

# Luis Gravano's Curriculum Vitae

Last updated: April 6, 2026

## Contact Information

- Address:  
Computer Science Department  
Columbia University  
1214 Amsterdam Avenue  
New York, NY 10027, USA
  - Phone: +1-212-853-8465
  - Email: [gravano@cs.columbia.edu](mailto:gravano@cs.columbia.edu)
  - Homepage: <https://www.cs.columbia.edu/~gravano> (<https://www.cs.columbia.edu/~gravano>)
- 

## Education

- Ph.D. Computer Science, September 1997. Stanford University, Stanford, California.  
Dissertation: Querying Multiple Document Collections across the Internet (advisor: Hector Garcia-Molina).
- M.S. Computer Science, March 1994. Stanford University, Stanford, California.
- B.S. Computer Science ("Licenciatura en Informática"), July 1991. Escuela Superior Latinoamericana de Informática (ESLAI), Argentina.

## Professional Employment

- July 2013 - Present: Professor. Computer Science Department, Columbia University, New York City, New York
- July 2002 - June 2013: Associate Professor. Computer Science Department, Columbia University, New York City, New York
- September 1997 - June 2002: Assistant Professor. Computer Science Department, Columbia University, New York City, New York
- September 2018 - June 2019: Visiting Faculty Researcher. Google Inc., New York City, New York (on sabbatical from Columbia University)
- January - August 2001: Senior Research Scientist. Google Inc., Mountain View, California (on leave from Columbia University)
- August 2000; July 2002: Consulting Researcher. Microsoft Research, Redmond, Washington
- June - July 2000: Academic Consultant. Google Inc., Mountain View, California
- July 2000: Consultant. Gigabeat Inc., Palo Alto, California
- July - August 1999: Visiting Professor. Computer Science Department, University of Buenos Aires, Argentina

- June - July 1999: Consulting Researcher. Microsoft Research, Redmond, Washington
- 1992 - 1997: Research Assistant. Computer Science Department, Stanford University, Stanford, California
- 1995 - 1996: Research Intern. Hewlett-Packard Laboratories, Palo Alto, California
- 1995: Supplemental Research Associate. IBM Almaden Research Center, San Jose, California
- July 1994: Instructor. Computer Science Department, University of Buenos Aires, Argentina
- 1993 - 1994: Teaching Assistant. Computer Science Department, Stanford University, Stanford, California
- 1990 - 1992: Resident Researcher. IBM Argentina, Buenos Aires, Argentina
- 1990 - 1992: Student Visitor. IBM Almaden Research Center, San Jose, California

## Honors and Awards

- 2025 ACM SIGMOD "Test of Time" Award, for paper "k-Shape: Efficient and Accurate Clustering of Time Series," by John Paparrizos and Luis Gravano
- "Distinguished Faculty Teaching Award," Columbia Engineering Alumni Association, Columbia University, 2012
- "Distinguished Teacher Award," Computer Science Department, Columbia University, 2011
- "Best Paper" Award, 2006 ACM SIGMOD International Conference on Management of Data (SIGMOD 2006), 2006
- "Best Paper" Award, 21st IEEE International Conference on Data Engineering (ICDE 2005), 2005
- "Best Student Paper" Award, 19th IEEE International Conference on Data Engineering (ICDE 2003), 2003
- CAREER Award, National Science Foundation (NSF), 1998
- "Most Original Paper" Award, International Conference on Parallel Processing (ICPP '92), 1992

## Grants and Gifts

- NSF. III: Medium: Adaptive Information Extraction from Social Media for Actionable Inferences in Public Health, IIS-15-63785, with D. Hsu (CoPI). \$1,196,617 for September 2016-August 2023.
- Microsoft Research. Unrestricted Cash Gift. \$50,000. 2015.
- Bloomberg. Unrestricted Cash Gift. \$30,000. 2014.
- Microsoft Research. Unrestricted Cash Gift. \$50,000. 2014.
- Microsoft Research. Unrestricted Cash Gift. \$50,000, 2013.
- Google Research Award. Information Extraction from Social Media: Detecting Disease Outbreaks. \$84,295, 2013.
- NSF. IGERT: From Data to Solutions: A New PhD Program in Transformational Data & Information Sciences Research and Innovation, with J. Hirschberg (project leader) et al. \$3,000,000 for July 2012-June 2017 (Gravano's exact share to be determined).
- Microsoft Research. Unrestricted Cash Gift. \$30,000, 2012.
- ODNI/IARPA. FUSE: Discovering and Explaining Technical Emergence through Analysis of the Language and Structure of Scientific Publications, with K. McKeown (project leader) et al.

\$11,513,044 for 2011-2016 (conditional renewal at various stages of project based on evaluation; Gravano's approximate share: \$1,118,641).

- Google Research Award. Capturing Real-World Event Content Across Social Media Sites. \$72,071 for 2011-2012.
- Microsoft Research, Data Management, Exploration, and Mining Group. Unrestricted Cash Gift. \$25,000, 2011.
- Amazon Web Services in Education Research Grant. \$7,500 for 2011-2013.
- NSF. III: Small: Collaborative Research: Detection and Presentation of Community and Global Event Content from Social Media Sources, IIS-10-17389, with M. Naaman (Cornell University). Columbia's share: \$249,551 for September 2010-August 2015.
- Yahoo! Faculty Research and Engagement Gift. Specialized Extraction of Entities and their Attributes. \$10,000 for 2010-2011.
- Microsoft Research, Data Management, Exploration, and Mining Group. Unrestricted Cash Gift. \$24,000, 2010.
- Google Research Award. Finding and Characterizing the World's Event Media. \$70,000 for January-December 2010.
- Google Research Award. Google Desktop Meets DejaView: Display-Centric Desktop Search, with Jason Nieh. \$70,000 for September 2009-August 2010.
- Yahoo! Faculty Research and Engagement Gift. User-Specific Extraction of Entity Lists and Attributes. \$7,500 for 2009-2010.
- NSF. III-COR-Small: Beyond Keyword Search: Enabling Diverse Structured Query Paradigms over Text Databases, IIS-08-11038. \$448,976 for September 2008-August 2013.
- Google Research Award. Searching for Events of All Sizes, Everywhere, with Hila Becker. \$72,779 for September 2008-August 2009.
- Microsoft Research, Data Management, Exploration, and Mining Group. Unrestricted Cash Gift. \$35,000 for 2007-2008; \$35,000 for 2006-2007; \$35,000 for 2005-2006; \$35,000 for 2004-2005; \$50,000 for 2003-2004; \$50,000 for 2002-2003; \$50,000 for 2001-2002; \$50,000 for 2000-2001.
- NSF. KDD: Tools for Monitoring Online Information Sources, with K. McKeown, J. Hirschberg, and O. Rambow. Started in June 2002 (part of project through December 2004). \$695,000 for Year 1, \$500,000 for Year 2.
- NSF. DLI-Phase 2: A Patient Care Digital Library: Personalized Retrieval and Summarization of Multimedia Information, IIS-98-17434, with K. McKeown et al. \$5,002,375 for 1999-2004.
- Lucent Technologies. Research Grant. \$20,000 for 2000-2001.
- NSF. Digital Government: The CARDGIS Energy Data Collection, EIA-98-76739, with S. Stolfo et al. \$1,631,623 for 1999-2002 (part of project from September 1999 through August 2000).
- Microsoft Research. Unrestricted Cash Gift. \$24,000 for 1999-2000.
- Microsoft Research, Database Group. Unrestricted Cash Gift. \$42,000 for 1999-2000.
- NSF. CAREER: Querying Information Sources Across the Internet, IIS-97-33880. \$299,985 for September 1998-August 2002.
- NSF. An Environment for Illustrated Briefing and Follow-up Search Over Live Multimedia Information, IRI-96-19124, with K. McKeown and S.-F. Chang. \$732,056 for March 1997-

February 2000 (joined on-going project in September 1997).

