Describing a Knowledge Field with Topic Maps

Artemis Chaleplioglou[1,2], Ioannis A. Papadakis[2,†] and Sozon Papavlasopoulos[2]

[1]Library Department, Biomedical Research Foundation of the Academy of Athens, Athens, Greece
[2]Department of Archives, Library Science and Museology, Faculty of Information Science & Informatics, Ionian University, Corfu, Greece

[†]This author passed away on August 30, 2017.

Abstract: Semantic web technologies are not focus intrinsically to the terms describing a scientific field but extend further towards meaningful relationships, a task that in traditional systems of knowledge representation is allocated to humans. In a standard approach the most valuable tool in the study and comprehension of a scientific disciple is the concept map. This map is a well-structured and classifiable graph that represents scientific information optically. Evolution of this representation in the knowledge organization systems are the Topic Maps. In Topic Maps the different concepts are hierarchical and relational connected in a virtual map, while URIs ensure the interoperability of this form of ontology. Topic Maps allow to the developer a high degree of expressiveness, especially in the case of multidimensional relationships. Herein, we generated a Topic Map for the Cardiology field, using a minimum but sufficient set of terms to cover almost 70% of cardiovascular terminology. Cardiology is in focus because the cardiological disorders remain the first cause of morbidity and mortality worldwide, while the available semantic description systems for this field are limited. We build an experimental cardiological ontology by utilizing XAMPP, Topincs, Wandora and Ontopia as developmental tools. The resulting map is currently under evaluation and enrichment to include in addition to the clinical entities, cell biology, implantable instruments, and pharmaceuticals. We anticipate that in future this tool would be a valuable addition in information science and cardiology which will assist both clinical and experimental scientists.

Keywords: Semantic web, Topic Maps, Cardiology

1. Introduction

The Semantic Web is based on Tim Berners-Lee bipartite vision for the Web as a collaborative medium of distinct databases and as a machine readable, understandable and processable information matrices (Berners-Lee, and

Received: 2.5.2018 / Accepted: 28.5.2018 ISSN 2241-1925

Hendler, 2001). Thus, data content can be used by machines, not just for display purposes, but also for automation, integration and reuse across applications. However, the Semantic Web per se is not a distinct entity but rather a logical extension of the current Web. The transition from HTML Web pages to Semantic markup ones, and from ordinal data to increasingly intelligent meta-data respectively, lead the web to semantics. The technologies needed for the development of the Semantic Web are: semantically enriched markup languages of a particular knowledge domain, ontologies, semantic search, semantic markup of pages and services like reasoners (Nelson, Avraham, Shoemaker, May, Ware, and Gessler, 2010).

Topic Maps are the International Organization for Standardization (ISO) standard for describing knowledge structures and associating them with information resources (ISO/IEC 13250:2003). They serve the semantic web as highly descriptive ontologies that annotate documents with conceptual information (Noy, Sintek, Decker, Crubézy, Fergerson, and Musen, 2001; Park, Hunting, and Foreword By-Engelbart, 2002). The central idea behind them, the concept mapping, was originally developed by Novak and the members of his research group as a means of representing frameworks for the interrelationships between concepts (Novak, Bob Gowin, and Johansen, 1983). The theoretical background of concept maps is grounded in Ausubel's knowledge psychology theory which focuses on how individuals integrate new learning into existing conceptual frameworks by making explicit, conscious connections between concepts as a way to integrate information into memory (Ausubel, 1968). The trajectory of concept mapping consists in the connection of concept words or phrases with linking words or phrases to form complete thoughts called 'propositions' which gradually build up into visual conceptual graphs (Novak, 1990). However, Topic Maps are far more enriched since they represent both a knowledge organization system (KOS), similar to keyword systems, classification systems, thesauri or taxonomies, and a computer language for the explicit description of concepts in the form of semantic graphs (Pieterse, and Kourie, 2014; Slavic, 2005). In a topic map the concepts are classified both hierarchically and relationally while they are connected with internal and external content. Each concept, a topic, is described into at least two levels, the knowledge one, which is internal to the map, and the informative one, which is located external to the resources (Ahmed, 2000). There is a clear distinction between topic-to-topic and topic-to-resource relationships. Within the topic map space, the topics are connected with associations. A topic may also be related to any number of resources by its occurrences. This partitioning allows a topic map developed for one set of resources to be reused to index a different set of resources. Thus, a topic map can be considered to be a portable form of knowledge (Ahmed, 2000). A third level of information may exist, the reference of the identity of the subject, while within the topic map a set of rules may be set up including roles, association types and scopes of the topics (Johnsen, 2010). The URIs of RDF resources are equated with the subject indicators of topics in topic maps (Garshol, 2001). Noteworthy, many knowledge oriented webpages such as Wikipedia (https://en.wikipedia.org/wiki/) (Mesgari, Okoli, Mehdi, Nielsen, and Lanamäki, 2015; Nielsen, 2012; Rafi, Hassan, and Shaikh, 2012) and Stanford Encyclopedia of Philosophy (https://plato.stanford.edu/index.html) (Hjørland, 2015; Kim, Choi, Shin, and Kim, 2007), with structural organized into subject groups content, they are using topic map semantic transformation of information.

