

- [76] Combi C, Parise P, Sala P, Pozzi G. Mining approximate temporal functional dependencies based on pure temporal grouping. In: Proc. of the ICDMW. 2013. 258–265. [doi: 10.1109/ICDMW.2013.100]
- [77] Fan WF, Geerts F, Tang N, Yu WY. Conflict resolution with data currency and consistency. Journal of Data and Information Quality, 2014,5(1-2):6. [doi: 10.1145/2631923]
- [78] Abedjan Z, Akcora CG, Ouzzani M, Papotti P, Stonebraker M. Temporal rules discovery for Web data cleaning. In: Proc. of the VLDB. 2016. 336–347. [doi: 10.14778/2856318.2856328]
- [79] Li MH, Li JZ. A minimized-rule based approach for improving data currency. Journal of Combinatorial Optimization, 2015. 1–30. [doi: 10.1007/s10878-015-9904-8]
- [80] Li MH, Li JZ. Algorithms for improving data currency. Journal of Computer Research and Development, 2015,52(9):1992–2001 (in Chinese with English abstract).
- [81] Libkin L. Incomplete data: What went wrong, and how to fix it. In: Proc. of the PODS. 2014. 1–13. [doi: 10.1145/2594538.2594561]
- [82] Liu H, Zhang S. Noisy data elimination using mutual k -nearest neighbor for classification mining. Journal of Systems & Software, 2012,85(5):1067–1074. [doi: 10.1016/j.jss.2011.12.019]
- [83] Tian J, Yu B, Yu D, Ma S. Missing data analysis: A hybrid multiple imputation algorithm using gray system theory and entropy based on clustering. Applied Intelligence, 2013,40:376–388. [doi: 10.1007/s10489-013-0469-x]
- [84] Van Buuren S. Flexible Imputation of Missing Data. Boca Raton: CRC Press, 2012.
- [85] Zhang S. Shell-Neighbor method and its application in missing data imputation. Applied Intelligence, 2011,35(1):123–133. [doi: 10.1007/s10489-009-0207-6]
- [86] Zhang S. Nearest neighbor selection for iteratively k NN imputation. Journal of Systems & Software, 2012,85(11):2541–2552. [doi: 10.1016/j.jss.2012.05.073]
- [87] Zhang S, Jin Z, Zhu X. Missing data imputation by utilizing information within incomplete instances. Journal of Systems & Software, 2012,84(3):452–459. [doi: 10.1016/j.jss.2010.11.887]
- [88] Zhu X, Zhang S, Jin Z, Zhang Z, Xu Z. Missing value estimation for mixed-attribute data sets. IEEE Trans. on Knowledge & Data Engineering, 2011,23(1):110–121. [doi: 10.1109/TKDE.2010.99]
- [89] Song S, Zhang A, Chen L, Wang J. Enriching data imputation with extensive similarity neighbors. In: Proc. of the VLDB. 2015. 1286–1297. [doi: 10.14778/2809974.2809989]
- [90] Wu S, Feng X, Han Y, Wang Q. Missing categorical data imputation approach based on similarity. In: Proc. of the IEEE Int'l Conf. on Systems, Man, and Cybernetics (SMC). 2012. 2827–2832. [doi: 10.1109/ICSMC.2012.6378177]
- [91] Gummadi R, Khulbe A, Kalavagattu A, Salvi S, Kambhampati S. SMARTINT: Using mined attribute dependencies to integrate fragmented Web databases. Journal of Intelligent Information Systems, 2012,38:575–599. [doi: 10.1007/s10844-011-0169-0]
- [92] Koutrika G. Entity reconstruction: Putting the pieces of the puzzle back together. Technical Report, Palo Alto: HP Labs, 2012.
