An Ontology Based CBR Architecture for Knowledge Management in BULCHINO Catalogue

Nadezhda Govedarova, Stanimir Stoyanov, Ivan Popchev

Abstract: *In this paper we present otology-based architecture for knowledge management in BULCHINO web-based catalogue for Bulgarian cultural heritage. Our approach for realizing content-based search and retrieval in the catalogue implies the application of the CBR technology for the representation of metadata characterizations of the cultural objects and the integration of ontology for semantic formalization of these characterizations. We provide an insight of the technical aspects of the architecture by enumerating the technological requirements, where we pay specific attention to the CBR framework jCOLIBRI and its features for implementing concept-based similarity assessment.*

Key words: *CBR, knowledge management, jCOLIBRI framework, BULCHINO*

INTRODUCTION

Case Based Reasoning (CBR) [1, 20] is widely discussed in the literature as a technology for building information systems to support knowledge management, where metadata descriptions for characterizing knowledge items are used. These characterizations are called cases and are stored in a case base. A main drawback of this technology, however, is that CBR systems are often isolated and closed in the sense that they are not developed with respect to cooperation with other systems and consequently do not allow an easy knowledge exchange. Current research of distributed CBR [9, 18, 26] shows how CBR systems can benefit from a standardized shared knowledge representation that implies unambiguous interpretation of cases and in this way enable the development of systems that are able to search across multiple case-bases.

Since 2005 we have been working on the project "CBR Platform for Ontology Based Knowledge Management", funded by the Bulgarian Ministry of Education and Science. The project aims at the design of a flexible and extensible CBR-based platform for knowledge management and knowledge processing**.** We intend to apply it by the implementation of prototype applications in the fields eLearning (Selbo2) [6] and cultural heritage tourism (BULCHINO) [2, 30]. The goal of our current work is the development of an intelligent e-service to perform the content-based search in the above mentioned systems. We make an issue of the retrieval process and study how the synergy of ontology and CBR technology could improve its efficiency**.**

In this paper we present architecture of the search layer in the particular system BULCHINO Catalogue, a web-based catalogue for the Bulgarian cultural heritage. We outline its main components and describe how they collaborate. For the concrete technological implementation we chose the CBR framework jCOLIBRI [9, 16, 18, 19]. The results of our ongoing work on the adaptation of the framework are presented in [32].

The rest of the paper is organized as follows. Next Section describes the setting of Cultural heritage domain, the research problems and current work in it. Then we present briefly the BULCHINO catalogue and our concept for extending the search layer of the system by integrating an intelligent e-service in it to perform the search process over different types of information repositories. We provide an insight of the technical aspects of the application by enumerating the technological requirements and the associated architecture, where we pay specific attention to the CBR framework jCOLIBRI and its features for implementing the reasoning process over ontologies. Finally we outline the directions of our future work.

INTELLIGENT INFORMATION PROCESSIG IN CULTURAL HERITAGE DOMAIN

Cultural heritage appreciation is a privileged area for the application of innovative, knowledge intensive e-services that provide a flexible and efficient method for searching cultural objects and guarantee the user with a set of results actually related to his/her

interest. In this section we discuss some of the opportunities and challenges in this domain with a specific view of intelligent information processing that takes into account the semantics of the knowledge items (cultural objects). Some of the up-to-date projects that address these issues are eCULTURE [10], HUBUSKA [14], CATCH [4], CHIP [5], STICH [28], FinnOnto [11], MICHAEL [22] to mention just a few of them.

However, despite large investments and efforts have been made, there are still a lot of unsolved problems, the most urgent of which concern the slow digitalization process, the distributed databases, the heterogeneous description of the objects in the different schemes and systems and an information presentation that slightly corresponds to the user requirements. A main bottleneck here is the lack of uniform semantics interpretation standards and technologies. Hence, interoperability can be only achieved by some kind of semantics unification.

A common approach for knowledge management is the abstraction of knowledge items to a characterization by means of metadata descriptions, which are used for further processing. However this characterization is not enough to guarantee unambiguous interpretation. One has also to provide a formal model for restricting the possible interpretations of metadata annotations, i.e. to provide the necessary shared conceptualization of the domain among computer agents and users of the system so that a semantics-based access to the knowledge items is possible. Therefore the problem of semantic interpretation could be brought on the one hand to the definition of abstract characterization of a knowledge item and on the other hand to the choice of common accepted vocabulary to formalize it.

