JavaScript Exercise 2 :

Extract Data from Wikipedia based on a Search String.

Purpose.

The purpose of this exercise is to start retrieving data from an external source via a REST Webservice call. We extend our HTML and JavaScript of the first exercise by adding a new service 'Search Wiki'. We allow to enter a string for which we search in Wikipedia. The search results (JSON) returned by the service will be shown as table (popup) in our DHBW portal.

Pre-Requisites

We make our modifications in the former Ex1JSON.JS (there are no more templates) which we rename to something like myJavaScript.JS.

- ../Starter/public/js/myJavaScript.js (former Ex1JSON, don't forget to change the HTML for loading it)
- ../Starter/public/htm/webex2.htm (or whatever you used as your latest HTML file)
- ../Starter/public/Stylesheet/your latest stylesheet

Exercise

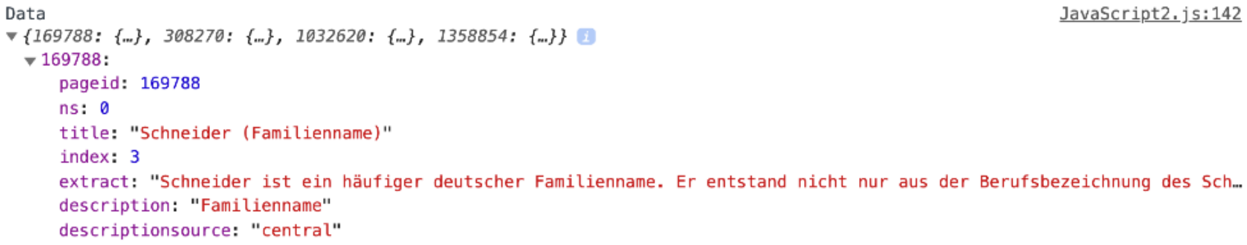
Wikipedia offers REST APIs (HTTPS Get) in order to retrieve Wikipedia data. After a complete revision by end of 2019 the best description is here.
https://www.mediawiki.org/wiki/API:Query

The URI you need to construct looks like this

- URI = https://de.wikipedia.org/w/api.php (The API Server)
  - Query String (A Query string starts with ? And each key/value pair is separated by a &)
    - action=query (requests a search)
    - generator=prefixsearch (This allows you enter your search string a bit fuzzy)
    - gpslimit=4 (limits the # of results to maximal 4)
    - format=json (The return format as JSON String)

- prop=extracts%7Cdescription (requests text (extract) and a main line(description))
- exintro=1 (requests only the introduction chapter)
- explaintext=1 (the returned data is just a plain text)
- exsentences=3 (we limit the extract to max. 3 sentences (or dots found))
- **gpssearch=<search> (the search string, the data you entered)**

The returned JSON (if you are successful) is as follows.



```
Data                                                                                          JavaScript2.js:142
▼{169788: {…}, 308270: {…}, 1032620: {…}, 1358854: {…}} ℹ
  ▼169788:
     pageid: 169788
     ns: 0
     title: "Schneider (Familienname)"
     index: 3
     extract: "Schneider ist ein häufiger deutscher Familienname. Er entstand nicht nur aus der Berufsbezeichnung des Sch…
     description: "Familienname"
     descriptionsource: "central"
```

It returns an object having maximal 4 key/values the value being again an object describing the data (the search hit) we need.

Each hit is an object consisting of:

- pageid (the unique number of the wiki page found, we need this number to construct the url.)
- title (a string showing how the search string has been translated)
- description (a string we can consider as a main line of the page)
- extract (a string with a longer text (which is limited until 3 dots are found)

The URL to the Wiki page you found is

https://de.wikipedia.org/?curid=pageid

Your task is to request a search argument (the search string) from your client (as HTML form), issue the REST call, and move the JSON data to a HTML table (similar to the previous exercise).

In this HTML table you should show the title, the description, the extract and the constructed URL as column (as shown earlier).

