

Improving BERT-based Query-by-Document Retrieval with Multi-Task Optimization

Amin Abolghasemi¹, Suzan Verberne¹, and Leif Azzopardi²

¹ Leiden University, The Netherlands

{m.a.abolghasemi,s.verberne}@liacs.leidenuniv.nl

² University of Strathclyde, UK

leif.azzopardi@strath.ac.uk

Abstract. Query-by-document (QBD) retrieval is an Information Retrieval task in which a seed document acts as the query and the goal is to retrieve related documents – it is particular common in professional search tasks. In this work we improve the retrieval effectiveness of the BERT re-ranker, proposing an extension to its fine-tuning step to better exploit the context of queries. To this end, we use an additional document-level representation learning objective besides the ranking objective when fine-tuning the BERT re-ranker. Our experiments on two QBD retrieval benchmarks show that the proposed multi-task optimization significantly improves the ranking effectiveness without changing the BERT re-ranker or using additional training samples. In future work, the generalizability of our approach to other retrieval tasks should be further investigated.

Keywords: Query-by-document retrieval · BERT-based ranking · Multi-task optimization.

1 Introduction

Query by document (QBD) [38,37], is a widely-used practice across professional, domain-specific retrieval tasks [33,35] such as scientific literature retrieval [25,9], legal case law retrieval [2,3,30,34], and patent prior art retrieval [13,28]. In these tasks, the user’s information need is based on a seed document of the same type as the documents in the collection. Taking a document as query results in long queries, which can potentially express more complex information needs and provide more context for ranking models [16]. Transformer-based ranking models have proven to be highly effective at taking advantage of context [10,11,24], but the long query documents pose challenges because of the maximum input length for BERT-based ranking models. Recent work showed that transformer-based models which handle longer input sequences are not necessarily more effective when being used in retrieval tasks on long texts [3]. We, therefore, direct our research towards improving retrieval effectiveness while acting within the input length limitation of ranking models based on large scale pre-trained BERT models [12]. We posit that the representations learned during pre-training

have been tailored toward smaller sequences of text – and additional tuning the language models to better represent the documents in this specific domain and query setting, could lead to improvements in ranking.

To investigate this, we first focus on the task of Case Law Retrieval (i.e. given a legal case find the related cases), and employ multi-task optimisation to improve the standard BERT-based cross-encoder ranking model [15] for QBD retrieval. We then explore the generalizability of our approach by evaluating our approach on four QBD retrieval tasks in the academic domain. Our approach draws upon multi-task learning to rank – where a shared structure across auxiliary, related tasks is used [1,20,29,8]. Specifically, in our method, we employ document-level representation learning as an auxiliary objective for multi-task fine-tuning (MTFT) of the BERT re-ranker. To our knowledge, there is no prior work on using representation learning directly as an auxiliary task for fine-tuning a BERT re-ranker. We show that optimizing the re-ranker jointly with document-level representation learning leads to consistently higher ranking effectiveness over the state-of-the-art with greater efficiency i.e., with the same training instances on the same architecture.

2 Preliminaries

BERT-based Ranking. Pre-trained transformer-based language models [12] have shown significant improvement in ranking tasks [10,11,24,18]. In this work, we use the BERT re-ranker proposed by Nogueira and Cho [26], which is a pre-trained BERT model followed by a projection layer W_p on top of its $[CLS]$ token final hidden states. The BERT re-ranker, which is a cross-encoder neural ranking model, uses the concatenation of a query and candidate document as the input to a fine-tuned pre-trained BERT model. The output of the model is used to indicate the relevance score s of the document d for the input query q , such that:

$$s(q, d) = BERT([CLS] q [SEP] d [SEP])_{[CLS]} * W_p \quad (1)$$

BERT-based Representation Learning. BERT was originally pre-trained on two tasks, namely Masked Language Modeling and Next Sentence Prediction [12]. These tasks, however, are not meant to optimize the network for document-level information representation [9] which may make the model less effective in representation-focused [14] downstream tasks [10]. Previous works have shown that leveraging a Siamese or triplet network structure for fine-tuning BERT could optimize the model for document-level representation [9]. Following Devlin et al [12], we use the final hidden state corresponding to the $[CLS]$ token to encode the query q and the document d into their representations r_q , and r_d :

$$r_q = BERT([CLS] q [SEP])_{[CLS]} \quad r_d = BERT([CLS] d [SEP])_{[CLS]} \quad (2)$$

