



EJBI 2009

ISSN 1801 - 5603

European Journal for Biomedical Informatics

Volume 5 (2009)

www.ejbi.eu



Aims and Scope

The European Journal for Biomedical Informatics reacts on the great European need to share the information in the multilingual and multicultural European area. The journal publishes peer-reviewed papers in English and other European languages simultaneously. This opens new possibilities for faster transfer of scientific-research pieces of knowledge to large international community of biomedical researchers, physicians, other health personnel and citizens.

The generally accepted translations of the English version of the paper to the official European languages and other European languages.

List of European languages

	ISO 639-1 code
Albanian	sq
Armenian	hy
Azerbaijani	az
Belarusian	be
Bosnian	bs
Bulgarian	bg
Catalan	ca
Croatian	hr
Czech	cs
Danish	da
Dutch	nl
English	en
Estonian	et
Finnish	fi
French	fr
Georgian	ka
German	de
Greek	el
Hungarian	hu
Icelandic	is
Irish	ga
Italian	it
Kazakh	kk
Latvian	lv
Lithuanian	lt
Luxembourgish	lb
Macedonian	mk
Maltese	mt
Norwegian	no
Polish	pl
Portuguese	pt
Romanian, Moldavian, Moldovan	ro
Romansh	rm
Russian	ru
Serbian	sr
Slovak	sk
Slovenian	sl
Spanish	es
Swedish	sv
Turkish	tr
Ukrainian	uk

Cooperating journals

Methods of Information in Medicine
Lékař a technika

(language)

(English)
(Czech)



Editors and Management

Editor in Chief: Jana Zvárová, Czech Republic
Managing Editor and Graphic Design: Petra Přečková, Czech Republic
Graphic Design PDF Version: Dana Vynikarová, Czech Republic
Sales and Marketing Manager: Libor Seidl, Czech Republic

Editorial Board

Ammenwerth, Elske (de)	Austria
Blobel, Bernd (de)	Germany
Bobrowski, Leon (pl)	Poland
Bureš, Vít (cs)	Czech Republic
Degoulet, Patrice (fr)	France
Dostálová, Tatjana (cs)	Czech Republic
Eryilmaz, Esat Nadir (tr)	Turkey
Hanzlíček, Petr (cs)	Czech Republic
Iversen, Irma (no)	Norway
Kern, Josipa (hr)	Croatia
Lukosevicius, Arunas (lt)	Lithuania
Mansmann, Ulrich (de)	Germany
Martin-Sánchez, Fernando (es)	Spain
Masic, Izet (bs)	Bosnia and Herzegovina
Mazura, Ivan (cs)	Czech Republic
McCullagh, Paul (en)	United Kingdom
Mihalas, George (ro)	Romania
Naszlady, Attila (hu)	Hungary
Nykänen, Pirkko (fi)	Finland
Paralič, Ján (sk)	Slovakia
Pisanelli, Domenico M. (it)	Italy
Sharp, Mary (ga)	Ireland
Sousa Pereira, Antonio (pt)	Portugal
Svačina, Štěpán	Czech Republic
Valenta, Zdeněk (cs)	Czech Republic
Vinarova, Jivka (bg)	Bulgaria
de Lusignan, Simon (en)	United Kingdom

Publisher

EuroMISE s.r.o.
Paprskova 330/15
CZ-14000 Praha 4
Czech Republic
EU VAT ID: Cz25666011

Office

EuroMISE s.r.o.
Paprskova 330/15
CZ-14000 Praha 4
Czech Republic

Contact

Karel Zvára
zvara@euromise.com,
Tel: +420 226 228 904
Fax: +420 241 712 990

Instructions to Authors

MANUSCRIPT SUBMISSIONS

European Journal of Biomedical Informatics (EJBI) is an international, peer-reviewed journal that publishes papers in the broad field of biomedical informatics. Manuscripts accepted for the electronic publication in EJBI are original contributions, reviews, brief reports, special communications, commentaries, and many other categories of papers. Due to special multilingual features of EJBI, these types of papers are published in English, but national language versions can be published simultaneously (see below).

Manuscripts should be sent electronically to the e-mail address: manuscripts@ejbi.org. Each manuscript should be submitted in MS Word, plain-text, HTML, TeX or LaTeX format and accompanied by image attachments and

- Identification Form: the name of the corresponding author with his contact address, phone number, fax number and e-mail address.
- Submission Requirement: the statement that the submitted paper has not been published in, nor has been submitted to, any other journal.
- Authorship Form: the covering letter signed by corresponding author, by which the author approves publication of the paper in the European Journal for Biomedical Informatics on behalf of all authors.
- Transfer of Copyright: All authors of the manuscript must have agreed to its publication and are responsible for its content and must also have agreed that the corresponding author has the authority to act on their behalf in all matters pertaining to publication of the manuscript. The corresponding author is responsible for informing the co-authors of the manuscript status throughout the submission, review, and publication process.

Submission Requirement, Authorship Form and Transfer of Copyright should be sent by fax +420 241 712 990 and by the surface mail to EuroMISE s.r.o., EJBI Editorial Office, Paprskova 330/15, 140 00 Prague 4, Czech Republic simultaneously.

Authors may supply a PDF file of the manuscript additionally. Submission of the PDF file only is not sufficient.

PUBLISHER

The European Journal of Biomedical Informatics is published by EuroMISE s.r.o., Prague, Czech Republic (tax identification number: CZ25666011). Papers published in the Internet version of EJBI may be republished in printed digest of EJBI.

LANGUAGE

All manuscripts should be written in English in an easily readable style. English version of the manuscript will be peer-reviewed. However, it is possible to submit another language version of the same paper additionally. The Internet version of the EJBI makes available all language versions submitted. It is also possible to publish the English version of the paper that has been already

published in the national journal, when the agreement of national publisher is obtained (copyright agreement). The English version follows a standard peer-review procedure.

ELECTRONIC FORM OF A MANUSCRIPT

Recommended formats of text files are: .DOC (Word version 6.0, 7.0, 97, 2000 or 2003), .RTF (Rich Text Format). The name of the text file should be identical with the first author's surname, e.g. NOVAK.DOC. Names of attached images, drawings or tables (i.e. figures and tables of the manuscript) should be composed of author's surname and figure or table number, e.g. NOVAKTAB3.BMP. Manuscripts along with attached files should be sent by e-mail to address manuscripts@ejbi.org.

ELECTRONIC IMAGE DOCUMENTATION

Figures and tables (bitmaps) should be submitted in BMP, GIF, TIF, PNG, EPS or JPEG format. JPEG files should have best compressive rate (10-20) for a good quality. All images (although included in the manuscript file) must be also submitted as separate files because the resolution of embedded images is often insufficient. Scanned images should have the resolution of at least 600x600dpi. Submissions that do not meet the Instructions for Authors will be returned.

ABBREVIATIONS AND NOMENCLATURE

Generally known abbreviations do not need to be explained. Abbreviations for symbols and expressions for terms should be spelled and they should be comprehensibly explained in brackets. Every such abbreviation should be explained only after its first occurrence in the text.

Measurements of length, height, weight and volume should be reported in metric units or their decimal multiples. All haematological and clinical chemistry measurements should be reported in the metric system in term of International System of Units (SI). Chemical substances should be described by their own systematic name or expression, medicines by general names. Commercial names of chemicals, medicines or technical innovations may be used after they have been defined by their scientific names. Radionuclides (radioisotopes) will be symbolised by the atomic number.

REVISED MANUSCRIPT SUBMISSION

When revision of a manuscript is requested, authors should return the revised version of their manuscript as soon as possible. The prompt action may ensure fast publication if a paper is finally accepted for publication in EJBI.

FINAL PROOFREADING

The Publisher will send the accepted paper to its author for final proofreading in the PDF format. The author may then correct printing errors only. No other changes or additions will be accepted. Author should send corrected and signed paper back to the Publisher by fax to +420 241 712 990 or by surface mail to: EuroMISE s.r.o., EJBI Editorial Office, Paprskova 330/15, 140 00 Prague, Czech Republic.

ORGANIZATION OF THE MANUSCRIPT

Title page. The first (title) page should contain the title of the paper, names and workplaces of all authors. Individual workplaces are necessary to be graphically differentiated (preferably by numeral as the upper index).

Abstracts and keywords. At the beginning the author puts an abstract and keywords. The abstract should be in the extent of 250-300 words. There should be 4 to 7 keywords, according to author's consideration, preferably from MeSH index.

Main text of the paper. General rules for writing manuscripts recommend use of simple and declarative sentences; avoid long sentences, in which meaning may be lost by complicated construction. All acronyms and abbreviations should be explained when they first appear in the text. The main text of the paper should follow the style of selected type of paper.

Acknowledgement. Acknowledgements, if any, should be given at the end of the paper, before bibliographic references.

References. References should be cited in the text by their index number according to the order of appearance in the manuscript. Each reference should be marked by its index number in square bracket corresponding to bibliography section. It is possible to include references to dissertation works and technical reports. It is obligatory to include information sufficient to look up referenced text.

Examples of references in bibliography section:

[1] Knaup P., Ammenwerth E., Brandner R., Brigl B., Fischer G., Garde S., Lang E., PilgramR., Ruderich F., Singer R., Wolff A. C., Haux R., Kulikowski C.: Towards Clinical Bioinformatics: Advancing Genomic Medicine with Informatics Methods and Tools. *Methods Inf Med* 2004; 43, pp. 302-307

[2] Blobel B., Pharow P.: A Model-Driven Approach for the German Health Telematics Architectural Framework and the Related Security Infrastructure. In: Connecting Medical Informatics and Bio-Informatics. Proceedings of MIE2005 (Eds. R. Engelbrecht, A. Geissbuhler, C. Lovis, G. Mihalas), Vol. 116, Amsterdam, IOS Press, 2005, pp. 391-396

[3] <http://www.infobiomed.org/>

Tables and Figures. Authors should use tables only to achieve concise presentation, or where the information cannot be given satisfactory in another way. Tables should be numbered consecutively using Arabic numerals and should be referred to in the text by numbers. Each table should have an explanatory caption that should be as concise as possible. Figures should be clear, easy to read and of a good quality. Styles and fonts should match those in the main body of the paper. All figures must be mentioned in the text in consecutive order and should be numbered with Arabic numerals. Authors should indicate precisely in the main text where tables and figures should be inserted, if these elements are given only separately or at the end in the original version of the manuscript.

Content

English version

- en 2 - 3 Editorial: Comprehensive eHealth
J. Zvárová, B. Blobel
- en 4 Architectural Approach to eHealth for Enabling Paradigm Changes in Health
B. Blobel
- en 5 Putting Biomedical Ontologies to Work
B. Smith, M. Brochhausen
- en 6 Subword-based Semantic Retrieval of Clinical and Bibliographic Documents
P. Daumke, S. Schulz, M. L. Müller, W. Dzeyk, L. Prinzen, E. J. Pacheco, P. Secco Cancian, P. Nohama, K. Markó
- en 7 Semantic Interoperability Adheres to Proper Models and Code Systems
A Detailed Examination of Different Approaches for Score Systems
F. Oemig, B. Blobel
- en 8 Ensuring the Continuity of Care of Cardiorespiratory Diseases at Home
Monitoring Equipment and Medical Data Exchange over Semantically Annotated Web Services
B. Spyropoulos, A. Tzavaras, M. Botsivaly, K. Koutsourakis
- en 9 Efficiency Evaluation of Autonomic Heart Control by Using the Principal Component Analysis of ECG P-Wave
A. Krisciukaitis, R. Simoliuniene, M. Tamasiunas, V. Saferis, A. Vainoras, L. Gargasas
- en 10 Voice-supported Electronic Health Record for Temporomandibular Joint Disorders
R. Hippmann, T. Dostállová, J. Zvárová, M. Nagy, M. Seydlova, P. Hanzlíček, P. Kriz, I. Smidl, J. Trmal
- en 11 Omics-based Medicine and Systems Pathology. A New Perspective for Personalized and Predictive Medicine
H. Tanaka
- en 12 Semantic Interoperability in Czech Healthcare Environment Supported by HL7 Version 3
M. Nagy, P. Hanzlíček, P. Přečková, A. Říha, M. Dioszegi, L. Seidl, J. Zvárová
- en 13 Architectural Approaches for HL7-based Health Information Systems Implementation
D. M. López, B. Blobel

Czech version

- cs 2 - 3 Předmluva: Komplexní pohled na eZdraví
J. Zvárová, B. Blobel
- cs 4 - 8 Hlasově ovládaná elektronická zdravotní dokumentace pro poruchy temporomandibulárního kloubu
R. Hippmann, T. Dostállová, J. Zvárová, M. Nagy, M. Seydlova, P. Hanzlíček, P. Kriz, L. Šmíd, J. Trmal
- cs 9 - 22 Medicína založená na omics vědách a systémová patologie - Nová perspektiva personalizované a prediktivní medicíny
H. Tanaka

German version

- de 2 - 9 eHealth-Ansatz für einen Paradigmenwechsel im Gesundheitswesen Architekturelle Herausforderungen
B. Blobel
- de 10 - 16 Semantiche Interoperabilität erfordert die richtigen Modelle und Code-Systeme:
Eine Prüfung verschiedener Ansätze für Score Systeme
F. Oemig, B. Blobel
- de 17 -22 Biomedizinische Ontologien für die Praxis
M. Brochhausen, B. Smith

Greek version

- el 2 - 9 Εξασφαλίζοντας την συνέχεια της κατ' οίκον περίθαλψης καρδιοαναπνευστικών παθήσεων:
Εξοπλισμός επιτήρησης και λογισμικό ανταλλαγής ιατρικών δεδομένων μέσω σημασιολογικά
εμπλουτισμένων υπηρεσιών του Ιστού
B. Spyropoulos, A. Tzavaras, M. Botsivaly, K. Koutsourakis

Lithuanian version

- lt 2 - 7 Autonominio širdies veiklos reguliavimo efektyvumo vertinimas EKG P bangos pagrindinių komponenčių analizės būdu
A. Krisciukaitis, R. Simoliuniene, M. Tamosiunas, V. Saferis, A. Vainoras, L. Gargasas

Portuguese version

- pt 2 - 9 Recuperação Semântica de Documentos Clínicos e Bibliográficos Baseada em Subwords
P. Daumke, S. Schulz, M. L. Müller, W. Dzeyk, L. Prinzen, E. J. Pacheco, P. Secco Cancian, P. Nohama, K. Markó

Slovak version

- sk 2 - 11 Sémantická interoperabilita v prostredí českého zdravotníctva podporovaná HL7 v3
M. Nagy, P. Hanzlíček, P. Přečková, A. Říha, M. Dioszegi, L. Seidl, J. Zvárová

Spanish version

- es 2 - 9 Arquitecturas para la Implementación de Sistemas de Información en Salud Basadas en el Estándar HL7
D. M. López, B. Blobel



EJBI 2009 ISSN 1801 - 5603

European Journal for Biomedical Informatics

Volume 5 (2009), Issue 1

English version

Cooperating journal – Methods of Information in Medicine
Editor-in-Chief: R. Haux (Germany)
Publisher: Schattauer
www.methods-online.com

www.ejbi.eu



Content

English version

- en 2 - 3 Editorial: Comprehensive eHealth
J. Zvárová, B. Blobel
- en 4 Architectural Approach to eHealth for Enabling Paradigm Changes in Health
B. Blobel
- en 5 Putting Biomedical Ontologies to Work
B. Smith, M. Brochhausen
- en 6 Subword-based Semantic Retrieval of Clinical and Bibliographic Documents
P. Daumke, S. Schulz, M. L. Müller, W. Dzeyk, L. Prinzen, E. J. Pacheco, P. Secco Cancian, P. Nohama, K. Markó
- en 7 Semantic Interoperability Adheres to Proper Models and Code Systems
A Detailed Examination of Different Approaches for Score Systems
F. Oemig, B. Blobel
- en 8 Ensuring the Continuity of Care of Cardiorespiratory Diseases at Home
Monitoring Equipment and Medical Data Exchange over Semantically Annotated Web Services
B. Spyropoulos, A. Tzavaras, M. Botsivaly, K. Koutsourakis
- en 9 Efficiency Evaluation of Autonomic Heart Control by Using the Principal Component Analysis of ECG P-Wave
A. Krisciukaitis, R. Simoliuniene, M. Tamosiunas, V. Saferis, A. Vainoras, L. Gargasas
- en 10 Voice-supported Electronic Health Record for Temporomandibular Joint Disorders
R. Hippmann, T. Dostállová, J. Zvárová, M. Nagy, M. Seydlova, P. Hanzlíček, P. Kriz, I. Smidl, J. Trmal
- en 11 Omics-based Medicine and Systems Pathology. A New Perspective for Personalized and Predictive Medicine
H. Tanaka
- en 12 Semantic Interoperability in Czech Healthcare Environment Supported by HL7 Version 3
M. Nagy, P. Hanzlíček, P. Přečková, A. Říha, M. Dioszegi, L. Seidl, J. Zvárová
- en 13 Architectural Approaches for HL7-based Health Information Systems Implementation
D. M. López, B. Blobel

Editorial

Comprehensive eHealth

J. Zvárová¹, B. Blobel²

¹Center of Biomedical Informatics, Institute of Computer Science AS CR, Prague, Czech Republic

²University Hospital Regensburg, eHealth Competence Center, Regensburg, Germany

Current demographic, economic and social conditions in developed countries require a paradigm change for delivering high quality and efficient health services. In that context healthcare systems have to turn from organization-centered to process-oriented and finally towards life-long, individualized, shared patient's care, based on eHealth platform services. This has been broadly considered, e.g. by Haux [1]. Interoperability requirements for ubiquitous personalized health services reach beyond current concepts of health information integration among professional providers of health care. Future personalized health platforms have to maintain particularly semantic interoperability among systems using different modalities and technologies, different knowledge representation and domain experts' languages as well as different coding schemes and terminologies to include home care as well as personalized and mobile systems. This development is not restricted to regions or countries, but appears globally, requiring a comprehensive multidisciplinary and international collaboration. The new translational characteristics of medicine and the consequences for health informatics have been extendedly discussed by Kuhn et al. [2] as well as Kulikowski [3].

This special issue summarizes a nowadays view on selected eHealth topics presented during the International Conference on eHealth in Regensburg, 2-5 December 2007. Within the International Experts Summit, the conference offered invited speeches from internationally leading experts representing all domains involved in eHealth. Although generally demonstrating and facilitating collaboration, the last day of the event offered specific seminars, workshops and symposia addressing the

collaborative potential between Europe and Latin America (ELAN), analyzing the cross-border activities among Germany, Austria, the Czech Republic, and Switzerland. The results of this conference comprehensively covering eHealth were published in [4], [5]. Several contributions to either the Experts Summit or the open conference have been selected by the Scientific Program Committee to be extended and updated for an English publication in a Special Issue of Methods of Information in Medicine (Methods) and in parallel for a national language publication in the European Journal for Biomedical Informatics (EJBI), both in cooperation with Schattauer Verlagsgesellschaft mbH.

The series is introduced by a paper covering the basic constituents and architectural challenges of eHealth interoperability solutions, also tackling the ontological aspects of the multidisciplinary approach to eHealth (Blobel [6]). This ontology challenge is discussed deeper in the paper from Smith and Brochhausen [7], focusing on formal ontology representation, while Daumke and co-authors [8] discuss semantic interoperability of nonformal concept representation in medicine. The ontology aspect has been practically highlighted by Oemig et al. [9] at the Score Systems example. Spyropoulos and co-authors [10] demonstrate the continuity of care in pHealth by the inclusion of home care for cardio-respiratory diseases, putting the biomedical engineering dimension on the table. Several papers focus on the multimodal challenge of pHealth, such as Krisciukaitis et al. [11], who tackle the algorithmic challenge of signal analysis, or Hippmann and co-authors [12] considering the application of voice communication in documentation. Tanaka [13] includes the new domain of omics-based medicine and systems pathology for opening new

perspectives for personalized and predictive care. The loop is finally closed by Nagy et al. [14], discussing the complex interoperability challenge impacted by international nomenclatures, communication standards and enterprise programming techniques as well as by Lopez' paper [15] on the automated implementation of interoperable eHealth solutions, based on an architecture-centric and formal development process.

The editors of the 2009 volume (J. Zvárová and B. Blobel) would like to thank all authors for their extraordinary efforts to provide papers of a high scientific level far beyond the original scope of the conference. Furthermore, we are indebted to thank the Methods Editorial Board and Schattauer Printing House for enabling the preparation of this joint publication with EJBI.

References

- [1] Haux R. Individualization, globalization and health – about sustainable information technologies and the aim of medical informatics. *Int J Med Inf* 2006; 75: 795–808.
- [2] Kuhn KA, Knoll A, Mewes H-W, Schwaiger M, Bode A, Broy M, Daniel H, Feussner H, Gradinger R, Hauner H, Höfler H, Holzmann B, Horsch A, Kemper A, Krcmar H, Kochs EF, Lange R, Leidl R, Mansmann U, Mayr EW, Meitinger T, Molls M, Navab N, Nüsslin F, Peschel C, Reiser M, Ring J, Rummey EJ, Schlichter J, Schmid R, Wichmann HE, Ziegler S. *Informatics and Medicine: From Molecules to Populations*. *Methods Inf Med* 2008; 47: 283–295.
- [3] Kulikowski CA. The Micro-Macro Spectrum of Medical Informatics Challenges – From Molecular Medicine to Transforming Health Care in a Globalizing Society. *Methods Inf Med* 2002; 41: 20–24.

- [4] Blobel B, Pharow P, Nerlich M (eds). *eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge – Global Experts Summit Textbook. Series Studies in Health Technology and Informatics, Vol. 134.* Amsterdam: IOS Press; 2008.
- [5] Blobel B, Pharow P, Zvárová J, Lopez DM (eds). *eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge – CeHR Conference Proceedings 2007.* Amsterdam: IOS Press; 2008, and Berlin: AKA GmbH; 2008.
- [6] Blobel B. Architectural Approach to eHealth for Enabling Paradigm Changes in Health. *Methods Inf Med* 2010; 49 (2): 123–134.
- [7] Smith B, Brochhausen M. Putting Biomedical Ontologies to Work. *Methods Inf Med* 2010; 49 (2): 135–140.
- [8] Daumke P, Schulz S, Müller ML, Dzeyk W, Prinzen L, Pacheco E.J, Secco Cancian P, Nohama P, Markó K. Subword-based Semantic Retrieval of Clinical and Bibliographic Documents. *Methods Inf Med* 2010; 49 (2): 141–147.
- [9] Oemig F, Blobel B. Semantic Interoperability Adheres to Proper Models and Code Systems – A Detailed Examination of Different Approaches for Score Systems. *Methods Inf Med* 2010; 49 (2): 148–157.
- [10] Spyropoulos B, Tzavaras A, Botsivaly M, Koutsourakis K. Ensuring the Continuity of Care of Cardiorespiratory Diseases at Home – Monitoring Equipment and Medical Data Exchange over Semantically annotated Web Services. *Methods Inf Med* 2010; 49 (2): 156–160.
- [11] Krisciukaitis A, Simoliuniene R, Tamosiunas M, Saferis V, Vainoras A, Gargasas L. Efficiency Evaluation of Autonomic Heart Control by Using the Principal Component Analysis of ECG P-Wave. *Methods Inf Med* 2010; 49 (2): 161–167.
- [12] Hippmann R, Dostalova T, Zvárová J, Nagy M, Seydlova M, Hanzlicek P, Kriz P, Smidl L, Trmal J. Voice-supported Electronic Health Record for Temporomandibular Joint Disorders. *Methods Inf Med* 2010; 49 (2): 168–172.
- [13] Tanaka H. Omics-based Medicine and Systems Pathology – A New Perspective for Personalized and Predictive Medicine. *Methods Inf Med* 2010; 49 (2): 173–185.
- [14] Nagy M, Hanzlicek P, Precková P, Riha M, Dioszegi M, Seidl L, Zvárová J. Semantic Interoperability in Czech Healthcare Environment Supported by HL7 Version 3. *Methods Inf Med* 2010; 49(2): 186–195.
- [15] López DM, Blobel B. Architectural Approaches for HL7-based Health Information Systems Implementation. *Methods Inf Med* 2010; 49 (2): 196–204.

Contact

Jana Zvárová, PhD, DSc.

Full Professor

Center of Biomedical Informatics,
Institute of Computer Science AS CR,
Pod Vodarenskou vezi 2,
182 07 Prague
Czech Republic
E-mail: zvarova@euromise.cz

Architectural Approach to eHealth for Enabling Paradigm Changes in Health

B. Blobel¹

¹ eHealth Competence Center, University Hospital Regensburg, Regensburg, Germany

Summary

Objectives

For improving safety and quality of care as well as efficiency of health delivery under the well-known burdens, health services become specialized, distributed, and therefore collaborative, thereby changing the health service paradigm from organization-centered over process-controlled to personal health (pHealth).

Methods

Personalized eHealth services provided independent of time and location have to be based on advanced technical paradigms of mobile, pervasive and autonomous computing, enabling ubiquitous health services. Personalized eHealth systems require a multidisciplinary approach including medicine, informatics, biomedical engineering, bioinformatics and the omics disciplines but also legal and regulatory affairs, administration, security, privacy and ethics, etc. Interoperability between different components of the intended system must be provided through an architecture-centric, model-driven, formalized process.

Results

In order to analyze, design, specify, implement and maintain such an interactive environment impacted by so many different domains, a formal and unified methodology for system analysis and design has been developed and deployed, based on an overall architectural framework. The paper introduces the underlying paradigms, requirements, architectural reference models, modeling and formalization principles as well as development processes for comprehensive service-oriented personalized eHealth interoperability chains, thereby exploiting all interoperability levels up to service interoperability. A special focus is put on ontologies and knowledge representation

in the context of eHealth and pHealth solutions. Furthermore, EHR solutions, security requirements, existing and emerging standards, and educational challenges for realizing personalized pHealth are briefly discussed.

Conclusion

For personal health, bridging between disciplines including ontology coordination is the crucial demand. All aspects of the design and development process have to be considered from an architectural viewpoint.

Keywords

Ontology, Knowledge Representation, semantic interoperability, Personal health, system architecture, universal logic, ubiquitous care

DOI

10.3414/ME9308

The full English version can be found in /Methods of Information in Medicine 2010 49 2: 123-134/ or <http://www.schattauer.de/de/magazine/uebersicht/zeitschriften-a-z/methods/issue/special/manuscript/12617/show.html>.

Putting Biomedical Ontologies to Work

B. Smith^{1,2}, M. Brochhausen²

¹Department of Philosophy and New York State Center of Excellence in Bioinformatics and Life Sciences

²Institute of Formal Ontology and Medical Information Science, Saarland University, Saarbrücken, Germany

Summary

Objectives

Biomedical ontologies exist to serve integration of clinical and experimental data, and it is critical to their success that they be put to widespread use in the annotation of data. How, then, can ontologies achieve the sort of user-friendliness, reliability, cost-effectiveness, and breadth of coverage that is necessary to ensure extensive usage?

Methods

Our focus here is on two different sets of answers to these questions that have been proposed, on the one hand in medicine, by the SNOMED CT community, and on the other hand in biology, by the OBO Foundry. We address more specifically the issue as to how adherence to certain development principles can advance the usability and effectiveness of an ontology or terminology resource, for example by allowing more accurate maintenance, more reliable application, and more efficient interoperation with other ontologies and information resources.

Results

SNOMED CT and the OBO Foundry differ considerably in their general approach. Nevertheless, a general trend towards more formal rigor and cross-domain interoperability can be seen in both and we argue that this trend should be accepted by all similar initiatives in the future.

Conclusions

Future efforts in ontology development have to address the need for harmonization and integration of ontologies across disciplinary borders, and for this, coherent formalization of ontologies is a pre-requisite.

Keywords

Quality Assurance, SNOMED CT, Biomedical ontologies, ontology harmonization

DOI

10.3414/ME9302

The full English version can be found in
/Methods of Information in Medicine 2010
49 2: 135-140/ or
<http://www.schattauer.de/de/magazine/uebersicht/zeitschriften-a-z/methods/issue/special/manuscript/12619/show.html>.

Subword-based Semantic Retrieval of Clinical and Bibliographic Documents

P. Daumke^{1,2}, S. Schulz², M. L. Müller³, W. Dzeyk⁴, L. Prinzen⁵, E. J. Pacheco^{6,7}, P. Secco Cancian⁶, P. Nohama^{6,7}, K. Markó^{1,2}

¹AVERBIS GmbH, Freiburg, Germany

²Freiburg University Hospital, Medical Informatics Group, Freiburg, Germany

³Freiburg University Hospital, Department of Dermatology, Freiburg, Germany

⁴German National Library of Medicine, Cologne, Germany

⁵Bertelsmann Stiftung, Gütersloh, Germany

⁶Paraná Catholic University, Health Informatics Laboratory, Curitiba, Brazil

⁷Universidade Tecnológica Federal do Paraná, CPGEI, Curitiba, Brazil

Summary

Objectives

The increasing amount of electronically available documents in bibliographic databases and the clinical documentation requires user-friendly techniques for content retrieval.

The full English version can be found in /Methods of Information in Medicine 2010 49 2: 141-147/ or <http://www.schattauer.de/de/magazine/uebersicht/zeitschriften-a-z/methods/issue/special/manuscript/12759/show.html>.

Methods

A domain-specific approach on semantic text indexing for document retrieval is presented. It is based on a subword thesaurus and maps the content of texts in different European languages to a common interlingual representation, which supports the search across multilingual document collections.

Results

Three use cases are presented where the semantic retrieval method has been implemented: a bibliographic database, a department EHR system, and a consumer-oriented Web portal.

Conclusions

It could be shown that a semantic indexing and retrieval approach, the performance of which had already been empirically assessed in prior studies, proved useful in different prototypical and routine scenarios and was well accepted by several user groups.

Keywords

Consumer Health Informatics, Electronic health record, cross-language information retrieval, Thesauri, semantic information retrieval, document indexing, bibliographic databases

DOI

10.3414/ME9303

Semantic Interoperability Adheres to Proper Models and Code Systems

A Detailed Examination of Different Approaches for Score Systems

F. Oemig¹, B. Blobel²

¹Agfa HealthCare GmbH, Bonn, Germany

²eHealth Competence Center, Regensburg, Germany

Summary

Objectives

Achieving semantic interoperability requires not only the use of communication standards like HL7 with its underlying models and specifications, but also to constrain those models to instances including permitted attributes, data types, values and code systems. Even the application of both strategies may lead to different modeling approaches and therefore incompatible results, however.

[z/methods/issue/special/manuscript/12618/show.html](#)

Methods

This paper analyzes the different ways to create a model exemplified at score and assessment systems.

Results

The different approaches have advantages and disadvantages. The presented results allow for transmitting the same basic information facilitating HL7 v2.x and V3 in a way reducing implementation efforts.

Conclusions

Establishing a generic approach to communicate the details of score systems driven by an appropriate set of codes is the best solution for implementers.

Keywords

HL7, Health telematics, semantic interoperability, communication standard, conformance, implementation guides, scores and assessments

DOI

10.3414/ME9304

The full English version can be found in /Methods of Information in Medicine 2010 49 2: 148-155/ or <http://www.schattauer.de/de/magazine/uebersicht/zeitschriften-a>

Ensuring the Continuity of Care of Cardiorespiratory Diseases at Home: Monitoring Equipment and Medical Data Exchange over Semantically Annotated Web Services

B. Spyropoulos¹, A. Tzavaras¹, M. Botsivaly¹, K. Koutsourakis¹

¹Medical Instrumentation Technology Department, Faculty of Engineering, Technological Educational Institution (TEI) of Athens, Athens, Greece

Summary

Objectives

A significant portion of care related to cardiorespiratory diseases is provided at home, usually but not exclusively, after the discharge of a patient from hospital. It is the purpose of the present study to present the technical means which we have developed, in order to support the adaptation of the continuity of care of cardiorespiratory diseases at home.

Methods

We have developed an integrated system that includes: first, a prototype laptop-based portable monitoring system that comprises low-cost commercially available components, which enable the periodical or continuous monitoring of vital signs at home; second, software supporting medical decision-making related to tachycardia and ventricular fibrillation, as well as fuzzy-rules-based software supporting home-ventilation optimization; third, a typical continuity of care record (CCR) adapted to support also the creation of a homecare plan; and finally, a prototype ontology, based upon the HL7 clinical document architecture (CDA), serving as basis for the development of semantically annotated web services that allow for the exchange and retrieval of homecare information.

Results

The flexible design and the adaptable data-exchange mechanism of the developed system result in a useful and standard-compliant tool, for cardiorespiratory disease-related homecare.

Conclusions

The ongoing laboratory testing of the system shows that it is able to contribute to an effective and low-cost package solution, supporting patient supervision and

treatment. Furthermore, semantic web technologies prove to be the perfect solution for both the conceptualization of a continuity of care data exchange procedure and for the integration of the structured medical data.

Keywords

continuity of care, Homecare, cardiorespiratory diseases, semantically annotated web-services

DOI

10.3414/ME9305

The full English version can be found in /Methods of Information in Medicine 2010 49 2: 156-160/ or <http://www.schattauer.de/de/magazine/uebersicht/zeitschriften-a-z/methods/issue/special/manuscript/12616/show.html>.

Efficiency Evaluation of Autonomic Heart Control by Using the Principal Component Analysis of ECG P-Wave

A. Krisciukaitis¹, R. Simoliuniene², M. Tamosiunas¹, V. Saferis², A. Vainoras³, L. Gargasas³

¹Institute for Biomedical Research, Kaunas University of Medicine, Kaunas, Lithuania

²Kaunas University of Medicine, Kaunas, Lithuania

³Institute of Cardiology, Kaunas University of Medicine, Kaunas, Lithuania

Summary

Background

Cardiac output is controlled by the autonomic nervous system by changing the heart rate and/or the contractions of the heart muscle in response to the hemodynamic needs of the whole body. Malfunction of these mechanisms causes the postural orthostatic tachycardia syndrome and/or the chronic fatigue syndrome. Evaluation of functionality and efficiency of the control mechanisms could give valuable diagnostic information in the early stages of dysfunction of the heart control systems and help to monitor the healing process in rehabilitation period after interventions.

Objectives

In this study we demonstrate how P-wave changes evoked by an orthostatic test could be quantitatively evaluated by using the method based on the principal component analysis.

Methods

ECG signals were recorded during an orthostatic test performed according to the typical protocol in three groups of volunteer subjects representing healthy young and older persons, part of which had transient periods of supraventricular arrhythmias. Quantitative evaluation of P-wave morphology changes was performed by means of principal component analysis-based method.

Results

Principal component-based estimates showed certain variety of P-wave shape during orthostatic test, what revealed a possibility to evaluate the properties of parasympathetic heart control.

Conclusions

Quantitative evaluation of ECG P-wave changes evoked by an orthostatic test by

using a newly developed method provides a quantitative estimate for functionality and efficiency of the heart rate control mechanisms. The method could be used in eHealth systems.

Keywords

principal component analysis, autonomic heart control, ECG P-wave

DOI

10.3414/ME9306

The full English version can be found in /Methods of Information in Medicine 2010 49 2: 161-167/ or <http://www.schattauer.de/de/magazine/uebersicht/zeitschriften-a-z/methods/issue/special/manuscript/12758/show.html>.

Voice-supported Electronic Health Record for Temporomandibular Joint Disorders

R. Hippmann^{1,2}, T. Dostálová^{1,2}, J. Zvárová^{3,2}, M. Nagy^{3,2}, M. Seydlova^{1,2}, P. Hanzlícek^{3,2}, P. Kriz¹, I. Smidl⁴, J. Trmal⁴

¹Department of Paediatric Stomatology, Second Faculty of Medicine of Charles University, Prague, Czech Republic

²Centre of Biomedical Informatics, Prague, Czech Republic

³EuroMISE Centre, Department of Medical Informatics, Institute of Computer Science, v.v.i., Academy of Sciences of the Czech Republic, Prague, Czech Republic

⁴Department of Cybernetics, University of West Bohemia, Plzen, Czech Republic

Summary

Objective

Biomedical ontologies exist to serve integration of clinical and experimental data, and it is critical to their success that they be put to widespread use in the annotation of data. How, then, can ontologies achieve the sort of user-friendliness, reliability, cost-effectiveness, and breadth of coverage that is necessary to ensure extensive usage?

Methods

Our focus here is on two different sets of answers to these questions that have been proposed, on the one hand in medicine, by the SNOMED CT community, and on the other hand in biology, by the OBO Foundry. We address more specifically the issue as to how adherence to certain development principles can advance the usability and effectiveness of an ontology or terminology resource, for example by allowing more accurate maintenance, more reliable application, and more efficient interoperation with other ontologies and information resources.

Results

SNOMED CT and the OBO Foundry differ considerably in their general approach. Nevertheless, a general trend towards more formal rigor and cross-domain interoperability can be seen in both and we argue that this trend should be accepted by all similar initiatives in the future.

Conclusions

Future efforts in ontology development have to address the need for harmonization and integration of ontologies across disciplinary borders, and for this, coherent formalization of ontologies is a pre-requisite.

Keywords

Quality Assurance, SNOMED CT, Biomedical ontologies, ontology harmonization

DOI

10.3414/ME9302

The full English version can be found in /Methods of Information in Medicine 2010 49 2: 168-172/ or <http://www.schattauer.de/de/magazine/uebersicht/zeitschriften-a-z/methods/issue/special/manuscript/12310/show.html>.

Omics-based Medicine and Systems Pathology A New Perspective for Personalized and Predictive Medicine

H. Tanaka¹

¹Graduate School of Biomedical Science, Tokyo Medical and Dental University, Tokyo, Japan

Summary

Objectives

Recent important advances in the human genomics and post-genomic “omics” are now bringing about a new medical care which we call “omics-based medicine”. In this article, we investigated the development and future possibilities of omics-based medicine.

Methods

We divided the development of omics-based medicine into three generations in order to clarify the main clinical goals and characteristics of informatics method of each generation, together with its future possibilities.

Results

The first generation of omics-based medicine started with “genomic medicine” based on the inborn individual differences of genome. It has opened the study of genetic polymorphism of the diseases and promoted the personalized medication based on the pharmacogenetic/pharmacogenomic difference of the drug response. In the second generation of omics-based medicine, owing to the advances in the high-throughput technology, vast amount of the various post-genomic disease omics data containing comprehensive molecular information of diseased somatic cells has become available. It reflects the ongoing state of diseases more closely and enables the predictive medicine such as prognosis prediction of disease by applying the data-driven analysis. Finally, due to the rapidly growing knowledge about the cellular molecular network, system-level understanding of the disease, called systems pathology, becomes possible. It can fully exploit the substantial contents of disease omics and will lead to a comprehensive understanding of disease process by using model-driven analysis.

Conclusion: Omics-based medicine and systems pathology will realize a new personalized and predictive medicine.

Keywords

Genomics, pharmacogenomics, systems pathology, Omics-based medicine, omics

DOI

10.3414/ME9307

The full English version can be found in /Methods of Information in Medicine 2010 49 2: 173-185/ or <http://www.schattauer.de/de/magazine/uebersicht/zeitschriften-a-z/methods/issue/special/manuscript/12757/show.html>.

Semantic Interoperability in Czech Healthcare Environment Supported by HL7 Version 3

M. Nagy¹, P. Hanzlíček¹, P. Prečková¹, A. Říha¹, M. Dioszegi¹, L. Seidl¹, J. Zvárová¹

¹Center of Biomedical Informatics, Institute of Computer Science of the Academy of Sciences of the Czech Republic, Prague, Czech Republic

Summary

Objectives

The data interchange in the Czech healthcare environment is mostly based on national standards. This paper describes a utilization of international standards and nomenclatures for building a pilot semantic interoperability platform (SIP) that would serve to exchange information among electronic health record systems (EHR-Ss) in Czech healthcare. The work was performed by the national research project of the „Information Society“ program.

Methods

At the beginning of the project a set of requirements the SIP should meet was formulated. Several communication standards (openEHR, HL7 v3, DICOM) were analyzed and HL7 v3 was selected to exchange health records in our solution. Two systems were included in our pilot environment: WinMedicalc 2000 and ADAMEKj EHR.

Results

HL7-based local information models were created to describe the information content of both systems. The concepts from our original information models were mapped to coding systems supported by HL7 (LOINC, SNOMED CT and ICD-10) and the data exchange via HL7 v3 messages was implemented and tested by querying patient administration data. As a gateway between local EHR systems and the HL7 message-based infrastructure, a configurable HL7 Broker was developed.

Conclusions

A nationwide implementation of a full-scale SIP based on HL7 v3 would include adopting and translating appropriate international coding systems and nomenclatures, and developing implementation guidelines facilitating the migration from national standards to international ones. Our pilot study showed

that our approach is feasible but it would demand a huge effort to fully integrate the Czech healthcare system into the European e-health context.

Keywords

information storage and retrieval, Electronic health record, HL7, semantic interoperability, communication standards

DOI

10.3414/ME09-02-0018

The full English version can be found in /Methods of Information in Medicine 2010 49 2: 186-195/ or <http://www.schattauer.de/de/magazine/uebersicht/zeitschriften-a-z/methods/issue/special/manuscript/12381/show.html>.

Architectural Approaches for HL7-based Health Information Systems Implementation

D. M. López^{1,2}, B. Blobel²

¹Telematics Research Group, University of Cauca, Popayán, Colombia;

²eHealth Competence Center, Regensburg University Hospital, Regensburg, Germany

Summary

Objective

Information systems integration is hard, especially when semantic and business process interoperability requirements need to be met. To succeed, a unified methodology, approaching different aspects of systems architecture such as business, information, computational, engineering and technology viewpoints, has to be considered. The paper contributes with an analysis and demonstration on how the HL7 standard set can support health information systems integration.

Methods

Based on the Health Information Systems Development Framework (HIS-DF), common architectural models for HIS integration are analyzed. The framework is a standard-based, consistent, comprehensive, customizable, scalable methodology that supports the design of semantically interoperable health information systems and components.

Results

Three main architectural models for system integration are analyzed: the point to point interface, the messages server and the mediator models. Point to point interface and messages server models are completely supported by traditional HL7 version 2 and version 3 messaging. The HL7 v3 standard specification, combined with service-oriented, model-driven approaches provided by HIS-DF, makes the mediator model possible. The different integration scenarios are illustrated by describing a proof-of-concept implementation of an integrated public health surveillance system based on Enterprise Java Beans technology.

Conclusion

Selecting the appropriate integration architecture is a fundamental issue of any software development project. HIS-DF provides a unique methodological approach guiding the development of healthcare integration projects. The mediator model – offered by the HIS-DF and supported in HL7 v3 artifacts – is the more promising one promoting the development of open, reusable, flexible, semantically interoperable, platform-independent, service-oriented and standard-based health information systems.

Keywords

Systems Integration, HL7, semantic interoperability, system architectures, middleware, J2EE

DOI

10.3414/ME09-02-0012

The full English version can be found in /Methods of Information in Medicine 2010 49 2: 196-204/ or <http://www.schattauer.de/de/magazine/uebersicht/zeitschriften-a-z/methods/issue/special/manuscript/12775/show.html>.



EJBI 2009 ISSN 1801 - 5603

European Journal for Biomedical Informatics

Volume 5 (2009), Issue 1

Czech version

Cooperating journal – Lékař a technika

Editor-in-Chief: Karel Roubík (Czech Republic)

Publisher: Společnost biomedicínského inženýrství a lékařské informatiky ČLS JEP
a Fakulta biomedicínského inženýrství ČVUT v Praze

www.ejbi.eu



Content

Czech version

- | | |
|-----------|--|
| cs 2 - 3 | Předmluva: Komplexní pohled na eZdraví
J. Zvárová, B. Blobel |
| cs 4 - 8 | Hlasově ovládaná elektronická zdravotní dokumentace pro poruchy temporomandibulárního kloubu
R. Hippmann, T. Dostálová, J. Zvárová, M. Nagy, M. Seydlová, P. Hanzlíček, P. Kříž, L. Šmíd, J. Trmal |
| cs 9 - 22 | Medicína založená na omics vědách a systémová patologie - Nová perspektiva personalizované a prediktivní medicíny
H. Tanaka |

Předmluva

Komplexní pohled na eZdraví

J. Zvárová¹, B. Blobel²

¹Centrum biomedicínské informatiky, Ústav informatiky AV ČR, v.v.i., Praha, Česká Republika

²University Hospital Regensburg, eHealth Competence Center, Regensburg, Germany

Současné demografické, ekonomické a sociální podmínky ve vyvinutých zemích vyžadují změnu paradigmatu pro poskytování vysoce kvalitních a efektivních zdravotnických služeb. V tomto kontextu se musí zdravotnické systémy proměnit ze systémů zaměřených hlavně na organizaci na systémy orientované na procesy a nakonec směrem k celoživotní, individualizované, sdílené péci o pacienty, založené na službách platformy eZdraví. O tom hluboce přemýšlel např. Haux [1]. Požadavky interoperability na všedypřítomné personalizované zdravotnické služby přesahují stávající koncepty integrace zdravotnických informací mezi profesionálními poskytovateli zdravotnické péče. Budoucí personalizované zdravotnické platformy musí zejména zachovávat sémantickou interoperabilitu mezi systémy používajícími různé modality a technologie, různou reprezentaci znalostí a doménové expertní jazyky, stejně tak i různá kódovací schémata a terminologie, aby mohla být zahrnuta domácí péče i personalizované a mobilní systémy. Tento vývoj není omezen na regiony nebo země, ale objevuje se celosvětově a vyžaduje komplexní multidisciplinární a mezinárodní spolupráci. Nová translační charakteristika medicíny a důsledky pro zdravotnickou informatiku byly podrobně diskutovány Kuhnem a kol. [2] a také Kulikowskim [3].

Toto vydání shrnuje dnešní pohled na vybraná téma eZdraví, která byla prezentována na mezinárodní konferenci o eZdraví v Regensburgu, 2.-5. prosince 2007. V rámci summitu mezinárodních expertů nabídla konference zvané přednášky od mezinárodně renomovaných expertů reprezentujících všechny domény zastoupené v eZdraví. Poslední den této konference nabídla specifické semináře, kurzy a symposia adresované potenciálu spolupráce mezi Evropou

a Latinskou Amerikou (ELAN), které také analyzovaly přes-hraniční aktivity mezi Německem, Rakouskem, Českou republikou a Švýcarskem. Výsledky konference, která komplexně pokrývala eZdraví, byly publikovány v [4], [5]. Několik příspěvků ze summitu expertů nebo z veřejné konference bylo vědeckým programovým výborem vybráno k rozšíření a aktualizaci, aby mohly být publikovány v anglickém jazyce ve speciálním vydání časopisu Methods of Information in Medicine (Methods) a současně aby byly publikovány v národním jazyce v European Journal for Biomedical Informatics (EJBI), při spolupráci s Schattauer Verlagsgesellschaft mbH.

Tento ročník časopisu EJBI je uveden článkem pokrývajícím základní složky a architektonické problémy řešení interoperability eZdraví a zabývajícím se také ontologickými aspekty multidisciplinárního přístupu k eZdraví (Blobel [6]). Tato ontologická výzva je hlouběji diskutována v článku od Smithe a Brochhausena [7], kteří se zaměřují na reprezentaci formální ontologie, zatímco Daumke a jeho spoluautoři [8] diskutují sémantickou interoperabilitu reprezentace neformálních konceptů v medicíně. Ontologický aspekt byl zdůrazněn Oemigem a kol. [9] na příkladu Score systémů. Spyropoulos a jeho spoluautoři ukazují kontinuitu péče v pZdraví zahrnutím domácí péče pro kardio-respirační onemocnění a přidávající dimenzi biomedicínského inženýrství. Několik článků se zaměřuje na multimodální problém pZdraví, jako např. Krisciukaitis a kol. [11], kteří řeší algoritmický problém analýzy signálu nebo Hippmann a jeho spoluautoři [12], kteří se zamýšlejí nad aplikací hlasové komunikace v dokumentaci. Tanaka [13] začleňuje novou doménu medicíny založenou na omics vědách a systémové

patologii, která otevírá nové perspektivy pro personalizovanou a prediktivní péči. Cyklus uzavírá Nagy a kol. [14], kteří diskutují komplexní problém interoperability ovlivněného mezinárodními nomenklaturami, komunikačními standardy a komerčními programovacími technikami, stejně tak jako článek Lópeze [15], který se zabývá automatizovanou implementací interoperabilních řešení eZdraví, založenou na procesu soustředujícího se na architekturu a na procesu formálního vývoje.

Editoři ročníku 2009 časopisu EJBI (J. Zvárová a B. Blobel) by rádi poděkovali všem autorům za jejich mimořádnou snahu poskytnout články na vysoké vědecké úrovni, značně přesahující rámec konference. Dále děkují redakční radě Methods a nakladatelství Schattauer, že umožnily připravit tuto společnou publikaci s EJBI.

Literatura

- [1] Haux R. Individualization, globalization and health – about sustainable information technologies and the aim of medical informatics. *Int J Med Inf* 2006; 75: 795–808.
- [2] Kuhn KA, Knoll A, Mewes H-W, Schwaiger M, Bode A, Broy M, Daniel H, Feussner H, Gradinger R, Hauner H, Höfler H, Holzmann B, Horsch A, Kemper A, Krcmar H, Kochs EF, Lange R, Leidl R, Mansmann U, Mayr EW, Meitinger T, Molls M, Navab N, Nüsslin F, Peschel C, Reiser M, Ring J, Rummeny EJ, Schlichter J, Schmid R, Wichmann HE, Ziegler S. *Informatics and Medicine: From Molecules to Populations. Methods Inf Med* 2008; 47: 283–295.
- [3] Kulikowski CA. The Micro-Macro Spectrum of Medical Informatics Challenges – From Molecular Medicine to Transforming Health Care in a Globalizing Society. *Methods Inf Med* 2002; 41: 20–24.

- [4] Blobel B, Pharow P, Nerlich M (eds). *eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge – Global Experts Summit Textbook. Series Studies in Health Technology and Informatics, Vol. 134.* Amsterdam: IOS Press; 2008.
- [5] Blobel B, Pharow P, Zvárová J, Lopez DM (eds). *eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge – CeHR Conference Proceedings 2007.* Amsterdam: IOS Press; 2008, and Berlin: AKA GmbH; 2008.
- [6] Blobel B. Architectural Approach to eHealth for Enabling Paradigm Changes in Health. *Methods Inf Med* 2010; 49 (2): 123–134.
- [7] Smith B, Brochhausen M. Putting Biomedical Ontologies to Work. *Methods Inf Med* 2010; 49 (2): 135–140.
- [8] Daumke P, Schulz S, Müller ML, Dzeyk W, Prinzen L, Pacheco E.J, Secco Cancian P, Nohama P, Markó K. Subword-based Semantic Retrieval of Clinical and Bibliographic Documents. *Methods Inf Med* 2010; 49 (2): 141–147.
- [9] Oemig F, Blobel B. Semantic Interoperability Adheres to Proper Models and Code Systems – A Detailed Examination of Different Approaches for Score Systems. *Methods Inf Med* 2010; 49 (2): 148–157.
- [10] Spyropoulos B, Tzavaras A, Botsivaly M, Koutsourakis K. Ensuring the Continuity of Care of Cardiorespiratory Diseases at Home – Monitoring Equipment and Medical Data Exchange over Semantically annotated Web Services. *Methods Inf Med* 2010; 49 (2): 156–160.
- [11] Krisciukaitis A, Simoliuniene R, Tamosiunas M, Saferis V, Vainoras A, Gargasas L. Efficiency Evaluation of Autonomic Heart Control by Using the Principal Component Analysis of ECG P-Wave. *Methods Inf Med* 2010; 49 (2): 161–167.
- [12] Hippmann R, Dostalova T, Zvárová J, Nagy M, Seydlova M, Hanzlicek P, Kriz P, Smidl L, Trmal J. Voice-supported Electronic Health Record for Temporomandibular Joint Disorders. *Methods Inf Med* 2010; 49 (2): 168–172.
- [13] Tanaka H. Omics-based Medicine and Systems Pathology – A New Perspective for Personalized and Predictive Medicine. *Methods Inf Med* 2010; 49 (2): 173–185.
- [14] Nagy M, Hanzlicek P, Precková P, Riha M, Dioszegi M, Seidl L, Zvárová J. Semantic Interoperability in Czech Healthcare Environment Supported by HL7 Version 3. *Methods Inf Med* 2010; 49(2): 186–195.
- [15] López DM, Blobel B. Architectural Approaches for HL7-based Health Information Systems Implementation. *Methods Inf Med* 2010; 49 (2): 196–204.

Contact

Prof. RNDr. Jana Zvárová, DrSc.
 Centrum biomedicínské informatiky,
 Ústav informatiky AV ČR, v.v.i.,
 Pod Vodárenskou věží 2,
 182 07 Praha 8
 Česká Republika
 E-mail: zvarova@euromise.cz

Hlasově ovládaná elektronická zdravotní dokumentace pro poruchy temporomandibulárního kloubu

R. Hippmann^{1,2}, T. Dostálová^{1,2}, J. Zvárová^{1,3}, M. Nagy^{1,3}, M. Seydllová^{1,2}, P. Hanzlíček^{1,3}, P. Kříž^{1,2}, L. Šmíd¹, J. Trmal⁴

¹Centre of Biomedical Informatics, Czech Republic

²Department of Paediatric Stomatolgy, Second Faculty of Medicine of Charles University, Prague, Czech Republic

³EuroMISE Centre, Department of Medical Informatics, Institute of Computer Science, v.v.i., Academy of Sciences of the Czech Republic

⁴Department of Cybernetics, University of West Bohemia, Plzen, Czech Republic

Souhrn

Cíl: Vytvořit podporu pro vkládání strukturovaných dat do elektronické zdravotnické dokumentace u poruch temporomandibulárního kloubu.

Metody: Jsou popsány metody stukturování dat ve stomatologii a je představena interaktivní DentCross komponenta. Na několika případech je demonstrováno hlasem ovládané vkládání strukturovaných dat do zdravotnické dokumentace ve stomatologii. Dále je vysvětleno propojení této komponenty s elektronickým zdravotním záznamem MUDRLite.

Výsledky: Představujeme použití aplikace DentVoice složené z elektronického zdravotnického záznamu MUDRLite a hlasem kontrolované interaktivní komponenty DentCross ke sběru stomatologických informací nutných pro léčbu poruch temporomandibulárního kloubu.

Závěry: Aplikace DentVoice s komponentou DentCross prokázala praktické využití pro podporu léčení temporomandibulárních poruch.

Klíčová slova: elektronická zdravotnická dokumentace, ukládání strukturovaných dat, stomatologie, poruchy temporomandibulárního kloubu.

1. Úvod

Elektronická zdravotnická dokumentace (EHR) je definována jako zdroj informací týkajících se péče o zdravotní stav subjektu v počítačem zpracovatelné formě, ukládané, bezpečně sdílené a přístupné pro více autorizovaných uživatelů [1]. Její nejdůležitější cíl je podpořit plynulou, účinnou a vysoce ucelenou zdravotní péci pomocí poskytování plnohodnotné informace o jednotlivci. Obsahuje nejenom data o současném a minulém zdravotním stavu pacienta, medicínských okolnostech, vyšetřeních, ošetřeních a medikaci, ale nabízí také více pokročilé zpracování těchto dat a také funkci pro podporu rozhodování např. pro temporomandibulární poruchy [14].

použití soustředěné na péci a v reálném čase [6], zdokonalení technik utajení dat zlepšující bezpečnostní aspekty [7], [8], semantickou součinnost založenou na ontologických přístupech [9] nebo systém pro podporu rozhodování [10]. Komplexní a vyhovující EHR systém nebyl dosud mezinárodně uvedený na trh žádnou společností, proto je hodnocení různých EHR systémů velice důležité [11]. Jednou z možností jak vylepšit EHR je zakomponování hlasového ovládání do systému, což by zjednodušilo celý proces sběru dat.

Použití hlasových příkazů je experimentálně zkoušeno od roku 1990 [12] jako vhodná náhrada kontroly počítače pomocí klávesnice a myši. Nezbytnost použití lidského hlasu ke kontrole počítače nebo jiných zařízení vychází z prostředí, kde dochází k zaměstnání obou rukou, jako jsou chirurgie a stomatologie [13].

Hlasem ovládaná elektronická zdravotnická dokumentace také podporuje plynulou, účinnou a vysoce ucelenou zdravotní péci pomocí poskytování plnohodnotné informace o jednotlivci. Obsahuje nejenom data o současném a minulém zdravotním stavu pacienta, medicínských okolnostech, vyšetřeních, ošetřeních a medikaci, ale nabízí také více pokročilé zpracování těchto dat a také funkci pro podporu rozhodování např. pro temporomandibulární poruchy [14].

Hlavní důvody, proč neexistuje stálý žádný konsensus v klasifikaci poruch temporomandibulárního kloubu, je jejich nejasná etiologie, stejná klinická zjištění vyplývající z různých příčin a prokázaná souvislost mezi poruchami temporomandibulárního kloubu a psychologickými faktory [15, 16]. Všechna zmíněná fakta dělají z temporomandibulárních poruch velice

komplikovanou skupinu onemocnění pro tvorbu vhodného a kompaktního EHR systému. Jedno z mnoha běžně používaných diagnostických schémat určených pro výzkumné účely je Research Diagnostic Criteria for TMD (RDC/TMD). RDC/TMD standardizuje klinický výzkum u pacientů s poruchou temporomandibulárního kloubu, zlepšuje reproducibilnost mezi klinickými pracovníky a usnadňuje porovnávání výsledků mezi výzkumníky [15]. Temporomandibulární poruchy (TMD) a jejich nejčastější projevy a symptomy jako jsou bolest, limitované otvírání úst a zvukové kloubní fenomény mohou být způsobené poruchami svalů, intrakapsulární poruchou součástí temporomandibulárního kloubu a degenerativními změnami kostních součástí kloubu [15]. Poruchy temporomandibulárního kloubu jsou pokládány za podskupinu muskuloskeletálních poruch [15], [16], [17]. To může vysvětlit zprávy o úspěšném použití fyzikální terapie v léčbě poruch temporomandibulárního kloubu.

Naše aplikace byla vytvořena jako uživatelsky přátelský program, který sjednocuje celý žvýkací aparát a jeho problematiku, protože všechny jeho části jsou propojeny. Systém byl ještě rozšířen o modul automatického rozpoznávání řeči.

2. Multimedia distribuující elektronickou zdravotnickou dokumentaci

Vývoj elektronické zdravotní dokumentace v EuroMISE centru začal v roce 2000 na základě inspirace a zkušenostech s existujícími CEN/TC251 standardy a několika evropskými projekty, převážně projekty I4C a TripleC [18]. modifikací nastavení shromažďovaných atributů bez jakékoli změny struktury databáze.

Hlavní požadavek na navrhovaný systém bylo ukládání strukturovaných dat kombinovaných s volným textem s možností dynamického rozšíření a modifikací nastavení shromažďovaných atributů bez jakékoliv změny struktury databáze. Z výzkumu vzešla pilotní EHR aplikace nazvaná MUDR (Multimedia Distributed EHR). Jiná související elektronická zdravotní dokumentace nazvaná MUDRLite [21] byla vyvinuta v roce 2003 s důrazem na potřeby prostředí malého měřítka. MUDRLite působí jako převaděč povelů, zpracovává instrukce do MLL (MUDRLite Language) [22], který je založený na XML a ovládá rozhraní databáze stejně jako vizuální aspekty a chování uživatelského rozhraní MUDRLite. Pilotní aplikace byla připravena pro oblast stomatologie [13].

Hlavní architektura MUDR EHR byla založena na třívrstvé architektuře, která se skládá z rozhraní databáze, rozhraní aplikace a uživatelského rozhraní. Sada shromážděných atributů a relací mezi nimi je uložena v řízené diagramové struktuře, která se nazývá „znalostní báze“. Vrcholy diagramu popisují shromážděné atributy pomocí jejich jedinečných identifikátorů, interních jmen, fyzikálních typů dat a jiných auxiliárních informací, okraje diagramu popisují vztahy mezi atributy. Dominantní okraj typu „inferior“ definuje hierarchickou strukturu stromu znalostní databáze, takže znalostní databáze může být popsána přímým lesem s několika stromy. Tyto stromy jsou také nazývány „domény znalostní báze“. Každý strom v diagramu popisuje data jednoho pacienta. Každý uzel ve stromě popisuje jeden příklad lékařského konceptu ze znalostní báze. Rozhraní aplikace se skládá ze čtyř základních komponent: HTTP server používaný pro komunikaci s klientskou aplikací, EHR-AppL služba vykonávající hlavní logiku aplikace, CGI-script (potencionálně více z nich) složící jako rozhraní mezi HTTP serverem a HER-AppL službou a eventuální moduly medicínských pokynů.

MUDRLite architektura je založena na dvou vrstvách. První je relační databáze (e.g. MS SQL) a druhá je vrstva MUDRLite uživatelského rozhraní (MUDRLite UI). Schéma databáze koresponduje s detailními potřebami na cílovou doménu

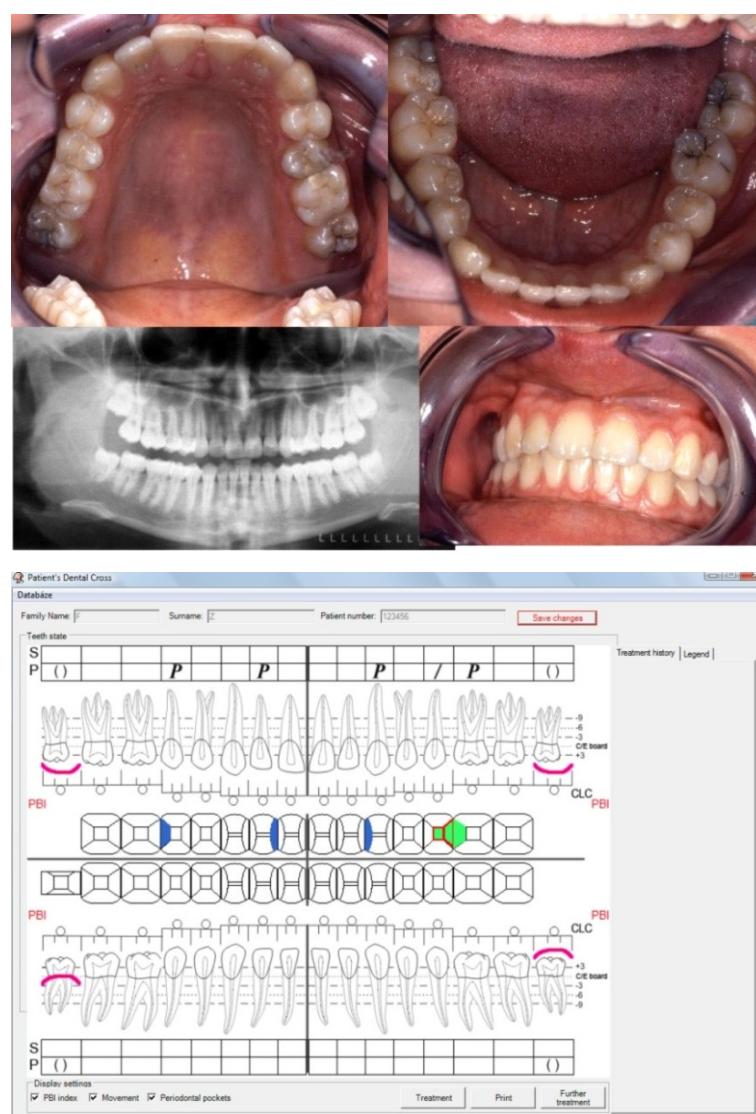
a proto se mění signifikantně v rozdílných prostředích, protikladně k fixnímu schématu databáze ve vrstvě MUDR data. Jádro MUDRLite MUDRLite Interpreter je schopno pracovat s různými schématy databáze. Tento rys často zjednoduší proces aktualizace dříve uložených dat v jiných databázích nebo složkách. Vizuální aspekty stejně jako chování MUDRLite UI jsou kompletně popsány pomocí XML konfigurační složky. Konečný uživatel může vidět sadu modelů s různými na nich umístěnými ovládáními díky odpovídajícím XML prvkům.

3. Interaktivní DentCross komponenta s nástrojem pro automatické rozpoznávání řeči (ASR)

Pro možnost využití MUDRLite ve stomatologii byla vyvinuta vysoce vyspělá

komponenta interaktivní zubní kříž, která představuje grafickou část stomatologické dokumentace [13]. DentCross komponenta vznikla jako samostatná knihovna DentCross.dll, kompletně vyvinutá pro .NET Framework platformu za použití vývojového nástroje Microsoft Visual Studio.NET 2003.

Uživatelem definovaná složka je vkládána pomocí zákaznického elementu MLL s následujícími závaznými atributy: „dll“ specifikuje jméno souboru, kam je prvek zařazován a „class“ specifikuje jméno hlavní třídy obsaženého prvku. Tento prvek je plně interaktivní. Stomatolog má možnost výběru z asi 60 různých postupů, ošetření nebo parametrů zubů, které jsou zobrazeny graficky a v jednoduše organizovaném stylu (Obr. 1).



Obr. 1. Interaktivní DentCross.

Schéma základní databáze elektronické zdravotnické dokumentace reflekтуje strukturu podskupiny rozsáhlé hierarchicky uspořádané znalostní báze, vytvořené pro obor stomatologie v českém i anglickém jazyce. Databáze obsahuje charakteristiky každého samostatného zuba lokalizace (meziálně, incizálně, incizálně-distálně, incizálně-meziálně, distálně, cervikálně-vestibulárně, cervikálně-orálně, vestibulárně a orálně, třídy dle Mounta 1, 2, 3 a 4) a typ (primární, sekundární) kazivé léze, ageneze, periodontitida, nekróza, gangréna, charakteristiky ošetření lokalizaci a typ materiálu (amalgám, kompozit, kompomer, skloionomer, provizorní výplň) výplně, rozdílné typy implantátů, korunky, inleje, lokalizace a materiál můstků, extrakce a přidatné charakteristiky jako indexy krvácivosti, zubní kámen a atributy periodoncia. Rozsáhlá znalostní báze používá různé známé klasifikace jako jsou Baumeho klasifikace pulpítid, Blackova klasifikace kavit, Angleovu klasifikaci okluze, Voldřichovu klasifikaci defektů po ztrátě zuba, Kennedyho klasifikaci částečně bezzubých stavů a částečných náhrad, klasifikace onemocnění dásní a periodoncia dle WHO atd.. Technologie representace znalostní báze byla patentována Český národní patent č. PV 2005-229.

Dále byl přidán nástroj automatického rozpoznávání řeči (ASR) jako samostatně stojící aplikace běžící na pozadí. Protokol komunikace serveru je patentovaný a běží na vrcholu zásobníku TCP/IP. Komunikační protokol umožnuje začátek a ukončení rozpoznávacího procesu, konfiguraci rozpoznávací úlohy v průběhu stejně jako přijímání rozpoznaných frází klientem.

