IntelWiki: Recommending Resources to Help Users Contribute to Wikipedia

Mohammad Noor Nawaz, Andrea Bunt

Department of Computer Science, University of Manitoba, Winnipeg, Manitoba, Canada
{chowdmnn, bunt}@cs.umanitoba.ca

Abstract. We describe an approach to facilitating user-generated content within the context of Wikipedia. Our approach, embedded in the *IntelWiki* prototype, aims to make it easier for users to create or enhance the free-form text in Wikipedia articles by: i) recommending potential reference materials, ii) drawing the users' attention to key aspects of the recommendations, and iii) allowing users to consult the recommended materials in context. A laboratory evaluation with 16 novice Wikipedia editors revealed that, in comparison to the default Wikipedia design, IntelWiki's approach has positive impacts on editing quantity and quality, and perceived mental load.

Keywords: User-Generated Content, Recommendations, Wikipedia.

1 Introduction

User-generated content (UGC) is content generated by people who voluntarily contribute data, information, articles, or media on the web. Despite the explosion of UGC in recent years, the percentage of the population that contributes content tends to remain relatively small. Most community content follows the "1% rule", where approximately 1% of internet users create content, 9% enhance it, and the remaining 90% simply consume it [5], [12]. This participation imbalance is a concern for a number of reasons, including both the amount of work required of contributors to uphold content standards and a potential underrepresentation of the views and interests of a large percentage of the population [12].

While there are many factors that influence participation rates, including community politics [19], a significant barrier to participation is simply the amount of effort required to do so. In particular, in his article on participation inequity in UGC, Nielsen's number one suggestion on how to increase participation rates is: "Make it easier" [12]. This assertion is supported by studies indicating that editing effort can indeed affect participation rates [2], [8], [20].

In this paper, we propose an approach for facilitating contributions to Wikipedia, one of the most widely accessed forms of user-generated content. Like other community content repositories, only a small percentage of Wikipedia users contribute content. For example, in September 2013, Wikipedia had over 500 million

unique visitors; however, only 0.05% of these visitors made at least one edit and only 0.015% were considered "active contributors" (i.e., with five or more edits) [18], [23].

Prior work suggests that an attribute of Wikipedia articles that makes them particularly difficult to edit in relation to some other forms of UGC (e.g., movie reviews) is the need for background research [20]. To address this challenge, our approach provides users with streamlined access to recommended reference materials -- recommendations that are personalized to the individual article. To illustrate our approach, we designed and implemented the *IntelWiki* prototype, which automatically generates resource recommendations, ranks the references based on the occurrence of salient keywords, and allows users to interact with the recommended references within the Wikipedia editor. A second contribution of this work is a formal laboratory evaluation exploring the potential for our approach to ease the editing burden in comparison to the default Wikipedia editor. Our results indicate that having streamlined access to resource recommendations increased the amount of text participants were able to produce (with time held constant) and that this text was both more complete and more accurate than when using the default editor. Participants also reported experiencing significantly lower mental workload and preferred the new design.

2 Related Work

Prior to describing our approach and its evaluation, we begin by overviewing related work. User-generated content in general and Wikipedia in particular, has been a widely studied phenomenon, including studies on what motivates contributions (e.g, [2], [13]), how editing roles evolve over time (e.g., [17]), and statistical analyses of Wikipedia data (e.g., [9]). We focus our coverage on two areas: systems designed to improve Wikipedia articles, either through completely automated means or by helping potential editors, and systems for helping people choose their editing tasks.

2.1 Enhancing the Text of Wikipedia Articles

The content of Wikipedia articles, and other similar UGC environments, can often be classified into two primary forms: 1) content that is structured, and 2) free-form content. Structured information has a pre-defined schema, such as the information found in a standard Wikipedia article's infobox (see Fig. 1, left). The bodies of the articles contain free-form content, including prose, images, links and references.

