

Wikipedia Dynamics

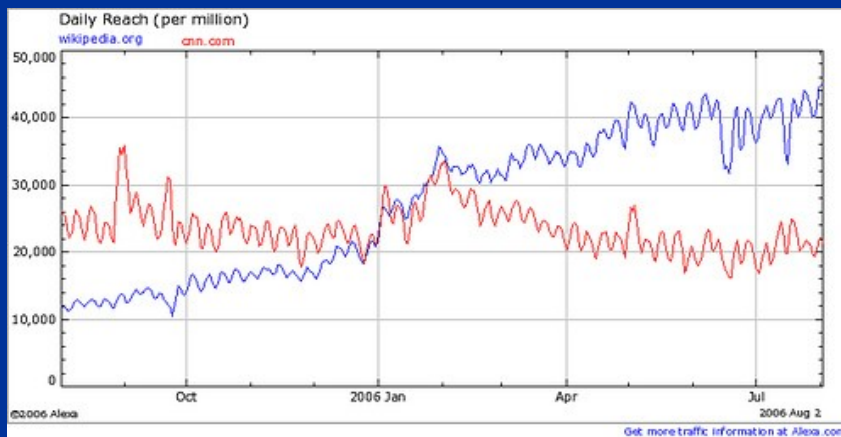
- with open problems

Mingli Yuan

August 27, 2006

Jimmy Wales at Wikimania 2006

- According to Alexa data, Wikipedia passed CNN on page views.
- *Nature* article on Wikipedia and Britannica.
- **One Laptop Per Child** project will integrate Wikipedia content.



- For English Wikipedia, turn our attention away from growth to quality.
- Image: limiting fair use and encouraging free images
- Stable versions

Overview

- Wikipedia: mission and characteristics
- Criticisms on Wikipedia
- Wikipedia: a dynamic system
- Growth of Wikipedia
- Featured articles and good articles
- Stub percentages
- Words and Revisions per Article
- Articles per Wikipedian
- Wikipedian
- How do Wikipedians create featured articles?
- My personal conclusions
- References

Wikipedia: mission and characteristics

■ Mission

- “Imagine a world in which every single person is given free access to the sum of all human knowledge.” – by Jimmy Wales [1]

■ Characteristics [2]

- free to redistribute and reproduce
- constant and plentiful updates
- comprehensive, diverse coverage
- absent of advertisement
- versions in numerous languages

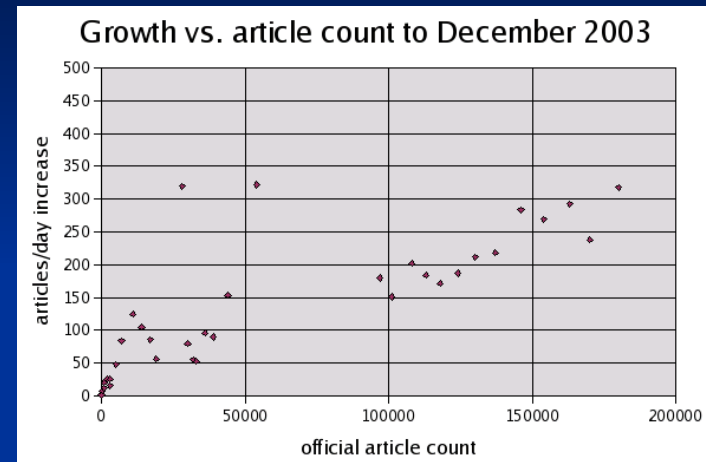
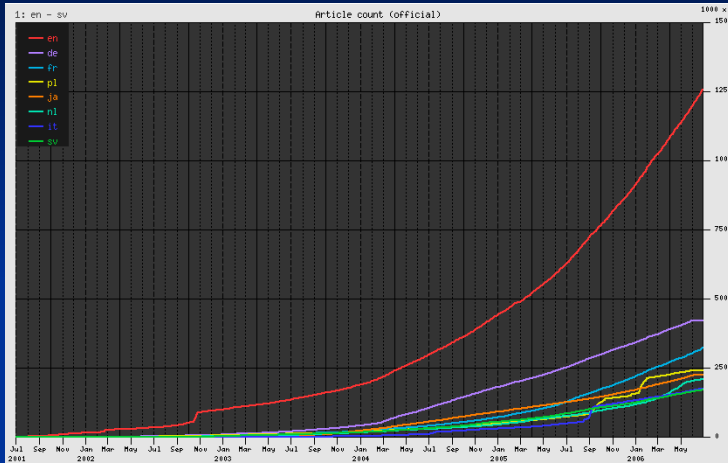
Criticisms on Wikipedia

