The future of Wikipedia

In his book *The Revenge of Gaia*,[1] James Lovelock advocates the compilation of a book summarizing man's accumulated scientific and technological knowledge, to be used in the event of of some global catastrophe that would destroy both conventional libraries and digital storage. Nowadays, many people would say that this compilation already exists in the form of Wikipedia. Indeed many University undergraduates no longer buy textbooks, or maybe just buy one covering their core subject for convenient reference, because whenever they are called upon to find out and verify facts, they can swiftly—and, they imagine, reliably—look them up on Wikipedia.

Probably few readers of this editorial are unaware of Wikipedia and what it is. In a sense it embodies an apotheosis of the World Wide Web. Anyone connected to the Internet can create or edit an article; and anyone connected to Internet can access and read an article. Its mission is "to make human knowledge available to every person on the planet".

Wikipedia is extremely firmly embedded in the ethos of our age, perhaps even more so than phenomena such as Facebook and Twitter. It enjoys widespread popularity. Some quantitative measures of this are the numbers of articles and of users (i.e., readers of articles). It is certainly easy, if not perhaps wholly objective, to consult Wikipedia itself (Wikipedia:Statistics) for this information. There are about five million articles in English (albeit that well over half of them are mere "stubs", which, according to Wikipedia's own quality assessment criteria, "provide very little meaningful content; [they] may be little more than a dictionary definition"). Wikipedia: Statistics asserts that on average about 800 new articles are created daily; historical data suggests that, although initially (Wikipedia started in 2001) growth was exponential, with a doubling time of between one and two years, during the last 3 or 4 years growth has indeed looked linear; the overall kinetics seem to be following something like a logistic law with an asymptote of around 6 million articles, which may reflect fatigue rather than the limits of human knowledge.

Most contributors to Wikipedia content are "registered users", of which there are over 25 million. "Active editors" seem to be defined as those making at least one contribution per month;[2] their number is relatively tiny (less then 250,000); about half of them spend at least an hour a day editing, which seems to epitomize *acharnement*. The volume of visits to Wikipedia pages to consult content is hard to quantify, but it could be of the order of 10^{11}–10^{12} per annum. These statistics place Wikipedia in sharp contrast to an early vision of the future in which everyone is interconnected in weblike fashion.[3] In Forster's story, the main activity of the population appears to be giving lectures but the audiences are small, of the order of ten listeners;

[1] J. Lovelock, *The Revenge of Gaia. Why the Earth Is Fighting Back—and How We Can Still Save Humanity*. London: Allen Lane (2006).
[2] See the Wikipedia article "Wikipedia community".
[3] E.M. Forster, *The Machine Stops*. Oxford and Cambridge Review (1909).

Nanotechnology Perceptions **11** (2015) 131–135
Received 30 November 2015

doi: 10.4024/N13RA15E.ntp.15.03

one may infer that everyone gives roughly the same number of lectures as they listen to. This at least manifests a healthy equilibrium of society, regardless of the merits or otherwise of the actual content of the lectures, and is surely better than the vast preponderance of Wikipedia readers over writers, just as it is surely better for every town to have its own active football team than 3 milliard people (the estimated television audience of the last World Cup final) watching a single football match.

Be that as it may, the concept of Wikipedia, namely universal access with respect to both contributing and reading, seems unexceptionable, even laudable at first sight. Wikipedia has no fixed rules as such, but it is expected that errors in articles will be subsequently corrected by editors with more accurate knowledge. This may be called a "law of large numbers" approach, in which the cumulative sum of many small, incremental steps eventually leads to an impressive edifice. Our present interest is to enquire whether this implicit claim is, in fact, an arrogation. In other words, can such a "statistical" approach ever yield the accuracy associated with the traditional mode of compiling encyclopaedias? The encyclopaedia appears to have been invented by the Chinese. Enormous encyclopaedias were commissioned by emperors in order to capture the extent of knowledge in their country. To what extent the later European encyclopaedists such as Diderot and d'Alembert, were inspired by those projects is not clear. In our own time, the zenith of this tradition is embodied in the great 11th edition (1911) of *Encyclopaedia Britannica* (now out of copyright and available on the Internet); articles on topics that have not noticeably advanced since then remain paragons of excellence to this day. It was the task of the editor or editors to persuade the leading experts in each field to contribute an article, which was duly signed by the author. In contrast, even registered Wikipedia editors are anonymous (unless they choose to reveal their identity, but only a tiny proportion so chooses). An editor must provide an authentic e-mail address upon initial registration, but there seems to be no way in which a casual reader could determine the identity of an editor.

