

# CURRICULUM VITAE

## CHIN-YEW LIN

Web Intelligence Group  
Microsoft Research Asia  
5F, Beijing Sigma Center, No. 49, Zhichun Road, Haidian District,  
Beijing 100190, P.R. China  
Phone: +86 (10) 5896-3481, Fax: +86 (10) 8809-7306, Email: cyl@microsoft.com  
August 2009

### EDUCATION

University of Southern California, August 1997: Ph.D. (Computer Engineering)  
University of Southern California, December 1991: M.S. (Computer Engineering)  
National Chiao Tung University, June 1987: B.Sc. (Electrical and Control Engineering)

### RESEARCH INTERESTS

My research interests are text summarization, question answering, opinion extraction, information extraction, machine translation, and machine learning. Recently, my main focus is developing social/scalable automatic question answering and distillation system – SQuAD. This project aims to mine existing question and answer pairs on the web, create a searchable question and answer repository, generate unseen questions and acquire their answers automatically.

Previously, I have developed technologies for summarization and machine translation. Basic Element (BE), a joint effort with E.H. Hovy, L. Zhou, J. Fukumoto from Ritsumeikan University, Japan, K. McKeown and A. Nenkova from Columbia University, was made freely available to the research community in April 2005 and it was used in Document Understanding Conference 2005 and 2006, a worldwide large-scale summarization evaluation effort sponsored by the US government. Before BE, I have created the ROUGE automatic summarization evaluation package. It was delivered to the research community in 2004. More than 150 research sites worldwide have downloaded this package so far. ROUGE has been chosen as the official automatic evaluation package for Document Understanding Conference 2004, 2005, and 2006.

My past projects include: *NeATS*, an automated multi-document text summarization system; *TextMap*, an intelligent question and answering assistant; *Webclopedia*, automated question answering systems; *SEE*, an environment written in Java in which assessors can evaluate the quality of one text (the peer text) as compared to an ideal (the model text). SEE has been used by the Document Understanding Conference to support its human evaluation protocol.

I am the architect of NeATS. NeATS is a multi-document summarization system that attempts to extract relevant or interesting portions from a set of documents about some topic and present them in coherent order. NeATS is one of the 2 top performers in the large scale summarization evaluation, Document Understanding Conference 2001 and 2002, sponsored by the US government.

I design the overall system architecture, the information retrieval front end, and the answer reranking module of Webclopedia. Webclopedia is one of the top 3 performers in the question answering track of the Text REtrieval Conference (TREC) in 2000 and 2001.

I am the principal designer of a multilingual information retrieval, summarization, and translation system, MuST. MuST integrates state-of-the-art off-the-shelf information retrieval engines, multilingual automated text summarization technologies from SUMMARIST project, and rapid-prototype shallow machine translation engine QuTE. MuST enables users to search Internet or local free text database of multiple languages (currently, English, Arabic, Bahasa Indonesia, Japanese, and Spanish), summarize documents to different lengths, and translate the full texts or summaries into English. I have demonstrated a prototype system to several government and private agencies and received very good reviews. Several sites have requested it for field testing and technology transfer.

I have been developing an automated text summarization system prototype, SUMMARIST (an overview paper describing the current status of SUMMARIST was presented at the ACL97/ECAL97 Intelligent Scalable Text Summarization Workshop), under the supervision of Dr. Hovy. The goal of SUMMARIST is to provide both extracts and abstracts for arbitrary (English, Arabic, Bahasa Indonesia, Japanese, Spanish, and later other languages) input text. SUMMARIST combines symbolic world knowledge (embodied in WordNet, dictionaries, and similar resources) with robust natural language processing (using Information Retrieval and statistical techniques) to overcome the problems prevalent in either approach alone. SUMMARIST uses several key modules developed during my Ph.D. work at the University of Southern California.

My Ph.D. work focused on the development of a robust automated topic identification system. Topic identification is the task of identifying what is being talked about in a text. Although topic identification is a crucial step for many text processing tasks such as text summarization, text filtering, text routing, etc., it has not been studied as a stand-alone task. I investigated several topic identification methods and developed three major results: the Position Method, the Topic Signature Method, and the Concept Generalization Method. These methods used combinations of robust linguistic tools, an available knowledge base, and statistical techniques.

Prior to this, I was a research assistant for the SNAP project in the Parallel Knowledge Processing Laboratory under the supervision of Dr. Moldovan, then of USC. I was responsible for testing a 144-processor parallel processing machine prototype (SNAP), porting natural language processing software to the machine, and participating the fourth and fifth Message Understanding Conferences (MUC-4 and MUC-5).

