Benny Kimelfeld

August 2018
Curriculum Vitae

Associate Professor of Computer Science at Technion, Israel

CONTACT INFORMATION

Taub 703, Faculty of Computer Science
Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel
Phone: +972-77-887-5528
Email: bennyk@cs.technion.ac.il
http://benny.net.technion.ac.il

RESEARCH INTERESTS

Data management: information extraction and retrieval; inconsistent, incomplete and uncertain probabilistic data; graph mining. **Natural language processing:** text correction, normalization, and extraction. **Algorithms:** ranked enumeration.

EMPLOYMENT HISTORY

Currently:	Associate Professor at Technion – Israel Institute of Technology
2014-2015:	Computer Scientist at LogicBlox, Inc.
2010–2014:	Research Staff Member at IBM Research – Almaden (Search and Analytics Department)
2008–2010:	Postdoctoral Scholar at IBM Research – Almaden (Theory Group).
2004–2008:	Teaching Assistant at The Hebrew University of Jerusalem, Israel.
2004–2005:	Tutor at The Open University, Israel.

EDUCATION

2005–2008: Ph.D. in *Computer Science,* The Hebrew University of Jerusalem
- Research advisor: Yehoshua Sagiv
- Thesis topic: *Querying Paradigms for the Web*
- *Summa cum laude*

2003–2004: M.Sc. in *Computer Science,* The Hebrew University of Jerusalem
- Research advisor: Yehoshua Sagiv
- Thesis topic: *Interconnection Semantics for XML*
- Thesis grade: 100 (cumulative average: 98)
- *Summa cum laude*

2000–2002: B.Sc. in *Computer Science* (major) and *Mathematics* (minor), The Hebrew University of Jerusalem
- Cumulative average: 99.12
- *Summa cum laude*

PUBLIC PROFESSIONAL ACTIVITIES

Leading Organization

- Program committee chair of the *International Conference on Database Theory* (ICDT), 2018
- Group chair at ACM SIGMOD 2017
- Co-chair of ACM SIGMOD 2016 workshop on *Web and Databases* (WebDB 2016)
- Co-chair of ACM SIGMOD/PODS 2014 workshop on *Big Uncertain Data* (BUDA 2014)

Journal Editorial Boards

- Associate Editor of the *Journal of Computer and System Sciences* (JCSS), since 2014.

AWARDS AND RECOGNITION

- The ACM SIGMOD Research Highlight Award, 2018.
- Taub Leaders in Science Fellowship, 2015.
- Outstanding Reviewer Award in WebDB 2015.
- Invited tutorial speaker for the 2014 ACM Symposium on Principles of Database Systems (PODS).
- Extended scholarship program for excellent doctorate students of The Hebrew University, 2007–2008.
- The Dimitris N. Chorafas Foundation Prize for research on Data Extraction on the Internet, 2006.
- Leibniz student fellowship, 2005–2007.
- Dean's prize, The Hebrew University, 2002.
- The Israeli-Knesset Prize for academic excellence, 2001 and 2004.
- Rector's prize, The Hebrew University, 2001 and 2004.

RESEARCH GRANTS

- **January 2017 – December 2017:** German-Israeli Foundation (GIF), *Extending Databases with Fundamentals of Text Extraction.* PI: Benny Kimelfeld. Amount: EURO 15,000.
- **October 2015 – October 2019:** Israel Science Foundation, *Incorporating Priorities in the Management of Inconsistent Databases.* PI: Benny Kimelfeld. Amount: NIS 960,000.
- **October 2015 – October 2017:** United States – Israel Binational Science Foundation (BSF), *Aggregation Methods for Partial Preferences.* PIs: Benny Kimelfeld and Julia Stoyanovich. Amount: USD 75,000.
- **October 2015 – October 2017:** Israel Science Foundation, *Equipment Grant.* PI: Benny Kimelfeld. Amount: NIS 305,000.

