

Python średnio zaawansowany

Dzień 2



Blok nr 2:

Akwizycja danych



AGENDA

- Podstawy formatu HTML
- Podstawy formatu JSON
- Wymiana informacji w sieci: protokół HTTP
- Narzędzia
- Ciasteczka

Podstawy formatu HTML

HTML (HyperText Markup Language)

HTML nie jest językiem programowania lecz *językiem znaczników* (ang. *markup language*).

HTML składa się z serii znaczników (**tagów**), których używa się do zamknięcia, *opakowania* różnych części treści, tak aby wyglądały i/lub działały w określony sposób.

HTML

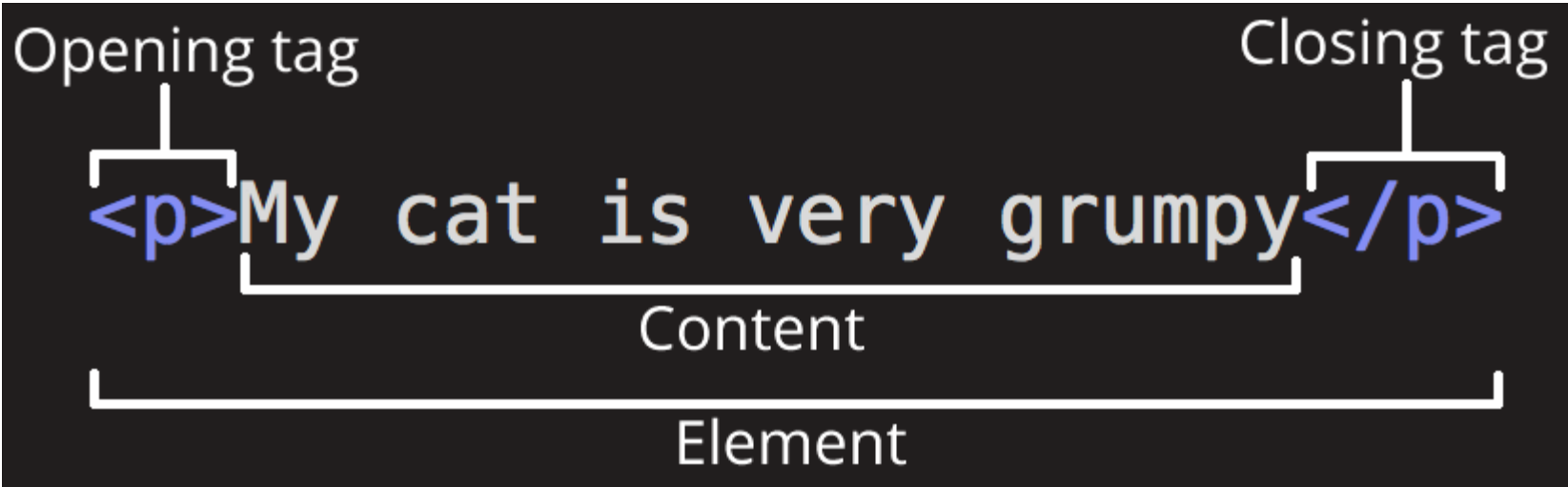
Dla poniższej treści:

My cat is very grumpy


Jeśli chcemy utworzyć z niej paragraf, wystarczy użyć odpowiednich tagów:

```
<p>My cat is very grumpy</p>
```

HTML



HTML



```
<p class="editor-note">My cat is very grumpy</p>
```

Atrybuty zawierają dodatkowe informacje o elemencie, które nie są widoczne dla odwiedzających stronę. W tym przykładzie `class` jest *nazwą* atrybutu, a `editor-note` jest jego *wartością*. Atrybut `class` pozwala nadać elementowi identyfikator, który potem może być użyty do nadania stylu elementowi przez CSS lub do innych celów.

HTML

Elementy możesz umieszczać wewnątrz innych elementów kodu HTML — nazywa się to **zagnieżdżeniem**.

```
<p>My cat is <strong>very</strong> grumpy.</p>
```

W powyższym przykładzie najpierw otwieramy element <p>, następnie element , dalej musimy zamknąć element a dopiero potem <p>.

HTML

Poniższy przykład jest błędny:

```
<p>My cat is <strong>very grumpy.</p></strong>
```

Elementy muszą zaczynać się i kończyć we właściwy sposób, tak aby jeden zawierał się w drugim. Jeśli elementy będą niewłaściwie domknięte, jak w przykładzie wyżej, przeglądarka będzie próbowała zgadnąć co miałeś na myśli, co może prowadzić do nieoczekiwanych rezultatów.

