Julia Stoyanovich, PhD

• Computer and Information Sciences Department • University of Pennsylvania • 3330 Walnut Street •
• Philadelphia, PA 19104 • Mobile: 1 (917) 470-8199 • Email: jstoy@cis.upenn.edu •

Research Interests My research is motivated by the *data management* needs of life sciences applications and of social information processing. I am working on incorporating semantic context into search, ranking, and data exploration in large complex datasets. I am also interested in managing provenance in scientific workflows, and in the related *privacy and security* considerations.

Affiliation

	Department of Computer and Information Science University of Pennsylvania, Philadelphia, PA, USA
11/2011 – currently	Visiting Scholar
11/2009 –10/2011	Postdoctoral Researcher, Computing Innovations Fellow

Education

PhD in Computer Science
Columbia University
09/2004 – 10/2009

- ✓ Thesis "Search and Ranking in Semantically Rich Applications."
- ✓ Advised by Professor Kenneth A. Ross

MS in Computer Science
Columbia University
09/2003 – 05/2004

- ✓ Fully funded through a research assistantship
- ✓ GPA: 4.0/4.0

BS in Computer Science
BS in Mathematics and Statistics
University of Massachusetts Amherst
09/1995 – 08/1998

- ✓ Graduated magna cum laude, GPA 3.73/4.0
- ✓ Phi Kappa Phi Honor Society
- ✓ Recipient of the 1997 Du Bois Research Scholarship

Publications

1. "Viewing the Web as a Distributed Knowledge Base", S. Abiteboul, E. Antoine, J. Stoyanovich, *ICDE 2012* (invited talk by S. Abiteboul).
2. "Putting Lipstick on Pig: Enabling Database-style Workflow Provenance", Y. Amsterdamer, S.B. Davidson, D. Deutch, T. Milo, J. Stoyanovich, V. Tannen, *PVLDB 5(4), 2011*.
3. "Deriving Probabilistic Databases with Inference Ensembles", J. Stoyanovich, S.B. Davidson, T. Milo, V. Tannen, *ICDE 2011*(19.8% accepted).
4. "Making Interval-based Clustering Rank-aware", J. Stoyanovich, S. Amer-Yahia, T. Milo, *EDBT 2011* (27.7% accepted).
5. "Keyword Search in Workflow Repositories with Access Control", S.B. Davidson, S.M. Lee, J. Stoyanovich, *AMW 2011*.
6. "On Provenance and Privacy", S.B. Davidson, S. Khanna, S. Roy, J. Stoyanovich, V. Tannen, Y. Chen, *ICDT 2011* (invited talk by S.B. Davidson).
7. "Enabling Privacy in Provenance-aware Workflow Systems", S.B. Davidson, S. Khanna, S. Roy, J. Stoyanovich, V. Tannen, Y. Chen, T. Milo, *CIDR 2011*.
8. "*AnnotCompute*: Annotation-Based Exploration and Meta-analysis of Biological Experiments", J. Zheng, J. Stoyanovich, J. Liu, E. Manduchi, C.J. Stoeckert, Jr., *The Journal of Biological Databases and Curation, 2011.*
9. *"SkylineSearch*: Semantic Ranking and Result Visualization for PubMed", J. Stoyanovich, M. Lodha, W. Mee, K.A. Ross, *SIGMOD 2011* (demonstration).

10. "Semantic Ranking and Result Visualization for Life Sciences Publications", J. Stoyanovich, W. Mee, and K.A. Ross, *ICDE 2010* (12.5% accepted).
11. "Exploring Repositories of Scientific Workflows", J. Stoyanovich, B. Taskar, S. Davidson, *WANDS 2010*.
12. "Rank-Aware Clustering for Structured Datasets", J. Stoyanovich, S. Amer-Yahia, *CIKM 2009* (short paper).
13. "Efficient Network-Aware Search in Collaborative Tagging Sites", S. Amer-Yahia, M. Benedikt, L. Lakshmanan, J. Stoyanovich, *PVLDB(1), 2008* (16.5% accepted).
14. "Schema Polynomials and Applications", K.A. Ross, J. Stoyanovich, *EDBT 2008* (16.7% accepted).
15. "From del.icio.us to x.qui.site: Recommendations in Social Tagging Sites", S. Amer-Yahia, A. Galland, J. Stoyanovich, C. Yu, *SIGMOD 2008* (demonstration).
16. "Leveraging Tagging to Model User Interests in del.icio.us", J. Stoyanovich, S. Amer-Yahia, C. Marlow, C. Yu. *AAAI Spring Symposium on Social Information Processing (AAAI-SIP) 2008*.
17. "EntityAuthority: Semantically-Enriched Graph Based Authority Propagation", J. Stoyanovich, S. Bedathur, K. Berberich, G. Weikum, *WebDB 2007* (25% accepted).
18. "MutaGeneSys: Making Diagnostic Predictions Based on Genome-Wide Association Data in Genotype Studies", J. Stoyanovich, I. Pe'er, *Bioinformatics, 12/2007* (5.039 impact factor).
19. "A Faceted Query Engine Applied to Archaeology", K.A. Ross, A. Janevski, J. Stoyanovich, *Internet Archaeology, 04/2007*.
20. "A Faceted Query Engine Applied to Archaeology", K.A. Ross, A. Janevski, J. Stoyanovich, *VLDB 2005* (demonstration).
21. "Symmetric Relations and Cardinality Bounded Multisets in Database Systems", K.A. Ross and J. Stoyanovich, *VLDB 2004* (16.1% accepted).

