# Document engineering

**Uwe M. Borghoff · Sebastian Rönnau**

Document engineering is the computer science discipline that investigates systems for documents in any form and in all media. As with the relationship between software engineering and software, document engineering is concerned with principles, tools and processes that improve our ability to create, manage, and maintain documents.

Apparently, document engineering is an important yet very diverse field of research. It covers aspects of document standards, authoring systems and workflows, as well as long-term archiving. Further on, layouting as well as an automatic adaptation of documents are important aspects. But also document migration and transformation, as well as synchronization are crucial topics in document engineering. In recent years, the importance of active documents and multimedia documents became more and more important.

The ACM Symposium on Document Engineering (DocEng) is an annual meeting of researchers active in document engineering. It is sponsored by ACM by means of the ACM SIGWEB special interest group. Up to now, it has been held ten times, with an steadily increasing popularity.

This Special Issue of the Journal on Computer Science—Research and Development consists of extended versions of the best papers of ACM DocEng 2009 which has been held at the Universität der Bundeswehr in Munich, Germany. The best papers have been selected in a two-step approach: firstly, a vote by the Program Committee based on the peer-reviewing results narrowed the numbers down, then, secondly, an anonymous audience vote after the presentation of the papers in question gave the final ordering. Thanks to all who have participated!

The selection reflects the broad range of topics in document engineering. Before presenting the articles, we wish to gratefully thank Peter King and Boris Chidlovksii. Being head of the Steering Committee, Peter King has brought the DocEng community to the point where it is today. Boris Chidlovskii served as Program Chair of DocEng 2009. He ensured the high quality of the proceedings by accepting only 37 papers out of 119 submissions, resulting in an acceptance rate of 31%.

The paper "XCC: Change Control of XML Documents" by Sebastian Rönnau and Uwe M. Borghoff presents an architecture for finding and encoding the differences between XML documents. These differences can be used to reconstruct document versions or to merge two versions of a document. Their work has a broad applicability, as many document formats rely on XML as meta-language. The paper is theoretically well-founded, presenting a novel tree-to-tree correction algorithm that is highly efficient. The authors of this paper have been granted the Best Paper Award of ACM DocEng 2009.

"Reconstituting typeset Marriage Registers using simple software tools" by David Brailsford shows a complete workflow of simple standard tools that are used to perform a highly complex use case. Printed marriage registers are analyzed using OCR techniques and post-processed, thus creating a computable output. This paper is an important contribution to practice and experiences in document engineering.

With their paper "Interactive Web Documents", John M. Boyer, Charles F. Wiecha, and Rahul P. Akolkar present a universal architecture for storing and managing interactive documents within the Web. They present a novel format for interactive documents relying on existing standards. Most importantly, they also present a corresponding protocol for

U.M. Borghoff · S. Rönnau (✉)
Institute for Software Technology, Faculty of Computer Science,
Universität der Bundeswehr München, 85577 Neubiberg,
Germany
e-mail: sebastian.roennau@unibw.de

the interaction with the document. As interactive documents become more and more important in business, this work may lay the basis for future standards in this domain.

Luiz Gustavo Fernandes, Thiago Nunes, Mariana Kolberg, Fabio Giannetti, Rafael Nemetz, and Alexis Cabeda have investigated the impact of increasing personalization of documents on high performance print queues. In their paper "Job Profiling and Queue Management in High Performance Printing", they present a profiling method for PDF documents that estimates the ripping cost to increase the efficiency of the printing queues. This paper has a high practical relevance as it helps to improve the return of investment of printing engines.

All these papers demonstrate the diversity and density of the document engineering community—a vibrant community which seeks for even more participation and contribution.