- Criticisms [2]
 - susceptibility to vandalism
 - uneven quality
 - inconsistency
 - systemic bias
 - preference for consensus or popularity over credentials

Wikipedia: a dynamic system

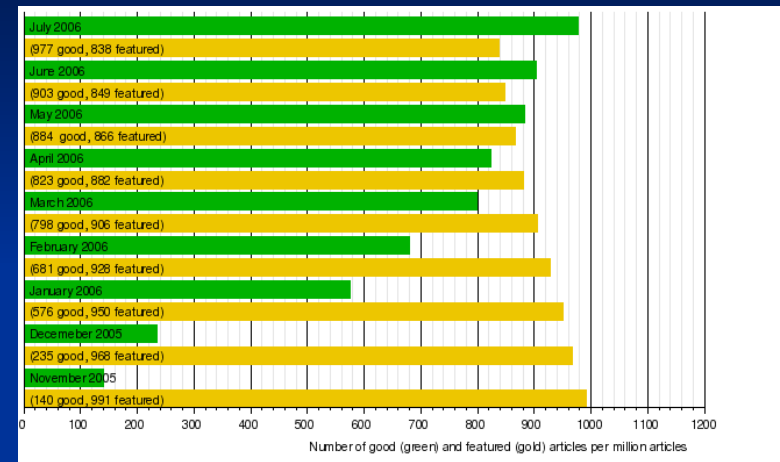
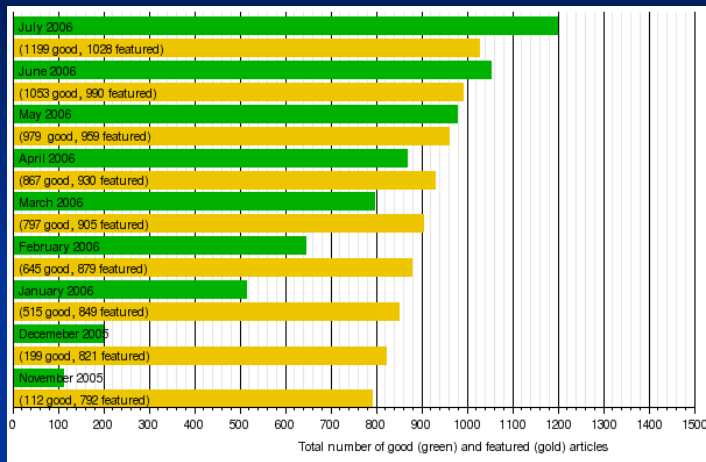
- Wikipedia: a complex social phenomena
- Problems:
 - How will Wikipedia evolve?
 - Is it feasible to establish a **high quality** knowledge repository in the Wikipedia way?

