Web-scale Retrieval Experimentation with `chatnoir-pyterrier`

Jan Heinrich Merker[1], Janek Bevendorff[2], Maik Fröbe[1], Tim Hagen[3], Harrisen Scells[3], Matti Wiegmann[2], Benno Stein[2], Matthias Hagen[1], and Martin Potthast[3,4]

[1] Friedrich-Schiller-Universität Jena, Germany
[2] Bauhaus-Universität Weimar, Germany
[3] University of Kassel and hessian.AI, Germany
[4] ScaDS.AI, Germany

Abstract The IR community has always aimed to improve the realism of retrieval experiments by increasing the size of the document collections. As collection sizes grow from megabytes to giga-, tera-, and maybe soon petabytes, IR labs are challenged to keep pace. Herein, we describe our work on integrating ChatNoir with `ir_datasets` and PyTerrier to create `chatnoir-pyterrier`, a Python package for using ChatNoir in multi-stage pipelines. ChatNoir provides BM25-based first-stage retrieval on all ClueWeb crawls and all MS MARCO variants with a collective index size of about 20 TB. This improves inclusivity by lowering the barrier to entry for web-scale IR, and reduces redundant first-stage indexing overhead across IR labs. We show how `chatnoir-pyterrier` simplifies a wide range of re-ranking approaches and facilitates retrieval-augmented generation setups against large corpora.

1 Introduction

Despite the ability of modern information retrieval (IR) systems to support web-scale collections, access to and use of such collections is limited. Firstly, web-scale collections are too large to be distributed over the Internet and instead often require multiple hard drives to be sent by physical mail [31]. This is not only slow, but also expensive, especially if the collections are used by many research groups. Secondly, the infrastructure required to index and search these collections poses a high barrier to entry for those without access to hundreds of terabytes of storage and high performance computing infrastructure [2]. This limits research on web-scale information retrieval to only a fraction of IR researchers. In this paper, we present `chatnoir-pyterrier`,[5] a Python library that significantly lowers this barrier, enabling anyone to conduct research on large information retrieval collections. With `chatnoir-pyterrier`, users can immediately and easily search collections such as the ClueWebs [14,15,30] and MS MARCOs [28], without having to index these collections themselves. The integration with PyTerrier [26] also allows to build advanced ranking pipelines on top of ChatNoir.

[5] Code and examples: `https://github.com/chatnoir-eu/chatnoir-pyterrier`

Table 1: Datasets available via `chatnoir-pyterrier`, (compressed) dataset size, index size, and link to the ChatNoir UI. In gray, we list shared tasks that use the dataset.

Dataset	ID	Dataset Docs.	Dataset Size	Index Docs.	Index Size	Link
ClueWeb09 [14]	`clueweb09`	1.0 B	4.0 TB	536 M	6.2 TB	🔗
Shared tasks: TREC Million Query 2009 [8], TREC Web 2009–2012 [9–12]						
ClueWeb12 [15]	`clueweb12`	733 M	4.6 TB	503 M	5.1 TB	🔗
Shared tasks: TREC Web 2013/2014 [16,17], CLEF eHealth [32,37], NTCIR WWW 2017–2020 [24,27,35], TREC Health Misinformation 2019 [1], CLEF Touché 2020–2022 [3,5,6]						
ClueWeb22 cat. B [30]	`clueweb22/b`	200 M	12 TB	187 M	6.0 TB	🔗
Shared tasks: CLEF Touché 2023 [4], TREC Lateral Reading 2024 (ongoing)						
MS MARCO v1 [28]	`msmarco-document`	3.2 M	8.5 GB	3.2 M	189 GB	🔗
↪ segmented	`msmarco-passage`	8.8 M	1.1 GB	8.8 M	12 GB	🔗
Shared tasks: TREC Deep Learning 2019/2020 [18,21]						
MS MARCO v2 [28]	`msmarco-document-v2`	12 M	35 GB	12 M	479 GB	🔗
↪ segmented	`msmarco-passage-v2`	138 M	22 GB	138 M	251 GB	🔗
Shared tasks: TREC Deep Learning 2021–2023 [19,20,22]						
MS MARCO v2.1 [28]	`msmarco-document-v2.1`	11 M	29 GB	11 M	539 GB	🔗
↪ segmented	`msmarco-passage-v2.1`	114 M	26 GB	114 M	975 GB	🔗
Shared tasks: TREC RAG 2024 (ongoing)						
∑ **Total**	—	**2.2 B**	**21 TB**	**1.5 B**	**20 TB**	—

```python
from chatnoir_pyterrier import ChatNoirRetrieve
chatnoir = ChatNoirRetrieve("msmarco-document-v2.1")
chatnoir.search("python library")
```

Listing 1: Searching for the query `python library` in MS MARCO v2.1.

