

Bibliography

- Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of databases*, volume 8. Addison-Wesley Reading, 1995.
- Katrin Affolter, Kurt Stockinger, and Abraham Bernstein. A comparative survey of recent natural language interfaces for databases. *CoRR*, abs/1906.08990, 2019. URL <http://arxiv.org/abs/1906.08990>.
- Lars Ahrenberg, Arne Jnsson, and Nils Dahlbck. Discourse representation and discourse management for a natural language dialogue system. In *Proceeding of the Second Nordic Conference on Text Comprehension in Man and Machine*, 1990.
- Bogdan Alexe, Laura Chiticariu, and Wang Chiew Tan. SPIDER: a schema mapping debugger. In Umeshwar Dayal, Kyu-Young Whang, David B. Lomet, Gustavo Alonso, Guy M. Lohman, Martin L. Kersten, Sang Kyun Cha, and Young-Kuk Kim, editors, *Proceedings of the 32nd International Conference on Very Large Data Bases, Seoul, Korea, September 12-15, 2006*, pages 1179–1182. ACM, 2006. URL <http://dl.acm.org/citation.cfm?id=1164238>.
- Alibaba. Alibaba cloud tablestore. <https://www.alibabacloud.com/help/doc-detail/91977.htm>, 2021. [Online; accessed April-1-2021].
- Moustafa Alzantot, Yash Sharma, Ahmed Elgohary, Bo-Jhang Ho, Mani Srivastava, and Kai-Wei Chang. Generating natural language adversarial examples. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2890–2896, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1316. URL <https://aclanthology.org/D18-1316>.
- National Archives and Records Administration. The soundex indexing system, 2007. URL <https://www.archives.gov/research/census/soundex.html>.
- Yoav Artzi and Luke Zettlemoyer. Bootstrapping semantic parsers from conversations. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 421–432, Edinburgh, Scotland, UK., July 2011. Association for Computational Linguistics. URL <https://aclanthology.org/D11-1039>.
- Yoav Artzi and Luke Zettlemoyer. Weakly supervised learning of semantic parsers for mapping instructions to actions. *Transactions of the Association for Computational Linguistics*, 1:49–62, 2013. doi: 10.1162/tacl_a_00209. URL <https://aclanthology.org/Q13-1005>.

- Yoav Artzi, Nicholas Fitzgerald, and Luke Zettlemoyer. Semantic parsing with Combinatory Categorical Grammars. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts*, Doha, Qatar, October 2014. Association for Computational Linguistics. URL <https://aclanthology.org/D14-2003>.
- Kevin Atkinson. Gnu aspell, 2022. URL <http://aspell.net>. [Online; accessed May-4-2022].
- Ricardo A. Baeza-Yates, Carlos A. Hurtado, and Marcelo Mendoza. Query recommendation using query logs in search engines. In Wolfgang Lindner, Marco Mesiti, Can Türker, Yannis Tzitzikas, and Athena Vakali, editors, *Current Trends in Database Technology - EDBT 2004 Workshops, EDBT 2004 Workshops PhD, DataX, PIM, P2P&DB, and ClustWeb, Heraklion, Crete, Greece, March 14-18, 2004, Revised Selected Papers*, volume 3268 of *Lecture Notes in Computer Science*, pages 588–596. Springer, 2004. doi: 10.1007/978-3-540-30192-9_58. URL https://doi.org/10.1007/978-3-540-30192-9_58.
- Baidu. baidu.com, 2021. [Online; accessed November-19-2021].
- Christopher Baik. Fragment-driven natural language interaction with databases. In *10th Conference on Innovative Data Systems Research, CIDR 2020, Amsterdam, The Netherlands, January 12-15, 2020, Online Proceedings*. www.cidrdb.org, 2020. URL http://cidrdb.org/cidr2020/gongshow2020/gongshow/abstracts/cidr2020_abstract23.pdf.
- Christopher Baik, H. V. Jagadish, and Yunyao Li. Bridging the semantic gap with SQL query logs in natural language interfaces to databases. In *35th IEEE International Conference on Data Engineering, ICDE 2019, Macao, China, April 8-11, 2019*, pages 374–385. IEEE, 2019. doi: 10.1109/ICDE.2019.00041. URL <https://doi.org/10.1109/ICDE.2019.00041>.
- Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The berkeley framenet project. In Christian Boitet and Pete Whitelock, editors, *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, COLING-ACL 98, August 10-14, 1998, Université de Montréal, Montréal, Quebec, Canada. Proceedings of the Conference*, pages 86–90. Morgan Kaufmann Publishers / ACL, 1998. doi: 10.3115/980845.980860. URL <https://aclanthology.org/P98-1013/>.
- Anja Belz, Michael White, Dominic Espinosa, Eric Kow, Deirdre Hogan, and Amanda Stent. The first surface realisation shared task: Overview and evaluation results. In *Proceedings of the 13th European Workshop on Natural Language Generation*, pages 217–226, Nancy, France, September 2011. Association for Computational Linguistics. URL <https://aclanthology.org/W11-2832>.

- Mike Bennett. The financial industry business ontology: Best practice for big data. *Journal of Banking Regulation*, 14:255–268, 2013.
- Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. Semantic parsing on Freebase from question-answer pairs. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1533–1544, Seattle, Washington, USA, October 2013. Association for Computational Linguistics. URL <https://aclanthology.org/D13-1160>.
- Núria Bertomeu, Hans Uszkoreit, Anette Frank, Hans-Ulrich Krieger, and Brigitte Jörg. Contextual phenomena and thematic relations in database QA dialogues: results from a Wizard-of-Oz experiment. In *Proceedings of the Interactive Question Answering Workshop at HLT-NAACL 2006*, pages 1–8, New York, NY, USA, June 2006. Association for Computational Linguistics. URL <https://aclanthology.org/W06-3001>.
- Shreyas Bharadwaj, Laura Chiticariu, Marina Danilevsky, Samarth Dhingra, Samved Divekar, Arnaldo Carreno-Fuentes, Himanshu Gupta, Nitin Gupta, Sang-Don Han, Mauricio A. Hernández, Howard Ho, Parag Jain, Salil Joshi, Hima Karanam, Saravanan Krishnan, Rajasekar Krishnamurthy, Yunyao Li, Satishkumaar Manivannan, Ashish R. Mittal, Fatma Ozcan, Abdul Quamar, Poornima Raman, Diptikalyan Saha, Karthik Sankaranarayanan, Jaydeep Sen, Prithviraj Sen, Shivakumar Vaithyanathan, Mitesh Vasa, Hao Wang, and Huaiyu Zhu. Creation and interaction with large-scale domain-specific knowledge bases. *Proc. VLDB Endow.*, 10(12):1965–1968, 2017. doi: 10.14778/3137765.3137820. URL <http://www.vldb.org/pvldb/vol10/p1965-Danilevsky.pdf>.
- Nikita Bhutani, Kun Qian, Yunyao Li, H. V. Jagadish, Mauricio Hernandez, and Mitesh Vasa. Exploiting structure in representation of named entities using active learning. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 687–699, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/C18-1058>.
- Hervé Blanchon. A solution for the problem of interactive disambiguation. In *COLING*, 1992.
- Kurt D. Bollacker, Colin Evans, Praveen K. Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In Jason Tsong-Li Wang, editor, *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10-12, 2008*, pages 1247–1250. ACM, 2008. doi: 10.1145/1376616.1376746. URL <https://doi.org/10.1145/1376616.1376746>.
- Vadim Borisov, Tobias Leemann, Kathrin Seßler, Johannes Haug, Martin Pawelczyk, and Gjergji Kasneci. Deep neural networks and tabular data: A survey, 2021. URL <https://arxiv.org/abs/2110.01889>.

- Jordan Boyd-Graber, Samuel Carton, Shi Feng, Q. Vera Liao, Tania Lombrozo, Alison Smith-Renner, and Chenhao Tan. Human-centered evaluation of explanations. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Tutorial Abstracts*, pages 26–32, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-tutorials.4. URL <https://aclanthology.org/2022.naacl-tutorials.4>.
- Ralf D. Brown. Human-computer interaction for semantic disambiguation. In *COLING*, 1990.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020a.
- Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020b. URL <https://arxiv.org/abs/2005.14165>.
- Christopher Bryant, Mariano Felice, Øistein E. Andersen, and Ted Briscoe. The bea-2019 shared task on grammatical error correction. In *BEA@ACL*, 2019.
- Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. MultiWOZ - a large-scale multi-domain Wizard-of-Oz dataset for task-oriented dialogue modelling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, October–November 2018. Association for Computational Linguistics.
- Peter Buneman, Sanjeev Khanna, and Wang Chiew Tan. Why and where: A characterization of data provenance. In Jan Van den Bussche and Victor Vianu, editors, *Database Theory - ICDT 2001, 8th International Conference, London, UK, January 4-6, 2001, Proceedings*, volume 1973 of *Lecture Notes in Computer Science*, pages 316–330. Springer, 2001. doi: 10.1007/3-540-44503-X_20. URL https://doi.org/10.1007/3-540-44503-X_20.
- Michael J. Cafarella, Alon Halevy, Daisy Zhe Wang, Eugene Wu, and Yang Zhang. Webtables: Exploring the power of tables on the web. *Proc. VLDB Endow.*, 1(1): 538–549, August 2008. ISSN 2150-8097.

- Michael J. Cafarella, Alon Y. Halevy, Hongrae Lee, Jayant Madhavan, Cong Yu, Daisy Zhe Wang, and Eugene Wu. Ten years of webtables. *Proc. VLDB Endow.*, 11(12):2140–2149, 2018. doi: 10.14778/3229863.3240492. URL <http://www.vldb.org/pvldb/vol11/p2140-cafarella.pdf>.
- Fei Cai and M. de Rijke. A survey of query auto completion in information retrieval. *Found. Trends Inf. Retr.*, 10:273–363, 2016.
- Qingqing Cai and Alexander Yates. Large-scale semantic parsing via schema matching and lexicon extension. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 423–433, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL <https://aclanthology.org/P13-1042>.
- Ruisheng Cao, Lu Chen, Zhi Chen, Yanbin Zhao, Su Zhu, and Kai Yu. Lgesql: Line graph enhanced text-to-sql model with mixed local and non-local relations. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2541–2555, 2021.
- Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. *Readings in information visualization - using vision to think*. Academic Press, 1999. ISBN 978-1-55860-533-6.
- Thiago Castro Ferreira, Diego Moussallem, Emiel Krahmer, and Sander Wubben. Enriching the WebNLG corpus. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 171–176, Tilburg University, The Netherlands, November 2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-6521. URL <https://aclanthology.org/W18-6521>.
- Asli Celikyilmaz, Elizabeth Clark, and Jianfeng Gao. Evaluation of text generation: A survey, 2020. URL <https://arxiv.org/abs/2006.14799>.
- Joyce Y. Chai and Rong Jin. Discourse structure for context question answering. In *Proceedings of the Workshop on Pragmatics of Question Answering at HLT-NAACL 2004*, pages 23–30, Boston, Massachusetts, USA, May 2 - May 7 2004. Association for Computational Linguistics. URL <https://aclanthology.org/W04-2504>.
- Adriane Chapman and H. V. Jagadish. Why not? In Ugur Çetintemel, Stanley B. Zdonik, Donald Kossmann, and Nesime Tatbul, editors, *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2009, Providence, Rhode Island, USA, June 29 - July 2, 2009*, pages 523–534. ACM, 2009. doi: 10.1145/1559845.1559901. URL <https://doi.org/10.1145/1559845.1559901>.
- Eugene Charniak. A maximum-entropy-inspired parser. In *1st Meeting of the North American Chapter of the Association for Computational Linguistics*, 2000. URL <http://aclanthology.org/A00-2018>.

