Natural Language Interfaces to Data

Other titles in Foundations and Trends® in Databases

Algorithmic Aspects of Parallel Data Processing
Paraschos Koutris, Semih Salihoglu and Dan Suciu
ISBN: 978-1-68083-406-2

Data Infrastructure for Medical Research
Thomas Heinis and Anastasia Ailamaki
ISBN: 978-1-68083-348-5

Main Memory Database Systems
Franz Faerber, Alfons Kemper, Per-Ake Larson, Justin Levandoski,
Thomas Neumann and Andrew Pavlo
ISBN: 978-1-68083-324-9

Query Processing on Probabilistic Data: A Survey
Guy Van den Broeck and Dan Suciu
ISBN: 978-1-68083-314-0

Big Graph Analytics Platforms
Da Yan, Yingyi Bu, Yuanyuan Tian and Amol Deshpande
978-1-68083-242-6

Natural Language Interfaces to Data

Abdul Quamar
IBM Research AI
ahquamar@us.ibm.com

Vasilis Efthymiou
FORTH-ICS
vefthym@ics.forth.gr

Chuan Lei
Instacart
chuan.lei@instacart.com

Fatma Özcan
Systems Research@Google
fozcan@google.com

the essence of knowledge
Boston — Delft

Foundations and Trends® in Databases

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

A. Quamar *et al.*. *Natural Language Interfaces to Data*. Foundations and Trends® in Databases, vol. 11, no. 4, pp. 319–414, 2022.

ISBN: 978-1-63828-029-3

Foundations and Trends® in Databases
Volume 11, Issue 4, 2022

Editorial Scope

Topics

Foundations and Trends® in Databases publishes survey and tutorial articles in the following topics:

- Data Models and Query Languages
- Query Processing and Optimization
- Storage, Access Methods, and Indexing
- Transaction Management, Concurrency Control and Recovery
- Deductive Databases
- Parallel and Distributed Database Systems
- Database Design and Tuning
- Metadata Management
- Object Management
- Trigger Processing and Active Databases
- Data Mining and OLAP
- Approximate and Interactive Query Processing
- Data Warehousing
- Adaptive Query Processing
- Data Stream Management
- Search and Query Integration
- XML and Semi-Structured Data
- Web Services and Middleware
- Data Integration and Exchange
- Private and Secure Data Management
- Peer-to-Peer, Sensornet and Mobile Data Management
- Scientific and Spatial Data Management
- Data Brokering and Publish/Subscribe
- Data Cleaning and Information Extraction
- Probabilistic Data Management

Information for Librarians

Foundations and Trends® in Databases, 2022, Volume 11, 4 issues. ISSN paper version 1931-7883. ISSN online version 1931-7891. Also available as a combined paper and online subscription.

Contents

Natural Language Interfaces to Data

Abdul Quamar[1], Vasilis Efthymiou[2], Chuan Lei[3] and Fatma Özcan[4]

[1]*IBM Research AI, USA; ahquamar@us.ibm.com*
[2]*FORTH-ICS, Greece; vefthym@ics.forth.gr*
[3]*Instacart, USA; chuan.lei@instacart.com*
[4]*Systems Research@Google, USA; fozcan@google.com*

ABSTRACT

Recent advances in natural language understanding and processing have resulted in renewed interest in natural language interfaces to data, which provide an easy mechanism for non-technical users to access and query the data. While early systems evolved from keyword search and focused on simple factual queries, the complexity of both the input sentences as well as the generated SQL queries has evolved over time. More recently, there has also been a lot of focus on using conversational interfaces for data analytics, empowering a line of business owners and non-technical users with quick insights into the data. There are three main challenges in natural language querying: (1) identifying the entities involved in the user utterance, (2) connecting the different entities in a meaningful way over the underlying data source to interpret user intents, and finally (3) generating a structured query in the form of SQL or SPARQL.

There are two main approaches in the literature for interpreting a user's natural language query. Rule-based systems make use of semantic indices, ontologies, and knowledge

Abdul Quamar, Vasilis Efthymiou, Chuan Lei and Fatma Özcan (2022), "Natural Language Interfaces to Data", Foundations and Trends® in Databases: Vol. 11, No. 4, pp 319–414. DOI: 10.1561/1900000078.

graphs to identify the entities in the query, understand the intended relationships between those entities, and utilize grammars to generate the target queries. With the advances in deep learning-based language models, there have been many text-to-SQL approaches that try to interpret the query holistically using deep learning models. Hybrid approaches that utilize both rule-based techniques as well as deep learning models are also emerging by combining the strengths of both approaches. Conversational interfaces are the next natural step to one-shot natural language querying by exploiting query context between multiple turns of conversation for disambiguation. In this monograph, we review the background technologies that are used in natural language interfaces, and survey the different approaches to natural language querying. We also describe conversational interfaces for data analytics and discuss several benchmarks used for natural language querying research and evaluation.

1

Introduction

Natural language interfaces provide an easy way to query and interact with data, and enable non-technical users to investigate the data sets without the need for knowing a query language like SQL. As a result, natural language interfaces have been an active area of research for many decades. With the advances in natural language processing (NLP) technologies, and language models like BERT (Devlin *et al.*, 2019), there is renewed research interest. Even limited forms of such interfaces are now becoming available in commercial products (*Ask Data / Tableau Software* 2021; *Power BI Platform* 2021; *Cognos Assistant* 2021).

Many business users and line of business owners rely on technical people to query and gain insights from their data. These technical people are experts on using complex query languages such as SQL or SPARQL. Today, it is vital for non-technical uses to derive insights from their data as quickly as possible to make effective business decisions. Most often business owners do not have direct access to the data, instead relying on application interfaces with pre-defined queries or dashboards to access and examine the data. Usually, technical users close the gap by creating the dashboards and the canned queries needed, but this introduces delays. Today, there is an increasing need for rapid data access and insights as well as quick exploration of data as soon as it lands in the database. Natural language interfaces provide this functionality, giving rise to the augmented consumer (Richardson *et al.*, 2021). Gartner predicts that the future analytics experiences will be consumer-focused, augmented in context as well as conversational.

