Wikipedia(s) on the language map of the world

ANDREW DALBY

A remarkable explosion in editing and knowledge management

WIKIPEDIA, the open-access encyclopedia, builds on the facilities offered by the Internet, but – in Internet terms – it is a strange beast. Like the Web in general, anyone can add to it; unlike the Web in general, anyone can edit or delete the contributions of others. I'm sure I wasn't the only person who decided, on first hearing how it worked, that it would never be worth much. I was wrong.

Introduction

Wikipedia began in 2001. By 1 March 2006, the English-language Wikipedia had reached its millionth article ('Jordanhill railway station') and on 24th November it reached 1,500,000 ('Kanab ambersnail,' an endangered species of the United States). The growth rate is still increasing. The Wikimedia Foundation is highly ambitious, in its fairly benevolent way, and shows little inclination to restrict the expansion of the galaxy of Wikis that are attached in various ways to the English Wikipedia, which remains the largest and the best-known of the family.[1]

The average surfer will seldom meet them, but there are many other independent sites using the same wiki structure. Grandmother of all the rest is the venerated WikiWikiWeb, started in 1995 as a collaborative encyclopedia for software developers.[2] The three most evident things that the software (the 'wiki engine') has to do are these:

1 to allow a surfer to search for, create, and edit articles within an ordinary Web browser
2 to manage the security of the resulting 'encyclopedia', if necessary by restricting the right to edit: the Wikipedias are essentially open to all, but some of the independent wikis are not[3]
3 to keep the links between articles functioning.

This task increases exponentially as a wiki site grows: within and among the 250 language Wikipedias, the total is already well over 100 million links, any of which a user must be able to select at any moment with a click. The number has recently been increasing at the rate of 10 new-language Wikipedias per month.

This article will not try to describe the whole Wikimedia galaxy. It will stick to Wikipedia in English, and that's ambitious enough. The English-language Wikipedia, by far the biggest of them, now (28th November 2006) contains 1,506,659 articles. The German Wikipedia reached 500,000 articles on 23rd November (note in passing: the English Wikipedia has added that many articles to its total in just six months), while the French Wikipedia reached the 400,000 milestone on 27th November. The newest and smallest Wikipedia, number 250, is

ANDREW DALBY studied at Bristol Grammar School and the University of Cambridge, and has a PhD in ancient history from Birkbeck College, the University of London (1987–93). His latest book, 'Language in Danger', considers the consequences of the current world language crisis: a language dying every two weeks. He edits and creates Wikipedia articles as a way of building up knowledge of topics about which he plans to write, such as the histories of language and food. He is in effect an Interwikipedian, having undertaken Wikipedia articles in eight languages: 'If I have skill in this area, it's simply that languages don't frighten me.' He lives in France.

DOI: 10.1017/S0266078407002027
 © 2007 Cambridge University Press

https://doi.org/10.1017/S0266078407002027 Published online by Cambridge University Press

in the Lak language of Dagestan, in the Caucasus, with one article and 20 users.

One more statistical measure will show how much Wikipedia matters. People who Google already know that for a great many Google searches one or more Wikipedia entries will turn up high on the first page of the results. They don't all know that Wikipedia now comes eleventh in alexa.com's traffic ranking of world websites. For a strictly non-commercial site with relatively academic content, that is astonishing success; what's more, the trend is steadily upwards, though it will be hard to overtake the top four: yahoo.com, msn.com, google.com, and the highly popular Chinese search engine, baidu.com.

So, how is Wikipedia affecting the languages of the world?

Other languages: 249 and growing

Thanks to the hospitality of the Foundation, any small band of users can ask to be allowed to create their own online encyclopedia in their own language. Although 'Wikipedia's volunteer developers are overworked and understaffed', the request is likely to be accepted with minimal delay. On each successive occasion these developers have plenty to do, in cooperation with the band of language enthusiasts, to make the basic processes work smoothly for each new language, whatever its script. We might call this activity the 'localization' of a new Wikipedia. Once it's done, it's up to local users to act as authors and editors.

