Proceedings

Second Annual Symposium on Document Analysis and Information Retrieval

April 26 - 28, 1993
Las Vegas, Nevada, USA

Sponsored by:
Information Science Research Institute
and
Howard R. Hughes College of Engineering

In Cooperation with

University of Nevada, Las Vegas
4505 Maryland Parkway Box 454021
Las Vegas, Nevada 89154-4021
# 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 Members ........................................................... v

## INVITED PAPERS

Calibration of Document Image Defect Models. ............................................. 1  
   Henry Baird

Machine-Tractable Dictionaries and Large-Scale Sense-Resolution in Text......... 17  
   Yorick Wilks

From Pixels to Paragraphs: the Use of Models in Text Recognition ................. 47  
   Sargur N. Srihari

English Document Database Design and Implementation Methodology ............... 65  
   Robert M. Haralick

## CONTRIBUTED PAPERS

Word Recognition for Information Retrieval in the Image Domain ................. 105  
   John M. Trenkle and Robert C. Vogt

Automatic Inference of Textline Orientation ........................................... 123  
   David J. Ittner

A New Model for Context-Driven Word Recognition .................................... 135  
   Eric Lecolinet

Word Recognition Result Interpretation Using the Vector Space Model for  
   Information Retrieval ........................................................................... 147  
   Jonathan J. Hull and Yanhong Li

ADIIR: A Testbed for Experiments in Document/Image Indexing and Retrieval ..... 157  
   Paul Thompson and Robert Kellogg

Improving OCR Accuracy With Linguistic Knowledge ................................... 169  
   Deborah A. Dahl, Lewis M. Norton and Suzanne Liebowitz Taylor

Exporting Phrases: A Statistical Analysis of Topical Language ..................... 179  
   Amy M. Steier and Richard K. Belew
A Boolean Approximation Method for Query Construction and Topic Assignment in TREC ................................................................. 191
Paul S. Jacobs, George R. Krupka and Lisa F. Rau

Information Retrieval with Semantic Representation of Texts ..................... 201
Sung H. Myaeng and Elizabeth D. Liddy

Keyword Location in Noisy Document Images ..................................... 217
Siamak Khoubyari and Jonathan J. Hull

Development of a Japanese Document Recognition System ....................... 233
Masaaki Mizuno, Takeshi Kamimura, Toshiyuki Tanaka, Haruhiko Tanaka, Toshinao Ide, Jun Tsukumo, Tsutomu Temma

Recognition of Multifont Latin Texts ............................................. 243
Adnan Amin and Andrew Foti

Spatial Similarity Based Retrieval In Image Databases .......................... 255
V. Naidu Gudivada and Vijay V. Raghavan

Effects of Query Term Weights Modification in Document Retrieval .......... 271
Jing-Jye Yang and Robert R. Korfhage

N-Gram-Based Matching for Multi-Field Database Access in Postal Applications. 287
William B. Cavnar and Alan J. Vayda

Numerical Descriptor of 2D Forms Applied to Line Drawings ................... 299
Lala Andriamampianina, Georges Stamon, Georges Karavias
Maurice Poulenard

Automatic Segmentation of Engineering Drawings with Symbols and Connections ............................................................ 317
Yuhong Yu, Ashok Samal and Sharad Seth

Conversion of Scanned Cartographic Maps to Geographic Information Systems Using Semantic Models ............................. 339
Gerd Maderlechner, Helmut Mayer and Christian Heipke

An Efficient Directory System for Document Retrieval Using Parallel Processing. 349
Ajay Gupta and Dalia Motzkin

The Scientist's Assistant: Structure-Based Navigation of Technical Texts .......... 367
Susan Gauch and Robert P. Futrelle
Experimental Comparison of Three Methods for Personal Information Retrieval. . 379
   Robert Godin, Marc Davidson, Rokia Missaoui and Hafedh Mili

Hand-Printed Digit Recognition Using Nearest Neighbour Classifiers. ............ 397
   Michael Sabourin, Amar Mitiche, Danny Thomas and George Nagy

Hierarchical Curve Retrieval with Application to Aerial Image Alignment and
   Cursive Handwriting Recognition. ................................................. 411
   Jianying Hu and Theo Pavlidis

A Contextual Reasoning System for the Interpretation of Machine Printed Address
   Block Images. ................................................................. 429
   Alan J. Vayda, Michael P. Whalen, Daniel J. Hepp and Andrew M. Gillies

INFOCLAS: Classifying the Message in Printed Business Letters. ................. 443
   Rainer Hoch and Andreas Dengel

Hyperdocument Structuration, Management and Exchange in an Object Oriented
   Database. ................................................................. 457
   Claude Chrisment, Catherine Comparot, Christine Julien,
   Pierre-Yves Lambolez and Florence Sedes

A Multimodal Approach for Information Retrieval. .................................. 475
   Catherine Mendiboure, Josiane Mothe and Jacques Thomazeau

Author Index ................................................................. 491