

# AIRS 2010

NTU, Taipei, Taiwan December 1-3

**BROCHURE**

# Organization

## **Honorary Conference Chair:**

Si-Chen Lee, President, National Taiwan University

## **Conference Chair:**

Pu-Jen Cheng, National Taiwan University

## **Program Committee Co-Chairs:**

Lam Wai, Chinese University of Hong Kong

Min-Yen Kan, National University of Singapore

## **Publication Co-Chair:**

Preslav Nakov, National University of Singapore

## **Finance Chair:**

Lun-Wei Ku, National Taiwan University

## **Webmasters:**

Chien-Wen Chen, National Taiwan University

Yen-Chieh Huang, National Taiwan University

## **Area Chairs:**

### **IR Models and Applications**

Min Zhang, Tsinghua University

Bin Wang, Chinese Academy of Sciences

### **User Study, IR Evaluation, and Interactive IR**

Tetsuya Sakai, Microsoft Research Asia

William Webber, University of Melbourne

### **Web IR, Scalability, and Adversarial IR**

Joemon Jose, University of Glasgow

### **Multimedia IR**

Winston Hsu, National Taiwan University

Kazunari Sugiyama, National University of Singapore

### **Natural Language Processing for IR**

Timothy Baldwin, University of Melbourne

Atsushi Fujii, Tokyo Institute of Technology

### **Machine Learning and Data Mining for IR**

Tie-Yan Liu, Microsoft Research Asia

Yi Zhang, University of California Santa Cruz

## **Welcome from the Conference & Program Chairs**

Welcome to the sixth Asia Information Retrieval Societies Conference (AIRS) 2010. Welcome to Taipei. The AIRS conferences aim to bring together international researchers and developers to exchange new ideas and the latest results in information retrieval. AIRS 2010 continues the conference series that grew from the Information Retrieval with Asian Languages (IRAL) workshop series, started in 1996. It has become a mature venue for information retrieval work, finding support from the ACM Special Interest Group on Information Retrieval (SIGIR), the Association for Computational Linguistics and Chinese Language Processing (ACLCLP) and the Information Processing Society of Japan, Special Interest Group on Information Fundamentals and Access Technologies (IPSI SIG-IFAT).

This year saw a sharp rise in the number of submissions over the previous year. A total of 120 papers were submitted, representing work by academics and practitioners not only from Asia, but worldwide such as Australia, Europe and North America. The high quality of the work made it difficult for the dedicated program committee to decide which papers to feature at the conference. Through a double-blind reviewing process, 26 submissions (21%) were accepted as full oral papers and 31 (25%) were accepted as short posters. The success of this conference is only possible with the support of all of the authors who submitted papers for review. We thank them for their support for this conference.

For a conference to run smoothly, much behind-the-scenes work is necessary, most of which is largely transparent to the authors and delegates. We would like to thank Preslav Nakov, our publication chair, who painstakingly worked with each individual paper's authors to ensure formatting, spelling, diction and grammaticality were

completely error-free. We thank Lun-Wei Ku, our finance chair, who spent many hours to reduce the conference registrations rates. We also thank the 11 area chairs: Min Zhang, Bin Wang, Tetsuya Sakai, William Webber, Joemon Jose, Winston Hsu, Kazunari Sugiyama, Timothy Baldwin, Atsushi Fujii, Tie-Yan Liu, and Yi Zhang, the 135 program committee members and the 19 Secondary Reviewers, who constructively assessed the submissions and helped select quality contributions. We thank Chien-Wen Chen and Yen-Chieh Huang for website maintenance. Without all of their work, the conference would not have been possible. We would also like to thank Si-Chen Lee, President of the National Taiwan University, for accepting to be the honorary chair for the conference.

Finally, we are grateful to the keynote speakers Sung-Hyon Myaeng from KAIST and ChengXiang Zhai from UIUC for sharing their insight on exciting topics. And we never forget our sponsors for supporting the AIRS 2010 conference: Dept. of Computer Science and Information Engineering, National Taiwan University, National Science Council of Taiwan, Ministry of Education of Taiwan, and Institute of Information Science, Academia Sinica.

