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

Paolo Massa

Bruno Kessler Foundation
Via Sommarive, 18
Trento, Italy
massa@fbk.eu

Maurizio Napolitano

Bruno Kessler Foundation
Via Sommarive, 18
Trento, Italy
napolitano@fbk.eu

Federico Scrinzi

Bruno Kessler Foundation
Via Sommarive, 18
Trento, Italy
fscrinzi@fbk.eu

Michela Ferron

Center for Mind/Brain Sciences
University of Trento
Corso Bettini 31, Rovereto, 38068
m.ferron@unitn.it

Abstract

In this short paper, we present WikiTrip, a web tool we created and released as open source which provides a visualization over time of two kinds of information about the Wikipedians who edited a selected page: their location in the world and their gender. We also describe evidence that pages on a language edition of Wikipedia which receive most attention in terms of edits from countries where the language is not primarily spoken are about TV shows and stars, football teams or specific geographic locations.

Author Keywords

Wikipedia; web tool; visualization; location; gender; animation over time; open source

Introduction

Wikis are a simple and powerful technology for collective creation of resources. Wikipedia is the most striking example of the application of such a technology. For instance, since its creation in 2001, the almost 4 million article pages on English Wikipedia have received more than 526 million edits by more than 16 million registered users. All these contributions are the result of volunteer work since none of the editors is paid for contributing to Wikipedia, “the online

encyclopedia anyone can edit". It is even possible to change any page without logging in the system with the personal login and password and hence totally anonymously.

This extreme level of openness gives rise to questions about the reliability of each single article on Wikipedia [1]. For instance, while 53% of American Internet users look for information on Wikipedia as of May 2010 [2] probably making the online encyclopedia the most visited space for information needs, it is not clear how many of the readers of Wikipedia are aware of the peculiar characteristics of the information source they rely on so heavily. A survey of 165 undergraduate students of an American University who are regular readers of Wikipedia showed that only 57.0% of them answered correctly that "anyone can add a Wikipedia page on any topic" and 81.8% were wrong in believing that "You must create an account in order to edit or create content on Wikipedia" being unaware of the possibility of editing anonymously [3]. These results clearly highlight that the inner functioning of Wikipedia is unknown to most of its readers.

Wikipedia offers an unmatched level of transparency with regard to all the activity that happens on site. In fact it is possible to access the complete history of every page showing every single edit along with the text inserted and removed, the username of the editor and the time, simply by clicking on the "View history" textual link present on top of each page. However according to the same survey [3], 68% of students wrongly believed that "Wikipedia cannot keep track of my edits when I'm not signed-in to my account" and

only 20.6% knew that it was possible to go back in the edit history of a Wikipedia article up to its creation.

This evidence suggests the importance of "lifting the veil" on Wikipedia inner functioning in order to improve accountability and social transparency [1].

In this short paper we present our contribution in highlighting aspects of the creation of a Wikipedia page which are often unknown and difficult to gather. Precisely we present WikiTrip, a web tool we released as open source which provides an insightful visualization of two kinds of information about the Wikipedians who edited a selected page: their location in the world and their gender. The visualization is longitudinal since it shows the evolution of edits to the page from its creation up to present. Moreover it works for every page from every language editions of Wikipedia (while the English version is the oldest and largest one, at April 2012 there are editions of Wikipedia in more than 280 languages).

WikiTrip Web Tool

Wikidashboard [1] was the first attempt to complement Wikipedia pages by adding visual clues about the recent edit history of a page, such as sudden bursts of edit, and most active editors. The graph allows to examine patterns on how an article evolves over time between multiple editors.

With WikiTrip, we decided to focus on two other aspects of the creation process of a Wikipedia page which are even less evident and easy to discover: where in the world edits to the specific page comes from and gender of the person performing those edits.

