University of Alberta

Visualizing Wiki Author Contributions in Higher Education

by

Cristina Arias Torres

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Arts
Humanities Computing

Fall 2009
Edmonton, Alberta

Examining Committee

Dr. Stan Ruecker, Supervisor
Humanities Computing Programme
Department of English and Film Studies
University of Alberta

Dr. Ofer Arazy, Supervisor
Department of Accounting and Management Information Systems (AMIS)
School of Business
University of Alberta

Dr. Geoffrey Rockwell
Humanities Computing Programme
Department of Philosophy
University of Alberta

Abstract:

This thesis investigates how users contribute to wiki environments implemented in higher education settings. The type, extent and quality of users' edits is assessed by analyzing a student-generated wiki through a manual analysis carried by three raters. This project aims to present the relative contribution of wiki users as a way to motivate them to collaborate in a wiki-based group project. Manual as well as automatic programs are used to analyze the same set of wiki articles. Automatic algorithms are proposed to estimate users' relative contribution and ownership to the article. Suggested is the implementation of a visual representation known as a glyph (working as an add-on to the wiki) that shows statistics on participants' contributions. Several glyph designs were tested by participants through interviews. These interviews showed that participants identified the visualization as a motivation enhancer to contribute to wikis.

Acknowledgements

It has been a great privilege to have worked with my supervisors. I would like to thank them for all their patience, encouragement and financial support:

Dr. Stan Ruecker, for his consistent support and guidance throughout the programme.

Dr. Ofer Arazy, for adopting me as part of his team when I most needed it.

Special thanks to:

Dr. Eleni Stroulia, for letting me collaborate in one of her multiple projects; one of the best team leaders I have ever met. Janey Kennedy, for all her hard work at the Office of Interdisciplinary Studies. Terry Butler, for helping me discover my love for educational technologies. Geoffrey Rockwell, for serving as a member of the examining committee and also for his valuable feedback. The research participants, for sharing their knowledge and experiences.

I also want to thank my family, for believing in me. “Gracias a mi madre por todo su esfuerzo y sacrificio para darme una educación. Gracias a mi padre por su apoyo incondicional y sus oraciones. Les agradezco mis hermanas, mi madrina Luz y Polo que siempre me alentaron a seguir estudiando y por su buen sentido del humor.”

Finally, I would like to express my sincere gratitude to Garson Law for editing the first draft of this thesis.

Table of contents

Chapter 1: Introduction 1

1.1 Purpose 1

1.1.1 Research Questions 2

1.2 Contributions to current knowledge 4

Chapter 2: Literature Review 7

2.1 Introduction 7

2.2 Theoretical Framework 8

2.2.1 Constructivism and Social Constructivism 8

2.2.2 Connectivism 9

2.2.3 Social Exchange theory 11

2.3 Wiki environments in Higher Education 13

2.4 Automatic author-contribution calculations 14

2.5 Wiki visualizations as motivational approach 18

Chapter 3: Method 23

3.1 Introduction 23

3.2 Analysis of the wiki articles 24

3.2.1 Manual Analysis procedures 25

3.2.2 Description of the automated analysis of the wiki articles 28

3.3 Visualizations' interviews: usability and motivational tests 31

3.3.1 Interview procedures and protocol 33

3.3.2 Presentation of visualizations 36
3.3.3 Qualitative analysis of the interview data 41
Chapter 4: Findings 45
4.1 Part I: Analysis of wiki contributions 45
4.1.1 Introduction 45
4.1.2 Presentation of Results 46
4.1.3 Discussion 50
4.1.3.1 Inter-rater agreement and intra-class coefficient 50
4.1.3.2 Correlations manual analysis 52
4.1.3.3 Correlations Automatic vs. Manual metrics 55
4.2 Part II: Wiki users' Interviews 58
4.2.1 Introduction 58
4.2.1.1 Description of participants and experiences with wiki environments 59
4.2.2 Usability Test 62
4.2.2.1 "Sunword" usability 64
4.2.2.2 "Pie chart" usability 66
4.2.2.3 "CircleMagic" usability 69
4.2.2.4 "All wiki pages" usability 72
4.2.2.5 Discussion 75
4.2.3 Motivation Test 78
4.2.3.1 Participant perceptions on the motivational factors of visualizations 80

4.2.3.2 Ranking and competition as motivational determinant 85

4.2.3.3 Considerations on anonymity 89

4.2.3.4 Discussion 90

Chapter 5: Conclusions 92

References 99

Appendix A: Ethics certificate 104

Appendix B: Information/Consent Letter 105

Appendix C: Confidentiality Agreement 108

Appendix D: The Visualization Glyphs 109

Appendix E: Sample Recruitment Advertisement 113

Appendix G: Data collection 120

Appendix H: Example of control table for manual analysis (template) 122

Appendix I: Summary of the Method for the Manual analysis experiment 123

Appendix J: Interview Questionnaires and Scenarios 124

Appendix K: Participants' Experiences with Wiki environments 126

List of Tables

Table 1.Glyphs characteristics used to codify usability test. 42

Table 2. Inter-rater Reliability results on intraclass correlation coefficient. 47

Table 3. Correlations in Manual Analysis 48

Table 4. Correlations, manual and automated calculations..... 49

Table 5. Demographics 60

Table 6. Frequency of responses to "Sunword" glyph's usability 65

Table 7. Frequency of responses to "Pie chart" glyph's usability 68

Table 8. Frequency of responses to "CircleMagic" glyph's usability..... 71

Table 9. Frequency of responses to "All wiki pages" glyph's usability..... 74

Table 10. Comparison: percentages of responses to all glyph's usability 77

List of Figures

Figure 1. Glyph – "Sunword"..... 37

Figure 2. Glyph – "CircleMagic" 38

Figure 3. Glyph – "Pie chart"..... 39

Figure 4. Glyph – "All Wikipedia pages"..... 40

Figure 5. Response frequency to "Sunword" glyph's usability 66

Figure 6. Response frequency to "Pie chart" glyph's usability 69

Figure 7.Response frequency to "CircleMagic" glyph's usability..... 72

Figure 8. Response frequency to "All wiki pages" glyph's usability..... 75

Figure 9. Favorite visualization – Aesthetically appealing..... 78

Chapter 1: Introduction

1.1 Purpose:

Wikis are online collaborative systems that facilitate production of content and cooperation among authors. In higher education this web-based application supports the management of construction of shared knowledge. The purpose of this project was to study the ways in which wiki users, in a higher education context, write articles collaboratively and how we can reward their effort. More specifically, this study addresses a weighting of the relative contribution of each user, while presenting visual feedback on their performance. Another objective is to compare the statistical results from the analysis of each participant's collaboration to a student-generated wiki. These results are derived from two experiments: manual and automated analysis, which unveil some of the basic mechanics by which wiki environments are ruled. Proposed is the implementation of a visualization that accurately shows the nature and percentage of wiki contributions, as well as giving participants credit for their effort in order to motivate them.

Wiki articles, from an educational wiki, were manually analyzed. In this analysis the kinds of users' contributions were rated. These articles were part of a higher education wiki project/assignment implemented by the School of Business at the University of Alberta. This wiki site was integrated as part of the Management Information Systems 311 (Fall 2008) course into the Blackboard Academic Suite. The same sample of articles was also analyzed using automated analysis algorithms that were designed to catch relative contributions and ownership percentage. The statistical results derived from the manual analysis were expected to calibrate the automatic algorithms. These automated results are proposed as the source for rating participant contribution; the same results would be shown in the visualizations (also called glyphs). One of the main purposes was to test these visualization tools on people with experience collaborating in wikis in an educational context. The objective was to measure the chances of a possible implementation in the real world.

1.1.1 Research Questions:

The following are the research questions that were addressed in this thesis:

1. *What weight do human assessors assign to each contribution category in giving an overall contribution score?*

1.1 *How do the automatically calculated metrics proposed by Arazy & Stroulia (2009) correspond to the manually-assessed edit categories scores? Is there a difference between an educational setting and the results reported for the Wikipedia by Arazy & Stroulia (2009)?*

2. *To what extent can users understand the information visualized in the glyph proposed by Ruecker et al., (2008)?*

2.1 *Are there specific variations of the visualizations that users can more easily understand?*

2.2 *What level of detail do users require?*

3. *How would visualizations of the attribution metrics impact users' attitudes towards contributing to the wiki? What is the expected impact of the visualization in users' wiki behavior?*

3.1 *Are there differences between the diverse variations of the visualizations?*

The significance of this project lies in the fairness of showing participants of collaborative writing the impact of their contribution. Particular strengths

will be revealed when identifying each type of editorial contribution. As in all communities, the constant participation of the members maintains the structure of a social network. The variety of roles that everyone inside the network plays is very important, even if it is the smallest contribution. In the present study, wiki articles are placed in an educational context; that is why student-generated articles were chosen for our sample. The activity in educational wikis has boundaries because the final product could either represent part of a student's grade or a future research publication. Wiki technologies implemented as instructional technology tools are not expected to have some of the chaotic elements that are common in Wikipedia articles, such as editorial wars and vandalism. This study can be considered relevant by educational technologists looking into better practices and implementations of wiki environments into the classroom. It also offers the reader a review of new visualization tools that can be used as an add-on in wiki technologies as proposed by Ruecker et al., (2008).

1.2 Contributions to current knowledge:

The current research builds over two partner projects, carried out by researchers at the University of Alberta. The first project developed in the department of Computing Science, researches wiki attribution and relative contributions of wiki authors. This wiki attribution project works on the development and implementation of the algorithms we used in this thesis,

thanks to the collaboration of Tim Yau and Veselin Ganev by developing and implementing the automatic algorithms. This thesis' experiments were focused in educational wikis instead public wikis. Arazy and Stroulia's (2009) manual analysis performed on Wikipedia articles was complemented by rating not only extent of users' contributions but also by analyzing the quality of these contributions. For this manual analysis we also had access to the students' final grade on the assignment, giving us another dimension on the perception of relative contribution. The results of our manual analysis of wiki users' contributions was correlated to the algorithm results in order to determine the validity of the programs' results.

The second project is a partner project developed in the Humanities Computing program at the University of Alberta. It consists of the design and implementation of visualizations of relative wiki contributions (designed by Carlos Fiorentino). In this thesis a qualitative experiment was conducted to test the clarity of the visualizations' design and the effectiveness of the proposed wiki visualization concept in a higher education context. Some of the contributions of this thesis are:

- A well-designed experiment to rate objectively the relative importance of each edit category in a higher education context, while taking into consideration the 5 edit categories used in Arazy

and Stroulia's research (2009) as well as measuring the extent and quality of contributions.

- Statistic analysis for identifying correlations between the algorithms and the findings of the manual experiment.
- Determining user perspectives on wiki visualizations proposed by Ruecker et al., (2008) and study motivational issues towards the implementation of a visual graph aimed to stimulate participation of wiki contributors in higher education.

Chapter 2: Literature Review

2.1 Introduction:

Wiki environments have been adopted by educational institutions as an effective instructional technology. Their adaptability in many educational activities that involve sharing written content over the internet makes wiki environments a powerful tool for active knowledge transfer. Wiki technologies as part of Web 2.0 enable users to take an active role inside a community or a group with shared interests. Wikis as social software promote collaborative writing to a level that no other online tool has done in the past. The Wikipedia can be considered one of the most popular implementations of wiki technologies.

First, the theoretical framework that supports the educational value of wiki technologies is described. Learning theories such as: connectivism, constructivism, social constructivism, are discussed. A secondary supporting theory, the exchange theory in active learning, is expected to help understand why wiki users consider this a valuable instructional

technology. These theories also point to some of the motivational questions when writing collaboratively. Second, the concept of educational wikis implemented as instructional technologies is revised. Third, computational tools that are being used for ownership attribution are also reviewed. Finally, literature on the usage of wiki visualizations and their motivational effects is studied.

2.2 Theoretical Framework:

In this theoretical framework we consider learning theories to be the groundwork for the usage of wikis as instructional technologies. Three main paradigms were included: constructivism & social constructivism, connectivism and social exchange theory. An understanding of these theories aides in justifying the importance of educational wikis in higher education.

2.2.1 Constructivism and Social Constructivism:

Constructivism states that learners create knowledge as they gain experience (Parker & Chao, 2007) in which past knowledge is merged with new experiences. One can also say that “learners often select and pursue their own learning” (Siemens, 2004) not only when they are part of

an educational institution but through their whole lives. Wiki environments give users the freedom to become more involved in topics in which they have a deep interest.

Constructivism is one of the theories that focus on knowledge that is encouraged by engaging students in activities to reaffirm and continue the process of learning. Wiki software engages students in active discussions among authors, and its asynchronous nature enables students to process more accurate arguments.

Social constructivism is a popular theory that addressed the importance of educational social software; it states that learning happens around “social activities” where people enrich their own understanding with others’ knowledge and experiences (Parker & Chao, 2007, p. 59). Experiences and interaction within a social environment can have a great impact on learning. Knowledge can be effectively promoted in wiki environments where the exchange of ideas and information among people allow them to engage in reflective learning.

2.2.2 Connectivism:

Connectivism is a theory that talks about the learning that occurs outside one’s own self; more specifically: in the group’s consciousness. Siemens

argued that some “learning theories (such as cognitivism) are concerned with the actual process of learning, not with the value of what is being learned” (Siemens, 2004). In wiki environments, the value, accuracy and currency of information are constantly reviewed by the users that generated the content in the first place The next list is comprised of key points that will help to understand the importance of connectivist theories to explain learning facilitation in wikis. This passage was extracted from Siemens (2004) when describing the principles of connectivism:

- Learning and knowledge rests in diversity of opinions.
- Learning is a process of connecting specialized nodes or information sources.
- Learning may reside in non-human appliances.
- Capacity to know more is more critical than what is currently known
- Nurturing and maintaining connections is needed to facilitate continual learning.
- Ability to see connections between fields, ideas, and concepts is a core skill.
- Currency (accurate, up-to-date knowledge) is the intent of all connectivist learning activities.

As we can see the implementation of innovative Web 2.0 technologies in education are supported by theories that stimulate the practical use of knowledge as well as enriching learning activities inside social networks.

2.2.3 Social exchange theory:

One of the difficulties in designing social spaces on the World Wide Web is to maintain active participation of people that create content. In this project one of the research questions addressed motivational issues when working in a collaborative writing space. Social exchange theory deals with reasons of motivation and collaboration of people that contribute towards a similar learning objective. Social exchange theory proposes that "all human relationships are formed by users through subjective cost-benefit analyses and comparison of alternatives". (Moore, 2007)

Klamma et al., have listed the ways in which social exchange theory activates participation in networks:

- "Personal access, or anticipated reciprocity": learner has a pre-existing expectation that he will receive actionable and useful (extra) information in return;
- "Personal reputation": learner feels he can improve his visibility and influence to others in the network, e.g. leading to more work or status in the future;
- "Social altruism": learner perceives the efficacy of the LN (learning network) in sharing knowledge as a 'public good', especially when contributions are seen as important, relevant, and related to outcomes;

- “Tangible rewards”: learners negotiate to get some kind of more tangible asset (financial reward, bond, book, etc.) in return. (2007, p. 74)

Social exchange theory can be considered as a platform for the movement of resources and information within a network through social processes. Users of collaborative technologies evaluate possible risks and benefits they are exposed to by using these services; “when a web application’s cost and risk is higher than its value, the user will abandon that relationship” (Moore, 2007). This means that if the cost (e. g. effort, time, work) of investment put into social software doesn’t pay off with some kind of reward (e.g. money, community acceptance feeling, effort acknowledgment, etc); users will not feel motivated to collaborate.

According to the social exchange theories users may feel discouraged from participating if their contribution is not recognized by other members of the group. In the educational setting this peer recognition is important if they want to build a reputation; it helps in balancing the effort-reward perception. In educational social software one of the reasons that keep users actively engaged is the collaboration with others. Educational wikis used as a writing medium for group authoring enables users “to build and edit the document on a single, central wiki page”, making the process of writing collaboratively simpler (Parker & Chao, 2007, p. 61).

2.3 Wiki environments in higher education:

Traditional wikis such as the Wikipedia are open mediums where “every one of its articles can be edited by anyone, nobody knows users’ levels of expertise on the topic and the information is available to everyone” (Priedhorsky et al., 2007, p. 1-2). In this project we take higher education as our focus of implementation. Wikis are being used in many educational activities, such “writing assignments, group projects, and online/distance education” (Parker & Chao, 2007, p. 60), post-classroom discussion groups, team project resources management tools, and one of the most important, collaborative writing environments in formal academic writing.

These educational settings represent a much more controlled environment. Educational wikis are not public to the internet; the number of participants is usually not as large as on an open wiki. There are fewer possible threats to the information or to the system because participants know they are under their university’s code of behavior. Mitchell states that it is the “collaborative and creative nature of the Wiki that holds the real power for education” (2006, pp. 119-143). Wiki technologies, as part of the Web 2.0, inherited the dynamism that enables people to contribute with web content. Inside the pedagogical context, issues like intellectual property and ethical behavior have to be seriously addressed (Mitchell, 2006).

2.4 Automatic author-contribution calculations:

This research uses automatic algorithms to calculate authors' relative contribution to wiki articles. In order to investigate particular patterns of Wiki text, it is useful to employ computational algorithms. Wiki technologies store overwhelming amounts of information. For humans to perform analysis without help of such software would be almost impossible, due to the size of the sample that is needed to make meaningful interpretations. Automatic algorithms that deal with attribution of wiki contributions attempt to mimic peer-reviewing activities. Some algorithms perform simple counts of edits as contributor's score. While this approach can be sufficient in most cases, it lacks of the qualitative information needed for other levels of user performance such as ownership.

