Ant Colony Optimization method analyzing for
Sequential Pattern Mining (Case Study: iGracias
Telkom University)

1 Isma Dewi Liana, 2 Ibnu Asror and 3 Indra Lukmana Sardi
123 School of Computing, Telkom University, Indonesia
E-mail: 1ismadewiliana@student.telkomuniversity.ac.id,
2iasror@telkomuniversity.ac.id, 3indraluk@telkomuniversity.ac.id

Abstract. With the development of internet technology, information becomes very easy to obtain. Especially the use of web-based information systems are very widely used to disseminate information to the world. Similarly, in the field of education that makes the website as a source of information for all students, lecturers, to employees. In general, each user access to the website can form the access pattern. Then, the accessing pattern can be used to do the development on the website to facilitate the user get the desired menu. By using Ant Colony Optimization (ACO), the system can output the pattern of access based on the interests of all users. The selection of the Ant Colony Optimization (ACO) method due to the flow when the user accesses the website has in common with the ants while building the road to the food. In this experiment, pheromone will be modified with interest time and interest visit user to release the result that can approach the pattern of user access when accessing the website. Also, after testing the threshold was obtained 0.02 for the student group, 0.05 for the lecturer group, and 0.025 for the employee group which can produce a pattern that represents the user’s habit patterns.

1. Introduction
With the development of internet technology, information in and out happens to increase [1]. Moreover, the use of web-based information systems is very widely used, making it easier for someone to disseminate information throughout the world [2]. One field that applies the use of websites is in the field of education. Many universities use websites as a source of information for all students, lecturers, and employees. The used website can be developed to facilitate users when accessing so that it does not take long to go to frequently visited pages [3, 4]. In this research, the length of time a user and many user visits will be used to strengthen how user interest is on a page. Also, this study uses the Ant Colony Optimization method to get a pattern of user habits in accessing a website.

Ant Colony Optimization (ACO) is one of the parts of Swarm Intelligence, where the methods in Swarm Intelligence have a possible basis for exploring collective (or distributed) problem solving without centralised control or global model provision [5]. Problem-solving in Swarm Intelligence takes inspiration from the social behaviour of insects and other animals [6]. The ACO algorithm is inspired by the behaviour of a very simple ant colony when searching for food, and even ants can find the shortest path between the nest and the food found. In the process of
looking for food, ants will leave chemicals as traces called pheromones on the path that passed [1]. So that other ants can follow the path of most pheromones left by ants before.

When website users explore, this behaviour has in common with ant colonies when looking for food [7]. By using the concept of interest pheromone, the pattern of interest of users to access the website can be generated based on each group from the weblog to be more efficient and accurate. Not only are the patterns of interest that can produce, but the menus that are most in demand by users can also found in this research based on the selection of ants on the track.

ACO was initially applied to the problem of travelling salesman problems by Marco Dorigo in 1997 [8], which then developed very quickly. Now ACO can solve various discrete cases such as travel route construction, job scheduling, to data grouping [9].

This research aims to build a menu search system that is most favourite by users of each user group and patterns that most often happens with a factor time when the user accesses the menu and how many times the user visits the menu. Also, this research will compare the results between using the factors of visit time and many visits by not using these factors, as well as knowing the parameters that can improve system performance in order to get the best results.

In this research divided into four sessions. First introduction, this session explains the background of the research conducted. The second session of the proposed model, which in this session describes the model proposed in this research. The third session, Experimental results which tell the tests performed and the results obtained. The fourth session of conclusions is the final session of this research.

2. Proposed Model
In this section, time and frequency are used as factors of user interest on a page.

![Flowchart proposed model](image)

**Figure 1.** Flowchart proposed model.

2.1. Data
Data using the Igracias Telkom University event log, which has a time limit from September to October 2014. In this section, there are three groups used is students, lecturers, and employees.

