



ACM/SIGIR

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

September 9-14, 2001

New Orleans
LOUISIANA, USA

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



FINAL PROGRAM

Conference Schedule

Sunday, September 9, 2001 Pre-Conference

8:30-12:30 Tutorials

1:30-5:30 Tutorials

Monday, September 10 Main Conference

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

10:30-12:30 Paper Session 1A and 1B

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

4:00-5:00 Paper Session 3

6:00-10:00 Poster Sessions, Demonstrations,
and Reception

Tuesday, September 11 Main Conference

9:00-10:00 Keynote Address

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

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

4:00-5:30 Paper Session 6

7:00-10:00 Banquet

Wednesday, September 12 Main Conference

9:00-10:00 Keynote Address

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

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

3:00-4:30 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 Text
Summarization Workshop

The Radisson Hotel New Orleans



Social Events

- There will be **coffee breaks** at 10:00–10:30 a.m. Monday to Wednesday, and at 3:30–4:00 p.m. on Monday and Tuesday for the SIGIR Conference.
- On Monday night, there will be a **reception** at 6:00–10:00 p.m., which will include the **Poster Sessions** and the **Demonstration Sessions**.
- On Tuesday night, there will be the **SIGIR Banquet**, at 7:00–10:00 p.m.
- There will be a **box lunch** on Wednesday at 12:30–3:00 p.m. for the **ACM/SIGIR Business Meeting**.
- On Thursday, following the SIGIR Conference, there will be the **Workshops**, where a **box lunch** will be provided.
- In addition, for those who are in the **optional Workshop** on Friday, **box lunch** will be provided.

Registration and

Information Desk Hours

Sunday, Sept. 9 – 7:00 a.m. to 6:00 p.m.

Monday, Sept. 10 – 8:00 a.m. to 6:00 p.m.

Tuesday, Sept. 11 – 8:00 a.m. to 6:00 p.m.

Wednesday, Sept. 12 – 8:00 a.m. to 5:00 p.m.

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

Edie Rasmussen, University of Pittsburgh

Local Arrangements Chair

Carol Barry, Louisiana State University

Information Architect

David Robins, University of Pittsburgh

Sunday, September 9

TUTORIALS

8:30 a.m. – 12:30 p.m. Sessions

Introduction to Information Retrieval

Mark Sanderson (University of Sheffield)

Oak Room, Conference Center

Web Information Retrieval

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

Magnolia Room, Conference Center

Multilingual Information Access

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

Mimosa Room, Conference Center

Text Summarization

Dragomir Radev (University of Michigan)

Cypress Room, Conference Center

1:30 p.m. – 5:30 p.m. Sessions

***Combining Text and Link-Based
Information Retrieval on the Web***

Andrei Broder (Altavista) and Prabhakar Raghavan
(Verity)

Magnolia Room, Conference Center

***Text Classification and Text Mining
in Vivo***

David Lewis (Independent Consultant)

Mimosa Room, Conference Center

***Designing Information Architectures
for Search***

Marti Hearst (Univ. California, Berkeley)

Cypress Room, Conference Center

Practical Digital Libraries Overview

Ed Fox (Virginia Polytechnic University)

Oak Room, Conference Center

Monday, September 10

9:00 a.m. – 10:00 a.m.

Opening Session/ Keynote Address

Chair: W. Bruce Croft (U. of Massachusetts)

Grand Ballroom, Salon A

Conference Opening:

Donald Kraft (Louisiana State U.)

Keynote Address: *Mining the Web*

Tom Mitchell (WhizBang!)

10:30 a.m. – 12:30 p.m.

PARALLEL PAPER SESSIONS

Session 1A —Summarization 1

Chair: Elizabeth Liddy (Syracuse U.)