Hess, Kerr & Rickards developed a computer program that calculates the "similarity between all pairs of consecutive edits and builds a new results table containing the user who made the change, a timestamp, and the similarity between the present and previous versions of the page" (2006, p. 1). In this way they tried to track individual's behaviors when they contribute to a page.

Wiki collaborators have the freedom of choosing the tasks they take on. In an educational context the majority of wiki users are not professional editors but it is probable that they have experience writing and editing articles in collaboration with their peers. These tasks are done in a more intuitive way, converging with their own editing skills. In Hess et als' research "wiki user statistics are generated by making inferences about user activity based on relevant information about edits (similarity, longevity, and authorship)" (Hess et al., 2006, p. 1-2).

A model that uses algorithms to measure author contributions was also developed by Adler et al. (2008, p. 2). Their research was based in two approaches: the first one, KLOC (thousands of lines of code in software measurement) measures contributions of programmers in software development, "counting how many lines of code are written per week". The second approach they take into consideration is that the "quality of an article improves with the number of edits and the number of distinct authors that revise the article" (Wilkinson & Huberman, 2007 as cited in Adler et al., 2008, p. 1-2).

Adler et al. presented three measures of author productivity: "edit longevity, 'text longevity' and 'text longevity with penalty'" (2008, p. 4). Edit longevity, represents both the amount of change in the edit itself and the time it lasts in the article (measuring quality). The amount of change is

calculated comparing the article with its previous version. The edit longevity is then calculated with the multiplication of quantity and quality metrics of contributions.

“Text longevity” and “text longevity” with penalty are calculated “tracking text through revisions” (Adler et al., 2008, p. 2). The difference between them is that the first one doesn’t give proper credit to small edits or maintenance tasks (as “edit longevity”); the second has the advantage of ‘punishing’ low quality contributions or vandalism.

Arazy and Stroulia’s (2009) studies on wiki attribution present a comprehensive identification of the “nature of contributions” made by wiki users which lead to a categorization of edits. Contribution metrics were defined by Arazy and Stroulia (2009) as follows:

- *Additions*: Integration of one or more ideas to the article.
- *Deletion*: Erasing of one or more ideas from the article.
- *Proofreading*: Cleaning up any typographical, grammatical or spelling errors.
- *Addition of links*: Addition of references in hyperlink format to support the content of the article, this includes Internal and external references.

- *Structural Changes*: Rearranging of text to ensure fluency and transition of the article. (p. 2-3)

The same categorization was used in this thesis' manual experiments.

Once these metrics where defined, the approach for the development of the algorithm was based on the concept of "sentence ownership"; including the weight of the types of edits made by authors.

"Sentence ownership" program "identifies which of the sentences of the earlier release have remained essentially the same in the new release" and in subsequent revisions (Arazy & Stroulia, 2009, p. 3). One may notice that the structure of the "text longevity" algorithm proposed in Adler et al., (2008) and the "sentence ownership" algorithm developed in Arazy and Stroulia's studies is based in a similar concept. However, as said before, Adler's program was based in software management metrics (KLOC) while Arazy and Stroulia's uses the "Munkres method to estimate how much of the sentences' position has changed" over newer versions (Arazy & Stroulia, 2009, p. 3). Arazy's and Stroulia's experiments with the Wikipedia articles showed satisfactory results in the evaluation of the algorithms. Thus, these algorithms are expected to be appropriate in evaluating educational wikis.

2.5 Wiki visualizations as motivational approach:

In this section, examples of the kind of visualizations that have been implemented in wiki environments are presented. Usually these visualizations aim to process the large amount of data linked with the inherent collaborative processes in wikis. The first two cases presented are examples of visualizations of contribution patterns at the article level. The last example proposes a visual approach of users' relative contribution (Ruecker et al., 2008); upon which part of this thesis is based.

A popular visualization that has been used on wiki data is "history flows", defined by Wattenberg et al. , as a "tool for visualizing how collaborative documents evolve over time" (2004, p. 582). In their article, Wattenberg et al., conclude that "the efficacy of history flow in highlighting patterns of behavior suggests that visualization is a technique well-suited to records of social behavior" (2004, p. 582). Some of these behaviors include: vandalism and repair, negotiation and authorship.

Wattenberg et al., (2006) developed a second visualization called 'chronograms'. This new visualization technique shows the "patterns in tasks carried by the Wikipedia administrators" (Wattenberg et al., 2006, p. 1, 3). This visualization helps to see where these contributors invest more work, it "displays sequences of words and phrases by mapping text to

color using an alphabetical code" (Wattenberg et al., 2006, p. 5). These two examples are successful cases of visualizations that reveal certain patterns in wiki activity condensed from an article's history pages. These visualizations present the nature of the activity and effort (positive or negative) of a wiki community working together.

For this project I was interested in finding out particular trends in an individual's contribution instead of communal contribution. For this reason a third visualization approach, as suggested by Ruecker et al., (2008) was adopted. Ruecker et al., designed a "set of information glyphs that read the results from the automated analysis (as developed in Arazy & Stroulia, 2009) of relative contributions and display them in visual form" (2008, p. 1). In his article, Ruecker et al., (2008) exposed two visualizations; "Sunword" and "CircleMagic". "Sunword" is described as circular tag-cloud that contains the group's information, where the size of the (font) author's name represents ranking. "CircleMagic" (resembling a colorful pie chart), holds more detailed information on the kinds of contributions users make. This last glyph breaks down user contributions according the metrics defined by Arazy and Stroulia (2008) (number of additions, deletions, structural changes, proofreading changes and links).

Both of these visualizations ("Sunword" and "CircleMagic") were taken for further study in this thesis. It was also decided to include two other

variations of these glyphs; “Pie chart” (Individual version of “CircleMagic”) and “All wiki pages”. The “Pie chart” glyph is a new design based in the “CircleMagic” glyph; the difference resides in the isolation of one single user’s contributions at a time. The second new design, “All wiki pages”, shows a user’s contributions compared to the whole wiki. It was desirable to put these visualizations to the test and prove how beneficial it would be for students and researchers as a motivational tool. These visualizations are also expected to support group management and knowledge acquisition; not only providing wiki users with simple statistical information on the progress of the team but with a tool that would keep contributors motivated during collaborative writing. For a better understanding of these motivational challenges some relevant literature that will be presented in the next section was reviewed.

Social theories suggest that “showing users, different perspectives on the value they add to the community will lead to differing amounts of contribution” (Rashid et al., 2006, p. 955). Differences in the quality of contributions would be expected as well. Ling (2005) recommends that in order to avoid social ‘loafing’, participants have to feel that their contributions are important, identifiable and link them to the community with which they are working.

In a survey conducted by Majchrzak et al., (2006) of corporate wikis, it was discovered that some causes of motivation of the authors to

contribute depended of the role they took in the writing/editing process. They present two kinds of contributors: Synthesizers and Adders. The study concluded that “Synthesizers are more interested in impact (impact to the organization, to other wiki users, reputation), while Adders are more interested in accomplishing their immediate work responsibilities” (Majchrzak et al., 2006, p. 103-104).

Studies on wiki systems made by Hoisl et al., (2006) show us the use of social reward as a motivational method. The term “social reward refers to something that causes a behavior to increase in intensity; usually by the most active members” in a community (Hoist et al., 2006, p. 3).

The implementation of the glyphs proposed by Ruecker et al., (2008) and the two variations described in this thesis are mostly described as feedback tools and social reward. The capability to “evaluate, monitor, measure, and apply reinforcement is referred to by Kluger and DeNisi (1996) as feedback” (as cited in Tedjamulia et al., 2005, p. 6-7). In the same article Tedjamulia also describes ‘feedback’ as the “process of measuring a person’s performance, comparing it to a standard, and reinforcing action through incentives over a period of time” (2005, p. 7).

This approach is presented in the model developed by Tedjamulia et al., as “extrinsic incentives that can modify a person’s performance and

contribution” (2005, p. 5-6). The theories presented here provide justification for later research in the implementation of visualization tools that support wiki users learning, leading to better understanding of motivational issues in educational wikis; as well as higher collaboration with the implementation of wiki visualizations.

Chapter 3: Method

3.1 Introduction:

In this thesis project the main research questions and their respective sub questions, mentioned in the introduction chapter, were addressed. All experiments were conducted at the University of Alberta and only participants from this institution were asked to be part of the project. The project involves two main experiments:

- The first one consisted of a detailed analysis of a set of student generated wiki articles. These articles were analyzed manually and automatically in order to acquire data about the relative contribution of each participant to an article. Statistical correlations were carried on for further analysis.
- The second experiment studied the collection of data through formal interviews. The second experiment was designed to explore the efficacy of relative contributions' visualizations and the effects on users' motivation. The interviews took into consideration past

experiences of participants in wiki environments and their reactions to the introduction of new visualization designs.

3.2 Analysis of the wiki articles:

The manual analysis of student-generated wiki articles is described as the first stage of the project. The experiment addresses the first research question of how human raters identify the quantity and quality of each edit category; and the sub-question of how these manual ratings correlat to the automatic ones. The data sample was extracted from an instructional Wiki application called Team LX, developed by Learning Objects Inc. (Washington, D.C.) and hosted in the Blackboard Academic Suite which is a course management system that allows users to post online course content.

The data set consisted of Wiki articles generated by a group of nearly 240 students belonging to three different sections of the same class. Approximately 60 articles on various Management Information Systems' topics were generated by the students during the assignment. Around 15 students per article were assigned to research and write these wiki documents. The topics for the articles were randomly assigned to students. The assignments were graded by peers based on the quality of their contributions to the article. At the end of the assignment, each

student had to grade 3 articles (other than their own) and every collaborator who was involved in the making of the paper. The process of grading was divided into two sections: grading the overall quality of the final product; and grading individual contributors.

The raters graded their peers based on the quality and extent of their contributions. The final grade for each student in each article was obtained by the average of the aggregated scores given by raters. Ethics Approval (see Appendix A: Ethics certificate) was obtained for work with data involving human subjects. All personal information was removed. Once anonymised, the original files that contained students' personal information were destroyed. The data collection was not intrusive into students' work; it was taken after the students submitted the final version of their assignment.

3.2.1 Manual Analysis procedures:

In the manual evaluation, versions of an article were studied to report students' types of contributions. The analysis of the sample was performed by comparing consecutive historical Wiki versions, starting with the latest version of the article and ending with the earliest version, discarding any versions saved after the assignment's deadline,. The approach for the rating was to register the types of contributions made

from the latest saved version to the proceeding version (known as the current version) and so on, until the newest version was reached. Contributions were evaluated as follows: classifying the type of contribution in one of the 5 edit categories proposed by Arazy & Stroulia (2009): addition, deletion, structural changes, proofreading and links; rating from 1 (lowest) to 5 (highest) their a) quality and b) extent of contributions. Each version of the history page contained zero or more changes made by a user. In the case of zero changes, a rating of zero across all categories was given. This analysis had two types of evaluation:

- Determination of the extent of every edit made by a user in every version of the article.
- Determination of the quality of every edit made by a user in every version of the article.

Three Research Assistants (RAs) were included in the rating process of the project to achieve a higher level of reliability on the results. The RAs were involved in a training period in which misunderstandings of the procedures were clarified. During this training period RAs would reach a consensus in the rating procedures as well as estimate the time the actual analysis would take. A categorization document (see Appendix F: Categorization procedures of manual analysis) was created to simplify the

process of agreement among the Research Assistants. This document presents:

- A clear description and delimitation of edits, followed by,
- A guide on how to rate quality and quantity of each edit.
- A list of borderline cases.

The rating process relied almost entirely on this document and on the constant communication among the assistants. For training purposes, three random Wikipedia articles were selected. These articles were chosen for their high activity in terms of users' participation. These articles also represented a challenge for their ambiguous categorization of edits. After every training exercise, the research assistants' ratings were compared and the exercise repeated until consensus was achieved. The Inter-rater reliability was calculated after the analysis; this measure compares the research assistants' scores and shows the level of agreement among them (Stoner et. al, 2005). Three randomized lists of the same 60 articles were prepared for each RA to follow; this step was expected to reduce the consequences of 'stimulus order' and attempted to neutralize the subjectivity associated with preferences of our human raters and the order of the articles they would chose to rate (Eisenberg & Barry, 1988) .

After the manual analysis was carried out, the next step was to normalize the results. For instance, the rating of the extent of an edit was graded from zero; being no edit found, to five; representing a considerably big edit (no negative ranking was given to this classification). However, the data coming from the qualitative rating given to edits was negative and positive, depending on the quality perceived by the raters. Ratings of one or two on quality represented negative contributions; for example: addition of inaccurate information, deletion of relevant text, addition of broken links, vandalism, etc. For these reasons statistical normalization of the data was essential. (Appendix I contains a summary of the method of the manual experiment).

3.2.2 Description of the automated analysis of the wiki articles:

In this section a brief explanation of the automated analysis of the wiki articles' sample will be given. The procedure of this analysis was not part of this thesis' experiments. However, the automated analysis of the wiki was an important step after the manual analysis; the main objective was to compare the two results and see how they correlate. For the clarification of the project, it was considered crucial to describe the way the algorithms worked. Programs developed in the Computing Science department at the University of Alberta for a partner project on measuring author contributions to the Wikipedia (Arazy & Stroulia, 2009), were used. The

importation of the Blackboard-hosted wiki to Media-wiki allowed us to solve any compatibility problems, give more controlled access to the copied data to the RAs and finally, prepare the data for the automatic analysis. The features to import into Media-wiki for the manual analysis were:

- The content of the articles themselves.
- History pages which hold all the log-in information such as date, time of changes and all versions of the article.
- Anonymised user IDs.

To perform this analysis, the data was collected from the history pages on the wiki; these pages track users' activity. Most of the procedures were followed as described by Arazy & Stroulia (2009) in their experiment with Wikipedia articles.

The automatic algorithms were run over the same set of articles analyzed manually. One focus of this stage of the project was to find out the correlation between the automatically calculated metrics with the manually assessed results; and to see how well the algorithms represented human ratings. In the next paragraphs, we will present roughly how the computer programs performed the analysis. Automated algorithms calculated user-based statistics for each Wiki page, as follows:

Simple counts:

- Number of edits, current and total attribution.
- Number of links, current and total attribution.
- Number of proofreading changes, current and total attribution.
- Number of sentences deleted / number of releases with deletion.

Sentence ownership:

- Number of sentences owned in current Wiki version.
- Number of the total sentences owned by the user in total attribution.

The statistics generated by the programs were divided into current and total results. A “current edit count” computes the changes that remain in the final version of the article. Meanwhile “total edit count” takes into consideration the sum of all edits done by a user, even though some of these edits no longer exist, or do not belong to the same user anymore.

Overview of the two algorithms:

- The first algorithm calculates the total amount of work; it counts the number of contributions that a user has collaborated with, then it accommodates them within one of the 5 edit categorization as

explained before. The objective of this program is to have a cumulative score of individual contributions, even when they no longer exist in the current version of the article.

- Roughly, the ownership algorithm extracts the information from the wiki database and analyses sentences as fragments; the rationale for this metric as is set in Arazy and Stroulia's article considered, that a sentence is the smallest unit of language that has a semantic meaning or is a thought (2009). Therefore, the inclusion of ownership is better distinguished by sentences instead of individual word-level editing. Then, the algorithm identifies "which of the sentences are essentially the same from consecutive releases" and assigns ownership percentages to the users that worked on the sentence (Arazy & Stroulia, 2009, p. 3). After comparing these article releases, the algorithm provides a detailed tree-like measure output where changes between versions can be identified.

3.3 Visualizations' interviews: usability and motivational tests:

The second stage of the thesis consisted in the testing of the design of four wiki visualization models. These models' principal objectives were

drawn by the second and third research questions. The second research question directs us to examine the level of understanding users require, and overall understanding of the visualizations' design:

> ***2.*** *To what extent can users understand the information visualized in the glyph proposed by Ruecker et al., (2008)?*
>
> ***2.1*** *Are there specific variations of the visualizations that users can more easily understand?*
>
> ***2.2*** *What level of detail do users require?*

The third research question deals with the motivational factors that would encourage (or discourage) users to contribute if the proposed visualization is implemented in a wiki environment:

> ***3.*** *How would visualizations of the attribution metrics impact users' attitudes towards contributing to the wiki? What is the expected impact of the visualization in users' wiki behavior?*
>
> ***3.1*** *Are there differences between the diverse variations of the visualizations?*

To address these questions, we decided that in-depth interviews were the best way to gather the information from participants' experiences. These interviews allow the interviewer as well as the subject to redirect the interview if an important topic arises. At the same time the structure of the interview is controlled with questionnaires that ensure the uniformity of the information collected across participants. Principally, the interviews were designed to assess the design of the glyphs, their suitability as instructional tools, as well as their impact on motivation.

3.3.1 Interview procedures and protocol:

Recruiting of participants was done through snowball sampling, where subjects nominate additional participants from their acquaintances. A sample of 10 students and faculty, with a range of expertise in wiki environments, were asked to participate through e-mails and an informative poster asking for research subjects (see Appendix E: Sample Recruitment Advertisement). Participation was voluntary.

In order to work with human subjects Ethics approval was obtained (see appendix A: Ethics certificate). Participants were chosen by the following criteria:

- Participants with experience on wikis in an educational setting.

- Belonging to one of the following scenarios: 1. Being a student in a class where a wiki assignment was part of their grade. 2. Being a member of a team where a wiki environment was implemented for writing collaboratively.

The interviewer had two training sessions to rehearse the protocol. This ensured that participants were exposed to minimal stress during the interview process. Participants were given a consent letter to read and sign. Any questions about the project were addressed before the interview began. All conversation from this point was recorded. It was made clear to the participants that the visualizations were not designed by the interviewer, so they did not have any pressure when commenting about the glyphs.

