

***Word Sense Disambiguation Menggunakan
Algoritma Simplified Lesk Pada Kalimat Bahasa Inggris***

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh:

Najlah Afifah
NIM : 09021181621035

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2021**

LEMBAR PENGESAHAN SKRIPSI

Word Sense Disambiguation Menggunakan Algoritma Simplified Lesk Pada Kalimat Bahasa Inggris

Oleh :

Najlah Afifah
NIM : 09021181621035

Palembang, 31 Agustus 2021

Pembimbing I,

Novi Yusliani, S.Kom., M.T
NIP.198211082012122001

Pembimbing II,

Osvari Arsalan, S.Kom., M.T
NIP. 198806282018031001

Mengetahui,
Ketua Jurusan,



Alvi Syahrin Utami, M.Kom.
NIP 197812222006042003

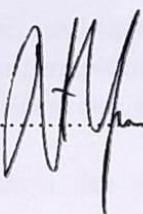
TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Sabtu tanggal 31 Juli 2021 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Najlah Afifah
NIM : 09021181621035
Judul : Word Sense Disambiguation Menggunakan Algoritma Simplified Lesk Pada Kalimat Bahasa Inggris

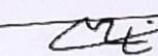
1. Pembimbing I

Novi Yusliani, M.T
NIP. 198211082012122001



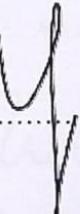
2. Pembimbing II

Osvari Arsalan, M.T
NIP. 198806282018031001



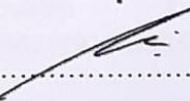
3. Pengaji I

Yunita, M.Cs
NIP. 198306062015042002



4. Pengaji II

Rizki Kurniati, M.T
NIP. 199107122019032016



Mengetahui,
Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Najlah Afifah

NIM : 09021181621035

Program Studi : Teknik Informatika

Judul Skripsi : Word Sense Disambiguation Menggunakan Algoritma Simplified
Lesk Pada Kalimat Bahasa Inggris

Hasil Pengecekan Software *iThenticate/Turnitin* : 19 %

Menyatakan bahwa Laporan Penelitian saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan penelitian ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 31 Agustus 2021



Najlah Afifah
NIM. 09021181621035

HALAMAN PERSEMBAHAN

MOTTO “KESEMPURNAAN HANYA MILIK ALLAH”

Skripsi ini dipersembahkan untuk :

- Orang Tua dan Keluarga Tercinta
- Sahabatku Tersayang
- Dosen Pembimbing
- Teman - Teman Seperjuangan IF REG A 2016
- Dosen - Dosen Jurusan Teknik Informatika
- Fakultas Ilmu Komputer, Universitas Sriwijaya

KATA PENGANTAR

Alhamdulillah, segala puji syukur dipanjangkan kepada Allah SWT atas segala rahmat dan karunia-Nya penulis dapat melaksanakan dan menyelesaikan Tugas Akhir yang berjudul “*Word Sense Disambiguation Menggunakan Algoritma Simplified Lesk Pada Kalimat Bahasa Inggris*” ini dengan baik. Tugas Akhir ini disusun sebagai salah satu syarat dalam menyelesaikan pendidikan program Strata-1 Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Selama penyelesaian Tugas Akhir ini tidak terlepas dari bimbingan dan bantuan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada seluruh pihak yang telah memberikan bantuan dan dukungan, baik dalam materil maupun moril selama proses pembuatan tugas akhir ini. Adapun penulis merujuk secara khusus pihak yang telah membantu sebagai berikut :

