Contribution ID: **3** Type: **not specified**

The Archive Query Log: Mining Millions of Search Result Pages of Hundreds of Search Engines from 25 Years of Web Archives

The Archive Query Log (AQL) is a previously unused, comprehensive query log collected at the Internet Archive over the last 25 years. Its first version includes 356 million queries, 166 million search result pages, and 1.7 billion search results across 550 search providers. Although many query logs have been studied in the literature, the search providers that own them generally do not publish their logs to protect user privacy and vital business data. Of the few query logs publicly available, none combines size, scope, and diversity. The AQL is the first to do so, enabling research on new retrieval models and (diachronic) search engine analyses. Provided in a privacy-preserving manner, it promotes open research as well as more transparency and accountability in the search industry.

Primary authors: REIMER, Jan Heinrich (Friedrich-Schiller-Universität Jena); SCHMIDT, Sebastian (Leipzig University); FRÖBE, Maik (Friedrich-Schiller-Universität Jena); GIENAPP, Lukas (Leipzig University); SCELLS, Harrisen (Leipzig University); STEIN, Benno (Bauhaus-Universität Weimar); HAGEN, Matthias (Friedrich-Schiller-Universität Jena); POTTHAST, Martin (Leipzig University and ScaDS.AI)

Presenter: REIMER, Jan Heinrich (Friedrich-Schiller-Universität Jena)

Session Classification: Poster session