## Patents

1. Systems and Methods for Using Anchor Text as Parallel Corpora for Cross-Language Information Retrieval, L. Gravano and M. Henzinger, United States Patent 8,631,010, issued January 14, 2014 (continuation of United States Patents 7,146,358, 7,814,103, 7,996,402, and 8,190,608)
2. Systems and Methods for Using Anchor Text as Parallel Corpora for Cross-Language Information Retrieval, L. Gravano and M. Henzinger, United States Patent 8,190,608, issued May 29, 2012 (continuation of United States Patents 7,146,358, 7,814,103, and 7,996,402)
3. Systems and Methods for Using Anchor Text as Parallel Corpora for Cross-Language Information Retrieval, L. Gravano and M. Henzinger, United States Patent 7,996,402, issued August 9, 2011 (continuation of United States Patents 7,146,358 and 7,814,103)
4. Systems and Methods for Using Anchor Text as Parallel Corpora for Cross-Language Information Retrieval, L. Gravano and M. Henzinger, United States Patent 7,814,103, issued October 12, 2010 (continuation of United States Patent 7,146,358)
5. String Predicate Selectivity Estimation, S. Chaudhuri, V. Ganti, and L. Gravano, United States Patent 7,149,735, issued December 12, 2006
6. Systems and Methods for Using Anchor Text as Parallel Corpora for Cross-Language Information Retrieval, L. Gravano and M. Henzinger, United States Patent 7,146,358, issued December 5, 2006
7. Method of Building Multidimensional Workload-Aware Histograms, S. Chaudhuri, N. Bruno, and L. Gravano, United States Patent 7,007,039, issued February 28, 2006
8. Method for Cost-Based Optimization over Multimedia Repositories, S. Chaudhuri and L. Gravano, United States Patent 5,806,061, issued September 8, 1998
9. Method of Packet Routing in Torus Networks with Two Buffers per Edge, R. Cypher and L. Gravano, United States Patent 5,444,701, issued August 22, 1995

## Editorships

- Associate Editor, ACM Transactions on Database Systems, 2004-2010
- Co-editor, SIGMOD Record Special Issue on "Managing Information Extraction" (Editor: AnHai Doan), vol. 37, no. 4, December 2008
- Associate Editor, ACM Transactions on Information Systems, 1997-2005
- Associate Editor, IEEE Data Engineering Bulletin, 2000-2002
  - Editor, special issue on "Text and Databases," Vol. 24, No. 4, December 2001
  - Editor, special issue on "Next-Generation Web Search," Vol. 23, No. 3, September 2000
- Associate Editor, ACM SIGMOD Digital Symposium Collection (DiSC), 1998-2001

## Program Committees

- 2024: 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (reviewer)

- 2019: The Web Conference 2019 (Web Search and Mining Track), International Workshop on Misinformation, Computational Fact-Checking and Credible Web (MisinfoWorkshop 2019)
- 2016: 25th International World Wide Web Conference (WWW 2016, Web Search Systems and Applications), ACM SIGMOD International Conference on Management of Data (ACM SIGMOD 2016, Tutorials Track), 25th International Joint Conference on Artificial Intelligence (IJCAI 2016, AI & Web Track)
- 2015: 24th International World Wide Web Conference (WWW 2015, Web Search Systems and Applications), 41st International Conference on Very Large Databases (VLDB 2015, Industrial Track)
- 2012: ACM SIGMOD International Conference on Management of Data (ACM SIGMOD 2012: Program Chair)
- 2011: 27th IEEE International Conference on Data Engineering (IEEE ICDE 2011, Industry Track), 37th International Conference on Very Large Databases (VLDB 2011: Workshop Program Co-Chair), 14th International Workshop on the Web and Databases (WebDB 2011)
- 2010: 26th IEEE International Conference on Data Engineering (IEEE ICDE 2010: Tutorials/Seminars Co-Chair), 19th International World Wide Web Conference (WWW 2010, Search Track), 36th International Conference on Very Large Databases (VLDB 2010), 2nd International Workshop on Keyword Search on Structured Data (KEYS 2010)
- 2009: ACM SIGMOD International Conference on Management of Data (ACM SIGMOD 2009), 18th International World Wide Web Conference (WWW 2009, Search Track), 12th International Workshop on the Web and Databases (WebDB 2009), VLDB 2009 Ph.D. Workshop
- 2008: 24th IEEE International Conference on Data Engineering (IEEE ICDE 2008: Program Committee Vice-Chair, for "Web Search and Deep Web" area; member of Best Research Paper Selection Committee), 34th International Conference on Very Large Databases (VLDB 2008)
- 2007: 33rd International Conference on Very Large Databases (VLDB 2007)
- 2006: ACM SIGMOD International Conference on Management of Data (ACM SIGMOD 2006), Workshop on Information Integration (IIWorkshop)
- 2005: ACM SIGMOD International Conference on Management of Data (ACM SIGMOD 2005)
- 2004: 7th International Workshop on the Web and Databases (WebDB 2004: Co-Chair), 13th ACM Conference on Information and Knowledge Management (ACM CIKM 2004: Program Committee Co-Chair, Databases), 20th IEEE International Conference on Data Engineering (IEEE ICDE 2004), 13th International World Wide Web Conference (WWW 2004), 27th International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR 2004)
- 2003: 26th International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR 2003), 12th ACM Conference on Information and Knowledge Management (ACM CIKM 2003), ACM SIGIR 2003 Workshop on Distributed Information Retrieval
- 2002: 28th International Conference on Very Large Databases (VLDB 2002), ACM SIGMOD International Conference on Management of Data (ACM SIGMOD 2002), 5th International Workshop on the Web and Databases (WebDB 2002)

- 2001: 24th International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR 2001), 17th IEEE International Conference on Data Engineering (IEEE ICDE 2001)
- 2000: 26th International Conference on Very Large Databases (VLDB 2000), 16th IEEE International Conference on Data Engineering (IEEE ICDE 2000), IEEE Advances in Digital Libraries (IEEE ADL 2000)
- 1999: ACM SIGMOD International Conference on Management of Data (ACM SIGMOD'99), 15th IEEE International Conference on Data Engineering (IEEE ICDE'99), 22nd International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR'99), 4th IFCIS Conference on Cooperative Information Systems (CoopIS'99), Symposium on Software Technology (SoST'99)
- 1998: 3rd ACM International Conference on Digital Libraries (ACM DL'98), Symposium on Software Technology (SoST'98)
- 1997: Symposium on Software Technology (SoST'97)

## Invited Talks

- 2012: Identificación en Twitter de Eventos del Mundo Real y de sus Contenidos Asociados, June 2012, Pragma Consultores, Buenos Aires, Argentina
- 2009: Querying Text Databases and the Web: Beyond Traditional Keyword Search, June 2009, keynote talk at the First International Workshop on Keyword Search on Structured Data (KEYS 2009), Providence, Rhode Island
- 2008: Information Extraction Over Text Databases: What's Ranking Got To Do With It?, April 2008, keynote talk at the Second International Workshop on Ranking in Databases (DBRank 2008), Cancun, Mexico; Information Extraction Over Text Databases: What's Ranking Got To Do With It?, July 2008, Pragma Consultores, Buenos Aires, Argentina
- 2004: Hidden-Web Databases: Classification and Search, March 2004, Polytechnic University, Brooklyn, New York
- 2003: Hidden-Web Databases: Classification and Search, August 2003, IBM T. J. Watson Research Center, Hawthorne, New York; Hidden-Web Databases: Classification and Search, April 2003, University of Waterloo, Waterloo, Ontario, Canada
- 2002: Hidden-Web Databases: Classification and Search, December 2002, Stern School of Business, New York University, New York City, New York; Hidden-Web Databases: Classification and Search, November 2002, Lucent-Bell Labs, Murray Hill, New Jersey; Text- and Web-Database Research at Columbia University, July 2002, Microsoft Research, Redmond, Washington; Web Mining Meets Web Search, June 2002, keynote talk at the ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD 2002), Madison, Wisconsin; Hidden-Web Databases: Classification and Search, March 2002, University of Pennsylvania, Philadelphia, Pennsylvania
- 2001: Characterizing Web Resources for Improved Searching and Browsing, April 2001, IBM Almaden Research Center, San Jose, California; Characterizing Web Resources for Improved Searching and Browsing, March 2001, Hewlett-Packard Laboratories, Palo Alto, California;

