

# Penerapan Algoritma Smith Waterman Untuk Mengukur Kemiripan Tugas Kuliah Mahasiswa

## *Implementation of Smith Waterman Algorithm In Measuring of Student Task Study*

Liliyana Amsir<sup>1)</sup>, Harlinda<sup>2)</sup>, Siska Anraeni<sup>3)</sup>, Poetri Lestari Lokapitasari Belluano<sup>4)</sup>

<sup>1,2,3,4</sup> Fakultas Ilmu Komputer, Universitas Muslim Indonesia

<sup>1,2,3,4</sup> Jl. Urip Sumohardjo Km.05 Gedung Fakultas Ilmu Komputer Kampus II UMI Tlp.(0411) 453009 Makassar 90231

E-mail: liliyanaamsir@gmail.com<sup>1)</sup>, harlinda@umi.ac.id<sup>2)</sup>, siska.anraeni@umi.ac.id<sup>3)</sup>, poetrilestari@umi.ac.id<sup>4)</sup>

**Abstrak** – Plagiat telah banyak terjadi dalam dunia pendidikan, khususnya di kalangan pelajar (mahasiswa). Maka dari itu, dibutuhkan sebuah sistem yang dapat mendeteksi kemiripan tugas kuliah mahasiswa. Metode yang diterapkan dalam sistem ini adalah algoritma smith Waterman. Proses pendeteksian menggunakan algoritma smith waterman menggunakan 3 tahap yaitu tahap penentuan skor, tahap penentuan jalur traceback dan tahap penentuan skor akhir. Sistem ini dapat dikatakan efektif, karena telah dapat melakukan 3 tahap dari proses perhitungan algoritma smith waterman dan telah dapat menampilkan skor akhir. Nilai skor akhir yang dihasilkan dalam sistem pendeteksian kemiripan tugas kuliah mahasiswa merupakan skor kemiripan tugas kuliah mahasiswa, semakin banyak karakter yang sama, maka semakin tinggi nilai skor yang dihasilkan. Sebaliknya, semakin banyak karakter yang tidak sama, maka semakin rendah nilai skor yang dihasilkan.

**Kata Kunci:** Traceback, Algoritma Smith Waterman, skor, gap

**Abstract-** *Plagiarism has been doing in education especially in students' community. So that why, system is needed what can detect of similarity of student's task. This research using Smith Waterman algorithm method through three steps such as determining of scores, determining of traceback route and determining of final score. This system can be claim as effective because this system can be past three steps of counting Smith Waterman algorithm until view the final score. The last score what produce from this similarity students task detection system is similarity score of student's tasks. More and more of similar character then similar score also be higher and etc.*

**Keywords:** *Traceback, Smith Waterman Score Algotirhm, Score, gap*

### PENDAHULUAN

Dalam suatu instansi akademik, tugas bukan lagi hal yang asing didengar di kalangan pelajar. pada proses belajar mengajar tentu saja tidak bisa lepas dari tugas. seperti halnya yang terjadi di Fakultas Ilmu Komputer Universitas Muslim Indonesia. Tugas-tugas yang diberikan oleh tenaga pengajar kepada mahasiswa biasanya diberikan tenggang waktu atau batas pengumpulan.

Salah satu kendala bagi mahasiswa adalah malasnya mengerjakan tugas yang diberikan oleh tenaga pengajar, sehingga tidak jarang diantara mereka yang hanya meniru sebagian bahkan ada beberapa yang meniru keseluruhan jawaban temannya.

Seiring perkembangan teknologi, mulai bermunculan metode-metode yang dapat diterapkan pada aplikasi untuk mengukur tingkat akurasi kemiripan tugas kuliah mahasiswa, salah satunya

adalah Algoritma Smith Waterman. Algoritma Smith Waterman adalah perluasan algoritma pencocokan string pada teks atau barisan sebagai salah satu implementasi program dinamis. Algoritma ini akan membandingkan keseluruhan panjang 2 sekuen (sederet pernyataan-pernyataan yang urutan dan pelaksanaan eksekusinya runtut) yang terbagi menjadi sub bagian untuk mendapatkan kesamaan tertinggi antara kedua sekuen.

Djafar, dkk (2014) dalam penelitiannya menjelaskan bahwa Algoritma Smith Waterman dapat digunakan dalam mendeteksi kesamaan dokumen. Dengan menggunakan metode yang sama, Darmawan, dkk (2009) dalam penelitiannya menjelaskan bahwa Algoritma Smith Waterman dapat digunakan untuk mengatasi plagiarism yang saat ini banyak terjadi.