But in fact it takes more than a small group to create a real, really useful encyclopedia. The German and Spanish Wikipedias have more than 250,000 registered users each. Usually only those who intend to edit Wikipedia bother to register, so *these figures mean a quarter of a million editors for each language*. A further twenty-four language versions have at least 10,000 registered users each: see Table 1. Admittedly, many lose enthusiasm after an hour of editing, but even if only a hundredth of those who sign up remain active, that adds up to a lot of volunteer contributors.

In all cases, however, the growth rate is impressive. It may seem curious that the French Wikipedia is so much smaller than the German, yet in the course of 2005 the French Wikipedia showed more traffic growth than any other French-language site, increasing its web traffic position from 137th to 37th. It may seem strange that the Spanish Wikipedia is so much smaller than the Portuguese, but this is partly because it has had Spanish-language competition in the *Enciclopedia Libre*, in spite of which it is now among the most-visited of all Wikipedias. The Chinese Wikipedia has suffered through most of its life from blocks imposed on access from mainland China, and now has strong competition from a Chinese wiki operated by baidu.com.

It becomes all the more astonishing that so many Wikipedias exist when, in most cases, each has only a few enthusiasts to keep it growing. In some minority languages there was relatively little reliable factual material available on the Web until that particular Wikipedia began its expansion. The language Wikipedias serve not only the growing number of native speakers who have access to the Internet: they also serve language learners and translators.

It's useful to these communities that many of the articles on the smaller Wikipedias are translated from English or one of the larger languages. If the translation has been done reasonably well, comparing it with the original will throw up a lot of useful vocabulary; in fact there are few better ways to pick up technical and scientific vocabulary in a minority language than to read relevant articles in Wikipedia. Those who pursue this method, however, soon encounter problems. Wikipedia is wholly compiled by volunteers, who cannot be blamed for deciding that it's more fun to name an article than to write it: therefore many articles on the smaller Wikipedias are only half translated, or machine-translated, or not translated at all, or are 'stubs' (consisting of a single not very informative sentence) or entirely empty. Brief articles about places can be generated, and transferred from one language to another, using a software utility: this has been done on many Wikipedias, and the result is that too many of the geographical articles tell you no more than you could learn from a telephone directory.

Still, these are early days. Experience shows that stubs do eventually turn into full and useful articles. Experience also shows that the number of Wikipedias will continue to grow.

The versions currently available in southeast Asian, African, and American indigenous languages are not very numerous and they are individually very small (see Table 2). This is not surprising since computer use among speakers of these languages is still rare. It will increase; and the fact that future students will

https://doi.org/10.1017/S0266078407002027 Published online by Cambridge University Press

TABLE 1 The 25 busiest Wikipedias (from the most to the least busy)

http://	Language	Registered users on 28 Nov 06	Articles on 30 Nov 2006
En.wikipedia.org	English	2 862 432	1 507 384
de.wikipedia.org	German	311 052	502 882
es.wikipedia.org	Spanish	266 553	174 099
fr.wikipedia.org	French	180 540	401 023
pt.wikipedia.org	Portuguese	155 076	199 380
zh.wikipedia.org	Chinese	147 758	103 013
it.wikipedia.org	Italian	112 608	218 699
ja.wikipedia.org	Japanese	96 161	292 694
pl.wikipedia.org	Polish	83 022	319 508
nl.wikipedia.org	Dutch	77 261	243 000
tr.wikipedia.org	Turkish	50 504	39 660
fi.wikipedia.org	Finnish	39 305	88 719
ar.wikipedia.org	Arabic	38 180	20 723
ru.wikipedia.org	Russian	33 725	118 757
id.wikipedia.org	Indonesian	30 363	45 274
he.wikipedia.org	Hebrew	28 559	48 491
no.wikipedia.org	Norwegian (bokmål)	27 754	87 758
sv.wikipedia.org	Swedish	26 213	195 985
ro.wikipedia.org	Romanian	23 658	42 132
fa.wikipedia.org	Persian	22 662	16 553
vi.wikipedia.org	Vietnamese	21 191	12 282
th.wikipedia.org	Thai	14 636	16 770
cs.wikipedia.org	Czech	12 201	50 998
hu.wikipedia.org	Hungarian	11 203	44 562
bg.wikipedia.org	Bulgarian	10 165	32 423
da.wikipedia.org	Danish	10 113	53 132
ko.wikipedia.org	Korean	10 018	29 653

