

## University of Southampton Research Repository ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

UNIVERSITY OF SOUTHAMPTON

Improving Users' Awareness Interactions in the  
Collaborative Document Authoring Process: The  
CAWS Approach

by

Ilaria Liccardi

A thesis submitted in partial fulfilment for the  
degree of Doctor of Philosophy

in the  
Faculty of Engineering and Applied Science  
Department of Electronics and Computer Science  
United Kingdom.

February 2010

## ABSTRACT

Awareness of individual and group activities is critical to successful collaborative authoring. Participants require knowledge of what other contributors are doing and have done, what meaningful changes have been made to a document, and who is editing each section of a document and why. With this information, group dynamics can be improved and members can work more efficiently toward the final product.

In this thesis, key problems in collaborative activities are identified through a review of previous research on the subject and from field research of authors engaged in collaborative work. From these initial observations we deduce that many problems in collaborative writing occur due to technology that hinders the proper distribution of information to members of the group. The concept of “awareness”, identified in past research, is discussed, and used as a model to explain the underlying causes behind these common problems.

As a specific example of the importance of communication and coordination mechanisms, an analysis is presented of the Wikibooks website, an online collaborative writing site that allows volunteers to work together to develop free textbooks. Statistical analysis of historical data from the site is used to correlate successful books with efficient use of planning, communication and coordination techniques. These results help to further cement the importance of communication and awareness channels.

From analysis of these issues, a set of requirements is defined for an effective collaboration tool, specifically the features that such a tool should include in order to support the types of awareness that are necessary for successful collaboration. Existing groupware systems are compared and judged against these requirements, with the discovery that most systems lack support for many different types of awareness.

To investigate the subject further, a prototype co-authoring system with features to support awareness (CAWS), developed as part of this research, is described. It is explained how these features attempt to reproduce some of the communications channels implicitly present within an office environment. The results of a usability study using the CAWS system are then presented, with particular reference to the

effectiveness of the features of the system. Feedback from participants was gathered with respect to usefulness and ease of gathering information about other users' progress and interactions with the workspace with these features present.

Finally the observations, findings and the implications of a real world groupware evaluation are presented (undertaken over a period of 17 weeks with 85 students divided into 15 groups). The groupware evaluation gives insight into the effectiveness of awareness mechanisms. This includes the role and effect of planning, the effect of the choice of tool on perceptions of awareness, the relative importance of awareness and how awareness contributes to a successful collaboration. We discuss the outcomes of the research with respect to the research questions and contribution, presenting how the research could be continued in the future.

# Acknowledgements

I would like to acknowledge several people who have provided inspiration as well as help throughout the development of this thesis.

This thesis would have not been able to see a conclusion without the continuous encouragement of my first supervisor Prof. Hugh C. Davis, who for the duration of the Ph.D allowed me to expand my knowledge and expose my work to the Computer Supported Collaborate Work (CSCW) and the Human Computer Interaction (HCI) communities. I would also like to thank my second supervisor Dr. Su A. White for help and support, comments and invaluable suggestions in writing this thesis.

I would also like to thank Prof. Carl Gutwin from the University of Saskatchewan for his insight and comments on my research and for his work in the CSCW community in the field related to awareness of users engaged in collaborative authoring.

I would like to thank the Web Science Research Initiative for awarding me the grant money to attend the WebScience Summer School at the Oxford Internet Institute at the University of Oxford, since these two weeks provided insight into the research area from a social and technical prospective. I would also like to thank the WSRI for funding my three month stay at MIT.

My deepest and warmest thanks go to Professor Hal Abelson at MIT. Apart from his continuous inspiration to me in the fields of privacy issues and web technology, he also provided extremely useful and inspirational insights into the research topic, specifically the suggestion of a large scale analysis of Wikibooks. He was also extremely helpful in the design of my experiments, and in addition to his helpful insights and comments he introduced me to Prof. Wanda J. Orlikowski at MIT Sloan School and Professor Wendy E. Mackay at Université de Paris-Sud (INRIA). I would like to thank Prof. Orlikowski for providing useful insight into the general evaluation of group dynamics.

I would also like to thank the all the members of the DIG Group at MIT, for their insights into the research in general and for how to prove/disprove the hypothesis. To name a few, these include Dr. Lalana Kagal, and Oshani Seneviratne for her comments on the CAWS tool and suggestions for how the semantic web might be used to improve

its effectiveness. I would also like to thank Prof. Sir Tim Berners-Lee for his insight into the features of the tool.

Professor Wendy MacKay was instrumental in the completion of this Ph.D, because of her suggestions and help in various parts of the work. I would like to thank Professor Mackay and Au Yeung Ching Man for collaborating in the Wikibooks work. Au Yeung Ching Man was instrumental in measuring redundancy behaviour in the book dataset (Chapter 4). Professor Mackay helped refine the idea as well as collaborate on the chapter. She provided constant insight in the last year of this research, providing different points of view on the research in general and was extremely helpful in designing the final experiment (Chapter 9).

I would also like to thank Dr. Dave Woods for his help and patience in the relation to the design of the experiments. He was extremely helpful in making me understand the different factors affecting groups of users and how these should be tested, and was instrumental in the design of the first usability study (Chapter 8) as well as his comments in the final experiment (Chapter 9).

I would like to thank Dr. Michael R. Poppleton and Dr Robert J. Walters for their help in running the experiments as well as the second year computer science students who patiently undertook each stage of the study.

I would also like my dear friend Asma for her support throughout the course of my Ph.D.

I would also like to thank my mother and father without whom I would not have been able to do this, for their love, their never-ending support, and for always understanding me.

My final and most important thanks goes to Simon, for his love, his never ending support and infinite patience through the entire course of study, without whom I would not have been able to have the satisfaction of writing this thesis.

*To Simon, my parents (Gianni and Maria), my brother  
Gianmaria, my uncle Giovanni, my aunt Giulia (who is missed  
dearly every day), my two cousins (Mariangela and Giuseppe)  
and the entire remainder of my family.*

# List of Publications (13)

## Article (2)

- **Liccardi, I.**, Ounnas, A., Pau, R., Massey, E., Kinnunen, P., Lewthwaite, S., Midy, M. A. and Sakar, C. (2007) *The role of social networks in students' learning experiences*. ACM SIGCSE Bulletin (December Issue) pp.224-237.
- **Liccardi, I.**, Davis, H. and White, S. (2008) CAWS: Visualizing awareness to improve the effectiveness of co-authoring activities. Special issue of Collaborative Computing in IEEE Distributed Systems Online pp. 6-12.

## Doctoral Consortium (1)

- **Liccardi, I.** (2007) *CAWS: Improving users' awareness for advancing effectiveness of co-authoring activities*. In Proceedings of ACM GROUP 2007 Doctoral Colloquium, Sanibel Island, Florida.

## Short papers (7)

- **Liccardi I.**, Hugh C. Davis, Su White, *Progressional awareness: designing a co-authoring tool to support the planning process*, ACM SIGDOC 2009, 5-7 October, Indianapolis, IL, USA, pages 115-118.
- **Liccardi I.**, Hugh C. Davis, Su White, Behavioral awareness for modeling social interactions in online collaborative authoring environments, WebSci 2009: Society On-Line, 18–20 March, Athens, Greece.
- **Liccardi, I.**, Davis, H. and White, S. (2008) *CAWS: an awareness based wiki system to improve team collaboration*. In: ICALT 2008: The 8th IEEE International Conference on Advanced Learning, 1-5 July, Santander, Cantabria, Spain, 265-267.
- **Liccardi, I.**, Davis, H. and White, S. (2007) *CAWS: A wiki system to improve workspace awareness to advance effectiveness of co-authoring activities*. In Proceedings of ACM CHI 2007 Extended Abstracts, San Jose, California, USA, pp. 2555-2560.



- Woods, D. C., Grove, D. M., **Liccardi, I.**, Lewis, S. M. and Frey, J. G. (2006) *An eLearning website for the design and analysis of experiments with application to chemical processes* (Speech). In Proceedings of 17th COMPSTAT Symposium of the IASC 2006, Rome, Italy.
- White, S. A. and **Liccardi, I.** (2006) *DISCIPLINARY DIFFERENCES – Frameworks For Better Learning Design*. In Proceedings of IEEE ICALT 2006 - The 6th International Conference on Advanced Learning Technologies, Kerkrade, The Netherlands.
- White, S. A. and **Liccardi, I.** (2006) *Harnessing Insight Into Disciplinary Differences To Refine E-Learning Design*. In Proceedings of 36th Annual ASEE/IEEE Frontiers in Education, S3J5-S3J10, San Diego CA.

### **Workshop Item (3)**

- Ching-man Au Yeung, **Ilaria Liccardi**, Kanghao Lu, Oshani Seneviratne, Tim Berners-Lee, *Decentralization: The Future of Online Social Networking*, W3C Workshop on the Future of Social Networking, 15-16 January, Barcelona, Spain.
- **Liccardi, I.**, Davis, H. C. and White, S. A. (2007) *CAWS: Improving users' awareness in collaborative authoring activities*. In Proceedings of ECSCW 2007 Workshop on The Challenges of Collaborative Work in Global Software Development, Limerick, Ireland.
- Ounnas, A., **Liccardi, I.**, Davis, H. C., Millard, D. E. and White, S. A. (2006) *Towards a Semantic Modeling of Learners for Social Networks*. In Proceedings of International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL) at the AH2006 Conference, pp. 102-108, Dublin, Ireland.

# Contents

<b>Chapter 1 Introduction.....</b>	<b>1</b>
1.1 Setting the scene .....	1
1.2 Research Aim.....	3
1.3 Overview of the dissertation.....	3
1.4 Research Questions .....	6
1.5 Hypothesis .....	8
1.6 Contributions of the research.....	8
1.7 Declaration.....	10
<b>Chapter 2 Collaborative Authoring .....</b>	<b>11</b>
2.1 Introduction.....	11
2.2 Traditional collaborative authoring .....	12
2.2.1 The form of collaboration .....	12
2.2.2 Collaborative activities .....	13
2.2.3 Collaboration Size .....	14
2.2.4 Gains and losses due to collaboration.....	14
2.2.5 Field Study .....	16
2.2.5.1 Stages of development and problems related to communication mechanisms .....	17
2.3 On-line Collaborative Authoring .....	22
2.3.1 The mode of collaboration .....	22
2.3.2 Why collaborate?.....	24
2.3.3 Communication mechanisms .....	24
2.4 Online and traditional collaboration compared .....	25
2.5 Conclusion .....	27
<b>Chapter 3 Awareness in Groupware.....</b>	<b>28</b>
3.1 Introduction.....	28
3.2 Types of awareness mechanism.....	29
3.2.1 Personal Awareness .....	30

3.2.2	Informal Awareness.....	31
3.2.3	Social Awareness.....	31
3.2.4	Group Awareness .....	32
3.2.5	Workspace Awareness.....	33
3.3	Discussion.....	33
3.4	Conclusion.....	35
<b>Chapter 4 Impact of poor communication and collaboration mechanisms - Example using Wikibooks .....</b>		<b>36</b>
4.1	Introduction.....	37
4.2	Measuring Semantic Similarity and Redundancy .....	37
4.3	Conjectures of Redundancy Occurrences.....	38
4.4	Data Source.....	40
4.4.1	Overview of Wikibooks.....	40
4.5	Description of Data for Analysis .....	42
4.5.1	Measure of Quality in Wikibooks .....	43
4.6	Method For Measuring Redundancy.....	44
4.7	Analysis .....	45
4.7.1	Contributors concentration.....	45
4.7.2	Changes in redundancy over time .....	46
4.7.3	Redundancy and planning .....	48
4.7.4	Redundancy and quality.....	52
4.8	Discussion.....	52
4.8.1	Limitation to the Approach .....	54
4.9	Conclusion .....	55
<b>Chapter 5 Collaborative Tools Requirements .....</b>		<b>56</b>
5.1	Conjectures and Hypothesis .....	56
5.2	Communication mechanisms .....	59
5.2.1	Commenting.....	59
5.2.2	Real time discussion .....	60
5.2.3	Asynchronous discussion.....	61
5.3	Awareness Characteristics.....	62

5.3.1	Personal Awareness .....	62
5.3.2	Social Awareness.....	63
5.3.3	Informal Awareness.....	63
5.3.4	Group Awareness .....	64
5.3.5	Workspace Awareness .....	64
5.4	Discussion and Conclusion.....	65
<b>Chapter 6 Available Collaborative Tools.....</b>		<b>66</b>
6.1	Introduction.....	66
6.2	Communication mechanisms .....	70
6.2.1	Commenting .....	70
6.2.2	Asynchronous discussion.....	70
6.2.3	Real time discussion .....	71
6.3	Awareness Characteristics .....	71
6.3.1	Personal Awareness .....	71
6.3.2	Informal Awareness.....	72
6.3.3	Group Awareness .....	73
6.3.4	Workspace Awareness .....	73
6.4	Discussion.....	74
6.4.1	Google Wave.....	78
6.5	Conclusion .....	79
<b>Chapter 7 CAWS – Co-Authoring Wiki-Based System .....</b>		<b>81</b>
7.1	Introduction.....	81
7.2	Example Scenario .....	82
7.3	Communication Mechanisms .....	84
7.3.1	The role of annotations .....	84
7.3.1.1	Supplementing participants’ points of view .....	84
7.3.1.2	Flagging appointed changes.....	85
7.3.1.3	Annotation filtering .....	86
7.3.1.4	Visualising authors’ opinions.....	87
7.3.2	Asynchronous Discussion.....	88
7.3.2.1	Facilitating annotation of sections of the manuscript.....	88

7.3.2.2	Discussion forum.....	89
7.3.2.3	Facilitating news updates with asynchronous discussion – Using a Blog.....	90
7.4	Awareness Features.....	91
7.4.1	Informational Awareness .....	91
7.4.1.1	History of users’ actions .....	91
7.4.2	Group Awareness .....	92
7.4.2.1	Division of Activities.....	92
7.4.3	Workspace Awareness .....	95
7.4.3.1	Users’ whereabouts .....	95
7.4.3.2	Spy Mode Feature.....	95
7.5	Conclusion .....	97
<b>Chapter 8</b>	<b>Experiment 1: Attitude based - Usability Study .....</b>	<b>99</b>
8.1	Methodology.....	99
8.2	Usage Patterns and Feedback .....	102
8.2.1	History of users’ actions .....	102
8.2.2	Users’ Whereabouts.....	103
8.2.3	Division of Activities.....	103
8.2.4	Use of colour for visualisation .....	105
8.3	Discussion.....	107
8.4	Conclusion .....	108
<b>Chapter 9</b>	<b>Experiment 2 - CAWS Groupware Evaluation.....</b>	<b>110</b>
9.1	Setting.....	110
9.2	Study Methodology.....	111
9.3	Observations from observational study and semi-structured interviews ...	113
9.3.1	Importance of awareness at each stage .....	113
9.3.2	Effect of document structure on the planning process.....	115
9.3.3	Breakdowns in communications .....	117
9.4	Observations from survey.....	119
9.4.1	Skill sets .....	119
9.4.2	Group structure and methodology .....	121

9.4.3	Perceived importance of different types of awareness .....	122
9.4.4	Tracking users' progress and document status (Informal Awareness) .....	124
9.4.5	Tracking up-to-the-minute progress and document status (Workspace awareness) .....	126
9.4.6	Effect of the collaborative authoring tool on the collaborative authoring process .....	127
9.5	Findings .....	128
9.5.1	Effectiveness of awareness mechanisms at each stage .....	128
9.5.2	Role of planning .....	130
9.5.3	Effect of planning on interactions .....	131
9.5.4	Effect of tool on perceptions of awareness .....	132
9.5.5	Relative importance of awareness mechanisms .....	133
9.5.6	Effect of awareness on success of collaboration .....	134
9.6	Implications .....	135
9.6.1	Progressional awareness .....	135
9.6.1.1	Who is assigned to edit which section? .....	136
9.6.1.2	Who is editing which section? .....	137
9.6.1.3	What are they writing? .....	137
9.6.1.4	What is complete? .....	138
9.7	Conclusion .....	138
<b>Chapter 10 Conclusion and Discussion.....</b>		<b>140</b>
10.1	Summary of Findings .....	140
10.2	Research Questions .....	142
10.3	Future Work - Awareness Mechanisms .....	144
10.4	Future Work - Improving Peer Review Process .....	145
10.4.1	Possible problems in the peer review process .....	146
10.4.2	Potential improvements to the peer review process .....	147
10.5	Conclusion .....	149
<b>References</b>		<b>151</b>
<b>Appendix A - Experiment 1 Materials: Usability Study .....</b>		<b>157</b>

Appendix A.1: User 1 SCENARIO.....	158
Appendix A.2: User 2 SCENARIO.....	172
Appendix A.3: User 3 SCENARIO.....	186
<b>Appendix B – Experiment 2 Materials: Semi-Structured Interviews.....</b>	<b>200</b>
Appendix B.1: Semi-Structured Interviews – Drafting Stage .....	201
Appendix B.2: Semi-Structured Interviews – Planning Stage.....	204
Appendix B.3: Semi-Structured Interviews – Developing Stage .....	207
Appendix B.4: Semi-Structured Interviews – Reviewing Stage.....	210
Appendix B.5: Semi-Structured Interviews – Formatting Stage .....	213
<b>Appendix C – Experiment 2 Materials: CAWS Groupware Evaluation User Questionnaire .....</b>	<b>216</b>
<b>Appendix D - Further features in CAWS.....</b>	<b>231</b>

# List of Figures

<b>Figure 1.1:</b>	Overview of the research approach).....	4
<b>Figure 2.1:</b>	Document Development stages.....	15
<b>Figure 3.1:</b>	Different types of Awareness (modified image from (Gutwin and Greenberg, 2002).....	26
<b>Figure 4.1:</b>	Possible patterns of changes of redundancy across time. (a) Redundancy increased over time. (b) Redundancy decreases over time. (c) Redundancy changes regularly over time. (d) Redundancy changes irregularly over time.....	40
<b>Figure 4.2:</b>	Redundancy, number of paragraphs and number of edits over time in four example books, (a) <i>Guitar</i> , (b) <i>This Quantum World</i> , (c) <i>C# Programming</i> , and (d) <i>New Zeland History</i> .....	49
<b>Figure 4.3:</b>	Latin Book main pages (–) and talk page (–) edit history (x-axis; display time; y-axis display: number of edits.....	50
<b>Figure 4.4:</b>	Latin talk page table of contents.....	51
<b>Figure 4.5:</b>	Interactions between contributors over the first six-months of the writing of <i>Arimaa</i> . Each symbol represents a different page within the book. The contributors(y-axis) used no discussion.....	51
<b>Figure 7.1:</b>	Annotation system within the CAWS System. (a) The annotation colour reflects the highlighted text in the manuscript (filtered by authors). (b) The annotation blinks yellow when clicked and the border of the annotation attach to it thickens .....	81
<b>Figure 7.2:</b>	Annotation filter box.....	82
<b>Figure 7.3:</b>	Different types of annotation viewed by users. The two colours represent the two authors that are reviewing the paper. When the same text is annotated by two users, the overalapping text becomes grey (a). Overlapping annotations made by the same author are shown with an increase of colour (b).....	82



<b>Figure 7.4:</b>	The CAWS annotation system is linked to a threaded discussion.....	84
<b>Figure 7.5:</b>	Forum feature in CAWS. Note that when a user is not online the icon next to it is greyed out providing awareness of who is online .....	85
<b>Figure 7.6:</b>	Blog featured in CAWS. The latest blog post can be seen from the document front page.....	85
<b>Figure 7.7:</b>	CAWS Document front page.....	86
<b>Figure 7.8:</b>	CAWS document structure editor. The document can be restructured by moving the section title with the mouse.....	87
<b>Figure 7.9:</b>	CAWS Planning feature: (a) Assignment of roles in relation to sections. (b) Deadline and target word count assignment. (c) Estimation by users of the remaining time needed for the completion of the section. (d) Summary of information entered and document status.....	88
<b>Figure 7.10:</b>	CAWS Document Editor highlighting the spy mode feature.....	90
<b>Figure 8.1:</b>	Results from analysing awareness of users' actions history. The number under the bars shows 100% correct answers to each of the questions given. Usefulness of the feature has an average rating of 8.1, while ease of gathering information about other users' actions has an average of 8.7.....	96
<b>Figure 8.2:</b>	Results of analysing awareness of users' whereabouts, showing 100% correct answers to all the questions given. The usefulness of the feature had an average rating of 7.9, while ease of gathering information scored 8.1 on average.....	97
<b>Figure 8.3:</b>	Results of analysis of awareness of roles and responsibilities. 79% of questions were correctly answered on average. Usefulness of the feature was rated 8.1 on average, while ease of gathering this information was rated 6.5 on average .....	98
<b>Figure 8.4:</b>	Results of analysis of awareness of the document deadlines showing 98% correct answers on average, and 8.5 average on usefulness of the feature, with 8.5 average on ease to gather information about the document deadlines.....	99

<b>Figure 8.5:</b>	Results of analysis of users' editing awareness showing 76% correct answers to all the questions given and 8.1 average rating on usefulness of the feature, with 7.8 average on ease of gathering information.....	99
<b>Figure 8.6:</b>	Results of analysis of users' annotation awareness focusing on understanding which parts of the document have more annotations, showing 88% correct answers on average, 6.5 average on usefulness of the feature, with 7.8 average on ease of gathering information.....	100
<b>Figure 8.7:</b>	Results of analysis of users' annotation awareness, focusing on understanding which authors has inserted more annotations to the document, showing 100% correct answers to all the questions given, with 7.0 on average for usefulness of the feature, with 9.0 on average for ease of gathering information.....	100
<b>Figure 9.1:</b>	Awareness types necessary at each stage of the document development process.....	121
<b>Figure 9.2:</b>	(a) Document structure; (b) Document progress (c) Editing plan for a section; (d) Summary of the plan for a section.....	128

# List of Tables

<b>Table 2.1:</b>	Purpose of different types of communication and the associated development stages. ....	19
<b>Table 2.2:</b>	Breakdown of participants' responses.....	22
<b>Table 2.3:</b>	Nested lagged regression analysis of the number of editors, coordination metrics (editor concentration and talk edits), and article lifecycle on change in article quality. (Note: *** $p < .001$ , ** $p < .01$ , * $p < .05$ ) as described in Kittur and Kraut (A. Kittur and R. E. Kraut, 2008).....	21
<b>Table 4.1:</b>	Descriptive statistics: mean, median and standard deviation in featured books .....	42
<b>Table 4.2:</b>	Descriptive statistics (mean, median and standard deviation) grouped by patterns of redundancy.....	43
<b>Table 6.1:</b>	Categorisation of tool used.....	68
<b>Table 6.2:</b>	Analysis of 21 different collaborative tools, compared against the set of features required for collaborative editing of academic publications. ■ indicates not supported, while ■ indicates supported. The legend below describes the feature represented in each column.....	78
<b>Table 8.1:</b>	Group Distribution, Sex, social relationship, current employment status and subject of study.....	100
<b>Table 9.1:</b>	Users' skill sets – Each user is assigned a random number from 1 to 6, representing their skills in writing and programming within the group.....	120
<b>Table 9.2:</b>	Mean and standard deviation of users' responses to questions about the importance of awareness mechanisms (using a ten-point Likert scale) within their collaborative environment. The responses of groups which used online tools and those which used offline tools (eg. Microsoft Word or OpenOffice) are shown separately.....	124

# **Chapter 1 Introduction**

Groupware systems are commonly used to allow people to work collaboratively from disperse locations. Potentially, geographically distant working groups can use the Internet to collaborate. However, almost all groupware systems suffer from usability flaws that inhibit their effectiveness. Compared to work conducted in person, work conducted through the use of a groupware tool can be inefficient and frustrating for the participants.

When designing a collaborative system, several aspects of collaboration must be taken into account. Users collaborate on numerous tasks, which require different levels of information (Rathwell and Burns, 1985). When collaborating in collocated environments, users are implicitly aware of the events that are taking place in the world around them. However, when collaboration takes place either partly or completely in an online environment, this information can be lost if the environment does not attempt to reproduce it, or if the users do not deliberately communicate it in detail (Carroll et al., 2003).

This research is based on the hypothesis that many of the usability flaws that commonly occur in collaborative authoring are caused by communications failures such as these. Past research is examined in the area of collaborative authoring that provides a theoretical model for believing that this is the case, and results presented from investigations into the problems affecting co-authoring. This knowledge is used to design a new co-authoring system (CAWS) with user interface features designed to support communication and awareness between authors, and results are presented of experiments performed using the system with other tools to provide a comparison.

## **1.1 Setting the scene**

Internet collaboration has become a commonplace activity. Communication with

colleagues or friends in other continents has become easy if not normal routine. However, when working and specifically when co-authoring online, most users do not use groupware applications. Microsoft Word and similar text editors are often used to work on a document, with email and/or shared directories used to distribute the document. Users may avoid groupware applications for several reasons: they may be unfamiliar with the interface (Sun et al., 2006), find the tool too difficult to use or simply believe that a normal text editor provides sufficient organisation.

Group writing differs greatly from individual writing. This is because individuals not only need to write their own contributions, but must make them coherent with the contributions of other group members. Many common problems have been identified as affecting the collaborative process. In order to address users' needs, developers of groupware applications must better understand collaborative writing activities (Noël and Robert, 2004a, Posner and Baecker, 1992). Collaborative writing not only needs an editor that allows users to work together, but also enhanced functionality to allow users to understand all aspects of the group writing activity.

Collaborative work entails cognitive aspects of communication (McGrath and Hollingshead, 1994): group members transmit, receive, and store various forms of information from each other and various other sources. Group members also affect and influence aspects of those same messages. In the ubiquitous small face-to-face group, each member can communicate with the others via a wide spectrum of communication modalities: verbal, para-verbal (e.g., voice inflection), and non-verbal (e.g., smiles and gazes). (Gutwin and Greenberg, 2002). These forms of communications are lost in an online environment.

It is important for people collaboratively authoring a document to be informed about the changes that have been made to the document between versions, when new parts are added to the document and by whom. Being aware of these changes helps the users to better understand the evolution of the document, to more easily cooperate with other users and avoid possible conflicts (Papadopoulou et al., 2006)

For example, collaborating on a common document in a student team can be challenging. Group members can have different schedules based on varying factors, such as lifestyle, part-time work, or sleeping patterns. Having a centralized document control system which allows access to the latest version of the document helps alleviate

some of these problems. However, when users work on the document from separate locations it can still be difficult to see what changes have been made. Furthermore, it is often only possible for a single user to edit the document at any one time.

In this research, I examine how the introduction of awareness mechanisms can improve group collaborative activities. These mechanisms are used for a variety of reasons, including coordination of actions, management of coupling, discussion of tasks, anticipating others peoples' actions, and to find opportunities to assist one another. When people are able to maintain awareness of one another, these activities are more natural, spontaneous, and unforced. Quality and productivity of distributed collaboration can therefore be improved if these mechanisms are integrated into a collaborative authoring tool.

## **1.2 Research Aim**

The aim of this research is to investigate techniques to improve collaborative authoring, specifically how the user interface in groupware systems can be designed to support awareness.

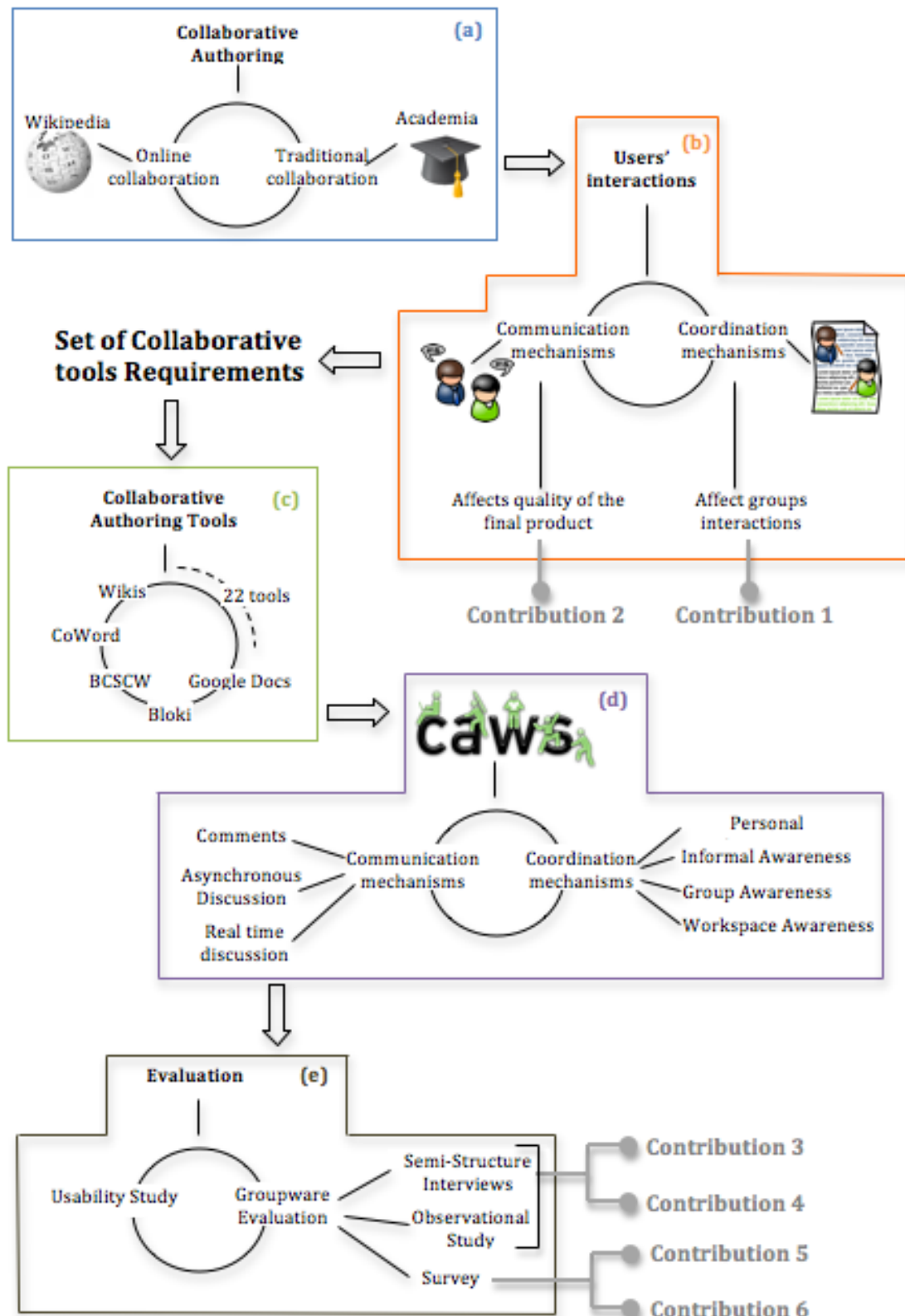
In order to achieve this aim, a prototype collaborative authoring tool (CAWS) has been developed that attempts to reproduce various different types of awareness information within an entirely electronic setting. This is accomplished by presenting up-to-date information about the activities of other users and about the state of the co-authoring process. The theoretical basis behind the design of this tool is explained in terms of existing theories of awareness, and how the tool aims to supplement this awareness. Through a series of experiments, the effectiveness of the tool is tested with groups of people performing co-authoring tasks.

## **1.3 Overview of the dissertation**

An effective design process must incorporate observations of users' experiences, experimentation, analysis of existing or similar developments and evaluation (Preece, 2002). To this end, the research process is divided into several distinct stages (Figure 1.1).

The first stage (Chapter 2) examines the types of collaborative authoring that exist, distinguishing between "traditional" face-to-face collaboration and online collaborative

authoring. In order to examine traditional collaboration, I examine past research in the



**Figure 1.1:** Overview of the research approach

area and present the results of a field study (Liccardi et al., 2007) of users engaged in

co-authoring activities either as part of their degree course (undergraduate and postgraduate students), or work (industrial researchers, professionals and academics).

This field study helps to give deeper insight into the collaborative authoring process and the common interactions that occur during co-authoring activities, in particular understanding common problems that affect users. The specific collaboration and communication mechanisms that are used during the collaborative authoring process are also identified.

The structure of “traditional” collaboration is then compared and contrasted with the comparatively newer phenomenon of online collaborative authoring. It is shown that the motivations and social relationships in an online collaboration are necessarily different due to the form that the two forms of collaboration take.

In Chapter 3, the concept of awareness is introduced. The specific types of awareness are presented, correlated with communication mechanisms that are commonly used to support them. A detailed examination is given of each type of awareness, including the type of information conveyed, details of past attempts to reproduce it, and its relevance in the context of collaborative authoring.

Chapter 4 examines the *wikibooks* website, a collaborative authoring site where volunteers work together primarily to develop free educational textbooks. A statistical study of historical data from the site is used to correlate the successful books on the site with the effective use of communications mechanisms by their authors. This further underscores the importance of communication for successful collaboration, regardless of the form that the collaboration takes.

Having identified the importance of communications mechanisms, and the detrimental effect on collaboration caused by lack of such communication, in Chapter 5 a set of requirements are devised that a collaboration tool should satisfy. The awareness types previously examined are used as the basis for these requirements. Following this, a review of the state of the art is presented in Chapter 6; existing and well known authoring tools are analysed with respect to the requirements. It is shown that none of the systems analysed fully satisfy all of the requirements or fully supports awareness.

Chapter 7 introduces the CAWS prototype system, developed as part of this research to act as a test bed for new user interface designs. The system is deliberately designed to include features intended to support awareness. Based on the design of a wiki, the



individual features are described and it is explained how these features are intended to benefit users of the system. Annotated screenshots of the system illustrate these features and their operation.

In order to evaluate the effectiveness of these features, it is necessary to test how successfully the system performs in real collaborative authoring scenarios. Chapter 8 and Chapter 9 describe the results of two such studies that were performed. In the first experiment, users performed a short sequence of tasks using the system as directed by a script and their reactions assessed. Although this was an artificial scenario, the results were helpful to the development of the system. This study also provided insights into how easily information about other users' activities were found and how effectively users responded to the supply of this information.

In the second experiment, undergraduate students used the system for a real collaborative authoring task as part of their studies. Not all of the groups used the CAWS system, allowing the efficacy of the system to be compared to other collaborative systems. This study provided information about effectiveness and importance of awareness throughout the document development stages. The results gave insight into how users collaborate with one another.

In this study we gathered quantitative and qualitative information about users' experiences within the collaborative authoring process. In an ideal world, it would have been useful to have performed a statistical analysis of the impact of each of the awareness features required by testing each feature enabled or disabled. However, this was not possible due to the difficulty in finding real life scenarios in which a large number of groups of users would be willing to undertake the same task and no ethical repercussions would have arisen. Hence a qualitative measure of users' experiences was gathered and compared with the different tools that they used in order to understand how awareness mechanisms affect the document development process.

## **1.4 Research Questions**

Since their introduction by Ward Cunningham (Leuf and Cunningham, 2001), wikis have achieved success in a number of problem domains (Cunningham & Cunningham). The wiki concept incorporates a number of features which intrinsically aid the collaborative authoring process. Examples include:

- A centralized point of access for the document, eliminating confusion caused by multiple versions of a document (Weng and Gennari, 2004).
- The common presence of a version history function, enhancing author awareness of changes to the document (Weng and Gennari, 2004).
- Inherent support for discussion, either through dedicated discussion pages or simply via inline comments.
- Universal and worldwide access from any computer with access to the web.

Despite these advantages, the use of wikis has nonetheless been generally restricted to a subset of applications within the larger field of collaborative authoring. The outcomes of this research will aim to determine if a wiki-based system with enhanced features can enhance the effectiveness of online authors' awareness. In particular, this research seeks to answer:

1. *What kinds of awareness mechanisms are more effective in each stage of the document development process? In particular we seek to understand:*
  - 1.1 What are the motivations behind this need?
  - 1.2 How can these awareness mechanisms enhance each stage of the document development process?
2. *What is the importance of planning in the collaborative authoring process?*
  - 2.1 How do participants in a collaboration approach the planning process?
  - 2.2 What kind of information is included in the planning process?
  - 2.3 How do they keep the plan up-to-date?
3. *What effect does the planning process in a collaborative authoring process have on interactions between participants?*
  - 3.1 Is it sufficient to keep a version of the plan online?
  - 3.2 What effect does a well-structured plan have on the collaborative process?
  - 3.3 What problems commonly encountered during the collaborative authoring process are due to planning?
  - 3.4 What kind of interface design interaction mechanisms are needed to overcome these problems?

4. *How do users' awareness' needs compare to the tool that they actually use?*
  - 4.1 Do users perceive the importance of awareness mechanisms differently depending on the tools that they have used?
5. *What type of awareness mechanism has the greatest effect upon the success of the collaborating authoring?*
6. *Do online tools which support users' awareness improve the collaborative authoring process?*

## 1.5 Hypothesis

With respect to the research questions, this research hypothesises the following:

- **Hypothesis 1:** Different awareness mechanisms are more effective in different stages of the collaborative authoring process.
- **Hypothesis 2:** Planning is crucial to the success of the collaborative authoring process and improves collaboration and coordination between authors.
- **Hypothesis 3:** The addition of awareness mechanisms to the collaborative workspace can help to improve the collaborative authoring process.

## 1.6 Contributions of the research

This research aims to develop techniques to improve interaction and productivity in collaborative authoring systems by enhancing the awareness of users. Experiments have led to a greater understanding of the quantitative and qualitative effects of communication and awareness mechanisms on collaboration.

This thesis documents several key contributions made to the field of Computer Supported Collaborative Work, Human Computer Interactions and Web Science.

**Contribution 1:** From a literature review and a user study, this research identifies common problems that occur during the collaborative authoring process. Specifically, these problems are mapped to the absence of awareness mechanisms within the collaborative environment (Chapter 2 and Chapter 3).

**Contribution 2:** Using a large-scale statistical analysis of a real world collaborative authoring system, this research demonstrates that when communication mechanisms

are properly used, desired results are achieved sooner than when they are not used (Chapter 4). This is used to estimate how types of communication and coordination mechanisms affect the collaborative authoring process, so that common problems can be avoided in the prototype system created by this research.

**Contribution 3:** This research identifies the crucial aspects of planning within the co-authoring process and the effect that they have on coordination mechanisms within the group (Chapter 9). This finding outlines the need for “progressional awareness” (Liccardi et al., 2009) which is defined as:

1. The connection between planning activities and up-to-date knowledge of the status of other users.
2. Knowledge of how the document is planned and how the document is progressing with relation to the plan.

Progressional awareness is an intersection of group, informal and workspace awareness since it deals with the information related to how roles and responsibilities affect the up-to-the-date status and progress of the document itself. For example, progressional awareness deals with the information about which sections have been completed, who completed them, the sections that have not yet been completed, who is supposed to complete them by what date, and the level of detail required. It deals with document status and the progress of users with respect to tasks assigned to them.

**Contribution 4:** This research identifies different types of awareness required at different stages in the document development process. The differences are identified and the importance of different awareness and communication mechanisms are explained, depending on the state and stage of the document throughout the development process with respect to feedback and knowledge required by users at each stage (Chapter 9).

**Contribution 5:** We show that users choose co-authoring tools according to the way they intend to interact with each other (varying from face-to-face to completely online, from centralized to decentralized), underlining the need for different coordination and communication mechanisms within a collaborative authoring tool to address the varying needs of different group profiles and attitudes (Chapter 9).

**Contribution 6:** This research identifies the effects that the presence and absence of awareness have on the document development process. It is shown that some problems

experienced by users of offline writing tools can be completely avoided by the use of online tools that provide better awareness (Chapter 9).

## **1.7 Declaration**

This thesis describes the research undertaken by the author while working within a collaborative research environment. This report documents the original work of the author.

I, Ilaria Liccardi, confirm that this work submitted for assessment is my own and is expressed in my own words. Any uses made within it of the works of other authors in any form (e.g. ideas, equations, figures, text, tables, programmes) are properly acknowledged at the point of their use. A full list of the references employed has been included.

## **Chapter 2 Collaborative Authoring**

This chapter lays the foundations for analysis of the collaborative authoring process. Different types of collaboration are described, such as traditional face-to-face, hybrid (combining face-to-face with online methods) and integrated online collaborative authoring where multiple functions are combined and supported. These different approaches are presented in order to analyse and understand the repercussions of each.

Past research has analysed how collaborative authoring is typically performed. A review of past research is presented to highlight the most important findings and to give insight into the process. The examined sources reveal a common pattern in how collaborative groups are structured. It shows the typical relationships between group members and common communications mechanisms that group's members use. The stages that the document development process passes through in developing the complete document are explained, and common problems are identified that occur during collaborative authoring.

To confirm these findings, the results are presented of a field study into collaborative authoring that was performed as part of this research. The findings of the field study are linked to past research on the subject.

### **2.1 Introduction**

It is now rare for a paper or document to be written by a single individual. Universities often teach with a focus on group activities and group training is common within industry. The popularity of research collaboration is easy to understand. It brings complementary backgrounds to a project, results in more publications, and educates graduate students and junior faculty members (Floyd et al., 1994).

There have been numerous studies and past research focused on different approaches to collaborative activity – including traditional face-to-face and online collaboration. While staying aware of others is something taken for granted in the everyday world (traditional face-to-face collaboration), maintaining this awareness has proven to be difficult in real-time distributed systems (online collaboration) where information resources are poor and interaction mechanisms are foreign (Gutwin and Greenberg, 2002).

Online writing differs greatly to traditional group writing. This is because individuals not only need to write their own contributions but must also make them coherent with the contributions of other group members. Group members may be unfamiliar with each other and activities may be difficult to coordinate due to different time zones or work habits.

Collaborative writing offers numerous advantages over individual writing: it can reduce individual effort, provide new ideas and viewpoints, and improve the overall quality of the document by allowing separate sections to be written by experts in different fields (Noël and Robert, 2004a). In order to address participants' needs, developers of groupware applications need to better understand collaborative writing activities (Noël and Robert, 2004a, Posner and Baecker, 1992). Collaborative writing requires not only an editor that allows participants to work together but also enhanced functionality to allow participants to properly understand the development stage of the group writing activity.

## **2.2 Traditional collaborative authoring**

Collaboration type covers three areas: the form of collaboration (e.g. peer similar vs peer different, mentor vs mentee); collaborative activities (creating documents, generating ideas, making decisions); and collaboration size (the differences between small groups and very large ones).

### **2.2.1 The form of collaboration**

Traditional collaborative writing can take several forms based on the ties between and location of authors. Thagard (Thagard, 1997) identifies different types of collaboration, reflecting on the different backgrounds and roles of the collaborators. These collaborations are either *dominant* relationships (e.g. employer/employee or

teacher/apprentice), *peer* relationships (e.g. among researchers of similar knowledge, background and level) or *peer-different* (i.e. researchers from different disciplines who have similar goals but different knowledge and skills).

1. **Employer/employee:** the weakest form of collaboration, in which an employee performs tasks under the direction of the employer. The employer is capable of performing the tasks, but delegates to the employee for efficient use of time.
2. **Teacher/apprentice:** similar to the employer/employee scenario, apprentices also aim to learn relevant skills from the teacher. This type of relationship is particularly useful for imparting certain complex skills that are better learned from direct experience of working with experienced researchers, as opposed to a book or lecture.
3. **Peer-similar:** researchers often find it useful to collaborate with other researchers with similar interests and knowledge. One notable historical example is the discovery of the structure of DNA by Watson and Crick. The researchers may have slightly different fields of expertise in order to bring a wider variety of skills and knowledge to the partnership.
4. **Peer-different:** researchers working in different subject areas may collaborate toward a common goal. Examples include cognitive science, which may involve a psychologist and a computer scientist, or physics collaborations between theoretical and experimental physicists.

These categories are not absolute. A real collaboration might include aspects of two or more different types of collaboration. Furthermore, authors can be either collocated in the same place or disperse over different locations. However, in all cases, the participants know each other personally.

### 2.2.2 Collaborative activities

As identified in Section 2.2.1, the type of activities performed by a group will vary depending on the form it takes.

Task lists are a popular and effective technique in teams with a division of many roles and responsibilities. Lists are appropriate because in such a group, the majority of activities are task-oriented, and group members can identify individual responsibilities, deadlines and progress by other members. A task list can help to identify incomplete



tasks that are holding up overall progress of the project. They can be supplemented through the use of workflow features such as alerts (eg. email alerts).

In other group structures, different techniques are more appropriate. In a master-apprentice relationship, for example, work is centred on meetings and the apprentice learning from the master. While the case of master-apprentice is a 1:1 relationship, the teacher-student case differs in that it is a 1:many relationship. Activities are based in the learning environment: for example, prerequisite courses that the student must have previously completed, information about where the activities will take place, and surveys to provide feedback to the teacher.

### **2.2.3 Collaboration Size**

The number of participants in a collaboration must inevitably affect the requirements for the collaborative workplace. The requirements for a group of seven, twenty or a hundred people are vastly different, for example. As the group size increases, the quality of content will increase up to a point, and then decrease as communication becomes focused on communication from the few to the many.

However, the effect of an increasing team size does depend on the type of group (Section 2.2.1) and activities (Section 2.2.2). For example, in the case of a large team, diversity increases as the number of team members grows, potentially introducing additional content and activity and coordination requirements. This can potentially introduce problems when attempting to maintain a creative network. By contrast, in a large peer group, a hierarchical group structure is the more likely result, with the group activities dominated by the most experienced and vocal group members. A peer group therefore has the potential to grow more quickly and effectively than a team.

### **2.2.4 Gains and losses due to collaboration**

Motivation behind parties involved in face-to-face collaboration can be measured in acquisition of tangible rewards in the form of a job as part of group project, an article publication, a grant funding. These advantages or disadvantages are different for different types of collaboration, and hence produce different gains and (occasionally) losses.

In the *employer/employee* relationship, gains and losses can depend on the experience

of the employee. As Thagard (Thagard, 1997) identifies, when a scientific researcher hires an employee such as a laboratory technician, research assistant, or computer programmer, it is probably unreasonable to expect increased reliability. With good employees, potential losses in reliability are compensated for by gains in overall power, speed, and efficiency. The effect should then be that the researcher gets more desired results (combined efforts), and gets them faster (speed). Hiring an employee increases the cost of research, thereby potentially reducing efficiency, but not nearly as much as hiring an additional researcher.

