RK-A - Feature #1544

Changes to match a token/word with hyphen with a header variant which does not have a hyphen but is exactly same

08/24/2021 09:18 AM - Nandini Bansal

Status:	Closed	**Start date:**	08/24/2021
Priority:	Normal	**Due date:**	
Assignee:		**% Done:**	100%
Category:		**Estimated time:**	4.00 hours
Target version:		**Spent time:**	0.00 hour

Description

This is a new feature we would like to add to our pipeline. We have observed a case where a token in the Whirlwind text file is "control-flow" and we have an exactly matching header "control flow" but they are not matched due to the presence/absence of hyphen.
We need to make changes to ensure that they are matched with high confidence.

The changes are to be made in the "get_candidates_for_variant" function inside the nested loops. We will have to add a condition that checks whether the current token contains a hyphen after a word i.e. follows the format "word1-word2" and "word1" is the same as the first word of header variant.

E.g.
the token is "word1-word2" and the header variant is "word3 word4 word5".
we will check if "word1 == word3", if it is, we will process the KP such that it becomes "word1 word2" and call "candidate_variant_distance_calculation" function for it.

Please note we will have to make changes in the candidate_variant_distance_calculation function such that for distance calculation, we will use "word1 word2" but while saving the candidates, we will use "word1-word2" format.

History

#1 - 08/26/2021 11:53 AM - Anonymous

- % Done changed from 0 to 100

#2 - 10/13/2021 12:33 PM - Anonymous

- Status changed from New to Resolved

#3 - 10/20/2021 08:37 AM - Nandini Bansal

- Status changed from Resolved to Closed

Files

Screenshot from 2021-08-24 14-28-57.png	51.6 KB	08/24/2021	Nandini Bansal