



**HAL**  
open science

## Passage Retrieval on Structured Documents using Graph Attention Networks

Lucas Albarede, Philippe Mulhem, Lorraine Goeuriot, Claude Le  
Pape-Gardeux, Sylvain Marié, Trinidad Chardin-Segui

► **To cite this version:**

Lucas Albarede, Philippe Mulhem, Lorraine Goeuriot, Claude Le Pape-Gardeux, Sylvain Marié, et al.. Passage Retrieval on Structured Documents using Graph Attention Networks. 44th European Conference on IR Research (ECIR 2022), Apr 2022, Stavanger, Norway. pp.13-21, 10.1007/978-3-030-99739-7\_2. hal-03626054

**HAL Id: hal-03626054**

**<https://hal.science/hal-03626054>**

Submitted on 22 Jul 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Passage Retrieval on Structured Documents using Graph Attention Networks

Lucas Albarede<sup>\*,\*\*</sup> Philippe Mulhem<sup>\*</sup> Lorraine Goeuriot<sup>\*</sup>  
Claude Le Pape-Gardeux<sup>\*\*</sup> Sylvain Marie<sup>\*\*</sup> Trinidad Chardin-Segui<sup>\*\*</sup>

<sup>\*</sup> Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France, Email:  
firstname.lastname@imag.fr

<sup>\*\*</sup> Schneider Electric Industries SAS, Email: firstname.lastname@se.com

**Abstract.** Passage Retrieval systems aim at retrieving and ranking small text units according to their estimated relevance to a query. A usual practice is to consider the context a passage appears in (its containing document, neighbour passages, etc.) to improve its relevance estimation. In this work, we study the use of Graph Attention Networks (GATs), a graph node embedding method, to perform passage contextualization. More precisely, we first propose a document graph representation based on several *inter*- and *intra*-document relations. Then, we investigate two ways of leveraging the use of GATs on this representation in order to incorporate contextual information for passage retrieval. We evaluate our approach on a Passage Retrieval task for structured documents: CLEF-IP2013. Our results show that our document graph representation coupled with the expressive power of GATs allows for a better context representation leading to improved performances.

**Keywords:** Passage Retrieval · Graph Attention Networks · Experiments · Document Representation.

## 1 Introduction

Passage Retrieval is a long lasting topic for Information Retrieval (IR) that is concerned with the retrieval of passages, i.e. small textual elements. This task is faced with one key problem: as passages are small excerpts of longer documents, their content is not always sufficient to adequately estimate their relevance. To cope with this phenomenon, current approaches resort to contextualization [1, 5–7, 21, 28]; that is, the consideration of a passage’s context in its relevance estimation. We study here how to perform passage contextualization with methods akin to neural IR, as their expressive power have created a gap in performances compared with classical methods [10]. Multiple approaches represent a passage’s context based on the various types of relations it has with other parts of the document [1, 4, 7, 15, 23, 24, 28]. We investigate how such a representation, encoded as a graph, may be leveraged by graph neural networks. Graph neural networks, successfully applied on different Information Retrieval tasks [11, 16, 34, 35], aim at computing the embedding of nodes in a graph by considering their relations

with other nodes. More precisely, we investigate the use of attention-based graph neural networks, known as *Graph Attention Networks* (GATs) [30]. GATs rely on dense representation (embeddings) of the nodes’ content. Many embeddings approaches have been proposed for text retrieval [12, 13, 32] and multiple embeddings [13, 18] have been proved to be the most effective. To our knowledge, this is the first time GATs have been exploited in a Passage Retrieval task. We present our proposal in section 2, describing the graph document representation and our two models leveraging GATs to perform passage contextualization. In section 3, we conduct an evaluation on a patent passage retrieval task and conclude in section 4.

