

Chen Li

Curriculum Vitae, as of April 12, 2023

Professor, Department of Computer Science, UC Irvine, CA 92697-3435

<https://chenli.ics.uci.edu/>

Education

- **1996 – 2001: Stanford University, Ph.D. in Computer Science.** Thesis: “Query Processing and Optimization in Information-Integration Systems.” Advisor: Professor Jeffrey D. Ullman.
- **1994 – 1996: Tsinghua University, M.S. in Computer Science.** Dissertation: “A Communication Interface based on SCI.” Advisor: Professor Meiming Shen.
- **1989 – 1994: Tsinghua University, B.S. in Computer Science.** Minor in Enterprise Management

Research

Data management, databases, information systems, search, data-intensive computing, ML-centric systems and data analytics, data integration, data quality, data visualization.

Experiences

- 7/2020 – current: Director of the ICS Master of Computer Science (MCS), UCI.
- 11/2012 – current: Professor, Department of Computer Science, UCI.
- 7/2007 – 11/2012: Associate Professor, Department of Computer Science, UCI.
- 10/2001 – 7/2007: Assistant Professor, Department of Computer Science, UCI.
- 8/2008 – 12/2014: Founder/CTO, SRCH2 Software Inc., Irvine, CA. Started the company specialized in information access for enterprises. Took a leave from UCI from 7/2013 to 12/2014.
- 9/2005–12/2007: Visiting Research Scientist, Google, Santa Monica/Irvine, CA. Full time before 12/31/2005, part time after 1/1/2006.
- 1996 – 2001: Graduate Student Researcher, Computer Science, Stanford University. Focused on database research, including data integration, web databases, data warehouses, and multimedia databases.
- 2000: Supplemental Research Associate, IBM Almaden Research Center. Researched on a web-based database system.
- 2000: Consultant, Common Object Inc., San Mateo, CA. Participated in the design and implementation of a corporate database system using data-integration technologies.
- Summer 1998: Research Intern, Hewlett Packard Labs, Palo Alto, CA. Ported a well-known main-memory database system to a shared-memory distributed system.
- 5/1995 - 9/1995: Student Visitor, Chinese University of Hong Kong. Studied how to reduce communication cost using a simplified transfer protocol in parallel processing.
- 1993 – 1996: Research Assistant, Tsinghua University. Researched on parallel and distributed processing and high-performance networking.
- 1996 – 2001: Research Assistant, Computer Science, Stanford University.

Recent Professional activities

- IEEE Fellow (since 1/1/2023)
- General Chair, ICDE 2023 in Anaheim, CA.
- ACM distinguished member (since 2018)
- Treasurer and board member of the VLDB Endowment, since 1/2020.
- Co-Chair, VLDB 2021 Industrial Track
- Committee members: recent SIGMOD and VLDB conferences
- Liaison for VLDB 2019 industrial sponsors
- Tutorial co-chair for Big Data 2019
- July 2016 – 2017: ICDE 2017 Area Chair
- Oct. 2015 – Sept. 2017: PVLDB advisory committee member
- Co-PC chair of VLDB2015

Awards and Funding

- 2022: NSF grant titled “An End-to-End Pandemic Early Warning System by Harnessing Open-source Intelligence.”
- 2021: NSF grant titled “Collaborative Research: III: Medium: Collaborative Machine-Learning-Centric Data Analytics at Scale.”
- 2019: NSF grant titled “CCRI: ENS: Collaborative Research: Supporting and Sustaining Apache AsterixDB for the CISE Research Community”.
- 2019: Research grant from Orange County Health Care Agency (OCHCA)
- 2018: NIH subawards on social media analysis
- 2018: Research grant from Orange County Health Care Agency (OCHCA)
- 2018: Research grant from Alibaba
- 2017 NSF EAGER grant.
- 2017: ICDE Best Demo Award.
- 2016, Research fund from ARL (Army Research Lab)
- 2013, NSF grant 1305430 titled “CI-ADDO-NEW: Asterix: A Community Software Platform for Big Data Research, Analysis, and Management.”
- 2013, Gift funds from HTC.
- 2013, DASFAA 10-year Best Paper Award.
- 2012, ACM SIGMOD 10-year Test-of-Time award.
- 2012, UCI ICS Dean’s Award for Graduate Student Mentoring.
- 2011, the third place at the Microsoft Speller Challenge competition.
- 2010, NIH grant (\$372,360) to support research on the iPubmed system
- 2010, Intel grant on compression of human genome data.
- 2010, NSF research award IIS 1030002 on powerful keyword search in a cloud-computing infrastructure.
- 2009, NSF research award 0910989 titled "DC: Large: Collaborative Research: ASTERIX: A Highly Scalable Parallel Platform for Semistructured Data Management and Analysis."

