Wikipedia's miracle

Niederer, S.M.C.

[Link to publication](#)

Interview with Sabine Niederer

Sabine Niederer is director of CREATE-IT, the applied research centre of the School of Digital Media and Creative Industries at the Amsterdam University of Applied Sciences, where she has recently founded the Citizen Data Lab. She is also analyst with the Digital Methods Initiative at the University of Amsterdam. In this interview, we discuss her research, which deals with the role of bots and software agents that co-author online content in the context of Wikipedia.

Nicolas Nova: Observing Wikipedia as a researcher, amongst other kinds of platforms, you mentioned that you were interested in social technical systems and also that you became interested in bots. Can you tell us more about how you became interested in this topic?

SN: Yes, I was very interested in Wikipedia, because it's such a place that has been heralded for its crowd effort. The idea of lot of people volunteering to spend time editing this encyclopaedia makes it an interesting phenomenon. But then, when I learnt that bots and other software agents were also participating, I was really struck because it is not commonly mentioned; and it's not often researched. Web 2.0 has been compared to the Mechanical Turk[1], a robot that played chess and nobody knew or noticed there was a person hiding inside. With Wikipedia, this Mechanical Turk seems to have been inverted. So you have everybody staring at the people and nobody notices that this 'crowd' is in fact part of a large machine. Also, it's not only a machine as in, a platform that enables you to interact in very specific ways, but many people are also using tools to participate in the encyclopaedia project.

There is a very strong interplay of this technology and the users, which I wanted to look at more closely, especially when I found the prominent role of bots in its system. For instance, lots of Wikipedians work with tools that alert them when an article they have worked on has been edited. Put differently, it's not only when you log in that you are connected to this platform. If you're really invested, you can set up your alert systems and your editing tools that help you maintain this project.

NN: So how would you define the role of these bots and their importance in Wikipedia? You gave some examples in the last question but more generally what is their role and how important are they?

SN: We immediately think of robots as creatures, but it's basically pieces of software. It's software robots, mainly in the description that Wikipedia gives itself: 'a piece of software that takes care of flow and repetitive tasks that would almost be impossible for a person to do'. It can be linking to other Wikipedia language versions of the same article, it can be deleting or reporting vandalism, it can be checking punctuation or formatting tasks. So it's a lot of that administrative work. Wikipedia is also very open about this, you can find a lot of pages on Wikipedia about bots and their history, bot policy, how can you

---

[1] A fake chess-playing machine constructed in the late 18th century, with a human chess master hiding inside to operate the machine.

actually create a bot, etc. But, overall, what you see in the hierarchy of Wikipedia users is that bots are quite important.

Bots have more permissions than registered users, like maybe you and I, or unregistered users. People who have just entered Wikipedia, who start working as a new user, have fewer permissions than a software robot. And that is quite striking. They are very important for the maintenance of this platform. And we have done, with Digital Methods, once, an experiment where we took out all the bots edits, and what we saw was a huge mess, with poor formatting and in the case of vandalism the article's content would be deleted and replaced by spammy text. This strongly illustrated how important these bots are for the maintenance of this platform.

And what I think is a bit less known, is that they also create articles. And there, you can think of it more as a feed: it takes a specific source, like the US census, and it 'feeds' articles about, for instance, about towns and cities and villages in the United States into Wikipedia. So it will create an article from that database and it will look like a very basic man-made Wikipedia article. Some bots, like Rambot[2], which is also the oldest bot in Wikipedia, has created thousands of articles. Of course these robots are fully dependent on their data source. It has gone wrong in the past, for instance for Rambot, when there were changes in the database and Rambot created some two thousand faulty articles in a very short time. This was fixed and Rambot still creates articles based on the US census.
So to sum up, not only are bots administrative workers for Wikipedia, they are also very much about weaving its content together and they create new content.

NN: What about updates to existing articles? For instance when a famous person with a Wikipedia article dies, would the bots be able to update the entry?

SN: The content-creating bots pull content from an existing database, and I am not aware of bots connected to databases of deceased celebrities, or whether there even are such databases. In the case of a deceased famous person with a Wikipedia article, there are a lot of people who will rush to edit the entry. The peaks in editing are often around newsworthy events.

NN: How does this bot contribution evolved over time? Did their importance increase over time?

SN: You definitely see an increase in bots over time, but they have actually been there from the start. What is really nice about Wikipedia, is that you have all these openly available statistics, so you can actually see the bot activity per language version, and you can also filter out all the bots to see only the human editor activity.
Looking at 2014, Wikipedia has twenty-two point four percent bot activity. In comparison, in 2009 this was The percentage of bot edits in all Wikipedias combined was 25,8% in February of 2015. Excluding the English-language Wikipedia, total bot activity counts up to over 35%. This shows that bot activity is unevenly distributed across language versions.