We sincerely hope you will enjoy the program and the conference.

**Pu-Jen Cheng**

Conference Chair  
National Taiwan University

**Lam Wai**

Program Committee Co-Chair  
Chinese University of Hong Kong

**Min-Yen Kan**

Program Committee Co-Chair  
National University of Singapore

# Organizer, Sponsors & Supporters

## Organizer & Sponsor

---



## Sponsors

---



## Supporters

---



ACLCLP



IPSJ SIG-IFAT





# General Information

## Climate:

Average temperature: 15°C(60°F)~24°C(75°F)

## Currency:

New Taiwan Dollar(TWD)

Average rate 1 USD = 30 TWD

## Electricity:

110V 60Hz



## Languages:

Official language: Chinese

Other languages often spoken: Taiwanese, English, Japanese

## Population:

Taipei city: 2,622,125 inhabitants

## Tipping:

Prices in hotels, restaurants and taxi fares include taxes and service.

For more information about Taipei, please visit

<http://www.taipei.gov.tw/> & <http://www.tpedit.taipei.gov.tw/>



# Social Program

## Welcome Reception:

Wednesday, December 1<sup>st</sup>, 16:30 ~ 18:30

Conference Hall

## Conference Banquet (& Best Paper Award Ceremony):

Thursday, December 2<sup>nd</sup>, 18:00 ~ 21:00

Leader Hotel – Taipei

(<http://www.leaderhotel.com/taipei/en/content/intro.htm>)

The conference dinner will be held on the 2nd floor of Leader Hotel.

Address: No. 83, Sec. 4, Roosevelt Rd., Taipei

Tel: +886-2-8369-2858

## Getting there

- From Conference Hall: About 20 minutes walk. Please reference to the MAP shown in the next page. Our staff will leave for Leader Hotel from the conference hall at 5:20 pm. You are welcome to go with us.
- By Metro: Get off the **GongGuan Metro Station** at Exit 3 and walk south along Roosevelt Rd. about 3mins, and then you can see Leader Hotel on your left side.

## Contact Phone Number during the Conference

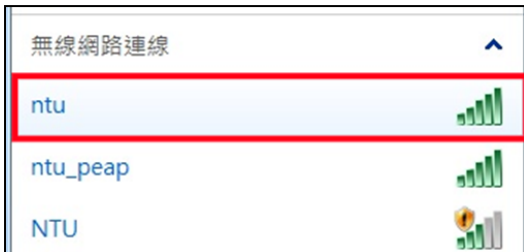
+886-987-885-847 or 0987-885-847

# From Conference Hall to Leader Hotel - Taipei



# Connecting to the Wi-Fi

1. Connect to the wireless network “ntu”



2. The key is “ntuwe”



3. Open your browser and enter the Wi-Fi account and password, which will be put in your badge holder.



Please contact our staff at the registration desk if you have any problem.

# Keynote Speeches

Wednesday, Dec. 1st, 2010, 9:00 am ~ 10:00 am

## Enabling Proactive Search

*Prof. Sung-Hyon Myaeng, KAIST, Korea*



More than ever, people use search technologies ubiquitously with smart devices including smart phones and TVs. People look for not only information but also objects like people, shops, and mobile applications, and even specific events and services. With more diversified search targets, it is not sufficient anymore to rely on users' search queries and react to them. A search system needs to behave proactively in anticipation of what's required for the user's task at hand. This talk will start from this premise and show our ongoing efforts to enable proactive search and action recommendations that are becoming increasingly important with smart devices.

