

|   |     |
|---|-----|
| RADU BADEA, MONICA LUPȘOR, HORIA STEFĂNESCU, CRISTIAN VICAȘ, DELIA MITREA, MIHAI SOCACIU<br>SERGIU NEDEVSCHI<br>Integrated procedures of evaluating the hepatic morphology, for the non-invasive detection and<br>quantification of steatosis and fibrosis based on the ultrasound diagnosis . . . . .  | 1   |
| SERGIU NEDEVSCHI, CRISTIAN VICAS, MONICA LUPSOR, RADU BADEA, MIRCEA GRIGORESCU<br>The employment of textural and non textural image analysis algorithms in assessing the<br>diffuse liver diseases. . . . .   | 12  |
| RADU BADEA, TEODORA POP, MIHAI SOCACIU, OFELIA MOȘTEANU, C. CARAIANI , MONICA LUPȘOR, PAULA<br>RAICA, MIHAELA MANISOR, LIVIU MICLEA, DELIA MITREA, MONICA LUPSOR, SERGIU NEDEVSCHI<br>The value of the ultrasound image processing procedures, combined with circulation<br>assessment and mathematical modelling techniques for HCC characterization, nature<br>specification and disease prognosis. . . . . | 18  |
| DELIA MITREA, SERGIU NEDEVSCHI, PAULINA MITREA, MONICA LUPSOR, MIHAI SOCACIU, TEODORA-ATENA<br>POP, OFELIA MOSTEANU, RADU BADEA<br>The Imagistic Textural Model of the Hepatocellular Carcinoma. Comparisons,<br>discussions, conclusions . . . . .   | 28  |
| SERGIU NEDEVSCHI, IOAN SALOMIE, RODICA POTOLEA, MIHAELA DINSOREANU, RADU BADEA, ALEXANDRU<br>FLOARES, IOAN COMAN, BOGDAN PETRUT, DELIA MITREA, CAMELIA VIDRIGHIN-BRATU, OVIDIU DANCEA,<br>CRISTINA-BIANCA POP<br>Intelligent system for assisting the therapeutic decision at patients with prostate<br>cancer – INTELPRO. . . . .  | 36  |
| CAMELIA VIDRIGHIN, BRATU, RODICA POTOLEA<br>Pre-processing Approaches for Improving Knowledge Extraction from Medical Data . . . . .  | 47  |
| MIHAELA DINSOREANU, IOAN SALOMIE, SERGIU NEDEVSCHI, IOAN COMAN, BOGDAN PETRUT,<br>VASILE STANCA, BOGDAN TINTILA, CRISTINA B. POP, SORIN L. SUCIU, TUDOR I. SALOMIE, EMANUEL<br>VARGA, DANIEL DANJI<br>Knowledge Engineering in the Prostate Cancer Domain. . . . .  | 54  |
| IOAN SALOMIE, MIHAELA DINSOREANU, SERGIU NEDEVSCHI, IOAN COMAN, BOGDAN PETRUT,<br>VASILE STANCA, BOGDAN TINTILA, CRISTINA B. POP, SORIN L. SUCIU, TUDOR I. SALOMIE, DIANA SZIRACZKI,<br>MIHAI TIMAR<br>Expert System Generator. A Case Study for the Prostate Cancer Domain. . . . .  | 58  |
| ALEXANDRU G. FLOARES, RADU BADEA, CARMEN FLOARES, IOAN COMAN, LUCIANA NEAMTIU, VIOREL<br>ARON, ALEXANDRA OLTEANU, MIHAI CIORNEI<br>Liver and Prostate i-Biopsy and the Related i-Scoring Systems. . . . .   | 63  |
| ALEXANDRU G. FLOARES, RADU BADEA, CARMEN FLOARES, IOAN COMAN, LUCIANA NEAMTIU, VIOREL<br>ARON, ALEXANDRA OLTEANU, MIHAI CIORNEI<br>RODES - Algorithm for Automatic Mathematical Modeling Complex Biological<br>Networks via Knowledge Discovery in Data . . . . .   | 69  |
| RADU DANESCU, SERGIU NEDEVSCHI, SORINA DANESCU, RODICA COSGAREA<br>Flexible Web-Based System for Storing and Retrieving Information about Patients<br>with Genodermatoses. . . . .  | 74  |
| SERGIU NEDEVSCHI, COSMIN D. PANTILIE, TIBERIU MARITA, DANIEL MOGA, SORIN M. DUDEA<br>Towards the Standardization of Ultrasound Elastography: a Complete Approach . . . . .  | 80  |
| GURZAU MIRCIA AND DORIAN POPA<br>Mathematical Models of the Heat Transfer in Tissues . . . . .  | 88  |
| IOAN SALOMIE, VIORICA R. CHIFU, SERGIU NEDEVSCHI, EMIL ȘT. CHIFU<br>Ontology Learning for Traceability in the Food Industry . . . . .   | 94  |
| IOAN SALOMIE, MIHAELA DINSOREANU, TUDOR CIOARA, IONUT ANGHIEL, IOAN TUDOR SALOMIE<br>Industrial Business Process Modeling, Verification and Simulation with Application in Food<br>Industry . . . . .   | 102 |
| IOAN SALOMIE, MIHAELA DINSOREANU, CRISTINA BIANCA POP, LIVIU SORIN SUCIU<br>Logistic Chain Modeling and Generation for SOA-based Total Traceability . . . . .   | 112 |
| G. SAPLACAN, M. RUSU, D. VLADOIU, M. CERGHIZAN, T. NICOARA, GH. SEBESTYEN, SERGIU NEDEVSCHI <sup>1</sup><br>Integrated systems and traceability for food-industry. Multiple input data sources . . . . .  | 118 |

|   |     |
|---|-----|
| DOINA BANCIU, GH. SEBESTYEN, TUNDE BALINT, BOGDAN MOSCALIUC<br>Strategies Concerning Research Development in Digital Library Field, Key Issues in Digital<br>Library Implementation . . . . . | 126 |
| KATALIN BÄNKESZI, PÉTER DIPPOLD<br>National Strategies for Digitization of Libraries, Museums and Archives in Hungary   | 133 |
| GABRIELA FLORESCU, GABRIELA DUMITRESCU, VALENTIN FLORESCU, CORNEL LEPADATU<br>A Digital Tool for Manuscripts Preservation . . . . .   | 136 |
| DORA COARDOȘ, VASILE COARDOȘ, CORNELIA LEPĂDATU, CORNEL LEPĂDATU<br>Support System for Libraries based on Business Intelligence . . . . .   | 140 |

## Integrated procedures of evaluating the hepatic morphology, for the non-invasive detection and quantification of steatosis and fibrosis based on the ultrasound diagnosis

Radu Badea<sup>1</sup>, Monica Lupșor<sup>1</sup>, Horia Ștefănescu<sup>1</sup>, Cristian Vicaș<sup>2</sup>, Delia Mitrea<sup>2</sup>, Mihai Socaciu<sup>1</sup>, Sergiu Nedevschi<sup>2</sup>

<sup>1</sup> *University of Medicine and Pharmacy Iuliu Hatieganu, Cluj-Napoca, Romania*

<sup>2</sup> *Department of Computer Science, Technical University of Cluj-Napoca, Romania*

rbadea@umfluj.ro

### Abstract

*Liver chronic diseases constitute a complex part of affections with different etiologies, more often overlapped. Their evolution often degenerates in fibrosis and reorganization of the parenchyma. The process is slowly developed, without knowing the periods of acceleration or stagnation. The clinic and functional biochemical diagnosis is specific enough, having a lower perception of the stage and the aggressiveness of the disease. The well-known diagnosis is taken by liver biopsy –an aggressive, invasive method, disliked by the patients. There are studies about the utility and the efficiency of noninvasive methods. They play a very important role in the diagnosis of diffuse liver diseases. The goals of the research are represented by the optimization of the ultrasound image with modern procedures such as: the unidirectional transitory elastography, the enhancement of the images using computerized techniques and automation of the diagnosis using artificial intelligence algorithms.*

*The studies made on large groups of biopsied patients led to the conclusion that the computerized analyze of the data extracted from the ultrasound images is conclusive for the examination and also facilitates the early diagnosis of some illnesses and even more accurate diagnosis of those diseases that have similar ultrasound images. The evaluation of the liver stiffness using the unidirectional transitory elastography (UTE) is an alternative to liver biopsy, where fibrosis is the parameter that needs to be evaluated. The diagnosis accuracy of UTE is very good, not only for the screening of severe fibrosis and the cirrhosis in chronic hepatic virus C disease, but also for more other liver chronic diseases, especially when the hepatic biopsy is very difficult, or even impossible. By using the combination of these methods and by correlation with other noninvasive methods of diagnosis we can approach more and more to “the hepatic virtual biopsy” which will allow a better accurate monitoring of the evolution of the disease.*

*This also will avoid as much possible, the noxious effects of some invasive manners of diagnosis.*

**Key-words:** fibrosis, cirrhosis, restructuring, ultrasonography, image-processing, elastography

**Abbreviation list:** HBV – hepatic B virus, HCV – hepatic C virus, LC – liver cirrhosis, HCC – hepatocellular carcinoma, NASH – nonalcoholic steatohepatitis, US - ultrasonography, UTE- unidirectional transitory elastography, AC – attenuation coefficient

### 1. Introduction: Importance of the issue. The etiology of diffuse liver diseases. Worldwide spread.

Liver chronic diseases represent a major problem of public health. The infections with hepatic virus B (HBV), C (HCV) or Delta (HVD), nonalcoholic steatohepatitis (NASH) or alcoholic hepatopathy (AH) constitute the main illnesses of the liver. Other diseases - genetic, metabolic or immunology related mechanisms - have a lower weight as a whole. The common feature of the liver diseases is represented by the slower evolution towards cirrhosis (LC), and then, towards hepatocellular carcinoma (HCC). Between viral affections, those with B virus are more often met (the risk of becoming chronic is 17.5%). The incidence and prevalence ciphers vary much worldwide. Thus, the areas with high incidence (15%) are: South-Saharan Africa and South-East Asia, where there is considered that between 70% and 90% from the population came into contact with the virus and those with intermediary incidence (2-7%) are concentrated in the Mediterranean basin and in Central - East Europe. (Here, approximate 20 - 55% of the population had passed through infection). In Nordic and Eastern Europe, North America and Australia is recording a cipher of incidence below 2% and less 20% of the population passed over the infection. The acute infection with HCV becomes chronic in approximate

