

# Defining and Discovering Cardinalities of the Temporal Workcases from XES-based Workflow Logs<sup>☆</sup>

Jaeyoung Yun<sup>1</sup>

Hyun Ahn<sup>1</sup>

Kwanghoon Pio Kim<sup>1\*</sup>

## ABSTRACT

Workflow management system is a system that manages the workflow model which defines the process of work in reality. We can define the workflow process by sequencing jobs which is performed by the performers. Using the workflow management system, we can also analyze the flow of the process and revise it more efficiently. Many researches are focused on how to make the workflow process model more efficiently and manage it more easily. Recently, many researches use the workflow log files which are the execution history of the workflow process model performed by the workflow management system. Our research group has many interests in making useful knowledge from the workflow event logs. In this paper we use XES log files because there are many data using this format. This paper suggests what are the cardinalities of the temporal workcases and how to get them from the workflow event logs. Cardinalities of the temporal workcases are the occurrence pattern of critical elements in the workflow process. We discover instance cardinalities, activity cardinalities and organizational resource cardinalities from several XES-based workflow event logs and visualize them. The instance cardinality defines the occurrence of the workflow process instances, the activity cardinality defines the occurrence of the activities and the organizational cardinality defines the occurrence of the organizational resources. From them, we expect to get many useful knowledge such as a patterns of the control flow of the process, frequently executed events, frequently working performer and etc. In further, we even expect to predict the original process model by only using the workflow event logs.

✉ keyword : workflow, workflow discovery, workflow rediscovery, workcase cardinality, workflow execution log, XES

## 1. Introduction

Workflow management system is a system that manages the workflow which defines the process of work in reality. We can define the workflow process by sequencing jobs which is related and influencing to each other. Using the workflow management system, we can also analyze the flow of the process and revise it more efficiently. Many researches are focused on how to make the workflow process model more efficiently and manage it more easily. Recently, workflow process model is becoming larger and

larger as the process is more complex. Also the workflow management system (WFMS) becomes larger due to the growth of computing power and techniques to design it [1]. Thus, it is an issue that to analyze a new type of requirements and demands concerning about workflow intelligence and quality.

Our research group focuses on the knowledge discovered from the execution logs of workflow process models to improve the quality of them with high consistent and efficiency. We called this behavior - discovering new knowledge from the workflow execution logs - process mining. There are workflow process mining technology in various perspectives such as control-flow, data-flow, perform and organizational resource.

In this paper, we introduce the instance cardinality, activity cardinality and organizational resource cardinality. Instance cardinality is the occurrence of the process instances in the event log file. Activity cardinality is the occurrence of the activities in the process instances. And organizational resource cardinality is the occurrence of the organizational resources in the process instances. Each cardinality has their

<sup>1</sup> Div. of Computer Science and Engineering, Kyonggi University, Gyeonggi, 16227, Republic of Korea

\* Corresponding author: (kwang@kgu.ac.kr)

[Received 29 December 2018, Reviewed 23 January 2019 (R2 5 March 2019, R3 11 April 2019), Accepted 30 April 2019]

☆ Preliminary version of this paper was presented at APIC-IST 2018.

☆ This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (Grant No. 2017R1A2B2010697). This work also was partially supported by Kyonggi University's Graduate Research Assistantship 2019.

own viewpoint which is needed to the process manager who has to analyze entirely nasty, large process. More details about these cardinalities are in chapter 3. In chapter 4, we show the examples of these cardinalities discovered from the real event logs. We used XES-based event logs in our experiments.

## 2. Related Works

### 2.1 XES

XES[3][4][5] is a standard format to define the workflow event log which is an execution history of the process model. It is an XML-based text file. Basically, it has three part using the XML tags. Log, trace, event are those. Log is the top level component of the XES. Log defines the whole process which was performed before making the XES log file. It is a top level of the XES file. The XES file begins with '<log>' and ends with '</log>'. In the log, there are many traces which define the instances of workflow process - the log. Each trace defines the instance which is the execution of the process. In a trace, there are many events which defines the activities in the instance. Each event defines the execution of an activity which is the unit of job in the process. As an example, there is a process of handling insurance claims. The log defines the process - handling insurance claims. The trace defines one specific instance of the process - one specific insurance claim. The event defines the execution of a activity - recording the client's personal information in the database has been completed. Log, trace and event have their own attributes to define the information about themselves. Moreover, Each attribute contains information by key with value. Key is the domain type of attributes. Therefore, it is easy to define and represent the execution history by using XES format.

