

A Case Based Reasoning Approach for Development of Intelligent Services

Stanimir Stojanov, Ivan Popchev, Daniela Chaushkova, Malina Trendafilova

Abstract: *A specific electronic catalog named "Cultural, historical and natural objects of Bulgaria" (BULCHINO) is developed in the E-Commerce Laboratory of the University of Plovdiv. In this paper we present a reengineering approach for the development of an intelligent electronic service. Furthermore two Case-Based Reasoning (CBR) models, which are implemented in the service, are described. The service offer tourist routes, which include cultural, historical and natural objects (selected from the catalog). The service will be integrated in the next version of the catalog.*

Key words: *Case-Based Reasoning, Case Retrieval Nets, Electronic Catalogues, Electronic Services.*

INTRODUCTION

In the last few years the number of the electronic catalogs, accessible over the Internet, was growing rapidly. The electronic catalogs are an essential ground of the most information systems that provide some electronic services over the communication networks. The variety of domains are represented by these systems, e.g. e-commerce, e-learning, e-government etc.

A specific electronic catalog named "Cultural, historical and natural objects of Bulgaria" (BULCHINO) is developed in the E-Commerce Laboratory of the University of Plovdiv. On the territory of Bulgaria there are over 40 000 significant cultural and historical monuments. Our aim is to build a technological environment by which we can embrace this wealth in electronic formats. Descriptions of more than thousand objects with fundamental relevance for our cultural and historical inheritance are saved in the first catalog version. BULCHINO have been integrated in a broad environment where the users have a free Internet access to the public parts of the catalog.

We intended to extend the existing catalog with new intelligent services which can be used for offering of cultural tourist routes. To do this we saw two possibilities: (i) direct implementation of the service guided by the definition of the abstract model or (ii) reengineering of an existing application, which provides a similar functionality. The choice of the latter approach was motivated by the following observation:

- Firstly, the obtaining of experience in the field of software reengineering was a topic of a joint project between the Humboldt University of Berlin and the University of Plovdiv. Furthermore a CBR-based system named TestManager has been implemented in the Humboldt University and we could take the opportunity.
- Secondly, we would like to prepare a new version of the CBR module implemented in Java (the TestManager was implemented in C++).
- Thirdly, we preferred to work with a real system that is proven in praxis instead with the abstract description of the model.

In this paper we present a reengineering approach for the development of an intelligent electronic service. Furthermore two Case-Based Reasoning (CBR) models, which are implemented in the service, are described. The service offer tourist routes, which include cultural-historical and natural objects (selected from the catalog). The choice criteria are extracted from the users' requests sent to the system. The service will be integrated in the next version of the catalog.

CREATION OF NEW GENERATION ELECTRONIC SERVICES

I. 'BULCHINO' ARCHITECTURE

The architecture of the first catalog version consists of four layers [19,20]:

- *User interface – the user interface, designed as an intelligent portal by using the JetSpeed framework [9], provides different services to the catalog users. The users do not have direct access to the information sources of the catalog. Instead they can*

only activate the selected service that acts as a mediator between the user and the system.

- *Catalog services* – the suggested catalog services are maintained in this layer. Moreover the interfaces to the back-end applications (integrated into the catalog) are arranged here as well.
- *Meta-catalog* – the meta-catalog includes the classification model of the subject catalogs (designed in accordance to the UNESCO definition) and the profiles of the subject catalogs, services and users. The meta-catalog is unique for the whole system.
- *Subject catalogs* – catalogs, kept as separate databases towards the classification model.

A significant difference from the “classical” e-commerce applications is the independence of the services from catalog architecture and application domains. The services can be customized, parameterized, and adapted to meet the individual requirements of the users.

II. BASIC CASE RETRIEVAL NETS

A. Brief Overview

The basic idea of Case-Based Reasoning [23] is to solve new problems by comparing them to old problems that already have been solved in the past. Figure 1 shows the common CBR approach for finding a solution to a new problem, where P_{new} is the problem to be solved, P_{old} is a well-known problem, S_{new} indicates the required solution, and S_{old} is the solution of P_{old} (known as well). A key assumption by the CBR is that if the presentations of two problems are similar, then solutions to these problems are often similar as well. In this way we define two operators: *sim* – a heuristic function of the similarity measures, and *adapt* – a generator procedure, which generate (by adapting) the new solution from the old one.

