

Pei Li

Department of Informatics, Systems and Communication
University of Milan – Bicocca (Unimib)
Viale Sarca 336/14, Milan, Italy 20126

Office: +39 (0)2 6448 7921
Mobile: +39 331 9450 345
pei.li@disco.unimib.it

Research Interests

Databases, with an emphasis on data quality and data integration, particularly interested in record linkage, also interested in large-scale data analytics and cloud computing.

Education

- Ph.D** Computer Science, University of Milan – Bicocca Expected in March, 2013
Dissertation: Linking Records with Value Diversity
Advisors: Professor Andrea Maurino & Dr. Xin Luna Dong
Ranked first in entrance exam
- M.S.** Computer Science, Beijing University of Posts and Telecommunications (BUPT) 2008
Dissertation: Design and Application of a Resource Semantic Model in Modern Service Systems
Advisor: Professor Meina Song & Prof. Junde Song
Ranked first in national entrance exam
- B.S.** Electronic Engineering, Beijing University of Posts and Telecommunications 2005
Top-1 out of 30 students

Scholarships & Awards

- ACM SIGMOD student travel award 2010, 2012
National Scholarship for Ph.D Students, Italy 2009-2011
Exchange Student Program Scholarship, University of Milan – Bicocca 2010
National Scholarship for Graduate Students, China (three times) 2005-2008
University-Level Scholarship at BUPT (three times) 2001-2004

Research & Work Experience

Research Assistant, University of Milan – Bicocca

2009 – present

Advisors: Prof. Andrea Maurino & Dr. Xin Luna Dong (external)

- Researching on record linkage, a critical problem in many data integration and data cleaning applications. Developed novel models to link records where tolerance to fairly high value variety is required, including linking records with contextual information stored in entity-relationships, linking temporal records that describe the same entity over time, and linking records that belong to the same group.
- Developed Chronos, a system that offers a useful tool for finding real-world entities over time and understanding history of entities in bibliography domain.
- Developed group linkage tool for identifying business listings that belong to the same business chains for AT&T Interactive.
- Participated in developing SAMEDUDE, a system for distinguishing inventors in European Patent database, which exploits structured relationships within the database and applied crowdsourcing approach to estimate accuracy of the results.
- Researched on quality evaluation of e-Government project planning that initializes design and development of e-services, to study the impacts of principles, policies and laws on information system evaluation.

Research Intern, AT&T Labs – Research

2010

Advisors: Dr. Xin Luna Dong & Dr. Divesh Srivastava

- Developed, implemented and experimented on a model for linking temporal records that describe the same entity over time.

Research Assistant, University of Milan – Bicocca

2008 – 2009

Advisors: Prof. Andrea Maurino & Prof. Flavio De Paoli

- Developed a probabilistic model that uses Constraint Satisfaction techniques to effectively select the optimal web-service composition considering quality of services.

Research Assistant, Beijing University of Posts & Telecommunications

2005 – 2008

Advisors: Prof. Meina Song & Prof. Junde Song

- Participated in National Key Project of the 11th Five-Period Scientific and Technical Supporting Programs in China, as a major contributor to national ratified specification – Service Access and Assembly Standard Version 0.93, which specifies the service discovery and operation restriction protocol for service composition in e-businesses.
- Developed an operation support system for an application provider that supports Mobile Instant Messenger.
- Designed and implemented a call queuing component for a private IT company that supports Computer-Telephony-Integration-based Call Center solutions.

Intern, InfowareLab Inc., Hangzhou China

2006

Designed and developed an operation support system for a service search engine operator that supports keyword-based service search via Internet and VoIP.

Teaching & Mentoring Experience

Mentor, University of Milan, Bicocca Spring, 2012
Haidong Wang, graduate student from Nankai University. Supervised development of the interface of Chronos, where the goal is to facilitate history discovery by linking temporal records.

Teaching Assistant, Beijing University of Posts & Telecommunications Spring, 2006
Basis of Electronic Circuit, 90 students. Graded assignments, helped students in office hours, assisted in preparing and grading exams.