Our approach implies the application of the CBR technology for metadata characterization of the cultural objects (presented as cases and stored in a case base) and the integration of ontology for semantics formalization of these characterizations. The usage of ontology is useful for the CBR community regarding different purposes [18]: persistence of cases and/or indexes using individuals or concepts that are embedded in the ontology itself; as the vocabulary to define the case structure, either if the cases are embedded as individuals in the ontology itself, or if the cases are stored in a different persistence media as a data base; as the terminology to define the query vocabulary; retrieval and similarity, adaptation and learning. At this stage we consider to use ontology only as vocabulary for defining the case structure (attribute-value pairs), which is based on the CCO (Cataloging Cultural Objects) standard [2].

Apart from specific domain knowledge every knowledge intensive system should include general knowledge, as well, to improve its reasoning power. Common vocabularies and thesauri in the Cultural heritage domain are Getty vocabulary databases (ULAN, TGN, AAT) produced and maintained by the Getty Vocabulary Program [13].They could be used as additional knowledge base that contain semantic networks that show links and paths between terms, names, and other information about people, places, things, and concepts relating to art, architecture, and material culture. These relationships can make retrieval more successful.

BULCHINO CATALOGUE

Bulgaria is the cradle of one of the most ancient civilizations, the Thracian one, monuments of which still reveal further aspects from the initial stages in the development of the European culture. The promotion of the diverse Bulgarian cultural heritage, therefore, will not only contribute to increase in the share of cultural tourism in Bulgaria but also to the enrichment of the shared European cultural area.

BULCHINO (BULgarian Cultural HIstorical and Natural Objects) is a web based catalogue for electronic representation of the Bulgarian culture-historical heritage that has been developed at the E-Commerce Laboratory in the University of Plovdiv. A major aspect to ensure the interoperability of the catalogue is the usage of standard, platform independent technologies – Semantic Web technologies, established standards for describing the cultural objects, such as CCO and web services. The access to the

available services and electronic content is accomplished via specialized portal, whose architecture is developed in compliance with the referent model of Delphi Group [6].The portal reference architecture includes the following nine different layers: presentation, personalization, e-services control, integration, collaboration, search, categorization, and loop process layer.

Our efforts aim to improve significantly the efficiency of the search process and extend the search layer with an intelligent e-service, so that intuitive information access and content-based retrieval is available. We consider that applying the Case Based Reasoning technology for the realization of the search task in BULCHINO is an appropriate approach [29, 31, 32].

TECHNICAL REQUIREMENTS&SOLUTIONS

In this section we describe briefly the design and implementation principles that lay in the base of the search layer extension. As mentioned in the previous sections, for the realization of the intelligent e-service we chose the framework jCOLIBRI-a java-based framework that supports the development of knowledge intensive CBR applications and help in the integration of ontology in them. Our motivation for choosing this framework is based on a comparative analysis between it and other frameworks, designed to facilitate the development of CBR applications. jCOLIBRI enhances the other CBR shells: CAT-CBR[3], CBR*Tools[21], IUCBRF [27], Orenge [24] in several aspects: availability (open source framework), implementation (the Java implementation is one of our main requirements with respect to the easy integration in the BULCHINO system which is implemented in J2EE [15] environment), GUI (the provided graphical tools facilitate the system design). Another decision criterion for our choice is connected with the fact that jCOLIBRI affords the opportunity to incorporate ontology in the CBR application to use it for case representation and content-based reasoning methods to assess the similarity between them. The ontology support of jCOLIBRI is built arround the OntoBridge library [23] that was developed by the GAIA research group [12] to easily manage ontologies and DLs (Description Logics) reasoners. It is base on the Jena framework for development of Semantic Web applications [17] and makes possible the connection with several actively supported DLs reasoners. Some of the actively supported once are presented in [8]: CEL, FaCT++, fuzzyDL, KAON2, MSPASS, Pellet, QuOnto. We chose Pellet [25] as inference engine in our architecture for several reasons. Apart from being open source and Java based, it supports Jena interface what enables the direct connection with the OntoBridge library of jCOLIBRI. For the manual generation and modeling of the domain ontology we chose the Protégé editor. A more detailed description of these components and the interaction between them is presented in the next section.

