Individual Project Report

INTRODUCTION

Organization: Wikipedia

Aspect: Sharing

This report has four sections. The research section aim to answer the question "How did you study your selected aspect? Please be as detailed as possible." It introduces the research methods, references and interview questions. The Finding and Discussion section talks about the findings and discusses about each finding. The finding section answers the question "What did you find?"; the conclusion section try to explain "What does it mean?"; the further discussion covers the question "What can we learn from?". Finally, the last section will have a discussion on "What can be improved?"

Different from other projects in our class, the different aspects of Wikipedia are closely related to each other. The Five pillars rule make a clear statement:

- Wikipedia is an encyclopedia: It combines many features of general and specialized encyclopedias, almanacs, and gazetteers.

- Wikipedia is written from a neutral point of view: We strive for articles that document and explain the major points of view, giving due weight with respect to their prominence in an impartial tone. We avoid advocacy and we characterize information and issues rather than debate them.

- Wikipedia is free content that anyone can use, edit, and distribute: Since all editors freely license their work to the public, no editor owns an article and any contributions can and will be mercilessly edited and redistributed.

- Editors should treat each other with respect and civility: Respect your fellow Wikipedians, even when you disagree.

- Wikipedia has no firm rules: Wikipedia has policies and guidelines, but they are not carved in stone; their content and interpretation can evolve over time.

These five rules create a great free space for Wikipedians to share their knowledge. Every piece of information on Wikipedia are born to be shared. It is hard to clearly distinguish any of the aspects from each other as they are all serving one purpose, to share human knowledge across language, country and race.

RESEARCH

This individual study focus on the sharing aspect of Wikipedia. The two main methods I use are literature review and Content analysis.

Besides enormous of information users daily search and read on Wikipedia, Wikipedia hold rich information about itself and its operator, Wikipedia Foundation, and the participant community Wikipedians. Wikipedia: about page(http://en.wikipedia.org/wiki/Wikipedia:About#Frequently_asked_questions_.28FAQ.29) gives a detailed introduction

to Wikipedia itself, from its founding, value, contribution to challenges and criticisms. So I spend the most time review the content within Wikipedia. It is interesting to see that Wikipedia provide quite rich and new information to each section of the about page. So after reading each section, I usually search this section on the Wikipedia, and it will lead me to other independent pages that contain much more detail information on this single point. So use About page as starting point, I review several pages that specific on sharing aspect. Wikipedia: Wikipedians is community page for Wikipedias. Wikipedians are people who write and edit the pages for Wikipedia, unlike readers who simply read the articles. They are the core force that create the information on Wikipedia, and keep information updated. They are also the source that the information shared from. I enter several Wikipedians community website, to see the daily discussion.In order to know the difference between difference language version, I select the item "iPhone", "Jintao Hu", "UT Austin", and switch them between English, Simplified Chinese and Traditional Chinese. I conduct a content analysis on the content of different language version of different items. First, I compared the sections that each version have, then go into the content to check the consistency across different version.

To acknowledge the previous research on sharing aspect of Wikipedia, I reviewed some previous research. Starting with simple search "Wikipedia+Share" on Library of University of Texas at Austin, it returns 1,824 Journal Article and Conference Proceeding. Then I do the same research on ACM digital library, it shows the valuable result is quite limited. I undo a search using a single word "Wikipedia" on ACM Digital Library, and it returns 22,084 results. I selected the results by the title and abstract. The list follows:

Stacey Kuznetsov. 2006. Motivations of contributors to Wikipedia. SIGCAS Comput. Soc. 36, 2, Article 1 (June 2006). DOI=10.1145/1215942.1215943 http:// doi.acm.org.ezproxy.lib.utexas.edu/10.1145/1215942.1215943

Shu-Mei Tseng and Jiao-Sheng Huang. 2011. The correlation between Wikipedia and knowledge sharing on job performance. Expert Syst. Appl. 38, 5 (May 2011), 6118-6124. DOI=10.1016/j.eswa.2010.11.009

