

로그 데이터를 이용한 기업 정보 시스템의 사용 패턴 분석

Utilization Pattern Analysis of an Enterprise Information System using Event Log Data

한관희

경상국립대학교 산업시스템공학부/공학연구원

Kwan Hee Han(hankh@gnu.ac.kr)

요약

최근 들어 기업정보 시스템의 성공적인 활용은 기업의 효과적인 전략 수행과 기업 경영 목적 달성에 핵심적인 역할을 하고 있다. 기업정보 시스템의 도입 성공 여부를 결정하는 인자 중의 하나가 시스템 활용도로서, 정보시스템 활용도를 측정하는 효과적인 방안을 마련하는 것은 기업 경영에서 매우 중요하다. 본 연구에서는 프로세스 마이닝 기법을 이용한 로그 데이터 분석을 통해 정보 시스템 활용도를 평가한다. 즉, 기본적인 접속 로그 통계뿐만 아니라 정보시스템을 실제로 사용하는 순서를 패턴화하고 유사 사용 기능을 군집화하여, 정보 시스템 설계시에 예상하였던 시스템 사용 순서/기능과 실제 현황을 비교함으로써 기존 정보시스템을 개선하거나 확장하는데 피드백을 제공하고자 한다.

■ 중심어 : | 로그 데이터 | 프로세스 마이닝 | 사용 패턴 | 기업 정보 시스템 | 군집화 |

Abstract

The success of enterprise information system(EIS) is crucial to align with corporate strategies and eventually attain corporate goals. Since one of the factors to information system success is system use, managerial efforts to measure the level of EIS utilization is vital. In this paper, the EIS utilization level is analyzed using system access log data. In particular, process sequence patterns and clustering of similar functions are identified in more detail based on a process mining method, in addition to basic access log statistics. The result of this research can be used to improve existing information system design by finding real IS usage sequences and function clusters.

■ keyword : | Log Data | Usage Pattern | Enterprise Information System | Process Mining | Clustering |

I. Introduction

In order to survive in the currently competitive and global business environment, most enterprises are struggling to adapt their existing business processes into more agile, product-focused and customer-oriented

structures. Within this environment, an enterprise information system (EIS) is a core element for strategic advantages. Therefore, the success of the EIS is a crucial factor to align with corporate strategies, and eventually attain corporate goals.

Since DeLone and McLean (2002) developed

접수일자 : 2022년 07월 05일

수정일자 : 2022년 07월 28일

심사완료일 : 2022년 07월 30일

교신저자 : 한관희, e-mail : hankh@gnu.ac.kr

their IS (Information System) success model called D&M model, there has been extensive research on the topic of success as well as extensions and tests of their model[1]. An updated D&M model (2008) consists of six major success dimensions, as depicted in [Figure 1][2]. One of these is 'system use'. This is the degree and manner in which staff and customers utilize the capabilities of an information system. The example of system usage includes the amount of use, frequency of use, nature of use, appropriateness of use, extent of use, and purpose of use.

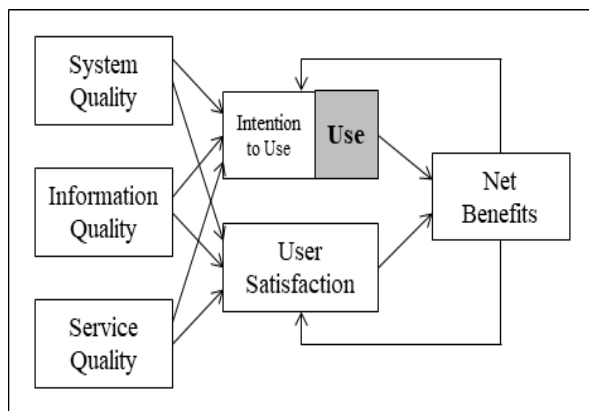


Figure 1. Updated D&M(DeLone and McLean) IS Success Model

As discussed above, since one of the factors to EIS success is system use, managerial efforts to measure the level of IS utilization is vital. Analysis of system log data recorded when users access the IS is widely used to measure the IS usage.

