AZON API - opis realizacji

6 maja 2020

Spis treści

1	Wst	Wstęp			
2	Biblioteka				
	2.1	Wyszu	kiwanie zasobów po ich autorze	2	
	2.2		kiwanie zasobów po stronach		
	2.3		owanie metadanych wyszukanych zasobów		
	2.4		etody oraz wersjonowanie biblioteki		
			Lista języków programowania		
		2.4.2	Lista centrów badawczych		
		2.4.3	Lista laboratoriów	,	
		2.4.4	Historia wersjonowania		
3	Aplikacja				
	3.1	Menu		;	
	3.2	Przetes	stowanie wersjonowania (upgrade i downgrade)	10	

1 Wstęp

Biblioteka została utworzona na platformie GitHub (link, gałąź *biblioteka*) w wyniku problemów z e-science'owym GitLab'em. Aplikacja korzystająca z biblioteki znajduje się tutaj, gałąź master. Na GitHub'ie znajduje się kilka wersji biblioteki - dostępne do wglądu tutaj. Opisy jak zrobić upgrade oraz downgrade biblioteki znajdują się w pliku *README.md* na GitLab'ie.

Ogólnie został utworzony jeden pakiet (package1) chociaż jest też drugi (package2), lecz w którym nic nie manie miałem pomysłu co by tam wrzucić. Wszystko co było wymagane jest w pakiecie package1, w tym drugim jest jedynie jakaś prosta funkcja, tylko w celu pokazania, że można korzystać z obu. W module api_handler.py są napisane funkcje pozwalające wyszukiwać zasoby po autorze oraz po stronach, w module bibtex_classes.py klasy odpowiadające poszczególnym typom wpisów w formacie BibTeX, a w module api_others.py są inne funkcje zwracające np. listę jezyków programowania, centrów naukowych lub laboratoriów.

Przetestowanie wersjonowania znajduje się na samym końcu.

2 Biblioteka

2.1 Wyszukiwanie zasobów po ich autorze

W celu wyszukiwania zasobów po autorze została napisana klasa Entry, która zawiera wszystkie podstawowe atrybuty-klucze, które są pod kluczem *result* w odpowiedzi pod na przykład endpoint'em *authors/entries/2871/*. Zostałą utworzona w celu uproszczenia przechowywania w pamięci uzyskanych wyników. Żeby uzyskać przykładowe obiekty klasy Entry została napisana funkcja *get_author_entries*, która zwraca wszystkie zasoby autora oraz je konwertuje na obiekty klasy Entry. Jeśli autor posiada więcej niż 100 zasobów to należy z Menu wybrać opcję wyszukania po stronie - funkcję *get_author_entries_by_page*:

```
def get_author_entries(author_pk, api_key):
          response = requests.get("https://api.e-science.pl/api/azon/authors/entries/"+str(
      author_pk)+"/", headers={'X-Api-Key': api_key})
          response.raise_for_status()
          json_data = response.json()
          results = json_data['results']
          entries_list=[]
          for item in results:
              entries_list.append(Entry(**item))
          return entries_list
11
      except HTTPError as http_err:
12
          print(f'HTTP error occurred: {http_err}')
13
14
  def get_author_entries_by_page(author_pk, api_key, page):
15
16
      try:
17
          response = requests.get("https://api.e-science.pl/api/azon/authors/entries/"+str(
      author_pk)+"/?limit=100&offset="
                                        + str(page) + "00", headers={'X-Api-Key': api_key})
         response.raise_for_status()
18
          json_data = response.json()
          results = json_data['results']
20
          entries list=[]
21
          for item in results:
              entries_list.append(Entry(**item))
23
24
          return entries_list
      except HTTPError as http_err:
26
          print(f'HTTP error occurred: {http_err}')
```

