



ACM/SIGIR

**24th Annual International
Conference on
Research and Development
in Information Retrieval**

September 9-13, 2001

to be held in
**New Orleans
LOUISIANA, USA**

In cooperation with
**AICA-GLIR (Italy), ASIS (USA)
BCS-IRSG (UK), GI (Germany)**

ADVANCE PROGRAM

Conference Schedule

Sunday, September 9, 2001 Pre-Conference

8:30-12:30 Tutorials

1:30-5:30 Tutorials

Monday, September 10 Main Conference

9:00-10:10 Conference Opening
and Keynote Address

10:30-12:30 Parallel Paper Sessions 1A and 1B

2:00-3:30 Parallel Paper Sessions 2A and 2B

4:00-5:00 Paper Session 3

5:30-10:00 Poster Session, Demonstration
Session, and Reception

Tuesday, September 11 Main Conference

9:00-10:00 Keynote Address

10:30-12:30 Parallel Paper Sessions 4A and 4B

2:00-3:30 Parallel Paper Sessions 5A and 5B

4:00-5:30 Paper Session 6

6:00-10:00 Banquet

Wednesday, September 12 Main Conference

9:00-10:00 Keynote Address

10:30-12:30 Parallel Paper Sessions 7A and 7B

12:30-3:00 ACM/SIGIR Meeting and Lunch

3:00-4:30 Parallel Paper Sessions 8A and 8B

Thursday, September 13 Post-Conference

8:00-5:00 Workshops

Friday, September 14

8:00-5:00 Optional Extension of One Workshop

Keynote Addresses

Monday, 9:00-10:00 *Mining the Web*

—Tom Mitchell (WhizBang!)

Tuesday, 9:00-10:00 *Wireless Graffiti & Annotating Physical Space with Digital Information*

—Tomasz Imielinski (Rutgers University)

Wednesday, 9:00-10:00 *Going Beyond Extract-Based Text Summarization*

—Eduard Hovy (USC Information Sciences Institute)

Welcome to the Advance Program for the **24th ACM/SIGIR International Conference on Research and Development in Information Retrieval.**

You are cordially invited to come, to join us, to participate in this major technical information retrieval conference. There will be tutorials, keynote addresses, technical paper sessions, poster papers, demonstrations, exhibits, and workshops that will advance the state of the art of information retrieval and discuss many of the main issues and applications.

The Program Committee, along with the Workshops Committee, the Tutorials Committee, the Poster Sessions Committee, the Demonstrations Committee, and the Exhibitions Committee, has put together an excellent combination of sessions to enrich anyone interested in information retrieval. The Local Arrangements Committee has made sure we have a variety of social programs and food offerings that promise to satisfy.

Let me suggest that as soon as you make the positive decision to come to this great conference, please act quickly. Book one of the limited number of rooms at the conference hotel, and register for the conference to take advantage of the early-bird price structure. Please check out the conference web site at "<http://www.sigir2001.org>" for an overall view of the conference and for details about the conference. If you have any questions, feel free to contact me.

I shall look forward to seeing you in the city "that care forgot," in NuAwlins, as they say down here. *Laissez les bon temps rouler!*

Donald H. Kraft, *Conference Chair*

Conference Organization

Conference Chair

Donald H. Kraft, Louisiana State University

Program Committee Chairs

W. Bruce Croft (USA), University of Massachusetts

David J. Harper (Europe/Africa), Robert Gordon
University

Justin Zobel (Asia and Australasia), RMIT University

Posters Chair

Efthimis Efthimiadis, University of Washington

Workshops Chair

Susan Gauch, University of Kansas

Demonstrations Chairs

Padmini Srinivasan, University of Iowa

Evangelos Triantaphyllou, Louisiana State University

Tutorials Chair

Raya Fidel, University of Washington

Mentoring Chair

David Lewis, Independent Consultant, Chicago

Awards Chair

David Lewis, Independent Consultant, Chicago

Exhibits and Corporate Sponsors Chair

David Grossman, Illinois Institute of Technology

Treasurer

Andrea Houston, Louisiana State University

Publicity Chair

Eddie Rasmussen, University of Pittsburgh

Local Arrangements Chair

Carol Barry, Louisiana State University

Information Architect

David Robins, University of Pittsburgh

Tutorials — Morning

Introduction to Information Retrieval

—Mark Sanderson (University of Sheffield)

Introduces the core concepts of Information Retrieval and provides attendees with extensive bibliographic information to allow them to pursue their interests further.