- Gloria Chatzopoulou, Magdalini Eirinaki, and Neoklis Polyzotis. Query recommendations for interactive database exploration. In Marianne Winslett, editor, *Scientific and Statistical Database Management, 21st International Conference, SSDBM 2009, New Orleans, LA, USA, June 2-4, 2009, Proceedings*, volume 5566 of *Lecture Notes in Computer Science*, pages 3–18. Springer, 2009. doi: 10.1007/978-3-642-02279-1_2. URL https://doi.org/10.1007/978-3-642-02279-1_2.
- Gloria Chatzopoulou, Magdalini Eirinaki, Suju Koshy, Sarika Mittal, Neoklis Polyzotis, and Jothi Swarubini Vindhiya Varman. The querie system for personalized query recommendations. *IEEE Data Eng. Bull.*, 34(2):55–60, 2011. URL <http://sites.computer.org/debull/A11june/Alkis.pdf>.
- Surajit Chaudhuri. An overview of query optimization in relational systems. In *Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, pages 34–43, 1998.
- Charles Chen and Razvan C. Bunescu. Context dependent semantic parsing over temporally structured data. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 3576–3585. Association for Computational Linguistics, 2019. doi: 10.18653/v1/n19-1360. URL <https://doi.org/10.18653/v1/n19-1360>.
- Danqi Chen and Christopher Manning. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 740–750, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1082. URL <https://aclanthology.org/D14-1082>.
- Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde, Jared Kaplan, Harrison Edwards, Yura Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray, Nick Ryder, Mikhail Pavlov, Alethea Power, Lukasz Kaiser, Mohammad Bavarian, Clemens Winter, Philippe Tillet, Felipe Petroski Such, David W. Cummings, Matthias Plappert, Fotios Chantzis, Elizabeth Barnes, Ariel Herbert-Voss, William H. Guss, Alex Nichol, Igor Babuschkin, S. Arun Balaji, Shantanu Jain, Andrew Carr, Jan Leike, Joshua Achiam, Vedant Misra, Evan Morikawa, Alec Radford, Matthew M. Knight, Miles Brundage, Mira Murati, Katie Mayer, Peter Welinder, Bob McGrew, Dario Amodei, Sam McCandlish, Ilya Sutskever, and Wojciech Zaremba. Evaluating large language models trained on code. *ArXiv*, abs/2107.03374, 2021a.
- Mingda Chen, Sam Wiseman, and Kevin Gimpel. WikiTableT: A large-scale data-to-text dataset for generating Wikipedia article sections. In *Findings of the Association*

- for Computational Linguistics: ACL-IJCNLP 2021*, pages 193–209, Online, August 2021b. Association for Computational Linguistics. doi: 10.18653/v1/2021.findings-acl.17. URL <https://aclanthology.org/2021.findings-acl.17>.
- Peter Pin-Shan Chen. The entity-relationship model—toward a unified view of data. *ACM transactions on database systems (TODS)*, 1(1):9–36, 1976.
- Wenhu Chen, Hongmin Wang, Jianshu Chen, Yunkai Zhang, Hong Wang, Shiyang Li, Xiyu Zhou, and William Yang Wang. Tabfact: A large-scale dataset for table-based fact verification, 2019. URL <https://arxiv.org/abs/1909.02164>.
- Wenhu Chen, Ming-Wei Chang, Eva Schlinger, William Wang, and William W. Cohen. Open question answering over tables and text, 2020a. URL <https://arxiv.org/abs/2010.10439>.
- Wenhu Chen, Jianshu Chen, Yu Su, Zhiyu Chen, and William Yang Wang. Logical natural language generation from open-domain tables. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7929–7942, Online, July 2020b. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.708. URL <https://aclanthology.org/2020.acl-main.708>.
- Wenhu Chen, Yu Su, Xifeng Yan, and William Yang Wang. KGPT: Knowledge-grounded pre-training for data-to-text generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8635–8648, Online, November 2020c. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.697. URL <https://aclanthology.org/2020.emnlp-main.697>.
- Wenhu Chen, Hanwen Zha, Zhiyu Chen, Wenhan Xiong, Hong Wang, and William Yang Wang. HybridQA: A dataset of multi-hop question answering over tabular and textual data. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1026–1036, Online, November 2020d. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.91. URL <https://aclanthology.org/2020.findings-emnlp.91>.
- Yang Chen, Jing Yang, and William Ribarsky. Toward effective insight management in visual analytics systems. In Peter Eades, Thomas Ertl, and Han-Wei Shen, editors, *IEEE Pacific Visualization Symposium PacificVis 2009, Beijing, China, April 20-23, 2009*, pages 49–56. IEEE Computer Society, 2009. doi: 10.1109/PACIFICVIS.2009.4906837. URL <https://doi.org/10.1109/PACIFICVIS.2009.4906837>.
- Zhiyu Chen, Wenhu Chen, Hanwen Zha, Xiyu Zhou, Yunkai Zhang, Sairam Sundaresan, and William Yang Wang. Logic2Text: High-fidelity natural language generation from logical forms. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2096–2111, Online, November 2020e. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.190. URL <https://aclanthology.org/2020.findings-emnlp.190>.

- James Cheney. Program slicing and data provenance. *IEEE Data Eng. Bull.*, 30(4): 22–28, 2007. URL <http://sites.computer.org/debull/A07dec/cheney.pdf>.
- Zhoujun Cheng, Haoyu Dong, Zhiruo Wang, Ran Jia, Jiaqi Guo, Yan Gao, Shi Han, Jian-Guang Lou, and Dongmei Zhang. HiTab: A hierarchical table dataset for question answering and natural language generation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1094–1110, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.78. URL <https://aclanthology.org/2022.acl-long.78>.
- Imran Chowdhury, Abdul Moeid, Enamul Hoque, Muhammad Ashad Kabir, Md. Sabir Hossain, and Mohammad Mainul Islam. Designing and evaluating multimodal interactions for facilitating visual analysis with dashboards. *IEEE Access*, 9:60–71, 2021. doi: 10.1109/ACCESS.2020.3046623. URL <https://doi.org/10.1109/ACCESS.2020.3046623>.
- Shumo Chu, Chenglong Wang, Konstantin Weitz, and Alvin Cheung. Cosette: An automated prover for sql. In *Proc. of the CIDR Conference*, 2017.
- Shumo Chu, Brendan Murphy, Jared Roesch, Alvin Cheung, and Dan Suciu. Axiomatic foundations and algorithms for deciding semantic equivalences of sql queries. *Proceedings of the VLDB Endowment*, 11(11), 2018.
- Alonzo Church. A set of postulates for the foundation of logic. *Annals of Mathematics*, 2:346–366, 1932.
- James Clarke and Sebastian Riedel, editors. *Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing*, Boulder, Colorado, June 2009. Association for Computational Linguistics. URL <https://aclanthology.org/W09-1800>.
- Jean claude Martin. Tycoon: Theoretical framework and software tools for multimodal interfaces. In *Intelligent Multimodality Multimedia Interfaces*, 1997.
- E. F. Codd. Seven steps to rendezvous with the casual user. In J. W. Klimbie and K. L. Koffeman, editors, *Data Base Management, Proceeding of the IFIP Working Conference Data Base Management, Cargèse, Corsica, France, April 1-5, 1974*, pages 179–200. North-Holland, January 1974.
- Edgar F. Codd. A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387, 1979.
- Philip R. Cohen, Mary Dalrymple, Douglas B. Moran, Fernando C Pereira, Joseph W. Sullivan, Robert A. Gargan, Jon Schlossberg, and Sherman W. Tyler. Synergistic use of direct manipulation and natural language. In *CHI '89*, 1989.

- Michael Collins. Three generative, lexicalised models for statistical parsing. In *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, pages 16–23, Madrid, Spain, July 1997. Association for Computational Linguistics. doi: 10.3115/976909.979620. URL <https://aclanthology.org/P97-1003>.
- Mariano P Consens and Alberto O Mendelzon. Expressing structural hypertext queries in graphlog. In *Proceedings of the second annual ACM conference on Hypertext*, pages 269–292, 1989.
- Pietro Crovari, Sara Pidò, Pietro Pinoli, Anna Bernasconi, Arif Canakoglu, Franca Garzotto, and Stefano Ceri. Gecoagent: A conversational agent for empowering genomic data extraction and analysis. *ACM Trans. Comput. Heal.*, 3(1):3:1–3:29, 2022. doi: 10.1145/3464383. URL <https://doi.org/10.1145/3464383>.
- Silviu Cucerzan and Eric Brill. Spelling correction as an iterative process that exploits the collective knowledge of web users. In *EMNLP*, pages 293–300. ACL, 2004.
- Richard Cyganiak, David Wood, Markus Lanthaler, Graham Klyne, Jeremy J Carroll, and Brian McBride. Rdf 1.1 concepts and abstract syntax. *W3C recommendation*, 25 (02):1–22, 2014.
- Deborah A. Dahl, Madeleine Bates, Michael Brown, William Fisher, Kate Hunicke-Smith, David Pallett, Christine Pao, Alexander Rudnicky, and Elizabeth Shriber. Expanding the scope of the ATIS task: The ATIS-3 corpus. *Proceedings of the workshop on Human Language Technology*, pages 43–48, 1994. URL <http://dl.acm.org/citation.cfm?id=1075823>.
- Veronica Dahl and Harvey Abramson. *Logic Grammars*. Springer, 1989.
- Xinyan Dai, Xiao Yan, Kaiwen Zhou, Yuxuan Wang, Han Yang, and James Cheng. Convolutional embedding for edit distance. In *SIGIR*, pages 599–608, 2020.
- Fred Damerau. A technique for computer detection and correction of spelling errors. *Commun. ACM*, 7(3):171–176, 1964.
- Marco Damonte. *Understanding and generating language with abstract meaning representation*. PhD thesis, 2020.
- Marina Danilevsky, Kun Qian, Ranit Aharonov, Yannis Katsis, Ban Kawas, and Prithviraj Sen. A survey of the state of explainable AI for natural language processing. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 447–459, Suzhou, China, December 2020. Association for Computational Linguistics. URL <https://aclanthology.org/2020.aacl-main.46>.

- Marina Danilevsky, Shipi Dhanorkar, Yunyao Li, Lucian Popa, Kun Qian, and Anbang Xu. Explainability for natural language processing. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 4033–4034, 2021.
- Donald Davidson. The individuation of events. In Nicholas Rescher, editor, *Essays in Honor of Carl G. Hempel*, pages 216–234. Reidel, 1969.
- Martin Davis. *The undecidable: Basic papers on undecidable propositions, unsolvable problems and computable functions*. Courier Corporation, 2004.
- Mostafa Dehghani, Sascha Rothe, Enrique Alfonseca, and Pascal Fleury. Learning to attend, copy, and generate for session-based query suggestion. In Ee-Peng Lim, Marianne Winslett, Mark Sanderson, Ada Wai-Chee Fu, Jimeng Sun, J. Shane Culpepper, Eric Lo, Joyce C. Ho, Debora Donato, Rakesh Agrawal, Yu Zheng, Carlos Castillo, Aixin Sun, Vincent S. Tseng, and Chenliang Li, editors, *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*, pages 1747–1756. ACM, 2017. doi: 10.1145/3132847.3133010. URL <https://doi.org/10.1145/3132847.3133010>.
- Naihao Deng, Yulong Chen, and Yue Zhang. Recent advances in text-to-SQL: A survey of what we have and what we expect. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2166–2187, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics. URL <https://aclanthology.org/2022.coling-1.190>.
- Xiang Deng, Huan Sun, Alyssa Lees, You Wu, and Cong Yu. TURL: table understanding through representation learning. *CoRR*, abs/2006.14806, 2020. URL <https://arxiv.org/abs/2006.14806>.
- Xiang Deng, Ahmed Hassan Awadallah, Christopher Meek, Oleksandr Polozov, Huan Sun, and Matthew Richardson. Structure-grounded pretraining for text-to-SQL. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1337–1350, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.105. URL <https://aclanthology.org/2021.naacl-main.105>.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL <https://aclanthology.org/N19-1423>.

- Kedar Dhamdhere, Kevin S. McCurley, Ralfi Nahmias, Mukund Sundararajan, and Qiqi Yan. Analyza: Exploring data with conversation. In George A. Papadopoulos, Tsvi Kuflik, Fang Chen, Carlos Duarte, and Wai-Tat Fu, editors, *Proceedings of the 22nd International Conference on Intelligent User Interfaces, IUI 2017, Limassol, Cyprus, March 13-16, 2017*, pages 493–504. ACM, 2017. doi: 10.1145/3025171.3025227. URL <https://doi.org/10.1145/3025171.3025227>.
- Shipi Dhanorkar, Christine T. Wolf, Kun Qian, Anbang Xu, Lucian Popa, and Yunyao Li. Who needs to know what, when?: Broadening the explainable AI (XAI) design space by looking at explanations across the AI lifecycle. In Wendy Ju, Lora Oehlberg, Sean Follmer, Sarah E. Fox, and Stacey Kuznetsov, editors, *DIS '21: Designing Interactive Systems Conference 2021, Virtual Event, USA, 28 June, July 2, 2021*, pages 1591–1602. ACM, 2021. doi: 10.1145/3461778.3462131. URL <https://doi.org/10.1145/3461778.3462131>.
- Chenhe Dong, Yinghui Li, Haifan Gong, Miaoxin Chen, Junxin Li, Ying Shen, and Min Yang. A survey of natural language generation. *CoRR*, abs/2112.11739, 2021. URL <https://arxiv.org/abs/2112.11739>.
- Haoyu Dong, Zhoujun Cheng, Xinyi He, Mengyu Zhou, Anda Zhou, Fan Zhou, Ao Liu, Shi Han, and Dongmei Zhang. Table pre-training: A survey on model architectures, pre-training objectives, and downstream tasks. In Lud De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 5426–5435. International Joint Conferences on Artificial Intelligence Organization, 7 2022. doi: 10.24963/ijcai.2022/761. URL <https://doi.org/10.24963/ijcai.2022/761>. Survey Track.
- Li Dong and Mirella Lapata. Language to logical form with neural attention. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 33–43, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1004. URL <https://aclanthology.org/P16-1004>.
- Li Dong and Mirella Lapata. Coarse-to-fine decoding for neural semantic parsing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 731–742, Melbourne, Australia, July 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-1068. URL <https://aclanthology.org/P18-1068>.
- Timothy Dozat and Christopher D. Manning. Deep biaffine attention for neural dependency parsing, 2016. URL <https://arxiv.org/abs/1611.01734>.
- Ondřej Dušek, Jekaterina Novikova, and Verena Rieser. Findings of the E2E NLG challenge. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 322–328, Tilburg University, The Netherlands, November