2.2. Preprocessing
Before entering the process deeper, the data first goes through the preprocessing process because the dataset is still irregular. In the preprocessing stage, the data will be prepared according to the user session ID and then look for the timestamp on each page visited by the user. Then, the pages contained in the dataset will be represented by words that are much easier like WA1, WA2 to many pages that exist. Example:

2.3. Influence Factor on Interest Pheromone
There are two factors that will affect the results of the calculation, that is Time and Frequency:
Table 1. Preprocessing example.

| No | Page | Activity |
|----|------|----------|
| 1  | WA1  | Login    |
| 2  | WA2  | Dashboard|

2.3.1. Time The length of time a user accesses a page reflects how interested the user is on that page. The longer the user visits on a page, then it can be concluded that the user gets the information needed. Likewise, on the contrary, the less user time on a page, there is the possibility of the page only as a bridge from the previous page to the page that user wants to go. Can be calculated using:

\[
Time(i) = \frac{TotalTime(i)}{Average_{i \in tabu} TotalTime(i)} \tag{1}
\]

\(TotalTime(i)\) as the total amount of user time when accessing page \(i\), and \(Average\) as the average time when the user accesses the page \(i\).

2.3.2. Freq Frequency is the number of users visiting the page. Every user has different times to visit the page visited. User interest on a page can calculate the frequency. If the user is interested or needs to visit the page, the intensity of the page will be even higher. Can be calculated using:

\[
Freq(i) = \frac{visit(i)}{\sum_{i \in tabu} visit(i)} \tag{2}
\]

Where \(visit(i)\) shows many visits on page \(i\), \(tabu_k\) shows a collection of all pages that have been visited by ants.

2.3.3. Interest Interest is built using two factors: frequency and time according to the formula above. Where will develop interest on every page, by using:

\[
Interest(i) = a \cdot Time(i) + b \cdot Freq(i) \tag{3}
\]

Where \(Interest\) \((i)\) as user interest on page \(i\). For \(a\) and \(b\) as the weight of interest for Time and Freq, and the amount of \(a + b = 1\). In this research, using \(a = 0.5\) and \(b = 0.5\).

2.4. Ant Colony Optimization
Which will build a pattern that is ant colony optimization, which must perform three stages of the process

2.4.1. Selection Preferences Selection preferences are the frequency of a user when accessing the path node \(i\) to node \(j\), with calculations using the following formula Selection preferences are the frequency of a user when accessing the path node \(i\) to node \(j\), with calculations using the following formula where \(C_{(ij)}\) is a lot of user visits on every page visited:

\[
\eta_{ij} = \frac{C_{ij}}{\sum_{k=1}^{n} C_{ik}} \tag{4}
\]
2.4.2. Path Selection  The path selection function is a calculation when the user selects the path from node i to node j, calculated using:

\[
P_{ij}(t) = \left[ \frac{\tau_{ij}(t)}{\sum_{k=1}^{n} \eta_{ij}} \right]^\alpha \left[ \eta_{ij} \right]^\beta
\]  

(5)

In the Path Selection, there has a threshold that will cut the page that formed in the pattern. The threshold used in each user group will be different.

2.4.3. Interest Pheromone  The \( t_{ij}(t) \) function indicates user interest in the path between nodes i to node j. This formula updates the pheromone path selected by the Path selection, with the following calculations:

\[
\tau_{ij}(t + 1) = (1 - \rho)\tau_{ij}(t) + \Delta \tau_{ij}
\]  

(6)

\[
\Delta \tau_{ij} = Q * \text{Interest}(j) \]  

(7)

In calculation (10), \( \rho \) is the evaporation factor of pheromone attraction, \( 1 < \rho \) is a pheromone storage factor, \( 0 < \rho < 1 \). And \( Q \) is a constant value [1, 2].

The final output obtained in the form of browsing patterns and algorithms ends.

3. Experimental Result

In this section displays the results of the testing that has done with several scenarios as well as the testing analysis.

3.1. Testing by Comparing System Calculations and Manual Calculations

The manual calculation is done to determine the accuracy of the success of the system was build. The test conducted with a sample of 25 students of the data from the user group. The results of the manual calculation compared with the results of the system calculation. Following are the results of the calculations that have made:

Table 2. Comparison of manual calculation results with the calculation of sequential pattern mining systems.

| Initial | System Calculation Pattern | Manual Calculation Pattern |
|---------|---------------------------|---------------------------|
| WA7     | WA7>>WA2>>WA3            | WA7>>WA2>>WA3            |
| WA3     | WA3                       | WA3                       |
| WA1     | WA1>>WA2>>WA3            | WA1>>WA2>>WA3            |
| WA2     | WA2>>WA3                  | WA2>>WA3                  |

Manual calculation testing with system calculations can see in table 2. By using the same page initials and using the same threshold of 0.1, both calculations have the same pattern results.