Prof. Sung-Hyon Myaeng is currently a professor in Department of Computer Science and the head of Web Science & Technology Division at Korea Advanced Institute of Science and Technology (KAIST). He is also the Director of KAIST-Microsoft Research Collaboration Center (KMCC). Previously he was on the faculty at Syracuse University, USA, where he was granted tenure. He earned his MS and Ph. D. from Southern Methodist University, Texas, USA in 1985 and 1987, respectively. His current research interests are: information retrieval (esp. contextual ad searching, mobile searching, & text analysis for unconventional search criteria), text mining (opinion mining, trend analysis, relation extraction, & categorization), and context-aware computing including extraction and use of commonsense knowledge, especially in human activities. He has been on editorial boards of

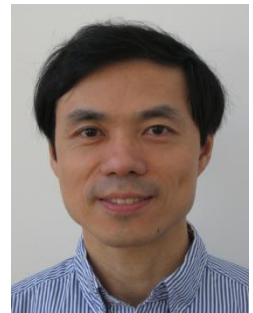
international journals including ACM Transactions on Asian Information Processing (TALIP) and Journal of Computer Processing of Oriental Languages (JCOPL), both as an associate editor and Information Processing and Management. He also has served on program committees of many reputable international conferences in the areas of information retrieval, natural language processing, and digital libraries, including his role as a co-program chair for ACM SIGIR, 2002 and 2008. In 2008, he won an award from Microsoft Research, based on global competition for the RFP 'Beyond Search – Semantic Computing and Internet Economics'. He received Digital Innovation Award from Hankookilbo (Korea Daily) in 2002 for the development of a new technique called Virtual Document Digital Library System.

---

Thursday, Dec. 2nd, 2010, 8:35 am ~ 9:35 am

## **Maximum Personalization: User-Centered Adaptive Information Retrieval**

*Prof. ChengXiang Zhai, UIUC, USA*



Precise understanding of a user's information need is critical for a search engine to deliver optimal search results to users. Unfortunately, the current search engines do not really know individual users well, limiting their ability to optimize the search results for each individual user or help users in case of long-tail difficult queries.

In this talk, I will present a new retrieval strategy called user-centered adaptive information retrieval (UCAIR) which would break this limitation and achieve maximum personalization. With UCAIR, each user would "own" a personalized intelligent search agent that would know the user very well through observing the user's information seeking behavior from the client-side, and would exploit its "complete"

knowledge about the user to interactively optimize the search results for a particular user.

As a theoretical foundation forUCAIR, I will present a general decision-theoretic framework for optimal interactive information retrieval, in which the system iteratively responds to each user action by choosing an optimal system action based on all the available user information and search context. I will present several algorithms for personalizing search results, and use them to show that such a general decision-theoretic view of retrieval can naturally suggest many interesting ways to achieve personalization in an interactive search system, and thatUCAIR can significantly improve the utility of the current document-centered search engines by maximally optimizing search results for each individual user.

ChengXiang Zhai is an Associate Professor of Computer Science at the University of Illinois at Urbana-Champaign, where he also holds a joint appointment at the Institute for Genomic Biology, Statistics, and the Graduate School of Library and Information Science. He received a Ph.D. in Computer Science from Nanjing University in 1990, and a Ph.D. in Language and Information Technologies from Carnegie Mellon University in 2002. He worked at Clairvoyance Corp. as a Research Scientist and a Senior Research Scientist from 1997 to 2000. His research interests include information retrieval, text mining, natural language processing, machine learning, and bioinformatics. He is an Associate Editor of ACM Transactions on Information Systems, and Information Processing and Management, and serves on the editorial board of Information Retrieval Journal. He is a program co-chair of ACM CIKM 2004, NAACL HLT 2007, and ACM SIGIR 2009. He is an ACM Distinguished Scientist, and received the 2004 Presidential Early Career Award for Scientists and Engineers (PECASE), the ACM SIGIR 2004 Best Paper Award, an Alfred P. Sloan Research Fellowship in 2008, and an IBM Faculty Award in 2009.

