AC 2007-1538: VISIT - VISUALIZATION AND INFORMATION TECHNOLOGY - A MULTI-TIER SYSTEM FOR INTERDISCIPLINARY EXPERIENCES IN DATA COLLECTION AND VISUALIZATION

Judith Challinger, California State University, Chico

Rachael Teasdale, California State University, Chico

VisIT - Visualization and Information Technology - A Multi-tier System for Interdisciplinary Experiences in Data Collection and Visualization

Abstract

Collaboration on interdisciplinary teams is an important experience for computer science students, and one that is too rarely available. Methods for data collection in the field and subsequent graphing and analysis are important skills for students in many different scientific disciplines. This paper presents a project currently underway by computer science and environmental science faculty at the California State University, Chico to address these needs. A multi-tier system is being developed to serve as a repository for data collected in the field by environmental science students, facilitate graphing and analysis of data, and provide a platform for interdisciplinary teams of students to collaborate on both software development and data analysis tasks.

Introduction

In 2005, Microsoft Research Cambridge brought together an international panel of scientists for the *Towards 2020 Science* workshop.[1] Their goal was to explore the increasing synergy between computing and the sciences, and to map out the goals and challenges ahead in the next fifteen years. One key challenge identified is the "need for better collaboration between computer scientists and other scientists…"

In order to meet this challenge, scientists must be exposed to interdisciplinary experiences during their formative years. The approach taken should benefit both the students of the computing sciences, and the students of the discipline with which they are collaborating. A Web-based system currently under development is intended to serve as a unifying focus for these interdisciplinary experiences.

VisIT, short for Visualization and Information Technology, is a multi-tier software platform that supports the establishment of projects with associated data dictionaries and graphing profiles. A pluggable architecture allows the addition of customized software for data analysis and visualization. The approach is Web-based and consists of a Linux server running the Apache HTTP server, Apache Tomcat servlet container, and MySQL. Java Servlet technology is used to provide the interface necessary to set up data repository and graphing projects. Users access the system using a Web browser. Plans to add support for mobile devices utilizing Java ME are underway. See figure 1 for a deployment diagram.

Pedagogical Applications

Our long-term goal is to support a variety of learning paradigms with VisIT. During the spring 2006 semester a senior-level computer science student built a simple prototype demonstrating the Web-based interface to a database and visualization framework. This prototype served as the basis of discussion among faculty of multiple disciplines. Three separate approaches were identified: the use of technology such as VisIT in a lower-division natural science course, the study of the architecture of a system such as VisIT by upper-division computer science students, and the extension of the capabilities and application of a system such as VisIT by interdisciplinary teams of upper-division students. All three approaches are being explored.

Use In General Education Courses

A more complete implementation of VisIT is now under development and will be utilized during the spring 2007 semester by lower-division students taking a General Education course in the Department of Geological and Environmental Science. Students in this GE course spend some part of the semester learning field methods for data collection, and techniques for the subsequent analysis of the data collected. One task they are assigned is to take measurements and samples from local rivers and streams. They will enter the data they collect into VisIT and use the graphing facilities to generate various types of graphs in order to analyze and visualize their data.

This "Rivers and Streams" project using VisIT will enhance the student experience by integrating technology and quantitative skills with fundamental principles of geology. Previously, students were introduced to geologic concepts related to stream flow characteristics by measuring stream parameters (width, depth, flow velocity) and then calculating simple properties such as discharge (volumetric flow rate) and sediment load capacities. In doing so, students determined current stream flow values, but had no appreciation for the contextual significance related to seasonal variations or comparisons with other streams. With VisIT, students will have access to data sets collected at different locations and at different times.

