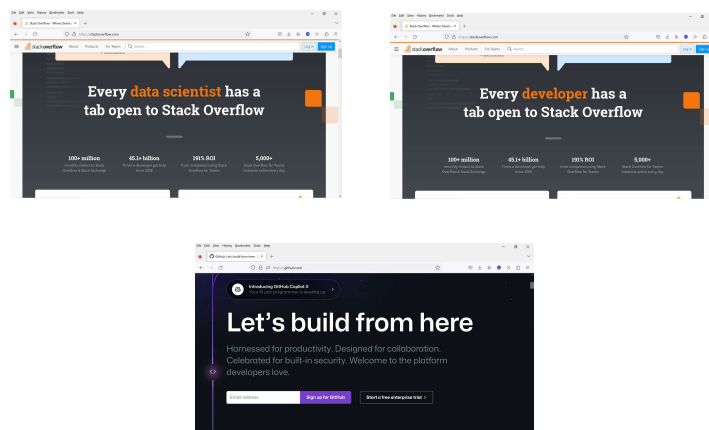
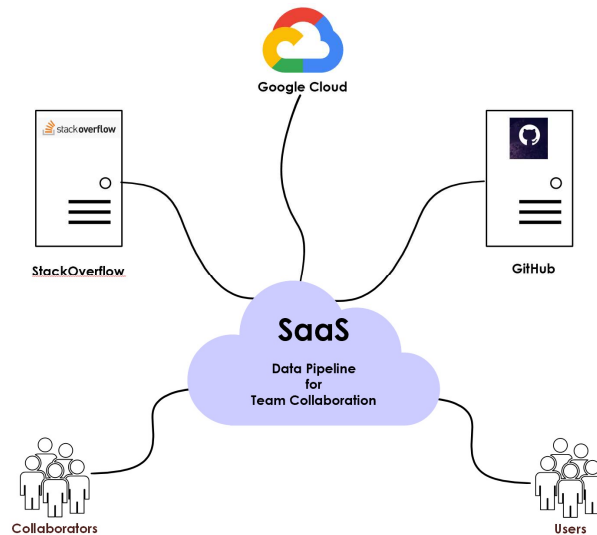


BugHub

Data Pipeline and Knowledge Graph For StackOverflow & GitHub Team Collaboration

Author: Atef Bader
Last Edit: 4/11/2023



Your Task is to use **GitHub** and **StackOverflow** data to create the data pipeline and the Knowledge Graph for **BugHub**



Project Overview Statement:

StackOverflow and **GitHub** decided to join forces to start a new SaaS (Software as a Service) data pipeline for the **BugHub** project to build the platform of the future for Team Collaboration and Knowledge Sharing.

The challenge for the **Repos' Contributors** is how to distinguish between a **general question** vs. a **bug**.

When you find a **bug** in a release, you open an **issue** on **GitHub**, and the **issue** will have a **number** and **labels** associated with it, see below an example:

The screenshot shows a GitHub repository page for `angular/angular-cli` with the 'Issues' tab selected. The page displays a list of issues. Two callout boxes provide information about the issue structure:

- Issue:**
 - Number
 - Description
- Issue:**
 - Labels

Arrows from the callouts point to specific elements on the issue list:

- The first callout points to the issue number `207` in the header of the first issue.
- The second callout points to the labels `area: devkit/build-angular` and `feature: under consideration` on the fifth issue.

The visible issues in the list are:

- `207 Open` `Compilation error after migrating from Angular 13 to 15` (opened 15 hours ago by `anandlingaraj`)
- `esbuild-builder: Sort output table by size` (opened yesterday by `enealiaho`)
- `esbuild serve not using configured outputPath, all assets missing` (opened yesterday by `jake58`)
- `support tailwind JIT from typescript tailwind.config.ts` (opened 4 days ago by `burner-account`)
- `esbuild watch mode keeps detecting changes and rebuilding` (needs more info)

