

Get Data from the Web

with JavaScript and Node.js

LFU, March 13 & 27 2018

Why Bother Learning Node.js and JavaScript?

TAKE THE CHALLENGE AND BECOME

PROUD DYKE.TECH

[Solve 9 tasks and prove you master web technologies]

ABOUT THE PROJECT

START THE CHALLENGE

<https://www.dyke.tech>



DOGODKI

Lezbične in kvirovske intimnosti v prijateljskih odnosih

1.2.2018 LFU

Lezbično-feministična univerza vas vabi na izkustveno delavnico "Lezbične in kvirovske intimnosti v prijateljskih odnosih", ki bo v ponedeljek, 19.2.2017 ob 19h v klubu Tiffany (ŠKUC-Kulturni center Q), Metelkova mesto. Na delavnici bomo skušale raziskati raznolike vidike naših lezbičnih in kvir prijateljstev, kolegialnosti, tovarištev in ostalih tesnih odnosov. Skušale bomo začrtati odprt, širši spekter intimnosti in [...]



IZJAVE IN PROTESTI

Srečno 2018!

26. 12. 2017 LFU

V letu 2018 vam želimo srečno na celotnem lezbičnem in feminističnem kontinuumu! Lezbično-feministična univerza P.S.: Zahvaljujemo se anonimni avtorici grafitu "Živel lezbični kontinuum!" na zunanjem zidu tovarne Rog (december 2017) in Adrienne Rich za "lezbični kontinuum" (1980).

[READ MORE >](#)

0



LEZBIČNO-FEMINISTIČNA UNIVERZA (LFU)

- × avtonomen izobraževalni, teoretski, umetniški & akcijski prostor
- × raziskovanje in praksa lezbičnega feminizma, antikapitalizma & antifazišizma.

DRUŽBENA OMREŽJA

[Facebook](#)

[Twitter](#)

[Goodreads](#)

Tweets by @lezfemuniverza



LFU
@lezfemuniverza



Vabljenje_i na Pusstny benefit za 19. festival Rdeče zore! @Red_Dawns 110.1. od 21h naprej v Modri kot! facebook.com/events/1900630...

<https://www.lezfemuniverza.org/>

Sektor Ž

Podcast radijske vsebine z Radia Študent.



[Zelo kratek uvod v črni feminizem](#)

February 17, 2018, 12:00 PM

Kratek uvod v črni in afroameriški feminizem z vidika filma, literature in glasbe

Media files


 [2018-02-17-sektor-ž-zelo-kratek-uvod-v-črni-feminizem.mp3](#) (audio/mpeg, 146 MB)

[ZAKAJ FEMINIZEM DANES?](#)

January 20, 2018, 12:00 PM

O feministični ediciji revije Dialogi

Media files

 [2018-01-20-sektor-ž-zakaj-feminizem-danes.mp3](#) (audio/mpeg, 146 MB)

[Feministično zavzetje Ljubljane](#)

December 16, 2017, 12:00 PM

Feministična tura po Ljubljani in shod Noč je naša!

Media files

 [2017-12-16-sektor-ž-feministično-zavzetje-ljubljane.mp3](#) (audio/mpeg, 143 MB)

[https://radiostudent.si/dru%C5%BEba/sektor-%C5%B
E/podcast](https://radiostudent.si/dru%C5%BEba/sektor-%C5%B
E/podcast)

Sinister Wisdom

A Multicultural Lesbian Literary & Art Journal

[Journal](#) [About](#) [Subscribe](#) [Donate](#) [Volunteer](#) [Submit](#) [Shop](#) [Contact](#) [Blog](#)

[Home](#)

Sinister Wisdom
Newsletter

**Subscribe to our
email list**

* Indicates required

Email Address *

First Name

Last Name

Name

[View previous
campaigns.](#)

[Subscribe](#)

Archive



<http://www.sinisterwisdom.org/archive>

What Will We Learn Today?

