

CALL FOR THE WORKSHOP

“CORPUS IN GISCIENCE: GOING BEYOND BUTTERFLY COLLECTING”

With the increasing availability of unstructured and semistructured text data in the form of user-generated content and digitized corpora, GIScience is recently actively exploring the potential of this data to answer some of the traditional as well as new questions. Interestingly, the word “corpus” is only used seldom in this research – “collection” is the preferred term. As the range of questions addressed through various corpora expands, it is high time for the GIScience community to gain an overview of emerging work as well as reflect upon some important corpus-related questions: Aren't we doing “butterfly collecting”¹ at best and being opportunistic at worst? Do concepts such as “language in use” and “representativeness” matter or should we take a pragmatic approach and only question, “Is corpus A good for task B”?

The aim of this workshop is to provide an overview of existing (types of) corpora, methods and research questions addressed through text corpora in GIScience, as well as to discuss the importance of aspects such as corpus characteristics or representativeness.

The scope of the workshop includes the following:

- characteristics of existing geospatial and general corpora (including Web as a corpus, digitized corpora)
- corpus-building strategies and frameworks
- making a corpus publicly available: tools and pitfalls
- methods for spatial and thematic exploration of a corpus (geographic information retrieval and beyond)
- development of spatial markup languages
- approaches to annotation of large corpora (including crowdsourcing)
- areas of application of corpus-based and –driven research in GIScience: from environmental monitoring to the investigation of variation in the use of spatial language

Submissions:

We invite contributions discussing (early) ongoing work relevant to one or more of the topics listed above. Submissions should be between 6 and 8 pages in length and should follow the GIScience 2018 formatting guidelines. All submissions will be peer-reviewed by members of the program committee, accepted papers will be included in the GIScience digital conference proceedings. The deadline for the submissions is **April 28, 2018**, review decisions will be provided to authors on **May 28, 2018**.

¹ Chomsky's famous critique of corpus linguistics (Chomsky, N. 1979. *Language and Responsibility: Based on Conversations with Mitsou Ronat*. New York: Pantheon. Translated by John Viertel, p.57)

For further questions, please contact the organizers: Ekaterina Egorova (ekaterina.egorova@geo.uzh.ch), Kristin Stock (K.Stock@massey.ac.nz), Lesley Stirling (lesleyfs@unimelb.edu.au)