

# Ziv Bar-Yossef – Curriculum Vitae

May 12, 2011

## CONTACT INFORMATION

Address: Department of Electrical Engineering, Technion, Haifa 32000, Israel.  
Tel: +972-54-2451022  
Email: [zivby@ee.technion.ac.il](mailto:zivby@ee.technion.ac.il)  
Homepage: <http://www.ee.technion.ac.il/people/zivby>

## EMPLOYMENT

- 2006 – Staff Research Scientist and Technical Lead Manager  
Google Israel Engineering Center.
  
- 2008 – Adjunct Senior Lecturer  
Department of Electrical Engineering,  
Technion - Israel Institute of Technology, Haifa, Israel.
  
- 2004 – 2008 Senior Lecturer  
Department of Electrical Engineering,  
Technion - Israel Institute of Technology, Haifa, Israel.
  
- 2004 – 2006 Consultant  
ContactSpeech LTD, Tel-Aviv, Israel.
  
- 2002 – 2004 Research Staff Member  
IBM Almaden Research Center, San Jose, CA, USA.
  
- 2000 – 2002 Research Intern  
IBM Almaden Research Center, San Jose, CA, USA.  
Supervisors: Dr. Sridhar Rajagopalan and Dr. D. Sivakumar
  
- 1999 – 2002 Research Assistant and Teaching Assistant  
Computer Science Division,  
University of California at Berkeley, Berkeley, CA, USA.  
Supervisor: Prof. Christos H. Papadimitriou
  
- 1997 – 1999 Research Assistant and Teaching Assistant  
Department of Computer Science,  
Hebrew University, Jerusalem, Israel.  
Supervisors: Prof. Noam Nisan and Prof. Yehoshua Sagiv

1994 – 1997 Software Developer  
Israel Defense Forces, Israel.

1989 – 1993 Military service  
Israel Defense Forces, Israel.

#### EDUCATION

1999 – 2002 PhD in Computer Science, University of California at Berkeley.  
Thesis: *The Complexity of Massive Data Set Computations*.  
Research advisor: Prof. Christos H. Papadimitriou.  
GPA: 3.95/4.0.

1996 – 1998 MSc in Computer Science, Hebrew University, *Cum laude*.  
Thesis: *Deterministic Amplification of Space-Bounded Randomized Algorithms*.  
Research advisor: Prof. Avi Wigderson.  
GPA: 96/100.

1993 – 1996 BSc in Mathematics and Computer Science, Hebrew University, *Summa cum laude*.  
Honors thesis: *Pointer Jumping Requires Concurrent Read*.  
Research advisor: Prof. Noam Nisan.  
GPA: 97/100.

#### HONORS AND AWARDS

- Google OC Award, Honorable Mention, for contribution to Google Instant, 2010.
- Runner up for Best Application Paper Award, Int'l Conference on Data Mining, 2006.
- Best Paper Award, International World-Wide Web Conference, 2006.
- The IBM Faculty Award, 2006.
- The Long Leong Lim Pre-Doctoral Prize, U.C. Berkeley, 2000.
- Regents Fellowship, U.C. Berkeley, 1999.
- The Knesset (Israeli parliament) honor for M.Sc. achievements, 1997.
- Rector's prize for M.Sc. achievements, Hebrew University, 1997.
- Rector's prize for B.Sc. achievements, Hebrew University, 1995.
- Dean's prize for B.Sc. achievements, Hebrew University, 1994.
- The Amirim special honors program, Hebrew University, 1993 – 1996.

#### TEACHING EXPERIENCE

2005–2006 Instructor, 046002: Design and Analysis of Algorithms, Technion.

2005–2006 Instructor, 049011: Algorithms for Large Data Sets, Technion.

2001 Teaching assistant, CS174: Combinatorics and Discrete Probability, U.C. Berkeley.

1998 Teaching assistant, Internet Programming Lab, Hebrew University.

## PHD STUDENTS

- Maxim Gurevich (joint with Idit Keidar), Technion, 2006–2010. Now at Yahoo! Research.  
PhD Thesis: *External Search Engine Mining*.
- Mirit Shalem (joint with Yaron Kanza), Technion, 2007–.

