DETEKSI KEMIRIPAN ANTAR DOKUMEN DENGAN METODE *CASE BASED REASONING* MENGGUNAKAN *COSINE SIMILARITY MEASURE* (STUDI KASUS SIMNG LPPM UNIVERSITAS SRIWIJAYA)

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

Oleh:

Nabila Febriyanti
NIM: 09021281823071

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2021

LEMBAR PENGESAHAN SKRIPSI

DETEKSI KEMIRIPAN ANTAR DOKUMEN DENGAN METODE *CASE BASED REASONING* MENGGUNAKAN *COSINE SIMILARITY MEASURE* (STUDI KASUS SIMNG LPPM UNIVERSITAS SRIWIJAYA)

Oleh:

Nabila Febriyanti

NIM: 09021281823071

Palembang, 30 Desember 2021

Pembimbing I

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

Pembimbing II

Osvari Arsalan, M.T.
NIP. 198806282018031001

Mengetahui,

Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom.

NIP. 197812222006042003

TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Kamis tanggal 30 Desember 2021 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama	:	Nabila Febriyanti
NIM	:	09021281823071
Judul	:	Deteksi Kemiripan Antar Dokumen dengan Metode *Case Based Reasoning* Menggunakan *Cosine Similarity Measure* (Studi Kasus SIMNG LPPM Universitas Sriwijaya)

1. Ketua Penguji

   Alvi Syahrini Utami, M.Kom. .....
   NIP. 197812222006042003

2. Pembimbing I

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

3. Pembimbing II

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

4. Penguji I

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

5. Penguji II

   M. Qurhanul Rizqie, M.T., Ph.D. .....
   NIDN. 0203128701

Mengetahui,
Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama	:	Nabila Febriyanti
NIM	:	09021281823071
Jurusan	:	Teknik Informatika
Judul Skripsi	:	Deteksi Kemiripan Antar Dokumen dengan Metode *Case Based Reasoning* Menggunakan *Cosine Similarity Measure* (Studi Kasus SIMNG LPPM Universitas Sriwijaya)

Hasil pengecekan *Software iThenticate/Turnitin* : 7%

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

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

Indralaya, 30 Desember 2021

Nabila Febriyanti
NIM. 09021281823071

MOTTO DAN PERSEMBAHAN

Motto:

- Apresiasi, bersangka baik dan bersyukur.
- Kerja keras tidak akan mengkhianati hasil.
- *Without any regret, do your best.*
- *Indeed, along with every hardship is relief.*
- *Always have faith in Allah. Indeed, Allah is with those that are patient.*

Kupersembahkan karya tulis ini kepada:

- Allah Subhanahu Wa Ta'ala dan Baginda Rasulullah SAW
- Mama dan Alm. Papa tercinta
- Keluarga Besar Yusman H. Aris dan Umar Ahmad
- Dosen pembimbing Akademik dan Skripsi saya
- LPPM Universitas Sriwijaya
- Sahabat dan teman-teman seperjuangan

ABSTRACT

LPPM Universitas Sriwijaya is an institution that coordinates academic research and community service inside Universitas Sriwijaya. In carrying out the duty, LPPM assesses every proposal's originality which would be impossible to do manually in the future due to massive data growth. Thus, automatization for the proposal's originality check is needed. The Case Based Reasoning method is used in this research because it allows the system to reuse the information that has been obtained to find documents that are similar to the test document. In this study, the data is represented in the form of the Vector Space Model and uses Cosine Similarity to measure document to document similarity. The data is represented by giving weight for each part of the tested documents. In this study, four formulas from previous research will be used for term weighting then the final result will be compared. The process begins by extracting data, separating parts of the document, figuring the similarity value of the test document to the case base utilizing Cosine Similarity Measure, results filtering with a certain threshold, summarizing the calculation results, and finally preserving the results obtained to be reused in the next calculation. The results of this study indicate that the text-similarity detection between documents has been successfully carried out using the proposed method with the best sensitivity level and the fastest computation time achieved in configuration II.

