

# 6X-01 Naviz: User Behavior Visualization System using Web Access Log

Bowo Prasetyo<sup>1</sup>, Iko Pramudiono<sup>1</sup>, Katsumi Takahashi<sup>2</sup>, Masashi Toyoda<sup>1</sup>, Masaru Kitsuregawa<sup>1</sup>  
<sup>1</sup>Institute of Industrial Science, University of Tokyo

## 1. Introduction

Navigational behavior of website visitors can be extracted from web access log files with data mining techniques such as sequential pattern mining[1]. Visualization of the discovered patterns is very helpful to understand how visitors navigate over the various pages on the site. Currently several web log visualization tools have been developed. However those tools are far from satisfactory. They do not provide global view of visitor access as well as individual traversal path effectively.

Here we introduce *Naviz*, a system of interactive web log visualization that is designed to overcome those drawbacks. It combines two-dimensional graph of visitor access traversals that considers appropriate web traversal properties and facilities for filtering traversal paths. The tool also provides support for modern dynamic web pages. We apply the tool to visualize results of data mining study on web log data of Mobile Townpage, a directory service of phone numbers in Japan for i-Mode mobile internet users. The results indicate that our system can effectively visualize the discovered patterns to identify interesting navigational behavior such as success paths, exit paths and lost paths.

## 2. Traversal Diagram

Two of the major problems with current web log analysis are difficulty to understand *what* visitors are trying to do on a website and *how* they are doing it. Web log visualization firstly should give understanding of global view of visitor access. By utilizing Graphviz's[2] drawing power, we found that it is very helpful to introduce two *appropriate web traversal properties* below when visualizing web log access:

1. Hierarchical structure regarding traversal traffic, i.e. more traversed edges are placed at the higher level and less traversed edges at the lower level position.
2. Grouping of related pages, for example, pages that have high degree of transitional probability among them are better to be placed together near to each other.

Another important factor regarding web log visualization is that today HTML pages are often dynamic pages that automatically created and destroyed whenever needed. A dynamic page can be thought of as a page class, which will create different instances for every access depending on input parameters. This kind of page requires different visualization technique compared with traditional static HTML pages. Class-instance view technique, i.e. a technique to visualize dynamic page either as a node of representative class or as many nodes of its instances, is one of the solutions for this problem.

To represent navigational behavior we use "traversal diagram", i.e. 2 dimensional directed graph which

represent pages as nodes and traversals as edges, using traversal path, a sequence of page classes traversed by visitors.

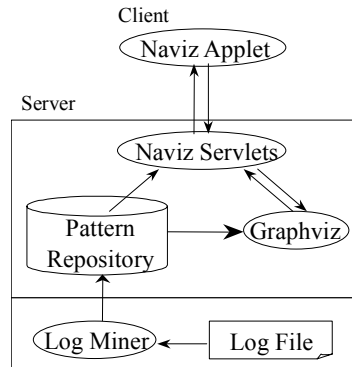


Fig. 1. Naviz Architecture

## 3. System Overview

As showed in Fig. 1 Naviz is a three-tier client-server application, consists of:

1. A java applet as user interface in the client side.
2. Java servlets in the server side, interface to allows communication between client and server.
3. Server-side applications include:

- (♯) Pattern repository containing discovered patterns.
- (♯) Graphviz that is responsible for drawing the graph layout according appropriate web traversal properties.
- (♯) Log miner to discover traversals/paths from web log data.

## 4. Features

Naviz has two operation modes, i.e. traversal diagram mode and traversal path mode. In traversal diagram mode Naviz displays traversal diagram that describes global visitor traversals on the entire site. In traversal path mode, Naviz displays traversal path on top of traversal diagram, in such a manner that only one path is showed at one time. Only part of paths of interests can be shown by specifying the pages that are visited and optionally by the number of hops required to reach those pages.

Naviz uses several strategies to form effective view of traversal diagram and traversal paths:

1. Consideration of appropriate web traversal properties.
2. Class-instance view as explained above.
3. Navigational behavior comparison: compare the visitor behavior of the same website regarding the different attribute, i.e. day behavior vs. night behavior.
4. Traversal path filtering.
  - (♯) The number of displayed paths will decrease/increase correspondingly as we specify the number of hops.
  - (♯) We can choose which pages are visited, and find the paths of visitors that traverse over those pages.
5. Interactive environment: layout of the traversal diagram, strategy of hierarchization, grouping, and filtering all can be changed interactively to select the best structure representing the web log data.

