PERSONALIZATION IN VIRTUAL ENTERPRISES

Claudio Biancalana, Fabio Gasparetti and Alessandro Micarelli
Department of Computer Science and Automation, Artificial Intelligence Laboratory, University Roma Tre
Via della Vasca Navale 79, 00146 Rome, Italy

Keywords: Knowledge management, Retrieval, Elicitation, Information, User modeling.

Abstract: Each business company collects, produces and exploits for its activities and goals large amounts of information. Most of the times this knowledge makes the intellectual capital for creating value and innovation. Knowledge management (KM) systems aim at manipulating knowledge by storing and redistributing corporate information that are acquired from the organizations members. In this context, Virtual Enterprises (VE) plays a crucial role as not permanent alliances of enterprises joined together to share resources and skills in order to better respond to business opportunities. The representation and retrieval of distributed knowledge is an important feature that information systems must provide in order to obtain advantages from this kind of enterprises. PVE (Personalized Virtual Enterprise) is an ongoing research project for developing a system able to extract and let different business companies access to collective knowledge required to achieve particular shared goals. In this paper, we report the most important features of this system, especially in the context of distributed knowledge representation and retrieval.

1 INTRODUCTION

Knowledge management (KM) (Alavi and Leidner, 2001) has been recognized as a fundamental asset in the global market place. Companies know-how and the accumulated knowledge must be collected and made easily accessible to enhance the efficiency and effectiveness of knowledge-intensive work processes and competitive advantages. Improve the capitalization on existing knowledge assets facilitates the creation of new knowledge, profit returns and innovation. In spite of the apparent simplicity of the term, there are not clear definitions and classifications of knowledge management. Some experts described the purpose of KM essentially as a document management system. Other experts prefer to focus on the process of handling unstructured knowledge, or more in general, technical and organizational initiatives to manage structured and unstructured knowledge in order to store and reuse the internal companys knowledge. In this context, Knowledge Management Systems (KMS) are defined as Information Technology-based systems to support and enhance the organizational processes of knowledge creation, storage/retrieval, transfer, and application (Alavi and Leidner, 2001). The process to capture and store knowledge in ad-hoc repositories to be able to quickly retrieve information according to the user needs and goals plays a predominant role in KMS. Knowledge engineers might help to extract knowledge by elicitation activities. Direct elicitation methods such as storytelling, interviewing and question answering provide required information directly from domain experts interviewed by knowledge engineers that know what knowledge will be elicited. In indirect elicitation methods, information is not directly obtained from domain experts but there is a further step where knowledge engineers are involved in the analysis of the results of elicitation sessions. This second approach is particularly useful when the knowledge engineers have not fully explored the current domain or some knowledge has been ignored during the direct elicitation sessions. In these circumstances, indirect methods help obtaining information that cannot be easily expressed directly. While elicitation methods are useful for capturing knowledge from users or groups working on particular tasks, the large amount of information stored in paper documents or databases are paramount sources of rich knowledge to exploit during companys activities. Examples of information maintained at the work group level or beyond includes: reports, procedures, pictures, video tapes, technical standards and databases. Many people believe that semi-structured or structured information play an important role in a companys knowledge management (Gulli and Signorini, 2005). Structured information is intended to unambiguous and explicitly represents concepts in formats that describes each necessary attribute and property, e.g., relational database table, while unstructured information usu-

Biancalana C., Gasparetti F. and Micarelli A.
PERSONALIZATION IN VIRTUAL ENTERPRISES.
DOI: 10.5220/0001842905810584
In *Proceedings of the Fifth International Conference on Web Information Systems and Technologies (WEBIST 2009)*, page
ISBN: 978-989-8111-81-4