Keywords: Case Based Reasoning, Text Similarity Detection, Cosine Similarity Measure, Vector Space Model

ABSTRAK

LPPM Universitas Sriwijaya adalah lembaga yang bertugas untuk mengoordinasikan penelitian dan pengabdian civitas akademika di lingkungan Universitas Sriwijaya. Dalam pelaksanaan penilaiannya, orisinalitas ajuan civitas akademika harus dikaji dan seiring dengan perkembangan waktu jumlah dokumen yang harus diproses menjadi semakin besar sehingga diperlukan sistem yang dapat mengukur secara otomatis orisinalitas dokumen ajuan dengan dokumen-dokumen ajuan terdahulu. Metode *Case Based Reasoning* digunakan dalam penelitian ini karena memungkinkan sistem untuk menggunakan kembali informasi yang pernah diperoleh untuk menemukan dokumen-dokumen termirip terhadap satu dokumen utama. Dalam penelitian ini data direpresentasikan ke dalam bentuk *Vector Space Model* dan kemiripannya diukur menggunakan *Cosine Similarity Measure*. Representasi data dilakukan dengan memberikan bobot pada setiap bagian dokumen yang akan diukur kemiripannya. Dalam penelitian ini, akan digunakan empat formula pembobotan yang kemudian akan dibandingkan hasil akhirnya. Proses diawali dengan melakukan ekstraksi data, pemisahan bagian dokumen, menghitung nilai kemiripan dokumen uji terhadap basis kasus dengan *Cosine Similarity Measure*, melakukan penyaringan dengan ambang batas tertentu, merangkum hasil perhitungan dan terakhir menyimpan hasil yang didapatkan untuk dapat digunakan kembali pada perhitungan selanjutnya. Hasil penelitian ini menunjukkan bahwa deteksi kemiripan teks antar dokumen berhasil dilakukan menggunakan metode yang diajukan dengan tingkat kepekaan terbaik dan waktu komputasi tercepat dicapai pada konfigurasi II.

Kata Kunci: *Case Based Reasoning,* Deteksi Kemiripan Teks, *Cosine Similarity Measure, Vector Space Model*

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT, Tuhan Semesta Alam atas berkat, rahmat, rahim dan karunia-Nya yang telah diberikan kepada penulis sehingga Tugas Akhir berjudul "Deteksi Kemiripan Antar Dokumen dengan Metode *Case Based Reasoning* (Studi Kasus SIMNG LPPM Universitas Sriwijaya)" dapat disusun dengan baik sebagai syarat dalam menyelesaikan studi Strata-1 program studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Sriwijaya.

Pada kesempatan ini penulis ingin mengucapkan ucapan terima kasih yang luar biasa besar kepada semua pihak yang memberikan dukungan, motivasi dan bimbingan selama penyusunan Tugas Akhir dan penelitian ini berlangsung. Secara khusus ucapan terima kasih ini ditujukan kepada:

1. Allah Subhanallahu Wa Ta'ala atas segala berkah, rahmat, rahim dan karunia-Nya.
2. Keluarga tercinta, Mama yang selalu menjadi pemberi dukungan terbesar yang tidak pernah menuntut kesempurnaan dariku anaknya namun selalu menuntun agar anaknya tetap dapat menjalani semua ujian. Alm. Papa, yang Insyaa Allah menyaksikan perjuanganku dari tempat terbaik bersama Allah SWT, kakak dan adikku juga keluarga besarku yang selalu memberi motivasi untuk bergerak maju dan memperbaiki diri.
3. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.

