Using command-line to label data with aizynthfinder

Installation

```
conda env create -f
https://raw.githubusercontent.com/MolecularAI/aizynthfinder/master/env-users.yml

conda env update -n aizynth-env -f
https://raw.githubusercontent.com/MolecularAI/aizynthfinder/master/env-users.yml

conda activate aizynth-env
```

Generate yml file

```
download_public_data my_folder
```

- `my_folder` is the folder that you want download to. This will creat a `config.yml` file.
- However, to keep the setting the same as the paper, we have to add more information.
  - open your `config.yml` file

  -
    ```
    properties:
      iteration_limit: 200
      return_first: true
      time_limit: 180
      C: 1.4
      cutoff_cumulative: 0.995
      cutoff_number: 50
      max_transforms: 7
    ```

    Add the code above on the top of the yml file, keep other parts unchanged.
  - the final `config.yml` file should be like:

    ```
    properties:
      iteration_limit: 200
      return_first: true
      time_limit: 180
      C: 1.4
      cutoff_cumulative: 0.995
      cutoff_number: 50
      max_transforms: 7
    policy:
      files:
        uspto:
          - D:\aizynthfinder\uspto_model.hdf5
          - D:\aizynthfinder\uspto_templates.hdf5
    filter:
      files:
    ```

```
    uspto: D:\aizynthfinder\uspto_filter_model.hdf5
stock:
  files:
    zinc: D:\aizynthfinder\zinc_stock.hdf5
```

- **Warning: yml files are sensitive to the space bar and cannot replace spaces with indentation**

Generate SMILES txt file

Generate a simple text file with SMILES (one on each row), e.g. smiles.txt

Analysis and generate hdf5 file

```
aizynthcli --config config.yml --smiles smiles.txt
```

Then, we will have the hdf5 output, whose name is "output.hdf5".

Use python to convert hdf5 into table

```
import pandas as pd
data = pd.read_hdf("output.hdf5", "table")
filt_data = data[["target", "is_solved"]]
```

Then, we will get the SMILES and corresponding labels.