Since IS utilization is multi-faceted, in this paper, it is further categorized into 3 types, as follows: (1) process: which sequence of activities does the participant perform on the IS?; (2) function: which functionalities does the IS user utilize?; and (3) user/organization: who uses the IS data and when does he or she use

the IS?

Analysis of the first type is a complex and difficult task, whereas analysis of the second and third type of system use is relatively easy using basic log statistics.

Process mining, which is one of system log data analysis technique, elicits useful information using accumulated event logs during business process execution[3][4]. Its main purposes are as follows: (1)derive the process model that enables the representation of the dependency relationship between business activities[5]; (2)extend or improve an existing business process model using information about the actual process recorded in some event log[4][6]; and (3)check the conformance of reality, as recorded in the logs, and conform to the pre-defined model[7].

The aim of this paper is to analyze the pattern of EIS utilization for IS success using log data, especially focusing on the elicitation of IS usage sequence patterns and function clustering by process mining.

II. Related Work

With regard to relationships between information usage levels and organizational performance, Barisic and Bach's research results (2019) revealed that the strongest impact to organizational performance is attained through the intensity of IS usage, measured by the number of different functionalities available in the software systems[8].

Cuellar et al. (2006) reviewed various attempts to measure system use and then proposed a re-conceptualization as "the level of incorporation of an information system within

a user's processes." They developed the concept of a Functional Interface Point (FIP) and four dimensions of system usage as follows: (1) automation level which is the proportion of the business process encoded by the information system, (2) extent which is the proportion of FIPs used by the business process, (3) frequency which is the rate at which FIPs are used by participants in the process, and (4) thoroughness which is the level of use of information/functionality provided by the system at an FIP[9].

Edwards (2007) explored the relationship between the most utilized screens and the changes requested by users of the system by simply counting screen hits and design changes [10]. Lopez and David (2006) proposed a data warehousing method based on click stream to analyze the usage of a web information system. They presented a dimensional model to monitor user behavior in Higher Education Web Information Systems and architecture for the extraction, transformation and load process. However, it resulted in simple analysis based on basic statistics such as page counts, user platform and used browser by user type[11].

Yan et al. (1996) described an approach for automatically classifying visitors of a web site based on their access patterns. User access logs are examined to discover clusters of users that exhibit similar information needs[12].

Kohavi (2001) proposed web mining of e-commerce sites based on web server logs and helped detect business opportunities that otherwise could remain unnoticed[13]. Chen et al. (1996) explored a data mining capability called path traversal patterns, which involves mining access patterns in a distributed information-providing environment where

documents or objects are linked together to facilitate interactive access[14].

As an event log analysis platform for information system management, Li et al. (2017) developed an integrated system called FIU Log Analysis Platform (FLAP) that aims to facilitate the data analytics for system event logs. FLAP provides an end-to-end solution that utilizes advanced data mining techniques to assist log analysts to conveniently, timely, and accurately conduct event log knowledge discovery, system status investigation, and system failure diagnosis[15]. Leemans et al. (2018) presented a software behavior exploration tool called the statechart workbench which provides a rich and mature integration of advanced mining techniques for the analysis of behavior, performance (timings), frequency (usage), conformance and reliability in the context of various formal models[16].

As an application of process mining approach, Jans et al. (2011) presented and applied a framework for internal fraud risk reduction, where risk reduction stands for both fraud detection and prevention. Their work was based on process mining aimed at uncovering a process model based on real transaction logs[17]. Han et al. (2015) presented a process mining technique to identify navigation patterns of web sites by analyzing user access logs. This analysis was used to improve the user interface design of web sites. They suggested that the proposed approach is expected to aid the information system administrator of the university in designing their website and to make it easy for users to use the university website[18]. Chiu and Jans (2019) showed that process mining enables auditors to detect potential risks, ineffective internal controls, and

inefficient processes from internal logs[19]. Caron et al. (2013) applied process mining techniques and tools to enterprise risk management, which minimizes the negative effects of uncertainty on the objectives. They provided a full exploration of the applicability of process mining in the context of the eight components of the COSO Enterprise Risk Management Framework[20]. De Weerd et al. (2013) showed real-life case studies suggesting methodologies to conduct process mining analysis and their benefits. They presented a methodological framework for multi-faceted analysis of real-life event logs based on process mining[21]. Choi et al. (2013) identified the failure pattern of a home appliance by applying a process mining method on repair log data. They suggested that product reliability can be enhanced through feedback of identified failure patterns to product development teams using the proposed method[22]. Sahraeidolatkhaneh and Han (2017) proposed integration framework of the diagnosis and (re)design phase of BPM(Business Process Management) life cycle by sharing automatically generated process model and basic statistics based on the process mining method[23].

