

# Marcado de textos

Rafael C. Carrasco  
Departamento de Lenguajes y Sistemas Informáticos  
Universidad de Alicante

## 1 Content models and regular expressions

- One-unambiguous regular languages (Brüggemann-Klein and Wood, 1998).
- Characterization of Glushkov automata (Caron and Ziadi, 2000).✕
- Ambiguity detection of RELAX grammars (Kawaguchi, 2001).
- On Subtyping of Tree-Structured Data: A Polynomial Approach (Bry et al., 2004).

## 2 Tree languages and automata

- Taxonomy of XML Schema Languages Using Formal Language Theory (Murata et al., 2001).✓
- Automata theory for XML researchers (Neven, 2002).✓.
- Query automata over finite trees (Neven and Schwentick, 2002).✕
- Incremental Schema Validation for XML Databases (Fisher and Wong, 2004).
- Efficient Incremental Validation of XML Documents (Barbosa et al., 2004).
- Incremental validation of XML documents (Balmin et al., 2004).

## 3 Structure comparison

- A survey on tree edit distance and related problems (Bille, 2005).✓
- X-Diff: An Effective Change Detection Algorithm for XML Documents (Wang et al., 2003).
- A Linear Tree Edit Distance Algorithm for Similar Ordered Tree (Touzet, 2005).
- Structure-Preserving Difference Search for XML Documents (Schubert et al., 2005).
- Detecting Changes in XML Documents (Cobena et al., 2002).

- A similarity between probabilistic tree languages: application to XML document families (Carrasco and Rico-Juan, 2003).✱

## 4 Databases

- eXist: An Open Source Native XML Database (Meier, 2002).
- XML to Relational Conversion Using Theory of Regular Tree Grammars (Mani and Lee, 2002).
- A performance evaluation of storing XML data in relational database management systems (Khan and Rao, 2001).
- A New Inlining Algorithm for Mapping XML DTDs to Relational Schemas (Lu et al., 2003).
- CPI: Constraints-Preserving Inlining algorithm for mapping XML DTD to relational schema (Lee and Chu, 2001).

## 5 DTD identification

- Automatic DTD simplification by examples (Bia and Carrasco, 2001).
- Generating Grammars for SGML Tagged Texts Lacking DTD (Ahonen et al., 1997).

## 6 Finite-state automata processing

- REX: XML Shallow Parsing with Regular Expressions (Cameron, 1998).✱
- Processing XML streams with deterministic automata and stream indexes (Green et al., 2004).
- Applications of Brzozowski derivatives to XML Schema processing (Sperberg-McQueen, 2005).

## 7 XML indexing and retrieval

- Updates of structure indexes (Kaushik et al., 2002)
- A survey in indexing and searching XML documents (Luk et al., 2002).
- Indexing XM documents: hybrid approach (Reiss, 2003).
- Searching structured documents (Trotman, 2004). Choosing document structure weights (Trotman, 2005).
- Proximal Nodes: A Model to Query Document Databases by Content and Structure (Navarro and Baeza-Yates, 1997).

