Background

The Department of Veterans Affairs provides healthcare, benefits, and burial services for American veterans. The VA VISN 2 Center of Excellence for Suicide Prevention is among the foremost federal research facilities in the area of suicide prevention. The CoE conducts epidemiological and clinical research relating to suicide prevention, and offers training and community outreach to first responders and practitioners.

The CoE has many paper forms and artifacts related to researchers and research programs. In order to simplify and streamline their process, they would like to digitize their ‘system’ and have a central online service to perform many of the functions that they currently do on paper or with Excel spreadsheets. The desired system would include the ability to group research artifacts into “programs,” update research grant status, and allow the project to be audited easily.

Higher-tier functionality for the system would be the ability to maintain a separate web crawler application which would be able to run either manually or automatically. The intention is to search the web for other research that cites CoE research as a source, so that the VA is able to see its research actually being read and put to use.

Process

The process that was used is a combination of evolutionary prototyping and the spiral model. Each spiral begins by identifying the win conditions of all success-critical stakeholders, then identifying alternate ways to satisfy the win conditions. After this, risks are identified and resolved, and then approval is obtained for the current spiral as well as a commitment to the next. Effort spent on the identified tasks is determined by the risk assessed in the planning stages. At the end of each spiral, a working prototype is delivered, and is reused and altered in making the next prototype. Specifically, we chose to release each prototype to the customer, receive feedback on each release and then return to the beginning of the spiral and perform these evaluations based on the customer’s feedback.

Future Work

The majority of the remaining work is centered around the web crawler. We have created a modular framework that can receive code modules for desired scholarly databases and academic websites. As there is a large variety of sites available, we have only implemented one module for Pennsylvania State University’s own open-source database, CiteSeerX. This serves as a functional example, able to provide both desired output and guidance in creating modules for other sites.

Due to difficulties in scholarly information gathering, primarily in the heavily-copyrighted nature of journal metadata, building a fully-functional citation-gathering web crawler is a hard problem to solve. Our architecture is structured to support the addition of data-mining modules - blocks of code dedicated to scraping particular sources of information - which would simplify the problem significantly if the legal status of journals and their metadata were to change.

Technologies

The technologies used in this project include .NET, C#, and SQL Server. Additionally, we utilized jQuery and Bootstrap for the client-side development, and MVC3 for the server-side development.