A notable example of improving structured Wikipedia content is the Kylin system, which automates the process of creating and completing Wikipedia article infoboxes (e.g., [8], [22]). An evaluation of a mixed-initiative version of Kylin revealed that recommending potential changes to the infoxboxes had positive impacts on both user contribution rates and infobox accuracy [8]. Sharing some similarities with our approach, Weld *et al.* proposed an extension to the system, where the information extraction used to improve the infoboxes is extended beyond Wikipedia articles to the general web [16]. As in our approach, this extension relied on web queries to find

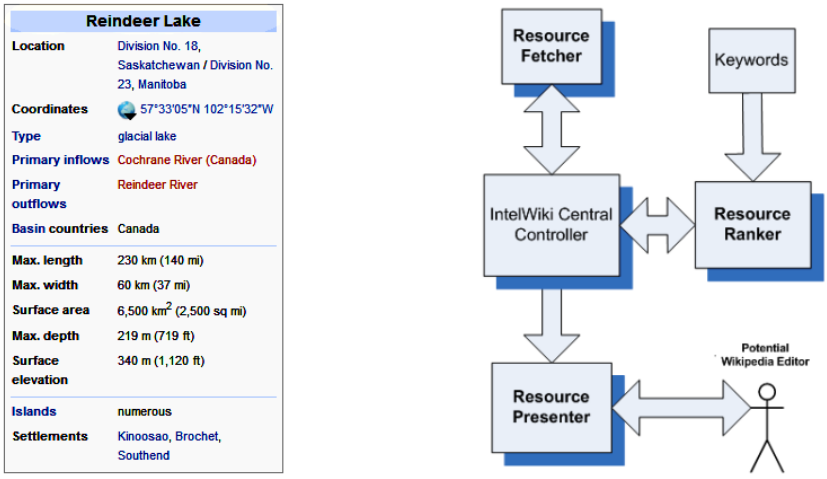


Fig. 1. (Left) Infobox for the Lake category (Right) The IntelWiki system architecture

useful resources, however, these resources were used by the learning algorithm only as opposed to presented to potential editors. As another example targeted at improving structured Wikipedia content, the WiGipedia tool helps users identify and correct inconsistencies among structured data spread across different articles [1].

Our work focuses on supporting edits to free-form Wikipedia text. In contrast to our approach, which aims to support human editors, most prior work in this area has tried to fully automate the process. For example, Okuoka *et al.*'s system links Wikipedia entries on news events with relevant videos from external sources [14]. WikiSimple takes Wikipedia articles as input and automatically produces articles re-written in simpler grammatical style (to enhance readability) [21]. Finally, Sauper *et al.* proposed a fully automated process for generating a multi-sectioned Wikipedia article [15]. Approaches that leverage human expertise have also been explored, but in the context of corporate wikis, where the focus has been on tools to support knowledge transfer from e-mails to wikis (e.g., [6], [10]).

3 IntelWiki Prototype

Our approach to facilitating user enhancements to free-form text in Wikipedia articles is to help editors locate and interact with relevant Web-based reference materials through article-tailored resource recommendations. To illustrate our proposed approach, we designed and implemented the *IntelWiki* prototype, which recommends pertinent resources to the user and streamlines the process of interacting with these recommended resources. In this section we overview the three main components in IntelWiki's framework (see Fig. 1, right): i) the *Resource Fetcher*, ii) the *Resource Ranker*, and iii) the *Resource Presenter*.

3.1 Resource Fetcher

IntelWiki's Resource Fetcher searches the web for resource material that could help a potential editor enhance a given Wikipedia article. To do so, IntelWiki uses Google's Custom Search Engine (CSE) API, submitting the article title as a search query. From the returned results, the Resource Fetcher then selects the top 60 pages (a configurable parameter) to submit to the Resource Ranker for further processing. From the set returned by the Google CSE, the Resource Fetcher removes any dead links or links to pages that are not easily machine readable (e.g., consist of solely images). These latter types of pages were removed as a simplification for this proof-of-concept prototype -- one could imagine extending this candidate set by embedding more sophisticated document processing capabilities within the system.

3.2 Resource Ranker

The Resource Ranker's role is to assess the suitability of each candidate resource, information that is then used by the Resource Presenter (described next) to emphasize the most promising resources. The Resource Ranker's assessment of suitability involves calculating a relevance score for each resource based on the number of occurrences of "pertinent keywords" within the resource. These relevance scores are then used to re-rank the resources from the ordering initially returned by the Google CSE.

By default, the Resource Ranker uses the article's infobox schema attributes as the set of pertinent keywords. Through experimenting with different article categories, we found that using the complete set of infobox attributes as the pertinent keywords typically provided a more personalized resource ranking than the Google CSE default ranking; however, we also noted the potential for improvement by using a widened set of keywords. Potential additions that we have found to improve rankings include: attribute synonyms, root words, and parts of speech variants, as well as units of measurement. Therefore, IntelWiki allows additional keywords to be specified on a per-category basis (articles in Wikipedia are grouped hierarchically according to category). We envision these tailored lists of keywords could be generated by a Wikipedia administrator, through crowdsourcing techniques, or by training the system to learn pertinent keywords from other (more complete) articles of the same category.

