

Schema Mapper: A Visualization Tool for DL Integration

Ananth Raghavan¹, Divya Rangarajan¹, Rao Shen¹, Marcos André Gonçalves², Naga Srinivas Vemuri¹, Weiguo Fan¹, Edward A. Fox¹

¹Digital Library Research Laboratory
Virginia Tech
Blacksburg, Virginia 24061 USA

²Department of Computer Science
Federal University of Minas Gerais
Belo Horizonte – MB Brazil 31270-901

{ananthr, divyar, rshen, mgoncalv, nvemuri, wfan, fox}@vt.edu

ABSTRACT

Schema mapping is a challenging problem. It has come to the fore in recent years; there are important applications like database schema integration and, more recently, digital library merging of heterogeneous data. Previous studies have approached the schema mapping process either from algorithmic or visualization perspectives, with few integrating both. With Schema Mapper we demonstrate a semi-automatic tool for schema integration that combines a novel visual interface with an algorithm-based recommendation engine. Schemas are visualized as hyperbolic trees (see Fig. 1), thus allowing more schema nodes to be displayed at one time. Matches to selections are recommended to the user, which makes the mapping operation easier and faster.

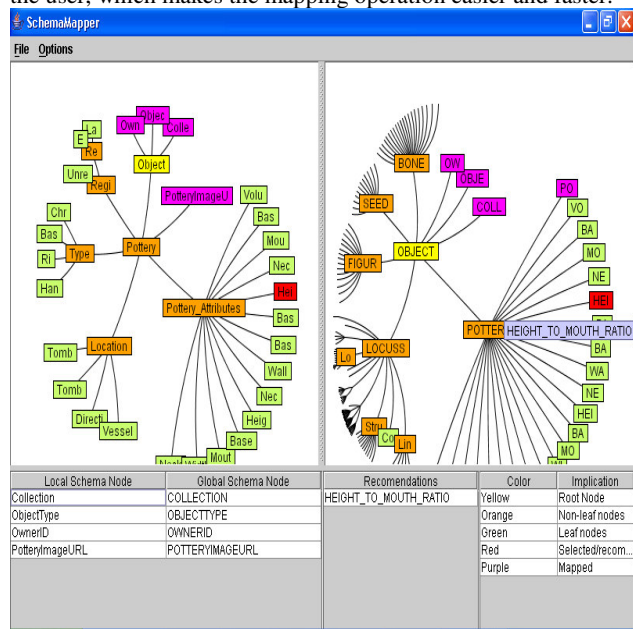


Figure 1. Schema Mapper

Schema Mapper allows editing the target schema, which includes deleting or renaming a node, or adding a sub-tree of the source schema as a child of a node in the target schema. The user has the option to undo mappings, or can choose to ignore recommended matches and create personal mappings. Once the user decides to save the mappings from the source to the target schema, an XSLT style sheet with those mappings is created. It can be applied to

files conforming to the source schema to transform them to files conforming to the target schema. Schema Mapper proves useful in domains like archaeology, wherein the global schema evolves as new collections are being integrated into the Archaeological DL. A pilot user study compared simple 1-1 schema mapping using Schema Mapper vs. MapForce (a commercial tool [2]). Initial results were positive [3]. Although we developed this tool as part of the ETANA-DL [4, 5] project, it can be extended to other applications. We will demonstrate the integration of a representative sample of local collections into the ETANA-DL union catalog, to show Schema Mapper's utility for DL integration. Meanwhile, we are working on: 1) recommendation algorithms based on mappings history and 2) usability studies.

Categories and Subject Descriptors

H3.7 [Information Systems]: Information Storage and Retrieval-Digital Libraries, H2.5 [Database Management]: Heterogeneous Databases

General Terms

Algorithms, Design, Experimentation, Human Factors.

Keywords

Schemas, hyperbolic trees, XSLT style sheet.

ACKNOWLEDGMENTS

ETANA-DL is funded in part by the National Science Foundation (ITR-0325579). Marcos Gonçalves was supported by an AOL fellowship and has support from CNPq. We would like to thank Dr. North and the members of the DLRL of Virginia Tech for their insightful comments during the prototyping of the tool.

REFERENCES

- [1] Schema Mapper, 2005. <http://feathers.dlib.vt.edu/~etana/integration>
- [2] Altova. *Mapforce*, 2005. http://www.altova.com/products_mapforce.html
- [3] Schema Mapper User Study Report, 2005. <http://feathers.dlib.vt.edu/~etana/Papers/Report.doc>
- [4] Ravindranathan, U. Prototyping digital libraries handling heterogeneous data sources - An ETANA-DL case study, Masters Thesis. Dept. Comp. Sci., Virginia Tech, 2004, <http://scholar.lib.vt.edu/theses/available/etd-04262004-153555>
- [5] Ravindranathan, U., Shen, R., Gonçalves, M.A., Fan, W., Fox, E. A. and Flanagan, J.W., Prototyping digital libraries handling heterogeneous data sources - The ETANA-DL case study. In Proc. ECDL, 2004, 186-197.

Copyright is held by the author/owner(s).

JCDL '05, June 7-11, 2005, Denver, Colorado, USA

ACM 1-58113-876-8/05/0006.