International Journal of Transformations in Business Management http://www.ijtbm.com

(IJTBM) 2014, Vol. No. 4, Issue No. III, Jul-Sep **ISSN: 2231-6868**

KNOWLEDGE MANAGEMENT WITH INNOVATIVE INTELLIGENT SYSTEMS OF HUMAN LANGUAGE TECHNOLOGY

***Prashant Vasudev, #Ashlyn Antony, @Kumar Siddhartha**

**Vice Principal, Manav Rachna School*
#BDO, Thmopson Reuters
@Former HOD, R D Engg College

ABSTRACT

In the past few years, knowledge management has received increasing attention from industry, academia, and government. Effective KM is often cited as a key capability for gaining a competitive advantage in global enterprises, and human language technology plays a central role in KM. This paper is based upon the interaction between human language technology and KM to achieve this goal. This paper discusses many techniques that have just found— or are currently finding—their way into KM applications. Elaborates on the challenges KM faces regarding information extraction and some of the techniques being explored. *However, they do need solid human language technology blocks on which to build. Second, human language technology should not presuppose any extremely specific input. Rather, it must adapt to text and speech input types and to its users in a flexible and efficient way. Although many of the building blocks of human language technology are ready to use, others still must find their way into overarching KM systems.*

INTRODUCTION

Knowledge management has changed the way we look at knowledge in the current economy; it is a key factor in an enterprise's success or failure. In contrast to what we as engineers typically love, KM puts people first, organizational issues second, and technology third. Let us explain this triad.

Company leaders want their employees to share knowledge in a way that supports creativity and increases profitability. Thus, the employees must quickly recognize KM's benefits, understanding its ability to offer more capabilities and less work and to better benefit the company. If employees don't recognize KM's value, they won't engage in the KM undertaking, possibly causing it to fail.

Organizational issues comprise several dimensions germane to the organization's objectives. The company's goal is to succeed in the market—otherwise, it wouldn't exist. It must deal with its value chain, taking care of core processes, core products or services, and its customers. KM that does not deal with these issues fails to contribute to the organization's success.

Finally, technology can rescue people from boring, redundant tasks. The question KM asks is, "What are these tasks?" Consider current KM systems—typically knowledge repositories— and how they carry knowledge and the tools they use to access it. These systems easily fall into one of two extremes.

Systems at the one extreme tend toward *rigid substance*— databases or knowledge bases in which algebras or logics carry the meaning of data bits and pieces. The corresponding tools are carefully crafted but for a surgeon rather than for a layman. Such systems are valuable but extremely hard to build and maintain. They are easy to destroy, and it isn't always easy to find knowledge in them or reuse knowledge from them. At the other extreme are systems that lean toward *fluid substance* —text content that cannot be gripped unless with a large bucket (the document), making it impossible to find the nuggets of valuable knowledge. Thus, current systems are bound to fail in many practical applications. Seriously considering these issues to produce a successful KM system leads to at least three requirements. We must

- Encourage employees to participate,
- Integrate KM with current organizational practice, and
- Provide the natural tools such that people can easily recognize the benefits, align with current organizational practices, and use the system.

KM CHALLENGES

KM enhances organizational performance through organizational knowledge sharing, learning, and application of expertise. Indicating KM's importance, many corporations that traditionally measured only the financial aspects of value are beginning to measure human and intellectual value as well.

A range of human language technologies can enable KM, including enhanced information retrieval, extraction, summarization, presentation, and generation. Moreover, human language technologies promise to enhance human access to information and human interaction by increasing our awareness of knowledge artifacts or activities intersecting our interests. Key KM elements include mapping existing knowledge, discovering expertise, and discovering new knowledge.

KNOWLEDGE MAPPING

A primary issue for many organizations is recognizing what they know. Even providing easy access to explicitly captured knowledge in artifacts such as written policies, strategies, documents, and presentations can provide individuals in organizations with tremendous power and efficiency. Often, however, an organization creates so much material that effectively organizing it is a daunting task.

We need tools that can automatically generate classifications or taxonomies of explicit corporate and world knowledge. The success of services such as Yahoo, Northern Light, and Quiver illustrate the value (and limitations) of current classification-based collections and collaborative filtering approaches.

