



**UvA-DARE (Digital Academic Repository)**

**Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system**

Niederer, S.M.C.; van Dijck, J.F.T.M.

*Published in:*  
New Media & Society

*DOI:*  
[10.1177/1461444810365297](https://doi.org/10.1177/1461444810365297)

[Link to publication](#)

*Citation for published version (APA):*

Niederer, S., & van Dijck, J. (2010). Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society*, 12(8), 1368-1387. DOI: 10.1177/1461444810365297

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <http://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*Full Title:* Wisdom of the Crowd or Technicity of Content? Wikipedia as a socio-technical system

*Authors:* Sabine Niederer and José van Dijck

Sabine Niederer, University of Amsterdam, Turfdragsterpad 9, 1012 XT Amsterdam, The Netherlands [s.m.c.niederer@uva.nl](mailto:s.m.c.niederer@uva.nl)

José van Dijck, University of Amsterdam, Spuistraat 210, 1012 VT Amsterdam, The Netherlands [j.f.t.m.vanDijck@uva.nl](mailto:j.f.t.m.vanDijck@uva.nl)

*Authors' Biographies*

Sabine Niederer is PhD candidate in Media Studies at the University of Amsterdam, and member of the Digital Methods Initiative, Amsterdam. She is also Managing Director of the Institute of Network Cultures, the new media research center at the Amsterdam University of Applied Sciences.

José van Dijck is a Professor of Media and Culture at the University of Amsterdam where she is currently Dean of Humanities. She has published widely in the area of media and science, (digital) media technologies, and television and culture. Her latest book is titled *Mediated Memories in the Digital Age* (Stanford University Press, 2007).

*Abstract*

Wikipedia is often considered as an example of ‘collaborative knowledge’. Researchers have contested the value of Wikipedia content on various accounts: some have disputed the ability of anonymous amateurs to produce quality information, while others have contested Wikipedia’s claim to accuracy and neutrality. Even if these concerns about Wikipedia as an encyclopedic genre are relevant, they misguidedly focus on human agents only. Wikipedia’s advance is not only enabled by its human resources, but is equally defined by the technological tools and managerial dynamics that structure and maintain its content. This article analyzes the socio-technical system—the intricate collaboration between human users and automated content-agents— that defines Wikipedia as a knowledge instrument.

*Keywords:* Wikipedia; collaborative knowledge; Web 2.0; socio-technical system; protocol.

*Word count:* 8092 Words, including keywords and abstract.

## Wisdom of the Crowd or Technicity of Content? Wikipedia as a socio-technical system

### *Introduction*

User-generated content has revived the idea of the Web as a place of human collaboration and a place of activity by ‘everybody’ (Shirky, 2008). The Wikipedia project is often considered as an example par excellence of ‘collaborative knowledge’, of ‘social media’ or the ‘wisdom of crowds’. Since 2001, a group of editors and volunteers have engaged in developing an online encyclopedia whereby everyone is invited to contribute and articles are open to continuous editing. Large numbers of contributors, the so-called ‘Wikipedians’, produce an online encyclopedia that is unprecedented in scale and scope. Researchers who have evaluated or contested the value of Wikipedia content, have almost unanimously focused on its human contributors. For instance, it has been disputed whether an encyclopedia that is produced by many (anonymous) minds results in quality information (Keen, 2008). Other critics have contested Wikipedia’s claim to accuracy and neutrality by pointing at the liability of allowing anonymous contributors whose interest or expertise remains undisclosed.

Even if these concerns about Wikipedia as an encyclopedic genre are legitimate and relevant, we argue that they misguidedly focus on human agents only while neglecting the role of technology. In this article, we focus on the intricate interrelation between human and technological tools which is at the heart of several debates concerning Wikipedia. The first debate revolved around the question whether Wikipedia is authored primarily by a few elite users or by many common contributors—an opposition we would like to question, for Wikipedia has a refined hierarchical structure in which contributing administrators, registered users, anonymous users and bots all have a distinct rank in an orderly system. Secondly, we would like to refocus the public debate on the quality of Wikipedia’s encyclopedic information—disputing whether its entries are

accurate and neutral—by shifting attention to the protocols and technologies deployed to facilitate consensus editing. A basic comprehension of Wikipedia’s automated editing systems as well as emerging tracking tools like the WikiScanner are needed to evaluate the encyclopedia’s ability to meet standards of neutrality and accuracy, while preventing overt bias and vandalism.

In the third section of this paper, we want to show how dependent various user groups and entries are on non-human content agents, or bots, that assist in editing Wikipedia content. Examining the variable dependency of human editors on bots for editing encyclopedic content per language Wikipedia, we explore how an automated system of *interwiki* and *interlanguage* bots help maintain overall content strategies of the online encyclopedia. Linking and networking specific language Wikipedias into one global system is less the result of people working together across languages and borders, as it is the product of collaboration between humans and bots.

From our analysis we conclude that any evaluation of Wikipedia’s qualities should acknowledge the significance of the encyclopedia’s dynamic nature as well as the power of its partially automated content-management system. It is the intricate collaboration between large numbers of human users and sophisticated automated systems that defines Wikipedia’s ultimate success as a knowledge instrument. In order to unravel this intricate human-technological interaction, we deploy several ‘natively’ digital methods for Web research (Rogers, 2009).<sup>1</sup> In his publication ‘The End of the Virtual’, Rogers calls for ‘research *with* the Internet’ that applies novel medium-specific methods and tools. Rather than importing existing analytical methods onto the Web, natively digital methods can ‘move beyond the study of online culture alone’ and help understand new media as the interplay between human and technological agents. The concrete goal of this analysis is to provide a better understanding of how Wikipedia’s automated technological systems and the management of large numbers of edits are inextricably

intertwined. More philosophically, we want to theorize human and machine contributions as complementary parts of a socio-technical system, which is at the heart of many Web 2.0 platforms.

*Many minds collaborating*

Wikipedia has been described in terms of ‘many minds’ (Sunstein, 2006) and similar notions such as ‘the wisdom of crowds’ (Surowiecki, 2004; Kittur and Kraut, 2008), ‘distributed collaboration’ (Shirky, 2008), ‘mass collaboration’ (Tapscott and Williams, 2006), ‘produsage’ (Bruns, 2008), ‘crowdsourcing’ (Howe, 2006), ‘Open Source Intelligence’ (Stalder and Hirsh, 2002), and ‘collaborative knowledge’ (Poe, 2006). The collectively written encyclopedia on a wiki platform is often heralded as an example of collaborative knowledge production at its best. In early 2008, an article in the *New York Review of Books* explained the compelling charm of Wikipedia:

So there was this exhilarating sense of mission—of proving the greatness of the Internet through an unheard-of collaboration. Very smart people dropped other pursuits and spent days and weeks and sometimes years of their lives doing ‘stub dumps,’ writing ancillary software, categorizing and linking topics, making and remaking and smoothing out articles—without getting any recognition except for the occasional congratulatory barnstar on their user page and the satisfaction of secret fame. Wikipedia flourished partly because it was a shrine to altruism—a place for shy, learned people to deposit their trawls (Baker, 2008).

Since the start of the Wikipedia project in 2001, the dedication of its contributors as well as the group effort as an alternative to the professional expert approach have been a source of both excitement and criticism. Even if Wikipedia now has become famous for its collaborative character of many minds producing knowledge, it is interesting to

remind ourselves that the project originally intended to be an expert-generated encyclopedia. Started by the name of ‘Nupedia’, a small team of selected academics was invited to write the entries, with the aim of creating a ‘free online encyclopedia of high quality’ (Shirky, 2008: 109). The articles would be made available with an open content license. Founder Jimmy ‘Jimbo’ Wales and his employee Larry Sanger put into place a protocol based on academic peer-review (Shirky, 2008; Poe, 2006). This expert approach failed, partly because of the slowness of the editing process by invited scholars. To speed up the process, Sanger suggested a wiki as a collective place where scholars and interested laypeople from all over the globe could help with publishing and editing draft articles. The success of Wikipedia and the commitment of the Wikipedians took them by surprise. Sanger became the ‘chief organizer’, a wiki-friendly alternative for the job of ‘editor in chief’ that he held for Nupedia. He made a great effort to keep Wikipedia organized, while at the same time providing space for some of the ‘messiness’ (edit wars, inaccuracies, mistakes, fights, etc.) that collaborative editing brings along. Early 2002, however, Sanger turned away from the epistemic free-for-all of Wikipedia, towards an expert-written encyclopedic model called Citizendium (Welcome to Citizendium, n.d.), while Wales chose to further pursue the Wiki-model.

