Undefined 1 (2014) 1–5
IOS Press

Ontology alignment for wearable devices and bioinformatics in professional health care

Editor(s): Name Surname, University, Country
Solicited review(s): Name Surname, University, Country
Open review(s): Name Surname, University, Country

Jack Hodges [a,*], Mareike Kritzler [a] and Florian Michahelles [a] and Stefan Lueder [a] and Erik Wilde [a]
[a] *NEC WOT-US, Siemens Corporate Technology, 2087 Addison, Berkeley, CA 94074, USA*
E-mail: {jack.hodges.ext, mareike.kritzler, florian.michahelles, stefan.lueder, erik.wilde.ext}@siemens.com

Abstract. Web Ontology Language (OWL) based models and triple stores hold great potential for access to structured information. Not only are OWL-based ontologies extremely versatile and extendable, but triple stores are robust against changes to ontologies and data. The biomedical field illustrates this value insomuch as it employs vast amounts of information distributed across different models and repositories. This paper presents a case study that sought to demonstrate the real-world value of linking disease, symptom, and anatomical models with wearable devices and physical property models and repositories. Integrating these models is both necessary and problematic; necessary to provide undifferentiated access to health care professionals, problematic because although the biomedical ontologies and repositories exist, they aren't semantically aligned and their designs make alignment difficult. This case study demonstrated that manually linking multiple biomedically-related models can produce a useful tool. It also demonstrated specific issues with aligning curated ontologies, specifically the need for compatible ontology design methodologies to ease the alignment. Although this study used manual ontology mapping, it is believed that systems can be developed that can work in tandem with subject matter experts to reduce mapping effort to verification and validity checking.

Keywords: ontology alignment, semantic bridge, professional health care, wearable devices, physical properties

1. Introduction

Consider a physician that has a patient diagnosed with Type II Diabetes. How might this physician locate wearable devices that could aid in tracking the patient's condition? Currently they would have to enumerate the symptoms associated with Diabetes, associate those with the kinds of measurements needed to qualify/quantify the symptom, and then search for wearable devices capable of measuring that quantity. Using a semantic integration it would be possible to do all this in a single application.

Achieving the promise of semantic technology involves a mixture of upper and domain ontologies, semantic repositories, and integration (linking and mapping) across ontologies. For ontologies such as QUDT (quantities, units, dimensions, and datatypes) [4] the related ontologies and repositories are fully integrated, meaning that they act as a single model. The vast majority of curated ontologies/repositories, however, remain stand alone. For example, the biomedical field has produced many ontologies; for diseases, biochemistry, symptoms, anatomy, pathology, etc. For the most part these ontologies are not integrated and thus the information they model cannot be shared across with each other.

This case study sought to demonstrate the value of semantic integration by showing how medical professionals could benefit by having integrated access to biomedical models/repositories that also integrate with wearable devices and the properties they measure. In the process of integrating the information sources that would allow the above example to be realized, both

*Corresponding author. E-mail: editorial@iospress.nl.

ontology and data mapping issues were encountered. This paper describes the mechanisms used to achieve this integration, manually, using a subset of five ontologies. The conclusion borne out by this study suggests that ontology mapping can be made easier with the use of computational techniques but still requires subject matter experts for validation and verification.

2. Goals

There were 4 primary goals in this study, all associated with the desire to show that integration of structured information sources would be useful for medical professionals:

- Demonstrate the integration of standardized and externally-curated models
- Demonstrate that linked biomedical models provide value through single-point access
- Demonstrate integration through the use of quantity/unit/value models
- Demonstrate the ability to associate wearable device models with biomedical models

The first goal saves us from having to take the time and energy to create models from scratch, but ultimately causes two problems inherent in any integration project; (1) mismatched design or implementation approaches, and (2) incomplete data. The remaining goals are illustrative of the functionality this study hoped to demonstrate; namely to bring together a set of ontologies that would connect wearable sensors and medical professionals.