In *teacher/apprentice* relationship where the apprentice is a graduate student, gains and losses are different from the *employer/employee* relationship as graduate students are an essential part of the research team. In a research scenario, work can often be delegated by researchers to less experienced workers such as graduate students. In the small scale, this can lead to a reduction in reliability as the apprentice will likely have less knowledge and experience than the researcher himself. However, this technique can lead to an overall increase in efficiency and ultimately faster results, as labor-intensive or time-consuming tasks can be delegated, allowing the researcher to concentrate on other matters.

In many respects, teacher/apprentice relationships are similar to employer/employee relationships. However, there are also fundamental differences between the two. While much can be learned from textbooks and lectures, effective knowledge of experimental techniques can typically be gained only from direct experience, and particularly from someone else who has such experience. The practice of science includes understanding of the design of experiments, construction of apparatus and how to interpret complex data; all of which are gained from experience. Thus, apprenticeships aim to produce new and effective researchers, rather than simply increasing efficiency of the work being performed.

In peer-similar collaborations, researchers with similar knowledge and interests are brought together. Mistakes can be avoided as other members of the team are able to identify mistakes that might not be noticed by a sole researcher. However, this type of collaboration can have detrimental side effects; research has identified that *groupthink* can lead to overconfidence in the actions of other collaborators (Janis, 1982).

In a peer-similar collaboration, contributors are most likely to gain from each other in

conceptual and theoretical contributions, as opposed to employer/employee or apprentice relationships, where gains are more likely from experimental work. Collaborators with similar levels of knowledge and experience can also benefit from being able to produce novel experimental designs. Researchers may also choose to collaborate for personal reasons; some researchers develop ideas better in conversation than by individual thought.

In peer-different collaborations, researchers from different fields can combine work to produce more reliable and robust results. However, this can be problematic in some scenarios, if different techniques or methodologies are used in different fields that are unfamiliar to researchers from other fields. Collaborators can therefore find difficulty in evaluating the work done in other fields.

### **2.2.5 Field Study**

In the past there have been numerous studies into collaborative authoring practises (Noël and Robert, 2004a) (Baecker et al., 1993), (Beck, 1993), (Beck and Bellotti, 1993), (Couture and Rymer, 1991), (Dillon and Maynard, 1995), (Ede and Lunsford, 1990), (Grudin, 1992), (Kim and Eklundh, 1998), (Kim and Eklundh, 2000), (Newman and Newman, 1992), (Rimmershaw, 1992), (Sharpley et al., 1993), (Tammaro et al., 1997), (Whitehead and Golland, 1999). Due to advances in technology such as faster Internet connections, video conferencing software, new tools and the introduction of wikis, a field study was conducted in order to identify the types of tool commonly used for collaborative authoring, common problems that affect authors engaged in collaborative authoring activities and technologies used to exchange information.

To gain insight into the attitudes, experience and perceptions of co-authoring activities, a field study was made of participants from Computer Science and Engineering backgrounds. The field study consisted of two stages. The first stage<sup>1</sup> consisted of interviews with 14 participants from academia and industry, who had collaborated with other people both in collocated environments and from disperse locations (supported by meetings - either online, or face-to-face). This helped to give insight into the stages of

---

<sup>1</sup> For this study we asked participants to describe the different stages of development and interactions during one of their most recent collaborative authoring activities.

the co-authoring process. In the second stage<sup>2</sup>, 53 more participants were interviewed in order to identify common problems which are encountered. Questions in the interviews required free-form answers to allow respondents to express themselves.

### **2.2.5.1 Stages of development and problems related to communication mechanisms**

In face-to-face collaboration, different methods of communication can be used in order to address specific stages within the document development process.

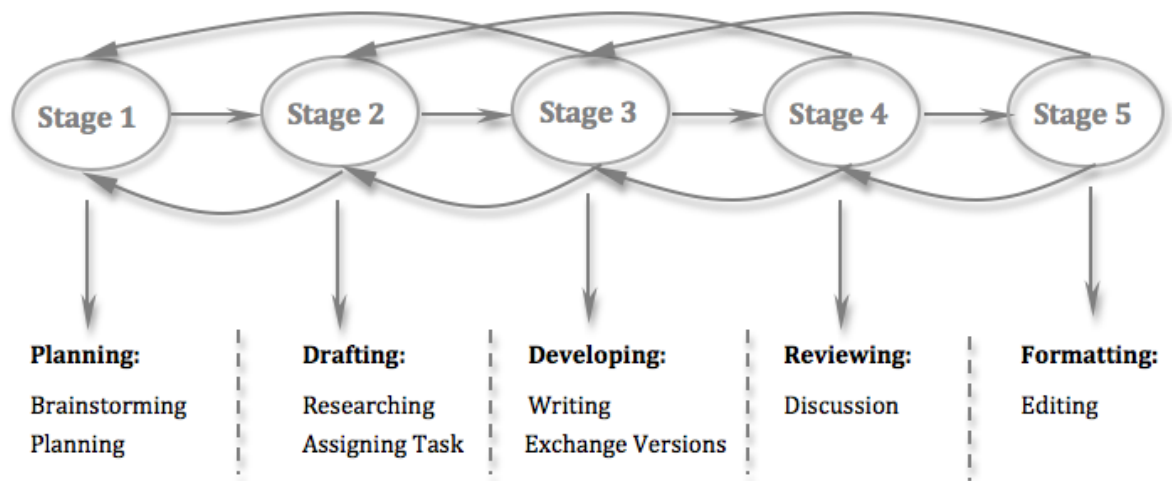
The first stage of the field study consisted of interviews with 14 participants: 6 academics from the same university and 8 professionals from two different companies. The interviews demonstrated the different ways participants approach the collaborative writing process, using different levels of communications at each stage of the document development process.

In all interviews the process can be subdivided into different stages depending on the activity to be completed (Liccardi et al., 2007). These are:

- **Planning:** In the planning stage, authors brainstorm ideas for possible content in order to plan the activities that need to be done to fulfil the objectives of the document.
- **Drafting:** In the drafting stage, authors investigate related research and decide on the content that each section should contain in more detail.
- **Development:** In the development stage, each author writes the different sections to which they have been assigned. Different versions of the document can be exchanged among contributors to see the current state of development.
- **Review:** In the review stage, authors discuss each others' sections in order to provide feedback and suggest improvements. Typically, in this stage repeated content is merged together, the work done by a single author for coherency.
- **Formatting:** In this stage, the document is formatted by a single author in order to conform to necessary stylistic/formatting guidelines for submission.

---

<sup>2</sup> For this stage we asked participants to describe the problems and motivations behind these problems during a recent collaborative authoring activity.



**Figure 2.1:** Document Development stages

These stages can occur iteratively, depending on the stage of the document. For example, after drafting the document, authors may realise that additional sections need to be added or removed in order to adjust the content of the document. The authors would then return to the planning stage and brainstorm the content and plan it again (Figure 2.1).

The results of the interviews yielded similar results to those found in past research (Noël and Robert, 2004a) (Baecker et al., 1993), (Beck, 1993), (Beck and Bellotti, 1993), (Couture and Rymer, 1991), (Dillon and Maynard, 1995), (Ede and Lunsford, 1990), (Grudin, 1992), (Kim and Eklundh, 1998), (Kim and Eklundh, 2000), (Newman and Newman, 1992), (Rimmershaw, 1992), (Sharples et al., 1993), (Tammaro et al., 1997), (Whitehead and Golland, 1999) that the most common methods of communication are face-to-face meetings, phone, fax, emails, mail and instant messaging. In the field study, it was found that users associate each method of communication with particular stages in the document development. This is summarized in Table 2.1 below.

COMMUNICATION METHODS	DOCUMENT DEVELOPMENT STAGE OVERVIEW	DEVELOPMENT STAGE
Face-to-face meeting	Used to decide the overall structure of the document.	Stage 1
	Used to assign particular tasks to authors.	Stage 2
Telephone Conversation	Used for clarification of comments and sections of the document which are not understood.	Stage 3

	Used to review parts of the document together when draft copies are distributed (stage 4).	Stage 4
<b>Fax</b>	Used to distribute the document with hand-written annotations. This is most common when more than one author is reviewing the document.	Stage 4
<b>Email</b>	Primary method of exchanging copies of the document, with discussion contained as part of the body of the email.	Stage 3
<b>Traditional Mail</b>	Used to exchange printed versions of the document with or without hand-written annotations.	Stage 3
<b>Instant Messaging</b>	Used for fast communication.	Stage 3
	Clarification of the document and comments	Stage 4

**Table 2.1:** Purpose of different types of communication and the associated development stages.

Table 2.1 shows the set stages of the document development. This process would be greatly simplified if authors could each write their individual sections in parallel, with integration in real-time. In this way, it would be possible for authors to be immediately aware of what other authors are doing and for them to accommodate each others' needs.

This field study showed that the major tools used for collaborative writing are Microsoft Word and Latex, with document exchange performed using email or a version control system. It was found that collaborative authoring in academia usually took the form of papers written for publication for conferences, journals or bids for government funding. Conversely, industrial work took the form of technical documentation for projects. Because of the need for publication, a different level of rigour was applied in industry and academia. In 76% of cases, major milestones were preceded by face-to-face meetings in which authors could discuss the document and make major changes as a group activity. In the remaining 24% of cases, face-to-face meetings were not required as document development was centred around a single author delegating work to other contributors.

In the second stage, 53 more participants were interviewed (PhD students, academic researchers and industrial researchers) to give further insight into the problems that participants in collaborative authoring activities encounter. Several common problems were encountered, many of which have also been highlighted by previous research (McDonald et al., 2004, Noël and Robert, 2004a, Noël and Robert, 2004b, Posner and

Baecker, 1992, Weng and Gennari, 2004). The following is a list of identified problems, which are summarized in Table 2.2. The percentage of users that reported such problems is shown in order to underline the numbers of users who encountered such problems.

1. **Communications degradation** (McDonald et al., 2004, Weng and Gennari, 2004): Participants in collaborative authoring rely most commonly on the use of email as the primary communications medium. Most email clients are deficient in representing threading of conversations; this is exacerbated when many participants are involved. In the authoring process, there typically exist multiple discussions and threads of conversation. This makes tracking a topic and the development of ideas difficult as the document develops. The common use of email attachments results in further work to explain changes to a draft.
  - 1.1. **Misinterpreted comments referring to sections:** 60.3% of participants in the study reported deficiencies in email clients when representing threading of conversations. This makes tracking a topic and the development of ideas difficult as the document develops. It is often difficult to trace how suggestions to further improve the document are incorporated as it develops. Misunderstandings commonly occur as authors are unable to see the reasoning behind the changes that the document undergoes (Weng and Gennari, 2004).
  - 1.2. **Discussions on specific sections are inadequately supported** (Weng and Gennari, 2004): There is often very little or no effort made to supporting group discussion. 22.6% of participants experienced problems with support for document discussion when using available technology. Consequently, conflicting revision suggestions are often difficult to resolve. It is often complicated for writers to track progress of any group without the appropriate notifications having been made.
  - 1.3. **Poor support of annotation** (Weng and Gennari, 2004): Writers often make use of comments for a variety of purposes; for example, to ask questions or suggest changes to draft versions of a document. 39.6% of participants identified problems with existing annotation systems. They noted that heavy use of annotations could cause the document to

become cluttered, making it difficult to read or track the reasons behind comments.

1.4. **Tracking of topics** (Weng and Gennari, 2004): As shown above, writers often have poor annotation tools, they may resort to writing comments in a separate note and emailing the note to the different group members. This leads to many emails reporting problems due to issues raised in comments and can cause a communications overhead to develop from comments. 47.1% of participants reported confusion to which comments apply to which versions as it develops in multiple versions of the source document.

1.4.1. **Duplication of effort:** 22.6% of these participants suggested that two authors might end up doing the same work due to poor support for tracking of required changes. Different conflicting versions of the document can result, since points might have already been made which suggested another change.

2. **Tracking previous versions** (Noël and Robert, 2004a): 24.5% of participants reported issues in tracking existing and previous version of the documents. Examples were given related to auditing, as users may be required to keep backup copies of previous versions of the document .
3. **Version Control Issues** (Sun et al., 2004): 39.6% of participants identified problems in identifying the latest version of the document. If two people are editing the same section of a document, version conflicts can arise when the document is redistributed through emails or uploaded in a common directory. For example, it may not be clear which version is the correct version to use.
4. **Merging Conflicts** (Sun et al., 2004): Merging different versions of a document can become problematic if multiple users edit concurrently. 49% of participants found that this was a major problem in their development process. In one occasion this became so frustrating that a locking system was implemented in order to lock the document completely. This lead to more problems as users were forced to work around each others' locks. The solution used was to split the document and merge it later.



Participants		22	13	18	53
Issues		PhDs	ARs	IRs	Total
Communication Degradation	Misinterpreted comments referring to sections	12	6	14	32
	Discussions on specific sections are inadequately supported	1	4	7	12
	Poor support of annotation	2	4	15	21
	Tracking topics to versions	4	8	13	25 (12)
Tracking previous versions		3	4	6	13
Version Control Issues		5	9	7	21
Merging Conflict		4	8	14	26

**Table 2.2:** Breakdown of participants' responses

The aim of this survey is to determine the types of technologies used by users engaged in online collaboration for work, and hence the types of problems that affect authors. The percentage of users that reported problems is shown in order to identify the numbers of users in the study who encountered such problems. From the field study it can be deduced that the three most encountered problems in collaborative authoring depend on the technology used by the users. In the field study it can be seen that the lack of support for threaded conversation within communication technologies is a major problem, since it was reported by 60.3% participants. This problem occurs since the document and the comments are maintained separately with different technologies.

Problems with merge conflicts were reported by 49% of the participants, since users within the user study used version control systems and emails to distribute the document. This type of problem is inevitable when using this type of distribution mechanism. Tracking topics across versions is another problem reported by 47.1% of participants, again due to the technology used by the group of users.

Other problems have been encountered in a lower rate and significantly less common than the previous three. Examples are Discussions on specific sections are inadequately supported (22.6%), Poor support of annotation (39.6%), Tracking previous versions (24.5%), Version Control Issues (39.6%).

## 2.3 On-line Collaborative Authoring

### 2.3.1 The mode of collaboration

Collaboration in an online community differs greatly from traditional face-to-face collaboration. This mainly happens because ties between authors are different. For

example, online projects such as Wikibooks and Wikipedia differ greatly to traditional methods such as those identified by Thagard (Thagard, 1997). One of the most significant differences is that content written for a site such as these is in a state of continuous development, rather than having a fixed deadline in which to be completed. The motivation for contributors is also different; as the authors are volunteers, it is implied that they have a personal interest in contributing to and improving the content.

The spread of authors is also likely to be different. In a traditional collaborative activity, it can be reasonably expected that each of the authors has contributed a significant amount of text towards the final product. This assumption does not hold in an online collaboration such as Wikibooks or Wikipedia; as authors do not hold any obligation to contribute, their contributions might consist of only a few sentences, or even a single spelling correction.

The social relationships between authors are also necessarily different. In a traditional collaborative activity, collaborators usually know one another or have worked together in the past. By contrast, while communication is possible in an online environment (including Wikibooks), the communal ownership of the manuscript means that there is not necessarily any communication between authors at all.

The issue of coordination between users is also of interest. In a traditional collaboration, careful planning of the work is often of vital importance to the activity. By contrast, a careful plan is not necessarily of the same importance in an online collaboration where contributors are under no obligation to follow the plan. However, one observation with Wikibooks is that authors instead use implicit planning tactics. One example is that authors can create “broken” links within pages to other pages that have not yet been created. This encourages other authors to create the missing page. This technique can even be expanded to cover the planning of an entire book. Any form of collaboration must inevitably involve coordination and communication between contributors. For example Wikibooks and Wikipedia provide a talk page attached to every page where users can attach comments and participate in discussion related to the content of the page. This allows the authors to plan the task of writing the books and resolve any disputes that may arise.

### 2.3.2 Why collaborate?

Collaboration in an online environment does not provide the same tangible gains and rewards that traditional face-to-face collaboration does. Online collaboration is typically performed by volunteers who share a common interest. Gains in such environments might not be tangible but might include increase in knowledge of a subject as well being part of a larger community of users who undertake projects of common interest.

### 2.3.3 Communication mechanisms

Online communication differs from traditional communication as the identity of collaborators might not be known. Users must rely on communication mechanisms provided within the environment. In environments such as wikis communication is only possible through discussion pages, supported by a version control system in case of problems and the need to undo changes. There has been extensive research into how communication mechanisms affect the development process and therefore the outcome of the collaborative work. In (Kittur and Kraut, 2008) it was found that communication mechanisms can improve the quality of an article when used properly. They used Heckman regressions to predict changes in quality of 23,619 articles over a period of 6 months, based on initial quality, article age, the number of editors working on an article and the coordination techniques they used (Table 2.3). Four different models were derived that highlighted the correlations between these factors and the effects they have on quality.

	Model 1 # Editors			Model 2 Editor Concentration			Model 3 Talk Edits			Model 4 All		
	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P
Intercept	.442	.025	***	.329	.031	***	.362	.028	***	.304	.033	***
Initial Quality	-.149	.005	***	-.138	.005	***	-.151	.005	***	-.146	.005	***
Article Age	-.024	.004	***	-.005	.004		-.017	.004	***	-.006	.004	
# Editors	.053	.002	***	.003	.003		-.005	.003		-.020	.003	***
Editor Concentration				.791	.041	***				.600	.038	***
Editors X Concentration				.248	.017	***				.216	.020	***
Quality X Concentration				-.236	.032	***				-.222	.035	***
Age X Concentration				-.066	.027	*				-.041	.028	
# Talk Edits							.113	.005	***	.087	.004	***
Editors X Talk							-.003	.001	*	-.010	.002	***
Quality X Talk							-.012	.003	***	-.001	.003	
Age X Talk							-.009	.003	**	-.003	.003	

**Table 2.3** Nested lagged regression analysis of the number of editors, coordination metrics (editor concentration and talk edits), and article lifecycle on change in article quality. (Note: \*\*\* p<.001, \*\* p<.01, \* p<.05) as described in Kittur and Kraut (Kittur and Kraut, 2008)

## 2.4 Online and traditional collaboration compared

Working collaboratively in a group allows different contributors to participate in the activity. These contributors can offer different skills and points of view, which can improve the overall product (Clearwater et al., 1991, Hill G.W., 1982). Adding additional contributors increases the resources available to accomplish the group's goals, i.e. time, energy and expertise. However, this may also reduce productivity due to the need to coordinate the work of more people (Brooks, 1975). Online collaboration offers many potential benefits over traditional face-to-face collaboration, and is the only option in fields such as open source software development. Lerner (Lerner et al., 2006) argues that open source development is driven by "hobbyists" rather than corporations. One of the most successful examples of volunteer collaboration is the Linux kernel, which was developed by a large number of contributors interested in the subject (Raymond E.S., 2001).

In traditional collaborations, contributors are linked by professional ties. Thagard (Thagard, 1997) identifies different kinds of collaboration, reflecting on the different backgrounds and roles of the collaborators. These collaborations are either *dominant* relationships (e.g. employer/employee or teacher/apprentice), *peer* relationships (e.g. among researchers of similar knowledge, background and level) or *peer-different* (i.e. researchers from different disciplines who have similar goals but different knowledge and skills).

In an online environment, these classifications can be more difficult to define. Particularly in wiki-based collaboration, it may be the case that there is no direct interaction among authors at all (beside the document). In some circumstances, one or more users may take control of the activity; however, this may be related less to their skills or knowledge and more to their commitment to the activity. Past research has shown that in environments such as Wikipedia, the quality of the article can be influenced by certain metrics. In (Kittur A. Kraut R.E., 2008) the number of contributors was found to affect quality only when proper coordination mechanisms were used. In (Lih, 2004), the number of edits and unique editors to an article were suggested as metrics for quality, but no justification was provided.

Other characteristics such as factual accuracy (Giles, 2005) (although disputed in (Encyclopaedia Britannica Inc., March 2006)), credibility (Chesney, 2006), revert times

(Viegas et al., 2004), and formality of language (Emigh and Herring, 2005) have been used to assess small samples of Wikipedia's articles and in some cases compare them to articles in traditional encyclopaedias.

The example of the software industry, where collaboration is comparatively mature and coherent, might offer insights into effective designs for collaborative writing systems. It has been observed that in software projects coordination costs can overwhelm the benefits of added personnel; a well-known example in the domain of software projects is Brooks' Law, which argues that *"adding manpower to a late software project makes it later"* (Brooks, 1975). The more contemporary open source "bazaar" method stands in stark contrast to Brooks' Law, suggesting that it is important to *"delegate everything you can, be open to the point of promiscuity"* (Raymond, 2001). In the "bazaar" method it is argued that using as many users as possible will improve the software more than having a few developers working on it. Coordination can therefore strongly influence the quality of a product.

Groups of different sizes might need to be managed in different ways to achieve such quality. Hence coordination mechanisms are domain- and group size-specific. Creating an online resource such as a book on Wikibooks or an article in Wikipedia presents potential benefits as the authors volunteer to write the book/article and hence have a personal interest in and commitment to the activity (Lerner et al., 2006). However, as contributors in most cases do not present a traditional collaborative arrangement (of the sort identified by Thagard (Thagard, 1997)), coordination mechanisms vary depending on the contributors and the requirements of the task. These mechanisms can impact the quality of the outcome.

Kittur and Kraut (Kittur and Kraut, 2008) highlight the relationship between coordination and quality in Wikipedia content. Using Wikipedia, Kittur and Kraut show that the use of appropriate coordination mechanisms is essential for harnessing the wisdom of the crowds. They show that articles edited by many authors are generally better than those with fewer authors only when implicit coordination techniques are used. Similarly, coordination by direct communication between editors only improves articles with relatively few editors but actually harms quality when many editors are involved (Kittur and Kraut, 2008). Stewart (Stewart, 2006) describes how larger teams generally perform better when they are engaged in low coordination work than when engaged in tasks requiring a high degree of coordination. In fact Kittur and Kraut

(Kittur and Kraut, 2008) describe that the degree of coordination required between authors as varying depending on the task at hand.

## **2.5 Conclusion**

In this chapter it has been shown that different styles of collaboration exist. These are used for different purposes; for example, traditional collaborative activities are commonly used for writing in academia and industry, while contemporary tools such as wikis are used for online collaborations that typically have many more participants. These different styles of collaboration are structured in radically different ways.

The results of the field study have helped to confirm the observations of past research which identified how traditional collaborations are structured. Specifically, it has helped to give insight into the stages that take place during the document development activity, and the communication methods that are used at each stage. Differences have been highlighted between how collaboration is performed in industry and in academia.

Importantly, this chapter has identified common recurring problems that affect participants in collaborative activities. By analysing the causes of these problems, it should be possible to suggest effective solutions to them. The insight that the study has given into the collaborative process will assist in understanding how enabling and enhancing awareness in technologically facilitated collaboration can improve the productivity of the collaborative objective. The next chapter will explore issues of awareness in greater depth.

## Chapter 3 Awareness in Groupware

Having analysed the structure of group collaborations in Chapter 2 and the problems that it can entail, this chapter examines the concept of awareness that has been identified from past research (Dourish and Bellotti, 1992), and considers how it affects the interactions between group members.

Established research into awareness considers the information that authors maintain within the collaborative context, classifying awareness around five classic forms, each of which has different effects on the interactions between authors and the collaborative authoring process as a whole. The nature of each of the types of awareness identified by past research is described.

The concept of awareness provides a useful theoretical model for analysing the problems in collaborative authoring. It is shown that the problems identified from the field study in Chapter 2 can be understood in terms of awareness issues. It is suggested that the roots of most problems affecting collaborative authoring are therefore social in origin, rather than technological. Technological solutions can nonetheless be sought to the problems by designing collaborative authoring systems to take issues of awareness into account and thereby enhance social interactions.

### 3.1 Introduction

Awareness of individual and group activities is always required to coordinate activities within a collaborative process (Dourish and Bellotti, 1992). There has been significant research in the areas of Computer Supported Collaborated Work (CSCW), and Human Computer Interactions (HCI) into the role of awareness in groupware systems. Dourish and Bellotti (Decouchant et al., 1995, Dourish and Bellotti, 1992, Greenberg et al., 1996, Gutwin and Greenberg, 2002, Mitchell et al., 1995) define awareness as *“understanding of the activities of others, which provides a context for your own*

*activity*". Gutwin and Greenberg (Gutwin and Greenberg, 2002) examine workspace awareness as a combination of the types of awareness that are present in daily life (Greenberg et al., 1996), but within an online system as "*the up-to-the-moment understanding of another person's interaction with a shared workspace*". This includes knowledge about what others are working on, what they are doing and their future plans (personal, social, group and informal awareness).

Awareness of participants' activities with respect to a collaborative context is therefore a critical issue for collaborative authoring systems (Dourish and Bellotti, 1992). Awareness can take several forms and can affect the group members as individuals or as a group. In this chapter we report a summary of the types of awareness that have been categorized by different researchers in the HCI and CSCW fields (Decouchant et al., 1995, Dourish and Bellotti, 1992, Greenberg et al., 1996, Gutwin and Greenberg, 2002, Mitchell et al., 1995).

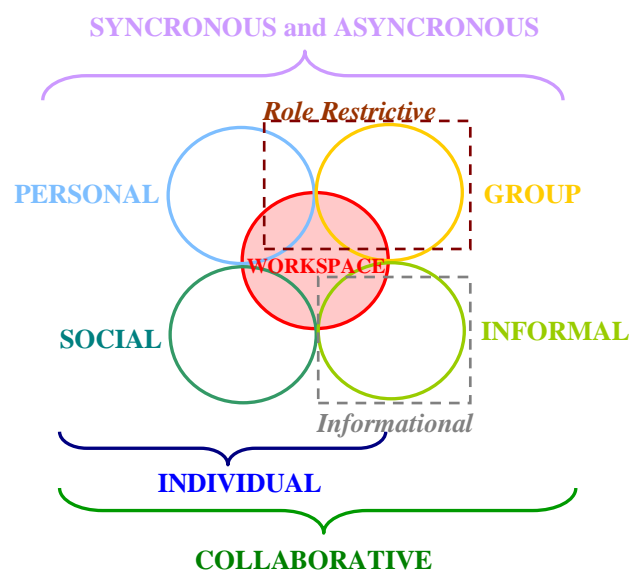
### 3.2 Types of awareness mechanism

Existing CSCW systems vary in the methods they use to support awareness: *informational* and *role restrictive*. In the informational method, explicit facilities are provided through which collaborators inform each other of their activities (Dourish and Bellotti, 1992). Informational methods allow users to identify changes made by others since the last time they used the system (Brush et al., 2002). Role-restrictive methods arise from explicit support for roles in collaborative systems; a role describes an individual's relationship to the shared work objects and to other participants (Dourish and Bellotti, 1992). Awareness can be synchronous and asynchronous; synchronous when authors are working together in real time, asynchronous when not working together in real time.

1. **Personal Awareness** refers to information that users maintain about themselves and their roles in the group. It can be synchronous (eg. current whereabouts within the system) or asynchronous (eg. where the user has been within the system) (Greenberg et al., 1996, Mitchell et al., 1995).
2. **Social awareness** refers to information that users maintain about others in social or conversational contexts: for example, whether a person is paying attention, their emotional state or level of interest. This kind of awareness can be asynchronous (eg.



- knowledge that a partner has replied to a comment), or synchronous (eg. actively paying attention to or replying to the queries of other users) (Greenberg et al., 1996).
3. **Informal awareness** involves knowledge of who is present and what they doing: the kind of awareness implicitly present in an office context (Greenberg et al., 1996). Morán et al (Morán A. L., 2001) refer to it as *informational awareness*.
  4. **Group awareness** gives an overview of other users' roles, activities, movements and status in the process. It includes peoples' roles and responsibilities, their positions on issues, current status, and group processes (Decouchant et al., 1995, Dourish and Bellotti, 1992, Greenberg et al., 1996).
  5. **Workspace awareness** concerns user presence in the workspace and what users are currently doing: up-to-the-minute knowledge about other people's interactions with the shared workspace. In face-to-face activities, workspace awareness is naturally present. Workspace awareness combines the types of awareness that people maintain when working in a group (Greenberg et al., 1996, Gutwin and Greenberg, 2002) (Figure 3.1) .



**Figure 3.1:** Different types of Awareness (modified image from (Gutwin and Greenberg, 2002))

### 3.2.1 Personal Awareness

Personal Awareness is the awareness that users maintain about themselves and their roles in the group. When a user is working independently, this awareness is based on memory, meaning that a user must remember what they are doing and extrapolate what

they are going to do next. However, when this information relates to other users within the collaborative environment this information is not easy to maintain. Although a person can know their task and role, this could change according to other users and tasks. For example if a user is undertaking an action which will require editing or modifying another user's text, this will mean that this particular user's task and role might be modified either for a certain amount of time or for ever meaning that another task and roles needs to be assigned to him. Hence maintaining this kind of awareness of own roles, responsibilities, tasks and whereabouts must be correlated with the other users within the group.

### **3.2.2 Informal Awareness**

Informal awareness refers to the implied knowledge that is available to people when they work together in the same building (e.g. (Bradner and Mark, 2002, Dourish and Bellotti, 1992, Kraut et al., 1990). Examples of this type of awareness include knowledge of who is present and their immediate availability. Informal awareness is often maintained as a result of day-to-day activities. Because of its nature, informal awareness has been traditionally dependent on physical proximity between participants.

The presence of informal awareness brings a number of important benefits. For example, it has been shown (Bradner and Mark, 2002) that collaboration is adversely affected when the distance between collaborators within a building increased, and that this is related to the number of casual interactions between the collaborators.

Multiple attempts have been made to attempt to replace informal awareness within distributed online environments, where these types of implicit interactions do not occur. For example, video links between contributors known as media spaces have been attempted (e.g. (Dourish et al., 1996, Dourish and Bly, 1992, Nakanishi et al., 1996)) as well as video conference systems (e.g. ((Roussel, 2002, Tang et al., 2004)).

### **3.2.3 Social Awareness**

Social Awareness arises from behavioral cues that occur during face-to-face conversation. These cues include facial expressions, eye contact and verbal intonation, and allow participants in the conversation to recognize that it is proceeding smoothly (Gutwin and Greenberg, 2002). Participants use this awareness to adjust their behavior as necessary.

Various different types of information are provided by social awareness. In the most basic sense, it allows a participant to determine mechanical questions as whether other people have heard and understood what he has said. For example, certain speech patterns indicate whether a speaker is willing to give up the floor (McLaughlin, 1984). Similarly, gestures and *back-channel responses* are used by people to indicate that they have understood what was said (Cassell et al., 1999). Social awareness allows participants in a conversation to determine whether other participants are paying attention to what is being sent, particularly through monitoring of eye contact (Goodwin, 1981).

In a deeper sense, social awareness allows participants to judge affective questions, such as whether other people believe what they have said.

Teleconferencing systems can reproduce some of the information gained via social awareness (Ou et al., 2003), provided that video and audio are of a sufficient quality. However, some information can still be lost. Another particular deficiency is the inability to determine eye contact due to the fact that the monitor and camera are typically offset. While subtle, these issues are significant enough to adversely affect the social protocols of normal conversation. This can become more problematic as the number of participants increases.

### 3.2.4 Group Awareness

Group awareness concerns the knowledge that participants maintain, while working in a group, of the group's structure. This includes information about the roles of participants and power structures within the group. While small groups often use more freeform dynamic structures, more formal structure is typically needed to coordinate larger groups. Group awareness is therefore essential for efficiency, particularly in scenarios such as formal meetings.

Different techniques have been used in groupware systems to attempt to incorporate this type of knowledge. Open source projects are commonly maintained through text-based communication (eg. mailing lists and chat systems) (Gutwin et al., 2004). Although this improves team performance, it has been proven to be effortful and time-consuming (Weisband, 2006).

### 3.2.5 Workspace Awareness

Gutwin and Greenberg (Gutwin and Greenberg, 2002) define workspace awareness as the “up-to-the-moment understanding of another person’s interaction with the shared workspace”. Despite the name, workspace awareness does not refer merely to awareness of the workspace itself, but also of other people and how they interact with it. Workspace awareness covers the visual workspace that the group is using to perform its task, and is limited to events that happen within this visual workspace. In this sense, workspace awareness is not covered by other types of awareness such as informal awareness that might otherwise describe it.

Workspace awareness is a special case of situation awareness, as a person working within a visual workspace such as a groupware tool interacts only with that workspace. However, in the case of a groupware tool, their situation awareness must expand to include awareness of the collaborative aspects of the scenario (Gutwin and Greenberg, 2002).

Groupware systems artificially reduce the information that can be gathered by participants, compared to a real-world scenario. Tasks that would otherwise be straightforward are made difficult by the participants’ reduced awareness. Problems of workspace awareness therefore concern the obtaining of this information. This includes ensuring that such information is presented in a useful form that helps promote awareness (Gutwin and Greenberg, 2002).

## 3.3 Discussion

With knowledge of the different types of awareness present between authors in a collaborative space, it is possible to frame some of the problems previously identified in Section 2.2.5 in terms of awareness issues.

Of the most commonly reported problems, issues related to communications degradation were some of the most widely reported, caused by participants experiencing confusion in determining the section of a document to which an email message refers. In a more general sense, this can be seen as a degradation in informal awareness; although email discussion may provide knowledge of the activities of other contributors, the lack of a direct link to the text being discussed implies that the depth and detail of this knowledge is restricted. Furthermore, if (as is common) work is

divided between authors, and an individual author assigned to write each section, these authors may not be aware that discussion affects the sections of text that they have been assigned to write. In scenarios such as these, the problem also potentially affects personal and group awareness.

Poor support for annotation was highlighted, where a communications overhead can develop when a large number of comments are used. This can be interpreted as a problem of workspace awareness; while working in a collocated scenario, comments may be discussed and resolved in a direct and instantaneous fashion, discussion is less fluid when using an asynchronous discussion system such as email. Version control issues where participants have difficulty identifying the latest version of the document can also be seen as a problem of workspace awareness, as participants are not aware of the up-to-the-minute status of the document.

It is often necessary to track previous versions of the document for auditing reasons; for example, a group of students working together may wish to confirm that each group member has contributed the amount expected of their role. In this scenario, informal awareness covers the historical information about the work that each group member has performed. This issue is also related to group awareness, as it involves knowledge of the roles, responsibilities and status of group members.

The issue of merging conflicts covers different types of awareness depending on the strategy used to deal with the conflicts. In general two methods can be used to deal with the problem of conflicting changes. Firstly, conflicts can be avoided entirely; in this case, coordination between members is needed to prevent conflicts from occurring; workspace awareness is necessary to provide the up-to-the-minute knowledge necessary in order to know what other group members are currently doing.

Secondly, conflicts can be resolved by merging changes from multiple sources to create a unified document. In this scenario, informal and group awareness are necessary to provide the necessary understanding of what each author has contributed so that their work can be combined.

In conclusion, awareness provides a useful model that can be used to classify and examine many of the common problems previously identified in collaboration authoring. Analysis such as this is essential in order to understand the root causes of the problems.

*Hence this research has identified common problems that occur during the collaborative authoring process. Specifically, it has been shown how these problems are mapped to the absence of awareness mechanisms within the collaborative environment (Contribution 1).*

### **3.4 Conclusion**

In this chapter it has been shown that awareness provides a useful model that can be used to classify and examine many of the common problems previously identified in collaboration authoring. The analysis performed in this chapter demonstrates how an awareness perspective can provide powerful insight and understanding of the root causes of problems and shortcomings of collaborative authoring interactions irrespective of whether it is conducted face to face or supported by technology. While in some cases, straightforward solutions to problems can be readily identified (eg. use of a version control system to eliminate confusion over multiple document versions), other problems require a deeper understanding of the social issues that are inherent to a collaborative process if effective solutions are to be devised.

Different types of awareness have different effects upon the document development process. Different types of awareness are also required for different tasks; for example, group awareness is required in order to understand the division of labour among authors, while informal and workspace awareness are needed to resolve merge conflicts. An effective and efficient groupware tool must therefore seek to reproduce each of these types of awareness within the collaborative workspace.

## **Chapter 4 Impact of poor communication and collaboration mechanisms - Example using Wikibooks**

The primary focus of past chapters has been on the effect of communications mechanisms within "traditional" collaborative writing activities, that is to say, a small group of people working together to develop a single document for publication. It was shown in Chapter 2 that online collaborative authoring websites such as Wikipedia have also become popular in recent years. This type of collaborative authoring is vastly different in aim, structure and method to traditional collaborative authoring. It is of interest because it also offers an unprecedented valuable source of data that evidences real-world, large-scale, collaborative authoring activities.

This chapter analyses the Wikibooks website, which allows contributors to work together to develop freely available textbooks. In contrast to a traditional collaborative writing activity, a typical book on the Wikibooks site might have hundreds or even thousands of contributors. It is shown that despite this difference of scale, communications and coordination mechanisms are still essential in order to achieve a high quality result.

To demonstrate this point, an analysis of successful books on the Wikibooks site is made, using textual redundancy within books as a measure for quality. Successful books are identified by using the set of featured books nominated by the editors of the site itself. By analysing patterns in the history of different books, it is shown that books reach featured status more quickly when richly featured communications mechanisms are used to coordinate development. This helps to further underscore the importance of communications between contributors in collaborative writing, regardless of the type of collaboration itself.

## **4.1 Introduction**

Technologies such as wikis have made it possible for authors from all over the world to engage in large-scale collaborative writing projects. Unlike traditional collaborative writing tools, systems such as Wikipedia enable strangers from diverse locations to collaborate on a shared topic, with associated advantages and disadvantages.

Manuscripts inevitably evolve during the writing process and certain discussions and arguments inevitably occur, due to different work practices, points of view, knowledge level, and writing experience among co-authors. The repetition of arguments has different effects, depending upon the current stage within the writing process. In the early stages, a certain level of redundancy is acceptable, but highly redundant text is not appropriate in the final product. Also, acceptable levels of redundancy differ depending on the activity. For example, scientific papers normally repeat information in the abstract and in the conclusion, whereas books might avoid such repetition.

This chapter examines the role of redundancy in large-scale collaborative writing projects. We base our research on a collection of books from the Wikibooks website that were written entirely on-line, with distributed co-authors. We have access to detailed historical data that permits us to analyse specific patterns of collaborative activity and their effects on the quality of the resulting book. We use argument repetition at different points in the writing process as a measure of the amount of similarity among arguments, at the level of paragraphs, sections and chapters.

## **4.2 Measuring Semantic Similarity and Redundancy**

Measuring the amount of redundancy in an article requires understanding the semantic meanings of the two sentences under comparison, which is still a very challenging task for modern computers. However, there are several different techniques for measuring the semantic similarity between two sentences, which can be used to provide an approximation to the amount of redundancy between them.

Semantic similarity between texts, which include documents, paragraphs, sentences and words, is actually extensively studied in existing literature, in particular in the area of natural language processing and information retrieval. The basic unit of measure for similarity between texts is words, and there are many word-to-word similarity measures based on string similarity (Islam and Inkpen, 2008), thesauruses (Pedersen et al., 2004)



or corpus statistics (Turney, 2001) In information retrieval, the “bag of words” approach is commonly used to measure similarity between documents. A document is usually characterized by a term vector of length  $n$ , the elements of which indicate whether a term is present in the document and its relative importance. Terms are usually weighted using the TF-IDF (term frequency-inverse document frequency) scheme (Kowalski, 1997). Similarity between documents depends on the cosine similarity between two vectors. Other methods have also been proposed to exploit word co-occurrence information instead of using exact word matching. For example, latent semantic analysis (Landauer et al., 1998) can be used to measure similarity between texts by computing higher-order word relations based on dimensionality reduction.

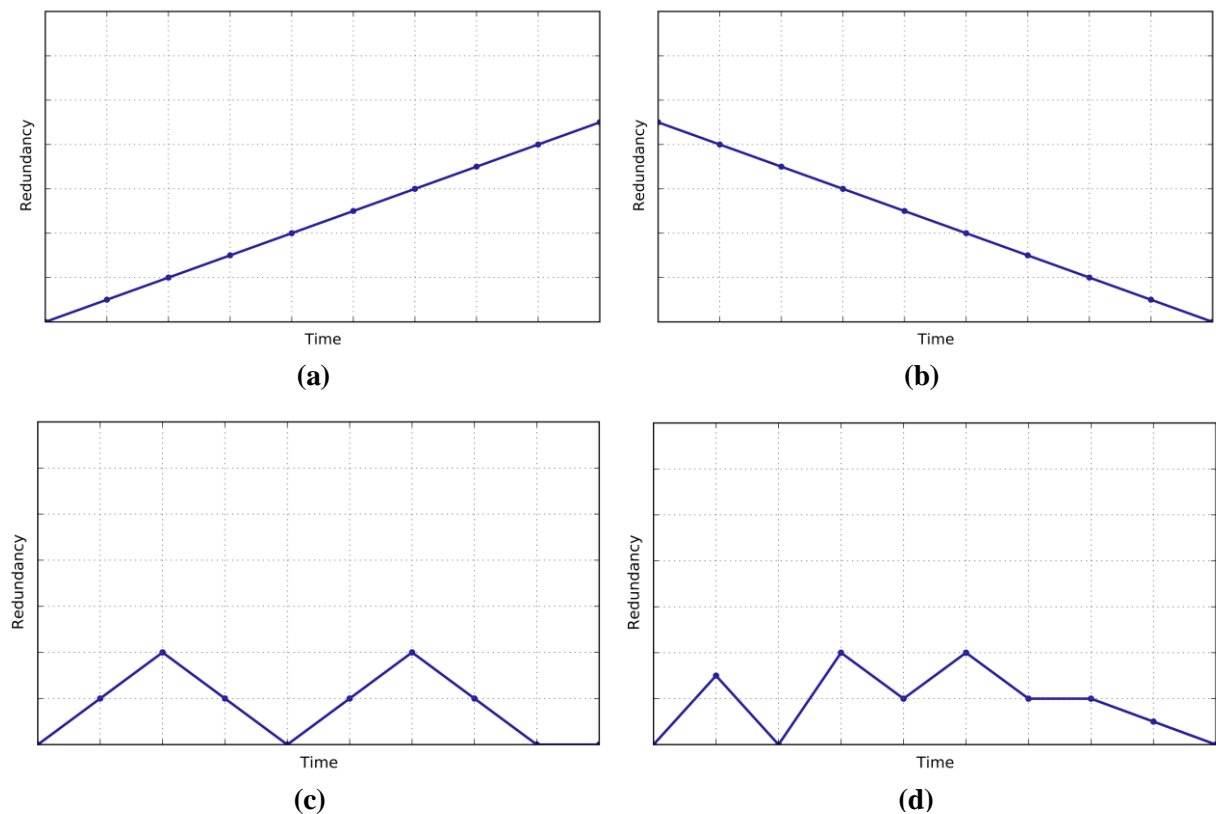
Various methods have also been proposed to measure semantic similarity between sentences by combining different techniques. For example, Li et al. (Li et al., 2006) propose a method that combines WordNet-based word similarity, corpus statistics and word order similarity. Islam and Inkpen (Islam and Inkpen, 2008) propose a similar approach based on substring matching of words, pointwise mutual information similarity, and word order similarity.

The problem of redundancy focuses on whether two sentences or paragraphs contain the same information. We are not aware of any studies that discuss this problem independently. Redundancy is sometimes discussed in automatic document summarization, in which redundancy in the summary or abstract produced for a document is undesirable. For example, Goldstein et al. (Goldstein et al., 2000) make use of cosine similarity of term vectors to measure redundancy between sentences extracted from the original document. Similarity, Li et al. (Li et al., 2009) attempt to reduce redundancy in generated summaries by incorporating sentence similarity and cluster memberships in the learning.

### **4.3 Conjectures of Redundancy Occurrences**

This research is interested in measuring redundancy within a manuscript over the period of time in which it is written. The redundancy pattern can be determined by analyzing the manuscript’s content at different points in time. We are particularly interested in the following issues:

1. **Changes in redundancy over time.** The amount of redundancy present in the manuscript usually changes over time during the process of collaborative writing. We are interested in understanding the different factors that correspond to the changes.
2. **Patterns of change in redundancy:** Change in redundancy over time may follow different patterns in different cases. This information is useful when we want to understand the stage of development of a manuscript. For example, it can reflect whether a book is close to completion or further revisions are needed.
3. **Effects of redundancy:** We are also interested in the effects of redundancy on the development of a manuscript. The presence of redundancy can be *constructive*, when there are different points of view on the same issue which can be merged in order to construct a more specific and appropriate result, or it can be *destructive*, when the redundancy involves only repetitive or conflicting content, which will need to be revised or deleted.



**Figure 4.1:** Possible patterns of changes of redundancy across time. (a) Redundancy increases over time. (b) Redundancy decreases over time. (c) Redundancy changes regularly over time. (d) Redundancy changes irregularly over time.

This research hypothesised that there could be several different possible patterns of change in redundancy. Redundancy within a manuscript can start from zero and increase over time (Figure 4.1(a)). In this case redundancy could be considered destructive as the amount present within the manuscript does not vary or diminish. This could mean that there has been a duplication of effort within the manuscript. In this case writers, might not be aware that this is happening. Similarly it can start at an arbitrary point if different users are collectively involved in the first draft of the document, and decrease over time (Figure 4.1(b)). In this case writers would have merge or delete the repetition.

Redundancy can increase and decrease over time regularly (Figure 4.1(c)). This is usually because contributions from different writers are merged and integrated into the manuscript by one or more leading writers at regular time intervals. Of course, this kind of restructuring and reorganisation may also happen at irregular time intervals, such as depicted in Figure 4.1(d).

## **4.4 Data Source**

### **4.4.1 Overview of Wikibooks**

In this research we analyzed data obtained from the English Wikibooks. Wikibooks<sup>3</sup> is a collaborative, wiki-based website started in 2003. The site is run by the Wikimedia Foundation, the same organization responsible for Wikipedia. However, unlike Wikipedia, the focus of Wikibooks is on developing free books, particularly textbooks for education (Sajjapanroj et al., 2007, Xiao et al., 2007). Since its beginning in 2003, the site has expanded to over 2000 books, the content of which is contributed by volunteers.

