Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Technology 5 (2012) 75 – 82

Procedia Technology

CENTERIS 2012 - Conference on ENTERprise Information Systems / HCIST 2012 - International Conference on Health and Social Care Information Systems and Technologies

Multilingual manager: a new strategic role in organizations

Jesús Cardeñosa[a,*], Jessica Herrero[a], Francisco Gisbert[a]

[a]*Universidad Politécnica de Madrid, Facultad de Informática Campus de Montegancedo, Boadilla del Monte (Madrid) 28660, Spain.*

Abstract

Today's knowledge management (KM) systems seldom account for language management and, especially, multilingual information processing. Document management is one of the strongest components of KM systems. If these systems do not include a multilingual knowledge management policy, intranet searches, excessive document space occupancy and redundant information slow down what are the most effective processes in a single language environment. In this paper, we model information flow from the sources of knowledge to the persons/systems searching for specific information. Within this framework, we focus on the importance of multilingual information processing, which is a hugely complex component of modern organizations.

Keywords: Language Management; Knowledge Management; Language Engineering; Multilingual Information Management; Content Representation; Content Management.

1. Introduction

The European Union (EU) has classed language technologies as a strategic sector for the 21st

* Corresponding author: Jesús Cardeñosa. Tel.: +34-91-336.42.88; fax: +34-91-352.48.19.
E-mail address: carde@fi.upm.es.

This work was supported by the Ministerio de Ciencia e Innovacion of Spain under Project TEC2010-21303-C04-02.

doi:10.1016/j.protcy.2012.09.009

century. Indeed, the EU's 25 official languages, and dozens of other languages used by minorities that are, nonetheless, equally important on cultural or group identity grounds, kerb the commercial expansion of EU countries. It is estimated that it takes a SME in a European country around a year to publicize a product in the rest of Europe, compared with more or less a week in the USA. The major reason behind this delay is the language and style question. And this is not easily avoidable. Even if a lingua franca (usually English) is used, people *live* in their mother tongues. People from all regions privately use their mother tongues, and this then influences their consumer decisions.

The language issue poses a problem for companies expanding abroad, and commercial systems provide some features (especially based on localization translation services) to counteract this phenomenon. Studying organizations or businesses introspectively, that is, their internal management, however, we rarely find multilingualism management policies. In the near future, any company or organization is likely to find that they need to define a language management position similar to the consideration given to information systems management as a strategic position reporting to the general management. Of course, there are companies that have more need of such systems than others. The advance of globalization means that communication among company employees in different countries is growing and, with this, the need for an internal language management policy.

Businesses and organizations have turned to designing Knowledge Management Systems (KMS), but they should know that their ability to effectively serve the organization will be limited unless the KMS includes language management. Workshops focusing on the problem of KM with human language technologies (HLT) were organized back in the early 21st century. A series of articles, originally titled *Human Language Technologies for Knowledge Management,* were published in the *Trends and Controversies* section of one such workshop (Maybury M.[1]; Ciravegna F., [2]; Moldovan, [3]; Nikfeld, [4]). They set out the massive problem of integrating language requirements within an organization. Knowledge sharing, one of the foundations of KM, is not such an easy thing to do if this knowledge, always written in some of textual medium, is in different languages. What engineering company has not experienced problems in accurately rendering a technical manual in more than one language without problems of mistranslation? What legal firm has not met with problems in defining contract clauses that have multiple translations in the target languages and preventing misunderstandings in commercial contracts among companies from several countries?

But taking a step further, from knowledge sharing to knowledge discovery, how can we discover knowledge that is supported in several languages? While it is mostly true that English is the working language, globalization obliges us to take into account multilingual KMS. There are multiple needs, ranging from information retrieval (extraction of documents according to search terms) to information extraction (extraction of specific information from a document). They include, of course, translation, the management of interactive system dialogues with users and automatic summarization, etc. Summarization is designed to convey the same information in less space, but systems sometimes contain exactly the same information in several languages. How can this be detected? How can a search in one language locate information in another? How can the multilingual question be built into a mobile device? A dictionary will not do the job… In short, what is the best error-free and low-cost way of addressing the problem of multilingualism?