- [93] Yakout M, Ganjam K, Chakrabarti K, Chaudhuri S. InfoGather: Entity augmentation and attribute discovery by holistic matching with Web tables. In: Proc. of the SIGMOD. 2012. 97–108. [doi: 10.1145/2213836.2213848]
- [94] Li Z, Qin L, Cheng H, Zhang X, Zhou X. TRIP: An interactive retrieving-inferring data imputation approach. IEEE Trans. on Knowledge and Data Engineering, 2015,27(9):2550–2563. [doi: 10.1109/TKDE.2015.2411276]
- [95] Li ZX, Shang S, Xie Q, Zhang XL. Cost reduction for Web-based data imputation. In: Proc. of the Database Systems for Advanced Applications. Springer Int'l Publishing, 2014. 438–452. [doi: 10.1007/978-3-319-05813-9_29]
- [96] Ye C, Wang HZ, Li JZ, Gao H, Cheng SY. Crowdsourcing-Enhanced missing values imputation based on Bayesian network. In: Proc. of the DASFAA. 2016. 67–81. [doi: 10.1007/978-3-319-32025-0_5]
- [97] Korn F, Saha B, Srivastava D, Ying SS. On repairing structural problems in semi-structured data. In: Proc. of the VLDB. 2013. 601–612. [doi: 10.14778/2536360.2536361]
- [98] Wang J, Song S, Zhu X, Lin X. Efficient recovery of missing events. In: Proc. of the VLDB. 2013. 841–852. [doi: 10.14778/2536206.2536212]
- [99] Wang S, Xiao X, Lee CH. Crowd-Based deduplication: An adaptive approach. In: Proc. of the SIGMOD. 2015. 1263–1277. [doi: 10.1145/2723372.2723739]
- [100] Gokhale C, Das S, Doan A, Naughton JF, Rampalli N, Shavlik JW, Zhu X. Corleone: Hands-Off crowdsourcing for entity matching. In: Proc. of the SIGMOD. 2014. 601–612. [doi: 10.1145/2588555.2588576]

- [101] Verroios V, Garcia-Molina H. Entity resolution with crowd errors. In: Proc. of the ICDE. 2015. 219–230. [doi: 10.1109/ICDE.2015.7113286]
- [102] Vesdapunt N, Bellare K, Dalvi NN. Crowdsourcing algorithms for entity resolution. In: Proc. of the VLDB. 2014. 1071–1082. [doi: 10.14778/2732977.2732982]
- [103] Whang SE, Lofgren P, Garcia-Molina H. Question selection for crowd entity resolution. In: Proc. of the VLDB. 2013. 349–360. [doi: 10.14778/2536336.2536337]
- [104] Hua W, Zheng K, Zhou XF. Microblog entity linking with social temporal context. In: Proc. of the SIGMOD. 2015. 1761–1775. [doi: 10.1145/2723372.2751522]
- [105] Shen W, Han JW, Wang JY. A probabilistic model for linking named entities in Web text with heterogeneous information networks. In: Proc. of the SIGMOD. 2014. 1199–1210. [doi: 10.1145/2588555.2593676]
- [106] Zhu X, Song S, Lian X, Wang J, Zou L. Matching heterogeneous event data. In: Proc. of the SIGMOD. 2014. 1211–1222. [doi: 10.1145/2588555.2588570]
- [107] Chiang YH, Doan AH, Naughton JF. Modeling entity evolution for temporal record matching. In: Proc. of the SIGMOD. 2014. 1175–1186. [doi: 10.1145/2588555.2588560]
- [108] Whang SE, Garcia-Molina H. Incremental entity resolution on rules and data. VLDB, 2014,23(1):77–102. [doi: 10.1007/s00778-013-0315-0]
- [109] Gruenheid A, Dong XL, Srivastava D. Incremental record linkage. In: Proc. of the VLDB. 2014. 697–708. [doi: 10.14778/2732939.2732943]