Natural language interfaces include natural language query (NLQ) systems, as well as dialogue (or conversational) systems. NLQ systems interpret a single user utterance and produce a SQL or SPARQL query. In other words, NLQ systems offer one-shot query answering, without any context between subsequent queries, whereas conversational systems allow multiple turns in question answering, while preserving some context between turns. This additional context information allows further disambiguation in interpretation.

There are several challenges in building natural language interfaces to data (Affolter *et al.*, 2019). Ambiguity in natural language is a big challenge, making it difficult to understand the semantics of the query and hence the user intent. Understanding the complex relationships between the entities in the user statement and generating a complex SQL query are also challenging. General purpose solutions that can be adapted quickly to any domain are difficult to build. Figure 1.1 shows the three important tasks that are involved in natural language querying of data. The first task in NLQ is semantic parsing and entity tagging, which identifies the entities involved in the user query. Identifying the relationships between these entities, associating them with the data elements in the database, and finally interpreting the user intent based on these entities and relationships is the most critical and challenging task in NLQ. There may be many interpretations that are valid and choosing the right one is also non-trivial. Finally, the last task in NLQ is generating the SQL query that corresponds to the chosen interpretation.

There are two main approaches to NLQ: rule-based and ML/DL-based techniques. Some systems (Saha *et al.*, 2016; Lei *et al.*, 2018; Sen *et al.*, 2019; Li and Jagadish, 2014b; Li and Jagadish, 2014a; Li and Jagadish, 2016; Blunschi *et al.*, 2012; Song *et al.*, 2015) use semantic indexes or ontologies to identify the entities in the query, and employ rule-based or grammar-based techniques for query interpretation and SQL generation. Machine learning (ML) and deep learning (DL) based

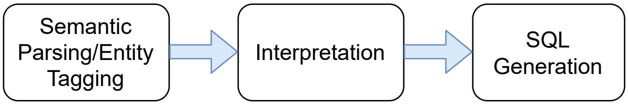

Figure 1.1: Tasks in natural language querying.

text-to-SQL techniques (Basik *et al.*, 2018; Weir and Utama, 2019; Zhong *et al.*, 2017; Xu *et al.*, 2017; Yu *et al.*, 2018a; Gur *et al.*, 2018; Zhang *et al.*, 2019), which encode user inputs into a feature embedding and train deep learning models to generate the SQL query in a holistic way, are widely used, and have become more popular recently. While rule-based approaches provide easier domain adaptation, text-to-SQL systems are more robust to paraphrasing of the input query. There are also some emerging hybrid solutions that mix rule-based and ML/DL-based techniques for different NLQ tasks. For example, Usta *et al.* (2021) provide a DL-based technique for entity tagging that can be plugged in any rule-based solution.

Natural language interfaces have been an area of active research in various communities for many years (Özcan *et al.*, 2020; Li and Rafiei, 2017; Affolter *et al.*, 2019; Katsogiannis-Meimarakis and Koutrika, 2021b; Gkini *et al.*, 2021). Figure 1.2 shows a historical timeline for many NLQ and conversational solutions. In particular, the search and NLP communities have worked on natural language interfaces by extending keyword search into templates and sentences. Many question answering systems are in this group. A question answering system allows the user to ask questions in natural language and to obtain direct answers that correspond to facts stored in the database. It can be considered as an enhancement to search systems. Instead of a simple keyword search over the data, question answering systems can provide more meaningful and insightful information in the form of short answers to the user's natural language questions. Similar to keyword search, the goal in these use cases is to find information about certain entities, such as the CEO of a company or the director of a movie. In these systems, the final structured query that gets generated is a simple lookup query. Examples include early systems (Aditya *et al.*, 2002; Tata and Lohman, 2008) that only allow a set of keywords, with very limited expressive

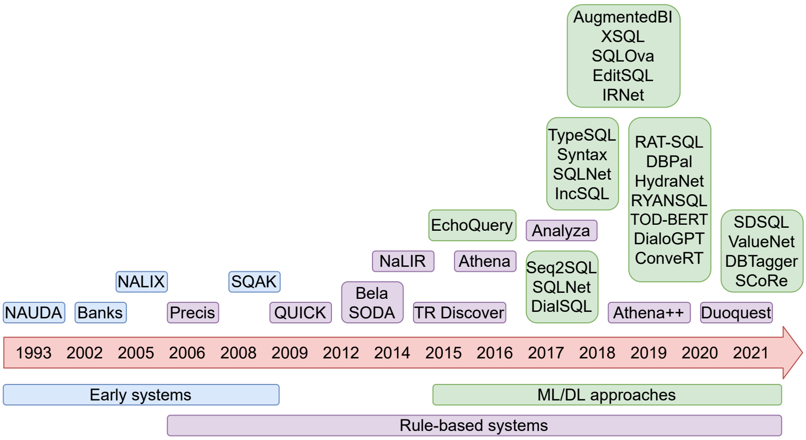

Figure 1.2: Historical perspective.

power, as well as systems (Blunschi *et al.*, 2012; Zenz *et al.*, 2009) that mostly focus on simple queries that access a single table using some selection criteria. Later works allow a full-blown English statement and try to disambiguate among the multiple meanings of the words and their relationships. There has been also work on building conversational systems (Yu *et al.*, 2019a; Quamar *et al.*, 2020a) that allow advanced search on well-curated databases.

The database community has focused on natural language interfaces for analytical queries, as such interfaces enable business users and analytics teams to quickly analyze the data, and understand reasons and key drivers for business behaviors. As predicted by Gartner (Richardson *et al.*, 2021), to become more widely used than pre-defined dashboards, these systems require complex SQL queries that are typical in analytical systems. With the recent advances in NLP (Young *et al.*, 2018), both the complexity of input natural language statements, as well as the generated SQL and SPARQL queries have increased over time. A lot of these systems (Li *et al.*, 2005; Saha *et al.*, 2016; Sen *et al.*, 2020; Basik *et al.*, 2018) have originated in the database research community and can generate complex SQL queries with many joins and aggregations, as well as nesting.