4. Ibu Alvi Syahrini Utami, M.Kom. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Dian Palupi Rini, M.Kom, Ph.D. selaku Dosen Pembimbing I yang senantiasa membagikan ilmu, membimbing, memberi arahan dan motivasi saya untuk menyelesaikan Tugas Akhir ini dengan baik.
6. Bapak Osvari Arsalan, M.T. selaku Dosen Pembimbingan II yang memberikan arahan dan bimbingan serta banyak ilmu untuk menyelesaikan Tugas Akhir ini.
7. Ibu Novi Yusliani, M.T. selaku penguji I Tugas Akhir yang telah memberikan nasihat dan saran yang membangun.
8. Bapak M. Qurhanul Rizqie, M.T., Ph.D. selaku penguji II Tugas Akhir yang telah memberikan nasihat dan saran yang membangun.
9. Ibu Rizki Kurniati, M.T. selaku Dosen Pembimbing Akademik yang senantiasa memberikan saran, arahan dan bimbingan kepada saya selama masa perkuliahan.
10. Seluruh Dosen Jurusan Teknik Informatika, juga Dosen Fakultas Ilmu Komputer yang telah banyak memberikan bekal, arahan, dan saran selama masa perkuliahan.
11. Bapak Samsuryadi, M.Kom., Ph.D. selaku Ketua LPPM Universitas Sriwijaya yang telah memberikan izin kepada penulis untuk melakukan penelitian pada Lembaga yang dipimpin.

12. Ibu Dwi Oktaria Sari, S.Sos. selaku staf Program Penelitian LPPM Universitas Sriwijaya yang telah sangat banyak membantu penulis selama melakukan penelitian di LPPM Universitas Sriwijaya.
13. Seluruh Staf Administrasi dan Pegawai yang telah membantu dalam urusan administrasi.
14. Sahabat – sahabatku, Shabrina, Tarisyah, Aulia, Nurul, Rahmavita, Aryo, Poedjo, Shandy, Tiansyah, Nadia, Ani, Ikhsan, Sholeh, Firman serta seluruh teman seperjuangan Teknik Informatika Angkatan 2018 yang selalu sedia menjadi rekan diskusi baik dalam pengerjaan Tugas Akhir maupun selama masa perkuliahan.

Penulis secara penuh menyadari akan kekurangan dalam penyusunan Tugas Akhir ini. Kekurangan ini semata – mata adalah karena keterbatasan pengetahuan juga pengalaman penulis. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan untuk menyempurnakan Tugas akhir ini sehingga dapat membawa manfaat lebih banyak dan luas. Akhir kata, semoga Tugas Akhir ini bermanfaat bagi kita semua.

Indralaya, 30 Desember 2021

Nabila Febriyanti

DAFTAR ISI

Halaman

DAFTAR TABEL

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1. Pendahuluan

Pada bab ini akan dijelaskan mengenai garis besar pokok-pokok pikiran dalam penelitian ini. Pokok-pokok pikiran dalam penelitian ini antara lain latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, manfaat penelitian dan batasan masalah. Pokok-pokok pikiran dalam penelitian ini akan menjadi acuan dalam menentukan metode penelitian.

1.2. Latar Belakang

Menurut Subroto & Selamat (2014) kemajuan teknologi mendukung pesatnya distribusi dokumen karya ilmiah secara publik sebagai referensi sebuah riset. Namun, luasnya distribusi dan kemudahan akses ini dapat menjadi celah terjadinya tindakan plagiarisme. Pada akademisi sanksi administratif yang dapat diberikan atas tindakan plagiarisme adalah teguran hingga yang paling berat pembatalan ijazah dan pemberhentian secara tidak hormat dari jabatan yang diduduki berdasarkan PERMENDIKNAS No. 17 tahun 2010 tentang Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi (Yuliati, 2012). Seiring pertumbuhan data dokumen yang sangat pesat, akan menjadi mustahil dilakukan inspeksi orisinalitas secara manual (Clough, 2000).

Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Sriwijaya adalah unsur pelaksana akademik di lingkungan Universitas Sriwijaya yang memiliki tugas melakukan perencanaan, koordinasi, pengawasan,

dan penilaian pelaksanaan dan hasil/luaran kegiatan penelitian dan pengabdian kepada masyarakat [1]. Untuk mendukung pelaksanaan tugas tersebut, LPPM Universitas Sriwijaya membangun Sistem Informasi Manajemen New Generation (SIMNG). Sistem ini bertujuan untuk mengkoordinasi dan mempermudah tenaga akademisi dalam memproses serta memperoleh informasi terkait pengajuan hibah penelitian dan pengabdian[2] sehingga kuantitas dan kualitas riset meningkat untuk tercapainya visi Universitas Sriwijaya menjadi perguruan tinggi berbasis riset[3].

