Empowering Language Model with Guided Knowledge Fusion for Biomedical Document Re-ranking

Appendix

A Related Work

Document re-ranking is one of the fundamental tasks of information retrieval that aims to rerank the document with respect to a query. We categorized the existing literature as follows:

A.1 Open-domain document re-ranking

Neural re-ranking methods take the output of a first-stage retrieval system, such as BM25, and reorder the retrieved documents to create a more accurate comparison. Currently, most of the neural methods for re-ranking follow the encoder-only or encoder-decoder paradigm, where the interactions between the query and the documents are modeled using the attention mechanism of the encoder. In order to better capture interactions between query and document terms, few of the existing works (6; 20; 3) use cross-attention models, combining the query and candidate document into a single string before feeding it to the model. However, deploying these models to large collections of documents is computationally expensive (7). To minimize huge computational burden, (7) designed a new objective function, combining a query generation loss with a cross-entropy loss on a specific token, which can achieve comparable performance to cross-attention models. (24) introduced a benchmark setup for information retrieval and shows BM25 still remains a strong baseline method and demonstrates the capability of the re-ranking method on multiple datasets from diverse domains. In a similar line, (8) trained the network with query generation as an auxiliary task and obtained a significant performance improvement. A few other prominent works include (22; 16) that have improved the passage retrieval task using a zero-shot question generation approach.

A.2 Biomedical domain document retrieval and reranking

The BioASQ (Large-scale biomedical semantic indexing and question answering) shared task challenge has spurred research in biomedical document retrieval (25). However, most of the systems proposed for the biomedical document retrieval task have primarily relied upon term-matching algorithms (such as TF-IDF and BM25). Some of the recent systems have made progress by leveraging neural

re-ranking of retrieved candidates (21; 1). For example, the best system (2) at BioASQ 6 challenge used a variation of DRMM (17) to re-rank the top 100 documents retrieved using the BM25. (14) employed language models pretrained on PubMed articles. Specifically, they propose a hybrid model for the retrieval stage that combines BM25 and a dual encoder model. In the re-ranking phase, they use a cross-attention model with ranking loss. Recently, (15) proposed Poly-DPR, TempQG, a template-based question generation method, and two new pre-training tasks designed for biomedical documents to overcome the drawbacks of traditional neural retrieval models.

B Methodology

B.1 Graph Representation:

Given the query-document pair sub-graph $\mathcal{G}_{q,d} = (\mathcal{V}_{q,d}, \mathcal{E}_{q,d})$ with nodes $\{n_{int}, n_1, n_2, \ldots, n_M\}$, we first compute the node embeddings $U^1 = \{u^1_{int}, u^1_1, u^1_2, \ldots, u^1_M\} \in \mathcal{R}^{(M+1)\times d_g}$ using the pre-trained knowledge graph embeddings (*cf.* Section C.2). We utilized the graph attention network (27) to compute the node representation by propagating the information across the nodes in the subgraph $\mathcal{G}_{q,d}$. The subgraph node representation U^l at l^{th} layer of GNN is passed to the $(l+1)^{th}$ layer of GNN to encode and obtain the representation U^{l+1}. Following this, we extracted the representation from GNN for $l = 1, 2, \ldots, S$:

$$u^{l+1}_{int}, u^{l+1}_1, \ldots, u^{l+1}_M = \mathsf{GNN}(u^l_{int}, u^l_1, \ldots, u^l_M) \tag{1}$$

Formally, at a given layer, $l+1$ GNN computes the node representation $u_k \in \{u^{l+1}_{int}, u^{l+1}_1, \ldots, u^{l+1}_M\}$ through the exchange of messages among neighboring nodes in the graph.

$$u^{l+1}_k = f\Big(\sum_{u_n \in N_{u_k} \cup u_k} \beta_{nk} m_{nk}\Big) + u^l_k \tag{2}$$

where f denotes the two-layer feed-forward network, u_n is the neighboring node representation of the node u_k, N_{u_k} is the list of all neighbors of the node u_k. m_{nk} is the message which neighbor u_n passes to u_k, and β_{nk} denotes the attention weight for message m_{nk}, which signifies the contribution of each neighbor's message by its importance. To compute the β_{nk} and m_{nk}, we first compute the node-relation joint embedding j_{nk} of the node n and k and the relation between them. Particularly, we concatenate the node embeddings (e_n, e_k) and relation embedding e_{nk} and pass them from a 2-layer feed-forward neural network to obtain j_{nk}. Thereafter, we compute message m_{nk} as follows:

