GIS (Geographic Information System): Mapping Locations

GIS: What is it?

A geographic information system (GIS) is a system that puts data on a geographic map. GIS connects data to a map, integrating location data (where things are, e.g., "Chicago, Illinois, USA") with all types of descriptive information (what things are like there, "I walked along lake Michigan in Chicago on a beautiful spring day.").

GIS in the NLP Suite: Three Steps

Going from text to maps *automatically* involves three steps.

1. First, you need to **extract the locations** mentioned in the text (via Stanford CoreNLP NER).
2. Second, you need to **geocode the locations**, i.e., find the latitude and longitude so that these locations can be mapped (via Nominatim or Google).
3. **Mapping** is indeed the last step in the process, a tool that, given, latitude and longitude, will place a location on a map, along with any desired information (e.g., the sentence where the specific location is mentioned, a photograph of a building). Mapping in the NLP Suite is done in Google Earth Pro (for pin maps) and Google Maps for heat maps.

The script GIS_main provides a complete pipeline to go from text to maps.
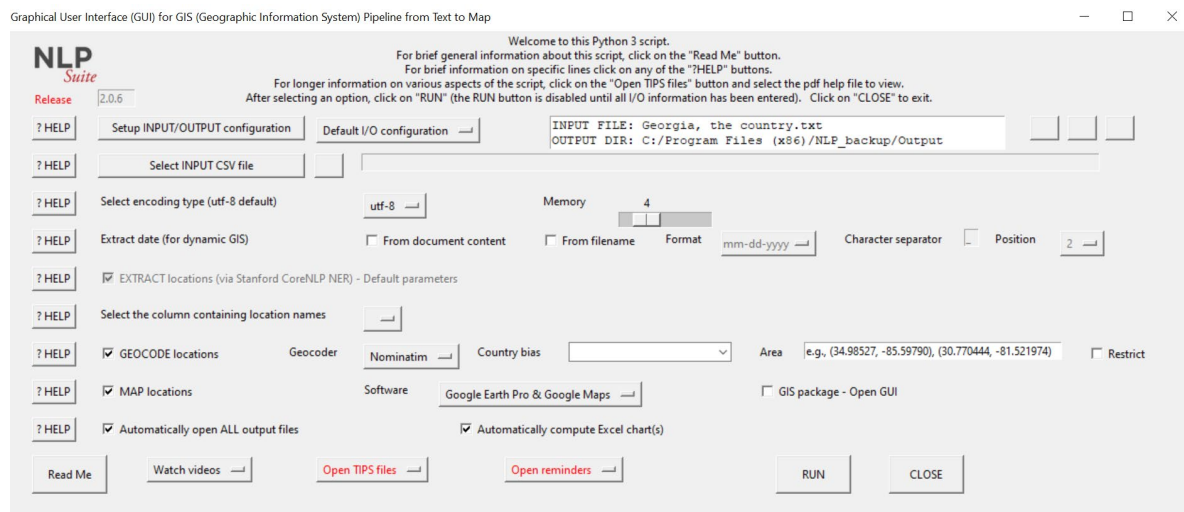
Other NLP Suite scripts (e.g., GIS_Google_Earth_main, SVO_main, whats_in_your_corpus_main), also have geocoding capabilities. GIS_Google_Earth_main, Stanford_CoreNLP_SVO_main, and whats_in_your_corpus_main rely on Nominatim as the default geocoding option. If you wish to use Google for geocoding, please, use the GIS_main script.

Whichever option is chosen for geocoding (Nominatim or Google) the NLP Suite will carry out the geocoding automatically.

Finding locations (via Stanford CoreNLP NER)

Given a text, you will need to find all the locations mentioned in the text. The NLP Suite relies on the Stanford CoreNLP NER (Named Entity Recognition) annotator. This annotator will extract automatically from texts, continents, countries, cities, provinces/states, and more generic locations.

The GIS pipeline produces in output, besides the maps, csv files with the extracted NER locations, so that they can be checked for accuracy.

