

Uniwersytet Jagielloński

Konwerter formatów mmCIF i PDB

Bioinformatyka Strukturalna

Kamil Malisz



2016

Spis Treści:

1. Informacje wstępne.	2
2. Opis Projektu.	2
2.1. Możliwości.	2
3. Funkcje.	2
3.1. Opis main.	3
3.2. Opis fileConversion.	3
3.3. Opis cutPDBFile.	3
3.4. Opis convert.	4
4. Wymagania.	4
4.1. Biblioteki.	4
4.2. Instalacja Bibliotek.	5
5. Obsługa wejścia, wyjścia oraz Obsługa błędów.	5
5.1. Wejście.	5
5.2. Wyjście.	6
6. Przykłady.	6
7. Podsumowanie.	6

1. Informacje Wstępne:

Celem projektu było stworzenie w języku python konwertera plików w formacie PDB do formatu mmCIF oraz odwrotnie. Konwerter miał w zamierzeniu ułatwić pracę z plikami poprzez poprawę kompatybilności z programami i skryptami które obsługują tylko jeden z wyżej wymienionych formatów. Cel został osiągnięty przy pomocy udostępnionego przez „Protein Data Bank Japan”(dalej PDBJ) konwertera online udostępnionego na ich serwerze pod adresem:

<http://mmcif.pdbj.org/converter/index.php?l=en>.

2. Opis Projektu:

Projekt umożliwia użytkownikowi korzystanie z konwertera udostępnionego na PDBJ z poziomu skryptu w języku python. Wybranie z pliku konkretnego łańcucha lub samych nagłówków.

2.1. Możliwości:

Przy pomocy tego skryptu użytkownik ma możliwość dokonania następujących konwersji:

- całego pliku,
- części nagłówkowej pliku,
- części z koordynatami,
- części z koordynatami z zawężeniem do interesujących.

2.2. Ograniczenia:

-przy konwersji pliku z formatu mmCIF do Formatu PDB w przypadku liczby atomów większej niż 99999 indeksy następnych atomów w pliku pdb to 99999.

3. Funkcje:

Projekt w celu utrzymania przejrzystości kodu oraz zmniejszenia redundancji podzielony został na kilka funkcji które zostaną pokrótce opisane poniżej.

3.1. Funkcja `main(argv)`

To główna wywoływana funkcja jej celem jest sprawdzenie poprawności parametrów istnienia wybranych plików oraz jeśli dane podane przez użytkownika są poprawne przekazanie ich do funkcji `fileConversion`. Jako argumenty przyjmuje parametry podane przy wywołaniu

3.2. Funkcja `fileConversion(infile, outfile, mode, log, chain)`

Rozpoznaje tryb konwersji i w zależności od tego jaki tryb został wybrany wywołuje z odpowiednimi parametrami funkcje `cutPDBFile` oraz `convert`. Jako parametry:

- `infile` – ścieżka do pliku wejściowego,
- `outfile` – ścieżka zapisu pliku wyjściowego,
- `mode` – wybrany tryb konwersji,
- `log` – jeśli `True` pobiera plik log z serwera
- `chain` – tekst z literami łańcuchów które chcemy wybrać.

3.3. Funkcja `cutPDBFile(infile, outfile, mode, chain)`

Wykorzystywana jest do przygotowania odpowiednio pliku PDB w celu konwersji gdy nie wybrano trybu konwersji całego pliku. Każda linia pliku musi spełnić określone dla danego trybu wymagania dla trybu 'h' pozostawiane są tylko elementy które pasują do wymienionych w instrukcji warunkowej zamieszczonej poniżej kryteriów:

```
if (line.startswith('HEADER') or line.startswith('OBSLTE') or
line.startswith('TITLE') or line.startswith('SPLIT') or
line.startswith('CAVEAT') or line.startswith('COMPND')
or line.startswith('SOURCE') or line.startswith('KEYWDS') or
```

```
line.startswith('EXPDTA') or line.startswith('NUMMDL') or  
line.startswith('MDLTYP')  
    or line.startswith('AUTHOR') or line.startswith('REVDAT') or  
line.startswith('SPRSDE') or line.startswith('JRNL') or  
line.startswith('REMARK') or line.startswith('END')):
```

Czyli w tym trybie pozostawiamy tylko te elementy które są nagłówkiem natomiast w trybie 'c' pozostawiamy w pliku wszystko to co nie jest nagłówkiem za wyjątkiem sekcji „HEADER”. Dzięki takiemu rozwiązaniu istnieje możliwość utworzenia własnych trybów konwersji lub modyfikacji już istniejących. Parametry funkcji:

- infile – ścieżka do pliku wejściowego,
- outfile – ścieżka zapisu pliku wyjściowego,
- mode – wybrany tryb konwersji,
- chain – tekst z literami łańcuchów które chcemy wybrać.