## MSc STUDENTS

- Maxim Gurevich, Technion, 2004–2006.  
MSc Thesis: *Random Sampling from a Search Engine's Index*.
- Uri Schonfeld (joint with Idit Keidar), Technion, 2004–2006.  
MSc Thesis: *Do Not Crawl in the DUST: Different URLs with Similar Text*.
- Ido Guy, Technion, 2004–2007.  
MSc Thesis: *Cluster Ranking with an Application to Mining Mailbox Networks*.
- Mirit Shalem, Technion, 2005–2007.  
Transferred to a direct PhD track.
- Li-Tal Mashiach, Technion, 2006–2008.  
MSc Thesis: *Local Approximation of PageRank and Reverse PageRank*.
- Naama Kraus (joint with Shaul Markovitch) , Technion, 2009–.  
Tranfering to a direct PhD track.

## PUBLICATIONS IN SCIENTIFIC JOURNALS

1. Z. Bar-Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. Information Statistics Approach to Data Stream and Communication Complexity. In *Journal of Computer and System Sciences (JCSS)*, Vol. 68, No. 4, pages 702–732, 2004. (Special issue on FOCS 2002.)
2. Z. Bar-Yossef, M. Fontoura, and V. Josifovski. On the Memory Requirements of Evaluating XPath Queries over XML Streams. In *Journal of Computer and System Sciences (JCSS)*, Vol. 73, No. 3, pages 391–441, 2007. (Special issue on PODS 2004.)
3. Z. Bar-Yossef, I. Guy, R. Lempel, Y. Maarek, and V. Soroka. Cluster Ranking with an Application to Mining Mailbox Networks. In *Knowledge and Information Systems (KAIS)*, Vol. 14, No. 1, pages 101–139, 2008. (Special issue on ICDM 2006.)
4. Z. Bar-Yossef, T.S. Jayram, and J. Kerenidis. Exponential Separation of Quantum and Classical One-Way Communication Complexity. In *SIAM Journal on Computing (SICOMP)*, Vol. 38, No. 1, pages 366–384, 2008.
5. Z. Bar-Yossef, R. Friedman, and G. Kliot. RaWMS - Random Walk based Lightweight Membership Service for Wireless Ad Hoc Networks. In the *ACM Transactions on Computer Systems (TOCS)*, Vol. 26, No. 2, pages 1–66, 2008.

6. Z. Bar-Yossef and M. Gurevich. Random Sampling from a Search Engine’s Index. In the *Journal of the ACM (JACM)*, Vol. 55, No. 5, 2008.
7. Z. Bar-Yossef, I. Keidar, and U. Schonfeld. Do not Crawl in the DUST: Different URLs with Similar Text. In the *ACM Transactions on the Web (TWEB)*, pages 3, 2009.
8. Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol. Index Coding with Side Information. In the *IEEE Transactions on Information Theory*, Vol. 57, No. 3, pages 1479–1494, 2011.

