On the analysis of Wikipedia activity through time

Nuno J. Silva
Instituto Superior Técnico
Universidade Técnica de Lisboa
Av. Rovisco Pais, 1, 1049-001 Lisboa
nunojsilva@ist.utl.pt

ABSTRACT

Wikipedia is a collaborative website which is growing in content and popularity. As its popularity grows, so does the likelihood of it reflecting the opinion of society as a whole. Thus, it is of interest to analyze Wikipedia content, editions, disputes and comments in order to draw conjectures about the public opinion. We present a visualization which condenses article content, metrics and other information, along with their evolution through time, in a single visualization. We then present some case studies, where we could find patterns and infer hypotheses on the activity of Wikipedia articles, and we also report findings which resulted from user tests.

Author Keywords

Wikipedia; User-generated content; Information visualization; Revision history

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces—*Graphical User Interfaces*; H.5.3. Information Interfaces and Presentation: Group and Organization Interfaces—*Collaborative computing, Computer-supported cooperative work, Web-based interaction*

INTRODUCTION

In 1995, Ward Cunningham[11] suggested a new approach to the web[4], where everyone is able to edit the hosted content. This approach, which became known as "wiki", allows users to provide their own feedback just by following an hyperlink, which leads users to a special page where they can submit their changes to the content. In some sites, such as Wikipedia, the previous versions are stored in a database and can be browsed by users. Accessing this data makes it possible to collect information on users, topics and their importance and relevance.

As this approach becomes more popular, this possibility becomes even more appealing, as behaviors and trends assessed from the Wikipedia edition history will converge to public

opinion, allowing us to extend the insights collected from the analysis of the Wikipedia edition history to the whole public opinion.

In this paper, we present a tool that enables the analysis and assessment of trends in Wikipedia, through the presentation of metrics and other information on (or from) the revision history of Wikipedia articles using a graphical visualization system, built upon the information made available by the MediaWiki API, which provides a consolidated display of the sequence of metrics computed for the article revisions. We also present a novel set of metrics, loosely based on insights from related works.

We also want to allow users to compare two articles, possibly from different Wikipedias, nourishing comparisons between the different communities of Wikipedia editors and of speakers of the involved languages.

The rest of this document is structured as follows: first, we briefly summarize related works, mentioning their contributions, strengths and weaknesses, and then we describe our solution. After that, we summarize the remarks and insights gained from user tests, and, finally, we present our conclusions.

RELATED WORK

Visualization of Wikipedia data has already been the focus of several academic works, employing different strategies.

WikiViz[14], ClusterBall[13] and Holloway et al.[16] present static depictions of topics and connections among them, which are not able to convey more than topics and connections, and also lack the depiction of the evolution of the visualized information through time.

Wikiswarm[26], Revert Graph[22] and Brandes et al.[6] propose similar network visualizations, but focused on authors, not topics, but still constrained with respect to the conveyable information, which, except for Brandes et al., does not include evolution through time. Even in Brandes et al., this information is disconnected from the network visualization.

WikiDashboard[23] and WikipediaViz[8] provide several metrics, some of them along with their variation along time, integrated in the display of the content of a single Wikipedia article version. iChase[20] presents similar information, but for a set of articles and using several visualization elements (including heatmaps).

Omnipedia[3], unlike the other works, explicitly focuses on

the comparison of the coverage of topics, and of relationships between topics, across the different editions of Wikipedia operated by the Wikimedia Foundation.

History flow[24] and ThemeRiver[15] present content-focused visualizations, where their evolution through time is a key part of the screen, being the only of the surveyed works that effectively enable users to analyze changes in the depicted information through time using the main element of the visualization.

We then surveyed works involving the analysis of content (both the visualization works above, along with some analysis-only works), assessing the metrics chosen by their authors, which can be grouped and described as follows:

- **Edit count** Used by four works[25, 26, 23, 20], refers to the use of the number of changes (possibly along time) as a heuristic;

- **Editor anonymity** Only employed by two works[19, 17], regards the identification of an author as anonymous, and influencing further decisions based on her anonymity.

- **Number of editors** Concerns using the number of different editors that changed the article as a metric[25, 26].

- **Number of reverts** The number of reverts that happened on an article. Used by only one work[22], which is a visualization of the revert network and, as such, this metric relies heavily on revert statistics to build its result.

- **Impact of deletions** The extent to which an edit changes the existing content[25].

- **Mutual controversy** is, as it name states, an assessment of the overall controversy expected to arise from two different users[25].

- **Compressibility** Involves the application of some compression algorithm and the computation of the resulting *compression ratio*, that is, the ratio between the length of the compressed content and the length of the original, uncompressed content[17]. This value (or its comparison with a reference value) can then be used to assess if the content deviates from the usual rate found in legitimate texts.

- Other **text statistics**, regarding word and letter frequency are also computed by some works[19, 17], and can be used in the same way as compressibility.

- **Link density** Is also a text statistic, but specific to hyperlinked text, concerning, in this case, the interpretation of Wikipedia markup to find the links present in the content[8]. Other than quality and vandalism, it also presents a measure of how tightly is the article connected to the link structure of Wikipedia.

- **Similarity** Concerns a more local statistical analysis, comparing the properties of the introduced changes with the scope where they were made, instead of comparing to a more global reference value[19, 24, 15].

- **Category similarity** Used to rule out or relate articles based on their categories, with articles that share categories being considered related[16, 14, 13, 3].

- **Length**[19, 17, 24, 15] and **age**[25, 2, 26, 6] Stand for those properties of an article and its content.

It is possible to conclude that unrelated works used different sets of heuristics, except for the *edit count*, *age*, *length*, *similarity* and *category similarity* metrics, which are used by three or more works. It should also be noted that Velasco[17] shares several heuristics with Potthast et al.[19] because the former work is based on the latter.

