Visualization of Interactions in Collaborative Writing

Robert P. BIUK-AGHAI¹, Christopher KELEN² and Hari VENKATESAN³

¹ Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau, e-mail: robertb@umac.mo
²,³ Department of English, Faculty of Social Sciences and Humanities, University of Macau, Macau, e-mail: (kitkelen,hariv)@umac.mo

Abstract—Wikis have become an important component of a collaboration infrastructure, particularly in loosely-coupled and self-organizing settings such as those of digital ecosystems. We report on our use of wikis in the education domain to support collaborative creative writing, as well as collaborative translation. This paper presents an analysis and visualization tool that we have developed as an aid for assessing both the process and the outcome of these collaborative writing tasks.

Index Terms—information visualization, collaborative writing, assessment, digital ecosystems.

I. INTRODUCTION

Digital ecosystems have emerged as a new way of conceptualizing and organizing loosely coupled, flexible, demand-driven interactive environments [2], and are mainly being applied to the business, education and health domains [6]. They comprise a diverse array of software components. As collaboration is a key element of digital ecosystems, collaboration software thus is among their key components.

Among these, wikis have become one of the widespread types of collaboration systems to assist collaborative knowledge production. The wiki concept was created in 1995 by Ward Cunningham and first implemented in the WikiWikiWeb system [11]. A wiki is a web site whose pages can be read and easily edited by anyone, although different implementations of the wiki concept may impose restrictions on reading and/or modifying wiki pages. Their practical foundation of user-as-editor has led to the emergence of practices such that the resulting complex socio-technical system can be regarded as an ecosystem in its own right. In this view, different groups of wiki users can be regarded as different species in the ecosystem. Typical groups include writers, who are the main contributors of content; editors, who edit content to conform to established formats; and tinkerers, who make minor corrections, improvements and clean-up to the content provided by writers. The co-existence of such different “species” within the same wiki site displays the four essential aspects of ecosystems [2]:

1. Interaction and engagement between different species for the overall good of the entire ecosystem (i.e. the entire wiki site).
2. Balance, through directing attention to where it is most needed, which is naturally maintained as individuals of each species perceive and attend to these needs.
3. Domain clustered and loosely coupled, as individuals of their own volition assume the roles and join the species they best identify with, without external pressure.
4. Self-organisation, as individuals are empowered to coordinate in order to take the necessary action demanded by a particular situation.

The underlying principles of wiki systems are thus congruent with those of digital ecosystems. The Wikipedia online user-contributed encyclopaedia (www.wikipedia.org) is perhaps the best known and most successful example of a wiki system, as it now contains millions of articles, many of them of exceptionally high quality (but also some of exceptionally poor quality), in over 250 languages. Wikis are nowadays being used in practically every application domain, both in the form of public sites (such as in the case of Wikipedia) and private sites (i.e. internal sites within an organization, be it a traditional or a virtual one).

Given their self-organising nature that lacks formal editorial procedures such as are common in traditional publishing processes, it may be difficult to assess both process and outcome of contributions made in a wiki system. Assessment of contribution is of particular interest in the education domain, where student’s contributions need to be assessed, but also in the business domain where assessment of contribution may be linked to reward structures. Assessment thus is of relevance to both digital education and digital business ecosystems. Within this context we have experimented with assessing the use of Wikis at our university. Over the past three years, we have employed wikis in two different domains: creative writing and translation. Our study has focused on assessing both the collaborative writing processes employed, as well as the outcomes of those processes. For this purpose we have developed analysis and visualization software that enables us to discern patterns of interaction among the collaborators. As this software is generic in nature it is not limited to the education domain, and can be applied to wiki data in all types of digital ecosystems.

The structure of the remainder of this paper is as follows: Section II briefly reviews wiki use in education. In Section III we present our experiences of using wikis in collaborative writing, and in Section IV we present our visualization tool. Finally, in Section V we give conclusions to this paper.

