

# Reminiscences on Influential Papers

*Kenneth A. Ross, editor*

I continue to invite unsolicited contributions. See <http://www.acm.org/sigmod/record/author.html> for submission guidelines.

---

**Frank Neven**, University of Limburg, [frank.neven@luc.ac.be](mailto:frank.neven@luc.ac.be).

[A. Brüggemann-Klein, M. Murata and D. Wood. Regular Tree and Regular Hedge Languages over Unranked Alphabets: Version 1, April 3, 2001. Technical Report HKUST-TCSC-2001-05.]

Already in 1996, the pre-XML era, my advisor Jan Van den Bussche suggested to work on structured document databases. In contrast with the at that time very popular graph-based semi-structured philosophy, we considered data to be trees defined by context-free grammars and wrote some papers on the use of attribute grammars as query languages. I wanted to extend these results to trees defined by extended context-free grammars (DTDs if you want) as was the case for SGML. Unfortunately, I had no clue how to deal in an elegant way with the inherent unrankedness of such trees. The question taunted me for quite a while until somewhere in the spring of 1998, Anne Brüggemann-Klein posted a message on the PODP email list (of which I already forgot I was subscribed to and I doubt still exists now), announcing a draft of a paper on unranked tree automata coauthored by Derick Wood. I was completely shocked when I read that paper. The mechanism they used for dealing with unrankedness was precisely what I had been looking for all that time. It became immediately clear to me that their technique could be useful in the study of many aspects of structured document databases. Their presentation of unranked tree automata did not only help me with my attribute grammars, it also had, and still has, a profound influence on all my research on foundations of XML. In my opinion, unranked tree automata as defined by Brüggemann-Klein, Murata and Wood, provide a fundamental and elegant tool for XML research.

The paper is available in its most recent form as a technical report coauthored by Makoto Murata. The latter author independently defined forest automata over unranked forests. The three authors then decided to join forces. As the initial draft attracted much attention, the authors decided to make it available as a research report thereby providing a pointer for database researchers to cite.

---

**Beng Chin Ooi**, National University of Singapore, [ooibc@comp.nus.edu](mailto:ooibc@comp.nus.edu).

[R. Bayer and E. McCreight. Organization and maintenance for large ordered indexes. *Acta Informatica*, 1(3):173-189, 1972.]

Most CS students would have learned the B<sup>+</sup>-tree either in a database or data structure course. Yet, the structure forms the basis for many researchers designing novel indexing structures. Indeed, in many advanced database application domains such as spatial databases, multimedia databases, temporal databases, and object-oriented databases, we can always find methods that are based on the B<sup>+</sup>-tree. In my own research, I have also built on the B<sup>+</sup>-tree for advanced application domains (e.g., iMinMax and iDistance).

What impresses me most of the B<sup>+</sup>-tree is that it is such a simple and yet elegant and powerful structure. It is simple because it is easy to understand, and the associated algorithms are easy to implement. It is elegant because it grows and shrinks gracefully; the ease of maintenance is one of the features required for any index to be long lasting and practically useful. It is powerful because of the performance and space utilization guarantees. Its strength in terms of robustness and scalability is simply an art, and the design principles behind the index provides me down-to-earth guiding principles in designing new indexes. I believe this thirty-year-old ubiquitous structure is here to stay for a long time. Its ability to adapt to “hostile climate” has made it one of the most fruitful tree in computer science research, and will continue to provide inspiration to myself and many.