A modular approach for Consolidating CCDs from multiple data sources

Masoud Hosseini, MSc[1,2], Jonathan Meade[3], Jamie Schnitzius[3], Brian E. Dixon, MPA, PhD[1,2,4]
[1] Indiana University School of Informatics and Computing, Department of BioHealth Informatics
[2] Regenstrief Institute, Center for Biomedical Informatics
[3] CreateIT Healthcare Solutions, Inc.
[4] Center for Health Information and Communication, Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service
Indianapolis, Indiana, USA

Abstract

Clinical information is fragmented across many different organizations and computer systems. To view complete, longitudinal medical records for a patient, clinical systems are starting to routinely, electronically request information from various sources in the form of a Continuity of Care Document (CCD). Given the heterogeneity and fragmentation of data across sources, systems are likely to receive multiple CCDs for a single patient. Current tools for viewing CCDs make it cumbersome for providers to explore different CCDs to find a specific data which can be duplicated or even conflicted. We have developed a modular approach for integrating and de-duplicating multiple CCDs into one consolidated document. Our system is designed to support nationwide and regional health information exchange efforts to support integration of standard clinical documents into clinical information systems for presentation to clinicians.

Introduction

The Continuity of Care Document (CCD) is an electronic document exchange standard for sharing patients' health information summary between providers and organizations which is required under Meaningful Use criteria. Given the fragmented and heterogeneous nature of clinical data, requests for complete medical records results in multiple CCDs from a variety of sources, however, care providers prefer to review consolidated information that represents a single, comprehensive picture of a patient's medical history and current condition, rather than multiple CCD documents that may include duplicate or conflicting information.

Methods

We designed a modular, open source system that aims to consolidate and de-duplicate received CCDs. The key component is the CCD Consolidation Engine, which executes a set of rules against a list of CCDs to produce a single, clinically useful CCD. Four types of rules are executed: 1) Pre-Format Rules examine incoming CCDs and inspect them for quality; 2) Primary Rules ensure the program returns valid information to the user; 3) De-duplication Rules merge, de-duplicate and handle conflicting information in the list of CCDs; and 4) Post-Format rules clean up the returned, consolidated and de-duplicated CCD. The system further audits all actions performed to comply with federal regulations. We further developed an application programming interface (API) to enable remote interaction through web services. The system is developed using the C# language and we used Microsoft .NET libraries to manipulate XML documents. Audit data is retained in a NOSQL database, such a Hadoop.

Results

A prototype was developed for a local health software competition, for which it took top honors. The prototype currently executes consolidation and deduplication rules against a de-identified reference dataset at a rate of approximately .009 to .03 seconds per rule depending on complexity. Evaluation of precision and recall using a larger dataset from an operational health information exchange is currently underway.

Conclusion

Given that meaningful use regulations and many HIE organizations are less than three years old, there do not yet exist a plethora of proven solutions for consolidating CCDs. Consolidation of CCDs, however, is challenging and requires processes that can interpret, merge, de-duplicate, and resolve conflicts across complex documents involving complex data types. Accuracy of the de-duplication and consolidation greatly depends on the logic used in the Consolidation Engine. The benefit of the audit model chosen for this project is that it can be used to analyze the logic and clinical usefulness of the output. Health informatics researchers and scientists will be able to improve the algorithms and methodologies used in the rule logic over time increasing the clinical usefulness of the CCD.