**Notice: while manually editing the json txt, pay attention to the commas between each item in the dictionary and the identation of “{}” and “[]” (ident = 4 when the json was created by code).**

**Verb + prepostion(file: NLP/lib/OpenIE/verb\_prep\_json.txt)**

In the csv output of certain corpus, you may find that a verb’s object is not extracted, and there’s a preposition between the verb and the object.

For example, in the sentence “Then she put on her lynx jacket, almost stepped over me with a conventional excuse in pure Latin American Spanish, left without even saying good-bye or at least thanking me for all I had done to make our night together a happy one, and disappeared into the sun of today in the Amazon jungle of New York.” (Strange Pilgrims, Gabriel Garcia Marquez, 1992), “me” is the object of the verb “step”, and there’s a preposition “over” between them. However, in the output, there’s no “step” in the verb column.

Table

Description automatically generated

In this case, the verb-preposition collocation “step over” can be added in to the json file to help the algorithm extract certain verbs and objects:

1. Open the file, and find if the object (“over” in this case) in the keys of the json

A picture containing text

Description automatically generated

2. If the preposition is in the keys, add “step” in the list, which is the content to each key.

3. If it’s not found, added the key and content (an empty list). It would be better to added the key at the correct position according to alphabetical order.

Graphical user interface, application, Word

Description automatically generated

4. add the verb.

Text

Description automatically generated

**LVC: verb + pseudo object + real object (file: NLP/lib/OpenIE/verb\_obj\_obl\_json.txt)**

In this sentence “The provisions of this MOU will not give rise to a right of any person to impede the execution of a request.”, “give rise to” should be considered as a light verb construction that plays the role of a verb, while “right” is its object. However, in the output, that light verb construction is not recognized:



In this case, the light verb “give rise to” can be added in to the json file to help the algorithm extract certain verbs and objects:

1. Go to this website (<https://stanfordnlp.github.io/CoreNLP/demo.html>) and click the demo link, then past the sentence containing that light verb and run:

2. Check the dependencies:

Graphical user interface, diagram

Description automatically generated

3. Open the txt file, and find if the verb (“give” in this case) in the keys of the json. If it is, add an empty dictionary to its content, which is a list. If not, add the key and an empty list as its content.

Text, letter

Description automatically generated

4. Check the dependencies:

i. since the dep of “rise” is obj and “give” is its syntactical head, add “obj”: “rise” to the dictionary

A picture containing logo

Description automatically generated

ii. Find the dep of the object (“obj:to”), then split the deprel at “:”, use the first part as the key and second part as content, and add it to the dictionary (“obl:to”)

A picture containing chart

Description automatically generated

iii. In some cases, the verb is not the direct syntactical head of the object, such as in the LVC “take charge of”:

Graphical user interface, diagram

Description automatically generated

In this case, the syntactical head of the real object “stage” is the pseudo object “charge”. If the real object is governed by a token whose syntactical head is the verb, add “downwards” as a key, and that token’s dep as the content (here the dep of “charge” is “obj”)

Text

Description automatically generated

**LVC that start is a link verb (file: NLP/lib/OpenIE/link\_verb\_LVC\_json.txt)**

Some LVC starts with a link verb, such as “be responsible for”, “be in charge of”. To recognize this LVCs:

1. Run Stanford CoreNLP online demo to visualize the dependency:

Diagram

Description automatically generated

Diagram

Description automatically generated

2. Find the “ROOT” within this LVC, which is the token that is the syntactical head of all other tokens (except for the preposition) as well as the real object. For instance, in the LVC “in charge of”, “charge” is the syntactical head of “is”, “in”, “investigation”, but not the preposition “of”

3. Add root (“charge”, “responsible”) into the dictionary as a key, with a content of an empty list including an empty dictionary.

Chart

Description automatically generated

4. Add other tokens in the LVC (except for the “preposition”) in to the empty dictionary to the dictionary in the form of token’s lemma : dep:

Text

Description automatically generated

5. Add the key “prep” in the dictionary, with the content of an empty list, then added the dep of the real object into the empty list.

Text, application

Description automatically generated

6. If sometimes an LVC whose information is stored into this file is not recognized, paste the original sentence to the online demo, and check the dep of the object:

Diagram

Description automatically generated

For instance, in this sentence, the dep of the real object of the LVC “be responsible for” is “advcl:for” instead of “obl:for”, and the new dep “advcl:for” can be added to the content of the key “prep” in that LVC’s dictionary

Text, letter

Description automatically generated