- **Source Code and HTML**
- **Node.js**
- **Code Editor - Visual Studio Code**
- **JavaScript**
 - **Variables**
 - **Array**
 - **forEach**
 - **Function**
 - **Object**
- **Data from web pages**
 - **Get HTML**
 - **Get data from HTML**
 - **Download files**

Source code in browsers

- Right click > View Page Source

Or

- Right click > Inspect Element
- Right click > View Page Source


```

1 <!doctype html>
2 <html lang="en-US">
3 <head>
4 <meta charset="UTF-8">
5 <meta name="viewport" content="width=device-width, initial-scale=1">
6 <link rel="profile" href="http://gmpg.org/xfn/11">
7 <title>Dyke.tech</title>
8 <meta name="description" content="Dyke.tech is a lesbian who masters computer science. Test yourself and solve 9 tasks to get a certificate Proud Dyke.tech!"/>
9 <link rel="preload" href="https://www.dyke.tech/wp-content/themes/dyke/fonts/montserrat-v11-latin-regular.woff2" as="font">
10 <link rel="preload" href="https://www.dyke.tech/wp-content/themes/dyke/fonts/passion-one-v8-latin-regular.woff2" as="font">
11 <link rel="preload" href="https://www.dyke.tech/wp-content/themes/dyke/style.css" as="style">
12 <link rel="preload" href="https://www.dyke.tech/wp-content/themes/dyke/js/jquery-3.2.1.min.js" as="script">
13 <link rel="manifest" href="https://www.dyke.tech/wp-content/themes/dyke/manifest.json">
14 <meta property="og:locale" content="en US"/>
15 <meta property="og:site_name" content="Dyke.tech"/>
16 <meta property="og:title" content="Proud Dyke.tech"/>
17 <meta property="og:url" content="http://www.dyke.tech"/>
18 <meta property="og:type" content="website"/>
19 <meta property="og:description" content="Dyke.tech is a lesbian who masters computer science. Test yourself and solve 9 tasks to get a certificate Proud Dyke.tech!"/>
20 <meta property="og:image" content="http://www.dyke.tech/wp-content/uploads/2017/10/dyketech.png"/>
21 <meta itemprop="name" content="Proud Dyke.tech"/>
22 <meta itemprop="description" content="Dyke.tech is a lesbian who masters computer science. Test yourself and solve 9 tasks to get a certificate Proud Dyke.tech!"/>
23 <meta itemprop="image" content="http://www.dyke.tech/wp-content/uploads/2017/10/dyketech.png"/>
24 <meta name="twitter:title" content="Proud Dyke.tech"/>
25 <meta name="twitter:url" content="http://www.dyke.tech"/>
26 <meta name="twitter:description" content="Dyke.tech is a lesbian who masters computer science. Test yourself and solve 9 tasks to get a certificate Proud Dyke.tech!"/>
27 <meta name="twitter:image" content="http://www.dyke.tech/wp-content/uploads/2017/10/dyketech.png"/>
28 <meta name="twitter:card" content="summary_large_image"/>
29 <style>html{font-family:sans-serif;line-height:1.15;-ms-text-size-adjust:100%;-webkit-text-size-adjust:100%;body{margin:0}header{display:block}h1{font-size:2em;margin:.67em 0}a{background-color:.col,.col-1,.col-10,.col-11,.col-12,.col-2,.col-3,.col-4,.col-5,.col-6,.col-7,.col-8,.col-9,.col-lg,.col-lg-1,.col-lg-10,.col-lg-11,.col-lg-12,.col-lg-2,.col-lg-3,.col-lg-4,.col-lg-5,.col-lg-6
30 .col,.col-1,.col-10,.col-11,.col-12,.col-2,.col-3,.col-4,.col-5,.col-6,.col-7,.col-8,.col-9,.col-lg,.col-lg-1,.col-lg-10,.col-lg-11,.col-lg-12,.col-lg-2,.col-lg-3,.col-lg-4,.col-lg-5,.col-lg-6
31 <link rel="dns-prefetch" href="//s.w.org" />
32 <link rel="canonical" href="https://www.dyke.tech/" />
33 </head>
34 <body class="home page-template-default page page-id-11 page-parent">
35 <div id="site" class="container-fluid nostyle">
36 <header class="row">
37 <div class="col-12">
38 <h1><a href="https://www.dyke.tech"><span>&lt;</span>DYKE.tech<span>&gt;</span></a></h1>
39 </div>
40 </header>
41 <div id="content" class="row justify-content-center">
42 <div class="col-10">
43 <p class="top">TAKE THE CHALLENGE AND BECOME</p>
44 <h2 class="shadow">PROUD Dyke.tech</h2>
45 <p class="tip">[ Solve 9 tasks and prove you master web technologies ]</p>
46 <div class="row justify-content-center">
47 <a class="btn btn-default btn-about" href="https://www.dyke.tech/about-the-project/">About the project</a><a href="https://www.dyke.tech/test/level01-9975280837471640/" class="btn btn-default"
48 </div>
49 </div>
50 </div>
51 </div>
52 <script>var _paq=_paq||[];_paq.push(['trackPageView']);_paq.push(['enableLinkTracking']);(function(){var u="//www.dyke.tech/piwik/";_paq.push(['setTrackerUrl',u+'piwik.php']);_paq.push(['setSite
53 window.addEventListener("load",function(event){
54     var elem_site=document.getElementById("site");
55     elem_site.classList.remove("nostyle");
56 });
57 </script>
58 </body>
59 </html>

