

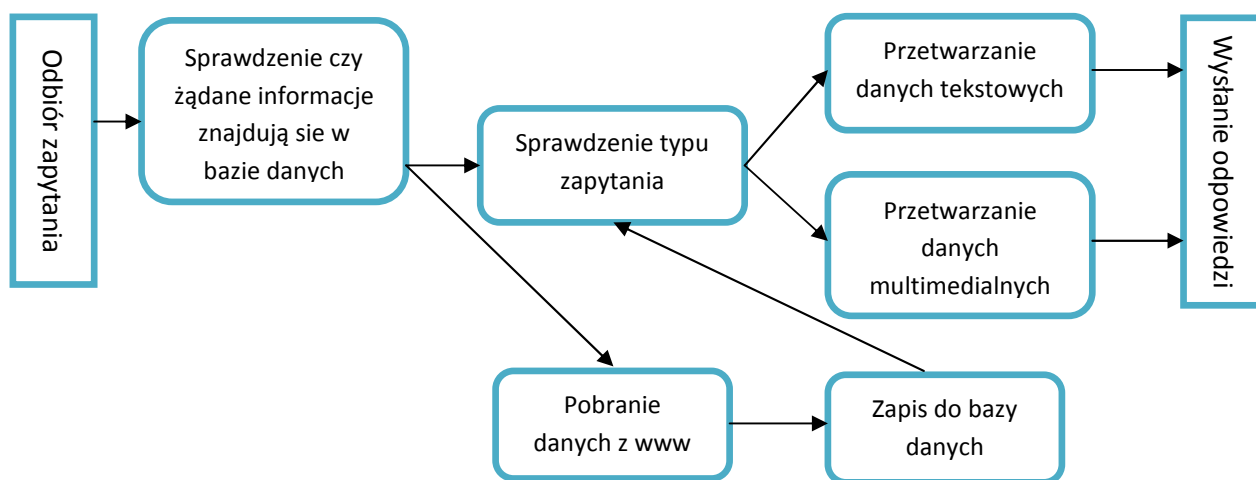
## Zadanie

Zadanie ma na celu praktyczne zapoznanie z wykorzystaniem **języka Python**. W szerszym ujęciu, rozwiązanie wskazanego problemu z wykorzystaniem zadanych technik/technologii.

**Zajęcia mają charakter projektowy.** Podczas kolejnych czterech zajęć laboratoryjnych każdy student powinien wykonać projekt oraz implementację aplikacji sieciowej napisanej w języku skryptowym Python. Uruchomienie aplikacji musi być możliwe na komputerach znajdujących się w laboratorium (sugerowana wersja języka to Python 2.7.x).

## Opis projektu

Stworzona aplikacja sieciowa będzie miała za zadanie pobrać zawartość strony www. Dla uproszczenia przyjmujemy, że będzie to strona angielskiej wikipedii (<http://en.wikipedia.org/>). Zapytanie będzie zawierało nazwę państwa, o którym informacje należy pobrać oraz opcjonalne szczegóły zapytania. Po odpowiedzi na zapytanie pobrane ze strony informacje należy zapisać w bazie danych MongoDB. W przypadku powtórzenia się identycznego zapytania dane należy odczytać z bazy danych (nie pobierać ponownie). Zapytania powinny być przesyłane za pomocą metod HTTP (post, get, itd.). Dodatkowo, aplikacja powinna umożliwić wyszukanie państwa na podstawie przesłanego obrazu przedstawiającego flagę państwa (obraz flagi należy pobrać z internetu).



Rys. Ogólny schemat przetwarzania danych aplikacji.

Wszystkie zapytania powinny być pisane zgodnie z formatem JSON. Zgodnie ze schematem:

```
data = {address: adres,          #adres na który należy przesłać odpowiedź, np. localhost
        port: port,             #port na który należy przesłać, np. 9091
        type: typ,              #typ zapytania; 'text' lub 'image'
        content: treść_zapytania #treść zapytania
    }
```

przykłady możliwych treści zapytań:

```
country(Poland)          #powinno zwrócić kompletny tekst opisu państwa - Poland
                           (pomijając elementy których nie można wyświetlić w formie
                           tekstowej oraz znaczniki html)

country(Poland);tag(Sea)  #jw. z tym, że należy zwrócić tylko te zdania w których znajduje
                           się słowo 'Sea' w formie listy elementów

country(Poland);getflag   #zapytanie zwraca adres pod którym jest dostępny obraz flagi
                           państwa Poland

checkflag(http://test.com.flag.gif) #zapytanie zwraca nazwę państwa którego flaga została
                                       przesłana w postaci parametru zapytania
```

Ocenie podlegać będą następujące elementy i funkcjonalności:

- Prawidłowe pobranie treści tekstowej strony internetowej (0-1).
- Wyszukanie informacji w pobranej treści tekstowej (0-1).
- Stworzenie odpowiedniej struktury dokumentów bazy danych MongoDB (0-1).
- Poprawny zapis i odczyt informacji do i z bazy danych (0-1pkt).
- Prawidłowe pobranie i przetworzenie treści multimedialnych, rozpoznanie flagi (0-2).
- Prawidłowy odbiór zapytania i odesłanie odpowiedzi metodą HTTP (0-2).

W nawiasach podano ilość możliwych do zdobycia punktów za każdy element aplikacji.

*Dokumentacja z kodzie programu zgodna z:*

*<https://www.python.org/dev/peps/pep-0008/#documentation-strings>*

*będzie bardzo mile widziana.*

## Wskazówki

### Język Python

Język Python (<https://www.python.org/>) jest obiekowym, skryptowym językiem programowania. Użycie tego języka jest (dosłownie) dziecinnie proste i dla osób, które nie miały z nim wcześniej styczności sprowadza się do zapoznania ze składnią języka. (Zakładamy, że student opanował w toku studiów przynajmniej jeden obiekowy język programowania.)

