

Free-link Topology Navigation on Statistical Table Objects: Metadata Schema and User Interface

Xiangming Mu and Gary Marchionini
Interaction Design Lab, School of Information and Library Science
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-3360
Phone: (919)962-8366; Fax: (919)962-8071
Email: {mux, march}@ils.unc.edu

Abstract

A table object, defined and described in XML, is treated as the basic unit in a free-link topology that enables multiple inheritances for each node. In this paper, a novel table navigation system called TableHunter is introduced. Each table object is presented in the TableHunter as a node of a map structure, and supports a context+focus view.

1. Introduction

Statistical data is an important type of information that is essential to the day-to-day lives of citizens. The table is the common format to represent such statistical data. An increasing amount of statistical tables is now moving to the WWW due to the fast development of the Internet. For example, in the FedStats (www.fedstats.gov) website, statistical information including tables, surveys, reports, and analysis from more than 100 federal agencies is available for public access. A novel free-link topology structure (FTS) is proposed to facilitate table searching tasks. Instead of classifying various “nodes” as roots, “hubs”, or “leaves” and forming a hierarchy “tree” structure, the FTS treats each node as an equal “table object” and connects each related table through links. “It is a web not a tree”. In this structure, each “table object” is self-contained and self-explainable: it maintains not only the raw table information, including the table itself, related webpages, documents, or graphics, but also “neighbor table objects” in terms of links. The advantage of adopting table objects is that the data is de-coupled from its presentation format and thus flexibility is increased. An application level system called TableHunter was developed based on the FTS. Part of the EIA hierarchy table index system was analyzed and implemented in the TableHunter prototype. Each table object is represented as a node and links are represented as lines between nodes. TableHunter provides a means for fast “trial-and-error” operations and thus speeds table hunting tasks.

2. Related work

Considerable research has been conducted to help users understand online tables. BEST (Marchionini & Mu, 2002), formerly called the table browser, provides frozen column and row headings in viewing large online tables. Metadata is integrated and presented with tooltips, submenus, and active links to aid table comprehension. FOCUS (Spence, Beilken, and Berlage, 1996) uses a focus + context technique to permit viewing a large table without scrollbars. Table Lens (Rao and Card, 1994) suggests a number of user operations to view and sort tables. Yang et al. (Yang and Luk, 2002) proposed a framework for web table mining using HTML tags and attribute-value pairs to extract online tables.

Another related research domain is web page navigation and visualization using a “map” structure. Durand et al. (Durand and Kahn, 1998) suggested that maps are the most basic visual orientation tool and present ten requirements for a map structure interface to guide the MAPA design. InXight

(www.inxight.com) is a commercial tool that visualizes the web hierarchy in the form of a “tree” structure. Similar technology was presented in other research (Robert and Lecolinet, 1998; Lamping and Rao, 1996) and has been referred to as Focus + Context technology (Mukherjear. And Hara, 1997)

However, the visualized map structure for web nodes is completely built on a hierarchy tree structure, which limits its applications. Limited visualization research has been conducted on the more natural free-link topology. In the above approaches, the basic unit for the map structure is a web page. Tables can only “belong to” a web page and cannot be identified independently. In this paper, we separate tables from the web pages and treat tables as well as webpages as the basic unit of our free-link topology. Each of these basic units is defined as an object. A metadata schema is suggested to “wrap” the table object and to add “link” attributes.

