

**IDENTIFIKASI BAHASA *JAWA VS INDONESIA*
DENGAN ALGORITMA PANJANG KATA BERBASIS
LEKSIKON**

Skripsi



oleh

Stephani Nugroho

71140022

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA

2019

**IDENTIFIKASI BAHASA *JAWA VS INDONESIA*
DENGAN ALGORITMA PANJANG KATA BERBASIS
LEKSIKON**

Skripsi



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

STEPHANI NUGROHO 71140022

PERNYATAAN KEASLIAN SKRIPSI

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

IDENTIFIKASI BAHASA JAWA VS INDONESIA DENGAN ALGORITMA PANJANG KATA BERBASIS LEKSIKON

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaaan saya.

Yogyakarta, 24 Juni 2019



71140022

HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN

Judul Skripsi : IDENTIFIKASI BAHASA JAWA VS INDONESIA
DENGAN ALGORITMA PANJANG KATA
BERBASIS LEKSIKON

Nama Mahasiswa : STEPHANI NUGROHO
NIM : 71140022
Matakuliah : Skripsi (Tugas Akhir)
Kode : TIW276
Semester : Genap
Tahun Akademik : 2018/2019

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 22 Mei 2019

Dosen Pembimbing I


Lucia Dwi Krisnawati, Dr. Phil.

Dosen Pembimbing II


Sri Suwamo, Dr. Ir. M.Eng.

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

IDENTIFIKASI BAHASA JAWA VS INDONESIA DENGAN ALGORITMA PANJANG KATA BERBASIS LEKSIKON

Oleh: STEPHANI NUGROHO / 71140022

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 17 Juni 2019

Yogyakarta, 24 Juni 2019
Mengesahkan,

Dewan Penguji:

1. Lucia Dwi Krisnawati, Dr. Phil.
2. Sri Suwarno, Dr. Ir. M.Eng.
3. Kristian Adi Nugraha, S.Kom., M.T.
4. Aditya Wikan Mahastama, S.Kom., M.Cs.



Dekan

(Budi Susanto, S.Kom., M.T.)

Ketua Program Studi

(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Dalam penelitian tugas akhir ini, penulis mendapatkan bantuan, saran, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada :

1. Bapak Budi Susanto, S.Kom., M.T. selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
2. Ibu Gloria Virginia, S.Kom., MAI. selaku Ketua Program Studi Teknik Informatika Universitas Kristen Duta Wacana.
3. Ibu Dr. Lucia Dwi Krisnawati, selaku dosen pembimbing I yang telah menyediakan waktu untuk melakukan konsultasi dan memberikan saran dan masukan mengenai pemrograman sistem, penyelesaian setiap kendala yang ada sampai akhir laporan.
4. Bapa Dr. Ir. Sri Suwarno, M.Eng. selaku dosen pembimbing II yang telah menyediakan waktu untuk memberikan bimbingan dan memberikan saran serta masukan mengenai penulisan laporan dan konsep sistem.
5. Teman – teman yang telah membantu dalam pengerjaan dokumen evaluasi.
6. Orangtua, saudara-saudara dan teman-teman terdekat Grup “Ingat Skripsimu”, “Gimbap Lovers” dan Grup “Penper Gass” yang selalu memberikan dukungan motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
7. Pihak – pihak lain yang tidak dapat penulis sebutkan satu per satu yang berperan secara langsung maupun tidak langsung selama pengerjaan tugas akhir.

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan kasih karunia-Nya sehingga penulis dapat menyelesaikan pembuatan sistem dan laporan tugas akhir dengan judul “IDENTIFIKASI BAHASA JAWA VS INDONESIA DENGAN ALGORITMA PANJANG KATA BERBASIS LEKSIKON” dengan baik.

Penulisan laporan tugas akhir ini diajukan sebagai salah satu syarat guna mencapai gelar Sarjana Strata Satu (S1) di Fakultas Teknologi Informasi Program Studi Teknik Informatika Universitas Kristen Duta Wacana Yogyakarta.