2.2 Wyszukiwanie zasobów po stronach

W przypadku wyszukiwania zasobów po stronach również została wykorzystana klasa Entry, lecz tym razem wykorzystany endpoint to *entry/filter/*. Została napisana metoda *get_entries_by_page*, która jako parametr przyjmuje klucz API oraz numer strony i zwraca listę obiektów klasy Entry z danej strony. Ciało funkcji wygląda następująco:

```
def get_entries_by_page(api_key, page):
    try:
        response = requests.get("https://api.e-science.pl/api/azon/entry/filter/?limit=100&
    offset=" + str(page) + "00", headers={'X-Api-Key': api_key})
        response.raise_for_status()
        json_data = response.json()
        results = json_data['results']
        entries_list=[]
        for item in results:
            entries_list.append(Entry(**item))
        return entries_list

except HTTPError as http_err:
        print(f'HTTP error occurred: {http_err}')
```

2.3 Generowanie metadanych wyszukanych zasobów

Każdy wyszukany zasób ma atrybut *entry_type_id*, który określa jaki typ reprezentuje dany zasób. Patrząc na odpowiedź serwera, po zapytaniu na endpoint *entry/types/index/*, wybrane zostały typy, których było najwięcej tak żeby pokryć niemalże całą bazę wszystkich zasobów:

- książka 1890, *entry_type_id* = 1,
- artykuł, rozdział 2682, entry_type_id = 2,
- czasopismo 7251, *entry_type_id* = 18,
- zbiór, baza danych 13174, entry_type_id = 15,
- zdjęcie 1956, *entry_type_id* = 6,
- prezentacja 107, entry_type_id = 7,
- inny dokument 964, *entry_type_id* = 17,
- kod źródłowy 62, entry_type_id = 24,
- wideo 3647, *entry_type_id* = 14,
- praca dyplomowa 373, *entry_type_id* = 4,
- analiza chemiczna 763, entry_type_id = 10,

W ten sposób uzyskano pewność przekonwertowania na format BibTeX 32869 zasobów z ogólnej liczby 41447 zasobów, czyli prawie 80% całej bazy. Jeśli zasobu nie da się przekonwertować to jest po prostu pomijany, a w logach konsoli wyskakuje informacja o KeyError. Została skonstruowana prosta funkcja *get_entries_details*, która jako parametr przyjmuje listę obiektów klasy Entry (*entries_list*) oraz klucz API (*api_key*) i za pomocą słownika *switcher* tworzy odpowiedni obiekt w zależności od *entry_type_id*. Został dodatkowo wprowadzone uśpienie wątku na 1 sekundę, gdyż przy mniejszych czasach (szczególnie jego braku) serwer sypał błędami typu 429 (*Too Many Requests*).

Jako ciekawostkę mogę podać, że ponad 25% wszystkich zasobów (prawie wszystkie typu zbiór, baza danych) jest autorstwa autora o imieniu Adrian Zakrzewski.

```
entries_BibTeXs = []
      for entry in entries_list:
          time.sleep(1)
          pk = entrv.pk
          entry_type = entry.entry_type_id
          response = requests.get("https://api.e-science.pl/api/azon/entry/"+str(pk)+"/", headers
      ={'X-Api-Key': api_key})
          response.raise_for_status()
          json_data = response.json()
          switcher = {
               '1': get_book,
10
              '2': get_article,
11
              '4': get_phdthesis,
12
13
               '6': get_photo,
              '7': get_misc,
14
               '10': get_techreport,
               '14': get_video,
16
              '15': get_misc,
17
               '17': get_misc,
               '18': get_magazine,
19
               '24': get_misc
20
          item = switcher[entry_type](json_data)
22
          entries_BibTeXs.append(item.get_BibTeX())
```

Każda z funkcji (*get_book*, *get_article* etc.) szuka innych kluczy w odpowiedzi (*json_data*) - takich które są potrzebne lub opcjonalne do powstania BibTeXa. Przykładowo jeśli zasób jest książką funkcja szuka takich kluczy jak poniżej i tworzy odpowiedni obiekt, który po utworzeniu jest zwracany:

```
def get_book(json_data):
       info={}
       item = json_data['item']
       info['authors']=json_data['authors']
       info['title']=json_data['title']
       info['year']=item['publish_time']
       info['publisher']=item['publisher']
info['isbn']=item['isbn']
       info['note']=json_data['comments']
       info['address']=item['publish_place']
info['edition']=item['numeration']
10
       info['series'] = item['series_name']
12
13
       info['pk']=json_data['pk']
       book = BibTeX_classes.Book(**info)
       return book
```

Funkcja get_entries_details przyjmuje jako parametr listę obiektów klasy Entry. Taką listę zwracają i funkcja wyszukania zasobów po autorze (get_author_entries) i funkcja wyszukiwania zasobów na poszczególnych stronach (get_entries_by_page). Dzięki temu aplikacja jest w stanie jest w stanie wygenerować BibTeX-y dla każdych wyszukanych zasobów aktualnie przechowywanych w pamięci.