```

HTML

Elements

- body
- H1
- a
- ...

Attributes

- class= .
- id= #
- style=
- src=

Node.js

Node is an open-source, cross-platform, runtime environment that allows developers to create tools and applications in JavaScript.

Download from

<https://nodejs.org>



The screenshot shows the Node.js website homepage. At the top is a dark navigation bar with the Node.js logo and links for HOME, ABOUT, DOWNLOADS, DOCS, GET INVOLVED, SECURITY, NEWS, and FOUNDATION. The main content area has a white background. It starts with a paragraph describing Node.js as a JavaScript runtime built on Chrome's V8 JavaScript engine, highlighting its event-driven, non-blocking I/O model and its package ecosystem, npm. Below this is a light green banner with the text "Spectre and Meltdown in the context of Node.js.". Underneath the banner, the text "Download for Linux (x64)" is centered. Two green buttons are displayed: "8.9.4 LTS Recommended For Most Users" and "9.5.0 Current Latest Features". At the bottom, there are links for "Other Downloads", "Changelog", and "API Docs" for both versions.

node.js

HOME | ABOUT | DOWNLOADS | DOCS | GET INVOLVED | SECURITY | NEWS | FOUNDATION

Node.js® is a JavaScript runtime built on **Chrome's V8 JavaScript engine**. Node.js uses an event-driven, non-blocking I/O model that makes it lightweight and efficient. Node.js' package ecosystem, **npm**, is the largest ecosystem of open source libraries in the world.

Spectre and Meltdown in the context of Node.js.

Download for Linux (x64)

8.9.4 LTS
Recommended For Most Users

9.5.0 Current
Latest Features

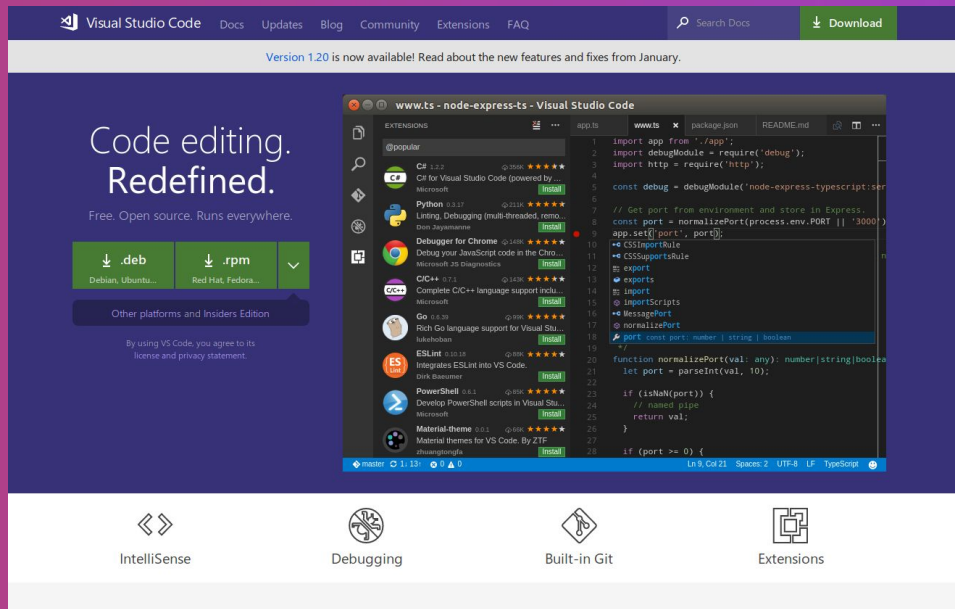
Other Downloads | Changelog | API Docs Other Downloads | Changelog | API Docs