- 2009, NSF grant 0844574 titled "Supporting efficient fuzzy queries on large text repositories using Hadoop."
- 2009, UCI ICS Dean's Award for excellence in Mid-Career Research, 2009.
- 2007, NSF Grant titled "SGER: Answering Approximate String Queries Using Variable-Length Grams"
- 2006, ICS Ted & Janice Smith Faculty Seed Fund
- 2006, Google Research Award, \$37.5K, renewable for a 2nd year.
- 2006: Gift Fund from Microsoft Research (7K).
- 2005: UCI Career Development Award.
- 2003 NSF CAREER Award, \$400K, PI, starting 9/1/2003, for 5 years.
- 2003 NSF ITR Award 0331707, "Responding to the Unexpected," \$12.5M for 5 years, Senior Investigator, starting 10/1/2003.
- Single Investigator Innovation Grant: CORCLR, UC Irvine, May 2003.
- Stanford Graduate Fellowship, Stanford University, 1997 – 2001.
- Merit-Based Scholarships, Tsinghua University, China, 1994 – 1996.
- Outstanding Student Award, Tsinghua University, June 1994.
- Entrance exams waived, Tsinghua University, 1994 and 1989.

Patents

- Li, C., Wang, B., Yang, X., Patent, "IMPROVING PERFORMANCE OF APPROXIMATE STRING QUERIES USING VARIABLE-LENGTH HIGH-QUALITY GRAMS", US Patent 7996369, <http://www.google.com/patents/US7996369>
- Li, C., Li, G., Ji, S., Wang, J., Feng, J., Patent, "SUPPORTING INTERACTIVE, FUZZY SEARCH ON STRUCTURED DATA EFFICIENTLY", US Patent 8631035, <http://www.google.com/patents/US8631035>

Publications

1. Tempura: a general cost-based optimizer framework for incremental data processing (Journal Version), Zuozhi Wang, Kai Zeng, Botong Huang, Wei Chen, Xiaozong Cui, Bo Wang, Ji Liu, Liya Fan, Dachuan Qu, Zhenyu Hou, Tao Guan, Chen Li, and Jingren Zhou, VLDB Journal 2023.
2. Fries: Fast and Consistent Runtime Reconfiguration in Dataflow Systems with Transactional Guarantees, Zuozhi Wang, Shengquan Ni, Avinash Kumar, Chen Li, to appear in VLDB 2023.
3. GSViz: Progressive Visualization of Geospatial Influences in Social Networks, Sadeem Alsudais, Qiushi Bai, Shuang Zhao, Chen Li, to appear in SIGSPATIAL 2022.
4. Maliva: Using Machine Learning to Rewrite Visualization Queries Under Time Constraints, Qiushi Bai, Sadeem Alsudais, Chen Li, Shuang Zhao, to appear in EDBT 2023.

