

The purpose of this thesis was to design and implement a hierarchical approach to visualization of the chemical space. Such visualization is a challenging yet important topic used in diverse fields ranging from material engineering to drug design. Especially in drug design, modern methods of high-throughput screening generate large amounts of data that would benefit from hierarchical analysis. One possible approach to hierarchical classification of molecules is a structure based classification based on molecular scaffolds. The scaffolds are widely used by medicinal chemists to group molecules of similar properties. A few scaffold-based hierarchical visualization methods have been proposed. However, to our best knowledge, there exists no tool that would provide a scaffold-based hierarchical visualization of molecular data sets on the background of known chemical space. In this thesis, such tool was created. First, a scaffold tree hierarchy based on ring topologies was designed. Next, this hierarchy was used to analyze frequency of scaffolds extracted from molecules in PubChem Compound database. Subsequently, the PubChem Compound scaffold frequency data was used as a background for visualization of molecular data sets. The visualization is performed by a client-server application implemented as a part of this thesis. It provides an interactive zoomable tree map based visualization of data sets, up to hundreds of thousands molecules large. The application is free to use and has been published under an open source license.