

VAST 2008 Wiki Editors Mini Challenge – Identifying Social Networks using Wiki.viz

Chi-Chun Pan*
The Pennsylvania State University

Donald Pellegrino†
Drexel University
Prasenjit Mitra§
The Pennsylvania State University

Chris Weaver‡
The Pennsylvania State University

ABSTRACT

We used Improvise visualization framework to build Wiki.viz, an interactive visualization of the 2008 VAST wiki edit mini-challenge. Users can extract and analyze data using automated text processing, visualization of wiki-revision networks and word senses

Index Terms: H.5.2 [Information Systems]: Information Interfaces and Presentation—User Interfaces;

1 INTRODUCTION

In this paper, we show how a visualization, Wiki.viz, based on the Improvise framework[1] can be used to (a) identify different groups of people and their relationships, and (b) extract information about any violent activities from the information available in Wikipedia edit pages. Such information can be valuable. The succinctness of wiki edit comments makes deriving social networks and violent activity information a challenging problem. Using our tool, we analyzed the 2008 VAST wiki edit mini-challenge consisting of the wiki page of the Paraiso Manifesto, a discussion page, and the edit history log of a wiki page related to the Paraiso Manifesto. In this paper, we summarize the key features of Wiki.viz and the analytic processes for the VAST 2008 wiki editors mini-challenge.

2 DATA PROCESSING

Before using the visualization tools, we processed the raw data. First, we created a set of rules to extract the revision patterns in the summary field of the edit history. The revision patterns reveal the opinions of the authors and identifies their leanings. We classified the revision patterns in to three types – *revert*, *revertTo*, and *undo*. When authors undo others work and returned the page to previous versions, we refer to the action as a *revert*. The difference between undo and revert is that by doing undo, users can revert a single edit without simultaneously undoing all constructive changes that have been made since. The third type of revision patterns is *revertTo*. Unlike the other two types of patterns, a *revertTo* edit might indicate that the authors edited the page think the changes made by other authors could be better than the current versions. Occasionally, a *revertTo* pattern may appear with a *revert* pattern. By extracting the patterns, we can create revision graphs by creating nodes as authors and linking nodes with the three types of patterns extracted from wiki history.

3 WIKI.VIZ

Wiki.viz is an interactive visualization of the wiki edit history data. Based on Improvise, Wiki.viz provides intuitive visualization for data and revision networks extracted from the revision patterns. Wiki.viz consists of 6 major components (see Figure 1). The table view of edit comments on the left shows the original edit history. On the right, the table views of edit and revert authors list all the authors who participated in the wiki page. By selecting authors from the list, we can visually analyze the social networks between authors on the graph view in the center. The calendar view and time series on the bottom show the activities throughout the time span. Finally the table views on the right bottom show the words used in the edit page.

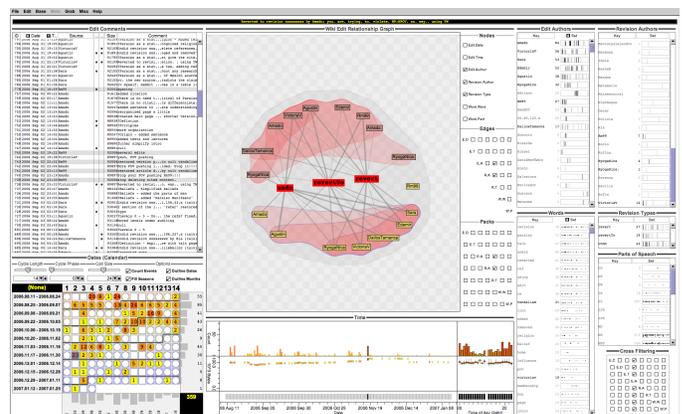


Figure 1: A screen shot of Wiki.viz. Wiki.viz consists of 6 major components. A table view of edit comments, table views of edit and revert authors, a graph view of social network, a calendar view and a time series show the activities, and table views of words used in the edit history and part of speech of those words.

4 ANALYTIC PROCESS

In order to identify groups in the wiki edit history, we started with the Discussion page. We first tried to categorize authors into three groups: supporters, challengers, and neutral contributors. By reading the discussion page, we obtained a list of users contributed to the discussion page (see Table 1) and their opinions regarding to the Paraiso Manifesto.

The supporters are the people who agree the Paraiso Manifesto. Among these users, VictoriaV is the most active one. Although, in the discussion, VictoriaV appeared to pretend that she is neutral with respect to the Paraiso Manifesto, she had many arguments with others in the discussion page that belie that claim. For example, VictoriaV argued about the source of Catalan enlightenment with Agustin and Edemir. In addition, Rm99 claimed that VictoriaV have lied about her background and her relation to the Paraiso

*e-mail: cpan@ist.psu.edu

†e-mail: don@drexel.edu

‡e-mail: cweaver@psu.edu

§e-mail: pmitra@ist.psu.edu

Table 1: Groups and their members found with our analysis

GROUP	NAMES
Supporters	VictoriaV, RyogaNica, Amado
Challengers	Edemir, Rm99, DailosTamanca
Neutral contributors	Agustin, Sara
Page removers	195.113.65.x, Alejo, 75.179.21.x, 84.158.202.x, Alejandrosanchez, 209.155.27.x, Molotover, 66.175.135.x, 201.226.51.x, Honoratas, 74.120.3.x, 85.135.211.x, 67.55.3.x, Rosamaria, Absalon, 204.52.215.x 74.130.152.x139.55.50.x, 69.14.85.x
Revert Only	Kurrop, Seina

Menifesto. It is obvious that she supports Catalanos point of view. Therefore, we tag VictoriaV as a supporter of the Paraiso movement.

