ATLANTIC: Structure-Aware Retrieval-Augmented Language Model for Interdisciplinary Science

Sai Munikoti, Anurag Acharya, Sridevi Wagle, Sameera Horawalavithana
Pacific Northwest National Lab, USA

Motivation

- Existing RAG models offers an effective solution to domain adaptation by retrieving context from external knowledge, but they ignore structural relationship between the documents and less explored in scientific domains.
- We propose a novel structure-aware retrieval augmented language model that accommodates document structure during retrieval augmentation, and a novel evaluation metrics to measure the quality of retrieved documents on scientific tasks.

Methodology

- Our model (ATLANTIC) is based on the ATLAS architecture[1], a state-of-the-art RAG model, consisting of a BERT-based Retriever model that retrieves top-k passages and feeds to the Reader, i.e., T5 Language model (LM).
- In ATLANTIC, given the input query, Retriever retrieves top-k passages from the input text corpus based on semantic relationship.
- Unlike ATLAS, which directly passes these top-k passages to the LM, we obtain their structural encodings (embeddings) by leveraging their structural relationships in the form of a Heterogeneous Document Graph (HDG).
- HDG for text corpus is constructed using co-citation, co-topic, co-venue, and co-institutions information.
- The structural embeddings of retrieved passages are computed using a GNN trained on HDG.
- The structural embeddings are then appended with their semantic counterparts as obtained via Retriever encoder, before feeding them to the LM.
- Fig. 1 depicts the overview of ATLANTIC architecture
- Structural encoding provides extra context to the LM for generation, and it also improves the Retriever model to retrieve better passages

Model	In-distribution Performance				Out-of-distribution Performance			
	Accuracy		Evidence Generation		Accuracy		Evidence Generation	
	EM	F1	Relevance	Diversity	EM	F1	Relevance	Diversity
T5	0.833	0.87	N/A	N/A	0.579	0.72	N/A	N/A
ATLAS	0.844	**0.92**	0.694	5E-5	0.591	**0.75**	0.69	60E-5
ATLAS-Science	0.847	**0.92**	0.564	8E-5	0.578	0.73	0.571	100E-5
ATLANTIC	**0.850**	0.89	**1.159**	**10E-5**	**0.595**	0.60	**1.163**	**120E-5**

Table 1: Models' performance (both Retriever and LM) on n in-distribution (SciDocs-FoS) and out-of-distribution (SciDocs-MAG) benchmarks.

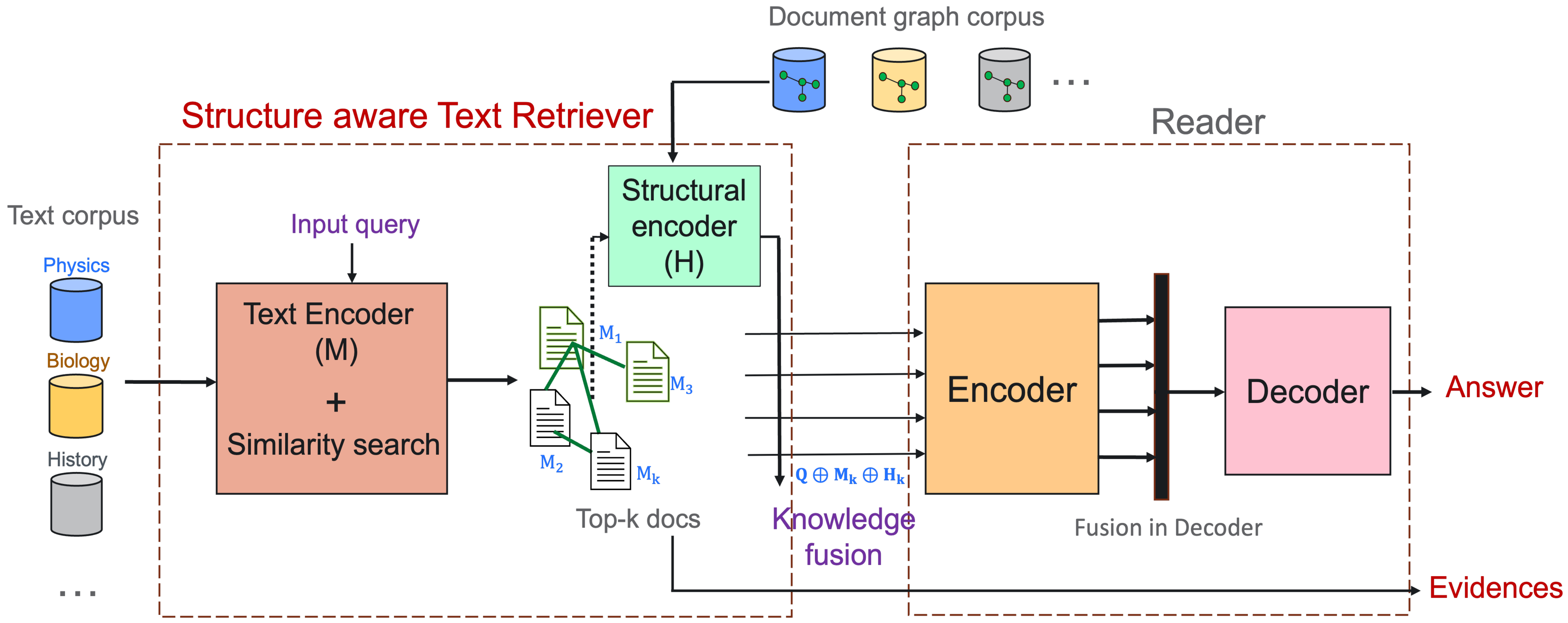
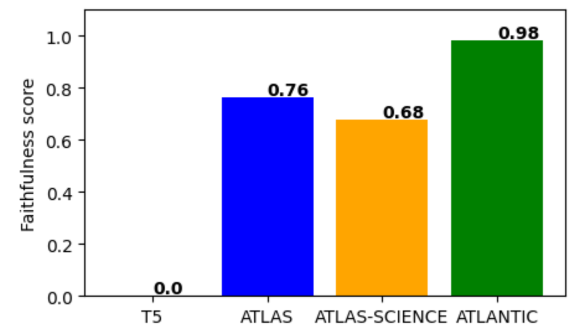