Finding knowledge in any format, anywhere and anytime still poses a challenge not only of interoperability but also of multilingualism. Two papers focusing on this issue were published back in 2001 and made the point that HLTs are the only technologies that can offer a solution to document management and thus the challenges of acquiring, retrieving and disseminating knowledge (Bontcheva et al., [5]; Budin and Melby, [6]). Even though HLTs are able to summarize, transcribe and translate huge amounts of information, the concept of multilingualism

adds a difficulty factor that sometimes renders the above systems unworkable.

Thus, both technological and methodological HLT solutions have been proposed for both information organization ,construed as a way of classifying documents, and knowledge discovery methods, targeting information search (Kao et al., [7]; Weikum et al., [8]). This applies to information location. At the other end of the spectrum, there is the problem of information delivery, that is, how the results of a search should be delivered to the client. These are knowledge delivery systems ranging from user dialogue systems to web portals. Some examples are described in (Ajila and Sun, [9]; Amardeilh and Francart, [10]).

The best known examples of information delivery systems are Internet search engines. However, the real success rate is low. It is useless to list hundreds of thousands of documents, ranked according to term frequency rather than the match between their contents and what the user is looking for. These systems generally offer users poor systems for expressing their requirements.

The design of more user-centred systems is already on the research agenda. These new systems are discussed in Davies et al.,[11]. Long-standing organizational problems, such as the reluctance of an organization's employees to use KMS on multiple grounds again resurfaced in 2005. One such ground is employees' fear of systems taking their place, and the second is the personal feeling that they and not the organization are the owners of what they know. This problem is beyond the scope of this paper, but should light the way towards the workforce's incorporation into the joint knowledge production system. For a good outline of this problem, see Atreyi, [12].

Later work focused on another problem. Support systems, whether they are document or pure databases, domain ontologies, and search engines alone are not enough. For knowledge to be accessible, we have to know what knowledge is available, and visual knowledge representation is a vital part of this process (Eppler and Burkhar, [13]). These systems associate questions with whoever is searching the information and with the formats in which it is to be delivered in just a few steps. Knowledge representation systems should be appealing and easy to use. Weblogs that are capable of creating, maintaining and sharing knowledge are a good example of such systems. The problem has been identified, but the solution is not straightforward, as the internal processes are complex (Jingjing, [14]).

It is often a matter not just of finding a document in a document database but of locating precise information in one such document, and, turning the screw further, finding information that can be inferred from, but is not explicitly specified, in the documents, and has to be pinpointed. These issues, ranging from data mining to text mining, are now a hot research topic for information and linguistic technologists (all these issues are what define what is known as linguistic engineering). EU projects have for years been concerned with topics like assuring that the right information is delivered to the right person at the right time (D'Atri et al., [15]). An example is ORCHESTRA (Cardeñosa,[16]), which generated systems capable of identifying the content of an unaddressed document sent to a company and forwarding it to the right person (Cardeñosa et al., [17]).

In view of the evolution of these issues, we can infer that unless KM includes language planning, it will fail to achieve many of its goals whenever there are multilingual parameters Language-independent knowledge repositories are still a challenging idea today, although there are several technologies that can solve the problem. It is a matter of determining how information flows from sources to users and what particular processes are in place in each organization.

This paper proposes a global information flow (IF) model enabling any organization to systematically deduce its HLT needs for its KM processes.

2. Information flows

KM is related to the management of different information formats (text, video, image, voice, data, etc.) and different languages. This activity often focuses almost exclusively on the management of different documents based on their formats and not on their content. This leads to new roles in the global content management process (some related to knowledge representation models) and, notably, validation processes associated with document management (for example, document detection in different languages with the same content).

The global information flow in any organization is based mainly on the KM concept. Today, there is no knowledge use model that defines the whole information flow process from the information sources to the user and vice versa.

Fig. 1. Content management flow

This paper aims to put forward a high-level proposal for identifying the organizational modules related to the presence of languages in this KM process.

2.1. Information sources

The modules described here are responsible for transforming information into knowledge. We appreciate that knowledge itself has no specific format or support. Knowledge will somehow have to be extracted from the information sources.

