

EKSTRAKSI KALIMAT BERBAHASA INDONESIA BERPOLA SPOK MENGGUNAKAN PEMROSESAN BAHASA ALAMI

Reza Fauzan¹⁾, Joni Riadi²⁾,
Muhammad Hafni³⁾

reza.fauzan@poliban.ac.id¹⁾, joni.riadi@poliban.ac.id²⁾,
debiyou@gmail.com³⁾

1, 2, 3) Teknik Informatika, Politeknik Negeri Banjarmasin

Abstrak

Penggunaan bahasa Indonesia yang baik dan benar di zaman sekarang sungguh memprihatinkan. Pengaruh sosial media dapat memenuhi aspek fungsi definisi bahasa Indonesia yang tepat. Dengan ekstraksi metadata memungkinkan untuk memisah sebuah kalimat menjadi kata-kata yang akan digunakan untuk mengelompokkan kata sesuai dengan SPOK. Pembangunan aplikasi ini bertujuan untuk mempermudah pembelajaran untuk kaum muda khususnya untuk generasi yang akan datang untuk mempelajari Bahasa Indonesia yang baik dan benar. Diharapkan dengan adanya aplikasi ini masyarakat Indonesia sadar akan pentingnya Bahasa Indonesia dan kesadaran untuk menggunakan Bahasa Indonesia yang baik dan benar akan meningkat.

Kata Kunci : Bahasa Indonesia, SPOK, *Natural Language Processing*

1. PENDAHULUAN

Dalam arti yang sangat singkat, bahasa adalah alat atau sarana untuk berkomunikasi. Bahasa dalam Kamus Besar Bahasa Indonesia adalah sistem lambang bunyi arbitrer, yang digunakan oleh anggota suatu masyarakat untuk bekerja sama, berinteraksi, dan mengidentifikasi diri. Beralih ke penggunaan bahasa di setiap bangsa atau negara, bahasa mengambil peran yang sangat penting dan merupakan identitas suatu bangsa. Indonesia mempunyai banyak bahasa, yang semakin memperjelas identitas negara Indonesia dengan negara lain, tetapi bahasa yang dapat menyatukan masyarakat Indonesia sendiri dan telah di akui sebagai bahasa nasional pada saat Sumpah Pemuda 28 Oktober 1928 adalah bahasa Indonesia.

Penggunaan bahasa Indonesia yang baik dan benar di zaman sekarang sungguh memprihatinkan. Kemajuan teknologi yang semakin berkembang, memaksa para kaum muda di zaman sekarang kurang memperdulikan penggunaan bahasa Indonesia yang tepat. Anak muda sekarang lebih cenderung menggunakan bahasa atau ungkapan yang sedang populer di seluruh dunia. Pengaruh sosial media dapat memenuhi aspek fungsi definisi bahasa Indonesia yang tepat. Sehingga ini membuat

kedudukan bahasa Indonesia semakin terjepit. Banyak orang berdalih bahwa berbahasa itu yang terpenting lawan berbicara dapat memahami informasi yang kita sampaikan, dan tidak harus menggunakan bahasa yang baik dan benar sebagaimana yang diatur dalam bahasa Indonesia. Pretensi itu berkembang menjadi sebuah aksioma di tengah masyarakat. Dampaknya, bahasa Indonesia menjadi terabaikan.

Dengan adanya aplikasi yang bisa mengelompokkan suatu kalimat menjadi kata-kata yang dikelompokkan sesuai dengan SPOK ini dapat meningkatkan kesadaran masyarakat Indonesia untuk menggunakan Bahasa Indonesia yang baik dan benar.

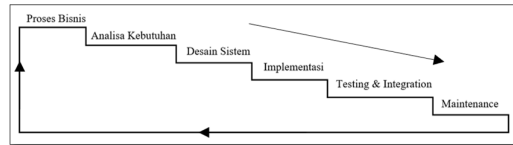
2. METODE PENELITIAN

Tahapan penelitian dalam pembuatan aplikasi ini adalah dengan menggunakan metode waterfall. Adapun fase-fase pada metode tersebut akan ditunjukkan pada Gambar 1.

