Proceedings of the 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), Larnaca, Cyprus, 11-13 November 2012

Electronic Health Record Diagnosis tool and a cross reference between coding standards

Harris Soteriades
Department of Computer Science
University of Cyprus
Nicosia, Cyprus
csoter03@cs.ucy.ac.cy

Kleanthis Neokleous
Department of Computer Science
University of Cyprus
Nicosia, Cyprus
kleneokl@cs.ucy.ac.cy

George Tsouloupas
The Cyprus Society for Medical Informatics
Nicosia, Cyprus
george.tsouloupas@gmail.com

Antonis P. Jossif
"Paedi" Center for Specialized Pediatrics
Nicosia, Cyprus
ajossif@paedi.org.cy

Schizas N. Christos
Department of Computer Science
University of Cypurs
Nicosia, Cyprus
schizas@ucy.ac.cy

Abstract— **Many computer applications have been released targeting the implementation of a general Electronic Health Record (EHR) system. The main goal of an EHR system is to digitally store useful medical information about a patient's history and to share it across different health care units. Furthermore, the EHR systems should provide important information to the medical doctors during or after an encounter of care that will facilitate their decisions on making the correct diagnosis. A major problem however, for adopting an EHR system, especially from older doctors corresponds to the time that is needed to become familiar with the tool and thus either prevents them from enjoying the full benefits of an advanced EHR system, or leads to a complete rejection of the system. Therefore one of the fundamental concepts for implementing an EHR system is to incorporate proper and easy to use tools that will enhance and guide the doctors' decisions in an optimal way without taking significant time of their usually busy schedule. In this paper we aim to tackle part of this problem by proposing an application tool that will help the doctor with the patient's diagnosis by suggesting a matching disease code based on the user's (doctor's) notes. In addition, the tool will offer cross reference functionality between two international standards, the International Classification of Diseases (ICD-10) and the Systematized Nomenclature of Medicine -- Clinical Terms (SNOMED-CT). This tool will be tested as a freemed-software module add-on and it will be implemented in a private group practice (the Center). With their cooperation and the feedback from the users the tool will be evaluated for the possibility to be used in other EHR applications.** *(Abstract)*

Index Terms—**Electronic Health Record, International Coding Standards cross reference, International Classification of Diseases, (ICD-10) Systematized Nomenclature of Medicine -- Clinical Terms (SNOMED-CT).**

I. INTRODUCTION

Patient oriented services in the healthcare system grow rapidly and the demand for a better and more sophisticated treatment becomes stronger almost every year [1]. Electronic Health Record systems contribute in the healthcare domain by gathering information from patients' records such as history of illnesses, current diseases and medications, billing, insurance claims and many other functionalities that can produce useful information [2]. The data gathered can be used to gain valuable knowledge and provide important information when shared with other facilities nationally and internationally for research or treatment aid [2,3]. Although in theory this seems to be an easy task, in practice a matter of data interoperability rises when trying to achieve this kind of collaboration between clinics and doctors.

Heterogeneous systems are being used worldwide having their distinct usability and customization according to their needs in each case. For these reasons some international standards are established from the World Health Organization (WHO), which is the directing and coordinating authority for health within the United Nations system [4].

In this article we focus on the two medical and disease coding standards that are supported by WHO and are used to classify diseases and other health problems recorded on many types of health and vital records including death certificates and health records [4]. The two medical standards are the ICD version 10 (upcoming version 11) and the SNOMED-CT. The ICD-10 stands for **I**nternational **C**lassification of **D**iseases and the SNOMED-CT stands for **S**ystematized **No**menclature of **Med**icine -- **C**linical **T**erms.

II. PROBLEM

Many of the heterogeneous systems that are used in healthcare include both ICD and SNOMED-CT for classifying diseases. This factor lead's to a diagnosis problem when there is an exchange of information between these two coding standards [5, 6, 7].

Additionally doctors require a guide tool for inserting semantically correct diagnosis, according to the two standards, and a tool that will facilitate their decisions towards a more accurate diagnosis.