Grand Ballroom, Salon A

***Applying Summarization Techniques for
Term Selection in Relevance Feedback***

Adenike Lam-Adesina and Gareth Jones (U. of Exeter)

Temporal Summaries of News Topics

James Allan, Rahul Gupta, and Vikas Khandelwal
(U. 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)

Session 1B —Systems

Chair: Ricardo Baeza-Yates (U. of Chile)

Grand Ballroom, Salon B

Vector-Space Ranking with Effective Early Termination

Vo Ngoc Anh, Owen Kretser, and Alistair Moffat
(The U. of Melbourne)

Static Index Pruning for Information Retrieval Systems

Aya Soffer (IBM Research Lab, Haifa), David Carmel (IBM Research), Doron Cohen (IBM Research), Ronald Fagin (IBM Almaden Research Center), Eitan Farchi (IBM Research), Michael Herscovici (IBM Research), and Yoelle S. Maarek (IBM Research)

Rank-Preserving Two-Level Caching for Scalable Search Engines

Patricia Correia Saraiva (U. Federal de Minas Gerais), Edleno Silva de Moura (Akwan Information Technologies, Brazil), Rodrigo Fonseca (U. Federal), Wagner Meira Jr. (U. Federal), Berthier Ribeiro-Neto (U. Federal), and Nivio Ziviani (U. Federal)

Using Event Segmentation to Improve Indexing of Consumer Photographs

Amanda Stent (U. of Rochester) and Alexander Loui (Eastman Kodak Company)

2:00 p.m. – 3:30 p.m.

PARALLEL PAPER SESSIONS

Session 2A —Evaluation

Chair: Pia Borlund (Royal School of Library and Information Science)

Grand Ballroom, Salon A

Ranking Retrieval Systems without Relevance Judgments

Ian Soboroff, Charles Nicholas, and Patrick Cahan
(U. 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)

Session 2B —Cross-Lingual

Chair: Sung Hyon Myaeng (Chungnam National U.)

Grand Ballroom, Salon B

***Improving Cross Language Retrieval with
Triangulated Translation***

Tim Gollins and Mark Sanderson (U. of Sheffield)

***Improving Query Translation for CLIR
Using Statistical Models***

Jianfeng Gao (Microsoft Research), Jian-Yun Nie
(U. of Montreal), Endong Xun (Microsoft), Jian
Zhang (Tsinghua U.), Ming Zhou (Microsoft),
and Changning Huang (Microsoft)

***Evaluating a Probabilistic Model for
Cross-lingual Information Retrieval***

Jinxi Xu, Ralph Weischedel, and Chanh Nguyen
(BBN Technologies)

4:00 p.m. – 5:00 p.m.

PAPER SESSION

Session 3 —Language Models 1

Chair: Norbert Fuhr (U. of Dortmund)

Grand Ballroom, Salon A

***Document Language Models, Query Models,
and Risk Minimization for Information
Retrieval***

John Lafferty and Chengxiang Zhai (Carnegie
Mellon U.)

Relevance-based Language Models

Victor Lavrenko and W. Bruce Croft (U. of Massa-
chusetts)

6:00 p.m. – 10:00 p.m.

Poster Sessions, Demonstrations
and Reception

POSTER SESSIONS

Grand Foyer (outside Grand Ballroom)

Cite Me, Cite My References?

Elana Broch (Princeton U.)

Construction of a Hierarchical Classifier Schema using a Combination of Text-Based and Image-Based Approaches

Mark Drew and Cheng Lu (Simon Fraser U.)

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, (U. of Sheffield)

Automatic Web Search Query Generation to Create Minority Language Corpora

Rayid Ghani (Carnegie Mellon U.), Rosie Jones (Carnegie Mellon U.), and Dunja Mladenic (J. Stefan Institute and Carnegie Mellon U.)

Intelligent Object-Based Image Retrieval using Cluster-Driven Personal Preference Learning

Kyoung-Mi Lee and W.N. Street (U. of Iowa)

Perceptual Consistency Improves Image Retrieval Performance

Wee Leow, Huizhong Long, and Wee Kheng Leow (National U. of Singapore)

Query Clustering Using Content Words and User Feedback

Jian-Yun Nie (U. 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, Ophir Frieder, David Grossman,
Catherine McCabe (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 (U. of Strathclyde)

***Probabilistic Combination of Content
and Links***

Susan Dumais (Microsoft Research) and Rong Jin
(Carnegie Mellon U.)