Dalam pembuatan laporan ini, penulis menyadari bahwa masih ada kekurangan, baik dari materi maupun teknik penyajiannya. Oleh karena itu, penulis sangat mengharapkan adanya kritik dan saran dari pembaca. Akhir kata penulis memohon maaf apabila dalam penulisan laporan ini, ada kalimat yang kurang berkenan. Semoga hasil dari penulisan tugas akhir ini dapat berguna dan bermanfaat bagi banyak pihak.

Yogyakarta, 24 Mei 2019

Penulis

INTISARI

IDENTIFIKASI BAHASA JAWA VS INDONESIA DENGAN ALGORITMA PANJANG KATA BERBASIS LEKSIKON

Identifikasi bahasa (*Language Identifier*) merupakan langkah awal dari sistem pemrosesan teks untuk menentukan suatu bahasa dimana teks yang digunakan tertulis. Sistem identifikasi bahasa yang akan dirancang lebih mengacu pada bahasa dengan sumber digital yang masih rendah (*under-resourced languages*) yaitu bahasa Jawa. Tujuan dari perancangan sistem ini adalah untuk mengidentifikasi dokumen kedua bahasa dengan algoritma Panjang Kata berbasis leksikon.

Pada penelitian ini, penulis menggunakan metode panjang kata berbasis leksikon untuk mempercepat pencarian kata dari pembangunan leksikon kemudian membandingkan dokumen uji dengan hasil sistem. Pra-pemrosesan dalam penelitian ini terbagi dalam 2 tahap yaitu normalisasi yang terdiri dari *case folding*, dan tokenisasi. Proses pengecekan dilakukan dengan menggunakan algoritma panjang kata yang terbagi menjadi 8 kategori karakter huruf. Hasil *output* sistem berupa kalimat yang telah teridentifikasi sesuai dengan bahasa yang ditemukan. Nilai akhir sistem didapatkan dari presentase jumlah kalimat terbesar yang didapatkan dari total masing – masing kalimat dibagi dengan seluruh kalimat di dalam dokumen.

Hasil akurasi yang didapatkan sebesar 100 % karena dokumen tersebut seluruhnya teridentifikasi kalimat baik Bahasa Indonesia maupun bahasa Jawa melalui penggunaan metode panjang kata berbasis leksikon. Kelemahan sistem adalah sistem mengidentifikasi kalimat Lain yang diperoleh jika jumlah kata bahasa Jawa sama dengan jumlah kata bahasa Indonesia. Hal ini disebabkan karena terdapat kata yang sama pada leksikon bahasa Jawa dan bahasa Indonesia sehingga teridentifikasi pada kedua bahasa.

Kata Kunci: *Language Identifier*, digital, leksikon, Panjang Kata

DAFTAR ISI

HALAMAN JUDUL.....	
PERNYATAAN KEASLIAN SKRIPSI.....	ii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN	v
UCAPAN TERIMA KASIH.....	vi
KATA PENGANTAR	vii
INTISARI	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiv
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah.....	2
1.3. Batasan Masalah	2
1.4. Tujuan Penelitian	2
1.5. Metodologi Penelitian.....	2
3.1 Sistematika Penulisan	3
BAB 2 TINJAUAN PUSTAKA	5
2.1. Tinjauan Pustaka	5
2.2. Landasan Teori	6
2.2.1. Metode Berbasis Leksikon (<i>lexicon</i>)	6
2.2.2. Pra-pemrosesan	7
2.2.2. Komputasi Matriks Biner.....	8
2.2.3. Metode Berbasis <i>Lexicon</i> dengan <i>Word Length Statistics</i>	8
2.2.4. Metode Evaluasi Sistem.....	9
BAB 3 PERANCANGAN SISTEM.....	12
3.1. Tahapan Pembangunan Sistem.....	12
3.1.1. Pra-pemrosesan	12
3.1.2. Pemrosesan	12
3.1.3. Tahap Evaluasi.....	13
3.2. Analisis Kebutuhan Sistem	13
3.2.1. Kebutuhan Fungsional	13