Install code editor

Visual Studio Code

Download from

code.visualstudio.com



Download files for the workshop

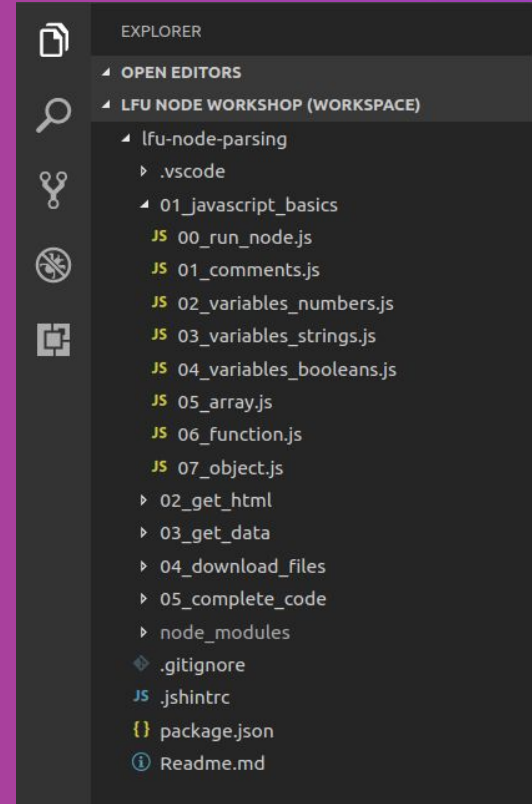
Download files from

goo.gl/vwvhBU

