

Report on Dagstuhl Seminar Towards Interoperability of Biomedical Ontologies

27.03.07 - 30.03.07, Seminar 07132

Barry Smith, IFOMIS, Universität des Saarlandes, Saarbrücken & Department of Philosophy, University at Buffalo

Mark A. Musen, Stanford Biomedical Informatics Research, Stanford University

Michael Schroeder, Professor in Bioinformatics, Biotec and Dept. of Computing, TU Dresden

The meeting focused on uses of ontologies, with a special focus on spatial ontologies, in addressing the ever increasing needs faced by biology and medicine to cope with ever expanding quantities of data. To provide effective solutions computers need to integrate data deriving from myriad heterogeneous sources by bringing the data together within a single framework. The meeting brought together leaders in the field of what are called ‘top-level ontologies’ to address this issue, and to establish strategies among leaders in the field of biomedical ontology for the creation of interoperable biomedical ontologies which will serve the goal of useful data integration.

One result of the meeting was to bring together representatives of different communities involved, especially the OWL/Semantic Web community, the qualitative spatial reasoning community, and the OBO Foundry and model organism database communities who have been associated with the Gene Ontology. The latter is, after 10 years, the most successful biomedical ontology when measured in terms of users, of associated tools, and of literature citations. Another result of the meeting was that, by contributing to the breaking down of barriers between these various communities, it gave significant impetus to the development and acceptance of new software tools which allow conversion of the OBO and OWL ontology formats. All OBO ontology content, including several hundreds of thousands of Gene Ontology Annotations, are now, as a result of this work, available on the Semantic Web in OWL format. The resulting potential for new kinds of exploitation of this data was outlined at the meeting by Alan Ruttenberg on behalf of the Semantic Web Health Care and Life Sciences Interest Group, who described new ways of integrating bio-ontology data, spatial data (for example pertaining to brain regions) with clinical data in neurology. Abstracts of talks presented at the meeting are appended below. They give some idea of the scope of the work presented.

Participants

1. Participants with presentations

Brandon Bennett, University of Leeds

Spatial Reasoning in Bio-Ontologies

The presentation reviewed the most promising methods for computing inferences involving spatial relations:

- first order logics (hard)
- compositional reasoning (good)
- concrete domains (promising)

- modal encodings (good)
- description logic encoding (promising)

The method of grounding ontologies in terms of spatial and material properties was introduced, and a specific methodology proposed. The issue of vague spatial boundaries (of for example biological organs) was raised. A method for solving this was presented, based on geometrical analysis of shape and the use of thresholding parameters.

Thomas Bittner, SUNY - Buffalo

The qualitative and time-dependent character of spatial relations in biomedical ontologies

The formal representation of mereological aspects of canonical anatomy (parthood relations) is relatively well understood. The formal representation of other aspects of canonical anatomy, such as connectedness and adjacency relations between anatomical parts, their shape and size as well as the spatial arrangement of anatomical parts within larger anatomical structures are, however, much less well understood and represented in existing computational anatomical and bio-medical ontologies only insufficiently. We outlined a methodology for incorporating this kind of information into anatomical and bio-medical ontologies by applying techniques of representing qualitative spatial information from Artificial Intelligence. In particular, we focus on how to take into account explicitly of the qualitative and time-dependent character of these relations. As a running example, we use the human temporomandibular joint (TMJ).

Elena Beisswanger, Universität Jena

BioTop and Chemistry Ontology

We describe the progress of our work on BioTop – a top-level ontology for linking biomedical domain ontologies. We start by making the case for the application of a common ontology to interface independent biomedical domain ontologies by introducing a set of more general classes. Then we briefly depict the relation of BioTop to the GENIA ontology as starting point of its initial development. Afterwards we propose our distinction of ontologies into top, top-domain and domain ontologies, and describe our approach to the integration of the top ontology BFO into BioTop. We conclude by detailing on some planned BioTop usages in the area of chemistry ontology.

Judith Blake, The Jackson Laboratory - Bar Harbor

Alignment of Ontologies for Biological Research

The development and use of bio-ontologies and controlled vocabularies for biology has exploded over the last 10 years in parallel with the use of high-throughput experimental procedures. The complexities of data integration necessitated the introduction of bio-ontologies to aid in robust data representation, recovery, and analysis. The Mouse Genome Informatics (MGI) system includes multiple controlled vocabularies and ontologies in its resource. Here I discuss the development of ontological views centered on specific diseases that provided cross-reference to core ontologies such as Mouse Adult Anatomy and GO Molecular Function. These views are utilized to co-curate mouse models of human disease and human clinical data. This work starts to address critical issues of interoperability of biomedical ontologies such as refinement of the Relation

Ontology, investigation of the intersection of large datasets, and methodology for providing project-specific ontological views.