Among 387 users, most users only made one edit. It would be difficult to judge their opinion about the Paraiso Menifesto. Therefore, for analyzing the revision graph, we only select users who made more than 3 edits. With time filtering, we can apply animation to replay the interaction between users in the revision graph. We can also filter on type of revision by selecting Revision Types on the visualization. We start our analysis with the users found in the previous analysis of discussion page. If we find two users revert/undo each other several times (the width of edges indicate number of revision occurred), we put them into different groups. On the other hand, if we see *revertTo* link between two users, we tag them as the same group. We found two more users in the supporters group, RyogaNica and Amado, with this approach. Based on the revision graph, we also found that RyogaNica made several revisions on Rm99s, DailosTamanca, and Edmirs edits (they are identified as challengers).

By tagging VictoriaV as a supporter, we could easily tag Rm99 and Edmir as challengers based on their wording in the discussion page. Rm99 questioned the credibility of the external links provided by VictoriaV. Also, Rm99 claimed that “The controversy surrounding Catalano has been downplayed considerably.” Edmir. In addition, based on the conversation in the Civility section of the discussion page, we suspect that Rm99 sometimes entered the wiki page with IP address 81.96.243.x. By analyzing the revision graph, we confirm that Edmir and VicotriaV have different agenda regarding to the Paraiso movement. We also indentify DailosTamanca as a challenger because DailosTamanca and RyogaNica make several reverts to each other

If we could not clearly judge a users agenda, we can select the user on the author list and filter the edits with the author name. This allows us to read all the edits made by the user. Using this approach, we think Agustin and Sara may be normal wiki users who tried to maintain the neutral point of view for the wiki page.

We also found some interesting patterns by creating time series visualization on the size of wiki page. We found several sudden decrease of the size of the page on the time series of page size. The time series on the bottom indicates the maximum, minimum, and average page size for everyday. By sorting the edits with the size of page on the table view of wiki edit history, we found a group of users who did not contribute to the content of the wiki page. Instead, they replaced the wiki page with short sentences. Most of them did not login with a registered username. We think those users may fit the profile of the protesters mentioned in the wiki page.

Finally, we analyze the frequency of a user appear in the revision graph and the number of edits made by the user. We found two persons (Kurrop and Seina) who only revise others edits. They did not add any information into the wiki page.

4.1 Violent Activities

In order to identify violent activities, we extract the words used by users in the wiki edit history. Our hypothesis is that violent ac-

tivities could be recognized with some cue words. By remove stop words and applying stemming algorithms, we obtained 1279 unique words. We further applied natural language processing tools on the comments in the wiki edit history and assigned part of speech tags for the 1279 words. By inspecting those 1279 words, we manually assign a real value between -1.0 and 1.0 to selected words (the default value for every word is 0.0). A negative value indicates that a word has higher possibility to be used to describe violent activities (such as hate and attack). A positive value indicates that a word has higher possibility to be used to describe non-violent activities (such as good and agree). By selecting users and words of interest from the table view of authors and words, the graph view becomes an author-word graph which provides an intuitive way to visualize who use what words in the wiki edit history. By analyzing the author-word graph and the summary of wiki edit history with the steps to identify groups of people, we found the edit which may link the Paraiso movement to violent activities.

(cur) (last) 03:16, 19 September 2006 Alphanzo (Talk — contribs) m (moved Paraiso to GUNNED DOWN SIX DOCTORS AND NURSES IN COLD BLOOD)

It is possible that someone who has firsthand knowledge about the execution tried to reveal Paraiso movements crime activity. By reading the posts before Alphanzos edit, we suspect the confrontation of Paraiso members and Dept. of Health may cause the violent activity.

(cur) (last) 03:09, 19 September 2006 Edemir (Talk — contribs) (97,530 bytes) (?Home Health Care - added confrontation of Paraiso members and Dept of Health)

We also suspect this message may be related to the incident that happened in the evacuation mini challenge.

Other wiki edits that point to potential violent activities include the reference to prosecution of Paraiso by Belgians and Catalano’s death. However, there is no direct evidence of violence associated with any of these events.

5 CONCLUSION

Wiki.viz is an intuitive system that enabled analysts to perform analysis of the data and the information extracted with automated text processes. Although the system is developed for analyzing the wiki edit history of the VAST 2008 mini-challenge, it should be able to use on any wiki edit history data.

ACKNOWLEDGEMENTS

This work is supported by the National Visualization and Analytics Center, a U.S. Department of Homeland Security program operated by the Pacific Northwest National Laboratory (PNNL). PNNL is a U.S. Department of Energy Office of Science laboratory.

REFERENCES

- [1] C. Weaver. Building highly-coordinated visualizations in improvise. In *INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization*, pages 159–166, Austin, Texas, USA, 2004. IEEE Computer Society.