## The employment of textural and non textural image analysis algorithms in assessing the diffuse liver diseases

Sergiu Nedevschi<sup>1</sup>, Cristian Vicas<sup>1</sup>, Monica Lupsor<sup>2</sup>, Radu Badea<sup>2</sup>, Mircea Grigorescu<sup>2</sup>

<sup>1</sup>Department of Computer-Science, Technical University of Cluj-Napoca, Romania

<sup>2</sup>Medical Clinic no. 3, University of Medicine and Pharmacy, Cluj-Napoca, Romania

### Abstract

*This paper is focused on current progress of our research in improving diagnosis value of ultrasound imaging in the context of diffuse liver diseases. Image features are computed on ultrasound images and these features are used to train a classifier. The classifier is able to distinguish between various pathology grades. Present study shows that, based on ultrasound images, steatosis can be accurately graded and a qualitative assessment can be made in case of fibrosis. Further improvements can be made if we include more patients and consider non-imagistic features like clinical and biochemical analysis of the patient.*

### 1. Introduction

There are two main diseases studied: nonalcoholic steatohepatitis and chronic hepatitis. Both diseases alter the liver by producing fatty infiltration (steatosis), fibrosis, inflammatory and necrosis processes. The way that these processes alter the liver varies from steatohepatitis to chronic hepatitis. For this reason, physicians use two scores in order to evaluate the liver biopsy [1]. First score is BRUNT and is used for steatohepatitis and the second is METAVIR and is used for chronic hepatitis. The aim of this research is to use noninvasive techniques (US imaging) to assess the BRUNT/METAVIR score or, at least, some components of these scores. Fibrosis is scored from 0 (healthy) to 4 (cirrhosis). Steatosis is graded from 0 (healthy) to 3 (>66% steatosis) Activity is graded also from 0 to 3.

We have four main patient groups. A clinically healthy group, steatohepatitis patients, chronic hepatitis patients (both groups have been diagnosed using liver biopsy) and a cirrhotic group which was only clinically diagnosed.

At liver tissue there are a number of pathological processes. Each of these processes could alter the aspect of the US images. As we shown in our previously work [2][3][4][5], steatosis has the biggest impact, followed by fibrosis and activity. Also we have shown that these processes can mask each other, i.e.

strong fibrosis can mask a low grade steatosis and vice-versa [3][4]. The activity can be detected only if we consider two patient lots, having the same steatosis and fibrosis grade but different activity grade. If the steatosis grade or fibrosis stage varies, the activity cannot be detected.

Another interesting aspect discovered in our previous work [2][4] is that the fibrosis detection is better if we select only patients with low steatosis. The opposite is also true, detection of steatosis is better if the fibrosis stage is kept low.

The paper is structured as follows: Section 2 describes the methodology, section 3 focuses on steatohepatitis disease and section 4 focuses on fibrosis stage detection at chronic hepatitis. Section 5 concludes the paper.

### 2. Methodology

The methodology proposed in our research consists in five main steps: image acquisition, the establishment of region of interest, image attributes computation, feature selection and classification.

The algorithms used to generate image features are shown in Table 1, along with the number of features generated by each algorithm. Image acquisition, ROI establishment and the attribute computation algorithms are widely explained in [2][3][4].

**Table 1. Algorithms used to generate image features and the number of features generated by each algorithm**

| Algorithm                                   | Number of features generated |
|---|------------------------------|
| Histogram statistics                        | 6                            |
| Grey tone difference matrix                 | 25                           |
| Grey level co-occurrence matrix             | 200                          |
| Multifractal differential box counting      | 90                           |
| Morphological multifractal exponents        | 464                          |
| Multiresolution fractal dimension           | 3                            |
| Law's texture energy measures               | 56                           |
| Wavelet transform                           | 255                          |
| Attenuation and backscattering coefficients | 12                           |

## **The value of the ultrasound image processing procedures, combined with circulation assessment and mathematical modelling techniques for HCC characterization, nature specification and disease prognosis**

Radu Badea<sup>1</sup>, Teodora Pop<sup>1</sup>, Mihai Socaciu<sup>1</sup>, Ofelia Moşteanu<sup>1</sup>, Cosmin Caraiani<sup>1</sup>, Paula Raica<sup>2</sup>,  
Mihaela Manisor<sup>2</sup>, Liviu Miclea<sup>2</sup>, Delia Mitrea<sup>3</sup>, Monica Lupşor<sup>1</sup>, Sergiu Nedevschi<sup>3</sup>

<sup>1</sup>*Ultrasonography Department, The 3<sup>rd</sup> Medical Clinic, University of Medicine and Pharmacy „Iuliu Hatieganu”, Cluj-Napoca, Romania;*

<sup>2</sup>*Department of Automation, Technical University of Cluj-Napoca, Romania*

<sup>3</sup>*Department of Computer Science, Technical University of Cluj Napoca, Romania*  
rbadea@umfcluj.ro

### **Abstract**

*The exploration of the liver tumors is based on progresses made on imagistic techniques. The evaluation complexity needs to take into account: the detection of bellows centimeters dimensions, the nature characterization and the appreciation of the diagnosis. The goal of the study was to identify and use complex criteria, processing procedures of ultrasound images, the evaluation of tumor circulation natively, or with i.v. contrast substances for consolidating a non invasive methodology hepatocellular carcinoma characterization. There have been explored, through ultrasound techniques, 190 patients with liver tumors, from which 120 with carcinoma developed on cirrhosis. The final diagnosis was specified through validated methods (clinical, functional-biochemical, imagistic, morphological). It was developed an imagistic/morphometric model of HCC and specific features were identified which cannot be detected by the examiner's eyes. A qualitative vascular model, as well as a model of tumor washing with contrast agent, which suggests the tumor nature, was defined. It was developed a mathematical model of the induced angiogenesis and a model of the tumor growth. The results show that a multidirectional approach of the ultrasound image allows the non-invasive diagnosis and the characterization of a liver tumor.*

**Key-words:** liver – liver cirrhosis – tumor – hepatocellular carcinoma – ultrasound images – image processing – i.v. contrast – mathematical modeling.

**Abbreviation list:** HCC – hepatocellular carcinoma – PV – Portal Vein — US – ultrasonography – EVD - the effective vessels density.

### **1. The importance of the issue.**

The hepatocellular carcinoma (HCC) is one of the most frequent forms of cancer in the world. It places on the 5<sup>th</sup> stage as occurrence and on the 3<sup>rd</sup> place as mortality. The prognosis of the National Institute of

Cancer from USA [1] shows that in the year 2008 in USA were being diagnosed with this disease 21 370 patients from which 18 410 risked to lose their lives because of the advanced detected stage or of some complications that define cirrhotic liver on which the tumor develops.

### **2. The HCC diagnosis**

The HCC diagnosis has known considerable progresses in the last years. Between the most performing procedures that even led to a significant amelioration of the prognosis of this disease, there are those imagistic. Imagistic methods had demonstrated their utility and performance in some more directions: a. precocious detection; b. diagnosis and staging; c. visual conducted biopsy; d. the therapy efficiency supervision through chemotherapy, ablation with radiofrequency, alcoholization, embolizations and s.o., e. planning the radiotherapy strategy; f. watching the disease evolution; g. the efficiency estimation of some medicines during clinic trials; h. oncologic research. [2]. Imagistic models are divided in two major classes: anatomical – containing conventional radiology (Rx), nuclear medicine (NM), computer based tomography (CT), nuclear magnetic resonance (NMR), ultrasonography (US), and functional - between which there are positrons emission tomography (PET) and Contrast ultrasonography (CEUS). These methods try, in different ways to identify textural or vascular tumor characteristics and also to distinguish them from normal liver tissue. The information vectors differ from one method to another, so that US uses acoustic waves with low energy, NMR uses low energy waves from electromagnetic field, in the same time Rx, CT and PET use high energy waves from the same field. Using contrast substances, which are unspecific, it raises the quality of the image and the combination of

## The Imagistic Textural Model of the Hepatocellular Carcinoma. Comparisons, discussions, conclusions

Delia Mitrea<sup>1</sup>, Sergiu Nedevschi<sup>1</sup>, Paulina Mitrea<sup>1</sup>, Monica Lupsor<sup>2</sup>, Mihai Socaciu<sup>2</sup>, Teodora-Atena Pop<sup>2</sup>, Ofelia Mosteanu<sup>2</sup>, Radu Badea<sup>2</sup>

<sup>1</sup> *Department of Computer-Science, Technical University of Cluj-Napoca, Romania*

<sup>2</sup> *Department of Ultrasonography, 3<sup>rd</sup> Medical Clinic, University of Medicine and Pharmacy, Cluj-Napoca, Romania*  
*Delia.Mitrea@cs.utcluj.ro*

### Abstract

*The purpose of this study is to elaborate a reliable imagistic textural model of HCC in order to perform a highly accurate automatic differentiation of this malignant tumor from other visually similar tissues like the cirrhotic parenchyma and from the benign liver tumors; thus, we aim to characterize the hepatocellular carcinoma (HCC), by detecting the relevant textural features obtained from ultrasound images, that have the major contribution concerning this differentiation. For the computation of the textural features the following methods are chosen in order to emphasize the chaotic nature of the malignant tumor: first and second order statistics of the grey levels, statistics of the local features, fractal-based methods and multi-resolution methods; specific methods for feature-selection are applied in order to determine the exhaustive set of independent and relevant features for each evolution phase. We focus on modeling the texture of the hepatocellular carcinoma, of the cirrhotic parenchyma on which HCC is evolved and of the benign liver tumors like hemangioma but we also watch the evolution from cirrhosis towards HCC. The final purpose is that of providing a reliable method and software tool for the non-invasive characterization and recognition of HCC.*