EXPERT AND COMMUNITY DISCOVERY

Knowing whom to call, who knows a key fact, or who has the know-how or skill to analyze, diagnose, or recommend solutions in a particular domain is a challenge. Manually created corporate-skills databases are costly and inconsistent across individuals and disciplines, and they quickly become obsolete. Finding experts or communities of experts rapidly can be a competitive advantage for a company. Indeed, it is an important function of new business models of virtual corporations.

KNOWLEDGE DISCOVERY

Having end users or communities of experts (individually or collectively) on hand to answer questions is advantageous. They can accomplish this through their own expertise, machine learning, or data mining, ultimately learning new knowledge, including *ontology induction* (learning new classes of knowledge and meaning representations).

HUMAN LANGUAGE TECHNOLOGY FOR KM

Having considered KM needs and some preliminary promise of human language technology to provide solutions to those needs, let's look at how human language technology can contribute to KM. A range of functional areas of human language technology that offer potential solutions to some required KM elements are outlined.

INPUT ANALYSIS

Analyzing user-spoken language and natural input is key to knowledge access. This is essential for applications such as natural language interfaces to databases, question and answering, and multimedia interfaces. In today's conventional interfaces, users are allowed to sequentially input mouse, keyboard, and speech input and perform limited natural language processing— for example, stemming, morphological analysis, and query expansion.

Challenges include dealing with imprecise, ambiguous, or partial input. Addressing these issues in multimodal (that is, text, speech, or gesture) and multiplatform (desktop, kiosk, or mobile) interfaces provides additional challenges, including the need to use potentially uncertain inputs from individual recognizers.

Input mechanisms that are intuitive and can adapt to different users and situations (or automatically adapt) promise to mitigate access complexity and user training, increasing broad availability of knowledge access.

RETRIEVAL

Retrieval technology has achieved approximately 80 percent precision with low recall (or 80 percent recall with low precision); by using relevance feedback, systems approach human-crafted queries.

The ability to leverage the advances in input processing—especially query processing— together with advances in content based access to multimedia artifacts (text, audio, imagery, and video) promises to enhance the richness and breadth of accessible material while improving retrieval precision and recall. Dealing with noisy, large scale and multimedia data from sources as diverse as radio, television, documents, Web pages, and human conversations will also offer challenges, but advances in this area will enhance document retrieval precision and recall, ease navigational burden for users, and reduce search time.

EXTRACTION

Extraction is the ability to identify and cull objects and events from multimedia sources. We currently can achieve 90 percent precision and recall when extracting named entities (people, organizations, or locations), 70 percent for relations among named entities (such as "father-of"), and 60 percent for events from text.

One challenge includes extracting entities within media and correlating those across it. This might include extracting names or locations from written or spoken sources and correlating them with associated elements within images. Whereas commercial products exist to extract named entities from text with precision and recall in the 90th percentile, domain-independent event extractors work at best in the 50^{th} percentile, and performance degrades further with noisy, corrupted, or idiosyncratic data. Achieving better extraction will provide direct access to information or knowledge elements—including specific types that might be user preferred—and will let us reuse media elements, enabling user tailored selection or presentations.

QUESTION ANSWERING

The single best performing system today provides approximately 75 percent precision and recall for a small question-and-answer corpus. Drawing on techniques from query processing, retrieval, and presentation, this important new class of systems is moving us from our current form of Web searches (type in keywords to retrieve documents) to more direct natural language searches that are answered directly by an answer extracted from the source. Challenges will include question analysis, response discovery, source selection, multi-perspectives, source segmentation, extraction, and semantic integration across heterogeneous sources of unstructured, structured, and semi-structured data. Eventually, by providing direct answers to questions, we'll be able to overcome the time, memory, and attention limitations currently required to shift through many returned Web pages.

TRANSLATION

Last year, for the first time, English was estimated as constituting less than half the material on the Web. Given that information will increasingly appear in foreign languages, there will be a need for systems to gist or skim content for relevance assessment—beyond the current ability to translate approximately 40 languages. We also need to improve current human-assisted machine translation to provide higher-quality translation for deeper understanding.