As this exercise extends the first we have a second entry in the pull down window asking for this search string.

Wikipedia does not allow CORS, that means that you cannot call above URL directly from the browser. However our local Node.js Server can be used as Proxy. That means we ask the local Node.js Server to invoke the Wikipedia on behalf of the client code. For that we just suffix above url to the following string http://localhost:6001/proxy/?....

to something like

```
http://localhost:6001/proxy/?https://de.wikipedia.org/w/api.php?action=query&generator=prefixsearch&format=json&gpslimit=4&prop=extracts%7Cdescription&exintro=1&explaintext=1&exsentences=3&redirects=1&gpssearch=Schneider
```

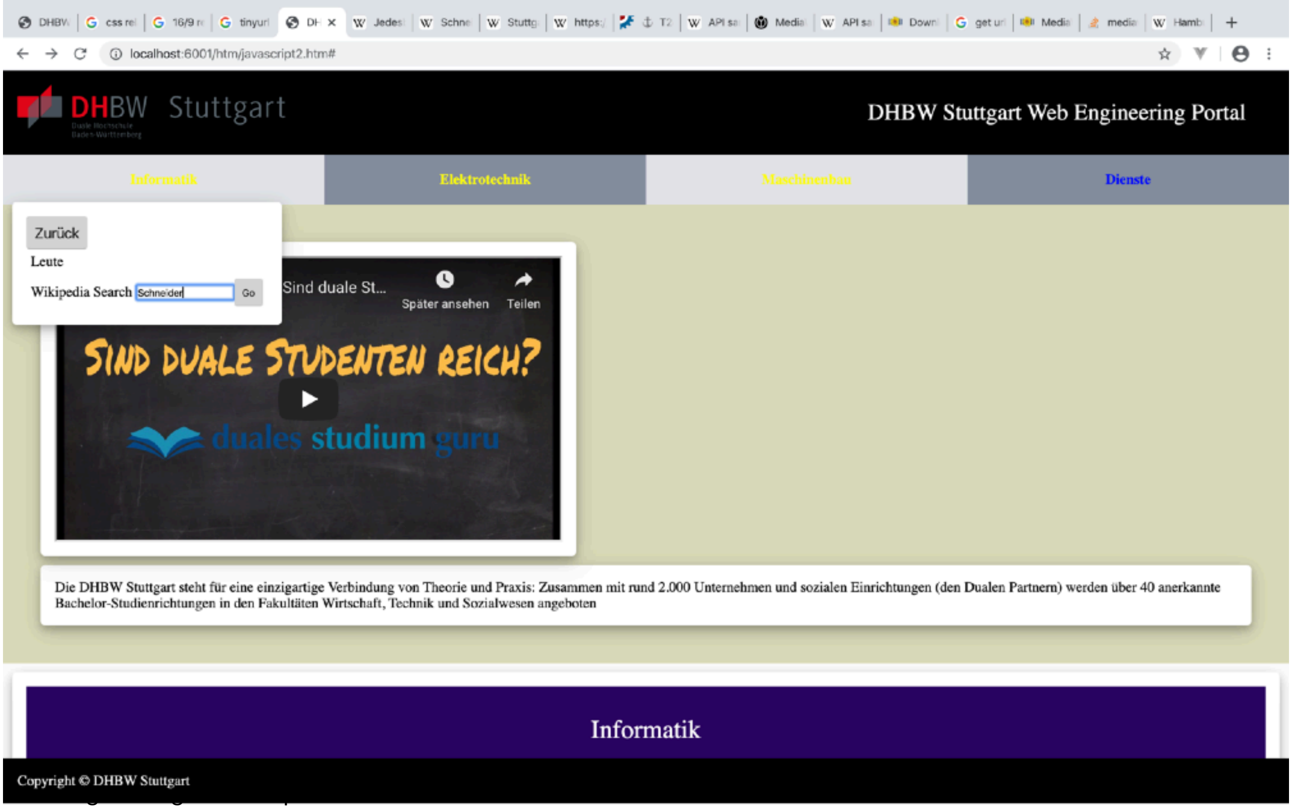
The first part sends the request to our local Node.js Server. This Server will then extract the real Wiki URL for us and forward that accordingly. The local Server will store the wiki response, as response object in an overall object as