2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-6539. URL <https://aclanthology.org/W18-6539>.
- Ondřej Dušek, Jekaterina Novikova, and Verena Rieser. Evaluating the state-of-the-art of end-to-end natural language generation: The e2e NLG challenge. *Computer Speech and Language*, 59:123–156, jan 2020. doi: 10.1016/j.csl.2019.06.009. URL <https://doi.org/10.1016%2Fj.csl.2019.06.009>.
- Ahmed Elgohary, Saghar Hosseini, and Ahmed H. Awadallah. Speak to your parser: Interactive text-to-sql with natural language feedback. In *Annual Conference of the Association for Computational Linguistics (ACL 2020)*, July 2020. URL <https://www.microsoft.com/en-us/research/publication/speak-to-your-parser-interactive-text-to-sql-with-natural-language-feedback/>.
- Michael Elhadad. The fuf functional unifier: User’s manual. Technical report, Technical Report CUCS-408-88, Columbia University, 1988.
- Erkut Erdem, Menekse Kuyu, Semih Yagcioglu, Anette Frank, Letitia Parcalabescu, Barbara Plank, Andrii Babii, Oleksii Turuta, Aykut Erdem, Iacer Calixto, Elena Lloret, Elena-Simona Apostol, Ciprian-Octavian Truică, Branislava Šandrih, Sanda Martinčić-Ipšić, Gábor Berend, Albert Gatt, and Grăzina Korvel. Neural natural language generation: A survey on multilinguality, multimodality, controllability and learning. *JAIR*, 2022.
- Dan Farber. Google search scratches its brain 500 million times a day, 2013. [Online: <https://www.cnet.com/tech/services-and-software/google-search-scratches-its-brain-500-million-times-a-day/>].
- Catherine Finegan-Dollak, Jonathan K. Kummerfeld, Li Zhang, Karthik Ramanathan, Sesh Sadasivam, Rui Zhang, and Dragomir Radev. Improving text-to-sql evaluation methodology. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 351–360, July 2018. URL <http://aclweb.org/anthology/P18-1033>.
- Nicolas Fiorini and Zhiyong Lu. Personalized neural language models for real-world query auto completion. In Srinivas Bangalore, Jennifer Chu-Carroll, and Yunyao Li, editors, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 3 (Industry Papers)*, pages 208–215. Association for Computational Linguistics, 2018. doi: 10.18653/v1/n18-3026. URL <https://doi.org/10.18653/v1/n18-3026>.
- Jeffrey Flanigan. *Parsing and Generation for the Abstract Meaning Representation*. PhD thesis, 2018.

- Pascale Fung, Yun-Nung (Vivian) Chen, Zhaojiang Lin, and Andrea Madotto. Neurips 2020 tutorial in deeper conversational ai. In *Tutorials of NeurIPS Thirty-fourth Annual Conference on Neural Information Processing Systems*, 2020.
- Yujian Gan, Xinyun Chen, Qiuping Huang, Matthew Purver, John R Woodward, Jinxia Xie, and Pengsheng Huang. Towards robustness of text-to-sql models against synonym substitution. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2505–2515, 2021a.
- Yujian Gan, Xinyun Chen, Jinxia Xie, Matthew Purver, John R Woodward, John Drake, and Qiaofu Zhang. Natural sql: Making sql easier to infer from natural language specifications. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2030–2042, 2021b.
- Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, and Karrie G. Karahalios. Datatone: Managing ambiguity in natural language interfaces for data visualization. In Celine Latulipe, Bjoern Hartmann, and Tovi Grossman, editors, *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, UIST 2015, Charlotte, NC, USA, November 8–11, 2015*, pages 489–500. ACM, 2015. doi: 10.1145/2807442.2807478. URL <https://doi.org/10.1145/2807442.2807478>.
- Cristina Garbacea and Qiaozhu Mei. Neural language generation: Formulation, methods, and evaluation, 2020. URL <https://arxiv.org/abs/2007.15780>.
- Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. The WebNLG challenge: Generating text from RDF data. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 124–133, Santiago de Compostela, Spain, September 2017. Association for Computational Linguistics. doi: 10.18653/v1/W17-3518. URL <https://aclanthology.org/W17-3518>.
- Albert Gatt and Emiel Krahmer. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation, 2017. URL <https://arxiv.org/abs/1703.09902>.
- Sebastian Gehrmann, Tosin Adewumi, Karmanya Aggarwal, Pawan Sasanka Ammanamanchi, Anuoluwapo Aremu, Antoine Bosselut, Khyathi Raghavi Chandu, Miruna-Adriana Clinciu, Dipanjan Das, Kaustubh Dhole, Wanyu Du, Esin Durmus, Ondřej Dušek, Chris Chinenye Emezue, Varun Gangal, Cristina Garbacea, Tatsunori Hashimoto, Yufang Hou, Yacine Jernite, Harsh Jhamtani, Yangfeng Ji, Shailza Jolly, Mihir Kale, Dhruv Kumar, Faisal Ladhak, Aman Madaan, Mounica Maddela, Khyati Mahajan, Saad Mahamood, Bodhisattwa Prasad Majumder, Pedro Henrique Martins, Angelina McMillan-Major, Simon Mille, Emiel van Miltenburg, Moin Nadeem, Shashi Narayan, Vitaly Nikolaev, Andre Niyongabo Rubungo, Salomey Osei, Ankur Parikh, Laura Perez-Beltrachini, Niranjan Ramesh Rao, Vikas Raunak, Juan Diego

- Rodriguez, Sashank Santhanam, João Sedoc, Thibault Sellam, Samira Shaikh, Anastasia Shimorina, Marco Antonio Sobrevilla Cabezero, Hendrik Strobelt, Nishant Subramani, Wei Xu, Diyi Yang, Akhila Yerukola, and Jiawei Zhou. The GEM benchmark: Natural language generation, its evaluation and metrics. In *Proceedings of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021)*, pages 96–120, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.gem-1.10. URL <https://aclanthology.org/2021.gem-1.10>.
- Sebastian Gehrmann, Abhik Bhattacharjee, Abinaya Mahendiran, Alex Wang, Alexandros Papangelis, Aman Madaan, Angelina McMillan-Major, Anna Shvets, Ashish Upadhyay, Bingsheng Yao, Bryan Wilie, Chandra Bhagavatula, Chaobin You, Craig Thomson, Cristina Garbacea, Dakuo Wang, Daniel Deutsch, Deyi Xiong, Di Jin, Dimitra Gkatzia, Dragomir Radev, Elizabeth Clark, Esin Durmus, Faisal Ladhak, Filip Ginter, Genta Indra Winata, Hendrik Strobelt, Hiroaki Hayashi, Jekaterina Novikova, Jenna Kanerva, Jenny Chim, Jiawei Zhou, Jordan Clive, Joshua Maynez, João Sedoc, Juraj Juraska, Kaustubh Dhole, Khyathi Raghavi Chandu, Laura Perez-Beltrachini, Leonardo F. R. Ribeiro, Lewis Tunstall, Li Zhang, Mahima Pushkarna, Mathias Creutz, Michael White, Mihir Sanjay Kale, Moussa Kamal Eddine, Nico Daheim, Nishant Subramani, Ondrej Dusek, Paul Pu Liang, Pawan Sasanka Ammanamanchi, Qi Zhu, Ratish Puduppully, Reno Kriz, Rifat Shahriyar, Ronald Cardenas, Saad Mahamood, Salomey Osei, Samuel Cahyawijaya, Sanja Štajner, Sebastien Montella, Shailza, Shailza Jolly, Simon Mille, Tahmid Hasan, Tianhao Shen, Tosin Adewumi, Vikas Raunak, Vipul Raheja, Vitaly Nikolaev, Vivian Tsai, Yacine Jernite, Ying Xu, Yisi Sang, Yixin Liu, and Yufang Hou. Gemv2: Multilingual nlg benchmarking in a single line of code, 2022. URL <https://arxiv.org/abs/2206.11249>.
- Alessandra Giordani and Alessandro Moschitti. Automatic generation and reranking of sql-derived answers to nl questions. In *Proceedings of the Second International Conference on Trustworthy Eternal Systems via Evolving Software, Data and Knowledge*, pages 59–76, 2012. URL https://doi.org/10.1007/978-3-642-45260-4_5.
- Orest Gkini, Theofilos Belmpas, Georgia Koutrika, and Yannis E. Ioannidis. An in-depth benchmarking of text-to-sql systems. In Guoliang Li, Zhanhuai Li, Stratos Idreos, and Divesh Srivastava, editors, *SIGMOD '21: International Conference on Management of Data, Virtual Event, China, June 20-25, 2021*, pages 632–644, 2021.
- Boris Glavic, Gustavo Alonso, Renée J. Miller, and Laura M. Haas. TRAMP: understanding the behavior of schema mappings through provenance. *Proc. VLDB Endow.*, 3(1):1314–1325, 2010. doi: 10.14778/1920841.1921003. URL http://www.vldb.org/pvldb/vldb2010/pvldb_vol13/R116.pdf.
- Boris Glavic, Alexandra Meliou, and Sudeepa Roy. Trends in explanations: Understanding and debugging data-driven systems. *Found. Trends Databases*, 11(3):226–318, 2021. doi: 10.1561/19000000074. URL <https://doi.org/10.1561/19000000074>.

- Google. google.com, 2021. [Online; accessed November-19-2021].
- Google. How our quality raters make search results better. <https://support.google.com/websearch/answer/9281931?hl=en>, 2022. [Online; accessed 18-August-2022].
- Goetz Graefe. Query evaluation techniques for large databases. *ACM Computing Surveys (CSUR)*, 25(2):73–169, 1993.
- Lars Grammel, Melanie Tory, and Margaret-Anne Storey. How information visualization novices construct visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):943–952, 2010. doi: 10.1109/TVCG.2010.164.
- Jim Gray. *Benchmark handbook: for database and transaction processing systems*. Morgan Kaufmann Publishers Inc., 1992.
- Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Comput. Linguistics*, 12:175–204, 1986.
- Jiaqi Guo, Zecheng Zhan, Yan Gao, Yan Xiao, Jian-Guang Lou, Ting Liu, and Dongmei Zhang. Towards complex text-to-SQL in cross-domain database with intermediate representation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4524–4535, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1444. URL <https://aclanthology.org/P19-1444>.
- Joseph Y. Halpern and Judea Pearl. Causes and explanations: A structural-model approach: Part 1: Causes. In Jack S. Breese and Daphne Koller, editors, *UAI '01: Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence, University of Washington, Seattle, Washington, USA, August 2-5, 2001*, pages 194–202. Morgan Kaufmann, 2001. URL https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=100&proceeding_id=17.
- Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques, 3rd edition*. Morgan Kaufmann, 2011. ISBN 978-0123814791. URL <http://hanj.cs.illinois.edu/bk3/>.
- Moshe Hazoom, Vibhor Malik, and Ben Bogin. Text-to-sql in the wild: A naturally-occurring dataset based on stack exchange data. In *Proceedings of the 1st Workshop on Natural Language Processing for Programming (NLP4Prog 2021)*, pages 77–87, 2021.
- Qi He, Daxin Jiang, Zhen Liao, Steven C. H. Hoi, Kuiyu Chang, Ee-Peng Lim, and Hang Li. Web query recommendation via sequential query prediction. In Yannis E. Ioannidis, Dik Lun Lee, and Raymond T. Ng, editors, *Proceedings of the 25th International Conference on Data Engineering, ICDE 2009, March 29 2009 - April 2 2009, Shanghai, China*, pages 1443–1454. IEEE Computer Society, 2009. doi: 10.1109/ICDE.2009.71. URL <https://doi.org/10.1109/ICDE.2009.71>.

- Marti Hearst. *Search User Interface*. Cambridge University Press, 2009.
- Marti Hearst and Melanie Tory. Would you like A chart with that? incorporating visualizations into conversational interfaces. In *30th IEEE Visualization Conference, IEEE VIS 2019 - Short Papers, Vancouver, BC, Canada, October 20-25, 2019*, pages 36–40. IEEE, 2019. doi: 10.1109/VISUAL.2019.8933766. URL <https://doi.org/10.1109/VISUAL.2019.8933766>.
- Jeffrey Heer, Frank van Ham, M. Sheelagh T. Carpendale, Chris Weaver, and Petra Isenberg. Creation and collaboration: Engaging new audiences for information visualization. In *Information Visualization*, 2008.
- Charles T Hemphill, John J Godfrey, and George R Doddington. The atis spoken language systems pilot corpus. In *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27, 1990*, 1990.
- Jonathan Herzig, Pawel Krzysztof Nowak, Thomas Müller, Francesco Piccinno, and Julian Eisenschlos. TaPas: Weakly supervised table parsing via pre-training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4320–4333, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.398. URL <https://aclanthology.org/2020.acl-main.398>.
- Jonathan Herzig, Thomas Müller, Syrine Krichene, and Julian Eisenschlos. Open domain question answering over tables via dense retrieval. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 512–519, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.43. URL <https://aclanthology.org/2021.naacl-main.43>.
- James Higginbotham. On semantics. *Linguistic Inquiry*, 16:547–593, 1985.
- Graeme Hirst. *Semantic Interpretation and the Resolution of Ambiguity*. Studies in natural language processing. Cambridge University Press, 1992.
- Daniel Hládek, Ján Staš, and Matúš Pleva. Survey of automatic spelling correction. *Electronics*, 9(10), 2020. ISSN 2079-9292. URL <https://www.mdpi.com/2079-9292/9/10/1670>.
- Enamul Hoque, Vidya Setlur, Melanie Tory, and Isaac Dykeman. Applying pragmatics principles for interaction with visual analytics. *IEEE Trans. Vis. Comput. Graph.*, 24(1):309–318, 2018. doi: 10.1109/TVCG.2017.2744684. URL <https://doi.org/10.1109/TVCG.2017.2744684>.
- Ehsan Hosseini-Asl, Bryan McCann, Chien-Sheng Wu, Semih Yavuz, and Richard Socher. A simple language model for task-oriented dialogue. *ArXiv*, abs/2005.00796, 2020.

