

RK-A - Feature #2585

Penalty to be proportional to the number of nouns in the KP Sentence and the simdoc's section hierarchy match_sent_head_nouns

08/09/2022 10:12 AM - Ram Kordale

Status:	Resolved	Start date:	08/09/2022
Priority:	Normal	Due date:	
Assignee:	Rohit Choudhary	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:	P1	Spent time:	0.00 hour

Description

Today, it is roughly proportional to the number of levels in the section hierarchy that have nouns. In a little more detail, it looks like this:

```
if (numberOfNounsInKpSentence > 0 or NumberOfNounsInSimilarDocHeader > 0 )
then penalty = 0.015
```

else:

```
-matched_words = check_context_match(numberOfNounsInKpSentence, NumberOfNounsInSimilarDocHeader );
```

```
-if matched_words are under 5k:
```

```
--if words are under 2k then 0.03 else: 0.015 penalty
```

```
--elif matched_words are empty(means no context matched):
```

```
---if sim_doc_header/parent_header_list length is 3 and in comparing first level with second level hierarchy Noun words are all same
----then 0.03 penalty
```

```
---else:
```

```
----penalty by decrease_score_by_context_match();
```

```
--else:
```

```
---no context match penalty
```

where decrease_score_by_context_match() consist of

```
penalty = 0.03 + (len(parent_header_list) - 1) * 0.01
```

and also penalties for some specific few cases.

Going forward, we would like to change the penalty to be proportional to numberOfNounsInKpSentence and NumberOfNounsInSimilarDocHeader.

History

#1 - 08/17/2022 05:27 AM - Ram Kordale

- Assignee set to Rohit Choudhary

#2 - 08/23/2022 01:30 PM - Rohit Choudhary

- Status changed from New to In Progress

#3 - 09/16/2022 06:11 PM - Rohit Choudhary

- Status changed from In Progress to Resolved