

## ANALISIS DAN IMPLEMENTASI WEB USAGE MINING MENGGUNAKAN ALGORITMA GRAPH PARTITIONING (STUDI KASUS : TUNEECA ONLINE STORE)

Sigit Bimo Wiweko<sup>1</sup>, Eko Darwiyanto<sup>2</sup>, Veronikha Effendy<sup>3</sup>

<sup>1</sup>Sistem Komputer, Fakultas Teknik Elektro, Universitas Telkom

---

### Abstrak

Peningkatan aktivitas kunjungan terhadap website menghasilkan data yang cukup banyak mengenai user dan interaksinya dengan website yang disimpan dalam web server log. Informasi yang bisa diperoleh salah satunya adalah pola navigasi user. Pola navigasi user menggambarkan aktivitas apa saja yang dilakukan user selama mengakses suatu website. Memahami pola navigasi user dalam mengakses suatu website dapat berguna untuk memahami tingkah laku user dalam mengakses websitetersebut. Sehingga dapat digunakan sebagai acuan dalam perbaikan kualitas website dan menjamin kepuasan user ketika menggunakan website tersebut. Pada ranah e-commerce, pola navigasi user dapat digunakan sebagai acuan untuk menentukan strategi bisnis berdasarkan tingkah laku user yang diperoleh. Dalam tugas akhir ini, web server logdari tuneeca online storeakan diproses dengan mengimplementasikan salah satu metode dalamweb usage mining yaitu clustering. Web usage mining merupakan salah satu pengaplikasian teknik data mining yang dapat digunakan untuk menemukan pola navigasi user. Data log tersebut akan melalui tahap preprocessing, kemudian dilakukan clustering terhadap page dengan menggunakan algoritma graph partitioning. Hasil penelitian menunjukkan bahwa penentuan parameter nilai minimum bobot mempengaruhi jumlah klaster yang dihasilkan serta nilai visit coherence yang diperoleh. Performansi dari algoritma graph partitioning cukup baik dalam membentuk klaster pola navigasi berdasarkan tingginya nilai modularization qualityyang diperoleh. Pola navigasi user yang dihasilkan dapat digunakan sebagai acuan untuk rekomendasi pengembangan web dari tuneeca online store.

**Kata Kunci :** web usage mining, pola navigasi user, web server log, graph partitioning, visit coherence, modularization quality

---

### Abstract

Increased activity of a visit to the website generates huge enough data about users and their interaction with a website that is stored in the web server logs. One of the information that can be obtained is user navigation patterns. User navigation patterns generated, could give an overview about what users actually do and need when access the website. Understanding the user navigation patterns can be useful for understanding user behavior in accessing the website. So it can be used as a reference in improving the quality of the website and ensure user satisfaction when using the website. In the domain of e-commerce, user navigation patterns can be used as a reference for determining a business strategy based on user behavior is obtained. In this final project, the web server logs of tuneeca online store will be processed by implementing clustering, one of web usage mining methods. Web usage mining is one of the application of data mining techniques that can be used to discover the user navigation patterns. The log data will be going through the preprocessing stage, then performed clustering to the pages by using graph partitioning algorithm. The result shows that determining the minimum weight value affects the number of clusters produced and the visit coherence value obtained. Performance of graph partitioning algorithm is quite good in forming clusters of navigation patterns based on high value of modularization quality obtained. User navigation patterns generated can be used as a reference for the recommendation of web development Tuneeca online store.

**Keywords :** web usage mining, user navigation patterns, web server log, graph partitioning, visit coherence, modularization quality

---

# 1. Pendahuluan

## 1.1. Latar Belakang

Perkembangan teknologi internet yang semakin pesat dapat dilihat dari semakin mudahnya dalam mendapatkan akses internet. Perkembangan internet tersebut juga diiringi dengan semakin banyaknya pemanfaatan website dalam berbagai bidang. Hal tersebut tentunya memudahkan dalam hal mencari informasi. Seiring dengan perkembangan pemanfaatan website ini, aktivitas *user* dalam penggunaan berbagai website semakin meningkat.

Peningkatan aktivitas kunjungan terhadap suatu website menghasilkan data yang cukup banyak mengenai *user* dan interaksinya dengan website yang disimpan dalam *web server log*. Informasi tersebut tentunya sangat bermanfaat bagi perkembangan website bila diolah secara tepat. Informasi yang bisa diperoleh salah satunya adalah pola navigasi *user*.

