BIO-ALGORITHMS AND MED-SYSTEMS



http://www.bams.cm-uj.krakow.pl

2006

EDITORIAL BOARD

EDITOR-IN-CHIEF

Professor IRENA ROTERMAN-KONIECZNA Medical College – Jagiellonian University, Krakow, st. Lazarza 16

HONORARY ADVISOR

Professor RYSZARD TADEUSIEWICZ AGH – University of Science and Technology Professor JAN TRĄBKA Medical College – Jagiellonian University

MANAGING EDITORS

BIOCYBERNETICS – Professor PIOTR AUGUSTYNIAK AGH – University of Science and Technology, Krakow, al. Mickiewicza 30

BIOLOGICAL DISCIPLINES – Professor LESZEK KONIECZNY Medical College – Jagiellonian University, Krakow, Kopernika 7

MEDICINE – Professor KALINA KAWECKA-JASZCZ Medical College – Jagiellonian University, Krakow, Pradnicka 80

PHARMACOLOGY – Professor STEFAN CHŁOPICKI Medical College – Jagiellonian University, Krakow, Grzegórzecka 16

PHYSICS – Professor STANISŁAW MICEK Faculty of Physics – Jagiellonian University, Krakow, Reymonta 4

MEDICAL INFORMATICS AND COMPUTER SCIENCE – Professor MAREK OGIELA AGH – University of Science and Technology, Krakow, al. Mickiewicza 30

TELEMEDICINE – Professor ROBERT RUDOWSKI Medical Academy, Warsaw, Banacha 1a

LAW (and contacts with business) – Dr SYBILLA STANISŁAWSKA-KLOC Law Faculty – Jagiellonian University, Krakow, Kanonicza 4

ASSOCIATE EDITORS

Medical College – Jagiellonian University, Krakow, Kopernika 7e

EDITOR-IN-CHARGE – PIOTR WALECKI E-LEARNING (project-related) – ANDRZEJ KONONOWICZ E-LEARNING (general) – WIESŁAW PYRCZAK DISCUSSION FORUMS – WOJCIECH LASOŃ ENCRYPTION – KRZYSZTOF SARAPATA

TECHNICAL SUPPORT

Medical College – Jagiellonian University, Krakow, st. Lazarza 16

ZDZISŁAW WIŚNIOWSKI – in charge WOJCIECH ZIAJKA ANNA ZAREMBA-ŚMIETAŃSKA

Polish Ministry of Science and Higher Education journal rating: 3.000

Punktacja KBN: 3.000

© COPYRIGHT BY INDIVIDUAL AUTHORS AND MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY

ISSN 1895-9091 (print version) ISSN 1896-530X (electronic version)

http://www.bams.cm-uj.krakow.pl

Contents / Spis treśći

OPENING PAPER / ARTYKUŁ OTWIERAJĄCY

3 Possible Uses of Information Technology (IT) in Anaesthesiology, Intensive Care and Emergency Medicine (eHealth) Janusz Andres

ARTIFICIAL INTELLIGENCE / SZTUCZNA INTELIGENCJA

- 5 Artificial neural networks (ANNs) as a tool of estimating relation between parameters of anti-oxidative ability and nutritional ingredients in the diet of pregnant women Emilia Kolarzyk, Aleksander Mendyk, Marek Stępniewski, Justyna Łyszczarz, Agata Pietrzycka, Agnieszka Ostachowska-Gąsio
- 13 Allergy Expert System (AES) a computer-assisted diagnosis Piotr Walecki, Magdalena Nowaczek, Grzegorz Porębski, Krystyna Obtułowicz

BIOINFORMATICS / BIOINFORMATYKA

- 21 The ligand database server Elżbieta Broniatowska, Olga Stępień, Irena Roterman
- 25 How specially adapted chromatographic bed material might be used for the controlled formation of immune complexes and studies of Congo red binding to bivalent antibodies Anna Drozd, Leszek Konieczny, Barbara Piekarska, Janina Rybarska, Barbara Stopa, Grzegorz Zemanek
- 29 Docker Analyzer: Program for comparison of docking results Artur Nielepiec, Wiktor Jurkowski

BIOMEDICAL ENGINEERING / INŻYNIERIA BIOMEDYCZNA

37 Estimation of exposure in electromagnetic fields of extremely low frequency Michał Zeńczak

PATTERN RECOGNITION / ROZPOZNAWANIE WZORÓW

43 Wavelet based classification of skin lesion images Grzegorz Surówka, Christian Merkwirth, Ewa Żabińska-Płazak, Andrzej Graca

TELEMEDICINE / TELEMEDYCYNA

- 51 DSS-MEDA: A web-based framework for video annotation in medical e-learning Paweł Mrozowski, Andrzej A. Kononowicz
- 57 Clinical Database Neonatology Krzysztof Sarapata, Joanna Woźniak, Piotr Walecki, Wojciech Kieczeń, Ryszard Lauterbach
- 65 Advanced system for organization of work and scientific data exchange Adam Piotr Żochowski, Wojciech Lasoń

E-LEARNING / E-NAUCZANIE

79 Web-based system for testing of students knowledges Lysenko Alexander, Lytvyn Igor 83 Decision support system for differential diagnosis of local edema and urticaria with allergic background Grzegorz Porębski, Piotr Walecki, Elżbieta Wardach, Krystyna Obtułowicz

MEDICINE / MEDYCYNA

91 The assessment of rehabilitation progress of patients with rheumatoid arthritis on the rheumatology ward in Provincial Hospital no. 2 in Rzeszow Krzysztof Styczeń, Mariusz Drużbicki

Possible Uses of Information Technology (IT) in Anaesthesiology, Intensive Care and Emergency Medicine (eHealth?)

JANUSZ ANDRES

Chair and Department of Anaesthesiology and Intensive Care Medicine, Jagiellonian University Hospital, Collegium Medicum, Kopernika Street 17, 31-501 Krakow, email: msandres@cyf-kr.edu.pl

Health is one of the most valuable social commodities. Every health service can derive benefit from the introduction of IT systems for management, education, diagnosis, treatment and monitoring purposes. This has been appreciated both by the US Congress and by the European Union (particularly Great Britain), which have allocated substantial funds to the development of e-Health technology within the health services of their respective countries^{1,2,3}.

Already in 2001, the world-renowned intensive care journal *Critical Care Medicine* published a review of the current possibilities for using the internet, telemedicine and other IT solutions in medicine, looking at developing widespread access to such technologies, with a view to using them in a variety of areas of healthcare⁴.

Over the past few years, many countries have experimented with electronic support within their health services, and the idea of e-Health is beginning to appear in healthcare systems at all levels. A number of basic requirements have been identified, which need to be met if IT in medicine is to offer a real improvement to the functioning of a health service^{5.8}. The first of these is **financial support**. In the long term, the introduction of IT systems in the health service will lead to considerable reductions in costs thanks to the improved quality of healthcare provided, and this is the main driving force behind the concept of e-Health. On the other hand, introducing this kind of technology entails considerable financial investment. In the USA, for instance, an initial sum of \$50m was allocated for this project, a meagre amount compared with the 6 billion pounds invested by Great Britain. The second requirement is the support and interest of leaders within the medical community: doctors, nurses, health service managers and scientists who are generally interested in IT solutions need to be presented with the aims of introducing the proposed systems and information about the areas in which they are to be implemented. In recent years, anaesthesiology, intensive care and emergency medicine have been the areas of specialisation in which IT solutions have been most widely introduced. There are already fully "electronic intensive care units" (eICU)⁵, where electronic monitoring- and treatment systems contribute to a significant improvement in the effectiveness of treating patients with life-threatening conditions. Great Britain plans to introduce a nationwide integrated IT system for both anaesthesiology and intensive care by the year 2008⁶. This system should improve patient safety as well as reduce the cost of treatment. Introducing IT systems can serve to improve the general running and efficiency of emergency departments⁷ by eliminating errors and the need for repeated hospitalisation. The third requirement for the introduction of an e-Health program is the proper **planning and organisation of the project**, which necessitates complex calculations, appropriate training and monitoring of the results as well as reliable analysis of the data obtained. The key to success (the fourth requirement) is **becoming accustomed to the new technology**. This is most likely to occur if and when the project reaches the stage of practical application. At this point, many innovative ventures do not get as far as long-term application. In situations where a project is carried out and enjoys both support and positive results, the next (the fifth) stage remains to be overcome, namely that **the health service and its employees need to adapt the project to everyday use**. The principal aim of introducing these new tools is to improve the end product, in other words better organisation of work and **a higher level of healthcare**. This final result (the sixth factor) is the most important since it concerns the patient, who needs to accept and agree to be cared for under the eHealth system. Among European countries at present, only Germany, Great Britain and Finland are relatively advanced in their implementation of IT systems in medicine, particularly **telemedicine**. Telemedicine, as a means of communication between patient and doctor(s), is becoming increasingly popular and, thanks to the state of technology available (quality of images and speed of data transmission), can be used for an ever wider range of medical specialisations.

The implementation of eHealth systems heralds the beginning of a new era in medicine and a sign that the third millennium is here. We are starting out on a journey which could lead to the improved use of our practical knowledge in the fields of molecular biology and genetics, and we already know that without the use of IT solutions this will not be possible.

Literature

- 1. US Department of Heath and Human Services. Available at: www.hhs.gov/news/press/2004pres.
- NHS Connecting for Health. A guide to the National Programme for Information Technology. Available at: www.connectingforhealth.nhs.uk.
 Commission of the European Communities. E-Health making healthcare better for European citizens: an action plan for a European e-Health area. Available at: http://europa.eu.int/eur-lex/lex/LexUriServ/site/en/com/2004.
- G. Gutierres. Medicare, the Internet, and the future of telemedicine, Crit. Care Med. 2001, Vol. 29, No. 8 (Suppl) N144-N149.
- G. Guierres. Medicare, the internet, and the future of telefieldine, Chi. Care Med. 2001, Vol. 29, Nol. 8 (Suppl) N144-N149.
 M.L. Bernstein et al. Five Constant of Information Technology Adoption in Healthcare, Hospital Topics, Winter 2007; 85, 1.
- R. Tackley. Integrating anaesthesia and intensive care into the National Care Record, Br J Anaesth 2006, 97, 69-76.
- M. Dinh, M. Chu. Evolution of health information management and information technology in emergency medicine. Emergency Medicine Australasia 2006, 18 (3), 289-294.
- 8. R.J.G. Chalmers. Health care terminology for the electronic era. Mayo Clin. Proc. 2006, 81, 729-731.

Janusz Andres Professor of medicine Jagiellonian University Medical College

ARTIFICIAL NEURAL NETWORKS (ANNS) AS A TOOL OF ESTIMATING RELATION BETWEEN PARAMETERS OF ANTI-OXIDATIVE ABILITY AND NUTRITIONAL INGREDIENTS IN THE DIET OF PREGNANT WOMEN

SZTUCZNE SIECI NEURONOWE (ANNS) JAKO NARZĘDZIE DO OKREŚLANIA RELACJI MIĘDZY PARAMETRAMI WYDOLNOŚCI ANTYOKSYDACYJNEJ I SKŁADNIKAMI DIETY KOBIET CIĘŻARNYCH

Emilia Kolarzyk¹, Aleksander Mendyk², Marek Stępniewski³, Justyna Łyszczarz⁴, Agata Pietrzycka ³, Agnieszka Ostachowska-Gąsior¹

¹Department of Hygiene and Ecology, Medical College Jagiellonian University, 7 Kopernika str., 31-034 Krakow, Poland ²Department of Pharmaceutical Technology and Biopharmaceutics, Medical College Jagiellonian University, 9 Medyczna str., 30-688 Krakow, Poland ³ Radioligand Laboratory, Pharmacy Faculty, Medical College Jagiellonian University, 9 Medyczna str., 30-688 Krakow, Poland

⁴ Chair of Orthodontia, Medical College Jagiellonian University, 4 Montelupich str., 31-155 Krakow, Poland.

Author to whom reprint requests are to be sent and all correspondence: Emilia Kolarzyk Sc.D, PhD Department of Hygiene and Ecology, Medical College, Jagiellonian University 7 Kopernika Str, 31-034 Kraków, Poland, Phone/fax: +48 (12) 4223720, e-mail: mykolarz@cvf-kr.edu.pl

Abstract: The aim of this study was to evaluate if there is a relation between superoxide dismutase (SOD) activity and total anti-oxidative status measured as Ferric Reducing Ability of Plasma (FRAP) concentration in saliva and intake of macro- and micro- nutritients in daily nutritional rations (DNR) in the first, second and third trimester of singleton uncomplicated pregnancy by using artificial neural networks (ANNs). Material and methods: Eighty women: 40 pregnant women aged 27.1+5.4 years (examined group) and 40 healthy women (the control group) were subjects of this study. Artificial neural networks (ANN) and clustering analysis with Ward's grouping method were used for evaluation of the relationship between FRAP and SOD in saliva and intake of macronutrients, minerals (Na, K, Ca, P, Mg, Fe, Zn, Cu, Mn), and vitamins (A, C, E, B1, B2, B6, PP) in DNR.

Results: During pregnancy FRAP and SOD values were lower than in the controls, but only for FRAP the differences were highly statistically significant (p<0.001). No functional relationships between FRAP and SOD and macro- and micronutrients were identified by ANN model. For whole pregnancy period cluster analysis identified 2 major clusters for which differentiating variables were SOD and retinoids intake but different patterns for each trimester of pregnancy were revealed.

Conclusion: ANNs and cluster analysis showed that there were no statistical relationships between the intake of microand macronutrients in DNR and SOD or FRAP level in saliva of pregnant women. **Streszczenie:** Celem pracy było określenie czy istnieje zależność między aktywnością dysmutazy ponadtlenkowej (SOD) i całkowitą wydolnością antyoksydacyjną (mierzoną wartością FRAP) w ślinie kobiet ciężarnych oraz podażą makro-i mikroelementów w diecie (w pierwszym, drugim i trzecim trymestrze ciąży).

Materiał i metodyka: Badaniami objęto 80 kobiet, w tym 40 ciężarnych w wieku 27.1+5.4 lat (grupa badana) oraz 40 zdrowych kobiet (grupa kontrolna). Sztuczne sieci neuronowe i analiza skupień (metodą Warda) były zastosowane dla oceny związku między wartościami FRAP i SOD w ślinie i podażą makroskładników, składników mineralnych (Na, K, Ca, P, Mg, Fe, Zn, Cu, Mn) i witamin (A, C, E, B1, B2, B6, PP) w diecie kobiet ciężarnych.

Wyniki: Podczas ciąży stwierdzono niższe wartości FRAP i SOD niż w grupie kontrolnej, ale tylko wartości FRAP różniły się w sposób statystycznie znamienny (p<0.001). Przy użyciu ANNs nie wykazano istotnych zależności między FRAP oraz SOD i składnikami diety. Dla całego okresu ciąży uzyskano metodą Warda 2 główne skupienia, w których zmiennymi różnicującymi było SOD i retinoidy, nie mniej jednak dla każdego trymestru ciąży uzyskano różne wzorce.

Wniosek: Zarówno metodą sztucznych sieci neuronowych jak i metodą analizy skupień wykazano, że nie ma statystycznie istotnego związku między podażą mikro- i makroelementów w diecie oraz stężeniem SOD i wartością FRAP w ślinie kobiet ciężarnych.

Key words: pregnancy, total antioxidant status (FRAP), superoxide dismutase (SOD), daily nutritional rations (DNR), cluster analysis, artificial neural networks (ANNs).

Introduction

Recent studies have suggested the existence of links between oxidative stress and the depletion of protective antioxidant agents, such as anti-oxidative vitamins and trace minerals (1,2,3). However, there is no consensus if maternal diet supports oxidant-antioxidant balance and increases the level of ROS sweepers in tissues and blood plasma (4,5).

In order to analyze a complex and large set of data derived from pregnant women examinations, Artificial neural networks (ANNs) were chosen. ANNs are the systems numbered among generally recognized artificial intelligence. They are defined as non-linear, information-processing systems designed in a manner similar to the biological neural structures. It is expressed in the structural and the functional composition of ANNs. The structure of ANNs is called architecture. It is usually organized in layers consisting of units, which are sometimes called "nodes". The nodes are artificial neurons responsible for information processing. Nodes from adjacent layers are usually fully interconnected by so-called "weights" simulating synaptic connections between artificial neurons. Such connectionist approach was proposed to simulate in the very simple manner cognitive functions of biological neural systems. Although it is very rough simulation of human brain ANNs are able to learn automatically on examples derived from data describing analyzed problem. This ability is main advantage of ANNs use in medical data analysis, since the complication of the problem and number of analyzed variables is not important for ANNs in such extent like for statistical methods. If there is a relationship between analyzed variables, ANNs are ideal tools to identify it, no matter what is the true nature of this particular relationship.

The aim of this study was to evaluate if ANNs may be suitable for analysis of multivariate date derived from pregnant women for the purpose of finding relationship between SOD activity and FRAP concentrations in saliva and intake of macro- and micronutrients in daily nutritional rations (DNR) in the first, second and third trimester of singleton uncomplicated pregnancy.

Material and Methods

The subjects of the study were 80 women: 40 women of the examined group and 40 women of the control group. The examined group consisted of 40 pregnant women, aged 27.1 (±5.4) years (±SD). They were clinically healthy and were pregnant for the first time. Health status during pregnancy was controlled by the same obstetrician. The subjects of the control group were matched for age, employment and life style to the subjects of examined group.

All subjects gave informed consent prior to their recruitment in the study. The Ethics Committee for Human Research of the Collegium Medicum of Jagiellonian University approved the study.

Słowa kluczowe: ciąża, całkowita wydolność antyoksydacyjna (FRAP), dysmutaza ponadtlenkowa (SOD), dieta, analiza skupień, sztuczne sieci neuronowe (ANNs)

Biochemical determinations

Saliva sampling was performed by the means of a dedicated system (Salivette, Sarsted, Germany) consisting of the tube with fitted polyethylene disposable filter and a cotton plug.

Total antioxidant power was assessed by the method described by Benzie et al. as FRAP (6). The assay is based on measuring ferric to ferrous ions reduction by a sample at low pH (300 mmol/l acetate buffer, pH 3.6). Coupled with tripyridyltriazine, blue-coloured complex is measured at 593 nm.

SOD activity was determined by the adrenaline method (according to Misra and Fridovich) based on generating free radicals in the process of conversion adrenaline to adrenochrome and the protein concentration was determined (7).

Nutritional study

The evaluation of actual food consumption of each woman was done using 24-hour nutritional recall. Recalls covered 2 non-consecutive week days and one Sunday. Recalls were the basis for assessing quantitative content, composition and nutritional value of average daily nutrition ration (DNR). The mean energy values and the levels of basic food components (proteins, total fat, carbohydrates, saturated fatty acids-SFA, monounsaturated fatty acids-MUFA, polyunsaturated fatty acids-PUFA) and their share in the daily diet of examined women were calculated. The level of minerals (sodium, potassium, calcium, phosphorus, magnesium, iron, zinc, cooper, manganese) and chosen vitamins (vit.A, vit.E, thiamine, riboflavin, niacin, vit.B6, ascorbic acid) in DNR were calculated without taking into account technological losses. All values from supplementation with mineral-vitamin formulations such as: Materna®, Prenatal®, Feminatal® (prescribed to all pregnant women by obstetrician) were added.

Data Analysis

At the very first stage of analysis artificial neural networks (ANNs) were chosen in order to create representative models for FRAP and SOD. Since there is no consistent theory of FRAP and SOD dependencies on all available data, empirical modelling using ANNs seemed to be logical choice regarding automatic mode of ANNs work. The latter is main advantage of neural modelling, which is exploited here: the machine learning paradigm allows to identify hidden and indirect relationships in the data.

Data transformation for artificial neural network (ANN) analysis.

ANN analysis was performed in two ways: a) for whole pregnancy period b) for I, II, III trimester separately

6

7

For whole pregnancy period a table of 29x120 numbers was collected and created, where 29 means parameters presented in table II and number 120 was obtained from 40 woman examined in the 1-st, 2-nd and 3-rd trimester of pregnancy (40x3). Although each woman was assayed three times during her pregnancy, it was not considered here, which data record was an assay for which woman – the dataset was treated as whole.

For particular trimester the independent dataset was created (29 parameters x 40 women) and labeled appropriately. In this case, for each one dataset associated with particular trimester, analysis was carried out separately.

Data preprocessing was used, which was specific scaling based on the following procedure:

$$\Pr epr(x_i) = \frac{x_i}{\sum_{k=1}^{N} (x_i^k)^2}$$
(1)

where:

Prepr(x) – preprocessed analyzed parameter (variable); x – original i-th parameter; i – parameter number; k – record number; N – total number of records

Scaling was based on Euclidean normalization procedure but with respect to cases (patients records) rather than variables. Such modification allowed to achieve independence of resulting data on the units range of particular variables. The procedure of data preprocessing was unified for both types of analyses: neural and statistical.

In case of neural modeling additional linear scaling was applied to ensure data working range to be between 0.2 and 0.8 or -0.8 and 0.8. Such procedure allowed fitting data to the working range of artificial neural networks, which are based generally on sigmoid-shape functions like i.e. hyperbolic tangent [9].

Artificial neural networks (ANNs) analysis

The objective of ANNs analysis was to create representative models describing FRAP and SOD dependencies on available input variables (table 2). Neural analysis was conducted on the native dataset as well as on the data preprocessed according to Eq. 1.

Various ANNs were prepared and tested (9,10) in order to create the best possible model for FRAP and SOD. One hundred different ANN architectures were tested. 10-fold cross-validation scheme was applied in order to ensure best quality of neural systems in terms of their ability to generalize beyond known database. 10-fold cross-validation procedure was as follows:

- selection of 10% of native dataset and exclusion of selected records to the test dataset
- 2) the use of remaining 90% as the training dataset

10 times repetition of p. 1 and 2 with respect to each time exclusion of different records of data. As a result of this procedure, the whole native dataset was covered by analysis in order to compute generalization error. It is easy to compute that 10 times 10% of dataset gives 100% and bearing in mind that each time different data records are included it supports claim that the generali-

zation error is be computed for the whole native dataset. The abovementioned procedure is a standard in neural analysis, which also allows to compare results between various researches.

Error measure was generalization error computed as root mean squared error (RMSE).

An attempt to reduce input vector was performed basing on the sensitivity analysis of best models (11).

All computations were performed with the use of ownwritten software **neuro_x** working under Linux environment and its Windows counterpart **Nets2004** (12,13). As hardware resources EGEE grid infrastructure was applied as well as 4 double CPU Pentium 1GHz workstations working locally.

Statistical analysis

Descriptive statistics: Since the most data had an abnormal distribution, non-parametric U-Mann-Whitney test was used for statistical evaluation of differences between compared medians. Nevertheless, for comparative purposes, in arithmetic means were also given in the tables. The significance level was set at $p \le 0.05$. The statistical package "Statistics for Windows" (Statsoft, Poland, version 6.0) was used.

Cluster Analysis: Clusterization was regarded here as model-independent approach allowing for an insight into the relationships governing analyzed problem. It was chosen in contrast to the ANNs and regression analyses which are model-dependent types of analysis. Euclidean distance and Ward grouping methods were applied in this type of analysis. Clustering was performed on the data preprocessed with the use of previously described scaling (Eq.1). Patients records were chosen as cases for clustering and 27 descriptive variables were the base for computations of clusters distances. Excluded variables were TAS, SOD and trimester description. They were treated as labels for particular cases classified to obtained clusters. In order to find differentiating features for obtained clusters there were computed average values for all variables in all obtained clusters separately. A comparison between clusters was performed using computed average variables. It was assumed that the differentiating variable would be such variable, for which the difference between average values derived from cluster A and cluster B is found to be on the high level. The differences between average variables were expressed as percentages of higher of two compared values.

Results

FRAP values were significantly higher in the 3^{rd} trimester (p<0.04) than in the 1^{st} and 2^{nd} trimesters, however significantly lower than in the control group (P<0.001). The lowest values of FRAP were observed in the second trimester (Fig.1A, Tab.1).

The differences in SOD activity in pregnant women and in the control group were not statistically significant. The considerable differences between arithmetic means and medians during 1st and 2nd trimester of pregnancy in comparison to the 3rd trimester, may be a consequence of large variability of obtained results (Fig. 2A)



Fig. 1. Total antioxidative status (FRAP) (A) and superoxide dismutase (SOD) activity (B) in pregnant women and controls.

Table 1. Values of significance levels of differences between
FRAP in pregnant women and controls.

	Control	1st Trimester	2nd Trimester	3rd Trimester
Control	Х	0,001	0,001	n.s.
1st Trimester	0,001	х	n.s.	0,04
2nd Trimester	0,001	n.s.	х	0,04
3rd Trimester	n.s.	0,04	0,04	х

Quantitative analysis of daily nutrition ration (DNR) is presented in Tab. 2

The mean intake of proteins calculated from the recalls and the amount of energy derived from proteins were in all trimesters in the range of recommended values. Fats always provided more than 30% of energy, especially in the first examination (35, 58% of energy) and it was due to the excessive intake of saturated fatty acids. The percentage of energy coming from polyunsaturated fatty acids was in the range of the recommended values (4,5% of total energy). The high intake of fats was concomitant with too low intake of carbohydrates.

The comparison of the content of minerals and vitamins in the diet of pregnant women with Polish recommendations showed low daily intake of calcium, iron, copper and riboflavin, niacin, vit. B6, especially in the first trimester. Lower intake of these minerals and vitamins was compensated by supplementation (Tab.2).

Neural analysis for whole pregnancy period

ANN modeling failed in creation of reasonable models describing relationships between input variables and FRAP and/or SOD.

It was revealed in RMSE errors values for 10-fold crossvalidation scheme, which were to high to classify the best obtained model as representative. The generalization errors were estimated as minimal RMSE for FRAP on the level 7.5 and for SOD on the level 155. A comparison with maximum values of these two variables (Table 2) ensures that neural models were far from being precise in generalization task

Additional proof of this claim was provided by correlation graphs between predicted and observed SOD/FRAP. The latter showed relatively stable output value of neural models in response to different inputs. It means that value of FRAP was always the same no matter which patient characteristics was presented to the ANN. Such behavior of neural model was inappropriate since in real life FRAP and SOD values were not constant for all patients as it was shown in Table 2. Therefore, it was a proof that this model was unable to perform its functions.

Cluster analysis for whole pregnancy period.

Since direct modeling failed in both cases for FRAP and SOD, a simple cluster analysis was applied in order to produce data patterns. There were 2 major clusters (Fig. 2) found as a result of cluster analysis.

Basing on the average values of parameters 6 differentiating variables were identified: SOD, retinol, vitamin A, beta-carotene, K and PUFA.

Selection of crucial variables based on the differences between their average values for particular clusters is shown in Table 3.

Table 3

Basing on the above results following relationships encoded in a manner of logical rules could be identified:

- 1) high SOD was associated with low intake of retinol, betacarotene, vit A, K and PUFA
- low SOD was associated with high intake of retinol, betacarotene, vit A, K and PUFA

There were no quantitative relationships between beta carotene, retinol, vitamin A and SOD, which could be identi-

No	Parameters	Average	Median	Min	Max	CV
1	FRAP (µmol/g of protein)	13.84	12.0	2	46	54.59%
2	SOD (U/g of protein)	112.88	64.5	6.213	1428.6	192.15%
3	Proteins (total) (g)	82.73	80.7	22.67	152.27	28.64%
4	Animal proteins (g)	53.79	53.4	7.74	144.16	40.43%
5	Plant proteins (g)	28.92	27.5	7.69	67.02	33.78%
6	Fats (g)	87.50	84.3	30.69	183.31	36.26%
7	Carbohydrates (g)	332.38	330.5	152.72	666.79	28.66%
8	Na (mg)	2032.84	1869.6	429	6257.54	43.05%
9	K (mg)	4310.25	4326.5	1291.73	9310.35	34.42%
10	Ca (mg)	1106.00	1017.4	409.93	2240.35	37.94%
11	P (mg)	1420.15	1395.4	475.38	2754.48	30.45%
12	Mg (mg)	349.49	341.4	133.94	793.02	31.40%
13	Fe (mg)	49.17	48.2	40.14	87.11	12.77%
14	Zn (mg)	31.38	31.1	23.71	42.23	10.50%
15	Cu (mg)	3.07	3.0	2.07	4.77	17.54%
16	Mn (mg)	6.59	6.3	2.56	12.64	30.11%
17	Vit A (µg)	2751.43	1781.4	740.23	39602.06	156.92%
18	Retinol (µg)	1131.24	466.3	96.28	37404.7	359.81%
19	Beta Carotene (µg)	7015.70	4414.8	235.55	44664.35	113.36%
20	Vitamin E (mg)	32.58	31.1	22.96	52.5	19.52%
21	Thiamine (mg)	4.56	4.5	3.54	6.84	12.33%
22	Riboflavin (mg)	5.58	5.4	4.19	15.41	22.15%
23	Niacin (mg)	38.24	36.5	23.34	70.82	22.63%
24	Vitamin B6 (mg)	12.46	12.4	10.6	15.49	7.63%
25	Ascorbic acid (mg)	322.52	290.7	133.03	815.96	44.07%
26	SFA(g)	34.34	32.9	8.35	82.7	38.11%
27	MUFA(g)	33.90	31.8	9.62	76.58	42.68%
28	PUFA (g)	11.94	11.0	2.19	44.7	53.63%
29	Cholesterol (mg)	399.40	309.4	101.2	1659	69.71%

 Table 2. Descriptive statistics of antioxidant and nutrient parameters for whole pregnancy period.

Table 3. Differences between values of antioxidant and nutrient parameters in two clusters obtained in the whole pregnancy period.

Parameters	Clus	ter A	Clus	Differer	
	Average	CV	Average	CV	
Retinol [µg]	456.93	50.11%	1537.68	332.98%	7
Vit A [µg]	1624.33	53.17%	3430.78	155.18%	Ę
Beta Carotene [µg]	4299.59	116.17%	8652.82	103.23%	!
SOD [U/g of protein]	157.20	196.86%	82.75	137.32%	4
PUFA [g]	8.12	49.05%	14.25	45.64%	4
K [mg]	3003.16	28.22%	5098.09	23.57%	4
Mg [mg]	251.08	23.44%	408.80	21.67%	3
Plant protein [g]	21.21	31.04%	33.56	24.88%	3
Ascorbic acid [mg]	239.52	34.03%	372.55	39.66%	3
Carbohydrates [g]	261.95	28.76%	374.83	21.30%	3
Mn [mg]	5.24	28.92%	7.40	24.14%	2
P [mg]	1131.95	33.88%	1593.85	22.81%	2
Na [mg]	1644.11	32.55%	2267.14	42.19%	2
MUFA [g]	27.53	46.64%	37.74	37.38%	2
Fats [g]	71.56	38.00%	97.12	31.43%	1
Cholesterol [mg]	327.90	54.24%	442.49	71.82%	1
Cu [mg]	2.56	8.82%	3.37	12.81%	2
Protein [g]	69.54	33.89%	90.68	22.12%	2
Vit E [mg]	28.05	11.62%	35.31	17.61%	2
SFA [g]	29.60	39.10%	37.19	35.46%	2
Niacin [mg]	33.01	19.85%	41.39	19.95%	2
Ca [mg]	974.32	46.12%	1185.38	32.25%	1
Animal protein [g]	48.28	49.58%	57.10	34.57%	1
Riboflavin [mg]	5.01	11.52%	5.92	23.61%	1
Thiamin [mg]	4.15	9.04%	4.81	10.61%	1
Fe [mg]	45.23	5.40%	51.55	12.98%	1
Zn [mg]	29.13	8.67%	32.74	9.01%	1
Vit B6 [mg]	11.65	4.98%	12.95	6.06%	1
FRAP [umol/g of protein]	13.56	56.09%	14.00	54.08%	

Parameters	Clus	ster A		Cluste	er B	Difference	
	Average	CV	1	Average	CV		
SOD [U/g of protein]	155.46	177.90%	1	46.23	58.43%	70.26%	
Mg [mg]	269.97	34.70%	1	419.44	18.90%	35.64%	
K [mg]	3392.81	35.87%	1	5169.75	24.26%	34.37%	
Mn [mg]	5.51	28.10%	1	8.14	18.78%	32.33%	
P [mg]	1062.41	32.40%		1566.14	15.09%	32.16%	
Plant protein [g]	24.62	26.32%	1	36.27	25.18%	32.13%	
Carbohydrates [g]	270.99	25.54%		385.98	21.45%	29.79%	
PUFA [g]	9.74	49.35%	1	13.66	47.61%	28.73%	
Protein [g]	67.45	34.00%		94.58	20.53%	28.68%	
Animal protein [g]	42.81	47.92%		58.28	31.58%	26.56%	
Ca [mg]	852.35	38.23%	1	1105.77	23.27%	22.92%	
Ascorbic acid [mg]	281.92	38.34%		360.37	46.19%	21.77%	
Cu [mg]	2.73	13.52%		3.40	13.98%	19.71%	
Niacin [mg]	35.05	22.21%		41.53	20.09%	15.61%	
Fats [g]	85.76	43.93%		101.13	22.46%	15.20%	
Vit E [mg]	30.26	16.21%		35.54	21.24%	14.85%	
Na [mg]	1958.23	64.98%		2280.57	27.03%	14.13%	
Thiamin [mg]	4.33	10.92%		5.02	9.14%	13.85%	
Retinol [µg]	456.29	44.93%		529.50	46.23%	13.83%	
Riboflavin [mg]	4.96	10.67%		5.72	10.33%	13.37%	
FRAP [µmol/g of protein]	14.55	43.16%		12.61	71.70%	13.30%	
MUFA [g]	34.96	49.94%		40.09	26.83%	12.80%	
SFA [g]	33.88	47.00%		38.76	23.75%	12.58%	
Zn [mg]	29.41	7.86%		33.02	7.68%	10.92%	
Beta Carotene [µg]	6334.61	115.50%		7056.38	97.99%	10.23%	
Fe [mg]	46.13	5.50%		50.77	6.30%	9.13%	
Vit A [µg]	1962.60	65.54%]	2155.93	53.02%	8.97%	
Vit B6 [mg]	11.87	7.13%]	12.98	6.23%	8.54%	
Cholesterol [mg]	327.51	48.82%]	343.13	44.90%	4.55%	

Table 4. Differences between values of antioxidant and nutrient parameters in two clusters obtained in the first trimester

fied neither in neural analysis nor in attempted linear regression. It is worth to mention that in the clusterization procedure SOD, FRAP and trimester variables were excluded from clusters' descriptions thus treated afterwards as labels for particular cases belonging to their clusters. Regarding this, high intake of retinoids is generally protective in pregnancy however no quantitative formula could be identified allowing



for assessment of how much of retinoids should be taken in order to lower SOD level for a particular quantity.

Results for I, II, III trimester separately

Inconsistency between ANN results and classical statistical methods were the cause of following analysis conducted with split datasets representing each trimester respectively. Following results were obtained with the use of preprocessed data only.

Cluster analysis for first trimester allowed to select two main clusters showing no particular difference between vit A, retinol and beta carotene intake, although SOD levels remained to be differentiating variable (Table 4).

For second trimester SOD levels remained differentiating variable but there were found also large differences between vit A, retinol and beta carotene average intake (Table 5).

For the third trimester a tendency to more uniform distribution of SOD levels but with lower average value was noticed, however vit A, retinol and beta carotene were remaining as differentiating factors with elevated levels of intake.

Another important feature of derived results has to be considered. Although, in the second trimester high uptake level of vit A, retinol and beta carotene was associated with low SOD level (Table 5), for the third trimester this relationship was shown to be completely reversed: high level of vit. A, retinol and beta carotene was associated with high level

Parameters	Clus	ter A	CI	Cluster B				
	Average	CV	Average	CV				
SOD [U/g of protein]	60.71	39.16%	277.11	167.69%	78.09%			
Retinol [µg]	1834.84	255.74%	442.10	46.97%	75.91%			
Vit A [µg]	3810.41	128.83%	1558.07	36.45%	59.11%			
Beta Carotene [µg]	9147.71	128.46%	3989.53	78.53%	56.39%			
PUFA [g]	13.20	39.18%	7.54	43.67%	42.86%			
Plant protein [g]	31.71	28.44%	18.81	31.82%	40.68%			
K [mg]	4931.99	22.37%	3172.34	27.97%	35.68%			
Mg [mg]	396.32	16.09%	259.03	16.74%	34.64%			
Carbohydrates [g]	369.86	23.53%	252.28	26.91%	31.79%			
Ascorbic acid [mg]	350.10	34.99%	239.20	37.84%	31.68%			
Mn [mg]	6.95	19.97%	4.77	24.13%	31.32%			
Na [mg]	2117.72	39.69%	1623.72	31.98%	23.33%			
Cu [mg]	3.31	12.74%	2.55	9.10%	22.86%			
P [mg]	1592.56	19.44%	1275.89	30.00%	19.88%			
Vit E [mg]	33.59	13.54%	27.37	8.86%	18.54%			
Ca [mg]	1281.63	32.17%	1078.35	48.26%	15.86%			
Niacin [mg]	41.20	20.53%	34.84	26.43%	15.45%			
MUFA [g]	33.94	38.92%	29.03	34.89%	14.46%			
FRAP [µmol/g of protein]	12.03	53.37%	13.86	71.93%	13.15%			
Riboflavin [mg]	5.96	18.11%	5.17	9.19%	13.15%			
Fats [g]	87.88	30.52%	76.62	28.96%	12.81%			
Protein [g]	87.42	16.89%	76.53	35.66%	12.46%			
Fe [mg]	51.53	16.13%	45.59	6.73%	11.52%			
Thiamin [mg]	4.58	8.64%	4.14	9.27%	9.43%			
Vit B6 [mg]	12.83	6.05%	11.78	6.41%	8.22%			
Zn [mg]	31.99	6.86%	29.62	7.56%	7.40%			
Cholesterol [mg]	440.17	69.47%	420.64	48.32%	4.44%			
Animal protein [g]	55.70	28.03%	57.68	52.83%	3.44%			

Table 5. Differences between values of antioxidant and nutrient parameters in two clusters obtained in the second trimester.

Table 6. Differences between values of antioxidant and nutrients parameters in two clusters obtained in the third trimester.

33.53%

33.17

28.87%

34.26

SFA [g]

Variable	Clus	ter A	Clus	Difference		
	Average	CV	Average	CV		
Retinol [µg]	2392.31	336.04%	530.59	53.29%	77.82%	
Vit A [µg]	4435.84	184.51%	1634.31	51.48%	63.16%	
Beta Carotene [µg]	9552.76	72.52%	3915.75	119.01%	59.01%	
SOD [U/g of protein]	114.19	146.60%	62.76	52.77%	45.04%	
Ascorbic acid [mg]	415.42	41.11%	244.00	34.41%	41.26%	
K [mg]	5398.61	25.20%	3175.14	30.71%	41.19%	
Cholesterol [mg]	525.14	82.29%	318.61	69.93%	39.33%	
Mg [mg]	438.08	25.11%	266.63	25.04%	39.14%	
Plant protein [g]	35.61	21.56%	22.03	24.15%	38.15%	
PUFA [g]	15.68	53.97%	9.80	56.62%	37.47%	
Mn [mg]	8.11	25.18%	5.32	25.11%	34.38%	
P [mg]	1746.38	25.39%	1166.06	34.67%	33.23%	
Fats [g]	100.05	33.60%	68.58	48.03%	31.46%	
SFA [g]	38.54	38.38%	26.49	47.76%	31.27%	
MUFA [g]	37.02	40.72%	26.27	56.91%	29.04%	
Na [mg]	2351.91	36.08%	1672.75	37.37%	28.88%	
Carbohydrates [g]	393.83	20.58%	282.42	25.68%	28.29%	
Protein [g]	96.41	23.29%	70.45	31.86%	26.93%	
Cu [mg]	3.52	12.31%	2.62	8.89%	25.53%	
Niacin [mg]	41.97	20.25%	32.51	14.40%	22.53%	
Animal protein [g]	60.78	38.66%	48.35	43.83%	20.45%	
Riboflavin [mg]	6.25	35.37%	5.11	13.86%	18.11%	
Ca [mg]	1235.98	34.50%	1016.83	43.86%	17.73%	
Thiamin [mg]	5.00	11.96%	4.14	6.74%	17.16%	
Vit E [mg]	36.19	18.29%	30.42	22.11%	15.96%	
Zn [mg]	34.39	10.21%	28.95	9.90%	15.83%	
Fe [mg]	53.60	13.89%	45.14	4.49%	15.79%	
FRAP [µmol/g of protein]	16.25	46.23%	13.97	49.97%	14.00%	
Vit B6 [mg]	13.15	6.30%	11.78	4.04%	10.39%	

11

3.17%

of SOD (Table 6). The levels of intake of K and PUFA were changing in the similar pattern.