### 1. Introduction

Liver chronic diseases constitute an important public health issue, due to their frequency and to the danger of their evolution towards cancer. The evolution of diffuse liver diseases is variable, but it usually has generally long term. Whatever the nature of the liver aggression, it seems to follow a pattern characterized by the successive stages: inflammation, necrosis, fibrosis, regeneration (cirrhosis), dysplasia, and, at the end, the hepatocellular carcinoma. The hepatocellular carcinoma (HCC) is the most frequent malignant liver tumor (representing 75% of the liver cancer cases), besides hepatoblastoma (7%), cholangiocarcinoma and cystadenocarcinoma (6%). The most frequent benign liver tumors are

hemangioma, adenoma and nodular focal hyperplasia. The visual aspect of the liver tissue, during the above-mentioned pathological stages, evolves from normal to inhomogeneous, due to the cirrhotic nodules, then the nodules increase giving birth to the hepatocellular carcinoma. The human observations are not enough in order to give a reliable diagnosis and the biopsy is an invasive method, dangerous for the patient. Thus, a more subtle analysis is due, and we aim to perform this by using computerized methods in order to extract the necessary information from ultrasound images. Texture is a very important visual feature in this context, as it provides a lot of information concerning the pathological state of the tissue. Although a commonly accepted definition does not exist, texture describes the regular arrangement of the grey levels in the region of interest, giving also information about the frequency and regularity of the local features, about the fundamental microstructures, being also able to provide multi-resolution information, appropriate for an accurate, subtle characterization of the tissue illness. Texture-based methods, in combination with classification methods have been widely used for the automatic diagnosis of various kinds of tumors. The Grey Levels Co-Occurrence Matrix, the Wavelet and Gabor transforms and fractal-based methods are only a few eloquent examples used in the characterization of the breast tumors, of the prostate adenocarcinoma, of the salivary gland tumors, of liver lesions, respectively of malignant and benign tumors. However, a systematic study concerning the relevant features and their specific values for the characterization of HCC based only on information extracted from ultrasound images is not done yet. We aim to do this in our work, which will consist in modeling the HCC tumor and the visually similar tissues through textural parameters. We also approach the problem of detecting the real margins of the HCC tumor, which are often invisible for the human eye. In order to build the imagistic textural model, some important steps had to be performed. At the beginning, we perform the *image analysis phase*, implying the computation of the textural features. First and second order statistics of the grey levels, edge-based statistics, statistics of the local

## **Intelligent system for assisting the therapeutic decision at patients with prostate cancer – INTELPRO**

Sergiu Nedevschi<sup>1</sup>, Ioan Salomie<sup>1</sup>, Rodica Potolea<sup>1</sup>, Mihaela Dinsoreanu<sup>1</sup>, Radu Badea<sup>3</sup>,  
Alexandru Floares<sup>2</sup>, Ioan Coman<sup>3</sup>, Bogdan Petrut<sup>3</sup>, Delia Mitrea<sup>1</sup>, Camelia Vidrighin-Bratu<sup>1</sup>,  
Ovidiu Dancea<sup>4</sup>, Cristina-Bianca Pop<sup>1</sup>

<sup>1</sup>*Department of Computer Science, Technical University of Cluj-Napoca, Romania*

<sup>2</sup>*SAIA, Cluj-Napoca, Romania*

<sup>3</sup>*University of Medicine and Pharmacy, Cluj-Napoca, Romania*

<sup>4</sup>*IPA SA – R&D, Institute for Automation*

*Sergiu.Nedevschi@cs.utcluj.ro*

### **Abstract**

*The main research goal of this project was the development and implementation of methods for prediction of the individual result of the radical prostatectomy at patients with clinically localized prostate cancer. The aim was to predict the postoperative evolution of the disease, after the radical prostatectomy, by finding the corresponding prognostic group of the patients, based on the evaluation of an exhaustive set of the relevant medical parameters. The data set included both the parameters obtained directly from medical analyses, respectively more refined information provided by the intelligent algorithms and image analysis methods. Data modeling, data collection, data mining applications, methods and tools for building the imagistic model of the prostatic malignant tumor, and an expert system for prostate cancer diagnosis are involved in this work.*

### **1. Introduction**

The prostate cancer is the most common neoplasy at men in the United States and the second cause of death through malignity at men older than 55. The prostatic adenocarcinoma (ADKP) is the most frequent malignant prostatic tumor (95% of the prostate cancer cases). The diagnostic elements are represented through: the anamnestic data, correlated with the clinical examination, the PSA value, the prostate needle biopsy (PBP) and the imagistic explorations. If we have a localized prostate cancer, without metastasis, the patient can undergo a very efficient treatment – the radical prostatectomy, performed through classical or laparoscopic surgery. This operation removes the prostate cancer in the majority of cases, but also can have a negative impact on the patient's life. Thus, it is very useful to detect the character and the evolution stage of the tumor, as well as to predict the evolution of the prostate cancer after

performing the radical prostatectomy, in order to make the right decision concerning the best treatment.

Our project consisted in the development of an intelligent system capable of predicting the evolution of the prostate cancers treated through radical prostatectomy. The prediction of the local or distant recurrence permits the individualization of the therapeutic strategy, leading to a reduction of the mortality, morbidity and costs. The objective of our project was that of using the methods of artificial intelligence in order to predict the individual result of the radical prostatectomy. We estimate that our system, built and trained using a big enough number of cases, brings a substantial contribution concerning the improvement of the prognostic capacity over the disease evolution and the decision over the most appropriate treatment.

### **2. The objectives of INTELPRO and the general description of the developed components**

The objective of this project consists in using the methods of artificial intelligence in order to predict the individual result of the radical prostatectomy to the patients with clinically localized prostate cancer.

The solving problematic consists in the creation of an intelligent system. The main purpose of this system is stated below:

- Predict the disease evolution for the patients with prostate cancer, treated through radical prostatectomy by integrating them in groups with different prognostic, according to the PSA post-therapeutic levels. [1]

The most important secondary objectives, that are inherent to the realization of a robust solution, are:

- Medical data modeling and collection, medical data preprocessing and enhancement from the quality and quantity points of view; the integration of heterogeneous data, combination of the data provided by different sources;

# Pre-processing Approaches for Improving Knowledge Extraction from Medical Data

Camelia Vidrighin Bratu, Rodica Potolea

*Department of Computer Science, Technical University of Cluj-Napoca, Romania*

*Camelia.Vidrighin@cs.utcluj.ro*

## Abstract

*Currently, machine learning schemes have reached a high level of maturity. There exist many robust methods which can be generally employed, as well as a series of approaches designed for particular domains. Experience has shown us that, besides algorithm selection and engineering, the data mining expert has to consider the quality of the training data. A low-quality data (missing values, noise, irrelevant/redundant information) can largely affect the learning capabilities of an algorithm. This paper presents the results obtained by two pre-processing approaches on medical data. Both data imputation and wrapper feature selection are shown to potentially improve the accuracy of classifiers.*

## 1. Introduction

Mining medical data has become one of the key concerns in the data mining community in the last years, and represents one of the most challenging application domains for knowledge extraction methods.

Among the various domain particularities encountered in medical data mining, the quality of the data is an important factor which affects the outcome of the knowledge extraction process.

This paper presents two pre-processing approaches explored while working on a system for assisting the diagnosis process in prostate cancer. It is structured as follows: section 2 introduces the need for improving the data quality, by presenting the results of robust classification methods on real prostate cancer data. Section three presents the two approaches considered, and introduces the idea of a combined pre-processing methodology for data mining. The conclusions are presented in the last part of the paper.

## 2. Data-related issues

The medical field possesses a series of domain particularities, which demands for special strategies to

be considered when mining such data. We have previously explored several complex approaches, which tackle different domain particularities [11].

The DST system assesses the baseline accuracy of a dataset, by combining the predictions of several different classifiers using principles from the mathematical theory of Dempster and Shafer, on belief functions and plausible reasoning [8]. Benchmark validations on medical data have shown that this method achieves stable results across different datasets, which makes it a good indicator of what we can expect in terms of accuracy, for a given dataset.

The prostate cancer data was provided by two different sources. The first source is a German clinic, and the data has been collected from patients which have undergone prostate surgery. After analyzing the data, the medical team has suggested as problem of interest the prediction of the value of post-operative PSA from pre-operative and operative data. The resulting dataset (after manual pre-processing) contains 399 instances, 15 attributes, and the class is nominal (3 different values). A second data source is the University of Medicine and Pharmacy in Cluj-Napoca. They have provided us with 50 more cases for which we were able to define the same problem of interest as for the German data.

On the available prostate cancer data the DST system has obtained the following accuracy levels:

| <b>ACCURACY LEVELS DST</b>     |               |
|--------------------------------|---------------|
| <b>German data:</b>            | <b>68.48%</b> |
| <b>Romanian data:</b>          | <b>51.88%</b> |
| <b>German + Romanian data:</b> | <b>65.58%</b> |

The accuracy is very low, unacceptable if we take into account the demands of such a domain. Also, when the 2 datasets are combined, the accuracy drops. This could be the result of the 2 datasets having different distributions.

The second approach is based on a PANE (Preceded by Artificial Neural network Ensemble) idea. The training instances are re-labelled using the model learned by an ensemble of artificial neural networks.



## Knowledge Engineering in the Prostate Cancer Domain

Mihaela Dinsoreanu<sup>1</sup>, Ioan Salomie<sup>1</sup>, Sergiu Nedevschi<sup>1</sup>, Ioan Coman<sup>2</sup>, Bogdan Petrut<sup>2</sup>,  
Vasile Stanca<sup>2</sup>, Bogdan Tintila<sup>2</sup>, Cristina Bianca Pop<sup>1</sup>, Sorin Liviu Suciu<sup>1</sup>, Tudor Ioan Salomie<sup>1</sup>,  
Emanuel Varga<sup>1</sup>, Daniel Danci<sup>1</sup>

<sup>1</sup>*Department of Computer Science, Technical University of Cluj-Napoca, Romania*

<sup>2</sup>*University of Medicine and Pharmacy, Cluj-Napoca, Romania*

*Mihaela.Dinsoreanu@cs.utcluj.ro*

### Abstract

*This paper addresses the problem of knowledge engineering in the medical field of prostate cancer. We propose a generic knowledge model and a methodology for representing and capturing knowledge from clinical practice guidelines on prostate cancer. Domain knowledge is represented as an ontology and as a set of rules. The rules are built according to a generic grammar that enables rules translation into specific representations required by different knowledge-based expert system cores. We validate our approach in a case study on prostate cancer diagnosis.*

### 1. Introduction

Medicine is a domain that would greatly benefit of intelligent systems in assisting physicians in their current activities. Developing such systems requires precise representation and capturing of the domain knowledge in order to obtain correct answers and solutions as a result of system inferences.

Knowledge engineering provides the methodology for building and processing knowledge bases.

Knowledge engineering in the medical domain aims to identify suitable methods for formalizing and capturing the information found in medical practice guidelines. The resulting knowledge should be machine processable.

