

Curriculum Vitae — AnHai Doan

(updated on Feb 11, 2014)

University of Wisconsin
Computer Sciences Department
1210 W. Dayton St., Madison, WI 53706

Office: (608) 262 9759
anhai@cs.wisc.edu
<http://www.cs.wisc.edu/~anhai>

RESEARCH INTERESTS

Managing heterogeneous, unstructured, and messy data; database systems; information integration; crowdsourcing; knowledge base construction; information extraction; Web and text management
Statistics: 8325 citations (by 2/11/2014 on Google Scholar), h-index of 40.

EDUCATION

Ph.D Computer Science, **University of Washington** 2002
Dissertation: *Learning to Translate between Structured Representations of Data*
(received ACM Doctoral Dissertation Award, 2003)
Advisors: Alon Halevy & Pedro Domingos

M.S. Computer Science, **University of Wisconsin-Milwaukee** 1996

B.S. Computer Science (*summa cum laude*), **Kossuth Lajos University**, Hungary 1993

ACADEMIC POSITIONS

Associate Professor, Computer Science, University of Wisconsin, Madison Jul. 2008-present

Assistant Professor, Computer Science, University of Wisconsin, Madison Aug. 2006-Jul. 2008

Assistant Professor, Computer Science, University of Illinois, Urbana Aug. 2002-Aug. 2006

Assistant Professor, Zero-time Appointment, Graduate School of
Library and Information Sciences, University of Illinois, Urbana Mar. 2004-Aug. 2006

Research Assistant, University of Washington, Seattle 1997-2002

Research Intern, Rockwell Science Laboratory, Palo Alto, CA 1996

Intern, Frontier Technologies Corp., Mequon, WI 1996

Research Assistant, University of Wisconsin, Milwaukee 1993-1996

Research Intern, Institute of Nuclear Research, Hungarian Academy of Sciences 1991

SELECTED INDUSTRIAL EXPERIENCE

Chief Scientist, @WalmartLabs May 2011-present

Member of Technical Advisory Board, Maana Corp. 2013-present

Member of Technical Advisory Board, Knode Corp. 2010-present

Chief Scientist, Kosmix Corp. Jun. 2010-Apr. 2011

Member of Technical Advisory Board, Transformic Corp. 2004-2005

AWARDS AND HONORS

Co-Organizer (with Mike Carey) of the Beckman Database Research Self Assessment Meeting (30 active database researchers discuss the prospects of the field for the next five years)	2013
Invitation to the Claremont Database Research Self Assessment Meeting (25 active database researchers discussed the prospects of the field for the next five years)	2008
Alfred P. Sloan Research Fellowship Award	2007
IBM Faculty Award	2007-2008
NSF CAREER Award	2004
CS Alumnus of the Year Award, University of Wisconsin-Milwaukee	2004
ACM Doctoral Dissertation Award	2003
William Chan Memorial Dissertation Award, University of Washington	2003
Incomplete List of Teachers Ranked as Excellent by Their Students, UIUC	2003-2004
Meritorious Service Award, Journal of Operations Research (for 25 of 300+ reviewers)	2003
Graduate School Fellowship, University of Wisconsin	1995-1996
University Fellowship, Kossuth Lajos University	1991-1993
America-Hungary Exchange Program Scholarship, Kossuth Lajos University	1993
Red Diploma (equivalent to <i>summa cum laude</i>), Kossuth Lajos University	1991
Government Scholarship for Undergraduate Studies in Hungary	1987
Member of the six-person team representing Vietnam at the 27th Int. Math. Olympiad	1986

SELECTED PROFESSIONAL SERVICE

Conferences and Workshops Organized

Co-Chair, Industrial Program, the Int. Conference on Very Large Databases (VLDB), 2015
Co-Organizer, the Beckman Database Research Self Assessment Meeting, 2013
Co-Chair, Workshop Program, the Int. Conference on Very Large Databases (VLDB), 2013
Chair, Industrial Program, the ACM SIGMOD Conference, Jun. 2012
Co-Chair, with Elaine Rich, the *AI Nectar track*, one of the four tracks that comprise the AAAI Conference, Aug. 2006
Co-Chair, with Frank Neven, the *Eighth Int. Workshop on the Web and Databases (WebDB)*, at the ACM SIGMOD Conference, Jun. 2005, with a focus on large-scale integration of Web data.
Co-Chair, with Fabio Ciravegna, Craig Knoblock, Nicholas Kushmerick, and Steffen Staab, the *Dagstuhl Seminar on Machine Learning for the Semantic Web*, Dagstuhl, Germany, Feb. 2005
Co-Chair, with Natalya Noy and Alon Halevy, the *First International Workshop on Semantic Integration*, at the Second International Semantic Web Conference (ISWC), Oct. 2003
Electronic Proceedings Chair, the *ACM SIGMOD Conference*, Jun. 2003

Ad-Hoc Committees

Member, the ICDE Ten Years Influential Paper Award Committee, since 2013

Special Issues Edited

Special Issue on Information Extraction, with Luis Gravano, Raghu Ramakrishnan, and Shivakumar Vaithyanathan, *ACM SIGMOD Record*, Fall 2008
Special Issue on Semantic Integration, with Natalya Noy and Alon Halevy, *AI Magazine*, 26(1), Mar. 2005
Special Issue on Semantic Integration, with Natalya Noy and Alon Halevy, *ACM SIGMOD Record*, 33(4), Dec. 2004

Tutorials

Crowdsourcing Platforms and Applications: A Data Management Perspective, with Michael Franklin (Berkeley), Donald Kossmann (ETH-Zurich), and Tim Kraska (Berkeley), *VLDB Conference*, 2011
Managing Information Extraction, with Raghu Ramakrishnan (Yahoo! Research) and Shivakumar Vaithyanathan (IBM Almaden Research), *SIGMOD Conference*, 2006.

