InchiKey pipeline Manual

Introduction

The inchiKey pipeline was created to quickly turn a large amount of SMILES in to InchiKeys with the help of molconvert. It accepts csv files containing ID,SMILES pairs with these it will then create a new csv file containing all the input data with added inchiKey(s). Since this tool is also used with another pipeline (cfm-id pipeline) the input smiles are also neutralised, so the output will contain depending on if a SMILES is neutral or not an ID,SMILES, InchIKey or an ID, neutralised SMILES, SMILES, neutralised InchIKey and InchIKey. In this guide the installation and usage of this pipeline is explained.

Installation

To install this pipeline and get it working you first download the installation bash script from: https://github.com/NP-Plug-and-Play-Scripts/Bash-scripts this also contains other bashscripts but the one you need is the inchikeyPipelineInstall.sh. Run this script while you are in the desired location for the inchiKeyPipeline workplace. This can be done by changing to the desired location with *"cd /path/to/location/"* followed up with running the bash script with *"sh inchiKeyInstall.sh"*.
This should have created a InchiKey_Workplace folder containing the folders; pythonscripts-master, smileFiles and jchem. The pythonscripts-master folder contains all the python scripts needed for the pipeline to run. The smileFiles folder will be the place to move your SMILES to making it so all the documents are kept in the same location (just for convenience they can be placed on other locations to). Finally the jchem folder is the location to install jchem (again just for convenience so everything is close to each other).

Next download jchem from https://chemaxon.com/products/jchem-engines/download . Log in or create an account if you don't have one yet and download the appropriate version of Jchem base. In the case of linux download the .sh file and install it using *"sh jchem_unix_18.28.sh"* (depends on the version of jchem) and this should start the installer. The installer should be able to guide you through the installation process, just make sure to install it in the jchem folder or on another accessible location.

This should be enough to run the pipeline. In case rdkit is not installed, either try to install it on the server or make a miniconda environment and install rdkit on it. Start the environment and then run the pipeline. https://conda.io/docs/user-guide/tasks/manage-environments.html or http://wiki.ab.wurnet.nl/index.php/How_can_I_install_software_for_my_own_use

Usage

In order to run the pipeline you will first need a .csv file containing ID, SMILES combinations (the id's can be whatever as long as they are unique) and place this in the smileFiles folder.

1. Next start the conda environment in case this is needed.
2. Run the pipeline with the following command *"python /path/to/InchiKey_Workplace/pythonscript-master/InchiKeyCreatorPipeline.py /path/to/InchiKey_Workplace/jchem/jchemsuite/bin/ /path/to/InchiKey_Workplace/smileFiles/ nameOfSmilesFile.csv"*
   This consists of the path to the script, path to the jchemsuite bin folder the path to the smileFiles folder and lastly the name of the file you want to turn in to inchiKeys.
3. Once the pipeline is done running check the smileFiles folder for a csv document called nameOfSmilesFile_full_datafile.txt. This should contain all the SMILES with inchiKey and if the Smiles was not neutral the neutral SMILES the SMILES the neutral InchiKey and InchiKey