THE "POPCULTURE" VISUALIZATION

We conceived a system which comprises two components: the main component, a visualization, and a backend component, concerned with the assessment of metrics, a high-level architecture which we present in figure 1.
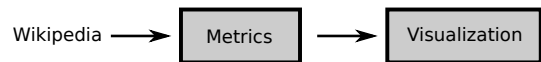

Wikipedia → Metrics → Visualization

Figure 1. High-level representation of the proposed system.

Metrics

It is in the metrics computation component that decisions are made on *what* to show in the visualization. We defined a set of metrics, following an approach similar to that of Mola-Velasco[17], where several metrics are aggregated together using techniques from machine learning, although we opted by using an unweighted average of the values, instead of gathering weights using a supervised learning step. The defined metrics are as follows:

- **Quality**, a score which highlights articles having many good words and little or no biased or offensive words, near-average compressibility and link ratios, and which follow the expected character distribution for human-readable text;

- **Vandalism**, which presents higher values for texts rich in offensive words, but poor in good words, and whose ratios (compressibility, link) along with the character distribution are not close to the reference values;

- **Controversy**, which attributes higher scores to articles which, while being rich in good words, showing a near-average character distribution, make use of biased words.

These metrics, along with the content length, version timestamp and authorship information, among other properties, are passed to the visualization component on request.

Visualization

The design of the visualization component (which is shown in fig. 2) was focused on *how* to show the information made available by the backend, presenting the retrieved information, a time series, using a plot with a timeline: article changes are explicitly marked over the timeline, along with information on the depicted time range.
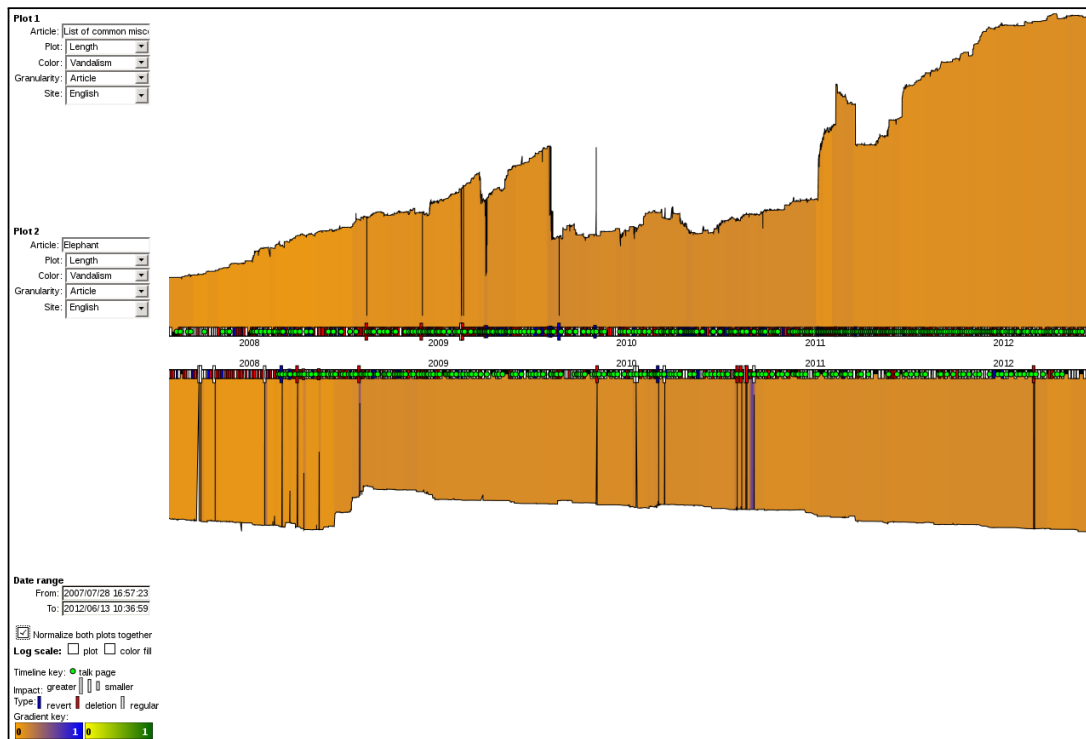
Figure 2. The "PopCulture" visualization

The visualization is made of three parts, one of which comprises two plots, used to depict the metrics extracted from two articles. These parts, except for the content pane, which is not shown by default, are visible in fig. 2 and are as follows:

The Option Pane
Located at the left side of the screen (fig. 3), provides options for the whole visualization (fig. 3C), options for each of the plots (fig. 3A,B) and a key explaining the meaning of some of the graphical elements used to convey information (timeline markers and gradients, fig. 3D).
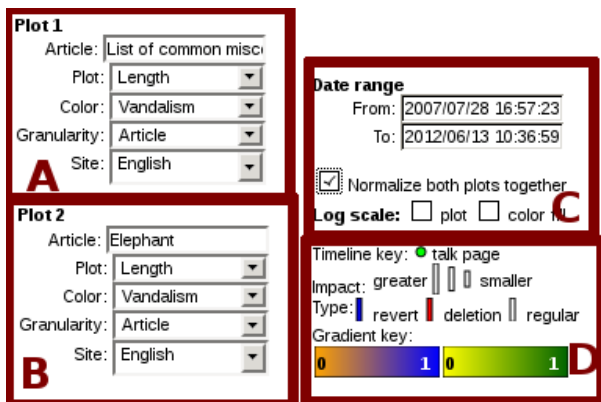

Figure 3. The options pane: A) Upper plot options; B) Lower plot options; C) General options; D) Visualization key.

For each plot, users are given an option pane, which allows them to

- **Change** the article under analysis, through an "Article" textbox, which is also equipped with real-time autocompletion, whose results are shown in a list which will appear below the textbox;

- Choose the metric used in the **plot**, one of "controversy", "vandalism", "quality" and "length".

- Choose the metric used to **fill** the plot, for which all the four metrics above are available, along with a fifth one, "by author", which colors revisions by their author (attributing colors to the twenty most common editors, and painting the remaining ones in gray).