Przykładowe wygenerowane BibTeX-y wyglądają następująco (dostępne też w pliku BibTeX.bib na gitlabie):

```
@inbook{47151,
   author = "El bieta Litwi ska, Jadwiga Brzuchowska, El bieta Ch dzy ska",
   title = "Computer models in town planning teaching",
   publisher = "Katedra Planowania Przestrzennego Wydzia Architektury Politechniki
        Wroc awskiej",
   pages = 2-13,
   year = 1993,
   address = "Wroc aw",
```

```
note = "Wyst pienie kongresowe. Ze zbioru "Planowanie i rodowisko w zmieniaj cej si
      Europie", Wroc aw 1993. '
10
  @misc{48241,
11
   authors = "['Jadwiga Brzuchowska', 'El bieta Litwi ska', 'Tomasz Ossowicz', 'Justyna Ma ko-
12
     Osiadacz']",
    title = "Mo liwo ci powi za wybranych technik wspomagania planowania przestrzennego baz
      danych typu GIS. Etap II",
    type = "inny dokument"
14
  }
16
  @book {49843,
17
   author = "Tomasz Ossowicz, El bieta Litwi ska, Jadwiga Brzuchowska",
18
    title = "Raporty Wydzia Architektury Politechniki Wroc awskiej: Modelowanie rozwoju
19
      przestrzennego Bydgoszczy przy pomocy modelu symulacyjno-decyzyjnego ORION",
    publisher = "Instytut Architektury i Urbanistyki Politechniki Wroc awskiej",
20
    year = 1997,
21
    series = "Raporty Instytutu Architektury i Urbanistyki PWr",
    address = "Wroc aw"
23
24
  }
25
  @book {26215,
26
   author = "Tadeusz Zipser, Jerzy S awski, El bieta Litwi ska, Tomasz Ossowicz, Jadwiga
     Brzuchowska",
    title = "Model symulacyjno-decyzyjny ORION",
28
    publisher = "Wydawnictwo Politechniki Wroc awskiej",
   year = 1994,
30
    address = "Wroc aw",
31
    isbn = "83-7085-115-0"
32
33
34
  @phdthesis{38884,
35
   author = "Jerzy S awski",
36
    title = "Model dynamiki system w osadniczych oparty na koncepcji potencja w. Zastosowanie
37
     wsp czesnych koncepcji dynamiki system w",
    school = "Politechnika Wroc awska",
39
    year = 1990,
    address = "Wroc aw",
40
    note = "Praca doktorska."
41
```

W Menu jest opcja zapisania do pliku aktualnie przechowywanych w pamięci BibTeX-ów.

Bibliografia dla przykładowych BibTeX-ów

- [1] Możliwości powiązań wybranych technik wspomagania planowania przestrzennego bazą danych typu gis. etap ii
- [2] Elżbieta Chądzyńska Elżbieta Litwińska, Jadwiga Brzuchowska. *Computer models in town planning teaching*, page 2. Katedra Planowania Przestrzennego Wydział Architektury Politechniki Wrocławskiej.
- [3] Jerzy Sławski. Model dynamiki systemów osadniczych oparty na koncepcji potencjałów. Zastosowanie współczesnych koncepcji dynamiki systemów. PhD thesis, Politechnika Wrocławska, Wrocław, 1990. Praca doktorska.
- [4] Elżbieta Litwińska Tomasz Ossowicz Jadwiga Brzuchowska Tadeusz Zipser, Jerzy Sławski. *Model symulacyjno-decyzyjny ORION*. Wydawnictwo Politechniki Wrocławskiej, Wrocław, 1994.
- [5] Jadwiga Brzuchowska Tomasz Ossowicz, Elżbieta Litwińska. Raporty Wydział Architektury Politechniki Wrocławskiej: Modelowanie rozwoju przestrzennego Bydgoszczy przy pomocy modelu symulacyjno-decyzyjnego ORION. Raporty Instytutu Architektury i Urbanistyki PWr. Instytut Architektury i Urbanistyki Politechniki Wrocławskiej, Wrocław, 1997.

