Realism-Based Ontology for Integrating Individually Compiled Biomedical Data Repositories

Werner CEUSTERS [a,1]

[a] *New York State Center of Excellence in Bioinformatics & Life Sciences, Buffalo, USA*

Abstract.. Ontologies, when designed appropriately, are ideal candidates to make individually compiled data collections comparable and fit for semantic integration. Whereas semantic technologies can check whether formal representations produce interpretations that are outside the intended model, they do not provide much help to build such a model. This is where Ontological Realism comes into play. In this tutorial, using clinical research data collections about pain and its impact on mental health as an example, attendees will learn how to use formal methods and Ontological Realism as reliable tools to build ontologies that are faithful to reality and allow separate data collections to be semantically integrated.

Keywords. formal representation, Ontological Realism, clinical research, database integration

Rationale

The goal of the OPMQoL-project funded by the National Institutes of Health (NIH) is to obtain better insight into the complexity of pain disorders, specifically concerning the assessment of different pain types in the orofacial region, as well as into pain-related disablement and its association with mental health and quality of life [1].

Five existing data collections compiled independently from each other in respectively the US, Germany, Sweden, the UK and Israël, and covering in total 2000 patients, have been made available for this study. The data collections cover the same domain, but are distinct in many respects.

In this tutorial, the OPMQoL project will be used as a concrete example of the sorts of analyses and developments that are required to integrate individually compiled data repositories by means of a realism-based ontology.

1. Target audience

The tutorial is suited for a wide range of participants with diverse backgrounds. Clinical researchers will acquire knowledge in how to compile and document de novo data collections that can be more reliably compared with similar resources, or learn

---

[1] Corresponding Author. Werner Ceusters. Ontology Research Group, New York State Center of Excellence in Bioinformatics & Life Sciences, and Department of Psychiatry, University at Buffalo, 701 Ellicott street, Buffalo NY 14204, USA; E-mail: ceusters@buffalo.edu.

how to make existing collections more suitable for that goal. Developers of semantic technologies, including terminologies and ontologies, will obtain valuable insight in how to apply the principles underlying Ontological Realism [2-3] as a methodology to avoid mistakes that cannot be detected by logical formalisms alone [4]. Users of eHealth applications in general, and clinicians in particular, specifically if they wish to function as champions ('super users') in EHR or data warehouse customization efforts will learn the requirements such technologies must adhere to in order to make optimal use of ontologies [5].

2. Prerequisite knowledge

This tutorial requires of the attendees a thorough background in either any of the healthcare professions (medicine, nursing, …) or informatics (computer science). Additional familiarity with healthcare informatics is advisable. No background in semantic technologies or ontology development is however necessary. It will be easier to follow the tutorial if attendees have read the references [2, 4-5]. Those who wish to come extremely well prepared will benefit from reading the other references as well.

3. Educational goal

After the tutorial, the audience will be able to (1) understand better the added value of the Ontological Realism principles over mere computational and logical frameworks [6], (2) apply the principles to build or evaluate ontologies, (3) assess how to optimally use such ontologies in eHealth Technologies such as EHR systems, clinical research systems and data warehouses and (4) make recommendations to clinicians and biomedical informaticists to improve the systems they are working with.

4. Contribution of each teacher

4.1. Teacher's credentials

The tutorial will entirely be given by Werner Ceusters, Director of the Ontology Research Group and of the National Center for Ontological Research, University at Buffalo. Ceusters has made prominent contributions to both theoretical and applied research in ontology over more than 15 years, including international standards efforts.

4.2. Tutorial format and layout

The tutorial will be given in an interactive fashion. Within in each topic outlined below, explanation of principles will be followed by an invitation towards the audience to solve easy to intermediate problems with immediate discussion thereof.

4.2.1. The problems of terminology-based ontologies

The semiotic/semantic triangle and the associated ambiguous notion of 'concept' underlying terminology-based ontologies render these artifacts inappropriate for

representing that part of reality which is denoted by its terms. It will be explained why and a classification of typical mistakes will be provided [4, 6].

4.2.2. Principles of Ontological Realism

A detailed account will be given of the realist methodology which is based on the idea that the most effective way to ensure mutual consistency of ontologies over time and to ensure that ontologies are maintained in such a way as to keep pace with advances in empirical research is to view ontologies as representations of the reality that is described by science [2].

4.2.3. From model-theoretic languages to reference-based languages

Model-theoretic semantics gives the truth conditions for a sentence: a model satisfies a logical sentence if and only if the sentence evaluates to true in the given model. However, the generic terms used therein to denote specific entities convey that some specific entity is denoted, but not enough to be clear about which one in particular. Referent Tracking is an alternative based on direct reference [5].

4.2.4. OBO Foundry Ontologies

An overview of relevant OBO Foundry ontologies, including the Information Artifact Ontology and the Ontology for General Medical Science will be given, thereby indicating how they relate to the principles of Ontological Realism.

4.2.5. Building an ontology for integrating clinical datasets about orofacial pain

One specific aim of the OPMQoL project is to make data collections comparable by building a realism-based reference ontology for pain-related disablement, mental health and quality of life (OPMQoL) following the principles of Ontological Realism [2-3]. It will be explained how this reference ontology is developed on how it relates to relevant terminologies, assessment instruments and data collections.

References

[1] Smith B, Ceusters W, Goldberg LJ, Ohrbach R. Towards an Ontology of Pain. In: Okada M, editor. Proceedings of the Conference on Logic and Ontology. Tokyo: Keio University Press; 2011;:23-32.

[2] Smith B, Ceusters W. Ontological Realism as a Methodology for Coordinated Evolution of Scientific Ontologies. Applied Ontology. 2010;5(3-4):139-88.

[3] Bittner T, Smith, B. . Normalizing Medical Ontologies Using Basic Formal Ontology. Tagungsband der 49 Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie. 2004.

[4] 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, editor. Ontologies in Medicine Studies in Health Technology and Informatics. Amsterdam, The Netherlands: IOS Press; 2004;:145-64.

[5] Ceusters W, Smith B. Strategies for Referent Tracking in Electronic Health Records. Journal of Biomedical Informatics. 2006 June;39(3):362-78.

[6] Brochhausen M, Burgun-Parenthoine A, Ceusters W, Hasman A, Leong TY, Musen M, et al. Discussion of "Biomedical Ontologies: Toward Scientific Debate". Methods of Information in Medicine. 2011;50(3):217-36.