

POS TAG BAHASA INDONESIA DENGAN TAGSET FLAIR



LAPORAN

**Disusun untuk Memenuhi Tugas Individu
Mata Kuliah Pemrosesan Bahasa Alami Semester VIII
yang diampu oleh Sukmawati Nur Endah, S.Si, M.Kom.**

DISUSUN OLEH:

LINGGAR MARETVA CENDANI (24060117120031)

**PROGRAM STUDI STRATA 1 INFORMATIKA
DEPARTEMEN ILMU KOMPUTER/INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG
2021**

DAFTAR ISI

DAFTAR ISI.....	2
BAB I.....	3
PENDAHULUAN.....	3
1.1 Permasalahan.....	3
1.2 Tujuan.....	3
BAB II.....	4
PEMBAHASAN.....	4
2.1 POS Tagging dengan Flair.....	4
2.2 Source Code Program.....	4
a. Mount Google Drive.....	4
b. Membaca File Berita.....	5
c. Install Library Flair.....	5
d. Import Semua Library Yang Akan Digunakan.....	5
e. Me-load Korpus, Menentukan Tipe Tag, Membuat Tag Dictionary, Membuat Word Embedding, Membuat Tagger, dan Melakukan Training.....	5
f. Me-load Tagger Yang Sudah Dibuat.....	6
g. Memberikan POS Tag Pada Berita 1.....	6
h. Memberikan POS Tag Pada Berita 2.....	6
2.3 Output Program	7
a. Hasil Training.....	7
a. Berita 1.....	8
a. Berita 2.....	9
2.4 Analisa Hasil	10
BAB III.....	13
PENUTUP.....	13
3.1 Kesimpulan.....	13

BAB I

PENDAHULUAN

1.1 Permasalahan

- 1.1.1. Apa itu Flair dan bagaimana cara POS Tagging dengan Flair?
- 1.1.2. Bagaimana source code program untuk melakukan POS Tagging dengan Flair?
- 1.1.3. Bagaimana output dari program dalam melakukan POS Tagging dengan Flair?
- 1.1.4. Bagaimana analisa hasil dari proses POS Tagging pada artikel berita dengan Flair?

1.2 Tujuan

- 1.1.1. Mengetahui apa itu Flair dan cara melakukan POS Tagging dengan Flair
- 1.1.2. Mengetahui source code program untuk melakukan POS Tagging dengan Flair
- 1.1.3. Mengetahui output program dalam melakukan POS Tagging dengan Flair
- 1.1.4. Mengetahui analisa hasil dari proses POS Tagging pada artikel berita dengan Flair

BAB II

PEMBAHASAN

2.1 POS Tagging dengan Flair

Word Clases merupakan kelas kata, biasa disebut juga *Part of Speech (POS)*. Proses pemberian label pada tiap kata kelas kata yang sesuai disebut **POS Tagging**. Pada program ini, digunakan sebuah library Python yang bernama **Flair** untuk melakukan POS Tagging.

Flair sendiri merupakan library bahasa pemrograman Python yang khusus menangani kasus - kasus di bidang *Natural Language Processing (NLP)*. Flair dapat digunakan untuk melakukan berbagai operasi NLP seperti *Word Embedding*, Tokenisasi, Klasifikasi Teks, dan tentu saja POS Tagging.

Flair memiliki koleksi korpus dari berbagai macam bahasa, termasuk Bahasa Indonesia. Korpus pada Flair sendiri terdiri dari 3 buah dataset korpus, yaitu *Development Corpus*, *Train Corpus*, dan *Test Corpus*. Maka dari itu, **Tagger** atau fungsi yang nantinya digunakan untuk melakukan POS Tagging dibuat dengan cara melakukan training pada korpus yang merupakan *tagset/ tag dictionary*.

Pada proses training, terdapat parameter yang dapat di-*tuning*, yaitu **learning_rate** dan **max_epoch**.

