

Projek Matakuliah Pemrosesan Bahasa Alami

Projek matakuliah Pemrosesan Bahasa Alami adalah membuat sebuah mesin penerjemah dari Indonesia – Inggris dan parsing tree dari sebuah sentences untuk mengetahui posisi sebuah kata dalam kalimat berdasarkan rule dan kata dasarnya.

1. Mesin Penerjemah Indonesia – Inggris dan sebaliknya

- Buat aplikasi mesin penerjemah dari bahasa Indonesia – Inggris serta sebaliknya dengan menggunakan POS Tagger, lexicon, dan aturan dari suatu kata dasar.
- POS Tagger

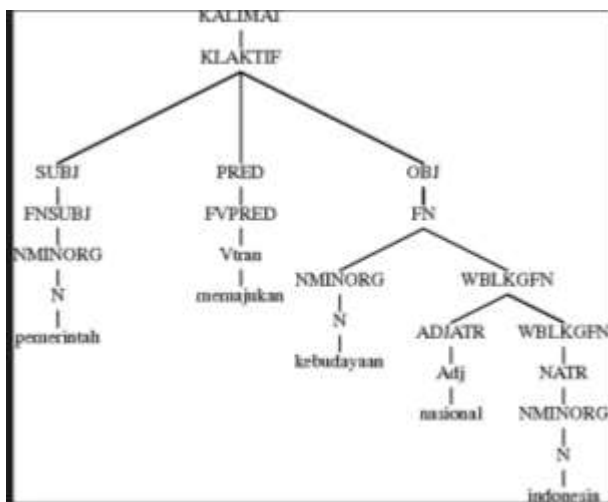
10.	KN	Noun	Meja, kerusi, mobil
11.	KK	Verb	Makan, tidur
12.	ADJ	Adjective	Hitam, cantik, dalam, besar
13.	KSN	Preposition	Di, ke, dari, dari pada
14.	KB	Auxiliary verb	Akan, belum, boleh
15.	KG	Pronoun	Aku, awak, beliau
16.	KH	Conjunction	Agar, bahawa, dan, atau
17.	ADV	Adverb	Bahawasanya, barangkali
18.	KT	Question	Apa, berapa, mengapa
19.	KDIL	Cardinal	Satu, beberapa, suku
20.	KPM	Narrator	Adalah, ialah
21.	KP	Command	Jangan, sila, tolong, harap
22.	KAR	Direction	Dalam, atas, bawah
23.	PW	Discourse mark	Ha, ha, maka
24.	KFP	Short form	UNCR, PBB
25.	KNF	Deny	Tidak, Bukan
26.	KNK	Proper noun	Abdullah Badawi
27.	KS	Interjection	Aduh, aduhai, ah
28.	DT	Determiners	Para, ini, itu
29.	RP	Particles	Adalah, ialah
30.	SEN	List Number	(i), (ii), (iii), etc
31.	SYM	Symbols	% & ' + " , } .

- Lexicon

```
indonesia_lexicon - Notepad
File Edit Format View Help
-- , -- ,
A NNP
a.l. NNP
a.n. NNP
a.s. NNPs
AA NNP
AAI NNP
AAMI NNP
abad NN
abadi VB JJ
ABAEI NNP
abah NN
abah-abah NN
abai VB
abaian NN
abar NN
abar-abar NN
ABC NNP
abdi VB NN
Abdul NNP
Abdullah NNP
```

- Aturan
Aturan dalam bahasa Indonesia bisa dilihat dari Buku Besar Bahasa Indonesia dengan memperhatikan posisi dalam suatu kalimat dan perubahan bila ada penambahan awalan atau akhiran.
- Basis data
Berisi kata dalam Indonesia – Inggris dan sebaliknya.

2. Parsing Tree Sentences (Kalimat)



NB : Untuk resources bisa mengakses NLP Stanford sebagai database untuk lexicon atau <http://bahasa.cs.ui.ac.id/postag/tagger>