Dalam pelaksanaan penilaiannya, orisinalitas ajuan dari tenaga akademik diperiksa satu per satu secara manual oleh tim penilai. Pada paruh kedua tahun 2020, tercatat hingga 500 ajuan penelitian dan pengabdian yang masuk ke LPPM Universitas Sriwijaya melalui SIMNG (Data ajuan masuk LPPM Universitas Sriwijaya, 2020). Perkembangan data yang sangat pesat karena kemudahan yang disediakan sistem membuat penilaian orisinalitas ajuan secara manual menjadi sangat sulit dilakukan. Karenanya, dibutuhkan sistem yang dapat mengukur secara otomatis orisinalitas dokumen-dokumen ajuan dengan ajuan terdahulu.

Penelitian terkait deteksi plagiarisme teks secara literal antar dokumen telah dilakukan dengan berbagai metode, diantaranya berbasis string dengan algoritma Rabin-Karp (Parwita *et al.*, 2019; Leman *et al.*, 2019) dan berbasis vektor dengan Vector Space Model (VSM) bersama ukuran kemiripan Jaccard Coefficient dan *Cosine Similarity* (Jiffriya *et al.*, 2014) serta metode hybrid pada VSM (Saptono *et al.*, 2018).

---

[1] http://lppm.unsri.ac.id/2020/visi-misi-lppm-unsri/
[2] http://lppm.unsri.ac.id/2020/wp-content/uploads/2020/08/user-Guide-simng-v1.pdf
[3] https://unsri.ac.id/main/visi_misi

Deteksi plagiarisme dengan algoritma Rabin-Karp memiliki kelemahan dalam menangani masalah nilai hash yang sama pada kata (Priambodo, 2018) dan membutuhkan waktu yang lama dalam komputasi dibandingkan dengan algoritma Levenshtein Distance (Purba & Situmorang, 2017). *Cosine Similarity* bekerja lebih baik dibandingkan dengan Jaccard Coefficient dalam pengujian menggunakan VSM trigram karena Jaccard Coefficient kurang mampu bekerja dengan baik dalam pemberian bobot lebih terhadap istilah unik (Jiffriya *et al.*, 2014). Dalam Saptono *et al.*, (2018) data direpresentasikan kedalam Vector Space Model kemudian mengombinasikan TF-IDF untuk pembobotan kata, *Cosine Similarity* dan probabilitas kemunculan kata untuk pengukuran kemiripan dalam kasus deteksi plagiarisme pada teks.

Metode Case Based Reasoning telah digunakan dalam kasus pencarian dokumen termirip dalam Mihajlovic & Xiong (2019). Dalam penelitian ini *Case Based Reasoning* digunakan karena memungkinkan sistem untuk menggunakan kembali informasi yang pernah diperoleh untuk menemukan dokumen-dokumen termirip terhadap satu dokumen utama. Dalam penelitian ini data direpresentasikan ke dalam *Vector Space Model*.

Representasi data teks menjadi bentuk vektor dilakukan dengan memberikan bobot pada setiap token kata yang akan menjadi elemen dari setiap vektor (Saptono *et al.*, 2018). Beberapa penelitian menggunakan pembobotan TF-IDF yang dapat mengukur derajat kepentingan dari setiap kata berdasarkan konteks dokumen secara keseluruhan (Mishra & Vishwakarma, 2016), diantaranya adalah

penelitian oleh Manning *et al.* (2008), Jiffriya *et al.* (2014), Xu *et al.* (2016) dan Saptono *et al.* (2018).

Berdasarkan penelitian-penelitian tersebut, penelitian ini akan menerapkan metode *Case Based Reasoning* dengan *Cosine Similarity Measure* untuk mendeteksi kemiripan teks antar dokumen ajuan hibah penelitian di lingkungan SIMNG LPPM Universitas Sriwijaya dan membandingkan pembobotan yang digunakan dalam beberapa penelitian sebelumnya dalam kasus ini.