ally requires a human interpretation in order to extract its intended meaning, e.g., natural language documents, audio, still images, and video. Structured information allows users to find, share and integrate information with more precision, also with the support of software agents (Jennings and Wooldridge, 1996). Nevertheless, unstructured information covers around 80 percent of all corporate information (Moore, 2002), and even several public available digital libraries that can be employed for business activities such as the Web, are mostly composed of unstructured information. If we consider Virtual Enterprises (VE), where strategic alliances amongst non-competing companies are settled for the accomplishment of specific goals, there importance of sharing and recovery of useful structured/unstructured information among partners becomes even more crucial A company may be not completely aware of the knowledge stored by the partners of the VE. Bringing collectively available complementary competencies and resources is one of the major achievements of a VE. In this paper we describe PVE (Personalized Virtuale Enterprise), an under development KMS able to retrieve and share knowledge shared by different companies grouped in a VE. One on the most important features of the proposed system is the ability to manage different kinds of information representations and degree of formality, i.e., structured, semi-structured or unstructured, providing an uniform access to several different information sources. Instead of long and costly elicitation processes to manually extract and annotate knowledge, PVE uses machine learning and information extraction techniques in order to draw knowledge from the information stored in the companies internal digital libraries and map it into the enterprise ontology. User modeling is employed to adapt the interaction with the information system in order to personalize the results of humans information seeking activities.

2 RELATED WORK

In recent years, knowledge management has become a focus of attention for many organizations. Knowledge is considered to be the key source for sustainable competitive advantage (Nonaka and Takeuchi, 1995), (Davenport and Prusak, 1997). Therefore, Knowledge management is aimed at locating, capturing, transferring, sharing and creating knowledge within and across organizations. It will be clear that conceptual modeling, as developed within the field of KBS construction, provide key techniques for knowledge management (Gaines et. al., 1997), (Schneider, 2000). A characteristic that turns out to be an advantage over other industries in terms of managing intellectual capital is that artifacts (documents) are already captured in electronic form and can easily be stored and shared. In addition, software engineers often have a friendly attitude towards using new technology. This means that a software organization that implements a knowledge management system could have a good chance to succeed with this mission. However, this remains a challenging task because a knowledge management system is more than just technology. There are only a few published works about initiatives to manage knowledge in software organizations, but all of them talk about the difficulty of achieving employees acceptance and implementing the KM system in a way that maximizes the help provided to its users (Schneider, 2000), (Schneider, 2001), (Brssler, 1999), (Johansson et. al., 1999).

3 KNOWLEDGE INDEXING

Before facing the problem of knowledge retrieval, it is essential to analyze how the system indexes the available information, that is, which representation has been chosen to guarantee an efficient and effective retrieval phase. In particular, the requirements are twofold: it is essential that knowledge is quickly retrieved by users, and this knowledge accurately satisfies the users information needs in terms of high precision. An indexing system for business companies must also be able to deal with different kinds of information representations, from unstructured documents based on natural language to ontology-based knowledge and relational databases. Moreover, it should provide a comprehensive and homogenous human-computer interface for knowledge retrieval. In order to provide the aforementioned prerequisite, it is necessary to consider different types of information and the degrees of information richness. Information based on ontological standards, for instance, expresses relationships between typically non-structured information, e.g., natural language text, and meta-data. These meta-data usually state features or classes related to given peaces of information. A typical example is the association between a document and one particular category in a predefined taxonomy. As for information stored in databases, we have an underlying relational model that clearly states the semantic meaning of each peace of informative unit, e.g. price, address, location, etc., and therefore facilitate the interpretation/recovery process. In order to define a unique representation that deals with the different types of available information, i.e., natural language, ontology-based and databases, we must define a subset of shared features that is possible to generalize, and automatic or semi-automatic methods and techniques for translating information from one of these representations to the internal one. Ontology languages capture high-quality relationships and meta-data content that enable logic-based agents to interpret and take decisions autonomously (Jennings and Wooldridge, 1996). The Semantic Web extends this vision to the Internet domain, where information is no longer based on HTML, but on semantic standards like OWL (Web Ontology Language), or preliminary standards like DAML+OIL. Nevertheless, research on these representations is not completed, and the logical engine able to automatically interpreting such information within single or multi-agent systems with traditional computational resources is yet to be defined. Moreover, much of the current available information is written by humans in natural language so additional effort is required to translate this information in the new ontology-based languages. Translating large amount of non-structured information into new formalisms is not an activity that can be accomplished manually. Research to develop methods and techniques for this goal has not reached completely satisfying results. As for unstructured information based on natural language poses many problems during the indexing, but it is also particularly problematic to retrieve it efficiently. The well-known vocabulary problem (Furnas et. al., 1987) for instance, points out further issues in terms of synonymity and polisemy of words that do not allow users to express univocally their information needs.

