

COP 5725
Database Management Systems

Project Deliverable 1

Group 21

Chen, Zhaoyang <zhaoyang.chen@ufl.edu>

Feng, Zhongxi <zhongxi.feng@ufl.edu>

Niu, Yicong <yicong.niu@ufl.edu>

Qi, Lin <linqi@ufl.edu>

Interactive city weather report

Contens

Overview and description of the application.....	3
Motivation of the database needs of the application and the potential user interest	3
The needed web-based user interface functionality	4
Application goals regarding trend analysis	5
Real-world data forming the basis of the application and the complex trend queries	6
Colloquial complex trend queries and their explanation	6
The intended use of the public domain and proprietary software	7

Overview and description of the application

We plan to do a visual analysis project of weather in multiple cities. The data stored in the database is selected from the hourly weather change records of 30 cities in the U.S. and Canada from 2012 to 2017, containing five attributes: humidity, pressure, temperature, wind direction, and wind speed. (The data is from <https://www.kaggle.com/selfishgene/historical-hourly-weather-data>)

Combining the knowledge of database and web design, we present the weather-related charts of different cities in the form of interactive web pages. The charts will describe various trends in weather over time, such as average the monthly temperature, number of occurrences of different weather throughout the year, and so on. The user will clearly see the weather conditions in each city, and can also filter the weather information according to their own wishes.

Motivation of the database needs of the application and the potential user interest

Every place, every day, there will be new weather information generated, which will be a huge amount of data. From the perspective of massive data storage, real-time efficiency, concurrency, redundancy, data security, and multi-user access are all issues that need to be considered in practice.

In our project, we need to store over one million records and realize the web-based application with a powerful backend which can respond to the visits of multiple users with low latency. Compared with traditional file-based storage, advanced database and database management systems are more likely to do this job.

Weather detection and recording are very important for people's travel. Moreover, people need to know whether a city is livable through weather reports. In our application, we will use interactive charts to present the weather conditions of a city from 2012 to 2017 in different aspects. We will also explore the various trends of the data over time so that users can have a very comprehensive understanding of the local weather conditions.

The needed web-based user interface functionality

In our case, the implementation of the web-based application should focus on usability and effectiveness. Also, for the purpose of data security, if put in a real-world context, authentication should be included too. The design of our application should include the above aspects.

The first feature is user login functionality. To access the data and our service, a user must log in with her member ID and password. This UI interface is designed as a box saying “User log in”, only by passing the correct combination of member ID and password stored in UserInfo could one log in successfully.

Moving forward, after successful authentication, the next page should be an introduction page/background showing what we do. A concise and beautiful background always captures people’s eyes. Considering that our dataset is a worldwide weather report, we’d put a world map on the entry page. The user could move a mouse over a specific area and thus trigger a detailed display of the weather report of that area.

In our dataset, the weather is described from four major features, namely humidity, temperature, pressure, and wind. In the UI design, we plan to include the four features in a tab on the top of the entry page. Due to a large number of possible queries that could be triggered by one general feature, it is necessary to specify the specific queries under each tab. For example, we need to design 5 complicated queries, so we could display the result here.

For each query, we plan to show the result by clicking a button or by submitting a web form. In the first case, if we want to show a pre-designed query that displays the weather trend, we will design a button. By clicking it, our service will return the result from the backend and display it on the web page.

In the latter case, if one wants to interact with the data, she could play with the database by inputting a customized query. We will provide the users with a dropdown list regarding each field. Say, listing city names, listing seasons, listing years. After selecting each value, one could hit the “Submit” button and the backend will handle the work. So, the user could find the humidity value in Vancouver in Spring in 2016 under the “Humidity” feature tab.

For trend display, we will provide a button that returns the results in a chart so one can tell the difference instantly, while for interactive data analysis, we will provide a user interface to satisfy various needs and show the result in a table since it is merely static data.

Application goals regarding trend analysis

In trend analysis, the hidden patterns need to be discovered from the past data. This means that we need to make a detailed and thorough research on our dataset and see what it could possibly contain.

Our dataset is a worldwide weather report on an hourly basis. So first, we could explore timely weather patterns. On an hourly basis, we could investigate how the weather changes throughout the day. Also, we could expand it to season level, say how the average temperature changes from the first Spring to the last Spring. Or even in the year level, finding how average temperature variations in different years.

