



# Evaluations of Museum Recommender System Based on Different Visitor Trip Times

Taweesak Sanpechuda<sup>ID</sup> and La-or Kovavisaruch\*<sup>ID</sup>

Department of LAI, National Electronics and Computer Technology Center, 12120, Thailand

## Abstract

The recommendation system applied in museums has been widely adopted owing to its advanced technology. However, it is unclear which recommendation is suitable for indoor museum guidance. This study evaluated a recommender system based on social-filtering and statistical methods applied to actual museum databases. We evaluated both methods using two different datasets. Statistical methods use collective data, whereas social methods use individual data. The results showed that both methods could provide significantly better results than random methods. However, we found that the trip time length and the dataset's sizes affect the performance of both methods. The social-filtering method provides better performance for long trip periods and includes more complex calculations, whereas the statistical method provides better performance for short trip periods. The critical points are defined to indicate the trip time for which the performances of both methods are equal.

**Index Terms:** Evaluation, Museum, Recommendation, Similarity, Statistic

## I. INTRODUCTION

Most museums have a wide variety of exhibits; however, there is limited visitation time. It is therefore difficult for visitors to see all exhibits. In addition, they might not know which exhibit to view first. Various efforts have recently been made to address this issue. In [1], the sound-augmented reality is applied to enhance museum visits. The authors provide an audio guide to visitors consisting of ambient sounds and comments associated with the exhibits. In [2], mobile devices are used to introduce exhibits focusing on the semantic relationship of the user preferences. The authors of [3] and [4] applied location-aware mobile services to provide visitors with cultural content, relying on Bluetooth low-energy beacons for proximity and localization capabilities. These studies have enhanced the use of wearable devices for faster and easier viewing of exhibits. In addition, the recording of a visited exhibit can be a resource for analyzing and

creating recommendations based on visitor preferences. In [5] and [6], a method is presented for analyzing the similarity of individual visitors from the same viewing information. The recommended lists are selected from a similar group of visitors. In addition, in [7], the authors offered recommendations based on an ontological formalization from knowledge of manipulated entities and graph-based semantics. In [8], it is suggested that the use of keywords of visitor interests can also be an alternative way to choose an exhibit. The exhibits are predefined into clusters based on keywords. By contrast, the authors in [9] proposed a recommender system for cultural tourism to minimize visitor dissatisfaction. Alternatively, in [10], a recommendation is presented based on visitor behaviors, such as crowd tolerance, speed of movement, and time spent on each exhibit. This suggests that using a standard overview of the logging information of an exhibit can be an easy way to create a recommended exhibit list for visitors. At the same time, in [11] and [12], the issue

Received 27 October 2021, Revised 23 November 2021, Accepted 10 December 2021

\*Corresponding Author La-or Kovavisaruch (E-mail: [la-or.kovavisaruch@nectec.or.th](mailto:la-or.kovavisaruch@nectec.or.th), Tel: +66 02-564-6900 ext. 2636)

Department of LAI, National Electronics and Computer Technology Center, 12120, Thailand.

Open Access <https://doi.org/10.6109/jicce.2022.20.2.131>

print ISSN: 2234-8255 online ISSN: 2234-8883

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © The Korea Institute of Information and Communication Engineering

of time constraint is raised as a condition for exhibit selection, and in [2], personalized recommendations are generated from contextual information (location and time) using visitor profiles. In addition, in [13], the artwork recommendation system, including digital collections such as movies, is proposed. The social filtering method uses personal visitation data to find similar demographics, whereas the statistical method uses general visitation data to recommend popular exhibits. We studied both methods at different trip times and used datasets based on evaluation scores to optimize the recommendation performance.

Several methods have been used to evaluate the performance of recommendation systems. In [14-15], the authors asked visitors to complete a questionnaire to assess their satisfaction. In addition, in [16], a recommendation of artwork sequences for a group of visitors is provided. Offline analysis of a pilot study conducted in a simulated museum environment is used to evaluate the implementation of a prototype. In [17], the authors focused on presenting exhibits according to the children’s preferences with the possibility of introducing new cultural topics and then letting a team of experts evaluate the recommendation list according to the specified criteria. Moreover, in [18], the average time that viewers view an exhibit was evaluated to adjust the length of the audio content. This study evaluated the performance of recommendations from a visitor behavior database. This evaluation was based on the F1 method [19], which was applied as a satisfaction assessment, considering both recall and precision variables used to cover all types of errors. Furthermore, we found that the length of the trip time affects the referral performance. The selection of appropriate guidance methods based on different trip times is presented in this study.

