Kannada Extractive text summarization Result with Rouge metrics

Rouge 1 results

No of sentences in the input	Total no of words in the system summary	Total no of words in the manual summary	F1 Score	Recall	Precision
11	56	45	0.6363636313785125	0.67307692307692 31	0.603448275862069
20	74	81	0.66666666616669182	0.771428571	0.6619718309859155
30	111	105	0.6376811544204066	0.64077669902912 63	0.6346153846153846
40	156	153	0.6293706243706295	0.729370629	0.6293706293706294
53	301	162	0.688405792101515	0.69090909090909 09	0.6859205776173285
Average			0.65169757278 75963	0.703663206	0.6430653396 902655

Rouge 2 results

	Rouge 2 results				
No of sentences in the input	Total no of words in the system summary	Total no of words in the manual summary	F1 Score	Recall	Precision
11	56	45	0.46296295797839504	0.49019607843137 253	0.438596491228 07015
20	74	81	0.7179856065110502	0.52173913043478 26	0.514285714285 7142
30	111	105	0.4682926779269483	0.67058823529411 764	0.466019417475 72817
40	156	153	0.4507042203521127	0.45070422535211 27	0.450704225352 1127
50	301	162	0.6309090859091571	0.53284671532846 72	0.528985507246 3768
Average			0.5461709097355326	0.5332148769681 705	0.47971827111 760046

Rouge -I results

No of sentences in the input	Total no of words in the system summary	Total no of words in the manual summary	F1 Score	Recall	Precision
11	56	45	0.6213592183127534	0.65306122448979 59	0.592592592 5926
20	74	81	0.6666666616678146	0.67692307692307 7	0.656716417910 4478
30	111	105	0.6421052581584488	0.64893617021276 6	0.635416666666 6666
40	156	153	0.6245353109868577	0.636363636363 64	0.613138686131 3869
50	301	162	0.6910569055704278	0.70247933884297 52	0.68
Average			0.64914467093 92604	0.663552689 366 4 501	0.6355728 726602188

Average of three

Rouge parameters	F1 Score	Recall	Precision
Rouge-1	0.65169757278759 63	0.703663206	0.643065 33969026 55
Rouge-2	0.54617090973553 26	0.533214876968 1705	0.479718 27111760 046
Rouge-I	0.64914467093926 04	0.663552689366 4501	0.635572 87266021 88
Average	0.61567105115412 98	0.633476924111 5402	0.586118 82782269 49

Screen shot showing summarized words and manual words

(system generated vs Reference words with nltk)