Pairwise Ranking Loss. In Learning-To-Rank tasks, pairwise loss minimizes the average number of pairwise errors in a ranked list [5,6]. Here, we aim to optimize

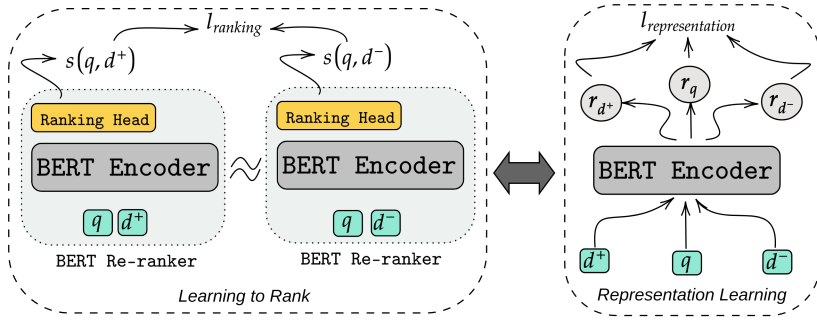


Fig. 1. The fine-tuning process. The same training triples (q, d^+, d^-) are used in each step. The BERT re-rankers are the same, and the BERT encoder is shared between the ranking and representation learning tasks.

the BERT re-ranker with a pairwise cross-entropy softmax loss function [5]:

$$l_{rank} = -\log \frac{e^{score(q, d^+)}}{e^{score(q, d^+)} + e^{score(q, d^-)}} \quad (3)$$

where the *score* function represents the degree of relevance between a query and a document computed as described in Eq. 1. In fact, this pairwise loss frames the ranking task as a binary classification problem in which, given a query (q) and a pair of relevant (d^+) and non-relevant (d^-) documents, the fine-tuned ranking model predicts the relevant one. However, at inference time the model is used as a point-wise score function.

Triplet Representation Learning Loss. In the context of representation learning with pre-trained transformers, a triplet loss function fine-tunes the weights of the model such that given an anchor query q , the representations of the query r_q and the document r_d (obtained as described in Eq. 2) are closer for a relevant document d^+ than for a non-relevant document d^- :

$$l_{representation} = \max\{ (f(r_q, r_{d^+}) - f(r_q, r_{d^-}) + margin), 0 \} \quad (4)$$

Here, f indicates a distance metric and $margin$ ensures that d^+ is at least $margin$ closer to q than d^- [31].

3 Multi-task fine-tuning of the BERT re-ranker

Our proposed re-ranker aims to jointly optimise both the l_{rank} and $l_{representation}$ – we shall refer to our BERT re-ranker model as MTFB-BERT. As shown in Figure 1, the Multi Task Fine Tuning (MTFT) is achieved by providing training instances consisting of triples (q, d^+, d^-) . To do so, we first feed the concatenation of q and d^+ , and the concatenation of q and d^- separately to the MTFB-BERT

re-ranker, as described in section 2, to compute the pairwise loss l_{rank} following Eq. 3. In the next step, we feed each of q , d^+ , and d^- separately to the shared encoder of the re-ranker to compute the $l_{representation}$ following Eq. 4. As distance metric f we use the L2-norm and we set $margin = 1$ in our experiments. The shared encoder is then fine-tuned with the aggregated loss as shown in Eq. 5 while the ranking head is only fine-tuned by the first term:

$$l_{aggregated} = l_{rank} + \lambda l_{representation} \quad (5)$$

The λ parameter balances the weight between the two loss functions. Later, we investigate the stability of our model under different values of λ . Since ranking is the target task, and the ranking head is only optimized by the ranking loss, we assign the regularization weight ($0 < \lambda < 1$) only to the representation loss. It is noteworthy that at inference time, we only use the ranking head of the MTFT-BERT re-ranker.