3.1 Internal Representation of Knowledge

As previously stated, the proposed internal representation of knowledge defines some common features shared among the three kinds of information briefly described. This sort of intermediate representation consists of traditional non-structured information with associated meta-information related to concepts of a taxonomy of the business domain for the given virtual enterprise (see figure 1). In a few words, each informative unit is classified in a subset of categories from a simplified ontology. Such meta-information can be exploited both in the retrieval phase, reducing possible ambiguities in the processed information, and to re-organize the knowledge in more efficient ways for further user search activities, e.g., online hierarchical clustering (Gulli and Signorini, 2005).

One further advantage of such a representation is

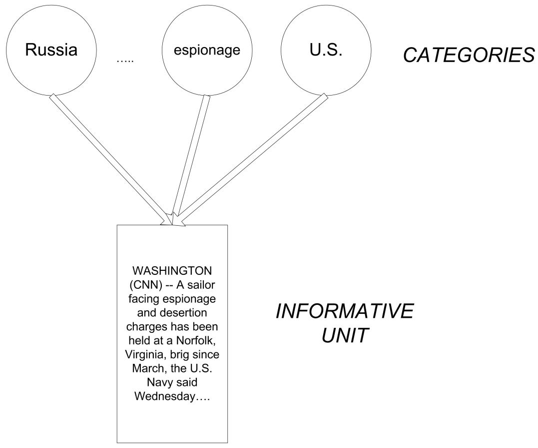

Figure 1: Each informative uint, e.g., a document or a segment of it, is associated with one or more categories in a given taxonomy.

the chance to exploit traditional and well-known indexing and retrieval techniques developed in the Information Retrieval field, as search engines based on the Vector space model (Salton and McGill, 1983). Such systems guarantee quick response time thanks to data structures appropriately studied for efficiently memorizing the input documents. Additional information can be easily indexed and retrieved together with the original data in appropriate fields that can be used during the recovery process. Even though the proposed representation simplifies the stored meta-data, i.e., there are no relationships between informative units such as IS-A or HAS-A relations in ontology-based languages; the burden is now on populating the internal knowledge base given the available information. In other words we have to define techniques and methodologies to transform the information represented in one of the three-above mentioned typologies, into the proposed intermediate representation.

4 EVALUATION TEST BED

We are currently defining the test-bed for the evaluation of PVE. We have assumed a 6-month lasting period, starting from a set-up phase, where we settle a configuration of the enterprise, i.e., number and typology of the companies, and install the various distributed modules. Afterwards, the VE prototypes will be instantiated and evaluated. Due to the nature of the prototype, i.e., a distributed environment with different kinds of knowledge representations and typologies of users, we have decided to evaluate the system in a real scenario. In this way, we are able to evaluate all the available features through precision and accuracy measures, and the analysis of ad-hoc user questionnaires.

5 CONCLUSIONS

The goal of PVE project is studying and devising methodologies and techniques for distributed management of knowledge in the context of a Virtual enterprise, favoring synergies and the interaction among different companies. In particular, we have introduced a knowledge indexing and retrieval engine used for manage the information stored in a company. This engine is used by the users to retrieve documents related to the current needs. The retrieval is enhanced with semantic metadata extracted by means of information extraction techniques. Moreover, the system employs a user modeling component to adapt the human-computer interaction during information seeking activities.

REFERENCES

M. Alavi and D.E. Leidner, Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. MIS Quarterly, Vol. 25 No. 1, pp. 107-136, 2001.

F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi and P.F. Patel-Schneider (Eds.), The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, 2003.

P. Brssler, Knowledge Management at a Software Engineering Company - An Experience Report, Workshop on Learning Software Organizations, LSO'99, Kaiserslautern, Germany, pp. 163-170, 1999.

P. Brusilovsky and M. Maybury, From adaptive hypermedia to the adaptive web. Communications of the ACM, Vol. 45 No. 5, pp.3033, 2002.

