Notifiable Diseases Knowledgebase with Protégé: From Print to Component

Haobo Ma, MD, MS; Timothy Doyle, MPH
Division of Public Health Surveillance and Informatics, Epidemiology Program Office,
Centers for Disease Control and Prevention; Atlanta, GA; hjm8@cdc.gov.

Background

Epidemiologists rely on standardized case definitions to collect and aggregate disease surveillance data. The descriptive nature of paper-based disease case definitions restricts their interpretation by computers. In an era of computer-based public health surveillance, a formalized knowledge representation for reportable diseases can provide computer-accessible definitions and increase system interoperability. With assistance from the Stanford Medical Informatics group, we developed a knowledgebase in Protégé of reportable public health conditions and related surveillance information.

Knowledge Representation of Case Definitions in Protégé

The reportable conditions are unique concepts in our knowledgebase and defined by their location in the hierarchical structure as well as their epidemiological components, such as pathogen, transmission mode, jurisdiction, clinical criteria, and laboratory criteria.

We represented the content of case definitions at three levels of computer accessibility. The first level is as free text components. We dissected the descriptive information and put it into appropriate slots. For example, “Clinical Criteria for Diagnosis” and “Laboratory Criteria for Diagnosis” contain corresponding parts in the free text case definition. At this level, we represent case definitions at a relatively coarse granularity, useful for re-creating a complete handbook of case definitions. The second level incorporates internal operational information, such as the “Notifiable Disease Code”, a

code previously maintained outside the case definition content and which is important for data aggregation. Within the knowledgebase, we expressed this code as a separate slot. This practice helps synchronize the maintenance of internal codes with the case definitions. The third level incorporates standard medical terminologies. We mapped case definition content to standard medical terminologies. Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) was used for disease, organism, symptom, and finding concepts; International Classification of Diseases (ICD) for disease; and Logical Observation Identifiers Names and Codes (LOINC) for laboratory orders related to the laboratory criteria. At this level, the knowledgebase provides maximal interoperability between surveillance, clinical, and laboratory information systems.

Applications of Notifiable Disease Knowledgebase

On the basis of the architecture of the knowledgebase, we designed two different applications using Protégé JAVA API and JESS Tab plug-ins. The first application is a web-based query interface for epidemiologists. The second is an automated classification module based on standard medical terminologies in electronic laboratory reporting systems.

Conclusion

The Notifiable Disease Knowledgebase developed in Protégé is a formal approach to knowledge sharing and reuse in the public health surveillance domain and holds promise for enhancing public health surveillance through the creation of computer-accessible case definitions.