

## **Ontologies applied in medicine: A digital library creation framework for health literacy**

Ramona Popa, Andrei Vasilateanu, Nicolae Goga, Andrei Doncescu, Paolo Darminio, Barbur Raluca-Maria, Ilie Catalin

► **To cite this version:**

Ramona Popa, Andrei Vasilateanu, Nicolae Goga, Andrei Doncescu, Paolo Darminio, et al.. Ontologies applied in medicine: A digital library creation framework for health literacy. 6th IEEE International Conference on E-Healthand Bioengineering (EHB 2017), Jun 2017, Sinaia, Romania. 4p., 2017, <<http://www.ehbconference.ro/>>. <hal-01565989>

**HAL Id: hal-01565989**

**<https://hal.laas.fr/hal-01565989>**

Submitted on 27 Jul 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# *Ontologies applied in medicine: A digital library creation framework for health literacy*

Ramona Popa<sup>1</sup>, Andrei Vasilateanu<sup>1</sup>, Nicolae Goga<sup>1,2</sup>, Andrei Doncescu<sup>3</sup>, Paolo Darminio<sup>4</sup>, Barbur Raluca-Maria<sup>5</sup>, Ilie Catalin<sup>5</sup>

<sup>1</sup>Department of Engineering in Foreign Languages, Universitatea Politehnica din Bucuresti, Spaiul Independentei 313, Bucuresti, Romania

<sup>2</sup>Faculty of Natural Sciences, University of Groningen, the Netherlands

<sup>3</sup>Laboratoire d'analyse et d'architecture des systems, University of Toulouse, France

<sup>4</sup>Anova Consulting Group, Madrid, Spain

<sup>5</sup>S.C. OSF GLOBAL SERVICES S.R.L., Blvd. Dacia, nr. 99, Bucuresti, Romania

popa.ramona91@yahoo.com, andrei.vasilateanu@upb.ro, n.goga@rug.nl, andrei.doncescu@laas.fr, paolo.darminio@anovagroup.es, raluca.barbur@osf-global.com, catalin.ilie@osf-global.com

*Abstract— The work of this paper is done within the context of European Project E! 9770 PrEmISES<sup>1</sup>. One of the central pieces of PrEmISES[16] is the use of automatic tool generation algorithms for the creation of ontologies. The objective of this work is to extend the work from PrEmISES for raising the medical literacy of empowered patients by creating a digital library that contains relevant articles from different medical domain. It should be noted that a large number of persons are interested in finding relevant information about their condition. We have chosen to discuss in this research the automated generation of ontologies for the medical domain. By the use of ontologies that gives key terms from a domain, the creation of the digital library can be done by crawling from the Internet articles which are related to the profile of the user and that are based on the key terms of the ontologies. The articles from the digital library can be used by a patient to raise his / her knowledge and literacy related to his/her diseases.*

**Keywords— Patient Empowerment; Automatic Ontology Generation; Medical Domain Ontology; Digital Library; Disease Information Services;**

## I. INTRODUCTION

Nowadays the capacity for adapting, learning and being informed is an important step in reducing the risk of many diseases. The work of this paper is done within the context of European Project E! 9770 PrEmISES. One of the central pieces of PrEmISES is the use of automatic tool generation algorithms for the creation of ontologies. The purpose of this framework is to extend the work from PrEmISES to empower patients by raising their medical literacy, by creating a personalized digital

library by selecting relevant articles for them through the use of ontologies.

A framework that works as a digital library containing relevant articles for the medical domain, can be seen as an efficient way to organize the information for researchers who must structure and share the information in a specific domain. Users without a formal education in medicine can be also encouraged to use this software framework in order to be constantly informed and gain basic medical knowledge relevant for their profile.

By the use of ontologies that gives key terms from a domain, the creation of the digital library can be done by crawling from the Internet articles which are related to the profile of the user and that are based on the key terms of the ontologies. Key terms - main relevant concept from ontologies – can be used to index medical articles and classify them. Digital library will contained the most relevant articles for a user.