III. RELATED WORK

To examine and diagnose a patient, doctors receive and study all the relevant information they can get. What if they miss an important piece of data and what will be the consequences?

978-1-4673-4358-9/12/$31.00 ©2012 IEEE

A lot of research has been done related to automating clinical coding [8], machine learning [9] algorithms and text-mining tools that aid domain experts on making the correct decision [10, 11]. Using text-mining, free text and important report information are used to train machine learning systems for clinical free text classification [12]. With the aid of normal language processors (NLP), semantic types are being separated and a specialized vocabulary is created [13, 14, 15] that helps recognizing important entities [16, 17] and producing useful and expert reports. Others researchers have applied the Naïve Bayes algorithm to build classifiers from several corpuses and diagnose patients [18] according to the accuracy of their sample training sets. Some rule based algorithms [19] assist on how experts plan to structure and build their rules for their decision support system [20, 21, 22] and match their domain requirements. Using this methodology experts can directly plan their tasks and workflows for their decisions such as diagnosing patients, diseases prevention, medication alerts etc. [22].

To achieve a better algorithm performance the tool must resolve problems that refer to the domain requirements (in our case the Center's requirements) [23].

IV. APPROACH

To clearly understand what the tool requirements are, further interaction with the medical doctors involved is required. The Center's staff is using the version v.0.8.4 of the open-source software "freemed" (available for download at [24]) and the European Paediatric Cardiac Coding (AEPC) scheme which is cross referenced with the ICD-10 standards. The application has a lot of functionality that meets the Center's requirements. For example, it provides an *extensible* patient records system, billing information, appointment list etc. One requested functionality missing from this system is the inability to aid and guide the user in correctly classifying the patient's diagnosis according to the international standards.

At the time of the medical examination of a patient, due to several factors such as the overloaded doctor's schedule is or that the examination happens to be an emergency, the user cannot rely on opening and searching databases or reference systems consisting of thousands of codes to locate the correct diagnostic phrase for the diagnosis because in this situation is time consuming for both the patients and the doctor. Furthermore, this procedure, frequently enough results in errors and wrong classifications.

It is accurate to explain here that the Center's coding scheme is more detailed from the ICD-10 canonical scheme thus some AEPC codes are cross referenced with generic ICD-10 codes.

To give an example scenario: The doctor examines the patient. The "patient encounter form" is opened and the appropriate notes are written in the form. The doctor concludes that the patient is suffering from either "Pulmonary atresia intact septum" or "Pulmonary atresia VSD". The doctor however, in the diagnosis area has written the ICD-10 code Q22.0 which stands for "Pulmonary valve atresia" which is more generic. Although the name is very similar, the physiology and treatment of the disease are quite different. Because of the limited time to access the coding scheme and verify that the diagnosis-code is indeed the same as it is used in the international standard or in this case the AEPC, the diagnosis is inserted incorrectly. In this scenario there are two problems. The first one refers to the incorrect coding and thus makes the data misleading for further medical investigation and the second problem corresponds to the wrong transfer of data for subsequent processes such as billing or updating of the patient's summary.

To resolve this problem we propose an additional functionality to be incorporated in the existing tool. At **first** an easy access to the coding system is needed so the doctor can instantaneously enter a keyword or a phrase that describe the diagnosis and automatically he or she will be presented with the suggested codings from the system. Subsequently, the doctor will select the appropriate codings that best correlate with the established diagnosis. Alternatively the algorithm will read and identify keywords through the doctor's text-notes and will produce medical keywords that will automatically present coding suggestions (from the coding standard database) to the doctor. The idea is inspired from some of the work that has been done using NLP [13, 15]. **Secondly** the tool will try and find the matched coding from the SNOMED-CT coding standard to give the cross reference functionality to the system.

