

Conference Program

September 4, 2008 (Thursday) 9 : 20~20 : 00		
Time	Session	Chair
08:45~09:20	Registration	
09:20~09:30	Opening Ceremony	Prof. Yuen-Hsien Tseng
09:30~10:30	Invited Talk.: Dr. Chin-Yew Lin (林欽佑) , MSRA	
10:30~10:50	Coffee Break	
10:50~12:10	Oral Session 1: Lexical and Semantic Information	Prof. Zhao-Ming Gao (NTU)
12:10~13:00	Lunch	
13:00~14:00	ACLCLP Meeting/Poster Session 1: NSC Projects	
14:10~15:50	Oral Session 2: Speech Analysis and Synthesis	Prof. Yuan-Fu Liao (NTUT)
15:50~16:20	Coffee Break	
16:20~17:40	Oral Session 3: Preliminary Language Processing and Applications	Prof. Wen-Hsiang Lu (NCKU)
18:00~20:00	Banquet	
September 5, 2008 (Friday) 9 : 00~16 : 40		
Time	Session	Chair
09:00~10:10	Invited Talk.: Prof. Wai Lam (林偉) , CUHK	
10:10~10:40	Coffee Break	
10:40~12:00	Oral Session 4: Statistical Techniques for Language Processing	Prof. Jing-Shin Chang (NCNU)
12:10~13:00	Lunch	
13:00~14:20	Poster Session 2 : ROCLING Papers	
14:20~14:40	Coffee Break	
14:40~16:20	Oral Session 5: Speech Detection and Recognition	Dr. Chun-Jen Lee (Telecommunication Laboratories)
16:20~16:40	Closing Ceremony and Best Paper Award	

The 20th ROCLING Conference has invited distinguished researchers, Prof. Wai Lam and Dr. Chin-Yew Lin, as our invited keynote speakers.

Time	09:30~10:30, Sep. 4, 2008 (Thursday)

Speaker	Dr. Chin-Yew Lin (林欽佑)
Title	Web-Scale NLP - A Challenge and An Opportunity
Abstract	<p>Web 2.0 presents us a challenge and also an opportunity. On one hand, it brings users to the forefront of the internet. Users are not just consumers of the web anymore. They are also producers. How do we deal with the ever increasing amount of user generated content becomes a real challenge. On the other hand, users' direct expressions as displayed on the internet also present us a rare opportunity to tap into what users really want and think. It also gives us a cost-effective way to harness users' collective intelligence.</p> <p>In this talk, I will use question answering as an example to illustrate how we at Microsoft Research Asia take this web-scale NLP challenge and turn it into an opportunity. Question answering has been a very active research field in information retrieval and natural language processing. Despite the success of TREC QA track, large scale robust QA systems are still yet to be found in the real world. In this talk, I will briefly introduce recent progress in SQuAD - a question and answering project aiming to crawl, index, and serve all question and answer pairs existing on the web. I will address six main challenges of the project and then focus on the topic of question search and recommendation.</p> <p>If time permits, I will also show a few demos about SQuAD and a couple of other demos developed at MSRA to illustrate our efforts in embracing the challenge of web-scale NLP.</p>
Vita	<p>Dr. Chin-Yew Lin (林欽佑) is a lead researcher and research manager at Microsoft Research Asia (MSRA). Before joining Microsoft in 2006, He was a senior research scientist at the Information Sciences Institute at University of Southern California (USC/ISI) where he worked in the Natural Language Processing and Machine Translation group since 1997. His research interests are automated summarization, opinion analysis, question answering, computational advertising, community intelligence, machine translation, and machine learning.</p> <p>Recently, his main focus is developing scalable automatic question answering and distillation system - SQuAD. He also developed automatic evaluation technologies for summarization, QA, and MT. In particular, He created the ROUGE automatic summarization evaluation package. It has become the de facto standard in summarization evaluations. More than 200 research sites worldwide have downloaded this package.</p> <p>Home page: http://research.microsoft.com/~cyl/</p>

Time	09:00~10:00, Sep. 5, 2008 (Friday)
Speaker	

	Prof. Wai Lam (林偉)
Title	Web Information Extraction Learning Based on Probabilistic Graphical Models
Abstract	<p>Building advanced Web mining applications typically requires precise text information extraction from a large number of different Web sites. Due to the diverse layout format and content, substantial human effort is needed for the text extraction task. We investigate learning frameworks automating and adapting the extraction task based on probabilistic graphical models which provide a principled paradigm harnessing the uncertainty during the learning process.</p> <p>One of our projects is to develop an unsupervised learning framework for simultaneously extracting and normalizing attributes of products from multiple Web pages originated from different sites. This framework can model the page-independent content information and the page-dependent layout information of the text fragments in Web pages. Previously unseen attributes can be discovered from the clue contained in the layout format of the text fragments.</p> <p>This framework can tackle both extraction and normalization tasks by jointly considering the relationship between the content and layout information. Dirichlet process prior is employed leading to another advantage that the number of discovered product attributes is unlimited.</p>
Vita	<p>Prof. Wai Lam (林偉) received a Ph.D. in Computer Science from the University of Waterloo. He obtained his BSc. and M.Phil. degrees from the Chinese University of Hong Kong. After completing his Ph.D. degree, he conducted research at Indiana University Purdue University Indianapolis (IUPUI) and University of Iowa. He joined the Chinese University of Hong Kong, where he is currently a professor.</p> <p>His research interests include intelligent information retrieval, text mining, digital library, machine learning, and knowledge-based systems. He has published articles in IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Knowledge and Data Engineering, ACM Transactions on Information Systems, etc.</p> <p>His research projects have been funded by the Hong Kong Government RGC Earmarked Grant and DARPA (USA). He also managed industrial projects funded by Innovation and Technology Fund (industrial grant) and IT companies.</p> <p>Home page: http://www.se.cuhk.edu.hk/people/wlam.html</p>