Program Committees

The IEEE Int. Conference on Data Engineering (ICDE), 2012
The Int. Conference on Very Large Databases (VLDB), 2011
The Int. Conference on Very Large Databases (VLDB), 2010
The ACM Conference on Data Management (SIGMOD), PC member and group leader, 2009
The Int. Conference on Very Large Databases (VLDB), 2009
The Int. Conference on Very Large Databases (VLDB), 2008
The IEEE Int. Conference on Data Engineering (ICDE), 2008
The Int. Conference on Very Large Databases (VLDB), PC of the Demo. Committee, 2007
The ACM Conference on Data Management (SIGMOD), 2007
The IEEE Int. Conference on Data Engineering (ICDE), 2007
The Int. Joint Conference on AI (IJCAI), 2007
The Int. Workshop on Collaborative Knowledge Management for Web Systems, 2007
The Int. Conference on Knowledge Discovery and Data Mining (KDD), 2006
The National Conference on AI (AAAI), the “AI and the Web” track, 2006
The Int. Conference on Information and Knowledge Management (CIKM), 2006
The 6th Workshop on Next Generation Information Technologies and Systems (NGITS), 2006
The IEEE Int. Conference on Data Engineering (ICDE), 2006
The 13th Int. Conference on Management of Data (COMAD), 2006
The Int. Conference on Very Large Databases (VLDB), 2005
The ACM Conference on Data Management (SIGMOD), 2005
The Int. Joint Conference on AI (IJCAI), 2005
The Int. Conference on Machine Learning (ICML), 2005
The IEEE Int. Conference on Data Mining (ICDM), 2005
The Pacific Conference on Data Mining (PAKDD), 2005
The AAAI Spring Symposium on Knowledge Collection from Volunteer Contributors (KVCV), 2005
The Int. Workshop on Data Integration and Semantic Web, at CAiSE, 2005
The Workshop on Contexts and Ontologies: Theories, Practices, and Applications, at AAAI, 2005
The National Conference on AI (AAAI), 2004
The IEEE Int. Conference on Data Mining (ICDM), 2004
The Int. Semantic Web Conference (ISWC), 2004
The Conference on Information and Knowledge Management (CIKM), 2004
The Workshop on Mining for and from the Semantic Web, at KDD, 2004
The Int. Workshop on Information Integration on the Web (IIWeb), at VLDB, 2004
FLAIRS, Special Track on AI and the Web, 2004
The Conference on Information and Knowledge Management (CIKM), 2003
The Int. Workshop on Information Integration on the Web (IIWeb), at IJCAI, 2003
The Second Int. Human.Society@Internet Conference, 2003

Reviewer for Journals and Handbooks

IEEE Transactions of Knowledge and Data Engineering (TKDE), 2009
Trends and Foundations, 2008
ACM Transaction on Database Systems (TODS), 2007
IEEE Transactions of Knowledge and Data Engineering (TKDE), 2007
Information Systems, 2006
ACM Transaction on Database Systems (TODS), 2006
VLDB Journal (VLDBJ), 2005
IEEE Transactions of Knowledge and Data Engineering (TKDE), 2005
ACM Transactions on Database Systems (TODS), 2004
IEEE Intelligent Systems, 2004
Journal of Web Semantics, 2004
ACM Transactions on Database Systems (TODS), 2003
ACM Transactions on Information Systems (TOIS), 2003
VLDB Journal (VLDBJ), 2003
IEEE Transactions on Knowledge and Data Engineering (TKDE), 2003
Journal of Operations Research, 2003
Handbook on Ontologies, Springer Verlag, 2002
World-Wide Web Journal, 2002
Journal of AI Research (JAIR), 2002
ACM Transactions on Information Systems (TOIS), 2002
INFORM Journal of Computing, 2001

Reviewer for Grants

Panelist, NSF IIS Division, 2013
Panelist, NSF IIS Division, 2009
Panelist, NSF IIS Division, 2005
Panelist, NSF IIS Division, 2003
External reviewer, NSF IIS Division, 2004

PROFESSIONAL MEMBERSHIPS

ACM SIGMOD, Special Interest Group in Data Management, 1998-present
ACM, 1996-present

GRANTS

Relational Data Processing over Distributed and Transactional Key-Value Stores, Google Focus Award, second year. PIs: Jeff Naughton, Jignesh Patel, AnHai Doan, 2014.

Seed Funding for a Crowdsourcing Research Lab at Wisconsin, sponsored research grant from WalmartLabs, 2014.

Seed Funding for a Crowdsourcing Research Lab at Wisconsin, sponsored research grant from WalmartLabs (received June 2013, with planned renewal for subsequent years).

Relational Data Processing over Distributed and Transactional Key-Value Stores, Google Focus Award, first year. PIs: Jeff Naughton, Jignesh Patel, AnHai Doan, Chris Re, 2013.

Crowdsourcing Knowledge Bases, gift money from WalmartLabs, 2012.

EAGER: Integrating Social Media Data, NSF 2-year program, PI: AnHai Doan, 2011.

III: Small: Enabling Technology for Best-Effort Data Integration Systems, NSF 3-year program, PI: AnHai Doan, Co-PI: Jeff Naughton, 2010.

CDI: Collaborative Research: New Knowledge from the Global Lake Ecological Observatory Network (GLEON), NSF 4-year program, PI: Paul Hanson, Co-PIs: AnHai Doan and Miron Livny, 2009.

Machine Reading, DARPA 4-year program, PI: Jude Shavlik, co-PIs: AnHai Doan and David Page.

Exploring Best-Effort Data Integration, Google grant, 2008.

Infrastructure for Text Analytics, Integration, and Search, Microsoft Jim Gray seed grant, 2008.

Infrastructure for Text Analytics, Integration, and Search, IBM Faculty Award, 2008.

Using Technology to Enhance Cancer Communication and Improve Clinical Outcomes, NIH grant, 2008-2012.

Exploring Best-Effort Data Integration, DARPA seedling project, 2007-2008. PI: AnHai Doan, co-PIs: Jeffrey F. Naughton and David J. DeWitt.

Infrastructure for the Development and Testing of Next-Generation, Data-Centric Cluster Management, NSF equipment grant, 2007-2008. PIs: Jeffrey F. Naughton and David J. DeWitt, co-PI: AnHai Doan.

Infrastructure for Text Analytics, Integration, and Search, gift money from IBM Research, 2007-2008. PI: AnHai Doan.

Toward Fast, Approximate, and Collaborative Data Integration, gift money from the Alfred P. Sloan Foundation, 2007-2009. PI: AnHai Doan.

Building Structured Data Portals for Web Communities, gift money from Yahoo! Research, 2006-2007. PI: AnHai Doan.

Multi-Modal Information Integration and Access, Department of Homeland Security grant to establish a 3-year research and development center at UIUC, 2006-2009. Main PI: Dan Roth, Co PIs: AnHai Doan, Kevin Chang, David Forsyth, Jiawei Han, and Chengxiang Zhai.

Evolving and Self-Managing Data Integration Systems, NSF CAREER, 2004-2009.

Privacy-Preserving Data Integration and Sharing, NSF ITR, 2004-2007.

