A Web-based Collaborative Document Reviewer

John F. Bremer Jr.  Michael W. Donovan  Ed Hougardy
The Boeing Company  The Boeing Company  The Boeing Company
John.f.bremer@boeing.com  Michael.w.donovan@boeing.com  ed.hougardy-ii@boeing.com

ABSTRACT

The Boeing Collaborative Document Reviewer is an enterprise web application supporting large scale asynchronous collaborative authoring through real-time shared commenting. The application is specifically targeted to supporting the review and critique of contractor draft product documents by United States government customers. The system supports an iterative In-Process Review (IPR) workflow with features that include project-oriented role-based security, document version control, in-context commenting, notification and comment search, comment life cycle, and full-text indexing and multiword term text mining with hyperlinked Key Word In Context (KWIC) search results. The application has been used successfully in production to review thousands of pages of HTML and PDF documents and CGM graphics by hundreds of reviewers distributed across the United States.

Author Keywords

ACM Classification Keywords
H5.3 [Information interfaces and presentation, HCI]
Group and Organizational Interfaces – Asynchronous interaction, Computer-supported cooperative work

INTRODUCTION

The Boeing Collaborative Document Reviewer (BCDR) is a web-based application that supports contractually required In-Process Reviews (IPR) of deliverable product documentation by the US Government. The BCDR application design was conceived in 2003 based on the experience of the development of an earlier stand-alone Internet Explorer based Interactive Electronic Technical Manual (IETM) Reviewer, Date Module Explorer (DME), which was deployed to NATO in 2000. The BCDR application was developed in 2004 by a team consisting of R&D Advanced Technologists, Production Web Application Developers, Technical Writers and Publications Managers and government reviewers. The design focused on fast downloading and updating of data through network firewalls and ease of installation and use in existing infrastructures which were known issues for DME and browser-based commercial IETM products. The application feature set was evolved with regular updates to the working application. BCDR was put into production in January of 2005 and has continued to be enhanced. Between January and August of 2005 there were five blockpointed upgrades to the application encompassing 96 change requests. (Regular upgrades to the application were possible because of the near zero client installation nature of the application.)

To date, the primary installation of the application has been used to review approximately 4 million words contained in 3,530 documents spread across 27 IPR projects with a total of 480 users at Boeing, the US Air Force and a Boeing subcontractor collaborating from locations on the West Coast, Midwest and East Coast of the United States.

There have been a total of 7,665 root comments created with an additional 30,022 threaded child comments associated with dispositioning the root comments through the comment life cycle.

As of August of 2007, in addition to the primary installation of the system, there were 4 other production installations of the system in company serving 6 different publications groups.

A general use version of the application is also available and has been employed by different groups in the company for a variety of uses including “shredding” resumes and survey results, and analyzing online technical help chat session histories.

FEATURES

The BCDR application is an automation and optimization of a legacy paper document IPR process. The legacy process consisted of a set of steps including iteratively creating a formal draft document and providing it to the customer, collecting comments on the draft during a fixed review period and then formally dispositioning the comments in a face to face meeting. In the face to face
meeting the comments were aggregated and cooperatively dispositioned in order to determine which comments should dictate changes in a future draft of the document.

BCDR supports an automation of this legacy process. In the automated process documents are authored in various digital authoring environments including ArborText Editor, MS Word, Interleaf and others. The documents are saved as HTML or PDF and graphics files are saved as CGM. They are then uploaded into an IPR project in BCDR where they are available for review on the web by customers who have been granted login access and review permissions on the IPR project. Users navigate to the BCDR web site, log in and create in-context, overlaid “sticky note” comments by highlighting text, rubber-banding part of a PDF graphic or zooming in on a specific portion of a CGM graphic and then typing a problem description and suggested re-write into a web form and submitting it back to the server.

Comments are saved separately from documents by the web server in a central database and upon creation are immediately broadcast to all other users of the application where they show up in the comments table of contents. Users who have been assigned contractor or government comment dispositioner roles on an IPR project disposition the comments, moving each comment through a series of status changes until Boeing and the government customer have cooperatively determined to accept or reject the comment. Subsequent comments associated with comment status changes are threaded and included with each original comment providing a complete view of the rationale behind the dispositioning of the comment.

RELATED WORK
BCDR is the result of a requirements driven software development project. The resulting application shares review features of commercial products found on the corporate desktop that support annotation such as MS Word and Adobe Acrobat and Microsoft Office. Like MS Word and Adobe Acrobat, BCDR has features that allow users to make in-context comments by selecting text. And like MS Office and Acrobat, BCDR also allows users to participate in shared reviews via the web.

