

# PENGEMBANGAN PENDETEKSIAN KAMUS JENIS KATA SEBAGAI SUMBER DAYA NLP BAHASA INDONESIA

Arief Adiguna Putra<sup>1</sup>, Ken Kinanti Purnamasari<sup>2</sup>

<sup>1,2</sup> Teknik Informatika – Universitas Komputer Indonesia

Jalan Dipatiukur No 114-116 Bandung 40132

E-mail : ariefadiguna@gmail.com<sup>1</sup>, guinev\_gs@yahoo.co.nz<sup>2</sup>

## ABSTRAK

Kamus adalah buku yang memuat kata-kata beserta keterangan maknanya, pemakaiannya atau terjemahannya. Seiring dengan perkembangan zaman, kamus pun dapat dijumpai dalam bentuk digital yang dapat diakses secara online. Pada pengembangannya kamus jenis kata telah dibangun dengan data yang bersumber dari Kamus Besar Bahasa Indonesia (KBBI). Di dalam Kamus Besar Bahasa Indonesia (KBBI) memiliki jumlah kata dasar dan kata berimbuhan sebanyak 65.857 lema yang terdiri dari jumlah lema pokok (kata dasar) sebanyak 41.250 dan kata berimbuhan sebanyak 24.607 lema. Pendeteksian jenis kata sudah diteliti pada penelitian sebelumnya dalam Pembangunan Kamus Jenis Kata Sebagai Sumber Daya NLP Bahasa Indonesia telah mendapat hasil sebanyak 38.870 lema [1]. Masalah dalam penelitian sebelumnya yang mempengaruhi jumlah lema yang terdeteksi adalah teknik konversi data masukan yang digunakan dan langkah – langkah pengambilan kata dan jenis kata yang masih kurang efektif.

Dari hasil analisis yang telah dilakukan, maka terdiri dari proses merapikan data masukan yang bertujuan untuk merapikan data masukan dan proses pendeteksian kata dan jenis kata yang bertujuan untuk mendeteksi kata dan jenis kata sesuai dengan format yang telah ditentukan yang terdiri dari kata dan jenis kata

Hasil dari penelitian ini telah mendapatkan hasil lema sebanyak 43.060 lema dengan persentase sebesar 65,38% dari total jumlah kata dasar dan turunannya sebanyak 65.857 lema. Berdasarkan hasil analisis, peningkatan hasil pendeteksian terjadi karena adanya optimasi pada teknik konversi data masukan per abjad dan penyelesaian masalah – masalah pada tahap pengambilan kata dan jenis kata.

Kata kunci : Kamus jenis kata, kamus online, pengklasifikasian kelas kata, sumber data NLP, *Wordnet*.

## 1. PENDAHULUAN

Jenis kata adalah golongan kata dalam satuan bahasa berdasarkan kategori bentuk fungsi dan makna dalam susunan sistem gramatikal. Untuk menyusun kalimat yang baik dan benar dengan berdasarkan pola-pola kalimat baku, pemakai bahasa harus mengenal jenis dan fungsi kelas kata. Kamus jenis kata sudah diimplementasikan pada beberapa penelitian dibidang NLP (*Natural Language Processing*) atau Pemrosesan Bahasa Alami seperti pengklasifikasian jenis kata dalam POS Tag, *Stemming*, Pembangunan *WordNet* dan lain – lain.

Seiring dengan perkembangan zaman, kamus pun dapat dijumpai dalam bentuk digital yang dapat diakses secara online. Dalam pengaksesnya kamus online tidak disertai dengan sumber daya, dengan kata lain pengguna hanya memakai dan tidak mempunyai hak akses untuk mengubah atau menambahkan. Akan tetapi pada pengembangannya kamus jenis kata telah dibangun dengan data yang bersumber dari Kamus Besar Bahasa Indonesia (KBBI). Di dalam Kamus Besar Bahasa Indonesia (KBBI) memiliki jumlah kata dasar dan kata berimbuhan sebanyak 65.857 lema yang terdiri dari jumlah lema pokok (kata dasar) sebanyak 41.250 dan kata berimbuhan sebanyak 24.607 lema.

Pendeteksian jenis kata sudah diteliti pada penelitian sebelumnya dalam Pembangunan Kamus Jenis Kata Sebagai Sumber Daya NLP Bahasa Indonesia telah mendapat hasil sebanyak 38.870 lema [1]. Masalah dalam penelitian sebelumnya yang mempengaruhi jumlah lema yang terdeteksi adalah teknik konversi data masukan yang digunakan dan langkah – langkah pengambilan kata dan jenis kata yang masih kurang efektif.