The pre-processing module is responsible for recognizing the input information in its different formats and languages. To do this, it has a multilingual module that is responsible for language management and language-dependent indexation. It also has multimedia and multiformat modules that are responsible for processing the different forms in which the information is stored and preparing the information for processing in the knowledge extraction system (KES).

The KES extracts the knowledge from the information output by the pre- processing module, which is stored in the knowledge repository. This knowledge repository is based on semantic/conceptual representation models that remove most of the restrictions derived from formats, media and languages. These modules are used on demand, that is, the fact that they exist does not mean that they are all used all the time. This calls for a fairly complex internal management system.

So, for example, if the system input is an MP3 file, the multilingual module will detect the language in which the metadata are stored, the multiformat module will recognize the format as audio, and the information will be prepared for processing by the KES. The KES will extract the knowledge and store it in a language- and format- independent repository. The information extracted by the KES will be entered in the knowledge repository (KR).

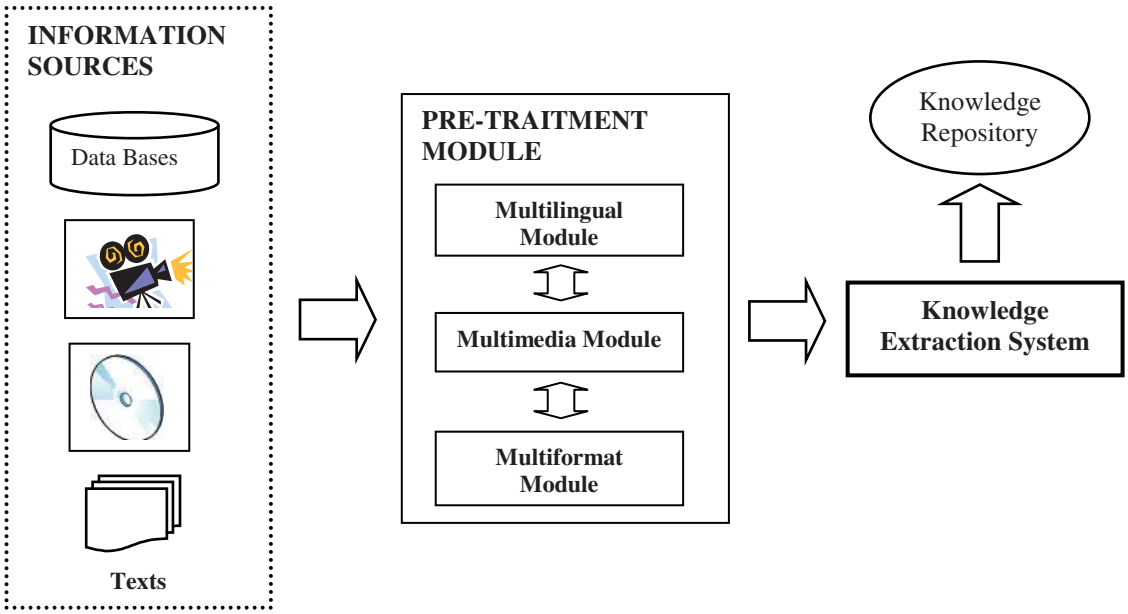

Fig. 2. Information side

2.2. User

Suppose, for brevity's sake, that users are people and not systems requesting information. The main function of the advanced dialog system (ADS) is to improve the question-answering system (QA) between the user and the KR. Figure 3 is a diagram showing these modules. Omitting details, such as the delivery system responsible for preparing and delivering the requested information, this system is supported by three main modules, namely, the multiformat module, the language manager and the multimodal module.

Fig. 3. User "side"

2.3. Knowledge repository (KR): new paradigm

The idea behind a KR is simple. It is the place where the language- or format- independent information is stored. We are not going to detail the knowledge representation models. This is still an active research area, weighing up ontologies, conceptual graphs, semantic nets and many other options. One of the challenges of interoperability is the equivalence of information from

sources with different formats.

Whatever the representation model, we know that it should be language independent. We call the conceptual framework of this single knowledge representation common conceptual representation (CCR).

Figure 4 shows a generic diagram of this repository.