total_words_system
['oemonds', 'ಜಿಯೋದಿಂದ', 'ಗ್ರಾಹಕರಿಗೆ', 'ಭರ್ಜರಿ', 'ಹೊಸ', 'ವರ್ಷದ', 'ಗಿಫ್ಟ್', 'ಇನ್\u200cಸ್ಟಾಗ್ರಾಮ್\u200cನಲ್ಲಿ', 'ಅಪಹಾಸ್ಯ',
'ವೀಡಿಯೊ', 'ಹಾಕ', 'ಜೈಲು', 'ಸೇರಿದ', 'ಯುವಕ', 'ಜಗತ್ತಿಗೆ', 'ಬರುತ್ತಿದ್ದಾನೆ', 'ಮತ್ತೊಬ್ಬ', 'ಹೊಸ', 'ದೇವರು..', 'ನಿಂತಲ್ಲೇ', 'ಸಿಗುತ್ತೆ',
'ವರ..', '!', 'ಏರ್\u200bಟೆಲ್\u200b', 'ದೀಪಾವಳಿ', 'ಧಮಾಕಾ', ':', '90', 'ದಿನಗಳವರೆಗೆ', 'ಅನಿಯಮಿತ', 'ಕರೆ', 'ಮತ್ತು', 'ಉಚಿತ',
'ಡೇಟಾ', 'ಪಡೆಯಿರಿ..', '!', 'ಶೀಘ್ರದಲ್ಲೇ', 'ಭಾರತದ', 'ಇಂಟರ್ನೆಟ್\u200bಗೆ', 'ಹೊಸ', 'ವೇಗ', 'ನೀಡಲಿದೆ', 'ಇಸ್ರೋ', 'ಸಂಸ್ಥೆ', 'ರೈಲು',
'ಪ್ರಯಾಣಿಕರ', 'ಗಮನಕ್ಕ', ':', 'ಇನ್ಮುಂದೆ', 'ಮಬೈಲ್\u200bಗಳ್ಲೇ', 'ಜನರಲ್', 'ಟಿಕೆಟ್', 'ಬುಕ್', 'ಮಾಡಿಕೊಳ್ಳಬಹುದು', 'Trending',
'ವಿಭಾಗವನ್ನೇ', 'ಕಿತ್ತೆಸೆದ', 'Facebook', 'ಈ', 'ಆರು', 'ಆ್ಯಪ್\u200bಗಳು', 'ಎಪ್ಸ್ಸ್ಬು200bಆ್ಯಪ್\u200b', 'ನು 'ಪೇಸ್\u200b', 'ಸಾರಾಗ'
ಬ್ರಿಂದ', 'ನಿಮ್ಮ', 'ಮಾಹಿತಿ', 'ಕದಿಯುತ್ತಿವೆ', ';', 'ಅವುಗಳನ್ನು', 'ಈಗಲೇ', 'ಡಿಲೀಟ್\u200b', 'ಮಡಿಡಿ', 'ವಗಿಗ', 'ವಂಚನೆಯಿಂದ', 'ಪಾರಾಗ'
ಬ್ರಿಂದ', 'ನಿಮ್ಮ', 'ಮೂಹಿತಿ', 'ಕಡಿಯುತ್ತಿವೆ', 'ಣುಚಿತ', 'ಡಬಿಟ್', 'ಹೂರ್ಡ್\u200b', 'ಬಿಡಿಯಾ', 'ನೀಡಲಿದೆ', '30', 'ಜಿಟಿರ್', 'ಇಂಟರ್\u200bನಟ್\u200bರನಟ್\u200bರನಟ್\u200bರನಟ್ಟಿಗೆ', 'ಅಪ್ಪಳಿಸಲಿದೆ', 'ಮರ್ಸಿಡಸ್, 'ಇಸ್ಕಿಸ್, 'ಇತಿಹಾಸ', '-', 'ಒಂದು', 'ಟ್ರಿಲಿಯನ್', 'ಡಾಲರ್', 'ಮುಟ್ಟಿಗ್, 'ಅವ್ಯುವಿತ',
'ಕರೆ', 'ಸ್ರೆಹ್ಮಿ, 'ಹೂಡಿಕೆ', 'ಮೂಡಿದರೂ', 'ಸಿಗುತ್ತೆ', 'ಅತ್ಯುತ್ತಮ', 'ಬಡ್ಡಿದರರ', 'ವಾಟ್ಸ್\u200bರ್ಟ್, 'ಮೊಡ್ರ್, 'ಮಾಹ್ನದಲ್ಲೇ', 'ಹಾರ್, 'ಆಫ್ರತ್ತಮ್, 'ಬಡ್ಡಿದರಿಗಾಗಿ', 'ದಕ್ಕಿಣ್, 'ಮಹ್ಯಾದಲ್ಲೇ', 'ಮಿಸ್ಕ್ಗೆ, 'ಅತ್ಯುತ್ತಮ', 'ಸುಂದ್, 'ಪೊಂಡಾ', '೧೯-೪', ':', 'ಸಂಪ್ರ್, 'ಗೂತ್ತಾ..', '?', 'ಗಾವಹಿಸಿದ್ದಿದರೆ', 'ಪೊಲೀಸ್, 'ಬಹ್ಯುಸ್ಗೆ', 'ಪಂಪಡೆಯುವುದು', 'ಈಗ', 'ಮತ್ತಮ್ಮ, 'ಸುಂಭ', 'ಸೈನ್', 'ದೇಶದ', 'ಪೂರ್ಷಕರು', 'ಮಾಹ್ಮ, 'ನೆರಮನೆಯವರ', 'ಬ
ಗ್ಗೆಯೂ', 'ಮಹಿತಿ', 'ನೀಡುತ್ತೆ', 'ಗೂಗಲ್', ', 'ಅಂಚ್', 'ಆಯ್ತು', 'ಈ, 'ಆ್ಯಪ್', '!', 'ಟ್ರಾಫಿಕ್', 'ಸಮಸ್ಕೆಗೆ', 'ಪರಿಂಹಾರ', ':', 'ಉ
ಬರ್(u200bನಿಂದ', 'ಹಾರುವ', 'ಸುರಿತ್,', 'ಸೇವೆ', 'ಮೋಟ್, 'ಪನ್\u200b', 'ಪ್ರೆಸ್', 'ಟ್ರಿಸ್', 'ಲೀಕ್\u200b', 'ಸಿ, 'ಸ್ರಿಪ್', 'ಪನ್', 'ಪರ್', 'ಪರ್ಸು', 'ಪ್ರಿಸ್', 'ಪರ್ಿ, 'ಪರ್ಿ, 'ಪರ್ಿ, 'ಪರ್ಿ, 'ಪರಿಸ್', 'ಪರ್ಿ, 'ಪ್ರಸ್, '!', 'ಪ್ರಸ್, 'ಟ್ರಿಸ್', 'ಪರ್ಿ, 'ಪರ್ಿ, 'ಪರ್ಿ, 'ಪರ್ಿ, total_words_system

