

## Seokhwan Kim

---

### CONTACT INFORMATION

Human Language Technology Department  
Institute for Infocomm Research  
1 Fusionopolis Way, #21-01 Connexis  
Singapore 138632

*Office:* +65 6408 2766  
*Mobile:* +65 9776 1740  
*E-mail:* kims@i2r.a-star.edu.sg  
*WWW:* <http://seokhwankim.com/>

### RESEARCH INTERESTS

Natural Language Processing, Spoken Dialogue Systems, Natural Language Understanding, Text Mining

### EDUCATION

**Pohang University of Science and Technology (POSTECH)**, Pohang, Korea

*Ph.D., Computer Science and Engineering*

*February 2012*

Dissertation Topic: “Cross-Lingual Weakly-Supervised Learning of Semantic Relations”

Advisor: Gary Geunbae Lee

Committee: Jong-hyeok Lee, Seung-jin Choi, Seung-won Hwang, Sung-Hyon Myaeng

*B.S., Computer Science and Engineering*

*August 2005*

### RESEARCH EXPERIENCE

**Institute for Infocomm Research (I2R)**, Singapore

*Scientist*

*January, 2012 - present*

**COLDAR project**, Developed components for multi-lingual natural language understanding and dialog management.

**TIRS project**, Developed components for natural language understanding and database query generation.

**DURIAN project**, Developed state tracking and variable sharing mechanisms across command, Q&A, dialogue and chat engines.

**Max Planck Institute for Informatics (MPI-INF)**, Saarbrücken, Germany

*Research Intern*

*April, 2011 - August, 2011*

*Advisor: Prof. Gerhard Weikum*

**DIDO project**, Developed a higher-order noun phrase chunker for recognizing bio-medical entities with environmental contexts.

**Pohang University of Science and Technology (POSTECH)**, Pohang, Korea

*Research Assistant*

*September, 2005 - December, 2011*

*Advisor: Prof. Gary Geunbae Lee*

**Development of Dialog-based Question Answering System for TV contents**, Dec 2010 - Oct 2011, LG Electronics. Developed components for robust natural language understanding.

**Development of Content Recommendation System for Digital TV**, Oct 2007 - Oct 2008, Samsung Electronics. Developed an automatic content feeding component for content-based recommendation algorithm.

**Semi-supervised Information Extraction for Automatic Content Feeding in Information Access Dialog Systems**, Apr 2007 - Mar 2008, Microsoft Research Asia. Developed an algorithm of semi-supervised information extraction.

**Implementation of Dialog System Components to Upgrade Chatbot**, May 2006 - Oct 2006, SK Telecom Co., Ltd. Developed components for natural language understanding and database management.

**Development of Natural Language Dialog System for Database Search and Entertainment**, Jul 2005 - Dec 2005, KT Corporation. Developed named-entity recognizer for

natural language understanding.

TEACHING  
EXPERIENCE

**Pohang University of Science and Technology (POSTECH)**, Pohang, Korea

*Teaching Assistant*

*Spring 2010*

*Professor: Gary Geunbae Lee*

**CS101, Introduction to Computing**

*Teaching Assistant*

*Spring 2006*

*Professor: Byoung-kee Yi*

**CS421, Database System**

PROFESSIONAL  
EXPERIENCE

**The Fifth Dialog State Tracking Challenge (DSTC5)**

*Organizing Committee Co-chair*

*2016*

**The Fourth Dialog State Tracking Challenge (DSTC4)**

*Organizing Committee Co-chair*

*2015*

**The Ninth Asia Information Retrieval Societies Conference (AIRS 2013)**

*Local Organizing Committee Co-chair*

*2013*

**ACL Special Workshop: Rediscovering 50 Years of Discoveries**

*Publicity Committee Member*

*2012*

**ACM-SIGIR, IEEE ICASSP, IEEE/ACM TASLP, IJCAI, IJCNLP, IWSDS, International Journal of Social Robotics, Interspeech, NAACL-HLT**

*Reviewer*

PUBLICATIONS

**International Journals**

*Seokhwan Kim, Minwoo Jeong, Jonghoon Lee, Gary Geunbae Lee.* **Cross-Lingual Annotation Projection for Weakly-Supervised Relation Extraction.** ACM Transactions on Asian Language Information Processing, 13:1 (3), Feb 2014.

