

BAB II. LANDASAN TEORI

2.1 Tinjauan Pustaka

Beberapa penelitian yang dijadikan rujukan dalam penelitian ini ditampilkan pada Tabel 2.1 sebagai berikut:

Tabel 2. 1 Penelitian terdahulu

No	Judul	Penulis/Jurnal	Univ/Tahun	Permasalahan	Metode	Hasil
1	Sistem Deteksi Plagiarisme Lintas Bahasa Menggunakan Algoritma TF-IDF	Stefania N. Lolyta, Rocky Y. Dillak, Folkes E. Laumal Jurnal Ilmiah Flash Sanditya Cakti Bavana	Politeknik Negeri Kupang 2019	Terjadinya tindak plagiarisme pada penulisan karya tulis	TF-IDF	Metode tersebut dapat mendeteksi kesamaan dokumen yang diuji dan dokumen pembanding yang menghasilkan selisih 0% antara perhitungan sistem dan manual. Banyak dokumen yang diproses berpengaruh terhadap lamanya waktu pemrosesan (Lolyta et al., 2019)

2	Implementation of <i>Winnowing</i> Algorithm with Dictionary English-Indonesia Technique to Detect Plagiarism	Sunardi Anton Yudhana, Iif Alfiatul Mukaromah International Journal of Advanced Computer Science and Applications (IJACSA)	Universitas Ahmad Dahlan Yogyakarta 2018	Pendeteksian plagiarisme otomatis belum ada yang dapat memeriksa kutipan dari bahasa asing.	<i>Winnowing</i>	Hasil akurasi yang dihasilkan jika menggunakan kamus Inggris-Indonesia lebih tinggi dibandingkan yang tidak menggunakan kamus Inggris-Indonesia (Yudhana et al., 2018).
3	Aplikasi Pendeteksi Tingkat Kesamaan Dokumen Teks: Algoritma <i>Rabin Karp</i> Vs. <i>Winnowing</i>	Sugiono, Herwin, Hamdani, Erlin Jurnal Teknologi Informasi & Komunikasi Digital Zone	STMIK Amik Riau 2018	Terdapat tindakan plagiat terhadap karya orang lain secara bebas melalui internet	<i>Rabin Karp</i> dan <i>Winnowing</i>	Algoritma <i>Rabin Karp</i> memiliki akurasi sebesar 37,5% dengan waktu proses selama 0,19 detik sedangkan algoritma <i>Winnowing</i> menghasilkan akurasi 88,89% dengan waktu proses selama 0,13 detik (Sugiono et al., 2018).

2.2 Dasar Teori

2.2.1 Plagiarisme

Plagiarisme merupakan tindakan tidak jujur yang disebabkan karena kurangnya kreatifitas dan inovasi dalam menciptakan karya yang asli atau orisinal (Leman et al., 2019). Dalam mencari kesamaan antara dokumen teks seperti jurnal dilakukan pengecekan plagiarisme yang berguna untuk mengecek keaslian dari jurnal tersebut (Pardede & Alvian, 2015).

Plagiarisme memiliki tiga tipe yaitu (Novantara, 2018):

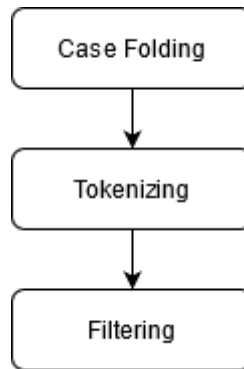
- a. *Word-for-word* plagiarism yang artinya setiap kata yang ada disalin secara langsung tanpa mengubahnya sama sekali.
- b. *Plagiarism of authorship* artinya menyatakan hasil karya orang lain sebagai miliknya dengan mengganti nama pemilik asli sebagai namanya.
- c. *Plagiarism of ideas* artinya menyatakan pemikiran atau gagasan orang lain sebagai miliknya.

Pada pendeteksian plagiarisme terdapat tiga macam metode yaitu metode perbandingan teks lengkap, metode dokumen *fingerprint* dan metode kesamaan kata kunci (Dwi Susanto, Achmad Basuki, 2016).

2.2.2 *Text Mining* dan *Text Preprocessing*

Text mining merupakan proses pencarian pola atau penggalian informasi dari data teks untuk menghasilkan informasi baru (Aida et al., 2018). Sumber data yang digunakan pada *text mining* adalah sebuah kumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur (Tria & Achmad, 2017).

