

Table of Contents

MESSAGE FROM THE ORGANIZERSX

KEYNOTE SPEAKERS

GOOGLE PRINT.....XX
Luc Vincent, Google

EDUCE: ENHANCED DIGITAL UNWRAPPING FOR CONSERVATION AND
EXPLORATION OF INACCESSIBLE TEXTS.....XX
W. Brent Seales and Yun Lin, University of Kentucky

SESSION 1 DOCUMENT ANALYSIS SYSTEMS

A FLEXIBLE EXPERIMENTATION AND CONFIGURATION PLATFORM FOR
MULTILINGUAL HMM OCR.....XX
Ehry MacRostie and Premkumar Natarajan, BBN Technologies

DOCUMENT INFORMATION PROCESSING: TOWARDS SOFTWARE AND TEST COLLECTIONS ...XX
G. Agam, S. Argamon, O. Frieder, D. Grossman and D. Lewis,
Illinois Institute of Technology, David D. Lewis Consulting**

DOCLIB: A DOCUMENT PROCESSING RESEARCH TOOL,XX
Kevin Chen, Stefan Jaeger, Guangyu Zhu and David Doermann,
Booz Allen Hamilton*, University of Maryland*

A PROCESS FLOW FOR REALIZING HIGH ACCURACY FOR OCR TEXT,XX
Kazem Taghza, Julie Borsack and Tom Nartker, University of Nevada, Las Vegas

SESSION 2 DOCUMENT PROCESSING AND ENHANCEMENT

A DOCUMENT IMAGE ENHANCEMENT MODULE: PERSPECTIVE WARP CORRECTION.....XX
C. Monnier, Vitaly Ablavsky, Steve Holden and Magnus Snorrason,
Charles River Analytics, Inc and Boston University**

DOCUMENT RECOGNITION AND TRANSLATION WITH HANDHELD MOBILE DEVICES.....XX
Esin Darici Haritaoglu and Ismail Haritaoglu, Polar Rain Inc.

SESSION 3 PROCESSING OF ARABIC DOCUMENTS

**A SYSTEM FOR DISCRIMINATING HANDWRITING FROM MACHINE PRINT ON
NOISY ARABIC DOCUMENTSXX**
John C. Femiani, Mariano Phielipp and Anshuman Razdan, Arizona State University

NLP-ENHANCED EXPLOITATION OF DEGRADED ARABIC TEXTSXX
Michael Shalev, Encyclopaedia Britannica

**INITIAL RESULTS IN OFFLINE ARABIC HANDWRITING RECOGNITION USING
LARGE-SCALE GEOMETRIC FEATURES.....XX**
Ilya Zavorin, Eugene Borovikov and Mark Turner, CACI International Inc.

CLASSIFICATION OF MACHINE PRINT AND HANDWRITING IN MIXED ARABIC DOCUMENTS ...XX
Karthik Sridharan, Faisal Farooq and Venu Govindaraju, CEDAR, SUNY at Buffalo

SESSION 4 EXTRACTION AND INFORMATION RETRIEVAL

ACCURATE DOCUMENT CATEGORIZATION OF OCR GENERATED TEXTXX
Robert Price and Anthony Zukas, Content Analysis LLC, and SAIC**

EFFECT OF DEGRADED INPUT ON STATISTICAL MACHINE TRANSLATIONXX
Faisal Farooq and Yaser Al-Onaizan, CEDAR, SUNY at Buffalo and
IBM T.J. Watson Research Center**

**MOBILE INTERACTIVE SUPPORT SYSTEM FOR TIME-CRITICAL
DOCUMENT EXPLOITATIONXX**
George Nagy and Daniel Lopresti, Rensselaer Polytechnic Institute,
Lehigh University**

SESSION 5 ARABIC DOCUMENT ANALYSIS

**HANDWRITTEN ARABIC WORD SPOTTING USING THE CEDAR ARABIC
DOCUMENT ANALYSIS SYSTEMXX**
*Sargur Srihari, Harish Srinivasan, Pavithra Babu and Chetan Bhole, CEDAR, University at
Buffalo, SUNY*

WORD SPOTTING IN ARABIC SCRIPT DOCUMENTS,.....XX
A. Lawrence Spitz and Jim Yaghi, DocRec Ltd.