Probe, Count, and Classify: Categorizing Hidden-Web Databases, February 2001, Google Inc., Mountain View, California

- 2000: Characterizing Web Resources for Improved Searching and Browsing, September 2000, Princeton University, Princeton, New Jersey; Computing Geographical Scopes of Web Resources, and Other Classification Problems, June 2000, Google Inc., Mountain View, California
- 1999: Internet Query Processing, December 1999, NEC Research Institute, Princeton, New Jersey; Query Processing and Data Quality over the Internet, with Martina Marré, ECI'99, July 1999, University of Buenos Aires, Argentina
- 1997: Searching over Autonomous Text Sources, FedWeb'97, October 1997, Bethesda, Maryland

## Invited Panels and Working Groups

- Information Fusion in Counter-Terrorism. Participant and invited speaker in workshop organized by the National Research Council's Computer Science and Telecommunications Board, Washington D.C., June 2002
- How Agencies and Universities Are Collaborating on Research to Solve a Federal Statistical Data Integration Problem. Panel in FedWeb 2000: Meeting the Growing Demand for Government E-Service, Bethesda, Maryland, May 2000
- Integrating Information Retrieval and Databases in the WWW, Internet, Wireless Era. Panel in the 1999 NSF Information and Data Management Workshop: Research Agenda for the 21st Century, Los Angeles, California, March 1999
- Resource Indexing and Discovery In a Globally Distributed Digital Library. Second meeting of one of five NSF-EU Digital Library Collaboratory Working Groups, Washington DC (one of five participants from the US), February 1998
- Resource Indexing and Discovery In a Globally Distributed Digital Library. First meeting of one of five NSF-EU Digital Library Collaboratory Working Groups, Budapest, Hungary (one of five participants from the US), November 1997
- InfoBus: Experience in Linking Heterogeneous Systems. Panel in the 2nd ACM International Conference on Digital Libraries (ACM DL'97), Philadelphia, Pennsylvania, July 1997

## Other Professional Activities

- NSF Information, Integration, and Informatics (NSF III) 2010 Workshop: Steering committee member
- North East DB/IR Day Workshop, April 18, 2008: Chaired one-day workshop, bringing together database and information retrieval researchers in the Northeastern United States.
- Grant Panels: NSF, 1998; NSF, 2001; NSF, 2005; NSF, 2009; NSF, 2011
- Journal Article Reviews (in addition to editorial boards): Information Retrieval, 2004; Journal of Computer and System Sciences, 2001; International Journal on Digital Libraries, 2000; ACM Computing Surveys, 2000, 2002; VLDB Journal, 1999; ACM Transactions on Database Systems, 1997; ACM Transactions on Information Systems, 1997.

- Conference Paper Refereeing (in addition to program committees): 25th International Conference on Very Large Databases (VLDB'99), ACM 1997 SIGMOD International Conference on Management of Data (ACM SIGMOD'97), ACM 1996 SIGMOD International Conference on Management of Data (ACM SIGMOD'96), 1st ACM International Conference on Digital Libraries (ACM DL'96), 21st International Conference on Very Large Databases (VLDB'95).
- STARTS Informal Standards Effort, 1996-1997: Coordinated an informal standards effort for Internet searching, which involved Netscape, Microsoft Network, Infoseek, Fulcrum, and Verity, among others, and produced STARTS, the Stanford Protocol Proposal for Internet Retrieval and Search. Organized a workshop at Stanford with 40 representatives of the participating organizations to agree on the final proposal.
- Student Representative to the Stanford Computer Forum Committee, 1995-1996. (The Stanford Computer Forum is an industrial affiliate program at Stanford University.) Co-chaired a poster session at the Forum's annual meeting, showing Computer-Science projects at Stanford.

## Papers in Refereed Journals

1. Interactive Machine Teaching by Labeling Rules and Instances, G. Karamanolakis, D. Hsu, and L. Gravano, in Transactions of the Association for Computational Linguistics, 12:1441–1459, 2024.
2. Discovering Foodborne Illness in Online Restaurant Reviews, T. Effland, A. Lawson, S. Balter, K. Devinney, V. Reddy, H. Waechter, L. Gravano, and D. Hsu, in Journal of the American Medical Informatics Association, vol. 25, no. 12, pages 1586–1592, Dec. 2018.
3. Fast and Accurate Time-Series Clustering, I. Paparrizos and L. Gravano, in ACM Transactions on Database Systems, vol. 42, no. 2, June 2017.
4. Sampling Strategies for Information Extraction over the Deep Web, P. Barrio and L. Gravano, in Information Processing & Management, vol. 53, no. 2, pages 309–331, Mar. 2017.
5. Predicting the Impact of Scientific Concepts Using Full-Text Features, K. McKeown and many others, in Journal of the Association for Information Science and Technology, vol. 67, no. 11, pages 2684-2696, Nov. 2016.
6. Answering General Time-Sensitive Queries, W. Dakka, L. Gravano, and P. Ipeirotis, in IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 2, pages 220-235, Feb. 2012.
7. Hip and Trendy: Characterizing Emerging Trends on Twitter, M. Naaman, H. Becker, and L. Gravano, in Journal of the American Society for Information Science and Technology, vol. 62, no. 5, pages 902–918, May 2011.
8. Classification-Aware Hidden-Web Text Database Selection, P. Ipeirotis and L. Gravano, in ACM Transactions on Information Systems, vol. 26, no. 2, art. 6 (66 pages), Mar. 2008.
9. Towards a Query Optimizer for Text-Centric Tasks, P. Ipeirotis, E. Agichtein, P. Jain, and L. Gravano, in ACM Transactions on Database Systems, vol. 32, no. 4, art. 21 (46 pages), Nov. 2007.
10. Modeling and Managing Changes in Text Databases, P. Ipeirotis, A. Ntoulas, J. Cho, and L. Gravano, in ACM Transactions on Database Systems, vol. 32, no. 3, art. 14 (38 pages), Aug.

2007.

11. Optimizing Top-k Selection Queries over Multimedia Repositories, S. Chaudhuri, L. Gravano, and A. Marian, in *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 8, pages 992-1009, Aug. 2004.
12. Evaluating Top-k Queries over Web-Accessible Databases, A. Marian, N. Bruno, and L. Gravano, in *ACM Transactions on Database Systems*, vol. 29, no. 2, pages 319-362, June 2004.
13. Learning to Find Answers to Questions on the Web, E. Agichtein, S. Lawrence, and L. Gravano, in *ACM Transactions on Internet Technology*, vol. 4, no. 2, pages 129-162, May 2004.
14. QProber: A System for Automatic Classification of Hidden-Web Databases, L. Gravano, P. Ipeirotis, and M. Sahami, in *ACM Transactions on Information Systems*, vol. 21, no. 1, pages 1-41, Jan. 2003.
15. Top-k Selection Queries over Relational Databases: Mapping Strategies and Performance Evaluation, N. Bruno, S. Chaudhuri, and L. Gravano, in *ACM Transactions on Database Systems*, vol. 27, no. 2, pages 153-187, Jun. 2002.
16. GLOSS: Text-Source Discovery over the Internet, L. Gravano, H. Garcia-Molina, A. Tomasic, in *ACM Transactions on Database Systems*, vol. 24, no. 2, pages 229-264, Jun. 1999.
17. The Stanford Digital Library Metadata Architecture, M. Baldonado, C.-C. K. Chang, L. Gravano, and A. Paepcke, in *International Journal on Digital Libraries*, vol. 1, no. 2, pages 108-121, Sep. 1997.
18. Data Structures for Efficient Broker Implementation, A. Tomasic, L. Gravano, C. Lue, P. Schwarz, and L. Haas, in *ACM Transactions on Information Systems*, vol. 15, no. 3, pages 223-253, Jul. 1997.
19. Storage-Efficient, Deadlock-Free Packet Routing Algorithms for Torus Networks, R. Cypher and L. Gravano, in *IEEE Transactions on Computers*, vol. 43, no. 12, pages 1376-1385, Dec. 1994.
20. Requirements for Deadlock-Free, Adaptive Packet Routing, R. Cypher and L. Gravano, in *SIAM Journal on Computing*, vol. 23, no. 6, pages 1266-1274, Dec. 1994.
21. Adaptive Deadlock- and Livelock-Free Routing with All Minimal Paths in Torus Networks, L. Gravano, G. Pifarre, P. Berman, and J. Sanz, in *IEEE Transactions on Parallel and Distributed Systems*, vol. 5, no. 12, pages 1233-1251, Dec. 1994.
22. Adaptive Deadlock- and Livelock-Free Routing in the Hypercube Network, G. Pifarre, L. Gravano, G. Denicolay, J. Sanz, in *IEEE Transactions on Parallel and Distributed Systems*, vol. 5, no. 11, pages 1121-1139, Nov. 1994.
23. Fully Adaptive Minimal Deadlock-Free Packet Routing in Hypercubes, Meshes, and Other Networks: Algorithms and Simulations, G. Pifarre, L. Gravano, S. Felperin, and J. Sanz, in *IEEE Transactions on Parallel and Distributed Systems*, vol. 5, no. 3, pages 247-263, Mar. 1994.

