FREIE UNIVERSITÄT BERLIN

Bots in Wikipedia: Unfolding their duties

Claudia Müller-Birn

TR-B-19-01
January, 2019

FACHBEREICH MATHEMATIK UND INFORMATIK
SERIE B • INFORMATIK

Abstract

The success of crowdsourcing systems such as Wikipedia relies on people participating in these systems. However, in this research we reveal to what extent human and machine intelligence is combined to carry out semi-automatic workflows of complex tasks. In Wikipedia, bots are used to realize such combination of human-machine intelligence. We provide an extensive overview on various edit types bots carry out in this regard through the analysis of 1,639 approved task requests. We classify existing tasks by an action-object-pair structure and reveal existing differences in their probability of occurrence depending on the investigated work context. In the context of community services, bots mainly create reports, whereas in the area of guidelines or policies bots are mostly responsible for adding templates to pages. Moreover, the analysis of existing bot tasks revealed insights that suggest general reasons, why Wikipedia's editor community uses bots as well as approaches, how they organize machine tasks to provide a sustainable service. We conclude by discussing how these insights can prepare the foundation for further research.

Keywords: Wikipedia, social computing, collective intelligence, distributed cognition, collaboration

This technical report is based on an unpublished article from January 2014.

1 Introduction

Over the past decade, systems have emerged in the Web that people use to solve a wide range of problems. These systems are often defined as crowdsourcing systems. Following [5], any system is a crowdsourcing system if it involves a crowd of humans in a problem-solving process. Examples include Wikipedia, Stackoverflow, or Mechanical Turk-based systems. However, these systems differ in various regards such as by possible types of user's contributions, or by possible combinations of user's contributions to solve the addressed problem.

Some of these system deal with problems that tend to be increasingly complex. For example, the Wikipedia project comprises more than 4 millions articles in the English language version today. Over 30 million registered users around the world collaboratively write this online encyclopedia. Even though, only about 30,000 of registered users are actively involved in the project, meaning they have made at least five edits in a given month, the challenges these people are facing to keep this project running are substantial. Especially because such crowdsourcing system are socio-technical in nature and all coordination and communication work are carried out by software. In the case of Wikipedia, the software is quite simple. The MediaWiki software follows the well-known

wiki idea [21], which can, in many parts, not easily adapted for such a complex project.

Previous academic research on Wikipedia has almost exclusively focused on social structures (e.g., [3], [1], [7] , [24]) which represents existing community practices. Wikipedia success and productivity are often explained by the sophisticated coordination system of human editors (e.g., [17], [19], [20], [18], [6]). Although the role of humans in such a complex problem solving processes is important, we want to reveal to what extent human and machine intelligence are combined to realize complex tasks in such project. Such realization of complex tasks can be understood as a cognitive process, which is distributed over individuals and (technical) artifacts and distributed over time [15]. Distributed cognition emerges in social interactions as well as in interaction of people with artifacts. Artifacts can even become part of the cognitive system itself. For example, a blind person's cane helps the person to perceive the world [15]. The artifact provides a description model on which the person can act on. In crowdsourcing systems, such artifacts correspond to software and its features. From the beginning, the needs of the Wikipedia project and the development path of the MediaWiki software have been highly symbiotic [14]. It means that ideas for many features developed for the MediaWiki software are originated in the community and their needs while curating and maintaining encyclopedic articles. The Wikimedia Foundation carries out the development of the software, but single editors have few possibilities to bring in new, needed functionality. Already in 2002, editors invented an alternative way to support their needs by software. They use bots, software programs that autonomously perform edits following their own pre-defined schedule.

The academic community has disregarded these algorithmic tools in their discussion. [8] differentiates three categories how researchers dealt with bots in the past. First, bots are used as a tool for collecting data (e.g., [2]), second, bots are installed to reach Wikipedia editors in an automatic way to influence their editing behavior (e.g., [4]), or third, researchers consider bots as either noise in the data [23] or not important enough for further analyses [16]).

Only recently, researcher started to emphasize their importance. For example [8] states: "such tools transform the nature of editing and user interaction". Building on prior research, [10] investigated the process of vandal fighting in Wikipedia by using trace ethnography[1]. They show how human editors and bots work together in a decentralized network to fight vandals in Wikipedia. They conjecture that such distribution of concerns to human and algorithmic editors might even change the moral order in Wikipedia. In another study, Geiger shows *"how a weak but pre-existing social norm was controversially reified into a technological actor"* [8]. He refers to the example of the HagermanBot, which has been implemented to sign unsigned discussion entries in Wikipedia. A guideline in Wikipedia recommends that signing a discussion entry: *"is good*