In this monograph, we first review the background technologies empowering the existing natural language interfaces to data in Section 2. Then, in Section 3, we discuss many rule-based and text-to-SQL systems, as well as hybrid solutions to natural language querying. We also describe how to extend the one-shot query approaches to dialogue, taking advantage of the context for disambiguation, in Section 4. In Section 5, we recount various benchmarks designed for evaluating natural language interfaces to data. Finally, we conclude with a discussion on challenges that need to be addressed before these systems can be widely adopted in Section 6.

References

Abzianidze, L. and J. Bos. (2017). "Towards Universal Semantic Tagging". *CoRR*. abs/1709.10381.

Aditya, B., G. Bhalotia, S. Chakrabarti, A. Hulgeri, C. Nakhe, P. Parag, and S. Sudarshan. (2002). "BANKS: Browsing and Keyword Searching in Relational Databases". In: *VLDB*. 1083–1086.

Affolter, K., K. Stockinger, and A. Bernstein. (2019). "A comparative survey of recent natural language interfaces for databases". *The VLDB Journal*. 28(5): 793–819.

Aghajanyan, A., J. Maillard, A. Shrivastava, K. Diedrick, M. Haeger, H. Li, Y. Mehdad, V. Stoyanov, A. Kumar, M. Lewis, and S. Gupta. (2020). "Conversational Semantic Parsing". *CoRR*. abs/2009.13655.

Ahmetaj, S., V. Efthymiou, R. Fagin, P. G. Kolaitis, C. Lei, F. Özcan, and L. Popa. (2021). "Ontology-Enriched Query Answering on Relational Databases". In: *AAAI*. 15247–15254.

Albawi, S., T. A. Mohammed, and S. Al-Zawi. (2017). "Understanding of a convolutional neural network". In: *2017 International Conference on Engineering and Technology*. 1–6.

Amazon. (2018). "Amazon Alexa". URL: https://developer.amazon.com/alexa.

Amazon. (2021). "Amazon QuickSight". URL: https://aws.amazon.com/quicksight/.

Apple. (2018). "Siri". URL: https://www.apple.com/ios/siri/.

Araci, D. (2019). "FinBERT: Financial Sentiment Analysis with Pretrained Language Models". *CoRR*. abs/1908.10063.

Asakura, T., J. Kim, Y. Yamamoto, Y. Tateisi, and T. Takagi. (2018). "A Quantitative Evaluation of Natural Language Question Interpretation for Question Answering Systems". In: *The 8th Joint International Semantic Technology Conference*. 215–231.

Asghar, N., P. Poupart, J. Hoey, X. Jiang, and L. Mou. (2017). "Affective Neural Response Generation". *CoRR*. abs/1709.03968.

"Ask Data | Tableau Software". (2021). URL: https://www.tableau.com/products/new-features/ask-data.

Auer, S., C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. G. Ives. (2007). "DBpedia: A Nucleus for a Web of Open Data". In: *ISWC*. 722–735.

Baader, F., I. Horrocks, C. Lutz, and U. Sattler. (2017). *An Introduction to Description Logic*. Cambridge University Press.

Bahdanau, D., K. Cho, and Y. Bengio. (2015). "Neural Machine Translation by Jointly Learning to Align and Translate". *CoRR*. abs/1409.0473.

Baik, C., Z. Jin, M. J. Cafarella, and H. V. Jagadish. (2020). "Duoquest: A Dual-Specification System for Expressive SQL Queries". In: *SIGMOD*. 2319–2329.

Banaee, H., M. U. Ahmed, and A. Loutfi. (2013). "Towards NLG for Physiological Data Monitoring with Body Area Networks". In: *Proceedings of the 14th European Workshop on Natural Language Generation*. 193–197.

Basik, F., B. Hättasch, A. Ilkhechi, A. Usta, S. Ramaswamy, P. Utama, N. Weir, C. Binnig, and U. Çetintemel. (2018). "DBPal: A Learned NL-Interface for Databases". In: *SIGMOD*. 1765–1768.

Bast, H. and E. Haussmann. (2015). "More Accurate Question Answering on Freebase". In: *CIKM*. 1431–1440.

Beltagy, I., K. Lo, and A. Cohan. (2019). "SciBERT: A Pretrained Language Model for Scientific Text". In: *EMNLP-IJCNLP*. 3615–3620.

Ben Abacha, A. and P. Zweigenbaum. (2015). "MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies". *Inf. Process. Manage.* 51(5): 570–594.

Bergamaschi, S., F. Guerra, M. Interlandi, R. T. Lado, and Y. Velegrakis. (2016). “Combining user and database perspective for solving keyword queries over relational databases”. *Inf. Syst.* 55: 1–19.

Bethard, S., P. V. Ogren, and L. Becker. (2014). “ClearTK 2.0: Design Patterns for Machine Learning in UIMA”. In: *LREC*. 3289–3293.

Beveridge, M. and J. Fox. (2006). “Automatic Generation of Spoken Dialogue from Medical Plans and Ontologies”. *J. of Biomedical Informatics*. 39(5): 482–499.

Bing-Hwang Juang and S. Furui. (2000). “Automatic recognition and understanding of spoken language - a first step toward natural human-machine communication”. *Proceedings of the IEEE*. 88(8): 1142–1165.

Bjerva, J., B. Plank, and J. Bos. (2016). “Semantic Tagging with Deep Residual Networks”. *CoRR*. abs/1609.07053.

Blunschi, L., C. Jossen, D. Kossmann, M. Mori, and K. Stockinger. (2012). “SODA: Generating SQL for Business Users”. *PVLDB*. 5(10): 932–943.

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. (2020). “Language Models are Few-Shot Learners”. *CoRR*. abs/2005.14165.

Brunner, U. and K. Stockinger. (2020). “ValueNet: A Neural Text-to-SQL Architecture Incorporating Values”. *CoRR*. abs/2006.00888.