INVITED TALKS

- From Keywords to SQL: Approaches to Search over Structured Data. In *WebDam-MoDaS Workshop*, Eilat, Israel, 2012.
- Database Foundations in Information Extraction. In *ACM SIGMOD/PODS*, Snowbird, Utah, USA, 2014.
- On Approaches to Search over Structured Data. In *SIGMOD Workshop on Automatic Creation and Curation of Knowledge Bases*, Snowbird, Utah, USA, 2014.
- Foundational Research Propelled by Text Analytics. In *The Jornadas Chilenas de Computación (JCC)*, Talca, Chile, 2014.
- Foundations of Database Systems for Text Analytics. In *L'entreposage de données et l'analyse en ligne (EDA)*, Brussels, Belgium, 2015.
- Extending Datalog Intelligence. In *International Conference on Web Reasoning and Rule Systems (RR)*, Berlin, Germany, 2015.
- Uncertainty and Statistics in Foundations of Data Management. In *Dagstuhl Perspectives Workshop on Foundations of Data Management*, Wadern, Germany, 2016.
- Database Principles in Text Analytics. In *Ronald Fagin Special Event at ACM SIGMOD/PODS*, San Francisco, CA, USA, 2016.
- Database Foundations for Modern Data. In *International Conference on Big-Data and Cloud Computing (ICBDCC'17)*, Coimbatore, Tamilnadu, India, 2017.

PATENTS

Published

- Tyler Baldwin, Howard Ho, Benny Kimelfeld, Yunyao Li, Congle Zhang. *Adaptive Parser-Centric Text Normalization.* Filed Dec. 2013. Published Jul. 2015.
- Benny Kimelfeld, Shivakumar Vaithyanathan. *Unsupervised Learning of Deep Patterns for Semantic Parsing.* Filed Aug. 2013. Published Feb. 2015.
- Benny Kimelfeld, David P. Woodruff. *Estimating the Total Sales over Streaming Bids.* Filed Sep. 2013. Published Nov. 2014.
- Benny Kimelfeld, Yunyao Li, Shivakumar Vaithyanathan. *SVO-Based Taxonomy-Driven Exhaustive Text Analytics.* Filed Sep. 2013. Published May. 2014.
- Zhuowei Bao, Benny Kimelfeld, Yunyao Li. *Automatic Suggestion of Query-Rewrite Rules.* Filed May 2013. Published Nov. 2014.
- Benny Kimelfeld, Yunyao Li, Huahai Yang. *Improving Search Quality via Query Provenance Visualization.* Filed May 2012. Published Dec. 2013.

Filed

- Benny Kimelfeld, Yunyao Li, Christopher R. Palmer, Jerome M. Pesenti. *Domain Centric Natural Language Query Answering.* Filed Oct. 2015.

Journal Articles

[1] Vince Bárány, Balder ten Cate, Benny Kimelfeld, Dan Olteanu, and Zografoula Vagena. Declarative probabilistic programming with datalog. *ACM Trans. Database Syst.*, 42(4):22:1–22:35, 2017.

[2] Ronald Fagin, Benny Kimelfeld, Frederick Reiss, and Stijn Vansummeren. Declarative cleaning of inconsistencies in information extraction. *ACM Trans. Database Syst.*, 41(1):6, 2016.

[3] Ronald Fagin, Benny Kimelfeld, Frederick Reiss, and Stijn Vansummeren. Document spanners: A formal approach to information extraction. *J. ACM*, 62(2):12, 2015.

[4] Benny Kimelfeld and Christopher Ré. Transducing markov sequences. *J. ACM*, 61(5):32, 2014.

[5] Benny Kimelfeld and Phokion G. Kolaitis. The complexity of mining maximal frequent subgraphs. *ACM Trans. Database Syst.*, 39(4):32, 2014.

[6] Marie Jacob, Benny Kimelfeld, and Julia Stoyanovich. A system for management and analysis of preference data. *PVLDB*, 7(12):1255–1258, 2014.

[7] Benny Kimelfeld, Jan Vondrák, and David P. Woodruff. Multi-tuple deletion propagation: Approximations and complexity. *PVLDB*, 6(13):1558–1569, 2013.

[8] Benny Kimelfeld, Jan Vondrák, and Ryan Williams. Maximizing conjunctive views in deletion propagation. *ACM Transactions on Database Systems (TODS)*, 37(4):24, 2012.