Discussion

The examination of obtained data showed decrease values of FRAP in the second trimester of pregnancy. In this period of pregnancy, women may be not fully adapted to increased demands for anti-oxidative mechanisms. In the third trimester increase of FRAP concentration was observed, though statistical analysis showed that it was not connected with maternal diet.

A failure of neural modeling is not a failure of the whole research since there are other non-modelling methods which might be used here. Main conclusion is that if there is no reasonable ANN model found after extensive automatic search for a such one, then maybe there is no strict, quantitative relationship between FRAP and patients' characteristics. The same is applicable to SOD. It is therefore the conclusion of lack of particular relationship, which bases on the ANNs features, where the failure of modeling cannot be assigned to the problem complexity or ineffectiveness of a priori selected equation. ANNs application helped, in spite of direct modeling failure, to discover some true aspect of analyzed problem. As a consequence of ANNs results a cluster analysis was performed which provided another information but this time of positive outcome.

Conclusions

- FRAP values were the lowest in the second trimester. It suggest that in this trimester of uncomplicated pregnancy women may be not fully adapted to increased demands for anti-oxidative mechanisms.
- ANNs and cluster analysis showed that there were no statistical relationships between the intake of micro-and macronutrients in daily nutritional rations and superoxide dismutase activity or total antioxidant status (FRAP) level in saliva of pregnant women.

References

- C. Biondi, B. Pavan, L. Lunghi, S. Florini, F. Vesce. The role and modulation of the oxidative balance in pregnancy. Curr. Pharmac. Design, 2005, 11, 2075-2089.
- J.M. Roberts, G. Pearson, J. Cutler, et al. Summary of the NHLBI Working Group on research on hypertension during pregnancy. Hypertension, 2003, 41, 437-445.
- J. Villar, J.M. Belizan. Same nutrients, different hypotheses: disparities in trials of calcium supplementation during pregnancy. Am. J. Clin. Nutr., 2000, 71, 1375S-1379S.
- I.R. Record, I.E. Dreosti, J. . Mc Inerney. Changes in plasma antioxidant status following consumption of diets high or low in fruit and vegetables or following dietary supplementation with an antioxidant mixture. Brit. J. Nutr., 2001, 85, 459-464.
- S. Bolissety, et al. Antenatal supplementation of antioxidant vitamins to reduced the oxidative stress at delivery – a pilot study. Early. Hum. Dev., 2002, 67, 91-92.
- I. Benzie, J. Strain. The Ferric Reducing Ability of Plasma (FRAP) as a measure of "Antioxidant power": the FRAP assay. Annal. Biochem., 1996, 239, 70.
- H.P. Misra, I. Fridovich. The role of superoxide anion in the autooxidation of epinephrine and a simple assay for superoxide dismutase. J. Biol. Chem., 1972, 247, 3170-3175.
- Ś. Ziemlański, B. Bułhak-Jachymczyk, J. Budzyńska-Topolowska, et al. Recommended dietary intake for Polish population (energy, protein, FAT, vitamins and minerals). New Med., 1995, 2, 1-27.
- J.M. Żurada. Introduction to Artificial Neural Systems. West Publishing Company USA 1992.
- R.R. Yager, D.P. Filev. Essentials of fuzzy modeling and control. John Wiley & Sons, Inc. 1994.
- J.M. Żurada, A. Malinowski, S. Usui. Perturbation Method for Deleting Redundant Inputs of Perceptron Networks, Neurocomputing, 1997, 14, 177-193.
- A. Mendyk, R. Jachowicz. Neural network as a decision support system in the development of pharmaceutical formulation – focus on solid dispersions. Expert Syst. Appl. 28:285-294, 2005.
- S. Polak and A. Mendyk, Artificial Intelligence Technology as a tool for initial GDM screening. Expert Syst. Appl., 2004, 26, 455-460.



ALLERGY EXPERT SYSTEM (AES) – A COMPUTER-ASSISTED DIAGNOSIS

ALERGOLOGICZNY EKSPERTOWY SYSTEM (AES) – KOMPUTEROWE WSPOMAGANIE DIAGNOZY

PIOTR WALECKI*, MAGDALENA NOWACZEK***, GRZEGORZ PORĘBSKI**, KRYSTYNA OBTUŁOWICZ**

* Department of Bioinformatics and Telemedicine, Collegium Medicum, Jagiellonian University, Kopernika 7, 31-501 Krakow, Poland

** Departament of Clinical and Enviromental Allergollogy, Collegium Medicum, Jagiellonian University, Sniadeckich 10, 31-531 Krakow, Poland

*** Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Reymonta 4, 30-060 Krakow, Poland

Abstract: This paper presents an expert system which assists medical diagnosis in allergy. After entering symptoms of a disease, the system supports the doctor in undertaking decisions and suggests the most probable diagnosis. Next, the system shows possible treatment and prevention for the diagnosed allergy. The system informs the doctor about needed examinations and may describe how a diagnostic inference should proceed.

Information about allergy is necessary for planning effective therapy and education, therefore the system is enlarged by educational elements, which may be used as informational materials for patients.

The expert system proposes diagnosis and treatment only; the final decision should be performed by a physician.

The system was built in the PC-Shell programming environment. The PC-Shell is a first Polish commercial shell expert system. It is a hybrid system, which combines different methods of solving problems and presenting results.

The project was designed in cooperation with the Department of Clinical and Environmental Allergy at the Jagiellonian University, Medical College.

Key words: expert system, allergy, medical diagnosis

Introduction

Expert systems are used in various domains of science and technology, but the most important usage is in medicine and healthcare. Medical expert systems are designed to deliver computer-assisted diagnosis and advisory systems (clinical guidelines). They work by using patients' clinical data as an input data for processing and the results are presented as a clinical recommendation.

Expert systems in medicine are used also in laboratory data analysis, data mining and automatic maintenance of medical devices.

Streszczenie: W pracy zaprezentowano system ekspertowy wspomagający diagnozę w alergologii. Po wprowadzeniu objawów zaburzeń system wspomaga lekarza w podejmowaniu decyzji i sugeruje najbardziej prawdopodobną diagnozę. Następnie przedstawia możliwe działania lecznicze lub profilaktyczne dla stwierdzonej choroby alergicznej. System również informuje o badaniach klinicznych, które należy przeprowadzić oraz może opisać jak przebiegała ścieżka diagnostyczna na podstawie podanych objawów.

Znajomość wiedzy o alergiach jest niezbędna w planowaniu efektywnej terapii i edukacji. Dlatego też system został wyposażony w elementy edukacyjne, które mogą zostać wykorzystane jako materiały informacyjne dla pacjenta.

Opisany system przedstawia tylko sugestie dotyczące postępowania, jednak ostateczne decyzje należą zawsze do lekarza. Całość systemu wykonana została w aplikacji PC-Shell, która jest szkieletowym systemem ekspertowym. System PC-Shell jest systemem o architekturze hybrydowej, tj. łączącej w sobie różne metody rozwiązywania problemów i reprezentacji wiedzy. Projekt powstał we współpracy z Zakładem Alergologii Klinicznej i Środowiskowej Collegium Medium Uniwersytetu Jagiellońskiego.

Słowa kluczowe: system ekspertowy, alergia, diagnoza medyczna

The most well known medical expert systems in use are:
 MYCIN – is an expert system developed in the early 1970s at the Stanford University. Its job was to diagnose and recommend treatment for certain blood infections. It was written in LISP and operated using a simple inference engine, and a knowledge base of ~500 rules [1].

 CASNET – is an expert system for the long-term management of diseases whose mechanisms are well-known.
 First was developed a Causal-ASsociational NETwork model for describing disease processes. CASNET is an expert-level consultation program in glaucoma [2], [3].

- SETH is an expert system which simulates expert reasoning, taking into account each toxicological class, delay, clinical symptoms and ingested dose. It generates accurate monitoring and treatment advice, addressing also drug interactions and drug exceptions [4].
- VentEx is a knowledge-based decision-support and monitoring system applied in ventilator therapy. The VentEx system is used both for monitoring and decision-support in patients with different kinds of imminent and obvious ventilatory insufficiencies [5].
- GIDEON (Global Infectious Diseases and Epidemiology Online Network) – is an easy to use, interactive and comprehensive web based expert system that helps overcome information overload while saving time through quick access to a vast knowledge database.
 GIDEON is used for diagnosis and reference in the fields of tropical and infectious diseases, epidemiology, microbiology and antimicrobial chemotherapy [6].

Other medical expert systems in use are:

- ADE (Adverse Drug Event) Monitor
- Apache III (Acute Physiology and Chronic Health Evaluation)
- CaDet (Early Cancer Detection)
- CADIAG-II (Computer-Assisted DIAGnosis)
- CHDR (Coronary Heart Disease Risk)
- CMD (Computerized Medical Diagnosis)
- FACTS (Finding Appropriate Clinical TrialS)
- HELP (Health Evaluation through Logical Processes)
- PEIRS (Pathology Expert Interpretative Reporting System)
- PEPID (Portable Emergency Physician Information Database)
- POEMS (Post Operative Expert Medical System)
- QMR (Quick Medical Reference)
- T-IDDM Project (Telematic Management of Insulin Dependent Diabetes Mellitus)
- VIE-PNN (Vienna Expert System for Parenteral Nutrition of Neonates)

Expert system building tools are:

- Special programming languages like LISP (LISt Processing) or PROLOG (PROgramming in LOGic).
- Special skeletal systems, shells, or AI tools like CLIPS (C Language Integrated Production System), MANDA-RAX, Jess or PC-Shell. The expert systems shells are a suite of software that allows constructing a knowledge base and creating interaction between knowledge base and an inference engine.

Building expert systems by using programming languages offers significant advantages: flexibility and specialization, but on other hand it is expensive and labour-intensive. Expert systems shells have simplified programming. A system can be built by entering into the shell all the necessary knowledge about a task domain. In general, they don't help with knowledge acquisition, a task currently performed by knowledge engineers. The large accumulation of high-quality knowledge is the most important for the efficacy of expert systems ("the more knowledge a system is given, the more competent it becomes"). Expert system consists of four principal parts (see fig. 1):

- knowledge base
- reasoning system (inference engine)
- explanation system
- knowledge acquisition system



Fig. 1. The building blocks of expert systems [7]

Material and Methods

The materials used in the project are comprised with relations between allergic diseases and their symptoms, diagnostic examinations and treatment & prophylactic. The medical data was prepared in cooperation with Departament of Clinical and Environmental Allergollogy, Collegium Medicum, Jagiellonian University.

The medical data is organized in four groups:

- 1. Symptoms
- 2. Allergic diseases
- 3. Treatment and prophylactic
- 4. Diagnostic examinations

Each group consists of six domains, which describe the anatomical localization:

- 1. Lungs
- 2. Nose and perinasal sinus
- 3. Skin
- 4. Eyes
- 5. Intestinal tract
- 6. Others

The expert system design and implementation is preformed in PC-Shell environment. PC-Shell is the first Polish commercial expert system shell. It was developed as a result of earlier research on creating expert system like PC-Expert (1985-87) and Diagnosta MC 14007 (1988). PC-Shell was awarded on International Software Trade Fair SOFTARG in 1994 and in 1996. It was often presented on international conferences and seminaries in Poland and other countries [8].

14

PC-Shell is called a knowledge base language. It serves as a formal description of domain-expert knowledge. This type of language is generally used by knowledge engineers, allowing them to build expert systems for a specific domain. Knowledge is programmed in PC-Shell by rules and facts. It is represented in the system thanks to a triad: object attribute - value. Object identifier and attributes are symbols that begin from small letter, after which may follow any sequence of alphanumeric signs, that includes letters, numbers, and also underline signs "_". Attributes may have values both real numbers and strings, and can be represented by variables.

Syntax of the triad: object – attribute – value (OAV) atrybut {obiekt} operator relacji liczba

or

```
atrybut {obiekt} = łańcuch_znakowy
or
```

atrybut {obiekt} = zmienna

In the triad, both, the object and the value, can be omitted, but the attribute is necessary because it is the fundamental element.

Knowledge in PC-Shell is described by five functional modules:

- Sources
- Facets
- Rules
- _ Facts
- Control

At least one module of facts or rules must appear in the system.

General structure of knowledge base description: knowledge_base_name sources source_knowledge_description end: facets facets_description end: rules rules__description end: facts facrs description end; control program end: end:

The PC-Shell uses blackboard architecture, therefore the system contains a main control module, and the whole knowledge must be placed in the source module. In the system at least one module of rules, facts or control must appear.

Sources of knowledge description

The PC-Shell is a hybrid system. This means for problem solving may use many heterogenic sources of knowledge. In actual PC-Shell version 4.1 we may use following types of knowledge bases:

- kb expert bases of knowledge
- neural_net applications based on neural nets _
- _ metaphor - data bases with a metaphor explanation _
 - what_is data bases with "what is?" explanation

Such a source knowledge description should include at least one description that consists of:

- name of this source, which is freely established by knowledge engineer
- specification of this source features, which consist of expressions including system key words

A general form of source_knowledge_description: characteristic _1 ... characteristic _n;

Type FILE determinate a file where source of knowledge is stored.

Facets description

In PC-Shell, a facet is a set of declarations that relate to chosen attributes. The facet module is a register of all attributes which are used in knowledge base and also included in knowledge sources. In the system not all attributes must be described through facets, but all must be declared in a facet module.

The facets description consists of global declarations ask and single, as well as of a set of attributes and connected with them facets. The ask declaration allows to put a question through the system to verify correctness of rules.

A general structure of facets description: [ask {yes | no};] [single {yes | no};] attribute_1 [declaration_facet_1];

attribute_n [declaration_facet_n];

Facts and rules description

Facts module allows writing fixed information which may be knowledge base parameters. A fundamental element that is used in facts description is the triad OAV (object - attribute - value). It differs however from general triad syntax OAV, because facts here can not include variables, and the only operator between the attribute and the value is the "=" sign.

In a PC-Shell language rules module fulfills a main role in a point of view of expert knowledge representation. In this language rules have a universal character that allows encoding of the knowledge from almost each domain. A standard rules syntax consists of conclusions and conditional part, that are separated by a key word if. The conditional part must contain at least one condition.

Programming in PC-Shell

In general, shell systems, comparing to programming languages, are much more efficient tools for building domain expert systems. However, it often comes at the cost of flexibility. In PC-Shell this problem is solved, because it is a hybrid system. It joins declarative knowledge representation with procedural languages. Control module contains all instructions used in a project. It allows separating expert knowledge from control.

- A program has two parts:
- variables declaration
- instructions set

Results

16

General project description

The aim of project was to create an expert system which assists medical diagnosis in allergy. After entering the symptoms of a disease, the system will support the doctor in undertaking decision and will suggest the most probable diagnosis. Next, the system shows possible treatment and prevention for the diagnosed allergy. The system informs about examinations that are to be conducted and may describe how diagnostic inference should proceed.

The system is designed mainly for physicians to aid diagnostic decisions. However, the expert system proposes diagnosis and treatment only, the final decision should be done by a physician.

Information about allergy is necessary for planning effective therapy and education. Therefore the system was enlarged by educational elements, which may be used as informational materials for patients.

Work with the system

The system works as a consultation with a user, that means, the system asks and eventually the user gets an answer. The consultation begins with opening a consultation window (see fig. 2). This window shows symptoms of allergy diseases. The system asks the user to choose symptoms that ail the patient.

Next, the system makes a diagnosis using rules contained in knowledge source. The system shows diagnosed



Fig. 2. Consultation window

allergy in a decision window, where we are presented with specifics of the disease, eg. photo (see fig. 3).



Fig. 3. Decision window

Verifing how the system made a diagnosis is simple. By clicking the "How?" button ("Jak?" in Polish) the system shows a decision tree window where steps of the diagnosis are presented (see fig. 4).

In this window we see the decision tree elements:

- A solution (conclusion) generated through the system
 - rule defined in the knowledge base
 - condition included in a rule

– fact included in the knowledge base or created as a result of question

9 JAK : choroba_alergiczna = "astma oskrzelowa"	? 🗙
 KONKLUZJA: choroba_alergiczna = "astma oskrzelowa" 1: choroba_alergiczna = "astma oskrzelowa" JEŚLI 1" Fakt : objawy = "duszność" 3" Fakt : objawy = "iwiszczący oddech" KONKUZJA: choroba_alergiczna = "astma oskrzelowa" JEŚLI 11: choroba_alergiczna = "astma oskrzelowa" JEŚLI 4" Fakt : objawy = "ściskania w klatce piersiowe]" 2" Fakt : objawy = "kaszel" 	
	~
Wyjaśnienia szczegółowe	[] 🥑

Fig. 4. Explanation - decision tree window

Second part of the system is educational. It allows getting information about a disease, its treatment, prophylactic and examinations that should be taken. It is possible by opening another window with a question: "Which disease was diagnosed?". In the next window needed information is shown. The system then asks to choose a treatment from the examinations list (see fig. 5).



Fig. 5. List of examinations

Structure of the system

Allergic diseases and their symptoms are the two main groups of the system elements. A process of assisting a diagnosis depends on relation (rules) between these two groups. These two groups, like the diagnostic examinations and the treatment of allergic diseases, are additional medical data that make the diagnosis deeper and propose adequate treatment.

Detailed schema of medial data thanks to which the system was created is show in fig. 6 and fig. 7.

The system implementation

All of project's medical data included in the knowledge source consists of two parts:

- Facets module
- Rules module

Facets module is a set of declarations relating to chosen attributes. Two types of facets are used in the project:

- ask
- pictures

The facet *ask* determinates whether the system can put a question concerning given attributes, e.g.:

facets

ask yes;

objawy: query "[A[0[B[G OKRESL JAKIE WYSTEPUJA OBJAWY [a[0[b[g:"

val someof {"duszność", "kaszel", "świszczący oddech", "ściskanie w klatce piersiowej", "katar", "kichanie", "przeziębienie", "zatkanie nosa", "ból głowy", "obrzęki w różnych lokalizacjach", "obrzęki powiek","podwyższona temperatura ciała", "obrzęki stóp i podudzia", "świąd skóry", "suchość skóry", "szorstka i zgrubiała skóra","osłabienie", "wysypka grudkowa", "wysypka pokrzywkowa(bąbel)",

"zaczerwienienie skóry(rumień)", "zaczerwienienie spojówki oka", "pieczenie oka", "obrzęk jamy ustne",

"obrzęk gardła i/lub krtani", "biegunka", "ból brzucha", "pogorszenie samopoczucia"}; The facet *picture* allows linking attributes with pictures (photos), e.g.:

choroba_alergiczna:

val someof {"astma oskrzelowa", "alergiczne zapalenie pęcherzyków płucnych", "alergiczny nieżyt nosa", "polip nosa", "atopowe zapalenie skóry", "obrzęk naczynioworuchowy wrodzony", "pokrzywka i obrzęk naczyniowy", "wyprysk kontaktowy",

"ponzywka i obrzęk naczyniowy , "wyprysk kontaktowy , "alergiczne zapalenie spojówek", "alergie pokarmowe", "anafilaksja", "nadwrażliwość na leki"} picture {"","","","","16.bmp", "18.bmp", "20.bmp", "21.bmp",

,", "", "", ""};

Rules module fulfills a main role in a point of view of expert knowledge representation. Rules descriptions consist of a set of rules preceded with a word *rules* and ended with a word *end*. Rules syntax consist of conclusions and a conditional part, that are separated with a key word *if*. rules

1: choroba_alergiczna="astma oskrzelowa" if objawy="duszność" & objawy="świszczący oddech"; 11: choroba_alergiczna="astma oskrzelowa" if objawy= "ściskanie w klatce piersiowej" & objawy="kaszel"; end;

Control module consists of two parts: variable declarations and instructions. The program searches needed information contained in the knowledge source through a function *solve*:

solve (choroby_alergiczne, ,choroba_alergiczna=Choroba_alergiczna");

Access to this knowledge is enabled by a *source* module: sources

choroby_alergiczne: type kb file "C:\\program1\\projekt.zw"; end;

Discussion

Allergy diseases belong to chronic diseases, and its diagnosis is usually long-lasting and laborious, it burdens the patient as well as the health care system. Allergy diseases have different base, symptoms and treatment. Therefore we created an expert system for general diagnosis and to help assist in choosing a therapy. The presented system seems to be the right first step towards using Al methods in medical practice. Introducing these advanced computer technologies may bring significant advantages to making better provision of healthcare services.

Advantages

The system is able to aid in physician's diagnosis as well as serves an educational role. It shows that this type of applications, easy for dynamic modification and development, may make examinations shorter, make decisions more objective as well as reduce costs. Physician's education and experience is the most important diagnostic factor that influences on effective therapy. However, an essence of expert systems is a direct use of physicians knowledge.

17



Fig. 6. Schema of relation (rules) between symptoms of allergic diseases and allergic diseases, and between allergic diseases and diagnostic examinations of allergic diseases

Another advantage is a separation of knowledge and control that allows simple adding new medical facts and rules.

Disadvantages

The biggest weakness of the system is not having an option of integration with the internet, and the internet is nowadays the largest source of information for patients.

Conclusion

Presented system has a huge possibility of development in the future. Proposed direction of development is a change in the programming language and the shell system into CLIPS and adding of case-learning modules.

References

- 1. Shortliffe E. H., MYCIN: A rule-based computer program for advising physicians regarding antimicrobial therapy selection. Ph.D. dissertation, Stanford University, 1974.
- 2. Kulikowski C. A., Weiss S., Computer-based Models for Glaucoma, CBM-TM-MS1, Rutgers University, 1971.
- 3. Weiss S., Kulikowski C., Amarel S., Safir. A., A Model-Based Method for Computer-Aided Medical Decision-Making, Artificial Intelligence 11, 1978, 145-172.
- 4. Darmoni S.J., Massari P., Droy J.M., Mahé N., Blanc T., Moirot E., Leroy J., SETH: an expert system for the management on acute drug poisoning in adults. Comput. Methods Programs Biomed. 1993; 43: 171-176.



Fig. 7. Schema of relation (rules) between allergic diseases and treatment and prophylactic of allergic diseases

- Lilić N., Obradović I., Stanković R., VENTEX: An Expert System for Mine Ventilation Systems Analysis, Mining Technology, vol.79, no.915, November 1997, pp.295-302.
- 6. http://www.gideononline.com/
- Zieliński J.S., Inteligentne systemy w zarządzaniu, PWN, Warszawa 2000
- 8. http://aitech.pl/

20

- Lasoń W., Walecki P., Pyrczak W., Sarapata K., Roterman-Konieczna I, Medycyna Oparta na Dowodach (EBM) – nowe podejście optymalizacyjne wykorzystujące systemy komputerowe, W: Tadeusiewicz R., Ligęza A., Szymkat M. (red): Computer Methods and Systems, 2005: 455-462.
- Martyniak J. (red), Podstawy informatyki z elementami telemedycyny: ćwiczenia dla studentów medycyny, Kraków, Wydaw. Uniwersytetu Jagiellońskiego 2005.
- Walecki P., Lasoń W., Porębski G., Obtułowicz K., Abductive Logic Programming in allergy diagnosis, Bio-Algorithms and Med-Systems, Vol. 1, No. 1/2, 2005, pp. 57-60.
- Walecki P., Sarapata K., Lasoń W., Pyrczak W., Roterman-Konieczna I., Balwierz W., Telemedical Database of Hodgkin Disease, In: Transformation of Healthcare with Information Technologies, Studies in Health Technology and Informatics, IOS Press, Amsterdam-Berlin-Oxford-Tokyo-Washington, DC 2004, s.51-60.
- Walecki P., Lasoń W., Wiśniowski Z., Roterman-Konieczna I, Abductive Logic Programming in medical diagnosis, Journal of Artificial Intelligence, Vol. 1, No 2(24)/2004, pp.107-115.
- Lasoń W., Walecki P., Trąbka J., Telemedyczne bazy danych – teoria i praktyka, W: Społeczeństwo informacyjne wizja czy rzeczywistość? Red. L. Haber, AGH Uczelniane Wydawnictwa Naukowo-Techniczne, Kraków 2004, s.349-360.
- Sarapata K., Usarz M., Trąbka J., Walecki P., Pyrczak W., Roterman-Konieczna I, Security of net servers in biological approach, Journal of Artificial Intelligence, Vol. 1, No 1(23)/ 2004, pp.31-38.
- Lasoń W., Walecki P., Pyrczak W., Sarapata K., Trąbka J., Zintegrowane telemedyczne bazy danych pacjentów z chorobami nowotworowymi, Sztuczna inteligencja – Organizacje wirtualne, red. M. Kłopotek, J. Tchórzewski, Wydawnictwo Akademii Podlaskiej, Siedlce 2003, s.35-40.
- Pyrczak W., Lasoń W., Walecki P., Sarapata K., Trąbka J., Teleinformatyczne systemy w edukacji medycznej, W: Komputer w edukacji, red. J. Morbitzer, Wyd. Naukowe Akademii Pedagogicznej, Kraków 2003, s.226-230.
- Walecki P., Trąbka J., Zastosowanie wnioskowania abdukcyjnego przy projektowaniu modeli semiotyki lekarskiej. W: Komputer w edukacji, red. J. Morbitzer, Wyd. Naukowe Akademii Pedagogicznej, Kraków 2003, s.285-292.

- Lasoń W., Walecki P., Pyrczak W., Trąbka J., Rzeczywistość wirtualna w ujęciu bionicznym, W: Proceedings of the Second International Conference on Philosophy and Computer Science Processes of Evolution in Real and Virtual Systems, Kraków 2002, s.11-15.
- Sarapata K., Walecki P., Pyrczak W., Lasoń W., Roterman I., Szybiński Z., Medical Database of thyroid cancer, W: E-he@lth in Common Europe, red. K. Zieliński, M. Duplaga, IEEE Kraków 2003, s.347-354.
- Walecki P., Pyrczak W., Lasoń W., Sarapata K., Roterman-Konieczna I., Klekawka T., Korab-Chrzanowska E., Balwierz W., Skoordynowana baza danych u chorobie Hodgkina (HD) zorientowana na diagnostykę w systemie rozproszonym, Przegląd Lekarski, t. 60, supl. 6, s.78.
- Walecki P., Lasoń W., Sarapata K., Biologizacja w obrębie nauk technicznych, Krakowskie Studia Małopolskie, 2002, t. 6, s. 313-330.
- Porebski G, Obtulowicz K, Obtulowicz P, Walecki P, Roterman-Konieczna I, Czarnobilska E, Krzanowska K, Kwapińska A, Targosz K, Internet as a potential source of information about allergy for patients, The XIXth World Allergy Organisation Congress, June 26 - July 1, 2005, Munich, Germany.
- Lasoń W., Walecki P., Pyrczak W., Sarapata K., Roterman-Konieczna I., Nauczanie medycyny w oparciu o paradygmat Evidence Based Medicine 15. Ogolnopolskie Sympozjum Naukowe nt. "Komputer w edukacji", Krakow, 23-24 wrzesnia 2005.
- Walecki P., Pyrczak W., Lasoń W., Sarapata P., E-Learning i telemedycyna – problemy strukturyzacji wiedzy, 15. Ogolnopolskie Sympozjum Naukowe nt. "Komputer w edukacji", Krakow, 29-30 wrzesnia 2006, s. 243-248.
- 26. Porębski G., Walecki P., Obtułowicz K., Czarnobilska E., Krzanowska K., Kwapińska A., Targosz K., Częstość wykorzystywania różnych źródeł wiedzy o alergii przez chorych ze schorzeniami atopowymi, Allergy Asthma Immunology 7th Central and Eastern European Conference June 8-11, 2005 Łódź, Poland.
- Walecki P., Lasoń W., Porębski G., Obtułowicz K., Skórna postać nadwrażliwości na aspirynę - pilotażowy telemedyczny system informacyjno-diagnostyczny, Konferencja "Modelowanie Cybernetyczne Systemów Biologicznych" MSCB 2005.
- Walecki P., Porębski G, Obtułowicz K., Zastosowanie logiki niemonotonicznej w systemie doradczym dla diagnostyki różnicowej schorzeń alergicznych. Alergol. Immunol., 2006 T. 3, nr 1-2, s. 10-13.
- Walecki P., Lasoń W., Pyrczak W., Roterman-Konieczna I., Sarapata K., Szpitalny System Informacyjny (HIS) oparty na metodologii EBM, w: Pozyskiwanie wiedzy i zarządzanie wiedzą, M. Nycz, M. Owoc, red., Prace naukowe Akademii Ekonomicznej we Wrocławiu, 2006, s.179-188.



THE LIGAND DATABASE SERVER

SERWER BAZY DANYCH LIGANDÓW

ELŻBIETA BRONIATOWSKA¹, OLGA STĘPIEŃ², IRENA ROTERMAN^{1*}

¹ Department of Bioinformatics and Telemedicine – Collegium Medicum – Jagiellonian University, Kopernika 7E, 31-530 Cracow, POLAND ² Jagiellonian University, Faculty of Physics, Astronomy and Applied Computer Science, Reymonta 4, 30-059 Cracow, POLAND *Corresponding Author

Abstract: The new database (library) of all smal molecules – ligands creating complexes with protein molecules deposited in the Protein Data Bank is provided. The library contains complete chemical information for all ligands making possible calculations for the complex; e.g. the ground state electronic structure and partial charges, both calculated by quantum-chemical methods and non-bonding parameters. These properties of ligand molecule are important for computer simulation of protein-ligand interactions. The ligand library can be useful in designing the biochemical experiments such as identification of possible ligands for proteins or drug design in the sense of similarity search to natural ligand for potential competitive inhibitors. The server is freely available via website: www.bioinformatics.cm-uj.krakow.pl/ligands.

Key words: protein-ligand complex, partial charges, optimal (ground state, equilibrium) electronic structure of molecule, quantum mechanical calculations.

Introduction

Some proteins require the presence of non-peptide molecules for their biological activity (co-factors, cosubstrates). The ligand molecule is usually low-mass chemical compound, which complexed to protein molecule participates in biological processes of the host-molecule. Term "ligand" may be extended also to the substrate, which creates the specific complex with protein molecule before the catalitic act takes place. In our terminology ligand is a nonprotein component covalently or noncovalently bound with protein.

Computer simulation of protein-ligand interaction requires the complete set of parameters e.g. equilibrium electronic structure, partial charges, force constants, nonbonding parameters for ligand energy terms included in particular force field. The parameters for amino acids are available for any force field applied in different programs. The ligands' molecules raise problems because of their large variability and sometimes high specificity. Presently, Streszczenie: Artykuł przedstawia nową bazę danych (bibliotekę) zawierającą ligandy – molekuły tworzące naturalne kompleksy z białkami, zebranymi w bazie Protein Data Bank (PDB). Prezentowana biblioteka zawiera pełną charakterystykę liganda, niezbędną w symulacjach komputerowych kompleksów białko-ligand, np.: ładunki cząstkowe, parametry niewiążące van der Waalsa oraz parametry potencjału torsyjnego. Informacje zawarte w bazie mogą być również wykorzystane w opracowywaniu biochemicznych eksperymentów, takich jak identyfikacja możliwych ligandów białek czy projektowanie leków, przez podobieństwo do naturalnych ligandów, jako potencjalnych inhibitorów kompetycyjnych. Baza jest dostępna na stronie internetowej: www.bioinformatics.cm-uj.krakow.pl/ligands.

Słowa kluczowe: kompleks białko-ligand, ładunki cząstkowe, optymalna (w stanie podstawowym, równowagowa) struktura elektronowa molekuły, obliczenia kwantowo-mechaniczne.

such detailed chemical characteristic of ligand (substrate) molecules is not available in the consistent form.

The Protein Data Bank (PDB) [1] contains information about proteins' crystallographic structures including these in form of complexes with natural or artificial ligands. These ligands extracted from PDB were used to construct the library of molecules bound to proteins (natural and engineered). The library described in this paper contains chemical parameters such as ground state electronic structure together with its energy and partial charges calculated according to the quantum chemistry methods and nonbonding parameters for all ligands extracted from the PDB. Therefore, the library provides the structural differences of ligands between their optimal, ground state geometry (obtained based on the quantum chemical calculations) and geometry occurring in the ligand-protein complex (as they appear in the PDB).

The ligand database is planned to be useful for quick identification of natural ligands of proteins and will give access to the set of parameters characterizing ligands molecules. The library enables also fast search of molecular structures similar to potential drug. It is especially useful in molecular similarity-based drug design. The search of ligands similar to drug metabolites is important in respect to disease-related pathological processes. The ligand library, which can be applied for drug designing, is planed prospectively to have application to the individual drug design for particular patients making the pharmacological treatment more efficient and eliminating the negative reaction to standard therapeutic procedures.

Computational Details

Data

The set of ligands originates from the PDB. The semistructural formulas of ligand molecules extracted from the Ligand Depot website [2] are included. All ligand molecules are described by individual ID in the library including: protein ID from the PDB, ligand ID from the PDB and the number corresponding to the frequency of ligand's occurrence in the protein. This ID is responsible also for linking to the more detailed information of this molecule.

Methods

The ligand's chemical properties are calculated using quantum chemical methods. The optimal geometry of ground state and the atomic partial charges of isolated ligand molecules as well as their energy are calculated using Gaussian package [3] in the 6-31+G^{**} extended basis set with B3LYP exchange-correlation functional. The atomic

charges are calculated according to Mulliken population analysis [4] and from the fitting of the electrostatic potential [5]. The algorithm generating the Gaussian input file for two ligand geometries (optimal and complexed) was also prepared.

Database

The library is based on MySQL system and is operated by PHP. The service is progressively available on the www.bioinformatics.cm-uj.krakow.pl/ligand website.

Major features of the session

Two main queries (Fig.1a and Fig.1b) of particular ligand and of particular protein are implemented in the library browser including the more detailed extended searching. The searching parameters for ligand molecule are: ligand ID from the PDB, chemical name, chemical formula, smile, charge, protein ID from the PDB and some other categories called: metal ion, acid, lipid etc. The searching parameters for protein are: its ID from the PDB, its name, function and ligands' ID interacting with particular protein. More detailed information about chemical properties of the ligand is available via the ligand ID defined within the library. This page (see Fig. 1c) contains the atomic Cartesian coordinates for the ground state and for the state as it occurs in complex, atomic partial charges calculated by Mulliken population analysis and by fitting the electrostatic potential and all essential computing details. Several searching cat-







Figure 1: An example of the ligand database pages: (a) searching for the particular ligand (here a-D-glucose, AGC), (b) searching for the particular protein (PDB ID: 2A0Z), (c) chemical information of the ligand and (d) the protein.

egories of proteins like: organism, tissue, cell, gene and cellular location are also included.

The link to the PDB via ID of protein is made from different levels of the session. The detailed information about the particular ligand molecule is accessible via the ligand ID in the ligand library from the protein searching result pages. The corresponding information about protein (see Fig. 1d) is accessible via the grey arrow (only the protein ID from the PDB serves as the PDB link) from the ligand searching result pages.

Future plans

The library is still being developed and, therefore, not all necessary information has been included yet but it is planned to be completed progressively. The quantum chemical calculation of the electronic structure of ligand molecules and their properties in their ground state are given as the first data. The next step is to derive the non-bonding interaction parameters and to calculate the torsional potential of all ligands for their optimal geometry.

The addition of three-dimensional visualization of the ligand structure in the forms: optimal (ground state) and complexed with protein is planned. The structures of these both ligand forms enable the calculation of the factor describing their structural similarity (e.g. Root Mean Square Deviation - RMSD). It is of importance due to quite large difference between ligand's ground state versus the one in the complexed form (e.g. flavin-adenine dinucleotide molecule, [6]).

The enclosure of similar quantum-chemical properties (the ground state electronic structure, partial charges applying Hirshfeld analysis [7]) computed by the Amsterdam Density Functional (ADF) [8] is taken into consideration for comparison.

Conclusion

The simulation of ligand-protein interaction requires definition of the ligand parameters which are essential for the particular force fields. It is known that determining these ligand parameters is ambiguous [9, 10] and troublesome. The implementation of different methods to calculate these parameters may help to solve this problem. The main advantage of the library is that a user can choose the best parameters for the interesting ligand molecule among all included parameterizations.

The availability of parameters may also speed up calculation of complexes, what is particularly important for pharmacological purposes. The ligand similarity search may be also useful for biochemical experiment design describing potential non-expected reactions or complexation between protein under consideration and chemical compound not necessarily expected as interacting with protein limiting the list of reagents for biochemical experiments. Acknowledgements: Chemical quantum calculations were performed at Academic Computer Centre Cyfronet AGH and Gaussian and at the Research Centre in Juelich (project no. 2246) and ADF (Amsterdam Density Functional).

References

- 1. www.rcsb.org
- 2. http://ligand-depot.rutgers.edu
- 3. Gaussian 03, Revision C.02, Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria G. E., Robb M. A., Cheeseman J. R., Montgomery Jr. J. A., Vreven T., Kudin K. N., Burant J. C., Millam J. M., Iyengar S. S., Tomasi J., Barone V., Mennucci B., Cossi M., Scalmani G., Rega N., Petersson G. A., Nakatsuji H., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Klene M., Li X., Knox J. E., Hratchian H. P., Cross J. B., Bakken V., Adamo C., Jaramillo J., Gomperts R., Stratmann R. E., Yazyev O., Austin A. J., Cammi R., Pomelli C., Ochterski J. W., Ayala P. Y., Morokuma K., Voth G. A., Salvador P., Dannenberg J. J., Zakrzewski V. G., Dapprich S., Daniels A. D., Strain M. C., Farkas O., Malick D. K., Rabuck A. D., Raghavachari K., Foresman J. B., Ortiz J. V., Cui Q., Baboul A. G., Clifford S., Cioslowski J., Stefanov B. B., Liu G., Liashenko A., Piskorz P., Komaromi I., Martin R. L., Fox D. J., Keith T., Al-Laham M. A., Peng C. Y., Nanayakkara A., Challacombe M., Gill P. M. W., Johnson B., Chen W., Wong M. W., Gonzalez C., and Pople J. A., Gaussian, Inc., Wallingford CT, 2004.
- Mulliken R. S. (1955) Electronic Population Analysis on LCAO-MO Molecular Wave Functions.I, *J. Chem. Phys.* 23, 1833-1840; (1955) Electronic Population Analysis on LCAO-MO Molecular Wave Functions.II. Overlap Populations, Bond Orders, and Covalent Bond Energies. *J. Chem. Phys.* 23, 1841-1846.
- Breneman C. M., Wiberg K. B., (1990) Determining Atom-Centered Monopoles Conformational Analysis. *J. Comp. Chem.* 11, 361-373.
- Hahn J., Michel-Beyerle M. E., Rösch N. (1998) Conformation of the Flavin-Adenine Dinucleotide Cofactor FAD in DNA-Photolyase: A Molecular Dynamics Study. *J. Mol. Model.* 4, 73-82
- Hirshfeld F. L. (1977), Bonded-atom Fragments for Describing Molecular Charge Densities. *Theo. Chim. Acta* 44, 129-138.
- te Velde G., Bickelhaupt F. M., van Gisbergen S. J. A., Fonseca Guerra C., Baerends E. J., Snijders J. G., Ziegler T. (2001) Chemistry with ADF, *J. Comput. Chem.* 22, 931-967; Fonseca Guerra C., Snijders J. G., te Velde G., and Baerends E. J. (1998) Towards an order-*N* DFT method *Theor. Chem. Acc.* 99, 391-403; ADF2006.01, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http:// www.scm.com.
- Roterman I. K., Gibson K. D., Scheraga H. A., (1989) A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. I. Conformational predictions for the tandemly repeated peptide (Asn-Ala-Asn-Pro) 9, *J. Biomolec. Structure* & *Dynamics*, 7, 391-419.
- Roterman I. K., Lambert M. H., Gibson K. D., Scheraga H. A. (1989) A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. II. Ø-φ maps for N-acetyl alanine N'-methyl amide: comparisons, contrasts and simple experimental tests, *J. Biomolec. Structure & Dynamics*, 7, 421-453.



HOW SPECIALLY ADAPTED CHROMATOGRAPHIC BED MATERIAL MIGHT BE USED FOR THE CONTROLLED FORMATION OF IMMUNE COMPLEXES AND STUDIES OF CONGO RED BINDING TO BIVALENT ANTIBODIES

JAK SPECJALNIE SKONSTRUOWANY CHOMATOGRAFICZNY MATERIAŁ MOŻE BYĆ WYKORZYSTANY DLA KONTROLOWANEGO TWORZENIA KOMPLEKSÓW IMMUNOLOGICZNYCH I ANALIZY BIWALENTNYCH ANTYGENÓW WIĄŻĄCYCH CZERWIEŃ KONGO

Anna Drozd, Leszek Konieczny, Barbara Piekarska, Janina Rybarska, Barbara Stopa, Grzegorz Zemanek

Chair of Medical Biochemistry, Collegium Medicum – Jagiellonian University, Kopernika 7, 31-034 Krakow, Poland

Abstract: Antibodies are the proteins that are designed biologically to recognize the foreign material in the organism and to initiate the process of its destruction and its removal. The functional connection of these two activities linked structurally to separate parts of antibody molecule is although critical, still poorly understood. The reason lies in the complexity of the phenomenon caused mainly by the natural heterogeneity of active immune complexes and their insolubility making the use of standard analytical techniques ineffective. A special analytical technique was hence elaborated to bypass the problem. It is based on using Congo red self-assembling dye as an indicator of binding bivalent antibodies to antigenic determinants fixed to the chromatographic bed selected for this aim. This technique allows for the controlled formation of immune complexes disclosing the dye binding effects as well as their easy removal for analytical purposes.