- Bo-June Paul Hsu and Giuseppe Ottaviano. Space-efficient data structures for top-k completion. *Proceedings of the 22nd international conference on World Wide Web*, 2013.
- Yushi Hu, Chia-Hsuan Lee, Tianbao Xie, Tao Yu, Noah A Smith, and Mari Ostendorf. In-context learning for few-shot dialogue state tracking. *arXiv preprint arXiv:2203.08568*, 2022.
- Binyuan Hui, Ruiying Geng, Lihan Wang, Bowen Qin, Yanyang Li, Bowen Li, Jian Sun, and Yongbin Li. S2sql: Injecting syntax to question-schema interaction graph encoder for text-to-sql parsers. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1254–1262, 2022.
- Wonseok Hwang, Jinyeong Yim, Seunghyun Park, and Minjoon Seo. A comprehensive exploration on wikisql with table-aware word contextualization, 2019. URL <https://arxiv.org/abs/1902.01069>.
- Hiroshi Iida, Dung Thai, Varun Manjunatha, and Mohit Iyyer. TABBIE: Pretrained representations of tabular data. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3446–3456, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.270. URL <https://aclanthology.org/2021.naacl-main.270>.
- Touseef Iqbal and Shaima Qureshi. The survey: Text generation models in deep learning. *Journal of King Saud University - Computer and Information Sciences*, 34(6, Part A):2515–2528, 2022. ISSN 1319-1578. doi: <https://doi.org/10.1016/j.jksuci.2020.04.001>. URL <https://www.sciencedirect.com/science/article/pii/S1319157820303360>.
- Srinivasan Iyer, Ioannis Konstas, Alvin Cheung, Jayant Krishnamurthy, and Luke Zettlemoyer. Learning a neural semantic parser from user feedback. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 963–973, 2017a. URL <http://www.aclweb.org/anthology/P17-1089>.
- Srinivasan Iyer, Ioannis Konstas, Alvin Cheung, Jayant Krishnamurthy, and Luke Zettlemoyer. Learning a neural semantic parser from user feedback. In *55th Annual Meeting of the Association for Computational Linguistics*, 2017b.
- Mohit Iyyer, Varun Manjunatha, Jordan Boyd-Graber, and Hal Daumé III. Deep unordered composition rivals syntactic methods for text classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1681–1691, Beijing, China, July 2015. Association for Computational Linguistics. doi: 10.3115/v1/P15-1162. URL <https://aclanthology.org/P15-1162>.

- Mohit Iyyer, Wen-tau Yih, and Ming-Wei Chang. Search-based neural structured learning for sequential question answering. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1821–1831, Vancouver, Canada, July 2017.
- Bernard J. Jansen, Amanda Spink, Chris Blakely, and Sherry Koshman. Defining a session on web search engines. *J. Assoc. Inf. Sci. Technol.*, 58(6):862–871, 2007. doi: 10.1002/asi.20564. URL <https://doi.org/10.1002/asi.20564>.
- Matthias Jarke and Jurgen Koch. Query optimization in database systems. *ACM Computing surveys (Csur)*, 16(2):111–152, 1984.
- Sai Muralidhar Jayanthi, Danish Pruthi, and Graham Neubig. NeuSpell: A neural spelling correction toolkit. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 158–164, Online, October 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-demos.21. URL <https://www.aclweb.org/anthology/2020.emnlp-demos.21>.
- Thomas C Jepsen. Just what is an ontology, anyway? *IT Professional Magazine*, 11(5): 22, 2009.
- HanQi Jin, Yue Cao, TianMing Wang, XinYu Xing, and XiaoJun Wan. Recent advances of neural text generation: Core tasks, datasets, models and challenges. *Science China Technological Sciences*, 63(10), 2020a.
- Zhijing Jin, Qipeng Guo, Xipeng Qiu, and Zheng Zhang. GenWiki: A dataset of 1.3 million content-sharing text and graphs for unsupervised graph-to-text generation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2398–2409, Barcelona, Spain (Online), December 2020b. International Committee on Computational Linguistics. doi: 10.18653/v1/2020.coling-main.217. URL <https://aclanthology.org/2020.coling-main.217>.
- Rogers Jeffrey Leo John, Navneet Potti, and Jignesh M. Patel. Ava: From data to insights through conversations. In *8th Biennial Conference on Innovative Data Systems Research, CIDR 2017, Chaminade, CA, USA, January 8-11, 2017, Online Proceedings*. www.cidrdb.org, 2017. URL <http://cidrdb.org/cidr2017/papers/p87-john-cidr17.pdf>.
- Rosie Jones, Benjamin Rey, Omid Madani, and Wiley Greiner. Generating query substitutions. In Les Carr, David De Roure, Arun Iyengar, Carole A. Goble, and Michael Dahlin, editors, *Proceedings of the 15th international conference on World Wide Web, WWW 2006, Edinburgh, Scotland, UK, May 23-26, 2006*, pages 387–396. ACM, 2006. doi: 10.1145/1135777.1135835. URL <https://doi.org/10.1145/1135777.1135835>.

- Dan Jurafsky and James H. Martin. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition, 2nd Edition*. Prentice Hall series in artificial intelligence. Prentice Hall, Pearson Education International, 2009. ISBN 9780135041963. URL <https://www.worldcat.org/oclc/315913020>.
- Mihir Kale and Abhinav Rastogi. Text-to-text pre-training for data-to-text tasks. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 97–102, Dublin, Ireland, December 2020. Association for Computational Linguistics. URL <https://aclanthology.org/2020.inlg-1.14>.
- Young Mo Kang, Wenhao Liu, and Yingbo Zhou. Queryblazer: Efficient query auto-completion framework. In Liane Lewin-Eytan, David Carmel, Elad Yom-Tov, Eugene Agichtein, and Evgeniy Gabrilovich, editors, *WSDM '21, The Fourteenth ACM International Conference on Web Search and Data Mining, Virtual Event, Israel, March 8-12, 2021*, pages 1020–1028. ACM, 2021. doi: 10.1145/3437963.3441725. URL <https://doi.org/10.1145/3437963.3441725>.
- George Katsogiannis-Meimarakis and Georgia Koutrika. Deep learning approaches for text-to-sql systems. In Yannis Velegrakis, Demetris Zeinalipour-Yazti, Panos K. Chrysanthis, and Francesco Guerra, editors, *Proceedings of the 24th International Conference on Extending Database Technology, EDBT 2021, Nicosia, Cyprus, March 23 - 26, 2021*, pages 710–713. OpenProceedings.org, 2021a. doi: 10.5441/002/edbt.2021.90. URL <https://doi.org/10.5441/002/edbt.2021.90>.
- George Katsogiannis-Meimarakis and Georgia Koutrika. A deep dive into deep learning approaches for text-to-sql systems. In Guoliang Li, Zhanhuai Li, Stratos Idreos, and Divesh Srivastava, editors, *SIGMOD '21: International Conference on Management of Data, Virtual Event, China, June 20-25, 2021*, pages 2846–2851. ACM, 2021b. doi: 10.1145/3448016.3457543. URL <https://doi.org/10.1145/3448016.3457543>.
- George Katsogiannis-Meimarakis and Georgia Koutrika. A survey on deep learning approaches for text-to-sql. *The VLDB Journal*, 2023.
- Mark D. Kernighan, Kenneth Ward Church, and William A. Gale. A spelling correction program based on a noisy channel model. In *13th International Conference on Computational Linguistics, COLING 1990, University of Helsinki, Finland, August 20-25, 1990*, pages 205–210, 1990.
- Nitish Shirish Keskar, Bryan McCann, Lav R. Varshney, Caiming Xiong, and Richard Socher. Ctrl: A conditional transformer language model for controllable generation, 2019. URL <https://arxiv.org/abs/1909.05858>.
- Michael Kifer, Arthur J Bernstein, and Philip M Lewis. *Database systems: an application-oriented approach*. Pearson Education India, 2005.

- Hyeonji Kim, ByeongHoon So, WookShin Han, and Hongrae Lee. Natural language to sql: Where are we today? *PVLDB*, 13(10):1737–1750, 2020.
- Paul Kingsbury and Martha Palmer. Propbank: The next level of treebank. In *Proceedings of Treebanks and Lexical Theories*, Växjö, Sweden, 2003.
- Ioannis Konstas and Mirella Lapata. Concept-to-text generation via discriminative reranking. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 369–378, Jeju Island, Korea, July 2012. Association for Computational Linguistics. URL <https://aclanthology.org/P12-1039>.
- Ioannis Konstas and Mirella Lapata. Inducing document plans for concept-to-text generation. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1503–1514, Seattle, Washington, USA, October 2013. Association for Computational Linguistics. URL <https://aclanthology.org/D13-1157>.
- Unni Krishnan, Alistair Moffat, and Justin Zobel. A taxonomy of query auto completion modes. In Bevan Koopman, Guido Zuccon, and Mark James Carman, editors, *Proceedings of the 22nd Australasian Document Computing Symposium, ADCS 2017, Brisbane, QLD, Australia, December 7-8, 2017*, pages 6:1–6:8. ACM, 2017. doi: 10.1145/3166072.3166081. URL <https://doi.org/10.1145/3166072.3166081>.
- Karen Kukich. Techniques for automatically correcting words in text. *ACM Comput. Surv.*, 24(4):377–439, 1992.
- Hideko S. Kunii. *Graph Data Language*. PhD thesis, 1983.
- Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, Kristina Toutanova, Llion Jones, Matthew Kelcey, Ming-Wei Chang, Andrew M. Dai, Jakob Uszkoreit, Quoc Le, and Slav Petrov. Natural questions: A benchmark for question answering research. *Transactions of the Association for Computational Linguistics*, 7:452–466, 2019. doi: 10.1162/tacl_a_00276. URL <https://aclanthology.org/Q19-1026>.
- Faisal Ladhak, Esin Durmus, Claire Cardie, and Kathleen McKeown. WikiLingua: A new benchmark dataset for cross-lingual abstractive summarization. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4034–4048, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.360. URL <https://aclanthology.org/2020.findings-emnlp.360>.
- Po-Ming Law, Subhajit Das, and Rahul C. Basole. Comparing apples and oranges: Taxonomy and design of pairwise comparisons within tabular data. In Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos, editors, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*,

- CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, page 179. ACM, 2019. doi: 10.1145/3290605.3300409. URL <https://doi.org/10.1145/3290605.3300409>.
- Po-Ming Law, Alex Endert, and John T. Stasko. Characterizing automated data insights. In *31st IEEE Visualization Conference, IEEE VIS 2020 - Short Papers, Virtual Event, USA, October 25-30, 2020*, pages 171–175. IEEE, 2020. doi: 10.1109/VIS47514.2020.00041. URL <https://doi.org/10.1109/VIS47514.2020.00041>.
- Rémi Lebret, David Grangier, and Michael Auli. Neural text generation from structured data with application to the biography domain. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1203–1213, Austin, Texas, November 2016. Association for Computational Linguistics. doi: 10.18653/v1/D16-1128. URL <https://aclanthology.org/D16-1128>.
- Doris Jung Lin Lee, Vidya Setlur, Melanie Tory, Karrie Karahalios, and Aditya G. Parameswaran. Deconstructing categorization in visualization recommendation: A taxonomy and comparative study. *CoRR*, abs/2102.07070, 2021. URL <https://arxiv.org/abs/2102.07070>.
- Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. Dbpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, 6:167–195, 2015.
- Chuan Lei, Fatma Özcan, Abdul Quamar, Ashish R. Mittal, Jaydeep Sen, Diptikalyan Saha, and Karthik Sankaranarayanan. Ontology-based natural language query interfaces for data exploration. *IEEE Data Eng. Bull.*, 41(3):52–63, 2018. URL <http://sites.computer.org/debull/A18sept/p52.pdf>.
- Fei Li and H. V. Jagadish. Constructing an interactive natural language interface for relational databases. *Proc. VLDB Endow.*, 8(1):73–84, 2014a. doi: 10.14778/2735461.2735468. URL <http://www.vldb.org/pvldb/vol8/p73-li.pdf>.
- Fei Li and H. V. Jagadish. Constructing an interactive natural language interface for relational databases. *Proceedings of the VLDB Endowment*, 8(1):73–84, September 2014b. URL <http://dx.doi.org/10.14778/2735461.2735468>.
- Fei Li and H. V. Jagadish. Constructing an interactive natural language interface for relational databases. *PVLDB*, 8(1):73–84, 2014c.
- Haoyang Li, Jing Zhang, Cuiping Li, and Hong Chen. Decoupling the skeleton parsing and schema linking for text-to-sql. *arXiv preprint arXiv:2302.05965*, 2023a.
- Jinyang Li, Binyuan Hui, Ge Qu, Binhua Li, Jiayi Yang, Bowen Li, Bailin Wang, Bowen Qin, Rongyu Cao, Ruiying Geng, Nan Huo, Chenhao Ma, Kevin C. C. Chang, Fei Huang, Reynold Cheng, and Yongbin Li. Can llm already serve as a database interface? a big bench for large-scale database grounded text-to-sqls, 2023b.

