

**KOMPUTASI KLASIFIKASI PADA BERITA *ONLINE* BERBAHASA ARAB
DENGAN PENDEKATAN SEMANTIK BERBASIS *FRAME***



TESIS

Diajukan Kepada Fakultas Adab dan Ilmu Budaya
UIN Sunan Kalijaga untuk Memenuhi Syarat Memperoleh Gelar
Magister Humaniora (M. Hum.)

Disusun Oleh:

Maulana Ihsan Ahmad

NIM: 23201011009

PROGRAM STUDI MAGISTER BAHASA DAN SASTRA ARAB
FAKULTAS ADAB DAN ILMU BUDAYA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA

2024

MOTTO

Nikmati prosesnya rasakan sensasinya.

- Maulana Ihsan Ahmad



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

Tesis ini saya persembahkan untuk semua orang yang sedang berjuang untuk tetap melanjutkan hidupnya.



PENGESAHAN TUGAS AKHIR

Nomor : B-2231/Un.02/DA/PP.00.9/11/2024

Tugas Akhir dengan judul : KOMPUTASI KLASIFIKASI PADA BERITA ONLINE BERBAHASA ARAB
DENGAN PENDEKATAN SEMANTIK BERBASIS FRAME

yang dipersiapkan dan disusun oleh:

Nama : MAULANA IHSAN AHMAD, S.Kom., S.Hum
Nomor Induk Mahasiswa : 23201011009
Telah diujikan pada : Kamis, 14 November 2024
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Adab dan Ilmu Budaya UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Dr. Moh. Kanif Anwari, S.Ag. M.Ag.
SIGNED

Valid ID: 673f0dbaaefed



Penguji I

Dr. Andi Holilulloh, S.Pd.I., M.A.
SIGNED

Valid ID: 67394c5dee0d2



Penguji II

Prof. Dr. H. Ibnu Burdah, S.Ag. M.A.
SIGNED

Valid ID: 673f0a69ad5df



Yogyakarta, 14 November 2024

UIN Sunan Kalijaga

Dekan Fakultas Adab dan Ilmu Budaya

Prof. Dr. Nurdin, S.Ag., S.S., M.A.
SIGNED

Valid ID: 674430cf0c42b

SURAT PERNYATAAN KEASLIAN TESIS

Yang bertanda tangan di bawah ini:

Nama : Maulana Ihsan Ahmad

NIM : 23201011009

Program Studi : Magister Bahasa dan Sastra Arab

Fakultas : Adab dan Ilmu Budaya

Menyatakan bahwa:

1. Tesis yang berjudul “Komputasi Klasifikasi pada Berita *Online* Berbahasa Arab dengan Pendekatan Semantik Berbasis *Frame*” merupakan hasil karya asli saya yang diajukan untuk memenuhi salah satu persyaratan memperoleh gelas magister (S2) di Program Studi Magister Bahasa dan Sastra Arab, Fakultas Adab dan Ilmu Budaya, UIN Sunan Kalijaga Yogyakarta.
2. Semua sumber yang saya gunakan dalam penulisan tesis ini telah saya cantumkan sesuai dengan ketentuan yang berlaku. Jika di kemudian hari terbukti bahwa karya ini bukan hasil karya asli saya atau hasil plagiat dari karya orang lain, maka saya bersedia menerima sanksi yang berlaku.

Yogyakarta, 18 Oktober 2024

Yang menyatakan,

The image shows a red official stamp of UIN Sunan Kalijaga Yogyakarta. To the right of the stamp is a handwritten signature in black ink. Below the stamp, the text 'METERAL TEMPEL' and the number '9EAMX032515345' are visible.

Maulana Ihsan Ahmad

NIM: 23201011009

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama : Maulana Ihsan Ahmad

NIM : 23201011009

Program Studi : Magister Bahasa dan Sastra Arab

Fakultas : Adab dan Ilmu Budaya

Menyatakan dengan sesungguhnya bahwa tesis ini secara keseluruhan adalah murni karya penulis sendiri dan bukan plagiasi sebagian atau keseluruhan dari karya orang lain, kecuali pada bagian-bagian yang dirujuk sebagai sumber pustaka sesuai dengan aturan penulisan yang berlaku. Apabila di kemudian hari terbukti bahwa tesis penulis ini merupakan plagiasi karya orang lain, penulis sanggup menerima sanksi akademik dari dosen yang bersangkutan. Demikian atas perhatiannya disampaikan terima kasih.

Yogyakarta, 18 Oktober 2024

Penulis



20
METERAI
TEMPEL
BCCAMX032515350

Maulana Ihsan Ahmad
NIM: 23201011009

NOTA DINAS PEMBIMBING

Kepada Yth.

Dekan Fakultas Adab dan Ilmu Budaya

UIN Sunan Kalijaga

Yogyakarta

Assalamu'alaikum wr. wb.

Setelah melakukan bimbingan, arahan, dan koreksi terhadap naskah tesis berjudul: *Komputasi Klasifikasi pada Berita Online Berbahasa Arab dengan Pendekatan Semantik Berbasis Frame*

Yang ditulis oleh:

Nama : Maulana Ihsan Ahmad

NIM : 23201011009

Program Studi : Magister Bahasa dan Sastra Arab

Sudah dapat diajukan sebagai salah satu syarat untuk memperoleh gelar Magister dalam Program Studi Magister Bahasa dan Sastra Arab Fakultas Adab dan Ilmu Budaya UIN Sunan Kalijaga.

Wassalamu'alaikum wr. wb.

Yogyakarta, 18 Oktober 2024

Dosen Pembimbing


Dr. Moh. Kanif Anwari, S.Ag. M.Ag.
NIP: 197107301996031002

PEDOMAN TRANSLITERASI ARAB-LATIN

Berdasarkan Surat Keputusan Bersama Menteri Agama RI dan Menteri Pendidikan dan Kebudayaan RI Nomor 158/1987 dan 0534b/U/1897, tanggal 22 Januari 1998.

A. Konsonan Tunggal

Huruf Arab	Nama	Huruf Latin	Keterangan
ا	Alif	Tidak dilambangkan	Tidak dilambangkan
ب	Ba'	B	Be
ت	Ta'	T	Te
ث	Ša	Š	Es (dengan titik di atas)
ج	Jim	J	Je
ح	Ḥa	Ḥ	Ha (dengan titik di bawah)
خ	Kha	Kh	Ka dan ha
د	Dal	D	De
ذ	Žal	Ž	Zet (dengan titik di atas)
ر	Ra'	R	Er
ز	Zai	Z	Zet
س	Sin	S	Es
ش	Syin	Sy	Es dan ye
ص	Šad	Š	Es (dengan titik di bawah)
ض	Ḍad	Ḍ	De (dengan titik di bawah)
ط	Ṭa'	Ṭ	Te (dengan titik di bawah)
ظ	Ẓa'	Ẓ	Zet (dengan titik di bawah)
ع	‘ain	‘	Koma terbalik di atas
غ	Gain	G	Ge
ف	Fa'	F	Ef
ق	Qaf	Q	Qi

ك	Kaf	K	Ka
ل	Lam	L	‘el
م	Mim	M	‘em
ن	Nun	N	‘en
و	Wawu	W	We
هـ	Ha’	H	Ha
ء	Hamzah	‘	Apostrof
ي	Ya’	Y	Ye

B. Konsonan Rangkap Karena Syaddah Ditulis Rangkap

متعددين	Ditulis	Muta’addidīn
عدة	Ditulis	’iddah

C. Ta’ Marbutah

1. Bila dimatikan ditulis h

هبة	Ditulis	Hibbah
جزية	Ditulis	Jizyah

(Ketentuan ini tidak diperlakukan terhadap kata-kata Arab yang sudah terserap ke dalam bahasa Indonesia, seperti salat dan zakat, kecuali bila dikehendaki lafal aslinya). Bila diikuti dengan sandang “al” dan bacaan kedua terpisah, maka ditulis dengan h.

كرامة الأولياء	Ditulis	Karāmah al-Auliyyā’
----------------	---------	---------------------

2. Bila ta’ marbutah hidup dengan harākat fathah, kasrah dan dammah ditulis t.

زكاة الفطر	Ditulis	Zakātul Fiṭr
------------	---------	--------------

D. Vokal Pendek

َ	Fathah	Ditulis	A
ِ	Kasrah	Ditulis	I
ُ	Dammah	Ditulis	U

E. Vokal Panjang

Fathah + alif جاهلية	Ditulis	Ā Jāhiliyyah
Fathah + ya' mati يسعى	Ditulis	Ā Yas'ā
Kasrah + ya' mati كريم	Ditulis	Ī Karīm
Dammah + wawu mati فروض	Ditulis	Ū Furūd

F. Vokal Rangkap

Fathah + ya' mati بينكم	Ditulis	Ai Bainakum
Fathah + wawu mati قول	Ditulis	Au Qaul

G. Vokal Pendek yang Berurutan dalam Satu Kata Dipisahkan dengan Apostrof

أأنتم	Ditulis	A antum
أأعدت	Ditulis	U'iddat
لألن شكرتم	Ditulis	La'in Syakartum

H. Kata Sandang Alif + Lam

1. Bila diikuti huruf qamariyah

القرآن	Ditulis	Al-Qur'ān
القياس	Ditulis	Al-Qiyās

2. Bila diikuti huruf syamsiyah ditulis dengan menggandakan huruf syamsiyah yang mengikutinya, serta menghilangkan huruf l (el)-nya.

السماء	Ditulis	As-samā'
الشمس	Ditulis	Asy-Syams

I. Penulisan Kata-Kata dalam Rangkaian Kalimat

ذوي الفروض	Ditulis	Zawī al-Furūd
أهل السنة	Ditulis	Ahl al-Sunnah

KATA PENGANTAR

Dengan memanjatkan puja dan puji syukur kehadiratan Allah Swt. yang telah melimpahkan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan penelitian ini sebagai salah satu syarat untuk menyelesaikan Program magister program studi Magister Bahasa dan Sastra Arab Fakultas Adab Ilmu Budaya Universitas Islam Negeri Sunan Kalijaga Yogyakarta.

Penulis menyadari bahwa tesis ini tidak mungkin terselesaikan tanpa adanya dukungan, bantuan, bimbingan, dan nasehat dari berbagai pihak selama penyusunan tesis ini. Pada kesempatan ini penulis menyampaikan terima kasih setulus-tulusnya kepada:

1. Prof. Noorhaidi, M.A., M.Phil., Ph.D. selaku Rektor Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
2. Prof. Dr. Nurdin, S.Ag., S.S., M.A. selaku Dekan Fakultas Adab Ilmu Budaya Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
3. Dr. Andi Holilulloh, S.Pd.I., M.A. selaku Ketua Program Studi Magister Bahasa dan Sastra Fakultas Adab dan Ilmu Budaya UIN Sunan Kalijaga Yogyakarta yang telah memberikan bimbingan dan kesempatan mengembangkan potensi diri kepada kami.
4. Prof. Dr. H. Mohammad Pribadi, M.A. M.Si. selaku dosen penasihat akademik selama berkuliah di Magister BSA UIN Sunan Kalijaga.
5. Dr. Moh. Kanif Anwari, S.Ag. M.Ag. selaku dosen pembimbing tesis. Terimakasih atas segala bimbingan, arahan serta saran yang diberikan kepada penulis sehingga tesis ini dapat diselesaikan dengan baik.
6. Seluruh dosen program studi Magister Bahasa dan Sastra Arab Fakultas Adab Ilmu Budaya yang telah memberikan ilmu pengetahuan yang tak ternilai selama penulis menempuh pendidikan di Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
7. Ayah dan ibu yang telah memberikan semua sampai saat ini dan seterusnya.
8. Teman dan sahabat dari Sabang sampai Merauke yang sudah memberikan dukungan saya dalam menyelesaikan skripsi ini.

ABSTRAK

Penelitian ini berfokus pada pengembangan metode klasifikasi otomatis untuk berita *online* berbahasa Arab dengan menggunakan pendekatan semantik berbasis *frame*. Latar belakang masalah penelitian ini didasarkan pada kebutuhan untuk meningkatkan efisiensi dan akurasi klasifikasi berita dalam bahasa Arab, yang dikenal memiliki struktur bahasa yang kompleks serta kaya akan morfologi. Tantangan utama dalam klasifikasi berita berbahasa Arab adalah bagaimana memahami konteks linguistik yang mendalam dan menyusun kategori berita dengan tepat. Penelitian ini bertujuan untuk mengeksplorasi penerapan teori semantik *frame* dalam proses klasifikasi berita *online* berbahasa Arab, guna meningkatkan pemahaman makna berita dan pengelompokan kategori berita secara otomatis. Metode penelitian ini menggunakan teknik web *scraping* untuk mengumpulkan data dari berbagai portal berita Arab dan memanfaatkan algoritma pembelajaran mesin, seperti *Support Vector Machine* (SVM), yang didukung oleh vektorisasi *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk ekstraksi fitur. Data yang dikumpulkan terdiri dari 170,934 berita dari 11 kategori berbeda. Hasil penelitian menunjukkan bahwa penerapan teori semantik berbasis *frame* meningkatkan efisiensi dan akurasi dalam klasifikasi berita, dibandingkan dengan pendekatan konvensional. Penggunaan semantik *frame* memungkinkan identifikasi makna yang lebih mendalam melalui analisis hubungan antar frasa dan kata kunci dalam konteks berita. Kesimpulannya, integrasi semantik *frame* dalam linguistik komputasi memberikan kontribusi signifikan dalam mengatasi kompleksitas bahasa Arab dan meningkatkan akurasi serta efisiensi klasifikasi berita *online*. Penelitian ini berpotensi diaplikasikan dalam pengembangan sistem pengelolaan berita digital di berbagai platform berbahasa Arab.

Kata kunci: Klasifikasi berita, semantik *frame*, linguistik komputasi.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

ABSTRACT

This study focuses on the development of an automatic classification method for Arabic online news using a frame-based semantic approach. The background of this research is driven by the need to improve the efficiency and accuracy of Arabic news classification, as Arabic is known for its complex language structure and rich morphology. The main challenge in classifying Arabic news lies in understanding the deep linguistic context and properly organizing news into appropriate categories. This research aims to explore the application of frame semantics theory in the classification process of Arabic online news to enhance the comprehension of news meaning and automate the categorization of news articles. The research method involves using web scraping techniques to collect data from various Arabic news portals and employing machine learning algorithms, such as Support Vector Machine (SVM), supported by Term Frequency-Inverse Document Frequency (TF-IDF) vectorization for feature extraction. The data collected consists of 170,934 news articles from 11 different categories. The results show that the application of frame semantics improves both the efficiency and accuracy of news classification compared to conventional approaches. Frame semantics enables deeper meaning identification through the analysis of relationships between phrases and keywords in the news context. In conclusion, the integration of frame semantics in computational linguistics provides significant contributions to addressing the complexity of the Arabic language and enhancing the accuracy and efficiency of online news classification. This research has the potential to be applied in the development of digital news management systems across various Arabic-language platforms.

Keywords: *News classification, frame semantics, computational linguistics.*

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

تجريد

تركز هذه الدراسة على تطوير طريقة التصنيف الآلي للأخبار الإلكترونية باللغة العربية باستخدام نهج قائم على الدلالات الإطارية. تعتمد خلفية هذه الدراسة على الحاجة إلى تحسين كفاءة ودقة تصنيف الأخبار باللغة العربية، التي تُعرف بتعقيد بنيتها اللغوية وغناها بالمورفولوجيا. يمثل التحدي الرئيسي في تصنيف الأخبار باللغة العربية في كيفية فهم السياق اللغوي العميق وتنظيم الأخبار ضمن الفئات المناسبة بشكل دقيق. تهدف هذه الدراسة إلى استكشاف تطبيق نظرية الدلالات الإطارية في عملية تصنيف الأخبار الإلكترونية باللغة العربية، لتعزيز فهم المعنى الإخباري وأتمتة عملية تصنيف الأخبار. تعتمد منهجية البحث على استخدام تقنيات تجميع البيانات عبر الويب لجمع البيانات من مختلف بوابات الأخبار العربية، وتوظيف خوارزميات التعلم الآلي مثل آلة الدعم المتجهة (SVM)، مدعومة بتمثيل التردد العكسي للمستندات (TF-IDF) لاستخراج الخصائص. تتألف البيانات التي تم جمعها من 170,934 خبراً موزعة على 11 فئة مختلفة. أظهرت النتائج أن تطبيق نظرية الدلالات الإطارية يحسن الكفاءة والدقة في تصنيف الأخبار مقارنة بالأساليب التقليدية. تمكّن الدلالات الإطارية من تحديد المعاني بشكل أعمق من خلال تحليل العلاقات بين العبارات والكلمات المفتاحية في سياق الأخبار. في الختام، يساهم دمج الدلالات الإطارية في اللغويات الحاسوبية بشكل كبير في معالجة تعقيدات اللغة العربية وتحسين دقة وكفاءة تصنيف الأخبار الإلكترونية. ويمكن تطبيق هذه الدراسة في تطوير أنظمة إدارة الأخبار الرقمية على مختلف المنصات الناطقة بالعربية.

الكلمات المفتاحية: تصنيف الأخبار، الدلالات الإطارية، اللغويات الحاسوبية.

DAFTAR ISI

HALAMAN SAMPUL.....	i
MOTTO	ii
HALAMAN PERSEMBAHAN	iii
PENGESAHAN TUGAS AKHIR.....	iv
SURAT PERNYATAAN KEASLIAN TESIS.....	v
SURAT PERNYATAAN BEBAS PLAGIASI.....	vi
NOTA DINAS PEMBIMBING.....	vii
PEDOMAN TRANSLITERASI ARAB-LATIN	viii
KATA PENGANTAR.....	xii
ABSTRAK	xiii
ABSTRACT	xiv
تجريد.....	xv
DAFTAR ISI.....	xvi
DAFTAR TABEL	xx
DAFTAR GAMBAR.....	xxi
 BAB I PENDAHULUAN.....	 1
A. Latar Belakang Masalah.....	1
B. Rumusan Masalah	10
C. Tujuan dan Kegunaan Penelitian.....	10
D. Tinjauan Pustaka	11
1. Berita Sebagai Objek Analisis	11
2. Implementasi Teori Semantik <i>Frame</i> dalam Analisis Berita.....	15
3. Otomatisasi Klasifikasi Teks dan Tantangan Masa Kini	18
4. Penelitian yang Akan Dikaji	21
E. Kerangka Teori.....	21
1. Teori Semantik Berbasis <i>Frame</i>	21
2. Linguistik Komputasi.....	30
3. Klasifikasi Berita.....	34
F. Metode Penelitian.....	36
1. Jenis Penelitian.....	36
2. Subjek dan Objek Penelitian	37

3. Sumber Data dan Data Penelitian.....	38
4. Pengumpulan Data	40
5. Proses Klasifikasi Berita	43
6. Analisis Data	45
G. Sistematika Penulisan.....	47
BAB II TREN <i>DATASET</i> BERITA BERBAHASA ARAB.....	49
A. Deskripsi <i>Dataset</i>	50
B. Data Berdasarkan Sumber Berita	51
C. Data Frekuensi N-Gram	53
D. Data Karakteristik Linguistik	58
E. Frekuensi Kata Tertinggi Berdasarkan Kategorinya.....	60
BAB III BENTUK SEMANTIK <i>FRAME</i> DALAM BERITA BERBAHASA ARAB	62
A. Analisis Semantik <i>Frame</i> untuk Kategori Berita	64
1. Analisis dan Penerapan Semantik <i>Frame</i> Berita Keamanan.....	64
2. Analisis dan Penerapan Semantik <i>Frame</i> Berita Ekonomi	67
3. Analisis dan Penerapan Semantik <i>Frame</i> Berita Teknologi	70
4. Analisis dan Penerapan Semantik <i>Frame</i> Berita Budaya.....	73
5. Analisis dan Penerapan Semantik <i>Frame</i> Berita Internasional	76
6. Analisis dan Penerapan Semantik <i>Frame</i> Berita Olahraga	79
7. Analisis dan Penerapan Semantik <i>Frame</i> Berita <i>Travel</i>	82
8. Analisis dan Penerapan Semantik <i>Frame</i> Berita Politik	85
9. Analisis dan Penerapan Semantik <i>Frame</i> Berita Kesehatan	88
10. Analisis dan Penerapan Semantik <i>Frame</i> Berita Sains	92
11. Analisis dan Penerapan Semantik <i>Frame</i> Berita Masyarakat/Komunitas.....	96
B. Pemilihan Frasa Kunci Berdasarkan Kategori	99
C. Daftar Semantik <i>Frame</i> Berdasarkan Kategori.....	100
1. Daftar <i>Frame</i> Keamanan.....	101
2. Daftar <i>Frame</i> Ekonomi	102
3. Daftar <i>Frame</i> Teknologi.....	103
4. Daftar <i>Frame</i> Budaya.....	105
5. Daftar <i>Frame</i> Internasional	106