## Book Chapter

1. XML & Data Streams, N. Bruno, L. Gravano, N. Koudas, and D. Srivastava. Chapter 4 in "Stream Data Management," edited by N. Chaudhry, K. Shaw, and M. Abdelguerfi, Series:

## Papers in Refereed Conferences

1. Geospatial and Geosocial Dimensions of Foodborne Illness as Reflected in Yelp Restaurant Reviews, E. Shaveet, S. Chowdhury, D. Hsu, and L. Gravano, 2024 International Conference on Social Media & Society, 2024.
2. Cross-Lingual Text Classification with Minimal Resources by Transferring a Sparse Teacher, G. Karamanolakis, D. Hsu, and L. Gravano, in Proc. of Findings of the 2020 Conference on Empirical Methods in Natural Language Processing (Findings of EMNLP 2020), 2020.
3. Leveraging Just a Few Keywords for Fine-Grained Aspect Detection Through Weakly Supervised Co-Training, G. Karamanolakis, D. Hsu, and L. Gravano, in Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP 2019), 2019 (23.8% accepted).
4. Ranking Deep Web Text Collections for Scalable Information Extraction, P. Barrio, L. Gravano, and C. Devellder, in Proc. of the 24th ACM Conference on Information and Knowledge Management (CIKM 2015), 2015 (18% accepted in "long paper" category in Knowledge Management Track).
5. k-Shape: Efficient and Accurate Clustering of Time Series, J. Paparrizos and L. Gravano, in Proc. of the 2015 ACM SIGMOD International Conference on Management of Data, 2015.
6. Learning to Rank Adaptively for Scalable Information Extraction, P. Barrio, G. Simões, H. Galhardas, and L. Gravano, in Proc. of the 18th International Conference on Extending Database Technology (EDBT 2015), pages 241-252, 2015.
7. When Speed Has a Price: Fast Information Extraction Using Approximate Algorithms, G. Simões, H. Galhardas, and L. Gravano, in Proc. of the VLDB Endowment, vol. 6, no. 13, pages 1462-1473, 2013.
8. Identifying Content for Planned Events Across Social Media Sites, H. Becker, D. Iter, M. Naaman, and L. Gravano, in Proc. of the 2012 ACM International Conference on Web Search and Data Mining (WSDM 2012), pages 533-542, 2012 (20.7% accepted; one of 30 papers, or 8.3% of submissions, selected for full-length plenary-session presentation).
9. Beyond Trending Topics: Real-World Event Identification on Twitter, H. Becker, M. Naaman, and L. Gravano, in Proc. of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011), pages 438-441, 2011 (short 4-page "poster" paper).
10. Selecting Quality Twitter Content for Events, H. Becker, M. Naaman, and L. Gravano, in Proc. of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011), pages 442-445, 2011 (short 4-page "poster" paper).
11. Learning Similarity Metrics for Event Identification in Social Media, H. Becker, M. Naaman, and L. Gravano, in Proc. of the 2010 ACM International Conference on Web Search and Data Mining (WSDM 2010), pages 291-300, 2010 (15.5% accepted).
12. Join Optimization of Information Extraction Output: Quality Matters!, A. Jain, P. Ipeirotis, A. Doan, and L. Gravano, in Proc. of the 25th IEEE International Conference on Data Engineering (ICDE 2009), pages 186-197, 2009 (16.8% accepted in "long paper" category).

13. Answering General Time-Sensitive Queries, W. Dakka, L. Gravano, and P. Ipeirotis, in Proc. of the 17th ACM Conference on Information and Knowledge Management (CIKM 2008), pages 1437-1438, 2008 (short 2-page "poster" paper; 16% accepted in "poster" paper category).
14. Optimizing SQL Queries over Text Databases, A. Jain, A. Doan, and L. Gravano, in Proc. of the 24th IEEE International Conference on Data Engineering (ICDE 2008), pages 636-645, 2008 (12.1% accepted in "full presentation" category).
15. Efficient Summarization-Aware Search for Online News Articles, W. Dakka and L. Gravano, in Proc. of the 2007 ACM+IEEE Joint Conference on Digital Libraries (JCDL 2007), pages 63-72, 2007.
16. Efficient Keyword Search Across Heterogeneous Relational Databases, M. Sayyadian, H. LeKhac, A. Doan, and L. Gravano, in Proc. of the 23rd IEEE International Conference on Data Engineering (ICDE 2007), pages 346-355, 2007 (19% accepted).
17. SQL Queries Over Unstructured Text Databases, A. Jain, A. Doan, and L. Gravano, in Proc. of the 23rd IEEE International Conference on Data Engineering (ICDE 2007), pages 1255-1257, 2007 (short 3-page "poster" paper).
18. To Search or to Crawl? Towards a Query Optimizer for Text-Centric Tasks, P. Ipeirotis, E. Agichtein, P. Jain, and L. Gravano, in Proc. of the 2006 ACM SIGMOD International Conference on Management of Data, pages 265-276, 2006 ("Best Paper" Award; 13% accepted).
19. Modeling and Managing Content Changes in Text Databases, P. Ipeirotis, A. Ntoulas, J. Cho, and L. Gravano, in Proc. of the 21st IEEE International Conference on Data Engineering (ICDE 2005), pages 606-617, 2005 ("Best Paper" Award; 13% accepted).
20. When one Sample is not Enough: Improving Text Database Selection Using Shrinkage, P. Ipeirotis and L. Gravano, in Proc. of the 2004 ACM SIGMOD International Conference on Management of Data, pages 767-778, 2004 (16% accepted).
21. Selectivity Estimation for String Predicates: Overcoming the Underestimation Problem, S. Chaudhuri, V. Ganti, and L. Gravano, in Proc. of the 20th IEEE International Conference on Data Engineering (ICDE 2004), pages 227-238, 2004 (14% accepted).
22. Categorizing Web Queries According to Geographical Locality, L. Gravano, V. Hatzivassiloglou, and R. Lichtenstein, in Proc. of the 12th ACM Conference on Information and Knowledge Management (CIKM 2003), pages 325-333, 2003 (15% accepted).
23. Efficient IR-Style Keyword Search over Relational Databases, V. Hristidis, L. Gravano, and Y. Papakonstantinou, in Proc. of the 29th International Conference on Very Large Data Bases (VLDB 2003), pages 850-861, 2003 (15% accepted).
24. Text Joins in an RDBMS for Web Data Integration, L. Gravano, P. Ipeirotis, N. Koudas, and D. Srivastava, in Proc. of the 12th International World Wide Web Conference (WWW 2003), pages 90-101, 2003 (13% accepted).
25. Querying Text Databases for Efficient Information Extraction, E. Agichtein and L. Gravano, in Proc. of the 19th IEEE International Conference on Data Engineering (ICDE 2003), pages 113-124, 2003 ("Best Student Paper" Award; 14% accepted).
26. Navigation- vs. Index-Based XML Multi-Query Processing, N. Bruno, L. Gravano, N. Koudas, and D. Srivastava, in Proc. of the 19th IEEE International Conference on Data Engineering

(ICDE 2003), pages 139-150, 2003 (14% accepted).

