
***Title* : Natural Language Processing (Almost)
from Scratch**

***Writer* : COLLOBERT, WESTON, BOTTOU,
KARLEN, KAVUKCUOGLU AND KUKSA**

***Citations* : 5359**

Mohammad Sabik Irbaz (160041004)

January 22, 2020

Types of NLP task :

Four standard NLP tasks are described : Part-Of-Speech tagging (POS), chunking (CHUNK), Named Entity Recognition (NER) and Semantic Role Labeling (SRL).

Part-Of-Speech Tagging :

POS aims at labeling each word with a unique tag that indicates its syntactic role, for example, plural noun, adverb, . . .

Chunking or Shallow Parsing :

Aims at labeling segments of a sentence with syntactic constituents such as noun or verb phrases (NP or VP). Each word is assigned only one unique tag, often encoded as a begin-chunk (e.g., B-NP) or inside-chunk tag (e.g., I-NP). Validation is achieved by splitting the training set.

Named Entity Recognition :

NER labels atomic elements in the sentence into categories such as 'PERSON' or 'LOCATION'. As in the chunking task, each word is assigned a tag prefixed by an indicator of the beginning or the inside of an entity.

Semantic Role Labeling

SRL aims at giving a semantic role to a syntactic constituent of a sentence. Feature categories commonly used by these system include (Gildea and Jurafsky, 2002; Pradhan et al., 2004):

- the parts of speech and syntactic labels of words and nodes in the tree;
- the node's position (left or right) in relation to the verb;
- the syntactic path to the verb in the parse tree;
- whether a node in the parse tree is part of a noun or verb phrase;
- the voice of the sentence: active or passive;
- the node's head word; and
- the verb sub-categorization.