# A Meta-Data Approach to Querying Multiple Biomedical Ontologies

Ravi Palla<sup>1,2</sup>, Dan Tecuci<sup>1</sup>, Vinay Shet<sup>1</sup>, Mathaeus Dejori<sup>1</sup>

<sup>1</sup>Siemens Corporate Research, Princeton, NJ, USA <sup>2</sup>School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA

**Abstract.** We present an approach for retrieving information spread across multiple large ontologies, with the goal of developing a biomedical question answering system that can assist physicians in diagnosis, treatment and therapy planning. The approach involves ontology integration and run-time SPARQL query generation, both of which are accomplished by defining a meta-ontology containing information about the properties and structure of the individual ontologies. The approach enables ontology integration with minimal changes and also supports ontology interoperability. We built a prototype of our approach that integrates the Foundational Model of Anatomy ontology, the human disease ontology, and an ontology that represents certain information from the Merck manual.

### 1 Introduction

Biomedical ontologies are rich sources of information that can be shared and used for reasoning within question answering (QA) systems. To assist physicians in making the correct diagnosis and prescribing the right medication, a QA system needs to have sufficient access to information regarding anatomy, pathology, pharmacology, and other related domains. While there are ontologies that cater to each of these individual domains, to the best of our knowledge, there is no ontology that sufficiently covers all these domains. Moreover, the ontologies for the individual domains do not necessarily contain all the information required by QA systems to effectively assist physicians. Therefore, such systems need to both enhance and integrate several biomedical ontologies.

In this paper, we present an approach for retrieving information spread across multiple ontologies in the context of building a question answering system. It involves ontology integration and run-time SPARQL query generation, both of which are accomplished by defining a meta-ontology that contains information about the various properties in the ontologies, the mapping between the properties, and the information needed to generate SPARQL queries for retrieving information with respect to these properties.

The approach abstracts away the actual ontologies as it refers only to the meta-ontology to retrieve the required information. This implies that ontologies can be integrated into simply system by updating meta-ontology. The approach also allows for interoperability between ontologies at the level of ontology alignment [1]. While this is a weak form of integration, we found it to be appropriate for QA systems that rely on several large ontologies. We tested our approach by considering the Foundational Model of Anatomy (FMA) ontology [3], the human disease ontology<sup>2</sup>, and an ontology that represents certain information from the Merck manual<sup>3</sup>. The metadata we use is tailored for QA systems, but the approach itself can be used for other applications.

### 2 The Approach

We define a set of high-level properties based on the types of questions to be answered and create an upper ontology that maps these high-level properties to properties of individual ontologies.<sup>4</sup> The mapping between the properties of individual ontologies can be derived from

http://sig.biostr.washington.edu/projects/fm/

<sup>&</sup>lt;sup>2</sup> http://www.obofoundry.org/

<sup>3</sup> http://www.merckmanuals.com/professional/index.html

<sup>&</sup>lt;sup>4</sup> This mapping is not necessarily one-to-one and can be a many-to-many.

their mapping to the high-level properties. Consider the case of answering questions of the form, "What is a [concept-name]?" For this, we define a high-level property "definition" and then update the upper ontology by including the mapping between "definition" and the properties of the individual ontologies that provide an appropriate answer for a definitional question.

Given this, in order to retrieve the URIrefs all properties necessary to answer definitional questions, the system can simply query the upper ontology. However, just retrieving these URIrefs is not sufficient to answer the question and correct queries need to be formulated to retrieve the definitions. In order to formulate the queries, the knowledge of the structure of the underlying ontologies is required, information that can also be included in the upper ontology. The following RDF description shows a possible mapping from the high-level property "definition" to the property providing definitions in the disease ontology. The description also contains the necessary information to generate the SPARQL queries that can be used to retrieve the definitions.<sup>5</sup>

```
<rdf:Description
rdf:about="OBOINOWL#hasDefinition">
   <hasProperty>definition</hasProperty>
   <hasQueryTarget>def</hasQueryTarget>
   <hasQueryLine>?x. ?x rdfs:label
   ?def.</hasQueryLine>
</rdf:Description>
```

The description above indicates that the SPARQL queries needed to retrieve definitions from the disease ontology are of the form

```
SELECT ?def WHERE {
[subject] <OBOINOWL#hasDefinition> ?x.
?x rdfs:label ?def.
```

where "[subject]" is the URIref of the concept in the disease ontology whose definition has to be retrieved.

