Published online: 2020-08-21

Appendix: Summaries of Selected Best Papers for the 2020 IMIA Yearbook, Section Knowledge Representation and Management

Burek P, Scherf N, Herre H

Ontology patterns for the representation of quality changes of cells in time

J Biomed Semantics 2019;10(1):16

In this paper, the authors investigate the fundamental problem of modeling changes of quality occurring over time in biomedical ontologies specified in OWL. The paper is the result of the lessons learned during the development of an ontology for the annotation of cell tracking experiments. They present, discuss, and evaluate six representation patterns for specifying cell changes in time. In particular, they discuss two patterns of temporally changing information: n-ary relation reification and 4d fluents. They analyze the performance of each pattern with respect to standard criteria used in software engineering and data modeling, *i.e.*, simplicity, scalability, extensibility, and adequacy. There is no ideal solution and the patterns behave differently depending on the temporal distribution of the information modeled.

Finally, modeling quality value change is not limited to cell tracking experiments. It is a common and non-trivial task across many biomedical domains. The presented patterns are domain-independent. Since a change of quality values is common to many biomedical domains, this work has possible applications to represent the evolution of a patient's condition according to his/her treatments.

Denaxas S, Gonzalez-Izquierdo A, Direk K, Fitzpatrick NK Fatemifar G, Banerjee A, Dobson RJB, Howe LJ, Kuan V, Lumbers RT, Pasea L, Patel RS, Shah AD, Hingorani AD, Sudlow C, Hemingway H

UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER

J Am Med Inform Assoc 2019;26(12):1545-59

In this paper, the authors present the CALIBER EHR platform for developing, validating, and sharing reproducible phenotypes from national structured EHR in United Kingdom with applications to translational research.

They implemented a rule-based phenotyping framework with a systematic validation based on up to six different approaches. This framework was applied on 15 million individuals based on national EHR data collection from four data sources: UK primary care EHR data, hospital care billing data, disease registry data, and national death records data. Five standard terminologies were used to record information: a subset of the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT Read terms) for primary care clinical data, a derivative of the NHS Dictionary of Medicines and Devices (dm+d) for prescription codes, the International Classification of Diseases (9th and 10th revisions) for secondary care diagnoses and causes of mortality, and the Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures (4th revision) for hospital surgical procedures. The authors created algorithms for 51 diseases, syndromes, biomarkers, and lifestyle risk factors, using the CALIBER phenotyping framework. Three are detailed in the article: Heart failure, Myocardial infarction, and Bleeding. Validation evidence was established on cross-EHR source concordance, clinical note review, etiology, prognosis, genetic associations, and external population.

The open-access CALIBER data platform has been used by 40 national and international research groups in 60 peer-reviewed publications. This EHR-based phenomics approach within the CALIBER platform is an important step towards the international use of UK EHR data for health research.

Rector A, Schulz S, Rodrigues J-M, Chute CG, Solbrig H

On beyond Gruber: "Ontologies" in today's biomedical information systems and the limits of OWL

J Biomed Inform: X 2019 Jun 1;2:100002

In this paper, the authors start with the precise definition of the term "ontology", the experience of commonly constructed ontologies, and the development of ontological reasoning. Since the introduction of ontologies, open-world reasoning systems based on description logics have been developed, OWL has become a standard, and philosophical issues have been raised.

The article highlights what OWL-DL statements mean and lists all the potential pitfalls of OWL-DL representations with examples. It illustrates in particular that a reasoning process is not an ontological classification task (open vs. closed world), and that a paradigm shift is needed for reasoning. These developments are described from a historical perspective that shows the Information Technology (IT) context present at the time the ontologies were created.

The authors discuss the confusion that has emerged from the evolution of research in this field with two broad usages for the word "ontology" in the biomedical informatics literature: i) its original usage as a general term for the "background knowledge base"; and ii) as a term for some subset of the background knowledge base that is considered fundamental, on logical and/or philosophical grounds. They advocate for using the invariants as the denomination for traditional ontological statements, and for developing symbolic representation systems as hybrid model combining these invariants and reasoning capabilities based on frames.

In this framework, the authors also state that the International Classification of Diseases, 11th version (ICD-11) is philosophically close to SNOMED CT with an hybrid architecture: an ontology at the base and transformation/linearization in classification for certain uses (e.g. coding).

This paper was rated as a best paper because it is a very interesting and well-documented article that encourages us to rethink about knowledge modeling and the future systems that we will be able to build.

Shen F, Zhao Y, Wang L, Mojarad MR, Wang Y, Liu S, Liu H

Rare disease knowledge enrichment through a data-driven approach

BMC Med Inform Decis Mak 2019;19(1):32

In this paper, the authors applied a data-driven approach to enrich existing rare disease resources by mining phenotype-disease associations from a 5-year collection of 12.8 million clinical notes from electronic medical records (EMRs) at the Mayo Clinic.

They used association rule mining algorithms on EMRs to extract significant phenotype-disease associations and enriched existing rare disease resources: Human Phenotype Ontology (HPO) and Orphanet. They generated three phenotype-disease bipartite graphs: the HPO-Orphanet graph, the EMR graph, and the enriched knowledge base HPO-Orphanet + graph. They conducted a case study on Hodgkin lymphoma to compare performance on differential diagnosis among these three graphs.

The disease-disease similarity generated by the eRAM (an existing encyclopedia of rare disease annotations mined from 10 million scientific publications and medical records) was used as gold standard to compare the three graphs with a sensitivity and specificity of 0.17, 0.36, 0.46 and 0.52, 0.47, 0.51, for the three graphs, respectively.

They also compared the top 15 diseases generated by the enriched knowledge HPO-Orphanet + graph with eRAM and another clinical diagnostic tool, the Phenomizer. The proposed approach was able to significantly enrich existing rare disease knowledge resources with phenotype-disease associations from EMRs. This work provides a solution for differential diagnosis across rare and non-rare diseases.