BULCHINO SEARCH LAYER ARCHITECTURE

The proposed architecture (Fig.1) is based on our approach for realization of content based retrieval of cultural objects by means of metadata characterizations and domain ontology inclusion. It implies to use ontology as vocabulary to define complex, multi-relational case structures to support the CBR processes. Standard cases are composed by several attributes with different simple data types (Integer, String). We use the Concept data type (supported by the jCOLIBRI framework) to indicate that an attribute is going to represent a concept of the ontology. The values of this attribute are going to be the corresponding instances of the linked concept.

Except from the Concept type, the architecture takes advantage, as well, of another feature of jCOLIBRI framework – the two-layer organization of the case base. The metadata descriptions of the cultural objects (cases) are abstracted from the details of their physical representation in the Electronic Catalogue (Persistency media) and are stored in the case base (In-memory organization). This way the same methods can operate over different types of information repositories. The mapping (the process is described in the following) between the two layers is realized by connectors – objects that

read the values of the data base columns and ontology and return them to the application, i.e. assign them to the attributes of the case. Basing on the same idea, the case base implements a common interface for the similarity methods to assess the cases. This way the organization and indexation of case base will not affect the implementation of the reasoning methods.

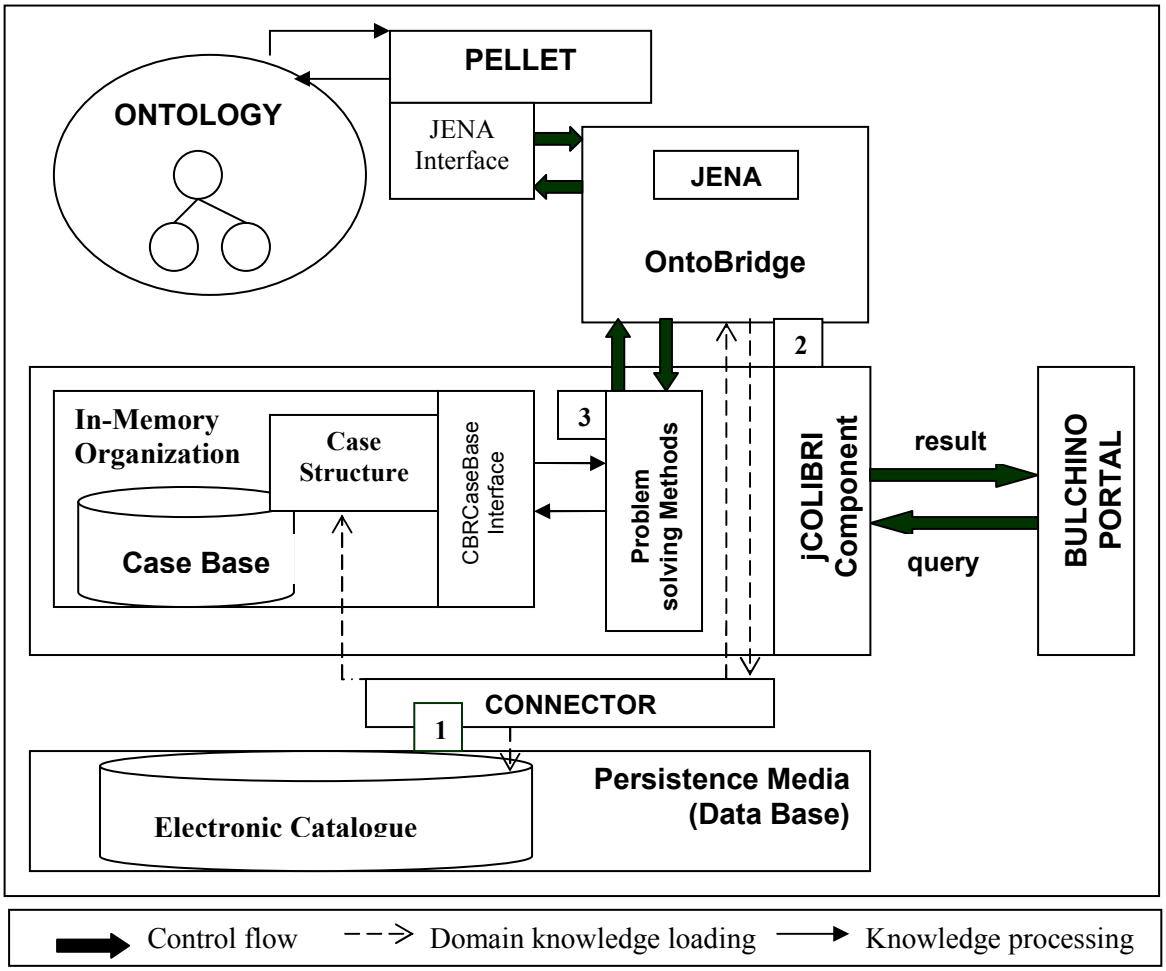

Fig.1. BULCHINO Search Layer Architecture

▪ Persistence Media - in this case we are not storing the whole case base into the ontology:

