OpenEMPI: An Open Source Enterprise Master Patient Index

Odysseas Pentakalos, Ph.D., Yimin Xie
SYSNET International, Inc.
odysseas@sysnetint.com, yimin.xie@sysnetint.com

Abstract

The Open Source Enterprise Master Patient Index (OpenEMPI) software has undergone considerable enhancements and redesign following the initial contribution by the California Health Care Foundation to provide an extensible architecture that can incorporate various matching, blocking, and string matching algorithms. The new data model along with the loosely-coupled architecture of the underlying services make it suitable for integration into environments with demanding workloads. Furthermore, its proven compliance with the IHE profiles makes it easy to integrate into heterogeneous environments.

Introduction and Overview

OpenEMPI is an Open Source Enterprise Master Patient Index (EMPI) that originated from the remnants of the Care Data Exchange software that were turned over to the open source community after the Santa Barbara County (California) Care Data Exchange (SBCCDE) ceased operations[1]. Since it was originally released it has undergone considerable refactoring and redesign in an effort to achieve the following goals:

- The overall architecture of the software is now based on the Service Oriented Architecture principles where individual loosely coupled components form comprise the overall system and interaction among components only takes place through well defined interfaces.
- Utilizing the loose coupling between the components, the matching algorithm that is at the core of every MPI service is abstracted behind an interface. This allows for multiple, alternative algorithms to be implemented. The software distribution currently includes both a simple, deterministic algorithm as well as a probabilistic matching algorithm, which is an implementation of the Fellegi-Sunter algorithm that uses Expectation-Maximization (EM) for estimating the marginal probabilities of the model[2,3]. The flexible architecture allow for additional algorithms to be developed and plugged into the service through configuration.
- In order for OpenEMPI to be suitable for implementation in an large-scale environment, it needs to be able to sustain intense workloads. The data model was redesigned with performance and extensibility as key design criteria.
- Other services that provide extensibility to the architecture are the Blocking Service, which reduces the number of record pairs that need to be compared for matching purposes, the String Comparison Service, which determines the measure of similarity between two strings, and the Standardization Service, which transforms patient attributes to improve matching performance.
- OpenEMPI during the Google Summer of Code program was integrated with OpenMRS to form a patient registry for a federation of health care providers using OpenMRS as their EMR. OpenEMPI provides record locator services for the federation of sites.

Software Availability

The software is available and supported through the Open Health Tools community at https://openhie.projects.openhealthtools.org and through http://openempi.kenai.com

Software Architecture

After undergoing considerable redesign and refactoring, the architecture of OpenEMPI is now based on the principles of Service Oriented Architecture (SOA) consisting of a number of independent services that are loosely coupled from one another. Each service is exposed to other services through a well defined interface. Figure 1 illustrates the high level architecture of OpenEMPI bringing out the multi-layered architecture of the software. The diagram of Figure 1 is not an exhaustive list of all the components that comprise the system but only shows a subset of them for the purpose of illustrating the layered architecture of the system and to bring out the loose coupling between the services.

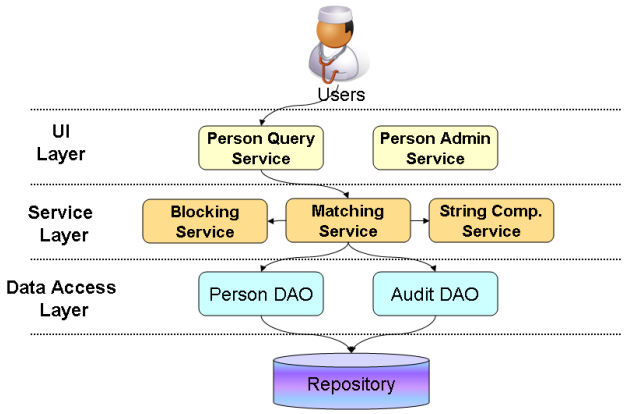

Figure 1. The architecture of OpenEMPI utilizes the principles of Service Oriented Architecture and Layering

Extensible through Pluggable Components

Another benefit of the service oriented architecture of OpenEMPI is that since the key components like the matching and blocking algorithms are only exposed to the other services of the system through a well defined interface, alternate implementations of those components can be implemented and plugged in independently of the rest of the system. This capability is utilized in the implementation of matching algorithms. Currently OpenEMPI provides two distinct implementation of the matching algorithm. A deterministic exact version and a probabilistic one based on the Fellegi-Sunter algorithm. Depending of your requirements you can choose which version you prefer to use by simple changing the Spring configuration of the system. Figure 2 illustrates the pluggable architecture of the system.

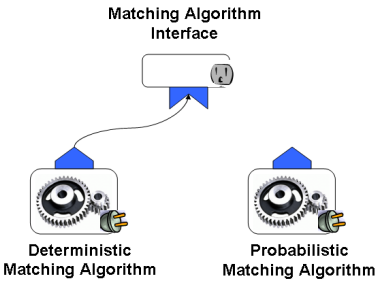

Figure 2. OpenEMPI can be easily extended by replacing one implementation of a component with another

Support for IHE PIX/PDQ Profiles

OpenEMPI is standards based and provides accessibility through the IHE PIX/PDQ profiles. Using an adapter it forms the core to the OpenPIXPDQ open source implementation of the PIX/PDQ profiles.

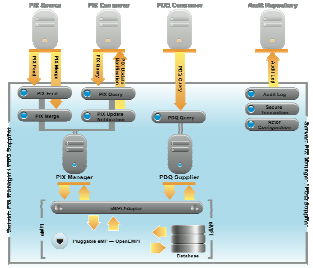

Figure 2. OpenEMPI is accessible through the IHE PIX/PDQ profiles

Future Projects

- We are currently working on integrating OpenEMPI as an alternate implementation of the MPI service in the adapter of the FHA Connect Gateway.
- Development of a user-friendly GUI for administration and query capabilities against the patient repository.
- Development of an data loader that simplifies the task of importing data into the patient repository.

References

1. Fellegi, I. P. and Sunter, A. B., "A Theory for Record Linkage", Journal of the American Statistical Association, 64, 1183-1210, 1969.
2. Winkler, E. W, "The State of Record Linkage and Current Research Problems", Statistical Research Division, US Bureau of the Census, 1999.
3. Cohen, W., Ravikumkar, P., and Fienberg, S., "A Comparison of String Metrics for Matching Names and Records", KDD Workshop on Data Cleaning and Object Consolidation, 2003.
4. https://openhie.projects.openhealthtools.org

Acknowledgments

We would like to thank the California Health Care Foundation and specifically Jonah Frohlich for their financial support of this effort.

Contact information

Odysseas Pentakalos, Ph.D.
Chief Technology Officer
SYSNET International, Inc.
odysseas@sysnetint.com

2930 Oak Shadow Drive
Oak Hill Virginia 20171
T: 703.855.2029
F: 703.716.9753