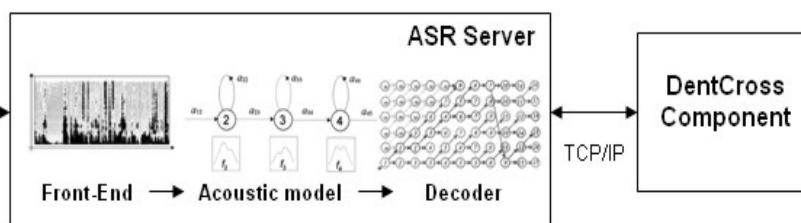
ASR systém je nezávislý na osobě mluvícího a používá statistické techniky. Schéma modulu ASR můžeme vidět na (Obr. 2). Řečový signál je digitalizovaný na 8kHz hladině. Poté je upravená akustická vlna segmentována do 25 milisekundových rámců s 15ms přesahem. Akustický model je založený na modelování trifónů. Každý individuální trifón je reprezentován třístavovým levoprávným HMM (Hidden Markov Model) se spojitou výstupní funkcí pravděpodobnostní density přiřazenou ke každému stavu. Dekoder využívá Viterbi hledací techniku kombinovanou s efektivním algoritmem prořezávání [23].

Prototyp aplikace DentVoice spojuje DentCross komponentu a TCP/IP klient ASR serveru s definujícím souborem hlasových příkazů. ASR klient užívá třídu DentCrossHandler, která implementuje funkčnost komponenty DentCross.

Hlasové rozpoznávání je aktivováno bezprostředně po spuštění DentCross komponenty. Proces rozpoznávání může být přerušen nebo zastaven speciálním hlasovým příkazem nebo přes uživatelské rozhraní. Hlasové příkazy mohou být rozděleny do dvou skupin: obecné příkazy a kontextově vázané příkazy.

Obecné příkazy jsou určeny k ovládání procesu rozpoznávání např. pauza, začít znova, stop atd. a k uzavírání oken hlášení otevřených aplikací k upozornění uživatele.

Kontextově vázané příkazy se vztahují k aktuálnímu stavu komponenty DentCross a mohou být dále rozděleny do 33 skupin příkazů korespondujících s 33 stavami (např. charakter ošetření zuba, umístění kazu, typ kazu, typ materiálu kořenové výplně).



Obr. 2. Funkčnost ASR.

4. Hlasově a počítačově podporovaná léčba pacientů s parafunkcí temporomandibulárního kloubu

Základ pro úspěšné ošetření je stanovení příčiny problémů pacienta [24]. Nová zjištění související s etiologií TMD vedly ke změnám v terapeutických postupech. K podpodpoře přirozené schopnosti hojení muskuloskeletálního systému je nyní preferován komplex konzervativních ošetření před irreverzibilní chirurgickou nebo protetickou terapií.

Kazuistika

Toto kazuistické sdělení ukazuje propojení automatického rozpoznávání řeči a komponenty zubního kříže u léčby TMD.

Na základě doporučení od zubního chirurga byl pacient vyšetřen pro časté kloubní zvukové fenomény a dlouhotrvající bolest v oblasti temporomandibulárních kloubů, pravděpodobně způsobených snížením skusu. Pacient netrpěl bruxismem nebo jinými parafunkcemi. Metody strukturování informací ve stomatologii mohou být ilustrovány na využití hlasem podporované interaktivní DentCross komponenty. Interaktivní DentCross komponenta pomáhá ve shromažďování stomatologických informací, přípravě terapeutického plánu a monitorování terapeutického procesu. Všechno je spojeno se sběrem anamnestických dat do speciálního formuláře, který je součástí zubního kříže.

Plán ošetření:

1. záznam do interaktivního zubního kříže stomatologické vyšetření a anamnestická data,
2. nácvik otevírání úst v ose, bez deviací,
3. stabilizační dlaha,
4. analýza pomocí počítačového obličejového oblouku,
5. protetická léčba,
6. dokončení léčby a měkká dlaha,
7. ověření výsledku a hodnocení terapie pacientem.

Předošetření

Během procesu předošetření byla použita komponenta DentCross k podpoře grafické analýzy horního a dolního zubního oblouku.

Záznamem celého vyšetření a anamnestických dat byla umožněna analýza celkové situace a komplexní přehled o stavu (Obr. 1).

Terapie

Pacientovi bylo doporučeno neotvírat ústa do maximální polohy, vyhýbat se ukusování tvrdé potravy a žvýkání žvýkaček. Pomocí tvrdé akrylátové stabilizační dlahy bylo docíleno překrytí zubů v horní čelisti, to učinilo jejich povrch hladký. To nám zajistilo centrální okluzi, eliminaci postranního působení, poskytnutí předního vedení pro frontální zuby a redukci neuromuskulární aktivity. Po 2 měsících tohoto režimu byla zjištěna redukce bolesti a frekvence kloubních zvukových fenoménů v levém TMK.

Poté byla iniciována protetická terapie nesprávného mezičelistního vztahu. Konečná fáze analýzy byla provedena v plně seřiditelném artikulátoru Protar Evo 7 (KAHO, Germany) a přenosem pomocí počítačového obličejoblého oblouku Arcus Digma (KAHO, Germany) (Obr. 3).

5. Diskuze

Dokončené analýzy nynějších komerčně dostupných programů a patentovaných technologií naznačují, že softwarová podpora pro ukládání stomatologické zdravotnické dokumentace se pomalu vyvíjí [18]. Nové trendy jsou nalézány v navrhování objektu příbuzných modelech založených na strukturovaných znalostních prvcích, které jsou dynamicky využitelné multimediálními nástroji pro dokumentaci jednotlivých případů, simulaci vývoje nemoci, a podporu rozhodování [19] a zaměření na přímé požadavky uživatele [20]. Prezentovaná interaktivní komponenta DentCross signifikantně zvyšuje kvalitu elektronické zdravotnické dokumentace. Grafický design DentCross UI je specificky konstruován, aby vypadal jako fotografie zubního oblouku kombinovaná s rentgenovým snímkem (t.j. obraz kořenového kanáku nebo implantátu). Index krvácivosti papil, viklavost, zubní kámen, kostní resorpce a periodontální choboty a informace o temporomandibulárním kloubu pacienta mohou být také dokumentovány.

Po protetické terapii byla zhotovena měkká dлага na zuby dolní čelisti a pacientka jí nosila vždy na noc. Měkká náhusná dлага a protetická terapie přispěly k poklesu zatížení TMK, což vedlo k redukci bolestivosti v oblasti TMK. Abrase byla prakticky eliminována.

TMK artralgie byla, jmenovitě v prezentovaném případě, pravděpodobně způsobena opakovanou traumatizací TMK pro nesprávný stereotyp otvírání úst, chronické přetěžování TMK jako následek nestabilní okluze a posunutí disku spojené s irritací retrodiskální tkáně [12], [16].

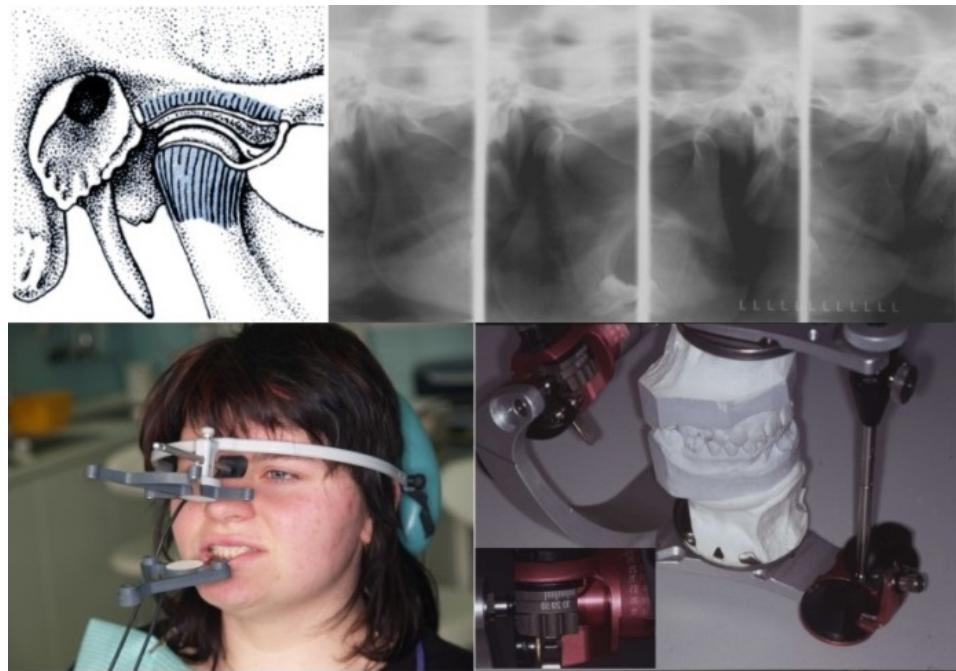
Naše zkušenosť s hlasem ovládaným vkládáním dat naznačuje, že tento typ kontroly se velice hodí pro aplikace v medicínském prostředí a jeho použitelnost není limitována pouze pro stomatologickou praxi. Hlasem ovládané vkládání dat je užitečné všude, kde by jiná varianta vkládání mohla negativně ovlivnit uživatelův výkon z pohledu nárůstu časové náročnosti na splnění zadaného úkolu nebo pocit nárůstu diskomfortu. Tyto situace často nástavají např. (ale ne vždy) během vyšetření pacienta, chirurgického výkonu nebo během manuální mikroskopické analýzy vzorků tkání nebo krve.

V těchto oblastech může být systém hlasového rozpoznávání s velkoobsahovým slovníkem velkým přínosem (HVV-SRS), který odstraní potřebu speciálního UI adaptéra pro hlasovou kontrolu frázového rozpoznávání. Avšak potřeba doménově specifického jazykového modelu pro HVV-SRS je hlavní výzvou v aplikování HVV-SRS do praxe.

Nastávající vývoj bude směrován k užití počítacem syntetizované řeči. DentCross by předčítal nahlas aktuální stav pacientova chrupu, jak byl uložen v databázi a zubní lékař by pouze kontroloval, zda je uložená informace odpovídající.

6. Závěr

MUDRLite EHR s interaktivní komponentou DentCross umožňuje přehledný zdravotnický záZNAM s celou denticí a provedení vyšetření pacienta v koncentrované formě. Stomatologická informace zaznamenaná v obvyklé grafické struktuře urychluje rozhodování lékaře a přináší kompletní pohled na shromážděnou informaci. Aplikace DentVoice (hlasem kontrolovaná verze DentCross) se nyní testuje při procesu získávání a hodnocení stomatologických dat.



Obr. 3. Počítačem podpořená analýza TMK.

Tento přístup nemusí jenom ulehčit ukládání strukturovaných dat ve stomatologické praxi, ale může také podpořit terapii TMK poruch. Výše popsaná aplikace MUDR EHR s komponentou DentCross tedy otvírá nové možnosti ukládání a klasifikace dat ve stomatologii.

Poděkování

Práce byla podpořena projektem 1M06014 Ministerstva školství, mládeže a tělovýchovy ČR a projektem AV0Z10300504 Ústavu informatiky Akademie věd ČR, v.v.i.

Literatura

- [1] Kalra, D.: Electronic Health Record Standards. IMIA Yearbook of Medical Informatics 2006; Methods Inf Medicine 2006; 45, Suppl 1; 136-144.
- [2] Los, R.K., van Ginneken, A.M., van der Lei, J.: OpenSDE: a strategy for expressive and flexible structured data entry. Int J Med Inform 2005, 74:481-490.
- [3] van Ginneken, A.M.: The computerized patient record: balancing effort and benefit. Int J Med Inf 2002, 65:97-119.
- [4] Bakker, A.: Access to EHR and access control at a moment in the past: A discussion of the need and an exploration of the consequences. Int J Med Inform 2004; 73:267-270.
- [5] Blobel, B.: Advanced EHR Architecture - Promises or Reality. Methods Inf Med 2006; 1: 95-101.
- [6] Reuss, E., Menoyyi, M., Buchi, M., Koller, J., Krueger, H.: Information access at the point of care: what can we learn for designing a mobile CPR system? Int J Med Inform 2004; 73:365-369 .
- [7] Pharow, P., Blobel, B.: Electronic signatures for long lasting storage purposes in electronic archives. Int J Med Inform 2005; 74:279-287.
- [8] Sax, U., Kohane, I., Mandl, K.D.: Wireless technology infrastructures for authentication of patients. PKI that rings. J Am Med Inform Assoc 2005; 12:263-268.
- [9] Min, Z., Baofen, D., Weeber, M., van Ginneken, A.M.: Mapping Open SDE domain models to SNOMED CT. Methods Inf Med 2006; 1: 4-9.8
- [10] Gallanter, W.L., Didomenico, R.J., Polkaitis: A trial of automated decision support alerts for contraindicated medications using physician order entry. J Am Med Inform Assoc 2005; 12:269-274.
- [11] Noehr, C.: Evaluation of electronic health record systems. IMIA Yearbook of Medical Informatics 2006. Methods Inf Med 2006; 45 Suppl.1: 107-113.
- [12] Grasso, M.A.: Automated speech recognition in medical applications. MD Computing 1995; 12: 16-23.
- [13] Zvarova, J., Dostalova, T., Hanzlicek, P., Teuberova, Z., Nagy, M., Pies, M., Seydlova, M., Eliasova, H., Simkova, H.: Electronic health record for forensic dentistry. Methods Inf Med 2008; 47: 8-13.
- [14] Dostalova, T., Seydlova, M., Zvarova, J., Hanzlicek, P., Nagy, M.: Computer-supported treatment of patients with the TMJ parafunction. eHealth: Combining telematics, telemedicine, biomedical engineering and bioinformatics to the edge. IOS press AKA, Berlin 2008; 171-177.
- [15] Dworkin, S.F., LeResche, L.: Research diagnostic criteria for temporomandibular disorders: review, examinations and specifications criteria. J Craniomand Disord Facial Oral Pain 1992; 6: 301-55.
- [16] McNeill, Ch.: History and evolution of TMD concepts. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 1997; 83: 51-60.
- [17] Goldstein, B.: Temporomandibular disorders: A review of current understanding. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 1999; 88: 379-85.
- [18] Van Ginneken, A.M., Stqam, H., van Mulligen, E.M., de Wilde, M., van Mastrigt, R., van Bemmel, J.H.: ORCA: the versatile CPR. Methods Inf Med 1999; 38:332-338.
- [19] Koch, S., Risch, T., Schneider, W., Wagner, I.V.: An object-relational model for structured representation of medical knowledge. Int J Comput Dent 2006; 9(3):237-252.
- [20] Koch, S.: Designing clinically useful systems: examples from medicine and dentistry. Adv Dent Res 2003; 17:65-68.
- [21] Hanzlicek, P., Spidlen, J., Heroutova, H., Nagy, M.: User Interface of MUDR Electronic Health Record. International Journal of Medical Informatics 2005; 74: 221-227.
- [22] Spidlen, J., Hanzlicek, P., Zvarova, J.: MUDRLite - health record tailored to your particular needs. Studies in health technology and informatics 2004; 105: 202-9.
- [23] Psutka, J., Muller, L., Matousek, J., Radova, V.: Mluvíme s počítačem česky. Praha, Academia, 2006 (in Czech).
- [24] Carlsson, G.E., Magnusson, T. (eds.): Management of temporomandibular disorders in the general dental practice. Quintessence Publishing Co, Inc. 1999; 108-129.

Medicína založená na omics vědách a systémová patologie Nová perspektiva personalizované a prediktivní medicíny

H. Tanaka¹

¹Graduate School of Biomedical Science Tokyo, Medical and Dental University, Tokyo, Japan

Shrnutí

Cíle: Poslední významné pokroky v lidské genomice a postgenomických „omics“ vědách nyní přinášejí nové možnosti zdravotní péče, kterou nazýváme „medicína založená na omics vědách“. V tomto článku jsme zkoumali vývoj a budoucí možnosti medicíny založené na omics vědách.

Metody: Rozdělili jsme vývoj medicíny založené na omics vědách do tří generací za účelem ujasnění hlavních klinických cílů a charakterizace informačních metod každé generace spolu s jejich budoucími možnostmi.

Výsledky: První generace omicsové medicíny začala s „genomickou medicínou“, založenou na vrozených individuálních rozdílech v genomu. Toto otevřelo studování genetických polymorfismů nemoci a napomohlo vývoji personalizovaných léků založených na farmakogenetických/ farmakogenomických rozdílech v reakci na lék. Ve druhé generaci omicsové medicíny, díky pokroku v technologii vysokorychlostního sekvencování, byl otevřen přístup k obrovskému množství různých omicsových dat postgenomických nemocí, obsahujících vyčerpávající molekulární informace o nemocech somatických buněk. Zobrazení průběžného stavu nemoci je tak mnohem věrnější a umožnuje tak prediktivní medicíně předpovězení prognózy nemoci uplatněním datové analýzy. A také, díky rychle se rozvíjejícím znalostem o buňkové molekulární síti, je nyní umožněno porozumět nemoci na systémové úrovni pomocí nového oboru, zvaného systémová patologie. Může plně využít významné znalosti omicsu nemoci a povede ke komplexnímu porozumění postupu nemoci použitím datové analýzy.

Závěr: Omicsová medicína a systémová patologie umožní nový druh personalizované a prediktivní medicíny.

Klíčová slova: Medicína založená na omics vědách, systémová patologie, genomika, omics vědy, farmakogenomika.

1. Úvod

Poslední rychlý pokrok v lidské genomice a následné studie postgenomické komprehensivní molekulární informace, souhrně nazývané „omics“ [1], [2], jako například transkriptomika, proteomika a metabolomika, dávají vzniknout novým možnostem v medicíně. Na základě těchto možností hraje rychle se vyvíjející informatika, nazývaná „klinická bioinformatika“ [3] nebo podle novějšího termínu „translační informatika“ [4], nezastupitelnou roli při odvozování klinicky významné informace z obrovského množství omicsových dat.

Až do současné doby bylo použití komplexní molekulární informace v medicíně označováno jako „genomická medicína“ [5], [6], která si dává za úkol poskytovat „personalizovanou lékařskou péči“, založenou na vrozených (germline) individuálních rozdílech nebo li „polymorfismech“ pacientovy genomické informace. Některé druhy personalizované lékařské péče už byly zavedeny do klinické praxe. Například „personalizovaná medicína“ se nyní klinicky používá u některých léků, u kterých je nutné před-predpisové genotypování nebo proteinová analýza. Tento před-diagnostický test zabraňuje nežádoucím vedlejším účinkům a zabezpečuje účinnost těchto léků u individuálních pacientů. Mezitím, díky posledním pokrokům ve vysokosekvenční technologii, jako například DNA mikroarrays [7], [11] nebo hmotnostní spektrometrie (MS) [12], [13], se postgenomické omicsové informace nebo prostě „omicsová“ data stala dostupná v klinickém kontextu. Mnoho různých studií prokázalo, že omicsová informace poskytuje komplexnější a významnější

informace vztahující se k postupu nemoci, než jen pouhou „genetickou dispozici“ k této nemoci, takže by mohla být použita ve vlastní klinické praxi na účely jako jsou například raná diagnóza, klasifikace poddruhů nebo předpověď onemocnění [14], [16]. Na tomto základě lze očekávat, že omicsové informace pozorované ve stavu nemoci („omics nemoci“) mohou přinést nový druh lékařské péče, který bude více prediktivní a preventivní než běžná genomická medicína. Tato nová etapa molekulární medicíny potřebuje nový název, aby se odlišila od genomické medicíny. Můžeme jí jednoduše nazývat „medicína založená na omics vědách“ nebo „omicsová medicína“. (omics based medicine) [17].

Při používání omicsu nemoci pro preventivní a prediktivní medicínu byla omicsová data nejprve analyzována přímo na své úrovni aplikací vytěžování dat nebo exploratorní statistiky („datová analýza omicsových dat“). Jako typický příklad můžeme uvést vhodně zvolené sady genů zvané „signatury“, které byly určeny pomocí aplikací získaných dat nebo pomocí exploratorní statistiky na profily genové exprese nemocných buněk za účelem stanovení předpovědi nemoci, jako například rekurence rakoviny během několika let po operaci [14], [16]. Jako jiný příklad můžeme uvést analýzu statistického vzorku proteomických dat. Extrakce charakteristických vzorků z hmotnostního proteinového spektra za použití statistické analýzy se provádí pro ranou detekci rakoviny [18], [19]. Tyto pokusy dosáhly v některých případech překvapivých úspěchů, ale ve většině případů nebyly výsledky pro další klinické aplikace dostatečné. Brzy se příšlo na to, že obrovské množství omicsových dat s jejich extrémně vysokou vzájemnou závislostí, často zabraňuje postupům založených na datech dosažení směrodatných výsledků.

Na základě rychle se rozvíjejících znalostí o celulární molekulární síti a jejích změnách při nemozech je zřejmé, že s výjimkou vzácných genetických nemocí není většina nemocí způsobena mutací jednoho nebo dvou genů, ale je spíše způsobena kooperativním účinkem aberací několika genů a proteinů. Tyto společné účinky způsobují změny v signálních kanálech nebo regulačních sítích genů, čímž produkují fenotypy nemocí. Takže bychom lépe porozuměli nemocem jako fenotypům, způsobených „systémovým narušením molekulární sítě“, díky vzájemně propojenému selhání genů a proteinů. Tento pohled na nemoci je obzvláště pravdivý pro běžné komplexní nemoci, jako například běžné typy rakoviny, které jsou nyní považovány za onemocnění dráhy („pathway disease“) [20]. Takže systémové porozumění nemocem založené na znalosti změn molekulární dráhy se nyní považuje za klíčové pro plné využití podstaty obsahu omicsu nemoci a ke komplexnímu porozumění toho, co tyto nemoci způsobuje („modelová analýza omicsových dat“).

My navrhujeme tento druh systémového přístupu k nemocem jako „systémovou patologii“ již několik let [21], [22] v tom smyslu, že se jedná o správnou aplikaci systémové biologie na nemoci [23], [24], [25]. Toto stádium omicsové medicíny, především té, která je založena na porozumění systémové patologii řetězce nemoci, by se mohlo nazývat „omicsová systémová medicína“ nebo ji můžeme jednoduše nazývat „systémovou medicínou“ jako to dělají její ostatní zástánci [26], pokud je zamýšlený význam tohoto konceptu správně chápán.

V každém případě bychom rádi zdůraznili, že tento nový druh medicíny, založený na obrovském množství omicsových dat a porozumění celému procesu nemoci na systémové úrovni, by nebyl možný a nebude se moci vyvíjet v budoucnosti bez spolupráce tak zvané (postgenomické) omics informatiky“ nebo „systémové patologie“.

V tomto článku popisujeme stěžejní rysy a vývoj omicsové medicíny a informatiky používané tak, jak jsme stručně naznačili výše. Není třeba zdůrazňovat, že

genomika je v omics běžně zahrnována. Takže když popisujeme vývoj omicsové medicíny, začneme stručným přehledem genomické medicíny, jakožto její první generace a potom přejdeme k postgenomické omicsové medicíně jako ke druhé a současné generaci, a nakonec popíšeme omicsovou systémovou medicínu jako třetí generaci s jejími možnostmi do budoucnosti. Domníváme se, že tento „trigenerační model“ je velmi vhodný pro debatu o podstatě a hlavních klinických cílech každé generace, spolu s informatikou specifickou pro každou generaci používanou pro dosažení těchto cílů.

2. Genomická medicína jako první generace omicsové medicíny stručný přehled

2.1 Genomická medicína pro personalizovanou péči

Jak je všeobecně známo, molekulární medicína začala „genomickou medicínou“, která představuje novou medicínu, jejichž snahou je uskutečnění personalizované péče založené na individuálních rozdílech v genomické informaci. Díky pokroku v projektu lidského genomu [27] se přišlo na to, že lidský genom má individuální vrozené rozdíly. V lidském genomu existují různé typy polymorfismů, jako například různý počet tandemových repeticí (VNTR), mikrosatelitů (opaková jednotka je jedna až několik bps) nebo minisatelitů (9 až desítky bps), inzerčně-delečních polymorfismů (indel), variace počtu kopí (CNV) a tak dále [28]. Ale v současné době je nejběžněji používaným polymorfismem jednonukleotidový polymorfismus (SNP), který se vyskytuje v průměru u každých 300-1000 nukleotidů v lidském genomu (přibližně 0,1-0,3 %), takže v každém individuálním lidském genomu se vyskytuje celkem 3 až 10 milionů SNP polymorfismů.

Z klinického pohledu jsou nejdůležitější dva druhy klinických fenotypů, způsobené výše zmíněnými genomickými polymorfismy:

1. náchylnost k nemoci (kauzalita) a
2. reakce na léky.

Bod číslo jedna vyjadřuje genetický risk výskytu nemoci, zatímco bod číslo dvě se zabývá rozdíly v individuální reakci na léky.

V poslední době jsme viděli mnoho vynikajících článků, jako například [5], [6] vysvětlující „genomickou medicínu“. Nebudeme ji zde popisovat v detailu, ale pouze si ji krátce představíme s důrazem na rozdíly od dalších (druhé a třetí) generací omicsové medicíny, které budou centrálním námětem tohoto článku.

2.2 Genomická medicína a studie genů nemoci

Genomické studie lidských nemocí v podstatě začaly v osmdesátých letech minulého století, kdy byly poprvé k dispozici genetické mapy celého genomu (i když pouze hrubé) a dostatečný počet DNA markerů (náhodně rozmístěných v genomu); nejdříve využitím RFLP (polymorfismus délky restrikčních fragmentů) [28] a později mikrosatelitů. Genetické mapy a DNA markery poskytují rámec pro určení lokace genů způsobujících nemoc v lidském genomu („poziční klonování“) [29]. Společně s pokrokem ve statistické genetice [30], jako je například Mendeliánská vazbová analýza založená na rodokmenech, uspělo několik genetických studií v pozičním klonování genů zodpovědných za nemoci, jako například za Huntingtonovu nemoc [29], [30], [31], Duchennovu svalovou dystrofií [32] (DMD/BMD-Duchenne/Beckerova muskulární dystrofie) nebo cystickou fibrózu [33]. Skutečnost je taková, že tyto úspěchy v identifikaci genů způsobujících nemoci přinesly motivaci začít s projektem lidského genu.

Ačkoliv rané genetické studie byly úspěšné v identifikaci různých kauzativních genů monogenových nemocí, jejich prevalence je relativně nízká. Je klinicky mnohem důležitější objasnit genetický původ polygenových nebo multifaktoriálních nemocí, jako jsou běžné komplexní nemoci jako například astma, rakovina, cukrovka, srdeční choroby nebo duševní poruchy. Ale na rozdíl od monogenních nemocí v těchto případech pozorujeme několik set „genů podezřelých z nemoci“ s nízkým relativním genotypickým riskem (GRR) [34], což znamená, že riziko spojené s určitým genotypem je u polygenních nemocí relativní k jinému genotypu (non-carrier, tj. ne-nositel).

Určitý SNP se může vztahovat ke genu náchylnosti k nemoci mechanismem zvaným „vazbová nerovnováha“ [35], což přináší vysoký společný výskyt SNP a genů náchylnosti k nemoci. V takovém případě může být tento SNP použit jako marker, který, pokud je nalezen, poskytuje zvýšený relativní genotypový risk pro tuto nemoc. Ale je obecně vzácné najít markery pro geny náchylnosti s relativně vysokým GRR, zvláště pro běžné komplexní nemoci.

Abychom překonali tyto překážky, byly podniknutы rozsáhlé studie zvané GWAS neboli celogenomové asociační studie [36], [37]. Studie GWA (GWAS) se provádí rychlým skenováním statisíců markerů v celých genomech mnoha lidí za účelem nalezení genetických variací spojených s určitou chorobou. Bylo provedeno mnoho rozsáhlých GWA studií. Velmi známé jsou například studie cukrovky, protože různé studie GWA objevily nový společný gen náchylnosti TCF7L2 [37], [38], [39]. Studie GWA začaly být možné díky DNA arrays (vzorky sekvence, mikročipy), které obsahují více než 500 tisíc SNP v jednom čipu. Velký rozdíl populace a srozumitelnost sond SNP v čipu [40] umožňuje GWA studiím vysoké možnosti detekce genů náchylnosti k nemoci.

Jiný přístup ke získání vyšších hodnot náchylnosti u polygenních nemocí je použití kombinovaného vzorku mnoha SNP, zvaných „SNP haplotyp“. Termín „haplotyp“, což je zkratka pro haploidní genotyp, je definován jako sada alel blízce svázaných loci na stejném chromozomu, které mají tendenci být zděděny společně. Podobně je SNP haplotyp definován jako určitý typ sekvenčních SNP na jednom chromozomu, které jsou statisticky příbuzné. Pro získání produktivního odhadu rizika nemoci pro polygenní nemoci je však jednotlivý SNP nedostačující. Použitím SNP haplotypu u studie asociace haplotypu a nemoci je možné získat přesnější relativní riziko nemoci. Hromadné sbírání tohoto druhu informace je náplní současné probíhajícího mezinárodního projektu Hapmap [41].

Studie genu nemoci za účelem určení mutací nebo polymorfismů genů

spojeného s nemocí (kausativní nebo susceptibilní gen) ukazují detailní individuální rozdíly v patogenezi nemoci. To znamená, že toto by přispělo k personalizované medicíně a k možnosti budoucího předpovídání výskytu nemoci u specifických jednotlivců. Avšak zdá se, že tento obor genomické medicíny není schopen okamžitého účinného přispěvku ke klinické praxi, protože dokonce „vysoká GRR“ susceptibility genu pro běžné nemoci je většinou zhruba 1,5, ale přesto se různé pokusy jako GWAS nebo SNP Haplotypová analýza stále zkoumají. Bylo by lepší se smířit s tím, že současné studie nemocných genů přispívají vědeckému ujasnění genetického původu nemocí, spíše než v nich vidět okamžité aplikace pro vlastní klinické využití.

Zde můžeme vidět výrazný rozdíl proti jinému oboru genomické medicíny „personalizované medicíně“, kde genomický polymorfismus s léky spojených genů někdy způsobí výrazné rozdíly v reakci na lék. Důvod by mohl být v tom, že lék se objevil relativně nově v lidském evolučním vývoji, takže genetické mutace nebo polymorfismy s léky spojených genů působících fatální reakce na léky stále přetrvávají, protože ještě neuplynul dostatečně dlouhý čas na jejich eliminaci lidskou evolucí. Naproti tomu genetické mutace a polymorfismy genů nemocí, které mají fatální účinky, byly z větší části eliminovány dlouhým vývojem lidské evoluce [42].

2.3 Personalizované užívání léků založené na genetickém polymorfismu

Na rozdíl od pomalého pokroku v personalizované péči založené na studii genů nemocí, personalizované užívání léků je okamžitě účinná a rychle se rozvíjející oblast genomické medicíny. Individuálně upravené léky se nyní zavádají do klinické praxe.

Některé genetické polymorfismy se mohou zúčastnit funkcí vztahujících se k účinku léku, čímž způsobují individuální rozdíly v reakci na lék a dávají tak podnět ke klasifikaci „respondér“ a „ne-respondér“ (citlivý/necitlivý na lék). Odvětví farmakologie, které studuje vlivy genomické variace na reakci na léky u pacientů, se nyní nazývá farmakogenomika nebo farmakogenetika (PGx je

zkratka pro obě odvětví). Koncept farmakogenetiky je používán již více než 40 let pro označení studií jednotlivých genů a jejich vlivu na inter-individuální rozdíly v enzymech, které metabolizují léky. Na rozdíl od toho je farmakogenomika relativně nový pojem, který vznikl po projektu lidského genomu a který označuje studii nejen jednotlivých genů, ale i funkce a interakce všech genů v genomu v celkové variabilitě reakcí na lék [43]. Nicméně rozdíl mezi těmito dvěma termíny je spíše arbitrární a tyto dva termíny mohou být nyní vzájemně zaměňovány.

Jak jsme se již zmínili výše, PGx je jedna z nejúspěšnějších oblastí genomické medicíny. „Journal of the American Medical Association“ (JAMA-časopis Americké lékařské společnosti) v roce 1998 napsal, že asi dva miliony pacientů ročně utrpí následky spojené s užíváním léků a asi 100 000 z nich zemře, což představuje čtvrtou nebo pátou příčinu smrti ve Spojených Státech [44]. Správa pro jídlo a léky (FDA) v USA [45] podporuje používání před-diagnostických testů DNA pro testování reakcí na léky. Tyto testy jsou u některých léků povinné, kvůli tomu, aby se zamezilo nežádoucím vedlejším účinkům.

Individuální rozdíly v reakci na léky se přisuzují především dvěma třídám genetických polymorfismů. První je nazýván farmakodynamický (PD) polymorfismus a je způsoben polymorfismy genů s proteiny, které se vztahují k účinkům léků a spojených s receptory nebo proteinovou kinázou. Druhý z nich se nazývá farmakokineticický (PK) polymorfismus a vyskytuje se díky genům obsahujícím proteiny zodpovědné za absorpci, distribuci, metabolismus a vylučování (ADME) léku, například enzymy metabolizující nebo přepravující léky.

PD polymorfismus reakce léků je často nalezen v molekulárně cílených lécích různých moderních protirakovinných prostředků. Za zmínu stojí Gefitinib (Iressa), což je lék proti rakovině plic (epidermální růstový faktor receptoru tyrosine kinase inhibitor, EGFR-TKI), který je zvláště účinný proti pokročilému nemalobuněčnému karcinomu plic (NSCLC) [46].

Tento lék má naprostě vynikající účinky u „super-respondéra“, jehož stav se dramaticky zlepší po aplikaci Gefitinibu, dokonce i když se plícní karcinom rozšířil po celé plíci [47]. Tento lék je účinný zvláště proti adenokarcinomům u asijských žen v rozsahu 25-30 %, ale jeho účinnost jsou pouhá 3 % u mužů v USA a 17 % u žen v USA. Avšak tento lék může způsobit vážné vedlejší účinky u 2 % Japonců, u kterých způsobí intersticiální plícní chorobu, následkem které jedna třetina pacientů zemře. Takže abychom zabránili nebezpečnému riziku tohoto léku u non-respondéra, genetické testování je nezbytné. Rozdíl v reakci na tento lék se nyní připisuje specifické mutaci receptoru EGF a bylo oznámeno, že účinek tohoto léku se u pacientů s touto mutací zvyšuje desetkrát [48].

Jiný dobře známý příklad je Trastuzumab (Herceptin), což je lék proti rakovině prsu, založený na monoklonální protilátkce. Tento lék je účinný pro pacienty s rakovinou prsu, kteří mají zvýrazněný receptor HER2 (lidský epidermální růstový faktor typu 2) [49,50]. Úřad pro potraviny a léky USA, FDA, rozhodl, že před-diagnostické testování DNA za použití imuno-histochemie (IHC) a fluorescenční *in situ* hybridizace (FISH) je před aplikací tohoto léku povinný.

Polymorfismus PK řadí pacientovu individuální reakci na určité léky od „extenzivního metabolizéra“, jehož trávicí enzymy eliminují lék příliš rychle, čímž potlačují jeho účinek, až do „slabého metabolizéra“, který stravuje lék špatně, takže lék zůstává v jeho organismu příliš dlouho a produkuje a způsobuje excesivní účinek. Nejznámější polymorfismus trávícího enzymu, všeobecně spojovaným s účinky léků, je skupina cytochromů P450 (CYP). K dnešnímu datu je u lidí známo více než 50 izoenzymů CYP. Klasifikace tohoto enzymu je založena na stupni homologie genetických sekvencí. Účinky mnoha důležitých léků, jako jsou například protirakovinný prostředek 5-fluorouracil (5-FU), Irinocetan nebo prostředky proti tuberkulóze jako Isoniazid atd., závisí na polymorfismu CYP. Je k dispozici mnoho vynikajících pojednání o polymorfismu CYP a jeho účincích na metabolismus léků. Nebudeme se zde pouštět do obecného popisu, pouze krátce

představíme některé aspekty personalizované medikace. Za tímto účelem si zvolíme jako příklad Warfarin.

Warfarin je nejrozšířenější antikoagulační lék, předepisovaný pro potlačování syntézy srážlivých faktorů, čímž zabraňuje infarktu, mrtvicím a srážení krve. Je také známý svým úzkým terapeutickým rozsahem; nadměrná dávka způsobuje krvácení, zatímco nedostatečná dávka způsobuje trombózu. Další problém je v tom, že se úzký terapeutický rozsah mezi jednotlivci podstatně liší a variace dosahuje až desetinásobku, takže udržovací dávka tohoto léku musí být pečlivě monitorována častými krevními testy, aby byla zajištěna adekvátní, ale současně bezpečná, dávka. Warfarin potlačuje schopnost podsložky 1 komplexu epoxidu reduktázy vitamínu K (VKORC1) syntetizovat vitamín K nutný pro koagulaci krve, čímž zabraňuje srážení krve. Polymorfismy týkající se účinku Warfarinu jsou CYP2CD (*1, divoký typ a *3 nízkoaktivní typ) u PK polymorfismu a VKORC1 (-1639G>A, alela G, s vysokou transkripční aktivitou) u PD polymorfismu. Nejnovější studie ukazují, že pacientova první reakce je určena zvláště genotypem VKORC1 [51]. Jak jsme se již zmínili, pokyny FDA doporučily před-diagnostické testy DNA; některé jsou povinné a jiné jsou buď doporučené nebo referenční, zaměřené na citlivost na tento lék. Co se týká Warfarinu, je před-diagnostický test DNA doporučen.

Používání na míru ušitých léků se zvyšuje spolu s pokrokem ve farmakogenomice a farmakogenetice [52].

3. Výhody a možnosti medicíny založené na postgenomické omicsové vědě druhá generace omicsové medicíny

Ještě na počátku devadesátých let minulého století, když ještě projekt lidského genomu (HGP) teprve probíhal, staly se postgenomické omicsové vědy, jako například transkriptomika nebo proteomika, skutečností a to díky rychlému pokroku v technologických vysokorychlostních sekvencování. Ačkoliv je každá z postgenomických omicsových věd původně odvozená ze stejné genomové sekvence, má svou vlastní původní

informaci získanou různými postgenomickými (post-transkripčními nebo post-translačními) postupy, jako je například alternativní rezání, post-translační modifikace nebo epigenetická modifikace. Nyní se postgenomické omicsové vědy vyvíjejí samostatně a některé aplikace byly již schváleny FDA pro klinické použití.

Mezi genomikou a postgenomickými omicsovémi vědami je velký rozdíl ve vztahu ke klinické medicíně. Genomická medicína používá kongenitální polymorfismus genomové sekvence „germ-line“, který je identický ve všech tkáních a zůstává stejný po celý život. Naopak většina postgenomických omicsové věd je závislá na komprehensivní molekulární informaci „nemocných somatických buněk“, která se může měnit během časového průběhu nemoci a také je různá u různých tkání, takže se má za to, že omics vědy mnohem lépe reflekují současný stav nemoci a tím poskytují mnohem přesnější informaci než je informace genomická.

Takže se očekává, že kdybychom byli schopni co nejlépe použít omicsové data v klinické medicíně, mohli bychom inovovat konvenční medicínu a otevřít novou fázi lékařské péče, kterou bychom mohli nazývat (postgenomická) omicsová medicína. V této sekci popíšeme, jaké nové druhy klinických možností omicsové informace přinesou do lékařské péče a tím vylepší současnou konvenční genomickou medicínu.

3.1 Omicsové vědy poskytují informace o detailním mechanismu nemoci

3.1.1 Určování poddruhů a prognostika nemoci

Postgenomická omicsová informace poskytuje detailní údaje týkající se struktury molekulárního vývoje nemocných buněk. Omicsová data nemocných buněk poskytují klinicky a patologicky nepozorovatelné informace, které mohou být použity pro detailní klasifikaci nemoci. Nejlepší příklad klasifikace nemocí založené na genovém profilování (transkriptomika) jsou DNA mikroarrays.

Studie Goluba a kolektivu [14] je prvním pokusem toto uskutečnit; na příkladu lidské akutní leukémie předvedli možnost podklasifikace rakoviny, založené čistě na vzorku genového projevu bez reference k žádným předešlým zdravotnickým znalostem. Alizadeh a kol. [15] zkoumali uspořádání genové exprese rozptýleného lymfomu velkých buněk B (DLBCL) a našli dva poddruhy, které nebyly nalezeny konvenčním klinicko-patologickým zkoumáním a ukázali, že každý z nich vykazuje zcela jinou prognózu.

Struktura genové exprese neposkytuje pouze mikroskopickou informaci o molekulární aktivitě nemocných buněk, ale také ukazuje úzký vztah k makroskopickým projevům nemoci, zvláště k prognóze nemoci, což nemůže být správně předpovězeno čistě na bázi „mezoskopického pozorování“, jako jsou například klinické a patologické nálezy. Rosenwald a kol. [16] použili profilování genetické exprese DLBCL k identifikaci tří podskupin, které přidali ke dvěma již známým podskupinám [15] a na základě tohoto seskupení vytvořili molekulární prediktor přežití pacientů po chemoterapii. Tento prediktor byl nezávislý na dřívějším indexu klinicko-patologické prognózy. Také se objevilo mnoho studií zabývajících se předpověditelností nemoci na základě transkriptomiky [53]. Sorlie a kol. [54] zkombinovali několik nezávisle provedených populačních studií zabývajících se profilováním genové exprese rakoviny prsu za účelem vytvoření klasifikace podskupin pacientů podle jejich specifické prognózy. Konvenční patologické testy nemocných tkání a klinická pozorování nebyly schopny poskytnout přesnou prognózu. V posledních letech schválila FDA dva prognostické testy založené na genové exprese („MammPrint“ [55] a „OncotypeDX“ [56]), které jsou schopny předpovědět výhled pacienta po operaci rakoviny prsu a jsou nyní komerčně dostupné.

3.1.2. Reakce na protirakovinný lék

Profily genové exprese mohou být také použity pro predikci reakce léku proti rakovině. Cheng a kol. [57] použili biopatické vzorky z primárních prsních nádorů odebraných před zásahem za účelem předpovědi reakce na léčení

docetaxelem, založené na profilu genetické exprese vzorků. Z jejich práce vyplynulo, že molekulární profily mohou být použity k vývoji klinických testů na citlivost na docetaxel. Ayers a kol. [58] zkoumali vhodnost vývoje multigenového prediktoru patologické kompletní odezvy (pCR) na neoadjuvantní chemoterapii kombinující fluorouracil, doxorubicin, cyklofosfamid a paclitaxel při léčení rakoviny prsu. Na základě profilování genové exprese nádoru byla dosažena prediktivní přesnost 78 %. Rozdíly v reakcích na protirakovinný lék se přisuzují nikoli kongenitálnímu polymorfismu enzymu metabolizujícího lék, jak je to popsáno v minulé sekci o genomické medicíně, ale charakteristice vzorku nemoci, nebo přesněji výchozí napadené molekulární dráhy, která se projevuje v profilu genové exprese.

3.2 Diagnostická schopnost proteomiky

Jelikož omicsová informace se více vztahuje ke stavu nemocných buněk, může být také použita pro ranou detekci nemoci. Co se týká rakoviny, máme mnoho dobré známých nádorových markerů, ale ty jsou použitelné v relativně pozdním stádiu, kde pouze slouží k potvrzení již zjištěné rakoviny. Na rozdíl od běžného nádorového markeru, proteomika nemoci měřená pomocí hmotnosti spektrometrie, může být použita na ranou detekci rakoviny. Mnoho studií ukázalo, že nejenom specifický hmotnostní spektrální vrchol, ale i celý hmotnostní spektrální útvar obsahuje informaci umožňující detekci rakoviny ve velmi časném stadiu. Tato informace může být získána pomocí statistiky nebo zpracování dat. Bylo například zjištěno, že rakovina vaječníku může být objevena ve velmi raném stadiu s přesností na 99 % pomocí proteomického testu pacientova séra při použití SELDI-TOF-MS [18].

Provedli jsme studii na porovnání rozlišovací schopnosti mezi proteomickou informací a konvenčními nádorovými markery s ohledem na detekci hepatocelulárního karcinomu (HCC) [59]. Je dobře známo, že u pacientů trpících cirhózou jater dojde k vývoji HCC. Sebrali jsme vzorky séra od pacientů trpících cirhózou jater (LC:20 subjektů), ranou HCC (E-HCC: 20 subjektů) a pokročilou

HCC (A-HCC:20 subjektů). K proteomickému měření bylo použito SELDI-TOF-MS na analýzu pacientova séra a u nádorových markerů bylo měřeno AFP a PIVKA-II. Z grafu hmotnostní spektrální analýzy byly získány 3 statistické komponenty použitím metody nejmenšího čtverce.

Klasifikační výsledky (Tabulka 1) ukázaly, že pokročilá HCC mohla být diagnostikována u všech tří modelů. Na druhé straně konvenční nádorové markery dopadly špatně v rané detekci HCC v porovnání s vrcholnými intenzitami získanými metodou SELDI-TOF-MS. Navíc nejvyšší senzitivita (0,875) na detekci HCC z LC byla dosažena kombinací proteomických dat a nádorových markerů. To znamená, že proteomický test doplňuje dobré známé nádorové markery a mohou být použity společně pro vývoj přesnější metody rané detekce pro pacienty s cirhózou jater.

3.3 Omicsová medicína přináší personalizovanou a prediktivní medicínu

Na základě různých klinických schopností omics nemocí zmíněných výše můžeme popsat důležité prvky postgenomické omicsové medicíny výrazněji. Jak jsme se zmínili na začátku této sekce, postgenomické omics nemocí mohou poskytnout komplexní molekulární informaci o nemocných somatických buňkách. To znamená, že tato informace bude lépe vyjadřovat současný stav nemoci a také se bude měnit s jejím průběhem. Na rozdíl od toho, genomové sekvence „germ-line“ v genomické medicíně zůstávají stejné po celý život.

Předpokládá se, že se omicsová informace nachází na středním stupni hierarchie mezi genotypem genomu a klinickým fenotypu, takže vztah ke klinickému fenotypu je mnohem bližší než genomická informace (Obr.1). Tato „střednost“ omicsových dat (blízko ke klinickému fenotypu a také k probíhajícímu postupu nemoci), poskytuje důležité klinické možnosti.

Na rozdíl od genomické medicíny, která „pouze“ počítá s možností objevení se nemoci v budoucnosti, postgenomická omicsová medicína obsahuje přesnější informace týkající se nemoci přímo z nemocných buněk a je tak schopná předpovídат mnohem přesněji, „kdy se nemoc objeví“ a „jak se nemoc bude vyvíjet“.

Další prvek omicsových dat, kromě jejich „střednosti“, je jejich srozumitelnost či jejich komplexnost, protože omicsová data původně pocházejí z genové sekvence, také v podstatě zahrnují celý genom. Většina současných měřicích technologií, jako je mikroarray genové exprese nebo SNP čip, pokrývají celý genom, ačkoliv nějaký druh omicsů, jako například proteomika, v současné době poněkud zaostávají v nějakém aspektu (molekulová informace o velmi velkých proteinech), mohou být prakticky použity na získání celkové struktury nemoci. Srozumitelnost omicsových dat zaručuje platnost predikcí nemoci, které jsou provedeny na základě omicsových věd. Všeobecně se soudí, že genomická medicína přinese personalizovanou lékařkou péči, zatímco druhá generace omicsové medicíny přinese prediktivní medicínu. Omicsy nemoci poskytují detailní informace pro identifikaci podtypu nebo jemné molekulární charakteristiky nemoci, což umožňuje přesnější předpovídání nemocí nebo jejich ranou detekci. Samozřejmě, že když používáme termín „omics“, zahrnuje to také genomickou informaci, takže omicsová medicína není omezena na postgenomické omics a obsahuje výhody první generace. Bylo by lepší říci, že očekáváme, že omicsová medicína v širším slova smyslu (včetně genomické medicíny) přinese personalizovanou a prediktivní medicínu. Naše charakterizace omicsové medicíny má několik společných prvků s medicínou 4P (personalizace, prevence, prediktivní a participatorní medicína), což L. Hood definoval jako koncept systémové medicíny [26].

Současný biotechnologický pokrok také výrazně změnil situaci v oborech příbuzných omicsové medicíně, kde příští generace sekvencorů umožňuje, že genom může být sekvencován během hodin až dní [60]. Sekvenční rychlosť, která

Tab. 1. Klasifikace přesnosti odhadnuté za použití příčné validace metodou „jeden vyměnit“, a) pouze špičky, b) pouze markery a c) modely „špičky a markery“ pro diagnózu hepatocelulárního karcinomu.

a) Prediction model using peak intensities (sensitivity: 0.850, specificity: 0.550 for HCC diagnosis)			
Prediction	Diagnosis		
	LC	E-HCC	A-HCC
LC	11	3	3
E-HCC	8	12	10
A-HCC	1	5	7
b) Prediction model using tumor markers (sensitivity: 0.800, specificity: 0.800 for HCC diagnosis)			
Prediction	Diagnosis		
	LC	E-HCC	A-HCC
LC	16	8	0
E-HCC	4	8	2
A-HCC	0	4	18
c) Prediction model using both peak intensities and tumor markers (sensitivity: 0.875, specificity: 0.700 for HCC diagnosis)			
Prediction	Diagnosis		
	LC	E-HCC	A-HCC
LC	14	4	1
E-HCC	6	14	4
A-HCC	0	2	15

byla dva miliony bázových páru za den na konci projektu lidského genomu v roce 2003, nyní dosahuje přes bilion bps za den. Příští generace sekvencorů nebude fungovat jenom pro sekvencování DNA genomů, ale také může být použita pro jiný druh sekvencování, například mRNA nebo CHIP sekvencování [61]. Sekvencování mRNA seq může být použito pro přímou digitální kvantifikaci sekvencí místo použití konvenční hybridizační techniky, takže je možné se vyvarovat hybridizačních chyb a bude možné získat kvantitativní data vyšší kvality. Mohli bychom nazvat tuto digitální sekvenční molekulární informaci „digitální omics“.

Společně s biotechnologickým pokrokem jsou náklady na sekvencování rapidně redukovány, takže můžeme očekávat příchod genomu za 1 000 dolarů, což bude znamenat nástup doby „osobního genomu“.

3.4. Omicsová informatika druhé generace

Typická výpočetní metoda používaná v této generaci může být nazvána datovým přístupem. Velké množství omicsových dat je analyzováno, s nebo bez odkazu na externí kritéria (většinou klinicko-patologické charakteristiky a pacientovy výsledky). Toto je provedeno statisticky nebo zpracováním dat (strojové učení), metodami jako například modelem Coxovy regrese nebo podpůrným vektorovým strojem (SVM support vector machine). Takto je extrahována sada genů nebo proteomický vzorek, které mohou poskytnout klinicky použitelnou předpověď, ranou detekci nebo podklasifikaci nemoci.

Ale většinou jsou omicsová data nepříliš vhodná, protože množství proměnných (například počet genů v mikroarray DNA) daleko převyšuje počet vzorků (například počet případů).

Také například zcela odlišné sady genů byly navrženy pro předpověď stejněho druhu rakoviny různými výzkumnými skupinami. V této generaci je několik typických studií, které provedly úspěšnou analýzu omicsových dat, ale ve většině případů obrovské množství omicsových dat se svojí extrémně vysokou interkorelací, zabraňuje získání jednoznačných výsledků, pokud je použit pouze přímý analytický postup. Tato nevýhoda datového přístupu k omicsovým datům musí být zvážena, aby bylo všeobecně zřejmé, že jakákoli znalost o postupech nemoci musí být zahrnuta k doplnění nerohodnosti datové analýzy omicsových dat. Tento trend vedl v jistém smyslu ke konceptu systémové patologie třetí generace omicsové medicíny a snažil se použít modelový přístup pro analýzu omicsových dat.

Co se týká informatiky druhé generace, rádi bychom navrhli slibný přístup k datové analýze. Abychom co nejlépe využili omicsové informace, měla by být vyvinuta nová bioinformatika nebo translační informatika orientovaná na nemoc, aby zvládla multi-omicsová data, a která by mohla analyzovat nejen omicsová data, ale také vzájemné vztahy mezi různými modalitami omicsových dat. Jelikož různé omicsové modality mají své vlastní informace, použití multimodálních omicsových dat nám dá nové informace, které vykompenzují nejednoznačnost výsledků získaných jednoduchými modálními omics. [62].

4. Systémová patologie a systémová omicsová medicína třetí generace omicsové medicíny

4.1 Systémová patologie zahrnuje celý postup nemocí

Tak jak vznikala systémová biologie jako součást vědy o životě, stejná situace vznikla v medicíně založené na omicsových vědách, kde systémový přístup je absolutně nutný, abychom plně porozuměli celému procesu nemoci. Toto je situace, která potřebuje systémový přístup k nemocem. Přístup, který se soustředí na molekuly, kde se hledají příčiny nemoci pro specifické molekuly, jako například kauzativní geny nebo mutantní proteiny, a který byl úspěšný v monogenních studiích nemocí, nebude

úspěšný u polygenních multifaktoriálních nemocí, jako jsou obecné komplexní nemoci. Jak jsme se již zmínili, rychlý pokrok ve znalostech celulární a molekulární sítě a jejich změn při nemocech, jasně ukázal, že s výjimkou vzácných genetických nemocí, je většina nemocí v podstatě zaviněna a podporována stabilně porušenými molekulárními sítěmi signálních drah nebo genových regulačních sítí.

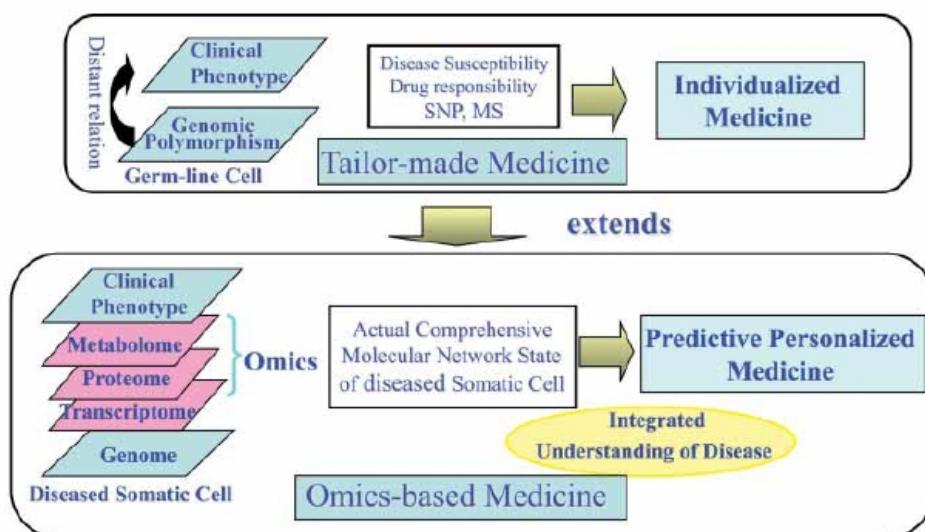
Ačkoliv tato distorze je vyvolána kooperativním účinkem aberací genů a proteinů, porušená dráha určuje stav nemoci více. Takže pochopení nemoci na systémové úrovni, založené na znalosti drah nemocných molekul, se nyní považuje za nezbytné, pakliže chceme plně porozumět informačnímu obsahu omicsů nemoci a následně celému procesu jako takovému (modelová analýza omicsových dat).

Navrhujeme nový koncept „systémové patologie“ k označení omicsové medicíny, jakožto obdobu systémové biologie v přírodních vědách. Prvek „celistvosti“ v omicsových datech umožňuje rozumět nemoci jako integrovanému celku, podtrhující omicsová data o nemoci. Při vývoji systémové patologie by bylo správné popsat, jak by se mělo nakládat s nemocemi jakožto s jednotnými systémy.

Popíšeme několik systémových prvků nemoci, které tomuto systémovému přístupu dají váhu. Následující prvky jsou většinou k vidění v obecných komplexních nemocech. Vzácné genetické nemoci v popisu těchto prvků vymezí.

Přeložené termíny k Obr. 1.:

- Distant relation* – vzdálený vztah
- Clinical Phenotype* – klinický fenotyp
- Genomic Polymorphism* – genomický polymorfismus
- Disease Susceptibility* – náchylnost k nemoci
- Drug responsibility* – odpovědnost za nemoc
- Individualized Medicine* – individualizovaná (osobní) medicína (na míru)
- Tailor-made Medicine* – lékařství na míru
- Extends* – pokračuje jako
- Clinical Phenotype* – klinický fenotyp
- Diseased Somatic Cell* – nemocná somatická buňka
- Actual Comprehensive Molecular Network State of Diseased Somatic Cell* – vlastní vyčerpávající stav molekulární sítě nemocné somatické buňky
- Predictive Personalized Medicine* – prediktivní personalizovaná medicína
- Integrated Understanding of Disease* – integrované porozumění nemoci
- Omics-based Medicine* – medicína založená na omicsových vědách



Obr. 1. Koncept omicsové medicíny: Postgenomické omics nemoci poskytuje komplexní molekulární informace o nemocných somatických buňkách. To znamená, že tato informace se mění během průběhu nemoci a liší se podle lokace nemoci, na rozdíl od genomových „germ-line“ (mikrobiálních) sekvencí v přizpůsobivé medicíně (na míru), kde zůstává stejná po celý život. Omicsová informace leží na středním stupni hierarchie, takže vztah ke klinickému fenotypu je mnohem bližší, což umožní nástup prediktivní medicíny.

4.2 Systémové prvky nemoci

4.2.1 Porušené molekulární dráhy jako primární příčiny nemoci

Díky pokroku ve studiích molekulárních sítí je všeobecně známo, že geny a proteiny spolupracují při formování molekulárních drah, které umožňují koherentní a komplexní biologické funkce. Takže bychom předpokládali, že s výjimkou vzácných genetických nemocí nejsou primární příčiny většiny nemocí dysfunkce genů, ale integrovaný důsledek pozměněných či porušených molekulárních drah, které vznikají vzájemným účinkem aberací několika genů a proteinů. Jinými slovy, nikoliv gen, ale dráha, je primární entita nemoci.

Když se příkladně podíváme na rakovinu, obecná rakovina vzniká změnami signálních drah nebo regulačních sítí genů, způsobená aberací několika (nikoli jednoho nebo dvou) onkogenů nebo genů potlačujících nádory, což následně vede k nekontrolovatelné proliferaci buněk. Všimněme si, že vznik nádoru je mnohakrokový proces, na kterém se podílí několik podstatných změn v genové dráze. Experimentální studie ukazují [63], že zavedení opičího viru „40 early region“ (SV40 ER), reverzní transkriptázy lidského telomeru (hTERT) a H-RAS do normální nakultivované buňky, definovalo sadu genetických změn, které jsou dostatečné k naprogramování tumorigenního fenotypu ve většině lidských kultivovaných buněk. SV40 ER zakóduje velký T antigen (LT) SV 40 a malý T antigen (ST), kde LT aktivuje retinoblastoma-protein (pRB) a dráhy supresoru nádoru p53 a ST potlačuje aktivitu proteinu fosfatázy 2A (PP2A), což vede k aktivaci MYC, což je transkripční faktor, který vyvolává proliferaci buněk.

Takže by se dalo říci, že existuje minimální soubor změn v molekulárních procesech, jako je imortalizace (hTERT), buněčný cyklus (pRB), apoptóza (p53), buněčný růst (MYC), angiogeneze, migrace a adheze (H-RAS), který je nutný pro vznik karcinomu. Také existuje mnoho variací ekvivalentní minimální genové sady, která způsobí nádor, ale změny v několika kritických dráhách, jako například těch, týkajících se růstu, mitózy a invaze, společně s potlačením apoptózy, se zdají být nutné pro aktivaci tumorigeneze.

Takže běžnou rakovinu bychom si mohli představit jako současný vznik aberace v těchto dráhách.

Toto se netýká pouze rakoviny, ale vztahuje se to i na jiné nemoci, zvláště běžné nemoci, kde se na vzniku nemoci podílejí stovky genů, jako je tomu například u hypertenze nebo u cukrovky. Je nepravděpodobné, že by každý z těchto genů sám byl schopen přímo ovlivnit spuštění nemoci, ale několik genů může spolupracovat na vzniku změn v jistých dráhách a společně způsobit aberaci těchto drah, což způsobí a podpoří nemoc.

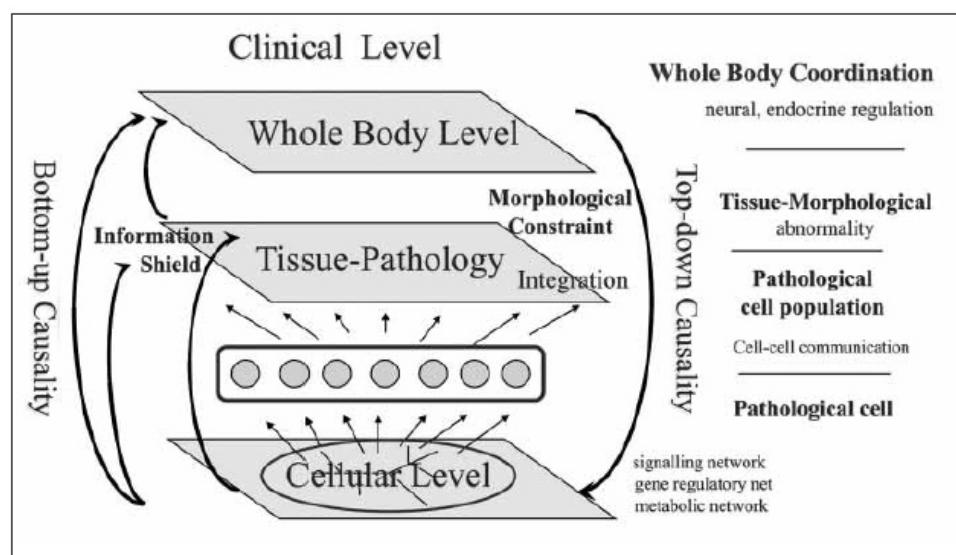
4.2.2 Hierarchická organizace a schopnost nemoci přetrávat

Aberace v molekulární síti nejsou omezeny na subcelulární síť, jako jsou například intracelulární signální dráhy nebo genová regulační síť. Komunikace mezi buňkami, neboli intercelulární síť, která je nezastupitelná pro koordinaci skupinového chování buněk nebo regulaci funkcí na tkáňové úrovni, je také „sídlo“ nemoci. Například intercelulární síť cytokinů, jako jsou interleukinky (IL) a TNF- α společně pracují v imunitním systému pro obranu hostitele, ale přehnaná exprese těchto genů způsobuje například chronická zánětlivá proliferaci onemocnění (CIPD) [64], jako je revmatická artritida nebo psoriáza.

Fyziologické řídící systémy větších rozdílů jako jsou oběhový, nervový nebo endokrinní systém, jsou také přirozeně spojené s nemocí pro formování klinického fenotypu na systemické úrovni.

Přeložené termíny k Obr. 2.:

Clinical Level	– klinická úroveň
Whole body level	– úroveň celého těla
tissue-Pathology	– patologie tkání
Cellular level	– buněčná úroveň
Bottom-up Causality	– kauzalita zdola nahoru
Top-down Causality	– kauzalita shora dolu
Information Shield	– informační bariera
Morphological Constraint	– morfologická omezení
Whole Body Coordination	– koordinace celého těla
Tissue-Morphological abnormality	– abnormalita morfologie tkání
Pathological cell population	– patologická populace buněk
Cell-cell communication	– mezibuněčná komunikace
Pathological cell patologická buňka	– signaling network signální síť
gene regulatory network	– síť genové regulace
metabolic network	– metabolická síť

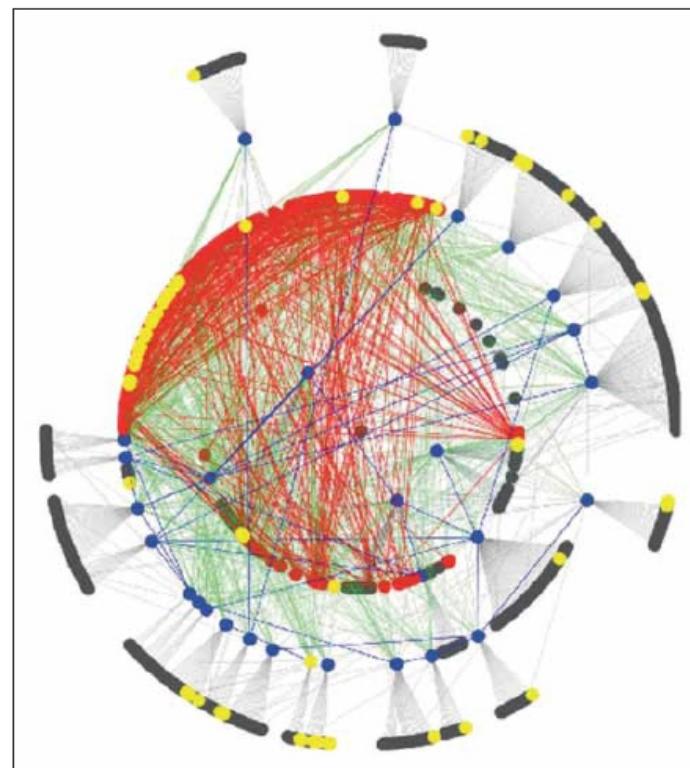


Obr. 2. Hierarchická organizace nemoci a schopnost přetrávat díky dvousměrné kauzalitě: nemoci jsou integrovaným multihierarchickým síťovým systémem, skládajícím se z podbuněčných molekulárních sítí, mezibuněčných komunikací, vazeb tkání/orgán a celkových tělesných koordinací. Nemoci jako běžná komplexní onemocnění mají soběstačný udržovací mechanismus díky smyčce dvousměrné kauzality.

Takže bychom mohli předpokládat, že nemoci jsou integrovaným multi-hierarchickým síťovým systémem, který se skládá ze subcelulární molekulární sítě, z mezibuněčné komunikace, vazeb tkání/orgán a systemické koordinace (Obr.2). Také jsou zde přítomny inter-hierarchické interakce. Kauzality shora dolů a zdola nahoru působí společně, takže nemoc představuje integrovaný celek. Kauzalita zdola nahoru by mohla být parafrázována jako „kumulativní a integrovaná kauzalita nasměrovaná do vyšších stupňů tkání a bio-objektů“. Naproti tomu kauzalita shora dolů by mohla být parafrázována jako nekumulativní a distributivní kauzalita pro tkáně, které jsou odvozeny od systémů nebo orgánů vyššího stupně. Také environmentální (na životním stylu závisející) faktory ovlivňují průběh nemoci, zvláště u běžných nemocí. Takže nemoci tvoří multihierarchickou poškozenou síť a dříve či později se projeví na úrovni systemických klinických fenotypů.

Běžné komplexní nemoci, jako jsou například hypertenze nebo cukrovka, jsou celoživotní nemoci, které trvají 20 až 30 let nebo i více. Aby stav nemoci mohl pokračovat po tak dlouhou dobu, musí existovat samoudržovací mechanismus, který má nějaký druh zpětné vazby.