This paper is divided as follows. Section II describes the recommendation methods, namely, social filtering and statistical methods. Section III introduces the evaluation method applied in this study, including the assessment and data pre-processing factors. Section IV presents the evaluation results. Finally, Section V provides a discussion and some concluding remarks regarding this study.

## II. RECOMMENDER SYSTEM APPROACHES

A benchmark for assessing the performance is referenced on a random method in this study. In the random approach, each exhibited object is randomly selected for presentation to the visitors with a uniform probability of  $1/n$ , where  $n$  is the total number of exhibits. The number of exhibits selected depends on the specified number of trips. For the same trip time, we used the evaluation score of the random method to reference the normalization of our studied approaches, i.e., the social-filtering and statistical methods. The trip time was calculated as the average time spent viewing each object.

## A. Social-Filtering Method

The social-filtering method employs a similarity analysis of individual visitor behavior as a reference [11]. This is based on the hypothesis that visitors who view the same exhibits tend to have similar preferences. Similarly, exhibits viewed by the same visitor are similar. For example, in Fig. 1, visitors 1 and 2 view the same exhibits (A, B, and D). They, therefore, weigh three on the visitor graph. It can be concluded that their behavior is more similar to that of other visitors. Likewise, in the exhibit view, exhibits A, B, and D have two visitors on the same point-of-interest (POI) graph. Thus, they are more similar to the other exhibits.

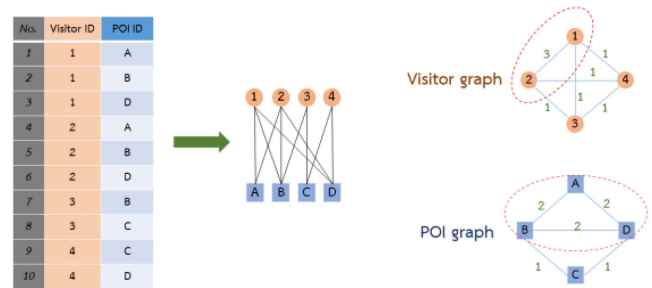


Fig. 1. The Activity log of visitors and exhibits (POIs) as visitor and POI graphs.

Let matrix  $R$  be the relationship between the visitor and exhibit, with dimensions of  $L \times C$ , where  $L$  is the number of visitors, and  $C$  is the number of exhibits. In addition,  $R_{ui}$ , an element in matrix  $R$ , is 1 when visitor  $u$  has viewed exhibit  $i$  and is 0 when visitor  $u$  has yet to view it, as shown in (1).

$$R = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \quad (1)$$

The similarity between the interested visitor  $m$  and other visitors  $u$  can be calculated using (2). We define  $r_m$  as the row vector of matrix  $R$  corresponding to visitor  $m$ , and  $r_u$  as the row vector of matrix  $R$  corresponding to other visitors. Moreover,  $\alpha$  is a weighting parameter of  $[0,1]$ .

$$\text{Similarity}(m, u) = \frac{r_m \cdot r_u}{\|r_m\|^{2\alpha} \times \|r_u\|^{2(1-\alpha)}} \quad (2)$$

$$R = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}^T = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \quad (3)$$

Here,  $r_m \cdot r_u = \sum_{i=1}^C r_{mi} \cdot r_{ui}$ , the dot product between vectors  $r_m$  and  $r_u$  when  $C$  is the number of exhibits. An exhibit list that anyone has visited will be compiled as a list of recommendations to other visitors. We can also apply (2) to calculate the similarity between exhibit  $m$  and  $u$  using the transpose matrix  $R^T$  in (3). However, we used visitor similar-

ities in our experiment.

### B. Statistical Method

This method uses visitor logging data to create a list of recommended exhibits, as shown in Fig. 2. The data were gathered from various repositories and processed to introduce exciting exhibits. The viewing frequency and duration are critical factors in selecting the recommendation list presented in [2]. Viewing frequency compares how many exhibits have been viewed without identifying the visitors. For example, POI (or exhibit) A has a viewing frequency of 3, and that of POI B is 1. In comparison, the viewing duration is the average viewing time from when a visitor first begins viewing an exhibit until viewing the following exhibit. In Fig. 2, visitor 1 starts viewing A at 12.00, and B at 12.05, thus spending 5 minutes viewing exhibit A. Using this calculation concept, we obtained viewing durations of 4 and 8 min for exhibit A by visitors 2 and 3, respectively.