***Query Optimizations for Vector Space
Problems***

Kazuo Goda (The U. of Tokyo), Takayuki Tamura
(Mitsubishi Electric Corporation), Masaru
Kitsuregawa (The U. 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 Lab/USAF Academy), Young-
Suk Lee (MIT Lincoln Lab), and Clifford J.
Weinstein (MIT Lincoln Lab)

***Query Expansion Based on Predictive
Algorithms for Collaborative Filtering***

Keiichiro Hoashi, Kazunori Matsumoto, Naomi
Inoue, and Kazuo Hashimoto (KDD R&D Laborato-
ries, Inc.)

***Unitary Operators for Fast Latent Semantic
Indexing (FLSI)***

Eduard Hoenkamp (Nijmegen Institute for Cogni-
tion and Information)

***Reading Time, Scrolling and Interaction:
Exploring Implicit Sources of User Prefer-
ences for Relevance Feedback During
Interactive Information Retrieval***

Diane Kelly and Nicholas J. Belkin (Rutgers U.)

***An Homogeneous Framework to Model
Relevance Feedback***

David Losada and Alvaro Barreiro (U. of A Corunna,
Spain)

***Anchor Text Mining for Translation
Extraction of Query Terms***

Wen-Hsinag Lu (National Chiao Tung U.), Lee-Feng
Chien (Institute of Information Science, Academia
Sinica), and Hsi-Jian Lee (National Chiao Tung U.)

***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 (U. 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 U.)

***Feature Selection for Polyphonic Music
Retrieval***

Jeremy Pickens (U. of Massachusetts)

***Structural and Content-Based Segmenta-
tion of Speech Transcripts***

Dulce Ponceleon and Savitha Srinivasan (IBM
Almaden Research Center)

***Flexible Pseudo-Relevance Feedback Using
Optimization Tables***

Tetsuya Sakai (U. of Cambridge/Toshiba) and
Stephen E. Robertson (Microsoft Research
Cambridge)

***A Data Fusion Experiment Involving a
Combined Document Representation***

Nicola Stokes and Joe Carthy (U. College Dublin,
Ireland)

***Interactive Phrase Browsing Within
Compressed Text***

Raymond Wan and Alistair Moffat (U. of Melbourne)

***Query-Biased Web Page Summarisation:
A Task-Oriented Evaluation***

Ryen White (U. of Glasgow)

***Automatic Information Extraction from
Web Pages***

Roland Yap and Budi Rahardjo (National U. of
Singapore)

DEMONSTRATION SESSIONS

Oak, Magnolia, Mimosa and Cypress
Rooms

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 (U. Mainz)

***Does Visualization Improve Our Ability
to Find and Learn from Internet Based
Information?***

Daniel Kauwell, Jim Levin, Hwan Jo Yu, Young Jim
Lee, and Arun Bhalla (U. of Illinois)

The HySpirit Retrieval Platform

Thomas Roelleke (HySpirit, Dortmund and Queen
Mary U. of London), Ralf Lubeck (HySpirit,

Dortmund), and Gabriella Kazai (Queen Mary U. of London)

Distributed Resource Discovery and Structured Data Searching with Cheshire II

Ray Larson (U. 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)

Tuesday, September 11

9:00 a.m. – 10:00 a.m.

Chair: Justin Zobel (RMIT U.)

Grand Ballroom, Salon A

Keynote Address: ***Wireless Graffiti & Annotating Physical Space with Digital Information***

Tomasz Imielinski (Rutgers University)

10:30 a.m. – 12:30 p.m.

PARALLEL PAPER SESSIONS

Session 4A —Categorization

Chair: David Lewis (Independent Consultant)

Grand Ballroom, Salon A

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 U.)

On Feature Distributional Clustering for Text Categorization

Ron Bekkerman (The Technion, Haifa), Ran El-Yaniv (The Technion), Naftali Tishby (The Hebrew University, Jerusalem), Yoad Winter (The Technion)

Session 4B —Retrieval Models

Chair: Mounia Lalmas (Queen Mary, U. of London)

Grand Ballroom, Salon B

Iterative Residual Rescaling: An Analysis and Generalization of LSI

Rie Kubota Ando and Lillian Lee (Cornell U.)