2.1 Penelitian Terdahulu

Data yang disiapkan terdiri dari 39.000 kalimat yang telah diberikan tanda POS tag sebelumnya. Pada setiap kata dalam kalimat

telah memiliki POS tag sendiri. Sebagai contoh : Indeks/NN tersebut/DT naik/NN 1.1/JJ %/SYM pada/IN kuartal/NN kedua/CDP /. Penjelasan tagging pada setiap kata terdapat pada Tabel 1



Gambar 1. Tahapan Waterfall

Tabel 1. Tagging

No	POS	POS Name	Contoh
1	OP	Open Parenthesis	{
2	CP	Close Parenthesis	}
3	GM	Slash	/
4	;	Semicolon	;
5	:	Colon	:
6	"	Quotation	"
7	.	Sentence Terminate	!?
8	,	Comma	,
9	-	Dash	-
10	...	Ellipsis	...
11	JJ	Adjective	Kaya, Manis
12	RB	Adverb	Sementara
13	NN	CommonNoun	Mobil
14	NNP	ProperNoin	Indonesia
15	NNG	Genitive Noun	Bukunya
16	VBI	IntransitiveVerb	Pergi
17	VBT	TransitiveVerbb	Pergi
18	IN	Preposition	Di, Dari, Ke
19	MD	Modal	Bisa
20	CC	Coor-Conjunction	Dan, Atau
21	SC	Subor-Conjunction	Jika, Ketika
22	DT	Determiner	Ini, Itu
23	UH	Interiection	Wah, Aduh
24	CDO	OrdinalNumerals	Pertama
25	CDC	Collective Numerals	Bertiga
26	CDP	PrimaryNumerals	Satu, Dua
27	CDI	IrregularNumerals	Beberapa
28	PRP	Personal Pronouns	Saya, Kamu
29	WP	WH-Pronouns	Apa, Siapa
30	PRN	NumberPronouns	Ke-dua
31	PRL	LocativePronouns	Sini, Situ
32	NEG	Negation	Bukan, Tidak

33	SYM	Symbols	@#%\$^&
34	RP	Particles	Pun, Kah
35	FW	ForeignWords	Foreign, Word

Data kalimat yang telah dikumpulkan akan dirubah menjadi berkas berbentuk .prop menggunakan *lexicon tagger*. Kemudian berkas tersebut digenerasi menggunakan kelas *MaxentTagger* pada *Library Stanford NLP*. Berikut adalah cara penggunaan kelas *MaxentTagger* pada menggunakan *Command Line*.

```

java -mx300m -classpath stanford-postagger.jar
edu.stanford.nlp.tagger.m
axent.MaxentTagger -prop <file
property>
  
```

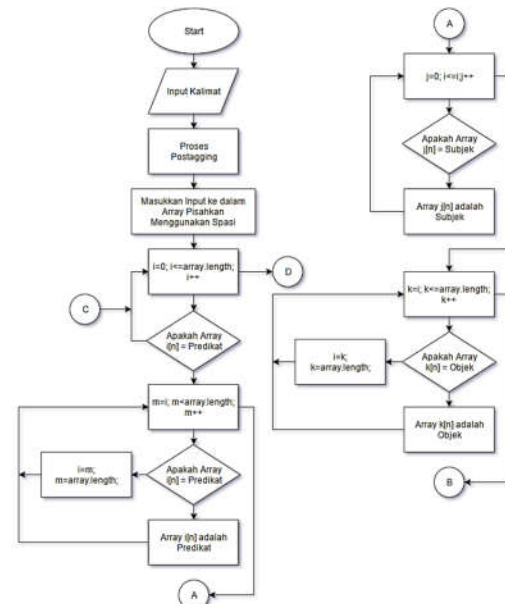
Contoh :

```

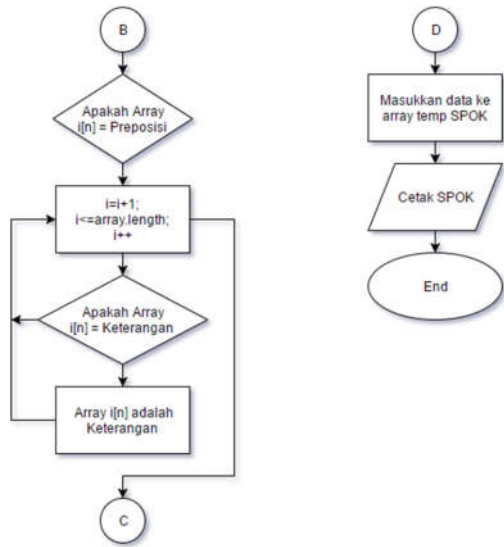
java -mx300m -classpath stanford-
postagger.jar
edu.stanford.nlp.tagger.m
axent.MaxentTagger -prop wsj-0-
18- left3words.tagger.props
  
```

2.2 Flowchart

Flowchart yang digunakan dapat dilihat dalam Gambar 2 dan Gambar 3 berikut :