[1] (Geiger and Ribes [10]) define trace ethnography as: *"a fine grained analysis of the various „traces that are automatically recorded by the project's software alongside an ethnographically derived understanding of the tools, techniques, practices, and procedures that generate such traces".*

practice, and facilitates discussion by helping identify the author of a particular comment."[2]. HagermanBot has been the first bot that facilitates the automatic execution of this guideline. [22] extends this line of research by showing that Wikipedia's governance system consists of social mechanisms (e.g., formal and informal rules) as well as algorithmic mechanisms that are defined by software features and bots. [13] shows that bots are not only responsible that existing guidelines are enforced at a larger scale, but also that their activities can have unexpected effects. For example, the number of reverts of newcomers' edits has elevated, while (surprisingly) the quality of those edits stayed almost constant. Editors increasingly apply algorithmic tools for monitoring edits of newcomers. In 2010, already 40 percent of rejections of newcomer's contributions are based on these algorithmic tools[3] [13]. This contradicts attempts of the community to engage more new editors. In a constitutive study [11] show, how a distributed cognitive network of human and algorithmic actors works efficiently together to detect and revert vandalism on Wikipedia. At the same time, efficiency seems to have its price. Wikipedia's quality control system especially relies on one bot - the ClueBot NG. They calculated that the median time-to-revert nearly doubled during ClueBot NG's downtime.

This more recent research suggests that bots are more important to the success of the Wikipedia project than expected previously. Existing research mainly concentrates on Wikipedia's quality system. In our research, we take a more general approach and attempt to provide a global overview about the various tasks that are carried out by the bot community. Over the years, Wikipedia's community has been experimenting with bots, gradually discovering what kinds of things they can and should be allowed to do. We present the current state of the bot community, which reveal the results of this experimentation.

2 The Wikipedia community and their algorithmic editors

First, we provide some contextual information on the existing governance system of bots in English Wikipedia. Second, we describe our data collection process and finally, we present basic statistics on this bot community.

2.1 Emergence of the bot policy and the approval group

In October 2002, the editor Ram-Man programmed the first officially accepted bot on Wikipedia called Rambot. Rambot created 30,000 stub-articles on towns in the US filled with basic statistics, such as current population and geographic

[2] `http://en.wikipedia.org/wiki/Wikipedia:Signatures`.

[3] Besides bots, so-called assisted editing tools exist as well that allow editors to carry out similar, often re-occurring edits more easily [13] . For example, editors often used tools such as Huggle for dealing with malicious edits or WPCleaner for repairing links to disambiguation pages. These tools still allow the editor to review and approve each change.

information from census data. The reactions of Wikipedia's editors varied considerably: many found the created articles useful, but many other complained about the number of successive edits on the *Recent Changes* page, which were caused by this bot. It became very difficult for editors to identify for them interesting edits (e.g., probably malicious edits) on this page. Thus, in December 2002, the *Bot User Group* was established in Wikipedia and edits from registered bots are automatically hidden from the *Recent Changes* page. During the same time, the bot operator Ram-Man initiated the process of creating a bot policy. The policy contained properties bots should exhibit, for example, "be harmless", and provides some examples in which ways a bot can be useful to the Wikipedia project. Most notable is the statement at the beginning of the policy, saying: *"One of Wikipedia's rules to consider: avoid using bots."*.

Three years later, end of 2005, the bot policy page listed almost 100 bots. They are carrying out tasks such as replacing categories, adding links to different language versions of articles, creating disambiguation pages for a given word or phrase. In March 2006, a group of experienced editors formed the *Bot Approval Group* (BAG) to cope with this ongoing growth. The group is amongst other things responsible for approving or denying bots. Members of this group nominate themselves and existing members of the BAG appoint them. In September 2013, the group consists of seven active, and eleven semi-active members. These members *"offer sound bot-related advice to bot operators, admins, and editors alike"*[4]. They are also responsible to block bots if they act outside the agreed reference. Over the years, bots became more accepted members in Wikipedia's community. In August 2006, the "historical anti-bot-ism" once written by Ram-Man disappeared from the policy page.

2.2 Process of data collection

The *Bot User Group* provides the user names for all officially registered bots in Wikipedia, but the data represent only the current number of bots. As one of our goal is to get as exact a picture as possible about the activities of bots in Wikipedia, we decided to determine bots by applying an alternative, albeit more complex approach. As mentioned in the previous section, every (semi-) automatic edit carried out by a bot or tool needs to be approved by a member of the Bot Approval Group. It means, before an operator can run her bot on Wikipedia, she has to consult members of the BAG who ensure that bots act within the framework of the bot policy[5]. Such a so-called task request is well documented and available to all interested community members. The operator has to use a template[6] for her task request (an example is given in Figure1). Such a template forces the requester to provide information such as a short functional description of the bot, the mode of automation (manual vs. automated), and

[4]More information is available at `http://en.wikipedia.org/wiki/Wikipedia:Bot_Approvals_Group`.

[5]For the most recent version look at: `http://en.wikipedia.org/wiki/Wikipedia:Bot_policy`.