3.2.2.	Kebutuhan Non – Fungsional	13
3.2.3.	Kebutuhan <i>Hardware</i> dan <i>Software</i>	13
3.3.	Rancangan Sistem	13
3.3.1.	Alur Kerja Sistem	14
3.3.2.	Alur Kerja Sistem Pra-pemrosesan.....	14
3.3.3.	Alur Kerja Sistem Algoritma Panjang Kata.....	15
3.3.4.	<i>Use Case Diagram</i>	18
3.3.5.	<i>User Interface</i>	20
3.3.6.	<i>Database</i>	24
3.4.	Kamus Data	25
3.5.	Rancangan Pengujian	25
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM		27
4.1.	Implementasi Sistem	27
4.1.1.	Implementasi Pembangunan Leksikon	27
4.1.2.	Implementasi <i>Database</i>	32
4.1.3.	Implementasi Antarmuka.....	32
4.2.	Implementasi Program	38
4.2.1.	Proses Pra-pemrosesan.....	38
4.2.2.	Implementasi Identifikasi Bahasa.....	39
4.3.	Evaluasi dan Pembahasan	41
4.3.1.	Hasil Evaluasi Sistem	42
4.3.2.	Pembahasan Analisis	46
BAB 5 KESIMPULAN DAN SARAN		48
5.1.	Kesimpulan.....	48
5.2.	Saran.....	48
DAFTAR PUSTAKA		49

DAFTAR TABEL

Tabel 3.1. Tabel Contoh Kasus untuk Kategori Panjang Kata	17
Tabel 3.2. Tabel Use Case Masukkan Data	19
Tabel 3.3. Tabel Use Case Perancangan Leksikon	20
Tabel 3.4. Tabel Use Case Evaluasi.....	20
Tabel 3.5. Kamus Data.....	25
Tabel 4.1. Data Statistik Sumber Data Pembangunan Leksikon.....	27
Tabel 4.2. Data Statistik Dokumen Uji.....	41
Tabel 4.3. Pengujian Presisi Dokumen Uji Jawaindo01 dan Dokumen Evaluasi	42
Tabel 4.4. Presisi 30 Dokumen Uji.....	43
Tabel 4.5. Pengujian Akurasi Dokumen Jawaindo01	45
Tabel 4.6. Hasil Pengujian Akurasi Sistem	45

DAFTAR GAMBAR

Gambar 2.1. Alur Pra-pemrosesan.....	7
Gambar 2.2. Contoh Proses Tokenisasi pada Kalimat.....	7
Gambar 2.3. Contoh Proses Case Folding	7
Gambar 2.4. Confusion Matrix dengan 2 kelas ("true", "false")	10
Gambar 3.1. Alur Kerja Sistem.....	14
Gambar 3.2. Tahap Pra-pemrosesan	15
Gambar 3.3. Flowchart Proses Algoritma Panjang Kata	16
Gambar 3.4. Use Case Diagram Sistem Identifikasi Bahasa Jawa Vs Indonesia	19
Gambar 3.5. User Interface Halaman Home.....	21
Gambar 3.6. User Interface Halaman Leksikon Kata	21
Gambar 3.7. User Interface Halaman Leksikon Singkatan.....	22
Gambar 3.8. User Interface Halaman Panjang Kata	22
Gambar 3.9. User Interface Halaman Panjang Kata Result.....	23
Gambar 3.10. User Interface Halaman Evaluasi.....	23
Gambar 3.11. User Interface Halaman Evaluasi Result	24
Gambar 3.12. Rancangan Database Leksikon	24
Gambar 4.1. Pseudocode Kategori Panjang Kata pada Leksikon.....	29
Gambar 4.2. Contoh Tabel Kata Bahasa Indonesia	30
Gambar 4.3. Contoh Tabel Kata dalam Bahasa Jawa.....	30
Gambar 4.4. Contoh Tabel Singkatan Bahasa Indonesia.....	31
Gambar 4.5. Contoh Tabel Singkatan Bahasa Jawa	31
Gambar 4.6. Desain Leksikon pada Database	32
Gambar 4.7. Tampilan menu Home.....	33
Gambar 4.8. Tampilan menu Leksikon Kata	34
Gambar 4.9. Tampilan menu Leksikon Singkatan.....	34
Gambar 4.10. Tampilan menu awal Identifikasi Bahasa	35
Gambar 4.11. Tampilan Pilih File Dokumen.....	35
Gambar 4.12. Tampilan Proses Progress Bar	36
Gambar 4.13. Tampilan Proses Progress Bar selesai.....	36
Gambar 4.14. Tampilan Hasil Dokumen Jawaindo01	37
Gambar 4.15. Tampilan Hasil Dokumen lanjutan Jawaindo01	37