## Publications

- Y. Cao, Chin-Yew Lin, Y.-I. Song, K. Sun, X. Song and X. Wang. Learning to Recommend Questions Based on User Rating, in *Proceedings of the 18<sup>th</sup> ACM Conference on Information and Knowledge Management (CIKM 2009)*, Hong Kong, China, November, 2-6, 2009.
- W.-Y. Yang, Y. Cao and Chin-Yew Lin. A Structural Support Vector Method for Extracting Contexts and Answers of Questions from Online Forums, in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, Singapore, August 6-7, 2009.
- M. Jeong, Chin-Yew Lin and G. G. Lee. Semi-Supervised Speech Act Recognition in Emails and Forums, in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP-2009 Poster)*, Singapore, August 6-7, 2009.
- M. Jeong, Chin-Yew Lin and G. G. Lee. Efficient Inference of CRFs for Large-Scale Natural Language Data, in *Proceedings of the Joint Conference of the 47<sup>th</sup> Annual Meeting of the Association for Computational Linguistics and the 4<sup>th</sup> International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2009 Short Paper)*, Singapore, August 2-7, 2009.
- Y. Cao, Chin-Yew Lin, Y. Yu, and H.-W. Hon. Recommending Questions Using the MDL-based Tree Cut Model, in *Proceedings of the 17th International World Wide Web Conference (WWW2008)*, Beijing, China, April 21-25, 2008.
- H. Duan, Y. Cao, Chin-Yew Lin, Y. Yu. Searching Questions by Identifying Question Topic and Question Focus. In *Proceedings of ACL-08: HLT*, Columbus, Ohio, USA, June 15-20, 2008.
- S. Ding, G. Cong, Chin-Yew Lin, and X. Zhu. Using Conditional Random Fields to Extract Contexts and Answers of Questions from Online Forums. In *Proceedings of ACL-08: HLT*, Columbus, Ohio, USA, June 15-20, 2008.
- G. Cong, L. Wang, Chin-Yew Lin, and Y.-I. Song. Finding Question-Answer Pairs from Online Forums. In *Proceedings of SIGIR 2008*, Singapore, July 20-24, 2008.
- Y.-I. Song, Chin-Yew Lin, Y. Cao, and H.-C. Rim. Question Utility: A Novel Static Ranking of Question Search. In *Proceedings of AAAI 2008 (AI and the Web Special Track)*, Chicago, Illinois, USA, July 13-17, 2008.
- Y. Liu, S. Li, Y. Cao, Chin-Yew Lin, D. Han, and Y. Yu. Understanding and Summarizing Answers in Community-Based Question Answering Services. In *Proceedings of COLING 2008*, Manchester, UK, August 18-22, 2008.
- X. Song, S. Ding, and Chin-Yew Lin. Better Binarization for the CKY Parsing. In *Proceedings of EMNLP 2008*, Hawaii, USA, Oct 25-27, 2008.
- Chin-Yew Lin. Automatic Question Generation from Queries. In *Proceedings of Workshop on the Question Generation Shared Task and Evaluation Challenge*, Arlington, USA, September 25-26, 2008.
- Y. Xue, Y. Hu, G. Xin, R. Song, S. Shi, Y. Cao, Chin-Yew Lin and H. Li. Web page title extraction and its application, *Information Processing & Management*, Volume 43, Issue 5, Patent Processing, September 2007, Pages 1332-1347.
- G. Sun, G. Cong, X. Liu, Chin-Yew Lin, M. Zhou. Detecting Erroneous Sentences using labeled sequential Patterns and Tree Patterns. In *Proceedings of AAAI 2007*, Vancouver, Canada, July 22 - 26, 2007.

