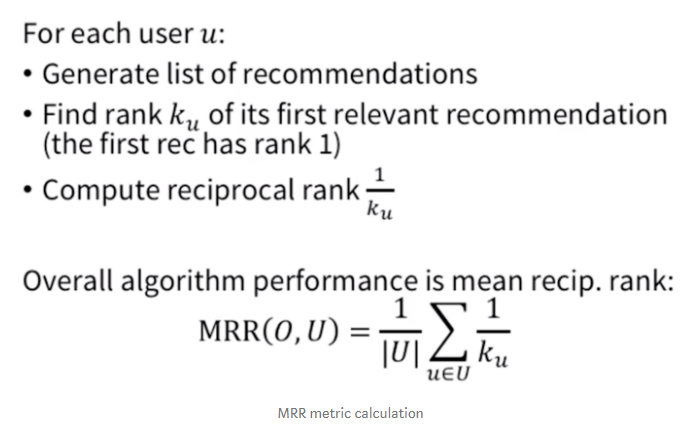
**Spell Checker**

**Theory:**

**MRR (Mean Reciprocal Rank):** The **mean reciprocal rank** is a statistic measure for evaluating any process that produces a list of possible responses to a sample of queries, ordered by probability of correctness.

**Formula for Calculations:**

****

**Code Description:**

**Non- Enhanced Model:** A simple N-Gram model with no changes.

**Enhanced Model:** Two constraints are applied to the Normal N-Gram Model.

* Length Constraint- Only those word are taken in the Query whose length varies from “length of input word” +2 to “length of input word”-2 .
* Bi-grams match constraint- The maximin bigrams match query and query less than that.

**Observations made:**

1. On addition of start and end symbol in the Non-enhanced model MRR increases by a factor of two.
2. Better results in the Enhanced Model.
3. More the MRR Value better the Results

**Additional Analysis:**

Slightly better results in case of removal of punctuations

**Stop Words**

**Observations made:**

1. NLTK class has a total of 179 stop words.
2. Machine was not able to plot bar graph for brown corpus as number of unique words, so in order to get the Threshold value I used another method mentioned above in the code.

**Results:**

**Result 1:** **Without removal of Special Characters.**

* Brown Corpus has 1161192 words and contains 56067 unique words.
* Brown Corpus has 155 stop words of the 179 stop words in English Language.
* 90 / 155 stop words found in the last 155 words of sorted brown corpus.
* Overlap Percentage = 57.692307692307686 %.

**Result 2: With removal of Special Characters.**

* Brown Corpus has 1040623 words and contains 43375 unique words
* Brown Corpus has 151 stop words of the 179 stop words in English Language
* 99 / 151 stop words found in the last 151 words of sorted brown corpus.
* Overlap Percentage = 65.13157894736842 %.

**Conclusion:**

When special characters were removed (especially ') new stop words like 'll','re','ve','t','d','s' came into existence so the number of stop words in brown corpus changes.