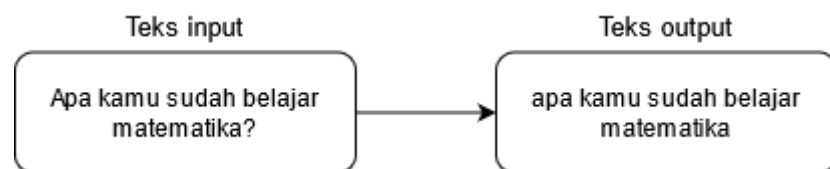
Text preprocessing adalah proses yang sering digunakan dalam melakukan *text mining* (Filcha & Hayaty, 2019). Tahap *preprocessing* terdapat 4 macam meliputi *case folding*, *filtering*, *stemming* dan *tokenizing* (Wibowo & Hastuti, 2016). Namun pada penelitian ini hanya dilakukan tiga tahap *preprocessing* seperti terlihat pada Gambar 2.1 berikut.



Gambar 2. 1 Tahap *Preprocessing*

a. *Case Folding*

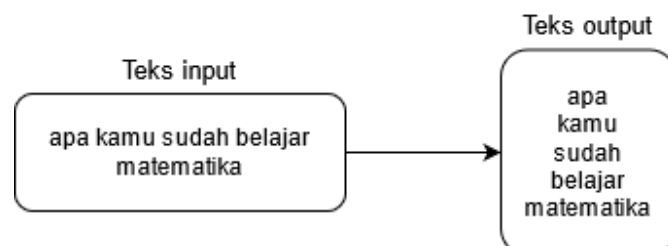
Case folding merupakan tahap mengubah semua huruf menjadi huruf kecil dan menghilangkan karakter yang tidak penting sehingga hanya menyisakan huruf a sampai z saja (Harijanto et al., 2018). Gambar 2.2 berikut merupakan contoh hasil dari proses *case folding*.



Gambar 2. 2 Proses *Case Folding*

b. *Tokenizing*

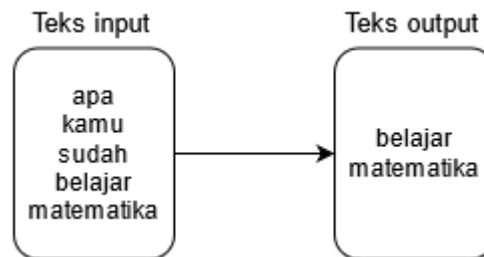
Tokenizing/parsing merupakan tahap di mana *string input* dipotong berdasarkan tiap kata penyusunnya. Hasil dari proses *tokenizing* ditampilkan seperti Gambar 2.3 berikut.



Gambar 2. 3 Proses *Tokenizing*

c. *Filtering*

Filtering merupakan tahap di mana suatu kalimat hanya diambil kata-kata yang penting atau memiliki arti relevan menggunakan algoritma *stopword* (Widiastuti et al., 2017). Contoh hasil dari proses *filtering* ditampilkan seperti Gambar 2.4 berikut.



Gambar 2. 4 Proses *Filtering*

2.2.3 Metode *Winnowing*

Metode *winnowing* merupakan suatu metode di mana mengambil atribut yang penting dari setiap *fingerprint* untuk mendeteksi kesamaan antar kalimat (Schleimer et al., 2003). Metode ini diperkenalkan pertama kali pada tahun 2003 oleh Schleimer et al. dengan jurnal yang berjudul *Winnowing: Local Algorithms for Document Fingerprinting*.

Sedangkan menurut jurnal yang disusun oleh Sunardi et al., metode *Winnowing* merupakan suatu metode yang berfungsi sebagai dokumen *fingerprint* atau metode yang digunakan untuk mendeteksi tindakan plagiarisme dengan menggunakan teknik *hashing* (Sunardi et al., 2018). *Fingerprint* merupakan hasil dari sekumpulan nilai *hash* yang digunakan sebagai dasar pembandingan antar *file* teks (Nurdin & Amin, 2017).

Langkah dalam perhitungan kemiripan teks dengan *Winnowing* adalah sebagai berikut (Nurdin & Amin, 2017):

- a. Tahap pertama yang dilakukan adalah membuang karakter yang tidak penting seperti tanda baca, spasi dan angka. Namun pada skripsi ini terdapat tambahan proses *case folding*, *tokenizing* dan *filtering* atau *stopword removal*. Pada *stopword removal* berguna untuk membuang kata yang tidak relevan. Misalnya pada kalimat *Hari ini cuacanya cerah* maka akan diubah menjadi *cuacanyacerah*.
- b. Tahap kedua adalah membentuk rangkaian gram dari hasil proses sebelumnya. *N-gram* merupakan hasil potongan dari tiap karakter. Perbedaan untuk hasil tiap *n-gram* terletak pada pembagian karakternya. Jika *n* yang digunakan 2 maka kata akan dibagi per 2 karakter, sedangkan jika *n* adalah 3 maka kata akan dibagi per 3 karakter (Najjichah et al.,

2019). Contoh hasil pembentukan rangkaian *n-gram* dengan nilai $n = 3$ adalah sebagai berikut *cua uac aca can any nya yac ace cer era rah*.

