CBR System for Innovation Management in Power Companies: CELESC case study

Tânia C. D. Bueno[1], César R. K. Stradiotto[13], Sonali P. M. Bedin[13], Vinícius O. Mirapalheta[1], Hugo C. Hoeschl[13], Thiago P. S. de Oliveira[1], Luiz A. P. Athayde Filho[2], Pedro A. S. de Carvalho Júnior[2]

[1]I3G - Instituto de Governo Eletrônico Inteligências e Sistemas, Florianópolis, SC, Brasil. http://www.i3g.org.br {tania.bueno; cesar.stradiotto; sonali.bedin; vinicius.mirapalheta; hugo.hoeschl; thiago.paulo}@i3g.org.br

[2]CELESC - Centrais Elétricas de Santa Catarina, Florianópolis, SC, Brasil. http://www.celesc.com.br {luizpaf@celesc.com.br; pedroascj}@celesc.com.br

[3]Programa de Pós Graduação em Engenharia e Gestão do Conhecimento-EGC/UFSC. Florianópolis, SC,

Abstract. The CELESC R&D Management Systems is a CBR system for information management whose retrieval process is based on knowledge representation by ontologies. The system was built to support R&D project management, to reduce losses on intellectual property over ideas generated inside the company. The company knowledge base has projects developed since 2001. Developed for web environments, the system allows a wide knowledge diffusion of its content and results.

Keywords: Knowledge management; Case-based reasoning, R&D Project Management, Electric Sector, Shared Management.

1 Introduction

In Brazil the Federal Government imposes, by means of the Law, minimal investments in R&D from electric power companies. CELESC is a governmental enterprise responsible for the electric sector, inside the state of Santa Catarina – Brazil. The present project was constructed a system based on a knowledge management framework developed by the Institute i3G [1, 2, 3, 4], and was created an Information Management Model. For this reason, to explain the developed research work, we divided this job on the following sections: In item 2 it is presented the CELESC R&D Management System. In item 3 it is presented the ontology-based information management model. In item 4 it is presented the construction of cases in

the range of the management system. In item 5 it is presented the information retrieval techniques. In item 6 are presented the conclusions and future wok. Finally the acknowledgements and the consulted references are presented.

2 The CELESC Innovation Management System

The system is composed by two environments based on Web: One for management, and the other for searching information on R&D documents. These technological structures will serve as management instruments to consult about the R&D information. With these environments, it will be possible to keep track – daily – about the produced job by all these people involved on production, storage and recovering of strategic information, inside the parameters defined by ANEEL. The search environment is composed by three sub-modules: **Search by similarity**, where the user entry a text in natural language, which is compared with the documents by ontologies; **Search by subject**, where the user looks for texts inside predefined specific domains organized by ontologies; and **Simple search**, where the system searches for an exact string.

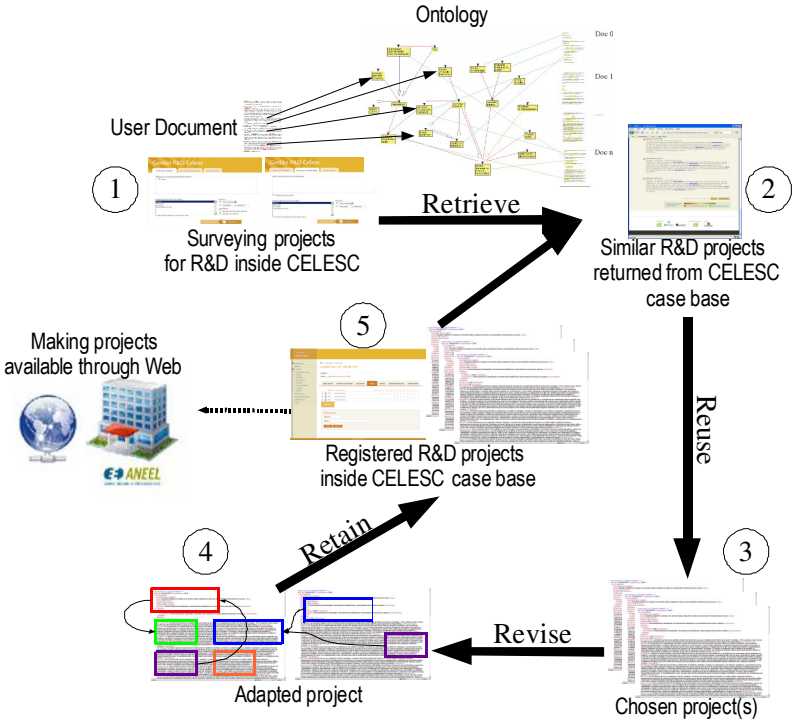

Figure 1: CBR Viewing for CELESC R&D Management System, showing the role of ontologies for case retrieval.