Maka dalam penelitian ini akan dilakukan analisis pada sumber data masukan dan pada setiap langkah – langkah pada proses pendeteksian kata dan jenis kata, dengan itu akan diketahui kekurangan – kekurangan yang mempengaruhi jumlah lema yang didapat. Sehingga akan dilakukan proses optimasi pada pendeteksian kata dan jenis kata yang tidak terdeteksi serta akan mengubah teknik konversi data masukan yang awalnya berada pada satu file dipecah menjadi perabjad.

## 2. ISI PENELITIAN

Rumusan masalah menurut latar belakang di atas adalah belum terdeteksinya semua kata (lema) di Kamus Besar Bahasa Indonesia (KBBI) dalam penelitian sebelumnya.

### 2.2 Kamus

Secara etimologi, kata kamus berasal dari kata dalam bahasa Arab, yaitu qamus (bentuk jamaknya qawamus). Bahasa Arab menyerap kata kamus dari kata dalam bahasa Yunani kuno, okeanos yang berarti lautan.

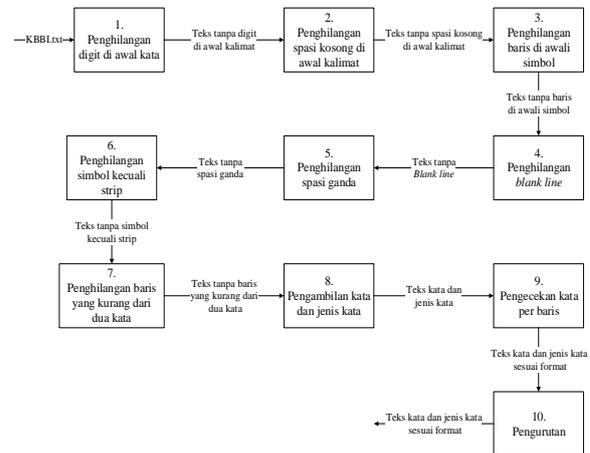
Klasifikasi kamus berdasarkan pengguna bahasa yang dapat ditulis dalam satu atau lebih dari satu bahasa terdiri dari beberapa jenis yaitu kamus eka bahasa, kamus dwi bahasa, dan kamus aneka bahasa. Pembangunan sumber daya kamus jenis kata yang akan dibangun termasuk ke dalam kamus eka bahasa, dimana kata-kata yang diolah hanya menggunakan satu bahasa. Kata adalah kumpulan beberapa huruf yang memiliki makna tertentu. Dalam kata adalah unsur bahasa yang diucapkan atau dituliskan yang merupakan perwujudan suatu perasaan dan pikiran yang dapat dipakai dalam berbahasa [2] Jenis kata adalah golongan kata dalam satuan bahasa berdasarkan kategori bentuk fungsi dan makna dalam susunan sistem gramatikal [2]. Klasifikasi kelas kata dalam pembangunan sumber daya jenis kata yang dipakai sebagai acuan adalah berdasarkan Tata Bahasa Baku Bahasa Indonesia yang memiliki persamaan dalam pengklasifikasian kelas kata seperti pada Kamus Besar Bahasa Indonesia yang menjadi sumber data masukan.

### 2.3 Analisis Masalah

Masalah yang diambil berdasarkan latar belakang yang dibahas sebelumnya adalah pendeteksian jenis kata sudah diteliti dalam Pembangunan Kamus Jenis Kata Sebagai Sumber Daya NLP Bahasa Indonesia telah mendapat hasil sebanyak 38.870 lema dengan dari total jumlah lema pada Kamus Besar Bahasa Indonesia (KBBI) yang memiliki jumlah kata dasar dan kata berimbuhan sebanyak 65.857 lema yang terdiri dari jumlah lema pokok (kata dasar) sebanyak 41.250 dan kata berimbuhan sebanyak 24.607 lema. [1]

Pada penelitian sebelumnya data masukan berupa Kamus Besar Bahasa Indonesia (KBBI) berada dalam satu file dengan format *plain text* (.txt) dan proses utama adalah tahap pengambilan kata dan jenis kata serta konversi format output. Tahapan proses pengambilan kata dan jenis kata terdiri dari sepuluh tahapan, yang terdiri dari proses penghapusan digit diagan baris yang kurang dari dua kata, pengambilan kata dan jenis kata, pengecekan kata per baris, dan pengurutan. Berdasarkan hasil analisis pada setiap prosesnya, penyebab banyaknya

lema (kata) yang hilang terjadi pada data masukan dan pada proses pengambilan kata dan jenis kata. [1] Pada tahapan pengambilan kata dan jenis kata pada penelitian sebelumnya, tidak adanya pengkategorian tahapan *pre-processing* dan proses utama. Berikut ini adalah tahapan pengambilan kata dan jenis kata pada penelitian sebelumnya.



