Using Document Type Definitions to Merge XML Documents

洪崇恩、邱紹豐

E-mail: 9314762@mail.dyu.edu.tw

ABSTRACT

XML provides the flexibility that users can define the structure for their own data. When merging different XML documents regarding to the same data, it becomes difficult due to the lack of fixed structure. In this research, we provide a mechanism to merge different XML document by examining their structures, or their DTD ’ s . By merging DTD ’ s, our system generates a new DTD that validates the original input XML documents and thus provides a new DTD for the merged document.

Keywords : minimum merged structured ; document type definition ; structure merge

Table of Contents

REFERENCES

[1] H. Garcia-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ullman, and Jennifer Widom, “ Integrating and Accessing Heterogeneous Information Sources in TSIMMIS, ” In Proceedings of the AAAI Symposium on Information Gathering, pp. 61-64, 1995.

[2] S. Chawathe, H. Garcia-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ullman, and J. Widom, “ The TSIMMIS Project: Integration of Heterogeneous Information Sources, ” In Proceedings of IPSJ Conference, pp. 7-18, 1994.

[3] P. M. D. Gray , L. Kerschberg , P. J. H. King, A. Poulovassilis, “ Functional Approach to Data Management - Modeling Analyzing and Integrating Heterogeneous Data, ” Springer, 2003.

[4] Y. Papakonstantinou, P. Velikhov, “ Enhancing Semistructured Data Mediators with Document Type Definitions, ” 15th International Conference on Data Engineering , 1999.

[5] Y. Papakonstantinou, H. Garcia-Molina, J. Ullman, “ Medmaker: A Mediation System Based on Declarative Specifications, ” In International Conference on Data Engineering, pp. 132 — 141, 1996.

[6] H. Garcia-Molina , Y. Papakonstantinou , D. Quass , A. Rajaraman , Y. Sagiv , J. Ullman , V. Vassalos , J. Widom, “ The TSIMMIS approach to mediation: Data models and Languages, ” In Journal of Intelligent Information Systems, 1997.

[7] Y. Papakonstantinou, P. and Velikhov, “ Enhancing Semi-structured Data Mediators with Document Type Definitions, ” In IEEE Data Engineering Conf., 1999.

[8] Y. Papakonstantinou, S. Abiteboul, H. Garcia-Molina. "Object Fusion in Mediator Systems," In VLDB Conference, 1996.

[9] Roy Goldman, JenniferWidom, “ Approximate DataGides, ” In Proc. Of the Workshop onQuery Processing for semistructured Data and Non-standard Data Formats, pp. 436-445, 1999.

[10] Andy S. Chiou, Chilan Lin, “ The study of index on semi-structured Data, “ Communications of Institute of Information and Computing Machinery 6, pp. 99-112, 2003.

[11] Hong Su, Harumi Kuno, and Elke Rundensteiner, “ Automating the transformation of XML documents, ” Workshop on Web Information and Data Management (WIDM'01), 2001.

[12] E. Bertino, G. Guerrini, M. Mesiti, I. Rivara, and C. Tavella, “ Measuring the Structural Similarity among XML Documents and DTDs, ” Technical Report DISI-TR-02-02, Dipartimento di Informatica e Scienze dell'Informazione, 2001.
[13] Francois Bry, Dan Olteanu, Sebastian Schaffert, “ Grouping Constructs for Semistructured Data, ” In Proceedings of DEXA , 2001.
[14] Mong Li Lee, Liang Huai Yang, Wynne Hsu, Xia Yang, “ XClust: Clustering XML Schemas for Effective Integration, ” in 11th ACM International Conference on Information and Knowledge Management (CIKM), McLean, Virginia, November 2002.
[15] S. S. Chawathe and H. Garcia-Molina. “ Meaningful Change Detection in Structrued Data, ” In SIGMOD, page 26-37, 1997 [16] S. S. Chawathe and H. Garcia-Molina. “ Change detection in hierarchically structrued information, ” In SIGMOD, page 493-504, 1996 [17] K. Zhang and D. Shasha. “ Simple Fast Alogrithms for the Editing Distance Between Trees and Related Problems. ” SIAM J. Comput. 18(6)1245-1262, DEC. 1989.
[18] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein, “ introduction to Algorithms, ” kingsinfo, 2002.