

LIST OF TERMS

Terms	Definition
monolingual alignment	task of finding and aligning the similar semantic units in a pair of sentences expressed in natural language
phrase	combination of more than one word
token-based alignment	one-to-one alignment
phrase-based alignment	non one-to-one alignment
Paraphrase Database (PPDB)	a collection of ranked English and Spanish paraphrase contain over 220 million paraphrase pairs derived by : extracting lexical, phrasal and syntactic paraphrases from large bilingual parallel corpora, computing distributional similarity score for each of the paraphrases using the Geogle n-grams and the Annotated Gigaword Corpus[1]
tokenization	splitting strings into their component words based on delimiters
normalization	elimination of stylistic differences due to capitalization, punctuation, word order, and characters not in the Latin alphabet
Part-of-speech (POS) tag	a grammatical category for words, such as noun, verb, adjective, etc.
Gold Standard	the phrase segments are created manually by referring to the rules of Indonesian grammar