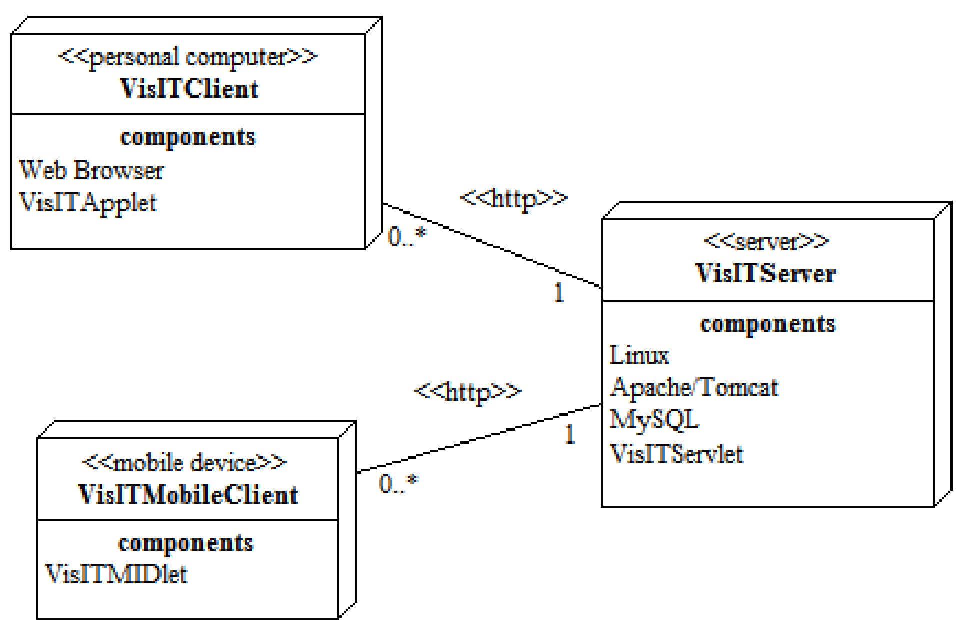


Figure 1 Deployment Diagram for VisIT

Students can use multiple data sets integrated with their own data, for activities such as the construction and interpretation of graphs and description of stream flow processes in a larger context. Once graphs are constructed, students can use their hydrographs (discharge vs. time) to detect any variation in flow rates (e.g. rainfall events) within the short term time frame of the semester in which they take the course, and potentially over longer periods of time (using data collected in previous semesters). VisIT allows more sophisticated interpretations of data sets and provides a platform for students to develop and improve their quantitative skills through the use of technology, both of which are programmatic goals of the physical science general education required coursework at CSU, Chico.

An additional benefit the application of technology such as VisIT provides is that the data collected by students each term will be available for other students to use in future sections of this course. Academic scheduling currently dictates that data is collected in August, November, January, and May, which allows for easy detection and interpretation of seasonal variations of stream flow data. With continued use of VisIT, a long-term database of flow properties of Big Chico Creek on the CSU, Chico campus will be established. Data collected at the on-campus site can also be compared to data collected upstream at the BIC^2 site maintained by the California Department of Water Resources. Data at this station is collected using automated techniques, which allows students to consider not only data variation in time and by location along the stream, but also in terms of collection techniques. The automated upstream site is also upstream of a diversion channel, which facilitates consideration and discussion of engineering techniques related to rivers and streams during flood stages.

In using VisIT, students will have access to new technologies (eventually hand held, with GPS capabilities) to learn about geologic processes. In doing so, they will learn quantitative skills of data collection, management, graphing, interpretation of multiple data sets, and analysis of multiple variables in using their data to understand a variety of stream flow processes. This provides a diverse array of ancillary skills building as well as provides students the opportunity to participate in all stages of the scientific method. The “Rivers and Streams” project will introduce lower-division science students to the power of computing technology in science, as well as getting them more comfortable using technology, thus setting the stage for later collaborations.

Use In Computer Science Courses

An upper-division elective course in the fundamentals of programming for the Web will give interested computer science students the opportunity to examine a fully deployed and actively used multi-tier Web application. Students in this course will examine the implementation of VisIT in detail. Projects will be undertaken for further development of VisIT. One project on the drawing board is the addition of tutorial modules for specific areas of scientific application. Such projects will necessitate interaction on the part of the computer science students with scientists of these other disciplines.

Another upper-division elective covers mobile/wireless programming. Students in this course may elect to investigate mobile applications that interface with and make use of VisIT. Of particular interest are data entry and use of GPS enabled devices.

