

IMPLEMENTASI PENGECEKAN PLAGIARISME PROPOSAL TUGAS AKHIR MAHASISWA TEKNIK INFORMATIKA UPN VETERAN YOGYAKARTA

Awang Hendrianto Pratomo ^{(1)*}, Andiko Putro Suryotomo ⁽¹⁾

⁽¹⁾Jurusan Teknik Informatika Universitas Pembangunan Nasional Veteran Yogyakarta
Jl. Babarsari 2, Tambakbayan, Yogyakarta 55281 Telp/Fax. (0274) 485323
e-mail: awang@upnyk.ac.id^{(1)*}

Abstract

The thesis proposal submission system currently used by the Department of Informatics Engineering at UPN "Veteran" Yogyakarta is still using conventional manual methods, so difficulties are encountered in checking the document's similarity for the student's final project proposal. The WInnowing algorithm can be used to do similarity checking in text-based documents. The WInnowing algorithm works by retrieving text data from the thesis proposal, then carrying out several steps such as preprocessing by finding the smallest hash value of each document used as a comparison to determine the similarity value between documents. The results showed that the smaller n-gram size would get a larger percentage similarity value (76% and 89%). This happens because a smaller n-gram size will result in smaller truncated strings, increasing the probability of finding the same string of characters. The bigger n-gram size will result in more characters than the smaller n-gram size, which causes fewer character strings to be found thus lowering the similarity value (down to 57% and 68%).

Keywords: thesis proposal, plagiarism, WInnowing algorithm, hash value, n-gram

Abstrak

Sistem informasi skripsi yang saat ini digunakan oleh program studi Teknik Informatika UPN "Veteran" Yogyakarta masih menggunakan cara konvensional, sehingga ditemui kesulitan dalam proses pemeriksaan kemiripan dokumen yang ada dalam proposal tugas akhir mahasiswa. Pemeriksaan kemiripan dokumen dapat dilakukan dengan menggunakan algoritme WInnowing. Algoritme WInnowing bekerja dengan cara mengambil data teks dari proposal, kemudian dilakukan beberapa tahapan seperti *preprocessing* kemudian dilakukan dengan mencari nilai *hash* terkecil dari setiap dokumen yang digunakan sebagai pembandingan untuk menentukan nilai kemiripan dokumen tersebut. Hasil penelitian diketahui bahwa pemilihan n-gram yang semakin kecil akan memperoleh nilai persentase kemiripan yang besar (76% dan 89%). Hal ini terjadi karena pada n-gram yang lebih sedikit, *string* yang dipotong lebih kecil sehingga kemungkinan untuk ditemukannya rangkaian karakter yang sama semakin besar. Semakin besarnya n-gram, maka mengandung karakter yang lebih banyak dibandingkan dengan n-gram yang lebih kecil sehingga menyebabkan rangkaian karakter yang ditemukan semakin berkurang sehingga menurunkan nilai similaritas (hingga 57% dan 68%).

Kata kunci: proposal skripsi, plagiarisme, algoritme WInnowing, nilai *hash*, n-gram

1. PENDAHULUAN

Jurusan Teknik Informatika UPN "Veteran" Yogyakarta memiliki sistem informasi skripsi yang telah dibuat sejak tahun 2007, beberapa fitur seperti pengajuan proposal skripsi, penetapan dosen pembimbing, penentuan waktu pengujian skripsi dan lain-lain sudah terdapat di sistem informasi skripsi tersebut. Namun jika berbicara mengenai skripsi tentu tidak lepas dengan yang namanya plagiat, karena skripsi merupakan karya ilmiah hasil riset asli dari mahasiswa tersebut sebagai syarat untuk memperoleh gelar sarjana.

Skripsi mahasiswa umumnya tidak lepas dari masalah copy paste, sehingga dokumen skripsi mahasiswa selalu ada kemiripan dengan dokumen yang sudah ada sebelumnya. Pengecekan kesamaan dokumen dapat dilakukan menggunakan aplikasi dekstop maupun online dengan menerapkan algoritme-algoritme atau metode-metode tertentu. Beberapa algoritme diantaranya yaitu algoritme Manber, WInnowing dan Rabin-Karp. Algoritme Manber memiliki jumlah langkah yang lebih sedikit dibandingkan dengan algoritme WInnowing sehingga pemrosesan dokumen menjadi lebih cepat (Djafar, 2015). Algoritme WInnowing memiliki tingkat keakuratan

pendeteksian lebih baik karena algoritme tersebut dapat mengetahui letak kesamaan dokumen (Kurniawati dan Wicaksana, 2008).