3.2. Testing Using Several Thresholds

Determination of the threshold in path selection must pay attention to the final results. If setting the threshold is too small then the pattern will be formed too much, and the selected menus have a small value. Where the pages that appear in the pattern only become a bridge to get to the user’s destination page. Moreover, a threshold that is too large will not succeed to form a pattern. Threshold will be different for each group, as it will adjust the desired result. In the system built using five ants and iterations ten times.
In each group experiment with several thresholds from 0 to 1, to get the right threshold. The threshold is determined according to the result of the length of the page produced is not too long and not too short. Each user group has a different threshold. For student groups using a threshold of 0.02, for the 0.05 group of lecturers, and for the group of employees 0.025. Here are the results of the patterns of each group using the respective threshold.

**Table 3.** Sequential Pattern Mining results group of students with a threshold of 0.02.

| Initial   | Pattern                                                                 |
|-----------|-------------------------------------------------------------------------|
| login     | login >> dashboard >> Presence                                           |
|           | login >> dashboard >> Registration-Schedule-Student Schedule >>          |
|           | Registration-Schedule-Student Exam Schedule >> Survey-Home >> Survey-   |
|           | Questionnaire                                                           |
| dashboard | dashboard >> Presence                                                   |
|           | dashboard >> Registration-Schedule-Student Schedule >> Registration-     |
|           | Schedule-Student Exam Schedule >> Survey-Home >> Survey-Questionnaire    |

**Table 4.** Sequential Pattern Mining results group of lectures with a threshold of 0.05.

| Initial   | Pattern                                                                 |
|-----------|-------------------------------------------------------------------------|
| login     | login >> dashboard >> Presence-Minutes of Lecture List >> Presence-     |
|           | Submit Minutes of Lecture >> Presence-Print Attendance List >> Presence-|
|           | Presence                                                                |
| dashboard | dashboard >> Presence-Minutes of Lecture List >> Presence-Submit Minutes|
|           | of Lecture >> Presence-Print Attendance List >> Presence-Print Presence  |

**Table 5.** Sequential Pattern Mining results group of employees with a threshold of 0.025.

| Initial   | Pattern                                                                 |
|-----------|-------------------------------------------------------------------------|
| login     | login >> dashboard >> HR Management-Employee-View Attendance >>         |
|           | HR Management-Payroll-Employee Payroll Detail >> HR Management-         |
|           | Employee-Data                                                           |
| dashboard | dashboard >> HR Management-Employee-View Attendance >> HR Management-   |
|           | Employee-Payroll-Employee Payroll Detail >> HR Management-Employee-     |
|           | Employee-Data                                                           |

3.3. Testing by Comparing Using Interest and Not Using Interest
Testing the system without using interest will not take into account the factor of user visit time on each page. The system will build a pattern only taking into account many user visits on a page so that the page that is only a bridge to the intended page can appear in the pattern. The test will use the same threshold as the threshold for testing by using interest in each group. The following test results for each group without using interest:

3.4. Testing by Comparing The Results of The Questionnaire
The questionnaire was carried out in two stages, first was a questionnaire about the page that became the user’s favourite where the questionnaire distributed before building the system. Also, the second questionnaire regarding the pattern that is often carried out by the user, where the questionnaire distributed after building the system.
Table 6. Results of Sequential Pattern Mining group of Students without using interest.

| Initial | Pattern |
|---------|---------|
| login   | login >> dashboard >> Score-Evaluation Study Report-View Semester Score >> Score-Home |
| dashboard | dashboard >> Registration-Schedule-Student Exam Schedule >> Survey-Home |

Table 7. Results of Sequential Pattern Mining group of Lectures without using interest.

| Initial | Pattern |
|---------|---------|
| login   | login >> dashboard >> Registration-Schedule-View Exam Schedule >> Registration-Schedule-Lecture Schedule |
| dashboard | dashboard >> Registration-Schedule-View Exam Schedule >> Registration-Schedule-Lecture Schedule |

Table 8. Results of Sequential Pattern Mining group of Employees without using interest.

| Initial | Pattern |
|---------|---------|
| login   | login >> dashboard >> HR Management-Organization Structure-View Organization Structure >> HR Management-View Organization Structure |
| dashboard | dashboard >> HR Management-Organization Structure-View Organization Structure >> HR Management-View Organization Structure |

3.4.1. Page Questionnaire  
In addition to testing the system, conducting surveys by distributing questionnaires to students, lecturers, and employees as users. These questionnaires will know the real condition. Then the results of the questionnaire will be compared with the results of the system to determine the accuracy of the truth of the system was build. The questionnaire uses a sampling area technique that will take samples in specific areas [10]. Favourite pages taken from the results of the questionnaire are only the top 15 menus.

