

Deliverables

Part 3

Stemming	Position Weighting	Local Collocation Modelling	$Tank_{acc}$	$Plant_{acc}$	PP_{acc}
False	Uniform	Bag-of-words	0.9025	0.9	0.78
True	Expndecay	Bag-of-words	0.795	0.91	0.7825
False	Expndecay	Bag-of-words	0.705	0.8725	0.7825
False	Expndecay	Adjacent-separate-LR	0.6675	0.865	0.77
False	Stepped	Bag-of-words	0.8075	0.9325	0.785
False	Linear-decay	Bag-of-words	0.8525	0.935	0.785

In Linear-decay, the adjacent word is weighted as 6, and the 2-away word is weighted as 5, and the 3-away word is weighted as 4, etc.

Part 4 - kNN extension

k	$Tank_{acc}$	$Plant_{acc}$	PP_{acc}
5	0.8475	0.745	0.6725
10	0.865	0.675	0.6125
20	0.8125	0.655	0.57