Web Information Retrieval

—Krishna Bharat (Google, Inc.) and David Hawking (CSIRO Canberra)

Intends to bring researchers, students and practitioners up to speed with both the theory and practice of Web search. Based on an understanding of challenges specific to the Web collection and user base, it will explore the state of the art in algorithms, deployment infrastructure and evaluation methodology. Participants will have basic familiarity with the Web and search engines, and exposure to IR concepts such as ranked retrieval and test collections.

Multilingual Information Access

—Carol Peters (IEI-CNR, Italy) and Paraic Sheridan (MNIS-TextWise Labs)

Provides an exhaustive overview and includes introductory material, a review of issues in multilingual text processing, techniques for cross-language text retrieval, evaluation of cross-language retrieval systems, and a study of some existing systems.

Text Summarization

—Dragomir Radev (University of Michigan)

Focuses on established and emerging techniques for producing automated summaries of text. The expected audience includes researchers and graduate students with interests in Information Retrieval and Natural Language Processing.

Tutorials — Afternoon

Combining Text and Link-Based Information Retrieval on the Web

—Andrei Broder (Altavista) and Prabhakar Raghavan (Verity)

Reviews the development of combined text and link-based retrieval, classification and clustering methods and examines what insights from web IR can (and can-

not) be abstracted back into enterprise settings. The tutorial is for researchers, students and professionals seeking an understanding of the similarities and differences between web and enterprise information retrieval, and it complements the one by Bharat and Hawking.

Text Classification and Text Mining in Vivo

—David Lewis (Independent Consultant)

Introduces practitioners and researchers to applications of text classification and related techniques that have emerged in text mining. The tutorial focuses on the priorities and constraints encountered in fielding this technology in operational settings. While it emphasizes automated approaches (particularly machine learning), it also discusses the tradeoffs between automated and more human-intensive methods.

Designing Information Architectures for Search

—Marti Hearst (Univ. California, Berkeley)

Introduces and explains a systematic approach to designing information architecture for web sites consisting of large collections of related information with the goals of producing websites with multiple views to reflect differences in user's preferred search and browsing methods, and incorporating search uniformly throughout the design of the site.

Practical Digital Libraries Overview

—Ed Fox (Virginia Polytechnic University)

Covers a variety of issues, including search, retrieval and resource discovery; multimedia/hypermedia; metadata (e.g., Dublin Core); electronic publishing; document models and representations; SGML and XML; database approaches; agents and distributed processing; 2D and 3D interfaces and visualizations; metrics; architectures and interoperability (e.g., OAI); commerce; educational (e.g., CSTC, NSDL, NDLTD) and social concerns; and intellectual property rights.

Papers

Paper Session 1A —*Summarization 1*
Applying Summarization Techniques for Term Selection in Relevance Feedback—Adnike Lam-Adesina and Gareth Jones (University of Exeter)

Temporal Summaries of News Topics—James Allan, Rahul Gupta, and Vikas Khandelwal (University of Massachusetts)

Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis—Yihong Gong and Xin Liu (NEC, C&C Research Laboratories)

A New Approach to Unsupervised Text Summarization—Tadashi Nomoto (National Institute of Japanese Literature) and Yuji Matsumoto (Nara Institute of Science and Technology)

Paper Session 1B —*Systems*

Vector-Space Ranking with Effective Early Termination—Vo Ngoc Anh, Owen Kretser, and Alistair Moffat (The University of Melbourne)

Static Index Pruning for Information Retrieval Systems—Aya Soffer (IBM Research Lab, Haifa), David Carmel (IBM Research Lab, Haifa), Doron Cohen (IBM Research Lab, Haifa), Ronald Fagin (IBM Almaden Research Center), Eitan Farchi (IBM Research Lab, Haifa), Michael Herscovici (IBM Research Lab, Haifa), and Yoelle S. Maarek (IBM Research Lab, Haifa)