2.2 Source Code Program

a. Mount Google Drive

```
#mounting Google Drive
from google.colab import drive
drive.mount('/content/drive')
```

b. Membaca File Berita

```
#read file
txt_berita_1 = open('/content/drive/MyDrive/Upload/Akademik/Tugas
2 PBA/Cetak Assist, Asnawi Mangkualam Masuk Best 11 K League 2.txt
', 'r').read()

txt_berita_2 = open('/content/drive/MyDrive/Upload/Akademik/Tugas
2 PBA/Egy Maulana Vikri Sumbang Assist dan Gol untuk Lechia Gdansk
II Sebelum Menghadap Shin Tae-yong.txt', 'r').read()
```

c. Install Library Flair

```
!pip install flair
```

d. Import Semua Library Yang Akan Digunakan

```
from flair.data_fetcher import NLPTaskDataFetcher, NLPTask
from flair.embeddings import TokenEmbeddings, WordEmbeddings, Stac
kedEmbeddings, BertEmbeddings
from typing import List
```

e. Me-load Korpus, Menentukan Tipe Tag, Membuat Tag Dictionary, Membuat Word Embedding, Membuat Tagger, dan Melakukan Training

```
# 1. get the corpus
corpus = NLPTaskDataFetcher.load_corpus(NLPTask.UD_INDONESIAN)

# 2. what tag do we want to predict?
tag_type = 'upos'

# 3. make the tag dictionary from the corpus
tag_dictionary = corpus.make_tag_dictionary(tag_type=tag_type)

# 4. initialize embeddings
embedding_types: List[TokenEmbeddings] = [
    WordEmbeddings('id-crawl'),
    WordEmbeddings('id'),
    #WordEmbeddings('glove'),
    #BertEmbeddings('bert-base-multilingual-cased')
]
```

```

embeddings: StackedEmbeddings = StackedEmbeddings(embeddings=embedding_types)

# 5. initialize sequence tagger
from flair.models import SequenceTagger
tagger: SequenceTagger = SequenceTagger(hidden_size=256,
                                       embeddings=embeddings,
                                       tag_dictionary=tag_dictionary,
                                       tag_type=tag_type,
                                       use_crf=True)

from flair.trainers import ModelTrainer

trainer: ModelTrainer = ModelTrainer(tagger, corpus)

# 7. start training
trainer.train('resources/taggers/example-universal-pos',
             learning_rate=0.1,
             mini_batch_size=32,
             max_epochs=10)

```

f. Me-load Tagger Yang Sudah Dibuat

```

from flair.data import Sentence
tag_pos = SequenceTagger.load('resources/taggers/example-universal-pos/best-model.pt')

```

g. Memberikan POS Tag Pada Berita 1

```

berita_1 = Sentence(txt_berita_1)
tag_pos.predict(berita_1)
print(berita_1.to_tagged_string())

```

h. Memberikan POS Tag Pada Berita 2

```

berita_2 = Sentence(txt_berita_2)
tag_pos.predict(berita_2)
print(berita_2.to_tagged_string())

```

2.3 Output Program

a. Hasil Training

● Overall Score :

```
2021-05-03 22:07:27,397 -----
2021-05-03 22:07:27,398 EPOCH 10 done: loss 8.4470 - lr 0.1000000
2021-05-03 22:07:30,259 DEV : loss 5.773402690887451 - score 0.9156
2021-05-03 22:07:30,317 BAD EPOCHS (no improvement): 1
2021-05-03 22:07:51,276 -----
2021-05-03 22:07:51,463 Testing using best model ...
2021-05-03 22:07:51,471 loading file resources/taggers/example-universal-pos/best-model.pt
2021-05-03 22:08:34,792 0.9215
2021-05-03 22:08:34,799
Results:
- F-score (micro): 0.9215
- F-score (macro): 0.8511
- Accuracy (incl. no class): 0.9215
```

● Score By Class :

By class:

	precision	recall	f1-score	support
NOUN	0.9043	0.8728	0.8883	2555
PROPN	0.8824	0.9498	0.9148	2211
ADP	0.9496	0.9644	0.9570	1153
NUM	0.9198	0.9582	0.9386	407
VERB	0.9193	0.9480	0.9334	1250
CCONJ	0.9582	0.9173	0.9373	375
ADJ	0.8476	0.6282	0.7216	425
PRON	0.9811	0.9398	0.9600	498
PUNCT	0.9960	0.9994	0.9977	1724
ADV	0.8468	0.8347	0.8407	490
AUX	0.9897	1.0000	0.9948	96
DET	0.9584	0.9251	0.9415	374
SCONJ	0.7500	0.8239	0.7852	142
PART	0.9333	0.8571	0.8936	49
SYM	0.9630	0.8667	0.9123	30
X	1.0000	0.0000	0.0000	1
accuracy			0.9215	11780
macro avg	0.9250	0.8428	0.8511	11780
weighted avg	0.9215	0.9215	0.9205	11780