We demonstrate that `chatnoir-pyterrier` simplifies the participation in shared tasks and retrieval-augmented generation experiments [23], and hence, enables students to participate in state-of-the-art IR research. In its current state, 1.5 billion documents can be easily searched (see Table 1). which would otherwise require about 20 TB of disk space just to store the indices—an unattainable amount of resources for individuals and many labs.

2 Simplified Retrieval with `chatnoir-pyterrier`

The `chatnoir-pyterrier` library uses the ChatNoir research search engine [2,33],[6] which employs a 130-node Elasticsearch cluster for retrieval and a 10 PB Ceph S3 storage cluster for random document access, both with sharding and replication. To build `chatnoir-pyterrier`, we extended ChatNoir to index any `ir_datasets` compatible dataset [25]. The standardized access provided by `ir_datasets` to document collections ensures compatibility with Cranfield-style IR evaluation [13]. On ChatNoir, we then maintain indices for each of these datasets, as listed in Table 1. Administrators can easily index new datasets to `chatnoir-pyterrier` as long as it is available through `ir_datasets` or a compatible extension package[7].

[6] Hosted by the Webis group: `https://chatnoir.eu`

[7] E.g., for ClueWeb22: `https://pypi.org/project/ir-datasets-clueweb22`

```
chatnoir = ChatNoirRetrieve("clueweb12", num_results=100)
mono_t5 = MonoT5ReRanker("castorini/monot5-base-msmarco")
duo_t5 = DuoT5ReRanker("castorini/duot5-base-msmarco")
pipeline = chatnoir % 100 >> mono_t5 % 5 >> duo_t5
```

Listing 2: Composing a multi-stage re-ranking pipeline for the ClueWeb12 with `chatnoir-pyterrier`, monoT5, and duoT5, using PyTerrier's operator syntax.

```
query = "How tall is the empire state building?"
search = ChatNoirRetrieve(features=Feature.SNIPPET_TEXT)
context = "\n\n".join(search.search(query)["snippet_text"])
prompt = f"""Answer the question using the given context:
Question: {query}\nContext: {context}"""
print(OpenAI(...).chat.completions.create(
    model="gpt-4o",
    messages=[{"role": "user", "content": prompt}],
).choices[0].message.content)
```

Listing 3: Simple RAG system using `chatnoir-pyterrier` and the OpenAI API.

For indexed datasets, access to ChatNoir's BM25 search is exposed as a typical search engine user interface and via the ChatNoir search API.[8] The `chatnoir-pyterrier` library exposes ChatNoir's API functionalities in a PyTerrier compatible, type-safe, and unit-/integration-tested Python interface. Our library handles pagination, parsing,[9] retries, rate limits, and allows keyword- or phrase search. Listing 1 shows how the core module, `ChatNoirRetrieve`, can replace PyTerrier's standard retrieval module, `Retriever` (for Terrier indices [29]).

Finally, the first-stage retrieval of `chatnoir-pyterrier` can easily be extended with neural re-rankers using PyTerrier's composition API [26]. In Listing 2, we build a multi-stage retrieval system using monoT5 and duoT5 cross-encoder models [34]. The system retrieves the top-100 documents from ChatNoir's ClueWeb12 index, re-ranks using monoT5, and again re-ranks the top-5 with duoT5.[10]

3 `chatnoir-pyterrier` for Retrieval-Augmented Generation

Retrieval-augmented generation (RAG), in which an LLM answers questions using the top retrieved passages as context, received much attention recently [23].[11] However, RAG on large indices remains limited to those with the resources to store and maintain them. With `chatnoir-pyterrier`, this barrier is lowered substantially, enabling RAG experiments even on resource-constrained services like Google Colab or GitHub Codespaces—especially when paired with projects like `llama.cpp`[12] which allow LLM inference on standard hardware. Listing 3 demonstrates a basic RAG system with `chatnoir-pyterrier` and OpenAI.