Heng-Li Yang and Cheng-Yu Lai. 2010. Motivations of Wikipedia content contributors. Comput. Hum. Behav. 26, 6 (November 2010), 1377-1383. DOI=10.1016/j.chb. 2010.04.011 http://dx.doi.org.ezproxy.lib.utexas.edu/10.1016/j.chb.2010.04.011

Ranjitha Gurunath Kulkarni, Gaurav Trivedi, Tushar Suresh, Miaomiao Wen, Zeyu Zheng, and Carolyn Rose. 2012. Supporting collaboration in Wikipedia between language communities. In Proceedings of the 4th international conference on Intercultural Collaboration (ICIC '12). ACM, New York, NY, USA, 47-56. DOI=10.1145/2160881.2160890 http://doi.acm.org.ezproxy.lib.utexas.edu/ 10.1145/2160881.2160890

Patti Bao, Brent Hecht, Samuel Carton, Mahmood Quaderi, Michael Horn, and Darren Gergle. 2012. Omnipedia: bridging the Wikipedia language gap. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12). ACM, New York, NY, USA, 1075-1084. DOI=10.1145/2207676.2208553 http:// doi.acm.org.ezproxy.lib.utexas.edu/10.1145/2207676.2208553

Raghavendra Udupa and Mitesh Khapra. 2010. Improving the multilingual user experience of Wikipedia using cross-language name search. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT '10). Association for Computational Linguistics, Stroudsburg, PA, USA, 492-500.

Morten Warncke-Wang, Anuradha Uduwage, Zhenhua Dong, and John Riedl. 2012. In search of the ur-Wikipedia: universality, similarity, and translation in the Wikipedia inter-language link network. In Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration (WikiSym '12). ACM, New York, NY, USA, , Article 20 , 10 pages. DOI=10.1145/2462932.2462959 http://doi.acm.org.ezproxy.lib.utexas.edu/10.1145/2462932.2462959

As a complementary to the literature review, I conduct a short interview via email and phone with friends back in China. The interview aims to analysis the usage of Wikipedia in region that not use English as first language, and find the reason that prevent or promote information sharing of Wikipedia in these region based on the feedback to first question. The questions are:

• Do you know Wikipedia? (continue if answer is yes)

• How often do you use Wikipedia?

• Have you ever edit any content on it?

• What is the first thing that prevent and promote you use Wikipedia?

• Do you have any alternative to Wikipedia?

After one-week period, it had 12 participants. All of the participants have Bachelor's degree and can communicate in English.

FINDING AND DISCUSSION

Language Diversity

Finding

The biggest effort the Wikipedia has made is the multilanguage support to most mainstream item.

From Figure1, it shows Wikipedia have more 1,000,000 items available in Deutsch, English, and Spanish, etc. version, and part of them are supported in Japanese, Chinese and Arabic, etc. As a Chinese, it is interesting to see that some item are even supported in Chinese minor languages which are rarely known in China.

List of Wikipedias page of Wikimedia shows that it contains a list of all 287 languages for which official Wikipedias have been created under the auspices of the Wikimedia Foundation. Content in other languages is being developed at the Wikimedia Incubator; languages that meet certain criteria can get their wikis. Starting with english, which has total 4,633,377 active articles, the table display all the languages that have been

officially supported by Wikipedia. The bottom of this table is Herero at the position 287 that has 0 Articles, but with 4,321 times of editing.

