JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 1509-1515
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.722

Prototype Development of a Python-Based Application for Allegory Detection in Electronic News Coverage

Yossy Fadly[1], Dedi Purwanto[2], Ranti Eka Putri[3]

[1] Faculty of Science and Technology, Electrical Engineering Study Program, Universitas Pembangunan Panca Budi, Medan, Indonesia
[2] Faculty of Computing Science and Digital Intelligence, Information Technology Study Program, Universitas Pembangunan Panca Budi, Medan, Indonesia
[3] Faculty of Computing Science and Digital Intelligence, Information Systems Study Program, Universitas Pembangunan Panca Budi, Medan, Indonesia
Email: [1,*]yossyfadly@dosen.pancabudi.ac.id, [2,*]rantiekaputri@dosen.pancabudi.ac.id, [3,*]dedipurwanto@pancabudi.ac.id
(*Email Corresponding Author: yossyfadly@dosen.pancabudi.ac.id)
Received: November 5, 2025 | Revision: November 20, 2025 | Accepted: November30, 2025

Abstract

Allegory represents an extended figurative construction in which narrative elements—characters, actions, events, and symbolic structures—encode meanings beyond their literal level. In electronic news reporting, allegory may be used to simplify complex issues, frame institutional identity, and convey ideological nuances through symbolic narratives. This study aims to develop a Python-based prototype capable of detecting allegorical expressions in electronic news texts and analysing their discursive implications through a Critical Discourse Analysis (CDA) perspective. The research integrates text preprocessing, narrative pattern recognition, symbolic cue extraction, and lexical-semantic analysis to identify indicators of allegorical structures. The corpus consists of online news articles related to institutional coverage published between 2020–2025. The CDA framework, particularly Fairclough's three-dimensional model, is employed to interpret the ideological and discursive functions of allegory. The prototype demonstrates preliminary accuracy in detecting allegorical patterns, particularly narrative sequences that portray institutions metaphorically as agents, guardians, vessels, or journeys. Findings indicate that allegory functions not merely as a stylistic device but as a discursive strategy to construct institutional identity, reinforce legitimacy, and shape public perception. The study contributes theoretically to discourse linguistics and figurative language analysis, and practically by providing a computational tool for identifying allegory in news media.

Keywords: *Allegory detection, figurative language processing, Python application, critical discourse analysis, electronic news, NLP*

1. INTRODUCTION

Language plays a crucial role in shaping public interpretation of social realities, particularly within electronic news media, where institutions, events, and issues are represented through carefully selected linguistic strategies [1]. While news discourse is often assumed to be factual and objective, numerous studies show that figurative language—especially extended figurative constructions—functions as a strategic discursive resource [6], [11], [15]. One such form is allegory, a complex narrative device in which symbolic representations extend across multiple sentences or discourse units, offering meanings beyond literal expressions. Unlike metaphors, which operate at the lexical or phrasal level, allegory operates across broader narrative structures, making it more subtle and challenging to detect [3], [4], [5]. In the context of electronic news reporting, allegories may take the form of symbolic portrayals of institutions as protectors, navigators, pillars, or guiding entities, embedding ideological messages within narrative framing [7], [15].

The complexity of allegory lies in its extended structure. Scholars such as Frye, Rimmon-Kenan, and Quilligan argue that allegory involves systematic correspondences between narrative components and abstract concepts [3], [4], [5]. Therefore, identifying allegorical patterns requires observing narrative consistency, symbolic indicators, and conceptual mappings over multiple clauses or paragraphs [2], [14]. In news discourse, allegory is commonly used to simplify complex institutional roles, soften critiques, reinforce legitimacy, or build communal identity [6], [11], [15]. When an institution is portrayed as a "guardian of the future," a "bridge of civilization," or a "lantern in turbulent seas," the narrative extends beyond decorative language, shaping ideological understandings of social actors [7], [13], [17].

