



APLIKASI PENDETEKSI PLAGIARISME BERBASIS MOBILE DENGAN METODE NATURAL LANGUAGE PROCESSING

Hudaimi

Informatika, Fakultas Teknologi Informasi, Universitas Teknologi Digital Indonesia
Email: hudaimi@gmail.com

Abstrak

Deteksi plagiarisme memiliki peran krusial dalam menjaga integritas akademis dan kejujuran intelektual di dunia pendidikan tinggi. Salah satu cara yang digunakan untuk mendeteksi plagiarisme adalah dengan menggunakan *natural language processing*. Dalam bidang akademik, salah satu syarat kelulusan untuk mendapat gelar akademik adalah membuat karya tulis, seperti skripsi, thesis dan disertasi. Akan tetapi tidak jarang, karya tulis yang dihasilkan tersebut adalah hasil plagiat. Plagiat adalah penjiplakan atau pengambilan karangan, pendapat, dan sebagainya dari orang lain dan menjadikannya seolah-olah karangan dan pendapat sendiri. Untuk pengecekan manual dengan jumlah tersebut membutuhkan waktu lama, oleh karena itu dibutuhkan aplikasi yang dapat membantu proses pengecekan agar lebih efisien. Tujuan dari penelitian ini adalah membuat aplikasi berbasis web yang dapat mendeteksi plagiat terhadap karya tulis menggunakan *natural language processing*.

Kata kunci : Karya Tulis, Plagiarisme, *Natural Language Processing* (NLP)

Abstract

Plagiarism detection has a crucial role in maintaining academic integrity and intellectual honesty in the world of higher education. One method used to detect plagiarism is by using natural language processing. In the academic field, one of the graduation requirements to obtain an academic degree is to write written work, such as a thesis, thesis and dissertation. However, it is not uncommon for the written work produced to be plagiarized. Plagiarism is plagiarizing or taking other people's essays, opinions, etc. and making them appear to be your own essays and opinions. Manual checking of this amount takes a long time, therefore an application is needed that can help make the checking process more efficient. The aim of this research is to create a web-based application that can detect plagiarism in written works using natural language processing.

Keywords : Written Work, Plagiarism, *Natural Language Processing* (NLP)

Article History

Received: Agustus 2024
Reviewed: Agustus 2024
Published: Agustus 2024

Plagiarism Checker No 234

Prefix DOI : Prefix DOI :
10.8734/Kohesi.v1i2.365

Copyright : Author

Publish by : Kohesi



This work is licensed
under a [Creative
Commons Attribution-
NonCommercial 4.0
International License](https://creativecommons.org/licenses/by-nc/4.0/)



1. PENDAHULUAN

Maraknya perkembangan teknologi informasi, khususnya Internet, telah memberikan banyak kemudahan bagi penggunaanya dalam mengakses berbagai sumber daya. Kemudahan ini sangat bermanfaat bagi civitas akademika, karena memudahkan pengembangan ilmu pengetahuan dengan menyediakan metode akses publikasi penelitian yang cepat dan mudah. Terbukanya akses informasi bagi dunia pendidikan sangat bermanfaat bagi para akademisi (termasuk mahasiswa, dosen, dan peneliti) dalam memperlancar kegiatan pengajaran, penyiapan karya ilmiah, dan penelitian. Persyaratan kelulusan bagi mahasiswa antara lain menyelesaikan penelitian dan menulis makalah akademik, tesis, atau disertasi. Akses mudah terhadap teknologi informasi berpotensi membantu siswa memenuhi tanggung jawab akademiknya. Namun dalam praktiknya, kemudahan ini sering disalahgunakan dalam bentuk plagiarisme yang meluas. Yang mengecewakan, tidak hanya mahasiswa tingkat sarjana, tetapi bahkan pada tingkat master dan doktoral, menjadi sasaran tren yang mengecewakan ini. Permasalahan ini menjadi perhatian besar di kalangan akademisi. Permasalahan ini sebagian besar bermula dari tersebar luasnya ketersediaan dokumen skripsi di internet.