### 2.2 Process Mining

Process mining means to find useful information from the workflow models. Basically, there are two resources in making valuable information through workflow management, first is the workflow model[6][7][8][9][10][11][12][13][14][15] and second is the workflow log[16][17][18][19].

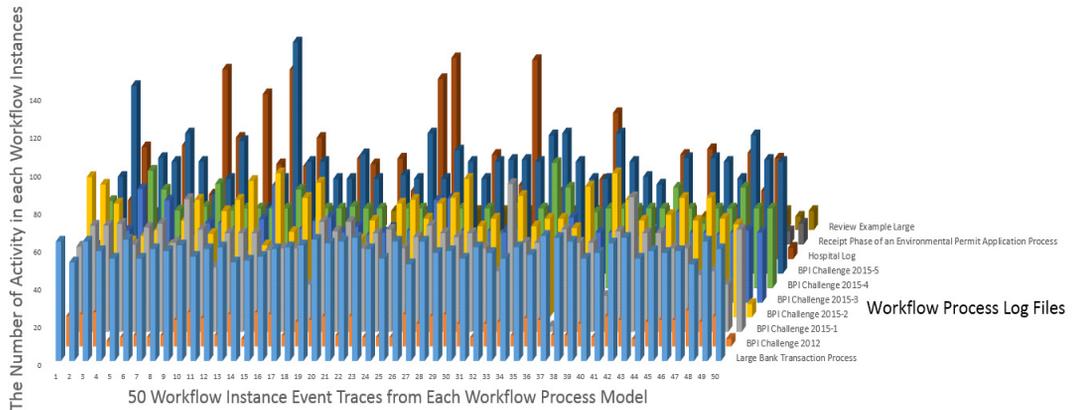
Recently, there are many interests in using workflow logs - we called it process mining - for example, there are some researches try to find out workflow models through workflow log and process mining from workflow log to get information. Process mining becomes important because it uses the real data which is execution data of the workflow process models. In this paper, we also use the workflow logs and we define and show some information got from the XES workflow logs i.e. our research is a kind of process mining.

## 3. Types of the Cardinalities

In this chapter, we introduce the definition of temporal workcases cardinalities using these information.

There are three temporal workcases cardinalities that can be discovered from XES-based workflow event logs and each cardinality has three types. Instance cardinality defines the occurrence of the workflow process instances which means the executed workflow process model. Instance cardinality can be divided into three types, instance cardinality per workflow model, instance class cardinality per workflow model and instance cardinality per instance class. Instance cardinality per workflow model defines how many times the workflow process instances occur in an execution of workflow process model. Instance class cardinality per workflow model defines how many types of the workflow process instances occur in an execution of workflow process model which is same as the redundancy removed instance cardinality per workflow model. Instance cardinality per instance class defines how many times the workflow control-paths occur in an execution of workflow process model. It means that what types of the activities in the workflow process instances without redundancy. The activity cardinality and the organizational resource cardinality each defines the occurrence of the activities, organizational resources. They also have three types each and their types have the same context as that in the instance cardinality. There is a definition of these temporal workcases cardinalities in [Definition 1].

[Definition 1] Temporal workcases cardinalities. A temporal workcases cardinalities that can be discovered from



(Fig. 1) A Workflow Warehouse of 50 Workflow Instances corresponding to 10 Workflow Processes

XES-based workflow process event logs.

#### Instance Cardinality

- Instance Cardinality per Workflow Model: The number of process instance for each workflow process model in the workflow execution log.
- Instance Class Cardinality per Workflow Model: The number of process instance's control-path for each workflow process model in the workflow execution log.
- Instance Cardinality per Instance Class: The number of process instance's control-path frequency for each workflow process model in the workflow execution log.

#### Activity Cardinality

- Activity Instance Cardinality per Workflow Model: The number of activity frequency that makes up each workflow process model in the workflow execution log.
- Activity Class Cardinality per Workflow Model: The number of activity that makes up each workflow process model in the workflow execution log.
- Activity Instance Cardinality per Instance per Workflow Model : The number of activity frequency that makes up each workflow instance in for each workflow process model in the workflow execution log.

