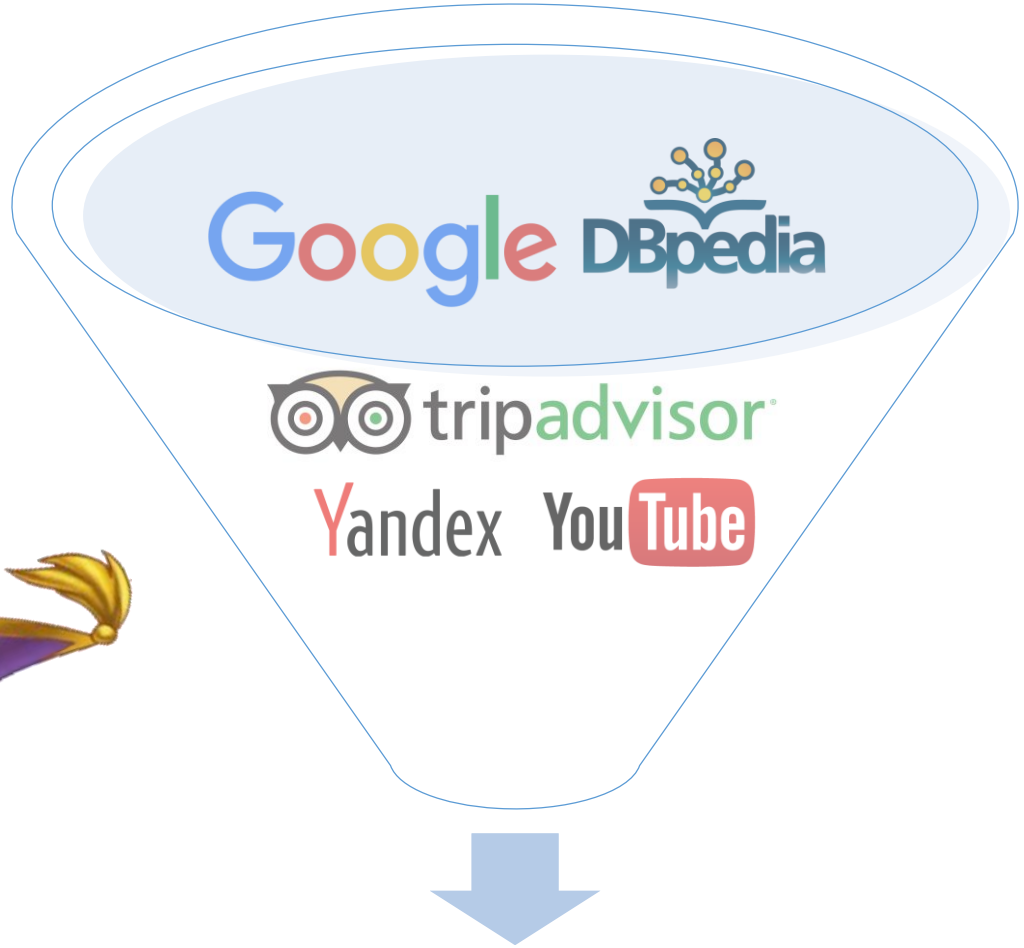


Enroute-Genie

A Coherent 1-Stop solution to explore the world

www.enroute-genie.com

https://github.com/raviraju/Enroute-Genie

Rashmi Nalwad

Ravi Raju Krishna

Aladdin's ancient carpet was found in the Cave of Wonders' treasure room

Our Software Carpet emerged from IIW course at WPH B27

Google

places to visit between los angeles and san francisco blogs

All | News | Maps | Images | Shopping | More | Settings | Tools

About 8,950,000 results (0.77 seconds)

Over 1000km you'll see stunning coastal views, seaside villages, untouched forest and make your way through some sunny wine making districts too.

- San Francisco. The Golden Gate Park is a great place to start this epic trip. ...
- San Jose. ...
- Santa Cruz. ...
- Monterey. ...
- Big Sur. ...
- San Simeon. ...
- San Luis Obispo. ...
- Santa Barbara.

More items...

13 Incredible Stops on a Pacific Coast Highway Road Trip - Gap Year
https://www.gapyear.com/articles/.../13-incredible-stops-on-the-pacific-coast-highway

About this result • Feedback

Five Great Stops between Los Angeles and San Francisco on the ...
www.hopper.com/.../five-great-stops-between-los-angeles-and-san-francisco-on-the-p...
May 14, 2014 - Discover the towns and attractions that dot the California coast between Los Angeles and San Francisco. ... Five Great Stops between Los Angeles and San Francisco on the Pacific Coast Highway - Hopper Blog ...

Planning a California Pacific Coast Highway Road Trip from San ...
independenttravelcats.com/.../planning-a-california-pacific-coast-highway-road-trip-fr...
Nov 24, 2015 - San Francisco* Half Moon Bay* San Jose / Silicon Valley (30-40 mile detour!) Santa Cruz* Monterey* Carmel-by-the-Sea* Big Sur* Ragged Point.

Drive the Pacific Coast Hwy from LA to San Francisco! - Blogger at Large
www.bloggeratlarge.com/.../drive-the-pacific-coast-hwy-from-la-to-san-francisco-2/
Jan 1, 2015 - Here is an itinerary of places to see and stay along the way. ... Drive the Pacific Coast Hwy

United States › California (CA) › San Luis Obispo County › San Simeon › Things To Do in San Simeon › Hearst Castle

Hearst Castle

4,990 Reviews | #2 of 14 things to do in San Simeon | Certificate of Excellence

Castles, Historic Sites, Sights & Landmarks, More

Overview | Tours & Tickets | Reviews (4,990) | Q&A (86) | Location | Book a Tour | Save | Write a Review

All visitor photos (3790)

Book In Advance

Santa Barbara, Solvang and Hearst Castle Day Trip from Los Angeles $169.00* and up — More Info

Why book on Viator?
- Tour highlights & full itinerary
- Easy online booking
- Lowest price guaranteed

More Information ›

Would this be a good cold day activity? Yes No Unsure

Get directions

Hearst Castle Visitor Center

Sponsored by:

Map data ©2016 Google

Address: 750 Hearst Castle Road, San Simeon, CA 93452-9740

Phone Number: +1 800-444-4445

Website

Improve this listing

Today 8:00 am - 3:00 pm Closed now

See all hours

TripAdvisor Reviewer Highlights

Read all 4,990 reviews

Visitor rating

Excellent 2745

Very good 1167

"Ornate history"