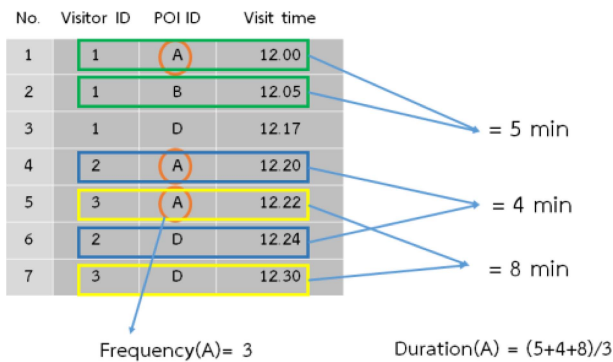


Fig. 2. Calculation of viewing frequency and duration.

Using this method, the selection of exhibits assumes that the higher the viewing frequency and the longer the viewing duration is, the more exciting the exhibits. The exhibit viewing duration is also used to determine the maximum time visitors spend at the museum (trip time).

## III. EVALUATION

To evaluate the performance of each recommendation method, we define the following factors.

- The total numbers of data and visitor log records in the database are determined. In this article, we analyzed three museums with different characteristics.
- The recommended duration (trip time) defines the period of recommendation based on the average viewing duration of different exhibits. The number of recommended exhibits could therefore vary.

### A. Data Cleaning

We used databases from two museums on our platform (museum pool), i.e., Chao Sam Phraya Museum (Thailand) and Shwedagon Museum (Myanmar), to evaluate the exhibit recommendation. Both museums have been in operation for more than two years. A summary of these databases is provided in Table 1. The *total records* show the number of visiting logs recorded within a specific period, and *total exhibits* and *total visitors* represent the whole numbers of exhibits and visitors, respectively. The *selected visitor* is the number of visitors found under the specified conditions, and we selected visitors with at least eight viewed exhibits in our evaluations. The log data from Shwedagon are noticeably fewer in number than those from Chao Sam Phraya.

Table 1. Summary of experimental database

Site	Total Records	Total Exhibits	Total Visitors	Selected Visitors
Chao Sam Phraya	34,256	96	2,926	671
Shwedagon	5,944	35	1,870	66

### B. Evaluation Implementation

To evaluate the performance of the recommender system, we use the F1-score, which consists of precision and recall, which are calculated from the variables shown in Fig. 3. The circular area shows the recommended exhibits, with the exhibits they visit placed on the left side. We define these variables as follows:

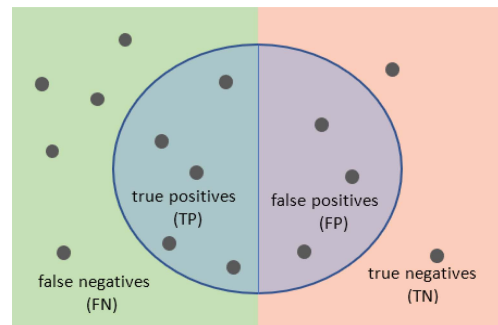


Fig. 3. Definition of variables used in the F1-score.

- *True positives* (TP) is the number of recommended exhibits that the user visits.
- *False Positives* (FP) is the number of recommended exhibits that a user has not visited.
- *False negatives* (FN) is the number of exhibits that have been visited but have yet to be recommended.
- *True negatives* (TN) is the number of exhibits that have yet to be visited and have not been recommended. The

TN value is typically invalid.

The precision is the proportion of recommended exhibits in which the user visits all recommended objects. We can calculate the precision using  $TP/(TP+FP)$ , where the recall is the ratio of recommended exhibits visited by users to the total number of visited exhibits, which we can calculate from  $TP/(TP+FN)$ . Finally, using the harmonic mean [20], we calculate the F1 value according to (4). In cases where the recommendation is most effective, the F1-score is 1.

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{TP}{TP + \frac{FP + FN}{2}} \tag{4}$$

#### IV. RESULTS AND DISCUSSION

In our evaluation, we defined the trip time as the sum of the viewing duration of our recommended exhibits. We assigned the experiment four different trip times (30, 60, 120, and 240 min). The F1-score of each recommended method proposed (4) is normalized to the random method, as shown in Table 2 for Chao Sam Phraya and Table 3 for Shwedagon. The results are the percentage of performance gain compared to the random method.

