ICBO: International Conference on Biomedical Ontologies · Buffalo, NY, USA
Working with Multiple Biomedical Ontologies Workshop, July 26, 2011

Multiple Ontologies in Healthcare Information Technology: Motivations and Recommendation for Ontology Mapping and Alignment

Colin Puri[1], Karthik Gomadam[1], Prateek Jain[2], Peter Z. Yeh[1], Kunal Verma[1]

[1]Accenture Technology Labs, San Jose, CA, USA
[2]Kno.e.sis Center, Wright State University, Dayton, OH, USA

Abstract. Electronic Health Records (EHR), Personal Health Records (PHR), data analysis and integration have emerged as key pieces in the delivery of quality health care. Integration of heterogeneous sources of patient information, domains of healthcare information, and associated ontologies brings about important questions. This paper enumerates upon some of the issues of ontology alignment, mapping, and motivations for the need of integration with respect to patient health care. No single ontology is sufficient to meet the growing needs of todays healthcare and the ontologies that exist today must themselves be integrated together for support of data integration and analysis. We also make a recommendation on one potential solution.

Keywords: Biomedical Ontologies, Ontology Mapping, Ontology Alignment, Healthcare Information Technology, BLOOMS

1 Introduction

As the healthcare industry moves towards wider adoption of Healthcare Information Technology (HIT) solutions such as Electronic Health Records (EHRs) and Personal Health Records (PHRs), data analysis and integration has emerged as a significant component in the delivery of quality healthcare services. For example, patient data can come in EMR systems, which capture treatments, symptoms, and diseases. It can also come from PHRs, such as Google Health[1] and Microsoft HealthVault[2], and other health and wellness applications, such as LiveStrong[3] and TrialX[4], that capture additional aspects of the patient's health such as lifestyle choices and diet. If these data sources are properly integrated, then we can begin to realize applications that will enable healthcare providers to effectively answer questions such as:

1. What treatments were administered to other patients with similar health conditions?
2. What was the efficacy of such treatments when administered to patients with a given physiological profile?
3. What medications are currently being prescribed to the patient and how do they constrain available treatment options?
4. How can one meaningfully find and utilize the vast amounts of medical knowledge, such as codified medical vocabularies, scientific publications, and findings from clinical trials, available in the public domain?
5. How can the health and wellness information stored by a patient in PHRs and other PHR-based applications be used to improve the quality of care?

Such applications can potentially save billions of dollars in healthcare costs [12], while improving the quality of care.

Biomedical ontologies provide a promising solution for integrating these heterogeneous data sources by providing a common vocabulary (and framework) to enable interoperability, resolve ambiguity, etc. However, no single ontology is sufficient. Instead, multiple ontologies must be combined in practice to fully

[1] http://health.google.com
[2] http://healthvault.microsoft.com
[3] http://livestrong.com
[4] http://trialx.com

realize meaningful integration and analysis of data in the healthcare domain. Hence, the ontologies themselves must first be integrated before they can support the necessary data integration and analysis.

This paper focuses on the technical challenges in integrating multiple ontologies, and takes the position that existing ontology alignment solutions can provide a viable solution to this end.

2 Ontology Mapping and Alignment: Motivation and Current Approaches

A patient's medical record captures multiple aspects of his/her health (e.g. medications, health conditions, prior treatments, etc) and can come from multiple sources (e.g. EMR systems, PHR applications, etc). Integrating this information into a coherent view requires combining multiple ontologies such as:

- SNOMED CT [16] is an systematic organization of medical terminology containing information related to medical conditions, procedures, pharmacueticals, etc.
- RxNorm [7] provides a vocabulary for normalized names for clinical drugs. It is intended to cover all prescription medication in the United States. It contains the active ingredients, strengths, and dose form comprising that drug.
- MeSH (Medical Subject Headings) [3] is a large and expansive controlled vocabulary for indexing medical journals, articles, and books.
- ICD-10 (International Statistical Classification of Diseases and Related Health Problems 10th Revision) [2] is a collection of codes specifying diseases, symptoms, findings, complaints, etc. as defined by the World Health Organization (WHO).
- Gene Ontology (GO) [4] is a unified ontology designed to represent the gene attributes across all species. Furthermore, the goal of this ontology effort is to develop a controlled vocabulary, annotate gene information, and provide a set of useful tools for access to the genetic information. The GO is a part of a larger initiative, the Open Biomedical Ontologies, to create controlled vocabularies for use between several biomedical domains.

A number of efforts such as UMLS [6], OpenGALEN [14], and 3M's Health-Care Data Dictionary [1] have tried to consolidate multiple ontologies. While these efforts do provide mappings between different biomedical ontologies, health-care providers still face several challenges when integrating their proprietary vocabulary and processes with third-party biomedical ontologies. These challenges range from syntactic differences (e.g. different terminologies, naming conventions, and formats) to deeper semantic differences (e.g. different granularity for modeling steps in a medical protocol). What is required are solutions that can generate these mappings either automatically or with minimal human effort.

