Introduction

• Challenge of Entity Resolution in the domain where information is scarcely available. Therefore, data is generated by the crowd over social multi-media platforms such as YouTube, Twitter, etc., Hence, data is ridden with subjective evaluations, opinions, and speculations.

• Data snippets contains multiple temple names and multiple temple location, therefore, additional problem of location disambiguation.

Methods

• We manually designed and wrote rules for parsing the textual data using JAPE Grammar in GATE tool.

• Extract temple name and temple location from the document di using disambiguation methods.

• Use Google Maps API to list temple names located around the extracted temple location. The temple names and temple location form a tuple ti stored in set T

• For each element t ∈ T, we calculate TF-IDF score for temple t over each document deD, where D is the indexed set of documents.

• We rank documents based on TF-IDF scores for each query ρT and map the top ranked d to the temple.

Results

• Our dataset consists of more than four hundred thousand temple names with their locations extracted from Google Places. It also contains more than two hundred thousand videos fetched from YouTube.

• We sample 1000 videos randomly from the complete video set to compute precision, recall and F-measure and evaluate the performance of videos mapped to the temple.

Precision = 0.863 , Recall = 0.89 , F = 0.876

Demonstration

• A demo of our system is accessible at tinyurl.com/templedemos

• Demo of the Snippet Ranking system can be accessed at tinyurl.com/entityr

References