[8] UI: `https://chatnoir.eu`; API: `https://chatnoir.eu/api`
[9] Metadata: doc. IDs, index ID, hostname, URL, cache URL, crawl date, page/spam rank, title, snippet, raw HTML/JSON, plain text, content type, explanation, language
[10] `https://hf.co/castorini/monot5-base-msmarco` and `duot5-base-msmarco`
[11] DBLP search for "retrieval-augmented generation": 60 papers in 2023; 539 in 2024.
[12] `https://github.com/ggerganov/llama.cpp`

Table 2: nDCG@5 effectiveness of ChatNoir BM25 and mono/duoT5 re-ranking (see Listing 2) on 15 shared tasks from TREC, CLEF, and NTCIR 2009–2022. Significant changes to ChatNoir baseline are bold (t-test, $p < 0.05$, Bonferroni-correction).

System	**TREC**									**CLEF**				**NTCIR**	
	Web						**HM**	**DL**		**eH.**	**Touché**		**T2**	**WWW**	
	’09	’10	’11	’12	’13	’14	’19	’19	’20	’16	’20	’21	’22	’17	’18
Query															
ChatNoir	0.33	0.30	0.32	0.16	0.35	0.35	0.45	0.20	0.14	0.16	0.54	0.38	0.00	0.53	0.24
+monoT5	0.39	0.35	0.34	0.21	0.38	**0.42**	0.45	0.22	0.14	0.16	0.54	0.41	0.00	0.52	0.25
+duoT5	0.39	0.35	0.34	0.21	0.39	**0.42**	0.45	0.22	0.14	0.16	0.53	0.41	0.00	0.52	0.25
Description															
ChatNoir	0.16	0.23	0.25	0.24	0.24	0.33	0.38	—	—	—	0.00	0.00	0.00	—	0.07
+monoT5	0.17	0.27	0.24	0.27	0.23	0.33	0.37	—	—	—	0.00	0.00	0.00	—	0.07
+duoT5	0.17	0.26	0.25	0.27	0.23	0.33	0.37	—	—	—	0.00	0.00	0.00	—	0.07

4 `chatnoir-pyterrier` for Shared Tasks Participation

As shared tasks involve ever larger and more complex document collections, participation has become increasingly difficult, especially for students. With `chatnoir-pyterrier`, we lower the barrier to participation in shared tasks such as those run at TREC, CLEF, and NTCIR. By using ChatNoir for first-stage retrieval, researchers no longer need to manage or index web crawls but can instead tune their own IR pipeline, e.g., by developing re-rankers. Table 2 evaluates ChatNoir's first-stage retrieval and the pipeline from Listing 2 on 15 shared tasks.

5 Conclusion and Future Work

We have presented `chatnoir-pyterrier`, a Python library that lowers the barriers for web-scale IR research by providing easy retrieval access to large document collections with more than 1.5 billion documents hosted on ChatNoir. By integrating with PyTerrier and `ir_datasets`, we support resource-constrained researchers to run web-scale IR experiments, to participate in a wide range of current and past shared tasks, and to build RAG applications. Further, `chatnoir-pyterrier` follows green IR research practices [36] by encouraging the reuse of existing indices and the sharing of computational resources. We plan to extend `chatnoir-pyterrier` by indexing additional collections through `ir_datasets`, ultimately aiming for full coverage of all shared tasks at TREC, CLEF, and NTCIR [7]. By making web-scale IR experiments more accessible, `chatnoir-pyterrier` will broaden participation and foster collaboration, to advance IR research.

Acknowledgements This publication has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101070014 (OpenWebSearch.EU, `https://doi.org/10.3030/101070014`).

