

Crosslingual Language Technologies for Knowledge Creation and Knowledge Sharing

Hans Uszkoreit

DFKI Language Technology Lab
Stuhlsatzenhausweg 3
66123 Saarbruecken, Germany
uszkoreit@dfki.de
www.dfki.de/~hansu/

A large and fast growing part of corporate knowledge is encoded in electronic texts. Although digital information repositories are becoming truly multimedial, human language will remain the only medium for preserving and sharing complex concepts, experiences and ideas. It is also the only medium suited for expressing metainformation. For a human reader a text has a rich structure, for a data processing machine it is merely a string of symbols. Classical information retrieval helps to sort and find information in large libraries of documents by matching strings of characters. Effective information management is a building block of modern knowledge management. However, language technology can contribute much more than methods for finding information.

A number of areas in which language technologies can improve knowledge management are described in Maybury (in this volume). We will concentrate on examples in which language technologies can facilitate the creation of new knowledge from large volumes of textual information and the sharing of knowledge across language boundaries.

1 Knowledge Sharing

One of the true challenges of KM is the development and implementation of schemes that make people share knowledge and use such shared knowledge in critical situations. Offering incentives for the sharing of knowledge is not sufficient. The valuable information needs to be offered in situations where it is needed. It also needs to be evaluated in such situations because

any effective incentive scheme might lead to information overflow if the quality of the provided information cannot be assessed. Language technology can provide means for associating shared knowledge with the relevant decision situations by automatically linking it to the critical elements within decision triggers, i.e. electronic documents in the workflow that demand and record a decision.

Together with some simple statistical methods this method can also support a scheme for evaluating shared information with a minimum of additional effort. The language technology that can be applied for this purpose we call automatic relational hyperlinking. Relational hyperlinks differ from the simple hyperlinks of HTML in that they are composed out of a number of named links that can be selected from a menu.

Language technology is needed for identifying and disambiguating the concepts in documents that need to be linked. To this end, techniques from information extraction are employed such as named entity recognition. When automatic hyperlinking associates information to decision situations, an evaluation can be enforced without an additional burden on the user.

Automatic hyperlinking can also be applied for transforming information into knowledge-like structures. By densely interconnecting informational elements, three criteria are met that distinguish knowledge from other forms of information: immediate accessibility, grounding of pieces of knowledge and associative

structure. The important fourth criterion is the suitability for inferencing, however in this application scenario inferencing is not performed by the machine but by the human user of the service.

This method has been applied in the system Hypercode of the DFKI LT Lab. The original purpose of this system which was developed for a large German bank is to facilitate work with legacy code. Hypercode provides dense associative relational hyperlinking to program code and documentation. By densely interlinking code and documentation, the knowledge encoded in the documentation becomes much more accessible and usable. The methods of Hypercode were also applied for enriching a new WWW-based information service of the Saarland State Government for start-up companies.

2 Crosslingual Knowledge Management

Globalization forces companies to become multilingual. The language of customer interaction should be the preferred language of the customer. The language for knowledge sharing should be preferred language of the experts who voluntarily provide the knowledge. On the other hand, the language of knowledge sharing has to be a language that the potential users of the information understand. The languages of provider and users may differ. Moreover, in a multinational enterprise there may be user communities that extend across several native languages. Translation is costly and may delay the exploitation of shared knowledge. Automatic translation offers alternative solutions. Even the best machine translation systems cannot translate unseen texts without grammatical or stylistic errors. However, for the purpose of knowledge sharing often a so called content translation or an indicative translation will suffice. Such a translation can be provided by existing translation systems. Factual errors can be

avoided by augmenting the general purpose translation systems with specialized terminology and transfer rules. We will exemplify the utilization of specialized indicative machine translation for multilingual expert groups by a project for a large multinational automobile manufacturer.

Finally we will provide an overview of other crosslingual language technologies and their potential for crosslingual knowledge management. In this context, we will point to a number of European R&D projects in which consortia composed of academic and industrial partners improve or adapt language technologies such as information retrieval, information extraction and summarization for knowledge management applications in multilingual applications scenarios.

References

- Glushko, R. J. (1989): Transforming Text Into Hypertext For a Compact Disc Encyclopedia, In: Proceedings of CHI '89, ACM Press.
- Jacobs, P. (1997) Text Interpretation: Extracting Information. In: R.A. Cole, J. Mariani, H. Uszkoreit, A.Zaenen, V. Zue (eds.): Survey of the State of the Art in Human Language Technology, Cambridge University Press and Giardini.
- Piskorski, J. and G. Neumann (2000) An Intelligent Text Extraction and Navigation System. In: Proceedings of 6th International Conference on Computer-Assisted Information Retrieval (RIA0-2000), Paris.
- Pustejovsky, J., B. Boguraev, M. Verhagen, P. Buitelaar, and M. Johnston (1997): Semantic Indexing and Typed Hyperlinking. In: Proceedings of the American Association for Artificial Intelligence Conference, Spring Symposium, NLP for WWW.. Stanford University, CA, 120-128.