	Location	NER Value	tokenBegi	tokenEnd	Sentence I	Sentence	Document	Document
1	Location	NER Value	tokenBegi	tokenEnd	Sentence I	Sentence	Document	Document
2	Baltimore	CITY	7	8	12	Happiness	1	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\A Spool of Blue Thread_Anne Tyler_Rebecca Pepper Si
3	Manderle	LOCATION	9	10	15	The house	1	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\A Spool of Blue Thread_Anne Tyler_Rebecca Pepper Si
4	Baltimore	CITY	26	27	24	(Tyler has	1	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\A Spool of Blue Thread_Anne Tyler_Rebecca Pepper Si
5	West	LOCATION	27	28	59	Her take o	1	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\A Spool of Blue Thread_Anne Tyler_Rebecca Pepper Si
6	Wyoming	STATE_OR	14	15	6	It â€™s th	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
7	Brokeback	LOCATION	19	21	6	It â€™s th	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
8	Maine	STATE_OR	17	18	15	She vivifie	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
9	Canada	COUNTRY	33	34	16	And while	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
10	Maine	STATE_OR	35	36	16	And while	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
11	New Zeala	COUNTRY	37	39	16	And while	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
12	Michigan	STATE_OR	40	41	16	And while	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
13	Canada	COUNTRY	4	5	17	On his arr	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
14	China	COUNTRY	6	7	20	Around 17	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
15	New	LOCATION	1	2	35	In New Fra	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
16	France	COUNTRY	2	3	35	In New Fra	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
17	Brazil	COUNTRY	18	19	36	And as Ku	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
18	Amsterda	CITY	2	3	38	In an Ams	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
19	New	LOCATION	18	19	38	In an Ams	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
20	England	COUNTRY	19	20	38	In an Ams	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
21	Scotland	COUNTRY	4	5	40	He trades	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
22	Boston	CITY	2	3	57	â€œ In Bo	2	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Barkskins_Annie Proulx_William T Vollmann_06-17-20
23	Afghanist	COUNTRY	21	22	13	He â€™s a	3	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Career of Evil_JK Rowling Writing as Robert Galbraith_
24	London	CITY	11	12	14	At the out	3	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Career of Evil_JK Rowling Writing as Robert Galbraith_
25	Denmark	COUNTRY	13	14	17	As â€œ Ca	3	C:\Users\rfranzo\Desktop\NLP-Suite\lib\sampleData\newspaperArticles\Career of Evil_JK Rowling Writing as Robert Galbraith_

Geocoding locations

Geocoding is the process of converting addresses (e.g., "Chicago, Illinois" or "1600 Amphitheatre Parkway, Mountain View, CA") into geographic coordinates (like latitude 37.423021 and longitude -122.083739 for 1600 Amphitheatre Parkway, Mountain View, CA), which you can use to place markers on a map or position the map.
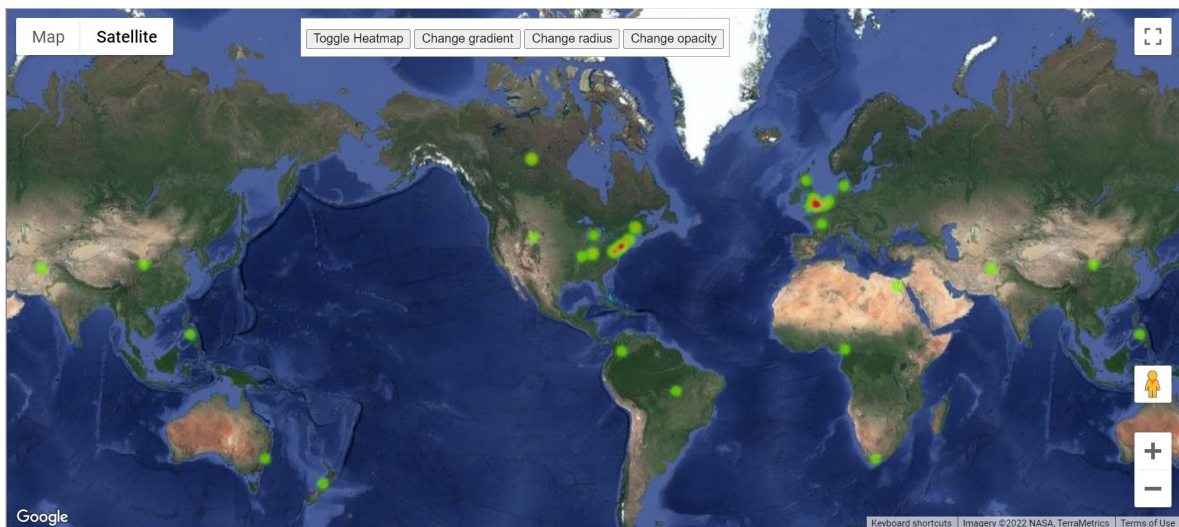
The NLP Suite offers two different freeware options for geocoding, via Nominatim and Google. Nominatim is an open-source, freeware geocoding tool. Nominatim uses **OpenStreetMap** (OSM) data to find locations on Earth by name and address (geocoding).