[6]`http://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/`

Figure 1: Example of a task request page.

the programming language (e.g., Python). A member of the BAG checks the usefulness of the task and the correct functioning of the bot. A task request is approved, if no concerns regarding the task type exist, the bot runs as expected, and the community has not expressed concerns. Otherwise, a task request is denied.

We parsed all 2,682 tasks requests that are in the final approval stage (i.e., open tasks have not been considered) in Dec 2012. In a pre-processing step, we aligned all template fields (e.g., cleaned different spellings) and subsequently saved all data in a database. Then, we selected all tasks requests that have been approved. These 1,639 approved task requests represent the different types of edits bots carried between Dec 2002 and Dec 2012. Hereinafter we call them tasks. In the next section, we compare the number of algorithmic edits to human edits and provide selective properties of bots.

2.3 Characteristics of bots

Based on the identified 824 bots, we measure the size and activity range of the bot community. In Figure 2, we present the monthly edit volume of bots over the years compared it to the edit volume of anonymous and registered users in Wikipedia. As opposed to human editors, whose total number of edits shows a downward trend since 2007 (as shown by [25]), the number of edits carried out by bots increased over the years and stays on a relatively stable level by the end of 2012. Thus, the proportion of bot edits grows compared to human editors in Wikipedia. Following [8], bots are important members of Wikipedia's collaboration system, and their role and influence should not be underestimated. However, the number of edits should not mistakenly equated with the edit volume. As we show later, the edits carried out by bots are mainly

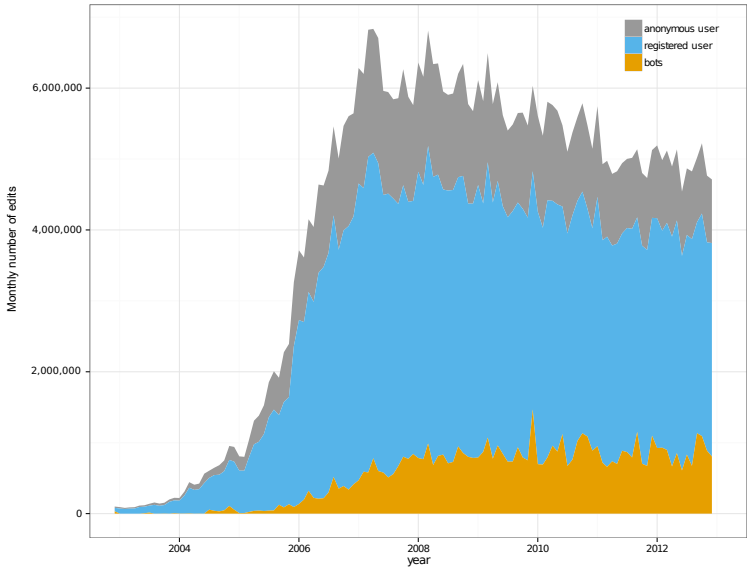

Figure 2: Number of edits per month and user group (yellow (low): bot edits, blue (middle): registered users, grey (top): anonymous users).

supportive in nature. Bots help to keep the administrative and maintenance system of this online encyclopedia project running. Although the sum of the edit count of all bots is relatively growing compared to human editors, the edit count of single bots is quite different. Table 1 give an overview about bot activities by the number of edits, number of namespaces, in which edits took place, number of (distinct) edited pages, and their age (number of days between the first and the last edit in Dec 2012). The number of edits and their number of active days are highly skewed which means that some bots have been inactive despite their status. However, one bot has been an active algorithmic editor during the whole period (Dec 2002 to Dec 2012). The same difference can be observed by comparing the number of edited pages. The bot with the highest edit activity edited almost two million different pages, which means that this bot theoretically made changes to every second article in Wikipedia.[7]

3 Types of algorithmic contributions in Wikipedia

In the next step, we analyzed each single bot request manually, to reveal a more comparable set of typical tasks, bots carry out in Wikipedia. In the following section, we describe this process more deeply. In the first step, we selected only task requests that provided context information regarding their scope. In the second step, we classified the provided functional descriptions of bot tasks by

[7]By August 2013, English Wikipedia has over 4.3 million articles.

	#edits	#namespaces	#edited (distinct) pages	#days between first and last edit
Minimum	1	1	1	0
Maximum	3,495,433	18	1,964,216	3,638
Mean	43,397	5	25,991	507
Minimum	872	3	376	202
Minimum	215,027	4	120,166	663

Table 1: Descriptive statistics of identified 824 bots.

action-object-pairs. Based on the resulting descriptions, we show typical bot activities by considering the two most often provided contexts.

