Multistage BiCross Encoder: Team GATE Entry for MLIA Multilingual Semantic Search Task 2

Iknoor Singh, Carolina Scarton and Kalina Bontcheva
The University of Sheffield, UK

The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

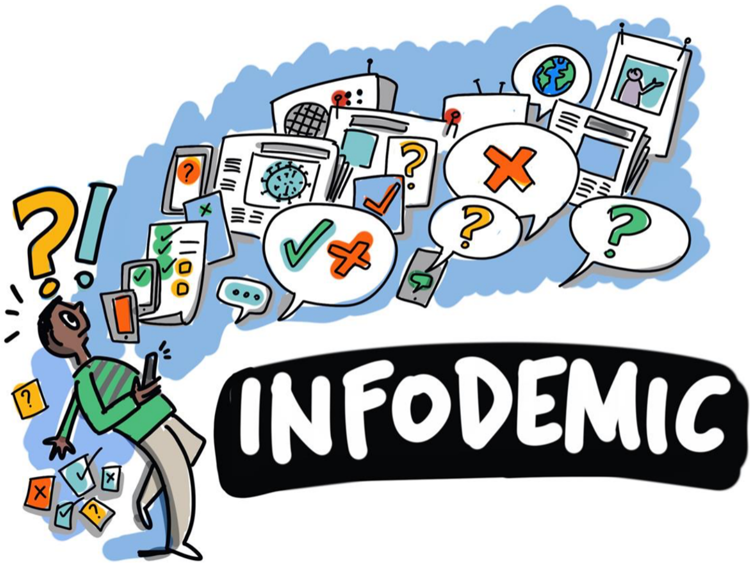
Overview

1. Introduction
2. Problem Statement
3. Proposed Approach
4. Evaluation and Results
5. Conclusion

The University Of Sheffield.

The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

1. Introduction



Ref. WHO Infodemiology Conference

- Coronavirus (COVID-19) pandemic has led to a rapidly growing 'infodemic' online.
- Researchers and journalists have been publishing a lot of content related to health, COVID-19 prevention methods, new laws and guidelines etc and this information comes from different parts of the world in multiple languages
- Accurate retrieval of reliable relevant data from millions of documents about COVID-19 has become urgently needed for the general public as well as for other stakeholders.

The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

2. Problem Statement

Ref. NIST Text Retrieval Conference (TREC) logo

How to build effective semantic search systems?

- Accurate retrieval of reliable relevant data

The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

2. Problem Statement

Ref. NIST Text Retrieval Conference (TREC) logo

How to build effective semantic search systems?

- Accurate retrieval of reliable relevant data
- Fast retrieval from millions of docs

The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

2. Problem Statement

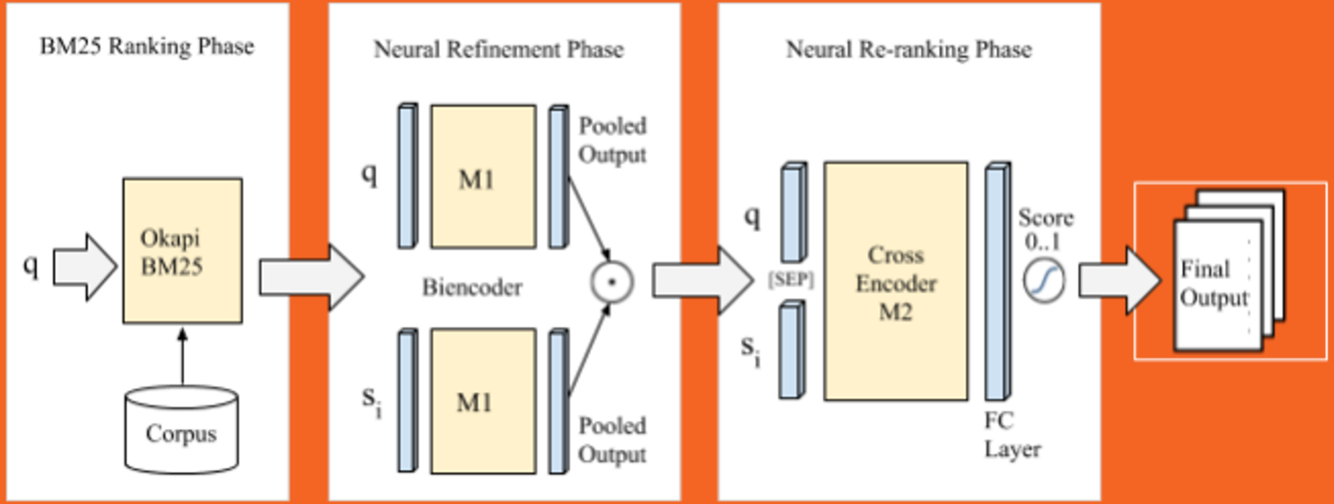
Ref. NIST Text Retrieval Conference (TREC) logo

How to build effective semantic search systems?

