Geophysical Research Abstracts
Vol. 17, EGU2015-8512, 2015
EGU General Assembly 2015

Mining the Geophysical Research Abstracts Corpus: Mapping the impact of Free and Open Source Software on the EGU Divisions

Peter Löwe (1), Jens Klump (2), and Jesse Robertson (2)

(1) German National Library of Science and Technology (TIB), Hannover, Germany , (2) CSIRO Australian Resources Research Centre, Kensington, Australia

Text mining is commonly employed as a tool in data science to investigate and chart emergent information from corpora of research abstracts, such as the Geophysical Research Abstracts (GRA) published by Copernicus.

In this context current standards, such as persistent identifiers like DOI and ORCID, allow us to trace, cite and map links between journal publications, the underlying research data and scientific software. This network can be expressed as a directed graph which enables us to chart networks of cooperation and innovation, thematic foci and the locations of research communities in time and space. However, this approach of data science, focusing on the research process in a self-referential manner, rather than the topical work, is still in a developing stage.

Scientific work presented at the EGU General Assembly is often the first step towards new approaches and innovative ideas to the geospatial community. It represents a rich, deep and heterogeneous source of geoscientific thought. This corpus is a significant data source for data science, which has not been analysed on this scale previously.

In this work, the corpus of the Geophysical Research Abstracts is used for the first time as a data base for analyses of topical text mining. For this, we used a sturdy and customizable software framework, based on the work of Schmitt et al. [1]. For the analysis we used the High Performance Computing infrastructure of the German Research Centre for Geosciences GFZ in Potsdam, Germany.

Here, we report on the first results from the analysis of the continuous spreading the of use of Free and Open Source Software Tools (FOSS) within the EGU communities, mapping the general increase of FOSS-themed GRA articles in the last decade and the developing spatial patterns of involved parties and FOSS topics.

References:
[1] Schmitt, L. M., Christianson, K.T, Gupta R..: Linguistic Computing with UNIX Tools, in Kao, A., Poteet S.R. (Eds.): Natural Language processing and Text Mining, Springer, 2007. doi:10.1007/978-1-84628-754-1_12.