**Table 2.** Performance of implemented methods against a random method at Chao Sam Phraya Museum for different trip times

Trip time (min)	Social filtering	Statistical
30	-39.86%	-16.22%
60	+24.61%	+29.23%
120	+29.19%	+11.67%
240	+33.20%	+11.58%

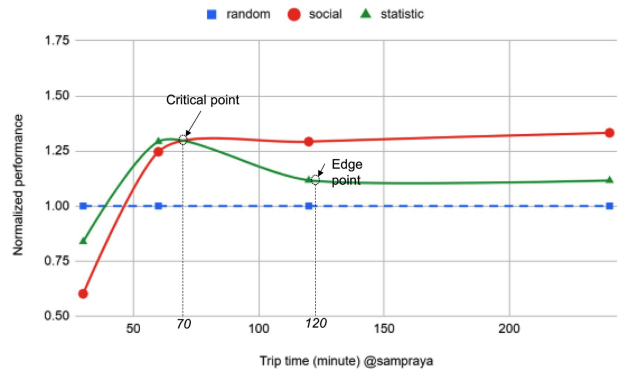
**Table 3.** Performance of implemented methods against the random method at Shwedagon Museum for different trip times.

Trip time (min)	Social filtering	Statistical
30	+48.45%	+34.70%
60	+50.79%	+14.11%
120	+49.05%	+12.79%
240	+52.43%	+15.35%

##### A. Chao Sam Phraya Museum

These cases belong to the Chao Sam Phraya Museum in Thailand, where the visiting data are from the most extensive dataset of 34,256 records, as mentioned in Table 1. The data include 2,926 visitors during approximately 3.5 years. After the data-cleaning process, 671 visitors met the conditions outlined in the previous section. Fig. 4 shows the normalized

F1-scores in the graphs for different trip times. The blue dashed line shows the reference for the normalized performance. The red line with a circular point represents the normalized performance of the social filtering method. Furthermore, the green line with a triangular point is the normalized performance of the statistical method.



**Fig. 4.** Normalized F1-scores of Chao Sam Phraya Museum.

It can be seen that the random methods initially achieve the best performance from Fig. 4, following statistical and social filtering methods in that order. Both statistical and social filtering methods outperformed the random method at a specific trip time. The critical point is the point where both social filtering and a statistical method perform the same. After reaching the critical point (at approximately 70 min), the social filtering performance increased and stabilized at approximately 120 min. Simultaneously, the statistical performance decreased and was stable for approximately 120 min. The average time spent visiting all exhibits in this museum was 120 min. After 120 min, there were no longer any exhibits to recommend to the visitors; therefore, we call this time the edge point. As a result, choosing a statistical method to guide the exhibit would be better for visitors with a trip time of less than 70 min. By contrast, the social filtering method should be applied to visitors with a trip time of more than 70 min.

##### B. Shwedagon Museum

These cases belong to the Shwedagon Museum in Myanmar. This is an indoor museum, similar to the Chao Sam Phraya Museum in the previous case. However, only 5,944 data records were available, with 66 of the 1,870 visitors meeting the criteria. The normalized F1-scores of the Shwedagon Museum are shown in Fig. 5.

From Fig. 5, we can see the trip times range in which both methods perform better than the reference approach. The edge point in the statistical method is approximately 60 min, which is close to the time it takes to view all objects at Shwedagon Museum. However, we have yet to identify a

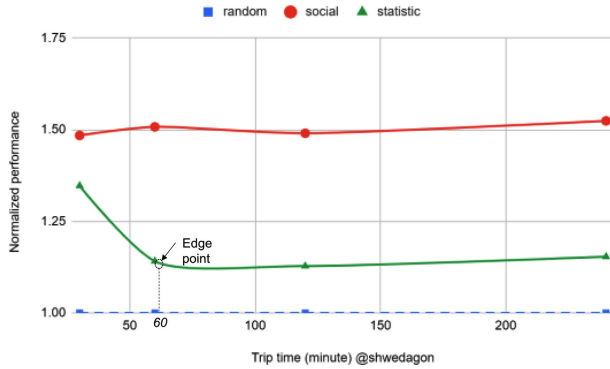


Fig. 5. Normalized F1-scores of Shwedagon Museum.

critical point for this experiment. However, the performance of the statistical method during a 30-min period increases to 34.70%, as shown in Table 3. The result of the approach reaches that of the social filtering method; thus, the critical point should be less than 30 min. This means that choosing a statistical method to guide the exhibit for visitors with a trip time of less than 30 min should be a better option. By contrast, the social filtering method should be applied to visitors with a trip time of more than 30 min.