3.1 Classification of bot tasks

During the coding process of all 1,639 tasks, we identified 575 tasks that provided context information in the form of links to pages, for instance, in the maintenance area, to policies and guidelines, or discussions and projects. We checked all links and assigned them to eleven contexts. In Figure 2, we show the defined contexts and the number of tasks that refer to them. We expected that bots are especially active in the maintenance area (i.e., Community Service).

Moreover, we categorized all functional descriptions of tasks. If a task consisted of several functions, we selected the function with the highest impact (i.e. most invasive edit). We categorized all 575 tasks by 67 action-object-pairs, which occurred during the following three-round process. During the first round, we collaboratively coded about 50 tasks until we obtained an almost stable set of actions and objects. For example, we merged the tasks "archive discussion pages" and "archive pages" to the action "move", because archiving pages in Wikipedia means moving them to sub-pages, and the object "page". In the second round, we separately coded the remaining tasks. In the third round, we checked all defined action-object-pairs and discussed ambiguous cases. Newly introduced action-object-pairs were cross-validated over all tasks. In the next two sections, we utilize the two most often provided contexts and the action-object-pairs to disclose various forms of algorithmic contributions to the encyclopedic project. In the first section, we look at bots that carry out tasks in the context of Community Service and Guideline or Policy.[8]

[8]When mentioning a bot, we provide the number of the according task request in parentheses. For example, AvicBot (11) relates to a bot with the name AvicBot and its task request 11. If only the bot name is used, then the bot conducts only one task.

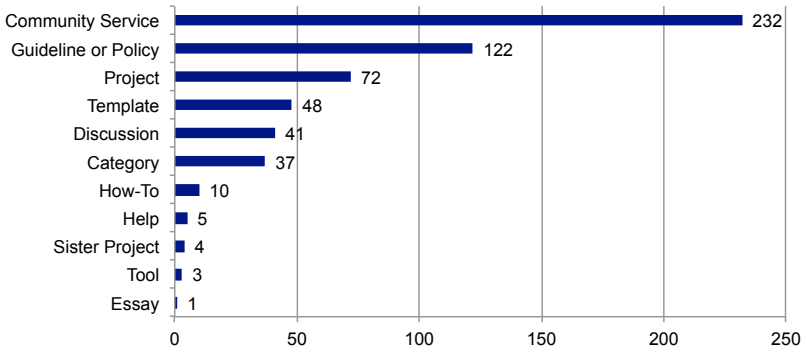

Figure 3: Defined contexts and the number of tasks that refer to them.

3.2 Bots as community servants and rule enforcers

In this section, we show how bots support Wikipedia's editors in their maintenance duties but also by ensuring the correct application of existing guidelines or policies. Table 2 contains the number of tasks that bots carry out in various areas in the context of *Community Service* and of *Guideline or Policy*. For the former, the listed pages differentiate into *Backlogs* (e.g., "Categories for Discussion") and *Noticeboards* (e.g., "Administrator intervention against vandalism"). A Wikipedia backlog is often simply a list of things that need to be done. Noticeboards relate to specific policies and guidelines in Wikipedia and editors can ask questions and request assistance from other editors who are more familiar with these norms. They contain often a backlog as well. The *Guideline or Policy* context comprises all policies such as deletion policies (e.g., "Criteria for speedy deletion"), legal policies (e.g., "Non-free content criteria"), and procedural policy (e.g., "IP block exemption") as well as guidelines such as behavioral (e.g., "Canvassing"), edit guidelines (e.g., "Redirect") content guideline (e.g. "Manual of Style") as well as norms and practices that are not a policy or guideline itself, but help to supplement or clarify existing guidelines and policies. The most often referenced guideline is *Manual of Style* and the most often linked policy is the *Criteria for speedy deletion*, which relates to the aforementioned backlogs.

This analysis shows that bots are especially active in the maintenance area of the Wikipedia project. Over the years, editors have built an extensive system of manual and semi-automatic workflows to deal with the growth and the complexity of the project. These workflows are partly supported by bots. The question regarding the application of bots is, whether the whole maintenance system became more efficient and editors can more concentrate on content creation or whether the increasing usage of bots is caused by a diminishing number of editors in this area. Moreover, the frequent usage of guidelines or policies suggests that bots also important for the governance system of the project. Bots are responsible that existing guidelines and policies are enforced at a larger scale, and the question is how this might effects the community. The results

	Type	# Tasks	# Bots
Community Service			
Categories for Discussion (CfD)	Backlog	30	26
Templates for Discussion (TfD)	Backlog	17	15
Articles for Deletion (AfD)	Backlog	17	16
Featured Articles (FA)	Backlog	8	7
Admin. Intervention against Vandalism (AIV)	Noticeboard	7	5
Changing Username (RENAME)	Noticeboard	7	6
Guideline or Policy			
Manual of Style (MoS)	Guideline	25	17
Criteria for Speedy Deletion (SPEEDY)	Policy	20	20
Redirect (R)	Guideline	6	5
Disambiguation (D)	Guideline	4	2
Image Use Policy (IUP)	Policy	4	4
Non-free Content Criteria (FUC)	Policy	2	2