1. Bapak Jaidan Jauhari, S.Pd., M.T selaku Dekan Dakultas Ilmu Komputer Universitas Sriwijaya.
2. Ibu Ali Syahrini Utami, M.Kom selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Muhammad Qurhanul Rizqie, S.Kom., M.T selaku Dosen Pembimbing Akademik yang telah memberikan arahan dalam urusan akademik.
4. Ibu Novi Yusliani, S.Kom., M.T dan Bapak Osvari Arsalan, S.Kom., M.T selaku Dosen Pembimbing I dan Dosen Pembimbing II yang dengan sabar telah memberikan bimbingan, arahan, kritik dan saran kepada penulis dalam menyelesaian tugas akhir.
5. Ibu Yunita, M.Cs dan Ibu Rizki Kurniati, M.T selaku Dosen Penguji I dan Dosen Penguji II yang telah memberikan masukan dan arahan kepada penulis dalam menyempurnakan tugas akhir ini.
6. Seluruh Doses Teknik Informatika Fakultas Ilmu Komputer yang telah memberikan ilmu kepada penulis selama masa perkuliahan.
7. Staff Jurusan Teknik Informatika beserta staff Fakultas Ilmu Komputer yang telah membantu dalam urusan administrasi dan akademik selama masa perkuliahan.

8. Kedua orang tua, Aba (Muhammad Shahab) dan Mama (Fatimah Tuzzahra) yang selalu mendoakan dan memberi dukungan kepada penulis agar terus giat dan berusaha untuk menyelesaikan tugas akhir ini sampai selesai.
9. Saudara kandung penulis. Luthfiah, Ummu Farwa Syabirah dan Ali Al-Hadi yang selalu memberikan semangat terhadap penulis agar cepat menyelesaikan tugas akhir ini.
10. Sahabatku tercinta selama kuliah, Rima Melati, Nur Sanifa Arisyah, Syamila Hanun dan Suhartini yang selalu ada membantu penulis dan berjuang bersama-sama dalam suka maupun duka.
11. Karina, Agustina Hutapea, Dina Eliyatun Nasuha, Cesil Oktavia Hersa dan Mayti Pratiwi, sahabat seperjuangan yang selalu bekerjasama membantu menyelesaikan tugas akhir ini.
12. Teman seangkatan IF Regular A 2016 yang selalu ada dan membantu dalam masalah perkuliahan, menciptakan momen selama masa perkuliahan yang tidak akan pernah terlupakan.
13. Semua pihak lainnya yang tidak dapat disebutkan satu persatu yang telah berperan bagi penulis dalam menyelesaikan tugas akhir. Terima kasih semuanya.

Semoga tugas akhir ini dapat bermanfaat bagi penulis khususnya maupun pembaca pada umumnya. Serta dapat menjadi referensi dan rujukan bagi hal-hal yang bermanfaat. Penulis menyadari bahwa dalam proses penyelesaian tugas akhir ini, terdapat beberapa penjelasan yang kurang sempurna. Oleh karena itu penulis mengharapkan kritik dan saran yang membangun agar tugas akhir ini menjadi karya tulis yang sempurna supaya terciptanya bekal pengetahuan yang baik bagi peneliti di masa depan.

Palembang, 31 Agustus 2021



Najlah Afifah

ABSTRACT

Every natural language has words that can have more than one meaning that could potentially cause doubt or ambiguity. This ambiguity can be an obstacle in communication causing differences in understanding. Therefore, a solution is needed to determine the meaning of ambiguous words in sentences so that they can be understood, namely Word Sense Disambiguation. The word sense disambiguation is carried out with the aim of identifying the right meaning, according to the context of the sentence. Various methods have been used for Word Sense Disambiguation, one of which is the Simplified Lesk algorithm. In this study, the data used were in the form of a collection of English sentences obtained from the FEWS website. The test was carried out on 500 data in the form of English sentences that have ambiguous words and their respective meanings which have been broken down into 100, 200, 300, 400 and 500 sentences. The test results on the Simplified Lesk algorithm only succeeded in determining the meaning of words less than 50% of the input sentences in the dataset. Where the results of the greatest accuracy are obtained when the number of input datasets is 100 sentences, which is 0.40. While the smallest accuracy results are obtained when the number of input datasets is 500 and 400 sentences, which is 0.34. And produce an average accuracy of 0.36.

Key Words : Word Sense Disambiguation, WordNet, Simplified Lesk.