[9] Konstantin Golenberg, Benny Kimelfeld, and Yehoshua Sagiv. Optimizing and parallelizing ranked enumeration. *PVLDB*, 4(11):1028–1039, 2011.

[10] Ronald Fagin, Benny Kimelfeld, and Phokion G. Kolaitis. Probabilistic data exchange. *Journal of the ACM (JACM)*, 58(4):15, 2011.

[11] Benny Kimelfeld, Yuri Kosharovsky, and Yehoshua Sagiv. Query evaluation over probabilistic XML. *The VLDB Journal*, 18(5):1117–1140, 2009.

[12] Sara Cohen, Benny Kimelfeld, and Yehoshua Sagiv. Incorporating constraints in probabilistic XML. *ACM Transactions on Database Systems (TODS)*, 34(3), 2009.

[13] Serge Abiteboul, Benny Kimelfeld, Yehoshua Sagiv, and Pierre Senellart. On the expressiveness of probabilistic XML models. *The VLDB Journal*, 18(5):1041–1064, 2009.

[14] Benny Kimelfeld and Yehoshua Sagiv. Efficiently enumerating results of keyword search over data graphs. *Information Systems*, 33(4-5):335–359, 2008.

[15] Sara Cohen, Benny Kimelfeld, and Yehoshua Sagiv. Generating all maximal induced subgraphs for hereditary and connected-hereditary graph properties. *Journal of Computer and System Sciences (JCSS)*, 74(7):1147–1159, 2008.

Conference Papers (Fully Reviewed)

[16] Ester Livshits, Benny Kimelfeld, and Sudeepa Roy. Computing optimal repairs for functional dependencies. In *PODS*, pages 225–237. ACM, 2018.

[17] Benny Kimelfeld, Phokion G. Kolaitis, and Julia Stoyanovich. Computational social choice meets databases. In *IJCAI*, pages 317–323. ijcai.org, 2018.

[18] Batya Kenig, Lovro Ilijasic, Haoyue Ping, Benny Kimelfeld, and Julia Stoyanovich. Probabilistic inference over repeated insertion models. In *AAAI*. AAAI Press, 2018.

[19] Dominik D. Freydenberger, Benny Kimelfeld, and Liat Peterfreund. Joining extractions of regular expressions. In *PODS*, pages 137–149. ACM, 2018.

[20] Uzi Cohen, Batya Kenig, Haoyue Ping, Benny Kimelfeld, and Julia Stoyanovich. A query engine for probabilistic preferences. In *SIGMOD*, pages 1509–1524. ACM, 2018.

[21] Ester Livshits and Benny Kimelfeld. Counting and enumerating (preferred) database repairs. In *PODS*, pages 289–301. ACM, 2017.

[22] Benny Kimelfeld and Christopher Ré. A relational framework for classifier engineering. In *PODS*, pages 5–20. ACM, 2017.

[23] Benny Kimelfeld, Ester Livshits, and Liat Peterfreund. Detecting ambiguity in prioritized database repairing. In *ICDT*, volume 68 of *LIPIcs*, pages 17:1–17:20. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017.

[24] Batya Kenig, Benny Kimelfeld, Haoyue Ping, and Julia Stoyanovich. Querying probabilistic preferences in databases. In *PODS*, pages 21–36. ACM, 2017.

[25] Batya Kenig, Benny Kimelfeld, Haoyue Ping, and Julia Stoyanovich. A database framework for probabilistic preferences. In *AMW*, volume 1912 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2017.

[26] Oren Kalinsky, Yoav Etsion, and Benny Kimelfeld. Flexible caching in trie joins. In *EDBT*, pages 282–293. OpenProceedings.org, 2017.

[27] Nofar Carmeli, Batya Kenig, and Benny Kimelfeld. Efficiently enumerating minimal triangulations. In *PODS*, pages 273–287. ACM, 2017.

[28] Adi Omari, Benny Kimelfeld, Eran Yahav, and Sharon Shoham. Lossless separation of web pages into layout code and data. In *KDD*, pages 1805–1814. ACM, 2016.