Karya tulis dapat diartikan sebagai sebuah hasil olah pikir manusia yang dituangkan ke dalam tulisan. Karya tulis dibagi menjadi 2 jenis, yaitu karya tulis fiksi dan karya tulis non fiksi atau ilmiah. Karya tulis fiksi adalah karya tulis yang merupakan rekayasa, khayalan, atau hasil imajinasi manusia. Dalam bidang akademik, salah satu syarat kelulusan untuk mendapat gelar akademik adalah membuat karya tulis, seperti: skripsi, thesis dan disertasi. Akan tetapi, tidak sedikit mahasiswa yang karya tulisnya adalah hasil plagiat (pencontek). Plagiat adalah penjiplakan atau pengambilan karangan, pendapat, dan sebagainya dari orang lain dan menjadikannya seolah-olah karangan dan pendapat sendiri.

Isu plagiarisme menjadi sumber kekhawatiran yang mendalam dalam menjaga integritas etika akademis dan kejujuran intelektual. Tindakan plagiarisme tidak hanya mencuri karya orang lain, tetapi juga merusak esensi kejujuran intelektual yang merupakan pondasi dari kemajuan ilmiah dan akademis. Dampaknya yang lebih luas mencapai kredibilitas lembaga pendidikan, karena tindakan plagiarisme mempertanyakan integritas proses pendidikan dan penelitian yang seharusnya menjadi landasan pembangunan pengetahuan yang sah. Hal ini mengancam kepercayaan masyarakat pada lembaga-lembaga tersebut dan mengarah pada keraguan tentang legitimasi ilmiah mereka. Oleh karena itu, penanggulangan plagiarisme tidak hanya penting untuk melindungi hak cipta dan penghargaan terhadap karya intelektual, tetapi juga untuk menjaga integritas dan kredibilitas lembaga pendidikan sebagai lembaga pembentuk dan penjaga kebenaran ilmiah

Meminimalisir isu plagiarisme dalam penulisan tugas akhir dilakukan dengan menerapkan sanksi kepada mahasiswa yang terbukti melakukan tindakan plagiarisme, sebagaimana diatur dalam Pasal 10 ayat (4). Sanksi ini diberlakukan secara berturut-turut mulai dari yang ringan hingga yang paling berat, termasuk teguran, peringatan tertulis, penundaan pemberian sebagian hak mahasiswa, pembatalan nilai satu atau beberapa mata kuliah, pemberhentian dengan hormat dari status sebagai mahasiswa, pemberhentian dengan tidak hormat dari status sebagai mahasiswa, dan bahkan pembatalan ijazah apabila mahasiswa telah lulus dari suatu program Selain menerapkan undang-undang, upaya penanggulangan plagiarisme telah melahirkan banyak perangkat lunak atau aplikasi yang



membantu dalam mengoreksi tulisan mahasiswa untuk mendeteksi kecurigaan plagiarisme. Beberapa contoh perangkat lunak atau aplikasi.

Salah satu alternatif untuk menanggulangi tindakan plagiarisme adalah penerapan sistem komputasi dengan metode *Natural Language Processing*. Metode *Natural Language Processing* menjanjikan tingkat akurasi yang sangat tinggi dalam mendeteksi plagiarisme dengan membandingkan kesamaan antara dokumen berdasarkan vektor representasi teksnya. Dengan menggunakan pendekatan ini, sistem dapat mengidentifikasi sejauh mana sebuah tulisan mirip dengan sumber lain, sehingga membantu dalam menanggulangi tindakan plagiarisme dengan lebih efektif.

Perkembangan teknologi telah menjadi kebutuhan dasar manusia dan merubah kondisi sosial terutama di dunia pendidikan (Escobar, 2016). Pemanfaatan teknologi dapat dilakukan dalam berbagai bidang kehidupan, tidak terkecuali di bidang pendidikan. Perguruan tinggi sebagai salah satu bagian dari sistem pendidikan harus mampu merespon perubahan lingkungan dan perubahan teknologi yang pesat (Azan et al., 2015). Persaingan yang semakin kompetitif sudah seharusnya perguruan tinggi meningkatkan kinerja pelayanan dengan menggunakan teknologi (Effendi & Tasrif, 2019).