Ontology mapping and alignment has been an active area of research. Various strategies, including machine learning, rule based mapping, and logic driven frameworks, have been adopted to address the challenge of ontology mapping. We briefly illustrate some of the research that have employed these techniques. Machine learning approaches have been used in Learning Source Description (LSD) [10]. LSD employs a multi-stage learning approach and exploits both the schema and the data. The Ontology Integration System (OIS) [8] adopts a query based approach and employs description logic based techniques. A hybrid approach, employing rules and learning is discussed in [11]. In addition to these techniques, ideas from the area of database schema matching have also been adopted in the context of ontology mapping. A survey of such approaches is presented in [15].

In general, ontology mapping can be classified into three categories [9]:

1. Global ontology view to local ontology view: An example of this would be the mappings between an ontology describing a provider's proprietary terminology and clinical pathways that use a view of SnoMed, with SnoMed.
2. Semantic mappings between local and target entities: An example would include

the mappings between an ontology for drug formulary and an ontology for clinical pathways, where a drug (a source entity) is mapped onto a medication (target entity). Once the mapping is done, the transformed entity captures the properties of the drug from the source ontology and the dosage information for a particular medical condition from the target ontology.

3. Mappings to enable ontology re-use by integration and alignment: An example would be the mappings to integrate multiple clinical pathway ontologies of similar chronic medical conditions that will help in identifying overlapping concepts and synonyms.

3 Recommendation

Approaches (described in the previous section) have shown promising results, but their application has been limited to a few public ontologies. There is one approach – i.e. BLOOMS [13] – which has been successfully applied to aligning disparate ontologies in the Linked Open Data Cloud [5] – a Web-scale effort to integrate vocabulary from diverse sources and providers, ranging from music ontology to consumer business categories. We recommend the consideration of BLOOMS as a possible engine for aligning disparate biomedical ontologies.

BLOOMS is a system for generating links between class hierarchies between two ontology schemas. In the context of the Linked Open Data Cloud (LODC), BLOOMS uses Wikipedia to create an initial set of categories for the given concepts and uses the comparison of the generated categories as a basis for link generation. Unlike many of the current approaches, BLOOMS uses the data resources available on the Web as a point of reference during the tasks of mapping and alignment. For example, BLOOMS uses Wikipedia to derive a category hierarchy for the concepts in the ontologies. The ability of the system to identify and leverage on non-traditional and open data sources makes it a flexible framework for alignment, while also reducing the dependency on the domain models. The latter advantage is significant and allows BLOOMS to deliver higher quality mappings.

References

1. 3M Healthcare Data Dictionary (3M HDD). http://www.3mtcs.com/products/ hdd.
2. International Classification of Diseases, 10th Revision. http://www.who.int/ classifications/icd/en/.
3. MeSH (Medical Subject Headings. http://www.ncbi.nlm.nih.gov/mesh.
4. The Gene Ontology. http://www.geneontology.org/.
5. C. Bizer, T. Heath, and T. Berners-Lee. Linked data-the story so far. *Int. J. Semantic Web Inf. Syst.*, 5(3):1–22, 2009.
6. O. Bodenreider. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research*, 32(suppl 1):D267, 2004.
7. O. Bodenreider. Visualization tools for the Unified Medical Language System (Sem-Nav), the Gene Ontology (GenNav), and RxNorm. In *Humans and the Semantic Web HCIL Workshop*, 2006.
8. D. Calvanese, G. De Giacomo, and M. Lenzerini. A framework for ontology integration. In *The Emerging Semantic WebSelected Papers from the First Semantic Web Working Symposium*, pages 201–214, 2002.
9. N. Choi, I. Song, and H. Han. A survey on ontology mapping. *ACM Sigmod Record*, 35(3):34–41, 2006.
10. A. Doan, P. Domingos, and A. Halevy. Learning to match the schemas of data sources: A multistrategy approach. *Machine Learning*, 50(3):279–301, 2003.
11. M. Ehrig, S. Staab, and Y. Sure. Bootstrapping ontology alignment methods with APFEL. *The Semantic Web–ISWC 2005*, pages 186–200, 2005.
12. R. Hillestad, J. Bigelow, A. Bower, F. Girosi, R. Meili, R. Scoville, and R. Taylor. Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Affairs*, 24(5):1103, 2005.
13. P. Jain, P. Hitzler, A. Sheth, K. Verma, and P. Yeh. Ontology alignment for linked open data. *The Semantic Web–ISWC 2010*, pages 402–417, 2010.
14. A. Rector, J. Rogers, P. Zanstra, and E. Van Der Haring. OpenGALEN: open source medical terminology and tools. American Medical Informatics Association, 2003.
15. P. Shvaiko and J. Euzenat. A survey of schema-based matching approaches. *Journal on Data Semantics IV*, pages 146–171, 2005.
16. K. Spackman, K. Campbell, et al. SNOMED CT: a reference terminology for health care. In *Proceedings of the AMIA annual fall symposium*, page 640. American Medical Informatics Association, 1997.