#### Organizational resource Cardinality

- Organizational resource Instance Cardinality per

Workflow Model: The number of organizational resource frequency that makes up each workflow process model in the workflow execution log.

- Organizational resource Class Cardinality per Workflow Model: The number of organizational resource that makes up each workflow process model in the workflow execution log.
- Organizational resource Instance per Instance per Workflow Model: The number of organizational resource frequency that makes up each workflow instance in for each workflow process model in the workflow execution log.

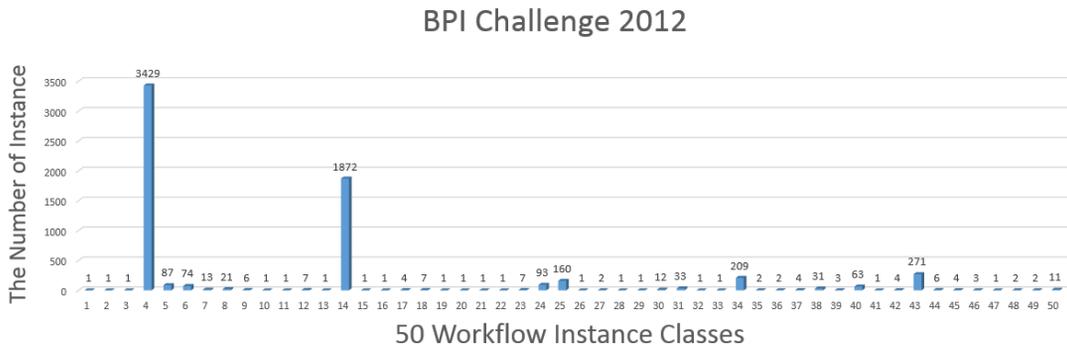
## 4. Experimental Results

We have made temporal workcases cardinalities from 10 XES event logs. BPI Challenge 2012, BPI Challenge 2015, Hospital Log, Large Bank Transaction Process, Receipt Phase of an Environmental Permit Application Process and Review Example Large are them. BPI Challenge 2012 is a log of loan application process, BPI Challenge 2015 is a log of building permit applications over a period of approximately four years. Other XES log files follow their names. Each log has its own attributes to contain the information of the process and has large amount of execution history.

Fig 1 shows a workflow warehouse of 50 workflow

(Table 1) Instance Cardinality and Instances Class Cardinality in 10 Workflow Process Logs

	BPI Challenge 2012	BPI Challenge 2015-1	BPI Challenge 2015-2	BPI Challenge 2015-3	BPI Challenge 2015-4	BPI Challenge 2015-5	Hospital Log	LBTP	RP-EPAP	Review Example Large
Instance Cardinality	13087	1199	832	1409	1053	1156	1143	10000	1434	10000
Instance Class Cardinality	4366	1170	828	1349	1049	1153	981	10000	116	4118



(Fig. 2) 50 Instance Cardinality per Instance Class in BPI Challenge 2012

instances discovered from the 10 workflow process log files. It shows how many events - execution of activities - in each of 50 workflow instances in the workflow log. In Fig 1, we can see that the large bank transaction process has about 60 events and in the hospital log there is a variation in number of events among the instances.

Table 1 shows the instance cardinality and instance class cardinality for each 10 workflow processes. The instance cardinality shows how many traces in the log file and the instance class cardinality shows how many types of control-path in the log file. In table 1, we can see that BPI Challenge 2012, RP-EPAP and Review Example Large have many duplicated control-path whereas LBTP has no duplicated control-path in other words, every control-path is different. Fig 2 shows the 50 instance cardinality per instance class in BPI Challenge 2012. In fig 2, we can see that the instance cardinality class number 4 and number 14 has many duplicated instances in BPI Challenge 2012 They takes up about 38% of total instances. With this kind of information - the instance cardinality - we can figure out

the proportion of loop cases or disjunctive cases.