- Change the granularity of the visualization, between article-level and section-level.

- Select a MediaWiki site to access.

Users can also access **general options** in the bottom left corner, concerning the visualization timespan ("Date range"), the normalization strategy, which can be set to normalize to the maximum of *both* plots, and toggling logarithmic scales ("Log scale"), both for the plot metric and for the filling color.

A **key** is also available, explaining the colors and symbols used in several parts of the visualization.

The plots
The central part of the visualization, lying on its center and extending to its right, are two plots along horizontal timelines, which encode information about edits, marked using small rectangles and color coded according to their kind (blue for reverts, red for deletions and white for regular edits, as

listed in the key), and edits to the corresponding talk page, marked with green circles. The impact of page edits is conveyed through the height of their markers, which is proportional to said *impact*.

The visualization can be zoomed into by selecting (click, drag and raise the left mouse button) a plot area, zoomed out, by right-clicking the plot, and panned (by dragging and dropping the area between the two timelines), allowing the users to navigate through time, and also provides users the ability to see the revision content on a side pane (fig. 5), which can be set to highlight changes between two consecutive revisions.

When the mouse pointer is over the plot, some textual information about the segment under the pointer is shown in a tooltip near the pointer (fig. 4). Clicking with the left mouse button inside the plot locks this information for further analysis, which can be unlocked by left clicking the plot again.
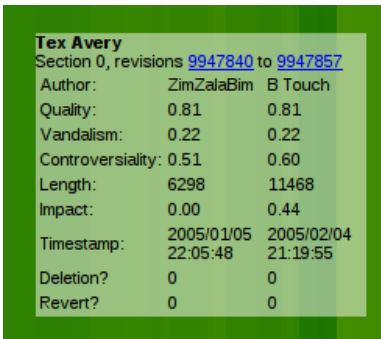

Figure 4. Tooltip being shown in the visualization of the English Wikipedia article on "Tex Avery".

This information includes the covered date ranges, the involved section (if the granularity is "by Section"), and the involved editors.

Interaction
Plots are updated in real time: after editing some option in the textboxes (article name and dates), the update is triggered by leaving the field or by hitting the "Enter" or "Return" keyboard keys.

The content pane is updated when moving the mouse pointer over the plot, unless the tooltip has been locked.

The plot conveys information through its line, using the distance between the plot line and the timeline, and through its plot area, which is filled with a color gradient. Both elements can be configured to depict different metrics (quality, vandalism, controversy and length), with the plot area also enabling users to fill it by author.

The visualization offers two sections, an upper plot and a lower plot, thus allowing users to compare either different metrics for the same article or to compare two different articles, possibly from different sites, on the same screen.

Each of these sections can be set to convey information for the entire article (using only one plot) or for each of its sections (rendering one plot by section).
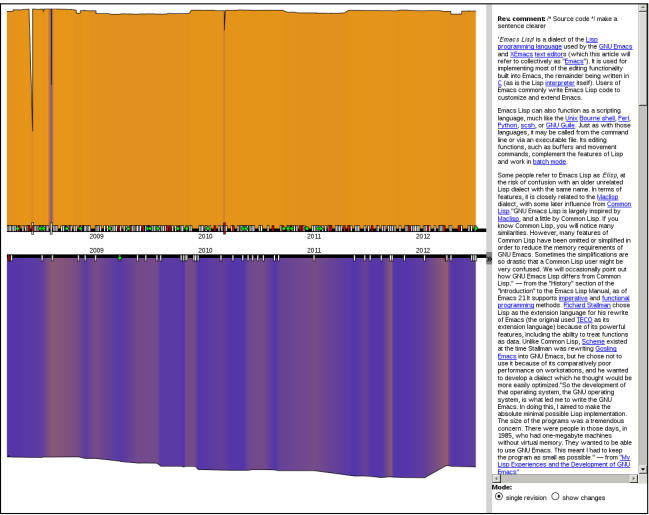
Implementation

Figure 5. Visualization with the content pane enabled

For the implementation of this visualization, we decided to build upon d3.js[5], a JavaScript visualization and graphics library, and jQuery, a generic library of JavaScript tools.

The visualization operates by issuing asynchronous requests for the data, which, when served, are preprocessed and used to build the `d3.js`-based plot, which involves the generation of SVG graphics from the metrics.

Plots are generated by creating a plot area from the data-points, which is then filled with a generated gradient. These steps are executed for each plot, as there may be several plots, if the granularity level is set to "by section". Finally, the plot timeline is updated and its markers are generated.

When drawing the plot line and the boundaries of plot areas, numeric metrics are processed in order to normalize the values so that the highest value corresponds to the maximum height. Plot lines are also colored black, in order to ensure the user is able to tell sections apart when the metric used in their fill presents the same value (resulting in an identical fill).

Gradients map values (which are in the range $[0; 1]$) in the same way, this time using a *color interpolator*, where 0 corresponds to orange (or yellow, in highlighted gradients) and 1 to blue (green, in highlighted gradients). Values between 0 and 1 are mapped to corresponding shades of orange and blue (or of yellow and green, for highlighted gradients).

Timestamps are passed to those objects in order to define the points of plot lines, boundaries of plot areas and gradient stops. Once again, `d3.js` scales are used to map timestamps to the visualization window. They are also passed to the timeline, in order to create the markers.

When several markers, points and stops are too close, they are merged into a single datum, whose value is the highest of them all, thus conveying the strongest impact among the merged data. In the timeline merged markers are colored as revert or deletion if and only if at least one of the merged markers are marked as such.

We designed the entire system in a site-agnostic way: other than the site language, for which only a small subset of the Wikimedia Foundation-operated sites are recognized, there is no tie to a specific site or language. The entire data flow is built upon a small set of parameters that identify the MediaWiki server: its hostname and the path, in the HTTP server, to the MediaWiki instance.