6. Daftar <i>Frame</i> Olahraga	108
7. Daftar <i>Frame Travel</i>	109
8. Daftar <i>Frame</i> Politik	110
9. Daftar <i>Frame</i> Kesehatan	112
10. Daftar <i>Frame</i> Sains	114
11. Daftar <i>Frame</i> Masyarakat/Komunitas.....	115
D. Integrasi Semantik <i>Frame</i> dengan Proses Linguistik Komputasi	118
1. Integrasi dalam Berita Keamanan	118
2. Integrasi dalam Berita Ekonomi.....	122
3. Integrasi dalam Berita Teknologi.....	125
4. Integrasi dalam Berita Budaya	130
5. Integrasi dalam Berita Internasional	133
6. Integrasi dalam Berita Olahraga.....	137
7. Integrasi dalam Berita <i>Travel</i>	141
8. Integrasi dalam Berita Politik.....	146
9. Integrasi dalam Berita Kesehatan.....	150
10. Integrasi dalam Berita Sains.....	154
11. Integrasi dalam Berita Masyarakat/Komunitas	159

BAB IV EFISIENSI DAN EFEKTIVITAS TEORI SEMANTIK BERBASIS *FRAME* DALAM PROSES KLASIFIKASI BERITA BERBAHASA ARAB

.....	164
A. Metodologi Evaluasi	164
1. Model Pembelajaran Mesin yang Digunakan	165
2. Metrik Evaluasi	165
3. Persiapan <i>Dataset</i>	166
4. Proses Evaluasi.....	167
B. Efisiensi Penerapan Teori Semantik <i>Frame</i>	168
1. Waktu Pemrosesan	168
2. Penggunaan Sumber Daya Komputasi.....	170
3. Perbandingan Efisiensi Antar Model	172
4. Pengaruh Teori Semantik <i>Frame</i> Terhadap Efisiensi	174
C. Efektivitas Penerapan Teori Semantik <i>Frame</i>	181
1. Akurasi	182
2. <i>Precision</i>	184

3. <i>Recall</i>	186
4. <i>F1-score</i>	188
5. Perbandingan efektivitas antar model	190
6. Pengaruh Teori Semantik <i>Frame</i> Terhadap Efektivitas	193
D. Analisis Model Klasifikasi Terbaik.....	197
1. Kinerja Model	198
2. Laporan Klasifikasi	200
3. Waktu Eksekusi.....	202
BAB V PENUTUP	203
A. Kesimpulan.....	203
B. Saran.....	204
DAFTAR PUSTAKA	206
DAFTAR RIWAYAT HIDUP PENULIS	216


 STATE ISLAMIC UNIVERSITY
 SUNAN KALIJAGA
 YOGYAKARTA

DAFTAR TABEL

Tabel 1. Struktur Data	40
Tabel 2. Frasa <i>bigram</i> frekuensi tertinggi berdasarkan kategorinya	53
Tabel 3. Frasa <i>trigram</i> frekuensi tertinggi berdasarkan kategorinya	56
Tabel 4. Daftar kata dengan frekuensi tertinggi berdasarkan kategorinya	60
Tabel 5. Frasa kunci dalam contoh berita keamanan	65
Tabel 6. Frasa kunci dalam contoh berita ekonomi	68
Tabel 7. Frasa kunci dalam contoh berita teknologi	71
Tabel 8. Frasa kunci dalam contoh berita budaya	74
Tabel 9. Frasa kunci dalam contoh berita internasional	77
Tabel 10. Frasa kunci dalam contoh berita olahraga	80
Tabel 11. Frasa kunci dalam contoh berita <i>travel</i>	83
Tabel 12. Frasa kunci dalam contoh berita politik	87
Tabel 13. Frasa kunci dalam contoh berita kesehatan	90
Tabel 14. Frasa kunci dalam contoh berita kesehatan	95
Tabel 15. Frasa kunci dalam contoh berita masyarakat/komunitas	97
Tabel 16. 10 frasa dalam <i>frame</i> keamanan dengan bobot tertinggi	101
Tabel 17. 10 frasa-frasa dalam <i>frame</i> ekonomi dengan bobot tertinggi	102
Tabel 18. 10 frasa dalam <i>frame</i> teknologi dengan bobot tertinggi	104
Tabel 19. 10 frasa dalam <i>frame</i> budaya dengan bobot tertinggi	105
Tabel 20. 10 frasa dalam <i>frame</i> internasional dengan bobot tertinggi	107
Tabel 21. 10 frasa dalam <i>frame</i> olahraga dengan bobot tertinggi	108
Tabel 22. 10 frasa dalam <i>frame travel</i> dengan bobot tertinggi	109
Tabel 23. 10 frasa dalam <i>frame</i> politik bobot tertinggi	111
Tabel 24. 10 frasa-frasa dalam <i>frame</i> kesehatan dengan bobot tertinggi	112
Tabel 25. 10 frasa dalam <i>frame</i> sains dengan bobot tertinggi	114
Tabel 26. 10 frasa dalam <i>frame</i> masyarakat/komunitas dengan bobot tertinggi ..	116
Tabel 27. Pembersihan angka dan tanda baca berita keamanan	119
Tabel 28. Pembersihan angka dan tanda baca berita ekonomi	122
Tabel 29. Pembersihan angka dan tanda baca berita teknologi	126
Tabel 30. Pembersihan angka dan tanda baca berita budaya	130
Tabel 31. Pembersihan angka dan tanda baca berita internasional	134
Tabel 32. Pembersihan angka dan tanda baca berita olahraga	138
Tabel 33. Pembersihan angka dan tanda baca berita <i>travel</i>	142
Tabel 34. Pembersihan angka dan tanda baca berita politik	147
Tabel 35. Pembersihan angka dan tanda baca berita kesehatan	151
Tabel 36. Pembersihan angka dan tanda baca berita sains	156
Tabel 37. Pembersihan angka dan tanda baca berita masyarakat/komunitas	160
Tabel 38. Perbandingan waktu eksekusi dan efisiensi model klasifikasi	173
Tabel 39. Perbandingan efektivitas model klasifikasi berita berbahasa Arab	191
Tabel 40. Laporan klasifikasi untuk model SVM	200

DAFTAR GAMBAR

Gambar 1. Elemen yang berkaitan dalam kategori cangkir	25
Gambar 2. Posisi linguistik komputasi.....	31
Gambar 3. Diagram teknik <i>web scrapping</i>	41
Gambar 4. Diagram klasifikasi berita	43
Gambar 5. Komposisi berita berdasarkan kategori	50
Gambar 6. Komposisi berita berdasarkan sumber	52
Gambar 7. Jumlah karakter berdasarkan jenis karakternya	58
Gambar 8. Jumlah <i>token</i> berdasarkan kategori	59
Gambar 9. <i>Wordcloud</i> hasil <i>tokenisasi</i> berita keamanan	119
Gambar 10. Hasil <i>vektorisasi</i> frasa berita keamanan	120
Gambar 11. <i>Wordcloud</i> hasil <i>tokenisasi</i> berita ekonomi	123
Gambar 12. Hasil <i>vektorisasi</i> frasa berita ekonomi	124
Gambar 13. <i>Wordcloud</i> hasil <i>tokenisasi</i> berita teknologi	127
Gambar 14. Hasil <i>vektorisasi</i> frasa berita teknologi	128
Gambar 15. <i>Wordcloud</i> hasil <i>tokenisasi</i> berita budaya.....	131
Gambar 16. Hasil <i>vektorisasi</i> frasa berita budaya.....	132
Gambar 17. <i>Wordcloud</i> hasil <i>tokenisasi</i> berita internasional.....	134
Gambar 18. Hasil <i>vektorisasi</i> frasa berita internasional.....	135
Gambar 19. <i>Wordcloud</i> hasil <i>tokenisasi</i> berita olahraga.....	139
Gambar 20. Hasil <i>vektorisasi</i> frasa berita olahraga	140
Gambar 21. <i>Wordcloud</i> hasil <i>tokenisasi</i> berita <i>travel</i>	143
Gambar 22. Hasil <i>vektorisasi</i> frasa berita <i>travel</i>	144
Gambar 23. <i>Wordcloud</i> hasil <i>tokenisasi</i> berita politik	147
Gambar 24. Hasil <i>vektorisasi</i> frasa berita politik.....	148
Gambar 25. <i>Wordcloud</i> hasil <i>tokenisasi</i> berita kesehatan.....	151
Gambar 26. Hasil <i>vektorisasi</i> frasa berita kesehatan	152
Gambar 27. <i>Wordcloud</i> hasil <i>tokenisasi</i> berita sains	157
Gambar 28. Hasil <i>vektorisasi</i> frasa berita sains	158
Gambar 29. <i>Wordcloud</i> hasil <i>tokenisasi</i> berita masyarakat/komunitas	161
Gambar 30. Hasil <i>vektorisasi</i> frasa berita masyarakat/komunitas	162
Gambar 31. Waktu eksekusi model klasifikasi.....	169
Gambar 32. Penggunaan sumber daya komputasi oleh model klasifikasi.....	171
Gambar 33. Perbandingan waktu eksekusi dengan dan tanpa semantik <i>frame</i>	175
Gambar 34. Perbandingan penggunaan memori dengan dan tanpa semantik <i>frame</i>	177
Gambar 35. Perbandingan penggunaan CPU dengan dan tanpa semantik <i>frame</i>	179
Gambar 36. Akurasi model klasifikasi berita berbahasa Arab.....	183
Gambar 37. <i>Precision</i> model klasifikasi berita berbahasa Arab	185
Gambar 38. <i>Recall</i> model klasifikasi berita berbahasa Arab	187
Gambar 39. <i>F1-Score</i> model klasifikasi berita berbahasa Arab.....	189
Gambar 40. Perbandingan akurasi model dengan dan tanpa semantik <i>frame</i>	193
Gambar 41. Perbandingan <i>precision</i> model dengan dan tanpa semantik <i>frame</i> ..	194

Gambar 42. Perbandingan <i>recall</i> model dengan dan tanpa semantik <i>frame</i>	195
Gambar 43. Perbandingan <i>F1-score</i> model dengan dan tanpa semantik <i>frame</i> ...	196
Gambar 44. <i>Confusion matrix</i> untuk model SVM	199



BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Era digital yang dipicu oleh perkembangan teknologi informasi dan komunikasi telah memperluas produksi dan konsumsi berita *online* secara global, termasuk di dunia Arab. Berita merupakan bentuk informasi yang berkenaan dengan suatu peristiwa yang ingin diketahui oleh umum karena bersifat aktual, melibatkan tokoh publik, terjadi di lingkungan pembaca, dan peristiwa tersebut dapat berpengaruh terhadap pembaca.¹ Volume besar berita yang dihasilkan setiap hari dari berbagai portal berita membutuhkan pengelolaan yang efisien, dan metode manual tidak lagi memadai untuk memenuhi kebutuhan ini. Komputasi dalam klasifikasi berita berbahasa Arab menjadi sangat penting untuk menyaring, mengelompokkan, dan menyajikan informasi secara cepat dan akurat. Proses otomatisasi ini diperlukan karena sifat berita yang selalu diperbarui, dan metode komputasi memungkinkan penanganan volume besar data berita dengan kecepatan dan konsistensi yang tidak bisa dicapai secara manual. Dengan bantuan teknologi komputasi, klasifikasi berita berbahasa Arab dapat dilakukan secara efisien, memastikan bahwa informasi yang disampaikan relevan dan terorganisir sesuai kategori, sehingga lebih mudah diakses oleh pembaca.

¹ Faith Gordon, "The significance and impact of the media in contemporary society," dalam *Children, young people and the press in a transitioning society* (London: Palgrave Macmillan UK, 2018), 17–46, https://doi.org/10.1057/978-1-137-60682-2_2.

Keunikan berita berbahasa Arab terlihat dari struktur kalimatnya yang kompleks² serta penggunaan gaya bahasa yang kaya³ dengan nilai budaya⁴ dan sudah ada sejak zaman klasik.⁵ Berita berbahasa Arab sering menggunakan metafora, retorika, serta istilah khusus yang sulit diterjemahkan ke dalam bahasa lain,⁶ sehingga membutuhkan pendekatan khusus dalam klasifikasinya.⁷ Pemilihan berita berbahasa Arab menjadi penting karena menawarkan perspektif yang unik dan mendalam, diperkaya oleh sejarah panjang, budaya yang kaya, dan struktur bahasa yang kompleks. Pemilihan berita berbahasa Arab menjadi penting karena konsep yang disusun dalam segala aspek bahasanya pada zaman Arab klasik tetap digunakan hingga zaman modern saat ini.

Pada umumnya, berita adalah informasi yang disebar dari banyak sumber sehingga melahirkan definisi yang banyak. Di dunia Arab, berita sering kali diterjemahkan berkali-kali dalam berbagai bahasa, menyebabkan distorsi makna asli karena tidak diterjemahkan oleh ahlinya.⁸ Walaupun demikian, makna dan arti dalam sebuah berita dapat disimpulkan. Sebuah berita terkadang mengandung satu

² Abdullah Zaitouni dan Latroush Charef, "Itālah Binā' Al-Jumlah 'An Tarīq At-Ta'āqib Fī Syi'ri Abī Abdullah Albū'abdalī," *Djoussour El-maarefa* 5, no. 1 (2019): 459–69.

³ Abdul Salam Hamid dan Ibrahim Sya'lan, "Anmāt Al-Jumlah Al-Ismiyyah Fī Namāzīj Min Jarīdah Ar-Rāyah Al-Qaṭriyyah: Dirāsah Naḥwiyyah Iḥṣā'īyyah," *Ḥauliyyāt Al-Ādāb Wa Al-Lughah* 10, no. 1 (2022): 9–35.

⁴ Muhammad Balebassi dan Rabha Adad, "Aṣar Al-Lughah Al-'Arabiyyah Fī Tarsīkh Muqawwamāt Al-Hawiyyah Al-Waṭaniyyah," *Dirāsāt Mu'āṣirah* 6, no. 2 (2022): 12–20.

⁵ Iman Fatima Al-Zahra Belqasm, "Al-Jumlah Al-Fi'liyyah Al-Baṣīṭah Al-Muṣbatah Fī Mu'allaqah Imri'i Al-Qais," *Majallah Kuliyyah Al-Ādāb wa Al-'Ulūm Al-Insāniyyah wa Al-Ijtimā'iyyah* 9, no. 4 (2011): 71–99.

⁶ Ganiah Bouhouche, "Tarjamah Ma'ānī Al-Qur'ān Al-Karīm Ilā Lugāt Ukhrā Ḍarūrah Da'wiyyah wa Mas'ūliyyah Syar'iyyah," *El-Mouzahir* 1, no. 2 (2020): 182–92.

⁷ Bilkhatir Bumdin, "Aṣar Ta'rīb Al-Mustalahāt Al-Qānūniyyah Al-'Ālamiyyah 'Alā Al-Fikr Al-Islāmī," *Majallah Ad-Dirāsāt Al-Islāmiyyah* 1, no. 1 (2012): 230–38.

⁸ Muhammad Ibrahim Jarah, "Aṣar At-Tarjamah Fī Tagyīri Al-Ma'nā wa Nasyr Al-Ma'lūmāt Az-Zā'ifah Fī Mawāqī'i At-Tawāṣul Al-Ijtimā'i Dirāsah Taṭbīqiyyah 'Alā Al-Lughah Al-'Arabiyyah," *Sémiotiques* 18, no. 2 (2023): 438–54.

atau lebih dari topik berita, padahal masih dalam satu judul berita.⁹ Di tengah kelimpahan informasi, klasifikasi berita menjadi aspek penting untuk memahami dan mengelolanya dengan lebih efisien. Banyaknya penggunaan media digital dalam menyampaikan berita berdampak pada banyaknya jumlah berita digital yang diterbitkan oleh beberapa portal berita setiap harinya menyebabkan berita tersebut berkaitan lebih dari satu kategori.¹⁰

Klasifikasi berita *online* menjadi aspek penting bagi pembaca karena tidak hanya menyediakan informasi tetapi juga sebagai kebutuhan untuk menyajikan informasi yang terstruktur. Klasifikasi berita juga merupakan aspek penting dalam *Natural Language Processing* (NLP) karena mengandung banyak informasi yang dibutuhkan.¹¹ Hal ini akan berhubungan dengan proses klasifikasi berita secara otomatis ke dalam kategori yang sudah ditentukan sebelumnya, sehingga data berita dengan jumlah yang sangat besar dapat diorganisir dengan baik. Klasifikasi berita ini akan dimanfaatkan oleh beberapa aplikasi dan sistem seperti sistem agregasi berita, sistem rekomendasi berbasis konten, hingga penyampaian berita yang dipersonalisasi.¹²

⁹ I Made Riartha Prawira, Adiwijaya, dan Mohamad Syahrul Mubarak, "Klasifikasi Multi-Label Pada Topik Berita Berbahasa Indonesia Menggunakan Multinomial Naïve Bayes," *e-Proceeding of Engineering* 5, no. 3 (2018): 7774–81.

¹⁰ B.H. Mahendra, A. Adiwijaya, dan U.N. Wisesty, "Kategorisasi Berita Multi-Label Berbahasa Indonesia Menggunakan Algoritma Random Forest," *e-Proceeding of Engineering* 6, no. 2 (2019): 9030–41.

¹¹ Jeffrey Pennington, Richard Socher, dan Christopher Manning, "Glove: Global vectors for word representation," dalam *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, ed. oleh Alessandro Moschitti, Bo Pang, dan Walter Daelemans (Stroudsburg, PA, USA: Association for Computational Linguistics, 2014), 1532–43, <https://doi.org/10.3115/v1/D14-1162>.

¹² M. Abualsaud dkk., "A deep learning model for Arabic news classification," *Journal of King Saud University-Computer and Information Sciences* 30, no. 4 (2018): 408–17.

Klasifikasi berita secara manual adalah metode tradisional yang dilakukan manusia. Proses tersebut akan memakan waktu yang panjang dan biaya yang mahal. Seiring dengan meningkatnya jumlah berita yang diterbitkan setiap hari oleh beberapa portal berita, metode manual tersebut menjadi tidak efisien dan tidak praktis. Oleh karena itu, penggunaan metode “pembelajaran mesin” atau yang biasa disebut dengan *Machine Learning* (ML) khususnya dengan algoritma *Deep learning*, menjadi metode yang populer dalam beberapa tahun terakhir ini karena menawarkan otomatisasi klasifikasi berita ke dalam berbagai topik.¹³ Khusus untuk berita berbahasa Arab, pentingnya klasifikasi yang akurat tidak bisa diabaikan, mengingat peran penting bahasa Arab dalam dunia informasi dan komunikasi.

Kebutuhan akan klasifikasi berita *online* berbahasa Arab didorong oleh jumlah berita yang terus meningkat. Berita berbahasa Arab umumnya menggunakan bahasa Arab *Fuṣḥā* (bahasa Arab standar), yang tidak terpengaruh oleh variasi dialek dan didasarkan pada uraian bahasa Arab klasik.¹⁴ Namun, bahasa Arab *Fuṣḥā* memiliki struktur yang kompleks dan kaya akan morfologi, yang membuat proses klasifikasi menjadi tidak mudah.¹⁵ Bahasa Arab merupakan bahasa yang memiliki karakteristik yang membedakannya dengan bahasa lainnya dan universal karena adanya kesamaan nilai antara bahasa Arab dan bahasa lainnya.¹⁶

¹³ Roobaea Alroobaea, “An Empirical Deep Learning Approach for Arabic News Classification,” (*IJACSA*) *International Journal of Advanced Computer Science and Applications* 14, no. 6 (2023): 1048–55, www.ijacsa.thesai.org.

¹⁴ Ali Benfettacha, “Min Adillah Irtibāt An-Nahw Al-’Arabī Bi Al-’Arabiyyah Al-Fuṣḥā,” *Majallah Isykalāt fi Al-Lughah wa Al-Adab* 10, no. 4 (2021): 561–77.

¹⁵ Usama Muhmali, “Āliyyāt Al-Mustasyriqīn Li Hadm Al-Lughah Al-’Arabiyyah Al-Fuṣḥā,” *Cahiers du Laboratoire la Poétique Algérienne* 2, no. 1 (2017): 28–42.

¹⁶ Faliqul Isbah dan M. Fairuz Rosyid, “Amaliyyah al-Andragogiya li Mu’allim al-Lughah al_juhūd li Tarqiyyah al-Kafā’ah al-Tarbawiyyah,” *Jurnal Alfazuna : Jurnal Pembelajaran Bahasa Arab dan Kebahasaaraban* 5, no. 02 (18 Juni 2021): 170–90, <https://doi.org/10.15642/alfazuna.v5i02.1042>.