27. Text Joins for Data Cleansing and Integration in an RDBMS, L. Gravano, P. Ipeirotis, N. Koudas, and D. Srivastava, in Proc. of the 19th IEEE International Conference on Data Engineering (ICDE 2003), pages 729-731, 2003 (short 3-page "poster" paper).
28. Distributed Search over the Hidden-Web: Hierarchical Database Sampling and Selection, P. Ipeirotis and L. Gravano, in Proc. of the 28th International Conference on Very Large Data Bases (VLDB 2002), pages 394-405, 2002 (16% accepted).
29. Evaluating Top-k Queries over Web-Accessible Databases, N. Bruno, L. Gravano, and A. Marian, in Proc. of the 18th IEEE International Conference on Data Engineering (ICDE 2002), pages 369-380, 2002 (19% accepted).
30. Extending SDARTS: Extracting Metadata from Web Databases and Interfacing with the Open Archives Initiative, P. Ipeirotis, T. Barry, and L. Gravano, in Proc. of the Second ACM+IEEE Joint Conference on Digital Libraries (JCDL 2002), pages 162-170, 2002 (33% accepted).
31. Approximate String Joins in a Database (Almost) for Free, L. Gravano, P. Ipeirotis, H. V. Jagadish, N. Koudas, S. Muthukrishnan, and D. Srivastava, in Proc. of the 27th International Conference on Very Large Data Bases (VLDB 2001), pages 491-500, 2001 (17% accepted).
32. Probe, Count, and Classify: Categorizing Hidden Web Databases, P. Ipeirotis, L. Gravano, and M. Sahami, in Proc. of the 2001 ACM SIGMOD International Conference on Management of Data, pages 67-78, 2001 (15% accepted).
33. STHoles: A Multidimensional Workload-Aware Histogram, N. Bruno, S. Chaudhuri, and L. Gravano, in Proc. of the 2001 ACM SIGMOD International Conference on Management of Data, pages 211-222, 2001 (15% accepted).
34. SDLIP + STARTS = SDARTS: A Protocol and Toolkit for Metasearching, N. Green, P. Ipeirotis, and L. Gravano, in Proc. of the First ACM+IEEE Joint Conference on Digital Libraries (JCDL 2001), pages 207-214, 2001.
35. PERSIVAL, a System for Personalized Search and Summarization over Multimedia Healthcare Information, K. McKeown, S.-F. Chang, J. Cimino, S. Feiner, C. Friedman, L. Gravano, V. Hatzivassiloglou, S. Johnson, D. Jordan, J. Klavans, A. Kushniruk, V. Patel, and S. Teufel, in Proc. of the First ACM+IEEE Joint Conference on Digital Libraries (JCDL 2001), pages 331-340, 2001.
36. Learning Search Engine Specific Query Transformations for Question Answering, E. Agichtein, S. Lawrence, and L. Gravano, in Proc. of the 10th International World Wide Web Conference (WWW10), pages 169-178, 2001 (20% accepted).
37. Computing Geographical Scopes of Web Resources, J. Ding, L. Gravano, and N. Shivakumar, in Proc. of the 26th International Conference on Very Large Data Bases (VLDB'00), pages 545-556, 2000 (15% accepted).
38. An Investigation of Linguistic Features and Clustering Algorithms for Topical Document Clustering, V. Hatzivassiloglou, L. Gravano, and A. Maganti, in Proc. of the 23rd ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'00), pages 224-231, 2000 (25% accepted).
39. Snowball: Extracting Relations from Large Plain-Text Collections, E. Agichtein and L. Gravano, in Proc. of the 5th ACM International Conference on Digital Libraries (DL'00), pages 85-94,

2000 (<33% accepted).

40. Evaluating Top-k Selection Queries, S. Chaudhuri and L. Gravano, in Proc. of the 25th International Conference on Very Large Data Bases (VLDB'99), pages 399-410, 1999 (15% accepted).
41. Merging Ranks from Heterogeneous Internet Sources, L. Gravano and H. Garcia-Molina, in Proc. of the 23rd International Conference on Very Large Data Bases (VLDB'97), pages 196-205, 1997 (15% accepted).
42. Metadata for Digital Libraries: Architecture and Design Rationale, M. Baldonado, C.-C. K. Chang, L. Gravano, and A. Paepcke, in Proc. of the 2nd ACM International Conference on Digital Libraries (DL'97), pages 47-56, 1997 (27% accepted).
43. STARTS: Stanford Proposal for Internet Meta-Searching, L. Gravano, C.-C. K. Chang, H. Garcia-Molina, and A. Paepcke, in Proc. of the 1997 ACM SIGMOD International Conference on Management of Data, pages 207-218, 1997 (21% accepted).
44. dSCAM: Finding Document Copies across Multiple Databases, H. Garcia-Molina, L. Gravano, and N. Shivakumar, in Proc. of the 4th International Conference on Parallel and Distributed Information Systems (PDIS'96), pages 68-79, 1996 (18% accepted).
45. Optimizing Queries over Multimedia Repositories, S. Chaudhuri and L. Gravano, in Proc. of the 1996 ACM SIGMOD International Conference on Management of Data, pages 91-102, 1996 (16% accepted).
46. Generalizing GLOSS to Vector-Space Databases and Broker Hierarchies, L. Gravano and H. Garcia-Molina, in Proc. of the 21st International Conference on Very Large Data Bases (VLDB'95), pages 78-89, 1995.
47. Precision and Recall of GLOSS Estimators for Database Discovery, L. Gravano, H. Garcia-Molina, and A. Tomasic, in Proc. of the 3rd International Conference on Parallel and Distributed Information Systems (PDIS'94), pages 103-106, 1994 (short paper).
48. The Effectiveness of GLOSS for the Text-Database Discovery Problem, L. Gravano, H. Garcia-Molina, and A. Tomasic, in Proc. of the 1994 ACM SIGMOD International Conference on Management of Data, pages 126-137, 1994 (15% accepted).
49. Requirements for Deadlock-Free, Adaptive Packet Routing, R. Cypher and L. Gravano, in Proc. of the 11th ACM Symposium on Principles of Distributed Computing (PODC '92), pages 25-33, 1992.
50. Adaptive, Deadlock-Free Packet Routing in Torus Networks with Minimal Storage, R. Cypher and L. Gravano, in Proc. of the 1992 International Conference on Parallel Processing (ICPP '92), pages 204-211, 1992 ("Most Original Paper" Award; 13% accepted).
51. Adaptive Deadlock- and Livelock-Free Routing with All Minimal Paths in Torus Networks, P. Berman, L. Gravano, G. Pifarre, and J. Sanz, in Proc. of the 4th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA '92), pages 3-12, 1992.
52. Adaptive Deadlock-Free Worm-Hole Routing in Hypercubes, L. Gravano, G. Pifarre, G. Denicolay, and J. Sanz, in Proc. of the 6th International Parallel Processing Symposium (IPPS '92), pages 512-515, 1992 (short paper).
53. Fully-Adaptive Routing: Packet Switching Performance and Worm-Hole Algorithms, S. Felperin, L. Gravano, G. Pifarre, and J. Sanz, in Proc. of Supercomputing '91, pages 654-663,

1991.

54. Fully-Adaptive Minimal Deadlock-Free Packet Routing in Hypercubes, Meshes, and Other Networks, G. Pifarre, L. Gravano, S. Felperin, and J. Sanz, in Proc. of the 3rd Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA '91), pages 278-290, 1991 (19% accepted).