4 Experimental Setup

Datasets. We first evaluate our proposed method on legal case retrieval. The goal of case law retrieval is to retrieve the relevant prior law cases which could act as supporting cases for a given query law case. This professional search task is a query-by-document (QBD) retrieval task, where both the query and the documents are case law documents. We use the test collection for the case law retrieval task of COLIEE 2021 [30]. This collection contains a corpus with 4415 legal cases with a training and a test set consisting of 650 and 250 query cases respectively. In addition, to evaluate the generalizability of our approach, we use another domain-specific QBD retrieval benchmark, called SciDocs [9]. SciDocs was originally introduced as a representation learning benchmark in the scientific domain while framing the tasks as ranking; we use the four SciDocs tasks: {citation, co-citation, co-view, and co-read}-prediction to evaluate our method. It is worth mentioning that while the original paper trains the model on a citation graph of academic papers, we take the validation set provided for each task and use 85% of it as training set and the rest as the validation set for tuning purposes.

Implementation. We use Elasticsearch¹ to index and retrieve the initial ranking list using a BM25 ranker. It was shown in prior work that BM25 is a strong baseline [32], and it even holds the state-of-the-art in case law retrieval on COLIEE 2021 [3]. Therefore, to make our work comparable, we use the configuration provided by [3] to optimize the BM25 with Elasticsearch for COLIEE 2021 case law retrieval. For query generation, following the effectiveness of term selection using Kullback-Leibler divergence for Informativeness (KLI) in prior work in case law retrieval [21,3], we use the top-10% of a query document terms scored with KLI² as the query for BM25 in our experiments. As the BERT encoders,

¹ <https://github.com/elastic/elasticsearch>

² Implementation from <https://github.com/suzanv/termprofiling/>

Table 1. The reranking results with BM25 and BM25_{optimized} as initial rankers for the COLIEE 2021 test data. † indicates the statistically significant improvements over BM25_{optimized} according to a paired t-test ($p < 0.05$). TLIR achieved the highest score in the COLIEE 2021 competition.

Model	Initial Ranker	Precision%	Recall%	F1%
BM25	-	8.8	16.51	11.48
TLIR [23]	-	15.33	25.56	19.17
BM25 _{optimized} [3]	-	17.00	25.36	20.35
BERT	BM25	10.48	18.80	13.46
MTFT-BERT	BM25	12.08	21.59	15.49
BERT	BM25 _{optimized}	14.40	24.63	18.17
MTFT-BERT	BM25 _{optimized}	17.44 [†]	29.99 [†]	22.05 [†]

Table 2. Ranking results on the SciDocs benchmark. HF is Huggingface. † indicates the statistically significant improvements according to a paired t-test ($p < 0.05$).

Model	Co-view		Co-read		Cite		Co-cite	
	MAP	nDCG	MAP	nDCG	MAP	nDCG	MAP	nDCG
SPECTER [9]	83.6%	0.915	84.5%	0.924	88.3%	0.949	88.1%	0.948
SPECTER w/ HF [36]	83.4%	0.914	85.1%	0.927	92.0%	0.966	88.0%	0.947
BM25	75.4%	0.874	75.6%	0.881	73.5%	0.876	76.3%	0.890
BM25 _{optimized}	76.26%	0.877	76.09%	0.881	75.3%	0.884	77.41%	0.896
BERT	85.2%	0.925	87.5%	0.940	94.0%	0.975	89.7%	0.955
MTFT-BERT	86.2% [†]	0.930 [†]	87.7%	0.940	94.2%	0.976	91.0% [†]	0.961 [†]

we use LegalBERT [7], and SciBERT [4], which are domain-specific BERT models pre-trained on the legal and scientific domains respectively. We train our neural ranking models for 15 epochs with a batch size of 32, and AdamW optimizer [22] with a learning rate of 3×10^{-5} . All of our models are implemented and fine-tuned using PyTorch [27] and the HuggingFace library [36].