New innovative applications include the translation of multilingual conversations, rapid creation of translingual corpora, and effective translingual retrieval, summarization, and translation. Other applications will involve verbalizing graphics and visualizing text. Cross-media or cross-mode information and knowledge access will enable broader access to global information sources using methods such as translingual information retrieval.

DIALOGUE MANAGEMENT

We currently can perform simple fact seeking dialogues in specific domains such as weather, travel planning, and inventory. However, knowledge workers will require systems that can support natural, mixed initiative human–computer interaction that deals robustly with context shift, interruptions, feedback, and shift of locus of control.

Open research challenges include tailoring the flow and control of interactions and facilitating interactions such as error detection and correction tailored to individual physical, perceptual, and cognitive differences.

Motivational and engaging lifelike agents also offer promising opportunities for innovation.

AGENT AND USER MODELING

Computers can construct models of user beliefs, goals, and plans. They can also model users' individual and collective skills by processing materials such as documents or user interactions and conversations. While raising important privacy issues, unobtrusively modeling users (or groups of users) from public materials or conversations can enable a range of important KM capabilities. For example, this might include expertise databases that can enhance organizational awareness and efficiency or track user characteristics, skills, and goals to increase interaction and help users or agents find experts.

SUMMARIZATION

Summarization aims to select content and condense it to present a compact form of the original source. Summaries can compress their content by 50 percent without losing information and can contain extracted information from or an abstract of original source material. They can be informative, indicative, or evaluative, and they offer knowledge workers access to larger amounts of material with less required reading.

Some related challenges include multimedia, multilingual, and cross-document summarization.

Addressing scale ability to large collections and user- or purpose-tailored summaries is also an active research area.

PRESENTATION

Effective presentations require selecting appropriate content, allocating content to appropriate media, and ensuring fine-grained coordination and realization in time and space. Discovering and presenting knowledge might require mixed media and mixed mode displays tailored to the user and context. This could include tailoring content and form to the user's specific physical, perceptual, or cognitive characteristics. It also might lead to new visualization and browsing paradigms for massive multimedia and multilingual repositories that reduce cognitive load or task time, increase analytic depth and breadth, or simply increase user satisfaction. We currently have highly complex systems with typically knowledge-rich methods of presentation planning and realization, but a grand challenge is to automatically generate coordinated speech, natural language, gestures, animation, and non speech audio, possibly delivered through interactive, animated, lifelike agents. Preliminary experiments suggest that independent of task performance, agents might be more engaging to younger or less experienced users.6

AWARENESS AND COLLABORATION

Our global Web provides unprecedented opportunity for worldwide collaboration, both asynchronously and synchronously. Users can use instant messaging and interact in place-based collaboration environments, but in the future they will need enhanced awareness of both emerging knowledge and one another's expertise. One such aid is detection and tracking of topics of interest to facilitate discovery and connection among communities. Another is creating expertise profiles based on publicly available information such as publications, interviews, or public conversations.

Human language technology promises to deliver great value to the KM challenge, but we need to address many fundamental scientific and technical challenges to ensure that this value accrues:

- *Heterogeneity*—dealing with the diverse nature of human language artifacts, both in form and semantic content.
- *Scalability*—addressing the size of corporate collections and global content.
- *Portability*—creating adaptive methods (such as corpus-based machine-learning approaches) that enable rapid retargeting of algorithms to new languages and media.
- *Complexity*—ensuring that the many content forms and presentational methods and devices do not overwhelm end users.
- *Security*—ensuring authentication to control access to source materials or ensuring the identity and integrity of source materials.
- *Privacy*—addressing the legal and social issues of maintaining privacy and control of a user's model extracted from public materials or interactions.

Overcoming these human language technology challenges is essential for KM to advance.

CHALLENGES IN INFORMATION EXTRACTION FROM TEXT FOR KNOWLEDGE MANAGEMENT

Nowadays, most knowledge is stored in an unstructured textual format. We can't query it in simple ways, thus automatic systems can't use the contained knowledge and humans can't easily manage it. The traditional knowledge management process for knowledge engineers has been to manually identify and extract knowledge—a complex and time-consuming process that requires a great deal of manual input. As an example, consider the collection of interviews with experts (*protocols*) and their analysis by knowledge engineers to codify, model, and extract the knowledge of an expert in a particular domain. In this context, *information extraction* from texts is one of the most promising areas of human language technology for KM applications.

