6.191 Proposal Outline

December 6, 2001

Implementing Database Solutions for the W3 Electronic Medical Record System

Clinical Decision Making Group

Laboratory of Computer Science

Patrick Cody

Prof. Peter Szolovits

Abstract

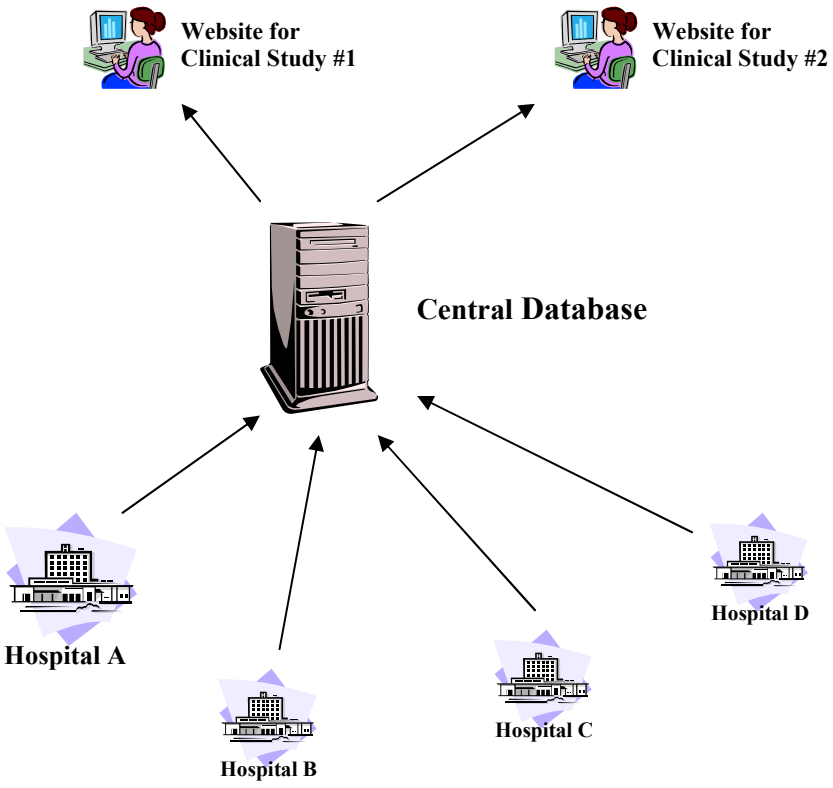
The Clinical Decision Making Group (CDM Group) recently completed a prototype of the World-Wide Web Electronic Medical Record System (W3-EMRS). This system is intended to facilitate the sharing of patient data between hospitals and clinical studies by providing a central and universal mechanism. The prototype system encompasses a variety of web-development and database components, whose capabilities need to be altered to reflect feedback from clinical study groups. The goal of this thesis proposal is to redesign and build a new database mechanism for EMRS system. The process of updating the existing system will include modularizing the web linkages, changing the database identity anonymization mechanism, and implementing improved tools for data error correction between Hospitals and study groups.

Background

The Clinical Decision Making Group at the MIT Laboratory for Computer Science is a research group dedicated and furthering the application of technology and artificial intelligence to clinical situations. Given the need for accurate and timely information to support clinical research projects, the group is involved in several projects the gathering, availability, security and use of medical information for use by research organizations.

Responding to a need from the medical research sector, the CDMG group designed and built a centralized database system for data sharing in clinical studies (the W3-EMRS project). Using this system, various hospitals could upload patient data to this database for use in clinical studies with sufficient assurances of patient privacy and authenticity. The goal was to facilitate and simplify the process of sharing medical information for clinical research, benefiting both hospitals and research groups by creating a standard transmission mechanism for dealing with the various Hospital databases.

Figure 1: The W3 EMRS architecture



The existing infrastructure (shown in figure 1) works along a simple process. Hospitals that are participating in clinical studies upload portions of their patient database into a central server, which then reversibly blinds the patient data, formats it into a standardized format, and makes this data available on an authenticated web-site for use by clinical study groups. Information is standardized in such a way that patient data cannot be correlated with the hospital, although clinical groups can request additional information through the central system from participating hospitals.

Problems with existing system

Feedback from clinical research groups that have examined this system reveals two major problems with the current design. These problems make the system unusable to for research purposes in its current form, and therefore should be corrected as soon as possible.

A major unanticipated problem with combining patient data from different hospitals is the duplication of individuals within study groups. Most clinical studies collect data over long periods of time, in which a patient involved in the study may have moved or been treated at another hospital. The movement of patients between hospitals may cause their clinical data to appear multiple times within a clinical study. Additionally, uncaught errors in a hospital's record keeping may create multiple instances of single patient within that hospital's database. The anonymization features of the CDMG's system prevent these problems from being solved in the combined database design.

The second major problem with the existing system is difficulties in communication between hospital and study groups. The current system applies a standardized formatting to hospital data so that the only way a clinical group can get additional information about specific patients is to contact the database system to get information about which hospital specific patients came from, then contact that hospital directly. This anonymization and communication mechanism has proved unwieldy, and both the hospitals and clinical groups would prefer direct communication without the database system being involved. Also, since different hospitals use different database

systems and accuracy measurement mechanisms, data from different hospitals often fails to standardize.

