Growing Patterns of Wikipedia' s Networks Analyzing Laws and Diversity in Japanese Wikipedia

Yuka Yamazaki,^{*1} Takaichi Ito,^{*1} Takashi Iba,^{*2} and Kenji Kumasaka^{*3}

*1 Graduate School of Media and Governance, Keio University *2 Faculty of Policy Management, Keio University *3 Faculty of Environment and Information Studies, Keio University

Abstract

In this research, we explored rules and diversities in Wikipedia' s growing patterns. Wikipedia is an encyclopedia on the Web. And, Wikipedia is becoming bigger and bigger compiled by a number of people.First, we analyzed the distributions of the frequencies of the linked words in several categories. The analysis gave a result that the distributions are raising to power law. But, the differences between the processes of growth still remain unclear.

In this research, we explored the following points; (i) common rules running through the whole categories, (ii) a huge variety of individual growing patterns. To be more precise, we analyzed the annual changes in distributions of the frequencies of the linked words in Japanese Wikipedia articles. We used 19,425 articles in 368 categories during the five years from 2004 to 2008 as data. Then, we plotted the data into double logarithm graphs, and calculated coefficient of regression to formulate entire trends. Additionally, by using the regression coefficient as parameters, we made clusters on Self-Organizing Maps. Analysis, using Self-Organizing Maps lead this research to find diversities in growing patterns of link words' distributions. In addition, we find that there are correlations between growing patterns and Wikipedia' s categories.

Findings

Laws - common rules running through the whole categories

Plotting frequencies of hyper-links on double logarithmic graph, we found they show power law distribution. In most of categories, there are power law in time-series. And, as years passes by, each inclination becomes gradually steep in any of the categories.

Diversity - variety of individual growing patterns

By using Self-Organizing Maps, the location of categories that have similar power index in time-series, is obtained as a solution.

Data

We use title of articles and hyper-links inside the articles from Japanese Wikipedia as data. The biggest letters mean article, and blue letters show hyper-links. We analyze 352 categories under the MAIN CATEGORIES. To see changes in time-series, we counts the frequency of hyper-links' appearance on 2004 - 2008 January 1st.



Method Plotting

1010

Plotting frequency of

hyper-links' appearance

to double logarithm graph.



Clustering

Making clusters using by Self-Organizing Maps, kind of neural networks of unsupervised learning.



To analyze, we colored power index and the increasing rate of articles, and calculated specialization coefficient.