INFORMATION EXTRACTION

IE is an automatic method for locating important facts in electronic documents—for example, information highlighting for enriching a document or storing information for further use (such as populating an ontology with instances). IE thus offers the perfect support for knowledge identification and extraction, because it can, for example, provide support in protocol analysis in either an automatic (unsupervised extraction of information) or semiautomatic way (helping knowledge engineers locate the important facts in protocols through information highlighting).

It is widely agreed that the main barrier to using IE is the difficulty in adapting IE systems to new scenarios and tasks. Most of the current technology still requires the intervention of IE experts. This makes IE difficult to apply, because personnel skilled in IE are difficult to find in industry, especially in small-to-medium-size enterprises. A main challenge is to enable personnel with knowledge of AI (for example, knowledge engineers) who have no or scarce preparation in IE and computational linguistics to build new applications and cover new domains. This is particularly important for KM. IE is just one of the many technologies for building complex applications: wider acceptance of IE will come only when IE tools don't require any specific skill apart from notions of KM.

Several machine learning-based tools and methodologies are emerging, but the road to fully adaptable and effective IE systems is still long. Here, I focus on two main challenges for IE adaptivity in KM that are paramount in the current scenario: automatic adaptation to different text types and human centered issues in coping with real users.

ADAPTIVITY TO TEXT TYPES

Porting IE systems means coping with four (often overlapping) main tasks:

1. *Adapting to the new domain information*— implementing system resources such as lexica, knowledge bases, and so forth, and designing new templates so that the system can manipulate domain-specific concepts.
2. *Adapting to different sublanguage features*—modifying grammars and lexica to enable the system to cope with specific linguistic constructions that are typical of the application or domain.
3. *Adapting to different text genres*—specific text genres such as medical abstracts, scientific papers, and police reports might have their own lexis, grammar, and discourse structure.

4. *Adapting to different types*—Web based documents can radically differ from newspaper-like texts. We need to be able to adapt to different situations.

Most of the literature on IE has focused on issues 1, 2, and 3, with limited attention to text types, focusing mainly on free newspaper-like texts. This is a serious limitation for portability, especially for KM, where an increase in the use of Internet and intranet technologies has moved the focus from free-texts-only scenarios (based on, for example, reports and protocols) to more composite scenarios including semistructured and structured texts such as highly structured Web pages produced by databases. In classical natural language processing (NLP), adapting to new text types is generally considered a task of porting across different types of free texts.

QUESTION-ANSWERING SYSTEMS IN KNOWLEDGE MANAGEMENT

At no time in history have so many people been able to access so much information.

Businesses rely on unstructured information available over the Internet, intranets, email, press releases, online newspapers, digital libraries, and other sources. Companies accumulate large quantities of written information with customer comments, trade publications, internal reports, competitor Web sites, and much more. Making sense of all this information and leveraging its advantages is crucial. New companies were formed with the purpose of helping other companies cope with this huge volume of information. These knowledge-supporting companies provide skills for integrating document management, workflow, workgroups, intranets, and knowledge portals. They might be general purpose or industry specialized, serving vertical markets such as financial, IT, telecommunications, travel, or media. Other companies have extended or created their own internal information management departments.

Yet, the technology to access and use information has dramatically lagged behind its growth rate. People have questions, and they need answers. Current Internet search engines let us locate documents that might have relevant information, but often the documents returned are too numerous to inspect or the answer simply isn't there. This has motivated the renewed interest in question-answering technology and natural language processing. The task in QA is to find correct answers to open-domain questions expressed in English or other natural languages by searching large collections of documents. These documents can come from a text collection, the Web, databases, digital libraries, or any other electronic source.

QA TECHNOLOGY'S POTENTIAL

QA technology will undoubtedly play a major role in knowledge management. Users can include casual questioners who ask simple factual questions; consumers who look for specific product features or prices; research analysts who are in the business of collecting specific information about market, competitors or business events; or professional information analysts such as police detectives, law enforcement officials, financial analysts, or intelligence analysts.