Jako příklad samoudržovacího mechanismu v hierarchicky organizovaném systému nemoci si zvolíme hypertenzi, kde kromě kauzality zdola nahoru, která působí z molekulární úrovně na systemickou, podporuje celkový stav nemocného těla genovou expresi pro zachování systemického stavu hypertenze. Při hypertenzi, pokud krevní tlak stoupá díky jistým příčinám jako je životní styl nebo neurohumorální faktory, a když hypertenze pokračuje po několik měsíců, genetická aktivita endotelialních buněk kapilár začne proces přestavby kapilár, začne je zpevňovat, aby vydřely vysoký krevní tlak [65]. Toto by mohl být samoudržovací mechanismus hypertenze. Tento jev také ukazuje, že interhierarchická dvousměrná aktivita způsobuje schopnost udržovat celkový stav nemoci. Toto je jistý druh „převzetí kontroly“ vrozenou homeostatickou morfogenetickou funkcí, využitou pro udržování stavu nemoci.



Obr. 3. Vícebodová topologie proteinové interakce sítě (PPI): PPI sestává ze tří vrstev: kmen (uzly vysokého stupně), střední vrstva (vnitřní prstenec) a větve (uzly nízké úrovně). Uzly kmene a větví jsou znázorněny jako vnější prstenec a uzly středního stupně jsou znázorněny jako vnitřní prstenec. Cílové proteiny léků jsou znázorněny žlutými skvrnami, které jsou koncentrovány ve vrstvách středních a nízkých uzlin (přečteno z PLoS Počítačové biologie) [85].

4.2.3 Nemoc jako dynamický systém

Zatím jsme diskutovali o „organizačních aspektech“ systému nemoci jako hierarchické organizaci a její schopnosti samostatně pokračovat. Ale stejně důležitý aspekt systému nemoci je, že se vyvíjí a sám sebe mění v průběhu času na bázi přirozené „dynamiky“. To znamená, že nemoci jsou dynamické systémy. Jako příklad můžeme uvést infekční nemoci způsobené retrovirem jako například AIDS, kdy temporální průběh nemoci po jejím spuštění je považován za koevoluční proces mezi hostitelovým imunitním systémem a lidským imunodeficitním virem (HIV). HIV se vyvíjí uvnitř hostitele (pacienta) milionkrát rychleji než je tomu u evoluce normálních organismů a vyvíjí se pod selektivním tlakem neutralizace hostitelova imunitního systému a protivirových prostředků. Bylo vyvinuto mnoho metod k popsání virálního evolučního procesu uvnitř hostitele, jako například longitudinální filogenní strom [66,67], včetně naší metody založené na seskupování virových druhů HIV [68]. Tyto studie by mohly vést k charakterizaci

intrinzzické dynamiky společné retrovirálním infekčním chorobám.

Dynamika vývoje rakoviny by také mohla být modelována z pohledu Darwinické evoluce rakovinných buněk. Podle současné teorie o rakovinných kmenových buňkách vznikly mutace již ve stadiu kmenových buněk a rozmnožily se vývojovým procesem z kmenových buněk do buněk dospělých. Rakovinné kmenové buňky, které postupně získají mutace výhodné pro proliferaci, expandují své klony (klonální expanze). Bylo navrženo mnoho dynamických modelů, které popisují vývoj rakoviny na základě tohoto konceptu [69].

Není mnoho nemocí, jejichž dynamika je nám známa. Ale pokud bychom dosáhli lepšího porozumění dynamickým strukturám nemoci na systémové úrovni, mohli bychom specifikovat společné charakteristiky dynamiky pro určité třídy nemocí, což by mohlo být použito pro kvalitativní předpovědi průběhů nemocí.

Na závěr se podíváme na slibnou oblast, kde bude přístup dynamických systému efektivní, což je „epiteliálně-mezenchymální transformace“ (EMT), která hraje důležitou roli při invazi nebo metastáze rakoviny [70]. Jak jsme se již zmínili, celulární molekulární síť pro regulaci genové exprese (transkripční regulační síť), je z velké většiny stejná pro všechny tkáně. Ale, v závislosti na typu buňky, exprese skupiny genů, jejichž funkce jsou nutné pro ten který typ buněk, jsou posílené; zatímco exprese jiných skupin genů, které nejsou nutné, jsou potlačeny za účelem dosažení správné funkce tohoto typu buněk. Tyto kombinace genů regulovaných buď směrem nahoru nebo dolů se liší podle typů buněk, ale tyto kombinace genové exprese jsou vcelku normální a stabilní pro ten který buněčný typ.

Jelikož máme 210 různých buněčných typů, a když budeme považovat transkripční regulační síť za dynamický systém, mohli bychom říci, že tato síť má 210 „stabilních řešení“, navzájem si odpovídajících. Každý profil exprese korespondující s typem buňky tvoří jedno stabilní řešení transkripční regulační sítě.

EMT je buněčný proces, kde si buňky v epitelu osvojují schopnost napadat bazální extracelulární matrix nebo sousední mezenchymální buňky. Během EMT si epiteliální buňky přisvojují mezenchymální vzhled včetně zvýšené hybnosti a útočnosti. Toto je normální vývojový proces a je používán při formaci třírozměrných struktur v embryu. Avšak u rakoviny se jedná o klíčový krok k invazi a metastáze. EMT je charakterizován kombinací ztráty pojivových proteinů epiteliálních buněk, jako například E-cadherin a integrin, a zvýšením počtu mezenchymálních markerů jako je například vimentin a fibronectin, které souvisejí s hybností buněk.

Z našeho pohledu systémové patologie může být EMT považován za „fázový posun“ (koherentní strukturální změnu) transkripční regulační sítě, což je přechod mezi různými stabilními sítovými řešeními korespondující s epiteliálním a mezenchymálním druhem buněk. Společně s touto fázovou tranzicí transkripční regulační síť nastává změna v genetické

aktivitě několika reprezentativních genů, jako jsou například E-cadherin (který těsně spojuje sousední buňky), integrin (který zajišťuje spojení k bazální membráně) nebo vimentin (vztahuje se k buněčným filamentům) a fibronectin (složka extracelulární matrix). EMT umožňuje rakovinným buňkám epiteliálního původu invazi mezenchymálních tkání pod bazální membránou, jako jsou například spojovací tkáně, svalové tkáně a krevní cévy tak, že předstírá vzhled mezenchymální buňky. Na základě konceptu, že EMT je „fázová tranzice“ molekulární sítě, můžeme získat představu, do jaké míry je aberace drah nutná a kritická, aby vyvolala tranzici fázi EMT pro invaze a metastáze.

Když vezmeme úvahu výše zmíněné systémové prvky, můžeme pochopit nemoc v jejím celém spektru. Nemoc narušuje biosystémy a normální dráhy jsou modifikovány tak, že vytvoří „udržitelné dráhy pro nemoc“.

4.3 Informatika systémové patologie plně využívá omics nemoci

Systémová patologie, konkrétně porozumění nemocem na úrovni systému, je efektivní rámcem pro získání komplexních omicsových dat. Na základě realizace systémové charakteristiky nemoci se nyní vyvíjejí různé důležité klinické aplikace, například systémová analýza, modelování a simulace nemoci.

S aplikací systémové patologie je nejčastěji studováno odhadování molekulární dráhy. Nejdříve byly vyuvinuty metody pro identifikaci normálních celulárních funkcí, aby mohli odhadnout molekulární dráhy z profilu genové exprese. Jedna z prvních klíčových studií prosazující tento přístup se snažila objevit genové regulační sítě v *Saccharomyces cerevisiae* z profilu genové exprese použitím Bayesiánské identifikační metody sítí [72]. Od té doby bylo vynalezeno mnoho druhů různých metod, z nichž se dnes používají hlavně následující dvě.

První je koexpressní model, což je neřízený pravděpodobnostní model, založený na korelačním koeficientu, parciálním korelačním koeficientu (Gaussův grafický model) [73] nebo vzájemné informaci (Relevantní síť) [74], ARANCNE [75], mezi

libovolnými páry genů. Druhý model je Bayesianský sítový model [76], což je řízený grafický model pro reprezentaci pravděpodobnosti kauzality mezi aktivitami genů. Ačkoliv některé tyto čistě datové přístupy dosahly úspěchu při objevení drah nemocí [77], předpokládá se [78], že čistě datové přístupy k mikroarray datů mají mnoho omezení; potřebují velký vzorek a zachytí pouze části biologicky relevantních sítí.

Kombinace transkriptomických dat a biologických znalostí průběhu se považuje za vhodnejší pro odhad průběhu nemoci. Například profil genové exprese nemocných buněk, které jsou superponovány na meziproteinovou interakční síť za účelem identifikace podsítě způsobující nemoc, se nazývá CGI (kombinace exprese genů a dat proteinové interakce). Chuang a kol. aplikovali tento přístup na profily genové exprese metastatických a nemetastatických pacientů s rakovinou prsu za účelem identifikace podsítí, jejichž úroveň exprese koreluje s prognózou metastázy [79].

Také jsme superponovali profil genové exprese na BIND (Biomolekulární interakční síť, <http://bond.unleashedinformatics.com/index.jsp>),

kde je meziproteinová interakce kombinovaná s interakcí genu (DNA) proteinu (transkripční faktor) za účelem identifikace kauzativního procesu pro opětovný výskyt hepatocelulárního karcinomu po operaci [80].

Další slibný přístup je eQTL [81], který považuje profil genové exprese za jeden z kvantitativních příznaků (fenotypů) a aplikuje statistickou genetiku pro výzkum aberací genetického lokusu. Přístup eQTL může sklobit profil genové exprese s omezeným počtem kauzativních genů. Kromě reverzního inženýrství, CGI a eQTL, bylo označeno mnoho metod pro systémovou patologii nemocí. Tyto v podstatě podobné koncepty jako „biologie rakovinných systémů“ [82] nebo „systémová genetika nemocí“ [83] předpokládají, že pouze integrovaný přístup k nemocem jako je systémová patologie může plně využít omicsů nemocí.

Topologický síťový přístup je slibnou oblastí systémové patologie. Jeong [84] zjistil, že stupeň (počet spojení uzliny) v síti meziproteinových interakcí se řídí „neomezenou“ (zákon sily) distribucí, což znamená strukturu kmen-větve. Mnoho topologických síťových studií bylo vykonáno například na síti nemocí nebo genové síti nemoci [42]. My jsme použili metodu pohyblivé stratifikace pro získání podsítí ve specifickém rozsahu za účelem nalezení hustě spojené střední vrstvy („páteř“) meziproteinové interakční sítě (PPI), kromě dobré známé struktury kmen větve [85]. Na této třívrstvové struktuře PPI jsme zmapovali cílové proteiny léků schválených FDA. Bylo zjištěno, že tyto proteiny se nalézají především ve střední vrstvě síti PPI (Obr. 3). Tento výsledek znamená, že kmenové uzliny (proteiny vysokého stupně) jsou nevhodné jako cíl léků, protože mají široké možnosti vedlejších účinků.

Avšak střední vrstva proteinů je vhodná jako cíl léků s méně vážnými vedlejšími účinky a vyšší účinností jako páteřní proteiny. Topologická analýza nevyžaduje numerické informace celulární sítě jako reakční koeficient, takže je vhodnější udělat si závěry o celulární síti, u které je těžké získat kvantitativní informace.

Co je v této oblasti stejně důležité a ještě nebylo zkoumáno, je kvantitativní analýza kvalitativních systémů a simulace [86] průběhu nemoci. Některé kvalitativní metody byly kdysi zkoumány v rámci běžného výzkumu.

Jak jsme se zmínili výše, kvantitativní informace se těžce získávají v celulární síti, zatímco máme spoustu kvalitativních strukturálních znalostí o signální dráze a regulační síti. Nová metodologie by měla strukturálně a kvalitativně analyzovat a předpovídat chování průběhu nemoci.

5. Diskuze a závěr

V tomto článku jsme popsali důležité prvky a vývoj omicsové medicíny a informatiky, která se k ní vztahuje. Při popisu vývoje omicsové medicíny jsme použili „trígenerační model“. V první generaci jsme začali genomickou medicínou a potom jsem po kráčovali k „(postgenomické) omicsové medicíně“, jakožto k druhé generaci, a nakonec jsme se dostali k medicíně založené na omicsových systémech, jakožto třetí generaci, která je stále v počátcích a čeká na budoucí vývoj.

Třígenerační model se zdá být vhodným konceptem pro popis vlastního vývoje omicsové medicíny. Na základě tohoto modelu je typická informační metoda použitá v každé generaci jasně stanovena; klasická statistika (statistická genetika) první generace byla nahrazena datovým přístupem jako je strojové učení, vytěžování dat nebo exploratorní statistika, která byla později nahrazena modelovým (hypotézovým) přístupem.

Další výhodou třígeneračního modelu je, že poskytuje jednotnou platformu, která může zahrnovat všechny tři koncepty molekulární medicíny:

1) „genomická medicína“ snažící se o uskutečnění personalizované medicíny na základě vrozené diference v genomu, 2) (postgenomická) omicsová medicína snažící se o realizaci prediktivní medicíny na základě profilů genové exprese a hmotnostního spektra proteinu, které se liší podle stádia nemoci, 3) systémová medicína usilující o celkové porozumění nemoci na základě znalosti celulárních drah a jejich změn při nemocech (Tab.2).

I když jsme spíše přecenili význam nově zavedených prvků u každé generace, měli bychom poznámenat, že u vlastní lékařské činnosti obsahuje další generace prvky generace předešlé. Například druhá generace omicsové medicíny zahrnuje personalizovanou péči genomické medicíny. Takže lékařské systémy třetí generace představují nejúplnejší medicínu včetně personalizované a prediktivní péče.

Nemoc je obecně chápána jako komplexní systém v tom smyslu, že to je hierarchicky organizovaný, soběstačný, dynamický systém, který se vyvíjí podle svých vlastních zákonů. Uvedli jsme různé inherentní prvky nemoci jakožto dynamického komplexního systému, jako jsou například soběstačnost, hypertenze, dvousměrná kauzalita, evoluční dynamika HIV, expanze rakovinných klonů, rakovinná metastáza (EMT) jakožto fázová tranzice systémů, možnost kvalitativního kauzálního zdůvodnění dynamiky nemoci a topologická lokace cílů léků v proteinové interakční síti.

Tab. 2. Prvky třígenerační omicsové medicíny.

Generace	Jméno	Hlavní informace	Hlavní cíl	Přístup informatiky	Metody
1	Genomická medicína	Mutace a polymorfismy genomu buněk germ-line	Personalizovaná (na míru) medicína	Statistický (geneticky)	Statistická genetika, farmakogenetika, farmakogenomika
2	(Postgenomická medicína) omicsová medicína	Celkové molekulární profily somatických buněk	Prediktivní a preventivní medicína	Na základě dat	Vytěžování dat, strojní učení, exploratorní statistika
3	Omicsová systémová medicína	Molekulární dráhy a sítě a jejich změny následkem nemocí	Komplexní medicína (personalizovaná, prediktivní, preventivní)	Na základě modelu	Identifikace systému, reverzní inženýrství, simulace systému, design systému

Mohli bychom očekávat, že se ve světle systémového přístupu objeví docela rozdílný pohled na nemoc. Avšak systémový přístup k nemoci je teoreticky v počátcích a v budoucnosti je potřeba udělat mnoho práce.

Naopak v praktických aplikacích systémového konceptu nemoci bylo předloženo mnoho metod informatiky, včetně použití reverzního inženýrství k odhadu celulární sítě nemoci z omicsových dat.

Na závěr uvádíme, že pochopení nemoci na základě procesů, což nazýváme systémovou patologií, je nejslibnější rámcem pro plné využití informačního obsahu omicsů nemoci a komprehensivní poznání procesu nemoci. Spolu s předešlými generacemi omicsových věd můžeme očekávat, že omicsové systémy podstatně změní současnou lékařskou péči a učiní ji více personalizovanou a prediktivní.

Poděkování

Tato práce vznikla za podpory Speciálního koordinačního fondu pro podporu vědy a technologie (Programy „Podpora talentu v nových oborech“ a „Program pro vývoj a výzkum kritických problémů“) Ministerstva vzdělání, kultury, sportu, vědy a technologie Japonska. Autor by rád poděkoval Dr. Kaoru Mogushi za pomoc při přípravě tohoto článku a také oponentům za jejich konstruktivní komentáře, které vylepšily tento článek.

Článek přeložila společnost Bio-Consult laboratories, spol. s.r.o., Centrum biomedicínské informatiky, s podporou projektu 1M06014 MŠMT ČR.

Literatura

- [1] Weinstein JN. Fishing expeditions. *Science* 1998; 282: 628-629.
- [2] Weinstein JN. 'Omic' and hypothesis-driven researchin the molecular pharmacology of cancer. *Curr Opin Pharmacol* 2002; 2: 361-365.
- [3] Knaup P, Ammenwerth E, Brandner R, Brigl B, Fischer G, Garde S, Lang E, Pilgram R, Ruderich F, Singer R, Wolff AC, Haux R, Kulikowski C. Towards clinical bioinformatics: advancing genomic medicine with informatics methods and tools. *Methods Inf Med* 2004; 43: 302-307.
- [4] Gaughan A. Bridging the divide: the need for translational informatics.
- [5] Guttmacher AE, Collins FS. Genomic medicine - a primer. *N Engl J Med* 2002; 347: 1512-1520.
- [6] Kumar D. Genomic medicine: a new frontier of medicine in the twenty first century. *Genomic Med* 2007; 1: 3-7.
- [7] Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; 270: 467-470.
- [8] Fodor SP, Rava RP, Huang XC, Pease AC, Holmes CP, Adams CL. Multiplexed biochemical assays with biological chips. *Nature* 1993; 364: 555-556.
- [9] Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA. Genome-wide location and function of DNA bindingproteins. *Science* 2000; 290: 2306-2309.
- [10] Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett-Engele P, McDonagh PD, Loerch PM, Leonardson A, Lum PY, Cavet G, Wu LF, Altschuler SJ, Edwards S, King J, Tsang JS, Schimmack G, Schelter JM, Koch J, Ziman M, Marton MJ, Li B, Cundiff P, Ward T, Castle J, Krolewski M, Meyer MR, Mao M, Burchard J, Kidd MJ, Dai H, Phillips JW, Linsley PS, Stoughton R, Scherer S, Boguski MS. Experimental annotation of the human genome using microarray technology. *Nature* 2001; 409: 922-927.
- [11] Xing Y, Kapur K, Wong WH. Probe selection and expression index computation of Affymetrix Exon Arrays. *PLoS ONE* 2006; 1: e88.
- [12] Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T. Protein and polymer analyses up to m/z 100000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 1988; 2: 151-153.
- [13] Hutchens TW, Yip TT. New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun Mass Spectrom* 1993; 7:576-580.
- [14] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; 286: 531-537.
- [15] Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; 403: 503511.
- [16] Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, López-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajcí P, Stokke T, Staudt LM. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002; 346: 1937-1947.
- [17] Tanaka H. Bioinformatics and genomics for opening new perspective for personalized care. *Stud Health Technol Inform* 2008; 134: 47-58.
- [18] Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002; 359: 572-577.
- [19] Meani F, Pecorelli S, Liotta L, Petricoin EF. Clinical application of proteomics in ovarian cancer prevention and treatment. *Mol Diagn Ther* 2009; 13:297-311.
- [20] Grubb RL, Deng J, Pinto PA, Mohler JL, Chinnaiyan A, Rubin M, Linehan WM, Liotta LA, Petricoin EF, Wulfkuhle JD. Pathway biomarker profiling of localized and metastatic human prostate cancer reveal metastatic and prognostic signatures. *J Proteome Res* 2009; 8: 3044-3054.
- [21] Tanaka H. System-theoretic approach to life, genome and disease. Proc of 3rd annual conference of Chemo-Bioinformatics 2002. p 3.
- [22] Tanaka H, Nakaya J. Research and education for biomedical informatics at Tokyo Medical and Dental University. *Yearb Med Inform* 2007. pp 157-162.
- [23] Kitano H. Systems biology: a brief overview. *Science* 2002; 295: 1662-1664.
- [24] Kitano H. Computational systems biology. *Nature* 2002; 420: 206-210.
- [25] Hood L, Heath JR, Phelps ME, Lin B. Systems biology and new technologies enable predictive and preventative medicine. *Science* 2004; 306: 640-643.
- [26] Auffray C, Chen Z, Hood L. Systems medicine: the future of medical genomics and healthcare. *Genome Med* 2009; 1: 2.
- [27] International human genome sequencing consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004; 431: 931-945.
- [28] Strachan T, Andrew PR. Human molecular genetics (3rd edition). New York: Garland Publishing; 2003.

- [29] Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 1983; 306: 234-238.
- [30] Balding DJ, Bishop M, Cannings C. *Handbook of Statistical Genetics* (3rd edition). Hoboken: Wiley; 2007.
- [31] Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993; 72:971-983.
- [32] Hoffman EP, Brown RH, Kunkel LM. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* 1987; 51: 919-928.
- [33] Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* 1989; 245: 1059-1065.
- [34] Schaid DJ, Sommer SS. Genotype relative risks: methods for design and analysis of candidate-gene association studies. *Am J Hum Genet* 1993; 53: 1114-1126.
- [35] Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. Linkage disequilibrium in the human genome. *Nature* 2001; 411: 199-204.
- [36] Hardy J, Singleton A. Genomewide association studies and human disease. *N Engl J Med* 2009; 360: 1759-1768.
- [37] Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet* 2007; 8: 657-662.
- [38] Diabetes Genetics Initiative. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; 316: 1331-1336.
- [39] The Wellcome Trust case control consortium. Genomewide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; 447: 661-678.
- [40] Teo YY, Small KS, Clark TG, Kwiatkowski DP. Perturbation analysis: a simple method for filtering SNPs with erroneous genotyping in genome-wide association studies. *Ann Hum Genet* 2008; 72: 368-374.
- [41] International Hapmap consortium. The International HapMap Project. *Nature* 2003; 426: 789-796.
- [42] Goh KL, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. *Proc Natl Acad Sci U.S.A.* 2007; 104: 8685-8690.
- [43] Evans WE, Relling MV. Moving towards individualized medicine with pharmacogenomics. *Nature* 2004; 429: 464-468.
- [44] Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* 1998; 279: 1200-1205.
- [45] FDA. Guidance for Industry Pharmacogenomic Data Submissions, 2003. <http://www.fda.gov/cder/guidance/5900dft.pdf>
- [46] Blackledge G, Averbuch S. Gefitinib ('Iressa', ZD1839) and new epidermal growth factor receptor inhibitors. *Br J Cancer* 2004; 90: 566-572.
- [47] Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004; 350: 2129-2139.
- [48] Asahina H, Yamazaki K, Kinoshita I, Sukoh N, Harada M, Yokouchi H, Ishida T, Ogura S, Kojima T, Okamoto Y, Fujita Y, Dosaka-Akita H, Isobe H, Nishimura M. A phase II trial of gefitinib as firstline therapy for advanced non-small cell lung cancer with epidermal growth factor receptor mutations. *Br J Cancer* 2006; 95: 998-1004.
- [49] Baselga J. Clinical trials of Herceptin (trastuzumab). *Eur J Cancer* 2001; 37 (Suppl 1): S18-24.
- [50] Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J, Norton L. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001; 344: 783-792.
- [51] Schwarz UI, Ritchie MD, Bradford Y, Li C, Dudek SM, Frye-Anderson A, Kim RB, Roden DM, Stein CM. Genetic determinants of response to warfarin during initial anticoagulation. *N Engl J Med* 2008; 358: 999-1008.
- [52] Aneesh TP, Sonal SM, Jose A, Chandrana L, Zachariaha SM. Pharmacogenomics: The Right Drug to the Right Person. *J Clin Med Res* 2009; 1: 191194.
- [53] van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415: 530-536.
- [54] Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Børresen-Dale AL, Botstein D. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U.S.A.* 2003; 100: 8418-8423.
- [55] Slodkowska EA, Ross JS. MammaPrint 70-gene signature: another milestone in personalized medical care for breast cancer patients. *Expert Rev Mol Diagn* 2009; 9: 417-422.
- [56] Asad J, Jacobson AF, Estabrook A, Smith SR, Boolbol SK, Feldman SM, Osborne MP, Boachie-Adjei K, Twardzik W, Tartter PI. Does oncotype DX recurrence score affect the management of patients with early-stage breast cancer? *Am J Surg* 2008; 196: 527-529.
- [57] Cheng JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R, Mohsin S, Osborne CK, Chamness GC, Allred DC, O'Connell P. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* 2003; 362: 362-369.
- [58] Ayers M, Symmans WF, Stec J, Damokosh AI, Clark E, Hess K, Lecocke M, Metivier J, Booser D, Ibrahim N, Valero V, Royce M, Arun B, Whitman G, Ross J, Sneige N, Hortobagyi GN, Pusztai L. Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J Clin Oncol* 2004; 22: 2284-2293.
- [59] Mogushi K, Nomura F, Tomonaga T, Sunaga M, Sogawa K, Tanaka H. Evaluation of feature selection and disease classification method for protein expression profile analysis using SELDI-TOF-MS. *Proc. 2nd Annual Meeting of Japan Society for Clinical Proteomics* 2006. p 20.
- [60] Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008; 26: 1135-1145.
- [61] Trapnell C, Salzberg SL. How to map billions of short reads onto genomes. *Nat Biotechnol* 2009; 27: 455-457.
- [62] Kleemann R, Verschuren L, van Erk MJ, Nikolsky Y, Cnubben NH, Verheij ER, Smilde AK, Hendriks HF, Zadelaar S, Smith GJ, Kaznacheev V, Nikolskaya T, Melnikov A, Hurt-Camejo E, van der Greef J, van Ommen B, Kooistra T. Atherosclerosis and liver inflammation induced by increased dietary cholesterol intake: a combined transcriptomics and metabolomics analysis. *Genome Biol.* 2007; 8: R200.
- [63] Zhao JJ, Roberts TM, Hahn WC. Functional genetics and experimental models of human cancer. *Trends Mol Med* 2004; 10: 344-350.
- [64] Hirano T. Cytokines in autoimmune disease and chronic inflammatory proliferative disease. *Cytokine Growth Factor Rev* 2002; 13: 297-298.
- [65] Schiffrin EL. Remodeling of resistance arteries in essential hypertension and effects of antihypertensive treatment. *Am J Hypertens* 2004; 17: 1192-2000.

- [66] Drummond A, Rodrigo AG. Reconstructing genealogies of serial samples under the assumption of a molecular clock using serial-sample UPGMA. *Mol Biol Evol* 2000; 17: 1807-1815.
- [67] Ren F, Ogishima S, Tanaka H. Longitudinal phylogenetic tree of within-host viral evolution from noncontemporaneous samples: a distance-based sequential-linking method. *Gene* 2003; 317: 89-95.
- [68] Hasegawa N, Sugiura W, Shibata J, Matsuda M, Ren F, Tanaka H. Inferring within-patient HIV-1 evolutionary dynamics under anti-HIV therapy using serial virus samples with vSPA. *BMC Bioinformatics* 2009; 10: 360.
- [69] Michor F, Iwasa Y, Nowak MA. Dynamics of cancer progression. *Nat Rev Cancer* 2004; 4: 197-205.
- [70] Lee JM, Dedhar S, Kalluri R, Thompson EW. The epithelial-mesenchymal transition: new insights in signaling, development, and disease. *J Cell Biol* 2006; 172: 973-981.
- [71] Liang S, Fuhrman S, Somogyi R. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac Symp Biocomput* 1998; 18-29.
- [72] Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. *J Comput Biol* 2000; 7: 601-620.
- [73] Edwards D. Introduction to Graphical Modelling. Heidelberg: Springer; 2000.
- [74] Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac Symp Biocomput* 2000; pp 418-429.
- [75] Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 2006; 7 (Suppl1): S7.
- [76] Friedman N, Koller D. Being Bayesian about network structure. *Math Learn* 2003; 50: 95-126.
- [77] Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nat Genet* 2005; 37: 382-390.
- [78] Markowetz F, Spang R. Inferring cellular networks a review. *BMC Bioinformatics* 2007; 8 (Suppl 6): S5.
- [79] Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. *Mol Syst Biol* 2007; 3: 140.
- [80] Tanaka S, Mogushi K, Yasen M, Noguchi N, Kudo A, Kurokawa T, Nakamura N, Inazawa J, Tanaka H, Arii S. Surgical contribution to recurrence-free survival in patients with macrovascular-invasionnegative hepatocellular carcinoma. *J Am Coll Surg* 2009; 208: 368-374.
- [81] Göring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, Jowett JB, Abraham LJ, Rainwater DL, Comuzzie AG, Mahaney MC, Almasy L, MacCluer JW, Kisseebah AH, Collier GR, Moses EK, Blangero J. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet* 2007; 39: 1208-1216.
- [82] Wang E, Lenferink A, O'Connor-McCourt M. Cancer systems biology: exploring cancer-associated genes on cellular networks. *Cell Mol Life Sci* 2007; 64: 1752-1762.
- [83] Sieberts SK, Schadt EE. Moving toward a system genetics view of disease. *Mamm Genome* 2007; 18: 389-401.
- [84] Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL. The large-scale organization of metabolic networks. *Nature* 2000; 407: 651-654.
- [85] Hase T, Tanaka H, Suzuki Y, Nakagawa S, Kitano H. Structure of protein interaction networks and their implications on drug design. *PLoS Comput Biol* 2009; 5: e1000550.
- [86] Kuipers B. Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge (Artificial Intelligence). Cambridge: The MIT Press; 1994. p 13.



EJBI 2009

ISSN 1801 - 5603

European Journal for Biomedical Informatics

Volume 5 (2009), Issue 1

German version

www.ejbi.eu



Content

German version

- de 2 - 9 eHealth-Ansatz für einen Paradigmenwechsel im Gesundheitswesen Architekturelle Herausforderungen
B. Blobel
- de 10 - 16 Semantische Interoperabilität erfordert die richtigen Modelle und Code-Systeme:
Eine Prüfung verschiedener Ansätze für Score Systeme
F. Oemig, B. Blobel
- de 17 - 22 Biomedizinische Ontologien für die Praxis
M. Brochhausen, B. Smith

eHealth-Ansatz für einen Paradigmenwechsel im Gesundheitswesen

Architekturelle Herausforderungen

B. Blobel¹

¹eHealth Competence Center, University of Regensburg Medical Center, Germany

Zur Verbesserung von Qualität und Effizienz der Gesundheitsversorgung unter den bekannten Zwängen muss das Gesundheitswesen von einem organisationszentrierten über ein prozessgesteuertes zu einem personalisierten Versorgungsparadigma (Personal Health pHealth) gewandelt werden. Solche personalisierten eHealth-Dienste unabhängig von Zeit und Ort erfordern die Erfüllung bestimmter technischer Paradigmen, Multidisziplinarität unter Einschluss von Medizin, Informatik, Biomedizintechnik und Bioinformatik sowie der „omics“-Disziplinen, aber auch die Gewährleistung aller Interoperabilitätsniveaus bis hin zur Service-Interoperabilität auf der Basis eines Architektur-Frameworks. Der Beitrag führt die zugrunde liegenden Paradigmen, Anforderungen, Architektur-Referenzmodelle, Modellierungs - und Formalisierungsprinzipien sowie Entwicklungsprozesse für umfassende, service-orientierte, personalisierte eHealth-Interoperabilitätsketten ein. Ein spezielles Augenmerk ist auf Ontologien und die Wissensrepräsentation im Kontext von eHealth und pHealth gelegt worden.

Stichwörter: Personal Health, Systemarchitektur, semantische Interoperabilität, Ontologie, Wissensrepräsentation, Universelle Logik, Ubiquitäre Versorgung.

Einleitung

Nicht nur in industrialisierten Ländern, sondern zunehmend auch in Ländern im Übergang unterliegen die Gesundheitssysteme einem grundlegenden Paradigmenwechsel, indem sie sich von der organisationszentrierten über die prozessgesteuerte oder verteilte hin zur personalisierten Versorgung wandeln. Um die Bedürfnisse, Erwartungen, Intentionen und Wünsche des Patienten - oder gar des Bürgers, bevor dieser zum Patienten wird in den Mittelpunkt des Gesundheitswesens zu rücken, sind eine Reihe

organisatorischer und technischer Voraussetzungen zu schaffen. Im Wesentlichen als Einführungs-Keynote der CeHR 2007 Konferenz „eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge“ präsentiert und veröffentlicht [1], führt der Beitrag die domänenspezifischen organisatorischen und technischen Paradigmen, Anforderungen und Lösungen für eine personalisierte, ubiquitäre Versorgung ein. Er betont die formalen Aspekte der Modellierung und Implementierung von eHealth- und pHealth-Interoperabilität und fokussiert auf die erforderliche Multidisziplinarität.

1. Material und Methoden

Im Folgenden werden das fortgeschrittene eHealth-Paradigma und seine Spezialisierung zu pHealth, die durch die aktive Einbeziehung des Patienten charakterisiert sind, die technischen Anforderungen für über Distanz realisierte personalisierte Gesundheitsdienste sowie eine generisches Referenzarchitektur-Modell kurz eingeführt. Des Weiteren werden Semiotik und Wissensrepräsentationsgesichtspunkte im eHealth-Kontext diskutiert. Als eine Schlussfolgerung dieses Abschnitts wird eine einheitliche Methodologie zur Erreichung der eHealth/pHealth-Geschäftsziele vorgeschlagen und teilweise im Ergebnisabschnitt exemplifiziert.

1.1 Das fortgeschrittene eHealth-Paradigma

In der Literatur kann man viele verschiedene Definitionen für eHealth finden, so z.B. die in [2]. Aus der architekturellen und formalen Perspektive, mit der sich dieser Beitrag befasst, umfasst eHealth Gesundheitsdienste, die dem Individuum zu jeder beliebigen Zeit unabhängig vom Aufenthaltsort der Akteure, Dienste und Ressourcen, die in

den Versorgungsprozess involviert sind, bereitgestellt werden. eHealth ist mit anderen Worten verteilte Versorgung, oder es kann pragmatisch durch gesundheitsbezogene Interaktionen beschrieben werden, die umfassend kommunizierende und kooperierende Akteure über Distanz verbinden. Eine Interaktion besteht aus Teilaktionen in einer Kette von Aktivitäten. Einige von ihnen werden technisch vermittelt, um eine geeignete und intendierte Zusammenarbeit zwischen den Akteuren zu ermöglichen [3]. Persönliches eHealth kontrastiert mit dem organisationszentrierten Ansatz, wo Policies, Prozesse oder Bedingungen für die Kommunikation und Kooperation sowie ihre organisatorische und technische Integration a priori definiert und geeignete Zwangsmaßnahmen durchgesetzt werden. Es ist auch verschieden vom prozessgesteuerten Ansatz, wo diese Aspekte verhandelt und zwischen den unterschiedlichen Organisationen im Kontext einer speziellen Krankheit harmonisiert werden. Personalisiertes eHealth ist durch den Status des zu Versorgenden, seine Bedürfnisse, Wünsche, Intentionen und Resultate in einer individuell zugeschnittenen Diagnostik und Therapie definiert. Im Ergebnis ist die Versorgungsumgebung eine virtuelle Organisation mit dynamischen Policies, Workflows, technischen sowie Umgebungsbedingungen. Interaktionen und Prozesse werden dynamisch erzeugt und realisiert, dabei Medizin, Informatik, Biomedizintechnik, Informations- und Kommunikationstechnologien, Bioinformatik und den Strauss der „omics“-Disziplinen wie Genomik, etc. integrierend. Die Beziehungen zur Epidemiologie und Public Health sind wesentlich [4], [5], [6].

1.2 Technische Paradigmen

Um eine individualisierte Gesundheitsversorgung an jedem beliebigen Ort zu erreichen, müssen drei technologische Paradigmen in die Praxis überführt werden: mobiles Computing, pervasives Computing und autonomes Computing (Bild 1). Mobiles Computing ermöglicht die permanente Erreichbarkeit der involvierten Principals, z.B. für Telekonsultationsdienste unabhängig vom Ort. Principals schließen nach der OMG/CORBA-Definition Akteure wie Personen, Organisationen, Systeme, Geräte und Applikationen ein [3], [7]. Pervasives Computing ermöglicht die durchgängige Servicebereitstellung durch jede Art von Principals über die Entfernung, dabei den zu Versorgenden direkt in das Kommunikations- und Kooperationsnetzwerk unter Verwendung von Sensoren, Akten, Body Area Networks (BAN) [8] oder Heimnetzwerke integrierend. Diese Dienste werden im allgemeinen Telemedizinservices genannt. Für eine personalisierte Versorgung müssen die Dienste flexibel sein und können nicht starr vordefiniert werden. Solch adaptives Design von Gesundheitsinformationssystemen, welches auf eine selbstorganisierte Umgebung abzielt, ist die Herausforderung für die aktuelle Forschung und Entwicklung zum autonomen Computing [9]. Mit der Kombination dieser Paradigmen wird ubiquitäre Versorgung gesichert.

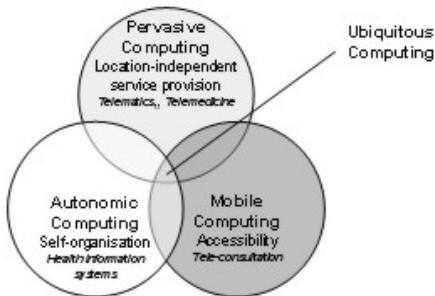


Bild 1. Im Kontext personalisierter Versorgung genutzte Computing-Paradigmen (nach Kim und Müller [10], verändert).

1.3 Das Generische Komponentenmodell

Ein komponenten-orientierter, modellgetriebener Ansatz, der einem einheitlichen Entwicklungsprozess folgt, ist erforderlich, um eHealth-Systeme zu

entwerfen und zu implementieren, die die folgenden funktionalen Anforderungen erfüllen: hohe Flexibilität, Skalierbarkeit, Service-Orientierung, Standardkonformanz, Nutzerfreundlichkeit, Gesetzeskonformanz, Integration aller erforderlichen Domänen, Interoperabilität auf gefordertem Niveau, Vertrauenswürdigkeit und Personalisierung. Um die Systemkomplexität einschließlich aller vorstehend beschriebenen Herausforderungen zu beherrschen, müssen fortgeschrittene Methodologien angewandt werden, die von der Systemtechnik oder von domänenübergreifenden Wissenschaften wie der Kybernetik übernommen wurden. Diese Disziplinen sind wohl begründet durch die Systemtheorie, Informationstheorie, Regelungs- und Steuerungstechnik, Logik, Statistik, usw. Diesen Grundlagen entsprechend, bestehen lebende, technische oder organisatorische Systeme aus Subsystemen, die in Wechselbeziehung stehen. Architektonisch können sie durch ihre Komponenten, deren Relationen und Verhalten beschrieben werden, welche den Komponentenspezialisierungen sowie Umgebungs- und Kontextbedingungen entsprechend instantiiert werden können, was unter Constraints zusammengefasst wird. Externe Beziehungen auf bestimmten Ebenen der Systemgranularität (z.B. Umgebungsvariable) werden durch Generalisierungen integriert. Das entspricht einer Systemaggregation, die zu einer höheren Komplexität führt, wodurch die angesprochenen Beziehungen intern repräsentiert werden. Andererseits können interne Aspekte zur Vereinfachung externalisiert werden.

Um Architektur, Multidisziplinarität und Lebenszyklus eines personalisierten eHealth-Systems, welches sowohl Umgebungsintelligenz als auch durchgängige Versorgung ermöglicht [11], zu beherrschen, kann das System in drei Dimensionen entsprechend dem Generischen Komponentenmodell (GCM) (Bild 2) [3], [12] vereinfacht werden, wobei das GCM als Architektur-Framework oder Referenzarchitektur dient. Die Entwicklung des GCM durch die deutsche OMG/CORBA Gemeinde begann in den frühen neunziger Jahren des vergangenen

Jahrhunderts. Sie wurde anschließend durch die Magdeburger Abteilung für Medizinische Informatik verfeinert und vollendet. Die erste Ebene der Vereinfachung betrifft die Beschränkung auf die jeweils interessierende Domäne. Beispiele für derartige Domänen sind die medizinische, die administrative, die technische, die rechtliche, etc. Innerhalb einer Domäne kann das betrachtete System in seine Komponenten zerlegt oder aus diesen zusammengesetzt werden, um das System zu analysieren oder zu entwerfen. Daraus resultieren unter Verwendung von Spezialisierungs- oder Generalisierungsbeziehungen unterschiedliche Ebenen der Granularität bzw. Komplexität. Im Generischen Komponentenmodell wurden die folgenden Granularitätsebenen spezifiziert: übergreifende Konzepte (übergreifende Geschäftsmodelle, z.B. Unternehmen), Beziehungsnetzwerke (z.B. Geschäftsfelder), Aggregationen (Basisdienste/Funktionen, z.B. Geschäftsprozesse) und Details (Basiskonzepte, z.B. Arbeitsschritte, Transaktionen). Architekturelle Beziehungen müssen zur Vervollständigung der Systemmodellierung sowohl innerhalb der als auch zwischen den Domänen betrachtet werden. Die dritte Dimension der generischen Systemarchitektur berührt verschiedene Sichten auf das System entsprechend dem Standard ISO/IEC 10746 Information Technology - Open Distributed Processing - Reference Model (ODP-RM) [13]. Hier werden die Geschäftsprozesse, Ziele, Policies durch die Unternehmenssicht (Enterprise View), der informationelle Ausdruck dieses Prozesses durch die Informationssicht (Information View) und die funktionelle Dekomposition des Systems aus einer Top-Down-Perspektive (oder der Aggregation von Algorithmen und Diensten in einer Bottom-Up-Sicht) durch die Computation-Sicht (Computational View) dargestellt. Die immer besser beherrschte Herausforderung ist die Harmonisierung der Geschäftsdomäne mit der IKT-Domäne [14]. Die vorstehenden Sichten werden durch plattformunabhängige Modelle des Systems beschrieben, wobei der logische Aspekt ausgedrückt wird.

Technologische Aspekte, z.B. plattformspezifische Details, werden durch die Ingenieurssicht (Engineering View) beschrieben, während die Technologiesicht (Technology View) technische und organisatorische Aspekte repräsentiert. Die Sichten-Dimension des GCM repräsentiert den Entwicklungsprozess des Systems, der aus den folgenden Schritten besteht: Anforderungsanalyse, Design, Implementierung, Evaluierung, Nutzung und Wartung, was dem Rational Unified Development Process (RUP) entspricht [15], [16], [17]. Die neuen Erweiterungen des beschriebenen ODP-RM [13] behandeln auch formal Software-Engineering-Aspekte, wobei sie UML und andere sichtenspezifische Sprachen benutzen [18], [19].

Innerhalb einer Domäne ist die Systemarchitektur durch die Konzepte der Systemkomponenten und ihre Wechselbeziehungen charakterisiert, die die Domänenontologie beschränken und so die Konzeptualisierung der Realität ermöglichen. [12], [20]. Üblicher Weise wird die Darstellung von Konzepten und Assoziationsregeln durch Constraint-Modelle erreicht, die von Referenzmodellen angeleitet werden und das Wissen über die Domäne und die entsprechenden Systeme reflektieren. Hier sollte angesprochen werden, dass der Term Ontologie die allgemeine, unveränderliche, kontextunabhängige Repräsentation der Realität das Sein wiedergibt. Ontologie beschreibt die ursprüngliche Natur der Realität, die von Nutzen für den Wissensgenerierungsprozess [22] sein kann und dafür zur Vermeidung von Fehlkonzeptionen, die mitunter im Kontext des Designs medizinischer Konzepte auftreten [21], verwendet werden sollte. Deshalb sollte eine „philosophische“ Ontologie ein maximal verallgemeinertes KategorienSchema anbieten, dass von allen domänen spezifischen Ontologien benutzt werden kann. Da eine Top-Level- oder Referenz-Ontologie die Realität unabhängig von unserem Wissen darstellt, unterstützt sie die Harmonisierung zwischen domänen spezifischen Ontologien, aber auch deren Konsistenz- und Korrektheitsprüfungen. Im IT-Kontext repräsentiert eine Ontologie die Objekte

einer Domäne der Realität durch systematische Korrelation dieser Objekte mit ihrer Beschreibung, die explizit, formalisiert und in einer für Domänenexperten verständlichen Weise bereitgestellt wird. Ein Ontologie-Modell spezifiziert das Wissen und strukturiert die Konzepte einer bestimmten Domäne. Aus einer stärker beschränkten Sicht auf menschen-generierte Systeme bietet eine Ontologie eine geeignete Domänenterminologie sowie Axiome für die Beschränkung spezifischer Interpretationen an. Mit den Worten nach [23] definiert eine Ontologie die Terme, die benutzt werden, um ein Wissensgebiet zu beschreiben und zu repräsentieren. Wissen stellt symbolisch den wirklichen, fiktiven, künftigen oder angenommenen Gegenstandsbereich einer interessierenden Domäne dar. Wissensrepräsentation reflektiert spezifische Angelegenheiten dieser Domäne in einer beschränkenden Weise.

Aus der Modellierungsperspektive auf ein in Frage stehendes System basiert die Wissensakquise zur Beschreibung der Konzepte und Wechselbeziehungen auf dem Verhalten des Modellierers (Erfahrung, Bildung, Spezialisierung, Einsicht, Wahrnehmung, etc.) und der Systemumgebung, d.h. der Natur der realen Welt. IKT-Systeme fokussieren gewöhnlich viel stärker auf die problemspezifischen Konzepte anstatt auf die problemunabhängigen (im Interaktionkontext die aufgabeneutralen) Konzepte oder Ontologien.

Ernsthaft gesprochen, sollte das Design eines Systems viel stärker auf die korrekte Entsprechung der realen Welt bezogen sein und weniger vom Weg abhängen, auf dem ein Designer sein Ziel verfolgt. Obwohl alle bezogenen Prinzipien einer formalen Ontologie, einer Konzeptanalyse und der Wissensrepräsentation ähnliche Methodologien (z.B. Wissensrepräsentationssprachen, formale Logik) benutzen, wäre es wichtig, sorgfältig zwischen ihnen bei der Beschreibung von Konzepten und Wechselbeziehungen zu unterscheiden.

Der nächste Abschnitt präsentiert weitere Einzelheiten der Konzeptmodellierung, wobei der Fokus mehr auf die Wissensrepräsentation und weniger auf Ontologien ausgerichtet wird. Alle Architekturelemente (Systemkomponenten, deren Funktionen und Beziehungen) müssen eindeutig identifiziert, zertifiziert und registriert werden. Das schließt jedwedes Principal als Akteurkomponente ein und verursacht eine Identifikationsherausforderung, die in Abschnitt 3. kurz diskutiert werden wird. Das resultierende eHealth/pHealth-Informationssystem erfüllt alle Charakteristika, die in den einführenden Sätzen zu diesem Abschnitt fixiert wurden. Organisatorische, kontextuelle, regelbezogene oder andere beschränkende Aspekte des Systems und seiner Komponenten werden durch an die Komponenten gebundene Policies und durch die geeignete Regelung der Komponentenaggregation ausgedrückt.

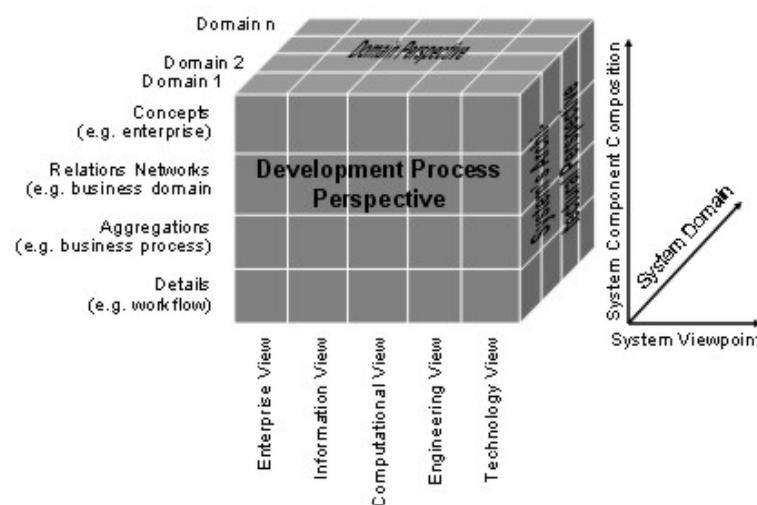


Figure 2. Das Generische Komponentenmodell (GGM).

2. Ergebnisse

Architektur in ihrem Verständnis als Summe der Systemkomponenten, ihrer Funktionen und ihrer Wechselbeziehungen muss durch die Darstellung der bezogenen Konzepte auf dem entsprechenden Granularitätsniveau sowie deren Aggregation auf ein höheres Level beschrieben werden. Unter Verwendung der Definition aus Abschnitt 1. werden Konzepte von Referenzmodellen durch Constraint-Modellierung abgeleitet. Da ein Objekt durch sein instantierendes Konzept zusammen mit einer geeigneten linguistischen Kennzeichnung repräsentiert wird, sind sowohl Referenzterminologien als auch -Ontologien zur Sicherung semantischer Interoperabilität erforderlich. Domänespezifische Ontologien wurden ursprünglich in natürlicher Sprache dargestellt, bewegen sich jetzt aber hin zu einem höheren Niveau an Expressivität und Abstraktion, um die aufgeführten Probleme überwinden zu können. Dieser Formalisierungspfad begann mit Glossaren und Thesauri, gefolgt von Meta-Daten-Beschreibungen und Datenmodellen. Die leistungsfähigsten Formen sind aber strikt formalisierte Beschreibungen (z.B. First-Order Logik (FOL) oder sogar universelle Logik) [24]. Wie schon in Abschnitt 1.4 generisch demonstriert, kann das Generische Komponentenmodell auch für die Analyse und Entwicklung von Kennzeichnungssystemen verwendet werden. Durch Harmonisierung und Integration unterschiedlicher Systeme mit verschiedenen Modalitäten (d.h. Text, Sprache, Bilder, Signale, etc.) aus unterschiedlichen Geschäftsdomänen, wobei unterschiedliche Domänenexperten sprachen (oder im Falle von pHealth-Versorgungsformen mit dem Bürger im Zentrum des Geschehens auch Laiensprachen) verwendet werden, kann ein konsistentes und kohärentes Entwicklungs-Framework für semantisch-interoperable Gesundheitsinformationssysteme breit gestellt werden. Das schließt auch den Aufruf von Komponenten sowie das Profiling von Anwendungen ein. Zur Beschreibung von Konzepten, Regeln und Wechselbeziehungen werden Meta-Sprachen wie die Unified Modeling Language (UML) und die entsprechende Object Constraint

Language (OCL), oder mit der Beschränkung auf strukturelle Informationen die eXtensible Markup Language (XML) Familie benutzt [25], [26]. Dabei wurden Spezialisierungen entwickelt, wie z.B. die von OASIS [27], [28] definierten datensicherheits- und datenschutzbezogenen Sprachen eXtensible Access Control Markup Language (XACML) und Security Assertion Markup Language (SAML). Getrieben durch den für den Anwender unsichtbaren Trend hin zu universeller Logik und allgemeiner Modellierungstheorie, werden der beschriebene Ansatz und seine Harmonisierung in Zukunft erleichtert werden. Das erfordert jedoch viel stärkere Bemühungen bei der Schaffung entsprechender Grundlagen sowie der Tools.

2.1 Die Integration

biomedizintechnischer Systeme

Die Integration von Biomedizintechnik für das Patientenmonitoring bzw. die Patientenversorgung wird typischer Weise unter Verwendung der Standards Sets CEN ISO/IEEE 11073 (abgeleitet vom früheren ENV 13734/13735 "VITAL" und IEEE 1073-x) [29] sowie CLSI (ex NCCLS) POCT-1A [30] realisiert. Biomedizinische Geräte können genauso skalierbar und flexible entworfen und integriert werden wie jedes andere Komponentensystem. Das gilt besonders für künftige mobile, modulare, personalisierte Systeme zur individuellen Versorgung von Patienten. Derartige Systeme können sowohl in klinischen Einrichtungen als auch in der häuslichen Umgebung der Person / des Patienten eingesetzt werden, aber auch dem Patienten folgen und dadurch den Übergang zwischen beiden Umgebungen ausgleichen [31].

Auf verschiedenen Entwicklungsstufen und unter unterschiedlicher Akzentuierung beinhalten typische Systemkonzepte die folgenden Basiskomponenten:

1. Hoch-integrierte Sensoren und Mensch-Maschine-Interface-Komponenten am Körper einerseits und intelligente, drahtlose Sensoren und tragbare, z.B. in Kleidung integrierte Komponenten andererseits;
2. Komponenten und Infrastrukturen für die Kommunikation zwischen diesen Subsystemen und entsprechenden

stationären Systemen und Diensten einschließlich des Übergangs zwischen primärer und sekundärer Versorgung (Body Artea Network, Mobiltelefon, tragbare Funknetze, drahtlose In-House-Funknetze, Arbeitsstationen mit Gateway-Funktionen in das Heim des Patienten einschließlich geeigneter Middleware);

3. Verteilte Funktionen zur Sensor-Signalverarbeitung, Zustandserkennung und -steuerung bis hin zur personenbezogenen und situationsspezifischen Aktivierung von Informations- und Interventionsangeboten (Alarmmanagement, kumulative Registrierung, Verarbeitung und Präsentation multipler Parameter unter Verwendung von PDAs und Workstations);
4. Entscheidungsunterstützungssysteme zur Erkennung und Behandlung von Notfallsituationen, für Patienteninformation sowie bedarfsweise für die Entscheidungsunterstützung für Heilberufler (Lokalisierung, Zugriff auf Referenzinformationen, personen-spezifische Unterstützung bei der Dateninterpretation, sicherer Zugriff auf sensitive Patienteninformationen wie Electronic Health Records oder Personal Health Records).

Die geplanten Systemfunktionen ermöglichen auch vernünftige und bedeutsame Eskalationsstrategien. Dabei sollten die personenbezogenen Systemkomponenten so selbstständig wie möglich sein, um die Kommunikation, Versorgungsanstrengungen und Stromverbrauch zu minimieren, aber auch stets bereit, um zuverlässig mit externen Systemkomponenten zu kommunizieren (z.B. im Fall des Überschreitens von Grenzwerten / Erkennen von Ausnahmesituationen, Notfällen oder Alarmen, aber auch bei der Routine-Datenübertragung). Der Ansatz kann leicht an das GCM adaptiert werden.

2.2 Standards für semantisch-interoperable Personal Health Systeme

Zur Ermöglichung semantischer Interoperabilität durch Etablierung eines einheitlichen Entwicklungsprozesses kann sowohl auf den Rational Unified Process [17] als auch alternative auf das HL7

Development Framework [32], am besten aber auf eine Kombination aus beiden Ansätzen [33] zurückgegriffen werden. Für domänenbezogene Spezifikationen sollten verfügbare domänenspezifische Konzept- und Prozessmodelle wie HL7 Domain Information Models (D-MIMs) oder die neueren HL7 Domain Models (DIMs), HL7 Refined Message Information Models (R-MIMs) und Common Message Element Types (CMETs), aber auch GEHR/openEHR Archetypes [34] wiederverwendet werden. Die Wissensrepräsentationsmodelle anderer Domänen (z.B. Micro- und Nanotechnologie) können ebenfalls benutzt werden [35]. Die anderen für eine umfassende eHealth/pHealth-Umgebung wichtigen Domänen müssen ebenfalls auf standardisierte Weise dargestellt werden. Hier können die in HL7 begonnenen und danach im ISO TC 215 fortgesetzten Arbeiten zum ISO TS 25720 "Health informatics - Genomic sequence variation markup language" [36] sowie die Biobank-Standardisierung als praktische Beispiele angeführt werden. Der Standardisierungsprozess zur Etablierung semantisch-interoperabler eHealth-Systeme muss den gesamten Entwicklungsprozess und alle angesprochenen Architekturparadigmen wie Modellierung, formale Sprachen, Referenz- und Domänenmodelle, Wissensrepräsentationswerkzeuge, domänenspezifische Terminologien und Ontologien, Datensicherheits- und Datenschutzdienste, Kommunikationsprotokolle, etc. [37] umfassen. Hier sind neben HL7 auch die HL7-Liaisons wie ISO TC 215, CEN TC 251, DICOM, OASIS, OMG, aber auch die IHTSDO zu nennen.

2.3 Electronic Health Record und Personal Health Record

Zukünftige fortgeschrittene und nachhaltige eHealth Architekturen für die individuelle Versorgung innerhalb regionaler oder europäischer Dimensionen sind im eHealth Action Plan der europäischen Kommission und der EU-Mitgliedsstaaten [38] beschrieben. Dieses herausfordernde Programm definiert den Electronic Health Record (EHR) als Kernapplikation jeglicher eHealth-Plattformen. Die unterschiedlichen Ansätze hin zu EHR-Implementierungen reichen vom Medikations-File in den Niederlanden und Großbritannien über

den gemeinsamen EHR in der finnischen Lösung bis hin zum umfassenden EHR in Dänemark. Langfristig werden alle Länder einen mehr oder weniger umfassenden, virtuellen EHR anstreben [39]. Im Rahmen des individualisierten Fokus mit der Person im Zentrum des Geschäfts und der Ermächtigung, eine bedeutende Rolle in seiner/ihrer Gesundheit zu spielen, wird die Person auch zur Dokumentation ihres Status' und der realisierten Prozesse beitragen, was in einem Personal Health Records (PHRs) [40] resultiert. EHRs und PHRs folgen dem gleichen architekturellen Ansatz, sind jedoch durch unterschiedliche Policies gesteuert und werden unterschiedlich implementiert.

Neben dem Electronic Health Record spielen die Verbesserung der Qualität und Sicherheit der Versorgung durch evidenzbasierte Medizin und Entscheidungsunterstützung eine außergewöhnliche Rolle. In diesem Kontext wurden das elektronische Rezept und die elektronische Verschreibung unter Verwendung von Entscheidungsunterstützungssystemen in Europa, aber auch in anderen eHealth-Regionen priorisiert.

2.4 Datensicherheits- und Datenschutz-Services in Personal Health Umgebungen

Eine umfassende Sicherheitsinfrastruktur ist die Grundvoraussetzung für jegliche verteilte Gesundheitsinformationssysteme oder Gesundheitsnetze. Dabei sind die Identifikation und Authentifizierung aller involvierten Principals, aber auch andere Sicherheitsdienste wie ein umfassendes ID-Management, Privilegmanagement und Zugriffskontrolle, Anonymisierung und Pseudonymisierung oder Audit zu nennen, die im nächsten Abschnitt in einiger Detailliertheit diskutiert werden. Für Sicherheitsdienste werden häufig Sicherheits-Tokens (z.B. Chipcards) benutzt. Datensicherheits- und Datenschutzdienste in Personal Health Umgebungen erfordern einen grundsätzlich anderen Ansatz im Vergleich zur gegenwärtigen Praxis. Sicherheitsdienste können weder vordefiniert noch durch Administratoren zur Laufzeit verwaltet und administriert werden. Sie müssen formal modelliert werden, wobei universelle Logik oder Meta-Sprachen einzusetzen sind, um ein Management

durch das System und seine Komponenten zu ermöglichen. Als eine Basisanforderung müssen alle Sicherheitsdienste entsprechend dem aktuellen Status, ablaufender Prozesse sowie Umgebungs- und Kontextbedingungen geregelt werden. Deshalb müssen die Datenschutz- und Datensicherheitsdienste durch flexible und intelligente Policies getrieben werden. Sie müssen als integrierte Architekturkomponenten realisiert sein. Die Vielfalt der Use Cases und der domänenüberspannenden Beziehungen kann nicht vollständig durch die Gesetzgebung geregelt werden. Deshalb erhält die Implementierung und Anwendung eines ethischen Frameworks eine zunehmende Bedeutung. Das ist mit speziellen administrativen und Bildungsherausforderungen verbunden.

Für die Integration von Datenschutz und Datensicherheitsdiensten in Personal Health Systeme müssen diese Dienste wie jedes andere System in Befolgung aller Dimensionen des GCM verwaltet werden. Als eine Konsequenz müssen die Konzepte der Sicherheitsservices formal ausgedrückt und mit den anderen Domänen repräsentierenden Komponenten verbunden werden. Dabei wird eine Aggregation von Komponenten innerhalb von und zwischen Domänen entsprechend dem Generischen Komponentenmodell realisiert. Das Binden von Policies betrifft alle Komponenten wie Akteure, Prozesse und Zielobjekte. Standards wie ISO TS 22600 "Health informatics Privilege management and access control" [41] folgen dem GCM-Ansatz.

In der Sicherheitsdomäne hat die Verwendung logischer Ausdrücke im Kontext von Privilegmanagement und Zugriffskontrolle eine lange Tradition. Indem man zunächst mit FOL begonnen hatte, basieren heutzutage viele Ansätze direkt oder indirekt auf Prädikatenlogik. Für weitere Informationen zu den Inhalten dieses Abschnitts siehe z.B. [27].

3. Identifikationsherausforderung in eHealth-Umgebungen

Kommunikation und Zusammenarbeit zwischen Systemen hängen in der Tat von der Identifikation jener Systeme und ihren Komponenten ab, die in die Interoperabilitätskette involviert sind. Es gibt zwei unterschiedliche Ansätze für die Identifikation von Systemen und Komponenten in einem Prozess: Identifikation auf der Basis eines verwalteten Identifikators oder Identifikation auf der Basis von Mustererkennung.

Wie schon in Abschnitt 1.4 vermerkt, ist die Identifikation und Authentifizierung aller Architekturelemente eine grundsätzliche Herausforderung für hoch verteilte, komponentenbasierte, selbst-organisierende Systeme. Das schließt alle Typen von Principals, aber auch Informationsmodelle, Konzepte (im IT-Kontext durch Klassen repräsentiert [20]), etc. ein, damit man stets die richtigen Komponenten aggregiert und verwendet. Dabei werden Signaturen zur konsistenten Identifikation solcher Komponenten benutzt. Das Konzept der Signatur grob gesagt der Bindung einer speziellen Charakteristik an ein Objekt ist sehr breit. Solch eine Signatur könnte als eine Frequenz in einem Spektrum zur Identifikation eines astronomischen Objekts, ein RFID (Radio Frequency Identification) Chip zum Etikettieren von Gütern, eine Charakteristik eines Moleküls in einer Lösung, eine individuelle biometrische Identifikation wie ein Fingerabdruck oder der Fußabdruck eines Tieres instantiiert sein. Sie könnte auch das Ergebnis eines Prozesses unter Verwendung eines individuellen Schlüssels und eines bezogenen kryptographischen Algorithmus sein, wie es bei der digitalen Signatur auf unterschiedlicher Vertrauensebene geschieht. Signaturen können sich auch auf einzelne Komponenten oder auf diese gruppierende Klassen beziehen. Signaturen werden nicht nur zur Überprüfung der Identität und Authentizität, sondern auch zur Integritätsprüfung von Komponenten sogar unter Einschluss der rechtlichen Dimension genutzt.

Die Herausforderung betrifft den gesamten Lebenszyklus dieser Komponenten. Identifikations- und Authentifizierungs-

dienste unterstützen Datenschutz und Datensicherheitsbelange, sind aber auch unverzichtbar zur Reduzierung von Sicherheitsrisiken. Dabei müssen die Identifizierung und Authentifizierung sowohl logisch als auch technisch unterstützt werden. In diesem Kontext spielen wie bereits vermerkt Identifikations- und Authentifizierungstokens eine bedeutende Rolle. Hier sind Smartcards zur Identifikation und Authentifizierung von Personen und spezielle individuelle (und persistente) Eigenschaften (z.B. Blutgruppe) aufzuführen. Beispiele sind Heilberufeausweise, elektronische Gesundheitskarten (Versichertenkarten), aber auch elektronische Pässe. Da die Information identifizierbar, verifizierbar und nachverfolgbar sein muss, können auch mit einem Identifikator etikettierte Güter authentifiziert und nachverfolgt werden. Deshalb werden Technologien wie RFID, die in einer globalisierten Umgebung verwendet werden, das „Internet der Güter“ genannt.

Gegenwärtig werden RFIDs für die Identifikation und Verfolgung von medizinischen Geräten, Proben, Patienten, Arzneimitteln usw. benutzt, wie das erfolgreich in den Asklepios-Kliniken in Deutschland erfolgreich demonstriert werden konnte [42]. Die FDA trachtet nach dem Einsatz von Gerät-Identifikatoren zum Verfolgen und zur Wartung von Medizingeräten. In Österreich wurden Herzschrittmacher eindeutig identifiziert, um diese Identifikation im Gerätemanagement zu nutzen und die im Nationalen Kardiologischen Register gespeicherten Gerätefunktionen vergleichbar und nachvollziehbar zu analysieren. [43]. Ein anderer Ansatz Referent Tracking betrifft die Verfolgbarkeit biologischer Objekte [44]. Die Verfolgbarkeit von Personen (Bewegungsprofil), aber auch von personen-identifizierenden Objekten wie Genoms, etc. bringt Datenschutzrisiken mit sich und kann durch die Gesetzgebung eingeschränkt werden.

4. Die Bildungsherausforderung

pHealth repräsentiert wirklich integrierte Versorgung unter Einschluss vieler Disziplinen. Die Führerschaft bei der Entwicklung und Anwendung fortge-

schrittenen pHealths erfordert breites und zugleich tiefes Wissen und Fertigkeiten auf allen involvierten Domänen, ihren Konzepten, Methodologien, Terminologien und Ontologien sowie geeignete Mittel, diese zu formalisieren und zu präsentieren. Gegenwärtige Bildungsprogramme erfüllen diese Herausforderungen nicht [45]. Um diese Schwäche zu überwinden, bereitet das eHealth Competence Center am Universitätsklinikum Regensburg gemeinsam mit renommierten internationalen Partnern aus den USA, Großbritannien, den Niederlanden und Belgien einen internationalen Masterkurs „International Master in eHealth“ vor. Dieser Kurs reflektiert in geeigneter Weise das umfassende eHealth-Konzept, wie es in diesem Beitrag präsentiert wird. Solche auf die Entwicklung von Eliten abzielende Bildungsprogramme bringen große Herausforderungen an die Studenten, die in den Kurs aufgenommen werden, mit sich. Folglich sind spezielle Aufnahmebedingungen, kleine Klassengrößen, fortgeschrittene Infrastrukturen, geeignete Praxispartner und hohe Qualifikationen der Lehrenden unverzichtbar [46].

5. Diskussion

Interoperabilität impliziert eine Reihe von unterschiedlichen Konzepten, z.B. funktionale Interoperabilität und Internet-Arbeit, semantische Interoperabilität und Anwendungs-Gateways. Die Integration von Gesundheitsinformationen (eHealth) bringt die Forderung nach Interoperabilität zwischen klinischen und anderen gesundheitsbezogenen Stakeholdern, Systemen, Prozessen und Abläufen mit sich. Domänenpezifische Kommunikations- und Interoperabilitätsstandards sind bereits gut etabliert; sie müssen aber für den domänenübergreifenden Gebrauch erweitert werden. Interoperabilitätskonzepte für Medizingeräte und für persönliche oder mobile Systeme erfordern die Einbeziehung aller 7 Schichten des ISO/OSI Referenzmodells, welches in Richtung des Generischen Komponentenmodells unter Einschluss von Terminologie- und Kodierungsaspekten weiterentwickelt werden muss.

