

Daftar Istilah

<i>Pre processing</i>	Tahapan awal dalam mengolah data masukan.
<i>Token</i>	Kumpulan kata yang dipecah-pecah dari suatu kalimat.
<i>Tokenization</i>	Tahapan awal yang ditujukan memecah suatu kalimat menjadi kumpulan kata.
<i>Lemma</i>	Bentuk dasar kata yang merepresentasikan arti spesifik.
<i>Lemmatization</i>	Proses menormalisasi suatu kata menjadi bentuk dasar kata.
<i>Alignment</i>	Penjajaran kata dari dua pasang kalimat untuk mencari kesamaan di antara dua pasang kalimat.
<i>Segmentation</i>	Proses pembagian atau pemisahan kata dalam suatu kalimat.
<i>Corpus</i>	Kumpulan data yang terkumpul dalam satu sumber yang tersusun secara sistematis.