Thus, although our work focuses on the analysis of Wikipedia articles, our system can effectively be used to analyze articles from other sites, such as Uncyclopedia[1], the official Gentoo Wiki[2] or any of the Wikimedia Foundation "Sister Projects", such as Wikibooks[3].

The site language is used in order to choose the list of offensive, biased and good words used in the computed metrics, with the current version providing lists for English, French, Portuguese and Spanish.

case studies

We used the visualization to explore some articles, in an attempt to discover patterns, similarities and differences:

When exploring articles expected to be highly vandalized, it became evident that choosing length as a metric made some acts of vandalism evident, as these translate into sudden, abrupt length changes, such as we observed in the English Wikipedia article on the year "42" of the Julian calendar (fig. 6), where the peak corresponds to an edition where some user added numerous repetitions of the expression "ITSTHE-MEANINGOFLIFE" to the first paragraph of the article.
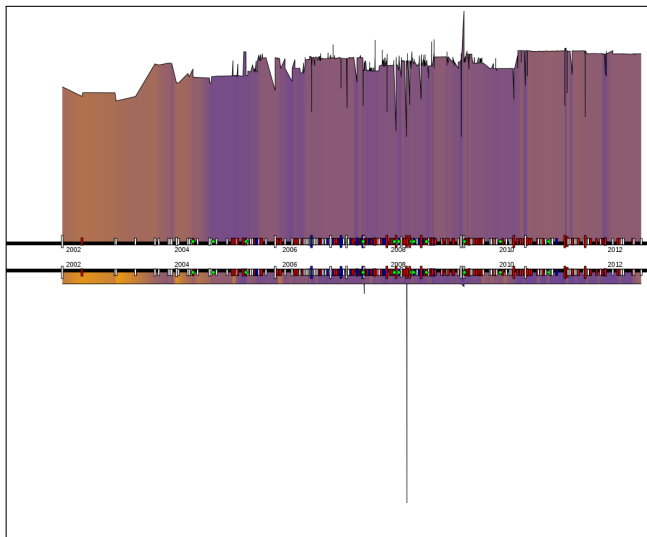

Figure 6. Visualization of the English Wikipedia article "42", where the lower plot is set to depict content length.

The visualization also enabled us to identify edit wars, through the "zigzag" pattern those sometimes impose on the length plot. An example is the English Wikipedia article on the Republic of Kosovo, where editors fought on whether to

redirect the page to the article on "Kosovo" or to have a separate article on the "Republic of Kosovo", as shown in fig. 7.
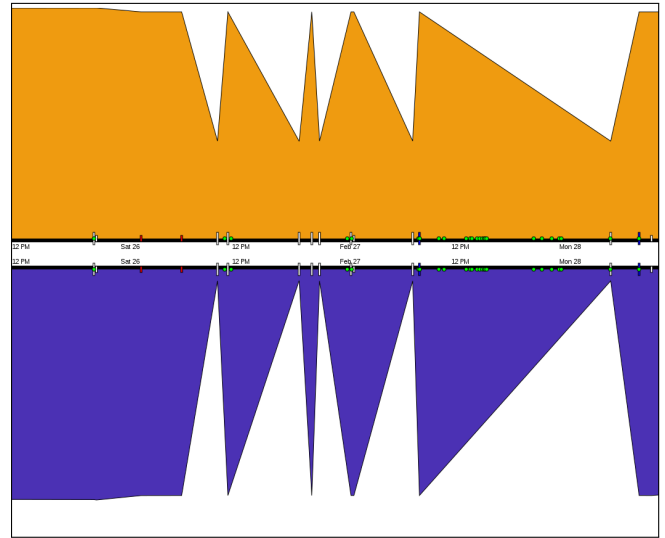

Figure 7. Visualization of the English Wikipedia article on the "Republic of Kosovo": an edit war regarding the split from "Kosovo" and increased talk page activity, where the lower plot is set to depict length.

It is also possible to assess the article popularity by looking at the timeline, such as in the article "List of common misconceptions" of the English Wikipedia, where it is possible, by looking at the big picture, to discover the article became popular in 2006, and has seen an increased edit rate ever since (fig. 8).
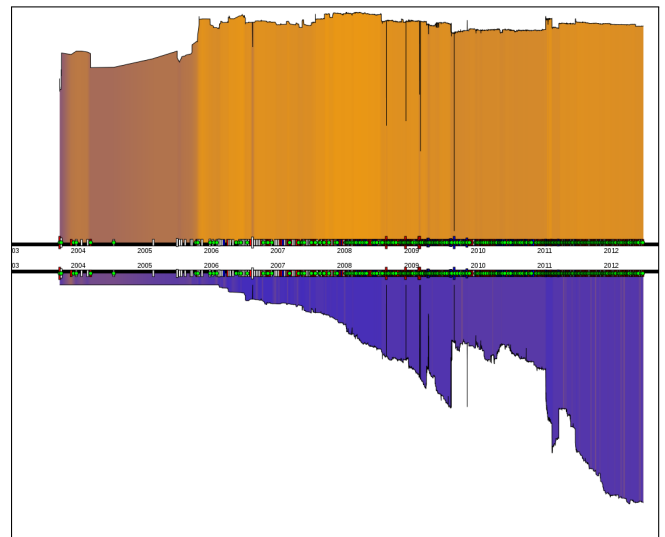

Figure 8. Visualization of the entire history of the English Wikipedia article on the "List of common misconceptions".

