

RK-A - Task #1898

Skip the repeating KPs within 50 words according to new scheme - Contd

11/17/2021 08:43 AM - Nandini Bansal

Status:	In Progress	Start date:	11/17/2021
Priority:	Normal	Due date:	
Assignee:	Rohit Choudhary	% Done:	0%
Category:		Estimated time:	3.00 hours
Target version:	P1	Spent time:	0.00 hour
Description			
The repeating KPs within 50 words should be skipped as per the new scheme:			
If multiple instances of the same KP are tagged within the threshold (50 words) like k1, k2, k3 (same order) Case 5) if $sk1 = sk2 = sk3$, all k1, k2, k3 will be tagged were previously getting tagged because their lemmatized forms are the same. i.e. if the true/original form of KP k1, k2, k3 are the same, we shall not tag all three but k1 only			
A hypothetical example to explain this would be, say k1 "allocators", k2 "allocation", k3 "allocations" and $sk1 = sk2 = sk3$			
Then k1 & k2 will be tagged			
k1 & (k2, k3) are originally different but getting tagged due to the same lemmatized forms Note: $skn = \text{sim_score of } kn$			
Changes to be done in "check_repetition_multi_occurrence" function			

History

#1 - 11/17/2021 08:43 AM - Nandini Bansal

- Tracker changed from Bug to Task

#2 - 11/17/2021 10:16 AM - Nandini Bansal

- Description updated

#3 - 11/22/2021 08:46 AM - Nandini Bansal

- Assignee set to Rohit Choudhary

#4 - 12/22/2021 02:36 PM - Rohit Choudhary

- Status changed from New to In Progress