1.3. Rumusan Masalah

Pada penelitian ini penulis akan meneliti penggunaan *Case Based Reasoning* dengan *Cosine Similarity Measure* pada pendeteksi kemiripan teks antar dokumen ajuan hibah penelitian SIMNG LPPM Universitas Sriwijaya. Rumusan masalah pada penelitian ini dibagi menjadi beberapa pertanyaan penelitian, yaitu:

1. Bagaimana melakukan deteksi kemiripan teks antar dokumen ajuan hibah penelitian SIMNG LPPM Universitas Sriwijaya menggunakan metode *Case Based Reasoning* dengan *Cosine Similarity Measure* dan melakukan perbandingan pembobotan yang digunakan dalam kasus tersebut*?*
2. Bagaimana kinerja metode *Case Based Reasoning* dengan *Cosine Similarity Measure* dan perbandingan pembobotannya pada pendeteksi kemiripan teks antar dokumen ajuan hibah penelitian SIMNG LPPM Universitas Sriwijaya?

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Membuat perangkat lunak deteksi kemiripan teks antar dokumen ajuan hibah penelitian SIMNG LPPM Universitas Sriwijaya menggunakan metode *Case Based Reasoning* dengan *Cosine Similarity Measure* dan konfigurasi pembobotan.
2. Mengetahui tingkat kinerja metode *Case Based Reasoning* dengan *Cosine Similarity Measure* dan konfigurasi pembobotannya pada masalah deteksi kemiripan teks antar dokumen.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Mengembangkan perangkat lunak deteksi kemiripan teks antar dokumen menggunakan metode *Case Based Reasoning* dengan *Cosine Similarity Measure* yang dapat digunakan untuk memproses dokumen ajuan hibah penelitian SIMNG LPPM Universitas Sriwijaya.
2. Memahami dan mempelajari kinerja metode *Case Based Reasoning* dengan *Cosine Similarity Measure* dan pembobotannya pada masalah deteksi kemiripan teks antar dokumen yang dapat menjadi rujukan penelitian terkait di masa depan.

1.6. Batasan Penelitian

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang akan diproses adalah teks berbahasa Indonesia.

2. Data yang digunakan adalah proposal pengajuan hibah yang diperoleh dari SIMNG LPPM Universitas Sriwijaya tahun 2020 sebagai basis kasus dan tahun 2021 sebagai data untuk pengujian.
3. Penelitian ini menggunakan Pustaka perangkat lunak PyMuPDF untuk ekstraksi teks dari dokumen dan PySastrawi pada proses pra-pengolahan teks (*stopword removal* dan *stemming*).
4. Ekstensi dokumen yang didukung oleh perangkat lunak adalah *.pdf*.
5. Perangkat lunak hanya menampilkan nilai kemiripan antara dokumen masukan dengan dokumen dalam repositori tanpa memberi label pasti status duplikasi.

1.7. Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini diuraikan latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian dan batasan masalah. Pokok-pokok pembahasan ini akan menjadi dasar pengembangan kajian pada bab selanjutnya.

BAB II. KAJIAN LITERATUR

Pada bab ini dibahas landasan teori yang digunakan di dalam penelitian, termasuk di dalamnya penelitian terkait, plagiarisme, Pra-pengolahan teks, *Case Based Reasoning*, *Cosine Similarity Measure* dan metode pengembangan perangkat lunak yang digunakan dalam penelitian ini.

BAB III. METODOLOGI PENELITIAN

Pada bab ini dibahas proses pengumpulan data dan tahapan-tahapan di dalam penelitian, termasuk pra-pengolahan data teks. Tahapan penelitian dibahas lebih rinci berdasarkan kerangka kerja tertentu. Di bagian akhir bab ini akan dimuat rancangan manajemen proyek penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini dibahas mengenai perancangan dan lingkungan implementasi deteksi kemiripan teks antar dokumen dengan pencarian dokumen termirip, implementasi program dengan metode CBR, hasil eksekusi dan hasil pengujian.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini hasil dari implementasi dan pengujian metode yang telah dirancang disajikan. Analisis diberikan sebagai dasar dari kesimpulan yang akan diambil di dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini kesimpulan dari semua uraian-uraian dalam penelitian ini disajikan. Selain itu, disajikan pula saran-saran yang diharapkan berguna untuk pengembangan pendeteksi kemiripan teks antar dokumen ini.