2.4 Inne metody oraz wersjonowanie biblioteki

W celu przetestowania, żeby zobaczyć jak to jest z wersjonowaniem zostało napisanych parę metod, a dodanie każdej powodowało utworzenie nowej wersji biblioteki. Dotychczasową wersją biblioteki była wersja v3.0.

2.4.1 Lista języków programowania

Po utworzeniu tej metody został nadany nowy tag v4.0.

```
def get_programming_languages(api_key):
2
      try:
         response = requests.get("https://api.e-science.pl/api/azon//programminglanguages/",
      headers = { 'X-Api-Key': api_key })
         response.raise_for_status()
          json_data = response.json()
          results = json_data['results']
          lang_list=[]
          for item in results:
             lang_list.append(item['name'])
          return lang_list
10
11
      except HTTPError as http_err:
12
          print(f'HTTP error occurred: {http_err}')
```

2.4.2 Lista centrów badawczych

Po utworzeniu tej metody został nadany nowy tag v5.0.

```
class ResearchCenters:
      def __init__(self,_id, name, number, director, tel, email, webpage):
          self.id=_id
          self.name=name
          self.number=number
          self.director=director
          self.tel=tel
          self.email=email
          self.webpage=webpage
10
      def __repr__(self):
          return f'Id - {self.id}\nNazwa - {self.name}\nNumer - {self.number}\nDyrektor - {self.
12
      director}\nTelefon - {self.tel}\nE-mail - {self.email}\nStrona - {self.webpage}\n'
13
  def get_pwr_reseach_centres(api_key):
14
15
          response = requests.get("https://api.e-science.pl/api/azon/databases/
16
      pwr_research_centers/", headers={'X-Api-Key': api_key})
          response.raise_for_status()
          json_data = response.json()
18
          results = json_data['results']
19
          rc_list=[]
          for item in results:
21
22
              info={}
              info['_id']=item['id']
23
              info['name']=item['name']
24
              info['number'] = item['number']
25
              info['director']=item['director']
26
              info['tel']=item['tel']
27
               info['email']=item['email']
28
              info['webpage']=item['webpage']
29
              rc_list.append(ResearchCenters(**info))
30
31
          return rc_list
32
```

```
except HTTPError as http_err:
print(f'HTTP error occurred: {http_err}')
```

2.4.3 Lista laboratoriów

Po utworzeniu tej metody został nadany nowy tag v6.0.

```
class Laboratory:
      def __init__(self,_id, web_address, acronym, full_name, partner):
          self.id=_id
          self.web_address=web_address
          self.acronym=acronym
          self.full_name=full_name
          self.partner=partner
      def __repr__(self):
          return f'Id - {self.id}\nStrona - {self.web_address}\nAkronim - {self.acronym}\nPe na
10
      nazwa - {self.full_name}\nPartner - {self.partner}\n'
11
12
  def get_laboratories(api_key):
13
      try:
          response = requests.get("https://api.e-science.pl/api/azon/databases/elaboratory/",
14
      headers = { 'X-Api-Key': api_key })
          response.raise_for_status()
15
16
           json_data = response.json()
          results = json_data['results']
          lab_list = []
18
19
           for item in results:
               info={}
20
               info['_id']=item['id']
info['web_address']=item['web_address']
21
22
               info['full_name'] = item['full_name']
23
               info['acronym']=item['acronym']
24
               info['partner']=item['partner']
25
               lab_list.append(Laboratory(**info))
26
           return lab_list
```