Oral Session 1: Lexical and Semantic Information

Session Chair: *Prof. Zhao-Ming Gao (NTU)*

1. [Measuring Text Readability by Lexical Relations Retrieved from WordNet](#)
Shuyen Lin, Chengchao Su, Yuda Lai, Lichin Yang, and Shukai Hsieh
 2. [A Semantic Composition Method for Deriving Sense Representations of Determinative-Measure Compounds in E-HowNet](#)
Chia-hung Tai, Shu-Ling Huang, and Keh-Jiann Chen
 3. [A Thesaurus-Based Semantic Classification of English Collocations](#)
Chung-chi Huang, Chiung-hui Tseng, Kate H. Kao, and Jason S. Chang
-

Oral Session 2: Speech Analysis and Synthesis

Session Chair: *Prof. Yuan-Fu Liao (NTUT)*

1. [以Fujisaki模型驗證連續語流中字調及韻律詞對應於階層性韻律架構HPG的意義](#)
Chiu yu Tseng and Zhao yu Su
 2. [基於ANN之頻譜演進模型及其於國語語音合成之應用](#)
古鴻炎、吳昌益
 3. [A New Novel Multi-Speaker Adaptation Architecture for Ubiquitous Speech Recognition System](#)
Po-Yi Shih, Yuan-Ning Lin, and Jhing-Fa Wang
 4. [Study of Modulation Spectrum Normalization Techniques for Robust Speech Recognition](#)
Chih-Cheng Wang, Wen-hsiang Tu, and Jeh-wei Hung
-

Oral Session 3: Preliminary Language Processing and Applications

Session Chair: *Prof. Wen-Hsiang Lu (NCKU)*

1. [形音相近的易混淆漢字的搜尋與應用](#)
劉昭麟、黃志斌、翁睿妤與莊怡軒
 2. [A Realistic and Robust Model for Chinese Word Segmentation](#)
Chu-Ren Huang, Ting-Shuo Yo, Petr Simon, and Shu-Kai Hsieh
 3. [國台語無聲調拼音輸入法實作](#)
Ming-Shing Yu and Cheng-Rong Tsai
-

Oral Session 4: Statistical Techniques for Language Processing

Session Chair: *Prof. Jing-Shin Chang (NCNU)*

1. [Propositional Term Extraction using Word Cohesiveness and Conditional Random Fields with Multi-Level Features](#)
Ru-Yng Chang and Chung-Hsien Wu
2. [利用統計方法及中文訓練資料處理台語文詞性標記](#)
楊允言、戴嘉宏、劉杰岳、陳克健與高成炎

3. Chinese NP Chunking: Experiments with Supervised, and Semi-supervised Learning

Zhao Ming Gao and Yen Hsi Lin

Oral Session 5: Speech Detection and Recognition

Session Chair: *Dr. Chun-Jen Lee (Telecommunication Laboratories)*

1. Study of the Improved Normalization Techniques of Energy-Related Features for Robust Speech Recognition

Chi-an Pan, Wen-hsiang Tu, and Jieh-weih Hung

2. Robust Features for Effective Speech and Music Discrimination

Fu Zhonghua and Wang Jhing-Fa

3. Robust Voice Activity Detection Based on Discrete Wavelet Transform

Kun-Ching Wang

4. Associative Cepstral Statistics Normalization Techniques for Robust Speech Recognition

Kuang-chieh Wu, Wen-hsiang Tu, and Jieh-weih Hung

Poster Session 2: ROCLING Papers

1. The Polysemy of PO in Mandarin Chinese

Harvey Hsin-chang Ho

2. One-Sample Speech Recognition of Mandarin Monosyllables using Unsupervised Learning

Tze Fen Li and Shui-Ching Chang

3. Examining the Lexical Effect on Categorical Perception of Stops in Taiwan Southern Min

Yunglin Tai

4. Automatic labeling of troponymy for Chinese verbs

Chiao-Shan Lo, Yi-Rung Chen, Chih-Yu Lin, and Shu-Kai Hsieh

5. 電腦輔助中學程度漢英翻譯習作環境之建置

賴敏華 與 劉昭麟

6. 以範例為基礎之英漢TIMSS試題輔助翻譯

張智傑 與 劉昭麟

7. 電腦輔助推薦學術會議論文評審委員之初探

陳禹勳 與 劉昭麟

8. 多領域文件集之詞彙概念擴展與知識架構之建立

Yong-Xiang Chen, Xiu-Ling Ke, Keh-Jiann Chen, and Chu-Ren Huang

9. 結合線上翻譯服務的跨語言專利檢索系統

Shun-Yuan Teng and Guo-Wei Bian

10. [Generating Patterns for Extracting Chinese-Korean Named Entity Translations from the Web](#)

Chih-Hao Yeh, Wei-Chi Tsai, Yu-Chun Wang, and Richard Tzong-Han Tsai