Teknologi yang memungkinkan dapat diterapkan pada bidang pendidikan yaitu *artificial intelligence* atau kecerdasan buatan (Pratikno, 2017). Pangsa pasar memperkirakan bahwa bidang pendidikan merupakan 10 industri teratas dalam penerapan kecerdasan buatan antara tahun 2018 hingga 2025 (Tractica, 2019). Teknologi tersebut memungkinkan mesin untuk berpikir dan mengambil keputusan secara komputasi (Baiti & Nugroho, 2013). Kecerdasan buatan diciptakan agar komputer dapat melakukan pekerjaan seperti yang dilakukan manusia (Nasution, 2012).

2. Kajian Pustaka

2.1. Karya Tulis

Perancangan adalah suatu kegiatan yang memiliki tujuan untuk mendesign sistem baru yang dapat menyelesaikan kebutuhan maupun masalah-masalah yang dihadapi perusahaan atau organisasi dalam mencapai tujuan yang diperoleh dari pemilihan alternatif sistem yang terbaik [3].

2.2. Plagiarisme

Sistem informasi adalah gabungan yang terorganisasi dari manusia, perangkat lunak, perangkat keras, jaringan komunikasi dan sumber data dalam mengumpulkan, mengubah, dan menyebarkan informasi dalam organisasi. Sistem informasi merupakan fokus utama dari studi untuk disiplin sistem informasi dan organisasi informatika. Dalam arti yang sangat luas, istilah sistem informasi yang sering digunakan merujuk kepada interaksi antara orang, proses algoritma, data, dan teknologi [4].

2.3. *Natural Language Processing* (NLP)

Jasa adalah aktivitas atau tindakan yang tidak berwujud fisik (Intangible) yang diberikan kepada pelanggan yang melibatkan anatra dua pihak dengan melakukan interaksi, yang dapat diidentifikasi, direncanakan, dan memenuhi permintaan pada pelanggan dan kepuasan pelanggan pada hasil jasa yang diberikan.



3. METODE

Penulisan karya ilmiah ini berbasis metode kualitatif yaitu berupa deskripsi mendalam terhadap fenomena kandungan dalam telaah pustaka. Kartanegara (2005) menjelaskan bahwa penelitian kualitatif adalah pengumpulan data pada suatu latar alamiah dengan maksud menafsirkan fenomena yang terjadi di mana peneliti adalah sebagai instrumen kunci, pengambilan sampel sumber data secara *purposive* dan *snowball*. Penelitian ini bertujuan untuk pengembangan sistem deteksi plagiarisme tugas akhir mahasiswa berbasis *Natural Language Processing*. Metode yang digunakan dalam penelitian ini seperti metode pengumpulan data, metode pengumpulan data terdiri dari studi literatur dan observasi. Pada studi literatur kegiatan yang dilakukan adalah mempelajari teori-teori pada buku dan jurnal ilmiah terkait pengembangan sistem deteksi plagiarisme dengan teknologi *Natural Language Processing* (NLP).

4. HASIL PENELITIAN

4.1 Jenis-jenis plagiarisme dan pendeteksiannya

Penggunaan internet merupakan berkah sekaligus kutukan bagi masyarakat umum. Kejahatan dalam berbagai dimensi terungkap melalui internet. Salah satu jenisnya adalah plagiarisme. Ini melibatkan penyalinan atau pencurian ide orisinal dan menerbitkan ide tersebut dengan mengaku sebagai miliknya.

Kerja keras dan kreativitas orang lain tidak dapat ditiru dengan lancar. Merupakan tindakan tidak bermoral bagi orang yang menuntut untuk menampilkan dirinya sebagai penulis kreatif. Plagiarisme terbagi lagi menjadi dua jenis, yaitu intrinsik dan ekstrinsik.

