RK-A - Task #1663

Adding new header variants from pos_nouns function with fullness_ratio 1.0

09/17/2021 08:59 AM - Nandini Bansal

Status:	Resolved	**Start date:**	09/17/2021
Priority:	Normal	**Due date:**	
Assignee:		**% Done:**	0%
Category:		**Estimated time:**	3.00 hours
Target version:	P1	**Spent time:**	0.00 hour

Description

The main goal behind this task is to add some meaningful header variants with fullness_ratio 1.0 if a large portion of the header is made up of common words. The threshold of common words for this task is 1K (unstemmed, with excluded_word_list included).

This task will be done in a couple of parts:

1. Inside the pos_noun function, we call variation_add_nouns. Currently, that list of header variants is directly appended into the final list. We need to return the list of header variants returned by this function (variation_add_nouns) separately as well. Not at all places but only where pos_noun function is being all with extract_single_uncommon_words function.

2. From the final list of header variants generated by the pos_noun function, we need to check if the header variant was generated by variation_add_nouns. If yes, check if "sn_tn" is TRUE and the header variant is not made up entirely of words within 1K CW.

3. Calculate the ratio of the count of words in the original header within 1K and the total word count of the original header. If this ratio is >= 0.6, set the fullness_ratio of the header variant as 1.0.

NOTE: The code changes will not be restricted to only one function.

Test with C-API, Whirlwind, Tutorial & Library Reference.

History

#1 - 10/13/2021 12:24 PM - Anonymous

- Status changed from New to In Progress

#2 - 10/20/2021 03:08 PM - Anonymous

- Status changed from In Progress to Resolved