## 2 Proposal: Merged and Late Interaction Models

### 2.1 Document graph representation

We represent a document corpus as a graph where nodes are document parts and edges are the *intra* and *inter* relations between these documents parts. The document parts come from the logical structure of documents. To cope with the variation of such structure, we define two types of document parts: *sections*, i.e., non-textual units with a title, and *passages*, textual units without titles. The *intra*-document relations considered are: **(1)** the order of passages [4, 7, 15, 28], **(2)** the hierarchical structure [1, 23, 24] and **(3)** internal citations. The *inter*-document relation considered is: the **(4)** citation of one document by another one. We also include the inverse relations of these four relations [24, 25].

Our graph document representation is therefore composed of two types of node: *passage* nodes that represent textual units and *section* nodes that represent titled structural units – and eight types of edge (one for each relation and its symmetrical): *order* characterizing the relation order between *passage* nodes (*order<sub>i</sub>* its symmetrical), *structural* characterizing the composition between a *passage* node and a *section* node or between two *section* nodes (*structural<sub>i</sub>* its symmetrical), *internal* characterizing the *intra*-document citations between nodes (*internal<sub>i</sub>* its symmetrical) and *external* characterizing the *inter*-document citations between nodes (*external<sub>i</sub>* its symmetrical). Formally, the document corpus is a directed graph  $G = (V, E, A, R)$  where each node  $v \in V$  and each edge  $e \in E$  are associated with their type mapping functions  $\tau(v) : V \rightarrow A$  and  $\phi(e) : E \rightarrow R$ , respectively. We have  $A = \{passage, section\}$  and  $R = \{order, order_i, structural, structural_i, internal, internal_i, external, external_i\}$ .

### 2.2 Models architecture

We explore ways to compute a passage’s score by taking into account information about its content and information about its context using GATs. We derive two models leveraging the power of GATs: Merged Interaction Model (MiM) and Late Interaction Model (LiM). They both follow a ColBERT-inspired [13] efficient design and they differ in the way they consider the content of a passage. We first describe the elements that are shared by our two models before focusing on the motivations behind their differences.

**Encoder** We use the multiple representation text-encoder taken from the ColBERT model [13, 18] to embed text into a dense semantic space. A passage is embedded using its text, a section using its title and a query is using its text.

**Graph Attention Network** GATs are multi-layer graph neural networks which compute an embedding for each node in a graph by taking into account information from its neighbours [30]. Each layer aggregates, for each node, its embedding with the embedding of its neighbours using attention functions [3]. Stacking  $n$  layers allows a node to gather information about nodes that are at a distance of  $n$  hops in the graph. One element worth mentioning is that GATs implicitly add *self-edges* connecting each node to itself to build the embedding of a node.

Our model uses *Attention is all you need* [29] definition of attention and uses one attention function  $MultiHead_r$  per type of edge  $r$ , so the model treats the interaction between nodes differently according to the type of their relation [31]. The model defines a learnable weight vector  $W_r$  for each type of edge in the graph, representing the global importance of the relation. For a node  $i$ , we define its neighbour nodes  $N_i$ , and the edge between nodes  $i$  and  $j$   $e_{ij}$  (with  $j$  in  $N_i$ ). For  $h_i$  and  $h_j$ , their respective intermediate representation, the output of a layer is computed as:

$$h'_i = \sum_{j \in N_i} softmax(W_{\phi(e_{ij})}) * MultiHead_{\phi(e_{ij})}(h_i, h_j, h_j) \tag{1}$$

**Query Similarity Measure** As described in [13], the similarity between  $E_q$  and  $E_p$ , the multiple representation embeddings from a query  $q$  and a passage  $p$  is:

$$sim(E_q, E_p) = \sum_{i \in [1, |E_q|]} \max_{j \in [1, |E_p|]} E_{q_i} \cdot E_{p_j}^T \tag{2}$$