- Data Base continues storing the values of the attributes. But now the column that stores the attribute linked with a concept is going to have the names (stored as strings of characters) of the instances of this concept. The electronic catalogue is developed in compliance with the CCO standard.
- Ontology stores information about cultural objects where concepts are types, or classes, individuals are allowed values, or objects and relations are the attributes describing the objects. Knowledge is formally represented and stored in OWL (Web Ontology Language) ontology that is managed with the Pellet reasoner.

▪ jCOLIBRI Component - is designed on the basis of jCOLIBRI framework.

It is the core of the architecture and is responsible for the knowledge and queries management. Since the problem solving methods are domain independent, the domain specific information should be first loaded from the persistence media so that processing with it is possible. The data base connector will read the values in the table (1) and if encounters a concept typed attribute it looks for an instance with the same name in the Ontology (through OntoBridge) (2). Once found the connector will fill the values of the attribute of each case with the corresponding instances of the ontology, loaded by the Pellet reasoner. It is used as well by the methods (3) to compute the content-based similarity between the concept typed attributes.

▪ OntoBridge - is a Java library that eases the management of the ontology in a jCOLIBRI-based application. It uses Jena library to implement most of the required methods for accessing the ontology, loaded in the reasoner. With this extension the jCOLIBRI component can acquire domain knowledge from ontology, defined in DLs, and achieve this way uniform case representation, what will enhance the interoperability of the whole system. This, namely, was our main goal when designed the jCOLIBRI component as BULCHINO search layer extension.

▪ Pellet - is the reasoner that we chose to load the ontology and implement inference over it.

CONCLUSIONS AND FUTURE WORK

In this paper we presented an ontology-based architecture for knowledge management in BULCHINO catalogue. A crucial role in it plays the jCOLIBRI-based component that is the cornerstone in the proposed architecture. Its functionality is briefly described, where specific attention to the incorporation of domain specific ontology is paid. Our main contribution is the described approach for integration of different types of information repositories and metadata characterizations of the cultural objects in BULCHINO to achieve semantic interoperability of the catalogue and the application of up-to-date Semantic Web technologies to realize it. This article discusses the issues involved, but the whole system implementation is far from its final form.

Our current efforts aim at the specification and implementation of the search layer architecture. To put our aims into practice we should first of all develop the domain ontology and study how the content-based similarity between the Concept typed attributes could be assessed. We have been working as well on the design of entirely ontology-based structure of the case and the development of our own reasoning methods to operate with it.

Further work will address the design of a service for generation of tourist routes, which will be base on the proposed architecture. This way the users of BULCHINO catalogue will be afforded the opportunity to search a particular cultural object (local queries) as well as an integrated tourist route, comprising several cultural objects (global queries). Our main challenge, however, remains the definition of the tourist rout concept and the its representation in the terms of the case-based reasoning technology.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the financial support of the Bulgarian Ministry of Education and Science and Research. Project Ref.No.MI-1502/2005.

REFERENCES

[1] Aamodt, A., Plaza, E., Case-base reasoning: Foundational issues, methodological variations, and system approaches, AI Communications, 7(1), 1994, 39-59

[2] Cataloging Cultural Objects, A guide to Describing Cultural Works and Their Images, 2005, http://www.vraweb.org/ccoweb/cco/index.html (to date)

[3] C.Abasolo, E.Plaza, and J.-L.Arcos, Components for case-based reasoning systems, Lecture Notes in Computer Science, 2504, 2002

[4] CATCH project, http://www.nwo.nl/nwohome.nsf/pages/NWOP_5XSKYG (to date)

[5] CHIP project, http://www.chip-project.org (to date)

[6] D. Mitev, I.Minov, O. Rahneva, Intelligent components in development environment for e-learning Selbo 2, In Proc. "Science, education and time as our concern" , 30 November - 1 December, Smolyan, Bulgaria, 2007 (in Bulgarian)

[7] Delphi Group, http://www.delphigroup.com (to date)

[8] Description Logics Reasoners, http://www.cs.man.ac.uk/~sattler/reasoners.html