Expressive Retrieval from XML Documents

Taurai T. Chinenyanga and Nicholas Kushmerick (U. College Dublin)

XIRQL: A Query Language for Information Retrieval in XML Documents

Norbert Fuhr and Kai Grossjohann (U. of Dortmund)

Empirical Investigations on Query Modification Using Abductive Explanations

Ian Ruthven (U. of Glasgow), Mounia Lalmas (Queen Mary, U. of London), Keith van Rijsbergen (U. of Glasgow)

2:00 p.m. – 3:30 p.m.

Session 5A —Summarization 2

Chair: Claire Cardie (Cornell U.)

Grand Ballroom, Salon A

Generic Summaries for Indexing in Information Retrieval

Tetsuya Sakai and Karen Sparck Jones (U. of Cambridge)

Automatic Generation of Concise Summaries of Spoken Dialogues in Unrestricted Domains

Klaus Zechner (Carnegie Mellon U.)

Enhanced Topic Distillation using Text, Markup Tags, and Hyperlinks

Soumen Chakrabarti, Mukul Joshi, and Vivek Tawde
(IIT Bombay)

Session 5B —User Studies

Char: Gary Marchionini (U. of North Carolina,
Chapel Hill)

Grand Ballroom, Salon B

***Transparent Queries: Investigating Users'
Mental Models of Search Engines***

Jack Muramatsu and Wanda Pratt (U. of California,
Irvine)

***Why Batch and User Evaluations Do Not
Give the Same Results***

Andrew Turpin (Curtin U. of Technology, Perth)
and William Hersh (Oregon Health Sciences U.)

***Evaluating a Content Based Image
Retrieval System***

Sharon McDonald (U. of Sunderland), Ting Sheng
Lai (National Palace Museum), and John Tait
(U. of Sunderland)

4:00 p.m. – 5:30 p.m.

Session 6 —The Web

Chair: David Hawking (CSIRO)

Grand Ballroom, Salon A

Evaluating Topic-Driven Web Crawlers

Filippo Menczer (U. of Iowa), Gautam Pant
(U. of Iowa), Padmini Srinivasan (U. of Iowa),
and Miguel Ruiz (TextWise)

***Effective Site Finding using Link Anchor
Information***

Nick Craswell (CSIRO, Canberra), David Hawking
(CSIRO), and Stephen Robertson (Microsoft
Research)

***Stable Eigenvector Algorithms for Link
Analysis***

Alice Zheng, Andrew Ng, and Michael Jordan (U.C.
Berkeley)

7:00 p.m. – 10:00 p.m.

BANQUET

Tulane Room, Mezzanine

Wednesday, September 12

9:00 a.m. – 10:00 a.m.

Chair: David Harper (Robert Gordon U.)

Grand Ballroom, Salon A

Keynote Address: *Going Beyond Extract Based Text Summarization*

Eduard Hovy (USC Information Sciences Institute)

10:30 a.m. – 12:30 p.m.

PARALLEL PAPER SESSIONS

Session 7A —Meta-Search

Chair: Mark Sanderson (U. of Sheffield)

Grand Ballroom, Salon A

Modeling Score Distributions for Combining the Outputs of Search Engines

R. Manmatha, T. Rath, and Fangfang Feng (U. 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 U.)

Session 7B —Learning

Chair: Howard Turtle (Chilid Publishing)

Grand Ballroom, Salon B

A Meta-Learning Approach for Text Categorization

Wai Lam and Kwok-Yin Lai (The Chinese U. 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 U.), Haym Hirsh (Rutgers U.), Foster Provost (NYU Stern School of Business), Ramesh Sankaranarayanan (NYU Stern School), and Vasant Dhar (NYU Stern School)

Discovering Information Flow Using a High Dimensional Conceptual Space

Dawei Song and Peter Bruza (U. of Queensland)

12:30 p.m. – 3:00 p.m.

ACM/SIGIR MEETING AND LUNCH
Grand Ballroom, Salon A

3:00 p.m. – 4:30 p.m.