Palembang, 31 Agustus 2021

Pembimbing I,

Novi Yusliani, S.Kom., M.T
NIP.198211082012122001

Pembimbing II,

Osvari Arsalan, S.Kom., M.T
NIP. 198806282018031001

Mengetahui,
Ketua Jurusan,



Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003

ABSTRAK

Setiap bahasa alami memiliki kata yang dapat bermakna lebih dari satu yang berpotensi menyebabkan keraguan atau ambigu. Keambiguan tersebut dapat menjadi kendala dalam komunikasi menyebabkan perbedaan pemahaman. Maka dari itu diperlukan solusi menentukan makna kata ambigu pada kalimat agar dapat dimengerti yaitu *Word Sense Disambiguation*. Dilakukannya *Word Sense Disambiguation* bertujuan untuk mengidentifikasi makna yang tepat, sesuai dengan konteks kalimat. Berbagai macam metode telah digunakan untuk *Word Sense Disambiguation* salah satunya adalah algoritma *Simplified Lesk*. Pada penelitian ini data yang digunakan berupa kumpulan kalimat berbahasa Inggris yang didapatkan dari website FEWS. Pengujian dilakukan pada 500 data berupa kalimat bahasa Inggris yang memiliki kata ambigu dan maknanya masing-masing yang telah dipecah menjadi 100, 200, 300, 400 dan 500 kalimat. Hasil pengujian pada algoritma *Simplified Lesk* hanya berhasil menentukan makna kata kurang dari 50% dari kalimat masukan pada dataset. Dimana pada hasil akurasi terbesar diperoleh saat jumlah dataset masukan 100 kalimat yakni sebesar 0,40. Sedangkan Hasil akurasi terkecil diperoleh saat jumlah dataset masukan 500 dan 400 kalimat yakni sebesar 0,34. Serta menghasilkan rata-rata akurasi sebesar 0,36.

Kata Kunci : *Word Sense Disambiguation, WordNet, Simplified Lesk*.

Palembang, 31 Agustus 2021

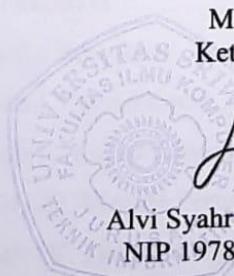
Pembimbing I,

Novi Yuslianii, S.Kom., M.T
NIP.198211082012122001

Pembimbing II,

Osvari Arsalan, S.Kom., M.T
NIP. 198806282018031001

Mengetahui,
Ketua Jurusan,



Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN SIDANG SKRIPSI.....	iii
HALAMAN PERNYATAAN.....	iv
HALAMAN PERSEMAHAN	v
KATA PENGANTAR	vi
ABSTRACT	viii
ABSTRAK	ix
DAFTAR ISI	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN.....	xvii
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah	I-1
1.3 Rumusan Masalah.....	I-3
1.4 Tujuan Penelitian.....	I-4
1.5 Manfaat Penelitian.....	I-4
1.6 Batasan Masalah	I-5
1.7 Sistematika Penulisan	I-5
1.7.1 Bab I. Pendahuluan.....	I-5
1.7.2 Bab II. Kajian Literatur.....	I-5
1.7.3 Bab III. Metode Penelitian	I-6
1.7.4 Bab IV. Pengembangan Perangkat Lunak	I-6
1.7.5 Bab V. Hasil Dan Analisa Peneltian.....	I-6
1.7.6 Bab VI. Kesimpulan Dan Saran	I-6
1.8 Kesimpulan	I-6