Submitted Publications and Working Papers

22. "Learning to Explore Scientific Workflow Repositories", P. Dhillon, J. Stoyanovich, B. Lyons, S.B. Davidson.
23. "Identifying Ranked Agreement in Large Structured Datasets", J. Stoyanovich, S. Amer-Yahia, T. Milo.

Patents

1. "Automatically and Adaptively Determining Execution Plans for Queries with Parameter Markers", W. Fan, G. Lohman, V. Markl, N. Megiddo, J. Rao, D. Simmen, J. Stoyanovich. US Patent 7,958,113. Assignee: IBM.
2. "Social Behavior Analysis and Inferring Social Networks for a Recommendation System", S. Amer-Yahia, E. Gabrilovich, B. Pang, J. Stoyanovich, C. Yu. US Patent 8,073,794 (Dec 6, 2011). Assignee: Yahoo! Inc.

Funding

1. NSF/CRA Computing Innovations Fellow: Data Exploration in Biological Repositories (2009-2011).
2. Google Research Award: Identifying ranked agreement among raters (2012).

Research Experience

11/2011 – currently Visiting Scholar
11/2009 – 10/2011 Postdoctoral Researcher / Computing Innovations Fellow
Department of Computer and Information Science
University of Pennsylvania, Philadelphia, PA, USA
Supervised by Professor Susan B. Davidson

- ✓ Scientific workflows and provenance
  - Exploring repositories of scientific workflows
  - Managing and querying workflow provenance
  - Security and privacy
- ✓ Search and ranking in semantic context
  - Rank-aware clustering
  - Modeling ranked agreement
- ✓ Probabilistic data management: deriving probabilistic databases from incomplete data

09/2003 – 10/2009 Graduate Research Assistant
Department of Computer Science
Columbia University, New York NY USA
Advised by Professor Kenneth A. Ross

- ✓ Doctoral thesis "Search and Ranking in Semantically Rich Applications"
- ✓ Research in data representation in complex domains, in scope of two projects
  - o Computational Tools for Modeling, Visualizing and Analyzing Archaeological Sites
  - o MAGNet: Multiscale Analysis of Genomic and Cellular Networks

Summer 2008 Research Intern
Summer 2007 Community Systems and Social Systems Groups
Yahoo! Research, New York NY USA
Supervised by Dr. Sihem Amer-Yahia

- ✓ Data representation and processing techniques for social information sites
- ✓ Rank-aware exploration of structured datasets

Summer 2006 Research Intern
Databases and Information Systems Group
Max-Planck Institut für Informatik, Saarbruecken, Germany
Supervised by Professor Gerhard Weikum

- ✓ Semantically enriched ranking for graphs such as Wikipedia and the Web

Summer 2005 Research Intern
Advanced Optimization Group
IBM Almaden Research Center, San Jose, CA, USA
Supervised by Jun Rao, Volker Markl, and Guy Lohman

- ✓ Novel query optimization techniques for the DB2 optimizer

02/1998 – 08/1998 Undergraduate Research Assistant
01/1996 – 03/1997 Center for Knowledge Communication
Department of Computer Science, UMass Amherst, MA USA
Supervised by Dr. Joseph Beck

- ✓ Example generation strategies and UI design for an Intelligent Mathematics Tutor