In addition to sharing common infobox schemas, articles in a given category are often very similar in structure. For example, articles in the "Lake" category typically contain sections describing Geography, Climate, History, Ecology and Geology, among others. Therefore, IntelWiki's Resource Ranker has the capability to leverage a keyword-to-section mapping, should one exist, to personalize its ranking of the resources based on the section the user is currently editing. Similar to the set of pertinent keywords, a keyword-to-section mapping could be defined on a per-category basis by a Wikipedia administrator, through crowdsourcing techniques, or through machine learning techniques.

Fig. 2. (Left) IntelWiki's callout. Clicking on "reference materials" will display recommended resources. (Right) A tooltip showing the occurrences of pertinent keywords within the resource.

3.3 Resource Presenter

The IntelWiki system's Resource Presenter makes the set of suggested resources available to a potential editor on demand (see Fig. 2, left). As shown in Fig. 3, when a potential editor asks to view the reference materials, the system adds two additional panes to the regular Wikipedia interface in both viewing and editing modes (shrinking the article to make room). The first is a "Suggested Resources" Pane, which lists the recommended resources. The second is a "Resource Viewer" Pane, which allows users to inspect and consult individual resources in place.

To promote the references that the system believes will be most helpful to the editing task, the Resource Presenter sorts the recommended resources using the relevance scores calculated by the Resource Ranker. Initially (or whenever the user is in the view mode) an article's recommended resources are sorted according to the per-article relevance scores. When the user goes to edit a particular section (i.e., in the edit mode), the list of suggested resources is reordered based on the section-specific relevance scores, if a keyword-to-section mapping exists for the article's category.

The system tries to further support resource selection in two ways. First, it displays the resources' relative relevance assessments (see the green bars in Fig. 2, right and Fig. 3) allowing the users to see which ones the system believes will be most useful. Second, to allow for additional inspection without having to open the resource, when the user hovers over a particular resource, the system displays a tooltip consisting of the keywords found in the resource and their respective frequencies (see Fig. 2, right).

The user can view the contents of a particular resource by either clicking on it or dragging it to the Resource Viewer Pane. To help the users locate relevant information within the resource, the system highlights all occurrences of the pertinent keywords within the resource (as shown in Fig. 3). In our initial design, we experimented with multiple resource viewer panes (up to four); however, pilot

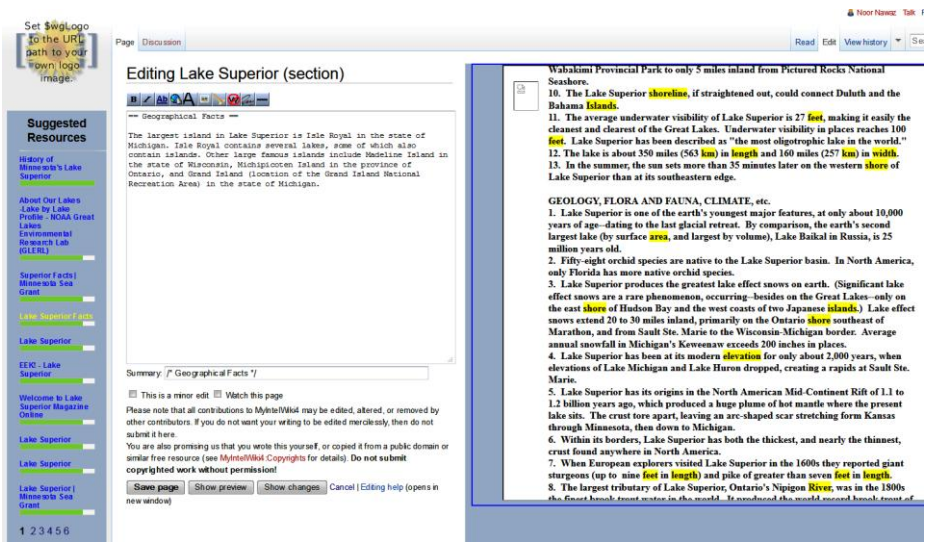


Fig. 3. Editing with IntelWiki Interface, with the "Suggested Resources" Pane (left) and the "Resource Viewer" Pane (right).

participants felt that they consumed too much screen real-estate and were difficult to manage.