Depending on the subset of the regional community that is interested in the topic, it is possible that an article is updated at two different Wikipedias in a similar way. In the English and Hungarian articles on former Hungarian president

Schmitt Pál[4], we could, by focusing on 2011 and 2012, find two activity bursts: January and April, which not only shows there is similar activity in both articles, but is also an example of a correlation between Wikipedia article activity and events outside Wikipedia: On 11 January 2012, a magazine accused Schmitt of plagiarizing his doctor thesis, explaining the first burst; On 27 March, the senate of his university was advised to withdraw his title, which it did on 29 March, and after which Schmitt announced his resignation from the office of President of Hungary, on 2 April, matching the second burst.
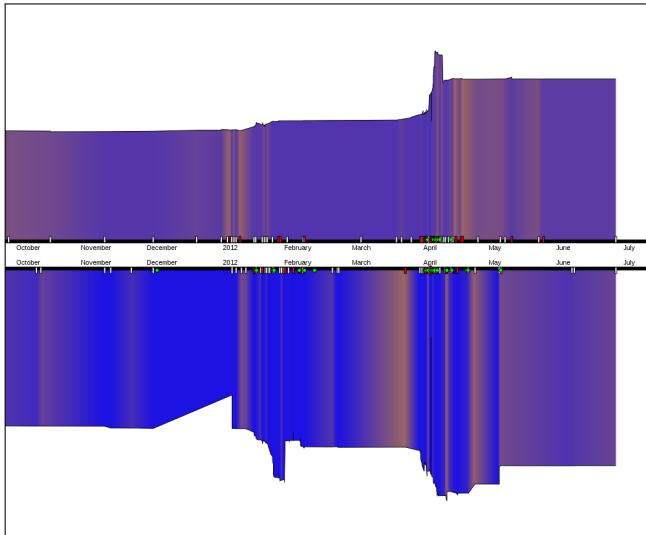

Figure 9. Visualization of the English Wikipedia (above) and Magyar Wikipedia (below) articles on "Schmitt Pál", focusing on the last months of 2011 and on 2012.

On the other hand, comparing the English article to the Portuguese one (under the same title), we get a different scenario: the claims of plagiarism did not translate into heavy activity bursts in the Portuguese article, although some editors promptly updated the article (fig. 10).

This type of bursts, triggered by events outside Wikipedia, can also be observed on the Portuguese article on "Fernando Nobre", while this also shows another example of different activity levels, as these activity spikes do not occur in the English version.

On 10 April 2011, Fernando Nobre accepted an invite to head the list of a political party for the electoral circle of Lisbon, with the explicit goal of being then elected president of the Assembly of the Republic[9]. The decision of the party leader to invite him, his acceptance and the bold way in which the leader of a party tried to choose the holder of the second most important political office in the country were heavily criticized and spurred controversy, ending with the deputies refusing to elect him[21].

Visualizing the entire lifespan of the articles on "Fernando Nobre" from the Portuguese and English Wikipedias, we found that there is a noticeable activity increase during 2011,

---

[4]Respectively, "Pál Schmitt" and "Schmitt Pál", where the former has his name written in western name order.
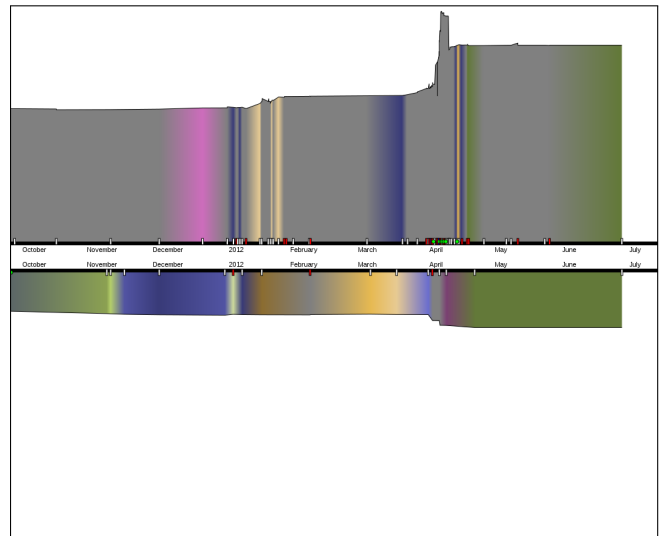

Figure 10. Visualization of the English Wikipedia (above) and Portuguese Wikipedia (below) articles on "Pál Schmitt", focusing on the last months of 2011 and on 2012, with both plots set to plot length and fill by author.

which, if we inspect closely, did only occur in the Portuguese article (fig. 11).
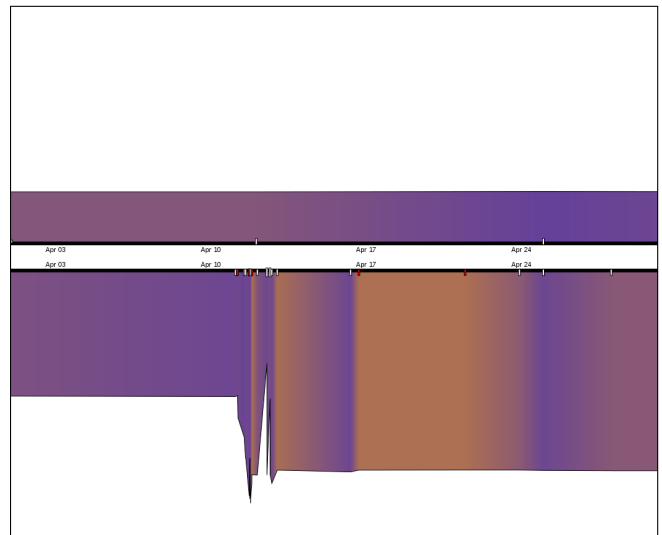
Figure 11. Visualization of the English Wikipedia (above) and Portuguese Wikipedia (below) articles on "Fernando Nobre" focusing on April 2011 (right). Both visualizations are set to plot length and fill by controversy.

