Using Mashup Technology to Integrate Medical Data for Patient Centric Healthcare

YoungWoo Pae[1,*], Gi-Cheol Bak[2], YoungJu Tak[1], KyungMee Park[1], and YongTae Shin[3]

[1] IBM Korea, Soul, Korea
{ywpae,yjtak,kmpark}@kr.ibm.com
[2] YangYoung Digital Highschool, Gyeonggi-do, Korea
neuron5@chol.com
[3] Soongsil University, Seoul, Korea
shin@ssu.ac.kr

Abstract. The information technology advances are driving the transformation of healthcare. The first characteristic of those modern trends is the large amount of publically available medical information. Harnessing the public medical knowledge for improving the patients' safety is inevitable. Physicians and patients can, literally, benefit from the vast public information and medical systems should incorporate it. The second characteristic is the greater demand of users to participate in their applications and to fully control them. There are different types of medical system users and building one interface that fits all is very difficult task. We design a healthcare system by leveraging the novel Mashup technology. It can overcome most of the above challenges. Our system uses Mashup to integrate public knowledge and provide customizable workspaces for the various end users. This comprehensive system is built with the aim of using advanced technologies for the patient centric healthcare.

Keywords: Mashups, data mashup, data integration, health information system, clinical decision support system, patient centric service.

1 Introduction

The scale and variety of public medical information are rapidly growing. Integrating information from those sources of information to create high quality customized information source becomes a challenge. It is much more challengeable to integrate the public information with medical data of patients or clinical studies.

Therefore, a platform for easy integration of the public information with medical data will be beneficial for patients, physicians and researchers. The lasts are expected to benefit from improved healthcare services that provide information that is more precise. Many previous works have attempted to integrate various sources of information with medical knowledge bases and Personal Health Record (PHR).

Several studies [2, 3] have shown that integrating health data from diverse sources are challenging because of the heterogeneity of the data structure, terminologies and

James J. (Jong Hyuk) Park et al. (eds.), *Future Information Technology*,
Lecture Notes in Electrical Engineering 309,
DOI: 10.1007/978-3-642-55038-6_11,

semantic. Those barriers prohibit the aspired analytical services that can be provided on top of the integrated clinical data and can supply to physicians, researchers and patients, significant clinical insights that will eventually improve the patients' safety as well as clinical treatments. A variety of technical approaches, like database federation [3], data warehousing [6], Web services and Semantic Web, have been developed to overcome the systems interoperability and the data integration challenges.

Online services, such as Healthline[1] or Drugs.com[2], operate domain experts that gather medical content from trusted medical publishers and manually consolidate the information. Such a manual process, though, consumes considerable amount of time and efforts and is limited to the integration of only few number of data sources.

In this paper, we present an agile data integration solution using a kind of Web-based data integration methodology called Mashup [1, 2]. Mashups have been designed to empower users, who are able to extract, format, and remix data from multiple data sources. In addition, Mashups have been designed to allow disparate data sources to be brought together to make a new Web application. Mashup provides tools for creating customized feeds of data from several sources. Those feeds can later on be used for providing dynamic services. However, the process of creating logical feeds becomes challenging because of the complexity of medical information. This complexity may explain the sparse number of data mashup works in the healthcare domain: Cheung et al. [2] implemented a data mashup that integrates cancer data with environment data; Belleau et al. [1] studied the usage of data mashups for bioinformatics domain. We propose taking these studies a step further by utilizing the PHR with drug knowledge from publically available data sources.

We design a patient centric healthcare system, named Medics, which integrates content from different online medical sources using Mashup technology to provide flexible Adverse Drug Events (ADE) alerts. We propose how the Mashup platform can create highly customizable workspaces for different types of end-users: physicians, patients, nurses, researchers, etc. Each type of user has its own unique requirements and preferences.

2 Patient Centric Healthcare

Medics is a patient centric healthcare system including clinical decision support system (CDSS) that brings the power of analytics to bear on everyday clinical decisions and provides personalized clinical guidelines and contents for patients and physicians. The management of medical knowledge in CDSS systems is a task of a great challenge, since the available knowledge in the healthcare domain is expanding continuously, and its types vary. In general, the knowledge can be divided as structured data, semi-structured data, and unstructured data. Medical semi-structured data such as data from the FDA, drugs.com, Wikipedia, etc. is publically available on the web. However, it is difficult to manage since the sources of semi-structured data are being

[1] http://healthline.com
[2] http://www.drugs.com

frequently updated and can be enriched by adding more datasets (e.g. LinkedData[3] and LODD[4]). There are even more difficulties with unstructured data. Unstructured data requires an additional step of transforming into a semi-structured format such as content analysis. The content analysis includes stemming, feature extraction, entity extraction and entity resolution. Search techniques include fuzzy search, faceted search, federated search and ranking. Medical knowledge is difficult to be centralized stored. Diverse data models, scale problems and legal issues prevent the simple solution of a centralized database. Thus, more solutions that are complex should be considered.

