

# Realizing the full potential of electronic health records: the role of natural language processing

doi:10.1136/amiajnl-2011-000501

Lucila Ohno-Machado, *Editor-in-chief*

Meaningful use of electronic health records (EHRs) for patient care or for research requires data to be comparable. Many portions of EHRs continue to be unstructured, presenting significant challenges for biomedical informatics. This issue of the journal displays several solutions to this problem that are based on natural language processing (NLP) techniques. A high-level review by Nadkarni (*see page 544*) is intended to introduce the main components of NLP for the novice, and to briefly describe machine learning methods that are successfully being employed in the field. It includes a discussion on Watson, a contestant on 'Jeopardy!', a popular question-and-answer TV show, and the ensuing speculations about its potential extensions to medical NLP. However, despite some notable examples of successful NLP applications in clinical care, progress in the field has been relatively slow.

Chapman and colleagues (*see page 540*) discuss the need to steer current NLP research efforts so that new developments can be accelerated, and research products can become readily usable in healthcare applications. The authors advocate for a concerted, collaborative effort to develop open-source software components and infrastructure to share annotated data and tools. The guest editorial raises important questions about the ability of the NLP community to collaborate in addressing big challenges that go beyond the scope of what can be accomplished by a single group. There are positive signs that the NLP community is ready for this: it has been able to share data and tools in the organization of software competitions, which are related to specific medical NLP tasks. The community may have reached a point in which researchers can start building on each other's achievements. *JAMIA* has historically featured the best papers from NLP competitions (*see pages 557–613*), and takes this opportunity to encourage readers, researchers, and developers in all informatics specialties to think beyond competitions, to envision a future in which most effort is spent on collaborative initiatives that

help assemble big teams to solve big problems, and to set a goal of developing new solutions that have a direct impact on healthcare.

Several NLP and information retrieval articles are featured in this issue: Garla (*see page 614*) provides an excellent example on how existing open-source components of an NLP pipeline, which was developed at another institution, can be enhanced. The authors evaluate their new components for the problem of detecting hepatic decompensation from findings contained in radiology reports. F Liu (*see page 625*) quantifies the improved performance of clinical query systems after speech recognition systems are adapted for medical language. Lokker (*see page 652*) reports on the value of the *Clinical Queries* filters in PubMed for retrieval of relevant articles for practicing physicians. Hunter (*see page 621*) reports on a novel application of data-to-text generation in the context of ICU summaries, where clinicians receive information in a familiar format.

Several other examples of how information from EHRs and the literature can be automatically processed for clinical decision support are featured in this *JAMIA* issue: Botsis (*see page 631*) compares the performance of several machine learning and rule-based classifiers for text-mining in the context of vaccine-adverse event reporting. Herasevich (*see page 639*) evaluates a rule-based system to detect sepsis in EHRs from an intensive-care unit. H Liu (*see page 645*) evaluates the adequacy of a proposed e-prescription standard, assessing whether it is ready for adoption given potential ambiguities and imperfect mapping to controlled terminologies. Other authors address the issue of structuring text from the biomedical literature: Huang (*see page 660*) describes a new approach for assigning MESH terms to documents based on nearest-neighbor documents. Shetty (*see page 668*) compares the performance of several filters for detection of drug-adverse events from information contained in PubMed.

Once data are structured, the ability to exchange them for clinical care or research

is greatly enhanced. Thus, without the wide adoption of structured EHRs, our field will not move forward. This issue also provides examples of realistic and timely accounts of critical issues in implementing EHRs: Banas (*see page 721*) describes the difficult road toward EHR implementation and the lessons learned that can help those who are starting their journey; Zheng describes problems with methods used to study clinical processes and workflows that need to be understood for successful EHR implementation; and Carroll (*see page 717*) provides a word of caution, showing that the ability of EHRs to accurately report critical information such as medication adherence is still limited. Finally, a perspective from Simborg (*see page 675*) addresses a way to regulate fraud in EHRs by calling on the Office of the National Coordinator for Healthcare Information Technology to regulate accountability.

Furthermore, health information exchange (HIE) presents its own set of challenges. Kuperman (*see page 678*) provides a lucid account on HIE efforts of the past decade, explaining why some efforts failed, while others have (at least partially) succeeded, and describes where current efforts are focused. However, there is evidence for optimism: Gadd (*see page 711*) reports on clinicians' high perception of usability of an HIE that has been operational since 2004. Li (*see page 683*) describes the Dolphin project, initiated in 1998, which is a system to exchange and translate records from two prefectures in Japan and one in China that overcame important logistical, cultural, and language barriers. LaBorde (*see page 698*) describes a method for analyzing patient crossover between different institutions that can be used by decision-makers to understand the need for HIE and the expected volume of transactions.

Biomedical informatics is evolving at a rapid pace, and by leveraging each other's efforts, we will achieve our collective goal of improving care and preventing disease through computation. It is exciting to see how far we have come as a scientific community and to document this progress within the pages of *JAMIA*.