Milestone 2- Project Outline

Github Link: https://github.com/estherxpr/CIS550_final_project

1. Motivation

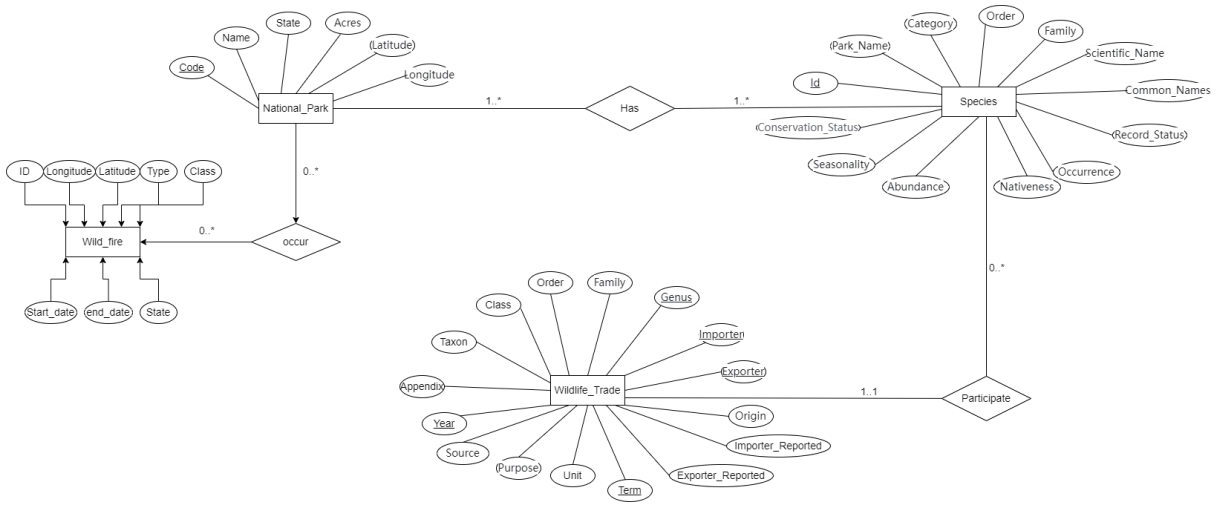
   Some of our members have backgrounds in biology. We all believe biodiversity and animals' survival circumstance in national parks are interesting topics. We are also curious about how wildlife trade affects the United states' biodiversity, and how emerging wildfires challenge the protection of endangered species in national parks. We want to investigate these questions and render a web-app to present the analysis result.
   We want to build a web app so that people can search species in which they are interested and get their habitat/distribution in the national parks. People can also search national parks and find what they can meet in the park.

2. List of Features / List of Pages(No.4 in outline rubric):
   a. Homepage
      i. Display each National park's basic info in a map. By hanging over each national park's icon in the map, users are able to view the park's abstract.
      ii. In the same map, if clicking on each state in the map, it would render a list of national parks in the state.
      iii. Two search bar: one for national park searching, and one for species searching
      iv. A second image layer/another page on visualizing the trending of wildfire on the map and show the overlap on national parks.
   b. National Park's page
      i. After either searching the name of a national park, or clicking on the national park in the homepage map, the corresponding national park's detail info will be rendered, including:
         1. species in the park
         2. Wildlife trading statistics in the park
         3. Wildfire threat level
   c. Species' page
      i. After searching the name of a species, or click a species' name in a national park's page, the information of the species will be rendered, including:

1. Common features of the species like origin, conservation status, etc.
2. The species's trading information.
3. Appearance in the national parks, allowing redirecting to the park page.

3. List of features that we might implement:
    a. We would like to write a web crawler to get all species pictures and present some pictures of each species, but we have to figure out how to store such a large database of pictures or just store pictures' cdn link.
    b. We would link each species to their related wiki pages so that people can get more information about it.
    c. We would like to make a user login using firebase, which supports google and facebook login and standard login.This might allow the public to contribute and edit our page.
    d. We will deploy the final app in AWS or Heroku and apply for a domain name for this application.
    e. We might integrate mapbox api so that in the homepage, we can display every national park in the map, and users can click the spot to navigate the park.
    f. We would like to integrate nps api in our app.
    g. We would like to build a recommendation system to allow continuous exploring across different fascinating species.

