|  |
| --- |
| Faculty of Applied Sciences and Technology |
| **XML/JSON Data Processing** |
| ITC5202 - Project |
|  |
| **Group 14: Kartik Nain (N01516607), Dhruvi Rajput (N01496176)**  **Submission Date** |
| **8/8/2022** |

|  |
| --- |
| This document explains how to process XML/JSON data …………………………. |

Table of Contents

[Step 1 2](#_Toc110423089)

[Step 2 3](#_Toc110423090)

[Step 3 4](#_Toc110423091)

[Step 4 5](#_Toc110423092)

[Step 5 6](#_Toc110423093)

[Summary 7](#_Toc110423094)

# Step 1 : Explore searching on reddit API: search functionality.

**What is the difference of the followings? Observe the output and explain**

[**https://www.reddit.com/search?q=ajax**](https://www.reddit.com/search?q=ajax)

**We'll search the most recent comments for the word "ajax" in the comment body. The word "ajax" will be found regardless of capitalization because this search is case-insensitive. The API sorts by most recent comments first by default.**

[**https://www.reddit.com/search.json?q=ajax**](https://www.reddit.com/search.json?q=ajax)

**In this, we are getting the above file in JSON format.**

**• Develop a simple web application that gets user input and searches for related articles on Reddit**

**o Display the article title, author, subreddit, URL, score, and flair, in a proper format**

**o Try to sort data based on the title**

**o Try to display 100 records (no of records in search result)**

**Solution:-**

[**https://www.reddit.com/search?q=ajax**](https://www.reddit.com/search?q=ajax)

We'll search the most recent comments for the word "ajax" in the comment body. The word "ajax" will be found regardless of capitalization because this search is case-insensitive. The API sorts by most recent comments first by default.

[**https://www.reddit.com/search.json?q=ajax**](https://www.reddit.com/search.json?q=ajax)

In this, we are getting the above file in JSON format.

**Code explanation:**

First, input from user is taken which is stored in a variable. Then, this variable is used in the fetch api link using template literals. The limit of no. of results is also set to 100. The data from the fetch api is passed to a function in which first, a table is created with the required headings. The data is sorted based on the title using sort method. Then, a for loop runs until the length of the “children” array which is found in the data received from the fetch api. Then, required information like title, author name etc. is extracted from the data and appended to the table.

Data validation has applied when the user doesn’t enter any input and errors arising from the api call are also handled.

Graphical user interface, text, application, table

Description automatically generated

# Step 2 : Explore using push shift API in order to search on comments and subreddits

* **Observe the output and explain how the following command works**
* **https://api.pushshift.io/reddit/search/comment/?q=science**
* **Read the API document for more info:** [**https://github.com/pushshift/api**](https://github.com/pushshift/api)

• **Develop a simple web application that gets user input and searches for related articles on Reddit.**

* **Display the article body, author, subreddit, URL, score, and flair, in proper**
* **Try to sort data based on title Try to display 100 records (no of records in search result)**

**Solution:-**

**Pushshift:**

Researchers now have access to Reddit data thanks to Pushshift, a social media data analysis and archiving platform. Real-time updates are made to the historical data in Pushshift's Reddit dataset.

**The Reddit comment and submission database can be accessed via one of two methods.**

One involves using the API directly, while the other entails gaining access to the Elasticsearch back-end search engine.

**https://api.pushshift.io/reddit/search/comment/?q=science**

We'll search the most recent comments for the word "science" in the comment body. The word "science" will be found regardless of capitalization because this search is case-insensitive. The API sorts by most recent comments first by default. There are 25 results returned by this search, which are displayed. This default size for searches can be modified using the size option. More details on this are provided in the parameters section. The "data" key contains the actual search results, which are returned in JSON format. Additionally, a "metadata" key offers additional information about the search, including the total number of objects found, processing time, etc. whenever aggregates.

**Code explanation:**

The logic for this is almost same as that of step 1. First, input from user is taken which is stored in a variable. Then, this variable is used in the fetch api link using template literals. Here wea re using pushshift api to search on comments. The limit of no. of results is also set to 100. The data from the fetch api is passed to a function in which first, a table is created with the required headings. The data is sorted based on the title using sort method. Then, a for loop runs until the length of the “children” array which is found in the data received from the fetch api. Then, required information like title, author name etc. is extracted from the data and appended to the table.