Wikibooks uses essentially the same MediaWiki software as is used by Wikipedia, in addition to the same "open editing" policy. However, it is considerably less popular than Wikipedia, likely due to its differing purpose in providing complete books. For example, while the largest article on Wikipedia is around 50,000 words in length, Wikibooks contains over 100 books that exceed this by word count. Although the main

---

<sup>3</sup> <http://en.wikibooks.org/>

focus of the site is on educational textbooks, the site includes books covering a wide range of subjects, including sports, religion, interpersonal relationships and even a guide to the Harry Potter series of novels.

The various books on the site are in varying stages of development. The editors of the site collectively identify books that are judged to be of a high quality, and these are added to a list of "featured books"<sup>4</sup>. Featured books are nominated using a democratic process in which any editors on the site can vote, and must also meet a set of basic criteria. As the content of the website undergoes constant change, the Wikibooks community actively removes the featured book status from books that are no longer judged to be of a suitable quality.

The complete history of each page on Wikibooks is stored in a database on the website, and it is possible to access every past revision of a page through the web interface. The Wikimedia Foundation provides downloadable versions of the database including the page history. The fact that Wikibooks is considerably less popular than Wikipedia makes analysis of its database more straightforward and less time-consuming.

Any form of collaboration must inevitably involve coordination and communication between contributors. Wikibooks provides a talk page attached to every page where users can attach comments and participate in discussion related to the content of the page. This allows the authors to plan the task of writing the books and resolve any disputes that may arise. Online projects such as Wikibooks differ greatly to traditional methods such as those identified by Thagard (Thagard, 1997). One of the most significant differences is that content written for a site such as Wikibooks is in a state of continuous development, rather than having a fixed deadline in which to be completed. The motivation for contributors is also different; as the authors of books on Wikibooks are volunteers, it is implied that they have a personal interest in contributing to and improving the book.

The spread of authors is also likely to be different. In a traditional collaborative activity, it can be reasonably expected that each of the authors has contributed a significant amount of text towards the final product. This assumption does not hold in an online collaboration such as Wikibooks; as authors do not hold any obligation to

---

<sup>4</sup> <http://en.wikibooks.org/wiki/Wikibooks:Featured-books>

contribute, their contributions might consist of only a few sentences, or even a single spelling correction.

The social relationships between authors are also necessarily different. In a traditional collaborative activity, collaborators usually know one another or have worked together in the past. By contrast, while communication is possible in an online environment (including Wikibooks), the communal ownership of the manuscript means that there is not necessarily any communication between authors at all.

The issue of coordination between users is also of interest. In a traditional collaboration, careful planning of the work is often of vital importance to the activity. By contrast, a careful plan is not necessarily of the same importance in an online collaboration where contributors are under no obligation to follow the plan. However, one observation with Wikibooks is that authors instead use implicit planning tactics. One example is that authors can create “broken” links within pages to other pages that have not yet been created. This encourages other authors to create the missing page. This technique can even be expanded to cover the planning of an entire book.

## 4.5 Description of Data for Analysis

Our data set comes from a database dump of the Wikibooks Website on 15 May 2009<sup>5</sup>. It contains 2,039 books. For each book, we have the metadata of edits, such as the writers who have contributed, the number of revisions and the time at which the book is edited. We also have the full text of the different pages of each book as well as the talk pages, on which writers directly communicate with one other to discuss how the book should be written/edited. We divided the data set into two subsets, namely *featured* and *non-featured* books, based on whether the book was included in the list of featured books. Books with only one page were excluded as they were unrepresentative. To facilitate our experiments, we parsed the metadata of each book to reconstruct the timeline of the development of the book, thus obtaining information such as which author edited the book or created a new page at each point in time.

In order to quantify book coordination mechanisms, we focused on several major characteristics of the books in Wikibooks. Firstly, we identified the number of

---

<sup>5</sup> The latest data set can be downloaded from <http://download.wikimedia.org/enwikibooks/>

*predicted pages*: that is, pages to which broken links were created before the page itself was created. This allows us to identify books where the broken link technique was used to form an implicit structure for the sections of the book. Secondly, we counted the number of edits made by different authors, and sorted the authors by their contributions, to identify the most prolific authors and the more casual authors. We calculated the inter-quartile range and the Gini coefficient (Gastwirth, 1972) of authors' contributions to measure the skew of contributions within each book. We applied these measures to talk pages and content pages separately.

We also identified the age of each book at the time it was featured. Special templates and page categories on Wikibooks are used to identify featured books; detection of these tags was used to identify the point at which the book was made featured. This was compared to the timestamp of the first revision to give the age.

### **4.5.1 Measure of Quality in Wikibooks**

Wikibooks does not provide a quantitative measure for book quality, so we chose an empirical and qualitative measure: whether a book is judged as a 'featured book' or not. We base our analysis on the definition of quality used within the Wikibooks community by its contributors. While their definition is imprecise, they do offer guidelines as to what constitutes a 'good' book that deserves to be featured. These include, for example, whether the book conforms to Wikibooks policies, the clarity of the text of the book, whether the book is complete, and whether the book is properly structured.

Previous studies of Wikipedia articles (Kittur et al., 2009a), (Kittur et al., 2009b) used a similar measure of quality, based on whether it was explicitly chosen to be featured by Wikipedia administrators. In others, users (participants to a user study) were asked to read an article and rate its quality. Since books are much longer than encyclopedia entries, it is not practical for the Wikibooks administrators to do this. Instead, the communities of people who read and assess these books nominate a small number to be featured, based on a set of well-defined guidelines. Once a number of users have nominated a book to be 'featured', an administrator reviews the strength of arguments in conjunction with the content of the book. If a book is made 'featured', the Wikibooks community then treats it as a high quality book. Thus this research bases its definition of quality on a guided, qualitative assessment by a large number of readers.

Although the guidelines do not explicitly mention the elimination of redundancy as a quality metric, it is implied in some of the characteristics of a featured book, such as not containing any cleanup or maintenance tags. Although users can vote for a book, featured status is only granted to a book based on the strength and quality of the arguments for the nomination. In addition, as only a small set of books are featured every month, the guidelines can be seen as stringent, and the quality of featured books assumed to be reliable. Hence, it is reasonable to use the status of a book as an indicator of its quality in our research.

## 4.6 Method For Measuring Redundancy

Redundancy within an article or a book can only be measured accurately if the meanings of individual sentences can be understood. However, this remains a difficult problem even for modern computers, and determining whether two sentences refer to the same piece of information is also difficult by implication. In most cases, semantic similarity between two pieces of text are used to approximate redundancy (e.g. (Goldstein et al., 2000, Li et al., 2009)). In this paper, we chose to use cosine similarity between term vectors constructed by using the TF-IDF weighting scheme (Kowalski, 1997) to measure redundancy between paragraphs.

Our goal is to examine how the number of contributors to a book and differing coordination techniques result in changes in redundancy over the lifetime of a book.

In formal notation, let  $T$  be the set of terms that appear in the book  $\beta$ . We employ standard IR pre-processing methods such as stop-word removal and stemming to produce the set  $T$ .  $\beta$  is divided into a set  $P$  of paragraphs. Each paragraph  $p$  is characterized by a term vector:

$$\vec{v}_p = (w_{p,1}, w_{p,2}, \dots, w_{p,k}, \dots, w_{p,|T|}) \quad (1)$$

where  $w_{p,k}$  is the weight of term  $t_k$ , given by the standard TF-IDF weighting scheme.

The similarity between two paragraphs can then be calculated by using the cosine similarity measure, given by:

$$\text{sim}(\vec{v}_{pi}, \vec{v}_{pj}) = \frac{\vec{v}_{pi} \cdot \vec{v}_{pj}}{\|\vec{v}_{pi}\| \times \|\vec{v}_{pj}\|} \quad (2)$$

The cosine similarity only tells us how similar two paragraphs are with respect to the terms they contain, but does not tell us how much redundancy is observed in the book. In this paper, we define redundancy as the proportion of pairs of paragraphs that attain a similarity value higher than a threshold  $\alpha$

$$redundancy(\beta) = \frac{\sum_{p_i, p_j \in P, i \neq j} \delta(p_i, p_j)}{2 \times C_2^{|P|}} \quad (3)$$

where

$$\delta(p_i, p_j) = \begin{cases} 1 & \text{if } sim(\vec{v}_{p_i}, \vec{v}_{p_j}) > \alpha \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

While a high level of similarity between two paragraphs does not always imply the existence of redundant information (since cosine similarity measures how similar two paragraphs are, not how redundant two paragraphs are), in practice we confirmed by testing and manual inspection that this is a good approximation of the amount of redundancy found in a revision of a book, and we see that in most cases pairs of paragraphs achieving similarity higher than 0.5 are redundant. Hence,  $\alpha$  is empirically chosen to be 0.5 in this research. By using the above definition of redundancy, we are then able to pinpoint specific periods in the timeline of a book to examine different coordination mechanisms and contributions by writers.

## 4.7 Analysis

### 4.7.1 Contributors concentration

Wikibooks includes 60 books nominated as “featured” by the community. These books are judged to have reached a high standard of quality. We analyzed statistical information from these books to understand the level of participation in each book from its beginning to the time that the book was made featured. Different authors participations are shown, representing authors who have contributed 25% of the book, 50% of the book, 75% of the book and the total number of authors. This is done in order to show the distribution of authors’ contributions within the books.

Different patterns of contribution can be seen within the histories of different books. From Table 4.1 we can observe the difference in distribution of contributions. Firstly,



many more authors contribute to books (mean = 401) than participate in discussion (mean = 41). Secondly we can see that for an average book, 50% of edits are made by a small number of authors (mean = 13.8), with the remaining 50% by a much larger number. Fewer people contribute to discussion pages, with 50 % of edits made by a mean of 4.8 authors.

Descriptive Statistics			
	Mean	Median	Std Dev
<b># Contributors in books</b>	<b>401</b>	<b>160</b>	<b>917</b>
<b>25% main edits</b>	4.47	1	2.28
<b>50% main edits</b>	13.8	3	28.7
<b>75% main edits</b>	80.5	17	220
<b># Contributors in talk pages</b>	<b>41</b>	<b>16</b>	<b>74</b>
<b>25% talk edits</b>	1.4	1	1.29
<b>50% talk edits</b>	4.8	2	7.9
<b>75% talk edits</b>	14.2	4	27.2
<b>Age (days)</b>	869.81	881	473.74

**Table 4.1:** Descriptive statistics: mean, median and standard deviation in featured books

## 4.7.2 Changes in redundancy over time

We measured changes in redundancy over time and analyzed the resulting patterns. Figure 4.2 for example shows the redundancy for the *Guitar*, *This Quantum World*, *C# Programming* and *New Zealand History books*, which reflect the different patterns of redundancy observed in the featured books set.

The pattern of changes in redundancy over time can be roughly classified into three different classes. Of the 60 featured books, 23 books present a pattern of a sharp initial increase in redundancy (as seen in Figure 4.2(a) and 4.2(b)) followed by a gradual decrease over time. Conversely, in 15 books, redundancy is initially low, increasing sharply at different times during the history (as seen in Figure 4.2(c)). Finally, 22 books present only negligible redundancy throughout the book's history (e.g. Figure 4.2 (d), the spike around September 2007 was not caused by redundancy but the presence of large amount of similar non-English words in a section written in the native language of New Zealand).

Table 4.2 presents statistics for books of each of these types, showing the distribution of edits over the books' histories. We also computed the percentage of predicted pages relative to the total number of pages in order to determine the amount of coordination through planning of the document. The age of the books, ie. the number of days since the book project was first created until it reached 'featured status', is also shown.

Books which presented a pattern of low redundancy in the beginning but experienced dramatic changes in the middle of the development take an average of two years to completely stabilize (we consider changes to be stabilized when subsequent changes in redundancy are within the 2% range). In some cases this can take longer, as seen for example in Figure 4.2 (a), where redundancy within the *Guitar* book increases in redundancy after its beginning in 2004, stabilizing by 2005 and increasing again before re-stabilizing by 2007, becoming featured two months later. In books where redundancy increases within the beginning of the book, it typically stabilizes within the first six months, without manifesting again (with the exception of *Guitar*).

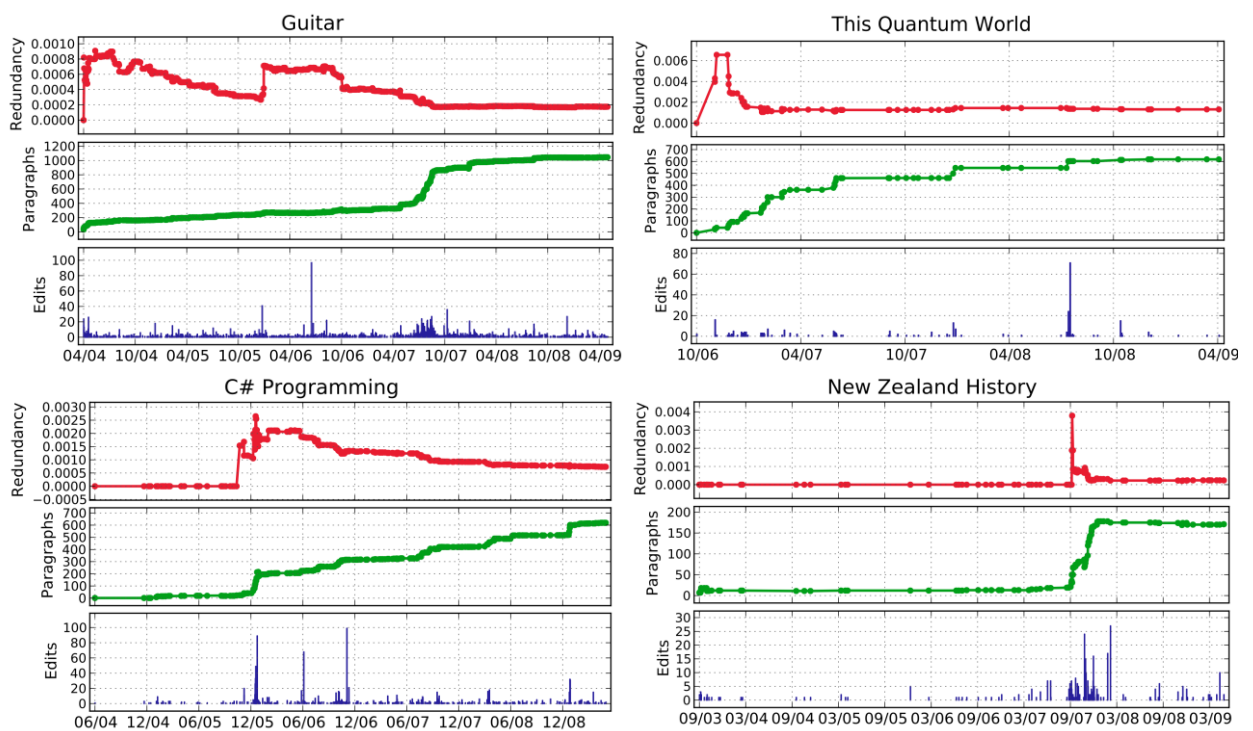
From Table 4.2 we see that books presenting little redundancy throughout their history

	Initial (23 books)			Sudden (15 books)			No Appearance (22 books)		
	Mean	Median	Std Dev	Mean	Median	Std Dev	Mean	Median	Std Dev
<b>Predicted pages (fraction of total pages)</b>	0.135	0.025	0.178	0.125	0.058	0.135	0.313	0.199	0.320
<b>#Total Editors</b>	<b>235</b>	<b>122</b>	<b>280</b>	<b>438</b>	<b>305</b>	<b>612.2</b>	<b>580</b>	<b>187</b>	<b>1370</b>
<b># 25%</b>	2.43	1	1.74	2.73	1	3,065	2.36	1	2.07
<b># 50</b>	8.04	5	8,84	19.13	4	37.54	18.63	3	34.64
<b>#75</b>	41.3	17	53.54	98.3	35	193.9	115.7	11.5	316.2
<b>Talk page Editors</b>	<b>29.1</b>	<b>16</b>	<b>30.7</b>	<b>63.0</b>	<b>40</b>	<b>85.76</b>	<b>57.5</b>	<b>9</b>	<b>125.0</b>
<b># 25%</b>	1.3	1	0.68	2.4	2	2.02	1.04	1	0.63
<b># 50</b>	4.1	2	4.06	8.66	4	13.3	3.5	2	5.4
<b>#75</b>	11.4	6	13.0	23.1	12	39.4	16.1	3	35.5
<b>Age (days)</b>	827	634	441.2	1031	1054	366	804	784	531
<b>Word count</b>	50804.3	43356	21719	80620	54315	62159	290734	69936	613427
<b>Talk page word count</b>	5110.5	2854	6002.2	12067	5551	12260	10213	2000	16928

**Table 4.2:** Descriptive statistics (mean, median and standard deviation) grouped by patterns of redundancy

have a higher mean number of predicted pages, with an average of over 30% of final pages created as broken links, thus facilitating collaborative writing and development of future sections of the book. This is significantly higher compared to 12.5% and 13.5% for other book types.

Books presenting sudden appearance in redundancy take longer to become featured – the average time for a book of this type is 1,031 days. This compares to 827 days and 804 days for the other book types. Interestingly, books which show little redundancy throughout their history take the shortest time to become featured, and are considerably larger by average word count (50,804 and 80,260 compared to 290,734 words).



**Figure 4.2:** Redundancy, number of paragraphs and number of edits over time in four example books, (a) *Guitar*, (b) *This Quantum World*(c) *C# Programming*, and (d) *New Zealand History*.

### 4.7.3 Redundancy and planning

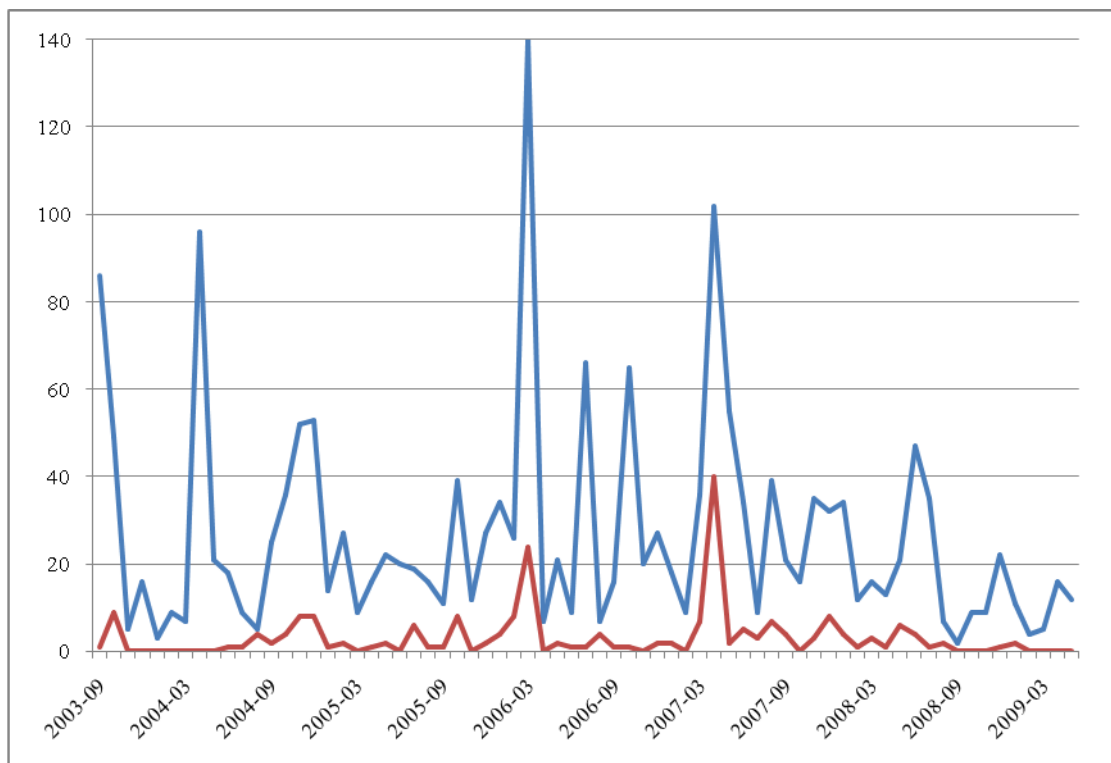
From the findings of the previous section we can conclude that the amount of redundancy observed in a book is closely related to the amount of time spent by the writers on planning.

To understand the effects of planning, we analyzed the number of predicted pages as a percentage of the total pages of a book, and the amount of discussion between authors in the book's talk pages.

The existence of predicted pages provides writers with a basis for division of labour, and writers are therefore less likely to spend time on writing the same or similar content. For example, of the books which present no redundancy, twelve have a predicted page percentage covering more than 68% of the total number of pages.

An example of this set is the book *Control Systems*, which has 65 predicted pages, accounting for a majority of the 75 pages in its latest form - containing 74,567 words. The amount of redundancy for this book remains low throughout its history, although it presents only negligible discussion (of a total of 1905 words, compared to an average of 10,213 word count for all books). The book included 93 total contributors with 10 contributors to the talk page discussion. Thus it can be seen that although there is little direct discussion between authors, redundancy does not occur within the book.

Of the books which present no redundancy, not all use predicted pages extensively. Ten of these books show a very low number of predicted pages; however, in these books, coordination is achieved through the use of talk page discussion. For example, while there were only two predicted pages for the *Latin* book, which had 87 pages in its



**Figure 4.3:** Latin Book main pages (—) and talk page (—) edit history (x-axis: display time; y-axis display: number of edits)

latest form, discussion of the book's structure occurred early in its development and continued to be used to coordinate efforts between authors (see Figure 4.3) on a variety of different topics (see Figure 4.4). Such collaboration mechanisms allowed 21 leading authors (accounting for 50% of edits) to coordinate the activities of 401 supporting authors. This is evident in the talk page where 77 authors actively discussed the status of the book.

In contrast, in books where no such kind of planning was performed (whether using broken links or talk page discussion), the redundancy curves rise sharply initially. 20 of the featured books show this pattern, including *Special Relativity* (3 out of 29 predicted pages), *Arimaa* (1 out of 53 predicted pages) and *XForms* (1 out of 154 predicted pages). This lack of coordination contributes to the appearance of redundancy.

Contents [hide]	
1	Introduction Completed
2	Some thoughts...
3	Systematic Approach
4	Latin Dictionary...
5	Formatting.
6	Speaking Latin???
7	Exercises
8	Comments from a Latin Student
9	declensions and cases
10	Imperatives
11	Answers To Exercises
12	Advanced
13	Learning unused languages
14	Long Vowels
15	Hard Going
16	Missing Stuff
17	Wikibooks:Naming policy
18	Comments from a different Latin student
19	And another student..
20	Orphans
21	Ideas for Revisions
22	Thinking of Helping
23	Pre/Suffixes
24	An alternative book
25	Wow!
26	Spoken Latin
27	Diagramming
28	Wikijunior
29	Serious lack of example sentences
30	Chapter five
31	Orphaned page

Figure 4.4: Latin talk page table of contents

In the first six months, *XForms* was developed mainly by one leading author (Gini coefficient  $G=0.936$ ), as was *Special Relativity* ( $G=0.874$ ). Initial development of *Arimaa* involved three leading and twelve supporting authors ( $G=0.687$ ). In these three examples, the increase in redundancy is not related to the number of authors who contributed to the books, but is instead linked to the approach to the writing activity taken by the authors themselves, who do not plan the different sections needed for the book.

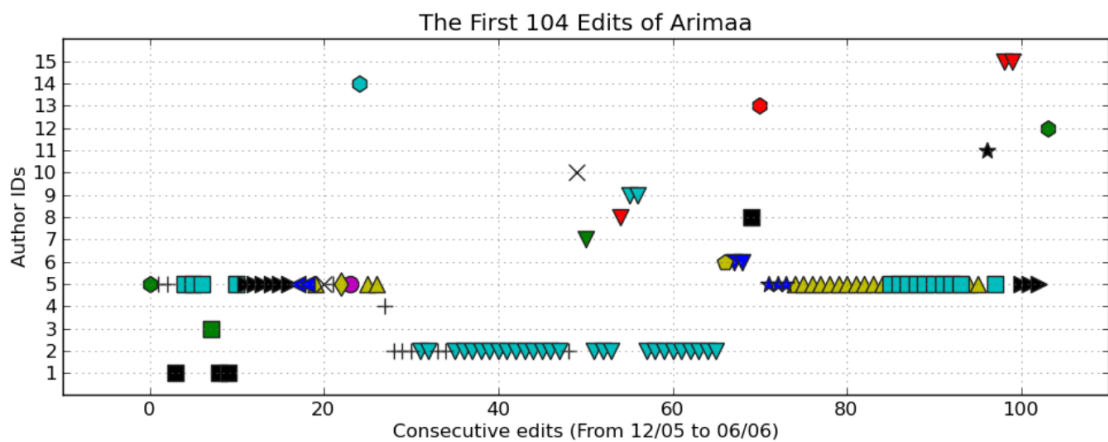


Figure 4.5: Interactions between contributors over the first six-months of the writing of *Arimaa*. Each symbol represents a different page within the book. The contributors (y-axis) used no discussion

Figure 4.5 shows how a number of authors contributed to different pages in the early stages of the writing of *Arimaa*. Although three authors led the activity in the first six months, twelve more authors contributed to the book. The three leading authors wrote separately without any interaction with one other or coordination of future activities. As a result, the other twelve authors who contributed to the book made rather chaotic edits, adding pages that contained redundant information (see also Figure 2(b) for changes in redundancy over time for this book category).

In these books, authors compensate for a lack of planning by restructuring the content of the books later in their development, ultimately leading to a decrease in redundancy. In *XForms*, for example, redundancy decreased after the book was restructured into different sections by a leading author. This same pattern of restructuring was seen in several other books. In *Special Relativity*, a leading author used the talk page to direct changes and amendments to the book, which were then carried out by other authors. The same was done in the *Arimaa* book along with 15 other featured books.

Discussion between authors is observed to be crucial in helping to reduce redundancy in a book. Taking the book *C# Programming* as an example of a book that shows a sudden spike in redundancy (Figure 4.2(c)), redundancy is seen to increase when four new active authors began to contribute to the book independently, apparently paying little attention to previous contributions by other authors. As the book did not already have a solid structure (having at this time 13 pages with 4 predicted pages), the contributions of these authors were not well organized and included no discussion. For example, two wrote about similar concepts, unaware of each other's actions. This led to an increase in redundancy in the book.

This overlap of information continued until other authors identified the problem and addressed it, by using talk pages to suggest changes to the structure. The result was a major restructuring of the book, leading to a decrease in redundancy. After this restructuring, 50 new authors began to contribute to the book, ultimately leading to an increase in the size of the book by 30%. This demonstrates that it is important that a book has been structured in a logical way such that authors have a clear idea of what each section should contain, without needing to read the large amount of text present in the book.

While redundancy within a book as a whole is undesirable, specific increases in redundancy can be beneficial, as they can lead to restructuring, resulting in a net decrease in redundancy. This is seen in 17 of the books which present sudden spikes of redundancy and 8 of the books which show initial redundancy. Another positive restructuring can be seen in the *European History* book. In this book, a sudden spike in redundancy is due to a casual contributor writing about material that was already present in other sections of the book. The section created by this author was nonetheless kept in the book and relevant text from other sections moved into it.

There are occurrences in which addition of text causes duplication of effort, as seen in *Formal Logic*, where a casual editor increased the redundancy in the book by adding to a page material already covered in other pages. Redundancy decreased sharply after text was moved by the main editor.

#### **4.7.4 Redundancy and quality**

In addition to observations discussed above, we find redundancy within a book to be indicative of the overall quality of the book. As discussed in previous sections, redundancy arises when coordination mechanisms are not properly used to coordinate the work of writers. From Table 4.2 we have seen that there is a correlation between redundancy and the time that a book will take to reach a good quality. This can be seen from the fact that books which present no redundancy throughout their history take a shorter time to become featured. From this, it is reasonable to assume that writers use a lack of redundancy as an indicator that a book contains high quality content.

For each of the 15 books where redundancy appeared suddenly, we compared the time when the book was made featured with the time when redundancy was stabilized. In 85% of the cases the book was nominated within two months from the point at which redundancy was stabilized. In other words, the occurrence of and stabilization of changes in redundancy can be considered an effective indicator of the quality of a book.

### **4.8 Discussion**

Although these books have a large number of contributors, authorship is usually differently concentrated. We have seen that within the set of featured books, on average 50% of the book is written by a mean of 14 authors with a larger number of supporting contributors for the remainder of the book.

Previous research has not examined how the number of authors affects coordination and textual quality. It might be presumed that the presence of many authors would lead to duplication of efforts and could introduce redundancy within the text. From our data set (Table 4.2) we see that number of contributors does not affect the appearance of redundancy or quality of the book. Duplication of effort only appears when communication mechanisms are not properly used, leading to chaotic collaboration (as seen in Figure 4.5). Books with a higher number of contributors but which use communication mechanisms early in the book's development become featured earlier than books which do not use coordination mechanisms and have fewer contributors.

We have seen that planning of how a book is to be structured has an impact on future coordination within the book. Books with proper planning require different levels of coordination, which ultimately reflects on the content of the book. If the book is not planned, we see a rapid initial increase in redundancy within the text, which only reduces when appropriate coordination mechanisms are used to compensate.

Different coordination mechanisms in Wikibooks hence affect the time taken for the book to reach featured status. Books with effective planning reach this status sooner than books that are not properly planned and coordinated.

However, the presence of redundancy within a book sometimes places a *positive* role with respect to the quality of the book. This is because it triggers restructuring within pages and discussion between editors. This effect can also attract new participants to the book (as seen in the *C# Programming* book). Redundancy can also provide different point of views, which are then used to create a better version of the book (as seen in *C# programming*, *Formal logic* and *US History*, for example).

Conversely, redundancy has a *negative* effect as it can increase the time taken for the book to become featured, due to the fact that more discussion is required to coordinate efforts even though fewer authors contribute to it overall (Table 4.2).

Hence this research has demonstrated that coordination mechanisms, whether in the form of explicit discussion or implicitly present in the form of broken links or section restructuring, have an impact on reducing redundant text.

In addition we have seen that coordination mechanisms have a strong association with textual quality from the beginning of the book's history. These findings are reflected in past research with Wikipedia (Kittur and Kraut, 2008), where a strong correlation was



found between article quality and coordination mechanisms. Using the measure of redundancy, we have seen that in books that do not use coordination mechanisms, an initial spike in redundancy occurs at the beginning of the book, stabilizing after coordination mechanisms are used. The earlier that coordination mechanisms are used, the sooner the book will become featured. We have shown that books that are properly coordinated from the beginning reach featured status earlier even though they contain more content on average than other books. These books also attract more authors. This phenomenon is reflected in other book types where the number of writers increases after coordination has been used to reduce redundancy within the text.

From the data we have noted that often a small group of authors set the direction of the book, which helps development of future sections. Our data suggests that it is important to start setting the direction, structure and scope of the article at the beginning of its development.

As the book matures and coordination requirements ease, tasks may be more effectively distributed to a larger group of contributors. These conclusions are consistent with observations from other research that explicit communication through coordination was most beneficial in an article's formative stages, when its structure is unconstrained (Kittur and Kraut, 2008). We have also seen that an increase in the number of authors can increase the amount of redundant text when coordination mechanisms are not properly specified.

*Hence through a large-scale analysis of a real world collaborative authoring system (Wikibooks), this research has demonstrated that when communication mechanisms are properly used, desired results are achieved sooner than when they are not used (Contribution 2).*

### **4.8.1 Limitation to the Approach**

While we have successfully studied different patterns of redundancy in collaborative writing and identified factors that affect the presence of redundancy, we are aware of some limitations of our approach. In particular, the method we use to measure redundancy is only an approximation and is not always accurate in identifying redundant pieces of text. In a few cases we observe sudden increases in redundancy

because a few paragraphs happen to describe similar concepts without actually being redundant.

Although it is impractical to manually check whether every case corresponds to genuine redundancy, this limitation poses a potential challenge to larger scale studies. For future work we therefore plan to consider more advanced techniques to identify redundancy in texts by looking at a meaningful analysis of the text within each book.

## **4.9 Conclusion**

While traditional books are typically written by a small number of authors, those written on-line, such as on the Wikibooks website, are typically a larger group effort. In some cases this group effort leads to a successful high-quality book.

In this chapter we studied the online coordination mechanisms used to develop 60 Wikibooks, designated as high-quality by the Wikibooks community. We undertook a study of textual redundancy to derive a measure of the quality of books, and analyze the different factors related to the occurrence of redundancy within the lifetime of a book and considered this against the set of communication mechanisms which were available to support the collaborative writing.

Our results have shown that changes in redundancy influence the quality of the book. When certain coordination mechanisms were not present, books developed redundant material, with a corresponding decrease in its perceived quality. This suggests a special role for co-ordinating mechanisms in collaborative environments related to enhancing the quality of the produced output.

Crucially, by analysing the history of successful books, it can be seen that changes in redundancy are correlated with the use of effective coordination mechanisms. This highlights the importance of coordination mechanisms to online collaborative activities.

These findings can be considered alongside the analysis in Chapter 3, where problems in traditional collaborative activities could be explained in terms of classic established analysis of the problems of awareness. Although the overarching objectives of these different types of collaboration are radically different in aim and approach, it remains the case that social factors and communication between contributors are key to effective collaboration.

# Chapter 5 Collaborative Tools Requirements

In Chapter 3, the concept of awareness was introduced and used to explore a theoretical model for the problems that affect the collaborative authoring process. It identified the value of associating different types of awareness as being necessary for accomplishing different tasks. It was therefore hypothesis that an effective collaborative tool would ideally provide a workspace that supports each type of awareness.

In this chapter, a set of requirements are defined for a richly featured collaborative tool designed to support social interactions between authors. The requirements cover both the communications mechanisms and awareness features that authors require. These choices are justified based on the previous analysis of awareness (Chapter 3), and the analysis of the collaborative process and common associated problems (Chapters 2 and 4).

The requirements in this chapter are deliberately kept at a high level. They define the type of information that should be provided, rather than the method by which it should be supplied to the user. Two core areas of functionality are incorporated communication mechanisms and awareness enhancing mechanisms.

These high level requirements will be used subsequently (Chapter 6) to analyse and compare existing systems, and to experiment with different techniques in order to measure the potential of the use of enhanced awareness to achieve the essential goals in collaborative authoring.

## 5.1 Conjectures and Hypothesis

From Chapter 2 (2.2.5.1) this researched has identified a set of problems that can occur within the collaborative authoring process. In Chapter 3 we mapped the occurrence of

these problems to the lack of awareness within the workspace. However some problems occur more commonly than others (especially within the collaborative authoring of documents and reports within academia and industry, since these are the sets of users who were investigated). Hence, although a tool should maintain different types of awareness as outlined in Chapter 3, the occurrences of problems are related to the different types of collaboration that take place during the collaborative authoring process.

As outlined in Chapter 2, there are different forms of collaborations (2.2.1), different forms of collaborative activities (2.2.2) and different collaboration size (2.2.3); in each of these, different levels of awareness might be more useful than others. In defining this set of requirements we suggest incorporating each type of awareness in order to create a tool that can support different forms, modes and size.

In today's available collaborative tools it is rare to find a collaborative tools that solves all the problems that might encounter in collaborative authoring. Task-specific tools are far more common. The MediaWiki wiki implementation, for example, was designed to allow multiple people to collaborate on a document with a very limited set of restrictions within the environment. Wikipedia and Wikibooks, which are based on MediaWiki, are examples of sites in which a large number of contributors work on same document together.

When beginning a new research document or Wikipedia/Wikibooks page, users might find it useful to plan the sections that need to be created, and coordinate future work through a small description of what each section should include (which is part of group awareness). However, this information might become irrelevant after the document has developed further, when authors are writing and extending sections of the document. ***Hence, it can be seen that different awareness mechanisms are more effective at different stages of the collaborative authoring process (Hypothesis 1).***

In Chapter 4 it was shown that redundancy within books on the Wikibooks website could be reduced through the use of communication and coordination mechanisms that support informal awareness. These findings are reflected in past research with Wikipedia (Kittur and Kraut, 2008), where a strong correlation was found between article quality and coordination mechanisms. Information about changes during development of the page could be beneficial for small task-focused groups of co-

authors since it is important that authors are aware of changes that were made since the document was last examined. This would help to give an overview of the ongoing issues and status of the document. However a feature such as this might be less desirable and useful in large-scale collaborations such as Wikipedia and Wikibooks.

Wikipedia and Wikibooks articles can change dramatically over the period of one day, due to the large population of voluntary contributors that works on them. Hence showing all changes that occur in the entire article since a user looked at, might be completely pointless, the users might only want to change a small portion of the document and does not care to understand the overall structure of the entire article. In fact in Wikipedia, document versions can be compared with one another in order to gain a sense of that is happening recently. Some communication systems such as annotations would be disruptive in a system like Wikipedia, since there are the high number of contributors would potentially lead to the document becoming cluttered with too many annotations.

In a research team, the peer-review process in which peers critique and review read each others' work is an important stage in the document development process. In Wikipedia, the use of discussion pages to plan future work (particularly at the beginning of development) has been proven to decrease the appearance of redundancy (as shown in Chapter 4) and can be linked to a better quality article as shown in previous research (Kittur and Kraut, 2008).

By contrast, discussion that is not attached to the document, and instead exchanged through a separate medium, has been shown in research to introduce additional problems. Examples are confusion in tracking conversation threads, as well as in identifying parts of the documents that were discussed, and whether information was up-to-date. Planning is however still a crucial aspect of collaboration in academia, and although it is done differently than in large-scale collaborations, a plan is a crucial aspect of collaboration, both in academia and in large scale online collaborations.

***Hence, planning is crucial to the success of the collaborative authoring process and improves collaboration and coordination between authors (Hypothesis 2)***

In academia, knowledge of roles and responsibilities (which are part of group awareness) is desirable, since authors need to track activity within the document and be able to identify sections that have been completed and sections that have not. It is

important to know who is working on a particular section in order to be able to reassign it or send comments directly to the author. This kind of awareness is redundant in large-scale systems such as Wikipedia in which roles and responsibilities are less well defined, and where the large number of editors makes precise planning of activities impractical.

Awareness of what has been change since the document was last viewed can be helpful in order to optimise the process and ensure that authors do not make redundant changes to the document itself. *So the addition of awareness mechanisms to the collaborative workspace can help to improve the collaborative authoring process (Hypothesis 3).*

## 5.2 Communication mechanisms

### 5.2.1 Commenting

In the context of collaborative authoring, a comment is typically a written remark expressing opinion or feedback on a piece of text. Akiko (Inaba et al., 2004) suggests representing interactions between users in utterance labels and utterance types since it is necessary to prepare an appropriate set of utterance labels, which can represent various types of collaborative learning process. These labels are needed to distinguish between each kind of interaction process. Since comments are utilized throughout the document development activity they can be used for (but are not limited to) different purposes:

- **Suggestion:** an author suggesting changes to a section of text. Examples might include suggesting the addition of missing material to ensure that the paper has a fluent argument.
- **Critique:** critiques may be made with respect to an argument made within the text of a paper. In some cases the comment might be used to propose a better solution.
- **General Comment:** an author proposing a general comment about either the structure or the content of the section. Examples might include formatting and spelling mistakes.

Comments can take different forms, such as a post it note, an electronic comment attached to the document or an email response. Comments are useful for different

purposes at different stages in development as described in Section 2.2.5. In the brainstorming and planning phase, where authors discuss their contributions and the plan of action for the document, comments can place the role of future “flags” to appoint changes and improvements to the document. During development, peer review and formatting stages, comments place a crucial role in identifying parts of the document in need of restructuring, whether related to formatting or content manipulation.

Comments can present problems depending on how they are used. If the comment is too large it can be disruptive to the development process as it can present different opinions which should be described individually. If too many comments are present, they can clutter the document and impair reading or writing of it.

If comments are not attached to the document (for example if sent by email) they can be misinterpreted or orphaned before they are addressed. Hence points of view of authors might be lost, or time might be wasted searching for the section to which a comment refers. Comments sometimes can bring up conflicting points of view that must be resolved by the authors, hence a commenting system should support in-depth discussion. Without support for proper discussion, multiple comments can be used for discussion, leading to further clutter.

In light of these problems we can identify that a commenting system should:

1. Include the ability to cope with large numbers of comments;
2. Include the ability to structure discussion without further cluttering the document;
3. Support comments that are attached directly to the document, rather than being exchanged through a secondary communications channel such as email.

### **5.2.2 Real time discussion**

Real time discussion is discussion where participants can exchange information instantaneously, as opposed to asynchronous discussion channels such as email. Examples of real time discussion are face-to-face communication and instant messaging. Real time discussion is needed at different stages within the development of a document. It is of particular importance in the *planning* stage in which authors brainstorm possible content. It is used in the *development* stage by authors who may need clarification on points of view as well as general queries, although not all authors

may participate. It can be used in the *review* stage in which authors often need to discuss each others' work in order to provide feedback and improvements.

Real time discussion is useful in collaborative authoring as it can provide immediate answers to authors' questions and resolve problems quickly. If real time discussion is not present, authors may resort to asynchronous communication such as emails or faxes. This can cause a time lag between when a question is posed and when it is answered. In a traditional face-to-face environment, real time discussion typically involves all parties involved in the activity meeting in the same place for discussion.

Real time discussion can present some disadvantages as it can be difficult to track how conversations evolve and how conclusions were reached, unless someone specifically keeps track of the points of view of and how they evolve. Real time discussion in an online environment can be limited compared to face-to-face communication as certain clues are not present. These include facial expressions and tone of voice, which can indicate of peoples' level of attention, as well as agreement or disagreement.

In light of these uses we can identify that an electronic real time discussion system should provide:

1. The ability to record a conversation thread for posterity;
2. The ability to allow multiple authors to discuss together;
3. The ability to simulate face-to-face interactions:
  - a. Keeping track of the focus of the authors on the discussion;
  - b. Providing authors with either sounds or icons to represent agreement or disagreement.

### **5.2.3 Asynchronous discussion**

Asynchronous discussion is discussion where participants can exchange information via a delayed information channel, such as email, fax, forum or a blog. Asynchronous discussion can be used when users are working at different times during the day or if they are based in different time zones. Comments are a form of asynchronous discussion. Asynchronous discussion can be used to support different types of awareness by allowing users to create topic threads that can be addressed by the users



who are not online at time. Asynchronous discussion can be used to compensate for differing time schedules of the authors as well as different time zones.

Asynchronous discussion can be used during the development of the document content in order to address issues that need to be discussed. Asynchronous discussion is usually logged within the system. This allows backtracking over information if required. If asynchronous discussion was not available it might be difficult to coordinate efforts if all the users are not available at the same time. In a traditional environment, emails, forums, blogs and wikis may all be used to coordinate efforts within the development environment.

Although asynchronous discussion is commonly used, it presents some problems due to the time delay required for the information exchange. For example, it might take a long time for all users in the activity to reply to a topic. At the same time a problem might have already been resolved by another user who does not take other users in the activity into consideration due to time constraints. Asynchronous discussion is a limited form of communication mechanism compared to face-to-face discussion. Asynchronous discussion might also require additional effort to read if many replies exist. In the case of emails, blogs and wikis, confusion might be created due to different topics discussed in a discussion thread, this can be caused by difference in discussion structured and overload of information.

Support for in-depth discussion of a manuscript within collaborative system could present various advantages. For example:

1. The ability for authors to focus their discussion on precise areas of the manuscript. Participants could discuss issues in a threaded system better suited to the discussion process.

## **5.3 Awareness Characteristics**

### **5.3.1 Personal Awareness**

As described in Section 3.2.1, personal awareness is the information that a user maintains about themselves, their roles and responsibilities, and the motivations behind their actions. Personal awareness in real life scenarios can be demonstrated with the use of a personal diary or calendar in which a user keeps tracks of meetings and tasks to be

completed during the day. In an online scenario such as a collaborative activity, this awareness can for example be maintained as a list of tasks that each user will need to accomplish. Personal awareness is important in order to understand the personal contributions and to make sure that all tasks are completed on time. If not present, users may confuse the work at hand and incur in a duplication of effort (as shown in Section 3.3).

Hence personal awareness should provide a way for users:

1. To know their roles with respect to each task;
2. To know their responsibilities;
3. To know their motivations.

### **5.3.2 Social Awareness**

As shown in Section 3.2.1, social awareness refers to information that users maintain about others in social or conversational contexts: for example, whether a person is paying attention, and their emotional state or level of interest. This kind of awareness is important as it can help to show the points of view of each author with relation to agreement and disagreement. Hence to maintain social awareness, a tool should provide:

1. The ability to simulate partly face-to-face interactions;
2. The ability to track whether authors are focused on the discussion;
3. The use of sounds or icons to represent agreement or disagreement.

### **5.3.3 Informal Awareness**

As shown in section 3.2.1, informal awareness refers to the information that users maintain of who is present and what they are doing. In a real life scenario, users can gain this awareness simply by observing the environment. In an online environment this awareness is completely lost if not maintained within the system. This kind of awareness is important as it helps users to coordinate efforts in relation to the presence of other users online. It is important when users need questions answered and they know who is actually working on the activity. In light of these facts informal awareness should support:

1. The ability for users to know which other users are working on the activity;
2. The ability for users to know what other users are doing.

### **5.3.4 Group Awareness**

Group awareness is the information that users maintain of other users' roles, activities, movements and status in the process. As shown in Section 3.2.1, it includes peoples' roles and responsibilities, their positions on issues, current status, and group processes.

In a real life scenario, group awareness is maintained, for example, as a plan of actions in which users are assigned roles and tasks depending on these roles. As the collaborative activity progresses, the plan might evolve to include users' position on issues as well as the current status of activities – such as if they are completed or delayed. As different versions of the plan are kept, an outline of the group's processes can be also derived.

Group awareness is important in order to coordinate efforts within the collaborative activity. This helps to ensure that no duplication of efforts occurs as well as helping to keep an overview of the collaborative process. It is important to estimate the time of completion for each task and to assign tasks to appropriate users in order to maximise group efforts. In the light of these factors group awareness should provide:

1. The ability to track or maintain users' statuses;
2. The ability to provide an immediate overview of each user's responsibilities;
3. Information on users' statuses and responsibilities through the entire collaborative activity.

### **5.3.5 Workspace Awareness**

Workspace awareness concerns a user's presence in the workspace and knowledge of what users are currently doing: up-to-the-minute knowledge about other people's interactions within the shared workspace. In a face-to-face scenario this kind of awareness is the knowledge of the up-to-date information about each user's position. If, for example one user is absent from their desk, this kind of information is easily gathered from a glance at the user's desk. At the same time, users can simply understand each others' status in the process by simply asking one another information.

