Inteligencia Artificial 19(58), 17-22
doi: 10.4114/ia.v19i58.1155

INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

An Intelligent System Prototype to support and sharing diagnoses of maligned tumors, based on personalized medicine philosophy

Víctor Flores, Aldo Quelopana
Department of Computing & Systems Engineering. Universidad Católica del Norte
Angamos Av. 0610, Antofagasta, Chile
vflores@ucn.cl, aldo.quelopana@ucn.cl

Abstract. Circulatory systems diseases are one of the most important causes of death in Chilean population according to a report presented by the Chilean National Bureau of Statistics (INE). Undoubtedly, these sad numbers arise an opportunity to analyze ways to improve this situation. Personalized Medicine is a new approach used by health professionals to adapt standard medical treatments to patients' individual characteristics. Currently, several types of personalized-medicine software applications are built using Artificial Intelligent techniques and supported by techniques as Cloud Computing and Big Data. This architecture provides complex and varied information access such as clinical data, genome data, patients' treatment or drugs information, among others. This document describes a proposal to produce a method for generating and sharing medical information, particularly of maligned tumors in Chile. The prototype will be developed within the framework of the personalized medicine philosophy.

Keywords: Intelligent System, Cloud Computing, Big Data, Knowledge representation of Dynamic Systems, Personalized Medicine, Evidence-Based Medicine.

1 Introduction

One of the nations' main concerns across the globe is to improve the public quality of life and health standards. Many countries in Latin-American are facing this challenge, being Chile one of the most involved with this commitment in the area. During the year 2013, circulatory system diseases were one of the most important causes of death in Chile according to a report presented by the Chilean National Bureau of Statistics (INE) last year. This report points out that the 29% of men and 27% of women in Chile have been affected by this type of disease, which means 27,600 cases among the Chilean population [1], [2].
Undoubtedly, these sad numbers arise an opportunity of analyzing ways to improve this situation by supporting health professionals in obtaining early diagnoses and more accurate treatments, and thus to increase the chances of patient's recuperation in Chile [2]. This proposal is focused on Chilean population and the framework resulting could in the future be used on hospitals and medical centers in Chile; therefore, this proposal could be extended to other parts in the world, for example the Europe Union, due to they have declared health and healthy ageing as one of their five main objectives [3].

1.1 Personalized Medicine, data sources and software systems

Recent advances in fields such as biomedicine, both technological and investigative level, have generated a change in the approach and practice of the Evidence-Based Medicine (EBM), which currently is the base of traditional and personalized medicine [4], [5]. Regardless the relevant development of the EBM during the last decades, it is undeniable that effective implementations on real-daily medical cases and personalized use, have been reached slowly and with difficulties. Furthermore, developed countries invest 84% of their public resources on medical treatments, whereas the average in Latin-America is only 12%, which has meant that people in those countries have increased their efforts to cover their own health spending [6].

ISSN: 1137-3601, 1988-3064(on-line)

Personalized Medicine is an approach defined as the adaptation of standard medical treatments to the individual characteristics of a patient [9]. Even though of the fact mentioned where adaptation relies on people's genetic and molecular profiles, the concept can be extended to any source of relevant and reliable information in order to provide a support on medical decisions [10]. Nevertheless, medical practices have increased their dependency on data bases over the years, and its corresponding accessibility to those resources, that is why new challenges are now opened for being resolved by Information and Communication Technologies. About personalized medicine, one of the most important challenges nowadays is to obtain flexible systems of information to provide accurate, update, suitable and interrelated knowledge, based on layer-base access to multiple sources of heterogeneous data [7], [8].

Unlike to EBM, the data sources regard to personalized medicine can be classified in three groups [8], [10]: (i) genomic data about particular information of an individual, (ii) clinical history of patients or similar cases captured from sources like sensors (e.g. Electronic Health Recorder - EHR) [11], [12], and (iii) public or private big biomedical data bases; they are able to contribute with enough evidences to apply intelligent and specialized algorithms in personalized treatments. Personalized medicine is related to the development of strategies to generate tailored treatments for strengthening traditional medicine by adding the patient's genetics profile and heterogeneous information available in repositories as indicated in (iii).

These strategies include prevention and diagnoses of diseases, besides of treatment strategies [13]. It is observed in the literature that some medical systems provide medical information to supports health professionals in tasks such as taking decisions, personalized diagnoses and healthy lifestyle advice for patients. Moreover, this type of systems helps medical professional to manage and interpret Big Data sources such as the cloud, which when the extraction is carried out properly, it is able to help producing efficient and personalized diagnoses (as indicated in [12], [14]).