The life cycle of the CBR-approach can be represented by the following control rule: **if** $sim(P_{new}, P_{old})$ **then** $S_{new} = adapt(S_{old})$. A pair (P,S) is called a *case*. The reasoning with cases remains in the same domain and there are no transformations of the solution space activated.

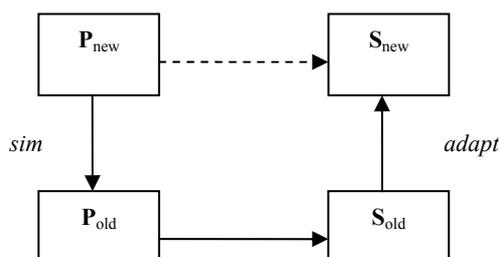


Figure 1. The CBR approach

There are different theoretical models that support the approach shown in Figure 1 [1,5,12]. By reasons mentioned above we have choice the Basic Case Retrieval Net (BCRN) model as a basis of our work.

B. BCRN Definition

A BCRN is a structure that consists of the following components:

IE – a finite set of *Information Entities Nodes*

CN – a finite set of *Case Nodes*

sim – a similarity function, *rel* – a relevance function, *prop* – a propagation function.

The performing of a request in a BCRN is realized by the next three main steps:

- Activating the IEs given in the request;

- Similarity propagation within the net of IEs;
- Relevance propagation to the associated case nodes.

A detailed description of the BCRN-model can be found in [3,6,14].

C. Routes Definition

We envisage the *route* as a structure whose elements are selected from the catalog based on some criteria. We distinguish the following kinds of routes:

- *Route* – a set of selected objects.
- *Structured route* – a route extended by a traversing scheme.

The criteria for the object selection and the information for determining the traversing scheme can be obtained from the users' requests, the meta-catalog, and the subject catalogs. In [19] a principled possibility for the presentation of the route objects in form of BCRN is shown. In order to implement the new intelligent service which support the generation of cultural routes on the basis of the electronic catalog BULCHINO we propose two extensions of the CRN model - the CR model and the the SCR model, which are described in the next sections.

III. CR MODEL

The CR model can be seen as a BCRN, where the local similarity is obtained by means of the meta-linguistic rules defined in this section. This model can be used for the implementation of a generator of cultural routes in BULCHINO.

A. Feature Categories

While the BCRN model is highly useful for investigating formal properties of a CBR-based retrieval approach, in case of a real application we need additional information which allows us to expand the net and to define a more flexible similarity relationship between the IEs. In many application domains the features describing a case can be grouped in three categories [15]: features with functionally describable relationships, features with extensionally describable relationships, features with internally describable relationships.

B. Meta-Linguistic Similarity Specification

There are application domains, where the features can't be associated with no one of the three categories. For example, in such a domain, the similarity of some features can be described functionally (i.e. them can be assigned numerical values) while the similarity of another features (in the same domain) has to be specified for each pair explicitly. In this case we need to be aware of how to compare the numerical values with the phrases of a natural language. The electronic catalog of the Old Bulgarian churches is an example for such a domain. The feature "building_time" that is specified in the domain can hold the following values: "about 1395", "XIV - XV c. (century)", "1495", "beginning of the XVII c.", "ending of the XVII c.", "the first half of the XVII c.", "the second half of the XVIII c.", "the middle of the XVI c".