4.2 Perbedaan Plagiarisme Intrinsik dan Ekstrinsik

Saat ini, plagiarisme secara luas diklasifikasikan menjadi plagiarisme ekstrinsik dan intrinsik. Jika berbicara tentang topik pemeriksaan plagiarisme, maka tugas utama yang dilakukan alat pemeriksa plagiarisme adalah menunjukkan plagiarisme ekstrinsik.

Dengan kata lain, ini hanyalah pemeriksaan sepintas pada konten tertentu dimana detail rumit seperti tata bahasa, *part-of-speech*, dan hal-hal lain sering kali terabaikan. Masalahnya disampaikan melalui internet, beserta kekurangannya. Namun, dengan hadirnya teknologi canggih seperti pemrosesan bahasa alami, skenario seperti itu dapat ditangani dengan baik. Plagiarisme lahiriah atau ekstrinsik relatif mudah dideteksi, sedangkan plagiarisme intrinsik cukup sulit dideteksi. Penggunaan kecerdasan mesin sangat penting di sini. Ini dapat mendeteksi jenis plagiarisme intrinsik apa yang digunakan dalam karya yang dipilih.

4.3 Pemanfaatan Aplikasi Natural Language Processing dalam Pemahaman Tulisan

Natural Language Processing seperti yang dijelaskan Jackson (2002) merupakan istilah yang digunakan untuk mendeskripsikan fungsi sebuah komponen perangkat lunak (*software*) dan atau keras (*hardware*) dalam sebuah sistem komputer yang mampu menganalisis atau mensintesis bahasa, baik dalam bentuk suara (*audio/speech*) dan atau tulisan (*written*). Julukan 'natural' membedakan antara bahasa yang dipahami manusia umum dengan komputer, seperti notasi matematika atau logis, atau bahasa komputer, seperti Java, LISP, dan sebagainya. Secara teori, NLP adalah cabang dalam ilmu kecerdasan buatan yang berfokus pada mengkuantifikasi bahasa sehari-hari manusia agar mampu diterima dan dipahami mesin. Kombinasi yang digunakan adalah kekuatan dari



ilmu linguistik dan algoritma komputer untuk mempelajari aturan dan struktur, leksikal, dan morfologikal bahasa dengan melalui sebuah sistem yang cerdas agar mampu memahami, menganalisis, dan mengekstrak makna dari masukan berupa tulisan (*text*) atau suara (*speech*). Klasifikasi fungsi NLP secara garis besar terbagi menjadi empat subjek dasar:

1. Klasifikasi Teks

Salah satu yang paling sering dijumpai, antara lain pencarian web, klasifikasi dan ranking dokumen atau folder, serta pencarian informasi. Teknis kerjanya, yaitu ketika pasca teks dan dokumen input diterima oleh sistem NLP, selanjutnya sistem tersebut akan membuat struktur algoritma agar mampu diperoleh untuk kebutuhan masa yang akan datang.

2. Pencarian Informasi

Dikenal juga dengan istilah “information retrieval”, pada dasarnya memberikan akses kepada pengguna dengan mencari dan mengklasifikasikan informasi dan data yang diminta. Hampir serupa dengan fungsi klasifikasi teks, melalui hal ini peran NLP selain memberikan efisiensi dalam proses pencarian informasi, juga mampu memperbaiki dan meminimalisir error manusia serta meningkatkan keakuratan data yang masuk.

3. Natural Language Generator

NLG berfungsi sebagai program perangkat lunak generator yang tersemat pada komputer yang mampu menghasilkan output berupa teks dan atau kalimat yang bisa dipahami manusia. Contohnya digunakan dalam penghasil laporan tugas yang baru dikerjakan komputer. Klasifikasi ini yang mendasari terbentuknya aplikasi Cek-Aja, program berbasis NLP bahasa Indonesia yang berguna memperbaiki dan menghasilkan revisi karya tulis berbentuk dokumen atau teks.