Knowledge engineering is a multi-stage process similar to software engineering. First of all, the domain of interest must be identified. Next, with the help of a domain expert, domain documents must be analyzed and information classified. The next step consists in identifying the proper means for knowledge representation followed by knowledge capturing, representation and validation. However, the process is iterative and incremental, in the sense that knowledge is updated and extended according to new inferred knowledge.

In [8], authors emphasize the idea that formalizing the information retrieved from the medical practice guidelines requires a permanent interaction between the knowledge engineer, who has computer science

skills, and the physician, who has the medical knowledge. The paper presents an intermediate knowledge representation, in the form of an ontology that may be used to mediate between the raw text available in the guidelines and several formal guideline representations.

In [7], authors present a methodology that automates parts of the clinical guidelines modeling process by providing a medical ontology, structuring and decomposing the guideline information (e.g. treatment processes, diagnosis methods, definitions), making the modeling process traceable, comprehensible and applicable for many guideline representation languages.

Studies of several approaches for creating computer-interpretable guidelines that facilitate decision support are presented in [8]. These approaches address the problems of medical guideline representation, verification and execution.

Medical knowledge can also be extracted from the patients' clinical records through data mining. This method is appropriate when large data sets are available.

Our paper proposes a knowledge model and a methodology for knowledge capturing from medical practice guidelines. For the captured knowledge we propose a representation consisting of domain ontology and a set of rules. The rules are based on a generic grammar that enables their integration within the knowledge bases of different expert systems. The knowledge extracted from the medical documents is validated by using two expert system cores.

The rest of the paper is organized as follows. Section 2 presents the knowledge model and representation. Section 3 illustrates the process of knowledge capturing. A case study for building the knowledge base for prostate cancer diagnosis is presented in section 4. The paper ends with our conclusions and future work proposals.

### 2. Knowledge Model and Representation

Domain knowledge is modeled as domain ontology and a set of rules that relate concepts from the domain ontology (Figure 1). Both the ontology and the set of

## Expert System Generator. A Case Study for the Prostate Cancer Domain

Ioan Salomie<sup>1</sup>, Mihaela Dinsoreanu<sup>1</sup>, Sergiu Nedevschi<sup>1</sup>, Ioan Coman<sup>2</sup>, Bogdan Petrut<sup>2</sup>,  
Vasile Stanca<sup>2</sup>, Bogdan Tintila<sup>2</sup>, Cristina Bianca Pop<sup>1</sup>, Sorin Liviu Suci<sup>1</sup>, Tudor Ioan Salomie<sup>1</sup>,  
Diana Sziraczki, Mihai Timar

<sup>1</sup>*Department of Computer Science, Technical University of Cluj-Napoca, Romania*

<sup>2</sup>*University of Medicine and Pharmacy, Cluj-Napoca, Romania*

*Ioan.Salomie@cs.utcluj.ro*

### Abstract

*This paper proposes a knowledge-based expert system generator and its application in a case study of knowledge engineering in the prostate cancer domain. The system is flexible and enables easy configuration and generation of expert system instances. By providing specialized translators, the raw data and the domain knowledge represented in a generic format can be easily translated into the specific representation required by the actual core of the expert system instance. We tested our solution in the context of the medical field of prostate cancer by instantiating two expert systems, one based on Prolog and one based on Jess core. Using the knowledge acquired from the medical practice guides and from the patients' medical records, the instantiated expert systems can assist doctors during their medical activities.*

### 1. Introduction

Advances in Artificial Intelligence have resulted in the development of knowledge-based expert systems. A knowledge-based expert system encompasses several components such as the knowledge base, inference mechanisms, explanation facility tools and others, that interact together in simulating the problem-solving process assisted by an acknowledged expert of the domain [1].

In order to better model the human reasoning and to act as experts for solving problems in particular domains, knowledge-based expert systems need to be trained with accurate knowledge and to have proper inference mechanisms attached. Generally, these inference mechanisms are based either on forward chaining – data-driven process, or on backward chaining – goal-driven process.

Significant effort has been directed towards identifying the proper ways to represent knowledge and to enable reasoning so that such systems may offer useful support in human decision-making.

MYCIN [2] [3] is one of the earliest expert systems used for medical diagnosis. Today, many commercial expert system shells use this approach. In order to deal

with uncertainty, MYCIN proposes the usage of *certainty factors*. The way these certainty factors are dealt with, should not be confused with the concept of probabilities, which have a completely different mathematical background, namely the probabilistic theory. The certainty factors are considered as "weights" or degrees of truth of a fact or rule, with the range between -100 and 100, the lower level being totally false while the upper one being considered as totally true. Values close to 0 are thought of as irrelevant. As far as the rules are concerned, the certainty factors in the conclusion part of the rule are based on the assumption that the premises are known with a certainty factor of 100. However, it is unlikely that a premise is perfectly known. The system determines the certainty factor of the premise as the minimum of the certainty factors of the individual sub goals in the premise. When the premise of a rule is uncertain, the certainty factor of the conclusion needs to be computed according to a specific formula.

CLAM [2] is an expert system shell written in Prolog, which deals with uncertainty in a very similar way to MYCIN. It is based on direct user interaction for fact acquisition and its knowledge base contains only attribute-value facts. The system stores all proven goals along the execution, thus increasing performance for repeated queries on the same goal.

OOPS [2] is another expert system shell written in Prolog. The use of Prolog terms for modeling facts and rules is the most interesting feature of this system which also takes advantage of the powerful Prolog unification mechanism and of its highly expressive knowledge representation.

This paper proposes a knowledge-based expert system generator, which can be easily instantiated and adapted to solve problems from several domains. The system enables the integration of different expert system cores that provide forward or backward chaining inference mechanisms. By using generic knowledge representation and specific translators, the system's knowledge base is adapted to the representation required by the currently considered expert system core. For the case study of prostate cancer domain we have considered two expert system cores, one based on Prolog that enables backward

## Liver and Prostate i-Biopsy and the Related i-Scoring Systems

Alexandru G. Floares<sup>1,2</sup>, Radu Badea<sup>3</sup>, Carmen Floares<sup>2</sup>, Ioan Coman<sup>3</sup>, Luciana Neamtiiu<sup>1,2</sup>,  
Viorel Aron<sup>1</sup>, Alexandra Olteanu<sup>1</sup>, Mihai Ciornei<sup>1</sup>

<sup>1</sup> SAIA, Cluj-Napoca, Romania, alexandru.floares@gmail.com

<sup>2</sup> Oncological Institute Cluj-Napoca, Cluj-Napoca, Romania,

<sup>3</sup> University of Medicine and Pharmacy Cluj-Napoca, Romania,

### Abstract

*An important medical goal is to replace invasive and painful procedures with non-invasive techniques for diagnosis. We investigated the capability of a knowledge discovery in data approach, based on artificial intelligence tools, to integrate information from various data sources - imaging data, clinical and laboratory data - to predict the results of the biopsy, with acceptable accuracy. The resulted intelligent systems, tested on 528 patients with chronic hepatitis C, and on 393 patients with prostate cancer, are based on C5.0 decision trees and boosting ensemble method. They predict the fibrosis stage results of the liver biopsy, according to the two largely accepted fibrosis scoring systems, Metavir and Ishak, with and without liver stiffness (FibroScan®), and the Gleason score of prostate biopsy. We also introduce the concepts of intelligent virtual biopsy or i-biopsy and that of i-scores for the corresponding scoring systems. To our best knowledge i-biopsy outperformed all similar systems published in the literature and offer a realistic opportunity to replace liver biopsy in many important medical contexts.*

### 1. Introduction

An important medical goal is to replace invasive and painful procedures with non-invasive techniques for diagnosis. In this paper we focused on biopsy, an important invasive procedure, and on two non-invasive techniques, namely ultrasonography and liver stiffness measurements (FibroScan®). We found that usually, the information content of the non-invasive imaging techniques is lower than that of the invasive techniques. As a consequence, the accuracy of the diagnostic based on these non-invasive techniques alone is not very high.

The main question addressed in our studies is: can we extract and integrate information from various (non-invasive) data sources, e.g. imaging data, clinical and lab-

oratory data, to reach an acceptable, say more than 90% diagnostic accuracy?

For any usual or traditional medical approach, the answer to this important question is *NO*. This is because our brain is very skilled in dealing with images, but the information content of the medical images does not seem to be enough for a diagnostic, and is less skilled in dealing simultaneously with many variables, and this seems to be the solution to the problem.

A knowledge discovery in data or data mining approach, based on artificial intelligence tools, could be the foundation for a positive answer to the above important medical question. The extraction and integration of information from various data sources is indeed possible, and the diagnostic accuracy of the resulted intelligent systems is high.

To illustrate this thesis, we present an overview of some of our recent investigations on building intelligent systems capable to predict the results of liver biopsy [4, 3], and prostate biopsy. We used several non-invasive approaches: routine laboratory tests and basic ultrasonographic features, with and without liver stiffness measurement by transient elastography (FibroScan®).

### 2 Intelligent systems for liver fibrosis stage and prostate Gleason score prediction

#### 2.1 Data integration and preprocessing

One of the key aspect of this studies consists in integrating various medical data: clinical, imaging and lab data. Our experiments showed that isolated data sources do not contain enough information for building accurate intelligent systems. The main problems we found, in mining the medical data bases used in these studies, were related to the small number of patients relative to the number of features, and to the extend of missing data. However, comparing to similar medical studies we investigated large datasets, with hundreds of patients.

# RODES - Algorithm for Automatic Mathematical Modeling Complex Biological Networks via Knowledge Discovery in Data

Alexandru G. Floares<sup>1,2</sup>, Carmen Floares<sup>2</sup>, Radu Badea<sup>3</sup>, Ioan Coman<sup>3</sup>, Luciana Neamtiiu<sup>1,2</sup>,  
Viorel Aron<sup>1</sup>, Alexandra Olteanu<sup>1</sup>, Mihai Ciornei<sup>1</sup>

<sup>1</sup> SAIA, Cluj-Napoca, Romania, alexandru.floares@gmail.com

<sup>2</sup> Oncological Institute Cluj-Napoca, Cluj-Napoca, Romania,

<sup>3</sup> University of Medicine and Pharmacy Cluj-Napoca, Romania,

## Abstract

*Modeling biological networks with ordinary differential equations is a sensible approach, but also very difficult. This paper describes an algorithm based on linear genetic programming which can be used to reverse engineer biological networks. The RODES algorithm automatically discovers the structure of the network in addition to estimating its parameters, and can even be used to identify the biophysical and biochemical mechanisms involved. The algorithm was tested on simulated time series data, generated using a realistic models of the tumor growing and angiogenesis. The resulting ODE systems are highly accurate, and results are obtained in a matter of minutes due to the use of domain knowledge to restrict the solution space. The problem of reversing a system of coupled differential equations is reduced to one of reversing individual algebraic equations. To our knowledge, this is the first time a realistic reverse engineering algorithm based on linear genetic programming has been applied to tumor growing and angiogenesis networks.*

## 1 Introduction

Biological networks come in many forms: neural networks, gene regulation and transcription networks, signal transduction cascades, metabolic systems, tumor vascularization, etc.. Modeling these networks through the acquisition of (high throughput) time series has become a major topic in biomedical informatics. The ordinary differential equations (ODE) approach to network modeling attempts to reproduce the exact nature of the feedback circuits and the underlying biochemical and biophysical mechanisms. Unfortunately, this technique is also difficult, tedious, expensive, and time-consuming.