In light of these information we can summarise that workspace awareness should provide the ability:

1. To maintain the up-to-date information about users' interactions with the collaborative activity.

## **5.4 Discussion and Conclusion**

This chapter has considered the high level requirements of a richly functioned collaborative authoring environment based on recognised and established needs for effective communication and awareness during the authoring process. Such a system must provide multiple channels of communication between authors, and support each of the types of awareness previously identified in Chapter 3. Eight different functions have been identified across the two areas.

This requirements analysis has deliberately avoided defining specific user interface techniques to achieve these aims. Numerous groupware systems exist, and the decision to keep the analysis at a high level is utilised to provide a framework for analysis of a set of selected collaborative authoring systems featured in Chapter 6. The requirements will also be the basis of subsequent experiments to measure the relative effectiveness of different supporting features in the collaborative environment.

## Chapter 6 Available Collaborative Tools

This chapter analyses existing collaborative tools with respect to the functionality as defined in the requirements in Chapter 5. This functionality incorporates support for group collaboration, as well as appropriate communication and awareness mechanisms. 21 contemporary collaborative authoring systems are analysed to give insight into the state of the art, specifically the extent to which existing systems address the needs for awareness and inter-author communication, as well as examining the user interface techniques that have been used to provide the required types of awareness.

The systems which are examined extend across a range of paradigms ranging from word processing and text editing tools, groupware systems, wikis and content managers. Particular attention is paid to wiki systems. We explore how the wiki concept can be used for collaborative authoring of professional academic papers. Wikis have proven valuable tools for collaboration and content generation over the Web. When examined against the awareness requirements, it is shown that wikis have numerous features that assist in their use for collaboration.

By analysing the state of the art, it is shown that no existing system fully supports each of the identified types of awareness, although different types of system excel more than others at providing specific types of awareness. Certain types of awareness do not appear to have been fully reproduced in any existing system.

### 6.1 Introduction

Different co-authoring tools take different approaches to supporting the collaborative authoring process. The biggest distinction between collaborative editors is between synchronous editors (where editing occurs in real time) and asynchronous editors such as Microsoft Word and OpenOffice, text editors, wikis and management systems. These are shown in Table 6.1.

		Name	Short Description
Synchronous	Microsoft Word/ OpenOffice - like	CoWord	CoWord enables multiple users to edit the same document over the Internet with Microsoft Word.
		Google Docs and Spreadsheet	Google Docs is a web-based office suite, including a word processor. Users can collaborate in real time to write documents.
	Text Editor	MoonEdit	MoonEdit is a synchronous text editor that allows multiple participants to edit a document simultaneously. The interface shows a separate cursor within the document for each participant, while a side pane shows the other users who are currently editing. Changes occur in real time and are propagated to other participants automatically.
		SubEthaEdit	SubEthaEdit is a text editor that allows multiple people to edit a document simultaneously. Participants can make changes at any time and are not blocked by the edits of other users.
		Gobby	Gobby is a text editor with a client-server interface that allows multiple documents to be worked upon simultaneously. The edits of different users are distinguished by the use of different colours. Users can perform discussion in real-time through an IRC-like chat interface.
	Asynchronous	Wiki based Systems	TikiWiki CMS Groupware
TellTable			TellTable allows OpenOffice document files to be edited from a web interface, with changes tracked and audited. However, it does not allow multiple people to edit simultaneously.
MediaWiki			MediaWiki is the wiki system used to develop Wikipedia. It includes full history tracking and the ability to “watch” pages (users watching a page are notified of changes to the page)
PmWiki			PmWiki is a wiki system that supports PDF output through the use of Latex; however, custom Latex style files are not supported
Instiki			Instiki is a wiki system that supports PDF output through the use of Latex; however, custom Latex style files are not supported.
UniWakka			UniWakka supports PDF output through the use of a Latex backend. It also contains a bibliography management system and citation support.
GroupSwim			GroupSwim is a wiki-based system that allows users to collaborate within a social environment that allows users can see each others’ works and interests. They can share projects or post notices about their work.
BusinessWiki			BusinessWiki is a wiki-based system that allows users

			within a company to share corporate knowledge. Unlike other wiki implementations, BusinessWiki allows users to save the pages created within the wiki environment into folders.
		<b>ProjectForum</b>	<b>ProjectForum</b> is designed around the wiki concept but presents features not found in other wiki implementations, such as a blog and a locking system that locks the document while editing (allowing each user to know who is editing the document). It also uses a “post” feature to allow comments and discussion on any page in the wiki.
		<b>Confluence</b>	<b>Confluence</b> is a wiki-like content management system. It integrates collaborative authoring techniques with office management tools. From the confluence dashboard (front page) users can see what other users are doing and access the wiki to discuss and exchange knowledge.
		<b>Bloki</b>	<b>Bloki</b> is designed around the wiki concept, which it extends with a blog and discussion forum. It also incorporates a system that locks the document while editing (allowing each user to know who is editing the document).
		<b>e-touch SamePage</b>	<b>SamePage</b> is a business-oriented wiki system that combines a wiki and blog to support teamwork.
<b>Management Systems</b>		<b>BSCW</b>	<b>BSCW</b> enables collaboration through shared objects such as document URLs, notes and calendars.
		<b>Workshare Professional</b>	<b>Workshare Professional</b> is a commercial tool used to share documents and to verify, secure and audit changes. It can compare two documents and identify changes.
		<b>Groove</b>	<b>Office Groove</b> is a collaboration software program that helps teams to work together remotely, or offline. It is a shared workspace in which users can upload or download documents. It also includes an integrated chat system.

**Table 6.1:** Categorisation of tool used

**CoWord** (Sun et al., 2006) provides an editing environment based upon Microsoft Word that can be used concurrently. **TellTable** (Adler et al., 2004) provides a collaborative environment based on OpenOffice.org, although it does not support synchronous editing. Similarly, tools such as **Gobby**<sup>6</sup>, **MoonEdit**<sup>7</sup> and **SubEthaEdit**<sup>8</sup> also present a synchronous shared text editor.

<sup>6</sup> 0x539 dev group, *Gobby a collaborative text editor*, 2005, <http://gobby.0x539.de/trac/> .

<sup>7</sup> Dobrowolski T., *MoonEdit*, 2006. <http://moonedit.com/indexen.htm>

<sup>8</sup> Ott, M., Pittenauer, M., Wagner, D., *SubEthaEdit 2.6*, 2007, <http://www.codingmonkeys.de/subethaedit/>

**Google Docs & Spreadsheets**<sup>9</sup> is semi-synchronous (as it refreshes the screen every two minutes). Tools such as **BSCW (Basic Support for Cooperative Work)**<sup>10</sup>, **Groove**<sup>11</sup> and **Workshare Professional**<sup>12</sup> use an alternative approach of providing only a shared space in which documents may be uploaded. Wikis are most commonly used as knowledge management systems; the typical wiki consists of a collection of hyperlinked pages containing information on different subjects (Wikipedia being the most well-known example). As such, most wiki systems are optimized for this purpose.

Various popular wiki systems exist, some of which contain some of the features needed for professional publications. Tools which are based on the wiki concept include **PmWiki**<sup>13</sup> and **Instiki**<sup>14</sup>, **GroupSwim**<sup>15</sup>, **BusinessWiki**<sup>16</sup>, **Confluence**<sup>17</sup>, **e-touch SamePage**<sup>18</sup>, **TikiWikiCMS Groupware**<sup>19</sup> and **UniWakka**<sup>20</sup>, although it is the **MediaWiki**<sup>21</sup> implementation that has gained the most prominence due to its use in Wikipedia. **Bloki**<sup>22</sup> is another wiki-based implementation which, like **ProjectForum**<sup>23</sup>, has been developed as a management system.

---

<sup>9</sup> Google Inc., *Google Docs & Spreadsheets*, Mountain View, CA, 2006.

<http://docs.google.com/>

<sup>10</sup> OrbiTeam, *BSCW 4.3.3*, in Fraunhofer FIT and OrbiTeam Software GmbH & Co. KG, ed., Boon, Germany, 2006. <http://public.bscw.de/>

<sup>11</sup> Microsoft Corporation, *Microsoft Office Groove*, 2007, <http://office.microsoft.com/groove/default.aspx>.

<sup>12</sup> Caton, M., *Workshare Professional 4.0*, eWeek.com, 2007, <http://www.workshare.com/products/wsprofessional/>

<sup>13</sup> PmWiki, <http://www.pmwiki.org/>

<sup>14</sup> Instiki, <http://www.instiki.org/>

<sup>15</sup> TN20, Incorporated, *GroupSwim*, <http://groupswim.com/>

<sup>16</sup> Divante Sp. z o.o. *BusinessWiki*, Wroclaw Poland <http://businesswiki.eu/>.

<sup>17</sup> Atlassian Pty Ltd., *Confluence*, 2008, <http://www.atlassian.com/software/confluence/>

<sup>18</sup> eTouch Systems Corp, *e-touch SamePage*, <http://www.etch.net/home/>

<sup>19</sup> TikiWiki CMS Groupware, <http://info.tikiwiki.org/>

<sup>20</sup> Rossato A., *UniWakka Wiki Farm*, 2007, <http://uniwakka.sourceforge.net/HomePage>.

<sup>21</sup> MediaWiki, <http://www.mediawiki.org/>

<sup>22</sup> Zapatec, I., *Bloki*, Berkeley, CA, USA, 2005. <http://www.bloki.com>

<sup>23</sup> CourseForum Technologies, *CourseForum and ProjectForum*, Guelph, Ontario, Canada, 2007, <http://www.projectforum.com/pf/>



## 6.2 Communication mechanisms

### 6.2.1 Commenting

In tools such as **Google Docs and Spreadsheets**, commenting is supported through inline comments. **CoWord** and **TellTable** are based on Microsoft Word and OpenOffice.org respectively. The commenting features provided by each are similar due to this design. As in Microsoft Word, commenting within the document is achieved with the use of annotations. Each user is distinguished by a colour and the name of the user who added the annotation.

**BSCW** provides the ability to attach comments to uploaded files, but does not provide any support for annotation of text within the documents. **Workshare** provides the ability to annotate changes made to documents, so that the purpose behind edits can be more easily understood when reviewing changes.

Wiki systems present different features. It is common for wiki systems to also allow a description to be attached to edits (as seen in **MediaWiki**, **PmWiki** and **Instiki**). **UniWakka** and **ProjectForum** allow comments to be attached to the end of a wiki page.

### 6.2.2 Asynchronous discussion

Wiki systems in general provide strong support for asynchronous discussion; it is an inherent feature of any wiki, as a “talk” page can be created to act as a discussion page where users can write comments. Systems such as **MediaWiki** provide explicit support for this concept, by attaching a dedicated talk page to every article. **ProjectForum**, in addition to its support for comments, provides blogging features and support for a discussion forum. Similar features are found in **Bloki** and **Confluence** and **SamePage**.

Non-wiki systems show greater variety in the forms of asynchronous discussion provided. **Google Docs**, **Gobby**, **SubEthaEdit** and **MoonEdit**, as synchronous editors, do not provide support for asynchronous discussion. Tools such as **BSCW (Basic Support for Cooperative Work)**, **Groove** and **Workshare Professional** use an alternative approach of providing only a shared space in which documents may be uploaded and where discussion is supported, either on individual uploaded documents or as a separate discussion.

### 6.2.3 Real time discussion

Various tools provide support for real time discussion. **Google Docs and Spreadsheets**, **Gobby** and **SubEthaEdit** provide built-in support for chat-style discussion between authors while editing. Discussion is implemented in different forms in the different tools. SubEthaEdit integrates with the iChat tool on Mac OS X, while **Google Docs** uses Google's own web-based chat system (interoperable with other systems via XMPP). While these tools implement simple person-to-person chat, Gobby provides an IRC-style shared discussion channel where all authors can participate in discussions.

**Groove** shows a list of users currently online within the shared workspace and allows online authors to communicate in real time through voice chat in addition to text-based chat.

In general, wiki systems do not tend to include support for real time discussion. **MediaWiki**'s support for user talk pages does allow users to be notified when a message has been written on their talk page, supporting a crude form of real time discussion; however, this is much less fluid than is provided by other systems.

## 6.3 Awareness Characteristics

### 6.3.1 Personal Awareness

Personal awareness is provided through a variety of different techniques in different collaborative tools. In **SubEthaEdit**, **MoonEdit** and **Gobby** for example, text is coloured according to the user who wrote the text, so the user can easily identify the text that he or she has written. This is specific personal awareness related to editing. **CoWord** uses a similar approach, but extends colouring to the Word "track changes" feature, so that comments and changes made to a document are highlighted according to the user who made the change. Similarly Workshare Professional presents personal awareness by allowing users to review recent changes made to the document.

Almost all wiki systems include the ability to view the past history of a page and to graphically compare the text between two versions (by using the differencing feature in which removed words are highlighted in red and added words in green). This provides a form of implicit historical personal awareness to the user. **MediaWiki** in particular includes the ability for the user to list all past changes that he or she has made.

Other tools use different techniques. **Confluence**, **ProjectForum** and **Google Docs** both allow the user to create a “to-do” list of work to be performed. **BSCW** also includes a to-do list, which it refers to as tasks; the age of each item is shown, allowing the user to identify tasks that have not been addressed. In wiki-based systems, users can create a page to act as a to-do list if they wish.

**Bloki** and **ProjectForum** include the ability to lock editing of pages. This allows users to know what they are editing. **Confluence** and **SamePage** include a user profile page and blog that show recent activity by the user. In **SamePage** this is tracked automatically by the system.

### 6.3.2 Informal Awareness

**CoWord**, **SubEthaEdit**, **MoonEdit** and **Gobby**, as synchronous editors, provide built-in visual indications of who is present within the shared workspace – a list of online users provides an immediate visual indicator. A similar feature is present in **Groove**, although it is not a synchronous system. **Google Docs** shows an indicator to the user if another user is present and editing the document. Other systems in general do not allow users to identify who is present within the system.

Synchronous editors such as **MoonEdit**, **Gobby** and **SubEthaEdit** allow the user to identify what other users are writing through the highlighting feature that shows sections of text colour-coded by the author that has written the text. This feature is not present in **Google Docs**, which instead uses a semi-synchronous approach and does not highlight text. Other, non-synchronous systems by their nature do not allow users to see activity as it occurs in real time.

Wikis in general are effective at providing historical informal awareness, as the common presence of page history page allows the changes made by other users to be listed and visualised in a straightforward manner. Systems such as **Bloki**, **GroupSwim**, **SamePage**, **ProjectForum** and **Confluence** provide an activity log that allows “recent changes” to be tracked, with events divided by type (eg. comments, text changes, attachments, etc). Users can also infer recent activity by using a page’s history page and viewing changes made since a certain date (eg. the last time they logged in).

Synchronous systems such as **CoWord**, **SubEthaEdit** and **Gobby** provide historical informal awareness through highlighting of added text coloured by author; however,

subsequent deletion of added text is not tracked, so a full history of changes is not visible. **CoWord** is a synchronous system that does not track the full history, allowing deleted text to be viewed. Similarly, TellTable tracks participants' current and past activity through the OpenOffice.org "track changes" feature, and allows participants to track who is editing the document at any time, as the system is based on a shared server that allows only one user to edit the document at a time.

**Workshare Professional** includes a feature that allows an author to merge changes made by another author. This provides informal awareness, as the author is intrinsically made aware of the changes that have been made. However, it does not present other informal awareness features (such as who is currently present within the system).

### 6.3.3 Group Awareness

Group awareness concerns awareness of roles and responsibilities. It was found that in general, support for group awareness is poor within most systems. In wikis, a page can be created to act as a plan for future work, to which authors can refer and amend as progress is made. This kind of awareness is completely user-centric as users will need to remember to update the list in order to supply other users with this information. In systems such as **ProjectForum**, **Google Docs** and **Confluence**, the to-do list and blogging features can be used for tracking progress and to plan future events. The to-do list can be used for the same purpose within **BSCW** (although this is not a wiki).

The presence of features to support informal awareness also assists group awareness. For example, the presence of an activity log (as seen in **Confluence** and **SamePage**) can allow users to infer information about roles and responsibilities.

### 6.3.4 Workspace Awareness

Workspace awareness concerns up-to-the-minute information about activities within the system. While synchronous systems (eg. **MoonEdit**, **Gobby**, **SubEthaEdit**, **CoWord**) provide features that support informal awareness and allow the user to determine what activities other users are engaged in, this information is not directly presented; the user must specifically examine the activities of other users and infer what they are attempting to achieve.

For non-synchronous systems the situation is typically worse, as no up-to-the-minute information is provided that relates to information about progress and interactions made with the workspace. In synchronous tools, such as **CoWord**, **SubEthaEdit** and **Gobby**, up-to-the-minute information about users' progress is maintained by the fact that each user is assigned a colour, allowing other users to visualise what they have done. Similarly up-to-the-minute information about document status is kept by providing the synchronous movements and the document status (this feature is also supported in **MoonEdit**, **Google Docs and Spreadsheets** and **TellTable**).

## 6.4 Discussion

From this examination it is quickly apparent that requirements needed for communication and collaboration mechanisms are partially satisfied by available tools. Table 6.2 summarizes the features present in each tool. An inability to discuss issues within the document in real time is likely the most fundamental inadequacy of most systems. From Table 6.2 it can be seen that only four tools (**Google Docs and Spreadsheets**, **Gobby**, **SubEthaEdit** and **Groove**) support this; however, these tools do not support asynchronous discussion, which is helpful to coordinate efforts when work and time schedules are different and users cannot be online simultaneously.

Only four tools (**CoWord**, **Google Docs and Spreadsheet**, and **TellTable**) support attaching annotations to specific paragraphs of the document, a feature that is important in order to provide insight into the changes that need to be made. If this feature is not present, the paragraph in question needs to be cited, and authors can encounter problems as a result (as seen in 2.2.5.1). Other systems provide some support for asynchronous discussion in the form of comments at the end of a page or discussion in a separate page (eg. 'talk pages' in wiki systems).

Some co-authoring tools designed specifically for collaborative authoring do not fully support all features needed within an online environment. Tools such as **TellTable**, **BSCW**, **CoWord**, **Groove**, **MoonEdit** and **Google Docs and Spreadsheet** lack adequate support for maintaining the collaboration mechanisms used to coordinate online work, eg. the ability to identify recent changes made to the document. Even though **CoWord**, **TellTable** and **Groove** have a global 'track changes' features that tracks changes made by individual authors, the list of changes can clutter the document, making it difficult to

differentiate recent changes from older ones. Changes accepted by one author can be missed by other authors accessing the document later.

The ability to know who is online and what each person is working on is essential for coordination of efforts, but is also another collaboration mechanism missing from most tools. While (semi-) synchronous tools (such as **Google Docs and Spreadsheets**, **Gobby**, **MoonEdit**, **SubEthaEdit**, **CoWord** and **Groove**) allow the user to see who is present within the workspace, determining what they are working on requires additional effort. Other tools do not support this feature at all, the result being that two authors could be working simultaneously without realising that they could discuss issues that they encounter. Another inadequacy within collaborative tools is the inability to assign roles and responsibilities to each user in order to plan and coordinate efforts, even though this is a common (even essential) step in the development of documents (as outlined in Section 3.3).

While some tools provide the ability to create ‘*to-do*’ lists, the information on these lists must be updated manually for users to be kept up-to-date with developments within the document. With the exception of synchronous tools such as **Google Docs and Spreadsheets**, **Gobby**, **MoonEdit**, **SubEthaEdit** and **CoWord**, most tools therefore lack the ability to maintain up-to-date information of activity within the workspace, which can hinder development (as seen in Section 2.2.5.1, Section 3.3 and Section 4.7.3).

From Table 6.1 it can be seen that the **UniWakka** tool does not present any advanced feature to help the collaborative authoring process. However, it does include other features such as the ability to create PDF files and maintain a bibliography, which are required to write research papers in academia.

Tools such as **CoWord** and **Groove** provide informal awareness through the track changes feature that allows recent changes to be viewed. **TellTable** has both a recent changes feature and also includes features for comparing versions and auditing changes to allow users to visualise changes. This is also supported further by a version control system that shows who has made each change to the document. However, users may only use the window of the document one at the time, which is not always convenient if multiple users need to edit the document.

Although wikis do not provide all the awareness required for proper coordination and communication throughout the document development process, they have proven helpful

for collaborative authoring; **MediaWiki**, for example, has been used to great success in the development of Wikipedia, which is edited by thousands of users daily and includes over three million pages. It has been shown in previous research (Kittur and Kraut, 2008) that the quality of the articles is correlated with the coordination mechanisms used to write them.

**Workshare Professional** allows users to review each others' changes, which can be helpful in industry where users need to write manuals and specification documents, and where the most important goal is to ensure that all requirements have been satisfied and contributions from all users have been integrated.

This analysis has covered **SubEthaEdit** and **TikiWiki CMS Groupware**, although these tools have proven more effective as tools for collaborative programming, rather than for collaborative authoring.

Table 6.2 shows how each of the tools scores against the requirements defined in Chapter 5. The maximum possible score is 19, although it can be seen that none of the compared tools satisfy even half of this.

Although most tools support asynchronous discussion, support for annotations and synchronous discussion is rare.

Only synchronous tools provide the information necessary to answer the questions, "who is around? what they are doing?". However, these tools lack the ability to distinguish between historical and recent changes.

Few tools include features to support group awareness. Some tools, such as Google Docs, **BSCW**, **ProjectForum** and **Confluence** allow ideas and tasks to be planned in advance with a "to do" list feature, while in others, especially wikis, this has to be performed in a separate page attached to the document. In both cases, to track progress, users must insert this information, which can be prone to human forgetfulness and errors.

In synchronous tools, up-to-the-minute information is difficult to track, since it can be difficult to determine the position of users, their interactions within the workspace and the status of the document unless the interface has been designed for this purpose. Only **TellTable** supplies each of the required types of up-to-date information; an individual user can use the server while other users watch the user's interactions with and movements within the workspace. This can however can be time consuming and wasteful as other participants cannot use the workspace at the same time.

**CoWord** allows users to view the positions of themselves and other members with a small page view and a highlighted box that shows the user’s position within the document. Users can then view what other users are typing in real time; different highlight colours are used to differentiate between users (a similar feature appears in **Gobby**, **SubEthaEdit** and **MoonEdit**).

	Communication Mechanism			Collaboration mechanism															Awareness Score
	Commenting	Asynchronous Discussion	Real-time Discussion	Personal Awareness			Informal Awareness				Social Awareness	Group Awareness				Workplace Awareness			
				A	B	C	A	B	C	D		A	B	C	D	A	B	C	19
Co Word																			7
TellTable																			6
Gobby																			7
MoonEdit																			4
SubEthaEdit																			7
Bloki																			7
ProjectForum																			9
BSCW																			5
Google Docs and Spreadsheet																			7
Workshare Professional																			5
Groove																			5
MediaWiki																			8
PmWiki																			7
Instiki																			7
UniWakka																			7
PBWiki																			6
GroupSim																			6
BusinessWiki																			7
Confluence																			9



e-touch SamePage	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	8	
TikiWiki CMS Groupware	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	6

**Table 6.2:** Analysis of 21 different collaborative tools, compared against the set of features required for collaborative editing of academic publications. ■ indicates not supported, while ■ indicates supported. The legend below describes the feature represented in each column.

#### Personal Awareness

A	B	C
---	---	---

A – “What am I doing?”

B – “What have I done?”

C – “What do I have to do next?” (own role/responsibility)

#### Informal Awareness

A	B	C	D
---	---	---	---

A – “Who is present?”

B – “What are they are doing?”

C – “What have other users done since I last looked at the document?”

D – “What have other users done in the past (historical changes)?”

#### Group Awareness

A	B	C	D
---	---	---	---

A – Tracking progress

B – Planning of future events

C – Knowledge of group members’ roles

D – Knowledge of group members’ responsibilities?

#### Workspace Awareness

A	B	C
---	---	---

A – Up-to-the-minute information about document status

B – Up-to-the-minute information about users’ interactions with the workspace

C – Up-to-the-minute information about users’ progress

### 6.4.1 Google Wave

Google recently introduced the Wave tool, which combines elements found in various different communications technologies (eg. e-mail, wikis, instant messaging) into a single system. Wave is not intended merely as a collaborative authoring tool, but can potentially be used for this purpose.

Users of the system edit documents called *waves*. Multiple users can subscribe to a wave and edit the text simultaneously. Discussion is supported by comment boxes called *blips*, which can be inserted into the document and used for threaded conversation between users. The system maintains an edit history of changes that are made to each wave, so that it is possible to view older versions of the document and view changes.

Because of these features, Wave provides several types of awareness. Informal awareness is supported, as it is possible to see the changes that other users make to the document as they edit it (a cursor with the other user's name shows where the user is editing), and it is also possible to view historical changes to the document through the edit history function. The system also supports workspace awareness, as its synchronous nature ensures that the current status of the document is instantly available, and the interactions of other users with the workspace can be viewed as they occur.

Support is also provided for both synchronous and asynchronous discussion through the same mechanism. New blips appear in real time, facilitating its use for synchronous discussion in a manner similar to instant messaging. If a user is not online to see the new blip as it is created, Wave highlights the new blip to the user on the next login.

However, the system does not provide support for personal or group awareness, as there is no direct support given for planning or tracking roles and responsibilities. The open-ended nature of the system means that this information could potentially be stored in a blip within the document.

At the time of writing, Wave is currently in a beta status with only limited access available. However, because of the features that it presents, it may potentially be of greater interest in the future.

## 6.5 Conclusion

In this chapter we have highlighted the features available in current tools with respect to communication and collaboration requirements identified in Chapter 5. We have shown that although tools have been designed to support such a process, many lack the ability to maintain information related to users' movements and activities within the workspace itself.

Several of the analysed tools are wikis. The wiki design is of particular interest as some of the wiki systems satisfy multiple requirements for supporting awareness and communications. These tools have been designed specifically for use in impersonal, online-only collaboration, which makes support for these awareness mechanisms desirable.

While most tools provide features that support one or more types of communication and awareness, each is deficient in particular areas. For example, CoWord supports personal awareness, but has no support for group awareness, while ProjectForum supports group

awareness but has no support for personal awareness. Of particular note is that no tool was found to support social awareness, suggesting that this is particularly difficult to reproduce in an electronic environment.

Although these tools have successfully been used to create documents, implementation of the enhancements identified by the requirements analysis could additionally provide an adaptive environment in which authors are aware of the actions of other users and ways in which these actions affect the progress of the document. The results of this survey demonstrate that much opportunity exists for improving upon the state of the art to move toward achieving this aim.

## **Chapter 7 CAWS – Co-Authoring Wiki-Based System**

In Chapter 5 a set of possible requirements were defined for an effective collaborative tool, and in Chapter 6 it was demonstrated through a survey of existing collaborative authoring tools that none of the widely used and widely available tools that were examined fully satisfied this set of requirements. There is therefore opportunity for improvement upon the state of the art to achieve this goal, and for experimentation with user interface techniques to support awareness mechanisms.

This chapter describes the design of the CAWS co-authoring system, developed as part of this research as a testbed to experiment with and evaluate the efficacy of user interface designs to support awareness. The features of CAWS are described, including the rationale behind their design and how they are intended to assist the user.

Analysis of existing systems revealed that wiki systems commonly include a number of features that support awareness. Because of this, and because of the ubiquity of web based collaborative authoring systems, CAWS is based on a wiki-like design. However, it also incorporates a number of features that are not commonly found in popular wiki systems, specifically designed to better support the awareness requirements.

### **7.1 Introduction**

In order to investigate experimental techniques to improve the collaborative authoring process, the CAWS prototype system was implemented. CAWS is designed to satisfy the requirements defined in Chapter 5.

The fundamental design of CAWS is based on a wiki, and incorporates many of the basic features found in a wiki. However, unlike a wiki that consists of a number of

separate hyperlinked pages, CAWS consists of documents containing sections that can be rearranged. The system also includes various features designed to support authors' awareness.

## 7.2 Example Scenario

A group of people, residing in different parts of the world, wish to work together to write a journal article on a common topic. Part of the group resides in Europe, while the others reside in the USA. Of the members, John and Mary reside in England, Sally and Chris in Scotland, Zoe in Belgium, George in California, and Stuart in Florida. They face not only dislocation in their geographical location but also in time zones. As a result, there are only few hours a day in which their work schedules overlap. They decide to use CAWS as a shared environment in which they can collaborate on the document.

John logs into CAWS and selects “create new document”, assigning the title “journal article”. Next, he invites the other users by entering their email addresses. The other users join the document after receiving the emails.

Later, John logs in to CAWS and selects the document, and is brought to the ‘front page’ of the document. This page displays the *status* of each user and the *activity log*. As the creator of the document, John has control over the rights granted to other group members. From the *user status* box he realizes that he is not the only group member currently online. Mary, Sally, Zoe and George are also online. Zoe and Sally are using the ‘Editor’ section of the system. Zoe is restructuring the document while Sally is editing the introduction to the document. Mary is in the ‘blog’ part of the system and is replying to a blog post called “part 1” previously created. George is in the ‘style’ part of the system and is customizing the style of the document. The user status box also shows that Chris and Stuart were online three and six hours before John logged in.

Since CAWS does not provide an instant messaging system, John logs into his instant messaging system, where he sees that Mary, Sally, Zoe and George are all online. John can hence discuss topics related to the work with the other members if he wishes.

From the *activity log*, John sees that Chris wrote a blog post called “part 1” and that Stuart replied to it and wrote an initial introduction to the document.

John starts to plan the activity by using the “planner” feature to assign roles and responsibilities. From the user status box shown in the editor, Zoe realises that John is using the planner. John assigns group members to sections. The other group members can insert their estimates for the time that it will take to finish their sections. Estimates can only be set by the user assigned the task. Development of the document can now begin.

Sally logs into CAWS to begin work on the section assigned to her. She realises that Zoe and John are also online and working on their individual sections. Sally’s section is closely related to what John is writing so before she continues, she examines John’s section to see what he is writing (she can view his changes in real time). Sally then goes to work on her own. While writing, she notices the other users finish and log out. After finishing her own work, she saves her changes and updates her remaining hours in the planner, before logging out.

Several hours later, George logs in. It is immediately apparent to him that nobody else is online from looking at the *user status* box. The *activity log* highlights the changes made to the document by the other users. He goes to the editor to continue to working on his section of the document. George takes one-hour lunch break, and after returning, realizes that Stuart has logged in. He finishes his changes and saves them. Before logging out, he returns to the document ‘front page’ and the *activity log* highlights that Stuart has replied to a comment that he previously added to Sally’s section. After reading the comment, he updates his hours remaining and logs out.

CAWS’ awareness mechanisms inform the group members of what happening in the document, both while they are logged in, and also what has happened while they were logged out. The planning system allows progress within the document to be tracked and notifications made to warn of upcoming deadlines.

After several weeks, John examines the *document summary* and realizes that in one section of the document, users have conflicting views. He also realises that one section is behind schedule, as the planner highlights the section in red. John creates a discussion in order to understand the different group members’ viewpoints on the controversy. John then asks the group members responsible for the delayed sections the reasons for the delay. Sally, the editor and second author, answers that she has

encountered a problem and asks to extend the deadline. As the administrator, John changes the schedules. The document is finished in time for the deadline.

## 7.3 Communication Mechanisms

### 7.3.1 The role of annotations

Providing support for annotation during development is crucial to the collaborative authoring process, since missing this feature can hinder the development of the document by causing communication degradation as shown in Section 2.2.5. The annotation system implemented in CAWS is therefore designed to satisfy the requirements outlined in Section 5.2.1. Authors need to be able to comment on the document, but this should not clutter or disrupt the process. In CAWS, annotations have been designed to serve different purposes throughout the document development process.

#### 7.3.1.1 Supplementing participants' points of view

Authors may possess different opinions within the manuscript, and may possibly disagree with one another. Disagreements such as these are hard to differentiate within a text editor or collaborative authoring system. The authors must express their point of view with respect to the arguments made within the manuscript, according to their personal knowledge (which can include comments, suggestions, critiques and agreement/disagreement). The lack of such feature can cause communication degradation in the form of inadequate support for discussion on a specific topic, a problem identified in Section 2.2.5.

The ability for the group members to define annotation types allows annotations to be structured into well-defined categories. This is useful to the authors who can use these categories to filter annotations to specific types in order to better understand reviews.

In order to understand different authors' opinions, it is important to facilitate commenting with the ability to differentiate types of annotation. In an academic environment, annotations can be categorized into various different types. For example:

- **Grammar:** grammatical corrections to the paper may be necessary. This is not relevant to the arguments made, but important for final publications.

- **Formatting:** formatting or style changes to the manuscript itself may be needed.
- **Critiques:** critiques may be made with respect to an argument made within the paper
- **Suggestions:** suggestions can help to ensure that the paper has a fluent argument.
- **Comments:** comments may identify particular issues that have not been addressed.

Different types of annotation may hold different meanings within the peer review process. For example, a simple formatting error does not influence the editor's decision to accept or reject the paper in the same way that a critique or comment could. These types of annotation are therefore more relevant to the editor when examining a review, or to authors when they are trying to understand why their paper was accepted or rejected. However, the type of an annotation is hard to extrapolate from a simple text-based review stored on the conference submission site.

In academia, research papers must make an argument, a statement that should be the central point for the manuscript. However, it is common that assumptions must be made within a paper, and on occasion authors may disagree with these assumptions.

### 7.3.1.2 Flagging appointed changes

Authors might require the other members of the group to make changes to the manuscript in order for it to be ready for acceptance.

This can help authors to flag other members' appointed changes in order to avoid the problems of '*misinterpreted comments referring to sections*' identified in Section 2.2.5, and to meet the requirements outlined in Section 5.2.1. Authors must address these changes and need to be able to flag the changes made in order that other group members may examine them. This helps group members to ensure that the changes made conform to their specifications.

From examining these issues, it is apparent that many of the problems stem from the fact that traditional collaborative authoring systems are not linked to the review process of the document itself.

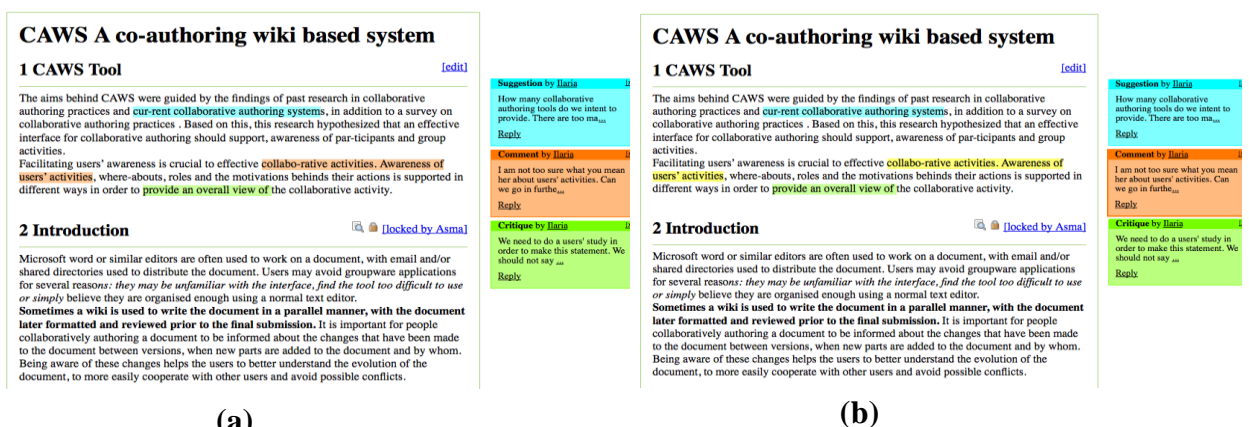


### 7.3.1.3 Annotation filtering

One previously highlighted problem is that of clutter introduced by the presence of too many annotations within a document (as outlined in Section 2.2.5.1). Providing a concise environment can allow authors to link annotations to individual paragraphs of the manuscript, allowing comments to be made in a more precise and directed fashion. Comment may be only partly shown; when the length of a comment exceeds the limit for the annotation type, an ellipsis (‘...’) is shown to indicate that the comment is only partly displayed. Less text is required to explain the context in which a comment applies. This design is intended to fulfil the requirements in Section 5.2.1. This helps authors when submitting a response, as it is possible to respond to individual annotations without quoting the specific paragraph or the comment made.

Within CAWS, parts of the manuscript can be directly annotated. Annotations are added by selecting the text to be annotated on-screen. This gives authors the ability to precisely identify the paragraph or section to which they are referring.

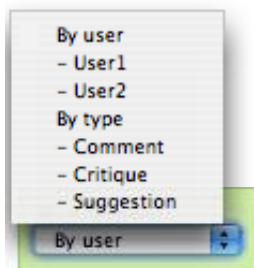
The ability to link annotations to the document also assists authors. An author may better identify the part of the manuscript to which an annotation refers (Figure 7.1 (a)). A visual aid within the tool helps accomplish this by flashing the annotated text when an annotation is selected (Figure 7.1(b)). Similarly, when the annotation is clicked the text to which it refers flashes; this is designed to assist the user if a large number of annotations are present. The annotation system also allows authors to respond to other members’ comments (if the process requires multiple interactions), as annotations are linked to a threaded discussion system in order to have detailed discussion of the annotation.



**Figure 7.1:** Annotation system within the CAWS System. (a) The annotation colour reflects the highlighted text in the manuscript (filtered by author). (b) The annotation blinks yellow when clicked and the border of the annotation attached to it thickens.

### 7.3.1.4 Visualising authors' opinions

In order to understand authors' opinions, it is important to facilitate commenting with the ability to differentiate types of annotations, as specified in the requirements in Section 5.2.1. In an academic environment, annotations can be categorized into various different types. The ability to insert and categorize reviewers' opinions is beneficial in several different ways, as this makes tracking a topic and the development of ideas easier as the document develops (addressing the '*communications degradation*' problem outlined in Section 2.2.5, by allowing users to visualise the topic of conversation in the annotation attached to the text). An annotation type (such as *suggestion* or *critique*) can be created to represent agreement or disagreement with the arguments made within the manuscript. The annotation text can be used by the authors to explain their opinion in greater detail.



The filter box allows users to filter annotations by user and by type.

The annotation types can be personalized to the scope of the manuscript.

**Figure 7.2:** Annotation filter box

CAWS provides the ability for the editor to configure the annotation types needed within the document. Once these annotation types are set, a type can be selected by the user when inserting an annotation. A drop-down list in the editor page allows annotations to be filtered by type or the user who inserted it (Figure 7.2).

This is helpful to the editor as the annotations made by a particular reviewer may be examined. Annotations made by multiple users may also be displayed together to compare the opinions of different reviewers (Figure 7.3).

CAWS annotations increase in colour intensity in areas of the text where multiple annotations are present (Figure 7.3(b)). When conflicting types are present on the same text, the highlight appears grey (Figure 7.3(a)). These features allow sections of the manuscript containing higher number of annotation to be identified.

Grey annotations help to identify areas of conflict between reviewers. For example, if two authors have commented on the same text, the overlapping section will be coloured grey. Similarly, if filtered by annotation type, and two authors have attached comments of different types to the same text, the text will also become grey.

The figure shows a screenshot of the CAWS Demonstration interface. On the left, there is a navigation menu with options: Document, Editor, Style, Planner, Blog, and Discussion. Below the menu is a filter box with the text: "The filter box allows users to filter annotations by user and by type". The filter box has a dropdown menu with options: By user, - User1, - User3, By type, - Critique, - Suggestion, and - Suggestion.

The main content area is titled "CAWS Demonstration" and contains sections: Abstract, 1 Introduction, 2 Related Research, and References. The text in the Introduction section is annotated with red and blue highlights. A filter box on the left shows the filter is set to "By user" and "User1".

On the right, there is a discussion panel with the following entries:

- Suggestion by Ilaria** (red header): This is a suggestion. Reply
- Critique by Ilaria** (red header): This is a critique. Reply
- Suggestion by Ilaria** (red header): This is a suggestion. Reply
- Suggestion by Rikki** (blue header): This is a suggestion. | comment | Reply
- Critique by Rikki** (blue header): This is a critique. Reply

Annotations (a) and (b) are highlighted with boxes and labels:

- (a) Overlapping comments from different users shown in grey to highlight differences.
- (b) Overlapping comments from the same user show an increase in shade.

**Figure 7.3:** Different types of annotation viewed by user. The two colours represent the two authors that are reviewing the paper. When the same text is annotated by two users, the overlapping text becomes grey (a). Overlapping annotations made by the same author are shown with an increase of colour (b).

## 7.3.2 Asynchronous Discussion

### 7.3.2.1 Facilitating annotation of sections of the manuscript

CAWS' annotations are linked to a full threaded discussion system (Figure 7.4), in order to facilitate discussion within the manuscript. Discussion can be used by authors throughout the activity. This feature addressed problems related to communication degradation as outlined in Section 2.2.5.

The discussion system is beneficial to the author, as it facilitates discussion of the annotations made by other members. An author can better judge other authors' points of view on each part of the document, and understand which parts of the document are under active discussion.

Authors may discuss their opinions with other members. In the case of a multi-stage process, in which two authors are responsible for the same section, authors can discuss annotations with the other respective authors. This gives the opportunity for both to respond in a concise and structured manner. It assists communication, as all authors can respond to annotations directly. This is in contrast to other collaborative authoring system where a review must continually reference the section of the manuscript to which it applies.

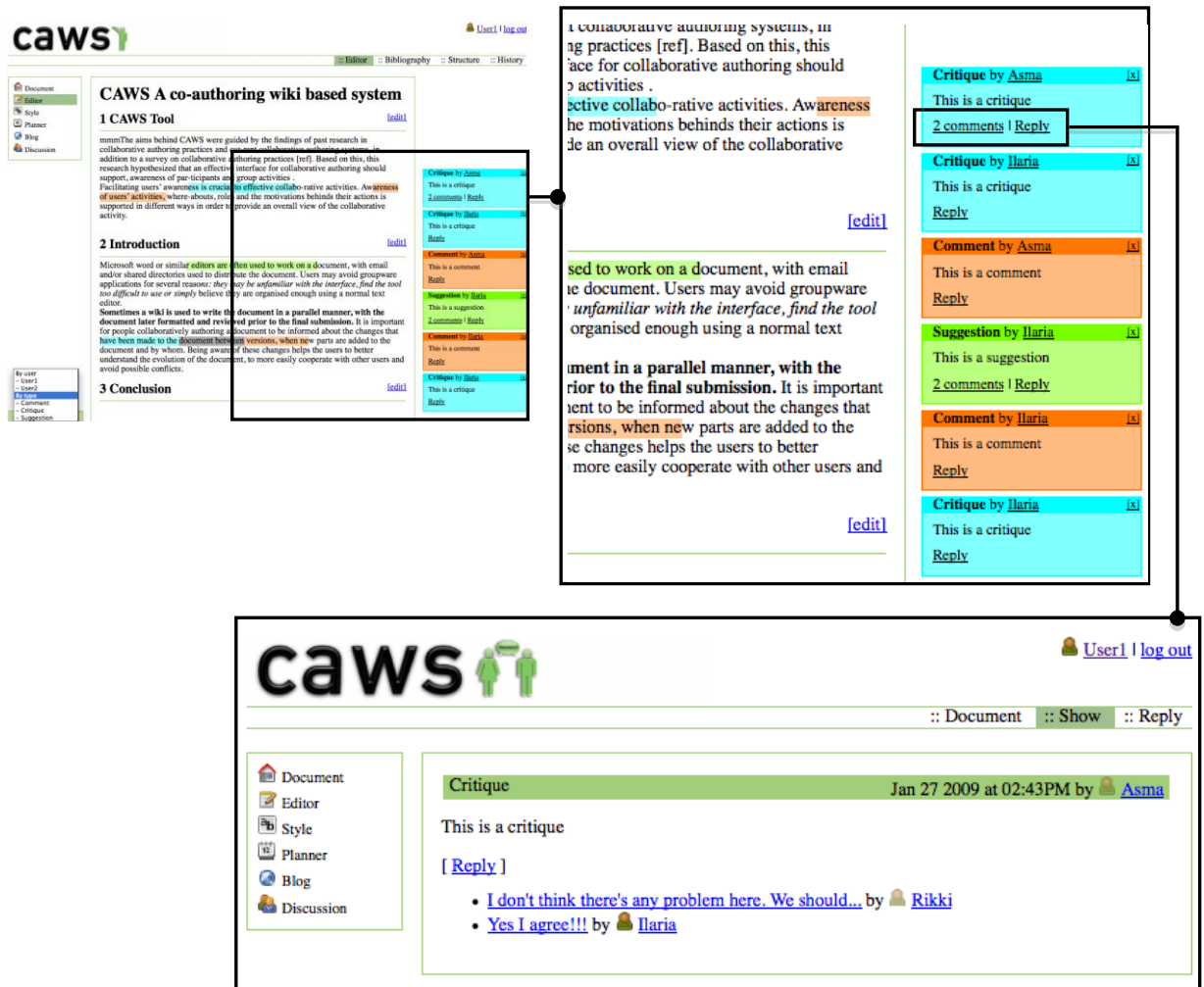
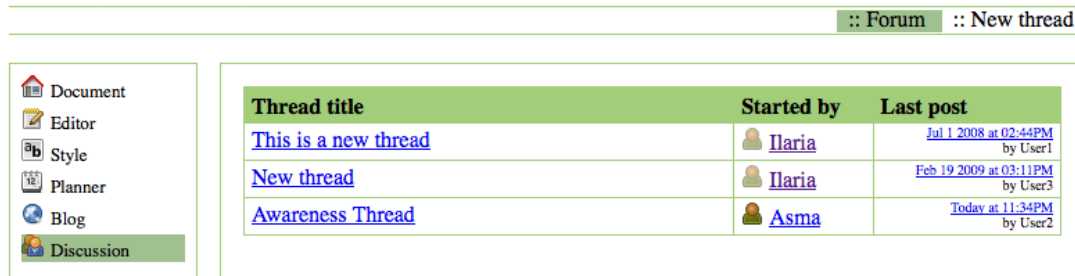


Figure 7.4: The CAWS annotation system linked to threaded discussion.