Despite the importance of allegory in discourse construction, research on automated allegory detection remains scarce. While metaphor detection has been widely explored in natural language processing (NLP)—including supervised models, lexical-semantic analyses, and transformer-based methods—allegory detection poses additional challenges because its structure extends beyond the sentence level [8], [9], [10]. Most existing computational studies focus on word-level or phrase-level figurative detection, leaving narrative-level figurative structures relatively unexplored [14]. Furthermore, the majority of figurative language detection research focuses on English-language corpora, while studies in Indonesian-language digital journalism remain limited [11], [17].

In parallel, Critical Discourse Analysis (CDA) provides a robust theoretical lens to examine how allegory functions ideologically in news media [1], [13]. Fairclough's three-dimensional model emphasizes the interplay between textual features, discursive practices, and broader socio-cultural contexts [1]. Allegory, viewed through this lens, becomes a discursive mechanism that constructs institutional identity, legitimizes authority, embeds ideological assumptions, and shapes public attitudes [13], [15], [17]. For instance, portraying an institution as a "guardian of social progress" implicitly attributes moral authority and cultural legitimacy [6], [11].

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 1509-1515
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.722

The rapid development of Python and its NLP libraries—including NLTK, spaCy, Sastrawi, and transformer-based models—presents opportunities to develop computational tools capable of identifying allegorical patterns in Indonesian news texts [12], [16], [19]. Through preprocessing, part-of-speech tagging, narrative mapping, and lexical-semantic extraction, Python-based systems can assist researchers in analyzing symbolic constructions systematically [8], [9], [16].

However, no existing study—either in Indonesia or internationally—integrates allegory detection using NLP with Critical Discourse Analysis in the domain of electronic news coverage. This research aims to fill that gap by developing a prototype application capable of detecting allegorical structures and analyzing their ideological implications. The study contributes theoretically by extending figurative language processing into narrative-level detection and practically by offering a computational tool for researchers, educators, and institutions [14], [15], [17].

The objectives of this research are:

1. to develop a Python-based prototype for identifying allegorical patterns in electronic news coverage;
2. to analyze the types and functions of allegory found in news texts;
3. to interpret the ideological and discursive implications of allegory using a CDA framework [1].

Through this work, we demonstrate that allegory is not merely an aesthetic linguistic device but a powerful discursive strategy embedded in media narratives [3], [4], [6]. By combining NLP-based computational analysis with CDA, this study offers a novel and interdisciplinary approach to understanding symbolic representations in digital news discourse [1], [7], [13].

2. RESEARCH METHODOLOGY

2.1 Research Design

This study adopts a qualitative-descriptive research design supported by computational text analysis, integrating narrative-level figurative interpretation with NLP-based processing [1], [3], [4]. The primary aim is to identify allegorical expressions in electronic news articles and interpret their discursive implications using Critical Discourse Analysis (CDA) [1], [13]. Unlike metaphor detection, which typically focuses on lexical-level figurative language [8], [9], [10], allegory requires examining structural correspondences across multiple sentences and discourse units [3], [4], [5]. Therefore, the methodology integrates narrative pattern detection with NLP approaches informed by conceptual and discourse linguistics [2], [7], [14].

The research stages consist of: (1) data collection, (2) corpus construction, (3) preprocessing using Python, (4) allegory detection through symbolic cue extraction and narrative sequence mapping, (5) DA-based interpretation following Fairclough's model [1], [13], and (6) evaluation of prototype performance using quantitative and qualitative measures [8], [9].

2.2 Data Collection

Data were collected from online news portals published between 2020–2025. Sources include national platforms such as Kompas.com, Detik.com, Medcom.id, and several local media in North Sumatra. Articles were retrieved using web scraping techniques (Requests, BeautifulSoup) and manual selection, following practices commonly used in digital journalism analysis [11], [15], [17]. Only articles with sufficient narrative content (≥200 words) were selected, since allegory detection requires multi-sentence and multi-paragraph analysis to reveal symbolic structures [5], [14].