*Seokhwan Kim, Minwoo Jeong, Gary Geunbae Lee.* **A Local Tree Alignment Approach to Relation Extraction of Multiple Arguments.** Information Processing and Management, 47:4 (593-605), Jul 2011.

*Cheongjae Lee, Sangkeun Jung, Seokhwan Kim, Gary Geunbae Lee.* **Example-based dialog modeling for practical multi-domain dialog system.** Speech Communications, 51:5 (466-484), May 2009.

*Sangkeun Jung, Cheongjae Lee, Seokhwan Kim, Gary Geunbae Lee.* **DialogStudio: A workbench for data-driven spoken dialog system development and management.** Speech Communications, 50:8-9 (683-697), , Aug-Sep 2008.

**International Conferences**

*Seokhwan Kim, Rafael E. Banchs, Haizhou Li.* **Exploring Convolutional and Recurrent Neural Networks in Sequential Labelling for Dialogue Topic Tracking.** To appear in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016), Berlin, Aug 2016.

*Seokhwan Kim, Luis Fernando D'Haro, Rafael E. Banchs, Jason D. Williams, Matthew Henderson.* **The Fourth Dialog State Tracking Challenge.** Proceedings of the 7th International Workshop on Spoken Dialogue Systems (IWSDS 2016), Saariseik, Jan 2016.

*Ridong Jiang, Rafael Banchs, Seokhwan Kim, Luis Fernando D'Haro, Andreea Niculescu, Kheng Hui Yeo.* **Configuration of Dialogue Agent with Multiple Knowledge Sources.** Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2015 (APSIPA ASC 2015), Hong Kong, Dec 2015.

*Luis Fernando D'Haro, Seokhwan Kim, Rafael E. Banchs.* **A Robust Spoken Q&A System with Scarce In-Domain Resources.** Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2015 (APSIPA ASC 2015), Hong Kong, Dec 2015.

*Ridong Jiang, Seokhwan Kim, Rafael E. Banchs, Haizhou Li.* **Towards Improving the Performance of Vector Space Model for Chinese Frequently Asked Question.** Proceedings of the 19th International Conference on Asian Language Processing (IALP 2015), Suzhou, Oct 2015.

*Seokhwan Kim, Rafael E. Banchs, Haizhou Li.* **Wikification of Concept Mentions within Spoken Dialogues Using Domain Constraints from Wikipedia.** Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), pp 2225-2229, Lisbon, Sep 2015. (24.0% acceptance)

*Seokhwan Kim, Rafael E. Banchs, Haizhou Li.* **Towards Improving Dialogue Topic Tracking Performances with Wikification of Concept Mentions.** Proceedings of the 16th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL 2015), pp 124-128, Prague, Sep 2015.

*Luis Fernando D'Haro, Seokhwan Kim, Rafael E. Banchs.* **Conversational Agent and Management Tools for Conference and Tourism Domain.** Proceedings of the 16th Annual Conference of the International Speech Communication Association (Interspeech 2015), Dresden, Sep 2015. (Show and Tell)

*Luis Fernando D'Haro, Seokhwan Kim, Kheng Hui Yeo, Ridong Jiang, Andreea I. Niculescu, Rafael E. Banchs, Haizhou Li.* **CLARA: a multifunctional virtual agent for conference support and touristic information.** Proceedings of the 6th International Workshop on Spoken Dialog System (IWSDS 2015), Busan, Jan 2015. (Demo)

*Kyusong Lee, Seonghan Ryu, Paul Hongsuck Seo, Seokhwan Kim, Gary Geunbae Lee.* **Grammatical Error Correction Based on Learner Comprehension Model in Oral Conversation.** Proceedings of the 2014 IEEE Spoken Language Technology Workshop (SLT 2014), South Lake Tahoe, Dec 2014.