Układ dokumentu HTML

```
<!DOCTYPE html>
<html>
  <head>
    <meta charset="utf-8">
    <title>My test page</title>
  </head>
  <body>
    
  </body>
</html>
```

Układ dokumentu HTML

`<!DOCTYPE html>` — doctype czyli typ dokumentu.

`<html></html>` — element `<html>` zawiera całą treść strony i czasem nazywany jest elementem bazowym (ang. root element) - wskazuje, gdzie zaczyna i kończy się kod HTML.

`<head></head>` — element `<head>` to tzw. nagłówek strony. Ten element działa jak kontener dla wszystkich elementów, które chcesz umieścić na stronie HTML, ale nie jej treści (styl CSS, deklaracje zestawu znaków, użytych skryptów i inne).

Układ dokumentu HTML

`<body></body>` — element, który zawiera *całą* zawartość, która ma być wyświetlana użytkownikom.

`<meta charset="utf-8">` — ten element ustawia zestaw kodowania znaków, którego Twój dokument powinien używać wyświetlania treści strony.

`<title></title>` — ustawia tytuł strony, który jest tytułem wyświetlanym na karcie przeglądarki.

Więcej informacji

<https://developer.mozilla.org/pl/docs/Web/HTML>

<https://www.w3schools.com/html/>

Podstawy formatu JSON

JSON (JavaScript Object Notation)

- lekki, tekstowy **format** wymiany danych (w odróżnieniu od XML),
- pomimo nazwy, JSON jest formatem niezależnym od konkretnego języka programowania,
- kodowany za pomocą UTF-8, który jest dla JSON domyślnym systemem kodowania,
- podobna postać do **zapisu** słownika Python



JSON

```
{  
  "name": "Luke Skywalker",  
  "eye_color": "Blue",  
  "active": true,  
  "gender": "Male",  
  "hair_color": "Blond",  
  "height": 172,  
  "films": [  
    "https://swapi.co/api/films/1/",  
  ],  
  "homeworld": "https://swapi.co/api/planets/1/",  
  "created": "2014-12-09T13:50:51.644000Z",  
  "edited": "2014-12-10T13:52:43.172000Z",  
}
```

str

boolean

int

list

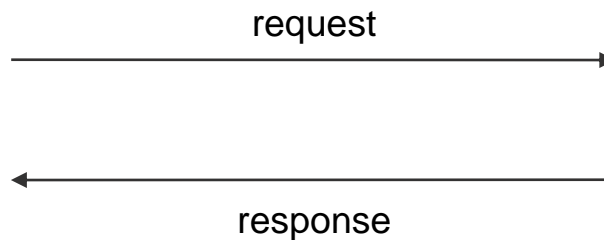
Typy danych przechowywane
w JSON po zdekodowaniu
będą odpowiadać typom
wbudowanym w język.

datetime

Wymiana informacji w sieci

Protokół HTTP

Komunikacja klient - serwer



Request




infoshare academy |



Szukaj w Google

Szczęśliwy traf

Response



[Wszystko](#)
[Grafika](#)
[Wiadomości](#)
[Mapy](#)
[Filmy](#)
[Więcej](#)
[Ustawienia](#)
[Narzędzia](#)

Okolo 59 100 wyników (0,37 s)

infoShare Academy: Kursy programowania
<https://infoShareacademy.com/> ▼
 Zostań programistą! Poznaj intensywne kursy programowania: Front-End, Java, Python, SQL i inne - w Gdańsku, Krakowie, Lublinie, Szczecinie i Warszawie.

Rodzaje kursów
 RODZAJE KURSÓW. Front-end. 320 h kurs dzienny. Szczegóły ...

Miasta i terminy
 Poznaj intensywne kursy programowania: Front-End ...

SQL od podstaw
 SQL od podstaw. 14 listopada 2018 - 16 stycznia 2019 ...

[Więcej wyników z infoShareacademy.com »](#)

Data Science
 Kurs Data Science dedykowany jest dla osób chcących ...



Kontakt
 Zostań programistą! Poznaj intensywne kursy ...

Zespół
 Zespół. Andrzej Kiesz CEO w InfoShare Academy. Andrzej ...

infoShare Academy - Strona główna | Facebook
<https://pl-pl.facebook.com> > Miejsca > Gdańsk ▼
 ★★★★★ Ocena: 4,8 - 30 głosów
 infoShare Academy. Gdańsk. 7,1 tys. osób lubi to. Nauczmy Cię programowania od podstaw. Z utalentowanego technologicznie entuzjasty zostaniesz...

InfoShare Academy - czy ktoś ukończył: 4programmers.net
<https://4programmers.net> > Forum > Szkolenia i konferencje ▼
 8 paź 2017 - 2 wpisy - 1 autor
 W internecie znalazłem 2 miesięczny kurs w szkole InfoShare academy. Koszt jest spory , bo 9900 zł i nie chciałbym utopić tych pieniędzy, ale ...

