IntelWiki - Recommending Reference Materials in Context to Facilitate Editing Wikipedia

by

Mohammad Noor Nawaz Chowdhury

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Computer Science
University of Manitoba
Winnipeg

Thesis advisor Author

Dr. Andrea Bunt **Mohammad Noor Nawaz Chowdhury**

IntelWiki - Recommending Reference Materials in Context to Facilitate Editing Wikipedia

Abstract

Participation in contributing content to online communities remains heavily skewed. Yet little research has focused on lowering the contribution effort. I describe a general approach to facilitating user-generated content within the context of Wikipedia. I also present the IntelWiki prototype, a design and implementation of this approach, which aims to make it easier for users to create or enhance the free-form text in Wikipedia articles. The IntelWiki system i) recommends article-relevant reference materials, ii) draws the users' attention to key aspects of the recommendations, and iii) allows 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. Participants also reported experiencing significantly lower mental workload while editing with IntelWiki and preferred the new design.

Publications

Some ideas in this thesis have appeared previously in the following publication by the author:

Mohammad Noor Nawaz and Andrea Bunt. IntelWiki: Recommending Resources to Help Users Contribute to Wikipedia. In *Proceedings of the 22nd International Conference on User Modeling, Adaptation, and Personalization*, (UMAP 2014), 12 pages, to appear.

Acknowledgments

Firstly, I would like to express my endless gratitude to Dr. Andrea Bunt for being an awesome supervisor for the entire duration of my masters program. Without her continuous guidance, precise directions and unceasing patience I would not be able to complete the thesis. She has been actively involved in each stage of my thesis from the beginning to the end, motivated and encouraged me in achieving the goal. I would also like to acknowledge the financial support I received from the department of Computer Science in the form of a Guaranteed Funding Package and would like to thank Dr. Bunt once again for the assistance provided.

Secondly, I would like to thank my committee members, Dr. James E. Young and Dr. Ken Ferens for their valuable times and efforts in perfecting my thesis through discerning suggestions and feedback.

Thirdly, I am highly indebted to my wonderful fellow lab-mates in the HCI lab who have always been very generous to support me in various ways. Matthew Lount, Khalad Hasan and Ashish Singh deserve special mention for their ideas, comments, feedback and help. I would also like to thank Adnan Alam Khan and Samiul Alam Anik for spending their time and effort in proof-reading my thesis.

Last but not least, I would like to acknowledge the unconditional love and support of my family staying overseas and dedicate this thesis to my parents Quazi Faruka and Mohammad Shah Nawaz Chowdhury.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

User-generated content (UGC) or community content is content generated by people who voluntarily contribute data, information, articles, blogs or media on the web [32]. This content then becomes available to other internet users (i.e., consumers) and can serve as a great source of *infotainment* to them. UGC has experienced rapid growth in recent years, due to the advances in Web 2.0 technologies [22]. Users generally supply content voluntarily [32] for various reasons, be it altruism, recognition, fun, or social needs [19, 25, 28, 35]. However, the percentage of the population that contributes content tends to remain relatively small. In particular, most community content follows the "1% rule", where approximately 1% of internet users create the content, 9% enhance it, and the remaining 90% simply consume it without contribution [13, 33, 34, 43]. 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 under-representation of the views and interests of a large percentage of the population [34].

There are many factors that influence participation rates. For example, community politics can make it difficult for newcomers to gain entry into the community [11]. While addressing community politics requires a cultural change, a significant barrier to participation is simply the amount of effort required to do so. In particular, Nielsen's number one suggestion on how to increase participation rates is: "Make it easier" [34]. This assertion is supported by studies indicating that editing effort can indeed affect participation rates [17, 27, 50].

My work focuses on supporting user contributions to Wikipedia. Wikipedia, the free internet encyclopedia, was launched on January 15, 2001 by Jimmy Wales and Larry Sanger [8]. It has experienced a steady growth in terms of article count since then [10]. Currently, Wikipedia is the sixth most globally visited website over the internet [1]. Most of the Wikipedia articles can be edited by any Wikipedia user [44], and about 100,000 people actively contribute to Wikipedia [55], which makes Wikipedia one of the most prominent examples of UGC. Like other community content repositories, only a very 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) [7, 54]. Priedhorsky et al. showed that, nearly half of the content in Wikipedia is created by the top 0.1% Wikipedia contributors [38].

Wikipedians generally edit articles on topics which they are familiar with [17, 35]. Despite this fact, searching for relevant online materials could be necessary to verify certain documented facts, find additional information and to provide references. Prior work indicates that one of the attributes that makes editing Wikipedia articles particularly difficult in relation to some other forms of UGC (e.g., movie reviews), is the need for background research prior to editing [50]. This thesis investigates whether an intelligent system that automatically generates resource recommendations could make the process of editing Wikipedia articles easier. This investigation involves the following research questions:

- How should a system generate resource recommendations?
- In what way should a system present the recommended resource materials to the user?
- Does having streamlined access to recommended resource materials make it easier for users to edit Wikipedia articles?

To answer these questions I designed and implemented the IntelWiki system. The IntelWiki system makes use of Google Custom Search Engine (CSE) API to search for online reference materials relevant to a certain Wikipedia article. It then ranks the resources measuring the volume and variety of information based on a set of keywords deemed important to that particular Wikipedia article. IntelWiki's editing environment also allows the users to consult any recommended resources they find useful within the context of the Wikipedia editor itself.

In order to test the IntelWiki system's impact on users' Wikipedia editing experience, I have conducted a formal laboratory evaluation with 16 novice Wikipedia editors comparing the system's approach to the default Wikipedia editor.

The participants were asked to edit articles using 1) the IntelWiki system and 2) The Wikipedia editor plus the Google search. The results of the evaluation provide evidence that IntelWiki's support does have the potential to facilitate the editing process. In particular, when editing an existing article, participants wrote significantly more text (in a fixed time period) with IntelWiki than with the default Wikipedia editor, wrote about a larger number of different concepts and were more accurate in their descriptions. Subjectively, participants reported experiencing significantly lower mental workload in the IntelWiki condition and their preferences were overwhelmingly in favour of IntelWiki.

The structure of the remainder of my thesis is as follows. In chapter 2 I explore related work. In chapter 3 I describe the design goals, process and implementation specifics of the IntelWiki system. In chapter 4 I describe the study I conducted to evaluate the IntelWiki system and present the results. I conclude in chapter 5 by summarizing the contributions of my thesis and discussing the limitations and possible future directions.

Chapter 2

Related Work

User-generated content (UGC) in general and Wikipedia in particular, has been a widely studied phenomenon. Researchers have focused their work on numerous aspects of UGC. I will be exploring work in this chapter from three main areas. I start with studies conducted to find the ways in which UGC systems can affect contributions. Then I discuss research most relevant to my topic: research on improving Wikipedia content (both structured and unstructured). Lastly, I explore work focusing on matching users to tasks in Wikipedia.

2.1 Influencing Contributions

Researchers have tried to analyze several aspects of online UGC contributions, including: studies on what motivates contributions [35, 37, 41, 42], how editing roles evolve over time [29, 31, 48], and participation inequalities [13, 33, 34, 38]. Below I discuss work that analyzes the different ways UGC systems guide contribution rates.

Bryant et al. conducted a study with 9 Wikipedians on their motivations and experiences with Wikipedia [17]. The interviews showed that readers arrive in Wikipedia searching for information and start contributing to topics they are knowledgeable on, by simply knowing that they are allowed to contribute (e.g., via the "Edit This Page" link). The advanced editors assume greater sets of responsibilities (e.g., administration, maintaining article integrities, removing vandalism) [17, 31]. A number of Wikipedia tools (e.g., the talk pages, the user pages, the WikiProject pages and the watchlist) help the advanced editors in carrying out their responsibilities.

Research has provided evidence that there is an inverse correlation between perceived editing effort and contribution rate. Wilkinson analyzed the contribution pattern in four collaboratively maintained UGC systems, namely: Wikipedia, Digg, Bugzilla and Essembly, and showed that the probability of a user ceasing to contribute is proportional to the effort required to contribute [50]. Hoffmann et al. showed that an intelligent Wikipedia system can encourage more participants to enhance the uninformative article infoboxes by offering help through the system's information extraction capabilities (elaborated in subsection 2.2.1) [27].

Inspired by these results that link contribution rates to editing effort, I have proposed the IntelWiki system which tries to make it easier for the Wikipedia editors by automatically suggesting online reference materials in context, relevant to the article being edited.

Figure 2.1: Sample Wikipedia Article for the "Lake" category (infobox specified by blue rectangle) [4]

2.2 Enhancing the Content of UGC Systems

Prior research has focused on enhancing the content of the UGC systems. A significant amount of research in this area has been based on Wikipedia. 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) unstructured or free-form content. Structured information has a pre-defined schema and can be organized using a relational database management system. One of the primary purposes of structuring information is to leverage searching and querying. Structured information is often presented in a tabular or graph-based format within the interface, such as the information found in a standard Wikipedia article's infobox (see Figure 2.1). An infobox is a compact, tabular summary of an article's salient points [9].

Wikipedia articles are organized into a collection of (hierarchical) categories. For example: the category "Lake" is under the parent category "Bodies of water", while the category "Film" is under the parent category of "Art Media". Both of these categories (i.e., "Lake" and "Film") have multiple subcategories. Figure 2.1 depicts a sample Wikipedia article for the "Lake" category. Articles within a category share a common infobox template, thereby summarizing key attributes of the articles belonging to that category. The unstructured information, on the other hand, consists of the free-form text written by human editors, including prose, images, links and references.

The IntelWiki system that I have designed as part of my thesis seeks to improve the free-form content in Wikipedia by drawing contribution from potential contributors through a reference-integrated environment.

2.2.1 Improvements to Structured UGC Content

A number of researchers have focused on using machine intelligence to improve the structured data in Wikipedia. A notable example is the Kylin system, which automates the process of creating and completing Wikipedia article infoboxes (e.g., [27, 52, 53]). An evaluation of a mixed-initiative version of Kylin revealed that having Kylin suggest potential changes to the infoxboxes had positive impacts on both user contribution rates and infobox accuracy [27]. Sharing some similarities with my approach, Weld et al. proposed an extension to the Kylin system, where the information extraction used to improve the infoboxes is extended beyond Wikipedia articles to the general web [47]. Like with my approach, this extension relied on web queries to find useful resources, however, these resources were used by the learning algorithm only as opposed to being presented to potential editors. As another example of improving structured Wikipedia content, the WiGipedia tool helps users identify and correct inconsistencies among structured data spread across different articles [16]. Finally, Adar et al. proposed a system, Ziggurat, which uses the cross-linkage between articles of different Wikipedias (e.g., English, Spanish, French and German) to facilitate information consistency between the infoboxes, by automatically completing the missing infobox attribute values [12].

Outside the context of Wikipedia, Shortipedia [45] pulls structured information from Linked Open Data [5], a Semantic Web [6] that combines a number of online sources (e.g., DBpedia [14], Freebase [15], Geo Names [49], ACM Digital Library, DBLP, CiteSeer [23], The New York Times, etc.). Shortipedia allows users to enhance this structured data by adding additional facts and/or aggregating the data. While Shortipedia relies on Semantic Web techniques (sameAs.org service and Sindice search engine) to collect facts from the linked open data, the resource fetching in my work is based on the entire web (via Google search engine).

In contrast to this work, my work focuses to enhance the free-form Wikipedia content by providing suitable reference materials.

2.2.2 Improvements to Free-Form Content

The majority of prior work on improving free-form Wikipedia content 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 [36]. WikiSimple takes Wikipedia articles as input and automatically produces articles re-written in simpler grammatical style (to enhance readability) [51]. Finally, Sauper et al. proposed a fully automated process for generating a multi-section Wikipedia article [40]. While this latter work shares some similarities with IntelWiki (i.e., collecting online content via search engine queries, and leveraging Wikipedia article category structure), IntelWiki aims to support an editor as opposed to automatically generating an entire article.