This was a great half-day experience. The visitor's center is large and has plenty to do, including a large-screen film on the building of Hearst Castle. I would highly recommend... read more

top attractions in san fra

https://www.youtube.com/results?search_query=top+attractions+in+san+francisco

Apps | Career | OnlineCourses | M/C LearnEd | DataMining | IBM_Watson | USC | ITS | IIW | NLP | Big Data Edu

YouTube

top attractions in san francisco

Home
Trending
History
Get YouTube Red

BEST OF YOUTUBE

Music
Sports
Gaming
Movies
TV Shows
News
Live
Spotlight

Filters

San Francisco Top 10 Travel Attract
California Travel Guide
geobeats
6 years ago • 357,802 views
Take a tour of Top 10 Travel Attractions of San Francis
the World's Greatest Attractions series by GeoBeats.
2:04

15 Travel Tips & Best Places in San
Reformatt Show
5 months ago • 27,107 views
Do you need the Best San Francisco Travel Guide? Lot
finding the most unique attractions, things to do, ...
SAN FRANCISCO 15 4:20

San Francisco: Top Ten Things To D

Carpinteria, California

https://en.wikipedia.org/wiki/Carpinteria_California

Apps | Career | OnlineCourses | M/C LearnEd | DataMining | IBM_Watson | USC | ITS | IIW | NLP | Big Data Edu | Big-O Algorithm Com | Bay Area Learning A | On Campus Recruitin

Not logged in Talk Contributions Create account Log in

Article Talk | Read Edit View history | Search Wikipedia

WIKIPEDIA The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export

Carpinteria, California

From Wikipedia, the free encyclopedia

Coordinates: 34°23'57"N 119°30'59"W

For other uses of "Carpentaria", see Carpentaria (disambiguation).
"Carpinteria" redirects here. For the town in San Luis Province, Argentina, see Carpinteria (San Luis).

Carpinteria /kɑ:rpinte´ri:ə/ is a small oceanside city located in southeastern Santa Barbara County, California, east of Santa Barbara and northwest of Ventura. The population was 13,040 at the 2010 census.

Carpinteria Beach is known for its gentle slope and calm waves in selected sandy areas but also good surfing swells in some of the more rocky areas.[18] Seals and sea lions can be seen in the area December through May at the rookery [19] in the nearby Carpinteria Bluffs,[12] as well as an occasional gray whale. Tidepools contain starfish, sea anemones, crabs, snails, octopuses and sea urchins. A marathon-length round trip north of the rookery along the beach to Stearns Wharf in Santa Barbara is possible, though passable only during low tide. A popular campground is located adjacent to the beach. There is bird watching at Carpinteria Salt Marsh Reserve, established in 1977 and administered by the Natural Reserve System of the University of California.[13] The Waldholme Torrey Pine, largest known Torrey pine tree on earth, is located in downtown Carpinteria.

Since 1987, the California Avocado Festival has been held in Carpinteria on the first weekend of October. The Santa Barbara Polo Club, one of the main equestrian polo fields in the country, is located in Carpinteria. The city is also home to Hollandia Produce, an organic produce company with 70 employees.[citation needed] lynda.com, an online software training company ranked as one of the fastest-growing private companies in the U.S. (according to *Inc.* magazine's 2010 500|5000 company listing) had its headquarters in Carpinteria. The company was purchased by LinkedIn in 2015 for $1.5 billion.[14]

Contents [hide]
1 History
2 Geography

City of Carpinteria
City


Sunset on the beach (end of Linden Ave)

hearst castle: 26 thousan

https://yandex.com/images/search?text=hearst%20castle

Apps | Career | OnlineCourses | M/C LearnEd | DataMining | IBM_Watson | USC | ITS | IIW | NLP | Big Data Edu | Big-O Algorithm Com | Bay Area Learning A | On Campus Recruitin

Yandex | Images | hearst castle | Search | Log in

Back to search | Size | Orientation | Type | Color | Format | Recent | Wallpaper 1366×768 | On site

Enroute Genie

www.enroute-genie.com

Career | OnlineCourses | M/C LearnEd | DataMining | IBM_Watson | USC | ITS | IIW | NLP | Big Data Edu | Big-O Algorithm Com | Bay Area Learning An | On Campus Recruiting

Enroute Genie Mashup

A Data Integration Solution powered by Google tripadvisor DBpedia Yandex YouTube

Start From: Select an end point... **End At:** Select another end point... Go Genie !!!

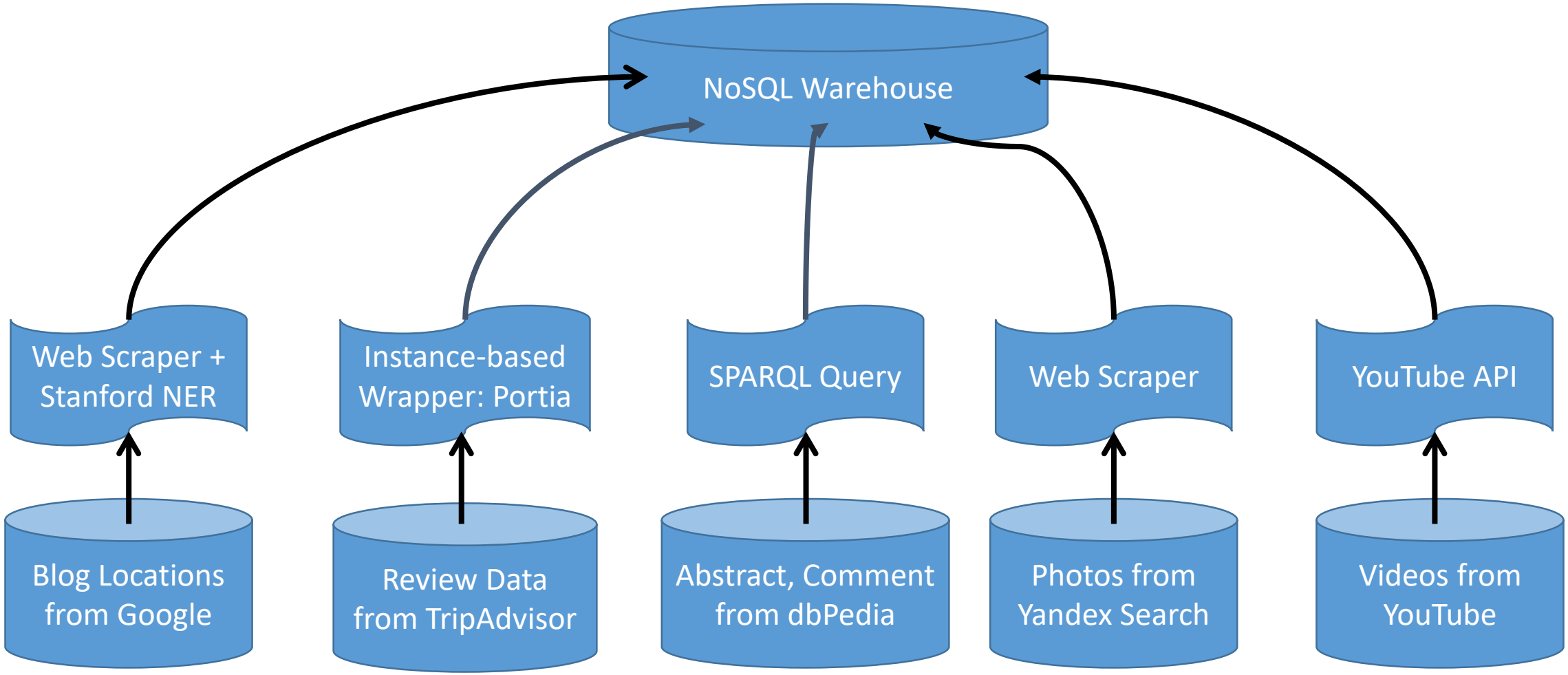
Learning Objectives of Course	Project Objectives
Understand the theory and techniques of traditional data integration, including logical view integration, schema mapping, and **record linkage**	Link Records from Google, TripAdvisor & DBPedia
Understand the foundations and techniques of the **Semantic Web**, including RDF, OWL, **SPARQL**, **linked data**	Query data from DBPedia – the semantic web version of Wikipedia, and add HTTP URI to rdf sources adhering to principles of linked data
Understand the theory and application of the state-of-the-art software and tools for **information extraction**	Scrape Location mention from blogs using customized NER
Understand the algorithms and techniques for **data cleaning**, source modeling, semi-structured extraction, and information extraction	Configure an Instance based Wrapper Induction Extraction system to fetch structured information from TripAdvisor and use Trifacta Wrangler to refine the results.
For **any given integration problem**, be able to select and apply the most relevant information integration techniques to solve that problem	Build Enroute-Genie

