Knowledge Management and Speech Recognition

by James Allan

Knowledge Management (KM) generally refers to techniques that allow an organization to capture information and practices of its members and customers, in order that the stored knowledge can be readily re-used later. This article is about how speech recognition technologies are related to knowledge management, and about the likely impact those technologies might have on KM. I will first describe what "knowledge management" means in this context, including listing several KM applications that might be impacted by speech recognition. I will argue that speech recognition will currently be most useful when the items being processed are not too short, and will highlight several of the open problems that remain—not the least of which is improving the quality of speech recognition for telephone conversations! I will conclude by briefly describing some KM-related technologies where speech recognition has been successful. The message I hope to convey is that speech recognition has good prospects in "KM as information technology," but that there is sufficient weakness in the recognition technology to view those prospects cautiously.

What is Knowledge Management?

Knowledge Management (KM) has been a major buzzword in the information and business communities over the last several years, to the extent that it has perhaps lost much of its meaning. Broadly speaking, KM is "[a] business process that formalizes management and leverage of a firm's intellectual assets" (Gartner 1999). Intellectual assets can be either information assets such as documents or the knowledge in people that is not stored elsewhere. This distinction between "information assets" and other knowledge can also be characterized as between objects and process (Allee 2001, Sveiby 2001).

In this article, I will focus on KM as being management of information "objects" rather than the more complex (and less well understood, or at least less well storable on a computer) management of process. This view of KM may be somewhat restrictive (because so much is omitted), but it fits a definition of KM that has been common in the information retrieval community, particularly during the late 1990's when KM and Internet companies were very hot topics. "Knowledge" software has been and is often a wrapper around information capture, indexing, and retrieval, with some careful profile crafting—either manually or automatically—to get the right information to users. This view is evidenced by old press releases (Verity 1997) and by current product announcements (Insight Technologies 2001).

Of course, this view of KM does not mean that people disappear from the picture. However, rather than explicitly trying to capture the human intelligence part of KM (Hildebrand 1999) people are likely to be represented by the documents that they create, read, and save. It is arguable that this is merely the first stage of KM, and that the interesting next steps involve incorporating the ideas from KM into people's workflow so that knowledge capture is a natural part of the process (Davenport 1999). However, providing capabilities for storing and accessing objects efficiently and meaningfully is still an important component of KM.

What are KM applications?

Given the definition of Knowledge Management above, as essentially information technology, what are the sorts of applications that fall into KM? The simplest capability is the storage and retrieval of documents in a wide variety of formats (this is, for example, a service that Lotus Notes provides). Other applications include:

- The creation of "agents" that can monitor information sources for items that are of interest to an individual or a group. These agents essentially provide a customizable query that indicates what type of documents should be retrieved as they pass by.

- "Indexing" of people by the documents they create and store, or by self-generated descriptions of their interests. Such a capability allows an organization to locate expertise on a particular topic rapidly.
- Representation of situations or cases by the documents and people that are associated with them. This type of indexing admits the possibility of finding documents or people that can help address a new situation that is similar to one in the past.
- Tracking the information flow within an organization. Observing where new information enters an organization and how it moves may highlight individuals or departments that provide important information to the rest of the organization, or may illuminate other information-based social structures for sharing knowledge.
- Automatic extraction of small pieces of information from arriving data. This application could include mining Web pages for price information as part of competitive analysis. The company WhizBang! Labs (whizbanglabs.com) is currently providing this type of service (their "job opportunity" Web page at flipdog.com is a public example of what they do).

Note that many of these applications are readily available in some limited form in commercial products today. However, *all* of them remain active areas of research in the community, and there is substantial room for improvement.

When is speech recognition appropriate for KM?

The point of this article is to talk about speech recognition as applied to knowledge management. The KM applications listed above seem to deal with text but they can just as easily be made to work with "documents" that were created by someone speaking. For example, an important part of a "situation" might be transcripts of related meetings. If the meetings were recorded, the *actual* discussion could be used for automatic indexing and retrieval, rather than relying on rapidly-created minutes that attempt to summarize the key points.

Speech could also play an important role by incorporating dictated information. "Knowledge workers" who do not have time to key in useful information might be willing to record it as they are on the move (Barth 1999). The resulting transcripts could be indexed as retrieval documents.

Similarly, automatically generated transcripts of pertinent telephone conversations might also capture valuable knowledge for the future. Indeed, telephone calls may be a major opportunity for capturing information that could contain knowledge. "Most customers still deal with your company through a technology that hasn't changed much in a century: the telephone" (Sherman 2001). Since voice will always be a natural way for people to interact—both within and outside an organization—there is potential value in capturing that information, too.