4 Evaluation

We conducted a formal laboratory study comparing the IntelWiki system described above to the default Wikipedia editor. The goal of the study was to explore if the IntelWiki system could make it easier for users to edit Wikipedia articles. We leave an assessment of recommendation quality to future work.

4.1 Participants

Sixteen participants completed the study (6 females, mean age 24.4), recruited through on-campus advertising. To ensure access to a wide enough pool, we did not screen according to previous Wikipedia editing experience. Our pre-study questionnaire revealed that all participants were regular Wikipedia visitors, but none had previous Wikipedia editing experience. Participants were provided with a $15 honorarium.

4.2 Design

Interface Type was the primary within-subjects factor with two levels:

1. **IntelWiki:** The complete IntelWiki system described in previous section.

2. **Default:** The Wikipedia Edit Interface plus the Google Search Engine.

Participants completed one task with each interface type (described in the next section). Therefore, task was a within-subjects control variable. Interface and task order were fully counterbalanced to account for potential learning effects.

4.3 Tasks and Procedure

After completing a demographics questionnaire, participants edited the "Geography" section in two articles on well-known lakes (one per condition). From these articles, we removed most of the content in the "Geography" section, leaving only three lines to provide initial guidance. We also removed the articles' infoboxes since they were populated with facts from the original Geography sections. Participants were provided with a list of example of attributes (using geography-related attributes from the infoboxes), but were told to edit the sections as they saw fit. To discourage direct plagiarism, we disabled copying and pasting.

Participants were asked to write the best Geography section that they could within the 25 minutes (i.e., editing time was fixed across all participants). Prior to editing, participants were briefly introduced to the interface in that condition, and completed a short practice task. Immediately following each condition, participants completed a NASA-Task Load Index (TLX) questionnaire [7] to measure their perceived mental workload. The experiment concluded with a post-session questionnaire and a short semi-structured interview. Each session lasted between 75-90 minutes.

In the IntelWiki condition, the system retrieved and assessed the recommended resources using a set of section-specific keywords related to "Geography", which consisted of the relevant infobox attributes and their units of measurement.

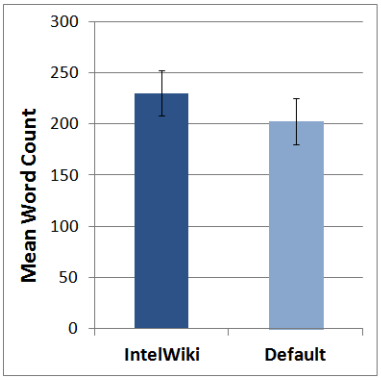
4.4 Results

In the analysis below, quantitative dependent measures were analyzed using a Repeated-Measures ANOVA with Interface Type (IntelWiki, Default) as the within-subjects factor. To check for asymmetric learning effects between two conditions, we also included Interface Order (IntelWiki_First, IntelWiki_Second) as a between-subjects factor in the analysis. Error bars on all graphs depict standard error.

Text Volume and Completeness.

Since editing time was fixed, we begin by examining text volume. Fig. 4(left) shows that participants wrote significantly more words with IntelWiki (229.9, s.e. 22.7) than with Default (202.8, s.e. 22.4; $F_{1,14} = 5.302$, $p = 0.037$, $\eta^2 = 0.275$).

We analyzed two measures of text completeness by having the first author code the text participants generated for: i) the number of facts described (Fact Count) and ii) the number of facts accurately described (Fact Accuracy). Any distinct piece of information was counted as a fact. A fact was coded as accurate if it i) was related to the topic of the section, and ii) was accurately reported (judged using the original infobox or article when possible, or the participant's source).



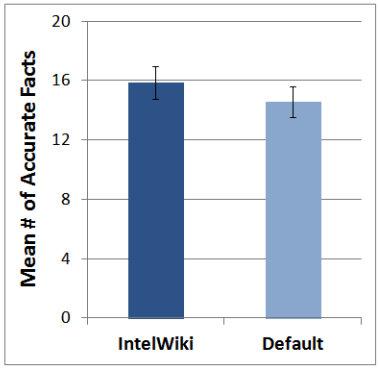


Fig. 4. (Left) Word count by condition. (Right) Fact accuracy by condition.