2. Motivation

In this report we are using the Topic Maps standard for the building of a cardiology topic map in order to achieve its semantic representation. To the best to our knowledge, although cardiovascular diseases, including coronary heart disease, stroke, hypertension, cardiomyopathies and heart failure, represent the predominant causes of morbidity and mortality worldwide (Gaziano, Bitton, Anand, Abrahams-Gessel, and Murphy, 2010; Organization, 2016), they are not sufficiently supported in the Semantic web. In cardiology and medicine in general, scientists from different disciplines that use different vocabularies and datasets, have to share information in order to answer system biology problems (Hudson, and Collins, 2015; Krumholz, Brindis, Brush, Cohen, Epstein, Furie, Howard, Peterson, Rathore, and Smith, 2006). In addition, the management of a huge volume of biomedical information as well as the complex nature and heterogeneity of data are challenging tasks for the information scientist (Bellazzi, 2014; Marx, 2013; Varmus, Klausner, Zerhouni, Acharya, Daar, and Singer, 2003). Concept maps have been proposed as the solution for the optimization of biomedical education from heterogeneous resources (Stewart, 1979; Wallace, and Mintzes, 1990). The building of semantic maps is based in the identification of the topics that best match and organize different databases, while setting up the rules for the transformation of data is necessary for the accurate assignment of the particular instances of each topic from one data resource to the next (Ramalho, Librelotto, and Henriques, 2005). The reuse of the long tradition of biomedicine in the systemic knowledge representation, organization and classification (Godfray, 2007; Graunt, 1977), may facilitate the development of a topic map. However, the generation of semantic topic maps and the setting of rules is performed manually and not automatically as it is feasible with other semantic web approaches. This is a setback in interoperability, one of the key concepts of the Semantic web, and adds a significant degree of difficulty in the management of biological Big Data. On the other hand Topic Maps standard offers unique freedom and descriptiveness to the publisher of an ontology which is of critical importance for the experts of a field like cardiology. Therefore, it could be anticipated that a cardiology topic map will improve the functionality of existing literature and data towards the assimilation of new discoveries, as well as the interplay between clinical and experimental cardiologists and pharmacologists.

3. **Methodology**

Our methodology for consolidating cardiological data evidence using the Topic Maps approach was based on the topic map ontology developmental guidelines by Garshol (Garshol, 2006). In brief, a five phases process was applied: (1) the startup phase, where the definition of the purpose of the project took place, as cardiological clinical and biological data mining from scientific literature and data with knowledge auto-discovery dynamics; (2a) the end-user phase, whereas the identification of the end-user needs took place, as the curated scientific information on specific cardiological topics that were collected from Medical Subject Headings (MeSH) (https://www.ncbi.nlm.nih.gov/mesh), on the basis of the American Heart Association (AHA) (https://professional.heart.org/) guidelines and statements; (2b) the analysis phase, where the aim was the understanding of the data and their resources after a thorough cardiological topic search in the open access biomedical database list of the Nucleic Acid Research (NAR) scientific journal (Galperin, Fernandez-Suarez, and Rigden, 2017) that allow the identification of the best fitting databases to the coverage, relevance and awareness criteria; (3) the drafting phase, where the topics were hierarchically clustered and analyzed with concept mapping tools graphically and in spreadsheets according to their associations in PubMed Central (https://www.ncbi.nlm.nih.gov/pmc/) scientific publications; (4) the interaction design phase, where the documentation of the topic map took place, with the assignment of all the components and interactions of the knowledge layer of the topic space, including association types, topic-to-topic associations, and roles, as a representation of the role played by a given topic as a member of an association; and (5) the verification phase that the topic map ontology actually matches the data in external sources and that topic-to-resource, occurrence types and occurrences, the information layer of the topic map, are valid. For the development of the cardiology topic map we used the Topincs (https://www.topincs.com) and Ontopia (https://github.com/ontopia/ontopia) applications in a PHP computation development environment set up with XAMPP (https://www.apachefriends.org), and the Wandora software (http://wandora.org/www/).