Chris Clifton and AnHai Doan assembled the team, which includes other co-PIs Ahmed K. Elmagarmid (Purdue U), Dan Suciu (U Washington), and Gunther Schadow (Regenstrief Inst. for Health care).

Research on Schema Matching, MITRE Corp., 2003-2004.

PATENTS

A patent on **Mupd8**, a system to process fast streaming data in a map-reduce style. It was used heavily at Kosmix to process tweet streams, and has now been opensourced at <http://github.com/walmartlabs/mupd8>.

SOFTWARE

I have been extensively involved in either building or working on significant components of various software systems in academia and industry.

Mupd8: a system to process fast streaming data in a map-reduce style. It was used heavily at Kosmix and at WalmartLabs to process tweet streams, and has now been opensourced at <http://github.com/walmartlabs/mupd8>. It was published as an industrial paper at VLDB-2012.

KCS: a giant knowledge base of entities and relationships in the world, as integrated from numerous public databases. This knowledge base formed the backbone of the software stack at Kosmix and was used to power the key Kosmix products, and now at WalmartLabs. It was published as an industrial paper at SIGMOD-2013, and I am in the process of trying to open source it.

Doctagger: a system that uses KCS to perform entity extraction, disambiguation, and linking for textual data. It was used heavily at Kosmix and later at WalmartLabs to process social media. It was published as an industrial paper at VLDB-2013, and I am in the process of trying to open source it.

DBLife: a prototype structured knowledge base and portal for the database research community (dblife.cs.wisc.edu). Its data and code have been used for many research projects. The system has helped spawning off PSOX, a sister project on community information management at Yahoo! Research.

INVITED TALKS AND PARTICIPATIONS

Keynote and Invited Talks at Conferences and Workshops

The 16th Int. Workshop on the Web and Databases (WebDB 2013), Jun 2013.

The 21st ACM Conference on Information and Knowledge Management (CIKM), Nov 2012.

The Int. Workshop on Information Integration, at SIGMOD-2012.

The Tenth Pacific Rim International Conference on Artificial Intelligence (PRICAI), Dec 2008.

The Int. ICDE Workshop on Information Integration Methods, Architectures, and Systems (IIMAS), Apr 2008.

The 5th Int. VLDB Workshop on Data Quality, Sep 2007, *Data Quality Challenges in Community Systems*

Academic, Industrial, and Research Labs

Between Jun 2010 and Dec 2013 I gave numerous invited talks at universities, conferences, companies, and government agencies about my research on data integration, social media, and crowdsourcing, as carried out at Wisconsin, Kosmix, and WalmartLabs.

Kosmix Corp., Feb 2010.

Johnson Control Inc., Oct 2009.

Microsoft SQL Server Groups, Jan 2009, Jan 2010.

The International Workshop on Developing Open-Source Data Integration Tools, Jun 2008, invited participation, Google, 2008.

Google Faculty Summit, Jun 2008, invited participation.

Microsoft AdLab, Jun 2008, *Managing Unstructured Data*, research seminar.

University of Waterloo, May 2008, *The Cimple Project on Community Information Management*.

University of Toronto, July 2007, *The Cimple Project on Community Information Management*.

The Northwest Database Society, Seattle, May 2007, *The Cimple Project on Community Information Management*.

IBM Almaden Research Center, Jan. 2007, *The Cimple Project on Community Information Management*.

Stanford University, Jan. 2007, *The Cimple Project on Community Information Management*.

Yahoo! Research, Oct. 2006, *The Cimple Project on Community Information Management*.

University of California, Berkeley, Oct. 2006, *The Cimple Project on Community Information Management*.

University of Washington/Microsoft Research Summer Institute on Infrastructure for Managing Imprecise Information in Relational Database Systems, organized by Surajit Chaudhuri and Dan Suciu, Aug. 2005.

University of Wisconsin-Madison, Dept. of Computer Science, Database Seminar, Apr. 2005. *Toward Fast, Collaborative, and Best-Effort Data Integration*.

CMU/Pitt Joint Database Group Meeting; also University of Pittsburgh, Computer Science, Dept. Seminar, Apr. 2004. *Evolving and Self-Managing Data Integration Systems*.

Information Science Institute (ISI), University of Southern California, Apr. 2004, *Evolving and Self-Managing Data Integration Systems*.

University of Southern California, Apr. 2004, *Schema and Ontology Matching: Current Research Directions*.

University of Maryland, College Park, Apr. 2003, *Autonomic Data Integration*.

University of Pennsylvania, Apr. 2003, *Schema Matching: A Machine Learning Approach*.

University of Illinois, Urbana, Graduate School of Library and Information Sciences, May 2003, *Evolving and Self-Managing Data Integration Systems*.

Purdue University, Nov. 2002, *Autonomic Data Integration*.

IBM Almaden Research, Jul. 2001, *Generic and Extensible Schema Matching with LSD*.

WatchMark Corp., Jul. 2001, *Reconciling Schemas of Disparate Data Sources: A Machine Learning Approach*.

Conferences, Workshops, and Meetings

The Beckman Database Research Self Assessment Meeting, Irvine, Oct 2013. Thirty active database researchers assess the progress and suggest future directions for database research.

SIGMOD-10 Panel, June 2010, *Crowds, clouds, and algorithms: exploring the human side of “big data” applications*, invited panel.

The Claremont Database Research Self Assessment Meeting, Berkeley, May 2008. Twenty

five active database researchers assess the progress and suggest future directions for database research.

The Data-Intensive Computing Symposium, organized by Carnegie-Mellone and Yahoo, Sunnyvale, CA, Apr 2008.

The 3rd DARPA Workshop on Data Integration, Tampa, Jan 2008.

Web 2.0 Panel, VLDB, Sep 2007, *What Can Databases Do for Web 2.0, and Vice Versa?*.

The 2nd DARPA Workshop on Data Integration, Seattle, May 2007.

The DARPA Workshop on Data Integration, (organized by DARPA to explore the possibility of setting up a data integration research program) Nov. 2006, *Recent Advances in Semantic Integration*.

The NSF/EPA/ONR/NARA/AHRQ/NCO Workshop on Data Integration, (sponsored by multiple government agencies to set up the data integration research agenda for the next decade) Oct. 2006, *Best-Effort Data Integration*.

Workshop on Collective Data Integration of Scholarly Databases, organized by Katy Borner, Aug. 2006.

The First Bertinoro Workshop on Data Mining, organized by Alin Dobra, Prabhakar Raghavan, and Raghu Ramakrishnan, Bertinoro, Italy, Oct. 2005.

