Proceedings

Fifth Annual Symposium on Document Analysis and Information Retrieval

April 15 - 17, 1996

Las Vegas, Nevada

UNLV

UNIVERSITY OF NEVADA LAS VEGAS

Sponsored by:

Information Science Research Institute

and

Howard R. Hughes College of Engineering

University of Nevada, Las Vegas
4505 Maryland Parkway, Box 454021
Las Vegas, Nevada 89154-4021

UB/TIB Hannover 89
114 017 166

Table of Contents

Conference Chairman's Message i
Message from the President, University of Nevada, Las Vegas ii
Message from the Dean, College of Engineering iii
Board of Advisors, Information Science Research Institute iv
Conference Committee v
Document Analysis Reviewers vi
Information Retrieval Reviewers vii
Debate Panel vii

INVITED PAPERS

Substituting Images for Books: The Economics for Libraries 1
Michael Lesk

Text Recognition - From Pixels to Meaning 17
Jürgen Schürmann

Information Retrieval - from Academic Research to Practical Applications 37
Hans-Peter Frei

CONTRIBUTED PAPERS

Minimum Spanning Trees for Text Segmentation 51
Antonio P. Dias

In-house Mail Distribution by Automatic Address and Content Interpretation 67
Thomas Brückner, Peter Suda, Hans Ulrich Block and Gerd Maderlechner

USeg: A Retargetable Word Segmentation Procedure for Information Retrieval 77
Jay M. Ponte and W. Bruce Croft

Text Categorization: A Symbolic Approach 87
Isabelle Moulinier, Gailius Raskinis and Jean-Gabriel Ganascia

Support Tools for Visual Information Management 101
Gökhan Kutlu, Bruce A. Draper, J. Eliot B. Moss and Edward M. Riseman

Edit Distance of Regular Languages 113
Horst Bunke

Language Identification: Examining the Issues 125
Penelope Sibun and Jeffrey C. Reynar

Fast Decision Tree Ensembles for Optical Character Recognition 137
Harris Drucker

Length Normalization in Degraded Text Collections 149
Amit Singhal, Gerard Salton and Chris Buckley

Extraction of Thematically Relevant Text from Images 163
Francine R. Chen and Dan S. Bloomberg

Measuring the Effects of Data Corruption on Information Retrieval 179
Elke Mittendorf and Peter Schäuble

Keyword-Based Browsing and Analysis of Large Document Sets 191
Ido Dagan, Ronen Feldman and Haym Hirsh

Tailoring a Retrieval System for Naive Users 209
Adrienne J. Kleiboemer, Manette B. Lazear and Jan O. Pedersen

Improving Full-Text Precision on Short Queries using Simple Constraints 217
Marti A. Hearst

Degraded Character Image Restoration 233
John D. Hobby and Henry S. Baird

Automatically-generated High-reliability Features for Dichotomies of Printed Characters 247
George Nagy and Xiaoyin Wang

Retrieval Strategies for Noisy Text 255
Daniel Lopresti and Jiangying Zhou

A General-Purpose Japanese Optical Character Recognition System 271
Sargur N. Srihari, Geetha Srikantan, Tao Hong and Brian Grom

OCR and Voting Shell Fulfilling Specific Text Analysis Requirements 287
Thorsten Jäger

Logotype Detection in Compressed Images using Alignment Signatures 303
A. Lawrence Spitz

Reliable Recognition of Handwritten Marks in Checkboxes 311
B. Latanzio and A. Garzotto

Generalized Form Registration Using Structure-Based Techniques 321
Michael D. Garris and Patrick J. Grother