---

# Technical Program

December 1st, 2010	
8:00-8:45	<b>Registration</b>
8:45-9:00	<b>Opening</b>
9:00-10:00	<b>Keynote Speech</b> <u>Enabling Proactive Search</u> <i>Prof. Sung-Hyon Myaeng</i> <i>Korea Advanced Institute of Science and Technology</i> <i>Chair: Prof. Hsin-Hsi Chen, National Taiwan University</i>
10:00-10:25	Coffee break
<b>Session 1: IR Models – I</b> <i>Session Chair: Dr. Benjamin Piwowarski, University of Glasgow</i>	
10:25-10:50	Relevance Ranking using Kernels
10:50-11:15	Mining YouTube to Discover Hate Videos, Users and Hidden Communities
11:15-11:40	Title-based Product Search - Exemplified in a Chinese E-commerce Portal
11:40-12:05	Relevance Model Revisited: With Multiple Document Representations
12:05-13:10	<b>Lunch</b>
<b>Session 2: Machine Learning for IR – I</b> <i>Session Chair: Prof. Qi Zhang, Fudan University</i>	
13:15-13:40	Multi-Viewpoint Based Similarity Measure and Optimality Criteria for Document Clustering
13:40-14:05	A Text Classifier with Domain Adaptation for Sentiment Classification
14:05-14:30	Coffee break

<b>Session 3: User Studies and Evaluation</b>	
<i>Session Chair: Prof. Masaharu Yoshioka, Hokkaido University</i>	
14:30-14:55	Effective Time Ratio: A Measure for Web Search Engines with Document Snippets
14:55-15:20	Investigating Characteristics of Non-click Behavior Using Query Logs
15:20-15:45	Score Estimation, Incomplete Judgments, and Significance Testing in IR Evaluation
16:30-18:30	<b>Reception and Poster Session (Conference Hall)</b>

<b>December 2nd, 2010</b>	
8:00-8:35	<b>Registration</b>
8:35-9:35	<b>Keynote Speech</b> <u>Maximum Personalization: User-Centered Adaptive Information Retrieval</u> <i>Prof. ChengXiang Zhai</i> <i>University of Illinois at Urbana-Champaign</i> <i>Chair: Prof. Lam Wai, Chinese University of Hong Kong</i>
9:35-10:00	Coffee break
<b>Session 4: NLP for IR</b>	
<i>Session Chair: Prof. Yoshihiko Hayashi, Osaka University</i>	
10:00-10:25	A Local Generative Model for Chinese Word Segmentation
10:25-10:50	Re-ranking Summaries Based on Cross-document Information Extraction
10:50-11:15	A Two-Stage Algorithm for Domain Adaptation with Application to Sentiment Transfer Problems
11:15-11:40	Domain-Specific Term Rankings Using Topic Models
11:40-12:05	Learning Chinese Polarity Lexicons by Integration of Graph Models and Morphological Features
12:05- 13:10	<b>Lunch</b>



<b>Session 5: Machine Learning for IR – II</b>	
<i>Session Chair: Dr. Jun Xu, Microsoft Research Asia</i>	
13:30-13:55	Learning to Rank with Supplementary Data
13:55-14:20	Event Recognition from News Webpages through Latent Ingredients Extraction
14:20-14:45	Tuning Machine-Learning Algorithms for Battery-Operated Portable Devices
14:45-15:15	Coffee break
<b>Session 6: IR Models – II</b>	
<i>Session Chair: Prof. Min Zhang, Tsinghua University</i>	
15:15-15:40	A Unified Iterative Optimization Algorithm for Query Model and Ranking Refinement
15:40-16:05	A Study of Document Weight Smoothness in Pseudo Relevance Feedback
16:05-16:30	Modeling Variable Dependencies between Characters in Chinese Information Retrieval
18:00-21:00	<b>Banquet and Best Paper Award Ceremony (Leader Hotel – Taipei)</b>

<b>December 3rd, 2010</b>	
8:30-9:00	<b>Registration (for both of conference and workshop)</b>
<b>Session 7: Web and QA</b>	
<i>Session Chair: Prof. Tyne Liang, National Chiao-Tung University</i>	
9:00-9:25	Mining Parallel Documents Across Web Sites
9:25-9:50	A Revised SimRank Approach for Query Expansion
9:50-10:15	On a Combination of Probabilistic and Boolean IR Models for Question Answering
10:15-10:40	Improving Web-Based OOV Translation Mining for Query Translation
10:40-11:00	Coffee break

