CS5560 Knowledge Discovery and Management

Problem Set (PS-3)

September 6th2018

Name:Lakshmi Korrapati

ClassID:4

1.Given the text example below, show the following NLP operations.

a. Triplet Extraction
b. Patterns of Triplets Based on Parts of Speech.

2. Summarize and draw a knowledge graph.

CHICAGO (AP) —Citing high fuel prices, United Airlines said Friday it has increased fares by $6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit AMR, immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL, said the increase took effect Thursday night and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Atlanta and Denver to San Francisco, Los Angeles and New York.

a. Triplet Extraction

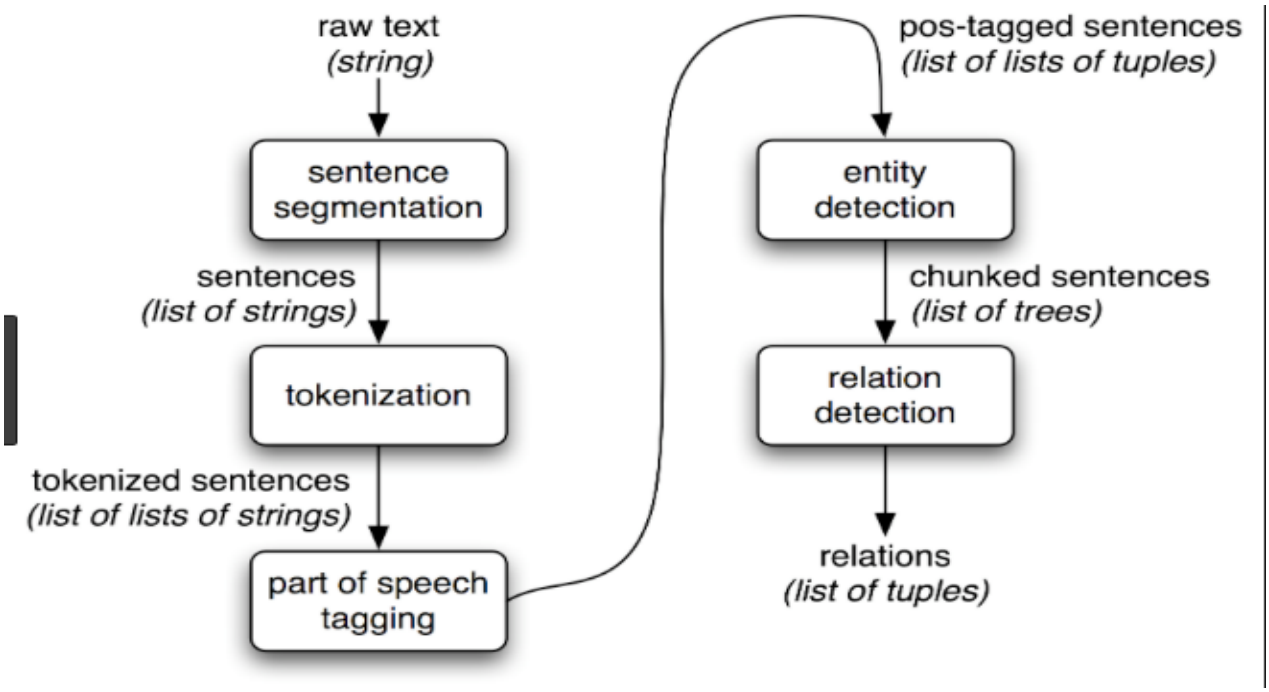
In this paper we present an approach to extracting subject-predicate-object triplets from English sentences. To begin with, four different well known syntactical parsers for English are used for generating parse trees from the sentences, followed by extraction of triplets from the parse trees using parser dependent techniques.

According to the approach presented in we define a triplet in a sentence as a relation between subject and object, the relation being the predicate. The aim here is to extract sets of the form {subject, predicate, object} out of syntactically parsed sentences, with four parsers, namely Stanford Parser, OpenNLP, Link Parser and Minipar. The work presented in used proprietary parser that already provided the logical form of a sentence including subject, object, predicate information, while our work is based on using publicly available parsers. Stanford Parser and OpenNLP both generate a treebank structure, and therefore will be discussed together. Furthermore, we present the results obtained using the Link Parser application, which is based on the link grammar. Finally, we describe the triplet extraction algorithm for the parse tree given by Minipar.

VB verb, base form
VBD verb, past tense
VBG verb, present participle or gerund
VBN verb, past participle
VBP verb, present tense, not 3rd person singular
VBZ verb, present tense, 3rd person singular
The type of adjective found
JJ adjective or numeral, ordinal
JJR adjective, comparative

a. Patterns of parts of speech

<verb-past-tense> <noun> <preposition> <adjective> <noun> <verb> <noun> <preposition> <noun>, <conjunction> <verb-past-tense> <pronoun> <preposition> <noun> <noun>.


raw text
(string)
sentence segmentation
sentences
(list of strings)
tokenization
tokenized sentences
(list of lists of strings)
part of speech tagging
pos-tagged sentences
(list of lists of tuples)
entity detection
chunked sentences
(list of trees)
relation detection
relations
(list of tuples)