Article published online: 2022-12-04

Content Summaries of Selected Best Papers for the 2022 IMIA Yearbook, Section Knowledge Representation and Management

Vogt L

FAIR data representation in times of eScience: a comparison of instance-based and class-based semantic representations of empirical data using phenotype descriptions as example

J Biomed Semantics 2021 Nov 25;12(1):20

In this article, Vogt places his research in the context of big data and he argues the interest of ontologies to represent data and to satisfy the criteria of machine usability (i.e. eScience-compliant) and therefore findable, accessible, interoperable and reusable (FAIR). But this is not the main interest of the article. The article is extremely complete, recalling all the concepts necessary to understand it and unfolding very precise examples of biological phenotype (insects with their heads and antennas).

Vogt then develops a long discussion on the different approaches to knowledge modelling, on the computability of each of the chosen representation modes, on the interest of metadata, on the questions raised by the Open World Assumption and on the use of reasoners.

A long part of the article compares the two approaches: the TBox semantic graph approach based on classes, called Semantic Phenotype, and the ABox semantic graph approach based on instances, called Phenotype Knowledge Graph.

He concludes that, from a conceptual point of view, the advantages of the instance-based ABox semantic graph approach outweigh its shortcomings and outweigh the advantages of the class-based TBox semantic graph approach. Therefore, he recommends the instance-based ABox approach as a FAIR approach for documenting and communicating empirical data and metadata in a knowledge graph.

Further to this conclusion, this article specifies descriptions and arguments for both types of representation and, in passing, specifies the interest of semantic representations in general and their computability difficulty. It is a thought-provoking article.

Keet CM, Grütter R

Toward a systematic conflict resolution framework for ontologies

J Biomed Semantics 2021 Aug 9;12(1):15

In their article, the authors propose a framework to systematically handle modelling conflicts in the ontology development and (re)use processes.

The work aims to address this by taking steps towards a framework to deal with the various types of modelling conflicts through meaning negotiation and conflict resolution in a systematic way. It proposes an initial library of common conflicts, a conflict set, typical steps toward resolution, and the software availability and requirements needed for it.

The interest of the article is to explain precisely the problems encountered and how to solve them: Meaning Negotiation which concerns deliberations to figure out the precise semantics one wants to represent in the ontology, and Conflict Resolution which concerns choosing one option among a set of two or more options, where that choice is deemed the 'lesser among evils' for that scenario. It necessarily involves a compromise and making it work requires reengineering something in at least one of the ontologies or as a whole.

The authors use numerous examples to show the difficulties that can be encountered when reusing ontologies. In particular, they explore the issues that arise when top ontologies are reused to build a domain ontology underneath. They also tested the proposed framework against a realistic case of epizootic disease outbreak in the Lemanic Arc (France-Switzerland). In this last use-case, problems arise when one wants to use together an epidemiology ontology and an administrative ontology with different expressivity.

The authors discuss tool requirements for the walk-throughs and the article concludes with specific system requirements to meet in order to achieve integrated support that will help users. Like the preceding one, it is a thought-provoking article.

Wang P, Hu Y, Bai S, Zou S

Matching biomedical ontologies: Construction of matching clues and systematic evaluation of different combinations of matchers

JMIR Med Inform 2021 Aug 19;9(8):e28212

The paper by Wang *et al.*, studies biomedical ontology alignment algorithms and their combination with the aim of finding the best combination strategies with respect to performance but also scaling criteria.

The experimental results show that considering distinguishable matching clues (in the following four dimensions: terminology, structure, external knowledge, and representation learning) in biomedical ontologies leads to a substantial improvement of all available information. Besides, incorporating different types of matchers with reliability results in a marked improvement, which is comparable to the state-of-the-art methods. The highlights of the article are:

- The dominant matchers achieve F1 measures of 0.9271, 0.8218, and 0.5 on Anatomy, FMA-NCI Thesaurus, and FMA-SNOMED CT data sets, respectively. The results are very different depending on the ontologies. This raises questions about the influence of the size and structure of ontologies on the performance of matchers.
- The authors extend their method, Reduction Anchors, for dealing with large-scale ontology matching to improve the performance of matching large-scale biomedical ontologies. It achieves a significant reduction in time complexity with little loss of the F1 measure at the same time. Here again, the results are very different depending on the data sets.

In conclusion, this is a very complete article that systematically analyzes and compares the effectiveness of different matching clues, matchers, and combination strategies.

Harrison JE, Weber S, Jakob R, Chute CG

ICD-11: an international classification of diseases for the twenty-first century

BMC Med Inform Decis Mak 2021 Nov 9;21(Suppl 6):206

The article from Harrison et al. provides an overview of the recently completed 11th revision of the ICD which proposes an information framework comprising a semantic knowledge base (the Foundation), a biomedical ontology linked to the Foundation and classifications derived from the Foundation. ICD-11 for Mortality and Morbidity Statistics (ICD-11-MMS) is the primary derived classification and the main successor to ICD-10.

Two other classifications, the International Classification of Functioning, Disability and Health, known more commonly as ICF (classification of health and health-related domains) and the International Classification of Health Interventions (ICHI – classification for reporting and analysing health interventions for clinical and statistical purposes) are, in their new version, derived from the Foundation. Thus, this architecture makes it possible to propose a homogeneous and coherent set of classifications, in particular, ICD-11-MMS, ICF and ICHI (https://icd.who.int/dev11/f/en).

Regarding the importance of ICD-11-MMS for coding, WHO proposes new features compared to ICD-10:

- Clustering: ICD-11-MMS provides a general mechanism to allow codes to be combined to form clusters for use where expressive power is required beyond that provided by any single category. For example in surgery, ICD-11-MMS allows for coding of the disease for which surgery was undertaken (that would be the subject of one cluster) and of the complication. A code cluster on the latter can record the harm sustained (e.g., marked nausea and vomiting after surgery), the medication involved (perhaps a particular anesthetic agent), and how the problem came about (e.g., dose too high or too low, or administered at the wrong time);
- Coding tool. WHO created a web-based coding tool that employs partial word-matching, word-order independence, synonym management, hierarchy traversal, and more. It is important to note that users do not have access to the Foundation relationships, but the coding tool constrains the cluster representations in the same way as if they were using semantic relationships with domains and ranges.

Finally, the similarities between ICD-10 and ICD-11-MMS and the great expressive potential of the new revision are reasons to expect that ICD-11-MMS will provide a good basis for activity-based systems.

We do not have to take everything written in the article as reported here at face value, but the proposed architecture is really promising and should correct the main defects of the ICD-10.