Madhura-Sreeja
Madhura Mathkar, Sreeja Pillai

Working of the Coreference Resolution System

The Coreference resolution system works collectively with the help of following:

- **POS TAGGING-** We used NLTK for part of speech tagging, these POS tags are further used for string matching, synonym search, finding the pattern of occurrence of a specific word.
- **NER TAGGING-** There are different NER tags such as PERSON, NORP, ORG, PRODUCT, DATE, TIME etc. These tags help identify the role of every word in a sentence. We have used NER for date and cardinal matching, full-forms.
- **STRING MATCHING** – At a very primitive level, we tried direct string matching of the coreference id along with other sentences, and found the sentences where a particular coreference occurs
- **DATE AND CARDINAL MATCH-** NER Tags helps us to identify the date and cardinal related words, we try matching it with the respective coreference, we are also able to map year and time for e.g. 1996 is mapped to year.
- **PRONOUN MATCH-** Pronoun match is implemented with the help of POS Tagging and NER Tagging. We find the 'PRP' POS tag and match it against the dictionary of 'PERSON', 'NORP' or 'ORG' NER tag which contains values such as he, she, it etc.
- **SYNONYM SEARCH-** We search for synonyms using NLTK synonyms and then search if a synonym of the coreference is occurring in a sentence
- **SEARCHING FOR FULLFORMS-** For full-forms we find the letters of a short form and search for words occurring consecutively with those letters. Then we make sure that the NER tag is an ORG and the POS Tag is NNP as full forms will be a proper noun which will be an organization in most cases
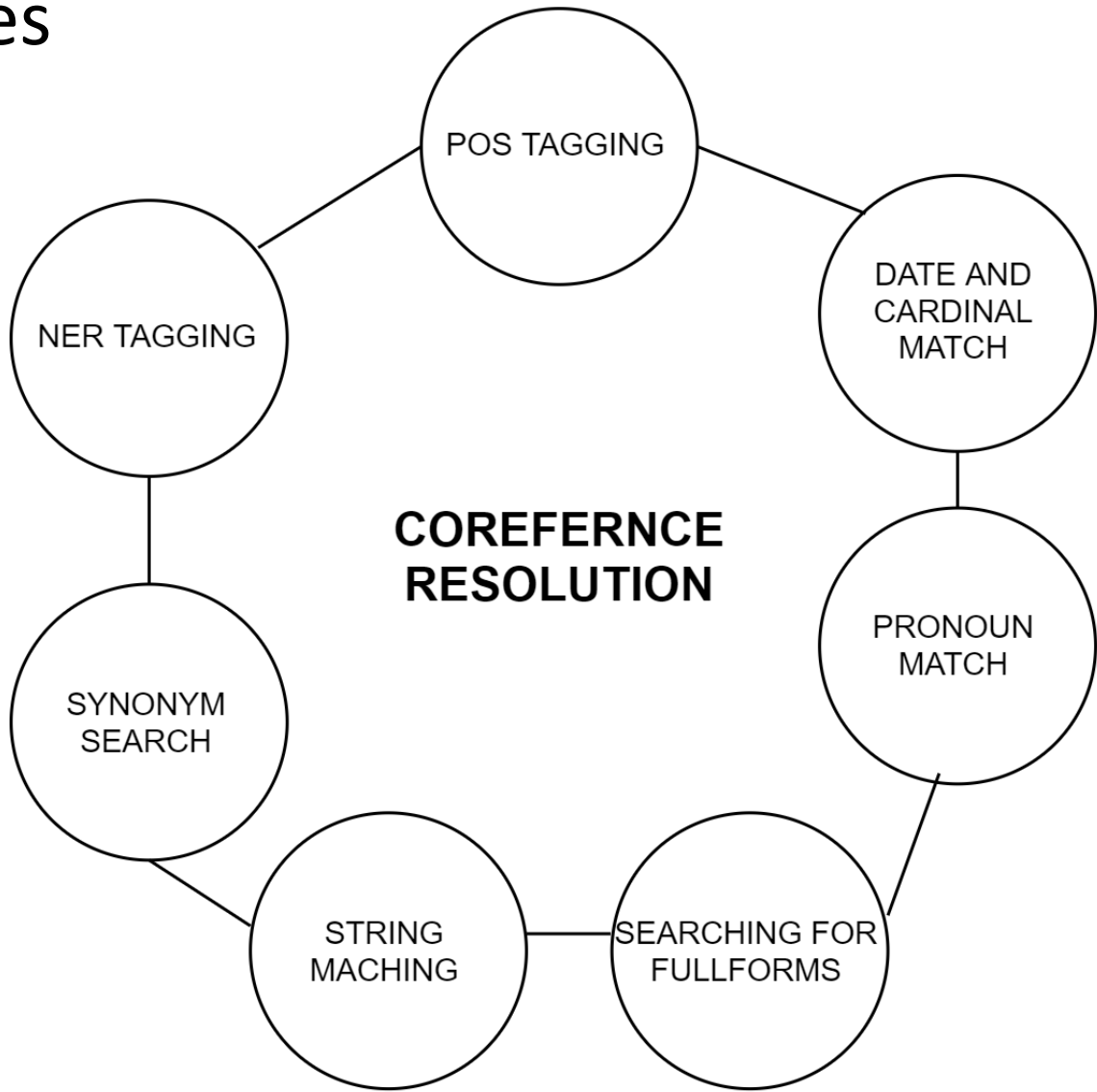

Fig1.
Representation of our Coreference Resolution System

Tools Utilized

- NLTK
- Spacy
- VS-Code

What works

We are able to generate the following-

- **Date:** for the given COREF ID we were able to match it in a sentence number 22
  <COREF ID="X2">03-27</COREF>
  {22}{03-27-96}
- **Full-form:** for the given COREF ID we were able to match it in a sentence number 11
  <COREF ID="X3">FAA</COREF>
  {12}{FAA}
  {19}{FAA}
  {11}{Federal Aviation Administration}
- **Pronoun:** for the given COREF ID we were able to match it in a sentence number 14 and 15
  <COREF ID="X7">George Auld</COREF>
  {17}{Auld}
  {14}{he}
  {15}{He}
- **String-Match:** for the given COREF ID we were able to match it in a sentence number 12
  <COREF ID="X12">investigation</COREF>
  {23}{investigation}
- **Synonyms:** for the given COREF ID we were able to match it in a sentence number 14
  <COREF ID="X3">the airplane</COREF>
  {14}{plane}
- **Indirect Mapping:** for the given COREF ID we were able to match it in a sentence number 11
  <COREF ID="X0">PANTEX</COREF>
  {5}{PANTEX}
  {11}{Pantex Weapons Plant}
  {14}{Pantex}
  {21}{Pantex}

Future Scope

- We hope to work on our indirect mapping, so that to find the sentences which does not contain the coreference explicitly for e.g.
  <COREF ID="X8">He and his wife</COREF>
  {25} {active parishioners}-Here we see that there is no mention of the coreference in the sentence but the matching happens with the help of context.