*Seokhwan Kim, Rafael E. Banchs.* **R-cube: a dialogue agent for Restaurant Recommendation and Reservation.** Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2014 (APSIPA ASC 2014), Siem Reap, Dec 2014.

*Rafael E. Banchs, Seokhwan Kim.* **An empirical evaluation of an IR-based strategy for chat-oriented dialogue systems.** Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2014 (APSIPA ASC 2014), Siem Reap, Dec 2014.

*Andreea Niculescu, Kheng Hui Yeo, Luis Fernando D'haro, Seokhwan Kim, Ridong Jiang, Rafael Banchs.* **Design and Evaluation of a Conversational Agent for the Touristic Domain.** Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit

and Conference 2014 (APSIPA ASC 2014), Siem Reap, Dec 2014.

*Rafael E. Banchs, Seokhwan Kim.* **Spoken Dialogue System for Restaurant Recommendation and Reservation.** Proceedings of the 15th Annual Conference of the International Speech Communication Association (Interspeech 2014), pp 1488-1489, Singapore, Sep 2014. (Show and Tell)

*Andreea Niculescu, Seokhwan Kim, Kheng Hui Yeo, Arthur Niswar, Ridong Jiang, Rafael E. Banchs.* **SARA - Singapore's Automated Responsive Assistant for the Touristic Domain.** Proceedings of the 15th Annual Conference of the International Speech Communication Association (Interspeech 2014), pp 2138-2139, Singapore, Sep 2014. (Show and Tell)

*Andreea Niculescu, Rafael E. Banchs, Seokhwan Kim, Kheng Hui Yeo, Arthur Niswar, Ridong Jiang.* **SARA - Singapore's Automated Responsive Assistant - A Multimodal Dialogue System for Touristic Information.** Proceedings of the 11th International Conference on Mobile Web Information Systems (MobiWIS 2014), pp 153-164, Barcelona, Aug 2014.

*Seokhwan Kim, Rafael E. Banchs, Haizhou Li.* **A Composite Kernel Approach for Dialog Topic Tracking with Structured Domain Knowledge from Wikipedia.** Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), Vol. 2, pp 19-23, Baltimore, Jun 2014. (25.0% acceptance)

*Seokhwan Kim, Rafael E. Banchs.* **Sequential Labeling for Tracking Dynamic Dialog States.** Proceedings of the 15th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL 2014), pp 332-336, Philadelphia, Jun 2014.

*Arthur Niswar, Rafael E. Banchs, Ridong Jiang, Seokhwan Kim.* **AIDA: An Avatar-Supported Multimodal Dialogue System.** Proceedings of the 27th International Conference on Computer Animation and Social Agents (CASA 2014), Houston, May 2014.

*Seokhwan Kim, Rafael E. Banchs, Haizhou Li.* **Wikipedia-based Kernels for Dialogue Topic Tracking.** Proceedings of the 39th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014), pp 131-135, Florence, May 2014.

*Ridong Jiang, Rafael E. Banchs, Seokhwan Kim, Kheng Hui Yeo, Arthur Niswar, Haizhou Li.* **Web-based Multimodal Multi-domain Spoken Dialogue System.** Proceedings of the 5th International Workshop on Spoken Dialog Systems (IWSDS 2014), Napa, Jan 2014.

*Rafael E. Banchs, Ridong Jiang, Seokhwan Kim, Arthur Niswar, Kheng Hui Yeo.* **AIDA: Artificial Intelligent Dialogue Agent.** Proceedings of the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2013), pp 145-147, Metz, Aug 2013. (Demo)

*Seokhwan Kim.* **A Graph-based Cross-lingual Projection Approach for Spoken Language Understanding Portability to a New Language.** Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), pp 8332-8336, Vancouver, May 2013.

*Injae Lee, Seokhwan Kim, Kyungduk Kim, Donghyeon Lee, Junhwi Choi, Seonghan Ryu, Geunbae Lee.* **A Two-Step Approach for Efficient Domain Selection in Multi-Domain Dialog Systems.** Proceedings of the 4th International Workshop on Spoken Dialog System (IWSDS 2012), pp 105-111, Ermenonville, Nov 2012.

