



Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe



Review Article

Biomedical ontologies—A review



Bogumil M. Konopka*

Institute of Biomedical Engineering and Instrumentation, Wrocław University of Technology, Wrocław, Poland

ARTICLE INFO

Article history:

Received 3 February 2014

Received in revised form

23 June 2014

Accepted 30 June 2014

Available online 17 July 2014

Keywords:

Biomedical ontologies

Data annotation

Information storage

ABSTRACT

Current societies undergo a transformation into information societies. “Digitalization” is progressing in every aspect of life, including health care. Handling the increasing flow of biomedical data presents a serious challenge to researchers and clinicians. Ontologies – controlled vocabularies that allow describing the meaning of data (its semantics) in a human and machine readable way are used more and more often to aid processing of information in biomedical research and in healthcare systems. The aim of this work is to bring closer the field of ontologies to the medical society. The theoretical basics are presented and exemplified with a range of ontologies used for describing diseases, medications, proteins, experimental procedures, etc. Currently the multitude of ontologies is an obstacle in further data integration. Unified Medical Language System (UMLS) and OBO Foundry (Open Biomedical Ontologies) are projects started to counteract this problem. UMLS aims at merging existing vocabularies, while OBO initiative is based on coordinated, harmonic development of new ontologies and reformation of existing ones. The pros and cons of both philosophies are presented. The final section of the article features examples of ontology applications.

© 2014 Nałęcz Institute of Biocybernetics and Biomedical Engineering. Published by Elsevier Sp. z o.o. All rights reserved.

1. Introduction

The tremendous progress in the domain of biological sciences especially in high-throughput experimental techniques leads to a rapid increase in the amount of available biological data. Also development of new diagnostic methods, therapies and medications is a source of a continuous information flow. Acquisition and processing of such amount of data cannot be performed manually. Even if it is supported by computer software it may still be problematic. On the one hand, natural language, which is often used in medical records, is ambiguous: (i) different words can be used to describe the same

objects or concepts (e.g. 'cellular membrane' = 'lipid bilayer') (ii) in different context the same words may have different meanings ('function of a protein', 'function of an organ'). On the other hand, heterogeneity of data formats and data structure makes it hard to integrate the information in larger systems, which hinders or even makes it completely impossible to reuse the data. Usually data is collected and primarily used by narrow groups of people in very specific contexts, it often forms hermetic silos which are hard to access and process by other researchers. If the data is to be useful it needs to be precisely defined and described – annotated. In order to address these issues so-called 'ontologies' are designed. Ontologies are computer science constructs that provide

* Corresponding author at: Institute of Biomedical Engineering and Instrumentation, Wrocław University of Technology, Wybrzeże St. Wyspińskiego 27, 50-370 Wrocław, Poland.

E-mail address: bogumil.konopka@pwr.wroc.pl

<http://dx.doi.org/10.1016/j.bbe.2014.06.002>

0208-5216/© 2014 Nałęcz Institute of Biocybernetics and Biomedical Engineering. Published by Elsevier Sp. z o.o. All rights reserved.

well-defined vocabularies which allow precise and machine-readable description of knowledge about a certain domain [1].

The success of using ontologies for biological data annotation can be confirmed by multiple examples [2–4]. Gene Ontology (GO) [5] is probably the most striking one. GO is used to describe three aspects of genes and proteins: their molecular function, the biological process they participate in and their sub-cellular localizations. Article which introduced GO was cited nearly 8000 times (as of June 2013 SCOPUS database), and until 2012 the ontology was used in over 3000 scientific projects (<http://geneontology.org/cgi-bin/biblio.cgi>).

So far, tens of biomedical ontologies have been designed. They are shared by the scientific community through two main resources: the BioPortal (<http://bioportal.bioontology.org/>) [6], which holds 370 ontologies (as of June 2014) that jointly comprise nearly 6 million terms (maintained by the NIH National Center for Biomedical Ontology [7]) and the Ontology Look-up System (OLS) (<http://www.ebi.ac.uk/ontology-lookup/>) [8] with 93 ontologies (as of June 2014) and over 2 million terms (maintained by the European Bioinformatics Institute (EBI)). Both resources are quite diverse in terms of content and it is not easy to characterize them. Often they include the same ontologies. In comparison to the BioPortal, OLS groups relatively more biological ontologies such as Yeast Phenotypes ontology or Zebrafish Anatomy and Development, which are useful to biologists and molecular biologists. Conversely, The BioPortal comprises a higher rate of ontologies useful in different branches of medicine. Out of its 370 ontologies, 85 are categorized as health-related. This includes 8 related to neurological disorders, 5 related to immunology. There are general ontologies such as the Pediatric Terminology, Ontology of Pneumology, Infectious Disease Ontology, Dermatology Lexicon, as well as more specific ones such as Bone Dysplasia Ontology or Alzheimer's disease ontology.

