

A Additional Experimental Results

2019 FINANCIAL OUTLOOK				
Lexical ROUGE	SMRS	TextRank	FWSR-BM25	FWSR-SIF
R-L F1	[11.77, 11.99]	[14.28, 14.57]	[13.79, 14.05]	[14.87, 15.11]
R-L Precision	[11.06, 11.28]	[13.05, 13.31]	[12.69, 12.92]	[14.63, 14.85]
R-L Recall	[18.69, 19.04]	[29.06, 29.51]	[33.58, 34.09]	[22.50, 22.97]
R-2 F1	[02.07, 02.23]	[03.65, 03.87]	[05.04, 05.29]	[02.96, 03.13]
R-2 Precision	[01.63, 01.75]	[02.62, 02.78]	[03.44, 03.62]	[02.37, 02.51]
R-2 Recall	[03.01, 03.23]	[06.65, 07.02]	[10.38, 10.92]	[04.88, 05.25]
R-1 F1	[14.47, 14.73]	[19.78, 20.10]	[19.75, 20.04]	[17.87, 18.13]
R-1 Precision	[11.91, 12.15]	[14.77, 15.05]	[13.95, 14.19]	[15.56, 15.79]
R-1 Recall	[20.01, 20.38]	[32.36, 32.81]	[36.63, 37.10]	[24.02, 24.50]

CLASSICAL LITERATURE				
Lexical ROUGE	SMRS	TextRank	FWSR-BM25	FWSR-SIF
R-L F1	[07.39, 07.56]	[16.15, 16.38]	[18.59, 18.81]	[13.10, 13.26]
R-L Precision	[16.96, 17.23]	[14.03, 14.22]	[15.82, 16.00]	[18.04, 18.27]
R-L Recall	[06.85, 07.00]	[19.02, 19.31]	[22.53, 22.81]	[12.95, 13.15]
R-2 F1	[00.37, 00.39]	[02.49, 02.58]	[03.16, 03.29]	[01.22, 01.27]
R-2 Precision	[00.68, 00.73]	[01.97, 02.04]	[02.87, 02.98]	[01.64, 01.71]
R-2 Recall	[00.27, 00.29]	[03.39, 03.51]	[03.51, 03.66]	[01.10, 01.15]
R-1 F1	[09.87, 10.07]	[19.49, 19.72]	[20.10, 20.30]	[16.48, 16.68]
R-1 Precision	[17.82, 18.10]	[18.43, 18.65]	[16.69, 16.88]	[20.33, 20.58]
R-1 Recall	[07.26, 07.42]	[22.10, 22.44]	[26.67, 26.97]	[14.83, 15.08]

Table 2: Lexical ROUGE performance for 2019 FINANCIAL OUTLOOK and CLASSICAL LITERATURE dataset. We reported the 95% confidence interval for the results. Results that are statistically better are bold faced and results that are statistically similar are colored as gray.

2019 FINANCIAL OUTLOOK				
Semantic ROUGE	SMRS	TextRank	FWSR-BM25	FWSR-SIF
R-L F1	[29.84, 30.18]	[26.02, 26.37]	[22.80, 23.13]	[34.36, 34.75]
R-L Precision	[32.70, 34.01]	[23.78, 24.09]	[20.97, 21.25]	[37.83, 38.97]
R-L Recall	[46.01, 46.33]	[52.45, 52.76]	[55.00, 55.33]	[48.66, 48.95]
R-2 F1	[55.45, 55.67]	[61.33, 61.53]	[61.72, 61.91]	[58.84, 59.00]
R-2 Precision	[52.59, 52.80]	[57.14, 57.33]	[56.27, 56.46]	[56.17, 56.34]
R-2 Recall	[59.16, 59.45]	[66.46, 66.74]	[68.73, 69.00]	[62.22, 62.49]
R-1 F1	[43.27, 43.52]	[48.16, 48.40]	[49.85, 50.09]	[45.54, 45.73]
R-1 Precision	[41.31, 41.55]	[44.80, 45.03]	[45.92, 46.14]	[43.23, 43.41]
R-1 Recall	[46.02, 46.33]	[52.45, 52.76]	[55.00, 55.33]	[48.66, 48.95]
CLASSICAL LITERATURE				
Semantic ROUGE	SMRS	TextRank	FWSR-BM25	FWSR-SIF
R-L F1	[31.83, 32.04]	[38.97, 39.33]	[38.89, 39.30]	[46.29, 46.91]
R-L Precision	[21.38, 21.94]	[37.69, 38.16]	[45.74, 46.49]	[48.37, 49.44]
R-L Recall	[33.25, 33.50]	[40.35, 40.58]	[42.19, 42.42]	[44.38, 44.62]
R-2 F1	[47.62, 47.84]	[52.87, 53.83]	[54.00, 54.20]	[60.41, 60.65]
R-2 Precision	[57.92, 58.19]	[56.11, 56.32]	[55.50, 55.71]	[66.39, 66.62]
R-2 Recall	[40.78, 41.01]	[49.98, 50.19]	[52.84, 53.07]	[55.42, 55.66]
R-1 F1	[38.62, 38.82]	[44.05, 44.24]	[42.33, 42.52]	[48.06, 48.29]
R-1 Precision	[50.06, 50.38]	[43.99, 44.17]	[42.72, 42.92]	[59.80, 60.04]
R-1 Recall	[31.83, 32.04]	[44.38, 44.62]	[42.19, 42.42]	[40.35, 40.58]

Table 3: Sematic ROUGE performance for 2019 FINANCIAL OUTLOOK and CLASSICAL LITERATURE dataset. We reported the 95% confidence interval for the results. Results that are statistically better are bold faced.