**Gambar 1. Tahapan Pengambilan Kata dan Jenis Kata pada Penelitian Sebelumnya**

Dari tahapan pengambilan kata dan jenis kata berikut telah dapat mendeteksi lema sebanyak 38.870 lema dengan presentase sebanyak 59,01%. Hasil perbandingan tahapan pengambilan kata dapat dilihat pada tabel berikut.

**Tabel 1. Perbandingan Proses Per Tahap**

No	Tahapan Proses di Penelitian Sebelumnya	Hasil Proses
1	Penghapusan Digit di Awal Kata	-
2	Penghapusan Spasi Kosong di Awal Kalimat	-
3	Penghapusan Baris Yang di Awali Simbol	-
4	Penghapusan Baris Kosong	-
5	Penghapusan Spasi Ganda	-
6	Penghapusan Simbol Kecuali Strip	-
7	Penghapusan Baris Yang Kurang Dari 2 Kata	-
8	Pengambilan Kata dan Jenis Kata	36.988 Baris/Lema
9	Pengecekan Kata Per Baris	38.870 Baris/Lema
10	Pengurutan	38.870 Baris/Lema
<b>Hasil Pendeteksian</b>		38.870 Baris/Lema
<b>Persentase</b>		59,01%

## 2.4 Analisis Solusi

Analisis solusi merupakan suatu tahapan dimana prosesnya untuk mengidentifikasi setiap masalah dan kebutuhan yang diperlukan untuk membuat percobaan dengan cara lain. Solusi yang diberikan dalam penelitian ini yang berpengaruh dalam bertambahnya jumlah lema adalah proses konversi data akan dilakukan 3 (tiga) tahap dari format semula *.pdf* maka akan merubah format data menjadi *.doc* dan terakhir merubah ke *.txt*. Proses konversi akan dilakukan perabjad atau satu persatu dari abjad A sampai Z. Lalu akan dilakukan optimasi pada langkah pengambilan kata dan jenis kata sehingga masalah – masalah dapat diselesaikan.

Pada analisis solusi/ sistem yang akan dijalankan memiliki tahapan merapikan data masukan dan pendeteksian kata dasar dan turunannya. Penjelasan setiap tahap dapat dilihat pada gambar dibawah ini.



**Gambar 2. Gambaran Arsitektur Sistem**

Pada gambar 2 Gambaran Arsitektur Sistem akan dijelaskan sebagai berikut.

1. Pada tahap awal pengguna memasukan file yang merupakan Kamus Besar Bahasa Indonesia yang berformat *plain text (.txt)* ke dalam sistem.
2. Pada tahap merapikan data masukan akan dilakukan proses penghapusan nomor makna, penyamaan jarak pemisah antar huruf, penghapusan baris majemuk, penghapusan simbol tanda baca kecuali tanda hubung, penghapusan baris yang kurang dari 2 (dua) kata.
3. Pada tahap pendeteksian kata dasar dan turunannya akan dilakukan proses pengambilan kata dan jenis kata, pemisahan baris setiap lema, pengecekan duplikasi, pengurutan.
4. Hasil dari tahap pendeteksian kata dasar dan jenis kata berupa list kata dan jenis kata dengan format data *plain text (.txt)*.
5. Selanjutnya, list format kata dan jenis kata yang merupakan hasil dari pengambilan kata dan jenis kata disimpan ke dalam *database*. Dalam hal ini *database* yang digunakan adalah MySQL.

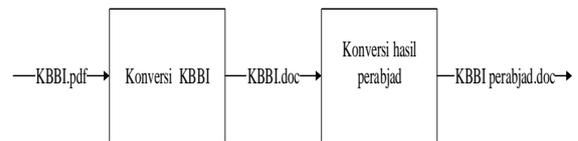
## 2.5 Analisis Data Masukan

Analisis data masukan pada sistem pendeteksian kamus jenis kata yaitu menjelaskan proses data masukan berupa data file teks yang diterima dan akan diproses oleh sistem.

## 2.6 Proses Konversi Data Masukan

Proses konversi format data awal yang berformat *.pdf* akan dilakukan dengan dua tahap, yaitu :

Proses perubahan sumber data awal yang berada dalam satu file menjadi per abjad dalam format *.doc* dan untuk proses konversi format dilakukan secara manual dengan menggunakan *software* Nitro Pro 10.



**Gambar 3. Proses Konversi Data .doc**

Berikut ini adalah contoh sebelum proses konversi yang dapat dilihat pada gambar berikut.

Name	Date modified	Type	Size
Kamus Indonesia	8/9/2016 8:29 AM	Adobe Acrobat D...	13,941 KB

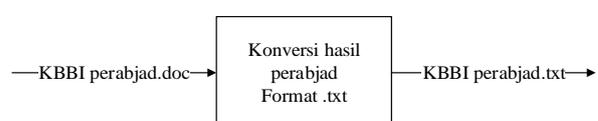
**Gambar 4. Data KBBBI Sebelum Konversi**

Berikut ini adalah contoh sesudah proses konversi yang dilakukan secara manual dengan menggunakan *software* Nitro Pro 10, hasil proses konversi berupa file KBBBI perabjad dari abjad A-Z dengan format *.doc* yang dapat dilihat pada gambar berikut.

