REFERENCES

- Agichtein, E & Ganti, V 2004, ‘Mining reference tables for automatic text segmentation’, Proceedings of the Association for Computing Machinery’s Special Interest Group on Knowledge Discovery and Data mining international conference on knowledge discovery and data mining, pp. 20–29, Seattle


- Bilenko, M & Mooney, R J 2002, ‘Learning to combine trained distance metrics for duplicate detection in databases’, Technical report, Computer Science Department, University of Texas, Austin


• Christen, P 2008, ‘Automatic Record Linkage Using Seeded Nearest Neighbor and Support Vector Machine Classification’, *Proceedings of Association for Computing Machinery’s Special Interest Group on Knowledge Discovery and Data mining*, pp.151-159, Las Vegas


• Christen, P & Goiser, K 2007, ‘Quality and Complexity Measures for Data Linkage and Deduplication’, *Studies in Computational Intelligence, Springer*, vol. 43, pp. 127-151


• Cohen, W 1998, ‘Integration of heterogeneous databases without common domains using queries based on textual similarity’, *Proceedings of Association for Computing Machinery’s Special Interest Group on Management of Data*, pp.201 – 212, Seattle

• Cohen, W W, Kautz, H A & McAllester, D A 2000, ‘Hardening soft information sources’, *Proceedings of the 6th Association for Computing Machinery’s Special Interest Group on Knowledge Discovery and Data mining International conference on Knowledge Discovery and Data mining*, pp. 255–259, Boston

• Cohen, W W, Ravikumar, P & Fienberg, S E 2003a, ‘A Comparison of String Distance Metrics for Name-Matching Tasks,” *Proceedings of the Association for Computing Machinery’s Workshop on Data Cleaning, Record Linkage and Object Identification*, pp. 73 – 78, Acapulco


• Cohen, W W, & Richman, J 2002, ‘Learning to Match and Cluster Large High-Dimensional Datasets for Data Integration’, *Proceedings of Association for Computing Machinery’s Special Interest Group on Knowledge Discovery and Data mining*, pp. 475-480, Edmonton

• Corral, A, Manolopoulos, Y, Theodoridis, Y & Vassilakopoulos, M 2000, ‘Closest pair queries in spatial databases’, *Association for Computing Machinery’s Special Interest Group on Management of Data*, pp. 189–200, Dallas


• Devendra Kumar Tiwary 2010, ‘Application of Data Mining In Customer Relationship Management (CRM)’, *Advances in Computational Sciences and Technology*, vol. 3, no. 4, pp. 527–540

• Do, H H & Rahm, E 2002, ‘A system for flexible combination of schema matching approaches’, *Proceedings of the 28th Very Large Data Bases Conference*, pp. 610-621, Hong Kong


• Faloutsos, C & Lin, K-I 1995, ‘Fastmap: A fast algorithm for indexing, data mining and visualization of traditional and multimedia datasets’, Association for Computing Machinery’s Special Interest Group on Management of Data, pp. 163–174, San Jose


• Koudas, N, Sarawagi, S, & Srivastava D 2006, ‘Record Linkage: Similarity Measures and Algorithms (Tutorial)’, *Proceedings of Association for Computing Machinery's Special Interest Group on Management of Data*, pp. 802-803

• Koudas, N & Sevcik, K C 1997, ‘Size separation spatial join’, *Association for Computing Machinery’s Special Interest Group on Management of Data*, pp. 324–335, Azusa


• Kukich, K 1992, ‘Techniques for automatically correcting words in text’, *Association for Computing Machinery’s Computing Surveys*, vol. 24, no. 4, pp. 377–439


• Lee, M L, Ling, T W & Low, W L 2000, ‘IntelliClean: a knowledge based intelligent data cleaner’, *Proceedings of the 6th Association for Computing Machinery’s Special Interest Group on Knowledge Discovery and Data mining International Conference on Knowledge Discovery and Data mining*, pp. 290-294, Boston
• Ley, M, DBLP bibliography, viewed 10 May 2012. http://dblp.uni-trier.de/pers/hs?q=SANG,
  http://dblp.uni-trier.de/pers/hd/c/Cha:Sang_Kyun.html


• Mamoulis, N & Papadias, D 1999, ‘Integration of spatial join algorithms for processing multiple inputs’, Proceedings of the Association for Computing Machinery’s Special Interest Group on Management of Data, pp. 1–12, Philadelphia


• McCallum, A, Nigam, K, & Ungar, L H 2000, ‘Efficient Clustering of High-Dimensional Datasets with Application to Reference Matching’, Proceedings of the 6th Association for Computing Machinery’s Special Interest Group on Knowledge Discovery and Data mining International conference on Knowledge Discovery and Data mining, pp. 169-178, Boston


- Sarawagi, S & Bhamidipaty, A 2002, ‘Interactive deduplication using active learning’, Proceedings of the 8th International Conference of Association for Computing Machinery’s Special Interest Group on Knowledge Discovery and Data mining, pp. 269-278, Edmonton


- Sevcik, K C & Koudas, N 2000, ‘High dimensional similarity joins: algorithms and performance evaluation’, Transactions on Knowledge & Data Engineering, vol. 12, no. 1, pp. 3–18


• Tejada, S, Knoblock, C A & Minton, S 2002, ‘Learning Domain-Independent String Transformation Weights for High Accuracy Object Identification’, *Proceedings of 8th Association for Computing Machinery’s Special Interest Group on Knowledge Discovery and Data Mining International Conference on Knowledge Discovery and Data Mining*, pp. 350 – 359, Edmonton


• Wang, J T L, Wang, X, Lin, K I, Shasha, D, Shapiro, B A & Zhang, K 1999, ‘Evaluating a class of distance-mapping algorithms for data mining and clustering’, *Association for Computing Machinery’s Special Interest Group on Knowledge Discovery and Data Mining*, pp. 307–311, San Diego


- Yu, H, Han, J, & Chang, C C 2004, ‘PEBL: Web Page Classification without Negative Examples’, IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 1, pp. 70-81

- Zhai & Liu, B 2006, ‘Structured Data Extraction from the Web Based on Partial Tree Alignment’, IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 12, pp. 1614-1628