### 7.3.2.2 Discussion forum

CAWS includes a discussion forum for each document (Figure 7.5), in which authors can raise points of discussion by creating a thread of arguments. This type of discussion does not provide the same type of fast response as real-time discussion or face-to-face

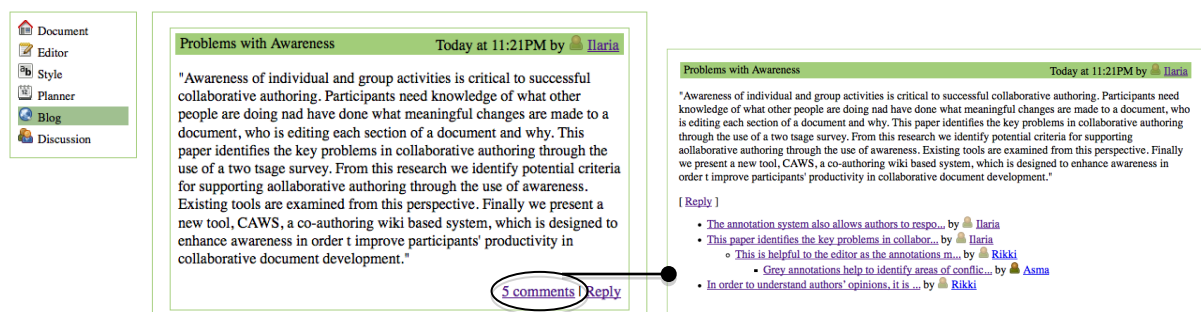
communication, as it relies on authors replying to threads in their own time. This feature is intended to satisfy the requirements for asynchronous discussion outlined in Section 5.2.3.



**Figure 7.5:** Forum feature in CAWS. Note that when a user is not online the icon next to it is greyed out, providing awareness of who is online.

### 7.3.2.3 Facilitating news updates with asynchronous discussion – Using a Blog

CAWS presents a blog feature in which users can post updates on the status of the document (Figure 7.6). Blogs have been used for various purposes such as reporting on personal activities, and updates for project development, (Penrod, 2007). The latest blog post is featured on the document front page, as part of the summary of users' recent activities.



**Figure 7.6:** Blog feature in CAWS. The latest blog post can be seen from the document front page

The blog includes a threaded discussion system, where users can respond to blog posts by posting comments. An indicator at the bottom of each blog post shows the number of comments that have been attached. This feature is intended to satisfy the requirements for asynchronous discussion in Section 5.2.3.

## 7.4 Awareness Features

### 7.4.1 Informational Awareness

#### 7.4.1.1 History of users' actions

One of the awareness mechanisms in CAWS is the activity log that maintains a history of users' actions. Each time a user performs an action within the system, the event is added to an "activity log", which is displayed on the 'front page' of the document (Figure 7.7). The group member's screen name is displayed along with the event. The events shown in the activity log include notifications of changes to the document itself, replies to comments, and changes to the document style. Events are displayed in chronological order, but related events are grouped together - for example, if two users both make changes to the same section, the activity log might display "George and John edited *Introduction*".

The purpose of this feature is to provide information on the order in which events occurred, interactions between users, and also the interactions between users and the system in order to provide informal awareness and hence satisfy the requirements outlined in Section 5.3.3 for group members to know what other members are doing.

The screenshot displays the CAWS Document front page for "CAWS Demonstration". The page includes a navigation menu on the left with options like Document, Editor, Style, Planner, Blog, and Discussion. The main content area is divided into several sections: a Welcome message with links for Editor, Revision History, PDF Download, RTF Download, and Latex Download; a Blog section titled "Showing the blog" by Ilaria; a Users section listing online and offline users (Ilaria, Asma, Rikki, Monica, Sam); and an Activity log section showing recent events such as "Asma replied to 'Showing the blog'", "Ilaria added a new blog post", and "Rikki and Ilaria modified 'Introduction'". An Owner Toolbox at the bottom right indicates that the user is the owner and can break section locks. Annotations with arrows point to the "Activity Log" and "User Status" sections, explaining their functions.

**Activity Log:** shows recent changes and other activity within the document. Events are highlighted if they occurred since the last time the user viewed the activity log.

**User Status:** Used to monitor users' statuses and position in the tool.

Figure 7.7: CAWS Document front page

Events are highlighted if they occurred since the last time the user viewed the activity log. This provides an effective means of identifying recent changes to the document, minimizing duplication of effort and providing an overview of what is going on in the document itself. Users joining the collaborative activity while it is in progress can view previous interactions between and the actions taken by other group members and gain a better understanding of the history of the document.

A history of users' actions is provided for several reasons. Firstly, it provides notification of activity during the development of the document (satisfying informal awareness by providing information on who is doing what). It provides a means of monitoring all users' actions since the start of the collaborative activity (satisfying past informal awareness by providing information on who has done what), which is a concern when operating in an academic environment (Liccardi et al., 2008).

Once users click the section, they are forwarded to the editor to read the changes that have been made. Users can click on the history tab and compare previous versions to see the sections that have been changed, as the differencing feature allows the added and removed words to be viewed. If a user is not happy with the changes made, they can roll back to the previous version of the document.

## **7.4.2 Group Awareness**

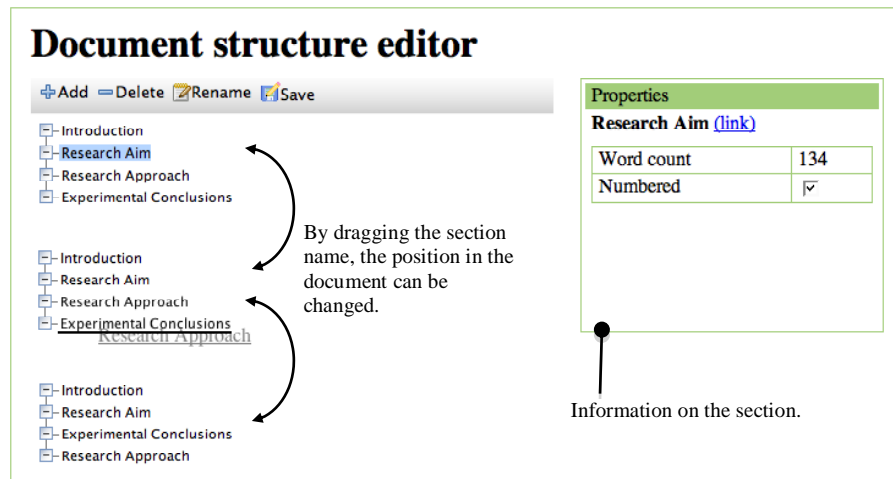
### **7.4.2.1 Division of Activities**

Many users that responded to the survey described the “*planning*” stage of the document as being crucial to its success. This includes explicitly dividing up the document into sections, assigning roles to users and assigning users to write specific sections of the document, and estimating time of completion. To address these issues, CAWS includes several mechanisms for managing roles and responsibilities within the tool (to satisfy the group awareness requirements related to the ability to track or maintain users' statuses, to provide an immediate overview of each user's responsibilities, and to have information on users' statuses and responsibilities through the entire collaborative activity as outlined in Section 5.3.4).

Traditional wikis are used for knowledge management and consist of multiple articles on different subjects. With an academic paper, the aim is to create a single document, rather than multiple documents. However, academic papers typically consist of

multiple *sections* that discuss different points related to the article. CAWS is based around this concept, with multiple sections that can be edited independently.

CAWS features a *document structure editor* (Figure 7.8) that allows the structure of the document to be viewed and manipulated in a tree form. This has two advantages. Firstly it allows the overall structure of the document to be defined at the beginning of the authoring process, providing a structured scaffolding in which authors can



**Figure 7.8:** CAWS document structure editor. The document can be restructured by moving the section title with the mouse

begin their work. Secondly, it allows the document to be restructured and sections rearranged while maintaining version history of the individual sections of the document.

Individual sections are locked for editing while a user is making changes; the system indicates which user is currently editing the section. In a traditional wiki, locking is undesirable as it prevents other users from making changes; however, when writing academic articles, it is common for a “divide and conquer” approach to be used in which authors are assigned to write specific sections. As a result, use of locking is more acceptable.

Planning of document activities is integrated into CAWS: users can be assigned roles in relation to sections created within the *document structure editor* (for example, “author” or “editor”) (Figure 7.9(a)). Users can create the amount and the type of roles that they wish to use within the activity. Once assigned a role, users can estimate the time needed for the completion of the section (Figure 7.9 (c)) The time cannot be changed by the



### Planner: Assignments

Section	Author	2nd Author	Editor
Abstract	Asma	Ilaria	Rikki
1 Introduction	Ilaria	Asma	None
2 Related Research	Asma	Ilaria	Monica
References	None	None	None

Update assignments

Users can be assigned roles in relation to sections (for example, “author” or “editor”). Users are selected from the list of users involved in the document.

(a)

### Deadlines

Section	Hours remaining	Deadline
Abstract	21	2010-02-07
1 Introduction	11	2010-02-14
2 Related Research	7	2010-02-25
References	0	2010-03-09
<b>Total</b>	<b>39</b>	<b>2010-03-31</b>

Update deadlines

Users can insert deadlines for each of the sections as well as inserting a deadline for the final version of the document. This also shows the amount of hours (inserted by the users) needed to complete the section

(b)

### Planner

Enter time estimates for the number of hours remaining for each of your assigned sections below.

Section	Author	2nd Author	Editor
Abstract	7 / 8 hrs	6 / 6 hrs	8 / 8 hrs
1 Introduction	10 / 10 hrs	1 / 12 hrs	Unassigned
2 Related Research	7 / 9 hrs	0 hrs	
References	Unassigned	Unassigned	Unassigned
<b>Total</b>	<b>24 / 27 hrs</b>	<b>7 / 18 hrs</b>	<b>8 / 8 hrs</b>

Update estimates

Once users are assigned to a section, they can estimate how long (in hours) each section will need to be completed. Once this information is entered, the amount of time becomes fixed and the user can decrease the amount of time.

(c)

### Planner : Progress graph

Section	Progress	Remaining	Deadline
Abstract	<div style="width: 5%;"></div>	5%	1 hrs / 4 days
1 Introduction	<div style="width: 50%;"></div>	50%	11 hrs / 11 days
2 Related Research	<div style="width: 23%;"></div>	23%	2 hrs / 22 days
References	<div style="width: 0%;"></div>	0%	34 days
<b>Total</b>	<div style="width: 27%;"></div>	<b>27%</b>	<b>14 hrs / 56 days</b>

CAWS tracks progress within the document and integrates it with the information that the users’ supply in order to show a summary of document status.

(d)

**Figure 7.9:** CAWS planning feature: (a) Assignment of roles in relation to sections. (b) Deadline and target word count assignment. (c) Estimation by users of the remaining time needed for the completion of the section (d) Summary of information entered and document status.

user in charge, only the administrator (ie. the user that created the document) can change them, or set a deadline and target word count (Figure 7.9(b)).

The planning part of CAWS captures this information in a graphical way (Figure 7.9(d)). This is used to summarise information about the document status. A section that has a deadline within the next week is displayed in red, while a section with a deadline within two weeks is displayed in yellow (Figure 7.9 (d)). A bar chart shows the percentage of work done with respect to the number of hours remaining (manually entered by the users), as well as information about deadlines and word count. This provides an overview of the document status (Figure 7.9(d)).

### **7.4.3 Workspace Awareness**

#### **7.4.3.1 Users' whereabouts**

Another awareness mechanism tracks users' whereabouts in the tool. As a user navigates through the system, their movements and actions are logged. This is used on the 'front page' of the document (Figure 7.7) to show the current status of all users (to satisfy the requirements for providing up-to-date information on users' movements in the collaborative workspace, as part of workspace awareness outlined in Section 5.3.5). Each group member's screen name is shown, along with their time of log in, status (set either manually by the user or automatically by the system) and up-to-date information about their location within the system. This information has the potential to provide clear and up-to-date information about other users' movements and the activities they are currently engaged in.

A smaller status box continues to show the information in a reduced form as the user navigates the various pages related to the document. This cut-down version only shows status information and activities of users who are online (Figure 7.7).

#### **7.4.3.2 Spy Mode Feature**

As the design of the CAWS system is based on that of a wiki, it inherits the feature of a wiki that a user making an edit must save his changes for other users to be able to see them. This affects workspace awareness as those other users do not have up to the minute knowledge of the work as it is being done. It is not possible to observe work as it progresses.

**CAWS A co-authoring wiki based system**

### 1 CAWS Tool

The aims behind CAWS were guided by the findings of past research in collaborative authoring practices and current collaborative authoring systems, in addition to a survey on collaborative authoring practices. Based on this, this research hypothesized that an effective interface for collaborative authoring should support awareness of participants and group activities. Facilitating users' awareness is crucial to effective collaborative activities. Awareness of users' activities, whereabouts, roles and the motivations behind their actions is supported in different ways in order to provide an overall view of the collaborative activity.

### 2 Introduction

You are watching this section being edited. Click the icon above again to stop.

Microsoft word or similar editors are often used to work on a document, with email and/or shared directories used to distribute the document. Users may avoid groupware applications for several reasons: they may be unfamiliar with the interface, find the tool too difficult to use or simply believe they are organised enough using a normal text editor. Sometimes a wiki is used to write the document in a parallel manner, with the document later formatted and reviewed prior to the final submission. It is important for people collaboratively authoring a document to be informed about the changes that have been made to the document between versions, when new parts are added to the document and by whom. Being aware of these changes helps the users to better understand the evolution of the document, to more easily cooperate with other users and avoid possible conflicts.

**CAWS Demonstration** [locked by Asma]

### Abstract

This paper describes ongoing research into collaborative authoring activities, specifically investigating how collaborative authoring tools can be engineered to improve interactions between authors. Consideration of these interactions is essential when designing an effective collaborative authoring system.

### 1 Introduction

The popularity of collaborative authoring has increased greatly in the last decade. The importance of collaboration is emphasized both in academia and in industry. Collaboration allows users with different experience and skill sets to combine their abilities toward a common goal. While remaining aware of the workspaces of other people is taken for granted in the everyday world, maintaining this awareness has proven to be difficult in distributed systems where behaviour can be hard to extrapolate and where interaction technologies are often poor. It is therefore important that authors are given effective tools to compensate for these differences and overcome dispersal of team members.

### 2 Related Research

[locked by Asma]

### References

(Autogenerated section)

Editing: Asma Editing Related Research

Show comments: By user

Critique by Asma: This is a critique, 2 comments | Reply

Critique by Ilaria: This is a critique, Reply

Comment by Asma: This is a comment, Reply

Suggestion by Ilaria: This is a suggestion, 4 comments | Reply

Suggestion by Ilaria: This is a suggestion, Reply

Suggestion by Ilaria: This is a suggestion, Reply

Suggestion by Rikki: This is a suggestion, 1 comment | Reply

Critique by Rikki: This is a critique, Reply

**Figure 7.10:** CAWS document editor, highlighting the spy mode feature.

In order to promote the up-to-the minute information needed for workspace awareness, a feature was included to allow users to observe other users as they are writing. If a user views the document while a section is being edited by another user, the section is displayed as locked, and an icon is displayed which, if clicked, allows the text being written by the editing user to be observed in real time (Figure 7.10).

This feature uses the fact that the CAWS system automatically saves draft versions of the text that a user is editing, so that work is not lost in the event of a web browser or computer crash, or a power cut. When a draft version is automatically saved, a copy is

transferred to other users who are viewing the section with the spy mode feature. Because of this, there can be a short delay between a user writing text and other users seeing the text.

The spy mode feature must be deliberately enabled; the system is not a synchronous editor and users viewing the document must explicitly choose to view the draft version of a section rather than the last version that was saved, hence once enable the section. This is for two reasons. Firstly, it encourages users to properly save the changes that they make so that the version history is properly retained. Secondly, users can read and edit the document without the view of the document being disrupted when text is added or removed by other users.

## 7.5 Conclusion

The CAWS system has been designed specifically to develop solutions to the problems of co-authoring identified in Section 2.1.3 and includes a set possible awareness mechanisms based on the model proposed by Gutwin and Greenberg (Gutwin and Greenberg, 2002) that are intrinsic to our everyday life. The requirements for a collaborative tool defined in Chapter 5 have been used as a guide for the features that the CAWS system incorporates. Some of these features are found in the existing systems that were analysed in Chapter 6, while others are new designs derived from analysis of models of collaborative working that are intended to provide a set of the types of awareness needed by authors during the document development process.

Awareness mechanisms have been thoroughly researched in the CSCW and HCI community and are not restricted to the ones proposed in this work. Hence CAWS could be considered a testbed for new ideas and could be improved in future work to incorporate different awareness mechanisms that could be automatically created by the system (for example, content based collaborative filtering (Ghali and Cristea, 2008)).

The rationale and methodology behind the design of these features has been described. However, as was identified in Chapter 3, it is not sufficient for a tool merely to provide information that authors can use to enable awareness; this information must be presented in an effective format to the user. The CAWS system must therefore undergo testing in order to understand and judge the effects that its features have on users. It

was purposefully intended to use feedback from the initial testing (Chapter 8) to iteratively refine the design of these features. The resultant modified system would then be evaluated in order to address the research question of whether collaborative authoring can be improved by increasing users' awareness.

## Chapter 8 Experiment 1: Attitude based - Usability Study

This chapter investigates the effectiveness of the features of the CAWS system (Chapter 7). It was essential that the system be tested with real users in order to investigate whether the features of the environment provide appropriate information about users' actions and movements within the tool. This was also done in order to test the system before its use in the real life groupware evaluation described in Chapter 9, to ensure that information was presented in a clear way. Feedback from the users is used in order to improve the tool.

The study was conducted in order to investigate: 1) Would users find CAWS helpful when collaboratively authoring a paper? 2) How does the system compare to their current collaborative strategies? 3) What visual strategies are useful to support awareness? 4) How can the design of CAWS be improved?.

### 8.1 Methodology

Fifteen subjects participated in the study, divided into five groups of three participants. These participants were each working in academia (all were either researchers or graduate students). Each group consisted of people who had socialized in the past either professionally or socially. Four of the users had never taken part in a collaborative authoring activity before. Table 8.1 details the group social relationships, gender and current employment status and subject of study. These details are important in order to show the variety in the subject participants.

Group	Users	Relationship	Sex	Employment	Subject
Group 1	3	Co-workers	F	Graduate	Computer Science
			M	Graduate	Computer Science

			F	Researcher	Computer Science
Group 2	3	F-M friends	F	Graduate	Computer Science
		F-M co-workers	M	Graduate	Computer Science
		M-M unfamiliar	M	Web Developer	Multimedia
Group 3	3	F-F friends	M	Graduate	Computer Science
		F/F – M unfamiliar	F	Graduate	Information Technology
			F	Graduate	Computer Science
Group 4	3	M-M friends	M	Graduate	Computer Engineering
		M/M – F unfamiliar	M	Graduate	Computer Science
			F	Graduate	Computer Science
Group 5	3	Co-workers	M	Graduate	Computer Science
			M	Graduate	Computer Science
			M	Graduate	Computer Science
Group 6	3	Co-workers	M	Graduate	Computer Science
			F	Graduate	Computer Science
			M	Post Doc Researcher	Computer Science

**Table 8.1:** Group distribution, sex, social relationship, current employment status and subject of study.

Each group was directed to the CAWS login page. Users were seated in different locations within the building so that they could not interact face-to-face. None of the users knew who was participating in the experiment with them (in some cases the participants were friends and they did not know that they were writing together). This was done deliberately since we did not want users to anticipate each others' actions due to any personal relationship with one another. Each subject was given a unique writing scenario (for example: write text by using the editor, restructure the document, reply to a blog post), and given an interaction questionnaire containing questions to answer such as: "Who is online?", "What are they doing?", and "What have they done?". Complete copies of the questionnaires are in Appendix A (Appendix A.1 for user 1, Appendix A.2 for user 2, Appendix A.3 for user 3). Each user had to complete the same sort of tasks although these were performed at different times in order to simulate group interactions. Each scenario was composed of seven parts:

- **User Status:** This is used to understand if the user's movements and current actions could be visualised properly in order to understand who is around and what

they are doing (informal awareness). This also covers workspace awareness since information is displayed as it happens, giving up-to-date information about users' movements and actions within the workspace.

- **Activity Log:** This was tested in order to understand if the users' activities could be easily viewed to provide an overview of recent and past activity within the document, and to understand who has done what (informal awareness).
- **Awareness in the Editor:** This tested users' editing awareness by judging the level of understanding of who is actively writing in the document (informal and workspace awareness) and hence what sections are being written.
- **Writing in the System:** This tested whether CAWS could be used to actually write the document, although this was mainly to test whether the editor contained bugs when multiple authors were working simultaneously.
- **Roles and Responsibilities:** This part of the usability study focused on understanding if group awareness related to roles and responsibilities is properly displayed.
- **Comment:** This part of the survey analysed informal awareness in the form of understanding which parts of the documents have the most annotations and understanding when users' points of view are similar in the document
- **Deadlines:** This part of the survey analysed whether information about group awareness in the form of document deadlines was presented properly

Users were asked to rate how straightforward they found it to locate the information requested, and the usefulness of the information itself. The questions within the scenario were designed to identify whether the features present in the tool were sufficient to make the participants aware of each others' activities and actions. Each group took fifty to sixty minutes to finish the experiment. At the end of the experiment, the subjects participated in semi-structured interviews regarding the tool, to elicit subjective feedback on how well they felt the system had helped them to accomplish the task. This was also done in order to gain insight into their typical strategies for collaborative authoring and to compare the CAWS tool to those strategies. This took fifteen to twenty minutes in total.

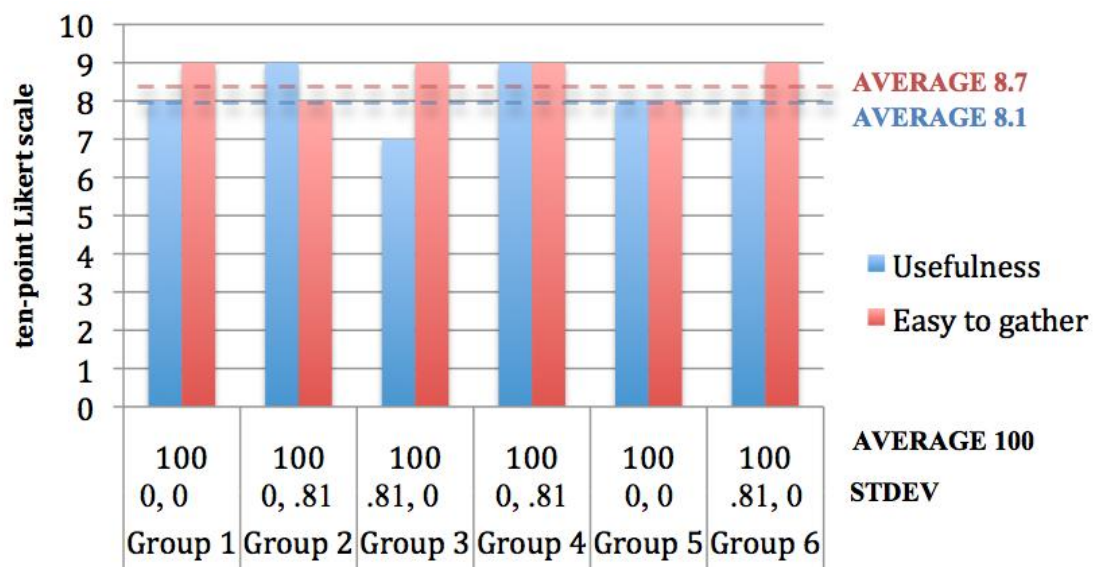


## 8.2 Usage Patterns and Feedback

### 8.2.1 History of users' actions

The 'activity log' feature (as described in Section 7.4.1.1), used to display users' actions within the tool, was considered a highly useful feature, with an average agreement rating of 8.1 on a ten-point Likert scale. Similarly, when asked if the feature made finding information easy, respondents agreed with an average rating of 8.7. Twelve participants commented that the activity log allowed them to be aware of exactly what happens in the document, as well as in the workspace in general. Thirteen participants agreed that having this information displayed with the date on which the action occurred allowed them to gain insight into which days saw the most intense activity. The activity log was found to serve its intended purpose of increasing awareness of users' actions; ten users commented that this information "allowed them to avoid duplicated effort" when noticing that other users had already edited a section in the document that was about to be done.

The test of the 'activity log' feature was one scenario where respondents were able to correctly answer all of the questions that were presented (Figure 8.1).

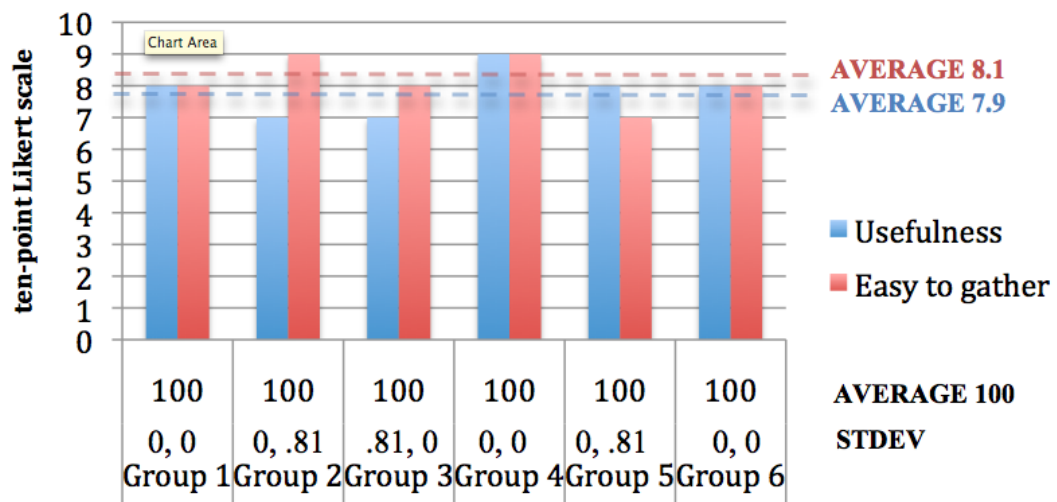


**Figure 8.1:** Results from analysing awareness of users' actions history. The number under the bars shows 100% correct answers to each of the questions given. Usefulness of the feature has an average rating of 8.1, while ease of gathering information about other users' actions has an average of 8.7.

## 8.2.2 Users' Whereabouts

The 'user status' feature (as described in Section 7.4.3.1), which displays users' whereabouts in the tool and their current activity, was considered a highly useful feature, with an average agreement rating of 7.9 on a ten-point Likert scale. When asked if the feature made it easy to find information, respondents agreed with an average rating of 8.1.

During the study, users were asked to switch between tasks to measure the effectiveness of this feature. Figure 8.2 shows that in all cases, users were able to identify the locations of their team members within the system (100%). Participants found this feature useful when using the editor (one participant found this information to be useful only in this case); participants agreed that "having the ability to know who is in the editor and amending sections is important to coordinate everybody's efforts and actions". Six users also agreed that "having this information is useful to be able to plan ahead".

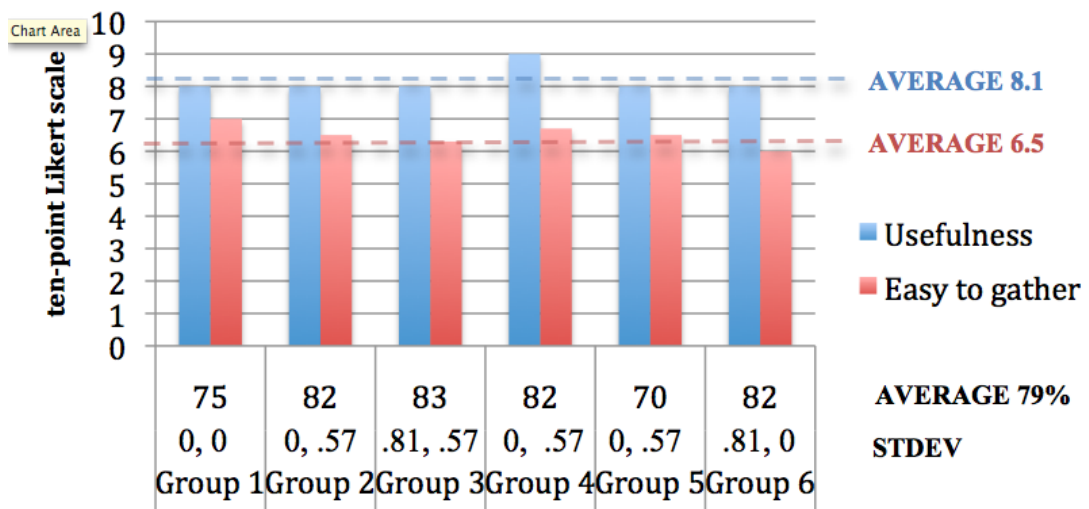


**Figure 8.2:** Results of analysing awareness of users' whereabouts, showing 100% correct answers to all the questions given. The usefulness of the feature had an average rating of 7.9, while ease of gathering information scored 8.1 on average.

## 8.2.3 Division of Activities

The built-in planning features of the tool (as described in Section 7.4.2.1), and the fact that sections are locked while being edited (as described in Section 7.4.3.2), provided participants with better awareness of roles and responsibilities. It also helped provide

awareness of deadlines and the number of hours of work needed to complete the sections. While the feature used to gather information about roles and responsibilities was rated as highly useful by participants (average of 8.1), information regarding users' roles and responsibilities relating to sections was difficult to access, rating 6.5 on a ten point Likert scale. Several users expressed frustration at the need to switch between the editor and the planner to view responsibilities. As a result of this issue, the measurement of correct answers for questions relating to the planner was only 79% on average (Figure 8.3), as questions related to roles and responsibilities were presented in

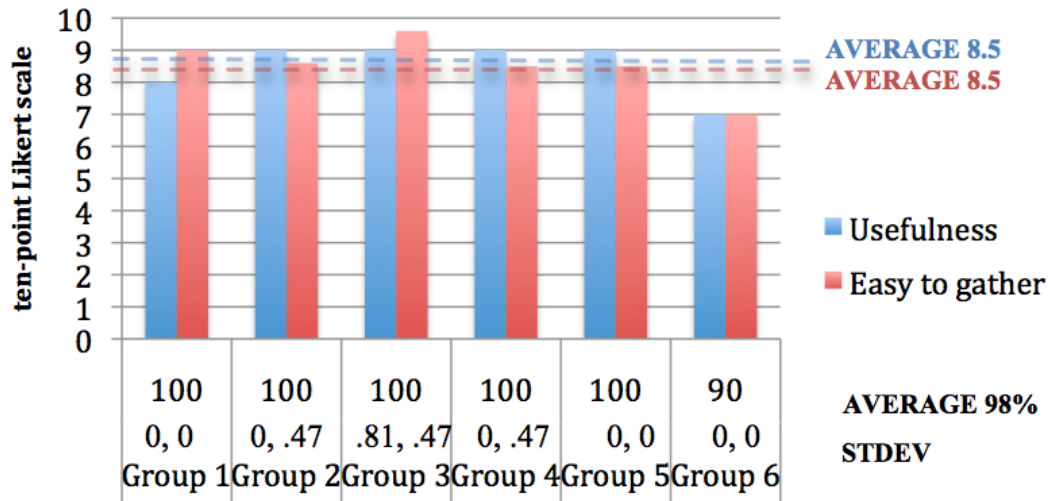


**Figure 8.3:** Results of analysis of awareness of roles and responsibilities. 79% of questions were correctly answered on average. Usefulness of the feature was rated 8.1 on average, while ease of gathering this information was rated 6.5 on average.

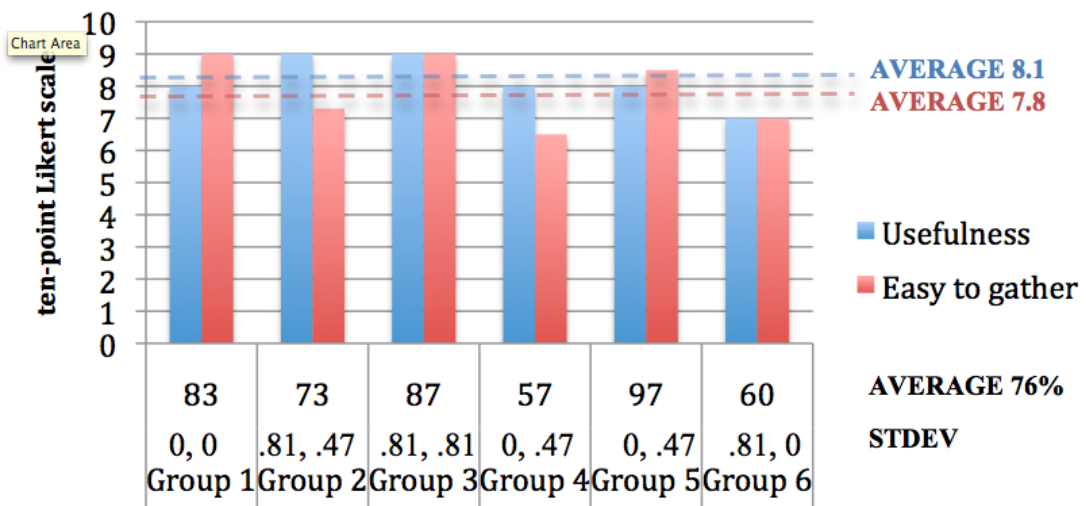
different parts of the questionnaire; six users got the wrong answer in the first instance.

The ability to monitor deadlines with respect to date and hours of work remaining was rated highly useful (8.5) and easy to access (8.5), with an average of 98% of questions correctly answered (members of group 6 were unsure where the information was stored at the beginning of the questionnaire so replied wrong to three questions each) (Figure 8.4).

The ability to see who is writing within the editor by locking the sections and displaying the user's name was rated highly useful (8.1) and easy to access (7.8). However the information was slightly out of sync with the user status box which caused confusion in responding correctly to the questions causing an average of 76% of correct answers (Figure 8.5).



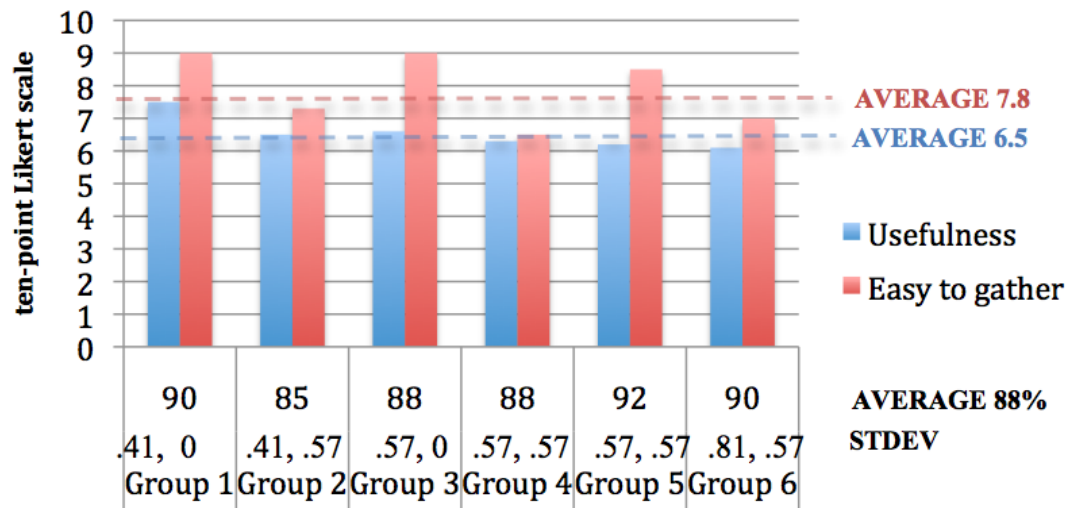
**Figure 8.4:** Results of analysis of awareness of document deadlines showing 98% correct answers on average, and 8.5 average on usefulness of the features, with 8.5 on average on ease of gather information about the document deadlines.



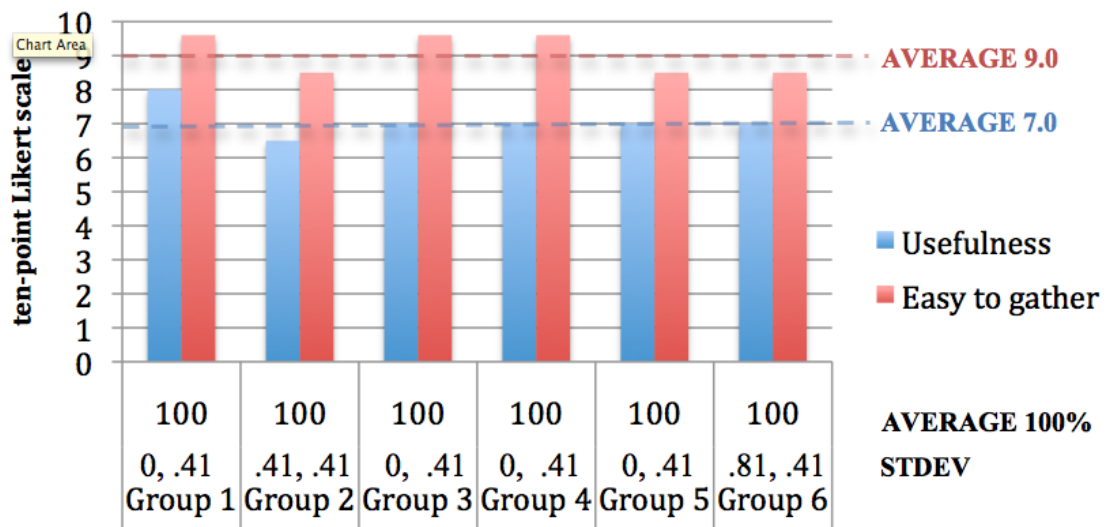
**Figure 8.5:** Results of analysis of users' editing awareness showing 76% correct answers to all the questions given on average, 8.1 average rating on usefulness of the feature, with 7.8 average on ease of gathering information.

### 8.2.4 Use of colour for visualisation

The CAWS tool presents awareness of annotations in two different ways (as described in Section 7.3.1); colour visualization and filtering. When comments are attached to a document in the form of annotations, the text is highlighted. The colour of the highlight



**Figure 8.6:** Results of analysis of users’ annotation awareness, focusing on understanding which parts of the document have more annotations, showing 88% correct answers on average, 6.5 average on usefulness of the feature, with 7.8 average on ease of gathering information



**Figure 8.7:** Results of analysis of users’ annotation awareness, when focused on understanding which author has inserted more annotations to the document, showing 100% correct answers to all the questions given, 7.0 on average for usefulness of the feature, with 9.0 on average for ease of gathering information.

increases in intensity when more comments are added to a portion of text. This feature rated an average of 7.5 in usefulness (on ten-point Likert scale) and 7.8 on ease of gathering information (Figure 8.6). Users commented “this information alerts me to portions of the document where there may be disapproval”. This information rated 88% for correctness when users where asked to identify portions of the document with more

annotations (comments in particular) and rated 100% correctness for portions of the document with more annotations by a specific user.

Average usefulness was rated at 7, while ease of gathering answers was rated 9 (Figure 8.7). Users commented that this information was useful when wanting to understand the opinions of a particular user, whom they might want to monitor because of agreement or disagreements in past collaboration.

### 8.3 Discussion

This chapter described a usability study of the CAWS tool. This usability study proved helpful in understanding the usefulness of the CAWS awareness mechanisms, how easy this information was easily accessed by the users, and how useful this information was rated by the participants. From the results of this study we can infer that CAWS may prove helpful in a real life collaborative authoring project since feedback from participants has suggested that the system has allowed participants to gather information about others participants movements, action, motivations, role, responsibilities and section deadlines.

In this study, user roles and responsibilities were the features that were hardest to access, although once users understood how to access them, they proved useful and the questions on roles and responsibilities were answered correctly. Users were not given any prior demonstration of how to use the tool. A greater number of correct answers therefore occurred when the users determined how to use the system. For this study this research did not identify this as a problem, since this information could be easily found in a help menu, or through a demonstration of the tool.

The information regarding users' statuses was slightly out of sync with the user status box, causing confusion in responses to questions related to users' whereabouts. Because of this, the status box feature was improved by adding immediate update of users' movements within the system.

Responses related to annotation awareness were rated more easy to gather than useful. This result stems from the fact that not many annotations were present within the user experiment scenarios, since the primary objective of the experiments was to ensure ease of gathering information. Information about usefulness of this feature might have been discovered through long term usage of the system.

In feedback from participants, it was commented that having this information about other participants made the user feel *“like I am actually in control of the document”*. Another participant commented, *“having this information can really improve the document writing, I really like it!”*, while another commented, *“I was not sure where to get the answer to the questions at first and then I glanced at the screen and saw all of them there, if I were to work with another person in a distant environment this would be the tool I use”*.

The results examined in the previous sections demonstrate that participants were in strong agreement in their responses, as their feedback in each case was within 0.81 standard deviations.

Following this study, in order to improve the information that users maintain about one another, a “spy mode” feature was inserted to allow users to see what other users are writing in real time.

## 8.4 Conclusion

This study has examined the reactions of users to the CAWS interface. Testing the system is essential in order to evaluate the effectiveness of the system and the principles on which it is designed. This test involved users in an artificial scenario that ran for a very limited amount of time, and because of this, the usefulness of the results is limited and cannot be used as a definitive or conclusive study of the system. However, the results are nonetheless useful as a source of initial feedback to identify major flaws in the system.

Several of the questions were correctly answered by all groups (Sections 8.2.1, 8.2.2, 8.2.4,), demonstrating that the system helps to satisfy the types of awareness with which these questions are associated. In each case, the information scored highly in terms of ease of gathering information (greater than 8 for all). This gives some early, albeit not definitive, evidence that the features to support these types of awareness are functioning as intended.

Other features did not score as highly. For example, identifying roles and responsibilities was considered difficult to identify by participants, specific comments expressing confusion in the need to switch between editor and planner to access the information (Section 8.2.3).

Overall, this experiment has provided an interesting preliminary study of the system, and has yielded some positive and encouraging results. More extensive experimentation is nonetheless required in order to give more conclusive results.



## Chapter 9 Experiment 2 - CAWS Groupware Evaluation

This chapter presents the results of an evaluation of the CAWS tool that involved computer science students involved in a group project. The aim of this exercise was to gather responses from users in order to identify the features of the tool that have the greatest effect upon the collaborative authoring process. The students were second year students and the study took place in their second semester over a period of 17 weeks. The study was a progressive evaluation of the tool, divided into three phases.

The first stage consisted of semi-structured interviews with the participants. This was done to gain insight into the coordination mechanisms used by users, so that details of the basic interactions between users and their modes of collaboration could be understood. In the second stage, an observational study was made of draft versions of documents through the development process, in order to observe the stages of development.

Finally, in the third stage, the participants responded to a survey designed to gauge the impact of awareness mechanisms on their collaborative activities. This survey also provided the participants with the opportunity to highlight conflicts and problems that had arisen from awareness problems.

### 9.1 Setting

**Set of user participants:** This research gathered the experiences of 85 students divided into fifteen groups (12 groups of six members, one group of five members and two groups of four members) engaged in collaborative writing activities. These groups consisted of 2<sup>nd</sup> year computer science students engaged in a group project in which they had to perform a programming task and collaboratively produce three documents.

Each group had previously worked together in an earlier study course. Information about the experiences of each group was gathered over a period of 17 weeks.

**Collaborative tasks:** The first report was delivered in the first two weeks of the course. This consisted of a project outline (5 pages) describing the different tasks assigned to each user. This was followed by group dynamics report (15 pages plus Appendix) and a project overview (20 pages plus Appendix), due in the last week of the semester (week 17).

**Tools used:** Different groups used different tools to write their documents. Seven groups used Microsoft Word (one group of four, one group of five and five groups of six members), two groups (of six members each) used MediaWiki and six groups (five groups of six, one group of four members) used CAWS (Liccardi et al., 2008). The MediaWiki users added one feature to their MediaWiki installation: the ability to identify when another user is editing a page, to avoid conflicts.

**Users' Skills:** Each group presented different skill sets, covering novice, average and expert writers and programmers. Groups were assigned by the lecturers in charge of the course using their first year overall marks to balance the skills present in each group. This was done in order to ensure that groups presented the different skills sets needed to complete the project.

## 9.2 Study Methodology

The aim of this research is to investigate different aspect of awareness and how they affect the collaborative document process. In this study we seek to understand:

1. *What kinds of awareness mechanisms are more effective at each stage of the document development process (as highlighted in chapter 2)? In particular we seek to understand:*
  - 1.1 What are the motivations behind this need? What types of information are exchanged and required at each stage?
  - 1.2 How can these awareness mechanisms enhance each stage of the document development process?
2. *What is the importance of planning in the collaborative authoring process?*
  - 2.1 How do participants in a collaboration approach the planning process?

- 2.2 What kind of information is included in the planning process?
- 2.3 How do they keep the plan up-to-date?
3. *What effect does the planning process in a collaborative authoring process have on interactions between participants?*
  - 3.1 Is it sufficient to keep a version of the plan online?
  - 3.2 What effect does a well-structured plan have on the collaborative process?
  - 3.3 What problems commonly encountered during the collaborative authoring process are due to planning?
  - 3.4 What kind of interface design interaction mechanisms are needed to overcome these problems?
4. *How do users' awareness needs compare to the tool that they actually use?*
  - 4.1 Do users perceive the importance of awareness mechanisms differently depending on the tools that they have used?
5. *What type of awareness mechanism has the greatest effect upon the success of the collaboration?*
6. *Do online tools which support users' awareness improve the collaborative authoring process?*

The study involved different phases in order to understand all aspects of the group activity:

1. **Semi-Structured Interviews:** Seven semi-structured in-person group interviews were conducted with the participants (Appendix B). The interviews consisted of 15-20 minute interviews which inquired into the approach taken to writing. In particular, investigations focused on how the groups divided their work to produce a coherent document. This included the activities they were involved in, the responsibilities of group members and their communication mechanisms.
2. **Observational Study:** The stages of development were observed by examining the draft versions of documents throughout the process. We had access to the different versions produced in CAWS and in the MediaWiki installation. Microsoft Word users stored their documents in version control systems such as CVS and Dropbox. We enquired into the stages of development and who had

contributed to them. Comments from group members and their opinions on the activity were collected.