Of course, speech recognition could also be used just to help someone interact with on-line information, whether from a telephone, a cell phone, or a desktop computer. Operating systems are now being shipped with small vocabulary voice recognition systems to allow some tasks to be completed by speech (e.g., "close this window"). This sort of functionality is not part of KM itself, but does provide an important entry point into the stored information.

What are the technical challenges and open problems?

It is clear, as outlined above, that spoken information has the potential to be at least as valuable as written material. One way to accomplish that would be to use an automatic speech recognition system to automatically transcribe all captured speech. It could then be treated identically to written information.

Will that work? Speech recognition systems are not perfect. High quality speech recordings (e.g., read by an announcer in a studio) might have a recognition error rate of under 10%. That means (roughly speaking) that close to one in a dozen words will be incorrectly recognized! Conversational speech, particularly on a telephone, will have error rates in the 30-40% range (Padmanabhan et al. 2001; Fiscus et al. 2000), probably on the high end of that in general.

Losing over a third of the words in a text seems like a showstopper. Fortunately, current indexing and retrieval technologies are very robust in the face of speech recognition errors (Allan 2001, Garafolo et al. 2000). Even in the presence of a 40% recognition error rate, a typical document retrieval system experiences a drop of only 10% in effectiveness. It appears that this pleasant result comes from several effects (Allan 2001):

- If spoken texts are long enough, then most important words are repeated (consider how often "knowledge" is repeated in this article). If they are not recognized one time, they will probably be recognized on some of the other occasions. This redundancy provides some resilience to recognition errors.
- Even if critical words are sometimes missed altogether, other strongly related words will usually be recognized. For example, if the word "speech" were never recognized in this article, the words "recognition," "voice," "spoken," and so on might be properly recognized. Those words would provide sufficient context that missing one key word would not be fatal.
- A large number of the unrecognized words are not content bearing. If "the" is omitted from a transcript, it is an error. Words such as "the" are not content bearing, and their absence—although contributing to the error rate—is not important for information storage and retrieval applications.

The implication of those items is that any information technology—hence, any aspect of KM—that deals with reasonably sized texts (a hundred words or more) is likely to be successful whether those texts were written or the result of a speech recognition system.

Of course, even with those points, a sufficiently large speech recognition error rate will prevent some spoken documents from being found. Fortunately, it seems that such an occurrence is rare enough not to be a large problem.

Some areas of success

What are some areas of KM (from the information technology perspective) that have succeeded in the presence of speech recognition errors? I will briefly discuss two such areas: document retrieval and organization and topic detection and tracking.

Document retrieval and organization. Automatic indexing and retrieval of documents is fundamentally based on the idea that documents using the same vocabulary talk about the same topics. More sophisticated systems provide elaborate query processing techniques to make it more likely that a system will find documents that are relevant to a query. Document comparison and grouping functions rely largely upon overlapping words for the core of their success. The central role that word overlap plays explains why it was originally felt that speech recognition errors would introduce a huge monkey wrench into the problem. However, for the reasons outlined above—i.e., redundancy in text—spoken documents have not been a problem. There remains some question of how well a system will do when it integrates both spoken *and* written documents into a single setting. Anecdotal evidence suggests that written documents will be more likely to be retrieved than will spoken documents (because the important words appear to occur more frequently in them since they were not misrecognized).

Topic detection and tracking. Another technology where spoken documents do not seem to be causing problems is that of automatically organizing news stories by the events that they describe. This set of tasks, investigated under the name of Topic Detection and Tracking (TDT), focuses much of its attention on television and radio news where speech recognition is the only way to acquire text transcripts (except in the case of some television that has closed captioning). TDT is a research program and is not integrated into any known KM systems, but its goals—tracking topics in the news, identifying the onset of new topics as rapidly as possible, and gathering news into groups of stories by their underlying topics—are all of great potential interest in many KM settings. The TDT research has demonstrated little impact on effectiveness when recognized speech is used as source material instead of written materials. As with document retrieval, TDT's robustness is most likely the result of there being fairly large stretches of speech being compared (i.e., a full news story).

Problems remaining

On the other hand, some KM applications could suffer more from recognition errors. When the span of recorded speech is short, the redundancy sort of effects listed above will not occur. For example, information retrieval systems are much more sensitive to recognition errors when the queries are spoken than when the documents are speech recognition output (Allan 2001).

The problem of mining text for small pieces of information to fill a database (e.g., the job listing work of WhizBang! mentioned above) is also likely to be less robust. Some related research toward just finding names of people, places, and organizations shows that at a recognition error rate of 40%, the error rate for finding such named entities was approximately 80% (Goel and Byrne 1999). Other results suggest that the problem may not be quite so severe (Kubala et al. 1998), but it is clear that finer-grained text analysis tasks are much more strongly impacted by recognition errors.