The development of the ontology can be seen as the construction of on a meta-ontology that is composed of more sub-ontologies, each sub-ontology corresponding to a specific disease. In order to develop the sub-ontologies, we have used an innovative technique that is based on the automatic generation of ontologies from unstructured text. This generation technique allows us to extract the most relevant terms for our domain of study from unstructured text.

We used asthma as a benchmark since ontologies exist for this domain. However the main advantage of our system is that it can create ontologies for conditions where such

---

<sup>1</sup> <http://www.premises-project.eu/>

ontologies do not exist or are too sparsely populated to help in retrieving relevant articles.

In this paper, we present the framework main flow and use cases for the digital library. Then we continue with the description of the building of the ontology and the validation of the ontology with a manual one.

Our paper is organized as follows. In Section 2 we present the state-of-the-art for ubiquitous learning (used by our framework) and the automatic generation of ontology from unstructured text. In the same section, we present differences between our system and other similar software. In Section 3 we describe the main flow of our framework. Section 4 shows our results and Section 5 draws the conclusions and our future work.

## II. LITERATURE REVIEW

**Ubiquitous learning** Health literacy [7] can be defined as the capacity of an individual to find, read and understand health-care information. The degree to which persons have the ability to use and understand health articles will help them to make appropriate health decisions and generally be more involved in their treatment process. Health literacy can be increased through ubiquitous learning.

According to [13] the main purpose of applying ubiquitous learning paradigm in e-learning is to build a ubiquitous learning environment that allows people to learn, more effectively, at any place (regardless of the individual's location), due to the portability of computers and to use the personal context in the learning process. Using ubiquitous learning software, one can deliver the right learning materials, at anyplace (regardless of the individual's location) and in a comfortable way. Ubiquitous learning is also known as u-learning. By using ubiquitous learning knowledge has become more available.

Our digital library can be seen as an u-learning system due to the fact that one can use it anywhere and anytime in order to gain medical knowledge.

An example of an ubiquitous learning platform is given in the article [12]. This platform, named CULP is based on a real-time help-seeking mechanism. Using CULP, individuals can use their mobile devices in order to have instant support in their learning activities. CULP system offers hints from qualified people in order to help individuals when they meet difficulties in their learning activities.

According [11] the application development for ubiquitous computing relies on software frameworks that provide higher level abstractions. Ubiquitous Computing paradigm called also the ubicom can be described as pervasive computing.

As any pervasive computing, the goal for our system is to make the devices smart by embedding a digital library that is semantically enriched. For this purpose ontologies are created. These ontologies are capable of collecting terms and based on that terms we obtain a relevance score for each concept.

Other common characteristics with the ubiquitous computing is that our framework involves internet connection and wireless communication. More than that, our framework can be accessed from any type of device: tablets, mobile, laptops or desktop computers and the personal context of an individual is taken into account- i.e. the specific disease of a person.

**Personal Health Records:** In [9], the authors present a personal health recorder, or PHR as a repository used in order to gather patient's data. A personal health record, is more than a simply a collection of information about patient's health. They improve patient's care and provide greater benefits by recording and tracking patient's health progress.

In the article [2] [15] a large number of PHR's potential advantages is presented. One of the most important benefits is that Personal Health Records' are universally available. Another benefit of PHR's is that they can store individual's health documents for many years to come and in a secure way.

Our software is an improved PHR software for educational purposes, based on the ubiquitous learning paradigm that uses ontologies in order to collect and store relevant documents about a specific disease and it will empower patients by raising their medical literacy.

**Automatic Ontology Generation:** Our digital library will crawl data from the internet and by the use of ontologies, the framework is capable to filter the most relevant articles which are related to the profile of the user.

Our software framework is based on a meta-ontology that is composed of sub-ontologies, each sub-ontology corresponding to a specific disease. Developing ontologies for the medical domain is hard to achieve manually. Our target is to build these ontologies automatically. One possible option for this is to use an ontology generation tool such as Text2Onto [8]. According to [10], [1] this tool contains many features. It contains rules for the generation of ontologies by statistically extracting concepts and using grammar rules to extract relations between terms. We have used our modified version of ontology generation tool built from Text2Onto while adding new relation discovery rules and similarity algorithms between terms [10][1].

