A Grid System for Timely Surveillance of Influenza/Pneumonia Using Death Records

Catherine Staes[a], Ronald C. Price[b], Kailah Davis[a], Barry Nangle[c], Jeff Duncan[c], Carol Friedman[d], Albert M. Lai[d, e], Lyudmila Shagina Ena[d], and Julio C. Facelli[a, b,]

[a] Department of Biomedical Informatics and [b] Center for High Performance Computing, University of Utah,
[c] Utah Department of Health, [d] Department of Biomedical Informatics, Columbia University
[e] Department of Biomedical Informatics, The Ohio State University.

Abstract and Objective

This poster demonstrates how to integrate grid data services, death certificates, analytical grid services, and natural language processing, to build a real time public health surveillance tool that uses data and services under the control of different administrative domains. The example used here provides insight on how a global public heath grid could be developed as a dynamically evolving ecosystem of grid enabled applications and data sources.

Keywords:

Public health surveillance, Grid computing, Public health grid

Introduction

Grid architecture is a promising methodology to aggregate and analyze disparate, heterogeneous data existing under independent administrative domains and to provide computational and analytic resources on demand [1]. The Utah team has made available de-identified death records from the Utah Department of Health as a data grid service. Columbia University has deployed a grid version of MedLEE, a natural language processing (NLP) tool [2]. This tool can be used to process the free text recorded in the 'cause of death' fields included in the death records and output either a controlled vocabulary or desired codes. Coded information about causes of death is needed to integrate death certificate information into real time surveillance tools in order to trigger alarms to initiate more detailed public heath surveillance. Currently, real-time surveillance of death certificate data is limited because coded death certificate data is typically not available for at least 1-3 months after the date of death [3].

Methods

This poster provides preliminary results to demonstrate the integration of public health data and analytic services using grid services to address a relevant public health problem: Surveillance of influenza/pneumonia using death records. Using three years of historical de-identified death records, we analyzed how a prototype system can be used for the stated purpose. We divided the available data into two years that were used as baseline data, and one year that is used as to simulate a real time data feed. The historic baseline data was used to establish baseline trends that can be compared with the simulated real time feed. The simulated real time feed was exposed as a grid service under the administrative domain of the Utah Department of Health. The MedLEE grid services, under the administrative domain of Columbia University, access the simulated daily feed of the death records and provides coded information on the cause of death. A third grid application is used to summarize the total deaths by age group and by date for each disease code of relevance to influenza and pneumonia, compare with the baseline data, visualize trends and develop water marks that can trigger automatic alerts.

Results

We demonstrate that it is possible to integrate grid data and analytic services under different administrative domains and to develop agile and easy to deploy applications for public health.

Conclusions

We believe that developing these types of applications will contribute to the establishment of a global public heath grid as a constantly evolving ecosystem of grid services, which under diverse administrative control can be combined into novel applications. We intent to implement an operational prototype for the 2010-2011 influenza season.

Acknowledgements

The Utah authors acknowledge funding by the CDC Rocky Mountain Center of Excellence in Public Health Informatics 1P01HK000069-10, National Library of Medicine Training grant LM007124 and NCRR CTSA award 1KL2RR025763-01. The Columbia University authors were partially supported by NIH grant R01LM008635.

References

[1] Staes C, Xu W, LeFevre S, et al. A case for using grid architecture for state public health informatics: the Utah perspective. BMC Medical Informatics and Decision Making. 2009; 9(1):32.

[2] Friedman C, Hripcsak G, DuMouchel W, Johnson SB, Clayton PD. Natural language processing in an operational clinical information system. J Natural Language Engineering. 1995; 1(1): 83-108.

[3] Friedman DJ. Assessing changes in the vital records and statistics infrastucture survey. National Association for Public Health Statistics and Information Systems, 2007. [Accessed Oct 13, 2009] (available from: http://www.naphsis.org/index.asp?bid=984).

Address for correspondence

Julio.Facelli@utah.edu