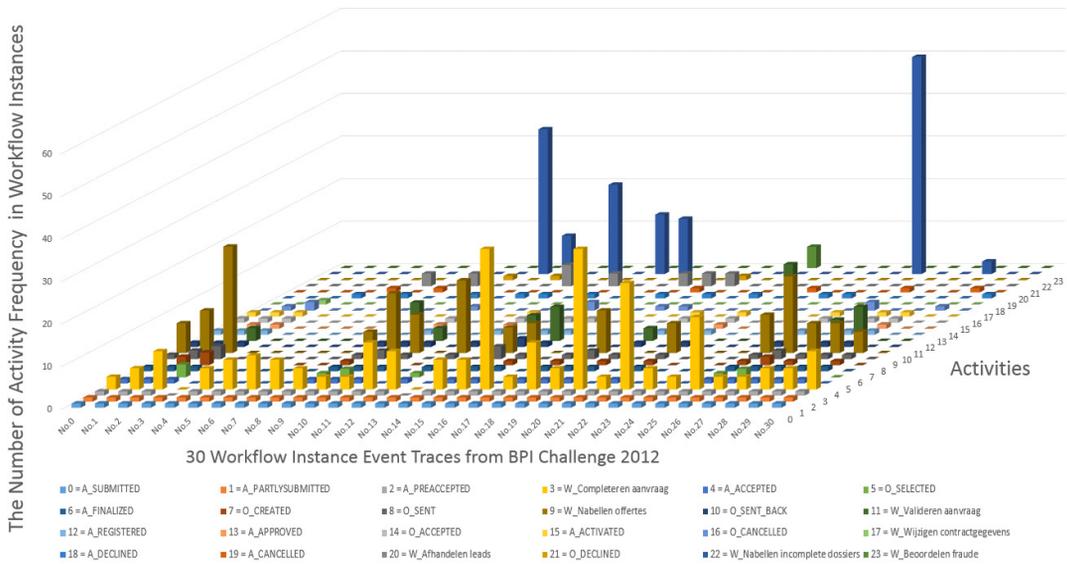
Table 2 shows the activity cardinality discovered from 10 workflow process log files. In Table 1 and 2, the logs with a small number of activity take a big proportion in duplicated instance class. However, in LBTP, there are small number of activity, it does not have any duplicated instance class. Fig 3 shows the activity instance cardinality in BPI Challenge 2012 log file. It shows the number of activity frequency of each activity. It is clear that the two activities which names are 'W\_Completeren aanvraag' and 'W\_Nabellenoffertes' were performed many times than other activities in BPI Challenge 2012. Fig 4 is an activity instance cardinality per instance of BPI Challenge 2012 log file. In Fig 4, we can see that there are many duplicated number of activities which are '3 = W\_Completeren aanvraag'(yellow), '9 = W\_Nabellen offertes'(brown) and '22 = W\_Nabellen incomplete dossiers'(dark-blue) among the 30 instances. 'No.25'instance even has activity 22 which is almost 50 times occur. With activity cardinality, we can get which activity is mostly executed and how many times of it occurs.

(Table 2) Activity Cardinality in 10 Workflow Process Logs

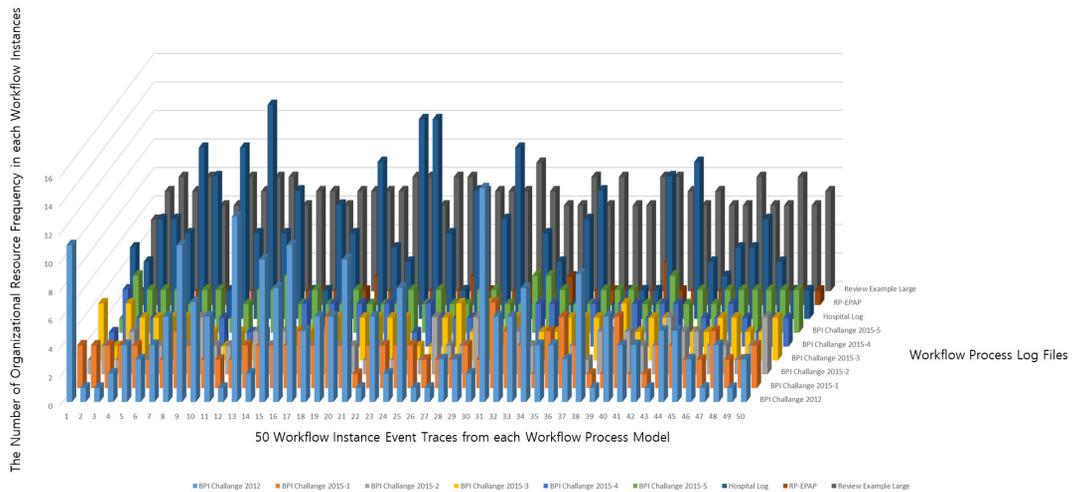
	BPI Challenge 2012	BPI Challenge 2015-1	BPI Challenge 2015-2	BPI Challenge 2015-3	BPI Challenge 2015-4	BPI Challenge 2015-5	Hospital Log	LBTP	RP-EPAP	Review Example Large
activity Cardinality	24	398	410	383	356	389	624	113	27	14