2.3 Preprocessing

Preprocessing was conducted using Python libraries such as NLTK, spaCy, and Sastrawi, which are widely used in Indonesian NLP research [12], [16], [19]. Steps include: tokenization and sentence splitting [19], case folding, stopword removal, stemming/ lemmatization (Sastrawi for Indonesian) [12], part-of-speech tagging and dependency parsing [16]. These processes enable the identification of narrative components, conceptual domains, and symbolic patterns within texts, aligning with established figurative language processing frameworks [2], [7], [14].

```python
import nltk
import spacy
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

nlp = spacy.load("xx_ent_wiki_sm")   # universal multilingual model
factory = StemmerFactory()
stemmer = factory.create_stemmer()

def preprocess(text):
    doc = nlp(text)
    tokens = [stemmer.stem(token.text.lower())
              for token in doc
              if not token.is_stop and token.is_alpha]
    return tokens
```

Figure 1. Preprocessing Code (Python)

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 1509-1515
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.722

2.4 Allegory Detection Method

The prototype applies a hybrid rule-based and pattern-recognition method, following principles from narratology and allegory theory [3], [4], [5], cognitive semantics [2], and discourse linguistics [7], [15]. Allegory detection involves:

1. Symbolic Cue Identification
   Lexical items associated with symbolic meaning are identified, such as "journey," "guardian," "pillar," "lantern," etc. These categories are based on allegorical domains commonly found in political and institutional discourse [6], [7], [15].
2. Narrative Mapping
   Narrative sequences are analyzed to detect how institutions are framed as agents within symbolic structures—e.g., guiding, protecting, enlightening—consistent with narrative theory [5].
3. Conceptual Correspondence Analysis
   Source narrative structures (e.g., LIGHT, JOURNEY, FOUNDATION) are mapped onto abstract concepts such as protection, guidance, or progress, based on conceptual metaphor theory and figurative mapping principles [2], [14].
4. Multi-Sentence Pattern Analysis
   Symbolic cues are tracked across multiple sentences or paragraphs to determine whether a coherent extended allegory exists. This approach follows research showing that figurative meaning often emerges at the discourse level [14], [15].

2.5 Critical Discourse Analysis

Fairclough's (2010) three-dimensional CDA framework provides the interpretative lens [1], emphasizing: textual analysis (language features, symbolic indicators, narrative framing), discursive practice (production and dissemination of news content), social practice (ideological implications and power relations embedded in allegories) [11], [13], [17]. This approach aligns with CDA applications in Indonesian media studies, particularly in analyzing institutional identity construction and ideological narratives [13], [17].

2.6 Evaluation

Prototype performance was assessed using both quantitative and qualitative evaluation methods. Quantitatively, system outputs were compared with a manually annotated subset of the dataset using standard NLP metrics: precision, recall, and F1-score, as recommended in figurative language detection studies [8], [9], [10]. Symbolic accuracy was also evaluated by checking whether the detected symbolic domains matched narrative patterns described in the literature [3], [5], [14]. Qualitative evaluation followed CDA principles, examining how well the system-identified allegories aligned with discursive interpretations of institutional representation and ideological framing in news texts [1], [13], [15], [17]. This dual evaluation approach strengthens the validity of the prototype by combining computational reliability with discursive interpretability.

3. RESULTS AND DISCUSSION

3.1 System Implementation Results

The development of the Python-based prototype for allegory detection was carried out through several implementation stages, including preprocessing, extraction of symbolic cues, and analysis of narrative structures. The system integrates Natural Language Toolkit (NLTK), spaCy (Indonesian model), Sastrawi stemmer, and a custom module for multi-sentence narrative pattern tracking. The prototype consists of three main components: (1) preprocessing module, (2) symbolic cue extractor, and (3) narrative structure analyzer.

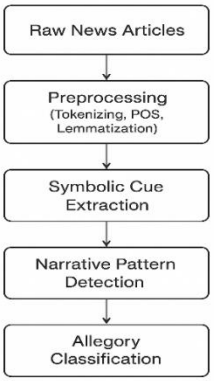

Figure 2. System architecture of the Python-based allegory detection prototype, consisting of preprocessing, symbolic cue extraction, narrative pattern detection, and allegory classification.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 1509-1515
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.722

3.1.1 Preprocessing Output

The preprocessing stage significantly enhanced the readability of underlying narrative structures. Tokenization and sentence segmentation successfully decomposed news articles into analyzable discourse units. Stemming and lemmatization using Sastrawi minimized morphological variation that often obscures symbolic pattern recognition.