Bahasa Arab memiliki sistem standar yang terdiri dari tata bunyi, tata kata, kalimat, gramatikal, hingga wacana. Setiap sub-sistem ini memiliki aturan khusus yang bekerja secara interdependen.¹⁷

Pendekatan yang cermat dalam *Natural Language Processing* (NLP) diperlukan untuk klasifikasi berita berbahasa Arab mengingat konsep bahasa yang berbeda dengan bahasa lainnya. Selain itu, bahasa Arab juga dikenal sebagai bahasa derivasi yang kaya akan bunyi, bentuk, dan ungkapan, serta memiliki teknik penyusunan kalimat yang kompleks.¹⁸ Berdasarkan tantangan tersebut, penelitian ini berfokus pada pengembangan metode klasifikasi yang memanfaatkan teori semantik *frame* untuk mengatasi tantangan ini dan meningkatkan akurasi serta efisiensi dalam pengelolaan berita berbahasa Arab. Dengan demikian, hasil penelitian diharapkan dapat memberikan kontribusi signifikan dalam pengelolaan informasi berbahasa Arab secara umum.

Penerapan teori semantik berbasis *frame* dalam penelitian ini menawarkan pendekatan konseptual yang dapat mengatasi masalah kompleksitas bahasa dan konteks. Pada konsepnya, teori tersebut memberikan landasan yang kokoh untuk melakukan identifikasi pola-pola makna yang muncul dalam teks berita *online*. Fillmore¹⁹ memperkenalkan gagasan tentang konsep *frame* sebagai sebuah pilihan pada sistem linguistik dan mengemukakan bahwa *frame* merupakan sistem kategori yang disusun sedemikian rupa untuk memberikan akses kepada orang lain untuk

¹⁷ Isbah dan Rosyid.

¹⁸ Wati Susiawati, "Kajian Bahasa Arab dari A Historis hingga Historis," *Alfaz (Arabic Literatures for Academic Zealots)* 7, no. 01 (18 Juni 2019): 43, <https://doi.org/10.32678/alfaz.Vol7.Iss01.1925>.

¹⁹ Charles J. Fillmore, "An Alternative to Checklist Theories of Meaning," dalam *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, 1975, 123–31.

memahami *frame* dan konteks pengetahuan tersebut.^{20,21} Teori Semantik *Frame* menyatakan bahwa setiap kata akan dikaitkan dengan serangkaian kata atau pemicu yang terkait dengan kata tersebut. Artinya, sebuah kata dapat membangkitkan semantik *frame* tertentu dalam suatu peristiwa atau wacana tertentu pula.²² Akibatnya, model dalam teori ini akan melibatkan identifikasi istilah-istilah penting atau unsur-unsur kebahasaan yang kemudian dihubungkan pada *frame*, yaitu kumpulan atau unit dalam *frame* tertentu dan berasal dari elemen itu sendiri.²³ Keseluruhan *frame* akan memberikan makna yang komprehensif ke wacana yang dimaksud. Dengan kata lain, memahami *frame*, yaitu struktur konseptual, memerlukan pemahaman kumpulan atau unit dalam *frame* turunannya.²⁴

Misalnya, dalam sebuah teks berita bahasa Arab ada kelompok kata yang dominan muncul berisi *أسعار صرف الدولار* (harga tukar dolar), *البورصة الرئيسية* (bursa utama), *دولار* (di pasar lokal), *في الأسواق المحلية* (harga jual menurun), *انخفضت أسعار البيع* (dolar Amerika), *إنفاق حكومي* (kebutuhan keuangan) dan *حاجة إلى مال* (belanja pemerintah). Secara tradisional, jika diperhatikan dengan saksama, kelompok kata

²⁰ Charles J. Fillmore, "Frame Semantics," dalam *Linguistics in the Morning Calm: Selected Papers from SICOL-1981*, ed. oleh The Linguistics Society of Korea (Seoul: Hanshin Publishing Company, 1982), 111–37.

²¹ Pernyataan Fillmore tentang *frame*: "system of categories structured in ... [which] some words exist in order to provide access to knowledge of such *frames* to the participants".

²² C. J. Fillmore dan B. T. Atkins, "Toward a Frame-Based Lexicon: The Semantics of RISK and Its Neighbors," dalam *Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organization*, ed. oleh A. Lehrer dan E. F. Kittay (Hillsdale, NJ: Lawrence Erlbaum Associates, 1992), 75–102.

²³ John Nerbonne, "Natural language processing in computer-assisted language learning," dalam *The Oxford handbook of computational linguistics (Oxford handbooks)*, ed. oleh Ruslan Mitkov, 2 ed. (USA: Oxford University Press, Inc., 2022), 634–98.

²⁴ Saba Kamil Ali dan Rana H. Al-Bahrani, "A Cognitive Linguistic Study of the Persuasive/Attention Strategies Used in Selected American English E-Advertisements," *Arab World English Journal* 7, no. 1 (15 Juli 2021): 279–98, <https://doi.org/10.24093/awej/call7.20>.

tersebut adalah kumpulan atau unit-unit dalam *frame* ekonomi, maka dapat diasumsikan bahwa berita tersebut berkategori ekonomi karena merupakan kumpulan atau unit dalam turunan yang sama. Hal ini membuktikan bahwa penggunaan semantik *frame* dapat membantu mengidentifikasi kategori dari suatu berita seperti politik, ekonomi, olahraga, kriminal, teknologi, sains, budaya, sosial, kesehatan, dan sebagainya. Ini membantu meningkatkan efisiensi dan akurasi dalam pengelompokan berita sesuai dengan kriteria tematik yang diinginkan.

Linguistik komputasi yang diterapkan dalam klasifikasi berita akan memperkaya proses dan hasil analisis dengan pendekatan berbasis komputer. Penerapan algoritma dan ML dapat memicu peningkatan efisiensi klasifikasi berita dan keakuratan hasilnya dengan memanfaatkan data besar yang sudah tersedia. Konsep semantik *frame* berperan penting untuk menjadi dasar bagi model-model NLP yang akan dibangun. Hausser²⁵ menyatakan bahwa linguistik komputasi menggabungkan metode-metode tata bahasa tradisional dan linguistik teoretis dengan metode yang seefektif mungkin mampu memverifikasi berbagai hipotesis dengan jelas melalui cara pengimplementasian tata bahasa formal dalam bentuk program komputer yang efisien dan mampu mengujinya secara otomatis dan realistis.²⁶ Kemudian hal tersebut dirincikan oleh Vogel²⁷ tentang hubungan antara komputasi dan linguistik dengan mengatakan bahwa linguistik adalah ilmu

²⁵ Roland Hausser, *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language* (Berlin; Heidelberg; New York; Barcelona; Hong Kong: Springer, 2001).

²⁶ "Computational Linguistics combines the methods of traditional grammar and theoretical linguistics with the method of effectively verifying explicit hypotheses by implementing formal grammars as efficient computer programs and testing them automatically on realistic"

²⁷ Carl (Dir.) Vogel, *Computer Science, Linguistics and a Language: Course Handbook 2006-2007* (Dublin: University of Dublin, 2006).

pengetahuan tentang bahasa manusia, baik bahasa ucap, bahasa tulis maupun bahasa dalam bentuk tanda, sedangkan linguistik komputer adalah ilmu pengetahuan tentang bahasa yang menitikberatkan perhatiannya secara khusus pada pemrosesan bahasa natural dengan berbagai kompleksitasnya, yang dipandu oleh rekayasa program yang secara kognitif dibuat oleh manusia.²⁸

Jika dikaitkan dua pernyataan di atas dengan proses klasifikasi berita dan teori semantik *frame*, dapat disimpulkan bahwa linguistik komputasi merupakan suatu pendekatan yang mempertemukan bahasa (manusia) dan komputer (mesin).²⁹ Hal tersebut akan menghasilkan sebuah kombinasi di mana komputer akan berpikir, mengelola, memproses bahasa seperti yang dilakukan manusia.³⁰ Pada awalnya, manusia mengklasifikasikan sebuah kategori berita dengan bantuan teori semantik *frame* dan dikelola secara tradisional. Sekarang, kehadiran linguistik komputasi mengubah cara tradisional tersebut menjadi cara yang lebih efisien dan efektif dalam segi waktu dan keakuratan hasil.³¹ Dengan memahami semantik *frame* kata-kata, sistem linguistik komputasi dapat mengidentifikasi makna kata-kata dalam konteksnya,³² mendukung pemrosesan bahasa yang lebih kontekstual dan

²⁸ “*Linguistics is the science of human language, whether spoken, written or signed; computational linguistics is the science of language with particular attention given to the processing complexity constraints dictated by the human cognitive architecture. Like most sciences, computational linguistics also has engineerin*”

²⁹ Guemaz Jamilah, “Al-Lisāniyāt Al-Ḥasūbiyah: Mafhūmuhā Manhajuhā wa Majallātu Istikhdāmihā,” *University of JIJEL* 8, no. 2 (2022): 8–17.

³⁰ Maryam Abū Bakar, “Qaḍāyā wa Taḥḍiyāt Aẓ-ẓakā’ Al-Iṣṭinā’ī fī Al-Lisāniyāt Al-’Arabiyah,” dalam Al-Mu’tamir Ad-Daulī Liṭalabah Ad-Dirāsāt Al-’Ulyā fī Ta’līmī Al-Lughah Al-’Arabiyah wal Ādāb wal Lisāniat 2023 (Universitas Malang, 2023), 286–97.

³¹ Muḥammad Saif Islām, “Al-Lisāniāt Al-Ḥasūbiyah wa Isykalāt Al-Manhaj wal Anẓimah fī Mizānīl Baḥsi-Al-Mu’alajah Taḥlīliyah li Ru’yi ‘Amaliyyah ‘Arabiyyah Mutamayyizah,” *Al-Mumarrasāt Al-Lugawīyyāh* 11, no. 2 (2020): 41–63.

³² Salahuddin Yahia dan Lamia Kedache, “Al-Lisāniyyāt Al-’Urfāniyyah wal Muḥṭawā Al-Ijra’ī Li Naẓriyyah Dalālah Al-Aṭar Fī Al-Madkhal Al-Mu’jāmiyyāh,” *Dirāsāt Al-Ma’āshirah* 5, no. 7 (2021): 249–64.

mendalam. Berangkat dari latar belakang tersebut, akan dilakukan penelitian berjudul “Komputasi Klasifikasi pada Berita *Online* Berbahasa Arab dengan Pendekatan Semantik Berbasis *Frame*” yang diharapkan dapat berkontribusi pada pengembangan linguistik komputasi sebagai strategi pemecahan masalah untuk meningkatkan efisiensi dan akurasi klasifikasi berita bahasa Arab.

Penelitian ini bertujuan untuk mengisi celah pengetahuan dalam domain klasifikasi berita berbahasa Arab dalam pemberitaan *online*, khususnya dalam konteks linguistik komputasi berbasis semantik *frame*. Meskipun telah ada kemajuan dalam penggunaan algoritma pembelajaran mesin untuk otomatisasi klasifikasi berita, penelitian yang fokus pada bahasa Arab masih terbatas. Bahasa Arab *Fuṣḥā* memiliki gramatikal yang kompleks dan kaya morfologi, yang menciptakan tantangan tambahan dalam mengenali frasa dan makna dalam berita. Oleh karena itu, penelitian ini akan mengembangkan metode klasifikasi berbasis linguistik komputasi yang memanfaatkan teori semantik *frame* untuk meningkatkan efisiensi dan akurasi klasifikasi berita berbahasa Arab secara otomatis. Dengan demikian, penelitian ini diharapkan tidak hanya memberikan pemahaman yang lebih baik tentang proses klasifikasi berita dalam konteks bahasa Arab, tetapi juga berkontribusi pada pengembangan teknik baru dalam bidang linguistik komputasi yang dapat diterapkan secara luas dalam analisis berita *online* dalam bahasa yang kompleks dan beragam.

B. Rumusan Masalah

Rumusan masalah dalam penelitian “Komputasi Klasifikasi pada Berita *Online* Berbahasa Arab dengan Pendekatan Semantik Berbasis *Frame*” diuraikan dengan mempertimbangkan langkah-langkah yang melibatkan data *train* dan *test* pada 170,934 berita *online* dengan kategori tertentu, serta proses pembersihan dan normalisasi data untuk mencapai hasil *output* yang sesuai dan akurasi yang baik. Adapun rincian rumusan masalah adalah sebagai berikut:

1. Bagaimana pola dan tren utama yang dapat ditemukan dalam dataset berita berbahasa Arab yang digunakan berdasarkan kategori berita yang telah ditentukan?
2. Bagaimana semantik *frame* yang terbentuk dalam berita berbahasa Arab oleh model klasifikasi berbahasa Arab?
3. Bagaimana efisiensi dan efektivitas teori semantik berbasis *frame* dalam proses klasifikasi berita berbahasa Arab?

C. Tujuan dan Kegunaan Penelitian

Penelitian ini memiliki tujuan dan kegunaan praktis dalam meningkatkan pemahaman dan klasifikasi berita berbahasa Arab. Adapun tujuan dan kegunaan penelitian “Komputasi Klasifikasi pada Berita *Online* Berbahasa Arab dengan Pendekatan Semantik Berbasis *Frame*” adalah sebagai berikut:

1. Mengidentifikasi pola yang terdapat dalam dataset berita berbahasa Arab sesuai dengan kategori berita yang telah ditentukan.
2. Mengidentifikasi dan menganalisis semantik *frame* yang terbentuk dalam berita berbahasa Arab oleh model klasifikasi teks.

3. Menilai efisiensi dan efektivitas penerapan teori semantik berbasis *frame* dalam proses klasifikasi berita berbahasa Arab dalam pemberitaan *online*.

Adapun kegunaan penelitian ini adalah:

1. Memberikan kontribusi dalam pengembangan teknologi linguistik komputasi berbahasa Arab dengan memanfaatkan teori semantik berbasis *frame* sebagai landasan.
2. Meningkatkan akurasi dan efisiensi dalam klasifikasi berita berbahasa Arab, memberikan manfaat praktis dalam mengelola dan menyajikan informasi berita secara lebih terstruktur.
3. Menyumbangkan pemahaman mendalam tentang faktor-faktor yang memengaruhi klasifikasi dan merancang strategi pemecahan masalah yang dapat diterapkan dalam konteks lain.
4. Memberikan pemahaman yang lebih mendalam terhadap proses klasifikasi berita berbahasa Arab, memperkaya pengetahuan dalam konteks linguistik komputasi.

D. Tinjauan Pustaka

1. Berita Sebagai Objek Analisis

Teks berita berbahasa Arab menjadi objek material utama dalam penelitian ini. Sebagai media komunikasi massa, berita memainkan peran penting dalam menyampaikan informasi kepada masyarakat, membentuk persepsi publik, dan memengaruhi opini. Oleh karena itu, analisis terhadap struktur, makna, dan elemen-elemen bahasa dalam teks berita menjadi

esensial untuk memahami cara berita berfungsi dalam membangun wacana dan menyalurkan pesan.

Berbagai penelitian sebelumnya telah mengaplikasikan metode analisis yang beragam untuk mengkaji berita sebagai objek kajian. Misalnya, metode analisis *framing* yang dilakukan oleh Frasticha & Pribadi,³³ Harnia dkk.,³⁴ Khoiruddin & Setiawan,³⁵ dan Komang dkk.,³⁶ menyoroti bagaimana pengaturan elemen-elemen berita dapat memengaruhi cara pandang pembaca terhadap isu yang diberitakan. Analisis ini memberikan wawasan tentang cara pemilihan kata dan struktur teks secara langsung dapat memanipulasi makna serta persepsi publik.

Selain itu, analisis wacana kritis seperti yang telah dikembangkan oleh Andayani dkk.,³⁷ Fadillah & Nurhadi,³⁸ dan Br Tarigan dkk.,³⁹

³³ Lussy Yuris Frasticha dan Farid Pribadi, "Bingkai Demokratisasi Isu Pengesahan UU Cipta Kerja (Analisis Framing Zhongdang Pan Dan Gerald M. Kosicki Pada Media Dalam Jaringan Kompas.Com Dan Tribunnews.Com)," *Jurnal Komunikasi dan Media* 2, no. 1 (2022): 1–9, <https://doi.org/10.24167/jkm.v2i1.3206>.

³⁴ Neng Tika Harnia, Ferina Meliasanti, dan Hendra Setiawan, "Analisis Framing Berita Perundungan pada Media Online Detik.Com dan Tribunnews.Com sebagai Bahan Ajar Teks Berita di SMP," *Edukatif: Jurnal Ilmu Pendidikan* 3, no. 5 (17 Agustus 2021): 3145–53, <https://doi.org/10.31004/edukatif.v3i5.1240>.

³⁵ Achmad Khoiruddin dan Hendra Setiawan, "Analisis Framing Berita Korupsi Maskapai Garuda Indonesia pada Media Online Cnbendonesia.com dan Sindonews.com sebagai Bahan Ajar Teks Berita di SMP," *Edukatif: Jurnal Ilmu Pendidikan* 4, no. 1 (30 Desember 2021): 778–85, <https://doi.org/10.31004/edukatif.v4i1.2005>.

³⁶ Ni Komang dkk., "Analisis Komponen Teks Berita Karya Siswa Kelas VIII SMP Negeri Satu Atap 1 Tejakula," *Jurnal Pendidikan Bahasa dan Sastra Indonesia Undiksha* 11, no. 2 (2021): 175–84, <https://doi.org/https://doi.org/10.23887/jjpbs.v11i2.35282>.

³⁷ Sri Andayani, Indra Tjahyadi, dan Hosnol Wafa, "Analisis Wacana Kritis Model Van Dijk Dalam Pemberitaan Krisis Energi di Media Daring cnbendonesia.com," dalam *Conference on Innovation and Application of Science and Technology (CIASTECH)*, 2021, 177–86, <https://doi.org/https://doi.org/10.31328/ciastech.v0i0>.

³⁸ Putri Fadillah dan Jatmika Nurhadi, "Analisis Wacana Kritis Teun A. Van Dijk pada Teks Berita KPK Respons Wacana Periksa Anies di Kasus Korupsi Munjul," *Jurnal Bahasa, Sastra, dan Pengajarannya* 7, no. 2 (2021).

³⁹ Anna Juliana Br Tarigan dkk., "Analisis Wacana Kritis Model Teun A. Van Dijk pada Teks Berita di Buku Bahasa Indonesia Kelas VII Berbasis Kurikulum Merdeka Terbitan

menggali lebih dalam konstruksi makna dan bagaimana teks berita dapat bertahan atau menanggapi kritik sosial yang muncul. Pendekatan ini memperlihatkan bagaimana berita tidak hanya menyampaikan informasi, tetapi juga merefleksikan nilai dan ideologi tertentu. Di sisi lain, analisis struktur teks dan kelengkapan unsur berita seperti yang dilakukan oleh Bangun dkk.⁴⁰ dan Karisna,⁴¹ berfokus pada aspek formal dan isi berita, yang penting untuk menilai kelayakan informasi yang disajikan.

Selanjutnya, analisis terhadap frasa dan diksi, yang dicontohkan oleh Aliyah dkk.,⁴² Dyah Hapsari dkk.,⁴³ dan Karwati & Wulansari,⁴⁴ menawarkan perspektif tentang bagaimana pemilihan bahasa dan pemakaian kata dapat membentuk karakter berita dan pengaruhnya terhadap audiens. Di lain pihak, analisis semantik seperti yang dilakukan oleh Dede

Kemendikbudristek Tahun 2021,” *Basastra* 12, no. 1 (2023): 17–34, <https://doi.org/10.24114/bss.v12i1.40645>.

⁴⁰ Eric Persadanta Bangun, Ferry V.I. A Koagouw, dan J.S. Kalangi, “Analisis Isi Unsur Kelengkapan Berita Pada Media Online Manadopostonline.com,” *Acta Diurna Komunikasi* 1, no. 3 (2019).

⁴¹ Deksen Karisna, “Analisis Unsur-Unsur Kelengkapan Berita dalam Teks Berita Siswa MTs. Muhammadiyah Lebung Itam,” *Wahana Didaktika : Jurnal Ilmu Kependidikan* 18, no. 1 (28 Januari 2020): 95–102, <https://doi.org/10.31851/wahanadidaktika.v18i1.4359>.

⁴² Hikmah Himmatul Aliyah, Roni Nugraha Syafroni, dan Suntoko Suntoko, “Analisis Deiksis Sosial pada Teks Berita Media Daring Detik News Seputar Covid-19,” *Jurnal Educatio FKIP UNMA* 8, no. 1 (2022): 22–26, <https://doi.org/10.31949/educatio.v8i1.1458>.