The multitude of available ontologies caused a repeated problem of data integration and annotation [9], i.e. if every research team has their own ontology for description of their data, the problems of ambiguity and heterogeneity of data remain. To counteract this process two different approaches have been proposed. The Open Biomedical Ontologies (OBO) consortium was started in order to guarantee a coordinated and consistent development of various ontologies that would ease data integration. An initiative by this consortium – OBO Foundry [9] – aims at providing and maintaining an evolving set of shared guidelines and principles for ontology development (<http://www.obofoundry.org/>). A different approach is based on The Unified Medical Language System (UMLS) – a system for merging and mapping vocabulary from different sources.

The overwhelming amounts of data resulting from the progress in biomedical sciences and the development of the information society which aims at digitization of most important aspects of life, especially health, induce the use of ontologies in biological research and healthcare systems. The aim of this work is to bring closer the field of ontologies to the medical society. First, theoretical basics of ontologies are shown (their foundations, classification and good practice design rules). This is followed by a survey of a selection of biomedical ontologies. Then, we present the UMLS and the OBO Foundry initiative, which are two conceptually different approaches to

solve the most current problem of ontology development – uncontrolled ontology proliferation. The last section of this work features examples of biomedical applications.

2. Theoretical foundations

Ontologies are constructs that appeared in Computer Science at the end of the 20th century [10]. Their theoretical foundations have their origins in Formal Ontology – a branch of philosophy that studies the classification of entities, their properties and general concepts that can be used to model them. The most often cited definition of an ontology was given by Gruber [11] and then extended by Borst [12]. An ontology is an explicit, formal specification of a shared conceptualization. The key elements of the definition are (1) the conceptualization and (2) formal specification.

2.1. Conceptualization

Although formal definitions of conceptualization differ at some points, they are consistent in general. Conceptualization is a simplified, generalized view of an observed universe or part of the universe [11]. According to Genesereth and Nilsson [13] a conceptualization is a tuple (D, R) , where D is the universe of discourse, which is a set of elements of the universe, and R is a set of relations between the elements. Every relation is a set of single elements or tuples which reflect a specific state of the world.

In order to better explain those concepts we will use an example of a small system of interacting proteins. Let us imagine a universe where only a few proteins and chemical compounds exist. This will be the universe of discourse $D = \{\text{ProteinA}, \text{ProteinB}, \text{ProteinC}, \text{ProteinD}, \text{CH}_4, \text{PO}_4^{3-}\}$. Particular chemicals are elements of the universe. What can be observed in this small universe is the following – if a combination of elements is put into a flask, some other elements start to appear after some time – a reaction takes place. For instance ProteinA and ProteinB together with CH_4 , produce ProteinC and ProteinB (Fig. 1).

Then according to Genesereth and Nilsson [13], a conceptualization for the world presented in Fig. 1 can be defined with $\{D, R\}$ shown in Table 1.

Concepts in the presented conceptualization are associated with the following natural language definitions:

- *Molecules* are all elements of the universe D .
- An *Enzyme* is a *Molecule* that allows a reaction but remains in the flask after the reaction takes place. Here ProteinB is an *Enzyme*.
- A *Substrate* is a *Molecule* which takes part in the reaction and disappears. ProteinA and CH_4 are *Substrates*.
- A *Product* is a *Molecule* which was not put into the flask but appears in it after the reaction takes place.
- *Interact-with* is a relation between two *Molecules* (a binary relation). Some physico-chemical forces acting between two *Molecules*, lead to a certain result.

We singled out the elements and phenomena in the universe D and defined concepts that can be used to describe

Download English Version:

<https://daneshyari.com/en/article/5198>

Download Persian Version:

<https://daneshyari.com/article/5198>

[Daneshyari.com](https://daneshyari.com)