INTELLIGENT TECHNIQUES FOR DATA INTEGRATION AND DECISION SUPPORT IN THE MEDICAL DOMAIN

Mirjana Ivanović[1], Hans Dieter Burkhard[2], Vladimir Kurbalija[1], Miloš Radovanović[1]
Danilo Schmidt[2], Gabriela Lindemann – v. Trzebiatowski[2], Marija Semnic[3]

[1]Department of Mathematics and Informatics, Faculty of Science
Trg D. Obradovića 4, 21 000 Novi Sad, Serbia
{mira, kurba, radacha}@dmi.uns.ac.rs
[2]Humboldt University of Berlin, Institute of Computer Science
Rudower Chausse 25, 10099 Berlin, Germany
hdb@informatik.hu-berlin.de, danilo.schmidt@charite.de, gabriela.lindemann@uv.hu-berlin.de
[3]Institute of Neurology, Medical School Novi Sad
H. Veljkova 1–3, 21 000 Novi Sad, Serbia
memoryns@gmail.com

ABSTRACT

Data commonly produced during medical practice often falls into the category of complex data, which may not comply with the traditional field structure of common types like alphanumeric, dates, etc., with prominent examples including text documents, time-series, and multimedia objects. In addition, medical data may not be mutually compatible and easy to integrate into a centralized repository, since it originates from heterogeneous sources which apply different assumptions, conventions and schemas. Having in mind these challenges, two research teams from Humboldt University, Berlin, Germany and University of Novi Sad, Serbia, organized the joint research project "Intelligent Techniques for Data Integration and Decision Support in the Medical Domain," with the aim of investigating and applying different techniques for reasoning, mining and retrieval to problems recognized in real-world scenarios within the medical domain. This paper describes the initial research efforts of the two groups towards exploring how to overcome difficulties in organizing different kinds of medical data, incorporating case-based reasoning/time-series techniques in the implementation of a reliable general-purpose decision-support framework, and generating special-purpose decision-support systems from the general framework. We expect the proposed research to provide scientific results and facilitate the implementation of appropriate intelligent software tools applicable in different domains as standalone applications, but also as components that can be integrated into already existing information systems and environments.

KEYWORDS

Case-based reasoning, time series, data integration, decision support, information systems, medicine.

1. INTRODUCTION

Data produced during medical practice often falls into the category of complex data, which may not comply with the traditional field structure of common types like alphanumeric, dates, etc. Prominent examples of complex data include text documents, time-series (i.e., sequences of data points measured at successive times), and multimedia objects (images, sound, etc.). Even in the cases where medical data conforms to classical database structures and schemas, the data may not be mutually compatible and easy to integrate into a centralized repository, since it originates from heterogeneous sources which apply different assumptions, conventions and schemas. Due to their usefulness, complex data are prevalent in many contemporary medical information systems (MIS). Management of complex data and integration of traditional data from different sources into a common database present numerous research challenges.

Having in mind the aforementioned challenges, background, and previous long-lasting cooperation between the research teams from Humboldt University, Berlin, Germany and University of Novi Sad, Serbia, in the domain of development and application of artificial intelligent techniques, the two teams proposed and received support for the joint research project "Intelligent Techniques for Data Integration and Decision Support in the Medical Domain" (2011–12).

The main objective of this project is investigation and application of different techniques for reasoning, mining and retrieval to problems recognized in real-world scenarios in the medical domain, stemming from the need to integrate data from different sources into a centralized repository and facilitate decision support, diagnosis and other analytical tasks. Different artificial intelligence (AI) methods have been applied to mentioned problems in the past (Yang and Wu 2000, Zhuang et al 2009). However, many methods generally have difficulties concerning large amounts of data, high dimensionality, the semantic gap, etc. Within the project we aim to find ways to overcome such difficulties and propose an appropriate general-purpose integrated framework that will offer and support quality decision actions in medical domains. It is planned to be realized as a stand-alone framework that will facilitate production of special-purpose decision support systems (DSS) for particular medical domains and usage. The produced special-purpose DSS can also be seen as components that can be incorporated into the wide spectrum of already existing MIS, bringing into them new flavor and functionalities.