Budzianowski, P., T. Wen, B. Tseng, I. Casanueva, S. Ultes, O. Ramadan, and M. Gasic. (2018). “MultiWOZ - A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling”. In: *EMNLP*. 5016–5026.

Castro Ferreira, T., I. Calixto, S. Wubben, and E. Krahmer. (2017). “Linguistic realisation as machine translation: Comparing different MT models for AMR-to-text generation”. In: *Proceedings of the 10th International Conference on Natural Language Generation*. 1–10.

Chaudhuri, S. and U. Dayal. (1997). “An Overview of Data Warehousing and OLAP Technology”. *SIGMOD Rec.* 26(1): 65–74.

Chen, Y.-N., A. Celikyilmaz, and D. Hakkani-Tür. (2018). "Deep Learning for Dialogue Systems". In: *Proceedings of the 27th International Conference on Computational Linguistics: Tutorial Abstracts*. 25–31.

Cho, K., B. van Merrienboer, D. Bahdanau, and Y. Bengio. (2014). "On the Properties of Neural Machine Translation: Encoder-Decoder Approaches". *CoRR*. abs/1409.1259.

"Cognos Assistant". (2021). URL: https://tinyurl.com/u3sdaxa.

Dai, Z., Z. Yang, Y. Yang, J. Carbonell, Q. Le, and R. Salakhutdinov. (2019). "Transformer-XL: Attentive Language Models beyond a Fixed-Length Context". In: *ACL*. 2978–2988.

Danilevsky, M., K. Qian, R. Aharonov, Y. Katsis, B. Kawas, and P. Sen. (2020). "A Survey of the State of Explainable AI for Natural Language Processing". *CoRR*. abs/2010.00711.

Dethlefs, N. (2014). "Context-Sensitive Natural Language Generation: From Knowledge-Driven to Data-Driven Techniques". *Lang. Linguistics Compass*. 8(3): 99–115.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *NAACL*. 4171–4186.

Dhamdhere, K., K. S. McCurley, R. Nahmias, M. Sundararajan, and Q. Yan. (2017). "Analyza: Exploring Data with Conversation". In: *IUI*. 493–504.

Dong, L. and M. Lapata. (2016). "Language to Logical Form with Neural Attention". *CoRR*. abs/1601.01280.

Dozat, T. and C. D. Manning. (2016). "Deep Biaffine Attention for Neural Dependency Parsing". *CoRR*. abs/1611.01734.

Dubey, M., D. Banerjee, A. Abdelkawi, and J. Lehmann. (2019). "LC-QuAD 2.0: A Large Dataset for Complex Question Answering over Wikidata and DBpedia". In: *ISWC*. Vol. 11779. *Lecture Notes in Computer Science*. 69–78.

Dusek, O. and F. Jurcicek. (2016). "Sequence-to-Sequence Generation for Spoken Dialogue via Deep Syntax Trees and Strings". In: *ACL*. 45–51.

Fagin, R., P. G. Kolaitis, R. J. Miller, and L. Popa. (2005). "Data exchange: semantics and query answering". *Theor. Comput. Sci.* 336(1): 89–124.

Ferragina, P. and U. Scaiella. (2010). "TAGME: on-the-fly annotation of short text fragments (by wikipedia entities)". In: *CIKM*. 1625–1628.

Fitzpatrick, K. K., A. Darcy, and M. Vierhile. (2017). "Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial". *JMIR Ment Health*. 4(2): e19.

Forney, G. (1973). "The viterbi algorithm". *Proceedings of the IEEE*. 61(3): 268–278.

Francia, M., E. Gallinucci, and M. Golfarelli. (2020). "Towards Conversational OLAP". In: *DOLAP@EDBT/ICDT*. Vol. 2572. 6–15.

Gao, J., M. Galley, and L. Li. (2018). "Neural Approaches to Conversational AI". *CoRR*. abs/1809.08267.

Gao, S., A. Sethi, S. Agarwal, T. Chung, and D. Hakkani-Tür. (2019). "Dialog State Tracking: A Neural Reading Comprehension Approach". *CoRR*. abs/1908.01946.

Garoufi, K. (2014). "Planning-Based Models of Natural Language Generation". *Language and Linguistics Compass*. 8(Jan.).

Gatt, A. and E. Krahmer. (2018). "Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation". *J. Artif. Int. Res.* 61(1): 65–170.

Giorgino, T., I. Azzini, C. Rognoni, S. Quaglini, M. Stefanelli, R. Gretter, and D. Falavigna. (2005). "Automated spoken dialogue system for hypertensive patient home management". *International Journal of Medical Informatics*. 74(2): 159–167.

Gkini, O., T. Belmpas, G. Koutrika, and Y. E. Ioannidis. (2021). "An In-Depth Benchmarking of Text-to-SQL Systems". In: *SIGMOD*. 632–644.

Google. (2021a). "Google Looker". URL: https://www.looker.com/google-cloud/.

Google. (2021b). "GoogleAssitant". URL: https://assistant.google.com.

Google. (2021c). "Lamda". URL: https://blog.google/technology/ai/lamda/.

Guo, J., Z. Zhan, Y. Gao, Y. Xiao, J. Lou, T. Liu, and D. Zhang. (2019). "Towards Complex Text-to-SQL in Cross-Domain Database with Intermediate Representation". *CoRR*. abs/1905.08205.

Gur, I., S. Yavuz, Y. Su, and X. Yan. (2018). "DialSQL: Dialogue Based Structured Query Generation". In: *ACL*. 1339–1349.

Hao, J., C. Lei, V. Efthymiou, A. Quamar, F. Özcan, Y. Sun, and W. Wang. (2021). "MEDTO: Medical Data to Ontology Matching Using Hybrid Graph Neural Networks". In: *SIGKDD*. 2946–2954.

He, H. and J. D. Choi. (2019). "Establishing Strong Baselines for the New Decade: Sequence Tagging, Syntactic and Semantic Parsing with BERT". *CoRR*. abs/1908.04943.