5 Results and Analysis

Ranking quality. Table 1 displays the ranking quality of the MTFT-BERT re-ranker in comparison to BM25, TLIR [23], BM25_{optimized}, and the original BERT re-ranker on COLIEE 2021. The cut-off k for all rankers is set to 5 during both validation and test since the train queries in COLIEE 2021 have 5 relevant documents on average. We report precision and recall besides F1, which is the official metric used in the COLIEE competition. It can be seen that the BERT re-ranker and the MTFT-BERT re-ranker can both achieve better quality over BM25 with default parameters as initial ranker. In contrast, when we use BM25_{optimized} as the initial ranker, the BERT re-ranker fails to yield improvement, while MTFT-BERT outperforms the state-of-the-art BM25_{optimized} [3] by a statistically significant margin of 8.3% relative improvement.

For comparability reasons on the SciDocs benchmark, we have included both the original paper results and the results reported in their official code repository³, which is achieved using Huggingface models like our implementations. As

³ <https://github.com/allenai/specter>

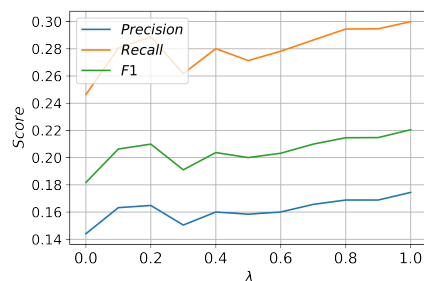


Fig. 2. The evaluation results of MTFT-BERT+BM25_{optimized} with various λ in the COLIEE 2021 case law retrieval task. $\lambda = 0$ indicates the BERT re-ranker.

Table 2 shows, while both the BERT re-ranker, and the MTFT-BERT re-ranker yield improvement over the SPECTER method, the MTFT-BERT re-ranker outperforms the BERT re-ranker which confirms the effectiveness of our method in an additional domain-specific QBD retrieval setting.

Robustness to varying λ . Task weighting is a widely used method in multi-task learning algorithms [19,17] where a static or dynamic weight is assigned to the loss of different tasks. Figure 2 displays the ranking quality of the MTFT-BERT re-ranker over different values of λ on the COLIEE test set, using BM25_{optimized} as the initial ranker. We bound λ at 1 since our target task is ranking, and we do not want the representation loss rate to have higher impact in the training. We can see that our model quality is relatively consistent across different values above 0.5 which indicates the robustness of our model in tuning this parameter.

Effect of re-ranking depth. We experimented with the ranking depth, i.e., number of documents re-ranked from the initial ranker result, by increasing it from 15 to 100 in steps of 5. We then analyzed the MTFT-BERT re-ranking quality relative to depth. We found that the ranking quality decreases rapidly after the lower ranking depths, to $F1 = 17.3$ at 100, which is lower than the original BM25_{optimized} ranking. While MTFT-BERT can improve over BM25 with a shallow re-ranking set, we confirm the findings by previous studies that BM25 is a strong baseline for case law retrieval [3,32].

6 Conclusion

This paper shows that it is possible to improve the BERT cross-encoder re-ranker quality using multi-task optimization with an auxiliary representation learning task. We showed that the resulting model named MTFT-BERT re-ranker obtains consistently better retrieval quality than the original BERT re-ranker using the same training instances and structure. While our focus was on query-by-document retrieval in professional search domains (legal and academic), as a future work, it would be interesting to study the effectiveness of MTFT-BERT re-ranker in other retrieval tasks where we have shorter queries.

7 Acknowledgments

This work is funded by the DoSSIER project under European Union’s Horizon 2020 research and innovation program, Marie Skłodowska-Curie grant agreement No. 860721.