STUDENTS & POSTDOCTORAL RESEARCHERS

Ph.D. Students (Current)

Adel Ardalan, post-qual

Sanjib Kumar Das, post-qual

Chaitanya Gokhale, post-qual

Paul Suganthan GC, post-qual

Pradap Konda

Ph.D and M.S. Students (Graduated)

Ba-Quy Vuong, Ph.D. (2012, first employment: WalmartLabs)

Xiaoyong Chai, Ph.D. (2012, first employment: Google)

Ting Chen, Ph.D. (2012, first employment: LinkedIn)

Yoonkyong Lee, Ph.D., (2010, first employment: Samsung, South Korea)

Fei Chen, Ph.D., (co-advised with Raghu Ramakrishnan, 2010, first employment: HP Labs, China)

Pedro DeRose, Ph.D., (co-advised with Raghu Ramakrishnan, 2009, first employment: Microsoft)

Warren Shen, Ph.D., (2009, first employment: Google)

Robert McCann Ph.D., (2007, first employment: Researcher, Microsoft Anti-spam Group)

Wensheng Wu Ph.D., (co-advised with Clement Yu, 2006, now postdoc at IBM Almaden)

Doug Burdick Ph.D., (co-advised with Raghu Ramakrishnan, 2007)

Chris Wolf M.S., (2005, now with MITRE Corp.)

Abhishek Tiwari M.S., (2005, continued in the Ph.D. program, UIUC)

Bedoor AlShebi M.S., (2005, continued in the Ph.D. program, UIUC)

Rishi Sinha M.S., (2004, continued in the Ph.D. program, UIUC)

Vanitha Varadarajan M.S., (2004, first employment in Chicago)

Olu Sobulo M.S., (2004, first employment in New York City)
 Robin Dhamanka M.S., (2004, first employment: Microsoft)
 Alex Kramnik M.S., (2003, first employment: Informatica)
 Reza Lesmana M.S., (2003, first employment in Indonesia)
 Arathi Aravi M.S., (2003, first employment in New York City)

Postdoctoral Researchers

Byron J. Gao, 2007-2008

TEACHING EXPERIENCE

Instructor, University of Wisconsin, Madison

CS 784: Advanced Database Management Systems (graduate), 11 students	Fall 2013
CS 784: Advanced Database Management Systems (graduate), 29 students	Spring 2013
CS 564: Database Management Systems (undergraduate), 80 students	Fall 2012
CS 784: Advanced Database Management Systems (graduate), 24 students	Fall 2011
CS 784: Advanced Database Management Systems (graduate), 14 students	Spring 2010
CS 564: Database Management Systems (undergraduate), 35 students	Fall 2009
CS 784: Advanced Database Management Systems (graduate), 20 students	Spring 2009
CS 784: Advanced Database Management Systems (graduate), 21 students	Spring 2008
CS 564: Database Management Systems (undergraduate), 38 students	Fall 2007
CS 764: Topics in Database Management Systems (graduate), 11 students	Spring 2007
CS 564: Database Management Systems (undergraduate), 11 students	Fall 2006

Instructor, University of Illinois, Urbana

CS 411: Database Management Systems, 140+ students	Spring 2006
CS 511: Advanced Topics in Data Management, 18 students	Fall 2005
CS 411: Database Management Systems, 13 students	Summer 2005
CS 411: Database Management Systems, 145 students	Spring 2005
CS 591AD: Seminar on Managing Distributed and Heterogeneous Data, 12 students	Fall 2004
CS 498AD: Machine Learning Approaches to Information Management, 18 students	Fall 2004
CS 491AD: Seminar on Managing Distributed and Heterogeneous Data, 15 students	Spring 2004
CS 311: Database Management Systems, 110 students	Spring 2004
CS 311: Database Management Systems, 130 students	Fall 2003
CS 497AD: Hot Topics in Data Integration, 18 students	Spring 2003
CS 491AD: Seminar on Data Integration, 20 students	Fall 2002

Teaching Assistant, University of Washington, Seattle

CSE 373: Data Structure and Algorithms, 70 students	Fall 1996
---	-----------

DEPARTMENTAL SERVICE

2013: Professional Program Committee (Chair). With the help of Mary Vernon (former department chair), Jeff Naughton (current department chair), and several other colleagues, I conceived, executed, and obtained approval for two new programs: Certificate in CS for Professionals and Professional MS in CS. We are in the process of implementing these programs.

2012: Professional Program Committee (Chair).

2010: Industrial Relations Committee (Co-Chair).
 2009: Industrial Relations Committee (Co-Chair).
 2008: Colloquium Committee (Chair), Industrial Relations Committee (Member).
 2007: Faculty Recruiting Committee, Department of Computer Sciences, UW-Madison (activities include attending the Grace Hopper Conference for recruiting purposes).
 2007: Graduate Student Admission Committee, Department of Computer Sciences, UW-Madison.
 2006: Advisory Committee, Department of Computer Science, UIUC.
 2006: Student Award Committee, Department of Computer Science, UIUC.
 2005: Organized job hunting activities for Ph.D. students. Served on the Student Award Committee.
 2004: Organized job hunting activities for Ph.D. students. The goal is to help create a lasting “student culture” on organizing job hunting activities each year. My services include: organizing the departmental job hunt seminar, organizing a two-hour talk on job hunt for graduating students, compiling a list of graduating students, attending practice job talks and giving advice, helping students organize a “What is Hot in CS” seminar, and creating the job hunt mailing list jobhunt@cs.uiuc.edu
 2004: Participated in the process of recruiting, evaluating the application materials of Vietnamese students, and mentoring them.
 2003: Fun Committee, Department of Computer Science.
 2003: Worked closely with CSGSO (the Computer Science Graduate Student Organization) to organize job hunt seminar for graduating Ph.D. students.
 2002: Created and maintained the departmental job hunt web page for Ph.D. students.
 2002: Fun Committee, Department of Computer Science.

COMMUNITY SERVICE

2001-date: Work on placing Vietnamese students at many U.S. universities (e.g., Berkeley, Illinois, Wisconsin, Maryland, Utah, Arizona). Part of this work is joint with the Vietnam Education Fund (www.vef.gov) and the U.S. Committee for Scientific Cooperation with Vietnam.
 1996-2001: Co-founder of VNSA (with Thanh Nguyen and Son Dam), the largest online forum for Vietnamese professionals and students abroad. To date, VNSA has obtained scholarships for Vietnamese students, found treatments in the U.S. for two leukemia patients from Vietnam, and collected tens of thousands of dollars for charities.