The question whether online encyclopedias and similar enterprises *should be* produced by few (expert) or many (amateur) minds has been the source of heated debate ever since the Sanger-Wales split-up. Internet critic Andrew Keen (Keen 2008: 186) applauded Sanger for coming to his senses about the debased value of amateur contributions in favor of expert-professionals. On the other end of the specter, many Wikipedia adepts have praised its democratizing potential as well as its ethos of community and collaboration, a source of knowledge free for everyone to read and write (Benkler, 2006; Jenkins, 2006). By the same token, the notion that Wikipedia is actually produced by ‘crowds’ has been regularly challenged, most notably by Wikipedia’s

founders. During the first five years of its existence, Wikipedia was largely dependent on the work of a small group of dedicated volunteers. Although they soon formed a thriving community, the notion of a massive collective of contributors was repeatedly downplayed by Wales. As he pointed out in a talk at Stanford University in 2006:

The idea that a lot of people have of Wikipedia, is that it's some emergent phenomenon -- the wisdom of mobs, swarm intelligence, that sort of thing -- thousands and thousands of individual users each adding a little bit of content and out of this emerges a coherent body of work. (...) [But Wikipedia is in fact written by] a community, a dedicated group of a few hundred volunteers. (...) I expected to find something like an 80-20 rule: 80% of the work being done by 20% of the users (...) But it's actually much, much tighter than that: it turns out over 50% of all the edits are done by just [0].7% of the users (Wales in Swartz, 2006).

As Wales asserts, until 2006 Wikipedia was largely written and maintained by a small core of dedicated editors (2% doing 73.4% of all the edits). The disproportionate contribution of (self-)designated developers versus 'common users' can also be found in research into the open source movement. Rishab Aiyer Ghosh and Vipul Ved Prakash were among the first to disaggregate the 'many minds' collaborating in the open software movement; their conclusion was that 'free software development is less a bazaar of several developers involved in several projects and more a collation of projects developed single-mindedly by a large number of authors' (Ghosh & Prakash, 2000: 1). In the open source movement, very few people were actually collaborating in developing software.

It would be a mistake, however, to dismiss the idea of Wikipedia's 'many contributors' as a myth. Starting in 2006, the online encyclopedia shows a distinct decline of 'elite' users while at the same time the number of edits made by novice users and 'masses' is steadily increasing. Various researchers have pointed at the dramatic shift in

workload to the common user (Kittur, Pendleton and Mytkowicz, 2008); instead of pitching the power of the expert versus the wisdom of the crowds, Kittur et al. speak of ‘the rise of the bourgeoisie’, a marked growth in the population of low-edit users between 2006 and 2008. Interestingly, these researchers explain this shift by describing Wikipedia in terms of a dynamic social system that evolves as a result of the gradual development, implementation and distribution of content management systems. After an initial period of being managed by a small group of high-powered, dedicated volunteers, the ‘pioneers were dwarfed by the influx of settlers’ (Kittur et al., 2008: 8). The early adopters select and refine technology and managerial systems, followed by a majority of novice users who begin to be the primary users of the system. Kittur and his colleagues observe a similar decline of elite users in Web 2.0 platforms and suggest that it may be a common phenomenon in the evolution of online collaborative knowledge systems. This tentative conclusion is underscored by other researchers who show that, in order to sustain the encyclopedia’s growing popularity, its organizers need to identify more productive workers and grant them ‘administrator’s status’ (Burke & Kraut, 2008).

Although these researchers correctly observe significant changes in the ‘wisdom of crowds’ paradigm, they seem to be stuck in the antagonism of (few) experts versus (many) common users; even if they notice the growing presence of non-human actors, such as software tools and managerial protocols, in the evolution of Wikipedia’s social dynamics, they tend to underestimate their importance. In fact, the increasing openness of Wikipedia to inexperienced users is *made possible* by a sophisticated techno-managerial system, which facilitates collaboration on various levels. Without the implementation of this strict hierarchical content management system, Wikipedia would most likely have become a chaotic experiment.

According to Alexander Galloway, the Internet and many of its (open source) applications are not simply open or closed, but *modulated*. Networked technology and



management style are moderated by protocol, which gains its authority ‘from technology itself and how people program it’ (Galloway, 2004: 121). Wikipedia, built as an open system and carried out by large numbers of contributors, appears to be a ‘warm, friendly’ technological space, but only becomes warm and friendly through what Galloway refers to as ‘technical standardization, agreement, organized implementation, broad adoption and directed participation’ (2004: 142).

This is exactly what happened during the first five years of Wikipedia, in which Wikipedia administrators developed strict protocols for distributing permission levels, imposing a hierarchical order in deciding what entries to include or exclude, what edits to allow or block. If we look more closely at Wikipedia’s organizational hierarchy (figure 1), we can distinguish various user groups, some of which are ‘global’ (in the sense that they edit across various language Wikipedias) and others are specific to a certain ‘local’ Wikipedia.

<i>Permission Level</i>	<i>Wikipedia Users</i>
Most Permissions	Developer / System Administrator
	Steward
	Checkuser
	Oversight
	Bureaucrat
	Administrator / Sysop
	Bot
	Registered User
	Newly Registered User
	Anonymous User
No Permissions	Blocked User

**Figure 1:** Schematic overview of global and local categories of Wikipedia users according to permission levels. URL (consulted January 2009):

[http://meta.wikimedia.org/wiki/User\\_groups](http://meta.wikimedia.org/wiki/User_groups)

Each user group maintains the same pecking order, regulating the distribution of permission levels: ‘blocked users’ have the least permissions, for they can only edit their own talk page. Anonymous users have fewer permissions than registered users, who in turn are at a lower level of permission than bots; bots are just below administrators (‘admins’), who occupy the highest level in the elaborate Wikipedia-bureaucracy. System administrators (or developers) have the most permissions, including server access. This is a small user group of only ten people who ‘manage and maintain the Wikimedia Foundation Servers’ (System Administrators, n.d.). Remarkable in this ranking system is the position of bots (short for ‘software robots’), whose permission level is just below that of administrators but well above the authority of registered users. We will return to the status of bots in the third section; for now, it is important to note the significant role of automated mechanisms in the control of content.

In fact, we could argue that the very success of the Wikipedia project lies in the regulation of collaborative production at *any level*, from a small edit or a single upload, to a more extensive contribution or even development of the platform or its content. Like any large public system, Wikipedia works through a system of disciplinary control by issuing rewards, such as granting a dedicated user the authority level of administrator (Burke & Kraut, 2008) and by blocking contributor’s rights of those users who deviate from the rules. A disciplinary system of power distribution in the digital age, however, can’t be regarded exclusively as a system of *social* control. As Gilles Deleuze (1990) has pointed out in his acute revision of Foucault’s disciplinary institutions, a ‘society of control’ deploys technology as an intricate part of its social mechanisms. Wikipedia’s content management system with distinct levels of permissions allows moreover for *protocological control*: a mode of control that is at once social and technological—one cannot exist without the other (Galloway, 2004: 17). Along the same lines, Bruno Latour

(1991:129) proposes to analyze technological objects and infrastructures as ‘socio-technical ensembles’, in which the strict division between ‘material infrastructure’ and ‘social superstructure’ is to be dissolved:

Rather than asking ‘is this social’ or ‘is this technical or scientific’ (...) we simply ask: has a human replaced a non-human? Has a non-human replaced a human? (...). Power is not a property of any of those elements [of humans or non-humans] but of a chain (1991: 110).

Similar to Latour’s search to dissolve the ‘technology/society divide’, we argue that the dynamic interwovenness of human and non-human content agents is an underrated yet crucial aspect of Wikipedia’s performance. The online encyclopedia’s success is based on socio-technical protocological control, a combination of its technical infrastructure and the collective ‘wisdom’ of its contributors. Rather than assessing Wikipedia’s epistemology exclusively in terms of ‘power of the few’ versus ‘wisdom of crowds’ we propose to define Wikipedia as a gradually evolving socio-technical system that carefully orchestrates all kinds of human and non-human contributors by implementing managerial hierarchies, protocols, and automated editing systems.

#### *Accurate and neutral encyclopedic information*

A similar disregard of technological aspects can be noticed in another heated debate that has haunted the online project from its inception: the question regarding the quality of Wikipedia’s encyclopedic information. Wikipedia entries have often been held against the standards of *accuracy* and *objectivity*, as set by reputed encyclopedias such as the Encyclopedia Britannica. Wikipedia entries are based on three core principles, which serve as leading rules for its contributors and aim at holding up the encyclopedia’s quality standards.