Interdisciplinary Use

Students in computer science are required to do a capstone project as a culminating activity. Students in most other science departments have the option of doing a senior project or independent study. Interested students will be encouraged to do an interdisciplinary senior project. Partnering with a student from another discipline, they will define a project, data dictionary, and graphing profiles, in VisIT. They will collect and enter data, develop custom software for their specific study needs, and generate visualizations specific to their project. In addition to fostering such projects, a potential new course that is interdisciplinary in nature is being investigated. In this upper-division course, interdisciplinary student teams would develop a project in VisIT that uses custom data analysis modules. Figure 2 presents potential use cases for VisIT. Anticipated as an added benefit of these activities is an increased level of interest on the part of the students - a feeling that they are doing something "real," and a feeling that they are contributing to both science and education in their community.

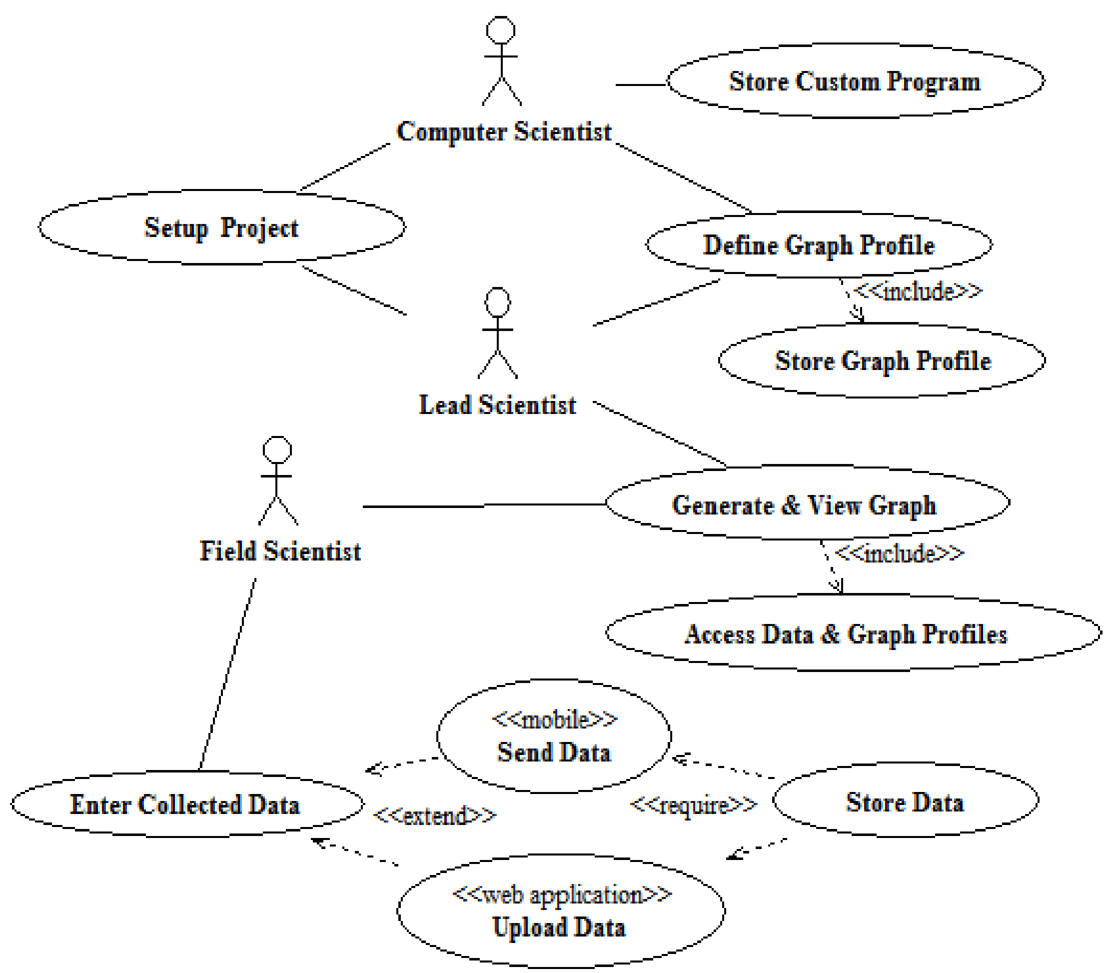


Figure 2 Some use cases for VisIT

Interdisciplinary teams are endorsed by the ACM Computing Curricula 2001 (CC2001).[3]