A mixed-initiative approach to enhancing free-form article text has been explored in the context of corporate wikis. Specifically, the Mail2Wiki system is designed to transfer knowledge automatically from the emails to wikis [24], with the VisualWikiCurator system providing users with tools to help them organize the transferred knowledge [30]. In contrast, IntelWiki automatically brings the information sources and presents them to the user facilitating an easy navigation through the materials, while allowing the users to transfer the knowledge by themselves.

2.3 Intelligent Task Routing in Wikipedia

Prior work has also sought to help users determine which articles to contribute to. Cosley et al. showed that intelligent task routing (i.e., matching people with tasks) can accelerate contribution in online communities [20], and proposed SuggestBOT [21], which matches an editor with appropriate Wikipedia editing tasks (based on interests and editing history). WikiTasks, on the other hand, is a system that helps potential editors focus their efforts by pulling tasks from article talk pages, user pages and WikiProject pages and displaying them within the context of the Wikipedia editor [31]. In contrast to the above prior work, IntelWiki focuses on helping editors once they have determined which articles/sections they wish to edit.

2.4 Summary

From the literature review, we see that significant research [12, 16, 27, 45, 47, 52, 53] has been done on enhancing the structured data over the web and UGC systems (particularly Wikipedia). In contrast, research work focusing on enhancing the free-form content [40, 51] has mainly focused on automating the content generation process. We also see work on suggesting tasks (articles to edit) to a user, with the user then responsible for writing the articles completely of their own. My work falls in between these two latter bodies of work. Assuming a user already knows what article s/he wants to work on, my system suggests relevant online reference materials to the user, providing them with useful information to edit the article.

Chapter 3

System Description

In this chapter, I present the IntelWiki prototype system, and describe the functionalities of different modules of the system. The IntelWiki system aims to make the process of enhancing free-form text in a Wikipedia article easier by helping editors locate and interact with relevant reference materials. More specifically, IntelWiki aims to reduce the time necessary to find pertinent resources and to streamline the process of interacting with those resources by integrating them within the Wikipedia editor.

To accomplish the above goals, the system relies on three primary components (displayed in Figure 3.1): i) the Resource Fetcher, which searches the Web for relevant reference materials; ii) the Resource Ranker, which processes the fetched resources to rank them according to the system's assessment of their suitability to the editing task; and iii) the Resource Presenter (the user interface component), which allows users to view and interact with the fetched resources. I describe each component in more detail below.

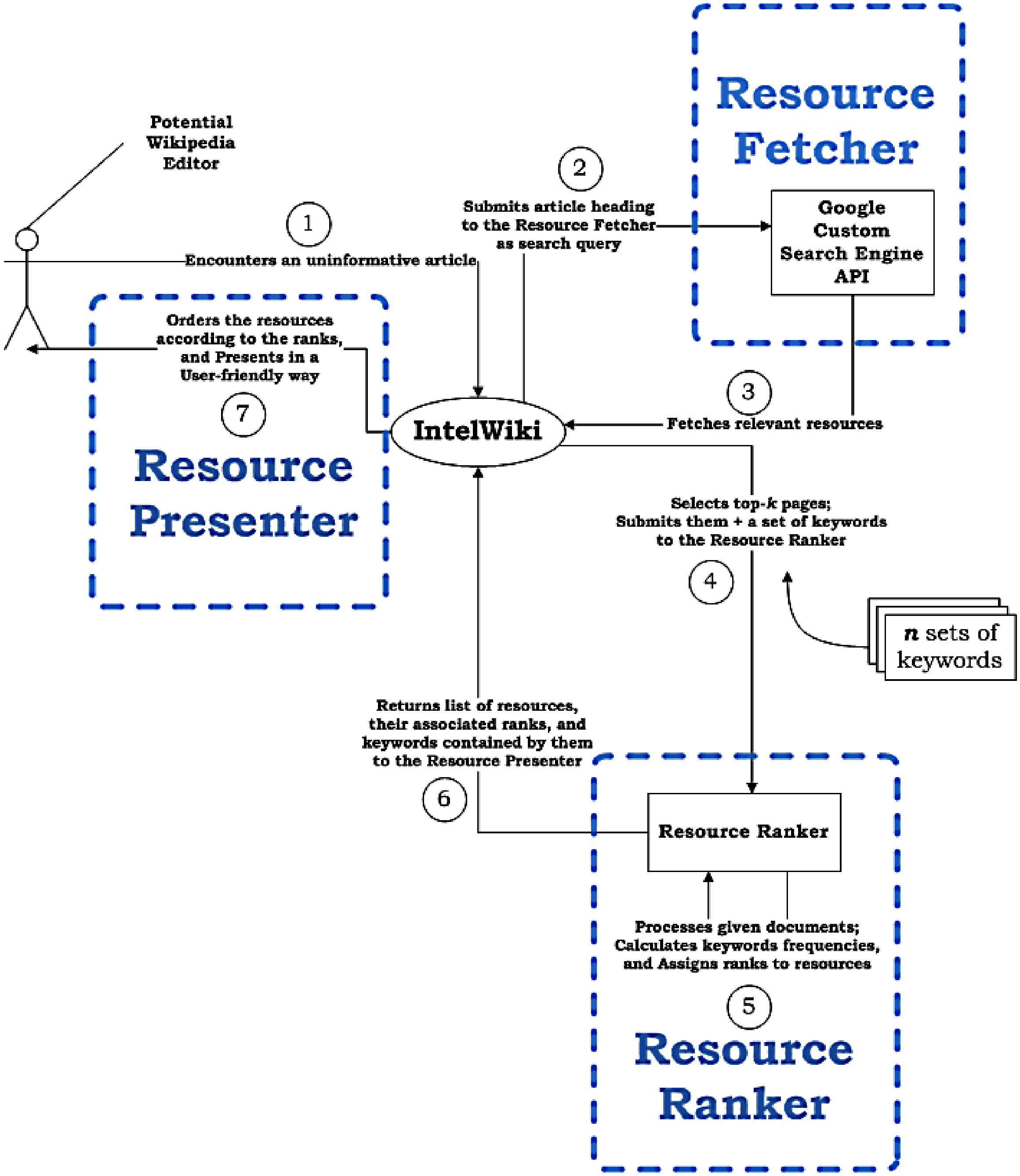

Figure 3.1: System Architecture (Clockwise from top: Resource Fetcher, Resource Ranker, Resource Presenter)

3.1 Resource Fetcher

As indicated above, IntelWiki's Resource Fetcher searches the web for resource materials that could help a potential editor enhance a given Wikipedia article. Inspired by prior research [46, 47, 52], IntelWiki uses Google's Custom Search Engine (CSE) API [3], submitting the article title as a search query. From the returned results, the Resource Fetcher then selects the top k pages. Presumably, too small a value (e.g., less than 10) of k would discard relevant information resources while too large a value (e.g., greater than 100) of k could include a lot of resources that are only vaguely relevant. Currently IntelWiki is using a value of 60 for k. However, further experimentation is required to find out the optimum value.

To help ensure that the selected resources are both potentially useful and ones that can be processed by the Resource Ranker, IntelWiki removes the following from the candidate set of URLs: the link to the article in question, pages with very long response times (greater than 30 seconds) and pages that are not machine readable (i.e., whose text is contained only in images). This final set of URLs is then stored into a database located in the same server as IntelWiki. These URLs are later used by the Resource Ranker module, which is described in the next section.

3.2 Resource Ranker

Based on my informal experimentation (i.e., searching with random article titles via Google), the relevant resources are seldom ordered according to their information richness. For example, as of October 2013, searching with the terms

"Dhanmondi Lake" (a natural lake in Dhaka, Bangladesh) and "Karnaphuli River" (a well-known river in Chittagong, Bangladesh) resulted in the corresponding articles from www.bpedia.org (one of the most comprehensive source for Bangladesh-related online articles) appearing in 15th and 17th positions respectively. Therefore, I perceived that sorting the resultant resources based on some key facts (which I call "pertinent keywords") associated with the search terms, would probably result in a better ordering (than using the Google CSE default ranking) for presentation by the IntelWiki system.

The Resource Ranker takes the candidate set of reference pages collected by the Resource Fetcher and examines each page to assess the page's potential relevance to the task at hand based on the occurrence of "pertinent keywords". In the subsections below, I elaborate on the nature of these keywords and how the Resource Ranker leverages them to produce its assessments.

3.2.1 Pertinent Keywords and Calculating Relevance Scores

To help assess the suitability of the resources in a way that is more tailored to the article than the default ranking returned by the Google CSE, the Resource Ranker ranks each resource according to the occurrence of pertinent keywords (described below). Let c_i denote the count of the article's i-th pertinent keyword. For each reference R_j, the relevance score (R_j_Score) is calculated as follows:

$$R_j\_Score = c_1 \times c_2 \times ... \times c_i \times ... \times c_n = \prod_{i=1}^{n} c_i$$

The formula uses the product of the keyword counts instead of sum to put more weight on the reference resources containing a higher number of distinct keywords.

All the reference resources are not likely to contain all the keywords. Therefore, in case of keywords with zero-count, the keywords are assigned a weight of 1 to prevent the relevance score from turning to 0.

As mentioned in the related work (section 2.2), the infobox in a Wikipedia article generally contains material elaborated elsewhere in the article. In the absence of additional information, the IntelWiki system uses the infobox schema attributes as the set of pertinent keywords.

Based on my informal experimentation I also noted the potential to improve the relevance score with a more complete set of pertinent keywords. Therefore, the IntelWiki system also allows additional keywords to be specified. Potential additions include: i) key attributes missing from the infoboxes (e.g., some categories simply inherit schemas from their parents, without proper tailoring); ii) attribute synonyms, root words, parts of speech variants; and iii) attribute values. The list of keywords could be generated by a Wikipedia administrator, through crowd-sourcing techniques, or through machine learning techniques that learn pertinent keywords from other articles of the same category.

3.2.2 Section-Specific Keyword Mappings and Relevance Scores

In addition to sharing common infobox schemas, articles in a given category are often very similar in structure. For example, Lake articles typically contain sections describing Geography (or Hydrology), 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. Figure 3.2 illustrates this mapping in formal terms. In particular, each section (S_i) could consist of a subset of the infobox attributes (A_i) plus an additional set of pertinent keywords (B_i). Table 3.1 shows a concrete example of what a section-specific keyword mapping might look like for articles in the "Lake" category.

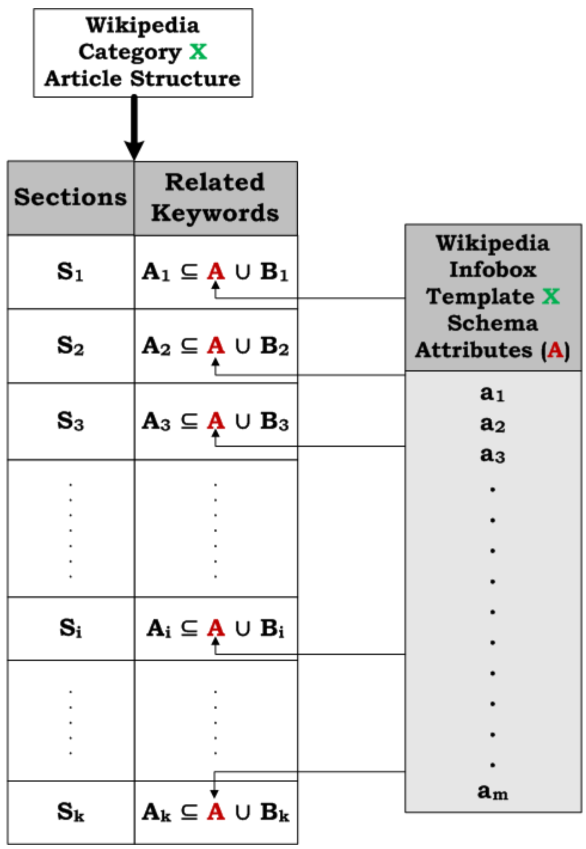

Figure 3.2: Mapping pertinent keywords to individual sections for articles of a given category.