Session 8A —Language Models 2

Chair: Keith van Rijsbergen (U. of Glasgow)

Grand Ballroom, Salon A

A Study of Smoothing Methods for Language Models Applied to ad hoc Information Retrieval

Chengxiang Zhai, John Lafferty (Carnegie Mellon U.)

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 (U. of Massachusetts)

Session 8B —Question Answering

Chair: Amit Singhal (Google, Inc.)

Grand Ballroom, Salon B

Exploiting Redundancy in Question Answering

Charles Clarke, Gordon V. Cormack, and Thomas R. Lynam (U. of Waterloo)

High Performance Question Answering

Marius Pasca and Sanda Harabagiu (Southern Methodist U.)

Searcher Performance in Question Answering

Ross Wilkinson (CSIRO, Melbourne), MingFang Wu (CSIRO), and Michael Fuller (RMIT, Melbourne)

Thursday, September 13

8:00 a.m. – 5:00 p.m.

WORKSHOPS

Open Archives: Communities, Interoperability, and Services

Ed Fox (Virginia Polytechnic Institute and State U.)
Oak Room, Conference Center

Mathematical/Formal Methods in IR

Sandor Dominich (U. of Veszprem), Mounia Lalmas (Queen Mary, U. of London) and Keith van Rijsbergen (U. of Glasgow)
Mimosa Room, Conference Center

Text Summarization

Donna Harman (National Institute of Standards and Technology) and Daniel Marcu (U. of Southern California)
Magnolia Room, Conference Center

IR Techniques for Speech Applications

Anni Coden, Savitha Srinivasan, and Eric Brown (IBM T.J. Watson Research Center)
Fleur de Lis Room 1, Conference Center

Operational Text Classification

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

Cypress Room, Conference Center

Recommender Systems

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

Fleur de Lis Room 2, Conference Center

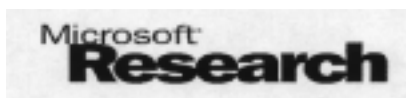
Friday, September 14

8:00 a.m. – 5:00 p.m.

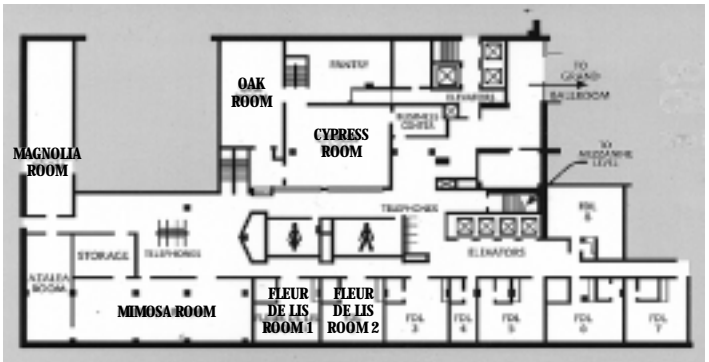
**Optional Second Day for Text Summarization
Workshop**

Location to be announced

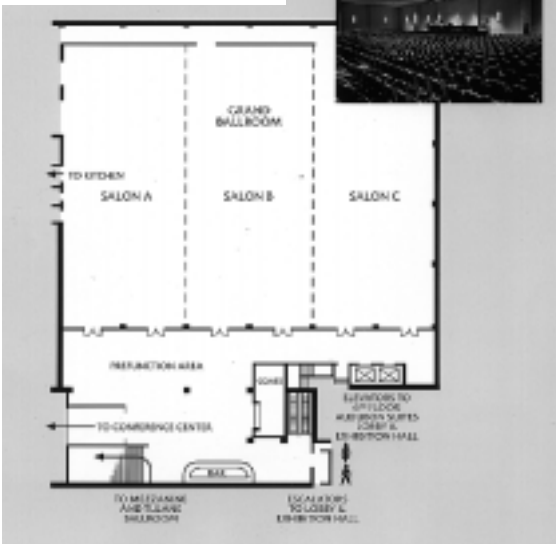
Conference Sponsors



The Radisson Hotel New Orleans Conference Center, 2nd Floor



Grand Ballroom
2nd Floor



Tulane Room
Mezzanine