Issues and Requirements

Several issues and requirements present themselves when considering the architecture of a system such as VisIT. One complicating factor requiring attention is the desire to use VisIT in a teaching environment where many of the users are transient (students). They will need to enter data and create graphs during a single term. In the next term there will be a new set of users (students). The instructor teaching the course may need to determine which data was entered by a certain student, or whether all students have completed the assignment. In the interest of minimizing the administrative workload associated with providing VisIT for use in this environment, particular uses of VisIT will be identified as a Project with a Project Manager. A VisIT Manager will establish the project in VisIT. The Project Manager will have the capability to add Project Scientists. For a classroom application the Project Manager may be the instructor or a teaching assistant or both. There may be one Project Scientist user shared by all students in the class, or one per class section, or one per student, as needed by the specific application.

Another issue related to classroom use is the desire to build up over time a valid set of data relating to regional environmental issues of interest (such as the health of rivers and streams). In order to protect the integrity of the database, values entered by a Project Scientist (typically a student) will need to be validated by the Project Manager before it is permanently added to the database. Data that has not yet been validated will be available to Project Scientists and the Project Manager for analysis and visualization, but will not be publicly available. Data that has been validated and added permanently to the database will be available for retrieval, analysis, and visualization by all other Projects.

A third requirement relates to the goal of dissemination. The architecture should support the use of guest access for data retrieval and graphing. Anonymous entry of data may be supported for graphing purposes only; the data will not be added to the database and will only be available during the guest user's session.

System Design

A deployment diagram for VisIT was given in figure 1. The database back end is provided by MySQL. Server side programming is done using Java servlet technology, and deployed using the Apache Tomcat servlet container. For visualization, the servlets use jFreeChart to generate graphs of the data retrieved from the database. On the client side, any Web browser can be used to access VisIT. In addition, development of a Java MIDlet suite is planned for access from a mobile client platform.

To support administrative functions in VisIT, two tables are defined in the schema: one for users and one for projects. Users are uniquely identified by their login, and are also given certain privileges. A user with the privileges of a VisIT Manager can add projects and users with the privileges of Project Manager. The Project Manager will typically be the instructor and/or teaching assistant, or lead scientist on a research project. The Project Manager can add, edit, and delete graph profiles and custom software modules for a specific project, and may validate data entries. A Project Manager can also add users with the privilege level of Field Scientist. A Field Scientist can enter data for a specific project, and can utilize existing graph profiles and custom software modules to analyze and visualize data. A login with the privileges of Field Scientist will typically be used by students. To initiate a project, a user must first be added to the users table and given the privileges of a Project Manager. Then a project is added to the project table, identifying a user as its Project Manager. A project is identified by a unique id.

Locations where data are collected are specified by a unique id, given a textual name, and stored in the locations table. Optionally, a description may be provided, as well has the latitude and longitude of the location. When support for GPS-enabled mobile devices is incorporated, a specification such as the GPS Exchange Format (GPX)[4] will be investigated. Various types of measures are defined and stored in the measures table. A user with the privilege level of a Project Manager may enter new measures. Actual measurements are stored in the measurements table. A measurement is identified by a unique id. It has a location, a date and time of collection, and is associated with a specific project. An optional description may be given. Graph profiles are defined by the Project Manager and stored in the profiles table. A graph profile defines a specific type of chart or graph of selected data from the database. Images are also supported and may be stored in the images table. An image has a unique id, a description, and is associated with a project and a location. The database schema is shown in figure 3.