If a keyword-to-section mapping exists for a particular article category, IntelWiki calculates a set of relevance scores on a per-section basis as follows:

Let c_i denote the count of Section X's (S_x) i-th pertinent keyword. The resource

R_j's relevance score for S_x ($R_j\_S_x\_Score$) is determined as follows:

$$R_j\_S_x\_Score = c_1 \times c_2 \times ... \times c_i \times ... \times c_n = \prod_{i=1}^{n} c_i$$

Table 3.1: An example of potential pertinent keywords for five sections of the Wikipedia article category "Lake"

"Lake" Sections	Section-Specific Pertinent Keywords
Geography	**Infobox Schema Attributes:** coords, inflow, outflow, catchment, basin_countries, length, width, area, depth, max-depth, water volume, residence-time, shore, elevation, islands. **Attribute Variants:** retention time, altitude, long, wide, deep. **Attribute Values:** meter, kilometer, feet, square meter, cubic meter.
Climate	**Infobox Schema Attributes:** frozen. **Additional Attributes:** lake-effect snow, surface temperature, fog, season. **Attribute Values:** maritime, summer, winter, fall.
History	**Infobox Schema Attributes:** date-built, date-flooded, cities. **Attribute Variants:** settlement. **Attribute Values:** ice age, paleolithic, year, century, millennium.
Geology	**Additional Attributes:** fossil fuel, mineral, soil, rock. **Attribute Values:** copper, iron, silver, gold, mine, lava, magma, rift, glacier.
Ecology	**Additional Attributes:** food web/chain, flora, fauna, species, carnivores, herbivores, decomposers. **Attribute Values:** fish, trout, catfish, pike, salmon, perch.

These section-specific relevance scores can then be used by the Resource Presenter interface component to re-order the list of resources as the user edits individual sections, based on the system's calculation of how valuable the resources will be to each section.

For my thesis, I leave the design of a component that determines the set of section-specific keywords as a "black box", focusing on exploring the value of the general approach of recommending resources. Similar to defining the set of pertinent keywords, this mapping could be defined by a Wikipedia administrator, through crowd-sourcing techniques, or through machine learning techniques. Note that this mapping would need to be done a per-category basis (as opposed to a per-article basis).

3.2.3 Impact of Pertinent Keywords on Rankings

To explore the impact of the pertinent keywords on IntelWiki's ranking, I selected two other categories (e.g., "Car" (automobile) and "Cricketer" (athlete)) and came up with sets of keywords for each of them. Then I randomly selected two articles from each of these three categories. To measure the difference between IntelWiki's ranking and Google's default ranking, I calculated the differences between the positions of each suggested resource in both rankings. Then I took the average of these differences to calculate the average absolute positional difference.

For each article, IntelWiki's rankings were calculated using three overlapping sets of keywords resulting in three different orderings. The first set of keywords comprised all the infobox schema attributes. The remaining two sets consisted of section-specific keywords. The first of the latter two sets of keywords were constructed using subsets of the infobox attributes, while the last set comprised subsets of the infobox attributes plus additional sets of pertinent keywords. The specifics of these measurements: the categories, articles, chosen sections and the average absolute positional differences resulting from different sets of keywords can be found in Table 3.2.

As demonstrated by Table 3.2, using the complete set of infobox schema attributes does lead to a substantially different ranking from Google's default ranking. However, the rankings calculated using more tailored information (i.e., section-specific attributes and keywords) varied only slightly from the rankings calculated using the complete set of infobox schema attributes.

Table 3.2: Average Absolute Positional Differences between Google's default ranking and IntelWiki's rankings (based on various sets of keywords).

Category	Section	Article	Average Absolute Positional Difference		
			All Infobox Attributes	Section-specific	
				Relevant Infobox Attributes	Pertinent Keywords
"Lake"	"Geography"	"Lake Winnipeg"	17.47	17	19.33
		"Lake Superior"	11.93	11.87	15.93
"Car"	"Features & Specifications"	"2010 Honda Accord"	19.3	19.03	19.53
		"2010 Chevrolet Impala"	18.3	17.9	18.2
"Cricketer"	"Cricket Career"	"Sachin Tendulkar"	13.5	13.2	14.23
		"Adam Gilchrist"	15.57	15.03	16.8

3.2.4 Resource Ranker Implementation Details

A PHP script was used to retrieve the stored URLs (by the Resource Fetcher module) from the database. For each URL, I extracted (discarding all the HTML tags and keeping the inner-texts of them) the plain text content from the page using the Simple HTML DOM Parser library [18]. Unreadable URLs and URLs with long response times were discarded. Then using the Simple HTML DOM Parser library and the set of pre-selected "Lake" relevant keywords (example in subsection 3.2.1), I parsed the resources counting the occurrence of each keyword. The relevance score of a resource (detailed in subsection 3.2.1) was calculated based on these counts and stored back into the database to be used by the Resource Presenter module.

3.3 Resource Presenter

All the collected and processed information by the earlier modules are presented to the user by the IntelWiki system's Resource Presenter module. One of the primary focuses of my thesis was to experiment with the system's ease of use. I wanted to optimize the way IntelWiki would present the processed information in order to minimize a user's editing effort.

3.3.1 Design Goals

I tried to follow three design principles in designing the IntelWiki interface layout.

1. Reduction of Context-Switch - I wanted the user to be able to browse and examine the suggested resources in the same screen (without needing to switch to another browser window) as the article being edited.

2. Consistency - I wanted to ensure a certain amount of consistency between the interface layouts of Wikipedia and IntelWiki. I also wanted to ensure that the IntelWiki features are displayed in the same fashion and location in both view and edit modes.

3. Persistence of Information Sources - I speculated that being able to display the information sources (the list of suggested resources and the currently opened resource(s)) persistently (i.e., without the user having to click or hover on them repeatedly) is advantageous.

3.3.2 Design Process

In designing the interface, I began by brainstorming on what different features the IntelWiki system interface would contain. The major candidate features of the interface were:

- A "Suggested Resources" pane containing:
  - The article-relevant resource titles the system would suggest.
  - The set of keywords contained by the resources and their respective counts.
  - The relevance score of the resource.
- A View pane to display a particular resource called the "Resource Viewer" pane.
- The Article Content pane.

Once the major features of the IntelWiki interface were identified, the question to ponder was how the different features would be accommodated following the current Wikipedia interface layout. A number of steps were followed before finalizing the look and feel of the IntelWiki interface.

Sketching

To explore a number of different design alternatives I began with a series of interface sketches of the two major features: the "Suggested Resources" pane and the "Resource Viewer" pane. I also sketched several IntelWiki interface variants having different layouts. The designs included (1) having the two panes in various fixed positions of the interface, and (2) having either one or both of them in a disposable layer (e.g., popup or tooltip) above the article content pane.

Low-Fidelity Prototypes

From the sketches, I selected the most promising aspects and implemented four low-fidelity prototypes. These prototypes can be found in Appendix A. The main differences in the prototypes were: the appearances and positions of the two major features (i.e., the "Suggested Resources" pane and the "Resource Viewer" pane).

Pilot Testing

I conducted a pilot test with three voluntary participants to finalize the look and feel of the InteWiki interface, based on these prototype interfaces. Using the pilot participants' feedback, I settled on various aspects of the interface, such as the appearances, dimensions and positions of the "Suggested Resources" and the "Resource Viewer" panes, and the manners of interaction between a user and the overall interface (elaborated below in subsection 3.3.3).

3.3.3 Final Interface

The IntelWiki system's Resource Presenter makes the set of suggested resources available to a potential editor on demand, as shown by the callout in Figure 3.3. When a potential editor asks to view the reference materials, the system adds the two additional panes (i.e., the "Suggested Resources" pane and the "Resource Viewer" pane) to the regular Wikipedia interface in both viewing and editing modes. The final IntelWiki prototype interface is shown in Figure 3.4.

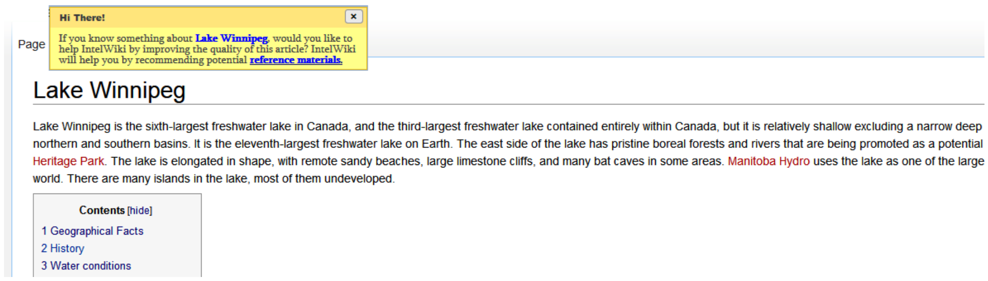

Figure 3.3: IntelWiki's callout. Potential editors can view IntelWiki's suggested resources by clicking on the link labelled "reference materials".

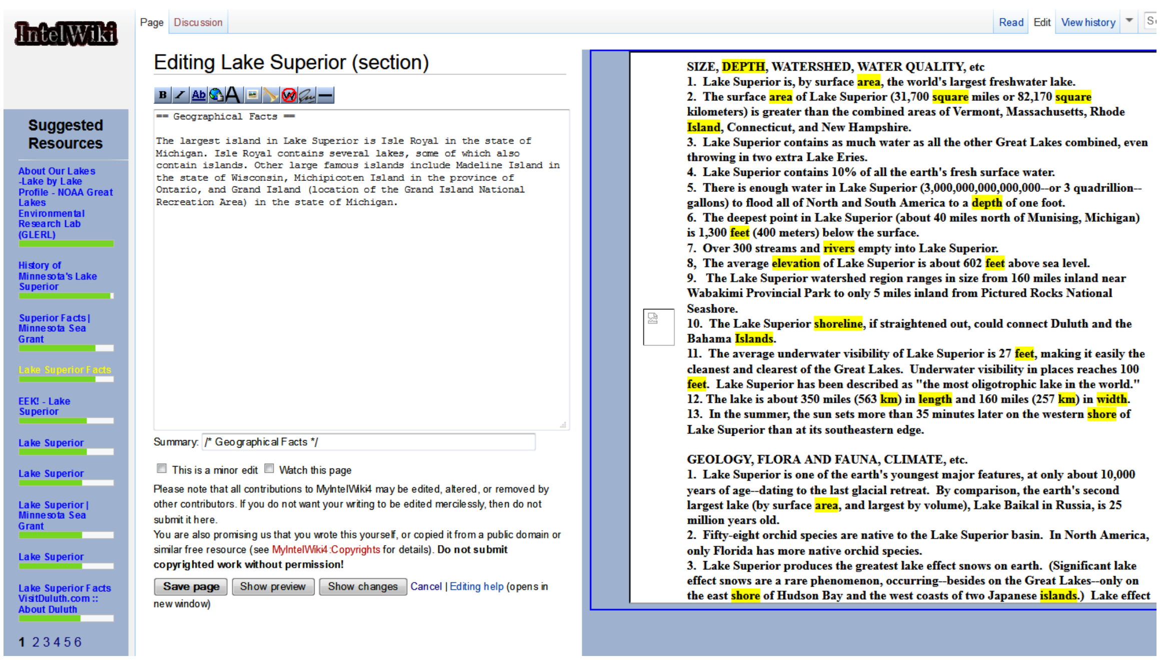

Figure 3.4: Editing with IntelWiki Interface, with the "Suggested Resources" pane (left) and the "Resource Viewer" pane (right).