Table 2: Most frequently provided tasks in Community Services and Guideline or Policy.

present a static picture, thus, a temporal analysis of existing relations might be interesting. In Figures 4 and 5 existing action-object-pairs are visualized for both contexts. This allows for understanding what kind of tasks bots carry out in as community service (Figure 4) or guideline or policies (Figure 5). The larger the area of a circle, the more tasks belong to this action/object pair. The prevalent tasks in the context of *Community Service* are "create reports", "notify users", and "add, remove or replace categories or templates". However, the overall concentration of tasks is quite different to the context *Guideline or Policy*. As opposed to bots that support community services, bots in this context manipulate categories and notify users to a much lower extent. At the same time, files, links and syntax is more often the target of their actions. Reports and templates seem to have an equal importance. We assume, bots are responsible for different types of tasks as in the former area.

Our results suggest that bots carry out specialized tasks that are adjusted to the respective context. The type of tasks needed in both contexts seems to be different. Further analyses might even reveal that certain action-object-pairs occur only in one or the other context. We leave this to future research. In the next section, we introduce selected bots, to present

- Reasons, why Wikipedia's editor community uses bots and,
- Approaches, how bots help to keep this collaborative project running.

We start each paragraph in the section with a brief description of the context, the underlying workflow, and the identified bots and conclude by a major insight we derived from the bot activities.

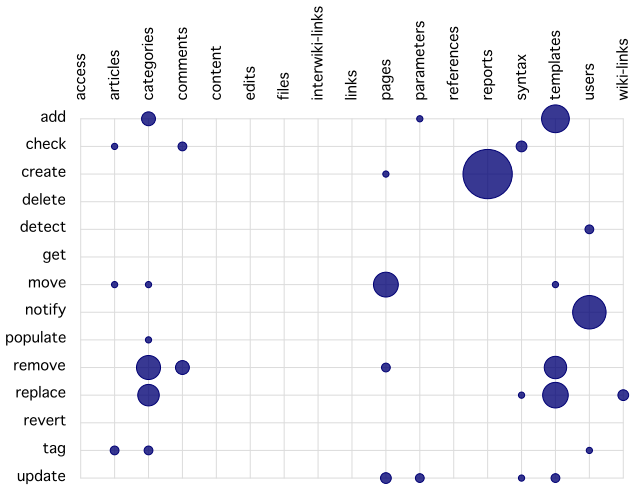

Figure 4: Number of tasks per action-object-pair in the context of *Community Service*. The largest circle represents 63 tasks and the smallest circles 1 task. Actions are vertically and objects are horizontally aligned.

3.3 Cause and effect of bot usage

The highest number of tasks is performed in the area of *Categories for Discussion* (CfD). In this backlog, editors discuss a proposed deletion, merging, and renaming of categories[9]. This workflow is triggered by adding a specific template at the top of the category page and finishes by another template that documents the community decision. Based on this decision, a category is deleted, merged or renamed. The category system in Wikipedia is very complex and more than 1,000 pages can be affected by such a decision. This can cause long lists of categories that need to be deleted, removed, or renamed. For a human editor, this work is very tedious and time-consuming; therefore, this task is often delegated to a bot.

Theoretically, a single bot is sufficient to carry out this task. The practice is different. The functioning of a bot depends on imponderabilities, such as technical (e.g., software is often hosted on local servers owned by their operators, which can breakdown) or social (e.g., a mal-functioning of the bot needs to be fixed by the operator, or a bot discontinue operations). That can lead to an unreliable bot service, which can cause the breakdown of the defined workflow.

[9]Categories are pages that belong to the Category namespace. They are used to arrange pages that belong to a similar topic. Editors carry out this classification by adding [[Category:XYZ]] to any page in Wikipedia. The page is then automatically listed on the category with name XYZ.

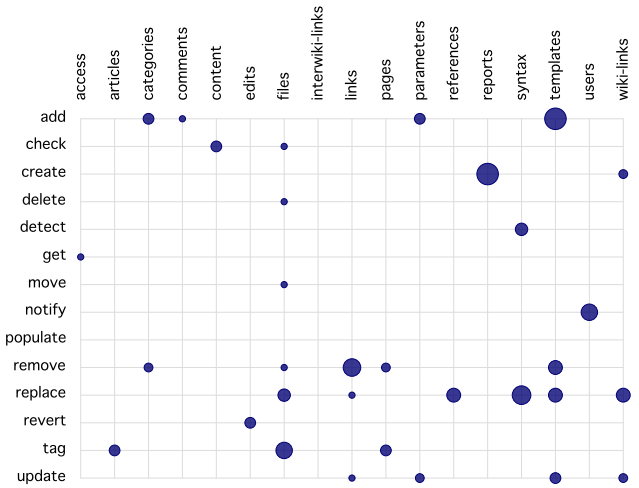

Figure 5: Number of tasks per action-object-pair in the context of *Guideline or Policy* (bottom). The largest circle represents 63 tasks and the smallest circles 1 task. Actions are vertically and objects are horizontally aligned.