⁴³ Intan Kumala Dyah Hapsari dkk., “Analisis Kesalahan Frasa Pada Teks Berita Covid-19 Koran Digital goriau.com,” *Lingua Rima: Jurnal Pendidikan Bahasa dan Sastra Indonesia* 10, no. 2 (2021): 35–42, <https://doi.org/10.31000/lgrm.v10i2.4740>.

⁴⁴ Dini Karwati dan Wulansari Wulansari, “Analisis Frasa Berdasarkan Golongan Kata Terhadap Teks Berita ‘Saat Internet Jadi Kambing Hitam Terhapusnya Rekaman CCTV Tragedi Kanjuruhan,’” *PENTAS: Jurnal Ilmiah Pendidikan Bahasa dan Sastra Indonesia* 8, no. 2 (23 November 2022): 27–34, <https://doi.org/10.52166/pentas.v8i2.3512>.

Hamzah & Reza Firdaus Tsani⁴⁵ dan Mardikantoro & Haryadi,⁴⁶ mengeksplorasi makna-makna tersembunyi yang mungkin tidak secara eksplisit disampaikan dalam teks, sehingga memberikan pemahaman lebih dalam tentang pesan implisit berita.

Selain itu, analisis resepsi oleh Santoso⁴⁷ memperlihatkan bagaimana berita diterima dan diinterpretasikan oleh pembaca, sementara analisis kesalahan sintaksis oleh Ariyadi & Utomo,⁴⁸ Augustin dkk.,⁴⁹ Talitha & Ferdianto,⁵⁰ dan Yani dkk.⁵¹ menyoroti pentingnya aspek gramatikal dan struktur kalimat dalam menunjang kejelasan dan ketepatan informasi yang disampaikan dalam berita.

Secara keseluruhan, pendekatan-pendekatan yang beragam ini menggambarkan cakupan luas dalam memahami teks berita, melibatkan aspek makna, sentimen, struktur, resepsi, dan elemen gramatikal serta

⁴⁵ Dede Hamzah dan Reza Firdaus Tsani, "Analisis Semantik Terhadap Teks Berita 'KKP Eksekusi RJ Lino ke Lapas Cipinang' Pada Media Online RMOL Jabar Terbitan 4 November 2022," *Ta'rim: Jurnal Pendidikan dan Anak Usia Dini* 3, no. 3 (2022): 113–26, <https://doi.org/10.59059/tarim.v3i3.53>.

⁴⁶ Hari Bakti Mardikantoro dan Haryadi Haryadi, "Korupsi Dalam Konstruksi Media : Analisis Struktur Mikro Semantik Teks Berita Korupsi Di Televisi," *Widyaparwa* 47, no. 1 (26 Agustus 2019): 104–15, <https://doi.org/10.26499/wdprw.v47i1.195>.

⁴⁷ Sofiana Santoso, "Analisis Resepsi Audiens Terhadap Berita Kasus Meiliana di Media Online," *Komuniti : Jurnal Komunikasi dan Teknologi Informasi* 12, no. 2 (22 Januari 2021): 140–54, <https://doi.org/10.23917/komuniti.v12i2.13285>.

⁴⁸ Ade Dufadhol Ariyadi dan Asep Purwo Yudi Utomo, "Analisis Kesalahan Sintaksis pada Teks Berita Daring berjudul Mencari Etika Elite Politik di saat Covid-19," *Jurnal Bahasa dan Sastra* 8, no. 3 (27 Desember 2020): 138–45, <https://doi.org/10.24036/jbs.v8i3.110903>.

⁴⁹ Rina Maulina Augustin, Isah Cahyani, dan Dadang Anshori, "Analisis Kesalahan Berbahasa Dalam Bidang Morfologi Pada Teks Berita Siswa SMPIT," dalam *Seminar Internasional Riksa Bahasa XIII* (Universitas Pendidikan Indonesia, 2020), 1497–1506.

⁵⁰ Stella Talitha dan Willi Ferdianto, "dAnalisis Kesalahan Berbahasa Pada Teks Berita Daring Dan Implikasinya Terhadap Pembelajaran Bahasa Indonesia SMP," *Jurnal Pendidikan Bahasa dan Sastra Indonesia* I, no. 1 (2022): 82–90.

⁵¹ Tri Andra Yani, Moh. Shofiuddin Shofi, dan Adelia, "Analisis Kesalahan Berbahasa Dalam Kumpulan Teks Berita Daring Radar Tegal," *DIALEKTIKA Pendidikan Bahasa Indonesia* 1, no. 2 (2021): 60–70.

linguistik lainnya. Tujuan utamanya adalah untuk memastikan bahwa berita disajikan secara informatif dan kohesif baik dalam segi makna maupun tata bahasa. Dengan demikian, kajian-kajian ini membangun fondasi yang kokoh bagi pengembangan sistem klasifikasi berita yang efektif, yang mampu memahami berita secara mendalam serta mempertimbangkan berbagai dimensi linguistik yang kompleks.

2. Implementasi Teori Semantik *Frame* dalam Analisis Berita

Proses klasifikasi berita berbasis teori semantik *frame* menjadi objek formal dalam penelitian ini. Semantik *frame*, sebagai teori yang menekankan hubungan antara kata-kata dan konteks penggunaannya, telah diaplikasikan secara luas dalam berbagai bahasa dan dengan beragam tujuan. Teori ini memungkinkan analisis yang mendalam terhadap makna teks berdasarkan hubungan antara elemen-elemen linguistik dalam suatu konteks tertentu, sehingga relevan untuk berbagai aplikasi linguistik, termasuk dalam kajian teks berita.

Penelitian-penelitian sebelumnya menunjukkan berbagai bentuk penerapan teori semantik *frame* dalam konteks bahasa yang berbeda. Beberapa penelitian, misalnya yang dilakukan oleh Guan dkk.,⁵² menggarisbawahi potensi teori ini dalam merangkum teks melalui pengelompokan konsep semantik berdasarkan *frame* yang relevan.

⁵² Yong Guan dkk., "Frame Semantics guided network for Abstractive Sentence Summarization," *Knowledge-Based Systems* 221 (Juni 2021): 106973, <https://doi.org/10.1016/j.knosys.2021.106973>.

Sementara itu, penelitian oleh Prasetyo⁵³ memperlihatkan kemampuan teori ini untuk mengklasifikasikan nama dan istilah tertentu, memperkaya pemahaman konteks yang lebih spesifik.

Selain itu, penerapan teori ini juga terlihat dalam penelitian yang fokus pada pengembangan kata kerja denominal baru, sebagaimana diungkapkan oleh Yu & Xu,⁵⁴ yang mengilustrasikan penggunaan teori semantik *frame* dalam pembentukan elemen bahasa baru. Aplikasi lain meliputi analisis kognisi manusia, seperti dalam penelitian mengenai kontrafaktual⁵⁵ dan pengaruh makna kata terhadap persepsi, yang membantu memahami bagaimana pembaca memproses makna dalam berita atau teks lainnya.⁵⁶

Dalam ranah leksikal dan frasa, penelitian yang dilakukan oleh Schafroth & Imperiale⁵⁷ dalam bahasa Italia, serta kajian konstruktivisme oleh Kim,⁵⁸ menunjukkan bagaimana teori semantik *frame* dapat mendukung analisis perbandingan antarbahasa. Sifat keuniversalan teori ini

⁵³ Indrawan Dwi Prasetyo, "Frame Semantics in Naming Military Vehicles," *Language Horizon* 6, no. 1 (2018): 28–39.

⁵⁴ Lei Yu dan Yang Xu, "Noun2Verb: Probabilistic Frame Semantics for Word Class Conversion," *Computational Linguistics* 48, no. 4 (1 Desember 2022): 783–818, https://doi.org/10.1162/coli_a_00447.

⁵⁵ Jochen Strobel, Jan de Vries, dan Friederike Wißmach, "Leserkognition und Wissensemergenz bei kontrafaktischem Erzählen aus Sicht der Frame-Semantik," *Zeitschrift für Literaturwissenschaft und Linguistik* 49, no. 1 (12 Maret 2019): 139–62, <https://doi.org/10.1007/s41244-019-00120-8>.

⁵⁶ Tamara M. Johnson dan Simone E. Pfenninger, "Tasty words: Using frame semantics to enhance consumer liking of potato chips and apples," *International Journal of Applied Linguistics* 31, no. 1 (25 Maret 2021): 79–94, <https://doi.org/10.1111/ijal.12323>.

⁵⁷ Elmar Schafroth dan Riccardo Imperiale, "Gebrauchsbasierte Phraseologie des Italienischen: Digitale Lexikographie zwischen Frame-Semantik und Konstruktionsgrammatik," *Lexicographica* 35, no. 2019 (1 Desember 2019): 133–67, <https://doi.org/10.1515/lex-2019-0004>.

⁵⁸ Jiwon Kim, "Eine Analyse der Werbung als Werbekonstruktionen - Interaktion zwischen Konstruktionsgrammatik und Frame-Semantik," *Journal for German Culture and Literature* 28 (31 Desember 2019): 117–41, <https://doi.org/10.32681/JGCL.28.5>.

juga tampak pada penelitian semantik-kognitif di berbagai bahasa, seperti dalam bahasa China,⁵⁹ Polandia,⁶⁰ dan Jerman⁶¹ yang menyoroti kesamaan prinsip semantik dalam beragam konteks budaya.

Penelitian lainnya, seperti yang dilakukan oleh Doğan⁶² mengenai derajat transivitas dalam bahasa Turki, dan pengembangan model berbasis span untuk parsing semantik *frame* oleh Su dkk.,⁶³ mengilustrasikan aplikasi teori ini dalam pemrosesan bahasa alami. Hal ini membuktikan relevansi teori semantik *frame* dalam konteks pemrosesan teks secara otomatis. Fokus pada bahasa dan budaya juga tercermin dalam penelitian Hussain dkk. pada klasifikasi bahasa Bengali,⁶⁴ Shao dkk. dalam bahasa China,⁶⁵ serta Chanakot & Sanrach dalam bahasa Thailand,⁶⁶ yang menunjukkan bagaimana teori ini dapat diterapkan dalam konteks yang beragam.

⁵⁹ Lee Ha-Yeong, "Semantic Analysis of '弄' through Frame Semantic Approach," *Chinese Studies* 82, no. 1 (31 Maret 2023): 121–38, <https://doi.org/10.14378/KACS.2023.82.82.7>.

⁶⁰ Maria Maćkowiak dkk., "Understanding of Dementia in the Polish Language: A Frame Semantic Approach," *Journal of Alzheimer's Disease* 91, no. 1 (3 Januari 2023): 389–406, <https://doi.org/10.3233/JAD-220633>.

⁶¹ Magdalena Szulc-Brzozowska, "Patriotismus – eine framebasierte kontrastive Analyse Deutsch-Polnisch anhand der Boulevardzeitung Bild und ihres polnischen Pendantes Fakt," *Roczniki Humanistyczne* 69, no. 5 (16 Juni 2021): 181–97, <https://doi.org/10.18290/rh21695-10>.

⁶² Nuh Doğan, "Formal and Semantic Hierarchy of Turkish Transitive Verbs," *İçtimaiyat* 6, no. 2 (2022): 688–703, <https://doi.org/10.33709/ictimaiyat.1195290>.

⁶³ Xuefeng Su dkk., "A Span-based Target-aware Relation Model for Frame-semantic Parsing," *ACM Transactions on Asian and Low-Resource Language Information Processing* 22, no. 3 (31 Maret 2023): 1–24, <https://doi.org/10.1145/3569581>.

⁶⁴ Md Gulzar Hussain dkk., "Comparison analysis of Bangla news articles classification using support vector machine and logistic regression," *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 21, no. 3 (1 Juni 2023): 584–91, <https://doi.org/10.12928/telkomnika.v21i3.23416>.

⁶⁵ Dangguo Shao dkk., "A news classification applied with new text representation based on the improved LDA," *Multimedia Tools and Applications* 81, no. 15 (15 Juni 2022): 21521–45, <https://doi.org/10.1007/s11042-022-12713-6>.

⁶⁶ Benjamin Chanakot dan Charun Sanrach, "Classifying thai news headlines using an artificial neural network," *Bulletin of Electrical Engineering and Informatics* 12, no. 1 (1 Februari 2023): 395–402, <https://doi.org/10.11591/eei.v12i1.4228>.

Secara keseluruhan, penelitian-penelitian ini memperlihatkan luasnya cakupan aplikasi teori semantik *frame* dalam analisis teks, baik untuk pemahaman makna secara manual maupun dalam sistem pemrosesan otomatis. Hal ini menegaskan peran teori semantik *frame* sebagai landasan dalam klasifikasi berita berbahasa Arab pada penelitian ini, memungkinkan analisis yang komprehensif dengan memperhatikan berbagai dimensi linguistik dan kognitif.

3. Otomatisasi Klasifikasi Teks dan Tantangan Masa Kini

Dalam perkembangan otomatisasi klasifikasi teks, berbagai peneliti telah memberikan kontribusi yang signifikan melalui beragam pendekatan dan metode untuk mengatasi kompleksitas analisis teks secara otomatis. Seiring berjalannya waktu, metode-metode ini semakin berkembang guna memenuhi kebutuhan yang terus berubah dalam bidang pemrosesan bahasa alami. Misalnya, langkah awal diambil oleh Lange dkk.⁶⁷ yang mengusulkan teknik ekstraksi dan normalisasi fitur-temporal sebagai metode utama untuk meningkatkan ketepatan klasifikasi teks.

Pendekatan tersebut kemudian dikembangkan lebih lanjut, seperti dalam penelitian Chanakot & Sanrach⁶⁸ yang menggunakan pembagian data 15-fold dengan metode Term Frequency-Inverse Class Frequency (TFICF),

⁶⁷ Lukas Lange, Omar Alonso, dan Jannik Strötgen, "The Power of Temporal Features for Classifying News Articles," dalam *Companion Proceedings of The 2019 World Wide Web Conference* (New York, NY, USA: ACM, 2019), 1159–60, <https://doi.org/10.1145/3308560.3315000>.

⁶⁸ Chanakot dan Sanrach, "Classifying thai news headlines using an artificial neural network."

sehingga meningkatkan ketelitian klasifikasi. Sholih'afif dkk.⁶⁹ berkontribusi melalui penerapan model klasifikasi teks berbasis algoritma Naïve Bayes, sementara Hendriyanto & Sari⁷⁰ mengeksplorasi Knowledge Discovery in Databases (KDD) sebagai landasan untuk memahami struktur konten secara lebih mendalam.

Sejalan dengan kemajuan dalam teknologi deteksi berita palsu, Trueman dkk.⁷¹ menciptakan model AC-BiLSTM, diikuti oleh Gupta & Gupta⁷² yang menerapkan model LSTM dalam klasifikasi berita berbahasa Punjabi, serta Shehatta⁷³ yang berfokus pada berita berbahasa Arab. Pendekatan ini mengilustrasikan pentingnya model jaringan saraf dalam mengidentifikasi pola-pola kompleks dalam teks.

Selanjutnya, tantangan otomatisasi dalam representasi dokumen turut diperhatikan oleh para peneliti. Meena⁷⁴ memanfaatkan otomatisasi untuk pengenalan tulisan digital dan kaligrafi Arab, sedangkan Yang &

⁶⁹ Muhammad Sholih 'afif dkk., "Text Mining Untuk Mengklasifikasi Judul Berita Online Studi Kasus Radar Banjarmasin Menggunakan Metode Naïve Bayes," *Kumpulan Jurnal Ilmu Komputer (KLIK)* 8, no. 2 (2021): 199–208.

⁷⁰ Muhammad Diki Hendriyanto dan Betha Nurina Sari, "Penerapan Algoritma K-Nearest Neighbor dalam Klasifikasi Judul Berita Hoax," *Jurnal Ilmiah Informatika* 10, no. 02 (15 September 2022): 80–84, <https://doi.org/10.33884/jif.v10i02.5477>.

⁷¹ Tina Esther Trueman dkk., "Attention-based C-BiLSTM for fake news detection," *Applied Soft Computing* 110 (Oktober 2021): 107600, <https://doi.org/10.1016/j.asoc.2021.107600>.

⁷² Varun Gupta dan Ekta Gupta, "Punjabi news multi-classification using language generation-based optimized long short-term memory networks," *Evolving Systems* 14, no. 1 (8 Februari 2023): 49–58, <https://doi.org/10.1007/s12530-022-09428-2>.

⁷³ Shehatta Usamah Sally, "Aẓ-Ẓakā' Al-Iṣṭinā'ī Wa Al-Taḥqīyyāt Wa Al-Adawāt Ar-Raqmiyyah Al-Mutakhaṣṣah Fī Al-Kasyfī 'An Al-Akhbār Az-Zā'ifah," *RAKMANA* 3, no. 2 (2023): 173–85.

⁷⁴ Aminah Meena, Imane Chacha, dan Hisamuddin Taouririt, "At-Ta'rīf Al-Ālī wa Al-Mu'ālajah Ar-Raqmiyyah Li Al-Khaṭ Al-'Arabī Baina Al-Ḥad Al-Lugawī wa Aḍ-Ḍabṭ At-Taḥqī," *Majallah Ta'līmiyyah* 13, no. 3 (2023): 495–503.

Wang⁷⁵ mengimplementasikan Linked Open Data (LOD) untuk memperkaya representasi teks. Shao dkk.⁷⁶ mengembangkan pendekatan representasi dokumen dengan model Latent Dirichlet Allocation (LDA), sementara Kumar Dutta dkk.⁷⁷ memperkenalkan inovasi melalui Weighted Extreme Learning Machine (WELM) dan Chaotic Ant Swarm (CAS) untuk klasifikasi berita palsu. Koloski dkk.⁷⁸ bahkan membawa dimensi baru dengan menggunakan representasi berbasis graf pengetahuan, yang semakin memperluas cakupan analisis otomatis dalam teks.

Meskipun telah ada banyak kemajuan, beberapa kekurangan masih tersisa, terutama dalam hal adaptasi teori Semantik Berbasis *Frame* secara terpadu dengan linguistik komputasi dalam konteks klasifikasi berita berbahasa Arab. Tantangan ini menggarisbawahi pentingnya pendekatan yang mampu menggabungkan elemen semantik dengan teknik komputasional guna meningkatkan efektivitas dan ketelitian dalam analisis teks yang kompleks.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

⁷⁵ Hsin Chang Yang dan Yu Chih Wang, "Content enrichment using linked open data for news classification," *Journal of Internet Technology* 21, no. 5 (2020): 1397–1407, <https://doi.org/10.3966/160792642020092105015>.

⁷⁶ Shao dkk., "A news classification applied with new text representation based on the improved LDA."

⁷⁷ Ashit Kumar Dutta dkk., "Optimal Weighted Extreme Learning Machine for Cybersecurity Fake News Classification," *Computer Systems Science and Engineering* 44, no. 3 (2023): 2395–2409, <https://doi.org/10.32604/csse.2023.027502>.

⁷⁸ Boshko Koloski dkk., "Knowledge graph informed fake news classification via heterogeneous representation ensembles," *Neurocomputing* 496 (Juli 2022): 208–26, <https://doi.org/10.1016/j.neucom.2022.01.096>.

4. Penelitian yang Akan Dikaji

Penelitian ini menyoroti pentingnya penggabungan teori semantik *frame* dan linguistik komputasi dalam klasifikasi berita berbahasa Arab. Dengan menjadikan teks berita sebagai objek material dan klasifikasi berbasis *frame* sebagai objek formal, penelitian ini berupaya menjawab kebutuhan klasifikasi otomatis yang akurat untuk berita berbahasa Arab. Metode ini diharapkan dapat mengisi celah dalam penelitian sebelumnya dan memberikan kontribusi signifikan pada pengembangan teknik klasifikasi dalam bidang NLP berbahasa Arab.

E. Kerangka Teori

1. Teori Semantik Berbasis *Frame*

Semantik *frame*, seperti yang dijelaskan oleh Charles J. Fillmore, adalah sebuah program penelitian dalam semantik empiris yang menganalisis makna kata dengan "menekankan kesinambungan, bukan ketidaksinambungan, antara bahasa dan pengalaman".⁷⁹ Semantik *frame* menentukan entri leksikal (dan mengamati maknanya) berdasarkan kerangka konseptual yang konsisten, di mana berbagai model kognitif manusia bersinergi.⁸⁰ Gagasan yang dikemukakan oleh kerangka kerja deskriptif semantik *frame* berbeda dengan pendekatan standar semantik formal yang didasarkan pada fitur semantik dan kondisi kebenaran. Teori ini mencoba untuk memberikan penjelasan yang komprehensif tentang

⁷⁹ Fillmore, "Frame Semantics."

