Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine, ITAB 2009, Larnaca, Cyprus, 5-7 November 2009

From Taxonomies to Folksonomies: a roadmap from formal to informal modeling of medical concepts and objects

Panagiotis D. Bamidis, *Member, IEEE*, Eleni Kaldoudi, *Member, IEEE,*
Costas Pattichis, *Senior Member, IEEE*

***Abstract*— The exploitation of nomenclatures, taxonomies, terminologies and classification of medical diseases, procedures, guidelines has been pivotal to advances in the healthcare domain. Parallel to those have been evolutions in educational technology and medical education where medical educational content is often shared among different educators and institutions. This paper compares the above parallel evolutions in the light of recent developments in web technology, as well as, perception of use like Web2.0 or Web3.0. The discussion is much influenced by recent progress within the general research arena of medical education and medical content repurposing in specific. It is concluded that novel approaches making use of social networking concepts that will govern the shaping of learning objects or their associated ontological descriptions are envisaged.**

Index Terms— coding, nomenclature, taxonomy, classification, ontology, Web 2.0/3.0, social networking, content sharing, medical education, standards, folksonomy

I. Introduction

HEALTH care practices continue to evolve with technological advances integrating computer applications and information management into systems. and protocols. Diseases and healthcare are predominantly social phenomena that shape much of the perspectives of key healthcare actors [1]. The phenomenon of people (patients or healthcare professionals in our case) making decisions about "health actions" depending on similar experiences of other people in the past or certain recommendations, has formed part of our everyday "social navigation" [2]. Throughout the centuries healthcare and disease treatment have always been considered as a problem affecting both an individual patient and a health professional. Recent technological advances, however, have induced a concept and paradigm shift from the 'localized expert', and opinion driven, physician approach to a 'global expert', through the notion of collaborative evidence based, approach that uses aggregates of information as the basis to individual patient care [3].

Much of this information is Internet based thereby allowing patients and health professionals to become much aware of diseases and treatments [4, 5]. Increasingly the impact of this information sharing ability is viewed as the so-called 'User driven healthcare' [6]. This evolution has been recently termed as "Health 2.0" describing healthcare which is Patient Empowered, whereby patients have the information they need to be able to make rational healthcare decisions.

Moreover, the aforementioned two categories (patients & relatives; care-givers/professionals) also form two categories of learners when talking about a collaborative learning environment. This notion affects seriously also the way medical education is currently designed and (hopefully) practiced (in the short-term future). Much of the latter evolution of medical education is attributed to the fact that medical educational content is now probably developed collaboratively by educators, or it originates partially from the user interaction (user-generated content), but it is also often shared among different educators and institutions [7].

The aim of this paper is to compare parallel evolutions in formal and informal descriptions of medical data in an effort to appreciate the already happening chances in the provision of medical education through the availability of social web artefacts and activities [8].

So, the remainder of this paper is organized as follows. In Section II, we briefly introduce basic definitions of medical data modelling efforts, while in Section III we focus on the paradigm shift witnessed by Web2.0 developments in an effort to compare traditional with contemporarily shaped approaches. The latter are further discussed in the final sections of the paper.

II. Formal Modelling of Medical Data

In this section we revisit some of the basic definitions of formal medical data modelling like the terms of nomenclatures, taxonomies, terminologies and classification

Manuscript received October 9, 2009. This work is benefited in part from the mEducator project (Contract Nr: ECP 2008 EDU 418006 mEducator) funded under the eContentplus programme, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

P. D. Bamidis is with the Medical Informatics Laboratory, School of Medicine, Aristotle University of Thessaloniki, Greece (corresponding author, phone:+30-2310-999310; fax: +30-2310-999263; e-mail: bamidis@med.auth.gr).

E. Kaldoudi is with the Medical Physics Laboratory, School of Medicine, Democritus University of Thrace, Greece (e-mail: kaldoudi@ med.duth.gr).

C. Pattichis is with the Dept of Computer Science, University of Cyprus, Nicosia, Cyprus (e-mail: pattichi@ cs.ucy.cy).

978-1-4244-5379-5/09/$26.00 ©2009 IEEE

of medical diseases, procedures, guidelines etc. by providing first a short look in their historical development.