This issue is addressed by entrusting various bots with the same task.

The CfD provides one example of such shared responsibility. For example, the AvicBot (11) moves pages from a category A to a category B, if merge request exist on CfD, or removes a category entirely if the requests is to empty the category. Bots such as the ArmbrustBot (3), CrismonBot, and Cydebot (4) provide similar functionality. Often, operators copy or re-implement needed functionality. Programming frameworks such as the Pywikibot[10] provide extensive functionality that can be easily replicated. We summarize:

Bots duplicate similar tasks concurrently or successively to ensure a reliable service.

On the backlog Articles for Deletion (AfD), editors propose and discuss articles that should be deleted because they meet criteria such as they violate copyright or provide obvious misinformation. Often, the editor who proposes an article for deletion is distinct from the editor that created that article. The editor who is proposing an article for deletion adds a specific template to the article and starts a discussion on the backlog.

The editor who has created the article does not necessarily notice this process. An editor can only track changes to created articles if she adds articles to her personal watch list. This list contains all recent changes that are made to those pages (and their associated talk pages). Even though an article is listed

[10]For more information please see `http://www.mediawiki.org/wiki/Manual:Pywikibot`

on the watch list, such minor change can easily be overlooked. Over time, the community has developed a number of bots to allow editors to receive a notification (entry on talk page) when their article is proposed for deletion. Examples are the CSDWarnBot, the Android_Mouse_Bot (2), the Erwin85Bot (8). All these bots provide a so-called opt-in mechanism that means that only editors are notified if they explicitly request this service. The need for such mechanism occurred during community discussions relating the HagermanBot (cf. complete case study in [9]. We conjecture:

Bots are created because of missing software features.

Featured articles (FA) in Wikipedia are the best articles created by the community. In this context, bots provide specific statistical reports to editors. For example, Jmax-bot counted daily the number of existing featured articles because another editor requested such list. The UcuchaBot (5) generates a list of FA that have not been shown on the main page yet. On the main page, every day another featured article is presented. The presented article needs to be proposed by the community and a double presentation should be avoided. The UcuchaBot (2) adds a notice to a discussion on FA nominations, if one of the participants is also involved in the WikiCup project[11]. This is a yearly championship on Wikipedia that can be won by excellent editing activities. All these bots have in common that they generating aggregated or related data to human editors. This data can be used for further decision-making. For example, by providing a list of featured articles that have not being showed on the main page, editors can more easily select the next article. We derive:

Bots provide statistics or lists for further decision-making that can more easily carried out by editors.

Administrator intervention against vandalism is a noticeboard on Wikipedia, which is used by editors and bots to report obvious spam or vandalism. For example, Mr.Z-bot directly notifies other editors on this board about users that are most likely spammers or vandals by continuously monitor the abuse log[12]. The abuse log is a list of edits that are selected based on a number of predefined filters. The HBC AIV helperbot, which is being operated with different duties by different editors ensure for instance that the listed vandals and spammers are not already blocked. If so, they are removed from the respective list. These two bots are jointly working on this noticeboard to help human editors to carry out their duties more efficiently. Geiger shows, that bots have become one of the predominant mechanisms for identifying vandals (Geiger, 2009). In subsequent research, [10] illustrate that blocking vandals in Wikipedia can even be interpreted as a distributed, cognitive process that is possible by a "complex network of interactions between humans, encyclopedia articles, software systems, and databases". We suggest:

Bots coordinate their tasks with humans to provide more sophisticated services.

Manual of Style (MoS) is the most often mentioned guideline. This guideline

[11] For more information see `http://en.wikipedia.org/wiki/Wikipedia:WikiCup`

[12] For more information see `http://en.wikipedia.org/wiki/Special:AbuseLog`.

documents Wikipedia's norms for language, layout, and formatting. The goal is twofold: (1) the used language should be clear and concise, (2) the style and formatting within articles should be consistent. The whole MoS is a very complex system of rules and sub-rules concerning topics such as quotations, punctuation, dates, images, and links. 17 bot operators explicitly related to MoS in their tasks. For example, DyceBot (4) replaced hyphens in page titles to en-dashes as stated in the MoS. WildBot (3) checks the wiki markup of new articles for links to dabbling pages because this is one of the most common errors in linking of new pages. MetsBot (5) formats dates according to MoS-norms. We hypotheses two reasons are responsible for the creation of these bots: first, norms have been developed over time in Wikipedia. As shown by [3] the number of norms has enormously grown over time. The application of bots has been necessary, to apply newly defined or adapted norms to existing articles. Second, as already seen, the editing, proofreading and layout creation of articles in Wikipedia is only supported by basic functionality of the underlying MediaWiki software. Simple spell or expression checking mechanisms has implemented in professional text or code editors do not exist. Editors built therefore mechanisms to support their work and to keep up with the expanding norms in Wikipedia. We suggest:

Bots provide services because community norms change over time and these changes need to be reflected in the project.