References

1. Ahmad, W.U., Chang, K.W., Wang, H.: Multi-task learning for document ranking and query suggestion. In: International Conference on Learning Representations (2018)
2. Althammer, S., Hofstätter, S., Sertkan, M., Verberne, S., Hanbury, A.: Paragraph aggregation retrieval model (parm) for dense document-to-document retrieval. In: Advances in Information Retrieval, 44rd European Conference on IR Research, ECIR 2022 (2022)
3. Askari, A., Verberne, S.: Combining lexical and neural retrieval with longformer-based summarization for effective case law retrieval. In: DESIRES (2021)
4. Beltagy, I., Lo, K., Cohan, A.: SciBERT: A pretrained language model for scientific text. 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. 3615–3620. Association for Computational Linguistics, Hong Kong, China (Nov 2019). <https://doi.org/10.18653/v1/D19-1371>, <https://aclanthology.org/D19-1371>
5. Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., Hullender, G.: Learning to rank using gradient descent. In: Proceedings of the 22nd international conference on Machine learning - ICML ’05. pp. 89–96. ACM Press, Bonn, Germany (2005). <https://doi.org/10.1145/1102351.1102363>, <http://portal.acm.org/citation.cfm?doid=1102351.1102363>
6. Cao, Z., Qin, T., Liu, T.Y., Tsai, M.F., Li, H.: Learning to rank: from pairwise approach to listwise approach. In: Proceedings of the 24th international conference on Machine learning. pp. 129–136 (2007)
7. Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., Androutopoulos, I.: LEGAL-BERT: The muppets straight out of law school. In: Findings of the Association for Computational Linguistics: EMNLP 2020. pp. 2898–2904. Association for Computational Linguistics, Online (Nov 2020). <https://doi.org/10.18653/v1/2020.findings-emnlp.261>, <https://aclanthology.org/2020.findings-emnlp.261>
8. Cheng, Q., Ren, Z., Lin, Y., Ren, P., Chen, Z., Liu, X., de Rijke, M.: Long short-term session search: Joint personalized reranking and next query prediction. In: Proceedings of the Web Conference 2021. pp. 239–248 (2021)
9. Cohan, A., Feldman, S., Beltagy, I., Downey, D., Weld, D.: SPECTER: Document-level Representation Learning using Citation-informed Transformers. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 2270–2282. Association for Computational Linguistics, Online (2020). <https://doi.org/10.18653/v1/2020.acl-main.207>, <https://www.aclweb.org/anthology/2020.acl-main.207>
10. Dai, Z., Callan, J.: Deeper text understanding for ir with contextual neural language modeling. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 985–988 (2019)

11. Dai, Z., Callan, J.: Context-aware term weighting for first stage passage retrieval. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1533–1536 (2020)
12. 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. Association for Computational Linguistics, Minneapolis, Minnesota (Jun 2019). <https://doi.org/10.18653/v1/N19-1423>, <https://aclanthology.org/N19-1423>
13. Fujii, A., Iwayama, M., Kando, N.: Overview of the patent retrieval task at the ntcir-6 workshop. In: NTCIR (2007)
14. Guo, J., Fan, Y., Ai, Q., Croft, W.B.: A deep relevance matching model for ad-hoc retrieval. In: Proceedings of the 25th ACM international on conference on information and knowledge management. pp. 55–64 (2016)
15. Humeau, S., Shuster, K., Lachaux, M.A., Weston, J.: Poly-encoders: Architectures and pre-training strategies for fast and accurate multi-sentence scoring. In: International Conference on Learning Representations (2020), <https://openreview.net/forum?id=SkxgmnNFvH>
16. Huston, S., Croft, W.B.: Evaluating verbose query processing techniques. In: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. pp. 291–298 (2010)
17. Kongyoung, S., Macdonald, C., Ounis, I.: Multi-task learning using dynamic task weighting for conversational question answering. In: Proceedings of the 5th International Workshop on Search-Oriented Conversational AI (SCAI). pp. 17–26 (2020)
18. Lin, J., Nogueira, R., Yates, A.: Pretrained transformers for text ranking: Bert and beyond (2021)
19. Liu, S., Liang, Y., Gitter, A.: Loss-balanced task weighting to reduce negative transfer in multi-task learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 9977–9978 (2019)
20. Liu, X., He, P., Chen, W., Gao, J.: Multi-task deep neural networks for natural language understanding. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 4487–4496. Association for Computational Linguistics, Florence, Italy (Jul 2019). <https://doi.org/10.18653/v1/P19-1441>, <https://aclanthology.org/P19-1441>
21. Locke, D., Zuccon, G., Scells, H.: Automatic query generation from legal texts for case law retrieval. In: Asia Information Retrieval Symposium. pp. 181–193. Springer (2017)
22. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019), <https://openreview.net/forum?id=Bkg6RiCqY7>
23. Ma, Y., Shao, Y., Liu, B., Liu, Y., Zhang, M., Ma, S.: Retrieving legal cases from a large-scale candidate corpus. In: Proceedings of the eighth International Competition on Legal Information Extraction/Entailment , COLIEE2021 (2021)
24. MacAvaney, S., Yates, A., Cohan, A., Goharian, N.: Cedr: Contextualized embeddings for document ranking. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1101–1104 (2019)

