

# **Growing Patterns of Wikipedia's Networks**

## **- Analyzing Laws and Diversity in Japanese Wikipedia -**

Y. Yamazaki<sup>1</sup>, T. Ito<sup>1</sup>, T. Iba<sup>2</sup> and K. Kumasaka<sup>3</sup>

<sup>1</sup>Graduate School of Media and Governance, Keio University, Endo 5322, Fujisawa, Kanagawa, Zip.252-8520, Japan

<sup>2</sup>Faculty of Policy Management, Keio University, Endo 5322, Fujisawa, Kanagawa, Zip.252-8520, Japan

<sup>3</sup>Faculty of Environment and Information Studies, Keio University, Endo 5322, Fujisawa, Kanagawa, Zip.252-8520, Japan

**Theme:** Economy, Communications and Management

### **Introduction**

In this research, we explored rules and diversities in Wikipedia's growing patterns. Wikipedia is an encyclopedia on the Web. Wikipedia was formally launched on January 15, 2001, as a single English-language edition. Multi-lingual version was launched on May 20th, 2001. Japanese version was introduced, as well as the other languages, but it did not have the function to edit with Japanese letters. On September 1st, 2002, the official version of Japanese Wikipedia was released. Like ordinary dictionaries made from paper, articles are classified by categories. 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.

### **Results**

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. 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 diversities in growing patterns of link words' distributions. In addition, we find that there are correlations between growing patterns and Wikipedia's categories.