References

1. Abualsaud, M., Lioma, C., Maistro, M., Smucker, M.D., Zuccon, G.: Overview of the trec 2019 decision track. In: TREC (2019)
2. Bevendorff, J., Stein, B., Hagen, M., Potthast, M.: Elastic chatnoir: Search engine for the clueweb and the common crawl. In: Pasi, G., Piwowarski, B., Azzopardi, L., Hanbury, A. (eds.) Advances in Information Retrieval - 40th European Conference on IR Research, ECIR 2018, Grenoble, France, March 26-29, 2018, Proceedings. Lecture Notes in Computer Science, vol. 10772, pp. 820–824. Springer (2018). https://doi.org/10.1007/978-3-319-76941-7_83, https://doi.org/10.1007/978-3-319-76941-7_83
3. Bondarenko, A., Fröbe, M., Beloucif, M., Gienapp, L., Ajjour, Y., Panchenko, A., Biemann, C., Stein, B., Wachsmuth, H., Potthast, M., Hagen, M.: Overview of Touché 2020: Argument Retrieval. In: Arampatzis, A., Kanoulas, E., Tsikrika, T., Vrochidis, S., Joho, H., Lioma, C., Eickhoff, C., Névéol, A., Cappellato, L., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. 11th International Conference of the CLEF Association (CLEF 2020). Lecture Notes in Computer Science, vol. 12260, pp. 384–395. Springer, Berlin Heidelberg New York (Sep 2020). https://doi.org/10.1007/978-3-030-58219-7_26
4. Bondarenko, A., Fröbe, M., Kiesel, J., Schlatt, F., Barriere, V., Ravenet, B., Hemamou, L., Luck, S., Reimer, J.H., Stein, B., Potthast, M., Hagen, M.: Overview of Touché 2023: Argument and Causal Retrieval. In: Arampatzis, A., Kanoulas, E., Tsikrika, T., Vrochidis, S., Giachanou, A., Li, D., Aliannejadi, M., Vlachos, M., Faggioli, G., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. 14th International Conference of the CLEF Association (CLEF 2023). Lecture Notes in Computer Science, vol. 14163, pp. 507–530. Springer, Berlin Heidelberg New York (Sep 2023). https://doi.org/10.1007/978-3-031-42448-9_31
5. Bondarenko, A., Fröbe, M., Kiesel, J., Syed, S., Gurcke, T., Beloucif, M., Panchenko, A., Biemann, C., Stein, B., Wachsmuth, H., Potthast, M., Hagen, M.: Overview of Touché 2022: Argument Retrieval. In: Barrón-Cedeño, A., Da San Martino, G., Esposti, M.D., Sebastiani, F., Macdonald, C., Pasi, G., Hanbury, A., Potthast, M., Faggioli, G., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. 13th International Conference of the CLEF Association (CLEF 2022). Lecture Notes in Computer Science, vol. 13390. Springer, Berlin Heidelberg New York (Sep 2022). https://doi.org/10.1007/978-3-031-13643-6_21
6. Bondarenko, A., Gienapp, L., Fröbe, M., Beloucif, M., Ajjour, Y., Panchenko, A., Biemann, C., Stein, B., Wachsmuth, H., Potthast, M., Hagen, M.: Overview of Touché 2021: Argument Retrieval. In: Candan, K., Ionescu, B., Goeuriot, L., Müller, H., Joly, A., Maistro, M., Piroi, F., Faggioli, G., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. 12th International Conference of the CLEF Association (CLEF 2021). Lecture Notes in Computer Science, vol. 12880, pp. 450–467. Springer, Berlin Heidelberg New York (Sep 2021). https://doi.org/10.1007/978-3-030-85251-1_28
7. Breuer, T., Voorhees, E.M., Soboroff, I.: Browsing and searching metadata of TREC. In: Yang, G.H., Wang, H., Han, S., Hauff, C., Zuccon, G., Zhang, Y. (eds.) Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024. pp. 313–323. ACM (2024).