I experimented with multiple "Resource Viewer" panes (shown in Appendix A, Figure A.4) to allow for simultaneous consulting of multiple resources. Placing multiple "Resource Viewer" panes in a single column, each having a minimum height to allow convenient reading, would force the cumulative height of the

"Resource Viewer" panes to be more than the screen height. Two of the three pilot participants commented that, while using this layout they had to scroll down the screen to view the bottom "Resource Viewer" pane entirely, which was inconvenient. Again, placing the "Resource Viewer" panes in multiple columns would significantly reduce the individual width which required both horizontal and vertical scrolling on the particular "Resource Viewer" pane, to locate a piece of information. All three pilot participants objected that horizontal scrolling was inconvenient. Therefore, in the final IntelWiki interface design I went for a single "Resource Viewer" pane (with height = 700 pixels) fitting entirely within the screen.

To minimize (if not eliminate) the need for horizontal scrolling in the "Resource Viewer" pane, I chose its width to be about 65% of the total article content pane. This reduced the width of the article content pane to about 35% of its original width. This design decision was well-received by two pilot participants provided that the article content pane was still wide enough for editing. Based on pilot participants' feedback, the "Suggested Resources" pane is located on top of the navigation panel in left sidebar so that it is entirely viewable within the screen space.

In the "Suggested Resources" pane, the system displays the rank (i.e., relevance score) of a particular resource by a green horizontal bar right below that resource link (see the green bars in Figure 3.5). To help the user browse the suggested resources, the system sorts the suggested resources using the relevance scores calculated by the Resource Ranker.

Since there can be large differences in the relevance scores, in particular, between the top few suggested resources and the remaining ones, the scores were scaled. If a resource R_j has a relevance score of R_j_Score, while the maximum relevance score of all the suggested resources is $R_j\_Score_{max}$, the resource R_j would have a displayed rank value $D\_R_j\_Score$ calculated by:

$$D\_R_j\_Score = \log_e(R_j\_Score) + \log_e(R_j\_Score_{max})$$

Figure 3.5: An example tooltip indicating the number of occurrences of each pertinent keyword within 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 counts (see Figure 3.5). Initially (or whenever the user is in the view mode) an article's recommended resources are sorted according to the per-article relevance scores (i.e., the R_j_Score described in the previous section). 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 pertinent keywords, if such a mapping exists for the article's category (i.e., based on the $R_j_S_x_Score$s described in subsection 3.2.1).

The user can view the contents of a particular resource by either clicking 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 relevant keywords within the resource (as shown in Figure 3.4).

3.3.4 Resource Presenter Implementation Details

The IntelWiki system Resource Presenter module is implemented over the MediaWiki framework, the platform used by Wikipedia. I installed the MediaWiki framework in a server, with PHP and MySQL.

Clicking on the link displayed by the IntelWiki callout invokes a JavaScript function which modifies the layout of the IntelWiki article page, loading the list of suggested resources in the left sidebar navigation panel, and an iframe to display the resources on the right side of the page. The majority of the JavaScript functions responsible for the IntelWiki interface layout are written in the Common.js file (a standard practice followed by the MediaWiki framework).

The Resource Presenter module pulls the URLs along with their calculated relevance scores from the database. Pulling all the information was facilitated using JSON.

A number of interactive components (e.g., rank-indicating bars, tooltips listing keywords and counts (see Figure 3.5), dragging and dropping links to the "Resource Viewer" pane, highlighting the keywords in the "Resource Viewer" pane) of the IntelWiki interface were implemented using JQuery libraries.

Opening a clicked (or dragged and dropped) URL in the "Resource Viewer" pane required addressing the same-origin policy for JavaScript [39]. Since the URLs were from remote locations, displaying their contents was facilitated by a simple server side PHP proxy script. Also, loading a complete HTML page in the "Resource Viewer" pane required using iframe.

During the first loading of an article page, all the information in the "Suggested Resources" pane were pulled from the database using JSON and stored in session (Mozilla Firefox sessionStorage [2]) to ensure faster performance in subsequent loading (e.g., toggling between view and edit modes) of that page.

3.4 Summary

In this chapter, I presented the IntelWiki prototype system. In designing the system, my primary focus was on providing users with streamlined access to a set of recommended reference materials – recommendations that are personalized to the individual article. The IntelWiki system automatically generates resource recommendations, ranks the references based on the occurrence of salient keywords, and allows users to interact with the recommended references directly within the Wikipedia editor. I followed three design goals for the IntelWiki interface. First, I wanted to allow the users to browse and examine the needed reference materials in the same screen as the Wikipedia editor. Next, I wanted to accommodate the features in the IntelWiki interface in such a way that the interface layout is consistent across view and edit modes, while maintaining consistency with the current Wikipedia interface. Lastly, I intended to display these extra pieces of information in a persistent way, so that a user could concentrate on the editing task without having to open up a resource repeatedly.

Chapter 4

Evaluation

In this chapter, I discuss my evaluation of the IntelWiki system. I conducted a formal laboratory study with 16 participants comparing the IntelWiki system to the regular (or 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. I begin by describing the study method. This is followed by a presentation of the study results and a discussion, which includes promising avenues of future work and limitations of the conducted study.

The study was approved by University of Manitoba's Research Ethics Board. The certificate of approval is included in Appendix B, Figure B.1.

4.1 Participants

18 participants (6 females) were recruited through on-campus advertising. Two participants (1 female) asked to stop the experiment early, stating fatigue as their

reason. The 16 participants who completed the experiment were between the ages of 18-32 (mean age 24.4). All participants were regular Wikipedia visitors, but none had previous Wikipedia editing experience. All participants were provided with a $15 honorarium for their time.

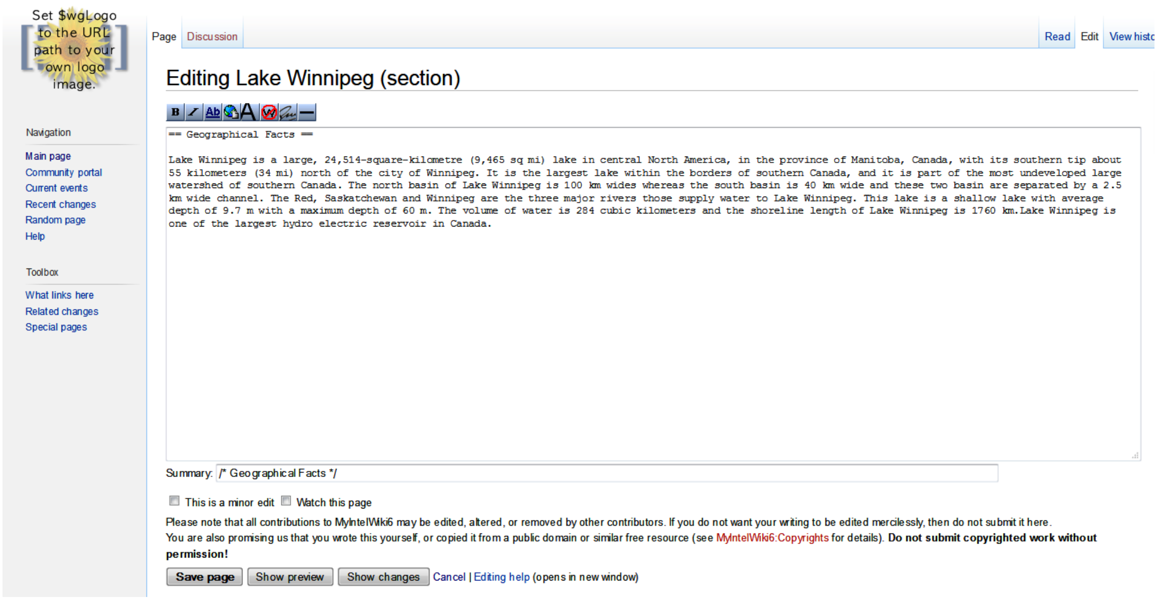

Figure 4.1: Regular Wikipedia interface in Edit Mode.

4.2 Apparatus

The experiment was conducted using a desktop machine with an Intel Core i7 2.93 GHz processor, 8 GB RAM, and a 23” monitor with a 1920x1080 resolution. The regular Wikipedia interface (displayed in Figure 4.1) was based on the MediaWiki platform (as the IntelWiki interface), but did not include the recommended resources and the resource viewer.

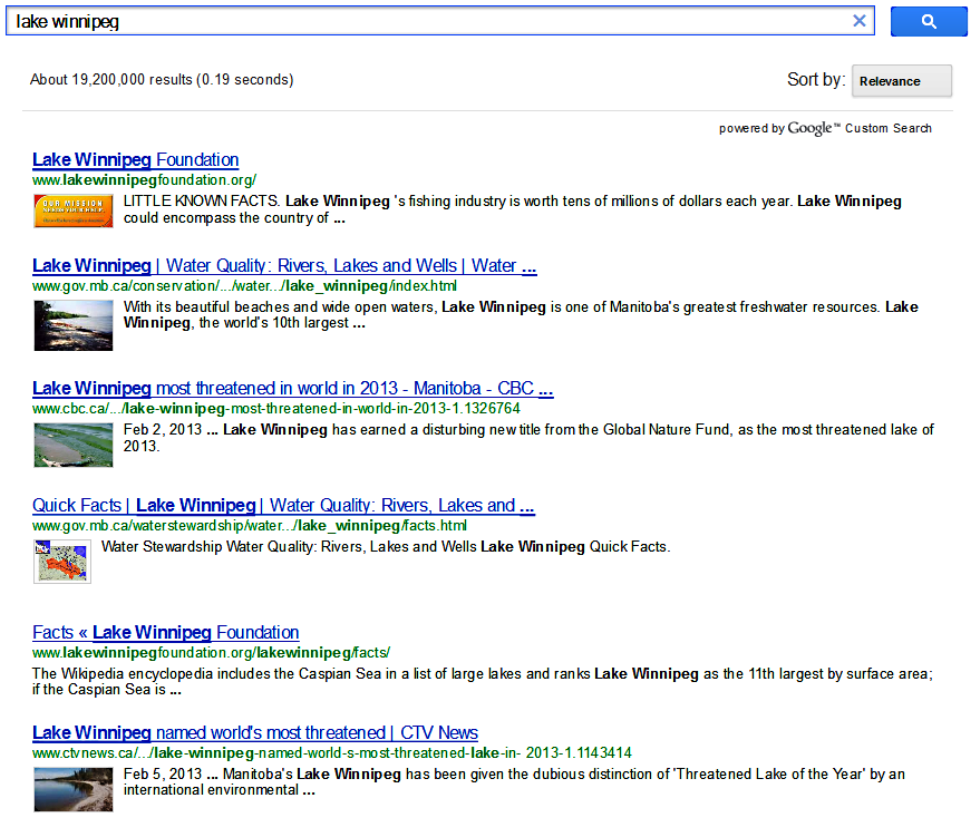

Figure 4.2: Google Custom Search interface

Both interfaces were loaded in the Firefox web browser. For the default Wikipedia interface, I used an instance of the Google Custom Search Engine API (the one used in the IntelWiki's Resource Fetcher module) instead of the regular Google Search engine to ensure that the link to the original Wikipedia article does not appear in the list of results. The default Wikipedia interface and the Google Custom Search interface (displayed in Figure 4.2) were loaded in two separate browser tabs.

4.3 Design

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

1. *Default:* The Wikipedia Edit interface plus the Google Custom Search interface.

2. *IntelWiki:* The complete IntelWiki system described in chapter 3.

To eliminate any preconceived bias, the IntelWiki was referred to as interface A, and the Default as interface B during the study. Participants completed one task (25 minutes long) with each interface type (described in the "Tasks" section). Therefore, task was a within-subjects control variable. Both the interface order and the task order were fully counterbalanced across participants to account for potential learning effects.