The discussion above makes it clear that better speech recognition of telephone and conversational speech is an important open problem for KM. But improved analysis techniques that are more robust in the presence of those errors are also needed.

Conclusion

If knowledge management is limited to the capture, indexing, and use of information, then speech recognition has an important role to play. A substantial amount of information is conveyed in meetings, on the telephone, or in casual settings. If that information can be captured and stored, it may provide access to much more of the underlying knowledge that an organization wants to preserve. The results of several document-based tasks when applied to spoken documents indicate that there is good potential for success.

There are, of course, huge problems that arise from such capture. All that additional information does not mean that the knowledge is readily available; KM requires techniques for *extracting* knowledge from the data. In addition, on a more social than technological side, the more information that is captured, the more privacy issues will begin to loom large and will have to be handled. Social problems can probably be dealt with (e.g., providing an "off the record" mode) if there is enough value in the captured information. Whether such value can be found remains to be seen.

References

All on-line references are known to be available as of October 15, 2001.

Allan, J. 2001. Perspectives on Information Retrieval and Speech. *Lecture Notes on Computer Science*. Springer Verlag. Forthcoming.

Allee, Verna. 2001. Knowledge Management. At Verna Allee's Web site: http://www.vernaallee.com/home/Knowledge%20Management.html

Barth, Steve. 1999. Tell It to the Machine: A new generation of recorders and software make it easier to digitize knowledge [online]. In *Knowledge Management* [online], as presented at Barthy's Web site: http://www.global-insight.com/KM-VR.htm

Davenport, Tom. 1999. Knowledge Management, Round Two [online]. From *CIO Magazine*, Think Tank, edited by Elaine M. Cummings, available at the Web site: http://www.cio.com/archive/110199_think_content.html

Fiscus, J.G., W.M. Fisher, A.F. Martin, M.A. Przybocki, and D.S. Pallett. 2000. 2000 NIST Evaluation of Conversational Speech Recognition Over the Telephone. *2000 Speech Transcription Workshop*, proceedings available on-line at http://www.nist.gov/speech/publications/tw00/html/abstract.htm#cts-10

Garafolo, J.S., C.G.P. Auzanne, and E.M. Voorhees. 2000. The TREC Spoken Document Retrieval Track: A Success Story. *Proc. of TREC-8 (1999)*. NIST special publication 500-246.

Gartner. 2001. Customer Relations Management Glossary [online]. From Gartner Interactive Web site: http://www.gartnerweb.com/public/static/crm/crm_glossary.html

Goel, V. and W. Byrne. 1999. Task Dependent Loss Functions in Speech Recognition: Application to Named Entity Extraction. *Workshop on Accessing Information in Spoken Audio*. Cambridge, U.K. Available on-line at http://www.clsp.jhu.edu/publications/pubs/CUWRKSHP99.pdf

Hildebrand, Carol. 1999. Does KM=IT? *CIO Enterprise Magazine* available at CIO.com's Web site: http://www.cio.com/archive/enterprise/091599_ic.html

Kubala, F., R. Schwartz, R. Stone, and R. Weischedel. 1998. Named Entity Extraction from Speech. *Proc. DARPA Broadcast News Transcription and Understanding Workshop*. Lansdowne, Virginia. Available on-line at http://www.nist.gov/speech/publications/darpa98/html/lm50/lm50.htm

Insight Technologies. 2001. Knowledge Insight version 5 [online]. From Insight Technologies Ltd's Web site: http://www.insighttechnologies.com/products/ki_v5.asp.

Padmanabhan, M., G. Saon, G. Zweig, J. Huang, B. Kingsbury and L. Mangu. 2001. Evolution of the performance of automatic speech recognition algorithms in transcribing conversational telephone speech. *IEEE Instrumentation and Measurement Technology Conf. (IMTC)*, IEEE.

Pallett, D., J.G. Fiscus, J.S. Garofolo, A. Martin, and M. Przybocki. 1999. 1998 Broadcast News Benchmark Test Results. *Proc. 1999 DARPA Broadcast NewsWorkshop*. Available on-line at http://www.nist.gov/speech/publications/darpa99/pdf/ov10.pdf

Sherman, Lee. 2001. Reality Check [online]. In *Knowledge Management* [online], from Destination CRM Web site: http://www.destinationcrm.com/km/dcrm_km_article.asp?id=973

Sveiby, Karl-Erik. 2001. What is Knowledge Management. From Sveiby's Web site: http://www.sveiby.com.au:80/KnowledgeManagement.html

Verity. 1997. Verity Ships IntelliServ, An Advanced Knowledge Management Application For Windows NT [online]. From Verity Inc.'s press release Web site: http://www.verity.com/press/1997/970723.html