PUBLICATIONS

Publications under Review or in Progress

1. “The Beckman Report on Database Research”, with many authors.
2. “Fast and Accurate Event Detection in the Twittersphere”, with A. Ardalan and J. Patel.
3. “Event Monitoring in the Twittersphere”, with P. Christopher.

4. “Tables, Sets, Lists: Managing Disparate Structured Data Islands in Wikis Using a Unified Approach”, with B. Vuong, X. Chai, and J. Naughton.
5. “Efficient Evaluation of Complex Regular Expressions In Information Extraction”, with Ting Chen and Shivakumar Vaithyanathan, under preparation.

Books

1. “Principles of Data Integration,” textbook, with Alon Halevy and Zachary Ives. Published by Morgan-Kaufmann in 2012, 520 pages.

Special Issues and Proceedings

2. A. Doan, L. Gravano, S. Vaithyanathan, R. Ramakrishnan, “Special Issue on Managing Information Extraction”, *SIGMOD Record*, Winter 08 issue.
3. AnHai Doan and Frank Neven, editors, “Proceedings of the Eighth ACM SIGMOD International Workshop on the Web and Databases (WebDB),” at the ACM SIGMOD Conference, 2005.
4. Fabio Ciravegna, AnHai Doan, Craig Knoblock, Nicholas Kushmerick, and Steffen Staab, editors, “Proceedings of the Dagstuhl Seminar on Machine Learning for the Semantic Web,” 2005.
5. AnHai Doan, Natalya Noy, and Alon Halevy, editors, “Special Issue on Semantic Integration,” *AI Magazine* 26(1), 2005.
6. AnHai Doan, Natalya Noy, and Alon Halevy, editors, “Special Issue on Semantic Integration,” *ACM SIGMOD Record*, 33(4), 2004.
7. AnHai Doan, Natalya Noy, and Alon Halevy, editors, “Proceedings of the First International Workshop on Semantic Integration,” at the Second International Semantic Web Conference (ISWC), 2003.
8. Alon Halevy, Zachary Ives, and AnHai Doan, editors, “Proceedings of the ACM SIGMOD Conference on Data Management,” 2003.

Publications in Journals

9. With many authors, *Social Media Analytics: The Kosmix Story*, IEEE Data Engineering Bulletin, September 2013.
10. A. Doan, R. Ramakrishnan, and A. Halevy, *Crowdsourcing Systems on the World-Wide Web*, Communications of the ACM, 2011. **Invited paper.**
11. A. Doan and many other authors, *The Claremont report on database research*, Communications of the ACM, 52(6):56-65, 2009.
12. A. Doan, J.F. Naughton, R. Ramakrishnan, A. Baid, X. Chai, F. Chen, T. Chen, E. Chu, P. DeRose, B. Gao, C. Gokhale, J. Huang, W. Shen, B. Vuong, “Information Extraction Challenges in Managing Unstructured Data”, *SIGMOD Record*, Winter 08, Special Issue on Managing Information Extraction.

13. AnHai Doan, Philip Bohannon, Raghu Ramakrishnan, Xiaoyong Chai, Pedro DeRose, Byron J. Gao, and Warren Shen, "User-Centric Research Challenges in Community Information Management Systems," *IEEE Data Engineering Bulletin*, Special Issue on Data Management in Social Networks, 2007. **Invited paper.**
14. AnHai Doan, Raghu Ramakrishnan, Fei Chen, Pedro DeRose, Robert McCann, Yoonkyong Lee, Mayssam Sayyadian, and Warren Shen, "Community Information Management," *IEEE Data Engineering Bulletin*, Special Issue on Probabilistic Databases, 29(1), 2006. **Invited paper.**
15. Mayssam Sayyadian, Yoonkyong Lee, AnHai Doan, and Arnon Rosenthal, "eTuner: Tuning Schema Matching Software using Synthetic Scenarios," *VLDB Journal*, *Special Issue of Best Papers of VLDB-05*, 2006. **Invited paper.**
16. AnHai Doan and Alon Halevy, "Semantic Integration Research in the Database Community: A Brief Survey," *AI Magazine*, Special Issue on Semantic Integration, 26(1), Spring 2005.
17. AnHai Doan, Ying Lu, Yoonkyong Lee, and Jiawei Han, "Object Matching for Data Integration: A Profile-Based Approach," *IEEE Intelligent Systems*, Special Issue on Information Integration on the Web, 18(5), pp. 2-7, 2003. **Invited paper.**
18. AnHai Doan, Jayant Madhavan, Robin Dhamankar, Pedro Domingos, and Alon Halevy, "Learning to Match Ontologies on the Semantic Web," *VLDB Journal*, 12(4), pp. 303-319, 2003.
19. AnHai Doan, Pedro Domingos, and Alon Halevy, "Learning to Match the Schemas of Databases: A Multistrategy Approach," *Machine Learning*, 50(3), pp. 279-301, 2003.
20. Vu Ha, AnHai Doan, Van Vu, and Peter Haddawy, "Geometric Foundations for Interval-Based Probabilities," *Annals of Mathematics and Artificial Intelligence*, 24(1-4), pp. 1-21, 1998.
21. Peter Haddawy, AnHai Doan, and Charles Kahn, "Decision-Theoretic Refinement Planning in Medical Decision Making: Management of Acute Deep Venous Thrombosis," *Journal of Medical Decision Making*, 16(4), pp. 315-325, 1996.

Publications in Rigorously Reviewed Conferences

22. C. Gokhale, S. Das, A. Doan, J. Naughton, R. Rampalli, J. Shavlik, and J. Zhu, "Corleone: Hands-Off Crowdsourcing for Entity Matching", *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD-14)*, 2014.
23. Y. Chiang, A. Doan, J. Naughton, "Modeling Entity Evolution for Temporal Record Matching", *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD-14)*, 2014.
24. Y. Chiang, A. Doan, J. Naughton, "Tracking Entities in the Dynamic World: a Fast Algorithm for Matching Temporal Records", *Proceedings of the Conference on Very Large Databases (VLDB-14)*, 2014.

