**Geospatial Analysis with Social & Web data**

# Basic Idea

The aim of this project will be to answer the following questions which are separated in 2 main categories:

**Land**

* **What distinguishes a land?**
  + The most prominent aspect of the land.
  + What category does the land belong to (entertainment, rest, educational, business, …).
* **Who visited/mentioned a land?**
  + Locations of the people who mentioned a land (geo-tagged social data).
  + Ranking of the areas where people mentioned a land (by number of mentions).

**People**

* **Where do people talk about the most?**
  + Ranking of places people in an area talks about (by number of mentions).
* **To what area these people belong to?**
  + Using geo-tagged social data to group people by their location.



Components

# Basic Requirements

* **App**
  + Easy to use user interface
  + Maps
  + Location Markers
  + Display Maps/Locations Information
  + Dashboard with different statistical data and information
  + Web Application
* **Data**
  + Social media data
  + Web data
  + Analysis of data
  + Categorization of data
  + Data Extraction scripts
  + Web Application

# Possible Tools

* **Normal Web App**
  + Angular web framework for user interface
  + Some map libraries such as leaflit.js
  + Python Django for server backend
  + NoSQL database such as MongoDB
* **ArcGIS & ArcGIS Online Web App**
  + ArcGIS Pro
  + Web App Builder
  + Geodatabase
  + Python Scripts
* **Data Extraction App**
  + Scrapping libraries for data extraction (selenium, …)

# Similar Aspect Applications

* **Google Maps**
  + **Provides**
    - Maps
    - Location Data Visualization
    - Location Information
    - Location Search
  + **Lacks**
    - Geospatial Analysis
* **Google Places**
  + **Provides**
    - Detailed Location Information
    - Location Attributes (photos, reviews, ..)
    - Location Search
  + **Lacks**
    - Geospatial Analysis
* **Human Behavior On Social Media (By Human Dynamics For Mobile Age - HDMA)**
  + **Project**
    - Determine urban land-use patterns in **Beijing**, **China** using 9.5 million geotagged social media messages from social media platform **Sina-Weibo** for six months in the urban core areas of **Beijing** and compared them with 385,792 commercial points of interests (POI) from **Datatang**, a Chinese digital data content provider. To estimate urban land-use types and patterns, the team created a grid measuring 400 x 400 meters to divide the urban core areas into 18,492 cells.
    - The team used Clustering techniques, Text Mining, Word Clouds, and the Distribution analysis of POI to identify seven types of land-use clusters in Beijing: residential areas, university dormitories, commercial areas, work areas, transportation hubs, and two types of mixed land-use areas.
    - Several computer programs to automatically or semi-automatically collect social media data from Twitter, Sina-Weibo, Google Places, and Reddit. They then saved it in MongoDB, a NoSQL Database
    - The center has also created two software toolsets to analyze and display the data it collects from social media platforms. A SMART Dashboard and a GeoViewer to manage and display their results.
  + **Reference**
    - <https://www.esri.com/about/newsroom/arcnews/human-behavior-on-social-media-is-big-data-and-gis-makes-it-actionable/>

# Basic Plan

The plan is divided into many phases:

**The 1st phase - Data Extraction**

This is the phase of data gathering where data will be gathered from different sources such as (Twitter, Reddit, Google Maps, Google Places). These data will then be reorganized in a database and used in the next phases.

**The 2nd phase – Geospatial Analysis on Data**

In this phase, using the data provided in the database, we start the main object of this project which is to analyze the data and deduce some conclusions from it. In order to do that we will first need to group some locations of buildings/sites and people to have a better overview on what distinguishes an area and how are people gathered in an area. This will result in population areas from people locations and zones from the groups of buildings/sites. This will be done using machine learning algorithms like clustering and some other grouping algorithms. After that, these results will be used to get an overview on which population area mentions/visits a site location the most (Top 10 sites mentioned in a population area) and which sites does a population area like and mentions a lot (Top 10 population areas who mentions a site).

**The 3rd phase - Data Visualization**

The last phase will be the visualization phase of all the data and results got from the previous phases. Here, a map will be used to show all the analysis results, sites information from social and web data, and any usable information that we might need to see. There will also be a search function to search for individual sites and show information about them. Finally, a dashboard will be shown with different information about the data with some charts and statistics.