⁸⁰ George Lakoff dan Mark Johnson, *Al-Isti'ārāt Allatī Nahyā Bihā*, ed. oleh Abdul Majid Juhfa (Casablanca, Maroko: Dār Tuqbal Lin Nasyr, 2009).

bagaimana makna disusun dan dikaitkan dengan kata-kata dalam struktur semantik dan bagaimana hal ini memberikan akses ke sistem konseptual kita, inventaris pengetahuan terstruktur yang kita gunakan untuk menavigasi dunia.⁸¹ Pandangan ini menjelaskan hubungan antara kata-kata yang tidak hanya dapat dianggap sebagai hubungan semantik struktural seperti hiponimi, sinonimi, atau antonimi. Sebaliknya, pandangan ini menggambarkan saling ketergantungan antara kata-kata berdasarkan latar belakang pengetahuan yang diperoleh manusia melalui pengalaman dan disimpan dalam memori jangka panjang. Mengutip kata-kata Fillmore:⁸²

With respect to word meaning, *frame* semantics can be thought of as the effort to understand what reason a speech community might have found for creating the category represented by the word, and to explain the word's meaning by presenting and clarifying that reason

Dari pernyataan ini, di samping keterkaitan sistemis antara kata-kata dan konsep-konsep yang mendasarinya, muncul satu karakteristik lebih lanjut dari semantik *frame*, yaitu bahwa pengetahuan didasarkan pada interaksi manusia dengan orang lain dan dengan dunia. Dengan demikian, *frame* berfungsi sebagai ekspektasi dan pengetahuan sebelumnya tentang dunia yang terus diperkuat, diperlemah, atau diubah sesuai dengan informasi yang diterima otak dari indera persepsi. *Frame* yang dimiliki diuji untuk meningkatkan prediksi dan jaringan *frame* yang berkembang dalam pikiran. Teori ini berusaha untuk mengamati keseluruhan struktur kognitif

⁸¹ Vyvyan Evans dan Melanie Green, *Cognitive linguistics: An introduction* (Lawrence Erlbaum Associates Publishers, 2006).

⁸² Fillmore, "Frame Semantics."

dan konseptual yang mengelilingi entri leksikal ini, karena *frame* ini berkontribusi dalam menentukan makna istilah dan merupakan sistem kognitif yang terkait dengan makna istilah dan strukturnya yang konseptual.

Makna leksikal harus sesuai dengan struktur konseptual dan kognitif. Teori semantik membahas hubungan entri dalam kamus, mengandalkan hubungan semantik antar kata dalam bidang tertentu. Ini mendukung pemahaman terintegrasi Fillmore, karena bidang semantik menggambarkan aspek perilaku manusia. Konsep bidang menunjukkan pentingnya menghubungkan entri yang membentuk bidang dengan *frame* berbasis pemahaman terintegrasi. Entri leksikal ditentukan oleh hubungan semantik antar kata melalui struktur linguistik dan kognitif.⁸³ Organisasi konseptual dari istilah memiliki sifat sebagai unit dalam pengetahuan, unit terminologi yang memiliki dimensi pengetahuan yang beragam (konsep), serta sebagai unit dalam bahasa dan unit dalam proses persepsi. Oleh karena itu, deskripsinya harus mencakup seluruh aspek tersebut, termasuk situasi.⁸⁴

Semantik *frame* dikembangkan untuk mengatasi beberapa masalah linguistik penting yang dihadapi dan tidak dapat dipecahkan oleh teori-teori sebelumnya. Pendekatan klasik terhadap semantik yang berasal dari Aristoteles didasarkan pada empat asumsi utama. **Pertama**, kategori linguistik didefinisikan dalam hal kondisi yang diperlukan dan cukup. Jadi, anak laki-laki akan didefinisikan sebagai [manusia], [laki-laki], [muda].

⁸³ Abdul Majid Al- Juhfa, *Madkhal Ilā 'Ilmi Ad-Dalālah Al-Hadīs* (Casablanca: Dar Toubkal, 2000).

⁸⁴ Mahyaddin Mohseb, *Al-Idrākiyyāt Ab'ād Istimūlūjiyyah Wajhāt Taṭbīqīyyah*, 1 ed. (Amman, Jordan: Dār Kunūz Al-Ma'rifah, 2017).

Dalam pandangan ini, makna dari suatu bentuk linguistik direpresentasikan dalam bentuk daftar syarat yang harus dipenuhi agar bentuk tersebut dapat digunakan secara tepat atau benar. Selain itu, pandangan ini berpendapat bahwa semua elemen yang memenuhi daftar periksa harus dianggap berasal dari kategori. **Kedua**, fitur-fitur itu bersifat biner. Dalam hal ini, anak perempuan akan digambarkan sebagai [manusia], [minus laki-laki], [muda]. **Ketiga**, kategori didefinisikan sebagai sesuatu yang memiliki batas-batas yang jelas dan anggota-anggota dalam sebuah kategori memiliki status yang sama.⁸⁵

Namun, asumsi pertama dan kedua mengalami masalah ketika mencoba mendefinisikan konsep yang kompleks seperti 'bujangan'. Syarat yang diperlukan dan cukup untuk itu adalah [manusia], [laki-laki], [dewasa], [tidak menikah]. Namun definisi ini akan memasukkan seorang 'Paus' sebagai bujangan, yang membuktikan kelemahan definisi tersebut. Selain itu, mendefinisikan kategori seperti "permainan" menimbulkan kesulitan jika ditinjau menggunakan asumsi pertama dan kedua.⁸⁶ Dalam kategori seperti permainan, tidak mungkin untuk menemukan seperangkat kondisi yang akan menjelaskan semua permainan mulai dari catur hingga sepak bola secara universal.

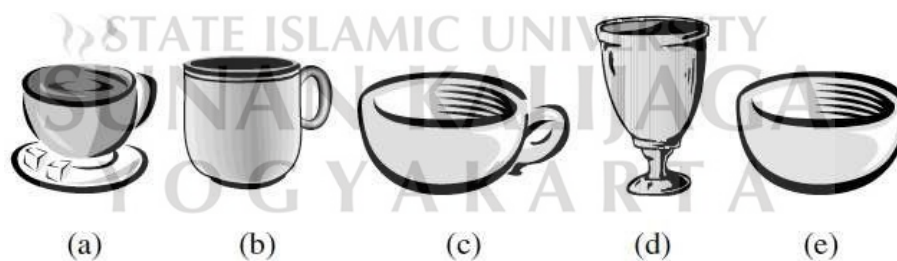
Mengenai asumsi **keempat**, masalah muncul dengan mempertimbangkan burung unta dan penguin sebagai hewan yang termasuk

⁸⁵ William Croft dan D. Alan Cruse, *Cognitive Linguistics* (Cambridge University Press, 2004), <https://doi.org/10.1017/CBO9780511803864>.

⁸⁶ Ludwig Wittgenstein, *Philosophische Untersuchungen*, 3 ed. (Harlow, London: Prentice Hall, 1958).

dalam kategori burung, dibandingkan dengan merpati atau burung gereja. Tentunya semua hewan ini secara akurat harus dimasukkan ke dalam kategori burung, namun, dua yang pertama tidak dapat dikatakan mewakili kategori tersebut seperti halnya dua elemen yang kedua. Hal ini telah didefinisikan sebagai masalah prototipe, yang dianalisis secara mendalam oleh Rosch⁸⁷ di antara para ahli lainnya. Menurut Rosch, beberapa entitas dalam kategori-kategori tersebut lebih prototipikal daripada yang lain.

Elemen-elemen ini lebih mudah tersedia ketika memikirkan kategori, mereka lebih mewakili kategori tersebut. Akhirnya, agak sulit untuk menyatakan bahwa kategori didefinisikan dengan jelas seperti yang ditunjukkan oleh Gambar 1 yang menunjukkan beberapa elemen yang berkaitan dengan kategori cangkir. Memang sangat sulit untuk mengidentifikasi kapan kategori cangkir berakhir dan kategori mangkuk dimulai seperti yang ditunjukkan oleh elemen (c) dan (e) pada Gambar 1 berikut.



Gambar 1. Elemen yang berkaitan dalam kategori cangkir

⁸⁷ Eleanor Rosch, "Principles of categorization," dalam *Cognition and Categorization*, ed. oleh Eleanor Rosch dan Barbara Lloyd (Hillsdale, NJ: Erlbaum, 1999), 27–48.

“Prototypes are understood as a fairly large slice of the surrounding culture against which the meaning of a word is defined and understood”.⁸⁸

Semantik *frame* dan teori prototipe berpadu dengan sangat apik, misalnya, untuk menjelaskan arti kata sarapan. Pertama, kita perlu memahami budaya makan tiga kali sehari, pada waktu yang kurang lebih tetap (namun berbeda dari satu budaya ke budaya lainnya) dan bahwa makanan pertama yang dimakan di pagi hari setelah tidur, memiliki menu yang khusus. Salah satu ciri utama dari teori prototipe adalah bahwa tidak semua kondisi yang mendefinisikan prototipe harus ada agar pembicara dapat menggunakankata tersebut tanpa kesulitan untuk dimengerti. Oleh karena itu, semua situasi berikut ini dapat dimasukkan ke dalam kategori ini: tidur sepanjang pagi, makan telur, roti panggang, kopi, dan jus jeruk pada pukul dua siang; pergi ke disko pada malam hari, begadang semalaman dan makan *croissant* dan kopi pada pukul tujuh pagi; tidur sepanjang malam, makan *pizza* dingin yang tersisa dari hari sebelumnya sebelum makan malam pada pukul tujuh pagi. Kata sarapan menyediakan sebuah kategori, yang dapat digunakan dalam berbagai konteks; konteksnya ditentukan oleh penggunaan prototipe kata tersebut; penggunaan prototipe adalah penggunaan yang dimilikinya saat kondisi situasi latar belakang sesuai dengan prototipe yang menentukan.⁸⁹

⁸⁸ Miriam R.L. Petruck, “Frame semantics,” dalam *Handbook of Pragmatics* (Amsterdam: John Benjamins Publishing Company, 1997), 1–13, <https://doi.org/10.1075/hop.2.fra1>.

⁸⁹ Fillmore, “Frame Semantics.”

Teori-teori lain telah mencoba menjelaskan semantik leksikal dengan hasil yang tidak memuaskan. Contohnya adalah teori medan semantik seperti yang dicirikan dan dipraktikkan oleh para ahli teori medan. Dalam teori medan semantik, sebuah kata didefinisikan dalam kaitannya dengan kata-kata lain dalam medan yang sama (dalam arti ‘medan peserta dalam perlombaan’). Dalam Semantik *frame*, sebuah kata didefinisikan dalam hubungannya dengan bingkai yang mendasarinya, bukan dalam hubungannya dengan kata-kata lain. Teori medan makna kata mendasarkan asumsinya pada keberadaan dan pengetahuan tentang kata-kata lain yang terkait secara semantik. Namun, pandangan ini tidak mengizinkan adanya bingkai dengan leksikal tunggal.

Teori linguistik lainnya telah mencoba untuk memisahkan pengetahuan mental bahasa dari penggunaan bahasa, membedakan antara semantik dan pragmatik. Ini adalah kasus kerangka kerja generatif. Chomsky, dan yang lainnya sebelumnya (sama halnya dengan perbedaan de Saussure antara *Langue* dan *Parole*)⁹⁰ dan setelahnya, berpendapat bahwa bahasa hanya dapat dijelaskan secara bermakna dari perspektif internalis (internal dalam pikiran individu) dan bukan dari perspektif (eksternalis) penggunaan bahasa. Ini adalah perbedaan antara kompetensi (pengetahuan) dan kinerja (penggunaan). Chomsky berargumen untuk keistimewaan kompetensi daripada kinerja sebagai subjek linguistik.

⁹⁰ Ferdinand de Saussure, *Course in General Linguistics* (New York: Columbia University Press, 2011).

Para ahli bahasa kognitif dan semantik *frame* menolak perbedaan antara kompetensi dan kinerja. Pengetahuan tentang bahasa berasal dari pola penggunaan bahasa dan bagaimana bahasa digunakan. Menurut psikolog Michael Tomasello,⁹¹ “struktur bahasa muncul dari penggunaan bahasa”,⁹² yang dikenal sebagai paradigma berbasis penggunaan. Perbedaan dalam makna sintaksis disebabkan oleh faktor linguistik dan struktur semantis, serta keberagaman makna satu tanda dan struktur mental konseptual. Charles Fillmore merangkum ini dalam teori semantik *frame*, yang menentukan kerangka hubungan semantis dan konseptual dalam bahasa.⁹³

Pada titik ini, beberapa contoh lebih lanjut dapat berguna untuk memperjelas posisi yang dipegang oleh pendekatan semantik *frame* dan teori-teori yang dibangun lebih lanjut di atas asumsi-asumsi ini. *Pertama*, peran pengalaman menjadi jelas bagi pemahaman makna kata-kata ketika menganalisis konsep alat seperti palu, jam, sepatu dan pensil atau objek pada Gambar 1. Menurut Fillmore, untuk mengetahui dan memahami konsep alat, kita harus mengetahui fonetik dan material fisik dari benda-benda tersebut. Namun dalam pandangannya, hal ini tidaklah cukup, dalam artian bahwa kita juga harus mengetahui untuk apa benda-benda tersebut

⁹¹ Michael Tomasello, *Constructing a Language: A Usage-Based Theory of Language Acquisition* (Harvard University Press, 2005), 5, <https://doi.org/10.2307/j.ctv26070v8>.

⁹² Kata Michael Tomasello sebagai ahli bahasa kognitif: “*language structure emerges from language use*”

⁹³ Lazhar Zaneed, “Al-Muta’alīq Bi Murātib Al-Ittisā’ Fi Ad-Dilālah Al-Mu’jāmiyyah,” dalam *Fuṣūl Fī Ad-Dilālah Mā Baina Al-Mu’ajjim Wa An-Naḥwi* (Beirut: Ad-Dār Al-‘Arabiyyah Lil ‘Ulūm Nasyirūn Wa Mansyūrāt Al-Ikhtilāf, 2010).

dibuat dan bahkan orang seperti apa yang menggunakannya dan mengapa mereka menggunakannya. Objek pada gambar 1, yang pada pandangan pertama mungkin terlihat tidak jelas bagi kebanyakan orang di dunia barat yang tidak memiliki pengetahuan tentang budaya *Ayurveda*, adalah pembersih lidah *Ayurveda*. Kita dapat dengan mudah membayangkan kasus di mana ketika diminta untuk mengambil alat pembersih lidah di kamar mandi umum di India, seorang turis yang ragu-ragu menatap rak kamar mandi, mencoba untuk membayangkan bentuk benda tersebut, mungkin tanpa menemukannya. Kasus ini memberi kita contoh miskomunikasi karena kehilangan *frame*. Di sini, penerjemah tidak dapat memanggil pengetahuan yang tepat dari ingatan, kerangka yang tepat, untuk memahami permintaan tersebut dan oleh karena itu tidak dapat menghasilkan perilaku yang diminta. Selain itu, contoh ini juga menunjukkan kepada kita bahwa ini bukanlah masalah pemahaman kata, LIDAH dan BERSIH, yang dapat diasumsikan dipahami oleh turis. Tetapi karena dia belum pernah menemukan objek atau kata atau konsep yang berhubungan dengannya sebelumnya, dia tidak dapat mengidentifikasi benda yang tepat.

Kasus sebaliknya adalah kata-kata yang tidak ada lagi karena bingkainya menghilang. Ini, misalnya, adalah kasus kata PHLOGISTON, yang berarti ‘zat tanpawarna, bau, atau berat, yang diyakini dikeluarkan saat terbakar oleh semua bahan yang mudah terbakar’.⁹⁴ Menurut semantik

⁹⁴ Fillmore, “Frame Semantics.”

frame, kata tersebut akan hilang dari kamus ketika bingkai yang digunakan untuk memahami konsep tersebut terbukti tidak akurat secara empiris.

Dalam konteks penelitian ini, teori semantik *frame* memberikan landasan penting dalam mengembangkan model klasifikasi berita berbahasa Arab. Dengan menerapkan teori ini, penelitian bertujuan untuk menangkap makna berita berdasarkan *frame* konseptualnya, sehingga memungkinkan pengkategorian berita secara otomatis ke dalam berbagai topik berdasarkan konteks makna yang terkandung dalam teks.

2. Linguistik Komputasi

Linguistik komputasi adalah studi ilmiah interdisipliner yang menggabungkan linguistik, ilmu komputer, dan matematika untuk menciptakan sistem yang memungkinkan komputer memproses bahasa alami secara otomatis, sebagaimana dilakukan oleh manusia.⁹⁵ Dalam penelitian ini, linguistik komputasi berperan dalam mendukung klasifikasi otomatis berita berbahasa Arab, yang memiliki kompleksitas unik dari segi morfologi dan sintaksis. Agar komputer mampu mengolah bahasa alami ini, diperlukan perancangan algoritma yang mengikuti aturan linguistik di berbagai tingkat dari fonologi hingga sintaksis dan semantik.⁹⁶

Menurut Stefano Crespi Reghizzi dan Ruslan Mitkov, linguistik komputasi melibatkan teori linguistik umum, teori komputer, serta

⁹⁵ Umar Mehdivi, *Taulīd Al-Asmā' Min Al-Juz'ūr As-Sulāsiyyah As-Sahīḥah fī Al-Lughah Al-'Arabiyyah - Muqārabah Lisāniyyah Ḥāsūbiyyah*, ed. oleh Abdul Gani Abu Al- 'Azm, vol. 1 (Casablanca: Wiḥdah 'Ulūm Al-Lughah Al-'Arabiyyah wa Al-Mu'jamiyyāt, 2008).

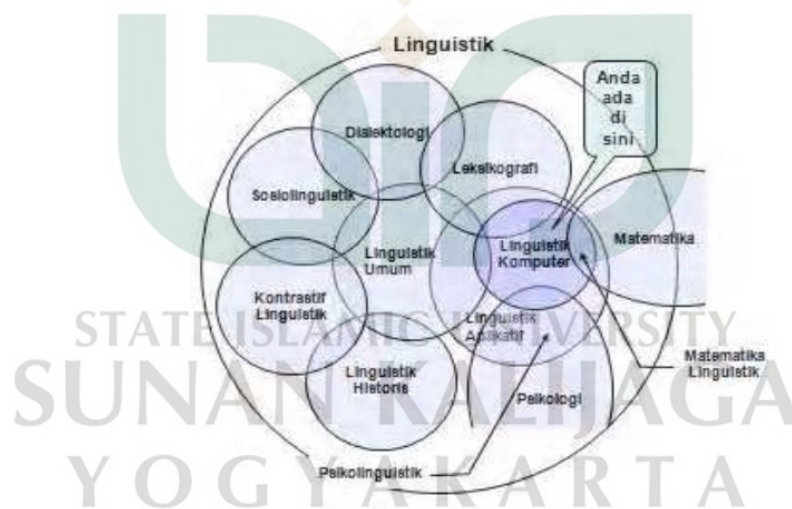
⁹⁶ Nihad Al- Mousa, *Al-Lughah Al-'Arabiyyah Naḥw Tauṣīf Jadīd fī Ḍau' Al-Lisāniyyāt Al-Ḥāsūbiyyah*, 1 ed. (Beirut: Al-Mu'assisah Al-'Arabiyyah Li Ad-Dirāsāt wa An-Nasyr, 2000).

matematika dan algoritma, menjadikannya bidang yang berfokus pada verifikasi hipotesis linguistik secara terprogram. Hausser menyatakan,⁹⁷

Computational Linguistics combines the methods of traditional grammar and theoretical linguistics with the method of effectively verifying explicit hypotheses by implementing formal grammars as efficient computer programs and testing them automatically on realistic “Linguistik komputer menggabungkan metode-metode tata bahasa tradisional dan linguistik teoretis dengan metode yang seefektif mungkin mampu memverifikasi berbagai hipotesis dengan jelas melalui cara pengimplementasian tata bahasa formal dalam bentuk program komputer yang efisien dan mampu mengujinya secara otomatis dan realistis”

Jika demikian, dimanakah letak dan posisi linguistik komputer (*computational linguistics*) di antara kajian-kajian linguistik yang lain?

Jawaban untuk hal ini dapat dilihat pada Gambar 2 berikut.



Gambar 2. Posisi linguistik komputasi

⁹⁷ Hausser, *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language*, 18.