Rank-Preserving Two-Level Caching for Scalable Search Engines—Patricia Correia Saraiva (Univ. Federal de Minas Gerais), Edleno Silva de Moura (Akwan Information Technologies, Brazil), Rodrigo Fonseca (Univ. Federal de Minas Gerais), Wagner Meira Jr. (Univ. Federal de Minas Gerais), Berthier Ribeiro-Neto (Univ. Federal de Minas Gerais), and Nivio Ziviani (Univ. Federal de Minas Gerais)

Using Event Segmentation to Improve Indexing of Consumer Photographs—Amanda Stent (University of Rochester) and Alexander Loui (Eastman Kodak Company)

Paper Session 2A —*Evaluation*

Ranking Retrieval Systems without Relevance Judgments—Ian Soboroff, Charles Nicholas, and Patrick Cahan (University of Maryland)

Evaluation by Highly Relevant Documents
—Ellen Voorhees (NIST)

Meta-scoring: Automatically Evaluating Term Weighting Schema in IR without Precision-Recall—Rong Jin, Christos Faloutsos, and Alex G. Hauptmann (Carnegie Mellon University)

Paper Session 2B —*Cross-Lingual*

Improving Cross Language Retrieval with Triangulated Translation—Tim Gollins and Mark Sanderson (University of Sheffield)

Improving Query Translation for CLIR using Statistical Models—Jianfeng Gao (Microsoft Research), Jian-Yun Nie (University of Montreal), Endong Xun (Microsoft Research), Jian Zhang (Tsinghua University), Ming Zhou (Microsoft Research), and Changning Huang (Microsoft Research)

Evaluating a Probabilistic Model for Cross-lingual Information Retrieval—Jinxi Xu, Ralph Weischedel, and Chanh Nguyen (BBN Technologies)

Paper Session 3 —*Language Models 1*

Document Language Models, Query Models, and Risk Minimization for Information Retrieval—John Lafferty and Chengxiang Zhai (Carnegie Mellon University)

Relevance-based Language Models—Victor Lavrenko and W. Bruce Croft (Univ. of Massachusetts)

Paper Session 4A —*Categorization*

A Statistical Learning Model of Text Classification with Support Vector Machines—Thorsten Joachims (GMD, Bonn)

A Study on Thresholding Strategies for Text Categorization—Yiming Yang (Carnegie Mellon Univ.)

On Feature Distributional Clustering for Text Categorization—Ron Bekkerman (The Technion, Haifa), Ran El-Yaniv (The Technion, Haifa), Naftali Tishby (The Hebrew University, Jerusalem), Yoad Winter (The Technion, Haifa)

An Empirical Study of Method Combination for Text Categorization and Routing—Hwee Tou Ng, Huey Ting Ang, and Wee Meng Soon (DSO National Laboratories)

Paper Session 4B —*Retrieval Models*
Iterative Residual Rescaling: An Analysis and Generalization of LSI—Rie Kubota Ando and Lillian Lee (Cornell University)

Expressive Retrieval from XML Documents
—Taurai T. Chinenyanga and Nicholas Kushmerick (University College Dublin)

XIRQL: A Query Language for Information Retrieval in XML Documents—Norbert Fuhr and Kai Grossjohann (University of Dortmund)

Empirical Investigations on Query Modification Using Abductive Explanations—Ian Ruthven (University of Glasgow), Mounia Lalmas (Queen Mary, University of London), Keith van Rijsbergen (University of Glasgow)

Paper Session 5A —*Summarization 2*
Generic Summaries for Indexing in Information Retrieval—Tetsuya Sakai and Karen Sparck Jones (University of Cambridge)

Automatic Generation of Concise Summaries of Spoken Dialogues in Unrestricted Domains
—Klaus Zechner (Carnegie Mellon University)

Enhanced Topic Distillation using Text, Markup Tags, and Hyperlinks—Soumen Chakrabarti, Mukul Joshi, and Vivek Tawde (IIT Bombay)