2 Context and identification challenge

Expert systems are a powerful tool used on many domains as biomedical or EBM. An expert system operates on a knowledge-base that contain domain facts, frequently it is elicited from human experts and this knowledge is represented using some formalisms as rules, frames, Bayesian networks, etc. [19]. Actually, Knowledge elicitation can be made from heterogonous sources (e.g. Internet or sensor networks) and this process can represent a complicate task. At the present time, the knowledge resulting of a heterogeneous elicitation process can be represented using combinations of formalisms as these mentioned above or others as ontologies. The ontology-based modeling approach is actually the most widely adopted intelligent-systems-knowledge-representation paradigm and it can be used for formal representation of knowledge through the definition of set of concepts within a domain by describing their corresponding relationships [20].

On medical and biological domains, knowledge has diverse structures such as concepts, data or text, treatment details, etc. The application contexts are several, but the focus of this proposal are the following: (1) facilitates the process of searching scientific information, related to standard treatments available on Internet, (2) generates algorithms to process knowledge that cause suggestions of early diagnoses, based on previous ones, but improving the process and results in order to apply them to the Chilean reality. This considers an interesting challenge due to the majority of the algorithms are created based on English language; and (3) provides new methods of presenting information in order to facilitate a comprehension based on previous researches performed by one of the researchers [7], [14].

To meet this challenge, Information and Communication Technologies (ICT) can contribute in several aspects, for example by using techniques such as Cloud Computing and Big Data. Cloud Computing brings paradigms that allow offering computing services via Internet or networks like Intranets [15]. Cloud Computing allows users to use various Internet resources on demand without worrying about where those resources are or how they are managed. Currently this work philosophy is allowing access to different types of applications related to health, which in turn makes possible accessing data (clinical records, genome, treatments, drug information, etc.) to analyze and produce relevant information for being used in fields such as personalized medicine [2]. This implies adequate means of storage, being Big Data the most suitable for this need.

When Big Data is personalized-medicine oriented, three factors are important [10], [11], [12]: (i) volume, which means the size of the datasets, (ii) speed, which means how much fast patients' data, treatment data, diagnoses and/or advice are generating; and (iii) variety, which means diversity and heterogeneity of data obtained from several sources and formats (e.g. texts, images, plots, etc.).

Having this in mind, the rest is summarized as the necessity of obtaining information from cloud in a suitable and opportune ways, using specific algorithms and framework in order to be used by Chilean health professionals.

This supposed being part of the dynamic trend of researching groups, normally shaped by multidisciplinary teams which developed flexible and open applications, able to integrate data and text throughout a set of techniques related to specific field, exceeding barriers such as nearness (physical) or availability of own data, and allowing the approach of the different resources available.

2.1 Related works

Some recent works are focus on problem of patients' vital data collection, distribution and processing, and how to make these processes more efficient [15]. For example, wireless sensor networks or utilities for collect and process information from the "cloud" are used in health care institutions [15], [16]. Cloud Computing represents a simple but powerful distributed computing model that also supports communication protocols with devices, secure data access protocols and information privacy, important aspects in fields as medicine [13]. Moreover, Cloud Computing generates and processes data related to personalized medicine and parameters that help customize treatments and/or recommendations for better health, it is possible to require large storage of information related to health.

In this sense, Big Data analysis methods are providing effective solutions in the field of medicine [5]. For example in [16] it is presented a study focused on the new challenges in biomedicine for processing and storing large volumes of data from magnetic resonance imaging, etc. According to [16], US President Obama (2015) announced "the Precision Medicine Initiative", efforts to focus attention on what they called "the medicine of precision" which is based in doctor's experience and capabilities, and also in the infrastructure and information technology available in the modern world. This initiative seeks to promote the development of infrastructures and innovative technologies to improve individual well-being health and treatment of diseases in society.

Due to technological advances, personalized medicine currently plays a relevant role in biomedical researching programs of several countries that is why prizes such as CTSA-Clinical and Translational Science Awards of the National Institute of Health of USA (http://www.ncats.nih.gov/ctsa) [17] are giving. Under this context, Computing Translational Medicine (CTM) is a clear example of use related to biomedical semantic, in where some developments (Health Care and Life Sciences Interest Group) have demonstrated the potential use of Semantic Web technology on CTM research [18].

