MLIA-LIP6@TREC-CAsT2021:
Feature augmentation for query recontextualization and passage ranking

Nawel Astouati[1]
nawel.astouati@etu.upmc.fr

Thomas Gerald[1]
thomas.gerald@lip6.fr

Maya Touzari[1]
maya.touzari@etu.upmc.fr

Jian-Yun Nie[2]
nie@iro.umontreal.ca

Laure Soulier[1]
laure.soulier@lip6.fr

[1]Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
[2]University of Montreal, Montreal, Canada

Machine Learning & Deep Learning for Information Access

TREC CAST 2021

Turn	Query
1	How do I build a cheap driveway?
2	Which is cheaper: concrete or asphalt?
3	Really? What type of product ?
4	Who knew? Which is more environmentally friendly ?
5	No. Which type of driveway is better for the environment ?
6	And most low-maintenance ?
7	Really? What about asphalt ?
8	Is sealing worth it ?

Reference to previous queries

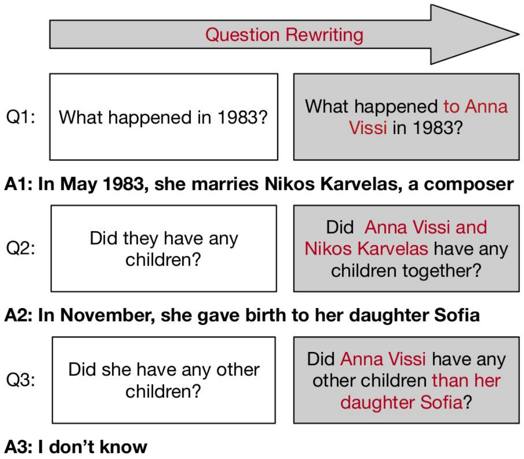
Reference to previous documents

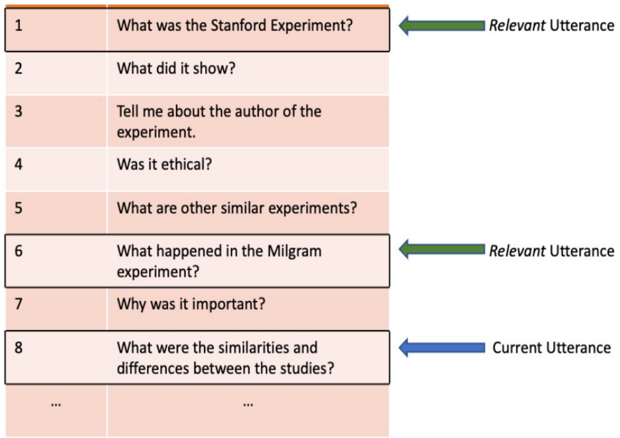
TREC-CAST 2021 MLIA-LIP6 submission

What information is required to contextualize query in conversational IR ?

- Focus on the reformulation part
- Evaluate pipeline (Reformulation and Ranking) based on different features
  - Raw queries
  - Previous reformulated queries
  - Previous documents
  - Queries generated from previous documents
- End2End version of the pipeline

Corpora

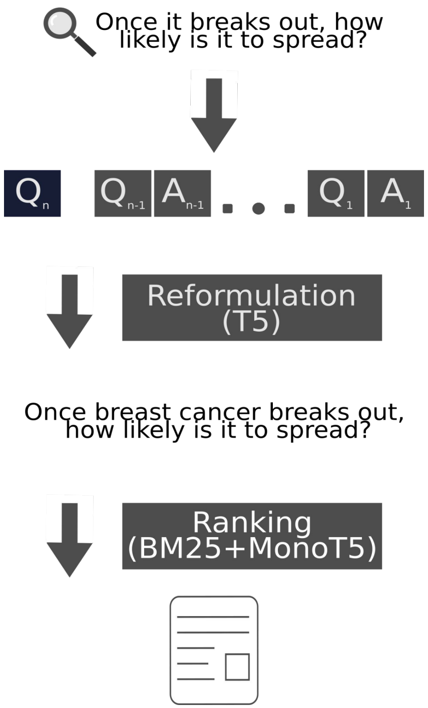




Copora

- CANARD [4]
  - Conversational queries corpus with reformulations and short answer.
- CASTuR [5]
  - Conversational queries with relevant utterances + our own manually rewritten queries
- MSMarco-Passage Ranking [6]
  - Passage ranking corpus with couples of query-document

Pipeline models



Training:

- Reformulation based on T5 model [7]

We fine-tune on the CANARD dataset + CASTUR dataset (manual query rewriting).