Key words: immunological signal, self-assembling ligand, immunological complex

Streszczenie: Antygeny są białkami przeznaczonymi przez ewolucie do rozpoznawania obcego dla danego organizmu materiału biologicznego w celu inicjacji procesów destrukcji obcego ciała i jego usuniecia. Funkcjonalne połaczenie dwóch aktywności (rozpoznania przeciwciała oraz inicjacji reakcji degradacji) jest krytyczne dla żywego organizmu. Proces ten nie jest jednak rozpoznany. Przyczyna tego jest dość duża złożoność tego procesu, czego przyczyna jest dość duża różnorodność kompleksów immunologicznych i ich nierozpuszczalność w warunkach eksperymentów. W tym celu opracowana została technika oparta na wiązaniu ciekłokrystalicznego liganda czerwieni Kongo, który okazał się być markerem dla wiązania antygenu angażującego oba ramiona cząsteczki IgG wiązanego do specjalnego podłoża chromatograficznego. Ta technika pozwala na kontrolowane tworzenie kompleksów immunologicznych ujawniając tworzenie kompleksów w sposób odwracalny, co ma istotne znaczenie dla celów analitycznych.

Słowa kluczowe: kompleksy immunologiczne, ligand ciekłokrystaliczny, sygnał immunologiczny

The idea of intramolecular signaling in immunoglobulins was born together with the understanding that the antigen binding site and the site responsible for starting the complement fixation are located in separate, distant parts of the antibody molecule [1,2]. Since the very beginning, the intra-molecular transduction of the signal was thought of to be highly attractive for many purposes. The attraction of such signaling transduced by specific structural alterations resulted in particular from the hope of possible finding the ligands which may fit to the new structural motifs and interfere with the resulted effects. The ligands, which could inhibit or activate the biological function of antibodies, may

have expectedly high applicability in medical practice playing the role of drugs. Unfortunately the identification of signal dependent changes seemed to be very difficult. The reason is three-fold: 1 – the detailed structure of complete, bivalent antibodies engaged in immune complexes and capable to induce effector activities remains unknown making any detailed considerations on this problem speculative; 2 – the polyclonal immune complexes of high effector activity are basically heterogeneous due to the diversity of antibody molecules participating in the complex. The diversity arises also from bivalent antibodies binding to randomly distributed antigenic determinants, not optimal usually to the natural arrangement of the Fab arms. This is why the different constraints and consequently structural changes are generated in immunoglobulin molecules. As they are heterogeneous and basically insoluble, the immune complexes escape from the most standard analytical techniques; 3 – the structural changes generated directly by the location and binding of antigenic determinants in the binding site and those induced by constraints in binding of bivalent antibodies over-lap making their differentiation highly complicated.

These difficulties became the reason of the failure of many efforts undertaken to solve the problem and finally raised the doubts concerning biological reality of the intramolecular way of signaling. It induced elaboration of alternative hypothesis of signal transduction which assumed the simple gathering of antibodies as the sufficient starting motif for complement fixation [3]. This hypothesis rejected the idea of the signal dependent intramolecular structural changes and consequently the hope for possible drug design which could modify the effects resulted from the antigen antibody complexation.

Also the crystallographic studies of Fab-antigen complexes currently available for analysis which had started with a great hope of finding the signal specific changes, failed to identify the expected specific alterations of common structural features in spite of complexation different antigens [4-8]. All that weakened the research activity in this field of studies.

The surprising comeback of the hypothesis concerning intramolecular signaling was caused however by finding that some self-assembling dyes, for example Congo red (Fig.1), interact selectively as specific ligands with complete, bivalent antibodies engaged in complexes but not with their fragments (Fab, or Fab) [9,10]. These non standard protein ligands penetrate the protein body outside the biological binding site interacting with polypeptides in its still preserved self-assembled form [11,12]. The specific local destabilization of immunoglobulin molecules upon constraints generated by binding to the antigen makes antibodies accessible for penetration and complexation of such ligands [11,12]. The finding that only complete bivalent antibodies undergo complexation indicates that structural changes induced directly by antigen binding and those generated by constraints derived from bivalent antibody binding differ. While the first are specific to the antigen structure and its direct fixation in the binding site of the antibody, the latter are related rather to the adjustment of bivalent antibody molecule to antigenic determinants and are initiated by the rotation of V domains. By affecting the biological function of antibodies in immune

Fig. 1. The dye Congo red – A - single molecule, B – its 3-D model, C – micellar – self-assembled form.

complexes, the supramolecular ligands support the idea of the occurrence of intra-molecular signaling [9, 10, 13-15]. Above all however, these dyes become indicators revealing the specific structural alterations in antibodies generated by the constraints which associate the fitting of the whole complete bivalent antibody to antigenic determinants on the antigen surface. The compounds which can recognize these structural changes by specific interaction may represent prospective drugs. Unfortunately, the understanding of the mechanism of this phenomenon is still incomplete and needs further study which must face all those technical barriers mentioned above. The most reliable analytical system corresponding to the expectations was based on the formation of immune complexes in conditions allowing easy observations of all standard biological effects connected with the immune complexation and as well the effects of binding the ligands.

A special technique was hence elaborated to study the complexation of supramolecular dyes (represented by Congo red) to antibodies in immune complexes. The technique was based on using the insoluble bed of low absorption property for Congo red, cross-linked to antigen determinants by easily split linkers. It allows for the removal of the whole immune complex together with the bound Congo red. CM-Sephadex A50 was chosen for this bed. TNP (trinitro-phenol) was attached as the antigen to react in the used model system with suitable polyclonal rabbit anti-TNP antibodies. The adsorption property of this antigenic bed to antibodies and Congo red, in the conditions studied appeared to be unsatisfactorily low. Such weak absorption property for Congo red, ensuring non-specific binding as

negligible, is necessary for the reliable analysis. The TNP used as an antigen was fixed to the bed by the cystamine spacer (Fig.2.). It allows for the easy removal of antigens as well as antibodies engaged in the immune complex by gentle reduction or sulphitholysis. Congo red used for complexation may also be removed together with antibodies and then the ratio of dye and antibodies may be easily measured by spectrophotometric techniques.

In general, such a bed seems suitable for the search and testing of proper supramolecular compounds which may be used as drugs.

Using the bed substituted by antibodies, the complement fixation phenomenon may also be studied. This could be done by measuring the percent of the remained complement activity after its passing through the bed containing immune complexes versus the initial activity. The loss of complement activity is equal in this system to the number of antibodies which are capable of binding C₁ pooling it out from the complement proteins.

Although the constructed antigenic bed may be used to study the effect of many different reagents interfering with the biological function of antibodies, its original design and present use concern supramolecular systems in particular. Its usage helps to overcome the essential problems of studies. The complexation of supramolecular ligands seems to be, at the moment, the only test disclosing the structural changes in antibodies generated specifically in bivalent antibody molecules and which probably represents the signaling phenomenon. The understanding of the nature of supramolecular ligand complexation introduces the practical aspects to these studies.



Fig. 2. The schematic model representing the binding of bivalent antibody to antigenic bed.

References

- G.M. Edelman (1970). The covalent structure of human ãG immunoglobulin. XI. Functional implications. Biochemistry 9, 3197-3205.
- N.J. Calvanico, T.B. Tomasi Jr (1979). Effector sites on antibodies. In: Atassi MZ Ed, Immunochemistry of proteins. Vol. 3. New York and London, Plenum Press, 1-85.
- H. Metzger (1990). General aspects of antibody structure and function. In: Fc receptors and the action of antibodies. Ed. H. Metzger. Am. Soc. Microbiol, 7-11.
- T.N. Bhat, G.A. Bentley, T.O. Fischmann, G. Boulot, R.J. Poljak (1990). Small rearrangements in structures of Fv and Fab fragments of antibody D1,3 on antigen binding. Nature 347, 483-485.
- U. Schulze-Gahmen, J.M. Rini, I.A. Wilson (1993). The detailed analysis of the free and bound conformations of an antibody. X-ray structures of Fab 17/9 and three different Fab peptide complexes. J. Mol. Biol. 234, 1098-1118.
- W.R. Tulip, J.N. Varghese, R.G. Webster, W.G. Laver, P.M. Colman (1992). Crystal structures of two mutant neuraminidase-antibody complexes with amino acid substitution in the interface. J. Mol. Biol. 227, 149-159.
- L.W. Guddat, L. Shan, J.M. Anchin, D.S. Linthicum, A.B. Edmundson (1994). Local and transmitted conformational changes on complexation of an anti-sweetener Fab. J. Mol. Biol. 236, 247-274.
- I.A. Wilson, R.L. Stanfield (1994). Antibody-antigen interactions: new structures and new conformational changes. Curr. Op. Struct. Biol. 4, 4857-4867.

- J. Kaszuba, L. Konieczny, B. Piekarska, I. Roterman, J. Rybarska (1993). Bis-azo dyes intereference with effector activation of antibodies. J. Physiol. Pharmacol. 44, 233-242.
- M. Król, I. Roterman, A. Drozd, L. Konieczny, B. Piekarska, J. Rybarska, P. Spólnik, B. Stopa (2006). The increased flexibility of CDR Loos generated In antibodies by Congo red complexation favors antygen binding. J. Biomol. Struct Dynam. 23, 407-415.
- B. Piekarska, A. Drozd, L. Konieczny, M. Król, W. Jurkowski, I. Roterman, P. Spólnik, B. Stopa, J. Rybarska (2006). The indirect generation of long-distance structural changes in antibodies upon their binding to antygen. Chem. Biol. Drug Des 68, 276-283.
- M. Król, I. Roterman, B. Piekarska, L. Konieczny, J. Rybarska, B. Stopa (2003). Local and long-range structural effects caused by the removal of the N-terminal polypeptide fragment from immunoglobulin L chain. Biopolymers 69, 189-200.
- J. Rybarska, L. Konieczny, I. Roterman, B. Piekarska (1991). The effect of azo dyes on the formation of immune complexes. Archiv. Immunol. Ther. Exp. 39, 317-327.
- B. Piekarska, L. Konieczny, J. Rybarska, B. Stopa, G. Zemanek, E. Szneler, M. Król, M. Nowak, I. Roterman (2001). The heat induced formation of a specific binding site for self-assembled Congo red in V domain of immunoglobulin chain I. Biopolymers 59, 446-456.
- B. Piekarska, L. Konieczny, J. Rybarska, B. Stopa, P. Spólnik, I. Roterman, M. Król (2004). Intramolecular signaling in immunoglobulins – New evidence emerging from the use of supramolecular protein ligands. J. Physiol. Pharmacol. 55, 487-501.



DOCKER ANALYZER: PROGRAM FOR COMPARISON OF DOCKING RESULTS

DOCKER ANALYSER: PROGRAM DLA ANALIZY PORÓWNAWCZEJ EFEKTÓW DOKOWANIA LIGANDA DO BIAŁKA

ARTUR NIELEPIEC^{1,2}, WIKTOR JURKOWSKI²

¹Faculty of Physics, Astronomy and Applied Computer Science – Jagiellonian University, Reymonta 4, 30-060 Krakow, Poland ²Department of Bioinformatics and Telemedicine - Collegium Medicum – Jagiellonian University, Kopernika 17, 31-501 Krakow, Poland

Abstract: Vast amount of docking algorithms and scoring functions available encourage development of methods aimed to adjust functionality of docking tools. A certain level of scoring functions precision is needed to recognize complexes of specific intermolecular interactions. Having this kind of model more detailed blind prediction of molecular complexes would be possible. To enhance specificity of docking results tools unifying application of many scoring functions should be developed.

A program for assessment of molecular docking: Docker Analyzer is presented as preliminary approach. Complex calculation of over ten scoring functions and its simultaneous use (multiple consensus scoring) is possible to improve default selection of best hits after standard molecular docking procedure.

Key words: docking procedue, protein-ligand complex, drug design

Introduction

Molecular docking methods aim at searching for functionally optimal protein – ligand spatial relation and has many successful applications^{1,2,3,4}. The optimal protein – ligand complex should reflect properties of natural systems verified by experimental methods (e.g. measured Inhibition Constant). It is considered that structures found in crystal form to some extend represents conditions characteristic for its natural activity. Protein structures obtained by X-ray or NMR techniques reveal protein – ligand contacts crucial for its function. These contacts give insight not only into arrangement of the protein binding site but also allow to distinguish specific parts in the ligand structure and define pharmacophore – set of structural elements crucial for given activity. This allows efficient analysis of many functionally related complexes to search for leading structure modificaStreszczenie: Dostępność wielu programów dokujących oraz różnorodność używanych w nich funkcji służących do oceny poprawności kompleksów białko-ligand skłania do ich oceny i porównań. Istotna wydaje się być możliwość dobrania funkcji oceniającej pod kątem specyfiki badanego układu, a więc wybór funkcji, która najpoprawniej w danych przypadku opisywałaby naturę oddziaływań miedzycząsteczkowych. Takie podejście powinno pomóc w poprawnym dokowaniu w przypadku kiedy brak jest wstępnej informacji strukturalnej (np. przy nieznanym miejscu wiążącym). Praca ta przedstawia program do oceny efektywności procedur dokowania będący wstępną realizacją takiego podejścia. Program Docker Analyzer umożliwia jednoczesne obliczenie kilkunastu popularnych funkcji oceniających i ich wzajemne porównanie, wyznaczenie na ich podstawie wskaźnika konsensusowego oraz prostą wizualizację kompleksów.

Słowa kluczowe: procedura dokowania, kompleksy białek z ligandami, projektowanie leków

tions (resulting with enhanced pharmacological properties) and ligand libraries construction⁵.

Molecular docking calculations are used in the case when structure of the complex is not known and aim on reproducing best fit of ligand and macromolecule: intramolecular contacts between aminoacid residues and ligand. This can be described on detailed atomic scale as a specific unique exemplification of molecular recognition phenomena but also on more rigid (and therefore more useful way) as mapping of specific pharmacophore on related binding site pattern. Both attempts require some basic knowledge of potential binding site – and function related protein cavity. In other case binding-site needs to be predefined. For some proteins this search may be homology based: proteins with similar AA sequences have in most cases similar structures and therefore function or structure based. Beyond homology search for functional sites includes molecular libraries searches, geometrical analysis of protein surface, solvent accessibility, and analysis of hydrophobicity distribution^{6,7,8}.

Proper description of ligand structure is also of high importance. On the basis of defined intermolecular degrees of freedom, random conformations are created producing data for further assessment. Thus the initial ligand preparation (ligand ground state, atomic parameters derivation) is a basic step in any docking algorithm.

FlexX⁹, Autodock¹⁰, Dock¹¹ and Schroedinger¹² may be mentioned as representatives among many presently accessible molecular docking tools. They represent different attempts in search of conformational space, representation of molecules and interactions in system. However, basically, procedure of docking can be divided into following steps:

- Creation of large sets of complexes for efficient search of conformational space (stochastic algorithms: Monte Carlo, genetics algorithms)
- Selections of complexes according to some defined rules (scoring functions)
- Final selection and ranking of structures (scoring functions)

In most cases final results require consecutive repetitions of first two steps followed by only once performed final scoring.

Molecular Docking may be used either for analysis of small number of complexes or for massive Virtual Screening procedures. In both cases automation o docking procedure is required to maintain high efficiency of calculations. Since present processor speed enables in-depth search of conformational space, the next milestone to be reached is fast and reliable assessment of produced protein - ligand complexes. Large number of currently accessible scoring functions needs to be compared and assessed. Currently they are some attempts to compare scoring functions basing on selected complexes produced by chosen docking tools¹³. These analyses to some extend shade a light on given scoring functions and reliability of its use in particular cases. What should be a matter of further discussions is the development of universal tool focused on scoring problem: allowing multiple scoring and choice of specific molecular function/properties related assessment method. One possible solution addressing this problem could be implementation of wide range of scoring functions representing different assessment methodologies (statistics based, experimentally parameterized, force field based). These functions when mutually used and tested on certain amount of specific molecular complexes with the help of classification methodology could be then used to create consensus functions focusing on relations between molecular function and recognition in molecular docking procedure.

This paper presents the Docker Analyzer: a multiplatform application for scoring and assessing results of protein-ligand docking simulations. Even though program is in early phase of development, it can be helpful in selection energetically stable and functionally favored complexes.

Implementation

Docker Analyzer was written in Python^{14,15,16}, an open source language with usage of GTK^{17,18} and OpenGL^{19,20,21} libraries. Considered as language which offers efficient I/O and files handling system, Python was chosen to process various formats of inputs necessary for determination of scoring functions. Python delivers object such as lists and tulps and also a dictionary with large amount of useful methods used in implementation of main modules. Furthermore Python excellently collaborates with third-party external libraries (e.g. GTK, OpenGL) what significantly improves application functionality.

Scoring functions were implemented basing on literature describing given software or methods used. In fact, there are no direct instructions of implementations for functions mentioned. Quite rarely, formulas describing scoring function in question are written explicitly. In most cases authors just more or less clearly introduce to the philosophy of given scoring function without taking necessary care for exhaustive explanation of used algorithms.

First version of presented software calculates scoring functions widely described in the literature: LUDI^{22,23}, ChemScore²⁴, PLP²⁵, LigScore^{26,27}, F-Score²⁸, D-Score²⁹, G-Score³⁰, Autodock³¹, PMF³², DrugScore³³ and HP-Score³⁴ for once prepared set of docked data. All functions mentioned are parameterized basing on the specific, quite narrow subset of protein - ligand complexes chosen from Protein Data Bank³⁵. Implementation of given function was treated as successful either if the score results have the same (or proportional) values or if the rank obtained has the same order. Information regarding formulas describing each scoring function in question (in the form used in this paper) is given on the web-site of the project³⁶. Some scaling factors used for standardization of obtained values are also in use. Main objective of each given scale is to reproduce correct ranking of hits yielded in docking procedure.

Program functions

Program is controlled by simple graphics interface (Figure 1.). Besides I/O options it enables direct choice of scoring function, tools (molecule visualization, parameters configuration), and calculation launching. First objective to point out is I/O data processing. Accepted input formats are: *PDB* and *MOL2* as the most popular structure formats used by docking software. This step requires good flexibility because of large amount of different data stored in files produced by docking tools (Autodock, DOCK, FlexX and LUDI) by which scoring functions were calculated originally.

The advanced configuration module enables definition of dictionaries with experimental data, weighting factors etc. and it does not require any user intervention with thirdparty scripts and parsers to prepare data for calculations. Scoring functions in question utilizes a large number of different parameters (e.g. hydrophobicity, H-bonds involved atom types, charges, rotamers) which are also free to configure.



Fig. 1. Main window of Docker Analyzer after software initialization. Graphic interface enables access all program functionalities by main upper menu and side-panel with buttons launching 11 given scoring functions.

Important Docker Analyzer option is the possibility of re-ranking analyzed sets of complexes basing on chosen set of scoring functions and accompanying weighting factors (consensus scoring). In present version calculation of normalized or non-normalized weighted mean is available.

Results may be saved as text files for further analysis (e.g. using statistics tools). Application maintains native project format keeping all inputs and originally used configuration files. This allows future work on changed configuration or set of scoring functions and further comparisons.

Both macromolecule and ligands may be displayed in visualization module: Docker 3D. This module enables basic display in window integrated with main module and convenient switch between consecutive docked conformations. Because of very simple display style this module does not replace full-featured visualization programs but is gives opportunity of fast preview of results and structure – score identification.

Additional features includes: parameters restore system (e.g. after accidental deletion of configuration files), errors and warnings notification (input data errors, lack of dictionary data, improper use of the software), system logs etc.

Example of use

Docker Analyzer compilations on MS Windows and UNIX/Linux platforms were tested on chosen Autodock 3.0 and FlexX (Sybyl based version) results. In presented attempt results obtained by Autodock 3.0 and FlexX software were re-used and re-ranked for testing purposes within Docker Analyzer tool. Final scores calculated using this implementation was compared with those brought by original programs. One should mark, that scoring function calculated within Docker Analyzer and by original programs may differ for the reason already mentioned.

As a test case a sildelafil (Viagra) molecule was docked to the Phosphodiesterase 5 enzyme (PDB: 1UDT) with usage of Autodock 3.0.5 program.



Fig. 2. Two conformation of docking protein-ligand structure inside docking pocked. Colors: red – 1UDT, green – main carbon a chain of 1UTD, blue – Viagra crystal form, yellow – Viagra run3_19 (the best place in consensus scoring ranking). Visualization created by using Docker 3D.



Fig. 3. Two conformation of docking protein-ligand structure outside docking pocked. Colors: red – 1UDT, green – main carbon a chain of 1UTD, blue – Viagra crystal form, yellow – Viagra run3_30 (the worst place in consensus scoring ranking). Visualization created by using Docker 3D.



Fig. 4. Two conformation of docking protein-ligand structure inside docking pocked. Colors: red – 1UDT, green – main carbon a chain of 1UTD, blue – Viagra crystal form, yellow – Viagra run3_22 (hit with the best RMSD). Visualization created by using Docker 3D.

Table 1. Ranking of 1UDT-viagra docking results, based on first (rigid) docking run. 50 best hits shown. First column gives names of conformations. Next 11 columns contain ranking position of given scoring function. Cons1 is average of all scoring functions. Table was sorted in relation to the column named Cons2 containing consensus scoring function created on base on 5 scoring function implemented in Docker Analyzer: Autodock, PLP, PMF, D-Score and HPScore. The last column contains RMSD factor calculated without hydrogen atoms for crystal form and docking hits.

Conformation	AutoDock	LigScore	PLP	PMF	LUDI	F-Score	G-Score	D-Score	ChemScore	DrugScore	HPScore	Cons1	Cons2	RMSD
1UDT-r1_3	1	45	18	10	39	43	49	2	44	49	9	5,1535	1,5306	9,303
1UDT-r1_4	12	51	9	8	33	23	11	10	24	43	1	4,4358	1,6913	8,454
Sildenafil	3	41	1	15	48	51	51	5	48	40	14	5,8508	1,9144	0,000
1UDT-r1_14	8	21	28	1	35	4	4	9	41	45	6	3,8446	1,9557	8,456
1UDT-r1_17	4	37	31	11	51	32	43	1	51	42	19	5,6123	2,4537	8,112
1UDT-r1_1	9	49	25	6	47	31	42	8	49	46	12	5,6088	2,4781	9,039
1UDT-r1_19	6	10	38	7	29	9	16	6	30	41	3	4,1722	2,5354	9,763
1UDT-r1_24	2	36	27	12	50	48	50	3	50	47	25	6,0147	2,7246	9,090
1UDT-r1_8	7	20	29	5	37	22	28	7	42	48	20	4,9098	2,7922	8,659
1UDT-r1_9	10	4	46	2	31	6	10	11	47	51	8	4,2357	2,8542	9,964
1UDT-r1_18	5	15	39	9	20	13	20	4	18	38	15	4,2793	3,0694	9,862
1UDT-r1_5	17	28	34	14	36	14	38	18	33	44	2	5,0421	3,2843	4,371
1UDT-r1_16	19	46	16	13	42	19	35	21	29	50	7	5,4380	3,2874	5,572
1UDT-r1_27	15	35	5	17	44	39	17	14	19	12	5	4,4537	3,3216	10,507
1UDT-r1_47	14	44	6	20	43	42	41	12	31	29	26	5,3740	4,1437	10,326
1UDT-r1_15	13	3	51	3	41	3	3	16	43	37	17	4,1689	4,2217	9,388
1UDT-r1_23	11	14	50	4	40	25	22	13	46	39	29	5,5091	4,3447	9,161
1UDT-r1 36	22	50	2	32	17	29	34	20	22	10	13	5,0371	4,3556	10,888
1UDT-r1 49	16	48	4	36	46	27	46	15	34	21	28	5,5227	4,4679	11,842
1UDT-r1 7	23	38	13	29	13	12	13	25	35	22	4	4,6221	4,4739	11,295
1UDT-r1 50	18	43	3	44	26	46	39	17	26	31	32	5.6200	5.1133	12.031
1UDT-r1 48	25	24	33	33	19	34	48	22	7	33	11	5.1946	5.2182	18,710
1UDT-r1 6	32	25	26	23	8	11	5	39	9	19	10	4.2906	5.3042	11.497
1UDT-r1 26	21	33	10	22	30	33	47	23	32	36	36	5.7475	5.3994	11.554
1UDT-r1 43	35	42	8	34	15	20	44	36	21	34	21	5.4956	5.5735	10.203
1UDT-r1 33	20	39	11	38	49	45	33	19	37	23	37	6.0028	5.5806	12,997
1UDT-r1 40	31	23	30	42	22	28	19	32	27	32	16	5,3844	5,8960	14,159
1UDT-r1 12	47	47	7	31	25	26	40	47	12	27	23	5,7932	6.0045	9,967
1UDT-r1_37	26	12	21	40	6	8	23	27	8	4	33	4,5878	6.0132	16,276
1UDT-r1 21	42	30	15	43	4	16	30	44	6	15	27	4.7356	6.0780	15,158
1UDT-r1 10	28	8	14	37	2	15	6	30	1	13	42	3,8880	6,1222	14.375
1UDT-r1 13	49	40	12	24	38	47	36	50	28	30	24	6.2189	6.2537	9,793
1UDT-r1 42	45	31	24	49	21	44	9	45	13	20	18	5.5075	6.2981	14.217
1UDT-r1 22	34	17	17	39	34	35	25	33	10	2	45	5,5056	6.4003	16,708
1UDT-r1 41	27	18	35	45	24	38	12	26	25	6	35	5.5118	6.4940	15.618
1UDT-r1 39	36	29	23	41	32	40	27	34	36	14	39	5.9966	6.5171	14.299
1UDT-r1 30	44	22	36	19	18	37	26	42	16	18	34	5.6502	6.5683	17.997
1UDT-r1 34	39	5	37	21	12	17	8	43	14	24	30	4,7949	6.6405	18,430
1UDT-r1 32	33	27	32	46	23	41	29	29	17	11	44	5.8050	6.6628	14.964
1UDT-r1 35	30	9	44	28	14	18	14	28	15	28	41	5.2058	6.7315	17.451
1UDT-r1 2	41	16	40	16	28	21	24	41	38	26	40	5.8158	6.7381	18.612
1UDT-r1 46	37	34	20	48	5	30	45	38	5	17	46	5,5789	6,7397	15,247
1UDT-r1 11	46	1	42	25	10	5	7	46	11	16	22	4,4713	6,7844	18.082
1UDT-r1 45	38	32	22	50	.0	36	37	37	3		48	5,5982	6,8815	16.378
1UDT-r1 44	29	13	47	35	11	7	18	31	39	25	43	5,4815	6.9619	19,909
1UDT-r1 28	43	19	19	47	27	49	32	40	20		50	6.0151	7.0142	16,908
1UDT-r1 25	40	26	43	26	45	50	31	35	40	3	47	6.4598	7,1400	16.617
1UDT-r1_38	24	6	49	30	1	2	2	24	2	7	49	3,6137	7,2241	19,710
1UDT-r1 20	48	2	48	18	7	1	1	48	45	35	31	4,5936	7,3329	17,563
1UDT-r1 31	50	11	41	27	16	24	21	49	23	5	51	6.0177	8.0260	16,489
1UDT-r1 29	51	7	45	51	3	10	15	51	4	1	38	5,3625	8,9682	16,603

32

Table 2. Ranking of 1UDT-viagra docking results, based on second (fine) docking run. 50 best hits shown. First column gives names of conformations. Next 11 columns contain ranking position of given scoring function. "Cons1" is average of all scoring functions. Table was sorted in relation to the last column named "Cons2" containing consensus scoring function created on base on 5 scoring function implemented in Docker Analyzer 1.1: Autodock, PLP, PMF, D-Score and HPScore. The last column contains RMSD factor calculated without hydrogen atoms for crystal form and docking hits.

Conformation	AutoDock	LigScore	PLP	PMF	LUDI	F-Score	G-Score	D-Score	ChemScore	DrugScore	HPScore	Cons1	Cons2	RMSD
1UDT-r3_19	17	44	6	25	32	13	25	12	37	27	1	4,846	2,062	7,931
Sildenafil	3	39	1	26	49	51	51	8	50	29	17	6,041	2,140	0,000
1UDT-r3_13	4	48	11	21	41	43	50	4	40	48	9	5,842	2,274	8,871
1UDT-r3_22	21	38	4	24	24	41	16	21	20	45	2	5,166	2,374	3,705
1UDT-r3_14	2	43	18	18	47	42	49	2	41	50	12	5,862	2,397	9,069
1UDT-r3_11	6	46	14	11	38	30	41	6	38	38	16	5,593	2,572	8,801
1UDT-r3_16	7	49	12	19	36	45	46	5	33	41	19	5,859	2,696	8,830
1UDT-r3_8	9	37	21	13	23	26	37	9	39	44	13	5,481	2,761	8,894
1UDT-r3_15	5	40	20	12	45	31	44	3	43	51	29	5,848	2,799	8,952
1UDT-r3_12	12	47	16	10	37	35	35	14	31	35	14	5,685	2,844	8,819
1UDT-r3_18	13	45	15	15	43	32	43	11	46	33	15	5,843	2,871	9,115
1UDT-r3_10	8	41	23	16	28	24	39	7	34	46	18	5,598	2,878	8,905
1UDT-r3 28	19	26	10	22	11	16	27	20	8	32	11	4,543	2,902	7,946
1UDT-r3_9	10	42	17	14	50	38	48	10	51	47	27	6,077	2,914	8,978
1UDT-r3 21	25	35	9	23	44	18	13	23	27	31	6	5.270	2.961	6.622
1UDT-r3 20	1	24	47	5	26	39	47	1	48	36	40	5.706	3.038	9.926
1UDT-r3 1	22	23	37	8	12	4	8	22	14	40	7	4.681	3.230	9.006
1UDT-r3 5	20	34	25	9	48	50	42	16	42	39	20	6.163	3.286	9,100
1UDT-r3 2	14	33	30	7	25	40	36	19	36	37	28	5.829	3.389	9.233
1UDT-r3 4	23	21	38	3	31	11	3	24	26	30	10	4,860	3,393	8,999
1UDT-r3 7	18	27	45	1	30	28	28	17	45	34	26	5.681	3,572	9.815
1UDT-r3_23	11	22	40	17	10	15	33	13	16	42	32	5 085	3 655	8,910
1UDT-r3 6	15	14	44	4	17	23	30	15	44	49	30	5,483	3,728	9.551
1UDT-r3_3	27	5	41	2	40	3	1	27	47	26	8	4 676	3,915	8,970
1UDT-r3 17	16	15	48	6	33	14	22	18	28	28	36	5.354	4,162	9,994
1UDT-r3 24	30	50	3	28	42	47	9	30	30	15	4	5 561	4 177	9.625
1UDT-r3_32	24	18	50	20	46	44	40	25	49	43	33	6,536	4 877	9.675
1UDT-r3 47	29	51	2	36	51	48	29	28	21	11	23	6,086	5 069	12 171
1UDT-r3_38	26	20	26	32	13	7	26	26	18	24	37	5 227	5 534	18 569
1UDT-r3_50	34	36	7	39	22	10	23	34	15	16	22	5 595	5 758	10,694
1UDT-r3_31	36	16	19	46	5	17	15	36	23	8	5	5 125	5,897	16 105
1UDT-r3_37	33	28	8	37	21	36	18	31	24	19	39	5 882	6 014	12,311
1UDT-r3 44	31	29	13	40	.9	34	45	33	19	14	31	5,969	6 195	21 195
1UDT-r3 27	32	25	29	31	29	25	24	32	35	20	42	6,100	6,350	19.094
1UDT-r3_33	28	31	42	33	7	5	14	29	17	22	38	5,571	6,376	19,512
1UDT-r3_45	35	32	28	27	39	49	34	35	22	25	43	6,650	6 4 9 6	18,961
1UDT-r3_34	37	30	5	38	35	29	31	37	32	13	45	6,303	6,508	12 721
1UDT-r3 29	38	19	24	29	34	33	21	39	25	23	34	6 165	6,638	16 468
1UDT-r3 25	40	1	51	30	1	1	2	40	10	17	3	3 888	6,650	19,959
1UDT-r3_36	41	10	31	34	18	37	19	41	9	21	21	5,855	6,855	19 446
1UDT-r3_40	43	3	33	43	19	6	4	42	11	3	24	5 193	7 284	14 253
1UDT-r3_41	44	9	22	41	6	22	7	44	4	1	25	5,226	7,365	14,350
1UDT-r3_43	47	7	36	47	2	2	6	48	2	4	35	4 939	8 174	14,652
1UDT_r3_35	30	12	30	50	15	8	10	38	12	10	50	6 022	8 195	14,002
111DT-r3 42	48	12	27	45	20	46	12	<u>4</u> 0	5	5	 	6 170	8 250	14 503
111DT_r2 20	50	10	30	11	1	10	20	16	2	<u>م</u>	11	5 509	8 2/2	1/ 200
111DT_r2 06	15	+ 6	2/	44 /19	+ Ω	20	5	40	6	7	44 17	5,000	8 / 25	1/ 120
111DT_r3 /19	46	11	35	- 1 0 51	1/	27	38	43	7	6	46	6 30/	8 5 2 7	14 771
111DT_r3 /0	40	17	42	40	27	21	30	42	20	18	51	7 012	8 702	14,805
111DT_r3_49	40	2	40	-72 10	2	0	17	50	1	2	48	5 300	0,720	15,003
111DT_r2 20	51	<u>2</u>	40	35	16	12	11	51	13	10	<u>40</u>	6 706	9 561	15 063
1001-10_00			+3	00	10	14		U 1	10	16	H U	0,100	0,001	10,000

Final ranking was created within Docker Analyzer v.1.1 with default configuration parameters and default dictionaries files. Table 1 show results obtained after first, rigid run of docking protocol. Calculations were not limited to any specified binding site and covered most part of protein chain. Table 2 contains consecutive run with strict restrictions of docking space. Both tables are accessible as supplementary materials on journal homepage. Files containing detailed results of these calculations for 50 best docked conformations can be downloaded from project home page³⁶.

The first column shows given conformation number. The next eleven represents results of calculated scoring functions. "Cons1" is average of all scoring functions. Ranking was sorted out in relation to the last column (marked as "Cons2") witch contains consensus scoring function created on base of 5 scoring functions: Autodock, PLP, PMF, D-Score and HPScore. In both Cons1 and Cons2 equal weights were used. The last column contains RMSD calculated directly without molecular fit (without hydrogen atoms), for crystal form and docking hits and was used to chose best conformation in this comparison (on top of Table 1 and Table 2). The worst results were filtered out as energetically unfavorable (data not shown).

When analyzing all the complexes obtained, two possible binding sites may be selected: correct one – characterized by crystal form (Fig. 2) and incorrect one (non native but energetically favored) with ligand bound in place distant from the native functional site (Fig. 3).

Conformations recognized as with the best agreement with the experiment (crystal structure marked as "Viagra") are: *1UDT-r3_13, 1UDT-r3_19* and *1UDT-r3_22*. Autodock built-in scoring function qualified the first one as the best solution with the highest binding energy as well as the lowest internal energy.

Also other functions select this conformation as energetically acceptable. In created consensus solution ranking (normalized means of 5 selected scoring functions) it may be found on third place. Comparison of this conformation to crystal form gives also satisfactory results. Residues in enzyme active site having contact with ligand molecule in these two cases were nearly the same.

Visualization of *1UDT* (red color) together with *1UDTr3_13* and crystal form of the ligand bound (color yellow and blue respectively), are shown in Figure 2 and depicts relatively small difference between locations of these two molecules in the proteins active site and relatively small RMSD value (8,871). The best geometrical agreement is found for model *1UDT-r3_22* and expressed by the lowest RMSD: 3.705. In fact, as it is seen in Figure 4, this molecule is located inside the native binding site and in correct spatial arrangement. Although, this conformation is not correctly assessed by Autodock scoring function (21st position in ranking) it is well ranked on 4th place by "Cons2" consensus scoring. This exemplifies discrepancy which may occur between the given native scoring function and experimental data.

Perspectives

This paper presents preliminary version of Docker Analyzer. Followed changes are planned to be introduced in coming upgrades:

- Supervised learning for automatic calculations of weights for consensus scoring function.
- Implementation of other chosen scoring functions, among them those introduced in new version of Dock program.
- Improvement of ligand display and manipulation. More flexible switching between conformations and enrichment of molecular display styles (e.g. sticks, bonds and surfaces).

Presented program will be also tested on a group of protein – ligand complexes, which structures were derived experimentally (taken from PDB). For each given class of complexes specific weighting factors will be derived to better rerank correct structures in each given class. This work is already under preparation and results will be published soon.

References

- J.L. Wang, D. Liu, Z.J. Zhang, S. Shan, X. Han, S.M. Srinivasula, C.M. Croce, E.S. Alnemri, Z. Huang. Structurebased discovery of an organic compound that binds Bcl-2 protein and induces apoptosis of tumor cells. Proc. Natl. Acad. Sci. USA, 2000, 97, 7124-7129.
- J. Leban, W. Saeb, G. Garcia, R. Baumgartner, B. Kramer. Discovery of a novel series of DHODH inhibitors by docking procedure and QSAR refinement. Bioorg. Med. Chem. 2004, 14, 55-58.
- H.-L. Liu, J.-C. Lin. Structural model of the voltage-gated potassium channel Kv1.1 and molecular docking of Tc1 toxin from *Tityus cambridgei* to KcsA and Kv1.1, Chem. Phys. Lett., 2003, 381, 592-597.
- B.K. Schoichet, S.L. McGovern, B. Wei, J.J. Irwin. Lead discovery using molecular docking. Curr. Op. Chem. Biol., 2002, 6, 439-446.
- J.S. Mason, D.L. Cheney. Ligand-Receptor 3-D Similarity Studies Using Multiple 4-Point Pharmacophores. Pacific Symposium on Biocomputing, 1999, 456-467
- R.A. Laskowski, J.D. Watson, J.M. Thornton. ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res., 2005, 33, W89-W93.
- M. Jambon, A. Imberty, G. Deléage, Ch. Geourjon. A new bioinformatic approach to detect common 3D sites in protein structures. Proteins, 2003, 52(2), 137-45.
- M. Brylinski, L. Konieczny, I. Roterman. Ligation site in proteins recognized in silico. Bioinformation, 2006, 1, 127-129.
- I. Schellhammer, M. Rarey. FlexX-Scan: Fast Structure-Based Virtual Screening, PROTEINS: Structure, Function, and Genetics, 2004, 57, 504-517.
- G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. J. Computational Chemistry, 1998, 19, 1639-1662.
- D.T. Moustakas, P.T. Lang, S. Pegg, E.T. Pettersen, I.D. Kuntz, N. Broojimans, R.C. Rizzo. Development and Validation of
a Modular, Extensible Docking Program: DOCK 5. J. Comput. Aided Mol. Design., 2006, 20, 601-609.

- R.A. Friesner, R.B. Murphy, M.P. Repasky, L.L. Frye, J.R. Greenwood, T.A. Halgren, P.C. Sanschagrin, D.T. Mainz. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. J. Med. Chem., 2006, 49, 6177-6196.
- R. Wang, Y. Lu, S. Wang. Comparative evaluation of 11 scoring functions for molecular docking. J. Med Chem, 2003, 46, 2287-2303.
- 14. D. Beazley. Advanced Python Programming. O'Reilly 2000.
- 15. W. Chun. Core Python Programming. Prentice Hall PTR 2000.
- 16. M. Lutz. Programming Python. 2nd Edition. O'Reilly 2001.
- 17. GTK+, http://www.gtk.org/ (3.11.2005).
- 18. Libglade Reference, http://developer.gnome.org/doc/API/ libglade/libglade.html (2.02.2006).
- 19. GtkGLExt, http://www.k-3d.org/gtkglext/ (2.02.2006).
- 20. OpenGL, http://www.opengl.org/ (15.02.2006).
- 21. NeHe, http://nehe.gamedev.net/ (15.02.2006).
- 22. H. Bohm. The development of a simple empirical scoring function to estimate the binding konstant for a protein-ligand complex of known three-dimensional structure. J. Comput.-Aided Mol. Des., 1994, 8, 243-256.
- User Guide. Cerius² 4.10L-Structure Based Focusing. Accelrys 2005
- 24 M. Eldridge, C. Murray, T. Auton, G. Paolini, R. Mee. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J. Comput.-Aided Mol. Des., 1997, 11, 425-445.
- 25 D. Gehlhaar, G. Verkhivker, P. Rejto, C. Sherman, D. Fogel, L. Fogel, S. Freer. Molecular recognition of the inhibitor AG-

1343 by HIV-1 Protease: Conformationally flexible docking by evolutionary programming. Chem. Biol., 1995, 2, 317-324.