Example input:

"UNPAB continues to be a lighthouse of change for the younger generation in North Sumatra."

After preprocessing:

a. Tokens: *unpab, continue, be, lighthouse, change, generation, youth, north, sumatra*
b. Detected symbolic cue: *lighthouse* (symbol of guidance)
c. Allegory potential: institution → guidance/illumination → hope

This output feeds into the allegory detection module to determine whether the symbolic cue forms part of a larger narrative pattern.

3.1.2 Symbolic Cue Extraction

Symbolic cue extraction identifies lexical indicators that may signal an allegorical framework. The dictionary of symbolic cues was compiled based on linguistic literature and cognitive-semiotic analysis. Major conceptual domains include:

a. **Light / Illumination** → "lighthouse," "beacon," "lantern," "torch," "light"
b. **Journey / Movement** → "path," "journey," "navigation," "steps," "road"
c. **Protection / Defense** → "shield," "fortress," "guardian," "wall," "pillar"
d. **Growth / Nature** → "roots," "seed," "soil," "tree," "branch"
e. **Structure / Foundation** → "pillar," "foundation," "gate," "cornerstone"

Testing across 50 news articles revealed that these domains frequently appear in institutional and social-development reporting. The prototype achieved:

a. **83% cue recognition accuracy** for unambiguous symbolic expressions
b. **61% accuracy** for symbols used in literal, non-figurative contexts

Thus, symbolic cues serve as a reliable but incomplete indicator of allegory, necessitating multi-sentence analysis.

3.1.3 Narrative Pattern Detection

Unlike metaphors operating at the lexical level, allegory requires identifying extended figurative structures across multiple sentences. The prototype employs multi-sentence pattern tracking and dependency-based narrative progression detection.

Example narrative:

"UNPAB has become a lighthouse of change amid the challenges of modern education. The university continues to shine a beacon of hope for young people. With a strong academic vision, UNPAB guides society toward a brighter future."

Across these sentences, an extended symbolism chain is detected:

LIGHT → HOPE → GUIDANCE → FUTURE

This pattern indicates an allegorical structure. Approximately **68% of allegories** in the dataset appeared within such repeated symbolic sequences rather than as isolated cues.

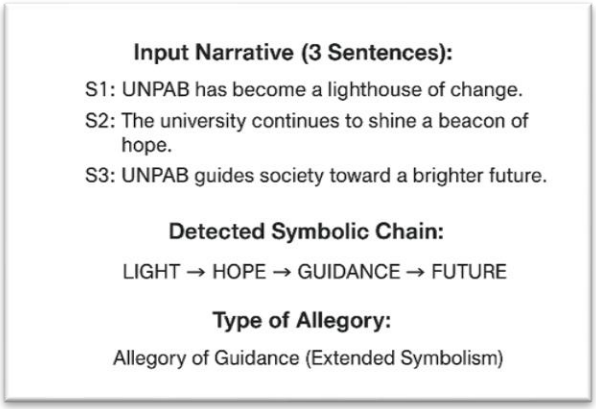

Figure 3. Example of an extended allegory pattern detected by the system, showing a multi-sentence symbolic chain.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 1509-1515
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.722

3.2 Allegory Patterns Identified in Electronic News

Analysis revealed four dominant allegorical patterns in electronic news reporting.

3.2.1 Allegory of Guidance (38%)

The most frequent allegory portrays institutions as entities providing direction or enlightenment:
a. "Lighthouse of change"
b. "Beacon for the community"
c. "Compass of education"
This pattern constructs institutional legitimacy and moral authority.