https://doi.org/10.1145/3626772.3657873,
https://doi.org/10.1145/3626772.3657873
8. Carterette, B., Pavlu, V., Fang, H., Kanoulas, E.: Million query track 2009 overview. In: Voorhees, E.M., Buckland, L.P. (eds.) Proceedings of The Eighteenth Text REtrieval Conference, TREC 2009, Gaithersburg, Maryland, USA, November 17-20, 2009. NIST Special Publication, vol. 500-278. National Institute of Standards and Technology (NIST) (2009), http://trec.nist.gov/pubs/trec18/papers/MQ09OVERVIEW.pdf
9. Clarke, C.L.A., Craswell, N., Soboroff, I.: Overview of the TREC 2009 web track. In: Voorhees, E.M., Buckland, L.P. (eds.) Proceedings of The Eighteenth Text REtrieval Conference, TREC 2009, Gaithersburg, Maryland, USA, November 17-20, 2009. NIST Special Publication, vol. 500-278. National Institute of Standards and Technology (NIST) (2009), http://trec.nist.gov/pubs/trec18/papers/WEB09.OVERVIEW.pdf
10. Clarke, C.L.A., Craswell, N., Soboroff, I., Cormack, G.V.: Overview of the TREC 2010 web track. In: Voorhees, E.M., Buckland, L.P. (eds.) Proceedings of The Nineteenth Text REtrieval Conference, TREC 2010, Gaithersburg, Maryland, USA, November 16-19, 2010. NIST Special Publication, vol. 500-294. National Institute of Standards and Technology (NIST) (2010), https://trec.nist.gov/pubs/trec19/papers/WEB.OVERVIEW.pdf
11. Clarke, C.L.A., Craswell, N., Soboroff, I., Voorhees, E.M.: Overview of the TREC 2011 web track. In: Voorhees, E.M., Buckland, L.P. (eds.) Proceedings of The Twentieth Text REtrieval Conference, TREC 2011, Gaithersburg, Maryland, USA, November 15-18, 2011. NIST Special Publication, vol. 500-296. National Institute of Standards and Technology (NIST) (2011), http://trec.nist.gov/pubs/trec20/papers/WEB.OVERVIEW.pdf
12. Clarke, C.L.A., Craswell, N., Voorhees, E.M.: Overview of the TREC 2012 web track. In: Voorhees, E.M., Buckland, L.P. (eds.) Proceedings of The Twenty-First Text REtrieval Conference, TREC 2012, Gaithersburg, Maryland, USA, November 6-9, 2012. NIST Special Publication, vol. 500-298. National Institute of Standards and Technology (NIST) (2012), http://trec.nist.gov/pubs/trec21/papers/WEB12.overview.pdf
13. Cleverdon, C.: The Cranfield tests on index language devices. In: ASLIB Proceedings. pp. 173–192. MCB UP Ltd. (Reprinted in Readings in Information Retrieval, Karen Sparck-Jones and Peter Willett, editors, Morgan Kaufmann, 1997) (1967)
14. The clueweb09 dataset (2009), https://lemurproject.org/clueweb09/
15. The clueweb12 dataset (2012), https://lemurproject.org/clueweb12/
16. Collins-Thompson, K., Bennett, P.N., Diaz, F., Clarke, C., Voorhees, E.M.: TREC 2013 web track overview. In: Voorhees, E.M. (ed.) Proceedings of The Twenty-Second Text REtrieval Conference, TREC 2013, Gaithersburg, Maryland, USA, November 19-22, 2013. NIST Special Publication, vol. 500-302. National Institute of Standards and Technology (NIST) (2013), http://trec.nist.gov/pubs/trec22/papers/WEB.OVERVIEW.pdf
17. Collins-Thompson, K., Macdonald, C., Bennett, P.N., Diaz, F., Voorhees, E.M.: TREC 2014 web track overview. In: Voorhees, E.M., Ellis, A. (eds.) Proceedings of The Twenty-Third Text REtrieval Conference, TREC 2014, Gaithersburg, Maryland, USA, November 19-21, 2014. NIST Special Publication, vol. 500-308. National Institute of Standards and Technology (NIST) (2014), http://trec.nist.gov/pubs/trec23/papers/overview-web.pdf