Data validation has applied when the user doesn’t enter any input and errors arising from the api call are also handled.

Graphical user interface, text, application, email

Description automatically generated

# 

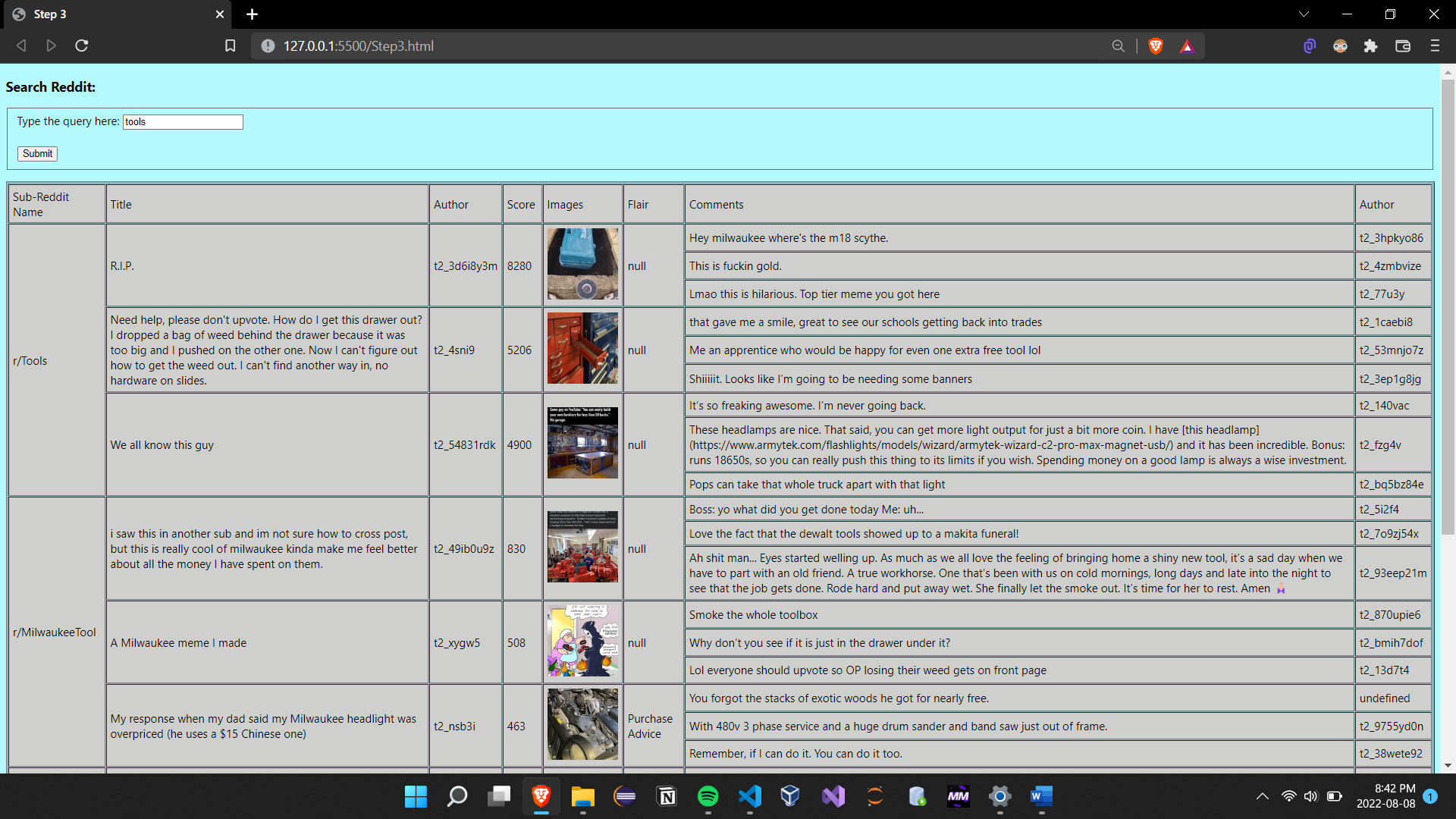
# Step 3: You are asked to develop a program which gets user input ‘subreddit’.