Das fortgeschrittene Konzept von pHealth erweitert eHealth durch den Einschluss von smarten Sensoren, am Körper getragenen mobilen Systemen und die situationsspezifische Aktivierung von Anwendungen und Mitarbeitern des Gesundheitswesens und realisiert dadurch personalisierte ubiquitäre Gesundheitsdienste. Body Area Networks und Mikrosysteme sind die Bausteine für zukünftige personalisierte Gesundheits-telematik-Infrastrukturen und erweitern vorhandene Interoperabilitätskonzepte. Bioinformatik und Genomik sind andere bedeutende Säulen von eHealth. Da pHealth ein personalisiertes Prozessmodell fordert, muss auch die zugrunde liegende Diagnostik und Therapie individualisiert werden. Das kann wie gesagt durch die Entwicklung und Anwendung fortgeschrittener Bioinformatik und Genomik erreicht werden, die in die Interoperabilitätskette integriert werden müssen. Dabei müssen Technologien und Methodologien zur Erfassung und Verwaltung persönlicher molekularer Informationen auf unterschiedlichem Granularitätsniveau von der Genomik über Epigenetik, Transkriptomik, und Posttranskriptomik bis hin zur Proteomik betrachtet und verbessert werden. In diesem Zusammenhang müssen Chips zur Untersuchung einzelner nukleotider Polymorphismen als Indikatoren für das individuelle Krankheitsrisiko, Arrays für den Nachweis spezifischer Nukleinsäuren, Mikroarrays für den Nachweis von Antikörpern oder anderen genetischen Bestandteilen auf unterschiedlichem Niveau molekularer Granularität genannt werden. Das Ergebnis wird in internationalen genetischen Datenbanken wie PubMed [47] und GenBank [48] gesammelt und genutzt.

Der Übergang zu pHealth-Informationsystemen mit prozessgesteuerten, service-orientierten, kontextsensitiven, semantisch-interoperablen Informations- und Kommunikationsarchitekturen erfordert offene, hoch-flexible, individuell zugeschnittene Anwendungssysteme für den Versorgten und die Versorger. Die Nutzung aller im Beitrag aufgeführten Paradigmen führt zu intelligentem, ubiquitarem Computing, was mit anderen Paradigmen und Trends wie Health Grids

verbunden ist. Personal Health erfordert auch ein adäquates rechtliches Rahmenwerk und eine Neuorientierung für traditionelle organisatorische Muster.

Zur Harmonisierung der Wissens-repräsentation unterschiedlicher Domänen sind Meta-Modelle und Transformationswerkzeuge unverzichtbar, zum einen für die Entwicklung und Pflege der Wissensbasen innerhalb der Domäne und zum anderen zur Realisierung des für ubiquitäre Versorgung erforderlichen interdisziplinären Ansatzes zwischen den Domänen [20], [49].

Ingenieursansätze wie Systemtheorie, Informationstheorie und Software Engineering, aber auch Biomedizintechnik unterstützen die technologiebezogenen Aspekte der Systemarchitektur, während Ontologien und Wissensrepräsentation auf der Basis universeller Logik abstrakt den Inhalt, also die Konzepte und Regeln der Komponenten und ihrer Wechselbeziehungen beschreiben. Meta-Sprachen helfen, zwischen den Domänenumgebungen einerseits und der abstrakten und/oder technologischen Welt andererseits zu vermitteln. Jedoch muss das strikt den Prinzipien der universellen Logik folgen. Es gibt verschiedene Meta-Sprachen, die diese Forderung noch nicht umfassend erfüllen. Die semiotischen Prinzipien jeder Sprache einschließlich der Meta-Sprachen erlauben jedoch, diese Defizite zu überwinden.

Danksagung

Der Autor seinen Kollegen von HL7, ISO TC 215, CEN TC 251, IMIA, EFMI und vielen anderen Organisationen für ihre freundliche Unterstützung zu Dank verpflichtet. Besonderer Dank gilt Stefan Schulz, Universität Freiburg, Bundesrepublik Deutschland, für die sorgfältige Durchsicht und Diskussion des Beitrags.

Literatur

- [1] Blobel, B.: Introduction into Advanced eHealth - The Personal Health Challenge. In: Blobel B, Pharow P, Nerlich M (Eds.) eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge - Global Experts Summit Textbook, Stud Health Technol Inform Vol. 134, Amsterdam, Berlin, New York, Tokyo: IOS Press, 2008: 3-14.
- [2] Eysenbach, G.: What is e-health? J Med Internet Res 2001;3(2):e20, <http://www.jmir.org/2001/2/e20/> last access: 6 February 2009
- [3] Blobel, B.: Analysis, design and implementation of secure and interoperable distributed health information system, Stud Health Technol Inform Vol. 89, Amsterdam: IOS Press, 2002.
- [4] Haux, R.: Individualization, globalization and health - about sustainable information technologies and the aim of medical informatics. Int J Med Inf 2006; 75: 795-808
- [5] Kulikowski, C.A.: The Micro-Macro Spectrum of Medical Informatics Challenges - From Molecular Medicine to Transforming Health Care in a Globalizing Society. Methods Inf Med 2002; 41: 20-24
- [6] Maojo, V., Martin-Sanchez, F., Billhardt, H., Iakovidis, I., Kulikowski, C.: Establishing an Agenda for Biomedical Informatics. Methods Inf Med 2003; 42: 121-125
- [7] Object Management Group, Inc.: <http://www.omg.org> last access: 6 February 2009
- [8] Fraunhofer Gesellschaft. Body Area Network. <http://www.ban.fraunhofer.de/> last access: 6 February 2009
- [9] Lanfranchi, G., Della Peruta, P., Perrone, A., Calvanese, D.: Toward a new landscape of systems management in an autonomic computing environment. IBM Systems Journal 2003; 42, 1: 119-128
- [10] Kirn, S., Müller, G.: Ministudie zu den F&E-Perspektiven auf dem Gebiet der Gesundheitstelematik. Hohenheim: University of Hohenheim, 2005.
- [11] Kuhn, K.A., Knoll, A., Mewes, H.W., Schwaiger, M., Bode, A., Broy, M., Daniel, H., Feussner, H., Gradinger, R., Hauner, H., Höfler, H., Holzmann, B., Horsch, A., Kemper, A., Krcmar, H., Kochs, E.F., Lange, R., Leidl, R., Mansmann, U., Mayr, E.W., Meitinger, T., Molls, M., Navab, N., Nüsslin, F., Peschel, C., Reiser, M., Ring, J., Rummeny, E.J., Schlüchter, J., Schmid, R., Wichmann, H.E., Ziegler, S.: Informatics and Medicine: From Molecules to Populations. Methods Inf Med 2008; 47: 283-295
- [12] Blobel, B.: Application of the Component Paradigm for Analysis and Design of Advanced Health System Architectures. Int J Med Inf 2000; 60, 3: 281-301.
- [13] ISO 10746 "Information Technology - Open Distributed Processing": <http://www.iso.org> last access: 6 February 2009
- [14] Vetere, G., Lenzerini, M.: Models for semantic interoperability in service-oriented architectures. IBM Systems Journal 2005; 44, 4: 887-903
- [15] Brown, A.W., Delbaere, M., Eeles, P., Johnston, S., Weaver, R.: Realizing service-oriented solutions with the IBM Rational Software Development Platform. IBM Systems Journal 2005; 44, 4: 727-752

- [16] Brown, A.W., Iyengar, S., Johnston, S.: A Rational approach to model driven development. IBM Systems Journal 2006; 45, 3: 463-480
- [17] International Business Machines Corporation (IBM). Rational Unified Process. www.ibm.com/software/awdtools/rup/ last access: 6 February 2009
- [18] ISO/IEC 19793 "Information technology - Open Distributed Processing - Use of UML for ODP system specifications": <http://www.iso.org> last access: 6 February 2009
- [19] ISO/IEC 15414 "Information technology - Open distributed processing - Reference model - Enterprise language": <http://www.iso.org> last access: 6 February 2009
- [20] Akerman, A., Tyree J.: Using ontology to support development of software architectures. IBM Systems Journal 2006; 45, 4: 813-825
- [21] Smith, B., Ceusters, W., Temmerman, R.: Wüsteria. In: Engelbrecht R, Geissbuhler A, Lovis C and Mihals G (Edrs.): Connecting Medical Informatics and Bio-Informatics, Proceedings of MIE2005, Stud Health Technol Inform Vol. 116, Amsterdam, Berlin, New York, Tokyo: IOS Press, 2005: 647-652
- [22] Guarino, N.: Formal Ontology, Conceptual Analysis and Knowledge Representation. In: Guarino N and Poli R: Special issue on Formal Ontology, Conceptual Analysis and Knowledge Representation. www.loa-cnr.it/Papers/FormOntKR.pdf
- [23] W3C, OWL Web Ontology Language - Use Cases and Requirements. W3C Recommendation 10 February 2004
- [24] Russel, S., Norvig, P.: Artificial Intelligence A Modern Approach, 2nd ed., Upper Saddle River, NJ: Pearson Education, 2003
- [25] Shabo, A., Rabinovici-Cohen, S., Vortman, P.: Revolutionary impact of XML on biomedical information interoperability. IBM Systems Journal 2006; 45, 2: 361-372
- [26] World Wide Web Consortium, <http://w3c.org> last access: 6 February 2009
- [27] Blobel, B., Pharow, P.: Formal Policies for Flexible EHR Security. In: Bos L, Roa L, Yugesan K, O'Connell B, Marsh A, Blobel B, editors. Medical and Care Compuetics 3, Stud Health Technol Inform Vol. 121. Amsterdam: IOS Press; 2006.307-316.
- [28] Organization for the Advancement of Structured Information Standards (OASIS), <http://www.oasis-open.org> last access: 6 February 2009
- [29] Institute of Electrical and Electronics Engineers (IEEE), Inc., <http://www.ieee1073.org> last access: 6 February 2009
- [30] National Committee on Clinical Laboratory Standards (NCCLS), <http://www.nccls.org> or <http://www.clsi.org> last access: 6 February 2009
- [31] Blobel, B., Pharow, P., Norgall, T.: How to Enhance Integrated Care towards the Personal Health Paradigm? In: Kuhn KA, Warren JR, Leong T-Z (Edrs.) MEDINFO 2007, Amsterdam, Berlin, Oxford, Tokyo, Washington, DC: IOS Press, 2007: 172-176.
- [32] Health Level Seven Inc.: <http://www.hl7.org> last access: 6 February 2009
- [33] Lopez, D.M., Blobel, B.: A Development Framework for Semantically Interoperable Health Information Systems. Int J Med Inf 2008; 78, 2: 83-103.
- [34] openEHR Foundation: <http://www.openehr.org> last access: 6 February 2009
- [35] Becks, T., Dehm, J.: IMEX - A New Knowledge Platform for Microsystems in Medicine. <http://www.vde-mikromedizin.de>
- [36] ISO TS 25720 "Health informatics - Genomic sequence variation markup language" International Organization for Standardization / TC 251 Medizinische Informatik <http://www.iso.ch/tc215>
- [37] Engel, K., Blobel, B., Pharow, P.: Standards for Enabling Health Informatics Interoperability. In: Hasman A, Haux R, van der Lei J, De Clercq E, Roger-France F (Edrs.) Ubiquity: Technology for Better Health in Aging Societies, Stud Health Technol Inform Vol. 124, Amsterdam: IOS Press, 2006: 145-150.
- [38] European Commission. The right prescription for Europe's eHealth. http://ec.europa.eu/information_society/activities/health/policy/index_en.htm
- [39] Blobel, B.: Comparing EHR Models. In: eHealth Conference 2007 - Experts' Special Interest Sessions, Schriftenreihe der GVG, Bd. 58, Bonn: nanos Verlag, 2008: 63-80.
- [40] ISO TS 18308 "Health informatics - Requirements for an Electronic Health Record Architecture"
- [41] ISO TS 22600 Health informatics - Privilege management and access control
- [42] RFID Atlas. RFID Anwenderbeispiel Future Hospital. <http://www.rfidatlas.de>
- [43] Merl, O., Peschorn, R., Nurnberg, M., Steinbach, K.: Austrian pacemaker registry: improving reporting compliance in 60center national registry. Computers in Cardiology 1999; 317-318
- [44] Ceusters, W., Elkin, P., Smith, B.: Negative findings in electronic health records and biomedical ontologies: A realist approach. Int J Med Inf 2007; 76S: S326-S333
- [45] Blobel, B.: Educational Challenge of Health Information Systems' Interoperability. Methods Inf Med 2007; 46: 52-56.
- [46] eHealth Competence Center. Int. Master Course. <http://www.ehealth-cc.de> last access: 6 February 2009
- [47] National Library of Medicine. PubMed. <http://www.ncbi.nlm.nih.gov/pubmed/> last access: 6 February 2009
- [48] National Institute for Health. GenBank. www.ncbi.nlm.nih.gov/Genbank/ last access: 6 February 2009
- [49] Czarnecki, K., Helsen, S.: Feature-based survey of model transformation approaches. IBM Systems Journal 2006; 45, 3: 621-645

Semantische Interoperabilität erfordert die richtigen Modelle und Code-Systeme: Eine Prüfung verschiedener Ansätze für Score Systeme

F. Oemig¹, B. Blobel²

¹Agfa HealthCare GmbH, Bonn, Germany,

²eHealth Competence Center, University of Regensburg Medical Center, Germany

Zusammenfassung: Das Erreichen der semantischen Interoperabilität erfordert nicht nur den Einsatz von Kommunikationsstandards wie HL7 mit den dazugehörigen Modellen und Spezifikationen, sondern für die Instanziierung auch die Einschränkung dieser Modelle in ihren Eigenschaften, Datentypen und den verwendeten Werten und Codesystemen. Die Anwendung dieser Strategien kann aber trotzdem zu unterschiedlichen und damit ggf. inkompatiblen Modellen führen. Dieses Papier gibt einen kurzen Überblick über unterschiedliche Modellierungsansätze, die anhand der Score und Assessment-Systeme erläutert werden. Es werden die Vor- und Nachteile der verschiedenen Ansätze demonstriert. Die präsentierten Ergebnisse berücksichtigen die Übermittlung derselben Basisinformationen mittels HL7 v2.x und V3, um die Implementierungsaufwendungen insgesamt zu reduzieren.

Stichwörter: Gesundheitstelematik, Kommunikationsstandard, HL7, Semantische Interoperabilität, Konformanz, Implementierungsleitfaden, Scores und Assessments.

1. Einführung

Die Nutzung von Kommunikationsstandards wie HL7 [1], [2] ist eine Grundvoraussetzung, um semantische Interoperabilität überhaupt erreichen zu können. Nationale oder anwendungspezifische Anforderungen können in Form von "constraint Profiles" ausgedrückt werden, die den zugrundeliegenden Standard weiter einschränken. Wenn ein Hersteller mit diesen Anforderungen konformant sein will, sollte er eine präzise Aussage über das treffen, was seine Schnittstellen erfordern bzw. voraussetzen [3]. Dies wird dann über sog. "implementable Profiles" ausgedrückt, die

Einschränkungen der vorab genannten constrainable Profiles sind. Erfahrung zeigt, dass ein Standard die Bereitstellung einer derartigen Dokumentation erzwingen sollte, damit ein Hersteller eine HL7-Konformität überhaupt erklären darf [13].

Andererseits sind die genutzten Kommunikationsstandards (HL7 v2.x und V3) sehr mächtig und nicht immer selbst erklärend. Sie bieten einen riesigen Satz von Optionen an, um in unterschiedlichen Domänen anwendbar zu sein. Die Schaffung von Implementierungsleitfäden, die Teile des Standards erklären, ist eine unabdingbare Voraussetzung [4], [5]. Sie stellen eine klare Anleitung dar und helfen, den Standard zu verstehen und auf eine einheitliche Weise umzusetzen, um Inkompatibilitäten zu reduzieren. Sie enthalten Use Cases, Storyboards, dynamische und statische Modelle sowie Beispiele. Semantische Interoperabilität hängt von einem gemeinsamen Verständnis und Verwendung der Informationen und deren Instanzen ab, die auf gemeinsamen Konzepten und deren Interpretation einschließlich einer Konzept-Grammatik und einer ontologiebasierten Terminologie basieren. Sogar in einer deterministischen Welt von quantitativen Messungen führt eine Interpretation vielleicht zu unterschiedlichen Ergebnissen. Dies wird noch deutlicher, wenn nicht-deterministische Messungen oder zusätzlich Algorithmen benutzt werden, die ebenfalls harmonisiert werden müssen. Um Instanzen der realen Welt vergleichbar zu machen, werden sie transformiert oder zu einem neuen Ausdruck gruppiert, der Score genannt wird. Hierbei gilt es einige Anforderungen zu erfüllen, um semantische Interoperabilität auf Instanzenebene garantieren zu können.

In einigen Fällen können unterschiedliche Implementierungsleitfäden zum gleichen Thema eine Interoperabilität sogar verhindern: Die einfache Frage eines Entwicklers, wie ein "APACHE"-Score mit HL7 v2.x kommuniziert werden soll, führt zu erstaunlichen Ergebnissen, die nun aufgeführt werden sollen.

2. Methoden

Die obengenannte Frage initiierte eine Untersuchung über Scoresysteme im allgemeinen. Eine erste Suche innerhalb des Internets enthüllte viele (>100) unterschiedliche Scoresysteme, die in der medizinischen Domäne für unterschiedliche Zwecke und Fachrichtungen im Gebrauch sind: Apache, Apgar, Braden, Barthel, BMI, GCS, Gleason, MODS, SAPS, TISS, um nur einige bekannte Scoresysteme zu nennen. Begleitende Informationen sind in unterschiedlichen Formen verfügbar: als herunterladbares Dokument oder Webseiten, die die Inhalte beschreiben, als wissenschaftliche Dokumente über Nutzbarkeit und Ergebnis-Interpretation, und nicht zuletzt als MS-Excel-Tabelle oder aktive Webseiten für eine online Berechnung. Manchmal variieren die Spezifikationen für das gleiche Scoresystem ein klein wenig und stellen deshalb unterschiedliche Varianten dar.

Dieses Papier konzentriert sich weder auf die Nachprüfung von Scoresystemen noch auf die Kalkulation (Algorithmen) von den resultierenden Scoreergebnissen. Der wichtige Aspekt ist die Analyse und die Spezifikation der Details, die notwendig sind, um ein bestimmtes Scoresystem und seine Details innerhalb einer Nachricht oder eines Dokumentes zu identifizieren.

Für Interoperabilitätszwecke zwischen unterschiedlichen Anwendungen müssen geeignete Codes für beides zugewiesen werden: das Scoresystem und jedes seiner Details. Gegenwärtig stellt nur Snomed CT ([9]) für einige Scoresysteme Details mit einigen Codes in Form individueller Konzepte bereit. Diese Konzepte werden innerhalb der internen Achsen von Snomed CT in Form von Beziehungen zu anderen Konzepten organisiert, aber sie werden nicht als zu einem bestimmten Score gehörig identifiziert.

Das neuste Release des RELMA Werkzeuges für LOINC ([12]) stellt jetzt umfassende Informationen über einige Scoresysteme bereit. Man kann sogar urheberrechtliche Bemerkungen und Codes finden.

Beide Codesysteme stellen allerdings keine Anleitung bereit, wie dieses Codes - vorausgesetzt man kann sie auffinden mit Kommunikationsstandards kombiniert werden können, um eine interoperable Kommunikation zu unterstützen.

2.1 Der "holländische" Ansatz

Vor vier Jahren haben holländische HL7 Kollegen von NICTIZ begonnen, mehrere Scoresysteme zu analysieren (ca. 20) und Informationen über ihre Details zu sammeln. Basierend auf diesen Informationen entwickelten sie individuelle Domänen-Modelle für jedes Scoresystem unter Verwendung der HL7 V3 Methodologie.

Die folgende Grafik ist ein Beispiel für den "Barthel Index". Hier steht das Scoresystem in der Mitte und die zehn Komponenten arrangieren sich darum (Abb. 1).

Für jedes Scoresystem wurden vollständig ausmodellierte R-MIMs geschaffen. Innerhalb eines solchen Modells wird jeder individuelle Messwert mit seiner eigenen Klasse dargestellt. Jede Klasse in einem Modell wird von seinem eindeutigen Namen und dem Code identifiziert. In der HL7 V3 Methodologie ist der Code wichtig, da er benutzt wird, um die Inhalte zu identifizieren, während in den meisten Implementierungen der Name der Klasse (dargestellt durch ein XML Element) für

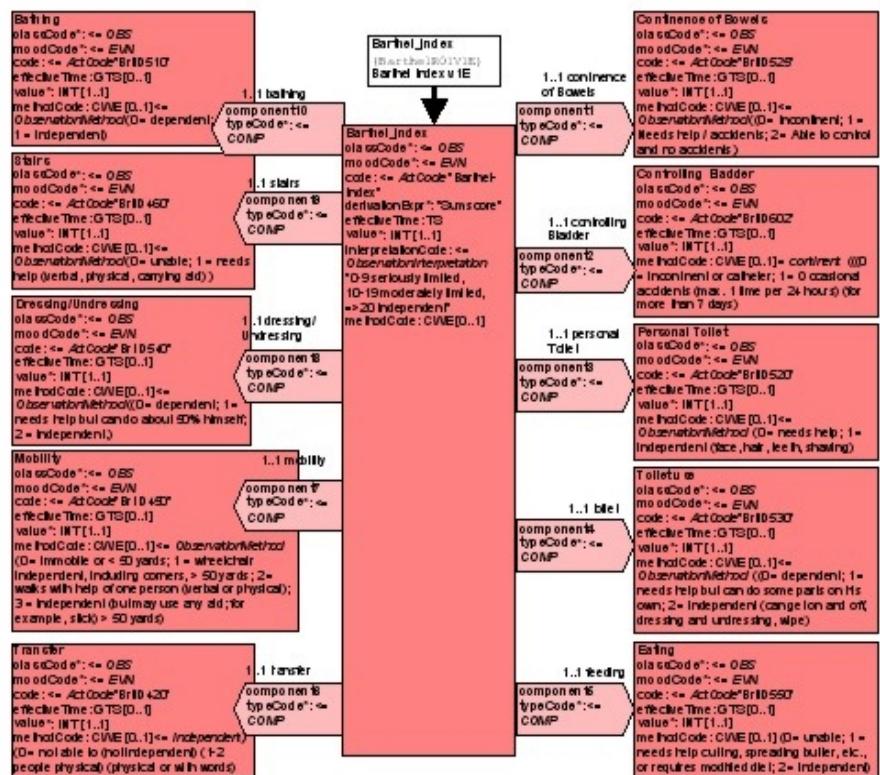


Abbildung 1. Der holländische Ansatz.

Validierungen benutzt wird. Dieser Ansatz hat offensichtlich den Nachteil, dass er individuelle Implementierungen für jedes Scoresystem erfordert.

Das zweite Problem ist nicht spezifisch für diesen Ansatz. Die Autoren der Modelle haben den Klassen proprietäre Codes zugewiesen. Das generische Modell, das im folgenden beschrieben wird, löst dieses Problem durch das Zuordnen von verschiedenen Codes für das gleiche Konzept und das Eintragen in der Datenbank. Auf diesem Weg können die proprietären Codes auf LOINC und Snomed CT gemappt werden.

Die Analyse der Definitionen jeweils eines Score-Systems hat gezeigt, dass sie sich unterscheiden können. Der holländische Barthel Index bspw. benutzt andere Werte, um die Ergebnisse darzustellen. Die Gesamtsumme reicht deshalb von 0 bis 30. Der deutsche Barthel Index hat hingegen einen Interpretationsbereich von 0 bis 100. Deshalb hat der Ergebniswert von 30 im holländischen System eine andere Interpretation als der gleiche Score für den deutschen Barthel Index.

Als Konsequenz müssen sich die Konzepte unterscheiden, um das gesamte Scoresystem zu definieren. Dies kann deshalb nur mit anderen Codes realisiert werden.

3. Ergebnisse

Das allgemeine HL7 V3 Paradigma nutzt Codes als ein Mittel eindeutiger Identifikation. Diese Codes werden für Objektinstanzen wie Personen, ambulante und stationäre Aufenthalte, individuelle Diagnosen oder Maßnahmen benutzt. Sie haben gemeinsam, dass eine Interpretation während der Verarbeitung erforderlich ist. Offenbar scheint ein interpretativer Ansatz am besten geeignet zu sein, um Scoresysteme zu beschreiben. Dies führt zu einem generischen Ansatz.

3.1 Der generische Ansatz

Es sollte deshalb ein generisches Modell entwickelt werden, das ein Beschreiben aller Details fast aller Scoresysteme erlaubt. Die folgende Liste stellt einen Überblick über die Details bereit, die abgedeckt werden sollten:

- Scoresystem:
 - Name (englisch und deutsch) inkl. Kurzbeschreibung.
 - Autor.
 - Referenzen/Hinweise (Hyperlinks zu Webinhalten).
 - Zugeordnete Fachrichtungen.
 - Status-Informationen.
 - Grobe Berechnungsart.
 - Hinweise auf Copyrights und Trademarks.
 - Score Details:
 - Mittels rekursiver Definition.
 - Beschreibung (englisch + deutsch).
 - Datentyp (Constraints) für:
 - HL7 v2.x,
 - HL7 V3.
 - (erwartete) Werte für jedes Konzept.
 - die Interpretation von Werten.
 - zugeordnete Codes (proprietär, LOINC [10] und Snomed CT [9] gleichzeitig) inkl. LOINC-spezifischer Details, die das Anfordern von neuen Codes für neue Interpretationen erlaubt.

und/oder ihre Details können zu Referenzbereichen in Beziehung gesetzt werden. Die daraus resultierenden Probleme werden später in diesem Papier diskutiert. Nicht zuletzt hat jedes Scoresystem und/oder jedes Detail vielleicht seine eigenen involvierten Personen (Ausführenden, Autor, Prüfer), welche durch Participations zu den entsprechenden Common Message Element Types (CMET) inkludiert werden.

Der Implementierungsleitfaden [11] legt die Bindung der Codes zu den entsprechenden Segmenten fest.

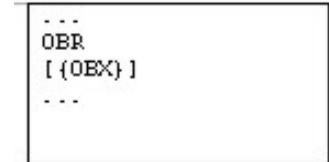


Abbildung 3. HL7 v2.x Modell.

3.3 Das HL7 v2.x Domänen Modell

HL7 v2.x hat eine vereinfachte Darstellung, die nur 2 Ebenen berücksichtigt. Das Scoresystem wird als Auftrag (Order) und die Details als Beobachtungen ausgedrückt (Abb. 3).

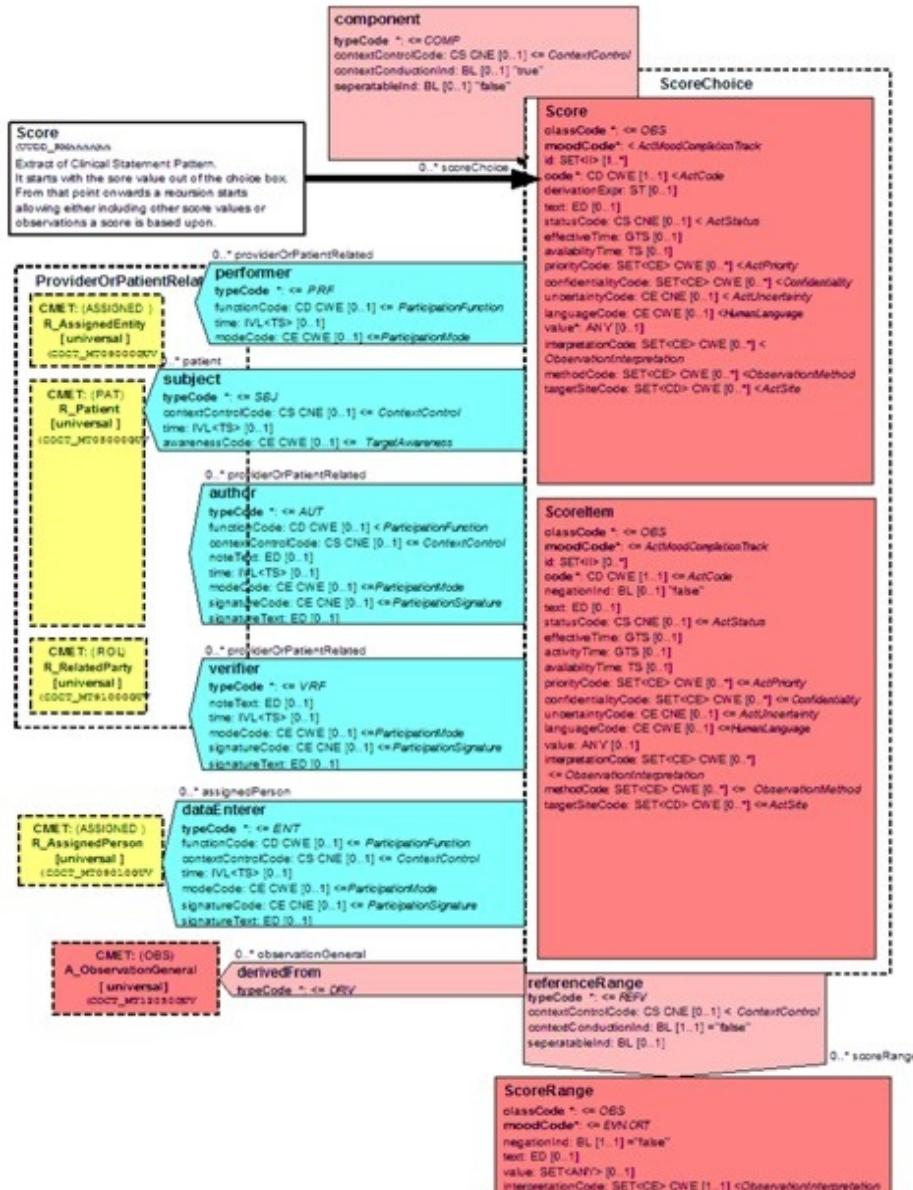


Abbildung 2. HL7 V3 DMIM für Scoresystems.

3.4 Interpretation von Werten

Ein weiteres Problem ist die Spezifikation von Referenzbereichen für Beobachtungen. Normalerweise hat eine einzelne Beobachtung einen Referenzbereich mit einer unteren und einer oberen Grenze. In den meisten Fällen aber nicht in allen werden alle Werte innerhalb dieses Bereiches als normal, alle anderen entweder als zu niedrig oder zu hoch behandelt. Dies führt zu einer unabhängigen Markierung (Flag), die die Interpretation angibt (Abb. 4).

Bei Scores sind alle möglichen Werte im voraus bekannt und können aufgezählt werden. Nur einige Scores (wie BMI wenn man es als Score auffasst) benutzen wirkliche Messwerte, für die der aktuelle Mechanismus gültig ist. Es gibt keine abnormalen Werte außerhalb eines solchen Bereiches. Alle möglichen Werte können klassifiziert werden, damit für jede Klasse eine bestimmte Interpretation zugewiesen werden kann (Abb. 5).

Die Konsequenz für unser obiges Modell ist die Übermittlung eines Satzes von Referenzbereichen, wo jeder einzelne seine eigene Interpretation hat, die wiederum in Form eines Codes ausgedrückt werden kann.

3.5 Klassifikations-Systeme

Scoresysteme bestehen aus Klassen (oder Werten), die an einen eindeutigen Identifikator oder Code gebunden werden. Zum Ausdrücken von Beziehungen und Abhängigkeiten sind multi-axiale Klassifikationen eingeführt worden. Ein Beispiel für so ein multi-axiales Klassifikationssystem ist Snomed III, ein Vorgänger des aktuell genutzten Snomed CT. Eine Ontologie führt dann Beziehungen zwischen den klassifizierten Konzepten ein. Ein wichtiger Teil der präsentierten Arbeit ist die Zuweisung von Codes, die die Identifikation des übermittelten Konzepts ermöglichen. Dabei können unterschiedliche Codesysteme benutzt werden:

- LOINC,
- SNOMED CT,
- ICD 10,
- proprietäre Codes.



Abbildung 4. Einfache Interpretation von Werten.

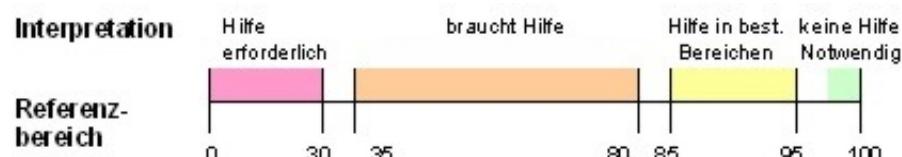


Abbildung 5. Verbesserte Referenzbereiche

Das am einfachsten zu nutzende Codesystem (neben proprietären Codes) ist LOINC, das vom Regenstrief Institute eingeführt wurde [12]. Es hat einen einfachen Mechanismus zur Definitionen von neuen Konzepten. Wenn kein Konzept gefunden werden kann, müssen die Details für die Anforderung eines neuen Codes parallel dazu festgehalten werden. Die Datenbank (s.u.) kann deshalb mehrere Codes von unterschiedlichen Codesystemen parallel verwalten.

übermittelt wird. Letzteres sorgt für eine zusätzliche 1:1 Beziehung.

3.6 Die Score-Datenbank

Die erste Zusammenfassung der Score Details wurde in einer MS-Excel-Tabelle dargestellt. Die ersten Schritte in der Entwicklung eines Implementierungsleitfadens haben gezeigt, dass die Details mehrdimensional sind, so dass eine "echte" Datenbank als geeigneter erschien. Das führte zur Entwicklung eines ER-Modells, um die Informationen in einer relationalen Datenbank zu speichern. Eine GUI hilft in der effektiven Verwaltung der o.g. Informationen. Auswertungen /Berichte (Reports) helfen, um die Daten an Interessenten zu verteilen sowie Tabellen zu generieren, um neue LOINC Codes vom Regenstrief Institut anzufordern.

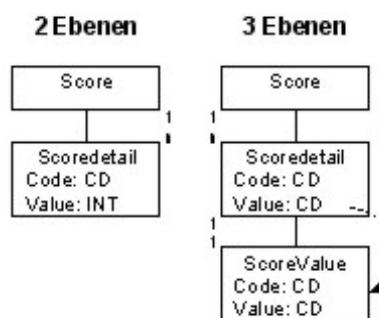


Abbildung 6. Gegenüberstellung des 2 bzw. 3 Ebenenansatz.

Snomed CT ([9]) ist ein sehr ausdifferenziertes Codesystem, das die Spezifikation von Konzepten mitunter für jeden individuellen Wert erlaubt. Das konzeptionelle Domänen-Modell für Scoresysteme erlaubt prinzipiell die Übersendung der Ergebnisse entweder in einem zwei oder drei Ebenen Ansatz (Abb. 6). Der zwei Ebenen Ansatz sendet einen Wert für den beobachteten Messwert, wohingegen innerhalb des drei Ebenen Ansatzes noch ein Code als Komponente

Ein HTML-Generator erlaubt die Bereitstellung der Informationen im WWW. Ein weiterer Generator, der gerade entwickelt wird, wird die Daten in einem R-MIM im Definition-Mood erzeugen, so dass diese Informationen direkt in einem Dokument oder einer Nachricht verwendet werden können. Dieses R-MIM enthält so die Beschreibung/Detailspezifikation des eigentlichen Score-Systems.

3.7 IHE PCC

Das IHE Patient Care Coordination (PCC) TC hat eine Spezifikation geschaffen [14], um Informationen für einige wenige Scoresysteme innerhalb eines CDA Dokumentes zu übersenden: Functional Status, Pain Scale, Braden Score und Geriatric Depression-Scale (GDS).

Leider konzentriert sich die Spezifikation nur auf ein Template, das bei der Identifikation hilft, ob ein entsprechender Teil in solch einem Dokument enthalten ist oder nicht. Schematron Regeln verifizieren, ob alle erforderlichen Teile mit so einer Template ID und den zugeordneten LOINC Codes in einem Dokument vorkommen. Die Spezifikation selbst kümmert sich nicht um weitergehende Details, weder in textueller Form noch als Level 3 strukturierter Entries.

Erweiterungen zu der bereits erwähnten Datenbank erlauben auch das Erzeugen von Score-spezifischen Patterns mit mehr Details. Dies umfasst basierend auf zusätzlichen Informationen nicht nur die textuelle Spezifikation als ein Dokument selbst, sondern auch das dazugehörige HL7 V3 Template XML Schema und Schematron Regeln [15].

Die einzige offene Frage ist die Herausforderung, um die Inhalte dieser Datenbank in das gegenwärtige bestehende IHE PCC Technical Framework zu integrieren, um Verwaltungskonflikte zu vermeiden. Gemäß der Entscheidung des IHE PCC Technical Committee von 2008 wird es ein zweiteiliges Whitepaper geben, das später als eigenständiges Integrationsprofil bereitgestellt werden kann. Der erste Teil erläutert den Aufbau und das Mapping auf die Kommunikationsstandards v2.x, V3 und CDA, der zweite wird mit den Detailinformationen aus der Datenbank generiert.

3.8 UML-Modell

Das ER-Schema für die Datenbank basiert auf einem vereinfachten Modell für die erforderlichen Fakten. Des Weiteren stellt die Datenbank nur die Details bereit, die notwendig sind, um die Inhalte zu spezifizieren. Deshalb stellt sie kein Feld bereit, um den resultierenden Wert zu speichern. Für das Kommunikations-

Szenario wird der Wert wichtig, womit es im UML Modell enthalten ist.

Der wichtigste Umstand ist die Tatsache, dass ein Score selbst eine Beobachtung ist. Deshalb wird er als eine Spezialisierung davon dargestellt (sehen Sie a) in Abb. 7). In der Datenbank werden diese zwei Klassen zusammengeführt, um die Daten einfach verwalten zu können. Dies macht es leichter, die Rekursion zu verwalten, d.h. eine Beobachtung kann andere Beobachtungen als Komponenten enthalten. Daher stellt jeder Eintrag in dieser Tabelle ein eigenes Konzept dar, welches in anderen Definitionen wiederverwendet werden kann.

Die nächste wichtige Tatsache ist die Verwendung von Codes, um das semantische Konzept darzustellen. In Abbildung 7 stellt die Klasse CatalogEntry ein Konzept mit einem Code und einer Beschreibung mit Bezug zu einem Katalog/Codesystem dar.

Ein solches Konzept könnte vielleicht aus einem Katalog wie LOINC [10] oder Snomed CT [9] stammen. Die zwei verschiedenen Spezialisierungen als ObsCatEntry und Interpretation helfen, die Referenzierungen korrekt vorzunehmen. Die erste Klasse soll Konzepte für Codes von Beobachtungen identifizieren, wohingegen die zweite Klasse Konzepte für Interpretationen darstellt.

Jedes Konzept für eine Beobachtung kann mit einem Referenzbereich verbunden werden, die Informationen über die erwarteten Beobachtungswerte bereitstellt (sehen Sie e) in Abb. 7). Ein Referenzbereich hingegen wird mit einer Interpretation kombiniert (sehen Sie b) in Abb. 7). Typischerweise kann eine Beobachtung für Scores unterschiedliche Referenzbereiche mit unterschiedlichen Interpretationen besitzen (sehen Sie 3.4 Werte Interpretation oben).

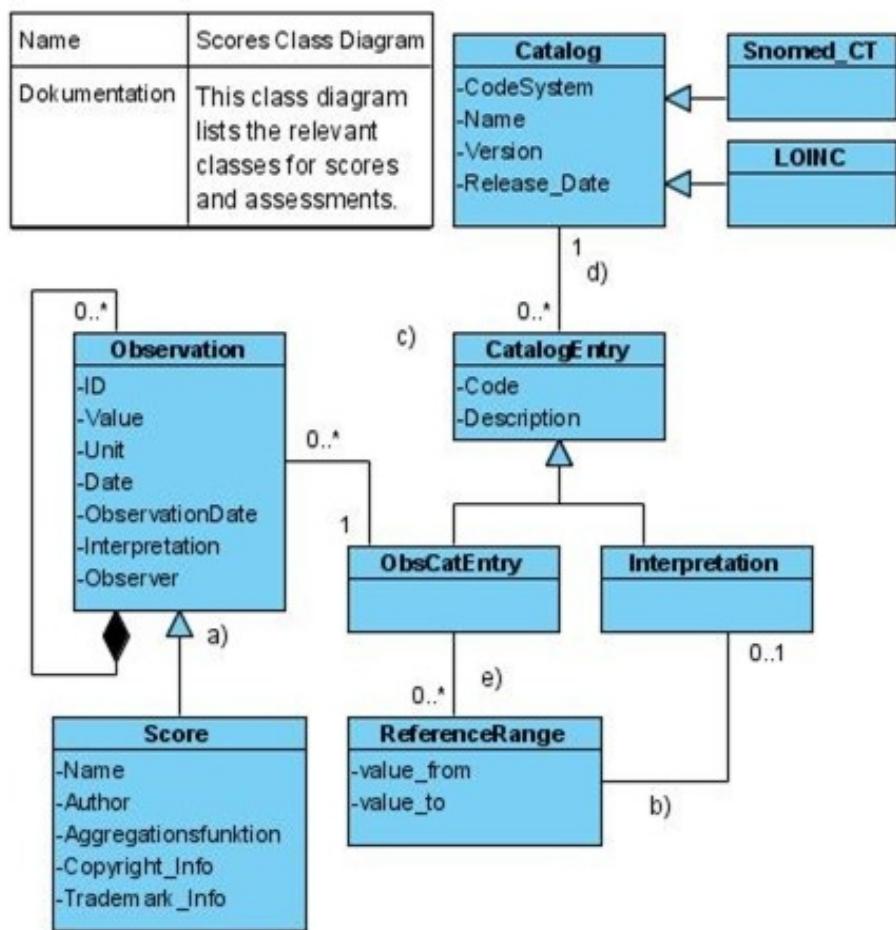


Abbildung 7. UML Modell.

3.9 UML-OWL-Modell

Um mit dem in dem oben erwähnten UML Modell enthaltenen Wissen zu argumentieren, sollte es in eine OWL-Darstellung konvertiert werden. Im Prinzip sind beide ganz ähnlich. Der wichtigste Unterschied ist die Abstammung von einer gemeinsamen Klasse, genannt owl:Thing. Des weiteren werden einige der Eigenschaften als Beziehungen zu anderen Klassen dargestellt. Eine Person (Beobachter oder Autor) als eine Spezialisierung von LivingSubject ist ein gutes Beispiel dafür.

Anmerkung: die Anmerkungen für Abb.7 gelten auch für Abb. 8.

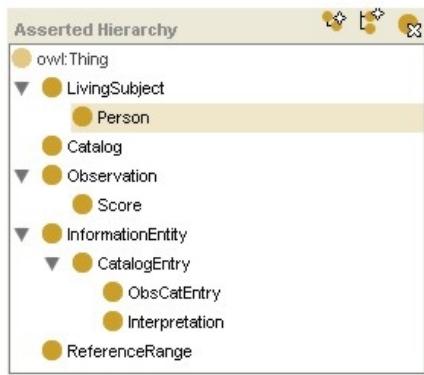


Abbildung 9. Protégé OWL-Modell.

Abbildung 9 stellt die Klassen-Hierarchie in Protégé dar [16].

3.10 Integritätsregeln für das OWL-Modell

Diese Klassenhierarchie allein ist nicht ausreichend. OWL erlaubt auch, den Klassen Beschränkungen zuzuweisen, die helfen, die Beziehungen unter ihnen zu verifizieren. Das Beispiel in Abbildung 10 drückt aus, dass jeder Score eine Beobachtung haben muss. Weiterhin erbt es die Beschränkung von Observation, um einen Code aus CatalogEntry zu haben.

Alle Beziehungen können als Beschränkungen ausgedrückt werden, die von sog. Reasonern verwendet werden können, um die Konsistenz der Ontologie überprüfen.

3.11 Handhabung von Übersetzungen

Gegenwärtig stellt die Datenbank textuelle Informationen über die Scoresysteme und ihre Details in deutscher und englischer

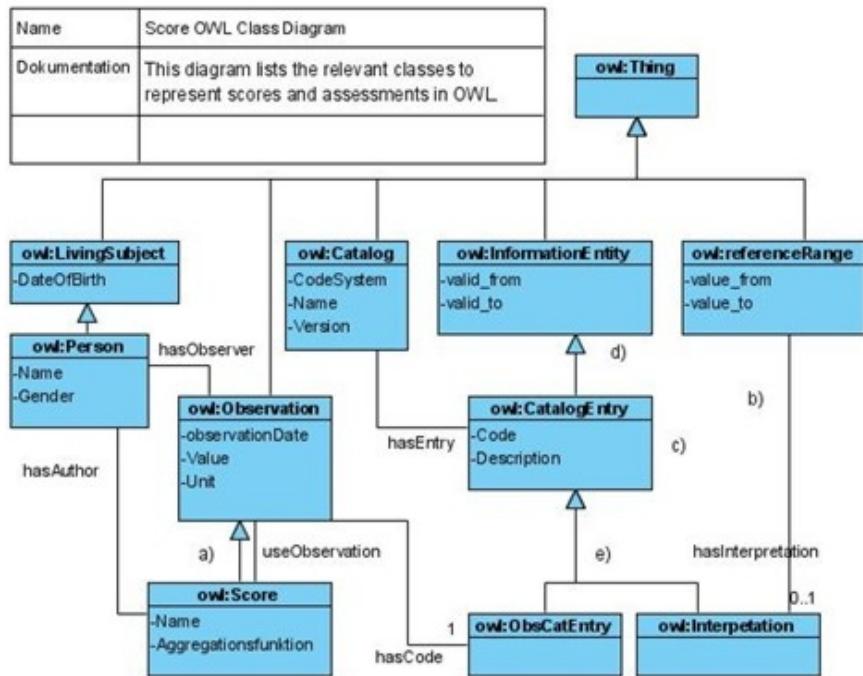


Abbildung 8. UML OWL Modell.

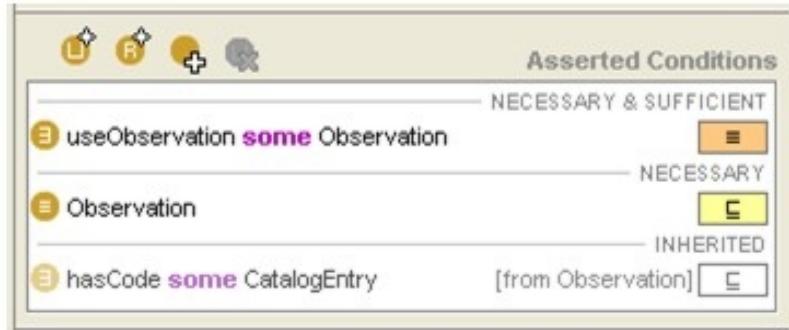


Abbildung 10. Beschränkungen für Scores in Protégé.

Sprache bereit. Die meisten Felder in der Datenbank sind zu diesem Zweck dupliziert worden. Der erste Autor hat darüber nachgedacht, hier ebenfalls einen generischen Ansatz für Übersetzungen zu implementieren. Dies würde bei dem UML-Modell anfangen. Allerdings führte das Wissen um das mühsame Sammeln der notwendigen Codes zu dem Entschluss dass gegenwärtig nicht mehr als zwei Sprachen berücksichtigt werden sollten: Die Verfügbarkeit der Codes wird den Übersetzungsprozess signifikant vereinfachen, da in den übersetzten Originalkatalogen nachgeschlagen werden kann.

3.12 Einsatz der Datenbank

Die Datenbank wurde auf Basis von MS-Access entwickelt, da hiermit alle Daten (Daten über die Scores, Formulare und Berichte) in Form einer einzelnen Datei bereitgestellt werden. Während der ersten Entwicklungsschritte des Projektes haben die Beteiligten sich diese Datenbank per eMail zugeschickt. Mit dem Anwachsen des Projektes ist dieses Vorgehen nicht mehr weiter aufrecht zu erhalten. Der Inhalt sollte daher auf der DCM Webseite [18] als generierte HTML-Dateien veröffentlicht werden. Alle Korrekturen und Ergänzungen sollten deshalb an den Autor geschickt werden. Dieses neue Vorgehen erlaubt die Protokollierung und Verifikation von Änderungen.

3.13 Projektstatus

Gegenwärtig (Jan. 2009) haben intensive Diskussionen mit den Beteiligten stattgefunden. Die HL7 Patient Care Work Group hat dieses Material (R-MIM) zur Abstimmung als Draft Standard for Trial Use (DSTU, HL7 MB21 Jan. 2009) gestellt und als Ergebnis eine prinzipielle Zustimmung erhalten.

Darüber hinaus hat das IHE PCC Technical Committee der Erstellung eines Whitepaper zugestimmt.

4. Schlussfolgerung

Die Details, die oben aufgezählt werden, werden jetzt in der beschriebenen Datenbank verwaltet. Gegenwärtig enthält sie mehr als 80 Scoresysteme. Für eine Reihe von ihnen werden auch schon die erforderlichen Details bereitgestellt. Die meisten von ihnen erfordern eine Verifikation durch die verantwortlichen Organisationen und Fachverbände. Weiterhin wird der vorgestellte Ansatz mit den verantwortlichen Personen und Technical Committees innerhalb der SDOs diskutiert. Trotz der schon existierenden ausführlichen Spezifikationen für ein paar Scoresysteme setzen sich alle Beteiligte für die weitere Ausarbeitung des generischen Ansatzes ein, der beim Verwalten der Details helfen wird. Die Bereitstellung eines abgestimmten Satzes

von Codes auch wenn es mehrere Codes aus verschiedenen Codesystemen für dasselbe Konzept sind ist die wichtigste Aufgabe, um eine semantische Interoperabilität zu erreichen.

Anerkennungen

Die Autoren würden gern Frau Dr. Sylvia Thun (DIMDI), Hr. Dr. Rainer Röhrig und Frau Ricarda Rüth (beide Universitätskrankenhaus Gießen) für ihre Unterstützung bei der Ausarbeitung der Modelle und dem Füllen der Datenbank danken. Des weiteren Herr Dr. William Goossen (Acquest, NL) für die Unterstützung des generischen Ansatzes innerhalb der HL7 Patient Care Work Group.

Literatur

- [1] Health Level Seven, Inc.: <http://www.hl7.org>
- [2] Hinchley, A.: Understanding Version 3 A primer on the HL7 Version 3 Communication Standard. Köln: Verlag Alexander Mönch, 2003.
- [3] Oemig, F., Blobel, B.: HL7 Conformance: How to do proper messaging? In: Bos L. and Blobel B. (Edrs.): Medical and Care Compunetics 4, p. 298-307. Series "Studies in Health Technology and Informatics", Vol. 127, IOS Press., Amsterdam, The Netherlands , 2007<http://www.icmcc2007.net>
- [4] Deutsche Nachrichtenprofile für HL7 v2.5, <http://www.hl7.de>
- [5] VhitG-Arztbrief, <http://www.vhitg.de>
- [6] NICTIZ, Nationaal ICT Instituut in de Zorg, <http://www.nictiz.nl>
- [7] Website mit Informationen über HL7 V3 Care Provision, <http://www.zorginformatiemodel.nl>
- [8] Oemig, F., Blobel, B.: Does HL7 Go towards an Architecture Standard? In: Engelbrecht R., Geissbuhler A., Lovis Ch. Mihalas G. (Edrs.): Connecting Medical Informatics and Bio-Informatics. Proceedings of MIE 2005, 761-766, Series "Studies in Health Technology and Informatics", Vol. 116, IOS Press, Amsterdam, Berlin, Oxford, Tokyo, Washington DC, 2005
- [9] SNOMED: Systemized Nomenclature of Medicine, <http://www.snomed.org>
- [10] LOINC: Logical Observation Identifiers Names and Codes, <http://www.loinc.org> .
- [11] Score System Implementation Guide, 13th draft, <http://www.hl7.org>
- [12] Regenstrief Institute, <http://www.regenstrief.org> .
- [13] DICOM, <http://www.rsna.org>
- [14] IHE PCC (Patient Care Coordination) Technical Framework, Version 3.0, http://www.ihe.net/Technical_Framework/index.cfm#pcc
- [15] Schematron: A language for making assertions about patterns found in XML documents, <http://www.schematron.com/>
- [16] Protégé: The Protégé Ontology Editor and Knowledge Acquisition System, <http://protege.stanford.edu>
- [17] ICD: International Classification of Diseases, <http://www.who.org> or <http://www.dimdi.de>
- [18] DCM: Detailed Clinical Information Models, <http://detailedclinicalmodels.org>

Biomedizinische Ontologien für die Praxis

M. Brochhausen¹, B. Smith^{1,2},

¹Institute of Formal Ontology and Medical Information Science, Saarland University, Germany,

²Department of Philosophy and New York State Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, USA

Hintergrund: Biomedizinische Ontologien existieren unter anderem zur Integration von klinischen und experimentellen Daten. Um dies zu erreichen ist es erforderlich, dass die fraglichen Ontologien von einer großen Zahl von Benutzern zur Annotation von Daten verwendet werden. Wie können Ontologien das erforderliche Maß an Benutzerfreundlichkeit, Zuverlässigkeit, Kosteneffektivität und Domänenabdeckung erreichen, um weitreichende Akzeptanz herbeizuführen?

Material und Methoden: Wir konzentrieren uns auf zwei unterschiedliche Strategien, die zurzeit hierbei verfolgt werden. Eine davon wird von SNOMED CT im Bereich der Medizin vertreten, die andere im Bereich der Biologie und Biomedizin von der OBO Foundry. Es soll aufgezeigt werden, wie die Verpflichtung zu speziellen Kriterien der Ontologieentwicklung die Nützlichkeit und Effektivität der Ontologien positiv beeinflusst, indem die Pflege der terminologischen Systeme und ihre Interoperabilität vereinfacht werden.

Ergebnisse: SNOMED CT und die OBO Foundry unterscheiden sich grundlegend in ihren Ansätzen und Zielen. Unabhängig davon kann jedoch ein allgemeiner Trend zur strengeren Formalisierung und Fokussierung auf Interoperabilität zwischen unterschiedlichen Domänen und ihren Repräsentationen beobachtet werden.

Stichwörter: Biomedizinische Ontologien, Harmonisierung, Qualitäts sicherung, SNOMED CT.

Einleitung

In der Entwicklung terminologischer Ressourcen für Anwendungen in der Biomedizin können wir vier wesentliche Einflussfaktoren unterscheiden, die hier in grober chronologischer Folge aufgelistet werden:

1. Der Einfluss von Bibliotheks wissenschaften und elektronischen

Terminologiesystemen, der sich deutlich zum Beispiel bei MeSH, einer Indizierungsressource, die von der National Library of Medicine gepflegt wird [1], zeigt.

2. Der Einfluss aus der Datenbankgestaltung und dem conceptual modeling, der durch HL7 illustriert wird [2].
3. Der Einfluss der Molekularbiologie, der sich am deutlichsten in der Gene Ontology (GO) [3] und anderen Ontologien der Open Biomedical Ontologies (OBO) Foundry [4], [5] nachverfolgen lässt.
4. Der Einfluss einer Orientierung hin zu stärkerer logischer Formalisierung. Dieser Trend zeigt sich bei den momentanen Entwicklungen in SNOMED CT und im Bereich des Semantic Web [6].

Im Folgenden befassen wir uns mit den Fortschritten, die im Hinblick auf die Punkte 3) und 4) gemacht wurden. Mehr und mehr können Ontologien und verwandte terminologische Ressourcen von ihren Vorgängern durch folgende Faktoren unterschieden werden:

- Die Beschäftigung mit Aspekten der Interoperabilität von Ontologien, die für inhaltlich zusammenhängende, doch unterschiedliche Domänen geschaffen werden.
- Die Bereitschaft, kohärente Verfahren zur Qualitätssicherung von Ontologien zu entwickeln, die auf dem Feedback durch Nutzer und auf empirischen Tests beruhen.
- Die Bereitstellung von Techniken zur Pflege von Ontologien, die die Evolution von Ontologien gemäß dem wissenschaftlichen Fortschritt garantieren.
- Ein zunehmender Grad an Formalisierung, der durch Fortschritte im Bereich der Entwicklung von Softwareanwendungen zur Pflege,

Validierung und Verknüpfung von Ontologien beschleunigt wird.

- Eine gesteigerte Sorgfalt bei der akkurate Darstellung biologischer Sachverhalte und damit der Realität insgesamt, die durch die terminologischen Komponenten des repräsentationalen Artefakts, z. B. der Ontologie, abgebildet werden.

Damit ist das Folgende aufs Engste verbunden:

- Die Bereitschaft genauer zwischen Daten und Informationen auf der einen Seite und Objekten in der klinischen Realität, die durch Daten und Informationen abgebildet werden, auf der anderen zu unterscheiden.

Einige Aspekte hiervon haben wir bereits ausführlich in [7] besprochen, wobei der Fokus dort auf das Problem der Harmonisierung von Ontologien und Terminologien gelegt wurde. Hier konzentrieren wir uns auf zwei unterschiedliche Strategien zur Erhöhung der Nützlichkeit von Ontologien bzw. Terminologien. Für diese beiden Strategien werden als Beispiele SNOMED CT und die OBO Foundry angeführt.

Vor allem existieren biomedizinische Ontologien, um der Integration von klinischen und experimentellen Daten zu dienen, sind die Kriterien für deren Erfolg im Grunde die gleichen, die für ein Telefonnetzwerk gelten: Wichtig ist die Menge der Nutzer. Ein Netzwerk, das nur wenige Kunden hat, ist ein Misserfolg, unabhängig davon wie gut die verwendeten Technologien sind. In diesem Artikel werden die beiden oben genannten Strategien auf ihre Anlagen hin untersucht, die Nutzer für die terminologischen Ressourcen gewinnen und nachhaltig daran binden zu können.

1. Anwachsen der formalen Strenge biomedizinischer Ontologien

Eine wichtige Tendenz bei der Entwicklung von Ontologien ist der gesteigerte Wert, der formallogischen Verfeinerungen von biologischen Terminologien und verwandten Artefakten beigemessen wird. Sicherlich gibt es immer noch Zweifel in einigen Kreisen, wenn es um die Frage geht, ob verstärkte formale Strenge ein Gewinn für medizinische Ontologien ist. Manche Parteien behaupten, dass medizinisches Wissen zu intuitiv ist (Medizin als Kunst, nicht als Wissenschaft) und außerdem in einem zu großen Maße von subjektiven Erfahrungen und örtlichen Traditionen abhängt, um wissenschaftsbasierte, einheitliche Terminologien oder Ontologien schaffen zu können.

Aus der Entwicklung der molekularen und genetischen Medizin und den damit verbundenen Technologien entsteht die Notwendigkeit, Daten aus verschiedenen Disziplinen, die sich auf unterschiedliche Granularitätsebenen beziehen, zu integrieren. Diese Entwicklungen werden sicherlich dazu beitragen, dass solche Argumente langsam aber sicher als überholt erkannt werden.

Die Entwicklung medizinischer Ontologien und Terminologien sieht sich zunehmend mit Fragen über formale Methoden konfrontiert, was z.B. in der Arbeit der Semantic Web Health Care and Life Sciences Interest Group [8] deutlich wird, aber ebenso bei der Entwicklung einer description logic Infrastruktur für Vokabulare wie GALEN [9], den National Cancer Institute Thesaurus [10] und von grundsätzlicher Bedeutung für diese Arbeit den neusten Versionen des SNOMED Vocabulary [11].

Auf Seiten der Biologie haben die zunehmende Bedeutung der Bioinformatik und die enorme Steigerung der Masse der den Forschern zur Verfügung stehenden Daten zu bemerkenswerten Anstrengungen darin geführt, formal stärkere Ontologieressourcen zu entwickeln, die jetzt in der Arbeit des National Center for Biomedical Ontology [12] zusammengefasst werden.

Der Erfolg der Gene Ontology führte dazu, dass viele Biologen das OBO Format (ehemals das GO Format) [13] als Repräsentations-sprache für die Ontologieentwicklung bevorzugten. Das OBO Format ist hinsichtlich formaler Möglichkeiten ebenfalls wachsender Aufmerksamkeit unterworfen. Das Ziel lautet, Wege für neue Möglichkeiten algorithmischen Reasonings zu finden, die sowohl die biomedizinische Forschung als auch die klinische Versorgung unterstützen. Die OBO Foundry stellt 53 Ontologien für die biologischen und biomedizinischen Disziplinen zur Verfügung, von denen 49 im OBO Format vorliegen, zusammen mit einer großen Menge an Daten, die mit ihrer Hilfe annotiert wurden, inklusive über 11 Millionen frei zugänglicher Annotationen, die Produkte von Genen (Proteine und funktionale RNA) mit Ausdrücken in der GO in Verbindung setzen [14]. Außerhalb der Biologie ist die Web Ontology Language (OWL) die bevorzugte Repräsentations-sprache der Ontology Community. OWL ist ein W3C Standard, und seine algorithmischen Eigenschaften machen es zu einem attraktiven Werkzeug für die Softwareentwicklung. Allerdings ist der ontologische Gehalt, der im ursprünglichen OWL Format repräsentiert werden kann mitunter sehr unausgereift. Dank der Entwicklung eines OBO-OWL Konverters sind ontologische Inhalte von OBO jetzt ebenfalls in auf OWL basierenden Anwendungen möglich. Die OBO Ontologien und die damit verbundenen Annotationen dienen also jetzt als wichtiges Medium für die Ausweitung des Semantic Web in den Gebieten der Lebenswissenschaften.

2. Wachsende Konzentration auf Biologische Realität

Während Informatiker sich früher vornehmlich mit Daten beschäftigten, ist ein zweiter Trend bei der jetzigen Entwicklung von Ontologien die steigende Beschäftigung mit der durch diese Daten abgebildete Wirklichkeit. Ontologien werden nach wie vor häufig von Personen aus dem Knowledge Management erstellt. Dadurch wurden Ontologien zunächst hauptsächlich als Möglichkeit zur Abbildung von Information, mitunter unabhängig von einer zugrundeliegenden

Wirklichkeit interpretiert. Jetzt aber, wo Ontologien zunehmend von Biologen und Kliniken für ihre eigenen fachspezifischen Bedürfnisse geschaffen werden, tritt die Frage nach der genauen Abbildung der Wirklichkeit immer stärker in den Vordergrund.

Diese neue Fokussierung auf die biologische Realität seitens der Bioontologen wird unter anderem durch die letzten Revisionen des SNOMED CT Vokabulars [11], [18], [19] und durch die Entwicklung der OBO Foundry deutlich. Diese stehen für eine Verschiebung von der frühen Beschäftigung mit Daten und Informationen um ihrer selbst Willen hin zu etwas, was wir das realistische Paradigma nennen.

Letzteres basiert auf einer Trennung zwischen drei Ebenen der Realität:

- Mentale Repräsentationen (Ideen oder Gedanken in unserem Bewusstsein, die bestimmte Teile der Realität abbilden).
- Repräsentationale Artefakte (einschließlich Ontologien, Textbücher, usw.), die entwickelt werden, um unsere mentalen Repräsentationen für andere zugänglich zu machen.
- Die Realität selbst, die den Gegenstand unserer mentalen und physikalischen Repräsentationen, vor allem in der Wissenschaft, darstellt.

Für einen Realisten hängt der Erfolg bei der Entwicklung einer Ontologie davon ab, die Trennung dieser drei Ebenen einzuhalten [20]. Diese Trennung der Ebenen impliziert, dass die Realität, für deren Abbildung unsere Repräsentationen entwickelt werden, unabhängig von diesen Repräsentationen existiert. Der konzeptualistische Standpunkt, der unter Informationswissenschaftlern immer noch verbreitet ist, sieht Ontologien als Repräsentationen von Begriffen, die, kurz gesagt, Einheiten von Wissen (oder Bedeutungen) im Bewusstsein von Menschen darstellen. Realisten definieren eine Ontologie im Gegensatz dazu folgendermaßen: Repräsentation der Typen von Entitäten, die in einem bestimmten Feld der Wirklichkeit existieren und deren Beziehungen zueinander.

Typen sind die Invarianten oder Muster in der Welt, die von den einzelnen Wissenschaften durch Experimente (auch klinische Experimente) ausgemacht werden. Die Beziehung zwischen Typ und Einzelding ist die der Instantiierung. Typen werden an unterschiedlichen Orten und zu unterschiedlichen Zeiten in unterschiedlichen Partikularien oder Einzeldingen instantiiert und normalerweise mit allgemeinen Ausdrücken wie „Hund“ oder „Diabetes“ bezeichnet [20]. „Hund“ ist der Name des Typs, der von meinem Hund Fido und von deinem Hund Rover instantiiert wird. Da Ontologien für Realisten repräsentationale Artefakte sind, die entwickelt werden, um wissenschaftliche Untersuchungen oder ähnliche Unternehmungen zu unterstützen, sind diese mehr auf allgemeine Gesetzmäßigkeiten und daher eher auf Typen als auf Instanzen bezogen.

SNOMED CT: Die Tendenz hin zu stärkerem Realismus zeigt sich bei SNOMED CT an der aktuellen Deaktivierung von Konzepten, die als „not otherwise specified“ (NOS, nicht weiter spezifiziert) qualifiziert werden, zum Beispiel „Brain injury NOS (disorder)" (26286008.) Bereits Cimino [21] hat herausgestellt, das solche Qualifikationen wie „NOS“ Probleme bereiten müssen. Von einem realistischen Standpunkt aus können diese Probleme darauf zurückgeführt werden, dass es keine solche Entität wie eine „nicht weiter spezifizierte Gehirnverletzung“ gibt; und zwar weder als Typ noch als Instanz. Wenn es zum Zeitpunkt t eine solche Entität geben würde und es würde zu einem späteren Zeitpunkt t' eine Spezifikation dieser Entität erfolgen, so müssten wir entweder davon ausgehen, dass die ursprüngliche Entität durch den schlichten Akt der Aufnahme dieser neuen Information zerstört würde, oder dass diese Entität mit der geplanten neuen Nachfolgeentität ungleich ist.

Wie bereits in [22] ausgeführt, referieren „NOS“ und ähnliche Terme nicht auf eine bewußtseins-unabhängige Realität, vielmehr handelt es sich um konfus formulierte Repräsentationen unseres Wissens über die Realität. Selbstverständlich ist es für biomedizinische Informationssysteme unerlässlich die

fortschreitenden Veränderungen, denen unser Wissensstand, etwa bei einer medizinischen Untersuchung, unterworfen ist, aufzuzeichnen, zum Beispiel im Hinblick auf die Zuverlässigkeit der Information. Es müssen also beide Arten von Information gesammelt werden: Informationen über die Realität und Informationen über unser Wissen über die Realität. Wenn aber Kodierungsschemata algorithmisches Reasoning in dem Umfang, in dem es in der Biomedizin in Zukunft erforderlich sein wird, unterstützen sollen, dann ist es ebenso wichtig zwischen diesen beiden Informationsarten deutlich zu unterscheiden.

Ihre Entwickler haben inzwischen einige Schritte in die richtige Richtung unternommen, aber SNOMED CT enthält immer noch Konzepte wie „unknown living organism“ (SNOMED: 89088004) und „presumed viral agent“ (SNOMED: 106551006). Diese werden in SNOMED behandelt, als seien es spezifische Arten von Organismen bzw. Viren, tatsächlich handelt es sich aber in solehen Fällen um Repräsentationen eines Wissensstandes.

