

RK-A - Task #2371

Potential KPs ('immutable' within 'immutable sequence') that are part BRIR need to be given higher score

05/24/2022 06:02 AM - Ram Kordale

Status:	New	Start date:	05/24/2022
Priority:	Low	Due date:	
Assignee:	Rohit Choudhary	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:	P1	Spent time:	0.00 hour
Description			
<p>In Python 3 Doc Library Reference > Built-in Types > Text Sequence Type — str, 'immutable' should have score >0.9 since it is actually part of the string "immutable sequence". However, it looks like we are not including "sequence" because it is BRIR. We have a variant immutable sequence but sequence is BR tag that's why only immutable is tagged. Given that it's fullness ratio is 0.333, it also has a very low sim score.</p> <p>ideally, we should consider match with 'immutable sequence', give the corresponding high score (because of better fullness ratio) and then make 'immutable' as the KP.</p> <p>Another example is 'common' in 'common sequence operations' in Python 3 Doc Library Reference > Built-in Types > Text Sequence Type — str > String Methods.</p>			

History

#1 - 05/24/2022 06:02 AM - Ram Kordale

- Subject changed from Potential KPs that are part BRIR need to be given higher score to Potential KPs ('immutable' within 'immutable sequence') that are part BRIR need to be given higher score

#2 - 05/24/2022 08:14 AM - Ram Kordale

- Description updated

#3 - 08/16/2022 05:58 AM - Ram Kordale

- Assignee set to Rohit Choudhary

#4 - 08/16/2022 06:13 PM - Rohit Choudhary

Current Investigation:

For doc_id 37 we have words in first few lines in cumulative_noun_adj_sect1_keyword.csv

Preprocessed text : ". textual data in python is handled with %# objects, or strings . strings are immutable %# of unicode code points. string literals are written in a variety of ways: single quotes: "

original text : "text sequence type — str . textual data in python is handled with str objects, or strings . strings are immutable sequences of unicode code points. string literals are written in a variety of ways: single quotes:

The BR tags are converted into "%#" and Preprocessed text are used to search for candidate, doc_id 37 has only word "immutable" because current code do not track "%#" words original form.

#5 - 08/16/2022 06:15 PM - Rohit Choudhary

Rohit Choudhary wrote in [#note-4](#):

Current Investigation :

For doc_id 37 we have words in first few lines in cumulative_noun_adj_sect1_keyword.csv

Preprocessed text : ". textual data in python is handled with %# objects, or strings . strings are immutable %# of unicode code points. string literals are written in a variety of ways: single quotes: "

original text : "text sequence type — str . textual data in python is handled with str objects, or strings . strings are immutable sequences of unicode code points. string literals are written in a variety of ways: single quotes:

The BR tags are converted into "%#" and Preprocessed text are used to search for candidate, doc_id 37 has only word "immutable" because current code do not track "%#" words original form.

#6 - 08/17/2022 05:05 AM - Ram Kordale

- Priority changed from Normal to Low