Conclusion and Further Discussion

Multi-language support is a good enhancement for the sharing of Wikipedia. As the Wikipedia's purpose states, Wikipedia's purpose is to benefits readers by acting as an encyclopedia, a comprehensive written compendium that contains information on all branches of knowledge. Wikipedia was intended to be the largest, most comprehensive, and most widely-available encyclopedia ever written. It has greatly helped Wikipedia that multi-language support can attract editor from different language and culture background. It diverse the content on the Wikipedia, which will contribute the comprehension and availability of Wikipedia. Multilingual editor can also help Wikipedia to convert the English version to other language, which help the knowledge sharing.
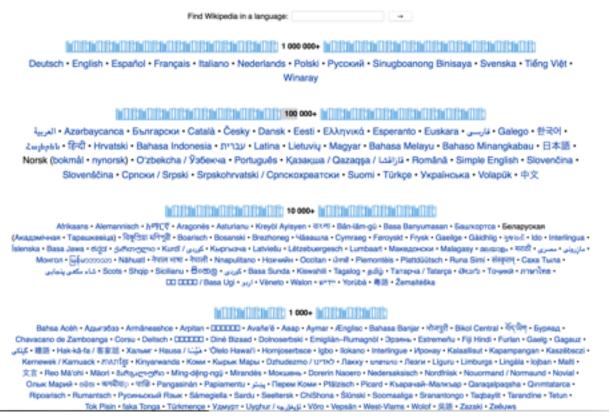

Table 1 Linguistic Asymmetry of Wikipedia

Languages Gap

Finding

It is exciting to see more than 200 languages have already been supported on Wikipedia. However, the unbalancing development of different have become a problem that have already made a negative impact on the spread of Wikipedia.

As Bao mentions, Wikipedia has over 270 language editions, dozens of which have more than 100,000 articles. Each language edition of the online encyclopedia has been shown to have a surprising amount of information not contained in any other language edition. It is true of both large Wikipedias like English and smaller ones such as Catalan and Finnish. Udupa also has also made a great illustration to this situation. He include a table in his paper, see Table 1.

Language	Speakers	Contributors	Articles
English	1500M	47.1%	3,072,373
Russian	278M	5.2%	441,860
Hebrew	10M	0.7%	97,987
Hindi	550M	0.06%	50,926
Bangla	230M	0.02%	20,342
Tamil	66M	0.04%	19,472
Kannada	47M	0.02%	7,185

Table 1 Linguistic Asymmetry of Wikipedia

Besides, the big amount difference between language, the content of the same item across different language is not same. It is quite obvious that English is the dominant language, so it is not surprising that the content of English is usually richer and updated more often than other languages. An example is an item "UTexas". The English version of "UTexas" has 11 sections, and each section may contain 2 to six subsection, which contribute to totally 9,685 words, including references. In comparison, the simplified Chinese version, it only contains 3,376 words. It is not perfectly accurate to say that English version have more contents than Chinese one since Chinese have distinct grammar and letter than English. However, I make some point that English content are more than Chinese. Anther noticeable fact is that the English version have 144 references listed, which is much more than 13 references on Chinese pages. The last modified time for English version is 27 October 2014, and the time for Chinese version is 11 August 2014.

According to the interview result, all 12 participants have heard and used Wikipedia. However, only 3 of them have used them more than once a week, in which only one of them are using it more than once a week. Moreover, none of them has ever edited the content on Wikipedia. Since all of the participants can communicate in English, they will

refer to English version when they read Chinese version. The reason they refer to English version is same as discussed, because English version is more updated and contains much more details. The top reason that prevent them using Wikipedia is Chinese version in lack of content. The best alternative to Wikipedia, according to the participant, is Hudong Baike(baike.com). It seems like the Chinese version of Wikipedia that focus on providing Chinese online encyclopedia.

Conclusion

Bao gives an explanation to it. "The language-induced splintering of information in Wikipedia poses both an opportunity and a challenge. On the one hand, Wikipedia embodies an unprecedented repository of world knowledge diversity in which each language edition contains its cultural viewpoints on a large number of topics. On the other hand, the language barrier serves to silo knowledge, slowing the transfer of less culturally imbued information between language editions and preventing Wikipedia's 422 million monthly visitors from accessing most of the information on the site."

