# Neural Database Operator Model

James Thorne[1,2], Majid Yazdani[2], Marzieh Saeidi[2], Sebastian Riedel[2], and Alon Halevy[2]

[1]University of Cambridge
[2]Facebook AI
jt719@cam.ac.uk, {myazdani,marzieh,sriedel,ayh}@fb.com

## 1 Introduction

Our goal is to answer queries over facts stored in a text memory. The key challenge in NeuralDBs (Thorne et al., 2020), compared to open-book NLP such as question answering (Rajpurkar et al., 2016, *inter alia)*, is that possibly thousands of facts must be aggregated to provide a single answer, without direct supervision. The challenges represented in NeuralDBs are important for both the NLP and database communities alike: discrete reasoning over text (Dua et al., 2019), retriever-based QA (Dunn et al., 2017) and multi-hop QA (Welbl et al., 2018; Yang et al., 2018) are common components.

## 2 Problem Definition and Model

Let $D$ be a database of facts stored in free-form text $\{d_1, \ldots, d_i\}$. The NeuralDB is a function $f : (x, D) \rightarrow y$ yielding an answer $y$ to natural language query $x$. To generate $y$, relevant facts $Z \subseteq D$ are aggregated (e.g. count, argmax), or used for boolean or extractive question answering.

Two limitations preclude direct application of modern neural architectures, such as transformers (Vaswani et al., 2017): lack of stable numerical reasoning prevents computation, and the memory complexity of self-attention limits maximum sequence length. The former can obviated through constructing shallow programs (e.g. Andor et al. (2019)) over extracted information and the latter can be mitigated through optimizations and linearizaitons such as Fusion in Decoder (Izacard and Grave, 2020) and LinFormer (Wang et al., 2020).

The architecture we present mitigates both limitations, enabling discrete reasoning over large databases, by combining retrieval, query-based derivation of facts, and discrete operators to answer a range of query types over large databases. The core component of our NeuralDB (Figure 1) is a seq2seq model that conditionally transduces a fact stored in text to a machine readable representation that is input into a non-neural operator for aggregation if needed to answer a query.

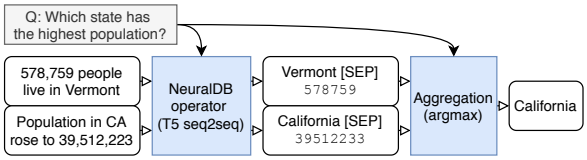


Figure 1: NeuralDB architecture showing unary projection of facts stored in texts to an intermediate representation with an argmax computation (selected with a query type classifier)

## 3 Results and Discussion

We generate training data (facts, questions and answers) from 26 Wikidata relations with 100k sampled entities using templates. Subjects are disjoint between the training and test sets.

**Baseline end-to-end neural model:** with small DBs ($< 50$ facts), concatenating facts and the query gives high EM on atomic queries and simple joins (99%) but low EM for counting (79%) and returning a set (90% F1). For large aggregations, token limits prevent encoding long sequences causing errors. Using fusion in decoder (Izacard and Grave, 2020) yielded no improvements as intra-sentence attention is necessary in encoding. We scale to larger databases (10k facts) using TFIDF or DPR for fact selection with similar results.

**NeuralDB operator:** performs high-precision selection and projects facts to a machine readable intermediate representation. For projections, we attain 98% EM for in-domain relations. Errors are induced when generating numbers and incorrectly resolving entity names. False negatives on out-of domain data are a limitation. Fine-tuning on 13 out of 26 relations and evaluating on all 26, EM falls to 86%. LM pre-training aids generalization as a randomly initialized model only attains 55% EM.

**Discussion:** NeuralDBs are a novel research area: querying unstructured data can be exploited by the database community. However, limitations of transformers limit application for neural aggregation. We propose a unary operator as a workaround, generating an intermediate form dependent on the query. LM pre-training aids generalization of this model to unseen relations. Further work is required to move from templated data to *natural* language.

## References


Daniel Andor, Luheng He, Kenton Lee, and Emily Pitler. 2019. Giving Bert a calculator: Finding operations and arguments with reading comprehension. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 2:5947–5952.

Dheeru Dua, Yizhong Wang, Pradeep Dasigi, Gabriel Stanovsky, Sameer Singh, and Matt Gardner. 2019. DROP: A Reading Comprehension Benchmark Requiring Discrete Reasoning Over Paragraphs. 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 2368–2378, Minneapolis, Minnesota. Association for Computational Linguistics.

Matthew Dunn, Levent Sagun, Mike Higgins, V. Ugur Guney, Volkan Cirik, and Kyunghyun Cho. 2017. SearchQA: A New Q&A Dataset Augmented with Context from a Search Engine.

Gautier Izacard and Edouard Grave. 2020. Leveraging Passage Retrieval with Generative Models for Open Domain Question Answering.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. SQuAD: 100,000+ Questions for Machine Comprehension of Text. pages 2383–2392.

James Thorne, Majid Yazdani, Marzieh Saeidi, Fabrizio Silvestri, Sebastian Riedel, and Alon Halevy. 2020. Neural databases.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Lilon Jones, Aidan Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA.

Sinong Wang, Belinda Z. Li, Madian Khabsa, Han Fang, and Hao Ma. 2020. Linformer: Self-Attention with Linear Complexity. 2048(2019).

Johannes Welbl, Pontus Stenetorp, and Sebastian Riedel. 2018. Constructing datasets for multi-hop reading comprehension across documents. *Transactions of the Association for Computational Linguistics*, 6:287–302.

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. 2018. HotpotQA: A dataset for diverse, explainable multi-hop question answering. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2369–2380, Brussels, Belgium. Association for Computational Linguistics.