

WikiTrip: animated visualization over time of gender and geo-location of Wikipedians who edited a page

Paolo Massa, Maurizio Napolitano, Federico Scrinzi
{massa, napolita, fscrinzi}@fbk.eu
Bruno Kessler Foundation
Trento, Italy

WikiTrip allows to have a trip in the process of creation of any Wikipedia page from any language edition of Wikipedia.

WikiTrip is an interactive web tool empowering its users by providing an insightful visualization of two kinds of information about the Wikipedians who edited the selected page: their location in the world and their gender. If you want to investigate, for example, where in the world are Wikipedians who edited the page “Peace”, WikiTrip is the right tool. And you can check also the origin of edits for the equivalent page “سلام” in the Arabic Wikipedia or “Amani” in the Swahili Wikipedia. Moreover, if you have ever wondered if a specific page was edited more by male or female Wikipedians, WikiTrip allows to explore this information as well. Visualization of both information is available over time so that you can appreciate the evolution of the page over years, from its creation up to the present. WikiTrip is available at <http://sonetlab.fbk.eu/wikitrip/>

Locations in the world are visualized on a zoomable and scrollable map and, in order to deal with large datasets of points, they are clustered at runtime in bubbles of varying dimensions depending on the number of points in that location. The 10 countries from which most edits came are visualized also in a specific bar plot. The location is visualized only for edits made by anonymous users since they are identified by their IP address and this can be mapped by WikiTrip to the place in the world from which the user edited Wikipedia. For privacy reasons, this IP address is not made available by Wikipedia for registered users.

The other information visualized by the WikiTrip interface is related to gender and allows to see how many edits to the selected page have been done respectively by male and female Wikipedians. The gender of Wikipedians is available only for registered users who have set this information in their preferences, it is a feature that was introduced in Mediawiki in 2009 and currently it is not used by all users (around 17% of users in the English Wikipedia expressed their gender as of August 2011). Nevertheless, the number of edits by males and females, visualized as an additional bar plot on the right side of the interface, allows to insightfully investigate the balance (or imbalance) of genders in Wikipedia.

Both information (geo-locations and genders) are shown in a longitudinal manner as they evolve over time, so that is possible to appreciate the entire history of the selected Wikipedia page from its creation up to the present day. The animation of the history of the edits to the page starts automatically so that the user can possibly grasp how the interface works in a more intuitive way by seeing it in action. However it is possible to stop and restart the visualization, to make it unfold faster or slower and to scroll to any point in time simply by dragging the time slider. Another plot, on top of interface, shows two time series, the number of edits performed by registered users (for which gender information is shown if available) and by anonymous users (for which geo-location is shown) so that it is possible to see the quantitative evolution of the page over time.

The user of WikiTrip can choose, by means of a checkbox, to see every information in two different ways: (1) in a cumulative way so that the last time window shows all the edits performed since the

creation of the page, or (2) as the distribution of edits over time so that each time window only shows edits performed in a given period of few months computed at runtime depending on the creation and current dates. In this second way, it is possible to better appreciate changes over time, for example it might allow to discover that soon after its creation the page was edited mainly by males and from USA while later the page was edited mostly by females and from Italy or that at a certain point in time there was a spike in editing activity, possibly from a specific location in the world.

Each wikiTrip visualization offers the possibility to view the equivalent page in any other language edition of Wikipedia by exploiting the interwiki links present on the page. The interface provides a dropdown list showing all the other language editions in which the equivalent page exists. In this way it is possible to answer to questions such as “is the page Melbourne on English Wikipedia edited most from Melbourne area?” and “from where the page Wikipedia on Latin Wikipedia is edited the most?”, “What about the Catalan Wikipedia? And Esperanto? And Wolof or Swahili?”. The structure of the URL is transparent and clear. For example the wikiTrip for the English Wikipedia page “Peace” is <http://sonetlab.fbk.eu/wikitrip/#|en|Peace> while for the German Wikipedia equivalent page “Frieden” is at <http://sonetlab.fbk.eu/wikitrip/#|de|Frieden> and the Arabic equivalent is at <http://sonetlab.fbk.eu/wikitrip/#|ar|سلام>. For example, for the page “Peace” (used in the attached screenshots as an example) there are, as of August 2011, interwiki linked pages in 94 different language editions of Wikipedia.

