

**KEMIRIPAN SEMANTIK DOKUMEN TUGAS AKHIR TERHADAP
ONTOLOGI BIDANG ILMU INFORMATIKA MENGGUNAKAN
METODE WU PALMER**

**Diajukan Sebagai Syarat untuk Menyelesaikan
Pendidikan Program Strata-1 pada
Jurusan Teknik Informatika**



Oleh:

Vanya Terra Ardani
NIM: 09021282126076

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2025**

HALAMAN PENGESAHAN

SKRIPSI

**Kemiripan Semantik Dokumen Tugas Akhir Terhadap Ontologi Bidang
Ilmu Informatika Menggunakan Metode Wu Palmer**

Sebagai salah satu syarat untuk penyelesaian studi di

Program Studi S1 Teknik Informatika

Oleh:

VANYA TERRA ARDANI

09021282126076

Pembimbing 1 : Desty Rodiah, S.Kom., M.T.

NIP. 198912212020122011

Mengetahui

Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D

198004182020121001

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI


Pada hari Jumat tanggal 14 Maret 2025 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Vanya Terra Ardani
NIM : 09021282126076
Judul : Kemiripan Semantik Dokumen Tugas Akhir Terhadap Ontologi Bidang Ilmu Informatika Menggunakan Metode Wu Palmer

dan dinyatakan LULUS.

1. Ketua Penguji

Dian Palupi Rini, M.Kom., Ph.D.
NIP. 197802232006042002



.....

2. Penguji

Julian Supardi, S.Pd., M.T., Ph.D.
NIP. 197207102010121001



.....

3. Pembimbing

Desty Rodiah, S.Kom., M.T.
NIP. 198912212020122011



.....



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Vanya Terra Ardani
NIM : 09021282126076
Program Studi : Teknik Informatika
Judul Skripsi : Kemiripan Semantik Dokumen Tugas Akhir Terhadap
Ontologi Bidang Ilmu Informatika Menggunakan Metode
Wu Palmer

Hasil pengecekan *Software iThenticate/Turnitin*: 10%

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 Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

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



Palembang, 14 Maret 2025



METERAI
TEMPEL
99E0FAMX229945671

Vanya Terra Ardani
09021282126076

MOTTO DAN PERSEMBAHAN

“Maka sesungguhnya bersama kesulitan ada kemudahan. Maka apabila engkau telah selesai (dari suatu urusan), tetaplah bekerja keras (untuk urusan yang lain). Dan hanya kepada Tuhanmulah engkau berharap.”

(QS. Al-Insyirah: 6-7)

Ku persembahkan karya tulis ini kepada:

- Allah SWT
- Nabi Muhammad SAW
- Orang Tua dan Keluarga
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

The increasing number of multidisciplinary undergraduate theses in Informatics field presents challenges in categorizing these documents accurately. This study develops a system to measure the semantic similarity of final project documents to the ontology of Informatics using the Wu Palmer method as a solution to this problem. This method utilizes ontology tree structures and taxonomic depth to measure semantic similarity between concepts. The system employs BidangIlmuInformatika.owl as a knowledge base to represent hierarchical structures and relationships between concepts in Informatics. The classification process involves computing the similarity between document text and ontology concepts. Testing was conducted on 200 undergraduate thesis abstracts, categorized into four fields: Data Science and Pattern Recognition, Distributed Systems, Natural Language Processing, and Graphics and Visualization. The evaluation results show that the system correctly classified 105 out of 200 documents, achieving an accuracy of 53%. However, the broad scope of concepts in the Data Science and Pattern Recognition (DSPR) domain led to higher similarity values in this field, affecting the overall classification performance.

Keywords: Semantic Similarity, Undergraduate Thesis, Wu Palmer, Ontology.

ABSTRAK

Peningkatan jumlah tugas akhir yang bersifat multidisiplin di bidang Informatika menimbulkan tantangan dalam pengelompokan tugas akhir mahasiswa secara akurat. Penelitian ini mengembangkan sistem untuk mengukur kemiripan semantik dokumen tugas akhir terhadap ontologi bidang ilmu Informatika menggunakan metode Wu Palmer sebagai solusi dari masalah tersebut. Metode Wu Palmer memanfaatkan struktur pohon ontologi dan kedalaman taksonomi untuk mengukur kemiripan semantik antara dua konsep. Ontologi yang digunakan merupakan BidangIlmuInformatika.owl sebagai basis pengetahuan untuk merepresentasikan struktur hierarki dan keterkaitan antar konsep dalam bidang Informatika. Proses klasifikasi dilakukan dengan menghitung nilai *similarity* antara teks dokumen dengan representasi konsep dalam ontologi. Pengujian dilakukan pada 200 abstrak tugas akhir mahasiswa yang dikategorikan ke dalam empat bidang: *Data Science* dan Pengenalan Pola, Sistem Terdistribusi, Pemrosesan Bahasa Alami, serta Grafik dan Visualisasi. Hasil evaluasi menunjukkan bahwa sistem berhasil mengklasifikasikan 105 dari 200 dokumen dengan benar dengan tingkat akurasi 53%. Hal ini disebabkan oleh luasnya cakupan konsep dalam domain Data Science dan Pengenalan Pola (DSPR), sehingga memengaruhi kinerja klasifikasi sistem.