b. Berita 1

```
berita_1 = Sentence(txt_berita_1)
tag_pos.predict(berita_1)
print(berita_1.to_tagged_string())

Jakarta <PROPN> , <PUNCT> CNN <PROPN> Indonesia <PROPN> -- <PUNCT> Asnawi <PROPN>
```

```
#Cetak Assist, Asnawi Mangkualam Masuk Best 11 K League 2
Jakarta <PROPN> , <PUNCT> CNN <PROPN> Indonesia <PROPN> --
<PUNCT> Asnawi <PROPN> Mangkualam <PROPN> masuk <VERB> dalam
<ADP> daftar <NOUN> Best <PROPN> 11 <NUM> K <PROPN> League
<PROPN> 2 <NUM> pekan <NOUN> kedelapan <NUM> setelah <ADP>
membuat <VERB> assist <NOUN> kemenangan <NOUN> Ansan <PROPN>
Greeners <PROPN> atas <ADP> Daejeon <PROPN> Citizens <PROPN> ,
<PUNCT> Sabtu <PROPN> ( <PUNCT> 24 <NUM> / <SYM> 4 <NUM> )
<PUNCT> . <PUNCT> Mantan <NOUN> pemain <NOUN> PSM <PROPN>
Makassar <PROPN> itu <DET> dipilih <VERB> masuk <VERB> dalam
<ADP> tim <NOUN> terbaik <ADJ> oleh <ADP> Federasi <PROPN> Sepak
<PROPN> bola <NOUN> Profesional <PROPN> Korea <PROPN> pada <ADP>
27 <NUM> April <PROPN> . <PUNCT> Ini <DET> merupakan <VERB> kali
<NOUN> pertama <NUM> Asnawi <PROPN> masuk <VERB> dalam <ADP> tim
<NOUN> terbaik <ADJ> setelah <ADP> empat <NUM> kali <NOUN>
memperkuat <VERB> Ansan <PROPN> di <ADP> laga <NOUN> K <PROPN>
League <PROPN> 2 <NUM> . <PUNCT> Dalam <ADP> pertandingan <NOUN>
akhir <NOUN> pekan <NOUN> lalu <ADP> tersebut <DET> , <PUNCT>
Asnawi <PROPN> tampil <VERB> cukup <ADV> militan <ADJ> . <PUNCT>
Ketika <SCONJ> menjadi <VERB> penyerang <NOUN> sayap <NOUN>
kanan <ADJ> pada <ADP> babak <NOUN> pertama <NUM> , <PUNCT>
Asnawi <PROPN> beberapa <DET> kali <NOUN> mengirim <VERB> umpan
<NOUN> ke <ADP> kotak <NOUN> penalti <NOUN> . <PUNCT> Selain
<ADP> itu <DET> Asnawi <PROPN> juga <ADV> cukup <ADV> aktif
<VERB> membantu <VERB> pertahanan <NOUN> . <PUNCT> Sempat <ADV>
gagal <VERB> mencetak <VERB> gol <NOUN> ketika <SCONJ> mendapat
<VERB> peluang <NOUN> yang <PRON> sangat <ADV> terbuka <ADJ> di
<ADP> menit-menit <NOUN> akhir <NOUN> babak <NOUN> pertama
<NUM> , <PUNCT> Asnawi <PROPN> yang <PRON> menjadi <VERB> bek
<NOUN> kanan <ADJ> pada <ADP> babak <NOUN> kedua <NUM> kemudian
<ADV> membuat <VERB> assist <NOUN> pada <ADP> 10 <NUM> menit
<NOUN> akhir <NOUN> pertandingan <NOUN> . <PUNCT> Umpan <PROPN>
Asnawi <PROPN> yang <PRON> dituntaskan <VERB> Shim <PROPN> Jae
<PROPN> Min <PROPN> menjadi <VERB> gol <NOUN> kemenangan <NOUN>
Ansan <PROPN> atas <ADP> Daejeon <PROPN> . <PUNCT> Sebelumnya
<ADV> Ansan <PROPN> tidak <PART> pernah <ADV> menang <VERB> atas
<ADP> Daejeon <PROPN> sejak <ADP> pertengahan <NOUN> 2019
<NUM> . <PUNCT> Ansan <PROPN> sempat <ADV> menempati <VERB>
peringkat <NOUN> kedua <NUM> setelah <ADP> kemenangan <NOUN>
atas <ADP> Daejeon <PROPN> , <PUNCT> namun <SCONJ> kini <ADV>
klub <NOUN> berjudul <VERB> Green <PROPN> Wolf <PROPN> itu <DET>
merosot <VERB> ke <ADP> peringkat <NOUN> keempat <NUM> menyusul
<ADP> tiga <NUM> poin <NOUN> yang <PRON> didapat <VERB> Jeonnam
<PROPN> Dragons <PROPN> dan <CCONJ> FC <PROPN> Anyang <PROPN> .
<PUNCT> Pada <ADP> pekan <NOUN> mendatang <VERB> , <PUNCT> Ansan
<PROPN> akan <ADV> berhadapan <VERB> dengan <ADP> Seoul <PROPN>
E-Land <PROPN> yang <PRON> menempati <VERB> peringkat <NOUN>
```