3 Ontology Construction and Case Representation

The participation of the ontology structure in the Innovation Management systems occurs in three moments. First, the system extracts information from different previously selected sources. Each of these documents is indexed based on ontologies defined by the specialists and knowledge engineers during the knowledge engineering process: The system will mark the documents with all indicative expressions found in the text, storing them in an organized way in the knowledge base. Thus, it is possible to make a pre-classification of the cases in the base according to what was defined in the knowledge organization helped by ontologies. In a second moment, the ontologies are used in the analysis interface available to the user. The process starts when user types the input text for search. At this point, the indicative expressions defined by the user, which are inside the ontology, are identified. These expressions in the entry case determine the stream of relations that will be used by the system. The cited relationships are the following: Synonyms, Related Terms: Representing terms appearing together in a great amount of texts. A type of: Same as class relation. A part of, also known as the Meronimy relation. These relations are used to compare terms from a user document and the documents inside the knowledge base.

3.1 Indexing Process

The present system was developed using CBR method, which allows the representation of texts in form of cases, through the use of indexes. Each case is structured as a set "indicative expressions" defined by ontologies and all words in the documents from current cases and target cases. Also, it was defined indexes, representing characteristics of the cases that indicate your usefulness in a specific situation. Those indexes have been determined to facilitate the retrieval process, which consists in four attribute-value pairs to index: Date, Source, List of Ontologies and List of Dictionary Words. The system uses the ontologies created in the editor to realize the information retrieval of *Projects, Documents in Library, Innovations and R&D Needs*. The system converts the stored information in a word vector, removing unuseful words to our information retrieval model (stop words), like pronouns, preposition, articles, gender, plural and special characters. After that, the system processes the word stemming. The storage uses an open structure established in language XML, directly related with the database indexed automatically in an intelligent way. The ontologies assist the process of automatic indexation (case definition) and supply the indexes and expressions for the similarity degree evaluation between the query and the case (document) in the database. After the automatic indexation each case is valuated, the cases are classified in accordance with the values obtained with the metric of similarity.

4 Case Constructions and Application of IR Techniques

The documents are represented in form of cases, which consist of the original document text and a set of indexes in form of attribute-value pairs, and a general

domain knowledge is included in form of an ontological representation. As a result, the indexed ontologies and relations represent information related to the content of the texts, and the information necessary to identify the tuples (source, number and date) is represented. Documents manually entered into the system follow a sequence of literal operations, which reduces the complexity for representation and allow the transformation of texts into a set of terms and indices. After the reduction in the document size, the system extracts the significant information using the dictionary of ontologies, created for a specific domain. Based on this information, a formal description is generated by the definition of the indices for a specific case. The description is partially compared with all the cases in the base using CBR techniques [5], in order to identify the most equivalent case. To compute the similarity for each case, the Ontology Dictionary is used and it indicates the similarity between the individual terms and each relation between the Indicative Expressions.

5 Conclusions and Future Work

The use of ontologies is responsible for the high quality of results in the information retrieval process. Ontologies are important in the analysis interface, available to the user. Although there are a model to represent knowledge from texts inside CELESC R&D knowledge base, this base is not completed with enough documents to go for experiments. In future works we are going to fill it with R&D related documents to make the tests for the representation with ontologies.

6 Acknowledgements

Our thanks to Centrais Elétricas de Santa Catarina – CELESC, and to i3G Team: Mr. Thiago Paulo de Oliveira, Angela Iara Zotti and Claudia Bueno, which helped us to finish this job.

References

1. BUENO, T.C.D., Wangenheim, C. G., Mattos, E. S., Hoeschl, H. C., Barcia, R. M.: JurisConsulto: retrieval in jurisprudential text bases using juridical terminology. ICAIL 1999: 147-155
2. BUENO, T.C.D., BEDIN, S., CANCELLIER, F., HOESCHL, H. C.: Evaluating the Effectiveness of a Knowledge Representation Based on Ontology in Ontoweb System. FQAS 2006: 625-633
3. BUENO, T. C. D. ONTOJURIS. Multi-Language Ontology Edition. http://www.i3g.org.br/ontojuris/ Acessado em Maio de 2009.
4. STRADIOTTO, C, R. K., BUENO, T.C.D., HOESCHL, H. C. Ontology Graphical Editor for Multilingual Document Search System. FQAS 2009: 453-464.
5. WANGENHEIM, A. V., WANGENHEIM, C. G. V. RACIOCÍNIO BASEADO EM CASOS. ISBN 85-204-1459-1 Editora Manole. Barueri. São Paulo. 2003.