Text2Onto requires large corpus of unstructured data in order to generate relevant concepts for our medical domain. We have extracted the data for the corpus from the web, by filtering the most relevant text for our domain. Each ontology contains the concepts that have a relevance score larger than a given threshold. The threshold was chosen in a dynamic way by taking into account the size of the corpus.

**Similar systems:** According to [4] TELUS HEALTH company provides many health platforms for their customers such as Home Health Monitoring and Personal Health Recorders. The goal of this platform is to reduce the health-care costs and in the same time to help patients that are diagnosed with chronic disease by aiding them to learn about how to manage their own condition. Using TELUS healthcare portals, patients are able to upload, track and modify their health data in a secure way. The staff caring the patients are able to add health information about them and to customize the page view of their patients.

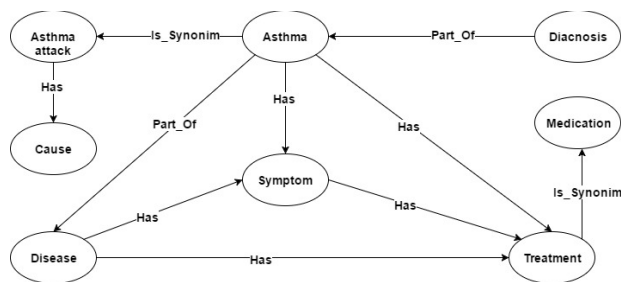


Fig. 2 Sample of generated ontology

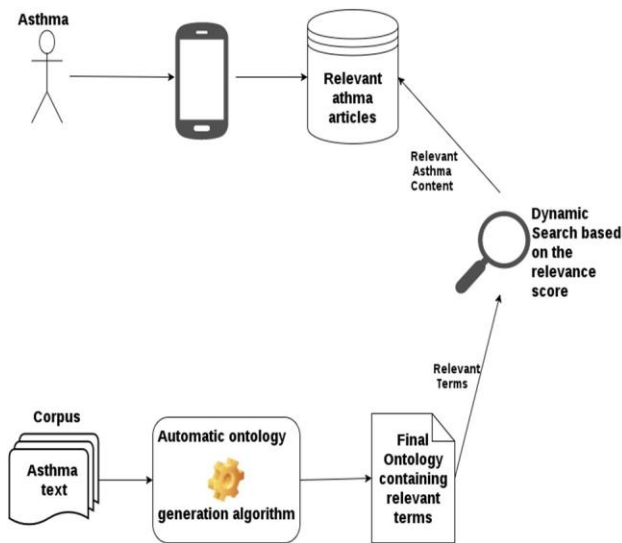


Fig. 1 System flow

In [5] a healthcare system framework named WHO is described. The system consists of six building blocks and their primary goal is to support and promote health. According to [3] the biggest challenge for the modern health systems is to find an equilibrium among the finest quality of care and the optimal use of diminished resources. Our framework finds this equilibrium and provides high quality healthcare literacy due to the fact that it's semantically enriched. Thanks to the system's ontology component the framework is able to filter the most relevant documents from the dwindling resources that are available on the web.

### III. METHODOLOGY AND FRAMEWORK

Our digital library has the following main components:

- **Ontology component:** a meta-ontology with ontologies from several medical domains. Ontologies for the different domains are created by automatically generating algorithms from corpus of text.
- **Search and indexing component** crawls the web based on the relevant terms stored in the ontologies for getting relevant articles related to a medical domain.
- **User Interface:** used by the user to complete a profile and to read the relevant articles from search and indexing component.

Current article will focus on the ontology component – respectively meta-ontology.

In Fig. 1 the system flow is highlighted (for the domain of asthma).