II. WIKIS IN EDUCATION

In the field of education, the past few years have witnessed increased adoption of wikis. Generally they are being used in two principal ways: as knowledge repositories for instructors making lecture material and course information available [14]; and for the collaborative production of con-
tent by students, such as assignments, projects, essays and other assessment material [4],[12]. Examples of the diverse uses which wikis were put to include: collaborative construction of encyclopedia entries by upper secondary school students [12] and university students [4]; collaborative creation of course content [16] or a shared artefact [8] by university students; project management in a project-oriented computer science course [18]; project-oriented product development in an online graduate course [10]; essay writing in a university ESL course [17], and for both short and semester-long assignments in a graduate course [3].

The wiki concept as developed by Cunningham is an open one, where anyone can access and modify any page. This is for example largely the way that the MediaWiki system (www.mediawiki.org) underlying the Wikipedia site works – although there are facilities for protecting pages from editing. For educational purposes some degree of access control is usually necessary, e.g. to protect pages that should only be editable by instructors but not by students. Moreover, in order to allow identification of contributions and to prevent changes by outsiders, anonymous editing may need to be disabled and students may be required to login [13]. It has also been observed that a single tool integrating all required functions for communication, project management and authoring is preferable to a set of separate tools [10]. Basic wiki technology has thus been extended with several different functions specifically for use in education: protecting and/or locking pages, creating read-only snapshots of an entire wiki site, and others [7],[16].

Following this brief review of the use of wikis in education, next we report on our experiences of using wikis.

III. OUR USE OF WIKIS IN COLLABORATIVE WRITING

At the University of Macau we have experimented with the use of wiki systems to support collaborative writing work by students since early 2005. In the process we have tried two different wiki systems: the eGroupWare Wiki application, and the MediaWiki system; as well as experimented with an email list as an alternative to wikis. Because of limitations we have experienced with the other systems, we currently use a MediaWiki-powered system.

We have used Wikis in two different settings: for collaborative creative writing of stories; and for collaborative translation of texts between English and Chinese. Both involved small groups of students (usually between three and eight, but in some cases more) working together over a period ranging from one week to a whole semester to collaboratively create a piece of text.

A. Wikis in Creative Writing

Creative writing refers to the imaginative production of text, in our case of fictional stories. It is a subject which traditionally assesses most work on an individual student basis, in this way imitating an “industry standard” of production (publication and performance) of works by individual authors. Importantly though, and increasingly, many real world creative writing applications do call for collaborative story production (or the collaborative production of other creative text types). This is especially true today of production of texts for screen or stage, committee reporting work and translation.

To support creative writing, we set up the “CoW wiki” site (“CoW” being short for “Collaborative Writing”). Although particular interactions through the wiki system has varied through the imposition of different rule-sets (for instance with regard to turn-taking, initiation and control/“ownership” of text), essentially the rules have been kept as straightforward as possible. Students were instructed to conduct all discussion related to the collaborative writing effort through the tools provided (i.e. online and not face-to-face or by means of other technology which would leave us no digital record, e.g. telephone, instant messenger, etc). The objective has in general been to allow the freest possible interaction, giving the maximum scope to spontaneous and imaginative interaction.

Student writing in the wiki consisted of two “genres” – one being the writing of the story itself, the other being a meta-discursive thread “about the story”. The former allows students to compose directly and spontaneously, the latter allows for reflection and forward planning. In our CoW wiki system, being based on MediaWiki, these two threads were placed in the familiar “article” and “discussion” tabs at the top of each page, corresponding to the writing and discussion threads, respectively. From the text production point of view, these simple tools are adequate for the collaborative story writing strategies thus far employed.

B. Wikis in Translation

Following our positive experiences with using wikis in creative writing, we expanded its use to translation. We set up a customized version of MediaWiki, named “TransWiki”, to facilitate translation pedagogy and learning over three semesters for third year translation students at the University of Macau. We observed that the quality of translations produced by students improved quite tangibly when working collaboratively through the wiki. Collaborative learning with wiki as a platform also served as a useful pedagogical tool for teachers of the subject.

As TransWiki is a customized version of MediaWiki, every wiki page has an associated discussion page accessible through the “discussion” label at the top of the page. Students were advised to use this for discussions regarding translations. Typically source texts were posted on TransWiki with links for different groups of students. Each link led to a page that members of each group could edit. A deadline for submission was specified in each case. Students were required to conduct discussions regarding their translation only on the discussion page. Before the deadline, each group had to post their final version on their group’s wiki page. Translations from each group were then compared and discussed in class.