$$m_{nk} = W(U(u^l_k \oplus e_n \oplus j_{nk}) + b_1) + b_2 \tag{3}$$

To compute the importance weight β_{nk}, we first compute the query vector $a_n = \mathsf{FFN}(e_n \oplus u^l_n)$ and key vector $b^n_k = \mathsf{FFN}(e_k \oplus u^l_k \oplus j_{nk})$, thereafter, importance weight is computed as follows:

$$\beta_{nk} = \frac{\mathsf{exp}(a^T_n b^n_k / \sqrt{d})}{\sum_{u_m \in N_{u_k} \cup u_k} \mathsf{exp}(a^T_m b^m_k / \sqrt{d})} \tag{4}$$

where d is the dimension of the node representation.

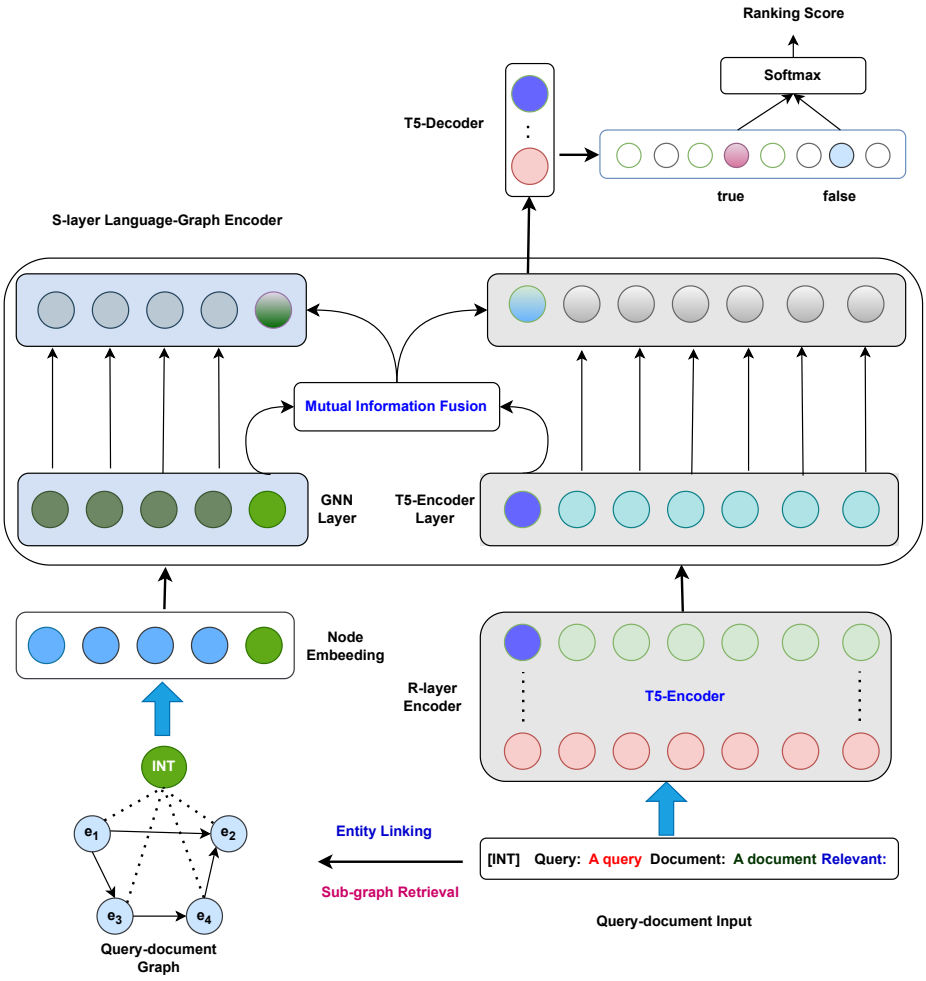

Fig. 1: Illustration of the proposed GraphMonoT5 model.