- error (object is set when there was a problem with the proxy). The particular problem will be explained in error.message
- status (Status of the Proxy request, could be ignored in our case)
- request (the original request object, for our debugging)
- response (The Wiki Response object, which we need to further process)

In our lesson we talked for quite some time about the JavaScript Implementation of a REST call (AJAX). It is now your turn to code such an invocation as part of your new JavaScript file.

https://www.w3schools.com/js/js_ajax_intro.asp

When you succeed, you will see the following pages.



localhost:6001/htm/javascript2.htm#

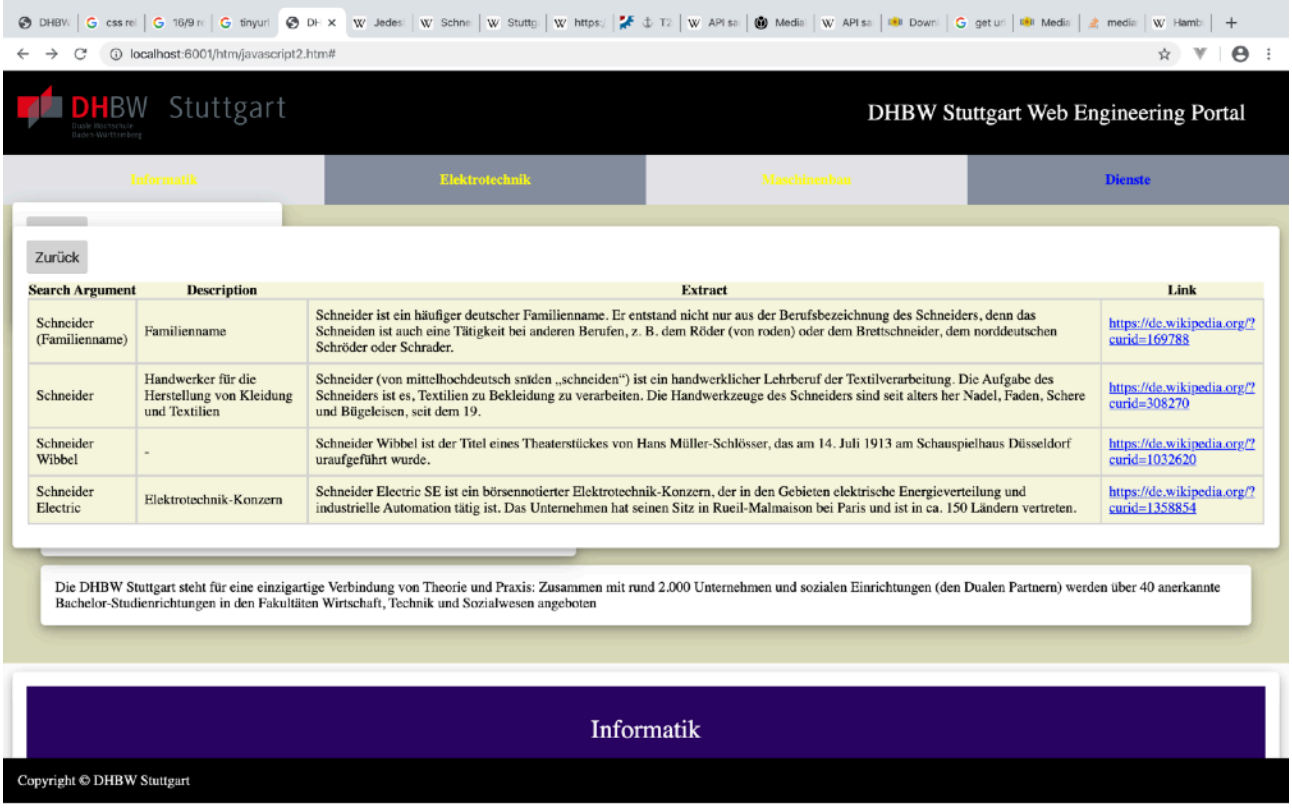
DHBW Stuttgart Web Engineering Portal

Wikipedia Search Schneider

Die DHBW Stuttgart steht für eine einzigartige Verbindung von Theorie und Praxis: Zusammen mit rund 2.000 Unternehmen und sozialen Einrichtungen (den Dualen Partnern) werden über 40 anerkannte Bachelor-Studienrichtungen in den Fakultäten Wirtschaft, Technik und Sozialwesen angeboten

Copyright © DHBW Stuttgart

In this case I requested a search with “Schneider” as string. And the result is like


localhost:6001/htm/javascript2.htm#
DHBW Stuttgart
DHBW Stuttgart Web Engineering Portal
Informatik
Elektrotechnik
Maschinenbau
Dienste
Zurück
Informatik
Copyright © DHBW Stuttgart

Search Argument	Description	Extract	Link
Schneider (Familienname)	Familienname	Schneider ist ein häufiger deutscher Familienname. Er entstand nicht nur aus der Berufsbezeichnung des Schneiders, denn das Schneiden ist auch eine Tätigkeit bei anderen Berufen, z. B. dem Röder (von roden) oder dem Brettschneider, dem norddeutschen Schröder oder Schrader.	https://de.wikipedia.org/?curid=169788
Schneider	Handwerker für die Herstellung von Kleidung und Textilien	Schneider (von mittelhochdeutsch snīden „schneiden“) ist ein handwerklicher Lehrberuf der Textilverarbeitung. Die Aufgabe des Schneiders ist es, Textilien zu Bekleidung zu verarbeiten. Die Handwerkzeuge des Schneiders sind seit alters her Nadel, Faden, Schere und Bügeleisen, seit dem 19.	https://de.wikipedia.org/?curid=308270