25. A. Gattani, D. Lamba, N. Garera, M. Tiwari, X. Chai, S. Das, S. Subramaniam, A. Rajaraman, V. Harinarayan, and A. Doan, "Entity Extraction, Linking, Classification, and Tagging for Social Media: a Wikipedia-Based Approach", industrial paper, *Proceedings of the Conference on Very Large Databases (VLDB-13)*, 2013.
26. O. Deshpande, D. Lamba, M. Tourn, S. Das, S. Subramaniam, A. Rajaraman, V. Harinarayan, and A. Doan, "Building, Maintaining, and Using Knowledge Bases: A Report from the Trenches", industrial paper, *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD-13)*, 2013.
27. W. Lam, L. Liu, STS Prasad, A. Rajaraman, Z. Vacheri, and A. Doan, "Muppet: MapReduce-Style Processing of Fast Data", industrial paper, *Proceedings of the Conference on Very Large Databases (VLDB-12)*, 2012.
28. F. Niu, C. Re, A. Doan, and J. Shavlik, "Scaling Up Statistical Inference in Markov Logic Networks using RDBMSs", *Proceedings of the Conference on Very Large Databases (VLDB-11)*, 2011.
29. A. Baid, I. Rae, J. Li, A. Doan, J. Naughton, "Toward Scalable Keyword Search over Relational Data", *Proceedings of the Conference on Very Large Databases (VLDB-10)*, 2010.
30. A. Baid, I. Rae, A. Doan, J. Naughton, "Toward Industrial-Strength Keyword Search over Relational Databases", *Proceedings of the IEEE Int. Conference on Data Engineering (ICDE-10)*, 2010. Short paper. 20% accepted (for both long and short papers).
31. F. Chen, B. Gao, A. Doan, J. Yang, R. Ramakrishnan, "Optimizing complex extraction programs over evolving text data", *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD-09)*, 2009. 63/397 = 16% accepted.
32. E. Chu, A. Baid, X. Chai, A. Doan, and J. Naughton, "Combining keyword search and forms for ad hoc querying of databases", *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD-09)*, 2009. 63/397 = 16% accepted.
33. X. Chai, B. Vuong, A. Doan, and J. Naughton, "Efficiently incorporating user feedback into information extraction and integration programs", *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD-09)*, 2009. 63/397 = 16% accepted.
34. A. Doan, J. F. Naughton, A. Baid, X. Chai, F. Chen, T. Chen, E. Chu, P. DeRose, B. Gao, C. Gokhale, J. Huang, W. Shen, B. Vuong, "The Case for a Structured Approach to Managing Unstructured Data", *Proceedings of the Biannual Conference on Innovative Database Research (CIDR-09)*, 2009. Refereed essay. Unknown acceptance rate.
35. A. Jain, P. G. Ipeirotis, A. Doan, L. Gravano, "Join Optimization of Information Extraction Output: Quality Matters!", *Proceedings of the IEEE Int. Conference on Data Engineering (ICDE-09)*, 2009. Full paper. 93/554 = 16.8% accepted.
36. R. Thonangi, H. He, A. Doan, H. Wang, J. Yang, "Weighted Proximity Best-Joins for Information Retrieval", *Proceedings of the IEEE Int. Conference on Data Engineering (ICDE-09)*, 2009. Full paper. 93/554 = 16.8% accepted.

37. Warren Shen, Robert McCann, Pedro DeRose, AnHai Doan, and Raghu Ramakrishnan, "Toward Best-Effort Data Integration", *Proceedings of the ACM Conference on Data Management (SIGMOD-08)*, 2008, 78/435 = 17.9% accepted.
38. Jiansheng Huang, Ting Chen, AnHai Doan, Jeffrey F. Naughton, "On the provenance of non-answers to queries over extracted data", *Proceedings of the Conference on Very Large Databases (VLDB-08)*, 2008. 49/296 = 16.5% accepted.
39. Xiaoyong Chai, Mayssam Sayyadian, AnHai Doan, Arnon Rosenthal, Len Seligman, "Analyzing and revising data integration schemas to improve their matchability", *Proceedings of the Conference on Very Large Databases (VLDB-08)*, 2008. 49/296 = 16.5% accepted.
40. Pedro DeRose, Xiaoyong Chai, Byron Gao, Warren Shen, AnHai Doan, Philip Bohannon, and Jerry Zhu, "Building Community Wikipedias: A Human-Machine Approach," *Proceedings of the IEEE Int. Conference on Data Engineering (ICDE-08)*, 2008. 75/617 = 12.1% accepted.
41. Fei Chen, AnHai Doan, Jun Yang, and Raghu Ramakrishnan, "Efficient Information Extraction over Evolving Text Data," *Proceedings of the IEEE Int. Conference on Data Engineering (ICDE-08)*, 2008. 75/617 = 12.1% accepted.
42. Alpa Jain, AnHai Doan, and Luis Gravano, "Optimizing SQL Queries over Text Databases," *Proceedings of the IEEE Int. Conference on Data Engineering (ICDE-08)*, 2008. 75/617 = 12.1% accepted.
43. Robert McCann, Warren Shen, Alex Kramnik, and AnHai Doan, "Web 2.0 Style Schema Matching," *Proceedings of the IEEE Int. Conference on Data Engineering (ICDE-08)*, 2008. 119/617 = 19.2% accepted.
44. Warren Shen, AnHai Doan, Jeffrey Naughton, and Raghu Ramakrishnan, "Declarative Information Extraction using Datalog with Embedded Extraction Predicates," *Proceedings of the Int. Conference on Very Large Databases (VLDB-07)*, 2007. 45/276 = 16.3% accepted.
45. Douglas Burdick, AnHai Doan, Shivakumar Vaithyanathan, and Raghu Ramakrishnan, "OLAP over Imprecise Data with Constraints," *Proceedings of the Int. Conference on Very Large Databases (VLDB-07)*, 2007. 45/276 = 16.3% accepted.
46. Pedro DeRose, Warren Shen, Fei Chen, AnHai Doan, and Raghu Ramakrishnan, "Building Structured Web Community Portals: The Case for an Incremental and Compositional Approach," *Proceedings of the Int. Conference on Very Large Databases (VLDB-07)*, 2007. 45/276 = 16.3% accepted.
47. Eric Chu, Akanksha Baid, Ting Chen, AnHai Doan, and Jeff Naughton, "A Relational Approach to Incrementally Extracting and Querying Structure in Unstructured Data," *Proceedings of the Int. Conference on Very Large Databases (VLDB-07)*, 2007. 45/276 = 16.3% accepted.
48. Mayssam Sayyadian, AnHai Doan, and Luis Gravano, "Efficient Keyword Search across Heterogeneous Relational Databases," *Proceedings of the IEEE Int. Conference on Data Engineering (ICDE-07)*, 2007. 122/659 = 18% accepted.