Pola navigasi *user* merupakan kemiripan dalam tingkah laku pengaksesan suatu website oleh sekelompok *user*. Dikarenakan terdapat kemungkinan bahwa beberapa *user* memiliki ketertarikan yang sama dalam navigasinya pada suatu web, pola navigasi ini diharapkan dapat menangkap kesamaan informasi yang dibutuhkan oleh beberapa *user* tersebut. Sehingga bisa didapatkan pola navigasi yang bisa merepresentasikan tingkah laku *user* dalam mengakses suatu website. Memahami pola navigasi *user* dalam mengakses suatu website dapat berguna untuk melakukan perbaikan kualitas website sehingga dapat menjamin kepuasan *user* ketika menggunakan website tersebut.

*Web usage mining* merupakan salah satu pengaplikasian teknik *data mining* yang dapat digunakan untuk menemukan pola navigasi user tersebut. Salah satu teknik permodelan yang dapat digunakan adalah *clustering*. *Clustering* adalah pengelompokan objek-objek yang berkaitan atau memiliki kemiripan. Dalam *web usage mining*, *clustering* dapat dilakukan untuk mengelompokkan *user* berdasarkan kriteria tertentu atau untuk mengelompokkan halaman website yang memiliki keterkaitan berdasarkan pola yang didapat dari navigasi yang dilakukan oleh *user*.

Dalam tugas akhir ini, digunakan *web log server* dari salah satu website *e-commerce* yaitu *tuneecca.com*. Website *e-commerce* dipilih karena *user* dapat bebas mengakses konten yang ada pada web tersebut sehingga klaster yang dihasilkan

dapat lebih optimal. Web log tersebut akan melalui tahap *preprocessing*, kemudian dilakukan pembentukan kluster dengan menggunakan algoritma *graph partitioning*.

*Graph Partitioning* merupakan salah satu algoritma *clustering* yang dapat digunakan untuk mencari kesamaan pola navigasi suatu website. Algoritma ini akan memodelkan keterhubungan antar halaman website berdasarkan waktu pengaksesan dan frekuensi kemunculan halaman website dalam suatu sesi. Sehingga algoritma ini digunakan untuk mengolah data log dari *tuneece online store*.

*Visit coherence* digunakan untuk mengetahui akurasi kluster yang dihasilkan berdasarkan koherensinya dengan data yang digunakan untuk membentuk pola navigasi. Performansi dari algoritma *graph partitioning* dalam membentuk kluster pola navigasi akan dievaluasi dengan menggunakan *modularization quality*. Metrik ini merupakan metrik yang biasa digunakan pada proses evaluasi *graph clustering*. Proses evaluasi ini dilakukan dengan melihat kualitas kluster yang dihasilkan berdasarkan kuatnya relasi antar anggota dalam kluster.

## 1.2. Perumusan Masalah

Berdasarkan uraian latar belakang, maka permasalahan yang dapat dirumuskan pada penelitian ini adalah:

1. Bagaimana menerapkan algoritma *Graph Partitioning* dalam membentuk pola navigasi *user* ?
2. Bagaimana performansi dari algoritma *Graph Partitioning* serta pengaruh penentuan parameter bobot dalam membentuk pola navigasi *user* ?
3. Bagaimana rekomendasi yang dihasilkan untuk pengembangan kualitas website berdasarkan pola navigasi *user* yang terbentuk ?

## 1.3. Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dari penelitian ini adalah :

1. Menerapkan algoritma *Graph Partitioning* sebagai bagian dari *pattern discovery* pada *web usage mining* dalam melakukan kluster terhadap data log dari *tuneece online store*.

2. Menganalisis pengaruh penentuan parameter bobot algoritma *Graph Partitioning* dalam membentuk pola navigasi *user* serta pada nilai *visit coherence* yang diperoleh.
3. Menganalisis performansi dari algoritma *Graph Partitioning* dalam mengklaster halaman web dengan menggunakan *Modularization Quality*.
4. Menghasilkan rekomendasi untuk pengembangan kualitas website *tuneeca online store* berdasarkan pola navigasi *user* yang terbentuk.

