

## RK-A - Bug #1755

### In partial\_header\_match, increase penalty for some KPs where start and end words are same as header variant

10/14/2021 01:41 PM - Nandini Bansal

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |          |                        |            |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|------------------------|------------|
| <b>Status:</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Resolved | <b>Start date:</b>     | 11/03/2021 |
| <b>Priority:</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | Normal   | <b>Due date:</b>       |            |
| <b>Assignee:</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |          | <b>% Done:</b>         | 0%         |
| <b>Category:</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |          | <b>Estimated time:</b> | 2.00 hours |
| <b>Target version:</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | P1       | <b>Spent time:</b>     | 0.00 hour  |
| <b>Description</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |          |                        |            |
| <p>There are cases where word count of KP &gt; word count of header variant and the uncommon word in the KP is VERB. The POS tag of the KP is to be found using Spacy. We need to increase the penalty of such cases to 0.5 as they are quite horrible looking KPs that shouldn't be tagged in the final annotated file.</p> <p>E.g.<br/>KP   Header Variant<br/>1) modules provide interfaces   module interface<br/>[['modules', 'NOUN'], ('provide', 'VERB'), ('interfaces', 'NOUN')]<br/>[['module', 'NOUN'], ('interface', 'NOUN')]</p> <p>2) module contains functions   module functions<br/>[['module', 'NOUN'], ('contains', 'VERB'), ('functions', 'NOUN')]<br/>[['module', 'NOUN'], ('functions', 'NOUN')]</p> <p>3) python validates bytecode   python bytecode<br/>[['python', 'PROPN'], ('validates', 'VERB'), ('bytecode', 'VERB')]<br/>[['python', 'PROPN'], ('bytecode', 'PROPN')]</p> <p>The KPs are from the Library Reference book. Recommended to create a dummy dataset.</p> |          |                        |            |

#### History

#1 - 10/14/2021 01:44 PM - Nandini Bansal

- Description updated

#2 - 11/02/2021 01:45 PM - Nandini Bansal

- Assignee set to Anonymous

#3 - 11/02/2021 01:45 PM - Nandini Bansal

- Start date changed from 10/14/2021 to 11/03/2021

#4 - 11/03/2021 07:09 AM - Anonymous

- Status changed from New to In Progress

#5 - 11/12/2021 01:35 PM - Anonymous

- Status changed from In Progress to Resolved