Zudem enthält SNOMED CT einige Terme, wie zum Beispiel „Abszess“, die zweifach eingetragen sind, zum einen als Befund, zum anderen als morphologische Anormalität:

128477000 Abscess (disorder),

44132006 Abscess (morphologic abnormality).

In solchen Fällen muss es zu Problemen kommen, wenn SNOMED CT als Kodierungsschema genutzt wird, denn es gibt hinsichtlich dieser beiden Terme keinen eindeutigen Unterschied in ihrer Referenz [23], [24].

OBO Foundry: Die Gene Ontology und ihre Schwesterontologien im OBO Repository wiesen eine Reihe antirealistischer Komponenten auf, die zu strukturellen Problemen in der GO führten, wie in [25], [26] dargestellt.

Die Gründung der OBO Foundry führte zu einer Reihe tiefgreifender Reformen dieser Ontologien. Zum Hauptziel der Initiative wurde eine akkurate Darstellung des

biologischen Gegenstandsbereiches, und, damit verbunden, wurde ein realistischer Standpunkt übernommen.

Jede Ontologie, die Mitglied der OBO Foundry ist, enthält strukturierte Repräsentationen der Typen, die in einer spezifizierten Domäne der Realität existieren. Ziel ist es, die Typen gemäß der besten wissenschaftlichen Theorie zum gegenwärtigen Zeitpunkt abzubilden. In diesem Sinne ist jede Ontologie selbst analog zu einer wissenschaftlichen Theorie: sie hat einen vereinheitlichten Gegenstandsbereich, der aus Entitäten besteht, die unabhängig von der Ontologie existieren. Weiterhin wird durch die Ontologie an einer Optimierung der deskriptiven oder repräsentationalen Adäquatheit hinsichtlich dieses Gegenstandsbereiches in den Grenzen, die durch formale Strenge und Computerisierbarkeit gesetzt werden, gearbeitet.

Ein einfaches Beispiel für die Reformen, denen sich die Ontologien der OBO Foundry unterwerfen mussten, ist der Verzicht auf Aussagen wie:

1. biological process is_a Gene Ontology, die bislang gemeinsam mit biologisch zutreffenden Aussagen wie
2. cell development is_a biological process bestanden.

Aussagen wie 1. zeigen die Verwendung der *is_a* Relation in Zusammenhängen, in denen sie ungefähr „ist eine terminologische Komponente von“ bedeuten soll. Dies ist jedoch inkonsistent mit der Hauptverwendung dieser Relation in der GO, die die Bedeutung „hat Untertyp (-klasse)“ hat. Aussage 2. kann in dem Moment zu Problemen führen, in dem die GO mit anderen Ontologien zusammengeführt wird, zum Beispiel mit solchen, die Informationsobjekte betreffen. Bei der Gültigkeit z.B. von:

3. Gene Ontology has_file_format RDF-XML
- kann, bedingt durch die Transitivität von *is_a*, aus 2. und 3. geschlossen werden, dass
4. cell development has_file_format RDF-XML.

Dies ergibt natürlich biologisch betrachtet keinen Sinn.

Um solche Probleme in der GO und ihren Schwesternontologien zu eliminieren, wird jetzt das Prinzip angewandt, dass alle Aussagen in den Ontologien Relationen verwenden, die eindeutig definiert sind, wie dies in den Definitionen der Relationen in der OBO Relation Ontology musterhaft zu sehen ist [27], [28]. Die Anwendung dieses Prinzips stellt sicher, dass die Bedeutung der Begriffe in Relationen in allen Ontologien die gleiche ist, was die logisch schlüssige Integration solcher Ontologien für *Reasoning*-Anwendungen erleichtert.

Ein weiteres Anzeichen für die höhere Gewichtung, die dem Realismus zukommt, ist die Aufnahme des Foundational Model of Anatomy (FMA) in die OBO Foundry. Ziel des FMA war es vom Anfang seiner Entwicklung an, von einem realistischen Standpunkt aus die Repräsentationen von Typen und Relationen vorzunehmen, die notwendig sind, um die Strukturen des Säugetierkörpers, im Besonderen des menschlichen Körpers, computerisierbar zu hinterlegen, und zwar in einer Form, die auch für Menschen verständlich ist [29], [30].

Die Formulierung von Definitionen im FMA, die auch in der gesamten OBO Foundry angenommen wurde, folgt Prinzipien, die erstmalig von Aristoteles niedergelegt wurden. Die Definition einer repräsentationalen Einheit „A“ in einer Ontologie, die auf einen Typ A referiert, muss durch die Referenz auf die Oberklasse von A, nämlich B, in der gleichen Ontologie und eines unterscheidenden Kriteriums C, dass von allen und nur von A erfüllt wird, formuliert werden. Aus diesem Prinzip resultieren Definitionen der Form:

ein A = ein B, das C erfüllt.

Diese Methode ist in der Biologie jedem aus dem Bereich der Taxonomie bekannt, da sich hier nach aristotelischem Muster der Speziesname aus dem Genusnamen einer *Differentia* zusammensetzt, z.B. *Homo sapiens* als Bezeichnung der Spezies von Mitgliedern der Gattung *Homo*, die vernunftfähig sind.

Beispiele hierfür aus dem FMA sind:
 a cell = def. an anatomical structure which consists of cytoplasm surrounded by a plasma membrane
 a plasma membrane = def. a cardinal cell part that surrounds the cytoplasma

Durch das Durchhalten dieser Strategie wird nicht nur die Konsistenz der Verwendung von Begriffen unterstützt, sondern sie trägt auch zur Schlüssigkeit der *is_a* Hierarchie bei.

3. Modularität biomedizinischer Ontologien

SNOMED CT und die OBO Foundry verfolgen zwei unterschiedliche Strategien gerade auch im Hinblick auf Modularisierung. Diese Strategien unterscheiden sich sowohl in ihren theoretischen, wie auch in ihren praktischen Konsequenzen.

SNOMED CT: Die ambitionierteste Initiative im Bereich der Vereinheitlichung biomedizinischer Begriffssysteme wurde gerade von der International Health Terminology Standards Development Organization (IHTSDO) übernommen [31]. Diese versucht SNOMED CT als internationale Hauptterminologie für den klinisch-medizinischen Bereich zu etablieren, die die Gesamtheit dieser Domäne in einer Vielzahl von Sprachen abdeckt. Die Basis für dieses Vorhaben ist die englische Version von SNOMED CT, die mehr als 315 000 Konzepte beinhaltet und bereits teilweise in andere Sprachen übertragen worden ist. Die Hauptstränge von SNOMED CT sind in *Description Logics* darstellbar [32], wenn auch nicht in einer der W3C Standard *description logics*. Weiterhin wird an *Mappings* für andere existierende Standards, zum Beispiel die International Classification Of Disease (ICD) 10 der WHO, gearbeitet. Der zentrale Vorteil von SNOMED CT liegt in der umfassenden Abdeckung des Gegenstandsbereiches, die durch 21 Hierarchien, die zum Beispiel Klinische Befunde, Prozeduren, Substanzen, Sozialen Kontext etc. garantiert werden soll.

Das Ziel all dieser Bemühungen ist es für den gesamten Bereich der klinischen Medizin eine einzige terminologische

Ressource bereitzustellen. Die Verwendung einer solchen „Master“-Terminologie würde zahlreiche positive Effekte zeitigen, darunter weniger redundante Daten, bessere Durchführbarkeit von longitudinalen Studien und Meta-Analysen, sowie die Sicherstellung der Konsistenz von Patientendaten über die gesamte Lebensspanne hinweg. Die Existenz einer einzigen Ressource ermöglicht die Benutzung gemeinsamer Tools und die Durchführung gemeinsamen Trainings. Trotz all dieser Vorteile, die vor allem der klinischen Entscheidungsfindung und der Datensammlung im Rahmen von Public Health Programmen zu Gute kommen würden, ist es schwer die notwendigen finanziellen Mittel zu mobilisieren, zumal SNOMED CT gut dokumentierte Schwachstellen enthält [33].

Die Tatsache, dass immer noch wenige Institutionen des Gesundheitswesens SNOMED CT zum Kodieren klinischer Daten verwenden, deutet darauf hin, dass die Initiative noch weiter investieren muss, um die Vorteile auch tatsächlich einzufahren [34]. Es steht weiterhin die Frage im Raum, inwieweit die hohen Kosten, die die Implementierung und Nutzung einer hochentwickelten Terminologie wie SNOMED CT im Klinikbereich verursacht, zu amortisieren sind. SNOMED CT zum Beispiel ist nicht wie viele andere Systeme entwickelt worden, um die Rechnungsstellung zu simplifizieren. Letzteres ist jedoch ein zentraler Aspekt des Krankhausmanagements, der aber von wesentlich weniger komplexen Terminologien abgedeckt werden kann und auch wird, die zudem schon sehr weit verbreitet sind.

Wie bereits erwähnt, ist auch die Existenz interner Fehler einer der Gründe für die langsame Ausbreitung von SNOMED CT. Es gibt Lücken in der Terminologie, Fehler in der Strukturierung und Probleme mit der Konsistenz, die aus der Anwendung unterschiedlicher Prinzipien bei der Entwicklung unterschiedlicher Zweige der Terminologie herrühren. In letzter Zeit wurden Beweise dafür, dass diese Schwächen effizientes Kodieren verhindern, publiziert [23], [24]. Diese Qualitätsprobleme werden inzwischen von den Kuratoren angegangen.

Ein Einwand ganz anderer Art zielt auf die Frage, ob die Strategie einer kompletten Abdeckung des Gegenstandsbereiches in einer Ressource richtig ist oder nicht.

OBO Foundry: Im Gegensatz zu SNOMED CT verfolgt OBO einen modularen Ansatz. Die Strategie, die der OBO Foundry [5] zugrunde liegt, ist eine Reihe von interoperablen Ontologien bereitzustellen, die unterschiedliche, nicht überlappende Domänen der biomedizinischen Realität auf der Basis von gemeinsamen Prinzipien der Ontologieentwicklung abbilden. Nur Ontologien, deren Entwickler diese Prinzipien anwenden, können in die OBO Foundry aufgenommen werden. Es gibt zwei Hauptziele der Prinzipien: zum einen sollen sie die Harmonisierung existierender Domänenontologien vorantreiben, zum anderen sollen sie als Leitfaden zur Entwicklung neuer Ontologien und als Beispiel guter wissenschaftlicher Praxis für Neulinge in der Ontologieentwicklung dienen. Immer mehr fallen im biomedizinischen Bereich Teilbereiche auf, für die völlig neue Ontologien entwickelt werden müssen. Leitlinien sind gerade deshalb besonders wichtig, um die Interoperabilität mit Ontologien benachbarter Bereiche sicherzustellen.

Das Ziel ist es, eine Situation herbeizuführen, in der die Abdeckung von Wissensbereichen durch terminologische Ressourcen schrittweise über wachsende Domänen der biomedizinischen Realität vergrößert wird. Diese Strategie ist selbstverständlich das exakte Gegenteil der Strategie, die von SNOMED CT verfolgt wird.

Aus Sicht der Foundry ist das Ziel von Ontologieentwicklung nicht, eine einzige große Ontologie bereitzustellen, die alles abdeckt, was überhaupt existiert. Vielmehr gilt es bessere Wege zu finden kleine Domänenontologien zunächst getrennt zu entwickeln, um so die Vorteile der Arbeitsteilung und die Bildung von domänenspezifischen Kompetenzteams zu fördern. Im Hinblick auf die Harmonisierung, die einen unabdingbaren zweiten Schritt darstellt, muss ein gewisses Maß an zentraler Organisation stattfinden, um die Einhaltung von allgemein anerkannten Leitlinien

sicherzustellen. Die Tatsache, dass Versuche unternommen werden, Harmonisierungen von oben durchzusetzen, zeigt, dass Harmonisierungsbestrebungen evolutionär aufgesetzt werden müssen und auch die Leitlinien von Zeit zu Zeit einer Revision unterworfen werden sollten. Die Dokumentation der Leitlinien und ihrer Entwicklung ist jedoch eine wichtige Voraussetzung dafür, auf die Bedürfnisse unterschiedlicher Interessengruppen einzugehen zu können und somit ihre Unterstützung zu gewinnen.

Die Gruppen, die die OBO Foundry vorantreiben, stimmen darüber ein, dass eine minimale gemeinsame *Upper Ontology* ausreichend ist, um Harmonisierung zu ermöglichen. Eine solche Ontologie kann dazu beitragen von Beginn an Grundlagen einer *best practice* der Ontologieentwicklung zu realisieren. Diese Vorteile kann eine Upper Ontology allerdings nur bringen, wenn sie selbst nach gut fundierten Prinzipien entwickelt und hinreichend getestet wurde. Ist dies der Fall, dann tritt als weiterer positiver Effekt bei der Benutzung einer Upper Ontology erhebliche Zeit- und Ressourcenersparnis hinzu, denn es muss nicht mehr in die Repräsentation der Strukturen der Metaebene investiert werden, die für Fachwissenschaftler unerlässlich sind um ihre Ontologie zu organisieren. Zudem gehen Upper Ontologies in der Regel auf Prinzipien zurück, die außerhalb der Erfahrung von Fachwissenschaftlern liegt.

Die Basic Formal Ontology (BFO), die als Kern des Upper Levels der OBO Foundry Ontologien dient, basiert (ebenso wie SUO und DOLCE) auf der grundlegenden Unterscheidung von continuants und occurrents. Erstere sind Entitäten, die über die Zeit hinweg existieren und Bestand haben. Sie bleiben mit sich selbst identisch, auch wenn sie sich auf unterschiedliche Weisen verändern. Letztere ereignen sich, was bedeutet, dass sie neben den räumlichen Dimensionen auch die vierte, zeitliche haben. Occurrents (zum Beispiel Prozesse) entwickeln sich in der Zeit in einer Art, die es erlaubt, sie in zeitliche Bestandteile oder Phasen zu unterteilen. Sie haben einen Anfang, eine Mitte und ein Ende. Continuants hingegen (zum Beispiel Organismen) existieren die gesamte Zeit

in ihrer Gänze, wenn sie überhaupt existieren. Gleichwohl verlieren sie kontinuierlich Teile, oder erhalten sie dazu im Rahmen von Entwicklung und Wachstum [35], [36].

Die Benutzung einer Upper Ontology fördert die Harmonisierung von Ontologien, indem sie eine einheitliche und kohärente Methodik zur Wirklichkeitsrepräsentation auf der höchsten Organisationsebene vorgibt. Selbstverständlich finden sich die Terme, die für die Praxis von vorrangiger Bedeutung sind, auf den unteren Ebenen. Kriterien, wie sie von einer Upper Ontology nahe gelegt werden, vereinfachen jedoch die Organisation dieser unteren Level immens [37], wobei zu berücksichtigen gilt, dass die Benutzung einer Upper Ontology die Nützlichkeit der Informationsressource insgesamt für den Endbenutzer nicht einschränken darf.

Neben diesen qualitätsbezogenen Aspekten widmet sich die OBO Foundry auch Fragen der Motivation. Ontologieentwicklung ist ein kostspieliges Unternehmen, zumal, um ein optimales Ergebnis zu erzielen, auch die Domänenexperten viel Zeit aufwenden müssen, um die Arbeit der Ontologen zu unterstützen. Die bisherigen Erfahrungen zeigen, dass Experten bereit sind, sich im Rahmen der Foundry zu engagieren, da sie, gerade vor dem Hintergrund der Arbeitsteiligkeit und Modularität, direkten Einfluss auf die Ressourcen haben, die sie in der Zukunft verwenden werden [38]. Insofern ähnelt die Strategie der OBO Foundry der in der open source Softwareentwicklung [39].

Ergebnisse

Sowohl IHTSDO als auch die OBO Foundry zielen darauf, eine vollständige Repräsentation der klinischen Realität zu schaffen. IHTSDO hat den Vorteil einer großen, einheitlichen Organisation, die erhebliche finanzielle Ressourcen mobilisieren kann und einem systematischen Plan folgt.

Die OBO Foundry, für die die klinische Realität nur eine Untermenge der biomedizinischen Realität ist, hat den Vorteil einer Strategie, der Mobilisierung von Fachexperten und einer großen, etablierten Benutzergruppe, letztere allerdings vor allem im Bereich der Biologie und noch nicht so sehr unter den Klinikern. Es ist noch zu früh endgültig zu beurteilen, welche der beiden Strategien die anstehenden dringlichen Probleme lösen wird. Aber diese Probleme Benutzerfreundlichkeit, Reliabilität, Kosteneffektivität und größtmögliche Abdeckung der Domäne müssen gelöst werden, wenn die effektive Integration biomedizinischer Daten gesichert werden soll.

Danksagung

Barry Smiths Beitrag zu dieser Arbeit wurde teilweise durch das National Center for Biomedical Ontology, dass durch das National Institut of Health finanziert wird, im Rahmen der NIH Roadmap for Medical Research, Grant 1 U 54 HG004028 gefördert. Informationen über die National Centers for Biomedical Computing können unter <http://nihroadmap.nih.gov/biomedicine> abgerufen werden.

Literatur

- [1] <http://www.ncbi.nlm.nih.gov/sites/entrez?db=mesh>
- [2] <http://www.hl7.org>
- [3] <http://www.geneontology.org>
- [4] <http://obofoundry.org>
- [5] Smith B., Ashburner M., Rosse C., et al.: The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*; 25 (11), 2007, pp. 1251-1255.
- [6] Ruttenberg A., Clark T., Bug W., et al.: Advancing translational research with the Semantic Web. *BMC Bioinformatics*, 8, 2007.
- [7] Smith B., Brochhausen M.: Establishing and Harmonizing Ontologies in an Interdisciplinary Health Care and Clinical Research Environment. In B. Blobel, P. Pharow, M. Nerlich, eds. "eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge, IOS Press, Amsterdam, 2008, pp: 219-234.
- [8] <http://www.w3.org/2001/sw/hcls>
- [9] <http://www.opengalen.org>
- [10] <http://ncitterms.nci.nih.gov/NCIBrowser/Dictionary.do>
- [11] <http://www.ihtsdo.org/snomed-ct>
- [12] <http://bioontology.org>
- [13] <http://www.geneontology.org/GO.format.shtml>
- [14] <http://www.ebi.ac.uk/GOA>
- [15] <http://www.w3.org/2004/OWL>
- [16] <http://www.berkeleybop.org/ontologie>
- [17] Moreira D. A., Musen M. A.: OBO to OWL: a Protégé tab to read/save OBO ontologies. *Bioinformatics*, 2007 Jul; 23(14), pp. 1826-70.
- [18] Ceusters W., Spackman K.A., Smith B.: Would SNOMED CT benefit from Realism-Based Ontology Evolution? In American Medical Informatics Association 2007 Annual Symposium Proceedings, Biomedical and Health Informatics: From Foundations to Applications to Policy, J.M. Teich, J. Suermondt, C. Hripcak, Eds. Chicago IL, 2007, pp. 105-109.
- [19] Schulz S., Suntisrivaraporn B., Baader F., Boeker M.: SNOMED reaching its adolescence: Ontologists' and Logicians' Health Check. *Int J Med Inform*, 2008; Sep, 10 [Epub ahead of print].
- [20] Smith B., Kusnierzyc W., Schober D., Ceusters W.: Towards a Reference Terminology for Ontological Research and Development in the Biomedical Domain. KR-MED 2006.
- [21] Cimino J.J.: Desiderata for controlled medical vocabularies in the Twenty-First Century. *Methods Inf Med*; 37(4-5), 1998, pp.394-403.
- [22] Bodenreider O., Smith B., Burgun A.: The ontology-epistemology divide: A case study in medical terminology. *FOIS (Formal Ontology and Information Systems)*, 2004, pp. 185-95.
- [23] Andrews J.E., Richesson R.L., Krischer J.: Variation of SNOMED CT coding of clinical research concepts among coding experts. *j Am Med Inform Assoc* 2007, 14, 4, p. 497-506.
- [24] Chiang M.F., Hwang J.C. , Yu A.C., Casper D.S. , Cimino J.J., Starren J.: Reliability of SNOMED-CT coding by three physicians using two terminology browsers. *AMIA 2006 Symposium Proceedings*, 2006, p. 131-135.
- [25] Smith B., Williams J., Schulze-Kremer S.: The Ontology of the Gene Ontology. In *Biomedical and Health Informatics: From Foundations to Applications, Proceedings of the Annual Symposium of the American Medical Informatics Association*, Washington DC, November 2003, pp. 609-613.
- [26] Smith B., Köhler J., Kumar A.: On the Application of Formal Principles to Life Science Data: A Case Study in the Gene Ontology. In *Proceedings of DILS 2004 (Data Integration in the Life Sciences)*, (Lecture Notes in Bioinformatics 2994), Berlin: Springer, 2004, pp. 79-94.
- [27] <http://www.obofoundry.org/ro>
- [28] Smith B., Ceusters W., Klagges B., et al.: Relations in Biomedical Ontologies. *Genome Biology*, 6:R46, 2005.
- [29] <http://sig.biostr.washington.edu/projects/fm>
- [30] Rosse C., Mejino J.V.L.: A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform*. 2003, 36, pp. 478-500.
- [31] <http://www.ihtsdo.org>
- [32] Donnelly K.: Multilingual documentation and classification. In B. Blobel, P. Pharow, M. Nerlich, eds. "eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge, IOS Press, Amsterdam, 2008, pp: 235-44.
- [33] <http://www.hiww.org/smcs2006/talks/Rector.ppt>
- [34] <http://www.chcf.org/documents/chronicdisease/SantaBarbaraLessonsLearned.pdf>
- [35] <http://www.ifomis.org/bfo>
- [36] Grenon P., Smith B., Goldberg L.: Biodynamic Ontology: Applying BFO in the Biomedical Domain. In *Ontologies in Medicine*, D.M. Pisanielli, Ed. Amsterdam, IOS Press, 2004, pp. 20-38.
- [37] Smith B.: From Concepts to Clinical Reality: An Essay on the Benchmarking of Biomedical Terminologies. *Journal of Biomedical Informatics* 39(3), 2006, pp. 288-298
- [38] Smith B.: Ontology (Science). In C. Eschenbach and M. Gruninger (eds.), *Formal Ontology in Information Systems. Proceedings of FOIS 2008* Amsterdam: IOS Press, 21-35.
- [39] Weber S.: The Success of Open Source. Cambridge, MA: Harvard University Press, 2004.



EJBI 2009 ISSN 1801 - 5603

European Journal for Biomedical Informatics

Volume 5 (2009), Issue 1

Greek version

www.ejbi.eu



Content

Greek version

- el 2 - 9** Εξασφαλίζοντας την συνέχεια της κατ' οίκον περίθαλψης καρδιοαναπνευστικών παθήσεων: Εξοπλισμός επιτήρησης και λογισμικό ανταλλαγής ιατρικών δεδομένων μέσω σημασιολογικά εμπλουτισμένων υπηρεσιών του Ιστού
B. Spyropoulos, A. Tzavaras, M. Botsivaly, K. Koutsourakis

Εξασφαλίζοντας την συνέχεια της κατ' οίκον περίθαλψης καρδιοαναπνευστικών παθήσεων: Εξοπλισμός επιπήρησης και λογισμικό ανταλλαγής ιατρικών δεδομένων μέσω σημασιολογικά εμπλουτισμένων υπηρεσιών του Ιστού

B. Spyropoulos¹, A. Tzavaras¹, M. Botsivaly¹, K. Koutsourakis¹

¹Medical Instrumentation Technology Department, Faculty of Engineering, Technological Educational Institution (TEI) of Athens, Athens, Greece

Περίληψη: Ένα σημαντικό μέρος της φροντίδας της σχετικής με τις καρδιοαναπνευστικές παθήσεις παρέχεται κατ' οίκον, συνήθως αλλά όχι αποκλειστικά, μετά από την έξοδο του ασθενούς από το νοσοκομείο. Έχουμε αναπτύξει ένα ολοκληρωμένο σύστημα, που υποστηρίζει τη συνέχεια της κατ' οίκον περίθαλψης των καρδιοαναπνευστικών παθήσεων, που περιλαμβάνει: Κατ' αρχάς, ένα πρωτότυπο φορητό σύστημα παρακολούθησης, βασισμένο σε φορητό υπολογιστή (laptop), που αποτελείται από χαμηλού κόστους διαθέσιμα στο εμπόριο εξαρτήματα, και επιτηρεί την κατάσταση του ασθενούς, μέσω περιοδικών ή συνεχών ελέγχων ζωτικής σημασίας παραμέτρων, δεύτερον, λογισμικό που υποστηρίζει την λήψη ιατρικής απόφασης σχετικά με τυχόν Ταχυκαρδία και Κοιλιακό Ινιδισμό, καθώς επίσης και λογισμικό βελτιστοποίησης του κατ' οίκον μηχανικού αερισμού ασθενών, βασισμένου σε κανόνες ασαφούς λογικής (Fuzzy-rules), τρίτον, ένα χαρακτηριστικό αρχείο συνέχειας της φροντίδας του ασθενούς (Continuity of Care Record, CCR), που προσαρμόζεται για να υποστηρίζει επίσης τη δημιουργία ενός σχεδίου κατ' οίκον φροντίδας και τέλος, μια πρωτότυπη οντολογία, που βασίζεται στη κλινική αρχιτεκτονική εγγράφων HL7 (Clinical Document Architecture, CDA), που χρησιμεύει ως η βάση για την ανάπτυξη των σημασιολογικά εμπλουτισμένων Διαδικτυακών Υπηρεσιών, που επιτρέπουν την ανταλλαγή και την ανάκτηση δεδομένων σχετικών με την κατ' οίκον φροντίδα. Ο ευέλικτος σχεδιασμός του υλικού και ο προσαρμόσιμος μηχανισμός ανταλλαγής δεδομένων του συστήματος, συνιστούν ένα χρήσιμο και σύμφωνο με τα ισχύοντα ή αναδύομενα διεθνή πρότυπα εργαλείο, για την

κατ' οίκον φροντίδα των πασχόντων από καρδιοαναπνευστικές παθήσεις.

Λέξεις κλειδιά: Κατ' οίκον νοσηλεία, καρδιοαναπνευστικές παθήσεις, αρχείο συνέχειας της φροντίδας του ασθενούς, σχέδιο κατ' οίκον νοσηλείας, σημασιολογικά εμπλουτισμένες υπηρεσίες του Ιστού, οντολογίες, CCR, HL7, CDA, POCT.

1. Εισαγωγή

Όπως έχει ήδη αρχίσει να γίνεται αντιληπτό, ο 21ος αιώνας φαίνεται να παρέχει ένα αλλαγμένο περιβάλλον επαγγελματικής δραστηριότητας, ενώ έχει αρχίσει παράλληλα να διαμορφώνεται και ένας νέος τρόπος αλληλεπίδρασης ιατρού-ασθενούς διαμορφώνεται, μέσω της σύντηξης της Βιοϊατρικής Τεχνολογίας, των συστημάτων Ιατρικής Πληροφορικής και των Διαδικασιών Λήψης Ιατρικών και Επιχειρησιακών Αποφάσεων σε ένα ενιαίο σύνολο. Η προσαρμογή των μεθόδων λήψης ιατρικών αποφάσεων και επεξεργασίας [1], [2], [3], [4], [5] στο αναδύομενο υψηλής τεχνολογίας περιβάλλον της κατ' οίκον νοσηλείας και η κατάρτιση και ο εξοπλισμός του ιατρού, με τις κατάλληλες συσκευές και το αντίστοιχο ιατρικό και διαχειριστικό λογισμικό, είναι ένας ουσιαστικός όρος, προκειμένου να εξασφαλιστεί, πρώτον, βιώσιμη υψηλής ποιότητας υγειονομική περίθαλψη του γηράσκοντος πληθυσμού, δεύτερον, αποκατάσταση και βοήθεια για ομάδες ατόμων με ειδικές ανάγκες και τέλος, η απαραίτητη ψυχοσωματική υποστήριξη του πληθυσμού [6].

Ο σκοπός της παρούσας εργασίας είναι να παρουσιάσουμε τα τεχνικά μέσα που έχουμε αναπτύξει, προκειμένου να υποστηριχθεί η προσαρμογή και η

συνέχεια της περίθαλψης των καρδιοαναπνευστικών παθήσεων στο σπίτι. Ένα σημαντικό και συνεχώς αυξανόμενο κομμάτι της φροντίδας της σχετικής με τις καρδιοαναπνευστικές ασθένειες παρέχεται ήδη στο σπίτι συνήθως, αλλά όχι αποκλειστικά, μετά από την έξοδο ενός ασθενούς από το νοσοκομείο. Για τον λόγο αυτό αναπτύξαμε ένα ολοκληρωμένο σύστημα, που υποστηρίζει την συνέχεια της περίθαλψης των καρδιοαναπνευστικών παθήσεων στο σπίτι που περιλαμβάνει: Πρώτον, ένα πρωτότυπο σύστημα παρακολούθησης βασισμένο σε φορητό υπολογιστή που αποτελείται από χαμηλού κόστους διαθέσιμα στο εμπόριο εξαρτήματα, και επιτηρεί την κατάσταση του ασθενούς, μέσω περιοδικών ή συνεχών ελέγχων ζωτικής σημασίας παραμέτρων, δεύτερον, λογισμικό που υποστηρίζει την λήψη ιατρικής απόφασης σχετικά με τυχόν Ταχυκαρδία και Κοιλιακό Ινιδισμό, τρίτον ο χαρακτηριστικό αρχείο συνέχειας της φροντίδας του ασθενούς (Continuity of Care Record, CCR), που προσαρμόζεται για να υποστηρίξει επίσης τη δημιουργία ενός σχεδίου κατ' οίκον φροντίδας και τέλος, μια πρωτότυπη οντολογία, που βασίζεται στη κλινική αρχιτεκτονική εγγράφων HL7 (Clinical Document Architecture, CDA), η οποία χρησιμεύει ως η βάση για την ανάπτυξη των σημασιολογικά εμπλουτισμένων Διαδικτυακών Υπηρεσιών, που επιτρέπουν την ανταλλαγή και την ανάκτηση δεδομένων σχετικών με την κατ' οίκον φροντίδα.

Ο ευέλικτος σχεδιασμός του υλικού και ο προσαρμόσιμος μηχανισμός ανταλλαγής δεδομένων του συστήματος, συνιστούν ένα χρήσιμο και σύμφωνο με τα ισχύοντα ή αναδύομενα διεθνή πρότυπα εργαλείο, για την κατ' οίκον φροντίδα των πασχόντων από καρδιοαναπνευστικές παθήσεις και θα παρουσιαστεί λεπτομερώς στα ακόλουθα μέρη αυτής της εργασίας.

2. Η κατ' οίκον

Καρδιοαναπνευστική Διαγνωστική

Οι *in vivo* διαγνωστικές διαδικασίες, οι οποίες βασίζονται στην απαγωγή και καταγραφή ηλεκτρικών και μη Βιοσημάτων, σχετίζομενων με συγκεκριμένες λειτουργίες του ανθρωπίνου σώματος, τείνουν πλέον να γίνουν συστήματα βασιζόμενα σχεδόν αποκλειστικά στη χρήση προσωπικών ή φορητών ηλεκτρονικών υπολογιστών. Στα πλαίσια αυτά, σχεδιάστηκε και αναπτύχθηκε από την ομάδα μας ένα σύστημα καταγραφής και ελέγχου, το οποίο μπορεί να χρησιμοποιηθεί σε οικιακό περιβάλλον και το οποίο επιτρέπει την συνεχή ή παροδική καταγραφή των ζωτικών σημάτων ενός ασθενούς. Η λειτουργία του συστήματος βασίζεται, αφ' ενός, στην χρήση χαμηλού κόστους εξαρτημάτων, τα οποία είναι στο σύνολό τους εμπορικά διαθέσιμα και αφ' ετέρου, στην χρήση ενός απλού οικιακού προσωπικού ηλεκτρονικού υπολογιστή. Ο εξοπλισμός του συστήματος αποτελείται από τα εξής στοιχεία:

- Μία διάταξη απαγωγής Ηλεκτροκαρδιογραφήματος (ΗΚΓ), η οποία είναι ιδιοκατασκευή, και είναι εξ ιπλίσμενη με την ηλεκτροκαρδιογραφία του Η/Υ και την επιχειρησιακή αναπνευστική συχνότητα του Η/Υ.
- Έναν ανιχνευτή μέτρησης παλμικής οξυμετρίας δακτύλου, τύπου Nellcor, ο οποίος χρησιμοποιείται για μετρήσεις Κορεσμού Οξυγόνου (SpO_2) μέσω πληθυσμογραφίας, καθώς επίσης και για τον προσδιορισμό του Καρδιακού και του Αναπνευστικού Ρυθμού.
- Μία διάταξη καταγραφής των Καρωτιδικών Ήχων, η οποία είναι επίσης ιδιοκατασκευή και χρησιμοποιείται για τον προσδιορισμό του Καρδιακού και του Αναπνευστικού Ρυθμού.

Η διάταξη απαγωγής ΗΚΓ είναι κατάλληλα σχεδιασμένη ώστε να επιτρέπει την εφαρμογή των απαγωγών κατά Eindhoven I, II και III. Η διάταξη αποτελείται από ένα προενισχυτή, ένα φίλτρο ζώνης διέλευσης στην περιοχή 0.5-150 Hz, έναν ενισχυτή, έναν αναλογικό σε ψηφιακό μετατροπέα και έναν μικροεπεξεργαστή, ο οποίος επιτρέπει την ασύρματη ψηφιακή μετάδοση του ΗΚΓ. Η διάταξη του δέκτη αποτελείται από έναν ψηφιακό δέκτη ραδιοσυχνοτήτων, ο οποίος είναι συντονισμένος στην κατάλληλη συχνότητα, και από έναν ελεγκτή, ο οποίος μετατρέπει κατάλληλα τα ψηφιακά δεδομένα βάσει του πρωτοκόλλου RS232. Η διάταξη Παλμικής Οξυμετρίας βασίζεται στην μη επεμβατική μέτρηση της απορρόφησης φωτός από το αρτηριακό αίμα, η οποία χρησιμοποιείται για τον προσδιορισμό του κορεσμού του οξυγόνου (SpO_2).

Η διάταξη καταγραφής των Καρωτιδικών Ήχων βασίζεται στην καταγραφή, με τη βιόθεια ενός στηθοσκοπίου, των ήχων που παράγονται από τη ροή του αρτηριακού αίματος. Η καταγραφή γίνεται μέσω ενός μικροφώνου, το οποίο είναι προσαρμοσμένο στο στηθοσκόπιο και συνδεδεμένο με την κάρτα ήχου του Η/Υ. Παρ' όλο που η αρχιτεκτονική της διάταξης είναι πάρα πολύ απλή και το κόστος της είναι αιμελητέο, η απόδοση του συστήματος είναι εξαιρετικά ικανοποιητική, επιτρέποντας όχι μόνο την καταγραφή των κυματομορφών του καρδιακού και του αναπνευστικού ρυθμού, αλλά επιτρέποντας επίσης πλήρη τηλε-ακροαστική εξέταση του ασθενούς, σε οιονεί πραγματικό χρόνο. Θα πρέπει βέβαια να αναφερθεί ότι η σωστή λειτουργία της διάταξης μπορεί να διαταραχθεί από τυχαίους εξωτερικούς θορύβους, όπως και στην παρακλίνια εξέταση.

Θα πρέπει τέλος να αναφερθεί ότι έχει δοκιμασθεί και η αντικατάσταση του Η/Υ ο οποίος χρησιμοποιείται από το σύστημα, από μία διάταξη απαγωγής και επεξεργασίας ΗΚΓ (Eindhoven απαγωγές I-III) βασισμένη σε μικροεπεξεργαστή (C-programmable controller Rabbit Semiconductor®). Η διάταξη αυτή είναι επίσης ιδιοκατασκευή, βασίζεται στην χρήση ενός JFET τελεστικού ενισχυτή (μοντέλο TL074 με 300 Hz συχνότητα

δειγματοληψίας), και είναι προαιρετικά εφοδιασμένη με μία σύνδεση ραδιοσυχνοτήτων της τάξης των 433 MHz μεταξύ του ενισχυτή και ενός επιπλέον εξωτερικού Η/Υ. Με την αντικατάσταση αυτή αναμένεται να επιτευχθεί περαιτέρω τόσο η σμίκρυνση, όσο και η ευελιξία χρήσης της συσκευής.

3. Συνεχής εικονική ιατρική επίβλεψη

Στο σύστημα επιτήρησης που περιγράψαμε, επιχειρούμε να προχωρήσουμε λίγο παραπέρα, όσον αφορά την εικονική ιατρική επιτήρηση ευαίσθητων και υψηλού κινδύνου πληθυσμών, όπως αυτό που υποφέρουν από καρδιαγγειακές ή/και αναπνευστικές παθήσεις. Έτσι, τα δεδομένα τα οποία συλλέγονται από το σύστημα, υφίστανται επεξεργασία ώστε να είναι σε κατάλληλη μορφή για να χρησιμοποιηθούν σε διαδικασίες υποστήριξης λήψης ιατρικής απόφασης. Τα συλλεγόμενα ακατέργαστα δεδομένα του συστήματος υποβάλλονται σε κατάλληλη επεξεργασία, για να παράγουν τα στοιχεία υποστήριξης απόφασης. Οι κυματομορφές ECG και ο καρδιακός ρυθμός (HR) εξάγονται άμεσα από το ECG-module, ενώ ο HR μπορεί επίσης να ληφθεί άμεσα από τον καρωτιδικό ανιχνευτή. Ο αναπνευστικός ρυθμός (RR) μπορεί επίσης να υπολογιστεί άμεσα από τον καρωτιδικό ανιχνευτή, ή να εξαχθεί από την ανάλυση συχνότητας και το φιλτράρισμα του αρχικού σήματος φωτοπληθυσμογραφίας (PPG). Το PPG που λαμβάνεται από την άκρη δάχτυλων εκτείνεται από 0.001Hz-6Hz, αφαιρείται το συνεχές περιεχόμενο, και η περιοχή 0.5-2.0 Hz συσχετίζεται με το μέσο καρδιακό ρυθμό. Τα στοιχεία συλλέγονται σε οιονεί πραγματικό χρόνο και υποβάλλονται σε στοιχεία πραγματικό χρόνο και υποβάλλονται σε γρήγορο μετασχηματισμό Fourier (Fast Fourier Transformation) στο πεδίο συχνοτήτων, με τη χρήση του τυποποιημένου λογισμικού Mat-Lab. Η εκτίμηση του αναπνευστικού ρυθμού και του καρδιακού ρυθμού, βασίζεται στον προσδιορισμό των τοπικών αιχμών συχνότητας στις σειρές 0.01-0.50 Hz και 0.5-2.0 Hz.

Αναφορικά με την αναγκαιότητα εφαρμογής Αυτόματης Εξωτερικής Απινίδωσης, η αναγνώριση των κυματορφών του Κοιλιακού Ινιδισμού (VF)

και της Κοιλιακής Ταχυκαρδίας (VT) για την ειδοποίηση του πλησιέστερου οικείου ή διασώστη, πραγματοποιείται στο σύστημά μας μέσω δύο τεχνικών, οι οποίες αναπτύχθηκαν από την ομάδα μας [7]. Η πρώτη τεχνική είναι αυτή της Ανάλυσης Εικόνας, σύμφωνα με την οποία το ΗΚΓ αντιμετωπίζεται ως εικόνα και διαιρείται σε έναν αριθμό Περιοχών Ενδιαφέροντος (ROI) με τη διαίρεση της εικόνας σε 20 περιοχές για τον άξονα του χρόνου και έπειτα σε τέσσερις περιοχές, που καθορίζονται από τα όρια Vmax, $\frac{1}{2}$ Vmax, 0, - $\frac{1}{2}$ Vmax και Vmax, στον άξονα πλάτους (τάσης). Η διάκριση μεταξύ των ρυθμών είναι δυνατή με τη μέτρηση της οπτικής πυκνότητας, δηλαδή τον αριθμό «πλήρων» εικονοκυττάρων (pixels) σε αυτά τα ROIs. Ο κανονικός ρυθμός έχει μια συγκεκριμένη κατανομή των οπτικών πυκνοτήτων σε αυτές τις περιοχές, ενώ οι αντίστοιχες κατανομές των ρυθμών VF και VT είναι εντυπωσιακά διαφορετικές. Αυτές οι 80 τιμές οπτικής πυκνότητας, χρησιμοποιούνται για την ταξινόμηση των εικόνων ECG από λογισμικό που αναπτύξαμε (fuzzy-neural software). Η δεύτερη τεχνική είναι η CDF-SKEW τεχνική, η οποία είναι ένας συνδυασμός δύο περιγραφικών στατιστικών συναρτήσεων, δηλαδή της συσσωρευτικής συνάρτησης κατανομής πιθανότητας (Cumulative Probability Distribution Function, CDF) και του συντελεστή εκτροπής (Coefficient of Skewness, SKEW). Αυτή η τεχνική επιτρέπει μια γρήγορη επισκόπηση για ένα σύνολο στοιχείων, χωρίς να πρέπει να εξεταστεί κάθε παρατήρηση, ή στοιχείο χωριστά. Αυτές οι λειτουργίες χρησιμοποιήθηκαν ως κριτήρια για την ταξινόμηση των καρδιακών ρυθμών ως απαιτούντων απινίδωσηή μη.

Όσον αφορά στην κατ' οίκον Μηχανική Υποστήριξη της Αναπνοής, είναι ευρέως αποδεκτό ότι η μέθοδος αυτή μπορεί να εφαρμοσθεί στην κατ' οίκον νοσηλεία ασθενών, οι οποίοι είναι εξαρτημένοι από τη χρήση αναπνευστήρα, πλην όμως βρίσκονται σε σταθερή κατάσταση [8], [9]. Η κατ' οίκον νοσηλεία των ασθενών αυτών κοστίζει κατά 70% λιγότερο από ότι η μηχανική υποστήριξη της αναπνοής σε έναν θάλαμο Εντατικής Θεραπείας, ενώ οι ασθενείς έχουν καλύτερη ποιότητα ζωής [10], διατηρούν κοινωνικές σχέσεις, και έχουν μικρότερη πιθανότητα

ενδονοσοκομειακών λοιμώξεων. Απαραίτητες προϋποθέσεις για την κατ' οίκον νοσηλεία αυτών των ασθενών αποτελούν ένα κατάλληλο οικογενειακό και οικιακό περιβάλλον, η προετοιμασία των συγγενών και ένα πλάνο νοσηλείας το οποίο θα πρέπει να είναι κατάλληλα διαμορφωμένο από ειδική ομάδα επιστημόνων.

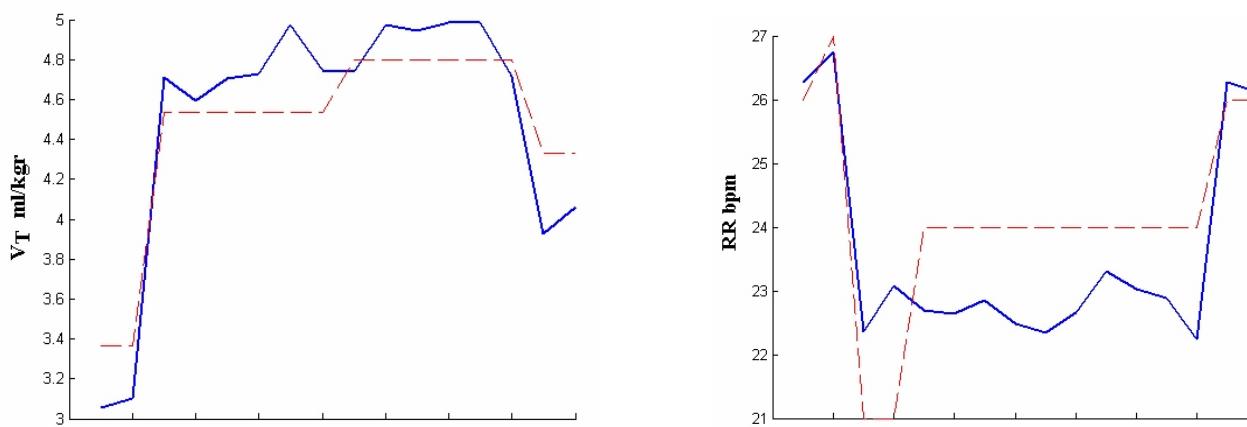
Προκειμένου να συμβάλουμε σε αυτήν την περίπλοκη προσέγγιση, αναπτύξαμε μία εφαρμογή, η οποία βασίζεται σε έναν αλγόριθμο ασφούς λογικής (Fuzzy Logic), ο οποίος είναι κατάλληλος για την διαχείριση της κατ' οίκον Οξυγονοθεραπείας και του Μηχανικά Υποστηριζόμενου Άερισμού.

Οι ασθενείς με χρόνια αποφρακτική πνευμονοπάθεια (Chronic Obstructive Pulmonary Disease, COPD) χαρακτηρίζονται από αυξημένο έργο αναπνοής (work of breathing, WOB) και από δυσλειτουργία των αναπνευστικών μυών. Ο μηχανικός εξαερισμός εφαρμόζεται για να ελαττωθεί το WOB. Οι υπόλοιποι αναπνευστικοί μύες μειώνουν την αρτηριακή μερική πίεση του διοξειδίου του άνθρακα (PCO_2) και διαχειρίζονται την υποξαιμία. Δεδομένου ότι οι αινάγκες των ασθενών δεν είναι στατικές, οι ρυθμίσεις των αναπνευστήρων πρέπει να ελέγχονται τακτικά.

Ο κορεσμός οξυγόνου (SpO_2), καθώς και ο καρδιακός και αναπνευστικός ρυθμός, όπως καταγράφονται από το σύστημα που περιγράφηκε στις προηγούμενες παραγράφους, αποτελούν τις εισόδους του ασφούς συστήματος. Η παράμετρος της μεταβολής του κορεσμού Οξυγόνου συναρτήσει του χρόνου ($d\text{SpO}_2$) χρησιμοποιείται από το σύστημα ως εσωτερικός δείκτης αποτελεσματικότητας, καθώς μπορεί να θεωρηθεί ως κεντρικός παράγοντας ενδεικτικός της κατάστασης του ασθενούς, ενώ η απόκριση του συστήματος ορίζεται από ένα σύνολο κανόνων ασφούς λογικής, οι οποίοι βασίζονται στις αρχές της φυσιολογίας της αναπνοής. Η έξοδος του συστήματος είναι στην πράξη μία πρόταση για την μεταβολή της ροής του οξυγόνου προς τον ασθενή.

Ο ελεγκτής (αλγόριθμος) ασφούς λογικής αναπτύχθηκε και εκπαιδεύθηκε για τον προσδιορισμό των κατάλληλων ρυθμίσεων παλιρροϊκού όγκου (tidal volume, VT) και συχνότητας αναπνοής (respiration frequency, RR) για τους ασθενείς με χρόνια αποφρακτική πνευμονοπάθεια (COPD) [11]. Σαράντα τρεις (43) ώρες καταγραφής μη επεμβατικά ελεγχόμενων παραμέτρων από τέσσερις (4) διαφορετικούς ασθενείς με COPD, που υφίσταντο ελεγχόμενο (control mode) μηχανικό αερισμό, συλλέχθηκαν σε δύο Γενικά Νοσοκομεία στην Ελλάδα. Ο ελεγκτής αναπτύχθηκε και εκπαιδεύθηκε, με την υιοθέτηση εξήντα τοις εκατό (60%) των διαθέσιμων καταγραμμάτων στοιχείων. Η αρχιτεκτονική του συστήματος χρησιμοποιεί τις ακόλουθες μη επεμβατικά μετρούμενες παραμέτρους: τον κορεσμό Οξυγόνου (S_pO_2), την ευενδοτότητα των πνευμόνων (lung compliance) C, την αντίσταση (resistance) R, τη μέγιστη πίεση εισπνοής (Peak Inspiratory Pressure, PIP) και την πίεση πλατώ (Plateau Pressure, PPlateau), για την πρόβλεψη κατάλληλων ρυθμίσεων VT και RR. Η αξιολόγηση εκτελέσθηκε με τα καταγραμμένα στοιχεία που δεν χρησιμοποιήθηκαν κατά τη διάρκεια της φάσης κατάρτισης του ελεγκτή. Ο ελεγκτής εξετάστηκε απέναντι στο σύνολο αξιολόγησης. Το μέσο τετραγωνικό λάθος του παλιρροϊκού όγκου VT ήταν 0.222 ml/Kg και του ρυθμού αναπνοής RR ήταν 1.21 αναπνοές ανά λεπτό (bpm) αντίστοιχα.

Η εφαρμογή της διαδικασίας επιτρέπει τον καλύτερο έλεγχο των παραμέτρων της κατ' οίκον μηχανικής υποστήριξης της αναπνοής, ενώ το σύστημα επιτρέπει επίσης την εκ του μακρόθεν ιατρική επίβλεψη της υποστήριξης της αναπνοής και την άμεση επέμβαση του ιατρού μέσω τηλεφώνου στους συγγενείς του ασθενούς, στην περίπτωση που οι τιμές αγγίζουν κάποια όρια συναγερμού. Το προς το παρόν εργαστηριακά αξιολογούμενο σύστημα μπορεί να οδηγήσει στο βελτιωμένο έλεγχο των ρυθμών ενός «οικιακού αναπνευστήρα», κάτω από κλινικές συνθήκες, και θα μπορούσε να επιτρέψει την εξ αποστάσεως ιατρική επίβλεψη και αξιολόγηση των μακροπρόθεσμων αναπνευστικών τάσεων στοιχείων,



Σχήμα 1. Μια περίπτωση των προτεινόμενων (πλήρης γραμμή) και καταγραμμένων τιμών (παύλες) του παλιρροϊκού όγκου VT (μέσο τετραγωνικό σφάλμα: 0.222 ml/Kg) και του ρυθμού αναπνοής RR (μέσο τετραγωνικό σφάλμα: 1.21 bpm).

ή των βραχυπόθεσμων. Θα πρέπει τέλος να αναφερθεί ότι η μετάδοση των δεδομένων τα οποία προκύπτουν από τις παραπάνω διαδικασίες μπορεί να γίνει είτε μέσω ασύρματων συνδέσεων, είτε μέσω modem και τηλεφωνικών γραμμών, είτε μέσω ασφαλών συνδέσεων στο διαδίκτυο, είτε, τέλος, μέσω δορυφορικών συνδέσεων. Με τον τρόπο αυτό παρέχεται στον ιατρό τόσο η δυνατότητα της επί μακρόν καταγραφής των αναπνευστικών δεδομένων όσο και η άμεση και αυτόματη ειδοποίησή του και, κατά συνέπεια, και η άμεση προσαρμογή της θεραπείας σε επείγουσες περιπτώσεις.

4. Η συνέχεια της κατ' οίκον περίθαλψης καρδιοαναπνευστικών παθήσεων

Είναι μάλλον κοινώς αποδεκτό ότι η αύξηση του προσδόκιμου επιβίωσης, σε συνδυασμό με την αλματώδη αύξηση των νοσοκομειακών δαπανών, θα οδηγήσει αναπόφευκτα σε μία μεταβολή του τρόπου παροχής της φροντίδας υγείας, ενθαρρύνοντας ιδιαίτερα κάποια αποκεντρωμένα σχήματα όπως είναι η κοινοτική ιατρική και η κατ' οίκον νοσηλεία. Παρ' όλα αυτά, υπάρχουν ορισμένα βάσιμα ερωτηματικά τα οποία αφορούν αυτά τα αποκεντρωμένα μοντέλα. Τα ερωτηματικά αυτά αναφέρονται κυρίως στην ασυνέχεια στην παρεχόμενη φροντίδα, στον κατακερματισμό των ιατρικών δεδομένων των ασθενών, καθώς επίσης και στον βέλτιστο τρόπο με τον οποίο αυτά τα μοντέλα νοσηλείας θα επιβλέπονται

ιατρικά και θα αποζημιώνονται οικονομικά.

Το ASTM, το οποίο είναι ένας Αμερικανικός Οργανισμός Θέσπισης Προτύπων, ενέκρινε πρόσφατα το πρότυπο E2369-05, το οποίο αφορά τη δημιουργία του Αρχείου Συνεχιζόμενης Φροντίδας (Continuity of Care Record - CCR) [12], [13]. Η θέσπιση του αρχείου αυτού στοχεύει στη διασφάλιση της διακίνησης ενός ελαχίστου απαραίτητου συνόλου πληροφοριών, όταν ο ασθενής παίρνει εξήτηριο από ένα Νοσηλευτικό Ίδρυμα ή όταν μεταφέρεται σε διαφορετικό Ίδρυμα ή σε διαφορετικό τμήμα του ίδιου ίδρυματος.

Επίσης, ο διεθνής οργανισμός για την τυποποίηση (ISO) έχει εγκρίνει πρόσφατα νέα πρότυπα για την επικοινωνία και τη σημασιολογική διαλειτουργικότητα ηλεκτρονικών αποσπασμάτων αρχείων υγείας [14]. Αυτό το πρότυπο είναι βασισμένο στην αρχιτεκτονική διπλού μοντέλου, όπου ένα απλό γενικό πρότυπο αναφοράς ορίζεται για την αντιπροσώπευση των δεδομένων και ένα αρχέτυπο μοντέλο εφαρμόζεται ώστε να επιτρέπει την αναπαράσταση των σύνθετων περιοχών εννοιών του ηλεκτρονικού αρχείου υγείας.

Το σύστημα που αναπτύξαμε και που περιγράφεται στην εργασία αυτή χρησιμοποιεί το πρότυπο E2369-05 (CCR) και αποτελείται από δύο ενότητες. Η πρώτη ενότητα είναι αρμόδια για τη δημιουργία ενός χαρακτηριστικού CCR που περιέχει τα κατάλληλα δημογραφικά

και διοικητικά στοιχεία, καθώς επίσης και τις σχετικές κλινικές πληροφορίες, ενώ η δεύτερη ενότητα είναι αρμόδια για τη δημιουργία ενός σχεδίου κατ' οίκον νοσηλείας (homecare plan), το οποίο θα περιληφθεί σε ειδικό τμήμα του CCR. Το σύστημα προορίζεται να χρησιμοποιηθεί για την μετάβαση ενός ασθενή από το νοσοκομείο στην κατ' οίκον νοσηλεία, αν και η πρώτη ενότητα θα μπορούσε να χρησιμοποιηθεί σε κάθε περίπτωση μετάβασης ή παραπομπής, από ένα σημείο φροντίδας σε άλλο. Η χαρακτηριστική ενότητα του CCR μπορεί, είτε να συλλέξει τα απαραίτητα στοιχεία από ένα εγκατεστημένο ήδη σύστημα EHR, είτε να επιτρέψει στο χρήστη να εισαγάγει τα στοιχεία με το χέρι, με την πλήρωση των ειδικών υποδειγμάτων (forms). Εν πάσῃ περιπτώσει, ο χρήστης αποφασίζει ποια μέρη του ιατρικού αρχείου του ασθενή είναι απαραίτητα ή σημαντικά για την περιγραφή της τρέχουσας κατάστασης της υγείας του ασθενή και πρέπει να περιληφθούν στο CCR.

Η δεύτερη ενότητα είναι αρμόδια για τη δημιουργία ενός σχεδίου κατ' οίκον νοσηλείας, σχετιζόμενο με παθήσεις του καρδιοαναπνευστικού, με τη δημιουργία ενός δομημένου υποσυνόλου των στοιχείων που περιέχει τις αναγκαίες δραστηριότητες (διαγνωστικές, επιτήρησης, θεραπείας και νοσηλείας), που πρέπει να υιοθετηθούν κατά τη διάρκεια της κατ' οίκον νοσηλείας. Το πρότυπο που αναπτύχθηκε επιτρέπει σε κάθε τμήμα νοσοκομείου ή ιατρική / νοσηλευτική ομάδα, να ορίσει ένα

κατάλληλο σύνολο δραστηριοτήτων κατ' οίκον νοσηλείας, που αντιστοιχίζεται στην διάγνωση του εξιτηρίου, με βάση την κωδικοποίηση των ομοειδών διαγνωστικών ομάδων (Diagnosis Related Group, DRGs). Πρέπει να αναφερθεί εδώ ότι το σύστημα κωδικοποίησης DRGs χρησιμοποιήθηκε επειδή προσφέρει ένα επαρκές και καλά δομημένο ιατρικό διαγνωστικό σύστημα ταξινόμησης, ενώ, συγχρόνως, επιτρέπει τη οικονομική διαχείριση των περιπτώσεων κατ' οίκον νοσηλείας, με βάση την κύρια διάγνωση που οδηγεί σε αυτή.

Τα σχεδιασμένα σύνολα δραστηριοτήτων κατ' οίκον νοσηλείας μπορούν να εκτελεσθούν στο οικιακό περιβάλλον και είναι πλήρως ατομικά διαμορφώσιμα. Κάθε χρήστης, δηλαδή, κάθε ιατρός αρμόδιος για την χορήγηση εξιτηρίου από το νοσοκομείο σε έναν ασθενή, μπορεί να οργανώσει τα κατάλληλα σχέδια για την κατ' οίκον φροντίδα του. Χορηγώντας το εξιτήριο ενός ασθενή, ο ιατρός μπορεί να χρησιμοποιήσει ένα από τα ήδη

προκαθορισμένα σύνολα δραστηριοτήτων κατ' οίκον νοσηλείας, που αντιστοιχίζεται στην διάγνωση του εξιτηρίου, με βάση την κωδικοποίηση των ομοειδών διαγνωστικών ομάδων (Diagnosis Related Group, DRGs). Πρέπει να αναφερθεί εδώ ότι το σύστημα κωδικοποίησης DRGs χρησιμοποιήθηκε επειδή προσφέρει ένα επαρκές και καλά δομημένο ιατρικό διαγνωστικό σύστημα ταξινόμησης, ενώ, συγχρόνως, επιτρέπει τη οικονομική διαχείριση των περιπτώσεων κατ' οίκον νοσηλείας, με βάση την κύρια διάγνωση που οδηγεί σε αυτή.

Το σύστημα προβλέπει επίσης την παραγωγή εντύπων ή οθονών εισαγωγής, που θα συμπληρώνονται από το νοσηλευτικό προσωπικό κατά τη διάρκεια των κατ' οίκον επισκέψεων, προκειμένου

να τεκμηριωθούν οι δραστηριότητές τους. Τα συμπληρωμένα έντυπα επιστρέφονται στον αρμόδιο ιατρό, που τα αξιολογεί και, ανάλογα με την αξιολόγησή του, μπορεί να τροποποιήσει το σχέδιο της κατ' οίκον νοσηλείας του συγκεκριμένου ασθενή με οποιοδήποτε κατάλληλο τρόπο. Η δομή και τα στοιχεία του παραχθέντος CCR συμμορφώνονται με τις προδιαγραφές ASTM E2369-05 για το αρχείο συνεχιζόμενης φροντίδας, ενώ χρησιμοποιείται η γλώσσα XML για την αναπαράσταση των δεδομένων.

Η αναπαράσταση XML γίνεται σύμφωνα με το W3C σχήμα XML, που προτείνεται από το ASTM [15]. Το CCR που παράγεται από το σύστημα αυτό ματατα μετασχηματίζεται σε μορφή HTML, χρησιμοποιώντας την γλώσσα XSL (Extensive Stylesheet Language), προκειμένου να καταστεί ορατό και εκτυπώσιμο. Πρέπει να αναφερθεί εδώ ότι οι διαγνωστικές και θεραπευτικές δραστηριότητες είναι ταξινομημένες σύμφωνα με τη διεθνή ταξινόμηση ICD-9 (International Classification of Diseases Version 9),

When a patient is discharged a DRG-code is assigned, according to the principal diagnosis. The user then has to:

The screenshot shows the 'Patient Discharge' window. At the top, it displays patient data (Surname: Γιανναδηλίδης, Name: Ιωάννης, FatherName: Γεώργιος, ID: M123301, Date Of Birth: 27/07/1970) and a discharge summary section. Below this, there's a note: 'Ο ασθενής ξρίζει κατ' οίκον νοσηλείας για διάστημα τουλάχιστον 20 ημερών'. The main area is divided into 'Home - Care Details' and 'Recommended Procedures'.

- 1. Select a home-care procedure/activity from the profile defined for the specific DRG-code**: A red callout points to the 'Available Procedures' dropdown menu, which lists various medical procedures like 'Ηλεκτροκαρδιογράφημα, στο σπίτι' and 'Ένεση υδρούρια / ενδομική : στο σπίτι'.
- 2. Determine the date for the procedure/activity to be executed**: A red callout points to a calendar showing the month of May 2005, with the date 'Today: 26/4/2005' highlighted.
- List of already recommended (scheduled) home-care procedures/activities**: A red callout points to the 'Recommended Procedures' table, which lists several scheduled procedures with their codes, dates, and statuses.

Procedure	Code	Date Scheduled	Status	Comments
Ηλεκτροκαρδιογράφημα, στο σπίτι	03.01.00.002	28/4/2005	Ordered	
Ιατρική εξέταση : στο σπίτι (ημέρα 08:00-22:00, καθημερινή)	02.01.00.005	28/4/2005	Ordered	
Ηλεκτροκαρδιογράφημα, στο σπίτι	03.01.00.002	3/5/2005	Ordered	
Ιατρική εξέταση : στο σπίτι (ημέρα 08:00-22:00, καθημερινή)	02.01.00.005	3/5/2005	Ordered	
Ένεση υδρούρια / ενδομική : στο σπίτι	02.02.00.002	7/5/2005	Ordered	

Σχήμα 2. Επιλογή δραστηριοτήτων κατ' οίκον νοσηλείας για έναν ασθενή.

ενώ τα Αυστραλιανά DRGs (Australian Refined DRGs) χρησιμοποιήθηκαν για την κωδικοποίηση των περιπτώσεων, και τέλος η ταξινόμηση νοσηλευτικών παρεμβάσεων του κλινικού συστήματος ταξινόμησης φροντίδας CCC (Clinical Care Classification) [16] χρησιμοποιήθηκε για την τεκμηρίωση των νοσηλευτικών δραστηριοτήτων.

5. Η σημασιολογικά εμπλουτισμένη υπηρεσία Ιστού

Η διαλειτουργικότητα των πληροφορικών συστημάτων υγειονομικής περίθαλψης έχει γίνει μια από τις κριτιμότερες και πλέον επίκαιρες πτυχές στην παροχή υπηρεσιών υγείας [17], [18]. Η ολοκλήρωση ιατρικών δεδομένων είναι ένα δύσκολο έργο, δεδομένου ότι τα υπάρχοντα συστήματα πληροφοριών υγείας, λειτουργούν ακόμα με απομονωμένο τρόπο. Κάθε σύστημα πληροφοριών προς το παρόν χρησιμοποιεί τα δικά του λεξιλόγια και βάσεις γνώσεων και αναπαριστά τα δεδομένα του με διαφορετικό τρόπο (format) [19]. Οι κλινικές ορολογίες και τα λεξιλόγια, όπως το SNOMED, το ICD-9, το ICD-10, και το LOINC είναι ήδη σε λειτουργία για αρκετά χρόνια και παρέχουν μια λίγο-πολύ καθιερωμένη περιγραφή της ιατρικής γνώσης. Άλλα πρότυπα υγειονομικής περίθαλψης, όπως το HL7 [20] και το openEHR [21] προσεγγίζουν το πρόβλημα της ανταλλαγής ηλεκτρονικών στοιχείων μεταξύ των διαφορετικών πληροφοριακών συστημάτων και του καθορισμού του περιεχόμενου των ιατρικών αρχείων των ασθενών. Αυτά τα πρότυπα παρέχουν έναν ορισμένο βαθμό διαλειτουργικότητας και είναι ήδη σε λειτουργία από τις πολυάριθμες οργανώσεις υγειονομικής περίθαλψης, με το HL7 (έκδοση 2) να είναι σήμερα το ευρύτερα εφαρμοζόμενο παγκόσμια πρότυπο (standard) ιατρικών πληροφοριών [18].

Πιο συγκεκριμένα, η σύμφωνη με το HL7 κλινική αρχιτεκτονική εγγράφων (Clinical Document Architecture, CDA) [22], [23] είναι ήδη σε χρήση σε διάφορες χώρες [24], δεδομένου ότι επιτρέπει μια κοινή αναπαράσταση των κλινικών εγγράφων, καθώς και την ανταλλαγή και τη διαχείριση εαυτών [25]. Η αρχιτεκτονική HL7-CDA συμβάλλει σημαντικά στη

σημασιολογική διαλειτουργικότητα, επιτρέποντας τη δομημένη χρήση ελεγχόμενων ορολογιών και προωθώντας την συλλογική κατανόηση της δομής και της σημασίας των κλινικών εγγράφων, που δημιουργούνται από διαφορετικά συστήματα πληροφοριών. Εντούτοις, το HL7-CDA δεν μπορεί να είναι μια λύση στο πρόβλημα της διαλειτουργικότητας, δεδομένου ότι είναι μη ρεαλιστικό να αναμένεται, ότι όλοι οι πάροχοι υπηρεσιών υγείας θα συμφωνήσουν στην υιοθέτηση ενός μοναδικού προτύπου, ενώ λείπει προς το παρόν μια παγκόσμια αποδεκτή ορολογία.

Η λύση για την ολοκλήρωση ιατρικών δεδομένων φαίνεται να είναι η χρήση των υπολογιστικών τεχνολογιών, που να είναι σε θέση να κατανοήσουν τη σημασία των περιεχομένων δεδομένων [17], [18], [19], [26]. Ο αναδυόμενος σεσημασμένος σημασιολογικά Ιστός, που θα χρησιμοποιεί σημασιολογικά εμπλουτισμένες υπηρεσίες του Ιστού, στις οποίες οι πληροφορίες θα έχουν μια καλά καθορισμένη και ερμηνεύσιμη από τη μηχανή έννοια, εμφανίζεται προς το παρόν ως η πιο ελκυστική προσέγγιση προς αυτήν την κατεύθυνση [26], [27]. Συγχρόνως, οι καθιερωμένες προσπάθειες τυποποιήσης, όπως τα κλινικά λεξιλόγια και τα προαναφερθέντα πρότυπα υγειονομικής περίθαλψης, δεν πρέπει να αγνοθούν.