The models of interest are generally nonlinear, high-

dimensional, stiff, and poorly constrained. Their numerical solution requires initial values for all parameters, but the number of unknowns is extremely large. In this case, one usually chooses arbitrary parameter values that reproduce the qualitative behavior of the system. For large models, however, finding an appropriate set of values may be difficult. Although it is possible in principle to sample the whole parameter space, degenerate solutions may be expected. There are many correlations between the parameters, due to the fact that biological systems have built-in regulation mechanisms. This makes them robust to changes in their parameter values. It is the network structure, not the precise values of the parameters, that confers stability.

This paper expands on the traditional hypothesis-driven approach by proposing a reverse engineering algorithm for biological networks. Our approach is based on the idea of complementary knowledge discovery in data mining. We employ the machine learning technique of linear genetic programming (LGP) to explore the solutions space [1], [3].

Similar methods have been proposed in the past, mainly for the related problem of reverse engineering genetic network models from gene expression time series data. Most of these applications require a predefined model structure, however, and are limited to parameter estimation. The model structure is often built using traditional mathematical modeling strategies (e.g., [11]). An alternative is the S-system model, which refers to a particular type of ODE system in which the component processes are power-law functions [13], [14]. Despite the elegance and computational simplicity of the S-system model, this formalism has its limitations even for biochemical networks (e.g., [2]). Furthermore, this ODE structure cannot be applied to other important biological networks like neural networks or tumor vascularization networks.

Our algorithm discovers the network model structure and connectivity, estimates its parameters, and even identifies the biophysical and biochemical mechanisms. It takes ex-

## Flexible Web-Based System for Storing and Retrieving Information about Patients with Genodermatoses

Radu Danescu<sup>1</sup>, Sergiu Nedevschi<sup>1</sup>, Sorina Danescu<sup>2</sup>, Rodica Cosgarea<sup>2</sup>

<sup>1</sup>*Department of Computer Science, Technical University of Cluj Napoca, Romania*

<sup>2</sup>*Department of Dermatology, University of Medicine and Pharmacy, Cluj-Napoca, Romania*  
radu.danescu@cs.utcluj.ro

### Abstract

*The genodermatoses are a very vast and complex group of diseases, having a large number of symptoms and associated problems. The observation, diagnosis and treatment of these diseases are complex and experimental processes that require and generate large amounts of data, which has to be organized, stored and retrieved in a relevant and efficient way. We have designed and implemented a software system that allows a flexible definition of the structure of patient data, adapted to the specifics of each genetic dermatological disease. Text, images and binary files can be stored and retrieved easily once the forms are defined. A search engine allows the users to freely search in the database, allowing for quick access to the relevant information.*

### 1. Introduction

The international scientific activity in the field of genodermatoses is oriented towards understanding, diagnosis, treatment and prevention. The fulfillment of these goals is sought by approaching a multitude of research directions (molecular diagnosis, understanding the genotype-phenotype correlations, genetic advice, symptomatic and causal treatment).

The genodermatoses (genetic diseases of skin, GD) received lately a special attention from the international research community, due mostly to the progress achieved in the field of human genome study. One of the most documented genodermatoses is the epidermolysis bullosa, a large part of the genes and mutations involved in its pathogenesis being already known. The study of epidermolysis bullosa and the steps required for its molecular diagnosis creates the foundation of genodermatoses research.

Epidermolysis bullosa (EB) is a hereditary skin disorder characterized by blistering of the skin and mucous membranes, following minor trauma. The severity of skin manifestations can be highly variable, so that in one end of the spectrum the skin findings may be relatively minor blistering tendency primarily on the hands and feet, whereas at the other end of the spectrum the compromised integrity of the skin can

result in early postnatal demise of the affected individual [1]

The main objectives regarding GD are related to the better understanding of the mechanisms and causes of these diseases, to the improvement of diagnosis methodologies (increasing the efficiency, shortening the analysis period, reduction of costs), to the development of causal treatments besides the symptomatic and preventive ones. The immediate objectives are related to the enhancement of the knowledge base related to genes and mutations (discovery of new genes and associated mutations), and associated proteins, to synthesis of substitution proteins for causal treatment, and to the development of complex patient record systems.

On the international level, the research and medical practice in the field of epidermolysis bullosa have achieved important results in:

- establishing the etiology of the disease by the analysis of the involved genes, proteins and mechanisms
- molecular diagnosis for the forms of EB for which the mutations are discovered
- prenatal diagnosis (chorion villus biopsy or amniocentesis)
- therapy of epidermolysis bullosa: skin lesion treatment, plastic surgery treatment of the hand, surgical removal of squamous carcinoma, cutaneous grafts, nutritional therapy, genetic advice

To date defects in 10 genes are known for different forms of EB. The incidence of all forms of EB is based only on estimates; the exact prevalence is not known. According to estimates, it occurs in 1 in 50000 to 1 in 100000 births [2].

Currently epidermolysis bullosa has been divided into three major categories (EB simplex, junctional and dystrophic), based on the level of blister formation in the dermal-epidermal junction zone [3]. Epidermolysis bullosa simplex shows intraepidermal blistering and is caused by defects in the genes for keratin 5, keratin 14 and plectin. Junctional EB is defined by junctional blistering and caused by mutations in the genes for laminin 5,  $\alpha 6\beta 4$  integrin and collagen XVII.

Dermal blistering and scarring are the hallmarks of dystrophic EB (DEB). It can be inherited by autosomal dominant or autosomal recessive transmission.

## Towards the Standardization of Ultrasound Elastography: a Complete Approach

Sergiu Nedevschi<sup>1</sup>, Cosmin D. Pantilie<sup>1</sup>, Tiberiu Marița<sup>1</sup>, Daniel Moga<sup>1</sup>, Sorin M. Ducea<sup>2</sup>

<sup>1</sup>*Department of Computer Science, Technical University of Cluj-Napoca, Romania*

<sup>2</sup>*University of Medicine and Pharmacy, Cluj-Napoca, Romania*

*Sergiu.Nedevschi, Cosmin.Pantilie, Tiberiu.Marita@cs.utcluj.ro, Daniel.Moga@aut.utcluj.ro, ducea@clicknet.ro*

### Abstract

*Elastography is a newly developed dynamic ultrasound technique for measuring tissues' elasticity. Like all ultrasound based medical examination techniques, is extremely operator dependent. Beside image interpretation, elastography introduces another variability factor: manual driving of the transducer leads to a great variability in the force and frequency of the applied displacement. The severe limitations due to the high motion sensitivity and inter-operator dependency allow the method to provide only qualitative information, not having, until now, any real quantification means. In this paper we present two complementary approaches that improve the diagnostic value of elastography: the first - an electro-mechanical device for inducing controlled vibrations of the transducer and the second involving development and use of statistical methods for automatic segmentation of the elastographic images.*

### 1. Introduction

Elastography is a relatively new ultrasound based non-invasive method of imaging tissue stiffness. The elastographic images are produced by applying small, external compressions over an area of the body, and measuring the displacement of the speckle nuclei in the tissue, on the corresponding ultrasound image, as a result of the applied external pressure [1]. The specific information is displayed in color hues superimposed on a grayscale sonographic image. Pathologic changes are generally associated with changes in underlying tissues stiffness. However many disease processes, although altering the structure and mechanical properties of the affected tissue, are not accompanied by visible changes in the ultrasound images [1, 2]. Assessment and visualization of elasticity can provide potentially vital information for disease diagnosis.

The efforts made over the past two decades resulted, among other benefits, in the emergence of ultrasound scanners that also offer the elastographic imaging feature. Besides the initial application for

tumor detection, a variety of recent applications have emerged in the last few years. These include poroelastic property estimation, High Intensity Focused Ultrasound (HIFU) lesion detection and computation several tissue properties like the Poisson's ratio, the mobility of lesions and the elastic modulus, which can be determined by knowing its displacement and strain. Another significant application is the depiction of tissue components within a normal organ of a certain organized structure (for example prostates and kidneys) [3].

The extremely high motion sensitivity of the technique, although extremely useful, induces diagnostic limitations. The lack of standardization of the magnitude, direction and frequency of the applied force by manual driving of the transducer is a source of uncertainty in the color encoding procedure, providing only qualitative information.

In such conditions, the quantitative diagnostic capabilities of the method are limited [4]. For improving the diagnostic value of Ultrasound Elastography a solution is proposed in this paper.

### 2. Motion-induced errors and their correction

The above mentioned limitations affect the measurement process and therefore the resulting elastographic images. This problem becomes clearly visible when analyzing a video sequence of recorded elastographic images, while the operator constantly applies external pressure over the same area of the patient's body. Repeated results of this experiment show that the results stored in each frame of the video sequence vary across the length of the video, as depicted in figure 1.

Measurement theory explains this unusual result. Each time a quantity is measured, the measurement is affected by errors. We identified two major types of external errors that are independent of the ultrasound scanner and greatly influence the final result.

Accurate measurement is crucial as recent research [3, 5] showed that in the case of benign tumors, the

## Mathematical Models of the Heat Transfer in Tissues

Octavian Mircia Gurzău, Dorian Popa

Department of Mathematics, Technical University of Cluj-Napoca, Romania  
gurzau@math.utcluj.ro

### Abstract

In this paper we construct a mathematical model of the heat transfer on an artificial structure based on TWMB - thermal wave model of bioheat transfer- which is an extension of the classical Pennes equation. A numerical method is used to solve this equation in one dimensional case. Also we obtain an approximation of the freezing front propagation in a tissue, during a process of freezing and heating, using two identical sources. The theoretical results obtained applying a numerical procedure are compared with the experimental data from an experiment made to the Clinical Institute of Urology and Renal Transplantation Cluj-Napoca

### 1. Introduction

Cryosurgery is known as a surgical procedure from 1945 where James Arnott exposed a malign tumor of the skin to a frozen saline solution at a temperature of  $-22\text{ }^{\circ}\text{C}$ . As an effective method of treatment it was introduced by Cooper and Lee in 1961 when they presented the first cryostat for medical application.

