Paper for Literature Review CS520

Boris Glavic

February 1, 2020

Contents

1 Overview

These are the options for the literature review for CS520. Each group will review one paper. You can access the pdfs for the papers here: `http://cs.iit.edu/~cs520/files/cs520-2020-spring-papers.zip`

2 Data Cleaning and Curation

- Cheng, R., Chen, J., & Xie, X. (2008). Cleaning Uncertain Data With Quality Guarantees. PVLDB, 1(1), 722–735. `http://dx.doi.org/10.14778/1453856.1453935`

- He, J., Veltri, E., Santoro, D., Li, G., Mecca, G., Papotti, P., & Tang, N. (2016). Interactive and deterministic data cleaning. In , Proceedings of the 2016 International Conference on Management of Data (pp. 893–907). : .

- Yu, Z., & Chu, X. (2019). PIclean: A probabilistic and interactive data cleaning system. In P. A. Boncz, S. Manegold, A. Ailamaki, A. Deshpande, & T. Kraska, Proceedings of the 2019 International Conference

on Management of Data, {SIGMOD} Conference 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019 (pp. 2021–2024). : ACM.

- Rekatsinas, T., Chu, X., Ilyas, I. F., & Christopher R\'e (2017). Holoclean: holistic data repairs with probabilistic inference. PVLDB, 10(11), 1190–1201. `http://dx.doi.org/10.14778/3137628.3137631`

- Qahtan, A. A., Tang, N., Ouzzani, M., Cao, Y., & Stonebraker, M. (2019). ANMAT: automatic knowledge discovery and error detection through pattern functional dependencies. In P. A. Boncz, S. Manegold, A. Ailamaki, A. Deshpande, & T. Kraska, Proceedings of the 2019 International Conference on Management of Data, {SIGMOD} Conference 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019. (pp. 1977–1980). : ACM.

- Heidari, A., McGrath, J., Ilyas, I. F., & Rekatsinas, T. (2019). Holodetect: few-shot learning for error detection. In P. A. Boncz, S. Manegold, A. Ailamaki, A. Deshpande, & T. Kraska, Proceedings of the 2019 International Conference on Management of Data, {SIGMOD} Conference 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019. (pp. 829–846). : ACM.

- Bertossi, L., & Chomicki, J. (2004). Query answering in inconsistent databases. In (Eds.), Logics for emerging applications of databases (pp. 43–83). : Springer.

- Raman, V., & Hellerstein, J. M. (2001). Potter's wheel: an interactive data cleaning system. In P. M. G. Apers, P. Atzeni, S. Ceri, S. Paraboschi, K. Ramamohanarao, & R. T. Snodgrass, VLDB} 2001, Proceedings of 27th International Conference on Very Large Data Bases, September 11-14, 2001, Roma, Italy (pp. 381–390). : Morgan Kaufmann.

3   Integration, Matching, and Mappings

- Atzeni, P., Bellomarini, L., Papotti, P., & Torlone, R. (2019). Metamappings for schema mapping reuse. PVLDB, 12(5), 557–569. `http://dx.doi.org/10.14778/3303753.3303761`

- Zhu, E., Deng, D., Nargesian, F., & Ren\'ee J. Miller (2019). JOSIE: overlap set similarity search for finding joinable tables in data lakes. In

P. A. Boncz, S. Manegold, A. Ailamaki, A. Deshpande, & T. Kraska, Proceedings of the 2019 International Conference on Management of Data, {SIGMOD} Conference 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019 (pp. 847–864). : ACM.

- Nargesian, F., Zhu, E., Pu, K. Q., & Ren\'ee J. Miller (2018). Table union search on open data. PVLDB, 11(7), 813–825. `http://dx.doi.org/10.14778/3192965.3192973`

4 Provenance

- Freire, J., & Chirigati, F. (2018). Provenance and the different flavors of computational reproducibility. Data Engineering, (), 15.

- Xu, J., Zhang, W., Alawini, A., & Tannen, V. (2018). Provenance analysis for missing answers and integrity repairs. Data Engineering, (), 39.

- Psallidas, F., & Wu, E. (2018). Smoke: fine-grained lineage at interactive speed. Proceedings of the VLDB Endowment, 11(6), 719–732.

- Senellart, P. (2018). Provenance and probabilities in relational databases. ACM SIGMOD Record, 46(4), 5–15.

- M\"uller, Tobias, Dietrich, B., & Grust, T. (2018). You say'what', i hear'where'and'why':(mis-) interpreting sql to derive fine-grained provenance. Proceedings of the VLDB Endowment, 11(11), 1536–1549.