The information types this study sought to connect are shown in Figure 1:

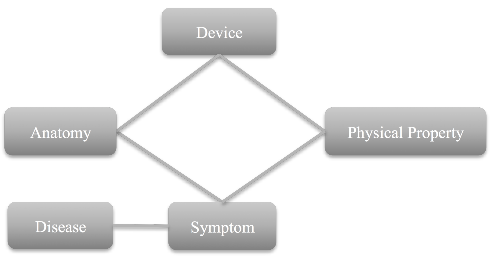

Fig. 1. Targeted integration of 5 information types

Figure 1 shows five information types. First, it was desirable to integrate biomedical information that would be useful in helping medical professionals evaluate/select wearable devices for patients: (1) human diseases, (2) disease symptoms, and (3) human gross anatomy. For the physician, being able to diagnose a patient, and relate the associated disease or symptoms with associated anatomy, would allow multiple access points to integrate with wearable devices.

Second, symptoms are associated with physical properties (e.g., cardiac disease and blood pressure), so integrating symptom models with properties that a wearable device might measure is necessary.

Finally, to integrate the wearable device to the biomedical models, both the anatomical parts where the device is worn and the properties that it measures are needed.

2.1. Choice of Ontologies

The choice of ontologies used in this study was based on the goal of demonstrating an integrated value. Three ontologies from the Open Biological and Biomedical Ontologies (OBO) Foundry ontologies were selected due to the content they modeled as well as the fact that they had unique identifiers that allowed them to be cross referenced. Disease information was represented using the OBO Disease Ontology (DOID) [12], symptom information was represented using the OBO Symptoms Ontology (SYMP) [13], and anatomical information was represented using the OBO Foundation Model of Anatomy (FMA) ontology [9][1] were used.

The QUDT[4] models were used because they represent an integrated approach to quantities, units, dimensions, and datatypes, but other models exist that might have been used, such as the Model Library for Quantities, Units, Dimensions, and Values (QUDV) [10], the Library for Quantity Kinds and Units (QU)[15]. QUDT was deemed to be more comprehensive set of models, and had representations for biomedical quantities needed to model wearable devices.

The Semantic Sensor Network (SSN) [7] ontology is an emerging standard for device modeling and seemed a good integration point for the project.

3. Approach

From a functional point of view the study would be successful if all 5 semantic models can be integrated into effectively a single model and searched from mul-

[1] All available through the Open Biological and Biomedical Ontologies Foundry

tiple entry points in a single interface. The integration required minimally 5 property alignments (and associated mappings) between the ontologies:

- Diseases (DOID) → symptoms (SYMP)
- Symptoms (SYMP) → anatomy (FMA)
- Symptoms (SYMP) → properties (QUDT)
- Wearable devices (SSN) → quantities (QUDT)
- Wearable devices (SSN) → anatomy (FMA)

These alignments are based on functional dependencies between the models. For example, there would be no need to align wearable devices with diseases or symptoms because there is no functional relationship between them. On the other hand, an alignment could be modeled between diseases and anatomy, but since there is a direct relationship between disease and symptom, and between symptom and anatomy, the relationship between disease and anatomy can be inferred and need not be explicitly modeled.

The ontological integrations used to support semantic search in this study are shown in Figure 2:

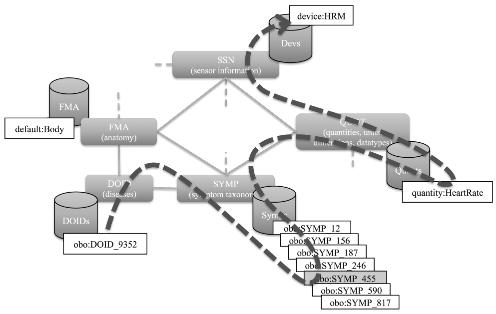

Fig. 2. Integration of 5 ontologies and sample graph traversal

Figure 2 illustrates the ontological alignments used in this study along with some values for a particular search task (in this case, Type II Diabetes). The large dotted line shows one path (disease → symptom → property → device) a search might take through these models to relate Type II Diabetes to a wearable device capable of measuring heart rate/blood pressure for the symptom abnormal weight gain. Solid lines between the repositories represent the ontological alignments. The data sets are depicted as repositories to illustrate the integrative/federated nature of the study. Square boxes represent items in the repositories along the depicted inference path.