Moving away from the timeline markers and length, the other metrics also allow us to identify unusual edits: focusing on the visualization of "Chicken" from September 4 to October 20, 2011, we can see how the several metrics are able to depict the article revision 452274414[5] (where user "Wearefictional" replaced the article content with "The chicken Delicious"), which translates into a lower quality value, a shorter length and a higher vandalism value, as depicted in fig. 12,

---

[5]`http://en.wikipedia.org/w/index.php?oldid=452274414`

where the vandalism revision clearly stands out in the vandalism fill gradient (marked in blue against a uniform orange fill).
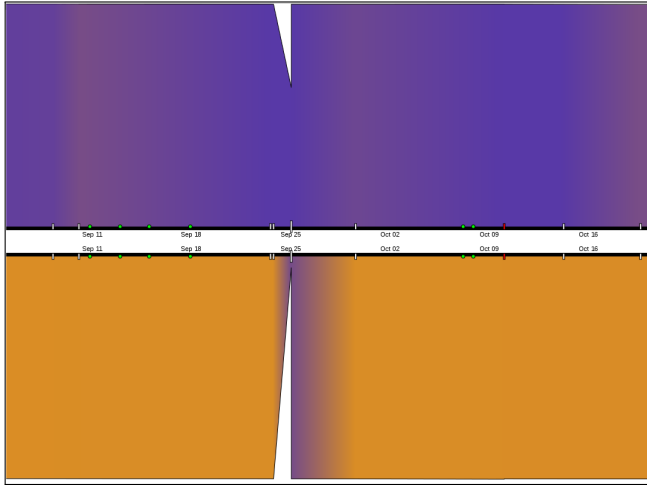

Figure 12. Visualization of the English Wikipedia article on "Chicken". The upper visualization is set to show length and quality and fill by controversy, while the lower one is set to plot length and fill by vandalism.

USER TESTS

We employed user tests to analyze the adequateness of our visualization. Tests were carried out with 20 users: 16 male, 4 female, with an average age of 26 years, $\sigma = 7.8$; 8 high school finalists, 9 bachelor degree holders and 3 master degree holders.

Users were asked to execute tasks spread across several scenarios, which involved previously chosen articles.

These tasks were presented to the users after some introductory steps, which included a verbal explanation of the goals and of the motivation behind the visualization; an introduction to and demonstration of the visualization features and elements; and some minutes of unmonitored experimentation time, where users were allowed to use the visualization and ask questions on its elements, in order to make sure they were already used to the most important features when executing the tasks.

During the execution of the tasks, information was collected regarding the duration of the task, the number of mouse button clicks and keyboard key presses and the final outcome of each task (whether the user was successful and whether they committed any error).

Additional information was then collected through a written questionnaire, split in three parts:

- *personal questions* asked in order to assess the demographics of the inquired population, comprising gender, age, education level, computer experience and Wikipedia experience

- *scaled questions on the system usability* asked using the System Usability Scale introduced by John Brooke[7], a generic Likert scale that allows a numeric evaluation of

the system usability and its comparison with other systems evaluated using the same scale.

- *open questions* regarding the visualization usefulness and the fitness of the provided metrics, and any other comments and suggestions the inquired users want to share on the visualization.

This protocol was then tested and refined by doing two preliminary user tests, whose results were not included in the analysis presented in this chapter.

Scenarios were as follows:

Scenario 1
This scenario features tasks that were built around the English Wikipedia article "List of common misconceptions". References in several popular media[18, 12] during early 2011 led to a sudden increase in the edit rate and article size. Tasks were:

1. Visualize the English Wikipedia article "List of Common Misconceptions";

2. Identify article activity changes and when did those happen;

3. Identify the contributors whose contributions survive longer.

With the second task, we wanted to find out whether users could identify the activity peak caused by the references to the article during January 2011. Although users may possibly try to use our metrics, as we depict the occurrence of changes over the timeline, this task is mostly intended to test whether users are able to understand the timeline, that is, identify parts of it which present an higher density of markers.

In the third task, we ask users to identify the authors whose contributions survive for longer periods of time, thus assessing whether the visualization enables users to assess authorship information, through its "color: by author" plot area fill setting, ideally identifying the most active authors by their non-gray colors, and then assessing their names using the tooltip.

Scenario 2
A scenario which involves an article comparison task, and the extraction of information from the content. It is based on the English Wikipedia articles "42" and "73", articles about the years 42 and 73 of the Julian calendar. Even if not directly related, we found the article 42 to be a frequent target of vandalism and other controversial changes, regarding the popularity of the number, which plays a central role in Douglas Adams' novel "The Hitchhiker's Guide to the Galaxy", where it is presented as "The answer to the Ultimate Question of Life, the Universe, and Everything"[1].

1. Visualize the English Wikipedia article "42";

2. Visualize the English Wikipedia article "73";

3. Rate the activity level of the article "42", when compared with "73" (faint, normal, excessive);

4. Identify an explanation for the activity level rating you chose.

With the third task, we want to find out whether the visualization enables users to successfully compare articles, by checking whether users consider that the article "42" is more active than the article "73", by comparing the density of markers in the timeline along time.

The fourth task, is designed to test the content pane, as it is concerned with the assessment of a justification for the observed difference, in the case the user reports a difference.

Scenario 3

The third scenario concerns the identification of activity rate changes and the extraction of information from the periods when the activity rate changed, along with authorship information. That is, information extraction at a deeper level than in the two previous scenarios, where authorship and information were assessed for the article as a whole.

Tasks are focused on the English Wikipedia article "Elephant", which was vandalized on 31 July 2006, following a request made by Stephen Colbert on his television show "The Colbert Report"[10].

1. Visualize the English Wikipedia article "Elephant";

2. Look for abnormal activity regions;

3. Identify an explanation for the abnormal activity;

4. Identify the contributor who triggered the abnormal activity.

Scenario 4

Then, we tested if users were able to spot activity changes when comparing two articles on the same topic from different Wikipedias: the English Wikipedia and Hungarian Wikipedia articles on Schmitt Pál ("Pál Schmitt" and "Schmitt Pál", respectively). These articles, previously shown as case studies, were found to exhibit two peaks, on January 2012 and on March/April 2012, which occur on both articles, starting first on the Hungarian Wikipedia.