The first core rule is ‘Verifiability’, i.e. readers have to be able to retrieve Wikipedia content in reliable sources. Therefore, referring to published articles and verifiable resources is necessary to have the article (or edits) accepted (Verifiability, n.d.). A second, related, core rule, is called ‘No Original Research.’ Wikipedia simply does not accept ‘new’ (unpublished) research or thought (No Original Research, n.d.). Again, reliability on Wikipedia means citing proven, published sources. Thirdly, articles have to be written from a ‘Neutral Point of View’ (NPOV) to avoid bias, meaning the articles have to be based on facts and facts about opinions, but not on opinions (Neutral Point of View, n.d.). All contributors, whether single anonymous users or administrators, are required to comply with these rules.<sup>2</sup> Compliance is regulated by the above mentioned core rules, and noncompliance is punished by removal of edits. In past debates on Wikipedia’s standards of accuracy and neutrality, the emphasis has been on whether they can be kept up by crowds of human users. The more profound question in line with our research thesis is, though, how these standards are maintained and controlled through the organization and mechanics of Wikipedia’s content management system.

Initially, the quality debate concentrated mainly on accuracy, or, more precisely, on the lack thereof due to the impossibility of verifying and authenticating sources. With so many anonymous and amateur contributors, the likeliness of vandalism, inaccuracy, and downright sloppiness in factual details was more than real. As danah boyd (2005) observes, Wikipedia ‘lacks the necessary research and precision’ and ‘students are often not media-savvy enough to recognize when to trust Wikipedia and when this is a dreadful idea’. Other researchers entered the quality-of-content debate by testing Wikipedia’s robustness in terms of content vandalism. Alexander Halavais (2004) intentionally contributed incorrect information to existing articles. For his ‘Isuzu Experiment’, he inserted thirteen mistakes into thirteen different articles, expecting that most of the errors would remain intact. Much to Halavais’ surprise, his wrongful edits

did not last long, but were all corrected within a couple of hours.<sup>3</sup>

However, the explanation for the speed at which his vandalism was detected lies less in human acuity than in technological perspicacity. The fact that Halavais had made all changes from the same username and IP-address arguably made it all too easy for Wikipedians and their tools to undo his edits. Making thirteen changes in thirteen different articles in a short timeframe obviously attracts attention from automated bots, and even human Wikipedians, after spotting one mistake, would have probably looked into his other edits in the other articles, and could have easily retrieved the other mistakes ‘by association.’ Philosopher of science P.D. Magnus therefore provided a corrective to Halavais’ research method with his 2008 study in which he inserted inaccuracies distributed across IP-addresses and fields of expertise. He found that one third of the errors were corrected within 48 hours, and most others were ‘corrected by association’, as was the case with Halavais’ experiment whereby the Wikipedians probably started checking his other edits after initially finding three mistakes. Some researchers conclude from these tests that the online encyclopedia is robust to vandalism due to its huge numbers of watchful community members (Poe, 2006); instead, we argue that it is rather the strict implementation of protocological control and the use of automated bots that account for Wikipedia’s vigilance. With their experiments, Halavais and Magnus have less proven the reliability of Wikipedians than the reliability of the encyclopedia’s techno-managerial system.

The notion of a community of vigilant users has continued to feed the accuracy debate, particularly by academic researchers who questioned the reliability of Wikipedia’s supposed egalitarian approach. Collaboratively written amateur content, as these critics contend, finds itself at odds with knowledge production. If not written by known experts, how accurate and reliable are these encyclopedia’s entries? In December 2005, the first academic research that systematically compared the accuracy of Wikipedia and

Encyclopedia Britannica was published in *Nature* (Giles, 2005). Researchers compared the two encyclopedias by checking 42 science articles in both publications. The reviewers were academics, who checked the articles without knowing its source. They found Wikipedia and Britannica to be almost equally accurate; not surprisingly, the news was triumphed as ‘Wikipedia Survives Research Test’ (BBC News, 2005). With this outcome, Wikipedia was recognized as an encyclopedia, at least on the level of its accuracy. But this was not the symbolic end of the reliability discussion. On the contrary, the debate heated up and more research followed. In 2006, information systems researcher Thomas Chesney (2006) conducted more empirical research into the credibility of Wikipedia, asking a total of 258 experts (academics) and non-experts to fill out a survey about a Wikipedia article from their area of expertise (or, for the laymen, in their realm of interest). The respondents found mistakes in 13% of the Wikipedia articles. But Chesney also found that the experts gave the Wikipedia articles a higher credibility rating than the non-experts did. Contrary to what Sanger described as the ‘perceived inaccuracy of Wikipedia’, the respondents expected (and found) Wikipedia to be a reliable source of information on the Web.

In response to this accuracy debate, centering on the assumed polarity between (known) experts and (unknown) laypersons, few academics proposed to redirect its focus from *product* to *process*, and from the abilities of *people* to the qualities of its technological *tools*. Historian Roy Rosenzweig (2006), who conducted a thorough analysis of Wikipedia biographical entries and compared them to entries from the American National Biography Online (written by known scholars), concludes that the value of Wikipedia should not be sought in the accuracy of its published content at one moment in time, but in the dynamics of its continuous editing process—an intricate process where amateurs *and* experts collaborate in an extremely disciplined manner to improve entries each time they are being edited. Rosenzweig notices the benefits of many edits to the factuality of

an entry. As he points out, it is not so much crowds of anonymous users that make Wikipedia a reliable resource, but a regulated system of consensus-editing that bares *how* history is written: ‘Although Wikipedia as a product is problematic as a sole source of information, the process of creating Wikipedia fosters an appreciation of the very skills that historians try to teach’ (2008: 138). One of the most important features, in this respect, is the website’s built-in history page for each article, that lets you check the edit-history of an entry. According to Rosenzweig, the history of an article as well as personal watch lists and recent changes pages are important instruments that give users additional clues to determine the quality of individual Wikipedia entries.

Part of the discussion disputing the accuracy and neutrality of Wikipedia’s content concentrated on the inherent unreliability of *anonymous* sources. How can an entry be neutral and objective if the encyclopedia accepts copy-edits from anonymous contributors who might have a vested interest in its outcome? Critics like Keen (2008) and Denning et al. (2005) have objected to the distribution of editing rights to all users. What remains unsaid in this debate is that the impact of anonymous contributors is clearly restricted due to technological and protocological control mechanisms. For one thing, every erroneous anonymous edit is systematically overruled by anyone who has a (similar or) higher level of permission (which is anyone except for blocked users). Since anonymous users are very low in the Wikipedia pecking order, their edit longevity is likely to be short when they break the rules of objectivity and neutrality.

On top of that, there is an increasing availability of ‘counter-tools’ that allow for checking the identity of contributors, or at least their location of origin. On the History page of each Wikipedia entry, we can find the timestamp and IP-address for every anonymous edit made. The WikiScanner, a tool created by California Institute of Technology student Virgil Griffith in 2007, makes it possible to geo-locate anonymous edits by looking up the IP addresses in a IP-to-Geo database, listing the IP addresses and

the companies and institutions they belong to. It facilitates the tracking of anonymous users by revealing who and where they actually are. The WikiScanner has proven to be a powerful tool for journalists trying to localize and expose biased content. In the WikiScanner FAQ on his website, Griffith states he created the WikiScanner (among other reasons) to ‘create a fireworks display of public relations disasters in which everyone brings their own fireworks, and enjoys.’ The WikiScanner was designed to reveal bias, and Griffith collects the most spectacular results on his website.<sup>4</sup>

The debates concerning Wikipedia’s accuracy and neutrality have been dominated by fallacious oppositions of human actors (experts versus amateurs, registered versus anonymous users) and have favored a static evaluation of its content (correct or incorrect at one particular moment in time). Both qualifications, however, are ill-suited when applied to a dynamic online encyclopedia such as Wikipedia, mostly because a debate grounded in such parameters fails to acknowledge the crucial impact of a non-human actor: Wikipedia’s dynamic content management system and the protocols by which it is run. Arguably, Wikipedia is neither the often-advertised platform for many minds, nor is it a space for anonymous knowledge production. The WikiScanner has made the revealing of anonymous users much easier by matching IP-addresses with contact information.<sup>5</sup> Bias can now be identified, tracked and if necessary, reverted. But there is more to the technicity of Wikipedia content than fast users armed with notification feeds and monitoring devices. The technicity of Wikipedia content, as we will show in the next section, lies in the totality of tools and software robots used for creating, editing, and linking entries, combating vandalism, banning users, scraping and feeding content and cleaning articles. It is the complex collaboration not of crowds, but of human *and* non-human agents combined that defines the quality standards of Wikipedia content.

