The following paper was originally published in the
Proceedings of the Fifth Annual Tcl/Tk Workshop
Boston, Massachusetts, July 1997

Using Tcl/Tk in Biology Research Application Development

Ellen R. Bergeman and Mark Graves
Department of Cell Biology, Baylor College of Medicine
Houston, TX

Using Tcl/Tk in Biology Research Application Development

Ellen R. Bergeman and Mark Graves

Department of Cell Biology, Baylor College of Medicine

One Baylor Plaza, Houston, TX 77030

phone: +1 (713) 798-8105; fax: +1(713) 798-3759

email: erb@bcm.tmc.edu

Biology research is an area which presents many problems for software application development. Biology application development occurs within a framework of rapidly changing needs; unspecified or incomplete requirements; short term versus long term needs; complex, often incompletely defined concepts; and multiple platforms with varying infrastructure. Biology researchers rely heavily on applications developed to analyze data, access databases, and store ongoing research data. Often the bottlenecks in research projects are not in the laboratory but instead are in the lack of software needed to handle the data.

Biology research changes rapidly and new laboratory techniques can be brought into a laboratory quickly, becoming integral elements in a very short time. Often a new technique is brought in because it will speed up the work in the lab, and therefore it changes the requirements for software support. For biology application developers, the need for an application is often not known until there is a bottleneck in the research process.

Another aspect of rapid change within biology research is that there are no clear-cut representations of the concepts within biology. The point of biology research is to explore unknown concepts to understand and clarify them. Application development should proceed in the same way as biology research so that applications are built to support current concepts and are allowed to evolve as research clarifies and changes the concepts.

We have found that Tcl and Tk are useful for solving some of the problems such as capturing complex concepts and improving communication between researcher and application developer. Complex concepts can be captured as data types which are initially developed in Tcl, then implemented in C as extensions to the Tcl language when they have been refined. Widgets which provide user-interface functionality for the data types can be implemented using TK. Using these data types and widgets allows for rapid development of small applications which can be used to improve communication between biologist and computer scientist.

A key to successful biology application development is having a framework in which a developer is able to try out different computer science solutions to the same problem without incurring substantial cost in development. Tcl and Tk provide a way for computer scientists to follow the same research paradigm as biologists: explore data representations for concepts as the concepts are evolving, while providing useful solutions to current software needs.

A major problem in biology software development is in the area of databases. It is difficult to design databases which capture the complex concepts and relationships of biology. Biology concepts have a graph-like structure, and we have developed a data model based on graphs which is useful for capturing biology data. Along with the data model, we have developed a graph database which stores and manipulates data as graphs. We have developed an extension to Tcl which includes the data types used in the graph database. These data types include an in-memory data storage graph and graph abstraction for complex queries. We have also developed widgets which provide graphical interfaces for the data types.