

A Correspondence Analysis Framework for Workflow-supported Performer-Activity Affiliation Networks

Hyun Ahn*, Chungun Park**, Kwanghoon Pio Kim*

*Dept. of Computer Science, KYONGGI UNIVERSITY, Suwonsi Kyonggido, Korea

**Dept. of Mathematics, KYONGGI UNIVERSITY, Suwonsi Kyonggido, Korea

{hahn, cgpark, kwang}@kgu.ac.kr

Abstract—The workflow-supported affiliation network is a special type of the workflow-supported social network, and represents affiliated relationships between two different entity types in workflow models. In this paper, we propose a theoretical framework for particularly analyzing the performer-activity correspondence measurements on a workflow-supported affiliation network. The proposed framework formalizes a series of theoretical steps from discovering a workflow-supported performer-activity affiliation network to analyzing the performer-activity association on the discovered affiliation network using correspondence analysis. As an operational example, we apply a typical example of the information control net workflow model to the steps of the procedural framework and show its results of the performer-activity correspondence analysis.

Keywords: workflow-supported affiliation network, correspondence analysis, workflow intelligence, ICN-based workflow model, organizational knowledge

I. INTRODUCTION

Until a few years ago, it may be fairly said that the research and development efforts in the workflow literature have focused on defining and implementing the basic functionality and services of workflow and business process automation technologies. However, recently the workflow literature[1][2] brings into being interested in “social networking knowledge.” It begins from the strong belief that social relationships and collaborative behaviors among people who are involved in enacting workflow models affect the overall performance in the real businesses and the working productivities, as well, over workflow-supported enterprises[3]. Therefore, the authors research group has been doing research on applying the concept of social networking services and its analysis methods to workflows named workflow-supported social network[4] and workflow-supported affiliation network[5]. Particularly, this paper is concerned with a workflow-supported affiliation network represents affiliations between two entity types in workflow models.

In this paper, we propose a theoretical framework for particularly analyzing the performer-activity correspondence measurements on a workflow-supported affiliation network. A little more precisely speaking, the proposed framework formalizes a series of theoretical steps from discovering a workflow-supported performer-activity affiliation network to

analyzing the performer-activity association on the discovered affiliation network using correspondence analysis (CA) which is a powerful multivariate statistical method that provides a means of displaying or summarizing a set of data in two-dimensional graphical form. Ultimately, through the calculation steps of CA in the framework, we can explore associations between two types of nodes in a workflow-supported affiliation network.

II. FORMAL DEFINITION OF WORKFLOW-SUPPORTED AFFILIATION NETWORKS

This section shortly introduces the basic concept of ICN-based workflow model[6] as a technological background and then describes a performer-activity affiliation network model as an example of formal definition of workflow-supported affiliation networks.

A. Information Control Net

An ICN-based workflow model can be defined by capturing the notations of workflow procedures, activities and their control precedence, invoked applications, roles, performers, and input/output repositories, as shown in Figure 1. In this subsection, we define the basic concept of workflow model with respect to the formal descriptions of ICN-based workflow model.

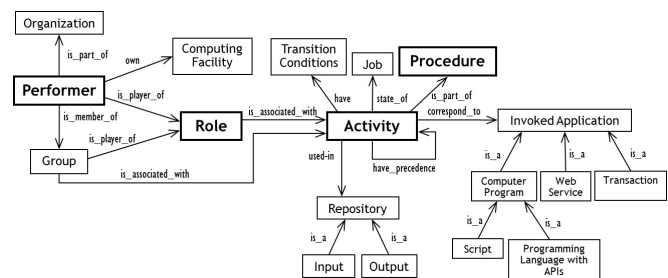


Fig. 1. The workflow meta-model

The following [Definition 1] is a formal definition of ICN-based workflow model, and its functional components to be used for retrieving workflow-related information, such as

activity precedence (control flow), activity-role association, activity-relevant data association (data flow), activity-invoked application association, activity-transition condition association, and role-performer association information. Based upon these types of information, it is possible to retrieve several types of derived workflow-related information like activity-performer association, relevant data-invoked application association, role complexity, performer complexity information, and so forth.