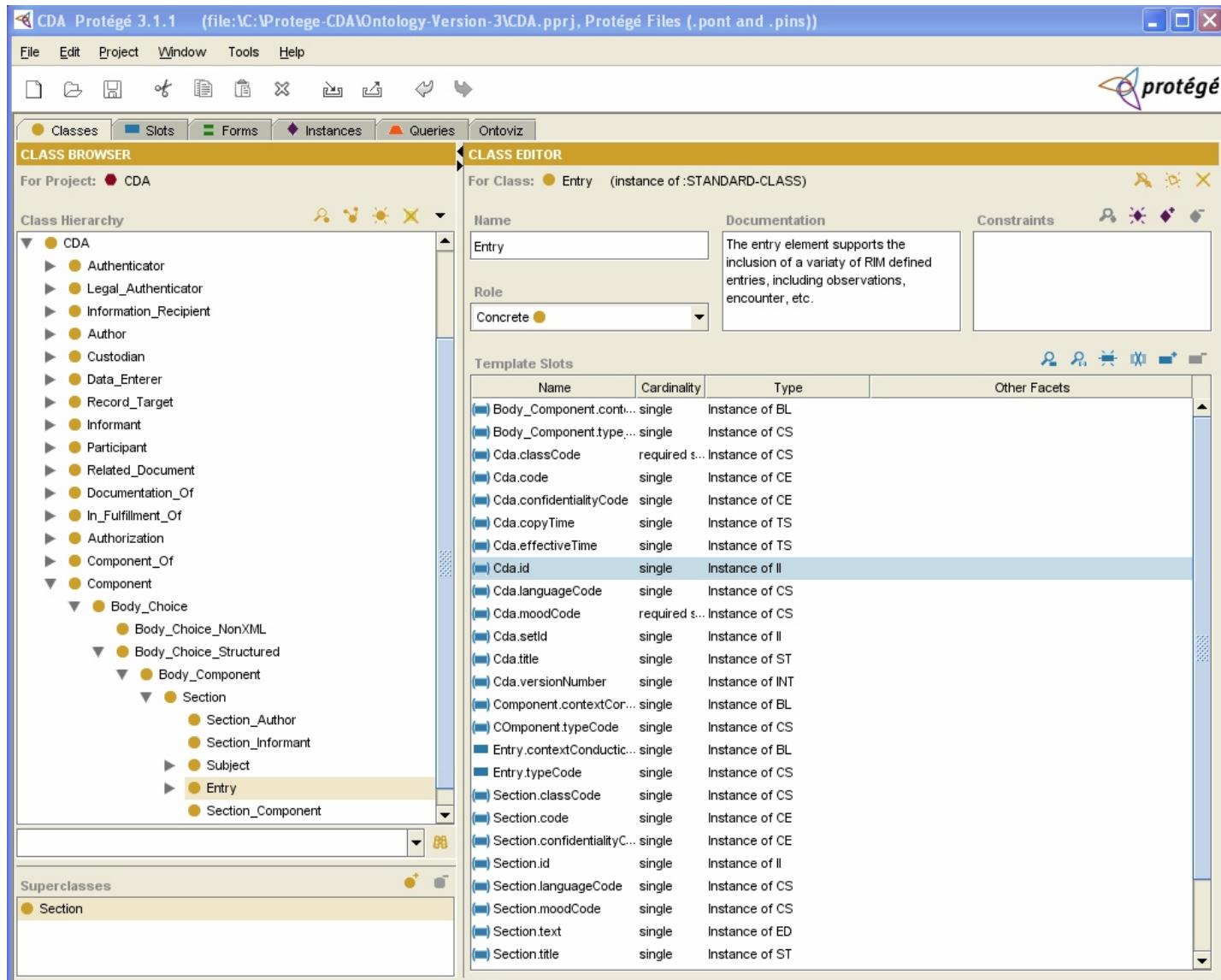
Το σύστημα που αναπτύχθηκε προσεγγίζει τα CDA-συμβατά έγγραφα ως περιοχές γνώσης, που περιγράφουν τα συγκεκριμένα γεγονότα μιας περίπτωσης, όπως, παραδείγματος χάριν, μια παραπομπή για αγγειογραφία των στεφανιάδων αγγείων. Η κατάλληλη αναπαράσταση των εννοιών αυτών των εγγράφων, από τη σκοπιά μιας οντολογίας, παρέχει την δυνατότητα αμοιβαίας κατανόησης του εγγράφου, και επιτρέπει τη δημιουργία κατάλληλα σχεδιασμένων σημασιολογικά εμπλουτισμένων υπηρεσιών Ιστού, που υπερβαίνουν τόσο τα προβλήματα των ασύμβατων τυποποιήσεων (formats) στα μηνύματα, όσο και αυτήν της χρήσης των διαφορετικών λεξιλογίων [28].

Το σύστημα που σχεδιάστηκε αποτελείται από μια πρωτότυπη οντολογία που βασίζεται στο HL7-CDA, και μια εφαρμογή

που μετατρέπει τα έγγραφα παραπομπής σε ένα CDA-συμβατό format και το περιεχόμενο των CDA-συμβατών εγγράφων σε συγκεκριμένες περιπτώσεις της οντολογίας. Μια κατάλληλα σχεδιασμένη σημασιολογικά εμπλουτισμένη υπηρεσία Ιστού αναλαμβάνει τη διανομή των εγγράφων μέσω του Διαδικτύου, με τον εντοπισμό των υπαρχουσών περιπτώσεων της οντολογίας, όταν δεχτεί τέτοιο αίτημα.

Για τους σκοπούς αυτής της εργασίας, η οντολογία παραπομπής σχεδιάστηκε, ενσωματώνοντας τα πρότυπα HL7-CDA. Η ιεραρχία της οντολογίας καθορίστηκε χρησιμοποιώντας την HL7 συμβατή μέθοδο πληροφοριών αναφοράς (HL7 Reference Information Mode, RIM) [29], τους HL7 συμβατούς τύπους δεδομένων και λεξιλογίων και την HL7-CDA R2 ιεραρχική περιγραφή (HL7-CDA R2 Hierarchical Description). Αυτές οι τέσσερις έννοιες αποτελούνται τις ανώτατες κατηγορίες (top-classes) της οντολογίας και αναλύονται περαιτέρω σε μια ιεραρχία υποκατηγοριών, οι οποίες περιγράφουν τις έννοιες που ανήκουν σε αυτές τις κύριες κατηγορίες. Τα λεξιλόγια που χρησιμοποιήθηκαν ήταν το ICD-9 και το LOINC, τα οποία εισήχθησαν στην οντολογία, ως περιπτώσεις των αντίστοιχων κατηγοριών λεξιλογίων. Μπορεί βέβαια να υιοθετηθεί οποιοδήποτε κλινικό λεξιλόγιο, όπως π.χ. το SNOMED κλπ.

Τα έγγραφα παραπομπής, περιέχουν συνήθως στατικές ή/και δυναμικές εικόνες, που εκφράζονται υπό την μορφή αντικειμένων συμμορφουμένων με την αρχιτεκτονική HL7-CDA, τα οποία δημιουργούνται σύμφωνα με τη δεύτερη έκδοση (CDA-R2) της τυποποιημένης προδιαγραφής HL7-CDA (CDA-R2) [30]. Το περιεχόμενο των συμμορφουμένων με την αρχιτεκτονική CDA εγγράφων παραπομπής μετατρέπονται σε περιπτώσεις οντολογίας. Μια κατάλληλα σχεδιασμένη σημασιολογικά εμπλουτισμένη υπηρεσία Ιστού αναλαμβάνει τη διανομή των εγγράφων μέσω του Διαδικτύου, με τον εντοπισμό των υπαρχουσών περιπτώσεων της οντολογίας, όταν δεχτεί ένα τέτοιο αίτημα.



Σχήμα 3. Η κατηγορία CDA της οντολογίας.

Η υπηρεσία που αναπτύχθηκε είναι προς το παρόν αρκετά απλή και επιτρέπει τον εντοπισμό των υπαρχουσών περιπτώσεων της οντολογίας, ανταποκρινόμενη σε ερώτηση με βάση τον προκαθορισμένο κατάλληλο αριθμό ταυτότητας ασθενούς.

6. Τελικές παρατηρήσεις

Η τρέχουσα εργαστηριακή δοκιμή του συστήματος που σχεδιάστηκε και αναπτύχθηκε δείχνει ότι είναι σε θέση να συμβάλλει ουσιαστικά στην δημιουργία ενός αποτελεσματικού, χαμηλού κόστους και βασισμένου στη σύγχρονη και διαθέσιμη στο εμπόριο τεχνολογία, συστήματος υποστήριξης της κατ' οίκον περίθαλψης καρδιοαναπνευστικών παθήσεων.

Επίσης, η χρήση των οντολογιών και των σημασιολογικά εμπλουτισμένων υπηρεσιών του Ιστού, σε συνδυασμό με την χρήση καθιερωμένων προτύπων και ιατρικών λεξιλογίων, είναι ζωτικής σημασίας, πρώτον, για την σύλληψη σε επίπεδο αρχής μιας διαδικασίας ανταλλαγής δεδομένων ιατρικής και νοσηλευτικής φροντίδας, και δεύτερον, για την ολοκλήρωση των δομημένων ιατρικών δεδομένων και άλλων στοιχείων μιας παραπομπής, σε οποιοδήποτε σύστημα, ηλεκτρονικών ή μη, ιατρικών αρχείων ασθενών. Επομένως, μπορούμε να υποθέσουμε, ότι αν και υπάρχει ακόμα ένα πολυσχιδές σύνολο προβλημάτων προς περαιτέρω διερεύνηση, υπάρχουν αρκετές ενδείξεις, ότι ο εύκαμπτος σχεδιασμός του υλικού και ο

προσαρμόσιμος μηχανισμός ανταλλαγής δεδομένων του συστήματος που παρουσιάστηκε, παρέχει ένα χρήσιμο και συμμορφωμένο με τις πλέον διαδεδομένες τυποποιήσεις (standards) εργαλείο, για την υποστήριξη της συνέχειας της κατ' οίκον περίθαλψης καρδιοαναπνευστικών παθήσεων.

Αναφορές

- [1] Meyer, G.S., Gibbons, R.V.: House calls to the elderly-a vanishing practice among physicians, *N Engl J Med* 337 (1997) 1815-1820.
- [2] Frederiks, C.M.A., te Wierik, M.J.M., van Rossum, H.J.L. et al.: Why do Elderly People seek Professional Home Care? Methodologies Compared, *Journal of Community Health* 17 (1992) 131-141.
- [3] Gilman, L.: A Doctor in the House, *Time* Sep. 24 (2006).
- [4] Spyropoulos, B., Papagounos, G.: A theoretical approach to Artificial Intelligence Systems in Medicine, *Artificial Intelligence in Medicine* 7, (1995), 455-465.
- [5] Spyropoulos, B. et al.: Application Tools supporting Decision Making in the Modern Hospital, *IJSR* 16 (2006) 129-133.
- [6] Editorial: Can House Calls Survive? *N Engl J Med* 337 (1997) 1840-1841.
- [7] Botsivaly, M., Koutsourakis K., Spyropoulos B.: A quasi real time Biosignal Acquisition processing and Transmission System. 4th European Symposium on Biomedical Engineering, Patras, Greece 2004.
- [8] Hillberg, R.E., Douglas, C.J.: Non Invasive Ventilation, *N Engl J Med* 337 (1997) 1746-1752.
- [9] Simonds, A.K.: Home ventilation, *Eur Respir J* 22 (2003) Suppl. 47 38s-46s.
- [10] Kohorst, J.: Transitioning the Ventilator-Dependent Patient From the Hospital to Home, *Medscape Pulmonary Medicine* 9 (2005) www.medscape.com/514735.
- [11] Tzavaras, Weller, P., Spyropoulos B.: A Neuro-Fuzzy Controller for the Estimation of Tidal Volume and Respiration Frequency Ventilator Settings for COPD Patients Ventilated in Control Mode, Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, August 23-26, 2007, Lyon, France.
- [12] ASTM (www.astm.org) : E2369-05, Standard Specification for Continuity of Care Record
- [13] Tessier, C.: Continuity of Care Record, ASTM E31- WG on CCR, 21st TEPR 2004 (Towards an Electronic Patient Record), May 16-18, 2005, Salt Lake City, USA
- [14] ISO 13606-1:2008. Health informatics - Electronic health record communication - Part 1: Reference model. <http://www.iso.org/>
- [15] ASTM (www.astm.org): ADJE2369-Adjunct to E2369-05 Continuity of Care Record (CCR).
- [16] Saba, V. K.: Home Health Care Classification of Nursing Diagnoses and Interventions. Washington, DC: Georgetown University, 1994.
- [17] Dogac, A., Laleci, G., Kirbas, S., Kabak, Y., Sinir, S., Yildiz, A., Gurcan Y.: Artemis: Deploying Semantically Enriched Web Services in the Healthcare Domain. *Information Systems Journal*, Vol. 31, Issues 4-5, June-July 2006, Pages 321-339
- [18] Bicer, V., Laleci, G., Dogac, A., Kabak, Y.: Providing Semantic Interoperability in the Healthcare Domain through Ontology Mapping. *eChallenges 2005*, Ljubljana, Slovenia.
- [19] Eccher, C., Purin, B., Pisanelli, D.M., Battaglia, M., Apolloni, I., Forti S.: Ontologies supporting continuity of care: The case of heart failure", *Comput Biol Med.*, Vol 36, No 7-8, pp 789-801, 2006.
- [20] Health Level 7 (HL7), <http://www.hl7.org>
- [21] The openEHR Foundation. Available from: <http://www.openehr.org/>.
- [22] Dolin, R. H.: Clinical Document Architecture", HL7 International Affiliates Joint Meeting. August 2000, Dresden, Germany.
- [23] Dolin, R.H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F.M., Biron, P.V., Shabo, A.: The HL7 Clinical Document Architecture. *J Am Med Inform Assoc.*; Vol 13, pp 30-39, 2006
- [24] Müller, M., Ückert, F., Bürkle, T., Prokosch H.: Cross-institutional data exchange using the clinical document architecture (CDA)", *International Journal of Medical Informatics*, Vol 74, pp 245-256, 2005.
- [25] Paterson, G. I., Shepherd M., Wang X., Watters C., Zitner D.: Using the XML-based Clinical Document Architecture for Exchange of Structured Discharge Summaries. *Proceedings of the 35th Hawaii International Conference on System Sciences*, 2002.
- [26] Della Valle, E., Corizza, D., Bicer, V., Kabak, Y., Laleci, G. B., Lausen H.: The Need for Semantic Web Service in the eHealth. *W3C Workshop on Frameworks for Semantics in Web Services*, 2005. [Available from: <http://www.w3.org/2005/04/FSWS/accepted-papers.html>].
- [27] Kamel Boulos, M.N., Roudsari, A.V., Carson E.R.: Towards a semantic medical Web: HealthCyberMap's tool for building an RDF metadata base of health information resources based on the qualified Dublin Core Metadata Set. *Med Sci Monit*, 2002; 8(7): MT 124-136.
- [28] Botsivaly M., Malamateniou F., Vassilacopoulos G., Macris A.: A semantic web-based system for information support to emergency healthcare delivery, Accepted for publication, *Computers in Human Behavior*, Elsevier, 2007.
- [29] HL7 Reference Information Model, http://www.hl7.org/library/data-model/RIM/modelpage_mem.htm, 2006.
- [30] Dolin R.H., Alschuler L., Boyer S., Beebe C., Behlen F.M., Biron P.V., Shabo A.: HL7 Clinical Document Architecture, Release 2. *J Am Med Inform Assoc.*; Vol 13, pp 30-39, 2006



EJBI 2009 ISSN 1801 - 5603

European Journal for Biomedical Informatics

Volume 5 (2009), Issue 1

Lithuanian version

www.ejbi.eu



Content

Lithuanian version

lt 2 - 7

Autonominio širdies veiklos reguliavimo efektyvumo vertinimas EKG P bangos pagrindinių komponenčių analizės būdu
A. Krisciukaitis, R. Simoliuniene, M. Tamasiunas, V. Saferis, A. Vainoras, L. Gargasas

Autonominio širdies veiklos reguliavimo efektyvumo vertinimas EKG P bangos pagrindinių komponenčių analizės būdu

A. Krisciukaitis¹, R. Simoliuniene², M. Tamosiunas¹, V. Saferis², A. Vainoras³, L. Gargasas³

¹Institute for Biomedical Research, Kaunas University of Medicine, Lithuania,

²Kaunas University of Medicine, Lithuania,

³Institute of Cardiology of Kaunas University of Medicine, Lithuania

Širdies veikla yra kontroliuojama vegetacinės nervų sistemos, keičiant širdies ritmo dažnį ir/arba širdies raumens susitraukimą jėgą, pagal viso kūno hemodinaminius poreikius. Reguliavimą atlieka pastoviai konkuruojančios simpatinė ir parasimpatinė nervų sistemos. Šiu mechanizmų sutrikimai sukelia ortostatinę tachikardiją ir/arba nuolatinio nuovargio sindromą. Reguliacinių mechanizmų funkcionalumo bei efektyvumo įvertinimas gali sutekti labai vertingos diagnostinės informacijos apie širdies veiklos reguliavimo sutrikimus ankstyvoje susirgimų stadijose bei padėti kontroliuoti gijimo procesus, reabilitaciją po intensyvaus gydymo. Kiekybinis EKG P-bangos formų vertinimas ortostatinio mėginio (kuris staiga sutrikdo pusiausvyrą tarp simpatinės ir parasimpatinės nervų sistemos) metu, naudojant pagrindinių komponenčių analizės pagrindu sukurtą metodą, leidžia išskirti kiekybinius širdies veiklos reguliacinių mechanizmų funkcionalumo bei efektyvumo įverčius. Šis metodas galėtų būtų naudojamas e-Sveikatos sistemoje.

Raktažodžiai: Pagrindinių komponenčių analizė, autonominis širdies reguliavimas, EKG P-banga.

1. Įvadas

Širdies veikla yra kontroliuojama vegetacinės nervų sistemos, keičiant širdies ritmo dažnį ir/arba širdies raumens susitraukimą jėgą, pagal viso kūno hemodinaminius poreikius. Simpatinis poveikis intensyvina širdies veiklą, stiprina širdies raumens susitraukimus, tuo tarpu parasimpatinės poveikis sukelia priešingus efektus. Širdies ritmas yra sąlygojamas spontaninė elektrinė aktyvumė turinčių ląstelių, daugiausia esančių sinusiniame mazge, ritmo. Šis ritmas yra pastovios sąveikos tarp simpatinio ir parasimpatinio poveikio pasekmė [1]. Reguliacinių

mechanizmų veikimo sutrikimai sukelia ortostatinę tachikardiją ir/arba nuolatinio nuovargio sindromą [2]. Kai kurios sisteminės ligos gali paveikti autonominės širdies veiklos reguliavimą. Autonominės širdies veiklos reguliavimo sutrikimai stebimi įvairiose diabetinės neuropatijos stadijose [3], [4]. Defibriliacija mėginant sustabdyti pavojingus širdies ritmo sutrikimus, neišvengiamai silpnina autonominės širdies veiklos reguliavimą [5], [6]. Tokios pat pasekmės tikėtinos ir po chirurginių intervencijų, kateterinių abliacijų ir t.t. Reguliacinių mechanizmų funkcionalumo bei efektyvumo įvertinimas gali sutekti labai vertingos diagnostinės informacijos apie širdies veiklos reguliavimo sutrikimus ankstyvoje susirgimų stadijose ir padėti kontroliuoti gijimo procesus bei reabilitacinių periodą po intervencijų.

Skirtinga simpatinių ir parasimpatinių nervų mazgų topologija širdyje [7], kaip ir skirtinė sistemų veikimo mechanizmai [8], lemia skirtinės dinamines simpatinio ir parasimpatinio širdies ritmo reguliavimo charakteristikas. Šie skirtumai leidžia identifikuoti mechanizmus iratskirai įvertinti juos, naudojant širdies ritmo variabilumo kriterijus [8]. Išsamesni tyrimai grindžiami EKG P-bangos morfologijos analize [9]. P-bangos forma įtakoja elektrinio sujaudinimo sklidimo prieširdžiuose, kuris gali būti įtakojamas širdies ritmo reguliacinių mechanizmų. Širdies elektrinis sujaudinimas kyla iš didžiausią spontaninę elektrinę aktyvumą turinčių ląstelių grupės, tikrojo ritmo vedlio, esančio sinusiniame mazge. Egzistuoja ir kitos ląstelių grupės, turinčios spontaninę elektrinę aktyvumą, esančias tiek sinusiniame mazge, tiek ir už jo. Jas aktyvuja elektrinis sujaudinimas, sklidantis nuo tikrojo ritmo vedlio. Šios ląstelių grupės vadinamos latentiniai ritmo vedliais, o didžiausia jų koncentracija

yra sinusiniame mazge [10]. Neuromediatorių išsiskyrimas neuronų terminalėse įtakoja artimiausių ritmo vedlių grupių spontaninę aktyvumą. Kaip rašoma [7], nervų rezginių topologija rodo, kad parasimpatinio aktyvumo metu neuromediatoriai yra išskiriami tikrojo ritmo vedlio vietoje ir jo spontaninis aktyvumas yra slopinamas. Taigi tikėtina, kad koks nors latentinis ritmo vedlys, kurio aktyvumas buvo mažiau įtakotas, prisiiams tikrojo ritmo vedlio vaidmenį. Skirtingas atskirų sinusinio mazgo bei prieširdžio dalių ritmo vedlių ja utrumas parasimpatiniams poveikiui [11], patvirtina šią idėją apie egzistuojantį ritmo vedlio migracijos mechanizmą sinusiniame mazge ar net už jo ribų. Šis reiškinys jau buvo nagrinėtas [12], [13]. Ritmo vedlio judėjimas lemia specifinius širdies ritmo pokyčius [14], [15], [16], [17], [11]. Prieširdžiais sklinda elektrinio sujaudinimo frontas, prasidėdamas nuo tikrojo ritmo vedlio. Kintanti tikrojo ritmo vedlio vieta įtakos elektrinio sujaudinimo sklidimą bei atispindės EKG P-bangos morfologijos pokyčiais [9]. Kuo didesni yra tikrojo ritmo vedlio poslinkiai, tuo ryškesni pokyčiai tikėtini P-bangos morfologijoje. Tačiau tik pirmasis P-bangos trečdalnis atspindi elektrinio sujaudinimo fronto sklidimą dešiniajame prieširdyje. Sujaudinimo sklidimo pokyčiai dešiniajame prieširdyje gali nebūtinai įtakoti sklidimą kairiajame prieširdyje, tuo pačiu ir likusios P-bangos dalies formą. Todėl stebimi P-bangos morfologijos pokyčiai suteikia mums labai ribotas galimybės sukurti kokius nors paprastus empirinius kriterijus kiekybiniam vertinimui. Suskaitmenintos EKG P-bangos formą atspindi signalo atskaitymai. Tokio atvaizdavimo dimensiškumas yra lygus atskaitymų skaičiui vektoriuose (priklasomai nuo diskretizacijos dažnio, jis gali būti lygus šimtui ar daugiau).

Pagrindinių komponenčių analizės metodas (PCA) mažina atvaizdavimo dimensiškumą išlaikydamas didžiausią pradinių duomenų variacijos dalį [18]. Mes siūlome šį metodą naudoti P-bangos morfologijos pokyčius atspindinčiu požymiu išskyrimui. PCA transformuoja pradinių daugiamatių atvaizdavimą į naujų koordinacių sistemą naudojant vektorius (pagrindines komponentes), kurie yra tarpusavyje ortogonalūs ir koncentruoja savo informaciją, atspindėtą keletu tarpusavyje koreliuotų kintamujų pradiniuose duomenyse. Gautosios pagrindinės komponentės suskaičiuojamos taip, kad pirmosios jų atspindi didžiausią pradinių duomenų variacijos dalį. Taigi, yra įmanoma sumažinti atvaizdavimo dimensiškumą, naudojant tik keletą pirmųjų pagrindinių komponenčių. Kiekvienas P-bangos vektorius x_i yra atvaizduojamas tiesine pagrindinių komponenčių ūk, ir jų koeficientų $w_{i,k}$ kombinacija:

$$x_i = \sum_{k=1}^p w_{i,k} \alpha_k.$$

Koeficientų įvairovė $w_{i,k}$ atspindi formos įvairovę P-bangų masyve. Vadinas, yra įmanoma sukonstruoti kiekybinį kriterijų, kuris atspindėtų P-bangos pokyčius. Šiame darbe mes parodome kaip P-bangos pokyčiai, sukelti ortostazės mėginiu (kuris staiga sutrikdo pusiausvyrą tarp simpatinio ir parasympatinio širdies reguliavimo), galėtų būti kiekybiškai įvertinti, naudojant pagrindinių komponenčių analizės metodą. Šis metodas galėtų būti naudojamas kaip e-Sveikatos sistemos dalis, gaunant vertingą diagnostinę informaciją apie širdies veiklos reguliacinių mechanizmų funkcionalumą bei efektyvumą. Mūsų preliminarūs tyrimo rezultatai buvo paskelbti [19].

2. Metodai

EKG signalai buvo registruojami ortostazės testo metu, atlanko, remiantis įprastu protokolu. Testas buvo pradėtas pacientui gulint horizontalioje padėtyje bei nejudant 3 minutes, po to pacientas atsistoja į vertikalią padėtį ir turi 2 minutes ramiai stovėti, po to vėl paguldomas. Pirmają kontrolinę grupę sudarė 20 sveikų abiejų lyčių savanorių studentų, kurių amžius buvo tarp 20 ir 24 metų. Antrają grupę

sudarė 10 vyresnio amžiaus, abiejų lyčių pacientų, kurių amžius buvo nuo 48 iki 70 metų. Trečiojoje grupėje buvo 10 vyresnio amžiaus, abiejų lyčių pacientų, kurių amžius buvo nuo 48 iki 70 metų, tačiau jie buvo patyrę laikinus prieširdžių virpėjimo periodus.

Dvylikos derivacijų EKG signalai buvo sinchroniškai registruojami 12 bitų skiriamaja geba ir 500 Hz diskretizacijos žingsniu, naudojant EKG analizės sistemą, vadintam "Kaunas-Krūvis" [20]. Analizei atrinkti tik tokie signalai, kurie neturėjo ekstrasistolinių dūžių, prieširdžių virpėjimo periodų, ar kitų ritmo sutrikimų. Analizavome tik geriausios kokybės po vieną kiekvieno iš 40 tirtų asmenų EKG įrašą. Viso analizuota 40 įrašų.

R-bangos maksimumas buvo pasirinktas kaip atraminis kiekvieno kardiociklo taškas. Jis nustatytas, remiantis tokia procedūra: R-banga buvo preliminariai fiksuota ten, kur pirma EKG išvestinė viršijo 0,5 savo maksimalios reikšmės slenkančiame 3 sekundžių lange antrojoje derivacijoje. Atraminis kardiociklo taškas rastas, naudojant tokią funkciją:

$$a_i = \sum_{j=1}^{12} x_{i,j}^2,$$

kur $x_{i,j}$ yra i-tasis atskaitymas j-toje EKG derivacijoje. Funkcija yra apskaičiuojama, naudojant 80 stebėjimų langą, esančią po preliminariai rastos R-bangos. Funkcijos a maksimumas buvo pažymėtas kaip atraminis kardiociklo taškas. R-R intervalai buvo skaičiuojami, remiantis atraminiais taškais. Septyniasdešimt atskaitymų, pradedant 80-uoj, esančių prieš atramini kardiociklo tašką, buvo laikomi P-bangos atskaitymais. Mes nusprendėme nenustatinėti P-bangos pradžios ir galo taškų ir tikėjomės, kad P-banga bet kuriuo atveju pateks į šį intervalą galimų P-Q intervalo pasikeitimų metu ir kad tokie pasikeitimai taip pat atispindės kaip P-bangos morfologijos kaita.

Visi P-bangos atskaitymų vektoriai iš kiekvienos iš 12 EKG derivacijų (po 70 atskaitymų kiekvienas), buvo sujungti nuosekliai į vektorių, iš viso turintį 840 atskaitymų. Gautas vektorius aprašė vieno kardiociklo P-bangą. Visi po 840 atskaitymų turintys vektoriai sudarė

matricą X, aprašančią visas P-bangas vieno EKG įrašo ortostazės mėginių metu:

$$X = \begin{matrix} x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ \dots & \dots & x_{i,j} & \dots \\ x_{p,1} & x_{p,2} & \dots & x_{p,n} \end{matrix}$$

Čia $x_{i,j}$ yra i-tasis j-tojo kardiociklo atskaitymas. p-visada buvo 840 tuo tarpu n kito nuo 120 iki 150 priklausomai nuo ortostazės testo metu registratorių kardiociklų skaičiaus. Taigi, P-bangos forma kiekvieno kardiociklo metu buvo aprašoma 840 parametrų (atskaitymų), kurių reikšmės ortostazės mėginių metu kito, kartu atspindėdamos P-bangos formos pokyčius. Pagrindinių komponenčių analizė (PCA) buvo panaudota šio atvaizdavimo dimensiškumo mažinimui. Suskaičiuotos pirmosios pagrindinės komponentės aprašo didžiausią pradinių duomenų variacijos dalį. Todėl vieno kardiociklo P bangų vektorius x atvaizduojamas naudojant tik keletą pirmųjų pagrindinių komponenčių, suskaičiuotų iš viso kardiociklų rinkinio. Ši procedūra yra žinoma kaip Karhunen-Loeve transformacija [21]. Bazines funkcijas (pagrindines komponentes) mes skaičiavome kaip kovariacinės matricos Rx nuosavus vektorius:

$$R_x = E[X \cdot X^T].$$

Kovariacinė matrica skaičiuojama, naudojant MatLab™ funkciją "COV", kuri, atėmus vidurkį iš kiekvieno matricos stupelio, apskaičiavo matematinę viltį E.

Nustatant minimalų, tačiau pakankamą, pagrindinių komponenčių skaičių, kuris būtų optimalus P-bangos atskaitymų matavimų skaičiaus sumažinimui, naudotas kriterijus, randamas pagal prognozuojamą skirtumų kvadratų sumą PRESS (PREDiction Sum of Squares) [18]:

$$PRESS(m) = \sum_{i=1}^n \sum_{j=1}^p (\hat{x}_{ij} - x_{ij})^2,$$

kur \hat{x}_{ij} yra originalaus signalo įvertis, naudojant ne visas, bet tik pirmąsias m pagrindinių komponenčių, x_{ij} originalaus signalo atskaitymai. Tokia procedūra, taikoma dideliems duomenų rinkiniams, reikalauja didelių skaičiavimo resursų. Yra parodyta [22], kad dideliems duomenų rinkiniams $PRESS(m)$ vėl beveik lygi

$$\sum_{k=m+1}^p l_k$$

, kur l_k - tikrinė reikšmė, atitinkanti tikrinj vektorių \hat{O}_k . Kadangi mūsų turimi duomenų rinkiniai yra palyginti dideli (daugiau nei 100 kardiociklų, turinčiu po 840 atskaitymų ir suskaičiuotų 840 pagrindinių komponentinių), mes naudojome šią, supaprastintą procedūrą. Norėdami nuspresti, ištraukti ar ne m -tajį iš eilės einantį tikrinj vektorių (pagrindinę komponentę) į atvaizdavimą, remémés kriterijumi, pasiūlytu Wold [23]:

$$PR(m) = \frac{PRESS(m)}{PRESS(m-1)}.$$

Deja, griežtas pagrindinių komponenčių ištrauktin mažinant dimensiškumą, skaičiaus nustatymas tebelieka problema. Ankstesniuose mūsų darbuose [24] siūloma minimalų pakankamą komponentinių skaičių pasirinkti tokį, ties kuriuo vizualiai stebimas specifinis trūkio taškas PR kriterijaus (6) grafike. Šį pasirinkimą patvirtina ir tai, kad būtent tokio skaičiaus pirmųjų pagrindinių komponenčių koeficientų kaita buvo stebima sinchroniškai su specifine širdies ritmo kaita ortostatinio mēginio metu. Kadangi pagrindinių komponenčių (kurios yra ortonormalios) koeficientų pokyčiai kiekybiškai atspindi P-bangos formos pokyčius, tai skirtumas tarp pradinių koeficientų reikšmių ramioje horizontalioje padėtyje ir reikšmių ortostazės metu ar stovimoje padėtyje kvadratų suma kiekybiškai atspindi P-bangos pokyčius šio testo metu. Kriterijų suskaičiavome naudodami (7) formulę:

$$C_j = \sum_{i=1}^k (w_{i,j} - w_{i,j\text{init}})^2,$$

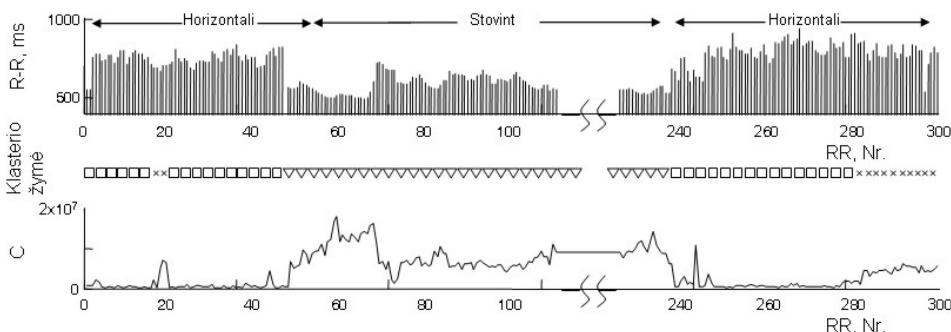
kur $w_{i,j}$ yra i-osios j-tojo kardiociklo pagrindinės komponentės koeficientas,

$w_{i,\text{init}}$ yra pradinis i-osios bazinės funkcijos koeficientas, apskaičiuojamas kaip šio koeficiente vidurkis pirmųjų 20 kardiociklų metu. k -minimalus tačiau pakankamas bazinių funkcijų skaičius, kuris nustatomas pagal PR kriterijų.

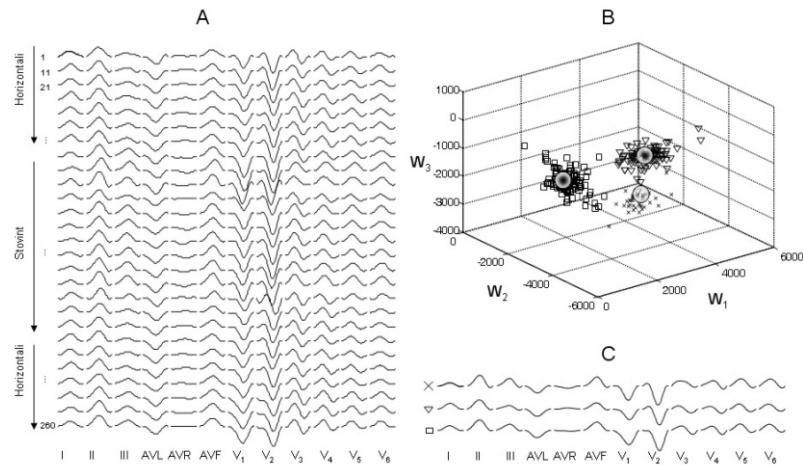
PCA procedūra buvo atskirai atlikta kiekvieno įrašo P-bangos duomenų matricai.

3. Rezultatai

Visų pacientų įrašuose buvo stebėti įvairaus dydžio specifiniai R-R intervalų pokyčiai ortostatinio mēginio metu. P-bangos formos pokyčiai ortostatinio mēginio metu dažnai buvo pastebimi ir vizualiai. Viršutinis grafikas 1 pav. iliustruoja tipinius R-R intervalo pokyčius testo metu. 2A pav. parodyta visų 12 EKG derivacijų P-bangos formos įvairovė. Aiškumo dėlei paveiksle pavaizduotos tik kas 10-to kardiociklo P-bangos. Testo laiko skalė yra parodyta kairiajame krašte. Vizualūs, tik maži P-bangos formos pokyčiai gali būti stebimi stabilioje horizontalioje ir stabilioje stovimoje padėtyse. Esminiai P-bangos formos skirtumai yra stebimi, lyginant kardiociklų formas horizontalioje ir stovimose padėtyse.



Pav.2. Tipinio įrašo visų 12 derivacijų P-bangų formų įvairovė (kad 10-tas kardiociklas parodytas A), trimatis P-bangos formos vaizdas, naudojant pirmų trijų pagrindinių komponenčių koeficientus (B), P-bangos formos apskaičiuotos, naudojant koeficientų klasterių svorio centru koordinates (C).



Kriterijaus PR (6) reikšmės praktiškai visuose įrašuose turėjo trūkio tašką ties 7-a pagrindine komponente. Todėl tolimesnei analizei naudojome pirmąsias 7 pagrindines komponentes.

Kriterijaus C reikšmės (7) kiekybiniam P-bangos formos pokyčių įvertinimui yra pavaizduotos 1 paveiksle po R-R intervalais. Pažymėti šio kriterijaus reikšmių pokyčiai vizualiai koreliuoja su kūno padėties pokyčiais ortostazės mėginių metu.

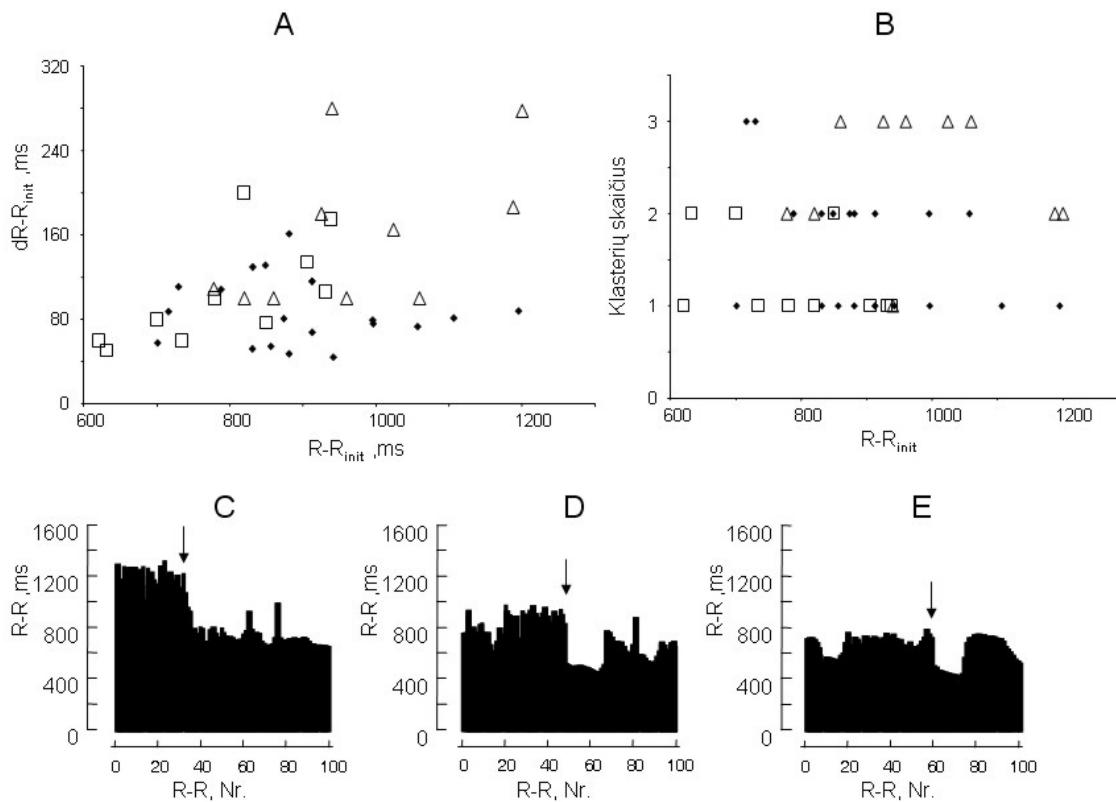
Suskaičiuoti pagrindinių komponenčių koeficientai kiekybiškai atvaizduoja kiekvienos P-bangos formą kaip tašką n-matėje erdvėje, kur n yra panaudotų pagrindinių komponenčių skaičius. Tai iliustruota 2B pav. Tačiau, atsižvelgiant į vizualizavimo galimybes, yra atvaizduoti tik pirmi 3, o ne visi 7 šiam įrašui panaudoti koeficientai. Vizualiai galime nustatyti mažiausiai tris šių taškų klasterius. Hierarchinis klasterizavimo metodas, naudojantis Euklidio atstumą (MatLab funkcija CLUSTERDATA),

nustatė tris klasterius. Klasterių identifikavimui buvo patikrinta statistinė hipotezė apie duomenų tolygų pasiskirstymą [25].

Jiems priklausantys taškai yra pažymėti kryžiukais, trikampiais ir kvadratais. Klasterių, kuriems priklauso P-bangos įverčiai, pavadinimai yra parodyti 1 paveiksle po R-R intervalais. Kaip matome, daugumos P-bangų, registruotų horizontalioje padėtyje, formos įverčiai priklauso klasteriui, pažymėtam kvadratais. P-bangų, registruotų vertikalioje pozicijoje, formos įverčiai daugiausia priklauso klasteriui, pažymėtam trikampiais. Keleto P-bangų, registruotų įrašo pabaigoje horizontalioje padėtyje, formos įverčiai priklauso klasteriui, pažymėtam kryžiukais. Visi maži P-bangos formos svyravimai apie tris pagrindines pozicijas pažymėtam kryžiukais. Visi maži P-bangos formos svyravimai apie tris pagrindines pozicijas formuoja tam tikrą n-matę kūną (debesj), apimantį visus taškus, priklausančius kiekvienam klasteriui.

Tokių n-maćių kūnų centrus atitinkančios koeficientų reikšmes buvo suskaičiuotos, panaudojant MatLabTM programe realizuotų dirbtinių neuroninių tinklų pagalba. Šių svorio centrų koordinatės, paimtos kaip bazinių funkcijų koeficientų reikšmės, galėtų atitinkti vidutinę kiekvieno atitinkamo klasterio P-bangos formą (parodyta 2C pav.).

Visuose įrašuose buvo stebimi širdies ritmo pokyčiai. Vis dėlto ne visi įrašai parodė tipinę akceleraciją, aprašytą daugelio autorų. Aukšta akceleracijos amplitudė buvo rasta įrašuose su lėtu pradiniu širdies ritmu. Tai iliustruoja trys R-R intervalo fragmentai atitinkantys tris skirtinges įrašus 3 pav. apačioje esančiuose grafikuose (C-E). Kairajame grafike (C) akceleracija yra matoma nuo R-R intervalo 1300 ms iki beveik 800 ms, tuo tarpu dešiniajame grafike (E) galima matyti pradinius apie 750 ms R-R intervalus horizontalioje padėtyje bei trumpalaikį pagreitėjimą beveik iki 420 ms.



Pav.3. Širdies ritmo akceleracijos (RR trumpėjimo) priklausomybė nuo pradinio širdies ritmo (A); klasterių skaičiaus priklausomybė nuo pradinės RR intervalų reikšmės visuose įrašuose (B); Trijų įrašų trys RR intervalų fragmentai: nuo labiausiai pastebimos akceleracijos (C), per (D), iki mažiausiai pastebimos akceleracijos (E). Visuose trijuose grafikuose atvaizduotas tik ortostazės momentas (parodyta strėlyte), tačiau juose nevaizduojamas grįžimo į horizontalią padėtį momentas.

Visuose trijuose grafikuose atvaizduotas tik ortostazės momentas (pažymėta rodykle), tačiau juose nevaizduojamas grįžimo į horizontalią padėtį momentas. Lėtėjimas viduriniame ir dešiniajame grafike nesisieja su paciento grįžimu į gulimą padėtį. Tokia priklausomybė stebėta tik vyresnių pacientų įrašuose (2 ir 3 grupėse).

Akceleracijos priklausomybė lyginant su pradiniu širdies ritmu yra pavaizduota 3A pav. Didžiausia akceleracijos amplitudė stebėta pacientams, patyrusemiems laikinus prieširdžių virpėjimo periodus (3 grupė, pažymėta trikampiais). Ne visi P-bangos formos atvaizdai 7-matėje pagrindinių komponenčių erdvėje formavo keletą klasterių kaip kad pateikta pavyzdys. Kai kuriuose įrašuose jie formavo tiktais vieną klasterį. 3B pav. yra pavaizduota klasterių skaičiaus priklausomybė nuo pradinės R-R intervalų reikšmės visuose įrašuose. Tikėtasi, kad aukštėsnį parasimpatinį tonusą atspindės lėtesnis širdies ritmas pacientui esant horizontalioje padėtyje bei bus stebimi ir ženklesni širdies ritmo pokyčiai ortostazės teste metu. Tai galėtų būti susiję su ženklesniais P-bangos formos pokyčiais bei didesniu P-bangos formų klasterių skaičiumi, arba bent jau aiškesniu skirtumu tarp klasterių. Vis dėlto, buvo stebėta vizuali neigiamo priklausomybė tarp širdies ritmo pacientui gulint ir klasterių skaičiaus kontrolinės grupės įrašuose (pažymėta rombais). Vyresnių pacientų, turėjusių laikinus prieširdžių virpėjimo periodus, P-bangos formos sudarė 2 arba 3 klasterius (pažymėta trikampiais), tuo tarpu tiktais 1 arba 2 klasteriai buvo stebimi sveiki vyresnių pacientų įrašuose (2 grupė, pažymėta kvadratais).

4. Aptarimas

Pasiūlytas pagrindinių komponenčių analizes pagrindu sukurtas metodas leidžia kiekybiškai įvertinti P-bangos formos pokyčius, koreliuojančius su širdies ritmo pokyčiais ortostazės teste metu. Suskaitmeninti visų 12 derivacijų P-bangos atskaitymai sujungti į bendrą vektorių duoda perteklinį, tačiau išsamų sujaudinimo sklidimo prieširdžiai vaizdą. Pagrindinės komponentės, suskaičiuotos iš tokų vektorių visumos, koncentruoja savyje viso įrašo variaciją. Taigi pasiūlytas C kriterijus (7 formulė) yra informatyvus P-

bangos pokyčių įvertis, atspindintis visus galimus pakitimus visose 12 derivacijų. P-bangos formos vaizdavimas pagrindinių komponenčių koeficientais kaip taškais n-matėje erdvėje, atskleidė naujas galimybes detaliai P-bangos formos morfologijos analizei ir rezultatų interpretacijai. Tikėtina, kad taškų klasteriai atspindi tam tikras ritmo vedlių - aktyviausių ląstelių lokalizacijos vietas dešiniajame prieširdyje. Klasterio dydį daugiamatėje erdvėje gali salygoti daugelis faktorių: signalo triukšmai, širdies ašies judėjimas kvėpavimo metu ir t.t. Tačiau šiuo metu mes neturime pakankamų įrodymų, pagrindžiančių minėtus faktus. Klasterių skaičiaus skirtumai, rasti mūsų įrašuose, galėtų būti paaškinti skirtingu autonomicio širdies reguliavimo efektyvumu. Stabėta didelė klasterių skaičiaus įvairovė nuo 1 iki 3 klasterių jauniems sveikiems pacientams (1 grupė) ir tik nuo 1 iki 2 klasterių sveikiems vyresnio amžiaus pacientams (2 grupė). Tai atspindi amžinį autonomicio širdies reguliavimo mažėjimą minimą [26]. Du ar trys klasteriai 3 grupės pacientams rodo padidintą parasimpatinį aktyvumą, kuris galėtų būti siejamas su prieširdžių virpėjimu, kaip aprašoma [27]. Mes negalėjome suskaičiuoti universalų pagrindinių komponenčių sujungdamai visus visų įrašų P-bangų rinkinius. Tam visų pirma reikėtų pakankamai tiksliai nustatyti anatominę bei fizinę tikrojo ritmo vedlio vietą registravimo metu. Nežiūrint to, kad registravimo prietaisai buvo gerai kalibravoti, skirtingu įrašų signalų amplitudės būdavo pakankamai skirtinges dėl fizinių tiriamų asmenų kūno skirtumų. Tai tik patvirtino mūsų apsisprendimą netesti universalų pagrindinių komponenčių paieškos. Metodo kalibracija galėtų būti atlikta naudojant kateterines navigacijos sistemas pvz. CARTO® ar pan., pritaikant tiesioginę tam tikros prieširdžių vietas elektrinę stimuliaciją. Tada šis siūlomas neinvazinis metodas leistų tiesiogiai stebeti tikslią tikrojo ritmo vedlio vieta.

5. Išvados

Kiekybinis EKG P-bangos pokyčių, sukeltyų ortostazės mėginio metu įvertinimas, naudojant naujai sukurtą pagrindinių komponenčių analizės metodą, suteikia vertingos diagnostinės informacijos apie autonomicinės širdies veiklos reguliacinių

mechanizmų funkcionalumą bei efektyvumą.

Detalesnė kiekybinių EKG P-bangos įverčių analizė galėtų parodoti potencialius širdies ritmo reguliavimo mechanizmus.

Padėka

Lietuvos valstybinis mokslo ir studijų fondas rėmė šį darbą.

Literatūra

- [1] Hainsworth R.: The control and physiological importance of heart rate, In: Heart Rate Variability, edited by M. Malik, and A. J. Camm. New York: Futura, 1995, p. 3-19.
- [2] Stewart J.M.: Autonomic nervous system dysfunction in adolescents with postural orthostatic tachycardia syndrome and chronic fatigue syndrome is characterized by attenuated vagal baroreflex and potentiated sympathetic vasomotion. Pediatr Res. 2000 Aug; 48(2):218-26.
- [3] Ziegler D., Dannehl D., Muhlen H., Spuler M., Gries F.A.: Prevalence of cardiovascular autonomic dysfunction assessed by spectral analysis, vector analysis, and standard tests of heart rate variation and blood pressure responses at various stages of diabetic neuropathy. Diabet Med 9:806- 814, 1992
- [4] Jacob G., Costa F., Biaggioni I.: Spectrum of autonomic cardiovascular neuropathy in diabetes. Diabetes Care. 2003 Jul; 26(7):2174-80. Erratum in: Diabetes Care. 2003 Sep; 26(9):2708.
- [5] Ito M., Pride H.P., Zipes D.P.: Defibrillating shocks delivered to the heart impair efferent sympathetic responsiveness. Circulation. 1993 Dec; 88(6):2661-73.
- [6] Rigden L.B., Mitrani R.D., Wellman H.N., Klein L.S., Miles W.M., Zipes D.P.: Defibrillation shocks over epicardial patches produce sympathetic neural dysfunction in man. J Cardiovasc. Electrophysiol. 1996 May; 7(5):398-405.
- [7] Pauza D.H., Skripka V., Pauziene N., Stropus R.: Morphology, distribution, and variability of the epicardiac neural ganglionated subplexuses in the human heart. AnatRec. 2000 Aug 1; 259(4):353-82.
- [8] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Eur Heart J 1996;17:354-81.

- [9] Krisciukaitis A., Bukauskas F., Adomonis V., Lukosevicius K., Muckus K.: Changes of Orthogonal Leads during Pacemaker Migration. *Electrophysiology and Surgery of Cardiac Arrhythmias*. Vilnius, "Mokslas", 1987, 41 - 46.
- [10] Boyett M.R., Honjo H., Kodama I.: The sinoatrial node, a heterogeneous pacemaker structure. *Cardiovasc Res*. 2000 Sep; 47(4):658-87. Review.
- [11] Kodama I., Boyett M.R., Suzuki R., Honjo H., Toyama J.: Regional differences in the response of the isolated sino-atrial node of the rabbit to vagal stimulation. *J Physiol*. 1996 Sep 15; 495 (Pt 3):785-801.
- [12] Adomonis V.M., Bredikis Iu.Iu., Bukauskas F.F., Lukoshiavichius K.K., Mutskus K.S.: [Transposition of the pacemaker in the right atrium during stimulation of the vagus nerve in the dog] *Biull Eksp Biol Med*. 1987 Apr; 103(4):387-90. Russian.
- [13] Shibata N., Inada S., Mitsui K., Honjo H., Yamamoto M., Niwa R., Boyett M.R., Kodama I.: Pacemaker shift in the rabbit sinoatrial node in response to vagal nerve stimulation. *Exp Physiol*. 2001 Mar; 86(2):177-84.
- [14] Bouman L.N., Gerlings E.D., Biersteker P.A., Bonke F.J.: Pacemaker shift in the sino-atrial node during vagal stimulation.
- Pflügers Archiv (1968) 302, 255-267.
- [15] Levy M.N., Martin P.J., Tesse S.L.: Neural regulation of the heart beat. *Annual Reviews in Physiology* (1981) 43, 443-453.
- [16] Krisciukaitis A., Bukauskas F., Puodzius S.: Investigations on hypoxic pacemaker shifts. *Abs. 10-th International Biophysics Congress, Vancouver, July 29 - August 3 1990, Canada*, p.405.
- [17] Pappano A.J.: Vagal stimulation of the heartbeat. Muscarinic receptor hypothesis. *Journal of Cardiovascular Electrophysiology* (1991) 2, 262-273.
- [18] Jollife I.T.: Principal component analysis (Second edition), (Springer New York, 2002) (ISBN 0-378-95442-2).
- [19] Krisciukaitis A., Simoliuniene R., Tamosiunas M., Saferis V., Vainoras A., Gargasas L.: Estimation of Autonomic Heart Control by Means of Cluster Analysis of Quantitative Estimates of ECG P-Wave Shape. *Proc.Int.Conf. CeHR 2007*, AKA Berlin (2008): 179-184.
- [20] Vitartaitė A., Vainoras A., Sedekerskiene V., Poderys J.: The influence of aerobics exercise to cardiovascular functional parameters of 30-40 year old women. *Medicina (Kaunas)*. 2004;40(5):451-8.
- [21] Sörnmo L., Laguna P.: Bioelectrical Signal Processing in Cardiac and Neurological Applications. Academic Press (June 15, 2005), ISBN: 0124375529.
- [22] Besse P., Ferre L.: Sur l'usage de la validation crois 'ée en analyse en composantes principales, *Rev. Stat. Appl.* 41 (1993) 71-76.
- [23] Wold S.: Cross-validatory estimation of the number of components in factor and principal component models. *Technometrics*, (1978) 20 pp. 397-405.
- [24] Krisciukaitis A., Tamaiunas M., Jakuska P., Veteikis R., Lekas R., Saferis V., Benetis R.: Evaluation of ischemic injury of the cardiac tissue by using the principal component analysis of an epicardial electrogram. *Comput Methods Programs Biomed*. 2006 May; 82(2):121-129.
- [25] Saferis V., Vilkauskas L.A.: Cluster Analysis by Testing the Statistical Hypothesis of Uniformity, *Statistics in Medicine*, Vol. 15, 817-821, (1996).
- [26] Kuo T.B., Lin T., Yang C.C., Li C.L., Chen C.F., Chou P.: Effect of aging on gender differences in neural control of heart rate. *Am J Physiol*. 1999 Dec; 277(6 Pt 2):H2233-9.
- [27] Olshansky B.: Interrelationships between the autonomic nervous system and atrial fibrillation. *Prog Cardiovasc Dis*. 2005 Jul-Aug; 48(1):57-78. Review.



EJBI 2009 ISSN 1801 - 5603

European Journal for Biomedical Informatics

Volume 5 (2009), Issue 1

Portuguese version

www.ejbi.eu



Content

Portuguese version

pt 2 - 9

Recuperação Semântica de Documentos Clínicos e Bibliográficos Baseada em Subwords

P. Daumke, S. Schulz, M. L. Müller, W. Dzeyk, L. Prinzen, E. J. Pacheco, P. Secco Cancian, P. Nohama, K. Markó

Recuperação Semântica de Documentos Clínicos e Bibliográficos Baseada em Subwords

P. Daumke^{1,2}, S. Schulz², M. L. Müller³, W. Dzeyk⁴, L. Prinzen⁵, E. J. Pacheco^{6,7}, P. Secco Cancian⁶, P. Nohama^{6,7}, K. Markó^{1,2}

¹VERBIS GmbH, Freiburg, Germany,

²Freiburg University Hospital, Medical Informatics Group, Freiburg, Germany,

³Freiburg University Hospital, Department of Dermatology, Freiburg, Germany,

⁴German National Library of Medicine, Cologne, Germany,

⁵Bertelsmann Stiftung, Gütersloh, Germany,

⁶Paraná Catholic University, Health Informatics Laboratory, Curitiba, Brazil,

⁷Universidade Tecnológica Federal do Paraná, CPGEI, Curitiba, Brazil

Objetivos: Com o incremento da disponibilidade de documentos eletrônicos em bases bibliográficas e de documentação clínica, a recuperação dessas informações requer a adoção de técnicas e interfaces amigáveis ao usuário.

Métodos: Apresenta-se neste artigo uma estratégia específica para a recuperação de documentos, no domínio da indexação semântica de textos. A estratégia é baseada em um thesaurus de subwords e no mapeamento de textos, em diferentes linguagens, para uma representação única, interlíngua, permitindo a busca entre diferentes coleções de documentos.

Resultados: Três casos são apresentados empregando a metodologia desenvolvida: uma base de dados bibliográfica, um sistema departamental de prontuário eletrônico de pacientes e um portal Web.

Conclusões: A utilização da metodologia de indexação e recuperação semântica revelou-se útil em diferentes protótipos e rotinas de trabalho, tendo sido bem aceita por diferentes grupos de usuários. Por fim, ressalta-se que a avaliação da performance da indexação já fora realizada, empiricamente, em estudos anteriores.

Palavras-Chave: Thesauri, Recuperação Semântica de Informação, Indexação de Documentos, Recuperação de Informação Inter-Língua, Banco de dados bibliográficos, Informática em saúde, Registro eletrônico do paciente.

1. Introdução

Atualmente, mais da metade de toda a informação produzida encontra-se disponível digitalmente [1]. Para o domínio da saúde, o contexto não é diferente, haja

vista a disponibilidade crescente de textos eletronicamente disponíveis em diferentes sistemas de informação clínica e publicações. Esta é uma das razões que justificam o desenvolvimento de sistemas específicos para a recuperação de informação (IR). Inúmeros grupos de pesquisa estiveram comprometidos na adaptação de tais sistemas para o domínio do tratamento médico [2], [3].

Em se tratando de termos médicos, há um elemento complicador, especialmente no idioma alemão, pois eles são freqüentemente caracterizados por formas complexas de composição, derivação e inflexão, como também pela geração constante de acrônimos novos, abreviações e nomes próprios. Além disso, regras gramaticais nem sempre são respeitadas. Em função dessas peculiaridades, técnicas atuais de recuperação de informação, que normalmente baseiam-se em comparação simples de palavras inteiras, não são adequadas, visto produzirem resultados quase sempre incompletos, inexatos, ou fora do escopo desejado [4].

Considerando a necessidade de recuperação de informação em outros tipos de documentos além de narrativas clínicas, como artigos relacionados à área de saúde, manuais e *guidelines*, é importante considerar a prevalência do inglês como idioma fonte em tais documentos de natureza técnica. Neste domínio, é comum identificar que pessoas não-nativas do idioma inglês, apresentam dificuldades para formular questões específicas e precisas em inglês, condição *sine qua non* para a adequada recuperação de documentos em um cenário de IR [5]. Este

contexto reforça a relevância da aplicação de técnicas que permitam a recuperação de documentos em idiomas diferentes da consulta formulada (multi e interlíngua).

Para a recuperação de documentos interlíngua, três diferentes estratégias podem ser aplicadas [6], [7]:

- **Tradução da consulta:** somente os termos da consulta são traduzidos; desta forma, os elementos utilizados para a indexação do documento não necessitam de modificação quando novas línguas são incorporadas ao sistema. O esforço computacional desta técnica é baixa [8].
- **Tradução de documento:** pode apresentar desempenho superior quando comparada à técnica de tradução de consultas, considerando que em um mesmo documento normalmente ocorrem variações morfológicas e sinônimos dos termos empregados [9].
- Para a aplicação de técnicas de interlíngua, tanto os documentos quanto as consultas são traduzidas para uma representação comum. Para tanto, torna-se necessário um léxico independente de linguagem, com suporte ao conjunto de idiomas cobertos.

Na seqüência, apresenta-se um sistema de recuperação interlíngua conhecido como "MorphoSaurus", que considera as diversas particularidades idiomáticas do alemão e permite a busca intra e interlíngua em dados biomédicos. Na Seção 3, duas aplicações reais, baseadas no sistema MorphoSaurus, serão descritas.

2. O Sistema MorphoSaurus

2.1 Subwords como identificadores atômicos de significado

O sistema MorphoSaurus [4], [5], [10], [11] mapeia o conteúdo de textos de domínio específico para um formato interlíngua. Isto requer simplificação e padronização do conteúdo dos documentos e das consultas, de forma a facilitar a recuperação de documentos em coleções multilíngües. O sistema MorphoSaurus utiliza *subwords* como unidades léxicas. Uma *subword* é definida como uma unidade léxica mínima para um termo significante em um domínio, ou seja, o termo não admite outras subdivisões. Assim, é possível considerar o termo "hepat + itis" como uma composição de duas *subwords*, "hepat" e "itis", pois o significado do termo em questão é a soma do significado de seus componentes. Um contra-exemplo é "hipófise", cujo significado não pode ser derivado definitivamente de "hypo" + "physis". *Subwords* tendem a apresentar menor granularidade do que morfemas lingüísticos e serem mais curtas que palavras.

No sistema MorphoSaurus, cada entrada de *subword* tem atributos, como idioma (atualmente, inglês (en), alemão (de), francês (fr), espanhol (es), português (pt), sueco (se) e italiano (it)) e tipo (raiz (ST), prefixo (PF), sufixo (SF), invariante (IV)). As *subwords* estão contidas em léxicos de *subwords*, específicos para cada idioma suportado.

2.2 Estrutura do Tesauro

A camada semântica do sistema MorphoSaurus é representada pelas

classes de equivalência, identificada por identificadores específicos denominados de MIDs (MorphoSaurus *identifiers*). Cada elemento léxico é associado exclusivamente a uma única classe de equivalência. Nas classes de equivalência, são agrupadas variações léxicas, sinônimos e traduções. A composição de todos os MIDs forma o tesauro independente de idioma.

O tesauro do sistema MorphoSaurus utiliza dois tipos de relação: "has_word_part" e "has_sense", conforme ilustrado na Figura 1.

- A relação sintagmática "has_word_part" relaciona um MID com no mínimo outros dois MIDs, objetivando "mascarar" composições semânticas. Normalmente, é utilizado para termos que não podem ser divididos corretamente pela rotina de segmentação, por exemplo, quando da falta de caracteres na composição do termo ("urinalysis").
- A relação paradigmática "has_sense" relaciona um MID ambíguo com, no mínimo, dois outros. Este tipo de relação é utilizado para o relacionamento dos possíveis significados para um mesmo termo.

2.3 Particularidades da base léxica do sistema MorphoSaurus

O processo decisório de construção do tesauro é guiado não apenas pelos critérios da lingüística formal, mas também pelo correto funcionamento da função de segmentação. Isto é especialmente relevante para palavras longas e compostas, com diferentes segmentações possíveis.

Por exemplo, "nephrotomy" pode ser segmentada em *nephri*^[en,ST] (#kidney) + *o*^[en,sp,pt,IN] + *tomy*^[en,PS] (#incision), mas também em *nephri*^[en,ST] + *oto*^[en,ST] (#ear) + *my*^[en,ST] (#muscle). Somente com um complexo processo de análise de contexto e de rotinas de processamento lingüístico (que normalmente não estão disponíveis), é possível garantir que a segunda opção (errônea) de segmentação NÃO seja a preferida. A solução pragmática é incluir, adicionalmente, outras variações de lexemas (sinônimos). Para o exemplo apresentado, isto significa que o significado de #kidney não é representado somente por *nephri*^[en,ST], mas também por *nephro*^[en,ST] (como também por *nefr*^[sp,pt,ST] e *nefro*^[sp,pt,ST]).

A delimitação de *subwords* e classes semânticas (MIDs) é uma atividade que requer considerável conhecimento do domínio da terminologia e não por ser totalmente automatizada [4]. Para a construção do léxico e do tesauro foram investidos (até novembro/2008) sete anos de trabalho, inicialmente focado no inglês e alemão e, depois, no português, espanhol, sueco e francês. Neste processo, foi possível identificar que existe uma substancial similaridade entre os termos médicos em diferentes idiomas (cognatos). Explorando esta característica observada, desenvolveu-se um processo semi-automatizado para a aquisição de entradas lexicais para novos idiomas, otimizando a atividade de aquisição [4]. Este procedimento baseia-se em uma tabela de regras de substituição, parcialmente apresentada na Tabela 1.

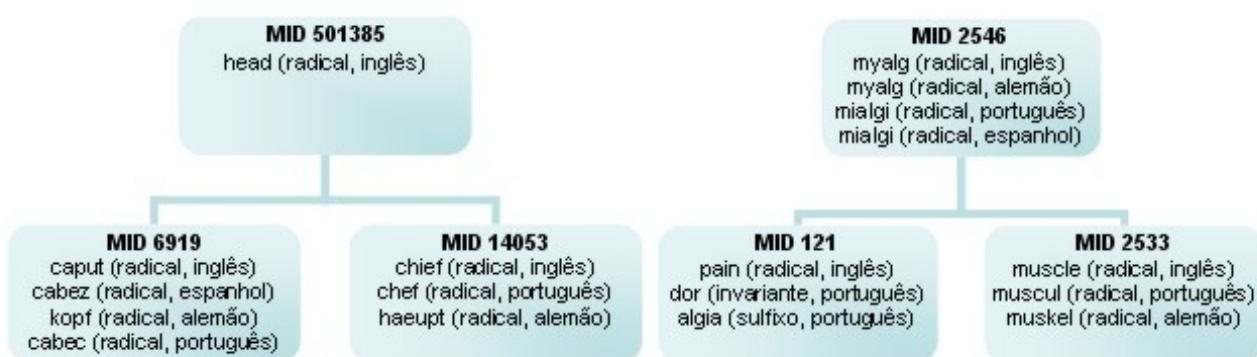


Figura 1. Relações semânticas suportadas pelo tesauro do sistema MorphoSaurus.

Com a aplicação dessas regras, um conjunto de candidatos para tradução foi gerado e validado contra uma lista de palavras na língua objeto da aquisição. Subseqüentemente, utilizaram-se corpora paralelos para enriquecer o vocabulário na língua objeto de trabalho.

2.4 Arquitetura do MorphoSaurus

A arquitetura do sistema MorphoSaurus é ilustrada na Figura 2. Primeiramente, na etapa de normalização ortográfica removem-se todas as palavras insignificantes e aplica as substituições de caractere (como, por exemplo, a eliminação da capitalização e acentos).

Tabela 1. Algumas regras de substituição e exemplos.

44 Regras:	Português	Espanhol	7 Regras:	Inglês	Sueco
ss → s	fracass	fracas	c → k	cramp	Kramp
lh → j	mulher	mujer	ph → f	phosphor	Fosfor
+ça → za	cabeça	cabeza	ce → s	iceland	island

Seqüencialmente, um parser morfossintático divide cada palavra restante em subwords. A estratégia com melhor performance para a extração de subwords em bases compostas por um conjunto muito grande de textos é a aplicação de técnicas baseadas em uma

máquina de estados finitos, para a decomposição léxica, derivação e deflexão, conforme descrito em [4]. Na etapa de normalização semântica, essas unidades são relacionadas com um MID.



Figura 2: Arquitetura do sistema MorphoSaurus.

Tabela 2. Pipeline de indexação morphosemântica.

