

	Feature	Description
Distance	Manhattan	Manhattan distance from two text fragments
	Euclidean	Euclidean distance from two text fragments
	Cosine similarity	Cosine similarity distance
	Matching coefficient	Matching coefficient of two text fragments
	Dice coefficient	Dice coefficient of two text fragments
	Jaccard	Jaccard distance of two text fragments
	Jaro	Jaro distance of two text fragments
	Damerau-Levenshtein	Damerau Levenshtein distance of two text fragments
	Levenshtein	Levenshtein distance of two text fragments
Statistical	Lcs	The longest common sub string of two text fragments
	Average of TF-IDF	Term frequency-inverse document frequency
	Avg-TF of Q and S	Avg-TF of words in a Q appearing in a S
	Avg-TF of S and Q	Avg-TF of words in a S appearing in a S
	Word overlapping	# word overlapping in a Q appearing in a article
	Average of Word2Vec	Average of word2vec similarity