BAB II KAJIAN LITERATUR.....	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Ambiguitas	II-1
2.2.2 <i>Word Sense Disambiguation</i>	II-2
2.2.3 Wordnet.....	II-4
2.2.4 <i>Preprocessing</i>	II-5
2.2.5 Algoritma <i>Simplified Lesk</i>	II-6
2.2.6 Pengukuran Hasil Penelitian	II-7
2.2.7 <i>Rational Unified Process</i>	II-7
2.3 Penelitian Lain yang Relevan.....	II-9
2.4 Kesimpulan	II-10
 BAB III METODE PENELITIAN	III-1
3.1 Pendahuluan	III-1
3.2 Pengumpulan Data.....	III-1
3.2.1 Jenis dan Sumber Data.....	III-1
3.2.2 Metode Pengumpulan Data.....	III-2
3.3 Tahapan Penelitian	III-2
3.3.1 Menentukan Kerangka Kerja Penelitian	III-3
3.3.2 Menentukan Kriteria Pengujian.....	III-5
3.3.3 Menetapkan Format Data Pengujian	III-5
3.3.4 Menentukan Alat Bantu Penelitian.....	III-6
3.3.5 Melakukan Pengujian Penelitian	III-6
3.3.6 Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan Penelitian.....	III-6
3.4 Metode Pengembangan Perangkat Lunak.....	III-7
3.4.1 Fase Insepsi	III-7
3.4.2 Fase Elaborasi	III-8
3.4.3 Fase Konstruksi	III-8

3.4.4 Fase Transisi.....	III-8
3.5 Rencana Jadwal Penelitian.....	III-9
3.6 Kesimpulan	III-16
 BAB IV PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1 Pendahuluan	IV-1
4.2 Fase Insepsi	IV-1
4.2.1 Pemodelan Bisnis	IV-1
4.2.2 Kebutuhan Sistem.....	IV-2
4.2.2.1 Fitur Pra-Pengolahan Data.....	IV-2
4.2.2.2 Fitur Word Sense Disambiguation dengan Algoritma Simplified Lesk	IV-3
4.2.3 Analisis dan Desain	IV-3
4.2.3.1 Analisis Kebutuhan Perangkat Lunak	IV-3
4.2.3.2 Analisis Data.....	IV-4
4.2.3.3 Analisis Pra-Pengolahan.....	IV-4
4.2.3.4 Analisis Word Sense Disambiguation dengan Algoritma Simplified Lesk	IV-7
4.2.3.5 Analisis Hasil Penelitian.....	IV-13
4.2.4 Implementasi	IV-13
4.3 Fase Elaborasi.....	IV-18
4.3.1 Pemodelan Bisnis	IV-18
4.3.1.1 Perancangan Data.....	IV-19
4.3.1.2 Perancangan Antarmuka.....	IV-19
4.3.2 Kebutuhan Sistem.....	IV-20
4.3.3 Analisis dan Perancangan	IV-21
4.3.3.1 Diagram Aktivitas	IV-21
4.3.3.2 Sequence Diagram	IV-23
4.4 Fase Konstruksi	IV-25
4.4.1 Kebutuhan Sistem.....	IV-25
4.4.2 Diagram Kelas.....	IV-26

4.4.3 Implementasi	IV-27
4.4.3.1 Implementasi Kelas	IV-27
4.4.3.2 Implementasi Antarmuka	IV-28
4.5 Fase Transisi.....	IV-28
4.5.1 Pemodelan Bisnis	IV-29
4.5.2 Kebutuhan Sistem.....	IV-29
4.5.3 Rancangan Pengujian.....	IV-29
4.5.4 Implementasi	IV-31
4.6 Kesimpulan	IV-34
 BAB V HASIL DAN ANALISIS PENELITIAN	V-1
5.1 Pendahuluan	V-1
5.2 Data Hasil Penelitian	V-1
5.2.1 Konfigurasi Percobaan.....	V-1
5.2.2 Data Hasil Konfigurasi	V-2
5.3 Analisis Hasil Penelitian	V-3
5.4 Kesimpulan	V-4
 BAB VI KESIMPULAN DAN SARAN	VI-1
6.1 Pendahuluan	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran	VI-2
 DAFTAR PUSTAKA	xviii
LAMPIRAN	L-1