III. Utilization Pattern Analysis

IS utilization pattern analysis was conducted on a supply chain management information system for subcontractors of a Korean manufacturing company (called "H company" in this paper). This information system (called "R system") was developed by H company for the purpose of enhancing the productivity of subcontractors, and supports the tasks of over

50 subcontractors, which supply parts and assemblies to H company. Major functionalities are as follows: (1) receiving order and delivery management; (2) shop floor control support including process control and daily labor reporting, and (3) communication services such as e-mail and bulletin boards.

IS utilization pattern in this paper can be classified into 2 types: (1) basic statistics, and (2) process sequence pattern and clustering analysis. The total number of records written in log data is 125,783, accessed by users of H company's 50 subcontractors between 2017-08-01 and 2017-09-30.

1. Basic Statistics about IS Usage Patterns

In this category, analysis results can be further categorized into 3 types: (1) who used the IS?; (2) When did they use the IS?; and (3) Which functionalities did they use?

First, since only four companies occupied 70.0% of the transactions in terms of access numbers, it was concluded that the R system is not widely used by subcontractor companies. In other words, a small number of subcontractor companies used the R system actively whereas many subcontractor companies have not used it yet. Therefore, it is urgent to increase the overall utilization level of R system.

Second, an M-shaped pattern is identified in terms of daily access number over one week, as depicted in [Figure 2]. In other words, transactions increase on Monday (the start of the week) and Thursday (near the end of the week). Overall transactions was increasing over time, as depicted in [Figure 2]. In terms of access frequency by time zone, usage of the R system is concentrated in the time slots of 8 a.m. to 11 a.m. and 1 p.m. to 4 p.m., as depicted

in [Figure 3].

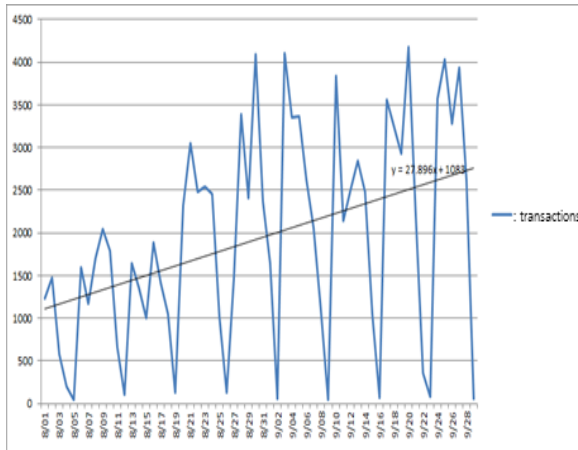


Figure 2. Daily Access Frequency

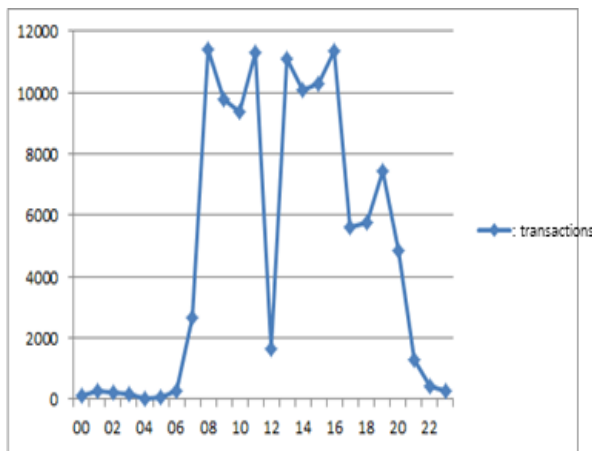


Figure 3. Access Frequency by Time Zone

Third, the shop floor control function is the most utilized (76.8%) IS support function, as depicted in [Table 1]. R system functions are divided into 3 modules: (1) order/delivery management (ODM); (2) shop floor control (SFC); and (3) communication services (CS). It was concluded that shop floor control functions are highly utilized since they occupied 76.8% of all transactions, whereas the number of screens of the shop floor control function makes up to 60.4%. Since communication services such as e-mail, bulletin boards and on-line forums make up 8.6%, this function was considered to

have a low utilization rate.