5. Demonstration of Collaborative and Interactive Workflow-Based Data Analytics in Texera, Xiaozhen Liu, Zuozhi Wang, Shengquan Ni, Sadeem Alsudais, Yicong Huang, Avinash Kumar, Chen Li, VLDB 2022.
6. Demonstration of Accelerating Machine Learning Inference Queries with Correlative Proxy Models, Zhihui Yang, Yicong Huang, Zuozhi Wang, Feng Gao, Yao Lu, Chen Li, X. Sean Wang, VLDB 2022.
7. Optimizing Machine Learning Inference Queries with Correlative Proxy Models, Zhihui Yang, Zuozhi Wang, Yicong Huang, Yao Lu, Chen Li, and X. Sean Wang, VLDB 2022.
8. Rainbow: A Rendering-Aware Index for High-Quality Spatial Scatterplots with Result-Size Budgets, Qiushi Bai, Sadeem Alsudais, Chen Li, Shuang Zhao, EGPGV 2022.
9. Demo of VisBooster: Accelerating Tableau Live Mode Queries Up to 100 Times Faster, Qiushi Bai, Sadeem Alsudais, Chen Li, EDBT/ICDT Workshops 2022.
10. JEDI: These aren't the JSON documents you're looking for..., Thomas Hütter, Nikolaus Augsten, Christoph Kirsch, Michael Carey, and Chen Li, SIGMOD 2022.
11. The Social Amplification and Attenuation of COVID-19 Risk Perception Shaping Mask Wearing Behavior: A Longitudinal Twitter Analysis, Suellen Hopfer, Emilia J. Fields, Yuwen Lu, Ganesh Ramakrishnan, Ted Grover, Quishi Bai, Yicong Huang, Chen Li, and Gloria Mark, PLOS ONE, 2021.
12. Why Do People Oppose Mask Wearing? A Comprehensive Analysis of US Tweets During the COVID-19 Pandemic, Lu He, Changyang He, Tera Leigh Reynolds, Qiushi Bai, Yicong Huang, Chen Li, Kai Zheng, and Yunan Chen. JAMIA 2021.
13. Tempura: A General Cost-Based Optimizer Framework for Incremental Data Processing, Zuozhi Wang, Kai Zeng, Botong Huang, Wei Chen, Xiaozong Cui, Bo Wang, Ji Liu, Liya Fan, Dachuan Qu, Zhenyu Hou, Tao Guan, Chen Li, Jingren Zhou, VLDB 2021.
14. Grosbeak: A Data Warehouse Supporting Resource-Aware Incremental Computing, Zuozhi Wang, Kai Zeng, Botong Huang, Wei Chen, Xiaozong Cui, Bo Wang, Ji Liu, Liya Fan, Dachuan Qu, Zhenyu Hou, Tao Guan, Chen Li, Jingren Zhou, SIGMOD 2020 (Demo).
15. Demonstration of Interactive Runtime Debugging of Distributed Dataflows in Texera, Zuozhi Wang, Avinash Kumar, Shengquan Ni, and Chen Li, VLDB 2020 (Demo).
16. Amber: A Debuggable Dataflow System Based on the Actor Model, Avinash Kumar, Zuozhi Wang, Shengquan Ni, and Chen Li, VLDB 2020.
17. Marviq: Quality-Aware Geospatial Visualization of Range-Selection Queries Using Materialization, Liming Dong, Qiushi Bai, Taewoo Kim, Taiji Chen, Weidong Liu, Chen Li, ACM SIGMOD, 2020.
18. Similarity query support in big data management systems, Taewoo Kim, Wenhai Li, Alexander Behm, Inci Cetindil, Rares Vernica, Vinayak Borkar, Michael J. Carey, Chen Li, Information Systems, Volume 88, February 2020.
19. Use of Twitter data to improve Zika virus surveillance in the United States during the 2016 epidemic, Shahir Masri, Jianfeng Jia, Chen Li, Guofa Zhou, Ming-Chieh Lee, Guiyun Yan and Jun Wu, BMC Public Health.
20. Inves: Incremental Partitioning-Based Verification for Graph Similarity Search, Jongik Kim, Dong-Hoon Choi, and Chen Li, EDBT 2019.

