



## **Applications of Latent Semantic Analysis to Track Trends in Geoscience Research**

Siri Jodha Khalsa, Luis Lopez, and Allen Pope  
CIRES, Univ. Colorado, Boulder, United States (sjsk@nsidc.org)

Organizers of large scientific conferences are faced with the challenge of selecting topics around which to organize sessions. Topics need to be specific enough to attract submissions while covering in aggregate the full range of activities within the discipline. The program should also aim to foster exchange of ideas across disciplinary boundaries. Finally, the program needs to evolve year to year as new topics emerge.

Latent Semantic Analysis (LSA) is a well-established technique in natural language processing and information retrieval that can help with these problem. We describe the application of Latent Dirichlet Allocation (LDA), a generative probabilistic form of LSA, to discover topics and trends in scientific abstracts submitted to the European Geophysical Union (EGU) General Assemblies between 2009 and 2018, and to abstracts submitted to the Arctic Science Summit Week in 2018 (ASSW'18). We compare topics derived from EGU abstract content against the sessions those abstracts were assigned to, and assess the degree of mismatch. We also evaluate the appropriateness of co-organized sessions. Topics derived from ASSW'18 abstracts are analyzed to identify interdisciplinary themes around which to offer sessions at the ASSW meeting in 2020.

This study developed Jupyter notebooks utilizing open source software routines for text preparation, analysis and visualization. These are being made available to the community for adaptation to any other document corpus.