# And displays the top 3 matching subreddits. In each subreddits, displays three top score posts, their authors, score, any related images, and flair of each post

# • For each post, displays few top comments and its authors (i.e display 3 comments, ordered by their score, or date)

**Code Explanation:-**

First, input from the user is taken which is passed to the fetch API that searches for top 3 relevant subreddits. Top 3 are displayed by adding limit = 3 to the URL. Then, a for loop runs three times for these 3 subreddits. In each loop, another fetch API gets data for the top 3 posts of the particular subreddit. The name of the subreddit is passed to the API using a variable and “top” is added to the URL to display the top 3 posts of each subreddit. Then, another for loop runs 3 times and gets various forms of each post of each subreddit. Each time, the info is appended to one row of the table. When both the loop finishes, we have 9 rows containing information. These rows are not closed yet as we still must append comments with them. These 9 rows are stored in a variable “output” which is then stored in an array of size 9 using a counter “count for Array”. Then, another fetch API is used to get data of the top 3 comments of each post of each subreddit. Now the table and its headings are declared inside the then function of this third fetch API. The table is declared inside this function as we will close the table in this then function and append it to a variable “output2”. The declaration is done only one time at the start using a counter which is “countForTable”. Then, information from the array is retrieved here using the same counter and again a for loop runs for three times which gets information of top 3 comments of each post of each subreddit. When the loop runs for the third time, table is closed, then information from next index of array is retrieved and again the for loop runs and then closes the row after third time. So in total we have 27 rows of comments. When the last row is closed, then the table is also closed and is appened to “output2” which is the finally displayed.



# Step 4 : You are asked to design the app as follow:

**• User has the option to search on reddit, choose the number of records in result-set, and choose the reddit-sort attribute**

**• The search result should have a title or a summary of the reddit content, + a picture + score + subreddit + a link (read more) to actual article.**

**• In your application, for formulating the URL for API call (and other purposes), use “JavaScript Interpolation/Variable substitution (https://www.w3schools.com/js/js\_string\_templates.asp)**

**• Apply proper data validation on user input and error handling on API calls and search operation**

**Code Explanation:**

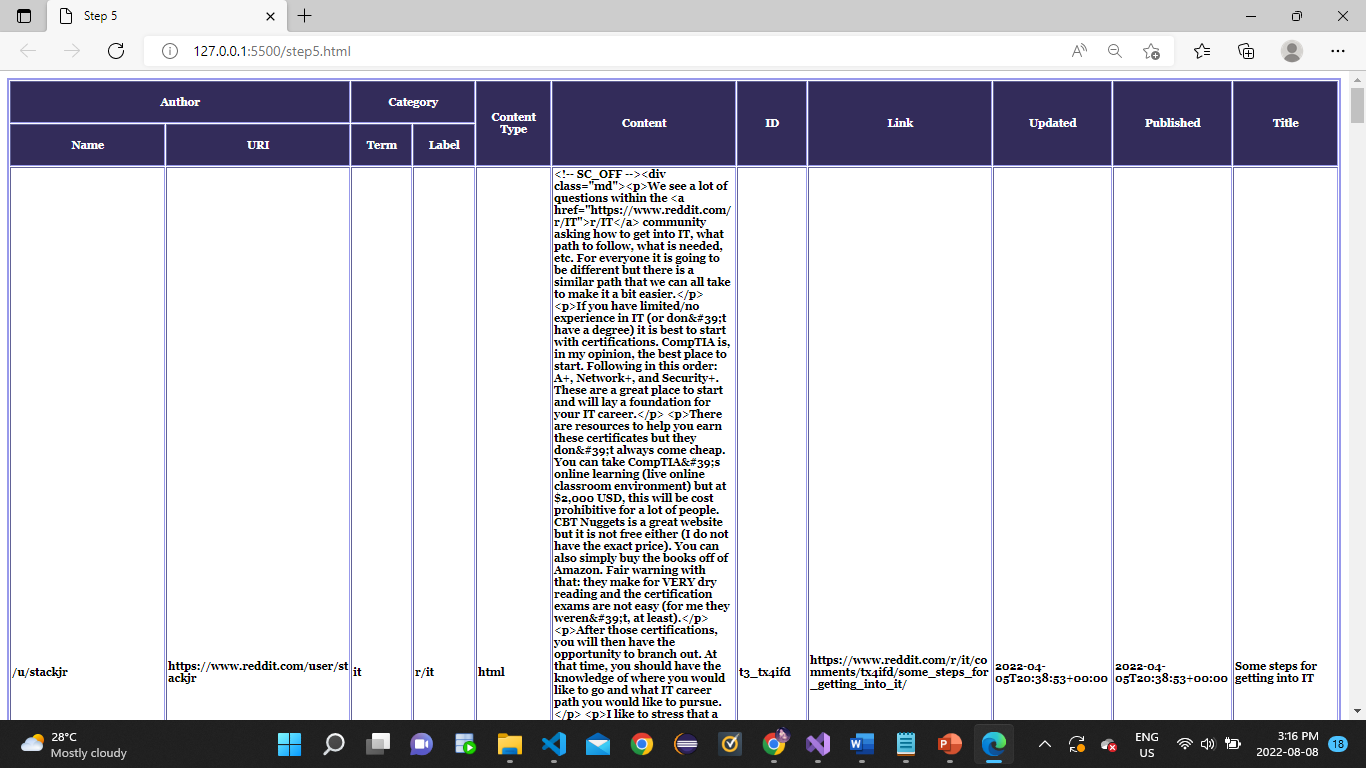
First, three inputs are asked from the user which are reddit search query, sort by and no. of results to display(limit). Then this information is passed to the fetch API which retrieves the data. Then, for each loop retrieves various information for each subreddit and displayed in a proper format using the Bootstrap library.