Table 1. Access Frequency by Module

Module	Number of Screens	Access Frequency
Shop Floor Control	55(60.4%)	89.1(76.8%)
Order & Delivery Management	20(22.0%)	16.9(14.8%)
Communication Service	16(17.6%)	10.0(8.6%)
Total	91(100.0%)	116.0(100.0%)

As shown in [Table 2], in terms of the access frequency of each screen, the most utilized screens are BOM & routing information maintenance and daily labor report system, which occupied 70.9 % of all access logs

2. Process Sequence Pattern Analysis

It is not clear whether system users utilize the IS following the proper designed sequence or not. Therefore, the usage sequence pattern and clustering is analyzed based on the process mining technique in this paper.

Process mining is a process for extracting meaningful knowledge from the "event log" generated when users perform certain tasks, and is also focused on the automated discovery of common usage patterns and process models from event log data[24][25].

Table 2. Access Frequency by Screen

Name of Screens	Access Frequency	Percent (%)
Change of routing data	15,113	13.0
Registration of daily labor	12,385	10.7
Inquiry of routing data	11,922	10.3
Routing list	11,903	10.3
Inquiry of detail order data	10,510	9.1
BOM list	8,252	7.1
Main menu screen of daily labor	7,087	6.1

report		
creation of routing data	2,688	2.2
Work instruction list	2,424	2.1
Sub-total	82,265	70.9
The rest	33,766	29.1
Grand Total	116,031	100.0

In general, a process is a structured set of activities designed to produce a specified output for a particular stakeholder.

The purpose of process sequence pattern analysis of information system usage is as follows: (1) to change the current process sequence to the desired process by investigating the difference between real practice and the designed process; and (2) to improve the menu structure or navigation route between screens by identifying the IS usage sequence.

The log data source consists of 125,783 records accessed by users of H company’s 50 subcontractors between 2017-08-01 and 2017-09-30. Among these records, after excluding the access records of the communication services and consecutive records of the same page, 79,735 records are analyzed.

ProM is used for process mining. This is an open source process mining tool, originally developed at Eindhoven University of Technology[26].

The log data format for process mining is shown in [Table 3]. Based on the 79,735 log data, ProM fuzzy mining algorithms were employed to determine the IS usage process sequence pattern. If there are many activities with a small frequency number, it is difficult to identify a distinct process sequence. The fuzzy mining method allows clustering of associated activities into one group.

Table 3. Log Data Format for Process Mining

Process Mining Attribute	Case Id.	Event Type	Time Stamp	Originator
Identifier in log data format	Screen number	“complete”	yy:mm:dd:hh:mm:ss	User Id.
Number of types	75	1	-	297

[Figure 4] shows the results of the R system usage process sequence pattern. The rounded rectangle represents a major activity, and an octagon represents a clustered group of associated activities.

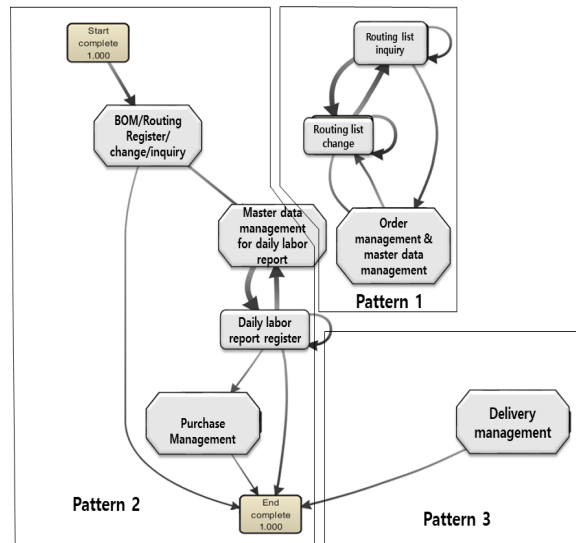


Figure 4. Process Sequence Pattern of R System Usage

The thickness of the arc between nodes (activities) represents the connection frequency between activities (i.e., screens), and a directed arc represents a precedence relationship between information activities. Furthermore, by uncovering the cluster in [Figure 4], a more detailed activity sequence is identified. As depicted in [Figure 4], three process sequence patterns are identified in this case study as follows: (1) Order & master data management process in the production preparation stage. A