4. **Results**

Using the AHA recommendations as a pathfinder to identify the specific information needs of cardiology experts at the end-user phase of the topic map building, we determine a set of 115 terms that sufficiently cover the vast majority, almost 70% of cardiovascular vocabulary. These terms and their hierarchies were retrieved from MeSH, namely: heart, aorta, aortic aneurysms, coronary arteries, epicardium, myocardium, endocardium, sinoatrial node, atrioventricular node, diabetes mellitus, hypertension, electrocardiography, echocardiogram, cardiac computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), restrictive cardiomyopathy, ventricular hypertrophy (VH), channelopathies, arrhythmias, atrial fibrillation (AF), supraventricular tachycardia (SVT), ventricular tachycardia (VT), ventricular fibrillation (VF), torsades de pointes (TdP), sudden cardiac arrest, sudden cardiac death (SCD), syncope, premature ventricular contraction (PVC), cardiopulmonary resuscitation (CPR), artificial cardiac pacemaker, implantable cardioverter defibrillator (ICD), cardiac defibrillator devices, cardiac monitors, antiarrhythmic agents, anticoagulants, antiplatelet agents, dual antiplatelet therapy (DAPT), angiotensin-converting enzyme (ACE) inhibitors, angiotensin II receptor blockers (ARB), angiotensin-receptor neprilysin inhibitors (ARNIs), beta blockers, combined alpha and beta-blockers, calcium channel blockers, cholesterol-lowering medications, digitalis preparations, diuretics, vasodilators, inotropic agents, heart failure, heart valve, mitral valve, tricuspid valve, aortic valve, pulmonary valve, valvular heart disease, aortic insufficiency, mitral insufficiency, pulmonary insufficiency, tricuspid insufficiency, congenital heart disease, aortic regurgitation, endocarditis, rheumatic heart disease, atherosclerosis, calcification, pericarditis, dyslipidemia, myocardial infarction, cardiac fibrosis, aortic stenosis, brain natriuretic peptide (BNP), coronary artery disease (CAD), cerebrovascular insult, thrombosis, nephropathy, fibrinolysis, angioplasty, artificial heart valve surgery, atherectomy, bypass surgery, cardiomyoplasty, heart transplantation, radiofrequency ablation, stent placement, transmyocardial revascularization (TMR), Holter monitoring, exercise cardiac stress test (ECST), cardiac enzymes test, troponin, creatine kinase, C-reactive protein (CRP), fibrinogen, homocysteine, lipoproteins, cholesterol, high density lipoprotein cholesterol (HDL-C), low density lipoprotein cholesterol (LDL-C), triglycerides, prothrombin, cardiac catheterization, atrial flutter, cardiomyocyte, mitochondria, sarco(endo)plasmic reticulum, hemodynamics, cardiac input, heart rate, end-systolic volume, end-diastolic volume, sonographer, suction volume, left ventricular ejection fraction (LVEF). These concepts were subsequently assigned, as topics and subtopics, to anatomical, disease, intervention or diagnostic topic types.

During the analysis phase, the open access biomedical databases enlisted in NAR were analyzed to identify topic-to-resource, occurrence types and occurrences and select the best fitting ones to define the external content of our developing cardiology topic map. Fifteen biomedical databases matched to our criteria, in specific: the bibliographic resources, PubMed (http://www.ncbi.nlm.nih.gov/pubmed/), Europe PubMed Central (http://europepmc.org/) and the eMedicine (Medcape) (http://emedicine.medscape.com/); the clinical resources, Central Cardiac Audit Database (http://www.ucl.ac.uk/nicor), Clinical Trials Registry (http://www.clinicaltrials.gov/), Qresearch (http://www.qresearch.org/), Clinical Practice Research Datalink (CPRD) (http://www.cprd.com/), The Health Improvement Network (THIN) (http://www.csdmruk.imshealth.com/) and the Disease Database (http://diseasesdatabase.com/); the pharmacological, ChEMBL database of bioactive drug-like small molecule (https://www.ebi.ac.uk/chembldb), the DrugBank resource (http://www.drugbank.ca/), the PubChem (http://pubchem.ncbi.nlm.nih.gov/)

and the Small Molecule Pathway Database (SMPDB) (http://www.smpdb.ca/); and the genetic resources, Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.ad.jp/kegg) and Online Mendelian inheritance in man (OMIM) (http://www.omim.org).

