Using the CSD Python API for interactive analytics and data mining of the Cambridge Structural Database.

Paul C. Sanschagrin, Cambridge Crystallographic Data Centre, Piscataway, NJ

Abstract:

The Cambridge Structural Database (CSD) is the world's primary repository for small molecule organic and metal-organic crystallographic structures and currently contains over 875,000 entries. Traditionally, this database has been accessed and analyzed using the desktop software developed by the Cambridge Crystallographic Data Centre (CCDC). As of July 2015, CCDC made available the first version of the CSD Python API to interact with the database and some of the analytic tools in a programmatic fashion. This has enabled researchers to make use of the CSD in new ways and to integrate with other tools and interfaces. An example tool for visual analytics and data mining of the CSD taking advantage of the CSD Python API and publicly available Python tools as well as integration with the Mercury visualizer will be presented.