- Accurate retrieval of reliable relevant data
- Fast retrieval from millions of docs
- Support for multilingual search

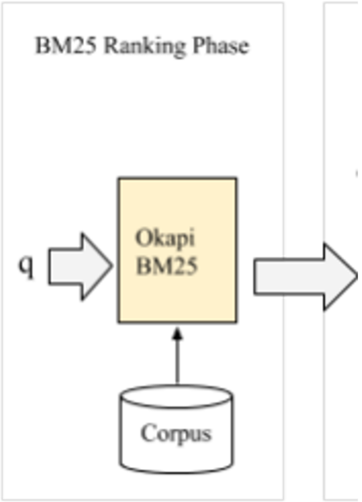
The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

3. Multistage BiCross Encoder



The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

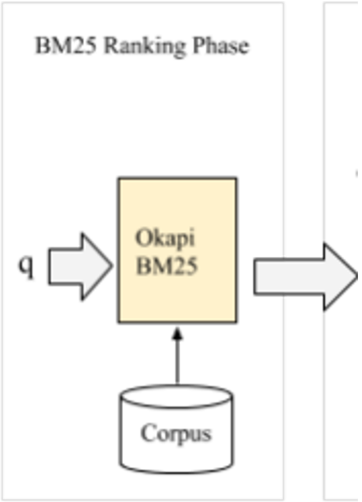
3.1 BM25 Ranking Phase



- Elasticsearch was used for initial retrieval using BM25 Okapi algorithm. It reduce the search space from a large number of documents (e.g. 1.4M in case of English documents) to a small set of possibly relevant documents.

Zhang, E., et al.: Covidex: Neural ranking models and keyword search infrastructure for the covid-19 open research dataset. arXiv preprint arXiv:2007.07846 (2020)
Nogueira, R., Lin, J., Epistemic, A.: From doc2query to docttttquery. Online preprint (2019)

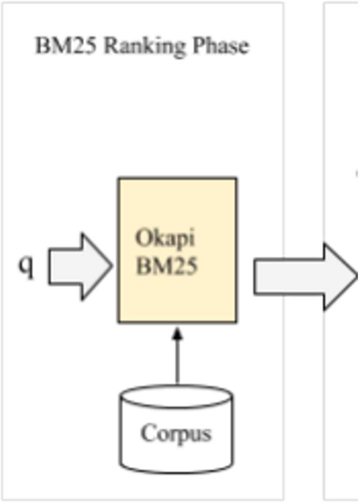
3.1 BM25 Ranking Phase



- Elasticsearch was used for initial retrieval using BM25 Okapi algorithm. It reduce the search space from a large number of documents (e.g. 1.4M in case of English documents) to a small set of possibly relevant documents.
- Only considered text inside the <p> tags and have excluded all the boilerplate tags. Text pre-processing methods such as stopwords removal and lemmatisation have been used before indexing the documents.

Zhang, E., et al.: Covidex: Neural ranking models and keyword search infrastructure for the covid-19 open research dataset. arXiv preprint arXiv:2007.07846 (2020)
Nogueira, R., Lin, J., Epistemic, A.: From doc2query to docttttquery. Online preprint (2019)

3.1 BM25 Ranking Phase

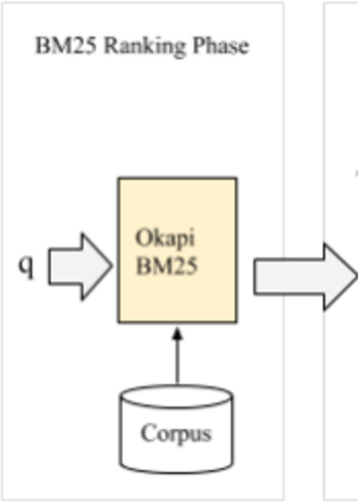


- Elasticsearch was used for initial retrieval using BM25 Okapi algorithm. It reduce the search space from a large number of documents (e.g. 1.4M in case of English documents) to a small set of possibly relevant documents.
- Only considered text inside the <p> tags and have excluded all the boilerplate tags. Text pre-processing methods such as stopwords removal and lemmatisation have been used before indexing the documents.
- Three types of query tried in our experiments,

  1. Concatenatenated keyword and conversational field

Zhang, E., et al.: Covidex: Neural ranking models and keyword search infrastructure for the covid-19 open research dataset. arXiv preprint arXiv:2007.07846 (2020)
Nogueira, R., Lin, J., Epistemic, A.: From doc2query to docttttquery. Online preprint (2019)

3.1 BM25 Ranking Phase

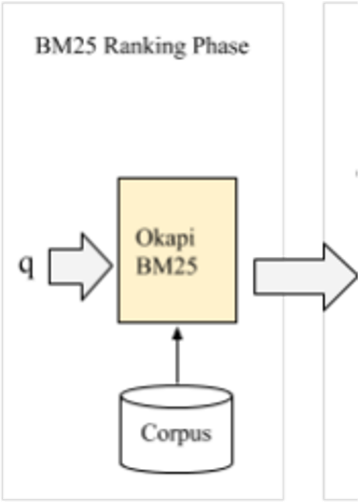


- Elasticsearch was used for initial retrieval using BM25 Okapi algorithm. It reduce the search space from a large number of documents (e.g. 1.4M in case of English documents) to a small set of possibly relevant documents.
- Only considered text inside the <p> tags and have excluded all the boilerplate tags. Text pre-processing methods such as stopwords removal and lemmatisation have been used before indexing the documents.
- Three types of query tried in our experiments,