DAFTAR TABEL

	Halaman
Tabel II-1. Tabel Hasil Perhitungan	II-7
Tabel III-1. Data yang akan digunakan.....	III-2
Tabel III-2. Rancangan Tabel Hasil Pengujian	III-5
Tabel III-3. Rancangan Tabel Performa Hasil Penelitian	III-7
Tabel III-4. Tabel Rencana Jadwal Penelitian.....	III-10
Tabel IV-1. Kebutuhan Fungsional Perangkat Lunak	IV-2
Tabel IV-2. Kebutuhan Non-Fungsional Perangkat Lunak	IV-2
Tabel IV-3. Contoh Data.....	IV-4
Tabel IV-4. Hasil <i>Case Folding</i> dari Data	IV-5
Tabel IV-5. Hasil <i>Tokenizing</i> dari Data	IV-5
Tabel IV-6. Hasil <i>Stopword Removal</i> dari Data	IV-6
Tabel IV-7. Hasil <i>Stemming</i> dari Data.....	IV-6
Tabel IV-8. Proses Penentuan Kata Ambigu oleh Sistem.....	IV-7
Tabel IV-9. Proses Penentuan Kata Ambigu dari Data	IV-8
Tabel IV-10. Makna Kata Ambigu.....	IV-8
Tabel IV-11. Perhitungan <i>Overlap</i> Kalimat Pertama	IV-10
Tabel IV-12. Perhitungan <i>Overlap</i> Kalimat Kedua.....	IV-10
Tabel IV-13. Perhitungan <i>Overlap</i> Kalimat Ketiga.....	IV-11
Tabel IV-14. Perhitungan <i>Overlap</i> Kalimat Keempat	IV-11
Tabel IV-15. Perhitungan <i>Overlap</i> Kalimat Kelima.....	IV-12
Tabel IV-16. Perbandingan Hasil Makna Ambigu	IV-13
Tabel IV-17. Tabel Definisi Aktor	IV-14
Tabel IV-18. Tabel Definisi <i>Use Case</i>	IV-15
Tabel IV-19. Skenario <i>Use Case</i> User Melakukan Masukan.....	IV-16
Tabel IV-20. Skenario <i>Use Case</i> Memproses Data	IV-17
Tabel IV-21. Skenario <i>Use Case</i> Mendisambiguasi Makna Kata Menggunakan Algoritma <i>Simplified Lesk</i>	IV-18
Tabel IV-22. Tabel Implementasi Kelas	IV-27

Tabel IV-23. Rencana Pengujian <i>Use Case</i> Melakukan Masukan	IV-30
Tabel IV-24. Rencana Pengujian <i>Use Case</i> Memproses Data	IV-30
Tabel IV-25. Rencana Pengujian <i>Use Case</i> Melakukan Disambiguasi Makna Kata Menggunakan Algoritma <i>Simplified Lesk</i>	IV-30
Tabel IV-26. Pengujian <i>Use Case</i> Melakukan Masukan	IV-31
Tabel IV-27. Pengujian <i>Use Case</i> Memproses Data	IV-32
Tabel IV-28. Pengujian <i>Use Case</i> Melakukan Disambiguasi Makna Kata Menggunakan Algoritma <i>Simplified Lesk</i>	IV-33
Tabel V-1. Tabel Hasil Disambiguasi Makna Kata.....	V-2
Tabel V-2. Data Hasil Evaluasi Disambiguasi Makna Kata	V-2

DAFTAR GAMBAR

	Halaman
Gambar II-1. Tahapan Pada Preprocessing	II-5
Gambar II-2. Arsitektur Rational Unified Process	II-8
Gambar III-1. Tahapan Sistem Word Sense Disambiguation	III-3
Gambar IV-1. Diagram Use Case	IV-14
Gambar IV-2. Rancangan Antarmuka Perangkat Lunak	IV-20
Gambar IV-4. Diagram Aktivitas Memproses Data	IV-22
Gambar IV-5. Diagram Aktivitas Melakukan Disambiguasi Makna Kata Menggunakan Algoritma Simplified Lesk	IV-23
Gambar IV-6. Sequence Diagram Melakukan Masukan	IV-24
Gambar IV-7. Sequence Diagram Proses Data	IV-24
Gambar IV-8. Sequence Diagram Melakukan Disambiguasi Makna Kata	IV-25
Gambar IV-9. Diagram Kelas Perangkat Lunak	IV-26
Gambar IV-10. Implementasi Antarmuka Perangkat Lunak	IV-28