The concept of cryosurgery as a minimal invasive technique, using multiple cryoprobes, was introduced in the middle of '80-'90, as a consequence of the development of the imagine technology in medicine, based on ultrasound and MNR (Onik-1985, Rubinsky-1993, Gilbert-1998). The first cryosurgical procedure that became a routine of treatment was in prostate cancer. The efficiency of the treatment depends on the precise placement of the cryoprobes into the tissue, which usually is irregular in shape, with different dimensions in each individual. For the planning of the cryosurgical procedure and the evaluation of the treatment effects is necessary to have a mathematical model and a computer program that can determine the dimensions of the ice ball and the distribution of the temperature in the tissue.

Modern methods of treatment in medicine as cryosurgery, cryopreservation, and therapy by radiation need understanding of thermal phenomena and the variation of temperature of living tissue. Mathematical analysis and prediction of temperature distribution in living tissues, in heating or freezing processes, are used in the study and optimization of cryosurgical procedures or irradiation procedures. In the last years, one of the most frequently approaches of the heat

transfer in living tissue are based on Pennes equation (bio-heat equation):

$$C\rho\frac{\partial T}{\partial t} = \nabla(k\nabla T) - C_b\rho_b Jf(T - T_b) + S \quad (1)$$

where  $T$  is the tissue temperature,  $T_b$  is the artery temperature,  $C$  is the specific heat,  $\rho$  is the density of the tissue,  $C_b$  and  $\rho_b$  are the specific heat and the density of the blood,  $J$  is the blood flow rate per unit tissue volume,  $K$  is the thermal conductivity of the tissue and  $S$  is the rate of metabolic heat generation. Cofactor  $f$ , ranging from 0 to 1, is due to heat exchange between arterial and venous blood flowing through the nearest vessels.

The equation of Harry H. Pennes had a great influence on the study of the heat transfer in living tissue in the last fifty years.

Propagation of the freezing front is usually described by Pennes equation with the following boundary conditions:

$$\begin{aligned} v_n\rho L &= -(k\nabla_n T)|_{\Gamma_+} + (k\nabla_n T)|_{\Gamma_-} \\ T|_{\Gamma_+} &= T|_{\Gamma_-} = T_f \end{aligned} \quad (2)$$

where  $L$  is the latent heat of the fusion,  $\Gamma_+$  and  $\Gamma_-$  are the boundaries of the freezing front of the living part and the frozen part of the tissue and  $T_f$  is the freezing temperature.

Analytical solution of Pennes one-dimensional equation was obtained by J.Liu., X.Chen and X.Xu [4]. Some important results on the problem (1)+(2) using different numerical methods were obtained in the papers [3], [4], [7], [8], [10], [11], [12], where are presented some mathematical models of the heat transfer and computer programs for the simulation of the cryosurgical procedure.

Applying the concept of the finite heat propagation velocity, Vernotte [9] and Cattaneo [1] formulated a modified unsteady heat conduction equation as:

$$q(\vec{r}, t) + \tau \frac{\partial q(\vec{r}, t)}{\partial t} = -K\nabla T(\vec{r}, t) \quad (3)$$

where  $\tau$  is defined as a thermal relaxation time in homogeneous substances, which ranges from  $10^{-8}$  to  $10^{-14}$  s. The value of relaxation time in biological systems is 20-30s. Using equation (3) Liu et al. [4] obtained the following equation (TWMB - thermal wave model of bioheat transfer):

## Ontology Learning for Traceability in the Food Industry

Ioan Salomie, Viorica R. Chifu, Sergiu Nedevschi, Emil Șt. Chifu  
*Department of Computer Science, Technical University of Cluj-Napoca*  
 {Ioan.Salomie, Viorica.Chifu, Sergiu.Nedevschi, Emil.Chifu}@cs.utcluj.ro

### Abstract

*Ontologies are considered as the most important building blocks of semantic Web. Building such ontologies is a time consuming and difficult task, which requires a high degree of human intervention. In this paper we describe a method for the enrichment of Romanian language domain taxonomies by using a text-mining approach, as well as a method to learn a domain specific ontology from scratch for the Romanian language by using machine learning techniques. In the first method, we exploit Romanian domain specific texts in order to automatically extract terms and relationships among them. A shallow parser is used to chunk texts on which we identify taxonomic relations with the help of Romanian language lexico-syntactic patterns. In the second method, the taxonomy learning process is based on an unsupervised neural network.*

### 1. Introduction

The ontologies can either be manually built or (semi) automatically derived. There is a growing need to automate the ontology acquisition process. Otherwise, building ontologies requires a high degree of human supervision and is still a bottleneck in the development of the semantic web technology.

The process of domain ontology enrichment has two inputs, the existing ontology and a domain text corpus. The former consists of the most generic concepts agreed upon in the domain and thus it plays the role of background knowledge. The aim of our method is to automatically enrich the hierarchical backbone of an existing ontology, i.e. its taxonomy, with new domain-specific knowledge [8].

In this paper we present a novel approach for the ontology enrichment of Romanian language domain ontologies. The following methods have been developed: (i) extraction and selection of domain related terms, and (ii) discovery of taxonomic relationships between terms. The term extraction process is based on recognizing linguistic patterns (essentially noun phrases) in the domain corpus documents while the taxonomic relations are identified with the help of Romanian language lexico-syntactic patterns. The whole enrichment methodology has been

implemented in an experimental prototype. It has been evaluated for the food traceability domain of knowledge, obtaining good quality results.

This paper also presents the automatic learning of a domain specific taxonomy from scratch, out of textual descriptions from Web sites [16, 17] of Romanian meat industry companies. Our taxonomy learning method is based on SOTA (Self-Organizing Tree Algorithm) [8] and extended WordNet. Extended WordNet is obtained by populating WordNet with concepts for the Romanian language.

The paper is organized as follows. Section 2 reviews several existing systems for ontology enrichment/learning. Section 3 details the taxonomy enrichment method. Section 4 gives an evaluation of the experimental results with the enrichment method. Then section 5 describes the taxonomy learning method. The conclusions are presented in section 6.

### 2. Related work

Nowadays, the multidisciplinary field of ontology learning and enrichment includes a variety of approaches and tools [2, 6]. In this section, we consider several of these methods as the closest to our methods.

[9] use a hierarchical self-organizing neural model in order to arrive at a hierarchy having only concept labels at the leaves. Then labels for the intermediate (more generic) nodes of the taxonomy are found in a bottom-up process by querying WordNet for common hypernyms of brother nodes. So, WordNet is used as an initial taxonomy to be enriched, by helping to give term names to more generic concepts in a domain. Indeed, generic terms occur less frequently in a domain corpus, and thus it isn't worth clustering them together with specific ones. Some generic terms even don't occur at all. The generic terms are acquired by querying WordNet, rather than by extracting them from the corpus followed by clustering.

In what regards other approaches for ontology enrichment, the work in [12] is worth mentioning. Here, the terms extracted from the corpus are first organized in several separate taxonomies. These specific trees are then integrated together by being appended under the appropriate nodes of the initial core domain taxonomy. Finally, the result ontology is



## **Industrial Business Process Modeling, Verification and Simulation with Application in Food Industry**

Ioan Salomie, Mihaela Dinsoreanu,  
Tudor Cioara, Ionut Anghel, Ioan Tudor Salomie  
*Department of Computer Science, Technical University of Cluj-Napoca, Romania*  
*Ioan.Salomie@cs.utcluj.ro*

### **Abstract**

*This paper addresses the problems of industrial business process modeling, verification and simulation. The paper proposes a methodology for building industrial business models, organized on layers of increasing complexity, from production line elementary machines and sensors to complex business workflows. The resulted process model is a workflow that corresponds to a certain product manufacturing technology and that can be translated, deployed and executed by any workflow execution engine. For workflow verification purposes, the industrial process model is represented in Process Algebra (PA) formalism which allows for reasoning and checking the correctness of the business process model and early identification of any logical faults in the model design phase. In order to simulate the execution of industrial workflow models for integration testing purposes, we propose a physical machine simulator based on nondeterministic, probability-based state machines. For each physical machine, a behavioral model is constructed using operational scenarios followed by its translation into a state machine representation.*

### **1. Introduction**

Business processes recently became the main elements of large scale industrial systems. Therefore, the need for their modeling and simulation became a fundamental problem. In the internet distributed computing, business processes are represented as workflows and should be able to collaborate and to be executed.

Business processes (BPs) should be defined according to business domain rules and can be classified in intra-business processes and inter-business processes. The **inter-business processes** (business to business cooperation) imply an active collaboration of a set of business partners based on specific business rules. The **intra-business processes** are modeling complex company-specific activities such as industrial workflows that drive industrial production lines. For product manufacturing traceability the intra-business process model should also contain traceability features.

The model should allow for both upstream and downstream product traceability. Upstream traceability starts from raw materials and concludes to the final product while downstream traceability takes the product and decomposes it into sub-products and traces them down until the raw materials.

The direct execution of industrial business process (intra-business processes) workflows on production line machines before the integration testing can be very expensive and may lead to improper machine operation and even to unrecoverable faults. The problem can be overcome by inserting a machine-into-workflow integration testing phase between design and run-time.

To simulate an industrial process which includes physical machines, a model that reproduces the real physical machine has to be used. Relevant simulation results are obtained only if the physical machine model accurately reproduces real machine's machine behavior. Due to their level of complexity, modeling real physical machines is not always a simple task, being difficult to describe them with precise deterministic or mathematic models.

This paper focuses on modeling, verification and simulation of intra-business processes representing industrial production lines consisting of a set of cooperating physical machines. The final result should be a verified workflow corresponding to manufacturing technology of a certain product. The paper also provide the means for simulating industrial machines, so that workflow integration testing can be done in a virtual environment. To obtain a relevant physical machine model we capture its behavioral model using machine's operational scenarios which are then transformed into an equivalent state machine representation. After the simulation, the industrial process reengineering can be used to remodel and correct the workflow faulty processes detected during the simulation process.