3.1. Alignment Problems

There are three aspects to the model integration problem encountered in this study. First, the semantic structure aligning the models must exist. That is, there must be classes and class properties that support the alignment. Second, the data allowing for an integration must exist. Finally, given the requisite semantic alignment, a mapping of the data between the models must be exist or be constructed.

3.1.1. Semantic Structural Alignment

Model integration requires that appropriate classes and properties exist between the models or that new *mapping models* be created. In the models used in this study there were a number of semantic structural alignment issues. In each case, since the original models are curated by separate entities, new mapping models (aka semantic bridge ontologies [8,5] were created and then merged with the originals:

- DOID → SYMP: Semantic bridge and mapping required
- SYMP → FMA: Semantic bridge required
- SYMP → QUDT: Semantic bridge required
- Wearable Device: Subclass SSN SensingDevice
- SSN → QUDT: Semantic bridge required
- SSN → FMA: Semantic bridge required

The simplest scenario was between DOID and SYMP. The DOID ontology provided a property doid:has_symptom that might have been used to map the two ontologies, but it wasn't being used (no data mappings). The OWL version of the model also had no root Disease class so a new semantic bridge ontology was created with a Disease class. The original disease classes were subclassed to Disease, see Figure 3:

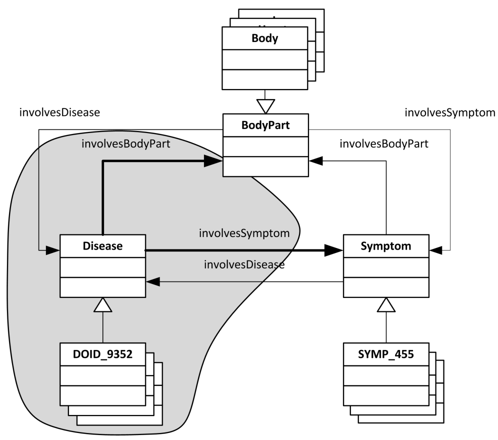

Fig. 3. A simple semantic bridge between DOID, SYMP, and FMA

The *involvesSymptom (involvesBodyPart)*, etc. properties were then created that link the Disease and

Symptom (and BodyPart, respectively) classes. Figure 3 shows the DOID→SYMP semantic bridge ontology highlighted in grey, with the property links associated with the bridge darkened.

Similar problems were evidenced in the SYMP and FMA ontologies. Though symptoms are generally associated with one or more anatomical entities (e.g., cardiac disease with heart), there was no alignment between these ontologies. Semantic bridges were also used to link the SYMP and FMA ontologies by creating the Symptom and BodyPart classes and related properties (also shown in Figure 3.

In the case of SSN, the sensing device model was very general so it was subclassed to support wearable devices using another semantic bridge, as shown in Figure 4.

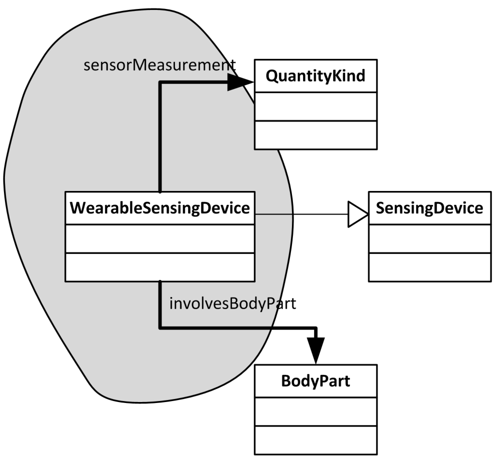

Fig. 4. WearableSensingDevice subclass from ssn:SensingDevice

The subclassing to WearableSensingDevice allowed a mapping from a device to the QUDT QuantityKind using the SSN sensorMeasurement property. Properties were added to support wear location, company, etc. Data was acquired for sample wearable devices by scraping the Vandrico [14] web site.