*Co-authored by bots*



The significant presence of bots appears counter to the common assumption that Wikipedia is authored by human ‘crowds’. In fact, human editors would never be able to keep up the online encyclopedia if they weren’t assisted by a large number of software robots. Bots are pieces of software or scripts that are designed to ‘make automated edits without the necessity of human decision-making’ (Wikipedia Bot Policy, n.d.). They can be recognized by a username that contains the word ‘bot’, such as SieBot or TxiKiBoT. Bots are created by Wikipedians, and once approved, they obtain their own user page; they form their own user group with a certain level of access and administrative rights, made visible by flags on a user account page. One year after Wikipedia was founded, bots were introduced as useful helpers for repetitive administrative tasks (History of Wikipedia Bots, n.d.). Since the first bot was created on Wikipedia, the number of bots has grown exponentially. In 2002, there was only one active bot on Wikipedia; in 2006, the number had grown to 151, and in 2008 there were 457 active bots (Editing Frequency of All Bots, n.d.).

In general, there are two types of bots: editing (or ‘co-authoring’) bots and non-editing (or administrative) bots. Each of the bots has a very specific approach to Wikipedia content, related to its often narrow task. Administrative bots are most well known and well liked among Wikipedia users: they are deployed to perform policing tasks, such as blocking spam and detecting vandalism. Vandalism combat bots come into action when ‘vandalism-like’ edits are made. Vandalism is recognizable, for it often means a large amount of deleted content in an article, or a ‘more than usual’ change in content. Spellchecking bots check language and make corrections in Wikipedia articles. Ban enforcements bots can block a user from Wikipedia, and thus take away his or her editing rights, which is something a registered user is not able to do. Non-editing bots are also data miners, used to extract information from Wikipedia, and copyright violation identifiers; the latter compare text in new Wikipedia entries to what is already available

on the Web about that specific topic, and report this to a page for human editors to review. Most bots, being created to perform repetitive tasks, make many edits. In 2004, the first bots reached the record number of 100,000 edits.

The second category of editing or co-authoring bots seems to be much less known by Wikipedia users and researchers (for otherwise it would certainly have played a role in the debates about reliability and accuracy). While not every bot is an author, all bots can be classified as ‘content agents’, as they all actively engage with Wikipedia content. The most active Wikipedians are in fact bots; a closer look at various user groups reveal that bots create a large number of revisions with high quality (List of Wikipedians by Number of Edits, n.d.). Adler et al. (2008) discovered that the two highest contributors in their edit-longevity-survival-test were bots. As mentioned before, bots as a user group have more rights than registered users and also a very specific set of permissions. For instance, bot edits are by default invisible in recent changes logs and watch lists. Research cited above has already pointed out that Wikipedians rely on these notification systems and feeds for the upkeep of articles.

Describing Wikipedians in bipolar categories of humans and non-humans, however, doesn’t do justice to what is in fact a third category: that of the many active users assisted by administrative and monitoring tools, also referred to as software-assisted human editors. Bots are Wikipedians’ co-authors of many entries. One of the first editing bots to be deployed by Wikipedians was rambot, a piece of software created by Derek Ramsey (Ram-Man User Page, n.d.). Rambot pulls content from public databases and feeds it into Wikipedia, creating or editing articles on specific content, either one by one or as a batch. Since its inception in 2002, rambot has created approximately 30,000 articles on U.S. cities and counties on Wikipedia, using data from the CIA World Factbook and the U.S. census. And since content produced by authoring bots relies heavily on their source, errors in the data set caused rambot to publish around

2,000 corrupted articles. In the course of time, bot-generated articles on American cities and counties were corrected and complemented by human editors, following a strict format protocol: history, geography, demographics, etc. The articles appear strikingly tidy and informative and remarkably uniform. If we compare, for instance, an article on La Grange, Illinois, as created by rambot in 2002 with a more recent version of this article from 2009, it clearly shows the outcomes of a collaborative editing process; the entry has been enriched with facts, figures and images (figure 2). The basic format, however, has remained the same. To date, it still is rambot's main task to create and edit articles about US counties and cities, while human editors check and complement the facts provided by this software robot.

**La Grange, Illinois**  
From Wikipedia, the free encyclopedia

This is an old revision of this page, as edited by RamBot (Talk) (contrib) at 17:38, 18 October 2002. It may differ significantly from the current revision.

La Grange is a village located in Cook County, Illinois. As of the 2000 census, the village had a total population of 15,608.

### Geography

According to the United States Census Bureau, the village has a total area of 6.5 km<sup>2</sup> (2.5 mi<sup>2</sup>), 6.5 km<sup>2</sup> (2.5 mi<sup>2</sup>) of it is land and 0.0 km<sup>2</sup> (0.0 mi<sup>2</sup>) of it is water. The total area is 0.00% water.

### Demographics

As of the census of 2000, there are 15,608 people, 5,624 households, and 4,049 families residing in the village. The population density is 2,400.9/km<sup>2</sup> (6,220.7/mi<sup>2</sup>). There are 5,781 housing units at an average density of 888.3 persons/km<sup>2</sup> (2,304.1 persons/mi<sup>2</sup>). The racial makeup of the village is 91.02% White, 6.02% African American, 0.26% Native American, 1.00% Asian, 0.02% Pacific Islander, 0.99% from other races, and 3.86% from two or more races. 3.66% of the population are Hispanic or Latino of any race.

There are 5,624 households out of which 27.9% have children under the age of 18 living with them, 60.3% are married couples living together, 9.2% have a woman whose husband does not live with her, and 26.0% are non-families. 24.5% of all households are made up of individuals and 9.7% have someone living alone who is 65 years of age or older. The average household size is 2.67 and the average family size is 3.23.

In the village the population is spread out with 28.5% under the age of 18, 4.8% from 18 to 24, 28.4% from 25 to 44, 23.8% from 45 to 64, and 13.9% who are 65 years of age or older. The median age is 38 years. For every 100 females there are 94.2 males. For every 100 females age 18 and over, there are 87.4 males.

The median income for a household in the village is \$80,342, and the median income for a family is \$95,554. Males have a median income of \$62,030 versus \$41,260 for females. The per capita income for the village is \$34,867. 4.0% of the population and 3.2% of families are below the poverty line. Out of the total people living in poverty, 4.3% are under the age of 18 and 4.4% are 65 or older.

**La Grange, Illinois**  
From Wikipedia, the free encyclopedia

Coordinates: 41°48′00″N 87°52′44″W﻿ / ﻿41.80000°N 87.87889°W﻿ / 41.80000; -87.87889

**La Grange**, a suburb of Chicago, is a village in Cook County, in the U.S. state of Illinois. The population was 15,608 at the 2000 census. The name La Grange is French for "the barn."

Country	United States
State	Illinois
County	Cook
Coordinates	<span><span><span><span><span>41°48′00″N</span> <span>87°52′44″W</span></span></span><span><span>﻿</span> / <span>﻿</span></span><span><span>41.80000°N 87.87889°W</span><span><span>﻿</span> / <span>41.80000; -87.87889</span></span></span></span></span>
Area	2.5 sq mi (6.5 km <sup>2</sup> )
Population	15,608 (2000)
Density	6,220.7 /sq mi (2,400.9/km <sup>2</sup> )
Timezone	CST (UTC-6)
- summer (DST)	CST (UTC-5)
Postal code	60525
Area code	708

### History

The area around La Grange was first settled in the 1830s, when Chicago residents, already fed up with the rapid population increase in that city in the decade since its incorporation, moved out to the west. The first settler, Robert Litch, came to what is now La Grange in 1830, a full seven years before the city of Chicago was incorporated in March 1837. La Grange's location, at approximately thirteen miles from Chicago's Loop, is not considered far at all from the city by today's standards, but in that time the residents enjoyed the peace of rural life without much communication with urban residents.

Incorporated on June 11, 1879, the Village of La Grange was the dream of Franklin Deight Cossitt, born in Granby, Connecticut and raised in Tennessee, who moved to Chicago in 1862 and built a successful wholesale grocery business.