[Definition 1] Information Control Net (ICN) for formally defining workflow model. A basic **ICN** is 9-tuple $\Gamma = (\delta, \rho, \gamma, \lambda, \varepsilon, \pi, \kappa, \mathbf{I}, \mathbf{O})$ over a set of **A** activities (including a set of group activities), a set of $\mathbf{E} \subseteq (\mathbf{A} \times \mathbf{A})$ edges (pairs of activities), a set **T** of transition conditions, a set **R** of repositories, a set **G** of invoked application programs, a set **P** of roles, and a set **C** of performers (including a set of performer groups), where, $\wp(\mathbf{A})$ represents a power set of the activity set, **A**:

- **I** is a finite set of initial input repositories, assumed to be loaded with information by some external process before execution of the ICN;
- **O** is a finite set of final output repositories, perhaps containing information used by some external process after execution of the ICN;
- $\delta = \delta_i \cup \delta_o$
where, $\delta_o : \mathbf{A} \rightarrow \wp(\mathbf{A})$ is a multi-valued mapping function from an activity to its sets of (immediate) successors, and $\delta_i : \mathbf{A} \rightarrow \wp(\mathbf{A})$ is a multi-valued mapping function from an activity to its sets of (immediate) predecessors;
- $\rho = \rho_i \cup \rho_o$
where $\rho_o : \mathbf{A} \rightarrow \wp(\mathbf{R})$ is a single-valued mapping function from an activity to its set of output repositories, and $\rho_i : \mathbf{A} \rightarrow \wp(\mathbf{R})$ is a single-valued mapping function from an activity to its set of input repositories;
- $\gamma = \gamma_i \cup \gamma_o$
where $\gamma_o : \mathbf{R} \rightarrow \wp(\mathbf{A})$ is a single-valued mapping function from a repository to its set of outdegree activities, and $\gamma_i : \mathbf{R} \rightarrow \wp(\mathbf{A})$ is a single-valued mapping function from a repository to its set of indegree activities;
- $\lambda = \lambda_a \cup \lambda_g$
where $\lambda_g : \mathbf{A} \rightarrow \mathbf{G}$ is a single-valued mapping function from an activity to its invoked application program, and $\lambda_a : \mathbf{G} \rightarrow \wp(\mathbf{A})$ is a single-valued mapping function from an invoked application program to its set of associated activities;
- $\varepsilon = \varepsilon_a \cup \varepsilon_p$
where $\varepsilon_p : \mathbf{A} \rightarrow \mathbf{P}$ is a single-valued mapping function from an activity to a role, and $\varepsilon_a : \mathbf{P} \rightarrow \wp(\mathbf{A})$ is a single-valued mapping function from a role to its set of associated activities;
- $\pi = \pi_p \cup \pi_c$
where, $\pi_c : \mathbf{P} \rightarrow \wp(\mathbf{C})$ is a single-valued mapping function from a role to its set of associated performers, and $\pi_p : \mathbf{C} \rightarrow \wp(\mathbf{P})$ is a single-valued mapping function from a performer to its set of associated roles;
- $\kappa = \kappa_i \cup \kappa_o$
where $\kappa_i : \mathbf{E} \rightarrow \wp(\mathbf{T})$ is a single-valued mapping func-

tion from an edge to a set of control-transition conditions; and $\kappa_o : \mathbf{T} \rightarrow \wp(\mathbf{E})$ is a single-valued mapping function from a control-transition condition to a set of edges.

B. The Formal Definition of Performer-Activity Affiliation Network Model

In order to represent the workflow-supported performer-activity affiliation networking knowledge, the paper formally defines the performer-activity affiliation network model which is abbreviated as PAANM, and consists of two types of nodes—a set of performers and a set of activities—and a set of relations between these two nodal types. Thus, the performer-activity affiliation network is a two-mode network, through which it used to accomplish the following dual objectives:

- to uncover the relational structures of performers through their joint involvement in activities.
- to reveal the relational structures of activities through their joint participation of common performers.

Additionally, those relational structures can be weighed to measure the extent of their strengths by assigning a value to each of relations between nodal types. Therefore, there are two types of performer-activity affiliation networks—binary performer-activity affiliation network and valued performer-activity affiliation network. In the binary performer-activity affiliation network, its value (0 or 1) implies a binary relationship of involvement (or participation), while values in the valued performer-activity affiliation network may represent various implications according to their application domains; typical examples of values might be stochastic (or probabilistic) values, strengths, and frequencies. The formal knowledge representation of PAANM is defined in [Definition 2].