Our approach on verifying industrial business processes models is presented in the context of the Food Trace research project [22]. The Food Trace project aims to study and design an integrated IT system for food industry processing organizations, in response to the EU requirements regarding food traceability and quality assurance. The system models

# Logistic Chain Modeling and Generation for SOA-based Total Traceability

Ioan Salomie, Mihaela Dinsoreanu, Cristina Bianca Pop, Liviu Sorin Suci  
*Department of Computer Science, Technical University of Cluj-Napoca, Romania*  
{Ioan.Salomie, Mihaela.Dinsoreanu, Cristina.Pop, Sorin.Suciu}@cs.utcluj.ro

## Abstract

*In this paper we propose a logistic chain model enhanced with traceability features that addresses the problem of supplying products from manufacturers to the end consumers. The model formalizes the main actors, the processes they perform and the generated traceability data. We also propose a broker-based service oriented architecture that instantiates the model and provides logistics support by planning, executing and monitoring logistic chains. The architecture takes traceability aspects into consideration by providing means for accessing and collecting the traceability data from the actors of a supply chain.*

## 1. Introduction

The lifecycle of a product, from its manufacturing to its consumption, implies a complex supply chain and multiple logistic activities. Managing these activities is a difficult task as customer-imposed constraints must be taken into consideration. Among these constraints, cost-effectiveness is one of the most important. In these conditions, systems that provide logistic support in complex supply chains are essential.

In [1], authors argue that a logistics-providing system must plan, implement and control the flow of goods in a cost-effective manner. Also, a logistics-providing system should perform collaborative order management by (i) sourcing the resources (identifying resource providers based on customer constraints), (ii) scheduling and monitoring the activities that ensure the products delivery, storing and returning [2]. In this context, stand-alone, local process management information systems are not able to efficiently manage all the logistic activities among all the parties involved. We consider that a Service Oriented Architecture represents a solution for logistic chains' management as it enables inter-business communication in a cost-effective manner and reduces human intervention in the process.

Product traceability plays an important role in the management of a logistic chain. In order to ensure the product's safety and quality assurance, traceability operations must be performed throughout the whole supply chain. Traceability can be also used to obtain

vital information that may help in improving the management of product delivery, recall and recycling. The total traceability of a supply chain refers to both internal and external traceability. Internal traceability is carried out inside the infrastructure of a business partner and follows the processing flow of raw materials, intermediate and final products. External traceability keeps track of product deliveries between different business partners in the supply chain.

When performing traceability, a system must track and trace the products throughout the supply chain. Tracking follows the paths of a product along the supply chain, while tracing determines the origin and characteristics of a product [3].

Logistic chain modeling and coordination, as well as the identification of proper means for providing rapid and efficient access to traceability data spanned across a distributed network have attracted the interest of research communities.

This paper presents a logistic chain model enhanced with traceability features and a broker-based technical solution for instantiating the model. The proposed technical solution relies on SOA technologies that enable modeling a logistic chain as a composition of Web services associated to each partner involved in the chain. By adding semantic markup to the Web services' descriptions we provide an automated solution for logistic chain construction.

The rest of the paper is organized as follows. Section 2 presents other approaches for supply chain modeling and traceability. In Section 3 we present the logistic chain model enhanced with total traceability features. Section 4 describes the broker-based SOA technical solution for logistic chain generation and total traceability. The paper ends with our conclusions and future work proposals.

## 2. Related Work

In [4], authors promote the use of Web services in supply chain coordination. Authors argue that nowadays supply chains are more complex and dynamic and that in this context, older information technologies are not suitable. They propose the use of Web services as they offer real time and dynamic solutions. The authors' arguments rely on a prototype system that enables B2B procurement. A main

## Integrated systems and traceability for food-industry. Multiple input data sources

G. Saplacan<sup>3</sup>, M. Rusu<sup>2</sup>, D. Vladioiu<sup>2</sup>, M. Cerghizan<sup>3</sup>, T. Nicoara<sup>3</sup>, Gh. Sebestyen<sup>1</sup>, S. Nedevschi<sup>1</sup>

<sup>1</sup>Department of Computer Science, Technical University of Cluj-Napoca, Romania

<sup>2</sup>Technical University of Cluj-Napoca, PhD, students

<sup>3</sup>Company for Applied Informatics SA, office@cianet.ro

*gheorghe.sebestyen@cs.utcluj.ro*

### Abstract

Nowadays international and Romanian regulations impose to all actors involved in the food industry to implement systems that prove the quality of their products and offer the means to trace products (their origin, processing mode and quality) along the whole production chain, from the primary producer until the final consumer.

This paper describes the results of an interdisciplinary research project named FoodTrace that combines knowledge from the domains of food processing, quality assurance, business modeling and ITC technologies in order to offer a general traceability solution for food industry. The main goal of this project was to design an integrated traceability and quality assurance information system for a generic food processing unit. Our integrated traceability system offers the following main functionalities: food processing lines' monitoring and supervision, traceability and quality documents generation and recording and business-to-business data exchange.

In order to assure scalability and flexibility a multi-layer architecture was adopted, with a data management layer, an intra/inter-business layer and a unified GUI layer. The system is based on a Service Oriented Architecture, uses Web Services for data exchange and combines other new ITC technologies such as: MS.NET, SOAP, XML, WSDL, UDDI, and BPEL.

### 1. Introduction

The problems of food risks and food products commerce globalization brought a new world scale challenge in food quality and safety assurance generating large institutional efforts to find solutions. The Institute of Food Technologists (IFT) [23] and the European Federation of Food Science and Technology (EF-FoST) [24] jointly launched the Global Harmonization Initiative (GHI)[25] with the goal of eliminating untenable differences by "achieving consensus on the science of food regulations and legislation to ensure the global availability of safe and wholesome food products for all consumers".

Standards and regulations were created to be applied globally in this activity field. To do that, flexible and distributed food traceability information systems are needed in order to manage the complex issues regarding monitoring of food products along the whole chain,

from the primary producers through the processing units until the end consumers.

In order to have real-time access to relevant information concerning the quality of food products during the whole chain a complex distributed monitoring information system is needed. Such a system must record and offer access to the relevant parameters of a product as follows: origin, organo-leptic features, processing procedures and their parameters, preservation locations and conditions, entities and personnel involved in the product's chain, etc.

At international level there are several regulatory and legislation initiatives that are meant to establish the legal framework for the development of quality and traceability assurance systems for the food industry. The European Union through its standardization and legislation boards adopted some important documents in the field, such as: The White Charta of food safety, The White paper on the quality systems HACCP, ISO-9000, ISO-22000, Directive of Europe's Council: nr.178/2002 "General Food Law", Directive 2001/95/EC "General Product Safety Requirement, etc. The compulsoriness of assuring the required information for traceability of food products determined a significant research and development effort in which important academic and industry entities were involved[8][9]. For instance TRACE [16] is a five year project involving more than 50 European institutions and organizations. The project aims to develop generic and sector-specific traceability systems for use in the food industry. In order to fulfill the safety, quality and traceability requirements, Romania as member of the European Union initiated a set of legislative actions [21][22] and funded a number of research projects.

In the field of automatic product identification new coding systems (e.g. EPC – Electronic Product Code, UPC Universal Products Code) based on bar code or RFID technologies were proposed [1][2][6]. The Auto-ID project developed at MIT [15] is an excellent example of product identification procedure for traceability purposes.

An automated monitoring and traceability system [3][11][13] must provide customers and different supervising authorities, relevant information concerning critical stages of the products' lifecycle. Two information streams could be identified: downstream and up-

## **Strategies Concerning Research Development in Digital Library Field, Key Issues in Digital Library Implementation**

Doina Banciu\*, Gheorghe Sebestyen\*\*, Tunde Balint\*\*, Bogdan Moscaliuc\*\*

\* University of Bucharest

\*\*Technical University of Cluj-Napoca,

***Abstract*** – *During the last decade a number of national research projects aimed to develop a general framework for the implementation of library information systems and digital libraries. This paper gives an overview on these efforts emphasizing the issues and possible solutions. It also presents a digital library model dedicated for scientific and technical purposes built on a GRID infrastructure. As important functionality of this model some full text data search and retrieval techniques are presented.*

***Keywords:*** *digital libraries, metadata, content-based retrieval, technical ontology, GRID*

### **1. Strategic steps towards digital libraries**

Taking into consideration the growing importance of digital libraries and in more general terms that of digital content management systems, starting with year 2000 a number of national research projects were initiated, with the main goal of enhancing library facilities and services, through information and communication technologies. These projects aimed to develop strategies, ICT tools and infrastructures for the implementation of digitized library systems.

The SIINFO project (Information culture and communication in the information society) proposed a coherent electronic information environment that allows exchanges of cultural and educational information between the main sources such as institutions, universities, libraries, schools and other organizations.

The NUSIDOC project set as its goal the development of a national information and documentation system in science and technology; as part of this system a national virtual catalog was created that allows readers to retrieve information about bibliographic documents regardless of their geographical or institutional location. The ROLINEST portal is the unified user access point for all the libraries included in the system.

The SINRED project went further in the digitization process, its goal being the development of a digital library framework for scientific and technical purposes. This paper is based mainly on the results obtained during this project.

The latest project initiated in this field and with national impact is SIPADOC; this project tries to re-evaluate the national cultural patrimony through the newest digital technologies. Its main goal is to develop interactive tools for paper-based documents' digitization, annotation and cataloging.

The architecture of the future digital libraries (DL), as outlined in DELOS project [1], should be able to allow any users transparent access and modification of all the digital content anytime from anywhere in an efficient, effective and consistent way. The term *digital library* has a variety of potential meanings, ranging from a digitized collection of documents that one might find in a traditional library to collections of all kind of digital information along with the services that make the information useful to all possible users (e.g. Internet search engines, library systems).

A first necessary step in the process of designing a DL system is the specification of goal(s) and required features and functionalities. It is also important to establish the group of users or entities (e.g. an organization or company) to whom the DL is designated. This analysis revealed that a DL management system should offer powerful user functionalities (such as search, browse, annotation) and that it should guarantee quality of service as well (such as availability, scalability, performance). Moreover, it should be extensible, easy to install and maintain and it should assure inter-operability with other DL applications.

Current developments on Service-oriented Architectures [10], Peer-to-Peer and Grid computing promise more open and flexible architectures for DLs. With the help of these technologies researchers wish to solve the problems which arise due to increased heterogeneity of the content, services and metadata, as well as due to the omission of a central control instance.