InfoShare Academy – Bootcampy programistyczne
<https://bootcampu.pl/firma/infoShare-academy/> ▼

Zobacz zdjęcia

infoShare Academy sp. z o.o.

Strona Trasa Zapisz

Centrum szkoleniowe w Gdańsku

Adres: piętro 1, aleja Grunwaldzka 472B, 80-309 Gdańsk

Godziny otwarcia: Otwarte - Zamknijcie: 17:00 ▼

Telefon: 730 833 804

[Zaproponuj zmianę](#)

Opinie z internetu

4,8/5 Facebook · 30 głosów

Podobne wyszukiwania

Fundacja infoShare infoShare Conferen... O4 Coworking Pomorska Fundacja Schibsted Tech Polska

Pokaż jeszcze ponad 5

Wiadomości HTTP

- Zapytania i odpowiedzi współdzielą podobną strukturę składającą się z:
 - Linii początkowej zawierającej informację o zapytaniu, statusie, czy operacja przebiegła pomyślnie lub nie
 - Opcjonalny zbiór nagłówków opisujących typ danych przesyłanych lub opisujących ciało wiadomości
 - Pustej linii oznaczającej pomyślne przesłanie wszystkich metadanych
 - Opcjonalne ciało wiadomości przesyłane w zapytaniu (np. zawartość formularza html), lub dokument zawierający odpowiedź serwera

Wiadomości HTTP

Requests

```
POST / HTTP/1.1
Host: localhost:8000
User-Agent: Mozilla/5.0 (Macintosh;... )... Firefox/51.0
Accept: text/html,application/xhtml+xml,..., */*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Connection: keep-alive
Upgrade-Insecure-Requests: 1
Content-Type: multipart/form-data; boundary=-12656974
Content-Length: 345
```

```
-12656974
(more data)
```

Responses

```
HTTP/1.1 403 Forbidden
Server: Apache
Content-Type: text/html; charset=iso-8859-1
Date: Wed, 10 Aug 2016 09:23:25 GMT
Keep-Alive: timeout=5, max=1000
Connection: Keep-Alive
Age: 3464
Date: Wed, 10 Aug 2016 09:46:25 GMT
X-Cache-Info: caching
Content-Length: 220
```

```
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML
2.0//EN">
(more data)
```