detailed process sequence is shown in [Figure 5].

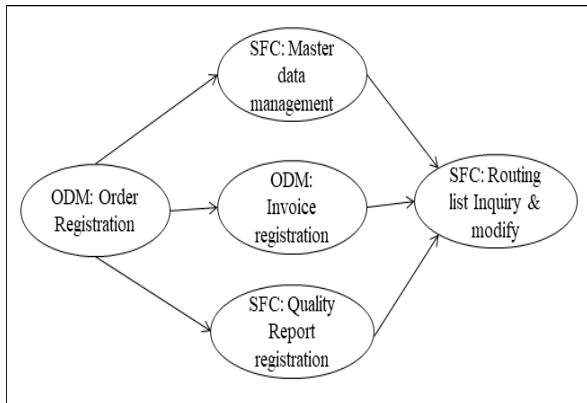


Figure 5. Cluster of Order & Master Data Management Process

(2) Daily labor report & purchase management process in the production stage. As depicted in [Figure 6], three sub-patterns are identified within this pattern as follows:

- ① Single cluster island for BOM/routing data management
- ② Daily labor report & purchase management process, which has longest steps
- ③ Sub-set process of the (2)-② pattern

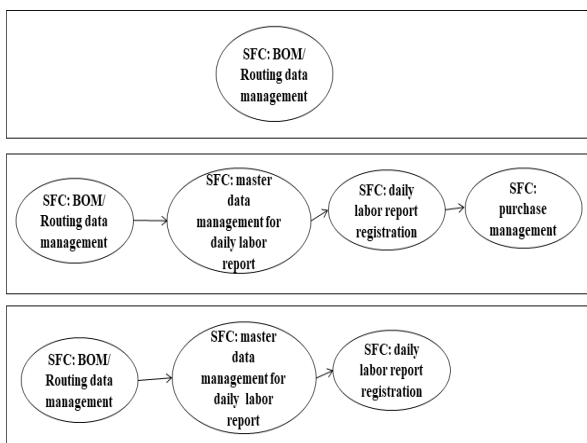


Figure 6. Cluster of Daily Labor Report & Purchase Management Process

(3) Delivery management process in the post-production stage. Within this pattern, there are two weakly connected major activities, as depicted in [Figure 7].

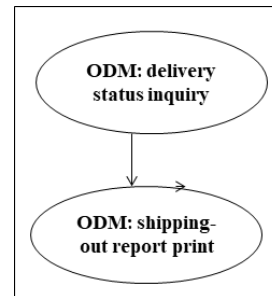


Figure 7. Cluster of Delivery Management Process

Table 4. Top 5 Longest Sojourn Time

Name of Screen	Average Sojourn Time (Min.)
BOM List	66.0
Shipping-out report print	46.0
Incoming/shipping-out items list	45.0
Inventory status inquiry	38.0
Delivery status inquiry	36.0

Besides the process sequence pattern, there are other useful functionalities in ProM for analyzing event log data of enterprise information system.

For example, by using the basic performance analysis of ProM, a user's average sojourn time in each screen is calculated in [Table 4], which shows the top 5 screens having the longest sojourn time.

Additionally, the user's access frequency is analyzed by screen name, as shown in [Figure 8]. At the 4th row of this table, LC00017 represents the user's id., and A-C-R is the name of screen.

It can be seen that user LC00017 in the 4th row of [Figure 8] accessed the A-C-R screen

313 times and the A-C-S screen 720 times over 2 months and so on.

