

# Visualizing Author Contribution Statistics in Wikis Using an Edit Significance Metric

Peter Kin-Fong Fong  
Dept. of Computer and Information Science  
Faculty of Science and Technology  
University of Macau  
Macau S.A.R., China  
ma86515@umac.mo

Robert P. Biuk-Aghai  
Dept. of Computer and Information Science  
Faculty of Science and Technology  
University of Macau  
Macau S.A.R., China  
robertb@umac.mo

## ABSTRACT

Wiki articles tend to be edited multiple times by multiple authors. This makes it difficult to identify individual authors' contributions by human observation alone. We calculate an edit significance metric, using different weights for different types of edits, which reflect the different value placed on them by wiki community members. We then aggregate edit significance values and present them as visualizations to the user to aid in perceiving extent and patterns of contributions.

## Categories and Subject Descriptors

H.3.5 [Information Storage and Retrieval]: Online Information Services—*web-based services*; H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces—*collaborative computing*; I.3.8 [Computer Graphics]: Applications

## General Terms

Measurement, Design, Experimentation

## Keywords

Wiki, Wikipedia, revision history, edit significance, information visualization

## 1. INTRODUCTION

An article in a wiki may have many authors, and may have been revised many times. This is especially true when the wiki is a large public one, such as Wikipedia. Sometimes wiki users wish to know who are the major contributors of an article. For example, a manager may look for domain experts familiar with a certain subject by identifying major authors of articles related to that subject in the company-internal wiki. In an open wiki community like Wikipedia, members may want to “reward” writers who contribute to a much requested topic. Although it is possible to manually check the revision history of related articles and examine all revisions one by one to find out major contributors, this would be an impractical approach given the large number of authors and revisions. An automatically calculated metric, expressing author contributions as a numerical value, would be convenient in such cases.

Copyright is held by the author/owner(s).  
WikiSym '11 Oct 03–05 2011, Mountain View, CA, USA  
ACM 978-1-4503-0909-7/11/10.

In 2010 we proposed a framework for edit significance calculation [1]. *Significance* is a subjective measure of how important a given editor's contribution is. It depends on both the *volume* and *type* of an edit. Instead of treating all kinds of edits equally, we distinguish different kinds of edits, e.g. adding content, inserting references, correcting spelling, moving content around, and many others. Different weights are assigned to different edit actions, to reflect their relative importance when calculating edit significance values. When appropriate weights are chosen, such a distinction produces significance values that are more in line with human intuition as to what is or is not a significant contribution. In this paper, we present several visualizations of our calculated edit significance values. Section 2 describes our visualizations, and Section 3 makes conclusions.

## 2. VISUALIZATIONS

Our edit significance calculations are performed by the edit history analyzer we proposed in 2010. Details of the calculation steps are described in [1].

To communicate the results of edit significance calculations in an easily perceptible way, we employ several simple visual representations. We have extended the Mediawiki “history” page to embed an edit significance bar inside the detail line of each revision, as shown in Figure 1. The bar width represents the significance value, with a full-width bar representing the largest significance value recorded in the entire wiki's database. Given that significance values tend to vary greatly, a log scale is used to map values to bar width. The actual significance value of a given revision is displayed when hovering the mouse pointer over the edit significance bar, as is shown in the case of the second bar in Figure 1. Clicking the bar brings up a page that shows a detailed breakdown of edit actions for that revision.

Furthermore, stored edit significance values can be aggregated for a given article to produce a total significance value of all its revisions, as well as for a given user to produce a total value of all



