P.O.Box 8718, Beijing 100080, China	Journal of Software, July. 2004,15(7):1099-1106
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2004 by The Editorial Department of Journal of Software

用于XML模式和DTD规范化设计的层次模式设计

吴永辉

吴永辉1,2 1(复旦大学 计算机科学与工程系,上海 200433)2(中国科学院 软件研究所 计算机科学重点实验室,北京 100080)
作者简介: 吴永辉(1966－),男,浙江温岭人,博士,副教授,主要研究领域为数据库,数字图书馆.
联系人: 吴永辉 Phn: +86-21-55664413, Fax: 86-21-65642820, E-mail: slwu@fudan.edu.cn, http://www.cs.fudan.edu.cn
Received 2003-05-21; Accepted 2003-11-27

Abstract
Normalization design of XML Schemas and DTDs (document type definitions) is to produce a set of XML schemas or DTDs that can well represent data dependencies and eliminate redundancies. Now there are a few researches on it, and the existing researches are still at its initial stage. Provost proposed the idea of applying the theory of relational database to XML schemas normalization design. This idea has not been put into practice. The paper shows algorithms of hierarchical schemas design for XML schemas and DTDs normalization design based on Provost's idea. Firstly the paper analyzes hierarchy decomposition based on Provost's idea. Then it presents an algorithm producing a decomposition tree to eliminate redundant schemas. Finally it shows an algorithm of hierarchical schemas design for XML schemas and DTDs normalization design to get over deficiencies for Provost's idea. With respect to other researches on normalization design for XML schemas and DTDs, the set of full and embedded MVDs in hierarchical schemas produced by these algorithms are implied by the given set of MVDs (multivalued dependencies), and the hierarchical schemas eliminate redundant ones and satisfy the lossless join property.

Wu YH. Hierarchical schemas design for XML schemas and DTDs normalization design. *Journal of Software*, 2004,15(7):1099~1106.
http://www.jos.org.cn/1000-9825/15/1099.htm

摘要
XML模式和DTD(document type definition)规范化设计是给出一个很好地表示数据间依赖关系并消除了冗余的XML模式或DTD的集合.目前在这一方面开展的研究还不多,而且才刚起步.Provost提出将关系数据库理论应用于XML模式规范化设计的思想,这一思想还没有付诸实施.在Provost思想的基础上给出用于XML模式和DTD规范化设计的层次模式设计的算法.首先分析了基于Provost思想的层次分解;然后给出用于消除冗余模式的分解树设计算法;最后给出用于XML模式和DTD规范化设计的层次模式设计算法,这一算法克服了Provost思想的缺陷.相应于其他XML模式和DTD规范化设计的研究,在算法产生的层次模式中,完全MVD(multivalued dependency)和嵌入MVD的集合由给出的MVD集合导出;并且产生的层次模式具有消除冗余模式和满足无损联接的特性.

基金项目：Supported by the National Natural Science Foundation of China under Grant No.69933010 (国家自然科学基金); the Foundation of Laboratory of Computer Science, Institute of Software, the Chinese Academy of Sciences under Grant No.SYSKF0306 (中国科学院软件研究所计算机科学实验室开放课题基金)

References:

[1] Provost W. Normalizing XML. http:// www.xml.com/pub/a/2002/11/13/normalizing.html

[2] Arenas M, Leonid L. A normal form for XML documents. In: Lucian P, ed. Proc. of the 21st ACM SIGACT-SIGMOD-SIGART Symp. on Principles of Database Systems (PODS 2002). Madison: ACM Press, 2002. 85?96. http://www.cs.toronto.edu/~marenas/

[3] Wu X, Ling TW, Lee SY, Lee ML. NF-SS: A normal form for semistructured schema. In: Tjoa AM, ed. Proc. of the Int'l Workshop on Data Semantics in Web Information Systems 2001. Yokohama: IEEE Computer Society, 2001. http://www.comp.nus.edu.sg/~lingtw

[4] Wu X, Ling TW, Lee SY, Lee ML, Gillian D. Designing semistructured database using ORA-SS model. In: Tamer M, ed. Proc. of the 2nd Int'l Conf. on Web Information System Engineering. Kyoto: IEEE Computer Society, 2001. 17~28. http://www.comp.nus.edu.sg/~lingtw

[5] Lee ML, Ling TW, Wai LL. Designing functional dependencies for XML. In: Simonas S, ed. Proc. of the VIII. Conf. on Extending Database Technology (EDBT 2002). Prague: Springer-Verlag, 2002. 124~141. http://www.comp.nus.edu.sg/~lingtw

[6] Lee SY, Lee ML, Ling TW, Kalinichenko LA. Designing good semi-structured databases and conceptual modeling. In: Jacky A, ed. LNCS 1728. Paris: Springer-Verlag, 1999. 131~145. http://www.comp.nus.edu.sg/~lingtw

[7] Tari Z, Stokes J, Spaccapietra S. Object normal forms and dependency constraints for object-oriented schemata. ACM Trans. on Database Systems, 1997,22(4):513~569.

[8] Mok WY, Ng YK, Embley DW. Using NNF to transform conceptual data models to object-oriented database designs. Data & Knowledge Engineering, 1998,24(3):313~336.

[9] Wu Y, Zhou A. Research on properties of a set of object dependencies. Journal of Computer Research and Development, 2001, 38 (12):1491~1498 (in Chinese with English abstract).

[10] Wu Y, Zhou A. The normal object scheme forest with respect to conflict-free dependencies. Journal of Software, 2002,13 (8):1488~1493 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/13/1488.pdf

[11] Roth MA, Korth HF. The design of non-1NF relational databases into nested normal form. In: Yuan L, ed. Proc. of the 1987 ACM SIGMOD Conf. San Francisco: ACM Press, 1987. 143~159.

[12] Roth MA, Korth HF, Silberschatz A. Extended algebra and calculus for nested relational databases. ACM Trans. on Database System, 1988,13(4):389~417.

[13] Ozsoyoglu ZM, Yuan L. On the normalization in nested relational databases. In: LNCS 361. New York: Springer-Verlag, 1987. 243~271.

[14] Ozsoyoglu ZM, Yuan L. A new normal form for nested relations. ACM Trans. on Database System, 1987,12(1):111~136.

[15] David M. The Theory of Relational Databases. New York: Computer Science Press, 1983.

附中文参考文献:

[9] 吴永辉,周傲英.对象依赖集合的性质的研究.计算机研究与发展,2001,38(12):1491~1498.

[10] 吴永辉,周傲英.相应于无冲突依赖的规范化对象模式森林.软件学报,2002,13(8):1488~1493. http://www.jos.org.cn/1000-9825/13/1488.pdf