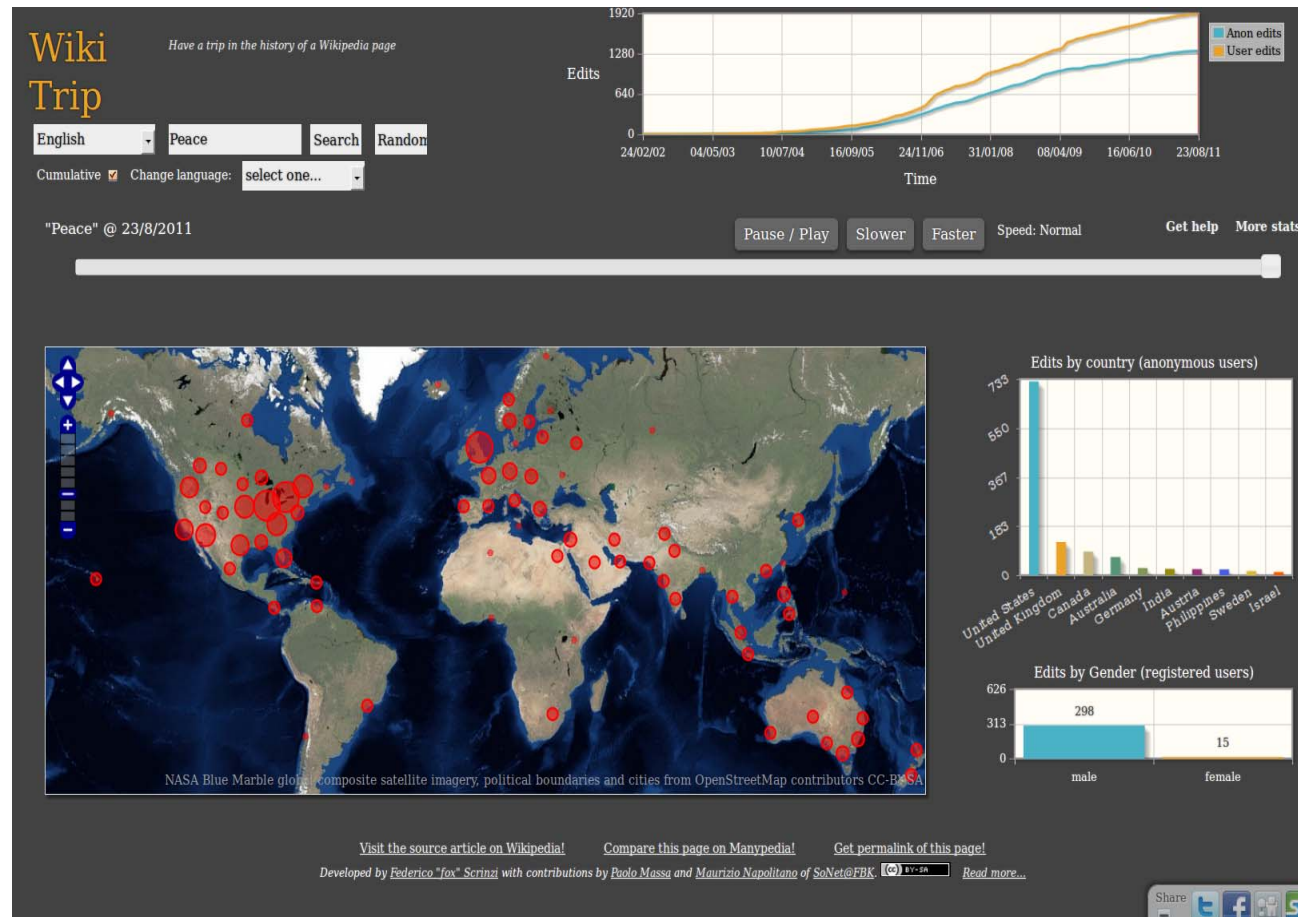


Figure 1. Screenshot of WikiTrip for the page "Peace" from English Wikipedia

Figure 1 shows a screenshot of 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 18% of users in the English Wikipedia expressed their gender as of January 2012). 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 (this feature is similar to what Wikidashboard [1] offers). 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” there are, as of April 2012, 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.

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. Moreover, 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.

The visualized information does not come from pre-computed datasets but it is computed live on toolserver.org, a cluster of servers offering a replicated database of Wikimedia's projects. The visualized information are served via three easy-to-use APIs and any other developer can build on them (documentation on WikiTrip site). For example, the HTTP GET request http://toolserver.org/~sonet/api_gender.php?article=سلام&lang=ar returns a JSON representation of timestamp and gender for any edit to the specific page

in the specific language edition of Wikipedia by registered users that specified their gender.

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.

