

RESULTS

Evaluations performed on the Portuguese newspaper dataset cited earlier produced the following tabulated results with the BLEU measure.

METHOD	BLEU SCORE
TF-IDF	0.433344438493
TEXTRANK	0.442393260308
SEMANTIC SIMILARITY METHOD	0.461430413696
SENTENCE EXTRACTION THROUGH CLUSTERING	0.558961140132

The following table tabulates the ROUGE unigram, bigram and trigram evaluation scores for the same dataset.

METHOD	ROUGE SCORE		
	Unigram	Bigram	Trigram
TF-IDF	0.641472390138	0.336487986654	0.235492595162
TEXTRANK	0.722071777132	0.429468582465	0.311777109304
SEMANTIC SIMILARITY METHOD	0.711440561912	0.408933331741	0.291565821891
SENTENCE EXTRACTION THROUGH CLUSTERING	0.74681767234	0.452319114024	0.328814256317

F1 score is measured as the harmonic mean of the BLEU and average ROUGE scores.

METHOD	F1 SCORE
TF-IDF	0.418417
TEXTRANK	0.463976
SEMANTIC SIMILARITY METHOD	0.465993
SENTENCE EXTRACTION THROUGH CLUSTERING	0.532986

From the above results, it is clear that sentence extraction based on k-means clustering algorithm performs the best with the given dataset in the evaluation metrics used. However, a more detailed discussion of these results is provided in the next section.