1.8. Kesimpulan

Berdasarkan uraian di atas, akan dikembangkan perangkat lunak deteksi kemiripan teks antar dokumen menggunakan metode *Case Based Reasoning* dengan *Cosine Similarity Measure.* Metode ini diharapkan dapat diimplementasikan dengan baik ke dalam perangkat lunak.

DAFTAR PUSTAKA

Adriani, M., Asian, J., Nazief, B., Tahaghoghi, S.M.M. & Williams, H. 2007. Stemming Indonesian: A confix-stripping approach. ACM Trans. Asian Lang. Inf. Process., 6.

Agnar, A. & Plaza, E. 1994. Case-Based reasoning: Foundational issues, methodological variations, and system approaches. AI Communications, 7(1): 39–59.

Albakush, I.H. 2017. Plagiarism Detection Between Theory And Practical Calculations. IOSR Journal of Electronics and Communication Engineering, 12(02): 60–67.

Alfikri, Z.F. & Purwarianti, A. 2014. Detailed Analysis of Extrinsic Plagiarism Detection System Using Machine Learning Approach (Naive Bayes and SVM). TELKOMNIKA Indonesian Journal of Electrical Engineering, 12(11): 7884–7894.

Alzahrani, S.M., Salim, N. & Abraham, A. 2012. Understanding plagiarism linguistic patterns, textual features, and detection methods. IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, .

Budiman, A.E. 2020. Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir. Jurnal Teknik Informatika dan Sistem Informasi, 6: 475–488.

Clough, P. 2000. Plagiarism in natural and programming languages: an overview of current tools and technologies. Finance, (July): 1–31. (http://www.dcs.shef.ac.uk/nlp/meter/Documents/reports/plagiarism/Plagiarism.pdf).

Furlan, B. & Batanovi, V. 2013. Semantic similarity of short texts in languages with a de fi cient natural language processing support. 55: 710–719.

Goyena, R. & Fallis, A.. 2019. Pengembangan Aplikasi Pendeteksi Plagiarisme Pada Dokumen Teks Menggunakan Algoritma Rabin-Karp. Journal of Chemical Information and Modeling, 53(9): 1689–1699.

Jiffriya, M.A.C., Jahan, M.A.C.A. & Ragel, R.G. 2014. Plagiarism detection on electronic text based assignments using vector space model. 2014 7th International Conference on Information and Automation for Sustainability: "Sharpening the Future with Sustainable Technology", ICIAfS 2014.

Kruchten, P. 2003. The Rational Unified Process An Introduction, Third Edition. 3 ed. Addison-Wesley Professional.

Leman, D., Rahman, M., Ikorasaki, F., Riza, B.S. & Akbbar, M.B. 2019. Rabin Karp and Winnowing Algorithm for Statistics of Text Document Plagiarism Detection. *2019 7th International Conference on Cyber and IT Service Management, CITSM 2019*.

Leonardo, B. & Hansun, S. 2017. Text Documents Plagiarism Detection using Rabin-Karp and Jaro-Winkler Distance Algorithms. Indonesian Journal of Electrical Engineering and Computer Science, 5(2): 462–471.

Liu, B. 2011. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications). 2 ed. Springer.

Manning, C.D., Raghavan, P. & Schutze, H. 2008. Introduction to Information

Retrieval. Cambridge: Cambridge University Press. (http://ebooks.cambridge.org/ref/id/CBO9780511809071, Diakses 20 Maret 2021).

Mihajlovic, M. & Xiong, N. 2019. Finding the most similar textual documents using Case-Based Reasoning. arXiv.

