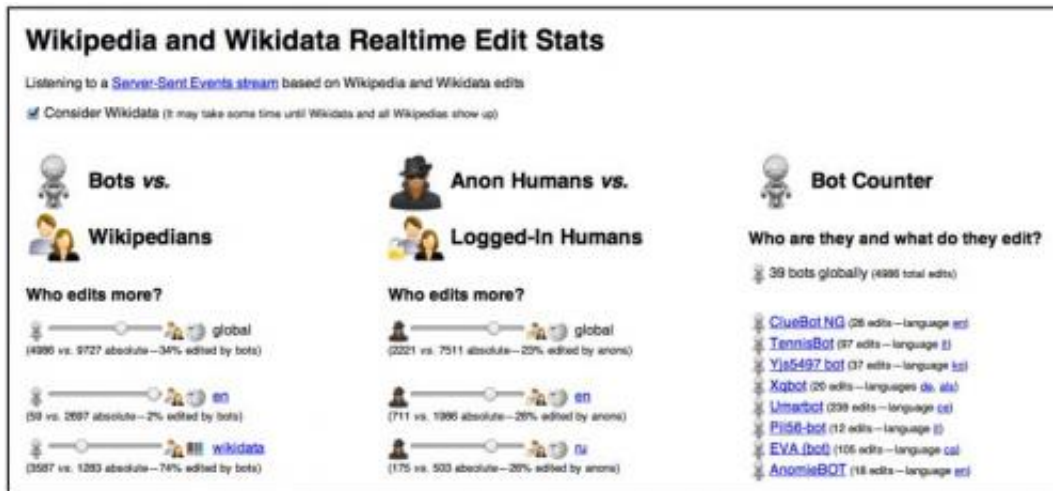# Google engineer creates application that monitors Wikipedia content bots

February 19 2014, by Bob Yirka

Screenshot of the application. Credit: arXiv:1402.0412 [cs.DL]

(Phys.org) —Thomas Steiner, a Customer Solutions Engineer at Google Germany GmbH, Hamburg has created an application that shows in a very clear way, how much of Wikipedia entries are being created or edited by bots, versus humans. He's also written a paper describing his efforts and posted it on the preprint server *arXiv*.

Many people may not realize it, but some of the information appearing on Wikipedia is put there by bots, rather than human beings. This is because Wikipedia has grown too large to be managed by people alone, especially when noting it's still mostly a volunteer effort.

To keep entries coming and to keep them updated, bots have been created—they grab information from one place and post them into another, thus, they're not actually writers or composer, they're more like auditors updating files automatically. Also, many people may not know that the folks at Wikipedia have also created another information repository—Wikidata—it's a database whose sole purpose is to share data amongst the difference language versions of Wikipedia. If a user in the U.S. enters information about the results of the New York Marathon into a Wiki entry, for example, that data can be automatically ported to Wikidata, where other bots can retrieve it, convert it to the pertinent language and post it to another language version of Wikipedia—all rather seamlessly to readers.

Because of all the automation, some have begun to wonder what portion of Wiki pages are generated by humans versus bots. That's where Steiner comes in—he's written an application that can be accessed and used by anyone to see—in real time—what percentage of pages are being written by humans, versus bots.

The application also allows for noting other aspects of Wikipedia—a quick glance, for example reveals that bots are doing a lot more of the work adding information to pages in non-English speaking countries, which suggests that the majority of Wikipedia content is still being created by real human beings in the U.S. and the U.K. The application also monitors activity on Wikidata, for those who are interested and also displays the data for both in a way that shows which bots are most active.

Steiner has also published the code for the application, making it open source. That should allow those who are interested in the murky world of bots to gain an insider's perspective, and perhaps, to add to the applications utility.

  More information: Bots vs. Wikipedians, Anons vs. Logged-Ins,

Abstract

Wikipedia is a global crowdsourced encyclopedia that at time of writing is available in 287 languages. Wikidata is a likewise global crowdsourced knowledge base that provides shared facts to be used by Wikipedias. In the context of this research, we have developed an application and an underlying Application Programming Interface (API) capable of monitoring realtime edit activity of all language versions of Wikipedia and Wikidata. This application allows us to easily analyze edits in order to answer questions such as "Bots vs. Wikipedians, who edits more?", "Which is the most anonymously edited Wikipedia?", or "Who are the bots and what do they edit?". To the best of our knowledge, this is the first time such an analysis could be done in realtime for Wikidata and for really all Wikipedias—large and small. Our application is available publicly online at the URL this http URL, its code has been open-sourced under the Apache 2.0 license.