21. Visually Analyzing A Billion Tweets: An Application for Collaborative Visual Analytics on Large High-Resolution Display, Simon Su, Michael An, Vincent Perry, Jianfeng Jia, Taewoo Kim, Te-Yu Chen, and Chen Li, Workshop on Analysis of Large-scale Disparate Data. BigData 2018: 3597-3606.
22. Heatflip: Temporal-Spatial Sampling for Progressive Heat Maps on Social Media Data, Niklas Stoehr, Johannes Meyer, Volker Markl, Qiushi Bai, Taewoo Kim, De-Yu Chen, Chen Li. Workshop on Big Social Media Data Management and Analysis (BSMDMA2018), BigData 2018: 3723-3732.
23. ZigZag: Supporting Similarity Queries on Vector Space Models, Wenhai Li, Lingfeng Deng, Yang Li, and Chen Li, SIGMOD 2018.
24. End-to-End Machine Learning with Apache AsterixDB, Wail Alkowaileet, Sattam Alsubaiee, Michael J. Carey, Chen Li, Heri Ramampiaro, Phanwadee Sinthong, Xikui Wang, DEEM Workshop 2018 (Co-located with SIGMOD).
25. Supporting Similarity Queries in Apache AsterixDB, Taewoo Kim, Wenhai Li, Alexander Behm, Inci Cetindil, Rares Vernica, Vinayak Borkar, Michael Carey, Chen Li, EDBT 2018.
26. Drum: A Rhythmic Approach to Interactive Analytics on Large Data, Jianfeng Jia, Chen Li, Michael J. Carey, IEEE Big Data 2017.
27. Caching Geospatial Objects in Web Browsers, Taewoo Kim, Vidhyasagar Thirumaraiselvan, Jianfeng Jia, Chen Li, ACM SIGSPATIAL 2017 (Demo Paper).
28. Visual Analytics Ecology for Complex System Testing, Simon Su, Michael Barton, Michael An, Vincent Perry, Chen Li, Jianfeng Jia, Brian Panneton, Visualization in Practice 2017 at IEEE VIS 2017.
29. Twitter Coverage of Climate Change and Health before and after the 2016 US Presidential Election, Suellen Hopfer, Miryha Runnerstrom, Jianfeng Jia, Taewoo Kim, Chen Li, American Public Health Association 2017.
30. A Demonstration of TextDB: Declarative and Scalable Text Analytics on Large Data Sets, Zuozhi Wang, Flavio Bayer, Seungjin Lee, Kishore Narendran, Xuxi Pan, Qing Tang, Jimmy Wang, Chen Li, ICDE 2017 Demo. (Best Demo award).
31. A Comparative Study of Log-Structured Merge-Tree-Based Spatial Indexes for Big Data, Young-Seok Kim, Taewoo Kim, Michael J. Carey, Chen Li, ICDE 2017 Poster.
32. Hobbes3: Dynamic Generation of Variable-Length Signatures for Efficient Approximate Subsequence Mappings, Jongik Kim, Chen Li, Xiaohui Xie, ICDE 2016.
33. Towards Interactive Analytics and Visualization on One Billion Tweets, Jianfeng Jia, Chen Li, Xi Zhang, Chen Li, Michael J. Carey, Simon Su, ACM GIS 2016 (Demo).
34. Negative Factor: Improving Regular-Expression Matching in Strings, Xiaochun Yang, Tao Qiu, Bin Wang, Baihua Zheng, Yaoshu Wang, Chen Li, ACM TODS, 2016.
35. LSM Based Storage and Indexing: An Old Idea with Timely Benefits, Alsubaiee, S., Carey, M. J., Li, C. GeoRich (2015).
36. AsterixDB: A Scalable, Open Source BDMS, Sattam Alsubaiee, Yasser Altowim, Hotham Altwaijry, Alexander Behm, Vinayak R. Borkar, Yingyi Bu, Michael J. Carey, Inci Cetindil, Madhusudan Cheelang, Khurram Faraaz, Eugenia Gabrielova, Raman Grover, Zachary Heilbron, Young-Seok Kim, Chen Li, Guangqiang Li, Ji Mahn Ok, Nicola Onose,

- Pouria Pirzadeh, Vassilis J. Tsotras, Rares Vernica, Jian Wen, Till Westmann: PVLDB 7(14): 1905-1916 (2014).
37. Improving read mapping using additional prefix grams, Jongik Kim, Chen Li, Xiaohui Xie. BMC Bioinformatics 15: 42 (2014)
 38. Storage Management in AsterixDB, Sattam Alsubaiee, Alexander Behm, Vinayak R. Borkar, Zachary Heilbron, Young-Seok Kim, Michael J. Carey, Markus Dreseler, Chen Li, PVLDB 7(10): 841-852 (2014).
 39. Efficient instant-fuzzy search with proximity ranking, Inci Cetindil, Jamshid Esmaelnezhad, Taewoo Kim, Chen Li. ICDE 2014: 328-339.
 40. Supporting Search-As-You-Type Using SQL in Databases, Guoliang Li, Jianhua Feng, Chen Li, IEEE Trans. Knowl. Data Eng. 25(2): 461-475 (2013)
 41. Efficient direct search on compressed genomic data, Xiaochun Yang, Bin Wang, Chen Li, Jiaying Wang, Xiaohui Xie, ICDE 2013: 961-972.
 42. Improving regular-expression matching on strings using negative factors, Xiaochun Yang, Bin Wang, Tao Qiu, Yaoshu Wang, Chen Li, SIGMOD Conference 2013: 361-372.
 43. String similarity measures and joins with synonyms, Jiaheng Lu, Chunbin Lin, Wei Wang, Chen Li, Haiyong Wang, SIGMOD Conference 2013: 373-384.
 44. SKIF-P: a point-based indexing and ranking of web documents for spatial-keyword search, Ali Khodaei, Cyrus Shahabi, Chen Li, Geoinformatica 16(3): 563-596 (2012)
 45. Supporting Efficient Top-k Queries in Type-Ahead Search, Guoliang Li, Jiannan Wang, Chen Li, Jianhua Feng, SIGIR 2012.
 46. Inside "Big Data Management": Ogres, Onions, or Parfaits? Vinayak Borkar, Michael J. Carey, and Chen Li, EDBT 2012.
 47. ASTERIX: An Open Source System for "Big Data" Management and Analysis, Sattam Alsubaiee, Yasser Altowim, Hotham Altwaijry, Alexander Behm, Vinayak R. Borkar, Yingyi Bu, Michael J. Carey, Raman Grover, Zachary Heilbron, Young-Seok Kim, Chen Li, Nicola Onose, Pouria Pirzadeh, Rares Vernica, Jian Wen. PVLDB 2012 (demo).
 48. Speeding Up Chemical Searches Using the Inverted Index: The Convergence of Chemoinformatics and Text Search Methods, Ramzi Nasr, Rares Vernica, Chen Li, Pierre Baldi. Journal of Chemical Information and Modeling 52(4): 891-900 (2012)
 49. Big data platforms: what's next? Vinayak R. Borkar, Michael J. Carey, Chen Li. ACM Crossroads 19(1): 44-49, 2012.
 50. Location-Based Instant Search, Shengyue Ji, Chen Li, SSDBM 2011: 17-36.
 51. Hobbes: optimized gram-based methods for efficient read alignment, Athena Ahmadi, Alexander Behm, Nagesh Honnalli, Chen Li, Lingjie Weng, and Xiaohui Xie, Nucleic Acids Research 2011; doi: 10.1093/nar/gkr1246.
 52. SKIF-P: a point-based indexing and ranking of web documents for spatial-keyword search, Ali Khodaei, Cyrus Shahabi, and Chen Li, Geoinformatica, Springer, 2011.
 53. Supporting BioMedical Information Retrieval: The BioTracer Approach, Heri Ramampiaro and Chen Li, In Transactions on Large-Scale Data- and Knowledge-Centered Systems (TLDKS), 2011, No.4. Vol. 6990, Springer. pp. 73-94.
 54. CHIME: An Efficient Error-Tolerant Chinese Pinyin Input Method, Yabin Zheng, Chen Li, Maosong Sun, IJCAI 2011, 2551-2556.