2.4.4 Historia wersjonowania

```
C:\Users\Radek\Desktop\6semestr\DPP\lab_9_pip\github>git log --oneline --graph
 4771908 (HEAD -> biblioteka, tag: v6.0, origin/biblioteka) Add get_laboratories method e31c952 (tag: v5.0) Add get_pwr_research_centres method
 73eaf38 (tag: v4.0) Add get_programming_languages_method
 eeea553 (tag: v3.0) Distribute version v3.0
 80b1345 Finish method saving BibTex to the file
 6387b6a (tag: v2.0) Finish converting all types of entries to bibtex
 dc020eb Change package1 structure
 5302799 Add method for returning author entries
 6a0a583 Distribute second version
 8eb9456 Add second package
 3568060 Update README.md
 79c871a Distribute package with something else
 375fc7a Change word in return statement
 71765fe (tag: v1.0) Add return to say hello method
 142864d Distribute first version of library
 5ccdf23 (origin/master, origin/HEAD, master) Initial commit
```

Rysunek 1: Historia logów

Na podstawie powyższej historii można spodziewać się, że wersja *v3.0* nie będzie posiadała żadnej z trzech dodanych metod i idąc tak po kolei dopiero wersja ostateczna - czyli *v6.0* będzie udostępniać wszystkie metody. Będzie się dało o tym przekonać podczas testowania aplikacji.

3 Aplikacja

W celu skorzystania z biblioteki został napisany prosty moduł aplikacji, który miał korzystać metod udostępnianych przez bibliotekę.

3.1 Menu

Utworzone menu pozwala na przetestowanie wszystkich wcześniej opisanych metod:

```
PS C:\Users\Radek\Desktop\6semestr\DPP\lab_9_pip\e-science> ${env:DEBUGPY_LAUNCHER_PORT}='57602'; & 'C:\Users\Radek\AppOata\Local\Programs\debugpy\wheels\debugpy\launcher' 'c:\Users\Radek\Desktop\6semestr\DPP\lab_9_pip\e-science\main.py'

1: Wyszukaj zasoby po ID autora.
2: Wyszukaj zasoby po ID autora i stronie
3: Wyszukaj zasoby po stronie
4: Wyszukaj jezyki programowania
5: Wyszukaj centra badawcze PWR
6: Wyszukaj laboratoria
7: Wygeneruj BibTeX-y dla obecnie wyszukanych zasobów
8: Zapisz do pliku obecne BibTex-y
9: Przetestuj pakiet package2
10: Wyjście z programu

Co chcesz zrobić?:
```

Rysunek 2: Utworzone menu

```
Tytuł - Próba zastosowania modelu podsystemu decyzyjnego "ORION" w planowaniu zagospodarowania turystycznego
Id - 46792
Autor - Barbara Jaskólska
Typ - książka
Partner - Politechnika Wrocławska
Tytuł - Computer models in town planning teaching
Id - 47151
Autor - Elżbieta Litwińska
Typ - artykuł, rozdział
Partner - Politechnika Wrocławska
Razem 6 zasobów
        1: Wyszukaj zasoby po ID autora.
        2: Wyszukaj zasoby po ID autora i stronie
        3: Wyszukaj zasoby po stronie
4: Wyszukaj języki programowania
        5: Wyszukaj centra badawcze PWR
        6: Wyszukaj laboratoria
        7: Wygeneruj BibTeX-y dla obecnie wyszukanych zasobów
        8: Zapisz do pliku obecne BibTex-y
        9: Przetestuj pakiet package2
        10: Wyjście z programu
Co chcesz zrobić?: 7
Wygenerowano razem 6 BibTeX-ów
        1: Wyszukaj zasoby po ID autora.
        2: Wyszukaj zasoby po ID autora i stronie
        3: Wyszukaj zasoby po stronie
        4: Wyszukaj języki programowania
        5: Wyszukaj centra badawcze PWR
        6: Wyszukaj laboratoria
        7: Wygeneruj BibTeX-y dla obecnie wyszukanych zasobów
        8: Zapisz do pliku obecne BibTex-y
        9: Przetestuj pakiet package2
        10: Wyjście z programu
Co chcesz zrobić?: 8
Zapisano razem 6 BibTeX-ów
        1: Wyszukaj zasoby po ID autora.
        2: Wyszukaj zasoby po ID autora i stronie
        3: Wyszukaj zasoby po stronie
4: Wyszukaj języki programowania
        5: Wyszukaj centra badawcze PWR
        6: Wyszukaj laboratoria
        7: Wygeneruj BibTeX-y dla obecnie wyszukanych zasobów
        8: Zapisz do pliku obecne BibTex-y
        9: Przetestuj pakiet package2
        10: Wyjście z programu
Co chcesz zrobić?:
```