4.4 Tasks

In choosing Wikipedia articles for the editing tasks, I sought topics on which participants were likely to be equally knowledgeable, in an attempt to minimize variability owing to prior knowledge. To this end, I selected two articles about well-known lakes ("Lake Winnipeg" and "Lake Superior"). The articles were of similar level of complexity and belonged to the same category (i.e., sharing the same "Lake" infobox template). The articles were also selected so that the amount and the quality of relevant online information was roughly equivalent. IntelWiki used the section-specific keywords as defined in the "Geography" row in Table 3.1.

With each article, I asked participants to complete the Geography section (which I labelled "Geographical Facts"), whose content I removed from the original articles. I did however, leave three lines of text in those sections to provide some initial guidance as to what type of content could be included. The other modification I made to the articles was to remove the infoboxes since they were populated with facts from the original articles' Geography sections.

In the IntelWiki condition, participants were encouraged to use the system suggested resources. However, participants were allowed to perform external searches if they could not find some information they were looking for. This allowed me to evaluate the mixed-initiative style interaction that would be present in any field deployment (i.e., IntelWiki users would always have external search available).

To help participants get started with the editing, I provided them examples of attributes that they could describe (using geography-related attributes selected from the infoboxes, included in Appendix C). However, participants were told that they could edit the sections as they saw fit. I disabled copying and pasting, to discourage direct plagiarism from internet resources.

I asked participants to produce the best piece of text that they could within the 25 minutes allotted for each task, in terms of both text amount and quality. As motivation, participants were told that the top three performers would be awarded $15.

4.5 Procedure

After a brief introduction to the experiment, participants completed a demographics questionnaire, which included questions on their prior Wikipedia experience. Participants then edited the two articles, one with each condition. Prior to editing each article, participants were given a brief overview of the interface in that condition, and were asked to complete a short practice editing task (writing about a single geographical fact) with another lake article ("Lake Huron").

Participants were given 25 minutes to edit each article. After the 25 minutes, participants were asked to complete a NASA-TLX worksheet [26] to measure their perceived mental workload prior to beginning the next condition.

After completing the tasks with both conditions, participants were asked to complete a post-session questionnaire consisting of a number of Likert-scale questions (included in Appendix D). Lastly, participants took part in a semi-structured interview to collect further qualitative data on their perceptions of the system.

Sessions were typically 75 minutes long, with none exceeding 90 minutes.

4.6 Measures

Since editing time was fixed, my primary objective dependent measures concerned text volume and text completeness. In particular, I analyze the following for each condition:

- Word Count: the number of words written.
- Fact Count: the number of different facts that participants described in their articles.
- Fact Accuracy: the number of facts that participants described that were both related to the tasks (i.e., related to lake geography) and accurately documented.

I created liberal coding rules for measures: Fact Count and Fact Accuracy. Any distinct piece of information was counted as a fact. A fact was coded as accurate if it 1) was related to the topic of the section, and 2) was accurately reported. This latter criterion was judged using the original infobox or article when possible, or alternatively, the source that the participant used.

I also collected subjective data (e.g., mental workload and interface preference) through the questionnaires and interviews.

4.7 Results

The quantitative dependent measures were analyzed using a Repeated-Measures ANOVA with *Interface Type* (i.e., *IntelWiki* and *Default*) as the within-subjects factor. To check for asymmetric learning effects between two conditions, I also included *Interface Order* (i.e., *IntelWiki_First* and *IntelWiki_Second*) as a between-subjects factor in the analysis. Error bars on all graphs depict standard error. To remove the between-subjects variability from the standard errors, I subtracted the subject average from each observation, and added the grand average (i.e., the average of all the $16 \times 2 = 32$ cells), for each measure according to the following formula:

$$value_{new} = value_{old} - subject\_average + grand\_average$$

4.7.1 Timing Data

The amount of time the participants took to complete the tasks often deviated slightly from the 25 minutes limit, only to complete their last sentences. Therefore, I begin my analysis by examining the timing data. I note that the deviation in timing data was not significant as the participants took 25.3 seconds in the IntelWiki condition and 25.5 seconds in the Default condition ($F_{1,14} = 0.972$, p = 0.341, $\eta^2 = 0.065$).

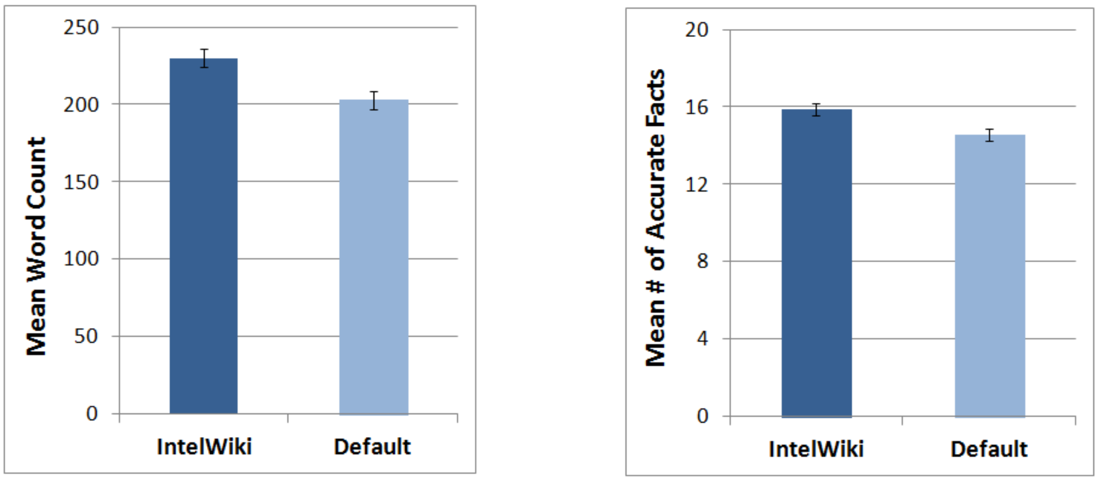

Figure 4.3: (Left) Mean Word Count by Condition. (Right) Mean Facts Accuracy by Condition.

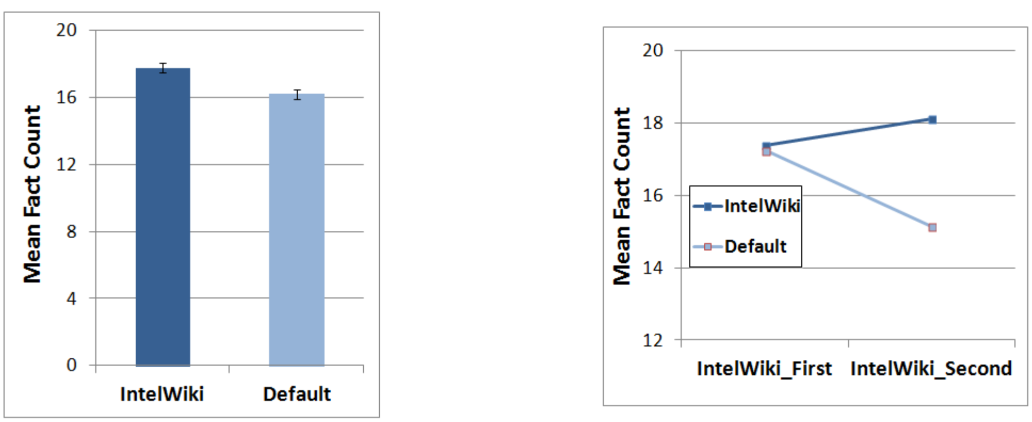

Figure 4.4: (Left) Mean Facts Count by Condition. (Right) The Interaction Effect between Interface Type and Interface Order.

4.7.2 Text Volume and Completeness

Next, I examine text volume which is measured by Word Count. Figure 4.3 (left) shows that participants contributed significantly more words with *IntelWiki* (229.9, s.e. 5.9) than with the *Default interface* (202.8, s.e. 5.9; $F_{1,14} = 5.302$, p = 0.037, $\eta^2 = 0.275$). 12 out of 16 participants contributed more words with IntelWiki than with Default (see Figure 4.5).

In terms of text completeness (measured by Fact Count and Fact Accuracy), IntelWiki outperformed the Default interface for both dependent measures (shown in Figure 4.4 (left) and Figure 4.3 (right)). For Fact Count, participants wrote about 17.8 (s.e. 0.29) different facts with IntelWiki as compared to 16.2 (s.e. 0.29) with Default ($F_{1,14} = 7.304$, p = 0.017, $\eta^2 = 0.343$). 11 out of 16 participants' edits contained more facts in the IntelWiki condition than in the Default condition, while two participants' edits contained equal number of facts in both conditions (see Figure 4.5).

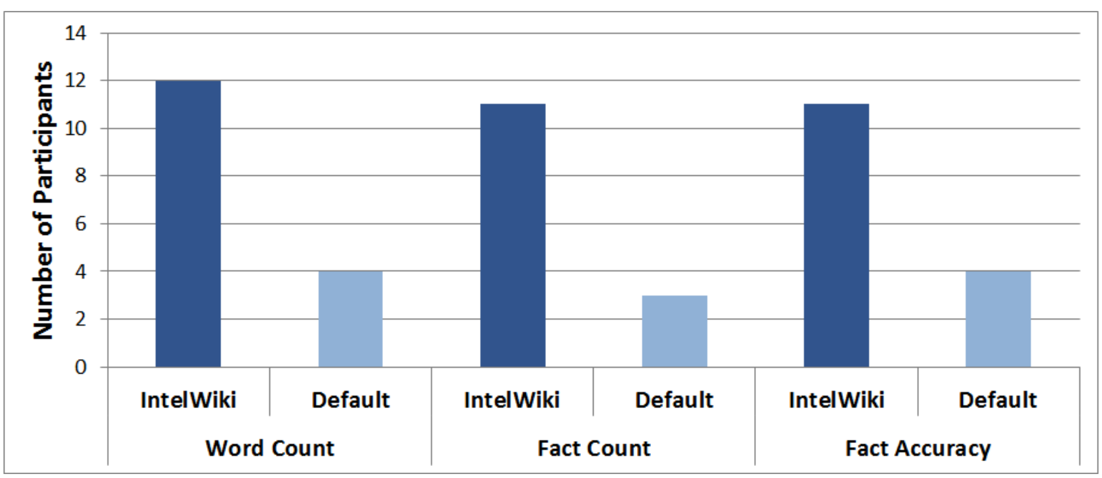

Figure 4.5: Number of Participants with Higher Contribution Count by Condition.

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 Figure 4.4 (right), the primary benefit of the IntelWiki system came for those who experienced this condition second. Those who edited with IntelWiki first, wrote about roughly the same number of facts in each condition. I suspect that in this latter case, IntelWiki helped the 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, Figure 4.3 (right) shows that trend was in IntelWiki's favour with 15.9 (s.e. 0.31) different accurate facts with IntelWiki as compared to 14.6 (s.e. 0.31) different accurate facts with Default interface ($F_{1,14}$ = 4.520, p = 0.052, η^2 = 0.244). For this measure, too, 11 out of 16 participants documented more accurate facts in IntelWiki condition than in the Default condition, while one being equal in both conditions (see Figure 4.5).

