Chapter 2.4
A Framework for Privacy Assurance and Ubiquitous Knowledge Discovery in Health 2.0 Data Mashups

Jun Hu
University of Ottawa, Canada

Liam Peyton
University of Ottawa, Canada

ABSTRACT

Knowledge discovery is a critical component in improving health care. Health 2.0 leverages Web 2.0 technologies to integrate and share data from a wide variety of sources on the Internet. There are a number of issues which must be addressed before knowledge discovery can be leveraged effectively and ubiquitously in Health 2.0. Health care data is very sensitive in nature so privacy and security of personal data must be protected. Regulatory compliance must also be addressed if cooperative sharing of data is to be facilitated to ensure that relevant legislation and policies of individual health care organizations are respected. Finally, interoperability and data quality must be addressed in any framework for knowledge discovery on the Internet. In this chapter, we lay out a framework for ubiquitous knowledge discovery in Health 2.0 based on a combination of architecture and process. Emerging Internet standards and specifications for defining a Circle of Trust, in which data is shared but identity and personal information protected, are used to define an enabling architecture for knowledge discovery. Within that context, a step-by-step process for knowledge discovery is defined and illustrated using a scenario related to analyzing the correlation between emergency room visits and adverse effects of prescription drugs. The process we define is arrived at by reviewing an existing standards-based process, CRISP-DM, and extending it to address the new context of Health 2.0.

DOI: 10.4018/978-1-61350-323-2.ch2.4

INTRODUCTION

Knowledge discovery is a critical component in improving health care. Web 2.0 technologies provide us an increasing ability to integrate and share data from a wide variety of sources on the Internet. Health 2.0 leverages those technologies to create on-line communities within which patients, caregivers, medical professionals, and other health stakeholders can collaborate and share data. In particular, both Google Google Health, 2009) and Microsoft Healthvault (HealthVault, 2009) are proposing services to maintain personal electronic health records on behalf of consumers, and share them as needed with physicians, clinics, and other health care providers. Consolidating all health care records across a broad spectrum of health care providers, provides a potentially very valuable source of information about health, disease and health care especially with respect to detecting trends in the spread of disease (SARS, Swine Flu) as well as measuring the efficacy of prescription drugs and the number of side effects. But to leverage that collective source of information, a framework for knowledge discovery is needed which defines a managed process for collecting, analyzing and publishing data relevant to healthcare decisions based on large, statistically significant data patterns within a Health 2.0 data architecture. Such a framework can promote a continuous process of evidence-based medicine and performance management to ensure that health care services are meeting the needs and objectives of the communities which use them.

There are a number of issues which must be addressed before knowledge discovery can be leveraged effectively and ubiquitously in Health 2.0. Trust is of paramount concern. Health care data is very sensitive in nature so privacy and security of personal data must be protected. Normally, in a single enterprise, data can be collected from multiple data sources within a single organization into a single data warehouse in a controlled fashion to build a model of the overall enterprise that can support knowledge discovery through analysis and data mining. In a Health 2.0 environment, though, different organizations each have their own independent sources of data that could be published, shared and linked over the Internet to support knowledge discovery.

Interoperability and data quality are also major issues. Typically, each organization stores its data in a different format, according to different standards, or in an ad hoc manner. Most family doctors still maintain paper records and write notes in natural language. In order to effectively process data in a consistent automated fashion, mechanisms must be put in place to standardize the data that is collected and linked. One of the biggest challenges is finding a mechanism for reliably identifying patients from different data sources, while still protecting privacy.

More importantly business agreements and regulatory oversight must be in place amongst the different organizations in order to collect and analyze data in this manner to ensure that the relevant legislation and policies of individual health care organizations are respected. As well, patient consent forms and access controls to ensure compliance with privacy laws (HIPAA, 1996; PIPEDA, 2000; PHIPA 2004) must be in place. Special care must be given, if public access is granted to data and knowledge discovery within a Health 2.0 network. Decisions based on health care data are complex and life affecting, so continuous monitoring to ensure validity, reliability and compliance of a Health 2.0 framework must be as ubiquitous as the data sharing it supports.

In this chapter, we lay out a framework for ubiquitous knowledge discovery in Health 2.0 based on a combination of architecture and process. First, we provide a summary and background of relevant technologies and trends in Health 2.0, paying attention to how privacy is protected via federated identity management. Then we introduce our framework. Emerging Internet standards and specifications for creating a Circle of Trust are used to define an enabling architecture for knowledge