Each group of students was formed such that the maximum possible variety of views and opinions would be available for discussion. Local Macau students and mainland
Chinese students were put in the same group wherever possible, as were students originating from high schools employing different mediums of instruction.

Evaluation of translations was based on two grades with equal weight. One was the quality of the translation produced (same for the entire group) and the other was the level of participation in discussions (different for each individual). This provided an incentive to students to participate actively in discussions.

C. Questions

In our use of wikis in the two domains described above, we were interested in knowing more about the patterns of interaction that occurred in the production of collaboratively written texts. Given that the text production happened as an assessable contribution of each student, yet being part of a larger piece of work involving several others with varying number and volumes of contributed text, one question that arose early was how to assess the final contribution. That is:

Q1: How much has each student contributed to the final product?

Because of the open nature of the writing process as it is conducted through a wiki system, where anyone can add their own contributions as well as modify and delete the contributions of others, assessing the contributions of an individual student is not necessarily a simple matter. For instance, a student may have contributed much text early, which however through subsequent editing by others may have all but disappeared in the final version. What should count in assessing this student’s contribution? If only the final result counts then a potentially large volume of text contributed earlier would be disregarded, raising concerns of fairness. On the other hand it could be argued that the text was removed for a reason, presumably because it did not live up to the quality requirements of later editors. Yet even a contribution that was later removed could have contained ideas and concepts that later sparked contributions of others, thus being responsible to some degree for those later contributions. Another concern is whether only contributions to the actual text produced should be considered in assessment, or also contributions to the discussion. It is possible that a student has actively contributed to the discussion accompanying a text, helping clarify goals and directions or infusing the subsequent writing process with ideas that may in no small part have had an influence on the final outcome. Thus clearly contributions to discussions are not to be discounted in assessing an individual’s contributions. In the following section we show how this complex issue is addressed.

Another concern in studying the use of wikis in collaborative writing is to assess the ways students work, or fail to work, together. Are there students who dominate the writing process or is participation more egalitarian? Is the process sluggish and intermittent or smooth and continuous? Moreover, do students reflect on the writing process through discussions or is the entire interaction based only on the writing? That is:

Q2: What is the process of collaboration?

Finally, we are interested to assess the depth of the collaboration. By depth we mean the extent to which students engage with the writing of others. Where depth of collaboration is low, students make individual contributions to the jointly authored text, but refrain from altering contributions of others. The final outcome is a patchwork of individually contributed sentences and paragraphs. However, where depth of collaboration is high, students engage intensively with each other’s contributions and do not hesitate to alter or delete contributions of others. In the final outcome a single sentence may consist of contributions from multiple students. The question thus is:

Q3: What is the depth of collaboration?

In order to address these questions we have decided to employ information visualization. This is introduced in detail in the following section.

IV. VISUALIZATION OF COLLABORATIVE WRITING

To obtain answers to the above questions is not very easy when faced solely with the wiki texts alone. Although the wiki system provides a history of each article and discussion page, which makes it possible to see who wrote what and when, it is a tedious effort to review a history of dozens or hundreds of versions and piece together an understanding of the writing process that can help answer the above questions. To assist in this effort we have developed an information visualization system that retrieves the entire version history of any article or discussion directly from the live wiki database and represents it in graphical form. This graphical representation allows the user to get “the big picture” of the entire writing process involved in an article, but also details of individual steps of this process, and thereby facilitates the assimilation of the related information. The operational wiki system interacts with its database which stores the latest version of an article as well as all previous versions. Our visualization program reads data from this database and displays it in a suitable visual form, using text and graphics.

To obtain different insights into the evolution of an article and the writing process, our visualization program provides several main types of visualizations:

Text comparison: display of the differences between two consecutive versions. Differences are colour-coded.

Text authors: display of the authors of each word in a version of an article. Different authors are colour-coded.

Analysis graph (all users): display the evolution of an article, as well as its accompanying discussion, of all users over the entire history of an article.

Analysis graph (single user): display the contributions of a specific user to an article, as well as its accompanying discussion, over the entire history of an article, showing both total accumulated contributions and remaining contributions after editing.