- [110] Wildani A, Miller EL, Rodeh O. HANDS: A heuristically arranged non-backup in-line deduplication system. In: Proc. of the ICDE. 2013. 446–457. [doi: 10.1109/ICDE.2013.6544846]
- [111] Li X, Dong XL, Lyons KB, Meng W, Srivastava D. Scaling up copy detection. In: Proc. of the ICDE. 2015. [doi: 10.1109/ICDE.2015.7113275]
- [112] Whang SE, Marmaros D, Garcia-Molina H. Pay-as-You-Go entity resolution. IEEE Trans. on Knowledge and Data Engineering, 2013,25(5):1111–1124. [doi: 10.1109/TKDE.2012.43]
- [113] Li LL, Li JZ, Wang HZ, Gao H. Context-Based entity description rule for entity resolution. In: Proc. of the CIKM. 2011. 1725–1730. [doi: 10.1145/2063576.2063825]
- [114] Li LL, Li JZ, Gao H. Rule-Based method for entity resolution. IEEE Trans. on Knowledge and Data Engineering, 2015,27(1):250–263. [doi: 10.1109/TKDE.2014.2320713]
- [115] Wang FD, Wang HZ, Li JZ, Gao H. Graph-Based reference table construction to facilitate entity matching. Journal of Systems and Software, 2013,86(6):1679–1688. [doi: 10.1016/j.jss.2013.02.026]
- [116] Altowim Y, Kalashnikov DV, Mehrotra S. Progressive approach to relational entity resolution. In: Proc. of the VLDB. 2014. 999–1010. [doi: 10.14778/2732967.2732975]
- [117] Altwaijry H, Kalashnikov DV, Mehrotra S. Query-Driven approach to entity resolution. In: Proc. of the VLDB. 2013. 1846–1857. [doi: 10.14778/2556549.2556567]
- [118] Wang HZ, Li JZ, Gao H. Efficient entity resolution based on subgraph cohesion. Knowledge Information Systems, 2016,46(2):285–314. [doi: 10.1007/s10115-015-0818-7]
- [119] Li Q, Li YL, Gao J, Su L, Zhao B, Demirbas M, Fan W, Han JW. A confidence-aware approach for truth discovery on long-tail data. In: Proc. of the VLDB. 2015. 425–436. [doi: 10.14778/2735496.2735505]
- [120] Prokoshyna N, Szlichta J, Chiang F, Miller RJ, Srivastava D. Combining quantitative and logical data cleaning. In: Proc. of the VLDB. 2016. 300–311. [doi: 10.14778/2856318.2856325]
- [121] Zhao Z, Cheng J, Ng W. Truth discovery in data streams: A single-pass probabilistic approach. In: Proc. of the CIKM. 2014. 1589–1598. [doi: 10.1145/2661829.2661892]
- [122] Interlandi M, Tang N. Proof positive and negative in data cleaning. In: Proc. of the ICDE. 2015. 18–29. [doi: 10.1109/ICDE.2015.7113269]
- [123] Ding XO, Wang HZ, Zhang XY, Li JZ, Gao H. Association relationships study of multi-dimensional data quality. Ruan Jian Xue Bao/Journal of Software, 2016,27(7):1626–1644 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5040.htm> [doi: 10.13328/j.cnki.jos.005040]
- [124] Fan W, Geerts F, Tang N, Yu W. Inferring data currency and consistency for conflict resolution. In: Proc. of the ICDE. 2013. 470–481. [doi: 10.1109/ICDE.2013.6544848]

- [125] Yang DH, Li NN, Wang HZ, Li JZ, Gao H. The optimization of the big data cleaning based on task merging. Chinese Journal of Computers, 2015,39(1):97–108 (in Chinese with English abstract).