IV. Conclusions

Many enterprise information systems are being introduced to enhance overall enterprise performance. Therefore, it is critical to measure the level of IS usage to determine the success of the IS.

originator	A-C-R	A-C-S	A-C-T	A-C-U	A-D-V	A-A	A-E-Z	A-	A-	A-	A-	A-	A-	A-	A-	A-	A-	A-
LC00010	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LC00014	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LC00016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LC00017	313	720	389	396	0	0	410	2	0	0	2	0	10	58	42	20	1353	1
LC00018	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LC00019	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LC00020	5	68	6	6	0	0	8	0	0	0	1	0	0	5	1	3	0	2
LC00021	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LC00050	0	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MIR0001	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MIR0003	1	10	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
MIR0005	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MKT0001	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAY0001	0	58	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	1
NAY0003	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAY0005	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NAY0006	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAY0007	5	48	1	3	0	0	3	1	6	0	2	5	0	4	12	4	2	13
NAY0008	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAY0028	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAY0044	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAY0049	1	186	0	1	0	0	22	0	0	0	1	0	0	2	0	0	0	69

Figure 8. Screen Access Frequency by user id.

However, it is a challenging task to measure the level of IS usage in collaborative information systems such as supply chain management, in which the number of involved organization is many, and users are distributed geographically.

In this paper, IS utilization level is analyzed using access log data. In particular, process sequence patterns and clusterings are identified in more detail in addition to basic access log statistics.

The result of this research belongs to process mining approach which is a subset of data mining discipline, and will be useful to IS

analysts for redesigning existing system by finding real IS usage sequences and clusters of similar functions.

However, even though real process sequence patterns of IS usage are identified, a comparison of the real process with the designed process was not conducted. Therefore, as a further research, design improvement of the IS based on log data analysis is required.

참고 문헌

- [1] S. Petter, W. DeLone, and E. McLean, "Measuring Information Systems Success: Models, Dimensions, Measures, and Interrelationships," *European Journal of Information Systems*, Vol.17, No.3, pp.236-263, 2003.
- [2] W. H. Delone and E. R. McLean, "The DeLone and McLean Model of Information Systems Success: a Ten-Year Update," *Journal of Management Information Systems*, Vol.9, No.4, pp.9-30, 2008.
- [3] W. M. Van der Aalst and A. J. M. M. Weijters, "Process Mining: a Research Agenda," *Computers in Industry*, Vol.53, No.3, pp.231-244, 2004.
- [4] W. Van der Aalst, "Process Mining: Discovery, Conformance and Enhancement of Business Processes," Springer Science & Business Media, 2016.
- [5] W. Van der Aalst, T. Weijters, and L. Maruster, "Workflow Mining: Discovering Process Models from Event Logs", *IEEE Transactions on Knowledge and Data Engineering*, Vol.16, No.9, pp.1128-1142, 2004.
- [6] W. M. Van der Aalst, H. A. Reijers, and M. Song, "Discovering Social Networks from Event Logs," *Computer Supported Cooperative Work (CSCW)*, Vol.14, No.6, pp.549-593, 2005.
- [7] A. Rozinat and W. M. Van der Aalst,