3. **Survey:** Following the conclusion of the project, we gathered information about the experiences of users with an online questionnaire (Appendix C). Particular attention was given to the activities of each member. Group dynamics were examined, specifically the planning activities and how those activities reflected the original plans. Where problems occurred, the reasons were queried, in order to gather the information needed to avoid those problems in future. In addition we asked each group member to rate different awareness mechanisms by order of importance to understand if different modes of collaboration affect the information that needs to be exchanged between group members.

### 9.3 Observations from observational study and semi-structured interviews

In the semi-structured and observational part of the study, we aim to investigate:

1. The role of planning in the writing process, and specifically the effect of awareness on planning (questions 2 and 3 in Section 9.2).
2. What interaction mechanisms are used more than others in different stages of the writing process (question 1 in Section 9.2).

For this investigation, we interviewed users over the timeline of the project in order to gather their experiences, and to understand the types of information required at each stage.

#### 9.3.1 Importance of awareness at each stage

In Section 2.2.5 we described the stages within the collaborative authoring process. From analysis of the student groups (using results from the semi-structured interviews and the observation study) it was observed that different types of awareness are more necessary in some stages than others.

For example, in the *planning stage* (stage 1, Figure 9.1), while workspace awareness, group awareness and personal awareness rated (on a ten point Likert scale) 4.87, 5.46 and 5.01 respectively, informal awareness (who is around and what they are doing) rated 7.62. In responses to interviews, “*inserting group members’ suggestions into the*

*document*” was often mentioned by groups, followed by “*making sure that everybody is assigned part of the project and that they are happy with it*”, and “*making sure that each member’s strengths and knowledge are properly used*”. It therefore follows that informal awareness, consisting of information on who is around and what they are doing, should be important in order to be able to plan the document properly.

By contrast, workspace awareness, consisting of information about changes to the document itself, is of little use in the early stages of the document, as few changes will be occurring. Similarly, group awareness and personal awareness, which relate to the roles and responsibilities of group members, received low scores, as the planning stage is where these roles are assigned.

In the *drafting stage* (stage 2, Figure 9.1), group awareness and personal awareness received the much higher ratings of 7.73 and 7.55 respectively. Users commented that knowledge of each others’ roles and responsibilities during this stage was important in order to provide balance by ensuring that everyone is working equally. In a real world scenario, different members of a group will contribute different amounts of work to a project. However, in the case of the student groups, it was a requirement that each member spend at least 100 hours of time in order to be awarded a mark. In a real life scenario it would therefore be useful to know the amount of work each group member has committed to the task.

In three of the fifteen groups, participants rotated roles (editor, second author, reviewer and formatter) from section to section, to ensure that all group members participated equally.

In this stage, workspace awareness and informal awareness rated 4.95 and 5.10 on average respectively. Users highlighted that in this stage, it is not necessary to know what progress has been made, it is only necessary to have up-to-date knowledge of users’ roles and responsibilities.

In the *developing stage* (stage 3, Figure 9.1), workspace awareness received a rating of 7.73 while informal awareness rated 7.9, demonstrating that this kind of awareness of users’ actions and movements within the workspace is considered important during this stage. Users commented that knowledge of who is present and their current activities is useful when resolving conflicts. This knowledge can also help to prevent duplication of effort, particularly towards the end of this stage (see Section 9.3.3).

Group awareness at this stage received an average rating of 5.84. When questioned, users commented that knowledge of who is present and their current activity provided sufficient information to determine who was supposed to write each section of the document. Personal awareness received a rating of 6.4, with users commenting that information about personal roles and responsibilities helped when keeping track of which sections were assigned to each user.

In the *reviewing stage* (stage 4, Figure 9.1), the document itself undergoes review by the authors, who insert comments on issues to be addressed and raise points of discussion. In this stage, group and workspace awareness were rated 7.5 and 7.16 respectively. Users highlighted the need for knowledge of members' roles and responsibilities so that questions and comments can be addressed to the author who is most appropriate. Users also need up-to-date information on the state of the document with respect to changes made as a result of discussion.

Informal awareness at this stage was rated at an average of 6.54. Users commented that knowledge of who is present can facilitate synchronous rather than asynchronous discussion, which has the potential to be more efficient. Personal awareness was rated 4.15 on average. Users responded that in this stage, knowledge of individual roles and responsibilities is not necessary since participants are examining other peoples' work rather than their own.

In the *Formatting stage* (stage 5, Figure 9.1), workspace and informal awareness received ratings of 6.41 and 7.44 respectively. Users expressed a particular need for this type of awareness, as it can be used to identify problems that have not yet been addressed; participants can then avoid duplication of effort. Group awareness was rated slightly lower at 6.10; useful, but not necessary at this stage. Users expressed the desire to know who is in charge of changes in order to expedite any formatting fixes. By contrast, personal awareness rated only 4.13 at this stage, and is of little use since information about other members is more important.

### **9.3.2 Effect of document structure on the planning process**

From the semi-structured interviews and the observational study, all groups were found to have produced a detailed plan of actions for the activity ahead. All groups drew up a written plan, regardless of the tool used. Planning was typically done separately from

the main writing activity, documented either online (in the case of MediaWiki) or distributed by email or version control.

In responses from users, there was a strong consensus that suggested that planning of a document is linked to the process of designing its initial structure. Typically the structure is first outlined, with authors then assigned to each section. Assigning an author to each section allows the group to define the responsibilities of each of the individual members of the group. This was deemed an essential stage in the process, without which collaboration would be impossible.

Planning of various tasks was found to be connected to the document structure. The following planning tasks were identified:

- a. **Roles:** Users (58%) reported that each section needed different roles attached to different users. Several participants reported the need for an editor to ensure that a section was completed (37%). Others (21%) reported that a section might need an author and separate proof-reader
- b. **Hours:** Users (42%) reported the usefulness of knowing the number of hours needed to write each section so that effort could be estimated: to quote one user, “*we know that everybody is doing more or less the same amount of work*”. 21% of participants believed that a time estimate helped to indicate the depth of the detail required for a section.
- c. **Word requirements** <sup>24</sup>: 45% highlighted the importance of estimating the required word count for each section. This was judged important for two major reasons. Firstly for ‘*overall estimation*’; 23% of participants believed it important in order to gain an overall estimate of the completeness of the work. Secondly for ‘*detail estimation*’: 22% believed that a per-section word count is useful when judging the detail required for a section.

---

<sup>24</sup> It has to be noted that since the requirement for this activity was based on word count, the student highlighted this granularity of information. In research environments the page count is more important than the word count, since conference require not to exceed a certain amount of pages and pictures, references, tables can increase the size of the document. Hence even though it was not highlighted in this survey, it is the opinion of the authors that a page length should be added to the granularity of information required within a collaborative authoring system.

- d. **Deadline:** 44% of participants considered the defining of deadlines for sections to be important. 23% of participants pointed out that it is necessary to set a deadline for each section to coordinate group efforts in cases where one section must be completed before another can be written. Secondly, 21% of participants commented that a deadline is useful to monitor progress, as it is possible to know if sections should be reassigned if not completed.
- e. **Short Section Description:** 30% of the users attached short descriptions to sections when defining the initial structure. This helped to clarify the intended content and to ensure that there was no repetition in the document.

### 9.3.3 Breakdowns in communications

Participants highlighted problems when maintaining planning information in a separate document, requiring that the plan be checked in conjunction with the document in order to see who was assigned to each section.

36% of participants (2 MediaWiki users, 29 of Microsoft Word / Latex-style tools) reported experiences where they did not believe the project plan to be up to date. For example, one user reported, *“our plan was done three weeks ago, so I was not sure if [xx] was supposed to write it, so I did”*. 27% of participants (using Microsoft Word / Latex-style tools) reported writing sections that they believed were overdue. 17% of those participants wrote sections not assigned to them because they believed them to be related to their own work. For example, three users reported that they *“knew someone else was assigned to the section but as it was related to the one that they were writing, they started writing it themselves”*. In some cases this behaviour led to duplication of effort, with two participants writing the same section. 10% of users (Microsoft Word users) noted mistakes in the version of the plan they looked at, two users commenting, *“my document headings were not in sync with the plan”*.

43% of users (14 CAWS users, 3 MediaWiki users, 9 of Microsoft Word / Latex-style tools) reported spending time maintaining an updated plan in order to avoid duplicated effort. Some groups wrote the name of the users in charge of the section next to the document headings to identify the authors assigned to each section.

17% of users (14 users using Microsoft Word / Latex-style tools and 2 users using MediaWiki) reported that as editors they were responsible for deciding when the



document was complete. In this role it is necessary to read the entire document in order to check for redundancy. Editors would therefore wait until the last possible minute before reading the document. 11% of participants (2 MediaWiki users, 7 of Microsoft Word / Latex-style tools) identified the need to ask other participants when sections were deemed complete, in order to know when to begin proofreading the document.

The features of the CAWS system were found to be useful in some situations. Several participants not using the CAWS tool reported “panic” when approaching the deadline due to uncertainty over whether their colleagues would complete the sections assigned to them. One user reported “[...] did not reply to my emails, so I did not know if he was working on it. I could not wait and I wrote it”. These concerns were not present amongst CAWS users, as the centralized design ensures that the current status of a document is immediately accessible. In the survey, 52% of users (10 CAWS users, 10 MediaWiki users, 25 of Microsoft Word / Latex-style tools) agreed that people will not necessarily write a section simply because it is assigned to them.

Similarly, some users reported delays to their plans due to an inability to see each others’ work until it was complete. 44% of users of Microsoft Word / Latex-style tools reported dependencies between sections, such that they were forced to wait for someone else before writing their own contribution. Similarly 27% (4 MediaWiki users, 19 of Microsoft Word / Latex-style tools) of these users reported agreed that the ability to see work progressing in real time would have helped. These concerns were not present amongst CAWS users.

CAWS includes features designed to assist planning and management of the writing activity. These features were designed to track the authors assigned to individual sections. However, these features were not used by the participants. It was identified that users did not want to use the planning features as they were not directly accessible from within the editor, only from a separate page, and did not contain all of the information required for the planning. One user commented: “*We were assigning the work during our weekly meetings. We were also writing comments [in the form of a summary of content that needed to be included into the sections] at the top of each of the sections, we were keeping it on a separate paper since CAWS did not have the ability to add comments to headings*”.

One group member reported a misunderstanding in the plan of activity saying: “*one week our plan was out of date since the user in charge of it forgot to upload the right version in SVN, but since A popup in CAWS tells you who is also online. There were no conflicts as each member knew which section they were working on*”. In this case, informal awareness provided by the tool helped to work around deficiencies in the plan that the group had devised.

Although one group tried to use CAWS for the planning of the document, they used other features since the planning feature lacked the ability to attach a description to sections. A group member commented, “*Each group member was assigned work during our weekly meeting. We were also putting comments in italics into the documents on CAWS to clarify things for others*”. Another user commented that it was not as important for his group to plan in detail, “*since changes could easily be tracked, it wasn't a problem*”.

## 9.4 Observations from survey

In the survey part of the study, we aim to investigate:

1. Users’ perception of importance of awareness (question 4 in Section 9.2).
2. The effect of awareness mechanisms on the collaborative authoring process (question 5 in Section 9.2).
3. The effect of tool used on the collaborative authoring process (question 6 in Section 9.2).

For this investigation, participants responded to the survey presented in Appendix C.

### 9.4.1 Skill sets

Each group<sup>25</sup> presented different skill sets, ranging from having novice, average and expert writers and programmers. The groups were assigned by the lecturers of the course using second year marks as an estimation of skill sets in order to create balanced groups. However users were not questioned about their individual skills hence in some

---

<sup>25</sup> In order to respect the privacy of participants, the group numbers do not match the group numbers that were assigned to them during the length of the course.

groups there were more experienced programmers than others, as well as more experienced writers. For this research, users were questioned about their skills and each member's skill was rated. The skill sets and tool used are described in Table 9.1.

		Writer			Programmer			Tools
		Novice	Average	Expert	Novice	Average	Expert	
<b>Groups</b>	<b>1</b>	1,2	3,4	6,5	1,4	3,5	6	CAWS
	<b>2</b>	2	1,3,4,6	5	5,6	3	1,2,4	Word
	<b>3</b>		1,2,3,4,6	5		1,2,3,5,6	4	Word
	<b>4</b>		1,2,3	4,5,6	2,5	3,4,6	1,	CAWS
	<b>5</b>	4	6	1,2,3,5	4,5,6		1,2,3	MediaWiki
	<b>6</b>	6	1,3	2,4,5	2,5	6	1,3,4	CAWS
	<b>7</b>	2,3	1,4,6	5	4,5	1,6	2,3,	OpenOffice
	<b>8</b>		1,2,3,4,5	6		3,	1,2,4,5,6	Word
	<b>9</b>	1	2,4,5	3	2,4	3,	1,5	Word
	<b>10</b>		1,2,3,5,6	4	3	1,2,3,4	5,6	Word
	<b>11</b>		1,3,4,6	2,5		1,4,6	2,3,5	MediaWiki
	<b>12</b>	5	1,2	3,4,6	5,	4,6	1,2,3	CAWS
	<b>13</b>	4,5		1,2,3,6		6	1,2,4,3,5	CAWS
	<b>14</b>		1,3,4	2	4	1,3	2	Word
	<b>15</b>		1	2,3,4		2,3,4	1	CAWS

**Table 9.1:** Users' skill sets – Each user is assigned a random number from 1 to 6, representing their skills in writing and programming within the group.

In order to maximise users' strengths, work in each group was assigned according to these strengths. Firstly, from Table 9.1 it is apparent when there are multiple expert writers (who are in charge of the document), the groups opt for a tool such CAWS or MediaWiki that can be used online by multiple users simultaneously, while in scenarios where there is only one expert writer, tools such as Microsoft Word or OpenOffice are chosen (with the support of emails or version control systems). Since there is only one user who will be in charge of the document, the expert writer either works alone or integrates other users' work into the document, functioning as a central point for other members of the group. In this case the expert writer participated slightly to the programming exercise and concentrated the majority of their hours in writing documents to hand in. This decision was made entirely by the group members according to users' preferences and skill sets.

## 9.4.2 Group structure and methodology

Different group structures and techniques were used by different groups to perform their work, depending on their skill sets. In general, the group structures can be categorised as centralised or distributed.

One user in a group with a centralised structure described their technique as follows: *“One person collated all the work from siothers. Versions were sent to that person directly with the sender indicating whether it was final”*. Another user from a different group stated that they *“avoided merging conflicts by having a Document Manager who would edit the main document, after individuals had written their part”*. Others commented *“Note/draft forms were often passed up to [xxx] who would then compile these into the document ensuring only one person was ever editing the final .doc, this also gave [xxx] the chance to write the stuff in a similar style”* and *“Emails being sent around to each other, specifying what had been changed and what not, and then reaching a consensus on final version”*.

Users who used version control software used naming conventions to differentiate documents, for example, one user commented, *“we had folders for each of the deliverables and every time we updated a new version of the document we changed the name of the previous document to old\_version”*. Members from another group commented, *“we committed it with a different name”* and that they would *“append ‘final’ to the file name”*.

This mode of collaboration created a large overhead of emails for the central user; one document manager commented, *“I was in charge of merging and formatting - I did find material which was repeated across the document which took some doing, and took slightly longer then expected”*. Another user from a different group commented, *“When a conflict arose, I merged the two documents together to create a document that included both mine and the other group members’ additions”*. Another document manager reported problems when using a version control system: *“Most documents were shared using DropBox. When there are conflicts, copies of the file are made, with the user's name appended to the filename. This proved problematic with MS Office programs as they modify metadata when opened, meaning the very act of reading them modified them and caused conflicts”*.

One group which used Microsoft Word decided to work together in labs in order to avoid any kind of conflicts: *“Mainly because our group would meet to write all documentation in the Zepler labs so as to avoid these types of conflict.”* In this case users needed to keep a tight schedule for meetings, taking into account lecture schedules and extra-curricular activity (which was not always feasible in other groups).

By contrast, users who used online tools needed only to log in into their system to access to the latest version of the document. No emails or merging were necessary since they always knew what each member had written. Users commented for example, *“It was easy in CAWS to see the current version of the document”, “You could clearly see it in the CAWS editor”, “You could see who was editing in the CAWS Editor, the spy feature was good to see what other people were writing”*. Similarly, users from one of the groups which used MediaWiki stated that they *“wrote a big warning at the top of the wiki page before starting to edit the page since [xxx] might lock the page as well.”* Another group using MediaWiki installed a plug-in for the system that showed when another user was editing the page in order to avoid merge conflicts. Users of MediaWiki also stated that they could see all the changes made to the document at any given time stating *“The wiki page will always show the latest (correct) version of the document, if you wish to view previous versions, you can click the History tab”*.

From these observations of the users’ collaboration methods, it can be seen that groups which used Microsoft Word or OpenOffice (Latex, Notepad), supported by email or version control, encountered problems or delays due to the overhead required to send versions to one another and to ensure that these were the right versions to use (average rating of 6.92 on ten-point Likert scale). In the case of online systems such as CAWS or MediaWiki, these issues were not encountered at all (average rating of 9.02 on ten-point Likert scale with the MediaWiki rating 8.6 and CAWS rating 9.3). Using online systems assists in version tracking by avoiding merge conflicts. In a system that relies on distribution of documents it is difficult or impossible to tell if another author is editing the same section of text.

### **9.4.3 Perceived importance of different types of awareness**

In Table 9.2 we report the mean and standard deviation of users’ opinions on the importance of different types of awareness within the collaborative authoring

environment. From this table we can see that users who used online tools such as MediaWiki and CAWS regarded awareness information higher than those who used offline tools (eg. Microsoft Word, Latex and OpenOffice).

Awareness Types	Survey Questions	Mean (all)	Online tools				Offline tools	
			CAWS		Media Wiki		Latex, Microsoft Word OpenOffice	
			Mean	STD	Mean	STD	Mean	STD
Personal Awareness	How important was it to know your individual role?	8.47	9.29 (↑)	0.98	9.08 (↑)	0.95	7.56 (↓)	2.07
Personal Awareness	How important was it to know your individual responsibilities?	8.90	9.58 (↑)	0.69	9.25 (↑)	0.82	8.20 (↓)	1.74
Group Awareness	How important was it to know other members' roles?	8.32	9.20 (↑)	0.83	8.58 (↑)	0.95	7.48 (↓)	1.89
Group Awareness	How important was it to know other members' responsibilities?	8.28	9 (↑)	1.05	8.41 (↑)	1.11	7.61 (↓)	2.08
Workspace Awareness	How important was it to know which version is the final version of the document?	8.41	9.17 (↑)	0.98	9 (↑)	0.81	7.56 (↓)	2.28
Workspace Awareness	How important was it to be able to know who has changed the document since you last viewed it?	7.94	8.64 (↑)	0.96	9.08 (↑)	0.75	6.97 (↓)	2.34
Informal Awareness	How important was it to know what edits have been made to the document since you last viewed it?	8.37	9.11 (↑)	0.96	8.91 (↑)	0.95	7.56 (↓)	2.20
Informal Awareness	How important was it to have an overview of the recent document activity (i.e. who changed the document, at what time, what changes have been made, what comments, etc.)?	8.15	9.14 (↑)	0.80	8.91 (↑)	0.95	7.05 (↓)	2.13
Informal Awareness	How important was it to be able to see	7.94	8.67	1.15	8.83 (↑)	0.89	7.02 (↓)	2.52

	who is working at the same time as you?		(↑)					
Workspace Awareness	How important was it to know details of how other members were currently engaged with the document at a specific time?	7.6	8.76 (↑)	1.13	8.75 (↑)	0.72	6.23 (↓)	2.28
Workspace Awareness	How important was it to know which parts are finished?	8.35	9 (↑)	1.05	8.5 (↑)	0.64	7.74 (↓)	2.0
Informal Awareness	How important was it to have an overview of current work on the document?	7.92	8.70 (↑)	0.98 1.02	8.58 (↑)	0.86	7.05 (↓)	2.31
Group Awareness	How important was it to know each member's contributions to the document?	7.35	8.88 (↑)	1.02	7.91 (↑)	0.75	5.84 (↓)	2.47
Group Awareness	How important was it to be able to plan work in advance and know what each user should contribute to the document?	7.95	8.76 (↑)	1.43	8.25 (↑)	0.92	7.15 (↓)	2.31
Workspace Awareness	How important was it to be able to monitor the progress of assigned tasks?	8	9.11 (↑)	0.86	8 (↑)	0.81	7.02 (↓)	2.30

**Table 9.2:** Mean and standard deviation of users' responses to questions about the importance of awareness mechanisms (using a ten-point Likert scale) within their collaborative environment. The responses of groups which used online tools and those which used offline tools (eg. Microsoft Word or OpenOffice) are shown separately.

#### 9.4.4 Tracking users' progress and document status (Informal Awareness)

For users to have a clear understanding of what is going in within the document development process, they need to be able to track the status of the document at each stage. In groups where CAWS or MediaWiki was used, users found this information trivial to retrieve. One user (that used MediaWiki) commented, "*Our Wiki has a system that indicates when the latest version of the document was uploaded [...] these changes are relatively quick and easy to identify*", while a user from another group that

used MediaWiki commented, *“by comparing versions. Wiki gives you all the changes in a nice way so you can clearly see what was changed, by whom, etc.”*

Similarly, users of CAWS commented, *“You could clearly see it in the CAWS revision history”*. In this situation, users which used online collaborative tools such as CAWS or MediaWiki did not seem to encounter the same problems as users using offline collaborative tools since the versions of the document were kept online. The tools helped in the intuition of information due to the fact that the tool retained a history of users’ movements and changes to the document.

Among users who used offline tools, examples of methods used for tracking the progress of users and the status of the document were (users from the same group did not necessarily use the same methods) :

1. Comparing versions (6 users), which rated 5.63 on a ten-point Likert scale.
2. Discussion over email (10 users) which rated 6.78
3. Using a document log which users needed to update themselves (6 users), which rated 7.23
4. Using the log feature of the version control software that they used (7 users) which rated 6.72
5. By glancing at the document (5 users) which rated 4.33

In these groups, a user commented *“When I see a new version in the SVN of a document/page, I look at the comment. And then I compare 2 different versions, to see what's different between them”*. Others commented that changes to the document were identified, *“by reading it or looking at log files.”* and that *“We (especially me) gave detailed messages when committing changes to our SVN repository.”*

In another group a user stated this was achieved, *“often by word of mouth and email”*. One group that used a document control system explaining that, *“the document control page was updated after any large (non spelling/grammar) changes”*. Similarly another user from a different group stated that changes were identified, *“by looking at the text in the document. Unfortunately people rarely filed their work under the correct headings”*.

A clear pattern is visible from these responses. Some users often were not aware of the progress that had been made by other members. In some cases, understanding what



had been changed in the document was potentially a time consuming process, or based on examining information passed between group members, whether by email, document control files or logs. By contrast, in online systems such as CAWS and MediaWiki, this information was easily accessed. One user of CAWS commented that, *“you could clearly see it in the CAWS revision history and by using the “compare” facility.”* (which was rated 9.02). Similarly, in the MediaWiki software this was done by examining the page’s history view (which was rated 9.83). This information was therefore easily accessed and rated highly useful by users.

#### **9.4.5 Tracking up-to-the-minute progress and document status (Workspace awareness)**

Tracking up-to-the-minute information in online systems can only be achieved if the system presents a way to view activity in a synchronous manner. In one of the groups that used MediaWiki, users extended the system with a feature that allowed them to track who was currently editing each page in order to avoid merge conflicts. One of these users commented that they were, *“aware that someone else was editing at the same time as me. To avoid that we were doing the write ups in notepad and then pasting it to the wiki. That would only take a second”*.

Similarly, users of CAWS commented that, *“there was no problem with the coordination of work in CAWS as sections being edited by one user were locked to the others”*. Another user commented, *“I just looked at the ‘user status’ feature, very easy”*; another stated, *“CAWS shows who is logged on and working, very easy and useful”*. On average, users which used this feature rated it 9.2.

Other groups used tools similar to Microsoft Word or OpenOffice, supported by SVN or emails. In these groups, this information was harder to gather. One user stated, *“we didn’t coordinate, we tried to stick to the plan that was outlined in the weekly meetings”* while another commented, *“I did not need to know who was working at the same time as me”*.

Some groups in this category used instant messaging tools to coordinate their work; comments on this included, *“we used IM to coordinate, although it did slow the progress of the document, but only slightly”*, *“Coordination over instant messaging software. People changed their bits locally and added into the main document in our shared Dropbox folder when finished”*. Other users used a collocated environment to

coordinate their work, one commenting that, “*we did not coordinate in this way. People mainly stuck to their assigned sections until the final editing stage which was done in the lab.*”

#### **9.4.6 Effect of the collaborative authoring tool on the collaborative authoring process**

24 out of the 34 users who used CAWS and 12 out of the 12 who used MediaWiki stated that although both tools caused problems with formatting during the final stages (the groups resorted to formatting in Microsoft Word), the systems improved their collaborative experience, since having the document online informed them of who was writing and what the latest changes were, accessible by either looking at the history page in the MediaWiki software, or just by glancing at the latest version in CAWS. The remaining 10 users (one group of six and one of four members) used the CAWS style feature, which included the required style format for the document as a predefined option.

One CAWS user stated, “*the tool was good for the aim of the project. However, we had quite a few problems with formatting (not being able of change table size, text size, font size, underlining etc.) and we ended up modifying the documents in Word before submitting. The collaborative writing was very useful though.*” Another user commented “*Yes, even for my own purposes. Makes collaborative writing much easier. Nice text and PDF export features. Could work in more browsers though<sup>26</sup>.*” Another user commented “*Definitely yes (to the question if the tool improved your document development activity). I will certainly use this tool in the future even for my own purposes. Great for tracking progress and for collective writing. PDF export was a great feature to.*”

Users of MediaWiki also commented “*The Wiki made writing as a group fairly easy provided that people didn't need to write the same section at the same time*”, “*Mediawiki was a good tool which provided us with the features we needed to collaborate together and keep track of each other*”, and “*It was really useful for our report writing. Everyone knew what has been changed etc. Much easier than sending*

---

<sup>26</sup> CAWS was developed to work with Mozilla Firefox, not other browsers.

*emails*". These features helped users to understand what had been going on in the document, and hence what was left to be done.

Users also emphasised the fact that discussion was achieved more easily since document sections could be annotated in CAWS and talk pages used in MediaWiki; it was helpful to have discussion together with the document rather than in email threads. One user who used CAWS<sup>27</sup> stated that they thought the system was, "*great for building up a document easily amongst a team of contributors*", while another stated that "*using CAWS allowed for the ease of discussion and editing between the group*". Another user stated that, "*it was useful as updates happened straight away without having to reload files*". Similarly, users who used MediaWiki stated that, "*having the document as a wiki page helped development since we always knew which one was the latest version [...] also it helped the (online) discussions since they were attached to the document in the talk pages*".

Among the 39 students who used Microsoft Word or OpenOffice, when asked if the tool improved their group writing experience, 18 replied "*no*", 16 stated "*it made no difference*" and 5 stated "*yes*" commenting that "*having a localised copy with SVN comments attached to it was useful to have since it avoided crashes*". From this we can identify that the centralised design of the online tools helped the communication and coordination mechanisms of authors who used them, while for users who used offline tools with a version control repository, the concern was more focused on making sure that the document was saved rather than on coordination mechanisms.

## 9.5 Findings

### 9.5.1 Effectiveness of awareness mechanisms at each stage

The following question was posed in Section 9.2:

---

<sup>27</sup> Users of CAWS did express frustration at the fact that they could not format the document as easily as in Microsoft Word. In this particular case, the groups had to export the document into Microsoft Word and format it by hand before submission.

1. *What kinds of awareness mechanisms are more effective at each stage of the document development process (as highlighted in chapter 2)? In particular we seek to understand:*
  - 1.1 *What are the motivations behind this need? What types of information are exchanged and required at each stage?*
  - 1.2 *How can these awareness mechanisms enhance each stage of the document development process?*

The results of the semi-structured interviews helped to identify the relative importance of different awareness mechanisms at each stage in the document development process. These results were reported in Section 9.3.1. Figure 9.1 summarises the most important awareness mechanisms at each stage.

These results have given insight into the reasons why these different types of awareness are required. In the planning stage, informal awareness provides the most important information needed by participants. After the plan has been devised by the group, group and personal awareness become important in order to create the draft version of the document.

Later, in the development stage, workspace and informal awareness increase in importance, as information about roles and responsibilities and up-to-the-minute knowledge of the status of the document are needed to coordinate authors and avoid conflict. The importance of roles and responsibilities continues in the review stage. In the final formatting stage, informal awareness is necessary to identify problems that have not yet been addressed.

For the final stages of authoring, it would be useful to provide features to assist the editor, so that the status of a particular section can be flagged and so that it is clear when the document is complete. This also has the potential to assist coordination between multiple members contributing to a section.

*This research identified different types of awareness required at different stages in the document development process. The differences have been identified and the importance of different awareness and communication mechanisms have been explained, specifically how these depend on the state and stage of the document throughout the process with respect to users' feedback and knowledge required at each stage (Contribution 4).*

## 9.5.2 Role of planning

The following question was posed in Section 9.2:

2. *What is the importance of planning in the collaborative authoring process?*
  - 2.1 *How do participants in a collaboration approach the planning process?*
  - 2.2 *What kind of information is included in the planning process?*
  - 2.3 *How do they keep the plan up-to-date?*

The importance of planning is highlighted from the results of the observational study, where it was observed that every group created a detailed plan of action before beginning work (Section 9.3.1). Similarly, the ability “to plan work in advance and know what each user should contribute to the document” was rated one of the most important factors in responses to the survey (Section 9.4.3). This was the case regardless of whether the collaboration occurred online or face-to-face, or whether a centralised or distributed development approach was used.

Section 9.3.2 highlighted the importance of the document structure to the planning process. The responses of participants make it clear that the process of defining a document’s structure is key to the initial planning stage of the document. Not only does this provide a useful skeletal framework in which writing can begin, but the process also serves as a foundation for division of labour between the participants. The list of section headings acts as a natural list of tasks to be assigned to participants (Section 9.3.1).

Participants identified the need to associate a number of pieces of information with each section, including *roles* (potentially multiple roles per section), *number of hours* (estimation of required effort), *expected word count* (potentially required to estimate the required level of detail), *deadline* and a *short description* of the intended contents of the section. The required level of detail is reflected in the fact that the planning features in CAWS were insufficient and not linked to the document itself (Section 9.3.3).

Different techniques were used by different groups to keep their plans up-to-date. In some cases, groups maintained a plan separate from the document itself, leading to breakdowns in communication due to group members believing the plan was out of date (Section 9.3.3). The preference that planning be integrated with the document was

reflected in the fact that one of the reasons that the CAWS planning system was not used was its lack of such integration.

*This research has identified the crucial aspects of planning within the co-authoring process and the effect that they have on coordination mechanisms within the group (Contribution 3).*

### **9.5.3 Effect of planning on interactions**

The following question was posed in Section 9.2:

3. *What effect does the planning process in a collaborative authoring process have on interactions between participants?*
  - 3.1 *Is it sufficient to keep a version of the plan online?*
  - 3.2 *What effect does a well-structured plan have on the collaborative process?*
  - 3.3 *What problems commonly encountered during the collaborative authoring process are due to planning?*
  - 3.4 *What kind of interface design interaction mechanisms are needed to overcome these problems?*

The effect of the planning process is demonstrated by the problems that occur in its absence. Common problems reported by users included breakdowns in communication leading to duplication of effort (Section 9.3.3), redundant material added to the document and “panic” when approaching deadlines. Crucially, the plan must be kept up to date; if group members lose confidence in the accuracy of the plan, it can encourage them to disregard it altogether.

Keeping a version of the plan online and accessible to the group members can be sufficient to avoid many problems; coordination can be kept smooth if the plan is accurate and maintained. However, mistakes can still occur that lead to confusion even when great care is taken to meticulously update the plan. Even when students update plans accurately, some problems such as section heading mismatches between the plan and the document (related to who was in charge of the section), can be traced to the plan being maintained as a separate document that can become desynchronised from the main document. In this case, sections of the documents can be reassigned to other

members of the group through the plan, but this change is not immediately obvious within the document itself, resulting in confusion related to who is supposed to write these sections.

From this analysis, some suggestions can be devised for how the interface might be designed to avoid common problems. Firstly, as was previously identified, the interface to any planning system should be personalisable and support a high granularity of information, so that all of the information needed to plan the development of the document can be captured. This was seen in the responses in Section 9.3.2 where users reported on needing different sets of granularity of information (some groups required the authors assigned to each sections, while others were happy in only attaching a description/summary of contributions). Secondly, it is important that any planning interface is linked to the document itself so that common problems associated with the use of a separate plan can be avoided.

The ability to see work progressing and to know that necessary sections have been written is also important. Without this form of awareness, participants can be tempted to disregard the work plan and duplicate work that they believe has not been written. These issues were avoided in the CAWS system, where it can immediately be seen if a section has been written or is in the process of being written. Users of MediaWiki reported having written their document sections in a separate word processor, only updating the wiki page for few minutes at time. This was due to the fact that conflicts would arise if multiple authors were editing simultaneously.

*This finding therefore outlines the need for “progressional awareness” (Contribution 3) (Liccardi et al., 2009) which is defined as:*

3. *The connection between planning activities and up-to-date knowledge of the status of other users.*
4. *Knowledge of how the document is planned and how the document is progressing with relation to the plan.*

#### **9.5.4 Effect of tool on perceptions of awareness**

The following question was posed in Section 9.2:

4. *How do users’ awareness needs compare to the tool that they actually use?*

4.1 *Do users perceive the importance of awareness mechanisms differently depending on the tools that they have used?*

It can be seen that in many cases, the tools that users chose did not satisfy their awareness needs. This was particularly true in the case of users of offline tools. Numerous examples from both the observational study and the survey demonstrate the problems that users encountered. These include problems caused by confusion over the plan of action and users' roles and responsibilities (Section 9.3.3), duplication of effort, problems when dealing with versions and merging changes from multiple users (Section 9.4.2), delays due to inability to see work until completed (Section 9.3.3) and the need to resolve redundancy within the document.

Section 9.4.3 compared the responses of users who used online tools against the responses of those who used offline tools. This demonstrated the varying perceptions of the two groups; the users of offline tools perceived the importance of awareness mechanisms as less on average than did the users of online tools. This was the case for every question presented to the users on the survey.

However, although awareness mechanisms were perceived as less important by users of offline tools, this does not mean that these awareness mechanisms are less important. Rather, the need for awareness is typically concentrated on a single person, the document manager. The comments by users who were in this role (Section 9.4.2) demonstrate that most users of offline tools were likely not aware of the problems faced by the group as a whole. This is also reflected in a higher standard deviation among responses from users of offline tools (Table 9.2).

### **9.5.5 Relative importance of awareness mechanisms**

The following question was posed in Section 9.2:

5. *What type of awareness mechanism has the greatest effect upon the success of the collaboration?*

It has been shown that each type of awareness is needed at some stage in the collaboration and that the required type of awareness changes from stage to stage (Section 9.5.1). It is therefore important that each type of awareness be supported for a successful collaboration. However, from the results of this analysis it can nonetheless



be seen that workspace and informal awareness have consistently high scores over the lifetime of the document.

These results highlight the importance of up-to-the-minute knowledge when working in a group. In some cases, this type of information can be used to work around deficiencies in other areas. For example, informal awareness helped users to resolve problems caused by errors in an out of date plan document (Section 9.3.3).

Informal awareness also includes knowledge of the changes that have been made to a document. This information is important in a number of potentially problematic situations, and can help to avoid problems of textual redundancy (duplication of effort), resolving merge conflicts (Section 9.3.3) and tracking progress (Section 9.4.4).

The importance of informal and workspace awareness also provides a conclusive demonstration that the exclusive use of an offline tool is insufficient for effective collaboration. An offline tool is inherently incapable of providing the up-to-the-minute information needed to support informal awareness, and at a minimum must be supplemented by other collaborative tools or communication channels.

*This research has therefore identified the effects that the presence and absence of awareness have on the document development process. It is shown that some problems experienced by users of offline writing tools can be completely avoided by the use of online tools that provide better awareness (Contribution 6).*

### **9.5.6 Effect of awareness on success of collaboration**

The following question was posed in Section 9.2:

6. *Do online tools which support users' awareness improve the collaborative authoring process?*

Section 9.4.6 highlights the responses of users to their choice of tool. In general the users of online tools were positive about their experiences (78%) and believed that their choice helped the collaborative process. By contrast, users of offline tools were in general negative about their choice of tool (46%), or believed that it “made no difference” (41%).

The positive effects of awareness are further highlighted in Section 9.4.3, where users of online tools better recognised the importance of awareness features, compared with

users of offline tools. Specific evidence of the positive effects of awareness can also be found in Section 9.3.3, where several problems that affected users of offline tools (in some cases even inducing “panic”), were not reported by users of online tools at all.

Comments and feedback from users have demonstrated several awareness-supporting features in the online tools that helped them during development: the presence of a document history and the ability to compare versions (Section 9.4.4), the use of informal awareness to resolve problems with an out of date plan (Section 9.3.3), workspace awareness providing information on progression and hence avoiding confusion and duplication of effort (Section 9.3.3), and avoidance of merge conflicts (Section 9.4.2).

*This research has therefore shown that users choose co-authoring tools according to the way they intend to interact with each other (varying from face-to-face to completely online, from centralized to decentralized), underlining the need for different coordination and communication mechanisms within a collaborative authoring tool to address the varying needs of different group profiles and attitudes (Contribution 5).*

## **9.6 Implications**

### **9.6.1 Progressional awareness**

To model the effects of awareness on planning of collaborative activities, we introduce a form of awareness called ‘progressional awareness’. This concept is linked to the previously-identified concept of document writing and planning awareness.

Progressional awareness relates to the connection between planning activities and up-to-date knowledge of the status of other users. It concerns the knowledge that users maintain of how the document is planned and how the document is progressing with relation to that plan. As has been discussed, the lack of this type of awareness can lead to problems relating to the organization of the activity, including duplicated effort, “panic” as to how the document has progressed, or to the plan itself simply being disregarded completely.

We propose that this type of awareness can be supported by presenting relevant planning information within the collaborative authoring tool used by the authors. However, it is important that this information is presented in a way that is properly

integrated with the tool; as it has been found that the information will be ignored if not properly presented.

To investigate these concepts, the CAWS system used in this study is in the process of being extended to add new features designed to support progression awareness. The following is a discussion of the new features along with the reasoning behind their design.

### 9.6.1.1 Who is assigned to edit which section?

Planning a collaborative writing process is crucial to its success as users coordinate their efforts according to the plan. This includes explicitly dividing up a document into sections, assigning roles to users and assigning users to write specific sections of the document, and estimating time of completion.

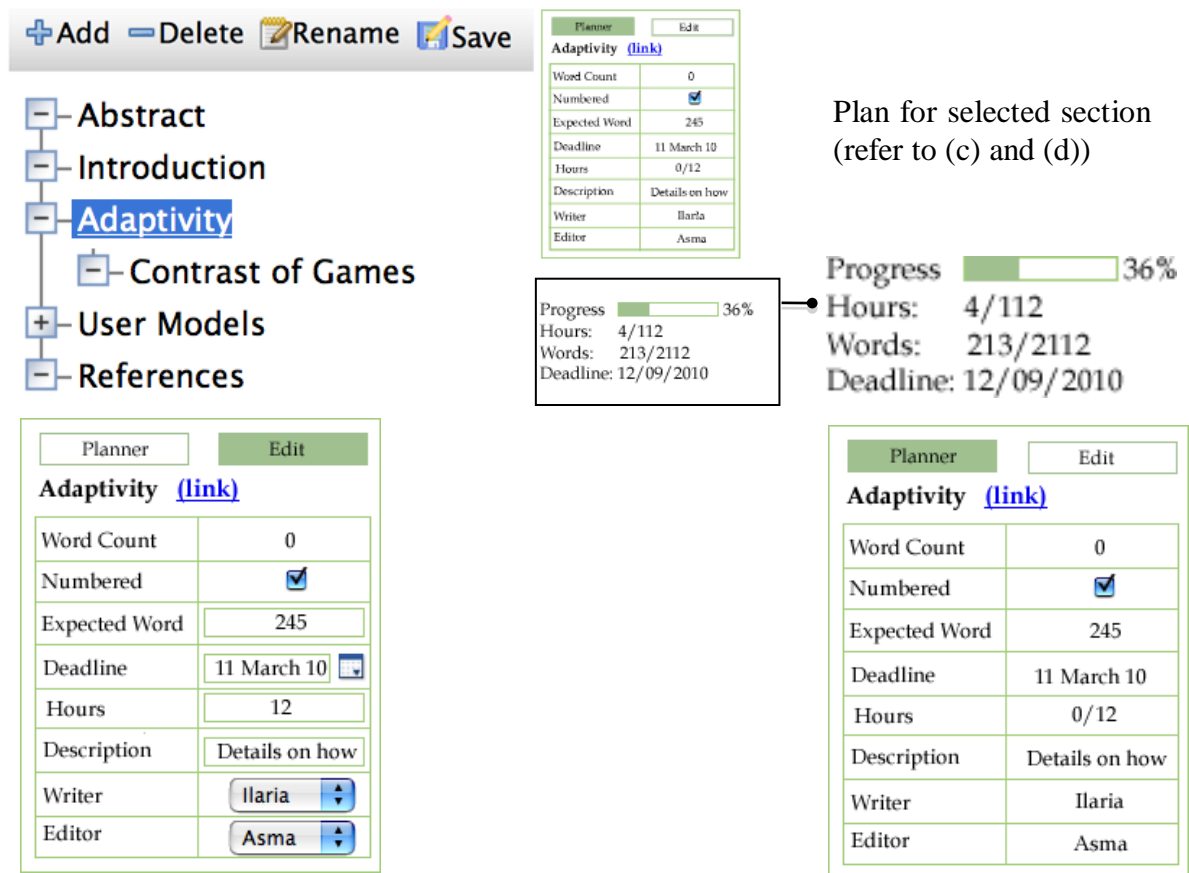
The existing “*structure*” mechanism (Figure 7.7) allows the group members to design an initial structure for the document, defining the sections into which the document is to be divided. This could be extended in order to allow users to be assigned roles in relation to sections (for example, “writer” or “editor”) (Figure 9.2(a)). Users can estimate the time needed for the completion of the section and set a deadline, target word count, section summary (Figure 9.2(c)). This information can be amended as the details of a section can evolve during the process (Figure 9.2(d)). Permission must be granted for this action.

The information entered is shown within the editor in a summarized form under each heading within the document (collapsible so as not to interfere with the text). By integrating the planning information with the document, it is readily available to users while they are working.

In order to give a visual indicator for deadlines, section headings are coloured; sections with a deadline within a week are shown red, while sections with a deadline within two weeks are shown yellow. Similarly, the users assigned to a section are shown along with their roles. These features provide an overview of the section status. The same information is also displayed in a summarized form in the document structure editor, providing an overview of the document plan.

### 9.6.1.2 Who is editing which section?

While an author is editing a particular section, that section is locked and cannot be edited by other authors, to avoid duplication and conflicts. While this is happening, the user's name is displayed next to the section heading. In this way, it is possible to see that work on necessary sections is progressing as required. It is also possible to know that the author assigned to a section is the author writing the section. This feature is present in CAWS (Section 7.4.3.2)



**Figure 9.2:** (a) Document structure; (b) Document progress; (c) Editing plan for a section; (d) Summary of the plan for a section.

### 9.6.1.3 What are they writing?

Knowing what other members are doing is crucial to coordinating efforts. However, understanding in detail what is going on can be difficult. CAWS allows users to view what other users are writing in real time. This is used to give a deeper understanding of other users' contributions to the document. Users being watched should also be informed, with a message displayed to show who is viewing (CAWS does not currently do this).

#### 9.6.1.4 What is complete?

It is useful for authors to be able to judge how far a document is from being completed. For this purpose, users assigned to sections can estimate the progress of that section. In the study, users did not make use of this feature even though it could be used to track progress. In order for this information to be effective, it should be shown next to the section heading. It should also be possible to flag sections as complete or ready for review.

### 9.7 Conclusion

This study has given valuable feedback into the effectiveness of the features of the CAWS tool, and has also given direct insight into the collaborative process and the problems which affect users. The survey has given quantitative values for the importance that users attach to awareness mechanisms (Sections 9.4.1, 9.4.3).

In some cases the results have reconfirmed common problems previously identified in Chapter 2 (Section 2.2.5.1). In Chapter 3 these problems were framed in terms of deficiencies in awareness among group members. In this study, it has been possible to provide evidence to demonstrate that this is the case. For example, duplication of effort was previously identified as a problem, and it has been demonstrated that workspace awareness mechanisms can eliminate this problem almost entirely (Section 9.3.3),

The analysis of the relative importance of different types of awareness at different stages in development (Section 9.5.1) has shown that radically different awareness types are needed at different stages, and that the presence of each type of awareness is therefore necessary for the effectiveness of the overall process. However, the particular importance of informal awareness (Section 9.5.5) demonstrates the key need for up-to-the-minute information to be presented within the workspace. This also demonstrates that the exclusive use of an offline tool is insufficient for effective collaboration.

The study has served as a useful large-scale test of the CAWS system and the principles on which the system is designed. Some of the features of the system have proved effective in their goal and demonstrated that awareness can improve the collaborative authoring process (Section 9.5.6). However, some features, such as the planning system

built into the tool, have been shown as flawed and insufficient for their intended task. This knowledge will allow the design of the tool to be refined to improve its effectiveness.

These findings have helped to define the concept of progressional awareness (Section 9.6.1) which it is believed will prove helpful to understanding the role of planning within the document development process. The use of planning to avoid common problems has been highlighted as particularly important (Sections 9.5.2, 9.5.3). This has been used to devise plans for how the CAWS tool might be extended to better support planning.

## Chapter 10 Conclusion and Discussion

This thesis has explored the subject of collaborative authoring systems and experimented with new user interface designs which can potentially be used to improve upon the current state of the art. These new designs have the potential to address common problems that occur in collaborative writing scenarios. The motivations behind this research were twofold; firstly, the relevance of collaborative authoring and its increase in popularity in recent decades, and secondly, the observation that collaborative authoring activities are afflicted by common problems that impair their effectiveness (Section 2.2.5.1).

This chapter concludes the dissertation. The useful conclusions that have been drawn as a result of this research will now be summarised. The objectives behind the research will be re-examined with respect to how successfully they have been achieved, and finally, directions will be described for future research.

