

Task Description

1. Use elastic search technology to fetch datasets from ISI service for human trafficking.
2. Tag <name> from the fetched datasets, and train a classifier to identify if it's a person name.
 - a) Actual <name> can get from <name> filed of JOSN for validation
 - b) Write or find machine learning algorithm in python that can get nice result.

Outputs

Project Folder Structure

RVIE is the root folder for this project, and you can find samples at test.py

```
| - data                                # main folder to store outputs
    | - json                            # raw data from elastic search
        | - sentences                  # all sentences contain names (JSON)
        | - names.json                # all names (JSON)
    | - raw                             # raw data for reverb, output from json parser
        | - sentences                 # sentences with period in the end
        | - names.raw                # all names
    | - reverb                          # reverb output
        | - rever.output              # single file contain all data from sentences
| - res                                # store resource
    | - reverb.jar                    # excutable jar file for reverb
| - plugin                             # plugins for core functions
    | - api                           # support for api functions
        | - js_parse.py               # parse json file
        | - curl_request.py           # manage curl connect
        | - query_builder.py          # generate query string for elastic search
        | - rv_extractor.py           # helper functions for ReVerb
| - core                               # core functions
    | - parser.py                     # parse json file for reverb
    | - streamer.py                   # fetch data by elastic search
    | - extractor.py                  # use reverb to extract information
| - api.py                             # main entrance
| - session.py                        # track the session of connection and process
```

ReVerb Output Evaluation

Examples that names cannot be extracted from sentences shown below

Hi	Hi babe im a 24yr old 42D Italian 145lbs busty trusty fun girl ...im available as of now until late ...ask for my specials & menu ...call me at 215-5839 - Amber\n
babe	
im	
a great physical therapist	Italian latina Amber here on b.p for a satisfying upscale clientele date ...message me for menu & donations im a great physical therapist i give sensual massages ...Calls preferred ...Texting to a minimal thnx ?
give	
sensual massages	