18. Craswell, N., Mitra, B., Yilmaz, E., Campos, D.: Overview of the TREC 2020 deep learning track. In: Voorhees, E.M., Ellis, A. (eds.) Proceedings of the Twenty-Ninth Text REtrieval Conference, TREC 2020, Virtual Event [Gaithersburg, Maryland, USA], November 16-20, 2020. NIST Special Publication, vol. 1266. National Institute of Standards and Technology (NIST) (2020), https://trec.nist.gov/pubs/trec29/papers/OVERVIEW.DL.pdf
19. Craswell, N., Mitra, B., Yilmaz, E., Campos, D., Lin, J.: Overview of the TREC 2021 deep learning track. In: Soboroff, I., Ellis, A. (eds.) Proceedings of the Thirtieth Text REtrieval Conference, TREC 2021, online, November 15-19, 2021. NIST Special Publication, vol. 500-335. National Institute of Standards and Technology (NIST) (2021), https://trec.nist.gov/pubs/trec30/papers/Overview-DL.pdf
20. Craswell, N., Mitra, B., Yilmaz, E., Campos, D., Lin, J., Voorhees, E.M., Soboroff, I.: Overview of the TREC 2022 deep learning track. In: Soboroff, I., Ellis, A. (eds.) Proceedings of the Thirty-First Text REtrieval Conference, TREC 2022, online, November 15-19, 2022. NIST Special Publication, vol. 500-338. National Institute of Standards and Technology (NIST) (2022), https://trec.nist.gov/pubs/trec31/papers/Overview_deep.pdf
21. Craswell, N., Mitra, B., Yilmaz, E., Campos, D., Voorhees, E.M.: Overview of the TREC 2019 deep learning track. In: Voorhees, E.M., Ellis, A. (eds.) Proceedings of the Twenty-Eighth Text REtrieval Conference, TREC 2019. NIST Special Publication, vol. 500-331. National Institute of Standards and Technology (NIST) (2019), https://trec.nist.gov/pubs/trec28/papers/OVERVIEW.DL.pdf
22. Craswell, N., Mitra, B., Yilmaz, E., Rahmani, H.A., Campos, D., Lin, J., Voorhees, E.M., Soboroff, I.: Overview of the TREC 2023 deep learning track. In: Soboroff, I., Ellis, A. (eds.) Proceedings of the Thirty-Second Text REtrieval Conference, TREC 2023. NIST Special Publication, National Institute of Standards and Technology (NIST) (2023), https://trec.nist.gov/pubs/trec32/papers/Overview_deep.pdf
23. Lewis, P.S.H., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., Kiela, D.: Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual (2020), https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html
24. Luo, C., Sakai, T., Liu, Y., Dou, Z., Xiong, C., Xu, J.: Overview of the NTCIR-13 we want web task. In: The 13th NTCIR Conference, Evaluation of Information Access Technologies, National Center of Sciences, Tokyo, Japan, December 5-8, 2017. National Institute of Informatics (NII) (2017), https://research.nii.ac.jp/ntcir/workshop/OnlineProceedings13/pdf/ntcir/01-NTCIR13-OV-WWW-LuoC.pdf
25. MacAvaney, S., Yates, A., Feldman, S., Downey, D., Cohan, A., Goharian, N.: Simplified data wrangling with ir_datasets. In: Diaz, F., Shah, C., Suel, T., Castells, P., Jones, R., Sakai, T. (eds.) SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021. pp. 2429–2436. ACM (2021). https://doi.org/10.1145/3404835.3463254, https://doi.org/10.1145/3404835.3463254