#### PUBLICATIONS IN REFEREED CONFERENCES

1. Z. Bar-Yossef and N. Nisan. Pointer Jumping Requires Concurrent Read. In the *Proceedings of the 29th Annual ACM Symposium on the Theory of Computing (STOC)*, pages 549–558, 1997.
2. Z. Bar-Yossef, O. Goldreich, and A. Wigderson. Deterministic Amplification of Space-Bounded Probabilistic Algorithms. In the *Proceedings of the 14th Annual IEEE Conference on Computational Complexity (CCC)*, pages 188–198, 1999.
3. Z. Bar-Yossef, Y. Kanza, Y. Kogan, W. Nutt, and Y. Sagiv. Querying Semantically Tagged Documents on the World-Wide Web. In the *Proceedings of the 4th Workshop on Next Generation Information Technologies and Systems (NGITS)*, pages 2–19, 1999.
4. Z. Bar-Yossef, A. Berg, S. Chien, J. Fakcharoenphol, and D. Weitz. Approximating Aggregate Queries about Web Pages via Random Walks. In the *Proceedings of the 26th International Conference on Very Large Data Bases (VLDB)*, pages 535–544, 2000.
5. Z. Bar-Yossef, R. Kumar, and D. Sivakumar. Sampling Algorithms: Lower Bounds and Applications. In the *Proceedings of the 33rd Annual ACM Symposium on the Theory of Computing (STOC)*, pages 266–275, 2001.
6. Z. Bar-Yossef, K. Hildrum, and F. Wu. Incentive-Compatible Online Auctions for Digital Goods. In the *Proceedings of 13th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 964–970, 2002.
7. Z. Bar-Yossef, R. Kumar, and D. Sivakumar. Reductions in Streaming Algorithms, with an Application to Counting Triangles in Graphs. In the *Proceedings of the 13th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 623–632, 2002.
8. Z. Bar-Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. Information Theory Methods in Communication Complexity. In the *Proceedings of the 17th Annual IEEE Conference on Computational Complexity (CCC)*, pages 93–102, 2002.
9. Z. Bar-Yossef, O. Reingold, R. Shaltiel, and L. Trevisan. Streaming Computation of Combinatorial Objects. In the *Proceedings of the 17th Annual IEEE Conference on Computational Complexity (CCC)*, pages 165–174, 2002.

10. Z. Bar-Yossef and S. Rajagopalan. Template Detection via Data Mining and its Applications. In the *Proceedings of the 11th International World Wide Web Conference (WWW)*, pages 580–591, 2002.
11. Z. Bar-Yossef, T.S. Jayram, R. Kumar, D. Sivakumar, and L. Trevisan. Counting Distinct Elements in a Data Stream. In the *Proceedings of the 6th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, pages 1–10, 2002.
12. Z. Bar-Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. Information Statistics Approach to Data Stream and Communication Complexity. In the *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 209–218, 2002.
13. Z. Bar-Yossef. Sampling Lower Bounds via Information Theory. In the *Proceedings of the 35th Annual ACM Symposium on the Theory of Computing (STOC)*, pages 335–344, 2003.
14. Z. Bar-Yossef, T.S. Jayram, and J. Kerenidis. Exponential Separation of Quantum and Classical One-Way Communication Complexity. In the *Proceedings of the 36th Annual ACM Symposium on the Theory of Computing (STOC)*, pages 128–137, 2004.
15. Z. Bar-Yossef, A. Broder, R. Kumar, and A. Tomkins. Sic Transit Gloria Telae: Towards an Understanding of the Web’s Decay. In the *Proceedings of the 13th International World Wide Web Conference (WWW)*, pages 328–337, 2004.
16. Z. Bar-Yossef, M. Fontoura, and V. Josifovski. On the Memory Requirements of Evaluating XPath Queries over XML Streams. In the *Proceedings of the 23rd ACM Symposium on Principles of Database Systems (PODS)*, pages 177–188, 2004.
17. Z. Bar-Yossef, T.S. Jayram, R. Krauthgamer, and R. Kumar. The Sketching Complexity of Pattern Matching. In the *Proceedings of the 8th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, pages 261–272, 2004.
18. Z. Bar-Yossef, T.S. Jayram, R. Krauthgamer, and R. Kumar. Approximating Edit Distance Efficiently. In the *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 550–559, 2004.
19. Z. Bar-Yossef, M. Fontoura, and V. Josifovski. Buffering in Query Evaluation over XML Streams. In the *Proceedings of the 24th ACM Symposium on Principles of Database Systems (PODS)*, pages 216–227, 2005.
20. Z. Bar-Yossef and M. Gurevich. Random Sampling from a Search Engine’s Index. In the *Proceedings of the 15th International World-Wide Web Conference (WWW)*, pages 367–376, 2006. **Winner of best paper award.**
21. Z. Bar-Yossef, R. Friedman, and G. Kliot. RaWMS - Random Walk based Lightweight Membership Service for Wireless Ad Hoc Networks. In the *Proceedings of the 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC)*, pages 238–249, 2006.

22. Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol. Index Coding with Side Information. In the *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 197–206, 2006.
23. Z. Bar-Yossef, I. Guy, R. Lempel, Y. Maarek, and V. Soroka. Cluster Ranking with an Application to Mining Mailbox Networks. In the *Proceedings of the 2006 IEEE International Conference on Data Mining (ICDM)*, pages 63–74, 2006. **Runner up for best application paper award.**
24. Z. Bar-Yossef and M. Gurevich. Efficient Search Engine Measurements. In the *Proceedings of the 16th International World-Wide Web Conference (WWW)*, pages 401–410, 2007.
25. Z. Bar-Yossef, I. Keidar, and U. Schonfeld. Do not Crawl in the DUST: Different URLs with Similar Text. In the *Proceedings of the 16th International World-Wide Web Conference (WWW)*, pages 111–120, 2007.
26. M. Shalem and Z. Bar-Yossef. The Space Complexity of Processing XML Twig Queries over Indexed Documents. In the *Proceedings of the 24th International Conference on Data Engineering (ICDE)*, pages 824–832, 2008.
27. Z. Bar-Yossef and M. Gurevich. Mining Search Engine Query Logs via Suggestion Sampling. In the *Proceedings of the 34th International Conference on Very Large Data Bases (VLDB)*, pages 54–65, 2008.
28. Z. Bar-Yossef and L. Mashiach. Local Approximation of PageRank and Reverse PageRank. In the *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM)*, pages 279–288, 2008.
29. Z. Bar-Yossef and M. Gurevich. Estimating the ImpressionRank of Web Pages. In the *Proceedings of the 18th International World-Wide Web Conference (WWW)*, 2009, pages 41–50.
30. Z. Bar-Yossef and N. Kraus. Context-Sensitive Query Auto-Completion. In the *Proceedings of the 20th International World-Wide Web Conference (WWW)*, 2011, pages 107–116.

#### REFEREED POSTERS

1. U. Schonfeld, Z. Bar-Yossef, and I. Keidar. Do not Crawl in the DUST: Different URLs with Similar Text. In the *Proceedings of the 15th International World-Wide Web Conference (WWW)*, pages 1015–1016, 2006.
2. Z. Bar-Yossef and L. Mashiach. Local Approximation of PageRank and Reverse PageRank. In the *Proceedings of the 31st Annual International ACM SIGIR Conference*, pages 865–866, 2008.

## TECHNICAL REPORTS

1. Z. Bar-Yossef, T. Kanungo, and R. Krauthgamer. Focused Sampling: Computing Topical Web Statistics. *Technical Report*, IBM T.J. Watson Research Center, RJ 10339, 2005.
2. Z. Bar-Yossef, R. Friedman, and G. Kliot. RaWMS - Random Walk based Lightweight Membership Service for Wireless Ad Hoc Networks. *Technical Report*, Computer Science Department, Technion, CS-2006-05, 2006.
3. Z. Bar-Yossef and M. Gurevich. Random Sampling from a Search Engine's Corpus. *Technical Report*, Department of Electrical Engineering, Technion, CCIT Report #598, 2006.
4. Z. Bar-Yossef, I. Keidar, and U. Schonfeld. Do not Crawl in the DUST: Different URLs with Similar Text. *Technical Report*, Department of Electrical Engineering, Technion, CCIT Report #601, 2006.
5. Z. Bar-Yossef, I. Guy, R. Lempel, Y. Maarek, and V. Soroka. Cluster Ranking with an Application to Mining Mailbox Networks. *Technical Report*, Department of Electrical Engineering, Technion, CCIT Report #608, 2006.
6. Z. Bar-Yossef and L. Mashiach. Local Approximation of PageRank and Reverse PageRank. *Technical Report*, Department of Electrical Engineering, Technion, CCIT Report #690, 2008.
7. Z. Bar-Yossef and M. Shalem. The Space Complexity of Processing XML Twig Queries Over Indexed Documents. *Technical Report*, Department of Electrical Engineering, Technion, CCIT Report #693, 2008.
8. Z. Bar-Yossef and M. Gurevich. Efficient Search Engine Measurements. *Technical Report*, Department of Electrical Engineering, Technion, CCIT Report #743, 2009.
9. Z. Bar-Yossef and N. Kraus. Context-Sensitive Query Auto-Completion. *Technical Report*, Department of Electrical Engineering, Technion, CCIT Report #779, 2011.