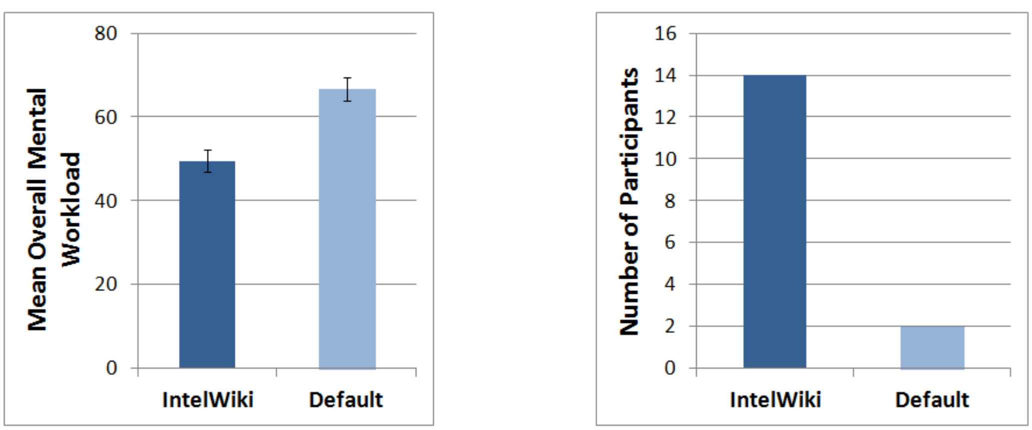

Figure 4.6: (Left) Cumulative Mental Workload by condition. (Right) Participants' interface preferences.

4.7.3 Perceived Mental Workload

According to the results of the NASA-TLX, participants reported experiencing significantly lower mental workload overall (see Figure 4.6 (left) for the cumulative total) when using IntelWiki (49.5, s.e. 2.69) than when using the Default interface (66.7, s.e. 2.69, $F_{1,14} = 10.212$, p = 0.006, $\eta^2 = 0.422$). Cohen's d was measured as 0.877. Table 4.1 illustrates that there were significant differences between the two conditions for all but two of the individual subscales (Mental Demand and Performance).

Table 4.1: Participants' perceived mental workload according to the NASA-TLX subscales (20-point scale).

Mental Workload	**Mean**		**F(1,14)**	**p**	η^2
	IntelWiki	**Default**			
Mental Demand	10.8	12.7	1.487	0.243	0.096
Physical Demand	8.4	10.8	5.790	0.031	0.293
Temporal Demand	9.9	12.7	6.803	0.021	0.327
Performance	6.3	8.5	2.991	0.106	0.176
Effort	10.0	14.9	11.63	0.004	0.454
Frustration	3.9	7.9	12.94	0.003	0.480

4.7.4 Subjective Preference

On the post-session questionnaire, participants responded positively both to IntelWiki's individual features and the system in comparison to the default Wikipedia interface. Table 4.2 suggests that participants appreciated the reference ordering, the tooltips and the keyword highlighting, and did not find these features distracting. In terms of overall preference, 14 out of the 16 participants preferred the IntelWiki interface over the Default one ($\chi^2 = 9.000, p = .003$). No participants gave a neutral response (see Figure 4.6 (right)).

Table 4.2: Participants' responses to a number of statements regarding IntelWiki's design. (Strongly Disagree = 1, Strongly Agree = 5).

Statement	**Mean**	**Std.**
Ordering the reference materials based on the keywords was useful.	4.44	0.51
Keywords and their associated count containing tooltips were useful.	4.44	0.73
The tooltips were distracting.	1.88	0.81
Highlighting the keywords was useful.	4.13	0.81
Highlighting the keywords was distracting.	1.88	0.81

4.7.5 Results with Window Placement

One might argue that a control condition where users could search and edit side-by-side (in the default condition) would have made for a more compelling experiment. I note that participants could place their windows in whatever configuration they felt most comfortable with, and according to the videotapes, 6/16 participants chose a side-by-side configuration. Even with this small sample size, Fact Count remained significant in IntelWiki's favour: 18.33 (s.e. 0.118) different facts with IntelWiki as compared to 16.83 (s.e. 0.118) with Default ($F_{1,4}$ = 40.5, p = 0.003, η^2 = 0.910). Although the difference in Word Count was not significant (p = 0.746), trends were in IntelWiki's favour for (1) overall mental workload: 49.83 (s.e. 4.12) in IntelWiki condition as compared to 72.5 (s.e. 4.12) in Default condition ($F_{1,4}$ = 7.56, p =

0.051, $\eta^2 = 0.654$), and (2) Fact Accuracy: 16.0 (s.e. 0.4) different accurate facts with IntelWiki as compared to 14.67 (s.e. 0.4) different accurate facts with Default interface ($F_{1,4} = 2.78$, p = 0.171, $\eta^2 = 0.410$).

4.8 Qualitative Data

In the semi-structured exit interviews, I elicited participants' impressions of the IntelWiki system, including what they liked and did not like about its approach. As I describe below, participants liked the system's integrated environment, the manner in which it supported resource inspection and evaluation, and its ability to save them from doing their own internet searches. Participants, however, expressed mixed views as to their willingness to trust and rely upon the system's support.

4.8.1 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 quotes illustrate:

> *I preferred [IntelWiki] because the screen was shared. [...] Yes, it was very very very advantageous. [...] 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*

> *[I preferred IntelWiki] because it gives me everything in one window. - P7*

> *In the default case I was always skeptical about how much correct information I could transfer from [the reference material] to [the article being edited], whereas in this interface it is opened before my eyes in the same screen. - P8*

4.8.2 Supporting Resource Inspection and Evaluation

Participants also liked the ability to quickly inspect the recommended resources through the tooltips to evaluate whether or not an individual resource would meet their current editing needs prior to opening it.

> *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*

> *The keywords in the tooltip helped me understand which article will be more suitable for geography. - P7*

> *I liked the ranking portion [with the tooltips]. It tells you which one has which keywords. So obviously if I click one of them I find out information about those keywords. It is helpful. - P15*

Similarly, participants liked the ability to find relevant information from an opened resource through the keyword highlighting:

> *When I looked into the resources I would always look for those keywords [...] If those keywords were highlighted I could go easily to those points and find*

whether the relevant information is there. So it helped me to locate and find the information very quickly. - P1

[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

4.8.3 Replacing Independent Search

Most participants responded positively to the idea of system recommended resources, with a number of participants commenting that it saved them the effort of performing the searches themselves. For example:

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

[IntelWiki] helped me a lot, because I don't have to go to Google search. It gives me the keywords, and I just click on [the links] and get my resources. - P9

One participant felt that providing editors with recommended resources could help improve Wikipedia's quality (and perception of its quality), since it would help editors focus on pertinent information:

It was a very wonderful interface. I was thinking about it and I was like if this can actually be incorporated into Wikipedia properly then I think, so many people will no longer have the problem of disregarding Wikipedia for not being a scholarly source. Yes, it is going to be scholarly but then we need to limit

> *certain information to just what is needed in that particular pane [Suggested Resources], so that we do not have a lot of things distracting the reader at the same time. - P12*

Not all participants, however, felt that IntelWiki would completely remove the need to search for resources as the recommendations would not always be sufficient or accurate:

> *The [suggested] resources point out most of the information but I feel, to efficiently contribute to the articles in Wikipedia, you have to also, on your own find more information. - P4*

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

4.8.4 Lack of Trust

For the two participants who preferred the default design, their preference seemed to be primarily based on lack of trust that an intelligent system could consistently provide an appropriate set of resources:

> *I don't usually trust the highlight and these things. I prefer to study by myself. - P6*

> *I preferred the [Default interface] because I can get more facts – I don't really have to base my knowledge on a limited set of resources. It makes me more elaborate in my research. [IntelWiki] too, on the other hand is good based on the fact that if you don't have any understanding of geography then you won't know what to look for, right? You just have the resources provided for you and then you go through them and see what you could make out of them. - P10*

In the quote above, P10 felt that the system's set of recommended resources would not likely be comprehensive enough, preferring to search independently. She did feel, however, that the IntelWiki system could be beneficial for somebody with limited knowledge of the article's topic.

4.9 Discussion

The proof-of-concept evaluation described above provides encouraging evidence in favour of IntelWiki's approach. With editing time fixed, participants were able to contribute significantly more text and experienced significantly lower mental workload 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). Specifically, an interaction effect suggested that the scaffolding IntelWiki provided to participants who interacted with it first, carried over to when the support was removed. Participants also expressed a strong preference for IntelWiki's support.

The evaluation demonstrated that IntelWiki's overall approach improves editing performance and subjective impressions as compared to the default editor, but did not isolate the impacts of the different aspects of the system's design. Isolating these impacts (e.g., integrated editing, resource retrieval, resource ranking, keyword previews, keyword highlighting) in a single study would have been extremely challenging given my desire to use a within-subjects design, which I felt was important due to its ability to 1) account for individual variability and 2) elicit comparative statements.

The analyses of most of my dependent measures produced statistically significant results in favour of IntelWiki. But, as the size of the differences in between the means for the two conditions (IntelWiki and Default) for the different measures indicate, the practical significances of the results were relatively low except for the mean overall mental workload. However, I expect to see a larger impact of this system on user performance in case of a live deployment, due to the following reasons: Firstly, during the tasks the participants were instructed to produce the best pieces of text that they could within the allotted time in each condition. All the participants were compensated for their times and efforts with a promise of additional benefit to the top performers. This would have motivated the participants to perform at their best regardless of the condition. In contrast, in a real world scenario Wikipedians contribute voluntarily without any remuneration. Therefore, the individual differences between the amount of contributions produced by IntelWiki and the Default design within a fixed time are more likely to increase with IntelWiki extracting higher amounts of contributions by reducing the editing effort. Secondly, in my study each task duration was 25 minutes. While a lot of participants produced considerable amount of texts, few, if any, could actually complete the section. While writing an actual article in Wikipedia, even 25 minutes could prove short for writing even one section of an article. According to my results, the participants' experienced mental workload was quite higher in the Default condition in comparison to the IntelWiki condition. Hence, In a real world scenario IntelWiki would probably be able to hold a user's attention span for a longer time in comparison to the default Wikipedia, and a more complete piece of text could be produced with IntelWiki.

The interview revealed that participants were enthusiastic about the system's integrated environment, the manner in which the system helped them evaluate potential resources at a glance and how the system directed their attention to relevant parts of the articles. Participants' comments also suggested that the system's recommended resources should be used as a complement to other research strategies rather than as a replacement, since a number of participants indicated that they would not be comfortable delegating that responsibility fully to the system. However, during the task with IntelWiki condition, the participants relied almost exclusively on IntelWiki's resources, with only one participant performing a single external search. Thus, while some participants wanted the option to search independently, this was rarely done in the study. There is still room, however, to further extend the system's mixed-initiative capabilities to allow the user to incorporate his/her independently discovered resources.

While the results of this evaluation are an important first step in exploring the utility of such an approach, further evaluations are required to determine the generalizability of the results. 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. This study design decision raises two potential questions. The first concerns the feasibility of using either the Wikipedia community or machine learning approaches to generate such a list. My evaluation provides support for further exploration on this topic. Second, my results do not speak of the value of the system's support in the case of less accurate and specific resource rankings and keyword highlighting.

To control for participant expertise, while still giving me access to a wide enough participant pool, I asked users to edit articles on topics that they were familiar with (i.e., popular lakes), but not for which they were experts. Exploring the value of the system with participants having more article-related expertise is an important area of future study. One question in this regard is how IntelWiki might impact editing confidence. Bryant et al. found that initially novice editors edit articles on topics that they are experts in, but eventually branch out to a wider range of articles [17]. It would be interesting to explore whether or not IntelWiki helps facilitate this transition.

Finally, given the laboratory nature of the study I also am not able to determine if the system's support does in fact lead to increased participant rates. Such an investigation would require a field deployment.

4.10 Summary

In this chapter I presented the study that I conducted to evaluate the IntelWiki system against a baseline (i.e., default Wikipedia interface and Google Search engine), which involved measuring the editing performance and experience of several human participants in these two conditions. According to the results, the participants' performance was significantly better in the IntelWiki condition in terms of amount of text produced, and the IntelWiki successfully reduced their mental workload to a significant extent. The participants in general liked the IntelWiki system compared to the default Wikipedia interface, and the individual IntelWiki features were well-received.