1. Visualize, in one of the plots, the English Wikipedia article "Pál Schmitt";

2. Visualize, in the other plot, the Hungarian ("Magyar") Wikipedia article "Schmitt Pál";

3. Look for abnormal activity regions;

4. Of the regions you found, identify those that only occur in one of the Wikipedias;

5. For the regions which occur on both Wikipedias, identify the one where the abnormal activity started first.

With the third task, we once again intend to test whether the visualization enables users to spot changes in the activity rate, by pointing at least one of the 2012 peaks.

On the fourth task, we ask users if these peaks span both Wikipedias, thus assessing whether users are able to compare articles using the visualization.

Scenario 5

A scenario involving article comparison together with information extraction, on the Portuguese Wikipedia articles on the Portuguese comedians "Zé Diogo Quintela" and "Ricardo de Araújo Pereira". Users were asked to

1. Visualize, in one of the plots, the Portuguese Wikipedia article "Ricardo de Araújo Pereira";

2. Visualize, in the other plot, the Portuguese Wikipedia article "Zé Diogo Quintela";

3. Check if there is any correlation between the activity rate of the two articles and, if so, identify a possible reason.

With the first task, we once again test whether users can compare two articles and find relations among their activity rates, by inspecting the timeline marker density and the evolution of article length along time, while the fourth task is intended to test information extraction with the help of the content pane.

Results

These tasks are spread across four major high-level goals, as follows[6]:

- *Spotting activity rate changes*, a goal covered by three tasks: scenario 1 task 2 (where 17 out of 20 users were successful), scenario 3 task 2 (18 of 20) and scenario 4 task 3 (20 of 20). That is, users were mostly successful, with an average success rate of, approximately, 91.67%;

- *Assessing authorship*, where only half of the users were able to correctly obtain information on the most frequent authors of an article (scenario 1 task 3, 10 users out of 20, success rate: 50%);

- *Comparing articles*, comprising four tasks (scenario 2 task 3 (where 19 users were successful, in 20), scenario 4 task 4 (part 1, with 8 successful users in 13; part 2 with 12 in 14), scenario 4 task 5 (part 1, 7 in 8; part 2 with 12 in 14) and scenario 5 task 3 (part 1: 17 on 20)), where users were, again, successful, with an average success rate of 83%.

- *Information extraction*, which encompasses two tasks (scenario 2 task 4 (14 users out of 19), scenario 5 task 3 (part 2: 17 of 17)) where users were asked to retrieve justifications or explanations for the observed patterns. Once again, users were successful, leading to an average success rate of 86%.

Tests have, thus, shown that the proposed visualization fulfills its goals concerning the identification of activity patterns, comparison of articles and information extraction, tasks which involve, mainly, the timeline and its markers and the content pane, also involving the analysis and comparison of the numeric metrics.

On the other hand, tests also showed that users had problems assessing authorship information (plot areas filled with "color: by author") This may be due to the the great volume of information available through the visualization, which

---

[6]We omit tasks 3 and 4 from the third scenario, as no user identified the specific peak we intended to study, thus rendering the results useless.

hampers the fulfillment of this task, but is also evidence that this visualization feature needs to be redesigned, as its current approach is clearly inefficient.

The time-series approach we chose to guide the design of our visualization turned out to work as expected, effectively condensing an enormous amount of sequential data in a single screen.

Tests also highlighted the need for some graphical optimization, in order to handle hidden data and resolve time and space complexity issues in the rendering stage.

User feedback

All users considered the visualization fulfilled its goals, with one user noting that it will have to be used together with other methods if we want to confirm our findings.

Users were then asked which metrics did they find to be the most useful, with length being chosen by 11 users, followed by quality (7), controversy (6), vandalism (5) and color by author (4).

Some users found metrics hard to understand, suggesting that we add some explanations or review the chosen metric names. As many scenarios involved increased activity, it was suggested that we include information on visit counts.

Several users suggested the addition of help screens or explanations, in order to help first-time users understand the visualization elements, possibly along with some information on what action is currently associated to each of the mouse buttons and operations.

Two users suggested the addition of a zoom-fit feature, that would change the visualization date range to encompass the entire article data. It was also suggested that a tooltip is shown when hovering a timeline marker.

Some usability problems were pointed out, such as the detailed sensitivity of the tooltip locking feature, which, when the pointer is moved before releasing the mouse button, even if just for a couple pixels, triggers zoom instead; the lack of a way to cancel a pending request; and the unability to hide one of the plots when it is not used (although not essential, this would help users focus on the plot they are analyzing). A user also suggested that the date input fields should be complemented with a calendar-like mouse-based date picker.

Comments were also raised on whether it would be possible to find a better alignment between the plot options and the plots. Users complained the tooltip would be sometimes rendered unreadable, as there is no code in place to avoid getting its content cut at the bottom and right ends of the screen. One user found the gradient colors confusing and unintuitive.

During one of the tests, the idea of adding, in the tooltip, the exact date under the pointer was suggested. Although, in most cases, this date would not correspond to any edition, this would make it easier for users to extract dates from the visualization.

Some users were expecting the color "by author" gradient to follow some key. Although the main idea was just to highlight

the most frequent authors, we welcome the suggestion, which could translate at least on a key where each frequency is always assigned the same color, and possibly in a key which explicitly lists the authors currently associated with each color.

A user pointed out that the system becomes harder to interact with and use when it tries to convey large amounts of information. Some optimizations, introduced after user tests, fix some of the speed and visualization issues caused by the huge volumes of information associated with some articles, although this does not exclude the possibility of applying additional optimizations, in order to improve the visualization even more. It was also suggested that the visualization should convey derivatives of the metrics, or provide some derivative-based metric.

Zoom was also pointed as an operation that could be improved, as it only works over the plot area, not working on the white background, along with the unability of the AJAX search box to show the entire text of the results, as those are limited to the width of the selection list widget, two flaws which we fixed in later versions.

CONCLUSIONS

Driven by the belief that the growth of Wikipedia, given its openness (anyone can edit (most of) the articles), translates into a convergence of the trends and behaviors of its users towards public opinion, we decided to conceive a new visualization that shows these trends and behaviors in an innovative way.