Kata Kunci: Kemiripan Semantik, Tugas Akhir, Wu Palmer, Ontologi.

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT atas limpahan nikmat, rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Skripsi yang berjudul “Kemiripan Semantik Dokumen Tugas Akhir Terhadap Ontologi Bidang Ilmu Informatika Menggunakan Metode Wu Palmer” ini dengan baik. Skripsi ini disusun untuk memenuhi salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya.

Dalam menyelesaikan skripsi ini, penulis telah menerima bantuan, bimbingan, serta dukungan dari banyak pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih yang tulus kepada:

1. Orang tua tercinta penulis, terutama Umi Suriyaningsih yang selalu mendoakan, mendukung, memberikan semangat, dan memotivasi penulis dalam segala aspek kehidupan. Terimakasih atas semua usaha dan perjuangan yang umi berikan sampai saat ini. Semoga umi senantiasa diberi kesehatan, rezeki, dan umur yang panjang agar bisa melihat anak-anaknya sukses, aamiin.
2. Adik-adik penulis, Anindya Petta Wardhana, Mahira Nazeefa, Syauqi Bilal Baihaqi yang penulis sayangi. Semoga kakak senantiasa bisa menjadi contoh dan panutan yang baik dalam kehidupan kalian.
3. Bapak Prof. Dr. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.

4. Bapak Hadipurnawan Satria, Ph.D. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Desty Rodiah, S.Kom., M.T. selaku dosen pembimbing akademik sekaligus pembimbing skripsi yang telah membimbing dan memberikan arahan kepada penulis selama proses kegiatan perkuliahan dan pengerjaan skripsi.
6. Seluruh Dosen, Admin, dan Staff Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Sahabat penulis, Tias Lufiani yang telah menjadi teman bercerita dan berkeluh-kesah, serta telah menemani penulis dalam segala suka dan duka semenjak SMA sampai sekarang.
8. Teman kost Lability, Allecyia Riany Allvemy, Musdalifah Salsabila, dan Zahra Aulia Rahma. Terimakasih sudah menghibur kehidupan bertetangga penulis selama di perantauan dengan berbagai cerita dan tingkah randomnya.
9. Risky Armansyah yang telah memberikan motivasi, saran, serta menjadi teman bercerita penulis selama masa pengerjaan skripsi.
10. Teman seperjuangan Project Penting, Ahmad Azhari, Dellin Irawan, Hanif Syahri Ramadhani, dan Yolendri Anisyahfitri yang telah memberi dukungan dan motivasi, serta membantu dan menemani perkuliahan dan kehidupan perantauan penulis.
11. Teman – teman Teknik Informatika angkatan 2021 yang telah menemani, memberikan motivasi, dan semangat selama pengerjaan skripsi penulis.

12. Vanya Terra Ardani, apresiasi sebesar-besarnya karena telah bertanggung jawab untuk menyelesaikan apa yang telah dimulai. Terimakasih telah bekerja keras dan berjuang sejauh ini.
13. Pihak – pihak lain yang tidak dapat penulis sebutkan satu-persatu.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dikarenakan keterbatasan

Indralaya, 14 Maret 2025

Penulis,

Vanya Terra Ardani

DAFTAR ISI

	Halaman
HALAMAN PENGESAHAN.....	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
<i>ABSTRACT</i>	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xv
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN.....	xix
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah.....	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian	I-4
1.5 Manfaat Penelitian	I-5
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan	I-6
1.8 Kesimpulan	I-7
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori.....	II-1
2.2.1 Klasifikasi Teks.....	II-1
2.2.2 Ontologi	II-1
2.2.3 <i>Text Preprocessing</i>	II-2
2.2.4 <i>Query Processing</i>	II-4
2.2.5 <i>Semantic Similarity</i>	II-6
2.2.6 <i>Confusion Matrix</i>	II-10