55. Answering Approximate String Queries on Large Data Sets Using External Memory, Alexander Behm, Chen Li, and Michael Carey, ICDE 2011.
56. ASTERIX: towards a scalable, semistructured data platform for evolving-world models. Alexander Behm, Vinayak R. Borkar, Michael J. Carey, Raman Grover, Chen Li, Nicola Onose, Rares Vernica, Alin Deutsch, Yannis Papakonstantinou, Vassilis J. Tsotras, Distributed and Parallel Databases, 2011, 29(3), 185-216.
57. Efficient fuzzy full-text type-ahead search, Guoliang Li, Shengyue Ji, Chen Li, Jianhua Feng. VLDB J. 20(4): 617-640 (2011).
58. qSpell: Spelling Correction of Web Search Queries using Ranking Models and Iterative Correction. Yasser Ganjisaffar, Andrea Zilio, Sara Javanmardi, Inci Cetindil, Manik Sikka, Sandeep Katumalla, Narges Khatib, Chen Li, Cristina Lopes, Spelling Alteration for Web Search Workshop, July 2011. , [Dataset] (The authors won the third place in the Microsoft's speller challenge in 2011.)
59. The Flamingo Software Package on Approximate String Queries. Chen Li, DASFAA Workshops 2011, 477.
60. Interactive and Fuzzy Search: A Dynamic Way to Explore MEDLINE, Jiannan Wang, Inci Cetindil, Shengyue Ji, Chen Li, Xiaohui Xie, Guoliang Li, Jianhua Feng, Journal of Bioinformatics, 2010.
61. Supporting Location-Based Approximate-Keyword Queries, Sattam Alsubaiee, Alexander Behm, and Chen Li, ACM GIS 2010.
62. Hybrid Indexing and Seamless Ranking of Spatial and Textual Features of Web Documents, Ali Khodaei, Cyrus Shahabi, Chen Li, DEXA 2010.
63. Efficient Parallel Set-Similarity Joins Using MapReduce. Rares Vernica, Michael J. Carey, Chen Li, SIGMOD 2010.
64. Seaform: Search-As-You-Type in Forms, Hao Wu, Guoliang Li, Chen Li, Lizhu Zhou, VLDB 2010 (Demo).
65. Search-As-You-Type: Opportunities and Challenges, Chen Li, Guoliang Li, IEEE Data Eng. Bull. 33(1): 37-45 (2010).
66. Fuzzy Keyword Search on Spatial Data, Sattam Alsubaiee, Chen Li: DASFAA, Excellent Demo Award, 2010: 464-467.
67. SAIL: Structure-aware indexing for effective and progressive top-k keyword search over XML documents, Guoliang Li, Chen Li, Jianhua Feng, Lizhu Zhou: Inf. Sci. 179(21): 3745-3762 (2009).
68. Human genomes as email attachments. Scott Christley, Yiming Lu, Chen Li, and Xiaohui Xie, Bioinformatics 25: 274-275 (2009). It was the most downloaded article on the Web site of the Journal of Bioinformatics for two months.
69. Rewriting Queries using Views, Chen Li: Encyclopedia of Database Systems 2009: 2438-2441.
70. Type-Ahead Search on Relational Data: a TASTIER Approach, Guoliang Li, Shengyue Ji, Chen Li, and Jianhua Feng, SIGMOD 2009.
71. Efficient Interactive Fuzzy Keyword Search, Shengyue Ji, Guoliang Li, Chen Li, and Jianhua Feng, WWW 2009.