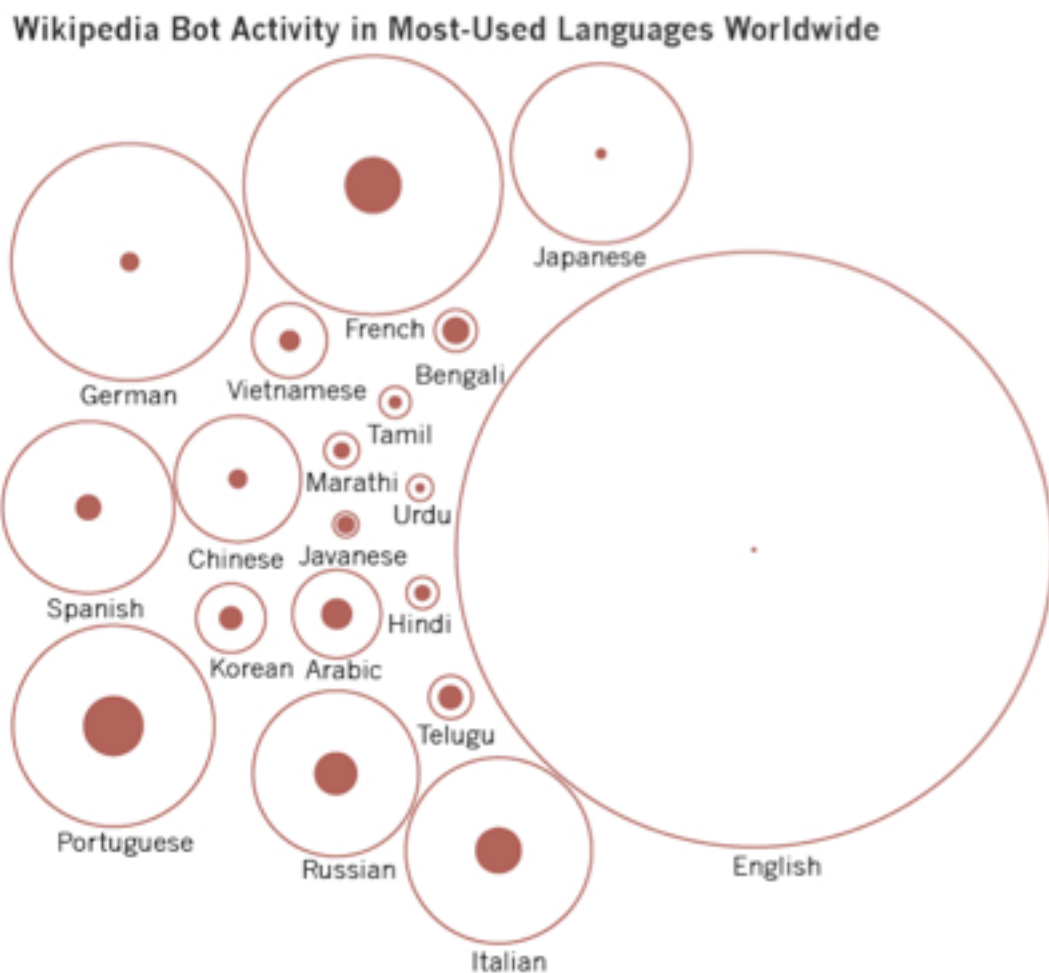
In 1870, Cossitt purchased several hundred acres of farmland in Lyons Township, along the Chicago-Oxton Road, known today as Oxton Avenue (U.S. Highway 34). Oxton Avenue, on the site of a defunct Native American trail, was also referred to as the "Old

**Figure 2:** Bot-created article compared to a human-edited article. The upper screenshot is the La Grange, Illinois article as created by rambot on December 11, 2002. The lower screenshot shows the same article on January 12, 2009. Source: Wikipedia, URL: [http://en.wikipedia.org/wiki/La\\_Grange,\\_Illinois](http://en.wikipedia.org/wiki/La_Grange,_Illinois)

But how dependent is Wikipedia on the use of bots for the creation and editing of content? What is the relative balance of human versus non-human contributions in the online encyclopedia? Peculiarly, the answer to this simple question turns out to be layered and nuanced. When we started to look for answers, we found there to be striking differences between various language Wikipedias. As a global project, Wikipedia features over ten million articles in over 250 languages—two million in English—serving a large number of language communities.<sup>6</sup> The fact that Wikipedia distinguishes between local and global user groups already suggests that bot activity might differ across local Wikipedias, which turns out to be the case.<sup>7</sup> Specific language Wikipedias not only greatly vary in size and number of articles, but also in bot activity. The Wikipedia Bot Activity Matrix, Wikipedia's own meta-data record, offers an overview of total bot activity as well as bot activity per language (Bot Activity Matrix, n.d.). The percentage of bot edits in all Wikipedias combined is 21,5%. Excluding the English-language Wikipedia, total bot activity counts up to 39%, which means that bot activity is unevenly deployed across languages and communities.

In order to account for the differences in bot activity versus human activity, it is interesting to compare bot activity in the most-used language Wikipedias (English, Japanese, German) to bot activity in endangered and revived language Wikipedias (e.g. Cornish, Oriya, Ladino). In the Dorling maps (figures 3 and 4), the thin-lined outer circle depicts the language Wikipedia, sized according to the amount of articles in total. The inner dot represents the share of bot activity in that language Wikipedia. The English,

Japanese, and German Wikipedias show that by far most of its editing is done by human editors. The German Wikipedia, for instance, has only 9% bot activity, the English version even less. Wikipedias of small and endangered languages show a high dependency on bots and relatively small percentage of human edits. Oriya, for instance, depends for 89% on automated software programs; one small Wikipedia, in the language ‘Bishnupriya Manipuri’, has seen 97% of its edits made by bots (Wikipedia Statistics Bishnupriya Manipuri, n.d.).



**Figure 3:** Visualization of Wikipedia bot activity in most-used languages worldwide, overview and detail. The outer circle is the Wikipedia size in that specific language; the inner dot depicts the percentage of bot activity in that language Wikipedia. Analysis by Rogers, Niederer et al. Graphic by Auke Touwslager using the Dorling Map Tool, Digital

Methods Initiative, Amsterdam, 2009. URL:

<http://wiki.digitalmethods.net/Dmi/NetworkedContent>



**Figure 4:** Bot activity in endangered language Wikipedias. Analysis by Rogers, Niederer et al. Graphic by Auke Touwslager using the Dorling Map Tool, Digital Methods Initiative, Amsterdam, 2009. URL:

<http://wiki.digitalmethods.net/Dmi/NetworkedContent>

Further analysis of bot activity versus human activity reveals that variety of bot dependency can be an indicator of the state of the language Wikipedia—if not the state of that language—in the global constellation. Looking at the types of bots, we may notice that Wikipedias are maintained mainly by bots that network the content, so-called interwiki and interlanguage bots. Phrased differently, the bots active in these spaces take care of linking articles to articles in Wikipedias, to prevent them from becoming ‘orphans’ or dead ends. Wikipedia policy states that articles should be networked and be part of the Wikipedia Web. This core principle is summarized as: ‘Link articles sideways to neighbors, upwards to categories and contexts, and downwards to sub-articles to create a useful web of information’ (Build the Web, n.d.). Not only are ‘good’ Wikipedia

articles full of links to reliable sources, they should also link to related Wikipedia articles and sub-articles, and *be linked to*. Articles that only refer to each other, but are not linked to or linking to other articles, are also considered a threat to the principle of building the web.

We can analyze a language's state of interconnectedness using the Wikipedia statistics pages, featuring lists of the most active bots per language Wikipedia. They reveal that most-used language Wikipedias, which obviously contain much more content than the smaller language Wikipedias, have bot activity distributed across administrative tasks. In German, for instance, the top 45 of most active bots features 27 interwiki bots and 18 bots that are meant to edit content, add categories and fix broken links.<sup>8</sup> In the smaller language Wikipedias, bots significantly outnumber human editors and are mostly dedicated to linking articles to related articles in other Wikipedias; they make sure the content, however scarce, is networked. The Cornish Wikipedia's top 45 of most active bots, for instance, shows at least 35 interwiki bots, and the remainders are bots with unspecified functions. These interwiki bots, such as 'Silvonenbot', a bot that adds interlanguage links, make connections between various language Wikipedias; smaller language Wikipedias thus make sure that every article is properly linked sideways, and prevent the language Wikipedia from becoming isolated.