[Definition 2] Performer-Activity Affiliation Network Model (PAANM). The performer-activity affiliation network model is formally defined as a 3-tuple formula, $\Lambda = (\sigma, \psi, \mathbf{S})$, over a set **C** of performers, a set **A** of activities, a set **V** of weight-values, a set $\mathbf{E}_p \subseteq (\mathbf{C} \times \mathbf{A})$ of edges (pairs of performers and activities), and a set $\mathbf{E}_a \subseteq (\mathbf{A} \times \mathbf{C})$ of edges (pairs of activities and performers), where, $\wp(\mathbf{C})$ and $\wp(\mathbf{A})$ represent power sets of the performer set **C** and the activity set **A**, respectively:

- **S** is a finite set of work-sharing performers or groups of some external performer-activity affiliation network models;
- $\sigma = \sigma_p \cup \sigma_v$ /* Involvement Knowledge */
where, $\sigma_p : \mathbf{C} \rightarrow \wp(\mathbf{A})$ is a single-valued mapping function from a performer to its set of involved activities; $\sigma_v : \mathbf{E}_p \rightarrow \mathbf{V}$ is a single-valued mapping function from an edge ($\in \mathbf{E}_p$) to its weight-value;
- $\psi = \psi_a \cup \psi_v$ /* Participation Knowledge */
where, $\psi_a : \mathbf{A} \rightarrow \wp(\mathbf{C})$ is a single-valued mapping function from a activity to a set of participated performers; and $\psi_v : \mathbf{E}_a \rightarrow \mathbf{V}$ is a single-valued function from an edge ($\in \mathbf{E}_a$) to its weight-value;

And the graphical knowledge representation is depicted by a bipartite graph[7]. So, the performer-activity affiliation

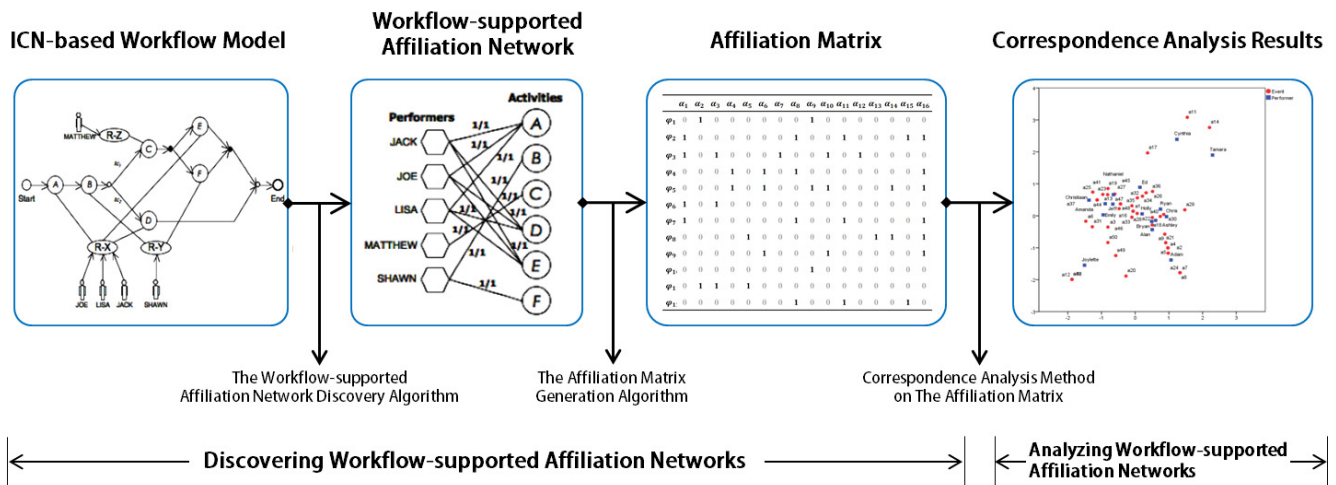


Fig. 2. Discovery and analysis phases of the framework

network's graphical model consists of two types of graphical nodes—a set of performers (shaped in hexagon) and a set of activities (shaped in circle)—and a set of non-directed edges between these two nodal types, which means that a workflow-supported affiliation network is a non-directed graph.