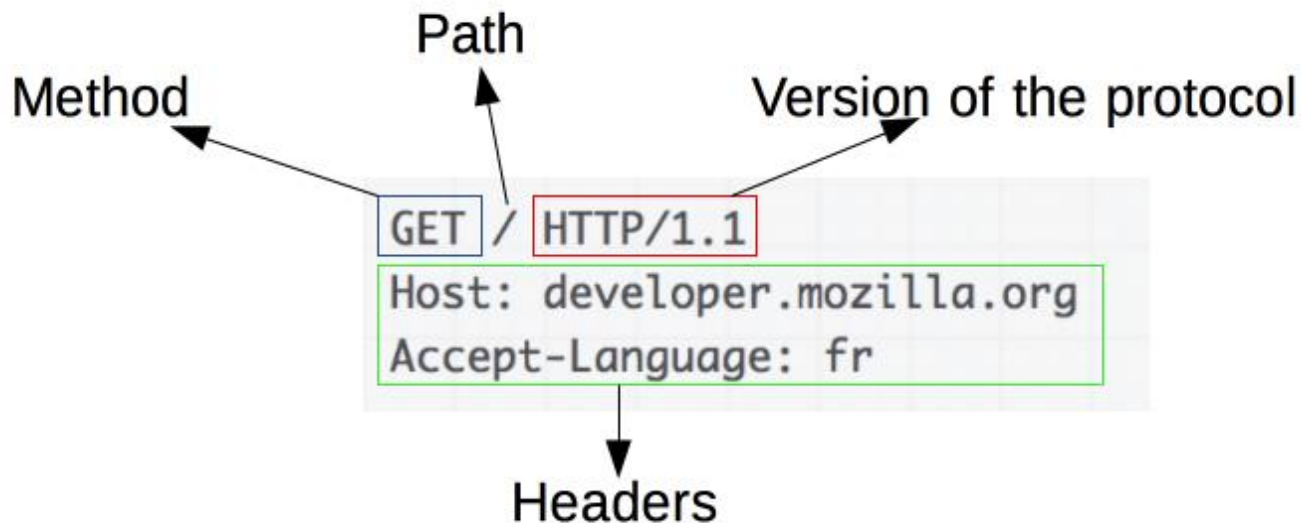
start-
line

HTTP
headers

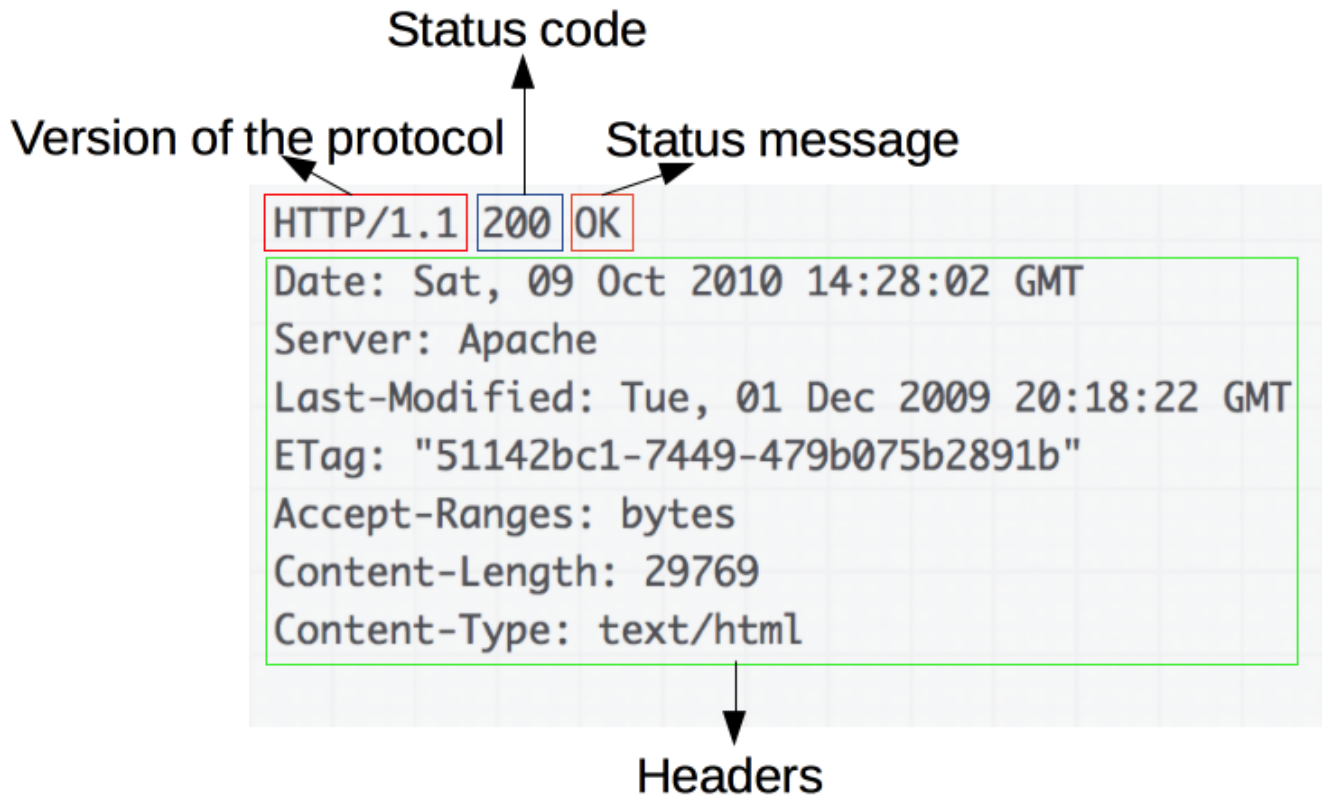
empty
line

body

Wiadomości HTTP - request



Wiadomości HTTP - response



Kody odpowiedzi HTTP

- 1xx – kody informacyjne
- 2xx – kody powodzenia
- 3xx – kody przekierowania
- 4xx – kody błędu aplikacji klienta
- 5xx – kody błędy serwera HTTP

Więcej na: https://en.wikipedia.org/wiki/List_of_HTTP_status_codes

Metody HTTP

Najpopularniejsze metody HTTP:

- **GET** - pobranie wskazanego zasobu
- **POST** - przyjęcie danych przesyłanych od klienta do serwera, w celu utworzenia kolejnego zasobu
- **PUT** - przyjęcie danych przesyłanych od klienta do serwera, aby zaktualizować wartość zasobu lub go stworzyć, jeśli nie istniał
- **DELETE** - żądanie usunięcia zasobu

Więcej na: <https://developer.mozilla.org/en-US/docs/Web/HTTP/Methods>

Narzędzia



Postman

Bardzo przydatny klient protokołu HTTP/S, który jest stosowany przy:

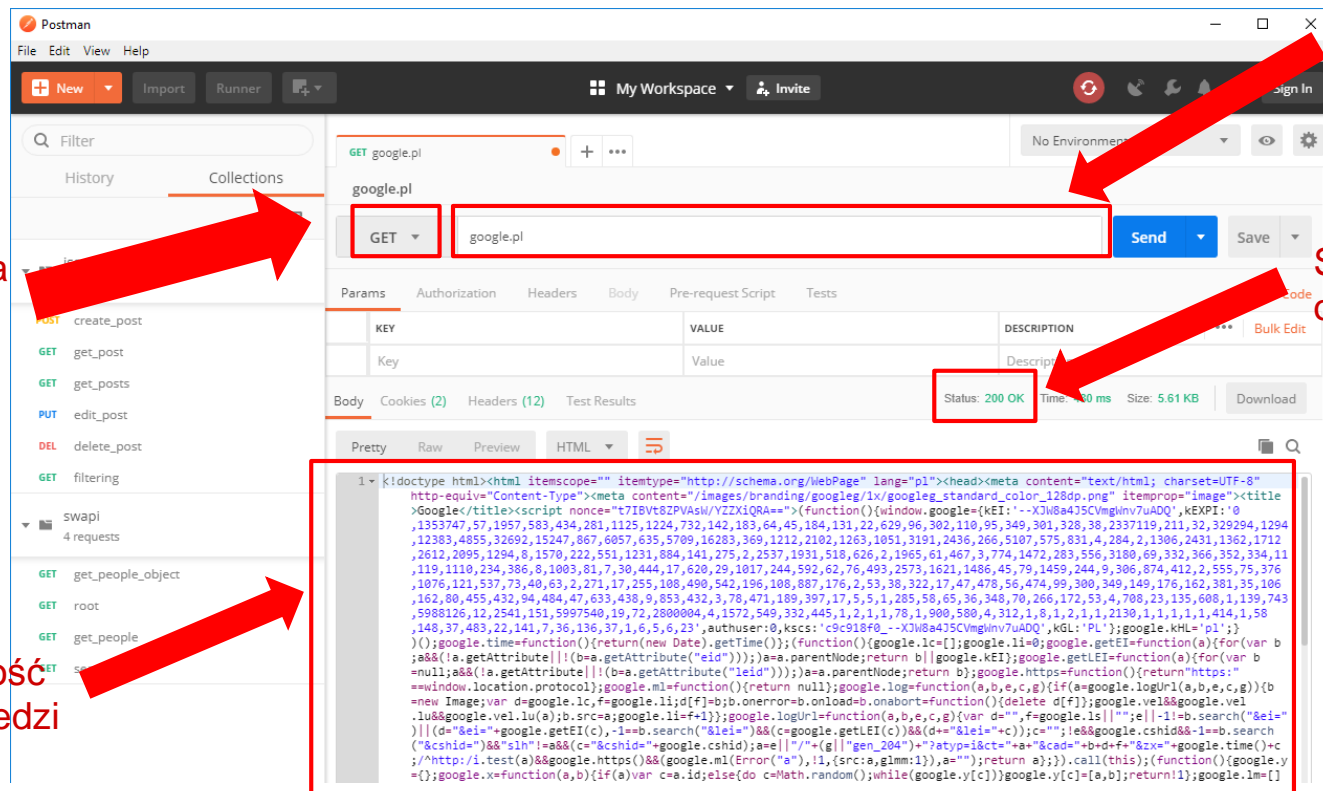
- testowaniu API usług sieciowych (format danych)
- testowaniu poprawności działania usług/serwisów (czy działa?)
- przygotowywaniu przykładów i dokumentowaniu API
- symulowaniu działania usług