In [3] the authors evaluate the feasibility of using a Grid infrastructure as support for a digital content management system. The outcome of this paper is that Grid computing can be used mainly for computer-intensive, in-depth search, on huge volumes of data.

This paper presents the main requirements and guidelines for the implementation of a digital content management system and offers an architectural solution for a generic digital library dedicated for scientific and technical purposes. As central part of this

# National Strategies for Digitization of Libraries, Museums and Archives in Hungary<sup>1</sup>

Katalin Bánkeszi, Péter Dippold  
*National Széchényi Library*

## 1. Introduction

To realize the main program of the i2010 initiative, the European Digital Library, participating libraries and other public collections should by all means harmonize their digitization projects on national level. The Hungarian cultural government plans to do this on the level of public collections: libraries, archives, museums and even the National Film Archive.

Digitization is in progress in Hungarian libraries, archives and museums for years, but common approach and coordination have not yet been achieved even in the specific branches. Lack of resource is only one reason for this phenomenon, lack of new approaches in certain branches, and lack of adequate competence also needs to be mentioned.

The most important task of every public collection in the field of digitization is to digitize the records (catalogues, register of fonds, item registers of museums, etc.) with a standard aspect. Only this can guarantee equal access to the documents. From this point of view, the distribution of different sectors is unequal. While almost all of the most important public and special libraries (630 libraries) have an electronic catalogue, museums have not even agreed upon a standard register system. Archives process their documents uniformly in the eArchive system on a national level. In the field of libraries, a national joint catalogue has been created. Another problem is that public collections lag behind with the so called "retrospective conversion" the digitization of card catalogues and other records. The National Széchényi Library takes immense efforts to make the card catalogue of records before 1952 accessible in electronic form as soon as possible. Presently only the data of records after 1952 are found in the OPAC of the national library.

The present official statistics of archives and museums do not contain information regarding the quality and quantity of digital documents. We have only partial knowledge of the already digitized material from individual experience and pieces of data. Only in 2007 did the questionnaires used for making library statistics include a question regarding digital documents in the

field of special libraries, but the data cannot yet be evaluated.

Although the Hungarian legal deposit act of 1998 contains a regulation regarding the submitting of electronic documents, this works only in the case of electronic documents in circulation, the creation of a regulation for the submission of e-books is still in progress in the national library.

## 2. History of the coordination of digitization in libraries and public collections

Being familiar with the Hungarian and international trends, cultural politics realized that sporadic and uncoordinated digitization in public collections sooner or later leads to chaos. That is why the Library Institute and the John von Neumann Digital Library got a commission to prepare a study dealing with the mapping and typifying of the national public asset to prepare for the coordinated digitization. The study gave commendations of the technological requirements in digitizing different types of documents, measured and typified the documents in public collections, film archives, radio- and television stations. Unfortunately, the study gave no space for the precise survey of digital contents.

In 2005, with the cooperation of the Library Institute, the John von Neumann Digital Library and the National Széchényi Library, relying on a broad team of professionals, designed the national strategy plan for library digitization between 2006-2013. Parallel to this, archives also made their IT strategy plans for the area of the archives, part of which deals with digitization. Museums continue their digitization work on institutional levels, they failed to make a branch strategy because of the diverse contents.

Branch strategies are indispensable for the planning of national digitization. However, plans remain plans, if there is not enough financial support for the project. In Hungary the government supported digitization in public collections only with smaller tenders for special areas, for a comprehensive project there was not enough resource available. This situation has changed with the start of the development supporting projects of the European Union in 2007-2008. As we all know, the EU

<sup>1</sup> This paper was also presented in the Colloquium of Library Information Employees of the V4+ Countries 6 – 8 July 2008, Brno, Czech Republic (<http://colloquium.mzk.cz/programme.php>)

## A Digital Tool for Manuscripts Preservation

Gabriela Florescu\*, Gabriela Dumitrescu\*, Valentin Florescu\*, Cornel Lepadatu\*\*

*\*National Institute for Research and Development in Informatics*

*\*\*Romanian Academy Library*

***Abstract - This paper is a presentation of a tool entitled "individual virtual manuscript collection" (RODOCIST). This tool is an innovative support for researchers for their studies of old manuscripts from archives with environmental vulnerabilities in Romania. This proposed tool is a safer alternative for the documents' owners in preserving and managing their collections or archives. It proposes an original consultation of private collections not altering the individual vulnerable archive. The method is new as it proposes a user-friendly tool for constructing RODOCIST. This method is based on secure use and interaction with digital document mirroring the original document. The method could be used interactively in creating both the digital collection and in disseminating it without involving high costs for IT resources, infrastructure or special personal expertise on IT. The access to this RODOCIST is modelled by the manuscripts' owner himself and it ranges from restricted individual access to full public access.***

### 1. Introduction

An important part of Romanian cultural heritage is represented by the old documents, manuscripts, icons, draws. They attract scientists, clerical staff and general public due to their content as well as due to their beauty and art [3]. These precious documents are kept in special places and under special environmental conditions either in archives, churches stacks or collections, so in most situations in a non accessible place for public. How to make more available these documents considering that, with time, they became more and more vulnerable due to environmental threats and the natural depreciation? The IT tool represents one good solution to be taken into account.

The attempt to digitize these old documents and manuscripts in order to make them available for electronic consultation and documentation is one possible way to increase individual or public accessibility to the manuscript collection, this important cultural information [1]. The idea is not new as there are many successful projects and results addressing the vulnerable documents such as those for the Library of Vatican [9], for United Kingdom

libraries [10] or the multinational initiative Manuscriptorium [11], to mention only some important results.

As important cultural information is also owned by Romanian monasteries and private owners we considered that a first step, towards these documents and manuscripts preservation and further dissemination, could be a computer-based collection of manuscripts, connected or not to the Internet at local or national level. At international level, the integration of this cultural collection could be obtained also pending some compatibility fulfilments as those further described.

A recommendable integration way is the Conceptual Reference Model standard (CRM), namely the International Committee for Documentation Conceptual Reference Model standard - CIDOC CRM - which is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework for modelling the cultural heritage information. International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM) had supervised the research in the field of conceptual reference model for cultural information and since 9/12/2006 CIDOC CRM is an official recommended standard ISO 21127:2006 [2] for compatibility.

The CIDOC CRM is intended to be a common language at international level for cultural heritage domain experts and implementers. It is used to formulate requirements for information systems and as a guide for conceptual modelling. In this way, CIDOC CRM can provide the framework to mediate between different cultural heritage information sources, such as that published by museums, libraries, archives or individuals. The portal Europeana – connecting cultural heritage - <http://www.europeana.eu/> - is a very important framework in which this standard could be used by digital collection owner.

Considering these international trends [4] in approaching the cultural information modelling, our aim was to design our tool for a further potential integration with other existing systems in the world. So we developed an experimental system for interactive building of a Virtual Manuscripts Collection - RODOCIST - as an alternative to the direct

## Support System for Libraries based on Business Intelligence

Dora Coardoș\*, Vasile Coardoș\*, Cornelia Lepădatu\*, Cornel Lepădatu\*\*

\*INCDI-ICI, Bucharest

\*\*Romanian Academy Library

**Abstract** - Reference centers in providing access to information and knowledge as a foundation of the culture have been and remain the institutions of cultural memory: libraries, museums, the information documentary archives. In this paper we emphasize some results of our scientific research related to the usage of multidimensional analysis tools OLAP-like (On-Line Analytic Processing) in the culture of knowledge. These results emerged from some objectives of the CEEEX research project entitled "Decision support systems in the knowledge culture based on solutions and tools in the Business Intelligence field – SSCBI". Solutions for Business Intelligence (BI) intended to improve the data analyzing process are emphasized. Multidimensional models that are intended to analyze various problems in the management of a library are presented. Finally an implementation of these models using services offered by Microsoft SQL Server 2005 is presented.

### 1. Introduction

In the current stage of evolution of mankind the information is treated as an element of infrastructure, essential in working and progress of human society, the stage is frequently defined by *the collocation information age*. In this age the academic environment already distinguished three stages of evolution: the pre-informational or informational, information society and post-information society or knowledge society.

For the knowledge society the following characteristics are essential: it is a society of science, by all its areas; is a learning society; is a durable and sustainable society; it is based on information and communication technology; knowledge becomes the key economic factor; is a society of humanistic culture.

In the field of cultural and scientific activities the essence is, according to research specialists, the transfer, dissemination, searching and receiving such information. In the field of art information is only the support intended to transport on the route of sensation, perception, intellect-awareness, sentiment-spirit the elusive emotion. In education the transfer of information and knowledge, the volume and quality are essential, but they finally subsume effect to the formative effect.

In all cases the processes that support the transfer of information aims identifying the target categories of users and their specific requirements, identifying information resources, selection of appropriate

resources, the acquisition and including in the system of the selected documents, organizing of documentary, storage, protection and preservation of their quality, the development of information products, communication and dissemination of documents and finally using of these document by users according to their goals.

Reference centers that provide access to information and knowledge as a foundation of the culture have been and remain the institutions of cultural memory: libraries, museums, the information documentary centers, archives. If in 70 years computers started to enter in the libraries, after 90 years the process reverts: libraries are starting to "enter" in computers. The new technologies of information dissemination have led to key changes in the structure, organization and functioning of services offered by institutions of cultural memory, the emphasis is transferring the functions of conservation and availability of documents toward the dissemination of information.

Demands related to the competence of the staff involved in the culture of knowledge, tasking into account the evolution in the field of library science and information technology targets the basic skills in decision-making processes in the dissemination, respectively: recognition of the need for information, identify ways of addressing the "lack" of information, building the strategies for location and accessing of information, comparing and evaluating the *information* provided by various sources, extracting the most appropriate information to satisfy the need for information.

The interest to the use of the facilities offered by information technology in the areas of decision-making aims at assisting the preparation of decisions based on analysis of data stored. The solutions and instruments offered by decisional informatics support the processes that transform data into information and information into knowledge, in substantially helping institutions in devising development strategies in the current classification of activities in these strategies, to improve operational effectiveness. Information obtained aims to the greatest extent the needs of customers, competitiveness and overall trends in the industrial, technological and economic field.

The developed algorithms, to hierarchies bibliometric entities in the system (authors, publications, publishing houses, topics) based on the indexes, by grouping documents on the basis of