In the third phase of our ontology development, the drafting phase, the knowledge layer of our topic map was set up though hierarchical clustering and topic-to-topic associations. The associations were extracted from the data, utilizing the pivotal Systematized Nomenclature of Medicine- Clinical Terms (SNOMED CT) upper concepts experience (Stearns, Price, Spackman, and Wang, 2001), and their evolution over time (Schulz, Suntisrivaraporn, Baader, and Boeker, 2009). In particular the concepts used in the classification of topics were: body structure, clinical finding, event, observable entity, pharmaceutical/biologic product, physical force, physical object, procedure, specimen, staging and scales, substance; and the concepts used in topic-to-topic associations were: active ingredient of, associated with, causative agent of, component of, definitional manifestation of, dependent of, direct morphology of, direct substance of, dose form of, has interpretation, has method, has occurrence, has part, has pathological process, has property, has route of administration, has scale type, has severity, has surgical approach, ICD-O-3 Code, inactivation indicator, occurs after, occurs before, occurs in, part of, pathological process of, property of, revision status of, severity of, surgical approach, and type; some of which were selected to satisfy compatibility and interoperability with other formats of structured information. The topics and associations were used for the design of the topic map working space during the interaction design phase, but the associations allow the connection of topics with their occurrences, stored in the previous mentioned relative databases. Thus, our developing topic map was interconnected with the external resources and the validity of these connections was tested in the final phase of the topic map ontology building.

5. Conclusions

The amounts of data available on the Web represent a goldmine for data-driven applications and services (Berners-Lee, and Shadbolt, 2011) but at the same time they represent a minefield of non-curated, possibly harmful or misleading information (Chow, and Golle, 2009). In addition, the data is available in different formats, stored in different databases. The biosciences are no different, while at the same time have to retain a fast pace of scientific discovery to cure human disease. Cardiology fights the most common causes of morbidity and mortality worldwide. Herein, we describe the building of a cardiological topic map ontology that allows the retrieve of information from multiple resources through the TAO principle of topics, associations, and occurrences (Pepper, 2000). The presented experimental model provides many opportunities for possible direct applications, but also has to be extended to cover more cardiological entities and concepts. Moreover, its applicability is strongly depended on the availability of third party semantic tools with user-friendly interfaces that currently are unavailable. A limitation of, the highly descriptive and useful in data modeling, Topic Maps standard is the lack of sufficient support by software developers because of the transition of their interest to other more common and simplest in format ontology building languages.

References

Ahmed, K. (Year). "Topic Maps for repositories." Paper presented at the Proceedings of the XML Europe, 2000.

Ausubel, D. P. (1968). *Educational psychology; a cognitive view*. Holt, New York,.

Bellazzi, R. (2014). Big data and biomedical informatics: a challenging opportunity. *Yearbook of medical informatics* 9, 8.

Berners-Lee, T., and Hendler, J. (2001). Publishing on the semantic web. *Nature* 410, 1023-4.

Berners-Lee, T., and Shadbolt, N. (2011). There's gold to be mined from all our data. *The Times, London*

Chow, R., and Golle, P. (Year). "Faking contextual data for fun, profit, and privacy." Paper presented at the Proceedings of the 8th ACM workshop on Privacy in the electronic society, 2009.

Galperin, M. Y., Fernandez-Suarez, X. M., and Rigden, D. J. (2017). The 24th annual Nucleic Acids Research database issue: a look back and upcoming changes. *Nucleic Acids Res* 45, D1-D11.

Garshol, L. M. (2001). Topic maps, rdf, daml, oil. *A comparison*

Garshol, L. M. (Year). "Towards a methodology for developing topic maps ontologies." Paper presented at the International Conference on Topic Map Research and Applications, 2006.

Gaziano, T. A., Bitton, A., Anand, S., Abrahams-Gessel, S., and Murphy, A. (2010). Growing epidemic of coronary heart disease in low- and middle-income countries. *Curr Probl Cardiol* 35, 72-115.