4. Natural Language Understanding

Salah satu implementasi dari cabang ilmu NLP adalah NLU atau pemahaman bahasa alamiah. Sederhananya, NLU bertugas dengan memberikan peluang terhadap teknologi sehari-hari, seperti komputer rumah atau telepon pintar mampu memahami, menganalisis, dan menyimpulkan, serta mengeksekusi tugas melalui perintah yang disampaikan dengan bahasa sehari-hari.

4.4 Mekanisme Kerja Natural Language Processing dalam Memahami dan Memberi Saran Revisi Tulisan

NLP mulai diaplikasikan sejak empat dekade lalu telah mengalami sejumlah perubahan mekanisme dan teknis dengan menyesuaikan pada perkembangan dan kapabilitas sistem *platform* serta kecerdasan buatan yang selalu mengalami peningkatan. Namun, prinsip teknis dalam proses analisis dan pemahaman NLP selalu tersusun dalam empat rangkaian penting, dimulai dengan *input* atau masukan dalam bentuk teks, kalimat, dan atau dokumen sejenisnya yang di-*import* ke dalam program yang kemudian mengalami proses sintesis dan analisis berakhir pada tersedianya *output* atau hasil akhir yang siap digunakan dalam mengeksekusi perintah, memahami tujuan dan maksud, serta perbaikan dalam sebuah kalimat atau teks tersebut.

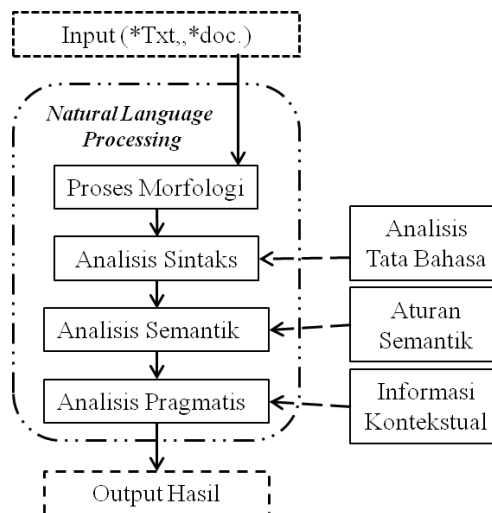
Berdasarkan hasil dari kajian literatur yang dilansir dari berbagai sumber, dapat



diketahui bahwa dalam proses pemahaman dan analisis sintesis terhadap masukan (*input*) hingga menghasilkan keluaran (*output*) ini terbagi dalam empat tahapan sebagai berikut:

1. Analisis Leksikal/Morfologi

Tahapan ini berfungsi menganalisis kalimat dan teks dibagi menjadi bagian-bagian kecil disebut "Token". Token ini akan mampu mengidentifikasi tipe, struktur, dan informasi substantif yang akan digunakan pada langkah berikutnya.



Gambar 1. Model mekanisme program

2. Analisis Sintaks

Pada tahap ini, token diuraikan dengan struktur kalimat yang sesuai dan benar berdasarkan susunan algoritma komputer (*Syntax*) dengan tujuan sebagai pengesahan dan memeriksa validitas serta interpretasi makna kalimat atau teks untuk dianalisis pada tahapan berikutnya. Bagian ini mencakup dua fase, mengoreksi bahwa teks atau kalimat sudah tersusun atau belum serta membaginya menjadi sejumlah struktur yang memberikan susunan sintaks antar susunan kata.

3. Analisis Semantik

Tahap ini bertujuan menganalisis dan mensintesis makna dalam sebuah teks atau kalimat. Untuk memaksimalkan fungsinya, pada tahap ini sistem akan berkolaborasi dengan *cloud server* terhadap aturan semantik yang sesuai dengan struktur dan fungsi *input*. Setiap teks dan kalimat akan ditelaah secara terkomputerisasi, menganalisis konteks dalam teks sekitarnya, dan menganalisis struktur untuk menginterpretasikan makna kata yang tepat yang memiliki lebih dari satu definisi, sehingga menghasilkan substantif yang sesuai dan kaidah bahasa.