Algoritme *Winnowing* memiliki teknik pengelompokan berdasar dari nilai hash menggunakan *window*. Waktu pemrosesan algoritme Rabin-Karp memiliki kemampuan yang cukup baik dalam memproses dokumen. Algoritme Rabin-Karp memiliki kelemahan yaitu ketidak mampuan untuk mengelompokkan nilai hash (Djafar, 2015). Agar dapat algoritma pendeteksian kemiripan dokumen dapat dilakukan dengan menambahkan metode fingerprint. Dokumen fingerprint merupakan metode yang digunakan untuk mendeteksi keakuratan kesamaan antar dokumen. Cara kerja metode dokumen fingerprint dengan menggunakan teknik hashing.

Algoritme yang menggunakan metode fingerprint salah satunya adalah algoritme Manber, algoritme *Winnowing* dan algoritme Rabin-Karp. Cara kerja algoritme Manber hampir sama dengan algoritme *Winnowing*, perbedaan kedua algoritme tersebut menurut Kurniawati & Wicaksana, 2008 yaitu jumlah langkah yang dibutuhkan oleh algoritme Manber lebih sedikit yaitu satu tahap sedangkan algoritme *Winnowing* menggunakan 3 tahap, algoritme *Winnowing* lebih informatif karena dapat memberitahukan letak fingerprint daripada algoritme Manber, dan keduanya mengimplementasikan pemilihan *fingerprint* yang berbeda. Pada algoritme Manber, *fingerprint* dipilih dari setiap nilai *hash* yang memenuhi persyaratan $H \bmod P = 0$, di mana H adalah nilai *hash* dan P adalah nilai pembagi yang digunakan, sementara pada algoritme *Winnowing* dipilih nilai hash minimum dalam setiap *window*.

Sistem Informasi Skripsi yang dimiliki oleh Teknik Informatika UPN "Veteran" Yogyakarta memiliki kelebihan dan keterbatasan. Permasalahan mengenai pendeteksian plagiarisme pada proposal skripsi dan efisiensi waktu penanganan pengajuan proposal skripsi menjadi permasalahan yang harus diselesaikan. Pada permasalahan tersebut solusi yang digunakan dalam penelitian ini adalah mengembangkan Sistem Informasi Skripsi tersebut dengan menggunakan algoritme *Winnowing* sebagai algoritme pendeteksi dokumen teks. Untuk menghitung tingkat persentase kesamaan dokumen menggunakan persamaan koefisien *Jaccard* dengan cara menentukan bobot dari suatu dokumen dan mengurutkannya berdasarkan nilai persamaanya (Fadelillah dkk, 2017) dan untuk mencari referensi artikel yang memiliki tingkat kesamaan dengan dokumen uji menggunakan teknik web *scrapping* yang bertujuan untuk mendapatkan data artikel ilmiah secara cepat dari halaman-halaman website (Josi dan Abdillah, 2014). Hal tersebut diharapkan dapat membantu penyelesaian masalah dalam mendeteksi plagiarisme dan memudahkan dalam penanganan pengajuan proposal yang ada di lingkup Teknik Informatika UPN "Veteran" Yogyakarta.

2. METODE PENELITIAN

Prinsip dari metode kesamaan kata kunci yaitu mengekstrak kata kunci pada dokumen uji dan membandingkan dengan kata kunci yang ada pada dokumen pembanding.

Sebuah algoritme harus memiliki salah satu persyaratan berikut untuk melakukan pendeteksian plagiarisme dokumen teks (Schleimer et al., 2003) yaitu:

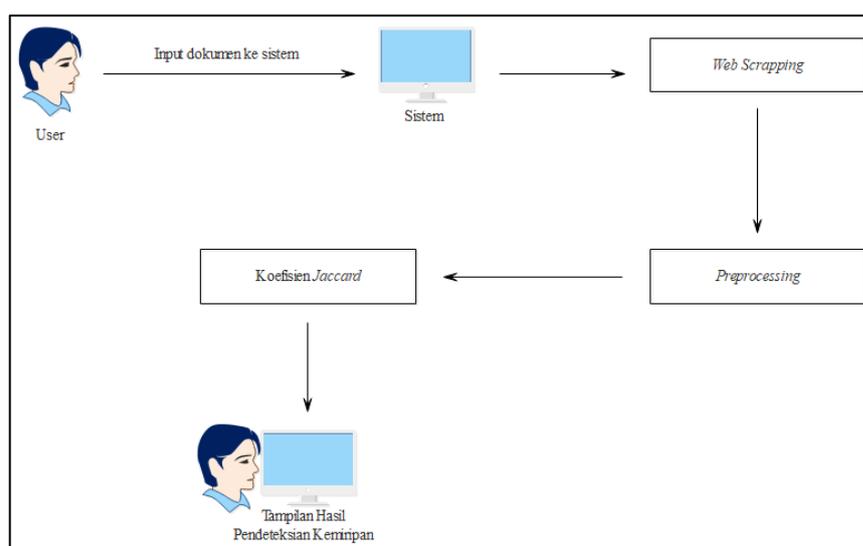
1. *Whitespace Insensitivity*, yaitu dalam melakukan pencocokan terhadap dokumen teks, seharusnya tidak terpengaruh oleh spasi, kapital atau huruf normal, tanda baca, simbol-simbol dan sebagainya.
2. *Noise Supression*, yaitu sebisa mungkin menghindari penemuan kecocokan kata yang kurang relevan atau bukan kata yang umum digunakan dan kata yang terlalu pendek, misalnya: "the". Panjang kata digunakan untuk membuktikan bahwa kata-kata tersebut merupakan plagiarisme atau bukan.
3. *Position Independence*, yaitu untuk mencari kesamaan tidak harus bergantung pada posisi kata-kata, yang berarti walaupun berada pada paragraf yang berbeda harus dapat ditemukan kecocokan atau kesamaannya.