The main difficulty QA technologists face is the broad range of questions a system must answer. Simple questions are relatively easy to answer. These are often the *Who*, *When*, *Where* types of questions:

• Who was the first American in space?

• When did the Neanderthal man live?

• Where is John Wayne airport?

Other questions require reasoning on knowledge bases, collecting pieces of evidence from multiple documents, and then combining together the answer, or other advanced AI techniques. Examples include (where A and B can be concepts or relations):

• How is A related to B?

• What causes A?

• What are the effects of A?

• What can damage A?

• How do you prevent A?

Even more difficult questions exist, which require considerable world knowledge and powerful reasoning capability. For example, to answer, "How likely is it that the federal government will lower interest rates next month?" the system would first have to find out what usually influences the government's decision and then compare the status of such parameters with previous situations for which the outcome is known.

QA technology alone is not enough to provide solutions to information-management problems. It is merely a powerful tool that, when embedded into other larger software systems currently used for information management, can let companies deliver fast, effective, and affordable results. It also lets companies analyze textual documents, extract desired information, rank and link important concepts, and personalize information.

Furthermore, it might help with automation within specific department, marketing, or customer support. Marketing departments might use QA to learn trends, discover customer interests, or identify customers most likely to buy specific products and services. Customer-support departments might use QA to give consumers better and faster product information and services, browse through technical information, or extract customer profiles.

TECHNICAL CHALLENGES

Open-domain QA is a complex task that encompasses many natural language processing, information retrieval, and AI techniques. The inherent difficulties in finding answers to open-domain questions pose many serious technical challenges. The Roadmap Research in Question Answering document has identified these challenges, some of the most important being

• understanding questions including ambiguities and implicatures,

• understanding questions and finding answers within a given context,

• extracting distributed answers that require answer fusion,

• providing answer justification and proof of correctness,

• offering interactive question answering,

• offering real-time question answering, and

• extracting answers from a wide range of document formats.

KNOWLEDGE MANAGEMENT AND MOBILE VOICE INTERFACES

Information technology has already made quick and easy access to remote data and information possible. An individual sitting in an office in front of a desktop PC can cross boundaries of space over a network of computers to retrieve data from another computer or communicate electronically with another person sitting in an office across the globe. At first, this possibility provided a quantitative improvement of information access, but it has also triggered qualitatively new patterns of work and knowledge exchange, as in virtual teams.

When the entry-points to the networks become mobile and independent of the office environment, spatial restrictions of data access relax. This enables mobile-data applications—for example, in the form of mobile-data services in the upcoming third generation telecommunication networks.

Existing and planned end-user terminals for mobile application scenarios vary in size and available input and output channels, ranging from notebooks, PDAs, smart phones, and WAP phones to plain telephones without display. In general, the smaller and simpler the terminal, the more impoverished the visual input and output.

Many small terminals also lack a comfortable keyboard for alphanumeric input. These restrictions of conventional visual interface techniques raise the attractiveness of voice communication—not only of person-to- person voice telephony, which will remain preeminent as a mobile communication type, but also of voice interfaces that use automatic speech recognition (ASR) or text-to-speech (TTS) to provide access to data applications.

We can use such voice-enabled mobile data applications even when our hands or eyes are busy—for example, when driving a car.

Mobile data services for use in cars (referred to as *telematics*) have emerged as one of the most promising markets for the wireless Internet. This is a real economic opportunity to apply speech technology, and general human language technology, to the potentially huge markets of telematics, mobile data applications, and 3G data services.

Some of the key technical issues for mobile voice interfaces have been resolved or are close to resolution. The mobile terminals do not need to support specialized voice processing beyond a standard voice call. A dedicated telephony server in the infrastructure picks up the voice call. In a speech-in, speech-out application, the telephony server sends the speech to an ASR server, which uses hidden Markov models to implement speaker-independent LVCSR (large vocabulary continuous speech recognition).

The results from ASR, usually in the format of ASCII text, are passed on for further linguistic processing.

In existing systems, this linguistic processing is often rudimentary. The input constitutes the user's choice from among a dozen or so options specified in a command grammar. Each recognized input text triggers a system response, which is either predefined text or a text template that contains dynamic information from a database query.