- 26. Accelrys Inc, http://www.accelrys.com (17.06.2006).
- A. Krammer, P. Kirchhoff, X. Jiang, C. Venkatachalam, M. Waldman, LigScore: a novel scoring function for predicting binding affinities. J. Molec. Graph. and Mod., 2005, 25, 395-407.
- 28 M. Rarey, B. Kramer, T. Lengauer, G. Klebe. A fast flexible docking method using an incremental construction algorithm. J. Mol. Biol., 1996, 261, 470-489.
- T. Ewing, S. Makino, A. Skillman, I. Kuntz. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. J. Comput.-Aided Mol. Des., 2001, 15, 411-428.
- G. Jones, P. Willett, R. Glen, A. Leach, R. Taylor. Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol., 1997, 267, 727-748.
- S. Weiner, P. Kollman, D. Case. A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc., 1984, 106, 765-784.
- I. Mueegge. Effect of ligand volume correlation on PMF scoring. J. Comput. Chem., 2001, 22, 418-425.
- H. Gohlke, M. Hendlich, G. Klebe. Knowledge-based scoring function to predict protein-ligand interactions. J. Mol. Biol., 2000, 295, 337-356.
- R. Wang, L. Lai, S. Wang. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J. Comput.-Aided Mol. Des., 2002, 16 (2), 11-26.
- H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, P. Bourne. The Protein Data Bank. Nucl. Acids Res., 2000, 28, 235-242.
- Docker Analyzer Home Page, http://www.bioinformatics.cmuj.krakow.pl/docker_analyzer.html (02.02.2007).



Bioinformatics

ESTIMATION OF EXPOSURE IN ELECTROMAGNETIC FIELDS OF EXTREMELY LOW FREQUENCY

OCENA EKSPOZYCJI NA ELEKTROMAGNETYCZNE POLE O EKSTREMALNIE NISKIEJ CZĘSTOTLIWOŚCI

MICHAŁ ZEŃCZAK

Institute of Electrical Engineering, Technical University of Szczecin, mzenczak@ps.pl

Abstract: In order to answer the question if electromagnetic fields of 50 Hz are harmful for living systems examinations in vivo and in vitro, laboratory tests and epidemiological studies are carried out. The exact estimation of exposure should be known by such the investigations. The paper contains the analysis of problems related to exposure in electric and magnetic fields of 50 Hz in various conditions. The methods of estimation of exposure for some cases are presented.

Keywords: electric field, magnetic field, power transmission line, biological effect, exposure.

1. Introduction

Electromagnetic fields of extremely low frequency (the most popular frequency is 50 Hz) influence natural environment and occupational environment. For these frequencies the quasi-stationary condition is fulfilled. Therefore both of components can be analyzed separately. Electric field (EF) exists between object with different potentials, while magnetic fields (MF) exists near circuits with electric currents. Lines of EF intensities and MF intensities have very often complicated geometrical shapes. Besides r.m.s values of intensities show big time variability (especially MF intensity).

The problem of harmfulness of these fields has been analyzed for many years. Four kinds of investigations referring to biological effects and harmfulness are carried out [1]:

- in vitro,
- in vivo,
- laboratory with exposed organisms,
- epidemiological.

In order to draw correct conclusions it is necessary to know exact value of exposure of examined living objects or whole population. Investigations in vivo, in vitro and laboratory create easier conditions for exact determination of exposure. During epidemiological investigations as the exposed populations the whole groups of workers of substa**Streszczenie:** W celu określenia wpływu pola elektromagnetycznego o częstotliwości 50 Hz został opracowany test dla oceny epidemiologicznej. Zaprezentowano analizę efektów wpływu pola o częstotliwości 50 Hz w różnych warunkach.

Słowa kluczowe: pole elektryczne, pole magnetyczne, sieci wysokiego napięcia, efekty biologiczne, narażenie

tions or inhabitants living near power transmission lines are analyzed without designation the intensities of fields and time of exposure [1]. There are classifications on electrical and non-electrical professions. When there is more information about the time and field intensity, exposure is divided into "high", "medium" and "low". Drawing conclusions about dependence between biological effects and value of exposure is embarrassing especially during epidemiological investigations. The paper presents the problems with estimation of exposure in electric and magnetic fields of extremely low frequency for different conditions and situations. The results are compared with the existing regulations [2, 3].

2. Classifications of field sources

Sources of EF and MF are very heterogeneous. Many different criterions of classifications can be introduced. With regard to possibility of mathematical modeling it can be recognized:

- Sources with simple geometry, for instance: power transmission lines,
- Sources with irregular geometry for instance: switching devices, household devices.

According to constant geometry there are:

- Sources with constant geometry in time,
- Sources with time varying geometry.

Taking into consideration variability of effective value there are:

- Sources with constant effective value,
- Sources with time varying effective value. According to value of voltage and current:
- Sources with high intensities,
- Sources with low intensities.

Taking into consideration ranges of influence the following fields can be defined:

- Sources of fields with big ranges including much more bigger area than dimensions of human organism (power transmission lines),
- Sources of local fields changing considerably value of intensity in area comparable to dimensions of human organism (household devices, low voltage substations). In consideration of accessibility there are:
- Sources in generally accessible areas (under power lines),

 Sources in prohibited (not for staff) areas (substations). Besides sources can be movable or immovable or exposed objects can be movable or not.

3. Definition of exposure dose

Value of field intensity K in any point of area near n sources of fields can be expressed as the functional:

$$K = F[q_1(t),.,q_n(t),x_1(t),y_1(t),z_1(t),.,x_n(t),y_n(t),z_n(t)]$$
(1)

where: $q_k(t)$ - source number k, while $x_k(t)$, $y_k(t)$, $z_k(t)$ - coordinates of this source.

Exposure dose (ED) can be defined as the following relationship:

$$ED = \int_{0}^{T} K^{p}(t) dt \approx \sum_{k=1}^{n} K^{p}_{k} \Delta t_{k}$$
(2)

where: K_k -average intensity of field in time period t_k . Usually p=2, what shows the energetistic effect. It is conformable to polish regulations [3].

ED can have deterministic or random character. It depends on character of field and exposed object. ED can be analyzed as the deterministic only in one case, when field and object are deterministic. Fig. 1 presents deterministic and random characters of ED.



Fig. 1. Deterministic and random characters of dose

In case of deterministic analysis the most unprofitable conditions are assumed and the ED is determined by equation (2). But for the estimation the real ED the random attributes should be taken into consideration. The random attributes of ED are very important during long term observations in changeable fields.

4. Methods of estimation of exposure doses

Generally the following methods of estimation of exposure doses can be distinguished: measurement, calculation and combination both of them.

4.1. Measurement method

In order to estimate ED it is necessary to equip the exposed person with recorder of field intensity K. The average value of field intensity should be estimated for each time period t. As the result daily chart, month chart or year chart are obtained. This chart can be base for estimation the value of ED in agreement with equation (2). Recorder can calculate the current value of equation (2) and signal the overpassing value of ED in agreement with standards [2, 3], or signal only overpassing of a permissible value of field intensity in agreement with standards [2, 3].

Such the measurement is easy for MF, because living organism does not deform it. Therefore measured intensities of MF are the same like without presence of exposed person.

The situation is much more complicated for EF, because living organism deforms EF, mainly near sharp conductive edge (limbs, head, feelers), where EF intensity is much more higher, than primary homogenous field. In this case depending on principle of operation recorders notice different value than (1) and (2). Therefore it is difficult to compare the results of exposition to values of standards from regulations [2, 3], which refer to undeformed fields.

There is possibility to introduce some special coefficients of correction for K and ED, depending on position of the organism. On the other hand ED defined according to [3] does not reflect real effect of field on living organism, especially local effect in places, where there are congestion of EF lines, for example near head.

The above mentioned principles refer to field of big range, which are homogenous in area comparable to dimensions of human organism.

Measurements should be local for local fields. Although according to standards [3] the local exposure is mentioned only for MF and refers to limbs. In this case the permissible values of fields are five time higher and values of ED can be 25 time higher.

4.2. Calculation methods

Calculation methods base on analytical or numerical estimation of field intensity according to (1) and ED according to (2). These relations should take into consideration properties of fields mentioned in point 2 and showed in fig. 1. For the fields with simple geometry it is possible to create the analytical relationships for K and ED. The most popular sources of EF and MF in natural environment are power transmission lines. There are many algorithms for calculation EF and MF intensities near lines [4, 5, 6].

Table 1 presents parameters of typical polish line 400 kV on the pylon Y52.

Wire	Cross- section [mm ²]	Bundle [m]	Distance from axis [m]	Height of suspension [m]
L1	2×525	0,4	- 10,3	7,7
L2	2×525	0,4	0	7,7
L3	2×525	0,4	10,3	7,7
01	70	-	-8,2	13,7
O2	70	-	8,2	13,7

Table 1. Parameters of line 400 kV on pylon Y52

The highest value of voltage for 400 kV line was assumed: 420 kV. Height of suspension was assumed according to polish old standard [7]: 5+Un/150. Using old standard is quite justifiable, because all polish lines were built according to old regulation [7].

EF intensity under line is complicated function of distance from axis of line. This complication is connected with presence of vertical and horizontal components of vector of EF intensity and with fact, that both components have different time displacement.

Fig. 2 presents EF intensity under 400 kV line on pylon Y52.

EF distribution under power line with the symmetrical flat formation of wire can be approximated by polynomial of the 6th degree and in some distance from axis of line by the power function:

$$W = \begin{cases} Ax^{6} + Bx^{4} + Cx^{2} + D & dla \ x \in (-x_{g}; x_{g}) \\ A_{1}x^{-2} & dla \ x \le -x_{g} & lub \ x \ge x_{g} \end{cases}$$
(3)

where: x - distance from axis of line, $x_g - boundary$ value between polynomial and power function. Boundary value x_g occurs under extreme wires. In fig. 2: Ea - results of exact calculations [4], W - results of approximation.



Fig. 2. EF intensity under line 400 kV on pylon Y 52. Ea – results of exact calculations [4], W – results of approximation

EF distribution under 400 kV line can be approximated by following functions:

$$W = \begin{cases} -0,0000177x^{6} + 0,0039x^{4} - 0,202x^{2} + 9,4 \\ dla \ x \in (-12,0;12,0) \\ 1318x^{-2} \\ dla \ x \le -12,0 \ lub \ x \ge 12,0 \end{cases}$$
(4)

In point of view of protection of natural environment the range is interesting where EF intensity exceeds 1 kV/m. In this area error does not exceed 46%, but in majority of area error does not exceed 20%. For other line 220 kV and 110 kV results are similar.

In point of view of protection of occupational environment the area is interesting where EF intensity exceeds 5 kV/m. In this area the error does not exceed 21%.

The same can be done for MF. Distribution of MF under power transmission line can be calculated using law of Biot-Savarte and superposition method [4]. Obtained relationships are complicated too. But they can be approximated by polynomial of 4th degree and power function.

Fig.3 presents distribution of MF under power line 400 kV on pylon Y 52. In fig. 3: Ha – result of calculation, W – result of approximation:

$$W = \begin{cases} -0,001625x^{4} + 0,206x^{2} + 21,38\\ dla \ x \in (-11,5;11,5)\\ 3200x^{-2} \ dla \ x \le -11,5 \ lub \ x \ge 11,5 \end{cases}$$
(5)

Error of approximation does not exceed 10%.

MF under power line shows big time variability because of variability of current (from 0 A to maximal current capacity). Values of MF intensity in fig. 3 were calculated for I =1000 A.



Fig. 3. Distribution of MF intensity under 400 kV line on pylon Y 52 for I = 1000 A

In order to obtain MF intensity for other current I_{rob} , the results from fig 3 should multiplied by coefficient k:

$$k = \frac{I_{rob}}{1000}$$
(6)

For some assumptions ED can be calculated using (2). If the man goes with the constant velocity v in direction perpendicular to axis of line from "minus side", then his coordinate can be expressed by simple formula:

$$\mathbf{x}(t) = \mathbf{x}_{\rm lm} + \mathbf{v} \cdot \mathbf{t} \tag{7}$$

where x_{lm} is the place, where the essential values of field intensity begin, for example: 1 kV/m for EF and 1 A/m for MF. Formula (7) can be conversed assuming, that movement performs in both direction from axis of line and the time in"minus direction" has the sign "-":

$$\mathbf{x}(t) = \mathbf{v} \cdot t \tag{8}$$

Taking into account (8) in (3):

$$W = \begin{cases} A_{v}t^{6} + B_{v}t^{4} + C_{v}t^{2} + D \, dla \, t \in (-t_{g}; t_{g}) \\ A_{1v}t^{-2} & dla \, t \leq -t_{g} \, lub \, t \geq t_{g} \end{cases}$$
(9)

where: A_v , B_v , C_v ...new coefficients obtained by multiplying A, B, C ... by adequate powers of v. Expression (9) enables to estimate value of ED under power transmission line:

$$ED = 2 \int_{0}^{t_{m}} W^{2}(t) dt$$
 (10)

where:

$$t_{m} = \frac{X_{m}}{V}$$
(11)

When the distribution of field is nonsymmetrical (for example: 110 kV line on pylon P), then ED can be obtained from formula:

$$ED = \int_{0}^{t_{ml}} W^{2}(t) dt + \int_{0}^{t_{mp}} W^{2}(t) dt$$
(12)

where: t_{mi} and t_{mp} – moments, when man is in points, where field intensity reaches essential value.

Calculation of integral (10) is very simple.

4.3. Combination of methods

Combination of methods is based on measurements of field intensities in each point of analyzed area. Knowing values of field intensities in each point and knowing ways or time and place of stay it is possible to calculation ED using expression (2).

5. Polish standards

There are special safety rules of protection against electromagnetic fields of 50 Hz frequency. There are so called environmental rules [2] and occupational ones [3].

The permissible value of EF intensity in natural environment must not exceed 10 kV/m [2]. The permissible value of magnetic field intensity in natural environment is equal to 60 A/m. But in places appropriated for the public building the highest value of electric field intensity must not be higher than 1 kV/m and magnetic field intensity 60 A/m.

In compliance with the occupational standards the area of field interaction has been divided into four zones: the dangerous, hazardous, intermediate and safe. In the dangerous zone staying the workers is prohibited. In hazardous zone staying is permissible when the doses $(ED_{\rm F}=E^2t,$

 $ED_{H}=H^{2}t$, where t8 h) do not exceed the permissible values $(ED_{Ep}=800(kV/m)^{2}h, ED_{Hp}=0.32(kA/m)^{2}h)$. During the whole shift workers can stay in intermediate zone. Safe zone is the area beyond remaining, where there is no limit. There are three limiting values for electric field E_{0} , E_{1} , E_{2} and for magnetic field H_{0} , H_{1} , H_{2} :

 $\rm E_{_0}$ and $\rm H_{_0}\,$ limiting values between safe zone and intermediate one,

E₁ and H₁ limiting values between intermediate zone and hazardous one,

E₂ and H₂ limiting values between hazardous zone and dangerous one.

For 50 Hz there are following values: $E_0 = 5 \text{ kV/m}$, $E_1 = 10 \text{ kV/m}$, $E_2 = 20 \text{ kV/m}$ i $H_0 = 66 \text{ A/m}$, $H_1 = 200 \text{ A/m}$, $H_2 = 2000 \text{ A/m}$.

In hazardous zone ED should fulfill following relationships:

$$ED_{E} ED_{E}$$

 $ED_{H} ED_{H_{F}}$
WE 1,

where: ED_{E} – real dose of EF, ED_{H} – real dose of MF, ED_{Ep} , ED_{Hp} – permissible values of ED for EF and MF, WE – indicator of exposure:

$$WE = \frac{ED_E}{ED_{Ep}} + \frac{ED_H}{ED_{Hp}}$$
(13)

There is some disagreement between regulation [2] and [3]. If EF intensity is below 5 kV/m, there is safe zone and it is no limits according to [3]. According to [2] there is no limits, if EF intensity is below 1 kV/m. There is some proposal of changing regulation in order to equalize the permissible value of natural environmental exposition with the value of safe zone [8].

6. Examples of exposure

EF intensity under polish transmission lines is equal to a few kV/m. Only under 400 kV line in unfavorable conditions EF intensity exceeds 10 kV/m. MF intensity does not exceed generally 30,40 A/m. Only under double circuit line MF intensity exceeds 60 A/m.

On order to compare exposure under lines to permissible values the relationships (10) or (12) can be used. For example the person walking perpendicular to axis of 400 kV line with speed 1 m/s is exposed to dose of EF 0,648 (kV/m)²h, and to dose of MF 5×10^{-6} (kA/m)²h. Indicator of exposure WE = 0,0008<<1.

The highest values of EF intensity are in substations. As an example can be area of bay in substation 400 kV (fig.4).

There are many other possibilities to stay in EF and MF. But fortunately the values of EF intensities near household devices are below 1 kV/m. But values of MF intensities near household devices are comparable to values under power lines. Values of MF intensity are presented in table 3.

Ships are very interesting objects, because of low voltage installation (3440/3220 V) with relatively high power (a few MW). For example tubular rudder has nominal power 590 kW. Therefore there are relatively high values of MF



Fig. 4. Electric field intensity in the bay of substation $\,400\;\text{kV}$

intensity, especially in control room, where there is switchgear. Watch engineer has behind him switchgear. Fig. 5. presents plan of control room.

Table 3. MF intensity near household devices

Device	H [A/m] case	H [A/m] in distance 10 cm
Mixer 180 W	32,7	2,0
Safety razor	30,0	2,0
Monitor 150 W	21,1	9,1
Micro. stove 2,5 kW	15,0	10,2
Vacuum cleaner 500 W	15,0	2,1
Iron 1,2kW	8,2	1,2
TV set 70 W	3,5	1,5
TV set 80 W	3,0	1,2
Hairdryer 250 W	2,8	<0,2
Kettle 2 kW	2,4	<0,2
Compact lamp 18W	2,2	0,4
Heater 750 W	1,4	<0,2
Toaster 750 W	1,2	<0,2
Electric oven 1 kW	1,1	0,3
Radio set 7 VA	0,9	<0,2

The highest values of MF intensity occur near switchbays 8, 9, 11 (generators) and near switch-bay 12 (supply of tubular rudder). Watch engineer is exposed to MF 1 A/m to 5 A/m. Exposed dose can exceed during 8 hours 0,0002 (kA/m)²h. The highest value of MF intensity on the height 1,8 m obtains 80 A/m during work of tubular rudder 590 kW. During 1 hour the ED of MF obtains 0,006 (kA/m)²h. Some area in control room is intermediate zone.

7. Conclusions

Basing on measurements carried out in many different places the following conclusions can be drawn:

- According to polish regulations only near high voltage installations there is necessity to control exposure dose of electric field,
- Exposure dose of magnetic field should be controlled near installations with high currents: generators, busbars,



Fig. 5. Plan of control room

- 42
- Presented methods can be used for estimation of exposure during investigations referring to biological effects and harmfulness of electromagnetic fields and to estimation of risk for general population or workers,
- Approximated formulas enable in easy way to estimate the exposure near power transmission lines.

8. References

- Zmyślony M., Aniołczyk H.: Oddziaływanie pól elektromagnetycznych na człowieka – metodyka prowadzenia badań i ocena wiarygodności ich wyników; Konferencja: Pola elektromagnetyczne 50 Hz w środowisku człowieka, Poznań, 27-29 października 2003 r., p. 21-30.
- Rozporządzenia Ministra Środowiska z dnia 30 października 2003 r. w sprawie dopuszczalnych poziomów pól elektromagnetycznych w środowisku oraz sposobów sprawdzania dotrzymania tych poziomów, Dz.U. nr 192, poz. 1883, Warszawa, 14 listopada 2003 r., poz. 1883.
- 3. Rozporządzenie Ministra Pracy i Polityki Społecznej z dnia 29 listopada 2002 roku w sprawie najwyższych dopuszczalnych

stężeń i natężeń czynników szkodliwych dla zdrowia w środowisku pracy, Dz.U. Nr 217, 18.12.2002, poz. 1833.

- Zeńczak M.: Analiza technicznych problemów związanych z dozymetrią pól elektromagnetycznych o częstotliwości przemysłowej; Prace Naukowe Politechniki Szczecińskiej, Instytut Elektrotechniki nr 543, Szczecin 1998.
- Krajewski W.: Elementy brzegowe i liniowe w analizie wybranych zagadnień EMC niskiej częstotliwości; Prace Instytutu Elektrotechniki, Zeszyt 224, Warszawa 2005.
- Wróblewski Z., Habrych M.: Ocena przydatności obliczeniowej metody wyznaczania rozkładu pola elektromagnetycznego pod liniami WN. Przegląd Elektrotechniczny, R. 80. Nr 5/ 2004. str. 5110-5114.
- PN-E-05100-1: Elektroenergetyczne linie napowietrzne. Projektowanie i budowa. Marzec 1998.
- Zmyślony M., Politański P.: Propozycje dopuszczalnych wartości natężeń pól elektromagnetycznych w paśmie od 0 Hz do 1 MHz dla celów bhp oraz w ramach ochrony środowiska. Przegląd Elektrotechniczny, nr 12/2005, p. 66-66.



WAVELET BASED CLASSIFICATION OF SKIN LESION IMAGES

KLASYFIKACJA OBRAZÓW ZMIAN SKÓRNYCH

GRZEGORZ SURÓWKA*, CHRISTIAN MERKWIRTH*, EWA ŻABIŃSKA-PŁAZAK**, ANDRZEJ GRACA*

Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Krakow, Poland, grzegorz.surowka@uj.edu.pl

^{**}Department of Dermatology, Collegium Medicum, Jagiellonian University, Kraków, Poland, zabinskae@wp.pl

Abstract: Visual examination of the early stages of the melanocytic skin cancer (melanoma) may often lead to a false diagnosis. Only the resection and then histologic examination of the lesion can fully detect malignant transformations of the skin. This is the reason why development of non-invasive methods for dermatological diagnosis, like dermatoscopy, is of key importance. We build a MLP-based binary classifier for discriminating melanoma from dysplastic nevus utilizing textural information contained in the skin lesion images taken in dermatoscopic examinations. Our analysis is based on the multiresolution wavelet-based decomposition of the images. Significant features of both classes are found by means of the Ridge regression models. Discriminating melanoma from dysplastic nevus with this method yields a sensitivity and specificity of 89.5% and 90%, respectively.

Key words: melanoma, dermatoscopy, wavelet transform, machine learning, automatic diagnosis

Introduction

Epidemiologists report that incidence of melanoma has grown all over the world [1]. Although the non-melanocytic types of the skin cancer are more common, melanoma is the most malignant human cancer and its mortality is very high [2]. Early detection and mass screening is the key of effective treatment of melanoma [3].

Melanoma is caused by malignant transformations of melanocytes (pigment cells) in the epidermis. The following degrees of this progress are possible: benign melanocytic nevus, dysplastic nevus which reveals some geometrical and cytologic atypia and then some phases of malignant melanoma. The latter are the radial growth phase where the lesion expands horizontally within the epidermis and the vertical growth phase where the cells expand to the dermis [4].

There are some measures that describe and help diagnose skin lesions. The measures can be divided into descriptive ones which are based on a visual examination and some laboratory-based variables. Streszczenie: Wizualna ocena wczesnych stanów procesu nowotworzenia skóry może prowadzić do błędnej diagnozy. Jedynie resekcja oraz histologiczna ocena może ocenić obecność procesu nowotworzenia. Stąd potrzeba nieinwazyjnej oceny w dermatologii jest potrzebą chwili. Zbudowaliśmy bazujący na MLP binarny klasyfikator dla dyskryminacji melanoma w oparciu o obrazy uzyskane dermatoskopowo. Metoda bazuje na dekompozycji obrazu. Model regresji Ridge'go został zaadaptowany dla klasyfikacji obrazu co dało specyficzność oceny rzędu 89.5% i 90%.

Słowa kluczowe: melanoma, dermatoskopia, uczenie maszyny, automatyczna diagnoza

Descriptive systems include the ABCD(E) rule and the 7-Point Checklist to mention the most common. The ABCD(E) rule [5,6] is a semi-quantitative analysis of five criteria: (A)symmetry, (B)order, (C)olors and (D)ifferential Structural Components found in the lesion. As an additional factor to the traditional ABCD classification, (E)volutionary change of the lesion is also taken into account. Each criterion is weighted to contribute to the total dermatoscopic score (TDS) classifying the lesion into three fuzzy categories: benign, suspicious and melanoma-like lesion.

Another system of visual classification of the pigmented lesions is the 7-Point Checklist [7]. The lesions are evaluated according to three major (atypical pigment network, gray-blue areas, atypical vascular pattern) and four minor (radial streaming, irregular pigmentation/blotches, dots and globules, regression pattern) criteria where each major (minor) criterion receives two (one) points. The total score greater than three is a distinctive feature of melanoma. The sensitivity and specificity of the descriptive rules is subject to research and discussions [8]. Laboratory tests, on the other hand, assume resection of the lesion, its preparation and view under the microscope. It should be emphasized that biopsy is the only fully reliable method to identify nevi and melanoma lesions. For melanoma two main staging schemes have been proposed based on histologic criteria: the Clark's level and the Breslow's depth.

The Clark's level (I-V) differentiates the degree of tumor penetration quantitatively while the Breslow's depth is an actual micrometer measurement of the lesion depth i.e. its vertical spread to the dermis. The latter is grouped into four categories (<0.75mm, 0.75-1.5 mm, 1.5-4.0mm and >4.0mm) which determine the prognosis of the case.

Invasive treatment of all suspicious skin lesions is not feasible and development of non-invasive methods for dermatological diagnosis is of key importance [9]. Epiluminescence Microscopy (ELM) (=surface microscopy, dermatoscopy) is a non-invasive technique that has been developed for decades. Dermatoscopy consists in visual examination of skin lesions that are optically enlarged and illuminated by halogen light. In order to get rid of the light reflections through the air gap immersion oil is used on the skin. The magnified field of the lesion can be digitally photographed (plain dermatoscopy) or displayed on a computer screen on-line (videodermatoscopy) for analyses of its surface structure. The most common (video)dermatoscopy systems nowadays are Dermogenius and Molemax II [12]. Some dedicated instruments [13] allow not only for the surface illumination (perpendicular to the skin surface) but also for trans-illumination or a combination of both. In the transillumination mode the light is directed into the skin at an angle of 45° making it possible to penetrate deeper layers of the skin and thus attempting to reveal its 3D structure.

Digital imaging has a great advantage to record how lesions develop in time. Computer acquisition systems integrated with conventional dermatoscopy (=DELM, Digital ELM) can store, retrieve and transmit dermatoscopic images. This approach can improve monitoring changes in pictures of the same lesion and introduce remote diagnosing by experienced specialists, which can generally increase accuracy of the diagnosis [14,15]. Telemedicine is future for mass screening of the skin cancer and some national tele-diagnostic centers for dermatoscopy have been established by now [16].

It is reported [10,11] that sensitivity and specificity of the descriptive rules for melanoma detection is generally enhanced by means of dermatoscopy. However the early detection of cutaneous neoplasmas with high accuracy is still unsatisfactory. In the earliest stages of malignancy even experienced specialists cannot distinguish between benign and malignant lesions which is the necessary condition of successful treatment and can be a life-saving factor [17].

New identification methods based on spatial and frequency information found in the skin texture have been proposed. They assume that some neighborhood properties of pixels in dermatoscopic images can be a sensitive probe of different pigmented skin atypia and the melanoma progression. The most promising methods are those decomposing different frequency scales of the texture. Such multiresolution analysis is well suited to determine distinctive signals characterizing the class of the image [18,19].

Wavelet transforms have been widely studied as tools for a multi-scale pattern recognition analysis. Wavelets are functions well localized both in space and frequency. The wavelet theory is closely related to the theory of digital filtering [21,22] so the properties of the decomposition filters (the choice of the basis, degree of regularity, the sub-bands of interest) play an important role in texture characterization.

Performance of the wavelet analysis of images depends on various factors [26]. i) For feature extraction it is important to choose between recursive decomposition of the lowfrequency branch [22] (the pyramidal algorithm) and, on the other hand, a more selective tree-structured analysis where the further decomposition is applied to the output of any channel [19,20]. The latter is advocated by observations that the significant sub-bands of the pigmented skin texture are those in the middle frequency range [18, 19]. ii) The order of the wavelets is another important property. The decomposition should be performed over an optimal finite range of resolutions [20, 22]. iii) Different wavelet bases have diverse impact on the texture classification [21,25]. They have certain constraints and should, in principle, be orthonormal (energy preserving, non-redundant). Biorthogonal (near orthogonal) bases can produce, on the other hand, more compact and symmetric wavelets and the requirement for orthogonality is not superior [22].

iv) Images are two-dimensional signals so separable or non-separable sampling is possible. In the former case a one-dimensional decomposition algorithm is applied to the rows and to the columns of the image [27]. The resulting signal is a tensor product of the separate 1D filters. The non-separable case, on the contrary, is based on lattice sampling [23,24]. Both representations have advantages and disadvantages.

In the following sections we present our direct motivation for the wavelet based classification of dermatoscopy images, describe all the steps of the analysis and present the results.

Motivation

We have briefly explained the overall motivation for the melanoma screening program. Now we present our objectives for this work.

Our working hypothesis is that dermatoscopic images can be efficiently classified in a computer manner. For the case of skin textures i.e. signals consisting mainly of middle frequency bands it has been shown that the most adequate approach to the wavelet-based multiresolution analysis is the concept of wavelet packets [20]. In this approach an image is decomposed along dominant frequency channels forming a parent-child structure of a tree. A certain path of the tree is chosen according to the average energy content of the channel referenced to the highest energy of a channel on the same decomposition level. The determined subbands form a feature set for classification. This general idea has been recently developed. The main bias of the method in [20] is the threshold value of the channel energy that triggers subsequent decompositions. Since features used for selection of melanoma and dysplastic lesions are clustered into rough sets, the structure of the tree may be effected by an arbitrary choice of the threshold. In the adaptive wavelet-based tree structure analysis (ADWAT) [18,19] decision about the further decomposition of the channel is taken by statistical analysis of average energy, maximum energy ratio and fractional energy ratio among all the features at each level of decomposition. The resulting tree structured models of melanoma and dysplastic nevus are patterns for semantic comparison with unknown skin lesion images. The sensitivity and specificity of this method is further improved by introducing Gaussian and Bell shaped fuzzy clustering of the features [18].

Our approach is based on the idea presented in [19] to perform a tree-like selective wavelet analysis of the skin lesion images. Unlike in [19] our feature set is built from all of the analyzed channels and the decision to include a given branch of the tree does not come from ADWAT. Instead, all the wavelet based features are input to the Ridge classification model [31], which determines a minimum–bias optimal vector of features. Those selected features are used to teach a three-layer back-propagated neural network. Taught from examples the neural network serves as a binary classifier of the dermatoscopy images with melanoma and dysplastic nevus lesions. It should be stressed that this research aims at developing an aid to the physician and not any diagnosing system all alone.

Procedure

Now we explain the tasks we have performed to build the binary classifier discriminating melanoma and dysplastic lesions.

Our database of dermatoscopic images consisted of 19 cases of melanoma and 20 cases of dysplastic nevi. The selection was done on the grounds of biopsy and histopathologic examination. The data set was collected using a Minolta Dimage Z5 digital camera equipped with an epiluminescence lense with white halogen lighting. The working resolution was set to 2272x1704 pixels and the quantization depth to RGB-8 (24bits). The images were selected to two categories: melanoma (label=1) and dysplastic nevus (label=0). The actual representation was subject to some geometrical transformations. Since the wave-



Fig. 1. Cropped and enlarged texture of an analyzed anonymous skin lesion

let decomposition downscales the input image by a factor of 2 in rows or columns every iteration, the width and length had to be powers of 2, so the source images were filled with black pixels in due number of rows and columns.

In order to apply any transform on the picture, the compressed image (JPG) has to be changed into a 2D matrix of numbers. Three ways of extracting the numerical values of pixels are possible:

(normal): binary values of the three channels R, G, B are put together (in this order) to compose one 24-bit long binary integer. This value is subtracted from the 24-bit long all-'1' binary number yielding a negative integer.

(average): an average value of R, G and B is calculated and stored as a floating point number.

(RGB): value of R, G, and B is stored independently in separate matrices. This approach assumes independent processing of the three channels.

The 2D=1Dx1D wavelet transform was applied to the values of pixels extracted from the discussed dermatoscopic images. The class of the filter was Daubechies 3 [28] and its efficient algorithm was taken from [29]. The choice for the filter was made concerning simplicity, performance and possible comparisons with [19].

One iteration of the wavelet algorithm produces 4 subimages which can be considered as LL, LH, HL and HH filters, where L and H denote the respective low-pass and high-pass filters. One subimage is a product of the wavelet transform acting on each raw and then on each column of the parent image. Each iteration reduces half of the rows and half of the columns from the parent image. This procedure is depicted in Fig.2.



Fig. 2. Pattern for the image decomposition. In each iteration four filters out of one input image are produced. One iteration is composed of two steps: row-wise and column-wise decomposition and 2:1 reduction

The algorithm presented in [29] is a pyramidal one i.e. it decomposes the image along the LL band in each iteration. An example of the decomposition process with 3 iterations is shown in Fig.3. Here the smooth and the detail part resulting from any iteration can be easily recognized.

According to the idea presented in the motivation of this work, we adapted the pyramidal algorithm to a selective analysis of any subband of the parent image. Altogether 3 iterations were calculated. The proper sequence of filters in that decomposition is shown in Fig.4. Following the original idea [20] and its extended applications [18,19] a feature set at a given level of decomposition was built. One iteration (resulting in 4 subimages) produced 11 coefficients calculated from the energy of the pixels (e_1 , e_2 , e_3 , e_4), the maximum energy ratio (e_i/e_{max} , i=any three out of four ex-



Fig. 3. Multi-band image i.e. a set of images of the same mole from its different frequency segments. Here recursive decomposition of the low-pass band takes place. For the sequence of the pyramid see Fig. 4. Note that the luminosity of this summary image is offset otherwise it would be too dark for a presentation

cept the subimage with the maximum energy) and the fractional energy ratio $(e_1/(e_2+e_3+e_4) + its three permutations)$. Energy means here the sum of the absolute values of the pixels, where M and N denote the actual dimensions of the subimage:

$$e = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} |f(m,n)|$$

Since the algorithm was applied to the image recursively, the number of coefficients amounts to (1+4+16)x11=231. 1, 4 and 16 mean here the number of subimages to filter in each iteration.

		1 23	1.4		.1 2.2	1111121212222211 212221222
	2	1.1	1.1 1.2 2	2.1		1.13 1.13 1.231 2.4213 2.14 2.23 2.24
1	-	42 44		24	12111221411422211222241241	
		1.5	1.3 1.4	2.3	2.3 2.4	133134143144233234243244
			2.0	11	42	0.11112031032411412421423
3	3 4	3.1	3.2	9.1	4.6	212232423232323424234144223424
	2.22	23	.3 3.4	3.4 4.3	12 11	221 222 241 42 431 432 441 442
		4.5			4.4	121324143544413434443444

Fig. 4. Recursive structure of the decomposition process. Every iteration has the same arrangement of filters with (sub)labels denoting 1=low-low, 2=low-high, 3=high-low and 4=high-high subband

In the approaches presented in [18,19] the calculated features were discriminated according to their distributions within a particular channel. Only features that generated bimodal distributions separating melanoma and dysplastic images were kept in the feature set. All other were rejected. In our case the features were not used to control the development of the tree structure signatures. Instead, all 231 coefficients were accepted as potentially significant discriminating signals. Since the number of dermatoscopy images taken to our analysis was limited to only 39 (19 cases of melanoma + 20 cases of dysplastic nevus), and the 231 coefficients affect the classification in a different way, we

tried to pre-select the coefficients by their importance. The selection was twofold and was done in a Matlab toolbox ENTOOL [31]. This toolbox is a statistical learning package that has diverse tools for classification and regression analysis.

First one Ridge regression model with 40 penalty vectors was used to determine an optimal representation of the images in a numeric form. As we remember three ways of extraction from the raster were possible. Ridge regression constructs a linear model $y = \mathbf{X}\beta + \beta_0$ (X-data matrix, y-vector of categories), but instead of minimizing the sum of squared residuals $(y - \mathbf{X}\beta + \beta_0)^T (y - \mathbf{X}\beta + \beta_0)$, it minimizes the regularized loss function (Tikhonov regularization):

$$\operatorname{RSS}_{\text{pen.}} = (y - \mathbf{X}\beta + \beta_0)^T (y - \mathbf{X}\beta + \beta_0) + \lambda \beta^T \beta$$

The additional penalty $\lambda \beta^T \beta$ shrinks the regression coefficients $\hat{\beta}$ towards zero, thereby moderately increasing bias while considerably decreasing variance of the constructed models. The penalty parameter $\lambda \ge 0$ controls the amount of shrinkage and can be used to fine tune the biasvariance tradeoff. For this study, the optimal ridge penalty λ is automatically determined by Leave-One-Out Cross Validation on each training fold individually. To apply ridge regression to a binary classification problem, training outputs were coded as y = 1 (melanoma), y = 0 (dysplastic) and a threshold of 0.5 was applied to discriminate between both classes when doing predictions. Prior to the model construction, input variables were normalized by removing the mean and dividing by the standard deviation for each variable separately.

Once we had chosen the best representation of data (measured in terms of the classification performance) we applied an ensamble of one hundred Ridge models with 60 penalty vectors each, to select the most significant features.

The generalization step of our classification system was done in a MLP neural network The neural network consisted of three layers. The input layer was composed of ten linearoutput neurons with constant unity weights. The hidden layer made of ten neurons and an output layer formed by one neuron had logistic-like activation functions. The topology of the network was subject to tests to determine an optimal configuration for maximizing its performance (both quality of the classifier and minimal training time). At the beginning the weights were set to values randomly chosen from the range of [-1,1] and then modified in the learning process. Evaluation of the features according to their rank was crucial in effective teaching of the neural network. There always is a concern that small statistics of individuals plus too many coefficients (weights) of a single individual could lead to the network overtraining effect. i.e. the perceptron almost memorizes the training set but the generalization error is quite large.

To teach the neural network 17 images of melanoma and 18 images of dysplastic nevus were used each time. The remaining set of two melanomas and two dysplastic nevi were used for testing. There were ten such repartitions so 10-fold cross validation was applied to test the accuracy on the training set. In one training cycle some 150 000 iterations were performed until the network could classify the input with a defined precision.

Results

Accuracy of the Ridge regression and that of the classification model was graphically presented by means of the Receiver Operating Characteristics (ROC) [30]. ROC shows the sensitivity i.e. ratio of correctly identified melanoma lesions in relation to the specificity i.e. ratio of correctly identified dysplastic nevi. Area under ROC (AUC) is a numerical measure of the performance.

At the very first step we determined the optimal procedure of extracting numerical values of pixels from the raster. The criterion was to maximize the classification performance of 20 most significant features. The following values of AUC were obtained: (normal)=85%, (average)=90.8% and (RGB)=93.2%. The RGB model appeared to be the most specific, which suggests that the other two representations reduce the information content relevant for constructing the feature set. This observation allowed us to concentrate on the (RGB)-based feature set. Separate wavelet processing of R, G, B color channels produced three different values of the channel energy. They had to be summed up before the 11 (per itaration) coefficients were calculated.

 Table 1. Ten most significant features of the (RGB)-representation. They were obtained from a statistical analysis in ENTOOL

 [31]. Indices denote the subchannels of the three different decomposition levels

average energy	maximum energy ratio	fractional energy ratio
e _{4.1.2}	$\frac{e_{1.2.2}}{e_{1.2.1}}$	$\frac{e_{2.3.3}}{(e_{2.3.1} + e_{2.3.2} + e_{2.3.4})}$
e _{2.1.1}	$\frac{e_{4.2}}{e_{4.1}}$	$\frac{e_{3.1.1}}{(e_{3.1.2} + e_{3.1.3} + e_{3.1.4})}$
e _{3.3.2}	$\frac{e_{4.3}}{e_{4.1}}$	$\frac{e_{3.1.1}}{(e_{3.1.2} + e_{3.1.3} + e_{3.1.4})}$
e _{3.3.1}	-	-

The second step was limited to the analysis of the most efficient representation. We applied an ensamble of 100 Ridge models with 60 penalty vectors each to the full (RGB) data. Now the number of features was reduced to only ten (out of 231) most discriminating. Those features are presented in Tab.1. For that feature set AUC was increased to 97.4%. ROC is shown in Fig.5.