[9] Díaz-Agudo, B., González-Calero, P.A., An Architecture for Knowledge Intensive CBR Systems, In Blanzieri, E., Portinale, L., (Eds.): Advances in Case-Based Reasoning (Procs. of the 5th European Workshop on Case-Based Reasoning, EWCBR 2000),

Lecture Notes in Artificial Intelligence, 1898, Springer, 2000.
[10] eCULTURE project, http://e-culture.multimedian.nl (to date)
[11] FinnOnto project, http://www.seco.tkk.fi/projects/finnonto (to date)
[12] GAIA Group, http://gaia.fdi.ucm.es/index.html (to date)
[13] Getty Vocabulary Program, http://www.getty.edu/research/conducting_research/vocabularies (to date)
[14] HUBUSKA project, http://www.hubuska.com/ (to date)
[15] J2EE, java.sun.com/j2ee/1.4/docs/tutorial/doc (to date)
[16] jCOLIBRI Theoretical Foundations, http://gaia.fdi.ucm.es/projects/jcolibri/docs.html/
[17] Jena Framework, http://jena.sourceforge.net/index.html (to date)
[18] Juan A. Recio-García, Belén Díaz-Agudo, Pedro A. González-Calero, and Antonio Sánchez, Ontology based CBR with jCOLIBRI, Applications and Innovations in Intelligent Systems XIV.Proceedings of AI-2006, the Twenty-sixth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, pages 149–162, Cambridge, United Kingdom, December 2006. Springer.
[19] Juan A. Recio-García, Belén Díaz-AgudoPedro, González-Calero, jCOLIBRI2 Tutorial, Document version 1.1, Jenuary 22, 2008
[20] Lenz, M., Bartsch-Sporl, B., Burkhard, H., and Wess, S. Case-Based Reasoning Technology – From Foundation to Applications,Lecture Notes in Artificial Intelligence 1400, Springer Verlag, 1998
[21] M.Jaczynski and B.Trousse, CBR*Tools: An object-oriented framework for the design and the implementation of case-based reasoners, In Proceedings of the 6th German Workshop on Case-Based Reasoning, 1998
[22] MICHAEL project, http://www.michael-culture.org/en/about/project (to date)
[23] OntoBridge, http://gaia.fdi.ucm.es/grupo/projects/ontobridge (to date)
[24] Orenge, http://www.ovitas.com/PDF/orengeWhitepaper.pdf (to date)
[25] Pellet, http://pellet.owldl.com (to date)
[26] Ralph Bergman, Martin Schaaf, On the Relation between Structural Case-Based Reasoning and Ontology-Based Knowledge management, In Proc. of German Workshop On Experience Management, April, 2003
[27] S.Bogaerts and D.Leake, IUCBRF: A Framework For Rapid And Modular Case-Based Reasoning System Development, November, 2004,
[28] STICH project, http://www.cs.vu.nl/STITCH (to date)
[29] Stoyanov S., I. Popchev, D. Chaushkova, M. Trendafilova, A Case based reasoning Approach for Development of Intelligent Services. Journal "Information Technologies and Control", No. 3/2004, Year II, Pp. 31-34, ISSN 1312-2622
[30] S. Stoyanov, M. Trendafilova, E-Catalogue "Culture-historical heritage and nature objects in Bulgaria", Conference "New education technologies", 16-17 May, 2003, Sofia, 289 – 298 (in Bulgarian)
[31] S.Stoyanov, M. Trendafilova, CBR-Search in Electronic Catalogues, In Proc. of the International Conference "Automatics and Informatics '03", vol.1, Pp.65-68, 6-8 October, 2003, Sofia, Bulgaria. ISBN 954-9641-34-1
[32] S.Stoyanov, N.Govedarova, I.Popchev, CBR-based Search in BULCHINO Catalogue, In Proc CS&P`07, vol. 2, Pp 521-533

ABOUT THE AUTHORS

Nadezhda Govedarova, PhD student, Institute of Information Technologies, Bulgarian Academy of Science, Phone + 885 945 669, E-mail: nadig@abv.bg

Assoc.Prof. Stanimir Stojanov, PhD, E-commerce Laboratory, University of Plovdiv, Phone: +359 888 318 164, E-mail: stani@uni-plovdiv.bg

Academician Ivan Popchev, Dr.Sc., Institute of Information Technologies, Bulgarian Academy of Science, Phone: +359 2 979 32 33, E-mail: ipopchev@iit.bas.bg