## Papers in Refereed Workshops and Demonstration Sessions, and Other Refereed Publications and Presentations

1. Quantifying the Effects of COVID-19 on Restaurant Reviews, I. Cao, Z. Liu, G. Karamanolakis, D. Hsu, and L. Gravano, in Proc. of the 9th International Workshop on Natural Language Processing for Social Media (SocialNLP@NAACL 2021), 2021.
2. Detecting Foodborne Illness Complaints in Multiple Languages Using English Annotations Only, Z. Liu, G. Karamanolakis, D. Hsu, and L. Gravano, in Proc. of the 11th International Workshop on Health Text Mining and Information Analysis (LOUHI@EMNLP 2020), 2020.
3. Weakly Supervised Attention Networks for Fine-Grained Opinion Mining and Public Health, G. Karamanolakis, D. Hsu, and L. Gravano, in Proc. of the 5th Workshop on Noisy User-Generated Text (W-NUT 2019), 2019.
4. Training Neural Networks for Aspect Extraction Using Descriptive Keywords Only, G. Karamanolakis, D. Hsu, and L. Gravano, in Proc. of the 2nd Learning from Limited Labeled Data Workshop (LLD 2019), 2019.
5. Detecting Foodborne Disease Outbreaks Using Social Media (demonstration), F. Psallidas, L. Gravano, and many others, in NYC Media Lab's Annual Summit, 2014.
6. Information Extraction from Social Media for Public Health, N. Elhadad, L. Gravano, D. Hsu, S. Balter, V. Reddy, and H. Waechter, in KDD at Bloomberg Workshop, Data Frameworks Track (KDD 2014), 2014.
7. REEL: A Relation Extraction Learning Framework (poster), P. Barrio, G. Simões, H. Galhardas, and L. Gravano, in Proc. of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2014), 2014.
8. Using Online Reviews by Restaurant Patrons to Identify Unreported Cases of Foodborne Illness — New York City, 2012–2013, C. Harrison, M. Jorder, H. Stern, F. Stavinsky, V. Reddy, H. Hanson, H. Waechter, L. Lowe, L. Gravano, and S. Balter, in Centers for Disease Control and Prevention Morbidity and Mortality Weekly Report (CDC MMWR), vol. 63, no. 20, pages 441-445, May 2014.
9. Quality Impact of Value Matching and Scoring in Top-k Entity Attribute Extraction, M. Solomon, L. Gravano, and C. Yu, in Proc. of the 5th International Workshop on Ranking in Databases (DBRank 2011), 2011.
10. Automatic Identification and Presentation of Twitter Content for Planned Events (demonstration), H. Becker, F. Chen, D. Iter, M. Naaman, and L. Gravano, in Proc. of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011), pages 655-656, 2011.
11. Popularity-Guided Top-k Extraction of Entity Attributes, M. Solomon, C. Yu, and L. Gravano, in Proc. of the ACM SIGMOD Workshop on the Web and Databases (WebDB 2010), 6 pages,

2010 (32% accepted).

12. Exploiting Social Links for Event Identification in Social Media (poster), H. Becker, B. Xiao, M. Naaman, and L. Gravano, in Proc. of the 3rd Annual Workshop on Search in Social Media (SSM 2010), 2 pages, 2010.
13. Event Identification in Social Media, H. Becker, M. Naaman, and L. Gravano, in Proc. of the ACM SIGMOD Workshop on the Web and Databases (WebDB 2009), 6 pages, 2009 (33% accepted).
14. Modeling Query-Based Access to Text Databases, E. Agichtein, P. Ipeirotis, and L. Gravano, in Proc. of the ACM SIGMOD Workshop on the Web and Databases (WebDB 2003), pages 87-92, 2003 (25% accepted).
15. QXtract: A Building Block for Efficient Information Extraction from Text Databases (demonstration), E. Agichtein and L. Gravano, in Proc. of the 2003 ACM SIGMOD International Conference on Management of Data, page 663, 2003 (30% accepted).
16. Snowball: A Prototype System for Extracting Relations from Large Text Collections (demonstration), E. Agichtein, L. Gravano, J. Pavel, V. Sokolova, and A. Voskoboynik, in Proc. of the 2001 ACM SIGMOD International Conference on Management of Data, page 612, 2001 (~50% accepted).
17. PERSIVAL Demo: Categorizing Hidden-Web Resources (demonstration), P. Ipeirotis, L. Gravano, and M. Sahami, in Proc. of the First ACM+IEEE Joint Conference on Digital Libraries (JCDL 2001), page 454, 2001.
18. Automatic Classification of Text Databases through Query Probing, P. Ipeirotis, L. Gravano, and M. Sahami, in Proc. of the ACM SIGMOD Workshop on the Web and Databases (WebDB'00), pages 117-122, 2000 (29% accepted). Also in LNCS Series no. 1997, Springer, pages 245-255, 2001.
19. Combining Strategies for Extracting Relations from Text Collections, E. Agichtein, E. Eskin, and L. Gravano, in Proc. of the ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD 2000), pages 86-95, 2000 (58% accepted).
20. Exploiting Geographical Location Information of Web Pages, O. Buyukkocuten, J. Cho, H. Garcia-Molina, L. Gravano, and N. Shivakumar, in Proc. of the ACM SIGMOD Workshop on the Web and Databases (WebDB'99), pages 91-96, 1999 (29% accepted).

## Invited Papers

1. k-Shape: Efficient and Accurate Clustering of Time Series, I. Paparrizos and L. Gravano, in SIGMOD Record, Special Issue on "2015 ACM SIGMOD Research Highlights," vol. 45, no. 1, pages 69-76, March 2016.
2. Effective Event Identification in Social Media, F. Psallidas, H. Becker, M. Naaman, and L. Gravano, in IEEE Data Engineering Bulletin, vol. 36, no. 3, pages 42-50, September 2013.
3. Building Query Optimizers for Information Extraction: The SQoUT Project, A. Jain, P. Ipeirotis, and L. Gravano, in SIGMOD Record, Special Issue on "Managing Information Extraction," vol. 37, no. 4, pages 28-34, December 2008.
4. Query- vs. Crawling-based Classification of Searchable Web Databases, L. Gravano, P. Ipeirotis, and M. Sahami, in IEEE Data Engineering Bulletin, vol. 25, no. 1, pages 43-50, March

2002.

5. Using q-grams in a DBMS for Approximate String Processing, L. Gravano, P. Ipeirotis, H. V. Jagadish, N. Koudas, S. Muthukrishnan, L. Pietarinen, and D. Srivastava, in IEEE Data Engineering Bulletin, vol. 24, no. 4, pages 28-34, December 2001.
6. Simplifying Data Access: The Energy Data Collection Project, J. L. Ambite, Y. Arens, E. Hovy, A. Philpot, L. Gravano, V. Hatzivassiloglou, and J. Klavans, in IEEE Computer, vol. 34, no. 2, pages 47-54, February 2001.
7. Database Research at Columbia University, S.-F. Chang, L. Gravano, G. Kaiser, K. Ross, and S. Stolfo, in SIGMOD Record, vol. 27, no. 3, pages 75-80, September 1998.
8. Mediating and Metasearching on the Internet, L. Gravano and Y. Papakonstantinou, in IEEE Data Engineering Bulletin, vol. 21, no. 2, pages 28-36, June 1998.
9. The Stanford InfoBus and Its Service Layers: Augmenting the Internet with Higher-Level Information Management Protocols, M. Roscheisen, M. Baldonado, C.-C. K. Chang, L. Gravano, S. Ketchpel, and A. Paepcke, in Digital Libraries in Computer Science: The MeDoc Approach, LNCS Series no. 1392, Springer, pages 213-230, 1998.
10. Optimizing Queries over Multimedia Repositories, S. Chaudhuri and L. Gravano, in IEEE Data Engineering Bulletin, vol. 19, no. 4, pages 45-52, December 1996.
11. Routing Techniques for Massively Parallel Communication, S. Felperin, L. Gravano, G. Pifarre, and J. Sanz, in Proceedings of the IEEE, vol. 79, no. 4, pages 488-503, April 1991.

## Position Papers, Meeting Reports, and Miscellaneous Publications

1. Using Restaurant Review Websites to Identify Unreported Complaints of Foodborne Illness, C. Harrison, M. Joarder, H. Stern, F. Stavinsky, V. Reddy, L. Gravano, and S. Balter. Poster in 2013 CSTE Annual Conference, Pasadena, California, June 2013.
2. Characterizing Web Resources for Improved Search, L. Gravano. Position paper for the First NSF-DELOS Workshop on Information Seeking, Searching, and Querying in Digital Libraries, Zurich, Switzerland, December 2000.
3. Resource Indexing and Discovery In a Globally Distributed Digital Library, L. Gravano. Position paper for the NSF-EU Digital Library Collaboratory Working Group, Budapest, Hungary, November 1997.
4. Informal Internet Standards at Stanford, L. Gravano, C.-C. K. Chang, H. Garcia-Molina, A. Paepcke. Position paper for the 1996 World-Wide Web Consortium (W3C) Distributed Indexing/Searching Workshop, May 1996.