Name	Date modified	Type	Size
A	8/31/2016 1:31 PM	Microsoft Word D...	686 KB
B	9/26/2016 1:30 PM	Microsoft Word D...	821 KB
C	9/26/2016 1:55 PM	Microsoft Word D...	313 KB
D	9/26/2016 1:57 PM	Microsoft Word D...	462 KB
E	9/26/2016 2:31 PM	Microsoft Word D...	198 KB
F	9/26/2016 1:59 PM	Microsoft Word D...	136 KB
G	9/26/2016 2:00 PM	Microsoft Word D...	523 KB
H	9/26/2016 2:01 PM	Microsoft Word D...	341 KB
I	9/26/2016 2:02 PM	Microsoft Word D...	233 KB
J	9/26/2016 2:03 PM	Microsoft Word D...	301 KB
K	9/26/2016 2:05 PM	Microsoft Word D...	1,252 KB
L	9/26/2016 2:07 PM	Microsoft Word D...	675 KB
M	10/22/2016 12:35 ...	Microsoft Word D...	622 KB
N	10/22/2016 12:43 ...	Microsoft Word D...	154 KB
O	9/26/2016 2:15 PM	Microsoft Word D...	108 KB
P	9/26/2016 2:17 PM	Microsoft Word D...	1,041 KB
Q	9/26/2016 2:17 PM	Microsoft Word D...	8 KB
R	9/26/2016 2:18 PM	Microsoft Word D...	648 KB
S	9/26/2016 2:21 PM	Microsoft Word D...	1,468 KB
T	9/26/2016 2:22 PM	Microsoft Word D...	1,143 KB
U	9/26/2016 2:22 PM	Microsoft Word D...	183 KB
V	9/26/2016 2:24 PM	Microsoft Word D...	63 KB
W	9/26/2016 2:25 PM	Microsoft Word D...	99 KB
X	9/26/2016 2:26 PM	Microsoft Word D...	13 KB
Y	9/26/2016 2:27 PM	Microsoft Word D...	14 KB
Z	9/26/2016 2:28 PM	Microsoft Word D...	42 KB

**Gambar 5. Data KBBBI Sesudah Konversi Ke .doc**

1. Selanjutnya hasil data kamus perabjad dengan format *.doc* maka akan diubah ke format *plain text (.txt)* dan untuk proses konversi format dilakukan secara manual dengan *software* Nitro Pro 10.



**Gambar 6. Proses Konversi Data .txt**

Berikut ini adalah contoh sebelum proses konversi yang dapat dilihat pada gambar berikut.

Name	Date modified	Type	Size
A	9/31/2016 1:31 PM	Microsoft Word D...	686 KB
B	9/26/2016 1:30 PM	Microsoft Word D...	822 KB
C	9/26/2016 1:55 PM	Microsoft Word D...	315 KB
D	9/26/2016 1:57 PM	Microsoft Word D...	462 KB
E	9/26/2016 2:31 PM	Microsoft Word D...	198 KB
F	9/26/2016 1:59 PM	Microsoft Word D...	136 KB
G	9/26/2016 2:00 PM	Microsoft Word D...	523 KB
H	9/26/2016 2:01 PM	Microsoft Word D...	343 KB
I	9/26/2016 2:02 PM	Microsoft Word D...	233 KB
J	9/26/2016 2:03 PM	Microsoft Word D...	301 KB
K	9/26/2016 2:05 PM	Microsoft Word D...	1,252 KB
L	9/26/2016 2:07 PM	Microsoft Word D...	675 KB
M	10/22/2016 12:35 ...	Microsoft Word D...	622 KB
N	10/22/2016 12:43 ...	Microsoft Word D...	154 KB
O	9/26/2016 2:15 PM	Microsoft Word D...	108 KB
P	9/26/2016 2:17 PM	Microsoft Word D...	1,041 KB
Q	9/26/2016 2:17 PM	Microsoft Word D...	8 KB
R	9/26/2016 2:18 PM	Microsoft Word D...	648 KB
S	9/26/2016 2:21 PM	Microsoft Word D...	1,469 KB
T	9/26/2016 2:22 PM	Microsoft Word D...	1,143 KB
U	9/26/2016 2:22 PM	Microsoft Word D...	183 KB
V	9/26/2016 2:24 PM	Microsoft Word D...	63 KB
W	9/26/2016 2:25 PM	Microsoft Word D...	99 KB
X	9/26/2016 2:26 PM	Microsoft Word D...	13 KB
Y	9/26/2016 2:27 PM	Microsoft Word D...	14 KB
Z	9/26/2016 2:28 PM	Microsoft Word D...	42 KB

**Gambar 7. Data KBBI Sebelum Konversi Ke .txt**

Berikut ini adalah contoh sesudah proses konversi menjadi KBBI perabjad dengan format *plain text* (.txt) yang dapat dilihat pada gambar berikut.

