RK-A - Bug #1752

In partial_header_match, reduce penalty for some KPs where start and end words are same as header variant

10/14/2021 01:20 PM - Nandini Bansal

Status:	Resolved		**Start date:**	10/14/2021
Priority:	Normal		**Due date:**	
Assignee:	Rohit Choudhary		**% Done:**	0%
Category:			**Estimated time:**	2.00 hours
Target version:	P1		**Spent time:**	0.00 hour

Description

In partial_header_match, we have a filter after the generation of candidates where we penalize the KPs because they start and end with the same words header variants. But it has been observed that some cases are being penalized unnecessarily.

We need to check the POS tags of the KPs with Spacy where word count of header variant < word count of KP and the uncommon word between header variant and KP is "ADJ" and rest of the words are "NOUN". Reduce the penalty of such cases to 0.05.

For e.g.
1) tkinter standard dialog |%| tkinter dialogs
[('tkinter', 'NOUN'), ('standard', 'ADJ'), ('dialog', 'NOUN')]
[('tkinter', 'NOUN'), ('dialogs', 'NOUN')]

2) ascii lower-case character |%| ascii characters
[('ascii', 'NOUN'), ('lower', 'ADJ'), ('-', 'PUNCT'), ('case', 'NOUN'), ('character', 'NOUN')]
[('ascii', 'NOUN'), ('characters', 'NOUN')]

3) ascii white-space characters |%| ascii characters
[('ascii', 'NOUN'), ('white', 'ADJ'), ('-', 'PUNCT'), ('space', 'NOUN'), ('characters', 'NOUN')]
[('ascii', 'NOUN'), ('characters', 'NOUN')]

History

#1 - 10/14/2021 01:22 PM - Nandini Bansal

- Description updated

#2 - 10/25/2021 09:54 AM - Nandini Bansal

- Assignee set to Rohit Choudhary

#3 - 10/25/2021 10:55 AM - Rohit Choudhary

- Status changed from New to In Progress

#4 - 11/02/2021 12:45 PM - Rohit Choudhary

- Status changed from In Progress to Resolved