

RK-A - Feature #1615

Add the str_between of variation_middle_parenthesis to processed_full_header list

09/06/2021 01:19 PM - Nandini Bansal

Status:	Closed	Start date:	09/06/2021
Priority:	Normal	Due date:	
Assignee:	Rohit Choudhary	% Done:	0%
Category:		Estimated time:	4.00 hours
Target version:	P1	Spent time:	0.00 hour

Description

In the C-API book, I have seen cases where the string extracted by the variation_middle_parenthesis is being tagged with correct context but the assigned score is too less due to the length of the header variant.

For e.g.

tls -> thread local storage (tls) api -> 0.81

tss -> thread specific storage (tss) api -> 0.81

We need to make sure that these KPs are assigned a higher similarity score. This will be possible if the str_between is added to the processed_full_headers list and assigned a fullness_ratio of 1.0. The string should be a single word and not within the 12K unstemmed CW list (feel free to tweak this CW threshold to include/exclude desirable cases: we'll discuss the cases for which you wish to tweak the threshold).

Cases like (de)compression of files should be left untouched. Only for those cases where the parenthesized string is surrounded by spaces.

Find the str_between which shall be added to processed_full_headers list for Whirlwind, Tutorial, C-API & Library Reference. Based on the strings extracted from the header, test the final changes with C-API and Library Reference with ADR.

History

#1 - 10/13/2021 12:34 PM - Rohit Choudhary

- Status changed from New to Resolved

#2 - 10/20/2021 08:41 AM - Nandini Bansal

- Status changed from Resolved to Closed