[29] Benny Kimelfeld, Ester Livshits, and Liat Peterfreund. Recognizing determinism in prioritized repairing of inconsistent databases. In *AMW*, volume 1644 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2016.

[30] Nofar Carmeli, Batya Kenig, and Benny Kimelfeld. On the enumeration of tree decompositions. In *AMW*, volume 1644 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2016.

[31] Vince Bárány, Balder ten Cate, Benny Kimelfeld, Dan Olteanu, and Zografoula Vagena. Declarative probabilistic programming with datalog. In *ICDT*, volume 48 of *LIPIcs*, pages 7:1–7:19. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016.

[32] Benny Kimelfeld and Christopher Ré. A database framework for classifier engineering. In *Alberto Mendelzon International Workshop (AMW)*, 2015.

[33] Ronald Fagin, Benny Kimelfeld, and Phokion G. Kolaitis. Dichotomies in the complexity of preferred repairs. In *PODS*, pages 3–15. ACM, 2015.

[34] Balder Ten Cate, Benny Kimelfeld, and Dan Olteanu. PPDL: Probabilistic programming with Datalog. In *Alberto Mendelzon International Workshop (AMW)*, 2015.

[35] Devora Berlowitz, Sara Cohen, and Benny Kimelfeld. Efficient enumeration of maximal k-plexes. In *SIGMOD*, pages 431–444. ACM, 2015.

[36] Molham Aref, Balder ten Cate, Todd J. Green, Benny Kimelfeld, Dan Olteanu, Emir Pasalic, Todd L. Veldhuizen, and Geoffrey Washburn. Design and implementation of the logicblox system. In *SIGMOD*, pages 1371–1382. ACM, 2015.

[37] Molham Aref, Benny Kimelfeld, Emir Pasalic, and Nikolaos Vasiloglou. Extending datalog with analytics in LogicBlox. In *Alberto Mendelzon International Workshop (AMW)*, 2015.

[38] Ronald Fagin, Benny Kimelfeld, Frederick Reiss, and Stijn Vansummeren. Cleaning inconsistencies in information extraction via prioritized repairs. In *PODS*, pages 164–175. ACM, 2014.

[39] Congle Zhang, Tyler Baldwin, Howard Ho, Benny Kimelfeld, and Yunyao Li. Adaptive parser-centric text normalization. In *ACL (1)*, pages 1159–1168. The Association for Computer Linguistics, 2013.

[40] Benny Kimelfeld and Yehoshua Sagiv. Extracting minimum-weight tree patterns from a schema with neighborhood constraints. In *ICDT*, pages 249–260. ACM, 2013.

[41] Benny Kimelfeld and Phokion G. Kolaitis. The complexity of mining maximal frequent subgraphs. In *PODS*, pages 13–24. ACM, 2013.

[42] Ronald Fagin, Benny Kimelfeld, Frederick Reiss, and Stijn Vansummeren. Spanners: a formal framework for information extraction. In *PODS*, pages 37–48, 2013.

[43] Sara Cohen, Lior Ebel, and Benny Kimelfeld. A social network database that learns how to answer queries. In *CIDR*. www.cidrdb.org, 2013.

[44] Benny Kimelfeld. A dichotomy in the complexity of deletion propagation with functional dependencies. In *PODS*, pages 191–202. ACM, 2012.

[45] Zhuowei Bao, Benny Kimelfeld, and Yunyao Li. Automatic suggestion of query-rewrite rules for enterprise search. In *SIGIR*, pages 591–600. ACM, 2012.

[46] Benny Kimelfeld, Jan Vondrák, and Ryan Williams. Maximizing conjunctive views in deletion propagation. In *PODS*, pages 187–198. ACM, 2011.

[47] Benny Kimelfeld and Yehoshua Sagiv. Finding a minimal tree pattern under neighborhood constraints. In *PODS*, pages 235–246. ACM, 2011.

[48] Ronald Fagin, Benny Kimelfeld, Yunyao Li, Sriram Raghavan, and Shivakumar Vaithyanathan. Rewrite rules for search database systems. In *PODS*, pages 271–282. ACM, 2011.