Paper Session 5B —User Studies

Transparent Queries: Investigating Users' Mental Models of Search Engines—Jack Muramatsu and Wanda Pratt (University of California, Irvine)

Why Batch and User Evaluations Do Not Give the Same Results—Andrew Turpin (Curtin University of Technology, Perth) and William Hersh (Oregon Health Sciences University)

Evaluating a Content Based Image Retrieval System—Sharon McDonald (University of Sunderland) Ting Sheng Lai (National Palace Museum), and John Tait (University of Sunderland)

Paper Session 6 —The Web

Evaluating Topic-Driven Web Crawlers—Filippo Menczer (University of Iowa), Gautam Pant (University of Iowa), Padmini Srinivasan (University of Iowa), and Miguel Ruiz (TextWise)

Effective Site Finding using Link Anchor Information—Nick Craswell (CSIRO, Canberra), David Hawking (CSIRO, Canberra), and Stephen Robertson (Microsoft Research)

Stable Eigenvector Algorithms for Link Analysis—Alice Zheng, Andrew Ng, and Michael Jordan (U.C. Berkeley)

Paper Session 7A —Meta-Search

Modeling Score Distributions for Combining the Outputs of Search Engines—R. Manmatha, T. Rath, and Fangfang Feng (University of Massachusetts)

Models for Metasearch—Javed Aslam and Mark Montague (Dartmouth College)

The Score-Distributional Threshold Optimization for Adaptive Binary Classification Tasks—Avi Arampatzis and Andre van Hameren (Katholieke Universiteit Nijmegen)

Maximum Likelihood Estimation for Filtering Threshold—Yi Zhang and Jamie Callan (Carnegie Mellon University)

Paper Session 7B —*Learning*

A Meta-Learning Approach for Text Categorization—Wai Lam and Kwok-Yin Lai (The Chinese University of Hong Kong)

Unsupervised and Supervised Clustering for Topic Tracking—Jeffrey McCarley, Martin Franz, Todd Ward, and Wei-Jing Zhu (IBM TJ Watson Research Center)

Intelligent Information Triage—Sofus Macskassy (Rutgers University), Haym Hirsh (Rutgers University), Foster Provost (NYU Stern School of Business), Ramesh Sankaranarayanan (NYU Stern School of Business), and Vasant Dhar (NYU Stern School of Business)

Discovering Information Flow Using a High Dimensional Conceptual Space—Dawei Song and Peter Bruza (The University of Queensland)

Paper Session 8A —*Language Models 2*

A Study of Smoothing Methods for Language Models Applied to ad hoc Information Retrieval
—Chengxiang Zhai and John Lafferty (Carnegie Mellon University)

Topic Segmentation with an Aspect Hidden Markov Model—David Blei (U.C. Berkeley) and Pedro Moreno (Compaq Research)

Finding Topic Words for Hierarchical Summarization—Dawn Lawrie, W. Bruce Croft, and Arnold Rosenberg (University of Massachusetts)

Paper Session 8B —*Question Answering*

Exploiting Redundancy in Question Answering
—Charles Clarke, Gordon V. Cormack, and Thomas R. Lynam (University of Waterloo)

High Performance Question Answering—Marius Pasca and Sanda Harabagiu (Southern Methodist University)

Searcher Performance in Question Answering
—Ross Wilkinson (CSIRO, Melbourne), MingFang Wu (CSIRO, Melbourne), Michael Fuller (RMIT, Melbourne)

Posters

Cite Me, Cite My References?—Elana Broch
(Princeton University)

Construction of a Hierarchical Classifier Schema using a Combination of Text-Based and Image-Based Approaches—Mark Drew and Cheng Lu (Simon Fraser University)

Comparison Between a Method Based on the Chi-Square Test and a Support Vector Machine for Document Classification—Robert Gaizauskas, M. Oakes, M. Beaulieu, H. Fowkes, Anna Jonsson, and Vincent Wan, (University of Sheffield)

Automatic Web Search Query Generation to Create Minority Language Corpora—Rayid Ghani (Carnegie Mellon University), Rosie Jones (Carnegie Mellon University), and Dunja Mladenic (J. Stefan Institute and Carnegie Mellon University)