The rest of paper is organized as follows. The second section describes the core background as a starting point for joint research of the two teams. In the third section, main joint research directions and necessary activities are presented. The architecture and essential functionalities of a general-purpose decision support system as a fundamental contribution of the joint research is presented in the fourth section. The fifth section concludes the paper and provides insights into planned future activities.

2. BACKGROUND

We are witnesses to the fact that retrieval and analysis of complex data, as well as integration of structured data from heterogeneous sources, are rapidly developing research fields whose results have numerous practical applications. In the context of the medical domain, with the growth of information gathered about patients, diseases, and other medical factors, it is required to have systems that can discover relevant information and analyze volumes of available medical data. In order to successfully perform these tasks, however, MIS need to be able to process not only large volumes of data, but meaningfully integrate data from different sources and deal with unstructured or semi-structured types of complex data such as text, time series and image data. Having the abovementioned in mind, two research teams decided to join efforts with the intention of considering the possibility to integrate several intelligent techniques into a unique framework applicable to different medical domains.

In the last decade both research teams, from the University of Novi Sad and Humboldt University, performed research and published papers on different AI techniques and their possible applications (Hildebrand et al 2008, Hildebrand et al 2009, Lindemann et al 2007, Schmidt and Lindemann – v. Trzebiatowski 2011, Kurbalija and Ivanović 2005, Kurbalija et al 2007, Kurbalija et al 2009). They have experience in applying AI techniques in the medical domain as well. Case-based reasoning (CBR) was recognized as the most appropriate technique by both teams, since knowledge of medical experts consists of a mixture of textbook knowledge and experience, which can be stored in the form of cases.

The German team possesses extensive experience in applying case-based reasoning within the nephrology and pathology domains (Gestewitz and Lindemann 2009, Lindemann et al 2007). Furthermore, application of ontologies as a suitable concept for organizing data from heterogeneous sources is also studied by this team (Gestewitz and Lindemann 2009). During the realization of several projects with Charité Hospital in Berlin they gathered experience and routine necessary to acquire, organize and use specific medical data. Also, they acquired experience in processes like: knowledge acquisition from medical experts, transformations and representations of collected knowledge, and communication with medical experts (Lenz et al 1998, Schmidt and Lindemann – v. Trzebiatowski 2011).

The Serbian team has experience in applying CBR in real-world environments (Kurbalija and Ivanović 2005, Kurbalija et al 2009). They developed an appropriate decision-support system and applied it on data relevant to the multiple-scleroses disease (Kurbalija et al 2007). This team also performed research in different fields of AI applicable in medicine, including: text mining (Radovanović et al 2009), time-series analysis (Kurbalija et al 2010, Radovanović et al 2010b), analysis of high dimensionality (Radovanović et al 2009, Radovanović et al 2010a), etc.

3. MAIN JOINT RESEARCH DIRECTIONS

Analyzing previous research experiences of both teams we decided to join efforts and examine issues that are essential for these rapidly evolving research fields, as well as explore the practical applications of these results in the medical domain. The main research focus has been placed on:

- Investigating how to overcome difficulties in organizing in appropriate way huge amount of different kinds of medical data (numbers, text, time-series),
- Incorporating case-based reasoning/time-series techniques in implementation of reliable general-purpose decision-support framework,
- Improvements of the performance of resulting special-purpose decision-support systems (generated from previously developed general one) in several medical fields (Internal medicine – with special emphasize on Nephrology and Hematology, and also in Pathology, and Neurology),
- Usage of data mining and ontologies to improve results of more complex data analyses.

Based on rigorous theoretical and empirical analysis we aim to perform research and obtain scientific results that will give insights into the process of medical knowledge acquisition, reasoning and different decision support tasks like diagnosis, and, eventually, therapy planning, etc. We also expect that joint research and results can provide benefits to a wide spectrum of medical information systems, not just those limited to the fields of nephrology, pathology and neurology (in which the German team has experience), since they will provide a deeper insight into the process of knowledge acquisition, reasoning, and various common medical data analysis tasks.

From the practical point of view, we are planning to produce several software products/support tools:

- General-purpose decision support framework which incorporates techniques for case-based reasoning and time-series analysis (in the future considering data mining and ontologies),
- Special-purpose system(s) (based on the abovementioned framework) for medical data analysis and decision support in diagnosis.

