

RULES AND DIVERSITY OF JAPANESE WIKIPEDIA'S GROWTH

Y. Yamazaki^a, S.Itoh^b, T.Iba^c and K. Kumasaka^d

^aGraduate School of Media and Governance, Keio University,
Endo 5322, Fujisawa, Kanagawa, Zip.252-8520, Japan
yuca@sfc.keio.ac.jp

^bGraduate School of Media and Governance, Keio University,
Endo 5322, Fujisawa, Kanagawa, Zip.252-8520, Japan
si@sfc.keio.ac.jp

^cFaculty of Policy Management, Keio University,
Endo 5322, Fujisawa, Kanagawa, Zip.252-8520, Japan
iba@sfc.keio.ac.jp

^dFaculty of Environment and Information Studies, Keio University,
Endo 5322, Fujisawa, Kanagawa, Zip.252-8520, Japan
ken@sfc.keio.ac.jp

In this research, we explored rules and diversity of Japanese Wikipedia's Growth. Wikipedia is an encyclopedia on the Web. It was formally launched on January 15th, 2001, as a single English-language edition. On September 1st, 2002, the official version of Japanese Wikipedia was released. And, Wikipedia is becoming bigger and bigger compiled by a number of people. As of July 25th 2008, there are half a million articles. We use all articles of Japanese Wikipedia as data.

We explored the following points; (1) common rules running through the whole data, (2) a huge variety of individual growing ways. To be more precise, we analyzed the annual changes in distributions of the frequencies of the linked words, number of articles and number of editors. We used 391,618 articles in 39,251 categories during the five years from 2004 to 2008 as data.

First, to analyze common rules running through the whole data, we plotted number of linked words, articles and editors in each category into cumulative distribution graph, then calculated regression of coefficients " μ ". Fig.1 and Fig.2 are cumulative distribution graphs about number of articles and

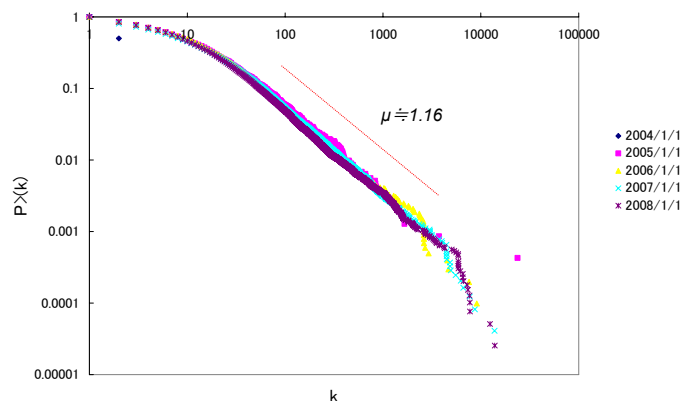


Figure 1 number of articles in each category's graph

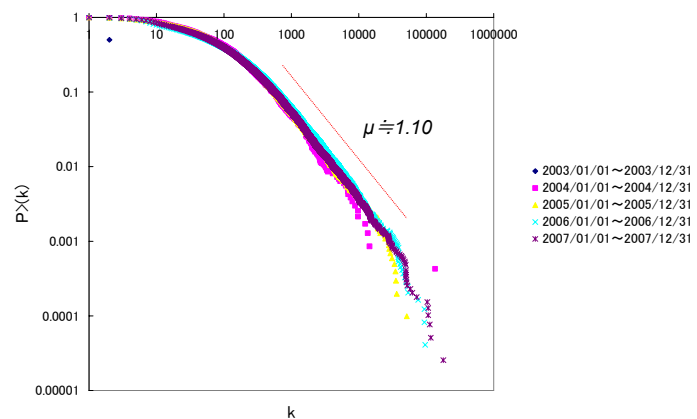


Figure 2 number of editors in each category's graph

editors in each category. Then we found that even if Wikipedia gets bigger and bigger, there are rules that power indexes ($\alpha = \mu + 1$) are almost all the same. There are a wide variety of categories in each year, for example, category in a few articles and editors and category in a great number of articles and editors. And, the component ratio of them are not change.

Secondly, to analyze a huge variety of individual growing ways, we made clusters on Self-Organizing Maps by using the regression coefficient of linked words, number of articles and number of editors as parameters. Self-Organizing Maps is a kind of neural network. By enabling to map multi dimensional data into two dimensional data, it is often used to visualize higher-dimensional space. Analysis, using Self-Organizing Maps lead this research to find diversity in growing patterns. To evaluate varieties of growing patterns, we colored parameters for visualizing. In addition, we find that there are correlations between growing patterns and Wikipedia's categories.

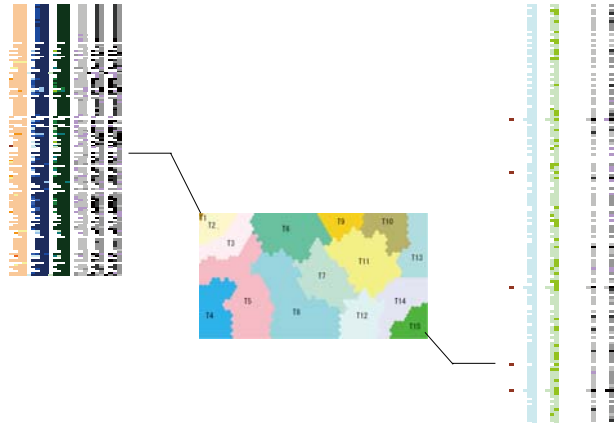


Figure 3 growing ways on SOM

Keywords

time series analysis, power laws, scale-free networks

References

- [1] A. Capocci, V. D. P. Servedio, F. Colaiori, F. S. Buriol, D. Donato, S. Leonardi, and G. Calderelli, "Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia" *Physical Review E*, v. 74, 036116, 2006.
- [2] T.Kohonen "Self Organizing Maps" Springer Verlag, 2001.
- [3] Y.Yamazaki, T.Ito, T.Iba, and K.Kumasaka, "Diversities in Wikipedia's growing patterns", *Fourth Joint Japan-North America Mathematical Sociology Conference*, CA, U.S., May 2008.
- [4] Y.Yamazaki, T.Ito, T.Iba, and K.Kumasaka, "Growing Patterns of Wikipedia' s Networks: Analyzing Laws and Diversity in Japanese Wikipedia", *poster, International Workshop and Conference on Network Science '08*, Norwich, UK, June 2008.
- [5] <http://ja.wikipedia.org/>