Growth of Wikipedia



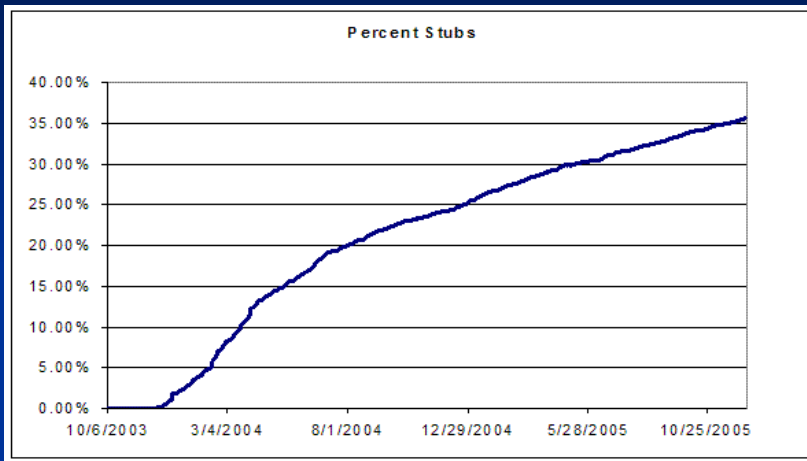
- Quasi-exponential growth pattern
- 2003 Model: using 2003 data to setup a model
- 2003 Model underestimated the actual growth
- Problem: how long will the exponential growth continue, or is it in reality merely the early phase of a logistic curve? [3]

Featured articles and good articles



- Statistics on featured articles and good articles [4]
 - The number of good articles has been rising faster than the number of featured articles.
 - The proportion of good articles is rising.
 - The proportion of featured articles is declining.
- Problems:
 - Will the proportion of good articles rise constantly?
 - Will the proportion of featured articles drop to near zero?

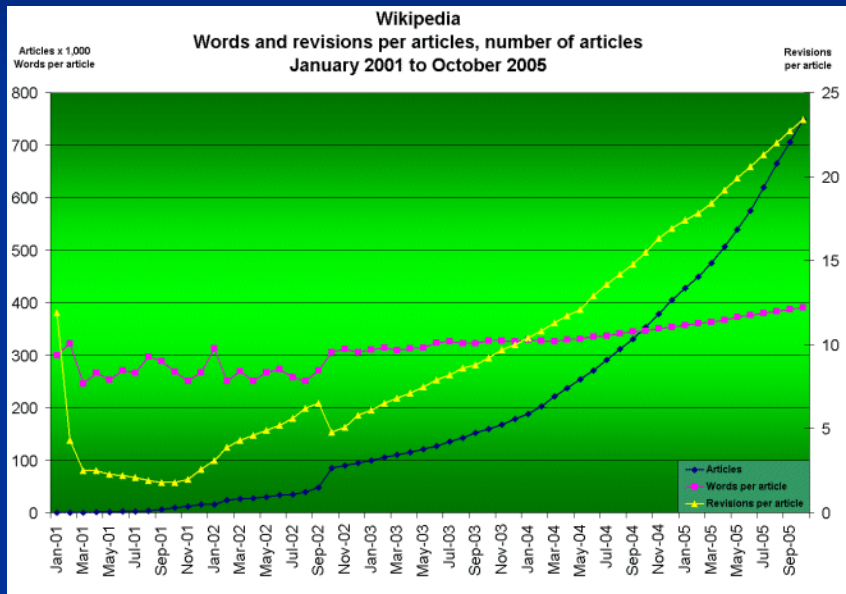
Stub percentages



- Stubs are Wikipedia entries that have not yet received substantial attention from the editors of Wikipedia, and do not yet contain sufficient information on their subject matter. [5]

- Statistical result: [6]
 - Stubs still comprise an increasingly large percentage of the articles on Wikipedia.
 - The rate of stub increase is slowing.
- The need to focus on stub expansion, not just article creation. [6]
- Problem: Does the stub percentage increase for ever?