A. Historical details of taxonomies – the ICD example

Between 1706-1777 de Lacroix conceived the idea of classification of diseases and published it under the title *Nosologia methodical*. Then Cullen (1710-1790) simplified the system for general use and published it under the title *Synopsis nosologiae methodicae.* Farr (1807-1883) was the first medical statistician who spelt out principles that should govern a statistical classification of disease and urged the adoption of a uniform classification internationally. The Health Organization of the League of Nations had also taken an active interest in vital statistics and appointed a Commission of Statistical Experts to study the classification of diseases and causes of death, as well as other problems in the field of medical statistics. In a conference held in Paris in 1900 a resolution for a detailed classification of causes of death was adopted consisting of 179 groups and an abridged classification of 35 groups. Decades after that, in 1948, the First World Health Assembly was held to mark, among others, the beginning of a new era in international vital and health statistics, by recommending the adoption of a comprehensive programme of international cooperation in the field of vital and health statistics. Finally, in 1948, the World Health Organization became responsible for the International List of Causes of Death, which marked the new era of ICD and the "internationalisation" of codification/classification of medical concepts like diseases.

B. Basic Definitions

Taxonomy is the theoretical study of classification, including its basic principles, procedures, and rules, in other words it is the science of classification.

Classification is a method of grouping of items scientifically according to purpose and codifying them with numerical (or alpha-numerical) identification according to certain principles. "A classification of diseases can be defined as a system of categories to which morbid entities are assigned according to established criteria…" [9]. The purpose of the well known effort of the International Classification of Diseases (ICD) was to permit the systematic recording, analysis, interpretation and comparison of mortality and morbidity data collected in different countries or areas and at different times. In current healthcare practice, ICD is used to translate diagnoses of diseases and other health problems from natural language words into an alphanumeric code, which permits easy storage, retrieval and analysis of the data, especially in electronic health records.

Moreover, *coding* refers to the process of assigning a single item to a category or to a set of categories, which are denoted by codes, which in turn take various forms/formats. In other words, coding may be realised as the meaning of an item's facet/aspect. For example, the annotation of terms in the Electronic Health Record (EHR) with codes from a coding system, by use of synonyms, translations, and perhaps hierarchies. Needless to denote of course that discerning and representing meaning is an extremely difficult exercise in the case of health terminology. However, within the various advantages of coding medical data, one includes data reduction, the usage of a standardized terminology, the enabling of statistical overviews and research, the provided support to management and planning, and finally the coupling with decision-support systems.

Classification on the other hand is for example the assignment of patients exhibiting certain features to a predefined class; classification is purpose oriented and culture dependent. The pity is of course that the above two processes are frequently mixed up in the biomedical community.

Furthermore, *nomenclature* refers to the assignment codes to medical concepts; these medical concepts can be combined according to specific rules to form more complex concepts.

In *thesauri,* a list of terms is made that is used for a certain application area or domain – the lack of a common health care terminology is noticeable in this case…

Moreover, *terminology* has two meanings: the discipline of *terminology management* is synonymous with terminology work (used in ISO 704 [10]; or the set of designations used in the special language of a subject field, such as the terminology of medicine (used in both the singular and plural [11]

Some well known examples of Medical Terminologies and Classification Standards are: SNOMED, Read Clinical Terms, UMLS, GALEN, MEDCIN, CPT-4, LOINC, ICPC-2, ICD-10.

But why are terminologies used in general? They are required upon the process of fixing/stabilizing the language within a domain and a linguistic community as well as the facilitation of unambiguous intra-(and extra)-community communication. More specifically, in Healthcare Information Technology, terminologies like the above listed ones, have facilitated Semantic Indexing, the Information exchange and linking between heterogeneous systems, and served as the basis for documentation through coding and classification.

Furthermore, an *ontology* (within the notion of the science of what things exist and the way they relate to each other) is a representation of some pre-existing domain of reality which [12]: (1) reflects the properties of the objects within its domain in such a way that there obtains a systematic correlation between reality and the representation itself, (2) is intelligible to a domain expert; (3) is formalized in a way that allows it to support automatic information processing.

Before going any further let us emphasise the distinction between "Terminology", which involves communication amongst humans as well as communication between human and machines, with "Ontology" which is a representation of "reality" inside a machine involving communication amongst machines, but also interpretation by machines.

In biomedical sciences the majority of concept/terminology-based systems are attempting to make implicit knowledge explicit. Ontologies aim to push explicitness further by harvesting the reasoning by machines, facilitating classification, enabling prediction, and allowing the triggering of alerts.

But generally speaking, concept-based terminologies and standardisations are mechanisms to improve communication of messages by humans, but it is probably not the right device to explain why reality is what it is, or to reason about reality, or to make machines understand what is real, or to integrate across different views, languages, and conceptualisations [12].

C. STANDARDS IN ONLINE MEDICAL EDUCATION

The e-learning community colloquially uses the word standard to denote either [13].