On each WikiTrip visualization, the interface also offers links to the related article in Wikipedia and to a comparison on Manypedia (<http://www.manypedia.com>), another web tool we developed whose goal is to ease comparison of the same pages from two language editions of Wikipedia.

All the visualized information is computed at runtime and nothing is precomputed so that the visualizations are always up to date and able to visualize last minute changes as well. This is especially interesting for recent events or pages which recently received a significant increase of attention. Data comes from scripts we wrote in PHP and run server-side on toolserver.org, a cluster of servers operated by Wikimedia Deutschland e. V. offering a replicated database of Wikimedia's projects. Note that sometimes toolserver machines and databases are down since the system is mainly maintained by volunteers and gender information are not always replicated on toolserver databases, especially for relatively less used language editions of Wikipedia. When toolserver machines and databases are down or don't have the required information, WikiTrip interface does not break but simply does not show the missing information. Some computation, for example about aggregating points on the map, happens also at runtime client-side via Javascript.

We placed special attention towards the openness of WikiTrip: all the software we wrote is released as open source and Free Software so that anyone can reuse and improve our initial effort, the visualizations themselves are released under Creative Commons so that anyone can reuse them or integrate them in other works. Visualized data are served through easy-to-use Application Programming Interfaces (API) so that independent developers can use them in order to build different visualizations and services.

WikiTrip is powered by PHP and Javascript code. PHP have been used on the server side (on toolserver.org) to query the databases and provide up to date information about evolution of genders and geo-locations over time. For associating IP addresses to world location, WikiTrip uses MaxMind GeoIP Country database and code which is released under LGPL. The code of the scripts we wrote is released under Affero GPL 3.0, so that anybody can read our code and improve it and, in case someone improves it and provides the modified version as “software as a service”, the improvements to the code must be redistributed to the community as well. PHP and other useful Python code is available at <https://github.com/volpino/toolserver-scripts/>

The interface uses the JQuery library and some of its plugins and the OpenLayers library. These libraries are open source. The map shown in the interface uses NASA Blue Marble global composite satellite imagery, in the public domain, and political boundaries and cities from OpenStreetMap contributors, released under CC-BY-SA. The Javascript and HTML code we wrote for the WikiTrip interface has been released as open source as well and is available at <https://github.com/volpino/wikipedia-timeline> We have also released as open source a separate plugin for the Openlayer framework so that others can reuse it independently of WikiTrip. Its specific purpose is to provide a way to show GeoJSON/GeoRSS data on a map through time, supporting cluster strategy for large datasets (OpenLayers.Strategy.Cluster) and time filtering. It is available at <https://github.com/volpino/openlayers-timeline> The visualizations themselves are released under Creative Commons By-Attribution Share-Alike license.

The visualized information does not come from pre-computed datasets but it is computed live from the Toolserver.org databases. Anyway the visualized information are served via three easy-to-use APIs and any other developer can build on them. APIs return values in JSON format and are explained with the three following examples but a more complete documentation is in the “Read more” page on WikiTrip site.

- `api_gender.php`: The HTTP GET request `http://toolserver.org/~sonet/api_gender.php?article=سلام&lang=ar` returns timestamp and gender for any edit to the specific page in the specific language edition of Wikipedia by registered users that specified their gender.
- `api_geojson.php`: The HTTP GET request `http://toolserver.org/~sonet/api_geojson.php?article=Peace&lang=en` returns a GeoJSON for anonymous edits on a specific page
- `api.php`: The HTTP GET request `http://toolserver.org/~sonet/api.php?article=London&lang=en&editors&max_editors=5` returns statistics about a page such as edits per month, unique registered and anonymous editors and more.

We believe our interactive visualization Web tool fulfills well the evaluation criteria as we explain in the following.

