Shark Tank Presentation

1. The goal of the project

Hope most of u have used Maps to travel around. Were U ever disappointed to missout on a wellKnown interesting place during ur journey.
Do Not Worry, just give our **EnrouteGenie** ur Src and Destination, it shall suggest u .

2. The data sources you plan to use.

EnrouteGenie gets its intelligence from all different forms of sources
I.e by
reading related blogs, which is an unstructured source,
extracts ratings and reviews from semi-structured source : tripadvisor,
Some factual data from structured source : dbPedia
pins these locations onto Map with the brief info, reviews, along with rich multimedia content : image and video.

---

3. The integration problem and integration techniques you will use.

Excited ??? Lets get into techniquealities.....

- We extract locations and labels(knownFor : Waterfall, Historical Place, Hindu Temple) using **NER and CRF classifier**.
- And irrelevant locations will be filtered with Google Distance Matrix API.
- Reviews about those locations are extracted from TripAdvisor using **Instance based Wrapper** in Portia.
- Some abstract about location from dbPedia using **sparql query**

The data collected thus far will be stored into a NoSQL Database integrated with an image and a youtube video links

---

4. Evaluation of project. How would you/we know that your project accomplished its goals.

Now Users can use Web App,Key-In Src and Destination, BAMMMMM
U shall get an embedded google Map will interesting locations pinned, which can be clicked to display Modal with
(image, brief info, knownFor label tags, youtube video, ratings and reviews)

Evaluation: Manually walk through and identify stuff : Training (8) & Test (2)
NER + distance matrix : #no_of_outlier locations threshold of 5, the least the better, more than 5, worse…
CRF : knownFor label : precision,recall,F-score
Wrapper : appropriate review comments and ratings
Sparql : manually cross-validate and report any discrepancies

Web App:

1) Search: Autocomplete feature of google Maps. Only City names to be entered.

Data Sources :

1) Google maps API : To get map path
2) dbPedia : Get all Cities of a State
   Abstract
3) Google search python module : To fetch blog links
4) Blogs : To get places suggested from people
5) Google Distance Matrix API : To filter irrelevant NER locations from blogs
6) TripAdvisor : To fetch Things_ToDo for relevant NER locations
7) Google Image : To display image in modal
8) Youtube : To display a video if it exists.

Steps:

Crawling & Extraction:

1) State of California: Get all Cities names from dbPedia
   http://dbpedia.org/page/List_of_counties_in_California
   http://dbpedia.org/page/List_of_counties_in_New_York
2) Extract Src,Destination Cities from Query
   a) Google Maps(lat,long) → Path to be displayed on Maps using Google maps API.
   b) Blogs : Fetch Top Results for Google Search “Src to Destination blogs”
3) Using combination of OpenNLP,CoreNLP to extract NER locations from blogs.
4) Filter irrelevant NER locations using Google Distance Matrix API.
5) CRF : to identify knownFor for relevant locations
6) Get Places of interest using Things to do of TripAdvisor for Relevant locations.
7) Fetch an image and video from google images and youtube
8) Faceted Search : knownFor

Mashup Visualization :

1) Map with path and places of interest pinned
   http://la.curbed.com/maps/best-los-angeles-picnic-spots
2) Click each pin to display modal of info
3) [OPTIONAL] : Sentiment Analysis of Reviews into +ve and -ve

Pipeline Design : For src,destination pairs which aren’t in database, record those into file and update database

DON’T KNOW:

Ontology Mapping, Entity Linking: Google Refine, **Building Knowledge Graph, ElasticSearch & GraphDB**