Modest Goals of data integration system

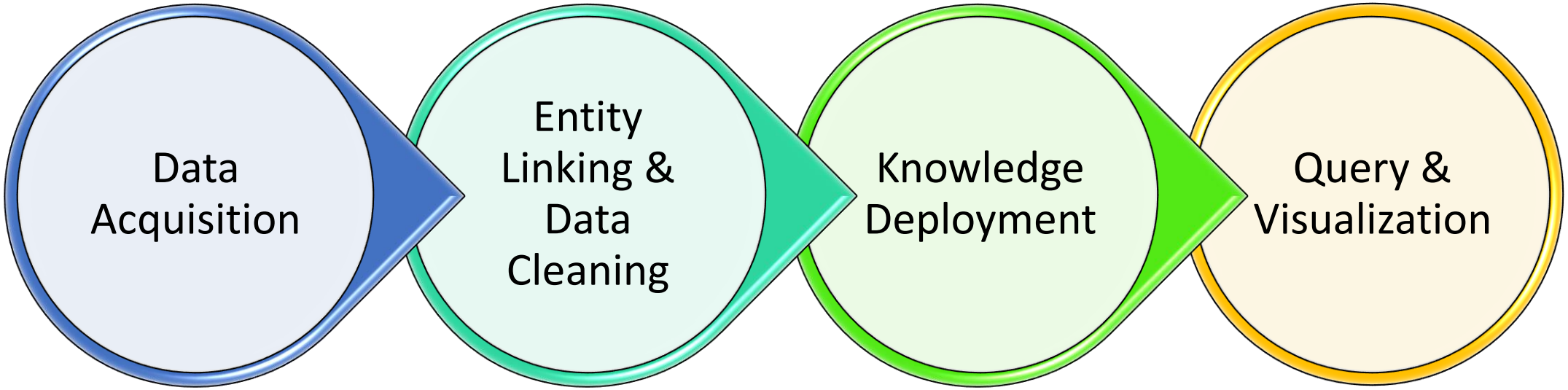
- User Effort : *To build tools that reduce the effort required to integrate a set of data sources. Make it easy to add new sources*
  - * Currently addressing State of California.
  - * Can be easily extended to other states.
  - * Triggered extension to Karnataka, India, during development.
- Accuracy : *To improve the ability of the system to answer queries in uncertain environments*.
  - * We are using current state of art for NER – StandfordCoreNLP to identify locations.
  - * Our accuracy improves adapting to future advances in NER

  [NOTE : See Evaluation Results for more info…]

Architecture



Pipeline


Data Acquisition
Entity Linking & Data Cleaning
Knowledge Deployment
Query & Visualization

Data Sources

Variety dimension in 3 V's of Big Data

1. Unstructured : Blog text Scraped from Google search results
2. Structured : Abstract and Comment of a city from dbPedia, TripAdvisor Review Data
3. Multimedia : Photos and YouTube Videos

Accessed:

Information Extraction

1. Scraper(Blog Links from Google & ImageUrls from Yandex)
2. Mine the Deep Web Data : Wrapper(TripAdvisor data)
3. SPARQL Query(dbPedia Info)
4. Web Service API(Youtube API)

Data Acquisition

Identify popular destinations of a State

For every pair Google for 'places to visit between ' + city_a + ' and ' + city_b + ' blogs'

Extract Content using **Beautiful Soup** and get NER Location Annotation from **Stanford CoreNLP**

Get geoCode using **GeoPy** and filter out distant/irrelevant location mentions

Fetch top 3 video URLs using **YouTube API**

City : top attractions in San Francisco

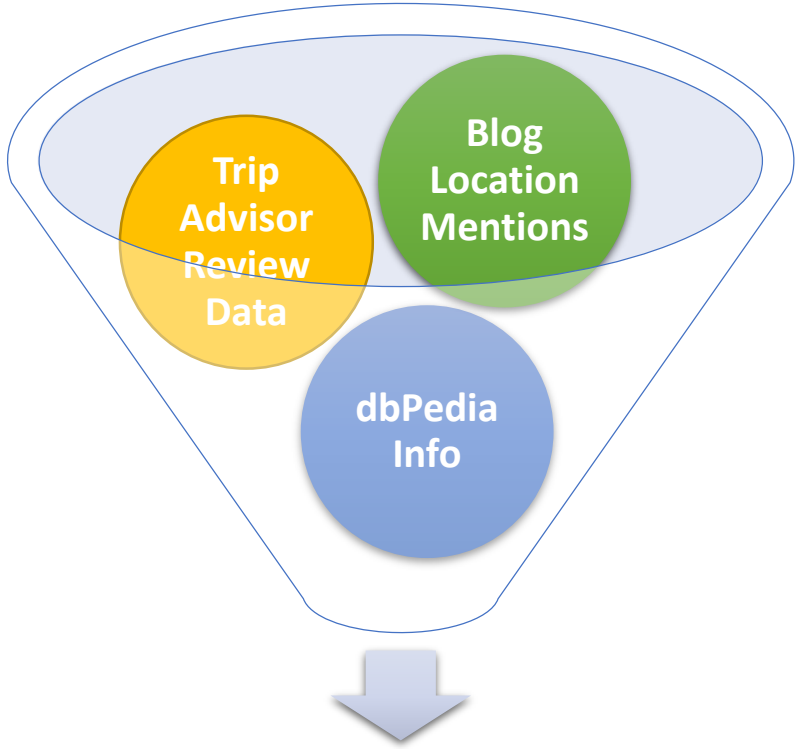
Location : Golden Gate Bridge

Configure an Instance-Based Data extraction wrapper-**Portia** to extract attraction review metadata

