

Instructor Name	Masashi Toyoda, associate professor	Laboratory Location	Institute of Industrial Science	Research Area	Web engineering
-----------------	-------------------------------------	---------------------	---------------------------------	---------------	-----------------

Large-scale Web analysis to explain social phenomena

The Web has a complex network structure where a large collection of documents, said to include one trillion or more pieces, is linked together. The structure is constantly changing on a daily basis through the creation, updating and deleting of documents. Recently, changes on the Web are increasingly likely to reflect phenomena in the real world.

In the Toyoda Lab, we are studying methods to acquire and analyze a huge amount of Web information and extract useful information based on the content of documents, link structures and their dynamic changes, thereby explaining social phenomena by leveraging visualization technology, etc. We are looking forward to the participation of students who are interested in new Web analysis techniques and the development of applications that are more sophisticated than mere search engines.

Main fields of study

Analysis of link structure of the Web:

On the Web, related pages tend to be tightly linked with one another by hyperlinks and a group of related pages known as a "Web community" can be extracted by using this characteristic. In our research, we are trying to develop a method to extract all the main communities from large-scale Web archives and create a map (Fig. 1) on which communities related to one another are connected with lines. The map provides information similar to an industry relational diagram and indicates how people recognize and categorize Web pages.

We also try to analyze and visualize the structure of Web spam, an attempt to unfairly boost the rank of one's Web site in a search engine (Fig. 2).

Analysis of time-series changes on the Web:

By observing time-series changes on the Web, it is possible to determine how new topics emerge and evolve. With the use of Web archives that are continuously collected, we are trying to develop techniques to track and visualize the evolution process of topics on the Web (Fig. 3). With these techniques, we have made it possible to determine changes in various topics from the beginning and identify influencers.

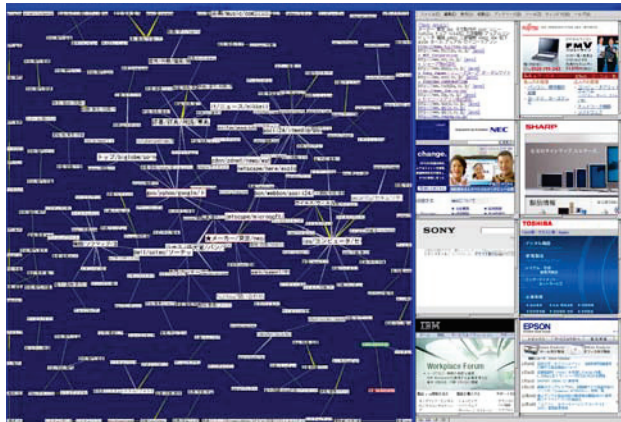


Fig. 1: Brower for Web community map

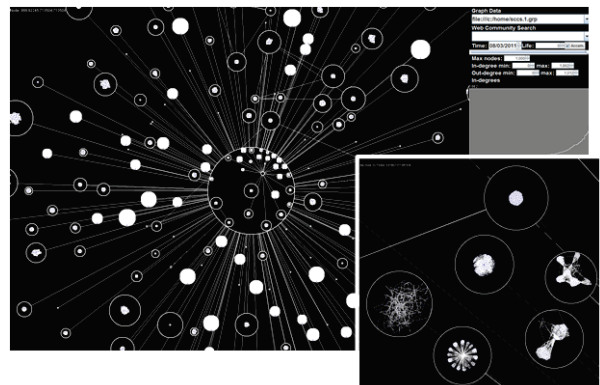


Fig. 2: Structure of search engine spam

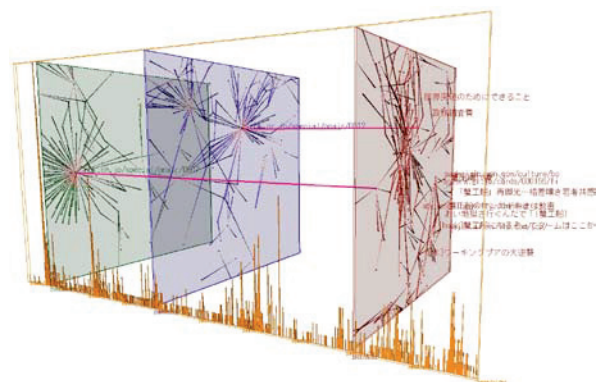


Fig. 4: Time-spatial analysis of topics on the Web

E-mail: toyoda@iis.u-tokyo.ac.jp
 HP: <http://www.tkl.iis.u-tokyo.ac.jp/~toyoda/>