[49] Sara Cohen, Benny Kimelfeld, Georgia Koutrika, and Jan Vondrák. On principles of egocentric person search in social networks. In *VLDS*, volume 880 of *CEUR Workshop Proceedings*, pages 3–6. CEUR-WS.org, 2011.

[50] Zhuowei Bao, Benny Kimelfeld, and Yunyao Li. A graph approach to spelling correction in domain-centric search. In *ACL*, pages 905–914. The Association for Computer Linguistics, 2011.

[51] Benny Kimelfeld and Christopher Ré. Transducing markov sequences. In *PODS*, pages 15–26. ACM, 2010.

[52] Ronald Fagin, Benny Kimelfeld, Yunyao Li, Sriram Raghavan, and Shivakumar Vaithyanathan. Understanding queries in a search database system. In *PODS*, pages 273–284. ACM, 2010.

[53] Ronald Fagin, Benny Kimelfeld, and Phokion Kolaitis. Probabilistic data exchange. In *ICDT*, pages 76–88. ACM, 2010.

[54] Sara Cohen and Benny Kimelfeld. Querying parse trees of stochastic context-free grammars. In *ICDT*, pages 62–75. ACM, 2010.

[55] Sara Cohen, Benny Kimelfeld, and Yehoshua Sagiv. Running tree automata on probabilistic XML. In *PODS*, pages 227–236. ACM, 2009.

[56] Foto Afrati, Rada Chirkova, Manolis Gergatsoulis, Vassia Pavlaki, Benny Kimelfeld, and Yehoshua Sagiv. On rewriting XPath queries using views. In *EDBT*, pages 168–179. ACM, 2009.

[57] Benny Kimelfeld and Yehoshua Sagiv. Revisiting redundancy and minimization in an XPath fragment. In *EDBT*, volume 261 of *ACM International Conference Proceeding Series*, pages 61–72. ACM, 2008.

[58] Benny Kimelfeld, Yuri Kosharovsky, and Yehoshua Sagiv. Query efficiency in probabilistic XML models. In *SIGMOD*, pages 701–714. ACM, 2008.

[59] Konstantin Golenberg, Benny Kimelfeld, and Yehoshua Sagiv. Keyword proximity search in complex data graphs. In *SIGMOD*, pages 927–940. ACM, 2008.

[60] Sara Cohen, Benny Kimelfeld, and Yehoshua Sagiv. Incorporating constraints in probabilistic XML. In *PODS*, pages 109–118. ACM, 2008.

[61] Benny Kimelfeld and Yehoshua Sagiv. Maximally joining probabilistic data. In *PODS*, pages 303–312. ACM, 2007.

[62] Benny Kimelfeld and Yehoshua Sagiv. Matching twigs in probabilistic XML. In *VLDB*, pages 27–38. ACM, 2007.

[63] Benny Kimelfeld and Yehoshua Sagiv. Combining incompleteness and ranking in tree queries. In *ICDT*, volume 4353 of *Lecture Notes in Computer Science*, pages 329–343. Springer, 2007.

[64] Benny Kimelfeld and Yehoshua Sagiv. Twig patterns: From XML trees to graphs. In *WebDB*, 2006.

[65] Benny Kimelfeld and Yehoshua Sagiv. Incrementally computing ordered answers of acyclic conjunctive queries. In *Next Generation Information Technologies and Systems (NGITS), Sixth International Workshop*, volume 4032 of *Lecture Notes in Computer Science*, pages 33–38. Springer, 2006.

[66] Benny Kimelfeld and Yehoshua Sagiv. Finding and approximating top-k answers in keyword proximity search. In *PODS*, pages 173–182. ACM, 2006.

[67] Benny Kimelfeld, Eitan Kovacs, Yehoshua Sagiv, and Dan Yahav. Using language models and the HITS algorithm for XML retrieval. In *Comparative Evaluation of XML Information Retrieval Systems, Fifth International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX)*, volume 4518 of *Lecture Notes in Computer Science*, pages 253–260. Springer, 2006.

[68] Sara Cohen, Itzhak Fadida, Yaron Kanza, Benny Kimelfeld, and Yehoshua Sagiv. Full disjunctions: Polynomial-delay iterators in action. In *VLDB*, pages 739–750. ACM, 2006.