CHALLENGES IN ADAPTING MACHINE-PRINT ARABIC OCR FOR HANDWRITINGXX
Steven G. Schlosser and Robert C. Vogt, NovoDynamics, Inc.

SESSION 6 CROSS DOMAIN APPLICATIONS

ROBOT NAVIGATION TECHNIQUES FOR ENGINEERING DRAWING ANALYSISXX
Thomas C. Henderson and Chimiao Xu, University of Utah

PERCEPTUAL ORGANIZATION IN SEMANTIC ROLE LABELING.....XX
P. Sarkar and E. Saund (Palo Alto Research Center)

**PICTOGRAPHY RECOGNITION TECHNOLOGY APPLIED TO DISTINCTIVE
CHARACTERISTICS OF HANDWRITTEN ARABIC TEXTXX**
Mark A. Walch and Donald T. Gantz, The Gannon Technologies Group and
George Mason University*)*

SESSION 7 DOCUMENT ANALYSIS APPLICATIONS

A DOCUMENT TRIAGE SYSTEM FOR ARMY APPLICATIONXX
F. Fisher, K. Marcus, R. Chang and J. Turner, Army Research Laboratory

MULTI-LANGUAGE HANDWRITING DERIVED BIOMETRIC IDENTIFICATION.....XX
Donald T. Gantz, John J. Miller and Mark A. Walch, George Mason University,
The Gannon Technologies Group**

**TRADEOFF STUDIES ABOUT STORAGE AND RETRIEVAL EFFICIENCY OF BOUNDARY
DATA REPRESENTATIONS FOR LLS, TIGER, AND DLG DATA STRUCTURESXX**
David Clutter and Peter Bajcsy, University of Illinois

DEMONSTRATION ABSTRACTS

**ABBYY SOFTWARE HOUSE OCR, FORMS PROCESSING AND DATA
CAPTURE TECHNOLOGY REVIEW275**
Imelda Valenzuela, ABBYY

VERUS A MIDDLE EASTERN LANGUAGE OCR.....XX
Steven G. Schlosser, NovoDynamics, Inc.

**IDG: A BUSINESS INFORMATION EXTRACTION, MANAGEMENT, AND ROUTING
FRONT-END FOR CONTENT MANAGEMENT SYSTEMSXX**
*Vikas Krishna, Savitha Srinivasan, Neil Boyette, Isaac Cheng, Jeffrey Kreulen and Tapas
Kanungo, IBM Almaden Research Center*

FAST SKEW AND SLANT CORRECTION FOR ARABIC WRITTEN WORD OR LINE.....XX
Essam M. Zaki, Sakhr Software USA Inc.

EVALUATION WORKSHOP ABSTRACTS

PERFORMANCE EVALUATION OF MULTILINGUAL DOCUMENT EXPLOITATION SYSTEMS.....XX
Steven G. Schlosser, NovoDynamics, Inc.

ISSUES WITH AUTOMATIC OCR EVALUATION AND METRICS.....XX
Kristian J. Concepcion.

EVALUATION ISSUES IN IMAGE REFINERXX
Kristen Summers and Eugene Borovikov, CACI

THE SPORADIC NATURE OF OCR EVALUATIONSXX
Tapas Kanungo, IBM Almaden Research Center

**GROUND TRUTH REPRESENTATION USED IN TESTING AND OPTIMIZATION OF THE
OPTICAL WORD RECOGNITION SYSTEM.....XX**
Mike Ladwig, Northrop Grumman Corporation

METRICS FOR WORD SPOTTINGXX
Sargur Srihari, CEDAR University of Buffalo

AUTHOR INDEX355