4. Analisis Pragmatis

Tahapan ini adalah tahap final dalam analisis data *input*. Fase ini bertugas adalah proses mengekstraksi informasi dari teks. Secara spesifik, bagian ini berfokus pada analisis struktur teks dan menentukan makna yang bisa dipahami dengan menghilangkan hal-hal yang memicu ambiguitas dan ketidakpadanan dalam sebuah teks dan kalimat. Selanjutnya, disaring dengan konten informasi yang menyesuaikan



pada konteks tertentu yang menghasilkan sebuah *output* hasil dari proses NLP.

5. KESIMPULAN

Salah satu alternatif untuk menanggulangi tindakan plagiarisme adalah penerapan sistem komputasi dengan metode *Natural Language Processing*. Metode *Natural Language Processing* menjanjikan tingkat akurasi yang sangat tinggi dalam mendeteksi plagiarisme dengan membandingkan kesamaan antara dokumen berdasarkan vektor representasi teksnya. NLP selalu tersusun dalam empat rangkaian penting, dimulai dengan *input* atau masukan dalam bentuk teks, kalimat, dan atau dokumen sejenisnya yang di-*import* ke dalam program yang kemudian mengalami proses sintesis dan analisis berakhir pada tersedianya *output* atau hasil akhir yang siap digunakan dalam mengeksekusi perintah, memahami tujuan dan maksud, serta perbaikan dalam sebuah kalimat atau teks.

6. DAFTAR REFRESNSI

- [1] R. Akerar, "Artificial Intelligence for Business," SpringerBrief in Business, number 978-3-319-97436-1, 2019.
- [2] A. Sanjaya, A. B. Setiawan, U. Mahdiah, I. N. Farida and A. R. Prasetyo, "Pengukuran Kemiripan Makna Menggunakan Cosine Similarity dan Basis Data Sinonim Kata," *Jurnal Teknologi Informasi dan Ilmu Komputer(JTIK)*, vol. 10, no. 4, pp. 747-752, 2023.
- [3] F. A. Nugroho, F. Septian, D. Pungkastyo and J. Riyanto, "Penerapan Algoritma Cosine Similarity untuk Deteksi Kesamaan Konten pada Sistem Informasi Penelitian dan Pengabdian Kepada Masyarakat," *Jurnal Informatika Universitas Pamulang*, vol. 10, no. 4, pp. 529-536, 2020.
- [4] Herlambang, J. Suwita and B. Tiara, "Analisa dan Perancangan Sistem Pendeteksi Plagiarisme Skripsi pada STMIK Insan Pembangunan Menggunakan Metode Cosine Similarity," *Jurnal IPSIKOM*, vol. 9, no. 1, pp. 10-22, 2022.
- [5] J. Joni and J. Halim, "Implementasi Metode Cosine Similarity dan Tf-Idf dalam Klasifikasi Pengaduan Masyarakat," *Jurnal Ilmiah Core IT: Community Research Information Technology*, vol. 10, no. 4, pp. 51-58, 2021.
- [6] S. Dwisanati and N. Fatonah "Penerapan Metode Cosine Similarity dalam Mendeteksi Plagiarisme pada Jurnal," *Jurnal Format*, vol. , no. 2, pp. 142-150, 2023.
- [7] M. Azmi, "Analisis Tingkat Plagiasi Dokumen Skripsi Dengan Metode Cosine Similarity dan pembobotan tf-idf," *TEKNIMEDIA*, vol. 2, no. 2, pp. 90-95, 2021.
- [8] S. H. Saniati, "Implementasi Algoritma Cosine Similarity untuk Mendeteksi Kemiripan Topik Judul," *In Jecsit*, vol. 1, no. 1, pp. 51-56, 2021.
- [9] R. Wakabayashi, "The Effects of the Peer Feedback Process on Reviewers' Own Writing," *Journal Of English Language Teaching*, vol. 10, no. 10, pp. 179-192. 2013, doi : 10.5539/elt.v6n9p177