Dengan menerapkan Algoritma Smith Waterman, akan diperoleh sebuah nilai yang kemudian

nilai yang dihasilkan akan diolah sehingga diperoleh persentase kemiripan tugas kuliah mahasiswa.

Kasus meniru tugas kuliah yang banyak dilakukan oleh mahasiswa jika dibiarkan terjadi dapat berpengaruh negative terhadap tenaga pengajar dan proses belajar mahasiswa. padahal peranan tenaga pengajar sangat dibutuhkan dalam menanggulangi permasalahan tersebut. keaktifan serta ketelitian tenaga pengajar dalam memeriksa tugas kuliah mahasiswa sangat diperlukan. oleh karena itu, penulis merancang sebuah sistem untuk mengukur persentase kemiripan tugas kuliah mahasiswa, sehingga dapat membantu tenaga pengajar dalam melihat persentase kemiripan tugas kuliah mahasiswa.

## METODOLOGI PENELITIAN

### 1. Penjajaran Sekuens

Penjajaran sekuens adalah proses untuk menyusun 2 atau lebih sekuens sehingga sekuens-sekuens tersebut memiliki hubungan fungsional, structural atau evaluasi. Hasil proses tersebut juga disebut sebagai sequence alignment atau alignment. Baris sekuens dalam suatu alignment diberi sisipan sedemikian rupa sehingga kolomnya memuat karakter yang identik atau serupa di antara sekuens-sekuens tersebut (Afif Bambang, 2017).

### 2. Algoritma Smith Waterman

Algoritma Smith Waterman adalah metode yang dapat mengidentifikasi local similarities (penjajaran sekuens) yaitu proses penyusunan dua local sequences (rangkaian/susunan atau rentetan). Berdasarkan fungsi proses penyejajaran sekuens, algoritma ini dapat diterapkan dalam pemrograman computer untuk pendeteksian kesamaan atau pengukuran tingkat kemiripan sebuah teks dengan dengan teks lain (Thalib dan Ratih, 2014).

Algoritma Smith Waterman adalah metode pemrograman dinamis untuk mendapatkan keselarasan lokal yang optimal antara dua urutan. Algoritma ini menemukan segmen serupa dalam dua langkah: (1) perhitungan matriks kesamaan H dan (2) backtracing (Yuma, dkk, 2014).

Algoritma Smith Waterman menyejajarkan dua urutan berdasarkan kecocokan/ketidakcocokan (juga dikenal sebagai substitusi), penyisipan, dan penghapusan. Baik penyisipan dan penghapusan adalah operasi yang memperkenalkan celah, yang diwakili oleh tanda hubung.

Algoritma standar Smith-Waterman yang dipakai untuk penghitungan local alignment dengan memperhatikan beberapa hal yakni (Djafar, dkk, 2009):

- 1 Menambahkan sebuah nilai pada setiap perbandingan.
  - a. Menggunakan nilai positif, apabila memiliki kemiripan.
  - b. Menggunakan nilai negatif, apabila tidak memiliki kemiripan.
- 2 Inisialisasi awal matriks dengan nilai 0 (nol).
- 3 Isi matriks penilaian menggunakan persamaan dibawah ini:

$$H_{ij} = \max \left\{ \begin{array}{l} H[i-1, j-1] + S(i, j), \\ H[i, j-1] + W, \\ H[i-1, j] + W \end{array} \right\} \dots \dots \dots (1)$$

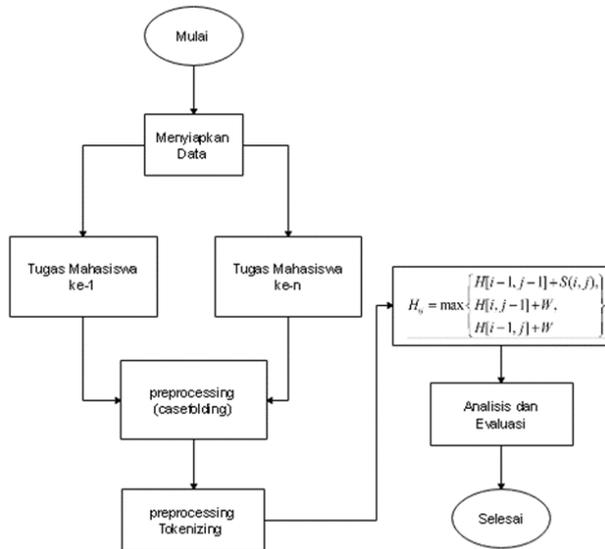
Dimana :