Name	Date modified	Type	Size
A	9/26/2016 12:58 PM	TXT File	314 KB
B	9/26/2016 2:30 PM	TXT File	379 KB
C	9/26/2016 2:30 PM	TXT File	140 KB
D	9/26/2016 2:31 PM	TXT File	216 KB
E	9/26/2016 2:32 PM	TXT File	81 KB
F	9/26/2016 2:33 PM	TXT File	64 KB
G	9/26/2016 2:33 PM	TXT File	232 KB
H	9/26/2016 2:34 PM	TXT File	157 KB
I	9/26/2016 2:34 PM	TXT File	109 KB
J	9/26/2016 2:35 PM	TXT File	132 KB
K	9/26/2016 2:35 PM	TXT File	370 KB
L	9/26/2016 2:36 PM	TXT File	304 KB
M	10/22/2016 12:45 ...	TXT File	295 KB
N	10/22/2016 12:44 ...	TXT File	40 KB
O	9/26/2016 2:42 PM	TXT File	38 KB
P	9/26/2016 2:43 PM	TXT File	412 KB
Q	9/26/2016 2:43 PM	TXT File	1 KB
R	9/26/2016 2:44 PM	TXT File	256 KB
S	9/26/2016 2:45 PM	TXT File	692 KB
T	9/26/2016 2:46 PM	TXT File	528 KB
U	9/26/2016 2:46 PM	TXT File	83 KB
V	9/26/2016 2:47 PM	TXT File	24 KB
W	9/26/2016 2:47 PM	TXT File	38 KB
X	9/26/2016 2:48 PM	TXT File	1 KB
Y	9/26/2016 2:49 PM	TXT File	3 KB
Z	9/26/2016 2:48 PM	TXT File	15 KB

**Gambar 8. Data KBBI Sesudah Konversi Ke .txt**

## 2.7 Data masukan

Data masukan dalam pendeteksian kamus jenis kata berupa data teks kata dan jenis kata perabjad yang didapatkan dari Kamus Besar Bahasa Indonesia dengan format *plain text* (.txt). Contoh data masukan yang akan diolah dapat dilihat pada tabel berikut.

**Tabel 2. Data Masukan**

1A, a n huruf pertama abjad  
 menghablur v menjadi hablur menghablurkan v  
 menjadikan hablur habuk n benda yg halus spt  
 tepung debu  
 a n are  
 Indonesia  
 2A n Ampere; lambang satuan ukuran arus  
 listrik

a n are

3a n are; nama satuan ukuran

ladang v, mengadang v 1 menghalangi  
 bisa

luas (= 100 m2)

a n are

ab ark n tabung atau kotak candu terbuat dr tanah

aba n ayah; bapak; (kadang-kadang juga berarti)  
 kakek

abadi a kekal selama-lamanya; tetap ada  
 sepanjang masa; tidak berkesudahan: ha- nya

Tuhan saja yg --; semua orang men-  
 dambakan kedamaian yg --;

aba-aba n kata-kata perintah atau ko-  
 mando, spt dl baris-berbaris, senam, atau

menyanyi bersama, msl siap!, kiri! kanan!,  
 satu! dua!, maju jalan!

abad n 1 masa seratus tahun: umurnya su-  
 dah setengah --; 2 jangka waktu yg la-  
 manya seratus

tahun: -- ke-20 dimulai dr tahun 1901 sampai  
 tahun 2000; 3 zaman (yg lamanya tidak tentu); 4

kl masa yg kekal, tidak berkesudahan;

-- kekal selama-lamanya; -- keemasan masa  
 kegemilangan dan kejayaan yg di- alami suatu

bangsa atau negara dl se- jarahnya; --  
 pertengahan zaman dl se- jarah Eropa antara

zaman purbakala dan zaman baru (tahun 476--  
 1492);

berabad-abad v beberapa abad lamanya; beratus-  
 ratus tahun: ~ lamanya bangsa kita dijajah;