Clean Data using **Trifacta Wrangler**

Query dbPedia for Abstract and Comment of locations partOf California Using SPARQL

Entity Linking – using FRIL



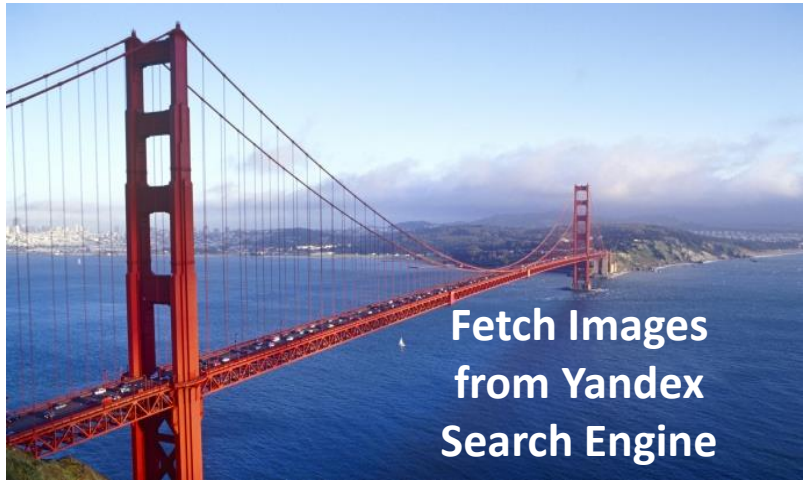
Record Linked Data

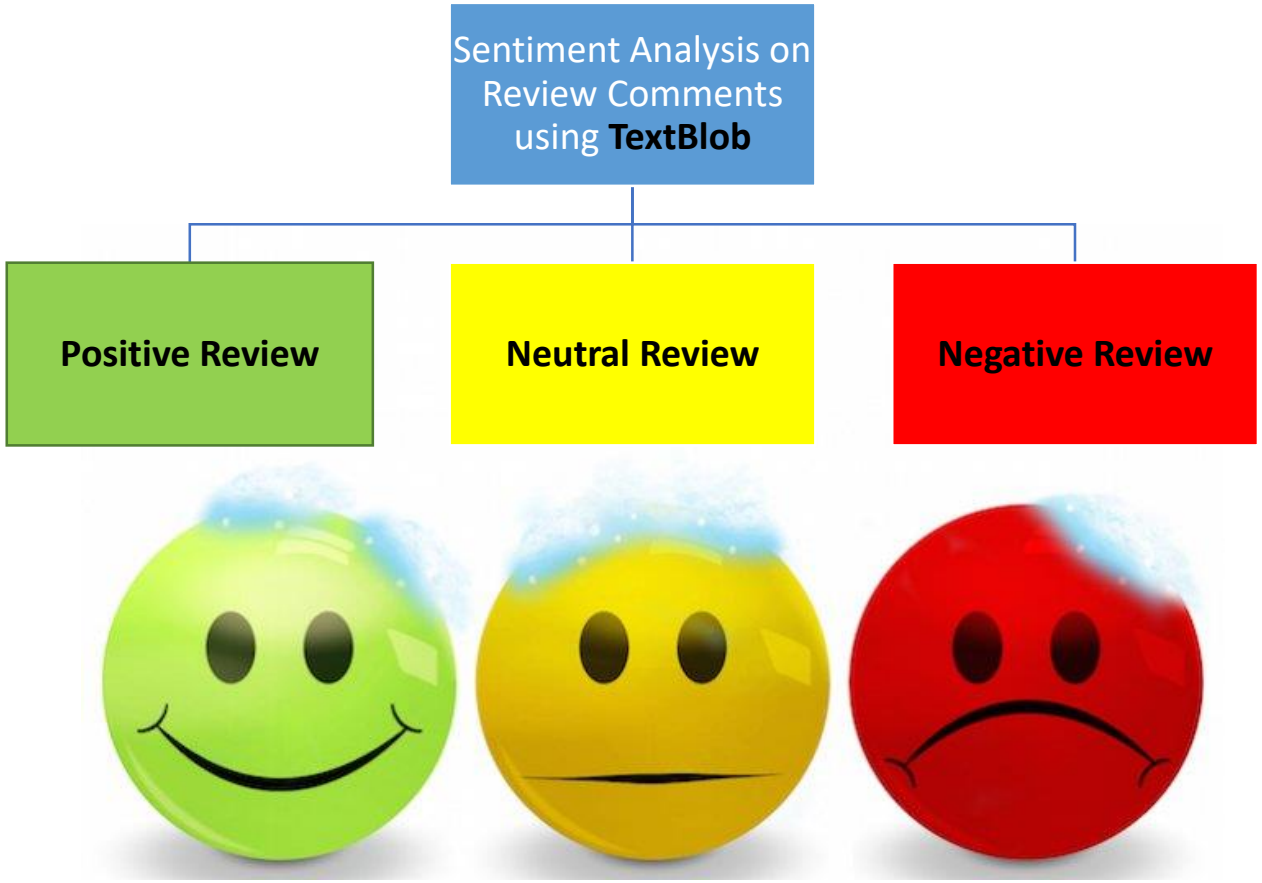
Edit-Distance : (mention_name,attraction_name),
(match-level-start=0.1, math-level-end=0.3)
weight="50"

Numeric-Distance : (mention_latitude, attraction_latitude)
exactMatch
weight="25"

Numeric-Distance : (mention_longitude, attraction_longitude)
exactMatch
weight="25"