72. Best-Effort Top-k Query Processing Under Budgetary Constraints, Michal Shmueli-Scheuer, Chen Li, Yosi Mass, Haggai Roitman, Ralf Schenkel, and Gerhard Weikum, ICDE 2009.
73. Space-Constrained Gram-Based Indexing for Efficient Approximate String Search, Alexander Behm, Shengyue Ji, Chen Li, and Jiaheng Lu, ICDE 2009. , [PPTX]
74. Efficient top-k algorithms for fuzzy search in string collections, Rares Vernica, Chen Li, KEYS 2009: 9-14.
75. Efficient Approximate Search on String Collections (Tutorial), Marios Hadjieleftheriou and Chen Li, VLDB 2009.
76. Efficient Approximate Search on String Collections (Tutorial), Marios Hadjieleftheriou, Chen Li, ICDE 2009.
77. SEPIA: Estimating Selectivities of Approximate String Predicates in Large Databases. Liang Jin, Chen Li, and Rares Vernica. VLDB Journal, Volume 17, Number 5, pages 1213-1229, August 2008.
78. Cost-Based Variable-Length-Gram Selection for String Collections to Support Approximate Queries Efficiently, Xiaochun Yang, Bin Wang, and Chen Li, ACM SIGMOD 2008.
79. Efficient Merging and Filtering Algorithms for Approximate String Searches, Chen Li, Jiaheng Lu, and Yiming Lu. ICDE 2008.
80. Data Exchange with Arithmetic Comparisons, Foto Afrati, Chen Li, and Vassia Pavlaki. EDBT 2008.
81. Quality-Aware Retrieval of Data Objects from Autonomous Sources for Web-Based Repositories, Houtan Shirani-Mehr, Chen Li, Gang Liang, Michal Shmueli-Scheuer, ICDE 2008 (poster).
82. Using Views to Generate Efficient Evaluation Plans for Queries Foto Afrati, Chen Li, and Jeff Ullman, Journal of Computer and System Sciences, Volume 73, Issue 5, pages 703-724, August 2007.
83. VGRAM: Improving Performance of Approximate Queries on String Collections Using Variable-Length Grams, Chen Li, Bin Wang, and Xiaochun Yang. VLDB 2007.
84. Processing Spatial-Keyword (SK) Queries in Geographic Information Retrieval (GIR) Systems, Ramaswamy Hariharan, Bijit Hore, Chen Li, Sharad Mehrotra, SSDBM 2007.
85. Protecting Individual Information Against Inference Attacks in Data Publishing, Chen Li, Houtan Shirani-Mehr, and Xiaochun Yang. DASFAA 2007.
86. Communication-Efficient Query Answering with Quality Guarantees in Client-Server Applications. Michal Shmueli-Scheuer, Amitabh Chaudhary, Avigdor Gal, Chen Li. WebDB 2007.
87. Rewriting Queries Using Views in the Presence of Arithmetic Comparisons, Foto Afrati, Chen Li, and Prasenjit Mitra, Theoretical Computer Science, Volume 368, Numbers 1-2, pages 88-123, 2006.
88. Supporting Efficient Record Linkage for Large Data Sets Using Mapping Techniques, Chen Li, Liang Jin, and Sharad Mehrotra, World Wide Web Journal, Volume 9, Number 4, pages 557-584, December 2006.