From a feature level, we could focus on the summits and valleys of the weather data, which is the most robust signal showing any extreme weather. Calculating the range of different attributes through a timeline, such as a day, a month or a year, gives us a better idea of city habitability. Besides, the average value and mode value tell the most authentic weather conditions in this area. Combining those features with the timely patterns, we aim to find a comprehensive weather analysis spanning 5 years in 30 cities with respect to the four features in our dataset.

Second, we could integrate geographic patterns in our trend analysis. Our dataset spans from Canada to the US. It's possible to find how it changes along the way from the north to the south in North America. Finding how the weather of cities of different altitudes changes through time could be another goal. At the same time, such investigation could benefit industry developments, such as agriculture, or studies concerning animal migration.

Third, our dataset includes four features, namely, humidity, temperature, pressure, and wind. How do these features relate to each other? In one city, could there be a relation between any features, and what is such a relationship? Positive correlation, negative correlation, or even be summarized in formula expressions? Though peripheral, such a trend would be meaningful enough to be analyzed.

We aim to discover these trends. After locating all trends mentioned before, it would be a good idea to glue everything together and tries to find a global connection between different fields.

Real-world data forming the basis of the application and the complex trend queries

The dataset was acquired using real-world data from Weather API on the OpenWeatherMap website. Weather API provides historical weather data for any location on the globe. Historical data can be requested by geographic coordinates, city name, and city ID. Besides, the dataset is available under the ODbL license(aka Open Database License)which is a license agreement intended to allow users to share, modify, and process a database.

When some data is missing or abnormal, it will be modified as the average of the context data. All of the data will be stored as per the regulation.

Since the dataset consists of the property of time, which records the weather variation by hours, with a wide range of 30 cities from 2012 to 2017 sorts of trends can be inferred.

The main properties of the complex trend queries will include the average data, maximum data, frequency of sorts of weather, the range of temperature crossing other attributes like geography (latitude and longitude). This can also be applied to the trend of maximum data.

In this project, the average data like temperature and humidity will be managed and related to the city by month. So the trend and curve will demonstrate monthly temperature changes over time in each city.

Another complex trend about the frequency of different weather will be formed this way: The number of occurrences of weather (like sunny, cloudy, rainy) will be recorded in each city over time so that it's clear to see what kind of weather occurs the most in a certain city in different months.

The weather in a different region can be extreme occasionally, we would like to record the specific attributes of extreme weather like duration. The comparison can form another complex trend to illustrate which month has the longest extreme weather.

Colloquial complex trend queries and their explanation

- Query: Monthly (average/highest/lowest) temperature/ humidity/pressure/wind speed changes over time in each city.

Example: For city Chicago, the average temperature in January is 10 Celsius, February is 13 Celsius, etc.

- Query: Monthly changes in the number of occurrences of certain weather in each city over time
Example: For city Dallas, clear weather occurs 10 times in March, 6 times in April, etc.
- Query: The monthly (average/highest/lowest) temperature changes with time when a certain weather occurs in each city
Example: For city Denver, the average temperature of all ‘few cloud’ days in October is 26 Celsius, in November is Celsius etc.
- Query: Monthly changes in the duration of certain wind speed in each city over time
Example: For city San Diego, the duration of wind whose speed is larger than 5 m/s in June is 100 hours, in July is 136 hours, etc.
- Query: Monthly changes in the duration of a certain wind direction in each city over time
Example: For the city San Diego, the duration of wind which moves towards north in June is 60 hours, in July is 186 hours, etc.
- Query: The monthly (average/highest/lowest) temperature of cities in different latitude scopes varies with time.
Example: In March, the highest temperature of cities which are between latitude 35N to 40N is 25 Celsius, in April is 27 Celsius, etc.

The intended use of the public domain and proprietary software

- For the user interface, we plan to use the R shiny to build an interactive web app. And we will extend the web app with CSS themes, HTML widgets and JavaScript actions.
- For the database interface, we plan to use the CISE Oracle database server.
- Other tools we will use: RStudio, ERDplus, Microsoft Office