Some specific works are present in the literature, for example, IPHealth is an intelligent system based on open, linked (related to medicine information) and big data, and designed to support tasks as making decision or learning on the health domain [10]. The purposes of IPHealth are to provide support to share diagnoses information from the Web and the other is to support learning activities on medicine. This tool has been developed in Spain and it represents one of the bases for the present proposal. Another example is OptiqueVQS described in [21], which is an approach designed to provide natural communication medium between end users and computers based on ontologies.

With respect to knowledge representation for medical data, there are many works that provide techniques very similar to this proposal. For example, CDKRM (Cancer Disease Knowledge Representation Model) is described in [22] as a model where concepts, relationships, restrictions and characteristics of cancer disease are represented; this document also describe MedTAS which is an information system approach based on an open-source framework that use natural language processing principles, machine learning and rules to discover relevant data and populate elements on the CDKRM from pathologist reports.

3 The proposal

Technological advances in translational medicine, and particularly in personalized medicine, are important and noticeable. Likewise, it is noteworthy that other advances in information and communication technologies facilitate the use of applications and information sharing as Internet media; where also a large amount of information can be used for specific tasks, such as provide specific information to support the customized diagnostic generating and tailored treatments to the patient's specific characteristics [22]. These elements can help to bring personalized medicine to Chilean citizens.

The proposed study combines multiple techniques for data acquisition and validation, knowledge representation, and sharing medical treatment information. The goal of this proposal is to determine how to present previous data, knowledge and relevant information related to diagnoses and treatments, when this information is hosted in the cloud and shared appropriately. The possibility of include all these elements in a computer system (that provides services to meet the needs of personalized treatments, availability of early diagnostic, it is considered to provide benefits to Chileans and facilitate the implementation of a more inclusive medicine. The scope of this project will be limited to hypertensive disease in Chile for its relevance as the third leading cause of circulatory disease, and because this proposal is related to a previous project conducted in Spain with results as [8], [10].

The architecture of the Intelligent System is detailed in Figure 1. It is suggested to use a combined architecture for sharing diagnoses (Intelligent Sharing Diagnoses -ISD), being the offline part, the communication module with the specialist ISD doctor. The specialist will be able to interact with the system by consulting and adding information about treatments, previous illnesses, and medical advances, generating information related to tailored diagnoses of patients.
In order to obtain support information for the diagnosis, the ISD module will have a communication with an online system which will contain functions of natural language recognition (Natural Processes Module - NPM module), and other functions of searching in free-access data on Internet. NPM module will prepare phrases to be searched in repositories of medical information as described in Figure 1.

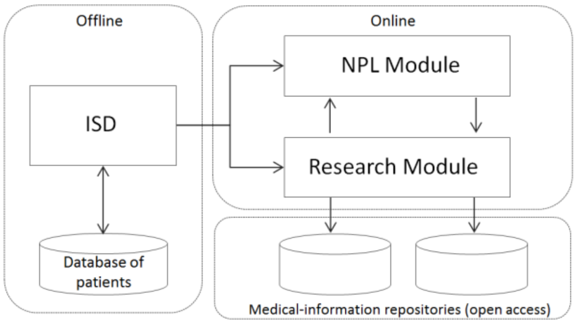

Figure 1. The ISD proposed architecture.

A knowledge representation based on domain ontology will be proposed to support the information retrieval related to biomedical data, diagnoses, and treatments, among others. This representation can be a specific ontology for Chilean patients which should be stored in the cloud. The notable growth of health information from Internet, particularly in text form, has created a great demand for software, able to turn textual information into useful knowledge. In order to carry out this challenging focused on Chilean health information, we describe our first idea of knowledge representation model, partially based on [22]. We propose this formalism for storing maligned tumors characteristics and their relations, including temporal information derived from medical treatments or evaluations, and its corresponding inferences (Figure 2). We aim to use this formalism to record patient's disease state, track disease progression and to related this data with information found in the Internet. In order to storing and information retrieved an international standard can be used (i.e. HL7[1]).
In order to share medical information, both ontology philosophy and a medical information standard such as HL7, can be used. The information could be managed by the software and stored in the ontology, being shared and visible for other users, who could use it under the instance of the class UserRole (inherited from the class User).