2.2.7	<i>Rational Unified Process</i>	II-13
2.3	Penelitian Lain yang Relevan.....	II-14
2.4	Kesimpulan	II-16
BAB III METODOLOGI 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-1
3.3	Tahapan Penelitian	III-2
3.3.1	Menganalisis Ruang Lingkup Permasalahan	III-2
3.3.2	Mempelajari Studi Literatur	III-3
3.3.3	Mengumpulkan Data	III-3
3.3.4	Menentukan Kerangka Kerja Penelitian	III-4
3.3.5	Menentukan Kriteria Pengujian	III-8
3.3.6	Menentukan Format Data Pengujian.....	III-8
3.3.7	Menentukan Alat Bantu Penelitian	III-8
3.3.8	Melakukan Pengujian Penelitian.....	III-9
3.3.9	Melakukan Analisis dan Menarik Kesimpulan Penelitian	III-9
3.4	Metode Pengembangan Perangkat Lunak.....	III-10
3.4.1	Fase Insepsi	III-10
3.4.2	Fase Elaborasi	III-11
3.4.3	Fase Konstruksi.....	III-11
3.4.4	Fase Transisi	III-11
3.5	Manajemen Proyek Penelitian	III-11
3.6	Kesimpulan	III-14
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
4.2.3	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 Text Preprocessing	IV-4
4.2.3.4	Analisis Query Processing	IV-8
4.2.3.5	Analisis Proses Klasifikasi	IV-10
4.2.3.6	Analisis Hasil Klasifikasi	IV-15
4.2.3.7	Desain Perangkat Lunak	IV-15
4.3	Fase Elaborasi	IV-21
4.3.1	Pemodelan Bisnis	IV-22
4.3.1.1	Perancangan Data	IV-22
4.3.1.2	Desain Antarmuka	IV-22
4.3.2	Kebutuhan Sistem	IV-24
4.3.3	Analisis dan Perancangan	IV-25
4.3.3.1	Activity Diagram	IV-25
4.3.3.2	Sequence Diagram	IV-29
4.4	Fase Konstruksi	IV-32
4.4.1	Kebutuhan Sistem	IV-32
4.4.2	<i>Class Diagram</i>	IV-33
4.4.3	Implementasi	IV-33
4.4.3.1	Implementasi Kelas	IV-34
4.4.3.2	Implementasi Antarmuka	IV-36
4.5	Fase Transisi	IV-39
4.5.1	Pemodelan Bisnis	IV-39
4.5.2	Rencana Pengujian	IV-39
4.5.3	Implementasi	IV-40
4.6	Kesimpulan	IV-44
BAB V HASIL DAN ANALISIS PENELITIAN		V-1
5.1	Pendahuluan	V-1
5.2	Hasil Penelitian	V-1
5.2.1	Konfigurasi Pengujian	V-1
5.2.2	Data Hasil Konfigurasi	V-2
5.3	Analisis Hasil Penelitian	V-9
5.4	Kesimpulan	V-11

BAB VI KESIMPULAN DAN SARAN	VI-1
6.1 Pendahuluan	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran.....	VI-2
DAFTAR PUSTAKA	xvii
LAMPIRAN.....	xxii

DAFTAR TABEL

	Halaman
Tabel II-1. Contoh Case Folding	II-3
Tabel II-2. Contoh Tokenizing	II-3
Tabel II-3. Contoh Stopword Removal.....	II-4
Tabel II-4. Contoh Stemming	II-4
Tabel II-5. Contoh Penambahan Sinonim WordNet.....	II-5
Tabel II-6. Contoh Remove Duplicate Query	II-5
Tabel III-1. Contoh Data Penelitian.....	III-3
Tabel III-2. Format Hasil Pengujian	III-8
Tabel III-3. Alat Bantu Penelitian	III-8
Tabel III-4. Rancangan Tabel Confusion Matrix	III-10
Tabel III-5. Rancangan Tabel Perhitungan Hasil Kinerja	III-10
Tabel III-6. Manajemen Proyek Penelitian.....	III-11
Tabel IV-1. Kebutuhan Fungsional	IV-2
Tabel IV-2. Kebutuhan Non-Fungsional.....	IV-3
Tabel IV-3. Sampel Dataset Tugas Akhir	IV-4
Tabel IV-4. Data Abstrak Tugas Akhir.....	IV-5
Tabel IV-5. Hasil Case Folding.....	IV-6
Tabel IV-6. Hasil Tokenizing	IV-7
Tabel IV-7. Hasil Stopword Removal.....	IV-7
Tabel IV-8. Hasil Stemming	IV-8
Tabel IV-9. Hasil Penambahan Sinonim WordNet.....	IV-9
Tabel IV-10. Hasil Remove Duplicate Query	IV-9
Tabel IV-11. Hasil Pencarian Vocab Ontologi pada Teks Dokumen	IV-11
Tabel IV-12. Hasil Perhitungan Nilai Similarity	IV-14
Tabel IV-13. Hasil Klasifikasi	IV-14
Tabel IV-14. Definisi Aktor	IV-16
Tabel IV-15. Definisi Use Case	IV-16
Tabel IV-16. Use Case Scenario Upload Dataset	IV-17