Tracing the collaboration between human and non-human agents in Wikipedias thus allows for an interesting and unexpected insight into the culturally and linguistically diverse makeup of this global project. Following the 'wisdom of crowds' paradigm, we might have been tempted to look for cultural-linguistic diversity in terms of many people across the world collaborating in different languages and from a number of cultural backgrounds. In line with this paradigm, British information scientists have demonstrated that the Internet—and Wikipedia in particular—is anything but a culturally neutral space; major aspects of collaborative online work are influenced by pre-existing

cultural differences between human contributors (Pfeil, Zaphiris & Ang, 2006). Adding a ‘natively digital’ analysis of the varied distribution of bot dependency across the wide range of language Wikipedias, we show that cultural differences in collaborative authoring of Wikipedia content cannot just be accounted for in terms of its *human* users; they reveal themselves perhaps more candidly in the relative shares of human and non-humans contributions as revealed through automated patterns of contributions. High levels of bot activity, mainly dedicated to networking content and to building the web, are an indicator of small or endangered languages; a richer variety of bot activity, largely subservient to human edit activity, could be considered an indicator of a large and lively language space.

### *Conclusion*

Wikipedia has most commonly been evaluated—either praised or detracted—for its collaborative knowledge production by many (anonymous) minds. From the start, researchers have pointed out that Wikipedia thrived by virtue of a small core of dedicated contributors rather than a large crowd of collaborators, even if the encyclopedia became more hospitable to common users after 2006. Wikipedia has never been the mythical egalitarian space either, for the various user groups have very distinct levels of permissions. Past Wikipedia research has focused mainly on the crowdsourcing of knowledge as well as on the reliability of Wikipedia content. Some researchers, such as Halavais and Magnus, tested Wikipedia by entering false information. These types of research have isolated Wikipedia content as a static product mainly by assessing it to other encyclopedic records.<sup>9</sup> Other research, mostly (investigative) journalism, was intent on ‘outing’ anonymous human editors, making use of ‘counter-technology’ like the WikiScanner to reveal the identity of contributors or the origin of an edit. More recently, we have seen different research approaches to Wikipedia, such as Rosenzweig’s who



acknowledges the encyclopedia's dynamic nature of content as well as the significance of its partially automated content-management system. For the most part, though, scholarly evaluations of Wikipedia have adhered to the human content-agent paradigm.

In this article, we have argued that Wikipedia's nature and quality should be evaluated in terms of collaborative qualities not only of its human users but specifically of its human *and* non-human actors. Since 2002, Wikipedia content has been maintained by both tool-assisted human editors and bots, and collaboration has been modulated by protocols and strict managerial hierarchies. Bots are systematically deployed to detect and revert vandalism, monitor certain articles, and, if necessary, ban users, but they also play a substantial role in the creation and maintenance of entries. As we have shown, bot activity may be analyzed as an indicator of the international or intercultural dimension of Wikipedia as a global project.

In the fall of 2009, Wikipedia has introduced WikiTrust, a MediaWiki extension developed by the WikiLab at the University of California in Santa Cruz. With WikiTrust, newly edited parts of Wikipedia articles are color-coded according to reliability based on the author's reputation, which is established by the lifespan of their other contributions. Instead of turning to the expert to check all articles, Wikipedia builds further on the combination of rules, hierarchies and editors. To understand Wikipedia's collaborative process, we need to unravel not simply Wikipedia's human agents, but the specificities of its technicity.

We propose to extend this kind of analysis from Wikipedia to various kinds of Web 2.0 infrastructures. Non-human actors and coded protocols are often overlooked in the many optimistic Web 2.0 theories that triumphantly claim the virtues of mass collaboration on Web 2.0 platforms (Tapscott & Williams, 2005). It is important to question the assumptions of the internet as a merely social laboratory of human interaction, instead analyzing and interrogating the socio-technical system which is at the

core of Web 2.0 platforms. Human and machine contributions are complementary parts of a society of control in which social interactions are increasingly facilitated by means of coded, automated processes. Human judgments such as reliability, accuracy or factuality are turned into machine-coded and regulated alert systems, as illustrated by Wikipedia's mechanisms for generating and checking content. Nicolas Carr compares Web 2.0 to a Mechanical Turk, which 'turns people's actions and judgments into functions in a software program' (Carr 2008: 218). A thorough and critical understanding of the automated processes that structure human judgments and decisions requires analytical skills and medium-specific methods which are crucial to a full understanding of how the Internet works. Instead of succumbing to the mechanisms of control, users should learn to critically analyze their interactions with technology and actively engage in technology's development (Zittrain, 2008: 245).

In line with David Beer's call in this journal for a more thorough understanding of the 'technological unconsciousness' of participatory Web cultures, we have tried to explain the 'performative infrastructure' of Wikipedia by focusing on the second level of analysis Beer proposes: unraveling software infrastructures and their applications (Beer, 2009: 998). In this article, we have deployed several natively digital methods to unravel in detail the close interdependency of human and technological agents. It is important to comprehend the powerful information technologies that shape our everyday life, and the coded mechanisms behind our informational practices and cultural experiences. The analysis of the Wikipedia platform as a socio-technical system is a first step in that direction.

---

<sup>1</sup> Digital Methods is a term for medium-specific methods for Web research, coined by Richard Rogers (2009). The research for this article was conducted with the Digital Methods Initiative and the Govcom.org Foundation, Amsterdam, URL (consulted nov. 09): <http://wiki.digitalmethods.net/>

<sup>2</sup> On top of the set of rules, there is a fourth important general principle emphasizing the ideals of openness and collaboration that lie at the core of the project: Ignore all rules. This general principle was written up by Adam Sanger to make clear that above all, Wikipedia is an open platform. Wikipedians should first and foremost strive to improve and maintain Wikipedia. (Ignore all Rules, n.d.)

<sup>3</sup> Halavais' approach was heavily criticized, mainly because he deliberately littered his object of study. After the event, Halavais regretted his approach, especially because the media attention for his experiment had encouraged others to test Wikipedia by inserting mistakes. His website now includes a call for testing Wikipedia in a 'non-destructive way.'

<sup>4</sup> In the summer of 2008, Virgil Griffith launched the WikiWatcher suite, a set of tools designed for monitoring and maintaining Wikipedia. The suite includes a tool that makes it possible to de-anonymize users with a username whose IP addresses match that of other user(name)s or companies/institutions in a IP-to-Geo database. This stretches the notion of anonymity from the unregistered to the registered with a username. See also <http://wikiwatcher.virgil.gr/> (consulted Jan. 2009).

<sup>5</sup> WikiScanner 2 works the other way around: enter a company name or URL and the tool shows you which Wikipedia articles were edited from that organization's IP-address.

<sup>6</sup> An overview of the available local Wikipedias is given on the Wikipedia portal page, [www.wikipedia.org](http://www.wikipedia.org). Wikipedia currently has 264 language versions, some of which only have a main page and no articles as of yet. The largest Wikipedia is in English, with more than two million articles; it is followed by the German, French, Polish and Japanese editions, each of which contain more than half a million articles. Seventeen other language editions contain 100,000+ articles, and more than 100 other languages contain 1,000+ articles; the overview also includes the smaller ones with only 100+ articles and even Wikipedias that have only a main page (Our Projects, n.d.).

<sup>7</sup> In a case study on Wikipedia as 'networked content' that Sabine Niederer conducted with Richard Rogers et al. (during the 10 Year Jubilee Workshop of the Govcom.org Foundation, 11-15 August 2008), the researchers noticed the great discrepancy between

---

bot activity in English and the other language versions of Wikipedia. URL (consulted Jan. 2009): <http://wiki.digitalmethods.net/Dmi/NetworkedContent>.

<sup>8</sup> The top 45 most active bots on the German Wikipedia consist of 27 interwiki bots, and 18 bots with various tasks, such as editing, fixing links and adding categories (Wikipedia Statistics German, n.d.).

<sup>9</sup> To this date, communication scholars like Halavais and Lackaff (2008) examine Wikipedia's reliability and completeness, assessing qualities of its users rather than those of its systems.