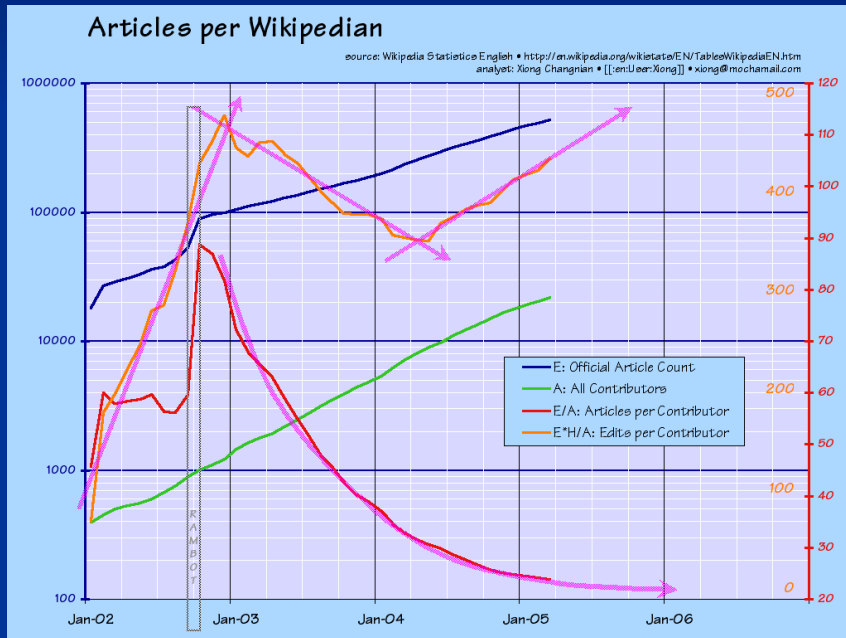
Words and revisions per Article



- the average number of words per page will increase at no more than a snail's pace. [7]
- it is even possible that the gradient might flatten or fall slightly if the rate of new stub addition eclipses the rate of expansion of existing articles. [7]

Articles per Wikipedian

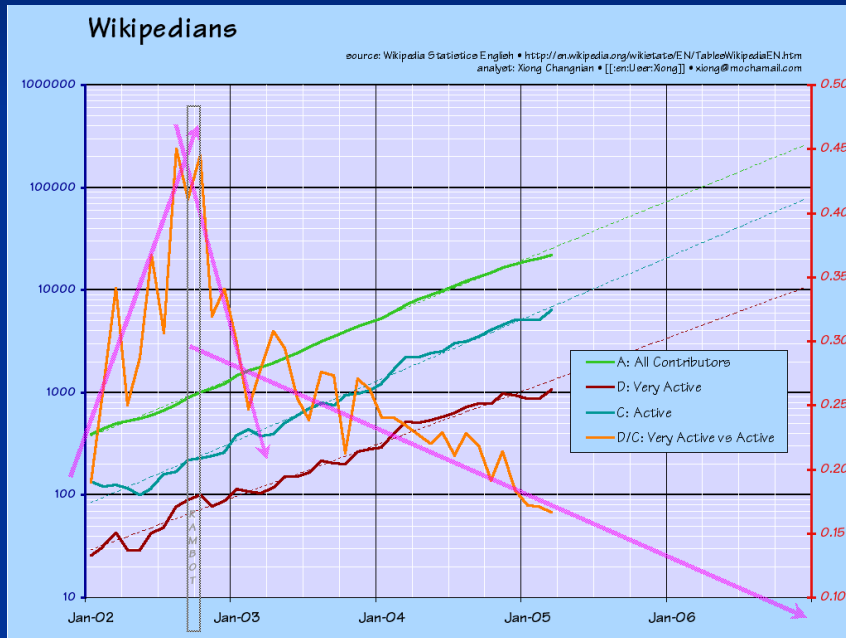
- Statistical result: [8]