Schneider Wibbel	-	Schneider Wibbel ist der Titel eines Theaterstückes von Hans Müller-Schlösser, das am 14. Juli 1913 am Schauspielhaus Düsseldorf uraufgeführt wurde.	https://de.wikipedia.org/?curid=1032620
Schneider Electric	Elektrotechnik-Konzern	Schneider Electric SE ist ein börsennotierter Elektrotechnik-Konzern, der in den Gebieten elektrische Energieverteilung und industrielle Automation tätig ist. Das Unternehmen hat seinen Sitz in Rueil-Malmaison bei Paris und ist in ca. 150 Ländern vertreten.	https://de.wikipedia.org/?curid=1358854

Die DHBW Stuttgart steht für eine einzigartige Verbindung von Theorie und Praxis: Zusammen mit rund 2.000 Unternehmen und sozialen Einrichtungen (den Dualen Partnern) werden über 40 anerkannte Bachelor-Studienrichtungen in den Fakultäten Wirtschaft, Technik und Sozialwesen angeboten

Lessons learned

- We use the xmlHTTPRequest Object in order to issue the Get and the onreadystatechange callback handler. This handler is asynchronous and is executed on each object status change. Therefore it must be established before the Open and Send method are invoked (both will invoke the handler). The handler is also called when the response data is available. For that case we need to code the processing of that data.
- This actually means that significant work is done in the handler as you have the data to further process only there. People call it the callback hell, because if you need to read other data based on some data of the first data a new request is triggered in the handler invoking again another callback and so on. (Google Chrome gives some warning logs if it takes too long)
- Look into your Node.js Server to see how we implemented the Proxy Function. It is actually quite simple as we take the real Wikipedia URL and make yet another HTTPS request using the Request middleware package. The data returned from that Request method is again returned to the client invocation.

- We know that we would need more coding to cover all potential error paths.