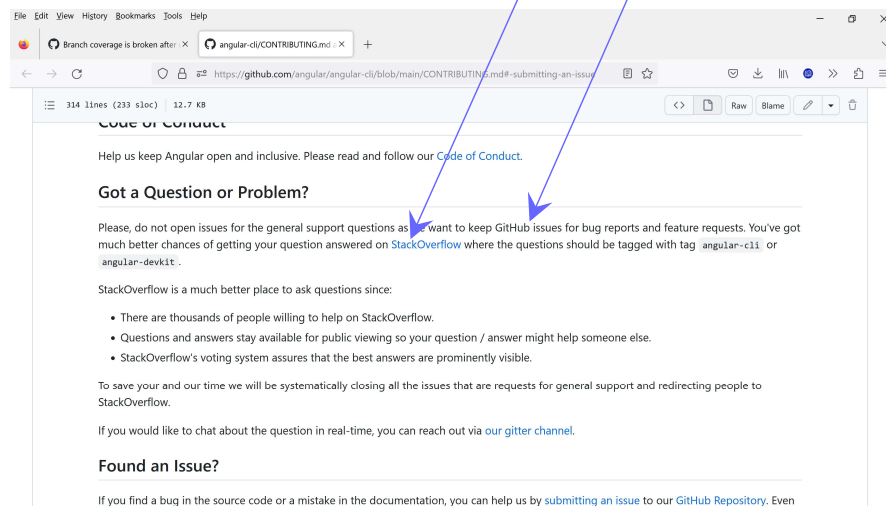
GitHub Repos' Contributors **urge** users to check **StackOverflow** for general questions before opening an **issue** on **GitHub**

StackOverflow for:

- General Questions
- Answers

GitHub Issues for:

- Bug Reports
- Feature Request



Use **Google** to search for the following error:

React Cannot read property 'displayName' of undefined

Google found it on:

- [GitHub](#)

Google found it on:

- [StackOverflow](#)

React Cannot read property 'disp X

React Cannot read property 'displayName' of undefined

About 38,200 results (0.29 seconds)

GitHub
https://github.com › etorybookje › etorybook › issues

Cannot read property 'displayName' of undefined #7390
Jul 11, 2019 — In my case it was because I imported the component as a named import instead of default import. So I fixed it from this: import { Button } from ...

Cannot read property 'displayName' of undefined · Issue #1168 Sep 25, 2018
Cannot read property 'displayName' of undefined · Issue #592 Jan 31, 2020
Uncaught TypeError: **Cannot read property 'displayName' of ...** May 16, 2018
Cannot read property 'displayName' of undefined #42 - GitHub Aug 6, 2017
[More results from github.com](#)

Stack Overflow
https://stackoverflow.com › questions › storybook-can...

Storybook cannot read property 'displayName' of undefined
Aug 4, 2021 — I got this error, I was working on an existing react library and making/editing things by hand. The solution was to find the `index.ts` in the ...

5 answers · Top answer: Faced the same issue. What is the mistake I have done is using the ...

reactjs - Cannot read property 'displayName' of undefined Aug 26, 2016
Cannot read properties of undefined (reading 'displayName' ... Aug 20, 2022
Uncaught TypeError: **Cannot read property 'displayName' of ...** Jun 30, 2021
reactjs - Unit testing : Error: TypeError: Cannot read property ... Dec 26, 2018
[More results from stackoverflow.com](#)

For this pilot project, imagine that **StackOverflow** and **GitHub** have contacted you to help them to develop a **data pipeline** to pull data from their servers concurrently and **cross-reference** the questions/answers/tags of StackOverflow to GitHub Repos/Branches/Issues/Labels/contributes/releases/etc.

The **data pipeline** will be used to collect the raw data from **StackOverflow** and **GitGub** and the data will be ingested from the two data sources into the data store. Before data flows into the data store, it will go through data preparation and preprocessing to filter and aggregate the data from the two data sources.

