

STEVEN EUIJONG WHANG

CONTACT INFORMATION

Address: 344 Olmsted Rd., #245, Stanford, CA 94305

Cell: 650-796-6006

E-mail: swhang@cs.stanford.edu

Homepage: <http://www.stanford.edu/~euijong>

EDUCATION

Stanford University

Ph.D. student in Computer Science, Advisor: **Hector Garcia-Molina**

M.S. in Computer Science with Distinction in Research (Database Specialization)

Thesis: Generic Entity Resolution with Negative Rules

Advisor: Jennifer Widom

Stanford, CA

3rd year of program

June 2007

Korea Advanced Institute of Science and Technology (KAIST)

B.S. in Computer Science

GPA: 4.05/4.3 (Class 2003: 1st among 96 in the Computer Science department)

Daejeon, Korea

Feb. 2003

HONORS AND AWARDS

- Stanford University School of Engineering Fellowship, 2007-8
- Korea Foundation for Advanced Studies (KFAS) Fellowship, 2005-7
- KAIST President's Prize for Academic Excellence at Commencement, KAIST, Feb. 2003

RESEARCH AND WORK EXPERIENCE

- **Stanford University, Stanford InfoLab member**, *June 2006-Present*
Advisor: Hector Garcia-Molina
My research at Stanford has been focused on various techniques for generic entity resolution (also known as deduplication or record linkage). I am currently interested in how entity resolution relates to privacy.
- **Yahoo! Research, Santa Clara, Research Intern in the APEX group**, *June 2008-Dec. 2008*
Mentors: Jayavel Shanmugasundaram, Ramana Yerneni
We developed efficient indexing techniques for complex boolean expressions using inverted lists. Applications include display advertising and, in general, publish/subscribe systems. Paper: VLDB '09
- **HP Labs, Palo Alto, Research Intern in the Advanced Business Intelligence Group**, *Summer 2007*
Mentors: Malu Castellanos, Umeshwar Dayal
We developed an upfront recommender for physical database design for the HP Neoview Data Warehouse product. Paper: ICDE '09 (demo)
- **Advanced Information Technology Research Center, KAIST, Research Assistant**, *Feb. 2005-May 2005*
We developed efficient techniques for detecting near-duplicate web pages for the ODYS web search engine.
- **Orom Info. Co., Ltd. (Daejeon, Korea), Software Engineer**, *Jan. 2003-May 2005*
We developed and maintained a nation-wide digital library web service for elementary and high schools in Korea.

TEACHING EXPERIENCE

- **Instructor for the Stanford course “CS245: Database System Principles,” Summer quarter, 2009**
Number of students: 34, Instructor mean score: 4.24/5, Course mean score: 4.29/5
Favorite course evaluation from anonymous student: *Steven definitely cared about learning. He came up problems and did a lot of prep outside of class. I also liked the way class was rather discussion oriented- this is a great teaching trait he shares with his advisor (who guest lectured a few times). Steven was also very available for consultation outside of class, and did a good job explaining the course concepts. If he stays in academia he will be a great professor.*
- **Course Assistant for the Stanford course “CS245: Database System Principles,” Winter quarter, 2009**

PUBLICATIONS

- **Evaluating Entity Resolution Results, Technical Report, 2009**
David Menestrina, Steven Euijong Whang, Hector Garcia-Molina
- **Indexing Boolean Expressions, VLDB, 2009**
Steven Euijong Whang, Chad Brower, Jayavel Shanmugasundaram, Sergei Vassilvitskii, Erik Vee, Ramana Yerneni, Hector Garcia-Molina
- **Entity Resolution with Iterative Blocking, SIGMOD, 2009**
Steven Euijong Whang, David Menestrina, Georgia Koutrika, Martin Theobald, Hector Garcia-Molina
- **Generic Entity Resolution with Negative Rules, VLDB Journal, 2009**
Steven Euijong Whang, Omar Benjelloun, Hector Garcia-Molina
- **QuickStart: An Upfront Client-based Design Advisor for Parallel Data Warehouses, ICDE, 2009 (Demo)**
Malu Castellanos, Ivo Jimenez, Neal Coddington, Hans Zeller, Steven Whang, Umeshwar Dayal
- **Swoosh: A Generic Approach to Entity Resolution, VLDB Journal, 2009**
Omar Benjelloun, Hector Garcia-Molina, David Menestrina, Qi Su, Steven Euijong Whang, Jennifer Widom (Algorithms published in Chapter 21.7 of the textbook “Database Systems: The Complete Book, 2nd Ed.”)
- **A Practitioner’s Approach to Normalizing XQuery Expressions, DASFAA, 2006**
Ki-Hoon Lee, Seo-Young Kim, Euijong Whang, Jae-Gil Lee

TALKS

- Stanford Entity Resolution Framework, Stanford-Berkeley exchange, Berkeley, Apr. 2009
- Swoosh: A Generic Approach to Entity Resolution, guest lecture given in the Stanford course “CS345C: Data Integration,” May 2008
- Generic Entity Resolution with Negative Rules, InfoLab/Hitachi Workshop, Stanford, Mar. 2008

COMPUTER SKILLS

- Programming Languages: Java, C/C++/C#, Python, Scheme, SQL, VB, ASP, JSP, Perl
- Database Applications: Oracle, SQL Server, MySql

HOBBIES

- 5th KAIST Alumni President at Stanford, 2007-8
- Violinist: Volunteer work in church for over 10 years
- Swimmer: Won a bronze medal in the 4th National Fin Swimming Competition, Korea, 2000
- Marathon Runner: Current half-marathon (21.097km) record is 1:59:09