- [126] Wang X, Dong XL, Meliou A. Data X-ray: A diagnostic tool for data errors. In: Proc. of the SIGMOD. 2015. 1231–1245. [doi: 10.1145/2723372.2750549]
- [127] Wang XL, Feng M, Wang Y, Dong XL, Meliou A. Error diagnosis and data profiling with data X-ray. In: Proc. of the VLDB. 2015. 1984–1995. [doi: 10.14778/2824032.2824117]
- [128] Prokoshyna N, Szlichta J, Chiang F, Miller RJ, Srivastava D. Combining quantitative and logical data cleaning. In: Proc. of the VLDB. 2016. 300–311. [doi: 10.14778/2856318.2856325]
- [129] Geerts F, Mecca G, Papotti P, Santoro D. Mapping and cleaning. In: Proc. of the ICDE. 2014. 232–243. [doi: 10.1109/ICDE.2014.6816654]
- [130] Chu X, Ilyas IF, Papotti P. Holistic data cleaning: Putting violations into context. In: Proc. of the ICDE. 2013. 458–469. [doi: 10.1109/ICDE.2013.6544847]
- [131] Li ZY, Wang HZ, Shao W, Li JZ, Gao H. Repairing data through regular expressions. PVLDB, 2016,9(5):432–443. [doi: 10.14778/2876473.2876478]
- [132] Zhang CJ, Chen L, Tong Y, Liu Z. Cleaning uncertain data with a noisy crowd. In: Proc. of the ICDE. 2015. 6–17. [doi: 10.1109/ICDE.2015.7113268]
- [133] Wang J, Song S, Lin X, Zhu X, Pei J. Cleaning structured event logs: A graph repair approach. In: Proc. of the ICDE. 2015. 30–41. [doi: 10.1109/ICDE.2015.7113270]
- [134] Volkovs M, Chiang F, Szlichta J, Miller RJ. Continuous data cleaning. In: Proc. of the ICDE. 2014. 244–255. [doi: 10.1109/ICDE.2014.6816655]
- [135] Fan WF, Li JZ, Tang N, Yu WY. Incremental detection of inconsistencies in distributed data. IEEE Trans. on Knowledge and Data Engineering, 2014,26(6):1367–1383. [doi: 10.1109/TKDE.2012.138]
- [136] Yakout M, Berti-Equille L, Elmagarmid AK. Don't be SCARED: Use SCalable automatic REpairing with maximal likelihood and bounded changes. In: Proc. of the SIGMOD. 2013. 553–564. [doi: 10.1145/2463676.2463706]
- [137] Dong XL, Gabrilovich E, Murphy K, Dang V, Horn W, Lugaresi C, Sun S, Zhang W. Knowledge-Based trust: Estimating the trustworthiness of Web sources. In: Proc. of the VLDB. 2015. 938–949. [doi: 10.14778/2777598.2777603]
- [138] Rekatsinas T, Dong XL, Srivastava D. Characterizing and selecting fresh data sources. In: Proc. of the SIGMOD. 2014. 919–930. [doi: 10.1145/2588555.2610504]
- [139] Pochampally R, Sarma AD, Dong XL, Meliou A, Srivastava D. Fusing data with correlations. In: Proc. of the SIGMOD. 2014. 433–444. [doi: 10.1145/2588555.2593674]
- [140] Li Q, Li YL, Gao J, Zhao B, Fan W, Han JW. Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation. In: Proc. of the SIGMOD. 2014. 1187–1198. [doi: 10.1145/2588555.2610509]
- [141] Chalamalla A, Ilyas IF, Ouzzani M, Papotti P. Descriptive and prescriptive data cleaning. In: Proc. of the SIGMOD. 2014. 445–456. [doi: 10.1145/2588555.2610520]
- [142] Dong XL, Berti-Equille L, Srivastava D. Integrating conflicting data: The role of source dependence. In: Proc. of the VLDB. 2009. 145. [doi: 10.14778/1687627.1687690]
- [143] Dong XL, Berti-Equille L, Srivastava D. Truth discovery and copying detection in a dynamic world. In: Proc. of the VLDB. 2009. 146. [doi: 10.14778/1687627.1687691]
- [144] Dong XL, Berti-Equille L, Hu YF, Srivastava D. Global detection of complex copying relationships between sources. In: Proc. of the VLDB. 2010. 1358–1369. [doi: 10.14778/1920841.1921008]
- [145] Dong XL. Solomon: Seeking the truth via copying detection. In: Proc. of the VLDB. 2010. 1358–1369. [doi: 10.1145/1966883.1966887]