Albert Burger, Heriot-Watt-University Edinburgh

Edinburgh Mouse Atlas: On the Interoperability Road

Anatomy ontologies play an increasingly important role in developing integrated bioinformatics applications. One of the primary relationships between anatomical tissues represented in such ontologies is part-of. As there are a number of ways to divide up the anatomical structure of an organism, each may be represented by more than one valid parthood (part-of) hierarchy. This raises the issue of how to represent and integrate multiple such hierarchies in order to ensure interoperability. We described one solution to this problem that is based on our work on an anatomy ontology for mouse embryo development as part of the Edinburgh Mouse Atlas Project (EMAP). We described the basic conceptual aspects of our approach and discussed strengths and limitations of the proposed solution.

Anthony G. Cohn, University of Leeds

Qualitative Spatial and Spatio-temporal Representations

I surveyed the field of qualitative spatial and spatio-temporal representations. I talked about spatial ontology – what kind of entities should be represented, mereology, mereotopology, orientation, shape and spatial change. I talked about how many relations can be defined from rather few primitives. I also illustrated some of the ideas with biological examples such as phagocytosis.

Karen Eilbeck, Lawrence National Laboratory, Berkeley

SO and Biological Sequence. Towards interoperability of biomedical ontologies

The Sequence Ontology (SO) is a structured controlled vocabulary for the parts of a genomic annotation. SO provides a common set of terms and definitions that will facilitate the exchange, analysis and management of genomic data. We showed how SO's consistent treatment of part-whole relationships can enable data described with its aid to become substrates for automated reasoning in virtue of the fact that instances of sequence features described by the SO can be subjected to the standard logical operations of extensional mereology. SO then serves interoperability of biomedical data by exploiting relations between underlying gene sequences used in common.

Robert Hoehndorf, Universität Leipzig

Interoperability, non-monotonicity and core ontologies

A large number of biomedical ontologies exist, covering diverse domains. Their integration into a common ontological framework is a difficult, yet rewarding task. In particular, there are ontologies that describe a canonical view of a domain. The Foundational Model of Anatomy (FMA) and many other ontologies. On the other hand, some other ontologies represent pathologies or exceptions to the canonical cases. Integration of these two kinds of ontologies is challenging because inconsistencies frequently arise, and must be avoided. We show where non-monotonic phenomena occur in the integration of the two types of ontologies we discuss, and propose a solution to the problem of their integration.

Waclaw Kusnierczyk, Univ. of Science & Technology - Trondheim

Taxonomic Partitioning of the Gene Ontology

The Gene Ontology (GO) project is intended as a species neutral controlled vocabulary for the annotation of (data describing) gene products. In some cases, only a generic or a species-specific subset of GO terms is required to annotate and analyze the results of a given biomedical experiment. We showed how one can define explicit links between terms in the GO and terms in a Taxonomy of Species (TS) in order to partition the GO automatically in light of a range of taxonomic criteria. We described some advantages of this new approach to the creation of species-specific GO slims are, including more adequate semantics of GO annotations, significant reduction of the effort needed to select GO terms appropriate for a particular taxonomic context, and ability to generate views of the GO even for taxa for which no explicit links with GO terms exist.

Patrick Lambrix, Linköping University

Ontology Alignment

The Semantic Web for life sciences shows promise for allowing effective and efficient querying of information needed in daily biomedical research. One of the key technologies for achieving this end is ontologies, which furnish the semantics of the Semantic Web. Since it is unlikely that there will be one single set of standard ontologies to which everyone will conform applications often need to deal with multiple overlapping ontologies whose heterogeneity then hampers interoperability. We addressed the technology of aligning ontologies which means: identifying relationships between different ontologies, as a means to overcome this problem. A number of alignment systems have been developed, in which various techniques have been proposed to facilitate identification of alignments. We also surveyed a range of issues which still need to be addressed to address the alignment problems we face.

Suzanna Lewis, Lawrence National Lab. - Berkeley

The Open Biomedical Ontologies Foundry

The OBO Foundry is based on the precept that ontology development is a cooperative enterprise. By doing so we can ensure that their classification strategies complement and augment one another. OBO is an umbrella organization within the broad domain of biology and medicine to produce well-structured ontologies. Those involved agree in advance to the adaptation of a growing set of principles specifying best practices in ontology development. These will ensure that the ontologies thereby improve in quality and formal rigor over time. This also assures that, for any particular domain, there is convergence by the community on the use of a – single – reference ontology. The prime principle of the OBO Foundry is that the ontology is open and available to be used by all without constraint.

Fabian Neuhaus, SUNY - Buffalo

CARO: Common Anatomy Reference Ontology

We described the project of creating a Common Anatomy Reference Ontology (CARO), whose goal is to facilitate interoperability between existing anatomy ontologies for different species and to provide a template for building new anatomy ontologies. CARO employs a structural axis of classification based on the top-level nodes of the Foundational Model of Anatomy. To provide coverage for developmental issues, CARO

will need to complement the developmental process sub-ontology of the GO Biological Process ontology. This in turn should provide a common framework for the model organism communities to classify developmental structures. CARO will support the coordination of cross-species ontologies at all levels of anatomical granularity from cell to whole organism, and will support the comparability of data gained through research on model organism with data gained through studies of human disease.