H : Skor dari Matriks  
 i : Baris  
 j : Kolom  
 S : Skor Perbandingan Karater  
 W : Gap Perbadingan Karakter

4. Memulai traceback dari nilai yang tertinggi yang ditemukan dimanapun pada matriks.
5. Menjajarkan sekuens berdasarkan traceback yang diperoleh. Jika arah traceback keatas, maka beri simbol garis datar (-) pada sekuens atas. . Jika arah traceback kesamping, maka beri simbol garis datar (-) pada sekuens bawah. . Jika arah traceback diagonal, maka tulis kedua sekuens.
6. Lakukan perhitungan skor dengan cara menambahkan nilai dari masih-masing skor yang dimiliki oleh sekuens.

### Alur Kerja Sistem

Berdasarkan flowchart pada gambar 1, dapat dilihat bahwa sebelum melakukan pendeteksian kemiripan tugas kuliah, mahasiswa harus terlebih dahulu menginputkan tugas. Tugas yang diinputkan oleh mahasiswa inilah yang akan masuk ketahap preprocessing kemudian dilanjutkan ke tahap pendeteksian skor kemiripan dengan menggunakan algoritma smith waterman



Gambar 1. Flowchart Sistem

|   |    |   |                    |  |
|---|----|---|--------------------|--|
| 5 | B1 | 1 | Technopren uership | Sistem Monitoring Ruang                  |
| 6 | B1 | 1 | Technopren uership | Absen Elektronik                         |
| 7 | B1 | 1 | Technopren uership | Sistem Resep Elektronik                  |
| 8 | B1 | 1 | Technopren uership | Sistem Informasi Percetakan              |
| 9 | B1 | 1 | Technopren uership | Aplikasi Lowongan Pekerjaan Berbasis WEB |

### 3 Pengumpulan Data

Data yang digunakan dalam penelitian ini sebanyak 11 data isi tugas kuliah mahasiswa dari 2 matakuliah yang ada di fakultas Ilmu Komputer, UMI (Metode Penelitian dan Technoprenuership). Pada proses perhitungan menggunakan algoritma smith waterman, yang menjadi data uji adalah sebagai berikut:

Tabel 1. DATA TUGAS MATA KULIAH MAHASISWA

| No | Kelas | Tugas Ke- | Matakuliah        | Isi Tugas   |
|----|-------|-----------|-------------------|---|
| 1  | B1    | 1         | Metode Penelitian | Aplikasi Rental Play Station Batara Indo Game                                     |
| 2  | B1    | 1         | Metode Penelitian | Pengontrolan Lampu Teras Dengan Sensor Intensitas Cahaya Berbasis Mikrokontroller |
| 3  | B1    | 1         | Metode Penelitian | Penerapan Metode Certainty Factor dalam Mendiagnosa Kanker Nasofaring             |
| 4  | B1    | 1         | Metode Penelitian | Sistem Informasi Percetakan   |

### 4 Tahap Preprocessing

Mengacu dari Tabel 1, setelah semua data uji selesai dikumpulkan, tahap selanjutnya adalah tahap preprocessing casefolding, yaitu mengubah semua huruf capital menjadi huruf non capital. Setelah tahap casefolding selesai, masuk kedalam tahap preprocessing tokenizing yaitu menghapus semua karakter special atau karakter yang tidak dibutuhkan. Berikut merupakan tabel hasil casefolding dan tokenizing

Tabel 2. Data Hasil Casefolding dan Tokenizing

| Isi Tugas   |
|---|
| aplikasirentalplaystationbataraindogame                                     |
| pengontrolanlampu teras dengansensorintensitascahayaberbasismikrokontroller |
| penerapanmetodecertaintyfactordalam mendiagnosakanker nasofaring            |
| sisteminformasipercetakan   |
| Sistemmonitoringruangan   |
| Absenelektronik   |
| Sistemresepelektronik   |
| sisteminformasipercetakan   |
| aplikasilowonganpekerjaanberbasisweb  |



Jumlahkan semua skor traceback untuk mendapatkan skor kemiripan tugas kuliah mahasiswa menggunakan algoritma smith waterman. berikut table skor akhir kemiripan tugas kuliah mahasiswa.

