



## Webové API

Projekt do predmetu WAP

## Wikipedia API

Samuel Slávka (xslavk02)

Marek Šalgovič (xsalgo00)

April 25, 2021

## Wikipédia API

API je napísané v jazyku **JavaScript**, s využitím frameworku **Express.js**[1]. Informácie sa získavajú spracovaním HTML stránky článkov Wikipédie použitím *webscraperu* **Cheerio.js**[2]. Na získanie HTML dokumentu, asynchrónne vysielame GET HTTP správu využitím knižnice **Axios**[3].

### Vyhľadanie článku

Prvý a jeden z najdôležitejších dotazov je vyhľadanie článku na Wikipédii. Na vyhľadanie článku je využité oficiálne Wikipédia API.[4]. Dotaz obsahuje jeden povinný *query* parameter **search**, ktorý obsahuje reťazec použitý na vyhľadanie článku. Dotaz vyzerá nasledovne:

```
GET /search?search={search string}
```

Odpoveď obsahuje 5 najviac vyhovujúcich výsledkov - ich názvy a reťazec, ktorý je potrebný použiť v ďalších dotazoch na API:

```
{
  "searched_by": "{string used to search}",
  "best_match_article_name": "{name of article with best match}",
  "best_match_article_url": "{article string to call API with}",
  "suggested_articles": [{names of articles found with search}],
  "suggested_urls": [{list of article strings to call API with}]
}
```

Ak žiadny článok nevyhovuje vyhľadávaniu je navrátený HTTP status 404.

### Popis článku

Tento dotaz vráti základný popis článku - prvý paragraf. Dotaz vyzerá nasledovne:

```
GET /articles/{article}
```

Odpoveď je JSON s jedným *key-value* párom:

```
{
  "result": "{article description}",
}
```

## Obsah článku

Výsledkom je štrukturovaný obsah článku. Dotaz vyzerá nasledovne:

GET /articles/{article}/contents

V odpovedi sa nachádza rekurzívne zanorený nadpis sekcie resp. podsekcie spolu s jej indexom. Príklad odpovede:

```
{
  "contents": [
    {
      "index": "1",
      "name": "{section name}",
      "subsections": [
        {
          "index": "1.1",
          "name": "{subsection name}",
          "subsections": []
        }
      ]
    },
    {
      "index": "2",
      "name": "{section name}",
      "subsections": []
    }
  ]
}
```

## Obrázky v článku

Ďalšou užitočnou informáciou, ktorú je možné okrem textu v článku na Wikipédii nájsť, sú obrázky. Zdrojové URL obrázkov je možné získať dotazom:

GET /articles/{article}/images

Formát odpovede je nasledovný:

```
{
  "image": {image from the article main panel},
  "images": [list of images from article]
}
```

Odpoveď sa môže líšiť pri zmene jazyka parametrom `lang`, ktorý je popísaný nižšie, pretože články iných jazykov môžu obsahovať iné obrázky.

## Jazyk článku

Wikipédia podporuje viacero jazykov/lokalizácií pre niektoré články. Naše API podporuje dobrovoľný *query* parameter **lang**. Predvolený jazyk pri nezvolenom parametri je angličtina (en). Príklad použitia parametra.

```
GET /articles/{článok}/contents?lang=cs
GET /articles/{článok}/images?lang=sk
GET /articles/{článok}?lang=sk
```

## Záver

Nami implementované API ponúka získavanie dát nachádzajúcich sa v jednotlivých článkoch, narozdiel od oficiálneho API od Wikipédie, ktoré svojimi dotazmi ponúka hlavne meta-dáta o článkoch. Na rozdiel od iných implementácií je naše API pomerne nenáročné na zdroje, prehľadné a jednoduché na používanie.[5]. Ďalším rozdielom od iných API je štandardná forma odpovedí v podobe validných JSON odpovedí.

# Bibliography

- [1] Express.js,  
<https://expressjs.com/>
- [2] Cheerio.js,  
<https://cheerio.js.org/>
- [3] Axios,  
<https://github.com/axios/axios>
- [4] Mediawiki API,  
[https://www.mediawiki.org/wiki/API:Main\\_page](https://www.mediawiki.org/wiki/API:Main_page)
- [5] Java Wikipedia Library,  
<https://dkpro.github.io/dkpro-jwpl/>