Intelligent Object-Based Image retrieval using Cluster-Driven Personal Preference Learning—Kyoung-Mi Lee and W.N. Street (University of Iowa)

Perceptual Consistency Improves Image Retrieval Performance—Wee Leow, Huizhong Long, and Wee Kheng Leow (National University of Singapore)

Query Clustering Using Content Words and User Feedback—Jian-Yun Nie (University of Montreal), Jirong Wen (Microsoft Research China), and Hongjiang Zhen (Microsoft Research China)

Modifications of Kleinberg's HITS Algorithm using Matrix Exponentiation and Web Log Records—Lesley Ward, Ayman Farahat, Thomas LoFaro, Joel Miller, Gregory Rae, and Fred Schaefer (Harvey Mudd College)

Analyses of Multiple-Evidence Combinations for Retrieval Strategies—Abdur Chowdhury (Illinois Institute of Technology)

Text Summarization via Hidden Markov Models—John Conroy and Dianne P. O'Leary (IDA/CCS, Bowie, MD)

Towards the Use of Prosodic Information for Spoken Document Retrieval — Fabio Crestani (University of Strathclyde)

Probabilistic Combination of Content and Links — Susan Dumais (Microsoft Research) and Rong Jin (Carnegie Mellon University)

Query Optimizations for Vector Space Problems — Kazuo Goda (The University of Tokyo), Takayuki Tamura (Mitsubishi Electric Corporation), Masaru Kitsuregawa (The University of Tokyo), Ophir Frieder (The Illinois Institute of Technology), and Abdur Chowdhury (The Illinois Institute of Technology)

Toward an Improved Concept-Based Information Retrieval System — Peter V. Henstock (MIT Lincoln Laboratory), Daniel J. Pack (MIT Lincoln Laboratory/USAF Academy), Young-Suk Lee (MIT Lincoln Laboratory), and Clifford J. Weinstein (MIT Lincoln Laboratory)

Query Expansion Based on Predictive Algorithms for Collaborative Filtering — Keiichiro Hoashi, Kazunori Matsumoto, Naomi Inoue, and Kazuo Hashimoto (KDD R&D Laboratories, Inc.)

Unitary Operators for Fast Latent Semantic Indexing (FLSI) — Eduard Hoenkamp (Nijmegen Institute for Cognition and Information)

Reading Time, Scrolling and Interaction: Exploring Implicit Sources of User Preferences for Relevance Feedback During Interactive Information Retrieval — Diane Kelly and Nicholas J. Belkin (Rutgers University)

An Homogeneous Framework to Model Relevance Feedback — David Losada and Alvaro Barreiro (University of A Corunna, Spain)

Anchor Text Mining for Translation Extraction of Query Terms — Wen-Hsinag Lu (National Chiao Tung University), Lee-Feng Chien (Institute of Information Science, Academia Sinica), and Hsi-Jian Lee (National Chiao Tung University)

Selecting Expansion Terms in Automatic Query Expansion —Hiroko Mano and Yasushi Ogawa (Ricoh Company, Ltd.)

Quantifying the Utility of Parallel Corpora —Jeffrey McCarley, Martin Franz, Todd Ward, and Wei-Jing Zhu (IBM TJ Watson Research Center)

An Experimental Framework for Email Categorization and Management —Kenrick Mock (University of Alaska, Anchorage)

Generic Topical Segmentation of Document Texts —Marie-Francine Moens and Rik De Busser (Katholieke Universiteit Leuven)

Metasearch Consistency —Javad Aslam and Mark Montague (Dartmouth College)

Browsing in a Digital Library Collecting Linearly Arranged Documents —Makoto Nakashima, Yanhua Qu, Keizo Sato, and Tetsuro Ito (Oita University)

Feature Selection for Polyphonic Music Retrieval —Jeremy Pickens (University of Massachusetts)

Structural and Content-Based Segmentation of Speech Transcripts —Dulce Ponceleon and Savitha Srinivasan (IBM Almaden Research Center)