## PATENTS

1. Z. Bar-Yossef, T.S. Jayram, R. Kumar, and R. Krauthgamer. System and Method for Detecting Matches of Small Edit Distance. *Filed patent application*. IBM Almaden Research Center, September 2005.
2. Z. Bar-Yossef and S. Rajagopalan. Method and System for Improving Data Quality in Large Hyperlinked Text Databases Using Pagelets and Templates. Issued as US Patent 6968331, November 2005.
3. M. Fontoura, V. Josifovski, and Z. Bar-Yossef. Running XPath queries over XML streams with incremental predicate evaluation. *Filed patent application*. IBM Almaden Research Center, April 2006.

4. Z. Bar-Yossef, T. Kanungo, and R. Krauthgamer. System, Method and Service for Using a Focused Random Walk to Produce Samples on a Topic from a Collection of Hyper-linked Pages. Issued as US Patent 7640488, December 2009.
5. Z. Bar-Yossef, A. Z. Broder, R. Kumar, and A. Tomkins. Method and Apparatus for Assessing Web Page Decay. Issued as US patent 7818312, October 2010.

#### TALKS AT CONFERENCES AND SEMINARS

1. Symposium on the Theory of Computing (STOC), May 1997.
2. Computer Science Theory Seminar, Hebrew University, June 1997.
3. Computer Science Theory Seminar, Hebrew University, May 1998.
4. Theory Lunch, U.C. Berkeley, February 1999.
5. IEEE Conference on Computational Complexity (CCC), May 1999.
6. IBM Almaden Research Center, March 2000.
7. Theory Lunch, U.C. Berkeley, March 2000.
8. International Conference on Very Large Data Bases (VLDB), September 2000.
9. Computer Science Theory Seminar, Hebrew University, September 2000.
10. Theory of Computing Seminar, IBM Almaden Research Center, May 2001.
11. Stanford Algorithms Seminar (AFLB), Stanford University, June 2001.
12. Microsoft Research, June 2001.
13. Symposium on the Theory of Computing (STOC), July 2001.
14. Computer Science Theory Seminar, Hebrew University, July 2001.
15. IBM Haifa Labs, July 2001.
16. Cryptography and Complexity Seminar, Weizmann Institute, July 2001.
17. Webfountain Seminar, IBM Almaden Research Center, July 2001.
18. Symposium on Discrete Algorithms (SODA), January 2002.
19. Theory Lunch, U.C. Berkeley, February 2002.
20. IEEE Conference on Computational Complexity (CCC), May 2002.
21. Theory seminar, U.C. Berkeley, May 2002.
22. Cryptography and Complexity Seminar, Weizmann Institute, September 2002.
23. Theory of Computing Seminar, IBM Almaden Research Center, February 2003.
24. Theory seminar, U.C. Berkeley, March 2003.
25. Theory lunch, Stanford University, April 2003.
26. Symposium on the Theory of Computing (STOC), June 2003.
27. Yahoo! Inc., September 2003.
28. Department of Computer Science Colloquium, Bar-Ilan University, December 2003.
29. Department of Computer Science Colloquium, Technion, December 2003.
30. Computer Science Seminar, Weizmann Institute of Science, December 2003.
31. Information Systems Seminar, Department of Industrial Engineering & Management, Technion, December 2003.
32. Department of Electrical Engineering Colloquium, Technion, December 2003.