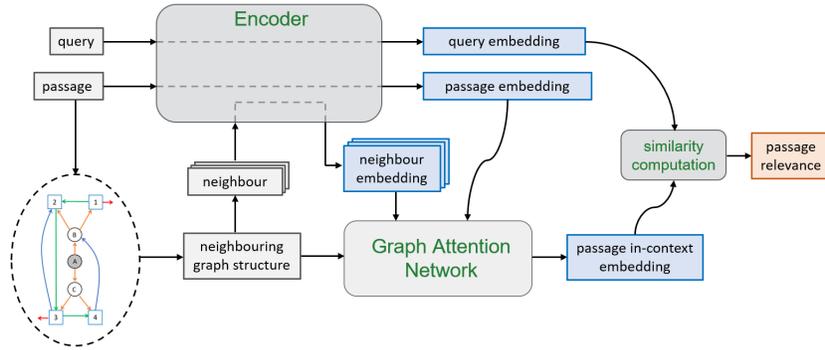


Fig. 1. Overview of the Merged Interaction Model.

**Merged Interaction Model** The *MiM*, described in Figure 1, proposes a *in-context* representation by simultaneously considering the content and the context for each passage. We compute such a representation of each passage by feeding

its embedding, its neighbours’ embedding and the neighbouring graph structure to the GAT. To obtain a passage’s relevance to a query, the model computes using the above-mentioned query similarity measure (2).

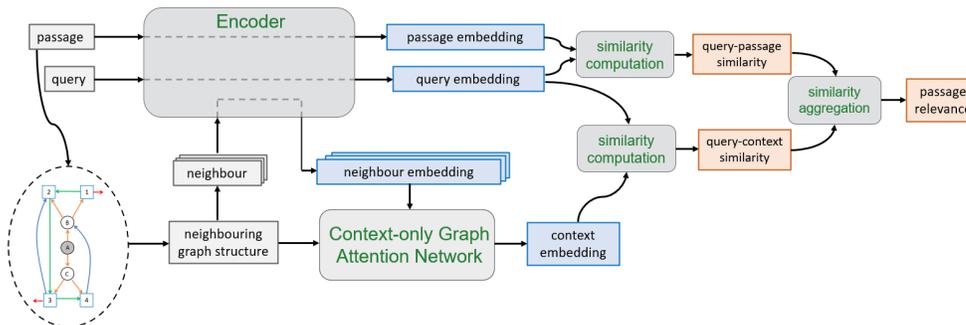


Fig. 2. Overview of the Late Interaction Model.

**Late Interaction Model** The *LiM*, shown in Figure 2, computes two embedding-based representations for a passage: one based on its content and one solely based on its context. Its relevance to a query is estimated using two similarity measures, computed using each representation. We are questioning here if mixing content and context information inside a single representation cannot impair its expressive power. For a passage  $p$ , the *LiM* computes a *content-based* embedding by feeding its text to the encoder, and a *context-based* embedding by making use of a Context-only Graph Attention Network (CGAT), a **modified Graph Attention Network** that does not consider  $p$ ’s content to compute its embedding.

To do that, a CGAT removes every edge going out of  $p$ ’s node in the graph. Note that this does not affect equation (1). To obtain a passage’s relevance to a query, the model first computes the *content-based* and the *context-based* query similarities with the above-mentioned similarity measure (2). Then, it aggregates the two similarities using a linear combination with a parameter  $\lambda \in [0, 1]$ :

$$relevance(q, p) = (1 - \lambda) * sim(E_q, E_p) + \lambda * sim(E_q, E_{context_p}) \quad (3)$$

where  $E_p$  is the *content-based* embedding and  $E_{context_p}$  is the *context-based* embedding of passage  $p$ .

## 3 Experiments

### 3.1 Experimental setup

We use a classical neural approach [13] to re-rank passages extracted from the top 1000 documents retrieved by BM25 [27] (with Terrier [17] default parameters).