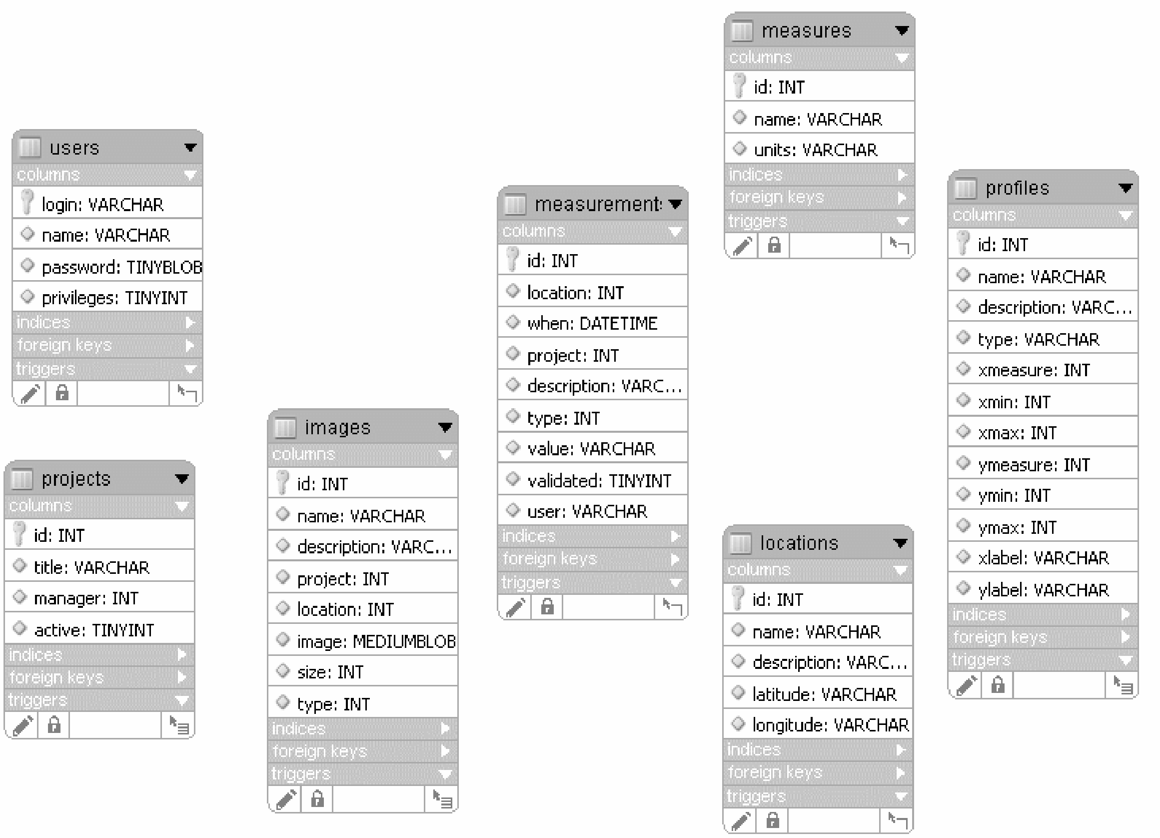
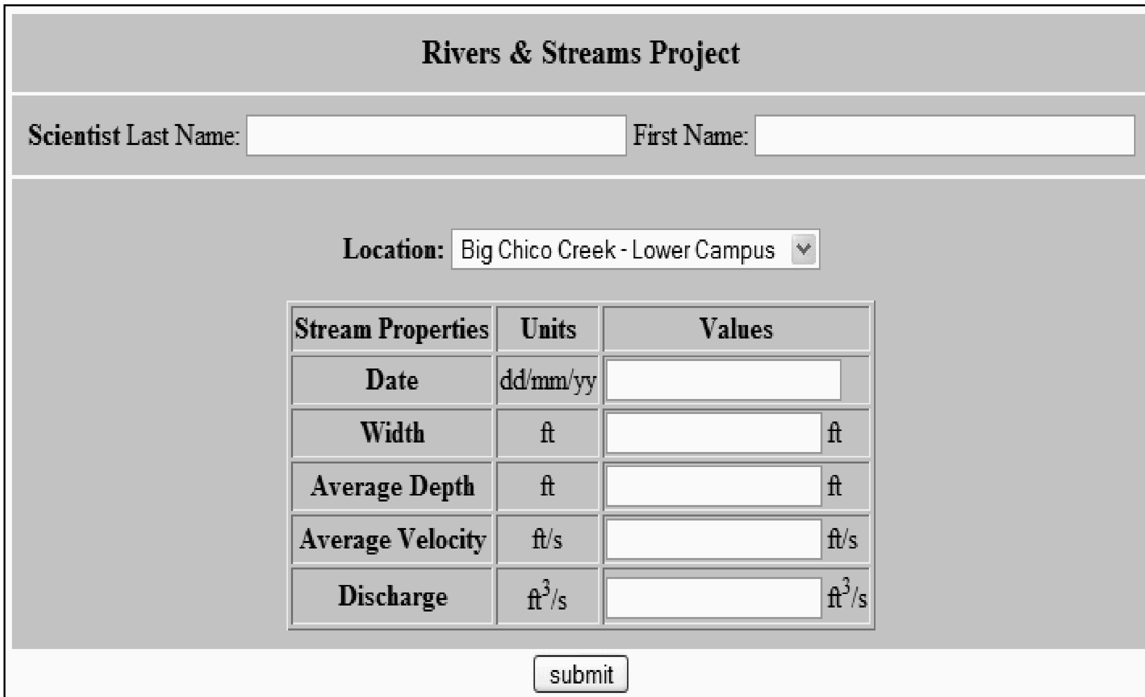