- Min Li, Marina Danilevsky, Sara Noeman, and Yunyao Li. DIMSIM: An accurate Chinese phonetic similarity algorithm based on learned high dimensional encoding. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, October 2018.
- Yuntao Li, Bei Chen, Qian Liu, Yan Gao, Jian-Guang Lou, Yan Zhang, and Dongmei Zhang. "what do you mean by that?" A parser-independent interactive approach for enhancing text-to-sql. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 6913–6922. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.emnlp-main.561. URL <https://doi.org/10.18653/v1/2020.emnlp-main.561>.
- Yunyao Li and Davood Rafiei. *Natural Language Data Management and Interfaces. Synthesis Lectures on Data Management*. Morgan & Claypool Publishers, 2018.
- Yunyao Li, Ishan Chaudhuri, Huahai Yang, Satinder P. Singh, and H. V. Jagadish. Enabling domain-awareness for a generic natural language interface. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada*, pages 833–838. AAAI Press, 2007a. URL <http://www.aaai.org/Library/AAAI/2007/aaai07-132.php>.
- Yunyao Li, Huahai Yang, and H. V. Jagadish. Nalix: A generic natural language search environment for XML data. *ACM Trans. Database Syst.*, 32(4):30, 2007b. doi: 10.1145/1292609.1292620. URL <https://doi.org/10.1145/1292609.1292620>.
- Yunyao Li, Huahai Yang, and H. V. Jagadish. Nalix: A generic natural language search environment for XML data. *ACM Trans. Database Systems*, 32(4), 2007c.
- Percy Liang, Michael Jordan, and Dan Klein. Learning dependency-based compositional semantics. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 590–599, Portland, Oregon, USA, June 2011. Association for Computational Linguistics. URL <https://aclanthology.org/P11-1060>.
- Dekang Lin and Patrick Pantel. Dirt - discovery of inference rules from text. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 12 2001.
- Xi Victoria Lin, Richard Socher, and Caiming Xiong. Bridging textual and tabular data for cross-domain text-to-SQL semantic parsing. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4870–4888, Online, November 2020a. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.438. URL <https://aclanthology.org/2020.findings-emnlp.438>.

244 BIBLIOGRAPHY

- Xi Victoria Lin, Richard Socher, and Caiming Xiong. Bridging textual and tabular data for cross-domain text-to-sql semantic parsing. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4870–4888, 2020b.
- Aiwei Liu, Xuming Hu, Lijie Wen, and Philip S Yu. A comprehensive evaluation of chatgpt’s zero-shot text-to-sql capability. *arXiv preprint arXiv:2303.13547*, 2023.
- Qian Liu, Bei Chen, Jiaqi Guo, Morteza Ziyadi, Zeqi Lin, Weizhu Chen, and Jian-Guang Lou. Tapex: Table pre-training via learning a neural sql executor, 2021. URL <https://arxiv.org/abs/2107.07653>.
- Tianyu Liu, Kexiang Wang, Lei Sha, Baobao Chang, and Zhifang Sui. Table-to-text generation by structure-aware seq2seq learning. *CoRR*, abs/1711.09724, 2017. URL <http://arxiv.org/abs/1711.09724>.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692, 2019. URL <http://arxiv.org/abs/1907.11692>.
- Qin Lyu, Kaushik Chakrabarti, Shobhit Hathi, Souvik Kundu, Jianwen Zhang, and Zheng Chen. Hybrid ranking network for text-to-sql, 2020. URL <https://arxiv.org/abs/2008.04759>.
- Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to information retrieval*. Cambridge University Press, 2008.
- Gary Marchionini and Ryen White. Find what you need, understand what you find. *Int. J. Hum. Comput. Interact.*, 23(3):205–237, 2007. doi: 10.1080/10447310701702352. URL <https://doi.org/10.1080/10447310701702352>.
- Abelardo Carlos Martínez Lorenzo, Marco Maru, and Roberto Navigli. Fully-Semantic Parsing and Generation: the BabelNet Meaning Representation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1727–1741, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.121. URL <https://aclanthology.org/2022.acl-long.121>.
- Christian MIM. Matthiessen and John A. Bateman. *Text generation and systemic-functional linguistics: experiences from English and Japanese*. Burns & Oates, 1991.
- Ryan McDonald, Fernando Pereira, Kiril Ribarov, and Jan Hajič. Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 523–530, Vancouver, British Columbia, Canada, October 2005. Association for Computational Linguistics. URL <https://aclanthology.org/H05-1066>.

- Alexandra Meliou, Wolfgang Gatterbauer, Katherine F. Moore, and Dan Suciu. The complexity of causality and responsibility for query answers and non-answers. *Proc. VLDB Endow.*, 4(1):34–45, 2010. doi: 10.14778/1880172.1880176. URL <http://www.vldb.org/pvldb/vol4/p34-meliou.pdf>.
- Microsoft. bing.com, 2021a. [Online; accessed November-19-2021].
- Microsoft. Narratives for power bi: <https://azuremarketplace.microsoft.com/en-us/marketplace/apps/narrativescience.narratives-for-power-bi?tab=Overview>, 2021b. [Online; accessed December-4-2021].
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013a.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119, 2013b.
- Simon Mille, Anja Belz, Bernd Bohnet, Yvette Graham, Emily Pitler, and Leo Wanner. The first multilingual surface realisation shared task (SR’18): Overview and evaluation results. In *Proceedings of the First Workshop on Multilingual Surface Realisation*, pages 1–12, Melbourne, Australia, July 2018a. Association for Computational Linguistics. doi: 10.18653/v1/W18-3601. URL <https://aclanthology.org/W18-3601>.
- Simon Mille, Anja Belz, Bernd Bohnet, and Leo Wanner. Underspecified Universal Dependency structures as inputs for multilingual surface realisation. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 199–209, Tilburg University, The Netherlands, November 2018b. Association for Computational Linguistics. doi: 10.18653/v1/W18-6527. URL <https://aclanthology.org/W18-6527>.
- Simon Mille, Anya Belz, Bernd Bohnet, Thiago Castro Ferreira, Yvette Graham, and Leo Wanner. The third multilingual surface realisation shared task (SR’20): Overview and evaluation results. In *Proceedings of the Third Workshop on Multilingual Surface Realisation*, pages 1–20, Barcelona, Spain (Online), December 2020. Association for Computational Linguistics. URL <https://aclanthology.org/2020.msr-1.1>.
- George A. Miller. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41, 1995. doi: 10.1145/219717.219748. URL <http://doi.acm.org/10.1145/219717.219748>.
- Jeff Mitchell and Mirella Lapata. Composition in distributional models of semantics. *Cogn. Sci.*, 34(8):1388–1429, 2010. doi: 10.1111/j.1551-6709.2010.01106.x. URL <https://doi.org/10.1111/j.1551-6709.2010.01106.x>.
- John X. Morris, Eli Lifland, Jack Lanchantin, Yangfeng Ji, and Yanjun Qi. Reevaluating adversarial examples in natural language. In Trevor Cohn, Yulan He, and Yang Liu,

- editors, *Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November 2020*, volume EMNLP 2020 of *Findings of ACL*, pages 3829–3839. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.findings-emnlp.341. URL <https://doi.org/10.18653/v1/2020.findings-emnlp.341>.
- Davide Mottin, Alice Marascu, Senjuti Basu Roy, Gautam Das, Themis Palpanas, and Yannis Velegrakis. IQR: an interactive query relaxation system for the empty-answer problem. In Curtis E. Dyreson, Feifei Li, and M. Tamer Özsu, editors, *International Conference on Management of Data, SIGMOD 2014, Snowbird, UT, USA, June 22-27, 2014*, pages 1095–1098. ACM, 2014. doi: 10.1145/2588555.2594512. URL <https://doi.org/10.1145/2588555.2594512>.
- Linyong Nan, Dragomir Radev, Rui Zhang, Amrit Rau, Abhinand Sivaprasad, Chiachun Hsieh, Xiangru Tang, Aadit Vyas, Neha Verma, Pranav Krishna, Yangxiaokang Liu, Nadia Irwanto, Jessica Pan, Faiaz Rahman, Ahmad Zaidi, Mutethia Mutuma, Yasin Tarabar, Ankit Gupta, Tao Yu, Yi Chern Tan, Xi Victoria Lin, Caiming Xiong, Richard Socher, and Nazneen Fatema Rajani. DART: Open-domain structured data record to text generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 432–447, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.37. URL <https://aclanthology.org/2021.naacl-main.37>.
- Linyong Nan, Lorenzo Jaime Yu Flores, Yilun Zhao, Yixin Liu, Luke Benson, Weijin Zou, and Dragomir Radev. R2d2: Robust data-to-text with replacement detection, 2022a. URL <https://arxiv.org/abs/2205.12467>.
- Linyong Nan, Chiachun Hsieh, Ziming Mao, Xi Victoria Lin, Neha Verma, Rui Zhang, Wojciech Kryściński, Hailey Schoelkopf, Riley Kong, Xiangru Tang, Mutethia Mutuma, Ben Rosand, Isabel Trindade, Renusree Bandaru, Jacob Cunningham, Caiming Xiong, Dragomir Radev, and Dragomir Radev. FeTaQA: Free-form table question answering. *Transactions of the Association for Computational Linguistics*, 10:35–49, 2022b. doi: 10.1162/tacl_a_00446. URL <https://aclanthology.org/2022.tacl-1.3>.
- Shashi Narayan and Claire Gardent. *Deep Learning Approaches to Text Production*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2020. doi: 10.2200/S00979ED1V01Y201912HLT044. URL <https://doi.org/10.2200/S00979ED1V01Y201912HLT044>.
- Shashi Narayan, Shay B. Cohen, and Mirella Lapata. Don’t give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1797–1807, Brussels, Belgium, October-November 2018.

- Association for Computational Linguistics. doi: 10.18653/v1/D18-1206. URL <https://aclanthology.org/D18-1206>.
- Arpit Narechania, Adam Fourney, Bongshin Lee, and Gonzalo A. Ramos. DIY: assessing the correctness of natural language to SQL systems. In *IUI '21: 26th International Conference on Intelligent User Interfaces, College Station, TX, USA, April 13-17, 2021*, pages 597–607. ACM, 2021. doi: 10.1145/3397481.3450667. URL <https://doi.org/10.1145/3397481.3450667>.
- Tahira Naseem, Austin Blodgett, Sadhana Kumaravel, Tim O’Gorman, Young-Suk Lee, Jeffrey Flanigan, Ramón Astudillo, Radu Florian, Salim Roukos, and Nathan Schneider. DocAMR: Multi-sentence AMR representation and evaluation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3496–3505, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.256. URL <https://aclanthology.org/2022.naacl-main.256>.
- Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant. The conll-2014 shared task on grammatical error correction. In *CoNLL*, 2014.
- Joakim Nivre, Johan Hall, and Jens Nilsson. MaltParser: A data-driven parser-generator for dependency parsing. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC’06)*, Genoa, Italy, May 2006. European Language Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2006/pdf/162_pdf.pdf.
- Jekaterina Novikova, Ondřej Dušek, and Verena Rieser. The E2E dataset: New challenges for end-to-end generation. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 201–206, Saarbrücken, Germany, August 2017. Association for Computational Linguistics. doi: 10.18653/v1/W17-5525. URL <https://aclanthology.org/W17-5525>.
- Tim O’Gorman, Michael Regan, Kira Griffitt, Ulf Hermjakob, Kevin Knight, and Martha Palmer. AMR beyond the sentence: the multi-sentence AMR corpus. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3693–3702, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics. URL <https://aclanthology.org/C18-1313>.
- Jessie Ooi, Xiuqin Ma, Hongwu Qin, and Siau Chuin Liew. A survey of query expansion, query suggestion and query refinement techniques. In *2015 4th International Conference on Software Engineering and Computer Systems (ICSECS)*, pages 112–117, 2015. doi: 10.1109/ICSECS.2015.7333094.
- Oracle. Oracle Analytics. <https://www.oracle.com/business-analytics/natural-language.html>, 2021. [Online; accessed March-28-2021].