Pada gambar di atas,⁹⁸ tampak jelas bahwa bahwa linguistik komputasi termasuk dalam lingkup linguistik terapan, tetapi melibatkan berbagai disiplin ilmu tambahan seperti matematika, psikologi, dan leksikografi untuk mendukung pemrosesan bahasa alami. Berikut adalah penjabaran beberapa disiplin penting yang berperan:

- a) Linguistik umum penting bagi linguistik komputer karena tema-tema dalam bidang kajian linguistik umum seperti fonologi, morfologi, sintaksis, semantik, dan pragmatik juga menjadi tema kajian linguistik komputer. Hanya saja, di dalam linguistik komputer semua tema dalam bidang kajian linguistik umum tersebut dikemas dalam kerangka program aplikasi pemrosesan bahasa natural oleh komputer (*Natural Language Processing*). Aturan-aturan yang digunakan oleh komputer di dalam pemrosesan bahasa natural juga didasarkan pada aturan-aturan yang ada di dalam linguistik umum.
- b) Linguistik umum penting bagi linguistik komputer karena tema-tema dalam bidang kajian linguistik umum seperti fonologi, morfologi, sintaksis, semantik, dan pragmatik juga menjadi tema kajian linguistik komputer. Hanya saja, di dalam linguistik komputer semua tema dalam bidang kajian linguistik umum tersebut dikemas dalam kerangka program aplikasi pemrosesan bahasa

⁹⁸ Igor A. Bolshakov dan Alexander Gelbukh, ed., *Computational linguistics: Models, resources, applications* (Mexico: Ciencia De La Computación, 2004).

natural oleh komputer (*Natural Language Processing*). Aturan-aturan yang digunakan oleh komputer di dalam pemrosesan bahasa natural juga didasarkan pada aturan-aturan yang ada di dalam linguistik umum.

c) Linguistik umum penting bagi linguistik komputer karena tema-tema dalam bidang kajian linguistik umum seperti fonologi, morfologi, sintaksis, semantik, dan pragmatik juga menjadi tema kajian linguistik komputer. Hanya saja, di dalam linguistik komputer semua tema dalam bidang kajian linguistik umum tersebut dikemas dalam kerangka program aplikasi pemrosesan bahasa natural oleh komputer (*Natural Language Processing*). Aturan-aturan yang digunakan oleh komputer di dalam pemrosesan bahasa natural juga didasarkan pada aturan-aturan yang ada di dalam linguistik umum.

d) Linguistik umum penting bagi linguistik komputer karena tema-tema dalam bidang kajian linguistik umum seperti fonologi, morfologi, sintaksis, semantik, dan pragmatik juga menjadi tema kajian linguistik komputer. Hanya saja, di dalam linguistik komputer semua tema dalam bidang kajian linguistik umum tersebut dikemas dalam kerangka program aplikasi pemrosesan bahasa natural oleh komputer (*Natural Language Processing*). Aturan-aturan yang digunakan oleh komputer di dalam pemrosesan bahasa

natural juga didasarkan pada aturan-aturan yang ada di dalam linguistik umum.

Linguistik komputasi bertujuan menciptakan komunikasi interaktif antara manusia dan komputer dalam bahasa yang alami, dan penelitian ini menerapkan prinsip-prinsip tersebut untuk membangun sistem klasifikasi berita otomatis berbasis semantik *frame*. Proses ini melibatkan pemrosesan teks berita dalam bahasa Arab dengan bantuan algoritma pembelajaran mesin untuk mendeteksi konteks semantik dan klasifikasi otomatis yang sesuai.⁹⁹

Dengan demikian, linguistik komputasi dalam penelitian ini memungkinkan penggunaan teori semantik berbasis *frame* dalam analisis dan klasifikasi berita, yang mempercepat pengenalan topik dan kategori dalam jumlah data berita yang besar.

3. Klasifikasi Berita

Klasifikasi adalah proses penyusunan data atau informasi secara sistematis ke dalam kelompok atau kategori berdasarkan kaidah atau standar yang telah ditetapkan.^{100,101} Berita adalah segala sesuatu yang

⁹⁹ Taher Brahim dan Abdelbasset Abdelssamed Benseddik, “Tauzīf Al-Lisāniyyāt Al-Ḥasūbiyyah fī Ṣinā’ah wa Taṭwīr Al-Aṭālas Ar-Raqmiyyah,” *Qaḍāyā Lugawiyyah* 3, no. 1 (2022): 22–36.

¹⁰⁰ Ebta Setiawan, “Arti kata klasifikasi - Kamus Besar Bahasa Indonesia (KBBI) Online,” Kamus Besar Bahasa Indonesia (KBBI) Online, April 2021, <https://kbbi.web.id/klasifikasi>.

¹⁰¹ Al-Jumhūriyyah Al-’Arabiyyah As-Sūriyyah Wizārah Aš-Šaqāfah, *Mukhtaṣar At-Taṣnīf Fī Al-Maktabāt wa Nizām Diwī Al-’Isyrī* (Damaskus: Maktabah Al-Asad, 2011).

ditransmisikan dan dibicarakan,¹⁰² baik secara lisan maupun tulisan.¹⁰³ Sedangkan dalam Kamus Besar Bahasa Indonesia (KBBI), berita adalah cerita, pemberitahuan, pengumuman atau keterangan mengenai kejadian atau peristiwa yang hangat.¹⁰⁴

Menurut para pakar, berita memiliki karakteristik yang membuatnya layak untuk diklasifikasikan. Dean M. Lyle Spencer mendefinisikan berita sebagai ide atau fakta yang memuat kebenaran, menarik minat pembaca, dan relevan bagi publik. Willard C. Bleyer dan William S. Maulsby¹⁰⁵ menambahkan bahwa berita adalah informasi terkini yang diproses oleh wartawan untuk disajikan kepada pembaca secara obyektif, tanpa keberpihakan, serta berisi fakta-fakta penting yang dapat mempengaruhi pemahaman pembaca. Michtel V. Charnley mengungkapkan bahwa berita harus memenuhi kriteria relevansi, kecepatan, dan faktualitas, yakni memberikan laporan yang cepat, menarik, dan sesuai dengan kepentingan pembaca.¹⁰⁶

Berdasarkan definisi-definisi ini, klasifikasi berita dapat dipahami sebagai proses pengkategorian berita ke dalam tema atau topik tertentu, seperti bisnis, hiburan, politik, olahraga, dan teknologi. Proses ini bertujuan

¹⁰² Louis Ma'luf, "Al-Mujid Fī Al-Lugah," dalam *Al-Mujid Fī Al-Lugah wa Al-A'lām*, 29 ed. (Beirut: Dar el-Machref Sarl, 2008), 176.

¹⁰³ *Jumhūriyyah Miṣr Al-'Arabīyyah*, *Al-Mu'jam Al-Wasīṭ*, 4 ed. (Messir: Maktabah Asy-Syurūq Ad-Dauliyyah, 2004).

¹⁰⁴ Ebta Setiawan, "Arti kata berita - Kamus Besar Bahasa Indonesia (KBBI) Online," Kamus Besar Bahasa Indonesia (KBBI) Online, April 2021, <https://kbbi.web.id/berita>.

¹⁰⁵ Totok Djuroto, *Manajemen Penerbitan Pers* (Bandung : Rosdakarya, 2004), 48.

¹⁰⁶ Sayyid Muhammad Habib dkk., "Klasifikasi Berita Menggunakan Metode Naïve Bayes Classifier," *JNKTl: Jurnal Nasional Komputasi dan Teknologi Informasi* 5, no. 2 (April 2022): 248–58.

untuk mengorganisir berita secara lebih efisien sehingga memudahkan pembaca dalam mencari dan memahami informasi yang sesuai dengan minat atau kebutuhannya.¹⁰⁷

Pada penelitian ini, klasifikasi berita akan dilakukan secara otomatis dengan memanfaatkan pendekatan linguistik komputasi dan teori semantik berbasis *frame*. Pendekatan ini memungkinkan sistem untuk mengenali topik berita berdasarkan pola kata dan frasa, sehingga dapat mengelompokkan berita dengan akurat ke dalam kategori tertentu secara efisien dan relevan dengan konteks bahasa Arab. Otomatisasi klasifikasi tidak hanya mempercepat proses pengelompokan, tetapi juga meningkatkan konsistensi dan akurasi dalam mengidentifikasi kategori berita yang besar dan beragam.

F. Metode Penelitian

1. Jenis Penelitian

Jenis penelitian ini adalah penelitian kuantitatif karena melibatkan data berupa nilai pada *vektorisasi* kata menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) dalam menentukan hasil prediksi klasifikasi berita. Desain penelitian ini menggunakan desain eksploratif¹⁰⁸

¹⁰⁷ Yash Veer Singh dkk., “News Classification System using Machine Learning Approach,” dalam *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* (IEEE, 2021), 186–88, <https://doi.org/10.1109/ICAC3N53548.2021.9725409>.

¹⁰⁸ Zahrah Zahrah dan Theresye Yoanita Octora, “Analisis konservasi energi pada dinding kaca ruang tunggu Bandara Ahmad Yani Semarang [Energy conservation analysis of the glass walls in the waiting room of Ahmad Yani Airport Semarang],” *Jurnal Teknik Transportasi* 2, no. 2 (2 November 2023): 115–27, <https://doi.org/10.54324/jtt.v2i2.29>.

dan eksperimental¹⁰⁹ untuk menggali potensi komputasi klasifikasi berita bahasa Arab dengan penggabungan teori semantik berbasis *frame* dan linguistik komputasi. Pendekatan ini memungkinkan eksplorasi mendalam tentang kontribusi masing-masing elemen dalam kerangka konseptual.

2. Subjek dan Objek Penelitian

a. Subjek penelitian (objek material)

Penelitian ini memfokuskan analisis pada berita berbahasa Arab yang tersebar melalui platform pemberitaan *online*. Subjek penelitian adalah entitas yang menjadi sumber data dalam penelitian ini yaitu teks berita yang dihasilkan dalam bahasa Arab dan dipublikasikan secara *online*. Entitas ini kemudian dikumpulkan dan dianalisis dalam penelitian ini. Metode yang digunakan dalam penelitian ini adalah linguistik komputasi, yang mencakup teknik-teknik pemrosesan bahasa alami dan pengolahan informasi yang melibatkan komputer. Selain itu, penelitian ini menitikberatkan pada aspek semantik berbasis *frame*, menunjukkan fokus pada penggunaan kerangka atau struktur semantik tertentu dalam menganalisis berita berbahasa Arab.

¹⁰⁹ Nurul Fatimah dkk., “Analisis konsep fisika osilasi terkopel pada sistem pegas bermassa [Analysis of the physics concept of coupled oscillations in a mass spring system],” *Jurnal Penelitian Fisika dan Terapannya (JUPITER)* 5, no. 1 (31 Juli 2023): 10–15, <https://doi.org/10.31851/jupiter.v5i1.11777>.

b. Objek penelitian (objek formal)

Objek utama penelitian adalah proses klasifikasi berita, yang melibatkan identifikasi dan kategorisasi berita berbahasa Arab ke dalam kelompok-kelompok atau kategori tertentu berdasarkan kontennya. Analisis ini dilakukan pada berita yang disajikan melalui platform pemberitaan *online*, seperti situs web berita atau portal berita, dengan menggunakan pendekatan linguistik komputasi, khususnya dalam konteks semantik berbasis *frame*. Dengan demikian, penelitian ini bertujuan untuk mengembangkan metode klasifikasi otomatis yang efektif untuk berita berbahasa Arab dalam pemberitaan online dengan memanfaatkan pendekatan linguistik komputasi berbasis semantik *frame*. Melalui pendekatan ini, diharapkan dapat tercapai peningkatan efisiensi dan akurasi dalam proses identifikasi dan pengkategorian berita secara otomatis.

3. Sumber Data dan Data Penelitian

a. Sumber data

Sumber data yang akan digunakan dalam penelitian ini berasal dari 6 berita *online* berbahasa Arab yaitu *alsumaria.tv*,¹¹⁰

assahraa.ma,¹¹¹

al-sharq.com,¹¹²

addustour.com,¹¹³

¹¹⁰ Alsumaria, "Akḥbār Al-'Irāq min Qanāh Al-Sūmariyyah Al-'Irāqiyyah," 2004, <https://www.alsumaria.tv/>.

¹¹¹ Assahraa, "Aṣ-Ṣahrā' Al-Magribiyyah," 2015, <https://assahraa.ma/>.

¹¹² Al-Sharq, "Jarīdah Asy-Syarq," 1997, <https://al-sharq.com/>.

¹¹³ Addustour, "Jarīdah Ad-Dustūr Al-Urdūniyyah," 1999, <https://www.addustour.com/>.

akhbarelsyaha.com,¹¹⁴ dan *sarayanews.com*.¹¹⁵ Keenam portal ini dipilih karena mewakili beragam sumber berita utama yang mempublikasikan konten aktual dengan topik yang luas, mencakup isu-isu penting dalam berbagai kategori berita. Setiap portal berita ini digunakan sebagai tempat pengumpulan teks berita dengan metode *web scraping*, yang memungkinkan penarikan data secara otomatis dan terstruktur dari masing-masing situs web.

b. Data penelitian

Data penelitian dalam studi ini terdiri dari kumpulan teks berita yang telah diambil dan diolah dari sumber data tersebut. Data ini mencakup 170,934 artikel berita berbahasa Arab, yang selanjutnya dikategorikan ke dalam 11 topik utama, yaitu: أمن (Keamanan), اقتصاد (Ekonomi), تكنولوجيا (Teknologi), ثقافة (Budaya), دوليات (Internasional), رياضة (Olahraga), سياسة (Politik), صحة (Kesehatan), علم (Sains), سياحة (Travel), dan مجتمع (Masyarakat/Komunitas). Setiap kategori memuat antara 10.000 hingga 20.000 artikel, sehingga total keseluruhan data penelitian berjumlah 42.661.503 kata, yang terdiri atas 754.584 kata dalam

¹¹⁴ Akhbarelsyaha, "Majallah Al-Akhbār As-Siyāḥah," 2018, <https://akhbarelsyaha.com/>.

¹¹⁵ Sarayanews, "Wakālah Anbā' Sarāyā Al-Akhbāriyyah," 2007, <https://www.sarayanews.com/>.

bahasa Arab, 74.727 kata non-Arab, serta 41.832.192 kata yang dikategorikan sebagai *stopwords*.

Proses *preprocessing* dilakukan terhadap data ini guna memastikan kualitas dan konsistensi teks yang digunakan dalam penelitian. Tahapan *preprocessing* meliputi pembersihan teks dari karakter tanda baca, sehingga menghasilkan data yang siap dianalisis. Hasil akhir dari *preprocessing* ini menghasilkan data dalam bentuk terstruktur yang siap untuk diklasifikasikan sesuai dengan tujuan penelitian. Struktur data yang telah diproses dan disiapkan sebagai data penelitian dapat dilihat pada Tabel 1, yang berfungsi sebagai input dalam pengembangan model klasifikasi berbasis teori semantik *frame* pada berita berbahasa Arab.

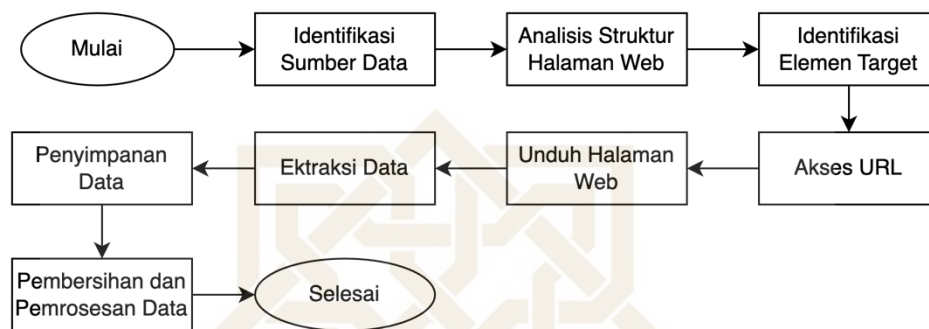
Tabel 1. Struktur Data

No	Y	X ₁	X ₂	...	X ₄₂₆₆₁₅₀₃
1	1	X _{1,1,1}	X _{1,1,2}	...	X ₄₂₆₆₁₅₀₃
2	1	X _{2,1,1}	X _{2,1,3}	...	X ₄₂₆₆₁₅₀₃
3	1	X _{3,1,1}	X _{3,1,4}	...	X ₄₂₆₆₁₅₀₃
4	1	X _{4,1,1}	X _{4,1,5}	...	X ₄₂₆₆₁₅₀₃
⋮	⋮	⋮	⋮	⋮	⋮
170,934	11	X _{152793,11,1}	X _{152793,11,2}	...	X _{152793,11,42661503}

4. Pengumpulan Data

Pengumpulan data dilakukan dengan mengimplementasikan teknik *web scraping* atau menggunakan API yang memungkinkan pengambilan data berita berbahasa Arab secara otomatis dan terstruktur. Implementasi tersebut menggunakan bahasa pemrograman *Python* dengan melibatkan beberapa *module* yang ada di dalamnya seperti *request* dan *sys* dan *library*

BeautifulSoup. Memastikan data yang terkumpul mencakup variasi yang cukup untuk melatih dan menguji model secara efektif. Adapun detail pengumpulan data pada diagram Gambar 3 berikut.



Gambar 3. Diagram teknik *web scrapping*

a. Identifikasi sumber data

Langkah awal dalam proses pengumpulan data adalah identifikasi web berita tersebut terlebih dahulu. Web berita yang akan diterapkan *web scrapping* akan dicek apakah diperbolehkan untuk di-*scrapping* atau tidak. Jika diperbolehkan, maka sudah diperbolehkan untuk analisis stuktur halam pada web tersebut seperti bagaimana pola *url* beritanya, elemen yang berisi judul berita, dan elemen yang berisi konten berita.

b. Identifikasi stuktur halaman web berita dan elemen target

Memeriksa struktur halaman web berita untuk memahami bagaimana informasi disajikan. Kemudian, dilakukan identifikasi pada bagian-bagian penting seperti judul, dan konten berita. Elemen HTML yang berisi data yang ingin diekstrak dipilih setelah proses

identifikasi. Misalnya, jika judul berita berada di dalam elemen `**<h1>**`, identifikasi elemen tersebut sebagai target.

c. Akses URL dan unduh halaman web

Program mengakses URL situs web yang dituju menggunakan modul seperti `requests` atau `urllib`. Konten halaman web diunduh menggunakan metode HTTP GET. Respons HTML dari server diperoleh untuk proses ekstraksi selanjutnya.

d. Ekstraksi data

Gunakan library parsing HTML seperti *BeautifulSoup* atau *lxml* untuk mengekstrak data yang diinginkan dari respons HTML. Misalnya, ekstrak judul, dan konten berita.

e. Penyimpanan data

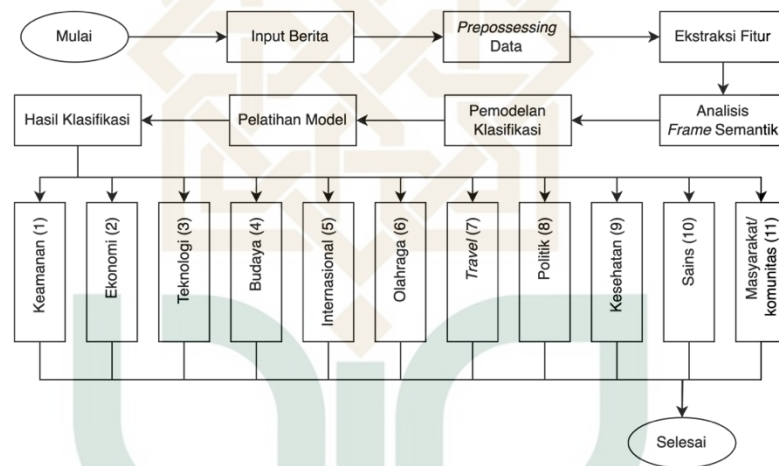
Data yang diekstrak disimpan dalam struktur data yang sesuai, seperti list atau dictionary, untuk mempermudah pengolahan dan pengelolaan data.

f. Pembersihan dan pemrosesan data

Lakukan pembersihan data jika diperlukan. Ini dapat mencakup langkah-langkah seperti penghapusan karakter yang tidak diinginkan, konversi format tanggal, atau pemrosesan tambahan sesuai kebutuhan.

5. Proses Klasifikasi Berita

Langkah pertama melibatkan analisis berita menggunakan Teori Semantik Berbasis *Frame* untuk mengidentifikasi semantik *frame* yang relevan. Selanjutnya, pendekatan linguistik komputasi akan diterapkan untuk melakukan analisis sintaktis dan pembelajaran mesin untuk klasifikasi berita ke dalam kategori yang sesuai. Adapun rincian pengumpulan data pada diagram Gambar 4 berikut.



