BAB II
DASAR TEORI

2.1 Algoritma

Kata Algoritma berasal dari nama matematikawan persia abad ke-9 Abu Abdallah Muḥammad ibnu Musa al-Khwarizmi. *Algorism* adalah kata pada awalnya, *Algorism* adalah istilah yang merujuk kepada aturan-aturan aritmetis untuk menyelesaikan persoalan dengan menggunakan bilangan *numeric*. Melalui terjemahan Al-Khwarizmi *"Algorithmi de numero Indorum"*, *algorism* menjadi *algorithm* pada abad ke-18 yang mencakup semua prosedur atau urutan langkah yang jelas dan diperlukan untuk menyelesaikan suatu permasalahan.

Algoritma dapat di artikan sebagai urutan langkah-langkah untuk menyelesaikan masalah yang tersusun secara sistematis. Algoritma sangat penting untuk proses data komputer. Banyak program komputer berisi algoritma yang detail dan instruksi yang spesifik untuk melaksanakan tugas tertentu , seperti menghitung gaji karyawan.

2.2 Algoritma Smith-Waterman

Algoritma Smith-Waterman pertama kali diusulkan oleh Temple F. Smith dan Michael S. Waterman pada tahun 1981. Menurut situs Wikipedia berbahasa Indonesia algoritma Smith-Waterman adalah algoritma klasik yang telah dikenal luas dalam bidang bioinformatika sebagai metode yang dapat mengidentifikasi *local similarities* yaitu proses penyusunan dua *local squences nucleotide* atau *protein sequences* sehingga kemiripan antara dua *sequence* tersebut akan terlihat. Berdasarkan fungsinya tersebut, maka algoritma ini dapat dikonversikan ke dalam pemrogaman komputer untuk membantu proses pendeteksian dokumen teks dengan melihat kesamaan urutan kata pada beberapa dokumen teks.

Algoritma Smith-Waterman yang dipakai untuk penghitungan *local alignment* berdasarkan thesis dari Erik Vermij *"Genetic sequence alignment on a supercomputing Platform"*,yaitu :

1. Menambahkan nilai pada setiap perbandingan
   - Apabila memiliki kemiripan akan menambahkan nilai.
   - Apabila tidak memiliki kemiripan akan mengurangkan nilai.
2. Nilai 0 (nol) akan diberikan sebagai nilai awal.
3. Apabila terdapat nilai kurang dari 0 (nol) pada matriks, maka akan dianggap bernilai 0 (nol).

Tabel 2.1 Contoh 2 kalimat sebelum dibandingkan.

		oleh	karenanya	banyak	negara	di	dunia	yang	menjadi
	0	0	0	0	0	0	0	0	0
oleh	0								
karenanya	0								
banyak	0								
negara	0								
di	0								
dunia	0								
yang	0								

Tabel 2.1 Setelah pembobotan oleh Algoritma Smith-Waterman. jika cocok +1, jika tidak -1.

		oleh	karenanya	banyak	negara	di	dunia	yang	menjadi
	0	0	0	0	0	0	0	0	0
oleh	0	1	0	0	0	0	0	0	0
karenanya	0	0	2	1	0	0	0	0	0
banyak	0	0	1	3	2	1	0	0	0
negara	0	0	0	2	4	3	2	1	0
di	0	0	0	1	3	5	4	3	2
dunia	0	0	0	0	2	4	6	5	4
yang	0	0	0	0	1	3	5	7	6

Dapat terlihat pada 2 tabel diatas (tabel 2.1 dan 2.2) dimana tabel 2.1 adalah contoh 2 kalimat sebelum dibandingkan dan setiap kata pada kalimat telah diberi nilai dan pada tabel 2.2 adalah hasil dari 2 kalimat yang telah dibandingkan. Dapat terlihat nilai pada tabel 2.2 dengan kolom berwarna adalah nilai dengan urutan kata terbanyak dengan kata yang sama, sehingga nilai 7 pada tabel 2.2 adalah bobot yang diperoleh dari perbandingan 2 kalimat tersebut.

Menurut Robert W. Irving skema pemrogaman algoritma Smith-Waterman, yaitu:

Didefinisikan S_{ij} menjadi nilai maksimum yang didapatkan dari proses perbandingan sebuah substring A pada posisi ke-i dengan sebuah substring B pada posisi ke-j. hubungan sekurens standar untuk S_{ij}, yaitu:

- Jika A(i) = B(j) maka $S_{ij} = S_{i-1j-1}+h$, atau
- Jika A(i) ≠ B(j) maka $S_{ij} = \max(0,S_{i-1j},S_{ij-1},S_{i-1j-1})-1$
- Dimana kondisi awal adalah $S_{i,0}=S_{0,j}=0$ untuk semua i,j.