89. Achieving Communication Efficiency through Push-Pull Partitioning of Semantic Spaces to Disseminate Dynamic Information, Amitabha Bagchi, Amitabh Chaudhary, Michael T. Goodrich, Chen Li, and Michal Shmueli-Scheuer. *IEEE Transaction on Knowledge and Data Engineering (TKDE)*, October 2006 (Vol. 18, No. 10).
90. Answering Queries Using Materialized Views with Minimum Size. Rada Chirkova, Chen Li, and Jia Li. *VLDB Journal* (2006), Volume 15, Number 3, 191-210.
91. Supporting Approximate Similarity Queries with Quality Guarantees in P2P Systems, Qi Zhong, Iosif Lazaridis, Mayur Deshpande, Chen Li, Sharad Mehrotra, Hal Stern, COMAD 2006, December 14-16, 2006, Delhi, India.
92. Relaxing Join and Selection Queries. Nick Koudas, Chen Li, Anthony Tung, and Rares Vernica. *VLDB 2006*, Seoul, Korea, 2006.
93. Selectivity Estimation for Fuzzy String Predicates in Large Data Sets, Liang Jin and Chen Li. *VLDB 2005*, Trondheim, Norway, August 30 – September 2, 2005.
94. Indexing Mixed Types for Approximate Retrieval, Liang Jin, Nick Koudas, Chen Li, Anthony K.H. Tung. *VLDB 2005*, Trondheim, Norway, August 30 – September 2, 2005.
95. Quality-Driven Approximate Methods for GIS Data Integration. Ramaswamy Hariharan, Michal Schmueeli-Scheuer, Chen Li, and Sharad Mehrotra. *ACM GIS 2005*, November 4-5th, 2005 Bremen, Germany.
96. Answering Aggregation Queries on Hierarchical Web Sites Using Adaptive Sampling. Foto Afrati, Paraskevas Lekeas, and Chen Li. Technical Report, UCI ICS, August 2005. A short version appears in *CIKM'2005*, 31st October – 5th November, 2005 Bremen, Germany.
97. XGuard: A System for Publishing XML Documents without Information Leakage in the Presence of Data Inference. Xiaochun Yang, Chen Li, Ge Yu, and Lei Shi. *Proc. of ICDE'2005*, demo track, Tokyo, Japan, March 2005.
98. Secure XML Publishing without Information Leakage in the Presence of Data Inference. Xiaochun Yang and Chen Li. *VLDB*, Toronto, Canada, August 29 – September 3, 2004.
99. NNH: Improving Performance of Nearest-Neighbor Searches Using Histograms. Liang Jin, Nick Koudas, Chen Li. *EDBT*, Crete, Greece, March 2004.
100. On Containment of Conjunctive Queries with Arithmetic Comparisons. Foto Afrati, Chen Li, Prasenjit Mitra. *EDBT*, Crete, Greece, March 2004.
101. RACCOON: A Peer-Based System for Data Integration and Sharing. Chen Li, Jia Li, Qi Zhong. *Proc. of ICDE'2004*, demo track.
102. Recent Progress on Selected Topics on Database Research — A Report from Nine Young Chinese Researchers Working in the United States. Zhiyuan Chen, Chen Li, Jian Pei, Yufei Tao, Haixun Wang, Wei Wang, Jiong Yang, Jun Yang, and Donghui Zhang. *The Journal of Computer Science and Technology*. Vol. 18, No. 5, Pages 538 – 552, September 2003.
103. Computing Complete Answers to Queries in the Presence of Limited Access Patterns. Chen Li. *The VLDB Journal* (2003) 12: 211-227.
104. Materializing Views with Minimal Size to Answer Queries. Rada Chirkova and Chen Li. *ACM PODS*, June 2003, San Diego, CA.