Działa jako aplikacja desktopowa lub jako aplikacja Chrome

<https://www.getpostman.com/>

Ciekawy serwis: <https://docs.postman-echo.com/>

Postman



URL

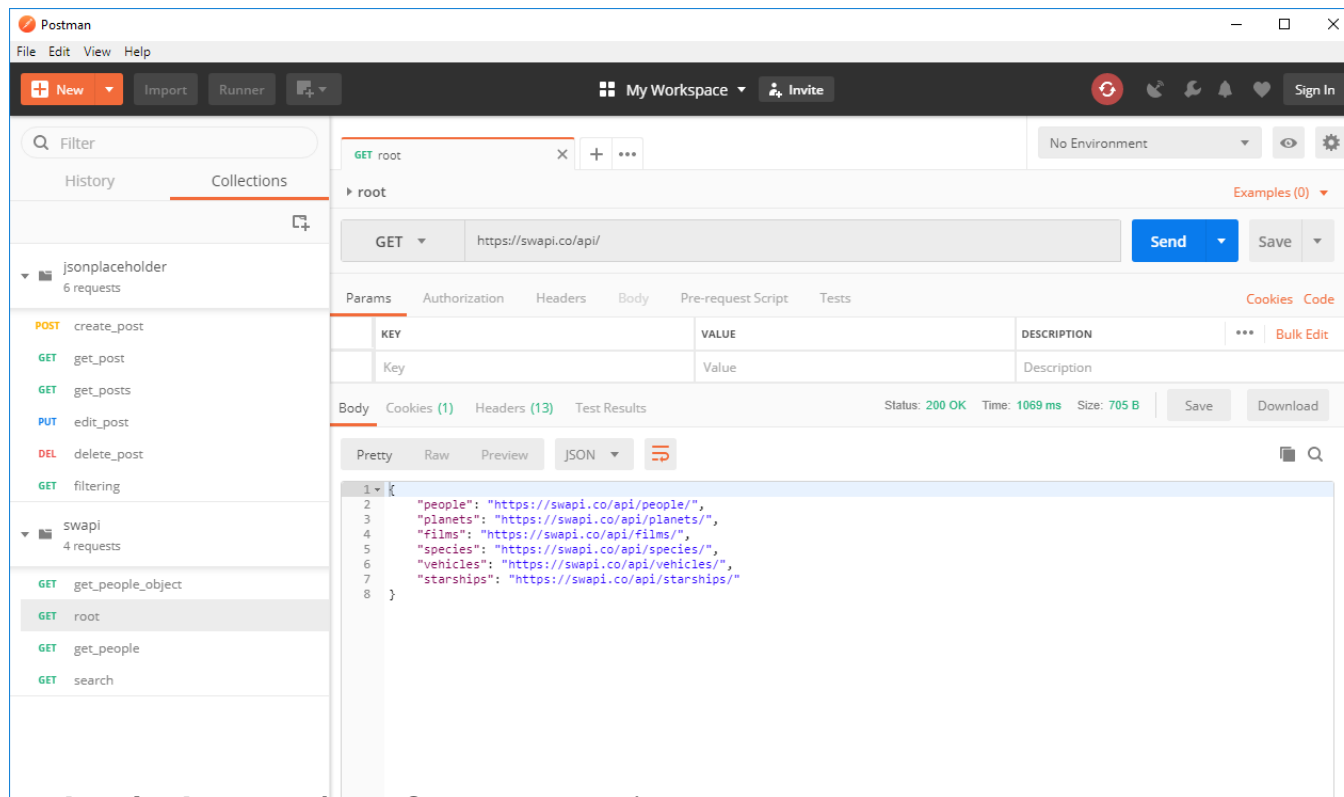
Metoda
GET