Related Work and Discussion on Possible Uses of WikiTrip

Wikiganda [4], previously known as Wiki Scanner, is an attempt which uses an intuition similar to ours. It converts IP addresses of anonymous edits into the corresponding Internet domain (for example, cia.gov or vatican.va) and it is hence able to show which pages receive the anonymous edit attention of different organizations, which might be called Wikiganda or propaganda through Wikipedia.

Other previous work focused on the geographical aspects of participation in Wikipedia. Hecht and Gergle [5,6] showed that each language edition of Wikipedia is characterized by a level of self-focus bias operationalized as number of links directed at articles located in the region of the world where that language is largely spoken. Moreover they found that each language Wikipedia exhibits a self-focus bias towards articles about regions where that language is largely spoken.

In both cases, their focus is at the level of characterizing the entire Wikipedia. On the other hand, by developing WikiTrip our goal was to create a Web tool empowering anyone to analyze where edits to a single specific page originated from.

Authors of "You Are Where You Edit: Locating Wikipedia Contributors Through Edit Histories" [7] showed that it is often possible to associate contributors with relatively small geographic regions, usually corresponding to where they were born or where they presently live, simply looking at the location of pages they edit. This phenomenon holds also for anonymous contributors [8]. This results indicate that it can be the case that topics that are highly relevant for a specific regional zone tend to be edited mainly from that zone leading to a self-focus bias [5] also at the level of a specific Wikipedia page. In fact while most edits to all language editions of Wikipedia tend to come from US, Germany, UK and France and, for example, there are more edits than originate in Hong Kong each quarter than the entire continent of Africa [9], it is also true that most edits to English Wikipedia comes from US and UK while for example Italian Wikipedia is edited mainly from

Italy, i.e. language editions of Wikipedia present a self-focus bias [5].

In order to investigate the self-focus bias at the level of single specific pages instead of an entire Wikipedia, we created a script listing all pages from English Wikipedia which were edited anonymously mainly from countries different from US and UK. Table 1 lists the 30 pages which received most anonymous edits. We manually classified their type and it is possible to observe that most of the pages which received a significant attention mainly from a specific part of the world where English is not the primary language were of three types: about TV or cinema stars and programs such as ABS-CBN Corporation programs edited mainly from Philippines, about football teams such as Fenerbahçe_S.K. From Turkey and about specific cities or countries such as Portugal edited mainly from this country.

The same script run on Italian, French and other Wikipedias (excluding countries in which the specific language is spoken most) shows the same results, i.e. pages about football teams, specific cities or countries and TV programs and stars attract attention from other parts of the world reflecting a regional bias.

Wikipedia page	Type	#Edits	Edits	#Anon Edits	Most anon edits	Seco mos anon t
List of programs broadcast by ABS-CBN Corporation	TV	18098	12732	Philippines	10180	US 1121
FC Barcelona	FB	17471	8278	Spain	1350	US 1267
Vijay (actor)	TV	14007	8010	India	3826	UK 756
List of programs broadcast by GMA Network	TV	14153	7959	Philippines	7233	US 330
Home and Away	TV	12597	7221	Australia	4603	UK 1781
Fenerbahçe_S.K.	FB	10196	5945	Turkey	3013	UK 982
Portugal	GEO	12606	4915	Portugal	1967	US 1497
Olympiacos F.C.	FB	9055	4618	Greece	3016	UK 396
S.L. Benfica	FB	8487	4567	Portugal	2309	US 876
Brazil	GEO	13126	4515	Brazil	1673	US 1624
Canada	GEO	16901	4327	Canada	1979	US 1745
Athens	GEO	9399	4299	Greece	1393	US 1327
List of One Day International cricket records	SPORT	5657	4238	India	2754	US 326
Toronto	GEO	11127	4156	Canada	2343	US 1526
know	OTHER	303401	4134	Canada	2979	US 551
Galatasaray S.K.	FB	6968	4030	Turkey	2131	Neth 459
Finland	GEO	9920	4028	Finland	1450	US 1376
Panathinaikos F.C.	FB	5923	4005	Greece	3213	UK 109
SkyCable	TV	6001	3980	Philippines	3766	US 144
List of Serbs	GEO	5689	3977	Canada	2758	US 368
Hyderabad, India	GEO	9502	3955	India	2238	US 1062
List of fraternities and sororities in the Philippines	GEO	6782	3925	Philippines	2533	US 1044
Hungary	GEO	9802	3894	Hungary	1536	US 962
Ajith Kumar	TV	8041	3890	India	2005	UK 552
AEK Athens F.C.	FB	7070	3798	Greece	2720	UK 239
New Zealand	GEO	11198	3786	New Zealand	1308	US 972
Singapore	GEO	11080	3760	Singapore	1361	US 957
Rajinikanth	TV	8891	3731	India	1983	US 768
Chiranjeevi	TV	7130	3726	India	1912	US 1365
Sao Paulo	GEO	7674	3555	Brazil	2257	US 522