As shown in Fig. 4(right) and Fig. 5(left), IntelWiki outperformed Default for both text completeness measures. For Fact Count, participants covered 17.8 (s.e. 1.1) different facts with IntelWiki as compared to 16.2 (s.e. 1.2) with Default ($F_{1,14}$ = 7.304, p = 0.017, η^2 = 0.343). Interestingly, there was also a significant Interface Type * Interface Order interaction effect ($F_{1,14}$ = 6.182, p = 0.026, η^2 = 0.306). As illustrated in Fig. 5(middle), the primary benefit of the IntelWiki system came for those who experienced this condition second. Those who edited with IntelWiki first covered roughly the same number of facts in each condition. We suspect that in this latter case, IntelWiki helped participants learn what types of facts to describe in the first condition, and that they were able to transfer this knowledge to the second editing task, even though the scaffolding was removed. For Fact Accuracy, Fig. 4(right) shows that participants were significantly more accurate with IntelWiki (15.9, s.e. 1.1) than they were with Default (14.6, s.e. 1.1, $F_{1,14}$ = 4.520, p = 0.052, η^2 = 0.244).

Perceived Mental Workload and Subjective Preference.

The results of the NASA-TLX indicate that perceived mental workload was significantly lower (see Fig. 5, right) when using IntelWiki (49.5, s.e. 6.1) than when using the Default interface (66.7, s.e. 3.1, $F_{1,14}$ = 10.212, p = 0.006, η^2 = 0.422). Participants also expressed a preference for its design, with 14 out of the 16 participants preferring the IntelWiki interface over the Default one (χ^2=9.000, p = .003).

Interview Comments.

While the above results suggest that IntelWiki's approach improves editing performance and lowers perceived mental workload, it does not isolate the value of its individual components. Therefore, in the semi-structured exit interviews, we elicited participants' impressions of the IntelWiki system, including what they liked and did not like about its approach.

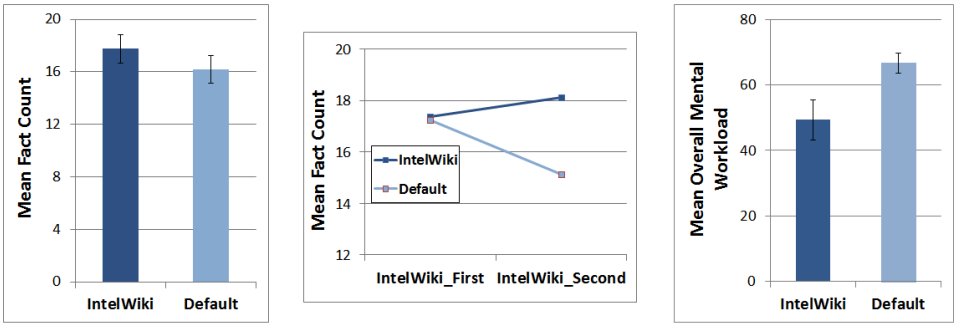


Fig. 5. (Left) Fact count (Middle) The Interface Type * Interface Order interaction effect, and (Right) Cumulative Mental Workload as measured by the NASA TLX.

***Integrating Editing and Background Research*:** For the majority of the users who preferred the IntelWiki system, it was for its ability to integrate the two tasks of background research and article editing. In particular, participants liked the fact that they did not have to switch windows to consult (or search for) reference material, as the following quote illustrates:

> *I preferred [IntelWiki] because the screen was shared. […] It gives you the ability to do two things at the same time: go through what you are going through and still edit what you are editing. – P5*

We note that in the Default condition, participants were able to place the windows in any configuration they wished, with the monitor used (23") providing ample space to place the editing and search interfaces side-by-side. When reviewing the session videotapes we found that six participants chose to place their windows in this configuration. An analysis of their data alone suggests that the value of IntelWiki's approach goes beyond integrated editing and resource viewing. For example, even with this small sample size, the difference in Fact Count remained significant ($p = 0.003$), with a trend in IntelWiki's favour for mental workload ($p = 0.051$).

