www.mitre.org/support/papers/tech_papers99_

Boykin, S. and Merlino, M. Feb. 2000. A. Machine learning of event segmentation for news on demand. *Communications of the ACM*. Vol 43(2): 35-41.

Breck, E., Burger, J. D., Ferro, L., House, D., Light, M., Mani, I. 2000. A Sys called Qanda. In Vorhees, E.. and Harman, D. The Eighth Text Retrieval Conference, NIST Special Pubs. February 2000.

D'Amore, R. and Konchady, M. 1999. Cooperative Searching on the Internet. Proceedings of WebNet 99 - World Conference on the WWW and Internet, Honolulu, Hawaii, USA, October 24-30, 1999.

Fenn, J. 1999. Skill Mining, An Emerging KM Technology, Gartner Group T-07-4233.

Kobsa. A. August 30, 2000. User Modeling, Privacy, and Security. Invited talk. First International Conference on Adaptive Hypertext, Trent, Italy.

Linton, F., Joy, D., & Schaefer, H-P. 1999. Building User and Expert Models by Long-Term Observation of Application Usage. In J. Kay (Ed.), UM99: User Modeling: Proceedings of the Seventh International Conference (pp. 129-138). New York: Springer Verlag. [Selected data are accessible from an archive on http://zeus.gmd.de/ml4um/]

Linton, F., Joy, D., Schaefer, H-P., & Charron, A. 2000. OWL: A Recommender System for Organization-Wide Learning. Educational Technology and Society. http://ifets.ieee.org/periodical/

Mattox, D., Maybury, M., and Morey, D. 1999. Enterprise Expert and Knowledge Discovery. Proceedings of 8th International Conference on Human-Computer Interaction (HCI International '99), August 22-27, 1999, Munich, Germany. 303-307.

Maybury, M. Feb. 2000. News on demand: Introduction. *Communications of the ACM*. Vol 43(2): 32-34.

Maybury, M., Merlino, A., and Morey, D. 1997. Broadcast News Navigation using Story Segments, ACM International Multimedia Conference, Seattle, WA, November 8-14, 381-391.

Merlino, A. and Maybury, M. 1999. An Empirical Study of the Optimal Presentation of Multimedia Summaries of Broadcast News. Mani, I. and Maybury, M. (eds.) *Automated Text Summarization*, MIT Press.

ACKNOWLEDGEMENTS

I am indebted to Lynette Hirschman at MITRE for the original idea and version Figure 1 of the ask questions/get answers vision, extended here to suggest extraction also of multimedia and possibly multiparty (e.g., collaborative filtering) sources. Communicator is led by Sam Bayer and Lynette Hirschman. David Day and Marc Vilain lead the Alembic and Alembic Workbench research at MITRE. Frank Linton is the originator of OWL, IUP, and IURS. Ray D'Amore et al. created the Scout system. David House, Inderjeet Mani, and Chris Elsaesser created ExpertFinder and Tim Frangioso performed the KEAN usage studies. QANDA was created principally by Marc Light and Eric Breck. Stanley Boykin and Warren Greif lead the BNN project.