Wikipedia, then, embodies an "epistemological law of large numbers", which asserts that for a sufficiently large number of editors e the deviation $T_e - m_e$, where T is the sum of the "truth" contributed by the editors ($T_e = \sum x_i$, where the x are increments of "truth"), and m the mean of the efforts of the editors ($m_e = \sum \mathrm{E}(x_i)$, where E signifies the expectation) is likely to be small in comparison to e. But there is nothing about the nature of knowledge that makes it reasonable for the totality of knowledge to be regarded as the *sum* of its components (other than the everyday, but unfortunately misleading, expression "the sum of our knowledge") nor for the mean to be the arithmetic mean. Knowledge is more likely to be multiplicative, engendering a distribution obeying very different laws. For a random multiplicative process the successive appending of terms leads to ever greater fluctuations and for complex systems generally there is no guarantee of convergence to a central "truth" by simply accumulating more and more data.

This criticism would apply even if the intentions of all contributing editors were altruistically motivated to provide the best possible information. In contrast to the previous encyclopaedic tradition, Wikipedia actually discourages leading experts to write articles about their core expertise, because in doing so they would have to cite themselves, and this is deemed to be self-promotion, which is anathaema to Wikipedia. It must also be mentioned that alongside this peculiar convention, no article is permitted to stand on its own merits, but must be based on already published work and, moreover, the sources must be accessible and, hence, verifiable.

Clearly, if anonymity is insisted upon, such a procedure is necessary, because the author of an article has no reputation to lose even if the most absurd nonsense is put forward. In essence, then, Wikipedia articles become a kind of "average" of all already-published knowledge on a topic; in our present era of excessive publication, there is probably very little that escapes scrutiny. It might be thought that a Wikipedia article thus has the character of a review in a journal like *Chemical Reviews* (published by the American Chemical Society), or the *Specialist Periodical Reports* published by the Royal Society of Chemistry, but this would be a great mistake. The ingredients of the Wikipedia "average" tend to be rather undiscriminating. Typically many of the quoted sources are little better than ephemeral pamphlets, with a great preponderance of online-only material. It is perhaps understandable that the mentality of the typical Wikipedian will tend to favour such sources, on the premiss that Wikipedia is itself an Internet source. Protection against such sources being removed from the web is in principle provided by subsequent editors who will replace the reference no longer existent by another one. Furthermore, articles are not supposed to cite original research; they should preferably refer only to reviews, inclusion of knowledge in such a review being deemed to constitute general acceptance of the validity of the quoted facts by the community. Hence, Wikipedia articles are "reviews of reviews" and in this sense resemble popular expositions (I presume that the style of *Chemical Reviews* would be considered to be far too technical for the average Wikipedia reader).

Unfortunately, these conventions or "guidelines" are evidently widely flouted. Most articles on topics dealing with science or technology cite original research papers as well as other sources (such as reports from obscure organizations available only on the Internet) whose reliability is questionable. In fact, the criterion for citability is not so much quality of scholarship but popularity. If an article is abundantly cited by others, there can be no grounds for refusing its inclusion in a Wikipedia article, for indubitably it reflects the popular view of the matter. This has very little to do with "truth"—as Galileo has remarked, "in the sciences, the authority of thousands of opinions is not worth as much as one tiny spark of reason in an individual man." This maxim is anathaema to the spirit of Wikipedia. On the contrary, if a solecism is repeated often enough, it acquires inalienable legitimacy in the eyes of Wikipedia.

Beyond that, unfortunately the foibles of human nature have contrived to introduce further distortion. Many editors, it has turned out, have personal agendas different from the disinterested dissemination of objective knowledge.[4] To know that this is the case, it suffices to look at Wikipedia's own guidelines, introduced in an attempt to dampen distortion of the ideal. Hence, for example, although any change made by an editor is deemed to be an "improvement", the reality is that some changes are destructive. It is also disquieting that the whole system is not actually self-regulating, but policed by a small number of "administrators"—apparently between 1000 and 2000, which is a tiny proportion of the active editors, who themselves constitute a tiny proportion of the registered users. This reminds one of the party members in a Communist society. Administrators seem to be chosen according to their zeal in creating new articles and editing existing ones. These administrators doubtless fulfil a valuable role by adjudicating—but

[4] The section "Motivation" (for editing) found in the article "Wikipedia community" makes no mention of such departures from the ideal.

in principle from ignorance; were they themselves to be experts in the field of the article in dispute, they would be deemed not to be objective—disputes over content. Hence, they must judge the outcome on extremely general criteria, with no guarantee that anything like "truth" will be upheld.