Tabel IV-17. Use Case Scenario Klasifikasi Berbasis Ontologi Menggunakan Semantic Similarity Wu Palmer.....	IV-19
Tabel IV-18. Use Case Scenario Klasifikasi dengan Input dari Pengguna.....	IV-20
Tabel IV-19. Daftar Library Python.....	IV-32
Tabel IV-20. Implementasi Kelas	IV-34
Tabel IV-21. Rencana Pengujian Use Case Upload Dataset.....	IV-39
Tabel IV-22. Rencana Pengujian Use Case Klasifikasi Dokumen Berbasis Ontologi Menggunakan Semantic Similarity Wu Palmer	IV-40
Tabel IV-23. Rencana Pengujian Use Case Klasifikasi dengan Input Teks dari Pengguna.....	IV-40
Tabel IV-24. Hasil Pengujian Use Case Upload Dataset.....	IV-41
Tabel IV-25. Hasil Pengujian Use Case Klasifikasi Dokumen Berbasis Ontologi Menggunakan Semantic Similarity Wu Palmer	IV-42
Tabel IV-26. Hasil Pengujian Use Case Klasifikasi dengan Input Teks dari Pengguna.....	IV-43
Tabel V-1. Hasil Pengujian	V-2
Tabel V-2. Confusion Matrix Hasil Pengujian.....	V-9
Tabel V-3. Hasil Kinerja Keseluruhan Sistem	V-9

DAFTAR GAMBAR

	Halaman
Gambar II-1. Contoh Konsep Hierarki Wu Palmer (Shenoy et al., 2012).....	II-8
Gambar II-2. Tabel Confusion Matrix (Artama et al., 2020).....	II-11
Gambar II-3. Arsitektur RUP (Perwitasari et al., 2020)	II-13
Gambar III-1. Rincian Tahapan Penelitian	III-2
Gambar III-2. Kerangka Kerja Penelitian.....	III-4
Gambar III-3. Hierarki Ontologi Bidang Informatika (Rodiah et al., 2024) ...	III-6
Gambar III-4. Hasil Implementasi Class dan Subclass Ontologi Domain Informatika pada Protégé (Rodiah et al., 2024)	III-7
Gambar IV-1. Diagram Use Case.....	IV-15
Gambar IV-2. Rancangan Antarmuka Halaman Dataset Classification	IV-22
Gambar IV-3. Rancangan Antarmuka Setelah Upload Dataset	IV-23
Gambar IV-4. Rancangan Antarmuka Setelah Klasifikasi Dataset Selesai ...	IV-23
Gambar IV-5. Rancangan Antarmuka Halaman Manual Input Classification....	IV-24
Gambar IV-6. Rancangan Antarmuka Setelah Klasifikasi Input Manual Selesai	IV-24
Gambar IV-7. Activity Diagram Upload Dataset.....	IV-26
Gambar IV-8. Activity Diagram Klasifikasi Dokumen Berbasis Ontologi Menggunakan Semantic Similarity Wu Palmer	IV-27
Gambar IV-9. Activity Diagram Klasifikasi dengan Input Teks dari Pengguna.	IV-28
Gambar IV-10. Sequence Diagram Upload Dataset	IV-29
Gambar IV-11. Sequence Diagram Klasifikasi Dokumen Berbasis Ontologi Menggunakan Semantic Similarity Wu Palmer	IV-30
Gambar IV-12. Sequence Diagram Klasifikasi dengan Input Teks dari Pengguna	IV-31
Gambar IV-13. Class Diagram.....	IV-33
Gambar IV-14. Antarmuka Halaman Dataset Classification	IV-36

Gambar IV-15. Antarmuka Dataset Classification Setelah Upload Dataset (1)..	IV-36
Gambar IV-16. Antarmuka Dataset Classification Setelah Upload Dataset (2)..	IV-37
Gambar IV-17. Antarmuka Dataset Classification Setelah Proses Klasifikasi (1)	IV-37
Gambar IV-18. Antarmuka Dataset Classification Setelah Proses Klasifikasi (2)	IV-37
Gambar IV-19. Antarmuka Halaman Manual Input Classification	IV-38
Gambar IV-20. Antarmuka Manual Input Classification Setelah Input Teks	IV-38
Gambar IV-21. Antarmuka Manual Input Classification Setelah Proses Klasifikasi	IV-38

