

Overview of LiLAS 2020 – Living Labs for Academic Search Workshop Lab

Extended Abstract*

Philipp Schaer¹[0000–0002–8817–4632], Johann Schaible²[0000–0002–5441–7640], and Leyla Jael Garcia Castro³[0000–0003–3986–0510]

¹ TH Köln - University of Applied Sciences, Germany
philipp.schaer@th-koeln.de

² GESIS - Leibniz Institute for the Social Sciences, Germany
johann.schaible@gesis.org

³ ZB MED - Information Centre for Life Sciences, Germany
ljgarcia@zbmed.de

1 Introduction and Background

In our previous work [7] and [8], we described the motivation and the outline for a new CLEF evaluation lab. The Living Labs for Academic Search (LiLAS) lab fosters the discussion, research, and evaluation of academic search systems by applying the concept of living labs to the domain of academic search [9]. This extended abstract summarizes the main ideas and contributions discussed in these previous articles.

The timeless challenge of Academic Search that the field of Information Retrieval has been dealing with for many years is more relevant than ever. With the rise of the COVID-19 pandemic it gained new momentum in the IR community with Initiatives like TREC COVID. However, test collections and specialized data sets like CORD-19 only allow for system-oriented experiments, while the evaluation of algorithms in real-world environments is only available to researchers from industry. In LiLAS, we open up two academic search platforms to allow participating researchers to evaluate their systems in a Docker-based research environment.

The need for innovation in academic search is shown by the stagnating system performance in controlled evaluation campaigns, as demonstrated in TREC and CLEF meta-evaluation studies [10,1]. User studies in real systems of scientific information and digital libraries show similar conditions. Although massive data collections of scientific documents are available in platforms like arXiv, PubMed, or other digital libraries, central user needs and requirements remain unsatisfied. The central mission is to find both relevant and high-quality documents - if possible, directly on the first result page. Besides this ad-hoc retrieval problem,

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2020, 22-25 September 2020, Thessaloniki, Greece.

* Extended abstract of the paper originally published in [7]

other tasks such as the recommendation of relevant cross-modality content including research data sets or specialized tasks like expert finding are not even considered here. On top of that, relevance in academic search is multi-layered [4] and a topic that drives research communities like the Bibliometrics-enhanced Information Retrieval (BIR) workshops [5]. CLEF and TREC hosted the Living Labs for Information Retrieval (LL4IR) and Open Search (TREC-OS) initiatives [2] that are the predecessors of LiLAS.

The goal of LiLAS is to expand the knowledge on improving the search for academic resources such as literature, research data, and the interlinking between these resources. LiLAS cooperates with two academic search systems providers from Life Sciences and Social Sciences. Both system providers support LiLAS by allowing participants of the lab to employ experimental search components into their production online systems. We will have access to the click logs of these systems and use them to employ A/B tests or more complex interleaving experiments. Our living lab platform STELLA makes this possible by bringing platform operators and researchers together and providing a methodological and technical framework for online experiments [3].

2 Evaluation Infrastructure

We use STELLA as our living lab evaluation infrastructure. STELLA is aiming to make it easier to evaluate academic retrieval information and recommendation systems [3]. Figure 1 shows an overview of how the steps flow from a researcher’s or developer’s idea to the evaluation feedback so the changes can be tuned and improved. It all starts with an idea, for instance adding synonyms to the keywords used by an end-user when searching for information. Developers will work on a modified version of the production system, including that change they want to analyze. Whenever an end-user goes to the system, everything will look as usual. Once the search keywords are introduced, STELLA will show end-users some results from the experimental system and some results from the regular production system. End-users will continue their regular interaction with the system. Based on the retrieved documents and the following interaction, STELLA will create an evaluation profile together with some statistics. Researchers and developers will then analyze STELLA’s feedback and will react accordingly to get the usage level they are aiming at.

STELLA’s infrastructure relies on the container virtualization environment Docker [6], making it easier for STELLA to run multiple experimental systems, i.e., a multi-container environment, and compare them to each other and the production system as well. The core component in STELLA is a central Application Public Interface (API) connecting data and content providers with experimental systems, a.k.a. participant systems or participants, encapsulated as Docker containers. Further information can be found at the project website⁴, including some technical details via a series of blogs published regularly.