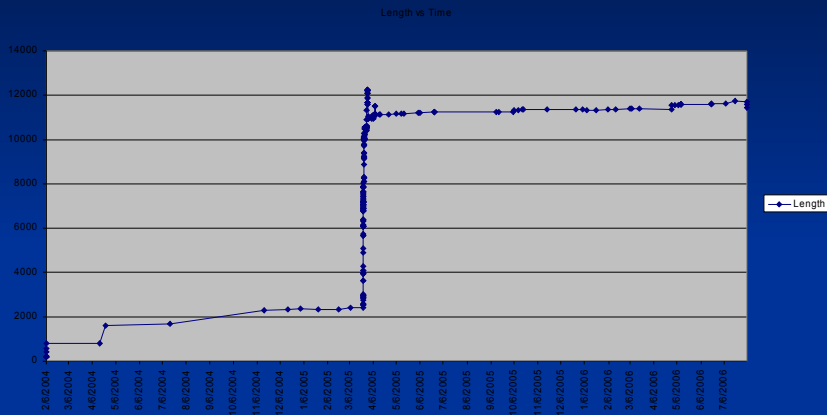
- Articles per Wikipedian is decaying along some function that appears either logarithmic or polynomial.
- It's uncertain whether this will level off somewhere above 20 articles per contributor.

Wikipedian

- Statistical result: [8]
 - Community membership has, like article count, been growing exponentially.
 - The trend of very active vs. active is dropping.



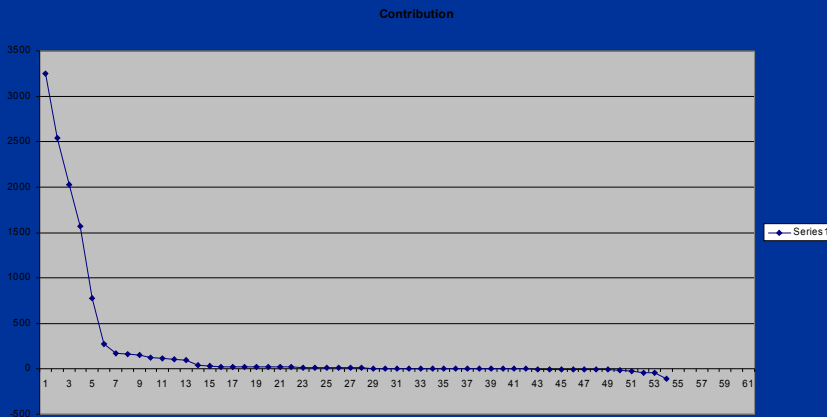
How do Wikipedians create featured articles?



- Leaps in Length vs. Time graph.

- Few Wikipedians contribute most of the content of the article.

- A pro-Am game



My personal conclusions

- The Wikipedia project is still in its early stage with exponential growth in contributors and articles but such growth will eventually decline to a constant rate.
- Articles per Wikipedian and edits per Wikipedian will be constant.
- The numbers of very active users and active users will approach certain constants, while the ordinary users will grow linearly.
- Feature articles, good articles and stub will grow linearly, but stub will grow faster than the other two; the stub percentage will approach a constant in the future.
- As a long-term project, Wikipedia may stand for hundreds of years, and become another Britannica in the Internet era.

References

- [1] Jimmy Wales Kicks off Wikimania
http://ross.typepad.com/blog/2006/08/jimmy_wales_kic.html
- [2] Wikipedia <http://en.wikipedia.org/wiki/Wikipedia>
- [3] Wikipedia: Modeling Wikipedia's growth
http://en.wikipedia.org/wiki/Wikipedia:Modelling_Wikipedia%27s_gr
- [4] Wikipedia: Good articles/Statistics
http://en.wikipedia.org/wiki/Wikipedia:Good_articles/Statistics
- [5] Wikipedia: Stub <http://en.wikipedia.org/wiki/WP:Stub>
- [6] User: Dantheox/Stub percentages
http://en.wikipedia.org/wiki/User:Dantheox/Stub_percentages
- [7] Wikipedia: Words per article
http://en.wikipedia.org/wiki/Wikipedia:Words_per_article
- [8] Wikipedia: Xiong's stats
http://en.wikipedia.org/wiki/Wikipedia:Xiong%27s_stats

Thanks!