seabad n satu abad; seratus tahun

2abuk n rambut: -- kaki; -- nya kusut masai  
 semua

3abuk v, mengabuk mengambil atau meng-  
 akui barang orang lain sbg miliknya

abuk-abuk n nama kue yg dibuat dr tepung aren;  
 kue abuk

abulhayat n bapak kehidupan; hujan

aba n ayah; bapak; (kadang-kadang juga berarti)  
 kakek

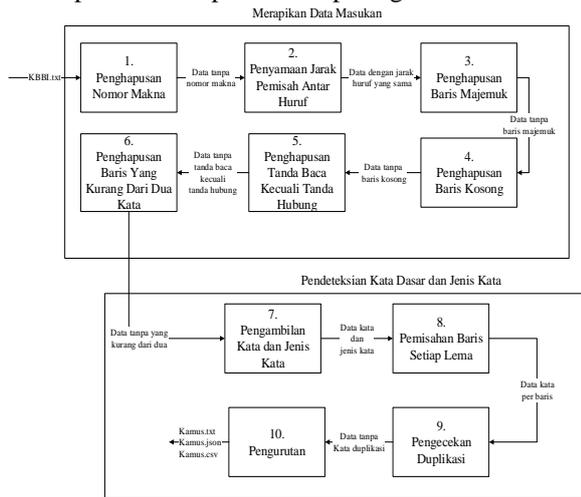
abun-abun ark n angan-angan;

## 2.8 Analisis Proses

Analisis proses terdiri dari proses merapikan data masukan dan proses pengambilan kata dan jenis kata. Tahap proses merapikan data masukan yang terdiri dari 6 (enam) tahap proses merapikan data masukan terdiri dari penghapusan nomor makna, penyamaan jarak pemisah antar huruf, penghapusan baris majemuk, penghapusan baris kosong,

penghapusan tanda baca kecuali tanda hubung, dan penghapusan baris yang kurang dari 2 (dua) kata. Dalam proses pengambilan kata dan jenis kata terdiri dari 4 (empat) proses yaitu pengambilan kata dan jenis kata, pemisahan baris setiap lema, pengecekan duplikasi, dan pengurutan.

Data masukan merupakan data yang bersumber dari Kamus Besar Bahasa Indonesia yang telah dilakukan proses konversi format sehingga menghasilkan data kamus perabjad dengan berformat *plain text* (.txt). Tahap tersebut dapat di lihat pada gambar berikut.



Gambar 9. Skema Proses

### 2.8.1 Proses Pendeteksian Kata Dasar dan Turunannya

Dalam tahap pengambilan kata dan jenis kata bertujuan untuk mendeteksi kata dan jenis kata sesuai dengan format yang terdiri dari 4 (empat) tahap yaitu pengambilan kata dan jenis kata, pemisahan baris setiap lema, pengecekan duplikasi, pengurutan. Tahapan pengambilan kata dan jenis kata dilakukan untuk mendapatkan struktur kalimat yang terdiri dari kata dan jenis kata (contoh : bisa n). Data masukan yang diolah secara terpisah perabjad dari abjad A sampai Z berasal dari Kamus Besar Bahasa Indonesia berformat *Plain Text* (.txt). Berikut ini adalah tahap proses pengambilan kata dan jenis kata.

#### 1. Pengambilan Kata dan Jenis Kata

Tahapan ini adalah pengambilan kata berdasarkan struktur kata yang ditentukan. Tujuan dari tahap ini adalah mengambil kata dan jenis sesuai dengan format yang telah ditentukan, dimulai dengan mencari kata dan jenis kata pada setiap barisnya, contoh: bisa n. Pada tahap pengambilan kata dan jenis kata penelitian sebelumnya, proses pengambilan kata dan jenis kata hanya bisa mengambil kata pada awal baris saja dan tidak bisa mengambil kata yang tidak diikuti jenis kata. Oleh karena itu, pada penelitian ini akan dilakukan optimasi agar masalah – masalah tersebut dapat di atasi. Sehingga akan meningkatkan hasil

pendeteksiannya. Untuk lebih jelasnya digambarkan dalam bentuk blok diagram dibawah ini.



Gambar 10. Blok Diagram Pengambilan Kata dan Jenis Kata

Tahapan pertama dimulai dengan membaca data hasil tahapan sebelumnya. Setelah itu melakukan penggantian label kata dengan spasi karena label kata tidak digunakan pada pendeteksian ini. Label kata yang dimaksud adalah ark, cak, ho, kas, dan kl. Berikut ini adalah contoh penggantian label kata dengan spasi yang dapat dilihat pada gambar berikut.

Tabel 3. Penggantian Label Kata

Data Sebelum Penggantian Label
ab ark n tabung atau kotak candu terbuat dr tanah
Data Sesudah Penggantian Label
ab n tabung atau kocak candu terbuat dr tanah

Proses selanjutnya yaitu pendeteksian kata dan jenis kata di setiap baris. Proses pendeteksian berdasarkan aturan atau pola yang telah ditentukan yaitu mencari di dalam setiap baris yang mengandung huruf a-z dengan minimal satu huruf lalu diikuti spasi dan diikuti simbol jenis kata yang telah ditentukan yaitu n,v,a,adv,num,p, atau pron. Jika ada yang sesuai dengan aturan tersebut maka akan dianggap sebagai kata dan jenis kata. Proses akan dilakukan secara berulang dan akan berhenti setelah tidak ditemukan kata yang tidak diikuti jenis kata pada suatu barisnya. Lalu disimpan ke dalam bentuk file .txt.

