An Ontology for Gastrointestinal Endoscopy

Shahim Essaid

Oregon Health & Science University, Portland, OR, USA

Abstract. The field of gastrointestinal endoscopy can benefit from an ontology for the purposes of data coding and data integration. This paper presents early results of an effort to develop such an ontology based on the OBO Foundry principles and existing OBO Foundry ontologies. Initially, the ontology will be limited to representing entities and relations currently implicit or hard-coded in the user interface of an existing endoscopy reporting system. The ontology will also be mapped to an existing database to evaluate the feasibility of ontology-driven queries. The long-term goal is to evolve this ontology to an application-independent terminology and information model for the domain of gastrointestinal endoscopy.

Keywords: gastrointestinal endoscopy, knowledge representation, data integration

1 Introduction

The practice of gastrointestinal endoscopy produces a significant amount of structured data that is captured in endoscopy reports. To encourage consistency in data collection, the World Endoscopy Organization maintains the Minimal Standard Terminology for gastrointestinal endoscopy (MST) [1]. The MST specifies a minimal set of terms and data structures needed to encode the majority of endoscopy data. However, the domain knowledge represented in the MST (in the form of terms, relations, and data structures) is not in a computable format and it could benefit from an ontological and logical analysis, and reorganization.

Another important effort in the field of gastrointestinal endoscopy is the Clinical Outcomes Research Initiative (CORI) [2]. CORI was established to assess the utilization and effectiveness of endoscopy procedures in clinical care. To meet its goals, CORI has developed an endoscopy reporting software and a central data repository for endoscopy reports. The reporting software is being used nationwide and the data repository currently receives over 250,000 reports annually. The data repository is primarily used for research purposes and to report on practice patterns and clinical outcome measures.

The endoscopy reporting software developed by CORI was initially based on the MST's representation of endoscopy data but it evolved to include additional terminology and data elements. Also, efforts are in place to integrate date generated by commercial reporting systems into the CORI data repository. These efforts have highlighted the need for a shared, stable, and computable terminology and information model for the field of gastrointestinal endoscopy to facilitate data integration while maintaining clear and consistent semantics.

2 Motivation and Planned Development

Recent advances in the area of biomedical ontology [3-4] can provide a foundation for a more formal and logical representation of entities and data elements needed to represent endoscopy data. Also, the existence of standardized knowledge representation languages and related inferencing capabilities can enable sophisticated querying of logically represented data and knowledge [5]. These advances have motivated an effort within the CORI project to develop an ontology for the field of gastrointestinal endoscopy.

The ontology will follow the Open Biomedical Ontology (OBO) Foundry development principles [6] and reuse entities from existing OBO ontologies when appropriate. The BFO will serve as an upper level ontology and other ontologies (IAO,

ogms, OBI, etc.) will be examined for middle level entities. Domain level entities will reference existing ontologies of anatomy, pathology, phenotypes, relations, and others when available. The ontology development project is hosted as a Google Code project [7].

3 Methods and Expected Difficulties

Development will start by identifying domain level terms and data elements hard-coded in the user interface of the existing CORI reporting software. These entities will initially form the main content of the ontology. This will decouple the domain knowledge from the application and allow for a more flexible evolution of the terminology and information model of the reporting software while still maintaining ontological and formal knowledge representation principles. The ontology will then be augmented by other entities from the MST, from free text entries in existing endoscopy reports, and from the endoscopy community. Also, as a proof of concept, the ontology will be mapped to the existing CORI data repository to evaluate the feasibility and benefit of ontology-driven queries compared to native SQL queries. The D2RQ Platform will be used for this part of the project [8].

A brief exploration of the user interface for the reporting software, and the MST, showed that difficult issues such epistemology vs. ontology, entities vs. statements, negation, and other related issues are frequent in clinical settings. Also, despite the relatively narrow focus of the practice of gastrointestinal endoscopy, endoscopy reports include information that ranges from current and past medical history, physical examination, visible endoscopy findings, and indirect findings through various imaging techniques. In addition to these various types of information, there is another epistemic layer that reflects the attitudes and judgment of clinicians in the form of assessments, diagnoses, plans, etc. To fully represent this information, an ontology will need a rich set of relationships that cover mereotopological, temporal, and modal relations, among others.

However, the primary use cases described below can be met by limiting our initial development efforts to an is_a hierarchy and a basic set of qualitative mereotoplogical relationships. The initial version will also be limited to representing endoscopic findings (polyps, ulcers, foreign bodies, etc.), their anatomical locations, and their clinical descriptions, according to the OBO Foundry ontology development principles.

4 Primary Use Case

The ontology will primarily serve as an interface terminology that supports data entry and enables consistent coding of endoscopy reports. The ontology will also be used to explore the value of ontology driven data retrieval by executing ontology driven queries against the current CORI dataset. These initial use cases can be met by a limited set of entities and relations and the remaining domain knowledge will be added as need arises.

References

1. Minimal Standard Terminology, http://www.worldendo.org/mst.html

2. Clinical Outcome Research Initiative, http://www.cori.org/

3. Bodenreider O, Stevens R. Bio-ontologies: current trends and future directions. Brief Bioinform. 7(3):256–274 (2006)

4. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat. Biotechnol. 25(11):1251-1255 (2007)

5. Stevens R, Aranguren ME, Wolstencroft K, Sattler U, Drummond N, Horridge M, et al. Using OWL to model biological knowledge. International Journal of Human-Computer Studies. 65(7):583-594 (2007)

6. OBO Foundry Principles, http://www.obofoundry.org/crit.shtml

7. Gastrointestinal Endoscopy Ontology, http://gieo.googlecode.com/

8. The D2RQ Platform – Treating Non-RDF Databases as Virtual RDF Graphs, http://www4.wiwiss.fu-berlin.de/bizer/d2rq/