Wikipedia assistant

Assignment Objective

Create an API to serve content from the Simple English Wikipedia . The API should expose two endpoints:

- One endpoint for querying the contents using arbitrary contents
- One endpoint to find the most outdated page in a category.

Requirements

1. Create a database based on the Simple English Wikipedia content. There are periodic dumps with the data that you can find here. And here, you can find a list of the tables and their schema. The tables that are relevant for this assignment are: page, pagelinks and categorylinks. The database should contain the following:

    - Basic metadata for every wiki page:
        - Page title
        - Categories of the page
        - Date of last modification

    - The links between the wiki pages, defined by the following:
        - The page which refers to another page
        - The referred page

    You're allowed to make simplifications (in case something is not trivial to implement) but you have to document your decisions about these simplifications.
    You can download and preprocess the data dumps as you wish, but we recommend using the SQL data dumps instead of the XML ones, since they are easier to work with.

2. Create an API to query the contents of the database. The API has to expose two endpoints:

    A. An endpoint that receives an arbitrary SQL query and returns the result of executing the query on the database.

    B. An endpoint that receives a category and returns the *most outdated* page for that category. A page is called outdated if at least one of the pages it refers to was

modified later than the page itself. The measure of this outdatedness is the biggest difference between the last modification of a referred page and the last modification of the page. This query can be a bit slow, so you should precompute the results for the top 10 categories with more pages. You can assume that this endpoint will only be called with one of the top 10 categories.

3. Create an example architecture diagram that depicts how the application could be deployed in one of the following cloud providers: AWS, Google Cloud, Azure. You can choose the cloud provider you are more comfortable with. Provide a brief explanation of your architecture.

Submission

Please submit your assignment as a GitHub repository. Include a README file with the necessary setup and usage instructions on how to set up and run your project.