Proses pencarian kemiripan dokumen dilaksanakan dengan menggunakan metode *web scrapping*. *Web scrapping* merupakan sebuah proses yang melibatkan dokumen semi-structured dari internet, umumnya pada halaman web menggunakan markup language seperti

HTML atau XHTML (Pramana & Mungguna, 2016). Web scrapping juga dikenal sebagai screen scrapping yaitu suatu teknik dalam program dimana komputer mengutip data dari tampilan keluaran program lain. Screen scrapping sering mengabaikan data biner (foto, data multimedia) dan format elemen, sehingga lebih cenderung ke data teks. Pada penelitian yang dilakukan oleh (Josi & Abdillah, 2014), teknik *web scrapping* merupakan suatu teknik yang sangat bermanfaat untuk mendapatkan data artikel ilmiah secara cepat dari halaman-halaman website. *Web scrapping* adalah legal selama tidak dilakukan untuk pencurian data karena *web scrapping* hanya berhubungan dengan metadata atau informasi terkait bibliography dari suatu artikel ilmiah. Sehingga dengan metode tersebut dapat diimplementasikan dalam system informasi Skripsi.

Arsitektur Aplikasi

Pada arsitektur aplikasi yang akan dibangun terdiri dari beberapa komponen yaitu terdapat *user* menginput dokumen teks yang akan diuji ke dalam sistem. Dokumen akan di proses untuk mencari dokumen pembanding dengan *web scraping*, fungsinya untuk mengambil hasil *searching* kalimat-kalimat dokumen uji di google *link* mana yang dikategorikan sama dengan dokumen uji. Teks dokumen dan url akan diproses terlebih dahulu sebelum masuk ke proses analisis kemiripan dokumen. Fungsi dari preproses yaitu untuk pengolahan data agar memproses pembuangan karakter-karakter yang tidak relevan seperti spasi, karakter-karakter tanda baca, selanjutnya akan di proses ketahap analisis kemiripan dokumen dengan metode Koefisien *Jaccard*. Gambaran arsitektur sistem disajikan pada Gambar 1.

