

RK-J - Task #3052

Sophisticated deduplication for getPurpleLinksInfoExtFull API results

05/30/2023 07:02 AM - Ram Kordale

Status:	Closed	Start date:	05/30/2023
Priority:	High	Due date:	
Assignee:	Venmuhilan B	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	1-RK-J-1	Spent time:	0.00 hour
Description			
Today, we do not send exact duplicates. Going forward, mark the following also duplicates			
- 'defining a function', 'define function' - 'threads', 'threading' - 'migrate', 'migration' - 'fetch the data', 'fetch data' - 'socket program', 'socket programming' - 'installing anaconda', 'install anaconda'			
Example URL that contains the above is https://www.youtube.com/watch?v=8O5kX73OkIY&list=PLsyeobzWxl7poL9JTVyndKe62ieoN-MZ3&index=54 .			
other than (a) adding articles like a, the in between words and checking, (b) it involves comparing stems. This and the (c) plurals work could use some java nlp library. you could pick the simplest from the list at https://xperti.io/blogs/java-natural-language-processing/ ?			

History

#1 - 05/30/2023 07:03 AM - Ram Kordale

- Priority changed from Urgent to High

#2 - 05/30/2023 07:12 AM - Venmuhilan B

- Status changed from New to In Progress

#3 - 06/14/2023 12:16 PM - Venmuhilan B

- Status changed from In Progress to Resolved

- % Done changed from 0 to 100

Documenting what was done for 3052 -

https://docs.google.com/document/d/1r_FSCpsZLa0NC6M9VC25OIHC5Oryo7YGUmReTsaOIeO/edit#heading=h.7quibv43tnmp

#4 - 06/14/2023 12:20 PM - Ram Kordale

- Status changed from Resolved to Closed