- G. Sun, X. Liu, G. Cong, M. Zhou, Z. Xiong, J. Lee and Chin-Yew Lin. Detecting Erroneous Sentences Using Automatically Mined Sequential Patterns. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007)*, Prague, Czech Republic, Jun 23 - 30, 2007.
- L.-C. Yu, C.-H. Wu, Chin-Yew Lin, E. Hovy and C.-L. Lin. Topic Analysis for Psychiatric Document Retrieval. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007)*, Prague, Czech Republic, Jun 23 - 30, 2007.
- J. Liu, Y. Cao, Chin-Yew Lin, Y. Huang and M. Zhou. Low-Quality Product Review Detection in Opinion Summarization. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2007)*, Poster Session, Prague, Czech Republic, Jun 23 - 30, 2007.
- Y. Ye, M. Zhou, and Chin-Yew Lin. Sentence Level Machine Translation Evaluation as a Ranking Problem: One Step aside from BLEU. In *Proceedings of the 2nd Workshop on Statistical Machine Translation*, Prague, Czech Republic, Jun 23, 2007.
- R. Mulkar, J. R. Hobbs, E. Hovy, H. Chalupsky, and Chin-Yew Lin. Learning by Reading: Two Experiments. In *Proceedings of the 3rd International Workshop on Knowledge and Reasoning for Question Answering (KRQA'07)*, Hyderabad, India, January 6, 2007.
- F. C. Gey, N. Kando, Chin-Yew Lin, and C. Peters. New Directions in Multilingual Information Access, SIGIR 2006 Workshop Report, In *SIGIR Forum*, December 2006, Volume 40 Number 2.
- F. C. Gey, N. Kando, Chin-Yew Lin, and C. Peters. New Directions in Multilingual Information Access: Introduction to the Workshop at SIGIR 2006, Seattle, Washington, USA, August 10, 2006.
- L. Zhou, Chin-Yew Lin, and E.H. Hovy. 2006. Re-evaluating Machine Translation Results with Paraphrase Support. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2006)*, Sydney, Australia, July 22 – 23, 2006.
- Chin-Yew Lin, G. Cao, J. Gao, and J.-Y. Nie. 2006. An Information-Theoretic Approach to Automatic Evaluation of Summaries. In *Proceedings of 2006 Human Language Technology Conference (HLT-NAACL 2006)*, New York, USA, June 5 – 7, 2006.
- L. Zhou, Chin-Yew Lin, D.S. Munteanu, and E.H. Hovy. 2006. ParaEval: Using Paraphrases to Evaluate Summaries Automatically. In *Proceedings of 2006 Human Language Technology Conference (HLT-NAACL 2006)*, New York, USA, June 5 – 7, 2006.
- E.H Hovy, Chin-Yew Lin, L. Zhou, and J. Fukumoto. 2006. Automated Summarization Evaluation with Basic Elements. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, May 24 – 26, 2006.
- L. Zhou, Chin-Yew Lin, and E.H. Hovy. 2006. Summarizing Answer for Complicated Questions. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, May 24 – 26, 2006.

- L. Zhou, Chin-Yew Lin, and E.H. Hovy. 2005. A BE-based Multi-document Summarizer with Query Interpretation. In *Proceedings of Document Understanding Conference (DUC 2005)*, Vancouver, B.C. Canada, Oct. 10-11, 2005.
- E.H. Hovy, Chin-Yew Lin, and L. Zhou. 2005. Evaluating DUC 2005 using Basic Elements. In *Proceedings of Document Understanding Conference (DUC 2005)*, Vancouver, B.C. Canada, Oct. 10-11, 2005.
- E.H. Hovy, Chin-Yew Lin, and L. Zhou. 2005. A BE-based Multi-document Summarizer with Sentence Compression. In *Proceedings of Multilingual Summarization Evaluation (ACL 2005 workshop)*, Ann Arbor, MI.
- Lin, Chin-Yew, L. Zhou, and E.H. Hovy. 2005. Multilingual Summarization Evaluation 2005: Automatic Evaluation Report. In *Proceedings of Multilingual Summarization Evaluation (ACL 2005 workshop)*, Ann Arbor, MI.
- Lin, Chin-Yew and Franz J. Och. 2004. "ORANGE: a Method for Evaluating Automatic Evaluation Metrics for Machine Translation. In *Proceedings of the 20<sup>th</sup> International Conference on Computational Linguistics (COLING 2004)*, Geneva, Switzerland, August 23 – August 27, 2004.
- Lin, Chin-Yew and Franz J. Och. 2004. "Automatic Evaluation of Machine Translation Quality Using Longest Common Subsequence and Skip-Bigram Statistics". In *Proceedings of the 42<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics (ACL 2004)*, Barcelona, Spain, July 24 – 26, 2004.
- Lin, Chin-Yew. 2004. "ROUGE: a Package for Automatic Evaluation of Summaries. In *Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004)*, Barcelona, Spain, July 25 – 26, 2004.
- J. Gao and Chin-Yew Lin. 2004. Introduction to the Special Issue on Statistical Language Modeling. In *ACM Transactions on Asian Language Information Processing (TALIP)*, Volume 3, Issue 2, June 2004.
- Lin, Chin-Yew. Looking for a Few Good Metrics: Automatic Summarization Evaluation – How Many Samples Are Enough? In *Proceedings of the NTCIR Workshop 4*, Tokyo, Japan, June 2 – 4, 2004.
- A. Leuski, Chin-Yew Lin, L. Zhou, U. Germann, Franz J. Och, and E.H. Hovy. Cross-lingual C\*ST\*RD: English Access to Hindi Information. In *ACM Transactions on Asian Language Information Processing (TALIP)*, Volume 2, Issue 3, September 2003.
- Lin, Chin-Yew and E.H. Hovy. 2003. Automatic Evaluation of Summaries Using N-gram Co-occurrence Statistics. In *Proceedings of the 2003 Human Language Technology Conference (HLT-NAACL 2003)*, Edmonton, Canada, May 27 – June 1, 2003.
- Lin, Chin-Yew. 2003. Improving Summarization Performance by Sentence Compression – A Pilot Study. In *Proceedings of the Sixth International Workshop on Information Retrieval with Asian Languages (IRAL 2003)*, Sapporo, Japan, July 7, 2003.
- Lin, Chin-Yew and E.H. Hovy. 2003. The Potential and Limitations of Sentence Extraction for Summarization. In *Proceedings of the Workshop on Automatic Summarization, post-conference workshop of HLT-NAACL-2003*, Edmonton, Canada, May 31 – June 1, 2003.

