MEDINFO 2017: Precision Healthcare through Informatics
A.V. Gundlapalli et al. (Eds.)

doi:10.3233/978-1-61499-830-3-1276

Asynchronous Query Distribution Between Multiple i2b2 Research Data Warehouses: Li2b2-SHRINE

Raphael W. Majeed [a, b], Mark R. Stöhr [a], Volker S. Thiemann [b], Rainer Röhrig [b] and Andreas Günther [a]

[a] *UGMLC, German Center for Lung Research (DZL), Justus Liebig University, Giessen, Germany*
[b] *Department of Medical Informatics, Carl von Ossietzky University, Oldenburg, Germany*

Abstract

Clinical Data Ware Houses are established sources for research and quality management. The open source data warehouse software i2b2 enjoys good reputation and wide-spread use in the international medical informatics community. We developed a novel infrastructure to allow queries to be distributed asynchronously between i2b2 data warehouses.

Keywords:

Biomedical Research; Information Storage and Retrieval

Introduction

Clinical Data Ware Houses are established sources for research and quality management. The open source data warehouse software i2b2 enjoys good reputation and wide-spread use in the international medical informatics community. Federated searches across different hospitals are possible with the "Shared Health Research Information Network" (SHRINE) [1] software which unfortunately has some limitations for hospitals with rigid IT security policies and counties with strict data protection – most notably requiring direct network access to participating hospitals as well as not allowing per query approvals by hospitals due to the synchronous query connections.

Aim of this project is the development of a software prototype which overcomes the above shortcomings and allows distribution of queries to federated data warehouses asynchronously without the need for direct network connections to the hospitals.

Methods

The development of an asynchronous network of i2b2 data warehouses can be broken down into three parts: (a) distribution infrastructure, (b) central query frontend and (c) integration software for all participating data warehouses.

For the decentralised query distribution infrastructure, we used software from the German emergency care registry [2]. The infrastructure is content-agnostic and supports any format or language for queries as well as query responses.

Central frontend and data warehouse integration is realized with li2b2-façade [3] with focus on minimal integration effort.

Results

We developed a standalone central server application which employs the original i2b2 webclient to formulate queries. Queries are stored and can be retrieved by data warehouse nodes at any time. Results are transferred back to the server and displayed in the web frontend.

Terminology and logic of the central queries can be adapted to local characteristics via customizable XML transformations.

Each additional SHRINE-node requires an API-Key in the central server configuration file. The distributed query network was successfully tested with 20 virtual data warehouse servers.

All source code is open source and available online at GitHub: https://github.com/li2b2/li2b2-shrine.

Discussion

In contrast to the original i2b2 SHRINE project [1], our solution does not require a full data warehouse installation for the central query interface. The server as well as the data warehouse connector run on any system without installation.

The presented software can be easily adapted to allow other data warehouse software to join a network of i2b2 data warehouse servers: Quick evaluation showed that a commercial data warehouse software with a RESTful query interface could be integrated with 200-300 lines of code (source link above).

Conclusion

Our software "li2b2-shrine" allows data warehouse queries to be distributed to multiple i2b2 data warehouses asynchronously without the need for firewall adjustments. Existing i2b2 data warehouses can be interconnected with minimal effort.

References

[1] G.M. Weber and et al, "The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories." *Journal of the American Medical Informatics Association* **16.5** (2009): 624-630.
[2] J. Ahlbrandt, D. Brammen, R.W. Majeed and et al, Balancing the need for big data and patient data privacy--an IT infrastructure for a decentralized emergency care research database. Stud Health Technol Inform. **205** (2014): 750-754.
[3] R.W. Majeed and et al.: Li2b2-façade: Simulation of i2b2 data warehouse server and client for interaction with other systems. *Stud Health Technol Inform* (2017).

Address for correspondence

e-mail: raphael.majeed@chiru.med.uni-giessen.de