Our joint intention for the realization of these tools will be performed through the following activities:

1. **Analysis of the wide spectrum of heterogeneous data appearing in medical domains, in order to identify characteristic ones.** It is necessary to identify general attributes of medical analysis and patient records. For crucial data, their attributes and particular values, an appropriate data representation needs to be proposed in order to facilitate application of sophisticated intelligent techniques.
2. **Analysis of essential functionalities for different intelligent techniques and ways of their integration into the general-purpose decision support framework.** An important aspect of any integrated medical data repository is support for fast and efficient retrieval and reasoning. In addition, the general-purpose framework needs to provide functionalities for data acquisition from different kinds of structured or unstructured sources which, e.g., involve techniques for time-series analysis, data mining, and ontology merging.
3. **Proposition of a general-purpose decision support framework and its implementation.** The German team has experience in implementation of medical systems coping with huge volumes of medical data and patient records. The Serbian team implemented an intelligent tool employing case-based recognition that is applicable in different domains. As a consequence, a general-purpose decision support framework that facilitates different medical tasks needs to be proposed and implemented.
4. **Implementation of two special-purpose decision support systems.** For the realization of these systems it will be important to support integration of data from different sources into a centralized repository and facilitate decision support, diagnosis and other analytical tasks for particular domains.

4. GENERAL-PURPOSE DECISION SUPPORT FRAMEWORK

Isolated application of either CBR or data mining cannot fully achieve the aims for intelligent decision support for medical domains (Zhuang et al 2009). While CBR is concerned with how to use previous knowledge to solve a new problem, data mining is dealing with discovering essential relations and knowledge from data. It seems that these two approaches can complement each other to better meet the requirements for intelligent decision support systems. Data-mining techniques have previously been combined with CBR (Yang and Wu 2000, Clerkin et al 2002, Arshadi and Jurisica 2005). Generally, the underlying philosophy of our proposed integrated approach is to establish a patient-centric knowledge repository of past experiences to support problem solving in new cases. In particular, once relevant knowledge is extracted using data-mining techniques, it can be retrieved and reused by the CBR process.

On the other hand, concerning the representation of relevant scientific (including medical) data, time-series appears as a very important and highly applicable concept. In different scientific fields, a time-series consists of a sequence of values or events obtained over repeated measurements of time. Analysis of time-series comprises methods that attempt to understand them, often either to understand the underlying context of the data points, or to make forecasts. Time-series databases are popular in many applications including medical treatments, etc. As a consequence, in the last decade there has occurred an increasing amount of interest in querying and mining such data, which resulted in a large amount of work introducing new methodologies for different task types, like: indexing, classification, clustering, prediction, segmentation, anomaly detection, etc.

Having in mind mentioned contemporary research in medical informatics, and also research activities and results that both teams conducted in last decade, the central result of our joint research is expected to be the proposal and implementation of a general-purpose decision support framework which incorporates the techniques of case-based reasoning, and time series analysis. Other aspects like data mining and ontologies will be considered in next phase of the project and will be realized as separate supporting components for special-purpose decision support systems.

The general-purpose decision support framework is the essential starting point for further results and development of special-purpose decision support systems for different medical domains of interest to both teams (at the moment Hematology, Nephrology and Neurology). In this way, the produced particular special-purpose systems could be used and incorporated as separate components in already existing medical information systems. Integration of such components into existing systems will bring new functionalities and increase their quality and operability. In the rest of this section, the general structure and basic components of the framework will be presented.

During the last several years, the group from Novi Sad developed the decision support system based on CBR technology – *CaBaGe* (Kurbalija et al 2005). One of the main advantages of this system is its domain independence. The input for the system consists of two parts: one is the description of attributes of the case and another is the database of already solved cases from a domain. On the basis of this data, the system creates a special memory structure – the Case Retrieval Net (CRN) (Lenz et al 1998) and is capable of solving new problems, i.e., cases (proposing solutions) from the domain using the already existing input database of previous cases.