*Seokhwan Kim, Gary Geunbae Lee.* **A Graph-based Cross-lingual Projection Approach for Weakly Supervised Relation Extraction.** Proceedings of the 50th Annual Meeting of

the Association for Computational Linguistics (ACL 2012), Vol. 2, pp 48-53, Jeju, Jul 2012. (20.0% acceptance)

*Hongsuck Seo, Jonghoon Lee, Seokhwan Kim, Kyusong Lee, Sechun Kang and Gary Geunbae Lee.* **A Meta Learning Approach to Grammatical Error Correction.** Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012), Vol. 2, pp 328-332, Jeju, Jul 2012. (20.0% acceptance)

*Junhwi Choi, Kyungduk Kim, Sungjin Lee, Seokhwan Kim, Donghyeon Lee, Injae Lee, Gary Geunbae Lee.* **Seamless error correction interface for voice word processor.** Proceedings of the 37th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012), pp 4973-4976, Kyoto, Mar 2012.

*Junhwi Choi, Kyungduk Kim, Seokhwan Kim, Donghyeon Lee, Injae Lee, Gary Geunbae Lee.* **Multi-domain spoken dialog system for information access in mobile environment.** Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop (IEEE ASRU 2011), Hawaii, Dec 2011. (Demo)

*Seokhwan Kim, Minwoo Jeong, Jonghoon Lee, Gary Geunbae Lee.* **A Cross-lingual Annotation Projection-based Self-supervision Approach for Open Information Extraction.** Proceedings of the 5th international joint conference on natural language processing (IJCNLP 2011), pp 741-748, Chiang Mai, Nov 2011. (27.0% acceptance)

*Donghyeon Lee, Cheongjae Lee, Minwoo Jeong, Kyungduk Kim, Seokhwan Kim, Junhwi Choi, Gary Geunbae Lee.* **Web-search enhanced contents retrieval for information access dialog system.** Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech 2011), pp 1297-1300, Florence, Aug 2011.

*Seokhwan Kim, Minwoo Jeong, Jonghoon Lee, Gary Geunbae Lee.* **A Cross-lingual Annotation Projection Approach for Relation Detection.** Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010), pp 564-571, Beijing, Aug 2010. (19.0% acceptance)

*Seokhwan Kim, Minwoo Jeong, Gary Geunbae Lee.* **A Local Tree Alignment-based Soft Pattern Matching Approach for Information Extraction.** Proceedings of the North American Chapter of the Association for Computational Linguistics/Human Language Technology (NAACL HLT 2009), pp 169-172, Colorado, May 2009.

*Seokhwan Kim, Minwoo Jeong, Gary Geunbae Lee.* **An Alignment-based Pattern Representation Model for Information Extraction.** Proceedings of the 31st Annual International ACM SIGIR Conference (SIGIR '08), pp 875-876, Singapore, Jul 2008.

*Seokhwan Kim, Minwoo Jeong, Gary Geunbae Lee. Kwangil Ko, Zino Lee.* **An alignment-based approach to semi-supervised relation extraction including multiple arguments.** Proceedings of the fourth Asian Information Retrieval Symposium (AIRS 2008), Harbin, Jan 2008.

*Cheongjae Lee, Seokhwan Kim, Minwoo Jeong, Sangkeun Jung, Donghyeon Lee, Gary Geunbae Lee.* **Example-based spoken dialog processing for guidance robots.** Proceedings of the 4th international conference on ubiquitous robots and ambient intelligence (URAI 2007), Pohang, Nov 2007.

*Seokhwan Kim, Minwoo Jeong, Gary Geunbae Lee.* **A semi-supervised method for efficient construction of statistical spoken language understanding resources.** Proceedings of the 8th Annual Conference of the International Speech Communication Association (Interspeech

2007), pp 2797-2800, Antwerp, Aug 2007.