2.1 System Overview

In Medics system, we leveraged data mashup technology to aggregate ADE knowledge and provide personalized contents to the users. Using this technology, ADE knowledge is gathered efficiently when the above technical issues are resolved. The following sections describe three main components of the system.

External ADE Knowledge Discovery

Fig. 1 illustrates the high-level architecture of the Medics external ADE knowledge discovery component. The purpose of this component is to utilize publically available external web knowledge bases and provide access to ADE related knowledge using semantic query language called SPARQL. This component augments the existing Medics ADE knowledge by discovering new potential ADEs and ultimately improves Medics ADE detection capabilities.

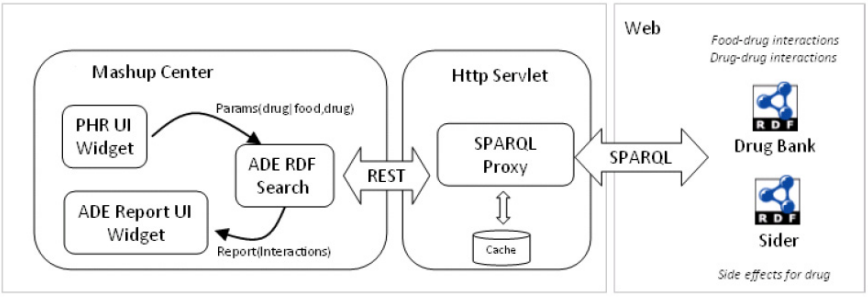

Fig. 1. Overview of external ADE knowledge discovery component

Search over External Text Sources

This component provides generic and flexible infrastructure for development of search-based services. Examples of search-based services are content discovery and recommendation. Fig. 2 depict the high-level system architecture of this component., and the details are presented with three features in the following.

[3] http://linkeddata.org
[4] http://esw.w3.org/HCLSIG/LODD

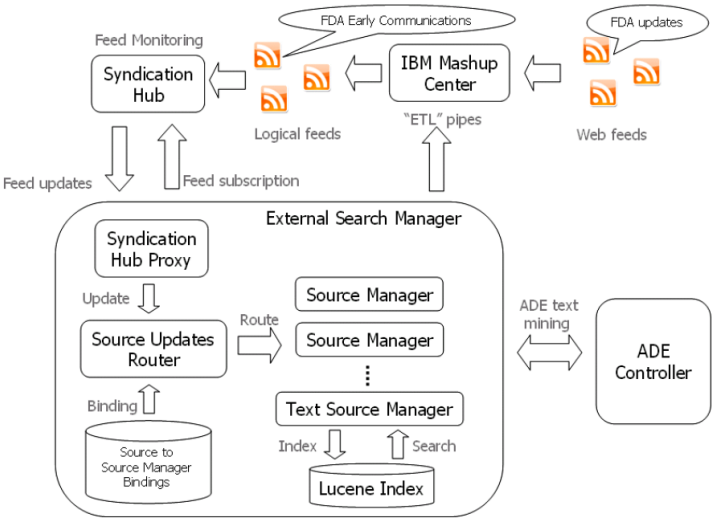

Fig. 2. System architecture for drug knowledge integration

Data gathering

Data is gathered using feed technology (e.g., RSS, ATOM). Feeds are monitored using feed aggregator called "Syndication Hub" which is capable of monitoring multiple feed feeding and timely delivering updates. Fig. 3 further illustrates one of such data mashups that is used to process and extract relevant ADE updates aggregated from both MedWatch[5] and FDA news feeds[6].

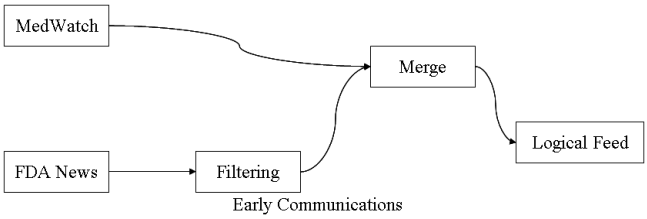

Fig. 3. Example of data mashup based on Mashup Hub

Source updates processing

Source update is processed by the source manager. Each manager can have its own processing logic. For instance, text source manager builds a textual document representation of the update text as well as facets and index the update in Lucene[7] index.

[5] http://www.fda.gov/safety/medwatch/default.htm
[6] http://www.fdanews.com
[7] http://lucene.apache.org/java/docs/

Search
The external search manager allows submitting text queries to any of the sources. Query results from different sources are further aggregated using federated search techniques.

Personalized Recommendations
This component provides personalized content recommendations based on patient's medical records. The medical contents of FDA early communications are recommended in a personalized view according to medications that patient prescribes. Medication names are highlighted yellow to increase ability in identifying the related information from the feed news.