Source <http://meta.wikimedia.org/wiki/List_of_Wikipedias>

be able to consult an online reference source in their own mother tongues means that some of those languages, now counted as endangered, will be better placed for survival. [One of the few novel developments I have come across in recent years regarding sustaining the world's small languages: *Ed.*]

A different question arises with regard to the Wikipedia versions: an astonishing number of them (see Table 3), in local languages and dialects of Europe (some of which have until recently had no standard written form and no reference publishing – in fact few publications of any kind). In terms of size and number of registered users, such versions are comparable to those in some Third World languages, but with a major difference: they have a much larger *notional* user base, and don't reach it. Most people in Lombardy who have computers don't ever think of using the Lombard Wikipedia, let alone editing it: they use the Italian or the English Wikipedia. I suspect that

https://doi.org/10.1017/S0266078407002027 Published online by Cambridge University Press

TABLE 2 Wikipedias in indigenous languages of Southeast Asia, Africa and the Americas and Oceania

Locations	**Over 1000 articles**, between 100 and 1000, *under 100*
South-East Asia	**Chinese**, **Thai**, **Vietnamese**, Min Nan, Cantonese, Wu, Tibetan, Classical Chinese, *Lao*, *Khmer*, *Burmese*, *Yi*, *Zhuang*, *Min Dong*
Island southeast Asia and Oceania	**Indonesian**, **Malay**, **Tagalog**, **Ilokano**, **Waray-Waray**, **Kapampangan**, **Cebuano**, **Sundanese**, **Javanese**, **Banyumasan**, Pangasinan, Tok Pisin, Maori, Tahitian, Nauruan, Tongan, Tetun, *Buginese*, *Chamorro*, *Bislama*, *Chavacano*, *Hiri Motu*, *Norfolk*, *Samoan*, *Hawaiian*, *Fijian*, *Marshallese*
Africa	**Swahili**, Amharic, Malagasy, Lingala, Wolof, Kongo, Bambara, *Somali*, *Hausa*, *Oromo*, *Afar*, *Tigrinya*, *Zulu*, *Yoruba*, *Kanuri*, *Lesotho*, *Kinyarwanda*, *Xhosa*, *Tswana*, *Tumbuka*, *Fulfulde*, *Twi*, *Igbo*, *Kirundi*, *Luganda*, *Chichewa*, *Shona*, *Ewe*, *Kikuyu*, *Kwanyama*, *Tsonga*, *Herero*, *Ndonga*, *Venda*, *Swati*
North America	Cherokee, Cree, *Inuktitut*, *Inupiaq*, *Kalaallisut*, *Cheyenne*, *Navaho*, *Choctaw*, *Creek*
Central America, Caribbean	Nahuatl, Haitian
South America	Quechua, *Aymara*, *Guaraní*

Source <http://meta.wikimedia.org/wiki/List_of_Wikipedias_by_language_family>

these European dialect Wikipedias are fated to be little more than the playground of small groups of enthusiasts, but I would love to be proved wrong.

The English Wikipedia

There are nearly three million registered users of this 'edition'. This number (however impressive at first sight) is not a measure of usage: most of those who consult it never think of registering. Three million is the number of those who have, or once had, the intention of editing.

For an encyclopaedia that has no print-on-paper form, and that began to exist only six years ago, its success is astonishing. Even its inventors have been astonished. Jimmy Wales conceived Nupedia in 2000 as a collaborative expert-written peer-reviewed encyclopedia, and it got nowhere;[4] he therefore unenthusiastically permitted Wikipedia as a supplement open to all, and its success surpassed his wildest dreams.

