National Center for Genome Resources

Business

The U.S. Human Genome Project was initiated by the Department of Energy and the National Institutes of Health in 1987 as a multi-disciplinary effort to understand the basis of human heredity. The specific focus of the project is the characterization of the human genome - the complete collection of human genetic material, including the estimated 50,000 - 100,000 genes contained in human DNA. This data is utilized by pharmaceutical companies, other government agencies, academic institutions and any researcher working in the field.

The results of the Human Genome Project will enable the creation of a new "molecular medicine" based on early detection and treatment of more than 4,000 diseases of genetic origin, effective preventative medicine, efficient drug development and personalized therapies.

Challenge

A significant portion of the information generated by the Human Genome Project is stored in the Genome Sequence Data Base (GSDB). It was initially developed at Los Alamos National Laboratory and in July, 1994 was moved to the NCGR in Santa Fe, New Mexico. Since the creation of the data base in 1979, the volume of data has been increasing exponentially, currently doubling every 18 months. Due to this massive increase in data, the need to track the information electronically was identified versus the traditional technique of paper journalizing. The database is composed of DNA sequences and annotation on the sequences that point out areas of biological significance.

Linda Smyth was hired as the Software Coordinator to manage all the software activities within the Genome Sequence Data Base Project. In the past, as software tools were being built for the data base, there was no process incorporated to version track what was being done, when and by whom, nor any bug tracking mechanism. Based on her 13+ years in the software industry, she recognized the need for a process to manage the bug tracking of the issues and traceability of linking issues and reasons for code changes.

Solution

After identifying what they needed and looking at what solutions were offered in the configuration management marketplace, the Razor product from Visible Systems was selected. Razor combines the functionality of

201 Spring Street • Lexington, MA 02421
(800) 6VISIBLE • Fa x (781) 778-0208
Email: sales@visible.com • www.visible.com

Visible
SYSTEMS CORPORATION

bug tracking and version control into one affordably priced product with the added bonus of an easy to use interface.

Over the last few months, Razor has been incorporated into a variety of software projects. It has been noted that efficiency, organization and accountability has increased significantly on these projects. It is still a cultural change for many of the software developers to use a software tool like Razor, but they are beginning to see the value in version control, for example being able to back track on the bug fixes.

The ease of use and short learning curve with Razor enables the software developers to access the necessary data more quickly. The initial benefits of a CM tool like Razor will come with the release of the GSDB Annotator in 3rd quarter 1995.

The GSDB Annotator will be used by molecular biologists in a variety of biotech industries and academic research settings to submit DNA sequence data and annotation to the database, as well as edit sequences already existing in the database. This tool will also make it easy to locate sequences, browse the data as well as incorporate findings from other prediction analysis tools in to the annotation.

201 Spring Street • Lexington, MA 02421
(800) 6VISIBLE • Fa x (781) 778-0208
Email: sales@visible.com • www.visible.com

Visible
SYSTEMS CORPORATION