Graphical user interface, website

Description automatically generated

# Step 5 : You are asked use the attached XML file which is based on one of the subreddit data in RSS/XML format.

# You are asked to develop a JS program which displays these records in proper format in the output.



**Code Explanation:**

Step 1: created a html file and added some style sheet to make it look nice.

Step 2:   //Create the XHR object

    var xhttp = new XMLHttpRequest();

    //Open the file

    xhttp.open("GET", "dataset\_r.xml", false);

    //Send the request

    xhttp.send();

    // Read data

    xmlDoc = xhttp.responseXML;

    console.log(xmlDoc);

Added this code to first create XHR object and open file which data we would like to print using JS. Then send the request and read the data from the XML document.

Step 3:  let output = "<table border='2'>";

        output+="<tr>";

        output+="<th colspan='2'>"+ "Author" + "</th>";

        output+="<th>"+ "Category" + "</th>";

        output+="<th>"+ "Content" + "</th>";

        output+="<th>"+ "Id" + "</th>";

        output+="<th>"+ "Link" + "</th>";

        output+="<th>"+ "Updated" + "</th>";

        output+="<th>"+ "Published" + "</th>";

        output+="<th>"+ "Title" + "</th>";

        output+="</tr><tr>";

        output+="<th>"+ "Name" + "</th>"

        output+="<th>"+ "Link" + "</th>"

        output+="<th></th>"

        output+="<th></th><th></th><th></th><th></th><th></th><th></th>"

        output+="</tr>"

Took a variable as output to store data inside that variable. We have used some html tags to make a table and do some formatting on the table.

Step 4:

var lengthofTagEntery = xmlDoc.getElementsByTagName("entry").length

    for(let i = 0; i<lengthofTagEntery ;i++)

    {

        let entry = xmlDoc.getElementsByTagName("entry")[i];

        output+=`<tr>`

        output+=`<td>${entry.children[0].children[0].innerHTML}</td>`;

        output+=`<td><a href='${entry.children[0].children[1].innerHTML}'>${entry.children[0].children[1].innerHTML}</a></td>`;

        output+=`<td>${entry.children[1].getAttribute("term")}</td>`;

        output+=`<td><html>${entry.children[2].innerHTML}</html></td>`;

        output+=`<td>${entry.children[3].innerHTML}</td>`;

        output+=`<td><a href='${entry.children[4].getAttribute("href")}'>${entry.children[4].getAttribute("href")}</a></td>`;

        output+=`<td>${entry.children[5].innerHTML}</td>`;

        output+=`<td>${entry.children[6].innerHTML}</td>`;

        output+=`<td>${entry.children[7].innerHTML}</td>`;

        output+=`</tr>`

    }

    document.getElementById("display").innerHTML=output;

In the last step, took another variable to map the length of the entry element from the XML file and store it in that variable. Then loop that until we fetch all the data, we want from the XML file. In the end, we are printing all data into the display.