### C. Discussion

The results showed that both methods were more effective than the random method when the trip time exceeded a specific value. The statistical method performs better when the trip time is less than the critical point. The performance then decreases to the edge point, close to the time it takes to view all objects in the museum. Before the critical point, the statistical method performed better because the interests of visitors were quite similar initially. However, as time passes, using overview statistics will not meet the needs of individual visitors. Simultaneously, the social filtering method guarantees better performance when the trip time exceeds the critical point. This means we can use the critical point to select the correct method for visitors with different trip times. We found that the critical point differed for each dataset. Our experiment showed that the critical point of the Chao Sam Phraya Museum is 70 min, while that of the Shwedagon Museum is less than 30 min. According to the experimental results, the critical point is always before the edge point. We can then estimate the critical point from the edge point based on the average viewing time of all objects in the museum.

## V. CONCLUSION AND FUTURE WORK

We studied the trip time on two different datasets that affect a recommendation performance. We evaluated the sat-

isfaction of two recommendation methods: 1) With the social filtering method, which analyzes the similarity among visitors using a visited exhibit database, visitors who visit the same exhibits are assumed to have the same preferences. Therefore, recommended exhibits should be selected from the same visitors with identical preferences. 2) The statistical method uses general statistical data, such as the most visited exhibits or exhibit that people have viewed the most. This information is used to select the objects exhibited to the visitor. Both methods have different principles for selecting exhibits, and we chose the F1-score as our assessment tool to evaluate their performance. These methods were normalized using the random method, and the exhibits were randomly presented to the visitors.

In a future study, other environmental information such as transportation, crowd density, facilities, and nearby attractions may be used as factors in introducing outdoor exhibits.

## ACKNOWLEDGEMENTS

We would like to thank Chao Sam Phraya Museum and Schwadogon Museum, essential resources for this study. We also wish to thank Professor Myint Myint Sein and her students from the University of Computer Study Yangon, who supported the findings and facilitated contact with resources in Myanmar.

## REFERENCES

- [1] F. Z. Kaghat, A. Azough and M. Fakhour, "SARIM: A gesture-based sound augmented reality interface for visiting museums," in *International Conference on Intelligent Systems and Computer Vision*, pp. 1-9, 2018. DOI: 10.1109/ISACV.2018.8354050.
- [2] I. Benouaret and D. Lenne, "Personalizing the museum experience through context-aware recommendations," in *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 743-748, 2015, DOI: 10.1109/SMC.2015.139.
- [3] S. Petros and N. Konstantinos, "BLE beacons for indoor positioning at an interactive IoT-based smart museum," *IEEE Systems Journal*, vol. 14, no. 3, pp. 3483-3493, 2020. DOI: 10.1109/JSYST.2020.2969088.
- [4] A. Stefano, C. Rita, F. Giuseppe, M. Luca, M. Vincenzo, P. Luigi, and S. Giuseppe, "An indoor location-aware system for an iot-based smart museum," *IEEE Internet of Things Journal*, vol. 3, no. 2, pp. 244-253, 2016. DOI: 10.1109/JIOT.2015.2506258.
- [5] C. C. Aggarwal, "An introduction to recommender system," in *Recommender Systems*, Springer., pp. 1-28, 2016.
- [6] T. Kuflik, E. Minkov, and K. Kahanov, "Graph-based recommendation in the Museum," in *CEUR Workshop Proceedings*, Bolzano: Italy, pp. 46-48, 2014. Available: <http://ceur-ws.org/Vol-1278/paper9.pdf>.
- [7] D. Louis and N. Yannick, "A graph-based semantic recommender system for a reflective and personalized museum visit: Extended abstract," in *12th International Workshop on Semantic and Social Media Adaptation and Personalization*, pp. 88-89, 2017. DOI: 10.1109/SMAP.2017.8022674.
- [8] D. Luh and T. Yang, "Museum recommendation system based on