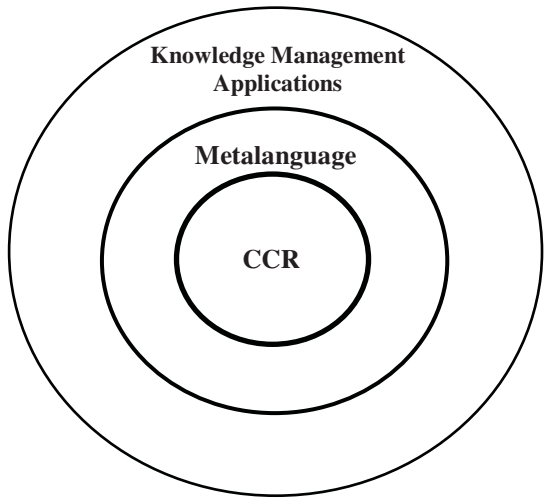

Fig. 4. Knowledge repository layers

The KR has three main layers. The primary layer responsible for maintaining the information without the adverse effects of formats, media and languages is the CCR. This kernel (CCR), which contains the pure knowledge, will interact with the different KM applications, with the above systems, using a metalanguage capable of mediating between the conceptual structures and languages, formats and media used by whoever requests or captures the information from the different sources.

The CCR should be based on modern content representation model approaches. These models are now being widely researched, and none so far substantially out performs the others. Some claim that the best model is an ontology, but this type of representation is somewhat removed from language representation. This makes assembling output documents tailored to a user and natural language (where query- tailored documents have to be generated) an enormously difficult task requiring a massive computational workload. Other proposals focus on semantic net or conceptual graph representations, which account for both language-based relations and data. These models, which are perhaps more usable in the near future, are similar to what are known as interlinguas (which have turned out to be ideal for representing language-independent documents).They are handy for representing written contents but require improvement in order to integrate ontological relations. Nowadays, when there is a pressing need to establish linked data, that is, structural relations between data from different systems, in order to effectively solve interoperability problems, representation models based on a single approach fall short. Ontological relations (mainly vertical), meaning-driven semantic relations (horizontal relations) and the need to link data from different systems suggest that the CCR model should be 3D.

3. Technologies

Each of the above organizational modules has a series of more or less language- dependent technologies. Thus, the pre-processing module and KES will be associated with voice recognition technologies, information extraction, multimodal and multimedia systems, multiformat management systems, natural language processing, machine learning, knowledge validation or language recognition technologies, all of which depend directly on the language in which the information is written.

On the other hand, the ADS will be associated with information delivery systems, automatic keyword determination systems, multilingual generation. Lastly, it will require knowledge management, distributed management, information repository and knowledge repository technologies.

Finally, content representation will have to be based on the most optimal integration of ontologies, interlinguas and conceptual graphs. Metalanguages capable of dynamically and adaptively converting ad hoc search system user queries, perhaps using intelligent agent techniques for specific users or types of information, are another worthwhile option. By reasons of space we cannot describe these technologies in more detail.

4. Conclusions

As we have seen, the maintenance of an organization's information management system is one of the most complex jobs and challenges in the 21st century. It calls for the effective integration of different types of technologies with enormous computational requirements. The only way to reduce, if not the complexity, at least the computational workload is to define either general or ad hoc models to represent language-independent information. Otherwise, the complexity of an organization's information management and use increases, the more languages are used. The systematic use of a lingua franca is not an option and even less so in document search engines for the simple reason that the number and use of languages is growing. Unless a KM system considers multilingualism as a matter requiring an organization's attention, the complexity of document management increases exponentially with the number of languages. So, we have briefly shown how a global KM system cannot obviate multilingual management as part of the organization's KM system. The idea of knowledge is neither format nor language dependent. There is a definite need for organizations to adopt the knowledge repository concept, which hinges on a multitude of technologies proper to the most advanced information systems.

References

[1] Maybury M. Challenges and Opportunities in Human language technologies for knowledge management. In Proceedings of the workshop on Human Language Technology and Knowledge Management - Volume 2001 (HLTKM '01) (2001).

[2] Ciravegna F. Challenges in Information Extraction from Text for Knowledge Management in Human language technologies for knowledge management: challenges and opportunities. In Proceedings of the workshop on Human Language Technology and Knowledge Management - Volume 2001 (HLTKM '01) (2001).