Some like the idea and some don't. But most surfers now find that they use Wikipedia, and for good reason: it is, for many topics, expected and unexpected, the best available source of reference. In December 2005, the journal *Nature* undertook a blind comparative test of 42 entries in the *Encyclopaedia Britannica* and Wikipedia, and found them to be of comparable quality, though Wikipedia had a higher average of errors per article (3.62 compared with 2.91).[5] It's not only science; in some areas of history and literature no printed encyclopedia matches the range and depth of Wikipedia. Across the whole range of subjects it continues to grow with dizzying speed.

Jimmy Wales has never been quite convinced, it seems. He and his Wikipedia Foundation continue to work on ideas for making Wikipedia magically error-free, for making the articles stand still when they have reached perfection,[6] for grading articles and publishing CD collections of the top scorers. They may do all these things, but so long as they keep the site going it seems that Wikipedia will live, prosper, and grow in accordance with their original, reluctant vision.

As long as it remains open to all, Wikipedia will never be free from errors. Anyone – expert, inexpert, illiterate or malicious – can add statements to Wikipedia articles at any time, and some of them will be wrong. In practice, as habitual Wikipedia editors know, the errors, illiteracies, and vandalism are confined to a limited number of articles, on topics that attract attention because they are in the news, or relate to some political controversy, or are studied in classrooms. In practice, vandalism in particular is easily and quickly 'reverted'.

One of the most interesting markers of the world reach of Wikipedia is the origin of its editors: see Table 4. Naturally, the majority of

https://doi.org/10.1017/S0266078407002027 Published online by Cambridge University Press

TABLE 3 European minority language Wikipedias

Celtic languages	Breton, Cornish, Irish, Manx, Scottish Gaelic, Welsh
Romance languages	Aragonese, Aromanian, Asturian, Catalan,[7] Corsican, Franco-Provençal, Emilian-Romagnol, Friulian, Galician, Ladino, Ligurian, Lombard, Neapolitan, Norman, Occitan, Piedmontese, Sardinian, Sicilian, Tarantino, Venetian, Walloon
Germanic languages	Alemannic, Bavarian, Faroese, Frisian, Limburgs, Low German, Low Saxon, Nynorsk, Ripuarian, Scots, West Flemish, Zeelandic
Slavic and Baltic languages	Kashubian, Serbocroatian,[8] Upper Sorbian, Samogitian
Turkic languages	Chuvash, Tatar
Uralic languages	Sami, Voro, Udmurt, Komi
Other language groups	Basque, Romani

Source <http://meta.wikimedia.org/wiki/List_of_Wikipedias_by_language_family>

TABLE 4 Origins of edits to the six busiest Wikipedias

English Wikipedia	United States: 52.1%, United Kingdom: 15.9%, Canada: 7.3%, Australia: 4.4%, Germany: 1.8%, Netherlands: 1.6%, France: 1.0%
German Wikipedia	Germany: 76.2%, Austria: 5.4%, United States: 4.8%, Switzerland: 4.4%, Netherlands: 3.2%
Spanish Wikipedia	Spain: 39.2%, Argentina: 10.7%, Chile: 8.8%, Netherlands: 8.4%, Mexico: 7.0%, Venezuela: 5.1%, Peru: 3.5%, United States: 3.1%, Colombia: 2.7%, Uruguay: 1.3%, Germany: 1.1%
French Wikipedia	France: 67.3%, Belgium: 6.2%, Netherlands: 5.4%, United States: 4.3%, Canada: 4.2%, Switzerland: 3.2%, Germany: 1.2%
Portuguese Wikipedia	Brazil: 61.3%, Portugal: 17.7%, Netherlands: 11.1%, Spain: 2.3%, France: 1.4%, Belgium: 1.2%, United States: 1.1%
Chinese Wikipedia	Hong Kong: 28.6%, Taiwan: 25.9%, United States: 13.7%, Netherlands: 8.2%, China: 6.1%, Germany: 3.6%, Belgium: 2.3%, Canada: 2.0%, France: 1.3%, Mongolia: 1.2%, Japan: 1.2%

