

ABSTRACT

Developing an application ontology for mining clinical reports: the extended syndromic surveillance ontology

M Conway, J Dowling, R Tsui, and W Chapman

Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA
 E-mail: michaelambroseconway@gmail.com

Objective

To develop an application ontology—the extended syndromic surveillance ontology (ESSO)—to support text mining of ER and radiology reports for public health surveillance. The ontology encodes syndromes, diagnoses, symptoms, signs and radiology results relevant to syndromic surveillance (with a special focus on bioterrorism).

Introduction

Ontologies representing knowledge from the public health and surveillance domains currently exist. However, they focus on infectious diseases (infectious disease ontology), reportable diseases (PHSKb—retired) and internet surveillance from news text (BioCaster ontology), or are commercial products (OntoReason public health ontology). From the perspective of biosurveillance text mining, these ontologies do not adequately represent the kind of knowledge found in clinical reports. Our project aims to fill this gap by developing a stand-alone ontology for the public health/biosurveillance domain, which (1) provides a starting point for standard development, (2) is straightforward for public health professionals to use for text analysis, and (3) can be easily plugged into existing syndromic surveillance systems.

Methods

The extended syndromic surveillance ontology (ESSO) uses the syndromic surveillance ontology (SSO) as a starting point. SSO is an ontology of chief complaints and consensus syndrome definitions developed by representatives from 10 currently functional surveillance systems.¹ In developing the ESSO, we tripled the number of concepts represented and added important new relations (for example, *has_symptom*). All new clinical concepts and relations were identified by an infectious disease physician (JD).

The ontology consists of 300 clinical concepts, each mapped to one (or more) of eight syndromes (see Table 1). Each concept is associated with a concept *type* (for example,

Table 1 Syndromes represented with example concepts

| Syndrome | Example concept |
|-----------------------------|-----------------|
| Rash | Measles |
| Hemorrhagic | Hematuria |
| Botulic | Slurred speech |
| Neurological | Seizures |
| Constitutional | Lethargy |
| Respiratory, sensitive | Sore throat |
| Respiratory, specific | Pneumonia |
| Gastrointestinal, sensitive | Dehydration |
| Gastrointestinal, specific | Vomiting |
| Influenza-like-illness | Prostration |

Note that respiratory and gastrointestinal syndrome are subdivided into *specific* and *sensitive* syndromes.

Table 2 Types of clinical concepts represented (with examples)

| Clinical concept type | Example concept |
|-----------------------|------------------|
| Diagnosis | Emphysema |
| Syndrome | Reactive airways |
| Symptom | Abdominal pain |
| Chest X ray finding | X ray pneumonia |
| Bioterrorism disease | Cholera |

diagnosis, bioterrorism disease—see Table 2) and, where possible, mapped to terms from other, related knowledge sources (for example, BioCaster, UMLS). Additionally, we include textual patterns (in the form of regular expressions), which help map clinical concepts to their textual instantiations (and *vice versa*). We encoded this structure as an OWL (web ontology language) file using the Protégé-OWL ontology editor (the current standard for ontology development).

Results

ESSO consists of 300 clinical concepts organized according to their clinical category (for example, *diagnosis*, *symptom*, or

radiology finding). Further, these 300 concepts are linked together through a series of relations (for example, DISEASE has_symptom Symptom, SYMPTOM isRelatedTo SYMPTOM). Preliminary evaluation in the influenza-like-illness domain has shown that the ontology has good domain coverage.²

Conclusion

The ESSO is a light weight, easily embedded ontology application designed to provide reasoning support for syndromic surveillance of clinical reports using text mining.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Okhmatovskaia A, Chapman W, Collier N, Espino J, Buckeridge DL. SSO: the syndromic surveillance ontology. 2009 *ISDS Annual Conference abstract* (unpublished).
- 2 Conway M, Dowling J, Chapman W. Developing a biosurveillance application ontology for influenza-like-illness. Proceedings of the OntoLex workshop: Beijing 2010.