Chapter 5

Conclusion

Despite the surge in user-generated content (UGC) in recent years, research shows that a small percentage of the online community are active contributors [7, 13, 33, 34, 38, 43, 54]. Research has also shown that reducing the editing effort has the potential to increase UGC contribution rates [17, 27, 50]. This thesis proposes the IntelWiki system, which aims to reduce the Wikipedia editing effort by automatically suggesting article-relevant resource materials. An evaluation of the IntelWiki system against the default Wikipedia editor shows that the IntelWiki system is successful in improving user performance and reducing the perceived mental workload.

5.1 Contributions

The first contribution of my thesis is a general approach of helping Wikipedia contributors in editing an article by automatically recommending article-relevant online resources within the Wikipedia interface. This approach intends to reduce the time and effort a contributor employs in searching for and consulting information online. To my knowledge, this approach of making editing Wikipedia (and any UGC system, in general) easier by means of integrating relevant resource materials is novel.

The second contribution of my thesis is the design and implementation of the general approach, embedded in the IntelWiki prototype. The IntelWiki system makes the automatic suggestions using Google Custom Search Engine (CSE) API, but instead of displaying the Google suggested resource list as is, IntelWiki re-ranks the list based on a set of pertinent keywords. The IntelWiki system allows a user to browse and examine the resource materials within the context of the Wikipedia editor.

The third contribution of my thesis is a formal laboratory evaluation with 16 participants, exploring the potential for my approach to ease the editing burden in comparison to the default Wikipedia editor. My 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. Subjectively, participants reported experiencing significantly lower mental workload and all but two of the 16 participants preferred IntelWiki's approach.

5.2 Limitations and Future Work

The general approach presented in my thesis is a resource recommendation strategy which emulates a possible natural work-flow of a new Wikipedia contributor (i.e., search for article-relevant online resources, rank based on a quick assessment of their information richness, and consult in the order of the ranks). While the IntelWiki prototype is an important first step in the manifestation of this general approach, it has certain limitations and has opportunities for enhancement in a number of areas.

The IntelWiki employs Google search technique using the article title as a search query. While this produces a reasonable set of relevant resources, using a larger set of search terms (e.g., the pertinent keywords, described in system description (section 3.2)) could have the potential to enhance the set of relevant resources. Since my evaluation revealed that users want the ability to search for their own resources to augment IntelWiki's recommended set, I would like to provide support for this process directly within the IntelWiki system. Promising areas for future work also include exploring ways to determine sets of pertinent keywords to guide IntelWiki's assessment of article relevance, including machine learning approaches and ways to elicit this information from the Wikipedia community. In addition to eliciting feedback from the community on relevant keywords, I would also like to explore ways to elicit and incorporate community feedback on the utility of IntelWiki's resource recommendations.

Due to the laboratory nature of the study it was not possible to determine if the system's support leads to increased participation rates. This could be an interesting study to conduct after deploying the system online. The study I conducted involved novice Wikipedia editors who had little to no knowledge of Wikipedia tools, policies and guidelines. Therefore, I would like to explore the value of IntelWiki's suggestions with existing Wikipedia contributors and subject-matter experts.

Finally, it would be interesting to explore whether or not the resource recommendation strategy followed by IntelWiki can be generalized to environments outside Wikipedia (e.g., writing articles/blogs in online communities, and writing research papers/essays using word processing software). The Google Search technique used to fetch relevant resources could be incorporated directly, while the search queries could be streamlined to produce more specific sets of resources. However, the "pertinent keywords"-based resource ranking process used in IntelWiki would require further research due to its dependency on the infoboxes which are native to Wikipedia. Also, the manner to present the reference materials to the user would need to be tailored to the particular environment to ensure interface integrity.

Bibliography

[1] Alexa ranking of wikipedia. http://www.alexa.com/siteinfo/wikipedia.org. Accessed: 2013/02/25.

[2] Dom storage guide - web developer guide |mozilla developer network. https://developer.mozilla.org/en-US/docs/Web/Guide/API/DOM/Storage. Accessed: 2013/10/22.

[3] Google apis client library for php. https://code.google.com/p/google-api-php-client/. Accessed: 2013/10/22.

[4] Lake winnipegosis - wikipedia. http://en.wikipedia.org/wiki/Lake_Winnipegosis. Accessed: 2014/03/10.

[5] Linked data - connect distributed data across the web. http://linkeddata.org/. Accessed: 2013/03/04.

[6] W3c semantic web activity homepage. http://www.w3.org/2001/sw/. Accessed: 2014/01/14.

[7] Wikimedia report card. http://reportcard.wmflabs.org/. Accessed: 2013/12/02.

[8] Wikipedia - wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Wikipedia. Accessed: 2014/04/16.

[9] Wikipedia: Infobox. http://en.wikipedia.org/wiki/Help:Infobox. Accessed: 2013/02/25.

[10] Wikipedia: Modelling wikipedia's growth. http://en.wikipedia.org/wiki/Wikipedia:Modelling_Wikipedia's_growth. Accessed: 2013/02/25.

[11] Wikipedia: Please do not bite the newcomers. http://en.wikipedia.org/wiki/Wikipedia:Please_do_not_bite_the_newcomers. Accessed: 2013/02/25.

[12] Adar, E., Skinner, M., and Weld, D. S. Information arbitrage across multi-lingual wikipedia. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining* (2009), ACM, pp. 94–103.

[13] Arthur, C. What is the 1% rule? http://www.guardian.co.uk/technology/2006/jul/20/guardianweeklytechnologysection2/. Accessed: 2013/02/25.

[14] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. Dbpedia: A nucleus for a web of open data. *The Semantic Web* (2007), 722–735.

[15] Bollacker, K., Evans, C., Paritosh, P., Sturge, T., and Taylor, J. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data* (2008), ACM, pp. 1247–1250.

[16] Bostandjiev, S., O'Donovan, J., Hall, C., Gretarsson, B., and Hllerer, T. Wigipedia: A tool for improving structured data in wikipedia. In *Proceedings of the 5th IEEE International Conference on Semantic Computing (ICSC 2011)* (2011), pp. 328–335.

[17] Bryant, S. L., Forte, A., and Bruckman, A. Becoming wikipedian: Transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work* (2005), pp. 1–10.

[18] Chen, S. C. Php simple html dom parser. `http://simplehtmldom.sourceforge.net/`. Accessed: 2013/10/22.

[19] Clerck, J.-P. D. Content marketing: What content people share and why they do it. `http://www.conversionation.net/2011/05/content-marketing-what-content-people-share-and-why-they-do-it/`. Accessed: 2013/02/25.

[20] Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. Using intelligent task routing and contribution review to help communities build artifacts of lasting value. In *Proceedings of the 2006 Conference on Human Factors in Computing Systems* (2006), pp. 1037–1046.

[21] Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. Suggestbot: Using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 2007 International Conference on Intelligent User Interfaces* (Jan. 2007), pp. 32–41.

[22] Daugherty, T., Eastin, M. S., and Bright, L. Exploring consumer motivations for creating user-generated content. *Journal of Interactive Advertising 8*, 2 (2008), 1–24.

[23] Giles, C. L., Bollacker, K. D., and Lawrence, S. Citeseer: An automatic citation indexing system. In *Proceedings of the 3rd ACM International Conference on Digital Libraries* (Jun. 1998), pp. 89–98.

[24] 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. In *Proceedings of the Fifth International Conference on Communities and Technologies* (2011), pp. 98–107.

[25] Harford, T. Why does anyone bother contributing to wikipedia? `http://www.ft.com/cms/s/2/9b1281be-c06b-11df-8a81-00144feab49a.html#axzz2JgUuUnII`. Accessed: 2013/02/25.

[26] 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.

[27] Hoffmann, R., Amershi, S., Patel, K., Wu, F., Fogarty, J., and Weld, D. S. Amplifying community content creation with mixed initiative information extraction. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (Apr. 2009), pp. 1849–1858.

[28] Howell, C. Why do people share online? http://socialmediatoday.com/cory-howell/353846/why-do-people-share-online. Accessed: 2013/02/25.

[29] Kittur, A., Chi, E., Pendleton, B. A., Suh, B., and Mytkowicz, T. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In *Proceedings of alt. CHI 2007* (2007).

[30] Kong, N., Hanrahan, B. V., Weksteen, T., Convertino, G., and Chi, E. H. Visualwikicurator: Human and machine intelligence for organizing wiki content. In *Proceedings of the 2011 International Conference on Intelligent User Interfaces* (2011), pp. 367–370.

[31] Krieger, M., Stark, E., and Klemmer, S. R. Coordinating tasks on the commons: Designing for personal goals, expertise and serendipity. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (Apr. 2009), pp. 1485–1494.

[32] Krumm, J., Davies, N., and Narayanaswami, C. User-generated content. *Pervasive Computing, IEEE 7*, 4 (Oct. 2008), 10–11.

[33] McConnell, B., and Huba, J. The 1% rule: Charting citizen participation. http://www.churchofthecustomer.com/blog/2006/05/charting_wiki_p.html. Accessed: 2013/02/25.

[34] Nielsen, J. Participation inequality: Encouraging more users to contribute. http://www.nngroup.com/articles/participation-inequality/. Accessed: 2013/02/25.

[35] Nov, O. What motivates wikipedians? In *Communications of the ACM*, vol. 50. Nov. 2007, pp. 60–64.

[36] Okuoka, T., Takahashi, T., Deguchi, D., Ide, I., and Murase, H. Labeling news topic threads with wikipedia entries. In *Multimedia, 2009. ISM'09. 11th IEEE International Symposium on* (2009), IEEE, pp. 501–504.

[37] Oreg, S., and Nov, O. Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Computers in Human Behavior 24*, 5 (2008), 2055–2073.

[38] Priedhorsky, R., Chen, J., Lam, S. T. K., Panciera, K., Terveen, L., and Riedl, J. Creating, destroying, and restoring value in wikipedia. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work* (2007), ACM, pp. 259–268.

[39] Ruderman, J. Same-origin policy - javascript |mozilla developer network. `https://developer.mozilla.org/en-US/docs/Web/JavaScript/Same_origin_policy_for_JavaScript`. Accessed: 2013/10/22.

[40] Sauper, C., and Barzilay, R. Automatically generating wikipedia articles: A structure-aware approach. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1* (2009), Association for Computational Linguistics, pp. 208–216.

[41] Schroer, J., and Hertel, G. Voluntary engagement in an open web-based

encyclopedia: Wikipedians and why they do it. *Media Psychology 12*, 1 (2009), 96–120.

[42] Solomon, J., and Wash, R. Bootstrapping wikis: Developing critical mass in a fledgling community by seeding content. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (2012), ACM, pp. 261–264.

[43] Stuckman, J., and Purtilo, J. Measuring the wikisphere. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (2009), ACM, p. 11.

[44] Vale, P. Wikipedia's jimmy wales speaks out on china and internet freedom. `http://www.huffingtonpost.co.uk/2011/08/29/wikipedias-jimmy-wales-sp_n_941239.html`. Accessed: 2013/02/25.

[45] Vrandecic, D., Ratnakar, V., Krtzsch, M., and Gil, Y. Shortipedia aggregating and curating semantic web data. *Journal of Web Semantics 9*, 3 (Sep. 2011), 334–338.

[46] Weld, D. S., Hoffmann, R., and Wu, F. Using wikipedia to bootstrap open information extraction. In *SIGMOD Record*, vol. 37. Dec. 2008, pp. 62–68.

[47] Weld, D. S., Wu, F., Adar, E., Amershi, S., Fogarty, J., Hoffmann, R., Patel, K., and Skinner, M. Intelligence in wikipedia. In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence* (2008), pp. 1609–1614.

[48] Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., and Smith, M. Finding social roles in wikipedia. In *Proceedings of the 2011 iConference* (2011), ACM, pp. 122–129.

[49] Wick, M. Geonames, 2008.

[50] Wilkinson, D. M. Strong regularities in online peer production. In *Proceedings of The 9th ACM Conference on Electronic Commerce* (2008), ACM, pp. 302–309.