B.2 Language-graph Interaction:

On a given layer $l \in S$, we aim to effectively fuse the modalities by using the interaction token representation h^l_{int} and interaction node representation u^l_{int}. Towards this, first, we obtained the fused representation $x^l = f(h^l_{int} \oplus u^l_{int})$ with a two-layer feed-forward network f. The fused representation x^l may contain redundant information. To overcome this issue, we introduce mutual information

(MI) based feature fusion which aims to minimize the MI $\mathcal{I}(x^l; z^l)$ between the compressed encoded representation z^l and the concatenated representation x^l. Formally given two random variables x^1 and z, their MI is defined as follows:

$$\begin{aligned}\mathcal{I}(x; z) &= D_{KL}(p(x, z)||p(x)p(z)) \\ &= \int p(x, z) \log \frac{p(x, z)}{p(x)p(z)} dxdz \\ &= \int p(x, z) \log \frac{p(z|x)}{p(z)} dxdz \\ &= \int p(x, z) \log p(z|x) dxdz - \int p(z) \log p(z) dz\end{aligned} \tag{5}$$

We know that KL divergence follows the property that $D_{KL}(p(z)||q(z)) \geq 0$; where $q(z)$ is a variational approximation to the distribution $p(z)$, therefore, $\int p(z) \log p(z) dz \geq \int p(z) \log q(z) dz$. Following this, we can rewrite Eq. 5 as follows:

$$\begin{aligned}\mathcal{I}(x; z) &= \int p(x, z) \log p(z|x) dxdz - \int p(z) \log p(z) dz \\ &\leq \int p(x, z) \log p(z|x) dxdz - \int p(z) \log q(z) dz \\ &\leq \int p(x)p(z|x) \log \frac{p(z|x)}{q(z)} dxdz \\ &\leq \alpha \mathbb{E}_{z \sim p(z|x)}[D_{KL}(p(z|x)||q(z))] \\ &\leq \alpha M(x; z)\end{aligned} \tag{6}$$

With the assumption that both the prior $q(z) = \mathcal{N}(0, I)$ and the posterior approximation $p(z|x) = \mathcal{N}(\mu, \sigma^2)$ are Gaussian. The approximated value (10) of the mutual information $M(x; z)$ is calculated using the Monte Carlo sampling (23) as follows:

$$M(x; z) = -\frac{1}{2|B|} \sum_{i=1}^{|B|} \left(1 + 2 \log \sigma - \mu^2 - \sigma^2\right) \tag{7}$$

where $|B|$ denotes the sample size (batch size), μ and σ are the mean and standard deviation of the Gaussian representing $p(z|x)$. Following Eq. 6, the fused representation x^l and encoded representation z^l on a given layer l, their MI is defined as follows:

$$\begin{aligned}\mathcal{I}(x^l; z^l) &= D_{KL}(p(x^l, z^l)||p(x^l)p(z^l)) \\ &\leq \alpha \mathbb{E}_{z^l \sim p(z^l|x^l)}[D_{KL}(p(z^l|x^l)||q(z^l))] \\ &\leq \alpha M(x^l; z^l)\end{aligned} \tag{8}$$

where, α is a constant and D_{KL} denotes the KL divergence. We model the $p(z^l|x^l)$ using a parameterized Gaussian distribution $\mathcal{N}(\mu_z^l, \Sigma_z^l)$ with mean μ_z^l

[1] For brevity, we drop superscript l from x and z.

and variance Σ_z^l which are learned during the network training. To compute the gradients through random variables, we follow the reparametrization trick (10) with standard normal distribution $\epsilon \sim \mathcal{N}(0, I)$ to calculate $z^l = \mu_z^l + \Sigma_z^l \epsilon$. Later, we split z^l into the $\tilde{h}_{int}^l$ and $\tilde{u}_{int}^l$ for further computation of the token and node, respectively. With the virtue of Transformer network (26) and GNN, the fused representation is mixed with the remaining tokens and nodes of the subgraph.

C Experimental Setups

C.1 Entity Extraction and Linking

We followed the work of (30) to extract the entities and link them to the knowledge-graph nodes. Particularly for biomedical datasets, we used the scispacy (19) model (`en_core_sci_md`) to extract the entities and later used the scispacy entity linker (with default parameters, except threshold=0.9) to link the entity to the UMLS concept, which is also the node in the biomedical knowledge graph used in this study. The performance of the entity extraction was 83.92% F1 on the BC5CDR (11) dataset, and entity linking was 88% recall@100 on the MedMentions dataset (18). For the open-domain dataset, we used the spacy model (`en_core_web_sm`) to extract the entities and later used the concept recognizer (12) to link the entity to the ConceptNet knowledge graph. The performance of the entity extraction was 85.00% F1 on the OntoNotes5 (28) dataset.