- Lin, Chin-Yew. 2002. The Effectiveness of Dictionary and Web-Based Answer Reranking. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002)*, Taipei, Taiwan, August 24 - September 1, 2002.
- E.H. Hovy, U. Hermjakob, Chin-Yew Lin, and Deepak Ravichandran. 2002. Using Knowledge to Facilitate Pinpointing of Factoid Answers. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002)*, Taipei, Taiwan, August 24 - September 1, 2002.
- U. Hermjakob, E.H. Hovy, and Chin-Yew Lin. 2002. Knowledge-Based Question Answering. In *Proceedings of the 6<sup>th</sup> World Multiconference on Systems, Cybernetics and Informatics (SCI-2002)*, Orlando, FL, U.S.A., July 14-18, 2002.
- Lin, Chin-Yew and E.H. Hovy. 2002. Manual and Automatic Evaluations of Summaries. In *Proceedings of the Workshop on Automatic Summarization* post-conference workshop of ACL-02, Philadelphia, PA, U.S.A., July 11-13, 2002.
- Lin, Chin-Yew and E.H. Hovy. 2002. From Single to Multi-document Summarization: A Prototype System and its Evaluation. In *Proceedings of the 40<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (ACL-02)*, Philadelphia, PA, U.S.A., July 7-12, 2002.
- Lin, Chin-Yew and E.H. Hovy. 2002. Automated Multi-document Summarization in NeATS. In *Proceedings of the 2002 Human Language Technology Conference (HLT2002)*, San Diego, CA, U.S.A., March 23-27, 2002.
- Hovy, E.H., U. Hermjakob, and Chin-Yew Lin. 2001. The Use of External Knowledge of Factoid QA. In *Proceedings of the 10<sup>th</sup> Text Retrieval Conference (TREC 2001)*, Gaithersburg, MD, U.S.A., November 13-16, 2001.
- Hovy, E.H., L. Gerber, U. Hermjakob, and Chin-Yew Lin and D. Ravichandran. 2001. Toward Semantics-Based Answer Pinpointing. In *Proceedings of the Human Language Technology Conference (HLT2001)*, San Diego, CA, U.S.A., March 18-21, 2001.
- Hovy, E.H., L. Gerber, U. Hermjakob, M. Junk, and Chin-Yew Lin. 2000. Question Answering in Webclopedia. In *Proceedings of the 9<sup>th</sup> Text Retrieval Conference (TREC9)*, Gaithersburg, MD, U.S.A., November 13-16, 2000.
- Lin, Chin-Yew and E.H. Hovy. 2000. The Automated Acquisition of Topic Signatures for Text Summarization. In *Proceedings of the 18<sup>th</sup> International Conference on Computational Linguistics (COLING 2000)*, Saarbrücken, Germany, July 31- August 4, 2000.
- Lin, Chin-Yew. Training a Selection Function for Extraction. In *the 8<sup>th</sup> International Conference on Information and Knowledge Management (CIKM 99)*, Kansa City, Missouri, November 2-6, 1999.
- Lin, Chin-Yew. 1999 Machine Translation for Information Access across Language Barrier: the MuST System. In *Machine Translation Summit VII*, Singapore, September 13-17, 1999.
- Hovy, E.H. and Chin-Yew Lin. 1999. Automating Text Summarization in SUMMARIST. In I. Mani and M. Maybury (eds.), *Advances in Automated Text Summarization*. MIT Press.