Table 1. 30 pages from English Wikipedia edited anonymously mainly from countries different from US and UK.

The other relevant piece of information “unveiled” by WikiTrip is the gender of editors of a specific page. There was a lot of attention to the gender balance (or imbalance) in Wikipedia after the January 2011 New York Times article “Define Gender Gap? Look Up Wikipedia’s Contributor List”, which make it well known that the percentage of women contributing to Wikipedia is around 11%, 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. The topic is of interest also for the research community given for instance the fact that the two best papers of Wikisym 2011 conference were both about gender issues in Wikipedia [11,12].

References

- [1] Bongwon S., Chi Ed H., Kittur A., and Pendleton B. A. 2008. 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, 10, USA, 1037-1040.
- [2] Pew Research Center. 2011. Wikipedia, past and present. A snapshot of current Wikipedia users.. Jan 13, 2011. Retrieved on August 4, 2011, from <http://www.pewinternet.org/Reports/2011/Wikipedia.aspx>
- [3] Antin, J., and C. Cheshire. 2010. “Readers are not Free-Riders: Reading as a Form of Participation on Wikipedia.” in Proceedings of the ACM 2009 conference on Computer supported cooperative work. Savannah, GA.

The difference with WikiTrip again is that it allows the user to investigate this important issue at the level of granularity of each single Wikipedia page and also to do it in a longitudinal manner so that it is also possible to appreciate changes in the balance of a Wikipedia page over time. Sadly, even with the automated scripts we wrote in order to compute the percentage of female editors for each page, it is very hard to find Wikipedia pages edited mainly by females.

Concluding, in this paper we have presented WikiTrip, a web tool we have developed and released as open source whose aim is to increase the social transparency behind the creation of specific Wikipedia pages from any language editions of Wikipedia by highlighting two little explored aspects: geolocations and gender of editors.

- [4] Chandy R. 2008. Wikiganda: Identifying Propaganda Through Text Analysis. Caltech Undergraduate Research Journal. Winter 2008-2009
- [5] Hecht, B., & Gergle, D. 2009. Measuring Self-Focus Bias in Community-Maintained Knowledge Repositories. Proceedings of Communities and Technologies 2009 (C&T 2009), pp. 11-20. New York: ACM Press.
- [6] Hecht, B. and Gergle, D. 2010. On The “Localness” of User-Generated Content. 2010 ACM Conference on Computer Supported Cooperative Work. Savannah, Georgia, USA
- [7] Lieberman, M. D. 2009. You Are Where You Edit: Locating Wikipedia Contributors Through Edit Histories - You Are Where You Edit - - ICWSM 2009, San Jose, CA
- [8] D. Hardy, J. Frew, and M.F. Goodchild. 2012. Volunteered geographic information production as a

spatial process. *International Journal of Geographical Information Science* 26

[9] Graham, M. 2012. Where do Wikipedia edits come from. Blog post retrieved on April 3, 2012 at <http://www.zerogeography.net/2012/02/where-do-wikipedia-edits-come-from.html>

[10] Define Gender Gap? Look Up Wikipedia's Contributor List. 2011. Retrieved on April 3, 2012 at <http://www.nytimes.com/2011/01/31/business/media/31link.html>

[11] Lam, S.K.; Uduwage, A.; Dong, Z.; Sen, S.; Musicant, D.R.; Terveen, L.; Riedl, J. 2011. WP:Clubhouse?: An Exploration of Wikipedia's Gender Imbalance. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*, Mountain View, California.

[12] Antin, J.; Yee, R.; Cheshire, C.; Nov, O. 2011. Gender Differences in Wikipedia Editing. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*, Mountain View, California.