kelima <NUM> . <PUNCT> Asnawi <PROPN> pun <ADV> berpeluang <VERB> kembali <ADV> dimainkan <VERB> , <PUNCT> namun <SCONJ> di <ADP> sisi <NOUN> lain <ADJ> pemain <NOUN> 21 <NUM> tahun <NOUN> itu <DET> juga <ADV> termasuk <VERB> dalam <ADP> pemain <NOUN> yang <PRON> dipanggil <VERB> ke <ADP> dalam <ADP> pemusatan <NOUN> latihan <NOUN> Timnas <PROPN> Indonesia <PROPN> jelang <PROPN> Kualifikasi <PROPN> Piala <PROPN> Dunia <PROPN> 2022 <NUM> . <PUNCT> Pelatnas <PROPN> akan <ADV> dimulai <VERB> pada <ADP> 1 <NUM> Mei <PROPN> hingga <ADP> 10 <NUM> Mei <PROPN> di <ADP> Jakarta <PROPN> , <PUNCT> setelah <ADP> itu <DET> pelatih <NOUN> Shin <PROPN> Tae <PROPN> Yong <PROPN> akan <ADV> membawa <VERB> anak <NOUN> asuhnya <NOUN> berlatih <VERB> di <ADP> Uni <PROPN> Emirat <PROPN> Arab <PROPN> sebelum <ADP> menghadapi <VERB> tiga <NUM> laga <NOUN> sisa <NOUN> Kualifikasi <PROPN> Piala <PROPN> Dunia <PROPN> pada <ADP> Juni <PROPN> . <PUNCT> Beberapa <DET> pekan <NOUN> lalu <CCONJ> Shin <PROPN> Tae <PROPN> Yong <PROPN> sudah <ADV> bertemu <VERB> dengan <ADP> perwakilan <NOUN> Ansan <PROPN> termasuk <VERB> membicarakan <VERB> perihal <ADP> pemanggilan <NOUN> Asnawi <PROPN> ke <ADP> Timnas <PROPN> Indonesia <PROPN> . <PUNCT>

c. Berita 2

```
berita_2 = Sentence(txt_berita_2)
tag_pos.predict(berita_2)
print(berita_2.to_tagged_string())
```

BOLASPORT.COM <PROPN> - <PUNCT> Pemain <NOUN> muda <ADJ> Indonesia <PROPN> , <PUNCT> Egy <PROPN>

#Egy Maulana Vikri Sumbang Assist dan Gol untuk Lechia Gdansk II Sebelum Menghadap Shin Tae-yong
 Jakarta <PROPN> , <PUNCT> CNN <PROPN> Indonesia <PROPN> -- <PUNCT> Asnawi <PROPN> Mangkualam <PROPN> masuk <VERB> dalam <ADP> daftar <NOUN> Best <PROPN> 11 <NUM> K <PROPN> League <PROPN> 2 <NUM> pekan <NOUN> kedelapan <NUM> setelah <ADP> membuat <VERB> assist <NOUN> kemenangan <NOUN> Ansan <PROPN> Greeners <PROPN> atas <ADP> Daejeon <PROPN> Citizens <PROPN> , <PUNCT> Sabtu <PROPN> (<PUNCT> 24 <NUM> / <SYM> 4 <NUM>) <PUNCT> . <PUNCT> Mantan <NOUN> pemain <NOUN> PSM <PROPN> Makassar <PROPN> itu <DET> dipilih <VERB> masuk <VERB> dalam <ADP> tim <NOUN> terbaik <ADJ> oleh <ADP> Federasi <PROPN> Sepak <PROPN> bola <NOUN> Profesional <PROPN> Korea <PROPN> pada <ADP> 27 <NUM> April <PROPN> . <PUNCT> Ini <DET> merupakan <VERB> kali <NOUN> pertama <NUM> Asnawi <PROPN> masuk <VERB> dalam <ADP> tim <NOUN> terbaik <ADJ> setelah <ADP> empat <NUM> kali <NOUN> memperkuat <VERB> Ansan <PROPN> di <ADP> laga <NOUN> K <PROPN> League <PROPN> 2 <NUM> . <PUNCT> Dalam <ADP> pertandingan <NOUN> akhir <NOUN> pekan <NOUN> lalu <ADP> tersebut <DET> , <PUNCT> Asnawi <PROPN> tampil <VERB> cukup <ADV> militan <ADJ> . <PUNCT> Ketika <SCONJ> menjadi <VERB> penyerang <NOUN> sayap <NOUN> kanan <ADJ> pada <ADP> babak <NOUN> pertama <NUM> , <PUNCT> Asnawi <PROPN> beberapa <DET> kali <NOUN> mengirim <VERB> umpan <NOUN> ke <ADP> kotak <NOUN> penalti <NOUN> . <PUNCT> Selain <ADP> itu <DET> Asnawi <PROPN> juga <ADV> cukup <ADV> aktif <VERB> membantu <VERB> pertahanan <NOUN> . <PUNCT> Sempat <ADV>