The pages contained in the pattern of test results on the system has similarities to the questionnaire by 33% for the group of students, 46% for the group of lecturers, and 41% for the group of employees.

Figure 2. Pie Chart Student Favorite Page.  
Figure 3. Pie Chart Lecture Favorite Page.


3.4.2. Pattern Questionnaire  After testing the system, the questionnaire distributed for the second time. In the second questionnaire, the survey conducted on the patterns commonly used by users. The contents of the questionnaire will provide several patterns from the results of the system and several other patterns.

In the student group pattern in the questionnaire, the results of the system testing are on option 1 and option 2. For the lecturer pattern group in the questionnaire, the results of system
testing are on option 1. Also, for the employee pattern group in the questionnaire the results of a testing system that is on option 3.

4. Conclusion
The Ant Colony Optimization method for Sequential Pattern Mining by paying attention to the visiting time factor and the factors of many visits on the page on the website, can form an access pattern that is often accessed by the user more efficiently by using a parameter that is $\alpha = 1$, $\beta = 1$, $\rho = 0.8$, $Q = 0.5$, $a = 0.7$, $b = 0.3$, $m = 5$, and $NCMax = 10$. The threshold determination when building access patterns can affect the short length of the pattern obtained. Is known after testing using several thresholds. If the threshold is too small, then the pattern that builds will be too long so that there are menus that are not too often visited by users in the pattern. If the threshold is too large, then the pattern that built will be too short and not too representative of the pattern of access that is often done by the user. The threshold used by the student group is 0.02, 0.05 lecturer group, and 0.025 employee group to get the results of the habit patterns of users accessing good Igracias. Forming a pattern of user impressions without using interest with the ant colony optimization method is difficult to find the final path through which the ants use the same parameters if using interest so that it requires much more iterations.

References
[1] X. Wei, Y. Wang, Z. Li, T. Zou, and G. Yang. Mining users interest navigation patterns using improved ant colony optimization. Intelligent Automation Soft Computing, 21(3):445454, 2015.
[2] Ling, H., Liu, Y. and Yang, S., 2007, May. An ant colony model for dynamic mining of users interest navigation patterns. In Control and Automation. 2007. ICCA 2007. IEEE International Conference on (pp. 281-283). IEEE.
[3] R. Rohdiniyah, I. Asror, and G. A. A. Wisudawan. Penggunaan metode berbasis graph untuk mining frequent sequential access pattern pada studi kasus: Website igracias universitas telkom. Indonesian Journal on Computing (Indo-JC), 2(1):91-100, 2017.
[4] Nissa, A.I.K., Asror, I. and Wisudiawan, G.A.A., 2017. ANALISIS DAN IMPLEMENTASI WEB USAGE MINING MENGGUNAKAN ALGORITMA SEQUENTIAL PATTERN DISCOVERY USING EQUIVALENCE CLASSES SPADE (Studi Kasus: Website iGracias Universitas Telkom). Jurnal Ilmiah Teknologi Informasi Terapan, 3(3).
[5] U. Jaiswal and S. Aggarwal. Ant colony optimization. International Journal of Scientific Engineering Research, 2(7):17, 2011.
[6] M. Dorigo, M. Birattari, and T. Stutzle. Ant colony optimization artificial ants as a computational intelligence technique. IEEE Comput. Intell. Mag, 1:2839, 2006.
[7] Q. Liu, M. Huang, B. Guo, and N. Chen. Research on browsing pattern of group users based on aco. 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), 5:V5394V5397, 2010.
[8] M. Dorigo and L. M. Gambardella. Ant colonies for the travelling salesman problem. Biosystems, 43(2):7381, 1997.
[9] S.-P. Tseng, M.-C. Chiang, and C.-S. Yang. L-nearest neighbors ant colony optimization for data clustering. In Machine Learning and Cybernetics (ICMLC), 2013 International Conference on, volume 4, pages 16841690. IEEE, 2013.
[10] U. Sekaran and R. Bougie. Research methods for business: A skill building approach. John Wiley Sons, 2016.