Enrich Record Linked Data



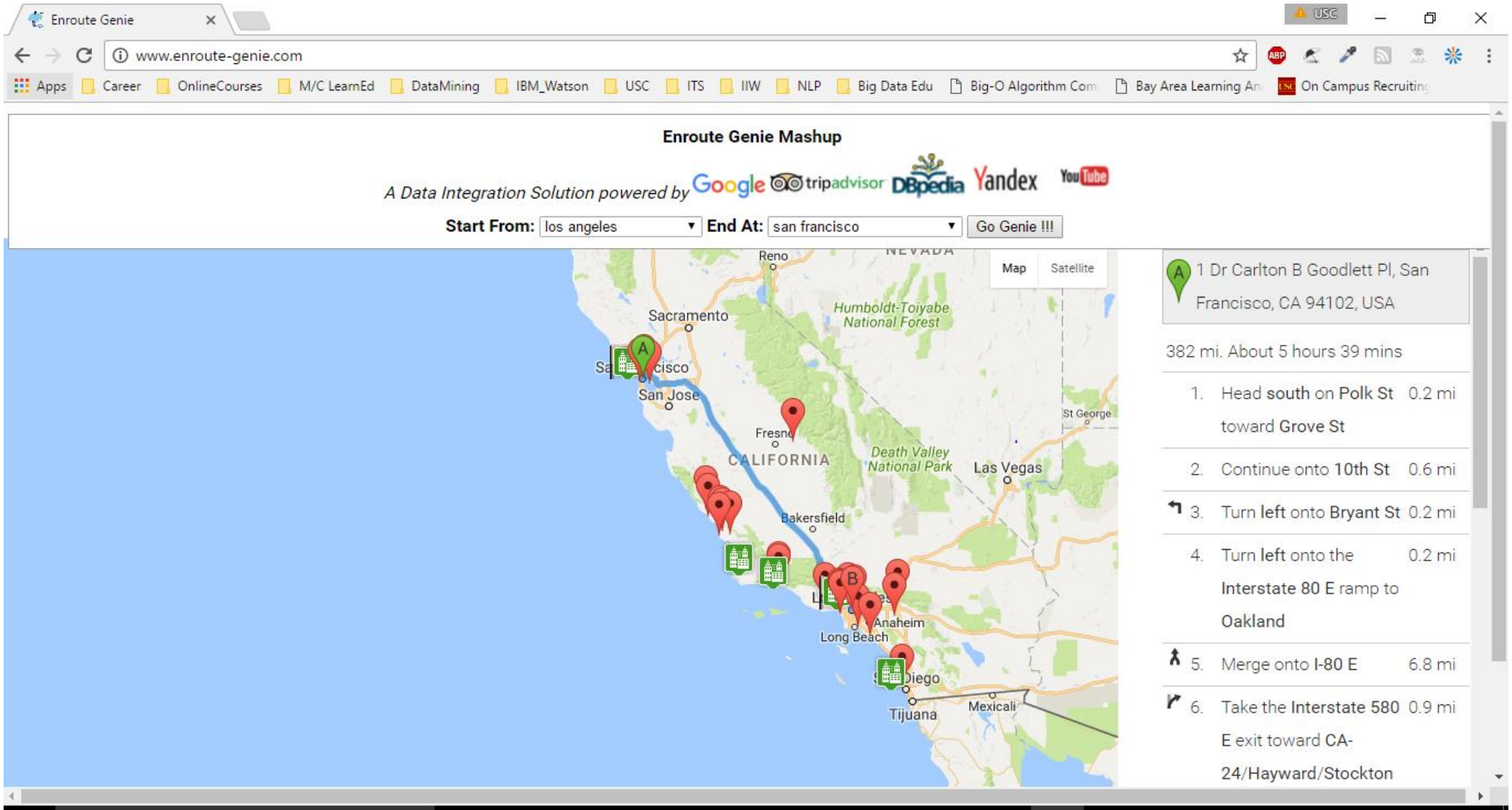


Knowledge Deployment

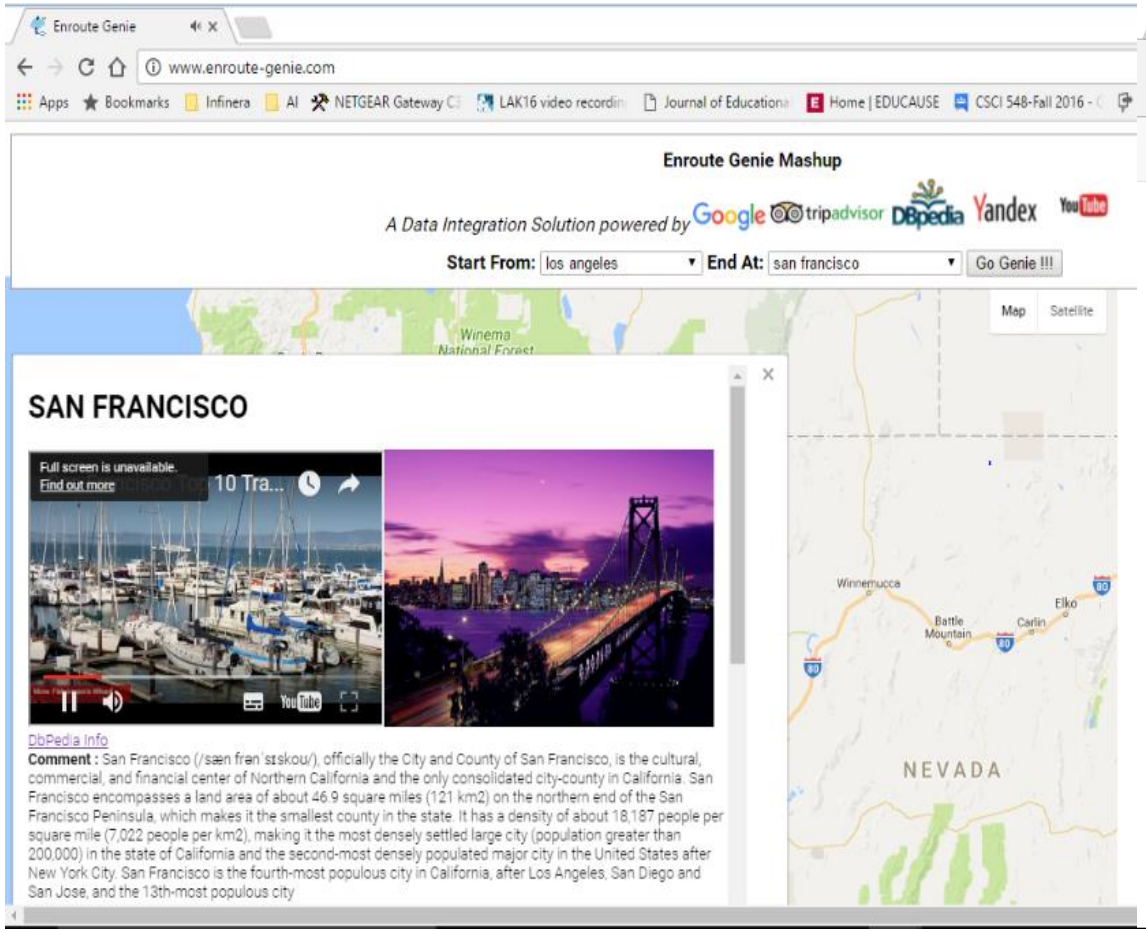
- Integrated Data loaded to NoSQL-CouchDB
- Web Service to share our integrated data with other applications:

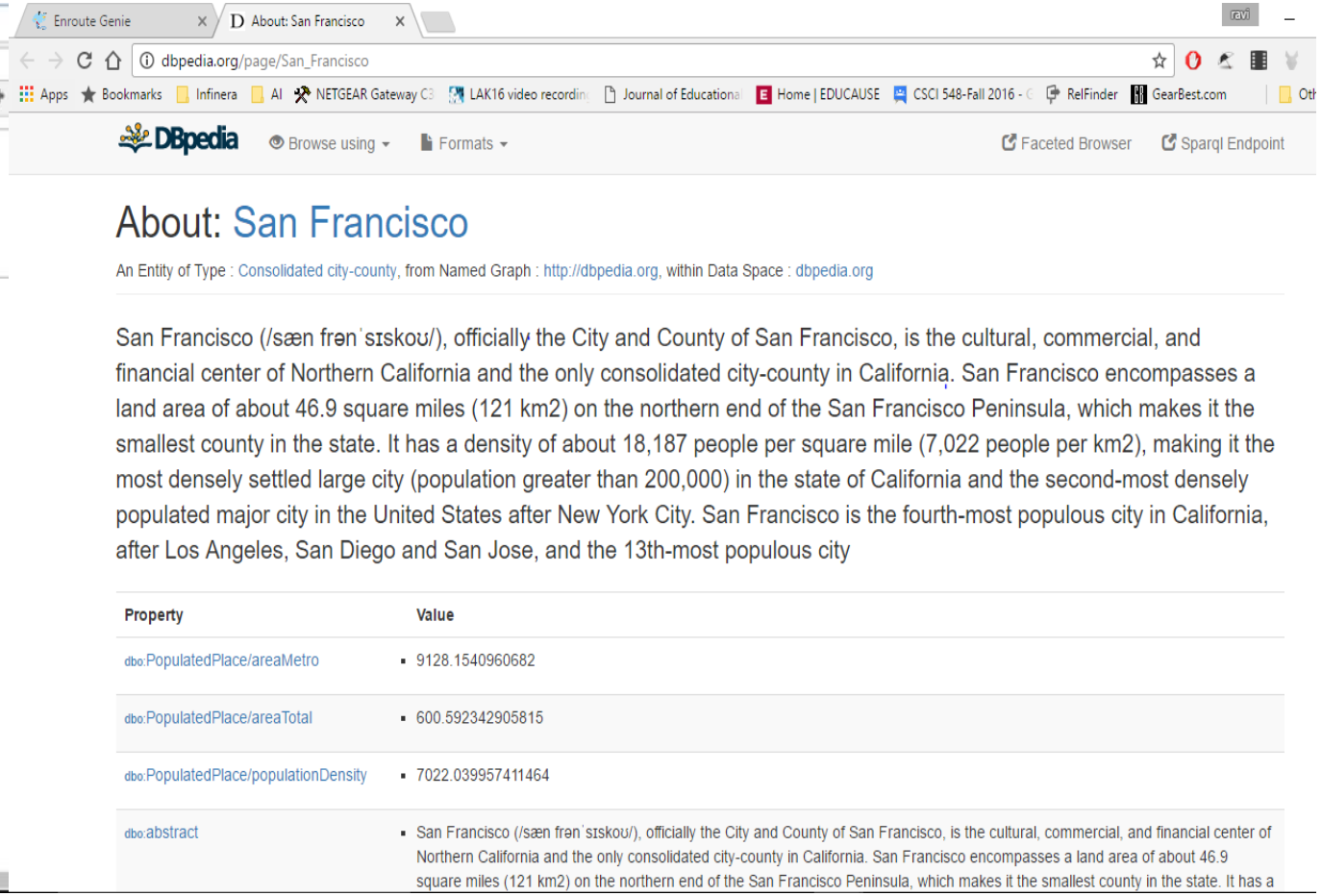
  HTTP API Access to our DB
- https://enroutegenie.cloudant.com/enroute_genie/_all_docs
- https://enroutegenie.cloudant.com/enroute_genie/los%20angeles_and_san%20francisco

