

RK-A - Task #1521

Changes to ensure that singular and plural forms of key phrases are also checked in 20K CW in `extract_single_uncommon_words` method

08/20/2021 08:03 AM - Nandini Bansal

| | | | |
|--|----------|------------------------|------------|
| Status: | Rejected | Start date: | 08/20/2021 |
| Priority: | Normal | Due date: | |
| Assignee: | | % Done: | 0% |
| Category: | | Estimated time: | 0.00 hour |
| Target version: | | Spent time: | 0.00 hour |
| Description | | | |
| <p>In <code>extract_single_uncommon_words</code> from <code>BR3_IR3_tagger.py</code>, all the single word keyphrases pass through the 20K common word filter (only those which are not present in 20K are allowed to be added as a header variant). Now, there are some cases where the word is present in the 20K common word list but its singular or plural form is not present.</p> <p>For example "datatype" is present in the 20K common word list but "datatypes" is not present in it. Using the <code>"singularize"</code> method from <code>common_tagging_functions.py</code> and the <code>"pluralize"</code> method from the <code>"pattern.text.en"</code> library, add an additional condition in <code>extract_single_uncommon_words</code> method.</p> <p>Import statement: from pattern.text.en import pluralize</p> <p>Test the changes with the following datasets:</p> <ol style="list-style-type: none">1. Python Whirlwind Tour.txt2. Python Tutorial.txt3. Python 3 - Library Reference.txt | | | |

History

#1 - 08/20/2021 08:04 AM - Nandini Bansal

- Status changed from New to Rejected