Enhancements in Functionality of the Interactive Visual Explorer for ATLAS Computing Metadata

Interactive Visual Explorer (InVEx) is a web application for the exploration of big volumes of multidimensional data. ATLAS computing metadata has become the research ground for this application.

Basic features:
- Machine learning methods for data analysis (clustering algorithms – Kmeans, DBSCAN)
- Interactive 3D visualization models of data
- Visualization of the results of clustering

Enhanced functionality:
- Dataset Info Panel: data sample features representation by statistical measurement types: numerical, ordinal, nominal, range non-categorical (string data, that can’t be treated as categorical)
- Added new clustering algorithms:
  - MiniBatchKMeans/sklearn
  - Hierarchical/sklearn
  - K-Prototypes/sklearn
- Implemented the Level-of-Detail Generator, providing grouping of an initial large data sample into clusters/groups to reduce the amount of data presented to the user simultaneously:
  - MiniBatchKMeans Clusterization method
  - Group by nominal/ordinal parameters
  - Group by numerical continuous parameters
- The LoD method hides the complexity of the initial data, and allows users to use 3D visual scene to select interesting objects/groups and investigate them separately in the new window.
- Storage backend with the ability to store all information about provided operations with data (clustering, grouping, change of the level of detail) and keep all stages of data derivation sequence.
- Interactive parallel coordinates graph with a linked data table for the exploration of data in all dimensions simultaneously.

Visualization of the clustering results

Dataset Info Panel

3D visualization of the grouped data sample (Level-of-Detail Generator)

Figure shows ATLAS BigPanDA jobs metadata for some period of time grouped by nucleus

Any group object can be selected for further investigation in the new window

Comparative analysis of the selected data groups in different dimensions

Parallel Coordinates graph with linked table (exploration of aggregated groups parameters)

Parallel Coordinates allows to explore trends of parameters for each cluster

The project "Visual analytics framework to monitor workflow management systems behavior in exascale era" is supported by RSCF grant №18-71-10003 for 2018-2021 years.