- Hovy, E.H. and Chin-Yew Lin. 1998. Automated Text Summarization and the SUMMARIST system. *TIPSTER Text Program Phase III Final Report*, October 1998.
- Lin, Chin-Yew. 1998. Assembly of Topic Extraction Module in SUMMARIST. In *Working Notes of AAAI Spring Symposium on Intelligent Text Summarization*. Stanford, California, March 1998.
- Lin, Chin-Yew. 1997. Robust Automated Topic Identification. *Ph.D. Thesis*. University of Southern California, August 1997.
- Hovy, E.H. and Chin-Yew Lin. 1997. Automated Text Summarization in SUMMARIST. In *Proceedings of the ACL97/EACL97 Workshop on Intelligent Scalable Text Summarization*. Madrid, Spain, July 1997.
- Lin, Chin-Yew and E.H. Hovy. 1997. Identifying Topics by Position. In *Proceedings of the 5<sup>th</sup> Conference on Applied Natural Language Processing (ANLP)*. Washington, D.C., March 1997.
- Lin, Chin-Yew 1995. Knowledge Based Automated Topic Identification. In *Proceedings of the 33<sup>rd</sup> Annual Meeting of the Association for Computational Linguistics*. Cambridge, Massachusetts, USA, June 1995.

## DEMONSTATION AND PRESENTATION

- L. Zhou, E. Shaw, Chin-Yew Lin, and E.H. Hovy. Classsummary: Introducing Discussion Summarization to Online Classrooms. In *Proceedings of 2005 Human Language Technology Conference (HLT/EMNLP 2005)*, Vancouver, B.C., Canada. □
- Leuski, A., Chin-Yew Lin, and E.H. Hovy. iNeATS: Interactive Multi-Document Summarization. In *Proceedings of the 41<sup>st</sup> Annual Meeting of the Association for Computational Linguistics (ACL 2003)*, Interactive Posters and Demos Session, Sapporo, Japan, July 7 – 12, 2003.
- Hermjakob, U., E.H. Hovy, and Chin-Yew Lin. Automated Question Answering in Webclopedia. In *Proceedings of the 40<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (ACL-02)*, Philadelphia, PA, U.S.A., July 7-12, 2002.
- Hermjakob, U., E.H. Hovy, and Chin-Yew Lin. Automated Question Answering in Webclopedia - A Demonstration. In *Proceedings of the Human Language Technology Conference (HLT2002)*, San Diego, CA, U.S.A., March 23-27, 2002.
- Lin, Chin-Yew. 2000. Text Summarization, Q&A, and Multilingual Information Access. In *Workshop on Interactive Searching in Foreign-Language Collections*, Human-Computer Interaction Laboratory, University of Maryland, MD, U.S.A., June 1, 2000.
- Lin, Chin-Yew. 1999. Multilingual Information Access: The MuST System and its Integration with Systran. In *Information Management and Intelligent Collaboration & Visualization Principal Investigators' Meeting*, San Diego, CA, U.S.A., November 2-5, 1999.
- Lin, Chin-Yew. 1999. Information Access across the Language Barrier: the MuST System. In *Proceedings of the 22<sup>nd</sup> International Conference on Research and Development in Information Retrieval*, Berkeley, CA, U.S.A, August 15-19, 1999.

- Hovy, E.H. and Chin-Yew Lin. 1998. SUMMARIST: Current Status and Future Work. In *Proceedings of TIPSTER Text Phase III 24-Month Workshop*, Baltimore, MD, U.S.A., October 13-15, 1998.
- Hovy, E.H. and Chin-Yew Lin. 1998. Current and Future Developments of SUMMARIST. In *Proceedings of TIPSTER Text Phase III 18-Month Workshop*, Fairfax, VA, U.S.A., May 4-6, 1998.
- Hovy, E.H. and Chin-Yew Lin. 1997. The Assembly of Topic Extraction Modules in SUMMARIST. In *Proceedings of TIPSTER Text Phase III 12-Month Workshop*, San Diego, CA, U.S.A., October 19-22, 1997.
- Lin, Chin-Yew and E.H. Hovy. 1997. Identifying Topics by Signatures. In *Proceedings of TIPSTER Text Phase III 6-Month Workshop*, Columbia, MD, U.S.A., May 11-14, 1997.