- Fatma Ozcan, Abdul Quamar, Jaydeep Sen, Chuan Lei, and Vasilis Efthymiou. State of the art and open challenges in natural language interfaces to data. In David Maier, Rachel Pottinger, AnHai Doan, Wang-Chiew Tan, Abdussalam Alawini, and Hung Q. Ngo, editors, *Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA], June 14-19, 2020*, pages 2629–2636. ACM, 2020. doi: 10.1145/3318464.3383128. URL <https://doi.org/10.1145/3318464.3383128>.
- Filipp Ozinov. JamsPELL, 2022. URL <https://jamsPELL.com>. [Online; accessed July-2-2022].
- Christos H Papadimitriou and Mihalis Yannakakis. On the complexity of database queries. *Journal of Computer and System Sciences*, 58(3):407–427, 1999.
- Ankur Parikh, Xuezhi Wang, Sebastian Gehrmann, Manaal Faruqui, Bhuwan Dhingra, Diyi Yang, and Dipanjan Das. ToTTo: A controlled table-to-text generation dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1173–1186, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.89. URL <https://aclanthology.org/2020.emnlp-main.89>.
- Dae Hoon Park and Rikio Chiba. A neural language model for query auto-completion. In Noriko Kando, Tetsuya Sakai, Hideo Joho, Hang Li, Arjen P. de Vries, and Ryen W. White, editors, *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*, pages 1189–1192. ACM, 2017. doi: 10.1145/3077136.3080758. URL <https://doi.org/10.1145/3077136.3080758>.
- Terence Parsons. *Events in the Semantics of English*. MIT Press, 1994. ISBN 9780262660938.
- Panupong Pasupat and Percy Liang. Compositional semantic parsing on semi-structured tables. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1470–1480, Beijing, China, July 2015.
- Baolin Peng, Chunyuan Li, Jinchao Li, Shahin Shayandeh, Lars Lidén, and Jianfeng Gao. Soloist: Few-shot task-oriented dialog with a single pre-trained auto-regressive model. *ArXiv*, abs/2005.05298, 2020.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1162. URL <https://aclanthology.org/D14-1162>.

- Laura Perez-Beltrachini and Mirella Lapata. Bootstrapping generators from noisy data. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1516–1527, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-1137. URL <https://aclanthology.org/N18-1137>.
- Xinyu Pi, Bing Wang, Yan Gao, Jiaqi Guo, Zhoujun Li, and Jian-Guang Lou. Towards robustness of text-to-sql models against natural and realistic adversarial table perturbation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2007–2022, 2022.
- Ana-Maria Popescu, Oren Etzioni, and Henry Kautz. Towards a theory of natural language interfaces to databases. In *Proceedings of the 8th International Conference on Intelligent User Interfaces*, pages 149–157, 2003. URL <http://doi.acm.org/10.1145/604045.604070>.
- Ana-Maria Popescu, Alex Armanasu, Oren Etzioni, David Ko, and Alexander Yates. Modern natural language interfaces to databases: Composing statistical parsing with semantic tractability. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, pages 141–147, Geneva, Switzerland, aug 23–aug 27 2004. COLING. URL <https://aclanthology.org/C04-1021>.
- Martin F. Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.
- Mohammadreza Pourreza and Davood Rafiei. Din-sql: Decomposed in-context learning of text-to-sql with self-correction. *arXiv preprint arXiv:2304.11015*, 2023.
- Danish Pruthi, Bhuwan Dhingra, and Zachary C. Lipton. Combating adversarial misspellings with robust word recognition. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5582–5591, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1561. URL <https://aclanthology.org/P19-1561>.
- Ratish Puduppully and Mirella Lapata. Data-to-text Generation with Macro Planning. *Transactions of the Association for Computational Linguistics*, 9:510–527, 05 2021. ISSN 2307-387X. doi: 10.1162/tacl_a_00381. URL https://doi.org/10.1162/tacl_a_00381.
- Ratish Puduppully, Li Dong, and Mirella Lapata. Data-to-text generation with content selection and planning. *CoRR*, abs/1809.00582, 2018. URL <http://arxiv.org/abs/1809.00582>.
- Ratish Puduppully, Li Dong, and Mirella Lapata. Data-to-text generation with entity modeling. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2023–2035, Florence, Italy, July 2019. Association for

- Computational Linguistics. doi: 10.18653/v1/P19-1195. URL <https://aclanthology.org/P19-1195>.
- Kun Qian, Nikita Bhutani, Yunyao Li, H. V. Jagadish, and Mauricio A. Hernández. LUSTRE: an interactive system for entity structured representation and variant generation. In *34th IEEE International Conference on Data Engineering, ICDE 2018, Paris, France, April 16-19, 2018*, pages 1613–1616. IEEE Computer Society, 2018. doi: 10.1109/ICDE.2018.00189. URL <https://doi.org/10.1109/ICDE.2018.00189>.
- Kun Qian, Poornima Chozhiyath Raman, Yunyao Li, and Lucian Popa. Learning structured representations of entity names using activelearning and weak supervision. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 6376–6383. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.emnlp-main.517. URL <https://doi.org/10.18653/v1/2020.emnlp-main.517>.
- Bowen Qin, Binyuan Hui, Lihan Wang, Min Yang, Jinyang Li, Binhua Li, Ruiying Geng, Rongyu Cao, Jian Sun, Luo Si, Fei Huang, and Yongbin Li. A survey on text-to-sql parsing: Concepts, methods, and future directions, 2022. URL <https://arxiv.org/abs/2208.13629>.
- Abdul Quamar, Vasilis Efthymiou, Chuan Lei, and Fatma Özcan. Natural language interfaces to data. *Foundations and Trends® in Databases*, 11(4):319–414, 2022. ISSN 1931-7883. doi: 10.1561/19000000078. URL <http://dx.doi.org/10.1561/19000000078>.
- Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. 2018.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67, 2020. URL <http://jmlr.org/papers/v21/20-074.html>.
- Nitarshan Rajkumar, Raymond Li, and Dzmitry Bahdanau. Evaluating the text-to-sql capabilities of large language models, 2022. URL <https://arxiv.org/abs/2204.00498>.
- Siva Reddy, Oscar Täckström, Michael Collins, Tom Kwiatkowski, Dipanjan Das, Mark Steedman, and Mirella Lapata. Transforming dependency structures to logical forms for semantic parsing. *Transactions of the Association for Computational Linguistics*, 4:127–140, 2016. doi: 10.1162/tacl_a_00088. URL <https://aclanthology.org/Q16-1010>.

- Sudeepa Roy and Dan Suciu. A formal approach to finding explanations for database queries. In Curtis E. Dyreson, Feifei Li, and M. Tamer Özsu, editors, *International Conference on Management of Data, SIGMOD 2014, Snowbird, UT, USA, June 22–27, 2014*, pages 1579–1590. ACM, 2014. doi: 10.1145/2588555.2588578. URL <https://doi.org/10.1145/2588555.2588578>.
- Sudeepa Roy, Laurel J. Orr, and Dan Suciu. Explaining query answers with explanation-ready databases. *Proc. VLDB Endow.*, 9(4):348–359, 2015. doi: 10.14778/2856318.2856329. URL <http://www.vldb.org/pvldb/vol9/p348-roy.pdf>.
- Ohad Rubin and Jonathan Berant. Smbop: Semi-autoregressive bottom-up semantic parsing. In *Proceedings of the 5th Workshop on Structured Prediction for NLP (SPNLP 2021)*, pages 12–21, 2021.
- James Rumbaugh, Ivar Jacobson, and Grady Booch. *The unified modeling language. Reference manual*, 1999.
- Diptikalyan Saha, Avrilia Floratou, Karthik Sankaranarayanan, Umar Farooq Minhas, Ashish R. Mittal, and Fatma Özcan. ATHENA: an ontology-driven system for natural language querying over relational data stores. *Proc. VLDB Endow.*, 9(12):1209–1220, 2016. doi: 10.14778/2994509.2994536. URL <http://www.vldb.org/pvldb/vol9/p1209-saha.pdf>.
- Rodrygo L. T. Santos, Craig Macdonald, and Iadh Ounis. Learning to rank query suggestions for adhoc and diversity search. *Inf. Retr.*, 16(4):429–451, 2013. doi: 10.1007/s10791-012-9211-2. URL <https://doi.org/10.1007/s10791-012-9211-2>.
- Torsten Scholak, Nathan Schucher, and Dzmitry Bahdanau. PICARD: Parsing incrementally for constrained auto-regressive decoding from language models. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9895–9901, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.779. URL <https://aclanthology.org/2021.emnlp-main.779>.
- Rolf Schwitter. Controlled natural languages for knowledge representation. In *COLING*, pages 1113–1121, 2010.
- Thomas Scialom, Paul-Alexis Dray, Sylvain Lamprier, Benjamin Piwowarski, and Jacopo Staiano. MLSUM: The multilingual summarization corpus. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8051–8067, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.647. URL <https://aclanthology.org/2020.emnlp-main.647>.
- P Griffiths Selinger, Morton M Astrahan, Donald D Chamberlin, Raymond A Lorie, and Thomas G Price. Access path selection in a relational database management system.

- In *Proceedings of the 1979 ACM SIGMOD international conference on Management of data*, pages 23–34, 1979.
- Jaydeep Sen, Chuan Lei, Abdul Quamar, Fatma Ozcan, Vasilis Efthymiou, Ayushi Dalmia, Greg Stager, Ashish Mittal, Diptikalyan Saha, and Karthik Sankaranarayanan. ATHENA++: Natural language querying for complex nested sql queries. *Proc. VLDB Endow.*, 13(11):2747–2759, 2020.
- Vidya Setlur, Sarah E. Battersby, Melanie Tory, Rich Gossweiler, and Angel X. Chang. Eviza: A natural language interface for visual analysis. In Jun Rekimoto, Takeo Igarashi, Jacob O. Wobbrock, and Daniel Avrahami, editors, *Proceedings of the 29th Annual Symposium on User Interface Software and Technology, UIST 2016, Tokyo, Japan, October 16-19, 2016*, pages 365–377. ACM, 2016.
- Vidya Setlur, Melanie Tory, and Alex Djalali. Inferencing underspecified natural language utterances in visual analysis. In Wai-Tat Fu, Shimei Pan, Oliver Brdiczka, Polo Chau, and Gaelle Calvary, editors, *Proceedings of the 24th International Conference on Intelligent User Interfaces, IUI 2019, Marina del Ray, CA, USA, March 17-20, 2019*, pages 40–51. ACM, 2019. doi: 10.1145/3301275.3302270. URL <https://doi.org/10.1145/3301275.3302270>.
- Vidya Setlur, Enamul Hoque, Dae Hyun Kim, and Angel X. Chang. Sneak pique: Exploring autocompletion as a data discovery scaffold for supporting visual analysis. In Shamsi T. Iqbal, Karon E. MacLean, Fanny Chevalier, and Stefanie Mueller, editors, *UIST '20: The 33rd Annual ACM Symposium on User Interface Software and Technology, Virtual Event, USA, October 20-23, 2020*, pages 966–978. ACM, 2020. doi: 10.1145/3379337.3415813. URL <https://doi.org/10.1145/3379337.3415813>.
- Lei Sha, Lili Mou, Tianyu Liu, Pascal Poupart, Sujian Li, Baobao Chang, and Zhifang Sui. Order-planning neural text generation from structured data. In Sheila A. McIlraith and Kilian Q. Weinberger, editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 5414–5421. AAAI Press, 2018. URL <https://ojs.aaai.org/index.php/AAAI/article/view/11947/11806>.
- Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(4):623–656, 1948.
- Peter Shaw, Ming-Wei Chang, Panupong Pasupat, and Kristina Toutanova. Compositional generalization and natural language variation: Can a semantic parsing approach handle both? In *Proceedings of the ACL Conference*, pages 922–938, 2021.

- Leixian Shen, Enya Shen, Yuyu Luo, Xiaocong Yang, Xuming Hu, Xiongshuai Zhang, Zhiwei Tai, and Jianmin Wang. Towards natural language interfaces for data visualization: A survey. *CoRR*, abs/2109.03506, 2021. URL <https://arxiv.org/abs/2109.03506>.
- Tianze Shi, Kedar Tatwawadi, Kaushik Chakrabarti, Yi Mao, Oleksandr Polozov, and Weizhu Chen. Incsql: Training incremental text-to-sql parsers with non-deterministic oracles. *CoRR*, abs/1809.05054, 2018. URL <http://arxiv.org/abs/1809.05054>.
- Tianze Shi, Chen Zhao, Jordan Boyd-Graber, Hal Daumé III, and Lillian Lee. On the potential of lexico-logical alignments for semantic parsing to SQL queries. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1849–1864, Online, November 2020a. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.167. URL <https://aclanthology.org/2020.findings-emnlp.167>.
- Tianze Shi, Chen Zhao, Jordan L. Boyd-Graber, Hal Daumé III, and Lillian Lee. On the potential of lexico-logical alignments for semantic parsing to SQL queries. *CoRR*, abs/2010.11246, 2020b. URL <https://arxiv.org/abs/2010.11246>.
- SM Shieber. An introduction to uni—fication—based approaches to grammar (vol. 4), 1986.
- Ben Shneiderman, Donald Byrd, and W. Bruce Croft. Clarifying search: A user-interface framework for text searches. *D Lib Mag.*, 3(1), 1997. URL <http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html>.
- Abraham Silberschatz, Henry F Korth, and Shashank Sudarshan. *Database system concepts*. McGraw-Hill, 7th edition, 2019.
- Anders Søgaard. *Explainable Natural Language Processing*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2021. doi: 10.2200/S01118ED1V01Y202107HLT051. URL <https://doi.org/10.2200/S01118ED1V01Y202107HLT051>.
- Robyn Speer, Joshua Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. In *AAAI*, 2017.
- Arjun Srinivasan and Vidya Setlur. Snowy: Recommending utterances for conversational visual analysis. In Jeffrey Nichols, Ranjitha Kumar, and Michael Nebeling, editors, *UIST '21: The 34th Annual ACM Symposium on User Interface Software and Technology, Virtual Event, USA, October 10-14, 2021*, pages 864–880. ACM, 2021. doi: 10.1145/3472749.3474792. URL <https://doi.org/10.1145/3472749.3474792>.