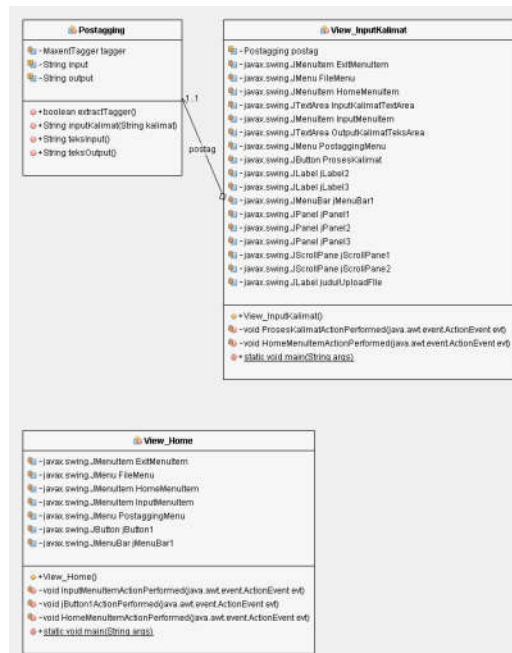
Gambar 2. Flowchart (1)



Gambar 3. Flowchart (2)

2.3 UML Diagram

Adapun UML diagram dalam sistem ini adalah sebagai berikut :

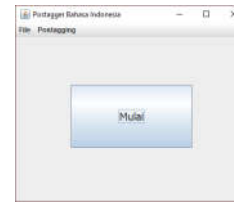


Gambar 4. UML Diagram

3. HASIL DAN PEMBAHASAN

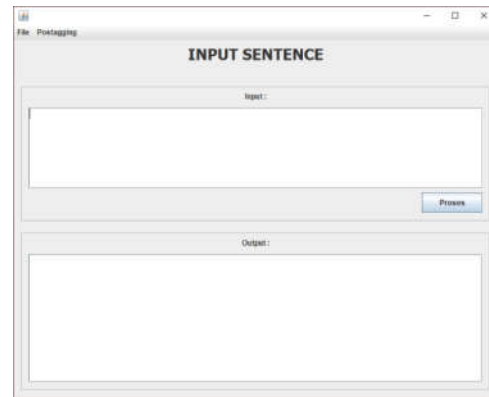
3.1 Hasil Implementasi

Berikut ini adalah implementasi dari aplikasi Ekstraksi Kalimat Berbahasa Indonesia Berpola SPOK Menggunakan Pemrosesan Bahasa Alami. Tampilan awal aplikasi menampilkan tombol mulai seperti Gambar 5. Tampilan awal aplikasi berikut :



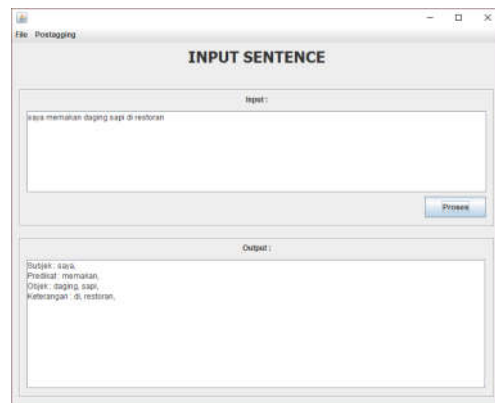
Gambar 5. Tampilan awal aplikasi

Pada Gambar 5 terdapat tombol mulai untuk memulai aplikasi ke tempat input output yang akan ditampilkan seperti pada Gambar 6. Tampilan Input Output berikut :



Gambar 6. Tampilan Input Output

Pada Gambar 6 terdapat kolom input untuk menginputkan data, tombol proses untuk memulai proses ekstraksi, dan kolom output untuk menampilkan hasil ekstraksi seperti yang ditampilkan pada Gambar 7. Hasil Output berikut :



Gambar 7. Hasil Output

3.2 Pengujian

Pengujian dilakukan dengan memasukkan contoh kata yang telah di proses manual oleh pakar. Hasil pengujian dibuat menjadi tabel yang dimana data pertama ialah Input untuk memasukkan salah satu bentuk kalimat, data kedua ialah Output untuk keluaran dari sistem yaitu berupa ekstraksi dari kalimat dan

di kelompokkan berdasarkan SPOK, dan data ketiga ialah validitas untuk ketepatan hasil output yang menjadi acuannya ialah benar atau salah.

Berikut ini adalah Tabel 2 Data dari Pakar :

Tabel 2. Data dari Pakar

No	Input	Output	Validitas
1	Mereka bermain bola	Subjek = Mereka Predikat = bermain Objek = bola Keterangan = null	Benar
2	Saya sedang sarapan bubur di rumah paman	Subjek = Saya Predikat = sedang sarapan Objek = bubur Keterangan = di rumah paman	Benar
3	Bapak sedang mencuci mobil	Subjek : Bapak Predikat : sedang mencuci Objek : mobil Keterangan : null	Benar
4	Diana mengerjakan PR di rumah	Subjek : Diana, Predikat : mengerjakan Objek : PR Keterangan : di rumah	Benar
5	Udin bersepeda di jalan	Subjek : udin Predikat : bersepeda Objek : Keterangan : di jalan	Benar
6	Saya sedang pergi berbelanja ke pasar	Subjek : saya Predikat : sedang pergi berbelanja Objek : Keterangan : ke pasar	Benar