3.1.2. Model Data Mapping

With model-level alignment it becomes possible to map between model data sets. When the data sets are complete, semantic tools exist to perform the alignment/mapping. However, when the data is incomplete, the mapping is complicated, requiring some form of web content mining for unstructured text [3](semi-automated methods exist to mine structured content).

The complexity of unstructured text data mining is illustrated with the relationship between diseases modeled in DOID and symptoms modeled in SYMP. Seven steps had to be taken to manually map diseases in DOID to symptom data in SYMP. Similar steps would be required to map any two ontologies:

- Identify each disease in DOID (e.g., label)
- Search web data sources for the disease
- Search the found data sources for symptom references
- Isolate and normalize the symptoms found
- Compare symptoms to those in SYMP
- Disambiguate possible matches
- Construct new graph relationships in DOID and SYMP

The first and last steps in this list can be satisfied with SPARQL queries against the models. The intermediate steps required search, compare, and disambiguation capabilities, (some) subject matter expertise, as well as language and reasoning abilities.

For example, in performing a manual mapping, we can perform a web search for "Type II Diabetes symptoms". Each result may provide text blocks that describe symptoms. For example,

> Type 2 diabetes develops when the body becomes resistant to insulin or when the pancreas stops producing enough insulin. Exactly why this happens is unknown, although genetics and environmental factors, such as excess weight and inactivity, seem to be contributing factors.[2]

As natural language readers/understanders, and with enough knowledge about the source and target domains (i.e., human diseases and disease symptoms), we can decide which content to use and which to ignore, how much of the content selected is appropriate, etc. Ultimately we must boil the information in a symptomatic description to labels that can be compared with the SYMP ontology. For example, we might be able to get:

- insulin resistance
- excess weight
- inactivity

from the provided quotation. If lucky, one might find something easier to parse, such as:

- Excessive thirst and appetite
- Increased urination (sometimes as often as every hour)

[2] See: http://www.mayoclinic.org/diseases-conditions/type-2-diabetes/basics/causes/con-20031902

– Unusual weight loss or gain
– Fatigue
– Nausea, perhaps vomiting
– Blurred vision
– Dry mouth
– Slow-healing sores or cuts
– Itching skin

[3]

Applying this approach across multiple sites, we can complete a manual map. For the current study data mappings were made for only four diseases. It was deemed an acceptable number to demonstrate the ability to disambiguate between diseases, symptoms, anatomy, properties, and devices.

4. Related Work

Maedche et al [8] used the notion of a semantic bridge to map ontologies. This same approach was used in the current study to link the DOID, SYMP, FMA, QUDT, and SSN ontologies without impacting the original ontologies or their repositories. Bozic, et al [1], demonstrated the use of semantic bridges with aspects of SSN and time-series climate change data.

Manual ontology data mapping requires sophisticated language understanding skills and subject matter expertise. Rance, et al [11] demonstrated an semi-automated approach could be used for data mapping. They sought to show a mapping between two specialized data sources for rare diseases (i.e., the Office of Rare Diseases Research - ORDR, and Orphanet), using the Unified Medical Language System (UMLS) as the mapping pivot. They used the online Mendelian Inheritance in Man (OMIM) as the reference. They used syntactic filters to normalize the data source results to match against the UMLS and OMIM data. Although Rance, et al, demonstrated an automatic mapping in 79%-95% of the cases (ORDR vs. Orphanet, respectively) it is clear that an automated approach can be used to reduce the amount of work needed by, but cannot entirely replace, the subject matter expert.

5. Conclusions

The case study described focused on integrating 5 currated ontologies for the purpose of demonstrating semantic search for health care professionals in a real-world scenario between wearable devices, physical properties, human anatomy, symptoms and diseases. Ontology choices were limited to those that used unique identifiers. Ontology alignment was achieved using single-property semantic bridge ontologies (DOID→SYMP, SYMP→FMA, SYMP→QUDT, SSN→QUDT, and SSN→FMA, and is an effective approach for aligning currated ontologies. Manual unstructured text data mining was used to populate the bridge ontologies. Only four of the 9,000 diseases in DOID, twelve of the 1,000 symptoms in SYMP, thirty of the roughly 240,000 anatomical entries in FMA, and twenty of the 1,400 quantities in QUDT were used in this study. Considering the size of these models, manual data mapping would be impractical to build out or maintain complete mappings so a combination of automatic candidate construction with human validation and verification would be appropriate.