3. Table Object and XML Schema

A table is defined as an object having the following attributes:

- Table description attributes, such as title, author, date, size, format (PDF, HTML or text), number of rows, fields, and data types.
- Table relational attributes: number of links, link URLs, and link Titles.
- Table contextual attributes: web page URLs

A structured metadata schema is employed in this paper in terms of the XML schema definition. “Structured metadata accompanying statistical tables from different NSI’s can be compared and their level of harmonization assessed automatically” (Petrakos, et al. 2001). The following is a sample schema.

```
<Program xmlns = http://www.ils.unc.edu/Table/ExampleSchema>
<DescriptionInfo>
  <Title> ... </Title>
  <Author> ... </Author>
  <Date format=TableDate> ... </Date>
  <Size > ... </Size>
  <Format value=“PDF, HTML, TEXT”> ... </Format>
  <RowNumber format=TableNumber> ... </RowNumber>
  <DataType value=“TableString, TableNumber”> ... </DataType>
  <Field format=“TableString” > ... </Field>
  <Field format=“TableString” > ... </Field>
  ...
</DiscriptionInfo>
<RelationInfo>
  <LinkNumber format=“TableNumber”> ... </LinkNumber>
  <Link>
    <LinkURL format=“LinkString”> ... </LinkURL>
    <LinkTitle format=“TableString”> ... </LinkTitle>
  </Link>
  ...
</RelationInfo>
<ContextInfo>
  <PageNumber format=“TableNumber”> ... </PageNumber>
  <Page>
    <PageURL format=“LinkString”> ... </PageURL>
    <PageType value=“Http,Ftp,Email”>.. </PageType>
  </Page>

```


The default TableHunter presents three levels of nodes: the current node is located in the center and represented in red. The direct links revealed from this current “table object” are treated as neighbors and represented in pink. The links from the neighbors are treated as a third level and represented in yellow. Users are allowed to present more than three levels by making selections from the drop-list entitled “Range” (fig.2) An alphabetic list of all available nodes is also available in the “Title” drop-list, which enables users to bring that node into the center of the map with a single click. The “Home” button will reset the map to the default map view. In the sample case, the default page is the home page of the EIA.

5. Discussion and Future Research

The prototype of the TableHunter was demonstrated for faculty members and graduate students in the School of Information and Library Science. Revisions were made and the revised version of the TableHunter was then informally tested by some students. Students conducted table searching tasks and then participated in a discussion about their experience with TableHunter. From our observations and the talks, we found that users were excited by the interface and its interactive functions. All the students claimed that it was easy to learn and use. Some students mentioned they preferred this way of displaying many nodes in a very small space and one student commented, “[TableHunter] made me quickly understand the relationships among these concepts and helped me find the right track for the table”. Compared to the homepage of EIA which can only present eight links, the default interface (three levels) of the TableHunter for EIA presents 53 nodes.

TableHunter is implemented in Java and presented as a Java applet, which means it is can be embedded into the web browser directly for access. TableHunter may also be integrated with other online table browsing tools such as the BEST (Marchionini & Mu, 2002). By adopting the current web structures and extract links, TableHunter does not change the web structure. The TableHunter will be demonstrated at the conference. As a prototype for visualizing table objects and their links, TableHunter needs formal evaluations in terms of both usability and system performance. Current research under consideration involves developing provisional means that would automatically authorize table objects from the database schema.

Acknowledgements

This work is supported by NSF Grant# EIA 0131824. Thanks to Amy Pattee for the help in the paper revision

References:

1. Durand, D., and Kahn, P., 1998, “*MAPA: a system for inducing and visualizing hierarchy in websites*”. Proceeding of the HyperText 98.
2. Marchionini, G. & Mu, X., (in press) “*User Studies Informing E-Table Interfaces*”, Information Processing & management. 2002
3. Petrakos, M., Farmakis, G., Koumanakos, G., Bisiela, P. "A structured approach to a statistical metadata base", NTTS 2001 International Conference in Official Statistics, Creta 2001.
4. Rao, R. and Card, S., “*The table lens: merging graphical and symbolic representations in an interactive focus+context visualization for tabular information*”, proceedings of CHI’94, pp.318-322.
5. Spenke, M., Beilken, C., and Berlage, T., “*Focus: the interactive table for product comparison and selection*”, proceedings of UIST’96, pp. 41-50, 1996
6. Yang, Y. and Luk, W.S., “*A framework for web table mining*”, WIDM’02, November 8, 2002.