

RK-A - Bug #1616

Some unwanted removals of KPs starting with VBG/IN

09/06/2021 02:24 PM - Nandini Bansal

Status:	Closed	Start date:	09/06/2021
Priority:	Normal	Due date:	
Assignee:	Rohit Choudhary	% Done:	0%
Category:		Estimated time:	5.00 hours
Target version:	P1	Spent time:	0.00 hour
Description			
<p>As per Issue #1514 & #1520, we discarded some KPs which were starting with the following POS tags "IN", "VBG", "VB". However, I just noticed that along with the bad ones like "preceding regular expressions", "unforgiving regular expressions", etc, we also losing some good ones like:</p> <ol style="list-style-type: none">1. defining a function -> defining and using functions2. defining a function -> defining functions3. importing standard library modules -> importing from python's standard library4. while true loop -> while loops <p>The examples listed above are from the Whirlwind book. I am sure there must be other examples in Whirlwind, C API & Tutorial books as well.</p> <p>Part A: Identify more such cases if there are any.</p> <p>Part B: Make changes to ensure the above cases and their likes are not skipped. We can leverage the info provided by POS tags for the same. A basic outline of the idea is to check if the first words of header variants and KPs are the same (in their stemmed forms) and apply some filters based on NOUNs.</p>			

History

#1 - 09/06/2021 02:24 PM - Nandini Bansal

- Estimated time set to 5.00 h

#2 - 10/13/2021 12:34 PM - Rohit Choudhary

- Status changed from New to Resolved

#3 - 10/20/2021 08:41 AM - Nandini Bansal

- Status changed from Resolved to Closed