  1. Concatenatenated keyword and conversational field
  2. Udels Query from TREC-COVID (Zhang et al. 2020)

Zhang, E., et al.: Covidex: Neural ranking models and keyword search infrastructure for the covid-19 open research dataset. arXiv preprint arXiv:2007.07846 (2020)
Nogueira, R., Lin, J., Epistemic, A.: From doc2query to docttttquery. Online preprint (2019)

3.1 BM25 Ranking Phase

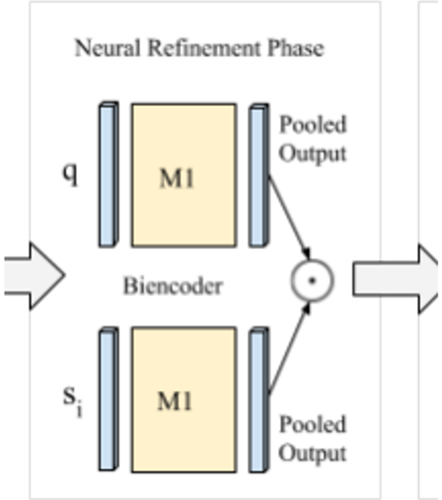


- Elasticsearch was used for initial retrieval using BM25 Okapi algorithm. It reduce the search space from a large number of documents (e.g. 1.4M in case of English documents) to a small set of possibly relevant documents.
- Only considered text inside the <p> tags and have excluded all the boilerplate tags. Text pre-processing methods such as stopwords removal and lemmatisation have been used before indexing the documents.
- Three types of query tried in our experiments,

1. Concatenatenated keyword and conversational field
2. Udels Query from TREC-COVID (Zhang et al. 2020)
3. T5 query using doc2query model (Nogueira et al. 2019)

Zhang, E., et al.: Covidex: Neural ranking models and keyword search infrastructure for the covid-19 open research dataset. arXiv preprint arXiv:2007.07846 (2020)
Nogueira, R., Lin, J., Epistemic, A.: From doc2query to docttttquery. Online preprint (2019)

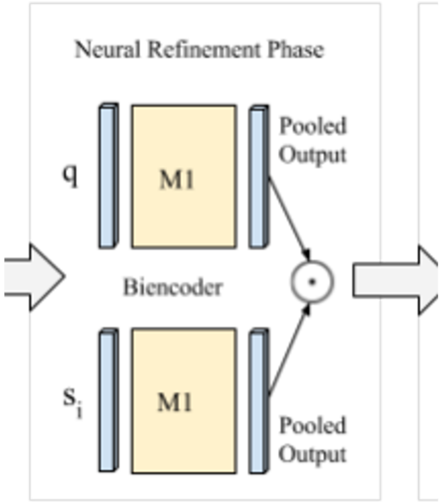
3.2 Neural Refinement Phase



- In this phase, bi-encoder is used to encode both query and document individually into contextualised representations. In the same vector space, query and relevant document lie in proximity of each other and can be efficiently retrieved using cosine similarity.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.
Yang, W., Zhang, H., Lin, J.: Simple applications of bert for ad hoc document retrieval. arXiv preprint arXiv:1903.10972 (2019)
TC+IFCN is a combined version of TREC-COVID data and IFCN COVID-19 debunk dataset. Cross_TC+IFCN is the translated version of TC+IFCN data.

3.2 Neural Refinement Phase

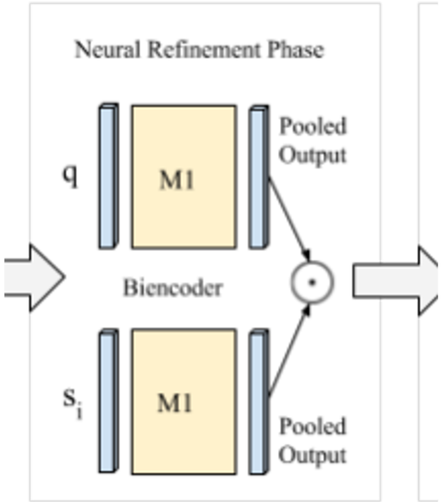


- In this phase, bi-encoder is used to encode both query and document individually into contextualised representations. In the same vector space, query and relevant document lie in proximity of each other and can be efficiently retrieved using cosine similarity.
- For training bi-encoder, TC+IFCN data was used to finetune transformer-based models using Siamese network architecture to get semantically meaningful sentence representations (Reimers et al. 2019). As the representations are separate, bi-encoder can store and reuse the encoded representation of inputs for faster predictions during inference.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.
Yang, W., Zhang, H., Lin, J.: Simple applications of bert for ad hoc document retrieval. arXiv preprint arXiv:1903.10972 (2019)

TC+IFCN is a combined version of TREC-COVID data and IFCN COVID-19 debunk dataset. Cross_TC+IFCN is the translated version of TC+IFCN data.