The second important advantage of the *CaBaGe* system is its highly efficient memory organization – CRN. CRNs represent a memory model that has been developed for efficient retrieval in large case bases. Functionality of the newly proposed general-purpose decision support framework is crucially based on the CRN structure which will be briefly described here. CRN has nodes for each case in the database and nodes for each value of every attribute – these values are called information entities (Figure 1). There exists an acceptance arc between every two corresponding information entities (information entities of the same attribute). Also, there exists a relevance arc from information entity nodes to the case nodes, if the corresponding information entity is relevant to the case. All the arcs in the network are weighted.

At the beginning of the problem-solving process, the user needs to enter his observations in the form of a query, which consists of values of existing attributes. For every attribute, a special numeric value called importance is defined. This value contains the information of how much is the corresponding attribute important for the problem-solving process. High values indicate high importance, while lower values indicate lower importance.

The retrieval process consists of evaluating acceptance values for each case. Acceptance values are computed by spreading activation in the network as follows: Information entity nodes are initially activated by the value of importance. The computation is performed by propagating activations along the acceptance arcs to further information entity nodes and from all activated information entity nodes over relevance arcs to case nodes. The acceptance value for every case is obtained by summing the activations gained from all information entities trough relevance arcs. The case node or nodes with highest activation are proposed as the solution of the actual problem. The previously described problem-solving method ensures efficient similarity-based problem resolution in only two steps regardless of the database size.

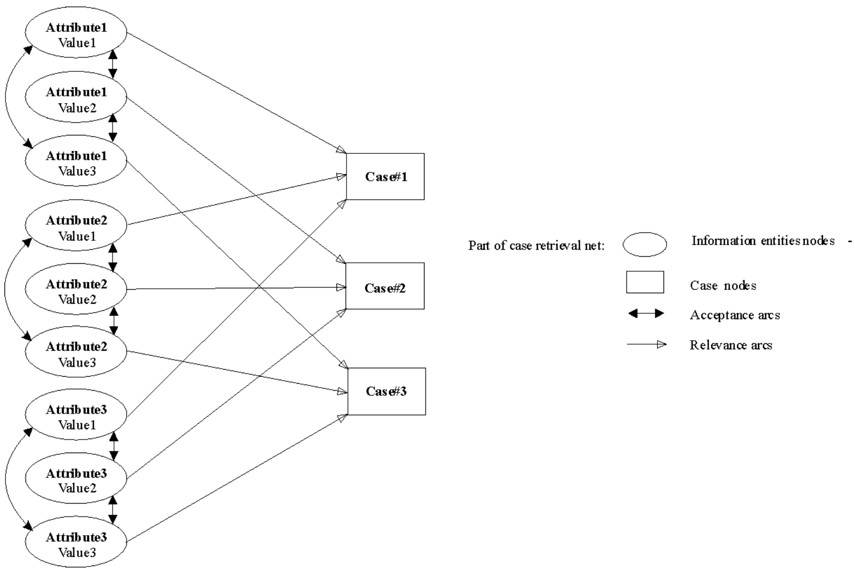

Figure 1. Part of a Case Retrieval Net

The process of CRN construction is out of scope of this paper. More information is provided by Lenz et al (1998). At this point it is important to emphasize that the weights of the relevance arcs are computed straightforwardly according to the algorithm explained by Lenz et al (1998), while the weights of acceptance arcs are not so formally covered. These weights (also known as local similarities) are essential for the functionality of CRNs because they are expressing the underlying similarities between different values of the same attribute. Local similarities can be easily calculated for ordered symbolical and numerical attributes, but for the more complex data this task is especially challenging.

In its original form, the system *CaBaGe* was designed only to work with attribute-value data. However, the main objective of the proposed general-purpose decision support framework is to provide functionalities for the complex data, which at this stage mainly includes the support for time-series.

Another aspect of our recent research which is crucial for the proposal of the general-purpose decision support framework is connected to deeper and more qualitative time-series analyses. There are three important concepts which need to be considered when dealing with time series: *pre-processing transformation*, *time-series representation* and *similarity/distance measure* (Kurbalija et al 2010).

Being motivated to improve and make easier processes of time series analyses, we have designed a multipurpose, multifunctional system FAP – Framework for Analysis and Prediction. FAP is under active development, and aims to support all mentioned concepts: representations, similarity measures and pre-processing tasks; with the possibility to easily change some existing or to add a new concrete implementation of any concept. At the moment we have implemented all major similarity measures for time series (15 different variations of 6 main similarity measures) (Kurbalija et al 2010), and conducted a set of experiments to validate their correctness. Such a system can be easily incorporated and used in our general-purpose decision support system for application of case-based reasoning in medical domains.

