LongEval Click-Model Relevance Judgements (Qrels) Readme

1. **About**

   The collection comprises the relevance judgments used in the 2023 LongEval Information Retrieval Lab (https://clef-longeval.github.io/), organized at CLEF. It consists of three sets of relevance judgments:

   1) Relevance judgments for the heldout queries from the LongEval Train Collection (http://hdl.handle.net/11234/1-5010).
   2) Relevance judgments for the short-term persistence (sub-task A) queries from the LongEval Test Collection (http://hdl.handle.net/11234/1-5139).
   3) Relevance judgments for the long-term persistence (sub-task B) queries from the LongEval Test Collection (http://hdl.handle.net/11234/1-5139).

   These judgments were provided by the Qwant search engine (https://www.qwant.com) and were generated using a click model. The click model output was based on the clicks of Qwant's users, but it mitigates noise from raw user clicks caused by positional bias and also better safeguards users' privacy. Consequently, it can serve as a reliable soft relevance estimate for evaluating and training models.

   The collection includes a total of 1,420 judgments for the heldout queries, with 74 considered highly relevant and 326 deemed relevant. For the short-term sub-task queries, there are 12,217 judgments, including 762 highly relevant and 2,608 relevant ones. As for the long-term sub-task queries, there are 13,467 judgments, with 936 being highly relevant and 2,899 relevant.

   Contact: galuscakova@gmail.com

2. **Authors**

   Petra Galuščáková, Université Grenoble Alpes
   Romain Devaud, Qwant
   Gabriela Gonzalez-Saez, Université Grenoble Alpes
   Philippe Mulhem, Université Grenoble Alpes
   Lorraine Goeuriot, Université Grenoble Alpes
   Florina Piroi, Research Studios Austria, Data Science Studio
   Martin Popel, Charles University in Prague

3. **Data**

   |-heldout-test.txt
   |- a-short-july.txt
   |- b-long-september.txt

|- readme.pdf

The qrels are published in the standard trec format, which allows application of the standard IR evaluation tools such as trec_eval (https://github.com/usnistgov/trec_eval):

`q06223196 0 doc062200204465 1`

where q06223196 is query id, doc062200204465 is document id and 1 refers to the information about the relevance. The document might be either highly relevant (2), relevant (1), or irrelevant (0) to the query.

4. Citation

Please refer the following article to cite this item:
Petra Galuščáková, Romain Deveaud, Gabriela Gonzalez-Saez, Philippe Mulhem, Lorraine Goeuriot, Florina Piroi, Martin Popel: LongEval-Retrieval: French-English Dynamic Test Collection for Continuous Web Search Evaluation, arXiv:2303.03229, 2023
https://arxiv.org/abs/2303.03229

Alternatively, you might also cite Lindat record:
http://hdl.handle.net/11234/1-5151

5. Licence

The collection is shared under the Qwant LongEval Attribution-NonCommercial-ShareAlike License.

6. Acknowledgements