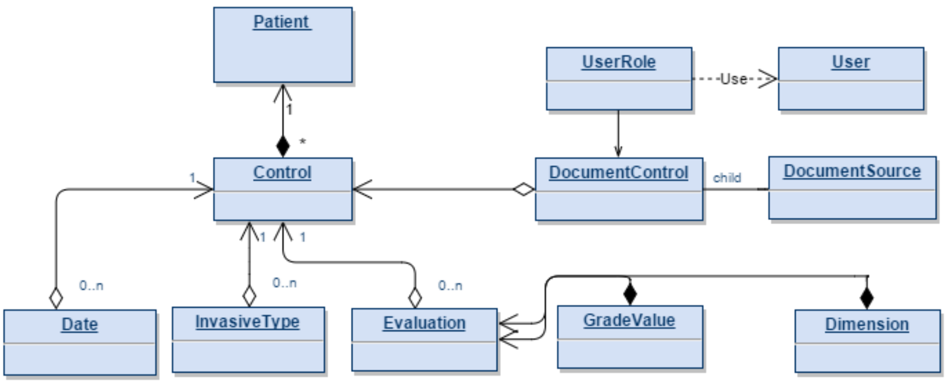

Figure 2. Maligned tumors knowledge representation model.

[1] http://www.hl7.org/implement/standards/

For the Internet information interpretation, text mining techniques can be used in combination with an inference process, using both the knowledge modeled and information retrieved from Internet. In order to support the intelligent System, a Web platform will be created. The Web could be useful for introduce new information, knowledge or important data that improve the inference process.
Figure 3 shows one process to select phrases from Internet using a key phrase (SourcePhrase), this is a coverage filter with a heuristic incorporated; the heuristic compares each key word with cloud-works and make a score of the source phase respects the cloud phases. If the two weights are similar or equal, a second check is performed to see whether the two phases are exact match and consequently the cloud phase can be selected.

```
Input:
  SourcePhrase
Output:
  Set of Internet Candidate Phrases (CandidatePhrases), and set of coincident phrases
  (InternetPhrases)
Procedure:
   1)  Select from Internet all candidate Phrases using SourcePhrase as key phrase. Allo-
       cate all selected phrases in the set CandidatePhrases
   2)  Score candidate phases:
        For each (CandidatePhrase)
       For each (word) IN (SourcePhrase)
         If CandidatePhrase has only one word
           {DELETE(CandidatePhrase) FROM CandidatePhrases}
   3)  Filter:
        For each (CandidatePhrase)
         If (EXACT_MACH(CandidatePhrase, SourcePhrase)
         {PUT(CandidatePhrase) ON (InternetPhrase)}
  Return:
    CandidatePhrases, InternetPhrase
```

Figure 3. Description of the coverage filter algorithm for select recommendations phrases from Internet.

The algorithm return two set of phrases, one of them with similar phrases (CandidatePhrases) and another one with exact similitude finding respect the source phrase (InternetPhrases). Internet-phrases selected by the algorithm described (Figure 1) can be showed on the Web framework for support the discussion between health professionals for support the diagnoses generation, personalized treatments, among others. The information resulting from the process previously describe can be uploaded to the Internet, using the Web framework.

4 Expected Outcome

The outcome of this proposal will be an Intelligence System able to assist health professionals in their labors of obtaining more comprehensible and accurate diagnoses. This prototype will combine the knowledge of doctors with information stored on large free-access repositories (free access) on Internet.
Health professionals could help each other through the system by uploading information to the cloud. This; medical information which might be used , later on , be used by treating physicians to provide personalized treatments, giving some of that information available in the Cloud to their patients and corresponding relatives.
It is expected that the resulting system of this proposal can help in tasks such as data integration, interactions, and treatments of patients. To carry out this proposal it has been made a search of previous works, verifying the feasibility of this Project. It is proposed to provide an architecture to be used by health professionals; a tool for supporting the retrieval of updated information (such as clinical care innovations), and encouraging discussion among colleagues about treatment of certain diseases, and being a tool medium for more accurate diagnoses, due to this information will be available for filters specialists (selected).
The development of this system will contribute towards to have a more inclusive medicine in Chile. Throughout the use of this computer application for medical-decision support, it is expected to improve the accuracy in diagnoses, get a better personalized application of drug treatments to people, and help doctors in rural areas to have information with the latest medical advances.

5 References

[1] “MORTALIDAD EN CHILE: 2003 y 2013” p. 2013, 2015.