Status
odpowiedzi

Zawartość
odpowiedzi

URL: google.pl

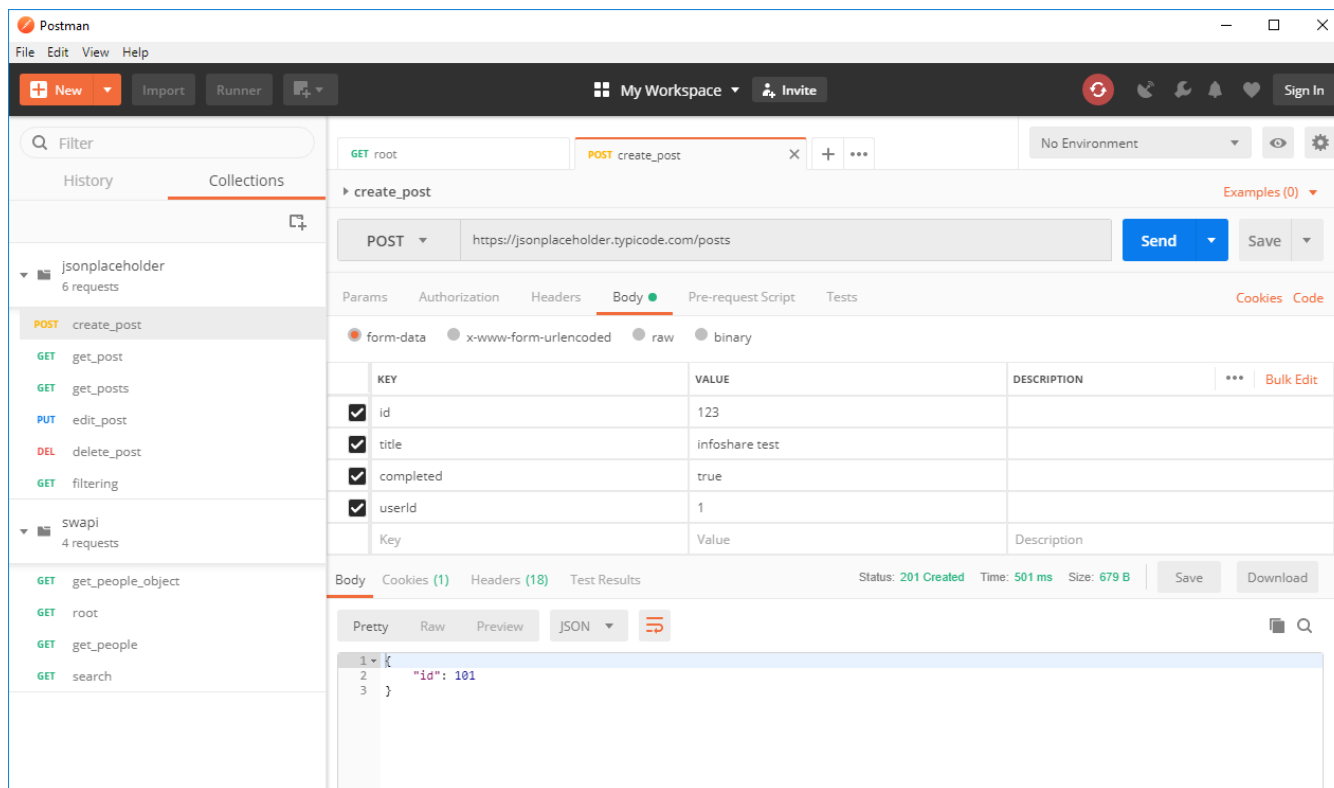
Postman - metoda GET



URL: <https://swapi.co/api>

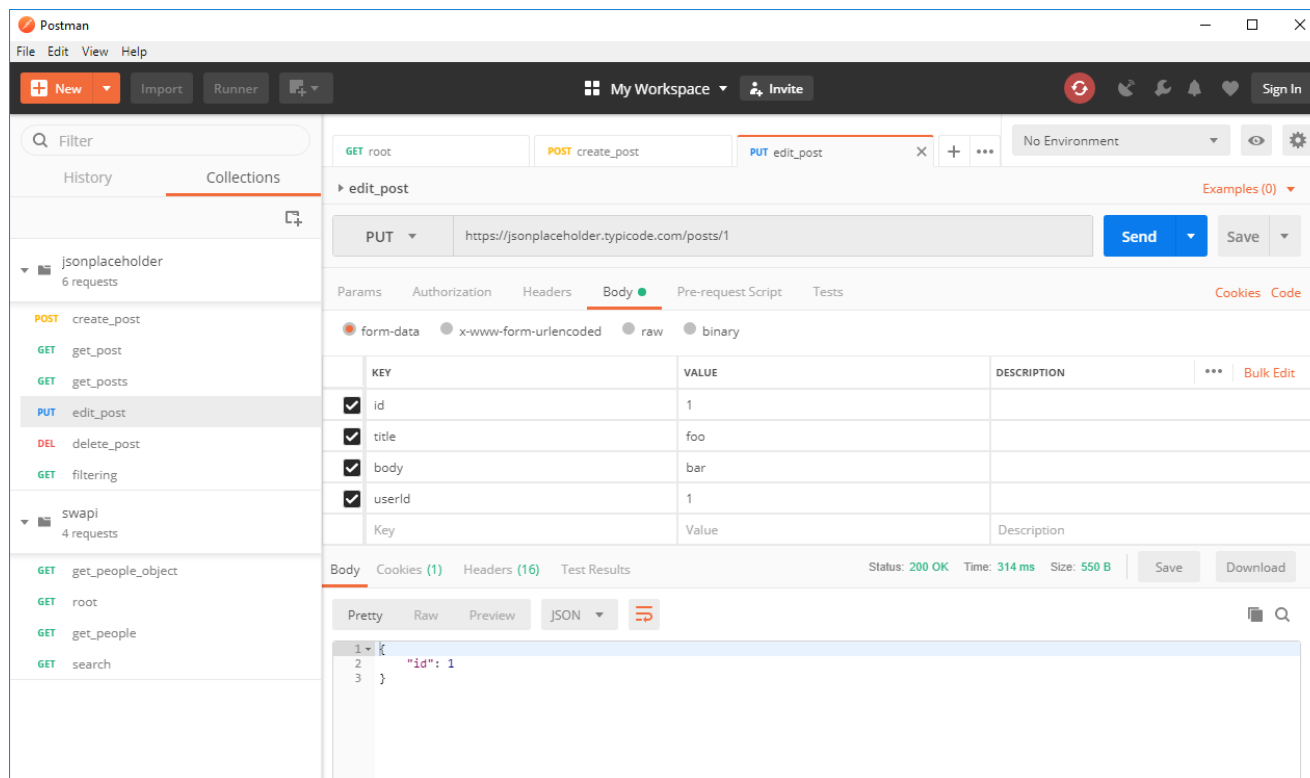
(The Star Wars API)

