Abstract

MedfoLink is a new software technology that applies novel design features to help solve issues regarding medical records processing that are part of the national agenda. This software overcomes the vocabulary and performance limitations of existing medical language processing technologies by employing high speed databases, Java, and the UMLS (Unified Medical Language Source). MedfoLink is a Java technology that enables a computer to accurately record and interpret data from patient records. MedfoLink’s design incorporates two novel approaches: (a) intelligent algorithms that with practice improve comprehension of medical records and (b) the ALI (Adaptive Learning Interface) architecture that enhances performance by integrating existing technologies with an object-oriented approach. Future applications of MedfoLink range from individual patient care, to clinical drug trials, to public health monitoring.

1. Introduction

Medical reports are generated for every hospital visit. These reports contain information in the form of text and charts that convey the condition of the patient. The text in medical reports can be utilized as the raw data for a software system that acts as a diagnostic tool for patient health. When reports are mass processed for a large patient group, vital statistics may be produced about the condition of the population and the spread of illness.

Automated medical report processing has wide-ranging applications, from clinical drug trials, to individual patient care, to public health monitoring. The automated analysis of medical reports may be accomplished through NLP (Natural Language Processing). The computer analysis of human language has benefited many fields. However, the application of NLP to the text of medical reports (medical language processing, MLP) presents many difficulties. First, physicians, clerks, and technicians write medical reports using a vast range of specialized vocabulary. Second, an extremely secure system is necessary in order to protect patient privacy. Third, medical language processors must be extremely accurate, as errors may decrease the quality of medical care.

Over the past sixteen months, our team at Columbia University has been developing a successful model for MLP. We are in the process of coding and testing MedfoLink, a patent-pending medical language processing system that will have striking advantages over current systems.

2. Bridging the Gap

Currently, medical language processors do not adequately address the challenges of medical report analysis. First, medical language processors lack the vast specialized vocabulary needed to understand medical text. Second, many medical language processing systems contain legacy components designed in declining programming languages vulnerable to security breaches. Third, medical language processors lack object-oriented standardization of components and normalized vocabularies.

The MedfoLink model has a Java core that intelligently integrates existing technologies in an object-oriented approach to provide the solutions to these issues. MedfoLink implements the latest Java technology to enable cross-platform, secure, high-speed medical language processing. The MedfoLink system acquires the technical vocabulary necessary to process medical reports from a relational database of medical terms, the National Library of Medicine’s UMLS (Unified Medical Language Source) database. An emerging resource in the medical language processing field, the UMLS is a complex relational database populated with over two million medical and standard language terms. By connecting to the most recent version of the UMLS Knowledge Server online in real-time, MedfoLink is able to overcome the vocabulary limitations of current medical language processing systems. Unique aspects of the UMLS database make it a powerful tool for natural language processing. The UMLS runs on a high-speed Oracle relational database. The database logically organizes all
2.2 million terms into approximately 875,000 unique concepts complete with lexical and relational information. The UMLS database also has hierarchical structures mapping concepts by semantic types and ancestry. We have developed a functional prototype that incorporates the UMLS into the MedfoLink Java core. MedfoLink accesses the UMLS Knowledge Server in real-time as needed.

In order to incorporate the UMLS into the MedfoLink system, we adapted the UMLS Java API (Application Programming Interface) to extract the data required by the natural language processor more efficiently. Uniquely designed algorithms intelligently traverse the UMLS database structure and extract the necessary lexical and relational data in real-time. Using object-oriented programming along with the JDBC (Java Database Connectivity) protocol, we integrated the UMLS Java API with the MedfoLink core and MedfoLink’s Adaptable Lexical Database.

The synergy of a medical reports database with the MedfoLink integrator, custom Java API, UMLS, and Adaptable Lexical Database provides MedfoLink with the ability to accurately process medical reports. With the integration of the aforementioned modules, we are able to use the content of the medical reports themselves to automatically populate the MedfoLink Adaptable Lexical Database. The manual population of medical vocabulary databases takes thousands of hours and requires intricate knowledge of medical terminology. Experts may take years to populate a database of sufficient content for accurate medical language processing of reports such as discharge summaries. ALI can create the lexical database needed for medical language processing in a matter of minutes through automated analysis of medical reports and querying the UMLS. MedfoLink’s ALI architecture allows it to understand an expansive list of terms, enabling it to process medical reports more accurately than previous MLPs.

ALI’s innovative architecture provides MedfoLink with advantages over current medical language processors. We have created a new paradigm by applying object-oriented programming techniques to medical language processing. ALI contains distinct modules with well defined interfaces. Thus, the code of each module may be changed as long as the interfaces remain the same. In the ALI architecture, one NLP may be interchanged with another to enable highly specific processing of different types of medical reports. To date, medical report processing systems have been designed for single NLPs that output to simple database structures. These medical report processing systems support the storage and analysis of data that compose only a small fraction of the information contained within the original medical reports. The ALI architecture will support sophisticated database management systems that will allow for the structured storage and optimized analysis of greater amounts of medical data. By allowing for the analysis of more raw data, the ALI architecture may allow for the identification of more trends and factors than other MLPs. Thus ALI has the potential to deliver more comprehensive statistics to health care professionals, enabling them to provide better care for the patient population.

MedfoLink is under development. We are currently testing MedfoLink with a natural language processor specific to the analysis of radiology reports. We are running a prototype of the system at the Columbia Presbyterian Hospital on the Windows platform. We are also planning to test MedfoLink on the UNIX, Linux, and Macintosh platforms.

3. Future Applications

In the future, MedfoLink may find important applications for public health monitoring and clinical drug trials. MedfoLink’s potential capacity for flexible report processing will make it a powerful system for generating statistics on patient populations. Both clinical drug trials and public health monitoring require accurate statistics on large groups of patients. These statistics facilitate the identification of trends with important implications for patient health.

MedfoLink might also provide real-time monitoring of public health by rapidly analyzing all medical records generated in a region. For example, MedfoLink could be used to identify a significant incidence of illnesses that may mark the early stages of an epidemic. MedfoLink may improve the drug development process by enhancing the detection of adverse side effects during clinical drug trials through a potentially more powerful analysis of patient data. We are developing MedfoLink as an information tool that may enable healthcare professionals to deliver better patient care and improve lives.