- an *official standard*: a set of definitions, requirements, formats, and design guidelines for elearning systems or their components documented and approved by a recognized standards organization.
- A *de facto standard*: similar to the above, but lacking formal approval from a recognized standardization body
- A *specification*, that is a document usually developed and promoted by organizations or consortia of partners from academia, industry, and educational institutions, capturing a rough consensus in the e-learning community
- A *reference model*, which is an adapted and reduced version of a combination of standards and specifications focusing on architectural aspects of an e-learning system, definitions of parts of the system, and their interactions

For example, Packaging Educational Standards that like the ADL's Sharable Content Object Reference Model (SCORM) and SCORM for Healthcare [14], which has been designed as a set of eXtensible Markup Language (XML) based specifications that can define, manage, access and deliver modular educational objects so that they are easily shared among different e-learning management systems. The Content Structure Format (CSF) is an XML-based representation of a course structure that can be used to define all of the course elements, structure, and external references necessary to move a course from one LMS (Learning Management System) environment to another. The IMS CP (Content Packaging) consists of a part for Metadata and a part for Content Structure organizations. The Metadata Part of the IMS CP includes metadata standards, like Healthcare LOM, IEEE LOM, IMS Metadata, or others.

For every standard, specification, or reference model that exists, a crucial question that arises whether it is going to be widely adopted and maintained through the years...[13]

D. THE WEB VISION

Tim Berners-Lee in his vision for the Web stated that "... a goal of the Web was that, if the interaction between person and hypertext could be so intuitive that the machine-readable information space gave an accurate representation of the state of people's thoughts, interactions, and work patterns, then machine analysis could become a very powerful management tool, seeing patterns in our work and facilitating our working together through the typical problems which beset the management of large organizations." [15]. And more specifically his vision for the semantic web was that "is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation"…

III. THE CURRENTLY WITNESSED WEB USE PARADIGM SHIFT & THE BIRTH OF FOLKSONOMIES

Current innovations in ICT systems and services mark a notable switch from an "information society", characterized by immense information seeking efforts largely dependent on the availability of pre-defined and standardized data, to a "knowledge society", that is, a society based on knowledge as a value [7]. This emphasizes essentially the cognitive advancement and involvement of each individual, and not only modifies quickly and habitually the way people work, but also the way the educational processes are designed and hopefully largely practiced.

In 2008 the World Wide Web Consortium (W3C) (http://www.w3.org/) has established the Semantic Web for Health Care and Life Sciences Interest Group (HCLS IG) (http://www.w3.org/2001/sw/hcls/) to develop and support the use of semantic web technologies to improve collaboration, research and development, innovation, and adoption in the domains of Health Care and Life Sciences [16]. A foreseen application and impact in the case of medical education is to enable more efficient information search and retrieval; what is more is the ability to construct personalized information searches tailored to a specific educational objective in addition to other administration and semantic linking of educational content amongst institutions and degrees [17].

However, it can be argued that the real predecessor of web 2.0 notions and technology is the programming paradigm of web services and service oriented architectures. Recent web services handling huge amounts of user generated content like Flickr or Del.icio.us, have employed a technique that has come to be known as *folksonomy* [18]. Folksonomy is a user created taxonomy where users provide short keywords (tags) about their uploaded content which are however generated by them rather selecting them from pre-fixed/pre-defined categories. As users can assign such decentralised keywords to their content, it is imperative that the semantics of the various information systems are self-evolved as there are no restrictions to what information one can submit. Users have absolute control re the organisation of the information resources, thereby making the system self maintainable and super flexible [1].

IV. PUTTING THE THREADS TOGETHER

The aforementioned notion of a folksonomy is somehow completely in contrast with prior taxonomy developments described in section II. A simple look at table 1 makes this evident.

Table 1 A comparison of classification, coding and Web2.0 communities (folksonomies - collective emergent ICT behavior).

Requirements for a classification	**Requirements for computer-assisted coding systems**	**Web 2.0 characteristics of user networks & communities**
Domain completeness	Allow for the use of synonyms	social networking
Nonoverlapping classes (mutual exclusiveness)	Allow for the use of lexical variations	participation
Homogeneous ordering (one principle per level)	Insensitive to spelling errors	personalisation
Clear criteria for class boundaries	Reliability	Openness and agility
Appropriate level of detail	consistent operation (insensitive to ordering of terms)	Collaboration and sharing
Unambiguous and complete guidelines for application	correct	Reuse and apomediation [19]

What is more, is that Web services describe themselves through a standardized Web Service Description Language (WSDL) document, and can be published to one or more Intranet or Internet repositories for potential users to locate through a standard Universal Description, Discovery and Integration (UDDI) registry. This core technology that supports Web 2.0 is continuously evolving and growing, as new specialized formats, standards, and protocols emerge. Examples include the RSS data/metadata XML format, the FOAF (Friend of a Friend) and XFN (XHTML Friends Network) protocols involved in social networking applications [7].

