

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penggunaan internet mengalami peningkatan yang sangat pesat, baik secara *mobile* ataupun tidak. Berdasarkan [3] menyatakan bahwa pengguna web secara *mobile* saja pada bulan Agustus 2010 mengalami peningkatan yang cukup pesat, halaman yang diakses meningkat lebih dari 14% dan transfer data bertambah hampir 10% tiap bulannya. Hal ini membuat para pengembang web berlomba-lomba untuk meningkatkan performa dan kualitas web masing-masing sehingga dapat memicu para pengembang untuk menggali *knowledge* yang ada dari aktivitas *user* dalam mengakses website, oleh karena itu munculah disiplin ilmu Web mining. Web mining adalah ilmu yang mempelajari penggalian data yang berhubungan dengan World Wide Web (WWW) [5].

Web mining dapat dikategorikan kedalam tiga kategori berdasarkan bagian mana dari sebuah website yang akan lakukan proses *mining*[11], Web Content Mining, Web Structured Mining, dan Web Usage Mining. Web Content Mining dan Web Structured Mining pada proses *mining*nya mencari informasi dari data yang tersedia secara *on-line*, sedangkan untuk Web Usage Mining data yang akan diolah dapat diperoleh secara *off-line* yaitu data log akses user yang tersimpan dalam *web log server*[4]. Pengolahan data secara *off-line* lebih mudah dikarenakan dalam pelaksanaan proses *mining* tidak mewajibkan tersediannya koneksi internet.

Web Usage Mining (WUM) merupakan proses untuk mengaplikasikan teknik web mining dalam melakukan analisa terhadap pola penggunaan (*usage pattern*) dan analisa aktivitas pengunjung web yang terekam dalam server log dari website. WUM memiliki tiga tahapan pengerjaan yaitu *preprocessing data*, *pattern discovery* (pencarian pola), dan *pattern analysis* (analisis pola yang terbentuk). Walaupun WUM dalam proses *mining* tidak dibutuhkan koneksi internet tetapi data yang tersedia tidaklah sedikit, dan proses pencarian pola terhadap data yang sangat banyak tersebut membutuhkan sebuah teknik pembahasan yang khusus. Salah satu teknik yang dapat digunakan adalah Frequent Sequential Pattern (FSP) mining, yaitu teknik pencarian pola sekuensial terhadap kecenderungan akses *user*. Menemukan Frequent Sequential Pattern (FSP) merupakan masalah yang penting dalam WUM [2]. Oleh karena itu digunakan sebuah metode pencarian pola sekuensial yaitu Adaptive Web Access Pattern Tree (AWAP Tree). AWAP Tree merupakan sebuah cara pencarian pola kecenderungan akses *user* dalam bentuk pohon sekuensial, dimana setiap transaksi akses user yang tercatat dalam web log akan dimasukkan kedalam pohon sekuensial. Metode ini juga menggunakan teknik *Huffman code* sehingga metode ini akan lebih baik dari metode sebelumnya dalam penggunaan memory.

1.2 Rumusan Masalah

Permasalahan yang menjadi objek dari penelitian tugas akhir ini, terdiri atas:

1. Bagaimana melakukan analisis terhadap pengaruh parameter minimum support dan jumlah transaksi serta pengaruh waktu terhadap performansi sistem.

2. Bagaimana menerapkan teknik *huffman code* dalam pengkodean *binary code position* untuk menemukan hubungan keterkaitan antar *item*.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah :

- a) Data server log yang digunakan berupa data aktivitas *user* dalam mengakses halaman-halaman website <http://www.itelkom.ac.id/id/>
- b) Web yang menjadi studi kasus dalam Tugas Akhir ini merupakan aplikasi aktif yang sudah dibangun dan berjalan pada jaringan internal maupun eksternal.
- c) Format data web server log yang digunakan adalah NCSA combined yang mengandung IP address, waktu, request, status, size, referrer, user_agent

1.4 Tujuan

Tujuan yang ingin dicapai dalam tugas akhir ini, yaitu :

1. Melihat pengaruh nilai Minimum Support dan jumlah transaksi serta pengaruh waktu terhadap performansi sistem
2. Mengetahui keterkaitan antar item dengan melihat *binary code item-item*

1.5 Metodologi Penyelesaian Masalah

Metode yang digunakan untuk menyelesaikan Tugas Akhir ini, yaitu :

1. Studi Literatur, yaitu dengan mempelajari literatur-literatur yang ada sesuai dengan permasalahan
2. Analisis masalah, dilakukan dengan tahapan analisis sebagai berikut :
 - a. Melakukan data selection dari server log, dengan mengambil data aktivitas akses user dari server aplikasi web
 - b. Melakukan preprocessing terhadap data mentah berupa server log, dengan melakukan tahapan: *parsing data* yaitu memisahkan data log menjadi bagian-bagian tertentu yang lebih kecil, *cleaning data* yaitu membuang data yang tidak diperlukan saat proses pencarian pola (sehingga menghasilkan atribut yang akan dianalisis seperti User IP dan halaman yang diakses) dan *data transformation* yaitu mengubah data menjadi data bertipe *discrete sequence* sebagai inputan pada tahap selanjutnya.
 - c. Melakukan pencarian pola terhadap data , dalam tahap ini akan dilakukan pencarian pola dengan algoritma AWAPT.
3. Perancangan perangkat lunak, yaitu dengan membuat pemodelan yang berorientasi objek yang dapat mengimplementasikan algoritma AWAPT .
4. Implementasi, membangun perangkat lunak berdasarkan perancangan perangkat lunak yang telah dilakukan sebelumnya dengan menggunakan bahasa pemrograman *c#*.
5. Pengujian dan analisis hasil pengujian perangkat lunak, dilakukan dengan cara menjalankan perangkat lunak dengan input berupa *raw log* yang telah dilakukan *preprocessing* sehingga menghasilkan output berupa pola

sekuensial pada pohon akses user, yang masing-masing nodenya telah memiliki kode binary, sesuai dengan parameter yang ada seperti parameter Minimum Support.

6. Penyusunan Laporan Tugas Akhir dan Kesimpulan Akhir.