III. THE PROPOSED FRAMEWORK

In this section, we give a general description of the proposed framework consist of the discovery and analysis phases. In the discovery phase, we describe how to automatically discover a PAANM from a workflow procedure defined in the ICN-based workflow model, and automatically generate an affiliation matrix from the discovered PAANM. And in the analysis phase, based upon the affiliation matrix, we are able to analyzing associations between the row and column values of the affiliation matrix through the calculation steps of CA. Figure 2 shows a view of the whole proposed framework and its conceptual phases with a series of associated algorithms.

A. Discovering Workflow-supported Affiliation Networks

At this moment, it is important to emphasize that workflow-supported affiliation networks will not be modeled or designed but automatically discovered from workflow procedures. So, we devised the algorithmic discovery methodology to automatically discover a PAANM by exploring the internal social perspective of an ICN-based workflow model in our previous work[8]. Figure 3 shows a example of ICN-based workflow model, and as a result for the discovery algorithm, a graphical representation of the PAANM discovered from the ICN-based workflow model of Figure 3, which is shown in Figure 4.

To apply CA to the discovered PAANM, the model with g performers and h activities should be transformed into an affiliation matrix with 2-dimension of $g \times h$. To this end, the algorithm for generating an affiliation matrix for workflow-supported performer-activity affiliation networks was devised in [9]. If a certain performer ϕ_i attends an activity α_j , then the entry in the i th and j th cell in the matrix equals to 1; otherwise the entry is 0. Denoting a binary performer-activity

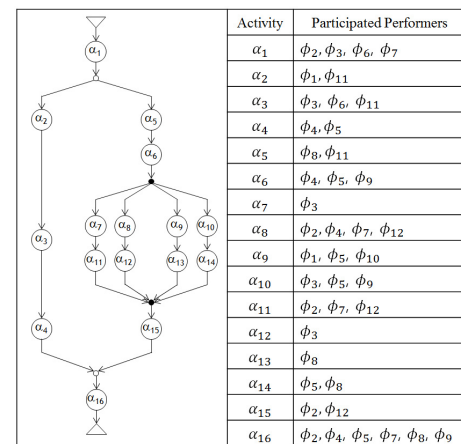


Fig. 3. Graphical representation of the ICN-based workflow model

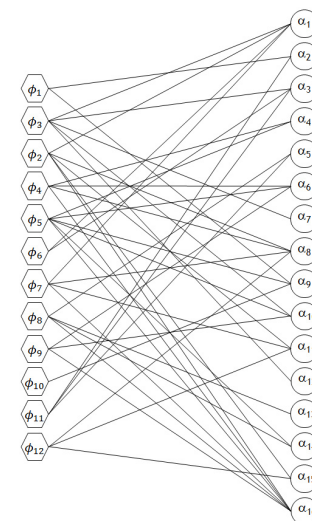


Fig. 4. Graphical representation (bipartite graph) of the discovered PAANM

	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	α_{11}	α_{12}	α_{13}	α_{14}	α_{15}	α_{16}	p_{i+}
ϕ_1	-	-	.0238	-	-	-	-	-	.0238	-	-	-	-	-	-	-	.0476
ϕ_2	.0238	-	-	-	-	-	-	.0238	-	-	.0238	-	-	-	.0238	.0238	.1190
ϕ_3	.0238	-	.0238	-	-	-	.0238	-	-	.0238	-	.0238	-	-	-	-	.1190
ϕ_4	-	-	-	.0238	-	.0238	-	.0238	-	-	-	-	-	-	-	.0238	.0952
ϕ_5	-	-	-	.0238	-	.0238	-	-	.0238	.0238	-	-	-	.0238	-	.0238	.1428
ϕ_6	.0238	-	.0238	-	-	-	-	-	-	-	-	-	-	-	-	-	.0476
ϕ_7	.0238	-	-	-	-	-	-	.0238	-	-	.0238	-	-	-	-	.0238	.0952
ϕ_8	-	-	-	-	.0238	-	-	-	-	-	-	-	.0238	.0238	-	.0238	.0952
ϕ_9	-	-	-	-	-	.0238	-	-	-	.0238	-	-	-	-	-	.0238	.0714
ϕ_{10}	-	-	-	-	-	-	-	-	.0238	-	-	-	-	-	-	-	.0238
ϕ_{11}	-	.0238	.0238	-	.0238	-	-	-	-	-	-	-	-	-	-	-	.0714
ϕ_{12}	-	-	-	-	-	-	-	.0238	-	-	.0238	-	-	-	.0238	-	.0714
p_{+j}	.0952	.0476	.0714	.0476	.0476	.0714	.0238	.0952	.0714	.0714	.0714	.0238	.0238	.0476	.0476	.1428	