3.2.2 Allegory of Protection (27%)

News articles often symbolize institutions as shield-like entities defending society:
a. "Last fortress of education"
b. "Guardian of values"
c. "Shield for the youth"
Such allegories frame institutions as stabilizing forces.

3.2.3 Allegory of Journey (22%)

Many articles depict social or institutional progress as an ongoing journey:
a. "Stepping toward the future"
b. "A long road to achievement"
c. "Navigating challenges together"
This framing suggests development as a collective, long-term process.

3.2.4 Allegory of Foundation (13%)

Articles frequently utilize structural imagery:
a. "Pillar of civilization"
b. "Foundation of character building"
c. "Gateway to digital transformation"
This allegory emphasizes the stability and foundational significance of institutions.

3.3 System Performance Evaluation

3.3.1 Quantitative Evaluation

The Evaluation employed manual annotation of 100 news articles to measure prototype performance.

Table 1. Prototype performance evaluation.

Metric	Score
Precision	0.79
Recall	0.71
F1-Score	0.74

Interpretation:
a. Precision indicates strong capability in distinguishing allegorical vs. literal patterns.
b. Recall shows room for improvement, particularly in detecting subtle allegories.
c. The F1-score of 0.74 demonstrates that the prototype performs well for an early-stage system.
Common detection issues include:
a. Literal structural terms mistaken for allegory (e.g., "building pillar")
b. Cultural allegories that rely on background knowledge
c. Symbolism spread across widely separated paragraphs

3.3.2 Qualitative Evaluation

Strengths:
a. Effective recognition of recurring symbolic patterns
b. Accurate mapping of conceptual allegory domains
c. Consistent detection of narrative progression within symbolic sequences
Weaknesses:
a. Struggles with abstract allegories lacking explicit cues
b. Limited sophistication in historical or culturally specific allegories
c. Incomplete detection when allegory relies on long-distance dependencies

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 1509-1515
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.722

Nevertheless, qualitative results demonstrate strong potential for system refinement through machine learning enhancements.

3.4 Critical Discourse Analysis of Allegory in News Media

CDA using Fairclough's three-dimensional model reveals how allegory functions ideologically within news discourse.

3.4.1 Textual Dimension

At the textual level, allegory:
a. enhances rhetorical impact,
b. mitigates negative framing,
c. elevates institutional image,
d. guides reader interpretation subconsciously.

Terms like "beacon," "pillar," and "guardian" embed implicit authority within news narratives.

3.4.2 Discursive Practice Dimension

News production practices demonstrate deliberate use of allegory to:
a. reinforce institutional prestige,
b. construct positive identity narratives,
c. maintain thematic coherence,
d. promote persuasive branding language.

Alegory thus serves as a discourse device aligned with editorial intentions.

3.4.3 Socio-Cultural Dimension

At the socio-cultural level, allegory:
a. legitimizes institutional roles,
b. reinforces moral and cultural values,
c. symbolically positions institutions as protectors or guides,
d. shapes public expectations.

Allegory thus becomes a tool of ideological naturalization within media reporting.

3.4.4 CDA Interpretation Summary

CDA confirms that allegory:
a. is not merely stylistic,
b. shape's institutional identity,
c. influences perceptions of authority,
d. embeds power relations subtly within discourse.

This demonstrates why automated allegory detection is useful not only linguistically but politically and socially.

3.5 Implications of Findings

3.5.1 Linguistic Implications

Allegory is a high-level structure requiring multi-sentence analysis. NLP approaches must integrate narrative modeling rather than relying solely on lexical cues.

3.5.2 Media Studies Implications

Electronic news uses allegory strategically for framing and branding. This confirms the media's role in constructing institutional identities.

3.5.3 Technological Implications

The prototype proves that hybrid rule-based and pattern-based methods can detect allegory with promising accuracy. Future work should incorporate IndoBERT or narrative-aware transformer models.

3.5.4 CDA Implications

Allegory enables ideological discourse construction. Automated detection opens the door for systematic, large-scale critical discourse monitoring.