C.2 Node Embedding

Following (31), we initialize the node embedding for the KG derived from UMLS and DrugBank using the pooled token representation of the node entity obtained from the SapBERT (13). To initialize the node embedding for the *ConceptNet* KG, we utilized the approach proposed by (5), which converts each KG triplets into sentences that passed to the BERT-Large (4) model to obtain the entity representation by applying the mean-pooling on entity mentions in the sentence.

C.3 Implementation & Training Details

We utilized the pre-trained T5-base model from HuggingFace[2] (29) to fine-tune it according to MonoT5 setup (20) where we consider the query and gold document as a positive question-document pair and randomly taken the two other document from corpus which are not part of the query's gold document to form the negative question-document pairs. We use Elasticsearch BM25 to report the lexical retrieval performance on all the datasets. In all our experiments, we re-rank the top 100 documents retrieved using BM25. For the BioASQ8B dataset, we use the $S = 3$ and $R = 9$, and the number of nodes in the subgraph is 10. For

[2] https://huggingface.co/t5-base

the HotPotQA dataset, we use the $S = 5$ and $R = 7$, and the number of nodes in the subgraph is 15. For both datasets, we find the optimal value of GNN hidden state representation size=200, the value of $\alpha = 0.01$, and the projection dimension of the feed-forward network is 100. The MonoT5 model is trained with batch size 16, and GraphMonoT5 is trained with batch size 8. We fine-tuned each model for 3 epochs on BioASQ8B and HotPotQA datasets. The maximum token length of the concatenated query and document is set to 512 for all the experiments. The model parameters are updated using Adam (9) optimization algorithm with the learning rate of $3e-4$ in all the experiments. We obtained the value of the optimal hyperparameters based on the respective development dataset performance in terms of the nDCG@10 score. We also computed the train/test(reranking) times on respective datasets. For the BioASQ8B and HotPotQA datasets, we reported the per epoch training time as 5.5 hrs and 11 hrs, respectively. Please note that the test dataset of TREC-COVID is evaluated in a zero-shot setting where we use the model trained on the BioASQ dataset to evaluate the performance. For the HotPotQA and HotPotQA datasets, we computed the per query reranking time as 840ms, 950ms, and 280ms, respectively, with 100 retrieved documents from the retrieval stage. We performed all the experiments on a single NVIDIA Tesla V100x GPU having 32GB memory.

D Analysis

We have provided the query-documents examples comparing MonoT5 and GraphMonoT5 in Table 2.

Table 1: Sample queries and gold document from the BioASQ dataset along with the top retrieved documents using BM25 and MonoT5 methods. Lexical and semantic matches considering context are shown in blue and pink, respectively. The highlighted texts in green represent the requirements of domain knowledge to retrieve the correct document.

Query	Top Retrieved Document (BM25)	Top Retrieved Document (MonoT5)	Gold Document
Ex1: Which algorithms have been developed for analysing CRISPR/Cas9 knockout screens data?	CRISPR/Cas9, an RNA guided endonuclease system is the most recent technology for this work. Here, we have discussed the major considerations involved in designing a CRISPR/Cas9 based screening experiment for identification of synthetic lethal targets.	We describe here in detail a simple, rapid, and scalable method for CRISPR-Cas9-mediated gene knockout and tagging in Leishmania. This method details how to use simple PCR to generate (1) templates for single guide RNA (sgRNA) transcription in cells expressing Cas9 and T7 RNA polymerase..	We propose the Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK) method for prioritizing single-guide RNAs, genes and pathways in genome-scale CRISPR/Cas9 knockout screens.
Ex2: What rare disease is associated with a mutation in the GPC6 gene on chromosome 13?	..We report the construction of a high-resolution 4 Mb sequence-ready BAC/PAC contig of the GPC5/GPC6 gene cluster on chromosome region 13q32.	The human gamma-sarcoglycan gene was mapped to chromosome 13q12, and deletions that alter its reading frame were identified in three families and one of four sporadic cases of SCARMD.	.. The proband had normal molecular analysis of the glypican 6 gene (GPC6), which was recently reported as a candidate for autosomal recessive omodysplasia. Mild rhizomelic shortening of the lower extremities has not been previously reported...

Table 2: Sample queries and top retrieved documents using MonoT5 and proposed GraphMonoT5 methods. For each query, the proposed model retrieved the relevant document while MonoT5 did not.

