



Contribution ID: 95

Type: **Sectional reports**

## Transformer-based Model for the Semantic Parsing of Error Messages in Distributed Computing Systems in High Energy Physics

*Thursday, 8 July 2021 17:00 (15 minutes)*

Large-scale computing centers supporting modern scientific experiments store and analyze vast amounts of data. A noticeable number of computing jobs executed within the complex distributed computing environments ends with errors of some kind, and the amount of error log data generated every day complicates manual analysis by human experts. Moreover, traditional methods such as specifying regular expression patterns to automatically group error messages become impractical in a heterogeneous computing environment without a well-defined structure of error messages. ClusterLogs framework for error message clustering was developed to address this challenge. The framework can discover common patterns in error messages from various sources and group them together. One of the essential results of this process is the clear automated description of the resulting clusters, which will be used for the analysis.

In this research, we propose that interpreting error messages as a natural language allows us to use transformer-based deep learning models such as BERT for this task. A model for extracting the relevant part of messages was trained and integrated into ClusterLogs to represent each cluster as a few actionable items, ensuring better interpretation and validation of the results of clustering.

### Summary

**Primary authors:** GRIN, Dmitry; Dr GRIGORIEVA, Maria (Moscow State University)

**Presenter:** GRIN, Dmitry

**Session Classification:** Research infrastructure

**Track Classification:** 2. Research infrastructure