Since the existing system *CaBaGe* already efficiently manipulates with attribute-value data by activation and propagation processes through the CRN, it would be natural to extend *CaBaGe* with functionality appropriate to time series. The most fundamental aspect of CRNs is the correct defining of local similarities between information entity nodes which is already implemented for attribute-value pairs in *CaBaGe*. As previously mentioned, all important up-to-date similarity measures for time-series are already implemented in our framework FAP. Therefore, the most convenient method for successful management of complex data would be the integration of FAP similarity measures with the existing system *CaBaGe*. To achieve this, it will be necessary to introduce a new type of information entity nodes as time-series nodes. The local similarities for such kinds of nodes will be calculated using the similarity measures provided by the FAP framework, while the other functionalities of the *CaBaGe* system (including CRN) will stay untouched.

With this approach, time-series data will be represented in the CRN in the same manner as attribute-value data – as nodes in the CRN. Furthermore, one of the main advantages of this approach is the variety of similarity measures for time-series that will provide the user of the newly proposed framework with the opportunity to test and validate different similarities appropriate for his/her domain and task type.

The structure of the general-purpose framework that we are proposing in our joint research is given in Figure 2. The framework integrates two existing components: *CaBaGe* (Kurbalija and Ivanović 2005) and FAP (Kurbalija et al 2010) and manages complex data in the form of attribute-value pairs and time-series. The system *CaBaGe* is responsible for the main retrieval algorithm based on CRNs, while the framework FAP offers its sophisticated similarity measure algorithms for the local similarities of time-series data.

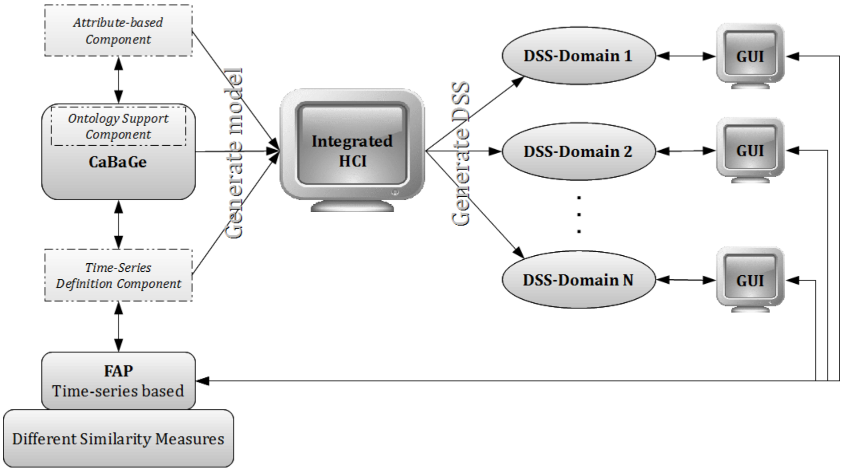

Figure 2. Structure of general-purpose decision support framework

In Figure 2, the *Attribute-based Component* and *Time-Series Definition Component* are responsible for correct manipulation of the corresponding data, while the *Ontology Support Component* is intended to be a formal model for definition of input data. These components, together with *CaBaGe* and *FAP*, are required for the generation of the exact model of a new decision support system from an arbitrary domain.

The input for the system is two-fold: the user needs to provide the description of the case and the database of already solved cases from the domain for which the special-purpose system is planned. The description of the case should contain the specification of the attributes important for the intended domain, and the specification of the temporal data represented as time series. The names, number and types of all significant attributes should be specified in the *Attribute-based Component.* The *Time-Series Definition Component* is responsible for the definition of names, number and types of needed temporal data, but also for some other important concepts for analysis and specification of time-series data like: offset translation, amplitude scaling, noise removal, equidistant or non-equidistant points, etc.