[2] M. de Salud, Estrategia Nacional de Salud 2011-2020. 2010.

[3] O. R. K. Programme. HORIZON 2020 WORK PROGRAMME 2014 – 2015. Health, demographic change and wellbeing (*European Commission Decision C 8631 of 10 December 2013*), vol. 2015, No-December 2013, 2015.

[4] E. H. Shortli and J. J. Cimino. *Biomedical Informatics*, 2014.

[5] G. H. Fernald, E. Capriotti, R. Daneshjou, K. J. Karczewski, and R. B. Altman. Bioinformatics challenges for personalized medicine. *Bioinformatics*, 27(13), pages 1741–8, 2011.

[6] A. Alyass, M. Turcotte, and D. Meyre. From big data analysis to personalized medicine for all: challenges and opportunities. *BMC Med. Genomics*, 8(1), p. 33, 2015.

[7] M. Molina and V. Flores. Generating multimedia presentations that summarize the behavior of dynamic systems using a model-based approach, *Expert Systems with Applications*, 39(3), pages 2759–2770, 2012.

[8] D. Gatchet, M. Buenaga and E. Puertas. Health Sensors Information Processing and Analytics using Big Data Approaches. *UEM Eds*. (Madrid), pages 1–6. 2011.

[9] D. McMullan, “What is Personalized Medicine?”. *Genome*, 1(1), pages 32–39, 2014.

[10] M. J. Maña, M. De Buenaga, and J. Mata. IPHealth : Plataforma inteligente basada en open, linked y big data para la toma de decisiones y aprendizaje en el ámbito de la salud, *Sociedad Española para el Procesamiento del Lenguaje Natural*, pages 161–164, 2015.

[11] E. A. Estape, M. H. Mays, and E. A. Sternke. Translation in Data Mining to Advance Personalized Medicine for Health Equity, *Intelligent Information Management*, 8, pages 9–16, 2016.

[12] S. Srivastava, R. Gupta, A. Rai, and A. S. Cheema, “Electronic Health Records and Cloud based Generic Medical Equipment Interface. *Proc. 9th National Conference on Medical Information* (*AIIMS*, New Delhi), 2014.

[13] R. Wu, Gail-Joon Ahn, and Hongxin Hu. Secure sharing of electronic health records in clouds. *Collaborate-2012*, pages 711–18, 2012.

[14] V. Flores. Generación de resúmenes de comportamiento de un sistema medioambiental a partir de información no-textual. *Intel. Artif.*, 15(49), pages 52–68, 2012.

[15] G. Dalpé and Y. Joly. Opportunities and challenges provided by cloud repositories for bioinformatics-enabled drug discovery. *Drug Dev. Res.*, 75(6), pages 393–401, 2014.

[16] B. Calabrese and M. Cannataro. Cloud computing in healthcare and biomedicine. *Scalable Comput*., 16(1), pages 1–18, 2015.

[17] M. S. Lebo, S. Sutti, and R. C. Green. “Big data” gets personal. Science Translational Medicine, 8(322), pages 8–11, 2016.

[18] L. Jordan. The problem with Big Data in Translational Medicine. A review of where we’ve been and the possibilities ahead. *Appl. Transl. Genomics*, 6, pages 3–6, 2015.

[19] F. Alonso, L. Martinez, A. Perez and J.P. Valente. Cooperating between expert knowledge and data discovered knowledge: Lessons learnied. *Expert Systems with Applications*, 39(2), pages 7524-7535, 2012.

[20] N. M. Tomaŝević, M. Batić, L. M. Blanes, M. M. Keane and S. Vraneŝ. Ontology-bases facility data model for energy management. *Advanced Engineering Informatics*, 29, pages 971-984, 2015.

[21] A. Soylu, M. Giese, E. Jimenez-Ruiz, G. Vega-Gorgojo and I. Horrocks. Experiencing OptiqueVSQ: a multi-paradigm and ontology-based visual query system for end users. *Universal Access in the Information Society*, 15, pages 129-152, 2016.

[22] A. Coden, G. Savova, I. Sominsky, M. Tanenblatt, J. Masanz, K. Schuler, J. Cooper, W. Guan and P. C. de Groen. Automatically extracting cancer disease characteristics from pathology reports into a Disease Knowledge Representation Model. *Journal of Biomedical Informatics*, 45(5), pages 937-949, 2009.