**Dataset** We use the CLEF-IP2013 Passage Retrieval task [26]. This dataset, split between a training and testing set, contains structured patent documents with both internal and external citations. To derive the structure of documents

and the internal citations, we use handcrafted features either based on XML tags, case or number of characters. Each query is a set of “claims” along with a full patent document, that must be transformed into more refined queries [2, 19, 20, 33]. We use a state-of-the-art method [20]: let  $d$  be a query patent document with a set of claims, we build a first form of the query with the top-10 words with highest  $tf-idf$  in  $d$ ’s abstract to perform document retrieval, and a second form composed of  $d$ ’s natural language claims to perform passage reranking. We report the five official evaluation measures of the task:  $PRES@100$ ,  $Recall@100$ , and  $MAP@100$  at the document-level;  $MAP(D)$  and  $PREC(D)$  at the passage-level.

**Model characteristics** We use a text encoder from the ColBERT model trained on MSMARCO [22] with a maximum passage length of 180 tokens and a maximum query length of 120 tokens. Our GAT and CGAT are composed of 3 layers, each layer having several attention function  $MultiHead_r$  with 8 heads with dropout=0.7. The retrieval filters documents based on the International Patent Classification codes they share with the query patent document [9].

**Learning process** The Adam optimizer [14] is used to jointly learn the GATs or CGATs parameters and to fine-tune the encoder. For the encoder, we use the advised learning rate of  $3 * 10^{-6}$  [13] and freeze the first six layers. For the graph-based model, the weight vectors  $W_r$  and the parameter  $\lambda$ , we use a learning rate of  $1 * 10^{-3}$ . Our learning process is as follows: given a triple  $\langle q, p^+, p^- \rangle$  with query second form  $q$ , positive passage  $p^+$  and negative passage  $p^-$ , the model is optimised via pairwise softmax cross-entropy loss over the computed scores of  $p^+$  and  $p^-$ . The negative passage sampling for a pair  $\langle q, p^+ \rangle$  is done as follows: we randomly sample a passage from the corpus with a probability of 0.6 and randomly sample a non-relevant passage from the set of relevant documents with a probability of 0.4. The model is therefore confronted with positive and negative passages having similar contexts. This sampling process is repeated 1000 times for each  $\langle q, p^+ \rangle$ , yielding a training on 3.5M triplets.

### 3.2 Results

We evaluate the performances of our two models ( $MiM$ ,  $LiM$ ), a SoTA non-neural passage contextualization model ( $QSF_{sectionPropagateAVG}$  [1]), and two baseline models that do not consider context (fine-tuned ColBERT [13], BM25 [27]). For the sake of comparison, we also evaluate the performances of a model ( $Mix-iM$ ) that mixes our two proposed approaches: it computes a  $MiM$  *in-context* passage embedding along with a  $LiM$  *content-based* passage embedding and combines them the same way as the  $LiM$  model. As explained earlier, each of these approaches has been used to rank the passages of the documents retrieved initially by the BM25 model. We also report the results of a state-of-the-art approach focusing on the query generation [2], namely *Query-gen*. Additionally, we report the learned value of parameter  $\lambda$  for the  $LiM$  and  $Mix-iM$  models. Table 1 shows that  $QSF_{sectionPropagateAVG}$  and our three approaches ( $MiM$ ,  $LiM$ ,  $Mix-iM$ ) improve upon BM25 and ColBERT, which focus solely on the content of passages, confirming the importance of taking context into account

during passage relevance estimation [1]. Comparing our models, we see that *MiM* falls behind both *LiM* and *Mix-iM*, especially on the passage-level metrics. We hypothesize that *MiM* fails to correctly condense content and context information into a single representation leading to an over-prioritization of the context, which causes the model to rank high every (relevant or not-relevant) passages appearing in a relevant document. *LiM* outperforms *Mix-iM*, indicating that the model benefits from fully separating content and context information. Finally, we see that two of our approaches (*LiM*, *Mix-iM*) outperform *QSF<sub>sectionPropagateAVG</sub>* (significantly) and *Query-gen* on every evaluation measures, showing the strength of our methods for the patent passage retrieval task.