Supporting Resource Inspection and Evaluation: Participants also liked the ability to inspect the recommendations through the tooltips, indicating that they were able to quickly evaluate the suitability of an individual resource:

> *Even before you open the resource in the viewer pane you know what you are expecting to see. When I am searching online [Google text snippets] show me a plethora of mostly useless information that would not directly give you what you are looking for. – P8*

Similarly, participants appreciated the manner in which IntelWiki's keyword highlighting streamlined their search for key information within an article:

> *[Keyword highlighting] was very helpful; didn't have to read the whole page, or even the paragraph, only the lines containing the highlighted words. – P17*

Replacing Independent Search: Most participants responded positively to the notion of system recommended resources, with many commenting that they were relieved of having to do their own searches. For example:

[IntelWiki] eliminated any need for [additional searches] because, virtually anything that's needed I think was provided in the [recommended resources] – P12

As the following quote illustrates, however, not all participants, however, felt that IntelWiki's recommendations were sufficient:

For most of the information I didn't need [Google]. But when I was looking for the "connected rivers", the "river" keyword was listed, but I did not find any information about connected rivers from that resource. So, I searched through Google. – P14

While participants had the option to supplement the recommended resources with external searches, P14 was the only participant to exercise this option (for a single external section). This suggests that when provided with a good set of pertinent keywords, the system is able to retrieve a useful set of resources. However, the above quote also suggests that allowing users to incorporate their own retrieved resources would be a useful extension to the system.

5 Discussion and Future Work

Our proof-of-concept evaluation provides encouraging evidence in favour of IntelWiki's approach. With editing time fixed, participants contributed significantly more text and experienced significantly lower perceived mental workload in doing so. In terms of text completeness, IntelWiki was particularly helpful for participants who experienced that condition second (i.e., after editing with the Default interface), with results suggesting that IntelWiki helped scaffold the editing process. Participants also expressed a strong preference for IntelWiki's design over the status quo.

Having established potential for the general approach, there are a number of promising directions for future research, one of which is assessing the accuracy of the system's recommendations. For "proof-of-concept" evaluation purposes, IntelWiki was provided with a set of hand-crafted section-specific pertinent keywords to help the system rank the resources. Future work could examine the feasibility of using crowdsourcing or machine learning approaches to generate such a list as well as the impact of list accuracy on the utility of the approach. Further evaluations are also needed to explore the relative utility of IntelWiki's different features. Finally, a field deployment would be necessary in order to explore the impact of IntelWiki's support on contribution rates.

Our decision not to screen for Wikipedia editing experience resulted in a set of participants without any Wikipedia editing experience. While this decision was primarily based on pragmatics, studying our approach with this participant group does align with the motivation of improving overall contribution rates by making it easier for newcomers to contribute. Given that IntelWiki's support is for background research as opposed to for wiki-editing mechanics, there is reason to be optimistic that the findings would generalize beyond novice editors. Similarly, to control for participant expertise while still having access to a wide enough participant pool, participants edited articles on topics that they were familiar with, but not for which they were experts. Therefore, exploring the value of the approach with participants with more article-related expertise is another important area of future study. It would

also be interesting to examine IntelWiki's impact on editing confidence, given Bryant *et al.*'s finding that novice editors initially edit articles on topics only which they are experts in, but eventually branch out as they gain confidence [2].

There are number of ways that the system could be extended to further personalize its recommendations. One promising approach would be to collect implicit and explicit relevance feedback for the recommended resources and to use this feedback to improve future recommendations. For example, one could image favouring resources previously used to edit other articles of the same category. To collect explicit feedback, editors could be allowed to "vote" on the utility of the different resources. For repeat editors, one could also weight the recommendations towards websites or domains that the editor has frequently consulted in the past.

Finally, it would be interesting to explore the generalizability of IntelWiki's resource recommendation strategy to other environments where background research is often required, such as writing articles/blogs in online communities, or writing research papers/essays using word-processing software. The Google Search technique used to fetch relevant resources could be incorporated directly, whereas developing streamlined queries and pertinent keywords would require further work. Further research would also be needed to determine effective ways to integrate recommendations within these new environments.

6 Summary

We presented an approach to facilitating user contributions to unstructured content within Wikipedia articles. This approach aims to reduce the amount of effort required to contribute to Wikipedia articles by helping users find and consult relevant resource materials. In a formal laboratory evaluation, we found that this approach, embedded in the IntelWiki prototype, affords a number of advantages in comparison to the default Wikipedia editor design. With IntelWiki, participants were able to write more text, describe a larger number of different facts and were more accurate in their descriptions. Subjectively, participants reported experiencing significantly lower mental workload and all but two of the sixteen participants preferred IntelWiki's approach. We have also identified a number of promising avenues of future work including automated pertinent keyword identification, exploring system extensions that leverage relevance feedback, and exploring the impact of the approach on contribution rates.