Gambar 4. Diagram klasifikasi berita

a. *Preprocessing* data

Teks berita diambil dan diproses untuk menghilangkan karakter khusus dan mengonversi huruf kecil agar teks lebih bersih dan dapat diolah lebih lanjut.

b. Ekstraksi fitur semantik

Fitur-fitur diekstrak dari teks berita, yang dapat mencakup informasi seperti kata-kata kunci, frekuensi kata, atau vektor fitur yang mencerminkan makna teks.

c. Analisis semantik *frame*

Analisis *frame* semantik pada teks berita dapat dilakukan untuk memahami struktur semantik yang lebih dalam, seperti entitas, hubungan, dan *frame* semantik lainnya.

d. Pemodelan klasifikasi

Model klasifikasi, seperti *Naïve Bayes*, *Support Vector Machine*, atau model *deep learning*, dipilih untuk melakukan klasifikasi. Model ini dapat menerima input fitur dan menghasilkan output klasifikasi.

e. Pelatihan model

Terdapat keputusan (*decision*) untuk melatih model menggunakan data pelatihan yang telah diberikan sebelumnya. Pelatihan model sangat penting untuk meningkatkan akurasi klasifikasi berita berbahasa Arab. Dengan melatih model pada data pelatihan, model dapat memahami pola bahasa dan konteks khusus, termasuk pola semantik *frame*, sehingga mampu mengklasifikasikan berita secara otomatis dan tepat sesuai kategori yang diinginkan.

f. Klasifikasi berita

Model yang sudah dilatih digunakan untuk mengklasifikasikan berita berdasarkan fitur yang telah diekstrak. Proses ini menghasilkan kategori atau label yang menunjukkan klasifikasi berita.

g. *Output* klasifikasi

Hasil klasifikasi ditampilkan kepada pengguna. Ini memberikan informasi tentang bagaimana berita tersebut diklasifikasikan berdasarkan model klasifikasi yang telah diterapkan.

6. Analisis Data

Data hasil klasifikasi akan dianalisis secara kuantitatif dan kualitatif. Pendekatan ini melibatkan evaluasi akurasi klasifikasi, identifikasi pola makna, dan interpretasi kontekstual dalam berita. Hasil analisis akan digunakan untuk mengukur efektivitas penggabungan teori semantik dan linguistik komputasi dalam klasifikasi berita Bahasa Arab. Analisis data dilakukan dengan memanfaatkan teknik-teknik pemrosesan bahasa alami (NLP) dan metode-metode klasifikasi yang sesuai dengan tujuan penelitian.

a. Penerapan teori semantik berbasis *frame*

Pertama, analisis data akan melibatkan penerapan Teori Semantik Berbasis *Frame* pada setiap berita. Identifikasi semantik *frame* yang terlibat akan dilakukan dengan memetakan makna kata

atau frasa ke dalam kerangka konseptual yang relevan. Proses ini akan menghasilkan representasi struktural yang memungkinkan pemahaman kontekstual makna dalam setiap berita.

b. Analisis linguistik komputasi

Langkah selanjutnya melibatkan analisis linguistik komputasi, termasuk pemrosesan bahasa alami dan penerapan model pembelajaran mesin. Analisis sintaktis akan digunakan untuk memahami struktur kalimat, sedangkan model pembelajaran mesin akan melibatkan klasifikasi berita ke dalam kategori berdasarkan pola makna yang ditemukan. Hasil analisis ini akan memberikan pemahaman tentang kemampuan linguistik komputasi dalam konteks klasifikasi berita.

c. Evaluasi akurasi klasifikasi

Hasil klasifikasi dari Teori Semantik dan linguistik komputasi akan dievaluasi untuk mengukur tingkat akurasi. Perbandingan hasil klasifikasi dengan kategori asli berita akan memberikan gambaran tentang sejauh mana penggabungan Teori Semantik dan linguistik komputasi dapat meningkatkan akurasi klasifikasi berita Bahasa Arab.

d. Analisis pola dan kontekstual

Analisis ini akan mencakup identifikasi pola makna yang dihasilkan oleh integrasi Teori Semantik dan linguistik komputasi.

Fokus pada aspek-aspek semantik dan kontekstual dalam berita akan memberikan wawasan tentang kemampuan pendekatan ini dalam memahami nuansa dan variasi makna dalam Bahasa Arab.

e. Interpretasi makna

Analisis data akan diakhiri dengan interpretasi hasil secara keseluruhan dalam konteks Bahasa Arab. Kesimpulan akan ditarik berdasarkan akurasi klasifikasi, identifikasi pola makna, dan pemahaman kontekstual. Hal ini akan memberikan pemahaman mendalam tentang potensi dan batasan penggabungan Teori Semantik dan linguistik komputasi dalam klasifikasi berita Bahasa Arab.

G. Sistematika Penulisan

Sistematika penulisan penelitian “Komputasi Klasifikasi pada Berita *Online* Berbahasa Arab dengan Pendekatan Semantik Berbasis *Frame*” dirinci dengan cermat untuk memastikan alur penulisan yang logis dan kohesif. Berikut adalah rincian lebih lanjut:

Bab I Pendahuluan berisi latar belakang dan permasalahan yang menjadi fokus kajian, tujuan dan manfaat penelitian, tinjauan pustaka, landasan teori, metode penelitian, dan sistematika pembahasan.

Bab II Tren *Dataset* Berita Berbahasa Arab berisi informasi pola yang ditemukan dalam *dataset* berita berbahasa Arab. Statistik yang disajikan mencakup

pengelompokan berdasarkan kategori berita serta karakteristik yang menonjol dalam *dataset*.

Bab III Bentuk Semantik *Frame* dalam Berita Berbahasa Arab berisi paparan yang membahas bentuk-bentuk semantik *frame* yang dihasilkan dalam berita berbahasa Arab melalui proses klasifikasi. Pembahasan mencakup bagaimana model klasifikasi mengenali dan mengelompokkan berita berdasarkan pola semantik yang muncul.

Bab IV Efisiensi dan Efektivitas Teori Semantik Berbasis *Frame* dalam Proses Klasifikasi Berita Berbahasa Arab berisi evaluasi performa model klasifikasi berita berbahasa Arab, dengan fokus pada efisiensi dan efektivitas penerapan teori semantik berbasis *frame* dalam meningkatkan akurasi dan kinerja klasifikasi otomatis.

Bab V Penutup berisi kesimpulan akhir dari penelitian dan saran.

Daftar Pustaka berisi daftar lengkap referensi dan sumber-sumber yang digunakan selama penelitian. Adapun penyajian daftar sesuai dengan gaya penulisan referensi yang diikuti yaitu *Chicago Manual of Style 17th edition (full note)*.

BAB V

PENUTUP

A. Kesimpulan

Penelitian ini telah berhasil mencapai tujuan utama yang dirumuskan, yaitu mengembangkan model klasifikasi berita *online* berbahasa Arab menggunakan pendekatan semantik berbasis *frame*. Pendekatan ini terbukti mampu meningkatkan efisiensi dan akurasi dalam pengelompokan berita ke dalam berbagai kategori, seperti keamanan, ekonomi, budaya, teknologi, internasional, olahraga, sains, kesehatan, politik, *travel*, dan masyarakat/komunitas. Model *Support Vector Machine* (SVM) yang digunakan dalam penelitian ini menunjukkan performa yang optimal dengan hasil evaluasi akurasi, *precision*, *recall*, dan *F1-score* yang baik. Penerapan teori semantik berbasis *frame* secara efektif menangkap makna tematik dalam berita dan mengelompokkan berita sesuai dengan kategori yang relevan.

Penelitian ini menjawab pertanyaan yang diajukan dengan hasil sebagai berikut. Pertama, analisis dataset menunjukkan adanya pola kata kunci dan frasa penting yang mencerminkan kategori berita. Kedua, *frame* semantik yang dihasilkan berhasil menangkap makna dan konteks secara mendalam, memperkuat pemetaan semantik pada tiap kategori. Ketiga, pendekatan semantik berbasis *frame* terbukti efisien dan efektif dalam meningkatkan akurasi dan kinerja model klasifikasi.

Selain itu, hipotesis yang diajukan bahwa penerapan teori semantik berbasis *frame* dapat meningkatkan efisiensi dan akurasi klasifikasi berita berbahasa Arab telah diterima. Berdasarkan hasil evaluasi, penggunaan teori ini memberikan keunggulan signifikan dibandingkan metode klasifikasi konvensional. Penerapan semantik *frame* memungkinkan pengelompokan berita yang lebih akurat dan efektif, terutama dalam menangani makna yang tersembunyi dalam teks berita.

Secara teoritis, penelitian ini memberikan kontribusi penting bagi pengembangan kajian semantik *frame* dalam konteks linguistik komputasi dan NLP, khususnya untuk bahasa Arab yang kompleks. Pendekatan ini memberikan pemahaman yang lebih mendalam tentang cara mengelola dan menganalisis makna dalam berita *online*. Secara praktis, hasil penelitian ini dapat diterapkan dalam sistem klasifikasi otomatis berita *online* berbahasa Arab, memberikan solusi yang efektif untuk mengelola volume berita yang besar dengan lebih terstruktur dan akurat.

Namun, penelitian ini juga memiliki beberapa keterbatasan. Kompleksitas bahasa Arab, terutama variasi dialek dan morfologi yang sangat kaya, menuntut sumber daya komputasi yang lebih besar dan metode yang lebih canggih. Penelitian ini juga terbatas pada penggunaan data dari berita berbahasa Arab standar (Fushhā), sementara variasi dialek lokal belum ditangani secara khusus.

B. Saran

Berdasarkan hasil penelitian, terdapat beberapa saran yang dapat diajukan untuk pengembangan lebih lanjut. Pertama, perlu adanya penanganan variasi dialek bahasa Arab, mengingat penelitian ini hanya berfokus pada bahasa Arab standar

(Fuṣḥā). Memasukkan dialek lokal akan membantu model klasifikasi mengakomodasi keragaman bahasa yang lebih luas. Kedua, penggunaan metode NLP yang lebih canggih, seperti *deep learning*, disarankan untuk meningkatkan akurasi dalam menangani kompleksitas bahasa Arab. Ketiga, penelitian mendatang sebaiknya menggunakan *dataset* yang lebih besar dan beragam agar model dapat diuji lebih luas dan menunjukkan generalisasi yang lebih baik. Keempat, penggunaan sumber daya komputasi yang lebih kuat diperlukan untuk menangani kompleksitas data dan meningkatkan efisiensi pemrosesan. Terakhir, evaluasi model klasifikasi pada berbagai domain berita, seperti keamanan, ekonomi, budaya, teknologi, internasional, olahraga, sains, kesehatan, politik, *travel*, dan masyarakat/komunitas, penting dilakukan untuk memperluas penerapan model dalam berbagai jenis konten berita. Dengan mengimplementasikan saran-saran ini, diharapkan penelitian selanjutnya dapat menghasilkan model yang lebih akurat dan efisien dalam klasifikasi berita berbahasa Arab.

DAFTAR PUSTAKA

- Abū Bakar, Maryam. “Qaḍāyā wa Taḥḍiyāt Az-ḡakā’ Al-Iṣṭinā’ī fī Al-Lisāniyyāt Al-‘Arabīyah.” Dalam *Al-Mu’tamir Ad-Daulī Liṭalabah Ad-Dirāsāt Al-‘Ulyā fī Ta’līmī Al-Lugah Al-‘Arabīyah wal Ādāb wal Lisāniyat 2023*, 286–97. Universitas Malang, 2023.
- Abualsaud, M., R. Elshawī, O. Alfarraj, dan M. Alshayeb. “A deep learning model for Arabic news classification.” *Journal of King Saud University-Computer and Information Sciences* 30, no. 4 (2018): 408–17.
- Addustour. “Jarīdah Ad-Dustūr Al-Urdūniyyah,” 1999. <https://www.addustour.com/>.
- Akhbarelsyaha. “Majallah Al-Akhbār As-Siyāḥah,” 2018. <https://akhbarelsyaha.com/>.
- Aliyah, Hikmah Himmatul, Roni Nugraha Syafroni, dan Suntoko Suntoko. “Analisis Deiksis Sosial pada Teks Berita Media Daring Detik News Seputar Covid-19.” *Jurnal Educatio FKIP UNMA* 8, no. 1 (2022): 22–26. <https://doi.org/10.31949/educatio.v8i1.1458>.
- Al-Jumhūriyyah Al-‘Arabīyyah As-Sūriyyah Wizārah Aṣ-Ṣaqāfah. *Mukhtaṣar At-Taṣnīf Fī Al-Maktabāt wa Niẓām Diwī Al-‘Isyrī*. Damaskus: Maktabah Al-Asad, 2011.
- Alroobaea, Roobaea. “An Empirical Deep Learning Approach for Arabic News Classification.” (*IJACSA*) *International Journal of Advanced Computer Science and Applications* 14, no. 6 (2023): 1048–55. www.ijacsa.thesai.org.
- Al-Sharq. “Jarīdah Asy-Syarq,” 1997. <https://al-sharq.com/>.
- Alsumaria. “Akhbār Al-‘Irāq min Qanāh Al-Sūmariyyah Al-‘Irāqiyyah,” 2004. <https://www.alsumaria.tv/>.
- Andayani, Sri, Indra Tjahyadi, dan Hosnol Wafa. “Analisis Wacana Kritis Model Van Dijk Dalam Pemberitaan Krisis Energi di Media Daring cnbndonesia.com.” Dalam *Conference on Innovation and Application of Science and Technology (CIASTECH)*, 177–86, 2021. <https://doi.org/https://doi.org/10.31328/ciastech.v0i0>.
- Ariyadi, Ade Dufadhol, dan Asep Purwo Yudi Utomo. “Analisis Kesalahan Sintaksis pada Teks Berita Daring berjudul Mencari Etika Elite Politik di saat Covid-19.” *Jurnal Bahasa dan Sastra* 8, no. 3 (27 Desember 2020): 138–45. <https://doi.org/10.24036/jbs.v8i3.110903>.

- Assahraa. "Aş-Şahrā' Al-Magribiyyah," 2015. <https://assahraa.ma/>.
- Augustin, Rina Maulina, Isah Cahyani, dan Dadang Anshori. "Analisis Kesalahan Berbahasa Dalam Bidang Morfologi Pada Teks Berita Siswa SMPIT." Dalam *Seminar Internasional Riksa Bahasa XIII*, 1497–1506. Universitas Pendidikan Indonesia, 2020.
- Balebassi, Muhammad, dan Rabha Adad. "Aşar Al-Lugah Al-'Arabiyyah Fī Tarsīkh Muqawwamāt Al-Hawiyyah Al-Waṭaniyyah." *Dirāsāt Mu'āşirah* 6, no. 2 (2022): 12–20.
- Bangun, Eric Persadanta, Ferry V.I. A Koagouw, dan J.S. Kalangi. "Analisis Isi Unsur Kelengkapan Berita Pada Media Online Manadopostonline.com." *Acta Diurna Komunikasi* 1, no. 3 (2019).
- Belqasm, Iman Fatima Al-Zahra. "Al-Jumlah Al-Fi'liyyah Al-Başīṭah Al-Mušbatah Fī Mu'allaqah Imri'i Al-Qais." *Majallah Kuliyyah Al-Ādāb wa Al-'Ulūm Al-Insāniyyah wa Al-Ijtīmā'īyyah* 9, no. 4 (2011): 71–99.
- Benfettacha, Ali. "Min Adillah Irtibāt An-Nahw Al-'Arabī Bi Al-'Arabiyyah Al-Fuṣḥā." *Majallah Isykālāt fī Al-Lugah wa Al-Adab* 10, no. 4 (2021): 561–77.
- Bolshakov, Igor A., dan Alexander Gelbukh, ed. *Computational linguistics: Models, resources, applications*. Mexico: Ciencia De La Computación, 2004.
- Bouhouche, Ganiah. "Tarjamah Ma'ānī Al-Qur'ān Al-Karīm Ilā Lugāt Ukhrā Ḍarūrah Da'wiyyah wa Mas'ūliyyah Syar'iyyah." *El-Mouzhir* 1, no. 2 (2020): 182–92.
- Br Tarigan, Anna Juliana, Muhammad Surif, Malan Lubis, Mara Untung Ritonga, dan Wisman Hadi. "Analisis Wacana Kritis Model Teun A. Van Dijk pada Teks Berita di Buku Bahasa Indonesia Kelas VII Berbasis Kurikulum Merdeka Terbitan Kemendikbudristek Tahun 2021." *Basastra* 12, no. 1 (2023): 17–34. <https://doi.org/10.24114/bss.v12i1.40645>.
- Brahim, Taher, dan Abdelbasset Abdelssamed Benseddik. "Tauzīf Al-Lisāniyyāt Al-Ḥāsūbiyyah fī Şinā'āh wa Taṭwīr Al-Aṭālas Ar-Raqmiyyah." *Qadāyā Lugawiyyah* 3, no. 1 (2022): 22–36.
- Bumdin, Bilkhatir. "Aşar Ta'rīb Al-Muṣṭalahāt Al-Qānūniyyah Al-'Ālamiyyah 'Alā Al-Fikr Al-Islāmī." *Majallah Ad-Dirāsāt Al-Islāmiyyah* 1, no. 1 (2012): 230–38.
- Chanakot, Benjamin, dan Charun Sanrach. "Classifying thai news headlines using an artificial neural network." *Bulletin of Electrical Engineering and Informatics* 12, no. 1 (1 Februari 2023): 395–402. <https://doi.org/10.11591/eei.v12i1.4228>.

- Croft, William, dan D. Alan Cruse. *Cognitive Linguistics*. Cambridge University Press, 2004. <https://doi.org/10.1017/CBO9780511803864>.
- Dede Hamzah, dan Reza Firdaus Tsani. “Analisis Semantik Terhadap Teks Berita ‘KKP Eksekui RJ Lino ke Lapas Cipinang’ Pada Media Online RMOL Jabar Terbitan 4 November 2022.” *Ta’rim: Jurnal Pendidikan dan Anak Usia Dini* 3, no. 3 (2022): 113–26. <https://doi.org/10.59059/tarim.v3i3.53>.
- Djuroto, Totok. *Manajemen Penerbitan Pers*. Bandung : Rosdakarya, 2004.
- Doğan, Nuh. “Formal and Semantic Hierarchy of Turkish Transitive Verbs.” *İctimaiyat* 6, no. 2 (2022): 688–703. <https://doi.org/10.33709/ictimaiyat.1195290>.
- Dyah Hapsari, Intan Kumala, Riska Warni Harahap, Alviansyah Bonde, dan Ibnu Adnan Cahya. “Analisis Kesalahan Frasa Pada Teks Berita Covid-19 Koran Digital goriau.com.” *Lingua Rima: Jurnal Pendidikan Bahasa dan Sastra Indonesia* 10, no. 2 (2021): 35–42. <https://doi.org/10.31000/lgrm.v10i2.4740>.
- Evans, Vyvyan, dan Melanie Green. *Cognitive linguistics: An introduction*. Lawrence Erlbaum Associates Publishers, 2006.
- Fadillah, Putri, dan Jatmika Nurhadi. “Analisis Wacana Kritis Teun A. Van Dijk pada Teks Berita KPK Respons Wacana Periksa Anies di Kasus Korupsi Munjul.” *Jurnal Bahasa, Sastra, dan Pengajarannya* 7, no. 2 (2021): 78–85.
- Fatihah, Nurul, Nikma Nurul Khomsati, Talitha Sahda Widyadhari, dan Khafidh Nur Aziz. “Analisis konsep fisika osilasi terkopel pada sistem pegas bermassa [Analysis of the physics concept of coupled oscillations in a mass spring system].” *Jurnal Penelitian Fisika dan Terapannya (JUPITER)* 5, no. 1 (31 Juli 2023): 10–15. <https://doi.org/10.31851/jupiter.v5i1.11777>.
- Fillmore, C. J., dan B. T. Atkins. “Toward a Frame-Based Lexicon: The Semantics of RISK and Its Neighbors.” Dalam *Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organization*, disunting oleh A. Lehrer dan E. F. Kittay, 75–102. Hillsdale, NJ: Lawrence Erlbaum Associates, 1992.
- Fillmore, Charles J. “An Alternative to Checklist Theories of Meaning.” Dalam *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, 123–31, 1975.
- . “Frame Semantics.” Dalam *Linguistics in the Morning Calm: Selected Papers from SICOL-1981*, disunting oleh The Linguistics Society of Korea, 111–37. Seoul: Hanshin Publishing Company, 1982.
- Frasticha, Lussy Yuris, dan Farid Pribadi. “Bingkai Demokratisasi Isu Pengesahan UU Cipta Kerja (Analisis Framing Zhongdang Pan Dan Gerald M. Kosicki Pada Media Dalam Jaringan Kompas.Com Dan Tribunnews.Com).” *Jurnal*

- Komunikasi dan Media* 2, no. 1 (2022): 1–9.
<https://doi.org/10.24167/jkm.v2i1.3206>.
- Gordon, Faith. “The significance and impact of the media in contemporary society.” Dalam *Children, young people and the press in a transitioning society*, 17–46. London: Palgrave Macmillan UK, 2018. https://doi.org/10.1057/978-1-137-60682-2_2.
- Guan, Yong, Shaoru Guo, Ru Li, Xiaoli Li, dan Hu Zhang. “Frame Semantics guided network for Abstractive Sentence Summarization.” *Knowledge-Based Systems* 221 (Juni 2021): 106973.
<https://doi.org/10.1016/j.knosys.2021.106973>.
- Gupta, Varun, dan Ekta Gupta. “Punjabi news multi-classification using language generation-based optimized long short-term memory networks.” *Evolving Systems* 14, no. 1 (8 Februari 2023): 49–58. <https://doi.org/10.1007/s12530-022-09428-2>.
- Hamid, Abdul Salam, dan Ibrahim Sya‘lan. “Anmāt Al-Jumlah Al-Ismiyyah Fī Namāzīj Min Jarīdah Ar-Rāyah Al-Qaṭriyyah: Dirāsah Naḥwiyyah Iḥṣā’iyyah.” *Hauliyyāt Al-Ādāb Wa Al-Lugah* 10, no. 1 (2022): 9–35.
- Harnia, Neng Tika, Ferina Meliasanti, dan Hendra Setiawan. “Analisis Framing Berita Perundungan pada Media Online Detik.Com dan Tribunnews.Com sebagai Bahan Ajar Teks Berita di SMP.” *Edukatif: Jurnal Ilmu Pendidikan* 3, no. 5 (17 Agustus 2021): 3145–53.
<https://doi.org/10.31004/edukatif.v3i5.1240>.
- Hausser, Roland. *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language*. Berlin; Heidelberg; New York; Barcelona; Hong Kong: Springer, 2001.
- Ha-Yeong, Lee. “Semantic Analysis of ‘弄’ through Frame Semantic Approach.” *Chinese Studies* 82, no. 1 (31 Maret 2023): 121–38.
<https://doi.org/10.14378/KACS.2023.82.82.7>.
- Hendriyanto, Muhammad Diki, dan Betha Nurina Sari. “Penerapan Algoritma K-Nearest Neighbor dalam Klasifikasi Judul Berita Hoax.” *Jurnal Ilmiah Informatika* 10, no. 02 (15 September 2022): 80–84.
<https://doi.org/10.33884/jif.v10i02.5477>.
- Hussain, Md Gulzar, Babe Sultana, Mahmuda Rahman, dan Md Rashidul Hasan. “Comparison analysis of Bangla news articles classification using support vector machine and logistic regression.” *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 21, no. 3 (1 Juni 2023): 584–91.
<https://doi.org/10.12928/telkomnika.v21i3.23416>.