Henderson, M., I. Casanueva, N. Mrksic, P. Su, T. Wen, and I. Vulic. (2019a). "ConveRT: Efficient and Accurate Conversational Representations from Transformers". *CoRR*. abs/1911.03688.

Henderson, M., I. Casanueva, N. Mrkšić, P.-H. Su, T.-H. Wen, and I. Vulić. (2020). "ConveRT: Efficient and Accurate Conversational Representations from Transformers". In: *EMNLP*. 2161–2174.

Henderson, M., B. Thomson, and S. Young. (2013). "Deep Neural Network Approach for the Dialog State Tracking Challenge". In: *SIGDIAL*. 467–471.

Henderson, M., B. Thomson, and S. J. Young. (2014). "Word-Based Dialog State Tracking with Recurrent Neural Networks". In: *SIGDIAL*. 292–299.

Henderson, M., I. Vulic, D. Gerz, I. Casanueva, P. Budzianowski, S. Coope, G. Spithourakis, T. Wen, N. Mrksic, and P. Su. (2019b). "Training Neural Response Selection for Task-Oriented Dialogue Systems". *CoRR*. abs/1906.01543.

Herzig, J., M. Shmueli-Scheuer, T. Sandbank, and D. Konopnicki. (2017). "Neural Response Generation for Customer Service based on Personality Traits". In: *Proceedings of the 10th International Conference on Natural Language Generation*. 252–256.

Hochreiter, S. and J. Schmidhuber. (1997). "Long Short-Term Memory". *Neural Computation*. 9(8): 1735–1780.

Holzinger, A., G. Langs, H. Denk, K. Zatloukal, and H. Müller. (2019). "Causability and explainability of artificial intelligence in medicine". *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 9(4).

Howard, J. and S. Ruder. (2018). "Universal Language Model Fine-tuning for Text Classification". In: *ACL*. 328–339.

Huang, Z., W. Xu, and K. Yu. (2015). "Bidirectional LSTM-CRF Models for Sequence Tagging". *CoRR*. abs/1508.01991.

Hui, B., X. Shi, R. Geng, B. Li, Y. Li, J. Sun, and X. Zhu. (2021). "Improving Text-to-SQL with Schema Dependency Learning". *CoRR*. abs/2103.04399.

Hussain, S., O. Sianaki, and N. Ababneh. (2019). "A Survey on Conversational Agents/Chatbots Classification and Design Techniques". In: 946–956.

Hwang, W., J. Yim, S. Park, and M. Seo. (2019). "A Comprehensive Exploration on WikiSQL with Table-Aware Word Contextualization". *CoRR*. abs/1902.01069.

Iyyer, M., W.-t. Yih, and M.-W. Chang. (2017). "Search-based Neural Structured Learning for Sequential Question Answering". In: *ACL*. 1821–1831.

Jammi, M., J. Sen, A. R. Mittal, S. Verma, V. Pahuja, R. Ananthanarayanan, P. Lohia, H. Karanam, D. Saha, and K. Sankaranarayanan. (2018). "Tooling Framework for Instantiating Natural Language Querying System". *PVLDB*. 11(12): 2014–2017.

Jiao, X., Y. Yin, L. Shang, X. Jiang, X. Chen, L. Li, F. Wang, and Q. Liu. (2020). "TinyBERT: Distilling BERT for Natural Language Understanding". In: *EMNLP*. 4163–4174.

Joshi, M., D. Chen, Y. Liu, D. S. Weld, L. Zettlemoyer, and O. Levy. (2019). "SpanBERT: Improving Pre-training by Representing and Predicting Spans". *CoRR*. abs/1907.10529.

Katsogiannis-Meimarakis, G. and G. Koutrika. (2021a). "A Deep Dive into Deep Learning Approaches for Text-to-SQL Systems". In: *SIGMOD*. 2846–2851.

Katsogiannis-Meimarakis, G. and G. Koutrika. (2021b). "Deep Learning Approaches for Text-to-SQL Systems". In: *EDBT*. 710–713.

Kaufmann, E. and A. Bernstein. (2010). "Evaluating the usability of natural language query languages and interfaces to Semantic Web knowledge bases". *J. Web Semant.* 8(4): 377–393.

Kim, W. (1982). "On Optimizing an SQL-like Nested Query". *ACM Trans. Database Syst.* 7(3): 443–469.

Kim, Y. (2014). "Convolutional Neural Networks for Sentence Classification". In: *EMNLP*. 1746–1751.

Kim, Y., Y. Jernite, D. Sontag, and A. M. Rush. (2016). "Character-Aware Neural Language Models". In: *AAAI*. 2741–2749.

Koutrika, G., A. Simitsis, and Y. E. Ioannidis. (2006). "Précis: The Essence of a Query Answer". In: *ICDE*. 69–78.

Kuchmann-Beauger, N., F. Brauer, and M.-A. Aufaure. (2013). "QUASL: A framework for question answering and its Application to business intelligence". In: *IEEE 7th International Conference on Research Challenges in Information Science (RCIS)*. 1–12.

Lai, G., Q. Xie, H. Liu, Y. Yang, and E. Hovy. (2017). "RACE: Large-scale ReAding Comprehension Dataset From Examinations". In: *EMNLP*. 785–794.

Lan, Z., M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut. (2019). "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations". *CoRR*. abs/1909.11942.

Lebret, R., D. Grangier, and M. Auli. (2016). "Neural Text Generation from Structured Data with Application to the Biography Domain". In: *EMNLP*. 1203–1213.

Lee, J.-O. and D.-K. Baik. (1999). "SemQL: A Semantic Query Language for Multidatabase Systems". In: *CIKM*. 259–266.

Lee, J., W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang. (2019). "BioBERT: a pre-trained biomedical language representation model for biomedical text mining". *Bioinformatics*. 36(4): 1234–1240.

Lei, C., V. Efthymiou, R. Geis, and F. Özcan. (2020). "Expanding Query Answers on Medical Knowledge Bases". In: *EDBT*. 567–578.