[69] Benny Kimelfeld and Yehoshua Sagiv. Efficiently enumerating results of keyword search. In *DBPL*, volume 3774 of *Lecture Notes in Computer Science*, pages 58–73. Springer, 2005.

[70] Benny Kimelfeld and Yehoshua Sagiv. Efficient engines for keyword proximity search. In *WebDB*, pages 67–72, 2005.

[71] Sara Cohen, Yaron Kanza, Benny Kimelfeld, and Yehoshua Sagiv. Interconnection semantics for keyword search in XML. In *CIKM*, pages 389–396. ACM, 2005.

Demo Papers (Fully Reviewed)

[72] Tal Yahav, Oren Kalinsky, Oren Mishali, and Benny Kimelfeld. elinda: Explorer for linked data. In *EDBT*, pages 658–661. OpenProceedings.org, 2018.

[73] Zhuowei Bao, Benny Kimelfeld, Yunyao Li, Sriram Raghavan, and Huahai Yang. Gumshoe quality toolkit: administering programmable search. In *CIKM*, pages 2716–2718. ACM, 2012.

[74] Hilit Achiezra, Konstantin Golenberg, Benny Kimelfeld, and Yehoshua Sagiv. Exploratory keyword search on data graphs. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD Conferece)*, pages 1163–1166. ACM, 2010.

[75] Benny Kimelfeld, Yehoshua Sagiv, and Gidi Weber. ExQueX: Exploring and querying XML documents. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD Conferece)*, pages 1103–1106. ACM, 2009.

Book Chapters

[76] Benny Kimelfeld and Pierre Senellart. Probabilistic XML: Models and complexity. In Zongmin Ma and Li Yan, editors, *Advances in Probabilistic Databases for Uncertain Information Management*, pages 39–66. Springer-Verlag, May 2013.

[77] Sara Cohen, Benny Kimelfeld, and Georgia Koutrika. A survey on proximity measures for social networks. In *SeCO Book*, volume 7538 of *Lecture Notes in Computer Science*, pages 191–206. Springer, 2012.

[78] Benny Kimelfeld and Yehoshua Sagiv. Efficient engines for keyword proximity search. In Ravi Kumar and Jain Bandamutha, editors, *Dynamics of Search Engines: An Introduction*, pages 13–30. The Icfai University Press, 2007.

Tutorials

[79] Avigdor Gal and Benny Kimelfeld. Entity resolution in the Big Data era: Probabilistic DB support to entity resolution, 2015. A tutorial given in *18th International Conference on Extending Database Technology (EDBT)*.

[80] Benny Kimelfeld. Database principles in information extraction. In *33rd ACM Symposium on Principles of Database Systems (PODS)*, 2014. Invited tutorial.

[81] Sara Cohen, Benny Kimelfeld, and Yehoshua Sagiv. Enumerating large query results, 2009. A tutorial given in *25th International Conference on Data Engineering (ICDE)*.

Review Articles (Published)

[82] Benny Kimelfeld. Technical perspective: Optimizing tree patterns for querying graph- and tree-structured data. *SIGMOD Record*, 46(1):14, 2017.

[83] Benny Kimelfeld. Extending datalog intelligence. In *RR*, volume 9209 of *Lecture Notes in Computer Science*, pages 1–10. Springer, 2015.

[84] Ronald Fagin, Benny Kimelfeld, Frederick Reiss, and Stijn Vansummeren. A relational framework for information extraction. *SIGMOD Record*, 44(4):5–16, December 2015.

[85] Benny Kimelfeld. Database principles in information extraction. In *PODS*, pages 156–163. ACM, 2014.

[86] Oktie Hassanzadeh, Anastasios Kementsietsidis, Benny Kimelfeld, Rajasekar Krishnamurthy, Fatma Ozcan, and Ippokratis Pandis. Next generation data analytics at IBM Research. *PVLDB*, 6(11):1174–1175, 2013.

[87] Benny Kimelfeld and Yehoshua Sagiv. Modeling and querying probabilistic XML data. *SIGMOD Record*, 33(4-5):335–359, 2008.