Alan L. Rector, Manchester University

The GALEN Experience and Simple Top Bio

The approach to ontology development behind GALEN and Pen&Pad is bottom-up development focused on the inferences that can be drawn from the ontology. Simple Top Bio is an exemplar teaching ontology of examples from the top domain demonstrating these principles in modern OWL notation. The guiding principles are deferring of commitment until evidence is available and that any distinction should make a difference by being in the domain or range of a high level property. An overview of the ontology was presented with examples.

Alan Ruttenberg, Science Commons

Living with BFO or Adventures on the Way to OBI. Some Puzzles, Some Solutions

The talk described two aspects of working on OBI, the Ontology for Biomedical Investigations. The “what is” game – going from a word loosely indicating some concept to identification of the appropriate classes in the ontology, leading towards success of BFO in assigning e.g. Specimen, Reagent as Roles and agreement in OBI about this assignment. I also discussed the connection and synergies between BFO and OWL in the context of multiple inheritance. The second part of the talk presented, to the best of my ability, a representation of a protocol following BFO. Introduced for discussion were protocols as plans, triggers for conditions, temporal future for planned times.

Susanna-Assunta Sansone, EBI - Cambridge

Working Towards Common Naming Conventions for use in Controlled Vocabularies and Ontology Engineering

A wide variety of ontologies relevant to the biological and medical domains are available through the OBO Foundry portal, and the number of such artifacts is growing rapidly. Integration of these ontologies, while requiring considerable effort, is extremely desirable. However, heterogeneities in format and style pose serious obstacles to such integration. In particular, inconsistencies in naming conventions can impair the readability and navigability of ontology class hierarchies, and hinder their alignment and integration. While other sources of diversity are tremendously complex and challenging, we argued that laying down an agreed set of common naming conventions is an achievable goal, particularly if those conventions are based on lessons drawn from pooled practical experience and surveys of community opinion. We outlined a proposed initial set of such conventions and described how they can help developers avoid flaws and inaccuracies when editing, and especially when interlinking, ontologies, thereby supporting ontology interoperability.

Michael Schroeder, TU Dresden

Towards Interoperability of Bio-Ontologies; Statistics vs Logic

Defining terms in ontologies is a difficult problem. The design and maintenance of ontologies should be linked to the literature underpinning the data to be annotated. Ontologies and text can be linked via textmining. While the extraction of terms from text is hard, statistical methods have been very successful with word sense disambiguation and gene identification achieving 80% success rates. Textmining can contribute to building ontologies by generating candidate terms, definitions and relationships. More importantly, textmining can show which parts of the ontology are used and which not.

On the other hand, meaning can be defined with logic. The problems of representation and reasoning arise. Description logics are widely used, but other logics are important too. Negative results are neglected, but useful. To represent them, explicit and implicit negation are needed. These give rise to non-monotonicity, which is useful to capture that some statements in the ontology are 'normally' the case.

Ultimately, logic and textmining should be used to allow authors of papers to express the main results in a controlled natural language, which is internally formally represented, lending itself to novel reasoning tasks.

Barry Smith

Biological Domain Ontologies and Basic Formal Ontology

The talk addressed certain problems with Simple Top Bio and other ambitious upper biological domain ontologies, and attempted to show how Basic Formal Ontology (BFO) resolves these problems. The challenges we currently face in integrating lower-level biological domain ontologies are increasing in severity. To address these problems a maximally efficient way requires an intuitive, upper level integrating ontology which does not overlap in its domain with the domains properly addressed by biological specialists. Thus an upper level integrating ontology should not, for example, contain its own small version of anatomy or biochemistry. The BFO ontology is precisely designed to address this need, and is being used to this end by the OBO Foundry and other ontology developers.

2. Other participants

Michael Ashburner, Cambridge University

Lindsay G. Cowell, Duke University, Durham, NC

Esra Erdem, Sabanci University, Istanbul

Carole Goble, Manchester University

Ian Horrocks, Manchester University

Ingvar Johansson, Universität des Saarlandes

Janet Kelso, MPI for Evolutionary Anthropology, Leipzig

Patty Kostkova, City University London

Holger Lewen, Universität Karlsruhe

Suzanna Lewis, Lawrence Berkeley National Lab

Yves Lussier, University of Chicago

Chris Mungall, Lawrence National Lab. - Berkeley

Dietrich Rebholz-Schuhmann, EBI - Cambridge

Philippe Rocca-Serra, EBI - Cambridge

Nigam Shah, Stanford University

Holger Stenzhorn, Universität des Saarlandes

Robert Stevens, Manchester University