Lei, C., F. Özcan, A. Quamar, A. R. Mittal, J. Sen, D. Saha, and K. Sankaranarayanan. (2018). "Ontology-Based Natural Language Query Interfaces for Data Exploration". *IEEE Data Eng. Bull.* 41(3): 52–63.

Li, F. and H. V. Jagadish. (2014a). "Constructing an Interactive Natural Language Interface for Relational Databases". *PVLDB*. 8(1): 73–84.

Li, F. and H. V. Jagadish. (2016). "Understanding Natural Language Queries over Relational Databases". *SIGMOD Record*. 45(1): 6–13.

Li, F. and H. V. Jagadish. (2014b). "NaLIR: an interactive natural language interface for querying relational databases". In: *SIGMOD*. 709–712.

Li, J., Y. Li, X. Wang, and W.-C. Tan. (2020). "Deep or Simple Models for Semantic Tagging? It Depends on Your Data". *PVLDB*. 13(12): 2549–2562.

Li, Y. and D. Rafiei. (2017). "Natural Language Data Management and Interfaces: Recent Development and Open Challenges". In: *SIGMOD*. 1765–1770.

Li, Y., H. Yang, and H. V. Jagadish. (2005). "NaLIX: An Interactive Natural Language Interface for Querying XML". In: *SIGMOD*. 900–902.

Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. (2019). "RoBERTa: A Robustly Optimized BERT Pretraining Approach". *CoRR*. abs/1907.11692.

Lyons, G., V. Tran, C. Binnig, U. Cetintemel, and T. Kraska. (2016). "Making the Case for Query-by-Voice with EchoQuery". In: *SIGMOD*. 2129–2132.

Lyu, Q., K. Chakrabarti, S. Hathi, S. Kundu, J. Zhang, and Z. Chen. (2020). "Hybrid Ranking Network for Text-to-SQL". *CoRR*. abs/2008.04759.

Mairesse, F. and S. Young. (2014). "Stochastic Language Generation in Dialogue using Factored Language Models". *Computational Linguistics*. 40(4): 763–799.

Mallios, S. and N. G. Bourbakis. (2016). "A survey on human machine dialogue systems". In: *7th International Conference on Information, Intelligence, Systems & Applications (IISA)*. 1–7.

Marneffe, M. de, B. MacCartney, and C. D. Manning. (2006). "Generating Typed Dependency Parses from Phrase Structure Parses". In: *LREC*. 449–454.

McTear, M. F. (2002). "Spoken Dialogue Technology: Enabling the Conversational User Interface". *ACM Comput. Surv.* 34(1): 90–169.

Microsoft. (2018). "Microsoft Cortana". URL: https://www.microsoft.com/en-us/windows/cortana.

"MicroStrategy". (2021). URL: https://community.microstrategy.com/s/article/Natural-Language-Query-in-A-Nutshell-MicroStrategy-11-0?language=en_US.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. (2013). "Efficient Estimation of Word Representations in Vector Space". In: *ICLR*.

Miner, A. S., A. Milstein, S. Schueller, *et al.* (2016). "Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health". *JAMA Internal Medicine*. 176(5): 619–625.

Mrksic, N., D. Ó. Séaghdha, T. Wen, B. Thomson, and S. J. Young. (2016). "Neural Belief Tracker: Data-Driven Dialogue State Tracking". *CoRR*. abs/1606.03777.

Negi, S. and P. Buitelaar. (2015). "Towards the Extraction of Customer-to-Customer Suggestions from Reviews". In: *EMNLP*. 2159–2167.

Net, C. (2021). "Concept Net a freely-available semantic network." URL: https://conceptnet.io/.

Orogat, A., I. Liu, and A. El-Roby. (2021). "CBench: Towards Better Evaluation of Question Answering Over Knowledge Graphs". *PVLDB*. 14(8): 1325–1337.

Özcan, F., C. Lei, A. Quamar, and V. Efthymiou. (2021). "Semantic Enrichment of Data for AI Applications". In: *Proceedings of the Fifth Workshop on Data Management for End-To-End Machine Learning*.

Özcan, F., A. Quamar, J. Sen, C. Lei, and V. Efthymiou. (2020). "State of the Art and Open Challenges in Natural Language Interfaces to Data". In: *SIGMOD*. 2629–2636.

Pasupat, P. and P. Liang. (2015). "Compositional Semantic Parsing on Semi-Structured Tables". In: *ACL*. 1470–1480.

Pennington, J., R. Socher, and C. Manning. (2014). "GloVe: Global Vectors for Word Representation". In: *EMNLP*. 1532–1543.

Peters, M. E., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. (2018). "Deep Contextualized Word Representations". In: *NAACL*. 2227–2237.

"Power BI Platform". (2021). URL: https://powerbi.microsoft.com/en-us/.

Preece, A. D. (2018). "Asking 'Why' in AI: Explainability of intelligent systems - perspectives and challenges". *Intell. Syst. Account. Finance Manag.* 25(2): 63–72.

Punjani, D., M. Iliakis, T. Stefou, K. Singh, A. Both, M. Koubarakis, I. Angelidis, K. Bereta, T. Beris, D. Bilidas, T. Ioannidis, N. Karalis, C. Lange, D. Pantazi, C. Papaloukas, and G. Stamoulis. (2020). “Template-Based Question Answering over Linked Geospatial Data”. *CoRR*. abs/2007.07060.

Qiu, X., T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang. (2020). “Pre-trained models for natural language processing: A survey”. *Science in China E: Technological Sciences*. 63(10): 1872–1897.

Quamar, A., C. Lei, D. Miller, F. Ozcan, J. Kreulen, R. J. Moore, and V. Efthymiou. (2020a). “An Ontology-Based Conversation System for Knowledge Bases”. In: *SIGMOD*. 361–376.

Quamar, A., F. Özcan, D. Miller, R. J. Moore, R. Niehus, and J. Kreulen. (2020b). “Conversational BI: An Ontology-Driven Conversation System for Business Intelligence Applications”. *PVLDB*. 13(12): 3369–3381.

Radford, A. and K. Narasimhan. (2018). “Improving Language Understanding by Generative Pre-Training”.