gagal <VERB> mencetak <VERB> gol <NOUN> ketika <SCONJ> mendapat <VERB> peluang <NOUN> yang <PRON> sangat <ADV> terbuka <ADJ> di <ADP> menit-menit <NOUN> akhir <NOUN> babak <NOUN> pertama <NUM> , <PUNCT> Asnawi <PROPN> yang <PRON> menjadi <VERB> bek <NOUN> kanan <ADJ> pada <ADP> babak <NOUN> kedua <NUM> kemudian <ADV> membuat <VERB> assist <NOUN> pada <ADP> 10 <NUM> menit <NOUN> akhir <NOUN> pertandingan <NOUN> . <PUNCT> Umpan <PROPN> Asnawi <PROPN> yang <PRON> dituntaskan <VERB> Shim <PROPN> Jae <PROPN> Min <PROPN> menjadi <VERB> gol <NOUN> kemenangan <NOUN> Ansan <PROPN> atas <ADP> Daejeon <PROPN> . <PUNCT> Sebelumnya <ADV> Ansan <PROPN> tidak <PART> pernah <ADV> menang <VERB> atas <ADP> Daejeon <PROPN> sejak <ADP> pertengahan <NOUN> 2019 <NUM> . <PUNCT> Ansan <PROPN> sempat <ADV> menempati <VERB> peringkat <NOUN> kedua <NUM> setelah <ADP> kemenangan <NOUN> atas <ADP> Daejeon <PROPN> , <PUNCT> namun <SCONJ> kini <ADV> klub <NOUN> berjuluk <VERB> Green <PROPN> Wolf <PROPN> itu <DET> merosot <VERB> ke <ADP> peringkat <NOUN> keempat <NUM> menyusul <ADP> tiga <NUM> poin <NOUN> yang <PRON> didapat <VERB> Jeonnam <PROPN> Dragons <PROPN> dan <CCONJ> FC <PROPN> Anyang <PROPN> . <PUNCT> Pada <ADP> pekan <NOUN> mendatang <VERB> , <PUNCT> Ansan <PROPN> akan <ADV> berhadapan <VERB> dengan <ADP> Seoul <PROPN> E-Land <PROPN> yang <PRON> menempati <VERB> peringkat <NOUN> kelima <NUM> . <PUNCT> Asnawi <PROPN> pun <ADV> berpeluang <VERB> kembali <ADV> dimainkan <VERB> , <PUNCT> namun <SCONJ> di <ADP> sisi <NOUN> lain <ADJ> pemain <NOUN> 21 <NUM> tahun <NOUN> itu <DET> juga <ADV> termasuk <VERB> dalam <ADP> pemain <NOUN> yang <PRON> dipanggil <VERB> ke <ADP> dalam <ADP> pemusatan <NOUN> latihan <NOUN> Timnas <PROPN> Indonesia <PROPN> jelang <PROPN> Kualifikasi <PROPN> Piala <PROPN> Dunia <PROPN> 2022 <NUM> . <PUNCT> Pelatnas <PROPN> akan <ADV> dimulai <VERB> pada <ADP> 1 <NUM> Mei <PROPN> hingga <ADP> 10 <NUM> Mei <PROPN> di <ADP> Jakarta <PROPN> , <PUNCT> setelah <ADP> itu <DET> pelatih <NOUN> Shin <PROPN> Tae <PROPN> Yong <PROPN> akan <ADV> membawa <VERB> anak <NOUN> asuhnya <NOUN> berlatih <VERB> di <ADP> Uni <PROPN> Emirat <PROPN> Arab <PROPN> sebelum <ADP> menghadapi <VERB> tiga <NUM> laga <NOUN> sisa <NOUN> Kualifikasi <PROPN> Piala <PROPN> Dunia <PROPN> pada <ADP> Juni <PROPN> . <PUNCT> Beberapa <DET> pekan <NOUN> lalu <CCONJ> Shin <PROPN> Tae <PROPN> Yong <PROPN> sudah <ADV> bertemu <VERB> dengan <ADP> perwakilan <NOUN> Ansan <PROPN> termasuk <VERB> membicarakan <VERB> perihal <ADP> pemanggilan <NOUN> Asnawi <PROPN> ke <ADP> Timnas <PROPN> Indonesia <PROPN> . <PUNCT>