49. Warren Shen, Pedro DeRose, Long Vu, AnHai Doan, and Raghu Ramakrishnan, "Source-aware Entity Matching: A Compositional Approach," *Proceedings of the IEEE Int. Conference on Data Engineering (ICDE-07)*, 2007. 122/659 = 18% accepted.
50. Alpa Jain, AnHai Doan, and Luis Gravano, "SQL Queries over Text Databases", *Proceedings of the IEEE Int. Conference on Data Engineering (ICDE-07)*, 2007. 182/659 = 28% accepted.
51. Wensheng Wu, AnHai Doan, and Clement Yu, "Learning from the Web to Match Query Interfaces on the Deep Web," *Proceedings of the IEEE Int. Conference on Data Engineering (ICDE-06)*, 2006. 89/456 = 20% accepted.
52. Mayssam Sayyadian, Yoonkyong Lee, AnHai Doan, and Arnon Rosenthal, "eTuner: Tuning Schema Matching Software using Synthetic Scenarios," *Proceedings of the Int. Conference on Very Large Databases (VLDB-05)*, 2005. 85/517 = 16% accepted.
53. Robert McCann, Bedoor AlShebli, Quoc Le, Hoa Nguyen, Long Vu, and AnHai Doan, "Maveric: Mapping Maintenance in Data Integration Systems," *Proceedings of the Int. Conference on Very Large Databases (VLDB-05)*, 2005. 85/517 = 16% accepted.
54. Warren Shen, Xin Li, and AnHai Doan, "Constraint-Based Entity Matching," *Proceedings of the Nat. Conference on AI (AAAI-05)*, 2005. 148/803 = 18% accepted.
55. Robert McCann, Alexander Kramnik, Warren Shen, Vanitha Varadarajan, Olu Sobulo, and AnHai Doan, "Integrating Data from Disparate Sources: A Mass Collaboration Approach," *Proceedings of the IEEE Int. Conference on Data Engineering (ICDE-05)*, 2005. 100/521 = 19% accepted.
56. Jayant Madhavan, Philip A. Bernstein, AnHai Doan, and Alon Y. Halevy, "Corpus-based Schema Matching," *Proceedings of the IEEE Int. Conference on Data Engineering (ICDE-05)*, 2005. 67/521 = 13% accepted.
57. Wensheng Wu, AnHai Doan, and Clement Yu, "Merging Interface Schemas on the Deep Web via Clustering Aggregation," *Proceedings of the IEEE Int. Conference on Data Mining (ICDM-05)*, 2005. 141/630 = 22% accepted.
58. Wensheng Wu, Clement Yu, AnHai Doan, and Weiyi Meng, "An Interactive Clustering-Based Approach to Integrating Query Interfaces on the Deep Web," *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD-04)*, pp. 95-106, 2004. 69/431 = 16% accepted.
59. Robin Dhamankar, Yoonkyong Lee, AnHai Doan, Alon Halevy, and Pedro Domingos, "iMAP: Discovering Complex Semantic Matches between Database Schemas," *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD-04)*, pp. 383-394, 2004. 69/431 = 16% accepted.
60. Oren Etzioni, Alon Halevy, AnHai Doan, Zack Ives, Jayant Madhavan, Luke McDowell, and Igor Tatarinov, "Crossing the Structure Chasm," *Proceedings of Conf. on Innovative Database Research (CIDR-03)*, pp. 121-132, 2003. 26/87 = 30% accepted.

61. AnHai Doan, Jayant Madhavan, Pedro Domingos, and Alon Halevy, "Learning to Map between Ontologies on the Semantic Web," *Proceedings of the World-Wide Web Conference (WWW-02)*, pp. 662-673, 2002. 72/454 = 16% accepted.
62. AnHai Doan and Alon Halevy, "Efficiently Ordering Query Plans for Data Integration," *Proceedings of the 18th IEEE International Conference on Data Engineering (ICDE-02)*, pp. 393-404, 2002. 54/287 = 19% accepted.
63. AnHai Doan, Pedro Domingos, and Alon Halevy, "Reconciling Schemas of Disparate Data Sources: A Machine Learning Approach," *Proceedings of the ACM SIGMOD Conf. on Management of Data (SIGMOD-01)*, pp. 509-520, 2001. 45/293 = 15% accepted.
64. AnHai Doan and Peter Haddawy, "Sound Abstraction of Probabilistic Actions in the Constraint Mass Assignment Framework," *Proceedings of the 12th National Conference on Uncertainty in AI (UAI-96)*, pp. 228-235, 1996.
65. AnHai Doan, "Modeling Probabilistic Actions for Practical Decision-Theoretic Planning," *Proceedings of the 3rd International Conference on AI Planning Systems (AIPS-96)*, pp. 62-69, 1996, top conference in AI planning.
66. Peter Haddawy, AnHai Doan, and Richard Goodwin, "Efficient Decision-Theoretic Planning: Techniques and Empirical Analysis," *Proceedings of the 11th National Conference on Uncertainty in AI (UAI-95)*, pp. 229-236, 1995.
67. AnHai Doan, Peter Haddawy, and C. Kahn, "Decision-Theoretic Refinement Planning: A New Method for Clinical Decision Analysis," *Proceedings of the 19th Annual Symposium on Computer Applications in Medical Care (SCAMC-95)*, pp. 299-303, 1995.
68. Peter Haddawy and AnHai Doan, "Abstracting Probabilistic Actions," *Proceedings of the 10th National Conference on Uncertainty in AI (UAI-94)*, pp. 265-272, 1994.

Other Publications

69. Michael Wick, Khashayar Rohanimanesh, Andrew McCallum, AnHai Doan, "A Discriminative Approach to Ontology Mapping", *Proceedings of the International Workshop on New Trends in Information Integration (NTII-08)*, 2008. Unknown acceptance rate.
70. Rakesh Agrawal, Anastasia Ailamaki, Philip A. Bernstein, Eric A. Brewer, Michael J. Carey, Surajit Chaudhuri, AnHai Doan, Daniela Florescu, Michael J. Franklin, Hector Garcia-Molina, Johannes Gehrke, Le Gruenwald, Laura M. Haas, Alon Y. Halevy, Joseph M. Hellerstein, Yannis E. Ioannidis, Henry F. Korth, Donald Kossmann, Samuel Madden, Roger Magoulas, Beng Chin Ooi, Tim O'Reilly, Raghu Ramakrishnan, Sunita Sarawagi, Michael Stonebraker, Alexander S. Szalay, Gerhard Weikum, "The Claremont report on database research", *SIGMOD Record 37(3): 9-19 (2008)*. Unknown acceptance rate.
71. Pedro DeRose, Warren Shen, Fei Chen, Yoonkyong Lee, Doug Burdick, AnHai Doan, and Raghu Ramakrishnan, "DBLife: A Community Information Management Platform for the Database Research Community," *Proceedings of the Conference on Innovative Database Research (CIDR-07)*, (demo) 2007.