	Model	PRES@100	Recall@100	MAP@100	MAP(D)	PREC(D)
Content only	<i>BM25</i> [27]	0.385	0.482	0.125	0.142 <sup>k</sup>	0.21 <sup>k</sup>
	<i>ColBERT</i> [13]	0.402	0.518 <sup>o</sup>	0.161 <sup>o</sup>	0.145 <sup>k</sup>	0.214 <sup>k</sup>
	<i>Query-gen</i> [2]	0.444	0.560	0.187	0.146	0.282
Content & Context	<i>QSF<sub>sectionPropagateAVG</sub></i> [1]	0.460 <sup>oi</sup>	0.609 <sup>oi</sup>	0.169 <sup>o</sup>	0.201 <sup>oik</sup>	0.237 <sup>ok</sup>
	<i>MiM</i>	0.470 <sup>oi</sup>	0.568 <sup>oi</sup>	0.181 <sup>o</sup>	0.104	0.141
	<i>Mix-iM</i> ( $\lambda = 0.518$ )	0.541 <sup>oijk</sup>	0.631 <sup>oijk</sup>	0.257 <sup>oijk</sup>	0.246 <sup>oijk</sup>	0.299 <sup>oijk</sup>
	<i>LiM</i> ( $\lambda = 0.307$ )	<b>0.564<sup>oijk</sup></b>	<b>0.651<sup>oijk</sup></b>	<b>0.296<sup>oijkl</sup></b>	<b>0.270<sup>oijkl</sup></b>	<b>0.322<sup>oijk</sup></b>

**Table 1.** Performance over CLEF-IP2013 (in boldface: best result in a column). *o*, *i*, *j*, *k* and *l* represent statistical significance (two tailed Student paired t-test,  $p \leq 5\%$ ) over *BM25*, *ColBERT*, *QSF<sub>sectionPropagateAVG</sub>*, *MiM* and *Mix-iM* respectively. Statistical significance over *Query-gen* could not be computed (data not available from [2]).

## 4 Conclusion

In this work, we investigated the use of GATs to perform passage contextualization. First, we proposed a document graph representation based on document parts and their relations. Then, we presented two models leveraging GATs on this representation to estimate the relevance of a passage (to a query) according to both its content and its context. We evaluated our proposals on the CLEF-IP2013 patent passage retrieval task. Our results show that, while the use of GATs for passage contextualization improves the results, separately considering the content and context information of a passage leads to a significant gap in performances. In the future, we would like to conduct a parameter study on both our models in order to analyse more precisely how the context is taken into consideration. It would also be interesting to investigate other negative sampling techniques [32]. Finally, we plan to extend our experiments on more datasets such as the *INEX* Wikipedia dataset [8].

## Acknowledgement

This work has been partially supported by MIAI@Grenoble Alpes (ANR-19-P3IA-0003), as well as the Association Nationale de la Recherche et de la Technologie (ANRT).