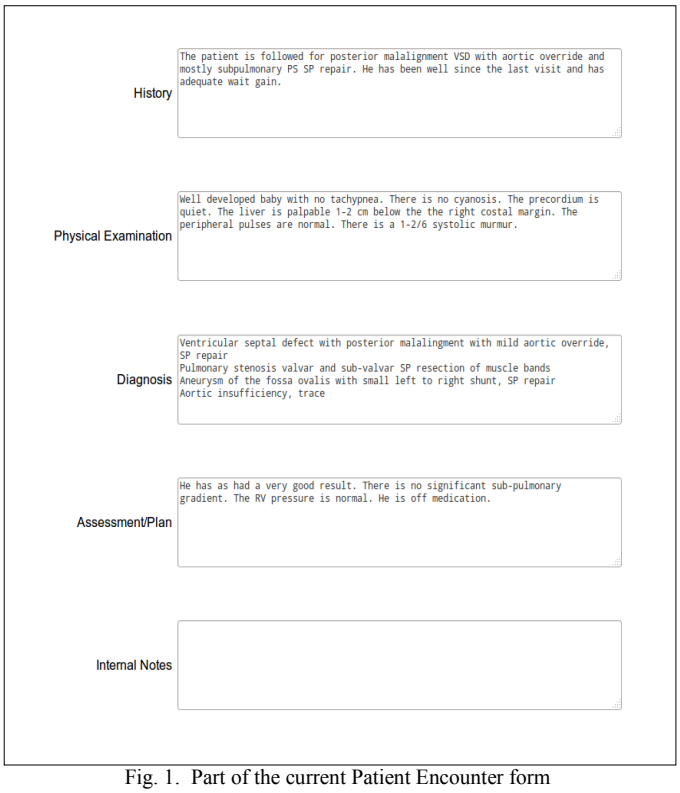

Fig. 1. Part of the current Patient Encounter form

V. TOOL DESIGN

The freemed v0.8.x is written in the "PHP" programming language using MySQL Database Management System (DBMS) on the server-site and html combined with JavaScript at the client-site.

Our design will be implemented on both, the server and the client sites and the approach that will be followed for the successful addition of the new functionalities are summarized in two steps. In the **first step,** a database to hold the coding standards of the AEPC will be produced (current format is in MS Excel). In the **second step,** a rearrangement of the "*patient encounter form"* will be performed with the following additions: a) The "***search***" element that will hold the keyword or phrase and will guide the user to the semantically correct diagnosis, b) the "***Add Diagnosis to Patient***" and "***Remove Diagnosis from Patient***", with the former functioning by adding the selected disease code from the retrieved query result of "***search***" and the latter to remove it from the patient and c)an element that will refer to the medical keywords of the doctor's notes regarding the patient's summary. These keywords will be combined together and a "***suggest***" functionality will present to the user the suggested coding list. The doctor will have the opportunity to choose from the two methods, by disabling the "***suggest***" functionality and going through the "***search***" function or vise versa.

For the second part the user will have the ability to map the selected diagnosis from ICD-10 to SNOMED-CT standard. The cross reference functionality will comprise a function that will introduce the matching SNOMED-CT coding to the former one.

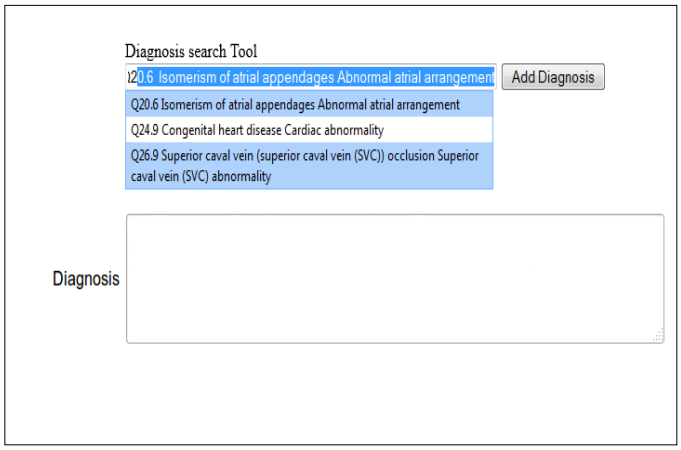

Fig. 2. The Add Diagnosis tool. When the user writes a phrase the tool searches to find matching diseases with their appropriate codings. (Part of the patient encounter form)