The first set of questions (To see the complete interview questionnaire refer to appendix J: Interview Questionnaires and Scenarios) served to collect demographic data and also to talk about the participants' experiences with wikis. After these questions were answered, a snapshot of an article from the Wikipedia (without visualization), was presented. The purpose of this page was to provide the participant with some context on the environment in which the new visualization would be implemented.

Four printed snapshots of the various visualizations were shown to the participant with the same order every time to minimize an order effect. In this context “the ‘order effect’ refers to the phenomenon that the temporal order in which information is presented affects the final judgment of an event” (Wang et al., 1998, p. 1-2). These glyph snapshots were shown at all times in both the second and the third sections of the questionnaire. The second set of questions contained the usability test, aimed to study the difficulty of using and understanding of the visualizations.

This section was also expected to clarify users’ preferences about the visualization, directing us to a better design for future implementation. Finally, the last section included the motivational test. This test intended to lead to a better understanding of wiki users’ participation and principally, to discover in which ways a visualization of their contributions would affect their collaboration.

Participants were asked, if possible, to answer every question bringing their current and past experiences on wiki projects to the interview. They were encouraged to talk freely about the extent to which these visualizations were likely to impact their wiki activity. The interview took approximately 40 minutes. Data from the interview was transcribed verbatim, and a qualitative analysis was performed.

3.3.2 Presentation of visualizations:

Four different visualizations were used in this project: “Sunword”, “CircleMagic”, “Pie chart” and “All wiki pages”. They were based on the glyphs proposed by Ruecker et al., where the principal goal was the “implementation of these graphics in wiki pages to present immediate feedback on contributions” (2008, p. 1).

One of the main objectives of this experiment was to test if users would welcome an implementation of a glyph as an informational tool, and also as a way to enhance their participation (See appendix D: The Visualization Glyphs). In the next illustrations (See Illustration 1 to 4) present a brief description of each one of the visualizations along with the snapshots used during the interview.

“Sunword”: this glyph shows all users’ names and the amount of contributions, the active user has a different font color. The size of the font gives a quick reference to who is the first-ranked contributor; this information is also supported by a rank and a percentage of the article the user owns. Shown in the middle is the name of the article and the total number of edits done to it by the printed date. In the lower part of the glyph window, a “details” button links to more specific information on the types of edits upon which the metrics were based.

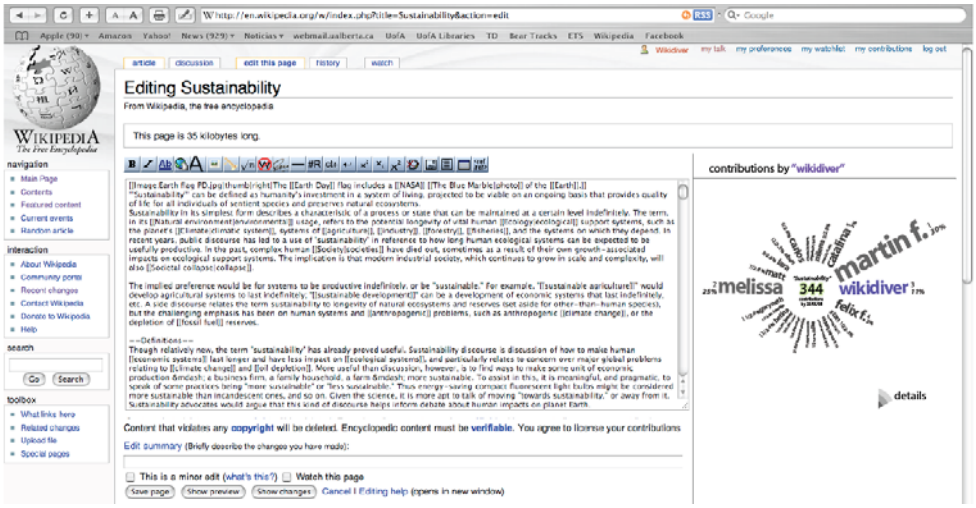


Figure 1: Glyph – “Sunword”

“CircleMagic”: this visualization presents the ranking of each user and the percentage of the article they own. Highlighted with a different colour is the active user. The most important characteristic of this glyph is the presentation of edit categories. The contribution to each type of edit is presented in different ring colours for the active user. Edits are categorized as follows: structural changes, links, structural changes, deletions, additions. The center of the glyph contains the name of the article and the total number of edits made to it until the printed date.

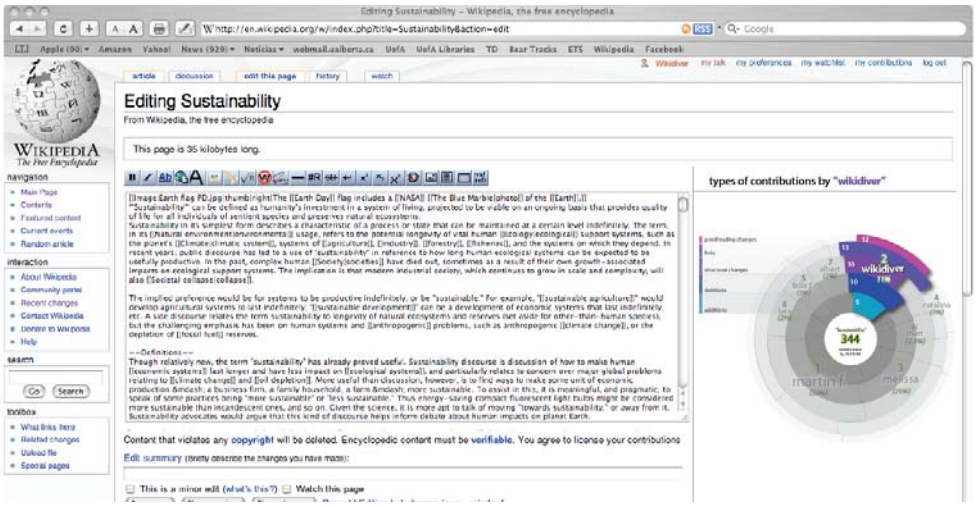


Figure 2: Glyph – “CircleMagic”

“Pie chart”: this glyph presents a single user’s contribution. Highlighted with a different colour is the active user. The glyph is intended to give the active user specific information on their own contributions; at the same time the gray areas are meant to represent other users’ contributions to the articles without giving specific information of those collaborating. The contribution to each type of edit is presented in different ring colours for the active user. Edits are categorized as follows: structural changes, links, structural changes, deletions, additions. The center of the glyph contains the name of the article and the total number of edits made to it until the printed date.

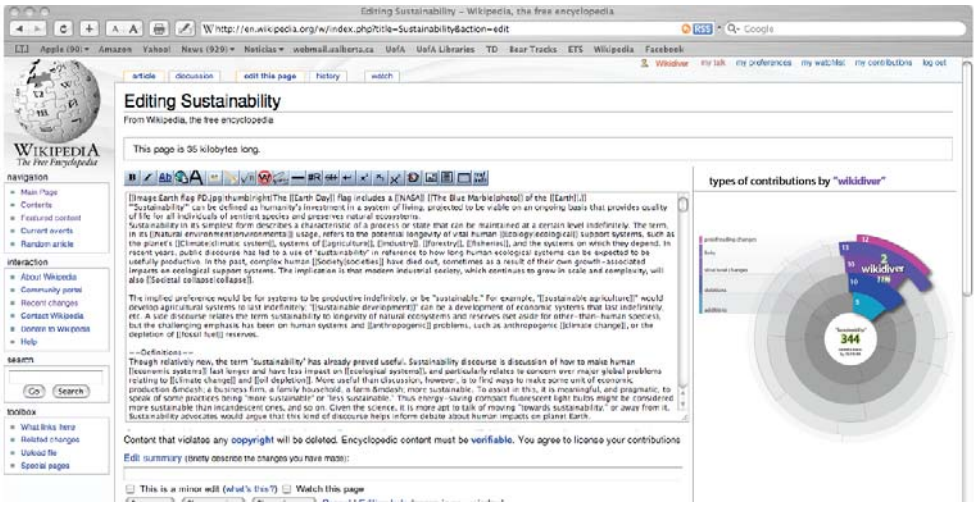


Figure 3: Glyph – “Pie chart”

“All Wikipedia pages”: it compares the whole Wikipedia content to the current article. The bigger ring in the background represents the ranking held by the user in reference to all the articles they have contributed to the wiki. The same ring also presents the total percentage the article’s content that the user owns. The smaller ring represents the ranking of the user within the current article and the ownership percentage.

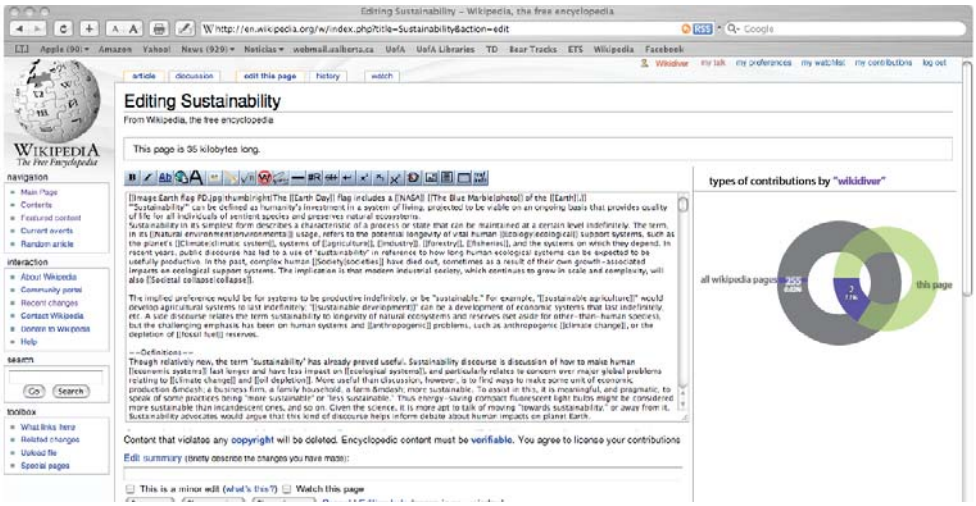


Figure 4: Glyph – "All Wikipedia pages"

By tracking contributors' activity, the visualizations are intended to reflect the effort of wiki users when writing collaborative documents. The usability questions are expected to give a sense of how easy it was for the participants to interpret glyph's information. Participants were asked to describe, one by one, all the characteristics, strengths and information they could identify in the glyphs.

The printed snapshots of the visualizations were also kept at hand during the motivational test. The motivational test questions were not aimed to get participant responses on every glyph. In these questions reference to a "graphic showing wiki contributions" or a "visualization tool" was made; and with this it was expected to discover the root of some of the motivational issues of wiki users in higher education.

Participant behavior was observed closely while answering the questionnaire, taking into account the possibility that valuable information could be overlooked if the results were based only on verbal responses. Particular body language (e.g. confusion, excitement) and contradictory responses were taken into account for the qualitative analysis.

3.3.3 Qualitative analysis of the interview data:

The first step in the qualitative analysis was the verbatim transcription of the interviews. Subsequently, all the files (audio and transcripts) were imported to NVivo 8.0. (QSR, Doncaster, AUS), Excel (Microsoft, Redmond, WA) was also used in the coding process. Software for qualitative research held and managed all information gathered during the interviews and permitted easy coding of the emergent themes. Transcriptions of the interviews consisted of four to five pages per person.

The Knowledge Organization Model (KO) proposed by Given & Olson (2003) was taken as reference. According to Given and Olson, “in studies that use inductive, qualitative, methodological approaches, thematic codes emerge from the information gathered during and following data collection” (2003, p. 158). In order to identify important themes to be coded, one has to “consider each research question, and decide how many themes will contribute to identifying relevant data” (Given & Olson, 2003, p. 170). While identifying pertinent themes and sub-themes, the researcher “must

go through each transcript (often many times) to code all instances" (Given & Olson, 2003, p. 171). Coded topics where chosen; first, for their relevance and direct relation to the questions addressed and secondly, by the number of occurrences found (Given & Olson, 2003).

For the usability test, the descriptions of the glyphs were used to create a table with all of the glyphs' characteristics (Table 1). In an ideal scenario, participants would have given a perfect description of the glyphs if they were able to correctly identify the next characteristics in each glyph:

Table 1: Glyphs characteristics used to codify usability test.

	Sunword	Pie Chart	Magic Wedge	All wiki pages
Visualization				
Users' names	x		x	
User name (individual)		x		x
Active user highlighted	x	x	x	
Font size = rank	x			
Users' ranking	x	x	x	x
Ownership %	x	x	x	x
Article's name	x	x	x	x
Article's total number of edits by date	x	x	x	
“Details” button	x			
Broken down edit categorization legend		x	x	
Contribution by user by type of edit		x	x	
Ranking/Ownership whole wiki				x

Participants were not directed if they misidentified the characteristics; they had the freedom to interpret the visualizations in any way. During the coding of the usability test, every reference to any characteristic of a glyph was marked as:

- Correctly identified: easy to understand or popular feature.
- Incorrectly identified: less intuitive or difficult to understand features.
- Not identified: overlooked or unpopular features.

Chapter 4: Findings

4.1 Part I: Analysis of wiki contributions

4.1.1 Introduction:

In this section of the chapter the results obtained through the manual analysis of 60 student-generated wiki articles are presented. This experiment, as described in the method, was performed by three research assistants, complying with ethics requirements and working with anonymised data. The procedures for this analysis were based in the method for manual analysis of Wikipedia articles described by Arazy and Stroulia (2009). The main objectives of the manual analysis after the rating of quantity and quality of the contributions were to:

- Explore the mechanics of writing/editorial tasks among students.
- Evaluate the accuracy of automatically calculated scores against the scores assessed by human raters.

The nature of educational wikis is different from public wikis such as the Wikipedia. For example: short time given for development of an article,

smaller number of versions, smaller number of contributors, retractable user information from contributions (adjusting data to human research ethics regulations) , are some characteristics taken into consideration when adapting Arazy and Stroulia's (2009) empirical analysis for our own purposes. As follows, the research questions that were addressed with this analysis are presented:

> ***1.*** *What weight do human assessors assign to each contribution category in giving an overall contribution score?*
>
> ***1.1*** *How do the automatically calculated metrics proposed by (Arazy & Stroulia 2009) correspond to the manually-assessed edit categories scores? Is there a difference between an educational setting and the results reported for the Wikipedia (in Arazy & Stroulia 2009)?*

In this first section of the chapter you will find: first, the tables with the statistical data coming from the experiment and a brief description; secondly, the interpretation and discussion of the results.

4.1.2 Presentation of Results:

Table 2: Inter-rater Reliability results on intra-class correlation coefficient.

	Intra-class correlation coefficient.
Manual analysis metrics	
Amount Structural Changes ……..0.664	
Amount Proofreading…… .0.811	
Amount Additions…….....0.906	
Amount Deletions ……....0.791	
Amount Links……...0.688	
Quality Changes ……0.717	
Quality Proofreading……0.552	
Quality Additions ……0.800	
Quality Deletions ……..0.506	
Quality Links……0.744	

The intra-class correlation coefficient intervals can be interpreted as follows: k = 0.7 - 0.9: Acceptable; k = 0.40 - 0.59: Moderate; k = 0.60 - 0.79: Substantial; k= > 0.80: Outstanding. This measure is also know as weighted kappa and is mainly used in methodology for psychology studies.

Table 3: Correlations in Manual Analysis of writing/editorial tasks (A*Q) of students in an educational wiki. All correlations are statistically significant at *P<0.01* (2-tailed t-test) when marked with '**'; otherwise marked as '*' at *P<0.05*

	Manual_ed	Struc_ch	Prfrd	Add	Dlt	Links
Manual Edits	1.000	.564**	.478**	.592**	.262**	.523**
Structural Changes	.564**	1.000	.325**	.488**	.279**	.333**
Proofreading	.478**	.325**	1.000	.069*	.299**	.181**
Additions	.592*	.488**	.069*	1.000	.076*	.599**
Deletions	.262**	.279**	.299**	.076*	1.000	.045*
Links	.523**	.333**	.181**	.599**	.045*	1.000

Manual_ed: Manual Edits, Struct_ch: Structural changes, Prfrd: Proofreading, Add: Additions, Dlt: Deletions, Links: Links

Table 4: Correlations, manual (A*Q) and automated calculations. All correlations are statistically significant at *P<0.01* (2-tailed t-test) when marked with '**'; otherwise marked as '*' at *P<0.05*

	Manual Analysis					
Automatic Algorithm	Manual_ed	Struc_ch	Prfrd	Add	Dlt	Link
Auto Edits	1.000**	.564**	.478**	.592**	.262**	.523**
Auto Current Owned	.336**	.475**	.146**	.545**	.072*	.386**
Auto Total Owned	.470**	.570**	.141**	.694**	.142**	.489**
Auto Current Proofread	.275**	.275**	.529**	.013	.169**	.097**
Auto Total Proofread.	.305**	.302**	.541**	.042	.205**	.120**
Auto Sentences Deleted	.163**	.261**	.190**	-.014	.475**	.015
Auto Releases with Dlt.	.535**	.449**	.649**	.176**	.409**	.262**
Auto Current Ext Links	.128**	.127**	.070*	.183**	-.017	.494**
Auto Total Ext Links	.185**	.206**	.120**	.234**	.041	.568**

Manual_ed: Manual Edits, Struct_ch: Structural changes, Prfrd: Proofreading, Add: Addition, Links: links, Dlt: Deletions, Auto Releases with Dlt: Auto releases with deletion, Auto Current Ext links: Auto current external links.