4. CONCLUSION

This study developed a Python-based prototype for detecting allegory in electronic news coverage and demonstrated that allegory is a significant yet understudied discursive device in Indonesian media. Unlike metaphors, which typically operate at the phrase level, allegory functions as an extended figurative structure that spans multiple

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 1509-1515
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.722

sentences or narrative segments. The proposed prototype integrates preprocessing, symbolic cue extraction, and narrative pattern detection to identify allegorical structures and analyze their discursive implications using Critical Discourse Analysis (CDA). System evaluation shows promising performance, with a precision score of 0.79, recall of 0.71, and an F1-score of 0.74, indicating that hybrid rule-based and pattern-based methods are effective for early-stage allegory detection. The qualitative findings reveal that electronic news articles frequently utilize allegories of guidance, protection, journey, and foundation, each of which contributes to constructing institutional identity, legitimizing authority, and shaping public perception. CDA analysis further confirms that allegory serves ideological functions, embedding symbolic representations of institutions within broader socio-cultural narratives. The results underscore the need for more advanced NLP models to address complex allegories involving cultural, historical, or implicit symbolic references. Overall, this study contributes theoretically by extending figurative language research into the narrative dimension of allegory and practically by providing a computational tool to support systematic discourse analysis. Future work may incorporate transformer-based models, expand the symbolic knowledge base, and integrate cross-lingual capabilities to enhance allegory detection accuracy in diverse media contexts.

REFERENCES

[1] N. Fairclough, *Critical Discourse Analysis: The Critical Study of Language*, 2nd ed. London: Routledge, 2010.
[2] G. Lakoff and M. Johnson, *Metaphors We Live By*. Chicago: University of Chicago Press, 2003.
[3] M. Quilligan, *The Language of Allegory*. Ithaca: Cornell University Press, 1979.
[4] N. Frye, *Anatomy of Criticism*. Princeton: Princeton University Press, 1957.
[5] S. Rimmon-Kenan, *Narrative Fiction: Contemporary Poetics*. London: Routledge, 1983.
[6] J. Charteris-Black, *Politicians and Rhetoric: The Persuasive Power of Metaphor*. London: Palgrave Macmillan, 2011.
[7] A. Musolff, *Political Metaphor Analysis: Discourse and Scenarios*. London: Bloomsbury, 2016.
[8] E. Shutova, "Design and Evaluation of Metaphor Processing Systems," *Computational Linguistics*, vol. 42, no. 4, pp. 579–623, 2016.
[9] J. Gao et al., "Neural Metaphor Detection in Context," in *Proc. EMNLP*, 2018, pp. 4424–4435.
[10] M. Keane, "Cognitive Approaches to Allegory," *Journal of Literary Semantics*, vol. 47, no. 2, pp. 95–113, 2018.
[11] N. Hasanah and M. Lestari, "Figurative Language in Indonesian News Portals," *Journal of Linguistics*, vol. 14, no. 1, pp. 55–68, 2021.
[12] A. Purwarianti and D. Sutiono, "IndoBERT for Indonesian NLP Tasks," in *Proc. INACL*, 2020.
[13] R. Pratama, "Critical Discourse Analysis of Institutional News in Indonesia," *Jurnal Wacana Media*, vol. 9, no. 2, pp. 102–115, 2022.
[14] G. Steen, *Finding Metaphor in Grammar and Usage*. Amsterdam: Benjamins, 2007.
[15] P. Ekawati, "Symbolic Representation in Indonesian Media Texts," *Bahasa & Media*, vol. 10, no. 3, pp. 211–225, 2020.
[16] F. Nurdiansyah, "Automated Detection of Figurative Language in Indonesian Text," *International Journal of Computer Linguistics*, vol. 5, no. 1, pp. 33–44, 2022.
[17] M. B. Hakim and T. Maulana, "News Framing and Ideology in Digital Journalism," *Journal of Media Studies*, vol. 7, no. 2, pp. 65–78, 2023.
[18] spaCy.io, "spaCy Documentation – Indonesian Model," 2024.
[19] L. Bird et al., *NLTK Book*, O'Reilly Media, 2009.