Teaching Experience

✓ Computer and Information Science Department University of Pennsylvania	Instructor, Spring 2012 *Data structures and algorithms with Java* (CIS 121) Instructor, Fall 2010 *Advanced topics in databases: information discovery in massive datasets* (CIS 650), co-taught with Susan Davidson
✓ Computer Science Department Columbia University	Instructor, Summer 2009 *Object-oriented programming and design in Java* (COMS 1007) Teaching Assistant, Spring 2006 and Spring 2007 *Advanced database systems* (COMS 6111)
✓ Joint EDBT / RussIR Summer School Saint Petersburg, Russia	Instructor, Summer 2011 *Top-K processing for search and information discovery in social applications*, co-taught with Sihem Amer-Yahia
✓ Encoda Systems, New York, NY	Instructor, Spring 2003 *Technical seminars on foundations of relational database systems*

- ✓ Science Enrichment Program, University of Massachusetts Amherst — Teaching Assistant, Summer 1996; *Introduction to programming,* a course for high-school students

Industry Experience

08/2001 – 08/2003	Database Developer and Administrator Encoda Systems, New York NY USA
08/2000 – 07/2001	Web and Database Developer MediaPartnerships, New York, NY USA
09/1998 – 08/2000	Database Developer Juno Online Services, New York NY USA
09/1997 – 01/1998	Programmer-Intern Hewlett-Packard, Medical Equipment R&D, Böblingen, Germany

Professional Activities

- ✓ Program committee member: WWW 2012, SSDBM 2012, AMW 2012, WebDB (2012, 2011), DBSocial 2012, SMANE 2012, ICDE 2011, EDBT 2011, CIKM 2010, SIASP 2010, VLDB 2009
- ✓ Reviewer: PVLDB 2012-13, CACM 2012, TODS (2012, 2011, 2008), TKDE 2011, Bioinformatics 2011, SNAM 2011, TOSEM 2010
- ✓ External reviewer: SIGMOD 2008, VLDB 2008, VLDB 2006
- ✓ Proceedings chair: EDBT/ICDT 2011
- ✓ Web chair: PersDB 2009
- ✓ Local arrangements chair: Spring 2008 North East DB/IR Day

Invited Talks

- ✓ VirginiaTech, CS colloquium, 3/20/2012
- ✓ Max Planck Institute for Software Systems, SWS colloquium, 3/15/2012
- ✓ University of Waterloo, CS colloquium, 2/15/2012
- ✓ Stevens Institute of Technology, CS colloquium, 1/23/2012
- ✓ Max Planck Institute for Informatics, Databases and Information Systems colloquium, 12/13/2011
- ✓ INRIA/Telecom ParisTech, DBWeb-Webdam seminar, 12/5/2011
- ✓ Drexel University, iSchool colloquium, 11/30/2011
- ✓ University of Pittsburgh, CS colloquium, 11/28/2011
- ✓ Harvard University, Systems Research at Harvard (SYRAH) colloquium, 10/28/2011
- ✓ MIT, Computer Science and Artificial Intelligence Laboratory (CSAIL) colloquium, 10/27/2011
- ✓ Boston University, Hariri seminar, 10/26/2011
- ✓ Yandex, Moscow, Russia, Research colloquium, 8/23/2011
- ✓ Northeastern University, CCS colloquium, 3/18/2011
- ✓ Illinois Institute of Technology, CS colloquium, 2/28/2011
- ✓ IBM Watson, Research Colloquium, 2/11/2011
- ✓ Yahoo! Research Barcelona, 12/16/2010
- ✓ Polytechnic Institute of NYU, CSE colloquium, 04/06/2010
- ✓ University of Southern California, CS colloquium, 03/01/2010
- ✓ University of Pennsylvania, Computational Biology and Informatics (CBIL) colloquium, 01/21/2010
- ✓ University of Pennsylvania, Database Group colloquium, 04/22/2009
- ✓ Technical University of Berlin, Database and Information Management (DIMA) colloquium, 01/19/2009

Honors and Awards

- ✓ NSF/CRA Computing Innovations Fellow, 2009-2010 and 2010-2011
- ✓ Recipient of 2008 Michelman award for service to the Computer Science Department, Columbia University
- ✓ Recipient of 1997 DuBois Undergraduate Research Scholarship, University of Massachusetts, Amherst

Leadership

- ✓ Senator, Columbia University Senate, member of the Senate Education Committee, 09/2006-05/2007
- ✓ Dean Search Committee, Columbia University, School of Engineering and Applied Science, Spring 2008
- ✓ Steering Committee Member, Columbia University Graduate Student Advisory Council, 09/2004 – 05/2009

Languages

fluent in English, German, Russian, and Serbian; intermediate Italian.

References

are available and will be furnished upon request.