Flexible Pseudo-Relevance Feedback Using Optimization Tables —Tetsuya Sakai (University of Cambridge/Toshiba) and Stephen E. Robertson (Microsoft Research Cambridge)

A Data Fusion Experiment Involving a Combined Document Representation —Nicola Stokes and Joe Carthy (University College Dublin, Ireland)

Interactive Phrase Browsing Within Compressed Text —Raymond Wan and Alistair Moffat (University of Melbourne)

Query-Biased Web Page Summarisation: A Task-Oriented Evaluation —Ryen White (University of Glasgow)

Automatic Information Extraction from Web Pages—Roland Yap and Budi Rahardjo (National University of Singapore)

Demonstrations

Find: A Web Image Search Engine—Zheng Chen, Liu Wenyin, Chunhui Hu, Mingjing, and Hongjiang Zhang (Microsoft Research China)

Building Interoperable Digital Library Services: MARIAN, Open Archives and NDLTD—Edward A. Fox, Robert France, Marcos André Goncalves, and Hussein Suleman (Virginia Tech)

AUTINDEX: An Automatic Multilingual Indexing System—Barbel Ripplinger (IAI, Saarbrücken) and Paul Schmidt (University Mainz)

Does Visualization Improve Our Ability to Find and Learn from Internet Based Information?
—Daniel Kauwell, Jim Levin, Hwan Jo Yu, Young Jin Lee, and Arun Bhalla (Univ. of Illinois)

The Retrieval Platform HySpirit—Thomas Roelleke, Ralf Lubeck, and Carsten Wich (HySpirit, Dortmund)

Distributed Resource Discovery and Structured Data Searching with Cheshire II—Ray Larson (University of California, Berkeley)

Searching the Deep Web – Distributed Explorer Directed Query Engine Application—Valerie Allen, and Abe Lederman (DoE Office of Scientific and Technical Information)

CROWSE: A System for Organizing Repositories and Web Search Results—Sudeshna Sarkar and Kinshuman (IIT, Kharagpur)

MS Read: User Modeling in the Web Environment—Natasa Milic-Frayling and Ralph Sommerer (Microsoft Research UK)

Workshops

Open Archives: Communities, Interoperability, and Services —Ed Fox (Virginia Polytechnic Institute and State University)

The Open Archives Initiative "<http://www.openarchives.org>" develops and promotes standards that aim to facilitate the efficient dissemination of content regardless of the type of content offered. Its goal is to serve communities wishing to share information by ensuring interoperability and componentized, layered services. OAI was launched in October 1999 to provide a forum to discuss and solve problems of interoperability among author self-archiving solutions. OAI aims to support archives, both those focused on e-prints (e.g., theses and dissertations, Web log files, and educational resources). The emphasis has been on allowing harvesting of metadata that describes diverse "records" of content, stored in managed repositories. This workshop will allow those involved in the OAI, and those wishing to become involved, to extend the Initiative through sharing of technology, description and demonstration of services, and community-based discussion of conventions that ensure interoperability. The workshop will include an introduction to OAI and provide technology sharing and community building opportunities.

Mathematical/Formal Methods in IR

—Sandor Dominich (University of Veszprem), Mounia Lalmas (Queen Mary, University of London) and Keith van Rijsbergen (University of Glasgow)

The previous workshop (ACM SIGIR 2000 MF/IR 2000 Workshop, Athens, Greece) showed that the mathematical/formal results achieved in Information Retrieval (IR) could be organized into a coherent theoretical framework, that they brought new knowledge to IR, and that mathematical/formal research in IR can stand as a specialized research area of IR. The purpose of the MF/IR 2001 workshop is, on the one hand, to continue and enhance the results obtained so far, and on the other hand, to present, discuss, analyze, integrate the newer/newest results. Therefore, MF/IR 2001 aims at promoting discussion and interaction among those with theoretical and applicative research interests in mathematical/formal aspects of Information Retrieval, and also at being a

forum for the presentation of both theoretical and applicative results (e.g., foundational issues; description and/or integration of models; retrieval applications; mathematical/formal techniques, properties and structures in IR; exigent and/or new theories and theoretical aspects).