6. Future Work

The production of a full-fledged demonstrator for wearable devices in professional health care requires that the mappings performed manually, in this study, for a few diseases, symptoms, etc. be fleshed out to the full ontologies. The approach would be to identify those properties that can be measured by wearable devices, map to the symptoms that can be associated with those properties, and then map to the anatomical parts and diseases that those symptoms are associated with. It would be appropriate to use semi-automated tools to achieve these mappings. For example, the content mining of unstructured text can be generalized to three steps that can be automated to varying degrees:

– Search for target information from first model's data
– Parse and disambiguate search results
– Normalize search results to second model's data

The search for target information can be implemented with a web crawler, bot, etc. Next, the resulting pages can be mined with existing parsing approaches. Third, the results can be normalized and compared to the target model's labels using staged syntactic filters such as those used by Rance, et al [11]. In some applications the acquisition, parsing, and normalization phases can be performed by a single or tiered approach. It is possible that systems such as

[3]See: http://www.webmd.com/diabetes/understanding-diabetes-symptoms

IBM's Watson[6], which is designed to perform targeted search on the web and parse/normalize the results, could be used in this capacity.

References

[1] B. Bozic and J. Peters-Anders and G. Schimak, *Ontology Mapping in Semantic Time Series Processing and Climate Change Prediction*, In International Environmental Modelling and Software Society (iEMSs), 7th Intl. Congress on Env. Modelling and Software, San Diego, CA, D. Ames and N. Quinn and A. Rizzoli eds, 2014.

[2] N. Choi and I. Song and H. Han, *A Survey on Ontology Mapping*, In SIGMOD Record, Vol. 35, No. 3, Sept. 2008.

[3] A. Herrouz and C. Khentout and M. Djoudi, *Overview of Web Content Mining Tools*, In The International Journal of Engineering and Science (IJES), Volume 2, Issue 6, 2013.

[4] Ralph Hodgson, *Quantities, Units, Dimensions, and Datatypes Ontologies*, 2012.

[5] Y. Kalfoglou and M. Schorlemmer, *Ontology mapping: the state of the art*, 2003.

[6] A. Kalyanpur and B.K. Boguraev and S. Patwardhan and J.W. Murdock and A. Lally and C. Welty and J.M. Prager and B. Coppola and A. Fokoue-Nkoutche and L. Zhang and Y. Pan and Z.M. Qiu, *Structured data and inference in DeepQA*, 2012.

[7] L. Lefort and C. Henson and K. Taylor , *Semantic Sensor Network XG Final Report*, 2011.

[8] A. Maedche and B. Motik and N. Silva and R. Volz, *MAFRA - A MApping FRAmework for Distributed Ontologies in the Semantic W*, In EKAW 2002, LNAI 2473, A. Gomez-Perez and V.R. Benjamins eds, Springer-Verlag pubs, pp. 235-250, 2002.

[9] Onard Mejino, *Foundational Model of Anatomy*, 2012.

[10] OMG, *Model Library for Quanties, Units, Dimensions and Values (QUDV)*, Version 1.2, In: OMG Document ptc/2009-08-16, OMG Systems Modeling Language (OMG SysML), 2009.

[11] B. Rance and M. Snyder and J. Lewis and O. Bodenreider, *Leveraging Terminological Resources for Mapping between Rare Disease Information Sources*, In: Medinfo, C.U. Lehmann et al eds., IMIA and IOS Press, 2013, pp. 529-533.

[12] Lynn Schriml, *DOID: The Human Disease Ontology*, 2012.

[13] Lynn Schriml, *SYMP: The Symptom Ontology*, 2012.

[14] Vandrico, *The Wearables Database*, 2014.

[15] W3C, *Library for Quantity Kinds and Units: schema, based on QUDV model OMG SysML, Version 1.2*, 2005.