- lifestyles,” in *9th International Conference on Computer-Aided Industrial Design and Conceptual Design*, Kunming, pp. 884-889, 2008. DOI: 10.1109/CAIDCD.2008.4730703.
- [9] P. George, “Apollo - A hybrid recommender for museums and cultural tourism,” in *International Conference on Intelligent Systems*, pp. 94-101, 2018. DOI: 10.1109/IS.2018.8710494.
- [10] I. Lykourantzou, C. Xavier, Y. Naudet, E. Tobias, A. Antonio, G. Lepouras, and C. Vassilakis, “Improving museum visitors; quality of experience through intelligent recommendations: A visiting style-based approach,” in *9th International Conference on Intelligent Environments*, Athens: Greece, pp. 507-518, 2013. DOI: 10.3233/978-1-61499-286-8-507.
- [11] L. Kovavisaruch, T. Sanpechuda, K. Chinda, T. Wongsatho, S. Wisadsud, and A. Chaiwongyen, “Incorporating time constraints into a recommender system for museum visitors,” *Journal of Information and Communication Convergence Engineering*, vol. 18, no. 2, pp. 123-131, 2020. DOI: 10.6109/jicce.2020.18.2.123.
- [12] M. C. Rodriguez, S. Ilarri, R. Hermoso, and R. T. Lado, “Towards trajectory-based recommendations in museums: Evaluation of strategies using mixed synthetic and real data,” in *Procedia Computer Science*, pp. 234-239, 2017. DOI: 10.1016/j.procs.2017.08.355.
- [13] G. Ignacio, “A hybrid approach for artwork recommendation,” in *IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering*, pp. 281-284, 2019. DOI: 10.1109/AIKE.2019.00055.
- [14] I. Keller and E. Viennet, “Recommender systems for museums: evaluation on a real dataset,” in *Fifth International Conference on Advances in Information Mining and Management*, Brussels: Belgium, pp. 65-71, 2015. Available: [http://72.52.166.99/articles/imm\\_2015\\_5\\_10\\_98004.pdf](http://72.52.166.99/articles/imm_2015_5_10_98004.pdf).
- [15] G. Pavlidis, “Towards a novel user satisfaction modelling for museum visit recommender,” in *Communications in Computer and Information Science*, Springer., pp. 60-75, 2018. DOI: 10.1007/978-3-030-05819-7\_6.
- [16] R. Silvia, B. Francesco, G. Clemente, and R. Luca, “Artworks sequences recommendations for groups in museums,” in *12th International Conference on Signal-Image Technology & Internet-Based Systems*, pp. 445-462, 2016. DOI: 10.1109/SITIS.2016.77.
- [17] E. P. Arias, C. A. Medina, B. V. Robles, B. Y. Robles, A. F. Pesntez, G. P. Solzano, and J. Ortega, “An expert system to recommend contents and guided visits for children: A practical proposal for the Pumapungo Museum of Cuenca,” in *IEEE International Autumn Meeting on Power, Electronics and Computing*, Ecuador, pp. 1-6, 2018. DOI: 10.1109/ROPEC.2018.8661377D.
- [18] L. Kovavisaruch, T. Sanpechuda, K. Chinda, and V. Somlertlamvanich, “Museum content evaluation based on visitor behavior,” in *13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, Thailand, pp. 1-5, 2016. DOI: 10.1109/ECTICon.2016.7561472.
- [19] M. W. David, “Evaluation: From precision, recall and f-score to roc, informedness, markedness & correlation,” *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37-63, 2011. DOI: 10.48550/arXiv.2010.16061.
- [20] Weisstein, Eric W. at Wolfram Research, Harmonic Mean [Internet], Available: <https://mathworld.wolfram.com/HarmonicMean.html>.



### Taweesak Sanpechuda

received BS and MS degrees from the Department of Electrical Engineering at Chulalongkorn University in Thailand. He is currently a researcher at the National Electronics and Computer Technology Center (NECTEC) in Thailand. His research interests include data analysis, wireless communications, and indoor positioning.



### La-or Kovavisaruch

received her BS degree from King Mongkut's Institute of Technology Ladkrabang, Bangkok Thailand in 1993, M.Sc. degree in electrical engineering from Imperial College, London, U.K., in 1995, M.Sc. degree in electrical engineering from the University of Southern California, USA, in 2001, and Ph.D. in electrical engineering from the University of Missouri-Columbia, USA, in 2005. She is currently working as a senior researcher at the Location and Auto-ID Laboratory in the National Electronics and Computer Technology Center, Thailand. Her research interests include indoor localization, wireless communication, inventory, and warehouse management technologies.