	Query	**Top Retrieved Document (MonoT5)**	**Top Retrieved Document (GraphMonoT5)**
(1)	What gene is mutated in Huntington's Disease patients?	We used PET scans with the tracers [18F]fluorodeoxyglucose (FDG) and [11C]raclopride (RACLO) to study glucose metabolism and dopamine D2 receptor binding in the caudate nucleus and putamen of 18 carriers of the Huntington's disease gene mutation (10 asymptomatic subjects and eight untreated symptomatic Huntington's disease patients in an early disease stage)...	The autosomal dominant spinocerebellar ataxias, commonly referred to as SCAs, are clinically and genetically heterogeneous neurodegenerative disorders. ... In some cases the clinical phenotype of SCA17 overlaps that of Huntington's disease (HD), hence the use of the term Huntington's disease-like. We screened 89 patients with a Huntington's disease-like phenotype without the HD-gene mutation and 178 patients with genetically unclassified cerebellar ataxia for the mutation in TBP...
(2)	What are manifestations of the Saint's Triad?	... Gastrointestinal malignancies are sometimes associated with paraneoplastic entities, isolated or manifested as syndromes, but neither Saint's triad or Heyde syndrome have been included. This patient persisted clinically stable during the preoperative period, but suddenly died;...	Yamanaka et al. described two case studies involving coexistent cholelithiasis, hiatal hernia, and umbilical hernias, and discussed clinical similarities with the classical features of the Saint's triad...
(3)	What are 3 symptoms of Waardenburg Syndrome?	Shah-Waardenburg syndrome is a rare congenital disorder with variable clinical expression, characterised by aganglionosis of the rectosigmo?d (Hirschsprung disease), and abnormal melanocyte migration, resulting in pigmentary abnormalities and sensorineural deafness (Waardenburg syndrome). Mutations in the EDN, EDNRB and SOX10 genes can be found in patients with this syndrome. SOX10 mutations are specifically associated with a more severe phenotype called PCWH: peripheral demyelinating neuropathy, central dysmyelinating leukodystrophy, Waardenburg syndrome, and Hirschsprung disease. ...	Waardenburg syndrome type 1 (WS1) is a rare autosomal dominant genetic disorder of neural crest cells (NCC) characterized by congenital sensorineural hearing loss, dystopia canthorum, and abnormal iris pigmentation. WS1 is due to loss-of-function mutations in paired box gene 3 (PAX3). Here, we identified a novel PAX3 mutation (c.808C¿G, p.R270G) in a three-generation Chinese family with WS1, and then analyzed its in vitro activities...

Bibliography

[1] Almeida, T., Matos, S.: Bit. ua at bioasq 8: Lightweight neural document ranking with zero-shot snippet retrieval. In: CLEF (Working Notes) (2020)

[2] Brokos, G., Liosis, P., McDonald, R., Pappas, D., Androutsopoulos, I.: Aueb at bioasq 6: Document and snippet retrieval. In: Proceedings of the 6th BioASQ Workshop A challenge on large-scale biomedical semantic indexing and question answering. pp. 30–39 (2018)

[3] Chen, X., Hui, K., He, B., Han, X., Sun, L., Ye, Z.: Co-bert: A context-aware bert retrieval model incorporating local and query-specific context. arXiv preprint arXiv:2104.08523 (2021)

[4] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: 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). pp. 4171–4186 (2019)

[5] Feng, Y., Chen, X., Lin, B.Y., Wang, P., Yan, J., Ren, X.: Scalable multi-hop relational reasoning for knowledge-aware question answering. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1295–1309 (2020)

[6] Han, S., Wang, X., Bendersky, M., Najork, M.: Learning-to-rank with bert in tf-ranking. arXiv preprint arXiv:2004.08476 (2020)

[7] Hui, K., Zhuang, H., Chen, T., Qin, Z., Lu, J., Bahri, D., Ma, J., Gupta, J., dos Santos, C., Tay, Y., et al.: Ed2lm: Encoder-decoder to language model for faster document re-ranking inference. In: Findings of the Association for Computational Linguistics: ACL 2022. pp. 3747–3758 (2022)

[8] Ju, J.H., Yang, J.H., Wang, C.J.: Text-to-text multi-view learning for passage re-ranking. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1803–1807 (2021)

[9] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015), http://arxiv.org/abs/1412.6980