In order to obtain the similarity between such values we need two functions: a distance function (for the numerical values) and a function that can get the semantic similarity between the textual phrases. A possible solution can be the tabular presentation of all possible pairs put together with their similarity values. We have adopted an approach that combines a linguistic-based model and a numerical comparison function. The idea of our approach is that the domain values will be initially processed and after that linguistic compared by help of meta-linguistic rules. If it is necessary during the linguistic comparison the numerical comparison function can be called. An important problem is the way of extracting of the meta-linguistic rules. The extracted rules from the electronic catalog of the Old Bulgarian churches are shown in Figure 2 in a meta-language (BNMF). The extracting process follows three steps: lexical analyzing, syntax analyzing, expansion of the rule set, generalization of the rules.

```

creationTime ::= startExpression | endExpression | middleExpression | intervalExpression | impleDecDigit
| simpleRomDigit ;

startExpression ::= 'Началото на' | 'Първата половина на' simpleDecDigit | simpleRomDigit ;
endExpression ::= 'Края на' | 'Втората половина на' simpleDecDigit | simpleRomDigit ;
middleExpression ::= 'Около' simpleDecDigit | simpleRomDigit ;
intervalExpression ::= simpleDecDigit '-' simpleDecDigit | simpleRomDigit '-' simpleRomDigit;

simpleDecDigit ::= lexDecDigit (e('.')?);
simpleRomDigit ::= lexRomDigit (e('.')?);

```

Figure 2. Example meta-linguistic rules

IV. SCR MODEL

By means of the CR model we can generate a route as a set of preferred objects. Usually the cultural routes only don't satisfy the users' requirements, i.e. the users would like to receive structured routes. The SCR model is an extension of the CR model, where a new microfeature level is integrated. This model can be used for the implementation of a generator of structured cultural routes in BULCHINO. Different extensions of the basic model are possible: Integration of Concept Nodes [4,16], Spreading Activation Nets [16], Rule-extended Corns [15], Integration of Object Models [17], Microfeature Corns (MFCRN)[16].

SCR Model- a Specific MFCRN

The establishment of the domain values as feature, which present the objects' characteristic directly, and as microfeatures, which hold additional information, is a suitable way for our application. The values we use for the obtaining of the route objects in the CR model can be specified as features. Each route object can be described by information for the geographical location and the region, the landscape, the reach possibilities (by car, by train, by plane), the leisure time facilities as well. The additional data can be specified as microfeatures.

In order to implement the above approach we extend the CR model, where a new level of processing is introduced. This level implements an arrangement scheme of the route objects, which uses background knowledge in form of microfeatures. The microfeature nodes of the "classical" MFCRNs can activate the IEs, which propagate the activation further to the cases along the relevance arcs. For our purpose the search activation in the microfeature nodes is of little use (the geographical location of an old Bulgarian church doesn't matter if the tourist shouldn't like to visit churches). We propose that the obtaining of a structured route achieves in two steps: firstly, we generate a route and then we start a secondary retrieval process in the microfeatures world. The scheme at the Figure 3 illustrates how the SCR model works. A case contains a set of information entities like *type*, *building_time*, *area*. Asking for a route which objects should be of type *church* and created 1844 year in the region *Plovdiv*, by similarity and relevance arcs are activated the cases: *St. Nikolaous church*, *St. Mary church* and *Bachkovo cloister*. A set of microfeature nodes are connected to IE nodes through importance arcs and are still not activated. After the retrieval the microfeature nodes are activated by the activated IEs and the route objects are ordered according to the new activation from the microfeature nodes.

V. OUR APPROACH

Into the new version of the catalog we intend to integrate services that will provide more flexible and intelligent opportunities for retrieval of information from very large databases, not only standard database approaches for storage and retrieval of objects.

In order to develop the routes offering service we propose an approach, which includes the following steps:

- Service kernel (BCRN model) implementation;