Source <http://meta.wikimedia.org/wiki/Edits_by_project_and_country_of_origin[9]>

edits in each language Wikipedia are made from countries where native speakers live. For English, this means the United States (more than 50%), Britain, Canada and Australia; but other countries play their part, beginning with those of western Europe in which computer use is particularly widespread. All English Wikipedia editors are aware of very active collaboration from people whose first language is not English.

Because of its wide range of contributors, for many of whom English is not a mother tongue, the English of Wikipedia articles is variable. On the whole an encyclopedic tone is adopted, and the house style rules are sensible and widely acceptable, but many articles have no consistent style. Some have British style and spelling, some American, and many have a mix of the two; some will suddenly switch because the latest editor so decided (although such changes are deprecated). Some articles, especially on international current affairs and the arts, are in poor English because they have never been revised by an English-speaking editor. And some additions are made by children or others who haven't learned to spell and punctuate.

As a result, Wikipedia is certainly not a reliable model of good English, nor of the English of native speakers. The other language Wikipedias, too, attract edits from non-native speakers. Wikipedia contains nonsense along-

https://doi.org/10.1017/S0266078407002027 Published online by Cambridge University Press

side the sense; it contains propaganda and error alongside the facts. It is fiercely up to date, except when it isn't. Wikipedia is an encyclopedia for the world as it is. It seems likely that it will continue to be the encyclopedia that the world deserves. ■

NOTE The etymology of *wiki*: The first wiki engine was given this name by Ward Cunningham, founder of the site WikiWikiWeb, because he remembered that at Honolulu Airport he had been told to catch the bus that runs *wikiwiki* between the terminals. *Wikiwiki* means 'fast' in Hawaiian.

[*Ed*: Probably Hawaiian Pidgin English, not the Hawaiian language as such: an adaptation of 'quick quick'. It would be good to have some clarification on this point.].

Notes

1 Links and statistics for all 713 Wikimedia projects can be found at: <http://s23.org/wikistats/wikimedias_html.php>.
2 'Lots of people have their first wiki experience here,' according to: <http://c2.com/cgi/wiki?WelcomeVisitors>.
3 For a list of those that use the same engine as Wikipedia, see: <http://www.mediawiki.org/wiki/Sites_using_MediaWiki>. Many are obscure and moribund, but one of the more active examples is the linguistics wiki at <http://home.unilang.org>. For a larger list of wikis, using different wiki engines, see <http://www.wikiindex.com/Welcome>.
4 'Before it ceased operating [in September 2003], Nupedia produced 24 articles that completed its review process ... and 74 more articles were in progress' (http://en.wikipedia.org/wiki/Nupedia).
5 Jim Giles, 'Internet encyclopaedias go head to head' <http://www.nature.com/news/2005/051212/full/438900a.html>.
The Encyclopaedia Britannica's publishers pointed to serious flaws in the survey method.
6 'Wales also talked about creating "stable" or "static" pages for entries that are considered complete': James Niccolai, 'Wikipedia taking on the vandals in Germany' <http:// www. pcadvisor.co.uk/news/index.cfm?newsid= 7177>.
7 With nearly 50,000 articles and over 4,500 registered users, the Catalan Wikipedia is much larger and busier than any of the others in this table
8 Separate from the Serbian, Croatian and Bosnian Wikipedias.
9 The strong showing of the Netherlands may be evidence of a hidden problem in collecting these statistics – or it may mean that the Dutch are even more multilingual than anyone realised! The weak showing of China (excepting Hong Kong) is because access to the Chinese Wikipedia is generally blocked in mainland China, though the block can be subverted by those who know how.

https://doi.org/10.1017/S0266078407002027 Published online by Cambridge University Press