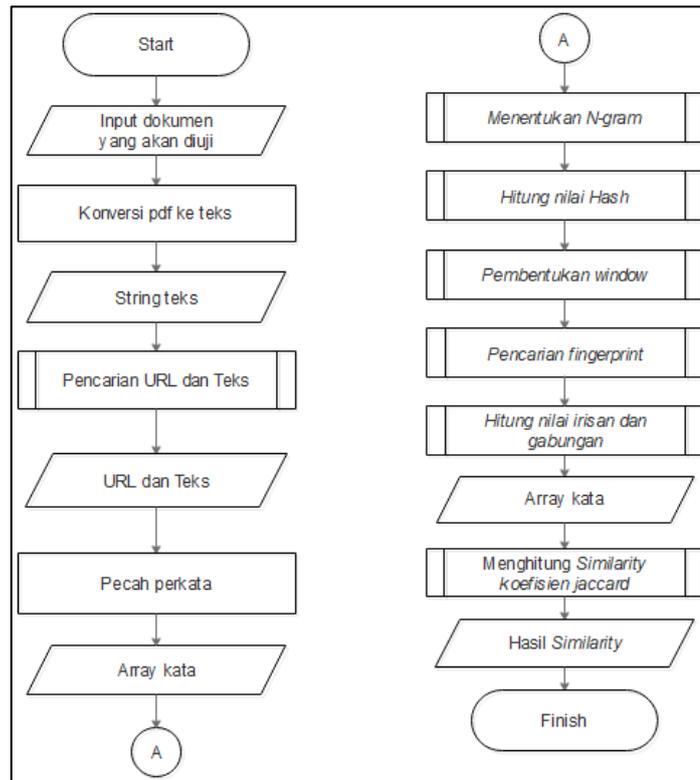


Gambar 1. Arsitektur aplikasi pendeteksian kemiripan dokumen

Implementasi Algoritme Wnnowing pada Pengecekan Plagiarisma

Algoritme Wnnowing digunakan dalam mendeteksi penjiplakan dokumen teks. Dalam alur perancangan proses deteksi kesamaan dokumen secara keseluruhan, dokumen dalam bentuk pdf dimasukkan kemudian diubah kedalam bentuk teks yang akan di jadikan sebagai kalimat dalam pencarian URL. Hasil URL akan dilakukan *preprocessing* untuk menghilangkan *noise* dalam kalimat yang didapat dan selanjutnya akan dihitung nilai *similarity*. Gambar 2 menampilkan alur proses pengecekan plagiarisme.

Diawali dengan input data teks proposal skripsi dan dilanjutkan dengan proses *web scrapping* untuk pencarian dokumen pembanding. Kemudian *preprocessing* dokumen dilakukan sehingga bisa digunakan untuk menentukan n-gram dan nilai *hash* dari masing-masing kata pada dokumen. Algoritme Wnnowing dimulai dari proses pembentukan *window* untuk melakukan pemilihan atau penyaringan kata-kata pada dokumen yang akan diproses. Selanjutnya dapat ditentukan fingerprint dari dokumen dan dokumen-dokumen pembanding sehingga dapat dihitung nilai similaritasnya dengan menggunakan koefisien Jaccard.



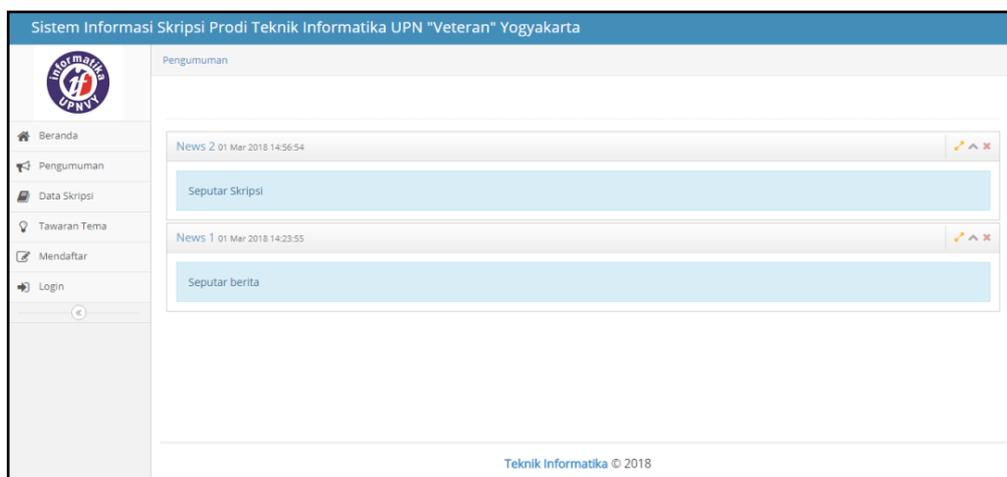
Gambar 2. Alur proses cek plagiarisme

3. HASIL DAN PEMBAHASAN

Aplikasi pengecekan plagiarisme yang dihasilkan memiliki fitur halaman pengumuman, halaman data skripsi, halaman proposal, dan halaman *review* proposal.

Halaman Pengumuman

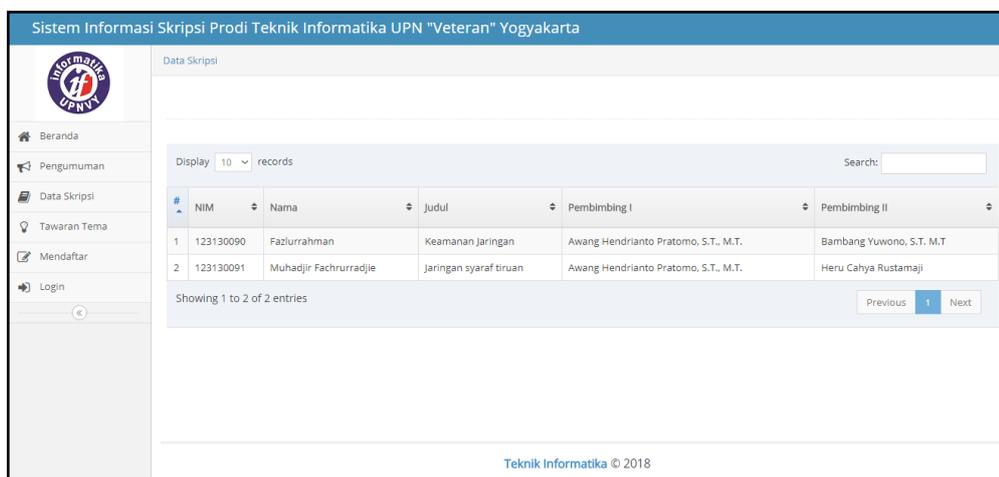
Halaman ini merupakan halaman pengumuman yang diinputkan oleh admin berisi informasi seputar skripsi, bisa berupa info pendaftaran seminar, pendaftaran pendadaran, dan juga seputar wisuda. Tampilan halaman pengumuman disajikan pada Gambar 3.