The data received from the two data sources will be processed and stored in EdgeDB as follows:

- **GitHub** data for the different **Repos** will be stored in **EdgeDB**
- **StackOverflow** data (Questions, Answers, Tags, etc.) will be stored in **EdgeDB**
- **BugHub** will have **GitHub Repo** data and attached to it **StackOverflow** data stored in **EdgeDB**

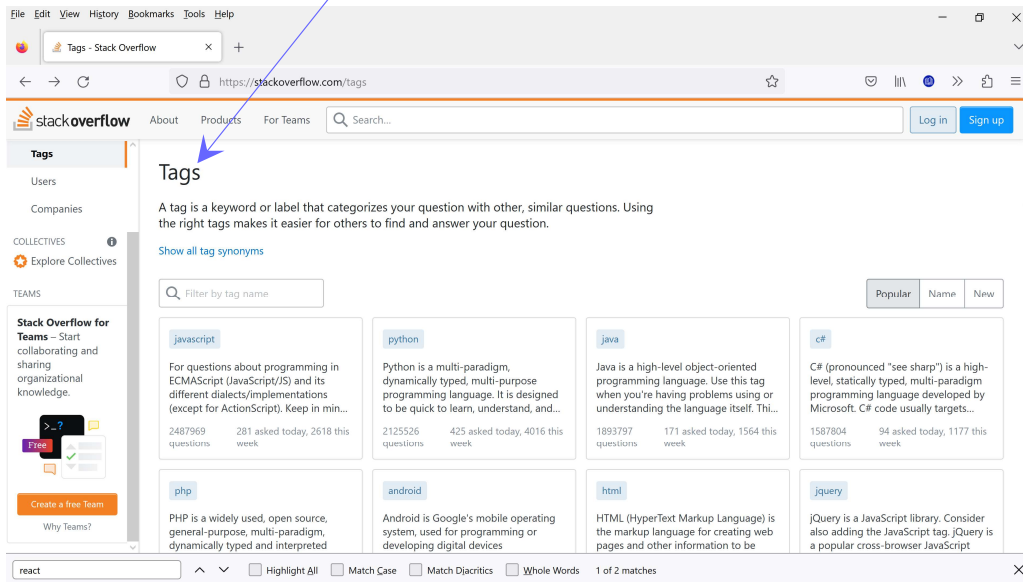
Online Documentations and Resources

After you review and analyze the documentation for **GitHub**, **Git**, and **StackOverflow** using the online educational resources listed below, provide your answers for the requirements listed in the **Requirements Specification** section below.

- <https://docs.github.com/en>
- <https://pygithub.readthedocs.io/en/latest/introduction.html>
- https://github.com/sks444/Pygithub-Examples/blob/master/get_repositories_details.py
- https://github.com/sks444/Pygithub-Examples/blob/master/get_issues_details.py
- <https://networkx.org/documentation/stable/tutorial.html#>
- <https://docs.github.com/en/get-started/quickstart/hello-world>
- <https://docs.github.com/en/get-started/quickstart/git-and-github-learning-resources>
- <https://docs.github.com/en/get-started/quickstart>
- <https://docs.github.com/en/get-started/quickstart/git-cheatsheet>
- <https://docs.github.com/en/get-started/quickstart/be-social>
- <https://docs.github.com/en/get-started/quickstart/github-flow>
- <https://docs.github.com/en/get-started/quickstart/communicating-on-github>
- <https://docs.github.com/en/account-and-profile/setting-up-and-managing-your-personal-account-on-github/managing-personal-account-settings/about-your-personal-dashboard>
- <https://docs.github.com/en/get-started/quickstart/be-social>
- <https://docs.github.com/en/get-started/quickstart/github-glossary>
- <https://stackoverflow.com/tags>

StackOverflow :