W zapoznaniu się z językiem Python pomogą:

- Kursy online:
  - [http://www.tutorialspoint.com/python/python\\_quick\\_guide.htm](http://www.tutorialspoint.com/python/python_quick_guide.htm)
  - <http://heather.cs.ucdavis.edu/~matloff/Python/PythonIntro.html>
  - <http://www.stavros.io/tutorials/python/>
- Książki:
  - Learning Python, 5th Edition, Mark Lutz, O'Reilly
  - Python for Kids, Jason R. Briggs, no starch press

Dla języka Python polecane są dwa IDE: PyCharm i Eclipse (z pluginem Pydev).

### Pobranie treści i parsowanie tekstu

Operacje z wykorzystaniem URL: moduł urllib2: <https://docs.python.org/2/library/urllib2.html>

Operacje na tekście: moduł string: <https://docs.python.org/2/library/string.html>

### Współpraca z baza danych

MongoDB jest to wieloplatformowa, obiektowa baza danych oparta na dokumentach. Nie jest to baza relacyjna, ani SQLowa. Dane przechowywane są w strukturze dokumentów w formacie BSON (binarnym odpowiednikiem formatu JSON). Pełną dokumentację można znaleźć na stronie: <http://www.mongodb.org/> w szczególności:

- tworzenie struktury dokumentów:  
<http://docs.mongodb.org/manual/core/data-modeling-introduction/>
- podstawowe operacje na nadanych:  
<http://docs.mongodb.org/manual/tutorial/getting-started/>

Zdecydowanie najlepszym sposobem na połączenie się z bazą danych MongoDB jest moduł pymongo: <http://api.mongodb.org/python/current/>

Baza MongoDB jest dostępna bezpłatnie na wielu hostingach, np. <https://mongolab.com/> lub użyć jej lokalną instancję.

### Operacje na obrazach

W celu porównania czy dwa obrazy są do siebie podobne można wykorzystać histogram rozkładu pikseli w obrazie. Następnie określić współczynnik korelacji dwóch wektorów i porównać, zakładając pewien stopień dopasowania jako próg.

Skuteczność identyfikacji flagi można poprawić między innymi przez:

- podział obrazu na wiele części i wyszukiwanie podobieństw w każdym z nich, w ten sposób rozbudowany zostaje zbiór cech obrazu używany w procesie porównywania,
- dodanie zależności zawierających informację o wzajemnym położeniu kolorów,
- wyszukiwanie charakterystycznych elementów flagi, używając deskryptor wizyjny (np. SIFT, SURF, FAST, ORB - dostępne w bibliotece OpenCV).

Nie można założyć, że dwa podobne obrazy to zawsze takie same pliki i posłużyć się np. sumą kontrolną.

Dla prostych operacji na obrazach można użyć biblioteki Pillow:

<https://pypi.python.org/pypi/Pillow/2.7.0>

Znacznie więcej możliwości daje biblioteka OpenCV:

<http://opencv.org/>

Obrazy przedstawiające flagi państw należy pobrać z internetu, np. ze strony:

<http://www.flags.net/fullindex.htm> lub <http://www.mapsofworld.com/flags/poland-flag.html>

Następnie odpowiednio parsując źródło strony www.

Ocenie podlegać będzie poprawna identyfikacja losowo wybranych flag.

Dla chętnych: Możliwe jest pobranie obrazu flagi bezpośrednio ze strony wikipedii, jednak w tym przypadku należy dokonać konwersji z formatu wektorowego (SVG) na rastrowy, aby umożliwić przetwarzanie obrazu. Można tego dokonać np. za pomocą biblioteki pyCairo.

### Serwer HTTP

W celu obsługi zapytań HTTP można użyć wbudowanego modułu serwera HTTP:

<https://docs.python.org/2/library/basehttpserver.html>

Bardziej eleganckim rozwiązaniem jest użycie dedykowanego modułu, np. Tornado

<http://www.tornadoweb.org/en/stable/>

Do wysyłania wiadomości metodami HTTP zaleca się użycie modułu: <http://docs.python-requests.org/en/latest/>

Przydatne pliki binarne dla Windows (np. pymogo, pillow, requests):

<http://www.lfd.uci.edu/~gohlke/pythonlibs/>

*Użycie powyższych wskazówek nie jest obligatoryjne. Zadanie można wykonać korzystając z wielu innych metod, modułów, bibliotek niż te wskazane powyżej. Oceniane będzie działanie implementacji wybranej metody.*

*Napisaną aplikację można testować za pomocą skryptu:*

*<http://kt.agh.edu.pl/~matiolanski/JPWP/clientHTTP.py>*

**Projekt proszę umieścić na serwerze:**

**FTP: [andrew.nstrefa.pl](ftp://andrew.nstrefa.pl)**

**login: [jpwp@andrew.nstrefa.pl](mailto:jpwp@andrew.nstrefa.pl)**

**password: zostanie podane na zajęciach**

**Każdy student powinien wgrać na serwer archiwum zawierające kod źródłowy oraz plik Readme zawierający instrukcję uruchomienia aplikacji oraz dane kontaktowe autora (email).**

**Nazywając archiwum proszę trzymać się konwencji: nrindeksu\_nazwisko, np. 234567\_kowalski.**

**Pliki powinny rozpakowywać się do katalogu o tej samej nazwie. Archiwum powinno zawierać jedynie kod źródłowy i dokumentację (bez plików graficznych, itp.).**