Populization - International Population Visualization System

Prathamesh Kulkarni, Nikhil Neroor, Harshini Bonam

***Abstract*— We propose Populization - a web application that visualizes international census information from 1950 to 2050 in an insightful manner. It would allow users to interact with the application through various interfaces and gain the desired in-depth knowledge of population. Users would apply filters to narrow down the scope of visual representations. Application would also offer the facility to download visual representations on client machine.**

I. Project Description

For many years, scientists have studied the effects population growth has had on the world. In recent times, the situation has started to go way out of our control. With the current growth in the number of people inhabiting our planet, and the fact that the availability of resources to each person is not increasingly correspondingly, the human race is in for an extremely tough time in the near future. If the population is not brought under control, then pretty soon our resources will run out and our livelihood will be ruined. Hence, in this project, we have created a user-friendly website which will help people visualize just how serious this situation is and will also give them an insight as to how we can remedy the current state of affairs.

The project has four stages: Gathering, Design, Infrastructure Implementation, and User Interface.

A. Stage1 - The Requirement Gathering Stage.

We have the international census data from 1950 to 2050 with recorded as well as projected values. We are planning to develop a web application that would visualize this international census in various, insightful formats. With this web application, users would interact, analyze and extract domain knowledge about the international population from raw data. Users will not have write access to the data which is reserved for developers of the website. This application aims to provide users only read access to the raw data in visual form.

- The dataset(CSV file) has three million records containing census information about every country from 1950 to 2050. The attribute year allows the data to be interpreted temporally. Since the dataset provided by United States Census Bureau has extrapolated values till year 2050, data streaming is no longer required for this data at rest.
- Entities in our database are as follows:
  - AgeSpecificFertilityRates
  - BirthDeathGrowthRates
  - CountryNamesArea
  - MidyearPopulation
  - MidyearPopulation5yrAgeSex
  - MidyearPopulationAgeSex
  - MidyearPopulationAgeSpecific
  - MortalityLifeExpectancy
- One to Many relationships are as follows:
  - From CountryNamesArea to AgeSpecificFertilityRates, MidyearPopulation, MortalityLifeExpectancy, BirthDeathGrowthRates
  - From MidyearPopulation to MidyearPopulation5yrAgeSex, MidyearPopulationAgeSpecific
- One to One relationship is as follows:
  - From MidyearPopulation to MidyearPopulationAgeSex
- In this project, based on the demographics of the scenario being visualized, visual representations range from basic scatter plots to heat maps and choropleth maps. rates at which the attributes such as birth rates, fertility rates, mortality rates, migration rates change and total counts of population would assist in presenting interesting scenarios that will be discussed in further sections.
- The fundamental scenarios that would be presented are the trends in population growth over years 1950-2050 across Trends of population growth as well as comparisons between trends of multiple countries will give the user an overview of the census scenario across the 228 countries and territories. These include a) The relationship between mother's age and child survival b) Distinction between overpopulated and underpopulated countries based on population density c) distribution of the ages of migrating people and the average age of a country
- The backend data in this project would be stored in MySQL database for the following reasons: a) Its a relational DBMS which fits our structured data storage needs. b) It's default size limit for a table is 4 GB which is more than sufficient for our 2GB data. c) It's an open-source DBMS
- Visual representations that would be displayed on a screen are scatter plots, bar charts, pie charts, line charts, histograms, heat maps, choropleth maps. User would be able to choose amongst alternate visual representations of the same data. User would also be able to apply filters such as countries, years, genders, ages, etc to the visual representations
- Interactivity

B. Stage2 - The Design Stage.

Transform the project requirements into a system flow diagram, specifyng the different algorithms, data types and

Fig. 1. Sample Data

Fig. 2. Sample Data

structures required for processing and their associated operations. The deliverables for this stage include the system flow diagram containing a graphical representation and textual descriptions of the corresponding data trasnformations, high level pseudo code of the overall system operation, and overall system time and space complexity.

Please insert your deliverables for Stage2 as follows:

- Short Textual Project Description. Please insert here the flow diagram textual description here together with its overall time and space complexity.
- Flow Diagram. Please insert your system Flow Diagram here.
- High Level Pseudo Code System Description. Please insert high level pseudo-code describing the major system modules as per your flow diagram.
- Algorithms and Data Structures. Please insert a brief description of each major Algorithm and its associated data structures here.
- Flow Diagram Major Constraints. Please insert here the integrity constraints:
  - Integrity Constraint. Please insert the first integrity constraint in here together with its description and justification.

  Please repeat the pattern for each integrity constraint.

C. Stage3 - The Implementation Stage.

-
- We are going to use Django web framework with programming languages Python 3.7 at the backend and HTML + CSS + Javascript at the frontend.
-
  - The backend data in this project would be stored in MySQL database for the following reasons: a) Its a relational DBMS which fits our structured data storage needs. b) It's default size limit for a table is 4 GB which is more than sufficient for our 2GB data. c) It's an open-source DBMS Data size: 1.7 GB; Disk Resident?; Streaming ?;
  - List the most interestng findings in the data if it is a Data Exploration Project. For other project types consult with your project supervisor what the corresponding outcomes shall be. Concentrate on demonstrating the Usefuness and Novelty of your application.
-
- We are going to use Django web framework with programming languages Python 3.7 at the backend and HTML + CSS + Javascript at the frontend.
-
  - The backend data in this project would be stored in MySQL database for the following reasons: a) Its a relational DBMS which fits our structured data storage needs. b) It's default size limit for a table is 4 GB which is more than sufficient for our 2GB data. c) It's an open-source DBMS Data size: 1.7 GB; Disk Resident?; Streaming ?;
  - List the most interestng findings in the data if it is a Data Exploration Project. For other project types consult with your project supervisor what the corresponding outcomes shall be. Concentrate on demonstrating the Usefuness and Novelty of your application.

D. Stage4 - User Interface.

Describe a User Interface (UI) to your application along with the related information that will be shown on each interface view (How users will query or navigate the data and view the query or navigation results). The emphasis should be placed on the process a user needs to follow in order to meet a particular information need in a user-friendly manner. The deliverables for this stage include the following items :

- The modes of user interaction with the data (text queries, mouse hovering, and/or mouse clicks ?).
- The error messages that will pop-up when users access and/or updates are denied
- The information messages or results that wil pop-up in response to user interface events.
- The error messages in response to data range constraints violations.
- The interface mechanisms that activate different views in order to facilitate data accesses, according to users' needs.
- Each view created must be justified. Any triggers built upon those views should be explained and justified as well. At least one project view should be created with a justification for its use.

Please insert your deliverables for Stage4 as follows:

- The initial statement to activate your application with the corresponding initial UI screenshot
- Two different sample navigation user paths through the data exemplifying the different modes of interaction and the corresponding screenshots.
- The error messages popping-up when users access and/or updates are denied (along with explanations and examples):
  - The error message:
  - The error message explanation (upon which violation it takes place): Please insert the error message explanation in here.
  - The error message example according to user(s) scenario(s): Please insert the error message example in here.
- The information messages or results that pop-up in response to user interface events.

  - The information message: Please insert the error message in here.
  - The information message explanation and the corresponding event trigger
  - The error message example in response to data range constraints and the coresponding user's scenario Please insert the error message example in here.
- The interface mechanisms that activate different views.
  - The interface mechanism: Please insert the interface mechanism here.

II. Project Highlights.