Let us consider a scenario for a patient having asthma. First when registering in the application he/she completes a short profile containing his/her conditions, level of education, learning habits (how many hours a day, preferred device). Once

Concept	Instance	Similarity	SubclassOf	InstanceOf	Relation	SubtopicOf
Label	Relevance					
asthma		0.0875785532825244				
symptom		0.032223559299371575				
airway		0.022195480679235192				
people		0.017114587511699426				
treatment		0.015376387217542453				
child		0.012568525203904266				
lung		0.01069661719481214				
condition		0.010429201764941836				
asthma attack		0.009626955475330927				
medicine		0.00909212461559032				
disease		0.008958416900655167				
time		0.00842358604091456				
attack		0.007888755181173953				
air		0.007888755181173953				
use		0.00762133975130365				
doctor		0.00762133975130365				
inflammation		0.0073539243214333465				
inhaler		0.0073539243214333465				
allergy		0.006551678031822436				
article		0.006551678031822436				
cause		0.005883139457146677				
adult		0.005883139457146677				

Fig. 3 Generating concepts from unstructured text

the profile is completed the application checks if the digital library contains the necessary articles, and starts delivering them if possible. If the digital library lacks the content, the generated condition ontology (in our case asthma) is used by the crawlers to retrieve relevant articles from approved sources and populate the library.

The advantage of this ontological approach is in the increased relevancy for the retrieved articles. Articles are not indexed only statistically, for example by using the TF-IDF algorithm [14], but also by taking into consideration the relations between terms and searching for this semantic information in the documents.

### IV. RESULTS

In the fig 2 there is a sample of our asthma ontology. This ontology was generated automatically by running Text2Onto tool on a corpus of unstructured text in order to acquire the most relevant asthma concepts. A snapshot from the application is visible in fig 3.

A number of 303 terms were generated for the asthma domain, by running our modified version of Text2Onto using unstructured text for its corpus. When building the final asthma ontology, we took into account, only the terms with a relevance mark that is greater than a considered threshold. In order to test if our ontology is relevant for the asthma domain we have chosen to compare these 303 terms, that were generated automatically, with the terms contained in other existing asthma ontologies.

We have searched for a sanctioned asthma ontology and found one on Biportal [6] that is composed of 289 asthma

concepts. A sample is presented in fig 4. The comparison between these ontologies is relevant because they contain almost the same number of concepts and both of these ontologies are for the asthma domain. We have compared the two ontologies by the term similarity between them. Firstly the ontology from Biportal needed some pruning as it contained terms with the same stem ( eg. occupation, occupational) and even after stemming the ontology contained many composed terms ( eg . “chronic asthma with fixed airflow obstruction” or “emergency asthma admission since last appointment”) due to poor ontology design. After pruning the ontology was left with 231 terms from the original 289. To find concepts that are similar in both ontologies we use the following methodology. A concept is considered to be a common concept if its concept label or any of its synonyms exist in both of the ontologies being compared. The common concepts set for the selected concept will be the result of the intersection between all the concepts in the manual ontology and the list comprising of the selected concept and its synonyms. These sets are then aggregated from each selected concept to form the overall common concepts set.

The similarities between these ontologies are notable: 50 out of the 231 terms are exact matches. If we also consider composed words (the automatic ontology contains all the words from a composed word in the manual ontology) we gain 45 more common terms. The automatic ontology lacks terms related to specific medication as instances of terms are more difficult to extract statistically.

Our conclusion is that the ontology that was generated automatically using Text2Onto is suitable for the domain of asthma and the method can be used with success for domains without ontologies.

We used asthma as a benchmark since sanctioned ontologies exist for this domain. However the main advantage of our system is that it can create ontologies for conditions where such ontologies do not exist or are too sparsely populated to help in retrieving relevant articles.

## V. CONCLUSION AND FUTURE WORK

The goal of this work was to describe a software framework that works as a digital library with medical articles for a given domain. The system can be used to empower patients by raising their medical literacy, by selecting relevant articles for them through the use of ontologies.

In this research, we describe the system’s flow and we present its main components. The emphasis is on the ontology component and on the automatic generation process from unstructured text. According to our experiments, the automated generation of ontologies for the medical domain is an innovative solution.

The next step in this research is to finish the meta-ontology by creating all the sub-ontologies, each sub-ontology corresponding to a specific disease. In the last phase, the meta-ontology will be integrated in the digital library system.

This system is currently a prototype done within context of European Project E! 9770 *PrEmISES*. In the project’s final version, doctors will be able to evaluate the articles, stored within our digital library, based on a grade system.

