Enabling the Health Internet - Market-Friendly Information Extraction of Online Healthcare Sources

Christan Grant[a], Tyrone Grandison[b], Alfredo Alba[b]

[a] *Database Research Center, Computer & Information Science & Engineering Department, University of Florida, Gainesville, FL, USA.*

[b] *Healthcare Informatics, IBM Almaden Research Center, San Jose, CA, USA*

Abstract

The Health Internet is the future of American healthcare. It is the patient-facing system with information about you and your condition. Admittedly, the vision of the Health Internet is evolving as this paper is being written and many aspects of it will no doubt change within the next few months. However, there are several things that will remain constant. The primary one being that the Health Internet will be a portal that integrates data from many different online sources. Unfortunately, current techniques for acquiring this information from these sources may have significant negative effects. In order to support the continued creation of novel healthcare applications (like the Health Internet), we define a process to ensure best practices for both Web publishers and end users. These techniques allow for continued innovation while respecting the source's terms and conditions. In turn, all parties involve will avoid a dramatic decline in both traffic and revenue, while enabling the integration of health information for the greater good – the patient's wellness.

Keywords:

Internet, Data collection, Information storage and retrieval

Introduction

Aneesh Chopra, the Chief Technology Officer of the United States, did two extremely important things in October of 2009. The first is stating the government's support for the creation of a financial environment where startups thrive. As part of this initiative, the Obama Administration has allocated a large amount of funding for the Health IT sector. There is already evidence of an increase in Healthcare websites and startups. The second is the announcement of the Health Internet project, which is to replace current NHIN (national Health Information Network) efforts.

In order for Web sites to protect their data, it is important to establish a set of best practices for Web publisher's in order to adjust to a changing online market.

Methods

Based on our experience creating the Health-e-Assistant system, we have observed many practices performed by healthcare applications, which would result in negative market impact. This work drove the development of a process model to ensure that the current and next generation health care web applications experience positive market feedback, i.e. legislative compliance, profitability, client safeguards and customer acceptance.

Discussion

Our separation of the world into data producers and data consumers offers us a nice, elegant abstraction that can be leveraged in multiple interesting ways outside of this work. For example, this allows the construction of simulations, games, models and workflows that could enable reasoning about optimal choices, tradeoffs and (dis-)incentives for sanctioned behavior. Additionally, as we presented the proper etiquette to be employed when publishing data and when obtaining data (whether by Web scraping or using application programming interface), which was employed in our real deployment [8], there are obvious research questions around countermeasures to stop inappropriate behavior. These will be the focus of future work. This changing market holds tremendous promise for healthcare startups and healthcare IT companies. For companies to continue to experience growth, data producers and consumers must show politeness while interchanging content. Our process model ensures that both producers and consumers exhibit a sufficient measure or level of Web etiquette, which will hopefully be a part of best practices for the field.

Conclusion

In this paper, we have shown, through our experience with healthcare efforts similar to the Health Internet, that the first goal of market-friendly information extraction from online healthcare sources is possible and we have provided guidance on how this task can be performed.