Gambar 4.16. Pseudocode Tahap Normalisasi	38
Gambar 4.17. Pseudocode Tahap Tokenisasi	39
Gambar 4.18. Pseudocode Pengecekan Kata dalam Kalimat	40
Gambar 4.19. Pseudocode Hasil Akhir Identifikasi Bahasa	41

©UKDW

DAFTAR LAMPIRAN

LAMPIRAN A Listing program	A-1
LAMPIRAN B Scan Kartu Konsultasi Tugas Akhir	B-1
LAMPIRAN C Formulir Perbaikan (Revisi) Skripsi	C-1

©UKDWN

INTISARI

IDENTIFIKASI BAHASA JAWA VS INDONESIA DENGAN ALGORITMA PANJANG KATA BERBASIS LEKSIKON

Identifikasi bahasa (*Language Identifier*) merupakan langkah awal dari sistem pemrosesan teks untuk menentukan suatu bahasa dimana teks yang digunakan tertulis. Sistem identifikasi bahasa yang akan dirancang lebih mengacu pada bahasa dengan sumber digital yang masih rendah (*under-resourced languages*) yaitu bahasa Jawa. Tujuan dari perancangan sistem ini adalah untuk mengidentifikasi dokumen kedua bahasa dengan algoritma Panjang Kata berbasis leksikon.

Pada penelitian ini, penulis menggunakan metode panjang kata berbasis leksikon untuk mempercepat pencarian kata dari pembangunan leksikon kemudian membandingkan dokumen uji dengan hasil sistem. Pra-pemrosesan dalam penelitian ini terbagi dalam 2 tahap yaitu normalisasi yang terdiri dari *case folding*, dan tokenisasi. Proses pengecekan dilakukan dengan menggunakan algoritma panjang kata yang terbagi menjadi 8 kategori karakter huruf. Hasil *output* sistem berupa kalimat yang telah teridentifikasi sesuai dengan bahasa yang ditemukan. Nilai akhir sistem didapatkan dari presentase jumlah kalimat terbesar yang didapatkan dari total masing – masing kalimat dibagi dengan seluruh kalimat di dalam dokumen.

Hasil akurasi yang didapatkan sebesar 100 % karena dokumen tersebut seluruhnya teridentifikasi kalimat baik Bahasa Indonesia maupun bahasa Jawa melalui penggunaan metode panjang kata berbasis leksikon. Kelemahan sistem adalah sistem mengidentifikasi kalimat Lain yang diperoleh jika jumlah kata bahasa Jawa sama dengan jumlah kata bahasa Indonesia. Hal ini disebabkan karena terdapat kata yang sama pada leksikon bahasa Jawa dan bahasa Indonesia sehingga teridentifikasi pada kedua bahasa.