4.1.3 Discussion

4.1.3.1 Inter-rater agreement and intra-class coefficient:

The manual analysis brought together the rates of three research assistants, as explained in the method chapter. Foster & Cone (1980) described the “Inter-rater agreement as a measure of the consistency of ratings among raters by comparing the raters’ scores with each other” (as cited in Wu et al., 2007, pp. 230–239). To measure the trustworthiness of the rating across the rates the intra-class correlation coefficient was measured, this calculation is widely used as methodology for psychology studies (Table 2) (Fleiss and Cohen, 1973). The intra-class coefficient is equivalent to weighted Kappa as explained by Fleiss and Cohen, in which this coefficient is “the proportion of agreement corrected for chance, and scaled to vary from -1 to +1 (where a negative value means ‘poorer than chance agreement’). A value of unity indicates perfect agreement” (1973, p. 613-619).

The results of the intra-class coefficient in Landis & Koch (1977) can be interpreted as follows: *K = 0.40 to 0.59* moderate inter-rater reliability, *0.60 to 0.79* substantial/reliable, and *<0.80* outstanding. In table 2, the coefficients on the quantitative analysis of the types of contributions ranged from reliable to outstanding, while the rating on the quality of contributions goes from moderate to outstanding.

We considered important to explain in case 1.1 and 1.2 why the metrics quality proofreading and quality deletions were under (K=0.60) a reliable agreement among raters.

Case 1.1: Quality proofreading: The migration of the articles from the Blackboard wiki to Media wiki generated noise in the html code; objects such as internal anchors and tables of contents were difficult to migrate cleanly and became unreadable. This noise was somehow misleading and usually it looked as if proofreading changes and the articles ended up ranked as low quality proofreading changes (quality=1, see Appendix F: Categorization Procedures). The same was true with paragraph breakdown. Some letters at the end of paragraphs had been cut; this could be blamed on the migration of the wiki too. At the time of saving the new version, these ends of paragraphs looked like inaccurate proofreading edits that also were categorized as mediocre (quality=1).

Case 1.2: Quality of deletions: Very large articles where users create many edits of many kinds. It is very difficult to tell if the deletion was a real deletion or just a structural change. In this manual analysis we compared the current version vs. the succeeding version, which caused that sentences that were presumed deleted were actually added back in the subsequent versions, making this a structural change instead a deletion. If this change would have been caught as a structural change in the first

place, it would have had more chances of a good rating on quality. Also, if the content that was presumed deleted was considered a good contribution, the rating of the deletion would be considered vandalism (quality=1 deletion of important information, see Appendix F: Categorization Procedures). Our suggestion is that the comparison has to be made across more than one version made consecutively by the same user. If the one user has a group of saved versions called “release”, it would be more efficient to rate quality of contributions, comparing differences between the first and last version of that release.

4.1.3.2 Correlations manual analysis:

Some interesting events concerning editorial tasks in which wiki users were involved were identified through the correlation of the various metrics of the manual analysis. In table 3 we can see the A*Q scores of each kind of contribution (amount x quality of additions, deletions, structural changes, proofreading changes and links) compared against the others. It can be assumed that in the student generated wiki, users who edited the wiki are very likely to do more than one kind of task during the development of the article.

The next is a discussion of interesting cases that were extracted from the tables presented above:

***Case 2.1*:** There is a strong correlation between *Additions* and *Links* (r=0.599, *p*=0.01) and *Additions* and *Structural Changes* (r=0.488, *p*=0.01). However, there is a much lower statistical correlation between *links* and *structural changes* (r=0.33, *p*=0.05). The main premise to explain the strong correlation between Additions and links is that, in order to give validity and support to an argument added by the writer, one or more links are inserted for reference or extra information on the topic. In the second circumstance, if a user wants to add new content to a section of the article, it usually takes some structural rearrangement of the content that is already there; so that one may "fit" the new addition. Table of contents, titles and subtitles were considered structural changes; consequently, these contributions are bound to the addition task (as addition of new content).

Case 2.2: Users that make *proofreading* changes are less likely to make *additions* than any other task (r=0.069, *p*=0.05). It shows also that the correlation to any other task is low. This could mean that if a "copy editor" user is proofreading a version of the article, this person is going to be mostly dedicated to this task alone.

In Arazy and Stroulia's experiment (2009, p. 4-5), one of the discoveries was the "highly correlated rating of top contributors with the 'add' class

(r=0.65), meaning that top contributors are appraised based on the amount of additions they contribute with, followed by links(r=0.24)". On the contrary, in the same experiment deleting and proofreading tasks are not connected to the perception of tasks top contributors do (Arazy & Stroulia 2009). The next two cases could be taken as complement on Arazy and Stroulia's conclusions:

Case 2.3: Users that do *addition*s are less likely to do *deletions* (r=0.076 *p*=0.05) than any other task. This result reiterates the findings in Arazy and Stroulia's article (2009). We can presume that if top contributors are concerned about keeping their score and they had already figured out what kind of contributions are the "most important", they will probably neglect other "dispensable" tasks, as is the case of proofreading edits.

Case 2.4: Users that do *deletions* are less likely to do *links* (r=0.045, *p*=0.05). If the assumption in case 2.1 is true, *links* are strongly correlated to *additions* because they add quality value to the new content. In this case we assume that users that delete content don't invest much time linking material to contributions that are not theirs. We noticed that if a user deletes content, usually they will have the courtesy of deleting any links to it as well.

4.1.3.3 Correlations Automatic vs. Manual metrics:

The automatic algorithms used in this project were intended to identify the next metrics: *ownership, proofreading, links* (current and total version),, *deletions* and *releases with deletion*. The current version represents only the contributions that have survived through time and are in the current version; while total scores keeps a sum of all contributions made by a user to the article. “Sentences deleted” gives a total of sentences or objects erased; meanwhile, “releases with deletion” tries to catch at least one sentence deleted out of a set of versions saved by one user. Here, in the case of vandalism, a “reverse” function ignores the version where the vandalism was committed and doesn’t give any credit to the user.

In the results of the correlations between manual and automatic analysis an interconnection between metrics can be found. The “current owned” and “total owned” algorithms are highly correlated to Additions (r=0.545 and r=0.694). “Current proofreading” and “total proofreading” are related proofreading category (r=0.529 and r=0.541). And finally, “current links” and “total links” seems to point correctly at the links of the manual classification (r=0.494 and r=0.568).

The direction the ownership (current and total) algorithm took was identifying not only *additions* but also *structural changes* (current=0.475,

total=0.570). This means that both algorithms ultimately will represent a combination of the two metrics correctly. Lastly, the metrics “sentences deleted” and “releases with deletion” do capture deletions (r=0.475 and r=0.409). However, the case of “releases with deletion” draws more attention because it has a high correlation with proofreading (r=0. 649) and structural changes (r=0. 449) also.

It is curious to see how these two cases present such different results. While “releases with deletion” captured structural and proofreading changes; sentences deleted did not. In the best of cases, one could say that this algorithm was measuring more than one task, which is very convenient, as one could combine the two algorithms to rate different metrics. This behavior could be attributed to the fact that having a comparison of a set of versions gives a bigger perspective of the kind of changes a user is actually making.

For instance, the steps that precede a structural change for the relocation of a sentence or paragraph are: 1) copy, 2) delete, 3) paste. Imagine the case of comparing two successive versions, and the user saved the article after the second step, the scores can be incorrectly assigned as: one deletion in version_1 and one addition in the version_2. With the analysis of the whole “release” made by a user the algorithm ultimately corroborates the real nature of the contribution.

One ultimate conclusion is the fact that the algorithms that summed up the total of contributions have higher correlations to the manual metrics.
This could mean that a considerable amount of edits that were important were still deleted, lost, reworded in the making of the article.

The results of this first experiment were indispensable for the future implementation of a visualization of relative contributions in a wiki article. It not only it gave a general picture of the ways in which students contributed to this educational wiki, but also gave more confidence to reliable scores that accurately represent the collaborative participation of wiki users, as this scores eventually will be fed to the wiki visualization. These results also lead to other possible situations that afflict public wikis and also educational wikis, such as vandalism. More about this topic will be discussed in the conclusion chapter, bringing together introspections from the qualitative results from the interviews and these results.

4.2 Part II: Wiki users' Interviews

4.2.1 Introduction:

In this section qualitative data coming from the interviews of 10 volunteers were analyzed. The gathering of data from interviews gave in-depth comments from wiki users for educational purposes. The opening questions of the questionnaire were aimed to collect some demographic data. After collecting this demographic data, participants were asked to describe their previous experiences with wiki environments in general. Data from these questions, as well as the description of participants, are presented in this introduction.

The questionnaire was divided in two main topics. The first one was the usability test, where participants were asked to describe the components and characteristics of the visualizations or glyphs shown to them during the interview. The second section contained a set of questions about motivational impact of the visualizations. Taking into account that we had two groups of people with different experiences on wikis, we wanted to

discover how an implementation of the proposed visualization would affect the contribution of participants in real educational settings.

4.2.1.1 Description of participants and experiences with wiki environments:

Pseudonyms were given to participants: BT, SH, MD, OR, MH, KA, MP, MK, EW, AV. They all were students and researches at the University of Alberta. All the of participants that were selected to participate belonged to one of the next two scenarios:

Scenario 1: Course grade

1. As part of your assignments in your undergrad course_____(a real course they've taken) you are asked to work collaboratively with your classmates in a wiki project. The project consisted in writing an article of quality about _____ (specific interest of the participant).

Scenario 2: Research team

2. As a member of a collaborative research team, you have been asked to contribute to the wiki project. As a researcher, you won't be graded, but you hope that your work on the project will eventually

lead to conference presentations and publication in peer-reviewed journals.

Five of the participants belonged to the first case scenario and the other half to the second case. The next table shows participants' gender and age information.

Table 5: Demographics

		Students (n = 5)	Researchers (n = 5)
1. Average age (years)		25	29
2. Gender	Male	3	3
	Female	2	2

Course grade volunteers:

Two of the five participants that had experience working on a wiki as part of the class were graduate students in the Humanities Computing program; they were part of a class were a wiki was developed as part of the class workload. The other three participants were undergraduate students from the School of Business that implemented wiki technologies

within the curricula. The assignment consisted in writing three different articles collaborating in groups of 15 to 20 students. Here one of the students describes his class assignment:

> *We were given a topic that we had to contribute to, for a class wiki on Management Information Systems, and several… about 20 students per topic were assigned and each had to edit pages and… make contributions and basically populate the wiki with information that we gather through our own research.*

Research team volunteers:

Four of the researchers were currently using wikis in their research projects to track progress of their work. One participant gave us this brief explanation on how their team uses the wiki:

> *We have a research wiki for our group, stuff for our engineering group and… anybody who is working on research keeps a log on there or keeps notes on there. If you are doing a project that is part of a team, other team members can come in and edit the same page.*

The case of the fifth participant belonging to this scenario was different; this interviewee used to be a professor of University of Sonora where a

wiki project was run. This was a University-wide project involving all first year students; they were assigned to write a set of articles about their community. Here is a fragment of what the participant said about his experience with his big wiki project:

> *The first time that we tried it, it wasn't… the resulting wiki wasn't as good as we had expected it to be. But it was again our inexperience on to how to use such a big wiki, but in the following courses instead of recreating [do it again from scratch] the wiki, we changed it so the students would have to input more information, do some more collaboration, fixing errors, shaping that a little bit* more.

For more examples on how the rest of the students and researchers are using wikis as collaborative tool or an educational technology, please refer to the Appendix K (this appendix contains participant experience with Wiki environments). The interviews' findings are laid out in the next two sections divided by: Usability questions and Motivation questions.

4.2.2 Usability Test:

Participants were shown 4 different glyphs. The design of these glyphs was based in the visualizations of wiki contributions proposed by Ruecker et al., (2008). The variations of the glyphs have characteristics in common, the first three: "Sunword", "Pie chart" and "CircleMagic", mainly

present the same information. The “All wiki pages” visualization brings a bigger picture to the table, including stats from the whole wiki.

The usability section of the interview was designed to address the next research question and sub-questions:

> ***2.*** *To what extent can users understand the information visualized in the glyph proposed by Ruecker et al., (2008)?*
>
> ***2.1*** *Are there specific variations of the visualizations that users can more easily understand?*

In order to address these questions, first, participants were asked to explain the data displayed in each of the visualizations. Second, they were encouraged to mention what they perceived to be the most evident strength of each visualization. Finally, they were directed to point out any particular piece of information they were interested in; and the glyph with the easiest design to find this information.

The answers of participants were sorted into three categories for the first question, when explaining the data contained by the glyphs. Responses would fall into one of the next categories: correctly identified, incorrectly identified and not identified. In the graphs (in figures 5 to 9 and tables 6 to

10) you can find the frequency of responses of the codification for the first and second usability question.

4.2.2.1 “Sunword” usability:

The results showed that 50% of the comments explaining the characteristics of the “Sunword” glyph were correct (incorrectly identified: 11.25%; not identified: 38.75%) (Table 6). The most evident attribute was the representation of ranking through the size of the font. This characteristic turned out to be the most popular and was considered as the strength of the glyph (See figure 5). One participant commented that she found this glyph easy to understand:

> *At a glance you can see the differentiation between contribution levels.*

On the other hand, the users’ ranking was often confused and identified as the number of edits a person had contributed. The details button was mostly ignored by all the participants; this condition could be blamed on the nature of the materials’ presentation; the glyphs’ prototypes shown on paper could have had less impact than an interactive digital presentation. One of the participants that correctly identified the “details button” said:

The ‘details’ (button) with a little part of a pie chart, looks like is asking you to click on that for more in-depth details on the contributions, perhaps if ‘wikidiver’ is selected and you click ‘details’ it would give you more information.

With the inclusion of a “details” link in the “Sunword” glyph we propose a combination of this and the “CircleMagic” glyph. This way, should users require more detailed information, they would have easy access to it.

Table 6: Frequency of responses to “Sunword” glyph’s usability question (1), glyph characteristics.

	Correctly identified	Incorrectly identified	Not identified
Sundword’s Characteristics			
Users' names	8	0	2
Active user highlighted	2	3	3
Font size = rank	9	3	5
Users' ranking	5	3	2
Ownership %	5	0	5
Article's name	2	1	7
Article's total number of edits by date	7	2	1
“Details” button	2	0	8

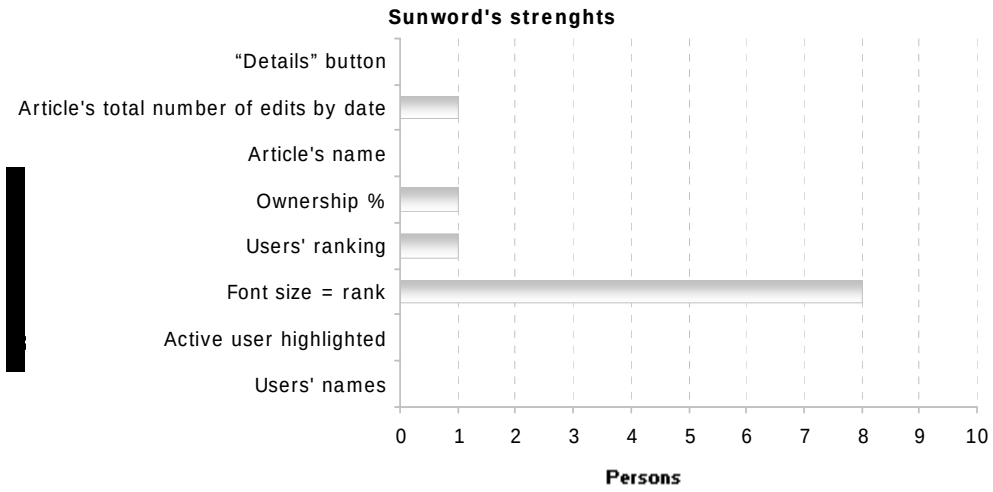


Figure 5: Response frequency to “Sunword” glyph’s usability question (2), glyph strengths.

4.2.2.2 “Pie chart” usability:

The study revealed that 40% of the comments on the description of the characteristics of the “Pie chart” glyph were correctly described (incorrect: 13.75%; not identified: 46.25%) (Table 7). The most popular characteristic was the broken down edit categorization’s legend. This legend is a description of the most common types of edits as explained by Arazy and Stroulia’s article (2009). Curiously, the number of contributions per type a user had scored was among the most mentioned strengths (Figure 6). Participants commented on the integration the broken down edit categorization:

It does have the breakout so it has a lot more granularity of the information on the contributions.

This other participant stated that the difference between this glyph and the previous (Sunword) one was the visualization of non numerical data:

It gives me qualitative information about my contributions, because the first one [Sunword] gave me quantitative information…

One can consider that the weight of the kinds of contributions users make would help understand better the mechanics of wikis. As seen in the manual analysis findings, interesting correlations among the different writing/editorial tasks users take over were discovered.

Table 7: Frequency of responses to “Pie chart” glyph’s usability question (1), glyph characteristics.