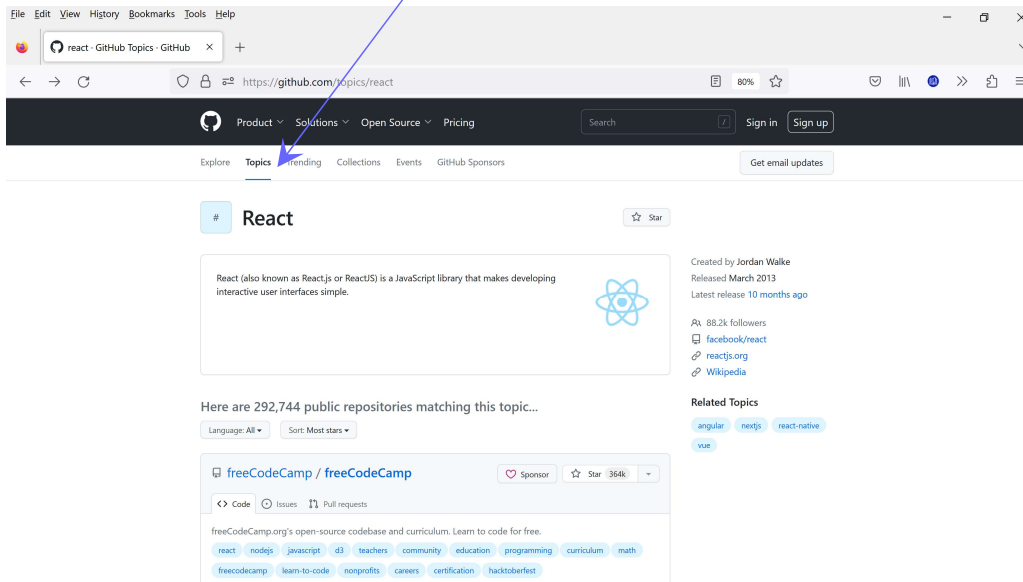
- Tags



The screenshot shows the Stack Overflow website's 'Tags' page. The browser address bar displays 'https://stackoverflow.com/tags'. The page features a sidebar on the left with navigation links like 'Tags', 'Users', 'Companies', 'Collectives', and 'Teams'. The main content area is titled 'Tags' and includes a definition: 'A tag is a keyword or label that categorizes your question with other, similar questions. Using the right tags makes it easier for others to find and answer your question.' Below this, there's a search bar 'Filter by tag name' and a grid of tag cards. Each card represents a programming language or technology, such as 'javascript', 'python', 'java', 'c#', 'php', 'android', 'html', and 'jquery'. Each card shows a brief description and statistics like 'questions' and 'asked today'. At the bottom, there's a search bar with the text 'react' and a '1 of 2 matches' indicator.

GitHub :

- Topics
- Repos



The screenshot shows the GitHub website's 'Topics' page for 'React'. The browser address bar displays 'https://github.com/topics/react'. The page features a dark header with navigation links like 'Product', 'Solutions', 'Open Source', and 'Pricing'. The main content area is titled '# React' and includes a description: 'React (also known as React.js or ReactJS) is a JavaScript library that makes developing interactive user interfaces simple.' Below this, there's a search bar 'Here are 292,744 public repositories matching this topic...' and a list of repositories. The first repository shown is 'freeCodeCamp / freeCodeCamp'. The page also includes a sidebar on the right with links to 'Related Topics' like 'angular', 'nextjs', 'react-native', and 'vue'.

Requirements:

Provide your answers for the following requirements:

1. List all objects, properties and links that are needed to represent the **graph data model** of the knowledge graph of **GitHub** topics, tags, repositories, and issues.
2. List all objects, properties and links that are needed to represent the **graph data model** of the knowledge graph of **StackOverflow** tags, questions, and answers.
3. Use **networkx** package to create the **graph data model** of the knowledge graph (Network) for **GitHub**
 - i. <https://networkx.org/documentation/stable/install.html>
4. Use **networkx** package to create the **graph data model** of the knowledge graph (Network) for **StackOverflow**
 - i. <https://networkx.org/documentation/stable/install.html>
5. Use **networkx** package to create the **graph data model** of the knowledge graph (Network) for the combined data, we will call it the **BugHub** knowledge graph, of **GitHub** and **StackOverflow**
 - i. <https://networkx.org/documentation/stable/install.html>
6. Write Python code (Use the **APIs** for **StackOverflow** and **GitHub**) to retrieve the data for **GitHub** and **StackOverflow** based on the knowledge graphs you created in Requirements 1, 2, 3, 4, and 5:
 - a. <https://github.com/topics/react>
 - i. <https://github.com/facebook/react>
 - ii. <https://stackoverflow.com/questions/tagged/reactjs>

- b. <https://github.com/topics/angular>
 - i. <https://github.com/angular/angular>
 - ii. <https://stackoverflow.com/questions/tagged/angular>
- c. <https://github.com/topics/visualization>
 - i. <https://github.com/d3/d3>
 - ii. <https://stackoverflow.com/questions/tagged/d3.js>