105. Efficient Record Linkage in Large Data Sets, Liang Jin, Chen Li, and Sharad Mehrotra, in the 8th International Conference on Database Systems for Advanced Applications (DASFAA 2003) 26 – 28 March, 2003, Kyoto, Japan. (33% accepted). Received DASFAA 2013 10-year Best Paper Award.
106. Schema-Guided Wrapper Maintenance for Web-Data Extraction. Xiaofeng Meng, Dongdong Hu, Chen Li. To appear in the Fifth International Workshop on Web Information and Data Management (WIDM 2003), New Orleans, Louisiana.
107. A Supervised Visual Wrapper Generator for Web-Data Extraction. . Xiaofeng Meng, Haiyan Wang, Dongdong Hu, Chen Li. COMPSAC 2003: 657-662.
108. Using Constraints to Describe Source Contents in Data Integration Systems. Chen Li. IEEE Intelligent Systems 18(5): 49-53 (2003).
109. Describing and Utilizing Constraints to Answer Queries in Data-Integration Systems. Chen Li. IJCAI 2003 workshop on Information Integration on the Web, August 2003, Acapulco, Mexico.
110. Clustering for Approximate Similarity Search in High-Dimensional Spaces. Chen Li, Edward Chang, Hector Garcia-Molina, and Gio Wiederhold. IEEE Transaction on Knowledge and Data Engineering, Volume 14, Number 4, pp.792-808, July/August 2002.
111. Executing SQL over Encrypted Data in the Database-Service-Provider Model. Hakan Hacigumus, Bala Iyer, Chen Li, and Sharad Mehrotra. In ACM SIGMOD, June 3-6, 2002 Madison, Wisconsin. (18% accepted). Received SIGMOD 2012 10-year Test-of-Time Award.
112. Answering Queries Using Views with Arithmetic Comparisons. Foto Afrati, Chen Li, and Prasenjit Mitra. In ACM Symposium on Principles of Database Systems (PODS), June 3-6, 2002, Madison, Wisconsin.
113. Answering Queries with Useful Bindings. Chen Li and Edward Chang. ACM Transactions on Database Systems (TODS), Volume 26 , Issue 3 (September 2001).
114. Generating Efficient Plans for Queries Using Views. Foto Afrati, Chen Li, and Jeff Ullman. In the Proc. of the 30th ACM SIGMOD Conference, Santa Barbara, CA, May, 2001.
115. Minimizing View Sets without Losing Query-Answering Power. Chen Li, Mayank Bawa, and Jeff Ullman. In the 8th International Conference on Database Theory (ICDT), London, UK, January, 2001.
116. On Answering Queries in the Presence of Limited Access Patterns. Chen Li and Edward Chang. In the 8th International Conference on Database Theory (ICDT), London, UK, January, 2001.
117. Query Processing and Optimization in Information-Integration Systems. Chen Li. Ph.D. Thesis, Computer Science Department, Stanford University, August, 2001.
118. Query Planning with Limited Source Capabilities. Chen Li and Edward Chang. International Conference on Database Engineering (ICDE), pages 401-412, San Diego, CA, February, 2000.
119. Towards Perception-Based Image Retrieval. Edward Chang, Beita Li, and Chen Li. Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries, p. 401-412, South Carolina, June, 2000.

120. Managing Parallel Disks for Continuous Media Data. Edward Chang, Chen Li, and Hector Garcia-Molina. A Book Chapter in Information Organization & Databases, p.107-120, Kluwer Publisher, 2000.
121. Answering Queries with Database Restrictions (Research Summary). Chen Li. Symposium on Abstraction, Reformulation and Approximation (SARA), pages 328 – 329, July, 2000, Horseshoe Bay (Lake LBJ), Texas.
122. I wrote a report for the Workshop on Data Mining in the Internet Age, which was held May 1 – 2, 2000, IBM Almaden Center, San Jose, California.
123. Computing Capabilities of Mediators. Ramana Yerneni, Chen Li, Hector Garcia-Molina, Jeffrey Ullman. SIGMOD'99, Philadelphia, PA, May 1999.
124. Optimizing Large Join Queries in Mediation Systems. Ramana Yerneni, Chen Li, Jeffrey Ullman, Hector Garcia-Molina. International Conference on Database Theory (ICDT), Jerusalem, Israel, January, 1999.
125. Searching Near-Replicas of Images via Clustering. Edward Chang, Chen Li, James Wang, Peter Mork, and Gio Wiederhold. Proc. of SPIE Symposium of Voice, Video, and Data Communications, Multimedia Storage and Archiving Systems VI, pages 281-292, Boston, MA, September, 1999.
126. Capability Based Mediation in TSIMMIS. Chen Li, Ramana Yerneni, Vasilis Vassalos, Hector Garcia-Molina, Yannis Papakonstantinou, Jeffrey Ullman, Murty Valiveti. Proc. of ACM SIGMOD 1998, demo track, pages 564 – 566, Seattle, WA, June, 1998.
127. RIME: A Replicated Image Detector for the World-Wide Web. Edward Chang, James Ze Wang, Chen Li, and Gio Wiederhold. Proceedings of SPIE Symposium of Voice, Video, and Data Communications, pages 58–67, Boston, MA, November 1998.
128. 2D BubbleUp: Managing Parallel Disks for Media Servers. Edward Chang, Hector Garcia-Molina, and Chen Li. The 5th International Conference of Foundations of Data Organization (FODO), pages 221-230, Kobe, Japan, 1998.
129. Performance Analysis of the Communication Mechanism for POE Workstation Cluster. Weiqiang Zhuang, Chen Li, Meiming Shen. Microcomputer & Micro-system, Jan, 1995
130. HiComm — A New Technique for Improving Communication Performance in Workstation Cluster. Chen Li, Weiqiang Zhuang, Meiming Shen, Dingxing Wang, Weimin Zheng, Proc. of International Workshop on Advanced Parallel Processing Technologies (APPT), October, 1995, Beijing, China.