The database of already solved cases is the most important source of domain knowledge, and it represents the basis for the creation and training of the CRN. This database can be provided in a form of a textual file or as a relational database. The quality of this database is crucial for the performance of the planned special purpose decision support system: the database should contain a large number of various successfully solved cases. In addition, the database should be consistent with the description of the case.

On the basis of the prepared model, the *Integrated HCI* component generates a special-purpose decision support system for a particular medical domain and prepared/existing case base. Each generated system has its own GUI which is based on the specification of the system (depending on the number and types of attributes). Furthermore, the GUIs of the newly produced systems will be connected with the FAP framework, offering the possibility to change the similarity measures between time series in at run-time through the FAP interface. This property gives the user of the special purpose system the opportunity to test different similarity measures for time series and determine which measure is the most convenient in their domain. Physicians, as potential users of generated special purpose systems, could observe different aspects of temporal data that represent the condition of the patient measured in time intervals. Furthermore, physicians could have different views on the data and the possibility for qualitative analysis of interdependence between different data.

Such an integrated framework, which encompasses *CaBaGe* and *FAP*, offers uniform management of complex data potentially originating from heterogeneous sources, which was our main objective. This approach has the ability to integrate data from different sources into a centralized repository of special-purpose decision support systems for particular medical domains. Once the special-purpose decision support system for a particular medical domain is generated, it represents a good baseline for further improvement and experiments: usage of different similarity measures on the same case base in order to discover higher-quality relationships and interdependences. To obtain better results, some other intelligent techniques can also be applied on the same data, in order to facilitate decision support, diagnosis and other analytical data-mining tasks.

The proposed system inherits numerous advantages:

- **Domain independence.** The CRN (as the basic component of the *CaBaGe* system) calculates similarity measures just on the basis of previous cases from any kind of domain. There is no need for some special domain knowledge (complex relations between attributes of the cases or values in the time-series), nor for the expertise of domain experts.
- **Incremental learning.** The process of acquiring of the new knowledge is extremely simple in both systems. New knowledge is incorporated in the knowledge base by simply saving the successfully solved case in the base of previously solved cases.
- **Platform independence.** Both systems are implemented in Java so they can work equally well on all kinds of platforms.
- **Fast retrieval algorithm (*CaBaGe*).** Case Retrieval Nets are used as a basic memory structure since they have very good retrieval performance. The solution is gained after only two iterations no matter how large the case base is.
- **Versatility and diversity.** The FAP framework offers a variety of similarity measures for time-series data designed for different kinds of problems. This convenience provides the possibility to test different measures and to empirically select the most suitable one for a particular application. Furthermore, FAP will include all important newly-introduced time-series similarity measures which will keep the new system up-to-date with modern trends in time-series analysis.

The proposed general-purpose framework could be useful for many reasons. Both underlying systems are domain-independent. Therefore, the resulting general-purpose decision-support system would also be independent from the domain. Furthermore, it would support the generation of a greater variety of special-purpose decision support systems for particular domains: it will cover the domains where scenarios could be specified with sets of attributes, but also the domains where cases are represented as curves or time-series.

5. CONCLUSION

Although there exist numerous scientific results concerning merging and data integration in general, as well as analysis of weakly structured complex data, they have seldom been rigorously tested and put into practical use within a centralized medical repository which brings together, under a single roof, application of wide range of artificial intelligence techniques. Providing an assessment of which techniques work well, within the context of a particular medical decision support task, will help advance the state-of-the art in the field, especially in the sense of providing guidelines for practical applications.

We expect that our research efforts will produce scientific results that will offer insight into the processes of knowledge acquisition, integration, and reasoning within the medical domain. Current research activities and the proposed framework bring several promising advantages:

1. The general-purpose decision support framework provides medical (but non-software) experts with several activities. He/She can easily and in a user-friendly environment realize the following:
   a) Define and model patterns for patient records (i.e., cases), specifying appropriate attributes and their referential values, time-series, defining relationships and similarities between them,
   b) Enter into a dedicated database real patient records (cases) and prepare the real case-base,
   c) Generate a special-purpose decision-support system for a particular medical domain including the prepared case-base. Using the available FAP system and its functionalities, the medical expert can perform different experiments on patients' case-base employing a wide-range of similarity measures supporting higher-quality data analyses.