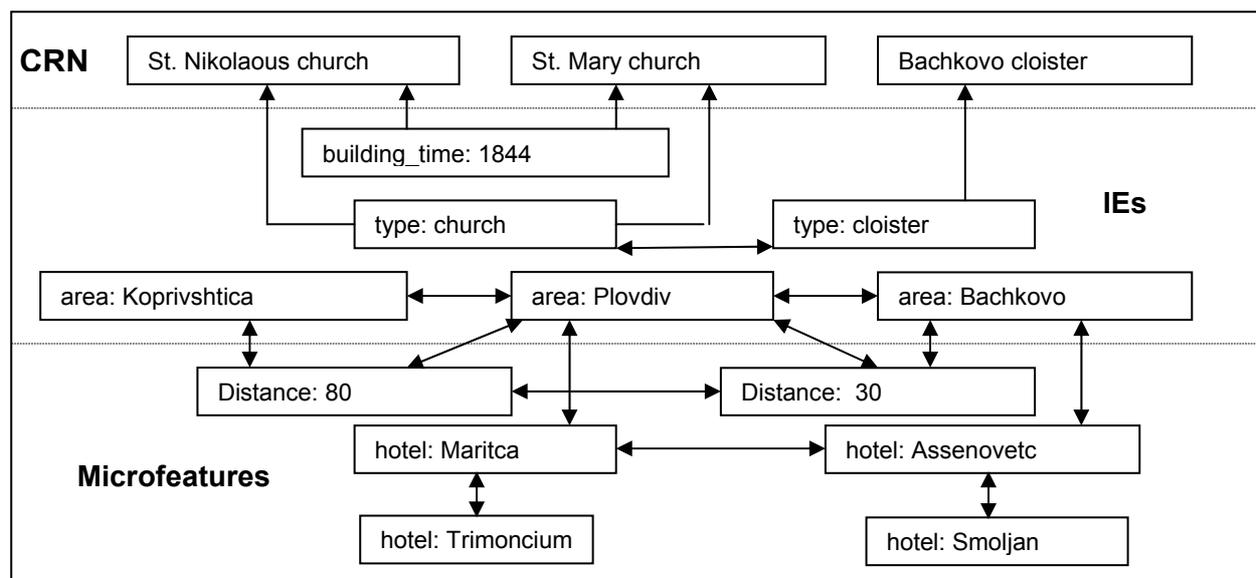


Figure 3. Example SCR – Search

- CR model development;
- SCR model development;
- Integration of the service in the catalog.

We intended to develop the BCRN kernel by reengineering of the existing TestManager [13]. On one hand, the TestManager is rather a textual CRN-based application as a direct implementation of the BCRN model (the BCRN implementation is the kernel of the TestManager) [18]. On the other hand, BULCHINO is a multi-layered application, which is integrated in a J2EE-based environment [21]. In order to keep the homogeneity of the system we needed a Java version of the BCRN model. The service kernel was extended further by new Java classes, which support the meta-linguistic similarity calculation (CR model) and the microfeatures processing (SCR model).

Furthermore we intend to integrate the service into the existing catalog version. We require that the integration have to satisfy the restriction resulted from the catalog architecture [7,8] which must remain unaffected. The next problem is that there isn't an explicitly set of IEs – the objects attributes are saved in form of relational structures (relational database), where they are split on different tables. By taking into account all these circumstances we propose the following integration approach: There are distinguished two kinds of objects: *route objects* and *catalog objects*. The route objects can be generated from the catalog objects by means of appropriate transformations. An explicit persistent layer is supported that (i) transfers data to the net from the database and vice versa and (ii) accomplishes transformations between the two spaces.

CONCLUSIONS AND FUTURE WORK

In the paper we show two possible models for developing of CBR-based intelligent services. In order to implement a real software version of a service we have to process the users' requests as well. We intend to implement a request analyzer, which will be integrated in the BULCHINO portal. The analyzer will extract the values, which can be used for the initialization of the cultural routes offering service.

REFERENCES

- [1] Aamodt, A., Plaza, E., Case-base reasoning: Foundational issues, methodological variations, and system approaches, *AI Communications*, 7(1), 1994, 39-59
- [2] Anderson, J. R., A theory of the origins of human knowledge, *Artificial Intelligence*, 40, 1989, Special Volume on Machine Learning, 1989