Kata Kunci: *Language Identifier*, digital, leksikon, Panjang Kata

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Identifikasi bahasa (*language identifier*) merupakan langkah awal dari sistem pemrosesan teks untuk menentukan bahasa alami dimana teks yang diberikan tertulis. Terkait dengan identifikasi bahasa, usaha terbaik yang dilakukan adalah mencoba dan menyimpulkan informasi dari teks itu sendiri (Beesley, 1988). Identifikasi bahasa juga banyak digunakan di dalam aplikasi seperti *Machine translation*, *Text Mining*, *Information Retrieval*, dan *Text Manipulation Used Online with Mobile Devices*. Beberapa contoh dari aplikasi di atas membutuhkan identifikasi bahasa sebagai langkah awal yang perlu di selesaikan terlebih dahulu.

Identifikasi bahasa di sini terbagi menjadi 2 bahasa, yaitu bahasa minoritas dan bahasa populer (Ali & Akosu, 2015). Bahasa populer adalah bahasa yang memiliki sumber *digital* (Internet) paling banyak dan bahasa yang sering digunakan, sedangkan bahasa minoritas atau bahasa kurang sumber daya (*under resources*) artinya bahasa yang diucapkan oleh beberapa orang, tetapi mendapatkan peran penting karena meluasnya penggunaan Internet dan kemungkinan bahasa tersebut digunakan untuk komunikasi melalui Internet. Sejauh ini, masih sedikit penelitian yang dilakukan pada identifikasi dari bahasa minoritas. Hal ini mungkin karena bahasa minoritas sebelumnya dianggap kurang penting dibandingkan dengan bahasa populer. Menurut (Ramisch, 2008) penerapan n-gram pada model bahasa membutuhkan 150.000 kalimat pelatihan dan 11.000 kalimat uji. Ukuran data seperti itu sulit ditemukan dalam bahasa yang kurang sumber daya (*under resources*). Tapi ada penelitian serupa yang dibuktikan berhasil, yaitu penelitian yang dilakukan oleh (Pienaar & Snyman, 2010) yang mengaplikasikan generasi kedua *spelling checkers* untuk mengidentifikasi 11 bahasa Afrika Selatan. Keputusan menggunakan bahasa Afrika ini berdasarkan fakta bahwa bahasa Afrika Selatan merupakan bahasa yang kurang sumber *digital* (Internet). Mereka memperoleh akurasi lebih dari 95% sehubungan dengan identifikasi bahasa yang terkait erat (beberapa bahasa resmi di Afrika Selatan berasal dari keluarga yang sama).

Dari uraian di atas dapat disimpulkan bahwa teknik pemeriksa ejaan yang digunakan berhasil mengidentifikasi menggunakan beberapa bahasa yang kurang sumber daya. Meskipun, hal ini harus diuji pada lebih banyak bahasa dalam kategori yang sama. Permasalahan tersebut dapat diatasi dengan menggunakan pendekatan algoritma panjang kata

dengan metode berbasis leksikon (Ali & Akosu, 2015). Penggunaan pendekatan berbasis leksikon diharapkan dapat menghasilkan suatu sistem untuk bahasa yang sumber datanya relatif berjumlah kecil (*under resources*) dan kedepannya dapat dikembangkan seperti aplikasi - aplikasi dengan kebutuhan kompleks.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang dijelaskan di atas, rumusan masalah yang didapatkan adalah sebagai berikut :

1. Bagaimana penerapan algoritma panjang kata (*word length*) berbasis leksikon (*lexicon*) dalam proses identifikasi bahasa.
2. Bagaimana akurasi sistem yang akan dibuat dalam mencocokkan identifikasi bahasa dengan 2 skala, yaitu skala global dan lokal.