The selected feature set was used to teach a three-layer back-propagated neural network. The cross validation set of 2 melanoma and 2 dysplastic nevi was shifted between the training runs. The MLP classification result is presented in Fig.6. As we can see, correctly identified lesions are those distributed around 1 on the ordinate for the first 19 indices (images) and around 0 for the other ones. On average two



Fig. 5. Ordinate of ROC is sensitivity, the abscissa is 1-specificity and the threshold values are taken between two successive data vectors. This ROC presents the classification performance of 10 most significant features derived with help of 100 Ridge regression models. AUC is 97.4%



Fig. 6. MLP binary classification of the lesions. The abscissa is the image number (1-19 melanoma, 20-39 dysplastic nevus), the ordinate shows the classification result (iput tags are: 1=melanoma, 0=dysplastic nevus)

dysplastic nevi and two melanoma images were misclassified yielding the sensivity (true positive fraction) of 89.2% and specificity (1-false positive fraction) of 90%. Apparent loss in quality between the performance of the feature set and that of MLP is a price for generalization.



Fig. 7. Distribution of the recognized image classes. Iput tags on the abscissa are: 1=melanoma, 0=dysplastic

Bimodal distribution of the MLP-classified lesions from Fig. 6 is displayed in Fig. 7. Lack of full separability of the two categories exhibits, as the major factor, the generalization error. Yet the inherent fuzzy nature of both feature sets also plays a role. The magnitude of both effects can be estimated by comparing appropriate performance factors above.

Due to statistics of our database (39 images) we decided to perform an experimentation with the oversampling technique. The (RGB) images limited to the pigmented spot of the mole (background-free) were divided into 64x64 blocks of pixels. Each block was decomposed with the Daubechies-3 wavelets for all possible subbands of frequency (3 iterations). When the blocks were treated as individual images the accuracy of the feature set was estimated by AUC=87.8%. Segments grouped to their parent images yielded AUC=92.4%. This method appeared not to significantly improve the accuracy of the classifier and was a burden for our computational resources.

The present accuracy of the classifier is a promising factor for further research in the field. It must be stressed, that dislike in [18] we do not take advantage of specialized instruments and the lesion images are taken in white light. The constructed model of features for melanoma and dysplastic lesions can evolve, after some improvements, to a computer-aided diagnostic system.

Conclusions and Outlook

19 (20) dermatoscopy images of the melanoma (dysplastic) lesions, all confirmed by histologic examinations, were classified using a tree-structured wavelet-based set of features. Discriminant power of those features were determined by the Ridge regression models and generalized in a three-layer back-propagated neural network. The sensitivity and specificity of the classifier amounts to 89.5 and 90%, respectively.

The method presented here and the significance of the results is provided for improvements. The following steps are expected:

 i) Our work is based on a database of anonymous images of the pigmented skin lesions. It is crucial to have a much better statistics for the learning phase sample.

ii) Since the features of melanoma and dysplastic nevus form two joint sets, it would be reasonable (as in [18]) to study soft membership functions to categorize the members of classification.

iii) Within the class melanoma a few subsets of different morphology can be considered. It is not clear now how treating them as separate categories affects the classification process.

iv) Fine tuning of the Ridge regression models and probing the performance of the neural network classifier is possible. All other machine learning methods can be easily tested for the classification performance. Especially the SVM technique is well suited to search for an optimal feature set by maximizing the margin between the classes.

v) Probing different bases of the wavelet decomposition should explain the role of the basis for optimal selection of the feature set.

vi) Evaluation of the method on different sources of skin lesion images should cast a light on the role of image acquisition and quality. Experimentations on issues addressed above will be reported in other publications.

Acknowledgements

We would like to thank dr Piotr Kanas for the database of the dermatoscopic images and fruitful discussions.

References

- Marks R., Epidemiology of melanoma, Clin. Exp. Dermatol. 25: 459-463 (2000).
- Thörn M., Pontén F., Bergström R. et al., Clinical and histopathologic predictors of survival in patients with malignant melanoma: a population-based study in Sweden. J Natl Cancer Inst 86: 761-769 (1994).
- Westerhoff K., McCarthy W.H., Menzies S.W., Increase in the sensitivity for melanoma diagnosis by primary care physicians using skin surface microscopy. Br J Dermatol 143: 1016-1020 (2000).
- Odom R.B., James W.H., Berger T.G., Melanocytic nevi and neoplasms, in: Andrews' Diseases of the Skin, 9th ed., 881-889 Philadelphia (2000).
- Dial W.F., ABCD rule aids in preoperative diagnosis of malignant melanoma, Cosmetic Dermatol. 8: 32-34 (1995).
- Carli P., De Giorgi V., Palli D. et al., Preoperative assessment of melanoma thickness by ABCD score of dermatoscopy, J. Am. Acad. Dermatol., 43:459-466 (2000).
- Johr R.H., Dermatoscopy: alternative melanocytic algorithms

 the ABCD rule of dermatoscopy, Menzies scoring method, and 7-point checklist, Clinics in Dermatology, 20: 240-247 (2002).
- Argenziano G., Fabbrocini G., Carli P. et al., Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. Arch. Dermatol. 134: 1563-70 (1998).
- Żabińska-Płazak E., Wojas-Pelc A., Dyduch G.: Videodermatoscopy in the diagnosis of melanocytic skin lesions, Bio-Algorithms and Med-Systems, 1: 333-338 (2005).
- Carli P., De Giorgi V., Gianotti B. et al., Dermatoscopy and early diagnosis of melanoma. Arch Dermatol 137: 1641-1644 (2001).
- Menzies S.W., Automated epiluminescence microscopy: human vs machine in the diagnosis of melanoma. Arch Dermatol 135: 1538-1540 (1999).
- http://www.dermogenius.com/ http://www.dermamedicalsystems.com/
- Dhawan A.P., Early detection of cutaneous malignant melanoma by three dimensional nevoscopy, Comp. Meth. Prog. Biomed. 21: 59-68 (1985).
- Piccolo D., Smolle J., Argenziano G. et al., Teledermatoscopy – results of a multicentre study on 43 pigmented skin lesions, J. Telemed. Telecare., 6: 132-137 (2000).
- Robinson J.K., Nickoloff B.J., Digital epiluminescence microscopy monitoring of high-risk patients, Arch. Dermatol. 140: 49-56 (2004).
- Piccolo D., Smolle J., Wolf I.H. et al., Face-to-face diagnosis vs telediagnosis of pigmented skin tumors: a teledermatoscopic study. Arch Dermatol. 135: 1467-1471 (1999).
- 17. Paine S., Cockburn J., Noy S. et al., Early detection of skin

cancer: knowledge, perceptions and practices of general practitioners in Victoria. Med. J. Aust. 161: 188-195 (1994).

- Patwardhan S.V., Dai S., Dhawan A.P., Multi-spectral image analysis and classification of melanoma using fuzzy membership based partitions, Computerized Medical Imaging and Graphics, 29: 287-296 (2005).
- Patwardhan S.V., Dhawan A.P., Relue P.A., Classification of melanoma using tree structured wavelet transforms, Computer Methods and Programs in Biomedicine, 72: 223-239 (2003).
- Chang T., Kuo C.C.J., Texture Analysis and Classification with Tree-Structured Wavelet Transform, IEEE Transactions on Image Processing, 2: 429-441 (1993).
- Kadiyala M., DeBrunner V., Effect of wavelet bases in texture classification using a tree structured wavelet transform, 33 Asilomar Conference on Signals, Systems, and Computers, 2: 1292-1296 (1999).
- Mallat S.G., A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, IEEE Transactions on pattern analysis and machine intelligence, 11: 674-693 (1989).
- Wang J.W., Chen C.H., Chien W.M., Tsai C.M., Texture classification using non-separable two-dimensional wavelets, Patt. Rec. Lett., 19: 1225-1234 (1998).

- Kovacevic J., Vatterli M., Nonseparable Multidimensional Perfect Reconstruction Filter Banks and Wavelet Bases for Rⁿ, IEEE Trans. Inf. Theor., 38: 533-555 (1992).
- Mojsilovic A., Popovic M.V., Rackov D.M., On the selection of an optimal wavelet basis for texture characterization, IEEE Transactions on Image Processing, 9: 2043-2050 (2000).
- An Analysis of Wavelet Characteristics in Image Compression, SPIE Conference on Wavelets: Applications in Signal and Image Processing (2003).
- Porter R., Canagarajah N., A Robust Automatic Clustering Scheme for Image Segmentation Using Wavelets, IEEE Transactions on Image Processing 5: 662-665 (1996).
- Daubechies I., Ten Lectures on Wavelets, S.I.A.M., Philadelphia, 1992.
- Numerical Recipes in C. The art of scientific computing, PWN 1999, pp. 591-606.
- Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology, European Journal of Radiology, 27: 88-94 (1998).
- 31. http://zti.if.uj.edu.pl/~merkwirth/entool.htm



Pattern recognition

DSS-MEDA A WEB-BASED FRAMEWORK FOR VIDEO ANNOTATION IN MEDICAL E-LEARNING

DSS-MEDA SIECIOWE ŚRODOWISKO ADNOTACJI MATERIAŁÓW WIDEO NA POTRZEBY E-NAUCZANIA MEDYCZNEGO

PAWEŁ MROZOWSKI, ANDRZEJ A. KONONOWICZ

Department of Bioinformatics and Telemedicine, Faculty of Medicine, Jagiellonian University, Krakow, Poland

Abstract: In this paper a video annotation system is presented, called DSS-MEDA, used for describing medical videos. The aim of this framework is to provide physicians with a tool for annotating video recordings of medical procedures destined for e-learning. The system manages media files, which are streamed by the Darwin Streaming Server. DSS-MEDA enables controlled access to video resources depending on roles assigned to users. The annotations can be of two kinds: single lined subtitles or complex, hierarchical descriptions in HTML format. The long-term goal of this study is to create a large multimedia database of medical videos.

Keywords: medical video annotation, content-based video retrieval, e-learning

1. Introduction

The amount of data sent and received by an average user of the Internet has multiplied several times in the last few years. Increasing Internet bandwidth and the availability of many network friendly multimedia codecs made it possible to develop new techniques of multimedia content distribution. High speed connections enable users to receive all kinds of multimedia data: text, images, sound and even high resolution video in a reasonable time. Additionally, a shorter transmission time needed for good quality multimedia data is obtained due to improved compression algorithms.

It is believed that the percent of understanding and memorizing information is better for materials in which multimedia is skillfully used, in comparison to approaches based on pure text only. The combination of network technologies, multimedia, and appropriate learning strategies may significantly contribute to better learning results. In order for multimedia e-learning systems to be successful, easy to use authoring tools are needed. This article describes one of such tools, named DSS-MEDA (Darwin Streaming Server Streszczenie: Praca prezentuje system DSS-MEDA służący do udostępniania przez Internet materiałów wideo z zabiegów medycznych oraz tworzenia ich opisu na potrzeby e-nauczania. Przechowywane w bazie danych filmy strumieniowane są za pomocą serwera Darwin Streaming Server. Dostęp do zasobów udzielany jest na podstawie systemu ról użytkowników. Opis materiałów wideo może być dwojakiego rodzaju: jednowierszowe podpisy oraz hierarchiczny opis w formacie HTML. Przedstawiony system jest częścią większego projektu mającego na celu stworzenia dużej medycznej bazy danych przechowującej edukacyjne materiały wideo.

Słowa kluczowe: adnotacja medycznych materiałów wideo, wyszukiwanie wideo, e-nauczanie

Medical Application), which facilities the annotation of videos for medical e-learning purposes. Combined with the Darwin Streaming Server and the MySQL database, DSS-MEDA forms a framework for storing, describing and presenting multimedia files on the Internet. The paper is organized as follows: Section 2 introduces the technical background of this work, including a short overview of the available streaming and annotation technologies. In section 3 we discuss the possible application of annotation technologies in medicine. Section 4 presents the architecture of the DSS-MEDA system. Section 5 contains a brief discussion. We conclude in section 6.

2. Background

Multimedia streaming technologies

The term *streaming* refers to the transmission of multimedia data like audio, video and any multimedia add-ons, by using a specialized server program, which prepares data by encapsulating it into a data stream, sends the stream in real-time to the recipient and controls the transmission process. Streaming technology is used in two modes: on-demand and webcast. On-demand means, that the user gains access to pre-recorded and stored data. This method is useful for the distribution of archive data. Webcast is a different approach in which data is streamed while the recording is taking place. The clients connected to the server receive the streamed data simultaneously. A simple analogy of this technology is TV broadcasting. Before sending multimedia to the network (webcast), the data needs to be processed. A slight delay between real-time events and creation of a data stream is required in which the data is compressed, packetized and buffered. After a while, which usually takes a few seconds (depending on the processing power of the compression system), the data is ready to be sent. For the purpose of the streaming technology specialized communication protocols were created. The most popular and widespread are: RTSP - Real Time Streaming Protocol (control, analyze of link state) [16][18] and RTP [17] - Real Time Transport Protocol (transport of data). The transportation layer of these protocols can be based upon TCP or UDP.

The provision of streamed data may occur in three different ways, depending on the number of recipients and technical conditions. The first type is called unicast and is a separate logical link creation from the data sender to the recipient. The advantage of this solution is its simplicity. The disadvantage is that each connection is treated as a separate one, and finally all connections summarize to the bandwidth of the network, which may impact on limited number of simultaneous clients. The latter mentioned limitation does not refer to multicast - the second type of data transmission. Multicast means sending the same data simultaneously to many recipients. There is only one stream of data, but each of the connected clients receives a copy of the data stream. This type of connection is also called point-to-multipoint. The third type of data distribution is broadcast in which the streamed data will be received by every device on the network. This method is not commonly used, because this kind of traffic is blocked by broadcast domains set by routers.

The term *architecture of streaming multimedia system* applies to a set of elements e.g. servers, clients, codecs, that combined create a multimedia providing solution. At the present moment, four main solutions are available on the market. These are: Real Media (Real Networks)[14], Windows Media (Microsoft)[21], Flash Media (Adobe/Macromedia) [11] and Quick Time (Apple Computers) [10].

The media server of the Real Media architecture is named Helix. It is available as a commercial or free version. Other products in the Real Media suite are e.g. Real One Player, and Helix Producer. The RealMedia format can store multimedia data in good quality with only 160 kbit/s, which is twice as compressed as MPEG-4.

The Windows Media (WM) architecture was developed by Microsoft. It consists of WM Encoder, WM Rights Manager, WM Services and WM Player. The transmission receivers can be end points (Players) or transit servers. The second option can be used as proxy server and data store. The streaming data can be prepared with different Microsoft tools, like Microsoft Producer for PowerPoint or Windows Movie Maker. The WM server is included in some versions of Microsoft Windows operating system: 2003 Std. Ed, Enterprise Ed., DataCenter Ed. The supported protocols of data exchange include RTSP and native Microsoft MMST, MMSU, MSBD.

Flash Media Server 2 is the main product offered by Macromedia to support multimedia streaming. The server enables users to develop new types of services e.g. multimedia chat, video blogs. The player for flash media is distributed as a plug-in for web browsers. The advantage of this solution is the platform independence. It also contributes to constant appearance of the player.

The Quick Time (QT) architecture originates from the Apple Macintosh computers, but it is now also available on other platforms (including PC). It consists of QT Player or QT Player Pro, QT Server, Darwin Streaming Server (DSS) and other software as e.g. the QT Broadcaster. The QT Server is a commercial product, developed for Mac OS X, whereas the Darwin Streaming Server is a multiplatform and open source solution. The main difference between the commercial and the free version is the presence of a module for administration of served data in the first one.

Video annotation technologies

The main reason of video annotations is to facilitate multimedia database searches. Automatic methods of content-based video data retrieval are not very successful because of the semantic gap problem. Even if the computer algorithms are able to automatically recognize a shape in a video, it is hard if not impossible, to decide without human intervention about the meaning of this object. Therefore manual data annotations are required which provide the missing semantics. The metadata (i.e. data about data) can consist of keywords from controlled vocabularies or free text descriptions. Keyword annotations combined with a proper ontology (i.e. formal model of the knowledge domain) enable the integration of artificial intelligence methods into search algorithms [7]. The use of free text complicates the search algorithms, but on the other hand, this method is often more convenient for those who are in charge of describing the multimedia resources. This kind of description is suitable for users who have no precise query concepts in mind and intend only to browse the database [3].

Several annotation specifications for creating metadata of multimedia files are already available. Among them are Dublin Core [19], RDF [15] and MPEG-7 [8][9]. Examples of systems using these standards can be found in [2][4][6]. Unfortunately, despite their great potential, the application of these technologies is still in experimental stages and is not very widespread.

3. Using video annotations in medical e-learning

Video annotations are used in many different areas. An important field of application is for instance linguistics where annotations are used in creating multimedia corpora (i.e. collections of written and spoken language stored for language research) [1][5].

Medicine is also an area which can benefit much from annotation technologies. The most obvious application example is medical e-learning. Students of medicine are always willing to learn on real world cases. Unfortunately, the great number of students makes it impossible for them to participate in all medical procedures. The sterility and privacy requirements and other (e.g. psychological) reasons were always a big barrier for students in accessing diagnostic or operation rooms. Video recordings of medical procedures have the potential to remedy this situation. After obtaining permission of the physician and patient to use the recording for educational purposes, the material should be made anonymous and then stored in a video database. Since uncommented videos are not very informative, the recordings could be annotated by physicians. Access to such content would be granted only to authorized students to avoid malpractices. Students gain by that a very attractive learning resource, which could be accessed almost everywhere and every time. In order to realize the outlined scenario a proper IT infrastructure is needed.

4. DSS-MEDA

The aim of the DSS-MEDA system is to provide physicians with a tool for annotating video recordings of medical procedures destined for e-learning and the students with a mean of browsing these materials. In the following section we outline the general architecture of the DSS-MEDA system and present some more detailed insights into its selected components.

Architecture

The architecture of the DSS-MEDA framework is presented in Fig.1. The system is composed of three main components: streaming server, database server and web application.

It has been decided to use the Darwin Streaming Server (DSS) for the role of the streaming server due to its openness, multi-platform availability, good performance results and stability. The shortcomings of DSS in comparison to its commercial counterpart QT Streaming Server (e.g. the missing user management tool) are to some extend compensated by features built into the DSS-MEDA web application. The role of the second architectural element is to store user data, access rights and video annotations. For this function we has been chosen the MySql database. The rationale for this decision is the servers' popularity, user-friendliness and its performance which is sufficient for our needs.

The center of the system is the third architectural element – the DSS-MEDA web application. The application has been implemented it in Java technology because of its portability, good performance results (in comparison to e.g. CGI scripts), and large choice of additional opensource libraries and tools. The server side of the application was implemented in the JSP (Java Server Page) technology and deployed on the Apache Tomcat servlet container. The client side requires a JavaScript enabled web browser with installed QuickTime plug-in.

User rights and roles

Three types of user roles have been distinguished: administrator, developer and user. A role grants specific rights to manage, annotate and perform media files.

Administrators can insert, update and delete information about users and videos in the database. This role allows setting access rights to media files. The administrator's functions are executed from within an administration panel (Fig.2.).

Dodawanie fil	mu do bazy danych:
Torakoskopa	tytul (wymagane)
przykładowy inkarz	twórca
9 mag 💙 2006	dzien-miesiac-rok
zahing wewsqirzidatanwy	krotki opis
tailing v	typ
torakosikopia mov	nazwa pliku (wymagane)
root 🛩	właściciel filmu
lonkoskopa,thoracoscopy	słowa kłucze (oddzielane ".")
Wirrowadz do bazy	Kami date

Fig. 2. DSS-MEDA Administrator's Panel – Adding a movie to the database

Developers can divide video clips into segments and describe them using the developer's panel (Fig. 3-4, the thoracoscopy video used in the example was taken from [20]).





Fig. 3. DSS-MEDA Developer's Panel – Adding subtitles



Fig. 4. DSS-MEDA Developer's Panel – Adding hierarchical annotations in HTML format

Owners of the role *User* can browse and play movies (Fig.5).



Fig. 5. DSS-MEDA User's Panel - Browsing video database

The roles may cumulate – i.e. a single user may have three types of privileges at the same time. Users with the *developer* role gain automatically *user* rights. The *developer's* and *user's* rights refer only to a single record in the database. It means that to create descriptions for other videos, proper access privileges must be granted.

Creating annotations

Video annotations in DSS-MEDA are of two kinds: short text subtitles linked to individual seconds in the movie (Fig. 3) and larger descriptions in HTML format referring to video segments (Fig. 4). The video segments are arranged in a 3 level hierarchy. The first level elements are called *sequences*, the following accordingly *scenes* and *frames*. Segments from different levels must be properly nested (e.g. if a sequence lasts from the 1. to the 45. second a nested

scene is limited by the segment's time interval and cannot intersect its boundaries).

Subtitles are entered by a plain text field. The content of this field is synchronized with the video playback by a JavaScript function using the QuickTime's JavaScript API [13]. HTML descriptions may be entered through a text form or using the Ekit editor [12]. Ekit is a simple open source HTML editor embedded into web pages as a Java applet (Fig. 6). It provides the user with basic commands like text editing, images insertion, table creation and list enumeration. It works in the WYSIWYG mode which means that the content of the editor during editing is similar to the final result.

File Edit View Font Format Search Insert Table Forms Help □ □ □ □ □ π >> B z u S A² E E I □ □ π <> (none) ✓ Amerika ✓ I <td< th=""><th colspan="7">👙 Ekit 📃 🗆 🔀</th></td<>	👙 Ekit 📃 🗆 🔀						
Elyit 11 ©2000-2006 heнidec codeн wheridec bodex NEW in 1.1 PARAM-based interface configs Images from URLs Seamlessly add pictures from the internet that will work anywhere New Translations Swedish, Polish and Ukrainian	File Edit View Font Fo	rmat Search Insert Table Forms Help B A2 E E E E E E E E E E E E E					
PARAM-based interface configs Specify menus & toolbars in applet mode from custom PARAM tag Images from URLs Seamlessly add pictures from the internet that will work anywhere New Translations Swedish, Polish and Ukrainian	Ehit 11 ©2000-2006 hexidec codex						
Images from URLs Seamlessly add pictures from the internet that will work anywhere New Translations Swedish, Polish and Ukrainian	PARAM-based interface configs Specify menus & toolbars in applet mode from custom PARAM tag						
New Translations Swedish, Polish and Ukrainian	Images from URLs Seamlessly add pictures from the internet that will work anywhere						
J	New Translations Swedish, Polish and Ukrainian						

Fig. 6. Ekit HTML editor

Once the edition is finished, the applet generates a HTML page and encodes it in BASE64 coding. Afterwards the coded information is stored in the database. It is also possible to insert images into the description. The images are uploaded by a servlet (Fig. 7). Users may browse the uploaded files. To minimize the download time of full images while browsing, their miniatures are created on the fly.

Fig. 7. Uploading images to the server

It has been decided to store the descriptions directly into the database without formatting it in a dedicated annotation standard as e.g. RDF or MPEG-7. However, if there is a need to exchange the annotated data with another system it is intended to implement an appropriate export mechanism.

5. Discussion

The system has not been deployed under real conditions yet, therefore we are not able to report any user experiences in using it. Nevertheless we are aware that the user interface needs some refinements to become more intuitive. Some of the flaws in the application ergonomics stem from the limitations of web applications in general. However, thanks to new emerging technologies – like AJAX – we hope to improve. We are willing to collaborate with physicians who are interested in publishing their videos in the way it has been described in the paper.

6. Summary

It is hoped that the presented system will contribute in building up the medical e-learning infrastructure. By DSS-MEDA we offer the physicians a tool for annotation of medical videos stored in a multimedia database. Students may learn by browsing these resources. The advantages of our solution are low runtime requirements (only a JavaScript enabled web browser with QuickTime plug-in installed), flexible user access control mechanisms, diverse annotation types and very low cost, since the system has been based our system on free components only.

Bibliography

- H. Brugman, A. Russel, D. Broeder, P. Wittenburg, *EUDICO*. Annotation and Exploitation of Multi Media Corpora, in: Proceedings of LREC 2000 Workshop, 2000.
- S. Carro, J. Scharcanski, A framework for medical visual information exchange on the WEB, Comp. Biol. Med. 36 (2006) pp. 327-338.
- 3. J. Fan, H. Luo, A.K. Elmagarmid, Concept-oriented indexing of video databases: toward semantic sensitive retrieval and

browsing, Image Processing, IEEE Transactions on, 13(7), (2004), pp. 974-992.

- J. Hunter, R. Iannella, *The Application of Metadata Standards to Video Indexing*, Proc. of the 2nd Europ. Conf. on Research and Advanced Technology for Digital Libraries, Springer Verlag London, UK, (1998) pp. 135-156.
- M. Kipp, Anvil A Generic Annotation Tool for Multimodal Dialogue in: Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech), Aalborg, (2001) pp. 1367-1370
- A.A. Kononowicz, Z. Wiśniowski, MPEG-7 as a metadata standard for indexing of surgery videos in medical e-learning, not published yet.
- H. Luo, J. Fan, *Building concept ontology for medical video* annotation, in: Proceedings of the 14th annual ACM international conference on Multimedia table of contents Santa Barbara, CA, USA, (2006), pp. 57-60
- J. Martínez, R. Koenen, F. Pereira, MPEG-7: The Generic Multimedia Content Description Standard, Part 1, IEEE MultiMedia, 9 (2) (2002) pp. 78-87.
- 9. J. Martínez, *MPEG-7: Overview of MPEG-7 Description Tools, Part 2*, IEEE MultiMedia, 9(3), (2002), pp. 83-93.
- 10. Apple Quick Time http://www.apple.com/quicktime (28.12.2006).
- Flash Media Server http://www.adobe.com/products/flashmediaserver/ (28.12.2006).
- Hexidec Codex Ekit http://www.hexidec.com/ekit.php (28.12.2006).
- JavaScript Scripting Guide for QuickTime, Apple Computer, Inc. 2004-2005.
- 14. RealNetworks http://www.realnetworks.com/ (28.12.2006).
- Resource Description Framework http://www.w3.org/RDF/ (29.12.2006).
- RFC 2326 Real Time Streaming Protocol, http://www.ietf.org/ rfc/rfc2326.txt (28.12.2006).
- RFC 3550 RTP: A Transport Protocol for Real-Time Applications http://www.ietf.org/rfc/rfc3550.txt (28.12.2006).
- RTSP Real Time Streaming Protocol Information and Updates, http://www.rtsp.org (28.12.2006).
- The Dublin Core Metadata Initiative http://dublincore.org/ (29.12.2006).
- Video Assisted Thoracoscopy, University of Southern California, http://www.cts.usc.edu/videos-mpeg-vidassistthoracoscopyall.html.
- Windows Media http://www.microsoft.com/windows/windowsmedia/default.mspx (28.12.2006).



Felemedicin

INSTRUCTION FOR AUTHORS

- 1. The original and one Photostat copy of the manuscript should be mailed to: Managing Editor Zdzislaw Wisniowski. Authors are strongly urged to submit a CD or floppies containing the manuscript in Word for PCs format along with the hard copies. Articles already published or those under consideration for publication in other journals or periodicals should not be submitted.
- 2. Manuscripts should be in English. Usage of correct language is the responsibility of the author.
- 3. Submission of manuscripts to us implies that the copyright of the entire material except those already copyrighted by other publications like tables, figures etc and those permission the author(s) have already obtained, belongs to the publisher.
 - PRESENTATION
- 4. The text may be prepared from a good quality laser printer or electronic typewriter. It is to be typed (one side only) on a regular sheet of paper A4 size preferred with single line spacing. New Times Roman font of 12 point size is recommended. Authors are requested to paste figures and tables (if possible) at appropriate places in the text. If this is not possible, they may be given separately, in which case their exact locations in the text may be marked by pencil.
- 5. The article should begin with a short ABSTRACT followed by a short INTRODUCTION. The rest of the article ma be titled and arranged as per the wishes of the author.
- 6. REFERENCES should continue the last section of the article. In the text, references to other papers or books should be cited using consecutive numbers in parenthesis (e.g. [1,2]) and they should be listed numerically in the last section.

Examples:

I. Pyrczak W., Sarapata K.: Instructions for authors: references. Bio-Alghoritms and Med-Systems, 1, 386, 2005 II. Tadeusiewicz R., Ogiela M.R.: Medical Image Understanding Technology. Springer Verlag, Heidelberg 2004 Authors are to make sure that each reference appearing in the text appears in the list of references at the end and vice versa.

- 7. TITLE PAGE AND FIRST PAGE The title page of the manuscript, the name(s) of author(s) and institution(s) where the work has been carried out should be typed out on a separate sheet of paper.
- 8. Section headings (ABSTRACT, INTRODUCTION etc.) are to be typed in upper case letters and placed on a separate line
- 1. Scientific names should be typed in italics or underlined
- 2. Tables should be numbered consecutively using Arabic numerals.
- 3. Illustrations should be limited to materials essential for the text. All figures should be numbered consecutively and should be submitted as sharp, high-quality prints. Figures that are to appear as a group should be photographed and mounted together. All figures and groups of figures should be trimmed at right angles and pasted directly on the typescript in the appropriate places. They should be of a size permitting photographic reduction by 25% together with text (not more than 15 cm across for page width and not higher than 20 cm). Legends to figures should appear directly below the respective figures.
- 4. In line drawings, all lines should be of uniform thickness; letters and numbers should be of professional quality and proper dimensions, approximately 1mm high when reproduced.
- 5. RUNNING TITLE Please make sure to mention the running title (short title) of the article separately in the first page where the title, names of authors, their affiliations, etc. are given.
- 6. PREFFERED LENGTH The length of the manuscript may be limited to 20-25 A4 size pages (20 cm width 26 cms height) through this is not binding.
- 7. PERMISSION Authors must obtain permission from other publishers if they use already published and copywrighted materials.
- COLOUR FIGURES Cost of printing color figures will be charged to the author at the rate of U.S. \$ 125 per plate of four color pages.
- 9. REPRINTS The order for reprints should be placed at the time of submission of the article itself, so that it may be out for printing without loss of time, if found suitable for order for not less than 100 reprints is to be made.

MANUSCRIPT SUBMISSION:

- 1. Electronic the <u>www.bams.cm-uj.krakow.pl</u> address is available for all authors to submit manuscript. The instruction is available at each step of submission process.
- 2. The CD with pdf or doc file with figures and tables incorporated into the manuscript body can be sent by email address: bams@cm-uj.krakow.pl
- 3. The CD (as above) can be also sent by post. Address: *Bio-Algorithms and Med-Systems*, 31-530 Krakow, Św. Łazarza 16, POLAND

CLINICAL DATABASE – NEONATOLOGY

KLINICZNE BAZY DANYCH – NEONATOLOGIA

Krzysztof Sarapata¹, Joanna Woźniak², Piotr Walecki³, Wojciech Kieczeń⁴, Ryszard Lauterbach²

^{1,3,4} Department of Bioinformatics and Telemedicine, Collegium Medicum Jagiellonian University
 ² Clinic of Neonatology, Collegium Medicum Jagiellonian University
 ⁴ The Faculty of Physics, Astronomy and Applied Computer Science Jagiellonian University

Abstract: This paper presents some aspects of the informatics implementation of the Neonatology Database, which was created in the Department of Bioinformatics and Telemedicine Collegium Medicum Jagiellonian Univ. in association with The Neonatology Clinic. The significance of this database for interdisciplinary collaboration is presented.

The database is mainly utilized to gather hospital data in form of accessing and processing. The main assumptions of implementation are: universal code to use it by other clinics' databases, a centralized location of the database server and minimize client computer required. Programming languages (PHP), system server (Linux) and projects (Apache, MySQL, openSSL) are on GNU GPL license¹.

Key words: medical data bases, neonatology, programming languages

Streszczenie: Praca opisuje wybrane aspekty implementacji bazy danych 'Neonatologia', która była zrealizowana w Zakładzie Bioinformatyki i Telemedycyny Collegium Medicum Uniwersytetu Jagiellońskiego we współpracy z Kliniką Neonatologii. Praca stanowi przykład interdyscyplinarnej współpracy podczas realizacji systemu bazodanowego.

Kliniczna baza danych przeznaczona jest zasadniczo do zbierania i przechowywania odpowiednio sformatowanych danych, umożliwia do nich dostęp i ich przetwarzanie. Głównym założeniem implementacji jest utworzenie szkieletu programistycznego (*framework*) systemu bazodanowego w celu ponownego jego wykorzystania na użytek innych klinik. Kolejne założenia głównie dotyczą uwarunkowań ekonomicznych: centralizacja położenia serwera i minimalizacja wymogów klienta-użytkownika systemu, wykorzystane języki programowania (PHP), system serwera (Linux) i serwery usług (Apache, MySQL, openSSL), które są na licencji GNU GPL.

Słowa kluczowe: medyczne bazy danych, neonatologia, języki programowania



Fig. 1. Web front-end database system – example 'Neonatology database'

Introduction

The database systems for any clinic have been produced in our Department of Bioinformatics and Telemedicine. Our intention is to create a universal environment (framework) for database creation and management.

Typical client-server architecture of computer networks has been implemented. This system is applicable for many clients' personal computers with internet browsers and centralized servers with WWW service. The only task for the client is to enter his or her patients' clinical data in a unified form. The database then becomes ready for permanent storage. The tasks for the client and the sever are distinct. The access to the database can be sheared among many users. We applied Apache server² with openSSL module³ as WWW service with HTTPS protocol, PHP language⁴ to dynamic controls content of WWW and database MySQL⁵ – appropriate for our application. Data storage formats is SQL. We are responsible to ensure proper security standard. Figure 1 presents the view of user database interface.

Implementation focus

Database Modeling

The term Data Modeling means a method mapping the database concepts of tables and relationships. This method is different from a basic design model (like data normalization⁶ – *is the process of restructuring the logical data model of a database to eliminate redundancy, organize data efficiently, and reduce the potential for anomalies during data operations*). Data Modeling allows analysis of the client's requirements and models database keys, triggers, constraints, referential integrity (RI) and other relational database features. The purpose of data modeling is to develop an accurate model, or graphical representation.

The steps in application development are:

- 1. to gather information and resources to determine the shape and form of the information to be processed
- 2. to utilize the Conceptual Data Model (CDM) : an entityrelationship model, Conceptual Data Model
- 3. Physical Data Model (PDM)
 - a) primary keys
 - b) foreign keys
 - c) unique columns
 - d) types of data

To create Data Modeling with so called entity-relationship (ER) we obtain Conceptual Data Model. ER model, originally proposed by Chen⁷, it consists of:

- entities that represent real-world objects
- relationships that capture associations among entities
 attributes that represent union properties of entities and relationships
- the degree of the relationship that means the number of entities

There are some diagrams noting the entity-relationship. One of them is the Merise⁸ notation (MERISE = Méthode d'Études et de Réalisation Informatique des. Systèmes d'Entreprises). Figure 2 presents part of the Conceptual Data Model of our database system.



Fig. 2. Part of entity-relationship diagram Neonatology database

The Physician Data Model was applied on the basis of a preview model utilizing table 1. Figure 3 presents that result. The next figure (fig. 4) presents view of structure database.

Tab. 1. Conversion CDM to PDM

Relationships	entity A		entity B			
(1)1:n (0n)	PRIMARY KEY		FOREIGN KEY xxx REFERENCES entity A(PRIMARY KEY) - xxx NOT NULL			
(0,1)1:n (0n)	PRIMARY KEY		 FOREIGN KEY xxx REFERENCES entity A(PRIMARY KEY) xxx NULL 			
n:m		entity A	_B			
	PRIMARY KEY	 PRIM yyy) FORI REFI A(PR FORI REFI B(PR 	VARY KEY (xxx, PRIMARY K EEIGN KEY xxx ERENCES entity RIMARY KEY) ERENCES entity RIMARY KEY)			
(1)1:1(1)	PRIMARY KEY		 FOREIGN KEY 3 REFERENCES (A(PRIMARY KE' - xxx NOT NULL UNIQUE KEY xx 	cxx entity Y) x		
(0,1)1:1(0,1)	PRIMARY KEY		 FOREIGN KEY 3 REFERENCES 6 A(PRIMARY KE' xxx NULL UNIQUE KEY xx 	xx entity Y) x		

A dynamic front-end web

A dynamic front-end web is generated by parser PHP language. PHP 5 has been implemented, which offers plenty of Objected Oriented capabilities. PHP 5 introduces the Standard PHP Library (SPL), which provides a number of ready-made classes and interfaces. For this reason PHP 5 can be compared with Java, the most popular OO language. PHP as a hybrid language can use objects or revert to procedural programming that was the dominant method in PHP 4.

We configured PHP with PHP Data Objects (PDO)⁹ which provides a *data-access* abstraction layer, which also means that, regardless of which database is used, you use the same functions to issue queries and fetch data. Moreover we use part of PEAR'a library¹⁰: *Spreadsheet_Excel_Writer*. A dynamic front-end web offers:

- configuration (setup) of own access to database
- input, edit, update of patient records
- review of patients records
- download stored data in .xls format file
- generated reports (RTF format) *Cart clinical treatment* An architecture for building applications the most popular Model View Controler (MVC)¹¹ allows to separate the

lar Model View Controler (MVC)¹¹ allows to separate the data (model) form user interface (view) and the processing (controller) form. We realize we do not have such strict architecture in our project, because the degree of advancing did not require that. The exception relates to the realizing of HTML form and their contents which are processed in the same steps.



Fig. 3. Part of database system - physical model



Fig. 4. Database structure



Fig. 5. Activity diagram - Patient Module

Figure 5 presents a necessary diagram (Activity diagram 'Patient Module') to realize interface with OOP language. An activity diagram is one method of specifying dynamic behavior of a model. It presents activities which are carried out by humans and the path of transitions between activities. It allows us to know the conditions and relationship of moving to another activity.

Authentication and authorisation system

A web browser and a web server are localized in the global computer network – Internet. For this reason the paper will focus on the security system of the data transfer process and the collection of medical data. Public Key Infrastructure (PKI)¹² and his standard X.509¹³ which de-

scribes providing and managing authentication using asymmetric cryptographic were applied. X.509 assumes a strict hierarchical system of certificate authorities (CAs) for issuing the certificates. This contrasts with a web of trust models, like PGP, where anyone (not just special CAs) may sign, and thus attest to the validity of others' key certificates. A web browser and server with SSL provide two X.509 based authentication mechanisms:

- Client Authentication, where a client application is authenticated by a server itself.
- Server Authentication, where the identity of a server is authenticated by a client.

A client application, specifically a web browser, manages a database of 'Trusted Root Certification Authorities' and 'Personal Certification'. An organization's trusted root certificates can be distributed to all employees so that they can use the company PKI system. Browsers such as Mozilla Firefox, Internet Explorer, Opera and others come with root certificates pre-installed, so SSL certificates from larger vendors who have paid for the privilege of being pre-installed will work instantly; in effect the browsers' owners determine which CAs are trusted third parties for the browsers' users. Although these root certificates can be removed or disabled, users rarely do so. In our case, a file in PKCS#12 format allows our CA place in correct store except during personal key installation process. Each of them obtain PKCS file with full need components (private and public key, certificate CA) and can install it in the web browser.

Polish law regulations (for electronic signatures by the Act of 18 September, 2001 on Electronic Signature) were introduced on August 16, 2002. We were also determined to create our own CAroot for internal use. Our CA is given below: Part of user certificate (.pem file)

Certificate:

Data:

Signature Algorithm: md5WithRSAEncryption

Issuer: C=PL, ST=Malopolska, L=Cracow, O=Jagiellonian Univ., OU=Department of Telemedicine, CN=Department of

Telemedicine CA/emailAddress=mysarapa@cyf-kr.edu.pl Validity

Not Before: Dec 13 06:57:02 2006 GMT

Not After : Dec 16 06:57:02 2006 GMT

Subject: C=PL, ST=Malopolska, L=Cracow, O=JUMC, OU=ZBiT, CN=Example/emailAddress=xxx@xxx.xx

Subject Public Key Info:

Public Key Algorithm: rsaEncryption RSA Public Key: (1024 bit)

Part of CA certificate (.pem file):

Data:

Signature Algorithm: md5WithRSAEncryption

Issuer: C=PL, ST=Malopolska, L=Cracow, O=Jagiellonian Univ., OU=Department of Telemedicine, CN=Department of Telemedicine CA/emailAddress=mysarapa@cyf-kr.edu.pl

Validity

Not Before: Feb 4 00:25:17 2006 GMT Not After : Feb 4 00:25:17 2007 GMT

Subject: C=PL, ST=Malopolska, L=Cracow, O=Jagiellonian Univ., OU=Department of Telemedicine, CN=*.xxx.cmuj.krakow.pl/emailAddress=mysarapa@cyf-kr.edu.pl

The user's authorization based on record in table 'users.bases'.