3.2 Neural Refinement Phase



- In this phase, bi-encoder is used to encode both query and document individually into contextualised representations. In the same vector space, query and relevant document lie in proximity of each other and can be efficiently retrieved using cosine similarity.

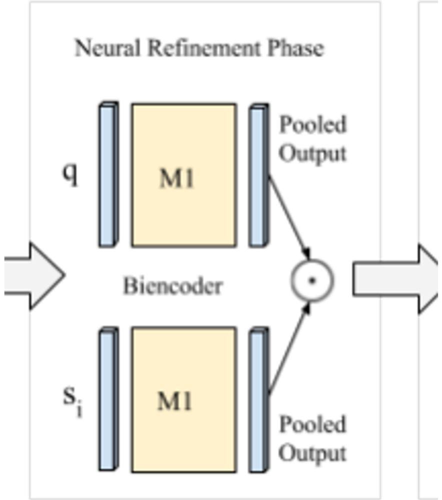
- For training bi-encoder, TC+IFCN data was used to finetune transformer-based models using Siamese network architecture to get semantically meaningful sentence representations (Reimers et al. 2019). As the representations are separate, bi-encoder can store and reuse the encoded representation of inputs for faster predictions during inference.

- Used sentence-level evidence where relevance score of each document is determined by combining the top k scoring sentences in the document (Yang et al. 2019)

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.
Yang, W., Zhang, H., Lin, J.: Simple applications of bert for ad hoc document retrieval. arXiv preprint arXiv:1903.10972 (2019)

TC+IFCN is a combined version of TREC-COVID data and IFCN COVID-19 debunk dataset. Cross_TC+IFCN is the translated version of TC+IFCN data.

3.2 Neural Refinement Phase



- In this phase, bi-encoder is used to encode both query and document individually into contextualised representations. In the same vector space, query and relevant document lie in proximity of each other and can be efficiently retrieved using cosine similarity.

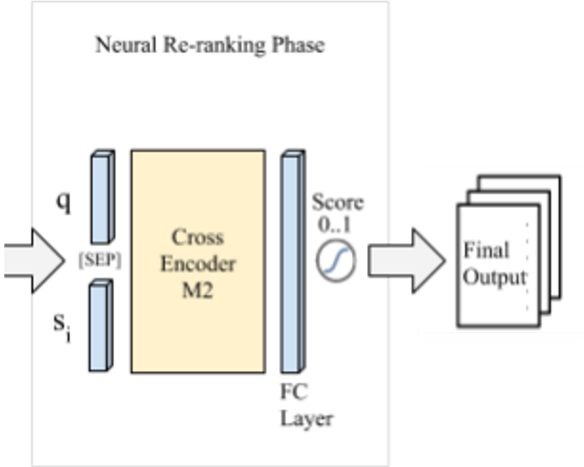
- For training bi-encoder, TC+IFCN data was used to finetune transformer-based models using Siamese network architecture to get semantically meaningful sentence representations (Reimers et al. 2019). As the representations are separate, bi-encoder can store and reuse the encoded representation of inputs for faster predictions during inference.

- Used sentence-level evidence where relevance score of each document is determined by combining the top k scoring sentences in the document (Yang et al. 2019)

- This stage will filter out all the semantically unrelated documents from BM25 retrieved documents.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.
Yang, W., Zhang, H., Lin, J.: Simple applications of bert for ad hoc document retrieval. arXiv preprint arXiv:1903.10972 (2019)
TC+IFCN is a combined version of TREC-COVID data and IFCN COVID-19 debunk dataset. Cross_TC+IFCN is the translated version of TC+IFCN data.

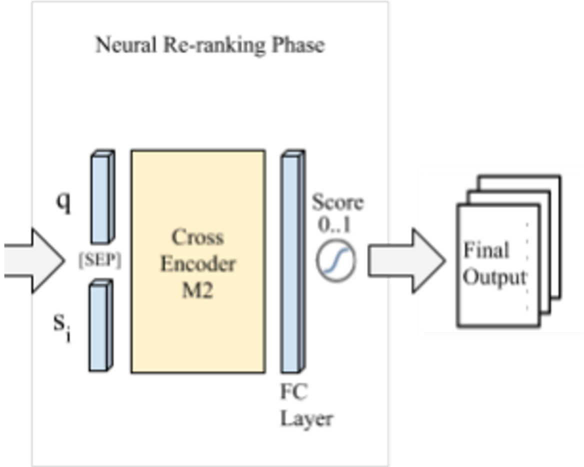
3.2 Neural Re-ranking Phase



- Transformer-based cross-encoder is used performs full self-attention over query and document pair to get the relevance score which is used to rank the final list documents with respect to the query

The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

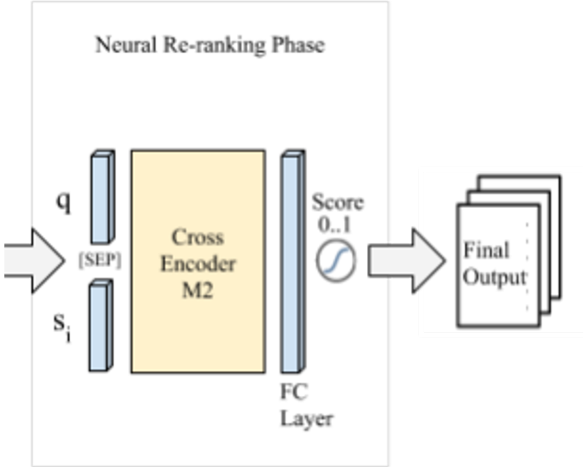
3.2 Neural Re-ranking Phase



- Transformer-based cross-encoder is used performs full self-attention over query and document pair to get the relevance score which is used to rank the final list documents with respect to the query
- In this, both query tokens and the document tokens separated by [SEP] token are passed to the model and the output of [CLS] token is passed to the linear layer with sigmoid activation to get relevance scores from 0 to 1 as illustrated.