1.3. Batasan Masalah

Indonesia memiliki begitu banyak bahasa daerah. Bahasa yang akan digunakan dalam penelitian ini adalah Bahasa Jawa dan bahasa Indonesia. Skala identifikasi berhubungan dengan perbandingan data yang digunakan di dalam identifikasi bahasa. Batasan masalah yang diambil di dalam perbandingan tersebut terbagi menjadi 2, yaitu identifikasi skala global pada tingkat dokumen dan identifikasi skala lokal pada tingkat kalimat. Dokumen uji, pelatihan dan korpus yang digunakan untuk bahasa Jawa diambil dari korpus TRAWACA, sedangkan untuk bahasa Indonesia diambil dari Kamus Besar Bahasa Indonesia (KBBI). Dokumen yang akan dimasukkan oleh pengguna berformat .txt.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah membangun sebuah sistem yang dapat mengidentifikasi bahasa dari dokumen teks untuk kebutuhan aplikasi – aplikasi yang lebih kompleks seperti *machine translator*, *web scrapping*.

1.5. Metodologi Penelitian

Proses penelitian Identifikasi *Language Identifier* Bahasa Jawa Vs Indonesia memiliki beberapa langkah. Langkah – langkah yang akan dilakukan adalah sebagai berikut:

1. Studi Pustaka

Studi Pustaka dilakukan dengan cara mempelajari teori-teori melalui artikel, jurnal, buku, dan bahan pendukung lainnya, kemudian mengumpulkan data dari artikel-artikel

pendukung mengenai Identifikasi Bahasa dan metode algoritma Panjang Kata berbasis Leksikon.

2. Pengumpulan Data

Pada tahap ini, proses pengumpulan data diambil dari 4 web berita yaitu www.kompas.com, www.detik.com, www.ceritawayangbahasajawa.blogspot.com dan www.pawartabahasajawa.blogspot.com dengan campuran bahasa (Jawa dan Indonesia).

3. Pembuatan Sistem

Pada tahap ini, proses pembuatan program pemrosesan kata dan kalimat dalam bahasa Jawa dan Indonesia menggunakan algoritma panjang kata. Bahasa pemrograman yang digunakan adalah bahasa pemrograman web PHP.

4. Evaluasi Dan Analisis

Pada tahap ini, proses evaluasi dilakukan dengan cara menghitung nilai presisi dan akurasi sistem. Pengujian sistem dilakukan dengan membandingkan hasil dokumen bahasa *Jawa vs Indonesia* dalam tingkat dokumen yang telah diproses oleh sistem dengan dokumen bahasa *Jawa vs Indonesia* dalam tingkat kalimat.

3.1 Sistematika Penulisan

Sistematika penulisan yang digunakan penulis untuk mengerjakan laporan tugas akhir ini sesuai dengan ketentuan dari buku panduan penulisan tugas akhir. Penulisan akan dibagi menjadi 5 bab, yaitu :

Bab I : Pendahuluan

Pada bab pertama akan dibahas mengenai latar belakang, rumusan masalah, batasan dari sistem yang akan dibangun, tujuan yang ingin dicapai oleh penulis, dan metodologi penelitian.

Bab II : Tinjauan Pustaka

Pada bab kedua akan diuraikan mengenai tinjauan pustaka dan landasan teori. Pada tinjauan pustaka akan berisi penelitian-penelitian yang sudah pernah dilakukan oleh orang lain berkaitan dengan penelitian yang akan dilakukan oleh penulis. Landasan teori akan berisi

tentang dasar-dasar teori yang akan digunakan sebagai pendukung pada penelitian ini, termasuk metode-metode yang akan digunakan.

Bab III : Perancangan Sistem

Pada bab ketiga penulis membahas mengenai perancangan sistem yang akan dibuat. Bab ini menjelaskan bagaimana tahapan – tahapan pembangunan sistem, analisis kebutuhan sistem, rancangan sistem, serta rancangan pengujian dari sistem.

Bab IV : Implementasi dan Analisis Sistem

Pada bab keempat penulis akan membahas mengenai hasil implementasi yang telah dilakukan. Hasil yang akan didapatkan dapat berupa hasil riset atau penelitian yang telah dilakukan oleh penulis.

