

Terna Engineering College
Computer Engineering Department
Program: Sem VIII

Course: Natural Language Processing


Experiment No. 9

A.1 Aim: Perform and analyse chunking operations using the virtual lab.

PART B
(PART B: TO BE COMPLETED BY STUDENTS)

Roll No. 50	Name: AMEY THAKUR
Class: BE COMPS B	Batch: B3
Date of Experiment: 01-04-2022	Date of Submission: 01-04-2022
Grade:	

B.1 Virtual Lab (Input & Output):

Virtual Labs
An MoE Govt of India Initiative

HOME PARTNERS CONTACT

Computer Science and Engineering > Natural Language Processing > Experiments

Aim
Theory
Objective
Procedure
Simulation
Assignment
References
Feedback

Building Chunker

Chunking is an analysis of a sentence which identifies the constituents (noun groups, verbs, verb groups, etc.) which are correlated. These are non-overlapping regions of text. Usually, each chunk contains a head, with the possible addition of some function words and modifiers either before or after depending on languages. These are non-recursive in nature i.e. a chunk cannot contain another chunk of the same category.

Some of the groups possible are:

1. Noun Group
2. Verb Group

For example, the sentence 'He reckons the current account deficit will narrow to only 1.8 billion in September.' can be divided as follows:


INP He | IVP reckons | INP the current account deficit | IVP will narrow | IPP to | INP only 1.8 billion | IPP in | INP September |

Each chunk has an open boundary and close boundary that delimit the word groups as a minimal non-recursive unit.

Aim
Theory
Objective
Procedure
Simulation
Assignment
References
Feedback

Building Chunker

The objective of the experiment is to know the importance of selecting proper features for training a model and size of training corpus in learning how to do chunking.



Building Chunker

Select a language
English

First step is to train a corpus, select the size of a corpus
1k

Select the algorithm for training
CRF


Now, select feature for training
only lexicon

Training is completed

Now, we have to do testing, a corpus of size 2k is chosen

Check the accuracy
Check Accuracy

Accuracy is: 72.28



Building Chunker

Select a language
English

First step is to train a corpus, select the size of a corpus
1k

Select the algorithm for training
CRF


Now, select feature for training
only pos

Training is completed

Now, we have to do testing, a corpus of size 2k is chosen

Check the accuracy
Check Accuracy

Accuracy is: 86.2



Building Chunker

Select a language

English

First step is to train a corpus, select the size of a corpus

1k

Select the algorithm for training

CRF

Now, select feature for training

lexicon and pos

Training is completed

Now, we have to do testing, a corpus of size 2k is chosen

Check the accuracy

Check Accuracy

Accuracy is: 86.72



Building Chunker

Select a language

English

First step is to train a corpus, select the size of a corpus

1k

Select the algorithm for training

HMM

Now, select feature for training

only lexicon


Training is completed

Now, we have to do testing, a corpus of size 2k is chosen

Check the accuracy

Check Accuracy

Accuracy is: 73.73



Building Chunker

Select a language

English

First step is to train a corpus, select the size of a corpus

1k

Select the algorithm for training

HMM

Now, select feature for training

only pos


Training is completed

Now, we have to do testing, a corpus of size 2k is chosen

Check the accuracy

Check Accuracy

Accuracy is: 88.83



Building Chunker

Select a language

English

First step is to train a corpus, select the size of a corpus

1k

Select the algorithm for training

HMM

Now, select feature for training

lexicon and pos

Training is completed

Now, we have to do testing, a corpus of size 2k is chosen

Check the accuracy

Check Accuracy

Accuracy is: 85.95

B.2 Observations and learning:

- Chunking of text involves dividing a text into syntactically correlated words. Each chunk has an open boundary and close boundary that delimit the word groups as a minimal non-recursive unit. This can be formally expressed by using IOB prefixes.

B.3 Conclusion:

- Thus, we were able to learn about chunking and the type of chunking with a practical approach.

B.4 Question of Curiosity:

Q1: Following are some erroneous chunks. Identify the errors, mark them into three categories: invalid chunk boundaries, not a minimal phrase and incorrect chunk label.

Do the necessary corrections. Note: The categories may be overlapping.

1. [NP She] [VP is] [AJDP a beautiful] [NP girl]

ANS: NO ERROR.

2. [NP He] [VP could not] [ADJP be tolerated]

ANS: VP could not-invalid chunk boundaries.

3. [NP They] [VP invited] [NP him to] [NP the dinner party]

ANS: NP him to- invalid chunk boundaries.

NP the dinner party- not a minimal phrase.

4. [NP He] [VP was shot] [NP by the swats]

ANS: NP by the swats- invalid chunk boundaries.

5. [NP He] [VP longed] [PP for] [NP a hot cup of coffee]

ANS: NP a hot cup of coffee - incorrect chunk label.