Gambar 3. Halaman pengumuman

Halaman Data Skripsi

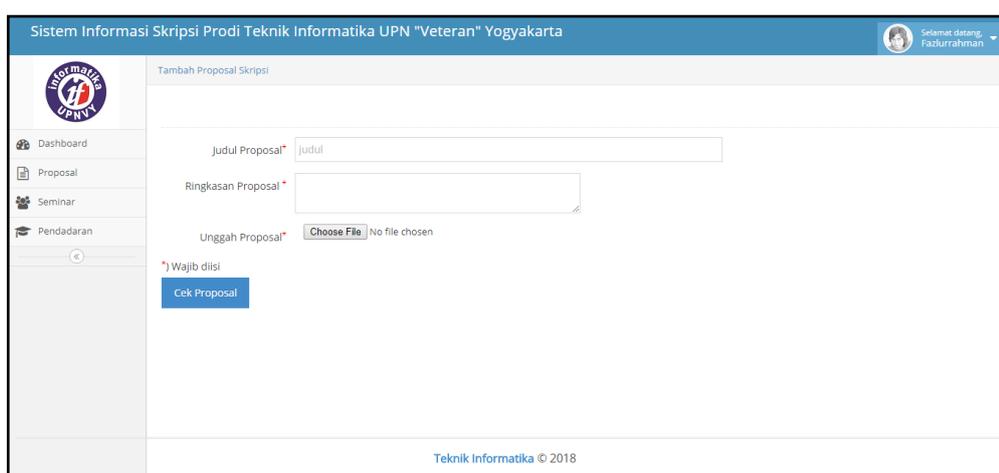
Halaman data skripsi berisi tentang informasi mahasiswa yang proposal skripsinya telah diterima. Terdapat info mengenai nim mahasiswa, nama, judul skripsi serta pembimbing mahasiswa tersebut.dengan judul skripsi dan pembimbing. Gambar 4 menunjukkan tampilan dari halaman data skripsi.



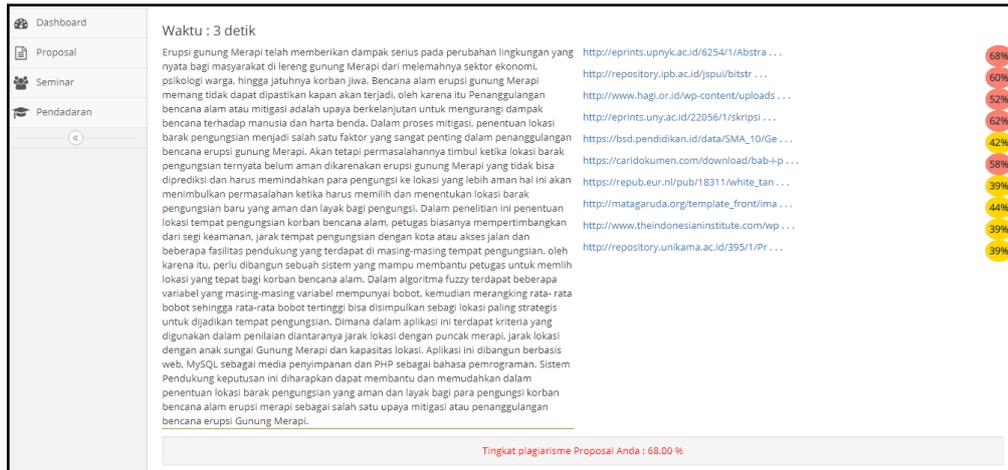
Gambar 4. Halaman Data Skripsi

Halaman proposal

Halaman proposal merupakan halaman yang digunakan dalam mengecek plagiarisme dan mengunggah proposal skripsi. Pada halaman ini juga digunakan untuk melihat informasi mengenai hasil dari pengajuan proposal skripsi yang telah diajukan, apakah diterima atau tidak. Informasi yang diberikan juga berisi tentang proposal skripsi, ringkasan proposal, kemiripan proposal tersebut dengan artikel lain, review dari dosen sebagai reviewer dan catatan dari koordinator skripsi serta surat pembimbing yang akan diberikan kepada dosen pembimbing yang bersangkutan. Gambar 5 menampilkan antarmuka dari halaman pengajuan proposal skripsi mahasiswa, sedangkan Gambar 6 menampilkan hasil dari pengecekan plagiarisme proposal skripsi mahasiswa.