Fig. 6. Structure table users.bases

There are specifications in 'grants' column which set general user's privileges and 'mask' allows the application of particular constraints within just one database. Independently, the user is informed to register obligatory $^{14}. \label{eq:14}$

Semantic controls

A lot of input variables from HTML form and their complexity links was a motivator to find a universal tool to realize semantic controls data in base. The syntactic controls data was simply accomplished using regular expression engine and using same automatic operations checking data on the level of their comparison (quantity or quality, equality-inequality like: less than, more then, etc.). Semantic controls and links are able to be described by any self created language in PHP and next parse them. The specification of this language is simple. It controls the specific data and checks condition for corrected value of data:

\$uniques=array(«pacjent_id data_morfologia», «pacjent_id data_jonogram», «pacjent_id data_bilirubina», «pacjent_id data_pr_watrobowe»);

\$conditions[«doba_tamponada»]=(«powiklania= tamponada_osierdzia»);

To realize semantic control, a so-called *parser generator* was needed (primitive in our implementation). A parser generator reads strings from a file and includes a description of semantic control variables and then analyzes them and by *eval()* function returns the result to the program.

SQL allows on the levels database to control specificity of any data by use UNIQUE KEY() on columns. Moreover in MySQL 5 Stored Procedure¹⁵ was implemented that includes data validation, integrated into the database structure. Stored procedures used for this purpose are often called triggers.



Figure 7 shows that the PHP level gives the best controls entered data. It's with its association with PHP that the server's language offers the biggest possibility of verifying data by its logical limitations, the possibility to access the database or integrate the data if necessary¹⁶ via connection and uses other dictionaries or data sources of other localization. Data integration is the problem of combining data residing at different sources and tries to create a unified view of these data.

Simultaneous database connection

Parallel access is a basic feature of databases. At the database level are some mechanism of resolving problems with simultaneous access and updating of records in tables. LOCK TABLE ¹⁷ one is problematic to use because it locks base tables (but not views) for the current thread. The second one - transactions - are useful in the case of realizing a sequence SQL queries that must be treated as one indivisible unit. (In MySQL 4 Innodb type of database is supported for transactional processing). We use transactions when an UPDATE of a patient's record is made on some (no one) tables. The primary problem is solved by editing the same record by the same users. We have to lock the record when another user tries to open it for editing. If the record was already opened by one user, it would tell the second user trying to access the same record that it was already being edited. To realize it, each table has a lock time column, where a timestamp is given. The global parameter 'locktime' (in /config/parameters.php file) determines the length of time for a user's edition of a particular record. Users distinction is relayed on session id (the same user cannot edit the same record on some client platform). All of those mechanisms were applied in PHP language.

The utilization in scientific methods

The clinical databases are a source of information for formal research hypotheses and science verification tools. The database is a form of informatics implementation to utilize: storing, gathering and reading data which are in electronic forms. The data is in symbolic form to write specific relationships physics property or natural phenomenon.

The database has a specific structure (it's distinguish hierarchic, web, relation, object and other structures of database), that is choose as the best model of data and relation among them. The databases in clinical activity meet two goals: pragmatic for health centre functionality and science to define science problems.

A practical diagram in scientific methods

Medical dissertations always include a statistical study of clinical data during creation or verification of scientific hypothesis. If it is not included it is theoretical, comparative or methodology work. The data for the statistical procedure originates from database as text (value, description) data, sound data, image (picture, video) data in the correct format file representation of collections of data (Fig. 8.).

The blue color of boxes means region of utility database (e.g. clinical database) in the scientific procedure. 'Gather information and resources' exposed practice and experience of physician and based on clinical database. The verification of hypothesis – 'Perform experiment and collected data' region is realized by e.g. *blind experiment* can use clinical data.

Neonatology database case study

The creators and designers have a difficult problem with the proper collection of data that will be gathered and stored.



Fig. 8. Scientific procedure¹⁸

The database gathering all possible clinical data and relations among them is impossible to realize. The obstacles are the large quantity of variables, their changeability in time domain (as e.g. periodical cycle) or any reason domain. As a result of having so many variables, a lot of measure values for each of them (frequently measured) are obtained. To make matters more complicated, as a rule, almost all of them are obtained by through statistical procedures.

The scientific experience allows reduction of number of positions in the record and their proper collection. The selection of variables is a compromise between objectivity – respect all data and subjectivity only them tail.

There is no technical constraint on the number of variables, the number of records, or data formats. Actual computers systems are scaleable and realize simultaneous database connections. There is no problem with the bandwidth network during transfer of AV files. Fundamental causes are much work to input data into database realized by hospital staff.

The constraints of statistical tools and statistical analysis, aren't crucial. There is a possibility to analyze 'ever' amount data records.

Preterm born infants accounts for about 7% of all live births. The percentage of newborns born with extremely low birth weight (<1250 g) amounts to 5%. About 160 newborns from the above mentioned group were born in the Neonatology Clinic in Krakow during last year. These patients require hospitalization at the neonatal intensive care unit (NICU), and what is more, they are the most demanding group. This is caused by specificity of diseases characteristic during this particular period of life. There exist numerous causes of preterm labour, risk factors of prematurity, clinical conditions which correlate with low birth weight. The interpretation of clinical data following EBM rules might be troublesome. Furthermore, the assessment of effectiveness of therapy is difficult, as the group has a low number of patients. That is the reason why collecting information should involve other clinics encountering similar problems. It is necessary to analyse results and data in cooperation, which is enabled when the form of gathering and exchanging information is unified.

Until we decided to create a proper system of collecting data that would function in the internet, the Neonatal Clinic had no safe electronic files.

While defining the range of data included into the files we wanted to keep a clear structure. The information connected with patients was grouped with the cause and effect relation. We also estimated the frequency of particular clinical events. To identify and register patients we used a unique personal date, such as the number of their clinic file, name and surname and date of birth. Other factors, such as sex, birth weight, week of gestation, prenatal condition may group patients in following analyses and statistical operation.

Crucial, and most difficult in drawing up the structure was the multitude and interrelation of different medical conditions. Schematically we divided information about diseases and grouped them in connection with particular organs and physiological systems. Therapeutic interventions, drugs used and prophylaxis were assigned to their proper folder. Some laboratory tests and measurements are multiple and to differentiate we labelled them by the date of collection. There are also specific filtering systems for protecting data against mistakes. Creating reports gives a helpful possibility of recording medical documents.

We test and verify clinical hypotheses connected with the effects of cure by the usage of the base.

Summary

We consider develop system database to improve same mechanism e.g. error handling, utilize *locale*¹⁹ to obtain multilingual version (The *gettext* functions implement an NLS (Native Language Support) API which can be used to internationalize your PHP applications). We intend to make an attempt at using ready framework like SMARTY (Template/ Presentation Framework)²⁰ or even move the system database into an environment operating system like eyeOS²¹.

The other development direction is to facilitate realization of a statistical operation on stored data. Our aspiration is to join option 'statistics' in the next step (including in content web front-end). It should be a simple SQL generator to receive a specification data form database as a substitute e.g. MySQL Workbench²². It can enable the carrying out of Data Mining²³.

The development of software adequate to the problem makes easy implementation a difficult procedure. For instance new version MySQL 5 entry useful features: Stored procedures, Views, Cursors, XA transactions, foreign keys (implemented in 3.23 for InnoDB), Triggers, Partitioning, Pluggable Storage Engine API, Row-Based Replication which should be taken into consideration. Typical uses for stored procedures include data validation, integrated into the database structure. Stored procedures are frequent, but not always, used as a method for executing SQL queries on the database objects in a way abstracted from a client application.

We should also remember that projects that are too grandiose might never have been finished.

References

Internet sources

- 1. http://www.gnu.org/licenses/gpl.html.
- 2. http://httpd.apache.org/.
- 3. http://www.openssl.org/.
- 4. http://php.net/.
- http://mysql.com/.
- E.F. Codd, A Relational Model of Data for Large Shared Data Banks. Communications of the ACM, June 1970, 13(6), 377-387. www.cs.nott.ac.uk/~nza/G51DBS/codd.pdf
- P.P. Chen, *The Entity-Relationship Model Toward a Unified View of Data*. ACM Transactions on Database Systems, 1976, 1(1), 9-36. http://csc.lsu.edu/news/erd.pdf
- http://www.supinfo-projects.com/fr/2005/etapes_merise_en/.
- 9. http://www.php.net/manual/en/ref.pdo.php.
- 10. http://pear.php.net/package/Spreadsheet_Excel_Writer.
- 11. http://struts.apache.org/primer.html#mvc.
- IEEE Std 1363-2000: Standard Specifications for Public Key Cryptography. IEEE, August 2000. http://grouper.ieee.org/ groups/1363/P1363
- 13. http://tools.ietf.org/html/rfc3280.
- 14. http://www.giodo.gov.pl/389/.
- http://dev.mysql.com/doc/refman/5.0/en/stored-procedures.html; http://dev.mysql.com/doc/refman/4.1/en/procedure-analyse.html.
- M. Lenzerini, Data Integration: A Theoretical Perspective. PODS 2002, 243-246.
- 17. http://dev.mysql.com/doc/refman/5.0/en/lock-tables.html.
- H.G. Hugh Jr., Scientific Method in Practice, Cornell University, New York 2002.
- 19. http://en.php.net/manual/pl/ref.gettext.php.
- 20 http://smarty.php.net/.
- 21. http://eyeos.org/.
- 22. http://mysql.com/products/tools/.
- 23. www.DataMining.DatabaseCorner.com.

Paper sources

- Walecki P., Sarapata K., Trąbka J., Biologizacja nauk informatycznych, W: Polskie doświadczenia w kształtowaniu społeczeństwa informacyjnego. Dylematy cywilizacyjno-kulturowe, red. L. H. Haber, Wydział Nauk Społecznych Stosowanych, Akademia Górniczo-Hutnicza, Kraków 2002, s.191-198
- Pyrczak W., Lasoń W., Sarapata K., Walecki P., Trąbka J., Rola spójności merytoryczno-formalnej w obiektowym projektowaniu systemów informatycznych, W: Techniki komputerowe w przekazie edukacyjnym, red. J. Morbitzer, Wydawnictwo Naukowe Akademii Pedagogicznej, Kraków 2001, s.148-152

- Pyrczak W., Lasoń W., Walecki P., Sarapata K., Trąbka J., *Teleinformatyczne systemy w edukacji medycznej*, W: Komputer w edukacji, red. J. Morbitzer, Wyd. Naukowe Akademii Pedagogicznej, Kraków 2003, s. 226-230.
- Walecki P., Sarapata K., Trąbka J., *Rzeczywistość wirtualna w ujęciu biocybernetycznym*, W: Proceedings of the Second International Conference on Philosophy and Computer Science Processes of Evolution in Real and Virtual Systems, Kraków 2002, s. 67-71.
- Sarapata K., Walecki P., Pyrczak W., Lasoń W., Roterman I., Szybiński Z., *Medical Database of thyroid cancer*, W: E-he@lth in Common Europe, red. K. Zieliński, M. Duplaga, IEEE Kraków 2003, s. 347-354.
- Walecki P., Pyrczak W., Lasoń W., Sarapata K., Roterman-Konieczna I., Klekawka T., Korab-Chrzanowska E., Balwierz W., Skoordynowana baza danych u chorobie Hodgkina (HD) zorientowana na diagnostykę w systemie rozproszonym, Przegląd Lekarski, t. 60, supl. 6, s. 78.
- Walecki P., Lasoń W., Sarapata K., *Biologizacja w obrębie* nauk technicznych, Krakowskie Studia Małopolskie, 2002, t. 6, s. 313-330.
- Lasoń W., Walecki P., Pyrczak W., Sarapata K., Roterman-Konieczna I, Medycyna Oparta na Dowodach (EBM) – nowe podejście optymalizacyjne wykorzystujące systemy kompu-

terowe, W: Tadeusiewicz R., Ligęza A., Szymkat M. (red): Computer Methods and Systems, 2005: 455-462.

- Martyniak J. (red), Podstawy informatyki z elementami telemedycyny: ćwiczenia dla studentów medycyny, (Walecki P. współautor sześciu rozdziałów), Kraków: Wydaw. Uniwersytetu Jagiellońskiego 2005.
- Walecki P., Sarapata K., Lasoń W., Pyrczak W., Roterman-Konieczna I., Balwierz W., *Telemedical Database of Hodgkin Disease*, In: Transformation of Healthcare with Information Technologies, Studies in Health Technology and Informatics, IOS Press, Amsterdam-Berlin-Oxford-Tokyo-Washington, DC 2004, s. 51-60.
- Lasoń W., Walecki P., Trąbka J., *Telemedyczne bazy danych* – *teoria i praktyka*, W: Społeczeństwo informacyjne wizja czy rzeczywistość? Red. L. Haber, AGH Uczelniane Wydawnictwa Naukowo-Techniczne, Kraków 2004, s. 349-360.
- Sarapata K., Usarz M., Trąbka J., Walecki P., Pyrczak W., Roterman-Konieczna I, Security of net servers in biological approach, Journal of Artificial Intelligence, Vol. 1, No 1(23)/ 2004, pp. 31-38.
- Lasoń W., Walecki P., Pyrczak W., Sarapata K., Trąbka J., Zintegrowane telemedyczne bazy danych pacjentów z chorobami nowotworowymi, Sztuczna inteligencja – Organizacje wirtualne, red. M. Kłopotek, J. Tchórzewski, Wydawnictwo Akademii Podlaskiej, Siedlce 2003, s. 35-40.



ADVANCED SYSTEM FOR ORGANIZATION OF WORK AND SCIENTIFIC DATA EXCHANGE

ZAAWANSOWANY SYSTEM WSPOMAGANIA PRAC ORGANIZACYJNYCH I WYMIANY DANYCH NAUKOWYCH

ADAM PIOTR ŻOCHOWSKI* **, WOJCIECH LASOŃ***

* Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Reymonta 4, 30-060 Krakow, Poland ** Bell Canada Enterprises Inc. ***Department of Bioinformatics and Telemedicine, Collegium Medicum, Jagiellonian University, Kopernika 7, 31-034 Krakow, Poland

Abstract: The purpose of this project was to create a system that would help in organizational and scientific data exchange. The system has been put into production at the Department of Bioinformatics and Telemedicine of the Medical College of Jagiellonian University.

The system is designed for the university staff, doctors, experts, students as well as anyone interested. The system helps to maintain fast and efficient data exchange of both administrative and scientific kind, particularly directed towards the medical and related fields such as biostatistics, bioinformatics and telemedicine.

This paper shows in detail: the design of the system, technology involved, functionality of the system, data flow, database structure, and security system, all matching the requirements of the project. The conclusion contains a discussion on the system's future growth for which the project is prepared for as well as direct examples of some of the possible new features.

Keywords: content management system, hashing, phonetic string matching, trust management

Streszczenie: Głównym celem pracy było stworzenie systemu wspomagania wymiany danych naukowych. System został skonfigurowany dla wymagań Zakładu Bioinformatyki i Telemedycyny Collegium Medicum Uniwersytetu Jagiellońskiego. System przeznaczony jest zarówno dla pracowników jednostki naukowej a także lekarzy, studentów oraz innych współpracowników. Umożliwia on pełniejszą i szybszą wymianę informacji zarówno organizacyjno-administracynych jak i naukowych. System został stworzony do zarządzania danymi pochodzącymi z nauk medycznych i innych pokrewnych dyscyplin w szczególności informatyki medycznej, telemedycyny, bioinformatyki.

Artykuł pokazuje nowatorskie rozwiązania zastosowane w skonstruowanym systemie zarządzania treścią (CMS), które znacznie poprawiają jego funkcjonalność. Szczególny nacisk został położony na kwestie bezpieczeństwa, poziomy dostępności danych i profile użytkowników. W systemie zostały zaimplementowane zaawansowane rozwiązania informatyczne dotyczące mechanizmów wyszukiwania danych jak fonetyczne porównanie i wielopoziomowe przydzielenie uprawnień.

Słowa kluczowe: system zarządzania treścią, kryptograficzne funkcje hashujące, fonetyczne porównanie słów, uwierzytelnianie

Introduction

Project's requirement was to design and build a system geared towards information exchange and data organization for the department of Bioinformatics and Telemedicine of the Medical College of the Jagiellonian University. It helps to organize publications, articles, books, research, grant proposals, news, presentations, and personal notes. Main benefits include remote access through the internet as well as helpful and easy to use interface. The multilayered permissions and security ensure high integrity of the system. Each of the involved departments yearly publishes dozens of articles. Projects vary in multiple ways from funding (personal, grant based) to involvement from other staff members, international cooperation and other third parties. Information on this work should be available to every person within the department, although not to the same extent to everyone.

Taken into the account were the following parameters:

- varied scientific directions in the departments
- interdisciplinary working on the cutting edge of science and technology

- varied cooperation
- large count of the courses and exercises
- physical separation between departments, lacking direct staff contact

The above factors heavily weighted into creating a single and cohesive system that was to be secure while at the same time to help the staff coordinate their work.

Requirements of the System

During the design of the system there were a series of requirements that needed to be fulfilled. These came from the analysis of workflow within the department. The main requirements were:

- remote access to the system
- integration to other systems within the department
- helpful and easy interface
- advanced methods for searching data
- archival of data
- verification of data
- hierarchic security
- time based restriction/granting of access
- ability to define actions allowed on a data set
- ability to manually add data
- ability to track comments and file attachments
- creating of data sets for specific users or groups of users

After further analysis of work being performed by the staff within departments it became apparent that the system could also track additional data, not just articles and publications, but also presentations (both: of student and conference origins), student works (bachelors and masters), as well as PhD dissertations. All the above requirements have been met, and the implementation specifics show originality in work. Thanks to elastic and adjustable framework it is easy to have the system change the way it works.

Users and Groups

Generally the staff members are split into groups based on the departments that they are in. Each faculty member (user of the system) is a member of a specific department, although it is possible for a staff member to coexist in a few departments. A member of one department can have access to data from another department. For example, a staff member of the Bioinformatics department can, but does not necessarily have to, have special access of the work of students from the Telemedicine department.

Different type of a group of users is represented by the students, understood here as people working with the department. Such users would have higher access than some anonymous internet guests; however, they would not have as much access as a staff member. Some students, due to cooperating work with a staff member, for a specific content area, might have escalated privileges. The least access would be given to the anonymous internet guest. Such user would have most limited access. For example, instead of having access to an article, they could only access its summary. Anonymous internet guest does not have to register to access content.

The system allows creation of an elastic user profile. Any user and/or group can be assigned to any data with any of the available permissions, and even that further can be limited by a time constraint.

Table 1. Groups based on departments and access levels

Biocybernetics Bioinformatics Biostatistics Telemedicin							
Administrator							
Staff Staff Staff Staff							
Student Student Student							
Anonymous Internet Guest							

Privileges

Users of the system are split into hierarchic groups. The structure has 4 levels of permission groups:

- administrators people controlling the well being of the entire system
- staff members the administrators and users within a department
- students and registered users
- anonymous internet guests

Further, the right to content is assigned per groups, but also can also be fine grained and assigned individually.

Data Categories

Another major aspect of the system is the ability to categorize data. Depending on data type tracked it is important have additional field to track on that content. In case of books this would be an ISBN number.

Publicati	OS	Projects		News		Work		Staff	
1.	Books	10.	Grants	17.	News	21.	Didactical	28.	Own Work
2.	Book Chapters	11.	Status	18.	Courses		Materials	29.	Projects
3.	Book Synopsis		Research	19.	Corespondece	22.	Presentations	30.	Other
4.	Scripts	12.	Private	20.	Other	23.	Student		
5.	Scripts		Research				Projects		
	Chapters	13.	Coordinated			24.	Master Thesis		
6.	Conferences		Research			25.	Doctorate		
7.	Summary of	14.	E-Learning				Dissertations		
	Conferences	15.	Web Pages			26.	Habilitations		
8.	Conference	16.	Other			27.	Other		
	News								
9	Other								

Table 2. Data categories

No matter what category data resides in, it has to have ability to link and be linked to from data from any other category. An article can refer to other articles (through keywords and explicitly) but also it can refer to books or conferences. Those on the other hand could link to picture galleries or other related articles. These links should operate between internal content, as well as external (for example to http://pubmed.org)

Implementation of the System

One of the postulates of the system was easy of use, both from the user perspective as well as maintenance. Choosing technology for the system required taking into account usefulness as well as longevity and price.

Technology

The technology was chosen that all the departments were most comfortable with. Both bioinformatics, biostatistics and telemedicine departments relied on PHP and MySQL hence these tools were chosen for this system. UTF8 font encoding was chosen to avoid future issues with various accented characters as well as to maintain backward compatibility with ASCII tools.

Describing Data

Any data that is sent has to be described. Filename usually tells us what to expect inside a file. Title of an article tells us what to expect within the article. The system was built to rely on various methods to describe data. Some descriptions human made, like titles or summaries, while others to help software deal with the data.

Line and Paragraph Descriptions

In articles, line description could be equalled to the title of the article, while the paragraph description would become article's summary, its synopsis. Article to which a user might have limited access might only see the summary, not the whole content. In the database they are defined:

article_title	varchar (128) not null
article_summary	varchar (1024) not null

Data Transfer: Errors and Security

When transferring data it could become that the destination does not receive what source was sending. This could occur due to an error during the transfer, or because it was done deliberately. To recognize that transfer did not happen as intended a naïve solution would be to send it over again. By comparing the first copy to the second copy we could hope to spot if an error has occurred. However, if some third party deliberately changed the data once, there is nothing stopping them from doing so again during the second transfer. Even if we do detect an error, then we will not know which of the two versions on our hands is proper. After all, the error could have occurred during the second transfer, not the first. Sending same data for the third time would solve this problem, but then we would be wasteful on the resources, both bandwidth and storage. Better solution, that does not rely on triplicate data transfers, is to write short algorithmic descriptions of the sent over data. The first such simple method relied on checking the parity. If the right number of bits set to one were sent. For example:

data transferred '0101 1111' has even parity as there are 6 ones sent

data transferred '0101 1101' has odd parity as there are 5 ones sent

In this example we can see how an error can be detected. Once we know there is an error, we can send data again and verify it again, if it was proper this time. No need to send data three times. However, it quickly became apparent that a lot depends on the type and amount of errors occurring. Notice two errors occurring at the same time:

data transferred '0101 1111' has even parity as there are 6 ones sent

data transferred '0111 1101' has even parity as there are 6 ones sent

We will not notice the error occurring. Many algorithms have evolved to help detect various errors. Some are geared for speed; others for error correction, or to stop and slow down a third party from tampering with the data. Our system relies on three popular hashing methods: CRC32, MD5 and SHA1 in the following way:

sprintf(«%08X», crc32(file_get_contents(\$file))); md5_file (\$file); sha1_file (\$file);

Any user of the system, either when sending or receiving, can ensure that the codes provided by the system for their files are same as on their own machines.

Keywords

Sometimes we need to treat similar data as if they were the same. For example, when user misspells a word, we might want to try to figure what was really meant.

There are few of different models of comparing words and finding how similar to each other they are. One such method is the Levenshtein string distance. It takes two strings and reports how many changes are required to transform one string to the other. The fewer the changes the more similar the words are. Another method is the Oliver [93] probability where we get a probability of how two words are similar to each other. In both of these methods there is a performance problem. To verify if two words are an exact match we need O(m) operations, while Levenshtein distance requires O(m*n) operations, and Oliver[93] probability requires O(m*m*m), where m is the length of the larger string, while n is the length of the other string.

At the beginning of the 20th century, long before first computers were being constructed, people noticed the need to find similar words. In United States, with large immigration influx, it was needed to track families, where it was known the pronunciation of a name, but not necessarily its spelling. For this purpose an algorithm titled soundex was created, that would convert string/words into special codes. Words that had similar pronunciation then would often share same soundex code, designating them as similar to each other.

For the last hundred years alternatives have also sprung up. One of such alternatives is metaphone and its younger brother double-metaphone.

Table	3.	Misspelled	words	and	their	phonetic	hashes
-------	----	------------	-------	-----	-------	----------	--------

Word	soundex	metaphone	double- -metaphone	Daitch- -Mokotoff soundex
Rzeka	R220	RSK	RSK, RTSK	450000, 945000
Zeka	Z200	SK	SK	450000
Rzka	R200	RSK	RSK, RTSK	450000, 945000
Druk	D620	TRK	TRK	395000
Drok	D620	TRK	TRK	395000
Adam	A350	ATM	ATM	036000
Aton	A350	ATN	ATN	036000

Quite problematic for Polish purposes is the lack of support of polish pronunciation rules of written words. There exists a version of soundex that takes into account Slavic phonetics, and it has been dubbed 'slavic soundex', although its official name is based on the designers, Daitch-Mokotoff soundex. Unfortunately, it treats words on their phonetics rules, and not grammatically. Hence two words 'przepadł' and 'spaść' would be considered as two different words, although both are created from the base word 'paść'. Other systems also work on the synonyms of words, their meaning similarity. For example, words 'good' and 'better' and 'well' are related, yet neither in writing, nor spelling, are they similar to each other.

The system has been designed to rely on soundex and metaphone algorithms.

Nodes

The main information container in the system has been dubbed a node. It is the elementary container that other related information refers to.

The Reason Behind

The system that is built includes varied types of content: articles, publications, dissertations, Masters Theses, news, conference presentations, all having some of the data

 Publication

 PK
 ID

 Author
 Creation Date

 Title
 Synopsis

 Text
 Source

 Citations

News		
РК	ID	
	Author Creation Date Title Synopsis Text	

in common. A possible, naïve, solution would be to store each information type in its container. (Fig 1)

Since every piece of information can relate to another. A published article can be linked to news or linked to Masters Theses. We notice that the number of pairings we get is related to the combination choose two: C(n,2). (Fig 2)

This further is complicated as each content type can have additional data attached, such as comments, or files. (Fig 3)

Creation of a Node

Keeping all data in own specialized containers creates problems with tracking the links. Alternatively, if common data was centralized it would make life simpler. After all, articles, news, Masters Theses, conferences, publications and presentations have common fields tracked: the author, creation date, title, and synopsis. This means, instead of creating a container for articles and another for news, we create a single table that has common data, named Node.

Now, when we want to attach comments it is just a single table as opposed to a whole set of them. For example: a node with additional tables like comments, attachments, ranking and keywords.

Referring to Nodes

Single unique object within a database needs to be tracked in a unique way. This is generally done through the use of sequence numbers, dubbed 'identity' or 'auto' numbers. The idea is that every new entry in the database gets its own serial number. That way, it is fairly easy to refer to each specific item. However, this often exposes information about the system. Knowing a serial number of 321 we can safely assume that 320 exists or has existed at one point in time, and have a good probability that 322 also exists.

UUIDs – Universally Unique IDentifiers (sometimes found under the name GUID – Global Unique IDentifier) tries to avoid predictability. Instead of providing numbers in sequence, serial numbers, an algorithm generates always random numbers. Since the number generated is huge (128 bits, allowing for 3.4*10³⁸), the probability of guessing UUID of an existing node is infinitesimal. With a million of nodes, probability is one to 3.4*10³². However, the length of UUIDs, 32digits in hexadecimal format, is not friendly for writing down or memorizing.

Popular systems found online use alternative tracking







Fig. 2. Pairings of data being categorized in own tables



Fig. 3. Attaching comments functionality to data that is categorized in own tables

methods. They have one method of tracking of content internally, while to the user, another method is presented, for example, by date and node's serial number of the day it was created in. RewriteEngine on RewriteRule ^node/(...)/(..)/(.*).html\$ node.php?year=\$1&month=\$2&day=\$3&serial=\$4

node/[year]/[month]/[day]/

Further, the internal handling of the system and parameters can be hidden, obfuscated, thanks to the use of URL rewriters, such as Apache's mod_rewrite. In Apache's configuration file, the following would then need to be written:

Keywords

It is a tedious job to define manually which node is related to another node, not without benefits. Such linking is as good as a human can be, but if some of the work can be automated then the important time is saved on doing actual research, or to attend students, instead of just linking data manually within the system. The solution is to define descrip-


Fig. 4. Implemented keywords structure

tions that computer can use and try to find best relations within the system. One such description relies on keywords. User enters words that are critical and define data the best, and then the system can search for other nodes with same or similar keywords. Not only can the system find related data, but also try to sort based on relation. (Fig. 4)

Common Keywords

If a node A has three keywords x, y and z then only other nodes, also with either keywords x, y and z are related. Some node that has all three keywords is more related to node A than another node that only matches one keyword. A node that has no common keywords can be assumed as totally unrelated. The SQL code can then look the following:

```
SELECT nodes.id
, COUNT ( nodes_keywords.keyword_id ) AS common
FROM nodes
INNER JOIN nodes_keywords ON
nodes.id = nodes_keywords.node_id
AND nodes_keywords.keyword_id IN ( x , y , z )
GROUP BY nodes.node_id
ORDER BY COUNT (nodes_keywords.keyword_id ) DESC
```

Keyword Popularity

An important aspect of a keyword is the similarity metric. How important is some word in relation to another word. For example, if we have a node A with keywords x, y and z, we match two related nodes to it, a node B through keyword x and node C through keyword y, then, which of the two matched nodes, B or C, is a closer match to node A? It depends on how popular are the keywords. If matching by keyword x is less likely event than matching by keyword y, then node B, through keyword x, should be promoted. In SQL we could do this through:

SELECT nodes.id , SUM (1000 / popTab.popWord) AS weight FROM nodes INNER JOIN nodes_keywords ON nodes.id = nodes_keywords.node_id AND nodes_keywords.keyword_id IN (x , y , z) LEFT JOIN (SELECT keyword_id , count(*) as popWord FROM nodes_keywords WHERE nodes_keywords.keyword_id IN (x , y , z) GROUP BY keyword_id) AS popTab ON popTab.keyword_id= nodes_keywords.keyword_id GROUP BY nodes.id

ORDER BY SUM (1000 / popTab.popWord) DESC

This way, we assign a weight to keyword match. A keyword that exists in the system 500 times will give a weight of 2, while a keyword that matches only 10 times will provide a weight of 100. We then sum the weight for all keywords of a matched node to get total weight of a matched node. In the end, the results are sorted descendingly by the total weight of the matched node.

Node Keyword Count

System could however become abused. A node could be defined with a copious amount of keywords making it a related match with virtually every other node. The system needs to promote nodes that match, while having few keywords, over nodes that match to anything.

For example, node A has three keywords, x, y and z. There are two related nodes, B and C, both matched by same keyword y. Which of the two matched nodes is a better match? If we look into how many keywords has in total node B and node C, we can promote the node that has less keywords over the one that has more. If node B has 10 keywords, while node C has only 2, then node C is more important match than node B. The reason is twofold. Node C, having less keywords, is less likely to be found in the system, hence it should be promoted higher in the results if it does actually match something. Secondly, node B having 10 keywords implies that each keyword, by itself, does not explain it well, while node C, with only two keywords, tells us that each keyword by itself would probably describe half of the node. In SQL the example solution is:

GROUP BY nodes.id ORDER BY SUM (1000 / popTab.popWord) DESC

Similar Keywords

Words sometimes can be misspelled. People not always remember proper spelling of a word; it could be something overheard. The system takes this into account and also tries to find a match based on soundex and metaphone algorithms. In the SQL we can see the additional subselect required to pull data through another indirection:

```
SELECT nodes.id
, COUNT ( nodes_keywords.keyword_id ) AS common
FROM nodes
INNER JOIN nodes_keywords ON
    nodes.id = nodes_keywords.node_id
    AND nodes_keywords.keyword_id IN IN (
        SELECT id
        FROM keywords
        WHERE soundex IN (
            SELECT soundex
        FROM keywords
        WHERE id IN ( x , y , z )
        )
    GROUP BY nodes.node_id
```

ORDER BY COUNT (nodes_keywords.keyword_id) DESC

Final Keywords Search

It is possible, that some related word is related through only one algorithm, either soundex or metaphone. Such word then needs to have a lower weight than if it is found to be related through both algorithms.

In the end, all the methods need to be combined to give user not only many matches, but also the best matches at the top. This means that we promote what has many words in common as it's a better match than something that has less. On the other hand, we want also to promote nodes that are rarely matched, as it is significant when such a match occurs.

Linking Related Nodes

A node can refer to any other node. A mapping table is required to accomplish this. This table would store references to any pair of nodes.

```
create table nodes_relationship (
node_id_parent integer not null
, node_id_child integer not null
, title varchar (80)
, author_id integer not null
```

- , creation_date timestamp not null DEFAULT
 - CURRENT_TIMESTAMP
- , primary key (node_id_parent , node_id_child)
- , foreign key (node_id_praent) references nodes (id)
- , foreign key (node_id_child) references nodes (id)
- , foreign key (author_id) references users (id)

```
) type=InnoDB CHARACTER SET utf8;
```

Security System

The security of every system is always a critical issue, however, often overly trivialized. Software should be cre-

ated in mind that some users could abuse weakness of a system, either for fun, or for more sinister goals. No matter what is the reason, such actions should always be curtailed and limited. The project was designed to be a non trusting, closed system as there is content that not everyone should have access to.

Logging In, Passwords

The first step in making a project secure is to ensure that there are doors with locks. A guest in a house will be welcomed only in the general entrance hall, but will not enter any of the rooms. The combination of username and password creates an access key that opens doors to sections that non privileged people don't have access to. Some doors could be accessed by many keys, while other would work only for one person.

Unfortunately, the passwords are problematic. If the password is overly long and complex it becomes very easy to forget. On the other hand, if the password is trivial and predictable then it is not secure. When the system generated a password it has to be both, secure yet simple to remember.

The problem is further complicated since users often prefer to remember one or two passwords. This means that there are many systems that a single username/password combination would grant access. If any of these systems is compromised, and the attacker gains user's passwords, then he could, with these passwords, try to compromise the other systems. How to build a system that is both secure but does not know user's passwords? How to verify a password without knowing it? The solution is to not store the password, but store a value derived from the password. For example, we could try to store SHA1 hash of the password. That way, when user enters the password the first thing the system performs is a hash of the password, and then compares the hash of the just entered password with the hash stored in the database. If at any time the system is successfully penetrated by a hacker then nothing but the hashed passwords is exposed. There is no easy way for the hacker to reverse the hash back to original password, hence other systems, that were secured with the same password, stay secure. The system relies on SHA1 to keep hashes secure.

User Access

The system, in initial state, assumes that no one has access to any node. The rights are gradually being granted. Even the anonymous internet guest has to be granted access to even see that a specific node exists. Although generally nodes are assigned with permissions to groups, sometimes, to avoid need of one person groups, users too can be assigned permissions to nodes.

User Groups, Departments

Assigning permissions to every user to every node would require Cartesian product of the m*n amount of permissions, where m represents number of users, while n the number of nodes. With 40 users and 80 nodes we end up with 3200 pairings. To minimize the amount of manage-



Fig. 5. Managing access to groups

ment of permissions it is much easier to group users and assign permissions to the groups.

Some such groupings would follow the physical department setup: Bioinformatics, Biostatistics and Telemedicine, however, other user grouping can occur, such us hierarchy based: administrator, staff-member or student. The groups are created independently. We can assume that groups are built like a directed graph, where each group is a node, and the group directional pairing defines which group is superior to the previous. This elasticity allows for creation of new hierarchal levels and new, even independent, groupings. (Fig. 5) The system, when it comes to node, recognizes the following permissions gradient hierarchy:

- no rights
- limited read access
- full read access right
- editor access
- administrator access

For example:

- anonymous internet guest has limited read access
- group a (users i, j, x and y) have full read access
- group b (users j, k, y and z) have editorial access
- users v, x, y and z have administrator access

Security - Gradient

As it was mentioned previously, by default no one has access to any content. The permissions are slowly granted. Slowly there is a set of permissions assigned to the node. With the above permission setup some users end up with multiple accesses to the system. The system grants the highest granted permission to the user.

G	rupa			start	koniec	gradient
Gość			zmień			Ograniczone
dod	aj grupę			start	koniec	gradient
Grupa A		(członkowie)	zmień			Pełne
Grupa B		(członkowie)	zmień			Edytorskie
dodaj u	żytkownika			start	koniec	gradient
&v		(nadrzedni), (suma praw)	zmień			Administratorskie
⁸ x		(nadrzedni) , (suma praw)	zmień			Administratorskie
&у		(nadrzedni), (suma praw)	zmień			Administratorskie
⁸ z		(nadrzedni) , (suma praw)	zmień			Administratorskie
pis gra	dientów					
poziom	opis					
D	Brak praw					
20	Ograniczone					
40	Pełne					
5 0	Edytorskie					
80	Administrato	skie				

	Guest	Group A	Group B	Explicit	Final
Guest	Limited				Limited
Н	Limited				Limited
I	Limited	Full			Full
J	Limited	Full	Editorial		Editorial
К	Limited		Editorial		Editorial
V	Limited			Admin	Admin
Х	Limited	Full		Admin	Admin
Y	Limited	Full	Editorial	Admin	Admin
Z	Limited		Editorial	Admin	Admin

 Table 4. Resolution process of action based permissions

The final permission that any user receives is the highest permission. For example, we can analyze how user j ends up with his editorial permission.

- since lowest permission assigned is to guest, user j at minimum gets anonymous internet guest permission, meaning limited read access
- however, user j is also a member of group A. Between the limited read access granted to anonymous internet guest and the group A's full read access, user j ends up with the full read access
- however, user j is also a member of group B. Between the full read access granted to group A and the group B's editorial access, user j ends up with the editorial access

Of the three possible permissions that user j has, the highest is granted. (Fig. 7)

In the database, the gradient permission is represented by an integer. The highest gradient to a node that can be traced to a user is used as user's access. However, gradient assumes that the higher access includes all permissions of a lower access. User with gradient at value 60 can do anything that users with lower accesses can do. This causes problems if we want to have two groups that have permissions mutually exclusive. For example, one group to have ability to attach files, but not comments, while another group vice-versa, ability to add comments, but not files.

Security - Actions

Sometimes there is a need to assigns permissions based on actions and not hierarchies. In gradient/hierarchy system, the higher the permission level the more access user is granted, however, this does not permit mutually exclusive actions. For example, let assume that on gradient level 20 the user is granted file attachment ability. If at level 40 we grant comment ability we end up with 3 types of users: below level 20 with nothing granted, between 20 and 40 with file attachment, and above 40 with comments and file attachment. There is no possible gradient level that would grant comments, but not file attachments. The solution is to track user's access based on actions he is granted. Can user attach file? Can user comment?

For example:

Group A (with users i, and j) can attach files Group B (with users j, and k) can comment (Fig. 8)

We end up with following action permission resolution table:

Table 5. Resolution process of action based permissions

	Guest	Grupa A	Grupa B	Final
Guest	Access			Acces
I	Access	File Attachment		Access File Attachment
J	Access	File Attachment	Comment	Access File Attachment Comment
К	Access		Comment	Access Comment
L	Access			Access



Fig. 7. Gradient based permissions

Uprawnienia Bieżące :

	Grupa			start	koniec	przegladac	dodac_komentarz	dodac_plik
Goś	ć		zmień					
	dodaj grupę			start	koniec	przegladac	dodac_komentarz	dodac_plik
Grup	a A	(członkowie)	zmień					~
Grup	ba B	(członkowie)	zmień				~	
do	daj użytkownika			start	koniec	przegladac	dodac_komentarz	dodac_plik





Fig. 9. Complete database diagram

Through assigning permissions to specific actions the permissions it is now possible to setup users who have permissions that are mutually exclusive. We have user i that can only do file attachments, user k that can only comment, user j that can do both, and I that can do neither.

Actions, such as commenting and attaching files, are not related; however, some actions can easily be dependant upon each other. For example we could build a system with two actions one for reading comments and another for writing new ones. In that case, adding new comments has little sense if user cannot read comments. Such actions are then dependant upon each other. This is where gradient comes in handy. We can define that one action is a gradient based. Where commenting action could be then defined as a gradient action, with levels: read comments, add comments and edit comments.

The Database

The final database ERD (Fig. 9).

Interface

Publikacje

Ksiażki

Skrypty

Wersje

Inne

Konferencje

Streszczenia Konferencie

Doniesienia

Projekty

Badania Statutowe

Badania Własne

Współpraca z Innymi Jednostkami

Strony WWW

E-Learning

Inne

Granty

The interface is generally a compromise between easy, familiar and understandable form that fits the user's needs, while trying to give a multitude of options and complex advanced functions. It is not an easy task to be at the same time complex and easy. The more choices the user is presented the more advanced the tool becomes, and becomes less understandable. Besides this problem with ease and complexity, aspects of physical display as well as hardware and software limitations need to be taken into account. The lower the requirements are set on the user the more likely that the user will end up using and relying on the system.