DAFTAR LAMPIRAN

Lampiran 1. Hasil Pengujian Data Uji	L-1
Lampiran 2. Kode Program.....	L-13

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab pendahuluan akan membahas tentang latar belakang masalah, rumusan masalah, tujuan, manfaat, batasan masalah, sistematika penulisan serta kesimpulan. Bab ini juga berisikan penjelasan mengenai gambaran umum dari keseluruhan kegiatan yang dilakukan dalam penelitian tugas akhir.

1.2 Latar Belakang Masalah

Setiap bahasa alami memiliki kata yang dapat bermakna lebih dari satu. Namun penggunaannya tergantung dari konteks kalimat yang menyertainya. Kata bermakna lebih dari satu tersebut, dapat berpotensi menyebabkan keraguan atau ambigu. Hal tersebut tentunya akan menjadi kendala dalam berkomunikasi menyebabkan keraguan dalam penafsiran kalimat itu sendiri. Maka dari itu diperlukan sebuah solusi menentukan makna kata ambigu pada kalimat agar dapat lebih dimengerti.

Penentuan makna kata yang ambigu termasuk salah satu topik *Natural Language Processing* (NLP) yang dikenal dengan *Word Sense Disambiguation*. *Word Sense Disambiguation* bertujuan mengidentifikasi makna kata yang tepat, sesuai dengan konteks kalimat (Basuki, Kholimi, Minarno, Sumadi, & Effendy, 2019).

Dalam *Natural Language Processing*, penggunaan *Word Sense Disambiguation* ini bermanfaat untuk perbaikan kata dalam mesin penerjemah.

Perbaikan kata ini dapat dilihat dalam *google translate*, dimana pengguna dapat membantu memperbaiki konteks kalimat yang telah diubah ke bahasa target dengan memilih kata yang dianggap ambigu. Selain itu dapat digunakan pula untuk *proof read*, dimana *proof read* adalah metode membaca ulang untuk memeriksa sebuah penulisan serta memastikan tidak ada penulisan yang salah, tidak konsisten dan tidak mengandung makna yang ambigu.

Berbagai macam metode telah digunakan untuk *Word Sense Disambiguation* salah satunya adalah algoritma *Simplified Lesk* yang diperkenalkan oleh Kilgariff dan Rosenzweig pada tahun 2000. Algoritma ini membandingkan definisi kata ambigu (*gloss*) beserta *examples* (contoh) kalimat dari *WordNet* dengan konteks kalimat dari kata ambigu tersebut (Suyuti, 2016).

Beberapa penelitian sebelumnya tentang *Word Sense Disambiguation* menggunakan algoritma *Simplified Lesk*, salah satunya (Basuki et al., 2019). Dalam penelitiannya menerapkan metode *Simplified Lesk* untuk *Word Sense Disambiguation* dalam penentuan arti kata homograf pada kalimat bahasa Indonesia. Pengujian dilakukan terhadap fungsionalitas dari sistem, dimana hasil pengujian sistemnya sesuai dengan data pengujian dari KBBI (Kamus Besar Bahasa Indonesia).

Penelitian lainnya oleh (Gitasari, 2007), menentukan makna kata yang tepat untuk menghilangkan disambiguasi makna kata dalam kalimat berbahasa Indonesia menggunakan algoritma *Simplified Lesk*. Hasil penelitian tersebut menilai bahwa kelengkapan basis data sangatlah perpengaruh dalam keberhasilan perangkat lunak. Menurut penelitian tersebut, algoritma *Simplified Lesk* tidak dapat menentukan

makna kata yang tepat jika tidak menemukan kata yang berelasi dengan kata dalam kalimat.