	Correctly identified	Incorrectly identified	Not identified
Pie chart’s characteristics			
Users' names (individual)	6	2	2
Active user highlighted	0	1	9
Users' ranking	3	3	4
Broken down edit categorization legend	9	0	1
Ownership %	5	2	3
Contribution by user by type of edit	4	2	4
Article's name	1	0	9
Article's total number of edits by date	4	1	5

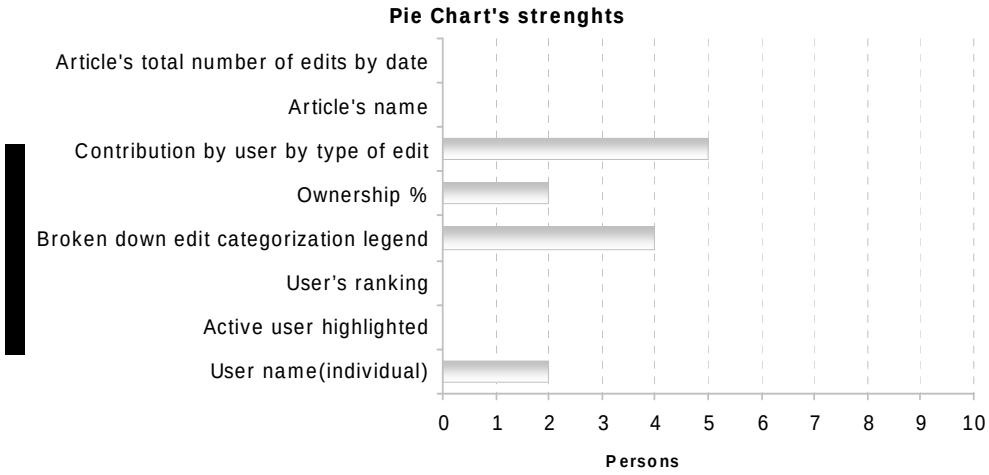


Figure 6: Response frequency to "Pie chart" glyph's question (2), glyph strengths.

4.2.2.3 "CircleMagic" usability:

The "CircleMagic" is a variation of the "Pie chart" glyph; the previous visualization was focused on one individual user's statistics, while this one was designed to show a group of wiki contributors. Their characteristics are very similar and this caused a predisposition from the interviewees to not mention some of the attributes that they had already identified in the "Pie chart" glyph. Whenever the glyphs' attributes were equivalent, it was decided to combine the results from the "Pie chart" into the "CircleMagic" coding table (Table 1 and Table 8). The final results showed that 53.70%

of the comments on the attributes were correctly identified (incorrect: 12.5%; not identified: 33.75%). The broken down categorization legend kept attracting participants' attention. Most of the participants also remarked that the integration of statistics from a group of users was convenient:

> *You can see a comparison between all the users*
> *They have put in all of the others' contributions to this particular article on sustainability.*

The participants pointed out that the visualization of all the members of the group was the CircleMagic's most evident virtue (Figure 7). Here is what a participant said:

> *It helps the visual metaphor, have that information there, to show that this is your contribution as part of the greater whole of everyone.*

The average timing when trying to understand this visualization was a bit longer in comparison with the first two glyphs; this was expected considering the amount of information and level of detail this visualization presents. The motivational test discussed some commented issues about anonymity and privacy .

Table 8: Frequency of responses to “CircleMagic” glyph’s usability question (1), glyph characteristics.

	Correctly identified	Incorrectly identified	Not identified
“CircleMagic” characteristics			
Users' names	9	0	1
Active user highlighted	1	1	8
Users' ranking	3	4	3
Broken down edit Categorization legend	9	0	1
Ownership %	7	2	1
Contribution by user by type of edit	7	2	1
Article's name	3	0	7
Article's total number of edits by date	4	1	5

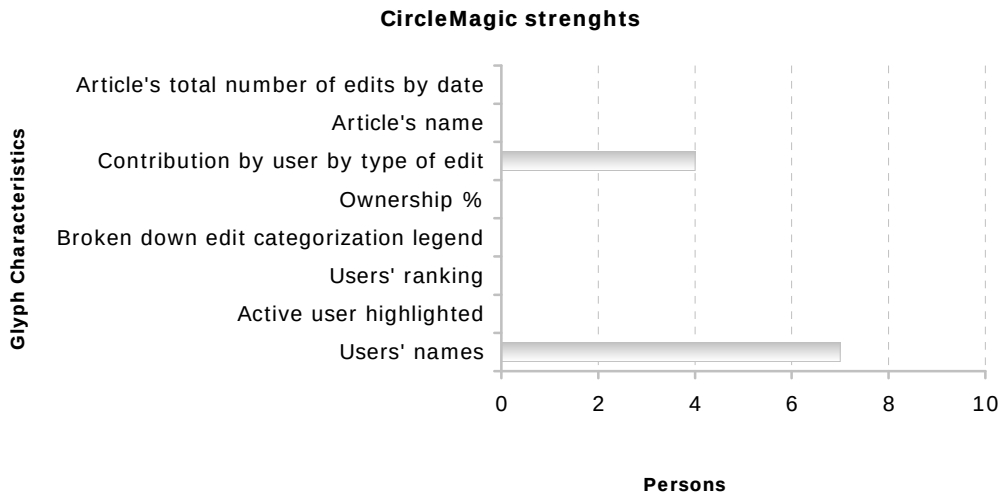


Figure 7: Response frequency to "CircleMagic" glyph's usability question (2), glyph strengths.

4.2.2.4 "All wiki pages" usability:

This glyph was the most controversial. Nearly all of the participants expressed that the design of the glyph was confusing. However, the results show that almost all of the characteristics of the glyph had been correctly identified by users. It seems that this glyph demanded an extra effort from participants to understand and explain the data displayed (Table 9). One of the main confusions expressed by participants was the overlapping of the two circles; they said that commonly, the overlapping of objects in a graph would represent a new layer of information that was not present in this glyph's design. It appears that the most distinguished component of this glyph was the inclusion of statistics from the whole wiki, which was not found in any other glyph (see figure 8). One of the

participants perceived that even when this glyph was presenting a different approach to contributions in the whole wiki, it was going to be an insignificant percentage:

> *I guess the only worthwhile thing of this one is the "all Wikipedia pages" ring, but I also assume for a regular user, the percentage is going to be very small and you'll have something like 0.0001; so this one doesn't struck me as useful, but maybe it would be if you had a small wiki.*

This last glyph helped to understand a bit more about participant preferences. Participants said they would like the information contained in this glyph to be accessible but not present all the time. They also mentioned that the information was too vague and over-simplistic to really matter.

Table 9: Frequency of responses to “All wiki pages” glyph’s usability question (1), glyph characteristics.

	Correctly identified	Incorrectly identified	Not identified
“All wiki pages” characteristics			
Gray ring represents the whole wiki	7	0	3
Ranking | Ownership percentage within the whole wiki (highlighted in blue)	8	2	0
Green ring represents the current article	6	0	4
Ranking | Ownership percentage in the current article (highlighted in blue)	7	1	2

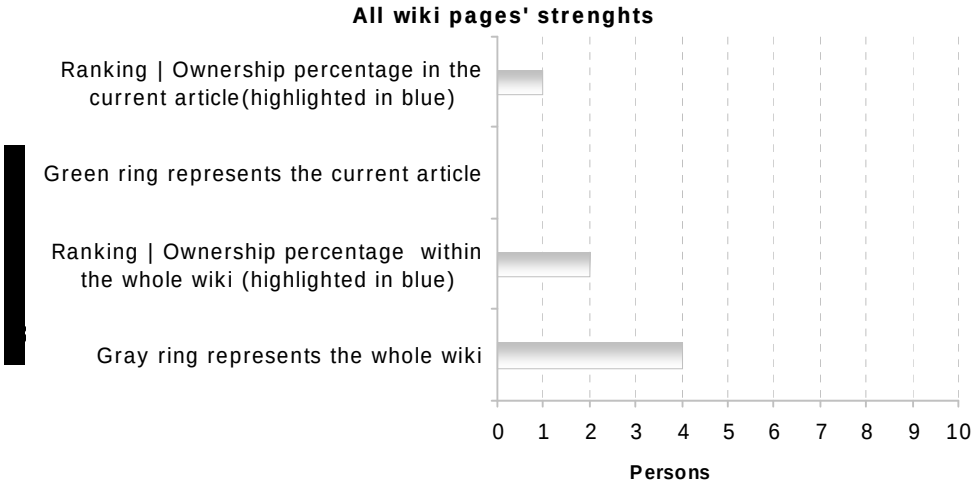


Figure 8: Response frequency to "All wiki pages" glyph's usability question (2), glyph strengths.

4.2.2.5 Discussion:

In the final stage of the usability questionnaire we asked participants what would be the most interesting data they would like to see in a visualization tool. According to the participants' answers there were three main pieces of information they cared about the most: a) Ranking, b) Percentage owned, and c) Contributions per type. Participants confirmed that the first two were easiest to find in the "Sunword" glyph. The third, contributions per type, was inherent to both the "Pie chart" and the "CircleMagic"; however, "CircleMagic" was selected as the easiest reference for this data.

The usability test was designed to determine if the design of the glyphs was positively understood. As shown in the previous tables (Tables 6 to 9), the “CircleMagic” visualization obtained the highest scores when people identified their characteristics, followed by “Sunword” (Table 10). Overall, participants had good comments about the “CircleMagic” glyph’s design. This visualization was voted as favorite by 6 out of the 10 participants (Figure 9). For instance, here is a fragment of what one of the researchers said:

> *It lists all the information in number three [Pie chart] with the categorized changes but also has information from number 2 [Sunword] in a nicer presentation, more aesthetically appealing; with the information from other users contributing from this page as well.*

Table 10: Comparison on percentage of responses to all glyph's usability question (1), glyph characteristics.

	Correctly identified	Incorrectly identified	Not identified
Glyphs			
Sunword	50%	11%	38%
Pie Chart	40%	13%	46%
CircleMagic	53%	12%	33%
All wiki pages	35%	7%	22%

In all the four glyphs, it was found that most of the confusions were caused by the numbers representing the rank and amount of contributions. Some of the participants expressed that they would like to see an explanatory descriptions the visualization, one student added:

> *Hum… well I don't actually know what the numbers mean, like I took a guess, but I don't actually know. So, it may be good if there is some kind of a menu or a little… legend?*

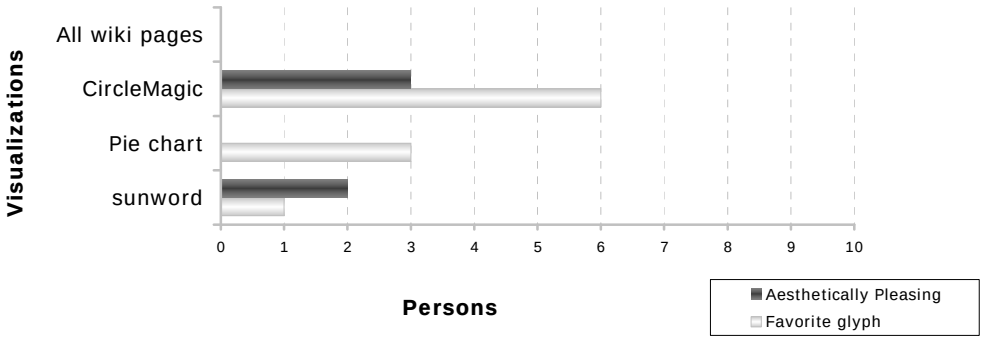


Figure 9: Favorite visualization and aesthetically appealing.

Most of the participants commented that it was interesting to see other persons' contributions. Nevertheless, some of them didn't seem too excited, knowing that other users could have access to their information as well. The anonymity topic will be discussed in the motivational section of this chapter. The majority of participants discarded the "All wiki pages" glyph as not useful. Some others recommended making a combination of the "Sunword" and "CircleMagic", the first one for quick reference and the second when details on the contributions were desired.

4.2.3 Motivation Test:

In this section you will find the qualitative results and discussion from the motivational and behavioral questionnaire. In the motivation questionnaire we were more concerned about the impact of a Wiki visualization implemented in an educational setting.

The information collected from this set of questions was devoted to address the next research question:

> 3. *How would visualizations of the attribution metrics impact users towards contributing to the wiki? What is the expected impact of the visualization in users' wiki behavior?*
>
> *3.1 Are there differences between the diverse variations of the visualization?*

During the usability test, participants became familiar with the visualizations; they now had a better understanding of the possible implementation of these glyphs in a real case scenario. First of all, we were particularly interested in the feelings a visualization like this would inspire to wiki users. We asked participants to allocate their feelings in a positive, neutral or negative inclination towards having one of the visualizations inserted in a wiki technology. Second, we asked them to tell us in which ways they would contribute to a wiki document if they had one of the visualizations available. Ultimately, we asked them to come up with a strategy to be the first-ranked contributor.

The objective of these questions was to find motivational factors that would stimulate or stop collaboration in the group when composing a wiki

article. From the emerging patterns discovered through the qualitative analysis of these responses, three relevant trends were identified:

- Participants' perceptions on visualizations' motivational factors
- Ranking and competition as motivational determinant
- Considerations on anonymity

4.2.3.1 Participant perceptions on the motivational factors of visualizations:

Most of the conversations revolved around the integration of the visualization as an incentive to write. The statistics provided by a visualization of relative contributions were supposed not only to show a wiki user's number of edits but to reflect on the quality of contributions made by the users.

Quality contributions represent all kinds of collaboration that enhance the overall quality of the article. Quality contributions are not always evident. It is difficult to recognize the real worth of what it is being added to an article. Frequently, in group work, the way to achieve quality is by peer reviewing; contributions are judged as valuable or not by the same authors. The concept of quality is subjective; and in this case, good contributions have to fit the whole group's perspective of quality. As previously explained in

this thesis, user contributions are evaluated by an automatic algorithm proposed by Arazy and Stroulia (2009) in which it is “assumed that the contributions that persist in time are more relevant.

Generally, wiki collaborators are not required to have any specialized knowledge on the topic they contribute. Nonetheless, it is usual that a few of the people who write and correct information are very knowledgeable about the subject. Some of the participants believed that expertise on a topic ensures high quality contributions and therefore higher ranking. One researcher stated that if she was aiming for quality contributions, she would do the following:

> *I would pick articles or topics in which I’m… of my expertise, because I can contribute more and edit, or make more changes to those articles faster... I would choose topics… for example: cultural studies, literature, maybe some authors that I really like to read or videogames that I’m including in my thesis or pages about my country of origin or places I know…*

The level of confidence and collaboration increases when wiki users write about topics which they consider themselves familiar with. They feel that the more they know the more their contributions are going to survive to other’s review. A student expressed concern about having his contributions deleted by the “experts” that were better at the subject:

> *There is a good chance that someone will know more of the subject than you and your stuff is going to be deleted, or much more could be added.*

As we can see, the notion of being an expert on the topic is connected with the assumption of quality product. Doing research before contributing gives the user more self confidence about their contribution. One student stated that, knowing the value of his contribution, he would try to learn more about the subject, so he can *"add more legitimately or delete more"*. Another participant emphasized that the glyph would show her if her effort really mattered, pushing her to study more the subject before contributing:

> *I would do my research, and make sure that what I have is correct before contributing. Because if somebody else is editing my stuff their score is going to increase… and then I'd be like 'all that for nothing'.*

Another motivational factor discussed by the participants came from the ownership concept. The next extracts from the interviews, explore two ways of thinking on how the sense of ownership affects participants' contribution. The first one presents an ideal example of how ownership would motivate wiki users to compose a quality end document. The second example shows us a case where the visualization of ownership could trigger improper behavior to increase the percentage of the article

they own, by contributing more content with less quality or to cheat.

In the first case we found that if a group of wiki collaborators care the way they see themselves reflected on their work; their attention to what is in the article would increase; even the little details. This is what a student said:

> *I think that you create ownership to it right? And if you contributed a lot on something then you want it to look the best it can. So you are going to go in and make those spelling changes… grammatical corrections or whatever you need to because it reflects you and them.*

The second case give us an example of the impression a student had about ownership, and what kind of behavior can be unleashed if students compete to posses the rights of the article for grading. This is what a student thought:

> *I would just try and you know…reword other things and contribute other… I'd probably contribute more but also go through and do things like: reword sentences so that I own them.*

In the time of the interview, participants seemed confused by the concepts of relative contributions in 'aggregated version' or 'final version' of an article. As explained in the methodology section, the final version of the

article is expected to have the contributions that had the most relevance to the content of the article; while the aggregated version counts all the contributions no matter what the quality is. After clearing up these concepts, most of the participants expressed how they wanted to see both scores in the visualization.

The majority felt that their final version's score was going to be smaller than the aggregated version's score. They thought there was a big chance their contributions were going to be reedited. They also said that collaborators are usually expecting to be rewarded by the effort they put into their work and every edit was part of that effort. One of the researchers manifested his apprehension about the concept of final version score. He explained:

> *I think the aggregated score, um it could be a bit misleading by… the problem I would have with having this score[final version] at all is that it may not tell the entire truth as to the quality of contributions that someone's made. So I think the aggregated score would be perhaps a more honest reflection of the work someone's put into it, I think is the time they've spent with the material, the time the spend working on it… not the portion of their work that makes it through to the end.*

In conclusion, a visualization of the final version score is good for reviewing your contributions to the final product; the aggregated score

would represent better individual efforts and would reflect much more the involvement of wiki users throughout the making of the article.

The presentation of the two scores would have a higher motivational reach, projecting those worked more towards supplement content and to those behind other editing tasks. The comparison between these two scores could also serve for adjusting individuals' involvement in different parts of the writing process participation in all editorial tasks.

4.2.3.2 Ranking and competition as motivational determinant:

The second relevant topic that rose during the interview was the evaluation of ranking and competition as motivational factors. The rank was given by the sentence ownership percentage; this means that in order to get a higher ranking, the user had to have a persistent contribution in the final version of the document. Participants mentioned that having the ranking in the visualization would motivate people to contribute. One of the students stated that seeing his ranking would have a positive effect on his contributions:

> *I think when people can see themselves stacked up against other people then they're going to try harder right away.*

Some of the participants tried to come up with clever ideas on how to get the first-ranked place. Two of them agreed that a good way to score high ranking would be starting a new article. Their supposition was that if you are the first contributor, chances are, you will write the foundation of the article and others will build up on that. Another participant was worried about the attitude of the students, should they be presented with a ranking. This participant thought ranking would increase low quality contributions or drive away the focus from the content of the article; he commented:

> *Try not to abuse the system just to get on top, people may try just to put more data just to get more ranking. So... I think… I think the effort stays the same. I mean the effort should be to get new Information… but just taking care of that and not just for the ranking.*

Competition, as well as ranking, has the power of changing persons' attitudes. In educational settings, competitive situations may affect people's performance. Depending on how serious the competition is taken, people's attitudes can go from beneficial to damaging a group's work. In the next paragraphs two examples where participants considered healthy competition as good motivation are presented. After that, we present the experience of students when competing for grades; they believed that if their grade depended on the results calculated from the algorithm, this could increment low quality contributions.