## Session 8: Multimedia IR

*Session Chair: Prof. Winston Hsu, National Taiwan University*

11:00-11:25	Emotion Tag Based Music Retrieval Algorithm
11:25-11:50	An Aesthetic-based Approach to Rank Web Images
11:50-12:00	<b>Closing Session</b>
12:00-12:50	<b>Lunch</b>

## Local IR workshop (December 3rd, 2010)

12:30-12:50	<b>Registration</b>
12:50-13:35	Query Reformulation for Information Retrieval <i>Prof. Pu-Jen Cheng, National Taiwan University</i>
13:35-14:20	A Kernel Approach to Matching of Query and Document in Search <i>Dr. Jun Xu, Microsoft Research Asia</i>
14:20-15:05	Recall Estimation with Assessor Error in E-Discovery <i>Mr. William Webber, University of Melbourne</i>
15:05-15:25	<b>Coffee Break</b>
15:25-16:10	Web-Scale Knowledge Discovery and Population from Unstructured Data <i>Prof. Heng A. Ji, The City University of New York</i>
16:10-16:55	User behavior analysis for Web search: with and beyond the Wisdom of Crowds <i>Prof. Min Zhang, Tsinghua University</i>

## Poster Session

**Poster session presenters can set up their presentation at 12:40 at the conference hall on Dec. 1st, 2010. We provide appropriate facilities for the setting up. For any questions, please contact our staff at the registration desk.**

- P01: Multi-Search: A Meta-Search Engine Based on Multiple Ontologies
- P02: Co-HITS-Ranking Based Query-Focused Multi-Document Summarization
- P03: Advanced Training Set Construction for Retrieval in Historic Documents
- P04: Ontology-Driven Semantic Digital Library
- P05: Revisiting Rocchio's Relevance Feedback Algorithm for Probabilistic Models
- P06: When Two is Better than One: A Study of Ranking Paradigms and Their Integrations for Subtopic Retrieval
- P07: Connecting Qualitative and Quantitative Analysis of Web Search Process: Analysis Using Search Units
- P08: Transliteration Retrieval Model for Cross Lingual Information Retrieval
- P09: The Role of Lexical Ontology in Expanding the Semantic Textual Content of On-Line News Images
- P10: Order Preserved Cost-sensitive Listwise Approach in Learning to Rank
- P11: Pseudo-Relevance Feedback Based on mRMR Criteria
- P12: An Integrated Deterministic and Nondeterministic Inference Algorithm for Sequential Labeling
- P13: FolkDiffusion: A Graph-based Tag Suggestion Method for Folksonomies
- P14: Effectively Leveraging Entropy and Relevance for Summarization
- P15: Machine Learning Approaches for Modeling Spammer Behavior
- P16: Research of Sentiment Block Identification for Customer Reviews Based on Conditional Random Fields
- P17: Semantic Relation Extraction Based on Semi-supervised Learning
- P18: Corpus-based Arabic Stemming using N-grams
- P19: Analysis and Algorithms for Stemming Inversion
- P20: Top-down and Bottom-up: A Combined Approach to Slot Filling
- P21: Relation Extraction between Related Concepts by Combining Wikipedia and Web Information for Japanese Language
- P22: A Chinese Sentence Compression Method for Opinion Mining

- P23: Relation Extraction in Vietnamese Text using Conditional Random Fields
- P24: A Sparse L2-Regularized Support Vector Machines for Large-scale Natural Language Learning
- P25: An Empirical Comparative Study of Manual Rule-Based and Statistical Question Classifiers on Heterogeneous Unseen Data
- P26: Constructing Blog Entry Classifiers using Blog-level Topic Labels
- P27: Finding Hard Questions by Knowledge Gap Analysis in Question Answer Communities
- P28: Exploring the Visual Annotatability of Query Concepts for Interactive Cross-Language Information Retrieval
- P29: A Diary Study-Based Evaluation Framework for Mobile Information Retrieval
- P30: Dynamics of Genre and Domain Intents
- P31: Query Recommendation Considering Search Performance of Related Queries