## TEACHING

- Chin-Yew Lin. 2006. Personalized Information Delivery with NLP and Search Technologies. In *Microsoft Research Asia Summer School 2006 on Next Generation World Wide Web Technologies*, Shenzhen, China, July 9 – 11, 2006.
- Chin-Yew Lin. 2005. Automated Text Summarization. In *Tutorial of the 2<sup>nd</sup> International Joint Conference on Natural Language Processing (IJCNLP-05)*, Jeju Island, Korea, October 11-13, 2005.
- Chin-Yew Lin. 2005. Text Summarization. In *Microsoft Research Asia Summer School 2005 on Information Extraction Technologies*, Harbin, China, June 1 – 3, 2005.
- Hovy, E.H. and Chin-Yew Lin and Daniel Marcu. 1999. Automated Text Summarization. In *Tutorial of the 22<sup>nd</sup> International Conference on Research and Development in Information Retrieval*, Berkeley, CA, U.S.A, August 15-19, 1999.

## WORK EXPERIENCE

- March 2006 — present: Group Manager of Web Intelligence Group  
 Microsoft Research Asia  
 5F, Beijing Sigma Center, No. 49, Zhichun Road, Haidian District,  
 Beijing 100080, P.R. China  
 +86 (10) 5896-3481  
 Developing advanced content analysis technologies, understanding user tasks and activities through analysis of user actions and fostering task-oriented social networks for next generation Internet applications. For more information, please see: <http://research.microsoft.com/en-us/groups/WIT/default.aspx>.
- August 1997 — February 2006: Senior Research Scientist  
 Information Sciences Institute  
 University of Southern California  
 4676 Admiralty Way  
 Marina del Rey, CA 90292-6695

(310) 448-8711

Developed new techniques for automatic learning knowledge from textbooks and automatic evaluations of summarization and machine translation, ROUGE and BE; Constructing prototypes for automated text summarization projects, SUMMARIST and NeATS; Integrating information retrieval, text summarization, and machine translation technologies in MuST; Conducting research to build targeted question and answering systems in Webclopedia and TextMap.

- June 1994 — July 1997: Graduate Research Assistant

*Advisor:* Dr. Eduard H. Hovy  
Information Sciences Institute  
University of Southern California  
4676 Admiralty Way  
Marina del Rey, CA 90292-6695  
(310) 822-1511 ext. 711

Developed topic identification modules for text processing tasks in the SUMMARIST project.

- January 1992 — May 1994: Research Assistant

*Advisor:* Professor Dan Moldovan  
Department of Electrical Engineering — Systems  
University of Southern California  
Hughes Aircraft Electrical Engineering Center  
EEB 100  
Los Angeles, CA 90089-2560

Joined the Parallel Knowledge Processing Laboratory (PKPL) at USC in January 1992. Maintained a 144-processor parallel processing machine prototype, SNAP I, and wrote system software for this machine. Participated in the fourth and fifth Message Understanding Conference (MUC-4 and MUC-5) with other PKPL team members using information extraction software developed for SNAP I.

- September 1991 — December 1991: Laboratory Assistant

Department of Electrical Engineering — Systems  
University of Southern California  
Hughes Aircraft Electrical Engineering Center  
EEB 100  
Los Angeles, CA 90089-2560

Was appointed laboratory assistant of the VLSI laboratory course of the Department of Electrical Engineering (course number EE577L). Was responsible for helping students use VLSI layout tool, MAGIC, and simulation tools, SPICE, IRSIM, and CRYSTAL.

- September 1989 — June 1990 : Teaching Assistant

Department of Electrical and Control Engineering  
National Chiao Tung University  
1001 Ta Hsueh Road

Hsinchu, Taiwan 30050

Was appointed teaching assistant in the Department of the Electrical and Control Engineering. Designed and taught sophomore- and junior-level microcomputer system laboratory courses including x86 assembly language and hardware implementation.

- January 1986 — June 1987: Research Assistant  
*Advisor:* Professor Pei-Chong Tang  
Department of Electrical and Control Engineering  
National Chiao Tung University  
1001 Ta Hsueh Road  
Hsinchu, Taiwan 30050

Developed a graphical user interface package on DOS using C for a joint project between the Computer Vision and Pattern Recognition group of the Department of Electrical and Control Engineering and the Energy and Resources group of the Industrial Technology Research Institute (ITRI) in Hsinchu, Taiwan.