WikiTrip visualization is innovative in presenting in a single, clean, interactive Web interface information about the process behind the formation of a Wikipedia page and its evolution in time. In particular it is the first tool (to the best of our knowledge) to give an easy opportunity to anyone who is not a programmer to investigate the gender balance for any page on any language edition of Wikipedia. There is a lot of attention recently to the gender balance (or imbalance) after the January 30, 2011 New York Times article “Define Gender Gap? Look Up Wikipedia’s Contributor List” and the Wikimedia Foundation goal to raise the share of female contributors to 25 percent by 2015 and the many interesting discussions on the “gendergap” mailing list recently created by the Wikimedia Foundation. We believe visualizing gender information at this level of granularity (each page from any language edition of Wikipedia) is novel and can be insightful in itself.

WikiTrip is also innovative in the sense that both information (gender and geo-location) are visualized as they evolve over time and users can explore the time dimension as well. We believe WikiTrip, along with other tools, can also be revealing for the non-Wikipedian of editing dynamics to Wikipedia pages that are in general hidden and unknown.

We believe the interface is self-explanatory, intuitive and clear. The fact the animation starts automatically (but can be stopped) should make it easier for the user to understand how to interact with the Web interface and which information are visualized. The conceptual elements on which it is built (space, time and plot bars) are among the simplest ones that can be used in visualizations. Moreover, by

clicking on the “Get help” link, different tooltips appear which contextually explain the different part of the interface, the information visualized and how to interact with them. Currently the interface textual elements are in English only. We didn't translate them providing to the user the possibility to see the interface in its preferred language for time reasons. However, since all the software is open source, it is easy to translate them especially as a result of a possible collaborative effort in which different people in the world provide a translation of the few text strings from English into their language. Interestingly, WikiTrip allows to explore the evolution of a Wikipedia page from any of the more than 280 language editions of Wikipedia, possibly offering interesting insights for cross-cultural studies.

The data on which WikiTrip is built are available to anyone and have been offered also as results of different APIs which can be freely queried by any developer for developing other tools. We like to deeply thank the Wikimedia Deutschland e. V and the Wikimedia Foundation for providing Toolserver.org, a cluster of servers offering a replicated database of Wikimedia's projects, from which we extract at run time the information WikiTrip visualizes.

Our special attention to openness (both in the tools we built on and the fact we released all our code as open source and Free Software) also makes it possible to improve and build on our initial effort. For example, while we don't have access to IP addresses of registered users for privacy reasons, few users highly trusted by the Wikimedia Foundation have access to this information (i.e. they have the checkuser flag). Would the Wikimedia Foundation consider WikiTrip interesting, the fact we release it as open source would make it totally possible to install it on some wikimedia.org servers and, modifying it by making it run under a user with checkuser flag, visualize geo-locations also for registered users, possibly by adding some random noise to the geo-location to protect privacy of each Wikipedia user. This is just one of the possibilities available given that we released the code as Free Software.

In attachment there are three screenshots (licensed under CC-BY-SA 3.0) of WikiTrip visualizing the pages “Peace” from English, German and Spanish Wikipedia. The URLs for direct access are <http://sonetlab.fbk.eu/wikitrip/#|en|Peace> , <http://sonetlab.fbk.eu/wikitrip/#|de|Frieden> , <http://sonetlab.fbk.eu/wikitrip/#|es|Paz|1261024534|0> (in the unlikely case that, during the evaluation of WikiTrip, the web site is down, we have provided a video of its functioning when browsing the previous 2 URLs. The video is available at http://www.gnuband.org/tmp/wikitrip_video_001.mpeg)

However it is possible to explore the WikiTrip visualizations of the page “Peace” from the 94 different language editions of Wikipedia in which this page currently exists. While the WikiTrip visualization for the English Wikipedia allows to appreciate how edits to the page “Peace” comes from all over the world and edits to the equivalent page on German Wikipedia allows to appreciate a non-negligible portion of edits by female Wikipedians, each of the 94 visualizations can shed a light on how Wikipedians concur to create this important concept. But of course this is just one exploration, WikiTrip allows to have an insightful trip on the creation process of any page of any language edition of Wikipedia.