<sup>2</sup>NTT Information Sharing Platform Laboratories

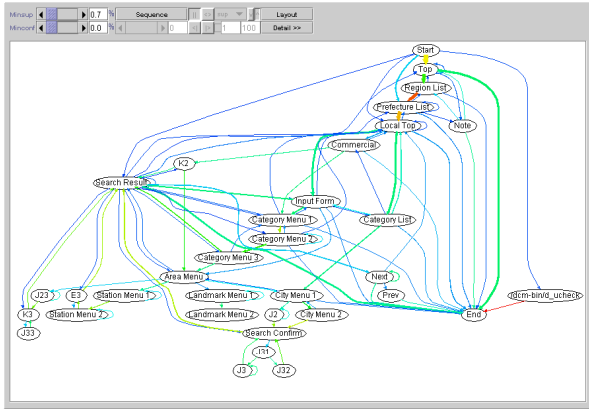


Fig. 2. Traversal Diagram

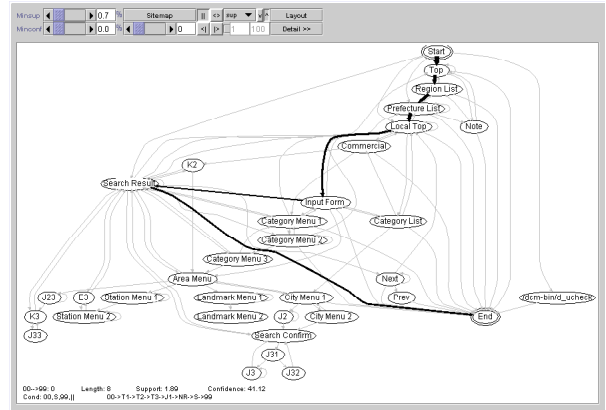


Fig. 3. Visitor's Success Path

### 5. Visualization of NTT i-Townpage served on i-Mode

Townpage is the name of "yellow pages", a directory service of phone numbers in Japan. It consists 11 million listings under 2000 categories. Data mining are performed on the web log data of Mobile Townpage site from 1 to 7 May 2000 with size of 15 GB, using minimum support threshold of 0.1%, which resulted in a set of traversal paths containing 1116 traversal between two pages and 4595 traversal paths among many pages (greater than two). Fig. 2 is the visualization result of traversal path between two pages by Naviz on traversal diagram mode, it gives the global view of visitor traversal on the entire site of Mobile Townpage.

In traversal diagram mode, Naviz considers appropriate web traversal properties to form view, so that we can think of visitors came into the site from top of the graph and went out from the bottom. We can see the most traversed edges, the thick ones, that are connecting pages "Top"→"Region List"→"Prefecture List"→"Local Top" etc. are placed in the upper position of the graph. Naviz also allows us to find related pages easily. As we can see there are several groups of related pages, which indicates that those pages have high degree of transition probability among them: i.e. group ("Top"→"Region List"→"Prefecture List"→"Local

Top"→"Commercial") etc.

Switching to traversal path mode, using Naviz we discovered some interesting user behavior on Mobile Townpage site. First, we try to search paths in which visitors are successful finding their goal, as the result Fig. 3 shows one of the success paths: i.e. 1.89 % of visitors successfully reached the "Search Result" page through the path "Top"→"Region List" →"Prefecture List"→"Local Top"→"Input Form"→"Search Result" and then exit from the site. Using comparison function, we found interesting differences of day behavior and night behavior of i-mode users. One of the differences is the second most visitor success path as showed in Fig. 4 and Fig. 5. With all of this knowledge website designers should be able to improve their site.

### 6. Conclusion and future works

We concluded that the important factors in visualizing the visitor navigational behavior from web log data is to consider appropriate web traversal properties as well as modern dynamic page, and utilize interactive environment to provide greater capability to analyze the discovered traversal paths.

In the future we plan to improve Naviz with capability of interactively communicate with log miner such that Naviz can control log miner about what data is to be mined, by what threshold it should be mined, etc.

### Acknowledgment

We would like to express our gratitude to NTT Directory Services Co. for providing us with web access log data of NTT i-Townpage served on i-Mode website.

### References

- [1] R. Srikant, and R. Agrawal. Mining Sequential Patterns: Generalizations and Performance Improvements. In Fifth Int'l Conference on Extending Database Technology (EDBT'96), pages 3-17, Avignon, France, March 1996.
- [2] <http://www.research.att.com/sw/tools/graphviz/>

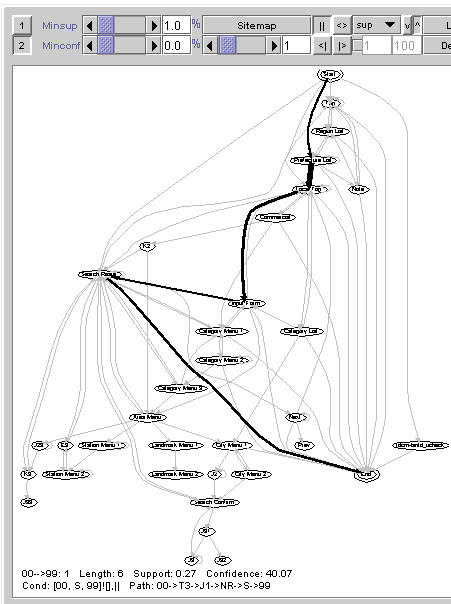


Fig. 4. Day Success Path

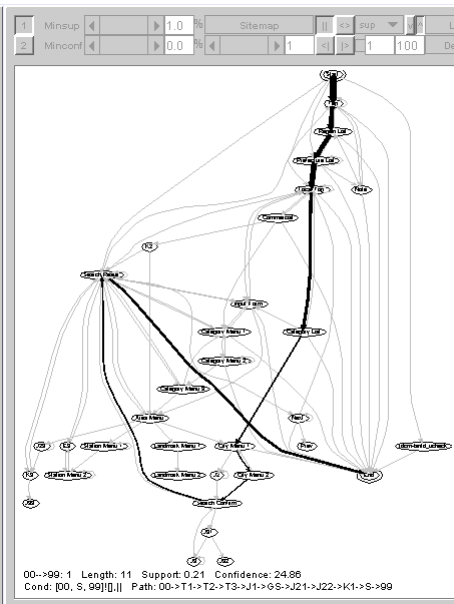


Fig. 5. Night Success Path