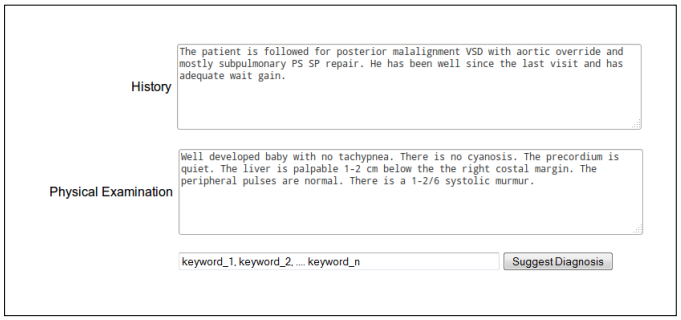

Fig. 3. The Suggest Diagnosis tool that is going to find a match according to the keywords that are extracted from the notes. (Part of the patient encounter form)

VI. CONCLUSIONS

The EHR functionalities are focused around the patient to enhance a better and more sophisticated treatment. In accordance each medical facility can seek for different and custom solutions in line with their requirements and needs.

The main goal is to conduct an approach-research regarding the requirements of a specific center, create a solution for the predefined needs and get the appropriate feedback from the field experts. We will test and measure the results of the gathered feedback to provide a more robust tool that will apply to all of the requirements mentioned in the previous sections. We hope that this research will build a foundation on the way that the whole progress should be made thus getting closer and closer to a better understanding on the specific domain reference-demand and provide a better solution.

ACKNOWLEDGMENT

This research was supported by grant 0311(BIE)/47 from the Cyprus Research Promotion Foundation.

REFERENCES

[1] Andy Podgurski, "Evidence-based validation and improvement of electronic health record systems," in proceedings of the FSE/SDP workshop on Future of software engineering research (FoSER '10), ACM, New York, NY, USA, pp. 287-292, 2010

[2] Kathrin M. Cresswell, Ann Robertson, and Aziz Sheikh, "Lessons learned from England's national electronic health record implementation: implications for the international community," in proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium (IHI '12), ACM, New York, NY, USA, pp. 685-690. 2012

[3] National Library of Medicine, National Institutes of Health, U.S. Department of Health and Human Services, "Mapping SNOMED-CT to ICD-10-CM final Release notes 2012," July, 2012.

[4] http://www.who.int/classifications/icd/en/

[5] Rodrigues JM, Kumar A, Bousquet C, Trombert B, "Using the CEN/ISO standard for categorial structure to harmonise the development of WHO international terminologies," European federation for medical informatics, 2009

[6] Schulz S, Rector A, Rodrigues JM, Chute C, Üstün B, Spackman K, "Ontology-based convergence of medical terminologies: SNOMED CT and ICD-11," in proceedings of the eHealth2012 Mai 10-11, Vienna, Austria, 2012.

[7] Craig E. Kuziemsky, James B. Williams, and Jens H. Weber-Jahnke,"Towards electronic health record support for collaborative

processes," in proceedings of the 3rd Workshop on Software Engineering in Health Care (SEHC '11), 2011.

[8] Stanfill MH, Williams M, Fenton SH, Jenders RA, Hersh WR, "A systematic literature review of automated clinical coding and classification systems," American Health Information Management Association, Chicago, Illinois, USA, 2010

[9] György Szarvas, Szilárd Iván, András Bánhalmi, János Csirik, "Automatic extraction of semantic content from medical discharge records," in proceedings of the 5th WSEAS international conference on System science and simulation in engineering (ICOSSE'06), M. Isabel Garcia-Planas and Concepcion Gonzalez-Concepcion (Eds.), World Scientific and Engineering Academy and Society (WSEAS), pp.178-183, 2006.