Tabel 4. Pendeteksian Kata dan Jenis Kata

Data Sebelum Proses
A a n huruf pertama abjad menghablur v menjadi hablur menghablurkan v menjadikan hablur habuk n benda yg halus spt tepung debu a n are
A n Ampere lambang satuan ukuran arus listrik a n are a n are nama satuan ukuran adang v mengadang v 1 menghalangi
luas 100 m2 a n are ab ark n tabung atau kotak candu terbuat dr tanah aba n ayah bapak kadang-kadang juga berarti kakek abadi a kekal selama-lamanya tetap ada sepanjang masa tidak berkesudahan ha- nya Tuhan saja yg -- semua orang men- dambakan kedamaian yg --

aba-aba n kata-kata perintah atau ko-  
 mando spt dl baris-berbaris senam atau menyanyi  
 bersama msl siap kiri kanan satu dua maju jalan  
 abad n 1 masa seratus tahun umurnya su-  
 dah setengah -- 2 jangka waktu yg la-  
 manya seratus tahun -- ke-20 dimulai dr tahun 1901 sampai tahun  
 2000 3 zaman yg lamanya tidak tentu 4 kl masa yg  
 kekal tidak berkesudahan  
 berabad-abad v beberapa abad lamanya beratus-  
 ratus tahun lamanya bangsa kita dijajah  
 seabad n satu abad seratus tahun  
 abuk n rambut -- kaki -- nya kusut masai

abuk v mengabuk mengambil atau meng-  
 akui barang orang lain sbg miliknya  
 abuk-abuk n nama kue yg dibuat dr tepung aren  
 kue abuk  
 abulhayat n bapak kehidupan hujan  
 aba n ayah bapak kadang-kadang juga berarti  
 kakek  
 abun-abun ark n angan-angan

**Data Sesudah Proses**

**A a n**  
 menghablur v  
 menghablurkan v  
 habuk n  
 a n  
 A n  
 a n  
 a n  
 adang v mengadang v  
 a n  
 ab n  
 aba n  
 abadi a  
 aba-aba n  
 abad n  
 berabad-abad v  
 seabad n  
 abuk n  
 abuk v  
 mengabuk v  
 abuk-abuk n  
 abulhayat n  
 aba n  
 abun-abun n

**2.8.2 Konversi Format Ouput**

Hasil dari beberapa proses pengambilan kata dan jenis kata dengan format *plain text (.txt)* akan di simpan ke dalam database yang menghasilkan format *.sql*, maka selanjutnya akan dilakukan konversi menjadi beberapa bentuk format seperti; *.txt*, *.json*, dan *.csv*.



**Gambar 11. Blok Diagram Konversi Format Output**

**2.9 Pengujian Sistem**

Tahap pengujian sistem bertujuan untuk menemukan kesalahan – kesalahan atau kekurangan – kekurangan pada sistem yang diuji. Pengujian bermaksud untuk mengetahui apakah program kamus kata dan jenis kata yang dibuat sesuai dengan tujuan penelitian.

**2.9.1 Pengujian Hasil Pendeteksian**

Pada pengujian akurasi akan dihitung nilai akurasinya dari jumlah lema yang didapat saat proses pengambilan kata dan jenis kata. Jumlah lema yang didapat sebanyak 43.060 lema dari total jumlah lema pada Kamus Besar Bahasa Indonesia (KBBI) sebanyak 90.049 lema dengan jumlah lema pokok (kata dasar) sebanyak 41.250 lema, jumlah lema berimbuhan (kata turunan) sebanyak 24.607, dan gabungan kata (kata majemuk) sebanyak 23.536 lema. Maka hasil perhitungannya sebagai berikut.

Jumlah total lema pokok dan berimbuhan : 65.857  
 Total lema yang didapat : 43060

Hasil Pengujian Pendeteksian :  

$$= ( \text{Jumlah Lema Yang Didapat} / \text{Jumlah Lema KBBI} ) * 100\%$$

$$= ( 43.060 / 65.857 ) * 100\%$$

$$= 65,38\%$$

Maka pada penelitian ini menghasilkan nilai akurasi sebesar 65,38%, dengan menghasilkan peningkatan sebanyak 4190 lema (kata) dengan persentase sebesar 6,36%.

Berikut ini adalah hasil (output) dari setiap tahapan proses pendeteksian kata dan jenis kata.