Radford, A., J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever. (2019). “Language Models are Unsupervised Multitask Learners”. *OpenAI blog*. 1(8): 9.

Radziwill, N. M. and M. C. Benton. (2017). “Evaluating Quality of Chatbots and Intelligent Conversational Agents”. *CoRR*. abs/1704.04579.

Rastogi, A., D. Hakkani-Tür, and L. P. Heck. (2017). “Scalable Multi-Domain Dialogue State Tracking”. *CoRR*. abs/1712.10224.

Reddy, S., D. Chen, and C. D. Manning. (2019). “CoQA: A Conversational Question Answering Challenge”. *Transactions of the Association for Computational Linguistics*. 7: 249–266.

Reiter, E. and R. Dale. (1997). “Building Applied Natural Language Generation Systems”. *Nat. Lang. Eng.* 3(1): 57–87.

Reiter, E. and R. Dale. (2000). *Building Natural Language Generation Systems*. USA: Cambridge University Press. ISBN: 0521620368.

Richardson, J., K. Schlegel, R. Sallam, A. Kronz, and J. Sun. (2021). “Top Trends in Data and Analytics for 2021: The Rise of the Augmented Consumer”. URL: https://www.gartner.com/doc/reprints?id=1-25H0EUUY&ct=210317&st=sb.

Rieser, V. and O. Lemon. (2016). "Natural Language Generation as Planning under Uncertainty Using Reinforcement Learning". *CoRR*. abs/1606.04686.

Rubin, O. and J. Berant. (2020). "SmBoP: Semi-autoregressive Bottom-up Semantic Parsing". *CoRR*. abs/2010.12412.

Saha, D., A. Floratou, K. Sankaranarayanan, U. F. Minhas, A. R. Mittal, and F. Özcan. (2016). "ATHENA: An Ontology-Driven System for Natural Language Querying over Relational Data Stores". *PVLDB*. 9(12): 1209–1220.

Sanh, V., L. Debut, J. Chaumond, and T. Wolf. (2019). "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter". *CoRR*. abs/1910.01108.

Santhanam, S. and S. Shaikh. (2019). "A Survey of Natural Language Generation Techniques with a Focus on Dialogue Systems - Past, Present and Future Directions". *CoRR*. abs/1906.00500.

Santos, C. N. dos and V. Guimarães. (2015). "Boosting Named Entity Recognition with Neural Character Embeddings". *CoRR*. abs/1505.05008.

See, A., P. J. Liu, and C. D. Manning. (2017). "Get To The Point: Summarization with Pointer-Generator Networks". In: *ACL*. 1073–1083.

Sekine, S. and C. Nobata. (2004). "Definition, Dictionaries and Tagger for Extended Named Entity Hierarchy". In: *LREC*.

Semarchy. (2021). "The SemQL Language". URL: https://www.semarchy.com/doc/semarchy-xdm/xdm/5.3/SemQL/overview.html.

Sen, J., C. Lei, A. Quamar, F. Özcan, V. Efthymiou, A. Dalmia, G. Stager, A. R. Mittal, D. Saha, and K. Sankaranarayanan. (2020). "ATHENA++: Natural Language Querying for Complex Nested SQL Queries". *Proc. VLDB Endow.* 13(11): 2747–2759.

Sen, J., F. Özcan, A. Quamar, G. Stager, A. R. Mittal, M. Jammi, C. Lei, D. Saha, and K. Sankaranarayanan. (2019). "Natural Language Querying of Complex Business Intelligence Queries". In: *SIGMOD*. 1997–2000.

Shao, Y., C. Hardmeier, and J. Nivre. (2016). "Multilingual Named Entity Recognition using Hybrid Neural Networks". In: *The Sixth Swedish Language Technology Conference (SLTC)*.

Sherstinsky, A. (2020). “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network”. *Physica D: Nonlinear Phenomena*. 404(Mar.): 132306.

Shi, T., K. Tatwawadi, K. Chakrabarti, Y. Mao, O. Polozov, and W. Chen. (2018). “IncSQL: Training Incremental Text-to-SQL Parsers with Non-Deterministic Oracles”. *CoRR*. abs/1809.05054.

Simitsis, A., G. Koutrika, and Y. E. Ioannidis. (2008). “Précis: from unstructured keywords as queries to structured databases as answers”. *VLDB J.* 17(1): 117–149.

Song, D., F. Schilder, C. Smiley, C. Brew, T. Zielund, H. Bretz, R. Martin, C. Dale, J. Duprey, T. Miller, and J. Harrison. (2015). “TR Discover: A Natural Language Interface for Querying and Analyzing Interlinked Datasets”. In: *ISWC*. 21–37.

Stoyanovich, J., S. Abiteboul, and G. Miklau. (2016). “Data Responsibly: Fairness, Neutrality and Transparency in Data Analysis”. In: *EDBT*. 718–719.

Stoyanovich, J., B. Howe, and H. V. Jagadish. (2020). “Responsible Data Management”. *PVLDB*. 13(12): 3474–3488.

Tata, S. and G. M. Lohman. (2008). “SQAK: Doing More with Keywords”. In: *SIGMOD*. 889–902.

Taylor, W. L. (1953). “Cloze Procedure: A New Tool for Measuring Readability”. In: *Journalism Quarterly*.

Thomson, B. and S. Young. (2010). “Bayesian update of dialogue state: A POMDP framework for spoken dialogue systems”. *Computer Speech and Language*. 24(4): 562–588.

Unger, C., L. Bühmann, J. Lehmann, A. N. Ngomo, D. Gerber, and P. Cimiano. (2012). “Template-based question answering over RDF data”. In: *WWW*. 639–648.

Usta, A., A. Karakayali, and Ö. Ulusoy. (2021). “DBTagger: Multi-Task Learning for Keyword Mapping in NLIDBs Using Bi-Directional Recurrent Neural Networks”. *PVLDB*. 14(5): 813–821.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. (2017). “Attention Is All You Need”. *CoRR*. abs/1706.03762.