One ex-professor expressed that when wiki collaborators have access to their score compared to others', and the general cooperation to the group's work gives them a sense of pride:

> *As far as how it would improve the information they were inputting, I think it makes a little bit of competition and it's always good, just to see who is doing more effort, just to try to add a little bit more. And I think it creates a sense of pride in your work, just to see how much effort you put into it and how much it matters in the end.*

When checking the strengths of visualizations, another participant noted that if she knew that other users had access to her statistics, it was more likely she would feel motivated to contribute more, here are her reasons:

> *I think that it [Sunword] would motivate the person to contribute more because you are not the only one that is going to be seeing it, there are other people that are going to be seeing it. It is kind of an ego booster in a way too, and you know ok… other people know when I'm contributing to the article.*

A student found the visualization to be risky if the competition between the students was taken too seriously and also if they were not clear about ranking and their individual grades.

> *I'd feel better about myself but I also think that it would become an endless competition. Um… because, people are going to assume that you are graded on what part or size of the chart you have, and like it had happen before; everyone wanted to edit the last minute.*

When the same participant was asked about the dynamics of the class if they felt they were competing for score, he stated:

> *Students would try and undermine each other just to get the largest piece of the pie. I would! If my mark was on the line and I had 11% I'd like do as much as I could to get 50% of the pie or bigger than anybody else, 'cause we are on a curve right?*

An interesting comment was sent by e-mail after the interview. This participant wrote that collaboration among undergraduate students is usually viewed as unachievable due to the competition for a better grade; this lessens their willingness to cooperate and help each other. Here is what this student wrote:

> *I think that the entire concept of wikis revolves around collaboration and the concept behind the assignments we're given is competition. Collaboration and competition don't really mix well together and I think it's ultimately flawed if the two are trying to be used together in an assignment.*

Ranking was one of the major topics that arose from the interviews. In

educational environments, ranking is a controversial idea. It holds social comparison and reputation, but “when rank has been earned and signifies excellence, then it's generally accepted” (Fuller, 2004). Ranking creates a competitive environment; it also pushes people to contribute even more.

4.2.3.3 Considerations on anonymity:

Is interesting to see how people would react to the fact that their personal score would be available to other members of the group. Do users take more responsibility for their contribution when they know that their statistics can be accessed by others? One of the participants said that she was sensitive to what her colleagues would think about her work: *“I don’t like to be the person that has the least contributions.”* Better performance is expected from users that care about reputation among their colleagues.

Another participant stated that as long as his personal information was not displayed it was fine, although he did not mind using his first name when working in a large group. One of the interviewees added that contributions in a group project are no secret and that part of the managing of the work group was to be aware of what others were doing: “*you know, you can always tell in a group, it’s not like you don’t know! so… this [visualization] just makes it easier to track I think.”*

The concept of wikis is strongly based in the freedom of collaboration and anonymity of contribution. Educational wikis are usually linked to the users' data in one way or another; in the end, these collaborators are expecting some kind of reward out of their job; this reward being part of a grade or copyright on the article. One of the researchers considered that anonymity was important for people to collaborate

> *I think if you start monitoring who's saying what, you… you start playing with the notion of collective wisdom, people think that they going to be molested, that their words are going to be tracked, it may act as a barrier to what people really think.*

4.2.3.4 Discussion:

In this chapter the results from the students/researchers interviews are presented. The overall response of participants was positive; they liked the tool we were proposing to implement in real case scenarios. It is worth mention two of the comments in which participants stated how they would use the visualization. One participant said that it would be great to have this visualization to balance the work done by every member of the team:

> *If I was in a class and I was working with fellow students on a project I think it would act as an equalizer, it'd make sure that the work that I'm doing is consistent with the work that other folks are doing in a class, and I would hope that it*

> *would act as an equalizer for everyone else as well. Umm… particularly if you use it as a set up as an expectation that everyone would be in a certain range of contributions it would be a useful tool to make sure that everyone is contributing to the group project.*

Another participant said that a combination of the “Sunword” and the “CircleMagic” would be the ideal amount of information on collaborators’ job. She thought that these stattistics could be used to manage not only current but future projects as well:

> *I think it’s nice to have access to the percent of other people as well… just to see, you know, if you wanted to do it again with a different article, maybe somebody is really good at proofreading and you can just say… Why don’t you just do all the proofreading edits? and then kind of assign things…make things a little more easily.*

From the qualitative analysis of the data, it was found that the majority of participants were more attracted to the visualization called “CircleMagic”. It was also discovered that the wiki visualization concept was well received among participants and that they find it a useful tool.

Chapter 5: Conclusions

Wiki technologies offer student-centered learning with more opportunities to succeed. The trend of implementing collaborative tools for writing, such as wikis, into the curriculum is growing rapidly. The notion of several students working on a class assignment as a group and then submitting a single report accompanied by a peer-reviewed presentation is not new. In this sense, wikis can be thought of as a natural progression in technology of this educational concept, brought forth into the 21st century.

The skepticism that appeared with the introduction of technologies in the classroom, far from stopping its momentum, helped to get more facilitators and students involved in the process of assessing these tools. This involvement, begun with apprehension, is now building trust among users online technologies like wikis are engaging, innovative and adaptable; the perfect tool to be brought to the classroom and other higher education settings.

However, after 10 years since the first development of the wiki concept, the challenge now is not to prove their efficacy in education; the real

challenge is to explore their potential for integration, keeping in mind the learning needs of the people that will be using them, their learning styles, learning objectives, and how wikis are going to aid in achieving these goals.

The novelty of the implementation of wiki environments in higher education is finally pushing the research community to investigate and unveil the secrets behind the capabilities of this tool as an educational technology. Much is still unknown and many opportunities for further research exist. This thesis is undertaken to address one small but crucial investigation: the attribution of relative contributions to authors in higher education. In order to present users with their relative contributions, issues related with authorship attribution were investigated.

This thesis presented the procedures and results of manual analysis in which 3 research assistants manually graded the contributions of students in a wiki assignment. Our manual analysis adapted from Arazy and Stroulia's experiment on Wikipedia articles (2009) made an improvement in the process of rating user's effort by:

- Giving wiki users proper recognition on their effort not only rating the extent of their edits but also the quality of each one of their contributions.

- Having the students' final grade in the assignment assigned by peer reviews, we were able to justify human raters' perception on relative contribution to the articles.

The results of the manual analysis lead to the following conclusions:

- The inter-rater agreement showed a high level of reliability; this demonstrates that the manual experiment was well designed and can be replicated in similar research (Appendix F). However, there is still room for improvement. At the time of migration of data from a wiki, related metadata will be present with noise in the exported data. It is recommended that if migration is imperative, the raters take extra precautions rating what appears as low quality contributions.

- The statistical correlation between editorial and writing tasks was highly significant and revealed some specific behaviors, supporting discoveries by Arazy and Stroulia (2009). Users that *add* content also insert *links* or make *structural changes*. However, Adders usually do not *proofread* or *delete*; and finally, users who *delete* content are not prompt to insert *links*.

- The correlation between manual and automatic analyses showed that: “Ownership” (current and total) algorithms efficiently detect *additions* and at the same time *structural changes*. The “Proofreading” (current and total) algorithm successfully detects the *proofreading* category. The algorithms for “sentences deleted” and “releases with deletion” do capture *deletions* as well as *proofreading*. And finally, the “current links” and “total links” algorithms identify *links*.

- Contrary to expectations, the statistical results suggested that the algorithms that presented a sum of the total contributions in each category better represent the human ratings of contributions, over algorithms that represent the current version.

A second experiment, an interview, was designed to test the introduction of a new visualization tool, glyphs. The glyphs were used to present wiki users’ relative contribution in a visual manner. User interviews were conducted on 10 participants to evaluate the usability of the visualization and to determine motivational issues related to the implementation of the proposed tool.

- Overall, participants were enthusiastic about the possibility of having the tool available. They considered that the visualization

would be helpful when working with wikis in educational settings. Participants stated that the most important characteristics to be included in visualizations are: ranking, ownership percentage, and scores on contributions per type.

- The “CircleMagic” visualization was favoured by 6 out of 10 participants; results showed that it was also the easiest to understand and use. The second most popular glyph was the “Sunword”. Participants were not impressed with the “All wiki pages” glyph which and received bad reviews, for its lack of useful information.

In order to achieve a better design, our recommendations are:

- Numerical scores need to be clarified; a mouse-over description containing rank, number of contributions, number of contributions per type, and ownership percentage would be sufficient.

- A combination of the “Sunword” and “CircleMagic” glyphs would contain all the information of interest to users.

- Users within a classroom setting should have the option to be anonymous to other users or to use nicknames.

The qualitative analysis of the motivational questionnaire showed that most of the participants were eager to have the visualization implemented in educational wikis. The conclusions are the following:

- Participants mentioned that the visualization would increase their collaboration. The visualization was recognized by users as good instrument for project management, for its informative characteristics.

- Drawbacks associated with the accuracy of this visualization tool were found. Some users said that they would develop specific strategies to get a higher contribution score. Some of the strategies mentioned were: to be well informed (an expert) in the topic they were collaborating on; to contribute big amounts of text regardless of the quality; to make many small proofreading changes that are long lasting; and to reword sentences to takeover ownership.

- Participants showed a positive reaction to the presentation of ranking, as they declared that some level of competition would definitely push them to contribute more. However, they also showed concern about the consequences of high competition in a collaborative environment where individual grades are given.

- In order to nullify bad behavior from contributors, we recommend an emphasis be given to students or work groups on the importance of collaboration over competition. It would also be beneficial to stress that authors are represented by the quality of the final product instead of the percentage they own.

The methodology presented in this thesis serves as a reference point to support further research. The next step of this investigation is the actual implementation of a visualization of relative wiki contributions in a real scenario in higher education. Its implementation in the real world would reveal the effectiveness of the tool, provide insight into other potential applications, and reveal other undesired behavior.

References:

Adler, B. T., Alfaro, L., Pye, I., & Raman, V. (2008). Measuring author contributions to the Wikipedia. *Paper presented at The International Symposioum on Wikis, WikiSym 2008.* Porto, Portugal: ACM.

Arazy, O., & Stroulia, E. (2009). Wiki Attribution: A Utility for Estimating the Relative Contributions of Wiki Authors*. Poster presented at the Third International Conference on Weblogs and Social Media.* Seattle, WA: ICWSM.

Baker, N. (2008, March 20). The Charms of Wikipedia. *The New York Review of Books*, *55*(4). Retrieved from: http://www.nybooks.com/articles/21131.

Dick, W. (1995). Instructional design and creativity: A response to the critics. *Educational Technology, 35*(5), 5–11.

Eisenberg, M., & Barry, C. (1988). Order effects: A study of the possible influence of presentation order on user judgments of document relevance. *Journal of the American Society for Information Science, 39*(5), 293-300.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, pp. 613-619.

Fuller, R. W. (2004). *Somebodies and Nobodies: Overcoming the Abuse of Rank.* Gabriola, B.C. : New Society Publishers, pp 2-60.

Given, L. M. & Olson, H. A. (2003). Knowledge organization in research: A conceptual model for organizing data. *Library & Information Science Research*, 25:157-176.

Hess, M., Kerr, B., & Rickards, L. (2006). *Wiki User Statistics for Regulating Behavior*, working paper. Retrieved from http://icd.si.umich.edu/684/files/684%20wikistat%20paper%202.pdf.

Hoisl, B., Aigner, W. & Miksch, S. (2006). Social Rewarding in Wiki Systems – Motivating the Community. *Paper presented at Second International Conference, OCSC 2007*. Berlin, Heidelberg: Springer, *4564*, 362-371.

Klamma, R., Chatti, M. A., Duval, E., Hummel, H., Hvannberg, E. H., Kravcik, M., et al. (2007). Social Software for Life-long Learning. *Educational Technology & Society*, *10*(3), 72-83.

Landis, J. R., & Koch, G. (1977, March). The Measurement of Observer Agreement for Categorical Data Source: Biometrics. *International Biometric Society Stable, 33*(1), pp. 159-174. Retrieved from http://www.jstor.org/stable/2529310.

Landow, G. (1992). *Hypertext: The Convergence of Contemporary critical Theory and Technology*. Baltimore: Johns Hopkins University Press.

Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., Cosley, D., et al. (2005). Using social psychology to motivate contributions to

online communities. *Journal of Computer-Mediated Communication, 10* (4). 212-121.

Majchrzak, A., Wagner, C., & Yates, D. (2006). Corporate Wiki Users: Results of a Survey. *Procedings at The International Symposioum on Wikis, WikiSym 2006.* Odense, Denmark: ACM.

Mitchell, P. (2006). Wikis in education. In J. Koblas (Ed.), *Wikis: tools for information work and collaboration* (pp. 119-143). Chandos Publishing (Oxford) Ltd.

Moore, L. (2007). *Social Exchange Theory and the Internet*. Retrieved December, 2007, from http://www.informationtechnologycrossing.com/article/370062/Social-Exchange-Theory-and-the-Internet/.

Parker, K. R., and Chao, J. T. (2007). Wiki as a Teaching Tool. *Interdisciplinary Journal of Knowledge and Learning Objects, 3,* pp. 58-67.

Priedhorsky, R., Chen, J., Lam, S. K., Panciera, K., Terveen, L., & Riedl, J. (2007). Creating, Destroying, and Restoring Value in Wikipedia. *Proceedings of the 2007 International ACM Conference on Supporting Group Work.* New York: ACM, 259–268. Retrieved from http://www-users.cs.umn.edu/~reid/papers/group282-priedhorsky.pdf\ .

Rashid A., Ling K., Tassone R., Resnick P., Kraut R. & Riedl J. (2006). Motivating Participation by Displaying the Value of Contribution.

Conference on Human Factors in Computing Systems. New York, NY: ACM, pp 955-958.

Ruecker S., Fiorentino, C., Arias, C., Bouchard, M., Ganev, V., Arazy, O., et al. (2008). Visualizing Relative Wiki Contributions. *Poster presented at the Third Annual Chicago Colloquium on Digital Humanities and Computer Science, DHCS 2008*. Chicago, IL: DHCS.

Sammons, M. (2003). Exploring the new conception of teaching and learning in distance learning. In M.G. Moore & W.G. Anderson (Eds.), *Handbook of distance education* (pp. 387-397). Mahwah, NJ: Erlbaum.

Siemens, G. (2004) *Connectivism: A Learning Theory for the Digital Age*. Retrieved December, 2008, from http://www.elearnspace.org/Articles/connectivism.htm.

Siemens , G. (2007). *Connectivism 101: for the curious* [Powerpoint slides]. Retrieved from Learning Technologies centre, University of Alaska Fairbanks: http://www.slideshare.net/gsiemens/connectivism-101.

Stoner, M. L. Easterbrooks, S. R., & Laughton, J. M., (2005). Handwritten and Word-Processed Story Retellings by School-Aged Students Who Are Deaf. *Journal of especial educational technology, JSET, 20*(3), 1-10.

Tedjamulia, S., Olsen, D., Dean, D., Albrecht, C. (2005). Motivating Content Contributions to Online Communities: Toward a More Comprehensive Theory. *Proceedings of the 38th Hawaii International Conference on System Sciences, HICSS* 2005. Hawaii: IEEE

Walter, H. (2006). Moving a printed-based editorial into electronic form. *Library of congress cataloging-in-publication data*, Electronic Textual Editing/edited by Lou Burnard. Pp 139-145.

Wang, H. (1998). *Order effects in human belief revision.* Ph.D. Dissertation, The Ohio State University.

Wattenberg, M., Viégas, F., Dave, K. (2004). Studying Cooperation and Conflict between Authors with history flow Visualizations. *Proceedings of International conference for human-computer interaction, CHI 2004***.** Vienna, Austria: ACM. Retrieved from http://alumni.media.mit.edu/~fviegas/papers/history_flow.pdf.

Wattenberg, M., Viégas, F., & Hollenbach, K. (2007). *Visualizing activity on Wikipedia with chromograms.* Proceedings of International Conference in Human-Computer Interaction, INTERACT 2007*. Rio de Janeiro, Brazil, pp. 272-287.*

Wu, S., Whiteside, U., & Neighbors, C. (2007, December). Differences in Inter-Rater Reliability and Accuracy for a Treatment Adherence Scale. *Cognitive Behaviour Therapy*, 36(4), 230-239. doi:10.1080/16506070701584367.

Appendix A: Ethics certificate

1783

Arts, Science & Law Research Ethics Board (ASL REB)

Certificate of REB Approval for Fully-Detailed Research Project

Applicant: Cristina Arias

Supervisor (if applicable): Stan Ruecker

Department / Faculty: Humanities Computing/Faculty of Arts

Project Title: Visualization of wiki users' contributions in higher education.

Grant / Contract Agency (and number):

Application number (ASL REB member) 1783 (CDH08-512)

Approval Expiry Date: June 2, 2009

CERTIFICATION of ASL REB Approval

I have reviewed your application for ethics review of your human subjects research project and conclude that your project meets the University of Alberta standards for research involving human participants (GFC Policy Section 66). On behalf of the *Arts, Science & Law Research Ethics Board* (ASL REB), I am providing expedited approval for your project.

Expedited research ethics approval allows you to continue your research with human participants, but is conditional on the full ASL REB approving my decision at its next meeting (***June 16, 2008***). If the full ASL REB reaches a different decision, requests additional information, or imposes additional research ethics requirements on your study, I will contact you immediately.