System Wymiany Danych

👍 - 🔿 - 🥃 🐵 🕋 🔝 http://localhost/magist 🔍 🕨 🖸 🏦 Strona Główna Collegium Medicum 🔎 zmień treść &System (2007-06-04 00:52:53) zmień uprawnienia 🗏 dodaj pliki Zainteresowania badawcze z zakresu bioinformatyki koncentrują 🗟 dodaj komentarz się na analizie układu białko-ligand na drodze symulacji Książki Rozdziały komputerowej. W szczególności interesują nas ligandy supramolekularne o właściwościach ciekłokrystalicznych. Różnią się one od typowych ligandów pochodzenia niebiologicznego sposobem i miejscem wiązania. Analizowane Skrypty Rozdziały obecnie przez naszą grupę ligandy supramolekularne należą do grupy związków bis-azo o Konferencje Pełne wydłużonych, sztywnych, symetrycznych cząsteczkach wykazujących silną zdolność do asocjacji w wodzie, zwłaszcza w roztworach soli. Do tej grupy związków należą: czerwień Kongo, błękit Evansa i inne. Problematyka badawcza obejmuje opracowanie analiz statystycznych dla badań i eksperymentów medycznych w oparciu o kliniczne bazy danych. Pragniemy również dostosować ustalone zasady prawne oraz metody przeprowadzania analiz statystycznych do warunków i ustaleń obowiązujących w Polsce. Budując specjalne modele regresji

(regresja logistyczna, Poissona, analiza przeżycia) oraz wykorzystując metody analizy wariancji zajmujemy się wyznaczaniem czynników ryzyka w sposób istotny wpływający na przebieg i natężenie wybranych chorób. Zespół zajmuje się również zastosowaniem metod wielowymiarowej statystyki (analiza dyskryminacji, analiza skupień) oraz sieci neuronowej do prognozowania i wybranych zagadnień klasyfikacji.

W ramach zagadnień telemedycznych zajmujemy się opracowywaniem zaawansowanych systemów przetwarzania danych medycznych oraz ich współdziałania z specjalnie zbudowanymi do celów diagnostycznych aplikacjami typu ekspert. Przy budowie i projektowaniu programów doradczych i wspomagających decyzje dla potrzeb praktyki

Since the most critical in the system is the actual data, the node, then the largest part of the screen is dedicated for it. Generally in such systems left side of the screen is dedicated towards system menu. On the top we remind users where they are.

Depending on granted permissions there might be a set of actions that can be performed on a node. These actions, not to be confused with system actions, are added in the node section of the screen, in the top right corner.

The figure nr 19 shows the actual screenshot taken from the system. The screenshot presents us with a node. On the left side of the screen we have the system menu that contains various categories that nodes are in. The node has a title and author and creation date present. In the top right corner of the node section we see additional menu of actions that can be performed on this particular node. (Fig. 10.)

Discussion

The ongoing changes and evolution of the departments (biostatistics, bioinformatics and telemedicine) as well as the interdisciplinary character of the work will force, in the future, changes upon the system. With this expected continuation of growth the project has been designed and implemented, welcoming any oncoming changes as opposed to fearing them. The elastic framework makes the system ready for adoption for new needs and requirements.

Elastic Framework

Deliberate extra effort was put in the design of the programming libraries. Such libraries, often referred to as framework, are a collection of code snippets that can be

used easily in many places for many purposes. For example: instead of writing HTML code to handle text fields and all of its related data (such as its description, sizes, etc); we could have a function that does the job for us:

Current implementation of the function uses the parameters to generate legend for the input field. When the field is defined as read only then the user is not presented with an input field, but an unchangeable text.

Since the function returns a text string, as opposed to sending it straight to the output buffer, the string can be used for other purposes. For example, the input field can be used to generate a download file, or sent through email.

Another example of the easy use of the system from the programmers point is the security. For the purpose of the example, let us assume we want to add ability to track who can delete files from a node. First we add to the permissions table new column for deleting files, so:

ALTER TABLE nodes ADD COLUMN

access_delete_file bit; ALTER TABLE nodes_accessby_groups ADD

COLUMN access_delete_file bit; ALTER TABLE nodes_accessby_users ADD

COLUMN access_delete_file bit;

This is enough for the system to start tracking the rights to delete files for groups and users. The interface for as-

signing permissions detects automatically the new tracked actions, so there is no need to change the administrative functions.

The only thing left is to add actual code that deletes the file in question, and verifies that user actually has the rights to do so:

```
if (canl(«nodes»,'id','##nodelD','delete_file'))
    unlink ( fileID2filename ( «##file» ) );
```

The code calls 'canl' function. The function takes 4 parameters, the first two help it work with the database, first defines table name, the second the primary column name. Third parameter is the node number in question. The last parameter is the action we want to perform on the node.

If the function confirms our inquiry, that we have the right to delete a file attached to that particular node, then it returns a value of true (in case of gradients, it returns the highest gradient value). This allows us very easy checking for actions and gradients, whether user should be allowed or not with his actions.

The canI function implicitly takes user's identifier from within the session, hence avoiding extra parameter.

The look of the system is defined through CSS (Cascading Style Sheets). CSS is a set of instructions that tell to the browser how to display content. The CSS also allows defining additional instructions as how to handle the content on other mediums. For example, a page could have one look on screen, but different on printout. Further, there are options for the visually impaired and also mobile devices that have constrained display screens.



Fig. 11. Expected growth of the keywords section of the database structure

Keywords Database Growth

Although the system was designed with keywords and similar search capabilities, they are not complete, nor fastest. The current design is generating the statistics dynamically, as seen through the subselects and various aggregates. Since these statistics rarely change they could be cached as columns in the database. The trade off requires ensuring that statistics are always up to date, so any changes to node's keywords would involve changing related statistics.

Currently the system relies on soundex and metaphone codes. However, double-metaphone and Daitch-Mokotoff soundex algorithms are also planned. These two methods are different in that they can return more than one hashed code for a keyword. This causes another set of indirection during lookup of a word to a keyword.

With the above changes the database would then start to look in the following way (Fig. 11):

The Future Perspectives

The described advanced system for data exchange is the culmination of many months of work. The needs and requirements were provided by the staff members of the Bioinformatics, Biostatistics and Telemedicine departments. After all they are the intended users of the system. Alas, there exist a lot of possibilities for growth. Some of it was noticed during the design and implementation stages, some during the testing phase when the staff members got the first hands on experience. Here is a list of features that are planned to be added in a near future:

- expanded commenting system: threaded / nested commenting system
- more statistics and searches with the administrator in mind
- calendar functionality (to show upcoming conferences / presentations)
- picture galleries
- video files
- expansion of what can be done through the user's profiles (such as private messages or user settings)

It is worthwhile to emphasize that already, in the current state, after the tests, the project has been not only accepted but welcomed by the department staff. This gives a good indication that the project has reached its goals and fulfilled the needs and requirements.

References

Internet sources

- 1. MD4 http://www.ietf.org/rfc/rfc1320.txt
- 2. MD5 http://tools.ietf.org/html/rfc1321
- 3. SHA1 http://tools.ietf.org/html/rfc3174
- 4. CSS http://www.w3.org/TR/CSS21/
- 5. HTML http://www.w3.org/TR/html4/
- Wikipedia http://pl.wikipedia.org/wiki/
 http://en.wikipedia.org/wiki/
- 7. MySQL http://www.mysql.com/
- PEAR PHP Extension and Application Repository – http://pear.php.net/
- 9. PHP: Hypertext Preprocessor http://www.php.net/
- 10. Apache HTTP Server Project http://httpd.apache.org/
- 11. Apache module mod_rewrite http://httpd.apache.org/docs/ 1.3/mod/mod_rewrite.html

Paper sources

- 1. P. DuBois. MySQL Cookbook. O'Reilly Media Inc., 2006.
- N. A. Good, A. Kent. Foundations of PEAR: Rapid PHP Development (Foundations). Apress, 2006.
- S. Schmidt, S. Stoyan, A. Wormus, C. Lucke. .PHP Programming with PEAR. Packt Publishing, 2006.
- A. Trachtenberg, D. Sklar. PHP Cookbook. O'Reilly Media, Inc., 2006.
- Mark Arnold, Jeff D. Almeida, Clint Miller. Apache administracja serwera. Wydawnictwo RM, Warszawa 2002.
- 6. R. Stones, N. Matthew. Bazy danych I MySQL. Helion, 2003.
- H. Garcia-Molina, J.D. Ullman, J. Widom. Implementacja systemów baz danych; Wydawnictwo Naukowo-Techniczne. Warszawa, 2003.



Felemedicin

WEB – BASED SYSTEM FOR TESTING OF STUDENTS KNOWLEDGES

WEB – БАЗОВАНА СИСТЕМА ТЕСТУВАННЯ ЗНАНЬ СТУДЕНТІВ

Alexander Lysenko, Igor Lytvyn Олександр Лисенко, Ігор Литвин

Ternopil State Economic University Тернопільський Державний Економічний Університет

1. At the traditional form of studies basic role in organization of educational process is taken to the teachers but in the systems of electronic education management of educational process is distributed evenly between a teacher and students, or it is fully subordinate to the latter. It allows the students to study independently in convenient time and place remote from a teacher. But efficiency of independent studies, in a greater measure than at a traditional form, depends on the methods of presentation of educational materials and on the control of students work from the side of teachers and frequency of contact with a teacher. Therefore for combining the advantages of traditional studies with the possibilities, given by the introduction of the newest information technologies and facilities of communication the Web - based system for testing knowledge of students has been developed at Ternopil State Economic University (http://www.tane.edu.ua). The primary objective for creating the system is approbation and introduction the computer educational systems and software products, electronic editions of the educational materials placed in an electronic library into the educational process with remote access. Most attention in using Web based system for testing knowledge of students was devoted to its integration into the ordinary forms of studies and into the computer multimedia systems of studies.

2. As economic effect of the introduction of the systems of electronic studies in great extent depends on the quantity and level of students' preparation. the system of testing the students' knowledge was developed and firstly approbated at the Faculty of Finances of Ternopil State Economic University in teaching "Informatics and computer technologies". The intensive use of new achievements of information technologies in a finance and bank sphere created favourable conditions for preparation of specialists of financial business with wide introduction of methods and systems of electronic studies via Internet. It provides the students with considerable increase of efficiency of studies, and also mastering skills necessary for subsequent increase of level of professional preparation [1].

1. При традиційній формі навчання основна роль в організації навчального процесу відводиться викладачу, а в системах електронного навчання ступінь керування навчальним процесом рівномірно розподіляється між викладачем і студентами, або повністю підпорядковується останнім. Це дозволяє студентами самостійно навчатися у зручний час та у віддаленому від викладача місці. Але ефективність самостійного навчання, в більшій мірі ніж при традиційній формі, залежить від способів подання учбових матеріалів та від контролювання роботи студентів зі сторони викладачів і частоти контактування з викладачем. Тому для поєднання переваг традиційного навчання з можливостям, закладеними у впровадженні новітніх інформаційних технологій і засобів комунікації в Тернопільському Державному Економічному Університеті (http:// www.tane.edu.ua) було розроблено Web - базовану систему тестування знань студентів. Головною метою створення системи є апробація та впровадження комп'ютерних навчальних систем та програмних продуктів, електронних видань навчальних матеріалів, розміщених в електронній бібліотеці з віддаленим доступом, в навчальний процес. Найбільшу увагу при використанні Web – базованої системи тестування знань студентів було приділено її інтеграції зі звичайними формами навчання та з комп'ютерними мультимедійними системами навчання.

2. Оскільки економічний ефект від впровадження систем електронного навчання багато у чому залежить від чисельності і якісного складу контингенту студентів, через це розроблена система тестування знань студентів була в першу чергу апробована на Факультеті Фінансів Тернопільського Державного Економічного Університету при вивченні курсу "Інформатика і комп'ютерна техніка". Інтенсивне використання нових досягнень інформаційних технологій у фінансово-банківській сфері створило сприятливі умови для підготовки фахівців фінансової справи із широким запровадженням

The characteristic peculiarity of the developed Webbased system of testing the students' knowledge is that it is aimed at the students interested in the effective obtaining of knowledge, conscientious implementation of the duties, in the complete use of those informative possibilities, which are given to them by the system for self-perfection, exposure to persistence, good organization, ability to work independently and have skills of work with a computer and telecommunication facilities. The system helps students to realize that success of their studies, level of knowledge, their use in future depends only on them. However there are some objective factors which make the process of using the system specific. For example, it was discovered that the students of senior courses were more active, and more regular participants of the process of self testing in comparison with their junior colleagues.

3. Using Web – based system of testing knowledge in teaching "Informatics and computer technologies" allows the students to understand managing principles, peculiarities, rules and criteria of modern educational technologies better. Besides for the providing of comprehensive analysis of the level of preparation and estimation of knowledge, the system allows the teachers to better adapt the educational programs to the base level of preparation of students. In particular, the system includes computer complexes with program facilities of the educational purpose, which allow the students:

- to provide self-control and correction of educational and cognitive activity;
- to form experience of making optimum decisions in the conditions of limited time;
- to obtain modern methods of e-education and to increase interest to the process of studies;

the teachers:

- to individualize the approach and differentiate the process of studies;
- to control the level of mastering the material, with diagnosis of mistakes and reverse link with lecture material;
- to conduct laboratory and practical classes, to do experiments and investigations in the conditions of virtual reality;
- to shorten time for labour-intensive and long-term calculations of statistical results of carried out laboratory and practical tasks.

The Possibilities of Web - based system of testing knowledges substantially simplify the task of conducting laboratory and practical classes by using multimedia technology, imitation design, Internet technologies. The system allows to realize educational methods which are very difficult or sometimes impossible to apply in ordinary conditions. The developed tests are well adjusted for self-control and very useful for individual classes.

The peculiarity of compiling test is their modularity which gives teachers the possibility to form educational plans from a set of independent educational courses - modules which correspond to the necessities of the individual or group needs of students (Fig.1). By using Internet technologies the system allows to conduct the operative questioning of методик і систем електронного навчання через мережу Internet. Це забезпечує студентам значне підвищення ефективності навчання, а також оволодіння навиками, необхідними для подальшого підвищенні рівня професійної підготовки [1].

Характерною особливістю розробленої Web базованої системи тестування знань студентів є те, що вона розрахована на студентів, зацікавлених в ефективному здобутті знань, сумлінному виконанні своїх обов'язків, в повному використанні тих інформаційних можливостей, що надаються їм системою для самовдосконалення, виявлення наполегливості, організованості, уміння працювати самостійно і мати навички роботи з комп'ютером і телекомунікаційними засобами. Система допомагає студентам усвідомити, що лише від них залежить успіх їх навчання, рівень знань, які вони здобудуть і майбутнє їх використання. Проте є ряд об'єктивних факторів, які роблять процес використання системи специфічним. Наприклад, було виявлено, що студенти старших курсів більш активні, є регулярнішими учасниками процесу самотестування у порівнянні з їх молодшими колегами.

3. Використання Web – базованої системи тестування знань при вивченні курсу "Інформатика і комп'ютерна техніка" дозволяє студентам краще зрозуміти управляючі принципи, особливості, правила і критерії сучасних освітніх технологій. Крім забезпечення всестороннього аналізу рівня підготовки та оцінки знань система дозволяє викладачам краще адаптувати навчальні програми до базового рівня підготовки студентів. Зокрема, система включає в себе комп'ютерні комплекси з програмними засобами навчального призначення, які дозволяють для студентів:

- забезпечити самоконтроль та корекцію учбовопізнавальної діяльності;
- формувати досвід прийняття оптимальних рішень в умовах обмеженого часу;
- дозволяє оволодіти сучасними методами електронного навчання та підвищити інтерес до процесу навчання;

для викладачів:

- індивідуалізувати підхід і диференціювати процес навчання;
- контролювати рівень засвоєння матеріалу, із діагностикою помилок та зворотнім зв'язком з лекційним матеріалом;
- проводити лабораторні і практичні заняття, робити експерименти і досліди в умовах віртуальної реальності;
- скоротити час на трудомісткі та довготривалі обрахунки статистичних результатів виконання лабораторних і практичних завдань.

Можливості Web – базованої системи тестування знань істотно спрощують задачу проведення лабораторних та практичних занять за рахунок використання мультимедіа-технологій, імітаційного моделювання, Internet-технологій і т.д. Система дозволяє реалізувати навчальні методи, які у звичайних умовах застосувати

⊨Back • → • 🙆 🕼 🚮					
	🛛 🐼 Search 🔛 Favorites 🐨 Media 🎲 🖾 🖕 🍰 🖂				
ddress 🙋 http://localhost/stud	d_test/get_subj.php				
Вибір теми для тест	v				
In the second se					
	Лисенко 0.0.				
	Литвин 1.С.				
	Xoma H.F.				
Виберіть автора					
	*не встановлено				
	I MY TANK I A MY TANK AN AVY TANK				
та тему для тесту	KHORKA MS Excel				
та тему для тесту	кнопки MS Excel Пакети прикладних програм				

Fig. 1. Web-page of a choice of a teacher and a module for testing. Рис. 1. Діалогова web-сторінка вибору викладача та модуля для тестування.

students' knowledge and to make up a protocol in which the date, time and results of the conducted testing, that shows the level of mastering of new material during practical seminars, laboratory classes and during independent preparation is stated (fixed) (Fig.2).

The developed system can be used for the current and final control of results of mastering material and for control of preparation for laboratory and practical classes, conducting of test papers, self-control of students' knowledge, final analysis, determination of rating.

At the development of Web – based system for testing knowledge of students the experience of introduction of the e-education systems in Centers of electronic studies at the universities of the USA was used, the level of development of e-education in the countries of Western Europe and Australia was studied. Very interesting and useful was acquaintance with the First Russian Center of Computer Studies at Moscow State Technical University (www.specialist.ru). The peculiarity of this Center is the possibility of payment for studies via Internet which is provided by CyberPlat, the universal integrated system of payments. In relation to Ukraine we can state that such electronic systems of education have not taken deserving place in education yet. Historically in major cases electronic education in Ukraine is used as a mean of delivery of the printed matters by e-mail to the students, which fulfill the task and instructions which are contained in these materials [2]. Only some time ago the e-education via Internet was introduced in Ukraine. The examples are the Center of distance learning in Ivano-Frankivsk (teaching is conducted by the Ukrainian-Canadian business-center - http//www.sbedif.if.ua.), and also the UniNet Center, set up at Kyiv National University (www.uninet.kiev.ua). One of the aims of these Centers is to coordinate actions of the interested organization, physical and legal bodies in the development of e-education. Basic Centers of distance learning in Ukraine are represented on the web-page of Ministry of Education and science of Ukraine (http://www.mon.gov.ua):

- udec.ntu-kpi.kiev.ua Ukrainian Center for Distance Learning;
- www.udl.org.ua Ukrainian System for Distance Studies;

Тема:	команди MS Excel		
рупа:	Φ-22		
1	Андронішин Олег Степанович	05.06.2003 11:22	3/5
		29.05.2003 11:10	4/5
2	Баран Андрій Григорович	19.06.2003 12:59	2/5
		22.05.2003 12:05	5/5
3	Баранчук Альона Ігорівна	19.06.2003 13:16	2/5
4	Бортник Марина Володимирівна	29.05.2003 11:37	3/5
5	Вовчук Іванна Дмитрівна	22.05.2003 11:55	3/5
6	Галич Олег Михайлович	19.06.2003 13:08	5/5
		05.06.2003 12:44	5/5
		29.05.2003 11:20	4/5

Fig.	2.	Web-page	with the	e protocol	of som	e results	s of tes	sting.
Рис	. 2	. Web-ctopi	нка з п	ротоколо	ом резу	пьтатів т	гестуе	ання

дуже складно або взагалі неможливо. Розроблені тести добре пристосовані для самоконтролю і дуже корисні для індивідуальних занять.

Особливістю побудови тестів є їх модульність – що надає викладачам можливість формувати навчальний план із набору незалежних навчальних курсів - модулів, які відповідають індивідуальним або груповим потребам студентів (Рис.1). Використовуючи технології Internet система дозволяє, провести оперативне опитування знань студентів та формулює протокол в якому фіксує дату, час та результат проведеного тестування, що показує рівень засвоєння нового матеріалу на практичних семінарах, лабораторних роботах та під час самостійної підготовки (Рис.2).

Розроблена система може бути використана для поточного та рубіжного контролювання результатів засвоєння матеріалу та для тестового контролю підготовки до лабораторних та практичних робіт, проведення контрольних робіт, самоконтролю знань, підсумкового аналізу, визначення рейтингу.

 При розробці Web – базованої системи тестування знань було використано досвід застосовування систем електронного навчання у Центрах електронного навчання при університетах США, вивчено рівень розвитку електронного навчання у країнах Західної Європи та Австралії. Цікавим і корисним було ознайомлення з Першим російським Центром Комп'ютерного Навчання при МДТУ ім. Н.Е.Баумана (www.specialist.ru). Особливістю цього центру є можливість оплати навчання через мережу Internet, що забезпечується універсальною міжбанківською інтегрованою системою платежів CyberPlat. Стосовно України, доводиться констатувати той факт, що такі електронні системи навчання ще не зайняли гідного місця у освіті. Історично в більшій частині випадків електронне навчання в Україні використовується, як засіб розсилання друкованих матеріалів електронною поштою студентам, які виконують завдання та інструкції, що містяться в цих матеріалах [2]. Нещодавно почали запроваджувати і навчання через Internet, прикладом може бути Центр дистанційного навчання, що знаходиться в Івано-Франківську (навчання проводиться за допомогою Українсько-канадського МБЕРІФ-Бізнес-центру – http://www.sbedif.if.ua.), а також створений при Київському Національному

- www.uapa-dlc.org.ua World Bank Center for Distance Studies;
- http://iuf.ntu-kpi.kiev.ua Distance Learning Courses at International University of Finances.

The electronic educational systems of these Centers are developed on the basis of Internet technology and allow the students of all forms of studies to promote efficiency of individual preparation. The use of Internet by these Centers allows to create virtual classes for students which are in different regions of Ukraine.

Conclusions

For the creation of systems of electronic studies, it is necessary to attract specialists with experience in technology and in educational methods which are well-informed with new education facilities of informative technologies. Such approach allows to solve problems of adaptation of the methods of studies, which is the primary task in the process of introduction of the systems of electronic studies.

The systems of electronic studies provide realization of the newst and the most effective forms of studies from the economic point of view. An economic effect from introduction of such systems shows up in better usage of experience of teaching staff, educational areas, technical and program facilities.

Reference:

- Danylchenko M, Lytvyn I. Technologies of electronic distance learning of specialists in financial business via Internet. Papers to the international scientific and methodological conference "Experience and ways of improving scientific and methodological providing of the process of preparation of specialists in financial and bank business and in taxation" Svalyava 2004.
- Lytvyn I. Methods and means of teaching informational society. Collection of papers to scientific Conference "Information technologies in economy, management and business ", Kiev: European University 2005.

Університеті імені Тараса Шевченка Центр UniNet (www.uninet.kiev.ua). Однією з цілей цього Центру є узгодження дій зацікавлених організації, юридичних та фізичних осіб. Основні Центри дистанційної освіти України представлені на офіційій web-сторінці Міністерства Освіти і науки України (http://www.mon.gov.ua):

- udec.ntu-kpi.kiev.ua Український центр дистанційної освіти
- www.udl.org.ua Українська Система Дистанційного Навчання
- www.uapa-dlc.org.ua Центр Дистанційного Навчання Світового Банку
- http://udec.ntu-kpi.kiev.ua/- НТУУ "КПІ"
- http://iuf.ntu-kpi.kiev.ua- Центр Дистанційного Навчання міжнародного фінансового університету.

Електронні навчальні системи цих Центрів розроблено на основі технології Internet і дозволяють студентам усіх форм навчання підвищити ефективність індивідуальної підготовки. Використання Центрів за допомогою мережі Internet дозволяє створювати віртуальні класи для студентів, що знаходяться в різних регіонах Українах, показало їх високу ефективність та надійність,.

Висновки

Для створення електронних систем навчання, необхідно залучати фахівців з досвідом роботи як у галузі освітніх методик так і обізнаних з новими засобами інформаційних технології. Такий підхід дозволяє вирішувати проблеми адаптації методик навчання, що є першочерговим завданням у процесі впровадження електронних систем навчання.

Системи електронного навчання забезпечують реалізацію найновітніших і найбільш ефективних з економічної точки зору форм навчання. Економічний ефект від впровадження таких систем проявляється у кращому використанні досвіду викладацького складу, навчальних площ, технічних та програмних засобів.

Література

- Данильченко М.Г., Литвин І.С. Технології електронного дистанційного навчання фахівців із фінансової справи через мережу Internet. Матеріали міжнародної науковометодичної конференції "Досвід і шляхи вдосконалення наукового та методичного забезпечення процесу підготовки фахівців із фінансової, банківської справи та оподаткування " м. Свалява, 2004р.
- Литвин І.С. Методи і засоби навчання інформаційного суспільства. Збірник праць наукової конференції "Інформаційні технології в економіці, менеджменті і бізнесі", Київ:-Європейський Університет. -2005.

DECISION SUPPORT SYSTEM FOR DIFFERENTIAL DIAGNOSIS OF LOCAL EDEMA AND URTICARIA WITH ALLERGIC BACKGROUND

SYSTEM WSPOMAGANIA DECYZJI W DIAGNOSTYCE RÓŻNICOWEJ OBRZĘKU MIEJSCOWEGO I OBRZĘKO-POKRZYWKI O PODŁOŻU ALERGICZNYM

www.hae.org.pl

GRZEGORZ PORĘBSKI**, PIOTR WALECKI*, ELŻBIETA WARDACH***, KRYSTYNA OBTUŁOWICZ**

* Department of Bioinformatics and Telemedicine, Collegium Medicum – Jagiellonian University, Kopernika 17, 31-501 Krakow, Poland ** Departament of Clinical and Enviromental Allergollogy, Collegium Medicum – Jagiellonian University, st. Sniadeckich 10, 31-531 Krakow, Poland *** Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Reymonta 4, 30-060 Krakow, Poland

Abstract: This paper presents a web-based decision support system (DSS) for differential diagnosis of local edema and urticaria with allergic background. This decision support system is a simply Clinical Practice Guidelines Online (CPGO) which refers to the Evidence Based Medicine (EBM) methodology. The system is integrated with Polish Association for Helping People with Angioedema website in the e-learning materials section.

Key words: decision support system (DSS), clinical guidelines, allergy, local edema, urticaria Streszczenie: W pracy zaprezentowano system wspomagania decyzji w diagnostyce różnicowej obrzęku miejscowego i obrzęko-pokrzywki o podłożu alergicznym, który dostępny jest za pośrednictwem sieci internetowej. Przedstawiany projekt jest przykładem prostego systemu wytycznych praktyki klinicznej, które oparte są o metodologię EBM. System jest zintegrowany z materiałami edukacyjnymi udostępnionymi na stronie Polskiego Stowarzyszenia Pomocy Chorym z Obrzękiem Naczynioworuchowym.

Słowa kluczowe: system wspomagania decyzji (DSS), ścieżki kliniczne, alergia, pokrzywka, obrzękopokrzywka

Introduction

Decision support system (DSS) is defined broadly as "a computer-based system that aids the process of decision making". DSS is a computerized system for helping to make decisions, which it is usually a choice between alternative decisions based on estimates of the values of those alternatives. A DSS can take many different forms, because the concept of decision-making is very broad and there is a wide range of domains in which decisions are made. In practice, DSS refers to computer applications that perform such a supporting role.

DSS is extensively used in clinical decision support systems for medical diagnosis. Clinical or practice guidelines are "user-friendly statements that bring together the best external evidence and other knowledge necessary for decision-making about a specific health problem" [1]. Good guidelines as "quality-improving strategies" come from Evidence-Based Medicine (EBM), which is the explicit and conscientious use of current best evidence in making decisions about the care of individual patients.

- The five steps of EBM are [2]:
- 1. convert clinical information needs into answerable questions
- 2. track down the best evidence with which to answer them
- critically appraise that evidence for its validity and usefulness
- 4. apply the results of this appraisal in clinical practice
- 5. evaluate your clinical performance

In the project were proposed advisory guidelines for diagnostic and therapeutic decisions, which are for optimal management, respecting situations of individual patients with edema and/or urticaria. It should be emphasized that this system based on inserted data, depicts suggestions of proceeding. However final decision depends on a patient's physician.

Project description

The advisory system for differential diagnosis of local edema and urticaria with allergic background was created based on web technologies. An access to the system is possible through web site and can be linked to medical portals. Actually it is integrated with the Polish Association for Helping People with Angioedema web site (<u>www.hae.org.pl</u>) (fig. 1). This site contains educational materials about angioedema in Polish and Ukrainian (fig. 2):

- clinical symptoms
- classification and pathogenesis
- diagnosis of C1-inhibitor deficiency
- treatment
- patients initiatives

The system has an open structure. It means that because of built-in editor any advisory path to a particular medical issue can be created. A physician, without help of a computer scientist, is able to manage with creating new advisory systems (fig. 3). The system was programmed with PHP script language and XML.

Medical background

Located edema and urticaria are common causes of suffering, one of the most frequently observed pathological symptoms. They afflict at least once in life of about 20% of general population (i.e. millions of people). These symptoms meet each physician in everyday practice, and its diagnosis is usually long-lasting and laborious. It loads the patient as well as the health care system. In the 1997 a cost



Fig 1. Screenshot from the system web page

Fig 2. Screenshot from the Ukrainian language web page

Pytanie 6	
Treść:	Czy obrzękowi towarzyszy pokrzywka lub świąd? Zapisz
	ТАК
-	
Końcowa odpow	edž:
	TAK <mark>Czy występuje obustronny obrzęk kończyn dolnych?</mark> > NE
Nie jest to obrzęk n	jscowy TAK < Czy występuje obustronny obrzęk okołooczodołowy?> NE
	lie jest to obrzęk miejscowy, rozważ: choroba nerek TAK < Czy występuje obrzęk jednej kończyny dolnej lub górnej?>
	TAK < Czy obrzęk utrzymuje się co najmniej tydzień?> NE
	TAK e Czy obrzekowi towarzyczy nokrzywka lub świad? SMF
	NECTOR AND A STREET
	THIS GO ULIQUIOSLYK

of its treatment in out-patient clinics was valued in seven European countries (Belgium, France, GB, Greece, Italy, Spain and Portugal) at 936 mln EUR (with 6,5 mln of visits and treatment expenses). Within the confines of the project arose the system designed for the edema and urticaria differential diagnosis based on anamnesis, physical examination, and additional laboratory examinational results.

Localized (local) edema – increase in soft tissue capacity, because of higher vascular permeability; not associated with too many body regions, but to specific organ or subcutaneous tissue.

Primal eruptions urticaria and angioedema (in connective tissue – especially lips, tongue, periorbital region, hands, scrotum) (see fig. 4).



85

Fig. 4. Pictures present samples of urticaria and angioedema on hands, eyelid and lips.

 Table 1. Advisory path of the differential diagnosis of local edema and urticaria

4. Does a face edema appear? FIRST STEP: Exclude of causes not connected with immunological system response. **YES** \rightarrow Go to the question 4a $NO \rightarrow Go$ to the question 6 1. Does a bilateral edema of lower limbs is appearing? 4a. Does the edema kept at least one week? **YES** \rightarrow This is not a local edema, consider: - Heart failure (worsening of symptoms after a few hours **YES** \rightarrow Consider: in stand position, usually accompanying short of breath Vena cava superior syndrome (chest X-ray: tumor and/ after exertion, hypertension) or lesion in mediastinum) Occlusion of vena cave inferior (symptoms of tumor in Occlusion in duct of parotid salivary gland abdominal cavity, ascites) Liver cirrhosis (decrease in albumin level in plasma, $\textbf{NO} \rightarrow \text{Go}$ to the question 5 ascites, alcoholism in history) Hypothyreosis **SECOND STEP:** Typical etiological factors $\textbf{NO} \rightarrow \text{Go}$ to the question 2 2. Does a bilateral, periorbital edema is appearing (espe-5. Does urticaria or itch accompanies the edema? cially in the morning)? **YES** \rightarrow Go to the question 6 $\textbf{YES} \rightarrow \textbf{This} \text{ is not a local edema, consider:}$ Kidney disease $NO \rightarrow Consider:$ 3. Does appeared an edema of one lower or upper limb? Hereditary angioedema - HAE (inherited C1-inhibitor deficiency): $\textbf{YES} \rightarrow \text{Go}$ to the question 3a Clinical features: self-limiting, $\textbf{NO} \rightarrow \text{Go}$ to the question 6 non-inflammatory subcutaneous angioedema, recurrent and usually lasting more than 12 hrs 3a. Does the edema kept at least one week? abdominal pain without clear organic etiology, recurrent and usually lasting more than 6 hrs **YES** \rightarrow Consider: • recurrent laryngeal edema Thrombosis of profound veins (typically: pain in calf or thigh, local edema and pain at pressure, risk fac-Laboratory features: tors: surgery, pregnancy, immobilization, neoplastic • C1 inhibitor antigenic level <50% of normal in badisease, increase in blood clotting) Lymphatic edema (primary, occlusion by tumor, at limb sal conditions (two separate determinations with paralysis, after mastectomy; typically: long-lasting, patients > 1 years old) painless, large) C1 inhibitor functional level <50% of normal in ba-Local / mechanical impairment sal conditions (two separate determinations with Rheumatological disease (symptoms in regions of arpatients > 1 years old) ticulations) at least 2 of 3 clinical features and one laboratory features are need to diagnose HAE $NO \rightarrow Go$ to the question 5

- Treatment of acute HAE attack: • C1-INH concentrate (Berinert P), fresh frozen plasma (when C1-INH concentrate is unavailable) Long-term prophylaxis: attenuated androgens (ie. Danazol) and antifibrinolytic agents avoidance of exogenous estrogens and ACE-inhibitors More information at http://www.haei.org/ Aquired angioedema Symptoms, C1 inhibitor antigenic and functional level similar as in HAE, but low level of C1q Further diagnostics: lymphoproliferative diseases (from benign gammapaty to malignant lymphomas) neoplastic disease or chronic infections **Others causes** \rightarrow Go to the question 6 6. Does appeared an urticarial eruptions / keeping in the same place for 24h and localized mostly at lower limbs? **YES** \rightarrow Consider a diagnosis in *urticaria vasculitis* direction: - Other symptoms are possible: pain in joints, slightly higher body temperature - Elevated level of Sedimentation Rate and/or CRP (C-reactive protein) Lower level of C1q and/or CH50 is possible \rightarrow indication to diagnostics of systemic disease Skin lesion biopsy: infiltration of multinuclear leukocvtes in blood vessels wall Management:

- Treatment: antihistaminic, anti-inflammatory (corticosteroids, colchicin, dapson et al.)
- Additional examination of kidneys, gastrointestinal tract, eyes for systemic disease detection

 $NO \rightarrow Go$ to the set of questions 7 (detailed anamnesis)

7. Detailed anamnesis is crucial because of high number of etiologic factors in edema and anamnesis

7.1. Does the symptoms of edema and/or anamnesis appeared after an insect sting or bite (by bee, wasp, and mosquito)?

YES \rightarrow Consider additional diagnostics:

- Skin prick tests
- Specific IgE

After the diagnosis confirmation:

- protection of a patient with antihistamines drug and if need be self-administered adrenalin
- in particular cases specific immunotherapy is possible

Check also $\rightarrow 7.2$

 $NO \rightarrow Go$ to the question 7.2

7.2. Does the symptoms of edema and/or anamnesis appeared after touch same substances or things (i.e. veg-

etables, fruits or drugs in ointments like *neomycin*, cosmetics, and products containing latex)?

- $\textbf{YES} \rightarrow \textbf{Consider}$ additional diagnostics:
 - Skin prick tests,
 - Specific IgE
 - Patch tests

After the diagnosis' confirmation:

- Protection of the patient with antihistamines
- Avoiding an exposition to harmful factors

Check also -> 7.3

 $NO \rightarrow Go$ to the question 7.3

7.3. Do the edema and/or urticaria appear in parts subjected to pressure (as: after carrying a shopping bag, a backpack, screw driving, on knees at long kneeling ones, on foots at wearing tight shoes ones)?

- $\textbf{YES} \rightarrow \textbf{Consider}$ additional diagnostics:
 - Dermographism
 - Pressure skin tests

This type of physical edema and urticaria appear usually at age of about 40, was not characterized at children.

Eruptions may appear directly or few hours after a stimuli effect (delayed pressure urticaria) and last a dozen hours or so.

After the diagnosis' confirmation:

- Protection of the patient with antihistamines
- In patients with delayed symptoms NSAIDs are used

Check also $\rightarrow 7.4$

 $\textbf{NO} \rightarrow \text{Go}$ to the question 7.4

4.5. Do the edema and/or urticaria symptoms appear during an exercises or when the person is sweated?

YES \rightarrow Consider supplement diagnostic:

- An exercise test causing new eruptions
- A bath in water of 42 degrees C through 15 minutes

Cholinergic urticaria and edema appear mainly at young men, much seldom at women.

Eruptions: small wheals (diameter about 1-3 mm) with red round border may initially appear on the sides of fingers on hands.

After the diagnosis' confirmation:

- Protection of the patient with antihistamines
- As additionally are sometimes used neuroleptics
- Caution: the symptoms appearance is possible only after food intake or drugs causing allergy preceding exercises.

Check also $\rightarrow 7.5$

 $NO \rightarrow Go$ to the question 7.5

7.5. Do the edema and/or urticaria appear after a cooling of the skin?

 $\begin{array}{lll} \textbf{YES} \rightarrow \text{Consider supplement diagnostics:} \\ & - & \text{Putting an ice cubs in plastic bag on skin or immers-} \end{array}$

ing forehand in water of about 20 C degrees for 10 minutes

- Wheals began to appear after the stimuli remove, while the skin is warming up.
- Negative provocation test does not exclude cold urticaria and edema, because at some patients larger parts need to be cooled to evoke the symptoms.

This type of urticaria characterized at children.

Eruptions might be caused by, i.e.:

- Leaving a home in winter
- A bath in cold water
- Eating ice-creams
- Strong cool wind

After the diagnosis confirmation:

- Protection of the patient with antihistamines (cyproheptadin – first choose drug)
- Possible tolerability development to cold can be produce (immersing in turn lower and upper limbs in cool water)

Check also $\rightarrow 7.6$

 $\textbf{NO} \rightarrow \text{Go}$ to the question 7.6

7.6. Do the symptoms of edema and/or urticaris appear after intake of NSAIDs (non-steroids anti-inflammatory drugs)?

YES \rightarrow Consider supplement diagnostics:

- Oral provocation test
- Diagnostic elimination diet

So called "aspirin urticaria" can be connected with aspirin triad (bronchial asthma, nasal polyps, aspirin hypersensitivity). Eruptions may appear after different drugs of NLPZ (crosshypersensitivity).

After the diagnosis confirmation:

- Protection of the patient with antihistamines
- Low-salicylic diet (if the symptoms appear after food intake as well)
- Not administer NSAIDs (but paracetamol is allowed) in particular cases desensitization is possible
- Caution: NSAIDs can only exacerbate previous skin lesions

Check also $\rightarrow 7.7$

 $\textbf{NO} \rightarrow \text{Go}$ to the question 7.7

7.7. Do the symptoms of edema and/or urticaris appear after intake of ACE-I (angiotensin convertase inhibitors)?

YES \rightarrow Consider supplement diagnostics:

- Remission after withdrawal ACE-I
- Oral provocation test

Urticaria and edema eruptions appear after different drugs of ACE-I (cross-hypersensitivity).

Occurrence 0,1-0,5% patients treated with drugs of this group, usually at head and neck, more often in Afro-Americans, recurring irregularly.

After the diagnosis confirmation:

- Withdrawal ACE-I (even after e few years of intake) after withdrawal symptoms can persist a few weeks
- Caution: first edema attack can occurs from a few hours to 8 years (sic!) after first ACE-I intake (after non-symptomatic period, without any adverse events)
- It is supposed that angiotensin II receptor blockers i.e. *losartan* can be used instead of ACE-I, because they do not elevate bradikinin level

Check also \rightarrow 7.8

 $NO \rightarrow Go$ to the question 7.8

7.8. Does appear symptoms of edema and/or urticaris depend on estrogens exogenous intake or endogenous production at women?