2.4 Analisa Hasil

Pada program ini, digunakan 2 buah berita yang memiliki tema Sepakbola, dengan berita 1 yang berjudul "**Cetak Assist, Asnawi Mangkualam Masuk Best 11 K League 2**", dan berita 2 yang berjudul "**Egy Maulana Vikri Sumbang Assist dan Gol untuk Lechia Gdansk II Sebelum Menghadap Shin Tae-yong**". Kedua berita berformat txt, dan di load dari Google Drive.

Korpus yang digunakan adalah korpus Bahasa Indonesia, yang di-load dengan memanggil **NLPTask.UD_INDONESIAN**, yang dimana terdapat 3 set data korpus yang tadi telah dijelaskan (*development, train, dan test corpus*). Digunakan jenis tag-nya adalah '**upos**' karena akan dilakukan POS Tagging. Tag Dictionary dibuat dari korpus yang tadi telah di-load, lalu Word Embedding yang digunakan adalah '**id-crawl**' dan '**id**'. Pada proses training, digunakan parameter **learning_rate** dengan nilai **0,1** dan **max_epochs**-nya **10**.

Proses training yang dilakukan pada korpus dengan jumlah data train 4477, development 559, dan test 557 kalimat didapatkan hasil terbaik yaitu besar **loss** sebesar **5.773402690887451** dan **score**-nya mencapai **0.9156**. Sedangkan F-score dan Akurasinya adalah sebagai berikut :

- F-score (micro): 0.9215
- F-score (macro): 0.8511
- Accuracy (incl. no class): 0.9215

Karena digunakan POS Tagging Universal, maka kelas kata yang kita miliki adalah sebagai berikut :

- ADJ : kata sifat
- ADP : preposisi
- ADV : keterangan
- AUX : kata bantu
- CCONJ : kata penghubung
- SCONJ : subordinating conjunctions
- DET : determiner
- NOUN : kata benda
- NUM : angka
- PART : partikel
- PROPN : proper noun (nama diri)

- PRON : kata ganti
- PUNCT : tanda baca
- SYM : simbol
- VERB : kata kerja
- X : lainnya

Dari hasil POS Tagging yang dilakukan pada kedua berita, menunjukkan hasil yang cukup bagus dan akurat. Karena kelas kata yang digunakan adalah universal, maka tiap kata diberi label kelas sesuai kelas universal yang sesuai.

Namun, karena hanya melakukan operasi POS Tagging dan belum melakukan operasi Name Entity Recognition, beberapa entitas yang memiliki lebih dari satu kata dan merupakan kesatuan diberikan label kelas secara terpisah, meskipun labelnya sama, namun pemberian label dilakukan per kata, bukan per entitas.

Namun secara keseluruhan hasil dari pemberian label kelas kata pada tiap kata telah sesuai dan dapat dikatakan sangat baik .

BAB III

PENUTUP

3.1 Kesimpulan

POS Tagging merupakan proses pemberian label pada tiap kata kelas kata yang sesuai disebut. Flair yang merupakan library Python yang dapat digunakan untuk melakukan POS Tagging memberikan hasil yang memuaskan.

Untuk proses tag POS pada dokumen Bahasa Indonesia, didapatkan hasil akurasi mencapai 92,15 %. Dan hasil pemberian tag POS pada dokumen berita memberikan hasil yang baik dan tiap kata diberi label sesuai jenis kelas katanya yang pada program ini digunakan kelas kata universal.