Commercial products exceed the capabilities of BCDR in some features, but lack other capability required for a real-world enterprise solution. Important areas where the commercial products fall short are security management, performance on low bandwidth, and support for multiple document types. Typically, commercial products provide a much richer set of annotation types than BCDR, which has only three. Commercial products also typically support only a limited comment dispositioning workflow to either accept or reject comments while BCDR supports table-driven customizable workflows like the contractor-governed workflow described in this paper.

BCDR shares features with other applications and designs described in the research literature [2,3,4]. For example, Weng and Gennari describe their annotation model for “Asynchronous Collaborative Writing through Annotions” [4] as solving four types of communication and collaboration problems. To paraphrase, these problems include loss of collaboration communication data, lack of rationale associated with document changes, poorly supported group discussion and coordination, and problems with keeping track of which comments apply to which versions of the documents. These problems are typical with traditional spread sheet style reviews in which users “log” a problem by adding a row to a table. Typically, the table includes the context of the problem, e.g., “in document X, paragraph 4”, the offending text or
image, and some comments which may include a suggested correction to the text or image. This kind of review continues to be pervasive in Boeing and the US Government.

In their annotation model Weng and Gennari suggest the following features are needed to address these problems.

“1. Support in-situ communication and decision-making”

BCDR includes shared inline comments and threaded annotations and a full-featured comment search capability with hyperlinks back from search results to the inline comments.

“2. Improve cross-role awareness”

BCDR supports notification, real-time comment status updates in the table of contents node text and full-featured comment search capability to insure “cross-role responsiveness”.

“3. Provide a rational version control mechanism”

BCDR maintains a comment history with each comment including mandatory comments associated with comment status changes made by contractor and government comment dispositioners. All comments are part of an IPR project which includes an archive of immutable draft documents which eliminates comment orphaning. Completed IPR projects can also be exported to a stand-alone archive format so that customers can archive DVD copies of reviews that can be used for later reference without having to access the online system. Users can view multiple IPR projects at the same time, including IPR projects for previous drafts of the same document.

“4. Improve shared workspace and group awareness”

BCDR provides real-time access to the state of the review at any given time. Users have the option to choose to be notified of activity undertaken by other collaborators or to use the search capability to search for specific activity related to specific documents, users, comment statuses or time frames.

As part of the group awareness the BCDR provides “up to the minute” comment status in the form of metrics reports. Two flavors of the report are available; a high level report which encompasses all the documents within the IPR and a detailed report showing the comment status on a per-document basis. The status reports are available for export out of the BCDR system.

While the design of BCDR supports most of the detailed features laid out in the Weng and Gennari design, there are some that it does not support. BCDR does not include any concept of comment “Rating,” as all comments are associated with perceived problems in the documents.

Cadiz et.al. [1] mention the reluctance of some reviewers at Microsoft to comment on trivial errors. BCDR users are encouraged to consider all comments on document problems “good,” but it is unavoidable that the total number of comments associated with a document will be taken as a measure of the document’s quality. In order to mitigate the potential for excess comments on trivial errors in the documents, Boeing authors initiated the practice of conducting limited peer IPR reviews of draft documents before including the drafts in an official IPR project involving the government reviewers.

“Urgency” is also a concept that is usurped in BCDR by the concept of an IPR project being open for review for a fixed time frame.

It is interesting to note that Weng and Gennari suggest that once an annotation has been incorporated it is no longer necessary to track it. This is consistent with the iterative IPR workflow supported by BCDR where each new draft effectively represents a new baseline of the document.

ARCHITECTURE

The success of the BCDR application in the Boeing enterprise and its continued proliferation can be attributed greatly to the application architecture that allows for significant improvements to legacy document authoring and review workflow. The application saves time and money in reduced travel and cycle-time, reduced duplication of review effort, reduced disruption of normal reviewer work activity, and reduced disruption of the existing culture and computing infrastructure.

Chief among the architecture attributes contributing to the success of the application include the following:

- The application requires near zero installation, excepting only the Acrobat Reader plugin for reviewing PDF documents and the Boeing Intelligent Graphics CGM Viewer plugin for reviewing CGM graphics.
- The application installs into the standard enterprise web application architecture of a Microsoft IIS web server, SQL Server database and Internet Explorer web browser. All data between the web server and web browser is transferred via the HTTPS protocol allowing it to pass securely through standard corporate and government firewalls.
- The application supports review of HTML, PDF and CGM document types which allows virtually any digital document to be accurately rendered and reviewed in the system.
- All documents under review are archived by the web server and are immutable. Existing
comments are stored separately from documents and are inserted dynamically into documents on-demand.