**Tabel 5. Hasil Pendeteksian Kata dan Jenis Kata**

Tahapan Penelitian Ini	Proses di	Hasil Proses
Penghapusan Nomor Makna		-
Penyamaan Jarak Pemisah Antar Huruf		-
Penghapusan Majemuk	Baris	-
Penghapusan Baris Kosong		-
Penghapusan Tanda Baca Kecuali Tanda Hubung		-
Penghapusan Baris Yang Kurang Dari Dua Kata		-
Pengambilan Kata dan Jenis Kata		40.878 Baris/Lema
Pemisahan Lema	Baris Setiap	46.184 Baris/Lema
Pengecekan Duplikasi		45.316

	Baris/Lema
Pengurutan	45.316 Baris/Lema
<b>Hasil Pendeteksian</b>	43.060 Baris/Lema
<b>Persentase</b>	65,38%

### 2.9.2 Analisis Hasil Pengujian

Berikut ini adalah analisis dari hasil pengujian hasil pendeteksian yang sudah dilakukan. Berdasarkan hasil perhitungan kata dasar dan turunannya yang dilakukan secara manual pada data masukan abjad A yang bersumber dari Kamus Besar Bahasa Indonesia (KBBI). Maka diperoleh jumlah kata dasar sebanyak 1774 lema dan turunannya sebanyak 853 lema. Sehingga total lema yang diperoleh sebanyak 2627 lema.

Hasil perbandingan antara perhitungan manual pada KBBI dengan hasil proses pendeteksian pada program dapat dilihat pada tabel berikut.

Perhitungan Manual	Pendeteksian Pada Program
2627	2544

Berdasarkan hasil perbandingan maka ada 83 lema yang belum dapat terdeteksi.

Dari hasil proses pendeteksian, kata majemuk/gabungan kata belum dapat terdeteksi seperti contoh pada gambar berikut.

**KBBI**

meja (meja) *n* perkakas (perabot) rumah yg mempunyai bidang datar sbg dasr mejanya dan berkaki sbg penyangga (bermacam-macam bentuk dan gayanya). -- bnda meja untuk bermain bulut. -- hi-jau pengabdian. -- makan meja yg khusus untuk marata hidangan. -- pempatan meja tempat pempatan (kepala keutor, pempatan rapat, dsb). -- tulis meja tempat kerja dan diperlekapngi dng laci-laci, ada yg dng rak tempat menyimpan buku atau berkas-berkas. -- dapur meja yg biasanya



Kata Pada KBBI	Kata Terdeteksi
Meja	Meja
Meja Bola	Tidak Terdeteksi
Meja Hujan	Tidak Terdeteksi
Meja Makan	Tidak Terdeteksi
Meja Pempatan	Tidak Terdeteksi
Meja Tulis	Tidak Terdeteksi

Gambar 12. Analisis Hasil Pengujian

## 3. PENUTUP

Pada bagian ini akan menjelaskan kesimpulan dan saran dari penelitian yang dilakukan.

### 3.2 Kesimpulan

Dalam hal ini pengembangan pendeteksian kamus jenis kata telah mendapatkan hasil lema sebanyak 43.060 lema dengan persentase sebanyak 65,38% dari total jumlah kata dasar dan turunannya sebanyak 65.857 lema. Berdasarkan hasil analisis, peningkatan hasil pendeteksian terjadi karena adanya optimasi pada teknik konversi data masukan per abjad dan penyelesaian masalah – masalah pada tahap pengambilan kata dan jenis kata. Namun peningkatan pendeteksian sekitar 4190 lema (kata) dengan persentase sebesar 6,36% karena dalam proses pendeteksian tidak mencakup kata majemuk dan pengambilan makna kata.

### 3.3 Saran

Saran untuk penelitian berikutnya adalah peningkatan atau pembaruan pengkodean agar dapat mendeteksi kata dan jenis kata yang dilengkapi makna kata dan label kata yang berfungsi untuk menentukan penggunaan kata tersebut. Pengembangan selanjutnya dilakukan dengan menyertakan kata majemuk dan makna kata. Dengan menambahkan makna kata pada kamus jenis kata, maka semakin baik dalam penggunaannya, selain itu, makna kata dapat dimanfaatkan untuk penelitian NLP lainnya yang membutuhkan makna kata seperti pembangunan *Wordnet* dan *Word Sense Disambiguation*.

## DAFTAR PUSTAKA

- [1] C G Ceppy Efraim Bolly, *PEMBANGUNAN KAMUS JENIS KATA SEBAGAI SUMBER DAYA NLP BAHASA INDONESIA*, vol. I, Agustus 2016.
- [2] Harimurti, Kridalaksana, *Kelas Kata Dalam Bahasa Indonesia*, 2nd ed. Jakarta: PT Gramedia Jakarta, 1994.