**Geospatial Intelligence 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:

**Districts & Spots**

* **What characterizes a District?**
  + The most prominent aspect of a District.
  + What category does this District belong to (ex: Entertainment, Rest, Educational, Business, …).
* **Who mentioned a Site/Location/Spot?**
  + Locations of the people who mentioned a Site (geo-tagged social data).
  + Ranking of the areas where people mentioned a Site (by number of mentions).
  + Top 10 Population Areas for each Site.

**People & Population Areas**

* **To what area people belong to?**
  + Using geo-tagged social data to group people by their location.
  + Population Areas.
* **What do people talk about the most?**
  + Ranking of the Sites/Locations people in an area talks about (by number of mentions).
  + Top 10 Sites/Locations for each Population Area.

Components

# Requirements

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

# Possible Tools

* **Web App**
  + Angular web framework for user interface
  + Some map libraries such as leaflet.js
  + Python Django for server backend
  + Python Data Extraction scripts
  + NoSQL database such as MongoDB
* **ArcGIS & ArcGIS Online Web App**
  + ArcGIS Pro
  + Web App Builder
  + Geodatabase

# Literature Review

The aim of this project is to eventually get a conclusion from all the results obtained from the different geospatial analysis techniques done on the social and web data gathered during the time of this project. Some of these techniques are based on what people have used and tested before, whereas the other techniques will be custom to our needs in this project. A similar study case can be used to prepare the base of this project and then expand on it as necessary to what the project needs. A helpful one like the HDMA (Human Dynamics For Mobile Age) study case where they determined the urban land-use patterns in Beijing, China using 9.5 million geotagged social media messages from different sources 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. In this study case, 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. This can help as a base to start from in this project in order to reach our end goal of geospatial intelligence based on social media and web data at the end.

**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 3 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, and Google Places. These data will then be organized 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 spots/sites and people to have a better overview on what characretirze 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 spots/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 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 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 display. 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.