(Fig. 3) Activity Instance Cardinality in the BPI Challenge 2012



(Fig. 4) 30 Activity Instance Cardinality per Instance of BPI Challenge 2012



(Fig. 5) Organizational Resource Frequency of 50 Workflow Instances Corresponding to 9 Workflow Processes

With the activity cardinality, we can figure out which activity is mostly performed and that activity probably is the most important.

Fig 5 shows the number of classes of organizational resources in each event traces. In Fig. 5, there are 9 workflow process log files which are little different from the log files of the instance cardinality and activity cardinality. It's because that Large Bank Transaction Process is a simulated log, so its organizational resource information is set to 'null'. Thus we didn't use Large Bank Transaction Process log files and use 9 other workflow process log files. Using the organizational resource cardinality, we can see that how many organizations are related to each event trace and which organization is mostly participated in the process. In fig 5, we can see that there is a clear variance in organizational resources among the workflow instances in BPI Challenge 2012 and Hospital Log. In RP-EPAP, we can see that very small number of organizations are participated in the process. Using these organizational resource cardinality, we can figure out how many organizations are participated in the process and which is mostly working. Furthermore, with this information, we can analyze who is taking charge in the job and who is consuming the time in the job.

## 5. Conclusions and Future Work

In this paper, we defined the instance cardinality, activity cardinality and organizational resource cardinality i.e. temporal workcase cardinality. Instance cardinality defines the occurrence of the workflow process instances which means the executed workflow process model. The activity cardinality and the organizational resource cardinality each defines the occurrence of the activities, organizational resources. We visualized the result of the discovery of temporal workcases cardinalities from the workflow event logs. With these cardinalities, we expect to get many useful knowledge such as a patterns of the control flow of the process, frequently executed events, frequently working performer and etc. In further, we even expect to predict the original process model by only using the workflow event logs.

## References

- [1] Hyung-Jin Ahn, and Kwang-Hoon Kim, "Design and Implementation of a Very Large-Scale Workflow Management System," Journal of Internet Computing and Services, vol. 10, no. 6, pp. 205-218, 2009.

