RK-A - Bug #1861

Finding the reason of low scores hyphenated KPs

11/09/2021 08:09 AM - Nandini Bansal

Status:	Resolved		**Start date:**	12/08/2021
Priority:	Normal		**Due date:**	
Assignee:			**% Done:**	0%
Category:			**Estimated time:**	3.00 hours
Target version:	P1		**Spent time:**	0.00 hour

Description

In MR https://gitlab.com/kordale/rk-ai-3/-/merge_requests/839 & https://gitlab.com/kordale/rk-ai-3/-/merge_requests/843, we worked on a problem of matching KPs with "-" (hyphen) with header variants that do not have hyphens. Just came across this issue in the "Numpy User Guide" book where a KP is being matched perfectly but the scores are below 0.5.

KP = "c-order"
Header Variant = "c order"

We need to investigate the reason for the low scores and fix it if possible.

Steps of investigation:
1) Identify where the score is reducing (BR3_IR3_tagger.py or tagging_utils.py)
2) Identify the reason for reduction (CW filter/POS tags/context matching)
3) Brainstorm the solution.

This ticket is primarily for steps (1) & (2).

History

#1 - 12/08/2021 09:42 AM - Nandini Bansal

- *Assignee set to Anonymous*

- *Start date changed from 11/09/2021 to 12/08/2021*

#2 - 12/08/2021 10:39 AM - Anonymous

- *Status changed from New to In Progress*

#3 - 12/10/2021 01:55 PM - Anonymous

- *Status changed from In Progress to Resolved*

Files

Screenshot from 2021-11-09 11-59-02.png	26.5 KB	11/09/2021	Nandini Bansal