If the full ASL REB reverses my decision, and if your research is grant or contract funded, the Research Services Office (RSO) will also be informed immediately. The RSO will then withhold further funding for that portion of your research involving human participants until it has been informed by the ASL REB that research ethics approval for your project has been granted.

This research ethics approval is valid for one year. To request a renewal after ***June 2, 2009***, please contact me and explain the circumstances, making reference to the research ethics review number assigned to this project. Also, if there are significant changes to the project that need to be reviewed, or if any adverse effects to human participants are encountered in your research, please contact me immediately.

ASL REB member (name & signature): C. Donald Heth

C. Donald Heth

Date: June 2, 2008

Appendix B: Information/Consent Letter

Introduction

You are invited to participate in the research project**:** Visualization of wiki users' contributions in higher education, being conducted at the University of Alberta. Your participation in the project is totally voluntary. This project is intended to test design issues and effectiveness of two visualization glyphs that display relative wiki contributions. It is hoped that this research will contribute to the design of a useful tool that will enhance the educational experience and participation when collaboratively writing wiki articles. This research is being conducted by Cristina Arias, a Master's student in Humanities Computing under the supervision of Stan Ruecker, an Assistant Professor in Humanities Computing.

Method

If you choose to participate in this study, you will be interviewed for approximately sixty minutes. A compensation of 20 CAN for your participation in the study, will be provided at the end of the interview. Participant's personal information will be made anonymous. All summaries and any direct quotations used will be anonymous in their attribution. We would like to interview people who have worked in wiki articles as part of their research or course assignment. This interview will be audio recorded. All interview data, including tapes, transcripts, and written notes, will be kept in a cabinet in Stan Ruecker's office located inside the Old Arts Building for at least five years, or when the project is complete, whichever is greater, after which the data will be destroyed (All the electronic data will be protected by a password and a firewall).

During the interview, first you will be asked some questions about your experience with wiki articles and collaborative writing; secondly, printed examples of the visualization graphics will be shown to you in order to

continue with the second set of questions about the effectiveness of the graphics and the possible impact of these on your motivation for participation. It is expected that wiki users in higher education will benefit from the findings of this research with better practices and understanding of motivational visualization.

There are no known risks associated with participation in the project.

The researchers associated with this project will comply with the University of Alberta Standards for the Protection of Human Research Participants. Data recorded in the course of this research will be available only to research collaborators who have signed confidentiality agreements.

Your Rights

You have the right not to participate in this study.

You have the right to withdraw from this study at any time, with no personal consequences. If you opt out, the data collected in your interview will not be used in the study.

You have the right to privacy, anonymity, and confidentiality.

You have the right to have any data collected in this study kept in a safe, secure place.

Other uses

The data collected in the course of this research project will be used principally in Cristina Arias' thesis research. In addition it will be used in research articles, scholarly presentations, and in other academic activities. Data for all uses will be handled in compliance with the University of Alberta Standards for the Protection of Human Research Participants, which can be read in full at: http://www.ualberta.ca/~unisecr/policy/sec66.html.

If you have any questions about this research project, please contact either:

Cristina Arias
Humanities Computing Master's Programme
Department of Interdisciplinary Studies
cga@ualberta.ca tel: (780)438-3067

Dr. Stan Ruecker
Assistant Professor
Humanities Computing Programme
Department of English and Film Studies
sruecker@ualberta.ca tel: (780) 9146372

Name of Participant (please print)
Signature of Participant
Date

Appendix C: Confidentiality Agreement

Researcher(s)/Transcriber Confidentiality Agreement

Project title: "**Visualization of wiki users' contributions in higher education"** Conducted by Cristina Arias and supervised by Stan Ruecker.

I, _____, the *Researcher/Transcriber*, agree to:

1. Keep all the research information shared with me confidential by not discussing or sharing the research information in any form or format (e.g., disks, tapes, transcripts, digital files) with anyone other than the *Researcher(s)*.

2. Keep all research information in any form or format (e.g., disks, tapes, transcripts, digital files) secure while it is in my possession.

3. Return all research information in any form or format (e.g., disks, tapes, transcripts, digital files) to the *Researcher(s)* when I have completed the research tasks.

4. After consulting with the *Researcher(s),* erase or destroy all research information in any form or format regarding this research project that is not returnable to the *Researcher(s)* (e.g., information stored on computer hard drive).

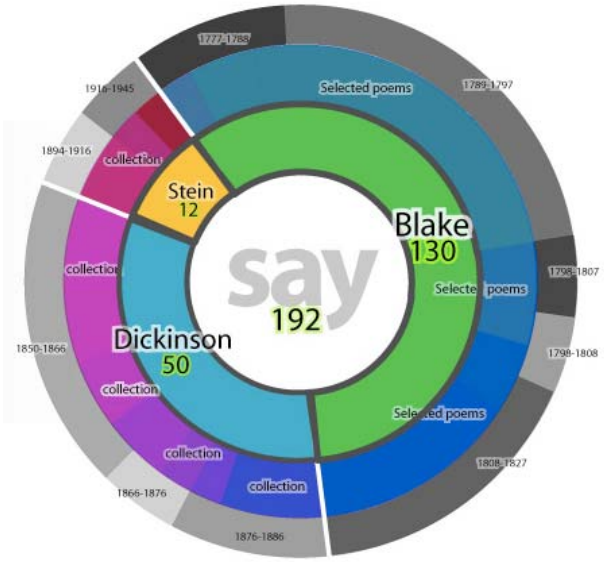
Researcher/Transcriber

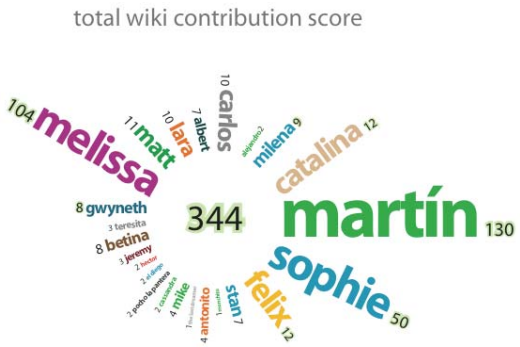
_____ _____ _____

(Print name) (Signature) (Date)

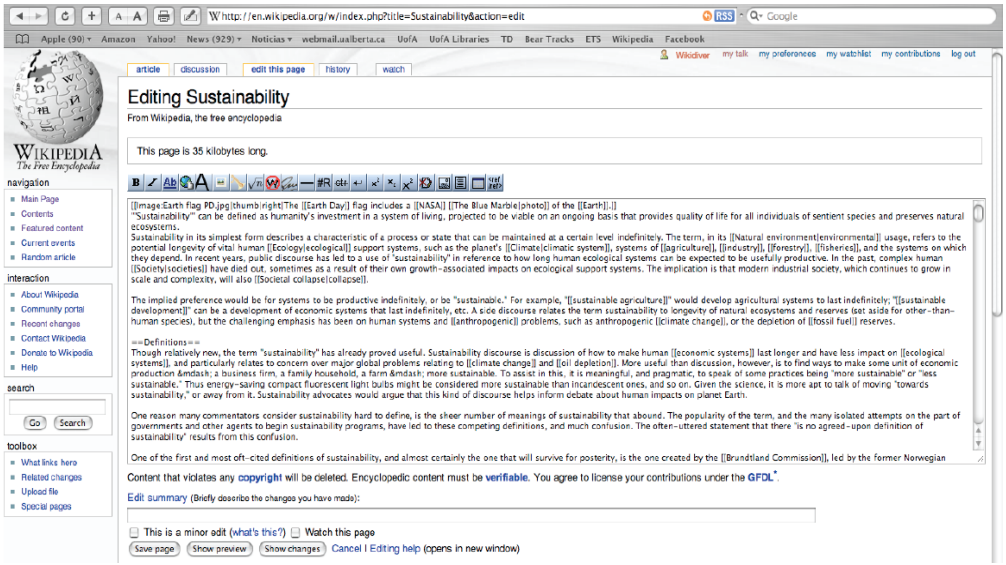
Appendix D: The Visualization Glyphs

Participants will be shown two information glyphs. The first one summarizes the total contribution of the top 25 authors (the sketch shown here is taken from another application that emphasizes word frequencies). The second glyph shows the same summary, but broken down by different kinds of contributions, such as additions, deletions, or structural changes to content, links, or citations. Note that each participant's name will be included so that it appears not as largest or smallest contributor, but somewhere in the middle range.

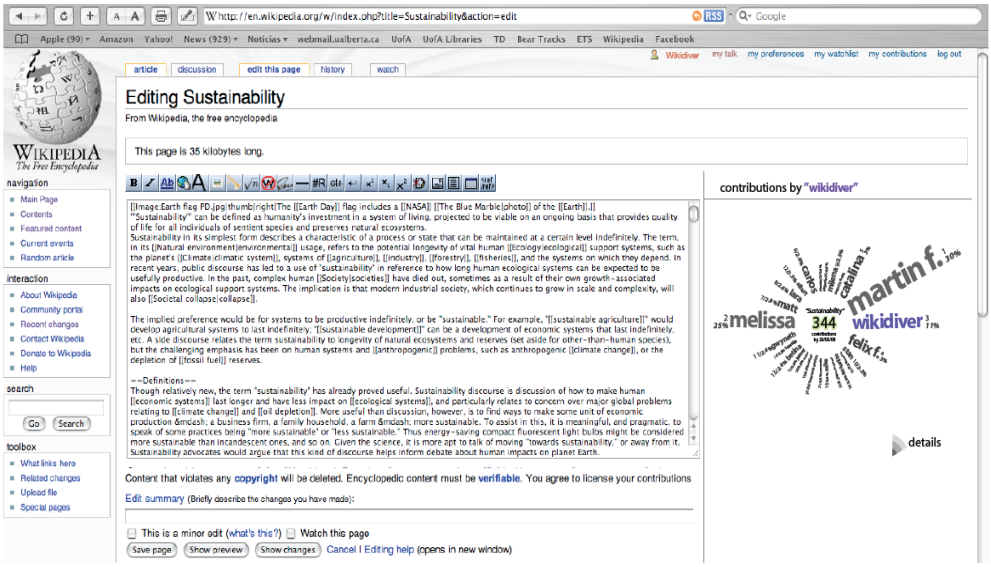




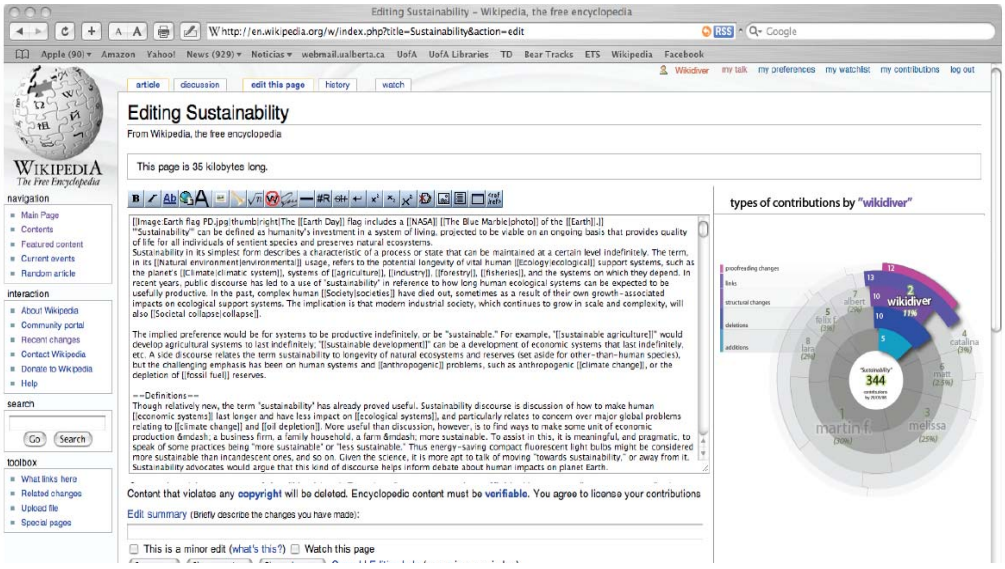
Page 1: Wikipedia screenshot without glyph



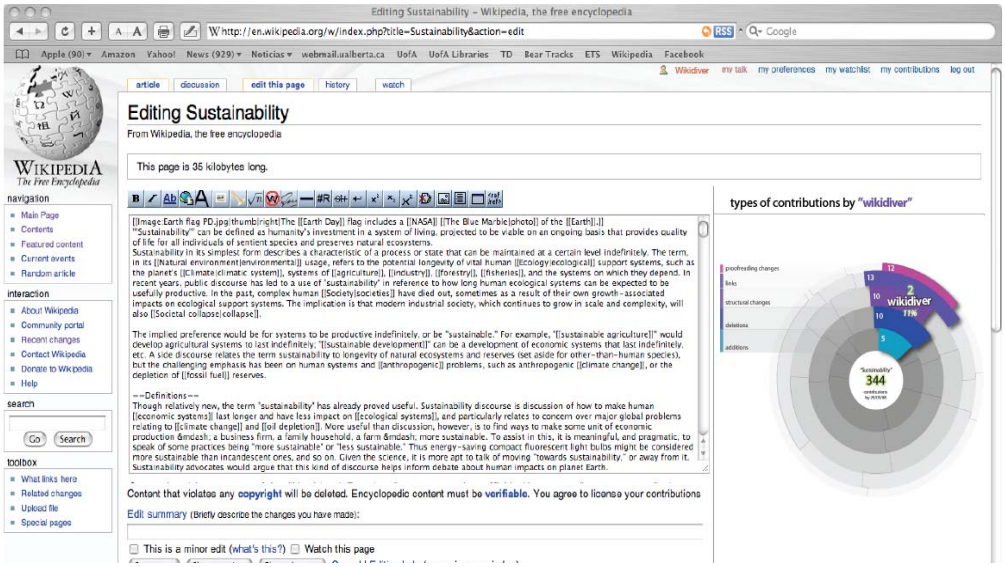
Page 2: “Sunword” glyph



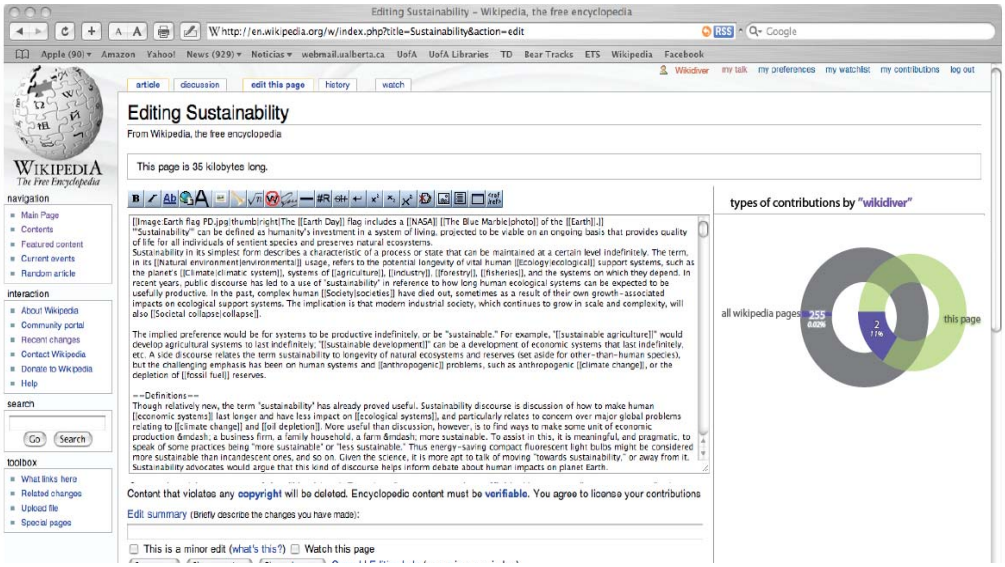
Page 3: “CircleMagic” glyph



Page 4: “Pie chart” glyph



Page 5: “All wiki pages” glyph



Graphic designer: Carlos Fiorentino

Appendix E: Sample Recruitment Advertisement

INTERESTED IN VISUALIZATION METHODS OF WIKI DATA?

WONDER WHAT IS BEING DONE TO DEVELOP BETTER SOFTWARE FOR VISUALIZATION OF USER'S CONTRIBUTIONS IN COLLABORATIVE WRITING?

Then sign up to test new wiki visualization glyphs being developed here at the U of A in the Humanities Computing department!

We're looking for students, faculty, and staff who have recent experience in wiki environments. We want your input into a new visualization gadget designed to show wiki user's contributions. If you are interested, we'll only ask 1 hour of your time, at your convenience.

Contact Dr. Stan Ruecker at 492-7816 or sruecker@ualberta.ca

Appendix F: Categorization procedures of manual analysis

Following we list the various categories of edit changes done to a Wikipedia article, explain each one, and list the procedures for categorization in terms of Amount and Quality.

Structural changes:

A structural change would include text that is rearranged to ensure fluency and transition of the article. Structural changes imply NO addition of new content (text) except for titles and subtitles. Example: Reversion of vandalism when greater than a sentence or contains an idea (links) is included in this category.