Hal tersebut sesuai dengan hasil penelitian lainnya yang dilakukan oleh (Putri, 2017) mengenai *Word Sense Disambiguation* menggunakan algoritma *Simplified Lesk* pada cerpen anak berbahasa Indonesia. Penelitian tersebut memiliki performa yang sangat kecil, yaitu tingkat akurasi sebesar 14%. Penyebab hasil performansi yang sangat kecil dikarenakan peneliti hanya menggunakan kamus kateglo sebagai sumber dalam menentukan kata yang ambigu.

Untuk mengatasi kondisi tersebut, peneliti akan menggunakan *WordNet*. *WordNet* dapat digambarkan sebagai sebuah database network semantik untuk bahasa Inggris yang dikembangkan di Princeton University. Perbedaan antara *WordNet* dengan kamus bahasa pada umumnya adalah kamus bahasa memfokuskan pada kata sedangkan *WordNet* lebih memfokuskan terhadap makna kata.

Oleh karena itu dalam penelitian ini akan dilakukan penelitian mengenai *Word Sense Disambiguation* pada kalimat berbahasa Inggris menggunakan algoritma *Simplified Lesk* dengan menggunakan *WordNet* sebagai acuan. Berdasarkan hal tersebut maka peneliti mengajukan penelitian dengan judul “***Word Sense Disambiguation Menggunakan Algoritma Lesk (Simplified Lesk) Pada Kalimat Bahasa Inggris***”.

1.3 Rumusan Masalah

Rumusan masalah dari penelitian ini adalah bagaimana cara membangun sistem penentuan makna dari kata yang ambigu (*Word Sense Disambiguation*)

menggunakan algoritma *Simplified Lesk* pada bahasa Inggris. Selanjutnya dirumuskan pertanyaan penelitian sebagai berikut:

1. Bagaimana membangun sistem perangkat lunak yang dapat menentukan makna dari kata yang ambigu menggunakan algoritma *Simplified Lesk*?
2. Bagaimana menentukan makna yang tepat dari kata yang ambigu pada kalimat bahasa Inggris yang sesuai dengan konteks kalimat menggunakan algoritma *Simplified Lesk*?
3. Bagaimana tingkat akurasi algoritma *Simplified Lesk* dalam menentukan makna kata yang sesuai dengan konteks kalimat?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menghasilkan perangkat lunak yang dapat menentukan makna dari kata yang ambigu menggunakan algoritma *Simplified Lesk*.
2. Menampilkan makna yang tepat dari kata yang ambigu pada kalimat bahasa Inggris yang sesuai dengan konteks kalimat menggunakan algoritma *Simplified Lesk*.
3. Mengetahui tingkat akurasi algoritma *Simplified Lesk* dalam menentukan makna kata yang sesuai dengan konteks kalimat.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Membantu pengguna mengetahui makna kata yang sesuai dengan konteks kalimat.
2. Hasil penelitian dapat digunakan sebagai rujukan untuk penelitian selanjutnya mengenai *Word Sense Disambiguation*.

1.6 Batasan Masalah

Batasan masalah antara lain:

1. Data yang digunakan dalam penelitian menggunakan kalimat Bahasa Inggris yang baku dan menggunakan bahasa formal.
2. *Dataset* yang digunakan berupa kalimat yang diambil dari *website* FEWS.
3. Format data masukan berupa file *csv*.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir ini mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya, antara lain:

1.7.1 Bab I. Pendahuluan

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah/ruang lingkup, dan sistematika penulisan.

1.7.2 Bab II. Kajian Literatur

Bab ini membahas seluruh dasar-dasar teori yang digunakan mulai dari definisi sistem, informasi mengenai domain, dan semua yang digunakan pada tahapan analisis, perancangan, dan implementasi.

1.7.3 Bab III. Metode Penelitian

Pada bab ini akan dibahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

1.7.4 Bab IV. Pengembangan Perangkat Lunak

Pada bab ini membahas mengenai analisis dan perancangan perangkat lunak yang akan digunakan sebagai alat penelitian. Dimulai dari pengumpulan dan analisa kebutuhan, rancangan dan konstruksi perangkat lunak serta pengujian untuk memastikan semua kebutuhan pengembangan perangkat lunak sesuai dengan kebutuhan.