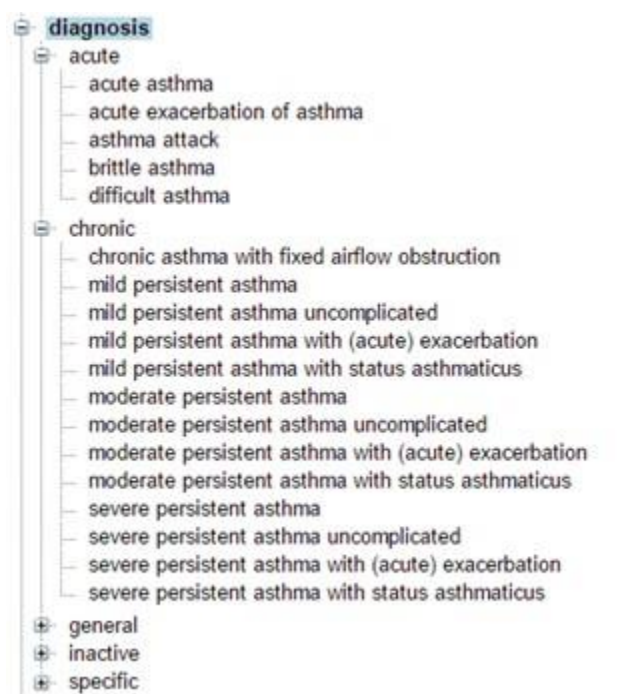


Fig. 4 Snapshot from Biportal ontology

## REFERENCES

- [1] Andrei Vasilăteanu, Nicolae Goga, Elena-Alice Tanase, Iuliana Marin, “Enterprise Domain Ontology Learning From Web-Based Corpus”, 6th ICCNT 2015 Conference Proceedings, USA, pp.112-117
- [2] Brittany Vance, Brent Tomblin, Jena Studney, Alberto Coustasse, “Benefits and Barriers for adoption of Personal Health Records”, Marshall University, Management Faculty Research, Spring 3-2015
- [3] Diane Levin-Zamir, Yitzhak Peterburg, “Health literacy in health systems: perspectives on patient self-management in Israel”, 01 March 2001
- [4] <https://www.telushealth.co/health-solutions/patient-and-consumer-health-platforms/products/personal-health-records/>
- [5] [http://www.wpro.who.int/health\\_services/health\\_systems\\_framework/en/](http://www.wpro.who.int/health_services/health_systems_framework/en/)
- [6] <http://biportal.bioontology.org/ontologies/AO>
- [7] Lynn Nielsen-Bohlman, Allison M. Panzer, David A. Kindig, “Health Literacy: A Prescription to End Confusion” 2004
- [8] P. Cimiano, and J. Völker, “Text2Onto,” Natural language processing and -information systems, Springer Berlin Heidelberg, pp. 227-238, 2005.
- [9] Paul C. Tang, Joan S. Ash, David W. Bates, J. Marc Overhage, Daniel Z. Sands, “Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption” Mar-Apr 2006
- [10] Popa Ramona-Cristina, Andrei Vasilăteanu, Nicolae Goga, “Ontology based multi-system for SME knowledge workers” IEEE ISSE, Edinburg 2016
- [11] Rodrigo Oliveira Spinola, Guilherme Horta Travassos, “Towards a framework to characterize ubiquitous software projects” vol . 54 issue 7, July 2012
- [12] Saadiyah Yahya, Erny Arniza Ahmad and Kamarularifin Abd Jalil, “The definition and characteristics of ubiquitous learning: A discussion”, International Journal of Education and Development using Information and Communication Technology (IJEDICT), 2010, Vol. 6, Issue 1, pp. 117-127
- [13] Gwo-Jen Hwang, Chih-Hsiang Wu, Judy C. R. Tseng, Iwen Huang, “Development of a ubiquitous learning platform based on a real-time help-seeking mechanism”, 10 September 2010
- [14] K. Sparck Jones. "A statistical interpretation of term specificity and its application in retrieval". Journal of Documentation, 28 (1). 1972.
- [15] -Vasilăteanu, Andrei, Luca-Dan Serbanati, and Mihai Isarioiu. "Integrating Clinical Information from a Personal Health Record into the Virtual Healthcare Record." HEALTHINF. 2011.
- [16] [www.premises-project.eu/the-project/](http://www.premises-project.eu/the-project/)