Text Summarization

— Donna Harman (National Institute of Standards and Technology) and Daniel Marcu (University of Southern California)

Summarization technology has the potential of adding significant value in the context of information retrieval applications. Summarization can provide, for example, an alternative display mode for retrieved documents; a new method of concentrating information for relevance feedback; and a means for presenting information specific to groups of related documents and web sites. There has been a long history of research in this area by both the retrieval and the natural language processing communities, with summarization papers being presented at SIGIR and ACL. However we still do not know what summarization techniques are most adequate for which purposes and what evaluation techniques are most appropriate for assessing the quality of a summary. The purpose of this workshop is two-fold. The first day of the workshop will serve as a focal point for presenting new results in this area. This will include invited presentations focusing on various open problems in summarization research, presentations of original scientific papers, and an overview of the goals and results from a new evaluation effort in summarization called DUC (Document Understanding Conference, "<http://www-nlpir.nist.gov/projects/duc/>"). The optional second day of the workshop will be devoted to presenting more detailed results from the first DUC evaluation and to discussing plans for the future of DUC.

IR Techniques for Speech Applications

— Anni Coden, Savitha Srinivasan, and Eric Brown (IBM T.J. Watson Research Center)

In the last few years automatic speech recognition has left the confines of the basic research lab and become a viable commercial application. Speech recognition tech-

nology has now matured to the point where speech can be used to interact with automated phone systems, control computer programs, and even create memos and documents. Moving beyond computer control and dictation, speech recognition has the potential to dramatically change the way we create, capture, and store knowledge. Advances in speech recognition technology combined with ever decreasing storage costs and processors that double in power every eighteen months have set the stage for a whole new era of applications that treat speech in the same way that we currently treat text. The goal of this workshop is to explore the technical issues involved in applying information retrieval and text analysis technologies in the new application domains enabled by automatic speech recognition.

Specifically, we would like to focus on: 1) What new IR related applications, problems, or opportunities are created by effective, real-time speech recognition? 2) To what extent are information retrieval methods that work on perfect text applicable to imperfect speech recognition? 3) What additional data representations from a speech engine may be exploited by applications? 4) Does domain knowledge (context/voice-id) help and can it be automatically deduced? 5) Can some of the techniques explored be beneficial in a standard IR application? 6) What constraints are imposed by real time speech applications? 7) Case studies of specific speech applications — either successful or not.

Operational Text Classification

— David Lewis (Independent Consultant) Susan Dumais (Microsoft) Ronen Feldman (Clearforest) and Fabrizio Sebastiani (CNR/IEI)

Text classification research and practice has grown dramatically during the last decade. Text classification algorithms have been discussed at numerous conferences, but less is known about the issues that arise in deploying operational text classification systems. The goal of our workshop is to bring together practitioners and researchers to discuss these issues. Topics include the costs and benefits of text classification systems, system architecture, resource usage, maintaining and modifying of classifiers over time, integration of automated and manual procedures, combining of prior knowledge with machine learning, and so on.

Recommender Systems

— John Herlocker (Oregon State University), Joaquin Delgado (TripleHop Technologies), David MacDonald (FX Palo Alto Laboratories), Douglas Oard (University of Maryland), and Ian Soboroff (University of Maryland-Baltimore County)

Recommender systems assist and augment the transfer of recommendations among members of a community. Recommendations can describe content in a way that is complementary to keyword terms and to metadata. A typical system collects preferences and opinions from individual users, then aggregates and transfers those as recommendations to other members of the community. The goal of this workshop is to bring together researchers and practitioners involved in the development, analysis, and deployment of recommender systems. This workshop will provide a forum for discussing current and recent research results, and develop a road-map for future recommender systems research.

Social Events

- There will be **coffee breaks** at 10:00–10:30 am and at 3:30–4:00 pm on Monday, Tuesday, and Wednesday for the SIGIR Conference.
- On Monday night, there will be a **reception** at 5:30–10:00 pm, which will include the **Poster Session** and the **Demonstration Sessions**.
- On Tuesday night, there will be the **SIGIR Banquet**, at 6:00–10:00 pm.
- There will be a **box lunch** on Wednesday at 12:30–3:00 pm for the **ACM/SIGIR Business Meeting**.
- On Thursday, following the SIGIR Conference, there will be the **Workshops**, where a **lunch** will be provided.
- In addition, for those who are in the **optional** extra **Workshop** on Friday, **lunch** will be provided.