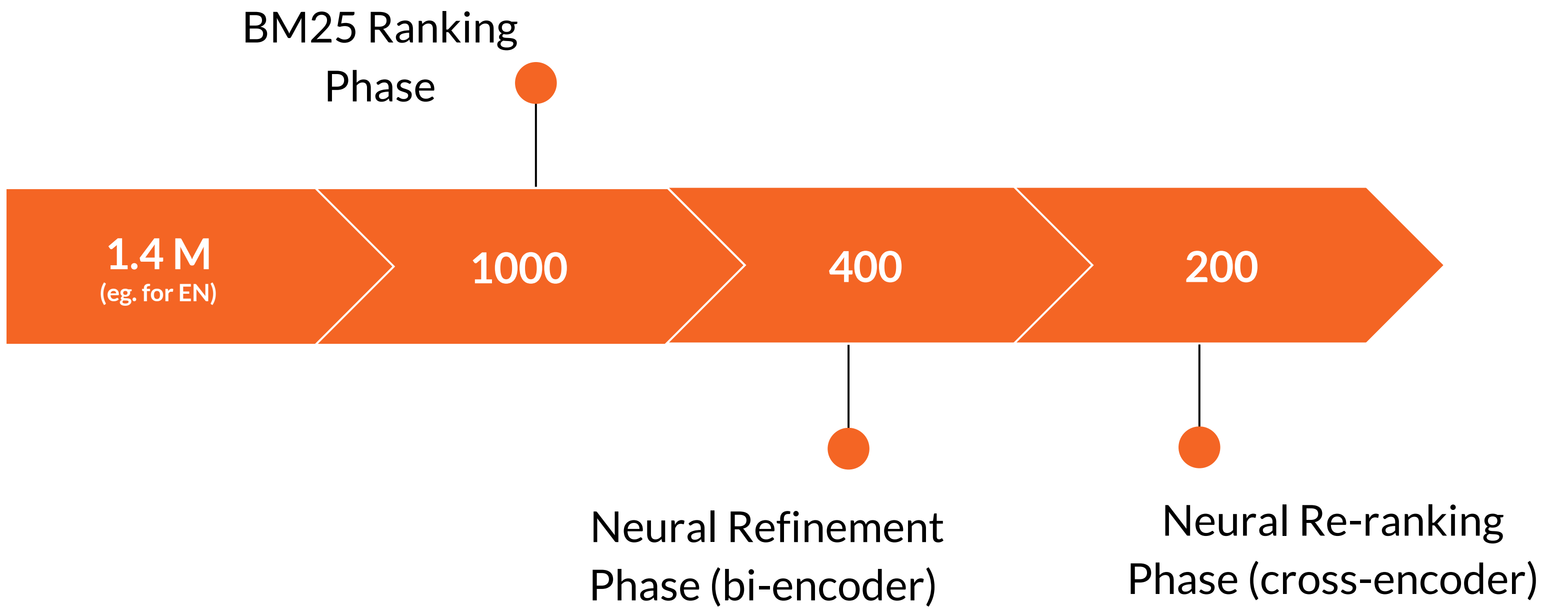
The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