## References

- Adler, T., L. De Alfaro, I. Pye and W. Raman (2008) 'Measuring Author Contributions to Wikipedia', in *Proceedings of WikiSym 2008*, Porto, September 8-10. New York: ACM. URL (consulted April 2009): <http://users.soe.ucsc.edu/~luca/papers/08/wikisym08-users.pdf>
- Baker, N. (2008) 'The Charms of Wikipedia', *New York Review of Books* 55(4), URL (consulted Jan. 2009): <http://www.nybooks.com/articles/21131>
- BBC News (2005) 'Wikipedia Survives Research Test', *BBC News* 15 December, URL (consulted Jan. 2009): <http://news.bbc.co.uk/2/hi/technology/4530930.stm>
- Beer, David (2009) 'Power through the algorithm? Participatory web cultures and the technological unconscious', *New Media & Society* 11(6): 985-1002.
- Benkler, Y. (2006) *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press.
- boyd, d. (2005) 'Academia and Wikipedia', *Corante* 4 January, URL (consulted Jan. 2009): [http://many.corante.com/archives/2005/01/04/academia\\_and\\_wikipedia.php](http://many.corante.com/archives/2005/01/04/academia_and_wikipedia.php)
- Bruns, A. (2008) *Blogs, Wikipedia, Second Life, and Beyond: From Production to Producership*. New York: Peter Lang. URL (consulted Jan. 2009): <http://produsage.org/>
- Burke, M. and R. Kraut (2008) 'Taking Up the Mop: Identifying Future Wikipedia Administrators', *Proceedings of the 2008 CHI Conference*, Florence, April 5-10, pp. 3441-

3446. New York: ACM. URL (consulted April 2009): <http://portal.acm.org/citation.cfm?id=1358628.1358871>

Carr, T. (2008) *The Big Switch. Rewiring the World, from Edison to Google*. New York: Norton.

Chesney, T. (2006) 'An empirical examination of Wikipedia's credibility', *First Monday* 11(11), URL (consulted Jan. 2009): [http://firstmonday.org/issues/issue11\\_11/chesney/](http://firstmonday.org/issues/issue11_11/chesney/)

Citizendium Beta, *Citizendium*, URL (consulted Jan 2009): [http://en.citizendium.org/wiki/Welcome\\_to\\_Citizendium](http://en.citizendium.org/wiki/Welcome_to_Citizendium)

Deleuze, Gilles (1990) 'Society of Control', *L'autre journal* (1), URL (consulted October, 2009): <http://www.nadir.org/nadir/archiv/netzkritik/societyofcontrol.html>

Digital Methods Initiative (2008), *Digital Methods Initiative*, URL (consulted Jan. 2009): <http://www.digitalmethods.net>

Galloway, A. (2004) *Protocol. How Control Exists after Decentralization*. Cambridge: MIT Press.

Govcom.org Foundation (2008), 'Govcom.org 10 Year Jubilee Workshop', *Digital Methods Initiative* 11-15 August 2008, URL (consulted Jan. 2009): [http://wiki.digitalmethods.net/Dmi/GovcomorgJubilee#Jubilee\\_Programme\\_archived](http://wiki.digitalmethods.net/Dmi/GovcomorgJubilee#Jubilee_Programme_archived)

Ghosh, R.A. and V. V. Prakash (2000) 'Orbiten Free Software Survey', in *First Monday* 5(7), URL (consulted Jan. 2009): [http://www.firstmonday.org/issues/issue5\\_7/ghosh/](http://www.firstmonday.org/issues/issue5_7/ghosh/)

Giles, J. (2005) 'Internet encyclopedias go head to head', *Nature* 438: 900-901, URL  
(consulted Jan. 2009):

<http://www.nature.com/nature/journal/v438/n7070/full/438900a.html>

Griffith, V. (2008) *WikiScanner Homepage*, URL (consulted Jan. 2009):

<http://virgil.gr/31.html>

Griffith, V. (2008) *Wikiwatcher Homepage*, URL (consulted March 2009):

<http://wikiwatcher.virgil.gr/>

Halavais, A. (2004) 'The Isuzu Experiment', *A Thaumaturgical Compendium Blog*, URL  
(consulted Jan. 2009): <http://alex.halavais.net/the-isuzu-experiment/>

Halavais, A. (2004) 'AlHalawi User Page', *Wikipedia*, URL (consulted Jan. 2009):

[http://en.wikipedia.org/wiki/User\\_talk:AlHalawi](http://en.wikipedia.org/wiki/User_talk:AlHalawi)

Halavais, A. and D. Lackaff (2008) 'An Analysis of Topical Coverage of Wikipedia',  
*Journal of Computer-Mediated Communication* 13: 429-40.

Howe, J. (2006) 'The Rise of Crowdsourcing', *Wired* 14(06), URL (consulted Sept. 09):

<http://www.wired.com/wired/archive/14.06/crowds.html>

Jenkins, H. (2006) *Convergence Culture. Where Old and New Media Collide*. Cambridge: MIT  
Press.

Keen, A. (2007) *The Cult of the Amateur: How blogs, MySpace, YouTube, and the rest of today's user-generated media are destroying our economy, our culture, and our values*. New York: Doubleday/Currency.

Kittur, A. and R. E. Kraut (2008) 'Harnessing the Wisdom of Crowds in Wikipedia: Quality through Coordination', *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work*, pp. 37-46. New York: ACM.

Kittur, A., E. Chi, B. Pendleton, B. Sun and T. Mytkowicz (2007) 'Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie', paper presented at CHI 2007, San Jose, April 28 - May 3.

Latour, B. (1991) 'Technology is Society Made Durable' in J. Law (ed.) *A Sociology of Monsters Essays on Power, Technology and Domination*, pp. 103-132. London: Routledge.

Negroponte, N. (1995) *Being Digital*. New York: Vintage Pbs.

Pfeil, U., P. Zaphiris, and C. S. Ang (2006) 'Cultural Differences in Collaborative Authoring of Wikipedia', *Journal of Computer-Mediated Communication* 12: 88-113.

Poe, M. (2006) 'The Hive', *The Atlantic Online* September, URL (consulted Jan. 2009): <http://www.theatlantic.com/doc/200609/wikipedia>

Rogers, R. (2009) *The End of the Virtual: Digital Methods*. Amsterdam: Vossiuspers UvA.



Rogers R., S. Niederer et al. (2008) 'Networked Content', *Digital Methods Initiative*, URL (consulted Jan. 2009): <http://wiki.digitalmethods.net/Dmi/NetworkedContent>

Rosenzweig, R. (2006) 'Can History Be Open Source? Wikipedia and the Future of the Past', *The Journal of American History* 93(1): 117-146.

Shirky, C. (2008) *Here Comes Everybody: The Power of Organizing without Organizations*. New York: Penguin Press.

Shirky, C. (2005) 'K5 Article on Wikipedia Anti-Elitism', *Corante*, URL (consulted Jan. 2009):  
[http://many.corante.com/archives/2005/01/03/k5\\_article\\_on\\_wikipedia\\_antielitism.php](http://many.corante.com/archives/2005/01/03/k5_article_on_wikipedia_antielitism.php)

Stalder, F. and J. Hirsch (2002) 'Open Source Intelligence', *First Monday* 7(6-3), URL (consulted Jan. 2009):  
<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/961>

Stevenson, M. and R. Rogers (2009) 'Digital Methods: First Steps', *Easst Review* 28(2), URL (consulted Aug. 2009): <http://www.easst.net/review/june2009/stevenson>

Sunstein, C. R. (2006) *Infotopia: How Many Minds Produce Knowledge*. Oxford: Oxford University Press.

Surowiecki, J. (2004) *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Societies and Nations*. New York: Doubleday.

Swarz, A. (2006) 'Who writes Wikipedia?', *Raw Thought Blog*, URL (consulted Nov. 2008):  
<http://www.aaronsw.com/weblog/howwriteswikipedia/>

Tapscott, D. and A.D. Williams (2006) *Wikinomics. How Mass Collaboration Changes Everything*. New York: Penguin.

