Co-operative Authoring and Collaboration over the World Wide Web

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Technology

in

Computer Systems Engineering

at Massey University, Palmerston North, New Zealand

Daniel Robert Hean

2000

Abstract

Co-operative authoring and collaboration over the World Wide Web is looking at a future development of the Web. One of the reasons that Berners-Lee created the Web in 1989 was for collaboration and collaborative design. As the Web has limited collaboration at present this thesis looks specifically at co-operative authoring (the actual creation and editing of web pages) and generally at the collaboration surrounding this authoring. The goal of this thesis is to create an engine that is capable of supporting co-operative authoring and collaboration over the Web. In addition it would be a major advantage if the engine were flexible enough to allow the future development of other access methods, especially those that are web related, such as WebDAV, WAP, etc.

Acknowledgements

Supervisor: W.H. Page, The Institute of Information Sciences and Technology

Supervisor 2: P. Lyons, The Institute of Information Sciences and Technology

Employer: School of Psychology, Massey University

Table of Contents

Table of Figures

1 Introduction

When I first discovered the Internet, back in the days before there was any overt commercial interest in the Internet, it was primarily a link between universities and other non-commercial groups. At that time one of the main uses of the Internet was for collaboration between people on a variety of different subjects. This collaboration was primarily based on text communications such as email, Usenet, and talk. They were mainly text based because of the limits of the bandwidth available at the time, although there were probably some groups connected with, what at that time would have been called, high speed. An example could be a high-speed backbone of 56Kbps, which was very high speed, when compared to some of the modem connect speeds at the time (for example 300 baud, that is 300 bits per second).

When I started to investigate the origins of the World Wide Web [Berners-Lee WtW], I came across some similar ideas for its use, that is, for collaboration on other projects. There was more to it than just collaboration on other projects though. Included in Berners-Lee's original visions for the Web was the idea of collaborating on and about the Web itself. This combined with other ideas from other areas (such as the open manner of collaboration that I had experienced over the Internet, older methods of group document creation and editing, etc) led me to look at the co-operative authoring of documents. This research concentrates on the co-operative authoring of documents, which in terms of this research means the actual collaborative work on a document. This can be compared with collaboration, which is one of the support mechanisms that allow people to work on documents together which will be looked at, but is not the main focus. This is an important distinction for this document, as it is specifically focused on the co-operative authoring of documents, rather than all the support mechanisms that surround that authoring (such as email, Usenet, discussion groups and the like).

The first area that is examined is co-operative authoring and collaboration in relation to the past and future of the Web. The second (and main) purpose of this research is to design and implement a vision of co-operative authoring and collaboration over the Web. Part of this vision dates back to the original proposal for the Web [Berners-Lee 1989].

The vision that I have for the future of the Web allows everyone the freedom to publish to the Web at anytime, from any location, using their favourite tool. Also to do so in a manner that allows co-operative authoring of documents and collaboration between the people involved, including full co-operative authoring of documents, rather than just sharing them. This is the basis for what I am calling a free flowing web of information. This may sound like an oxymoron, because a web is typically static, while it catches things that flow through it. In the sense of the Web though, information is flowing around it (currently not freely, as it mostly only flows from a webmaster to a web surfer), but at the same time, the Web is not static. The Web is continually changing, which is one of the challenges that people can face when trying to understand it. For example search engines and web maps will probably never cover the whole of the Web, because it is continually expanding at a tremendous rate, while other parts disappear. This means that not only is information moving through the Web, the Web itself is moving and changing as parts are changed, added and removed.

In conventional approaches to creating and publishing documents on the Web, one person or group assembles the information and then publishes it in "read-only" format on the web site. Many authors modify this idea and use automatically generated content, such as constantly updated prices, or a threaded web discussion. Generally this is separated from the rest of the site, either by position, or by some marking out. Some of this automatically generated content may be an integral part of the site, although this tends to change less often, for example, menus, page headers and page footers. The fully automatic generation of content and/or pages can be quite clearly defined as different from the static web pages (even though the distinction does blur a bit when you talk about server executed pages, such as SSI (Server Side Includes) or Java Servlets (Java [Java] running on the server), which include pieces of automatically generated content in otherwise static pages). The reason that fully static pages and fully dynamic pages can usually be quite clearly defined is primarily because the type of information present is quite different. Static pages typically have information that does not change all that often, whereas dynamic pages have information that can be changing continuously, or at least very often. This distinction can blur with the use of static pages as cache for a dynamic process (as can be the case in multi-tier web sites), or with the use of server executed pages where part of the page is static, and part is dynamic, which means that the page can contain both static and dynamic information.

The antithesis of the normal web publishing method is to allow anyone and everyone to change and update the web site. An example of this is WikiWiki [Cunningham], a web site created and maintained by W. Cunningham at http://c2.com/cgi-bin/wiki that implements some of the co-operative authoring vision for the Web and allows for a limited amount of free flowing of information. It can be described in many ways, such as this quote from its home page (current October 2000) "(It is) ... a fun way of communicating asynchronously across the network", or as a set of web pages that are open and free for anyone to edit as they wish. Wiki is not real-time; therefore people have time to think before they follow up a web page, often days or weeks, so they have time to consider what they write. It was created for the discussion of People, Projects, and Patterns (a pattern is a recurring solution to a common problem in a given context and system of forces [Alexander 1977][Alexander 1979]).

The mixing of ideas and possibilities that can occur in an environment such as WikiWiki is in essence very similar to how the Internet came into being, and it is a good example of one way of implementing a web site that is based on the free flow of information. It is completely free flowing, and follows no set rules; anyone can update anything, anytime, anywhere, with no barriers stopping people from doing what they want. This free flowing method of web publishing could be anarchy when compared with the current normal web publishing practice, but it shows up some interesting comparisons with the normal method of publishing on the Web. On a normal web site the expectation is that only those changes that are approved and have gone through the normal channels will ever be made to the web site. This can have the consequence that when a change that is detrimental for the web site is made through approved channels, it can take just as long to fix it (once it is known), as it took to implement. Another consequence is that because any changes have a set channel to go through, the owner can be lulled into a false sense of security that no unauthorised changes will ever be made. [Risks] Thus, if unapproved changes are made, it can cost a tremendous amount of time and money to fix. The free flow of information on a web site can also be the key to realising a whole new form of web collaboration. It is based partly on the

original vision for the Web (that has not yet been fully explored) and extends it using today's ideas and tools while keeping the compatibility to allow the majority of people on the Web to have access to and use it.

I have a vision of the Web as a free flowing web of information that flows in any direction. The implementation of this vision provides the freedom for people to have their views heard and seen by those that need to see them most. It enables people with unique ideas to co-author documents in an easy and commonly available format, and it allows for anyone to use the tools they are most productive with, at the time of day or week that they are the most productive.

The author has been working on a web site (http://www.YEdit.com/) that gives an overview of the ideas for a free flowing web, as well as an implementation. Web sites can be set up so that multiple people can update the open part of the sites, while groups can share information, and co-operatively and collaboratively write documents. This web site fits in nicely with the current crop of CSCW (Computer-Supported Co-operative Work), BSCW (Basic Support for Co-operative Work) [BSCW] and Groupware sites that are set up to enable collaboration about documents.