Inputs: reformulated queries + current query

- Ranking with Pretrained Mono-T5 [1]

t5_monot5 (Whole Conversation)

- Previous queries
- Previous documents
- Current Queries

Rewritt5_monot5 (online reformulation)

- Previous reformulated queries (automatic rewritten utterance)
- Previous documents
- Current Queries

t5_doc2query (intent on the context)

- Previous reformulated queries (automatic)
- Queries formulated on previous documents (Doc2Queries [2])
- Current queries

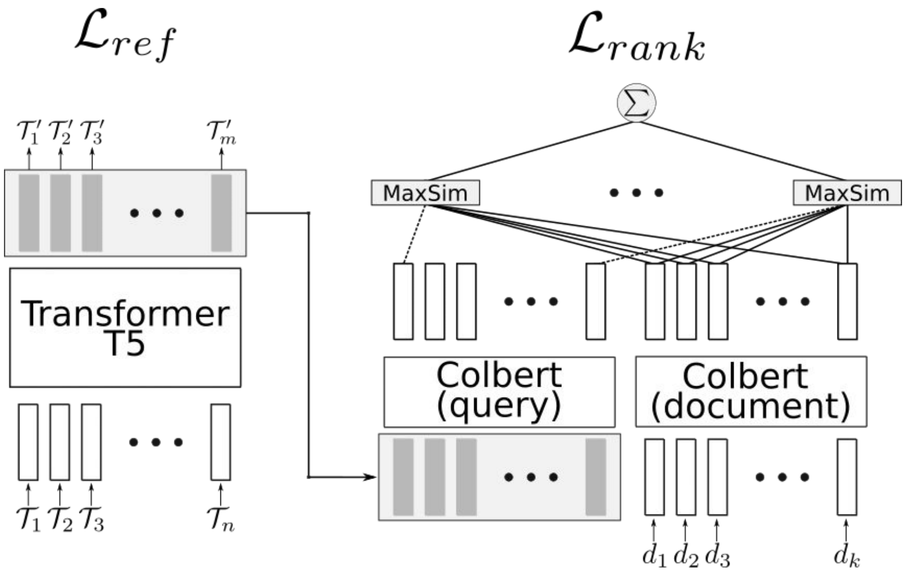
T5-Colbert

Objective

- Injecting IR signals in the reformulation module to solve the ranking task in a end-to-end fashion

The T5-Colbert model:

- Reformulation module (T5-pretrained)
- Feedforward (between T5 embedding and BERT input)
- Ranking Module (configuration Colbert [3])



Optimization

$$E(f \circ g) = \sum_{c_i \in C} \Big[\mathcal{L}_{ref}(f_\theta(c_i), r_i) \tag{1}$$

$$+ \sum_{d_p \in \mathbf{D}_{i,p}} \sum_{d_n \in \mathbf{D}_{i,n}} \mathcal{L}_{rank}\Big(g(f_{\theta,e}(c_i), d_p), g(f_{\theta,e}(c_i), d_n) \Big) \Big] \tag{2}$$

T5-Colbert : Training Corpus

Unified Corpus

- Conversational queries extracted from CANARD Corpus
- Documents are extracted from MSMarco Passages:
    a. Retrieved Top 200 documents from MonoT5 reranking
    b. Select first 4 documents as positive documents for the current query
    c. Last 4 documents as negative document for the current query

123.836 negative/positive pairs (30.959 queries) for training and 1152 for validation

Ranking Performances (x 100)

Model	NDCG@3	NDCG@5	NDCG@500	MAP@500
t5_monot5	**38.7**	**39.0**	**33.6**	19.5
Rewritt5_monot5	36.9	37.2	33.1	18.9
t5_doc2query	37.7	37.9	33.5	**19.7**
E2E (t5colbert)	15.3	15.8	31.4	10.1

Pipeline results

- T5-monoT5 for short range metrics
- T5-doc2query gets similar performances at longer range

E2E evaluation

- Using Sparse ranking (based on TREC baseline documents)
- Difficult to assess the relevance of the dataset (alignment of query-documents)

Results: Reformulation

Turn	Query
1	I heard Bernie Sanders was ill recently. What happened?
2	What did the doctors do?
3	How did it affect the campaign?
4	How did he attempt to regain the public's confidence?
5	Okay. What did the records say?