Gambar 5. halaman pengajuan proposal skripsi mahasiswa

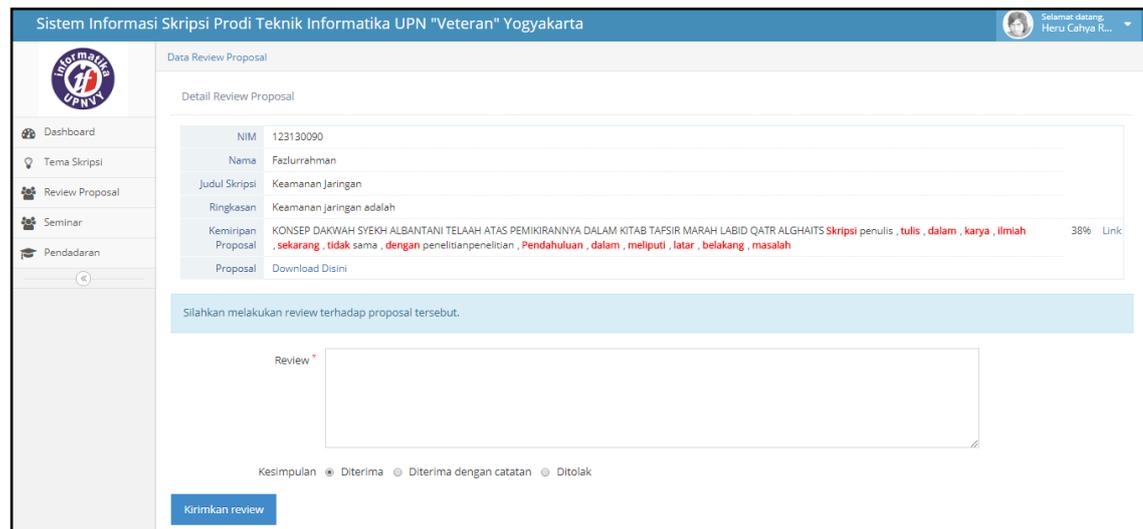


Gambar 6. Hasil cek plagiarisme proposal skripsi

Sebelum proposal dikirimkan ke admin selaku koordinator skripsi, proposal skripsi tersebut terlebih dahulu dilakukan pengecekan terhadap proposal tersebut. Jika persentase kemiripan proposal tersebut berada dibawah 50%, maka proposal tersebut dapat di kirimkan ke koordinator skripsi dan selanjutnya akan diproses. Namun jika persentase kemiripan proposal tersebut berada diatas 50%, maka proposal tersebut tidak dapat di kirimkan ke koordinator skripsi. Pada proses ini juga berisi informasi tentang data yang ada didalam proposal tersebut, link artikel yang memiliki kesamaan dengan proposal yang akan diunggah serta persentase kesamaan dari proposal tersebut dengan artikel yang didapat di internet.

Halaman *review* proposal

Halaman ini digunakan untuk melakukan proses *review* terhadap proposal mahasiswa yang telah di kirimkan oleh koordinator skripsi. Dosen yang bersangkutan dapat melakukan *review* dan memberikan penilaian terhadap proposal tersebut, apakah dapat diterima atau tidak. Gambar 7 menunjukkan tampilan dari halaman *review* proposal skripsi.



Gambar 7. Halaman dosen *review* proposal skripsi

Pengujian pengecekan plagiarisme

Pengujian pengecekan plagiarisme yang akan dilakukan yaitu membandingkan hasil pemrosesan pengecekan plagiarisme dari sistem sebelumnya (yang diteliti oleh Rina Amelia, S.T.) dengan sistem yang sedang dikembangkan. Pengujian berdasarkan pada nilai n-gram,

hasil yang dibandingkan dari kedua sistem tersebut terletak pada sisi persentase kemiripan dokumen dan kecepatan pemrosesan. Data pengujian menggunakan dua (2) buah data uji berupa proposal skripsi mahasiswa yang sudah diterima sebelumnya.

