



CSCE 2501

Fundamentals of Database Systems

Project - World Geopedia

Prof. Hossam Sharara
Dept. of Computer Science, AUC

Introduction

In this project, you will be building a world geopedia, that allows users to get useful geographical, economical, and political information about all countries in the world in a structured form.

One of the sites that provide such information is Wikipedia, which has a wealth of information about all the countries in the world. However, the website allows for viewing the information on a single country or a single city of interest at a time, which make it hard to draw any analytics or gross reporting on a global scale.

Your task will to build a database system backend and an application that provide the regional and global analytics and features described next.

Bonus (2%):

In addition to the information requirement on the following page, connecting to google news to retrieve the total covid-19 cases and vaccinations administered for each country (<https://news.google.com/covid19/map>)

Milestone I: Data Requirements

The database system you design should be based on the information provided on each country and its corresponding capital city listed on Wikipedia (e.g. <https://en.wikipedia.org/wiki/Egypt>)

Your database design should store the following information about each country:

Basic: name, official languages, population, driving side, calling code, and timezone

Geographical: continent, capital, area (km²), and water percentage

Political: capital, legislature, president / monarch, and HDI (human development index)

Economical: currency, GDP (Purchase Power), GDP (nominal), and Gini Index

For each president / monarch, you need to store his/her name, birthdate, the date he/she assumed office, and his/her political party. You also need to store information about capital city for each country, including its name, population, area / metro area, governor, and coordinates.

The system should also allow users to store their travel history for different countries, through registering on the system using their email addresses, and pick a username, gender, age, and birthdate. After registering, the users can add their travels for any country, providing their travel dates, their rating of the visits (1-10) and a textual review.

Milestone I: Database Design & Implementation

Tools: MySQL

Deadline: 17th October, 2021

Description:

In this milestone, you are required to design the database ERD, create your database and schema in MySQL server for the data requirements presented on the previous slides.

Deliverables:

- Entity-Relationship Diagram of the World Geopedia System
- Relational Model for your system
- A transcript for the SQL statements used to create your database and schema

Milestone II: Database Population

In this milestone, you are required to write a web crawler to crawl the pages from the Wikipedia website for all the world countries, parse the HTML you crawl and extract the relevant fields for populating the non-user tables in your schema. The countries in each continent is available on wikipedia at [\(https://en.wikipedia.org/wiki/List_of_sovereign_states_and_dependent_territories_in_<continent name>\)](https://en.wikipedia.org/wiki/List_of_sovereign_states_and_dependent_territories_in_<continent_name>)
E.g. The list of countries in Africa will be available at [\(https://en.wikipedia.org/wiki/List_of_sovereign_states_and_dependent_territories_in_Africa\)](https://en.wikipedia.org/wiki/List_of_sovereign_states_and_dependent_territories_in_Africa). You should use these pages as the seed node for the countries in each continent.

At the end of this milestone, you should have all your non-user tables populated with the extracted data. For the user tables, use sample test information to populate the relevant tables.

Milestone II: Web Crawling and Data Population

Tools: Scrappy, pycharm, python (or any other crawler library / language)

Deadline: 7th November, 2021

Description:

In this milestone, you are required to implement your crawler, extract the data from elcinema website and populate your database

Deliverables:

- Crawling script
- Populated MySQL database dump
- CSV files for the values in each table

Milestone III: Application Layer

In the final milestone, you are required to design a client application that is capable of connecting to the database hosted on MySQL server. The application should have the following functionalities:

- Add a new user review on a country
- View existing reviews on a given country
- Register a user
- Show all the countries that have a specific legislature
- Show the top 10 countries by GDP, population, area, density, GDP per capita, both globally and within each continent
- Show all the countries who drive on the right vs. on the left
- Query and view a given country / capital city information
- Query and view president / monarch's information
- Identify the country for a given phone number
- Identify the country of a given city
- (Bonus): Identify the top and bottom 5 countries in each continent in terms of covid cases and vaccination rate

Milestone III: Application Layer

Tools: Any language / platform to implement your application (Web-based, GUI, command-line are all acceptable)

Deadline: 28th November, 2021

Description:

In this milestone, you are required to implement an application to perform various transactions / queries on your backend database system. The database has to be hosted on an actual server (you can use any of the free DB hosting services available (e.g. www.db4free.net)) and make sure your application is communicating with this service not the localhost

Deliverables:

- Application Implementation / Source Code + Executable
- Latest Dump of your database
- Demo