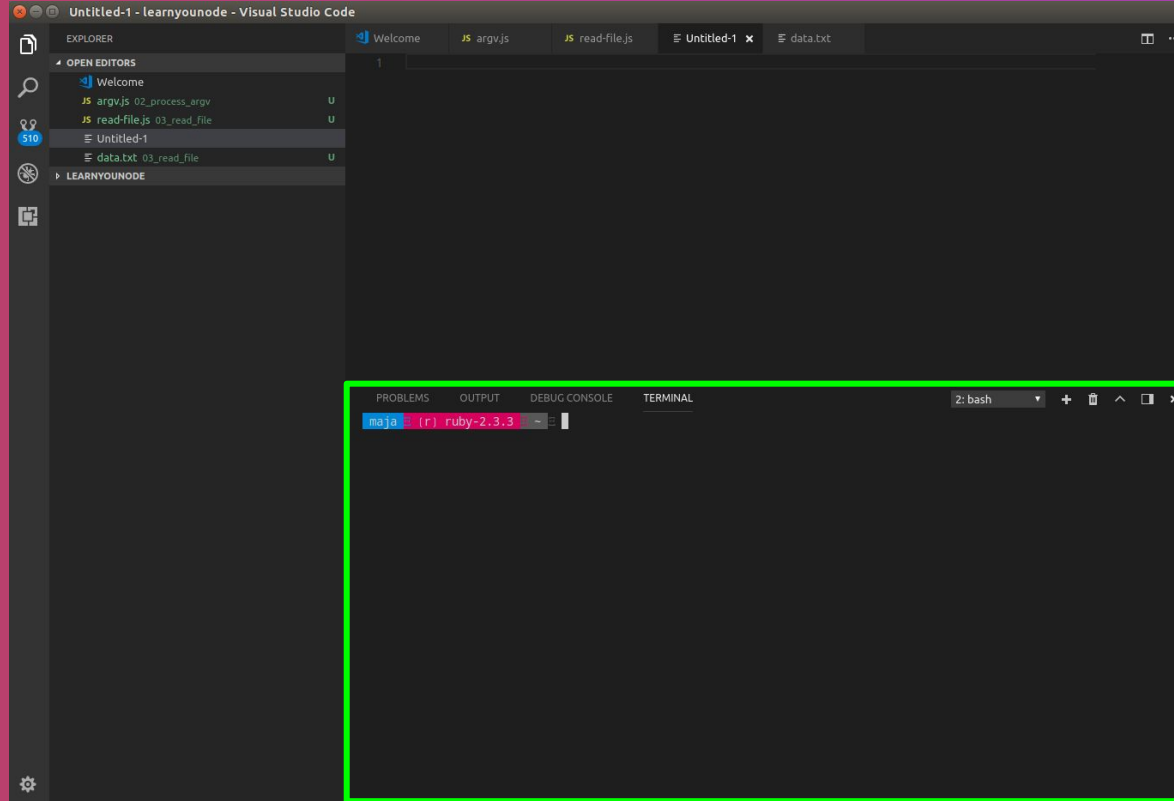
<https://github.com/22nds/lfu-node-parsing/archive/master.zip>

Unzip and Open folder with code in Code Editor

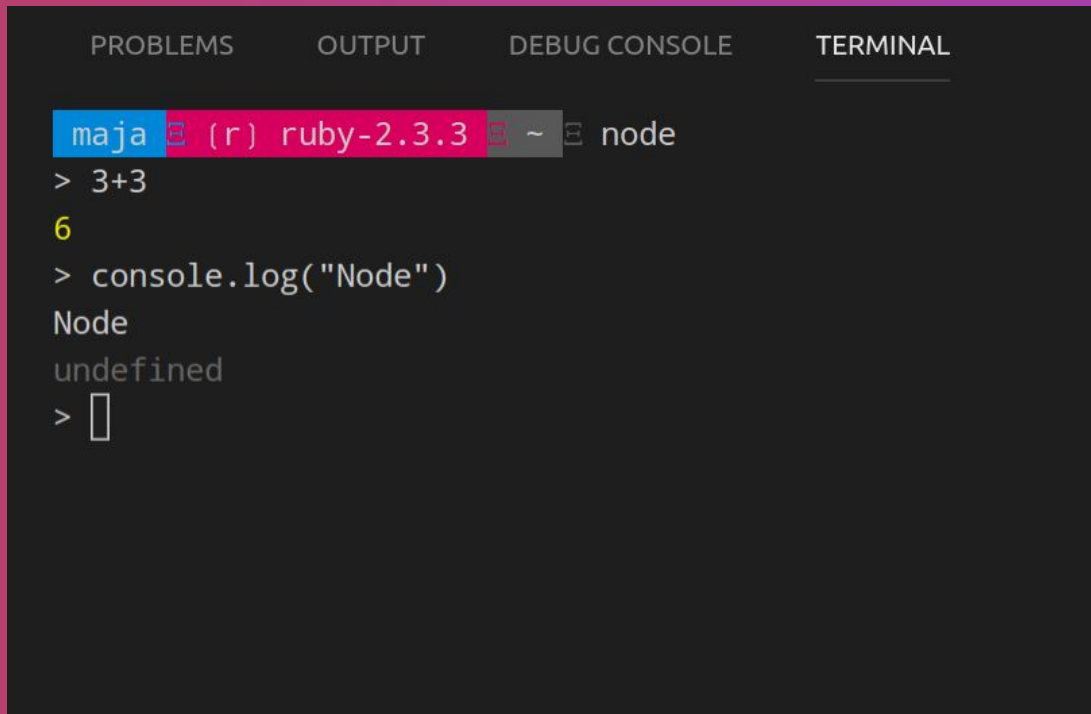
File > Open Folder...