3.2 Neural Re-ranking Phase



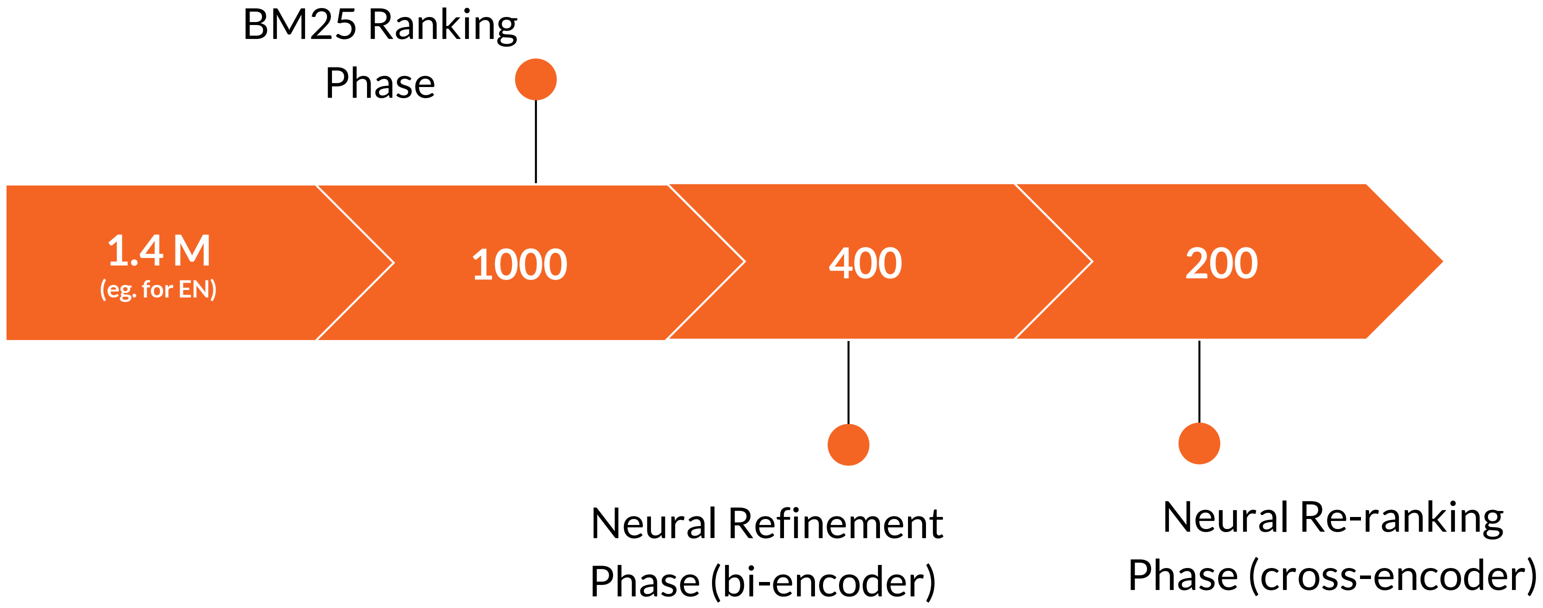
- Transformer-based cross-encoder is used performs full self-attention over query and document pair to get the relevance score which is used to rank the final list documents with respect to the query
- In this, both query tokens and the document tokens separated by [SEP] token are passed to the model and the output of [CLS] token is passed to the linear layer with sigmoid activation to get relevance scores from 0 to 1 as illustrated.
- Directly used the output of cross-encoder but for some runs, we also combined scores from previous stages using various fusion algorithms including score-based and rank-based fusion algorithms.

The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042



The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

The University Of Sheffield.



Neural Refinement Phase (bi-encoder)

- Representations can be cached
- Fast - can scale well for large scale search
- Separate encoding
- Less accurate

Neural Re-ranking Phase (cross-encoder)

- Representations cannot be cached
- Slow - doesn't scale well for large scale search
- Joint encoding
- More accurate

The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

The University Of Sheffield.

Evaluation and Results

Run ID	P@5	P@10	MAP	NDCG@10	NDCG	Rprec	Recall
gatenlp_run5	**0.9333**	**0.9000**	**0.2944**	**0.8331**	**0.5187**	**0.3486**	0.4382
gatenlp_run3	0.9200	0.8900	0.2912	0.8223	0.5155	0.3484	0.4375
gatenlp_run2	0.9000	0.7967	0.2560	0.7775	0.4925	0.3215	0.4278
gatenlp_run1	0.8867	0.8633	0.2776	0.8139	0.5067	0.3310	**0.4411**
gatenlp_run7	0.8667	0.8800	0.2719	0.8212	0.5014	0.3305	0.4292
CUNIMTIR_Run1	0.5933	0.4800	0.1145	0.4254	0.2802	0.1976	0.2613
CUNIMTIR_Run3	0.3600	0.3233	0.0609	0.2712	0.1444	0.1046	0.1278
CUNIMTIR_Run4	0.3533	0.3267	0.0530	0.2688	0.1422	0.0940	0.1239
ims_bm25_1k	0.3067	0.2433	0.0688	0.2391	0.2418	0.1579	0.2595
ims_bm25_2k	0.2400	0.1833	0.0478	0.1744	0.1789	0.1277	0.2028
ims_bm25_3k	0.2067	0.1633	0.0396	0.1413	0.1582	0.1075	0.1930
ims_bm25_4k	0.1933	0.1533	0.0367	0.1546	0.1483	0.1037	0.1677
gatenlp_run10	**0.9200**	0.8767	**0.3427**	0.8108	**0.6255**	**0.3711**	0.6334
ims_nlex	0.8933	**0.9000**	0.3055	**0.8365**	0.5740	0.3408	0.5593
gatenlp_run8	0.8867	0.8533	0.2999	0.8126	0.6092	0.3295	0.6334
ims_c-bm25	0.8600	0.8267	0.2771	0.7592	0.5945	0.3089	0.6482
ims_v-csum	0.8533	0.8233	0.2999	0.7693	0.6092	0.3450	**0.6516**
ims_bm25	0.7200	0.6900	0.2269	0.6202	0.5264	0.2673	0.6079
CUNIMTIR_Run5.	0.6867	0.6900	0.1908	0.5780	0.4574	0.2364	0.5160
CUNIMTIR_Run1	0.6800	0.5033	0.1659	0.4928	0.4450	0.2223	0.5148
ims_nsle	0.5067	0.5133	0.1595	0.4084	0.4145	0.2205	0.4837
CUNIMTIR_Run3	0.4800	0.3367	0.0882	0.2944	0.2500	0.1221	0.2986
CUNIMTIR_Run2	0.4667	0.3033	0.0646	0.3005	0.2379	0.1163	0.2683
CUNIMTIR_Run4	0.4267	0.3400	0.0658	0.2809	0.2200	0.1051	0.2662

Table 1. Performance of different monolingual English runs. Best results are bolded.

