RK-A - Bug #1520

Bug # 1514 (Resolved): Removing key phrases which are starting with IN pos tag or a preposition

Instead of removing entire key phrase starting with IN pos tag for all cases, we can process the keyphrase and discard just the word

08/20/2021 05:49 AM - Anonymous

Status:	Closed	**Start date:**	08/20/2021
Priority:	Normal	**Due date:**	
Assignee:	Rohit Choudhary	**% Done:**	100%
Category:		**Estimated time:**	3.00 hours
Target version:		**Spent time:**	0.00 hour

Description

This is an experiment that can be conducted to identify the right course of action for this task. While the KPs starting with IN pos tag can be really bad but there might be some cases where it may be useful. We will have to observe the patterns between the header variant and key phrase matched with it. Some cases that I was able to identify are:

KP -> header variants
Case 1:
1. for type hinting -> objects for type hinting
2. before the python initialization -> before python initialization
3. since unicode strings -> unicode strings

In case the continuous sequence of NN from the header variant is present in the KP, it may be useful to only drop the "IN" word.

Case 2:
1. for type -> for type hinting
2. in multiple -> in multiple directories
3. before python -> before python initialization

In case the continuous sequence of NN from the header variant is not present in the KP, it is more suitable to drop the KP.

For this sub-task,

Step 1: We can print all the KPs and their respective header variants satisfying the case where the KP starts with a word with an "IN" tag (or we can write the examples in a text file/log file). Some additional information may also be added such as the POS tags of both KP and header.

Step 2: Go over all the examples and decide some rules for elimination of the KP such that we can minimise the loss of good KPs and maximize the removal of bad looking KPs as well.

The decision is to be made by going over examples from multiple datasets (esp Std Lib)

History

#1 - 08/23/2021 11:52 AM - Anonymous

- Assignee set to Rohit Choudhary

#2 - 09/13/2021 07:12 AM - Anonymous

- % Done changed from 0 to 100

#3 - 10/20/2021 07:10 AM - Rohit Choudhary

- Status changed from New to Resolved

#4 - 10/20/2021 08:36 AM - Anonymous

- Status changed from Resolved to Closed