Acknowledgements. This work was supported by the GRAND Network Center of Excellence and the National Sciences and Engineering Research Council of Canada.

References

1. Bostandjiev, S., O'Donovan, J., Hall, C., Gretarsson, B., and Höllerer, T. WiGipedia: A Tool for Improving Structured Data in Wikipedia. Proc. of ICSC, (2011), 328–335.

2. Bryant, S. L., Forte, A., and Bruckman, A. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. Proc. of GROUP, (2005), 1–10.
3. Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. Using Intelligent Task Routing and Contribution Review to Help Communities Build Artifacts of Lasting Value. Proc. of CHI, (2006), 1037–1046.
4. Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. SuggestBot: Using Intelligent Task Routing to Help People Find Work in Wikipedia. Proc. of IUI, (2007), 32–41.
5. Croteau, D., Hoynes, W., & Milan, S. Media/Society: Industries, Images, and Audiences. Sage, 2011.
6. Hanrahan, B. V., Bouchard, G., Convertino, G., Weksteen, T., Kong, N., Archambeau, C., and Chi, E. H. Mail2Wiki: Low-Cost Sharing and Early Curation from Email to Wikis. Proc of C&T, (2011), 98–107.
7. Hart, S. G. and Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Human Mental Workload 1, (1988), 139–183.
8. Hoffmann, R., Amershi, S., Patel, K., Wu, F., Fogarty, J., and Weld, D. S. Amplifying Community Content Creation with Mixed Initiative Information Extraction. Proc. of CHI, (2009), 1849–1858.
9. Holloway, T., Bozicevic, M., and Borner, K. Analyzing and Visualizing The Semantic Coverage of Wikipedia and Its Authors. Complexity, 12(3), (2007), 30–40.
10. Kong, N., Hanrahan, B. V., Weksteen, T., Convertino, G., and Chi, E. H. VisualWikiCurator: Human and Machine Intelligence for Organizing Wiki Content. Proc. of IUI, (2011), 367–370.
11. Krieger, M., Stark, E., and Klemmer, S. R. Coordinating Tasks on the Commons: Designing for Personal Goals, Expertise and Serendipity. Proc. of CHI, (2009), 1485–1494.
12. Nielsen, J. Participation Inequality: Encouraging More Users to Contribute. http://www.nngroup.com/articles/participation-inequality/. Accessed: 2013/10/07.
13. Nov, O. What Motivates Wikipedians? CACM, 50(11), 2007, 60–64.
14. Okuoka, T., Takahashi, T., Deguchi, D., Ide, I., and Murase, H. Labeling News Topic Threads with Wikipedia Entries. ISM, (2009), 501–504.
15. Sauper, C. and Barzilay, R. Automatically Generating Wikipedia Articles : A Structure-Aware Approach. Proc. of ACL and AFNLP, (2009), 208–216.
16. Weld, D. S., Wu, F., Adar, E., Amershi, S., Fogarty, J., Hoffmann, R., Patel, K., and Skinner, M. Intelligence in Wikipedia. Proc. of AAAI, (2008), 1609–1614.
17. Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gery, G., and Smith, M. Finding Social Roles in Wikipedia. Proc. of iConference, (2011), 122–129.
18. Wikimedia Report Card. http://reportcard.wmflabs.org/. Accessed: 2013/12/02.
19. Wikipedia:Please Do Not Bite The Newcomers. http://en.wikipedia.org/wiki/Wikipedia:Please_do_not_bite_the_newcomers. Accessed: 2013/10/07.
20. Wilkinson, D. M. Strong Regularities in Online Peer Production. Proc. of EC, (2008), 302–309.
21. Woodsend, K. and Lapata, M. WikiSimple: Automatic Simplification of Wikipedia Articles. Proc. of AAAI, (2011), 927–932.
22. Wu, F., Hoffmann, R., and Weld, D. S. Information Extraction from Wikipedia: Moving Down The Long Tail. Proc of KDD, (2008), 731–739.
23. Zachte, E. Wikipedia Statistics. http://stats.wikimedia.org/EN/TablesWikipediaZZ.htm #editor_activity_levels. Accessed: 2013/12/02.