

COLLABORATIVE AUTHORIZING OF SCHOLARLY DOCUMENTS

organized by
Rob Beezer, Steve Blood, and David Farmer

Workshop Summary

Mathematics is increasingly collaborative, and one consequence is that mathematicians often write documents having multiple authors. Examples range from research papers involving two or more authors, to dynamic surveys and problem lists which are maintained by a group of people over time, to open source educational material such as an open textbook which could have many people contributing exercises and examples.

Currently the tools to facilitate collaborative authoring are inadequate. Small groups, authoring small documents, can circulate drafts via email, or use Dropbox, or Google Documents, or adapt a tool designed for open-source software. But even in the limiting case of two authors writing a short research note it is cumbersome to maintain multiple files, allow both authors to edit at the same time, and allow both authors easy access to the latest version.

The purpose of this workshop was to bring together mathematicians and people with experience in open source software and open source documents to chart a plan for developing a Collaborative Authoring Tool (CAT) that would address these issues. We found that the workshop was quite successful, determining the key issues, clarifying the current state of related projects, and identifying the exact nature of the missing piece which the CAT would provide. We elaborate on each point below.

The Issues

Workshop participants included research mathematicians who have written many collaborative papers and tried several methods for collaborative authoring: CVS (software version control), Dropbox, circulating files by email, GoogleDocs, wikis, keeping the files in a software repository, and others.

Participants also included several authors of open textbooks. This included successful authors with widely-used free textbooks, and well as new authors who are in various stages of writing such books.

These authors provided a variety of perspectives on their needs for collaborative authoring. Some of the less obvious issues were: the need for embedded video, the need to have sections of a book which an instructor can “hide” until a particular time in the semester, and the need to track dependencies between different parts of a document so that if a component is re-used, the necessary earlier parts can be identified.

Existing Efforts

Workshop participants were highly knowledgeable about existing efforts related to collaborative authoring. Examples include repositories (both for software and for documents) and browser-based editing systems. Much of the discussion was informed by the fact that the collaborative authoring problem has largely been solved for software. The conclusion

was reached, from several perspectives and for several different reasons, that existing software version control systems are not suitable for scholarly documents. Furthermore it is not feasible to modify such a system to work as the CAT.

The conversation struck a balance between stories about unsuccessful efforts and the lessons to be learned, and discussion about available resources and how they can be used. The key point, made repeatedly by people with software design experience, is that you should re-use whatever you can, and only create from scratch as a last resort. This applies to software, and also to data formats.

What is Needed?

It was established that there are places to store open documents (like Connexions at Rice), and formats for the storage and exchange of documents. But what is needed, and does not exist, is an interface for editing a collaborative document. That is the key piece which is missing and which the CAT should provide.

Specifically, the CAT needs to provide a way to display the document (in publication quality on the web), it needs to provide a way for a variety of people to edit the document, and it needs a way for the main author(s) to oversee the contributions of other people.

A lengthy discussion, which continued online after the workshop, is that there are a variety of possible users of the CAT, and it is not going to be possible to make everyone happy. For example, there is a disconnect between the user who just wants to write a brief classroom handout, and the user who is writing an NSF proposal or a textbook. In the latter case it is necessary to take time, even before much writing has been done, to specify the structure of the document. This is unproductive overhead for the user who just wants to write something quickly: they would be better off using something like Word.

It was determined that the authoring of structured scholarly documents was the target audience of the CAT. This would include research papers, proposals, and textbooks. With that decision made, the discussion turned to planning the next steps.

The Next Steps

A lesson recently learned in the software industry is that user interfaces should be designed before writing any back-end code. A mockup of the interface allows potential users to give feedback on the planned features, and those features can be tested without the expense of fully implementing them.

An expert in graphical interface design took us through the steps of creating an interface. He described the “wireframe” approach, which consists of creating a collection of web pages which have the basic layout of the final product, but without any styling, graphics, or fancy fonts. The wireframe is a graphic and functional specification that allows a user to click a submit button and go to a new page, however, the result is just a static page and no processing was done on the submitted data.

We spent about an hour going through part of one page of the CAT, with the participants providing feedback as the wireframe page was modified to better reflect the priorities of the users. (In this particular example, the users indicated that reverting to a previous version of a document was not a high priority, so that portion of the page was made much smaller.) It was estimated that it would take about two person-months of work by an interface designer to create a wireframe for the CAT, provided there was regular feedback from potential users. The resulting wireframe could then be used directly when back-end coding began.

The participants were enthusiastic about the creation of the CAT and many offered their projects for beta-testing. Creation of a wireframe will allow these users to provide more specific guidance about the design and function of the CAT. Furthermore, a wireframe will allow us to more easily demonstrate to outside groups the nature and purpose of this novel tool for the creation of scholarly documents.