## References

1. Albarede, L., Mulhem, P., Goeuriot, L., Le Pape-Gardeux, C., Marie, S., Chardin-Segui, T.: Passage retrieval in context: Experiments on patents. In: Proceedings of CORIA'21. Grenoble, France (Apr 2021), <https://hal.archives-ouvertes.fr/hal-03230421>
2. Andersson, L., Lupu, M., Palotti, J.a., Hanbury, A., Rauber, A.: When is the time ripe for natural language processing for patent passage retrieval? In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. p. 1453–1462. CIKM '16, Association for Computing Machinery, New York, NY, USA (2016). <https://doi.org/10.1145/2983323.2983858>, <https://doi.org/10.1145/2983323.2983858>
3. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate (2016)
4. Beigbeder, M.: Focused retrieval with proximity scoring. In: Proceedings of the 2010 ACM Symposium on Applied Computing. p. 1755–1759. SAC '10, Association for Computing Machinery, New York, NY, USA (2010). <https://doi.org/10.1145/1774088.1774462>, <https://doi.org/10.1145/1774088.1774462>
5. Bendersky, M., Kurland, O.: Utilizing passage-based language models for document retrieval. In: Proceedings of the IR Research, 30th European Conference on Advances in Information Retrieval. p. 162–174. ECIR'08, Springer-Verlag, Berlin, Heidelberg (2008)
6. Callan, J.P.: Passage-level evidence in document retrieval. In: Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 302–310. SIGIR '94, Springer-Verlag, Berlin, Heidelberg (1994)
7. Fernández, R., Losada, D., Azzopardi, L.: Extending the language modeling framework for sentence retrieval to include local context. *Inf. Retr.* **14**, 355–389 (08 2011). <https://doi.org/10.1007/s10791-010-9146-4>
8. Geva, S., Kamps, J., Lethonen, M., Schenkel, R., Thom, J.A., Trotman, A.: Overview of the *inex* 2009 ad hoc track. In: Geva, S., Kamps, J., Trotman, A. (eds.) *Focused Retrieval and Evaluation*. pp. 4–25. Springer Berlin Heidelberg, Berlin, Heidelberg (2010)
9. Gobeill, J., Ruch, P.: Bitem site report for the claims to passage task in CLEF-IP 2012. In: Forner, P., Karlgren, J., Womser-Hacker, C. (eds.) *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20, 2012*. CEUR Workshop Proceedings, vol. 1178. CEUR-WS.org (2012), <http://ceur-ws.org/Vol-1178/CLEF2012wn-CLEFIP-GobeillEt2012.pdf>
10. Guo, J., Fan, Y., Pang, L., Yang, L., Ai, Q., Zamani, H., Wu, C., Croft, W.B., Cheng, X.: A deep look into neural ranking models for information retrieval. *Information Processing Management* **57**(6), 102067 (2020). <https://doi.org/https://doi.org/10.1016/j.ipm.2019.102067>, <https://www.sciencedirect.com/science/article/pii/S0306457319302390>
11. Han, F., Niu, D., Lai, K., Guo, W., He, Y., Xu, Y.: Inferring search queries from web documents via a graph-augmented sequence to attention network. In: *The World Wide Web Conference*. p. 2792–2798. WWW '19, Association for Computing Machinery, New York, NY, USA (2019). <https://doi.org/10.1145/3308558.3313746>, <https://doi.org/10.1145/3308558.3313746>

12. Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., tau Yih, W.: Dense passage retrieval for open-domain question answering (2020)
13. Khattab, O., Zaharia, M.: Colbert: Efficient and effective passage search via contextualized late interaction over BERT. CoRR **abs/2004.12832** (2020), <https://arxiv.org/abs/2004.12832>
14. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2017)
15. Krikon, E., Kurland, O., Bendersky, M.: Utilizing inter-passage and inter-document similarities for reranking search results. ACM Trans. Inf. Syst. **29**(1) (Dec 2011). <https://doi.org/10.1145/1877766.1877769>, <https://doi.org/10.1145/1877766.1877769>
16. Li, X., de Rijke, M., Liu, Y., Mao, J., Ma, W., Zhang, M., Ma, S.: Learning better representations for neural information retrieval with graph information. In: Proceedings of the 29th ACM International Conference on Information Knowledge Management. p. 795–804. CIKM '20, Association for Computing Machinery, New York, NY, USA (2020). <https://doi.org/10.1145/3340531.3411957>, <https://doi.org/10.1145/3340531.3411957>
17. Macdonald, C., McCreddie, R., Santos, R.L., Ounis, I.: From puppy to maturity: Experiences in developing terrier. Proc. of OSIR at SIGIR pp. 60–63 (2012)
18. Macdonald, C., Tonellotto, N., Ounis, I.: On single and multiple representations in dense passage retrieval. CoRR **abs/2108.06279** (2021), <https://arxiv.org/abs/2108.06279>
19. Mahdabi, P., Gerani, S., Huang, J.X., Crestani, F.: Leveraging conceptual lexicon: Query disambiguation using proximity information for patent retrieval. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 113–122. SIGIR '13, Association for Computing Machinery, New York, NY, USA (2013). <https://doi.org/10.1145/2484028.2484056>, <https://doi.org/10.1145/2484028.2484056>
20. Mahdabi, P., Keikha, M., Gerani, S., Landoni, M., Crestani, F.: Building queries for prior-art search. In: Hanbury, A., Rauber, A., de Vries, A.P. (eds.) Multidisciplinary Information Retrieval. pp. 3–15. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
21. Murdock, V., Croft, W.B.: A translation model for sentence retrieval. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. pp. 684–691. Association for Computational Linguistics, Vancouver, British Columbia, Canada (Oct 2005), <https://www.aclweb.org/anthology/H05-1086>
22. Nguyen, T., Rosenberg, M., Song, X., Gao, J., Tiwary, S., Majumder, R., Deng, L.: MS MARCO: A human generated machine reading comprehension dataset. CoRR **abs/1611.09268** (2016), <http://arxiv.org/abs/1611.09268>
23. Norozi, M.A., Arvola, P.: Kinship contextualization: Utilizing the preceding and following structural elements. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 837–840. SIGIR '13, Association for Computing Machinery, New York, NY, USA (2013). <https://doi.org/10.1145/2484028.2484111>, <https://doi.org/10.1145/2484028.2484111>
24. Norozi, M.A., Arvola, P., de Vries, A.P.: Contextualization using hyperlinks and internal hierarchical structure of wikipedia documents. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management. p. 734–743. CIKM '12, Association for Computing Ma-

- chinery, New York, NY, USA (2012). <https://doi.org/10.1145/2396761.2396855>, <https://doi.org/10.1145/2396761.2396855>
25. Norozi, M.A., de Vries, A.P., Arvola, P.: Contextualization from the bibliographic structure (2012)
  26. Piroi, F., Lupu, M., Hanbury, A.: Overview of clef-ip 2013 lab. In: Forner, P., Müller, H., Paredes, R., Rosso, P., Stein, B. (eds.) *Information Access Evaluation. Multilinguality, Multimodality, and Visualization*. pp. 232–249. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
  27. Robertson, S., Walker, S., Jones, S., Hancock-Beaulieu, M., Gatford, M.: *Okapi at trec-3*. pp. 109–126 (1996)
  28. Sheerit, E., Shtok, A., Kurland, O.: A passage-based approach to learning to rank documents (2019)
  29. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need (2017)
  30. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks (2018)
  31. Wang, X., Ji, H., Shi, C., Wang, B., Cui, P., Yu, P.S., Ye, Y.: Heterogeneous graph attention network. *CoRR* **abs/1903.07293** (2019), <http://arxiv.org/abs/1903.07293>
  32. Xiong, L., Xiong, C., Li, Y., Tang, K.F., Liu, J., Bennett, P., Ahmed, J., Overwijk, A.: Approximate nearest neighbor negative contrastive learning for dense text retrieval (2020)
  33. Xue, X., Croft, W.B.: Automatic query generation for patent search. In: *Proceedings of the 18th ACM Conference on Information and Knowledge Management*. p. 2037–2040. CIKM '09, Association for Computing Machinery, New York, NY, USA (2009). <https://doi.org/10.1145/1645953.1646295>, <https://doi.org/10.1145/1645953.1646295>
  34. Yu, J., Lu, Y., Qin, Z., Liu, Y., Tan, J., Guo, L., Zhang, W.: Modeling text with graph convolutional network for cross-modal information retrieval (2018)
  35. Zhang, T., Liu, B., Niu, D., Lai, K., Xu, Y.: Multiresolution graph attention networks for relevance matching. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (Oct 2018). <https://doi.org/10.1145/3269206.3271806>, <http://dx.doi.org/10.1145/3269206.3271806>