## Ph.D. Thesis Advising

- Active Ph.D. Advisees:
- Graduated Ph.D. Advisees:
  - Eugene Agichtein
    - Defended thesis: November 2004
    - Deposited thesis: May 2005
    - First employment: Postdoc Researcher, Microsoft Research, Redmond, Washington

- Current employment: Professor, Department of Computer Science, Emory University, Atlanta, Georgia
- Pablo Barrio
  - Defended thesis: September 2015
  - Deposited thesis: October 2015
  - First employment: Software Engineer, Google, New York
  - Current employment: Staff Software Engineer, Google, New York
- Hila Becker
  - Defended thesis: September 2011
  - Deposited thesis: October 2011
  - First employment: Software Engineer, Google, New York
  - Current employment: Director, Software Engineering, Google, New York
- Nicolás Bruno
  - Defended thesis: April 2003 (with distinction)
  - Deposited thesis: May 2003
  - First employment: Researcher, Microsoft Research, Redmond, Washington
  - Current employment: Gray Systems Lab, Microsoft Research, Redmond, Washington
- Wisam Dakka
  - Defended thesis: May 2008
  - Deposited thesis: November 2008
  - First employment: Software Engineer, Google, New York
  - Current employment: Co-founder, Meemo, California
- Panagiotis Ipeirotis
  - Defended thesis: July 2004 (with distinction)
  - Deposited thesis: September 2004
  - First employment: Assistant Professor, Department of Information, Operations, and Management Sciences, Stern School of Business, New York University, New York
  - Current employment: Professor, Department of Information, Operations, and Management Sciences, Stern School of Business, New York University, New York
- Alpa Jain
  - Defended thesis: May 2008
  - Deposited thesis: September 2008
  - First employment: Scientist, Yahoo! Labs, Santa Clara, California
  - Current employment: Principal Software Engineer, Google, Mountain View, California
- Giannis Karamanolakis
  - Defended thesis: July 2022
  - Deposited thesis: August 2022
  - First and current employment: Applied Scientist, Amazon Alexa AI, New York
- Amélie Marian
  - Defended thesis: June 2005

- Deposited thesis: September 2005
- First employment: Assistant Professor, Department of Computer Science, Rutgers University, New Jersey
- Current employment: Associate Professor, Department of Computer Science, Rutgers University, New Jersey
- Ioannis Paparrizos
  - Defended thesis: January 2018
  - Deposited thesis: July 2018
  - *Honorable Mention, KDD 2019 Dissertation Award*
  - First employment: Postdoc Scholar, Department of Computer Science, University of Chicago, Illinois
  - Current employment: Assistant Professor, Department of Computer Science and Engineering, The Ohio State University, Ohio
- Gonalo Simões (in Portugal, co-advised by Helena Galhardas)
  - Defended thesis and graduated: June 2016
  - First employment: Software Engineer, Google, London
  - Current employment: Senior Software Engineer, Google, London

## Bridge to the Ph.D. Program Advising

- Graduated Advisee: Eden Shaveet

## Teaching at Columbia University

- COMS W3998, W4901, E6901 Projects in Computer Science, Fall 1997-Present (1 Ph.D. student, 54 M.S. students, and 41 undergraduates)
- COMS E6111 Advanced Database Systems (graduate level), Spring 2000 (58 students), Spring 2002 (66 students), Spring 2003 (46 students), Spring 2004 (52 students), Spring 2006 (27 students), Spring 2007 (34 students), Spring 2008 (41 students), Spring 2009 (46 students), Spring 2010 (71 students), Spring 2011 (121 students), Fall 2011 (84 students), Fall 2012 (71 students), Spring 2014 (85 students), Spring 2015 (86 students), Fall 2015 (79 students), Fall 2016 (85 students), Fall 2017 (80 students), Spring 2020 (98 students), Spring 2021 (81 students), Spring 2022 (170 students), Spring 2023 (174 students), Spring 2024 (100 students), Spring 2025 (77 students)
- COMS W4111 Introduction to Databases (advanced undergraduate and graduate level), Fall 2008 (108 students), Fall 2009 (128 students), Fall 2010 (140 students), Spring 2012 (150 students), Spring 2016 (156 students), Spring 2017 (150 students), Spring 2018 (152 students), Fall 2019 (138 students), Fall 2020 (158 students), Fall 2021 (200 students), Fall 2022 (209 students), Fall 2023 (203 students), Fall 2024 (156 students)
- COMS W4111 Database Systems (advanced undergraduate and graduate level), Fall 1999 (121 students), Fall 2000 (174 students), Fall 2001 (133 students), Fall 2002 (82 students), Fall 2003 (93 students), Fall 2005 (73 students), Fall 2006 (62 students), Fall 2007 (51 students)
- COMS W3139 Data Structures and Algorithms (undergraduate level), Fall 1998 (52 students)

- COMS E6113 Advanced Database Systems (graduate level), Spring 1998 (23 students), Spring 1999 (42 students)
- COMS W3203 Discrete Mathematics (undergraduate level), Spring 1998 (76 students)
- COMS E6998 Topics in Digital Libraries (graduate level), Fall 1997 (12 students), co-taught with Dragomir Radev

## Other Educational Activities

- Ph.D. Dissertation Committees: Zachary Huang, Modern Analytics over Wide-Tables, Columbia University, Computer Science Department, April 2025; Giannis Karamanolakis, Efficient Machine Teaching Frameworks for Natural Language Processing, Columbia University, Computer Science Department, July 2022; Lampros Flokas, Complaint-Driven Training Data Debugging for Machine Learning Workflows, Columbia University, Computer Science Department, October 2022; Wangda Zhang, Optimizing Query Processing under Skew, Columbia University, Computer Science Department, August 2020; Fotis Psallidas, Physical Plan Instrumentation in Databases: Design Principles and Applications, Columbia University, Computer Science Department, March 2019; Ioannis Paparrizos, Fast, Scalable, and Accurate Algorithms for Time-Series Analysis, Columbia University, Computer Science Department, January 2018; Orestis Polychroniou, Optimizing Analytical Query Execution Across All Layers of Modern Hardware, Columbia University, Computer Science Department, September 2017; Evangelia Sitaridi, GPU-Acceleration of In-Memory Data Analytics, Columbia University, Computer Science Department, May 2016; Yves Petinot, A Nearest-Neighbor Approach to Indicative Web Summarization, Columbia University, Computer Science Department, February 2016; Grzegorz Drzadzewski, An Online Analytical System for Multi-Tagged Document Collections, University of Waterloo, School of Computer Science, November 2015; Pablo Barrio, Ranking for Scalable Information Extraction, Columbia University, Computer Science Department, September 2015; Sara Rosenthal, Detecting Influencers in Social Media Discussions, Columbia University, Computer Science Department, July 2015; Kristen Parton, Lost and Found in Translation: Cross-Lingual Question Answering with Result Translation, Columbia University, Computer Science Department, September 2012; Hila Becker, Identification and Characterization of Events in Social Media, Columbia University, Computer Science Department, September 2011; Oren Laadan, A Mobile Personal Computer Recorder, Columbia University, Computer Science Department, September 2010; Corey Goldfeder, Data-Driven Grasping, Columbia University, Computer Science Department, August 2010; Julia Stoyanovich, Search and Ranking in Semantically Rich Applications, Columbia University, Computer Science Department, October 2009; John Cieslewicz, Architecture-Sensitive Database Query Processing on Chip Multiprocessors, Columbia University, Computer Science Department, December 2008; Alpa Jain, Query Processing over Relations Extracted from Text Databases, Columbia University, Computer Science Department, May 2008; Wisam Dakka, Faceted Searching and Browsing Over Large Collections of Textual and Text-Annotated Objects, Columbia University, Computer Science Department, May 2008; Knarig Arabshian, Ontology-based Context-aware Service Discovery in a Globally Distributed Network, Columbia University, Computer Science Department, May