The GIS pipeline produces in output, besides the maps, csv files with the geocoded locations, so that they can be checked for accuracy.

	Location	Latitude	Longitude	Address						
1	Location	Latitude	Longitude	Address						
2	Afghanista	33.76801	66.23851	Afghanistan						
3	Amsterdar	52.37276	4.893604	Amsterdam, North Holland, Netherlands						
4	Ashbury	-33.9001	151.1181	Ashbury, Inner West, Sydney, Canterbury-Bankstown Council, New South Wales, 2193, Australia						
5	Baltimore	39.29088	-76.6108	Baltimore, Maryland, United States						
6	Baltimore	39.29088	-76.6108	Baltimore, Maryland, United States						
7	Boston	42.35543	-71.0605	Boston, Suffolk County, Massachusetts, United States						
8	Boston	42.35543	-71.0605	Boston, Suffolk County, Massachusetts, United States						
9	Brazil	-10.3333	-53.2	Brazil						
10	Bridge	51.24559	1.125283	Bridge, Canterbury, Kent, South East England, England, United Kingdom						
11	Brokeback	4.646186	-74.0625	Brokeback Mountain Cafe Bar, 9-28, Calle 60, Chapinero, UPZ Chapinero, Localidad Chapinero, Bogota, Bogota Capital District - Municipality, 110231, Colombia						
12	Cafe	10.83661	122.5588	Cafe, Iloilo, Western Visayas, 5002, Philippines						
13	Canada	61.06669	-107.992	Canada						
14	Canada	61.06669	-107.992	Canada						
15	China	35.00007	104.9999	China						
16	Denmark	55.67025	10.33333	Denmark						
17	England	52.53102	-1.26491	England, United Kingdom						
18	Euston Sta	51.52917	-0.13478	Euston railway station, Euston Square, St Pancras, London Borough of Camden, London, Greater London, England, NW1 2RT, United Kingdom						
19	Flushing	51.49434	3.415006	Vlissingen, Zeeland, Netherlands						
20	France	46.60335	1.888334	France						
21	George	-33.9644	22.45972	George, George Local Municipality, Garden Route District Municipality, Western Cape, 6529, South Africa						
22	London	51.50732	-0.12765	London, Greater London, England, United Kingdom						
23	London	51.50732	-0.12765	London, Greater London, England, United Kingdom						
24	London	51.50732	-0.12765	London, Greater London, England, United Kingdom						
25	Maine	45.7091	-68.859	Maine, United States						

On geocoding, please read the following TIPS files.

TIPS_NLP_GIS_Geocoding.pdf

TIPS_NLP_GIS_Geocoding Nominatim.pdf

Mapping locations

The NLP Suite relies on Google Earth Pro and Google Maps to display maps, as pin maps (where each location is mapped with a specific pin) or heat maps, where most frequent locations are shown with different colors (from green, less frequent, to red, the "hottest", most frequent locations. Heat maps are particularly useful to visualize instantly the "hottest" spots of a text.

Google Earth Pro (Pin Map)

Google Maps (Heat Map)

On Google Earth Pro, see the following TIPS files.

TIPS_NLP_GIS_Google Earth Pro.pdf

TIPS_NLP_GIS_Google API Key.pdf

TIPS_NLP_GIS_Google Earth Pro Icon.pdf

TIPS_NLP_GIS_Google Earth Pro HTML.pdf

TIPS_NLP_GIS_Google Earth Pro From KML to Excel.pdf

TIPS_NLP_GIS_Google Earth Pro Description.pdf

References

There are several TIPS files in the NLP Suite that deal with GIS issues.

TIPS_NLP_GIS_Geocoding.pdf

TIPS_NLP_GIS_Geocoding Nominatim.pdf

TIPS_NLP_GIS_Google Earth Pro.pdf

TIPS_NLP_GIS_Google API Key.pdf

TIPS_NLP_GIS_Google Earth Pro Icon.pdf

TIPS_NLP_GIS_Google Earth Pro HTML.pdf

TIPS_NLP_GIS_Google Earth Pro From KML to Excel.pdf

TIPS_NLP_GIS_Google Earth Pro Description.pdf