72. AnHai Doan, "Best-Effort Data Integration," *The NSF/EPA/ONR/NARA/AHRQ/NCO Workshop on Data Integration*, (position statement) 2006.
73. AnHai Doan, Raghu Ramakrishnan, and Shivakumar Vaithyanathan, "Managing Information Extraction," (tutorial description), *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD-06)*, 2006.
74. Robert McCann, Warren Shen, and AnHai Doan, "Collaborative Development of Information Integration Systems," *Proceedings of the AAAI Spring Symposium on Knowledge Collection from Volunteer Contributors (KVC-05)*, 2005.
75. Wensheng Wu, AnHai Doan, Clement Yu, and Weiyi Meng, "Bootstrapping Domain Ontology for Semantic Web Services from Source Web Sites," *Proceedings of the VLDB-05 Workshop on Technologies for E-Services*, 2005.
76. Chris Clifton, Murat Kantarcioglu, AnHai Doan, Gunther Schadow, Jaideep Vaidya, Ahmed K. Elmagarmid, and Dan Suciu, "Privacy Preserving Data Integration and Sharing," *Proceedings of the 9th Int ACM SIGMOD Workshop on Data Mining and Knowledge Discovery (DMKD-04)*, pp. 19-26, 2004. $8/34 = 24\%$ accepted.
77. Robert McCann, Warren Shen, and AnHai Doan, "Collective Integration of Information for Virtual Organization," *Proceedings of the ACM SIGMOD Workshop on Databases in Virtual Organizations (DIVO-04)*, pp. 9-16, 2004.
78. AnHai Doan, Jayant Madhavan, Pedro Domingos, and Alon Halevy, "Ontology Matching: A Machine Learning Approach," *Handbook on Ontologies*, edited by Steffen Staab & Rudi Studer, Springer-Verlag, invited paper, pp. 397-416, 2004.
79. Hwanjo Yu, AnHai Doan, and Jiawei Han, "Mining for Information Discovery on the Web: Overview and Illustrative Research," *Intelligent Technologies for Information Analysis*, edited by Ning Zhong, Springer-Verlag, invited paper, pp. 135-168, 2004.
80. Mayssam Sayyadian, Azadeh Shakery, AnHai Doan, and ChengXiang Zhai, "Toward Entity Retrieval over Structured and Text Data," *the ACM SIGIR Workshop on the Integration of IR and Databases (WIRD-04)*, pp. 47-54, 2004.
81. AnHai Doan, Natalya Noy, and Alon Halevy, "Introduction to the Special Issue on Semantic Integration," *ACM SIGMOD Record*, 33(4), pp. 11-14, 2004.
82. AnHai Doan, Natalya Noy, and Alon Halevy, "Report on the Semantic Integration Workshop at the 2nd Int. Semantic Web Conf. (ISWC-03)," *ACM SIGMOD Record*, 33(1), pp. 138-140, 2004.
83. AnHai Doan, Natalya Noy, and Alon Halevy, "Report on the Semantic Integration Workshop at the Second International Semantic Web Conference (ISWC-03)," *AI Magazine*, 2004
84. Robert McCann, AnHai Doan, Vanitha Varadarajan, and Alex Kramnik, "Building Data Integration Systems: A Mass Collaboration Approach," *Proceedings of the International ACM SIGMOD Workshop on the Web and Databases (WebDB-03)*, pp. 25-30, 2003. $17/74 = 23\%$ accepted.

85. AnHai Doan, Ying Lu, Yoonkyong Lee, and Jiawei Han, "Object Matching for Data Integration: A Profile-Based Approach," *Proceedings of the IJCAI Workshop on Information Integration on the Web (IIWeb-03)*, 2003.
86. AnHai Doan and Robert McCann, "Building Data Integration Systems: A Mass Collaboration Approach," *Proceedings of the IJCAI Workshop on Information Integration on the Web (IIWeb-03)*, 2003.
87. Pedro Domingos, Yeuhi Abe, Corin Anderson, AnHai Doan, and 13 other co-authors, "Research on Statistical Relational Learning at the University of Washington," *Proc. of the IJCAI-03 Workshop on Learning Statistical Models from Relational Data*, 2003.
88. Marianne Winslett, Kevin Chang, AnHai Doan, Jiawei Han, ChengXiang Zhai, and Yuanyuan Zhou, "Database Research at UIUC," *ACM SIGMOD Record*, 31(3), pp. 97-102, 2002.
89. AnHai Doan, "Learning to Map between Structured Representations of Data," *Ph.D. Dissertation*, Department of Computer Science, Univ. of Washington, 2002.
90. AnHai Doan, Pedro Domingos, and Alon Levy, "Learning Source Descriptions for Data Integration," *Proceedings of the 3rd International ACM SIGMOD Workshop on the Web and Databases (WebDB-2000)*, pp. 81 - 86, 2000. 20/69 = 29% accepted.
91. AnHai Doan, Pedro Domingos, and Alon Levy, "Data Integration: A "Killer App" for Multi-Strategy Learning," *Proceedings of the ECAI Workshop on Multi-Strategy Learning (MSL-00)*, 2000.
92. AnHai Doan, Pedro Domingos, and Alon Levy, "Learning Mappings between Data Schemas," *Proceedings of the AAAI-2000 Workshop on Learning Statistical Models from Relational Data*, 2000.
93. AnHai Doan and Alon Levy, "Efficiently Ordering Query Plans for Data Integration," *The IJCAI-99 Workshop on Intelligent Information Integration*, 1999.
94. AnHai Doan, Peter Haddawy, and Charles Kahn, "Decision-Theoretic Planning for Clinical Decision Analysis," *The Working Papers of AI in Medicine Spring Symposium*, Stanford, 1996.
95. Charles Kahn, AnHai Doan, and Peter Haddawy, "Management of Acute Deep Venous Thrombosis of the Lower Extremities (abstract)," *American Roentgen Ray Society Meeting*, 1996.
96. AnHai Doan, "An Abstraction-Based Approach to Decision-Theoretic Planning," *Masters Thesis*, Department of Computer Science, University of Wisconsin, 1995.
97. AnHai Doan and Peter Haddawy, "Generating Macro Operators for Decision-Theoretic Planning," *AAAI Spring Symposium on Extended Theories of Action Representation*, Stanford, 1995.