Previous Queries Information

Previous Documents Information

Model	Reformulation
Query	How did he attempt to regain the public's confidence?
t5_monot5	How did Sen Bernie Sanders attempt to regain the public s confidence in his ability to serve
Rewritt5_monot5	How did Sen Bernie Sanders attempt to regain the public s confidence in his ability to serve as president ?
t5_doc2query	How did Bernie Sanders attempt to regain the public s confidence?
E2E	How did Bernie Sanders attempt to regain the public's confidence?
Query	Okay. What did the records say?
t5_monot5	What did the records say about Sen Bernie Sanders ?
Rewritt5_monot5	What did the records say about Sen Bernie Sanders ?
t5_doc2query	What did the records say about Sen Bernie Sanders ?
E2E	Okay. What did the records say about Bernie Sanders 's heart attack ?

Conclusion

Submitted approaches

- **Pipeline Model**
  - Better using non reformulated queries
- **T5-Colbert**
  - Low performances:
    - Training set pairing issues
    - Hyperparameters search
    - Missing time to effectively use dense retrieval

What's next ?

- Training with all configuration
- Better optimization of E2E model

References

[1] R. Nogueira, Z. Jiang, R. Pradeep, and J. Lin. Document ranking with a pretrained sequence-to-sequence model. In EMNLP, pages 708–718. Association for Computational Linguistics, 2020.

[2] R. Nogueira, W. Yang, J. Lin, and K. Cho. Document expansion by query prediction. CoRR, abs/1904.08375, 2019.

[3] O. Khattab and M. Zaharia. Colbert: Efficient and effective passage search via contextualized late interaction over BERT. In SIGIR conference on research and development in Information Retrieval, pages 39–48. ACM, 2020.

[4] A. Elgohary, D. Peskov, and J. Boyd-Graber. Can you unpack that? learning to rewrite questions-in-context. In Empirical Methods in Natural Language Processing, 2019.

[5] M. Aliannejadi, M. Chakraborty, E. A. Rıssola, and F. Crestani. Harnessing evolution of multi-turn conversations for effective answer retrieval. In Proceedings of 2020 Conference on Human Information Interaction and Retrieval (CHIIR), CHIIR ’20, 2020.

[6] T. Nguyen, M. Rosenberg, X. Song, J. Gao, S. Tiwary, R. Majumder, andL. Deng. MS MARCO: A human generated machine reading comprehension dataset. In T. R. Besold, A. Bordes, A. S. d’Avila Garcez, and G. Wayne. Neural Information Processing Systems (NIPS 2016)

[7] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou,W. Li, and P. J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer.J. Mach. Learn. Res., 21:140:1–140:67, 2020.

Different features (CAST 2020)

Modèle	map	recip_rank	ndcg
allHistory-R_T5	**0.136**	**0.294**	**0.213**
RighHistory-R_T5	**0.127**	**0.224**	**0.204**
allHistory+Passage_T5	**0.126**	**0.284**	**0.203**
RightHistory+Passage_T5	**0.110**	**0.256**	**0.181**
allHistory-R+Passage_T5	**0.132**	**0.289**	**0.219**
RightHistory-R+Passage_T5	**0.124**	**0.220**	**0.205**

Doc2Query Features

Modèle	map	recip_rank	ndcg
OnlyPassages_T5	**0.108**	**0.230**	**0.182**
allHistory-R+doc2query_T5	**0.116**	**0.258**	**0.183**
allRewrittenHistory_T5+doc2query_(BM25+MonoT5)	**0.078**	**0.177**	**0.137**

Modèle	map	recip_rank	ndcg
BM25	**0.050**	**0.142**	**0.135**
BM25+MonoT5	**0.060**	**0.153**	**0.111**
Raw_colbert	**0.126**	**0.262**	**0.183**
AllRewrittenH_T5+Colbert	**0.123**	**0.222**	**0.135**