25. Mysore, S., O’Gorman, T., McCallum, A., Zamani, H.: Csfcube—a test collection of computer science research articles for faceted query by example. arXiv preprint arXiv:2103.12906 (2021)
26. Nogueira, R., Cho, K.: Passage re-ranking with bert. arXiv preprint arXiv:1901.04085 (2019)
27. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d’Alché-Buc, F., Fox, E., Garnett, R. (eds.) *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc. (2019), <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
28. Piroi, F., Hanbury, A.: Multilingual patent text retrieval evaluation: Clef-ip. In: *Information Retrieval Evaluation in a Changing World*, pp. 365–387. Springer (2019)
29. Qu, C., Yang, L., Chen, C., Qiu, M., Croft, W.B., Iyyer, M.: Open-retrieval conversational question answering. In: *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*. pp. 539–548 (2020)
30. Rabelo, J., Kim, M.Y., Goebel, R., Yoshioka, M., Kano, Y., Satoh, K.: COLIEE 2020: Methods for Legal Document Retrieval and Entailment. In: Okazaki, N., Yada, K., Satoh, K., Mineshima, K. (eds.) *New Frontiers in Artificial Intelligence*, vol. 12758, pp. 196–210. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-79942-7_13, https://link.springer.com/10.1007/978-3-030-79942-7_13, series Title: *Lecture Notes in Computer Science*
31. Reimers, N., Gurevych, I.: Sentence-BERT: Sentence embeddings using Siamese BERT-networks. 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. 3982–3992. Association for Computational Linguistics, Hong Kong, China (Nov 2019). <https://doi.org/10.18653/v1/D19-1410>, <https://aclanthology.org/D19-1410>
32. Rosa, G.M., Rodrigues, R.C., Lotufo, R., Nogueira, R.: Yes, bm25 is a strong baseline for legal case retrieval. arXiv preprint arXiv:2105.05686 (2021)
33. Russell-Rose, T., Chamberlain, J., Azzopardi, L.: Information retrieval in the workplace: A comparison of professional search practices. *Information Processing & Management* **54**(6), 1042–1057 (2018)
34. Shao, Y., Mao, J., Liu, Y., Ma, W., Satoh, K., Zhang, M., Ma, S.: Bert-pli: Modeling paragraph-level interactions for legal case retrieval. In: Bessiere, C. (ed.) *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*. pp. 3501–3507. International Joint Conferences on Artificial Intelligence Organization (7 2020). <https://doi.org/10.24963/ijcai.2020/484>, <https://doi.org/10.24963/ijcai.2020/484>, main track
35. Verberne, S., He, J., Kruschwitz, U., Wiggers, G., Larsen, B., Russell-Rose, T., de Vries, A.P.: First international workshop on professional search. In: *ACM SIGIR Forum*. vol. 52, pp. 153–162. ACM New York, NY, USA (2019)
36. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T.L., Gugger, S., Drame, M., Lhoest, Q., Rush, A.M.: 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. Association for Computational Linguistics, Online (Oct 2020), <https://www.aclweb.org/anthology/2020.emnlp-demos.6>
37. Yang, E., Lewis, D.D., Frieder, O., Grossman, D.A., Yurchak, R.: Retrieval and richness when querying by document. In: DESIRES. pp. 68–75 (2018)
 38. Yang, Y., Bansal, N., Dakka, W., Ipeirotis, P., Koudas, N., Papadias, D.: Query by document. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining. pp. 34–43 (2009)