

ABSTRAK

Clustering merupakan salah satu bagian penting dalam penggunaan web mining untuk segmentasi pengunjung. Tindakan ini sangat penting untuk modifikasi web personalisasi. Dalam tulisan ini, kami melakukan pengelompokan pengunjung web menggunakan kombinasi metode *clustering* hirarki dan non-hirarki terhadap data log web *i-gracias*. Metode pengelompokan hirarki digunakan untuk menentukan jumlah *cluster*, dan metode pengelompokan non-hirarki digunakan dalam membentuk kelompok. Tahapan analisis *cluster* didahului oleh pra-pengolahan data dan analisis faktor. Dengan pendekatan ini, pemilik web lebih efektif dalam menemukan pola akses pengunjung web dan dapat memiliki pengetahuan baru tentang segmentasi pengunjung. Dari pengujian yang diterapkan pada data log web *i-gracias*, beberapa kelompok pengunjung web dihasilkan. Di antara beberapa *cluster*, terdapat jumlah *cluster* memiliki jumlah anggota terbesar pada setiap skenarionya kita dapat mengetahui menu apa saja yang diakses.

Informasi ini dapat berguna bagi manajemen *web* untuk memperhatikan pola perilaku anggotanya *cluster* yang memiliki anggota terbanyak. Jumlah *cluster* yang dihasilkan adalah 2-5 *cluster*. Dapat diketahui dari *cluster* tersebut yang memiliki anggota *cluster* yang paling banyak. Terdapat beberapa hari dimana banyak menu yang diakses oleh *user*. Pada minggu 1, lebih cenderung pada menu yang diakses adalah tentang menu Registrasi, baik itu dalam tagihan pembayaran, keterlambatan registrasi dan proses registrasi. Pada minggu 2, *user* cenderung dominan pada menu silabus dan akademik mahasiswa baik kehadiran maupun jadwal mahasiswa. Sementara pada minggu 3 *user* tidak cenderung pada beberapa menu saja melainkan banyak menu yang di kunjungi, namun pada minggu 3 ini hal yang paling diperhatikan adalah pada menu tentang Tugas Akhir/Proyek Akhir. Pada minggu 4 yang paling diperhatikan adalah pada menu tentang akademik mahasiswa baik kehadiran, presensi maupun jadwal.

ABSTRACT

Clustering is one important part of web usage mining for visitor segmentation purposes. This action is very important for web personalization or web modification. In this final project, web visitors *clustering* with combination of hierarchical clustering methods and non-hierarchical against web log data *i-gracias* has been performed. Hierarchical clustering method is used to determine the number of clusters, and non-hierarchical clustering method used in forming the group. Cluster analysis stage is proceed by pre-processing the data and factor analysis. With this approach, the owner of the web is more effective in finding patterns of access and web visitors can have new knowledge about visitor segmentation. From the tests applied to the *i-gracias* web log data, several groups of web visitors generated. Among several clusters, there are number of clusters having the largest number of members in each scenario and the variable which accessed could be known. This information can be useful for web management to pay attention to patterns of behavior that cluster members have the most members, to create a personalized or modification on the web.

This information can be useful for web management to pay attention to patterns of behavior that cluster members have the most members. The number of generated clusters are clusters 2-5. It can be seen from the cluster that has the most members of the cluster. There are some days where a lot of the menu accessed by the user. In week 1, more likely on the menu that is accessed is on the menu Registration, be it in the bill payment, delay registration and registration process. In week 2, the user tends to be dominant on the menu and academic syllabus good student attendance and student schedules. While at 3 weeks of the user is not likely to be on some menus alone but many menu visit, but at this 3 week the most attention is on the menu of Final / Final Project. At week 4 the most attention is on the menu of a good student academic, presence and schedule.

\