V. CONCLUSION

It is common knowledge that traditional medical taxonomies have not been as widely accepted as originally hoped; they haven't resolved some of the major medical problems either (at least not as they ought to). As folksonomies encompass a "self-organization" aspect, a crucial property of well functioning complex adaptive systems, it is imperative to hope that the allowed user relationship building and the formation of personal cognitive conceptual and operational models as a part of social cognitive interactions labeled social cognitive ontological constructs (SCOCs) [1] may find key applications in the area of medical education and content sharing [20, 21].

REFERENCES

[1] R Biswas, C M. Martin, J Sturmberg, K Mukherji, E Wen Huo Lee, S Umakanth, AS Kasthuri, "Social Cognitive Ontology and User Driven Healthcare", In S Hatzipanagos, S Warburton, "Handbook of Research on Social Software and Developing Community Ontologies", Information science reference Series, IGI Global, Hershey, New York, 2009

[2] Dourish, P., & Chalmers, M. (1994). Running Out of Space: Models of Information Navigation. Paper presented at the HCI'94.

[3] Biswas, R., Umakanth, S., Strumberg, J., Martin, C. M., Hande, M. & Nagra, J. S. (2007b). The process of evidence-based medicine and the search for meaning. Journal of Evaluation in Clinical Practice, 13, 529–532.

[4] Murray, E., Lo, B., Pollack, L., Donelan, K., Catania, J., White, M. et al. (2003). The impact of health information on the internet on the physician-patient relationship: patient perceptions. Arch Intern Med., 163, 1727-34.

[5] Larner, A. J. (2006). Searching the internet for medical information: frequency over time and by age and gender in an outpatient population in the UK. J Telemed Telecare 2006, 12 186-8.

[6] www. patientslikeme.com, last access Sept. 2009.

[7] E. Kaldoudi, S. Konstantinidis, P. Bamidis, "Web 2.0 Approaches for Active, Collaborative Learning in Medicine and Health", Peer-Reviewed Chapter in: S. Mohammed and J. Fiaidhi (eds.), " Ubiquitous Health and Medical Informatics: Advancements in Web 2.0, Health 2.0 and Medicine 2.0", IGI Global, Hershey, PA, USA (In press)

[8] S Hatzipanagos, S Warburton, "Handbook of Research on Social Software and Developing Community Ontologies", Information science reference Series, IGI Global, Hershey, New York, 2009.

[9] Worlf Health Organisation, www.who.org

[10] International Standards Organisation, www.iso.org

[11] Ceusters W, "MIE 2006 Tutorial on Standards and Ontology - Part 3: SNOMED - CT", MIE 2006 web page, www.mie2006.org, last access, Oct, 2006.

[12] Ceusters W, Smith B, Kumar A, Dhaen C. "Mistakes in Medical Ontologies: Where Do They Come From and How Can They Be Detected?", in: Pisanelli DM (ed) IOS Press, Studies in Health Technology and Informatics, vol 102, 2004

[13] Konstantinidis ST, Kaldoudi E, Bamidis PD, "Enabling content sharing in contemporary medical education: a review of technical standards" In Proc. Of ICICTH 2009.

[14] Smothers V, "SCORM for Healthcare, Specifications and Description Document", http://medbiq.org/working_groups/ learning_objects/ SCORMforHealthcareSpecification.pdf, last visited on January 2009

[15] Berners-Lee T, Hendler J, Lassila O, The Semantic Web, Scientific American, May 2001

[16] Vandervalk, B.P., McCarthy, E. L., & Wilkinson M.D. (2009). Moby and Moby 2: Creatures of the Deep (Web). Briefings in Bioinformatics, 10(2), 114-128.

[17] Ohler, J. (2008), The semantic web in education. EDUCAUSE Quarterly, 31(4), 7-9.

[18] Golder, S. A., & Huberman, B. A. (2005). The Structure of Collaborative Tagging Systems.

[19] Eysenbach G. (2007) From intermediation to disintermediation and apomediation: new models for consumers to access and assess the credibility of health information in the age of Web2.0. Studies in Health Technology and Informatics, 129(Pt 1):162-166.

[20] Dovrolis N, Konstantinidis ST, Bamidis PD, Kaldoudi E, "Depicting Educational Content Re-purposing Context and Inheritance", This volume.

[21] Kaldoudi E, Dovrolis N, Konstantinidis ST, Bamidis PD, "Social Networking for Learning Object Repurposing in Medical Education", The Journal on Information Technology in Healthcare 2009; 7(4): 233–243.