

WEB USAGE MINING MENGGUNAKAN ALGORITMA ADAPTIVE WEB ACCESS PATTERN TREE (AWAPT)

Fithratul Aini¹, Sri Widowati², Eko Darwiyanto .³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Saat ini perkembangan website dalam World Wide Web semakin pesat dan data yang berhubungan dengan itu semakin banyak, para pengembang berusaha memberikan yang terbaik untuk mendapat performansi yang lebih baik dengan mengolah data yang didapatkan, sehingga muncullah ilmu Web mining. Salah satu kategori dalam Web Mining adalah Web Usage Mining (WUM), WUM merupakan proses untuk mengaplikasikan teknik web mining dalam melakukan analisa terhadap pola penggunaan dan analisa aktivitas pengunjung web yang terekam dalam server log dari website. WUM memiliki tiga tahapan pengerjaan yaitu preprocessing data, pattern discovery, dan pattern analysis

Dalam Tugas Akhir ini akan diimplementasikan salah satu metode Frequent Sequential Pattern mining untuk melakukan pencarian pola sekuensial terhadap kecenderungan akses user dalam tahap pattern discovery, yaitu Adaptive Web Access Pattern Tree (AWAPT). Metode ini juga menggunakan teknik Huffman code sehingga metode ini akan lebih baik dari metode sebelumnya dalam penggunaan memory

Kata Kunci : Web Usage Mining, Frequent sequential Pattern mining, AWAPT

Abstract

Now days, developing of websites and these data in World Wide Web grow rapidly. Developers of the site try to give the best effort to get better performance by mining all data that they get, so that there is Web Mining issue. One of Web mining's categories is Web Usage Mining (WUM). WUM is process to analyze pattern usage and user's activities recorded in server log of the website. WUM has three phases, they are preprocessing data, pattern discovery, dan pattern analysis.

In this final task, one of Frequent Sequential Pattern mining's method will be implemented to find sequential pattern of user access in pattern discovery phase, it is Adaptive Web Access Pattern Tree (AWAPT). This method uses Huffman code technic so that will be better in memory usage

Keywords : Web Usage Mining, Frequent sequential Pattern mining, AWAPT

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.



BAB V

Penutup

5.1 Kesimpulan

Berdasarkan hasil pengujian dan analisis terkait dengan penelitian Tugas Akhir yang telah dilakukan, diperoleh beberapa kesimpulan antara lain :

1. AWAP Tree dibangun dari sekuens halaman yang paling sering diakses dengan membentuk general tree, yang mempunyai *binary code*, lalu dikonversi ke binary tree sesuai aturan *Huffman coding tree*
2. Nilai parameter nilai *minimum support* yang optimal adalah $0 < \text{minimum support} < 0,2$. Pengaruh nilai *minimum support* pada performansi sistem dilihat dari pola yang dihasilkan adalah semakin besar nilai *minimum support* maka semakin sedikit pola yang dihasilkan. Begitu juga sebaliknya.
3. Parameter jumlah transaksi yang semakin kecil akan semakin cepat waktu eksekusinya. Tetapi terdapat faktor lain yang mempengaruhi perbandingan jumlah transaksi dataset terhadap waktu, yakni jumlah 1-sequence frequent item yang dihasilkan. Untuk besarnya jumlah transaksi yang sama dari dataset yang berbeda maka jika jumlah 1-sequence frequent item semakin besar maka waktu eksekusi akan semakin besar, begitu juga sebaliknya.
4. Pengaruh waktu penggunaan data, dalam hal ini dipartisi perhari selama seminggu pola yang dihasilkan tiap harinya berbeda-beda, tetapi pada hari kerja Senin – Jumat pola transaksi pengaksesan lebih banyak dibandingkan non hari kerja (*weekend*).
5. Akurasi pola pengaksesan pada hari kerja Senin – Jumat lebih besar dibandingkan non hari kerja Sabtu- Minggu (*weekend*).
6. Hubungan keterkaitan antar item dapat diketahui dengan melihat *Huffman codenya*.

5.2 Saran

Diharapkan untuk penelitian selanjutnya dapat melanjutkan tahapan *pattern analysis* untuk menghasilkan sistem yang lebih baik, seperti menghasilkan rekomendasi link pada aplikasi web recommender system.

REFERENSI

- [1] Gomathi. C,M. Moorthi, K.Duraiswamy, 2008, *Binary Coded Web Access Pattern Tree in Education Domain*. www.ccsenet.org/journaal.html Vol.1,No.4 , November 2008
- [2] Gomathi. C,M. Moorthi, K.Duraiswamy, 2008,*Web Access Pattern Algorithms in Education Domain*, <http://ccsenet.org/journal/index.php/cis/article/view/1192/1154.pdf>
- [3] <http://swa.co.id/2010/10/penggunaan-mobile-web-terus-meningkat-setiap-bulan/> diakses pada tanggal 28 September 2010.
- [4] Iváncsy Renáta, István Vajk, 2006, *Frequent Pattern Mining in Web Log Data*. Acta Polytechnica Hungarica Vol.3, No.1, 2006.
- [5] Margaret H. Dunham, 2003, *Data Mining Introductory and Advanced Topics*, New Jersey , Prentice Hall.
- [6] Mobasher Bamshad, 2008, *Chapter 12: Web Usage Mining*.
- [7] Ningtyas. Dian Kusuma, Farah Virnawati, Prasetyo, Tirta Paramitta,I Wayan Simri, 2008, Analisis Perilaku pengguna sistem e-learning universitas gunadarma, Proceeding, Seminar Ilmiah Nasional Komputer dan Sistem Intelijen (KOMMIT 2008).
- [8] Paramita. Ayudya ,2009, Analisis pola kunjungan pengguna situs web IPB menggunakan algoritma Totally Fuzzy, Bogor, IPB
- [9] Sathiyamoorthi. V., V. Murali Bhaskaran, 2009, *Data Preparation Techniques for Web Usage Mining in World Wide Web-An Approach*, International Journal of Recent Trends in Engineering, Vol 2, No. 4, November 2009 , Academy Publisher
- [10] Vasumathi. D, A. Govardhan, 2009, *BC-WASPT Web Access Sequential Pattern Tree Mining*, IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.6, June 2009
- [11] Vijayalakshmi. S, V.Mohan and S.Suresh Raja, 2010, *Mining of Users' access behaviour for frequent sequential pattern from web logs*, International Journal of Database Management Systems (IJDMS) Vol.2, No.3, August 2010
- [12] Zhou. Baoyao, Siu Cheung Hui, Kuiyu Chang, *An Intelligent Recommender System using Sequential Web Access Patterns*, Singapore.