Godfray, H. C. J. (2007). Linnaeus in the information age. *Nature* 446, 259-260.

Graunt, J. (1977). Natural and political observations mentioned in a following index, and made upon the bills of mortality. In *Mathematical Demography*, pp. 11-20. Springer.

Hjørland, B. (2015). Theories are Knowledge Organizing Systems (KOS). *Knowledge Organization* 42,

Hudson, K. L., and Collins, F. S. (2015). Sharing and reporting the results of clinical trials. *Jama* 313, 355-356.

Johnsen, L. (2010). Topic Maps From Information to Discourse Architecture. *Journal of Information Architecture* 2, 5-17.

Kim, J.-M., Choi, B.-I., Shin, H.-P., and Kim, H.-J. (2007). A methodology for constructing of philosophy ontology based on philosophical texts. *Computer Standards & Interfaces* 29, 302-315.

Krumholz, H. M., Brindis, R. G., Brush, J. E., Cohen, D. J., Epstein, A. J., Furie, K., Howard, G., Peterson, E. D., Rathore, S. S., and Smith, S. C. (2006). Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation* 113, 456-462.

Marx, V. (2013). *Biology: The big challenges of big data*. Nature Publishing Group.

Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F. Å., and Lanamäki, A. (2015). "The sum of all human knowledge": A systematic review of scholarly research on the content of Wikipedia. *Journal of the Association for Information Science and Technology* 66, 219-245.

Nelson, R. T., Avraham, S., Shoemaker, R. C., May, G. D., Ware, D., and Gessler, D. D. (2010). Applications and methods utilizing the Simple Semantic Web Architecture and Protocol (SSWAP) for bioinformatics resource discovery and disparate data and service integration. *BioData Min* 3, 3.

Nielsen, F. Å. (2012). Wikipedia research and tools: Review and comments.

Novak, J. D., Bob Gowin, D., and Johansen, G. T. (1983). The use of concept mapping and knowledge vee mapping with junior high school science students. *Science education* 67, 625-645.

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of research in science teaching* 27, 937-949.

Noy, N. F., Sintek, M., Decker, S., Crubézy, M., Fergerson, R. W., and Musen, M. A. (2001). Creating semantic web contents with protege-2000. *IEEE intelligent systems* 16, 60-71.

Organization, W. H. (2016). Hearts: technical package for cardiovascular disease management in primary health care.

Park, J., Hunting, S., and Foreword By-Engelbart, D. C. (2002). *XML Topic Maps: creating and using topic maps for the Web*. Addison-Wesley Longman Publishing Co., Inc.

Pepper, S. (Year). "The TAO of topic maps." Paper presented at the Proceedings of XML Europe, 2000.

Pieterse, V., and Kourie, D. G. (2014). Lists, taxonomies, lattices, thesauri and ontologies: paving a pathway through a terminological jungle.

Rafi, M., Hassan, S., and Shaikh, M. S. (2012). Content-based text categorization using wikitology. *arXiv preprint arXiv:1208.3623*

Ramalho, J. C., Librelotto, G. R., and Henriques, P. R. (Year). "Metamorphosis–a topic maps based environment to handle heterogeneous information resources." Paper presented at the International Conference on Topic Map Research and Applications, 2005.

Schulz, S., Suntisrivaraporn, B., Baader, F., and Boeker, M. (2009). SNOMED reaching its adolescence: ontologists' and logicians' health check. *Int J Med Inform* 78 Suppl 1, S86-94.

Slavic, A. (2005). Knowledge organization systems, network standards and Semantic Web. In *Informacijske znanosti u procesu promjena*. Zavod za informacijske studije, Sveuciliste u Zagrebu.

Stearns, M. Q., Price, C., Spackman, K. A., and Wang, A. Y. (Year). "SNOMED clinical terms: overview of the development process and project status." Paper presented at the Proceedings of the AMIA Symposium, 2001.

Stewart, J. (1979). Concept Maps: A Tool for Use in Biology Teaching. *American biology teacher* 41, 171-75.

Varmus, H., Klausner, R., Zerhouni, E., Acharya, T., Daar, A., and Singer, P. (2003). *Grand challenges in global health*. American Association for the Advancement of Science.

Wallace, J. D., and Mintzes, J. J. (1990). The concept map as a research tool: Exploring conceptual change in biology. *Journal of research in science teaching* 27, 1033-1052.