Vinyals, O., M. Fortunato, and N. Jaitly. (2015). “Pointer Networks”. In: *Advances in Neural Information Processing Systems*. Vol. 28.

Walter, S., C. Unger, P. Cimiano, and D. Bär. (2012). "Evaluation of a Layered Approach to Question Answering over Linked Data". In: *ISWC*. 362–374.

Waltinger, U., D. Tecuci, M. Olteanu, V. Mocanu, and S. Sullivan. (2013). "USI Answers: Natural Language Question Answering Over (Semi-) Structured Industry Data". In: *IAAI*. 1471–1478.

Wang, B., R. Shin, X. Liu, O. Polozov, and M. Richardson. (2019). "RAT-SQL: Relation-Aware Schema Encoding and Linking for Text-to-SQL Parsers". *CoRR*. abs/1911.04942.

Wang, C., P. Huang, A. Polozov, M. Brockschmidt, and R. Singh. (2018a). "Execution-Guided Neural Program Decoding". *CoRR*. abs/1807.03100.

Wang, W., Y. Tian, H. Xiong, H. Wang, and W. Ku. (2018b). "A Transfer-Learnable Natural Language Interface for Databases". *CoRR*. abs/1809.02649.

Weir, N. and P. Utama. (2019). "Bootstrapping an End-to-End Natural Language Interface for Databases". In: *SIGMOD*. 1862–1864.

Wen, T.-H., M. Gašić, N. Mrkšić, P.-H. Su, D. Vandyke, and S. Young. (2015). "Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems". In: *EMNLP*. 1711–1721.

Williams, J. D. and S. Young. (2007). "Partially observable Markov decision processes for spoken dialog systems". *Computer Speech and Language*. 21(2): 393–422.

Wiseman, S., A. Backurs, and K. Stratos. (2021). "Data-to-text Generation by Splicing Together Nearest Neighbors". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 4283–4299.

Wu, C., S. C. H. Hoi, R. Socher, and C. Xiong. (2020). "ToD-BERT: Pre-trained Natural Language Understanding for Task-Oriented Dialogues". *CoRR*. abs/2004.06871.

Wu, Z. and M. S. Palmer. (1994). "Verb Semantics and Lexical Selection". In: *ACL*. 133–138.

Xu, X., C. Liu, and D. Song. (2017). "SQLNet: Generating Structured Queries From Natural Language Without Reinforcement Learning". *CoRR*. abs/1711.04436.

Yang, Z., Z. Dai, Y. Yang, J. G. Carbonell, R. Salakhutdinov, and Q. V. Le. (2019). "XLNet: Generalized Autoregressive Pretraining for Language Understanding". *CoRR*. abs/1906.08237.

Young, S. J., M. Gasic, B. Thomson, and J. D. Williams. (2013). "POMDP-Based Statistical Spoken Dialog Systems: A Review". *Proceedings of the IEEE*. 101(5): 1160–1179.

Young, T., D. Hazarika, S. Poria, and E. Cambria. (2018). "Recent trends in deep learning based natural language processing". *IEEE Computational Intelligence Magazine*. 13(3): 55–75.

Yu, T., Z. Li, Z. Zhang, R. Zhang, and D. R. Radev. (2018a). "TypeSQL: Knowledge-Based Type-Aware Neural Text-to-SQL Generation". In: *NAACL-HLT*. 588–594.

Yu, T., R. Zhang, H. Er, S. Li, E. Xue, B. Pang, X. V. Lin, Y. C. Tan, T. Shi, Z. Li, Y. Jiang, M. Yasunaga, S. Shim, T. Chen, A. R. Fabbri, Z. Li, L. Chen, Y. Zhang, S. Dixit, V. Zhang, C. Xiong, R. Socher, W. S. Lasecki, and D. R. Radev. (2019a). "CoSQL: A Conversational Text-to-SQL Challenge Towards Cross-Domain Natural Language Interfaces to Databases". *CoRR*. abs/1909.05378.

Yu, T., R. Zhang, A. Polozov, C. Meek, and A. H. Awadallah. (2021). "SCoRe: Pre-Training for Context Representation in Conversational Semantic Parsing". In: *ICLR*.

Yu, T., R. Zhang, K. Yang, M. Yasunaga, D. Wang, Z. Li, J. Ma, I. Li, Q. Yao, S. Roman, Z. Zhang, and D. R. Radev. (2018b). "Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task". In: *EMNLP*. 3911–3921.

Yu, T., R. Zhang, M. Yasunaga, Y. C. Tan, X. V. Lin, S. Li, H. Er, I. Li, B. Pang, T. Chen, E. Ji, S. Dixit, D. Proctor, S. Shim, J. Kraft, V. Zhang, C. Xiong, R. Socher, and D. R. Radev. (2019b). "SParC: Cross-Domain Semantic Parsing in Context". *CoRR*. abs/1906.02285.

Zenz, G., X. Zhou, E. Minack, W. Siberski, and W. Nejdl. (2009). "From keywords to semantic queries - Incremental query construction on the semantic web". *J. Web Semant.* 7(3): 166–176.

Zhang, R., T. Yu, H. Er, S. Shim, E. Xue, X. V. Lin, T. Shi, C. Xiong, R. Socher, and D. Radev. (2019). "Editing-Based SQL Query Generation for Cross-Domain Context-Dependent Questions". In: *EMNLP-IJCNLP*. 5337–5348.

Zhang, Y., S. Sun, M. Galley, Y.-C. Chen, C. Brockett, X. Gao, J. Gao, J. Liu, and B. Dolan. (2020a). "DIALOGPT : Large-Scale Generative Pre-training for Conversational Response Generation". In: *ACL*. 270–278.

Zhang, Z., Y. Wu, H. Zhao, Z. Li, S. Zhang, X. Zhou, and X. Zhou. (2020b). "Semantics-Aware BERT for Language Understanding". In: *AAAI*. 9628–9635.

Zhong, V., C. Xiong, and R. Socher. (2017). "Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning". *CoRR*. abs/1709.00103.