F. Y. Y. Choi, Advances in domain independent linear text segmentation, In Proc. of the 1st conference on North American chapter of the Association for Computational Linguistics. Seattle, Washington, pp2633, 2000.

C. W. Choo, Information Management for the Intelligent Organization: Roles and Implications for the Information professions. In Proc. of the 1995 Digital Libraries Conference, Singapore, March 1995.

T.H. Davenport and L. Prusak, Working Knowledge: How Organizations Manage What They Know, Harvard Business School Press, 1997.

P. Ferragina and A. Gulli, A personalized search engine based on web-snippet hierarchical clustering. In Proc. Of the 14th World Wide Web Conference, 2005.

G.W. Furnas, T.K. Landauer, L.M. Gomez, and S.T. Dumais, The vocabulary problem in human-system communication. Commun. ACM, Vol. 30 No. 11, pp. 964971, 1987.

B.R. Gaines, M.A. Musen and R. Uthurusamy, Artificial Intelligence in Knowledge Management, Technical Report SS-97-01, Stanford University, 1997.

S. Gauch, M. Speretta, A. Chandramouli and A. Micarelli, User Profiles for Personalized Information Access. In Brusilovsky, P., Kobsa, A., Nejdl, W., eds.: The Adaptive Web: Methods and Strategies of Web Personalization. Volume 4321 of Lecture Notes in Computer Science, pp.54-89. Springer-Verlag, Berlin Heidelberg New York (2007).

A. Gulli and A. Signorini, The indexable web is more than 11.5 billion pages. In the Proc. of the 14th international conference on World Wide Web (WWW05), pp902903, New York, USA, 2005.

R. Hartley, R. Kelsey and R. Webster, Consolidating Multi-Source and Multi-Media Knowledge. In AAAI Spring Symposium Articial Intelligence in Knowledge Management. AAAI, March 1997.

N. R. Jennings and M. Wooldridge, Software Agents. IEE Review. Vol. 42 No. 1, pp17-21. January 1996.

T. Joachims, Text categorization with support vector machines: learning with many relevant features. In Proc. of ECML-98, 10th European Conference on Machine Learning, pp.137142, Chemnitz, DE, 1998. Springer Verlag, Heidelberg, DE. Published in the Lecture Notes in Computer Science series, number 1398. 1998.

T. Joachims, Transductive inference for text classification using support vector machines. In Proc. of ICML-99, 16th International Conference on Machine Learning, pp.200209, Bled, SL, 1999. Morgan Kaufmann Publishers, SF, US, 1999.

C. Johansson, C. and P. C. M. Hall, Talk to Paula and Peter - They are Experienced, Workshop on Learning Software Organizations, LSO'99, pp. 171-185, 1999.

D. Kelly and J. Teevan, Implicit feedback for inferring user preference: a bibliography. SIGIR Forum, Vol. 37 No. 2, pp.1828, 2003.

G. Lawton, Knowledge Management: Ready for Prime Time IEEE Computer, Vol. 34, No. 2, pp. 12-14, 2001.

C. Moore, Diving into Data. InfoWorld. October 25, 2002.

I. Nonaka and H. Takeuchi, The Knowledge-Creating Company, Oxford University Press, New York, 1995.

P. Pirolli, P. Schank, M.A. Hearst and C. Diehl, Scatter/Gather Browsing Communicates the Topic Structure of a Very large Text Collection, In Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), May 1996.

G. Salton and M. McGill, An Introduction to modern information retrieval. Mc-Graw-Hill, New York, NY, 1983.

K. Schneider, LIDS: A Light-weight Approach to Experience Elicitation and Reuse, The Profes 2000 Conference, Oulo, Finland, 2000, pp. 407-424, 2000.

K. Schneider, Experience Magnets - Attracting Experiences, Not Just Storing Them, Product Focused Software Process Improvement, PrOFES'01, Kaiserslautern, Germany, pp. 126-140, 2001.

F. Sebastiani, Machine learning in automated text categorization. ACM Computer Surveys. 34(1): 1-47 (2002).