Conference Location

Located near the mouth of the Mississippi River, New Orleans in the state of Louisiana, USA, is a big city with a lot going on. The city “that care forgot” is the home of “A Streetcar Named Desire,” that you can actually take from the famous French Quarter uptown past the Garden District, the Audubon Zoo, Tulane and Loyola University, and more.

The conference hotel, the Hotel Radisson New Orleans, is located on Canal Street, just above the French Quarter, which is walking distance, although the hotel does offer shuttle service during the day and evening. New Orleans is home of great music, including Dixieland jazz, modern jazz, blues, rock and roll, Cajun, and much, much more. New Orleans is home of great food, both Cajun and Creole, with a plethora of fantastic restaurants. Moreover, there are a variety of clubs, including Pat O’Brien’s “home of the famous Hurricane” that are a must. And, right next to Pat O’Brien’s is the famous Preservation Hall, where \$1 gains you admission to the greats of Dixieland jazz just doing their thing as great as they always do.

New Orleans can offer access to swamp tours and tours of the antebellum mansions, as well as cruises up and down the Mississippi River. The concierge at the conference hotel will be more than glad to help you make reservations, for dinner, for entertainment, for tours, and more.

Check out the conference web site,

<http://www.sigr2001.org>

to find links to the best that New Orleans has to offer.

Travel

The New Orleans International Airport is served by most major airlines. ACM has arranged for discounts on the following airlines:

Delta –contact for information:

http://www.acm.org/sig_volunteer_info/delta.html

US Air –contact for information:

http://www.acm.org/sig_volunteer_info/usair.html

New Orleans can be reached by Amtrak train, recalling the famous song celebrating the “City of New Orleans” train from Chicago down to New Orleans. By car, New Orleans is reached via the interstate I-10 highway.



Weather

New Orleans in September can be warm, and as always, a bit humid. The normal high in September is 88°F / 31°C, and the average low is 70°F / 21°C, with an average of 5.5 inches of rain for the entire month. Bring clothing that is light, perhaps an umbrella, just in case. Incidentally, there are very few places (restaurants or clubs) that are so formal as to require a tie and jacket or skirt or dress.

Lodging

Conference Hotel:

Radisson Hotel New Orleans

1500 Canal Street

New Orleans, Louisiana 70112-2818 USA

The conference will be held in its entirety in the conference hotel. There are complimentary shuttles to go to the French Quarter (Vieux Carré) and back to the conference hotel during much of the day and evening, although you can walk if you so choose. The concierge at the conference hotel will be glad to help you with plans for the great food, fantastic music and entertainment, and cultural activities, in the French Quarter, in the Garden District, downtown, uptown, and all around.

If you are unable to make reservations at the conference hotel, check the conference web site, <http://www.sigir2001.org> for nearby alternatives.

Reservations at the conference hotel must be made directly with the hotel, but you may use the registration form on the following page.



Attending the **ACM/SIGIR 2001 Conference**
September 9-13, 2001

(Conference Rate: \$129.00 per night, plus tax,
for single, double, triple or quad)

Name (first, middle, last):

Address:

City:

State: _____ Country: _____

Zip Code: _____

Arrival Date: _____ Departure: _____

Special Requests (bed type, floor, smoking/non-smoking,
handicap access, etc.):

Number of people in room (circle one):

single double triple quad

- For Reservations by Fax: (504) 523-9617
- For Reservations by Telephone direct to hotel:
(504) 522-4500 Please mention you are attending
the ACM/SIGIR '01 Conference and request the
conference rate.
- For Reservations by Toll Free Number (US only):
1-800-333-3333
- For Reservations by Mail:
Radisson Hotel New Orleans – Reservations
1500 Canal Street
New Orleans, LA 70112
- Guaranteed Reservation (circle one):
American Express Master Card VISA Check

Credit Card Number: _____

Expiration Date: _____