Contribution summary graph: display a summary of all contributions made by all users, considering amount and/or number of times of contributions to article and/or discussion.

User participation graph: display a graph showing the history of a user’s contributions to all the articles they were involved in, and the pattern of switching between articles.
By using one or more of these visualizations we are able to find answers to the questions about patterns of interaction that we presented in the previous section. Below we show examples of their use for each of our questions.

A. Contributions per Student

The question here was: “How much has each student contributed to the final product?” In order to assess a student’s contribution, we consider both contributions to the article itself as well as contributions to its accompanying discussion, as activity in the discussion can provide a constructive influence on the development of the article. Moreover, we consider both total amount of contributions and the number of contributions made. The total amount is a count of the characters of text in the article and discussion. The number of contributions reflects the frequency of contributions. By taking this factor into consideration we not only take contributions to the final end-product (i.e. the article) into account, but also contributions to the process of reaching that end-product. Fig. 1 shows a contribution summary graph for the article “Alice’s story” in the CoW wiki. It shows that a total of 12 users were involved in the creation of the article and shows for each user the percentage of contribution. It is clearly visible that some users have contributed significantly more than others. For example, the three largest contributions are, respectively 25%, 20%, and 18% of the total, whereas the four lowest contributions each amount to only 1% of the total. This would have been difficult to perceive if consulting the wiki texts directly.

The calculation of the contribution percentage was performed according to this formula:

\[
\frac{\alpha \times (aca_u / aca) + \beta \times (dca_u / dca) + \gamma \times (acn_u / acn) + \delta \times (dcn_u / dcn) }{100}
\]

where:

\[
aca_u \quad \text{is article contribution amount of user } u
\]

\[
dca_u \quad \text{is discussion contribution amount of user } u
\]

\[
acn_u \quad \text{is number of article contributions of user } u
\]

\[
dcn_u \quad \text{is number of discussion contributions of user } u
\]

\[
\alpha, \beta, \gamma, \delta \quad \text{are weights where } \alpha + \beta + \gamma + \delta = 1.0
\]

We use a default value of 0.25 for the weights \(\alpha, \beta, \gamma, \delta\), i.e., we give equal weight to both the amount of contributions and number of contributions, for both articles and discussions. However, this is configurable and different weights can be used to emphasize different contribution factors.

Currently our assessment of contribution makes a purely quantitative evaluation of amount and number of times of contribution, and does not make any qualitative assessment of contributed content. However, recent work on determining reliability [5], quality [9] and trust [19] of Wikipedia articles has demonstrated that a quantitative, and thus automated, approach can be applied to assessing article content. We plan to integrate such methods in the next version of our analysis and visualization tool.

B. Process of Collaboration

The second question we were interested in answering was: “What is the process of collaboration?” Specifically, we were interested to know how the collaboration progressed over time, how students worked together in producing the final outcome, and how this differed among different groups of students.

The analysis graph (all users) is a visualization which helps show the overall evolution of an article, and the accompanying discussion. Fig. 2 is an example of an analysis graph for the article “Alice’s story”. On the x axis is time, ranging over the entire history of this article, and on the y axis is the size of the article in characters. The blue curve shows the growth of the article and makes clearly visible different patterns of growth. For example, the steep sections of the curve shortly after the beginning and shortly before the end indicate periods of intense work on the article, whereas the long period of a relatively slow increase in between those two periods indicates a period of more sluggish growth. The black dots overlaid on the blue line further confirm this pattern of growth. Each black dot corresponds to a single contribution. Thus it is visible in the two steep sections that not only were large gains made in the volume of writing, but also many contributions were marked in a relatively short period of time. Conversely, in the period in between the number of contributions was small and spaced much further apart. This pattern is typical of work performed...
over an extended period and with a fixed deadline. At the beginning there is enthusiasm to start work, which drives progress forward until some obstacles are faced and enthusiasm wanes. Only when the deadline nears, work resumes with renewed urgency until completion.