**Tsbel 6** SKOR AKHIR KEMIRIPAN TUGAS KULIAH MAHASISWA

| No | Sample   | Skor Akhir |
|----|----------|------------|
| 1  | Sample 1 | 125        |

**HASIL DAN PEMBAHASAN**

**1 Analisis Hasil Deteksi Kemiripan Tugas kuliah**

Rumus untuk memperoleh nilai persentase terhadap kemiripan tugas kuliah mahasiswa yang dibandingkan adalah sebagai berikut:

$$\text{persentase} = ((jkM / jka) + (jkM / jkb) / 2) * 100 )$$

keterangan :

- jkM : Jumlah karakter yang sama
- jka : Jumlah karakter tugas mahasiswa ke-1
- jkb : Jumlah karakter tugas mahasiswa ke-n

Nilai persentase yang dihasilkan untuk setiap tugas yang dibandingkan bergantung pada banyaknya karakter yang sama dan jumlah karakter dari masing-masing tugas yang dibandingkan. Nilai skor yang didapat dari perhitungan menggunakan algoritma smith waterman tidak dimasukkan dalam perhitungan untuk mendapatkan persentase kemiripan. Sehingga, nilai persentase yang dihasilkan nantinya belum bisa dikatakan akurat.

**1.1 Matakuliah Metode Penelitian**

Berikut ini merupakan data tugas kuliah pada matakuliah metode peneltian:

D1 = Aplikasi Rental Play Station Batara Indo Game

D2 = Pengontrolan Lampu Teras Dengan Sensor Intensitas Cahaya Berbasis Mikrokontroller

D3 = Penerapan Metode Certainty Factor dalam Mendiagnosa Kanker Nasofaring

D4 = Sistem Informasi Percetakan

Berdasarkan Tabel 7, skor tertinggi adalah pasangan D3 dan D3 dengan nilai 310. Hal ini membuktikan bahwa untuk pasangan D3 dan D3 pasangan karakter yang sama lebih banyak dari pada pasangan karakter yang tidak sama. Sedangkan skor terendah adalah D1 dan D2 / D2 dan D1 dengan skor -

139, hal ini membuktikan bahwa pasangan karakter yang tidak sama lebih banyak dari pada pasangan karakter yang sama. Dari keseluruhan hasil skor di atas, dapat dilihat bahwa yang paling mirip adalah pasangan D3 dan D3 sedangkan pasangan tugas yang paling tidak mirip adalah D2 dan D1/D1 dan D2.

**Tabel 7** Persentase KEMIRIPAN TUGAS MAHASISWA MATAKULIAH METODE PENELITIAN

| No | Tugas Kuliah Mahasiswa |         | Skor | Persentase (%) |
|----|------------------------|---------|------|----------------|
|    | Tugas 1                | Tugas 2 |      |                |
| 1  | D1                     | D1      | 195  | 100            |
| 2  | D1                     | D2      | -139 | 13             |
| 3  | D1                     | D3      | -96  | 10.11          |
| 4  | D1                     | D4      | -68  | 9.13           |
| 5  | D2                     | D1      | -139 | 10.61          |
| 6  | D2                     | D2      | 365  | 100            |
| 7  | D2                     | D3      | -124 | 21.76          |
| 8  | D2                     | D4      | -47  | 16.85          |
| 9  | D3                     | D1      | -96  | 8.68           |
| 10 | D3                     | D2      | -124 | 22.98          |
| 11 | D3                     | D3      | 310  | 100            |
| 12 | D3                     | D4      | -91  | 3.61           |
| 13 | D4                     | D1      | -68  | 10.56          |
| 14 | D4                     | D2      | -47  | 23.42          |
| 15 | D4                     | D3      | -91  | 4.81           |
| 16 | D4                     | D4      | 125  | 100            |

**1.2 Matakuliah Teknopreneuership**

Berikut ini merupakan data tugas kuliah pada matakuliah metode peneltian:

D1 = Aplikasi Lowongan Pekerjaan Berbasis WEB

D2 = Sistem Monitoring Ruangan

D3 = Absen Elektronik

D4 = Sistem Resep Elektronik

D5 = Sistem Informasi Percetakan

**Tabel 8** persentase KEMIRIPAN TUGAS MAHASISWA  
MATAKULIAH TECHNOPRENUERSHIP

| No | Tugas Kuliah Mahasiswa |         | Skor | Persentase (%) |
|----|------------------------|---------|------|----------------|
|    | Tugas 1                | Tugas 2 |      |                |
| 1  | D1                     | D1      | 180  | 100            |
| 2  | D1                     | D2      | -78  | 14.86          |
| 3  | D1                     | D3      | -63  | 0              |
| 4  | D1                     | D4      | -39  | 15.48          |
| 5  | D1                     | D5      | -58  | 19.11          |
| 6  | D2                     | D1      | -78  | 17.21          |
| 7  | D2                     | D2      | 115  | 100            |
| 8  | D2                     | D3      | -45  | 0              |
| 9  | D2                     | D4      | -36  | 40.37          |
| 10 | D2                     | D5      | -4   | 63.48          |
| 11 | D3                     | D1      | -63  | 0              |
| 12 | D3                     | D2      | -45  | 0              |
| 13 | D3                     | D3      | 75   | 100            |
| 14 | D3                     | D4      | 51   | 100            |
| 15 | D3                     | D5      | -57  | 0              |
| 16 | D4                     | D1      | -39  | 18.45          |
| 17 | D4                     | D2      | -36  | 41.61          |
| 18 | D4                     | D3      | 51   | 97.14          |
| 19 | D4                     | D4      | 105  | 100            |
| 20 | D4                     | D5      | -28  | 47.33          |
| 21 | D5                     | D1      | -58  | 21.56          |
| 22 | D5                     | D2      | -4   | 61.74          |
| 23 | D5                     | D3      | -57  | 0              |
| 24 | D5                     | D4      | -28  | 44.67          |
| 25 | D5                     | D5      | 125  | 100            |

Berdasarkan tabel VIII, skor tertinggi adalah pasangan D1 dan D1 dengan nilai 180. Hal ini membuktikan bahwa untuk pasangan D1 dan D1 pasangan karakter yang sama lebih banyak dari pada pasangan karakter yang tidak sama. Sedangkan skor terendah adalah D1 dan D2 / D2 dan D1 dengan skor -78, hal ini membuktikan bahwa pasangan karakter yang tidak sama lebih banyak dari pada pasangan karakter yang sama. Dari keseluruhan hasil skor di atas, dapat dilihat bahwa yang paling mirip adalah pasangan D1 dan D1 sedangkan pasangan tugas yang paling tidak mirip adalah D2 dan D1/D1 dan D2.

## KESIMPULAN

Berdasarkan penelitian yang dilakukan oleh penulis dalam membuat sistem pendeteksi

kemiripan tugas kuliah mahasiswa menggunakan algoritma smitha waterman maka penulis dapat menyimpulkan bahwa Algoritma Smith Waterman dapat digunakan dalam mengetahui persentase kemiripan tugas kuliah mahasiswa. Dimana setiap karakter akan dipasangkan kemudian dibandingkan. Nilai yang dihasilkan dalam sistem pendeteksi kemiripan tugas kuliah mahasiswa merupakan skor kemiripan tugas, semakin banyak karakter yang sama, maka semakin tinggi nilai skor yang dihasilkan (skor akhir bernilai positif). Sebaliknya, semakin banyak karakter yang tidak sama (mismatch), maka semakin rendah nilai skor yang dihasilkan (Skor akhir bernilai negatif).

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Purnawansyah, M.Kom, Huzain Azis, S.Kom., M.Cs., Dedy Atmajaya, S.Kom., M.Eng atas segala ilmu serta saran-saran yang sangat membantu penulis hingga penulis mampu menyelesaikan penelitian ini tepat pada waktunya.

## DAFTAR PUSTAKA

- Afif Bambang Prasetia. 2017. "Penerapan Algoritma Program Dinamis Pada Penyejajaran Sekuens Dengan Algoritma Smith – Waterman."Karya Tulis Ilmiah: Institute Teknologi Bandung.
- Djafar, Farid Bangkit, Agus Lahinta, dan Lillyan Hadjaratie. 2014. "Penerapan Algoritma Smith-Waterman Dalam Sistem Pendeteksi Kesamaan Dokumen."Tugas Akhir: Universitas Negeri Gorontalo.
- Darmawan, Helmy, Z Abdurahman Baizal, dan Rimba Widhiana Ciptasari. 2009. "Analisis Dan Implementasi Algoritma Smith-Waterman Pada Proses Identifikasi Kesamaan Dokumen." : 0-4. Tugas Akhir: Telkom University Bandung.
- Thalib Farid, dan Ratih Kusumawati. 2014. "Pembuatan Program Aplikasi Untuk Pendeteksian Kemiripan Dokumen Teks Dengan Algoritma Smith-Waterman."Universitas Gunadarma Depok, JawaBarat.
- Yuma Munekawa, Fumihiko Ino, Member, IEEE, dan Kenichi Hagihara. 2014. "Design and Implementation of the Smith-Waterman Algorithm on the CUDA-Compatible GPU."ResearchGate