In Warncke-Wang's page, he uses Tobler's law to explain the fact. As he mentioned, "The relationships between Wikipedia editions follow Tobler's first law of geography: similarity decreases with increasing distance. The number of articles in a Wikipedia edition is found to be the strongest predictor of similarity while language similarity also appears to have an influence. The English Wikipedia edition is by far the primary source of translations."

Language gap has become the primary fact that limits the wide-availability of Wikipedia, which wide-availability is the most component of Wikipedia's purpose. The Language gap limit that the sharing of information that created on Wikipedia platform. The information is constrained into English speaking region. In verse, the limit sharing will have a negative impact on the creation of information as the low-quality information will not attract user to contribute. As shown, 76% of contributors edit English version and 49% of users read English version, followed by the German Wikipedia as 20% and 12%, and Spanish Wikipedia as 12% and 6% respectively. More than half (51%) the editors contribute in two or more languages.

Further Discussion

Language gap is not problem only for Wikipedia, but many companies have faced the same problem. Some studies have already talked about potential solution.

Bao present Omnipedia, a system that allows Wikipedia readers to gain insight from up to 25 language editions of Wikipedia simultaneously. Omnipedia highlights the similarities and differences that exist among Wikipedia language editions, and makes salient information that is unique to each language as well as that which is shared more widely. We found that users actively sought information exclusive to unfamiliar language editions and strategically compared how language editions defined concepts.

From the perspective of computer science, Kulkarni presents a tool for supporting inter-cultural collaboration in Wikipedia. It describes an approach to checking the consistency of information contained in info boxes on pairs of pages devoted to the same topic but in

different language Wikipedias. The evaluation showed that there was a significant increase in recall after the concepts of homophones and synonyms were applied in addition to the direct text comparison. Two evaluations on two different language sets that cumulatively include English, German, Chinese, and Hindi, which have very different language structures, demonstrates that the proposed approach has some generality across languages. It can also be seen that the two domains considered namely Cities of India and Companies based in the United States were not similar and hence our method also has some generality across domains where the corresponding pages in Wikipedia include infoboxes.

Motivation of Sharing

Finding

During the research, I found many studies focus on the motivation of contributor. For Wikipedia, it is hard and unreasonable to separate sharing from creating information. In Wikipedia, the sharing and creation are embedded into each other, since every piece of information that created on Wikipedia are born be sharable. Visitor even without an account can review and edit the content with high freedom.

To secure the motivation of contribution, Wikipedia have open and restricted privacy policy in order to eliminate the worry of personal information exposure. As stated in the summary of the Privacy of Policy, "We believe that you should not have to provide personal information to participate in the free knowledge movement. You do not have to provide things like your real name, address, or date of birth to sign up for a standard account or contribute content to the Wikimedia Sites."

According to Wikipedia Community page, the six motivations it used were:

- **Values** – expressing values to do with altruism and helping others
- **Social** – engaging with friends, taking part in activities viewed favourably by others
- **Understanding** – expanding knowledge through activities
- **Career** – gaining work experience and skills
- **Protective** – e.g. reducing guilt over personal privilege
- **Enhancement** – demonstrating knowledge to others

To these six motivations he also added:

- **Ideology** – expressing support for what is perceived to be the underlying ideology of the activity (e.g. the belief that knowledge should be free)
- **Fun** – enjoying the activity

Conclusion

Considering the content of Wikipedia, this study assumed that individual participation in Wikipedia can be considered knowledge sharing behavior since individuals generally engage by contributing what they know to relevant entries. Studies of knowledge

sharing typically apply motivational theory to interpret individual knowledge-sharing behavior.

Kuznetsov conducted a comprehensive analysis of the motivations of Wikipedians using The Value Sensitive Design (VSD) approach. He concluded that the virtual realm of Wikipedia rewards its participants in ways that are unmatched by most establishments in the non-virtual world. Wikipedians enjoy a sense of accomplishment, collectivism, and benevolence, while working with exceptional freedom and ease. The values of reputation, community, reciprocity, altruism and autonomy are fostered both by the people and the technology, effectively creating a near-utopian society in which individuals voluntarily collaborate and learn together.