[51] Woodsend, K., and Lapata, M. Wikisimple: Automatic simplification of wikipedia articles. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence* (2011), pp. 927–932.

[52] Wu, F., Hoffmann, R., and Weld, D. S. Information extraction from wikipedia: Moving down the long tail. In *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining* (2008), pp. 731–739.

[53] Wu, F., and Weld, D. S. Autonomously semantifying wikipedia. In *Proceedings of the ACM Conference on Information and Knowledge Management* (2007), pp. 41–50.

[54] Zachte, E. Wikipedia statistics. `http://stats.wikimedia.org/EN/TablesWikipediaZZ.htm#editor_activity_levels`. Accessed: 2013/12/02.

[55] Zachte, E. Wikipedians who edited at least 10 times since they arrived. `http://stats.wikimedia.org/EN/TablesWikipediansContributors.htm`. Accessed: 2013/02/25.

Appendix A

Low-Fidelity Prototypes

Prototype Interface 1 In this prototype, both the “Suggested Resources” Pane and the “Resource Viewer” Pane were placed in the same tooltip, that displayed the callout. The prototype is displayed in Figure A.1 below.

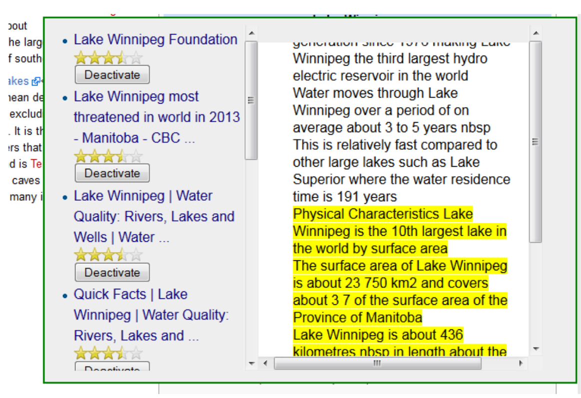

Figure A.1: Prototype Interface 1: Both Panes in tooltip

Prototype Interface 2 In this prototype, the “Suggested Resources” Pane was placed on the left sidebar navigation panel. The “Resource Viewer” Pane would appear in a popup window on clicking a resource title from the “Suggested Resources” Pane (displayed in Figure A.2).

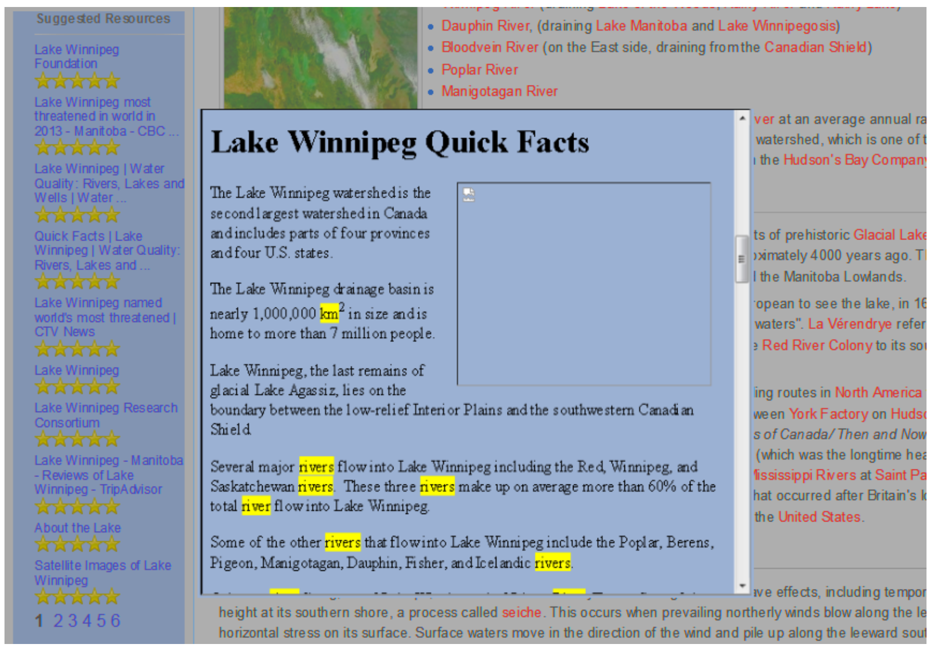

Figure A.2: Prototype Interface 2: “Resource Viewer” Pane as popup

Both prototype interface 1 and prototype interface 2 conformed to the first two design principles (listed in subsection 3.3.1), but somewhat violated the third design principle. Neither of the interface variants could present the “Resource Viewer” Pane persistently.

Prototype Interface 3 In this prototype, both the "Suggested Resources" Pane and the "Resource Viewer" Pane were added to the regular Wikipedia interface on the left and on the right of the article content Pane respectively, thereby shrinking the article to make room. It is displayed in Figure A.3.

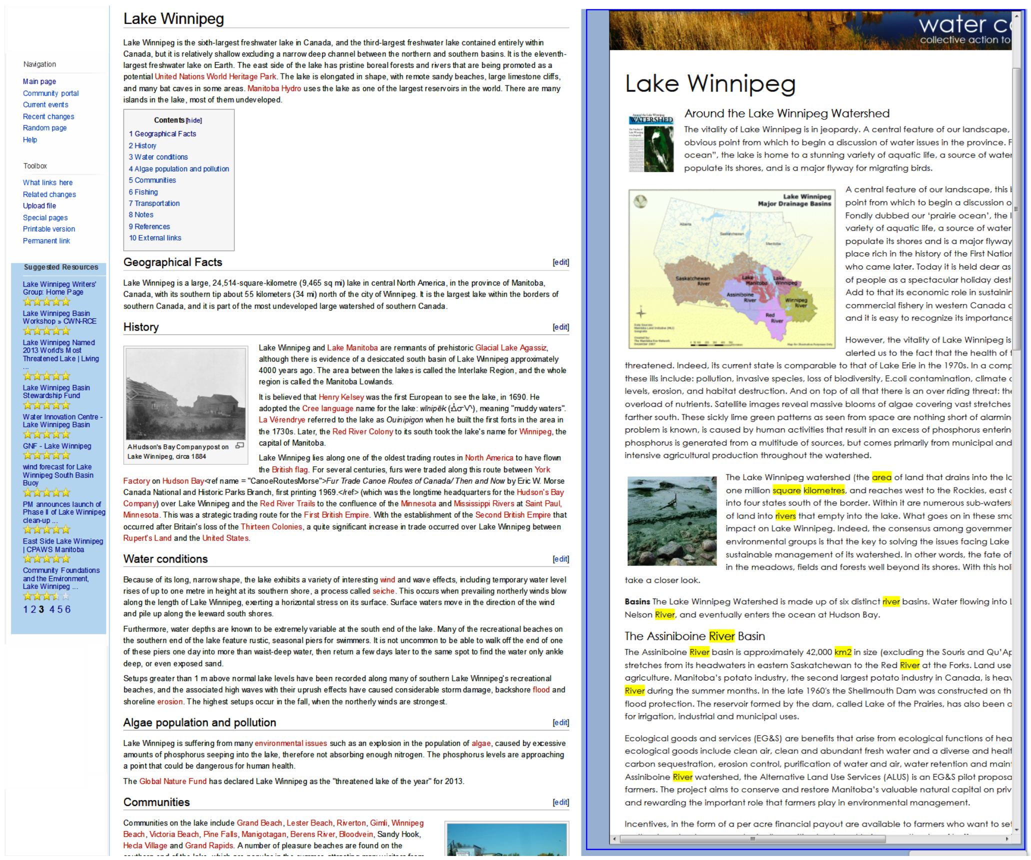

Figure A.3: Prototype Interface 3: Both the "Suggested Resources" Pane and the "Resource Viewer" Pane in fixed positions in the interface

The prototype interface 3 made it possible to display all the information sources persistently across different modes (View and Edit), while conforming to the first two design goals as well. As Figure 3.4 shows, the final IntelWiki interface is a slightly modified version of the prototype interface 3.

Prototype Interface 4 The only difference this prototype had with the above mentioned Prototype Interface 3 was the number of resource viewer panes. The reason behind this design was to allow the user to open multiple resources at a time and compare the information s/he is about to put. The layout of the "Resource Viewer" panes in this prototype is demonstrated by Figure A.4.

Figure A.4: Prototype Interface 4: Multiple "Resource Viewer" Panes

Similar to the prototype interface 3, prototype interface 4, too, conformed to all three design principles.

Appendix B

Ethics Approval Certificate

UNIVERSITY OF MANITOBA | Research Ethics and Compliance
Office of the Vice-President (Research and International)

Human Ethics
208-194 Dafoe Road
Winnipeg, MB
Canada R3T 2N2
Phone +204-474-7122
Fax +204-269-7173

APPROVAL CERTIFICATE

August 15, 2013

NSERC
35353

TO:	**Mohammad Noor Nawaz** Principal Investigator	(Advisor A. Bunt)
FROM:	**Susan Frohlick, Chair** Joint-Faculty Research Ethics Board (JFREB)	
Re:	**Protocol #J2013:114** **"MyIntelWiki: Intelligence-Driven User Generated Content Contribution Acceleration"**	

Please be advised that your above-referenced protocol has received human ethics approval by the **Joint-Faculty Research Ethics Board**, which is organized and operates according to the Tri-Council Policy Statement (2). **This approval is valid for one year only**.

Any significant changes of the protocol and/or informed consent form should be reported to the Human Ethics Secretariat in advance of implementation of such changes.

Please note:

- **If you have funds pending human ethics approval, please mail/e-mail/fax (261-0325) a copy of this Approval (identifying the related UM Project Number) to the Research Grants Officer in ORS in order to initiate fund setup. (How to find your UM Project Number: http://umanitoba.ca/research/ors/mrt-faq.html#pr0)**
- **if you have received multi-year funding for this research, responsibility lies with you to apply for and obtain Renewal Approval at the expiry of the initial one-year approval; otherwise the account will be locked.**

The Research Quality Management Office may request to review research documentation from this project to demonstrate compliance with this approved protocol and the University of Manitoba *Ethics of Research Involving Humans*.

The Research Ethics Board requests a final report for your study (available at: http://umanitoba.ca/research/orec/ethics/human_ethics_REB_forms_guidelines.html) **in order to be in compliance with Tri-Council Guidelines.**

umanitoba.ca/research

Figure B.1: Ethics Approval Certificate

Appendix C

Instructions to the Participants

Task: Please Write the empty "Geographical Facts *" section of the selected article with the given interface. Please use the system suggested resources as primary reference materials, while working with interface A.

Table C.1: Sample Facts You Can Look For

* Geographical Facts may contain	Synonyms/Keywords/Units
Average/maximum length	long, kilometre, km, metre
Average/maximum width	wide, metre, feet
Surface area	square km
Average/maximum depth	deep, metre, feet
Shore length	shoreline
Water volume	cubic kilometre/metre
Surface elevation	altitude
Residence time	retention time
Connected rivers	watershed, tributaries

Appendix D

Post-Study Questionnaire

1. Please indicate the extent to which you agree or disagree with the following statements. All questions refer to features of the interface A.

Table D.1: Post User Study Questionnaire

Statement	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
Ordering the reference materials based on the keywords was useful					
Listing the contained keywords and their counts in tooltip was useful					
The tooltips containing keywords and their counts distracted me from my task					
I preferred the interface featuring suggested resources and a viewer pane over the regular Wikipedia interface and Google Search					
Highlighting the keywords in the resource viewer pane was useful					
Highlighting the keywords in the resource viewer pane distracted me from my task					

2. For the interfaces below please indicate which was your more preferred one:

   - Interface A: Suggested resources + Resource Viewer pane
   - Interface B: Regular Wikipedia edit interface + Google Custom Search Interface