- "Conformance Checking of Processes Based on Monitoring Real Behavior," *Information Systems*, Vol.33, No.1, pp.64-95, 2008.
- [8] A. F. Barisic and M. P. bach, "The Intensity of Human Resources Information Systems Usage and Organizational Performance," *Interdisciplinary Description of Complex Systems*, Vol.17, No.3-B, pp.567-597, 2019.
- [9] M. J. Cuellar, E. R. McLean, and R. D. Johnson, "The Measurement of Information System Use: Preliminary Considerations," *Proceedings of the 2006 ACM SIGMIS CPR conference on Computer Personnel Research*, pp.164-168, ACM, Claremont, CA USA, 2006.
- [10] H. K. Edwards, "System Utilization and Changes in Implemented Information Systems: a Case Study," *IAENG International Journal of Computer Science*, Vol.33, No.1, pp.68-73, 2007.
- [11] C. T. Lopes, and G. David, "Higher Education Web Information System Usage Analysis with a Data Webhouse," *International Conference on Computational Science and Its Applications (ICCSA 2006)*, pp.78-87, Springer Berlin Heidelberg, 2006.
- [12] T. W. Yan, M. Jacobsen, H. Garcia-Molina, and U. Dayal, "From User Access Patterns to Dynamic Hypertext Linking," *Computer Networks and ISDN Systems*, Vol.28, No.7, pp.1007-1014, 1996.
- [13] R. Kohavi, "Mining E-commerce Data: the Good, the Bad, and the Ugly," *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, pp.8-13, ACM, San Francisco CA USA, 2001.
- [14] M. S. Chen, J. S. Park, and P. S. Yu, "Data Mining for Path Traversal Patterns in a Web Environment," *Proceedings of the 16th International Conference on Distributed Computing Systems*, pp.385-392, IEEE, Hong Kong, 1996.
- [15] T. Li, Y. Jiang, C. Zeng, B. Xia, and Z. Liu, "FLAP: An End-to-End Event Log Analysis Platform for System Management," *Proceedings of 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.1547-1556, Halifax, NS, Canada, 2017.
- [16] M. Leemans, W. M. P. Van der Aalst, and M. G. J. Van den Brand, "The Statechart Workbench: Enabling Scalable Software Event Log Analysis Using Process Mining," *2018 IEEE 25th International Conference on Software Analysis, Evolution and Reengineering*, pp.502-506, Campobasso, Italy, 2018.
- [17] M. Jans, J. M. van der Werf, N. Lybaert, and K. Vanhoof, "A Business Process Mining Application for Internal Transaction Fraud Mitigation," *Expert Systems with Applications*, Vol.38, No.10, pp.13351-13359, 2011.
- [18] K. H. Han, B. Hwang, and J. Jeon, "A Navigation Pattern Analysis of University Department's Websites Using a Processing Mining Approach," *Innovations in Education and Teaching International*, Vol.52, No.5, pp.485-498, 2015.
- [19] T. Chiu and M. Jans, "Process Mining of Event Logs: A Case Study Evaluating Internal Control Effectiveness," *Accounting Horizons*, Vol.33, No.3, pp.141-156, 2019.
- [20] F. Caron, J. Vanthienen, and B. A. Baesens, "Comprehensive Investigation of the Applicability of Process Mining Techniques for Enterprise Risk Management," *Computers in Industry*, Vol.64, No.4, pp.464-475, 2013.
- [21] J. De Weerd, A. Schupp, A. Vanderloock, and B. Baesens, "Process Mining for the Multi-Faceted Analysis of Business Processes—A Case Study in a Financial Services Organization," *Computers in Industry*, Vol.64, No.1, pp.57-67, 2013.
- [22] S. H. Choi, K. H. Han, and G. H. Lim, "Analysis of a Repair Processes Using a Process Mining Tool," *Journal of the Korea Contents Association*, Vol.13, No.4,

pp.399-406, 2013.

- [23] A. Sahraeidolatkhaneh and K. H. Han, "Integrated Framework of Process Mining and Simulation Approaches for the Efficient Diagnosis and Design of Business Process," Journal of the Korea Contents Association, Vol.17, No.5, pp.44-55, 2017.
- [24] D. R. Ferreira and E. Vasilyev, "Using Logical Decision Trees to Discover the Cause of Process Delays from Event Logs," Computers in Industry, Vol.70, pp.194-207, 2015.
- [25] W. M. Van der Aalst, M. H. Schonenberg, and M. Song, "Time Prediction Based on Process Mining. Information Systems, Vol.36, No.2, pp.450-475, 2011.
- [26] ProM Documentation (v6.11) https://www.promtools.org/doku.php?id=docs:start#prom_documentation, 2022.

저 자 소 개

한 관 희(Kwan-Hee Han)

정회원



- 1982년 2월 : 아주대학교 산업공학과(공학사)
- 1984년 2월 : KAIST 산업공학과(공학석사)
- 1996년 8월 : KAIST 자동화 및 설계공학과(공학박사)
- 2000년 3월 ~ 현재 : 경상국립대학교 산업시스템공학부 교수

〈관심분야〉 : 스마트팩토리, 비즈니스프로세스관리, 프로세스마이닝, Digital Manufacturing, Modeling & Simulation