Figure 3 VisIT Database Schema

Web-based user interface pages available or planned include:

- Administrative screens for
  - User management
  - Project management
- Project management screens for
  - Locations
  - Measures
  - Graph profiles
  - Custom software
  - Data validation
  - Management of Field Scientists
- Screens utilized by Field Scientists
  - Data entry
  - Data analysis
  - Data charting and graphing
  - Uploading of images
- Screens for general use
  - Login
  - Help
  - Tutorials

Server-side generation of graphs is accomplished using an open source charting and graphics package, jFreeChart.[5] This library supports a wide variety of chart types and is suitable for use in a servlet environment. A data entry form is shown in figure 4.



Rivers & Streams Project

Scientist Last Name: ______ First Name: ______

Location: Big Chico Creek - Lower Campus

Stream Properties	Units	Values
Date	dd/mm/yy	
Width	ft	ft
Average Depth	ft	ft
Average Velocity	ft/s	ft/s
Discharge	ft^3/s	ft^3/s

submit

Figure 4 Data Entry Form

An example line graph is shown in figure 5. In this exercise the students using the Rivers & Streams project are comparing the stream velocity at three different locations and at three different times of the year. The slope of each line gives the stream velocity at that particular time and place.

Conclusions

We are optimistic about this project as a motivator for students of many disciplines, but especially for computer science students. The application of computer technology to scientific inquiry is an important and fascinating area of study. Opportunities for students to experience the stimulation of working on an interdisciplinary team will further engage them in their discipline, while making them aware of the needs of other disciplines. A meaningful application domain such as environmental studies provides added attraction and importance. VisIT is designed for use in a teaching environment in both the computing sciences and other scientific disciplines. It