Pengujian Data 1

Proses pengujian yang dilakukan adalah mengunggah data uji pertama tersebut pada sistem lama dan sistem yang sedang dikembangkan. Pengujian menggunakan pengujian terhadap nilai n-gram 2 sampai 5, tujuan dari pengujian ini untuk melihat persentase kemiripan dan waktu proses tiap-tiap n-gram pada sistem yang berbeda. Hasil dari proses pendeteksian plagiarisme tersebut disajikan pada Tabel 1.

Tabel 1. Tabel pengujian data 1

No	Jumlah Kata Uji	n-gram	Waktu Pemrosesan (detik)		Persentase Kemiripan (%)	
			Sistem Lama	Sistem Baru	Sistem Lama	Sistem Baru
1	294	2	20	9	48.92	76.00
2	294	3	17	6	40.22	69.00
3	294	4	15	4	34.56	61.00
4	294	5	15	3	31.69	57.00

Pengujian Data 2

Proses pengujian yang dilakukan sama seperti yang dilakukan pada pengujian data 1. Namun pada pengujian 2 akan menggunakan data 2 untuk diuji. Pengujian data 2 juga menggunakan nilai n-gram 2 sampai 5, tujuan dari pengujian ini untuk melihat persentase kemiripan dan waktu proses tiap-tiap n-gram pada sistem yang berbeda. Hasil dari proses pendeteksian plagiarisme tersebut disajikan pada Tabel 2.

Tabel 2. Pengujian data 2

No	Jumlah Kata Uji	n-gram	Waktu Pemrosesan (detik)		Persentase Kemiripan (%)	
			Sistem Lama	Sistem Baru	Sistem Lama	Sistem Baru
1	255	2	21	8	71.09	89.00
2	255	3	17	6	64.78	82.00
3	255	4	14	5	58.39	74.00
4	255	5	10	2	50.50	68.00

Setelah dilakukan uji coba, dapat diketahui bahwa pemilihan n-gram yang semakin kecil cenderung mempunyai nilai persentase kemiripan yang besar. Hal ini karena pada n-gram yang lebih sedikit, *string* yang dipotong lebih kecil sehingga kemungkinan untuk ditemukannya rangkaian karakter yang sama semakin besar. Semakin besarnya n-gram, maka mengandung karakter yang lebih banyak dibandingkan dengan n-gram yang lebih kecil sehingga menyebabkan rangkaian karakter yang ditemukan semakin berkurang. Namun jika nilai n-gram terlalu kecil, ada kemungkinan memiliki persentase kemiripan yang tidak sesuai dengan kata aslinya. Oleh karena itu digunakan nilai n-gram yang tidak terlalu besar dan tidak terlalu kecil.

KESIMPULAN

Berdasarkan dari hasil analisis, perancangan dan implementasi yang telah dilakukan dapat disimpulkan bahwa pengembangan sistem informasi skripsi dengan menerapkan metode *Winnowing* untuk deteksi *similarity* dokumen (studi kasus Teknik Informatika UPN "Veteran" Yogyakarta) telah berhasil dirancang dan dibangun. Sistem ini dibangun untuk digunakan oleh

mahasiswa, dosen dan Koordinator Skripsi untuk mempermudah dalam mengajukan, memproses serta memberi penilaian terhadap proposal skripsi yang diajukan. Dari hasil penelitian diketahui bahwa pemilihan n-gram yang semakin kecil akan memperoleh nilai persentase kemiripan yang besar (76% dan 89%). Hal ini terjadi karena pada n-gram yang lebih sedikit, *string* yang dipotong lebih kecil sehingga kemungkinan untuk ditemukannya rangkaian karakter yang sama semakin besar. Semakin besarnya n-gram, maka mengandung karakter yang lebih banyak dibandingkan dengan n-gram yang lebih kecil sehingga menyebabkan rangkaian karakter yang ditemukan semakin berkurang sehingga menurunkan nilai similaritas (hingga 57% dan 68%).