[3] Moldovan Dan. Question Answering Systems in Knowledge Management in Human language technologies for knowledge management: challenges and opportunities. In Proceedings of the workshop on Human Language Technology and Knowledge Management - Volume 2001 (HLTKM '01) (2001).

[4] Nikfeld George, 2001. Knowledge Management and Mobile Voice Interface in Human language technologies for knowledge management: challenges and opportunities. In Proceedings of the workshop on Human Language Technology and Knowledge Management – Volume 2001 (HLTKM '01) (2001).

[5] Bontcheva K., Brewster C., Ciravegna F., Cunningham H., Guthrie L., Gaizauskas R., Wilks Y., 2001. Using HLT for acquiring, retrieving and publishing knowledge in AKT: position paper. In Proceedings of the workshop on Human Language Technology and Knowledge Management - Volume 2001 (HLTKM '01) (2001).

[6] Budin G., Melby A. K., 2000. Accessibility of Multilingual Terminological Resources– Current Problems and Prospects for the Future. Gavrilidou, M. et al. (Eds.): Second International Conference on Language Resources and Evaluation – LREC 2000, Proceedings Volume II. Athens: National Technical University of Athens Press, S. 837–844 (2000).
[7] Kao A., Quach L., Poteet S.,Woods S., 2003. User assisted text classification and knowledge management. In Proceedings of the twelfth international conference on Information and knowledge management (CIKM '03) (2003).
[8] Weikum G., Kasneci G., Ramanath M., Suchanek F., 2009. Database and information-retrieval methods for knowledge discovery. Commun. ACM 52, 4, 56-64 (April 2009).
[9] Ajila S.A., Sun Z., 2004. Knowledge management: impact of knowledge delivery factors on software product development efficiency. In Proceedings of the 2004 IEEE International Conference on Information Reuse and Integration, Las Vegas, NV, United States, pp. 320– 325 (2004).
[10] Amardeilh F., Francart T., 2004. A Semantic Web Portal with HLT Capabilities. In Actes du colloque Veille Stratégique Scientifique et Technologique (VSST2004), Toulouse, France, Vol. 2, p 481--492 (2004).
[11] Davies J., Duke A., Kings N., Mladenic D., Bontcheva K., Grcar M., Benjamins R, Contreras J., Blazquez Civico M., Glover T., 2005. Next generation knowledge access. Journal of Knowledge Management, Vol. 9 Iss: 5, pp.64 – 84 (2005).
[12] Atreyi Kankanhalli, Bernard C. Y. Tan, Kwok-Kee Wei, 2005. Contributing Knowledge to Electronic Knowledge Repositories: An Empirical Investigation. MIS Quarterly, Vol. 29, No. 1, pp. 113-143. Special Issue on Information Technologies and Knowledge Management (Mar., 2005).
[13] Eppler M. J., Burkhard R. A., 2006. Using Visual Representations in Knowledge Management – a Conceptual Framework and Application Examples. In: F. Lehner, H. Nösekabel, & P. Kleinschmidt (Eds.), Multikonferenz Wirtschaftsinformatik *2006* (Berlin, Germany: GITO Verlag) (2006).
[14] Jingjing Li, 2007: Sharing Knowledge and Creating Knowledge in Organizations: the Modeling, Implementation, Discussion and Recommendations of Weblog-based Knowledge Management. International Conference on Service Systems and Service Management, vol., no., pp.1-6, 9-11 (June 2007).
[15] D'Atri Alessandro, Marco Marco, Casalino Nuncio, D'Avanzo E., Elia, A., Kuflik, T., Lieto, A., Preziosi, R., 2008. Where Does Text Mining Meet Knowledge Management? A Case Study**.** Interdisciplinary Aspects of Information Systems Studies, pp. 311-317 (2008).
[16] Cardeñosa J, 1996. Final report. ORCHESTRA (ESPRIT III 8749), CEC. (1996).
[17] Cardeñosa, J., Pastor, G., Reyna, G., Espinosa, J. A, 1996. Managing Information in a Complex Organisation. Interdisciplinary Workshop on Using Complex Information Systems. Poitiers, France, 4-6 (September 1996).