- [146] Dong XL, Naumann F. Data fusion: Resolving data conflicts for integration. In: Proc. of the VLDB. 2009. 1654–1655. [doi: 10.14778/1687553.1687620]
- [147] Cheng SY, Li JZ. Sampling based (ϵ, δ) -approximate aggregation algorithm in sensor networks. In: Proc. of the IEEE ICDCS 2009. Piscataway, 2009. 273–280. [doi: 10.1109/ICDCS.2009.8]
- [148] Li JZ, Cheng SY. (ϵ, δ) -Approximate aggregation algorithms in dynamic sensor networks. IEEE Trans. on Parallel and Distributed Systems, 2012,23(3):385–396. [doi: 10.1109/TPDS.2011.193]

- [149] Cheng SY, Li JZ, Cai ZP. ϵ -Approximation to physical world by sensor networks. In: Proc. of the INFOCOM. Piscataway, 2013. 3184–3192. [doi: 10.1109/INFOCOM.2013.6567121]
- [150] Li JZ, Li GH, Gao H. Novel ϵ -approximation to data streams in sensor networks. IEEE Trans. on Parallel Distrib. System, 2015, 26(6):1654–1667. [doi: 10.1109/TPDS.2014.2323056]
- [151] Cheng SY, Li JZ, Liu Y. Location aware peak value queries in sensor networks. In: Proc. of the INFOCOM. Piscataway, 2012. 486–494. [doi: 10.1109/INFOCOM.2012.6195789]
- [152] Gao J, Li JZ. Composite event coverage in wireless sensor networks with heterogeneous sensors. In: Proc. of the INFOCOM. 2015. 217–225. [doi: 10.1109/INFOCOM.2015.7218385]
- [153] Li JZ, Cheng SY, Gao H, Cai ZP. Approximate physical world reconstruction algorithms in sensor networks. IEEE Trans. on Parallel and Distributed Systems, 2014,25(12):3099–3110. [doi: 10.1109/TPDS.2013.2297121]
- [154] Cheng SY, Cai ZP, Li JZ, Fang XL. Drawing dominant dataset from big sensory data in wireless sensor networks. In: Proc. of the INFOCOM. 2015. 531–539. [doi: 10.1109/INFOCOM.2015.7218420]
- [155] Data collection in multi-application sharing wireless sensor networks. IEEE Trans. on Parallel and Distributed Systems, 2015,26(2): 403–412. [doi: 10.1109/TPDS.2013.289]
- [156] Li JZ, Yu L, Gao H, Xiong SG. Grouping-Enhanced resilient probabilistic en-route filtering of injected false data in WSNs. IEEE Trans. on Parallel and Distributed Systems, 2012,23(5):881–889. [doi: 10.1109/TPDS.2011.217]
- [157] Yu L, Li JZ, Cheng SY, Xiong SG, Shen HY. Secure continuous aggregation via sampling-based verification in wireless sensor networks. IEEE Trans. on Parallel and Distributed Systems, 2014,25(3):762–744. [doi: 10.1109/TPDS.2013.63]
- [158] Altwaijry H, Mehrotra S, Kalashnikov DV. QuERy: A framework for integrating entity resolution with query processing. In: Proc. of the VLDB. 2015. 120–131. [doi: 10.14778/2850583.2850587]
- [159] Rezig EK, Dragut EC, Ouzzani M, Elmagarmid AK. Query-Time record linkage and fusion over Web databases. In: Proc. of the ICDE. 2015. 42–53. [doi: 10.1109/ICDE.2015.7113271]
- [160] Liu XL, Wang HZ, Li JZ, Gao H. Similarity join algorithm based on entity. Ruan Jian Xue Bao/Journal of Software, 2015,26(6): 1421–1437 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4610.htm> [doi: 10.13328/j.cnki.jos.004610]
- [161] Razniewski S, Korn F, Nutt W, Srivastava D. Identifying the extent of completeness of query answers over partially complete databases. In: Proc. of the SIGMOD. 2015. 561–576. [doi: 10.1145/2723372.2750544]
- [162] Savkovic O, Mirza P, Tomasi A, Nutt W. Complete approximations of incomplete queries. In: Proc. of the VLDB. 2013. 1378–1381. [doi: 10.14778/2536274.2536320]
- [163] Bharuka R, Kumar PS. Finding skylines for incomplete data. In: Proc. of the 24th Australasian Database Conf., Vol.137. Australian Computer Society, Inc., 2013. 109–117.