In [27]: 1 print("total_words_reference\n",total_words_reference)

total_words_reference

total_words_reference
['ಕನ್ನಡದಲ್ಲೂ', 'ಇನ್ಮುಂದೆ', 'ಸುಲಭವಾಗಿ', 'ಚಾಟ್\u200b', 'ಮಾಡಬಹುದು', 'ಇನ್\u200cಸ್ಟಾಗ್ರಾಮ್\u200cನಲ್ಲಿ', 'ಅಪಹಾಸ್ಕ', 'ವೀಡಿಯೊ',
'ಹಾಕಿ', 'ಜೈಲು', 'ಸೇರಿದ', 'ಯುವಕ', 'ಜಗತ್ತಿಗೆ', 'ಬರುತ್ತಿದ್ದಾನೆ', 'ಮತ್ತುಬ್ಬ', 'ಹೂಸ', 'ದೇವರು..', 'ನಿಂತಲ್ಲೇ', 'ಸಿಗುತ್ತ', 'ವರ..', '!',
'ಏರ್\u200bಟೆಲ್\u200b', 'ದೀಪಾವಳಿ', 'ಧಮಾಕಾ', ':', '90', 'ದಿನಗಳವರೆಗೆ', 'ಅನಿಯಮಿತ', 'ಕರೆ', 'ಮತ್ತು', 'ಉಚಿತ', 'ಡೇಟಾ', 'ಪ
ಡೆಯಿರಿ..', '!', 'ಐಡಿಯಾ', 'ನೀಡಲಿದ', '30', 'ಜಿಬಿ', 'ಇಂಟರ್\u200bನಟ್\u200bಮಾರುಕಟ್ಟೆಗೆ', 'ಅಪ್ಪಳಿಸಲಿದೆ', 'ಮರ್ಸಿಡಸ್', 'ಎಸ್',
'ಶ್ರೀಣಿಯ', 'ಹೊಸ', 'ಕಾರ್', 'ಕೇವಲ', '9', 'ರೂಪಾಯಿಗಳಿಗೆ', 'ಅನಿಯಮಿತ', 'ಕರ', ',', 'SMS', 'ಕೂಡಾ', 'ಫ್ರೀ', 'ಆ್ಯಪಲ್', 'ಹೊಸ',
'ಇತಿಹಾಸ', '-', 'ಒಂದು', 'ಟ್ರಿಲಿಯನ್', 'ಡಾಲರ್', 'ಮಟ್ಟಿದ', 'ವಿಶ್ವದ', 'ಮೊದಲ', 'ಕಂಪನಿ', 'ಒಂದುಗೂಡಿವೆ', 'ಜಿಯೋ', 'ಮ್ಯೂಸಿಕ್',
'-', 'ಸಾವನ್', ':', 'ಕಲಾವಿದರಿಗಾಗಿ', 'ದಕ್ಷಿಣ', 'ಏಷ್ಯಾದಲ್ಲೇ', 'ಅತಿದೊಡ್ಡ', 'ವೇದಿಕೆ', 'ಸ್ಯೃಷ್ಟಿ', 'ಹೂಡಿಕೆ', 'ಮಾಡಿದರೂ', 'ಸಿಗುತ್ತೆ', 'ಅ
ತ್ಯುತ್ತಮ', 'ಬಡ್ಡಿದರ', 'ವಾಟ್ಸ್\u200bಆ್ಯಪ್\u200b', 'ಮೇಲೆ', 'ನಿಗಾವಹಿಸಲಿದೆ', 'ಪೊಲೀಸ್', 'ಇಲಾಖೆ..', '!', '5', 'ಸ್ಟಾರ್\u200b',
'ಗಿಟ್ಟಿಸಿಕೊಂಡ', 'ಹೊನ', 'ಮೊಬೈಲ್', 'ಕುರಿತು\u200b', 'ಮಾಹಿತಿ', 'ಸೈಬರ್\u200b\u200b', 'ಕ್ರೈಮ್\u200b\u200b', ':', '50', 'ಮಿಲಿಯ

output length: 20 input length: 53 Number of words in system 218 number of words in reference 152 Out[28]: [{'rouge-1': {'f': 0.707920787079208, 'p': 0.7079207920792079, 'r': 0.7079207920792079}, 'rouge-2': {'f': 0.5671641741044776, 'p': 0.5671641791044776, 'r': 0.5671641791044776}, 'rouge-l': {'f': 0.7115902914962839, 'p': 0.7058823529411765, 'r': 0.717391304347826}}]