The transmission mechanism itself has some minor quirks which need to be corrected as well. While functional, the update mechanism into the SQL database server uses ssl http to transport large amounts of data, which has proven to be cumbersomely slow and problematic. In the process of changing the system design, minor issues will also be corrected.

Proposed Project

My proposed thesis project is to design and implement a new version of the W3 EMRS database backend that will satisfactorily resolve the stated problems with the current system. Since the new database backend would have significantly redesigned security and anonymization features, the security of the new design will have to be reviewed by the CDM Group and the clinical research entities. The design and review of the revised system will be conducted by the entire CDM group, including the designers of the original system, Min Wu and Prof. Peter Szolovits.

Once the design has been finalized, this MEng project will then be to build the database backend and the linkages into the rest of the system. Redevelopment of other parts will proceed simultaneously. The process of building a working prototype of the system will likely take several months and involve a significant amount of SQL database work, web programming, and implementing commercial data analysis packages. A firm description of the final system is not possible at this time, because the system is in the process of being redesigned.

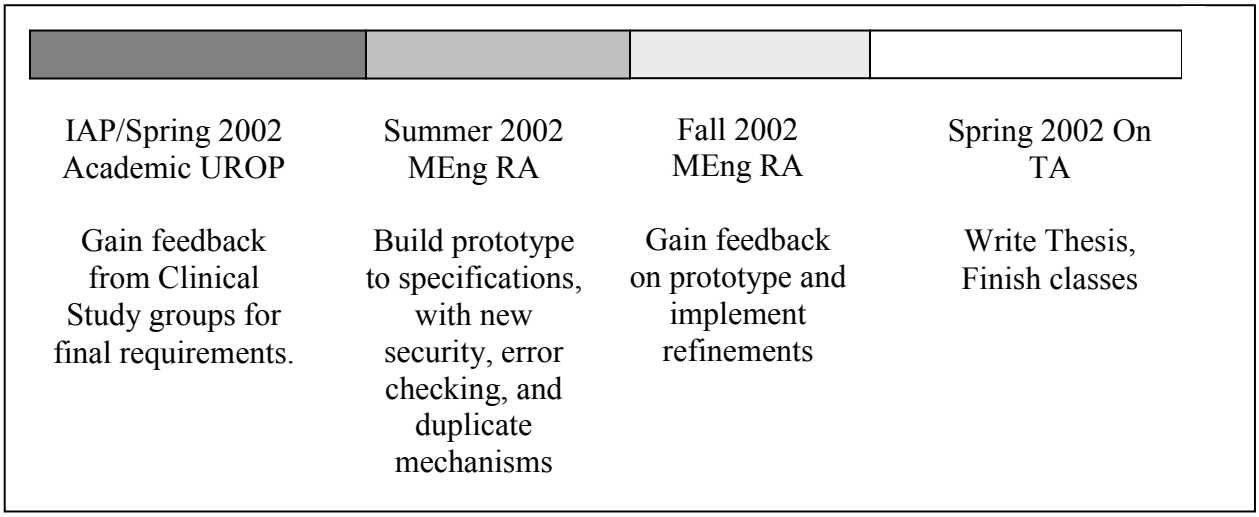
Specific Technologies

The E3 EMRS is built principally upon JSP and SQL, both of which will be used in the redeveloped version. The process of changing the underlying backend will necessitate changing the JSP-based front-end web page. The server itself will continue to use MS SQL 2000 for the database purposes, with Standard Query Language and JDBC acting as the interface between the web page and database. Data transportation is currently accomplished using SSL and Certificate authentication, although this may

change with the new design. I have previous work and UROP experience with all of these technologies.

The tools used in the duplicate identity checking system and privacy mechanisms are yet to be determined. Duplicate checking will likely be accomplished using a heuristic metric to mark records for human review. Patient privacy may be accomplished using unique individual identifiers or cryptographic techniques.

Timeline



Resources

This project relies heavily on input and feedback from medical research groups. The CDM Group already has contacts within the American Medical Informatics Association (AMIA) and the clinical study groups. A large body of published research exists on standardized medical electronic recordkeeping and database merging.

My proposed thesis project requires a minimal budget beyond tuition and stipend requirements. The principle tools this project would use, servers and licensed database software, have already been acquired by the CDM Group. Almost all of the work and review will be done at MIT LCS on computer and through the web. Minor

communication and travel expenses may be required for demonstration and gathering feedback of the revised system to hospitals and clinical study groups.

Risks

Pattern matching systems and electronic medical record systems recently received a great deal of commercial attention. Some commercial web projects are attempting to combine medical records online, with the rise of several prominent dot-com efforts. As such, some of the recent research into medical databases is protected by patents and prohibitive licensing requirements. While clinical study information sharing is considered too small of a niche for commercial companies, the possibility of commercial competition for this system does exist.

This thesis relies upon feedback from clinical research groups as a basis for its design goals. Difficulties in gathering additional feedback, while not crippling to the project, would make it more difficult to satisfy the project's overall goal of creating a useful mechanism to facilitate clinical research.