⁴ <https://stella-project.org/>

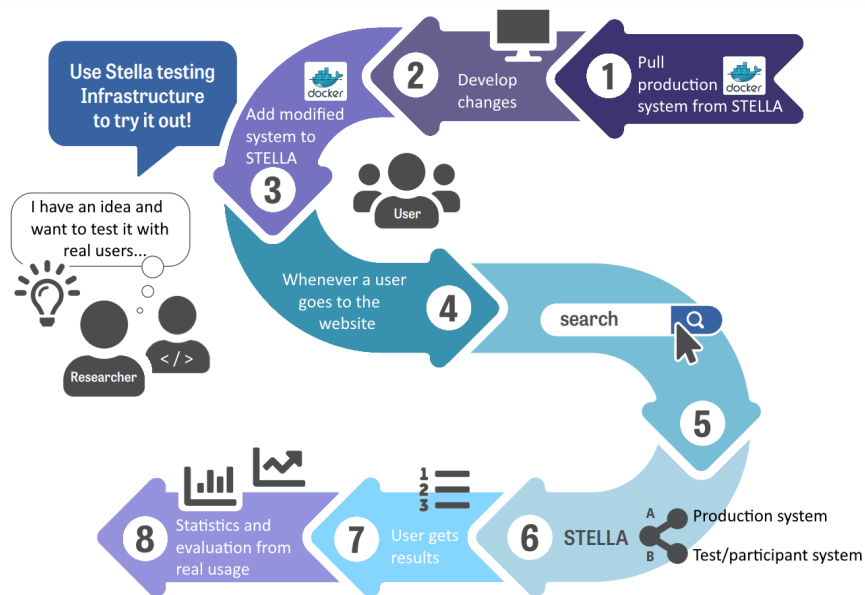


Fig. 1. STELLA workflow, an online living lab supporting testing from ideas to evaluation: Participants package their systems with the help of Docker containers that are deployed in the backend of academic information retrieval and recommendation systems. Users interact directly with the system, with a percentage diverted to the experimental features. Researchers and developers retrieve results and deliver feedback to tune and improve changes.

3 Conclusion and Outlook

Currently, STELLA supports two main tasks: *ad-hoc retrieval* and *recommendation*, both of them used within LiLAS. The two academic search systems LIVIVO and GESIS Search are from the two disjoint scientific domains life sciences and social sciences and include different metadata on research articles, data sets, and many other entities. For the next CLEF lab in 2021 we will focus on (1) ad-hoc retrieval for life science documents and (2) research data recommendations on social science topics. These tasks allow us to use the different data types available in the platforms and offer participants the unique opportunity for their solutions to be tested in real-time environments.

Acknowledgements

This work was partially funded by the German Research Foundation (DFG) under the project no. 407518790.

References

1. Armstrong, T.G., Moffat, A., Webber, W., Zobel, J.: Improvements that don't add up: ad-hoc retrieval results since 1998. In: Proceeding of the 18th ACM conference on information and knowledge management. pp. 601–610. CIKM '09, ACM, Hong Kong, China (2009). <https://doi.org/10.1145/1645953.1646031>
2. Balog, K., Schuth, A., Dekker, P., Tavakolpoursaleh, N., Schaer, P., Chuang, P.Y.: Overview of the TREC 2016 Open Search track. In: TREC. vol. Special Publication 500-321. National Institute of Standards and Technology (NIST) (2016)
3. Breuer, T., Schaer, P., Tavakolpoursaleh, N., Schaible, J., Wolff, B., Müller, B.: Stella: Towards a framework for the reproducibility of online search experiments. In: Proceedings of the The Open-Source IR Replicability Challenge (OSIRRC) @ SIGIR (2019)
4. Carevic, Z., Schaer, P.: On the connection between citation-based and topical relevance ranking: Results of a pretest using isearch. In: Proceedings of the First Workshop on Bibliometric-enhanced Information Retrieval co-located with 36th European Conference on Information Retrieval (ECIR 2014), Amsterdam, The Netherlands, April 13, 2014. CEUR Workshop Proceedings, vol. 1143, pp. 37–44. CEUR-WS.org (2014), <http://ceur-ws.org/Vol-1143/paper5.pdf>
5. Mayr, P., Scharnhorst, A., Larsen, B., Schaer, P., Mutschke, P.: Bibliometric-enhanced information retrieval. In: Advances in Information Retrieval - 36th European Conference on IR Research, ECIR 2014, Amsterdam, The Netherlands, April 13-16, 2014. Proceedings. Lecture Notes in Computer Science, vol. 8416, pp. 798–801. Springer (2014). https://doi.org/10.1007/978-3-319-06028-6_99
6. Merkel, D.: Docker: lightweight Linux containers for consistent development and deployment. *Linux Journal* **2014**(239), 2:2 (Mar 2014)
7. Schaer, P., Schaible, J., Garcia Castro, L.J.: Overview of LiLAS 2020 – Living Labs for Academic Search. In: Arampatzis, A., Kanoulas, E., Tsirikika, T., Vrochidis, S., Joho, H., Lioma, C., Eickhoff, C., Cappellato, L., Névéol, A., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction Proceedings of the Eleventh International Conference of the CLEF Association (CLEF 2020). Lecture Notes in Computer Science, vol. 12260 (2020)
8. Schaer, P., Schaible, J., Müller, B.: Living Labs for Academic Search at CLEF 2020. In: Jose, J.M., Yilmaz, E., Magalhães, J., Castells, P., Ferro, N., Silva, M.J., Martins, F. (eds.) Advances in Information Retrieval. Lecture Notes in Computer Science, vol. 12036, pp. 580–586. Springer International Publishing, Cham (2020)
9. Schaible, J., Breuer, T., Tavakolpoursaleh, N., Müller, B., Wolff, B., Schaer, P.: Evaluation infrastructures for academic shared tasks. *Datenbank-Spektrum* **20**(1), 29–36 (Mar 2020). <https://doi.org/10.1007/s13222-020-00335-x>
10. Yang, W., Lu, K., Yang, P., Lin, J.: Critically Examining the "Neural Hype": Weak Baselines and the Additivity of Effectiveness Gains from Neural Ranking Models. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR'19. pp. 1129–1132. ACM Press, Paris, France (2019). <https://doi.org/10.1145/3331184.3331340>