Normalização Ortográfica	Parser Morfosintático	Normalização semântica		
		Regras de normalização	Léxico de Subwords	Thesauros de subword
High TSH values suggest the diagnosis of primary hyperthyroidism	→ high tsh values suggest the diagnosis of primary hyperthyroidism	→ high tsh values suggest the diagnos is of primar y hyper thyroid ism	→ #up tsh #value #suggest #diagnost #first #small #thyre	
Erhöhte TSH-Werte erlauben die Diagnose einer primären Hypothyreose	→ erhoechte tsh-werte erlauben die diagnose einer primaeren hypothyreose	→ er hoeh te tsh wert e erlaub en die diagnose e einer primae ren hypo thyre ose	→ #up tsh #value #permit #diagnost #first #small #thyre	
A presença de valores elevados de TSH sugere o diagnóstico de hipotireoidismo primário	→ a presencia de valores elevados de tsh sugere o diagnostico de hipotireoidismo primario	→ a presenc a de valor es elevad os de tsh suger e o diagnost ico de hipo tireoid ismo primary o	→ #current tsh #value #up tsh #suggest #diagnost #small #thyre #first	

As consultas são processadas de forma análoga, permitindo a busca multilingual. Na Tabela 2, descrevem-se as principais etapas deste processo, com exemplos em inglês, alemão e português.

A ambigüidade léxica é tratada após o processo de normalização. Por exemplo, em certo contexto, a palavra *lobo^{PT}* pode denotar um animal (*wolf^{EN}*), mas em outro cenário denota uma parte do cérebro humano. O sistema MorphoSaurus usa uma rotina de desambiguação léxica, treinada em um *corpus* multilingual. Experimentos anteriores demonstraram que a desambiguação eleva, substancialmente, o desempenho global do sistema [12].

3. Aplicações

Apresentam-se três cenários de aplicação nos quais a tecnologia do sistema MorphoSaurus foi utilizada. O primeiro foca na recuperação multilíngue em literatura publicada, enquanto o segundo mostra como usuários da área clínica podem se beneficiar de técnicas de recuperação semântica para a busca em registros eletrônicos de pacientes.

3.1 Otimizando e avaliando uma máquina de busca médica

A biblioteca nacional alemã de medicina (ZB MED) é a maior biblioteca europeia de medicina. Mantém inúmeras bases bibliográficas, disponíveis através do portal MEDPILOT-Portal (<http://www.medpilot.de>), um desenvolvimento conjunto entre a ZB MED (German National Library of Medicine (ZB MED): <http://www.zbmed.de>) e o Instituto alemão de Documentação e Informação Médica (DIMDI) (German Institute of Medical Documentation and Information (DIMDI): <http://www.dimdi.de>). Com uma simples requisição, usuários do MEDPILOT podem, simultaneamente, pesquisar um amplo conjunto de fontes e bases médicas utilizando um sistema de busca.

Usando a estratégia de indexação do sistema MorphoSaurus, foi possível disponibilizar um sistema de inter e intra-língua para o MEDPILOT. Três diferentes bases foram disponibilizadas com a nova tecnologia: a versão de 2007 do Medline, Current Contents Medicine (CCMED)

e o catálogo público de acesso (OPAC) da biblioteca nacional de medicina da Alemanha, totalizando 15.7 milhões de itens bibliográficos que, basicamente, cobrem as línguas mais faladas na Europa.

O principal objetivo foi analisar as necessidades dos usuários que efetuavam buscas na nova ferramenta e a determinação de um padrão ouro para a realização de *benchmark* para medição de desempenho do sistema MEDPILOT, antes e depois da integração da nova tecnologia.

Três foram os estágios de avaliação:

- um analisador de consultas que explora as necessidades dos usuários e estabelece o escopo de conteúdo especial, como acrônimos, variações lingüísticas, o uso de termos químicos clássicos, marcas e nomes farmacêuticos;
- desenvolvimento e avaliação de um conjunto de coleções de teste para a comparação entre os dois sistemas, levando em consideração a relevância dos resultados obtidos [13];
- melhoria na usabilidade da interface com o usuário. A usabilidade é um fator importante na satisfação e na credibilidade do website [14], [15].

Para a análise de conteúdo, utilizou-se o log de consultas do MEDPILOT, como realizado por Herskovic et al., em 2007 [16]. Deste modo, foram extraídas as consultas que cobrem sete meses (142.922 consultas) e selecionado um conjunto aleatório de 10.000 consultas. Na continuidade desenvolveu-se um sistema de categorização com 24 classes que foram construídas por interação de procedimentos dedutivos e induktivos [17]. Por um lado, isto baseou-se no conhecimento sobre pesquisa médica [18] e por outro, o conteúdo do material foi analisado. A validação do sistema de categorização foi realizada com 150 consultas. Nesta atividade, o grau de confiança alcançou 88%, o que confirma a correlação realizada como suficiente [19]. Finalmente, cada uma das 10.000 consultas foi correlacionada com uma ou mais categorias por um profissional com conhecimento no domínio.

A análise foi realizada considerando o seguinte conjunto de questões:

- Que tipo de conteúdo os usuários do MEDPILOT estão interessados?
- Quão complexadas são as consultas de busca? Quantas são as palavras utilizadas para a busca de conteúdo médico?
- Qual o tipo de operador e campo é utilizado nas buscas booleanas?
- Qual o tipo de abreviações e acrônimos, no domínio médico, é utilizado nas consultas?
- Qual a quantidade e o tipo de erros ortográficos que ocorrem nas consultas?
- O que normalmente os usuários procuram para químicos clássicos: nome genérico, droga ou marcas?

A avaliação aconteceu com o apoio de experts em terminologia do domínio médico e biológico. Considerando que a determinação exata do valor de *recall* é impossível para um conjunto muito grande de documentos no Medline, foi adotada uma linha mais pragmática para a avaliação dos dados. Para este propósito, dois indicadores de performance de recuperação de informação foram utilizados: a quantidade e a qualidade (precisão) dos resultados. Cada coleção de teste utilizada para a comparação entre os sistemas (com e sem a aplicação do sistema MorphoSaurus) consistiu de 50 consultas e refletiu aspectos lingüísticos especiais, entre os quais, erros ortográficos, acrônimos, abreviações, sinônimos, composição de palavras simples, traduções, etc.

Realizou-se a comparação da quantidade (número de *hits*) e qualidade (número de *hits* relevantes entre os primeiros 20 resultados) dos resultados antes da pesquisa nas bases do Medline e CCMED, frente aos resultados da versão anterior do MEDPILOT. Futuramente, será construída uma coleção de testes na qual será possível trabalhar com um amplo conjunto de fenômenos lingüísticos.

A avaliação preliminar do MEDPILOT indicou que aproximadamente 35,9% das consultas consistem de uma única palavra, 30% das consultas contêm apenas duas palavras,

16% das consultas consistem de 3 termos e 6,6% com 4 termos. Cinco palavras são utilizadas em 3,7% das consultas e 7,8% das consultas contêm mais de 5 palavras. Em resumo, consultas com até duas palavras correspondem às estratégias de busca mais freqüentes para a busca no domínio médico. Ao examinar os arquivos de log, foi possível observar que operadores booleanos são raramente utilizados. Acrônimos e abreviações ocorrem em 5,0% das consultas e erros ortográficos estão presentes em 4,6% das consultas.

A análise do conteúdo das consultas de busca revelou que a categoria mais procurada é a de "doenças, síndromes, sintomas" com 30,9%, "métodos de tratamento, terapia e diagnósticos" aparece em segundo lugar, com 28,5% e, com 15,6%, o conteúdo de "medicina social, estatísticas, estudos e epidemiologia". Uma das razões para a investigação do conteúdo das consultas foi determinar se a busca por químicos clássicos (nome genérico) é comum. Ao analisar os resultados, foi encontrado apenas 2,1% das consultas com esta característica. Outras informações necessárias estão presentes em quatro outras categorias: bioquímica, drogas, farmacêuticas e marcas.

Preliminarmente, os resultados indicam que a tecnologia do sistema Morpho-Saurus apresenta vantagens significativas, tanto na qualidade quanto na quantidade de resultados recuperados. Um primeiro exame das bases do Medline e do CCMED revelou que a metodologia tradicional apresenta uma média de *hits* de 13,5 vezes menor do que a nova metodologia. Na mesma direção, ao se comparar a qualidade dos *hits* obtidos (precisão dos 20 primeiros resultados), a nova aplicação apresentou, em média, um incremento de 20 a 30% na relevância. Na média, 60% dos *hits* entre o conjunto dos 20 primeiros elementos foram relevantes.

3.2 Recuperação de informações em um Sistema de Prontuário Eletrônico do Paciente (PEP)

O sistema MorphoSaurus foi utilizado em um segundo cenário: a busca de informações em um banco de dados que armazena, em formato eletrônico, os

registros de saúde do Departamento de Dermatologia do Centro Médico Universitário, em Freiburg, na Alemanha. A busca e a recuperação de informações ocupam uma parte significativa do tempo dos médicos [20] que, muitas vezes, não ficam satisfeitos com as ferramentas que deveriam - em teoria - dar-lhes um acesso rápido tanto a informações clínicas sobre os pacientes quanto a informações científicas.

Soluções de busca nas quais os usuários podem facilmente navegar através de diferentes fontes de dados, utilizando uma única interface intuitiva ("Google para os médicos") são uma grande necessidade. Sistemas de Informação Hospitalar (HIS) tradicionais costumam ser centrados no paciente, considerando questões como o armazenamento de resultados laboratoriais, relatórios médicos, sumários de alta, códigos de procedimentos e diagnósticos, e outros documentos clínicos.

O PEP, "um sistema de armazenamento de informações clínicas computadorizado, associado a uma identificação pessoal" [21], [22], deve também cumprir o requisito de "ser seguro, operar em tempo real, estar disponível nos pontos de atendimento, proporcionar um recurso centralizador das informações do paciente para os médicos", de acordo com o HIMSS [23]. O PEP oferece suporte igualmente à busca de informações para outros fins que não estejam diretamente relacionados aos cuidados clínicos, como faturamento, gestão de qualidade, relatórios de aplicação de recursos, planejamento, vigilância de saúde pública e de informação. De acordo com essas definições, o acesso às informações médicas contidas nos sistemas atuais de Informação Clínica é essencialmente horizontal, centrada no paciente, ou seja: ao acessar o prontuário eletrônico do paciente, o médico recebe apenas as informações associadas ao paciente.

A definição do HIMSS sugere cenários adicionais que agregam informações de um grande número de PEPs de um modo verticalizado. Visto que os sistemas baseiam-se em conjuntos de dados estruturados como faturamento, codificação de procedimentos

e diagnósticos, resultados de exames laboratoriais ou de microbiologia, essas informações facilmente podem ser recuperadas com a aplicação de técnicas convencionais de mineração de dados. No entanto, as fontes de narrativas não estruturadas, como os relatórios de queixas do paciente dos sumários de alta, são da mesma importância para a assistência ao paciente: quanto mais informações o sistema armazena, mais interessantes são suas interdependências verticais, independentes do paciente.

Algumas questões que o médico poderia formular:

- "Que pacientes com a mesma doença foram tratados?"
- "Quais foram os resultados / efeitos adversos de determinado tratamento?"
- "Eu tive pacientes com doença X sintoma Y?"
- "Qual era o nome do paciente com sintomas Y1 que atendi três semanas atrás?"

Embora diversas tecnologias promissoras como a Clinical Document Architecture [24], [25] e as terminologias médicas tenham sido desenvolvidas com o objetivo de padronizar e estruturar as informações clínicas, ainda existe uma lacuna entre as necessidades clínicas e o que realmente se pratica.

No estudo realizado, foram extraídos aproximadamente 30.000 documentos clínicos da base de dados do sistema de informação hospitalar do Departamento de Dermatologia do Centro Médico Universitário em Freiburg, na Alemanha. Os documentos eram, em sua maioria, sumários de alta, mas existiam também relatórios cirúrgicos, relatos imuno-dermatológicos e vários outros tipos de relatos e relatórios. Utilizando a abordagem de busca semântica desenvolvida, todas as informações foram disponibilizadas aos médicos através de uma interface de pesquisa baseada no estilo Google. A fim de avaliar o seu impacto, um cenário de avaliação foi criado. Os usuários intermitentemente eram incentivados a avaliar os benefícios que perceberam no sistema e o potencial para seu trabalho.

Os seguintes tópicos foram examinados na avaliação:

- Como os resultados da pesquisa devem ser apresentados aos médicos?
- Quais características são importantes na rotina clínica?
- Que informações adicionais podem melhorar a usabilidade desse sistema para os médicos?
- Quão útil é a busca e a recuperação de informações para a rotina clínica, para a pesquisa científica ou o ensino?

Considerando que as imagens fotográficas dos pacientes desempenham um papel importante na formação dos registros dermatológicos [26], um grande repositório de imagens, composto de 90.000 fotografias, foi associado à interface web, a fim de aumentar o benefício global do sistema. Para a concepção do aplicativo, baseado em interface web, utilizou-se como princípio uma combinação de facilidade de utilização com opções específicas de pesquisa, de acordo com necessidades dos profissionais (ver Figura 3):

Uma caixa de edição permite pesquisa por texto livre em todos os documentos. Os resultados podem ser classificados por relevância, data, nome e data paciente. Na seção de resultados, a parte relevante do documento é apresentada, em conformidade com os termos da pesquisa. As palavras ou suas derivações são realçadas. Informações sobre o paciente, o autor e a data de criação do documento são apresentadas no respectivo contexto.

O documento original pode ser visto como um arquivo PDF associado ao registro.

Outro link permite acesso à biblioteca de imagens do respectivo paciente (Figura 3), onde todas as imagens são apresentadas juntamente com a localização e a data de criação. Isto é particularmente útil para os dermatologistas, permitindo que eles possam analisar as imagens no contexto do documento subjacente e, se necessário, associar as imagens, a resultados de exames laboratoriais.

A avaliação mostrou que aceitação por parte dos usuários foi extraordinariamente alta e que os mesmos mostraram-se entusiasmados desde o início: pela primeira vez, eles tiveram acesso a milhares de documentos e imagens, independentes da visão centrada no paciente. A economia de tempo foi comprovada para todas as etapas de documentação clínica. Além disso, o sistema facilitou a recuperação de relatos de casos para fins de pesquisa clínica-epidemiológica.

No questionário aplicado, 82% dos usuários afirmaram que o sistema poderia melhorar a seu desempenho clínico. Aproximadamente 89% dos usuários consideram que esse tipo de mineração de dados, com integração de textos e imagens dermatológicas, tem um impacto muito positivo sobre o seu trabalho científico. O impacto sobre a educação dermatológica teve uma avaliação inferior: 52% dos usuários estimam um potencial benefício.

Os usuários gostariam de uma integração com mais dados clínicos, como imagens de radiologia, relatórios, resultados de exames laboratoriais e outros achados clínicos. A integração do sistema com vocabulários padronizados (por exemplo, SNOMED ou MeSH) também poderia elevar significativamente o valor clínico. Informações sobre o paciente poderiam ser cruzadas e associadas a fontes externas de informação como Medline ou a base de dados Cochrane [27]. Este não foi o foco do trabalho aqui descrito, mas será um marco importante no futuro.

3.3 Facilitando acesso às informações médicas

A plataforma "Weisse Liste" (lista branca) [28] é uma associação entre o German Bertelsmann Foundation e as maiores organizações de pacientes e consumidores da Alemanha. A plataforma é centrada em tecnologia baseada em portal, livre e não comercial, que permite aos pacientes, seus familiares e conselheiros a busca de informações sobre instituições de saúde. Nesta primeira etapa, o portal oferece informações sobre cerca de 2000 hospitais na Alemanha.

O portal propicia aos usuários a formulação de consultas conforme seu conhecimento anterior e suas necessidades. O elemento central - um assistente interativo para doenças específicas - guia o usuário, passo-a-passos, na customização do resultado,

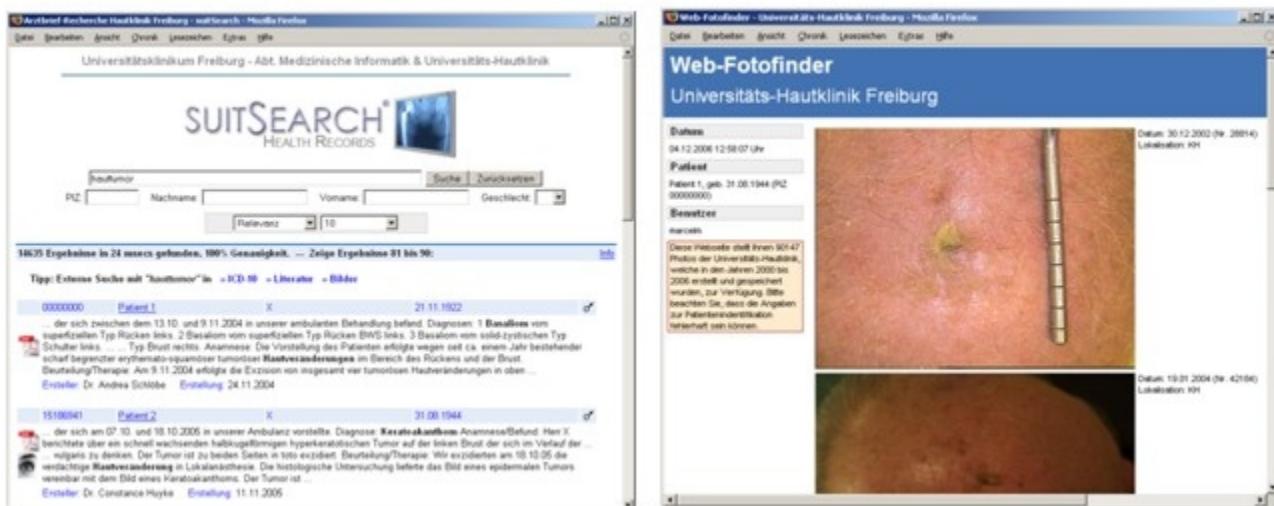


Figura 3. Interface web do Usuário e Recuperação de Imagem.

provendo uma maneira simples para a compreensão de informações complexas sobre o desempenho e a estrutura da qualidade de instituições de saúde.

As informações sobre os hospitais são obtidas por intermédio do "Relatório Estruturado de Qualidade" (SQB), um relatório que agrupa e normaliza informações dos hospitais na Alemanha. Essas especificações garantem que todos os hospitais, de forma pública e regular, fornecem detalhes que permitem a comparação entre a qualidade e os serviços prestados.

O SQB possui duas partes - a sessão básica contém serviços e dados de organização dos hospitais, que provê uma visão sistemática sobre o envolvimento nos processos de gerenciamento de qualidade. Indicadores de qualidade dos hospitais incluem estatísticas de diagnósticos e procedimentos utilizando o *German Modification of the International Classification of Diseases (ICD-10-GM)* e o *German procedure classification OPS* [29].

Um dos maiores desafios para o portal é oferecer aos leigos uma interface de fácil manuseio à base do SQB. Para este propósito, a "lista branca" usa uma adaptação do sistema MorphoSaurus, combinando com uma grande lista de sinônimos incluindo, aproximadamente, 20.000 termos técnicos e genéricos, permitindo a pacientes e demais pessoas que efetuam buscas com suas próprias palavras, 4.000 do ICD-10-GM e de códigos OPS.

O portal independe do interesse de companhias e provedores de saúde. O objetivo é fornecer um meio para apoiar, pacientes e seus familiares, no processo decisório e no suporte aos conselheiros e médicos, oferecendo uma interface bem organizada para acesso a informações sobre qualidade, antes disponível apenas para profissionais de saúde, o que facilita o diálogo entre médicos e pacientes, e permite uma participação ativa na seleção do provedor de saúde.

4. Conclusões

Neste artigo, apresentou-se um sistema

interlíngua de recuperação de informação, denominado MorphoSaurus, que considera muitas das particularidades do alemão e permite a busca, em bases biomédicas, intra e inter lingual. Três casos reais foram descritos.

O sistema MorphoSaurus baseia-se em um tesouro que utiliza *subwords* como unidades lexicais e uma camada semântica onde essas *subwords* são mapeadas para identificadores independentes da linguagem. As aplicações desta abordagem que já haviam sido previamente descritas em diferentes cenários, como o mapeamento de terminologias e recuperação de documentos em bases monolíngues e multilíngues; puderam agora ser demonstradas em grande escala e acessadas por usuários reais. O caso da biblioteca comprovou que uma ferramenta de recuperação de documentos independente do idioma é útil para explorar uma coleção heterogênea de itens bibliográficos no campo da biomedicina. A experiência com prontuários médicos tem mostrado que, em muitos casos, a recuperação horizontal de informações neste tipo de documento vai de encontro às necessidades dos usuários e viabiliza a análise de dados em novos cenários. Por fim, o caso do portal da saúde demonstrou que a abordagem de indexação semântica facilita o acesso a informações de qualidade no serviço de saúde, tanto usando terminologia médica quanto por leigos, usando suas próprias expressões.

Agradecimentos

Este trabalho foi desenvolvido com o suporte do CNPq, Brasil (550830/05-7) e o International Bureau of the BMBF, Alemanha (BRA 05/022).

Literatura

- [1] Scott R.E.: e-Records in health-Preserving our future. *International Journal of Medical Informatics*. 2007; 76(5-6):427-31.
- [2] Lobach D.F., Detmer D.E.: Research Challenges for Electronic Health Records. *American Journal of Preventive Medicine*. 2007; 32(5, Supplement 1):S104-S11.
- [3] Wen H-C, Ho Y-S, Jian W-S, Li H-C, Hsu Y-HE: Scientific production of electronic health record research, 1991-2005. *Computer Methods and Programs in Biomedicine*. 2007; 86(2):191-6.
- [4] Markó K., Schulz S., Medelyan O., Hahn U.: Bootstrapping Dictionaries for Cross-Language Information Retrieval. *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Salvador, Brasil. 2005: 528-535.
- [5] Honeck M., Hahn U., Klar R., Schulz S.: Text Retrieval Based on Medical Subwords. In: Surjan G, Engelbrecht R, McNair P (Hrsg.): *Health Data in the Information Society. Proceedings of the MIE2002*. Amsterdam: IOS Press, 2002; 241-245.
- [6] Oard D.W., He D., Wang J.: User-assisted query translation for interactive cross-language information retrieval. *Information Processing & Management*; 2008: 181-211.
- [7] Gey F.C., Kando N., Peters C.: Cross-Language Information Retrieval: the way ahead. *Information Processing & Management*. 2005; 41(3):415-31.
- [8] Kishida K.: Technical issues of cross-language information retrieval: a review. *Pergamon Press, Inc.* 2005:433-55.
- [9] Franz M., McCarley J.S., Ward R.T.: Ad hoc, cross-language and spoken document information retrieval at IBM. *TREC-8*. Gaithersburg, MD: National Institute of Standards and Technology, 2000.
- [10] Markó K., Hahn U., Schulz S., Daumke P., Nohama P.: Interlingual Indexing across Different Languages. 7th International Conference "Recherche d'Information Assistée par Ordinateur" (RIA0'04); 2004; Avignon, France: Proceedings of RIAO'04; 2004. p. 82-99.
- [11] Markó K., Daumke P., Schulz S., Klar R., Hahn U.: Large-Scale Evaluation of a Medical Cross-Language Information Retrieval System. *Proceedings of the 12th World Congress on Medical Informatics*, Brisbane, Australia (Studies in Health Technology and Informatics, vol. 129). 2007: 392-396.
- [12] Markó K., Schulz S., Hahn U.: Unsupervised Multilingual Word Sense Disambiguation via an Interlingua. *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI '05)*, Pittsburgh, Pennsylvania. 2005: 1075-1080.
- [13] Lewandowski D.: Mit welchen Kennzahlen lässt sich die Qualität von Suchmaschinen messen? In: *Die Macht der Suchmaschinen. / The Power of Search Engines*. Machill M, Beiler M, editors. Cologne: Herbert von Halem Verlag. 2007; 243-258.
- [14] Flavián C., Guinaliu M., Guerra R.: The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management*. 2006; 43:1-14.
- [15] Dzeyk W.: *Vertrauen in Internetangebote*. Saarbrücken: VDM Verlag. 2007.

- [16] Herskovic J.R., Tanaka L.Y., Hersh W., Bernstam E.V.: A Day in the Life of PubMed: Analysis of a Typical Day's Query Log. *J Am Med Inform Assoc.* 2007; 14:212-220.
- [17] Rustemeyer R.: Praktisch-methodische Schritte der Inhaltsanalyse. Eine Einführung am Beispiel der Analyse von Interviewtexten. Münster: Aschendorff. 1992.
- [18] Davies K.: The information-seeking behaviour of doctors: a review of evidence. *Health Information and Libraries Journal.* 2007; 24:78-94
- [19] Wirtz M., Caspar F.: Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen. Göttingen: Hogrefe. 2002.
- [20] Koch O., Kaltenborn R.: Mehr Zeit für Patienten durch bessere Information. *Deutsches Ärzteblatt* 2005; 28-29: A2008-2011
- [21] Waegemann C.P.: The vision of electronic health records. *J Med Pract Manage* 2002; 18(2): 63-65
- [22] Waegemann C.P.: The five levels of electronic health records. *MD Comput* 1996; 13(3): 199-203
- [23] HIMSS Electronic Health Record, Definitional Model. <http://www.himss.org/content/files/EHRAtributes.pdf>, last accessed: January 2009
- [24] Dolin R.H., Alschuler L., Boyer S., Beebe C., Behlen F.M., Biron P.V., Shabo S.A.: HL7 Clinical Document Architecture, Release 2. *J Am Med Inform Assoc* 2006; 13(1): 30-39
- [25] Müller M.L., Uckert F., Bürkle T., Prokosch H.U.: Cross-institutional data exchange using the clinical document architecture (CDA). *Int J Med Inform.* 2005; 245-56.
- [26] Scheinfeld N.: Photographic images, digital imaging, dermatology, and the law. *Arch Dermatol* 2004; 140(4): 473-476
- [27] The Cochrane library. <http://www.cochrane.org>.
- [28] Weisse Liste. <http://weisse-liste.de>
- [29] Klassifikationen im Gesundheitswesen. Deutsches Institut für Medizinische Dokumentation und Information (DIMDI). <http://www.dimdi.de/static/de/klassi/index.htm>



EJBI 2009 ISSN 1801 - 5603

European Journal for Biomedical Informatics

Volume 5 (2009), Issue 1

Slovak version

www.ejbi.eu



Content

Slovak version

sk 2 - 11

Sémantická interoperabilita v prostredí českého zdravotníctva podporovaná HL7 v3

M. Nagy, P. Hanzlíček, P. Přečková, A. Říha, M. Dioszegi, L. Seidl, J. Zvárová

Sémantická interoperabilita v prostredí českého zdravotníctva podporovaná HL7 v3

M. Nagy¹, P. Hanzlíček¹, P. Prečkova¹, A. Říha², M. Dioszegi¹, L. Seidl¹, J. Zvárová¹

¹Centre of Biomedical Informatics, Institute of Computer Science AS CR, Prague, Czech Republic,

²Department of Medical Informatics, Institute of Computer Science AS CR, Prague, Czech Republic

Ciele: Výmena dát v prostredí českého zdravotníctva je väčšinou založená na národných štandardoch. Tento článok popisuje využitie medzinárodných štandardov a nomenklátorov na skonštruovanie pilotnej platformy sémantickej interoperability (PSI), ktorá by slúžila na výmenu informácií medzi systémami elektronických zdravotných záznamov (EZZ) v českom zdravotníctve. Táto práca bola zastrešená národným výskumným projektom programu „Informačná spoločnosť“.

Metódy: Na začiatku projektu boli formulované požiadavky, ktoré by mala PSI splňovať. Bolo analyzovaných niekoľko komunikačných štandardov (openEHR, HL7 v3, DICOM) a HL7 v3 bol zvolený pre naše riešenie na výmenu medicínskych záznamov. Do pilotného prostredia boli zahrnuté dva systémy: WinMedical 2000 a EZZADAMEKj.

Výsledky: Na popis informačného obsahu oboch systémov boli vytvorené Lokálne Informačné Modely (LIM) založené na HL7. Koncepty z našich pôvodných informačných modelov boli namapované na kódovacie systémy podporované HL7 (LOINC, SNOMED CT and ICD-10) a výmena dát pomocou správ HL7 v3 bola naimplementovaná a otestovaná pomocou dotazov na pacientove administratívne údaje. Ako brána medzi lokálnymi systémami EZZ (sEZ) a infraštruktúrou založenou na správach HL7 slúžil nami vyvinutý konfigurovateľný HL7 broker.

Záver: Celonárodná implementácia PSI v plnej veľkosti založená na HL7 v3 by zahrňovala prijatie a preklad príslušných medzinárodných kódovacích systémov a nomenklátorov, vytvorenie implementačných postupov uľahčujúcich migráciu od národných štandardov po tie medzinárodné. Naša pilotná štúdia ukázala, že násprístup je uskutočniteľný, ale úplná integrácia systému českého

zdravotníctva do európskeho kontextu e-zdravia bude vyžadovať ešte veľké úsilie.

Kľúčové slová: Elektronický zdravotný záznam, HL7, sémantická interoperabilita, ukladanie a získavanie informácií, komunikačné štandardy.

1. Úvod

V modernej dobe majú lekári široké možnosti využívania rôznych diagnostických vyšetrení, liečebných procedúr a liekov. Pri mnohých z týchto činností vzniká veľké množstvo dát, ktoré musia byť spracované a posúdené kvalifikovaným odborníkom a ich výsledky musia byť prenesené k ošetrovajúcemu lekárovi, na oddelenie, prípadne do iného zdravotníckeho zariadenia. Niekoľko rôznych faktorov môže vplyvovať na výberu liečby. Možnosť lekára pristupovať k relevantným informáciám o pacientovi v primeranej forme, v správny čas a na správnom mieste, môže byť rozhodujúca pre dobro pacientov.

Po prijatí nových členov do Európskej únie sa ľudia začali viac stáhovať v porovnaní s minulosťou, a preto vznikla potreba cezhraničného prístupu k informáciám a ich integrácia. Problém je, že české zdravotníctvo nemá štandardný spôsob výmeny štruktúrovaných dát so zreteľom na medzinárodnú spoluprácu. Táto situácia predstavuje prekážku pre zahraničných pacientov, „medicínskych turistov“, ako aj pre dodávateľov zdravotníckych informačných systémov, ktorí plánujú vstúpiť na český trh. Viac informácií o využití informačných a komunikačných technológií v rámci komunikácie medzi zdravotníckymi organizáciami možno nájsť v [1].

Uvedenému problému sa venoval národný výskumný projekt programu „Informačná spoločnosť“ Akadémie vied Českej

republiky (AV ČR) s názvom „Informačné technológie pre rozvoj kontinuálnej zdieľanej zdravotnej starostlivosti“ (2004-2008). Projektový tím pozostával z Oddelenia medicínskej informatiky (OMI) Ústavu informatiky AV ČR, dvoch spoločností zaobrajúcimi sa vývojom softvéru a Mestskej nemocnice v Čáslavi.

V tomto príspevku predkladáme čiastkové výsledky tohto projektu z oblasti interoperability a skúsenosti z procesu pilotného zavádzania prostriedkov komunikácie medzi nemocničné informačné systémy. Sekcia 2 popisuje problematiku na pozadí spomínaného projektu a súbor požiadaviek kladených na výsledky tohto projektu. V oddieloch 3 a 4 zhŕnieme použité metódy a stručne predstavíme systémy EZZ (sEZ) podielajúce sa na našom riešení. Návrh platformy pre sémantickú interoperabilitu (PSI), jej časti a niektoré príklady komunikačných správ sú uvedené v sekcií 5. Výsledky nášho projektu sú popísané v oddiele 6 a budúce aktivity zamerané na širšie využitie HL7 v3 v Českej republike v sekcií 7.

2. Pozadie problematiky

Súčasné nemocničné informačné systémy (NIS) v českých zdravotníckych zariadeniach sú skôr zamerané na riadenie inštitúcie, než na samotnú zdravotnú starostlivosť. sEZ sú často založené na voľných textových správach a poskytujú len veľmi obmedzené funkcie pre inteligentné spracovanie uložených dát. Ďalším problémom je, že české NIS-y sú väčšinou koncipované ako monolitické systémy, a preto zdravotnícke zariadenia nemôžu kombinovať komponenty od rôznych dodávateľov.

V prostredí českého zdravotníctva je zdieľanie dát a ich opäťovné využívanie na dosť nízkej úrovni.

Väčšina NIS-ov v Českej republike komunikuje medzi sebou prostredníctvom národného komunikačného štandardu DASTA [2], ktorý je založený na národnej nomenklatúre tzv. Národnom číselníku laboratórnych položiek (NČLP) [2]. Tento štandard je vyvíjaný a spravovaný inštitúciami zaoberejúcimi sa vývojom zdravotníckych informačných systémov, medzi ktoré patria špecializované firmy, univerzity a výskumné inštitúcie v Českej republike. Vývoj tohto štandardu bol podporený najmä Ministerstvom zdravotníctva ČR. DASTA sa špecializuje predovšetkým na zasielanie žiadostí a výsledkov laboratórnych analýz. Aktuálna verzia DASTA je založená na XML a poskytuje aj funkcie pre odosielanie štatistických údajov do Ústavu zdravotníckych informácií a štatistiky Českej republiky a obmedzenú funkcionality pre výmenu klinických informácií vo forme voľného textu. Bohužiaľ, DASTA nemá takmer žiadny vzťah k medzinárodným komunikačným štandardom ako HL7 [3] alebo európskym normám, ako je EN13606 [4].

Použitie medzinárodných štandardov, ako HL7 vo verzii 2 alebo DICOM [5], je zapríčinené najmä požiadavkami v zdravotníckych zariadeniach, ktoré ich využívajú primárne pre komunikáciu s modernými prístrojmi. Bohužiaľ, takéto použitie predstavuje len malú časť celkového využívania mezinárodných komunikačných štandardov v českom zdravotníctve.

Jednou z dobre známych aktivít v Českej republike v rámci e-zdravia je projekt IZIP [6]. Hlavnou myšlienou IZIP-u je, že občania by mali mať prístup k centralizovanej kópii svojej zdravotnej dokumentácie (k tzv. „internetovej zdravotnej knižke“) a mali by byť schopní povoliť prístup k tomuto „skladu“ informácií svojmu praktickému alebo inému ošetrujúcemu lekárovi. Tento projekt zavedol istý stupeň sémantickej interoperability realizovaný ľuďmi avšak nie strojmi. Hlavnou prekážkou v dosiahnutí sémantickej interoperability medzi strojmi [7] bola predovšetkým orientácia IZIP-u na dátá v podobe textových dokumentov.

Hlavným cieľom nášho projektu bolo navrhnuť platformu pre výmenu údajov o zdravotnej starostlivosti v štruktúrovanej podobe s ohľadom na národnú a cezhraničnú komunikáciu. V tejto súvislosti bol formulovaný súbor požiadaviek:

- Vyuzitie medzinárodných komunikačných štandardov a nomenklatúr.
- Automatizované interakcie medzi systémami založené na dobre definovaných pravidlach a dátových tokoch.
- Nezávislosť na klinickej doméne cielová doména by nemala byť napevno spojená s výsledným riešením.
- Možnosť konfigurácie transformácie dát uložených v EZZ do podoby vyžadovanej vybraným komunikačným štandardom by mali byť definované skôr v konfiguračnom súbore než zapísané v programovacom jazyku.
- Rozšíriteľnosť a otvorenosť pridanie novej komunikujúcej strany by malo byť možné pohodlným spôsobom.

3. Materiály

V našom pilotnom projekte, ktorý je primárne demonštráciou možností, nebolo možné pokryť celú oblasť medicíny ako domény pre interoperabilitu. Hlavný riešiteľ projektu, OMI ÚI AV ČR, má dlhú história v interdisciplinárnom výskume orientovanom na kardiologiu. Preto bola vybraná ako doména pre pilotnú realizáciu PSI. Z týchto dôvodov bol pripravený Minimálny Dátový Model pre Kardiologiu (MDMK) [8] zástupcami Českej kardiologickej spoločnosti v spolupráci so štatistikmi špecializujúcimi sa na spracovávanie medicínskych dát. Je to množina dôležitých lekárskych pojmov v oblasti kardiologie, ktorá poslúžila ako základ pre informačné modely dvoch sEZD použitých v našom riešení.

3.1 WinMedicalc 2000 a ADAMEKj

WinMedicalc 2000 (WMC) [9] je stredne veľký komerčný modulárny nemocničný informačný systém, ktorý uprednostňuje využitie štruktúrovanej elektronickej zdravotnej dokumentácie pred voľným textom. WMC sa používa ako nemocničný informačný systém v Mestskej nemocnici v Čáslavi.

ADAMEKj je sEZD vyvinutý na OMI ÚI AV ČR s cieľom zhromaždiť genetické a klinické dátá kardiologického pacienta. Doménový model aplikácie je založený na MDMK. Užívateľské rozhranie je inšpirované predošloou aplikáciou ADAMEK [10]. Hlavný dôraz bol kladený na prijatie programu užívateľmi, preto boli lekári úzko zapojení do celého procesu vývoja, ktorý je detailnejšie popísaný v [11].

3.2 HL7 v3

HL7 v3 je všeobecne uznávaný štandard pre elektronickú výmenu zdravotníckych informácií.

Tento štandard sa okrem množstva artefaktov skladá najmä z Referenčného Informačného Modelu (RIM), pojmových slovníkov, zjednených modelov (D-MIM, R-MIM), scenárov udalostí a špecifikácií implementácie. RIM je popísaný v [12] takto: „RIM je informačný model, ktorý zahŕňa oblasti záujmu HL7 ako celok. Je to koherentný, zdieľaný informačný model, ktorý je zdrojom štruktúr pre dátový obsah všetkých HL7 správ. Ako taký poskytuje konzistentné údaje a koncepty znova použiteľné vo viacerých informačných štruktúrach, vrátane správ.“

Popri definíciách statických štruktúr je definícia dynamiky systému veľmi dôležitá. HL7 v3 prináša pojem „scenár udalostí“, čo je popis udalostí relevantných k zdravotnej starostlivosti definovaný ako postupnosť interakcií alebo prípadov použitia [13]. Scenáre udalostí sú v HL7 analógiami diagramov chovania z UML (diagramu prípadov použitia, stavového diagramu a sekvenčného diagramu).

Štandard HL7 v3 poskytuje špecifikáciu implementačnej technológie (ITS) pre XML. XML ITS špecifikuje spôsob kódovania správ, pravidlá pre vytváranie spojení a prenosov, a procedúry na spracovanie chýb.

4. Metódy

4.1 Sémantická interoperabilita

Podrobnejšiu prácu na tému sémantickej interoperability možno nájsť v [14], kde je popisaný vývojový rámec (nie samotná realizácia) pre vývoj sémanticky interoperabilných zdravotníckych informačných systémov. Avšak my sa zameriavame skôr na realizáciu PSI a na identifikáciu

prekážok v prostredí českého zdravotníctva.

V [7] autor popisuje rozdiel medzi sémantickou interoperabilitou medzi ľuďmi a sémantickú interoperabilitou medzi počítačmi (CSI - computable semantic interoperability), a uvádza štyri nutné, ale nie postačujúce podmienky („Štyri piliere“) pre dosiahnutie CSI:

1. všeobecný informačný model pokrývajúci všetky záujmové oblasti,
2. jednoznačnú sémantiku každého prenášaného dátového elementu definovanou jeho dátovým typom,
3. použitie slovníkov prepojenie informačných modelov a terminológií špecifických pre jednotlivé oblasti,
4. formálny proces tvorby správ zhora nadol obmedzujúci voliteľnosť.

Autor uvádza, že bez splnenia „Štyroch pilierov“ nie je CSI prakticky možné konzistentným spôsobom dosiahnuť a taktiež ukazuje, že štandard HL7 v3 je v súlade s týmito podmienkami.

4.2 Násť prístup

Našou úlohou bolo nájsť cestu k interoperabiliti a zdieľaniu informácií medzi českými zdravotníckymi zariadeniami a medzinárodnými inštitúciami. Na začiatku projektu (polovica roku 2004) bol uskutočnený prvotný výskum v oblasti komunikačných štandardov v zdravotníctve (viď sekciu 6). Štandard HL7 v3 bol vybraný z pomedzi kandidátov ako HL7, DICOM, openEHR [15] a ENV 13606 [16]. Naša prototypová implementácia spája WMC s EZZ ADAMEKj a je založená práve na štandarde HL7 v3. Kedže sa táto norma v tej dobe ešte stále využívala a nebola k dispozícii žiadna normatívna verzia (publikovaná až v roku 2005), nebolo ani veľa implementácií tohto štandardu, ktorými by sme sa mohli inšpirovať. Avšak, v priebehu času sa niektoré skúsenosti z procesu implementácie objavili v [17], [18] a [19].

Aby sme splnili podmienky sémantickej interoperability (1.-4.) vykonali sme nasledujúce kroky:

1. Prvá podmienka poukazuje na používanie spoločného modelu. „Svet HL7“ prichádza s modelom RIM. V našom riešení sme modelovali informácie uložené v použitých sEZZ

vytvorením Lokálnych Informačných Modelov (LIM) (príklad je na Obr. 1). Triedy LIM-u sú odvodené od základných tried RIM. Tento prístup predstavuje mierny odklon od zaužívaneho postupu vývoja správ v HL7 v3, ku ktorému sa ešte vrátíme neskôr.

2. Druhá podmienka bola splnená použitím dátových typov definovaných HL7 v3 pre všetky atribúty v modeli LIM. Dôležitú úlohu zohral dátový typ „kódované s ekvivalentmi“ (CE coded with equivalents) a jeho odvodené podtypy, ktoré nám umožnili spojiť pojmy z nášho modelu s konceptmi z číselníkov.
3. Tretia podmienka vyžaduje používanie pojmového slovníka. Preto sme museli namapovať pojmy z modelov LIM (založené na MDMK) na číselníky, ktoré sú podporované HL7. Kódovací systém LOINC (Logical Observation Identifier Names and Codes) [20] bol použitý ako preferovaný, pretože obsahoval veľkú podmnožinu pojmov používaných v modeloch LIM. Zvyšné koncepty boli kódované pomocou SNOMED Clinical Terms (SNOMED CT) [21] a MKN-10 [22].
4. Posledná podmienka vyžaduje použitie prísnej metodiky. Štandard HL7 v3 poskytuje metodiky pre definovanie štruktúr na výmenu dát s použitím prvkov len z RIM. Po vytvorení LIM

modelov a ich naviazaní na kódovacie systémy, sme ako ďalší krok v procese definície správy z nich museli odvodiť stromové štruktúry, tzv. LIM šablóny. Každá LIM šabloná predstavuje jednu ucelenú časť EZZ, ktorú LIM popisuje, napr. fyzikálne vyšetrenie, predpísané lieky a EKG. HL7 HMD [13] a LIM šablóny slúžia rovnakému účelu, t.j. definovať hierarchiu, poradie a voliteľnosť tried v R-MIM (Refined Message Information Model) resp. LIM.

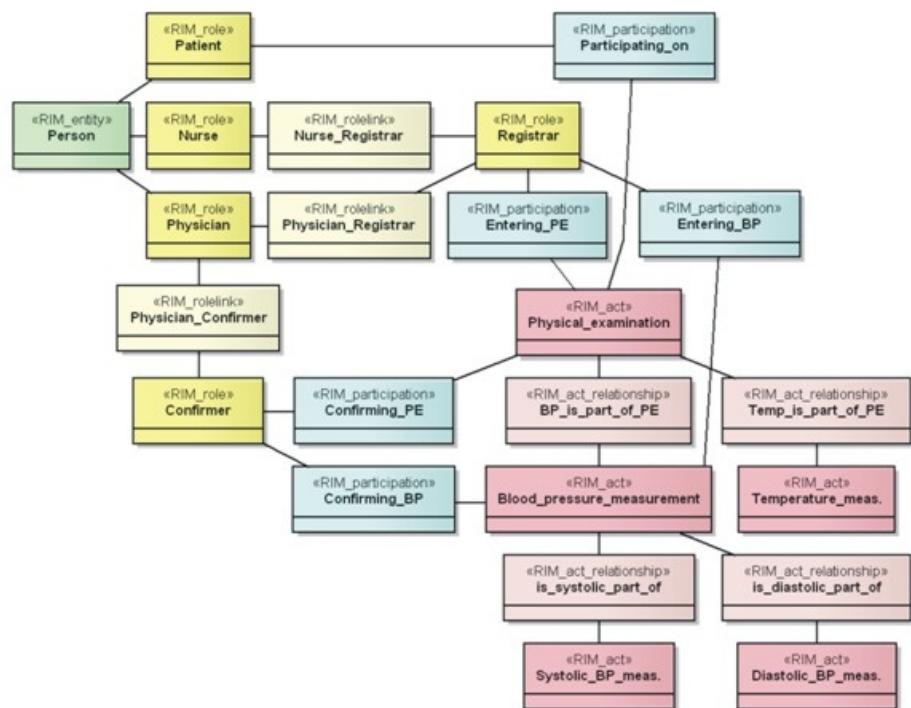
4.3 Testovacie prostredie

Po aplikácii všetkých vyššie uvedených metód, sme museli pripraviť prostredie na otestovanie našej implementácie. Testovacie prostredie bolo tvorené vygenerovanými dátami fiktívnych pacientov z každého sEZZ, rozšírenými verziami klientských aplikácií sEZZ, HTTPS spojeniami cez Internet a pilotnou realizáciou PSI pozostávajúcou z tzv. HL7 v3 brokerov.

5. Výsledky

5.1 Mapovanie konceptov MDMK na medzinárodné nomenklatúry

MDMK je minimálny súbor atribútov potrebných na zistenie, či je pacient rizikový, alebo trpí aterosklerotickým srdcovo-cievnym ochorením.



Obr. 1. Časť LIM modelu v systéme ADAMEKj.

Na nájdenie vhodného mapovania klinických pojmov z MDMK na medzinárodné terminológie bol použitý UMLS vedomostný server (UMLS Knowledge Source Server) [23]. Niektoré príklady tohto mapovania sú uvedené v tabuľke 1.

Pri analýze sme zistili, že približne 85 % všetkých pojmov MDMK bolo zahrnutých aspoň v jednom klasifikačnom systéme. Podarilo sa nám namapovať väčinu klinických pojmov z MDMK do LOINC a SNOMED CT. Počas mapovania sme museli vyriešiť niektoré problémy s pojмami s príliš malou alebo veľkou granularitou, s pojмami, ktoré boli synonymá avšak použité s mierne odlišným významom, či koncepty, ktoré nebolo možné nájsť v žiadnom dostupnom klasifikačnom systéme [24]. Proces mapovania je zhrnutý v tabuľke 2.

5.2 Platforma pre sémantickú interoperabilitu

Primárny výsledok nášho projektu bol návrh PSI založenej na medzinárodnom komunikačnom štandarde, ktorý je uvedený na Obr. 2.

Tab. 1. Zobrazenie namapovaných pojmov z MDMK do LOINC, SNOMED CT a MKN-10.

Popis kódovaného pojmu	Kód	Systém
Dychová frekvencia za minútu	9279-1	LOINC
Srdcová frekvencia za minútu	8893-0	LOINC
Telesná teplota	8328-7	LOINC
Diastolický krvný tlak	8462-4	LOINC
Meranie bielkovín v krvi	2885-2	LOINC
Subjektívne obtiaže pacienta	10154-3	LOINC
Liečenie ischemickej choroby srdečnej	C0585894	UMLS CUI
Zistenie hypertrofie ľavej komory	C0149721	UMLS CUI
Kašeľ po ACE inhibítorkach	C0740723	UMLS CUI
Tranzitorná ischemická príhoda	I61	MKN-10
Angina Pectoris	I20	MKN-10
Hypertrfia prostaty	N40	MKN-10

Tab. 2. Počty úspešne namapovaných klinických konceptov z MDMK.

Názov nomenklatúry použitej k mapovaniu	Počet namapovaných klinických konceptov
SNOMED CT	49
LOINC	34
MKN-10	21
žiadna	3
Celkom	107

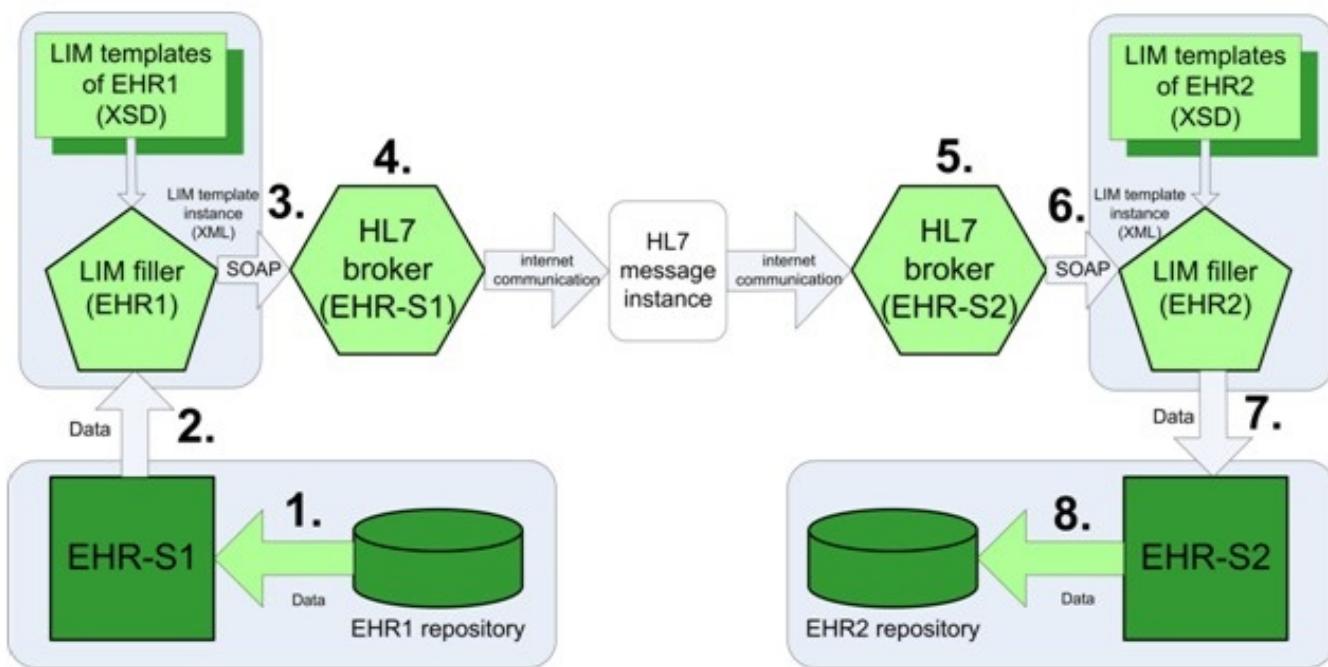
Tento návrh sa skladá z modulov „LIM filler“, HL7 v3 brokerov (skrátene HL7 broker) a pôvodných sEZZ. Všetka komunikácia je bez-stavová a realizovaná zabezpečene. Čísla na Obr. 2 ukazujú tok dát v modelovej situácii, kedy sEZZ (EHR-S1) pošle odpoveď na otázku položenú systémom EHR-S2.

Najskôr sa získajú potrebné dátá z EHR-S1 za pomocí modulu LIM filler, ktorý má prístup k úložisku EHR-S1. Ďalej sa vyberie príslušná LIM šablóna, ktorá obsahuje správne koncepty na prenos údajov. Modul LIM filler vyplní hodnoty

určitým atribútom tried v LIM šablóne (transformácia Dáta-LIM), čím sa vytvorí LIM správa. HL7 broker dostane LIM správu od LIM filler-a pomocou protokolu SOAP. Ďalej sa vykonáva ďalšia transformácia (LIM-HL7), v ktorej HL7 broker vytvára vhodné inštancie HL7 správ, ktoré sú odosielané zabezpečeným spôsobom ďalšiemu HL7 brokeru (prijímaciuemu). V tej chvíli prebieha proces transformácie dát v obrátenom poradí. HL7 broker prislúchajúci k EHR-S2 vytvára LIM správy (HL7-LIM transformácia) prijateľné pre modul LIM fillera systému EHR-S2, a odosielá ich cez

SOAP. Prijímaci modul LIM filler rozpozná prichádzajúcu LIM správu a vyextrahuje z nej údaje (LIM-dáta transformácia) vo forme vhodnej pre skladovanie v databáze EHR-S2. Nakoniec sa údaje natrvalo uložia v EHR-S2. V tomto príklade sme vynechali mechanizmy žiadostí a potvrdení zdôvodnenosti.

Modul LIM filler a HL7 broker budú podrobnejšie popísané v nasledujúcom teste.



Obr. 2. Návrh platformy pre sémantickú interoperabilitu založenej na medzinárodných komunikačných štandardoch.

Tento návrh sa skladá z modulov „LIM filler“, HL7 v3 brokerov (skrátene HL7 broker) a pôvodných sEZ. Všetka komunikácia je bez-stavová a realizovaná zabezpečene. Čísla na Obr. 2 ukazujú tok dát v modelovej situácii, kedy sEZ (EHR-S1) pošle odpoveď na otázku položenú systémom EHR-S2.

Najskôr sa získajú potrebné dátá z EHR-S1 za pomocí modulu LIM filler, ktorý má prístup k úložisku EHR-S1. Ďalej sa vyberie príslušná LIM šablóna, ktorá obsahuje správne koncepty na prenos údajov. Modul LIM filler vyplní hodnoty určitým atribútom tried v LIM šablóne (transformácia Dáta-LIM), čím sa vytvorí LIM správa. HL7 broker dostane LIM správu od LIM filler-a pomocou protokolu SOAP. Ďalej sa vykonáva ďalšia transformácia (LIM-HL7), v ktorej HL7 broker vytvára vhodné inštancie HL7 správ, ktoré sú odosielané zabezpečeným spôsobom ďalšiemu HL7 brokeru (prijímajúcemu). V tej chvíli prebieha proces transformácie dát v obrátenom poradí. HL7 broker prislúchajúci k EHR-S2 vytvára LIM správy (HL7-LIM transformácia) priateľné pre modul LIM fillera systému EHR-S2, a odosiela ich cez SOAP. Prijímaci modul LIM filler rozpozná prichádzajúcu LIM správu a vyextrahuje z nej údaje (LIM-dáta

transformácia) vo forme vhodnej pre skladovanie v databáze EHR-S2. Nakoniec sa údaje natrvalo uložia v EHR-S2. V tomto príklade sme vynechali mechanizmy žiadostí a potvrdení z dôvodov jednoduchosti.

Modul LIM filler a HL7 broker budú podrobnejšie popísané v nasledujúcom teste.

5.3 Modul LIM Filler

Prenášané dátá sú prevádzané medzi formátom lokálneho úložiska EZ a formátom LIM správ, ktoré sú inštanciami LIM šablón z oboch strán komunikácie. Túto činnosť vykonáva modul s názvom LIM filler. Modul LIM filler je prispôsobený pre lokálny sEZ k tvorbe LIM správ v súlade so štruktúrami v danom EZZ.

LIM filler môže byť buď implementovaný ako zásuvný modul sEZ alebo ako samostatná aplikácia. Pracuje v dvoch režimoch: odosielanie a prijímanie. V režime odosielania dát, vytvára LIM správy podľa požiadaviek užívateľov a posiela ich HL7 broker-u. V druhom režime sa cyklicky pýta HL7 brokera či neobsahuje nové relevantné správy pre LIM filler. V prípade, že je stiahnutá nová správa, vyextrahuje z nej dátu a ďalej

postupuje v súlade s konkrétnym scenárom udalostí alebo len uloží údaje pacienta v miestnom EZZ.

LIM filler musí dodržiavať bezpečnostné aspekty komunikačného protokolu. Komunikuje s HL7 brokerom prostredníctvom zabezpečeného HTTP kanála pomocou protokolu SOAP. LIM správy musia byť digitálne podpísané oboma zúčastnenými stranami a podpisy musia byť pred prevodom údajov z LIM správy overené.

5.4 HL7 Broker

Hoci sú na trhu dostupné niektoré HL7 v2 broker-y (napr. MergeCOM-3™, Interfaceware Chameleon™) v počiatocnej fáze projektu (v roku 2005) sme však nenašli žiadne broker-y pre verziu 3. Toto bol dôvod vytvorenia vlastného brokera pre nás projekt. HL7 broker je ťažiskovou súčasťou nášho riešenia. Pre zúčastnené sEZ slúži HL7 broker ako brána do sveta HL7 v3 a oddeluje ich od podrobností standardu HL7. Situácia je znázornená na Obr. 3. Výrobcovia sEZ musia iba specifikovať LIM šablóny založené na ich LIM modeloch, ktoré popisujú štruktúru ich databáz a stačí keď implementujú jednoduchý SOAP klient na odosielanie a prijímanie LIM správ na/od HL7 broker-a.

Okrem LIM šablón musia výrobcovia poskytnúť všetky interne používané číselníky s vhodným popisom nášmu online nástroju na medzi-slovnikové mapovanie.

HL7 broker rozkladá implementačné úsilie na výrobcov sEZZ ako aj na výrobcov HL7 broker-a. Čím ľahšie sú transformácie na strane sEZZ, tým viac práce je potrebné urobiť na strane HL7 broker-a. HL7 broker je potrebné nastaviť osobitne pre príslušný sEZZ. Je treba nahrať všetky LIM šablóny definované výrobcom EZZ (vo forme XSD schém všetkých správ LIM). Pre každý scenár udalostí vyžaduje HL7 broker inštrukcie na spracovanie dát (t.j. čo robiť, keď príde určitá LIM správa, keď je vyvolaná nejaká aplikáčna úloha alebo čo robiť pri chybe). Je tiež potrebné namapovať číselníky v EZZ na pojmový slovník HL7 alebo na iné číselníky.

Jadro HL7 broker-a je zodpovedné za prepis informácií z LIM do HL7, a naopak. To je realizované spracovávaním inštrukcií v jazyku XML pomocou triedou.

Komunikácia medzi sEZZ a HL7 brokerom je realizovaná pomocou webových služieb, ktoré využívajú SOAP cez protokol HTTPS. HL7 broker poskytuje iba tri metódy (sendLimMsg(), getLimMsg(), ackLimMsg()) na prenos dát vo forme LIM správy.

Zaviedli sme jednoduché schémy komunikácie „odosielanie na broker“

a „pravidelné pýtanie sa brokera“, ktoré umožňujú implementáciu SOAP servera len na strane broker-a a SOAP klienta iba na strane sEZZ. SOAP klient je typicky implementovaný jednoduchšie a lacnejšie ako SOAP server. Ako je znázornené na Obr. 3, vývojári môžu tiež monitorovať všetku komunikáciu medzi sEZZ a HL7 brokerom pomocou špeciálnej aplikácie napísanej v jazyku Java. Ukázalo sa to byť veľmi užitočné, najmä na začiatku implementácie SOAP klienta v sEZZ. Spolu s prenášanou LIM správou odosielame a potvrdzujeme kryptograficky podpísaný výťah (digest) zo správy za účelom potvrdenia pravosti prenosu pre každú stranu. HL7 broker môže byť vyvolaný zo strany sEZZ alebo ďalším HL7 v3 brokerom. Podľa typu LIM správy alebo identifikátora HL7 interakcie kombinovaný s identifikátorom LIM relácie prípadne identifikátorom odpovede na HL7 správu sa bude broker rozhodovať či sa nachádza na začiatku novej relácie (a ktorým scenárom začať), alebo či už daná relácia beží. Následne je vybraná vhodná akcia (napr. zostavovanie správy alebo vyvolá výnimku).

Jadro broker-a používa počas prepisu informácie z priatej správy do novej pomocné triedu a súvisiace mapovanie číselníkov.

Popis procesu transkripcie je uchovaný v protokole a výsledok prepisu môže byť prenesený späť odosielateľovi (ako kontrolná správa alebo HL7 LIM potvrdenie).

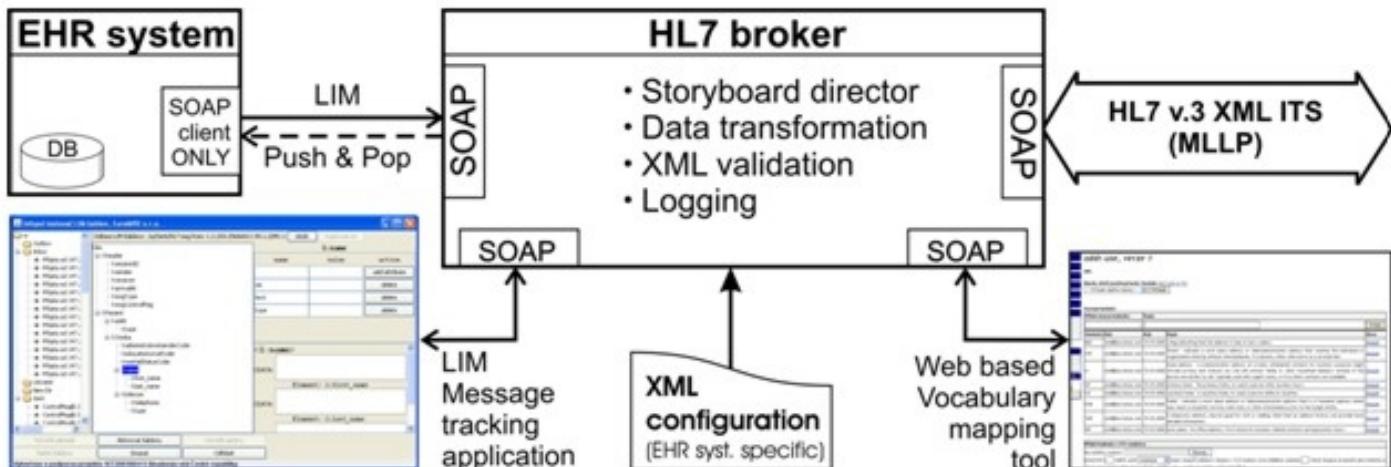
Transport HL7 v3 správ je založený na jednoduchej imitácii MLLP (Minimal Lower Layer Protocol) nad SOAP. HL7 správy sú serializované podľa XML IDS R1.

5.5 Príklady implementovaného scenára udalostí, LIM a HL7 správ

Dotazovanie je v našom riešení založené na HL7 scenároch udalostí. Napríklad môžeme spomenúť scenár z domény „Správa Pacienta (Patient Administration)“ s názvom „Register pacientov nájst kandidátov dotaz (Patient Registry Find Candidates Query)“ (kód artefaktu PRPA_ST201305), ktorý bol implementovaný za účelom hľadania pacientových administratívnych údajov. Sekvenčný diagram UML na Obr. 4 predstavuje činnosti vykonávané v súlade so scenárom „dotaz na pacienta“ s prídavkom dotazu ADAMEK a odpovedi WMC.

Dotazy vytvorené sEZZ a predané HL7 broker-u sa skladajú z „prázdnej“ LIM šablóny, v ktorej sú vyplnené len niektoré atribúty a tie slúžia ako parametre dotazu. Naše riešenie umožňuje vytváranie dotazov na všetky oblasti, ktoré sú pokryté LIM šablónami. HL7 broker vykoná príslušný scenár udalostí, ktorý vede k získaniu údajov požadovaných dotazujúcim sa modulom LIM filler.

Aby sa zlepšil čitateľov prehľad o HL7 správach a dotazoch popísaných v predchádzajúcom texte, uvádzame príklad správy zodpovedajúcej interakcii "Patient Registry Find Candidates Response" (kód artefaktu PRPA_IN201306) v podobe XML (Obr. 6) a dve pridružené LIM správy.

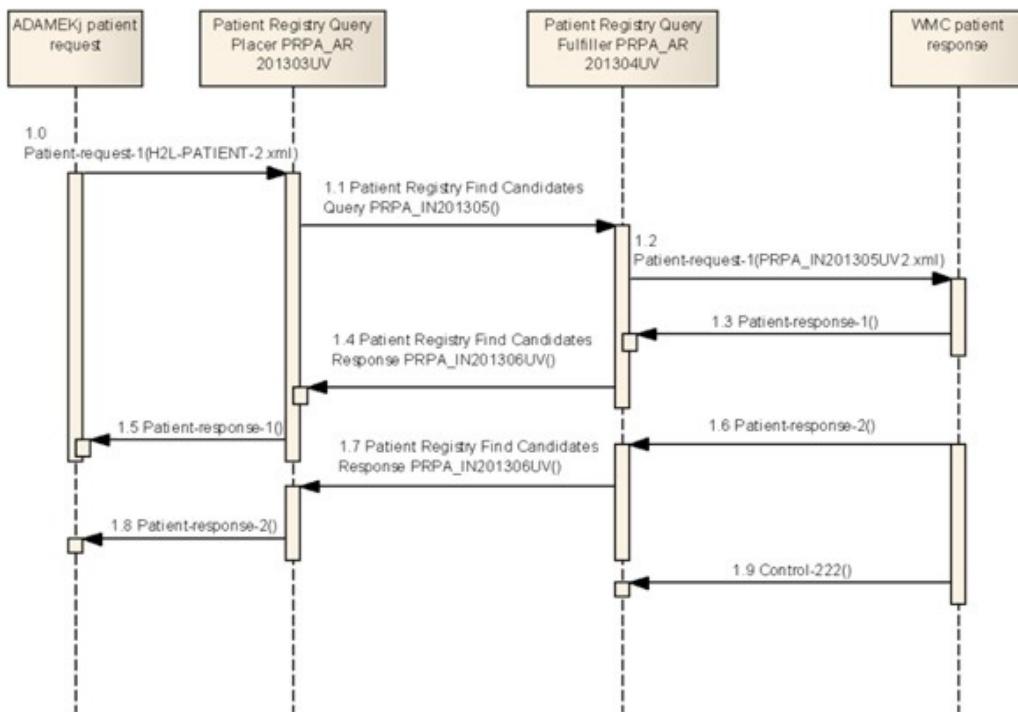


Obr. 3. Funkčná schéma HL7 v3 broker-a.

Prvá (viď Obr. 5 vľavo) slúži ako zdroj údajov pre LIM-HL7 transformáciu na strane odosielateľa (napr. WMC EZZ), a druhá (viď Obr. 5 vpravo) je určená

dátam vytoreným počas HL7-LIM transformácie na strane prijemcu (napr. ADAMEKj).

Predložené správy obsahujú údaje o administratívnych dátach pána Johna Smitha, ktorý bol predmetom dotazu podaného systémom ADAMEKj.