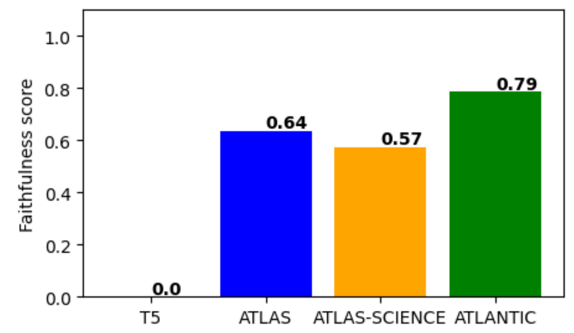
Fig 1: Proposed ATLANTIC framework (docs referred to passages). Structural embeddings (H_k) quantify the cross-document connections among the retrieved docs, which could be useful for multi-hop (multi document) reasoning.

Experiments and Results

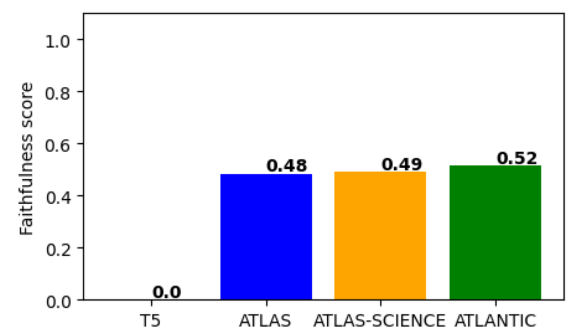
- **Dataset**: S2ORC[2] – (31.1M scientific papers across 19 domains)
- **Baseline**: T5-lm-adapt model[3] and original ATLAS model[1].
- **Benchmarks**: (i) SciRepEval[4]: Two classification tasks – Fields of study (FoS) and MAG. (ii) MMLU[5], 57 multi-choice question-answering from high school science topics.
- We pretrain the 220M ATLANTIC model on S2ROC corpus with query-side finetuning approach. It is also instruction finetuned for FoS task.
- We design three evaluation metrics to evaluate the relevance (*query relevance*), diversity of the extracted evidences (*diversity*) and *faithfulness score* to incorporate the performance of both retriever and language model .
- Research inference 1: Retrieving structural knowledge helps RAG models to perform better than just retrieving textual knowledge as illustrated via aggregated faithfulness scores in Fig 2, and further reinforced via individual performances (see Table 1).
- Research inference 2: Structure aware RAG models retrieve relevant passages to justify model predictions better than text-only models as evident via high relevance score in Tab 1.
- Specifically, structural knowledge helps the models to extract more faithful documents as evidence to support the model predictions



(a) FoS



(b) MAG



(c) MMLU

Fig 2: Faithfulness scores across FOS, MAG and MMLU benchmarks

Conclusions

- We propose new architecture (ATLANTIC) to integrate document structural knowledge into retrieval-augmented language models.
- We evaluate our model in multiple scientific benchmarks and demonstrate that retrieving structural knowledge helps retrieval-augmented language models to perform better overall than only retrieving textual knowledge.
- In the future, we will test our model on a wider range of scientific benchmarks and tasks (e.g., hypothesis generation)

Acknowledgement

This work was supported by the NNSA Office of Defense Nuclear Nonproliferation Research and Development, U.S. Department of Energy, and Pacific Northwest National Laboratory, which is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract DEAC05–76RLO1830.

Contact

Sai Munikoti
Pacific Northwest National Lab
Email:sai.munikoti@pnnl.gov
Website:https://github.com/pnnl/EXPERT2
Phone: +17853709027

References

1. Izacard, G., Lewis, P., Lomeli, M., Hosseini, L., Petroni, F., Schick, T., ... & Grave, E. (2022). Few-shot learning with retrieval augmented language models. arXiv preprint arXiv:2208.03299.
2. Lo, K., Wang, L. L., Neumann, M., Kinney, R., & Weld, D. S. (2019). S2ORC: The semantic scholar open research corpus. arXiv preprint arXiv:1911.02782.
3. Lester, B., Al-Rfou, R., & Constant, N. (2021). The power of scale for parameter-efficient prompt tuning. arXiv preprint arXiv:2104.08691.
4. Singh, A., D'Arcy, M., Cohan, A., Downey, D., & Feldman, S. (2022). SciRepEval: A Multi-Format Benchmark for Scientific Document Representations. arXiv preprint arXiv:2211.13308.
5. Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J. (2020). Measuring massive multitask language understanding. arXiv preprint arXiv:2009.03300.