Invited Talks

Linking Records in a Dynamic World

Graduate School of Library and Information Science, UIUC, Champaign IL USA, June 2012
Swiss Federal Institutes of Technology and Research Zurich (ETHZ), Switzerland, August 2012

Linking Temporal Records

International School on Mathematics “Guido Stampacchia” Workshop, Erice, Italy, Sept. 2011
International workshop on Big-Data Analytics for the Temporal Web, Paris, France, Nov. 2011

Professional Activities & Services

Journal Reviewer: Information System 2012 (special issue on data quality), Journal of Data and Information Quality (February, 2011)

Conference External Reviewer: ICSOC 2009, ICIQ 2009, SSDBM 2012

Journal External Reviewer: IEEE Internet Computing (May, 2009)

Workshop Committee Members: QDB (2010, 2011), WEWST (2010-2012)

Publications

Book Chapter:

1. **Pei Li** and Andrea Maurino: Record Linkage in Complex Contexts. In “Handbook on Research and Practice in Data Quality”. To be published by *Springer*.

Papers in Refereed Conferences:

2. **Pei Li**, Xin Luna Dong, Andrea Maurino and Divesh Srivastava: Linking Temporal Records. *In Int. Conf. on Very Large Data Bases (VLDB)*, pages 956–967, 2011.
3. **Pei Li**, Marco Comerio, Andrea Maurino and Flavio De Paoli: Advanced Non-Functional Property Evaluation of Web Services. *In IEEE European Conference on Web Services (ECOWS)*, pages 27-36, 2009.
4. **Pei Li** and Andrea Maurino: Schema Based Deduplication. *In Int. Conf. on Information Quality (ICIQ)*, page 262-264, 2009.

Invited Journal Paper:

5. **Pei Li**, Xin Luna Dong, Andrea Maurino and Divesh Srivastava: Linking Temporal Records. *Frontiers of Computer Science* 6(3): 293-312, 2012.

Conference Demo:

6. **Pei Li**, Haidong Wang, Christina Tziviskou, Xin Luna Dong, Xiaoguang Liu, Andrea Maurino and Divesh Srivastava: Chronos: Facilitating History Discovery by Linking Temporal Records. *In Int. Conf. on Very Large Data Bases (VLDB)*, 2012.

Papers in Refereed Workshops:

7. **Pei Li**: Linking Records in Dynamic World. *In SIGMOD/PODS PhD Symposium*, pages 51-56, 2012.
8. **Pei Li**: Schema Based Deduplication. *In SIGMOD PhD Workshop on Innovative Database Research*, pages 25-30, 2010.
9. **Pei Li** and Gianluigi Viscusi: Quality Evaluation for Strategic Alignment Engineering: An eGovernment Application. *In Int. Workshop on Business/IT-Alignment and Interoperability*, pages 1-15, 2010.
10. **Pei Li** and Andrea Maurino: Schema Based Deduplication. *In International Workshop on Quality in Databases (QDB)*, 2009.

Others:

11. **Pei Li**, Xin Luna Dong, Songtao Guo, Andrea Maurino and Divesh Srivastava: Linking Group Records. *Technical Report*, University of Milan – Bicocca, 2012.
12. **Pei Li**, Qian Wang and Meina Song: Two Ways of Calling Web Services with Ajax in Service Oriented Architecture, *Future Telecommunications Conference*, 2007.

References

Prof. **Andrea Maurino** (advisor)
Department of Informatics, Systems and Communication
University of Milan – Bicocca
Viale Sarca 336/14, Milan, Italy 20126
+39 (2) 6448 7897
maurino@disco.unimib.it

Xin Luna Dong (supervisor)
Data Management Dept
AT&T Labs--Research
Florham Park, NJ 07932
+1 (973)360-8508
lunadong@research.att.com

Divesh Srivastava (supervisor)
Data Management Dept
AT&T Labs--Research
Florham Park, NJ 07932
+1 (973) 360-8776
divesh@research.att.com