Obr. 4. Sekvenčný diagram dotazu „Patient Registry Find Candidates Query“.

```

<?xml version="1.0" encoding="utf-8"?>
<lim:lim xmlns:lim="urn:lim" xmlns:limdt="urn:limdt"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="urn:lim limdt-PACIENT.xsd urn:limdt limdt.xsd"
    type="template">
<lim:header>
    <limdt:sessionid>48ebd9734f52b9_54759825</limdt:sessionid>
    <limdt:sender>MC-1</limdt:sender>
    <limdt:receiver>UI-1</limdt:receiver>
    <limdt:arrivalat>151008150813</limdt:arrivalat>
    <limdt:msgtype>RESPONSE</limdt:msgtype>
    <limdt:msgcontrolflag>1</limdt:msgcontrolflag>
</lim:header>
<lim:patient type="class">
    <lim:addr>
        <limdt:typ>1</limdt:typ>
        <limdt:jmeno>Ing. CSc. Doc. John Smith</limdt:jmeno>
        <limdt:ulice>Mirové Náměstí 10</limdt:ulice>
        <limdt:psc>32300</limdt:psc>
        <limdt:mesto>Plzeň</limdt:mesto>
        <limdt:stat>203</limdt:stat>
    </lim:addr>
    <lim:osoba type="class">
        <lim:administrativegendercode>
            <limdt:value>M</limdt:value>
        </lim:administrativegendercode>
        <lim:birthtime>
            <limdt:datum format="D">1970-05-07</limdt:datum>
        </lim:birthtime>
        <lim:maritalstatuscode>
            <limdt:value>2</limdt:value>
        </lim:maritalstatuscode>
        <lim:name>
            <limdt:titul_pred>Ing. CSc. Doc.</limdt:titul_pred>
            <limdt:prijmeni>Smith</limdt:prijmeni>
            <limdt:jmeno>John</limdt:jmeno>
        </lim:name>
        <lim:telecom>
            <limdt:telefon typ="M">(+420) 377259020</limdt:telefon>
            <limdt:telefon typ="P">(+420) 737151760</limdt:telefon>
        </lim:telecom>
    </lim:osoba>
</lim:patient>
</lim:lim>

<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE limt SYSTEM "limt-20130301.dtd">
<limt:header>
    <limt:sessionID>1224076036984</limt:sessionID>
    <limt:sender>MC-1</limt:sender>
    <limt:receiver>UI-1</limt:receiver>
    <limt:arrivalat>151008150816</limt:arrivalat>
    <limt:msgtype>RESPONSE</limt:msgtype>
    <limt:msgControlFlag>1</limt:msgControlFlag>
</limt:header>
<limt:patient type="class">
    <limt:addr>
        <limt:typ>unknown</limt:typ>
        <limt:street>Mirov Náměstí 10</limt:street>
        <limt:postal>32300</limt:postal>
        <limt:city>Plzeň</limt:city>
        <limt:state>CZ</limt:state>
    </limt:addr>
    <limt:osoba type="class">
        <limt:administrativeGenderCode code="1" />
        <limt:maritalStatusCode code="ženatý" />
        <limt:birthDate>07.05.1970</limt:birthDate>
        <limt:name>
            <limt:prefix_name>Ing. CSc. Doc.</limt:prefix_name>
            <limt:first_name>John</limt:first_name>
            <limt:last_name>Smith</limt:last_name>
        </limt:name>
        <limt:telecom>
            <limt:telephone>(+420) 737151760</limt:telephone>
            <limt:type>T1</limt:type>
        </limt:telecom>
        <limt:telecom>
            <limt:telephone>(+420) 377259020</limt:telephone>
            <limt:type>T2</limt:type>
        </limt:telecom>
    </limt:osoba>
</limt:patient>
</limt>

```

Obr. 5. LIM správy obsahujúce administratívne dátia pacienta.

```

<?xml version="1.0" encoding="utf-8"?>
<h17:PRPA_IN101306UV02 xmlns:h17="urn:hl7-org:v3">
  <h17:id root="48f5eb3d5d7399.80032341" extension="48f5eb3d5d7462.58285204"/>
  <h17:creationTime>201008150814</h17:creationTime>
  <h17:versionCode code="V3PR1"/>
  <h17:interactionCode root="2.16.840.1.113883.1.6" extension="PRPA_IN101306UV02"/>
  <h17:processingCode code="P"/>
  <h17:processingModeCode code="T"/>
  <h17:acceptAckCode code="AL"/>
  <h17:acknowledgement>
    <h17:typeCode code="AA"/>
    <h17:TargetMessage>
      <h17:id root="48f5eb0f8d9773.81740782" extension="48f5eb0f8d9887.41094918"/>
    </h17:TargetMessage>
  </h17:Acknowledgement>
  <h17:ControlActProcess>
    <h17:classCode code="CACT"/>
    <h17:moodCode code="EVN"/>
    <h17:priorityCode root="2.16.840.1.113883.5.7" codeSystemName="ActPriority"/>
    <h17:QueryAck>
      <h17:queryResponseCode code="OK" codeSystem="2.16.840.1.113883.5.1067" codeSystemName="QueryResponse"/>
    </h17:QueryAck>
    <h17:Subject1>
      <h17:typeCode code="SUBJ"/>
      <h17:contextConductionInd value="true"/>
    </h17:RegistrationEvent>
    <h17:classCode code="REG"/>
    <h17:moodCode code="EVN"/>
    <h17:id root="2.16.840.1.113883.19.420.2" extension="cust1"/>
    <h17:statusCode code="active"/>
    <h17:Custodian>
      <h17:typeCode code="CST"/>
      <h17:contextControlCode code="AP"/>
    </h17:Custodian>
    <h17:Subject2>
      <h17:typeCode code="SBJ"/>
      <h17:IdentifiedPerson>
        <h17:classCode code="IDENT"/>
        <h17:id root="2.16.840.1.113883.19.420.1" extension="6501010001"/>
        <h17:statusCode code="active"/>
        <h17:Person>
          <h17:classCode code="PSN"/>
          <h17:determinerCode code="INSTANCE"/>
          <h17:name>
            <h17:prefix>Ing.</h17:prefix>
            <h17:prefix>CSc.</h17:prefix>
            <h17:prefix>Doc.</h17:prefix>
            <h17:given>John</h17:given>
            <h17:family>Smith</h17:family>
          </h17:name>
        </h17:Person>
      </h17:IdentifiedPerson>
    </h17:Subject2>
  </h17:RegistrationEvent>
  <h17:ControlActProcess>
    <h17:Receiver>
      <h17:typeCode code="RCV"/>
      <h17:Device>
        <h17:classCode code="DEV"/>
        <h17:determinerCode code="INSTANCE"/>
        <h17:id root="1.2.203.25666011.99.1.2" extension="UI-1"/>
      </h17:Device>
    </h17:Receiver>
    <h17:Sender>
      <h17:typeCode code="SEND"/>
      <h17:Device>
        <h17:classCode code="DEV"/>
        <h17:determinerCode code="INSTANCE"/>
        <h17:id root="1.2.203.25666011.99.1.2" extension="MC-1"/>
      </h17:Device>
    </h17:Sender>
  </h17:ControlActProcess>
</h17:PRPA_IN101306UV02>

```

Obr. 6. Výpis z komunikácie podľa scenáru udalostí PRPA_ST201305 XML reprezentácia odpovedi typu „Patient Registry Find Candidates Response“.

6. Diskusia

V tejto sekcií zopakujeme požiadavky naformulované v sekcií 2 a budeme diskutovať o stupni ich naplnenia v našom systéme.

Výmena dát pacientov medzi rôznymi zdravotníckymi zariadeniami v Českej republike je realizovaná väčšinou pomocou národnej normy DASTA s obmedzenou úplnosťou a s malým zreteľom na medzinárodné úsilie (štandardy a číselníky). Vzhľadom k jej postupnému vývoju zdola nahor, neobsahuje vhodný informačný model a má malé pokrytie domén zdravotnej starostlivosti (s výnimkou laboratórnych vyšetrení).

Kandidáti na primárny komunikačný štandard pre PSI boli openEHR, ENV 13606 (ENV 13606 ešte neboli v roku 2004 k dispozícii), DICOM a HL7 v3. OpenEHR je vhodnejší pre budovanie samotného sEZZ [25] než správ a navyše neboli tak vyzretý ako je dnes. Vzťah medzi špecifikáciami od openEHR a HL7 v3 je popísaný v [26]. ENV 13606 mala príliš mälo reálnych implementácií a nachádzala

sa v postupnom procese revízií, ktoré vyústili v mierne odlišnú EN 13606, ktorá ale nebola v tej dobe ešte dokončená. DICOM je zameraný predovšetkým na oblasť obrazových dát a nebol vhodný na nami plánované použitie. Ako najlepšie a najperspektívnejšie riešenie bol zvolený štandard HL7 v3 vďaka svojej širokej medzinárodnej akceptácii, prepracovaným metodikám [13] a dlhej a úspešnej histórii jeho predchádzajúcej verzie. Toto rozhodnutie viedlo k uspokojeniu prvého kritéria - *použitie medzinárodného komunikačného štandardu a číselníkov*.

Obr. 2 ukazuje, že oba sEZZ komunikujú bez akéhokoľvek ľudského zásahu. Sú riadené otázkami kladenými užívateľom sEZZ, ktorý začína komunikáciu. Z tohto pohľadu splňame aj druhé kritérium - *automatizované interakcie medzi systémami*.

Dôležitou črtou nášho riešenia je zavedenie LIM modelov a LIM šablón (vid' sekcia 4.2). HL7 broker umožňuje vytváranie dotazov zo všetkých oblastí, ktoré pokrývajú LIM šablóny. Všetky potrebné informácie o danej klinickej

doméne sa nachádzajú v konfiguračnom súbore HL7 broker-a. Týmto sú splnené požiadavky *nezávislosť na klinickej doméne a konfigurovatelnosť*.

Rozšíritelnosť riešenia bola dosiahnutá použitím medzinárodného komunikačného štandardu a zavedenie HL7 brokera. Komunikujúce systémy nemusia nič vedieť o dátových štruktúrach ich náprotívnej strany. Nový komunikujúci sEZZ je možné pridať po vykonaní nasledujúcich krokov: vytvorenie LIM modelu popisujúceho informácie uložené v EZZ, odvodenie LIM šablón z LIM modelu, implementácia transformačného modulu (napr. LIM filler) pre dátové konverzie medzi databázou EZZ a LIM šablónami, a nakoniec vytvorenie HL7 brokera resp. definičného súboru špecifického pre nový sEZZ. *Otvorenosť* nášho riešenia bola dosiahnutá poskytnutím našich výsledkov všetkým zainteresovaným stranám.

7. Záver

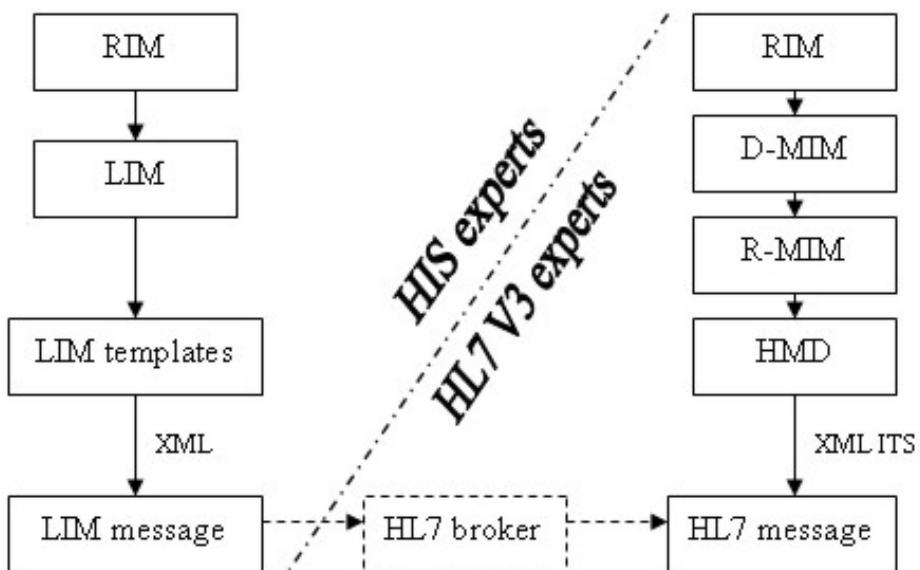
Proces vývoja správ odporúčaný HL7 v3 (vid' Obr. 7) bol zmenený rozdelením pracovného úsilia medzi implemntátorov sEZZ a štandardu HL7.

Tento nový prístup by mohol vývojárom pomôcť prekonať počiatočnú frustráciu, ktorá by mohla byť spôsobená ohromujúcou veľkosťou štandardu HL7 (RIM, množstvo artefaktov atď.). Vývoj jednotlivých LIM modelov úzko súvisiacich s vnútornou štruktúrou informácií v konkrétnom EZZ, jednoduchého komunikačného rozhrania medzi sEZ a HL7 brokerom založenom na LIM správach, SOAP a webových službách, sa zdá byť ľahšie realizovateľný menšími vývojárskymi tímmi ako implementácia celého štandardu HL7 v3. Toto tvrdenie je podporované skúsenosťami dvoch nezávislých vývojárskych tímov zúčastnených na našom projekte. Tento prístup môže ušetriť finančné aj personálne zdroje ak sa použije outsourcing HL7-špecifických úloh príslušným tímom vývojárov, ktorí sú skúsení v implementácii HL7 v3 štandardu.

V priebehu realizácie PSI bolo potrebné použiť simulované údaje o pacientoch, keďže reálne dátá pacientov nemohli byť využité na tieto účely z legislatívnych dôvodov. Výsledky vykonaných testov (napr. položenie dotazu na hľadanie administratívnych dát pacienta) neboli ovplyvnené tým, že údaje boli simulované a sú tiež platné pre skutočné údaje pacientov.

Použitie LIM modelov a LIM filler-a má za následok, že naše riešenie je univerzálne a nie je závislé na žiadnom konkrétnom komunikačnom štandarde (s výnimkou HL7 broker-a), aj keď LIM vychádza RIM HL7 v3. Súčasné moderné komunikačné štandardy majú niektoré dôležité charakteristiky spoločné: základný referenčný model, užívateľsky definované modely odvodené z tohto referenčného modelu pomocou striktnej metodiky, a napokon taktiež istý druh šablóny pomáhajúcich pri vytváraní novej správy alebo dokumentu. Porovnanie súčasných komunikačných štandardov možno nájsť v [27], [28] a [29].

Po vyhodnotení výsledkov projektu „Informačné technológie pre rozvoj kontinuálnej zdieľanej zdravotnej starostlivosti“ môžeme povedať, že HL7 v3 má mierne obmedzenú použiteľnosť v prostredí českého zdravotníctva (z



Obr. 7. Vpravo proces vývoja správ doporučený štandardom HL7 v3 a vľavo naše riešenie.

absencie českej meny v slovníkoch HL7). V súčasnej dobe nie je podporovaný vládnymi inštitúciami a len obmedzená podpora pochádza od výrobcov softvéru. Hlavným krokom k širšiemu využívaniu HL7 v3 v Českej republike by mal byť preklad a včlenenie NČLP medzi kódovacie systémy podporované štandardom HL7, alebo namapovanie NČLP na nejakú zavedenú medzinárodnú nomenklatúru akou je SNOMED CT. Ďalší zásadným krokom by bolo preloženie medzinárodných nomenklatúr do češtiny.

Podakovanie

Tento článok bol podporený projektom 1ET200300413 Akadémie vied ČR a projektom 1M06014 Ministerstva školstva ČR.

Referencie

- [1] Pirnejad H., Niazkhani Z., Berg M., Bal R.: Intra-organizational communication in healthcare Considerations for standardization and ICT application. Method Inform Med. 2008; 47(4): 336-345.
- [2] Ministerstvo zdravotníctva ČR (domovská stránka), Datový Standard MZ ČR DASTA a NČLP. Dostupný na: <http://ciselniky.dasta.mzcr.cz> (citované 2009.01.26).
- [3] Health Level Seven, Inc. (domovská stránka) Health Level 7. Dostupný na: <http://www.hl7.org> (citované 2009.01.26).
- [4] European Committee for Standardization (CEN), Technical Committee CEN/TC 251: European Standard EN 13606, „Health informatics Electronic health record communication“
- [5] NEMA Medical Imaging & Technology Alliance (domovská stránka), DICOM. Dostupný na: <http://dicom.nema.org>
- [6] IZIP s.r.o. (domovská stránka), Internetový prístup ke zdravotním informacím pacienta. Dostupný na: <http://www.izip.cz>
- [7] Mead Ch. N.: Data Interchange Standards Healthcare IT - Computable Semantic Interoperability: Now Possible but Still Difficult, Do We Really Need a Better Moustrap? J. of Healthcare Information Management 2006; 20:71-8..
- [8] Tomeckova M. et al.: Minimální datový model kardiologického pacienta. Cor et Vasa 2002; 4: 123.
- [9] Medicalc Software s.r.o. (domovská stránka), WinMedicalc 2000. Dostupný na: <http://medicalc.cz/winmedicalc>
- [10] Mares R., Tomeckova M., Peleska J., Hanzlicek P., Zvarova J.: Uživatelská rozhraní pacientských databázových systémů - ukážka aplikace určené pro sběr dat v rámci minimálního datového modelu kardiologického pacienta. Cor et Vasa 2002; 4: 76.
- [11] Nagy M., Hanzlicek P., Preckova P., Kolesa P., Misur J., Dioszegi M., Zvarova J.: Building Semantically Interoperable EHR Systems Using International Nomenclatures and Enterprise Programming Techniques. In: Blobel B, Pharow P, Zvarova J, Lopez D, editors. eHealth: Combining health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge. CeHR 2007: Conference Proceedings; 2007 Dec 2-5; Regensburg, Germany. Amsterdam: IOS press; 2008. p. 105-10.

- [11] Nagy M., Hanzlicek P., Preckova P., Kolesa P., Misur J., Dioszegi M., Zvarova J.: Building Semantically Interoperable EHR Systems Using International Nomenclatures and Enterprise Programming Techniques. In: Blobel B, Pharow P, Zvarova J, Lopez D, editors. eHealth: Combining health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge. CeHR 2007: Conference Proceedings; 2007 Dec 2-5; Regensburg, Germany. Amsterdam: IOS press; 2008. p. 105-10.
- [12] Hinchley A.: HL7 V3 Guide July 27, 2009. Available from:<http://www.hl7.org/v3ballot/html/help/v3guide/v3guide.htm>
- [13] Hinchley A.: Understanding Version 3 A primer on the HL7 Version 3 Communication Standard. Understanding HL7 Series. 3rd ed. Munich: Alexander Moench Publishing; 2005.
- [14] Lopez D. M., Blobel G. M. E.: A development framework for semantic interoperable health information systems. Int. J. of Medical Informatics 2009; 78:83-103.
- [15] openEHR (domovská stránka), openEHR future-proof and flexible EHR specifications. Dostupný na: <http://www.openehr.org>
- [16] European Committee for Standardization (CEN), Technical Committee CEN/TC 251: European Standard ENV 13606-1, „Health informatics Electronic healthcare record communication"
- [17] Spahni S., Lovis C., Mercille R., Verdel H., Cotten M., Geissbuhler A.: Implementing a new ADT based on the HL7 version 3 RIM. Int J of Med Inform. 2007; 76(2-3):190-4. Amsterdam: IOS press; 2008. p. 105-10.
- [18] Carlsson G., Hagberg T., Bogavac L., Wintell M.: How HL7 version 3 is used at the Sahlgrenska University Hospital to exchange information with a central archive. Pro Biomed Opt Imag - Proceedings of SPIE. 2008; vol.6919, art. no. 691902.
- [19] Orgun B., Vu J.: HL7 ontology and mobile agents for interoperability in heterogeneous medical information systems. Comput Biol Med. 2006; 36(7-8): 817-836.
- [20] Logical Observation Identifiers Names and Codes - LOINC (homepage on the Internet). Available from: <http://loinc.org>
- [21] SNOMED International (homepage on the Internet), Available from: <http://www.snomed.org>
- [22] WHO (homepage on the Internet), International Classification of Diseases. Available from: <http://www.who.int/classifications/icd/en>
- [23] US NIH National Library of Medicine (homepage on the Internet), UMLS Knowledge Source Server. Available from: <http://umlsks.nlm.nih.gov>
- [24] Hanzlicek P., Preckova P., Zvarova J.: Semantic Interoperability in the Structured Electronic Health Record. Ercim News 2007, 69:52-3.
- [25] Garde S., Knaup P., Hovenga E. J. S., Heard S.: Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for openEHR Archetypes. Method Inform Med. 2007; 46(3): 332-343.
- [26] The openEHR Foundation (homepage on the Internet), Relationship between openEHR Specifications and HL7v3 Standards. Available from: <http://www.openehr.org/206-OE.htm>
- [27] Blobel B. G., Engel K., Pharow P.: Semantic interoperability HL7 Version 3 compared to advanced architecture standards. Method Inform Med. 2006; 45(4):343-53.
- [28] Lopez D. M., Blobel G.: Enhanced Semantic Interoperability by Profiling Health Informatics Standards. Method Inform Med. 2009; 48(2): 170-177.
- [29] Eichelberg M., Aden T., Riesmeier J., Dogac A., Laleci G. B.: A Survey and Analysis of Electronic Healthcare Record Standards. ACM Comput Surv. 2005; 37(4): 277 - 315.



EJBI 2009

ISSN 1801 - 5603

European Journal for Biomedical Informatics

Volume 5 (2009), Issue 1

Spanish version

www.ejbi.eu



Content

Spanish version

es 2 - 9

[Arquitecturas para la Implementación de Sistemas de Información en Salud Basadas en el Estándar HL7](#)

D. M. López, B. Blobel

Arquitecturas para la Implementación de Sistemas de Información en Salud Basadas en el Estándar HL7

D. M. López^{1,2}, B. Blobel¹

¹eHealth Competence Center, University of Regensburg Medical Center, Germany,

²Telematics Research Group, University of Cauca, Colombia

Objetivo: El proceso de integración de sistemas de información en salud es complejo, especialmente cuando se deben cumplir requisitos de interoperabilidad en los niveles semántico y de la lógica del negocio. Para tener éxito en un proyecto de integración de estas características, es fundamental el uso de una metodología de desarrollo unificada que permita analizar de forma separada diferentes aspectos de la arquitectura del sistema como por ejemplo los procesos de la organización, la información a ser procesada, la distribución de componentes, así como aspectos específicos de las tecnologías a emplear. Este artículo contribuye con un análisis detallado y una demostración de cómo el estándar HL7 puede ser usado para soportar el desarrollo de proyectos de integración de sistemas de información en salud.

Métodos: La metodología de desarrollo unificada a utilizar es el Marco Integral de Desarrollo para Sistemas de Información en salud (HIS-DF). HIS-DF es una metodología basada en diferentes estándares, genérica, personalizable y escalable; que soporta el diseño de sistemas de información semánticamente interoperables. Basados en HIS-DF, se analizan tres modelos genéricos para la integración de sistemas de información en salud.

Resultados: Los modelos arquitectónicos para integración de sistemas analizados son: Interfaz Punto a Punto, Servidor de Mensajes y Mediación. Los modelos de Interfaz Punto a Punto y Servidor de Mensajes son completamente soportados por sistemas tradicionales de mensajería HL7 versión 2 y versión 3. El estándar HL7 versión 3, combinado con tecnologías orientadas a servicios y la metodología HIS-DF, hacen posible el modelo de Mediación. Los diferentes escenarios de integración son ejemplificados mediante el desarrollo de un sistema de vigilancia de la

salud pública basada en la tecnología de Enterprise Java Beans.

Conclusión: La selección de la arquitectura de integración es una decisión fundamental de cualquier proyecto de desarrollo de software. HIS-DF provee un enfoque metodológico único para guiar el desarrollo de proyectos de integración de sistemas de información en salud. El modelo de Mediación ofrecido por la metodología HIS-DF apoyada en artefactos del estándar HL7 versión 3 es el modelo más prometedor de los tres analizados, pues promueve el desarrollo de sistemas de información en salud abiertos, reutilizables, flexibles, semánticamente interoperables, independientes de la plataforma, orientados a servicios y basados en estándares.

Palabras Claves: Integración de sistemas, arquitectura de sistemas, interoperabilidad semántica, HL7, Middleware, J2EE.

1. Introducción

La integración de sistemas de información heredados con nuevos sistemas de información es una práctica muy común en el ámbito de los sistemas de información en salud, debido principalmente a las necesidades de migración de datos y a la propia evolución de los sistemas. Este proceso de integración es bastante complejo especialmente cuando se debe ofrecer soluciones de interoperabilidad en el nivel semántico y de servicios. Restricciones políticas y administrativas, escasa documentación, la necesidad de migrar grandes cantidades de datos que han sido recolectados durante años, el impacto en otros sistemas con los que se comparte la información, y la disponibilidad de personal capacitado para desarrollar o mantener las tecnologías de integración [1] , [2], son los principales factores que afectan dicha decisión.

Para tener éxito en un proyecto de integración de estas características, es muy importante el uso de una metodología unificada de desarrollo de sistemas de información en salud (SIS) que permita analizar de forma separada diferentes aspectos de la arquitectura como los procesos de la organización, la información a ser procesada, la distribución de componentes, así como aspectos específicos de las tecnologías a emplear [3], [4]. Las metodologías de desarrollo software existentes, así como nuevas herramientas y tecnologías soportan el desarrollo de soluciones de interoperabilidad. Sin embargo, para alcanzar interoperabilidad semántica, algunos estándares para modelar la información en salud deben ser especialmente considerados en el proceso de desarrollo. Algunas alternativas en estandarización incluyen por ejemplo los arquetipos de openEHR, sin embargo los autores de este artículo creen que la familia de estándares HL7 versión 3, y especialmente algunos elementos (artefactos) del Marco Integral de Desarrollo de HL7 (HL7 Development Framework, HDF), pueden ser utilizados de manera innovadora para apoyar el desarrollo aplicaciones en salud basadas en una arquitectura orientada a servicios. El objetivo de este trabajo es analizar y demostrar, desde el punto de vista de su implementación (esto es aspectos computacionales y tecnológicos de la arquitectura del sistema), la forma en que el conjunto de especificaciones HL7 pueden ser usados para soportar el desarrollo de sistemas de información en salud, soportados además en los paradigmas del desarrollo dirigido por modelos (Model-Driven Development, MDD), el desarrollo basado en componentes (component-based development) y la arquitectura orientada a servicios (Service Oriented Architecture, SOA).

2. Métodos y Materiales

El análisis de diferentes enfoques arquitectónicos basados en estándares HL7, se realiza teniendo en cuenta los modelos más comunes para la integración de sistemas propuestos en el dominio de los sistemas de información. Basados en estos modelos, el artículo analiza un escenario de integración en salud considerando los diferentes modelos arquitectónicos, centrado principalmente en aspectos de implementación del sistema. La implementación del escenario de integración está guiada por el Marco Integral de Desarrollo para Sistemas de Información en Salud (HIS-DF). A continuación se describen brevemente los modelos de integración y la metodología HIS-DF.

2.1 Modelos Comunes para la Integración de Sistemas de Información

Algunos enfoques que describen arquitecturas para la integración de sistemas de información han sido propuestos en la literatura, por ejemplo [5], [6], [7]. La Fig. 1 resume los más importantes, considerando la integración de un par de sistemas de información.

El modelo *Interfaz Punto a Punto* (Point to Point Interface) define un canal de comunicación directa entre cada par de aplicaciones. Este modelo es apropiado cuando un sistema heredado debe ser integrado a unos pocos sistemas de información. La extensibilidad del sistema, es decir, su capacidad de ser modificado para aumentar su capacidad funcional [8], se ve comprometida en este modelo porque para interconectar (n) aplicaciones, es necesario crear $n * (n-1)/2$ interfaces.

El modelo *Servidor de Mensajes* (Messages Server) está basado en la utilización de un enrutador de mensajes el cual actúa como un "puente", encargado de la recepción, mapeo y distribución de los mensajes intercambiados entre las aplicaciones. Este enfoque aumenta la extensibilidad de la solución porque cada vez que se desee integrar un nuevo sistema de información, este puede integrarse de manera sencilla mediante la adición de una sola interfaz hacia el servidor de mensajes. El principal

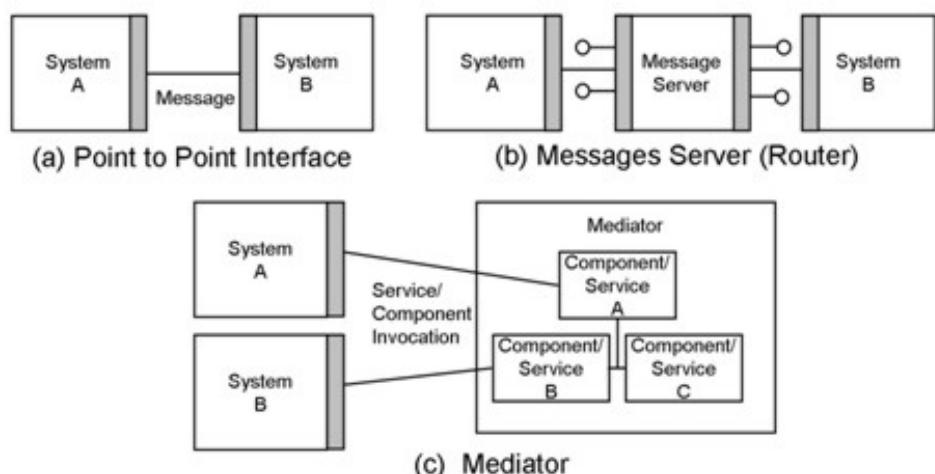


Fig. 1 Modelos Comunes para la Integración de Sistemas Heredados.

problema con este modelo es que toda la lógica de integración (enrutamiento, mapeo, transformación) está soportada en el Servidor de Mensajes, lo que dificulta la ampliación o creación de nuevas funcionalidades del sistema.

El modelo de *Mediación* (Mediator) está basado en un servicio central que gestiona las peticiones de las aplicaciones a ser integradas. A diferencia del modelo de Servidor de Mensajes, el modelo de Mediación expone su funcionalidad (lógica de negocio) como un conjunto de servicios o componentes que pueden ser invocados por las diferentes aplicaciones, siendo capaces de compartir su funcionalidad, mientras que esconden su implementación interna. Este modelo abstrae los conceptos de SOA o de Middleware. Un Middleware, desde una perspectiva tecnológica, define las interfaces y los protocolos entre aplicaciones distribuidas en un entorno empresarial, que se ejecutan a su vez en múltiples plataformas [9]. Por otro lado, SOA sería un estilo arquitectónico para el Mediador, definiendo el modelo de orquestación de servicios y la interoperabilidad.

Entre los diferentes modelos de integración, el modelo de Mediación es el enfoque más flexible, el que permite mayor reuso, así como la extensibilidad, mantenimiento, y garantiza la interoperabilidad semántica de los servicios /aplicaciones. Sin embargo por su complejidad y costos de implantación, este enfoque es efectivo sólo en soluciones empresariales de gran escala.

2.2 El Marco Integral de Desarrollo para Sistemas de Información en Salud

HIS-DF es una metodología genérica y personalizable para el análisis arquitectónico de sistemas de información en salud, desarrollado recientemente en el centro de investigación eHealth Competence Center Regensburg [4].

HIS-DF consta principalmente de una base de conocimiento de referencia que incluye las diferentes tareas, roles (miembros del equipo de desarrollo encargados de ejecutar las tareas), productos (artefactos resultantes de cada tarea), guías, fases, flujos de trabajo, etc., necesarios para el análisis de la arquitectura del sistema en proyectos de desarrollo de software de salud, especialmente orientado al desarrollo de soluciones que soporten interoperabilidad semántica. Esta metodología se basa en los principios de la teoría de sistemas tomando como referencia el modelo de componentes genéricos (GCM) [10]. Este enfoque permite modelar el comportamiento estático de los SIS como una agregación de componentes, así como su comportamiento dinámico, descrito como la interacción entre estos componentes. HIS-DF aborda también la complejidad intrínseca de sistemas de información en salud mediante un análisis por separado los diferentes dominios cubiertos por el sistema de información (por ejemplo separando el dominio médico del técnico),

además de ofrecer un método para abstraer los diferentes puntos de vista desde los que se puede describir la arquitectura de un sistema de información. Esta abstracción está soportada en el uso del modelo de referencia del estándar abierto para el procesamiento distribuido ISO 10746 (Open Distributed Processing Reference Model, RM-ODP) [11], el cual define cinco puntos de vista (viewpoints) y un lenguaje para cada vista, con los cuales se puede describir de manera formal los diferentes componentes de un sistema de información. Además, la Arquitectura Dirigida por Modelos (MDA) es también empleada dentro de este marco de desarrollo definiendo un proceso para la transformación de los modelos de la organización a un modelo independiente de la plataforma, y finalmente su transformación a un modelo específico de la plataforma, este último describiendo aspectos tecnológicos de implementación. En la metodología, la implementación de los modelos específicos de la plataforma hace parte del proceso de desarrollo a utilizar (e.g., RUP), sin embargo HIS-DF sugiere el uso de métodos de desarrollo dirigido por modelos (MDD) para facilitar portabilidad a cualquier entorno de desarrollo. Para permitir el uso y adaptación del MDD, unos Perfiles para el Lenguaje Unificado de Modelado (Unified Modeling Language, UML) son aplicados a los modelos de información HL7 [12]. Para una descripción completa de la metodología, se puede consultar las siguientes referencias [4], [13].

3. Análisis del Uso de la Familia de Estándares HL7 en el Diseño de Arquitecturas de Sistemas de Información en Salud

Un análisis de HL7 comparado con otras aproximaciones arquitectónicas como CORBA y DHE [5], [14] por un lado, y su análisis comparado con estándares para registros electrónicos en salud [15] por el otro, ha sido discutido previamente en la literatura. Para realizar un análisis de HL7 desde la perspectiva de implementación de sistemas de información en salud como el que se realiza en este artículo, la reciente Versión 3 de HL7 ha de tenerse en cuenta. Siguiendo las recomendaciones de la metodología HIS-DF, HL7 versión 3 servirá para apoyar el modelado de SIS

apoyado en su Modelo de Referencia de Información (RIM) y el conjunto de modelos de información derivados del mismo (e.g., DIM, CIM). Los modelos de información representan el universo de discurso de los mensajes intercambiados por los sistemas en sus respectivos dominios (e.g., administración de pacientes, laboratorios, salud pública, etc.). Es importante también mencionar que los artefactos definidos en la metodología de desarrollo de HL7 versión 3 (HDF) como por ejemplo diagramas de estado (state diagrams), eventos disparadores (trigger events) y las interacciones (interactions); pueden ser utilizados para describir el comportamiento dinámico de la aplicación, así como detalles funcionales del sistema. Sin embargo, más allá de las especificaciones para tecnologías de implementación (Implementable Technology Specifications, ITS), el estándar HL7 versión 3 no prescribe formalmente ninguna solución arquitectónica para la interoperabilidad de los sistemas de salud, ni tampoco una arquitectura de referencia o middleware de integración. Los estándares HL7 que más se acercan una solución tecnológica para la integración de aplicaciones son la especificación para el transporte de mensajes HL7 (HL7 message transport specifications), las cuales proporcionan guías de cómo trasportar mensajes HL7 usando tecnologías de comunicación bien conocidas (por ejemplo Web Services y ebXML); así como algunas especificaciones desarrolladas en el grupo de trabajo interno a HL7 sobre Arquitectura Orientada a Servicios (SIG SOA).

Concretamente, el uso de Servicios Web como transporte se define en la especificación Perfil para Servicios Web (HL7 Web Service Profile, WSP) [16]. WSP restringe el uso de protocolos de Servicios Web como una capa de transporte para el intercambio de mensajes HL7 versión 3. En esta especificación no se considerada, sin embargo, la especificación de la funcionalidad de los servicios (servicios del negocio), las interfaces, las operaciones y el gobierno de servicios.

Por su parte, SIG SOA es un grupo de trabajo comprometido con la definición y documentación de especificaciones funcionales de servicios de e-salud,

tratando de asegurar al máximo su conformidad con el estándar HL7. Este enfoque se basa en un paradigma diferente al del WSP, evolucionando desde la mensajería tradicional HL7 hacia la definición de una arquitectura SOA. El SOA SIG define una metodología para la especificación de servicios, y un conjunto de especificaciones funcionales para servicios básicos como el Servicio de Identificación de Entidades (Entity Identification Service, EIS), el servicio de Recuperación, Localización y Actualización (Retrieve, Locate, Update Service, RLUS), y el Servicio de Apoyo a la Toma de Decisiones (Decision Support Service, DSS). Cabe destacar que SOA SIG solamente desarrolla especificaciones funcionales al igual que lo hace por ejemplo el bastante difundido modelo funcional para registros clínicos electrónicos (EHR-S Functional Model Specification). La especificación de los servicios es responsabilidad de la OMG (Object Management Group). SOA SIG tiene por tanto un objetivo más amplio que WSP proveyendo una solución basada en servicios.

SOA SIG también viene trabajando en la definición de una arquitectura SOA de referencia para servicios de e-Salud, la cual define entre otras cosas una taxonomía, y las normas de información y referencia para la implementación de las especificaciones funcionales [17]. Esta arquitectura de referencia ofrecerá un marco genérico de referencia para la facilitar la interoperabilidad de servicios de e-Salud, definiendo una conjunto de servicios genéricos y reusables que puedan ser especializados (adaptados) a proyectos específicos de e-Salud (tecnologías de la información aplicada a salud).

4. El Rol de HL7 en los Diferentes Modelos de Integración

Con el objetivo de demostrar el papel de HL7 versión 3 en los modelos de integración anteriormente expuestos, diferentes alternativas para una solución de integración de sistemas de información heredados en el contexto del Sistema de Información Integrado de Salud en Colombia (SIIS) [18] se analizan a continuación.

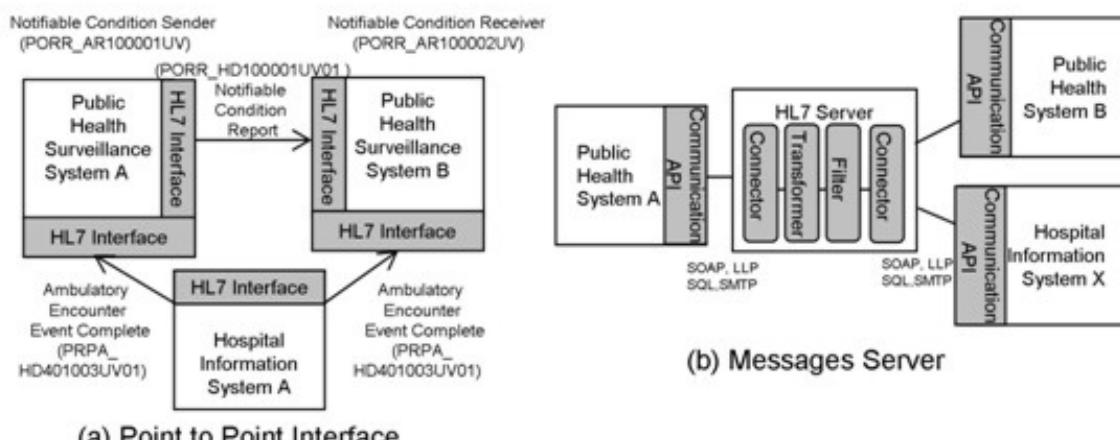
En Colombia, especialmente en la última década, varios sistemas de información computarizados han sido desarrollados para apoyar el Sistema Nacional de Vigilancia en Salud Pública (SIVIGILA) [19]. Estos sistemas son principalmente aplicaciones de escritorio, que usan diferentes tecnologías y son a su vez diversos en cuanto a funcionalidad. Recientemente, el Instituto Nacional de Salud (INS) como la oficina gubernamental encargada de la gestión del sistema SIVIGILA, ha diseñado, desarrollado y desplegado una aplicación cliente/servidor (denominada SIVIGILA 2007) para apoyar la recolección y reporte de enfermedades transmisibles a las autoridades de salud pública. La situación actual es que todas estas aplicaciones heredadas funcionan de forma totalmente

independiente uno de la otra, salvo algunos pocos sistemas que son capaces de intercambiar alguna información en forma de archivos de texto que son importados manualmente. La idea detrás de la propuesta de integración en este artículo, es apoyar el proceso de notificación de eventos de salud pública a las autoridades de salud locales, mediante el desarrollo de un componente de integración que permita la captura de información epidemiológica necesaria por el sistema SIVIGILA directamente de los registros clínicos existentes en los sistemas de información hospitalarios (por ejemplo los llamados Registro Individual de Prestación de Servicios de Salud, RIPS). Ninguno de los sistemas de información existentes actualmente soporta esta funcionalidad.

Considerando los modelos de integración discutidos en la sección anterior, y soportados en las especificaciones HL7 disponibles, las diferentes alternativas para la integración de aplicaciones heredadas para el escenario de integración del sistema de salud colombiano se analizan a continuación.

4.1 Interfaz Punto a Punto

El modelo de Interfaz Punto a Punto puede ser implementado utilizando mensajes HL7 versión 3 como se muestra en la Fig. 2(a). Para conectar los diferentes sistemas de información heredados que hacen parte del sistema de vigilancia en salud pública colombiano, cada sistema debe crear una interfaz HL7 con el fin de permitir el intercambio de información sobre eventos de salud pública.



(b) Messages Server

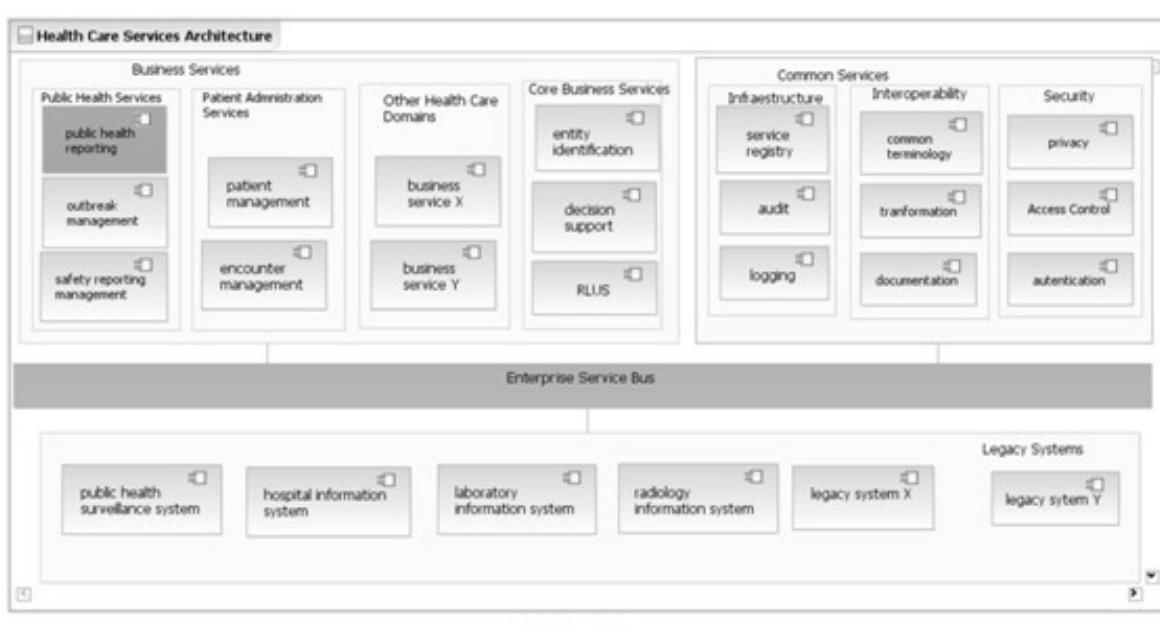
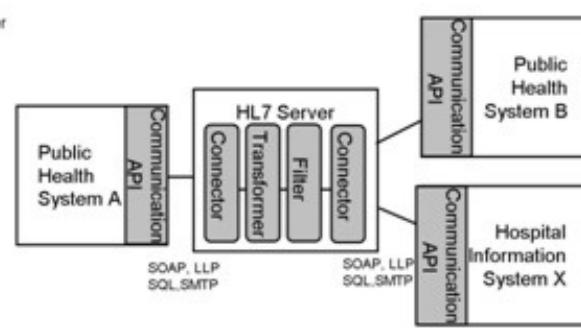


Fig. 2 Aproximaciones de Integración Comunes Usando HL7.

Un mensaje HL7 que puede soportar dicho intercambio de información es el Notifiable Condition Report HMD (PORR_HD100001UV01), definido dentro del Dominio HL7 de salud pública. Así mismo, para comunicar información de diagnósticos clínicos desde los cuales pueda extraerse la información de posibles casos (eventos) de salud pública, el mensaje HL7 Ambulatory Encounter Event Complete (PRPA_HD401003UV01) puede ser utilizado. Cada interfaz HL7 puede ser implementada usando interfaces de programación de aplicaciones (API) para HL7 como por ejemplo Java SIG API [20], que permite analizar y construir mensajes HL7 versión 3. Los mensajes HL7 pueden ser transferidos mediante cualquier protocolo tradicional de transporte (HTTP, TCP, UDP, SMTP, etc.) y usando mecanismos de transporte como por ejemplo una infraestructura de mensajería basada en Servicios Web.

4.2 Servidor de Mensajes

El modelo de servidor de mensajes de la Fig. 2(b) simplifica el intercambio de mensajes entre sistemas de vigilancia de la salud pública y sistemas de información clínica, mediante el uso de servidores HL7. La lógica necesaria para transformar y enrutar mensajes reside en el interior del servidor de mensajes, junto con los conectores a sistemas heredados. Varios productos comerciales están disponibles en el mercado que actúan como servidores de mensajes HL7, por ejemplo, Interfaceware Iguana, HL7Connect, Neointegrate, iWay, Cloverleaf, etc. Un servidor HL7 de código abierto muy prometedor es el ofrecido por el proyecto Mirth [21]. Este sistema ofrece funcionalidades de filtrado, transformación y enrutamiento de mensajes; además que recientemente ha incorporado dentro de sus funcionalidades la gestión de mensajes HL7 versión 3. Mediante configuraciones muy simples, es posible la comunicación de mensajes a través de protocolos LLP, SQL, JMS, SOAP Web Services, y FTP.

4.3 Mediación

La Fig. 2(c) muestra una solución basada en el modelo de Mediación. Esta aproximación está basada en una federación de servicios de acuerdo con la

arquitectura de referencia SOA para servicios de salud propuesta por HL7. En la arquitectura propuesta en este artículo, los servicios se interconectan a través de una solución de Bus de Servicios Empresariales (Enterprise Service Bus, ESB), que proporciona además acceso a/desde las aplicaciones heredadas. Los denominados Servicios del Negocio (Business Services) están agrupados de acuerdo los diferentes dominios de salud definidos en las especificaciones HL7 (HL7 Universal Domains) como por ejemplo, Salud Pública, Administración de pacientes, Atención en Salud, Contabilidad y Facturación, etc. Aquí, las especificaciones funcionales de servicios HL7 existentes se utilizan para definir los Servicios Básicos de Negocios (Core Business Services) incluyendo EIS, RLUS, y DSS. Finalmente, los Servicios Comunes (Common Services) incluyen servicios como Seguridad, Interoperabilidad, Servicios de Infraestructura.

Volviendo a la arquitectura del Sistema Integrado de Información Nacional de Colombia, el servicio propuesto para el apoyo a la notificación de eventos de salud pública es el Servicio de Salud Pública, incluido dentro de los servicios del negocio (representado en la figura en color gris oscuro). Este servicio puede ser invocado por los sistemas de información hospitalarios para ejecutar automáticamente la notificación obligatoria de los eventos de salud pública. De esta forma también los sistemas de vigilancia en salud existentes (por ejemplo, SIVIGILA2007) pueden invocar este servicio para efectuar el reporte de eventos en salud pública a la autoridad de salud pública correspondiente (por ejemplo, el sistema regional de salud pública). Esta arquitectura provee además una integración completa con otros sistemas fuera del sistema de salud pública, por ejemplo, proveyendo interconexión con sistemas de información de laboratorio para efectuar la confirmación clínica de casos dudosos. De esta forma, los servicios comunes se asegurarían interoperabilidad semántica, e interoperabilidad a nivel de servicios.

5. La Implementación del Proyecto de Integración

El modelo de Mediación es el más avanzado entre los enfoques anteriormente investigados dado que soporta interoperabilidad semántica, ofrece una arquitectura orientada a servicios, es flexible, abierto y facilita la conformidad con estándares de información en e-salud. Sin embargo, y especialmente para el escenario de integración expuesto, debe existir un equilibrio entre la disponibilidad y la complejidad de la solución de integración. Siendo el caso de países en desarrollo como Colombia, la principal desventaja para la implantación de una arquitectura robusta como la del modelo Mediador sería la necesidad de desarrollar (o adquirir) en primera instancia los servicios del negocio.

La solución de integración propuesta para el escenario Colombiano combina en una sola arquitectura los modelos de integración de Servidor de Mensajes y de Mediación. Se propone del modelo Servidor de Mensajes como una solución simple y eficiente a mediano plazo basada en la implementación de una interfaz HL7 versión 3 para la interoperabilidad con sistemas de información hospitalarios heredados. Junto a esta solución, se desarrolla un servicio de Salud Pública conforme al modelo de Mediación usando HL7 presentado en la Fig. 2 (c). En el futuro, cuando este servicio pueda ser desplegado y otros Servicios del Negocio y Servicios Básicos estén disponibles, esta solución puede ser integrada en una arquitectura ESB. Si se desean conocer los detalles de la plataforma independiente de la arquitectura utilizada para desplegar el Servicio de Salud Pública, puede consultarse la referencia [12]. En las subsecciones que siguen, se describen los detalles de implementación de una plataforma independiente de la arquitectura diseñada estrictamente usando la metodología HIS-DF.

5.1 La Solución de Integración en el Contexto de HIS-DF

Tal como es definido en HIS-DF, para la fase de implementación de la solución se ejecuta mediante la transformación de la arquitectura del sistema descrita en el modelo computacional (Modelo Independiente).

diente de Plataforma), a los respectivos modelos de ingeniería y tecnológicos, tal como se detalla en la Fig. 3.

En la figura, el modelo de componentes derivado del punto de vista computacional de HIS-DF se transforma en un modelo de componentes Java que representa el punto de vista de ingeniería. Esta transformación está soportada en el modelo de desarrollo dirigido por modelos (MDD) concretamente mediante el uso de transformaciones UML modelo a modelo y su posterior transformación en código fuente usando perfiles UML, por ejemplo, Perfiles UML para Web Services o Enterprise Java Beans (EJB).

Hay dos grandes decisiones que hay que tomar al abordar el punto de vista de ingeniería. La primera es elección de la plataforma de implementación (por ejemplo, CORBA, J2EE, Microsoft® COM) y la segunda corresponde a los estilos arquitectónicos (por ejemplo SOA, Modelo-Vista-Controlador (MVC) y N-nivel).

La plataforma debe apoyar la interacción distribuida entre los componentes del sistema, es decir, el apoyo a la gestión de transacciones, la persistencia entre objetos, multihilos, la seguridad y el repositorio de los datos. La plataforma seleccionada para el escenario de integración del sistema de información de salud pública es J2EE, debido a su gran aceptación de la industria, la disponibilidad de servicios de código fuente abierto, además de la experiencia previa del equipo de desarrollado en programación de EJBs. La plataforma tecnológica usada es IBM Websphere debido a su disponibilidad dentro del proyecto, además de la plena compatibilidad con la herramienta de modelado utilizado a lo largo del proyecto, es decir, IBM Rational Software Architect (RSA).

El estilo arquitectónico utilizado se ve reflejado en el modelo específico de la plataforma (Platform Specific Model, PSM) de la Fig 3(b). Esta arquitectura se basa en el MVC y la arquitectura multinivel definiendo al menos cuatro niveles: nivel de presentación, nivel del negocio, nivel de datos y nivel de integración.

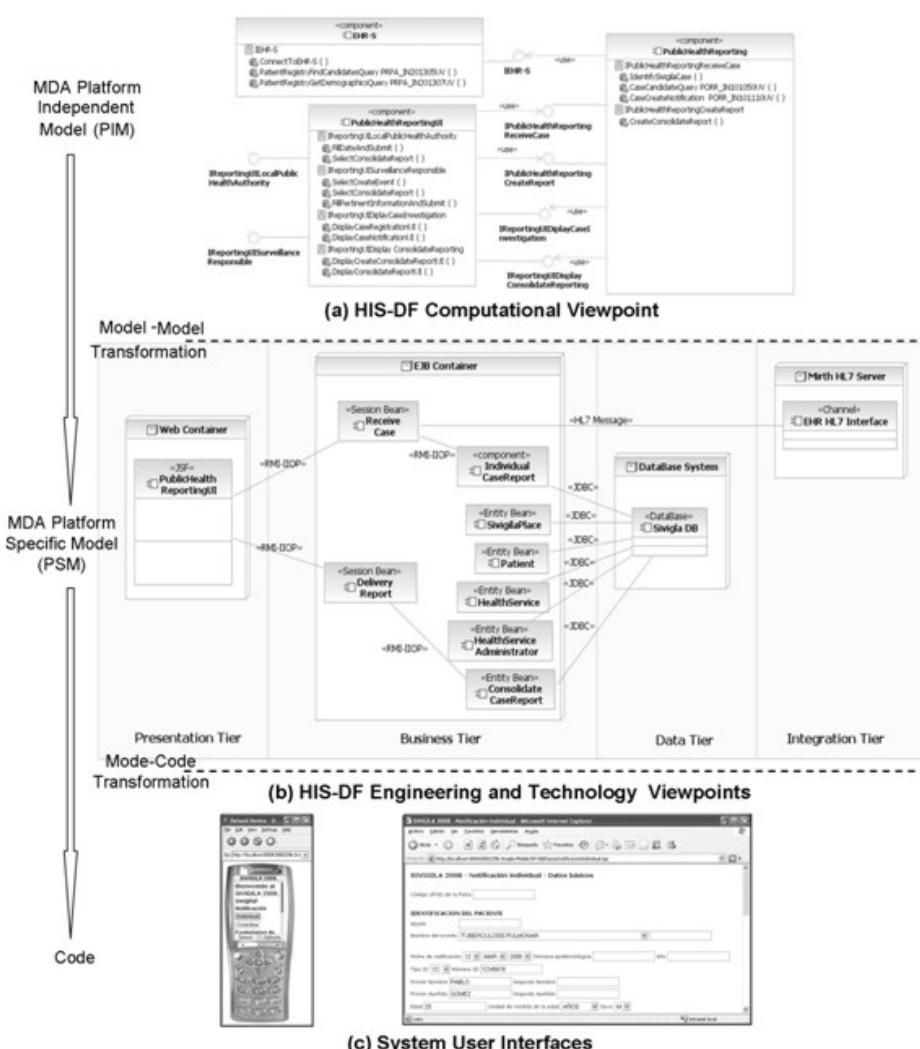


Fig. 3 Esquema del Proceso de Implementación.

5.2 La Arquitectura Multinivel

En el modelo de la Fig. 3 (b), el nivel de presentación (presentation tier) representa las interfaces de usuario implementadas utilizando la tecnología Java Server Faces (JSF). JSF es un conjunto de APIs de Java que permite desarrollo de interfaces de usuario flexibles e independientes del dispositivo de visualización, además que provee una separación entre la lógica de negocio y el nivel de presentación. JSF es importante en el contexto del proyecto, ya que permite crear fácilmente dos interfaces diferentes a partir de una sola página JSF: una para clientes Web tradicionales mediante una librería JSP (Java Services Pages) para la presentación en HTML, y la otro para clientes que usan dispositivos móviles mediante una librería JSF para la transformación a documentos WML (Wireless Markup Language).

El nivel de negocio (business tier) comprende la lógica de aplicación tal como se describe en el PIM del Servicio de Salud Pública (véase el diagrama de componentes de la Fig. 3 (a)). La lógica se implementa como un conjunto de Java Beans de Sesión y Entidad (Session and Entity Beans). Los Beans de Sesión "Receive Case" y "Delivery Report" encapsulan las operaciones necesarias para implementar la funcionalidad descrita en los respectivos Casos de Uso del servicio de Salud Pública. Los EJBs restantes son objetos que proveen persistencia al modelo de datos definido en el nivel de datos.

El nivel de datos (data tier) se implementa usando Apache Derby, que es una base de datos relacional de código abierto en Java. Esta base de datos almacena de forma permanente los objetos que representan la

información sobre eventos de salud pública (Casos) del sistema SIVIGILA, por ejemplo, Pacientes, Servicios de Salud, Lugar del Evento, Reporte de Caso y Reporte Consolidado (Patient, Health Service, Place, CaseReport, and ConsolidateReport).

Finalmente, el nivel de integración (integration tier) incluye las interfaces para la interoperabilidad del sistema con sistemas de información hospitalarios heredados. En el escenario, el EJB denominado "Receive Case" hace uso de la interfaz HL7 "EHR-S HL7 Interface" para intercambiar mensajes HL7 versión 3 con otros sistemas. Esta interfaz se implementa como un conjunto de canales usando el servidor Mirth 1.7, implementando así el modelo de integración de Servidor de Mensajes.

5.3 Transformaciones usando MDD

En el proyecto se utilizaron las funcionalidades de transformación de la herramienta RSA, con el fin de apoyar el proceso de transformación de modelos UML a código Java. La transformación de PIM a PSM (transformación del modelo de la fig. 3 (a) al modelo de la fig. 3 (b)) utiliza el perfil UML para EJB incluido en la herramienta RSA. Además, unos Perfiles UML especialmente creados para el HL7 RIM y algunos DIMs se utilizaron con el fin de "anotar" los modelos UML con la semántica de los modelos de información HL7 [13]. La herramienta de modelado RSA genera el esqueleto del modelo de clases y su correspondiente código Java para los Beans de Entidad, incluyendo de manera especial las operaciones y tipos de datos definidos por HL7. Cabe aclarar que la lógica de las operaciones es codificada posteriormente, de acuerdo a los requerimientos específicos de la aplicación. Detalles sobre esta transformación y los perfiles utilizados se describen en [13].

La comunicación entre los Beans de Entidad y la base de datos relacional es soportada mediante una transformación (mapping) entre los elementos del EJB, y los elementos del Modelo Relacional. Además, el código WML y HTML para las interfaces de usuario también fue generado de manera automática mediante transformaciones a partir de los

documentos JSF. Fig. 3(c) muestra una de las interfaces de usuario resultantes después de la transformación. Estas transformaciones también fueron ejecutadas con la herramienta de modelado de IBM.

6. Discusión

En un proyecto de integración de sistemas de información en salud la decisión acerca de qué estilos arquitectónicos y plataformas son más apropiados, depende en gran medida de las necesidades específicas de la organización y sus limitaciones. En consecuencia, es siempre necesario adaptar la arquitectura del sistema a los requisitos específicos del proyecto. El documento aporta un análisis exhaustivo sobre cómo pueden utilizarse el conjunto de estándares HL7 en la integración de sistemas heredados, sobre la base de algunos modelos genéricos de integración y soportado en aproximaciones arquitectónicas maduras como el desarrollo dirigido por modelos (MDD), el desarrollo basado en componentes y en SOA.

Después de analizar los diferentes modelos de integración, se constató que los modelos de Interfaz Punto a Punto y de Servidor de Mensajes pueden ser completamente soportados por mensajes HL7 versión 2 y versión 3; mientras que el modelo de integración de Mediación es parcialmente apoyado por algunos modelos de información del estándar HL7 versión 3. El modelo de integración de Servidor de Mensajes, a pesar de no ser la opción más avanzada, es muy valioso especialmente en proyectos de integración de mediana complejidad, debido a la facilidad ofrecida para implementar de forma rápida interfaces HL7.

Este artículo se ocupa exclusivamente de las perspectivas de la ingeniería y aspectos tecnológicos de las arquitecturas de sistemas de información en salud basados en estándares HL7. Hasta donde es conocido por los autores, estos aspectos no han sido abordados dentro de HL7, ni discutido con anterioridad en la literatura. La arquitectura propuesta se basa en la plataforma EJB que es un modelo de componentes en Java para el desarrollo y

despliegue de componentes orientado a objetos y distribuidos. La lógica de aplicación se implementa como un conjunto de EJBs de Sesión y Entidad, lo que permite la escalabilidad, distribución e interoperabilidad de la solución. El uso de EJBs es también es importante, debido a que estos pueden ser fácilmente agregados para el despliegue de Servicios Web (bottom-up development approach) en la implantación de una arquitectura SOA. Considerando que los EJB implementan una serie de operaciones que intercambian mensajes HL7, los sistemas desarrollados pueden interactuar directamente con sistemas de información heredados que implementen interfaces HL7, evitando así la creación de nuevas interfaces. Los servicios pueden ser fácilmente integrados en una infraestructura de ESB.

La aproximación arquitectónica presentada es avanzada porque facilita el desarrollo de sistemas de información en salud semánticamente interoperables, portables, escalables y basados en estándares. Es evidente que entre más detallada sea la arquitectura independiente de la plataforma (PIM), el esfuerzo necesario su para implementación es menor. Además, el uso de modelos y Perfiles UML en forma consistente facilita la reutilización de la arquitectura.

En cuanto a los desafíos metodológicos y tecnológicos abordados, se ha mostrado como puede completarse el ciclo de desarrollo propuesto en HIS-DF, afrontando la fase de implementación mediante la transformación de una arquitectura independiente de la plataforma en código ejecutable. El entorno de desarrollo utilizado fue la familia de herramientas de IBM Rational Software para el análisis, diseño e implementación de sistemas; junto con una base de datos de código abierto y un servidor de mensajes HL7. Las herramientas de IBM fueron elegidas debido a su disponibilidad, robustez, soporte y documentación. Sin embargo, esta no es el entorno de desarrollo óptimo en proyectos de pequeña y mediana escala con presupuestos limitados. La ventaja de usar las herramientas de IBM es que todas ellas se basan en la plataforma Eclipse,

existiendo siempre variantes de código abierto para las mismas.

El enfoque arquitectónico presentado demuestra la viabilidad de una solución escalable que combina mensajería HL7 tradicional, con una aproximación más innovadora basada en una arquitectura SOA. Una limitación importante de este estudio de viabilidad es el hecho de que debido a la escasez de recursos, y siendo principalmente un trabajo académico, los autores evaluaron por sí mismos los desarrollos software. Por lo tanto la comparación y evaluación que se presenta en este artículo corre el riesgo de estar sesgada. Como trabajo futuro es deseable una evaluación formal preferiblemente empírica a través de un experimento o un estudio de caso. Actualmente se está negociando la colaboración con el Instituto Nacional de Salud en Colombia (INS), entidad encargada de la gestión del sistema SIVIGILA, a fin de completar la aplicación y evaluar los resultados mediante el desarrollo de un estudio de caso.

7. Conclusión

La integración de sistemas heredados implica un análisis detallado de los procesos del negocio y de los requisitos de interoperabilidad. HIS-DF soporta dicho análisis ofreciendo una metodología detallada para el diseño de la arquitectura del sistema, soportado en un proceso de desarrollo dirigido por modelos (MDD). En el artículo se demuestra cómo, a través de un ejemplo de integración, de los paradigmas del MDD, del desarrollo basado en componentes, SOA, y un conjunto de especificaciones de HL7; es posible desarrollar un proyecto completo de integración de sistemas heredados. El análisis está basado en tres modelos que son los más comunes para la integración de sistemas de información.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el servicio de intercambio del gobierno alemán (German Academic Exchange Service, DAAD) y la Universidad del Cauca en Colombia. Los autores agradecen la colaboración de colegas dentro de la organización HL7 y otros organismos de estandarización en e-salud.

Referencias

- [1] Spahni S., Lovis C., Mercille R., Verdel H., Cotten M., Geissbühler A.: Implementing a new ADT based on the HL7 version 3 RIM. *Int. J. Med. Inform.* 2007; 76(2-3): 190-4.
- [2] Makki S.K.: The Integration and Interoperability Issues of Legacy and Distributed Systems. In: *Seventh International Conference on Web-Age Information Management Workshops*; 2006 Jun 17-19; Hong Kong, China. Wahington DC: IEEE Press; 2006. p. 21.
- [3] Mykkanen J., Porrasmäe J., Korpela M.: A process for specifying integration for multi-tier applications in healthcare. *Int. J. Med. Inform.* 2003; 70(2-3): 173-82.
- [4] López D.M., Blobel B.: A Development Framework for Semantically Interoperable Health Information Systems. *Int. J. Med. Inform.* 2003 (in press).
- [5] Andersson J., Johnson P.: Architectural integration styles for large-scale enterprise software systems. In: *Proceedings of the Fifth IEEE International Conference on Enterprise Distributed Object Computing*; 2001 Sept 1-4; Seattle, USA. Wahington DC: IEEE Press; 2001. p. 224-36.
- [6] Vetere G., Lenzerini M.: Models for semantic interoperability in service-oriented architectures. *IBM Syst. J.* 2005; 44(4): 887 - 903.
- [7] Bernstein K., Bruun-Rasmussen M., Vingtoft S., Andersen S.K., Nohr C.: Modelling and implementing electronic health records in Denmark. *Int. J. Med. Inform.* 2005; 74(2-4): 213-20.
- [8] IEEE Computer Society. IEEE Standard Glossary of Software Engineering Terminology, IEEE Std 610.12-1990 The Institute of Electrical and Electronics Engineers, Inc, 1990
- [9] Bernstein P.A.: Middleware: a model for distributed system services. *Commun. ACM.* 1996; 39(2): 86-98.
- [10] Blobel B.: Analysis, Design and Implementation of Secure and Interoperable Distributed Health Information Systems. Amsterdam: IOS Press; 2002.
- [11] ISO/IEC IS 10746-2. Information technology - Open Distributed Processing -Reference Model: Foundations. International Organization for Standardization; 1996.
- [12] Lopez D.M., Blobel B.: Enhanced Semantic Interoperability by Profiling Health Informatics Standards. *Methods Inf. Med.* 2009; 48: x-x, doi: 10.3414/ME9216 (In Press)
- [13] Lopez D.M., Blobel B.: Connecting public health and clinical information systems by using a standardized methodology. *Medinfo* 2007; 12(Pt 1):132-6. (Best Student Paper Award).
- [14] Blobel B., Holena M.: Comparing middleware concepts for advanced healthcare system architectures. *Int. J. Med. Inform.* 1997;46(2):69-85
- [15] Blobel B.G., Engel K., Pharow P.: Semantic interoperability--HL7 Version 3 compared to advanced architecture standards. *Methods Inf. Med.* 2006; 45(4):343-53.
- [16] Transport Specification - Web Services Profile, Release 2. [standard on the Internet]. Ann Arbor: Health Level Seven, Inc; 2008. [Cited 2008 May 18] Available from <http://www.hl7.org/v3ballot/html/infrastructure/transport/transport-wsprofiles.htm>
- [17] The HSSP Service Oriented Reference Architecture [homepage on the Internet]. Ann Arbor: Health Level Seven, Inc; 2008. [Cited 2008 May 18] Available from: <http://hssp.wikispaces.com/>
- [18] Resolución 2542 de 1998. Sistema. Integral de Información para el Sistema General de Seguridad Social en Salud-SIIS [monograph on the Internet]. Bogotá: Ministerio de la Protección Social, República de Colombia; 1998. [Online]. [Cited 2008 May 15]. Available from: <http://www.minproteccionsocial.gov.co>
- [19] Decreto 3518 de 2006. Sistema de Vigilancia en Salud Pública. [monograph on the Internet]. Bogotá: Ministerio de la Protección Social, República de Colombia; 1998. [Online]. [Cited 2008 May 15]. Available from: <http://www.minproteccionsocial.gov.co>
- [20] HL7 Java SIG Project [home page on the Internet] 2008 [Cited 2008 May 22]. Available form: <http://aurora.regenstrief.org/javasig>.
- [21] Mirth Project [home page on the Internet]. Irvine: WebReach Inc; 2008 [Cited 2008 May 22]. Open Source Cross-Platform HL7 Interface Engine. Available form: <http://www.mirthproject.org/>.