

Using cKASS to facilitate knowledge authoring and sharing for syndromic surveillance

Liqin Wang^{1*}, Mingyuan Zhang¹, Mike Conway², Peter Haug¹ and Wendy Chapman²

¹Biomedical Informatics, University of Utah, Salt Lake City, UT, USA; ²University of California, San Diego, La Jolla, CA, USA

Objective

We describe cKASS (clinical Knowledge Authoring & Sharing Service), a system designed to facilitate the authoring and sharing of knowledge resources that can be applied to syndromic surveillance.

Introduction

Mining text for real-time syndromic surveillance usually requires a comprehensive knowledge base (KB), which contains detailed information about concepts relevant to the domain, such as disease names, symptoms, drugs and radiology findings. Two such resources are the Biocaster Ontology (1) and the Extended Syndromic Surveillance Ontology (ESSO) (2). However, both these resources are difficult to manipulate, customize, reuse and extend without knowledge of ontology development environments (like Protégé) and Semantic Web standards (like RDF and OWL). The cKASS software tool provides an easy-to-use, adaptable environment for extending and modifying existing syndrome definitions via a web-based Graphical User Interface, which does not require knowledge of complex, ontology-editing environments or semantic web standards. Further, cKASS allows for—indeed encourages—the sharing of user-defined syndrome definitions, with collaborative features that will enhance the ability of the surveillance community to quickly generate new definitions in response to emerging threats.

Methods

We have developed a web-based prototype of the cKASS system that allows individual users or collaborative communities to access the service anytime and anywhere, without a complex technical configuration process (Fig. 1). Two types of databases are used to support cKASS. First, a relational database is used to store user information and KB descriptors (e.g., KB domain and status). Second, KBs are stored as RDF triples using triple store and queried using SPARQL, an RDF query language, with the Jena SDB (SPARQL database,) providing robust and scalable storage. Existing resources stored in standard RDF

and OWL formats can be easily loaded into the triple store and used as a basis for constructing new syndrome definitions.

The web interface is designed to support both individual and collaborative KB development. cKASS consists of two zones:

1. User workspace, where registered users can create, browse, modify and publish customized syndrome definitions constructed from either publicly available or user-created resources.
2. Community space, where anyone can browse and search shared KBs. Users who choose to share their KBs can make them available to the general community in this space.

cKASS also provides search/query capabilities at different levels. For example, the user can search within a specified domain or within a named KB for terms or concepts. Queries can either be simple strings or can consist of arbitrarily complex SPARQL and SQL queries. Further, users can import queried results into their syndrome definitions (for example, concept only, concept and its attributes or concept and all its subclasses). Finally, once created, KBs can be exported in standard formats, such as XML, CSV or RDF, for use with other tools.

Results

Currently, two existing syndromic surveillance oriented ontologies—Biocaster and ESSO—have been loaded into the cKASS triple store and can be used as a basis to construct new syndrome definitions. Both KBs can be queried using SPARQL and SQL.

Conclusions

cKASS offers public health professionals and clinicians an environment to support the extension and modification of existing KBs, without the need to use complex ontology editing environments and formalisms, allowing the user to rapidly develop or augment existing syndrome definitions and react quickly to the changing surveillance landscape.

Keywords

Syndromic surveillance; knowledge authoring; ontology

References

1. Collier N, Goodwin RM, McCrae J, Doan S, Kawazoe A, Conway M et al. An ontology-driven system for detecting global health events. In: Proc. 23rd International Conference on Computational Linguistics (COLING), Beijing, China, August 23-27; 2010:215-22.
2. Conway M, Dowling J, Chapman W. Developing an application ontology for mining free text clinical reports: the extended syndromic surveillance ontology. In: Proc. 3rd International Workshop on Health Document Text Mining and Information Analysis (LOUHI), Bled, Slovenia, July 6; 2011:75-82.

*Liqin Wang

E-mail: liqin.wang@utah.edu

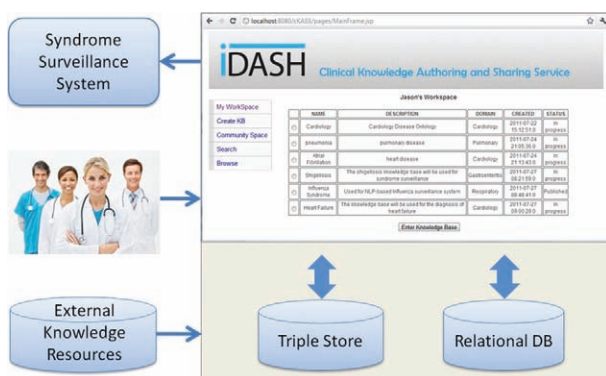


Fig. 1. The architecture of cKASS.