- Arjun Srinivasan and John T. Stasko. Orko: Facilitating multimodal interaction for visual exploration and analysis of networks. *IEEE Trans. Vis. Comput. Graph.*, 24(1):511–521, 2018. doi: 10.1109/TVCG.2017.2745219. URL <https://doi.org/10.1109/TVCG.2017.2745219>.
- Arjun Srinivasan, Bongshin Lee, and John T. Stasko. Interweaving multimodal interaction with flexible unit visualizations for data exploration. *IEEE Trans. Vis. Comput. Graph.*, 27(8):3519–3533, 2021. doi: 10.1109/TVCG.2020.2978050. URL <https://doi.org/10.1109/TVCG.2020.2978050>.
- Mark Steedman. *The Syntactic Process*.
- Vijayan Sugumaran and Veda C Storey. The role of domain ontologies in database design: An ontology management and conceptual modeling environment. *ACM Transactions on Database Systems (TODS)*, 31(3):1064–1094, 2006.
- Alane Suhr, Srinivasan Iyer, and Yoav Artzi. Learning to map context-dependent sentences to executable formal queries. In Marilyn A. Walker, Heng Ji, and Amanda Stent, editors, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pages 2238–2249. Association for Computational Linguistics, 2018a. doi: 10.18653/v1/n18-1203. URL <https://doi.org/10.18653/v1/n18-1203>.
- Alane Suhr, Srinivasan Iyer, and Yoav Artzi. Learning to map context-dependent sentences to executable formal queries. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2238–2249, New Orleans, Louisiana, June 2018b. Association for Computational Linguistics. doi: 10.18653/v1/N18-1203. URL <https://aclanthology.org/N18-1203>.
- Alistair G. Sutcliffe and Mark Ennis. Towards a cognitive theory of information retrieval. *Interact. Comput.*, 10(3):321–351, 1998. doi: 10.1016/S0953-5438(98)00013-7. URL [https://doi.org/10.1016/S0953-5438\(98\)00013-7](https://doi.org/10.1016/S0953-5438(98)00013-7).
- Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL <https://proceedings.neurips.cc/paper/2014/file/a14ac55a4f27472c5d894ec1c3c743d2-Paper.pdf>.
- Idan Szpektor, Aristides Gionis, and Yoelle Maarek. Improving recommendation for long-tail queries via templates. In Sadagopan Srinivasan, Krithi Ramamritham, Arun Kumar, M. P. Ravindra, Elisa Bertino, and Ravi Kumar, editors, *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India,*

- March 28 - April 1, 2011*, pages 47–56. ACM, 2011. doi: 10.1145/1963405.1963416. URL <https://doi.org/10.1145/1963405.1963416>.
- Lappoon R. Tang and Raymond J. Mooney. Automated construction of database interfaces: Integrating statistical and relational learning for semantic parsing. In *2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 133–141, 2000. URL <http://www.aclweb.org/anthology/W00-1317>.
- Nan Tang, Ju Fan, Fangyi Li, Jianhong Tu, Xiaoyong Du, Guoliang Li, Sam Madden, and Mourad Ouzzani. Rpt: relational pre-trained transformer is almost all you need towards democratizing data preparation. *Proceedings of the VLDB Endowment*, 14 (8):1254–1261, 2021.
- Toby J Teorey, Sam S Lightstone, Tom Nadeau, and HV Jagadish. *Database modeling and design: logical design*. Elsevier, 2011.
- Bernhard Thalheim. *Entity-relationship modeling: foundations of database technology*. Springer Science & Business Media, 2013.
- The New York Times. Microsoft bets big on the creator of ChatGPT in cace to dominate A.I., 2023. URL <https://www.nytimes.com/2023/01/12/technology/microsoft-open-ai-chatgpt.html>. [Online; accessed December-1-2023].
- Reuben Thomas. Enchant, 2022. URL <https://abiword.github.io/enchant/>. [Online; accessed July-2-2022].
- Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam M. Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, Yaguang Li, Hongrae Lee, Huaixiu Zheng, Amin Ghafouri, Marcelo Menegali, Yanping Huang, Maxim Krikun, Dmitry Lepikhin, James Qin, Dehao Chen, Yuanzhong Xu, Zhifeng Chen, Adam Roberts, Maarten Bosma, Yanqi Zhou, Chung-Ching Chang, I. A. Kriukon, Willard James Rusch, Marc Pickett, Kathleen S. Meier-Hellstern, Meredith Ringel Morris, Tulsee Doshi, Renelito Delos Santos, Toju Duke, Johnny Hartz Søraker, Ben Zevenbergen, Vinodkumar Prabhakaran, Mark Díaz, Ben Hutchinson, Kristen Olson, Alejandra Molina, Erin Hoffman-John, Josh Lee, Lora Aroyo, Ravindran Rajakumar, Alena Butryna, Matthew Lamm, V. O. Kuzmina, Joseph Fenton, Aaron Cohen, Rachel Bernstein, Ray Kurzweil, Blaise Aguera-Arcas, Claire Cui, Marian Croak, Ed Chi, and Quoc Le. Lamda: Language models for dialog applications. *ArXiv*, abs/2201.08239, 2022.
- James Thorne, Majid Yazdani, Marzieh Saeidi, Fabrizio Silvestri, Sebastian Riedel, and Alon Halevy. From natural language processing to neural databases. *Proc. VLDB Endow.*, 14(6):1033–1039, feb 2021a. ISSN 2150-8097. doi: 10.14778/3447689.3447706. URL <https://doi.org/10.14778/3447689.3447706>.

256 BIBLIOGRAPHY

- James Thorne, Majid Yazdani, Marzieh Saeidi, Fabrizio Silvestri, Sebastian Riedel, and Alon Halevy. Database reasoning over text. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3091–3104, Online, August 2021b. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.241. URL <https://aclanthology.org/2021.acl-long.241>.
- Melanie Tory and Vidya Setlur. Do what I mean, not what I say! design considerations for supporting intent and context in analytical conversation. In Remco Chang, Daniel A. Keim, and Ross Maciejewski, editors, *14th IEEE Conference on Visual Analytics Science and Technology, IEEE VAST 2019, Vancouver, BC, Canada, October 20-25, 2019*, pages 93–103. IEEE, 2019. doi: 10.1109/VAST47406.2019.8986918. URL <https://doi.org/10.1109/VAST47406.2019.8986918>.
- John W. Tukey. Exploratory data analysis. In *Addison-Wesley series in behavioral science : quantitative methods*, 1977.
- P. D. Turney and P. Pantel. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37:141–188, feb 2010. doi: 10.1613/jair.2934. URL <https://doi.org/10.1613%2Fjair.2934>.
- Arif Usta, Akifhan Karakayali, and Özgür Ulusoy. Dbtagger: Multi-task learning for keyword mapping in nlidbs using bi-directional recurrent neural networks. *Proc. VLDB Endow.*, 14(5):813–821, 2021. URL <http://www.vldb.org/pvldb/vol14/p813-usta.pdf>.
- Hossein Vahabi, Margareta Ackerman, David Loker, Ricardo Baeza-Yates, and Alejandro López-Ortiz. Orthogonal query recommendation. In Qiang Yang, Irwin King, Qing Li, Pearl Pu, and George Karypis, editors, *Seventh ACM Conference on Recommender Systems, RecSys '13, Hong Kong, China, October 12-16, 2013*, pages 33–40. ACM, 2013. doi: 10.1145/2507157.2507159. URL <https://doi.org/10.1145/2507157.2507159>.
- Jens E. L. Van Gysel, Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Tim O’Gorman, Andrew Cowell, William Croft, Chu-Ren Huang, Jan Hajič, James H. Martin, Stephan Oepen, Martha Palmer, James Pustejovsky, Rosa Vallejos, and Nianwen Xue. Designing a uniform meaning representation for natural language processing. 2021.
- Pavlos Vougiouklis, Hady Elsahar, Lucie-Aimée Kaffee, Christoph Gravier, Frederique Laforest, Jonathon Hare, and Elena Simperl. Neural wikipedia: Generating textual summaries from knowledge base triples, 2017. URL <https://arxiv.org/abs/1711.00155>.

- Pavlos Vougiouklis, Eddy Maddalena, Jonathon Hare, and Elena Simperl. Point at the triple: Generation of text summaries from knowledge base triples. *JAIR*, 2020. doi: <https://doi.org/10.1613/jair.1.11694>.
- Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. *Commun. ACM*, 57(10):78–85, 2014. doi: 10.1145/2629489. URL <https://doi.org/10.1145/2629489>.
- Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 353–355, Brussels, Belgium, November 2018a. Association for Computational Linguistics. doi: 10.18653/v1/W18-5446. URL <https://aclanthology.org/W18-5446>.
- Bailin Wang, Richard Shin, Xiaodong Liu, Oleksandr Polozov, and Matthew Richardson. Rat-sql: Relation-aware schema encoding and linking for text-to-sql parsers. In *ACL*, 2020a.
- Bailin Wang, Richard Shin, Xiaodong Liu, Oleksandr Polozov, and Matthew Richardson. RAT-SQL: Relation-aware schema encoding and linking for text-to-SQL parsers. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7567–7578, Online, July 2020b. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.677. URL <https://aclanthology.org/2020.acl-main.677>.
- Bailin Wang, Richard Shin, Xiaodong Liu, Oleksandr Polozov, and Matthew Richardson. Rat-sql: Relation-aware schema encoding and linking for text-to-sql parsers. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7567–7578, 2020c.
- Po-Wei Wang, Huan Zhang, Vijai Mohan, Inderjit S. Dhillon, and J. Zico Kolter. Real-time query completion via deep language models. In Jon Degenhardt, Giuseppe Di Fabbrizio, Surya Kallumadi, Mohit Kumar, Andrew Trotman, Yiu-Chang Lin, and Huasha Zhao, editors, *The SIGIR 2018 Workshop On eCommerce co-located with the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018)*, Ann Arbor, Michigan, USA, July 12, 2018, volume 2319 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018b. URL <http://ceur-ws.org/Vol-2319/paper24.pdf>.
- Sida Wang, Weiwei Guo, Huiji Gao, and Bo Long. Efficient neural query auto completion. In Mathieu d’Aquin, Stefan Dietze, Claudia Hauff, Edward Curry, and Philippe Cudré-Mauroux, editors, *CIKM ’20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, pages 2797–2804. ACM, 2020d. doi: 10.1145/3340531.3412701. URL <https://doi.org/10.1145/3340531.3412701>.

- Xingbo Wang, Furui Cheng, Yong Wang, Ke Xu, Jiang Long, Hong Lu, and Huamin Qu. Interactive data analysis with next-step natural language query recommendation. *CoRR*, abs/2201.04868, 2022a. URL <https://arxiv.org/abs/2201.04868>.
- Yun Wang, Zhida Sun, Haidong Zhang, Weiwei Cui, Ke Xu, Xiaojuan Ma, and Dongmei Zhang. Datashot: Automatic generation of fact sheets from tabular data. *IEEE Transactions on Visualization and Computer Graphics*, 26:895–905, 2020e.
- Yushi Wang, Jonathan Berant, and Percy Liang. Building a semantic parser overnight. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1332–1342, Beijing, China, July 2015. Association for Computational Linguistics. doi: 10.3115/v1/P15-1129. URL <https://aclanthology.org/P15-1129>.
- Zhiruo Wang, Haoyu Dong, Ran Jia, Jia Li, Zhiyi Fu, Shi Han, and Dongmei Zhang. Tuta: Tree-based transformers for generally structured table pre-training. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 1780–1790, 2021.
- Zhiruo Wang, Zhengbao Jiang, Eric Nyberg, and Graham Neubig. Table retrieval may not necessitate table-specific model design. In *Proceedings of the Workshop on Structured and Unstructured Knowledge Integration (SUKI)*, pages 36–46, Seattle, USA, July 2022b. Association for Computational Linguistics. doi: 10.18653/v1/2022.suki-1.5. URL <https://aclanthology.org/2022.suki-1.5>.
- David H.D. Warren and Fernando C.N. Pereira. An efficient easily adaptable system for interpreting natural language queries. *American Journal of Computational Linguistics*, 8(3-4):110–122, 1982. URL <https://aclanthology.org/J82-3002>.
- Ryen W. White and Resa A. Roth. *Exploratory Search: Beyond the Query-Response Paradigm*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, 2009. doi: 10.2200/S00174ED1V01Y200901ICR003. URL <https://doi.org/10.2200/S00174ED1V01Y200901ICR003>.
- World Health Organization (WHO). International Classification of Diseases, Eleventh Revision (ICD-11), 2019.
- Leland Wilkinson, Anushka Anand, and Robert L. Grossman. Graph-theoretic scagnostics. In John T. Stasko and Matthew O. Ward, editors, *IEEE Symposium on Information Visualization (InfoVis 2005), 23-25 October 2005, Minneapolis, MN, USA*, pages 157–164. IEEE Computer Society, 2005. doi: 10.1109/INFVIS.2005.1532142. URL <https://doi.org/10.1109/INFVIS.2005.1532142>.
- Sam Wiseman, Stuart Shieber, and Alexander Rush. Challenges in data-to-document generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural*