Query & Visualization

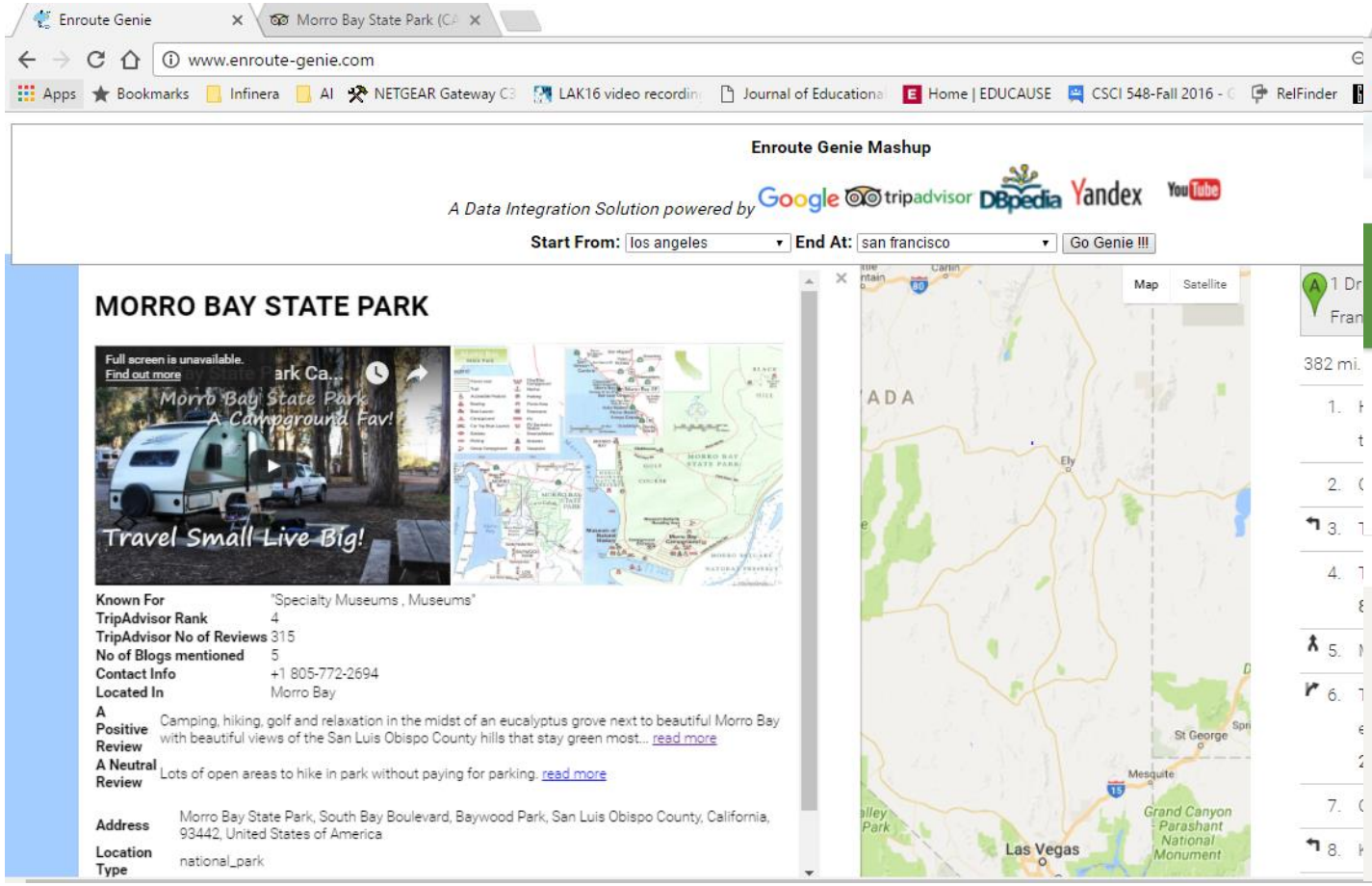


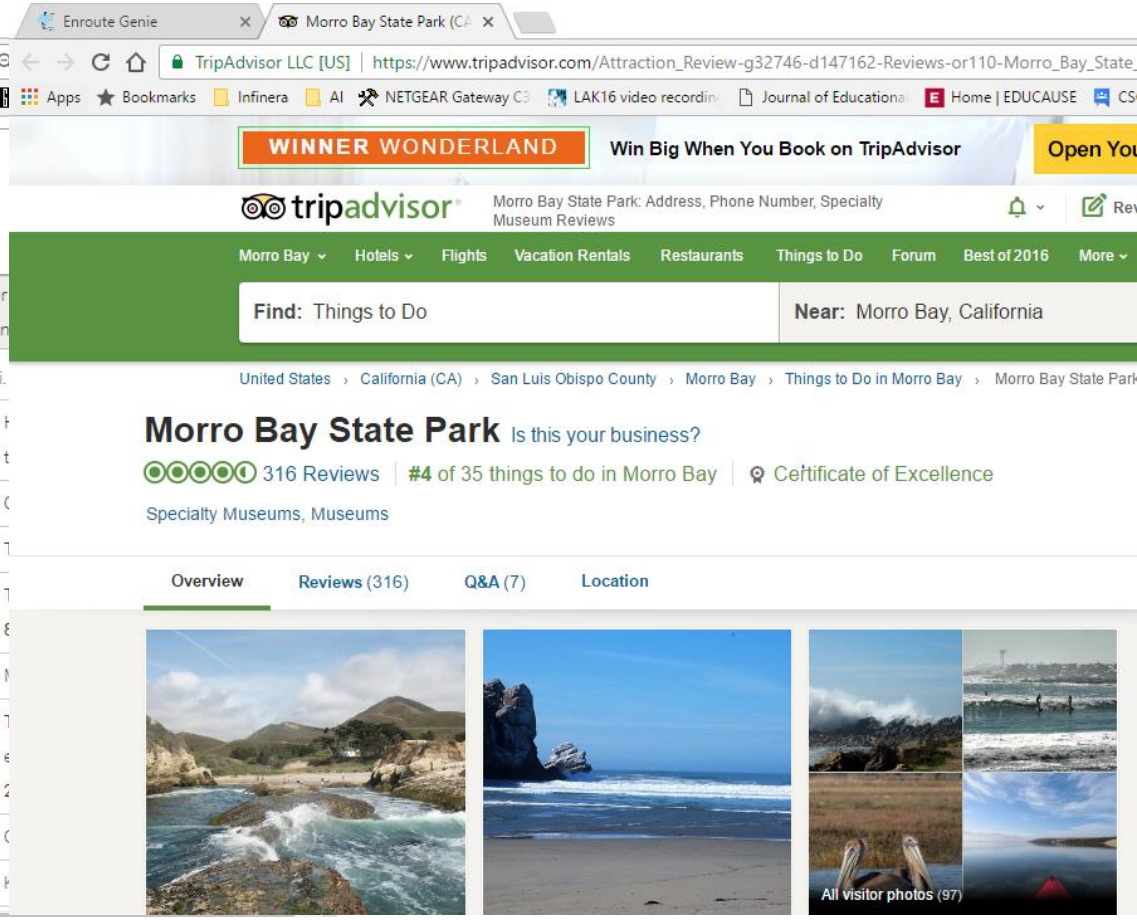
City Attractions



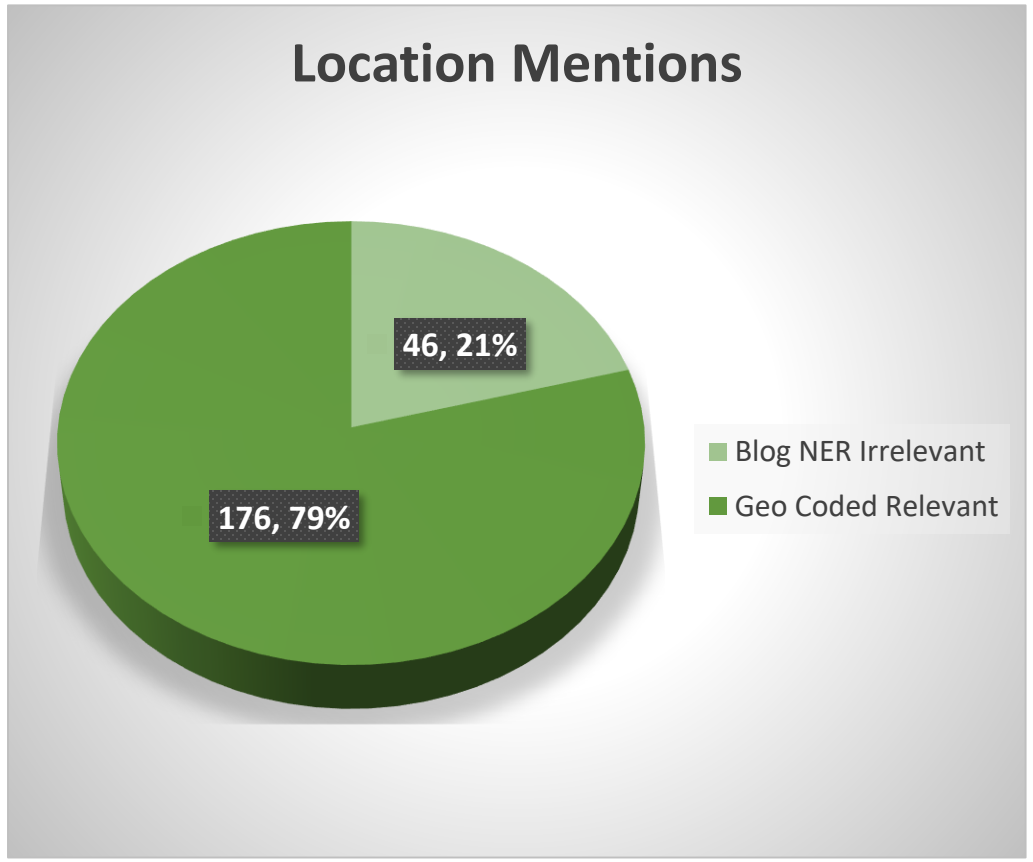


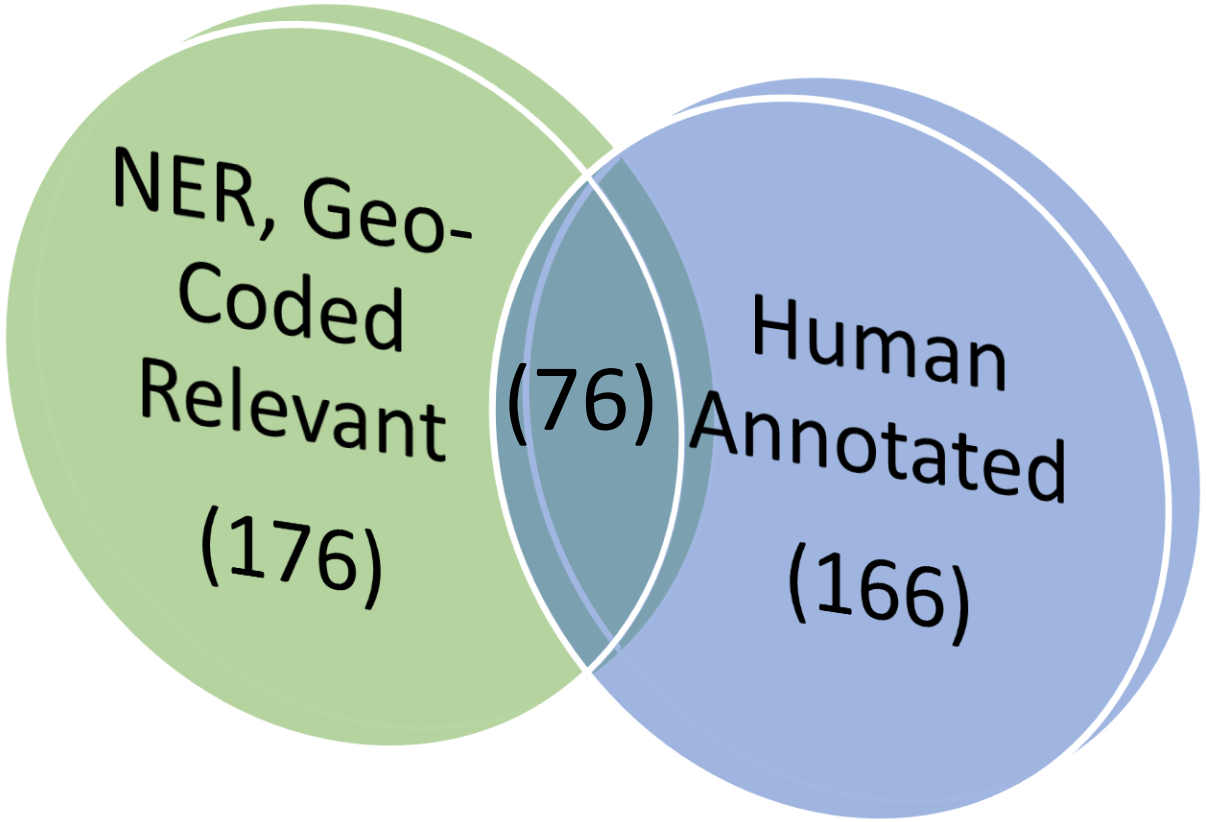
Location Mentions





Evaluation – Los Angeles and San Francisco





Data Integration or Tools Integration ?

Feature	Library	Reference
Blog Text	Google Scraper-"Bing Search Engine“ Beautiful Soup	https://github.com/NikolaiT/GoogleScraper https://www.crummy.com/software/BeautifulSoup/bs4/doc/
NER Location	Stanford CoreNLP	http://stanfordnlp.github.io/CoreNLP/
Geo Coding	GeoPy	http://geopy.readthedocs.io/en/latest/
TripAdvisor Data	Portia, Scraping Hub	https://doc.scrapinghub.com/ https://doc.scrapinghub.com/portia.html
Data Cleaning	Trifacta Wrangler	https://www.trifacta.com/start-wrangling/
Record Linking	FRIL	http://fril.sourceforge.net/
Sentiment Analysis	Text Blob	https://textblob.readthedocs.io/en/dev/api_reference.html#textblob.blob.TextBlob.sentiment
Videos	YouTube API	https://developers.google.com/youtube/v3/code_samples/python#search_by_keyword
Images	Google Scraper-“Yandex Search Engine“	https://github.com/NikolaiT/GoogleScraper

More on Grading

- This is a hard class, but you will learn a lot!
  - Principles and theory
    - Technical readings and lectures (quizzes, final exam)
  - Putting principles into practice
    - Homeworks and project!
- Grade distribution

94 - 100 = A	74 - 76 = C
90 – 93 = A-	70 - 73 = C-
87 – 89 = B+	67 - 69 = D+
84 – 86 = B	64 - 66 = D
83 – 83 = B-	60 - 63 = D-
77 – 79 =C+	Below 60 is an F

Questions ?