- Isbah, Faliqul, dan M. Fairuz Rosyid. “Amaliyyah al-Andragogiya li Mu'allim al-Lughah al_juhūd li Tarqiyyah al-Kafā'ah al-Tarbawiyah.” *Jurnal Alfazuna : Jurnal Pembelajaran Bahasa Arab dan Kebahasaaraban* 5, no. 02 (18 Juni 2021): 170–90. <https://doi.org/10.15642/alfazuna.v5i02.1042>.
- Islām, Muḥammad Saif. “Al-Lisāniāt Al-Hasūbiyah wa Isykalāt Al-Manhaj wal Anẓimah fī Mīzānil Baḥṣi-Al-Mu'ālajah Taḥlīliyah li Ru'yi 'Amaliyyah 'Arabiyyah Mutamayyizah.” *Al-Mumarrasāt Al-Lugawiyyāh* 11, no. 2 (2020): 41–63.
- Jamīlah, Guemmaz. “Al-Lisāniyāt Al-Hasūbiyah: Mafhūmuhā Manhajuhā wa Majallātu Istikhdamihā.” *University of JIJEL* 8, no. 2 (2022): 8–17.
- Jarah, Muhammad Ibrahim. “Aṣar At-Tarjamah Fī Tagyīri Al-Ma'nā wa Nasyr Al-Ma'lūmāt Az-Zā'ifah Fī Mawāqī'i At-Tawāṣul Al-Ijtimā'i Dirāsah Taṭbīqiyyah 'Alā Al-Lugah Al-'Arabiyyah.” *Sémiotiques* 18, no. 2 (2023): 438–54.
- Johnson, Tamara M., dan Simone E. Pfenninger. “Tasty words: Using frame semantics to enhance consumer liking of potato chips and apples.” *International Journal of Applied Linguistics* 31, no. 1 (25 Maret 2021): 79–94. <https://doi.org/10.1111/ijal.12323>.
- Juhfa, Abdul Majid Al-. *Madkhal Ilā 'Ilmi Ad-Dalālah Al-Hadīṣ*. Casablanca: Dar Toubkal, 2000.
- Jumhūriyyah Miṣr Al-'Arabiyyah. *Al-Mu'jam Al-Wasīf*. 4 ed. Miṣr: Maktabah Asy-Syurūq Ad-Dauliyyah, 2004.
- Kamil Ali, Saba, dan Rana H. Al-Bahrani. “A Cognitive Linguistic Study of the Persuasive/Attention Strategies Used in Selected American English E-Advertisements.” *Arab World English Journal* 7, no. 1 (15 Juli 2021): 279–98. <https://doi.org/10.24093/awej/call7.20>.
- Karisna, Deksen. “Analisis Unsur-Unsur Kelengkapan Berita dalam Teks Berita Siswa MTs. Muhammadiyah Lebung Itam.” *Wahana Didaktika : Jurnal Ilmu Kependidikan* 18, no. 1 (28 Januari 2020): 95–102. <https://doi.org/10.31851/wahanadidaktika.v18i1.4359>.
- Karwati, Dini, dan Wulansari Wulansari. “Analisis Frasa Berdasarkan Golongan Kata Terhadap Teks Berita 'Saat Internet Jadi Kambing Hitam Terhapusnya Rekaman CCTV Tragedi Kanjuruhan.'” *PENTAS: Jurnal Ilmiah Pendidikan Bahasa dan Sastra Indonesia* 8, no. 2 (23 November 2022): 27–34. <https://doi.org/10.52166/pentas.v8i2.3512>.
- Khoiruddin, Achmad, dan Hendra Setiawan. “Analisis Framing Berita Korupsi Maskapai Garuda Indonesia pada Media Online Cnbcindonesia.com dan

- Sindonews.com sebagai Bahan Ajar Teks Berita di SMP.” *Edukatif: Jurnal Ilmu Pendidikan* 4, no. 1 (30 Desember 2021): 778–85. <https://doi.org/10.31004/edukatif.v4i1.2005>.
- Kim, Jiwon. “Eine Analyse der Werbung als Werbungskonstruktionen - Interaktion zwischen Konstruktionsgrammatik und Frame-Semantik.” *Journal for German Culture and Literature* 28 (31 Desember 2019): 117–41. <https://doi.org/10.32681/JGCL.28.5>.
- Koloski, Boshko, Timen Stepišnik Perdih, Marko Robnik-Šikonja, Senja Pollak, dan Blaž Škrli. “Knowledge graph informed fake news classification via heterogeneous representation ensembles.” *Neurocomputing* 496 (Juli 2022): 208–26. <https://doi.org/10.1016/j.neucom.2022.01.096>.
- Komang, Ni, Widya Santi, Wayan Wendra, dan I Gede Nurjaya. “Analisis Komponen Teks Berita Karya Siswa Kelas VIII SMP Negeri Satu Atap 1 Tejakula.” *Jurnal Pendidikan Bahasa dan Sastra Indonesia Undiksha* 11, no. 2 (2021): 175–84. <https://doi.org/https://doi.org/10.23887/jjpbs.v11i2.35282>.
- Kumar Dutta, Ashit, Basit Qureshi, Yasser Albagory, Majed Alsanea, Manal Al Faraj, dan Abdul Rahaman Wahab Sait. “Optimal Weighted Extreme Learning Machine for Cybersecurity Fake News Classification.” *Computer Systems Science and Engineering* 44, no. 3 (2023): 2395–2409. <https://doi.org/10.32604/csse.2023.027502>.
- Lakoff, George, dan Mark Johnson. *Al-Isti‘ārāt Allatī Nahyā Bihā*. Disunting oleh Abdul Majid Juhfa. Casablanca, Maroko: Dār Tuqbal Lin Nasyr, 2009.
- Lange, Lukas, Omar Alonso, dan Jannik Strötgen. “The Power of Temporal Features for Classifying News Articles.” Dalam *Companion Proceedings of The 2019 World Wide Web Conference*, 1159–60. New York, NY, USA: ACM, 2019. <https://doi.org/10.1145/3308560.3315000>.
- Maćkowiak, Maria, Agnieszka Libura, Lyn Phillipson, Dorota Szcześniak, dan Joanna Rymaszewska. “Understanding of Dementia in the Polish Language: A Frame Semantic Approach.” *Journal of Alzheimer’s Disease* 91, no. 1 (3 Januari 2023): 389–406. <https://doi.org/10.3233/JAD-220633>.
- Mahendra, B.H., A. Adiwijaya, dan U.N. Wisesty. “Kategorisasi Berita Multi-Label Berbahasa Indonesia Menggunakan Algoritma Random Forest.” *e-Proceeding of Engineering* 6, no. 2 (2019): 9030–41.
- Mardikantoro, Hari Bakti, dan Haryadi Haryadi. “Korupsi Dalam Konstruksi Media : Analisis Struktur Mikro Semantik Teks Berita Korupsi Di Televisi.” *Widyaparwa* 47, no. 1 (26 Agustus 2019): 104–15. <https://doi.org/10.26499/wdprw.v47i1.195>.

- Ma'luf, Louis. "Al-Mujid Fī Al-Lugah." Dalam *Al-Mujid Fī Al-Lugah wa Al-A'lām*, 29 ed. Beirut: Dar el-Machref Sarl, 2008.
- Meena, Aminah, Imane Chacha, dan Hisamuddin Taouririt. "At-Ta'rīf Al-Ālī wa Al-Mu'ālajah Ar-Raqmiyyah Li Al-Khaṭ Al-'Arabī Baina Al-Ḥad Al-Lugawī wa Aḍ-Ḍabṭ At-Taqnī." *Majallah Ta'līmiyyah* 13, no. 3 (2023): 495–503.
- Mehdivi, Umar. *Taulīd al-asmā' min al-juzūr as-sulāsiyyah aṣ-ṣaḥīḥah fī al-lugah al-'Arabiyyah - muqārabah lisāniyyah ḥāsūbiyyah [The formation of names from the correct trilateral roots in the Arabic language - a computational linguistic approach]*. Disunting oleh Abdul Gani Abu Al- 'Azm. Vol. 1. Casablanca: Wiḥdah 'Ulūm Al-Lugah Al-'Arabiyyah wa Al-Mu'jāmiyyāt, 2008.
- Mohseb, Mahyaddin. *Al-Idrākiyyāt Ab'ād Istimūlūjiyyah Wajhāt Taṭbīqīyyah*. 1 ed. Amman, Jordan: Dār Kunūz Al-Ma'rifah, 2017.
- Mousa, Nihad Al-. *Al-Lugah Al-'Arabiyyah Naḥw Tauṣīf Jadīd fī Ḍau' Al-Lisāniyyāt Al-Ḥāsūbiyyah*. 1 ed. Beirut: Al-Mu'assisah Al-'Arabiyyah Li Ad-Dirāsāt wa An-Nasyr, 2000.
- Muhammad Habib, Sayyid, Elin Haerani, Siska Kurnia Gusti, dan Siti Ramadhani. "Klasifikasi Berita Menggunakan Metode Naïve Bayes Classifier." *JNKTI: Jurnal Nasional Komputasi dan Teknologi Informasi* 5, no. 2 (April 2022): 248–58.
- Muhmali, Usama. "Āliyāt Al-Mustasyriqīn Li Hadm Al-Lugah Al-'Arabiyyah Al-Fuṣḥā." *Cahiers du Laboratoire la Poétique Algérienne* 2, no. 1 (2017): 28–42.
- Nerbonne, John. "Natural language processing in computer-assisted language learning." Dalam *The oxford handbook of computational linguistics (Oxford handbooks)*, disunting oleh Ruslan Mitkov, 2 ed., 634–98. USA: Oxford University Press, Inc., 2022.
- Pennington, Jeffrey, Richard Socher, dan Christopher Manning. "Glove: Global vectors for word representation." Dalam *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, disunting oleh Alessandro Moschitti, Bo Pang, dan Walter Daelemans, 1532–43. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014. <https://doi.org/10.3115/v1/D14-1162>.
- Petruck, Miriam R.L. "Frame semantics." Dalam *Handbook of Pragmatics*, 1–13. Amsterdam: John Benjamins Publishing Company, 1997. <https://doi.org/10.1075/hop.2.fra1>.

- Prasetyo, Indrawan Dwi. "Frame Semantics in Naming Military Vehicles." *Language Horizon* 6, no. 1 (2018): 28–39.
- Prawira, I Made Riartha, Adiwijaya, dan Mohamad Syahrul Mubarak. "Klasifikasi Multi-Label Pada Topik Berita Berbahasa Indonesia Menggunakan Multinomial Naïve Bayes." *e-Proceeding of Engineering* 5, no. 3 (2018): 7774–81.
- Rosch, Eleanor. "Principles of categorization." Dalam *Cognition and Categorization*, disunting oleh Eleanor Rosch dan Barbara Lloyd, 27–48. Hillsdale, NJ: Erlbaum, 1999.
- Sally, Shehatta Usamah. "Aẓ-Ẓakā' Al-Iṣṭinā'ī Wa Al-Taḥqīqiyāt Wa Al-Adawāt Ar-Raqmiyyah Al-Mutakhaṣṣah Fī Al-Kasyfī 'An Al-Akhbār Az-Zā'ifah." *RAKMANA* 3, no. 2 (2023): 173–85.
- Santoso, Sofiana. "Analisis Resepsi Audiens Terhadap Berita Kasus Meiliana di Media Online." *Komuniti : Jurnal Komunikasi dan Teknologi Informasi* 12, no. 2 (22 Januari 2021): 140–54. <https://doi.org/10.23917/komuniti.v12i2.13285>.
- Sarayanews. "Wakālah Anbā' Sarāyā Al-Akhbāriyyah," 2007. <https://www.sarayanews.com/>.
- Saussure, Ferdinand de. *Course in General Linguistics*. New York: Columbia University Press, 2011.
- Schafroth, Elmar, dan Riccardo Imperiale. "Gebrauchsbasierte Phraseologie des Italienischen: Digitale Lexikographie zwischen Frame-Semantik und Konstruktionsgrammatik." *Lexicographica* 35, no. 2019 (1 Desember 2019): 133–67. <https://doi.org/10.1515/lex-2019-0004>.
- Setiawan, Ebta. "Arti kata berita - Kamus Besar Bahasa Indonesia (KBBI) Online." Kamus Besar Bahasa Indonesia (KBBI) Online, April 2021. <https://kbbi.web.id/berita>.
- . "Arti kata klasifikasi - Kamus Besar Bahasa Indonesia (KBBI) Online." Kamus Besar Bahasa Indonesia (KBBI) Online, April 2021. <https://kbbi.web.id/klasifikasi>.
- Shao, Dangguo, Chengyao Li, Chusheng Huang, Yan Xiang, dan Zhengtao Yu. "A news classification applied with new text representation based on the improved LDA." *Multimedia Tools and Applications* 81, no. 15 (15 Juni 2022): 21521–45. <https://doi.org/10.1007/s11042-022-12713-6>.
- Sholih 'afif, Muhammad, Muhammad Muzakir, Moh Iqbal Al, dan Ghifari Al Awalaen. "Text Mining Untuk Mengklasifikasi Judul Berita Online Studi

- Kasus Radar Banjarmasin Menggunakan Metode Naïve Bayes.” *Kumpulan Jurnal Ilmu Komputer (KLIK)* 8, no. 2 (2021): 199–208.
- Singh, Yash Veer, Piyush Naithani, Parvez Ansari, dan Pragya Agnihotri. “News Classification System using Machine Learning Approach.” Dalam *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 186–88. IEEE, 2021. <https://doi.org/10.1109/ICAC3N53548.2021.9725409>.
- Strobel, Jochen, Jan de Vries, dan Friederike Wißmach. “Leserkognition und Wissensemergenz bei kontrafaktischem Erzählen aus Sicht der Frame-Semantik.” *Zeitschrift für Literaturwissenschaft und Linguistik* 49, no. 1 (12 Maret 2019): 139–62. <https://doi.org/10.1007/s41244-019-00120-8>.
- Su, Xuefeng, Ru Li, Xiaoli Li, Baobao Chang, Zhiwei Hu, Xiaoqi Han, dan Zhichao Yan. “A Span-based Target-aware Relation Model for Frame-semantic Parsing.” *ACM Transactions on Asian and Low-Resource Language Information Processing* 22, no. 3 (31 Maret 2023): 1–24. <https://doi.org/10.1145/3569581>.
- Susiawati, Wati. “Kajian Bahasa Arab dari A Historis hingga Historis.” *Alfaz (Arabic Literatures for Academic Zealots)* 7, no. 01 (18 Juni 2019): 43. <https://doi.org/10.32678/alfaz.Vol7.Iss01.1925>.
- Szulc-Brzozowska, Magdalena. “Patriotismus – eine framebasierte kontrastive Analyse Deutsch-Polnisch anhand der Boulevardzeitung Bild und ihres polnischen Pendants Fakt.” *Roczniki Humanistyczne* 69, no. 5 (16 Juni 2021): 181–97. <https://doi.org/10.18290/rh21695-10>.
- Talitha, Stella, dan Willi Ferdianto. “dAnalisis Kesalahan Berbahasa Pada Teks Berita Daring Dan Implikasinya Terhadap Pembelajaran Bahasa Indonesia SMP.” *Jurnal Pendidikan Bahasa dan Sastra Indonesia* I, no. 1 (2022): 82–90.
- Tomasello, Michael. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press, 2005. <https://doi.org/10.2307/j.ctv26070v8>.
- Trueman, Tina Esther, Ashok Kumar J., Narayanasamy P., dan Vidya J. “Attention-based C-BiLSTM for fake news detection.” *Applied Soft Computing* 110 (Oktober 2021): 107600. <https://doi.org/10.1016/j.asoc.2021.107600>.
- Vogel, Carl (Dir.). *Computer Science, Linguistics and a Language: Course Handbook 2006-2007*. Dublin: University of Dublin, 2006.
- Wittgenstein, Ludwig. *Philosophische Untersuchungen*. 3 ed. Harlow, London: Prentice Hall, 1958.

- Yahia, Salahuddin, dan Lamia Kedache. “Al-Lisāniyyāt Al-‘Urfāniyyah wal Muḥṭawā Al-Ijrāi Li Nazriyyah Dalālah Al-Aṭar Fī Al-Madkhal Al-Mu‘jāmiyyāh.” *Dirāsāt Al-Ma‘āṣirah* 5, no. 7 (2021): 249–64.
- Yang, Hsin Chang, dan Yu Chih Wang. “Content enrichment using linked open data for news classification.” *Journal of Internet Technology* 21, no. 5 (2020): 1397–1407. <https://doi.org/10.3966/160792642020092105015>.
- Yani, Tri Andra, Moh. Shofiuddin Shofi, dan Adelia. “Analisis Kesalahan Berbahasa Dalam Kumpulan Teks Berita Daring Radar Tegal.” *DIALEKTIKA Pendidikan Bahasa Indonesia* 1, no. 2 (2021): 60–70.
- Yu, Lei, dan Yang Xu. “Noun2Verb: Probabilistic Frame Semantics for Word Class Conversion.” *Computational Linguistics* 48, no. 4 (1 Desember 2022): 783–818. https://doi.org/10.1162/coli_a_00447.
- Zahrah, Zahrah, dan Theresye Yoanita Octora. “Analisis konservasi energi pada dinding kaca ruang tunggu Bandara Ahmad Yani Semarang [Energy conservation analysis of the glass walls in the waiting room of Ahmad Yani Airport Semarang].” *Jurnal Teknik Transportasi* 2, no. 2 (2 November 2023): 115–27. <https://doi.org/10.54324/jtt.v2i2.29>.
- Zaitouni, Abdullah, dan Latroush Charef. “Itālah Binā’ Al-Jumlah ‘An Ṭarīq At-Ta‘āqib Fī Syi‘ri Abī Abdullah Albū‘abdalī.” *Djoussour El-maarefa* 5, no. 1 (2019): 459–69.
- Zaneed, Lazhar. “Al-Muta‘alliq Bi Murātib Al-Ittisā’ Fī Ad-Dilālah Al-Mu‘jāmiyyah.” Dalam *Fuṣūl Fī Ad-Dilālah Mā Baina Al-Mu‘ajjim Wa An-Naḥwi*. Beirut: Ad-Dār Al-‘Arabiyyah Lil ‘Ulūm Nasyirūn Wa Mansyūrāt Al-Ikhtilāf, 2010.