2. Integration of the general-purpose decision support framework or special-purpose decision-support system for a particular medical domain into an already existing MIS, bringing functionalities mentioned above.

We expect the proposed and ongoing research to provide scientific results and facilitate the implementation of intelligent software tools applicable in different domains as standalone applications, but also as components that can be integrated into existing information systems and environments.

The proposed general-purpose decision support system can serve as a good starting point to establish a research network and include other research groups to use such systems in everyday practice. As a future step, based on the obtained results, the development of a prototype with key functionalities for therapy planning will be considered. Further down the road, we plan to explore the possibilities for integration of more general data-mining techniques into the framework.

Acknowledgement. The authors gratefully acknowledge DAAD and the Serbian Ministry of Education and Science for support through the bilateral project "Intelligent Techniques for Data Integration and Decision Support in the Medical Domain" (2011–12).

REFERENCES

Arshadi, N. and Jurisica, I., 2005. Data mining for case-based reasoning in high-dimensional biological domains. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 8, pp. 1127–1136.

Clerkin, P. et al, 2002. Automated case generation for recommender systems using knowledge discovery techniques. Trinity College Dublin, Computer Science Department Technical Report, April.

Gestewitz, M. and Lindemann, G., 2009. Creating an ontology for the domains nephrology and pathology into the OpEN.SC project. *Proceedings of the 5th International Conference on Intelligent Computing and Information Systems (ICICIS)*, Kairo, Egypt.

Hildebrand, F. et al, 2008. Application of case retrieval nets in the OpEN.SC project. *Proceedings of the 17th International Workshop on Concurrency, Specification and Programming (CS&P)*, Groß Väter See, Germany, Vol. 1, pp. 218–228.

Hildebrand, F. et al, 2009. A dynamical data-structure for case retrieval nets in the OpEN.SC project. *Proceedings of the 5th International Conference on Intelligent Computing and Information Systems (ICICIS)*, Kairo, Egypt.

Kurbalija, V. and Ivanović, M., 2005. CaseBaseGenerator for intelligent systems. *Novi Sad Journal of Mathematics*, Vol. 35, No. 1, pp. 25–40.

Kurbalija, V. et al, 2007. Multiple sclerosis diagnosis – case-based reasoning approach. *Proceedings of the 20th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, Maribor, Slovenia, pp. 65–71 .

Kurbalija, V. et al, 2009. Case-based curve behaviour prediction. *Software: Practice and Experience*, Vol. 39, No. 1, pp. 81–103.

Kurbalija, V. et al, 2010. A framework for time-series analysis. *Proceedings of the 14th International Conference on Artificial Intelligence: Methodology, Systems, Applications (AIMSA)*. *LNAI*, Vol. 6304, pp. 42–51, Springer.

Lenz, M. et al, 1998. *Case-Based Reasoning Technology: From Foundations to Applications*. Springer.

Lindemann, G. et al, 2007. The system architecture of the Open European Nephrology Science Centre. *Proceedings of the World Academy of Science, Engineering and Technology (CESSE)*, Venice, Italy, pp. 422–427.

Radovanović, M. et al, 2009. Text categorization and sorting of Web search results. *Computing and Informatics*, Vol. 28, No. 6, pp. 861–893.

Radovanović, M. et al, 2010a. Hubs in space: Popular nearest neighbors in high-dimensional data. *Journal of Machine Learning Research*, Vol. 11, pp. 2487–2531.

Radovanović, M. et al, 2010b. Time-series classification in many intrinsic dimensions. *Proceedings of the 10th SIAM International Conference on Data Mining (SDM)*, Columbus, Ohio, USA, pp. 677–688.

Schmidt, D. and Lindemann – v. Trzebiatowski, G., 2011. Roll out of the OpEN.SC project. *Proceedings of the 20th International Workshop on Concurrency, Specification and Programming (CS&P)*, Pułtusk, Poland, pp. 454–464.

Yang, Q. and Wu, J., 2000. Keep it simple: A case-base maintenance policy based on clustering and information theory. *Proceedings of the Canadian AI Conference*. *LNAI*, Vol. 1822, pp. 102–114, Springer.

Zhuang, Z. Y. et al, 2009. Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners. *European Journal of Operational Research*, Vol. 195, pp. 662–675.