[10] Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: Bengio, Y., LeCun, Y. (eds.) 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings (2014), http://arxiv.org/abs/1312.6114

[11] Li, J., Sun, Y., Johnson, R.J., Sciaky, D., Wei, C.H., Leaman, R., Davis, A.P., Mattingly, C.J., Wiegers, T.C., Lu, Z.: Biocreative v cdr task corpus: a resource for chemical disease relation extraction. Database **2016** (2016)

[12] Lin, B.Y., Chen, X., Chen, J., Ren, X.: Kagnet: Knowledge-aware graph networks for commonsense reasoning. 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). pp. 2829–2839 (2019)
[13] Liu, F., Shareghi, E., Meng, Z., Basaldella, M., Collier, N.: Self-alignment pretraining for biomedical entity representations. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 4228–4238 (2021)
[14] Lu, J., Ma, J., Hall, K.: Zero-shot hybrid retrieval and reranking models for biomedical literature. In: CLEF (Working Notes) (2022)
[15] Luo, M., Mitra, A., Gokhale, T., Baral, C.: Improving biomedical information retrieval with neural retrievers. Proceedings of the AAAI Conference on Artificial Intelligence **36**(10), 11038–11046 (Jun 2022). https://doi.org/10.1609/aaai.v36i10.21352, https://ojs.aaai.org/index.php/AAAI/article/view/21352
[16] Ma, J., Korotkov, I., Yang, Y., Hall, K., McDonald, R.: Zero-shot neural passage retrieval via domain-targeted synthetic question generation. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. pp. 1075–1088 (2021)
[17] McDonald, R., Brokos, G., Androutsopoulos, I.: Deep relevance ranking using enhanced document-query interactions. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. pp. 1849–1860 (2018)
[18] Murty, S., Verga, P., Vilnis, L., Radovanovic, I., McCallum, A.: Hierarchical losses and new resources for fine-grained entity typing and linking. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 97–109 (2018)
[19] Neumann, M., King, D., Beltagy, I., Ammar, W.: Scispacy: Fast and robust models for biomedical natural language processing. In: Proceedings of the 18th BioNLP Workshop and Shared Task. pp. 319–327 (2019)
[20] Nogueira, R., Jiang, Z., Pradeep, R., Lin, J.: Document ranking with a pretrained sequence-to-sequence model. In: Findings of the Association for Computational Linguistics: EMNLP 2020. pp. 708–718 (2020)
[21] Pappas, D., Stavropoulos, P., Androutsopoulos, I.: Aueb-nlp at bioasq 8: Biomedical document and snippet retrieval. In: CLEF (Working Notes) (2020)
[22] Sachan, D.S., Lewis, M., Joshi, M., Aghajanyan, A., Yih, W.t., Pineau, J., Zettlemoyer, L.: Improving passage retrieval with zero-shot question generation. arXiv preprint arXiv:2204.07496 (2022)
[23] Shapiro, A.: Monte carlo sampling methods. Handbooks in operations research and management science **10**, 353–425 (2003)
[24] Thakur, N., Reimers, N., Rücklé, A., Srivastava, A., Gurevych, I.: Beir: A heterogeneous benchmark for zero-shot evaluation of information retrieval models. In: Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2) (2021)
[25] Tsatsaronis, G., Balikas, G., Malakasiotis, P., Partalas, I., Zschunke, M., Alvers, M.R., Weissenborn, D., Krithara, A., Petridis, S., Polychronopoulos, D., et al.: An overview of the bioasq large-scale biomedical semantic

indexing and question answering competition. BMC bioinformatics **16**(1), 1–28 (2015)

[26] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)

[27] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: International Conference on Learning Representations (2018)

[28] Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Pradhan, S., Ramshaw, L., Xue, N., Taylor, A., Kaufman, J., Franchini, M., et al.: Ontonotes release 5.0 ldc2013t19. Linguistic Data Consortium, Philadelphia, PA **23**, 170 (2013)

[29] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., et al.: Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations. pp. 38–45 (2020)

[30] Yasunaga, M., Ren, H., Bosselut, A., Liang, P., Leskovec, J.: Qa-gnn: Reasoning with language models and knowledge graphs for question answering. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 535–546 (2021)

[31] Zhang, X., Bosselut, A., Yasunaga, M., Ren, H., Liang, P., Manning, C.D., Leskovec, J.: Greaselm: Graph reasoning enhanced language models. In: International Conference on Learning Representations (2021)