- [164] Lofi C, El Maarry K, Balke WT. Skyline queries over incomplete data-error models for focused crowd-sourcing. In: Proc. of the Conceptual Modeling. Berlin, Heidelberg: Springer-Verlag, 2013. 298–312. [doi: 10.1007/978-3-642-41924-9_25]
- [165] Lofi C, El Maarry K, Balke WT. Skyline queries in crowd-enabled databases. In: Proc. of the 16th Int'l Conf. on Extending Database Technology. ACM Press, 2013. 465–476. [doi: 10.1145/2452376.2452431]
- [166] Miao X, Gao Y, Chen L, Chen G, Li Q, Jiang T. On efficient k -skyband query processing over incomplete data. In: Proc. of the Database Systems for Advanced Applications. Berlin, Heidelberg: Springer-Verlag, 2013. 424–439. [doi: 10.1007/978-3-642-37487-6_32]
- [167] Gao Y, Miao X, Cui H, Chen G, Li Q. Processing k -Skyband, constrained skyline, and group-by skyline queries on incomplete data. Expert Systems with Applications, 2014,41(10):4959–4974. [doi: 10.1016/j.eswa.2014.02.033]
- [168] Arefin MS, Morimoto Y. Skyline sets queries from databases with missing values. In: Proc. of the 22nd Int'l Conf. on Computer Theory and Applications. IEEE, 2012. 24–29. [doi: 10.1109/ICCTA.2012.6523542]
- [169] Markus E, Patrick R, Florian W, Alfons H, Werner K. Handling of null values in preference database queries. In: Proc. of the 6th Multidisciplinary Workshop on Advances in Preference Handling.
- [170] Kolaitis PG, Pema E, Tan WC. Efficient querying of inconsistent databases with binary integer programming. In: Proc. of the VLDB. 2013. 397–408. [doi: 10.14778/2536336.2536341]
- [171] Bertossi LE, Kolahi S, Lakshmanan LVS. Data cleaning and query answering with matching dependencies and matching functions. In: Proc. of the ICDT. 2011. 268–279. [doi: 10.1145/1938551.1938585]
- [172] Wang J, Krishnan S, Franklin MJ, Goldberg K, Kraska T, Milo T. A sample-and-clean framework for fast and accurate query processing on dirty data. In: Proc. of the SIGMOD. 2014. 469–480. [doi: 10.1145/2588555.2610505]

- [173] Xu C, Xia F, Sharaf MA, Zhou MQ, Zhou AY. AQUAS: A quality-aware scheduler for NoSQL data stores. In: Proc. of the ICDE. 2014. 1210–1213. [doi: 10.1109/ICDE.2014.6816743]
- [174] Chen YC, Li JZ, Luo JZ. ITCI: An information theory based classification algorithm for incomplete data. In: Proc. of the WAIM. 2014. 167–179. [doi: 10.1007/978-3-319-08010-9_19]
- [175] Liu XL, Li JZ. Consistent estimation of query result in inconsistent data. Chinese Journal of Computers, 2015,38(9):1727–1738 (in Chinese with English abstract).
- [176] Razniewski S, Nutt W. Completeness of queries over incomplete databases. In: Proc. of the VLDB. 2011. 749–760.
