Journal of Universal Computer Science, vol. 14, no. 18 (2008), 2909-2911
submitted: 22/9/08, accepted: 29/9/08, appeared: 1/10/08 © J.UCS

Further Advances in Document Engineering

J.UCS Special Issue

Rafael Dueire Lins
(Universidade Federal de Pernambuco, Refice, Brazil
rdl@ufpe.br)

1 Introduction

A Document is any sort of object that conveys relevant information. This wide definition of document goes far beyond paper documents, the most usual form of document, and encompasses all sorts of materials from bones of pre-historical animals to videos, etc. Document Engineering is the area of knowledge focused in principles, algorithms, tools and processes that allow creating, managing, store, compact, access, and maintaining digital documents. The World Wide Web (WWW) made the fields of document recognition and retrieval grow rapidly in recent years. New application areas such as the, digital libraries, and video- and camera-based OCR have appeared lately.

The main fields in Document Engineering are:

- Algorithms and systems for machine-printed and handwritten character and word recognition, especially for degraded documents (e.g., faxes);
- Character and word segmentation techniques;
- Identification and analysis of tables or equations;
- Page segmentation, including hierarchical decomposition of documents into text regions, halftones, colored/textured background, etc;
- Logical structure analysis and recognition, linguistic representation of document structure;
- Raster-to-vector conversion of line-art, maps, and technical drawings;
- Document image filtering, enhancement and compression techniques;
- Document degradation models;
- Video and camera based OCR;
- Applications of document recognition to the WWW and digital libraries;
- Techniques to support spoken language access to document text (audio browsing of doc. databases);
- Multilingual character recognition;
- Impact of recognition accuracy on retrieval effectiveness;
- Recovery and use of logical structure for retrieval;
- Relevance feedback techniques for document retrieval;
- Cross-language and multi-lingual retrieval;
- Categorization and summarization of text documents and image documents;

- Keyword spotting in document images;
- Approximate string matching algorithms for OCRs;
- Non-textual retrieval methods;
- Image and multimedia search;
- Interfaces for document retrieval;
- Benchmarking and evaluation issues

2 Contents of this issue

This volume opens with an invited contribution by Josep Lladós and his colleagues from the Computer Vision Center of the Universitat Autònoma de Barcelona, Spain entitled "A Generic Architecture for the Conversion of Document Collections into Semantically Annotated Digital Archives" addressing a central point in document engineering. Josep Lladós received the 2007 IAPR (International Association on Pattern Recognition) Young Distinguished Scientist Award for his contributions in the area of document analysis and recognition.

Along the same research line of the invited paper in this issue there is the contribution from Austria and Germany entitled "Systematic Characterisation of Objects in Digital Preservation: The eXtensible Characterisation Languages".

3 The Reviewing board

Experts of all areas of document engineering, from all over the world, composed the board that refereed and reviewed the papers for this issue:

Adel M. Alimi (*University of Sfax, Tunisia*)
Angelo Marcelli (*University of Salerno, Italy*)
Apóstolos Antonacopoulos (*University of Salford, UK*)
Alejandro C. Frery (*Universidade Federal de Alagoas, Brazil*)
Andreas Dengel (*Kaiserslautern University, Germany*)
Antony Wiley (*Hewlett Packard Labs., Bristol,UK*)
Aurélio Campilho (*Universidade do Porto, Portugal*)
Daniel P. Lopresti (*Lehigh University, USA*)
Brian Lawler (Cal Poly, USA)
David S. Doermann (*University of Maryland, USA*)
Dov Dori (*Technion, Israel Institute of Technology, Israel*)
Ethan Munson (*Univ. of Wisconsin – Milwaukee, USA*)
F.Heron de Carvalho Jr (*U. Federal do Ceará, Brazil*)
F.Mário Martins (*Universidade do Minho em Braga, Portugal*)
Flávio Bortolozzi (*OPET, Brazil*)
Graham Leedham (Nanyang Technical University, Singapore)
Henry S. Baird (*Lehigh University, USA*)
Hirobumi Nishida (*Ricoh SW Research Center, Japan*)
Horst Bunke (*University of Bern, Switzerland*)

J.Caldas Pinto (Instituto Superior Técnico, Portugal)
Jacques Facon (*Pontifícia Universidade Católica do Paraná, Brazil*)
Jean-Marc Ogier (*Université de la Rochelle, France*)
Jian Fan (*Hewlett Packard Labs., Palo Alto, USA*)
Jian Liang (*Media Management Tech, Seattle, USA*)
Jin H. Kim (*KAIST, Korea*)
João Marques de Carvalho (*UFCG, Brazil*)
Jonathan J. Hull. (*Ricoh California Res.Center, USA*)
Josep Llados (*Univ. Autonoma de Barcelona, Spain*)
Kazem Taghva (*University of Nevada, USA*)
Lawrence O'Gorman (*Avaya Labs, USA*)
Louisa Lam (*The HK Inst of Education, Hong Kong*)
Luis Corte-Real (*Universidade do Porto, Portugal*)
Luis Eduardo Oliveira (*Pontifícia Universidade Católica do Paraná*, *Brazil*)
Majid Mirmehdi (*University of Bristol, England*)
Marco Gori (*Università di Siena, Italy*)
Maria Feldgen (*Universidad de Buenos Aires, Argentina*)
Michael Perrone (*IBM T.J. Watson Research Center, USA*)
Mohamed Kamel (*University of Waterloo, Canada*)
Nasser Sherkat (*The Nottingham Trent University, UK*)
Nelson Mascarenhas (*Universidade Federal de São Carlos, Brazil*)
Pedro Rangel Henriques (*Universidade do Minho em Braga, Portugal*)
Pertti Vakkari. (*University of Tampere, Finland*)
Qian Lin (*Hewlett Packard Labs, Palo Alto, USA*)
Robert Sabourin (*Ecole de Technologie Supériure, Canada*)
Rolf Ingold (*University of Fribourg, Switzerland*)
Ricardo Queiroz (Universidade de Brasília, Brazil)
Salvatore Tabbone (*University of Nancy 2, France*)
Sargur Srihari (*State University of New York at Buffalo, USA*)
Seong-Whan Lee (*Korea University, Korea*)
Tan Chew Lim (NUS, Singapore)
Thierry Paquet (*Université de Rouen, France*)
Thomas Mandl (*University of Hildeshein, Germany*)
Tin Kam Ho (*Bell Labs, Lucent Technologies, USA*)
Umapada Pal (*Indian Statistical Institute, Kolkata, India*)
Utpal Garain (*Indian Statistical Institute, Kolkata, India*)
Venu Govindaraju (*State University of New York at Buffalo, USA*)
Weiler Finamore (*Pontifícia Universidade Católica do Rio de Janeiro, Brazil*)
Xiaoqing Ding (*Tsinghua University, China*)

Acknowledments

The editor and the authors of this volume are grateful for the enthusiasm of Prof. Dr. Herman Maurer and Dana Kaiser that made it possible.

Rafael Dueire Lins
Brazil, September 2008