Readers who, like myself, occasionally read a Wikipedia article when it is thrown up by a search engine, will probably only be aware of articles with scientific and technical content. There is, however, a vast number of other kinds of articles of more or less bizarre nature, not corresponding to anything in traditional encyclopaedias. For example, there are vast numbers of "Lists of" the most trivial items, such as "Russian islands" or "Finnish lakes". There are vast numbers of articles on culinary recipes and (of course) on any sporting or television personality likely to be widely known (however, one is discouraged from creating articles about one's friends unless they have a large following). Even more oddly, many scientific journals have Wikipedia articles (not, I can affirm, at least on the day this editorial was submitted, *Nanotechnology Perceptions*). Given that all these journals have their own official websites (I suppose), it is hard to understand the motivation of spending time to merely duplicate what is readily available elsewhere. Furthermore, cursory inspection of these "journal" articles reveals that they all seem to have a very similar, very brief and highly stereotyped format, as if they were created by the publishers themselves merely as a publicity tool which, of course, completely contradicts the ethos of Wikipedia.

One measure of the success of Wikipedia has been its proliferation in many languages other than English. Given that encouraging fact, it is surprising that so many badly written English articles are still tolerated. It is an impoverishment if an attractive style is not cultivated, and if the writing is so bad that its meaning becomes ambiguous the article is not even usefully informative. Surely this should be more strictly policed. It may be, of course, that the bad articles are indeed slowly being eliminated but in general it cannot be assumed that Wikipedia will inevitably get better and better; on the contrary, it seems more likely that Gresham's law is applicable to it.

Wikipedia is obsessed with quantity rather than quality. It does attempt to assess the quality of its articles (see Wikipedia:Statistics), but peering through the veil of the rather grandly named "official review", it appears that assessment is based primarily on popularity—the number of visits and web-connectedness. A page visit does not, in itself, tell us how much information was gained from the page. According to the Shannon–Wiener index, the gain in information equals the information possessed after visiting the page minus the information possessed before the visit. If the aim of every Wikipedia article is to be a perfect reflexion of the popular view, then the gain in information is likely to be zero, since anyone enjoying a convivial social life will be well enough aware of the popular view. Only an anchorite needs to consult an encyclopaedia in order to be apprised of the popular view. Perhaps that is what Wikipedia is anticipating—a society in which we are all essentially isolated from one another except via the Internet.[5]

The anonymity and discouragement of real experts to contribute to articles condemns Wikipedia to mediocrity. In my own experience, only a very small number of articles are of high

[5] The latest acolyte recruited to propagate Wikipedia's allure is a mass-produced doll called Cayla; it can respond to questions by finding an article in Wikipedia and extracting the answer from it. In the run-up to Christmas, it has been stated that Cayla is this year's most popular toy. One wonders whether page visits from Cayla also count towards the popularity index.

quality. They have largely been written by one person, either a true expert or someone who has, in the tradition of past popularizers of science, gone to the trouble to thoroughly understand a field even though he or she has not contributed original scholarship to it. But even such articles are subject to constant, populistic, "improvement". Could *Principia* ever have been created on Wikipedia? It is, actually, absurd even to ask the question, and should by some chance Newton ever have created a Wikipedia page, knowing his somewhat irascible nature he would probably have died of apoplexy after noticing that someone had edited it.

Returning to the idea with which this essay began, we conclude that Wikipedia has more in common with Flaubert's *Bouvard et Pécuchet* than with anything resembling a practically useful and reliable repository of knowledge. Will, then, Wikipedia eventually disappear, as has been predicted for Facebook?[6] If so, there is perhaps little point in criticizing it, other than to discourage undergraduates and others from relying too heavily on it, and discouraging would-be compilers of the totality of human knowledge from assuming that it has been done by Wikipedia.

J.J. Ramsden

[6] J. Cannarella and J.A. Spechler, Epidemiological modeling of online social network dynamics. arXiv:1401.4208 [cs.SI].