Structural Categorization Procedures

Amount	**Sentence/paragraph/image re-ordering**	**Title/Subtitle/TOC insertion**	**Text to bullet/table breakdown**
1	One object	One object (TOC insertion)	One sentence
2	Two objects	Two objects	Two sentences
3	Three objects	Three objects	Three sentences
4	Four objects	Four objects	Four sentences
5	Greater than four objects	Greater than four objects	Greater than four sentences
Quality			
1	Decreased fluency	Misleading description	Decreased fluency
2	No fluency improvement	No description improvement	No fluency improvement
3	Minimal fluency improvement, reversion of vandalism, position change of object/link	Minimal description improvement, reversion of vandalism ,position change of object, TOC insertion, title/subtitle insertion	Minimal fluency improvement, reversion of vandalism
4	Medium fluency improvement, breakdown of large paragraphs	Medium description improvement	Medium fluency improvement
5	High fluency improvement, complex structural improvements (breakdown of large paragraphs, position changes/organizing)	High description improvement	High fluency improvement

Proofreading:

Proofreading will include cleaning up of any typographical, grammatical or spelling errors, minor changes to the text/existing links, minor deletions and additions. Copyedit. Examples: Capitalization, Grammar, Spelling, abbreviations, use of italics, bold, hyphenation, spacing before and after dashes and symbols, punctuation. In case of English grammar errors the rearrangement at word level will be considered proofreading and not structural change. Vandalism and reversion of vandalism on a small scale (less than a sentence) will be categorized as proof reading.

Proof-Reading Categorization Procedures

Amount	**Proof-reading word edits**
1	One to three
2	Four to six
3	Seven to nine
4	Ten to twelve
5	Greater than twelve
Quality	
1	Inaccurate edit (object changed from correct to incorrect), and vandalism
2	No edit improvement (object changed with no difference in quality), (/), (*) or (#) added to TOC
3	Minimal edit improvement (evident error), reversion of vandalism, addition of reference numbers to text, one word changes, small style changes
4	Medium edit improvement (some knowledge of subject matter), complex sentence (many words) improvements

5	High edit improvement (High knowledge of subject matter), complex sentence (many words) improvements

Additions:

This includes addition of text and images. (Objects: text, images or descriptions to links) Complete sentences, taking into account that a sentence represents integrally an idea. Vandalism additions made by an author greater than a sentence will be included in this category.

Additions Categorization Procedures

Amount	**Additions of new ideas/sentences/images**
1	One to two
2	Three to four
3	Five to six
4	Seven to eight
5	Greater than eight
Quality	
1	Object(s) unrelated to topic, inaccurate information, or vandalism
2	Object(s) did not improve or decrease quality of article
3	Minimal article improvement (adds relevant information to existing ideas), new sentence added to existing paragraph, new sentence added to reference,
4	Medium article improvement (adds new and relevant ideas indirectly related to topic), new paragraph added with subtitles but lacks good structure.
5	High article improvement (adds new and relevant ideas directly related to topic), new paragraph added with good structure(subtitles, proper spacing, references)

Deletions:

This includes deletion of text, images, and links. Vandalism deletions of text that is greater than a sentence or contains an idea will be included in this category.

Deletions Categorization Procedures

Amount	Deletions of sentences/images/links
1	One to two
2	Three to four
3	Five to six
4	Seven to eight
5	Greater than eight
Quality	
1	Deletion of relevant text with good structure (subtitles and spacing), relevant links, Vandalism
2	Deletion of relevant information without good structure (did not improve or decrease quality)
3	Minimal article improvement, deletion of irrelevant text
4	Medium article improvement, deletion of vendor sites and vendor information
5	High article improvement

Links (Hyperlinks):

Addition of reference links to support the content of the article: Includes internal, external links, and vandalism links.

Links Categorization Procedures

Amount	Additions of working topic supporting links
1	One to three
2	Four to six
3	Seven to nine
4	Ten to twelve
5	Greater than twelve

Quality	
1	Addition of irrelevant links, broken links and vandalism links
2	a relevant Link that neither improves article knowledge nor provides inaccurate information
3	Minimal article improvement (relevant link, but is indirectly related to topic), other language links
4	Medium article improvement (relevant link that increases knowledge of topic), category links
5	High article improvement (relevant link that increases knowledge of topic and is directly related)

Borderline Cases

1. Structural vs. Deletions

Deletions: Deletions of titles

Structural: Additions of titles

2. Structural vs. Links

Structural: Addition of title/link to table of contents.

Link: Addition of “Back to top” link to body text

3. Proofreading Vs. Additions

Additions: A new idea and knowledge is created within an existing sentence.

4. Proofreading Vs. Links

Link: Brackets are inserted to existing text to create a link.

Proofreading: One bracket was inserted to existing text in order to fix a non-functional link (creation of link was intended by previous author).

Proofreading: Changes to existing table of contents which include addition of random symbols (/, #, []) and text.

Appendix G: Data collection

Both automatic and manual techniques will be used in estimating user contribution for the sample Wiki pages.
The sample will also include a rating of the overall contribution of each user, based on the student's assignment mark.

Automated algorithms will calculate user-based statistics for each Wiki page, as follows:

- Number of Edits
- Current Owned (sentences owned in last Wiki version)
- Total Owned (total sentences owned by the user)
- Total Proofread.
- Sentences Deleted
- Releases with Deletion
- Total Ext Links
- Current Ext Links

Manual analysis of the Wiki pages in the sample will be performed by analyzing each edit made to the article (by comparing historical Wiki versions), determining the category of the edit, and rating the a) extent and b) quality of the contribution (on a 1-5 scale). Being 5=high quality or great amount of contributions.
The categorization will be based on the classification framework as defined in Arazy and Stroulia's article (2009). The ad-hoc categories identified are:

- Additions

- Deletions
- Structural changes
- Proofreading
- Internal / External Links

Once all edits are analyzed, a cumulative contribution (A*Q=amount *quality of contributions)

Appendix H: Example of control table for manual analysis (template)

	Quality								
Links	Amount								
	Quality								
Deletions	Amount								
	Quality								
Additions	Amount								
	Quality								
Proofreading	Amount								
Structural Changes	Quality								
	Amount								
		15:51, 19 January 2008 0 (Ta	18:54, 16 February 2008 112	10:39, 26 February 2008 174	21:26, 4 March 2008 279 (Tal	14:42, 23 March 2008 219 (Ta	14:42, 23 March 2008 219 (Ta	14:45, 23 March 2008 219 (Ta	11:14, 24 March 2008 133 (Ta
Article: Biochips		60	59	58	57	56	55	54	53

Appendix I: Summary of the Method for the Manual analysis experiment

1. - A template of the database with the names of the chosen articles that was part of our sample (60 articles) was created for the manual analysis. This template was sent to the other two research assistants that worked simultaneously.

2. - For each article, this database contained both “amount” and “quality” cells of the edit categories, and a list of versions with date and hour of the contribution as well as contributor ID (Appendix H: Example of control table for manual analysis).

3. - The data from each article was collected, ranking the amount and quality of every edit in consecutive versions.

4. - Relative contribution’s statistics were calculated for each article.

5. - copy of the results taken from the above mentioned database was made in order to compare it with the automatic results. The algorithms were implemented on the articles.

6. – A comparison table with the results from the algorithm and manual analysis’ results was prepared to calculate statistical correlations among them

Appendix J: Interview Questionnaires and Scenarios

Pre-test Questionnaire - level of experience and demographic questions:

1. - What is your sex? Male Female

2. - What is your year of birth? _____

3. - What degree and major are you currently working on?

4. - Please describe your previous or current experience with wiki environments.

Please rate yourself on a scale of 1-5 for the following questions.

5. - I am very sensitive to how people perceive my relative contribution to a group project.

1	2	3	4	5
Strongly disagree				Strongly agree

6. – Getting high marks in courses is very important to me.

1	2	3	4	5
Strongly disagree				Strongly agree

Presentation of two scenarios:

We have developed two visualization tools that are intended to show relative contributions of different wiki authors. We would like to see what your reaction is to each of these visual items.

Scenario 1: Course grade

1. As part of your assignments in your undergrad course_____ (a real course they've taken) you are asked to work collaboratively with your classmates in a wiki project. The project consisted in writing an article of quality about _____ (specific interest of the participant).

Scenario 2: Research team

2. As a member of a collaborative research team, you have been asked to contribute to the project wiki. As a researcher, you won't be graded, but you hope that your work on the project will eventually lead to conference presentations and publication in peer-reviewed journals.

Usability Questions (after showing the visualizations):

1. Please, try to explain the data displayed.

2. What are the strengths of each visualization; what is it good for?

3. How would you identify specific data you are interested in (e.g. your contributions)?

4. What would it be your score in the final version compared to your aggregated score?

Impact Questions

5. How would you feel if a graphic showing user's contributions were to be implemented as an add-on in a wiki technology?

6. Now that you have the visualization tool inserted in wiki technologies how are you going to contribute?

7. What would be your strategy if you wanted to be the first-ranked contributor?

Appendix K: Participants' Experiences with Wiki environments
(Extracted from Interviews' transcripts)

Pre-test Questionnaire - level of experience

Q4. - Please describe your previous or current experience with wiki environments.

- *"I've always used Wikipedia for like research and like just like looking on like articles, and information and stuff. But for this research project itself it was basically categorizing conflicts. So I had to go through about 266 articles [she later corrected that instead of articles she meant Wikipedia discussion pages] and categorize the different kinds of conflict that I see and then my perceptions on this and you like rate it; that was the first research. Then the second one was categorizing the quality and quantity of the edits in different articles."*

- *"I've look thinks up on Wikipedia, and I've contributed to wiki articles for a class. MIS 311. Three different subjects, and three different pages that we had to edit, working between 8 and 11 people on each one, and would add maybe… two paragraphs… three paragraphs, or a subheading kind of thing. That's about it, no many structural changes or anything. More adding than anything, I didn't delete any. Graded by our peers, based on our contributions, how much you added or deleted or structural changes."*

- *"I'm a mayor public user of wikis. I'm using them left and right. I'm actually, well you think I have no life, but umm sometimes I just go and look something up on Wiki. If I think about something, something pops into my mind, I just go and read about it on Wiki. Wiki is my first source for most information. For finding any basic stuff to begin with. To find out about anything. I learn about a new book or I learn about a new movie and I go and read it, or I've seen a new movie and I want to see what people are saying about it, or a plot summary for places that I didn't understand, because I still can get confused with places with English I go on Wikipedia and I read the plot over there. As far as (anything goes? 2:34) I haven't really participated on anything a lot, really, actually the actual Wikipedia or*

any wiki outside the classroom I haven't participated at all, but hum I would on occasion I would report the vandalism or something like that. So I'm mostly a consumer not a contributor."

- *"For the past four years now, we've had a research wiki for keeping things neat. We've had a research wiki for our group, stuff for our engineering group and anybody who is doing research keeps a log on there or keeps notes on there. If you are doing a project that is part of a team, other team members can come in and edit the same page. Media wiki is a customized version of that wiki. Right now I'm creating a bunch of extensions for media wiki to help other people collaborate and we still have our research wiki going on too, so our research projects are kept on there. Pretty much all the notes we shared with each other are kept on wiki. So in a nutshell that's all my experience with a wiki*"

- *"I've used Wikipedia extensively, for research basically, looking up stuff that I don't know. Hum… I've contributed to Wikipedia once or twice but I stopped because it was kind of pointless. Hum... I've corrected things, um... I used to work for a company that wrote from scratch their own wiki and we uh… we had to basically populate it and populate it out you know with stuff… and uh… other that that in a class throughout the web, you know different types… a BB wiki [BlackBoard wiki]. Um…yeah… I'm pretty knowledgeable I guess about wikis, but I never actually built one myself. We had to do…uh… like we were given a topic that we had to contribute to for a class wiki on Management Information Systems, and several… about 20 students per topic were assigned and each had to edit pages and… make contributions and basically populate the wiki with information that we gather through our own research."*

- *"As a user I've edited many pages but, hum… I've learned more about wikis in this project and research management class. Hum.. in which we worked in teams. It was five, five students. Also we had support from three professors that were interested in helping us in developing this project. So they wanted us to write a proposal grant for NSERC or NSHERC I don't remember. So, we designed the methodology for developing the wiki. So, we started researching on wiki software, there are so many wiki software out there that it was a bit difficult for us to choose from the regular media wiki software, which is hum… from Wikipedia, or so many other open source software. Because we wanted software that was flexible enough to allow us have a user tag wiki. So people could tag the article. And also we wanted to have both options, to have: user*

tagging, but also hum… semantic classification. So that's a little bit more…like… customized. Yeah, it's more controlled by expert people who developed this classification, anthologies, etc,. and on the other hand we wanted this user tagging, and like… just uh… people who want to organize the Wikipedia, according to their interests. And that made us look for… that was what help us… trying to search for different software. At the end I think we decided just to stay with media wiki. Hmm so, hmm, the question was if we would use this semantic media wiki software, which is different, that is more controlled, that uses expert classification, controlled classification. And the purpose was… hmm, or our intention was to have people who programmed like an add-on that helped users to tag also. So one of the parts of the project was that. We included like a year or something, I don't remember how much time but it was like a year or two years for developing this new… it's like a…it's not a plug-in… like an extra feature for the semantic wiki. And that would be also available for other people for contribute. Later on we also designed a prototype of the interface. So in this interface we hmm… use the same Humanities Computing logo, and the regular wiki features, like search and how would an article look like. Hum…. Also we developed… we designed this user page, which was a little bit different than Wikipedia. And in this user page they ideally would have control on their user tags and their classification of words on… on the wiki. And, OK… so… the third part was hmm… a study about the results of the wiki. After the wiki was online and many people from Humanities Computing and Digital Humanities were invited from different universities to participate developing these wikis; our intention was to… analyze what were the topics they were interested in. And try to understand how…what is the structure of Humanities Computing as a discipline. So, because there is no convention on exactly what this discipline is, because is from interdisciplinary discipline, etc., all those things. So that was what these professors were interested in. And, so after a year or so these people were working on articles in Wikipedia. Then, we would hmm… conduct a study to try to understand how… the study in the tags, the user tags. So we would look at the user tags, look for frequency of some tags that were used, semantic preferences, syntactic preferences that were some kind of linguistics? And later on we would propose an ontology, which is like classification for Humanities Computing.

Me: very interesting and complex… so did you get a grade for that?

Yes. It was a very intense class, because it was so much work, we had no idea what the project was because it was not very well planned, it was just an idea, "we want a wiki for Humanities Computing". And then… so what… how do we want the wiki to be? and then we would started looking at this user tagging options and blah, blah, blah and… And at the end we had to justify why this tags were important for Humanities Computing as a discipline. OK, so… we at the end uh said… well, these tags will help us

defining our discipline and understanding what are the interests in different universities."

- *"I have done a couple of projects, not too very many that involves… and…OK, well, I've used Wikipedia a lot but I've haven't actually contributed to Wikipedia [Laughs]. I first tried to build my own wiki when I started HUCO to keep track of all of my notes and flies from classes, so it was a personal wiki. And then I did a project for a HUCO class, which was to propose a Humanities Computing wiki for all the researchers and people interested in Humanities Computing. So what we did was to write a grant proposal. And for Technology class, one of my library classes hum… we had to research a library technology and write a wiki on it that had certain amount of articles and then new articles that couldn't find in like Wikipedia or whatever on that topic."*

- *"Before I started my work in computing science department, I use wiki in general research, mostly I use Wikipedia for finding information, after I started my work… now I'm using the wiki more specifically, mostly I use wiki to collaborate with others students and also with my professors.*
Me: how many people do you have in your team?
I'm not sure because I'm now working with more or less 12 guys and my supervisor, I think there are maybe15.
Me: what kind of things do you write there?
Hum… I usually write weekly… my weekly progress, every week or sometimes my daily progress in my projects, I add some articles every week. And this part I have to say what I have done in that week. And my supervisor adds some comments… I read them and get feedback from her. Just for adding some comments and reading some comments.
Me: do you ever read the work from your partners?
MH: Yes. I go in my partner's wiki and read her comments because there were some things that I didn't know and I went to read her past comments [notes]."

- *"Wikipedia mainly as a reference material, and Wetpaint as a collaborative medium.*
Me: And what is your experience in a research team? How do you communicate with your supervisor?
Through Wetpaint, so… face to face, of course and then uh we use Wetpaint as brainstorming and collaborative medium. [With two co-supervisors]
Me: Have you collaborated to Wikipedia?

I have, to see how it worked. Hum… a couple of articles on geography I was collaborating on. It was fun, it was interesting actually, and it was very interesting.
Me: How were you collaborating?
Adding and editing, so hum the urban section on one of the geography sites needed material and I was just putting my thoughts down and try to synthesize what other people had written."

- *"Right now I think I would be just a user in Wikipedia, but before that I was part of a program in Hermosillo in University of Sonora where we used to teach fresh men about computer's use and internet… all these things. And we got to make one wiki that held all the information about our state, so we coordinated a university-wide [unclear] to create this wiki and have the students gathering information about their city of origin and surrounding areas.*
Me: so it was just one article?
No. no, it was the whole wiki, we created the whole wiki.
Me: did you participate programming that? Or what was your roll?
Well, we actually didn't program the whole wiki, we got one of those off the shelf wiki environments. We did, however, some modifications to it, so that the students wouldn't have much problem logging into it, and creating accounts. Which was paired up with a [Joongla] system, so they would use the same account and we set it up the way they would gather the information and put it into wiki… if that makes sense…?
Me: so your roll was more as an administrator?
In the technology department? Yes. But I was also teaching that course, and I was also telling them how to gather information… respecting the copyrights of certain pictures or text in the internet.
Me: you did the curriculum for that class… assignment?
we just got together and we went through how wikis may manage those things, and we tried to oversimplify it, because most of the students didn't know how to use a computer for example. So we were trying to introduce them to more fundamental aspects of the wiki.
Me: how was it at the end?
Well the first time that we tried it, it wasn't… the result in wiki wasn't as good as we had expected it to be. But it was again our inexperience on to how to use such a big wiki, but in the following courses instead of recreating [do it again from scratch] the wiki, we changed it so the students would have to input more information, do some more collaboration, fixing errors, shaping that a little bit."