The most referenced policy is the Criteria for speedy deletion which relates to the aforementioned backlogs. This policy defines cases, in which discussions about the deletion issue is omitted and pages or files can be immediately deleted. The policy lists a number of properties a page or file has to exhibit so that speedy deletion can be applied. For example, if files lack licensing information, files are not free and orphaned (no links to other pages), or blocked users create articles, these files or pages can be deleted immediately. Most of the bots that are actively enforcing this policy only collect candidates for deletion by tagging files or articles. They support administrators in their duties. In the context of this policy the usage of the Orphaned_image_deletion_bot has been discussed (in September 2009). The operator sought besides the usual bot flag also administrative rights for her bot. For deleting files or pages on Wikipedia administrator rights are necessary. Over the years, the editor community has collected a set of tasks in the bot policy (cf. Section 2.1) that prohibit bots to carry out such as the (semi-) automated creation of articles at large scale, the automated assignment of categories to person articles, and conducting only cosmetic mostly syntactical changes to articles. Also awarding administrative rights to bots belonged to this set of forbidden tasks for some time, and it is still controversial. A long discussion appeared on the task request page of the Orphaned_image_deletion_bot, but finally, the request has been approved. Today, bots with administrator rights can be approved in the regular process, if and only if the wider community approves them. Having more experiences with bots has changed community norms about bots over time. Attitudes have changed from initial resistance to having bots at all, to handing over more and more duties to them. We summarize:

Bots extend their task coverage because community's perception on what is acceptable changes.

The Image use policy revealed another approach, how bots help to keep the project running. The policy defines requirements on format, content, and copyright issues of images. OrphanBot is responsible for the latter issue. The bots removes images from articles that are tagged with a special template for example because they have an unknown source, or an unknown copyright status. Then an administrator can delete the respective image from Wikipedia. Originally, the OrphanBot was responsible for checking uploaded images as well. But the operator decided to move this task to another bot – the ImageTaggingBot. The ImageTaggingBot checks newly uploaded images and tags images if they have for example no description or copyright information. The bot additionally notifies the uploader of any problems it encounters. This approach of splitting up single tasks to bots is an example for the specialization strategy of bots. Similarly, there are operators that aggregate diverse tasks in one bot. For example, SmackBot has 33 tasks. We convey:

Bots follow different specialization strategies for employing tasks.

During the course of analyzing bots in their referred context, we revealed possible reasons, why Wikipedia's editor community started using bots, and which approaches have been developed by the bot community to provide a maintainable and sustainable service. Albeit, this part of our research is based on anecdotes, we are convinced that these anecdotes yield to new, still unanswered research questions.

4 Conclusion

In this research, we provided a global overview about the various tasks that bot carry out in Wikipedia. We shed light on the main characteristics of bots and revealed reasons for their application as well as their modes of working. This research is not conclusive. It helps to identify possible areas in which more research is appropriate. Such research can improve our understanding of interdependencies between human and algorithmic activities in Wikipedia. Existing research often focuses on the social infrastructure of the project. The question is how this social infrastructure shapes the technical infrastructure and vice versa. Each part of our article provides another perspective on the bot community. In the first part, we emphasize the importance of the bot community by comparing their edits counts to human editors. In this area, we need a more fine-granular analysis to show what kinds of edits bots carry out. In the second part of this article, we derive these edit types by existing task descriptions. In our discussion, we disregarded the edit volume behind each task, so it is quite challenging to define its importance. An automatic approach to classify existing bot edits might be more suitable. In the third part, we share patterns that emerged during the coding process of the tasks and differentiated two main types. Each of the suggested explanations needs further investigation and opens up for further research. In our discussion, we did not

address the probably existing cultural changes that might be caused by the usage of bots. [12] describe, how counterproductive the usage of assisted editing tools is for newcomers. They show a correlation between the application of these tools and the low retention rate of new editors in Wikipedia. Maybe similar effects can be revealed for bot usage as well. Moreover, community norms regarding bots have changed substantially over time. Exposure to bots, as well as the need to cope with the growing complexity of the Wikipedia technology has led to greater acceptance of some types of bots, and increasingly tight integration with the MediaWiki infrastructure. Future research should integrate the social and the technical dimension of Wikipedia's editor community. This can provide insights that can even be applied to other crowdsourcing systems.

References

[1] Ivan Beschastnikh, Travis Kriplean, and David W McDonald. Wikipedian Self-Governance in Action: Motivating the Policy Lens. In *International Conference on Weblogs and Social Media*, 2008.