DAFTAR LAMPIRAN

1. Kode Program
2. Surat Keterangan Validasi Dataset
3. Sample Dataset
4. Teks Dokumen
5. Hasil Case Folding
6. Hasil Tokenizing
7. Hasil Stopword Removal
8. Hasil Stemming
9. Hasil Penambahan Sinonim WordNet
10. Hasil Remove Duplicate Query
11. Hasil Pencarian Query pada Ontologi
12. Hasil Perhitungan Nilai Similarity

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini akan menjelaskan gagasan yang menjadi latar belakang penelitian. Pembahasan mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan skripsi akan dijelaskan pada masing-masing subbab pada bab ini.

1.2 Latar Belakang Masalah

Salah satu syarat bagi mahasiswa untuk memperoleh gelar sarjana atau diploma adalah dengan menyelesaikan tugas akhir, yang umumnya dalam bentuk skripsi sebagai laporan penelitian. Bertambahnya jumlah mahasiswa yang lulus dari perguruan tinggi, berbanding lurus dengan jumlah koleksi skripsi yang terus meningkat. Hal tersebut menimbulkan beberapa kesulitan dalam pengelolaan dan pengelompokan skripsi, terlebih jika dilakukan secara manual (Lestari et al., 2024). Dengan bertambahnya jumlah skripsi, mencari informasi yang spesifik dan relevan menjadi semakin sulit, terutama pada bidang Informatika yang sering melibatkan berbagai disiplin ilmu. Semakin banyak penelitian yang bersifat multidisiplin, semakin sulit pula untuk mengategorikan rumpun ilmu Informatika secara tepat (Rodiah et al., 2024), sehingga pengelompokan yang akurat menjadi tantangan. Di Jurusan Teknik Informatika Universitas Sriwijaya, setiap dosen memiliki bidang keilmuan tertentu, dan mahasiswa perlu memilih dosen pembimbing yang sesuai dengan topik skripsinya. Namun, tidak jarang mahasiswa mengalami kesulitan

dalam menentukan topik yang tepat dan keliru dengan memilih dosen pembimbing yang tidak sesuai dengan bidang keilmuannya. Oleh karena itu, diperlukan suatu sistem yang mampu mengukur kemiripan semantik dokumen tugas akhir terhadap bidang keilmuan yang telah ditentukan. Salah satu pendekatan yang dapat dilakukan adalah dengan membuat sistem klasifikasi dokumen.

Klasifikasi dokumen bertujuan untuk mengelompokkan dokumen secara otomatis berdasarkan isi dokumen ke dalam kategori yang telah diketahui sebelumnya (Utomo & Rainarli, 2019). Klasifikasi dokumen tersebut dapat dilakukan dengan metode ontologi yang tidak memerlukan data latih, sehingga dapat mengatasi permasalahan akan kebutuhan data latih pada proses klasifikasi (Lestari et al., 2024). Klasifikasi dokumen dapat digunakan sebagai metode evaluasi untuk melihat seberapa baik sistem dalam mengelompokkan dokumen tugas akhir berdasarkan kemiripan semantik terhadap ontologi bidang ilmu yang telah ditentukan. Dalam konteks teknologi web semantik, ontologi merupakan elemen utama yang berfungsi untuk menggambarkan konsep, menghubungkan entitas, dan mengategorikan berbagai objek (Lan et al., 2022). Ontologi menyimpan pengetahuan semantik dalam format yang dapat dipahami oleh mesin dan merepresentasikan pengetahuan suatu domain menggunakan kosakata yang terstruktur (Behr et al., 2023).