---

[2] http://en.wikipedia.org/wiki/User:Rambot

However, there are variations depending on languages: very small languages and very large languages. In a study about this, we noticed that, basically, the large languages like English, the English language version, relies much less on bots than smaller languages. We looked at the types of bots and they are very much about weaving it together within the larger Wikipedia framework. So, trying to find the same articles in different language versions and making sure there are no dead ends in these small language versions, and that's, of course, one of the aims of Wikipedia: to weave this web of articles across languages, of things that share the same topic. Bots clearly have a big role in that. And you could see, also, this phenomenon in an endangered language. We looked also at the really small languages such as Corsican and Scottish Gaelic and small, endangered languages, there are hardly any people engaged in creating articles in those languages.

NN: Apart from languages, are there any factors that could help you differentiate the role of bots in Wikipedia?

SN: For instance, there is a really interesting body of work done by a researcher called Stuart Geiger. He investigates what he calls the "social lives of bots" referring to the whole culture around this bot activity. He documented how people thank and reward these bots with messages on their user page. There are lists of bots on Wikipedia and you can always recognise bots them as well by searching for a user with 'bot' in his name, as it is mandatory to include 'bot' in the username of a software robot. If you go and visit a bot page, you will notice this whole social culture. Users thank bots for various reasons: for "keeping Wikipedia tidy", for "editing my article" and it's this whole social scene around bots creators and the people who appreciate their work.

Another thing that you can look at is bots that have not really been updated or looked at anymore, so they are called "zombie bots" which, essentially, are not so closely connected to the human users who created them anymore. Looking at bots as a specific community of users would be interesting and could include observing their evolution over time. For a specific article, you can zoom in and see, for instance a controversial topic, how many bots are working there against vandalism or other activities, trying to keep the discussion itself out of the article. In any case, there is a lot of work being done there on how to use Wikipedia for controversy analysis, such as the Contropedia project, which is well worth looking at.

NN: Are there any topics, subjects on Wikipedia for which bots are more important. I mean controversial subjects might be one, but are there any other relevant examples?

SN: Controversial topics often lead to bot activity but they can be very diverse. They can be about Michael Jackson or Julian Assange, it can be about events or a specific persons. If you look at the examples I gave before, the articles about all the cities and towns in the USA, I am sure someone would be interested in creating such articles, but it would simply take up too much time. So I think bots can also really help in creating content in projects where people are trying to be exhaustive. Bots also help making sure that Wikipedia remains usable. Generally, if a lot of people are jumping on a topic because of a news event, then bots can help to keep it legible.

NN: How does the Wikimedia Foundation or the local Wikipedia Chapters react to those bots. I guess they created some sort of framework on regulations or how, is this supported or what's their reaction?

SN: Yes, I would say they find it very normal. The Wikipedians, the people who maintain Wikipedia, as mentioned by Jimmy Wales quite often, is a very cohesive group. It's a few hundred people and they know each other and they work together. Bots have been part of Wikipedia from the start.

For Wikipedians, bots are necessary and you need policies to make sure that not everybody can create a bot and that the bot itself is not a spammy one. So you have to first demonstrate that it is harmless and useful and sort of close to the rules. But then, it's just part of the whole ecology that shapes Wikipedia. And I think that that is relatively unknown to people using it on a daily basis.

For now, there is one Wikipedia policy for bots and it's a very specific procedure that you need to follow to be able to have your bot accepted in the first place. As any user, it's hard enough to get anything on Wikipedia, and to have it stick and accepted by other Wikipedians, and the same goes for bots. It is not so easy to have a bot be accepted. I can't imagine specific Wikipedia chapters to stand up and say: "no more bots here!" This would be highly unadvisable because it would create a mess.

NN: What about the outcomes of your research? Can you summarize some of the results about these issues?

SN: The main thing that I found important to conclude from this research is that Wikipedia is fully co-authored by bots and that to understand it, we need to grasp that they are important authors and editors of this platform. What you see then, is that per language this differs very widely, and that also that in controversial topics, you can see that they are necessary, like I said, to keep it legible. I think that future research could maybe look very specific case studies. I'm currently studying climate change scepticism, on different web platforms, looking at Twitter, Google and Wikipedia. And there, you can also look at the specific commitment of Wikipedians –humans or bots-, to a specific issue, and see whether they migrate along with similar articles that are being created. So, for instance, for the issue of climate change, you have a lot of controversy, for instance about whether it is man-made or not, and whether global warming is unprecedented or not. And to deal with controversy, Wikipedians then tend to create new articles. So one can be about the climate change controversy and you try to do that, to take the debate outside of the main article on climate change. And then, it's nice to be able to see whether actors and bots migrate along to this new discussion space, rather than remain in the main article about climate change.

So this kind of Wikipedian-migration across the platform is something that I found very interesting, to see that you can actually study the 'issue commitment' of Wikipedians by looking at Wikipedia behind the scenes. It is a really good platform for issue analysis like that, you can actually see over time and across languages and across linked articles, it offers you this whole view on what's happening to the development of an issue and who are invested in participating, and whether they are human editors or software robots.