

Code examples

Document to Structure is a toolkit for extracting chemical structures out of text, HTML and PDF documents. Currently, it recognizes names, SMILES, and InChI. Its API class is `chemaxon.naming.DocumentExtractor`. Below is a list of real life use-cases and code examples that showcase the various ways to use it:

1. **Finding structures in text:**
Uses `DocumentExtractor`'s `processPlainText()` method to process a string.
2. **Finding structures in a live webpage:**
Downloads a live webpage and processes it using `DocumentExtractor`'s `processHTML()` method.
3. **Finding structures in a PDF document:**
Creates a `DocumentExtractor` instance that reads the text from the PDF document.
4. **Highlighting recognized structures in a webpage:**
Finds the recognized names in the HTML code and wraps them with a special element for highlighting.
5. **Saving results in SDF or MRV file:**
Saves the results and related information into a multi-molecule file for use in chemical editors.
6. **Storing results in a JChem structure table:**
Sets up a database connection and stores the hits in a chemical structure database for searching.
7. **Increasing processing speed by multithreading:**
Uses multithreading and breaks HTML pages into fragments.