## **PROFESSIONAL ACTIVITIES**

- Program Committee Co-chair AIRS-2009 ([www.kb.ist.hokudai.ac.jp/AIRS2009](http://www.kb.ist.hokudai.ac.jp/AIRS2009)).
- Area Chair of Summarization and Generation EMNLP-2009 (<http://conferences.inf.ed.ac.uk/emnlp09/>).
- Panelist of Social Network Mining and Search, Panel of 25<sup>th</sup> International Conference on Data Engineering (<http://i.cs.hku.hk/icde2009/panels.htm>).
- Program Committee Co-chair WWW 2009 Workshop on Web Search Result Summarization and Presentation (<http://www.wssp.info/2009.html>).
- Area Chair of Evaluation for ACL-2008 ([www.ling.ohio-state.edu/acl08](http://www.ling.ohio-state.edu/acl08)).
- Keynote speaker ROCLING 2008 ([forum.dmc.ntnu.edu.tw/~rocloing2008](http://forum.dmc.ntnu.edu.tw/~rocloing2008)).
- ACLIA Track Coordinator for NTCIR-7 ([ntcir.nii.ac.jp](http://ntcir.nii.ac.jp)).
- External Reviewer for the Research Grants Council (RGC) of Hong Kong. January 2006 – present.
- Member of the Program Committee for the Workshop on Analyzing Conversations in Text and Speech (ACTS, a HLT-NAACL-2006 workshop), June 8, 2006, New York City, USA.
- Co-coordinator of the Opinion Workshop in the NTCIR Workshop. November 2005 – present.
- Member of Advisory Board of the NTCIR Workshop. November 2005 – present.
- Member of the Editorial Board of the Journal of Computational Linguistics and Chinese Language Processing (CLCLP) May 2004 – April 2008.
- Area Chair of Question Answering and Summarization for ACL-2005.
- Co-chair of the Workshop on Intrinsic and Extrinsic Evaluation for MT and/or Summarization (MTSE) 2005.
- Member of the Program Committee for the Workshop on Crossing Barriers in Text Summarization Research, post-conference workshop of RANLP 2005.

- Member of the Program Committee for the 10<sup>th</sup> Conference on Artificial Intelligence and Applications (TAAI-2005).
- Guest Editor, ACM Transactions on Asian Language Information Processing (TALIP) – Special Issue: Recent Advances in Statistical Language Modeling – Beyond N-gram.
- Co-chair of the Workshop on Multilingual Summarization and Question Answering – Machine Learning and Beyond to be held from July 11 to 12, 2003, in conjunction with ACL-2003 in Sapporo, Japan.
- Member of the Program Committee for the Text Summarization Workshop to be held on May 31 and June 1, 2003, in conjunction with HLT-NAACL 2003 in Edmonton, Canada.
- Member of the Program Committee for the Special Track on Recent Advances in Natural Language Processing to be held with the 16<sup>th</sup> International FLAIRS Conference from May 11 to 15, 2003, in St. Augustine, Florida, U.S.A.
- Member of the Program Committee for the 13<sup>th</sup> International Workshop on Research Issues on Data Engineering: Multilingual Information Management (RIDE-MLIM'2003) to be held from March 10 to 11, 2003, in conjunction with International Conference on Data Engineering 2003 in Hyderabad, India.
- Co-chair of the Workshop on Multilingual Summarization and Question Answering held from August 31 to September 1, 2002, in conjunction with COLING-2002 in Taipei, Taiwan.
- Member of the Program Committee for the Workshop on Cross-Language Information Retrieval: A Research Roadmap held on August 15, 2002, in conjunction with the 25<sup>th</sup> International Conference on Research and Development in Information Retrieval (SIGIR-2002) in Tampere, Finland.
- Project affiliate of the CLSP Workshop 2001 on Automatic Summarization of Multiple (Multilingual) Documents held from June 25 to August 17, 2001, in the Center for Language and Speech Processing, John Hopkins University, Baltimore, MD, U.S.A.
- Co-chair of the Workshop on Automatic Summarization (WAS 2001) held in Pittsburgh, PA, U.S.A., on June 3, 2001.
- Member of the Question and Answering Roadmap Committee of DARPA TIDES program.
- Co-chair of the Workshop on Automatic Summarization (WAS 2000) held in Seattle, WA, U.S.A. on April 30, 2000.
- I have presented a tutorial, Automated Text Summarization, with Eduard Hovy and Daniel Marcu at the 22<sup>nd</sup> International Conference on Research and Development in Information Retrieval (SIGIR'99).
- I have reviewed papers for the ACM Transactions on Information Systems (TOIS), the ACM Transactions on Asian Language Information Processing (TALIP), the Journal of Information Processing and Management, the Journal of Information Retrieval, the Information Processing Letter (IPL), Fundamenta Informaticae, the Journal of Computational Linguistics and Chinese Language Processing (CLCLP), the South African Computer Journal, the International Joint Conference on Artificial Intelligence (IJCAI), the International Conference on Computational Linguistics (COLING), the Applied Natural Language Processing

- Conference (ANLP), the Annual Meeting of the Association for Computational Linguistics (ACL), the North American Association for Computational Linguistics conferences (NAACL), the Human Language Technology Conference (HLT), the International Joint Conference on Natural Language Processing (IJCNLP), Asia Information Retrieval Symposium (AIRS), and the Conference on Artificial Intelligence and Applications (TAAI).
- Member of the Association for Computational Linguistics (ACL), the ACL Special Interest Group on Chinese Language Processing (SIGHAN), and the Association for Computational Linguistics and Chinese Language Processing (ACLCLP).