[2] Moira Burke and Robert Kraut. Taking up the mop: identifying future wikipedia administrators. In *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, pages 3441–3446, New York, NY, USA, 2008. ACM.

[3] Brian Butler, Elisabeth Joyce, and Jacqueline Pike. *Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia.* the nature and roles of policies and rules in wikipedia. ACM, New York, New York, USA, April 2008.

[4] Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. SuggestBot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th international conference on Intelligent user interfaces*, pages 32–41, New York, NY, USA, 2007. ACM.

[5] Anhai Doan, Raghu Ramakrishnan, and Alon Y Halevy. Crowdsourcing systems on the World-Wide Web. *Commun. ACM*, 54:86–96, April 2011.

[6] Andrea Forte, Niki Kittur, Vanessa Larco, Haiyi Zhu, Amy Bruckman, and Robert E Kraut. Coordination and Beyond: Social Functions of Groups in Open Content Production. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, pages 1277–1286, New York, NY, USA, 2012. ACM.

[7] Andrea Forte, Vanesa Larco, and Amy Bruckman. Decentralization in Wikipedia Governance. *J. Manage. Inf. Syst.*, 26(1):49–72, July 2009.

[8] R Stuart Geiger. The social roles of bots and assisted editing programs. *Proceedings of the 5th International Symposium on Wikis and Open Collaboration, WikiSym '09*, pages 30:1–30:2, 2009.

[9] R Stuart Geiger. The lives of bots. In *CRITICAL POINT OF VIEW A Wikipedia Reader*, pages 78–93, Amsterdam, 2011.

[10] R Stuart Geiger and David Ribes. *The work of sustaining order in wikipedia: the banning of a vandal.* the banning of a vandal. ACM, New York, New York, USA, February 2010.

[11] Aaron Halfaker, R Stuart Geiger, Jonathan T Morgan, and John Riedl. The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline. *American Behavioral Scientist*, 57(5):664–688, 2013.

[12] Aaron Halfaker, Aniket Kittur, and John Riedl. Don't bite the newbies: how reverts affect the quantity and quality of Wikipedia work. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*, pages 163–172, New York, New York, USA, 2011. ACM.

[13] Aaron Halfaker and John Riedl. Bots and Cyborgs: Wikipedia's Immune System. *Computer*, 45(3):79–82, 2012.

[14] Sumana Harihareswara and Guillaume Paumier. MediaWiki. In *The Architecture of Open Source Applications*. lulu.com, 2012.

[15] James Hollan, Edwin Hutchins, and David Kirsh. Distributed Cognition: Toward a New Foundation for Human-computer Interaction Research. *ACM Trans. Comput.-Hum. Interact.*, 7(2):174–196, June 2000.

[16] A Kittur, E Chi, B A Pendleton, B Suh, and T Mytkowicz. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World Wide Web*, 1(2):1–9, 2007.

[17] Aniket Kittur and Robert E Kraut. Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 37–46, New York, NY, USA, 2008. ACM.

[18] Aniket Kittur and Robert E Kraut. *Beyond Wikipedia: coordination and conflict in online production groups.* coordination and conflict in online production groups. ACM, New York, New York, USA, February 2010.

[19] Aniket Kittur, Bryant Lee, and Robert E Kraut. Coordination in collective intelligence: the role of team structure and task interdependence. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 1495–1504, New York, NY, USA, 2009. ACM.

[20] Aniket Kittur, Bryan Pendleton, and Robert E Kraut. Herding the cats: the influence of groups in coordinating peer production. In *Proceedings of the 5th international Symposium on Wikis and Open Collaboration*, pages 7:1–7:9. ACM, 2009.

[21] Bo Leuf and Ward Cunningham. *The Wiki Way - Quick Collaboration on the Web.* Addison-Wesley, Boston, MA, USA, 2001.

[22] Claudia Müller-Birn, Leonhard Dobusch, and James D Herbsleb. *Work-to-rule: the emergence of algorithmic governance in Wikipedia.* ACM, New York, New York, USA, June 2013.

[23] Felipe Ortega and Jesus M Gonzalez Barahona. Quantitative analysis of the wikipedia community of users. In *WikiSym '07: Proceedings of the 2007 international symposium on Wikis*, pages 75–86, New York, NY, USA, 2007. ACM.

[24] Andreas Schroeder and Christian Wagner. Governance of open content creation: A conceptualization and analysis of control and guiding mechanisms in the open content domain. *J. Am. Soc. Inf. Sci. Technol.*, 63(10):1947–1959, October 2012.

[25] Bongwon Suh, Gregorio Convertino, Ed H Chi, and Peter Pirolli. The singularity is not near: slowing growth of wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, pages 8:1–8:10, New York, NY, USA, 2009. ACM.