7. Use **networkx** package to create the knowledge graph (Network) for every repo listed in requirement 6.

8. Go to **GitHub** and **StackOverflow** and inspect the following **topics**, **repos**, and **tags** and document the most relevant **repos** and **tags** for every topic:

- a. <https://github.com/topics/react>
 - i. <https://github.com/facebook/react>
 - ii. <https://stackoverflow.com/questions/tagged/reactjs>
- b. <https://github.com/topics/angular>
 - i. <https://github.com/angular/angular>
 - ii. <https://stackoverflow.com/questions/tagged/angular>
- c. <https://github.com/topics/visualization>
 - i. <https://github.com/d3/d3>
 - ii. <https://stackoverflow.com/questions/tagged/d3.js>

9. Use **GitHub API** to retrieve information of the past 2 years for the following repositories:

1. <https://github.com/angular/angular>
2. <https://github.com/angular/material>
3. <https://github.com/angular/angular-cli>
4. <https://github.com/SebastianM/angular-google-maps>
5. <https://github.com/d3/d3>
6. <https://github.com/facebook/react>

7. <https://github.com/tensorflow/tensorflow>
8. <https://github.com/keras-team/keras>
9. <https://github.com/pallets/flask>

10. Use **StackOverflow API** to retrieve information of the past 2 years for **Tags** related to the repositories listed in requirement 9.
11. Store your **GitHub** data in a DataFrame and retrieve from the DataFrame the **repo** that has the highest number of **GitHub** issues created.
12. Store your **StackOverflow** data in a DataFrame and retrieve from the DataFrame the **repo/tag** that has the highest number of **StackOverflow** questions created.
13. Store your **StackOverflow** data in a DataFrame and retrieve from the DataFrame the top 5 questions for **React** on **StackOverflow** that have the highest number of answers, votes, and views.
14. Use **EdgeQL/GraphQL** to create the complete object-relational data model for **GitHub** in **EdgeDB** using the **GitHub** data
15. Use **EdgeQL/GraphQL** to create the complete object-relational data model for **StackOverflow** in **EdgeDB** using the **StackOverflow** data
16. Use **EdgeQL/GraphQL** to create the complete object-relational data model for **BugHub** in **EdgeDB** using the combined **GitHub** and **StackOverflow** data; consider the following partial list of objects/attributes, it is your responsibility to provide the complete list of objects/properties/links:
 - git
 - GitHub
 - Repository
 - Project
 - Wiki

- Pulse
- Graph
- Code
- Issue
- commit
- branch
- clone
- remote
- fork
- pull
- log
- merge
- push
- status
- Tag
- Release
- histroy
- File
- General Comment
- Line Comment
- Review
- User Account
- Organization Account
- README.md
- CONTRIBUTING.md
- ISSUE_TEMPLATE.md
- System Configuration
- Global Configuration
- Local Configuration
- Etc.

17. Use **EdgeDB UI** to demonstrate how **StackOverflow** data (questions/answers/tags/etc) for the following tags/topics/repos are attached to **GitHub** repos in your **BugHub** object-relational database application:
- d. <https://github.com/topics/react>
 - i. <https://github.com/facebook/react>
 - ii. <https://stackoverflow.com/questions/tagged/reactjs>
 - e. <https://github.com/topics/angular>
 - i. <https://github.com/angular/angular>
 - ii. <https://stackoverflow.com/questions/tagged/angular>
 - f. <https://github.com/topics/visualization>
 - i. <https://github.com/d3/d3>
 - ii. <https://stackoverflow.com/questions/tagged/d3.js>
18. Use **EdgeQL/GraphQL** and **BugHub EdgeDB** to retrieve the repo that has the highest number of **GitHub** issues created.

19. Use **EdgeQL/GraphQL** and **BugHub EdgeDB** to retrieve the repo that has the highest number of **StackOverflow** questions created.
20. Use **EdgeQL/GraphQL** and **BugHub EdgeDB** to retrieve the top 5 questions for **React** on **StackOverflow** that have the highest number of answers, votes, and views.