2008; Sameer Maskey, Automatic Broadcast News Speech Summarization, Columbia University, Computer Science Department, December 2007; Hassan Malik, Efficient Algorithms for Clustering and Classifying High Dimensional Data Using Interesting Patterns, Columbia University, Computer Science Department, November 2007; Sasha Blair-Goldensohn, Long-Answer Question Answering and Rhetorical-Semantic Relations, Columbia University, Computer Science Department, January 2007; Luo Si, Federated Search of Text Search Engines in Uncooperative Environments, Carnegie Mellon University, School of Computer Science, May 2006; Noemie Elhadad, User-sensitive Text Summarization, Columbia University, Computer Science Department, January 2006; Amélie Marian, Evaluation of Top-k Queries over Structured and Semi-structured Data, Columbia University, Computer Science Department, June 2005; Eugene Agichtein, Extracting Relations from Large Text Collections, Columbia University, Computer Science Department, November 2004; Panagiotis Ipeirotis, Classifying and Searching Hidden-Web Text Databases, Columbia University, Computer Science Department, July 2004; Jingren Zhou, Architecture-Sensitive Database Query Processing, Columbia University, Computer Science Department, May 2004; Nicolás Bruno, Statistics on Query Expressions in Relational Database Management Systems, Columbia University, Computer Science Department, April 2003; Min-Yen Kan, Automatic Text Summarization as Applied to Information Retrieval: Using Indicative and Informative Summaries, Columbia University, Computer Science Department, November 2002; Eleazar Eskin, Sparse Sequence Modeling Applied to Computational Biology and Intrusion Detection, Columbia University, Computer Science Department, April 2002; James Shaw, Clause Combining: An Approach to Generating Concise Text, Columbia University, Computer Science Department, September 2001; Hongyan Jing, Cut and Paste Based Text Summarization, Columbia University, Computer Science Department, April 2001; David W. Fan, Cost-sensitive, Scalable and Adaptive Learning Using Ensemble-based Methods, Columbia University, Computer Science Department, December 2000; Jun Rao, Advanced Query Processing in Databases, Columbia University, Computer Science Department, May 2000; Jingshuang Yang, Extensible Transaction Service for WWW-based Collaborative Systems, Columbia University, Computer Science Department, November 1999; Andreas Prodromidis, Management of Intelligent Learning Agents in Distributed Data Mining Systems, Columbia University, Computer Science Department, September 1999; Michelle Zhou, Automated Generation of Visual Discourse, Columbia University, Computer Science Department, October 1998; Akira Kawaguchi, Implementation Techniques for Materialized Views, Columbia University, Computer Science Department, October 1997

- Ph.D. Thesis Proposal Committees: Zachary Huang, October 2023; Lampros Flokas, November 2021; Giannis Karamanolakis, January 2021; Wangda Zhang, December 2018; Fotis Psallidas, February 2018; Mohammad Sadegh Rasooli, November 2016; Ioannis Paparrizos, May 2016; Orestis Polychroniou, May 2016; Evangelia Sitaridi, November 2014; Pablo Barrio, May 2014; Yves Petinot, October 2012; Hila Becker, December 2009; Corey Goldfeder, March 2008; Julia Stoyanovich, December 2007; John Cieslewicz, December 2007; Sameer Maskey, June 2007; Wisam Dakka, February 2007; Alpa Jain, January 2007; Hassan Malik, April 2006; Sasha Blair-Goldensohn, February 2006; Luo Si, July 2004; Noemie

- Elhadad, March 2004; Amélie Marian, April 2003; Panagiotis Ipeirotis, May 2002; Jingren Zhou, May 2002; Eugene Agichtein, April 2002; Nicolás Bruno, April 2002; Eleazar Eskin, December 2000; Giuseppe Valetto, May 2000; Min-Yen Kan, December 1999; Hongyan Jing, February 1999; David W. Fan, September 1998; Jun Rao, May 1998; Andreas Prodromidis, January 1998; Steve Dossick, December 1997; James Shaw, December 1997
- Ph.D. Candidacy Exam Committees: Matthew Tolles, April 2024; Junyoung Kim, April 2023; Keyang Xu, December 2022; Yiru Chen, December 2022; Zachary Huang, November 2022; Lampros Flokas, May 2020; Giannis Karamanolakis, April 2020; Fotis Psallidas, February 2017; Wangda Zhang, February 2017; Bingyi Cao, December 2016; Ioannis Paparrizos, December 2015; Pablo Barrio, May 2013; Orestis Polychroniou, May 2013; Evangelia Sitaridi, May 2013; John Cieslewicz, April 2006; Julia Stoyanovich, October 2005; Sameer Maskey, August 2004; Alpa Shah, May 2004; Wisam Dakka, May 2004; Elena Filatova, December 2002; Amélie Marian, April 2002; Panagiotis Ipeirotis, April 2001; Jingren Zhou, April 2001; Eugene Agichtein, January 2001; Nicolás Bruno, December 2000; Junyan Ding, May 2000; Junxin Zhang, May 2000; Kazi Zaman, December 1998
  - External Student Supervision: Richard Lichtenstein (undergraduate student at Harvard University), Summer and Fall 2002, towards Harvard University's CS 91r-Supervised Reading and Research
  - Short Courses: Universidad Torcuato Di Tella, Inteligencia Comercial y Data Mining (elective course for MBA program), June 2012 (38 students); Universidad Torcuato Di Tella, Inteligencia Comercial y Data Mining (elective course for MBA program), May-June 2011 (44 students); Universidad Torcuato Di Tella, Data Mining para Business Intelligence, July 2008; Universidad de Buenos Aires, Distributed Databases, July 1994 (approximately 100 students)
  - Teaching Assistantships: Stanford University, Database Implementation (upper-level undergraduate students), Distributed Databases (graduate students), 1993-1994

## University Service

- LGBTQ+ Diversity Roundtable Discussion, University Life's Graduate Initiative for Inclusion and Belonging: Panelist, October 2021
- Research Initiatives in Science and Engineering (RISE) Program, Advisory Board Member: 2014
- Google Ph.D. Fellowship Program, Selection Committee: 2014
- Committee on Dual Career Policies and Resources (under the Vice Provost for Diversity Initiatives), Member, 2005-2006

## Computer Science Department

- New Directions for Curriculum Task Force, Member, 2006-2007
- Faculty Retreat, Co-organizer, October 2001, November 2000
- Ph.D. Admissions Committee, Chair 2023-2024, Chair 2022-2023, Chair 2021-2022, Member 2020-2021, Member 2019-2020, Chair 2017-2018, Chair 2016-2017, Chair 2015-2016, Chair 2011-2012, Chair 2010-2011, Chair 2009-2010, Chair 2008-2009, Chair 2007-2008, Chair

2006-2007, Chair 2005-2006, Chair 2003-2004, Associate Chair 2002-2003, Member 2001-2002, Member 1999-2000, Chair 1998-1999, Member 1997-1998

- M.S. Admissions Committee (application reviewing), 2025, 2024, 2023, 2022, 2021, 2020, 2019, 2018, 2017, 2016, 2015, 2014, 2008-2009
- Dual M.S. in Journalism and Computer Science Admissions Committee, Member 2015, 2014, Member 2011-2012, Member 2010-2011
- Faculty Recruiting Committee, Member 1998-2000, Co-chair 1997-1998
- Colloquium Series Organizer, 1997-1998: Approximately 40 talks scheduled

### **School of General Studies**

- Advisor for Computer Science Majors, 1997-2004, 2005-2012

### **Continuing Education and Special Programs**

- Advisor for Computer Science Majors in the Second-Majors Program, 1997-2004, 2005-2007
- Advisor for Postbaccalaureate Studies Program Computer Science students, 2008-2012

### **Columbia College**

- Advisor for Computer Science Majors: Spring 2015 (Sophomores), 2015-2016 (Juniors), 2016-2017 (Seniors), 2017-2018 (First Year Students), 2019-2020 (First Year Students), 2020-2021 (Sophomores), 2021-2022 (Juniors), 2022-2023 (Seniors), 2023-2024 (Sophomores), 2024-2025 (Juniors)
-