- [177] Savković O, Paramita M, Paramonov S, Paramonov S, Nutt W. MAGIK: Managing completeness of data. In: Proc. of the 21st ACM Int'l Conf. on Information and Knowledge Management. ACM Press, 2012. 2725–2727. [doi: 10.1145/2396761.2398741]
- [178] Savkovic O, Mirza P, Tomasi A, Nutt W. Complete approximations of incomplete queries. In: Proc. of the VLDB. 2013. 1378–1381. [doi: 10.14778/2536274.2536320]
- [179] Nutt W, Razniewski S. Completeness of queries over SQL databases. In: Proc. of the 21st ACM Int'l Conf. on Information and Knowledge Management. ACM Press, 2012. 902–911. [doi: 10.1145/2396761.2396875]
- [180] Nutt W, Razniewski S, Vegliach G. Incomplete databases: Missing records and missing values. In: Proc. of the Database Systems for Advanced Applications. Berlin, Heidelberg: Springer-Verlag, 2012. 298–310. [doi: 10.1007/978-3-642-29023-7_30]
- [181] Darari F, Nutt W, Pirrò G, Razniewski S. Completeness statements about RDF data sources and their use for query answering. In: Proc. of the Semantic Web (ISWC 2013). Berlin, Heidelberg: Springer-Verlag, 2013. 66–83. [doi: 10.1007/978-3-642-41335-3_5]
- [182] Darari F, Prasojo RE, Nutt W. CORNER: A completeness reasoner for SPARQL queries over RDF data sources. In: Proc. of the Semantic Web: ESWC 2014 Satellite Events. Springer Int'l Publishing, 2014. 310–314. [doi: 10.1007/978-3-319-11955-7_40]
- [183] Paramonov S. Query completeness—A logic programming approach. Technical Report, KRDB13-2, KRDB Research Center, Free University Bozen-Bolzano, 2013. <http://www.inf.unibz.it/kldb/pub/tech-rep.php>
- [184] Nutt W, Paramonov S, Savkovic O. An ASP approach to query completeness reasoning. Theory and Practice of Logic Programming, 2013,13(4-5):1–10.
- [185] Nutt W, Paramonov S, Savkovic O. Implementing query completeness reasoning. In: Proc. of the 24th ACM Int'l Conf. on Information and Knowledge Management. ACM Press, 2015. 733–742. [doi: 10.1145/2806416.2806439]
- [186] Cao Y, Deng T, Fan W, Geerts F. On the data complexity of relative information completeness. Information Systems, 2014,45: 18–34. [doi: 10.1016/j.is.2014.04.001]
- [187] Razniewski S, Montali M, Nutt W. Verification of query completeness over processes. In: Proc. of the Business Process Management. Berlin, Heidelberg: Springer-Verlag, 2013. 155–170. [doi: 10.1007/978-3-642-40176-3_13]
- [188] Marengo E, Nutt W, Savkovic O. Towards a theory of query stability in business processes. In: Proc. of the 8th Alberto Mendelzon Workshop on Foundations of Data Management. Cartagena de Indias, 2014.
- [189] Savkovic O, Marengo E, Nutt W. Query stability in data-aware business processes [Extended Version]. In: Proc. of the CoRR. 2015.
- [190] Savkovic O, Marengo E, Nutt W. Query stability in monotonic data-aware business processes. In: Proc. of the ICDT. 2016.
- [191] Wang HZ, Li JZ, Huo R, Jia L, Jin L, Meng XY, Xie H. HITCleaner: A light-weight online data cleaning system. DASFAA, 2013,2: 481–484. [doi: 10.1007/978-3-642-37450-0_41]
- [192] Ortona S, Orsi G, Buoncristiano M, Furche T. WADaR: Joint wrapper and data repair. In: Proc. of the VLDB. 2015. 1996–2007. [doi: 10.14778/2824032.2824120]
- [193] Haas D, Krishnan S, Wang JN, Franklin MJ, Wu E. Wisteria: Nurturing scalable data cleaning infrastructure. In: Proc. of the VLDB. 2015. 2004–2015. [doi: 10.14778/2824032.2824122]
- [194] Bergman M, Milo T, Novgorodov S, Tan WC. Query-Oriented data cleaning with oracles. In: Proc. of the SIGMOD. 2015. 1199–1214. [doi: 10.1145/2723372.2737786]
- [195] Khayyat Z, Ilyas IF, Jindal A, Madden S, Ouzzani M, Papotti P, Quiané-Ruiz JA, Tang N, Yin S. Big dancing: A system for big data cleansing. In: Proc. of the SIGMOD. 2015. 1215–1230.