Yang examined the knowledge sharing behavior of Wikipedians from a motivational perspective. The analytical results implied that, compared to the influence of conventional motivation, self-concept-based motivation has a greater influence on individual knowledge-sharing behavior. Specifically, compared to other motivational sources analyzed in this study, internal self-concept-based motivation is the most important motivation for sharing knowledge in Wikipedia. However, external self-concept motivation is not significantly related to knowledge sharing intention, probably due to the difficulty establishing strong links with reference groups in the real world as well as the lack of social interaction in Wikipedia. Second, due to the relatively limited interaction in Wikipedia as compared to that in the general virtual community, intrinsic motivation also has an insignificant impact on individual willingness to share knowledge in Wikipedia.

Further Discussion

User-generated content provides a new way to create, manipulate, and consume information online. Wikipedia is one of the most renowned user-generated content applications and the largest multilingual free-content encyclopedia in the world, is written by users collaboratively. Wikipedians (i.e., individuals who write and edit Wikipedia) are generally allowed to edit Wikipedia content by sharing their knowledge in relevant entries. Accordingly, anyone with Internet access can search and browse Wikipedia entries for free. Compared to conventional websites, in which content is provided by the vendor, the emergence of user-generated content has indeed changed conventional views of how information is created, shared, and used.

Wikipedia provides a great model how user-generated content information could be create, shared and accurate online. To study the motivation of contribution will not only benefit the participation of knowledge sharing on Wikipedia, but also other knowledge sharing platform, like Quora and Yahoo! Answer. As it is discussed in one of the papers, on this platform, only small group of users are contributing to the content, keep adding and updating. Most of them are self-motivated fro interesting, value and enhancement, etc. To acknowledge the motivation of sharing, It will help the platforms like Wikipedia establish effective stimulation system to encourage more participant, which will be the fundamental power that push the information sharing.

SUMMARY AND SUGGESTION

The Wikipedia's purpose is to benefits readers by acting as an encyclopedia, a comprehensive written compendium that contains information on all branches of knowledge. Wikipedia was intended to be the largest, most comprehensive, and most widely-available encyclopedia ever written. To achieve this goal, the Wikipedia have some problems with sharing to solve.

Language gap is the biggest problem the Wikipedia facing. It is understandable that English is the dominant language on Wikipedia or any other sharing platform, but Wikipedia seems to be in lack of effective method to suppress the trend that English has become over-dominant. As we discussed, the lack of non-English content will estrange non-English contributor, and the lack of contributor will result in the unbalanced development of Content.

To eliminate the gap, Wikipedia should enhance the widely-availability.One feasible solution could be cooperation with local agencies. My research found in China, there are some Wikipedians community that groups many contributors. They have a good resource that could promote spread Wikipedia in China. Most of the members are students from education institute. Bao mentions that students are the most active contributor member. Another solution is to combine the automatic and manual translation system to translate the English version to a different language. Since some participants are motivated by the aim to learn languages, Wikipedia could recruit volunteers to take charge of it.

The study has found the motivation of contribution. Therefore, future studies may examine aspects such as cost or quality of contribution to exploring individual sharing behavior in Wikipedia. Finally, because this study only focused on individual motivation, the effects of some variables such as personal characteristics may have been overlooked. Future research can extend this study model by incorporating such variables into the analysis of knowledge sharing behavior.

This series of studies will evaluate how motivation affects individual knowledge sharing behavior in Wikipedia. It would be a great inspiration for Wikipedia to acknowledge user sharing behavior. For user-generated content platforms, such as Wikipedia, users are the core of the platform. A detailed analysis will greatly help these platform attracting contributor, and keep their community active.