Run ID	P@5	P@10	MAP	NDCG@10	NDCG	Rprec	Recall
gatenlp_run37	**0.8333**	**0.7933**	0.2043	0.7263	0.3705	0.2806	0.3086
gatenlp_run25	0.8133	0.7767	0.2154	0.7455	**0.3808**	0.2795	**0.3111**
gatenlp_run28	0.8067	0.7767	0.2113	**0.7478**	0.3768	0.2758	0.3086
gatenlp_run34	0.7933	0.7833	0.2173	0.7383	0.3769	0.2858	0.3086
gatenlp_run31	0.7933	0.7867	**0.2246**	0.7362	0.3790	**0.2873**	0.3086
sinai_sinai1	0.5200	0.4867	0.0900	0.4629	0.2177	0.1557	0.1767
sinai_sinai2	0.4400	0.4067	0.0631	0.3868	0.1835	0.1284	0.1537
sinai_sinai4	0.3600	0.3067	0.0535	0.2904	0.1738	0.1243	0.1594
sinai_sinai3	0.2267	0.1733	0.0284	0.1820	0.1121	0.0786	0.1011
sinai_sinai5	0.2267	0.1733	0.0155	0.1832	0.0634	0.0407	0.0444
ims_bm25_1k	0.2067	0.1867	0.0577	0.1812	0.1944	0.1366	0.2142
ims_bm25_2k	0.2000	0.1800	0.0591	0.1745	0.2003	0.1402	0.2275
ims_bm25_3k	0.1733	0.1433	0.0444	0.1359	0.1744	0.1196	0.2072
ims_bm25_4k	0.0867	0.0800	0.0309	0.0793	0.1535	0.1046	0.1900
ims_c-bm25	**0.7000**	0.6933	0.1654	0.6346	**0.3993**	0.2224	**0.4084**
ims_v-csum	0.6867	**0.7133**	0.1697	**0.6604**	0.3797	0.2171	0.3612
ims_csum	0.6800	0.6200	**0.1720**	0.5822	0.3769	**0.2259**	0.3779
ims_bm25	0.6133	0.5800	0.1458	0.5263	0.3540	0.2020	0.3740
sinai_sinai1	0.5200	0.4867	0.1000	0.4629	0.2839	0.1560	0.2928
sinai_sinai2	0.4400	0.4067	0.0715	0.3868	0.2436	0.1285	0.2618
sinai_sinai4	0.3600	0.3067	0.0626	0.2904	0.2368	0.1247	0.2689
sinai_sinai5	0.2267	0.1733	0.0157	0.1832	0.0693	0.0408	0.0550
sinai_sinai3	0.2267	0.1733	0.0342	0.1820	0.1644	0.0788	0.1906

Table 2. Performance of different monolingual Spanish runs. Best results are bolded.

Run ID	P@5	P@10	MAP	NDCG@10	NDCG	Rprec	Recall
gatenlp_run26	**0.8800**	**0.7533**	**0.3505**	**0.7490**	**0.5672**	**0.3773**	**0.5267**
gatenlp_run29	0.8600	0.7400	0.3302	0.7324	0.5406	0.3651	0.4926
gatenlp_run32	0.8133	0.7367	0.3161	0.7180	0.5297	0.3593	0.4926
gatenlp_run35	0.8133	0.7267	0.3125	0.7116	0.5268	0.3541	0.4926
gatenlp_run38	0.7867	0.6400	0.2752	0.6436	0.5030	0.3269	0.4926

Table 3. Performance of different monolingual French runs. Best results are bolded.

Run ID	P@5	P@10	MAP	NDCG@10	NDCG	Rprec	Recall
gatenlp_run30	**0.9067**	**0.8767**	0.4537	**0.8234**	0.6403	0.4794	0.6253
gatenlp_run27	0.9000	0.8667	**0.4629**	0.8211	**0.6488**	0.4858	**0.6339**
gatenlp_run36	0.8733	0.8267	0.4442	0.7772	0.6377	0.4843	0.6253
gatenlp_run33	0.8733	0.8300	0.4531	0.7793	0.6399	**0.4972**	0.6253
gatenlp_run39	0.7733	0.7700	0.4227	0.7078	0.6200	0.4601	0.6253
ims_bm25_1k	0.1667	0.1633	0.0700	0.1475	0.2288	0.1413	0.3063
ims_bm25_2k	0.1667	0.1600	0.0793	0.1515	0.2176	0.1388	0.2769
ims_bm25_4k	0.1467	0.1433	0.0629	0.1396	0.1967	0.1120	0.2589
ims_bm25_3k	0.1400	0.1367	0.0650	0.1276	0.1924	0.1163	0.2488
ims_v-csum	**0.7267**	**0.6733**	**0.3447**	**0.6341**	**0.6174**	**0.3737**	0.7080
ims_csum	0.6267	0.5700	0.3072	0.5315	0.5731	0.3507	0.6940
ims_c-bm25	0.6133	0.5633	0.2890	0.5150	0.5667	0.3131	**0.7114**
ims_bm25	0.5933	0.5333	0.2869	0.4912	0.5572	0.3173	0.6924

Table 4. Performance of different monolingual German runs. Best results are bolded.