3.4. Funkcja **convert**(infile, outfile, log)

Służy do wysyłania przygotowanego pliku na serwer oraz do pobierania skonwertowanego pliku oraz logów konwersji. Jako parametry:

- infile – ścieżka do pliku wejściowego,
- outfile – ścieżka zapisu pliku wyjściowego,
- log – jeśli True pobiera plik log z serwera.

4. Wymagania:

Projekt do poprawnego działania wymaga:

- Python w wersji 3.*,
- biblioteki requests,
- biblioteki bs4.

4.1. Biblioteki

Biblioteka requests umozliwia nam wysyłanie zapytań na serwer www i przetwarzanie jego odpowiedzi. Wykorzystana do komunikacji z serwerem PDBJ.

Biblioteka bs4 wykorzystana została do przetworzenia odpowiedzi html z serwera www w celu zbudowania kolejnych zapytań.

4.2. Instalacja Bibliotek

Biblioteki można łatwo zainstalować na przykład przy pomocy pakietu pip wystarczy w konsoli ze znaną ścieżką do pliku python.exe wydać polecenia:

```
python -m pip install bs4
python -m pip install requests
```

5. Obsługa wejścia, wyjścia oraz błędów:

Parametry konwersji jak i ścieżki do plików podawane są przez użytkownika za pomocą argumentów wywołania. Skrypt ten wykorzystuje pliki tymczasowe w celu przesłania na serwer części plików. Wszelkiego rodzaju błędy wyświetlane są w postaci: „Error: Opis błędu”. Najczęstszymi są błędne parametry, nieistniejący plik wejściowy, plik w formacie nie obsługiwanym.

5.1. Wejście

Program można obsługiwać przy według następującego schematu:

```
main [-h] [-l] -i plikWejście -o plikWyjście [-m tryb]
      [-c chain1,chain2,...]
```

-h	wyświetla pomoc
-l	z tym parametrem log zostanie pobrany
-i plikWejście	ustawia plik wejściowy jako plikWejście
-o plikWyjście	ustawia plik wyjściowy jako plikWyjście
-m tryb	ustawia tryb konwersji jako tryb

domyślny tryb = a - cały plik
tryb = h - tylko nagłówek
tryb = c - tylko część koordynatową
-c chains tylko z [-m c] pozwala na wybranie konkretnych modeli łańcuchów oddzielonych [,]

5.2. Wyjście

Końcowym wyjściem programu jest przekonwertowany plik oraz jeśli wybrano plik log. Podczas działania skryptu mogą pojawić się następujące komunikaty kontrolne:

```
Input file name: nazwa pliku wejściowego  
Output file name: nazwa pliku wyjściowego  
Chosen mode: wybrany tryb  
Uploading file to server.. (postęp konwersji)  
Conversion in progress...  
Downloading file...  
Succeeded.
```

6. Przykłady:

6.1. main -i 1asy.cif -o 1asy.pdb

konwersja całego pliku 1asy.cif do pliku 1asy.pdb.

6.2. main -l -i 1asy.pdb -o 1asy.cif -m h

konwersja części nagłówkowej z pliku 1asy.pdb do pliku 1asy.cif z opcją utworzenia pliku log.

6.3. main -l -i 1asy.pdb -o 1asy.cif -m c -c A,B

konwersja części zawierającej koordynaty z zawężeniem do łańcucha A i B z pliku 1asy.pdb do pliku 1asy.cif wraz z utworzeniem pliku log.

7. Podsumowanie:

Projekt ten ma duży potencjał zastosowań praktycznych jest zbudowany w łatwy do dostosowania sposób oraz może zostać rozbudowany o dodatkowe funkcję na przykład podział pliku mmCIF w celu uniknięcia przekroczenia ilości atomów w pliku PDB.