- Language Processing*, pages 2253–2263, Copenhagen, Denmark, September 2017. Association for Computational Linguistics. doi: 10.18653/v1/D17-1239. URL <https://aclanthology.org/D17-1239>.
- Sam Wiseman, Stuart Shieber, and Alexander Rush. Learning neural templates for text generation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3174–3187, Brussels, Belgium, October–November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1356. URL <https://aclanthology.org/D18-1356>.
- Yuk Wah Wong and Raymond Mooney. Learning for semantic parsing with statistical machine translation. In *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, pages 439–446, New York City, USA, June 2006. Association for Computational Linguistics. URL <https://aclanthology.org/N06-1056>.
- William A. Woods. Progress in natural language understanding: an application to lunar geology. In *AFIPS National Computer Conference*, 1973.
- Bin Wu, Chenyan Xiong, Maosong Sun, and Zhiyuan Liu. Query suggestion with feedback memory network. In Pierre-Antoine Champin, Fabien Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis, editors, *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018*, pages 1563–1571. ACM, 2018. doi: 10.1145/3178876.3186068. URL <https://doi.org/10.1145/3178876.3186068>.
- Chien-Sheng Wu, Andrea Madotto, Ehsan Hosseini-Asl, Caiming Xiong, Richard Socher, and Pascale Fung. Transferable multi-domain state generator for task-oriented dialogue systems. In *ACL*, 2019.
- Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. Google’s neural machine translation system: Bridging the gap between human and machine translation, 2016. URL <https://arxiv.org/abs/1609.08144>.
- Tianbao Xie, Chen Henry Wu, Peng Shi, Ruiqi Zhong, Torsten Scholak, Michihiro Yasunaga, Chien-Sheng Wu, Ming Zhong, Pengcheng Yin, Sida I. Wang, Victor Zhong, Bailin Wang, Chengzu Li, Connor Boyle, Ansong Ni, Ziyu Yao, Dragomir R. Radev, Caiming Xiong, Lingpeng Kong, Rui Zhang, Noah A. Smith, Luke Zettlemoyer, and Tao Yu. Unifiedskg: Unifying and multi-tasking structured knowledge grounding with text-to-text language models. *CoRR*, abs/2201.05966, 2022. URL <https://arxiv.org/abs/2201.05966>.

- Xinyu Xing and Xiaojun Wan. Structure-aware pre-training for table-to-text generation. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2273–2278, 2021.
- Peng Xu, Wenjie Zi, Hamidreza Shahidi, Ákos Kádár, Keyi Tang, Wei Yang, Jawad Ateeq, Harsh Barot, Meidan Alon, and Yanshuai Cao. Turing: an accurate and interpretable multi-hypothesis cross-domain natural language database interface. In *ACL*, 2021.
- Xiaojun Xu, Chang Liu, and Dawn Song. Sqlnet: Generating structured queries from natural language without reinforcement learning. *arXiv preprint arXiv:1711.04436*, 2017.
- Navid Yaghmazadeh, Yuepeng Wang, Isil Dillig, and Thomas Dillig. Sqlizer: Query synthesis from natural language. In *International Conference on Object-Oriented Programming, Systems, Languages, and Applications, ACM*, pages 63:1–63:26, October 2017a. URL <http://doi.org/10.1145/3133887>.
- Navid Yaghmazadeh, Yuepeng Wang, Isil Dillig, and Thomas Dillig. Sqlizer: query synthesis from natural language. *Proceedings of the ACM on Programming Languages*, 1 (OOPSLA):1–26, 2017b.
- Navid Yaghmazadeh, Yuepeng Wang, Isil Dillig, and Thomas Dillig. Type-and content-driven synthesis of sql queries from natural language. *arXiv preprint arXiv:1702.01168*, 2017c.
- Yahoo! yahoo.com, 2021. [Online; accessed November-19-2021].
- Masaya Yamaguchi, Takeyuki Kojima, Nobuo Inui, Yoshiyuki Kotani, and Hirohiko Nisimura. Combination of an automatic and an interactive disambiguation method. In *ACL*, pages 1423–1427, 1998.
- Yandex. yandex.com, 2021. [Online; accessed November-19-2021].
- Huahai Yang, Yunyao Li, and Michelle X. Zhou. Understand users’ comprehension and preferences for composing information visualizations. *ACM Trans. Comput. Hum. Interact.*, 21(1):6:1–6:30, 2014. doi: 10.1145/2541288. URL <https://doi.org/10.1145/2541288>.
- Jingfeng Yang, Aditya Gupta, Shyam Upadhyay, Luheng He, Rahul Goel, and Shachi Paul. TableFormer: Robust transformer modeling for table-text encoding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 528–537, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.40. URL <https://aclanthology.org/2022.acl-long.40>.

- Pengcheng Yin and Graham Neubig. A syntactic neural model for general-purpose code generation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 440–450, Vancouver, Canada, July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1041. URL <https://aclanthology.org/P17-1041>.
- Pengcheng Yin, Zhengdong Lu, Hang Li, and Ben Kao. Neural enquirer: Learning to query tables. *CoRR*, abs/1512.00965, 2015. URL <http://arxiv.org/abs/1512.00965>.
- Pengcheng Yin, Graham Neubig, Wen-tau Yih, and Sebastian Riedel. TaBERT: Pre-training for joint understanding of textual and tabular data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8413–8426, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.745. URL <https://aclanthology.org/2020.acl-main.745>.
- Ori Yoran, Alon Talmor, and Jonathan Berant. Turning tables: Generating examples from semi-structured tables for endowing language models with reasoning skills. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6016–6031, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.416. URL <https://aclanthology.org/2022.acl-long.416>.
- Tao Yu, Zifan Li, Zilin Zhang, Rui Zhang, and Dragomir Radev. Typesql: Knowledge-based type-aware neural text-to-sql generation. In *Proceedings of NAACL*. Association for Computational Linguistics, 2018a.
- Tao Yu, Michihiro Yasunaga, Kai Yang, Rui Zhang, Dongxu Wang, Zifan Li, and Dragomir Radev. Syntaxsqlnet: Syntax tree networks for complex and cross-domain text-to-sql task. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018b.
- Tao Yu, Rui Zhang, Kai Yang, Michihiro Yasunaga, Dongxu Wang, Zifan Li, James Ma, Irene Li, Qingning Yao, Shanelle Roman, Zilin Zhang, and Dragomir Radev. Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-sql task. In *Proceedings of EMNLP 2018*, Brussels, Belgium, 2018c.
- Tao Yu, Rui Zhang, Kai Yang, Michihiro Yasunaga, Dongxu Wang, Zifan Li, James Ma, Irene Li, Qingning Yao, Shanelle Roman, Zilin Zhang, and Dragomir Radev. Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-sql task. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3911–3921, 2018d. URL <http://aclweb.org/anthology/D18-1425>.
- Tao Yu, Rui Zhang, Kai Yang, Michihiro Yasunaga, Dongxu Wang, Zifan Li, James Ma, Irene Li, Qingning Yao, Shanelle Roman, Zilin Zhang, and Dragomir R. Radev. Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing

- and text-to-sql task. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 3911–3921. Association for Computational Linguistics, 2018e.
- Tao Yu, Rui Zhang, Heyang Er, Suyi Li, Eric Xue, Bo Pang, Xi Victoria Lin, Yi Chern Tan, Tianze Shi, Zihan Li, Youxuan Jiang, Michihiro Yasunaga, Sungrok Shim, Tao Chen, Alexander Fabbri, Zifan Li, Luyao Chen, Yuwen Zhang, Shreya Dixit, Vincent Zhang, Caiming Xiong, Richard Socher, Walter Lasecki, and Dragomir Radev. CoSQL: A conversational text-to-SQL challenge towards cross-domain natural language interfaces to databases. In *Proceedings of EMNLP 2019*, pages 1962–1979, Hong Kong, China, November 2019a.
- Tao Yu, Rui Zhang, Heyang Er, Suyi Li, Eric Xue, Bo Pang, Xi Victoria Lin, Yi Chern Tan, Tianze Shi, Zihan Li, Youxuan Jiang, Michihiro Yasunaga, Sungrok Shim, Tao Chen, Alexander R. Fabbri, Zifan Li, Luyao Chen, Yuwen Zhang, Shreya Dixit, Vincent Zhang, Caiming Xiong, Richard Socher, Walter S. Lasecki, and Dragomir R. Radev. Cosql: A conversational text-to-sql challenge towards cross-domain natural language interfaces to databases. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 1962–1979. Association for Computational Linguistics, 2019b. doi: 10.18653/v1/D19-1204. URL <https://doi.org/10.18653/v1/D19-1204>.
- Tao Yu, Rui Zhang, Michihiro Yasunaga, Yi Chern Tan, Xi Victoria Lin, Suyi Li, Heyang Er, Irene Li, Bo Pang, Tao Chen, Emily Ji, Shreya Dixit, David Proctor, Sungrok Shim, Jonathan Kraft, Vincent Zhang, Caiming Xiong, Richard Socher, and Dragomir R. Radev. Sparc: Cross-domain semantic parsing in context. In Anna Korhonen, David R. Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28-August 2, 2019, Volume 1: Long Papers*, pages 4511–4523. Association for Computational Linguistics, 2019c. doi: 10.18653/v1/p19-1443. URL <https://doi.org/10.18653/v1/p19-1443>.
- Tao Yu, Rui Zhang, Michihiro Yasunaga, Yi Chern Tan, Xi Victoria Lin, Suyi Li, Irene Li, Heyang Er, Bo Pang, Tao Chen, Emily Ji, Shreya Dixit, David Proctor, Sungrok Shim, Vincent Zhang, Jonathan Kraft, Caiming Xiong, Richard Socher, and Dragomir Radev. Sparc: Cross-domain semantic parsing in context. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 2019d.
- Tao Yu, Chien-Sheng Wu, Xi Victoria Lin, Bailin Wang, Yi Chern Tan, Xinyi Yang, Dragomir Radev, Richard Socher, and Caiming Xiong. Grappa: Grammar-augmented pre-training for table semantic parsing, 2020. URL <https://arxiv.org/abs/2009.13845>.

- Tao Yu, Rui Zhang, Oleksandr Polozov, Christopher Meek, and Ahmed Hassan Awadallah. SCoRE: Pre-training for context representation in conversational semantic parsing. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=oyZxhRI2RiE>.
- Wenhao Yu, Chenguang Zhu, Zaitang Li, Zhiting Hu, Qingyun Wang, Heng Ji, and Meng Jiang. A survey of knowledge-enhanced text generation. *ACM Computing Surveys*, mar 2022. doi: 10.1145/3512467. URL <https://doi.org/10.1145/3512467>.
- Sina Zarrieß, Henrik Voigt, and Simeon Schüz. Decoding methods in neural language generation: A survey. *Inf.*, 12(9):355, 2021. doi: 10.3390/info12090355. URL <https://doi.org/10.3390/info12090355>.
- John M. Zelle and Raymond J. Mooney. Learning semantic grammars with constructive inductive logic programming. In Richard Fikes and Wendy G. Lehnert, editors, *Proceedings of the 11th National Conference on Artificial Intelligence. Washington, DC, USA, July 11-15, 1993*, pages 817–822. AAAI Press / The MIT Press, 1993. URL <http://www.aaai.org/Library/AAAI/1993/aaai93-122.php>.
- John M Zelle and Raymond J Mooney. Learning to parse database queries using inductive logic programming. In *Proceedings of the national conference on artificial intelligence*, pages 1050–1055, 1996.
- Luke Zettlemoyer and Michael Collins. Learning to map sentences to logical form: Structured classification with probabilistic categorical grammars. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, 2005.
- Luke Zettlemoyer and Michael Collins. Learning context-dependent mappings from sentences to logical form. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 976–984, Suntec, Singapore, August 2009. Association for Computational Linguistics. URL <https://aclanthology.org/P09-1110>.
- Rui Zhang, Tao Yu, He Yang Er, Sungrok Shim, Eric Xue, Xi Victoria Lin, Tianze Shi, Caiming Xiong, Richard Socher, and Dragomir Radev. Editing-based sql query generation for cross-domain context-dependent questions. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, Hong Kong, China, 2019a.
- Rui Zhang, Tao Yu, Heyang Er, Sungrok Shim, Eric Xue, Xi Victoria Lin, Tianze Shi, Caiming Xiong, Richard Socher, and Dragomir Radev. Editing-based SQL query generation for cross-domain context-dependent questions. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5338–5349, Hong Kong, China, November 2019b. Association for Computational Linguistics. doi: 10.18653/v1/D19-1537. URL <https://aclanthology.org/D19-1537>.

- Yusen Zhang, Xiangyu Dong, Shuaichen Chang, Tao Yu, Peng Shi, and Rui Zhang. Did you ask a good question? a cross-domain question intention classification benchmark for text-to-sql. *ArXiv*, abs/2010.12634, 2020.
- Ruiqi Zhong, Tao Yu, and Dan Klein. Semantic evaluation for text-to-sql with distilled test suite. In *EMNLP 2020*, 2020a.
- Ruiqi Zhong, Tao Yu, and Dan Klein. Semantic evaluation for text-to-sql with distilled test suites. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 396–411, 2020b.
- Victor Zhong, Caiming Xiong, and Richard Socher. Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning. *CoRR*, abs/1709.00103, 2017. URL <http://arxiv.org/abs/1709.00103>.