#### 1.4. Batasan Masalah

Dalam tugas akhir ini, permasalahan dibatasi dalam beberapa hal yaitu:

1. Data yang digunakan adalah *web server log* dari situs *tuneeca.com*
2. Data yang digunakan adalah *access log* selama satu minggu dari tanggal 19 Agustus 2014 hingga 25 Agustus 2014.
3. Url yang diperoleh dari *web server log* diklasifikasikan menjadi beberapa kategori sesuai dengan menu dan submenu pada web *tuneeca.com*, sehingga hasil klasternya adalah klaster kategori tersebut yang merepresentasikan page.

#### 1.5. Metodologi Penyelesaian Masalah

Metode penelitian yang digunakan dalam penelitian ini terdiri atas delapan tahap, yaitu :

1. Identifikasi Masalah

Pada penelitian ini, permasalahan yang ingin diselesaikan adalah bagaimana memodelkan pola navigasi *user* dengan mengelompokkan halaman web pada web *tuneeca.com* yang saling berkorelasi menggunakan algoritma *graph partitioning*.

2. Studi Literatur

Studi literatur yang dilakukan bertujuan untuk proses pembuatan laporan tugas akhir dengan mengacu referensi-referensi yang berkaitan dengan *Web Usage Mining*, *Graph Partitioning*, beserta dengan *tools* yang akan digunakan dalam membangun sistem.

3. Analisis dan Perancangan Aplikasi

Analisis dalam menentukan kebutuhan sistem, batasan-batasan sistem, pengumpulan data, dan menentukan perangkat lunak yang akan digunakan selama proses perancangan sistem.

#### 4. Pengumpulan Data

Penelitian ini dibangun dengan menggunakan data berupa *web server log* dari website yang akan diamati.

#### 5. Implementasi

Sistem yang akan dibangun ini akan diimplementasikan dengan menggunakan bahasa pemrograman Java serta tools berupa Eclipse IDE, dan MySQL.

#### 6. Pengujian Sistem

Pengujian sistem dilakukan dengan melihat hasil keluaran sistem berupa kluster halaman web pada satu website yang merepresentasikan pola navigasi *user* serta mengamati performansi dari sistem.

#### 7. Penyusunan Laporan

Penyusunan laporan tugas akhir merupakan tahap terakhir dalam pembangunan sistem yang dilakukan untuk mendokumentasikan penelitian yang telah dilakukan supaya dapat bermanfaat bagi berbagai pihak.

### 1.6. Sistematika Penulisan

Tugas Akhir ini disusun dengan sistematika pembahasan sebagai berikut :

#### **Bab 1 Pendahuluan**

Bab ini memaparkan latar belakang dilakukannya penelitian, tujuan yang ingin dicapai dari penelitian ini, batasan masalah, perumusan masalah yang akan dibahas, metode penyelesaian masalah, dan sistematika penulisan.

#### **Bab 2 Tinjauan Pustaka**

Bab ini memuat penjelasan mengenai teori-teori yang mendukung dan melandasi penulisan tugas akhir, yaitu tentang metode algoritma *graph partitioning* serta teori mengenai *web usage mining*.

#### **Bab 3 Perancangan Sistem**

Bab ini berisi mengenai analisis kebutuhan perangkat lunak maupun perancangan sistem secara global dengan menggunakan diagram alir, diagram konteks, serta *data flow diagram*.

#### **Bab 4 Implementasi Dan Pengujian Sistem**

Bab ini berisi tentang implementasi dari sistem yang dibangun dengan menggunakan bahasa pemrograman Java. Bab ini juga berisikan mengenai cara kerja dan fungsionalitas dari sistem yang akan dibangun, serta cara dan hasil pengujian yang dilakukan dari sistem tersebut.

#### **Bab 5 Kesimpulan Dan Saran**

Bab ini berisikan tentang kesimpulan dari hasil penelitian tugas akhir ini dan saran-saran untuk pengembangan lebih lanjut.