Fig. 5. The correspondence matrix \mathbf{P} for the performer-activity affiliation matrix \mathbf{X}

affiliation matrix as $\mathbf{X} = [x_{ij}]_{g \times h}$, its x_{ij} values meet these conditions:

$$x_{ij} = \begin{cases} 1 & \text{if performer } \phi_i \text{ is affiliated with activity } \alpha_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

B. Correspondence Analysis on the PAANM

CA is a multivariate statistical method to display a multidimensional representation of the association between the row and column categories of a two-way contingency table. Categories that are similar to each other appear close to each other in the plots. Therefore, the main purpose of the framework is to explore associations between performers and activities in ICN-based workflow models using CA. The calculation steps of CA on the PAANM were as follows.

1) **Correspondence Matrix:** We derive the correspondence matrix $\mathbf{P} = [p_{ij}]_{g \times h}$ from $\mathbf{X} = [x_{ij}]_{g \times h}$ by dividing its entries by their grand total n .

$$n = \sum_{i=1}^g \sum_{h=1}^h x_{ij} \quad (2)$$

$$p_{ij} = \frac{x_{ij}}{n} \quad (3)$$

Next, row and column masses can be defined as follows:

$$p_{i+} = \sum_{j=1}^h p_{ij} \quad (4)$$

$$p_{+j} = \sum_{i=1}^g p_{ij} \quad (5)$$

The row and column profile can be regarded as a mathematical vector which may be represented as a point in respective h

and g dimensional weighted Euclidean space. Each row profile r_i (or column profile c_j) consist of elements of correspondence matrix \mathbf{P} divided by their respective row masses p_{i+} (or column masses p_{+j}), which can be written as:

$$r_i = (p_{i1}/p_{i+}, \dots, p_{ih}/p_{i+}) \quad (6)$$

$$c_j = (p_{1j}/p_{+j}, \dots, p_{gj}/p_{+j}) \quad (7)$$

Figure 5 represents the correspondence matrix \mathbf{P} calculated from the performer-activity affiliation matrix \mathbf{X} of the PAANM that shown in Figure 4.

2) **Euclidean Distances:** Based on the profiles, coordinate distances among the profiles each other, are calculated by Euclidean distance method. The following equations 8 and 9 are for calculating Euclidean distances between the row profiles or column profiles.

$$d(r_a, r_b) = \sqrt{\sum_{j=1}^h \left(\frac{p_{aj}}{p_{a+}} - \frac{p_{bj}}{p_{b+}} \right)^2 / p_{+j}} \quad (8)$$

$$d(c_a, c_b) = \sqrt{\sum_{i=1}^g \left(\frac{p_{ia}}{p_{+a}} - \frac{p_{ib}}{p_{+b}} \right)^2 / p_{i+}} \quad (9)$$

Eventually, row and column profiles, Euclidean distance measures represented by 2-dimensional bi-plot or scatter plot through a mathematical step of singular value decomposition[7]. The result of CA of PAANM example is shown in Figure 6.

As you see, if a pair of performer and activity have a affiliation between them, their points in the CA plot are close together. Furthermore, The larger the number of joint participations (or involvements) between entities of same type,

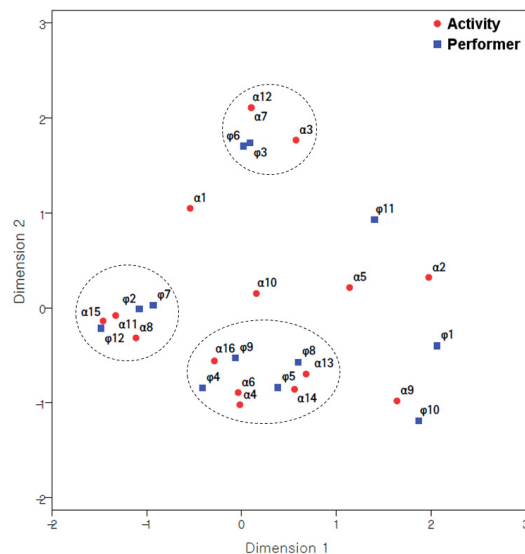


Fig. 6. The correspondence analysis result for the PAANM

Their points also tend to be close to each other. On figure 6, for example, performers $\phi 2$ and $\phi 7$, their points are closely located because they are jointly attended in three activities ($\alpha 1$, $\alpha 8$, $\alpha 11$).