Run ID	P@5	P@10	MAP	NDCG@10	NDCG	Rprec	Recall
gatenlp_en2es_run49	**0.8533**	0.7367	0.1579	0.7042	0.3214	0.2273	**0.2565**
gatenlp_en2es_run52	0.8200	0.7700	0.1666	**0.7368**	**0.3287**	0.2286	**0.2565**
gatenlp_en2es_run40	0.8067	**0.7733**	**0.1740**	0.7211	0.3277	**0.2380**	**0.2565**
gatenlp_en2es_run43	0.8000	0.6867	0.1538	0.6555	0.3155	0.2287	**0.2565**
gatenlp_en2es_run46	0.7733	0.6367	0.1439	0.6330	0.3120	0.2231	**0.2565**
gatenlp_en2fr_run53	**0.8400**	**0.7467**	**0.2870**	**0.7245**	**0.4993**	0.3220	**0.4452**
gatenlp_en2fr_run41	0.8267	0.7033	0.2811	0.6980	0.4966	**0.3294**	**0.4452**
gatenlp_en2fr_run44	0.7667	0.6667	0.2527	0.6622	0.4801	0.3107	**0.4452**
gatenlp_en2fr_run47	0.7400	0.6300	0.2378	0.6234	0.4633	0.2980	0.4360
gatenlp_en2fr_run50	0.7133	0.6700	0.2506	0.6521	0.4712	0.3054	0.4360
gatenlp_en2de_run42	**0.8000**	**0.7600**	**0.2776**	**0.7300**	**0.4546**	**0.3307**	**0.3950**
gatenlp_en2de_run54	0.7733	0.7267	0.2680	0.7007	0.4484	0.3221	**0.3950**
gatenlp_en2de_run51	0.7200	0.6867	0.2475	0.6568	0.4334	0.3029	0.3907
gatenlp_en2de_run45	0.7133	0.6700	0.2474	0.6444	0.4349	0.3076	**0.3950**
gatenlp_en2de_run48	0.6867	0.6433	0.2292	0.6099	0.4187	0.2978	0.3907

Table 5. Performance of different bilingual Spanish (en2es), French (en2fr) and German (en2de) runs. Best results are bolded.

The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

Conclusion

1. We found Multistage BiCross Encoder to be highly effective at achieving state-of-the art performance on a wide range of metrics, including precision (P@10 & P@10) and NDCG at top ranks, R-precision, mean average precision and high recall for all the retrieved documents. Also, fine-tuning bi-encoder and cross-encoder model on TC+IFCN (or Cross_TC+IFCN) data proved to be beneficial and helped in achieving the highest scores.

2. For monolingual English runs, fine-tuning the bi-encoder model trained STS data gave the highest scores. For monolingual runs other than English, where we used multilingual models, the scores of German runs are comparatively higher, followed by French and Spanish runs respectively.

3. Overall, we find that runs which use key_conv as query perform better than Udels query and t5 query. Apart from this, we couldn't find any single model which perform good for all the languages as different models and methods give distinct results for different metrics.

The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

Conclusion

1. We found Multistage BiCross Encoder to be highly effective at achieving state-of-the art performance on a wide range of metrics, including precision (P@10 & P@10) and NDCG at top ranks, R-precision, mean average precision and high recall for all the retrieved documents. Also, fine-tuning bi-encoder and cross-encoder model on TC+IFCN (or Cross_TC+IFCN) data proved to be beneficial and helped in achieving the highest scores.

2. For monolingual English runs, fine-tuning the bi-encoder model trained STS data gave the highest scores. For monolingual runs other than English, where we used multilingual models, the scores of German runs are comparatively higher, followed by French and Spanish runs respectively.

3. Overall, we find that runs which use key_conv as query perform better than Udels query and t5 query. Apart from this, we couldn't find any single model which perform good for all the languages as different models and methods give distinct results for different metrics.

For future rounds, we plan to make further improvements to our approach, as well as extensively explore BiCross encoder for document retrieval for future research.

The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

Thanks

Arxiv Preprint: https://arxiv.org/abs/2101.03013

The WeVerify and SoBigData++ projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 825297 and 871042

References

Zhang, E., et al.: Covidex: Neural ranking models and keyword search infrastructure for the covid-19 open research dataset. arXiv preprint arXiv:2007.07846 (2020)

Nogueira, R., Lin, J., Epistemic, A.: From doc2query to docttttquery. Online preprint (2019)

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.

Yang, W., Zhang, H., Lin, J.: Simple applications of bert for ad hoc document retrieval. arXiv preprint arXiv:1903.10972 (2019)