- c. Tahap ketiga melakukan perhitungan nilai hash dari setiap gram menggunakan *rolling hash*. *Rolling hash* adalah suatu cara untuk mentransformasikan sebuah *string* menjadi suatu nilai yang unik dengan panjang tertentu dan berfungsi sebagai penanda *string* tersebut. Dalam perhitungan nilai *hash* menggunakan rumus Persamaan 2.1 berikut.

$$H_{(ck)} = c1 * b^{(k-1)} + c2 * b^{(k-2)} + \dots + ck * b^{(k-k)} \quad (2.1)$$

Keterangan:

c = kode ASCII dari karakter

b = basis (bilangan prima)

k = banyak karakter

Hasil perhitungan nilai *hash* dengan besar basis 11 adalah 13363, 15323, 12923, 13156, 13068, 14738, 15807, 12927, 13204, 13572 dan 14965.

- d. Lalu tahap selanjutnya membentuk *window* dari nilai *hash*. *Window* merupakan pembagian atau pengelompokan beberapa nilai *hash* dengan ukuran yang ditentukan. Berikut merupakan hasil pembentukan *window* dengan nilai $w = 3$, {13363, 15323, 12923}, {15323, 12923, 13156}, {12923, 13156, 13068}, {13156, 13068, 14738}, {13068, 14738, 15807}, {14738, 15807, 12927}, {15807, 12927, 13204}, {12927, 13204, 13572}, {13204, 13572, 14965}.
- e. Tahap terakhir adalah memilih nilai terkecil dari setiap *window* untuk dijadikan *fingerprint*. *Fingerprint* merupakan hasil dari pembentukan *window* dari pemilihan nilai *hash* terkecil (Sunardi, A. Yudhana, 2017). Hasil *fingerprint* adalah 12923, 13068, 12927 dan 13204.

2.2.4 Jaccard Coefficient

Jaccard Coefficient merupakan sebuah persamaan yang digunakan untuk menghitung kemiripan (*similarity*) dari hasil pembentukan *fingerprint* sehingga dihasilkan persentase kemiripan teks (Alamsyah et al., 2019). Persamaan *Jaccard Coefficient* dituliskan pada Persamaan 2.2 berikut:

$$\text{Persentase Kemiripan} = \frac{\text{Jumlah fingerprint sama}}{\text{Total seluruh fingerprint}} \times 100\% \quad (2.2)$$

2.2.5 Dasar Hukum Plagiarisme

Undang-Undang yang mengatur tentang Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi diatur dalam UU No. 17 tahun 2010. Pada Bab II Pasal 3 dituliskan yang termasuk plagiator di perguruan tinggi adalah satu atau lebih mahasiswa, satu atau lebih dosen/peneliti/tenaga kependidikan atau satu atau lebih dosen/peneliti/tenaga kependidikan bersama satu atau lebih mahasiswa.

Salah satu ayat yang menjelaskan tindakan pencegahan terhadap plagiat dijelaskan pada Bab IV Pasal 7 yaitu pada setiap karya ilmiah yang dihasilkan di lingkungan perguruan tinggi harus dilampirkan pernyataan yang ditandatangani oleh penyusunnya bahwa:

- a. Karya ilmiah tersebut bebas plagiat
- b. Apabila kemudian hari terbukti terdapat plagiat dalam karya ilmiah tersebut, maka penyusunnya bersedia menerima sanksi sesuai ketentuan peraturan perundang-undangan.

Sanksi yang diberikan bagi mahasiswa apabila terlibat dalam tindakan plagiat dijelaskan pada Bab VI Pasal 12 ayat (1) sebagai berikut:

- a. Teguran
- b. Peringatan tertulis
- c. Penundaan pemberian sebagian hak mahasiswa
- d. Pembatalan nilai satu atau beberapa mata kuliah yang diperoleh mahasiswa
- e. Pemberhentian dengan hormat dari status sebagai mahasiswa
- f. Pemberhentian tidak dengan hormat dari status sebagai mahasiswa; atau
- g. Pembatalan ijazah apabila mahasiswa telah lulus dari suatu program.

Sedangkan sanksi yang diberikan bagi dosen/peneliti/tenaga kependidikan yang terbukti melakukan tindakan plagiat sebagaimana dimaksudkan pada Pasal 11 ayat (6) dijelaskan pada Bab VI Pasal 12 ayat (2) sebagai berikut:

- a. Teguran
- b. Peringatan tertulis
- c. Penundaan pemberian hak dosen/peneliti/tenaga kependidikan
- d. Penurunan pangkat dan jabatan akademik/fungsional

- e. Pencabutan hak untuk diusulkan sebagai guru besar/profesor/ahli peneliti utama bagi yang memenuhi syarat
- f. Pemberhentian dengan hormat dari status sebagai dosen/peneliti/tenaga kependidikan
- g. Pemberhentian tidak dengan hormat dari status sebagai dosen/peneliti/tenaga kependidikan; atau
- h. Pembatalan ijazah yang diperoleh dari perguruan tinggi yang bersangkutan.