26. Macdonald, C., Tonellotto, N., MacAvaney, S., Ounis, I.: Pyterrier: Declarative experimentation in python from BM25 to dense retrieval. In: Demartini, G., Zuccon, G., Culpepper, J.S., Huang, Z., Tong, H. (eds.) CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021. pp. 4526–4533. ACM (2021). https://doi.org/10.1145/3459637.3482013
27. Mao, J., Sakai, T., Luo, C., Xiao, P., Liu, Y., Dou, Z.: Overview of the ntcir-14 we want web task. In: Proceedings of NTCIR 2018 (2018)
28. Nguyen, T., Rosenberg, M., Song, X., Gao, J., Tiwary, S., Majumder, R., Deng, L.: MS MARCO: A human generated machine reading comprehension dataset. In: Besold, T.R., Bordes, A., d'Avila Garcez, A.S., Wayne, G. (eds.) Proceedings of the Workshop on Cognitive Computation: Integrating neural and symbolic approaches 2016 co-located with the 30th Annual Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, December 9, 2016. CEUR Workshop Proceedings, vol. 1773. CEUR-WS.org (2016), https://ceur-ws.org/Vol-1773/CoCoNIPS_2016_paper9.pdf
29. Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C., Johnson, D.: Terrier information retrieval platform. In: Losada, D.E., Fernández-Luna, J.M. (eds.) Advances in Information Retrieval, 27th European Conference on IR Research, ECIR 2005, Santiago de Compostela, Spain, March 21-23, 2005, Proceedings. Lecture Notes in Computer Science, vol. 3408, pp. 517–519. Springer (2005). https://doi.org/10.1007/978-3-540-31865-1_37, https://doi.org/10.1007/978-3-540-31865-1_37
30. Overwijk, A., Xiong, C., Callan, J.: Clueweb22: 10 billion web documents with rich information. In: Amigó, E., Castells, P., Gonzalo, J., Carterette, B., Culpepper, J.S., Kazai, G. (eds.) SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022. pp. 3360–3362. ACM (2022). https://doi.org/10.1145/3477495.3536321, https://doi.org/10.1145/3477495.3536321
31. Overwijk, A., Xiong, C., Liu, X., VandenBerg, C., Callan, J.: Clueweb22: 10 billion web documents with rich information. CoRR **abs/2211.15848** (2022). https://doi.org/10.48550/ARXIV.2211.15848, https://doi.org/10.48550/arXiv.2211.15848
32. Palotti, J.R.M., Zuccon, G., Jimmy, Pecina, P., Lupu, M., Goeuriot, L., Kelly, L., Hanbury, A.: CLEF 2017 task overview: The IR task at the ehealth evaluation lab - evaluating retrieval methods for consumer health search. In: Cappellato, L., Ferro, N., Goeuriot, L., Mandl, T. (eds.) Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017. CEUR Workshop Proceedings, vol. 1866. CEUR-WS.org (2017), https://ceur-ws.org/Vol-1866/invited_paper_16.pdf
33. Potthast, M., Hagen, M., Stein, B., Graßegger, J., Michel, M., Tippmann, M., Welsch, C.: Chatnoir: a search engine for the clueweb09 corpus. In: Hersh, W.R., Callan, J., Maarek, Y., Sanderson, M. (eds.) The 35th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '12, Portland, OR, USA, August 12-16, 2012. p. 1004. ACM (2012). https://doi.org/10.1145/2348283.2348429, https://doi.org/10.1145/2348283.2348429

34. Pradeep, R., Nogueira, R.F., Lin, J.: The expando-mono-duo design pattern for text ranking with pretrained sequence-to-sequence models. CoRR **abs/2101.05667** (2021), https://arxiv.org/abs/2101.05667
35. Sakai, T., Tao, S., Zeng, Z., Zheng, Y., Mao, J., Chu, Z., Liu, Y., Dou, Z., Soboroff, I.: Overview of the ntcir-15 we want web with centre (www-3) task. In: Proceedings of NTCIR 2020 (2020), https://api.semanticscholar.org/CorpusID:229384015
36. Scells, H., Zhuang, S., Zuccon, G.: Reduce, reuse, recycle: Green information retrieval research. In: Amigó, E., Castells, P., Gonzalo, J., Carterette, B., Culpepper, J.S., Kazai, G. (eds.) SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022. pp. 2825–2837. ACM (2022). https://doi.org/10.1145/3477495.3531766, https://doi.org/10.1145/3477495.3531766
37. Zuccon, G., Palotti, J.R.M., Goeuriot, L., Kelly, L., Lupu, M., Pecina, P., Müller, H., Budaher, J., Deacon, A.: The IR task at the CLEF ehealth evaluation lab 2016: User-centred health information retrieval. In: Balog, K., Cappellato, L., Ferro, N., Macdonald, C. (eds.) Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum, Évora, Portugal, 5-8 September, 2016. CEUR Workshop Proceedings, vol. 1609, pp. 15–27. CEUR-WS.org (2016), https://ceur-ws.org/Vol-1609/16090015.pdf