Based on the insights from related works, we devised a set of metrics we then compute on a revision-by-revision basis, after retrieving the article content from Wikipedia.

These metrics, which comprise a time series, are then used to build a visualization focused on depicting the evolution of these metrics through time, while also enabling quick access to revision content.

Plots are accompanied by their timelines, where article and talk page changes are marked along time, thus providing a quick visual depiction of the article activity along time.

We then analyzed several case studies, where we could find patterns, relate changes to real-world events and compare articles on different topics.

We carried user tests, finding out that, while our visualization effectively enables users to identify activity patterns and compare articles, some more active articles make it harder for users to spot activity changes. It was also observed that users could successfully extract information through the visualization.

Test results showed that the time-series approach we chose to guide the design of our visualization turned out to work as expected, condensing a possibly large volume of sequential data in a single screen.

REFERENCES
1. Adams, D. N. H2G2: The hitchhickers guide to the galaxy, 1979.

2. Adler, B. T., and de Alfaro, L. A content-driven reputation system for the wikipedia. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, ACM (New York, NY, USA, 2007), 261–270.

3. Bao, P., Hecht, B., Carton, S., Quaderi, M., Horn, M., and Gergle, D. Omnipedia: bridging the wikipedia language gap. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, ACM (New York, NY, USA, 2012), 1075–1084.

4. Berners-Lee, T., and Cailliau, R. Worldwideweb: Proposal for a hypertext project. *European Particle Physics Laboratory (CERN)* (1990). `http://www.w3.org/Proposal.html`.

5. Bostock, M., Ogievetsky, V., and Heer, J. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)* (2011).

6. Brandes, U., and Lerner, J. Visual analysis of controversy in user-generated encyclopedias. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on* (30 2007-nov. 1 2007), 179 –186.

7. Brooke, J. SUS — A quick and dirty usability scale. *Usability evaluation in industry 189* (1996), 194. `http://www.usabilitynet.org/trump/documents/Suschapt.doc`.

8. Chevalier, F., Huot, S., and Fekete, J.-D. Wikipediaviz: Conveying article quality for casual wikipedia readers. In *Pacific Visualization Symposium (PacificVis), 2010 IEEE* (march 2010), 49 –56.

9. Coelho, P. P. Candidato do PSD à presidência da assembleia da república, April 2011. `https://www.facebook.com/notes/pedro-passos-coelho/candidato-do-psd-%C3%A0-presid%C3%AAncia-da-assembleia-da-rep%C3%BAblica/10150153318796246`.

10. Colbert, S. The Colbert Report — The Word: Wikiality, July 2006. `http://www.colbertnation.com/the-colbert-report-videos/72347/july-31-2006/the-word---wikiality`.

11. Cunningham, W. Wiki wiki web, 1995. `http://c2.com/cgi/wiki?WikiWikiWeb`.

12. Doctorow, C. Wikipedia's list of common misconceptions, January 2011. `http://www.boingboing.net/2011/01/11/wikipedias-list-of-c.html`.

13. Harrison, C. clusterball, July 2008. `http://www.chrisharrison.net/projects/clusterball/index.html`.

14. Harrison, C. Wikiviz, September 2009. `http://www.chrisharrison.net/projects/wikiviz/index.html`.

15. Havre, S., Hetzler, E., Whitney, P., and Nowell, L. Themeriver: visualizing thematic changes in large document collections. *Visualization and Computer Graphics, IEEE Transactions on 8*, 1 (jan/mar 2002), 9 –20.

16. Holloway, T., Bozicevic, M., and Börner, K. Analyzing and visualizing the semantic coverage of wikipedia and its authors: Research articles. *Complex. 12* (January 2007), 30–40.

17. Mola-Velasco, S. M. Wikipedia Vandalism Detection Through Machine Learning: Feature Review and New Proposals - Lab Report for PAN at CLEF 2010. In *Notebook Papers of CLEF 2010 LABs and Workshops*, M. Braschler, D. Harman, and E. Pianta, Eds. (Padua, Italy, September 2010).

18. Munroe, R. XKCD #843: Misconceptions, January 2011. `http://www.xkcd.com/843/`.

19. Potthast, M., Stein, B., and Gerling, R. Automatic vandalism detection in wikipedia. In *Advances in Information Retrieval*, C. Macdonald, I. Ounis, V. Plachouras, I. Ruthven, and R. White, Eds., vol. 4956 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2008, 663–668.

20. Riche, N. H., Lee, B., and Chevalier, F. ichase: supporting exploration and awareness of editing activities on wikipedia. In *Proceedings of the International Conference on Advanced Visual Interfaces*, AVI '10, ACM (New York, NY, USA, 2010), 59–66.

21. Rodrigues, S., and Simas, N. Fernando Nobre renuncia ao mandato de deputado, July 2011. `http://publico.pt/1501403`.

22. Suh, B., Chi, E., Pendleton, B., and Kittur, A. Us vs. them: Understanding social dynamics in wikipedia with revert graph visualizations. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on* (30 2007-nov. 1 2007), 163 –170.

23. Suh, B., Chi, E. H., Kittur, A., and Pendleton, B. A. . 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*, CHI '08, ACM (New York, NY, USA, 2008), 1037–1040.

24. Viégas, F. B., Wattenberg, M., and Dave, K. Studying cooperation and conflict between authors with history flow visualizations, April 2004.

25. Vuong, B.-Q., Lim, E.-P., Sun, A., Le, M.-T., and Lauw, H. W. On ranking controversies in wikipedia: models and evaluation. In *Proceedings of the international conference on Web search and web data mining*, WSDM '08, ACM (New York, NY, USA, 2008), 171–182.

26. Wilkinson, J. Wikiswarm: visualize wikipedia page histories, January 2009. `http://jamiedubs.com/wikiswarm-visualize-wikipedia-page-histories`.