1.7.5 Bab V. Hasil Dan Analisa Peneltian

Pada bab ini diuraikan hasil pengujian berdasarkan langkah-langkah yang telah direncanakan. Tabel hasil pengujian serta analisisnya disajikan sebagai basis dari kesimpulan yang akan diambil dalam penelitian ini.

1.7.6 Bab VI. Kesimpulan Dan Saran

Pada bab ini berisi kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga saran yang diberikan berdasarkan hasil dari penelitian.

1.8 Kesimpulan

Dengan demikian berdasarkan masalah yang telah diuraikan diatas akan dilakukan penelitian mengenai *Word Sense Disambiguation* pada kalimat berbahasa Inggris menggunakan algoritma *Simplified Lesk*.

DAFTAR PUSTAKA

- Basuki, S., Kholimi, A. S., Minarno, A. E., Sumadi, F. D. S., & Effendy, M. R. A. (2019). Word Sense Disambiguation (WSD) for Indonesian Homograph Word Meaning Determination by LESK Algorithm Application. In *2019 12th International Conference on Information & Communication Technology and System (ICTS)* (pp. 8–15). IEEE.
- Blevins, T., Joshi, M., & Zettlemoyer, L. (2021). FEWS: Large-Scale, Low-Shot Word Sense Disambiguation with the Dictionary. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*. Retrieved from <https://blvns.github.io/papers/eacl2021.pdf>
- Faisal, E., Nurifan, F., & Sarno, R. (2018). Word sense disambiguation in Bahasa Indonesia using SVM. In *2018 International Seminar on Application for Technology of Information and Communication* (pp. 239–243). IEEE.
- Feinerer, I., & Hornik, K. (2020). wordnet: WordNet Interface. Retrieved from <https://cran.r-project.org/package=wordnet>
- Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Bradford Books.
- Gitasari, R. M. (2007). Penghilangan Ambiguitas Makna Kata Dalam Kalimat Berbahasa Indonesia Dengan Menggunakan Parser, Wordnet Dan Algoritma Lesk.
- Gorman, K., Mazovetskiy, G., & Nikolaev, V. (2018). Improving homograph disambiguation with supervised machine learning. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation ({LREC} 2018)*. Miyazaki, Japan: European Language Resources

- Association (ELRA). Retrieved from
<https://www.aclweb.org/anthology/L18-1215>
- Kruchten, P. (2000). *The rational unified process 2nd edition: An introduction.* Reading, MA: Addison Wesley Longman, Inc.
- Kruchten, P. (2004). *The rational unified process: an introduction.* Addison-Wesley Professional.
- Lelywiary, C. J. S., Widowati, S., & Lhaksamana, K. M. (2019). Deteksi Pola Ambiguitas Struktural pada Spesifikasi Kebutuhan Perangkat Lunak menggunakan Pemrosesan Bahasa Alami. *Indonesia Journal on Computing (Indo-JC)*, 4(3), 51–64.
- Putrayasa, I. G. N. K. (2017). Ketaksaan (ambiguitas) dalam bahasa indonesia. *LUGAS Jurnal Komunikasi*, 2(2), 14.
- Putri, D. R. (2017). Word Sense Disambiguation Menggunakan Metode Lesk Untuk Cerpen Anak Berbahasa Indonesia. Universitas Komputer Indonesia.
- Sutomo, R. M. (2020). Implementasi Algoritma Naïve Bayes untuk Word Senses Disambiguation dalam Bahasa Indonesia.
- Suyuti, M. (2016). Word Sense Disambiguation dengan Algoritma Lesk (Simplified Lesk). universitas muhammadiyah jember.
- Trismanto, T. (2018). Ambiguitas Dalam Bahasa Indonesia. *Bangun Rekaprima: Majalah Ilmiah Pengembangan Rekayasa, Sosial Dan Humaniora*, 4(1, April), 42–48.
- Wallace, M. (2007). Jawbone Java WordNet API. Retrieved from
<https://sites.google.com/site/mfwallace/jawbone>