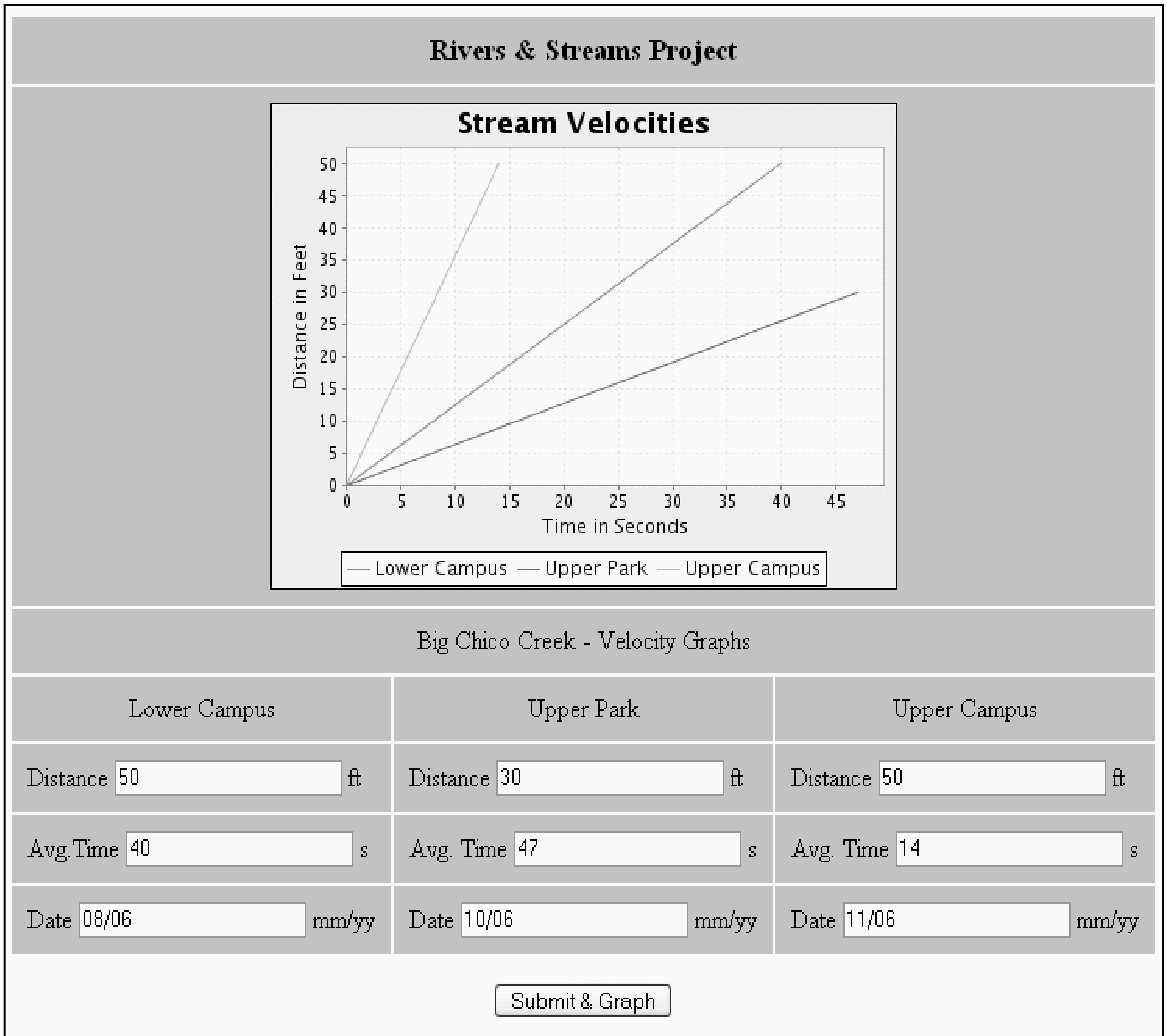


Figure 4 Sample Visualization

can be extended and customized by for a particular project by interdisciplinary teams of students. In addition, VisIT lends itself to dissemination to other institutions and promotes collaborative scientific studies.

Future Work

Enhancements planned for the next year include support for GPS-enabled mobile devices as a data entry mechanism, and the incorporation of publicly available data such as that provided by the California Department of Water Resources. In addition to continued development of VisIT as a platform for teaching and interdisciplinary research, an important focus of future work will be on the dissemination of this tool to other institutions. Instructors teaching lower-division GE science courses at other universities and community colleges will be invited to create a project in VisIT for their students. Future collaborations between institutions will increase the extent of the database, further enhancing the experiences and activities available for all students.

---

[1] http://research.microsoft.com/towards2020science/background_overview.htm
[2] http://cdec.water.ca.gov/cgi-progs/staMeta?station_id=BIC
[3] http://acm.org/education/curric_vols/cc2001.pdf
[4] http://www.topografix.com/gpx.asp
[5] http://www.jfree.org/jfreechart/