- [3] Burkhard, H-D. Case Retrieval Nets, Techn. Report, Humboldt University, Berlin, 1995
- [4] Burkhard, H-D. Cases, Information, and Agents. In: P. Kandzia, M. Klusch (eds), Cooperative Information Agents, Proceedings First Int. Workshop CIA'97, Lectures Notes in Artificial Intelligence, Vol 1202, Springer-Verlag, Berlin, 1997, 64-79
- [5] Burkhard, H-D. Extending some Concepts of CBR – Foundations of Case Retrieval Nets. In: Case-Base Reasoning Technology, From Foundations to Applications, Springer-Verlag, Berlin, 1998, 17-50
- [6] Burkhard, H-D. Information Gathering for Vague Queries Using Case Retrieval Nets. In: I. Baldderjahn, R. Mathar, S. Martin (eds), Classification, Data Analysis, and Data Highways, Studies in Classification, Data Analysis, and Knowledge Organization, Springer-Verlag, Berlin, 1998, 345-354
- [7] Hadjikolev, E. The Architecture of the Distributed eLearning Center, Scientific and practical conference “The new education technologies”, May, 16-17, 2003, Sofia (in Bulgarian)
- [8] Hadjikoleva, S., E. Hadjikolev, The Services in a Distributed E-Learning Environment, Scientific and practical conference “The new education technologies”, May, 16-17, 2003, Sofia (in Bulgarian)
- [9] <http://jakarta.apache.org/jetspeed/> (to date)
- [10] <http://www.omg.org/techprocess/meetings/schedule/adopt.html> (to date)
- [11] J2EE: <http://java.sun.com/j2ee>
- [12] Genter, D., Forbus, K. D., MAC/FAC: A model of similarity-based retrieval. In: Proceedings 13th Annual Conference of the Cognitive Science Society, 1991, 504-509
- [13] Kunze, M., TestManager, Software-Dokumentation, September, 1999, www.informatik.hu-berlin.de
- [14] Lenz, M, E. Auriol, Miming. Diagnosis and Decision Support. In: Case-Base Reasoning Technology, From Foundations to Applications, Springer-Verlag, 1998, 51-90
- [15] Lenz, M., Burkhard, H. D., Case retrieval Nets: Foundations, Properties, Implementation, and Results, <http://www.informatik.hu-berlin.de/>
- [16] Lenz, M., H.D. Burkhard, S. Brueckner, Case Retrieval Nets: Basic Ideas and Extensions. In: K. Goerz, S. Hoelldobler (eds), KI-96: Advances in Artificial Intelligence, Lectures Notes in Artificial Intelligence, Vol 1137, Springer-Verlag, Berlin, 1996, 227-239
- [17] Lenz, M., H.D. Burkhard, S. Brueckner, Applying Case Retrieval Nets to Diagnostic Tasks in Technical Domains. In: I. Smith, B. Faltings (eds), Advances in Case-Based Reasoning, Lectures Notes in Artificial Intelligence, Vol 1186, Springer-Verlag, Berlin, 1996, 219-233
- [18] Lenz, M., Huebner, A., Kunze, M., Textual CBR, In: Case-Base Reasoning Technology, From Foundations to Applications, Springer-Verlag, Berlin, 1998, 115-137
- [19] Stojanov, S., M. Trendafilova, CBR-Search in Electronic Catalogs, International Conference “Automatics and Informatics'03”, October 6-8, 2003, Sofia, 65-68
- [20] Stojanov, S., M. Trendafilova. Electronic Catalog BULCHINO, Scientific and practical conference “The new education technologies”, May, 16-17, 2003, Sofia
- [21] Sutcliffe, R. F., Representing meaning using microfeatures, In: Reilly, R. G., Sharkey, N.E. (eds.), Connectionist Approaches to Natural Language Processing, Lawrence Erlbaum Associates.
- [22] Vollrath, I., W. Wilke, R. Bergmann. Intelligent Electronic Catalogs for Sales Support, <http://www.wagr.informatik.uni-kl.de/>
- [23] Wess, S., K.D. Althoff, G. Derwand. Using K-D Trees to Improve the Retrieval Step in Case-Base Reasoning. In: S. Wess, K.D. Althoff, M.M. Richter (eds), Topics in Case-Base Reasoning, Springer Press, 1994

ABOUT THE AUTHOR

Acad., Prof. Ivan Popchev, PhD, Institut of Information Technologies - BAS, Phone: 716 851; 979 3233, E-mail: ipopchev@iit.bas.bg.

Assoc. Prof Stanimir Stojanov, PhD, Plovdiv University "Paisii Hilendarski" – holder of chair Computer Systems , E-mail: csstani@pu.acad.bg

Daniela Chaushkova, PhD Student, Plovdiv University "Paisii Hilendarski" – master of Informatics, E-mail: daniela_chaushkova@yahoo.com

Malina Trendafilova, PhD Student, Plovdiv University "Paisii Hilendarski" – master of Informatics, E-mail: Malina.Trendafilova@ecl.pu.acad.bg.