## References

- Ahonen, H., Mannila, H., and Nikunen, E. (1997). Generating grammars for SGML tagged texts lacking DTD. *Mathematical and computer modelling*, 26(1):1-??
- Balmin, A., Papakonstantinou, Y., and Vianu, V. (2004). Incremental validation of XML documents. *ACM Trans. Database Syst.*, 29(4):710–751.
- Barbosa, D., Mendelzon, A. O., Libkin, L., Mignet, L., and Arenas, M. (2004). Efficient incremental validation of XML documents. In *Proceedings of the 20th International Conference on Data Engineering, ICDE 2004, 30 March - 2 April 2004, Boston, MA, USA*, pages 671–682. IEEE Computer Society.
- Bia, A. and Carrasco, R. C. (2001). Automatic DTD simplification by examples. In *ACH/ALLC 2001. The Association for Computers and the Humanities, The Association for Literary and Linguistic Computing, The 2001 Joint International Conference*, pages 7–9, New York University, New York City.
- Bille, P. (2005). A survey on tree edit distance and related problems. *Theor. Comput. Sci.*, 337(1-3):217–239.
- Brüggemann-Klein, A. and Wood, D. (1998). One-unambiguous regular languages. *Information and Computation*, 142(2):182–206.
- Bry, F., Drabent, W., and Maluszynski, J. (2004). On subtyping of tree-structured data: A polynomial approach. In Ohlbach, H. J. and Schaffert, S., editors, *Principles and Practice of Semantic Web Reasoning, Second International Workshop, PPSWR 2004, St. Malo, France, September 6-10, 2004, Proceedings*, volume 3208 of *Lecture Notes in Computer Science*, pages 1–18. Springer.
- Cameron, R. D. (1998). REX: XML shallow parsing with regular expressions. Technical Report TR 1998-17, School of Computing Science, Simon Fraser University, Burnaby, BC, Canada.
- Caron and Ziadi (2000). Characterization of glushkov automata. *TCS: Theoretical Computer Science*, 233:75–90.
- Carrasco, R. C. and Rico-Juan, J. R. (2003). A similarity between probabilistic tree languages: application to XML document families. *Pattern Recognition*, 36(9):2197–2199.
- Cobena, G., Abiteboul, S., and Marian, A. (2002). Detecting changes in XML documents. In *Proceedings of the 18th International Conference on Data Engineering, 26 February - 1 March 2002, San Jose, CA*. IEEE Computer Society.
- Fisher, D. K. and Wong, R. K. (2004). Incremental schema validation for xml databases. Technical Report UNSW-CSE-TR-0423, School of Computer Science and Engineering, University of New South Wales, Sydney 2052 Australia.
- Green, T. J., Gupta, A., Miklau, G., Onizuka, M., and Suci, D. (2004). Processing XML streams with deterministic automata and stream indexes. *ACM Trans. Database Syst.*, 29(4):752–788.
- Kaushik, R., Bohannon, P., Naughton, J. F., and Shenoy, P. (2002). Updates for structure indexes. In *VLDB '02: Proceedings of the 28th international conference on Very Large Data Bases*, pages 239–250. VLDB Endowment.
- Kawaguchi, K. (2001). Ambiguity detection of relax grammars.

- Khan, L. and Rao, Y. (2001). A performance evaluation of storing XML data in relational database management systems. In *WIDM '01: Proceedings of the 3rd international workshop on Web information and data management*, pages 31–38, New York, NY, USA. ACM Press.
- Lee, D. and Chu, W. W. (2001). CPI: Constraints-preserving inlining algorithm for mapping XML DTD to relational schema. *Data Knowl. Eng.*, 39(1):3–25.
- Lu, S., Sun, Y., Atay, M., and Fotouhi, F. (2003). A new inlining algorithm for mapping XML DTDs to relational schemas. In *ER (Workshops)*, pages 366–377.
- Luk, R. W., Leong, H. V., Dillon, T. S., Chan, A. T., Croft, B. B., and Allan, J. (2002). A survey in indexing and searching XML documents. *J. Am. Soc. Inf. Sci. Technol.*, 53(6):415–437.
- Mani, M. and Lee, D. (2002). XML to relational conversion using theory of regular tree grammars. In *EEXTT*, pages 81–103.
- Meier, W. (2002). eXist: An open source native XML database. In *Web, Web-Services, and Database Systems*, pages 169–183.
- Murata, M., Lee, D., and Mani, M. (2001). Taxonomy of XML schema languages using formal language theory. In *Proceedings of the Extreme Markup Languages 2001 Conference, 12-17 August 2001, Montréal, Quebec, Canada*.
- Navarro, G. and Baeza-Yates, R. (1997). Proximal nodes: A model to query document databases by content and structure. *ACM Transactions on Information Systems*, 15(4):400–435.
- Neven, F. (2002). Automata theory for XML researchers.
- Neven, F. and Schwentick, T. (2002). Query automata over finite trees. *Theor. Comput. Sci.*, 275(1-2):633–674.
- Reiss, K. (2003). *Indexing XML documents: A hybrid approach*.
- Schubert, E., Schaffert, S., and Bry, F. (2005). Structure-preserving difference search for XML documents. In *Proceedings of the Extreme Markup Languages 2005 Conference, 1-5 August 2005, Montréal, Quebec, Canada*.
- Sperberg-McQueen, C. M. (2005). Applications of brzozowski derivatives to XML schema processing. In *Extreme Markup Languages®*.
- Touzet, H. (2005). A linear tree edit distance algorithm for similar ordered trees. In *CPM*, pages 334–345.
- Trotman, A. (2004). Searching structured documents. *Inf. Process. Manage.*, 40(4):619–632.
- Trotman, A. (2005). Choosing document structure weights. *Inf. Process. Manage.*, 41(2):243–264.
- Wang, Y., DeWitt, D. J., and yi Cai, J. (2003). X-diff: An effective change detection algorithm for XML documents. In Dayal, U., Ramamritham, K., and Vijayaraman, T. M., editors, *Proceedings of the 19th International Conference on Data Engineering*, pages 519–530, Bangalore, India. IEEE Computer Society Press.