### 10.1 Summary of Findings

This research began by exploring different types of collaborative authoring scenario (Chapter 2). A distinction was made between "traditional" face-to-face collaborative authoring scenarios, of the type used by small groups of authors, and large-scale online collaborations such as Wikipedia. An in-depth examination of the traditional format provided insight into how these types of collaboration are structured: analysing the typical relationships between authors, the stages into which they are divided, and the different communication methods used by authors to communicate with one another. The field study of authors with experience of collaborative activities helped to give detail to this analysis.

Importantly, the examination gave insight into the most common problems affecting collaborative authoring. Chapter 3 introduced the concept of awareness as a model

for understanding the knowledge that authors maintain in a collaborative context, describing the different types of awareness identified by previous research. It was then shown that the identified problems could be framed in terms of awareness issues (Chapter 3.2), and that the causes of these problems could be understood as stemming from the awareness information that is missing from online environments.

Chapter 4 analysed the Wikibooks website, using statistical measures to correlate successful books with the effective use of communication mechanisms. It was therefore shown that although online collaborative sites such as Wikibooks differ greatly in structure to traditional collaborative efforts, communication between authors remains a crucial aspect of a successful collaboration.

Having cemented the importance of communications to collaborative writing, and specifically the importance of awareness, Chapter 5 defined the requirements for an effective collaborative authoring system. These requirements were derived from the identified problems with existing systems and the awareness models previously discussed.

Following this, the requirements were analysed against existing tools in Chapter 6. This analysis covered a wide range of tools, with the ultimate conclusion that none of the analysed tools supported all of the required communication and awareness mechanisms.

To further continue this research's investigation into the effects of communication and awareness mechanisms, the CAWS system was developed as a prototype for new interface designs. Chapter 7 introduced CAWS and provided an overview of its features, linking the features of its design to the requirements set out in Chapter 5. The CAWS tool has a design similar to that of a wiki, as during the analysis of existing systems, many of the features commonly found within wiki systems were found to help satisfy the communication and awareness requirements.

Chapters 8 and 9 described the results of two usability studies of the CAWS tool, with particular reference to the effectiveness of the features of the design which were intended to support communications and awareness. While the first study used participants taking part in an artificial collaborative scenario that lasted only a short time, the results were useful in guiding the design of the CAWS system as it developed, and provided early quantitative evidence of the effectiveness of the system.



The second study tested the system more thoroughly, using participants working toward real goals over a much longer time frame. Through monitoring of the participants throughout the project it was possible to analyse the effects of the system on the groups at each stage.

Groups were divided between those who used online systems (such as CAWS and MediaWiki) and those who chose to rely on Microsoft Word for writing their reports. The results of this study demonstrated that features of the CAWS design were successful in addressing some of the problems identified earlier in the research - some of the issues reported by participants who did not use the system were not reported by participants who did use the system (Section 9.3.3). The study also provided insight into the varying importance of different types of awareness at different stages of the development process (Section 9.5.1).

Analysis of these provided insight into how the CAWS system might be further developed (Section 9.6.1).

In summary, this thesis has analysed the common problems affecting collaborative authoring scenarios, provided a theoretical model for understanding the causes of these problems and devised new techniques that attempt to address them. These techniques have been tested by practical experimentation and testing that has demonstrated their success in some situations, and provided useful feedback that will help to further refine the designs.

## **10.2 Research Questions**

Section 1.4 posed the research questions for this research. The identified objectives of this research were to determine whether specific communication and awareness mechanisms can be useful in improving the interactions within an online environment. The success of this research in answering this question can now be reviewed.

Chapter 4 analysed the role of communication mechanisms within the Wikibooks website. The results of this analysis demonstrated that books on the site which used talk pages (a form of asynchronous communication) and predicted pages (a form of planning) achieved featured status in less time than books which did not. It was further demonstrated that books only reached featured status after the authors began to make

more use of talk pages, and that when not used, redundancy would increase. This analysis therefore helped to demonstrate that the use of asynchronous communication mechanisms does improve interactions between users.

Similarly, this research also aimed to investigate the effects of synchronous (real-time) discussion systems. The results of the second user study (Chapter 9) gave some insight into this issue. Multiple users reported having used real-time discussion systems (eg. instant messaging) during development to positive effect.

Issues can often be resolved more quickly when using a synchronous discussion system such as instant messaging, instead of an asynchronous system such as e-mail. A synchronous system presents a lower communications overhead and lower latency. During testing, users of Microsoft Word needing to resolve conflicts were forced to send versions of documents over e-mail to exchange changes and resolve problems. By contrast, users of online systems such as CAWS were able to resolve problems more quickly as the system would instantly update.

The research questions queried the effectiveness of various types of awareness related to collaborative authoring. Users' actions awareness (concerning up to the minute knowledge of what people are currently doing) was represented within the design of CAWS by the user status box displayed in the corner of each page. During testing, users commented that this expedited communications with other users; for example, they could see if another user was writing a comment.

A subset of action awareness is users' editing awareness, which concerns the specific case where users can tell what other users are currently writing. Support for this type of awareness was included within the CAWS tool, via the user status window and the presence of the user's name next to the locked sections. The positive effect of this feature was seen within the second user study, where groups using CAWS and MediaWiki were able to use this information to adapt to out of date plans. By contrast, groups which used Microsoft Word suffered problems of duplication of effort. The "spy mode" feature also provided editing awareness that helped to avoid duplication of effort and redundancy within the text.

The advantages of synchronous collaboration mechanisms to resolve issues in an efficient manner have been previously discussed.

However, an interesting point that became evident during the second user study was that the ability to determine who was online helped in choosing to use synchronous discussion over asynchronous discussion. This demonstrates one identified use of status awareness.

Mechanisms to assist planning proved difficult to investigate, as participants in the second user study did not use the planning mechanisms built into the CAWS tool. It is clear that features to support planning should include detail and high granularity. The planning system built into CAWS was avoided as it was incapable of holding all of the information needed for planning. It was also identified that any planning system should be linked to the document itself. This insight has led to refined designs for a more effective planning system (see Figure 9.2).

However, the results of their collaborations did help to underscore the importance of awareness of users' roles and responsibilities. Confusion over roles and responsibilities was seen to occur regardless of the tool used; however, in the case of CAWS, the "spy mode" feature was successfully used to avoid duplication of effort that otherwise occurred in groups using Microsoft Word.

Similarly, results from the second user study highlighted the importance of awareness of deadlines; one group missed such a deadline due to confusion that caused a delay. Many of the student participants commented on the importance of this feature; however, it was made clear that in addition to tracking deadlines, it is important that the system also includes the ability to flag when a section has been completed.

Some of these questions, notably those concerning planning features, have been partially answered but need further investigation with refined designs in order to provide a more definitive answer.

### **10.3 Future Work - Awareness Mechanisms**

This research presents many opportunities for future work and investigation. The results of the two user studies have helped to partially answer most of the research questions that were originally posed. However, there is an opportunity for these issues to be investigated in greater depth by future work.

**Progressional Awareness:** Awareness of planning information is one particular example: the results obtained in this research can potentially be used to design a more

effective planning system, as the requirements for such as system are now better understood.

The CAWS system provides a useful testbed for development of new features to support awareness. As it currently stands, there are many opportunities for new features that could be added to the system and for further development.

It would be interesting to explore new ways to reproduce social awareness within a collaborative authoring tool. Of particular interest is the effect this awareness might potentially have on the interactions of collocated and disperse groups.

## **10.4 Future Work - Improving Peer Review Process**

Since CAWS allows different types of annotations to be created and to be attached to the document, CAWS could prove useful as a peer-review system, since having the document linked with the peer review system could therefore make the review and exchange of ideas more efficient. This system could be extended to allow the reviewers to compile the reviews as they read the paper. As a result, the amount of time needed to review the paper is decreased. The fact that the editor can examine reviewers' annotations directly also has the potential to increase transparency of reviewers' actions, increasing fairness within the peer-review process and reducing the overall required time.

The techniques that will be used in this further research will not try to simulate face-to-face interactions between the actors in the peer-review process; instead, the approach taken will be to analyze the peer review process and develop techniques that improve online interactions, making face-to-face communication less important.

The CAWS tool could differ from previous peer review tools for conference and journal publications in four ways:

1. It could integrate and expand upon a variety of observations and previous models of peer-review practices, including peer-review systems that are currently in use by major publishing houses.
2. It could assist editors in the peer review process by making the process of reviewing quicker, allowing editors to tailor annotations to suit particular review criteria within the publication scope.

3. It could assist the peer review process by making the process of reviewing quicker, allowing reviewers to attach annotations to parts of the manuscript itself. It is therefore not necessary to print a paper copy of the manuscript for annotating, an activity that slows the review process, forcing the reviewer to type up written annotations.
4. It could support editors when reviewing reviewers' points of view, ultimately increasing transparency of the review process and improving transparency and ultimately fairness.

### 10.4.1 Possible problems in the peer review process

The process of peer review of a manuscript can be achieved in different ways, such as through a peer review system or even a simple organized exchange of emails between editors and reviewers.

Regardless of the method used, restrictions sometimes apply to reviews. For example:

- Word limits are usually in place in order to provide a concise statement attached to the paper.
- The reviews might have a word limit, in which case the reviewer might waste words to quote the paragraphs of the paper.
- Similarly, a rebuttal, if submitted, might have a word limit. In this case the author might lose words quoting paragraphs from the review.

Reviews typically consist of free-form text with no fixed structure (reviewers might need to follow a protocol of items; however, these are not categorized). The editor (or meta-reviewer, expert editor), who typically moderates the submitted reviews, is tasked with understanding if the reviewers are in agreement. The editor must identify if multiple reviewers comment on the same issues or if sections of the paper have not been fully covered by the reviewers.

To accomplish this task, the editor must understand if there is conflict between the reviewers' opinions on the points and argument presented in the manuscript. Moreover, the review process must be focused and discussed over several iterations.

A rebuttal might need to address several points, leading to further discussion in future iterations of the review cycle. Long discussions may not be supported effectively by the

structure of the email or peer review system. The reviewers need to determine whether the arguments that were previously proposed by them were addressed in the updated version of the manuscript before reading it again and providing more feedback.

### **10.4.2 Potential improvements to the peer review process**

From this analysis of peer review, several potential issues can be identified. These problems do not always occur and may only be relevant over a long process of reviews. Furthermore, we do not state that peer review is impossible when these issues are not considered, but the process could be expedited and made easier to manage.

Online submission and peer-review systems have often been linked, primarily as doing so simulates the offline workflow previously used before online and web-based systems were introduced.

Peer review could potentially be improved if the protocols could provide:

- The ability to maintain a concise environment.
- The ability for reviewers to express their points of view more specifically with respect to:
  - Their opinions.
  - Whether they agree or disagree with arguments made within the manuscript.
- The ability to facilitate discussion between reviewers, authors and editors.
- The ability to flag suggested changes.

Reviews and rebuttals should ideally be presented in a concise form. Although this is usually considered good practice, in some cases it is necessary due to the presence of a word limit. Therefore, the ability to link reviews with individual paragraphs of the manuscript allows comments to be made in a more precise and directed fashion. Less text is required to explain the context in which a comment applies.

This helps authors when submitting a rebuttal, as it is possible to respond to individual annotations without quoting the paragraph or the comment made.

Reviewers possess different opinions within the manuscript, and may possibly disagree with one another. Disagreements such as these are hard to differentiate within a text editor or peer review system. The reviewers must express their point of view with

respect to the arguments made within the manuscript, according to their personal knowledge (which can include comments, suggestion, critique and agreement/disagreement), and the conference or journal guidelines.

The ability for the editor to define annotation types allows annotations to be structured into well-defined categories. This is useful to the author and editor who can use these categories to filter annotations to specific types in order to better understand reviews.

In order to understand reviewers' opinions, it is important to facilitate commenting with the ability to differentiate types of annotations. In an academic environment, annotations can be categorized into various different types. For example:

- **Grammar:** grammatical corrections to the paper may be necessary. This is not relevant to the arguments made, but important for final publications.
- **Formatting:** formatting or style changes to the manuscript itself may be needed.
- **Critiques:** critiques may be made with respect to an argument made within the paper
- **Suggestions:** suggestions can help to ensure that the paper has a fluent argument.
- **Comments:** comments may identify particular issues that have not been addressed.

Different types of annotation may hold different meanings within the peer review process. For example, a simple formatting error does not influence the editor's decision to accept or reject the paper in the same way that a critique or comment could. These types of annotation are therefore more relevant to the editor when examining a review, or to authors when they are trying to understand why their paper was accepted or rejected. However, the type of an annotation is hard to extrapolate from a simple text-based review stored on the conference submission site.

In academia, research papers must make an argument, a statement that should be the central point for the manuscript. However, it is common that assumptions must be made within a paper, and on occasion reviewers may disagree with these assumptions. A peer review system should support this, as:

- Editors must understand if the reviewers are in agreement with one another. This can include, for example, if two authors regard the point of an argument as a

suggestion to improve the paper or a critique which might lead to the rejection of the paper.

- Reviewers must indicate whether they agree or disagree with comments made by other reviewers.
- Authors can use this to construct their argument when submitting a rebuttal to reviews.

Support for in-depth discussion of a submitted paper within a peer review system could present various advantages. For example:

2. The ability for authors and reviewers to focus their discussion on precise areas of the manuscript. Participants could discuss issues in a threaded system better suited to the discussion process.
3. The ability to provide linked information in the form of videos or appendices to support arguments made. In many journal and conference articles, only the submission results are published. For review purposes, the methodology could be presented in order to provide further support for discussion.

Reviewers might require the authors to make changes to the manuscript in order for it to be ready for acceptance. This only appears in the rebuttal stage of a conference proceedings or journal.

This can help authors to flag reviewers' appointed changes as stated in the review process. Authors must address reviewers' changes and need to be able to flag the changes made for the reviewers to examine. This helps reviewers to ensure that the changes made conform to their specifications.

## **10.5 Conclusion**

This research has investigated the factors that affect the collaborative authoring process and the features that collaborative tools ought to incorporate in order to properly support the social relationships between participants in a collaboration.

It seems clear that designers of existing tools often have a poor understanding of the social aspects of collaboration. There remains much opportunity for improvement upon the state of the art.



Understanding the effects of communication and awareness mechanisms can help to guide the introduction of new features that better support the way that people work in group collaborations.

This research has demonstrated that such features can have a positive effect on the collaborative process. While not all the tested designs were successful, the experimental process has helped to provide deeper insight into the collaborative process that will assist in the development of better designs.

## References

- ADLER, A., NASH, J. & NOËL, S. 2004, TellTable: A Server for Collaborative Office Applications, The 6th international Workshop on Collaborative Editing Systems, CSCW 2004, Chicago, USA,
- BAECKER, R. M., NASTOS, D., POSNER, I. & MAWBY, K. 1993, The User-Centred Iterative Design of Collaborative Writing Software, *ACM Computer Human Interaction*, 399-405.
- BECK, E. E. (1993) A Survey of Experiences of Collaborative Writing. *In M. Sharples (ed.): Computer Supported Collaborative Writing. London, Springer-Verlag, 87-112.*
- BECK, E. E. & BELLOTTI, V. M. E. 1993, Informed Opportunism as Strategy: Supporting Coordination in Distributed Collaborative Writing, European Conference on Computer-supported Cooperative Work, Milan, Italy, 233–248.
- BRADNER, E. & MARK, G. (2002), Why distance matters: effects on cooperation, persuasion and deception, *ACM Computer supported cooperative work*, New Orleans, Louisiana, USA 226 - 235.
- BROOKS, F. P. (1975) The mythical man month. *Addison-Wesley Longman Publishing Co., Inc. Boston MA,*
- BRUSH, B., BARGERON, D., GRUDIN, J. & GUPTA, A. 2002, Notification for shared annotation of digital documents, *ACM CHI*, Minneapolis, Minnesota, USA, 89-96.
- CARROLL, J. M., NEALE, D. C., ISENHOUR, P. L., ROSSON, M. B. & MCCRICKARD, D. S. (2003) Notification and awareness: synchronizing task-oriented collaborative activity *Elsevier Science - Int. J. of Human Computer Studies*, 58, 605–632.
- CASSELL, J., BICKMORE, T., BILLINGHURST, M., CAMPBELL, L., CHANG, K., VILHJÁLMSOHN, H. & YAN, H. (1999), Embodiment in conversational interfaces: Rea, *ACM Human Factors in Computing Systems*, Pittsburgh, Pennsylvania, USA, 520-527
- CHESNEY, T. (2006), An empirical examination of Wikipedias credibility, *First Monday*, 11(11),
- CLEARWATER, S., HUBERMAN, B. & HOGG, T. (1991) Cooperative solution of constraint satisfaction problems *Science*, 254 (5035), 1181-1183.

- COUTURE, B. & RYMER, J. (1991) Discourse interaction between writer and supervisor: A primary collaboration in workplace writing. In *M.M. Lay and W.M. Karis (eds.): Collaborative Writing in Industry: Investigations in Theory and Practice*. Amityville, NY: Baywood Publishing Company, 87–108.
- CUNNINGHAM & CUNNINGHAM, I. Wiki Choice tree. <http://c2.com/cgi/wiki?WikiChoicetree>.
- DECOUCHANT, D., QUINT, V. & SALCEDO, M. R. 1995, Structured cooperative editing and group awareness, CHI,
- DILLON, A. & MAYNARD, S. (1995) Don't Forget to Put the Cat Out' – or Why Collaborative Authoring Software and Everyday Writing Pass One Another by! *The New Review of Hypermedia and Multimedia*, 1, 135-153.
- DOURISH, P., ADLER, A., BELLOTTI, V. & HENDERSON, A. (1996) Your place or mine? Learning from long-term use of Audio-Video communication. *Springer Netherlands, Computer Supported Cooperative Work (CSCW)*, 5, 33-62.
- DOURISH, P. & BELLOTTI, V. 1992, Awareness and Coordination in Shared Workspaces, CSCW, Toronto, Canada, 107-114.
- DOURISH, P. & BLY, S. (1992), Portholes: Supporting Awareness in a Distributed Work Group, ACM CHI, Monterey, California, USA, 541 - 547.
- EDE, L. & LUNSFORD, A. (1990) Singular Texts/Plural Authors: Perspectives on Collaborative Writing. *Carbondale, IL: Southern Illinois University Press*.
- EMIGH, W. & HERRING, S. C. 2005, Collaborative authoring on the Web: A Genre Analysis of Online Encyclopedias, Hawaii International Conference on System Sciences, Big Island, Hawaii, 99a.
- ENCYCLOPAEDIA BRITANNICA INC. (March 2006), Fatally flawed: refuting the recent study on encyclopedic accuracy by the journal Nature, [http://corporate.britannica.com/britannica\\_nature\\_response.pdf](http://corporate.britannica.com/britannica_nature_response.pdf) [accessed 08.09.09].
- FLOYD, S. W., SCHROEDER, D. M. & FINN, D. M. (1994) ONLY IF I'M FIRST AUTHOR": CONFLICT OVER CREDIT IN MANAGEMENT SCHOLARSHIP. *Academy of Management Journal*, 37, 734-747.
- GASTWIRTH, J. L. (1972) The Estimation of the Lorenz Curve and Gini Index. *The Review of Economics and Statistics* 54: 306,
- GHALI, F. & CRISTEA, A. (2008), Evaluation of Authoring for Adaptation and Delivery via Learning Management System, 11th IASTED International Conference on Computers and Advanced Technology in Education, Crete, Greece, 1-6.
- GILES, J. (2005) Internet encyclopaedias go head to head. *Nature*, 438, 900–901.
- GOLDSTEIN, J., MITTAL, V., CARBONELL, J. & KANTROWITZ, M. (2000) Multi-document summarization by sentence extraction. In *NAACL-ANLP 2000 Workshop on Automatic summarization (Association for Computational Linguistics)*, 40-48.

- GOODWIN, C. (1981) *Conversational Organization: Interaction Between Speakers and Hearers*, New York: Academic Press.
- GREENBERG, S., GUTWIN, C. & COCKBURN, A. 1996, Awareness Through Fisheye Views in Relaxed-WYSIWIS Groupware, ACM Graphic Interface, Toronto, Canada, 28-38.
- GRUDIN, J. (1992) Why CSCW Applications Fail: Problems in the Design and Evaluation of Organisational Interfaces. In D. Marca and G. Bock (eds.): *Groupware: Software for Computer Supported Cooperative Work*. Los Alamitos, CA: IEEE Computer Society Press, 552-560.
- GUTWIN, C. & GREENBERG, S. (2002) A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *JCSCW*, 411-446.
- GUTWIN, C., PENNER, R. & SCHNEIDER, K. (2004), Group awareness in distributed software development, ACM Computer supported cooperative work, Chicago, Illinois, USA 72-81
- HILL G.W. (1982) Are  $n + 1$  heads better than one? *Psychological bulletin*, 91, 517-539.
- INABA, A., OHKUBO, R., IKEDA, M. & MIZOGUCHI, R. 2004, MODELING LEARNER-TO-LEARNER INTERACTION PROCESS IN COLLABORATIVE LEARNING: An Ontological Approach to Interaction Analysis, *ISWC*, 23-32.
- ISLAM, A. & INKPEN, D. (2008) Semantic text similarity using corpus-based word similarity and string similarity. *ACM Trans. Knowl. Discov. Data*, 2(2), 1-25.
- JANIS, I. L. (1982) *Groupthink: Psychological studies of policy decisions and fiascoes (2nd ed.)* Boston: Houghton Mifflin,
- KIM, E. & EKLUNDH, K. S. (1998) How Academics Co-ordinate their Documentation Work and Communicate about Reviewing in Collaborative Writing, Interaction and Presentation Laboratory, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm University. Technical Report TRITA-NA-P9815, NADA, August 1998.
- KIM, E. & EKLUNDH, K. S. (2000) Change Representations in Collaborative Writing. Interaction and Presentation Laboratory, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm University. Technical Report TRITA-NA-P0005, NADA, March 2000.
- KITTUR, A. & KRAUT, R. E. (2008), Harnessing the wisdom of the Crowds in Wikipedia: Quality through Coordination, CSCW, San Diego, CA, USA, 37-46.
- KITTUR, A., LEE, B. & KRAUT, R. E. 2009a, Coordination in Collective Intelligence: The Role of Team Structure and Task Interdependence CHI, Boston, MA, USA, 1495-1504.
- KITTUR, A., SUH, B. & CHI, E. H. (2009b), What's in Wikipedia? mapping topics and conflict using socially annotated category structure., ACM 27th Annual CHI Conference on Human Factors in Computing Systems (CHI 2009), Boston MA. , 1509-1512
- KITTUR A. KRAUT R.E. (2008), Harnessing the wisdom of the Crowds in Wikipedia: Quality through Coordination, CSCW, San Diego, CA, USA, 37-46.

- KOWALSKI, G. (1997) Information retrieval systems: theory and implementation. *Kluwer Academic Publishers*,
- KRAUT, R., EGIDO, C. & GALEGHER, J. (1990) Patterns of Contact and Communication in Scientific Research Collaboration. IN INC, L. E. A. (Ed.) *Intellectual teamwork: social and technological foundations of cooperative work*. Hillsdale, NJ, USA.
- LANDAUER, T. K., FOLTZ, P. W. & LAHAM, D. (1998) An introduction to latent semantic analysis. . *Discourse Processes*, 25:, 259–284.
- LERNER, J., PATHAK, P. A. & TIROLE, J. (2006) The Dynamics of Open-Source Contributor. *The American Economic Review*. American Economic Association.
- LEUF, B. & CUNNINGHAM, W. (2001) *The Wiki Way*, Addison Wesley.
- LI, L., ZHOU, K., XUE, G.-R., ZHA, H. & YU, Y. 2009, Enhancing diversity, coverage and balance for summarization through structure learning, World wide web (WWW 2009), Madrid, Spain, 71–80.
- LI, Y., MCLEAN, D., BANDAR, Z. A., O'SHEA, J. D. & CROCKETT, K. (2006) Sentence similarity based on semantic nets and corpus statistics. 18(8):1138–1150, 2006. *IEEE Transactions on Knowledge and Data Engineering* 18(8), 1138–1150.
- LICCARDI, I., DAVIS, H. C. & WHITE, S. (2008) CAWS: Visualizing awareness to improve the effectiveness of co-authoring activities. *IEEE Distributed Systems Online*, 9, 1-6.
- LICCARDI, I., DAVIS, H. C. & WHITE, S. A. 2007, CAWS: Improving users' awareness in collaborative authoring activities., ECSCW 2007 Workshop on The Challenges of Collaborative Work in Global Software Development, Limerick, Ireland, 1-8.
- LICCARDI, I., DAVIS, H. C. & WHITE, S. A. (2009), Progressional Awareness: Designing a Co-authoring Tool to Support the Planning Process, ACM Proceedings of the 27th ACM international conference on Design of communication, Bloomington, Indiana, USA 115-118
- LIH, A. (2004), Wikipedia as participatory journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource, International Symposium on Online Journalism, Austin, TX, USA,
- MCDONALD, D. W., WENG, C. & GENNARI, J. H. 2004, The Multiple Views of Inter-organizational Authoring, CSCW, Chicago, IL, 564-573.
- MCGRATH, E. J. & HOLLINGSHEAD, A. B. (1994) *Groups Interacting With Technology*, Sage Publication , Inc.
- MCLAUGHLIN, M. (1984) *Conversation: How Talk is Organized*, Sage, Beverly Hills.
- MITCHELL, A., POSNER, I. & BAECKER, R. 1995, Learning to write together using groupware, CHI, Denver, 288-295.
- MORÁN A. L., F. J., MARTÍNEZ A. M., DECOUNCHANT D. (2001) Document Presence Notification Services for Collaborative Writing *IEE*, 125-133.

- NAKANISHI, H., YOSHIDA, C., NISHIMURA, T. & ISHIDA, T. (1996), FreeWalk: supporting casual meetings in a network, ACM Computer supported cooperative work, 308 - 314
- NEWMAN, J. & NEWMAN, R. (1992) Three Modes of Collaborative Authoring. In P.O. Holt and N. Williams (eds.): Computers and Writing: State of the Art. Oxford: Intellect Books. 20–28.
- NOËL, S. & ROBERT, J. M. (2004a), Empirical Study on Collaborative Writing: What do co-authors do, use, and like?, CSCW, Chicago, IL, 63-89.
- NOËL, S. & ROBERT, J. M. (2004b) How the Web is used to support collaborative writing. *Behaviour & Information Technology*, 22, 245-262.
- OU, J., FUSSELL, S. R., CHEN, X., SETLOCK, L. D. & YANG, J. (2003), Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks, ACM international conference on Multimodal interfaces, Vancouver, British Columbia, Canada, 242-249.
- PAPADOPOULOU, S., IGNAT, C., OSTER, G. & NORRIE, M. (2006), Increasing Awareness in Collaborative Authoring through Edit Profiling IEEE CollaborateCom, Atlanta, Georgia, USA, 1-9.
- PEDERSEN, T., PATWARDHAN, S. & MICHELIZZI, J. 2004, WordNet::Similarity - Measuring the Relatedness of Concepts, Artificial Intelligence, (AAAI-04), Cambridge, Boston, MA, 1024–1025.
- PENROD, D. (2007) *Using Blogs to Enhance Literacy: The Next Powerful Step in 21st-Century Learning*, Rowman & Littlefield, Lanham.
- POSNER, I. R. & BAECKER, R. M. 1992, "How people write together", System Science, Hawaii, 127-138.
- PREECE, J., SHARP, H., ROGERS, Y. (2002) *Interaction design: beyond human-computer interaction*, New York, John Wiley & Sons.
- RATHWELL, M. A. & BURNS, A. (1985) Information systems support for group planning and decision marking activities. *MIS Quarterly*, 9, 255-271.
- RAYMOND E.S. (2001) *The Cathedral and the Bazaar - Musings on Linux and Open Source by an Accidental Revolutionary*, O'Reilly Media Inc.
- RAYMOND, E. S. (2001) *The Cathedral and the Bazaar - Musings on Linux and Open Source by an Accidental Revolutionary*, O'Reilly Media Inc.
- RIMMERSHAW, R. (1992) Collaborative Writing Practices and Writing Support Technologies. *Instructional Science*. 21, 15-28.
- ROUSSEL, N. 2002, Experiences in the design of the well, a group communication device for teleconviviality, ACM international conference on Multimedia, 146-152
- SAJJAPANROJ, S., BONK, C. J., LEE, M. M. & LIN, M. F. (2007), The challenges and successes of wikibookian experts and wikibook novices: classroom and community collaborative experiences, AERA Annual Meeting, Chicago, IL.,
- SHARPLES, M., GOODLET, J., BECK, E., WOOD, C., EASTERBROOK, S. & PLOWMAN, L. (1993) Research Issues in the Study of Computer Supported

- Collaborative Writing. In M. Sharples (ed.): Computer Supported Collaborative Writing. London: Springer-Verlag, pp. 9–28.
- STEWART, G. L. (2006) A Meta-Analytic Review of Relationships Between Team Design Features and Team Performance *Journal of Management*, 32, 26-55.
- SUN, C., XIA, S., SUN, D. & CHEN, D. 2004,Operational Transformation for Collaborative Word Processing, CSCW, Chicago, IL, 437-446.
- SUN, C., XIA, S., SUN, D., CHEN, D., SHEN, H. & CAI, W. (2006) Transparent adaptation of single-user applications for multi-user real-time collaboration. *ACM Transactions on Computer-Human Interaction*, 13, 531-582.
- TAMMARO, S. G., MOSIER, J. N., GOODWIN, N. C. & SPITZ, G. (1997) Collaborative Writing is Hard to Support: A Field Study of Collaborative Writing. *Computer Supported Collaborative Work: The Journal of Collaborative Computing*, vol. 6, pp. 19–51.
- TANG, A., BOYLE, M. & GREENBERG, S. (2004),Display and presence disparity in Mixed Presence Groupware, ACM Australasian user interface, Dunedin, New Zealand 73-82
- THAGARD, P. 1997,Collaborative Knowledge, *Noûs* 31(2), 242-261.
- TURNEY, P. D. 2001, Mining the web for synonyms: Pmi-ir versus lsa on toefl. , *Machine Learning (EMCL-01)* 491–502.
- VIEGAS, F., WATTENBERG, M. & DAVE, K. 2004, Studying cooperation and conflict between authors with history flow visualizations., *CHI Vienna, Austria*,, 575–582.
- WEISBAND, S. (2006) Maintaining Awareness in Distributed Team Collaboration: Implications for Leadership and Performance. *Enabling Distributed Work*.
- WENG, C. & GENNARI, J. H. 2004,Asynchronous Collaborative Writing through Annotations, CSCW, Chicago, IL, 578-581.
- WHITEHEAD, E. J. & GOLAND, Y. Y. (1999) WebDAV: A Network Protocol for Remote Collaborative Authoring on the Web. *Proceedings Sixth European Conference on Computer-Supported Cooperative Work*, Copenhagen, Denmark, pp. 291–310.
- XIAO, Y., BAKER, P. B., O'SHEA, P. M. & ALLEN, D. W. 2007,Wikibook as college textbook: a case study of college students' participation in writing, editing and using a Wikibook as primary course textbook, AERA Annual Meeting, Chicago, IL., USA,

## **Appendix A - Experiment 1 Materials: Usability Study**

Appendix A includes materials used by users in the three different scenarios used to evaluate the usability of the tool.

- Appendix A.1 contains materials for User 1
- Appendix A.2 contains materials for User 2
- Appendix A.3 contains materials for User 3



## Appendix A.1: User 1 SCENARIO

This test is aimed to understand if users' awareness in a collaborative authoring environment is improved by providing information about users whereabouts, status, actions and movements inside the tool.

For the purpose of this test please use the following information to log in.

url: <http://caws.ecs.soton.ac.uk>

Username: User1 (with capital U)

Password: user1

Document: Scenario Article

Log in into the system and select the document "Scenario Article". Please answer the following questions and insert data into the system when required:

### Section 1 : Background Information

#### 1.1 Are you?

Male  Female

#### 1.2 Do you know what a wiki is?

Yes  No

#### 1.3 Have you ever co-author a paper before?

Yes  No

### Section 2 : User Status

Look at the "front page" of the document. This is the first page that appears after you selected the document you wish to edit.

#### 2.1 Who is online now?

User 2  User 3  User 4

#### 2.2 Where is user 2 in the system?

- Editor                       Style                       Discussion                       Blog                       Document

### 2.2.1 What exactly user 2 is doing?

- Editor and Writing

If is so which of the following section?

- Section 1: Introduction  
 Section 2: Background Research  
 Section 2.1: Research 1  
 Section 2.2: Research 2  
 Section 2.3 Research 3  
 Section 3: Conclusion  
 Section 4: Future Work  
 None

- Looking at the document history

Looking at difference in version?

Which one? \_\_\_\_\_

- Inserting a new Bibliography

- Restructuring the document

- Styling the document

Formatting the document style

Formatting the document page

Inserting a prefix style

- Blog

Replying to a blog post

If so which blog post

Blog post 1: Items on the agenda

Blog post 2: Social meeting in London

Adding a new blog post

If so what is the title? \_\_\_\_\_

- Discussion

Reading a discussion thread

If so which one?

Discussion 1: Assigning Roles

If so please give more information about who started the discussion:

User 1                       User 2                       User 3                       User 4

What time was it started?

---

Who else replied to the post?

User 1                       User 2                       User 3                       User 4

Who agreed?

User 1                       User 2                       User 3                       User 4

Who disagreed?

User 1                       User 2                       User 3                       User 4

Who is neutral?

User 1                       User 2                       User 3                       User 4

 Discussion 2: Items in the document

If so please give more information about who started the discussion:

User 1                       User 2                       User 3                       User 4

What time was it started?

---

Who else replied to the post?

User 1                       User 2                       User 3                       User 4

Who agreed?

User 1                       User 2                       User 3                       User 4

Who disagreed?

User 1                       User 2                       User 3                       User 4

Who is neutral?

User 1                       User 2                       User 3                       User 4

**2.3 Where is user 3 in the system?**

Editor                       Style                       Discussion                       Blog                       Document

**2.3.1 What exactly user 3 is doing?** Editor and Writing

If is so which of the following section?

Section 1: Introduction

Section 2: Background Research

Section 2.1: Research 1

Section 2.2: Research 2

- Section 2.3 Research 3
- Section 3: Conclusion
- Section 4: Future Work
- None

Looking at the document history

Inserting a Bibliography

Restructuring the document

Styling the document

- Formatting the document style
- Formatting the document page
- Inserting a prefix style

Blog

Replying to a blog post

If so which blog post

- Blog post 1: Items on the agenda
- Blog post 2: Social meeting in London

Adding a new blog post

If so what is the title? \_\_\_\_\_

Discussion

Reading a discussion thread

If so which one?

Discussion 1: Assigning Roles

If so please give more information about who started the discussion:

- User 1                       User 2                       User 3                       User 4

What time was it started?

\_\_\_\_\_

Who else replied to the post?

- User 1                       User 2                       User 3                       User 4

Who agreed?

- User 1                       User 2                       User 3                       User 4

Who disagreed?

- User 1                       User 2                       User 3                       User 4

Who is neutral?

- User 1                       User 2                       User 3                       User 4  
 Discussion 2: Items in the document

If so please give more information about who started the discussion:

- User 1                       User 2                       User 3                       User 4

What time was it started?

---

Who else replied to the post?

- User 1                       User 2                       User 3                       User 4

Who agreed?

- User 1                       User 2                       User 3                       User 4

Who disagreed?

- User 1                       User 2                       User 3                       User 4

Who is neutral?

- User 1                       User 2                       User 3                       User 4

## 2.4 Where is user 4 in the system?

- Editor                       Style                       Discussion                       Blog                       Document

### 2.4.1 What exactly user 4 is doing?

- Editor and Writing  
 If is so which of the following section?  
 Section 1: Introduction  
 Section 2: Background Research  
 Section 2.1: Research 1  
 Section 2.2: Research 2  
 Section 2.3 Research 3  
 Section 3: Conclusion  
 Section 4: Future Work  
 None
- Looking at the document history
- Inserting a Bibliography
- Restructuring the document
- Styling the document  
 Formatting the document style

- Formatting the document page
- Inserting a prefix style

Blog

- Replying to a blog post
  - If so which blog post
    - Blog post 1: Items on the agenda
    - Blog post 2: Social meeting in London

- Adding a new blog post
  - If so what is the title? \_\_\_\_\_

Discussion

- Reading a discussion thread

If so which one?

- Discussion 1: Assigning Roles

If so please give more information about who started the discussion:

- User 1                       User 2                       User 3                       User 4

What time was it started?

---

Who else replied to the post?

- User 1                       User 2                       User 3                       User 4

Who agreed?

- User 1                       User 2                       User 3                       User 4

Who disagreed?

- User 1                       User 2                       User 3                       User 4

Who is neutral?

- User 1                       User 2                       User 3                       User 4

- Discussion 2: Items in the document

If so please give more information about who started the discussion:

- User 1                       User 2                       User 3                       User 4

What time was it started?

---

Who else replied to the post?

- User 1                       User 2                       User 3                       User 4

Who agreed?

- User 1                       User 2                       User 3                       User 4

Who disagreed?

- User 1                       User 2                       User 3                       User 4

Who is neutral?

User 1

User 2

User 3

User 4

**2.5 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**2.6 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**Section 3 : Activity Log**

Look at the “front page” of the document. This is the first page that appears after you selected the document you wish to edit.

**3.1 What has happened yesterday in the document content? (Please summarise the changes and who has done them)**

---

---

---

---

---

---

---

---

---

---

**3.2 What has User 2 done today regarding comments?**

---

---

---

---

---

---

---

---

---

---

---

**3.3 What has user 3 done in the Editor (comments, edits)?**

---

---

---

---

---

---

---

---

---

---

**3.4 What has user 4 done today?**

---

---

---

---

---

---

---

---

---

---

---

**3.5 What has generally happened Today? (Please list everything including document content changes and document related features such as adding/comments, adding reference, inserting bibliography, add/reply to blog post or discussion, styling etc.)**

---

---

---

---

---

---

---

---

---

---

---

**3.6 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---



**3.7 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---

**Section 4 : Awareness in the Editor**

**4.1 Go to the “Editor” part of the system. When a user is editing a section in the document, the section is automatically locked. However any other user can see what they are writing in real time. In addition to that an administrator can save the changes and unlocked it if needs be.**

**By looking at the document select which sections are locked. Please also write which user is editing them and how long they have been editing it, their status (online, offline and so on), and who is assigned to the section (more than one user can be assigned to a section).**

	Locked	User who is editing	Status	Which user(s) is assigned to it
Section 1: Introduction	<input type="checkbox"/>	_____	_____	_____
Section 2: Background Research	<input type="checkbox"/>	_____	_____	_____
Section 2.1: Research 1	<input type="checkbox"/>	_____	_____	_____
Section 2.2: Research 2	<input type="checkbox"/>	_____	_____	_____
Section 2.3: Research 3	<input type="checkbox"/>	_____	_____	_____
Section 3: Conclusion	<input type="checkbox"/>	_____	_____	_____
Section 4: Future Work	<input type="checkbox"/>	_____	_____	_____

**4.2 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---

**4.3 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---

**Section 5 : Writing in the System**

**5.1 Find ‘Section 1: Introduction’ and press Edit. Please type the following text:**

*Microsoft Word or similar text editors are often used to work on a document, with email and/or shared directories used to distribute the document. Users may avoid groupware applications for several reasons: they may be unfamiliar with the interface, find the tool too difficult to use or simply believe they are organized enough using a normal text editor.*

*Sometimes a wiki is used to write the document in a parallel matter, with the document later formatted and reviewed prior to the final submission. It is important for people collaboratively authoring a document to be informed about the changes that have been made to the document between versions, when new parts are added to the document and by whom. Being aware of these changes helps the users to better understand the evolution of the document, to more easily cooperate with other users and avoid possible conflicts [insert reference 1].*

*reference 1: Papadopoulou, S., Ignat, C., Oster G., Norrie M., Increasing Awareness in Collaborative Authoring through Edit Profiling CollaborateCom, IEE, Atlanta, GA, USA, 2006, pp. 1-10*

**[HELP TO INSERT A REFERENCE:** To insert a reference you will need to add the citation to the bibliography of the document, which is a tab in the Editor part of the document. After saving it you can insert it by finding the position of the reference in the text and selecting the last icon on the section editor tool bar. You can find the reference by the title.]

**After you have finished DO NOT save the section.**

**5.2 Please rate how easy is to use the editor tool box (1 easiest and 10 hardest)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---

**5.3 Scroll down to ‘Section 2.1: Research 1’ and insert the following comment on the text “This is research 1, this research deals with awareness”.**

**Comment details:**

**Subject:** User 1 Comment

**Text:** This is the comment from User 1.

**[HELP TO INSERT A COMMENT:** To insert a comment, highlight the text where you want to insert the comment and then press C on the keyboard.]

**5.4** Scroll down to ‘Section 2.2: Research 2’ and insert the following critique on the text “This is research 2, this research deals with co-authoring”.

**Critique details:**

**Subject:** User 1 Critique

**Text:** This is the critique from User 1.

**[HELP TO INSERT A Critique:** To insert a critique, highlight the text where you want to insert the comment and then press C on the keyboard.]

**5.5** Please rate how easy is to insert comments (1 easiest and 10 hardest)?

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---



---

**Section 6 : Roles and Responsibilities**

**6.1** What are user 3’s role and responsibility?

	Editor	Writer 1	Writer 2	Other (Please specify)
Section 1: Introduction	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2: Background Research	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.1: Research 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.2: Research 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.3: Research 3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 3: Conclusion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 4: Future Work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____

**6.2** What are user 1’s (your own) role and responsibility?

	Editor	Writer 1	Writer 2	Other (Please specify)
Section 1: Introduction	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2: Background Research	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.1: Research 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____

Section 2.2: Research 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Section 2.3: Research 3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Section 3: Conclusion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Section 4: Future Work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

**6.3 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

**6.4 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

**Section 7 : Comments**

**7.1 By looking in the editor which section has the most comments?**

- Section 1: Introduction
- Section 2: Background Research
- Section 2.1: Research 1
- Section 2.2: Research 2
- Section 2.3: Research 3
- Section 3: Conclusion
- Section 4: Future Work

**7.2 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**7.3 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**7.4 By looking at the sections which paragraph has the most comments?**

---

---

**7.5 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**7.6 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**Section 8 : Deadlines**

**8.1 How many days are left to the project deadline?**

---

**8.2 How many days are left for completing the sections?**

Section 1: Introduction	_____
Section 2: Background Research	_____
Section 2.1: Research 1	_____
Section 2.2: Research 2	_____
Section 2.3 Research 3	_____
Section 3: Conclusion	_____
Section 4: Future Work	_____

**8.3 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

**8.4 Please rate the usefulness of this feature?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

## Appendix A.2: User 2 SCENARIO

This test is aimed to understand if users' awareness in a collaborative authoring environment is improved by providing information about users whereabouts, status, actions and movements inside the tool.

For the purpose of this test please use the following information to log in.

url: <http://caws.ecs.soton.ac.uk>

Username: User2 (with capital U)

Password: user2

Document: Scenario Article

Log in into the system and select the document "Scenario Article". Please answer the following questions and insert data into the system when required:

### Section 1 : Background Information

#### 1.1 Are you?

Male  Female

#### 1.2 Do you know what a wiki is?

Yes  No

#### 1.3 Have you ever co-author a paper before?

Yes  No

### Section 2 : Writing in the System

#### 2.1 Find 'Section 3: Conclusion' and press Edit. Please type the following text:

*Microsoft Word or similar text editors are often used to work on a document, with email and/or shared directories used to distribute the document. Users may avoid groupware applications for several reasons: they may be unfamiliar with the interface, find the tool too difficult to use or simply believe they are organized enough using a normal text editor.*

*Sometimes a wiki is used to write the document in a parallel matter, with the document later formatted and reviewed prior to the final submission. It is important for people collaboratively*

authoring a document to be informed about the changes that have been made to the document between versions, when new parts are added to the document and by whom. Being aware of these changes helps the users to better understand the evolution of the document, to more easily cooperate with other users and avoid possible conflicts [insert reference 1].

reference 1 Decouchant, D., Quint, V., Salcedo, M., R., Structured cooperative editing and group awareness, CHI, 1995.

**[HELP TO INSERT A REFERENCE:** To insert a reference you will need to add the citation to the bibliography of the document, which is a tab in the Editor part of the document. After saving it you can insert it by finding the position of the reference in the text and selecting the last icon on the section editor tool bar. You can find the reference by the title.]

After you have finished **DO NOT** save the section.

**2.2 Go to the structure of the document and insert another section. Call it “Abstract” and insert it before the section 1: “Introduction”. Make it so it is not numbered. Save the changes.**

**2.3 Please rate how easy is to use the editor tool box (1 easiest and 10 hardest)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---

**2.4 Scroll down to ‘Section 2.1: Research 1’ and insert the following comment on the text “This is research 1, this research deals with awareness”.**

**Comment details:**

**Subject:** User 2 Comment

**Text:** This is the comment from User 2.

**[HELP TO INSERT A COMMENT:** To insert a comment, highlight the text where you want to insert the comment and then press C on the keyboard.]

**2.5 Scroll down to ‘Section 2.2: Research 2’ and insert the following critique on the text “This is research 2, this research deals with co-authoring”.**

**Critique details:**

**Subject:** User 2 Critique

**Text:** This is the critique from User 2.

**[HELP TO INSERT A Critique:** To insert a critique, highlight the text where you want to insert the comment and then press C on the keyboard.]



**2.6 Please rate how easy is to insert comments (1 easiest and 10 hardest)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---



---

**Section 3 : User Status**

Look at the “front page” of the document. This is the first page that appears after you selected the document you wish to edit.

**3.1 Who is online now?**

User 1                       User 3                       User 4

**3.2 Where is user 1 in the system?**

Editor                       Style                       Discussion                       Blog                       Document

**3.2.1 What exactly user 1 is doing?**

Editor and Writing

If is so which of the following section?

- Section 1: Introduction
- Section 2: Background Research
  - Section 2.1: Research 1
  - Section 2.2: Research 2
  - Section 2.3 Research 3
- Section 3: Conclusion
- Section 4: Future Work
- None

Looking at the document history

Looking at difference in version?