*Seokhwan Kim, Cheongjae Lee, Sangkeun Jung, Gary Geunbae Lee.* **A spoken dialog system for electronic program guide information access.** Proceedings of The 16th IEEE International Symposium on Robot and Human Interactive Communication (IEEE RO-MAN 2007), pp 178-181, Jeju, Aug 2007.

*Seokhwan Kim, Yu Song, Kyungduk Kim, Jeong-Won Cha, Gary Geunbae Lee.* **MMR-based active machine learning for Bio named entity recognition.** Proceedings of the Human Language Technology/North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2006), pp 69-72, New York, Jun 2006.

## Patents

*Cheongjae Lee, Seokhwan Kim, Gary Geunbae Lee.* **Dialog management apparatus and method for chatting agent.** Registration #10-0818979, South Korea, 2008. 3. 27

*Seokhwan Kim, Yu Song, Kyungduk Kim, Gary Geunbae Lee.* **Method and system for recognizing biological named entity based on workbench.** Registration #10-0825687, South Korea, 2008. 4. 22

*Seokhwan Kim, Hyungjong Noh, Gary Geunbae Lee.* **Method and apparatus for automatically constructing ontology from non-structure web documents.** Registration #10-0917176, South Korea, 2009. 9. 7  
Registration #, South Korea,

*Seokhwan Kim, Donghyeon Lee, Hyungjong Noh, Gary Geunbae Lee.* **Method of automatically detecting dangerous situation using natural language extraction and apparatus thereof.** Registration #10-1023031, South Korea, 2011. 3. 10

*Donghyeon Lee, Hyungjong Noh, Seokhwan Kim, Gary Geunbae Lee.* **Method and apparatus for searching voice data from audio and video data under the circumstances including unregistered words.** Registration #10-1069534, South Korea, 2011. 9. 26

*Hyungjong Noh, Seokhwan Kim, Donghyeon Lee, Gary Geunbae Lee.* **User profile automatic creation apparatus through voice dialog meaning process, and contents recommendation apparatus using the same.** Registration #10-1072176, South Korea, 2011. 10. 4

*Gary Geunbae Lee, Seokhwan Kim, Kyungduk Kim, Donghyeon Lee, Junhwi Choi.* **Method for searching for information using the web and method for voice conversation using same.** PCT/KR2012/004405, 2012. 8. 28

*Gary Geunbae Lee, Seokhwan Kim, Kyungduk Kim, Donghyeon Lee, Junhwi Choi.* **Korean Open Type Information Extracting Method and Program Readable Recording Medium Capable of Increasing Korean Open Type Information Extracting Performances.** Registration #10-1180589, South Korea, 2012. 8. 31

*Gary Geunbae Lee, Junhwi Choi, Seokhwan Kim, Kyungduk Kim, Donghyeon Lee.* **Speech Processing Device Capable of Effective Input Modification and a Method Thereof.** Registration #10-1197010, South Korea, 2012. 10. 26

*Donghyeon Lee, Kyungduk Kim, Junhwi Choi, Seokhwan Kim, Gary Geunbae Lee.* **Information Search Method by Using the Web Capable of Using a Vector Space Database and a Voice Conversaion Method Using the Method.** Registration #10-1252397, South Korea,

2013. 4. 2

*Donghyeon Lee, Kyungduk Kim, Junhwi Choi, Seokhwan Kim, Gary Geunbae Lee.* **Conversation Classification Method Capable of Classifying Conversation Intention by Using a Hidden Markov Model.** Registration #10-1255468, South Korea, 2013. 4. 10

*Gary Geunbae Lee, Seokhwan Kim, Kyungduk Kim, Donghyeon Lee, Junhwi Choi.* **Entity Name Tagging Method Capable of Applying an Unsupervised Learning Method Including a Restriction Condition and a Device Thereof.** Registration #10-1255957, South Korea, 2013. 4. 11

#### SKILLS

**Programming Languages:** Python, C++, Java

**Machine Learning Toolkits:** scikit-learn, Theano, SVM-Light, CRFsuite

**NLP and IR Toolkits:** NLTK, Stanford CoreNLP, Lucene

**Databases:** MySQL, SQLite, MongoDB

#### REFERENCES

Available upon request