YES \rightarrow Urticaria and/or edema:

- After contraceptive drug intake or hormonal replacement therapy
- Related to periodic hormonal fluctuations
- During pregnancy (symptoms 14-21 day after conception)

After the diagnosis confirmation:

- Limit estrogen intake
- Caution: family angioedema in women was described (inherited estrogen-dependent angioedema)
- Estrogens can be found in meat

Check also \rightarrow 7.9

 $\textbf{NO} \rightarrow \text{Go}$ to the question 7.9

7.9. Do the symptoms of edema and/or urticaris appear after intake other drugs than ACE-I, NLPZ, estrogens?

- $\textbf{YES} \rightarrow \textbf{Consider}$ supplement diagnostics:
 - Remission after withdrawal ACE-I
 - Oral provocation test
 - Skin tests

Urticaria and angioedema eruptions related to drug evoke more often:

- Antibiotics, mostly beta-lactams, sulfonamides
- Amiodaron, drugs other than ACE-I administered in hypertension (i.e.: calcium channel blockers)
- Psychotropic
- Direct degranulation caused by opiats and contrast media

After the diagnosis confirmation:

- Drug withdrawal
- Protection of the patient with antihistamines
- Caution: antibiotics can be found in meat

Check also \rightarrow 7.10

 $NO \rightarrow Go$ to the question 7.10

7.10. Do the symptoms of edema and/or urticaris appear after eat <u>specific</u> food?

- **YES** \rightarrow Consider supplement diagnostics:
 - Skin tests with food commercial and native allergens
 - Specific IgE determination

- Elimination some products or groups of food suspected to be causative agents for 14 day - observation of symptoms Rice diet and introducing new groups of food every week diary After the diagnosis confirmation: - Elimination of allergen - Protection of the patient with antihistamines - Caution: most common cross-hypersensitivity between food and pollen (i.e.: celery - birch) • OAS oral allergy syndrome: after consumption allergic food - itch, erythema and edema of oral cavity, pharynx, lips, sometimes larynx, urticaria, rhinitis and bronchial asthma attack Check also \rightarrow 7.11 $\textbf{NO} \rightarrow \text{Go}$ to the question 7.11 7.11. Do the symptoms of edema and/or urticaris appear after eat nonspecific food (as food additives)?
- $\textbf{YES} \rightarrow \textbf{Consider}$ supplement diagnostics:
 - Remission during diet not-containing food additives and exacerbation
 - Oral provocation tests
 - Diary

Urticaria and angioedema eruptions related to food evoke:

- Sulphites
- Glutaminiane sodium
- Benzoate acid and benzoates (wide used antifungal and antibacterial chemical agents)

After the diagnosis confirmation:

- Avoiding an exposition to the products contain harmful substances
- Protection of the patient with antihistamines
- Caution: some food evoke symptoms because high level of histamine or histamine-release activity

Check also \rightarrow 7.12

$\textbf{NO} \rightarrow \text{Go}$ to the question 7.12

7.12. Do the symptoms of edema and/or urticaris appear with coexisting infection?

- $\textbf{YES} \rightarrow \textbf{Consider}$ supplement diagnostics:
 - Viral
 - Bacterial
 - Fungal
 - Parasitic infection

Urticaria and angioedema eruptions related to infections:

- Mononucleosis
- Hepatitis A, B, C
- Viral infections of respiratory tract
- Helicobacter pylori infection
- Chronic bacterial infection (i.e.: sinusitis, cholecystitis) fungal infections (especially foot) chronic parasitic infection of gastrointestinal tract

After the diagnosis confirmation:

- Specific treatment
- acute urticaria in children evoke adenovirus, enterovirus, EBV more often

Check also $\rightarrow 8$

 $NO \rightarrow Go$ to the question 8

THIRD STEP: Additional anamnesis and deep diagnostic

8. Does the physical examination (including lymph node) and routine laboratory results (blood cell count, erythrocyte sedimentation reaction, urine analysis, and liver function tests) indicate any abnormalities suggesting etiologic factor of edema and/or urticaria?

 $\textbf{YES} \rightarrow \text{Treatment}$ depend of diagnosed abnormalities and \rightarrow Go to the question 10

 $NO \rightarrow Go$ to the question 9

9. Do urticaria and/or edema occur shorter then 6 weeks?

 $\textbf{YES} \rightarrow \text{Recommend symptomatic treatment of urticaria and/ or edema:}$

- antihistamines
- Il generation (i.e.: loratadine, ceterizine, fexofenadine)
- If is no complete remission use I generation or II generation hydroxizine on demand or max. therapeutic doses
- If optimal dose H1-blockers don't give a appropriate control may use tricyclic antidepresant (i.e.: doxepin)
- combination of antihistamines (questionable from pharmacological point of view)
 - I generation drug
 - Il generation drug
 - Il generation drug morning + I generation drug evening
 - Doxepin + I or II generation drug
 - H1-blocker + H2-blocker
- If no change for the better corticosteroids (in minimal effective doses i.e.: 3 days 40mg of prednisone orally)
 adrenalin in emergency

Fix the term of control visit and when symptoms last \rightarrow See point 10

 $NO \rightarrow Go$ to the point 10

10.a. Conduct next detailed anamnesis and physical examination regarding to rare diseases that can cause urticaria and/or angioedema:

- Limphoprolipherative diseases (lymph nodes swollen, spleen or liver enlargement, loss in weight)
- Chronic inflammatory diseases (abnormalities in joints, kidneys, central nervous system, skin)
- Sarcoidosis, amyloidosis
- Hypoparathyreoidismus
- Systemic or skin mastocytosis
- Chronic hepatitis

10.b. Conduct:

 Intradermal test with autologous plasma (positive result suggests autoantibodies against IgE or IgE receptor presence)

- Antibodies against thyroid peroxidase determination (presence of them without hyperthyreoidismus suggests autoimmunologic etiology of urticaria)
- (i.e.: marrow or skin biopsy, computer tomography, calcium and parathormone level in blood)

If diagnosis is established

- Specific therapy (i.e.: thyroid hormones in autoimmunologic thyroid diseases can cause urticaria remission)
- Symptomatic treatment (see point 9)

If no recognized causes \rightarrow Go to the point 11

11. If no recognized causes of edema and/or urticaria on the basis of diagnostic steps it's necessary to:

- Accept this is an idiopathic edema and/or urticaria
- Periodically conduct new anamnesis
- Inform patient than in some number of ill (5-20%) is possible find cause of chronic edema and/or urticaria

Conclusion

In spite of its implementation and simplicity, the created advisory system meets requirements well. Particularly significant element of the system is an editor designed for easy creation of medical advisory paths of various types. Physicians with the web-based editor might create future advisory systems and clinical guidelines as the needs come.

The example of such system, that answers physicians` real needs, is the created one: differential diagnosis of local edema and urticaria with allergic background. It is available on educational and e-learning site devoted to hereditary angioedema (HAE). This site is a first in Polish (and Ukrainian) language source of HAE most aspects. The system, being available on Internet and thanks to easy creation of proceedings paths and advisory systems, fills in a gap in Polish medical e-learning sites.

References

- Sackett D. L. et al., Evidence-based medicine: How to practice and teach EBM. New York: Churchill Livingstone. Centre for Evidence-Based Medicine – Oxford University, 1997.
- 2. http://medicine.ucsf.edu/resources/guidelines/intro.html
- Porebski G., Obtulowicz K., Obtulowicz P., Walecki P., Roterman-Konieczna I, Czarnobilska E., Krzanowska K., Kwapińska A., Targosz K., Internet as a potential source of information about allergy for patients, The XIXth World Allergy Organisation Congress, June 26 - July 1, 2005, Munich, Germany.
- Lasoń W., Walecki P., Sarapata K., Roterman-Konieczna I., Nauczanie medycyny z wykorzystaniem paradygmatu Medycyny Opartej na Dowodach Naukowych (Evidence Based Medicine), 16. Ogólnopolskie Sympozjum Naukowe nt. "Komputer w edukacji", Kraków, 29-30 września 2006, s.125-130.
- Walecki P., Pyrczak W., Lasoń W., Sarapata K., E-Learning i telemedycyna - problemy strukturyzacji wiedzy, 16. Ogólnopolskie Sympozjum Naukowe nt. "Komputer w edukacji", Kraków, 29-30 września 2006, s. 243-248.
- Porębski G., Walecki P., Obtułowicz K., Czarnobilska E., Krzanowska K., Kwapińska A., Targosz K., Częstość wykorzystywania różnych źródeł wiedzy o alergii przez chorych ze schorzeniami atopowymi, Allergy Asthma Immunology 7th Central and Eastern European Conference June 8-11, 2005 Łódź, Poland.
- Walecki P., Lasoń W., Porębski G., Obtułowicz K., Skórna postać nadwrażliwości na aspirynę – pilotażowy telemedyczny system informacyjno-diagnostyczny, Konferencja "Modelowanie Cybernetyczne Systemów Biologicznych" MSCB 2005.
- Walecki P., Porebski G, Obtułowicz K., Zastosowanie logiki niemonotonicznej w systemie doradczym dla diagnostyki różnicowej schorzeń alergicznych. Alergol. Immunol., 2006 T. 3, nr 1-2, s. 10-13.
- Walecki P., Lasoń W., Pyrczak W., Roterman-Konieczna I., Sarapata K., Szpitalny System Informacyjny (HIS) oparty na metodologii EBM, w: Pozyskiwanie wiedzy i zarządzanie wiedzą, M. Nycz, M. Owoc, red., Prace naukowe Akademii Ekonomicznej we Wrocławiu, 2006, s.179-188.
- Lasoń W., Walecki P., Pyrczak W., Sarapata K., Roterman-Konieczna I, Medycyna Oparta na Dowodach (EBM) – nowe podejście optymalizacyjne wykorzystujące systemy komputerowe, W: Tadeusiewicz R., Ligęza A., Szymkat M. (red): Computer Methods and Systems, 2005: 455-462.



E-learning

THE ASSESSMENT OF REHABILITATION PROGRESS OF PATIENTS WITH RHEUMATOID ARTHRITIS ON THE RHEUMATOLOGY WARD IN PROVINCIAL HOSPITAL NO. 2 IN RZESZÓW

OCENA POSTĘPU USPRAWNIANIA CHORYCH Z RZS LECZONYCH NA ODDZIALE REUMATOLOGII SZPITALA WOJEWÓDZKIEGO NR 2 W RZESZOWIE

KRZYSZTOF STYCZEŃ*, MARIUSZ DRUŻBICKI* **

 * Rzeszów University, Institute of Physiotherapy, Address: ul. Warszawska 26A, 35-205 Rzeszów http://www.univ.rzeszow.pl/fizjoterapia/index.html styku1@gazeta.pl
 ** Provincial Hospital no. 2 in Rzeszow Address: Lwowska 60, 35-301 Rzeszów www.szpital2.rzeszow.pl, sekretariat@szpital2.rzeszow.pl

Abstract

OBJECTIVE: Rheumatoid arthritis (RA) is a chronic, inflammatory disease of skeleto-muscular system, leading to a progressive disability. People all over the world are affected by rheumatoid arthritis. In Poland it is more common in women than in men (1:3). The level of disability depends on many factors like: age, sex, duration of disease, intensity of pain, activity of the inflammation process, grade of joint destruction, psychical condition, economical and social conditions. Like in other diseases process and peculiarity are individual and very varied. Progress of rheumatoid arthritis is gradual. Functional restriction affects all aspects of patient's life like: work, free time, relations in the family, physiology, isolation and progressive depression. From the social aspect rheumatoid arthritis and other joint inflammatory diseases are a constant burden for health protection budget. Long-term high activity of inflammatory process, especially in the first years of disease duration, can strongly deteriorate life quality. With the time progress in joint destruction begins to play dominant role in the disability development. Rehabilitation combined of pharmacotherapy, kinesitherapy and physiotherapy plays a key role in rheumatoid arthritis treatment. The aims of rehabilitation are: reduction of the subjective pain assessment, increasement of the joint movement range and muscle strength, improvement of patient's functionality and stopping or regression of disease progress by pharmacotherapy. Pharmacotherapy and surgical therapy, which are started in a most suitable moment, are also very important in RA treatment. METHODS: The aim of our study is estimation of rheumatoid arthritis patient's rehabilitation progress at the rheumatology department at Provincial Hospital No. 2 in Rzeszów. Hypothesis of our study that we undertook is: "Complex treatment of patients with rheumatoid arthritis affects the improvement of

their functionality level". This hypothesis was tested on a group

of a 70 patients with a diagnosed rheumatoid arthritis and

hospitalized at rheumatology department at Provincial Hospital No. 2 in Rzeszów. Patients were tested for such factors

Streszczenie

Reumatoidalne Zapalenie Stawów jest przewlekła zapalna choroba układu ruchu, prowadzącą do postępującej niepełnosprawności i inwalidztwa. Na schorzenie to choruja ludzie wszystkich ras na całym świecie. W Polsce zdarza się częściej u kobiet niż meżczyzn (1: 3). Stopień niepełnosprawności zależy od wielu czynników: wieku, płci, czasu trwania choroby, nasilenia bólu, aktywności procesu zapalnego, stopnia zaawansowania destrukcji stawów, stanu psychicznego oraz warunków socialno-ekonomicznych. Jak w każdei chorobie przebieg i specyfika są indywidualne i bardzo zróżnicowane. Przebieg tej choroby charakteryzuje się również przebiegiem etapowym. Ograniczenie wydolności funkcjonalnej wpływa z kolei na wszystkie dziedziny życia pacjenta: pracę zawodową (inwalidztwo), wypoczynek, wzajemne relacje w rodzinie, zaspokajanie potrzeb fizjologicznych, osamotnienie i pogłębiającą się depresję. W aspekcie społecznym Reumatoidalne Zapalenie Stawów i inne zapalne choroby stawów stanowią ogromne i stale wzrastające obciążenie budżetu przeznaczonego na ochronę zdrowia. Rehabilitacja składająca się z farmakoterapii, kinezyaterapii oraz zabiegów fizykoterapeutycznych odgrywa kluczową rolę w leczeniu Reumatoidalnego Zapalenia Stawów. Celami rehabilitacji są: zmniejszenie subiektywnego odczucia bólu przez pacjentów, powiększenie ruchomości stawów i siły mięśniowej, polepszenie funkcjonalności pacjentów oraz zatrzymanie lub cofniecie postępu choroby poprzez farmakoterapię. W leczeniu RZS równie ważna jest farmakoterapia oraz leczenie chirurgiczne, które są wdrażane w odpowiednim momencie terapii.

Celem naszej pracy będzie ocena postępu usprawniania chorych z RZS leczonych kompleksowo na Oddziale Reumatologii Szpitala Wojewódzkiego nr 2 w Rzeszowie. Hipoteza, którą przyjęliśmy brzmi: "Kompleksowe leczenie chorych na Reumatoidalne Zapalenie Stawów wpływa na poprawę stanu funkcjonalnego". Hipoteza ta była sprawdzana w grupie 70 osób z rozpoznaniem Reumatoidalnego Zapalenia Stawów hospitalizowanych na Oddziale Reumatologii Szpitala Wojewódzas: the movement range of joints in lower limbs, muscle strength in muscle groups in lower limbs, subjective paint assessment, disease stage and functionality grade.

RESULTS: After ending the research a significant decrease of subjective pain assessment was noticed, also unimportant improvement of the joint movement range and muscle strength were noticed. Disease stage and functionality grade practically did not change.

CONCLUSIONS: On the basis of the carried out research we can say that subjective assessment of the complex treatment and rehabilitation was overestimated by the patients as the objective researches show that, for example: the range of joints movement or the muscle strength improvement were unimportant. This might be a result of an enormous decrease of patient's pain assessment, which in patient's estimation was a confirmation of an effective complex treatment.

Key words: rheumatoid arthritis, pain, rehabilitation, physiotherapy, movement range, functionality, efficiency.

1.1 Definition of rheumatoid arthritis

Rheumatoid arthritis (RA) is a chronic, progressive inflammatory and systemic disease, which is characterized mainly by joints involvement. The inflammation of the synovial membrane leads to destruction of bones and cartilages. Inflammatory process develops also in structures near the joint like ligaments and fascias where it causes damages. Besides the joints type of RA (which is the most common) there are other types of RA which affect subcutaneous tissue (rheumatoid nodes), vessels, nervous system, heart and kidneys [6,12].

1.2 Epidemiology of rheumatoid arthritis

Rheumatoid arthritis is one of the most common systemic disease of connective tissue. Morbidity rate in Poland is estimated at 1% of grown up population (there are about 400.000 people affected). Incidence of disease is 2-4 cases per 10.000 people, so it is about 8-16 thousand new ill patients in Poland a year. This disease is more common in women than in men (3-4:1), it might begin at random age, but the top of its occurrence is between the ages of 30 and 60. RA can have a chronic or periodic character with periods of exacerbation and remission. In some cases there might be a possibility of a spontaneous remission in a very early phase of disease. There is also a severe, progressive form of the disease which within a few years leads to invalidism and progress of this form is impossible to modify despite the medical treatment. This form occurs in about 20-30% of patients. Within 20 years a high grade of invalidism is observed. It is at a level of 80% in hospitalized patients and about 20% in people's population [6,12].

1.3 Diagnosis and diagnostic criteria of rheumatoid arthritis

The criteria are as follows:

 morning stiffness in and around joints lasting at least 1 hour before maximal improvement; kiego nr 2 w Rzeszowie. U pacjentów badano takie czynniki jak: zakres ruchu w stawach kończyny dolnej, siłę mięśniową grup mięśniowych kończyny dolnej, subiektywną ocenę stopnia bólu oraz stadium choroby i stopień funkcjonalności.

Po przeprowadzeniu badań zaobserwowano: znaczne zmniejszenie subiektywnego stopnia nasilenia bólu, niewielkie poprawy parametrów w zakresie siły mięśniowej oraz zakresu ruchu w stawach kończyny dolnej. Stadium choroby oraz stopień funkcjonalności pacjentów praktycznie nie uległy zmianie.

W wyniku przeprowadzonych badań można stwierdzić, że subiektywna ocena usprawniania i kompleksowego leczenia pacjentów jest znacznie zawyżona niż badania obiektywne np. zakresu ruchów w stawach czy siły mięśniowej. Jest to spowodowane najprawdopodobniej olbrzymim zmniejszeniem odczuć bólowych pacjentów, co w ocenie pacjentów przekładało się na skuteczność leczenia kompleksowego.

Słowa kluczowe: reumatoidalne zapalenie stawów, ból, rehabilitacja, fizjoterapia, zakres ruchu, sprawność, funkcjonalność.

- soft tissue swelling (arthritis) of 3 or more joint areas observed by a physician;
- swelling (arthritis) of the proximal interphalangeal, metacarpophalangeal, or wrist joints;
- 4) symmetric swelling (arthritis);
- 5) rheumatoid nodules;
- 6) the presence of rheumatoid factor; and
- radiographic erosions and/or periarticular osteopenia in hand and/or wrist joints.

Criteria 1 through 4 must have been present for at least 6 weeks. Rheumatoid arthritis is defined by the presence of 4 or more criteria, and no further qualifications (classic, definite, or probable) or list of exclusions are required [2].

1.4 The aim of the study, research problem and hypothesis

1.4.1 Aim, research issues and hypothesis

For our study we took the assumption that aim is a wanted state of things that the individual wants to achieve in his action.

The aim of the research: the estimation of improvement progress in patients with rheumatoid arthritis made during hospitalization on rheumatoid ward in provincial hospital no. 2 in Rzeszów

The main research problem: What influence does complex rehabilitation have on the hospitalization on rheumatoid ward of patients with rheumatoid arthritis?

Detailed research problems:

- If complex rehabilitation increased mobility range of main joints in lower limbs?
- How did patients pain perception change?
- If the complex rehabilitation increased the strength of main muscle groups in lower limbs?

The main research hypothesis [16]:

 Complex rehabilitation of people with rheumatoid arthritis causes improvement of functional state. The detailed research hypothesis [16]:

- The complex treatment has the influence on mobility range of main joints of lover limbs making patients with RA functioning better.
- The complex treatment has the influence on pain perception.
- Kinesitherapy has the influence on muscle strength increasement

2.0 Material and method

2.1 Material - research group characteristic

The research group was composed of 70 patients with confirmed rheumatoid arthritis hospitalized in the provincial hospital no. 2 rheumatology ward in Rzeszów. There were 16 men, what was 23% of the group and 54 women, what made 77% of the group.



Fig. 1. The research group percentage distribution

The age average of the research group is 64 years – the average age of men is 68 years and women – 63 years.



Fig. 2. Histogram of the research group age

The initial assessment of the disease advance stadium and functional state with the use of the HAQ Disability Score questionnaire showed that:

 The average value of disease advance stadium for the whole research group is 2, for men is 2 and also for women it is 2.

There are 9 patients in the research group in the 1st stadium of disease what is 13% of the whole research group. The amount of men in the 1st stadium of disease is 2 what is

22% of the 1st stadium group. In that group there are also 7 women in the 1st stadium what is 78% of this group.

There are 33 patients in the research group in the 2^{nd} stadium of disease what is 47% of the whole research group. The amount of men in the 2^{nd} stadium of disease is 9 what is 27% of the 2^{nd} stadium group. In that group there are also 24 women what is 73% of this group.

There are 28 patients in the research group in the 3rd stadium of disease what is 40% of the whole research group. The amount of men in the 3rd stadium of disease is 5 what is 14% of the 3rd stadium group. In that group there are also 23 women what is 86% of this group.



Fig. 3. The patent disease stadium percentage distribution

2.2 Methods, techniques and research tools

In the subjective examination of the group there were used:

- the goniometric examination of joint mobility range (hips, knees and jarrets)
- the muscle strength examination the Lovett's test
- the pain level estimation with the use of the ten gradual Likert's test
- the HAQ Disability Score questionnaire

Examination of joint mobility range was carried out according to rules given by Zembaty's [24].

Muscle strength examination was carried out according to modified Lovett's test. Lovett's test scale [24]:

- 1 no tension in active movement attempt the muscle is totally paralyzed
- 2 tonus can be felt in palpation, no active movement is possible - the muscle is functionally defective
- 3 the muscle in a proper position counteracts the gravity forces of the limb – the muscle is functionally efficient.
- 4 the muscle overcomes the resistance of the limb or a part of the limb with the indefinite external resistance
- 5 the muscle can overcome the resistance of the limb or a part of the limb with the sizeable external resistance

Likert's scale – visual Likert's scale is used when there is a need to have patient's estimation of pain, health or disease activity. This scale is presented in a form of a ruler on which patient points a spot. This spot is an estimation of the pain level. On the ruler the 0cm point means insignificant pain intensification and the 10cm point – maximal pain intensification. The higher value of centimeter indicated by the patient, the higher is pain intensification.

The HAQ Disability Score questionnaire – it was elaborated in 1978 by J. Fries and his collaborators from Stanford University. HAQ Disability Index is the most common used part of the questionnaire in the clinical practice. It estimates in a standardized way the disability level of physical fitness in patients with rheumatoid arthritis. The complete questionnaire includes additionally analogue pain estimation and disease activity scale, questions concerning the out of joint aliments, coexistence of other diseases, previous treatment and its side effects and questions about treatment costs.

The HAQ Disability Index questionnaire is divided into eight sections concerning different everyday activity spheres such as:

- 1) washing and getting dressed
- 2) morning getting up
- 3) eating
- 4) walking
- 5) personal hygiene
- 6) lifting
- 7) grabbing
- 8) other activities

The ill person estimates different activities from those sections in the four gradual scale deciding how much difficulties they have during those activities.

The questionnaire grading scale:

- no difficulties 0 points
- with little difficulties 1 point
- with big difficulties 2 points
- can not do the activity 3 points

The highest note in the given section is taken into consideration. In the questionnaire there are also questions verifying if the patients needs additionally special devices or help of another person during activities that were mentioned in the questionnaire. If such kind of help is needed then the section mentioning that activity scores 2 points (if that section has already 3 points, then the highest score is taken into consideration). You need to verify if the device is used only in one activity or is it used in many, for example walking stick can be used in walking (walking section), but might be also used in getting up from a chair (getting up section). On the other side the patient may use walking stick only because he feels safer with it although he has little problem with walking. To get the final questionnaire result all points from all eight sections should be added and then divided into 8. The received result is between 0 and 3 points [3].

The patients during their hospitalization on rheumatologic ward underwent the complex physiotherapy, kinesitherapy and also an individually selected pharmacotherapy. Kinesitherapy was composed of morning gymnastics, allevated activities and individual activities preformed in the remission period by a therapist. The approximate time of kinesitherapy during the day was about 1,5 hour. The physiotherapy was composed of kryotherapy, lasertherapy, hydromassages and ultrasounds. Both kinesitherapy and physiotherapy were adjusted to individual patients because of intolerance of some kinds of therapy by some patients [6,8].

2.3 The research course organisation

Before we began our researches in the group of patients we had adopted the following research organisation plan on the purpose of our work:

- the researches were carried out in the provincial hospital no. 2 in Rzeszów on the rheumatology ward, with the acceptance of the head of the department and every patient
- each patient was examined twice with the use of those same methods and tests – the first examination was carried out in the first or second day of the hospitalisation in the morning hours, the second examination was carried out in the last day of hospitalisation (also in the morning hours), to capture the changes that could have a significant influence on the rehabilitation process of patients with rheumatoid arthritis.

The planned time of examination of a single patient was about 40 minutes [18].

3.0 The research results presentation based on own researches

In the research group the investigations were carried out which were supposed to estimate the progress of rehabilitation of patients with rheumatoid arthritis hospitalised in the provincial hospital no.2 in Rzeszów. The investigations were carried out in the research group twice: once in the first or second day of hospitalisation, twice in the last day of hospitalisation. The investigation was composed of: goniometric mobility range examination and muscle strength examination – the Lovett's test of the bigger joints, HAQ Disability Score questionnaire with was elaborated in internal medicine teaching hospital and rheumatology WIM CSK MON in Warsaw based on Health Assessment Questionnaire by Stanford University School of Medicine, Division of Immunology and Rheumatology and also a pain level estimation with the use of the Likert's test.



Fig. 4. Confrontation of the initial and final examination of internal rotation in the hip joint for men and women (movement rage).

Abbreviations:

```
LLL – left lower limb
RLL – right lower limb
```



Fig. 5. Confrontation of the initial and final examination of external rotation in the hip joint for men and women (movement rage).



Fig. 6. Confrontation of the initial and final examination of flexion in the hip joint for men and women (movement rage).



Description like in fig. 4

Description like in fig. 4



Fig. 7. Confrontation of the initial and final examination of the muscle strength for flexion in the hip joint for men and women.

8 7 6 5 4 3 2 1 0 LLL - M RLL - LLL - F RLL - FLLL - M RLL LLL - FRLL - F М - MST М-- MST - MST MST Initial examination Final examination

Fig. 8. Confrontation of the initial and final examination of extension in the hip joint for men and women (movement rage and muscle strength).



Fig. 9. Confrontation of the initial and final examination of abduction in the hip joint for men and women (movement rage).

Description like in fig. 4

Abbreviations: MST – muscle strength

Other like in fig. 4



Fig. 10. Confrontation of the initial and final examination of the muscle strength for abduction in the hip joint for men and women.

Description like in fig. 4



Fig. 11. Confrontation of the initial and final examination of adduction in the hip joint for men and women (movement rage).

Description like in fig. 4

96



Fig. 12. Confrontation of the initial and final examination of the muscle strength for adduction in the hip joint for men and women.

Description like in fig. 4

- knee joints:



Fig. 13. Confrontation of the initial and final examination of flexion in the knee joint for men and women (movement rage).

Description like in fig. 4



Fig. 14. Confrontation of the initial and final examination muscle strength for flexion in the knee joint for men and women.

Description like in fig. 4

3.4 Likert's scale of pain estimation

The initial Likert's test pain estimation showed that:

 the average value of pain for the research group is 6,5, for men is 6 and for women it is 6,5.



Fig. 15. The initial examination pain estimation histogram (Likert's scale)

- The final Likert's test pain estimation showed that:
- the average value of pain for the research group is 4, for men is 4 and for women it is 4,5.



Fig. 16. Final examination pain estimation histogram (Likert's scale)

After the confrontation of the received results from the research group initial and final examination we can observe some changes between those examinations which present in a following way.

The difference between initial and final Likert's test pain estimation showed that:

the average value of pain that diminished during the rehabilitation was 2 for the research group, men and women.



Fig. 17. Confrontation of the pain estimation in the initial and final examination (Likert's scale)

3.5 The functional state and disease stadium assessment with the use of HAQ Disability Score questionnaire



Fig. 18. The disease stadium advance assessment and functional state with the use of the HAQ Disability Score questionnaire histogram in the initial examination.

The final assessment of the disease advance stadium and functional state with the use of the HAQ Disability Score questionnaire showed that:

 The average value of disease advance stadium for the whole research group is 1,5, for men is 1,5 and also for women it is 1,5.



Fig. 19. The disease stadium advance assessment and functional state with the use of the HAQ Disability Score questionnaire histogram in the final examination.

After the confrontation of the received results from the research group initial and final examination we can observe some changes between those examinations which present in a following way.

The difference between initial and final assessment of the disease advance stadium and functional state with the use of the HAQ Disability Score questionnaire that:

 the average value of disease advance level that diminished during the rehabilitation was 0,3 for the research group and women, for men it is 0,2.



Fig. 20. Confrontation of the assessment of the disease advance stadium and functional state with the use of the HAQ Disability Score questionnaire in the initial and final examination

4.0 Discussion

Goniometric examination of joint mobility range in the lower limb - the analysis of the research result in a group of patients with rheumatoid arthritis show a insignificant increasement of joint mobility range in the lower limb. This increasement is a result of a complex therapy composed of pharmacotherapy, kinesitherapy and physiotherapy. The main aim of the pharmacotherapy is slowing down the development of the disease and also diminishing the pain what had a significant influence on the increasing the joint mobility range during the passive and active exercises in the kinesitherapy. This result does not fully confirm the detailed research hypothesis, however it supports it. The biggest increasement of mobility range was observed in the knee the average increasement was 6° in the research group, for men it was 4,5° and women group was characterised by a 6,5°. The smallest increasement of mobility range was observed in abduction and extension of the hip joint. For the abduction the average increasement in the research group was 1,2°, for men it was 0,5° and women group was characterised by a 1,3°. For the extension the average increasement in the research group was 1.5°, for men it was 1,1° and women group was characterised by a 1,5°. Very often patients minimalise the movement to avoid the pain that accompanies the wider range of movement. This fact had influence on the final research result. However in the research group there was a small group of patients which made an evident improvement of mobility range even up to 50°. In the research group there was also a group of patients where a decreasement of mobility range was observed even up to 28°. This might be a result of an exacerbation of the disease during the hospitalisation. This refers especially to pain which is the most important factor in rheumatoid arthritis that restricts the movement. The joint movement rage is 20-30% smaller in the group of men than women. This might be caused by a psychical attitude towards disease and a smaller commitment in the active rehabilitation especially kinesitherapy. Women can better cope with the disease symptoms and complaints and are more active in the rehabilitation.

Muscle strength examination in the lower limb muscle groups - the analysis of the research result in a group of patients with rheumatoid arthritis show an increasement of muscle strength in the lower limb muscles groups. It is a result of a complex therapy mainly composed of kinesitherapy supported by physiotherapy and pharmacotherapy. The biggest increasement of muscle strength was observed in the group of muscles that is responsible for extention of hip joint - the average increasement was 0,3 in the research group, for men it was 0,6 and women group was characterised only by a 0.2. The smallest increasement of muscle strength was observed in the group of muscles that are responsible for abduction in the hip joint. The average increasement in the research group was 0,1, for men it was 0,2 and women group was characterised by the value lower than 0,1. The second examination also brings an insignificant increasement. This might be caused by a high average age in the research group what has a significant influence on development and increasement of muscle strength. This dependence can be described by following rule: the oldest the organism is, the slower and smaller are his abilities to develop and strengthen the muscle tissue. If the time of the hospitalisation was longer we could observe a bigger increasement of the muscle strength. With the two week hospitalisation, the small amount of physical exercises and disease exacerbation in some cases which is contraindication for intensive kinesitherapy and physiotherapy the received results are satisfactory. This fact is very significant in modulation the muscle strength parameters. However in the research group there is a small group of patients which characterises with a significant increasement of some muscle groups even up to 2,5. There were also patients with a decreasement of muscle strength in some muscles groups up to 3,5. But the biggest part of the research group showed no changes between initial and final examination of the muscle strength. This decreasement or lack of changes might be caused by disease exacerbation or a subjective increasement of pain. In the direct result comparison between men and women no significant domination of one group can be observed. In the particular muscle groups better results were achieved by men and in other groups women.

Pain level estimation with the use of the Likert's test – the analysis of the research result in a group patients with rheumatoid arthritis show a decreasement of subjectively felt pain by the patients. The main factor that modified the pain during the hospitalization was pharmacotherapy correctly adjusted to physiotherapeutic analgesic procedures like kryotherapy. The average diminision of the pain in the research group and men was 2 points in the Likert's scale and for women it was 2,1. Among the whole research group there were patients with pain decreasement of 4,7. Pursuant to those results we can observe the enormous effects of pharmacotherapy which by reducing the pain had also influence on the life quality improvement including the improvement of psychical attitude towards disease and tasks. This is a very significant aspect of rheumatoid arthritis treatment because it allows by diminishing the pain to carry out more intensive kinesitherapy which gives better effects in increasing the joint mobility range and muscle strength of all muscle groups.

Three gradual assessment of the disease advance stadium and functional state with the use of the HAQ Disability Score questionnaire – on the base of carried out researches we can observe a very slight changes in the assessment of the disease stadium. Those results are caused by a chronic character of the rheumatoid arthritis and long periods of remission and exacerbation when the changes in the development of the disease are hard to observe or are very rapid. The HAQ Disability Index questionnaire is divided into eight sections concerning different everyday activity zones like:

Patient assesses different kind of activities that are mentioned in those sections in the four gradual scale describing the level of difficulties with achieving those activities. The changes in walking with the use of walking devices (for example patient before the hospitalisation used a walking stick or crutches to walk) during two or three week hospitalisation are not in most cases observed. The reason of this phenomenon is that teaching the patient walking without the walking stick or crutches is a very long process especially when patient has a huge restrictions in the joint mobility range and has huge weakness of many muscle groups. This problem also concerns other everyday activities like morning getting up, grabbing or activities connected with going out of the house or flat. The negligence of regular limb exercises that some of the patients showed had a huge influence on those different kinds of everyday activities. The reason of this negligence was the age of the patients or feeling a severe pain during exercise attempt. In the research group there were only few patients who on their own hand exercised regularly what had a good reflection in the examination results. Those patients were characterised by a wider mobility range of majority of joints in lower limb. Also the muscle strength could have achieved bigger values and the patients would not complain about pain so often. The HAQ Disability Index questionnaire was supposed to define the patients disease stadium and check what influence had the complex rehabilitation on the rheumatology ward on the disease development.

The carried out examinations confirm the main research hypothesis. However the detailed research hypothesis was not confirmed due to the insignificant increasement of joints mobility range and muscle strength. The small improvement achieved is smaller than the expected one.

The achieved results indicate the insignificant improvement of lower limb dexterity in the joint mobility range and significant improvement of the muscle strength, quality of life and functioning of the patients with rheumatoid arthritis. The majority of patients were characterized by long time disease experience so the hospitalization was a continuance of the long term treatment of RA. The smallest changes during the complex rehabilitation gave the mobility range in the joints of lower limbs. The result could be better but the short time of hospitalization, the occurrence of disease exacerbation and lack of commitment in rehabilitation of some patients could have a significant influence on achieving such small improvement. Also the subjective character of examinations does not give the optimal picture of changes that occurred during complex treatment of the rheumatoid arthritis patients. The best results were achieved in changing the pain sense. It confirms the effectiveness of pharmacotherapy and the adjusted physiotherapy. The decreasement of pain had a huge influence on patient's mood and subjective assessment of the complex treatment. The majority of the patients during the final examination said that they were feeling better and were convinced that each part of the complex treatment was effective. They said that in their subjective opinion their joint mobility range in lower limb and muscle strength improved significantly. Some patients with the help of analgesics achieved a little movement in some planes in which there was no movement observed in the first days of hospitalization due to the pain. All this data confirm the importance of analgesics in the pharmacotherapy used with the appropriate drug that slower or control the development of the rheumatoid arthritis.

Some of the articles found in medic database like Pubmed and Cochrane, confirm our results, that complex rehabilitation has influence on improvement of functional condition of patients with rheumatoid arthritis. Ann Readapt in her work "Reconditioning in patients with rheumatoid arthritis" writes that: "dynamic and aerobic exercises do not aggravate joint inflammation and do not accelerate joint damage in such patients. The important goal of reconditioning patients with RA is the prevention of functional decline. RA patients need to be persuaded about the effectiveness and safety of moderate and even high-intensity exercise." Also in Bartels EM, Lund H and Danneskiold-Samsøe B. work "Pool exercise therapy of rheumatoid arthritis" it is mentioned that "pool therapy shows positive effects as part of the treatment of rheumatoid arthritis patients..." [1,5].

5.0 Conclusions

- Complex treatment including rehabilitation has influence on improvement of functional condition of patients with rheumatoid arthritis.
- The decreasement of pain allows to keep the joint mobility range and muscle strength on the same level.

Bibliography

- A. Readapt: Reconditioning in patients with rheumatoid arthritis, Med Phys., 2007
- F.C. Arnett, S.M. Edworthy, D.A. Bloch, D.J. McShane, J.F. Fries, N.S. Cooper, L.A. Healey, S.R. Kaplan, M.H. Liang, H.S. Luthra: *The American Rheumatism Association* 1987

revised criteria for the classification of rheumatoid arthritis, Arthritis Rheum., 1988, 31(3): 315-24.

- Bachta Artur elaboration on basis of: J. Fries: *Measurement* of patient outcome in arthritis, Arthritis and Rheumatism, 1980, 23: 137-45.
- Ch. Barnard: Jak żyć lepiej mimo zapalenia stawów i reumatyzmu, POLSKA DIRECT, Warszawa 1995.
- E.M. Bartels, H. Lund, B. Danneskiold-Samsře: *Pool exercise therapy of rheumatoid arthritis*, Ugeskr Laeger, 2001 Oct 1; 163(40): 55-7-13.
- K. Borysewicz: Znaczenie wczesnej rehabilitacji w reumatoidalnym zapaleniu stawów, materiały wewnętrzne Zakładu Reumatologii Akademii Medycznej we Wrocławiu, 1998.
- L. Brosseau, M.G. Judd, S. Marchand, V.A. Robinson, P. Tugwell, G. Wells, K. Yonge: *Transcutaneous electrical nerve stimulation (TENS) for the treatment of rheumatoid arthritis in the hand (Review)*, John Wiley & Sons, *The Cochrane Library*, Issue 4, 2005.
- L. Brosseau, V. Robinson, G. Wells, R. deBie, A. Gam, K. Harman, M. Morin, B. Shea, P. Tugwell: Low level laser therapy (Classes I, II and III) for treating, rheumatoid arthritis (Review), John Wiley & Sons, Issue 4, 2005.
- 9. W. Bruckle: *Reumatoidalne zapalenie stawów*, INTERSPAR, Warszawa 2000.
- J. Brzeziński: Metodologia badań pedagogicznych. PWN, Warszawa 2004.
- K.W.A. Chan et al..: The long time between onset of symptoms and diagnosis of rheumatoid arthritis. Arthritis Rheum., 1994, 37: 814-20.
- P. Emery, D.P.M. Symmons: What is early rheumatoid arthritis?, definition and diagnosis. Baillieres Clinical Rheumatology, 1997, 1: 13-26.
- A. Filipowicz-Sosnowska, E. Stanisławska-Biernat, A. Zubrzycka-Sienkiewicz: Wczesne reumatoidalne zapalenie stawów – implikacje kliniczne, Polskie Towarzystwo Reumatologiczne 2002.
- A. Kamiński: Metody i techniki badań w pedagogice, in: Pedagogika. Podręcznik akademicki. PWN, Warszawa 1974.
- A.W. Maszke: Metodologiczne podstawy badań pedagogicznych, Wydawnictwo URz, Rzeszów 2004.
- S. Nowak: Metodologia badań socjologicznych, PWN, Warszawa 1970.
- A. Paciorek: Reumatoidalne zapalenie stawów, materiały elektroniczne – www.resmedica.pl, 2001.
- T. Pilch, T. Bauman: Zasady badań pedagogicznych, "ŻAK", Warszawa 2002.
- J. Pazdur: Współczesne możliwości i perspektywy modyfikowania przebiegu reumatoidalnego zapalenia stawów, materiały wewnętrzne Kliniki Chorób Reumatycznych Instytutu Reumatologicznego w Warszawie 1999.
- V.A. Robinson, L. Brosseau, L. Casimiro, M.G. Judd, B.J. Shea, P. Tugwell, G. Wells: *Thermotherapy for treating rheumatoid arthritis (Review)*, John Wiley & Sons, Issue 4, 2005.
- 21. Z. Skorny: Prace magisterskie, WSiP, Warszawa 1984.
- 22. L.W.Smith, Psychika i ciało, Prószyński i S-ka, Warszawa 1998.
- A.P. Verhagen, S.M.A. Bierma-Zeinstra, J.R. Cardoso, R.A. de Bie, M. Boers, H.C.W. de Vet, *Balneotherapthy for the rheumatoid arthritis*, The Cochrane Database of Systematic Reviews 2004, Issue 1. Art. No.: CD000518. DOI: 10.1002/14651858.CD000518. 1rst online publishment: 26 January 2004 in Issue 1, 2004.
- 24. A. Zembaty: Kinzeyterapia. Vol. I. Kasper 2002.

Medicine