Bab V : Kesimpulan

Pada bab kelima penulis akan membahas mengenai kesimpulan dari hasil penelitian dan saran. Kesimpulan berisi tentang hasil yang telah didapatkan selama penelitian berlangsung. Sedangkan saran dapat berupa sesuatu hal yang dapat dilakukan berkaitan dengan pengembangan topik penelitian tersebut untuk kemajuan di masa mendatang.

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil pembahasan analisis pada Bab 4, maka dapat disimpulkan sebagai berikut:

1. Sistem sudah cukup baik untuk mengidentifikasi kalimat baik Bahasa Jawa maupun Bahasa Indonesia. Dapat dilihat pada hasil persentase presisi (ketepatan) dokumen nilai persentase terkecil adalah 0 % dalam kalimat bahasa Indonesia karena semua kalimat tersebut teridentifikasi bahasa Jawa sebanyak 100 % begitu juga dengan nilai persentase 0 % pada kalimat bahasa Jawa dikarenakan seluruh kalimat merupakan kalimat Bahasa Indonesia dengan persentase 100 %.
2. Hasil persentase nilai akurasi sistem terbesar 100 % tetapi nilai akurasi terkecil adalah 29,41 % dikarenakan sumber data leksikon pada bahasa Jawa dan bahasa Indonesia memiliki kata yang sama tetapi mempunyai arti kata yang berbeda sehingga kata tersebut muncul pada kedua bahasa.

5.2. Saran

Beberapa kekurangan yang sudah disampaikan pada Bab 4, sehingga nantinya dapat dikembangkan lebih lagi maka saran yang dapat diberikan antara lain:

1. *Stemming* kata bahasa Indonesia diperlukan agar nantinya sistem mampu mengenali kata dasar dari kata – kata berimbuhan, sehingga dapat mempengaruhi jumlah kalimat yang nantinya akan dipakai pada nilai presisi dan keakuratan sistem (akurasi).
2. Penambahan daftar kata dalam leksikon bahasa Jawa dan bahasa Indonesia perlu diperhatikan sumbernya sehingga percampuran kata dalam kedua bahasa dapat diminimalisir.
3. Meningkatkan kualitas dan kuantitas dokumen uji, dengan meminta ahli pakar yang sesuai dengan bahasa yang diuji untuk menuliskan sebuah artikel sebagai dokumen uji.

DAFTAR PUSTAKA

- Ali, S., & Akosu, N. (2015). Word-length Algorithm for Language Identification of Under-resourced Languages. *Journal of King Saud University - Computer and Information Sciences* , 457-569.
- Beesley, K. R. (1988). Language Identifier: A Computer Program For Automatic Natural-Language Identification of On-line Text. *Proceedings of the 29th Annual Conference of the American Translators Association*, (pp. 47-54).
- Botha, G., Zimu, V., & Barnard, E. (2007). Text-based Language Identification for the South African Languages. *SAIEE Africa Research Journal* , 7-13.
- Krisnawati, L. D., & Mahastama, A. W. (2018). A Javanese Syllabifier Based on its Orthographic System. *2018 International Conference on Asian Language Processing (IALP)* (pp. 244-249). Bandung: IEEE.
- Pienaar, W., & Snyman, D. (2010). Spelling Checker-based Language Identification for the Eleven Official South African Languages. *In: Proceedings of the Twenty-First Annual Symposium of the Pattern Recognition Association of South Africa*, (pp. 213-216). South Africa.
- Provost, F., & Kohavi, R. (1998). *On Applied Research in Machine Learning*. Boston: Kluwer Academic .
- Ramisch, C. (2008, December 21). N-gram Models for Language Detection. pp. 1-14.
- Winkelmolen, F., & Mascardi, V. (2011). Statistical Language Identification of Short Texts. *ICAART 2011 - Proceedings of the 3rd International Conference on Agents and Artificial Intelligence*, (pp. 498-503). Rome.