## TALKS

I gave keynote speeches at the NTICR Workshop 5, 2005, in Japan and the Workshop on Machine Translation Evaluation in Machine Translation Summit IX, 2003. I was also invited to give presentations at the 2<sup>nd</sup> Young Scholar Symposium on Natural Language Processing, (YSS-NLP'2005) in Harbin, China and NTCIR Workshop 4, 2003, in Japan. I have also delivered talks at a number of universities and organizations. These include the University of Southern California, Los Angeles; Information Sciences Institute, Marina del Rey; University of Maryland, College Park; Southern Methodist University, Dallas; Academia Sinica, Taipei, Taiwan; National Tsing-Hua University, Taiwan; National Taiwan University, Taiwan; National Chung Cheng University, Taiwan; National Cheng Kung University, Taiwan; National Taiwan Normal University, Taiwan; Behavior Design Corporation, Taiwan. I also presented papers at the Government's invitation-only TREC and TIPSTER conferences.

## SOFTWARE PACKAGES

- BE – Basic Element Automatic Evaluation Package for Summarization  
It will be used in Document Understanding Conference (DUC) 2005 and the Multilingual Summarization Evaluation (MSE) 2005 as the official automatic evaluation package. (<http://www.isi.edu/~cyl/BE>)
- ROUGE – An Automatic Evaluation Package for Summarization  
It was used by Document Understanding Conference (DUC) 2004 as the official automatic evaluation tool. Over 130 sites around the world have downloaded this package. It will also be used in DUC 2005 and the Multilingual Summarization Evaluation (MSE) 2005 as one of the two official automatic evaluation packages. (<http://www.isi.edu/~cyl/ROUGE>)
- C\*ST\*RD (Clustering, Summarization, Translation, Retrieval, and Display)  
Created with A. Leuski, L. Zhou, U. Germann, F.J. Och, and E.H. Hovy for DARPA TIDES Surprise Language Exercise to process Hindi language.
- SEE – Summary Evaluation Environment  
It is a graphical summary evaluation environment and has been used by NIST in DUC 2001, 2002, 2003 and 2004 to carry out manual evaluation. (<http://www.isi.edu/~cyl/SEE>)

## **PROGRAMMING LANGUAGES**

- C, C++, PERL, Ruby, Java, JavaScript, SQL, Tcl/Tk, LISP, PASCAL, FORTRAN, BASIC.

## **SYSTEM EXPERIENCE**

- UNIX, Linux, X Window System, Windows NT/2000, Windows 95/98/ME, Windows 3.x, OS/2, Mac OS X, DOS, Macintosh.

## **REFERENCES**

Available on request from:

### **Ms. Donna Harman**

Information Access Division (894)  
National Institute of Standards and Technology (NIST)  
100 Bureau Drive, Stop 8940  
Gaithersburg, MD 20899-8940  
Phone: (301) 975-3569  
Email: donna.harman@nist.gov

### **Dr. Eduard Hovy**

Information Sciences Institute (ISI)/University of Southern California  
Director of the Natural Language Group  
Deputy Directory of the Intelligent Systems Division of ISI  
Director of Research of the Digital Government Research Center (DGRC)  
4676 Admiralty Way  
Marina del Rey, CA 90292-6695, USA  
Phone: (310) 448-8731, Fax: (310) 823-6714  
Email: hovy@isi.edu

### **Dr. Hsiao-Wuen Hon**

Managing Director of Microsoft Research Asia  
5F, Beijing Sigma Center  
No. 49, Zhichun Rd, Haidian District  
Beijing 100190, China  
Phone: (86) 10-58968888  
Email: hon@microsoft.com

### **Dr. Keh-Yih Su**

Behavior Design Corporation  
2F, No. 5, Industry E. Road IV/Science-Based Industrial Park  
Hsinchu, Taiwan  
Phone: (886) 3-5770243  
Email: kysu@bdc.com.tw