

Help For HealthCare: Mapping Unstructured Clinical Notes To ICD-10 Coding Schemes

dataversity.net/help-healthcare-mapping-unstructured-clinical-notes-icd-10-coding-schemes/

11/26/2013

by [Jennifer Zaino](#)

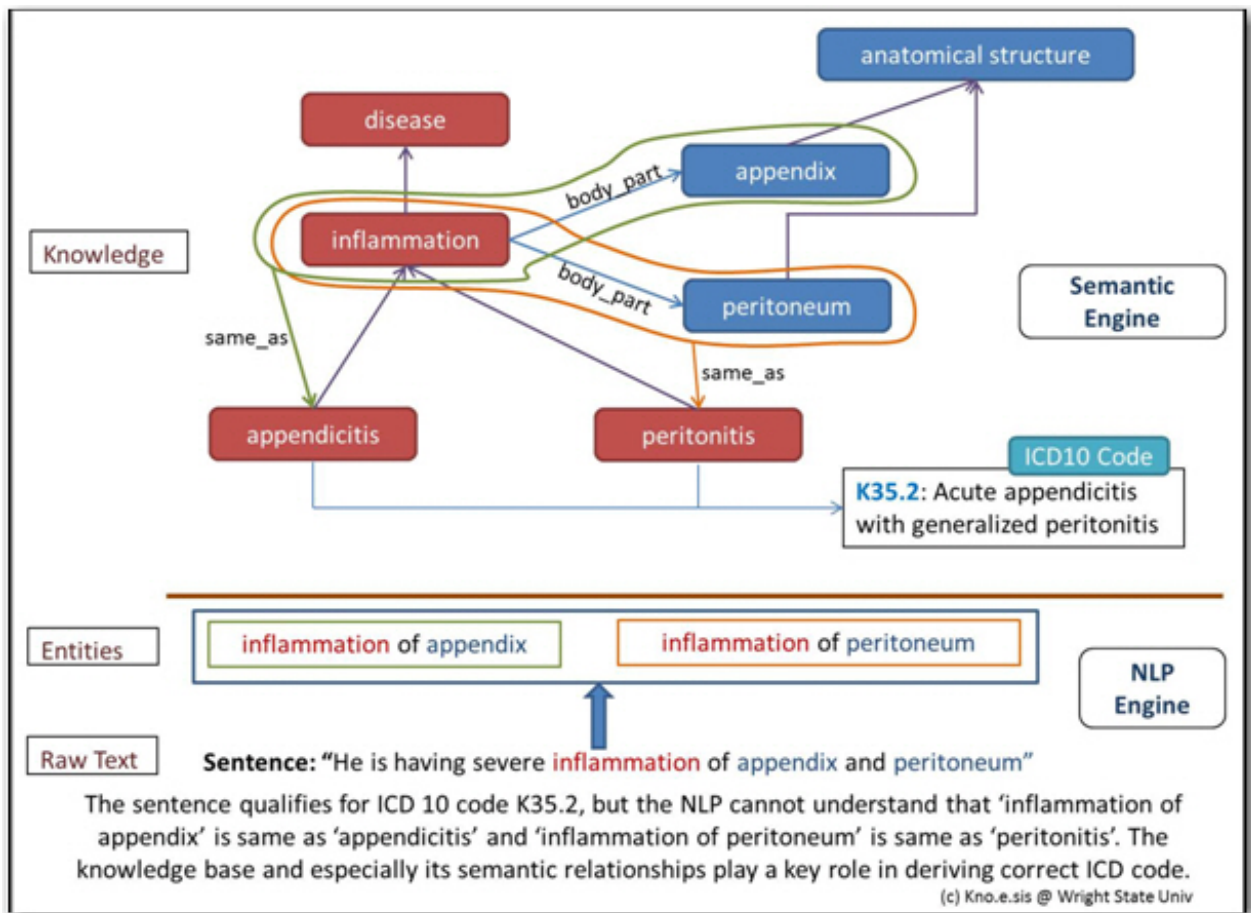
The health care industry – and the American citizenry at large – has been focused of late on the problems surrounding the implementation of the Affordable Care Act, the federal website’s issues foremost among them. But believe it or not, there are other things the healthcare industry needs to prepare for, among them the October 1, 2014 date for replacing the World Health Organization’s [International Statistical Classification of Diseases and Related Health Problems](#) ICD-9 code sets used to report medical diagnoses and inpatient procedures by ICD-10 code sets. ICD-9 uses 14,000 diagnosis codes which will increase to 68,000 in ICD-10, which is a HIPAA (*Health Insurance Portability and Accountability Act*) code set requirement.



Natural language processing has had the primary role in many solutions aimed at transforming large volumes of unstructured clinical data into information that healthcare IT application vendors and their hospital customers can leverage. But there’s an argument being made that understanding unstructured text of clinical notes that contain a huge stash of information and then mapping them to fine-grained ICD-10 coding schemes requires a combination of NLP, advanced linguistics, machine learning and semantic web technologies, and Amit Sheth, professor of computer science and engineering at [Wright State University](#) and director of the [Kno.e.sis Center](#) is making them. (See our story [yesterday](#) for a look at how the NLP market is evolving overall, including in healthcare.)

“ICD-10 has thousands of codes with millions of possible permutations and combinations. A rule-based approach is not effective to cover the huge number of ICD-10 codes.” Sheth says. Extracting the correct concepts, identifying the relationship between these concepts and mapping them to the correct code is a major challenge, with codes often formed by information from various sections of a clinical document that itself is subject to individual physicians’ style of recording information, among other factors.

Sheth gives one example of how easy it is to miss things: In the sentence: “He is having severe inflammation of appendix and peritoneum,” the typical rules-based NLP engine on its own doesn’t recognize that inflammation of appendix means appendicitis, and inflammation of peritoneum equals peritonitis (see graphic below). “Without a knowledge base, without an ontology, the NLP engine did not have enough context” to make the match to ICD-10 code K35.2, he says. When NLP is combined with semantic web technologies, however, it can help machines identify the right concepts, identify interconceptual relationships, understand and disambiguate abbreviations, and continuously learn and improve. He and a team of students at Kno.e.sis set about to raise the level, and their work today is the basis of the cloud-based coding product ezCAC from [ezDI](#), which launched late last month.



Rules-based NLP solutions also require that humans input the rules – and then re-input them when exceptions occur. That takes time, which means that users have to wait as long as months for a new version of their solution to be available, he notes. That wait can be eliminated when “instead you take the NLP engine and use an ontology which has the relationships that are the forms of rules, and you get the same benefit. All you have to do is to change the knowledge base,” Sheth says, an easier proposition than recoding. Once that’s done, the next time the NLP engine with an ontology in its rack looks at a text it is equipped to have the appropriate understanding and make the right identification. Additionally, this approach avoids the costs of having to maintain staff focused on maintaining the rules.

The work at Kno.e.sis also delivered unique IP in the way of an ability to assess the richness of a knowledge base with respect to a given corpus, determine the missing domain relationships in the knowledge base, and suggest the most plausible relationships that can fill the gap created by the missing relationships. “It’s highly semi-automated in the sense that humans make the ultimate call. But it helps for keeping the ontology up with new medical knowledge at very low cost.”

Translating clinical data to the ICD-10 codes matters because, as much as any other organizations, hospitals are driven by concerns about revenue cycle management. “Hospitals don’t get paid unless they encode the bill right and send it to the insurance company,” he notes. “They need some computer-assisted coding to improve their operations.”