MULTIMODALITY

To avoid repeating previous mistakes made by some human language technology advocates, it is nevertheless important to stress that visual interfaces will also play an important role in mobile data access. Visual interfaces are easy to implement and familiar to users. In addition, for some tasks, they are better suited than voice—for example, to represent spatial information using images.

One promising route for research is therefore to develop interfaces that combine visual elements and voice into so-called multimodal interfaces. The two modalities can then be offered as alternatives, so that the user can choose between a visual or a voice mode, depending on the situation. A second mode of combining voice and visual is to have both modalities active concurrently, without explicit coordination. However, only when coordinated, simultaneous use of the voice and visual channels is supported is the potential for a multimodal interface exploited fully. For example, using a terminal that offers visual input through a touchscreen, a user could draw a number of irregular shapes around areas on a city map in a tourism application and say, “Tell me about hotels in these areas.” The system would resolve the reference of “these” to the set of selected areas in the present time interval. It might then insert icons for the selected hotels on the map and read through TTS the names and details of the hotels that pay a registration fee to the service provider.

While reading the data for a hotel, the corresponding icon could blink on the display. In contrast to voice-only interfaces, technologies for mobile multimodal interfaces to data services are still at the research stage.

VOICE AND KM

Due to robustness issues of voice recognition technology, voice interfaces should be used in contexts and applications where they have a chance to work well. Not all applications in IT-based knowledge management appear suitable. For example, voice interfaces to unrestricted search services are problematic because of the huge branching factor in recognition—that is, the number of possible alternatives when recognizing the search terms. Yet, I am convinced that voice-only and combined voice and visual interfaces to applications can be useful beyond “nice to have” in situations that favor them, especially in mobile situations and on small communication terminals. Putting these interfaces in place should be a concern for knowledge-intensive organizations with mobile workers.

Whether the resulting new types of information access will also give rise to qualitatively new types of knowledge exchange will be interesting to watch.

CONCLUSION

The natural choice of “substance” for such a KM system is human language, and the required tools are based on human language understanding. Motivating people to use human language for knowledge sharing is easy, because they are used to it. Similarly, aligning human language with many organizational practices is easy, because much of the latter come with documents written in or entangled with human language. Here is where reality strikes. As we all know, comprehensive human language understanding is out of reach for the foreseeable future. Nevertheless, although the knowledge system’s substance is language and complete human language understanding is out of reach, the system need not be restricted to text nor the tools restricted to a keyword-based search. Human language technology can and should do better. QA technology will undoubtedly play a major role in knowledge management. Users can include casual questioners who ask simple factual questions; consumers who look for specific product features or prices; research analysts who are in the business of collecting specific information about market, competitors or business events; or professional information analysts such as police detectives, law enforcement officials, financial analysts, or intelligence analysts. Several machine learning-based tools and methodologies are emerging, but the road to fully adaptable and effective IE systems is still long.

REFERENCES

1. K. Bontcheva et al., "Using HLT for Acquiring, Retrieving and Publishing Knowledge in AKT,"
2. M. Maybury, R. D'Amore, and D. House, "Automating the Finding of Experts," *Int'l J. Research Technology Management*, vol. 43, no. 6, Nov./Dec. 2006, pp. 12–15.
3. F. Ciravegna, "Adaptive Information Extraction from Text by Rule Induction and Generalisation," *Proc. 17th Int'l Joint Conf. Artificial Intelligence* (IJCAI 2007), Morgan Kaufmann, San Francisco, 2001, pp. 1241–1256.
4. S. Harabagiu et al., "FALCON: Boosting Knowledge for Answer Engines," *Proc. 9th Text Retrieval Conf.* (TREC-9), NIST, Gaithersburg, Md., 2009, pp. 50–59.
5. Douthat, *The Message Understanding Conference Scoring Software User's Manual*, www.itl.nist.gov/iaui/894.02/related_projects.
6. *"Auvo Technologies and Vodafone's AirTel Móvil to Trial Multi-Modal Wireless Internet Platform," AUVO, Itasca, III., 2008, www. auvo.com/frmset5.html*