View > Integrated terminal



View > Integrated terminal



The screenshot shows the Integrated Terminal interface in VS Code. At the top, there are four tabs: PROBLEMS, OUTPUT, DEBUG CONSOLE, and TERMINAL. The TERMINAL tab is selected and underlined. The terminal window has a title bar that reads 'maja (r) ruby-2.3.3 ~ node'. The terminal content shows a series of commands and their outputs: a prompt '>' followed by '3+3' which outputs '6', then a prompt '>' followed by 'console.log("Node")' which outputs 'Node', then a prompt '>' followed by 'undefined', and finally a prompt '>' followed by an empty line.

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL
maja (r) ruby-2.3.3 ~ node
> 3+3
6
> console.log("Node")
Node
undefined
> 
```

Install dependencies in packages.json

npm install

```
{ } package.json x
1  {
2    "name": "lfu-node-parsing",
3    "version": "1.0.0",
4    "description": "Workshop - Learning web page parsing with Node.js",
5    "main": "index.js",
6    "scripts": {
7      "test": "echo \"Error: no test specified\" && exit 1",
8      "start": "node ./index.js"
9    },
10   "author": "Maja Kraljic",
11   "license": "ISC",
12   "dependencies": {},
13   "keywords": [],
14   "devDependencies": {
15     "async": "^2.6.0",
16     "axios": "^0.17.1",
17     "cheerio": "^1.0.0-rc.2",
18     "colors": "^1.1.2",
19     "download": "^6.2.5"
20   }
21 }
```


Open a file in Code Editor

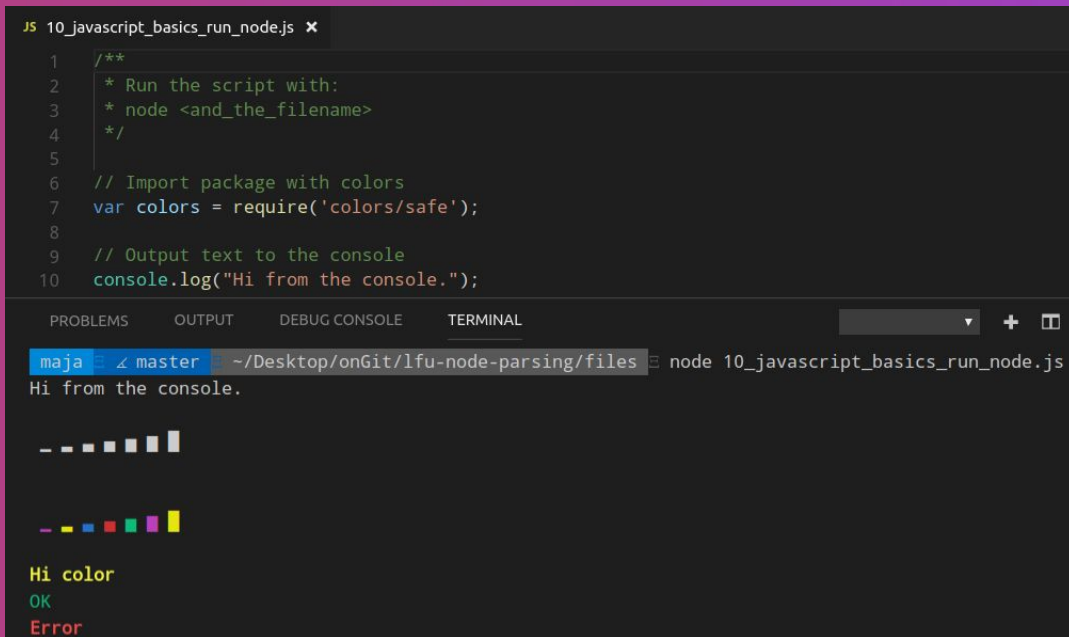
Change directories with:

`cd files`

Run script with:

`node <filename.js>`

Use **tab** for completion of
file/directory names



The screenshot shows the Visual Studio Code editor interface. The top pane displays a JavaScript file named `10_javascript_basics_run_node.js` with the following content:

```
1  /**
2   * Run the script with:
3   * node <and_the_filename>
4   */
5
6  // Import package with colors
7  var colors = require('colors/safe');
8
9  // Output text to the console
10 console.log("Hi from the console.");
```

The bottom pane shows the **TERMINAL** tab. The terminal prompt is `maja ~ master` and the command executed is `~/Desktop/onGit/lfu-node-parsing/files node 10_javascript_basics_run_node.js`. The output of the command is:

```
Hi from the console.

- - - - -

- - - - -

Hi color
OK
Error
```

JavaScript

Variables

Number

```
var number = 1;  
number = 5;  
console.log(number);
```

String

```
var word;  
word = 'LFU';  
word.length;  
console.log(word);
```

Boolean

```
var bool = true;  
bool = false;  
console.log(bool);
```

Array

```
var list = [1, 2, 3];  
list[0];  
list.length;  
console.log(list);
```

forEach


```
var list = [1, 2, 3];  
list.forEach(  
  item => console.log(item);  
);
```

```
var list = [1, 2, 3];  
list.forEach(  
    item => {  
        console.log(item);  
    }  
);
```

Function

```
function square(num) {  
    return num * num;  
}
```

```
var result = square(5);  
console.log(result);
```

Object

```
var myObj = {  
  name: "LFU",  
  location: "Ljubljana"  
};  
myObj.name
```

Let's go online!

Get HTML

with axios

<https://github.com/axios/axios>

Axios

```
const axios = require('axios');
axios.get("https://www.lezba.si")
  .then((response) => {
    console.log(response)
    console.log(response.config.url)    // Address of the page
    console.log(response.status)        // Server status 200 is OK
    console.log(response.data)          // All HTML of the site
  })
  .catch((err) => {
    console.log(err);
  })
```

Parse HTML

with cheerio

<https://github.com/cheeriojs/cheerio>

Get data: link URL

```
const cheerio = require('cheerio');

const $ = cheerio.load('<a  
href="https://www.dyke.tech">Dyke.tech</a>');
var link = $('a').attr('href').text();
console.log(link);
```

Cheerio

```
$( 'h2' )
```

```
$( 'div li' )
```

```
$( '#page' )
```

```
$( '.black' )
```

Get data: Title

```
const axios = require('axios');
const cheerio = require('cheerio');

axios.get("https://www.lezba.si")
  .then((response) => {
    const $ = cheerio.load(response.data);
    var title = $('h1').text();
    console.log(title);
  })
  .catch((err) => {
    console.log(err);
  })
```

Get data: mp3 links

```
const axios = require('axios');
const cheerio = require('cheerio');

axios.get("https://radiostudent.si/dru%C5%BEba/sektor-%C5%BE/podcast")
  .then((response) => {
    const $ = cheerio.load(response.data);
    var mp3s = [];
    $('enclosure').each(function(i, elem) {
      var link = $(this).attr('url');
      mp3s.push(link);
    });
    console.log(mp3s);
    console.log(mp3s.length);
  })
  .catch((err) => {
    console.log(err);
  })
```

Download Files

with progress-download

<https://www.npmjs.com/package/progress-download>

Download one file

```
const download = require('progress-download');

var file = 'https://radiostudent.si/sites/default/files/posnetki/Sektor%20%C5%BD/2012-12-15-sektor-%C5%BE-feministke-vstanite.mp3'

// download the file in folder downloads
download(file, './downloads').then(() => {
  console.log('Downloaded: ' + file);
});
```


Download multiple files