Which one? \_\_\_\_\_

Inserting a new Bibliography

- Restructuring the document
  
- Styling the document
  - Formatting the document style
  - Formatting the document page
  - Inserting a prefix style
  
- Blog
  - Replying to a blog post
    - If so which blog post
      - Blog post 1: Items on the agenda
      - Blog post 2: Social meeting in London
  - Adding a new blog post
    - If so what is the title? \_\_\_\_\_
  
- Discussion
  - Reading a discussion thread
    - If so which one?
      - Discussion 1: Assigning Roles
        - If so please give more information about who started the discussion:
          - User 1                       User 2                       User 3                       User 4
        - What time was it started? \_\_\_\_\_
        - Who else replied to the post?
          - User 1                       User 2                       User 3                       User 4
        - Who agreed?
          - User 1                       User 2                       User 3                       User 4
        - Who disagreed?
          - User 1                       User 2                       User 3                       User 4
        - Who is neutral?
          - User 1                       User 2                       User 3                       User 4
  - Discussion 2: Items in the document
    - If so please give more information about who started the discussion:
      - User 1                       User 2                       User 3                       User 4
    - What time was it started? \_\_\_\_\_
    - Who else replied to the post?
      - User 1                       User 2                       User 3                       User 4
    - Who agreed?

- |                                 |                                 |                                 |                                 |
|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| <input type="checkbox"/> User 1 | <input type="checkbox"/> User 2 | <input type="checkbox"/> User 3 | <input type="checkbox"/> User 4 |
| Who disagreed?                  |                                 |                                 |                                 |
| <input type="checkbox"/> User 1 | <input type="checkbox"/> User 2 | <input type="checkbox"/> User 3 | <input type="checkbox"/> User 4 |
| Who is neutral?                 |                                 |                                 |                                 |
| <input type="checkbox"/> User 1 | <input type="checkbox"/> User 2 | <input type="checkbox"/> User 3 | <input type="checkbox"/> User 4 |

### **3.3 Where is user 3 in the system?**

- Editor       Style       Discussion       Blog       Document

#### **3.3.1 What exactly user 3 is doing?**

- Editor and Writing  
If is so which of the following section?
- Section 1: Introduction
  - Section 2: Background Research
    - Section 2.1: Research 1
    - Section 2.2: Research 2
    - Section 2.3 Research 3
  - Section 3: Conclusion
  - Section 4: Future Work
  - None
- Looking at the document history
- Inserting a Bibliography
- Restructuring the document
- Styling the document
- Formatting the document style
  - Formatting the document page
  - Inserting a prefix style
- Blog
- Replying to a blog post  
If so which blog post
    - Blog post 1: Items on the agenda
    - Blog post 2: Social meeting in London

- Adding a new blog post

If so what is the title? \_\_\_\_\_

- Discussion

- Reading a discussion thread

If so which one?

- Discussion 1: Assigning Roles

If so please give more information about who started the discussion:

- User 1                       User 2                       User 3                       User 4

What time was it started?

---

Who else replied to the post?

- User 1                       User 2                       User 3                       User 4

Who agreed?

- User 1                       User 2                       User 3                       User 4

Who disagreed?

- User 1                       User 2                       User 3                       User 4

Who is neutral?

- User 1                       User 2                       User 3                       User 4

- Discussion 2: Items in the document

If so please give more information about who started the discussion:

- User 1                       User 2                       User 3                       User 4

What time was it started?

---

Who else replied to the post?

- User 1                       User 2                       User 3                       User 4

Who agreed?

- User 1                       User 2                       User 3                       User 4

Who disagreed?

- User 1                       User 2                       User 3                       User 4

Who is neutral?

- User 1                       User 2                       User 3                       User 4

### 3.4 Where is user 4 in the system?

- Editor                       Style                       Discussion                       Blog                       Document

#### 3.4.1 What exactly user 4 is doing?

- Editor and Writing

If is so which of the following section?

- Section 1: Introduction
- Section 2: Background Research
- Section 2.1: Research 1
- Section 2.2: Research 2
- Section 2.3 Research 3
- Section 3: Conclusion
- Section 4: Future Work
- None

Looking at the document history

Inserting a Bibliography

Restructuring the document

Styling the document

- Formatting the document style
- Formatting the document page
- Inserting a prefix style

Blog

Replying to a blog post

If so which blog post

- Blog post 1: Items on the agenda
- Blog post 2: Social meeting in London
- Adding a new blog post

If so what is the title? \_\_\_\_\_

Discussion

Reading a discussion thread

If so which one?

Discussion 1: Assigning Roles

If so please give more information about who started the discussion:

- User 1                       User 2                       User 3                       User 4

What time was it started?

\_\_\_\_\_

Who else replied to the post?

- User 1                       User 2                       User 3                       User 4

Who agreed?

- User 1                       User 2                       User 3                       User 4

Who disagreed?

- User 1                       User 2                       User 3                       User 4

Who is neutral?

- User 1                       User 2                       User 3                       User 4

Discussion 2: Items in the document

If so please give more information about who started the discussion:

- User 1                       User 2                       User 3                       User 4

What time was it started?

---

Who else replied to the post?

- User 1                       User 2                       User 3                       User 4

Who agreed?

- User 1                       User 2                       User 3                       User 4

Who disagreed?

- User 1                       User 2                       User 3                       User 4

Who is neutral?

- User 1                       User 2                       User 3                       User 4

**3.5 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

- 1       2       3       4       5       6       7       8       9       10

Please explain the motivation behind the rating

---



---



---



---

**3.6 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

- 1       2       3       4       5       6       7       8       9       10

Please explain the motivation behind the rating

---



---



---



---

**Section 4 : Activity Log**

Look at the “front page” of the document. This is the first page that appears after you selected the document you wish to edit.

**4.1 What has happened Today in the document content? (Please summarise the changes and who has done them)**

---

---

---

---

---

---

---

---

---

---

**4.2 What has User 1 done Today regards the comments?**

---

---

---

---

---

---

---

---

---

---

**4.3 What has user 3 done Today in the Editor?**

---

---

---

---

---

---

---

---

---

---

**4.4 What has user 4 done Today in the Editor?**

---

---

---

---

---

---

---

---

---

---

**4.5 What has generally happened Today? (Please list everything including document content changes and document related features such as adding/comments, adding reference, inserting bibliography, add/reply to blog post or discussion, styling etc.)**

---



---



---



---



---



---



---

**4.6 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---

**4.7 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---

## Section 5 : Awareness in the Editor

**5.1 Go to the “Editor” part of the system. When a user is editing a section in the document, the section is automatically locked. However any other user can see what they are writing in real time. In addition to that an administrator can save the changes and unlocked it if needs be.**

**By looking at the document select which sections are locked. Please also write which user is editing them and how long they have been editing it, their status (online, offline and so on), and who is assigned to the section (more than one user can be assigned to a section).**

Locked	User who is editing	Status	Which user(s) is assigned to it
--------	---------------------	--------	---------------------------------



Section 1: Introduction	<input type="checkbox"/>	_____	_____	_____
Section 2: Background Research	<input type="checkbox"/>	_____	_____	_____
Section 2.1: Research 1	<input type="checkbox"/>	_____	_____	_____
Section 2.2: Research 2	<input type="checkbox"/>	_____	_____	_____
Section 2.3: Research 3	<input type="checkbox"/>	_____	_____	_____
Section 3: Conclusion	<input type="checkbox"/>	_____	_____	_____
Section 4: Future Work	<input type="checkbox"/>	_____	_____	_____

**5.2 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---



---

**5.3 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---



---

**Section 6 : Roles and Responsibilities**

**6.1 What are user 3's role and responsibility?**

	Editor	Writer 1	Writer 2	Other (Please specify)
Section 1: Introduction	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2: Background Research	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.1: Research 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.2: Research 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.3: Research 3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 3: Conclusion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 4: Future Work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____

**6.2 What are user 1's (your own) role and responsibility?**

	Editor	Writer 1	Writer 2	Other (Please specify)
Section 1: Introduction	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2: Background Research	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.1: Research 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.2: Research 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.3: Research 3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 3: Conclusion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 4: Future Work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____

**6.3 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---



---

**6.4 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---



---

**Section 7 : Comments**

**7.1 By looking in the editor which section has the most comments?**

- Section 1: Introduction
- Section 2: Background Research
- Section 2.1: Research 1
- Section 2.2: Research 2
- Section 2.3: Research 3
- Section 3: Conclusion
- Section 4: Future Work

**7.2 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**7.3 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**7.4 By looking at the sections which paragraph has the most comments?**

---

---

**7.5 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**8.5 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**Section 8 : Deadlines**

**8.1 How many days are left to the project deadline?**

**8.2 How many days are left for completing the sections?**

Section 1: Introduction	_____
Section 2: Background Research	_____
Section 2.1: Research 1	_____
Section 2.2: Research 2	_____
Section 2.3 Research 3	_____
Section 3: Conclusion	_____
Section 4: Future Work	_____

**8.3 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

**8.4 Please rate the usefulness of this feature?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

## Appendix A.3: User 3 SCENARIO

This test is aimed to understand if users' awareness in a collaborative authoring environment is improved by providing information about users whereabouts, status, actions and movements inside the tool.

For the purpose of this test please use the following information to log in.

url: <http://caws.ecs.soton.ac.uk>

Username: User3 (with capital U)

Password: user3

Document: Scenario Article

Log in into the system and select the document "Scenario Article". Please answer the following questions and insert data into the system when required:

### Section 1 : Background Information

#### 1.1 Are you?

Male  Female

#### 1.2 Do you know what a wiki is?

Yes  No

#### 1.3 Have you ever co-author a paper before?

Yes  No

### Section 2 : Writing in the System

#### 2.1 Find 'Section 2.1: Research 1' and press Edit. Please type the following text:

Awareness of individual and group activities is critical to successful collaboration and is always required to coordinate group activities, whatever the task domain (Dourish and Bellotti, 1992). Many researchers in the CSCW field have studied the need of awareness in groupware applications. Dourish and Bellotti (Dourish and Bellotti, 1992) defined awareness as "*understanding of the activities of others, which provides a context for your own activity*". Gutwin and Greenberg (Gutwin and Greenberg, 2002) examined workspace awareness as an overlapping of the kind of awareness that are present in our daily life (Greenberg et al., 1996) but reported in an online system as

“the up-to-the-moment understanding of another person’s interaction with a shared workspace” which involves knowledge about where others are working on, what they are doing and what they are going to do next (personal, social, group and informal awareness).

[1] P. Dourish, Bellotti, V., *Awareness and Coordination in Shared Workspaces*, CSCW, ACM, Toronto, Canada, 1992, pp. 107-114.

[2] S. Greenberg, Gutwin, C., Cockburn, A., *Awareness Through Fisheye Views in Relaxed-WYSIWIS Groupware*, ACM Graphic Interface, ACM, Toronto, Canada, 1996, pp. 28-38.

[3] C. Gutwin, Greenberg, S., *A Descriptive Framework of Workspace Awareness for Real-Time Groupware*, JCSCW (2002), pp. 411-446.

**[HELP TO INSERT A REFERENCE:** To insert a reference you will need to add the citation to the bibliography of the document, which is a tab in the Editor part of the document. After saving it you can insert it by finding the position of the reference in the text and selecting the last icon on the section editor tool bar. You can find the reference by the title.]

After you have finished **DO NOT** save the section.

**2.2 Please rate how easy is to use the editor tool box (1 easiest and 10 hardest)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---



---

**2.3 Scroll down to ‘Section 2.2: Research 2’ and insert the following comment on the text “this research undertook backed previous research”.**

**Comment details:**

**Subject: User 3 Comment**

**Text: This is the comment from User 3.**

**[HELP TO INSERT A COMMENT:** To insert a comment, highlight the text where you want to insert the comment and then press C on the keyboard.]

**2.4 Scroll down to ‘Section 2.2: Research 2’ and insert the following critique on the text “collaborative authoring a document and most”.**

**Critique details:**

**Subject: User 3 Critique**

**Text: This is the critique from User 3.**

**[HELP TO INSERT A Critique:** To insert a critique, highlight the text where you want to insert the comment and then press C on the keyboard.]

**2.5 Please rate how easy is to insert comments (1 easiest and 10 hardest)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---

### Section 3 : Blog post

Go to the “Blog” part of the document. And create a new Blog post called User 3 – (insert time of the user study) – post.

Add this text to the blog post

*“Awareness of individual and group activities is critical to successful collaborative authoring. Participants need knowledge of what other people are doing and have done, what meaningful changes are made to a document, who is editing each section of a document and why. This paper identifies the key problems in collaborative authoring through the use of a two stage survey. From this research we identify potential criteria for supporting collaborative authoring through the use of awareness. Existing tools are examined from this perspective. Finally we present a new tool, CAWS, a co-authoring wiki based system, which is designed to enhance workspace awareness in order to improve participants' productivity in collaborative document development”*

### Section 4 : User Status

Look at the “front page” of the document. This is the first page that appears after you selected the document you wish to edit.

#### 4.1 Who is online now?

User 1                       User 2                       User 4

#### 4.2 Where is user 1 in the system?

Editor                       Style                       Discussion                       Blog                       Document

#### 4.3 What exactly user 1 is doing?

Editor and Writing

If is so which of the following section?

- Section 1: Introduction
- Section 2: Background Research
- Section 2.1: Research 1
- Section 2.2: Research 2
- Section 2.3 Research 3
- Section 3: Conclusion
- Section 4: Future Work

- None
  
  - Looking at the document history
    - Looking at difference in version?
    - Which one? \_\_\_\_\_
  
  - Inserting a new Bibliography
  
  - Restructuring the document
  
  - Styling the document
    - Formatting the document style
    - Formatting the document page
    - Inserting a prefix style
  
  - Blog
    - Replying to a blog post
      - If so which blog post
        - Blog post 1: Items on the agenda
        - Blog post 2: Social meeting in London
    - Adding a new blog post
      - If so what is the title? \_\_\_\_\_
  
  - Discussion
    - Reading a discussion thread
      - If so which one?
        - Discussion 1: Assigning Roles
          - If so please give more information about who started the discussion:
 

<input type="checkbox"/> User 1	<input type="checkbox"/> User 2	<input type="checkbox"/> User 3	<input type="checkbox"/> User 4
---------------------------------	---------------------------------	---------------------------------	---------------------------------
          - What time was it started? \_\_\_\_\_
- 
- Who else replied to the post?
 

<input type="checkbox"/> User 1	<input type="checkbox"/> User 2	<input type="checkbox"/> User 3	<input type="checkbox"/> User 4
---------------------------------	---------------------------------	---------------------------------	---------------------------------
- 
- Who agreed?
 

<input type="checkbox"/> User 1	<input type="checkbox"/> User 2	<input type="checkbox"/> User 3	<input type="checkbox"/> User 4
---------------------------------	---------------------------------	---------------------------------	---------------------------------
- 
- Who disagreed?
 

<input type="checkbox"/> User 1	<input type="checkbox"/> User 2	<input type="checkbox"/> User 3	<input type="checkbox"/> User 4
---------------------------------	---------------------------------	---------------------------------	---------------------------------
- 
- Who is neutral?
 

<input type="checkbox"/> User 1	<input type="checkbox"/> User 2	<input type="checkbox"/> User 3	<input type="checkbox"/> User 4
---------------------------------	---------------------------------	---------------------------------	---------------------------------



- Discussion 2: Items in the document

If so please give more information about who started the discussion:

- User 1                       User 2                       User 3                       User 4

What time was it started?

---

Who else replied to the post?

- User 1                       User 2                       User 3                       User 4

Who agreed?

- User 1                       User 2                       User 3                       User 4

Who disagreed?

- User 1                       User 2                       User 3                       User 4

Who is neutral?

- User 1                       User 2                       User 3                       User 4

#### 4.4 Where is user 2 in the system?

- Editor                       Style                       Discussion                       Blog                       Document

#### 4.5 What exactly user 2 is doing?

- Editor and Writing

If is so which of the following section?

- Section 1: Introduction  
 Section 2: Background Research  
 Section 2.1: Research 1  
 Section 2.2: Research 2  
 Section 2.3 Research 3  
 Section 3: Conclusion  
 Section 4: Future Work  
 None

- Looking at the document history

- Inserting a Bibliography

- Restructuring the document

- Styling the document

- Formatting the document style

- Formatting the document page
- Inserting a prefix style

Blog

- Replying to a blog post
  - If so which blog post
    - Blog post 1: Items on the agenda
    - Blog post 2: Social meeting in London

- Adding a new blog post
  - If so what is the title? \_\_\_\_\_

Discussion

- Reading a discussion thread

If so which one?

- Discussion 1: Assigning Roles

If so please give more information about who started the discussion:

- User 1                       User 2                       User 3                       User 4

What time was it started?

---

Who else replied to the post?

- User 1                       User 2                       User 3                       User 4

Who agreed?

- User 1                       User 2                       User 3                       User 4

Who disagreed?

- User 1                       User 2                       User 3                       User 4

Who is neutral?

- User 1                       User 2                       User 3                       User 4

- Discussion 2: Items in the document

If so please give more information about who started the discussion:

- User 1                       User 2                       User 3                       User 4

What time was it started?

---

Who else replied to the post?

- User 1                       User 2                       User 3                       User 4

Who agreed?

- User 1                       User 2                       User 3                       User 4

Who disagreed?

- User 1                       User 2                       User 3                       User 4

Who is neutral?

User 1

User 2

User 3

User 4

**4.6 Where is user 4 in the system?**

Editor

Style

Discussion

Blog

Document

**4.7 What exactly user 4 is doing?**

Editor and Writing

If is so which of the following section?

Section 1: Introduction

Section 2: Background Research

Section 2.1: Research 1

Section 2.2: Research 2

Section 2.3 Research 3

Section 3: Conclusion

Section 4: Future Work

None

Looking at the document history

Inserting a Bibliography

Restructuring the document

Styling the document

Formatting the document style

Formatting the document page

Inserting a prefix style

Blog

Replying to a blog post

If so which blog post

Blog post 1: Items on the agenda

Blog post 2: Social meeting in London

Adding a new blog post

If so what is the title? \_\_\_\_\_

Discussion

Reading a discussion thread

If so which one?

Discussion 1: Assigning Roles

If so please give more information about who started the discussion:

User 1                       User 2                       User 3                       User 4

What time was it started?

---

Who else replied to the post?

User 1                       User 2                       User 3                       User 4

Who agreed?

User 1                       User 2                       User 3                       User 4

Who disagreed?

User 1                       User 2                       User 3                       User 4

Who is neutral?

User 1                       User 2                       User 3                       User 4

Discussion 2: Items in the document

If so please give more information about who started the discussion:

User 1                       User 2                       User 3                       User 4

What time was it started?

---

Who else replied to the post?

User 1                       User 2                       User 3                       User 4

Who agreed?

User 1                       User 2                       User 3                       User 4

Who disagreed?

User 1                       User 2                       User 3                       User 4

Who is neutral?

User 1                       User 2                       User 3                       User 4

**4.8 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1       2       3       4       5       6       7       8       9       10

Please explain the motivation behind the rating

---



---



---



---

**4.9 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**Section 5 : Activity Log**

Look at the “front page” of the document. This is the first page that appears after you selected the document you wish to edit.

**5.1 What has happened Today in the document content? (Please summarise the changes and who has done them)**

---

---

---

---

---

---

---

---

---

---

**5.2 What has User 2 done Today regards the comments?**

---

---

---

---

---

---

---

---

---

---

**5.3 What has user 1 done Today in the Editor?**

---

---

---

---

---

---

---

---

---

---

**5.4 What has user 4 done Today in the Editor?**

---

---

---

---

---

---

---

---

---

---

**5.5 What has generally happened Today? (Please list everything including document content changes and document related features such as adding/comments, adding reference, inserting bibliography, add/reply to blog post or discussion, styling etc.)**

---

---

---

---

---

---

---

---

**5.6 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

---

**5.7 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

---

## **Section 6 : Awareness in the Editor**

**6.1 Go to the “Editor” part of the system. When a user is editing a section in the document, the section is automatically locked. However any other user can see what they are writing in real time. In addition to that an administrator can save the changes and unlocked it if needs be.**

**By looking at the document select which sections are locked. Please also write which user is editing them and how long they have been editing it, their status (online, offline and so on), and who is assigned to the section (more than one user can be assigned to a section).**

	Locked	User who is editing	Status	Which user(s) is assigned to it
Section 1: Introduction	<input type="checkbox"/>	_____	_____	_____
Section 2: Background Research	<input type="checkbox"/>	_____	_____	_____
Section 2.1: Research 1	<input type="checkbox"/>	_____	_____	_____
Section 2.2: Research 2	<input type="checkbox"/>	_____	_____	_____
Section 2.3: Research 3	<input type="checkbox"/>	_____	_____	_____
Section 3: Conclusion	<input type="checkbox"/>	_____	_____	_____
Section 4: Future Work	<input type="checkbox"/>	_____	_____	_____

**6.2 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---

**6.3 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---

**Section 7 : Roles and Responsibilities**

**7.1 What are user 3's role and responsibility (your own)?**

	Editor	Writer 1	Writer 2	Other (Please specify)
Section 1: Introduction	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2: Background Research	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.1: Research 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.2: Research 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.3: Research 3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 3: Conclusion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 4: Future Work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____

**7.2 What are user 1’s role and responsibility?**

	Editor	Writer 1	Writer 2	Other (Please specify)
Section 1: Introduction	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2: Background Research	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.1: Research 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.2: Research 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 2.3: Research 3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 3: Conclusion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____
Section 4: Future Work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> _____

**7.3 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---

**7.4 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---



---



---



---

**Section 8 : Comments**

**8.1 By looking in the editor which section has the most comments?**

- Section 1: Introduction
- Section 2: Background Research
- Section 2.1: Research 1
- Section 2.2: Research 2
- Section 2.3: Research 3
- Section 3: Conclusion
- Section 4: Future Work



**8.2 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**8.6 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**8.3 By looking at the sections which paragraph has the most comments?**

---

---

**8.4 Please rate the usefulness of this feature (10 extremely useful and 1 extremely useless)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**8.7 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**Section 9 : Deadlines**

**9.1 How many days are left to the project deadline?**

---

---

**9.2 How many days are left for completing the sections?**

Section 1: Introduction	_____
Section 2: Background Research	_____
Section 2.1: Research 1	_____
Section 2.2: Research 2	_____
Section 2.3 Research 3	_____
Section 3: Conclusion	_____
Section 4: Future Work	_____

**9.3 How easy was to find this information? (10 extremely easy and 1 extremely difficult)?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

**9.4 Please rate the usefulness of this feature?**

1     2     3     4     5     6     7     8     9     10

Please explain the motivation behind the rating

---

---

---

---

---

## **Appendix B – Experiment 2 Materials: Semi-Structured Interviews**

Semi-structured interviews were given at each stage in the document development process. In the following sections we give examples of questions, although questions were tailored to the answers received by the group members in order to properly understand what information was required at each stage.

1. Drafting
2. Planning
3. Developing
4. Reviewing
5. Formatting

## Appendix B.1: Semi-Structured Interviews – Drafting Stage

1. What kind of information do you require for this stage?

a. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

b. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

c. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

d. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. What kinds of interaction are required in this stage?

a. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

b. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

c. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

d. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. What are you planning on doing next?

---

---

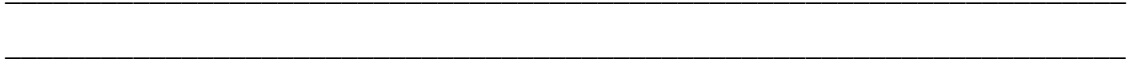
---

4. Any further comments/problems

---

---

---



## Appendix B.2: Semi-Structured Interviews – Planning Stage

1. How do you plan the document’s activity?

---



---



---



---

2. What kind (granularity) of information do you need (for example, deadline, roles, responsibilities etc.)?

---



---



---



---

3. What kind of information is required in order to coordinate for this stage?

a. \_\_\_\_\_

Rate this information

Extremely ineffective											Extremely Effective
1	2	3	4	5	6	7	8	9	10		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

b. \_\_\_\_\_

Rate this information

Extremely ineffective											Extremely Effective
1	2	3	4	5	6	7	8	9	10		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

c. \_\_\_\_\_

Rate this information

Extremely ineffective											Extremely Effective
1	2	3	4	5	6	7	8	9	10		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

d. \_\_\_\_\_

Rate this information

Extremely ineffective											Extremely Effective
1	2	3	4	5	6	7	8	9	10		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. What kinds of interaction are required in this stage?

a. \_\_\_\_\_

Rate this information

Extremely ineffective											Extremely Effective
1	2	3	4	5	6	7	8	9	10		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

b. \_\_\_\_\_

Rate this information

Extremely ineffective											Extremely Effective
1	2	3	4	5	6	7	8	9	10		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

c. \_\_\_\_\_

Rate this information

Extremely ineffective											Extremely Effective
1	2	3	4	5	6	7	8	9	10		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

d. \_\_\_\_\_

Rate this information



Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5. What are you planning on doing next?

---

---

---

5. Any further comments/problems

---

---

---

---

---

## Appendix B.3: Semi-Structured Interviews – Developing Stage

1. How are you writing the document (Word, CAWS, Wiki)?

---



---



---



---

2. How do you coordinate group efforts (email, chat etc.)?

---



---



---



---

3. What kind of information is required in order to coordinate for this stage?

a. \_\_\_\_\_

Rate this information

Extremely ineffective											Extremely Effective
1	2	3	4	5	6	7	8	9	10		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

b. \_\_\_\_\_

Rate this information

Extremely ineffective											Extremely Effective
1	2	3	4	5	6	7	8	9	10		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

c. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

d. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. What kinds of interaction are required in this stage?

a. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

b. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

c. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

d. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
--------------------------	--	--	--	--	--	--	--	--	--	------------------------

1	2	3	4	5	6	7	8	9	10
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5. What are you planning on doing next?

---

---

---

6. Any further comments/problems

---

---

---

---

---

## Appendix B.4: Semi-Structured Interviews – Reviewing Stage

1. How do you keep track of review activity?

---



---



---



---

2. What kind of information is required in order to coordinate for this stage?

a. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

b. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

c. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

d. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	

3. What kinds of interaction are required in this stage?

a. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
	1	2	3	4	5	6	7	8	9	10
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

b. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
	1	2	3	4	5	6	7	8	9	10
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

c. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
	1	2	3	4	5	6	7	8	9	10
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

d. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
	1	2	3	4	5	6	7	8	9	10
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. What are you planning on doing next?

---



---



---

7. Any further comments/problems

---

---

---

---

---

## Appendix B.5: Semi-Structured Interviews – Formatting Stage

1. How do you know when to format the document?

---

---

---

---

2. What kind of information is required in order to coordinate for this stage?

a. \_\_\_\_\_

Rate this information

Extremely ineffective											Extremely Effective
1	2	3	4	5	6	7	8	9	10		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

b. \_\_\_\_\_

Rate this information

Extremely ineffective											Extremely Effective
1	2	3	4	5	6	7	8	9	10		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

c. \_\_\_\_\_

Rate this information

Extremely ineffective											Extremely Effective
1	2	3	4	5	6	7	8	9	10		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

d. \_\_\_\_\_

Rate this information

Extremely											Extremely
-----------	--	--	--	--	--	--	--	--	--	--	-----------



ineffective										Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. What kinds of interaction are required in this stage?

a. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

b. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

c. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

d. \_\_\_\_\_

Rate this information

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. What are you planning on doing next?

---

---

---

8. Any further comments/problems

---

---

---

---

---

# Appendix C – Experiment 2 Materials: CAWS Groupware Evaluation User Questionnaire

## 1. Personal profile:

### 1. Personal Information

Name:

Surname:

Group Number:

### 2. What was your individual role in the group?

- a. Software Architect,
- b. Document Manager
- c. Strong Programmer
- d. Normal Programmer
- e. Weak programmer
- f. Strong Writers
- g. Normal Programmer
- h. Weak Writers
- i. Other (Please specify)

### 3. What activities did you perform? (Please try to be as specific and accurate as you can, including programming task and document headlines).

### 4. Did you finish everything that was assigned to you?

- a. Yes
- b. No
- c. Yes (but I needed help from \_\_\_\_\_, \_\_\_\_\_)
- d. No (\_\_\_\_\_ finished it for me)
- e. Please specify (explain the reason behind those situation)

5. Did you always meet the deadline?

<b>Always</b>	<b>Most of the times</b>	<b>Often</b>	<b>Rarely</b>	<b>Never</b>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please give an account to specify your answer (you might have missed deadline due to coursework, or the fact that your code was relying on some else to finish first, or other factors).

---



---



---

6. Rate your contribution to the project.

	Invaluable	Great Job	Normal	Average	Trivial
Programmer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Writer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please specify \_\_\_\_\_

7. Were all your relevant skills used within the project?

- a. Yes
- b. No

Please give a motivation \_\_\_\_\_

8. Rate the quality of your performance?

	Invaluable	Great Job	Normal	Average	Trivial
Programmer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Writer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please specify \_\_\_\_\_

### Group profile:

9. Categorise your group members according to these roles:

Name:

- a. Software Architect,
- b. Document Manager
- c. Expert Programmer (the person who has an overall of all the project components)
- d. Average Programmer (the person who has an understanding of the task at hand, but does not have a complete understanding of the overall of the project components)
- e. Novice programmer (the person who has basic programming skills and needs help in order to complete the task at hand)
- f. Expert Writer (the person who has an understand and can write the entire document without
- g. Average Writer
- h. Novice Writer
- i. Other Please specify \_\_\_\_\_

10. Did your group members finish everything that was assigned to them?

Name:

- a. Yes (but he/she needed help from \_\_\_\_\_)
- b. No (\_\_\_\_\_ finished it for \_\_\_\_\_, \_\_\_\_\_)
- c. Yes
- d. No

Please specify \_\_\_\_\_

Name

- e. Yes (but he/she needed help from \_\_\_\_\_)
- f. No (\_\_\_\_\_ finished it for \_\_\_\_\_, \_\_\_\_\_)
- g. Yes
- h. No

Please \_\_\_\_\_ specify

\_\_\_\_\_

Name

- i. Yes (but he/she needed help from \_\_\_\_\_)

j. No (\_\_\_\_\_ finished it for \_\_\_\_\_, \_\_\_\_\_)

k. Yes

l. No

Please specify \_\_\_\_\_

Name

m. Yes (but he/she needed help from \_\_\_\_\_)

n. No (\_\_\_\_\_ finished it for \_\_\_\_\_, \_\_\_\_\_)

o. Yes

p. No

Please specify \_\_\_\_\_

Name

q. Yes (but he/she needed help from \_\_\_\_\_)

r. No (\_\_\_\_\_ finished it for \_\_\_\_\_, \_\_\_\_\_)

s. Yes

t. No

Please specify \_\_\_\_\_

Name

u. Yes (but he/she needed help from \_\_\_\_\_)

v. No (\_\_\_\_\_ finished it for \_\_\_\_\_, \_\_\_\_\_)

w. Yes

x. No

Please specify \_\_\_\_\_

11. Did your group members meet the deadline?

Name

Always	Most of the times	Often	Rarely	Never	Please comment
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

Name

Always	Most of the times	Often	Rarely	Never	Please comment
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

Name

Always	Most of the times	Often	Rarely	Never	Please comment
--------	-------------------	-------	--------	-------	----------------

	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Name						
	Always	Most of the times	Often	Rarely	Never	Please comment
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

Name						
	Always	Most of the times	Often	Rarely	Never	Please comment
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

12. Rate the quality of your group members ' contribution to the project.

Name						
	Invaluable	Great Job	Normal	Average	Poor	Please comment
Programmer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Writer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

Name						
	Invaluable	Great Job	Normal	Average	Poor	Please comment
Programmer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Writer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

Name						
	Invaluable	Great Job	Normal	Average	Poor	Please comment
Programmer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Writer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

Name						
	Invaluable	Great Job	Normal	Average	Poor	Please comment
Programmer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Writer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

Name						
	Invaluable	Great Job	Normal	Average	Poor	Please comment
Programmer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Writer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

13. Where all their relevant skills used within the project?

Name						
	Always	Most of the times	Often	Rarely	Never	Please comment
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

Name						
	Always	Most of the times	Often	Rarely	Never	Please comment

	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Name						
	Always	Most of the times	Often	Rarely	Never	Please comment
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Name						
	Always	Most of the times	Often	Rarely	Never	Please comment
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Name						
	Always	Most of the times	Often	Rarely	Never	Please comment
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

14. Rate your group members quality of performance?

Name						
	Invaluable	Great Job	Normal	Average	Poor	Please comment
Programmer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Writer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Name						
	Invaluable	Great Job	Normal	Average	Poor	Please comment
Programmer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Writer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Name						
	Invaluable	Great Job	Normal	Average	Poor	Please comment
Programmer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Writer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Name						
	Invaluable	Great Job	Normal	Average	Poor	Please comment
Programmer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____
Writer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

**Rate the importance of users' awareness**

15. Importance to know your individual role?

Extremely

Extremely



ineffective										Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

16. Importance to know your individual responsibility?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

17. Importance to know other member role?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

18. Importance to know other members responsibilities?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

19. Importance of knowing which one is the final version of the document.

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

20. Importance of knowing who has change the document since you last view it?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

21. Importance of knowing what edit changes have been made to the document since you last viewed it?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

22. Importance of gaining an overview of the document recent activities (i.e. who changed the document, at what time, what changes have been made, what comments and so on.)?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

23. Importance of knowing who is working at the same time you are?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

24. Importance of knowing the details of how other members are currently engaged with the document at a specific time?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

25. Importance to know which parts are finished?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

26. Importance of knowing an overview of current work on the document?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

27. Importance of knowing what each members' contribution to the document?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

28. Importance of knowing what should be the users contribution to the document?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

29. Importance of monitoring task progress with respect to the task assigned?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

## Tools Used

30. What did you use to write the documents?

- a. Word
- b. Latex
- c. CAWS
- d. GoogleDocs
- e. Wiki (please specify which one \_\_\_\_\_)
- f. Other (please specify)

31. Did you always know which one was the final version of the document?

- i. Yes
- ii. No
- iii. Sometimes
- iv. Most of the time
- v. Rarely

Please Specify - How did you know which one was the final version of the document? (if you used a repository please explain how you avoided merging conflicts)

---



---

Rate this feature effectiveness is present in the tool you used.

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

32. By looking at the document after other members have been working on it, did you know what were the changes, added comments added that were made?

- i. Yes
- ii. No
- iii. Sometimes
- iv. Most of the time
- v. Rarely

Please Specify - How did you gain an overview of the document changes?

---



---

Rate this feature is present in the tool you used.

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

33. Did you have a way to review documents activities (changes, comments, restructuring) made to the document when you were not working on it? (For example using SVN you can add comments to the revisions)

- i. Yes
- ii. No
- iii. Sometimes

- iv. Most of the time
- v. Rarely

Please Specify - How did you gain the information related to what was going on when you were not working on the document?

---



---

Rate this feature is present in the tool you used.

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

34. Did you know who was working at the same time as you were?

- i. Yes
- ii. No
- iii. Sometimes
- iv. Most of the time
- v. Rarely

Please Specify - How did you coordinate to work at the same time and if coordination was not needed explain how did you know who is working at the same time you do?

---



---

Rate this feature is present in the tool you used.

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

35. Did you know the details on who and what has changed the document since you last viewed it?

- a. What were they working on?

- i. Yes
- ii. No
- iii. Sometimes
- iv. Most of the time
- v. Rarely

Please specify - How did you know on what they were working on?

---

Rate this feature is present in the tool you used.

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

b. How did they progress?

- i. Yes
- ii. No
- iii. Sometimes
- iv. Most of the time
- v. Rarely

Please specify - How did you know about their status in the process?

---



---

Rate this feature is present in the tool you used.

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

36. Did you know which parts were complete in the document?

- i. Yes

- ii. No
- iii. Sometimes
- iv. Most of the time
- v. Rarely

Please specify - How did you know which part are finished?

---



---

Rate this feature is present in the tool you used.

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

37. Did you have details regarding an overview of current work on the document?

- i. Yes
- ii. No
- iii. Sometimes
- iv. Most of the time
- v. Rarely

Please Specify - How did you know what is going on while you are working?

---



---

Rate this feature is present in the tool you used.

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

38. Did you know what are the individual contribution to the document?

- i. Yes

- ii. No
- iii. Sometimes
- iv. Most of the time
- v. Rarely

Please specify - How did you know what are the contribution to the document?

---



---

Rate this feature is present in the tool you used.

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

39. Did you know what should have been the users contribution to the document?

- i. Yes
- ii. No
- iii. Sometimes
- iv. Most of the time
- v. Rarely

Please specify - How did you know what should be the users contribution to the document?

---



---

Rate this feature is present in the tool you used.

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

40. Did you know when each member has been working on the document?

- i. Yes



- ii. No
- iii. Sometimes
- iv. Most of the time
- v. Rarely

Please specify - How did you know who was present in the workspace?

---

---

Rate this feature is present in the tool you used.

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

41. Rate the tool used?

Extremely ineffective										Extremely Effective
1	2	3	4	5	6	7	8	9	10	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

42. Did the tool improve your group writing process?

- a. Yes
- b. No
- c. Please specify your answer.

43. Please add any other comments you wish to provide about the group experience related to the tool used?

---

---

---

## Appendix D - Further features in CAWS

### PDF Output

One of the most important features of CAWS is its ability to export the document's contents in the form of a PDF file for publication. As previously noted, the ability to generate documents conforming to journal style guidelines is an important issue. In order to support this requirement, the back-end part of the system has been based on the Latex typesetting system.

Using Latex to typeset documents has several advantages. The majority of academic journals provide a Latex style sheet that can be used to apply the correct style settings to articles for submission. As a result, configuring a document for publication is usually a simple matter of uploading the Latex style file for the journal to which the article is to be submitted. For ease of use, the system has preconfigured style files for various journals that can be selected from a list.

The use of Latex does present some difficulties, however. Latex by default does not include support for the Unicode character set, necessitating the use of a Unicode add-on package for correct rendering of characters beyond the standard ASCII character set. Also, Latex as a system is designed to be written by a human; the generated Latex may not always be as optimally typeset as an experienced Latex user would be capable of creating.

In some rare situations it is necessary to be able to enter Latex code directly into the document for processing; for example, style files for certain journals may require the use of non-standard Latex commands in order to specify copyright or author information. CAWS provides a configuration page for these situations where custom Latex code can be entered to be inserted into the document header.

## Style Editor

It is not always possible to rely on the availability of a Latex style sheet; some conferences do not provide them. Similarly, a user may want to write an academic paper that they do not intend to submit to a conference. For this reason, CAWS includes a custom style editor that allows users to define their own custom document styles as an alternative to using a predefined style file.

The style editor works by automatically generating a style file for the document, based on settings entered by the user. The user can configure the page settings (margin, paper type, orientation) and text formatting settings (text alignment, font, style, line spacing, etc). The style editor shows a preview of how the resulting document will appear with the specified settings.

## Bibliography

A bibliography system is essential for a collaborative writing system for academic articles. CAWS includes a bibliography system that allows a library of citations to be created for use in the document being written. The bibliography editor allows the type of reference to be specified (eg. journal article, book, conference proceedings); appropriate fields for the type are then shown which can be completed by the user.

Once a reference has been entered into the bibliography, it can be cited from within the editor. The citation itself is tracked; if the details of the reference are amended within the bibliography, the citation is automatically updated within the text.

The citations are generated in the final document using Bibtex, the Tex-based system commonly used when writing Latex documents. The style editor allows the Bibtex citation style to be specified, as different journals require different citation styles.

## Security

In a conventional wiki, it is often desirable to impose as few security restrictions on access to the document as possible; this helps to encourage potential contributors to add information to the wiki by removing barriers to contribution. However, when writing an academic paper, it is desirable to be able to impose more stringent controls on who can contribute to the document.

CAWS includes fine-grained control of permissions. For example, it is possible to grant a user permission to comment on the document without granting permission to edit the document. This is useful for the peer review stage, for example, to stop reviewers from making their own changes to the document.

## **Automatic save**

An informal survey conducted previously as part of this research highlighted that academics were apprehensive about using a web-based system for writing their work, as they were concerned about losing their work in the event of network problems.

To address this concern, CAWS includes an automatic save mechanism that saves a draft copy of the current text to the server every ten seconds. An indicator notes that the text has been saved. If the connection to the server is interrupted, a warning message is displayed indicating that the text is not being saved. If the user closes the browser, or their computer crashes, the last saved draft copy is restored when the editor page is next accessed.

## **Revision history**

CAWS retains the standard revision history feature found in almost all wikis. Users can review a list of changes made to the document and a differencing system allows the specific textual changes to be highlighted for quick review.

As a CAWS document is based around multiple sections, the history function also allows the history for specific sections to be examined, along with the overall document history. The history system, in addition to tracking the document's contents, also tracks other changes to other properties of the document, including the style settings (for PDF output) and the bibliography.

## **Creating a PDF Document**

All data related to documents within CAWS is attached to a version. This includes the document text, along with style settings, citations, and document properties (such as the

title). In this way, it is possible to revert any changes and roll back to an older version of the document if necessary. All changes to the document are tracked along with a description of the change.

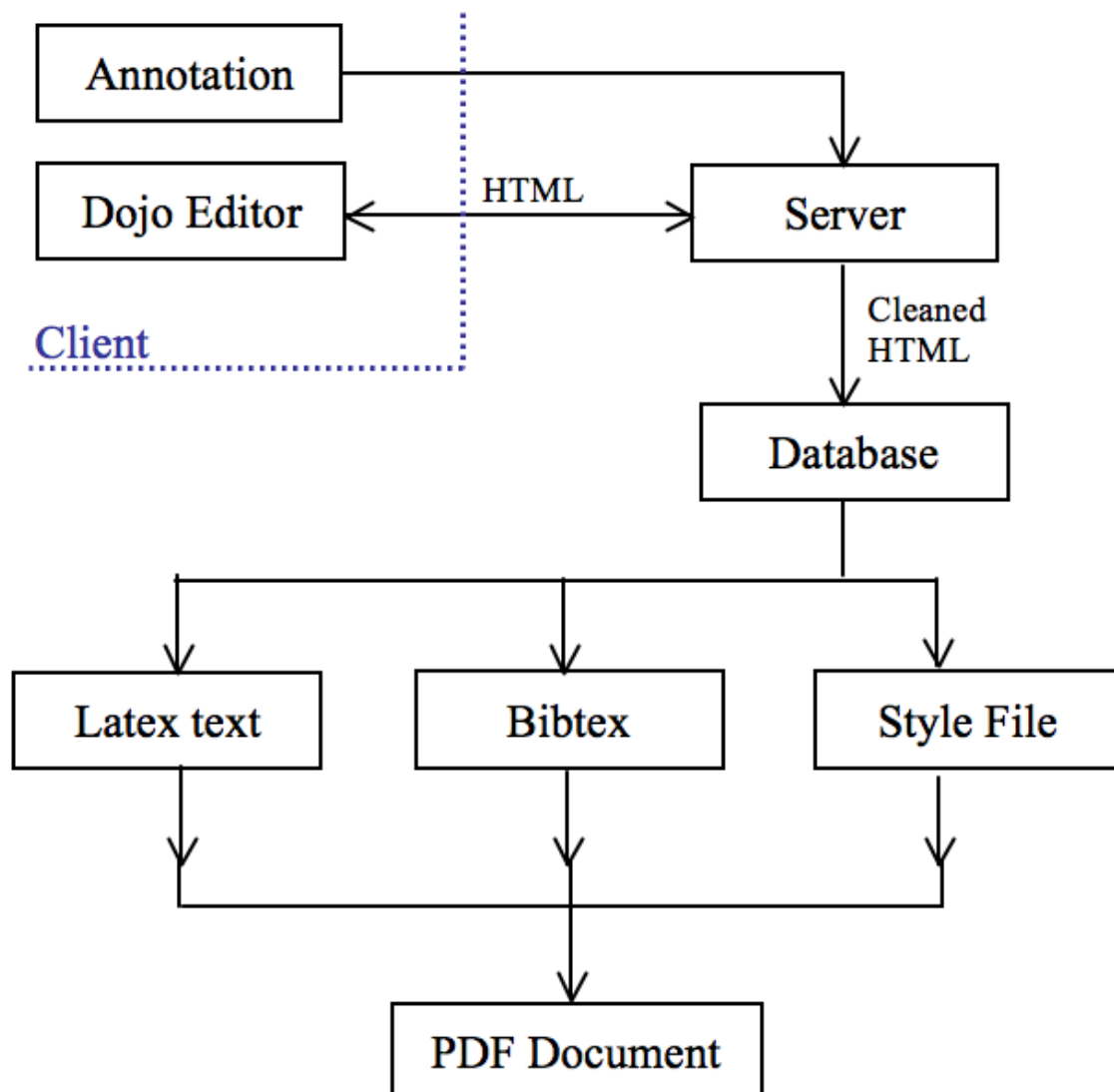
The text of each section of the document is stored independently, allowing the sections to be rearranged as desired. The structure of the document is stored separately from the document text, and, again, is also version-controlled. Certain sections can be tagged as special in purpose; this feature is used for sections that are automatically generated. Examples are the table of contents and bibliography.

Internally, the document text is stored in HTML format. The Dojo web toolkit is used for editing text, although this has been extended to allow insertion of images and citations within the text. Upon upload to the server, the HTML text is scrubbed to remove any unsupported HTML tags; this is particularly important, as extraneous and unsupported tags can be inserted into the text if it is copied and pasted from a website or another editor, such as Microsoft Word (Figure D8.1).

Annotations are stored through the use of start and end tags inserted in-line into the text. In this way, multiple annotations can overlap the same text. Javascript code on the client tracks which text nodes are assigned to which annotations and performs the highlighting and filtering functionality.

When generating a PDF file, the HTML text is translated into Latex format. This is complicated by the fact that Latex is designed to be crafted by hand, rather than machine-generated (Figure D.8.1). For example, Latex processing is deliberately run in a mode that is less fussy about line breaks, to prevent text overruns. In testing, URLs have been found to be a particular problem in causing overruns, so these are automatically detected and processed using the "URL" package. Widths of table columns must be automatically calculated, and the "longtable" package is used to allow tables to pass over page boundaries.

The list of citations for the document is similarly converted to a Bibtex format file, while a Latex style file is automatically generated from the user's defined style settings



**Figure D8.1:** CAWS Implementation Architecture

(if a predefined style file is uploaded, this is used). These are then combined to generate the final PDF version of the document (Figure D8.1).