

The Anserini Docker Image for OSIRRC at SIGIR 2019

Ryan Clancy and Jimmy Lin

David R. Cheriton School of Computer Science
University of Waterloo

Image Source: <https://github.com/osirrc/anserini-docker>

Docker Hub: <https://hub.docker.com/r/osirrc2019/anserini>

Commentary: My suggestion for these “docker papers” is to provide the URLs to the implementations and leave technical details there (i.e., don’t repeat stuff that’s better captured in the README of the source image repo).

1 OVERVIEW

Anserini [4, 5] is an open-source information retrieval toolkit built around Lucene to facilitate replicable research. The project grew out of the Open-Source IR Reproducibility Challenge from 2015 [3] and reflects growing community interest in using Lucene for academic IR research [1, 2]. As Lucene was not originally designed as a research toolkit, Anserini aims to fill in “the missing parts” that allow researchers to run standard *ad hoc* retrieval experiments “right out of the box”, including competitive baselines and integration hooks for neural ranking models. Given Lucene’s tremendous production deployment base (typically via Solr or Elasticsearch), better alignment between research in information retrieval and the practice of building research world search engines promises a smoother transition path from the lab to the “real world” for research innovations.

Commentary: Here, my suggestion is to provide an overview of the underlying system, techniques, etc.

2 TECHNICAL DESIGN

Supported Collections:

robust4, core17, core18, gov2, cw09b, cw12b

Supported Hooks:

init, index, search

As with Lucene itself, Anserini is implemented in Java. Thus, the Anserini Docker image primarily consisted of lightweight hooks implemented in Python that wraps underlying Java commands.

Commentary: Here, my suggestion is to provide an overview of the technical design of the image, without going into too much detail.

3 RETRIEVAL MODELS

The Anserini Docker image

Commentary: Here, my suggestion is to provide an overview of the retrieval models...

4 OSIRRC EXPERIENCE

Anserini was the first image that was created in conjunction with the jig...

Commentary: Like, how was the experience?

REFERENCES

- [1] L. Azzopardi, M. Crane, H. Fang, G. Ingersoll, J. Lin, Y. Moshfeghi, H. Scells, P. Yang, and G. Zuccon. 2017. The Lucene for Information Access and Retrieval Research (LIARR) Workshop at SIGIR 2017. In *SIGIR*. 1429–1430.
- [2] L. Azzopardi, Y. Moshfeghi, M. Halvey, R. Alkhawaldeh, K. Balog, E. Di Buccio, D. Ceccarelli, J. Fernández-Luna, C. Hull, J. Mannix, and S. Palchowdhury. 2017. Lucene4IR: Developing Information Retrieval Evaluation Resources Using Lucene. *SIGIR Forum* 50, 2 (2017), 58–75.
- [3] J. Lin, M. Crane, A. Trotman, J. Callan, I. Chattopadhyaya, J. Foley, G. Ingersoll, C. Macdonald, and S. Vigna. 2016. Toward Reproducible Baselines: The Open-Source IR Reproducibility Challenge. In *ECIR*. 408–420.
- [4] P. Yang, H. Fang, and J. Lin. 2017. Anserini: Enabling the Use of Lucene for Information Retrieval Research. In *SIGIR*. 1253–1256.
- [5] P. Yang, H. Fang, and J. Lin. 2018. Anserini: Reproducible Ranking Baselines Using Lucene. *JDIQ* 10, 4 (2018), Article 16.