[10] Aaron M. Cohen, William R. Hersh, "A survey of current work in biomedical text mining," Brief Bioinform vol 6 (1), pp. 57-71, 2005.

[11] D. E. Oliver and R. B. Altman, "Extraction of SNOMED concepts from medical record texts," in proceedings of the annual symposium of computer application Med Care, pp. 179-83, 1994

[12] Anastasia Sotelsek-margalef, Julio Villena-román, Iii Madrid, "MIDAS: An Information-Extraction Approach to Medical Text Classification," "MIDAS: Un enfoque de extracción de información," procesamiento del lenguaje natural, ISSN 1135-5948, Nº. 41, 2008.

[13] Richárd Farkas, György Szarvas, István Hegedűs, Attila Almási, Veronika Vincze, Róbert Ormándi, Róbert Busa-Fekete, "J Am Med Inform Assoc. Semi-automated Construction of Decision Rules to Predict Morbidities from Clinical Texts," journal of the American Medical Informatics Association Jul-Aug vol. 16(4), pp. 601–605, 2009.

[14] Bodenreider, O, "Biomedical ontologies in action: role in knowledge management, data integration and decision support," in Yearbook of Medical Informatics, Stuttgart: Schattauer pp. 67–79, 2008.

[15] Aurélie Névéol ,Won Kim, W. John Wilbur, and Zhiyong Lu, "Exploring two biomedcal text genres for disease recognition," in proceedings of the workshop on Current Trends in Biomedical Natural Language Processing (BioNLP '09), Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 144-152, 2009.

[16] Md. Faisal Mahbub Chowdhury and Alberto Lavelli, "Disease mention recognition with specific features," in proceedings of the 2010 workshop on Biomedical Natural Language Processing (BioNLP '10), Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 83-90, 2010.

[17] Yang, H., Spasic, I., Keane, J., Nenadic, G, "Combining Lexical Profiling, Rules and Machine Learning for Disease Prediction from Hospital Discharge Summaries," in proceedings of i2b2 Obesity Challenge 2008: i2b2 Obesity Challenge, 2008.

[18] Pakhomov SV, Buntrock JD, Chute CG, "Automating the assignment of diagnosis codes to patient encounters using example-based and machine learning techniques," journal of the American Medical Informatics Association, Sep-Oct, vol. 13(5) pp. 516-25, 2006.

[19] Iram Fatima, Muhammad Fahim, Donghai Guan, Young-Koo Lee, and Sungyung Lee, "Socially interactive CDSS for u-life care," in proceedings of the 5th international conference on Ubiquitous Information Management and Communication (ICUIMC '11), ACM, New York, NY, USA, Article 95, 2011.

[20] M. Williams, F. Wu, P. Kazanzides, K. Brady, and J. Fackler, "A modular framework for clinical decision support systems: medical device plug-and-play is critical," SIGBED Rev. 6, 2, Article 8, July, 2009.

[21] Kamran Sartipi, Mohammad H. Yarmand, and Douglas G. Down, "Mined-Knowledge and Decision Support Services in Electronic Health," in proceedings of the international workshop on Systems Development in SOA Environments (SDSOA '07), 2007

[22] Jinbo Bi and Arun Abraham, "An intelligent web-based decision support tool for enhancing asthma guideline adherence," in proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium (IHI '12), 2012

[23] Benoît Misset, Didier Nakache, Aurélien Vesin, Mickael Darmon, Maïté Garrouste-Orgeas, Bruno Mourvillier, Christophe Adrie, Sébastian Pease, Marie-Aliette Costa de Beauregard, Dany Goldgran-Toledano, Elisabeth Métais, Jean-François Timsit,"Reliability of diagnostic coding in intensive care patients," the Outcomerea Database Investigators Critical Care, vol. 12(4) R95, July 29, 2008

[24] http://freemedsoftware.org/download/