The analysis graph also shows the discussion that accompanies the evolution of the text as the green vertical lines on top of the time axis. In the case of Fig. 2 there are only few discussion messages posted and thus almost the entire writing process is performed with minimal reflection on the process, i.e. as far as interaction through the wiki system is concerned (unless, of course, this interaction took place outside the system, such as face-to-face or through other means, which the students were told to refrain from and which we have no knowledge of).

One further insight into the collaborative aspect of the writing process can be obtained from an analysis of the contributions of different users. Fig. 3 shows an example of a text author analysis view. Different portions of the text are colour-coded according to their authors. It clearly makes the different authorship of the text visible. This example also shows that one of the authors, SidraWong (represented by the steel blue coloured text), appears to be dominating the writing.

Another visualization which gives some insight into each student’s participation in the writing process is the user participation graph. Fig. 4 shows two user participation graphs for the students Alice and Graffarn. These graphs show time on the x axis and a list of stories to which the user has contributed on the y axis. A black dot indicates a contribution by this user, and a red line connects two consecutive contribution dots. The usefulness of this graph is to see the pattern of contribution to one or more articles. In the case of Alice (Fig. 4, top) the zigzag pattern indicates that this user, was concurrently actively contributing to three articles, as most contributions to one article were followed by contributions to a different article. The other student, Graffarn (Fig. 4, bottom), however, initially switched contributing from one article to another but very quickly settled on one article to which he contributed for the rest of the time. Alice can thus be characterized as a “multi-tasking” author whereas Graffarn is much more a “single-minded” author.

C. Depth of Collaboration

To address the question “What is the depth of collaboration?” we also make use of author text analysis. Depth of collaboration, i.e. the extent to which students engage with the writing of others, can be seen by the pattern of text that emerges from a period of collaborative writing. In Fig. 3 we saw one example of an author text analysis in which very little engagement with the writing of others took place. With only one exception in the first paragraph where one student changed four consecutive words in the writing of another student, the remainder of the text here has remained intact and each contributor simply added text to the existing one. Fig. 5 shows part of a story in which the depth of collaboration is much higher. Within the six lines of text, five students contributed, with as many as three students within the same sentence.

Another visualization that helps perceive the extent to which students engage with each other’s writing is the analysis graph for a single user. Fig. 6 shows an analysis graph for the user Graffarn in the article “Alice’s story”. The graph is similar to the analysis graph for all users, with an x axis representing time, and a y axis representing the size of the text in characters. However, there are two lines: the red line shows the cumulative contributions of this user whereas the blue line shows the remaining contributions after editing. Dots, as before, show instances of editing: cyan dots represent edits by the user him/herself, black dots represent edits by others. If no text was ever deleted, the blue and red lines should overlap. However, if editing has taken place and removed text, the blue line will fall below the red one. The extent to which this has happened can be seen by the distance...
tance between the lines, and the number of times this has happened can be seen by the number of times the blue line drops below its current level. Moreover, the extent to which other users have edited this user’s text can be seen by the number of times the blue line has dropped following a black edit dot. In Fig. 6 we can see 12 cyan dots, i.e. edits done by this user, and we can observe that in all cases the text has grown as a result. On the other hand, we can see that the text has shrunk on 14 occasions, and that all of these were edits by other users. That is, the user himself has never removed any of his own text, but others have engaged with his text and deleted from it more often than he edited his own text. This is a general observation we made – users tend to delete from the text of others far more often than deleting from their own text. Finally only about ¾ of the originally contributed text remains.

V. CONCLUSIONS

In this paper we have presented our experiences of using information visualization to help assess interactions in collaborative writing through wikis. While the examples presented were taken from the education domain, the analysis and visualization is not domain specific but rather generic to wikis in general and thus applicable in a wide range of domains. Moreover, our assessment tools are particularly well-suited for use in loosely coupled and self-organizing settings such as are typical of digital ecosystems, be they education- or business-related.

Related work on wiki visualization exists but has had a different focus: history flow visualizations [15] have a single document as the focus, WikiTrails context overview visualization [16] focuses on paths through a collection of linked documents, and WikiVis co-author network visualization [1] focuses on making relationships between documents visible. None of these is particularly well-suited to the needs of assessing author contributions and making the kinds of process characteristics visible that we do.

VI. REFERENCES