3 Experimental Scenarios

To prove how the data mashup can be leveraged in the real world we developed a system that uses the mashup technology to support several scenarios related to the ADE detection services. The system is based on a commercial product of IBM called Mashup Center. In the first scenario, we test the searching of medical information of FDA Early Communication alerts according to patient's medications. The user or patient manages the personal health record in the PHR UI widget which is shown in left upper-side widget of Fig. 4. Information includes list of medications, allergies, immunizations, social history, etc. Once the user enters the list of prescribed medications or imported it from an external system, he or she can interactively select one medication and get in response a list of relevant FDA Early Communication alerts that will be presented in the left lower-side widget of Fig. 4. The query behind the scenes is processed over dynamic data extracted from trusted resources, like Adverse Event Reporting Program, FDA News and the MedWatch site. These resources were collected by medical knowledge experts and stored in Medics system by intuitive tools that the platform provides. Due to this mechanism, a user acquires publically available medication knowledge in a more comprehensive way.

In the second scenario, as the right side widget of Fig. 4., we test how users benefit from fully personalized and customizable workspaces. Users have access to a collection of health widgets that can be dragged and dropped into their workspace to answer the unique needs of the user. Different types of users have various requirements and this kind of extremely customizable platform can fit the needs of each type.

We test a new customized user page, named Welcome, with five health widgets. The user can integrate widgets that wrap third-party existing applications such as the news feed from Korea Food and Drug Administration and the medical article composition site. User can also choose to show, hide, or remove widgets based on specific contexts. The process of creating this kind of page, from scratch, takes few minutes. All of the above actions for workspace personalization and customization can be done by users without any programming skills in a powerful yet intuitive way.

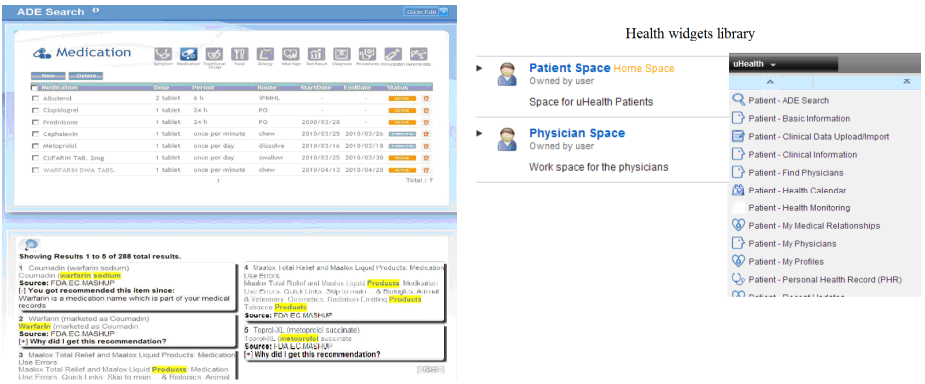

Fig. 4. Experimental Scenarios – health widgets

4 Conclusion

In this work, we prove the feasibility of using Mashup technology including Mashup Hub and Mashup Center to integrate the fusion of both structured data, which comes from the PHR system, and unstructured data, which is collected by monitoring and aggregating medical content in public resources. The main benefit is that this technology can be used by both patients and clinicians without programming experience to build a lightweight but powerful data mashup over the internet. It will accelerate the patient centric healthcare.

References

1. Belleau, F., Nolin, M., Tourigny, N., Rigault, P., Morissette, J.: Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. J. of Biomedical Informatics 41(5), 706–716 (2008)
2. Cheung, K., Yip, K.Y., Townsend, J.P., Scotch, M.: Methodological Review: HCLS 2.0/3.0: Health care and life sciences data mashup using Web 2.0/3.0. J. of Biomedical Informatics 41(5), 694–705 (2008)
3. Louie, B., Mork, P., Martin-Sanchez, F., Halevy, A., Tarczy-Hornoch, P.: Methodological Review: Data integration and genomic medicine. J. of Biomedical Informatics 40(1), 5–16 (2007)
4. Roitman, H., Carmel, D., Yom-Tov, E.: Maintaining dynamic channel profiles on the web. Proc. VLDB Endow. 1(1), 151–162 (2008)
5. Simmen, D.E., Reiss, F., Li, Y., Thalamati, S.: Enabling enterprise mashups over unstructured text feeds with InfoSphere MashupHub and SystemT. In: Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, SIGMOD 2009, pp. 1123–1126. ACM, New York (2009)
6. Zhou, X., Chen, S., Liu, B., Zhang, R., Wang, Y., Li, P., Guo, Y., Zhang, H., Gao, Z., Yan, X.: Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. Artificial Intelligence in Medicine 48(2), 139–152 (2010)