

Master Thesis Proposal

Prepared for: Prof Johannes Schoning Dr. Anusuriya Devaraju, Uwe Scindler

Prepared by: Ahmed AlShafeei

Title

The use of log analysis to investigate user search behaviour on PANGAEA data portal & reporting usage metrics to Datacite counter service

Objective of the proposal

The aim of this proposal is to use the current analytics provided by Dr Devaraju and improve it to cover new analytics questions related to data search. The current analytics cover the usage of meta data (eg. author name, event) and the top filter usage on both external and internal data search. My thesis would seek to cover more areas like visiting path, path traversed, page last visited, and success rate. These statistical information related to web log analysis have proved beneficial over researches as it allows the exploration of quality and performance of the PANGAEA data portal. The second part of this thesis will be working on sending out the metrics of the PANGAEA datasets usage information to Datacite counter service, as this would allow the PANGAEA to become a more reliable trustworthy source for different usages. The key challenge in this part is to identify the necessary metrics needed according to Datacite and working with their documentation to reach a reliable result that would send the data for the PANGAEA references.

Expected Results

The outcome of this thesis would be to provide more depth in log analysis related to data search and the sending of PANGAEA datasets usage information to Datacite counter service.

References

- Agarwal, Sonali, and Bakshi Rohit Prasad. "High speed streaming data analysis of web generated log streams." *2015 IEEE 10th International Conference on Industrial and Information Systems (ICIIS)*. IEEE, 2015.
- Zhang, Ying, Bernard J. Jansen, and Amanda Spink. "Time series analysis of a Web search engine transaction log." *Information Processing & Management* 45.2 (2009): 230-245.