33. Department of Computer Science Colloquium, Tel-Aviv University, December 2003.
34. Department of Computer Science Colloquium, Hebrew University, December 2003.
35. Research Seminar, IBM Haifa Labs, December 2003.
36. Computer Science Seminar, The Interdisciplinary Center, Herzliya, December 2003.
37. IEEE Symposium on Foundations of Computer Science (FOCS), October 2004.
38. Information Processing: From Theory to Practice Workshop, Technion, December 2004.
39. Department of Computer Science Colloquium, Hebrew University, January 2005.
40. Computer Science Theory Seminar, Tel-Aviv University, March 2005.
41. Computer Science Seminar, Ben-Gurion University, April 2005.
42. ACM Symposium on Principles of Database Systems (PODS), June 2005.
43. Information Retrieval Seminar, IBM Haifa Labs, December 2005.
44. Department of Electrical and Computer Engineering Colloquium, Ben-Gurion University, January 2006.
45. Department of Computer Science Colloquium, Technion, April 2006.
46. Computer Science Colloquium, University of Haifa, April 2006.
47. Computer Science Seminar, Weizmann Institute of Science, June 2006.
48. Department of Computer Science Colloquium, Hebrew University, June 2006.
49. Google Tech Talk, Google Campus, August 2006.
50. Workshop on the Law of Search Engines, University of Haifa, December 2006.
51. ClubNet, Department of Electrical Engineering, Technion, January 2007.
52. Web Interactions Workshop, University of Haifa, May 2007.
53. Information Retrieval Seminar, IBM Haifa Labs, June 2007.
54. Information Systems Seminar, Department of Industrial Engineering, Technion, July 2007.
55. Random Walks Seminars Day, Department of Computer Science, Ben-Gurion University, March 2008.
56. The Annual Meeting of the Israeli Statistical Association, Herzliya, June 2008.
57. Collegial colloquium, Jerusalem College of Engineering, March 2009.

#### PROFESSIONAL ACTIVITY

- **Editorial boards:**

- ACM Transactions on Internet Technology (TOIT), Associate Editor, 2009–.

- **Steering committee member:**

- International ACM Conference on Web Search and Data Mining (WSDM). I am the initiator and co-founder of the conference. First meeting was held in Stanford University in February 2008. See <http://www.wsdm2008.org>.

- **Program committees:**

- International World-Wide Web Conference (WWW), Search Track, 2005, 2006, 2007, 2008, 2009 (Vice Chair).

- International Conference on Web Search and Data Mining (WSDM), 2008 (Senior PC), 2009, 2010 (Senior PC), 2011 (Senior PC).
- SIGIR Conference, 2010.
- International Conference on Very Large Data Bases (VLDB), 2009.
- IEEE Symposium on Foundations of Computer Science (FOCS), 2005.
- International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX), 2009.

- **Reviewer:**

- Science Journal.
- Journal of Computer and System Sciences (JCSS).
- SIAM Journal on Computing (SICOMP).
- SIAM Journal on Discrete Mathematics (SIDMA).
- Theoretical Computer Science (TCS).
- Journal of Theory of Computing Systems (TOCS).
- ACM Transactions on Internet Technology (TOIT).
- ACM Transactions on the Web (TWEB).
- World-Wide Web Journal (WWWJ).
- Journal of the American Society for Information Science and Technology (JASIST).
- The Journal of Software – Practice and Experience.
- IEEE Symposium on Foundations of Computer Science (FOCS).
- ACM Symposium on the Theory of Computing (STOC).
- ACM Symposium on Discrete Algorithms (SODA).
- IEEE Conference on Computational Complexity (CCC).
- Symposium on Theoretical Aspects of Computer Science (STACS).
- European Symposium on Algorithms (ESA).
- Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS).
- ACM Symposium on Principles of Database Systems (PODS).
- International Conference on Data Engineering (ICDE).
- ACM Symposium on Principles of Distributed Computing (PODC).
- International Conference on Distributed Computing Systems (ICDCS).
- The US-Israel Binational Science Foundation (BSF).
- The Israel Science Foundation (ISF).

## GRANTS

- The Marie Curie International Re-Integration Grant, European Commission, €80,000, 2006–2007.
- The Israel Science Foundation (ISF), 142,000 NIS per year, 2007–2011.