Postman - metoda POST



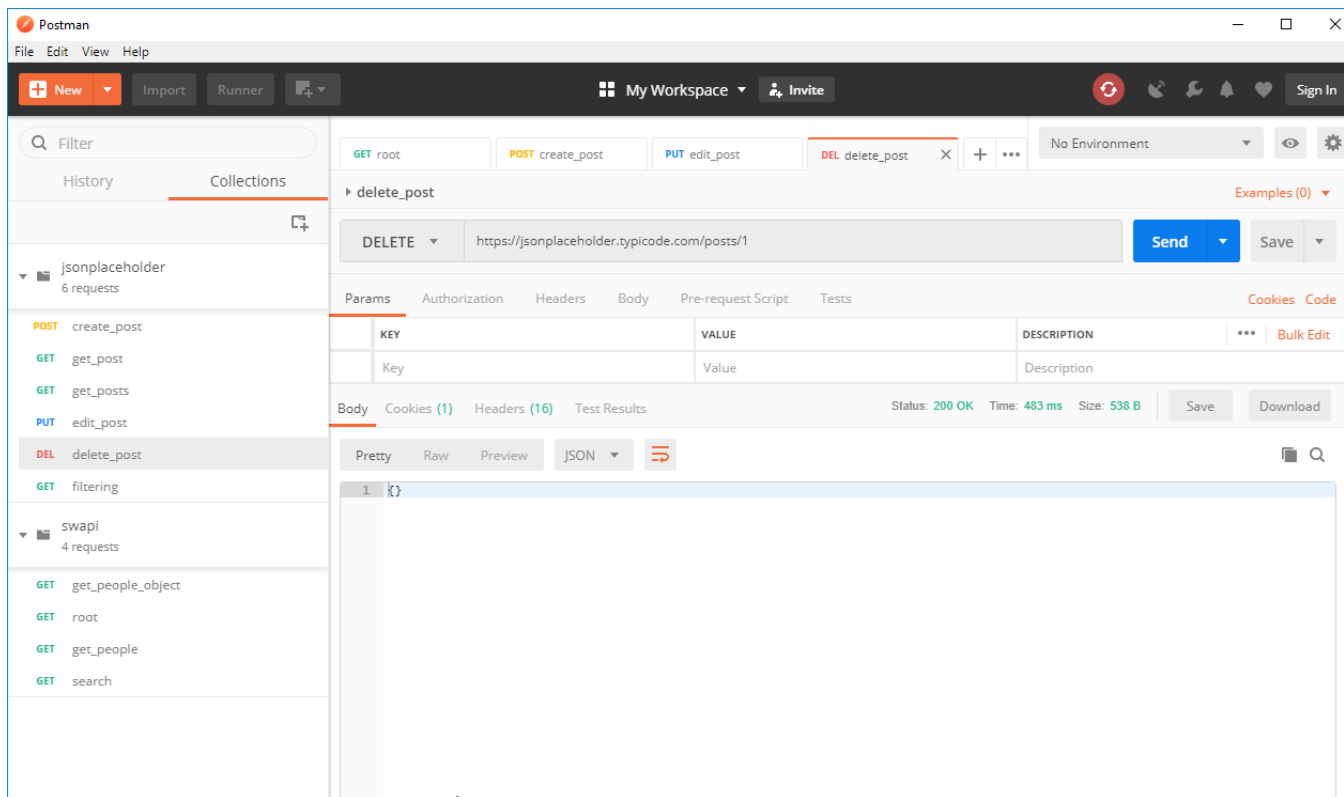
URL: <https://jsonplaceholder.typicode.com/posts>

Postman - metoda PUT



URL: <https://jsonplaceholder.typicode.com/posts/1>

Postman - metoda DELETE



URL: <https://jsonplaceholder.typicode.com/posts>

Narzędzia deweloperskie (F12)

- Podgląd i edycja zawartości strony oraz jej układu
- Konsola JavaScript
- Debugger JavaScript
- Monitor sieci
- Narzędzia monitorujące wydajność

Inne narzędzia

- Burp Proxy (<https://portswigger.net/burp>) – proxy i interceptor:
 - przechwytyje żądania,
 - umożliwia wielokrotne odtworzenie żądań,
 - zmianę treści
- curl – „scyzoryk” konsolowy
- wget – „scyzoryk” konsolowy

Ciasteczka

HTTP cookie



HTTP cookie (web cookie, browser cookie)

zawiera pewną niewielką część danych, którą serwer wysyła do przeglądarki użytkownika.

Przeglądarka może ją zapisać i przesłać przy kolejnym zapytaniu do tego samego serwera np. zapamiętanie czy użytkownik jest już zalogowany, koszyk w sklepie internetowym.