- The application includes multiword term entity and concept text mining, full-text search and Key Word In Context report generation, as well as fully parameterized searching and reporting on comment text.

- And finally, the application makes extensive use of AJAX (Asynchronous JavaScript And XML) techniques to transfer XML data between the web server and web browser in order to update web pages in the browser incrementally, without refreshing entire web pages. This enables the application to substantially reduce the amount of data being transferred and perform much faster than standard server-side web applications tasked with moving large amounts of data over the limited bandwidth of corporate and government firewalls.

The different kinds of data encoded and transferred as XML via HTTP includes Table of Contents nodes, comments, HTML document text, HTML document images, CGM images and CGM Intelligent Graphics, comment archives and PDF documents.

A variety of standard XML APIs are employed at the database, in the ASP Visual Basic code, in the browser JavaScript code and in internal Adobe Acrobat JavaScript code to both generate and consume XML data.

Novel programming techniques and algorithms are employed throughout the application in order to improve the speed with which the application responds to user actions. Features of the application that employ this advanced technology include the following:

- A database architecture and XML schema that allows tables of contents to grow dynamically in width and depth and table of contents node child nodes to be fetched “just in time” when a user clicks to expand a parent node.

- Fast, dynamic generation of hyperlinked Key Word In Context (KWIC) full-text search results with a user specified amount of leading and following text.

- Fast full-text index generation, text-mining for multiword entity and concept name spotting, and database loading.

- Directed Boolean AND Query construction for full-text search.

- Full integration of the Adobe Acrobat plugin into the web application.

- A point-in-time archive capability. This allows for historical record keeping.

**DISCUSSION**

The development of BCDR for its initial customer had the advantage of being an automation of an existing process with substantial potential savings in money and time while improving overall quality. However, as the application has been migrated to other organizations with other workflows, it has not been as readily accepted into active use. In light of this, it is interesting to try to make some observations about what it is about the application and about particular publications groups that make them amenable to the adoption of BCDR.

By working in an interactive collaborative environment, earlier exposure/detection to potential problems can result in reduced rework by not repeating the problem in data which is prepared but not yet posted for review.

A key design to the usability of the BCDR is its common user interface. Using a standard web browser with expected navigation methods (pull-down menus, hyper text linking, and choice boxes) allows the reviewers to quickly become comfortable with the system. This allows the Reviews to be focused on the reviewing of the IPR data rather than the system itself. Peter Morville in his book *Ambient Findability* quotes Calvin N. Mooers: “An information retrieval system will tend not to be used whenever it is more painful and troublesome for a customer to have information than for him not to have it.” [6] The same holds true for the BCDR; if it were too complex to use, reviewers would not use the system. Again echoed by Morville …[U]sability, user experience,
and user-centered design. Make it simple. Make it easy. Don't make me think!

Clearly the biggest motivator for acceptance of the application into a publications group is time and cost savings in the context of a contractual requirement. The amount time and money that can be saved depends on the level of publication activity. The larger the number of collaborators and greater the size and importance of the documents, the more attractive BCDR becomes. It is still a significant undertaking, however, for a large publications organization to automate portions of their daily workflow. In the case of the primary BCDR customer, there are hundreds of people working on thousands of pages of logistics documents, which support a product (an airplane) that will fly for 40+ years. The documents will be referenced on a daily basis and will be used to keep the product in service. So here we have an ideal case of a large distributed group of people collaborating on the creation of a large set of very important documents.

On the other hand, the system has also been deployed to a publications group to review small documents of a perishable nature. The group initially reviewed their documents using the system, but abandoned the application after a few months because using BCDR in their workflow did not account for any significant cost or time savings.

CONCLUSION
BCDR is an example of a successful collaborative web application on an enterprise scale. It facilitates valuable collaboration between subcontractor and company authors and their customer audience throughout the world. It employs an automated IPR process for organized and efficient creation of documents through the iterative review of formal draft documents avoiding the pitfalls of orphaned comments and the duplication of effort associated with the review of continuously creeping document drafts. BCDR employs an AJAX architecture allowing good application performance over low bandwidth firewalls and allowing for easy upgrades to the application with changes to code required only on the web server and not at the users desktop.

REFERENCES


5. Bremer, John F., Donovan, M., Folger D., Jackson, W. “AJAX Elements of The Boeing Collaborative Document Reviewer,” Boeing XML@Large Vignette, 2006