### 2.1 Interoperability: Handling Synonyms

The approach presented so far assumes that all the URIrefs corresponding to the concept names in the user's question have been identified. However, this is not trivial since

different ontologies might refer to a concept using different names, and the user can use any of these names in the question. For example, the Merck manual ontology contains the definition for "Atrioventricular block" and the user can ask the question, "What is a AV block?" The disease ontology contains the block" as a synonym for name "AV "Atrioventricular block" but does not contain the definition. So, if the system only gets the URIrefs corresponding to the name "AV block", it will not be able to answer the question. In order to answer the question, the system needs to retrieve the URIrefs corresponding to "Atrioventricular block" in the Merck manual ontology, and this can be done by first retrieving the synonyms of "AV block" from the disease ontology and then using them to obtain the corresponding URIrefs from the Merck manual ontology. In general, to answer any question about a concept, the system needs to first retrieve all the synonyms of the concept name used in the question and then use them to retrieve the corresponding URIrefs.

However, since different ontologies have different structures, querying for the synonyms is not straightforward. To address this, we define a high-level property "synonym" and use the upper ontology to represent information about querying for synonyms. The following description shows one way to represent information about retrieving synonyms from the disease ontology.

```
<rdf:Description
rdf:about="OBOINOWL#hasExactSynonym">
    <hasProperty>synonym</hasProperty>
    <hasQueryTarget>syn</hasQueryTarget>
    <hasQueryLine>?x. ?x rdfs:label
    ?syn.</hasQueryLine>
</rdf:Description>
```

In order to obtain the synonyms, the system can query the upper ontology to retrieve all the information required to formulate the SPARQL queries needed to retrieve the synonyms.

## 3 The QA System, High-Level Properties and their Mappings

We have implemented a prototype QA system that uses this approach to query multiple biomedical ontologies. The system answers questions of the form "What is a [concept-

Here, OBOINOWL is an abbreviation for "http://www.geneontology.org/formats/oboInOwl".

name]?" and "What is/are the [relationname(s)] of the [concept-name]?" With respect to such questions we defined high-level properties like "definition", "part", "location", "connections", and "affected organs", so that questions such as "What is the location of the heart?" and "What are the affected organs of atrial fibrillation?" can be asked.<sup>6</sup>

The table below shows some of the mappings for the high-level properties discussed above.

High-Level Property	Mapped To
definition	fma:definition, oboInOwl:hasDefinition, fma:location, merck:hasDefinition, fma:surrounded_by, rdfs:subClassOf
location	fma:location, fma:surrounded_by, fma:contained_in

The table suggests that in order to answer definitional questions of the form "What is a [organ]?", the system also retrieves information about the type of the organ and some information about the location of the organ. This is another advantage of our approach as it enables us to change the information that is retrieved by simply adding/deleting certain mappings.

### 4 Related Work and Conclusion

Building a comprehensive biomedical QA system requires some level of integration of multiple domain ontologies. There have been several approaches presented for integrating biomedical ontologies, like the Ontology of Biomedical Reality (OBR) framework [2] and the framework of the Open Biomedical Ontologies (OBO)<sup>7</sup> consortium that attempt to make the process of development of biomedical ontologies more formal, thereby allowing more interoperability between ontologies spanning several domains. Among other approaches, is

While our approach is still in its infancy, preliminary results show that it enables ontologies to be integrated into the QA system with minimal changes while supporting interoperability between ontologies with different structures as shown in section 2.1. While the interoperability supported is a weak form of integration, it seems appropriate for QA systems that need to consider several large ontologies of different structures.

We plan to extend the QA system to handle more question types, answering which would not only require retrieval of information but also deep reasoning so that the system can effectively assist physicians in their tasks.

### Acknowledgements

We are grateful to Michael Sintek and Patrick Ernst from the German Research Center for Artificial Intelligence (DFKI) for their contribution to this work, and to the anonymous referees for their valuable comments.

#### References

- Sowa, J.F.: Knowledge Representation: Logical, Philosophical and Computational Foundations. Brooks/Cole (2000).
- Rosse, C., Kumar, A., Mejino, J.L.V. Jr., Cook, D.L., Detwiler, L.T., and Smith, B.: A Strategy for Improving and Integrating Biomedical Ontologies. In Proceedings of AMIA Symp. (2005) 639–643.
- 3. Rosse, C., Mejino, J.L.V. Jr.: A reference ontology for biomedical informatics: the Foundational Model of Anatomy. J. Biomed. Inform. 36(6) (2003) 478–500.

the Linked Life Data<sup>8</sup> platform that has been used to integrate biomedical ontologies spanning multiple domains. While our approach also deals with ontology integration, albeit weak, it differs from the above approaches in that it is goal driven. We define a set of high-level properties of interest based on the kinds of questions we want the system to be capable of answering and do the integration with respect to these properties.

<sup>6</sup> The information regarding the affected organs is retrieved from an ontology obtained by a stronger integration of the FMA and the disease ontologies.

<sup>7</sup> http://www.obofoundry.org/

<sup>8</sup> http://linkedlifedata.com/