- [196] Chu X, Morcos J, Ilyas IF, Ouzzani M, Papotti P, Tang N, Ye Y. KATARa, a data cleaning system powered by knowledge bases and crowdsourcing. In: Proc. of the SIGMOD. 2015. 1247–1261. [doi: 10.1145/2723372.2749431]
- [197] Elmagarmid AK, Ilyas IF, Ouzzani M, Quiané-Ruiz JA, Tang N, Yin S. NADEEF/ER: Generic and interactive entity resolution. In: Proc. of the SIGMOD. 2014. 1071–1074. [doi: 10.1145/2588555.2594511]

- [198] Wang HZ, Li MD, Bu YY, Li JZ, Gao H, Zhang JC. Cleanix: A big data cleaning parfait. In: Proc. of the CIKM. 2014. 2024–2026. [doi: 10.1145/2661829.2661837]
- [199] Wang HZ, Li MD, Bu YY, Li JZ, Gao H, Zhang JC. Cleanix: A parallel big data cleaning system. SIGMOD Record, 2015,44(4): 35–40. [doi: 10.1145/2935694.2935702]
- [200] Stonebraker M, Bruckner D, Ilyas IF, Beskales G, Cherniack M, Zdonik SB, Pagan A, Xu S. Data curation at scale: The data tamer system. In: Proc. of the CIDR. 2013.
- [201] Wang HZ, Zhang XD, Li JZ, Gao H. ProductSeeker: Entity-Based product retrieval for e-commerce. In: Proc. of the SIGIR. 2013. 1085–1086. [doi: 10.1145/2484028.2484205]
- [202] Wang HZ, Liu XL, Li JZ, Tong X, Yang L, Li YK. EntityManager: An entity-based dirty data management system. DASFAA, 2013,2:468–471. [doi: 10.1007/978-3-642-37450-0_38]

附中文参考文献:

- [12] 李建中,刘显敏.大数据的一个重要方面:数据可用性.计算机研究与发展,2013,50(6):1147–1162.
- [13] 郭志懋,周傲英.数据质量和数据清洗研究综述.软件学报,2002,13(11):2076–2082. <http://www.jos.org.cn/1000-9825/20021103.htm>
- [20] 刘显敏,李建中.一种扩展条件函数依赖的发现算法.计算机研究与发展,2015,52(1):130–140.
- [21] 孙继洲,李建中.微函数依赖及其推理.计算机学报,录用待发表.
- [22] 苗东菁,刘显敏,李建中.概率数据库中近似函数依赖挖掘算法.计算机研究与发展,2015,52(12):2857–2865.
- [33] 李默涵,李建中,程思瑶.一种基于不确定规则的数据时效性判定方法.软件学报,2014,25(S2):147–156 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14033.htm>
- [51] 李默涵,李建中,高宏.数据时效性判定问题的求解算法.计算机学报,2012,35(11):2348–2360.
- [52] 刘永楠,邹兆年,李建中.数据完整性的评估方法.计算机研究与发展,2013,50(S1):230–238.
- [70] 张安珍,门雪莹,王宏志,李建中,高宏.大数据上基于 Hadoop 的不一致数据检测与修复算法.计算机科学与探索,2015,9(9): 1044–1055.
- [80] 李默涵,李建中.数据时效性修复问题的求解算法.计算机研究与发展,2015,52(9):1992–2001.
- [123] 丁小欧,王宏志,张笑影,李建中,高宏.数据质量多种性质的关联关系研究.软件学报,2016,27(7):1626–1644. <http://www.jos.org.cn/1000-9825/5040.htm> [doi: 10.13328/j.cnki.jos.005040]
- [125] 杨东华,李安宁,王宏志,李建中,高宏.基于任务合并的并行大数据清洗过程优化.计算机学报,2015,39(1):97–108.
- [160] 刘雪莉,王宏志,李建中,高宏.基于实体的相似性连接算法.软件学报,2015,26(6):1421–1437. <http://www.jos.org.cn/1000-9825/4610.htm> [doi: 10.13328/j.cnki.jos.004610]
- [175] 刘雪莉,李建中.不一致数据上查询结果的一致性估计.计算机学报,2015,38(9):1727–1738.



李建中(1950—),男,黑龙江哈尔滨人,博士,教授,博士生导师,主要研究领域为海量数据管理与计算,无线传感器网络,数据质量.



高宏(1966—),女,博士,教授,博士生导师,CCF 高级会员,主要研究领域为海量数据计算,无线传感器网络.



王宏志(1978—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为数据库,大数据,数据质量.