7	Mobil sedang diperbaiki itu di	Subjek : mobil Predikat : sedang Objek : Keterangan : di diperbaiki	Benar
8	Udin sedang mandi di pantai	Subjek : Udin Predikat : sedang mandi Objek : Keterangan : di pantai	Benar
9	Ayah sedang bekerja di kantor	Subjek : Ayah Predikat : sedang bekerja Objek : Keterangan : di kantor	Benar
10	Ina memasak daging sapi di dapur	Subjek : ina Predikat : memasak Objek : daging sapi Keterangan : di dapur	Benar

Dan berikut ini adalah Tabel 3. Data dari Aplikasi, yaitu data yang didapat dari aplikasi ekstraksi metadata.

Tabel 3. Data dari Aplikasi

No	Input	Output	Validitas
1	Mereka bermain bola	Subjek = Mereka Predikat = bermain Objek = bola Keterangan =	Benar
2	Saya sedang sarapan bubur di rumah paman	Subjek = Saya Predikat = sedang sarapan Objek = bubur Keterangan = di rumah paman	Benar

3	Bapak sedang mencuci mobil	Subjek : Bapak Predikat : sedang mencuci Objek : mobil Keterangan :	Benar
4	Diana mengerjakan PR di rumah	Subjek : Diana, Predikat : mengerjakan Objek : PR Keterangan : di rumah	Benar
5	Udin bersepeda di jalan	Subjek : udin Predikat : bersepeda Objek : Keterangan : di jalan	Benar
6	Saya sedang pergi berbelanja ke pasar	Subjek : saya Predikat : sedang pergi berbelanja Objek : Keterangan : ke pasar	Benar
7	Mobil itu sedang di perbaiki	Subjek : mobil Predikat : sedang Objek : Keterangan : di perbaiki	Benar
8	Udin sedang mandi di pantai	Subjek : Predikat : sedang mandi Objek : Keterangan : di pantai	Salah
9	Ayah sedang bekerja di kantor	Subjek : Predikat : sedang bekerja Objek : Keterangan : di kantor	Salah
10	Ina memasak daging sapi di dapur	Subjek : ina Predikat : memasak Objek : daging sapi Keterangan : di dapur	Benar

Nilai persentase yang didapat dari data percobaan diatas ialah :
Persentase : $8/10 \times 100\% = 80\%$

4. KESIMPULAN

Kesimpulan yang didapat adalah aplikasi ekstraksi kalimat berbahasa Indonesia berbasis SPOK menggunakan metode *NLP* dapat berjalan sesuai dengan masalah yang dihadapi yaitu bagaimana cara menggunakan Bahasa pemrograman java untuk melakukan ekstraksi pada kalimat yang dimana pada aplikasi ini menggunakan split atau memisahkan kata dengan acuan spasi pada kalimat untuk mengidentifikasi antara satu kata dengan kata yang lainnya.

5. DAFTAR PUSTAKA

- [1] Adi (2013) Sejarah Java dan Pengertian Java. Available at: http://rionoadi.blogspot.co.id/2013/04/sejarah-java-dan-pengertian-java_1651.html (Accessed: 7 March 2017).
- [2] Fauzan, R. (2014) 'POS TAGGING BAHASA INDONESIA DENGAN MAXIMUM ENTROPY MENGGUNAKAN STANFORD POSTAGGER', 2, pp. 26–30.
- [3] Rashel, F.(2014) 'Building an Indonesian Rule-Based Part-of-Speech Tagger', 2, pp. 1-4.
- [4] Haloho, D. S. (2015) Penggunaan Bahasa Indonesia Zaman Sekarang. Available at: https://www.kompasiana.com/dessyhaloho/penggunaan-bahasa-indonesia-zaman-sekarang_5517e55c813311ad689de520 (Accessed: 29 January 2018).
- [5] Novtani, D. (2012) *NLP (Natural Language Programming)*. Available at: <https://novtani.wordpress.com/2012/06/04/natural-language-processing-nlp/#more-972> (Accessed: 8 March 2018).
- [6] Youri (2015) Ciri-ciri S-P-O-K (Struktur Kalimat). Available at: <http://modulgurucerdas.blogspot.co.id/2014/04/ciri-ciri-s-p-o-k-struktur-kalimat.html> (Accessed: 8 March 2018).
- [7] Agus (2013) Metode Pengembangan Sistem Waterfall. Available at: <https://agusdar.wordpress.com/2013/04/13/metode-pengembangan-sistem-waterfall/> (Accessed 23 March 2018)
- [8] Standford Log-linear Part-of-Speech Tagger. Available at:

<https://nlp.stanford.edu/software/tagger.html>
(Accessed: 23 March 2018)