5. ER Diagram:



6. DDL:

```
CREATE TABLE National_Park (
        Code VARCHAR (4),
        Name VARCHAR (64),
        State VARCHAR (16),
        Acres INT,
        Latitude DECIMAL (5, 2),
        Longitude DECIMAL (5, 2),
        PRIMARY KEY (Code)
)

CREATE TABLE Species (
        Id VARCHAR (16),
        Park_Name VARCHAR (64) NOT NULL,
        Category VARCHAR (32),
        Order VARCHAR (32),
        Family VARCHAR (32),
        Scientific_Name VARCHAR (64),
        Common_Names VARCHAR (128),
        Record_Status VARCHAR (16),
        Occurrence VARCHAR (64),
        Nativeness VARCHAR (16),
        Abundance VARCHAR (8),
        Seasonality VARCHAR (16),
        Conservation_Status (32),
        PRIMARY KEY (Id),
        FOREIGN KEY (Park_Name) REFERENCES National_Park (Name)
)

CREATE TABLE Wildlife_Trade (
        Year VARCHAR (4),
        Appendix VARCHAR (8),
        Taxon VARCHAR (32) NOT NULL ,
        Class VARCHAR (32),
        Order VARCHAR (32) NOT NULL,
        Family VARCHAR (32) NOT NULL,
```

```
        Genus VARCHAR (64) NOT NULL,
        Importer VARCHAR (2),
        Exporter VARCHAR (2),
        Origin VARCHAR (2),
        Importer_Reported INT,
        Exporter_Reported INT,
        Term VARCHAR (32),
        Unit VARCHAR (2),
        Purpose VARCHAR (1),
        Source VARCHAR (1),
        PRIMARY KEY (Year, Genus, Importer, Exporter, Term),
        FOREIGN KEY (Taxon) REFERENCES Species (Category),
        FOREIGN KEY (Order) REFERENCES Species (Order),
        FOREIGN KEY (Family) REFERENCES Species (Family),
        FOREIGN KEY (Genus) REFERENCES Species (Scientific_Name)
)
CREATE TABLE Wildfire (
        short_ID INT NOT NULL,
        Fire_ID INT PRIMARY KEY,
        Source_system_type,
        Report_place VAR (64) NOT NULL,
        Report_place_ID INT NOT NULL,
        Fire_type VAR(32),
        Year INT NOT NULL,
        Discovery_Date float NOT NULL,
        Continuation Date float NOT NULL,
        Fire_cause VAR(32),
        Fire_cause_CODE INT,
        Fire_size DECIMAL (10, 2),
        Fire_size_class VAR(1),
        State VAR(2),
        County INT,
        FIPS_CODE INT,
        FIPS_name VAR(32),
        Latitude  DECIMAL (10, 2),
        Longitude  DECIMAL (10, 2),
        FOREIGN KEY (Latitude) REFERENCES National_park(Latitude)
        FOREIGN KEY (Longitude) REFERENCES National_park(Longitude)
```

7. Clean and pre-process the data.

For the species dataset, we find the Conservation Status column contains many empty grids. This is because the data only show endangered degree, for those rows with null values, these mean they are not in conservation status. For the Seasonality column, we find the reason that some rows are null is that these species are native, therefore they don't have seasonality. We will replace these Null grids with num 0. For the Abundance column, it is correlated with the occurrence column, we will fill the Null grid according to the occurrence column.

For the wildlife trade dataset, we will remove some redundant rows for rows that are in the same Taxon.

For the wildfire dataset, we will use the longitude and latitude of fire location together with the National park polygon file to build the 'occur' relationship with arcgis. We would also split the duration of fires to the months that it happened.

8. Technologies project will use:

Pandas as data clean and data analysis tool. MySQL as database, Express.js as backend controller, React as front-end view library

Firebase as NoSQL database to store user login and registration information.

We might use some crawler library such as BeautifulSoup to extract data from HTML.

We will use ArcPy to deal with geographic information and visualization.

9. Description of what each group member will be responsible for:

Peiran Xu: optimization for SQL query & data cleaning

Xiangchen Guan: query writing & database population

Zijian Xiao: front end component & write Crawler

Minzheng Zhang: front-end component & Express & rest api design