## 5. Kesimpulan dan Saran

### 5.1. Kesimpulan

Kesimpulan yang didapatkan dalam pengerjaan tugas akhir ini adalah :

1. Algoritma *Graph Partitioning* dapat diimplementasikan dalam pembentukan pola navigasi *user* pada *tuneeca online store* yang direpresentasikan oleh klaster halaman web. Akurasi klaster yang dihasilkan cukup baik berdasarkan nilai *visit coherence*, yaitu mencapai 88,21% untuk nilai *visit coherence* saat menghasilkan jumlah klaster yang paling optimum.
2. Penentuan parameter nilai minimum bobot mempengaruhi jumlah klaster yang diperoleh karena nilai parameter tersebut menentukan proses partisi dari graf yang terbentuk.
3. Penentuan parameter nilai minimum bobot juga mempengaruhi nilai *visit coherence* secara signifikan. Hal tersebut dibuktikan pada nilai *visit coherence* yang terus turun seiring dengan peningkatan nilai minimum bobot. Penurunan tersebut dikarenakan semakin meningkatnya jumlah *outlier* atau *page* yang tidak masuk ke klaster manapun.
4. Algoritma *Graph Partitioning* memiliki performansi yang cukup baik dalam membentuk klaster. Hal tersebut dapat diketahui dari klaster yang terbentuk memiliki hubungan antar *node* yang cukup kuat dan hubungan antar klaster yang lemah, yaitu dengan rata-rata nilai *modularization quality* sebesar 0,65.

### 5.2. Saran

1. Melakukan pengembangan website *tuneeca online store* berdasarkan pola navigasi yang diperoleh.
2. Metode *web usage mining* lainnya seperti *association rule* dapat diterapkan pada *web server log e-commerce* untuk mencari keterhubungan antar navigasi *user*.
3. Algoritma *clustering* lainnya dapat digunakan untuk membandingkan dengan pola navigasi *user* yang dihasilkan.



## Daftar Pustaka

- [1] Baraglia, Ranieri dan Paolo Palmerini. 2002. *SUGGEST : a Web Usage Mining System*. International Conference on Information Technology: Coding and Computing.
- [2] Borgas, Jose dan Mark Levene. 2000. *Data Mining of User Navigation Pattern*. Department of Computer Science, University College London.
- [3] Chiricota, Y., F. Jourdan, dan G. Melançon. 2003. *Software Components Capture using Graph Clustering*. 11th IEEE International Workshop on Program Comprehension. Portland, Oregon: IEEE / ACM.
- [4] Cooley, R., B. Mobasher, dan J. Srivastava. 1999. *Data preparation for mining World Wide Web browsing patterns*. Journal of Knowledge and Information Systems.
- [5] Han, Jiawei. 2006. *Data Mining: Concepts and Techniques Second Edition*. San Francisco : Morgan Kaufmann Publisher.
- [6] Jalali, M. et al. 2008. *A new clustering approach based on graph partitioning for navigation patterns mining*. 9th International Conference on Pattern Recognition.
- [7] Larose, Daniel T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey : John Wiley & Sons, Inc
- [8] Markov, Zdravko. dan Daniel T. Larose. 2007. *Data Mining The Web : Uncovering Patterns in Web Content, Structure, and Usage*. New Jersey : John Wiley & Sons, Inc.
- [9] Mitchell, B. et al. 1999. *Bunch: A clustering tool for the recovery and maintenance of software system structures*. International Conference on Software Maintenance, ICSM.
- [10] Pamutha, Thanakorn. et al. 2012. *Data Preprocessing on Web Server Log Files for Mining Users Access Patterns*. International Journal of Research and Reviews in Wireless Communications (IJRRWC) Vol. 2, No. 2.
- [11] Perkowitz, M. dan O. Etzioni. 1999. *Adaptive web sites: Conceptual Cluster Mining*. International Joint Conference on Artificial Intelligence, hal 264–269.



- [12] Sedgewick, Robert dan Kevin Wayne. 2014. *Undirected Graphs*. <http://algs4.cs.princeton.edu/41undirected/>, (diakses tanggal 20 Maret 2014).
- [13] Srivastava, J. *et al.* 2000. *Web usage mining: discovery and applications of usage patterns from web data*. SIGKDD Explore.
- [14] Srivastava, J., Prasanna Desikan, dan Vipin Kumar. 2005. *Web Mining Concepts, Applications & Research Directions*. Foundations and Advances in Data Mining - Studies in Fuzziness and Soft Computing, hal 275-307.
- [15] Zaidi, Faraz., Daniel Archambault, dan Guy Melançon. 2010. *Evaluating the Quality of Clustering Algorithm using Cluster Path Lengths*. 10th Industrial Conference.