- [2] WfMC, "Business Process Analytics Format (BPAF)", Workflow Management Coalisation Workflow Standard Document Number WfMC-TC-1015, 2008.
- [3] IEEE, Draft Standard for XES - eXtensible Event Stream - for achieving interoperability in event logs and event streams.
- [4] Christian W. Günther and Eric Verbeek, "XES Standard Definition", Technische Universiteit Eindhoven University of Technology, Netherlands, 2014.
- [5] Acampora, Giovanni, et al., "IEEE 1849: The XES Standard: The Second IEEE Standard Sponsored by IEEE Computational Intelligence Society [Society Briefs]", IEEE Computational Intelligence Magazine, 12.2: 4-8, 2017.
- [6] Daniela Grigori, Fabio Casati, Malu Castellanos, Umeshwar Dayal, Mehmet Sayal and Ming-Chien Shan, "Business Process Intelligence," JOURNAL OF COMPUTERS IN INDUSTRY, Vol. 53, Issue 3, 2004.
- [7] Fabio Casati, et al, "Business Process Intelligence," Technical Report, HPL-2002-119, HP Laboratories Palo Alto, 2002.
- [8] Kwang-Hoon Kim and Clarence A. Ellis, "Workflow Reduction for Reachable-path Rediscovery in Workflow Mining," Series of Studies in Computational Intelligence: the Foundations and Novel Approaches in Data Mining, Vol. 9, pp.289-310, Springer, 2006.
- [9] Aalst, W.P.M., Alves de Medeiros, A.K., and Weijters, A.J.M.M., "Process Equivalence: Comparing Two Process Models Based on Observed Behavior", BPM2006, Lecture Notes in Computer Science, Vol. 4012, pp. 129-144, 2006.
- [10] Ahn, H. and Kim, K., "A Stochastic Activity-to-Performer Affiliation Binding Formalism in ICN-based Workflow Models," ICICI Express Letters, Vol. 9, No. 12, 2015.
- [11] Min-Joon Kim, Hyun Ahn and Minjae Park, "A GraphML-based Visualization Framework for Workflow-Performers' Closeness Centrality Measurements," KSII Transactions on Internet and Information Systems, vol. 9, no. 8, pp. 3216-3230, 2015. <https://doi.org/10.3837/tiis.2015.08.028>.
- [12] Min-Joon Kim, Hyun Ahn and Min-Jae Park, "A Theoretical Framework for Closeness Centralization Measurements in a Workflow-Supported Organization," KSII Transactions on Internet and Information Systems, vol. 9, no. 9, pp. 3611-3634, 2015. <https://doi.org/10.3837/tiis.2015.09.018>.
- [13] Haksung Kim, Hyun Ahn and Kwanghoon Pio Kim, "Modeling, Discovering, and Visualizing Workflow Performer-Role Affiliation Networking Knowledge," KSII Transactions on Internet and Information Systems, vol. 8, no. 2, pp. 689-706, 2014. <https://doi.org/10.3837/tiis.2014.02.022>.
- [14] Jawon Kim, Hyun Ahn, Minjae Park, Sangguen Kim and Kwanghoon Pio Kim, "An Estimated Closeness Centrality Ranking Algorithm and Its Performance Analysis in Large-Scale Workflow-supported Social Networks," KSII Transactions on Internet and Information Systems, vol. 10, no. 3, pp. 1454-1466, 2016. <https://doi.org/10.3837/tiis.2016.03.031>.
- [15] Daeyong Jung, Taeweon Suh, Heonchang Yu and JoonMin Gil, "A Workflow Scheduling Technique Using Genetic Algorithm in Spot Instance-Based Cloud," KSII Transactions on Internet and Information Systems, vol. 8, no. 9, pp. 3126-3145, 2014. <https://doi.org/10.3837/tiis.2014.09.010>.
- [16] Pham Dinhlam, Hyun Ahn, and Kwanghoon Pio Kim, "A Temporal Work Transference Discovery Algorithm and Experimental Results on XES-Formatted Workflow Logs", 2018.
- [17] Kim Sang-Bae, Kim Hak-Seong, and Paik Su-Ki, "A Workcase Mining Mechanism using Activity Dependency," Journal of Internet Computing and Services, vol. 4, no. 6, pp. 43-56, 2003.
- [18] Min Jun-Ki, Kim Kwang-Hoon, and Chung Jung-Su, "A Control Path Analysis Mechanism for Workflow Mining," Journal of Internet Computing and Services, vol. 7, no. 1, pp. 91-100, 2006.
- [19] Kyoungsook Kim, Moonsuk Yeon, Byeongsoo Jeong and Kwanghoon Kim, "A Conceptual Approach for Discovering Proportions of Disjunctive Routing Patterns in a Business Process Model," KSII Transactions on Internet and Information Systems, vol. 11, no. 2, pp. 1148-1161, 2017. <https://doi.org/10.3837/tiis.2017.02.030>.

● 저 자 소 개 ●



**Jaeyoung Yun**

2019 B.S. in Computer Science, Kyonggi University

2019 ~ Present, M.S. Student in Computer Science, Kyonggi University

Research Interests : Workflow systems, discovery control flow, process mining, deep learning.

E-mail : lastmore9@kgu.ac.kr



**Hyun Ahn**

2011 B.S. in Computer Science, Kyonggi University

2013 M.S. in Computer Science, Kyonggi University

2017 Ph.D. in Computer Science, Kyonggi University

2018~Present, Assistant Professor of the Dept. of Computer Science and Engineering at Kyonggi University

Research Interests : Business Process Management, business process intelligence, process mining.

E-mail : hahn@kgu.ac.kr



**Kwanghoon Pio Kim**

1984 B.S in Computer Science, Kyonggi University

1986 M.S. in Computer Science, Chungang University

1994 M.S. in Computer Science, University of Colorado at Boulder

1998 Ph.D. in Computer Science, University of Colorado at Boulder

1998 ~ Present, Professor of the Dept. of Computer Science and Engineering at Kyonggi University

Research Interests : CSCW, workflow systems, Business Process Management, process mining, enterprise social network analysis.

E-mail : kwang@kgu.ac.kr