Rysunek 3: Przykładowy output konsoli

3.2 Przetestowanie wersjonowania (upgrade i downgrade)

W celu przetestowania wersjonowania został napisany prosty *bash'owy* skrypt instalujący odpowiednią wersję i sprawdzający czy metody są znajdowane poprzez złapanie odpowiedniego wyjątku *AttributeError*:

```
def search_laboratories():
    try:
        labs = api_others.get_laboratories(api_key)
    except AttributeError:
        return "Nie ma takiej metody jak get_laboratories"
    return f'Znaleziono {len(labs)} laboratori w'
```

Skrypt wygląda następująco:

```
#!/bin/bash
 pip3 install --upgrade git+git://github.com/radosz99/azon-api-library.git@v3.0#egg=azon-api-
     library
 echo ------>> results.txt
 python3 main.py >> results.txt
 pip3 install --upgrade git+git://github.com/radosz99/azon-api-library.git@v4.0#egg=azon-api-
     library
 echo ------>> results.txt
10
 python3 main.py >> results.txt
12
 pip3 install --upgrade git+git://github.com/radosz99/azon-api-library.git@v5.0#egq=azon-api-
13
14
 echo ------>> results.txt
 python3 main.py >> results.txt
16
17
 pip3 install --upgrade git+git://github.com/radosz99/azon-api-library.git@v6.0#egg=azon-api-
     library
19
 echo ------>> results.txt
 python3 main.py >> results.txt
21
22
 pip3 install --upgrade git+git://github.com/radosz99/azon-api-library.git@v5.0#egg=azon-api-
     library
24
 echo ------>> results.txt
 python3 main.py >> results.txt
26
 pip3 install --upgrade git+git://github.com/radosz99/azon-api-library.git@v4.0#egg=azon-api-
28
     library
 echo ----- >> results.txt
31
 python3 main.py >> results.txt
32
 pip3 install --upgrade git+git://github.com/radosz99/azon-api-library.git@v3.0#egg=azon-api-
33
     library
34
 echo ----- >> results.txt
35
 python3 main.py >> results.txt
```

Output - w celu zaprezentowania rezultatów - został przekierowany do pliku *results.txt*, którego treść po uruchomieniu skryptu wygląda następująco:

```
-----Wersja v3.0-----
2 Nie ma takiej metody jak get_programming_languages
3 Nie ma takiej metody jak get_pwr_reseach_centres
 Nie ma takiej metody jak get_laboratories
  -----UPGRADE-Wersja v4.0-----
7
 Znaleziono 30 jezykow programowania
 Nie ma takiej metody jak get_pwr_reseach_centres
 Nie ma takiej metody jak get_laboratories
  -----UPGRADE-Wersja v5.0-----
11
12 Znaleziono 30 jezykow programowania
 Znaleziono 15 centrow badawczych
13
 Nie ma takiej metody jak get_laboratories
15
  -----UPGRADE-Wersja v6.0-----
16
17
 Znaleziono 30 jezykow programowania
 Znaleziono 15 centrow badawczych
18
 Znaleziono 38 laboratoriow
19
20
  -----DOWNGRADE-Wersja v5.0-----
21
22
 Znaleziono 30 jezykow programowania
23
  Znaleziono 15 centrow badawczych
 Nie ma takiej metody jak get_laboratories
24
  -----DOWNGRADE-Wersja v4.0-----
26
27 Znaleziono 30 jezykow programowania
28 Nie ma takiej metody jak get_pwr_reseach_centres
29 Nie ma takiej metody jak get_laboratories
  -----DOWNGRADE-Wersja v3.0-----
32 Nie ma takiej metody jak get_programming_languages
 Nie ma takiej metody jak get_pwr_reseach_centres
34 Nie ma takiej metody jak get_laboratories
```

Treść zgadza się z tym co powinno być - w wersji v3.0 nie było żadnej z dodatkowych metod, w wersji v4.0 doszło wyszukiwanie języków programowania, w wersji v5.0 wyszukiwanie centrów badawczych, zaś w ostatniej - v6.0 - wyszukiwanie laboratoriów.