2.3 Dokumen

Dalam kamus bahasa Indonesia dokumen adalah sesuatu yang tertulis atau tercetak yang dapat dipergunakan sebagai bukti atau keterangan. Biasanya dokumen ditulis menggunakan media kertas dan informasinya menggunakan tinta. Menurut Suzzane Briet yang dikenal juga sebagai *"madame documentation"* konsep dokumen telah di definisikan sebagai *"any concrete or symbolic indication, preserved or recorded, for reconstructing or for proving a phenomenon, whether physical or mental"*. Dengan berkembangnya zaman dokumen dapat dibuat secara elektronik dan dapat disimpan dengan aman tanpa harus takut hilang atau rusak.

Dalam komputasi DOC / doc adalah ekstensi untuk dokumen pengolahan kata yang digunakan oleh Microsoft untuk format Microsoft word yang mereka miliki. Secara historis, ekstensi ini digunakan untuk dokumentasi dalam teks biasa, terutama dari program atau perangkat keras komputer pada berbagai sistem operasi. Selama

tahun 1980, WordPerfect menggunakan DOC sebagai hak milik ekstensi dari format mereka. Kemudian, pada tahun 1990, Microsoft memilih untuk menggunakan ekstensi DOC untuk format Microsoft Word.

2.4 Java

Java adalah bahasa pemrograman yang dapat dijalankan di berbagai komputer termasuk telepon genggam. Bahasa ini awalnya dibuat oleh James Gosling saat masih bergabung di Sun Microsystems saat ini merupakan bagian dari Oracle dan dirilis tahun 1995. Bahasa ini banyak mengadopsi sintaksis yang terdapat pada C dan C++. Ditambah sekarang beberapa perusahaan lebih membutuhkan programmer Java dibanding bahasa pemrograman yang lain. Karena fungsionalitasnya yang memungkinkan aplikasi Java mampu berjalan di beberapa *platform* sistem operasi yang berbeda, java dikenal pula dengan slogannya "Tulis sekali, jalankan di mana pun".

Kelebihan dari bahasa pemrogaman java, yaitu:

1. *Multiplatform*. Kelebihan utama dari Java ialah dapat dijalankan di beberapa *platform* / sistem operasi komputer. Dengan kelebihan ini pemrogram cukup menulis sebuah program Java dan dikompilasi lalu hasilnya dapat dijalankan pada beberapa platform tanpa perubahan.
2. Merupakan bahasa pemrogaman berorientasi objek
3. Perpustakaan Kelas Yang Lengkap, Java terkenal dengan kelengkapan *library* / perpustakaan yang sangat memudahkan dalam penggunaan oleh para pemrogram untuk membangun aplikasinya.

2.5 XAMPP

XAMPP adalah perangkat lunak yang mendukung banyak sistem operasi, merupakan kompilasi dari beberapa program. Fungsinya adalah sebagai *server* yang berdiri sendiri (*localhost*), yang terdiri atas program Apache HTTP Server, MySQL database, PHP dan Perl.

Asal kata XAMPP adalah singkatan yang masing-masing hurufnya adalah :

1. X. Program ini dapat dijalankan dibanyak sistem operasi seperti Windows, Linux, Mac OS, dan Solaris.
2. A (Apache). Merupakan aplikasi web server. Tugas utama Apache adalah menghasilkan halaman *web* yang benar berdasarkan kode PHP yang dituliskan oleh pembuat halaman *web*. jika diperlukan dapat suatu *database* diakses untuk mendukung halaman *web* yang dihasilkan.
3. M (MySQL). MySQL merupakan aplikasi *database server*. Perkembangannya disebut SQL yang merupakan kepanjangan dari *Structured Query Language*. MySQL dapat digunakan untuk membuat dan mengelola *database* beserta isinya. Kita dapat memanfaatkan MySQL untuk menambahkan, mengubah, dan menghapus data yang berada dalam database.
4. P (PHP). PHP adalah bahasa pemrograman web. Bahasa pemrograman PHP merupakan bahasa pemrograman untuk membuat web yang bersifat *server-side scripting*. PHP memungkinkan kita untuk membuat halaman web yang bersifat dinamis.
5. P (Perl). Perl adalah bahasa pemrograman yang dikembangkan pertama kali oleh Larry Wall.