Penelitian terdahulu oleh Ayu, dkk. membahas klasifikasi teks berbasis ontologi untuk dokumen tugas akhir berbahasa Indonesia. Penelitian tersebut menggunakan 500 data abstrak dari Tugas Akhir Teknik Informatika yang diperoleh dari *Digital Library* (DIGILIB) Universitas Muhammadiyah Malang

tahun 2014-2016. Tahap yang dilakukan adalah membangun dan memodelkan rule ontologi berdasarkan *Association for Computing Machinery (ACM) IEEE Computer Society*. Untuk menghitung kemiripan antar dokumen dari sebuah node ontologi, digunakan metode Dao. Hasil penelitian menunjukkan akurasi sebesar 87%, namun terdapat beberapa kendala berupa ketidaksesuaian topik pada sejumlah abstrak serta adanya beberapa abstrak yang teridentifikasi dengan lebih dari satu klasifikasi. Masalah tersebut disebabkan oleh tidak adanya relasi kesamaan antar kata dengan istilah pada ontologi dan nilai Dao yang sama pada ontologi. Hal ini dapat menyebabkan kesalahan klasifikasi, karena sistem tidak dapat membedakan antara dua konsep yang memiliki nilai Dao yang sama. Oleh karena itu, penelitian ini bertujuan untuk mengatasi masalah nilai Dao tersebut menggunakan metode pengukuran *Semantic Similarity*.

Semantic Similarity adalah sebuah metode untuk mengukur tingkat kemiripan makna antara dua kalimat atau kata (Abriani & Yaqin, 2019). Metode *semantic similarity* atau kemiripan semantik biasanya memberikan peringkat atau persentase kesamaan antara teks (Chandrasekaran & Mago, 2020). *Semantic similarity* terbagi menjadi beberapa jenis, yaitu *term similarity*, *sentence similarity*, *entity similarity*, *document similarity*, dan *word similarity* (Styawati et al., 2020).

Salah satu metode *semantic similarity* yang populer adalah algoritma Wu Palmer (WUP). Wu Palmer merupakan algoritma kemiripan semantik yang mampu mengukur derajat keterkaitan atau relevansi antar dokumen ataupun antar *term* (Abriani & Yaqin, 2019). Berdasarkan penelitian yang dilakukan oleh (Dilekh et al., 2023), metode Wu Palmer (WUP) menghasilkan skor kemiripan terbaik dengan

nilai korelasi tertinggi (0.82) dan *Mean Squared Error* (MSE) terendah dibandingkan dengan metode Path, Leacock & Chodorow (LCH), Resnik, dan Lin, dalam menentukan kemiripan kata di English WordNet (WN). Selain itu, Wu-Palmer juga terbukti mampu bersaing dengan metode *embedding* modern, terutama dalam pengelompokan teks berdasarkan *Semantic Similarity* (Marcínczuk et al., 2021).

Berdasarkan penjelasan di atas, penelitian mengenai pengukuran kemiripan semantik dokumen tugas akhir terhadap ontologi bidang ilmu informatika ini akan menggunakan metode pengukuran *Semantic Similarity* Wu Palmer.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, didapatkan rumusan masalah sebagai berikut:

1. Bagaimana mengembangkan sebuah sistem untuk mengukur kemiripan semantik dokumen tugas akhir terhadap ontologi bidang ilmu Informatika menggunakan metode Wu Palmer?
2. Bagaimana kinerja metode pengukuran *Semantic Similarity* Wu Palmer dalam mengukur kemiripan semantik dokumen tugas akhir terhadap ontologi bidang ilmu Informatika?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menghasilkan sebuah sistem untuk mengukur kemiripan semantik dokumen tugas akhir terhadap ontologi bidang ilmu Informatika menggunakan metode Wu Palmer.
2. Mengetahui kinerja metode pengukuran *Semantic Similarity* Wu Palmer dalam mengukur kemiripan semantik dokumen tugas akhir terhadap ontologi bidang ilmu Informatika.

1.5 Manfaat Penelitian

Berikut adalah manfaat dari penelitian yang dapat diperoleh :

1. Memperdalam pemahaman dan keterampilan dalam pengembangan sistem untuk mengukur kemiripan semantik dokumen tugas akhir terhadap ontologi bidang ilmu Informatika, terutama pada penerapan metode pengukuran *Semantic Similarity* Wu Palmer.
2. Memberikan kontribusi kepada pihak terkait dan masyarakat akademik dalam proses pengelompokan dokumen tugas akhir melalui sistem yang dihasilkan, sehingga dapat mempercepat akses terhadap informasi relevan di bidang penelitian dan pendidikan.

1.6 Batasan Masalah

Adapun batasan-batasan masalah yang akan didefinisikan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan dokumen tugas akhir berbahasa Indonesia dari program studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

2. Analisis yang dilakukan hanya berfokus pada abstrak tugas akhir, tanpa mempertimbangkan keseluruhan isi dokumen.
3. Analisis yang dilakukan hanya berfokus pada kemiripan semantik antara dokumen dengan ontologi.
4. Penelitian ini menggunakan metode pengukuran *Semantic Similarity* Wu Palmer untuk mengukur kemiripan antara dokumen dengan ontologi.

1.7 Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut.