Mishra, A. & Vishwakarma, S. 2016. Analysis of TF-IDF Model and its Variant for Document Retrieval. Proceedings - 2015 International Conference on Computational Intelligence and Communication Networks, CICN 2015, 772–776.

Mubarak, A., Muis, A., Studi, P., Informatika, T., Teknik, F., Ternate, U.K., Studi, P., Informatika, T., Komputer, F.I., Indonesia, U. & Makassar, T. 2020. Case-Based Reasoning Untuk Aplikasi Pemilihan Pestisida Hama Case-Based Reasoning for Web Based Selection of Rice Pesticides. 3(2): 119–124.

Parwita, W.G.S., Indradewi, I.G.A.A.D. & Wijaya, I.N.S.W. 2019. String Matching based Plagiarism Detection for. 2019 5th International Conference on New Media Studies.

Potthast, M., Stein, B., Eiselt, A., Barrón-Cedeno, A. & Rosso, P. 2009. Overview of the 1st international competition on plagiarism detection. CEUR Workshop Proceedings, 502: 1–9.

Priambodo, J. 2018. Pendeteksian Plagiarisme Menggunakan Algoritma Rabin-Karp dengan Metode Rolling Hash. Jurnal Informatika Universitas Pamulang, 3(1): 39.

Purba, A.H. & Situmorang, Z. 2017. Analisis Perbandingan Algoritma Rabin-Karp Dan Levenshtein Distance Dalam Menghitung Kemiripan Teks. Jurnal Teknik Informatika Unika St. Thomas (JTIUST), 02: 24–32.

Ratna, A.A.P., Purnamasari, P.D., Adhi, B.A., Ekadiyanto, F.A., Salman, M., Mardiyah, M. & Winata, D.J. 2017. Cross-language plagiarism detection system using latent semantic analysis and learning vector quantization. Algorithms, 10(2).

Richter, M.M. & Weber, R.O. 2013. Case-Based Reasoning. Case-Based Reasoning. Springer International Publishing.

Saptono, R., Prasetyo, H. & Irawan, A. 2018. Combination of cosine similarity method and conditional probability for plagiarism detection in the thesis documents vector space model. Journal of Telecommunication, Electronic and Computer Engineering, 10(2–4): 139–143.

Sommerville, I. 2011. Software Engineering. 9 ed. Pearson Education.

Subroto, I.M.I. & Selamat, A. 2014. Plagiarism detection through internet using hybrid artificial neural network and support vectors machine. Telkomnika (Telecommunication Computing Electronics and Control), 12(1): 209–218.

Tala, F.Z. 2003. A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. M.Sc. Thesis, Appendix D, pp: 39–46.

Tawfik, A.A., Alhoori, H., Keene, C.W., Bailey, C. & Hogan, M. 2018. Using a Recommendation System to Support Problem Solving and Case-Based Reasoning Retrieval. Technology, Knowledge and Learning, 23(1): 177–187.

Uysal, A.K. & Gunal, S. 2014. The impact of preprocessing on text classification. Information Processing and Management, 50(1): 104–112.

(http://dx.doi.org/10.1016/j.ipm.2013.08.006).
Widaningrum, I., Mustikasari, D. & Arifin, R. 2018. A review of detection plagiarism in indonesian language. 1(2): 65–75.
Xu, L., Sun, S. & Wang, Q. 2016. Text similarity algorithm based on semantic vector space model. hal.1–4.
Yuliati, Y. 2012. Perlindungan Hukum Bagi Pencipta Berkaitan Dengan Plagiarisme Karya Ilmiah Di Indonesia. Arena Hukum, 5(1): 54–64.
Zechner, M., Muhr, M., Kern, R., Granitzer, M. & Graz, K.-C. 2009. External and Intrinsic Plagiarism Detection Using Vector Space Models.
Zou, Y., Kiviniemi, A. & Jones, S.W. 2017. Retrieving similar cases for construction project risk management using Natural Language Processing techniques. Automation in Construction, 80(September 2016): 66–76. (http://dx.doi.org/10.1016/j.autcon.2017.04.003).