DAFTAR PUSTAKA

- Aulia, R. (2008). Analisis Algoritme Knuth Morris Pratt dan Algoritme Boyer Moore dalam Proses Pencarian String. Bandung: ITB.
- Djafar, I. M. (2015). Deteksi Kemiripan Dokumen Teks Menggunakan Algoritme Manber. Deteksi Kemiripan Dokumen Teks Menggunakan Algoritme Manber.
- Firdaus, H. B. (2008). Deteksi Plagiat Dokumen Menggunakan Algoritme Rabin-Karp. Program Studi Teknik Informatika Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung (ITB). Bandung. Diambil dari <http://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2007-2008/Makalah2008/MakalahIF2251-2008-076.pdf>
- Goenawan, W., Augustinus, R., & Sembiring, K. (2005). Penerapan Algoritme Edit Distance Pada Pendeteksian Praktik Plagiat. *Bandung: Institut Teknologi Bandung*.
- Hidayani, N., Sari, J. N., & Suhatman, R. (2012). Perancangan dan Implementasi Metode Brute Force untuk Pencarian String pada Website PCR. *Jurnal Aksara Komputer Terapan*, 1(2).
- Istiana, P., Purwoko, & Perpustakaan UGM. (2014, Januari 26). Panduan Anti Plagiarism. Diambil 31 Januari 2018, dari http://lib.ugm.ac.id/data/panduan_plagiarisme.pdf
- Josi, A., & Abdillah, L. A. (2014). Penerapan teknik web scraping pada mesin pencari artikel ilmiah. *arXiv preprint arXiv:1410.5777*.
- Rahma, Mivthahul. (2015). *Cara-Cara Konversi Sistem Informasi*. Bogor: Institut Pertanian Bogor.
- Kurniawati, A., & Wicaksana, I. W. S. (2008). PERBANDINGAN PENDEKATAN DETEKSI PLAGIARISM DOKUMEN DALAM BAHASA INGGRIS. *Proceeding, Seminar Ilmiah Nasional Komputer dan Sistem Intelijen (KOMMIT 2008)*. Diambil dari <http://repository.gunadarma.ac.id/977/>
- Mulyanto, A. (2010). ANALISIS EDIT DISTANCE MENGGUNAKAN ALGORITME DYNAMIC PROGRAMMING, Vol 5, No 2. Diambil dari http://repository.ung.ac.id/get/simlit_res/1/281/Analisis-Edit-Distance-Menggunakan-Algoritme-Dynamic-Programming.pdf
- Munir, R. (2004). *Algoritme Brute Force Bagian 2–Algoritme Brute Force (Lanjutan)*. ppt. ITB, Bandung.
- Pramana, Y., & Mungguna, W. (2016). Implementasi Web Scrapping pada Website Eduesia. Com untuk Pengukur Kesenjangan Jumlah Mahasiswa Perguruan Tinggi di Indonesia. *ULTIMA INFOSYS*, 6(2), 130–135.
- Purwitasari, D., Kusmawan, P. Y., & Yuhana, U. L. (2011). Deteksi Keberadaan Kalimat Sama Sebagai Indikasi Penjiplakan Dengan Algoritme Hashing Berbasis N-Gram. *Jurnal Ilmiah KURSOR*, 6(1), 37–44. Diambil dari http://old.its.ac.id/personal/files/pub/4745-diana-cs-kursor_Diana.pdf
- Saragih, M. A. (2013). Implementasi Algoritme Brute Force dalam Pencocokan Teks Font Italic untuk Kata Berbahasa Inggris pada Dokumen Microsoft Office Word. *Pelita Informatika: Informasi dan Informatika*, 4(3).
- Fadelillah, M., Subroto, I. M. I., & Kurniadi, D. (2017). Sistem Rekomendasi Hasil Pencarian Artikel Menggunakan Metode Jaccard's Coefficient. *TRANSISTOR Elektro dan Informatika*, 2(1), 1–14.
- Schleimer, S., Wilkerson, D. S., & Aiken, A. (2003). Winnowing: local algorithms for document fingerprinting. Dalam *Proceedings of the 2003 ACM SIGMOD international conference*

- on Management of data* (hlm. 76–85). ACM. Diambil dari <http://theory.stanford.edu/~aiken/publications/papers/sigmod03.pdf>
- Jumardi, R., Nugroho, E., & Hidayah, I. (2015). Analisis Kesuksesan Implementasi Sistem Informasi Skripsi pada Program Studi Teknik Informatika Universitas Pembangunan Nasional “Veteran” Yogyakarta. Dalam Seminar Nasional Aplikasi Teknologi Informasi (SNATI) (Vol. 1).
- Soelistyo, H. (2011). *Plagiarisme: pelanggaran hak cipta dan etika*. Penerbit Kanisius. Diambil dari <https://books.google.co.id/books?id=xhwptwAACAAJ>
- Palupi, Y. E., & Suroso, B. D. I. A. I. (2014). Pengembangan Sistem Informasi.
- Triawati, C. (2009). Metode Pembobotan Statistical Concept Based untuk Klustering dan Kategorisasi Dokumen Berbahasa Indonesia. *Institut Teknologi Telkom, Bandung*.
- Turland, M. (2010). *Php, Architect's Guide to Web Scraping*. Marco Tabini & Associates, Inc.
- Zuliarso, E., & Mustofa, K. (2009). Crawling Web berdasarkan Ontology. *Dinamik-Jurnal Teknologi Informasi, 14*(2).