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan penelitian ini.

BAB II. KAJIAN LITERATUR

Bab ini membahas landasan teori yang digunakan dalam penelitian, seperti definisi ontologi, perhitungan *Semantic Similarity*, metode Wu Palmer, serta beberapa literatur yang relevan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab ini membahas mengenai tahapan yang akan dilaksanakan pada penelitian. Masing-masing rencana tahapan penelitian akan dideskripsikan dengan kerangka kerja sebagai acuannya.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini membahas proses analisis dan perancangan perangkat lunak yang dikembangkan. Pembahasan dimulai dengan analisis kebutuhan, perancangan serta

implementasi perangkat lunak, dan diakhiri dengan pengujian untuk memastikan bahwa sistem yang dikembangkan telah sesuai dengan rancangan serta kebutuhan penelitian.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini menyajikan hasil pengujian yang diperoleh berdasarkan langkah-langkah yang telah dirancang sebelumnya. Analisis terhadap hasil pengujian digunakan sebagai dasar dalam penarikan kesimpulan dari penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab ini memaparkan kesimpulan yang diperoleh berdasarkan pembahasan pada bab-bab sebelumnya. Selain itu, saran juga diberikan berdasarkan hasil penelitian yang telah dilakukan.

1.8 Kesimpulan

Bab pendahuluan ini telah menjelaskan latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penelitian yang menjadi dasar pemikiran peneliti dalam pengembangan sistem untuk mengukur kemiripan semantik dokumen tugas akhir terhadap ontologi bidang ilmu Informatika menggunakan metode Wu Palmer.

DAFTAR PUSTAKA

- Abriani, G. U., & Yaqin, M. A. (2019). Implementasi Metode Semantic Similarity untuk Pengukuran Kemiripan Makna antar Kalimat. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, 1(2), 47–57. <https://doi.org/10.28926/ilkomnika.v1i2.15>
- Alian, M., & Awajan, A. (2020). Semantic Similarity for English and Arabic Texts: A Review. *Journal of Information & Knowledge Management*, 19(04), 2050033. <https://doi.org/10.1142/S0219649220500331>
- Amin, M. F. (2022). Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial. *Journal of Engineering Research*, 6(5).
- Amin, M. F. (2023). Confusion Matrix in Three-class Classification Problems: A Step-by-Step Tutorial. *Journal of Engineering Research*, 7(1), 0–0. <https://doi.org/10.21608/erjeng.2023.296718>
- Anandarajan, M., Hill, C., & Nolan, T. (2019). *Text Preprocessing* (pp. 45–59). https://doi.org/10.1007/978-3-319-95663-3_4
- Artama, M., Sukajaya, I. N., & Indrawan, G. (2020). Classification of official letters using TF-IDF method. *Journal of Physics: Conference Series*, 1516(1), 012001. <https://doi.org/10.1088/1742-6596/1516/1/012001>
- Behr, A. S., Völkenrath, M., & Kockmann, N. (2023). Ontology extension with NLP-based concept extraction for domain experts in catalytic sciences.

Knowledge and Information Systems, 65(12), 5503–5522.
<https://doi.org/10.1007/s10115-023-01919-1>

Chandrasekaran, D., & Mago, V. (2020). *Evolution of Semantic Similarity -- A Survey*. <https://doi.org/10.1145/3440755>

Dilekh, T., Boulahia, M. A., & Benharzallah, S. (2023). Assessing Semantic Similarity Measures and Proposing a WuP-Resnik Hybrid Metric for Enhanced Arabic Language Processing. *Revue d'Intelligence Artificielle*, 37(5), 1311–1322. <https://doi.org/10.18280/ria.370524>

Fellah, A., Zahaf, A., & Elçi, A. (2024). Semantic Similarity Measure Using a Combination of Word2Vec and WordNet Models. *Indonesian Journal of Electrical Engineering and Informatics (IJEI)*, 12(2). <https://doi.org/10.52549/ijeei.v12i2.5114>

Harispe, S., Ranwez, S., Janaqi, S., & Montmain, J. (2017). *Semantic Similarity from Natural Language and Ontology Analysis*. <https://doi.org/10.2200/S00639ED1V01Y201504HLT027>

Kaban, R., Sihombing, P., Pandia, M., & Simamora, P. (2023). Pemrosesan Query dan Pemingkatan Hasil dalam Information Retrieval: Sebuah Kajian Literatur. *Journal of Information System Research (JOSH)*, 4(3), 748–754. <https://doi.org/10.47065/josh.v4i3.2867>

Khairunnisa, S., Adiwijaya, & Al Faraby, S. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter

- (Studi Kasus Pandemi COVID-19). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), 406. <https://doi.org/10.30865/mib.v5i2.2835>
- Lan, G., Liu, T., Wang, X., Pan, X., & Huang, Z. (2022). A semantic web technology index. *Scientific Reports*, 12(1), 3672. <https://doi.org/10.1038/s41598-022-07615-4>
- Lestari, A. P., Maskur, & Hayatin, N. (2024). Klasifikasi Teks Berbasis Ontologi Untuk Dokumen Tugas Akhir Berbahasa Indonesia. *Jurnal Repositor*, 1(2). <https://doi.org/10.22219/repositor.v1i2.30328>
- Marcínczuk, M., Gniewkowski, M., & Walkowiak, T. (2021). *Text Document Clustering: Wordnet vs. TF-IDF vs. Word Embeddings*. <https://github.com/CLARIN-PL/wnsim>
- Mihindukulasooriya, N., Tiwari, S., Enguix, C. F., & Lata, K. (2023). *Text2KGBench: A Benchmark for Ontology-Driven Knowledge Graph Generation from Text*. <http://arxiv.org/abs/2308.02357>
- Orosoo, M., Govindasamy, S., Bayarsaikhan, N., Rajkumari, Y., Fatma, G., Manikandan, R., & Kiran Bala, B. (2023). Performance analysis of a novel hybrid deep learning approach in classification of quality-related English text. *Measurement: Sensors*, 28, 100852. <https://doi.org/10.1016/j.measen.2023.100852>
- Perwitasari, R., Afawani, R., & Anjarwani, S. E. (2020). Penerapan Metode Rational Unified Process (RUP) Dalam Pengembangan Sistem Informasi Medical Check Up Pada Citra Medical Centre. *Jurnal Teknologi Informasi*,

Komputer, Dan Aplikasinya (JTIKA), 2(1), 76–88.
<https://doi.org/10.29303/jtika.v2i1.85>

Qiu, J., Qi, L., Wang, J., & Zhang, G. (2018). A hybrid-based method for Chinese domain lightweight ontology construction. *International Journal of Machine Learning and Cybernetics*, 9(9), 1519–1531. <https://doi.org/10.1007/s13042-017-0661-0>

Rahayu, K., Fitria, V., Septhya, D., Rahmaddeni, & Efrizoni, L. (2023). Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis Machine Learning. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 108–114. <https://doi.org/10.57152/malcom.v3i2.780>

Rodiah, D., Miraswan, K. J., Kurniati, J., Irawan, D., & Ardani, V. T. (2024). Pengembangan Representasi Pengetahuan Ontologi Domain Bidang Ilmu Informatika. *Jurnal PROCESSOR*, 19(2). <https://doi.org/10.33998/processor.2024.19.2.1905>

Rosid, M. A., Fitriani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi. *IOP Conference Series: Materials Science and Engineering*, 874(1), 012017. <https://doi.org/10.1088/1757-899X/874/1/012017>

- Shenoy, M., Shet, K. C., & Acharya, U. D. (2012). A New Similarity Measure for Taxonomy Based on Edge Counting. *International Journal of Web & Semantic Technology*, 3(4), 23–30. <https://doi.org/10.5121/ijwest.2012.3403>
- Styawati, Yulita, W., & Sarasvananda, I. B. G. (2020). Survey Ukuran Kesamaan Semantic Antar Kata. In *JDMSI* (Vol. 1, Issue 5).
- Utomo, R. N. P., & Rainarli, E. (2019). *Klasifikasi Dokumen Menggunakan Metode Extreme Learning Machines (Elm) Pada Dokumen Berbahasa Indonesia*.
- Wu, Z., & Palmer, M. (1994). *Verb Semantics And Lexical Selection*.
- Yahya, L., & Mulyati, M. (2022). Rancang Bangun Sistem Informasi Menggunakan Metode Rational Unified Process (Rup) Pada Salon Mobil Scuto. *Klik - Jurnal Ilmu Komputer*, 3(2), 101–106. <https://doi.org/10.56869/klik.v3i2.383>
- Yunita. (2017). *Pemodelan Ontologi Web Semantik pada Pencarian Lowongan Pekerjaan Berdasarkan Profil Pencari Kerja*.
- Zulkifli, U. C., & Suadaa, L. H. (2019). *Pengembangan Modul PreprocessingTeks untuk Kasus Formalisasi dan Pengecekan Ejaan Bahasa Indonesia pada Aplikasi Web Mining Simple Solution (WMSS)* (Vol. 15, Issue 2).