Figure 1: Edit significance bar in page history view

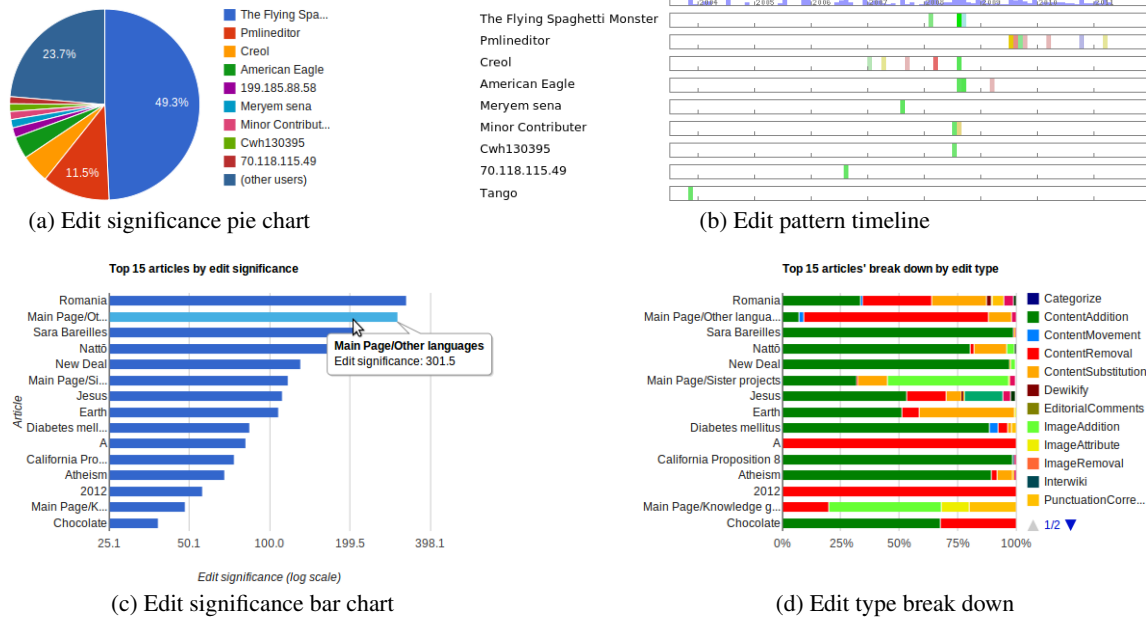


Figure 2: Different types of edit significance visualization

of this user’s contributions. It can also be used to find and rank an article’s most significant contributors, and the most significantly contributed-to articles of a given user. We have designed and implemented a MediaWiki extension that performs this kind of aggregation and visualization. A ranked list of edit significance values of top authors of an article is shown in Figure 2a. One can clearly see that almost half of all contributions came from a single author.

We also produce a timeline showing a temporal breakdown similar to WikiDashboard [3] but displaying edit type and significance (not just edit counts), as well as edit types per author and some other statistics (Figure 2b). These graphs give a general impression of the share of contributions made by top contributors, and the type of their contributions. The timeline visualizes different types of edits with different colours: green for content addition, red for content removal, and blue for maintenance edits. Mixes of these three colours represent the nature of edits in a highly concise way. Colour saturation represents the edit significance: the more saturated the colour, the higher the significance value.

User contributions are represented in similar charts, accessible through another MediaWiki special page, showing a ranked list of a user’s most significantly contributed-to articles (Figure 2c), useful in seeing the extent of contribution; and the types of contributions per article (Figure 2d), useful in perceiving differences of contribution types between articles.

All needed values are pre-calculated and stored in a database by a batch processor, so that preparation of statistics for visualization simply requires data retrieval and aggregation, which can be realistically performed in real time even on large wikis. Statistic charts are drawn using the Google Visualization API<sup>1</sup>.

### 3. CONCLUSIONS AND FUTURE WORK

Our visualizations of edit significance provide quantitative insight into both extent and patterns of author contributions. We be-

lieve that these graphs give useful insight into users’ topic areas of interest, and help classify editors by their predominant types of edits performed.

We continue exploring further ways to enhance information value. One way is to group contributions by general categories corresponding to the wiki’s category hierarchy, revealing users’ expertise and areas of interest. When combined with existing category aggregation methods [2], higher-level category contributions could be shown, stating for example that user  $u$  has mainly contributed  $x\%$  to category “Science”,  $y\%$  to “Culture” and  $z\%$  to “People”.

In earlier work we determined degrees of co-authorship between wiki users based on a simple edited revision count of co-authored articles [4]. Edit significance was only coarsely determined by distinguishing “minor” and “non-minor” edits, but without an ability of more accurate assessment. Applying our edit significance calculation would make such a more accurate assessment possible.

### 4. REFERENCES

- [1] P. K.-F. Fong and R. P. Biuk-Aghai. What did they do? Deriving high-level edit histories in wikis. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, WikiSym ’10, pages 2:1–2:10. ACM, 2010.
- [2] C.-I. Pang and R. P. Biuk-Aghai. A method for category similarity calculation in wikis. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, WikiSym ’10, pages 19:1–19:2. ACM, 2010.
- [3] B. Suh, E. H. Chi, A. Kittur, and B. A. Pendleton. Lifting the veil: Improving accountability and social transparency in Wikipedia with WikiDashboard. In *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, pages 1037–1040. ACM, 2008.
- [4] L. V.-S. Tang, R. P. Biuk-Aghai, and S. Fong. A method for measuring co-authorship relationships in MediaWiki. In *Proceedings of the 4th International Symposium on Wikis*, WikiSym ’08, pages 16:1–16:10. ACM, 2008.

<sup>1</sup>[http://code.google.com/apis/visualization/interactive\\_charts.html](http://code.google.com/apis/visualization/interactive_charts.html)