Wikimedia (n.d.) 'Blocked Users', *Wikimedia*, URL (consulted March 2009):  
[http://meta.wikimedia.org/wiki/Blocked\\_user](http://meta.wikimedia.org/wiki/Blocked_user)

Wikimedia (n.d.) 'Bot Activity Matrix', *Wikimedia Stats*, URL (consulted Nov. 2008):  
<http://stats.wikimedia.org/EN/BotActivityMatrix.htm>

Wikimedia (n.d.) 'Wikipedia Statistics German', *Wikimedia Stats*, URL (consulted Nov. 2008): <http://stats.wikimedia.org/EN/TablesWikipediaDE.htm>

Wikimedia (n.d.) 'List of Wikipedias', *Wikimedia*, URL (consulted Nov. 2008):  
[http://meta.wikimedia.org/wiki/List\\_of\\_Wikipedias#1.2B\\_articles](http://meta.wikimedia.org/wiki/List_of_Wikipedias#1.2B_articles)

Wikimedia (n.d.) 'Our Projects', *Wikimedia Foundation*, URL (consulted Nov. 2008):  
[http://wikimediafoundation.org/wiki/Our\\_projects](http://wikimediafoundation.org/wiki/Our_projects)

Wikimedia (n.d.) 'System Administrators', *Wikimedia*, URL (consulted March 2009):  
[http://meta.wikimedia.org/wiki/System\\_administrators](http://meta.wikimedia.org/wiki/System_administrators)

Wikimedia (n.d.) 'User Rights Manual', *Mediawiki*, URL (consulted March 2009):

[http://www.mediawiki.org/wiki/Manual:User\\_rights](http://www.mediawiki.org/wiki/Manual:User_rights)

Wikimedia (n.d.) 'Wikipedia user groups', *Wikimedia*, URL (consulted Jan. 2009):

[http://meta.wikimedia.org/wiki/User\\_groups](http://meta.wikimedia.org/wiki/User_groups)

Wikipedia (n.d.) 'Editing Frequency of All Bots', *Wikipedia*, URL (consulted Nov. 2008):

[http://en.wikipedia.org/wiki/Wikipedia:Editing\\_frequency/All\\_bots](http://en.wikipedia.org/wiki/Wikipedia:Editing_frequency/All_bots)

Wikipedia (n.d.) 'History of Wikipedia Bots', *Wikipedia*, URL (consulted Nov 2008):

[http://en.wikipedia.org/wiki/Wikipedia:History\\_of\\_Wikipedia\\_bots](http://en.wikipedia.org/wiki/Wikipedia:History_of_Wikipedia_bots)

Wikipedia (n.d.) 'Ignore all rules', *Wikipedia*, URL (consulted March 2009):

[http://en.wikipedia.org/wiki/Wikipedia:Ignore\\_all\\_rules](http://en.wikipedia.org/wiki/Wikipedia:Ignore_all_rules)

Wikipedia (n.d.) 'List of Wikipedians by number of edits', *Wikipedia*, URL (consulted Nov. 2008):

[http://en.wikipedia.org/wiki/Wikipedia:List\\_of\\_Wikipedians\\_by\\_number\\_of\\_edits#List](http://en.wikipedia.org/wiki/Wikipedia:List_of_Wikipedians_by_number_of_edits#List)

Wikipedia (n.d.) 'Neutral Point of View', *Wikipedia*, URL (consulted March 2009):

[http://en.wikipedia.org/wiki/Wikipedia:Neutral\\_point\\_of\\_view](http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view)

Wikipedia (n.d.) 'No Original Research', *Wikipedia*, URL (consulted March 2009):

[http://en.wikipedia.org/wiki/Wikipedia:No\\_original\\_research](http://en.wikipedia.org/wiki/Wikipedia:No_original_research)

Wikipedia (n.d.) 'Verifiability', *Wikipedia*, URL (consulted March 2009):

<http://en.wikipedia.org/wiki/Wikipedia:Verifiability>

Wikipedia (n.d.) 'Wikipedia Bot Policy', *Wikipedia*, URL (consulted Nov. 2008):


[http://en.wikipedia.org/wiki/Wikipedia:Bot\\_policy](http://en.wikipedia.org/wiki/Wikipedia:Bot_policy)

WikiTrust (n.d.) 'Main Page', *WikiTrust*, URL (consulted Sept. 2009):

[http://wikitrust.soe.ucsc.edu/index.php/Main\\_Page](http://wikitrust.soe.ucsc.edu/index.php/Main_Page)

Zittrain, Jonathan (2008) *The Future of the Internet*. New York: Penguin.

FIGURES AND CAPTIONS

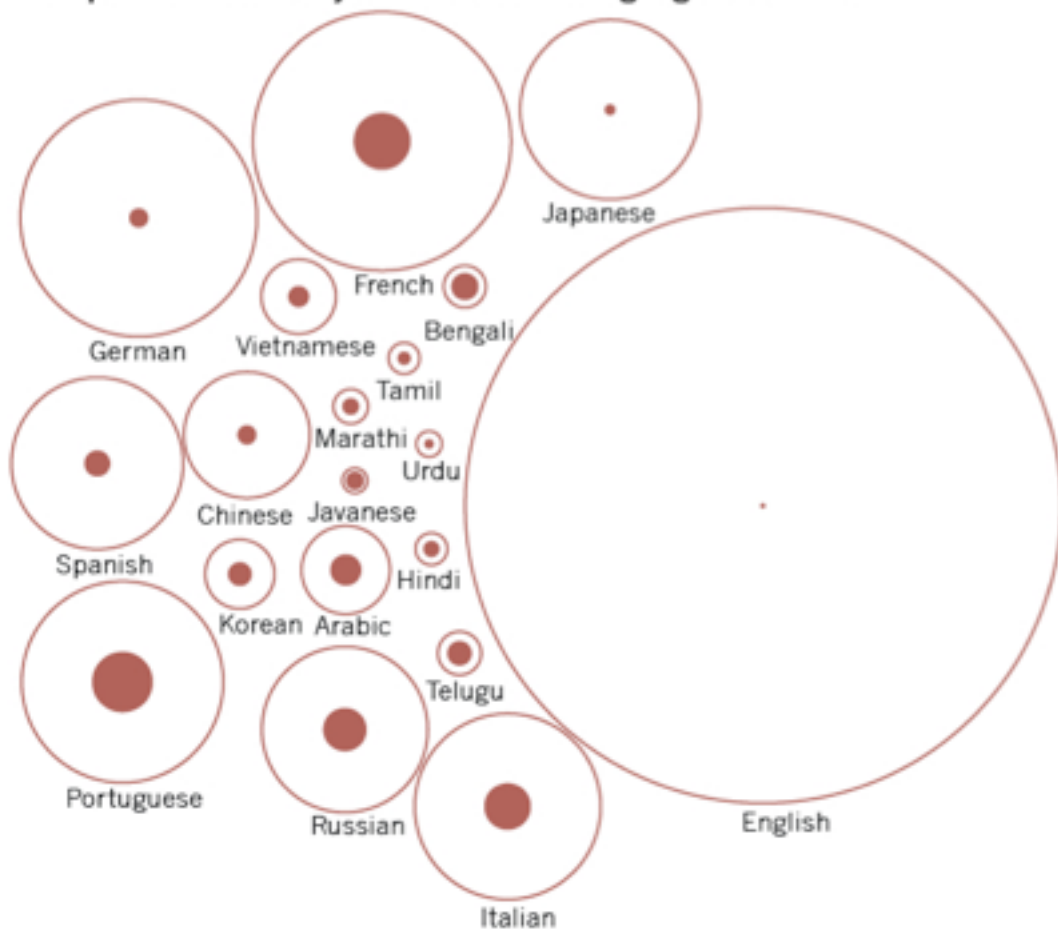
<i>Permission Level</i>	<i>Wikipedia Users</i>
Most Permissions	Developer / System Administrator
	Steward
	Checkuser
	Oversight
	Bureaucrat
	Administrator / Sysop
	Bot
	Registered User
	Newly Registered User
	Anonymous User
No Permissions	Blocked User

**Figure 1:** Schematic overview of global and local categories of Wikipedia users according to permission levels.



**Figure 2:** Bot-created article compared to a human-edited article. The upper screenshot is the La Grange, Illinois article as created by rambot on December 11, 2002. The lower screenshot shows the same article on January 12, 2009. Source: Wikipedia, URL: [http://en.wikipedia.org/wiki/La\\_Grange,\\_Illinois](http://en.wikipedia.org/wiki/La_Grange,_Illinois)

### Wikipedia Bot Activity in Most-Used Languages Worldwide



**Figure 3:** Visualization of Wikipedia bot activity in most-used languages worldwide, overview and detail. The outer circle is the Wikipedia size in that specific language; the inner dot depicts the percentage of bot activity in that language Wikipedia. Analysis by Rogers, Niederer et al. Graphic by Auke Touwslager using the Dorling Map Tool, Digital Methods Initiative, Amsterdam. URL:

<http://wiki.digitalmethods.net/Dmi/NetworkedContent>

### Wikipedia Bot Activity in Endangered Languages



Figure 4: Bot activity in endangered language Wikipedias. Analysis by Rogers, Niederer et al. Graphic by Auke Touwslager using the Dorling Map Tool, Digital Methods Initiative, Amsterdam. URL: <http://wiki.digitalmethods.net/Dmi/NetworkedContent>