```
const download = require('progress-download');  
var files = [  
  'https://www.lezfemuniverza.org/wp-content/uploads/2018/02/lesbians-1974-SF-Gay-Parade-Rink-Foto-1-e1517516581958.jpg',  
  'https://www.lezfemuniverza.org/wp-content/uploads/2017/12/grafit-zid-Rog-dec-2017.jpg',  
  'https://www.lezfemuniverza.org/wp-content/uploads/2017/11/Lesbian-couple-at-Le-Monocle-Paris-1932.jpg',  
  'https://www.lezfemuniverza.org/wp-content/uploads/2017/08/dyke-hard-2-e1502827128353.jpg'  
]  
files.forEach((file) => {  
  download(file, './images').then(() => {  
    console.log('Downloaded: ' + file);  
  });  
})
```

Download Big Files

one after another with async

<https://github.com/caolan/async>

Download big files one after another

```
const async = require('async');
var mp3s = [
  'https://radiostudent.si/sites/default/files/posnetki/Sektor%20%C5%BD/2018-02-17-sektor-%C5%BE-zelo-kratek-uvod-v-%C4%8Drni-feminizem.mp3',
  'https://radiostudent.si/sites/default/files/posnetki/Sektor%20%C5%BD/2017-06-17-sektor-%C5%BE-prihodnost-brez-pravic.mp3'
]
async.everySeries(mp3s, (mp3, callback) => {
  console.log("Downloading: " + mp3);
  download(mp3, './mp3s').then(() => {
    console.log('DONE. ');
    callback(null, true)
  })
}, function(err, result) {
  console.log('All files are downloaded. ');
});
```

What We Learned?

- **Source Code and HTML**
- **Installed Node.js**
- **Installed Code Editor Visual Studio Code**
- **JavaScript Basics**
- **How to get data/download files from web pages?**
 - **Get HTML**
 - **Get information we need**
 - **Download files**

**What Will You
Download Next?**

Resources



JavaScript

Languages

Edit

[Web technology for developers](#) > JavaScript

Related Topics

JavaScript

Tutorials:

- ▶ Complete beginners
- ▶ JavaScript Guide
- ▶ Intermediate
- ▶ Advanced

References:

- ▶ Built-in objects
- ▶ Expressions & operators
- ▶ Statements & declarations
- ▶ Functions
- ▶ Classes
- ▶ Errors
- ▶ Misc
- ▶ New in JavaScript

Documentation:

- ▶ Useful lists
- ▶ Contribute

JavaScript (JS) is a lightweight interpreted or JIT-compiled programming language with first-class functions. While it is most well-known as the scripting language for Web pages, [many non-browser environments](#) also use it, such as [Node.js](#), [Apache CouchDB](#) and [Adobe Acrobat](#). JavaScript is a prototype-based, multi-paradigm, dynamic language, supporting object-oriented, imperative, and declarative (e.g. functional programming) styles. Read more [about JavaScript](#).

This section is dedicated to the JavaScript language itself, and not the parts that are specific to Web pages or other host environments. For information about APIs specific to Web pages, please see [Web APIs](#) and [DOM](#).

The standard for JavaScript is [ECMAScript](#). As of 2012, all [modern browsers](#) fully support ECMAScript 5.1. Older browsers support at least ECMAScript 3. On June 17, 2015, [ECMA International](#) published the sixth major version of ECMAScript, which is officially called ECMAScript 2015, and was initially referred to as ECMAScript 6 or ES6. Since then, ECMAScript standards are on yearly release cycles. This documentation refers to the latest draft version, which is currently [ECMAScript 2018](#).

Do not confuse JavaScript with the [Java programming language](#). Both "Java" and "JavaScript" are trademarks or registered trademarks of Oracle in the U.S. and other countries. However, the two programming languages have very different syntax, semantics, and uses.

Tutorials

Reference

<https://developer.mozilla.org/en-US/docs/Web/JavaScript>



NODESCHOOL

Open source workshops that teach web software skills. Do them on your own or at a workshop nearby.

[Tutorials](#)[Events](#)[Chapters](#)[About](#)[Build](#)[Host](#)

Get started by installing one of our **core workshops** or subscribe to our free **email newsletter**.

<https://nodeschool.io/>

Questions?

maja.kraljic@gmail.com