IV. CONCLUSION

In this paper, we proposed the theoretical framework for the correspondence analysis on a workflow-supported affiliation network representing performer-activity affiliations in a workflow model. In order to verify the framework, we applied the devised algorithms to the typical example of the hiring workflow model, as an operational example. Based upon the proposed framework, we are able to implement a special workflow intelligence system that is able to quantitatively measure the correspondence relationships between performers and activities of workflow-supported affiliation networks.

ACKNOWLEDGEMENTS

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (Grant No. 2012006971), Republic of Korea.

REFERENCES

- [1] M. S. Wil M. P. van der Aalst, Hajo A. Reijers, "Discovering social networks from event logs," *COMPUTER SUPPORTED COOPERATIVE WORK*, vol. 14, no. 6, pp. 549–593, 2005.
- [2] e. a. Harri Oinas-Kukkonen, "Social networks and information systems: Ongoing and future research streams," *JOURNAL OF THE ASSOCIATION OF INFORMATION SYSTEMS*, vol. 14, no. 2, pp. 61–68, 2010.
- [3] J. Won, *A Framework: Organizational Network Discovery on Workflows*. Graduate School of Kyonggi University, 2008.
- [4] e. a. Jihye Song, "A framework: Workflow-based social network discovery and analysis," in *Proceedings of the 4th International Workshop on Workflow Management in Service and Cloud Computing*. IEEE, December 2010.
- [5] K. P. K. Hyun Ahn, Cheongeon Park, "A bpm performer-activity correspondence analysis method," *JOURNAL OF KOREAN SOCIETY FOR INTERNET INFORMATION*, vol. 14, no. 4, pp. 64–72, August 2013.

- [6] K. P. Kim and C. A. Ellis, *Section II / Chapter VII. An ICN-based Workflow Model and Its Advances, Handbook of Research on BP Modeling*. IGI Global, ISR, 2009.
- [7] S. Y. David Knoke, *SOCIAL NETWORK ANALYSIS - 2nd Edition, Series: Quantitative Applications in the Social Sciences*. SAGE Publications, 2008.
- [8] K. P. K. Sungjoo Park, "A closeness centrality analysis algorithm for workflow-supported social networks," *JOURNAL OF KOREAN SOCIETY FOR INTERNET INFORMATION*, vol. 14, no. 5, pp. 77–85, October 2013.
- [9] e. a. Alyeksandr Battsetseg, "Organizational closeness centrality analysis on workflow-supported activity-performer affiliation networks," in *Proceedings of the 15th International Conference on Advanced Communications Technology*. IEEE, January 2013, pp. 154–157.



Hyun Ahn Hyun Ahn is a full-time Ph.D. student of computer science department and a graduate member of the collaboration technology research laboratory at Kyonggi University, South Korea. He received B.S. and M.S. degrees in computer science from Kyonggi University in 2010 and 2012, respectively. His research interests include workflow systems, BPM, scientific workflow systems, workflow-supported social and affiliation networks discovery, analysis, and visualization.



Chungun Park Chungun Park is an assistant professor of mathematics department at Kyonggi University, South Korea. He received B.S. in mathematics from Kyonggi University, M.S. in Statistics from Chungang University, and Ph.D. in Statistics from the Texas A&M University, in 1993, 1995, and 2003, respectively. His research interests include mathematics, statistics, Bayesian estimations & approaches, big data analysis, and wavelet.



Kwanghoon Pio Kim Kwanghoon Pio Kim is a full professor of computer science department and the founder and supervisor of the collaboration technology research laboratory at Kyonggi University, South Korea. He received B.S. degree in computer science from Kyonggi University in 1984. And he received M.S. degree in computer science from Chungang University in 1986. He also received his M.S. and Ph.D. degrees from the computer science department at University of Colorado Boulder, in 1994 and 1998, respectively. He had worked as researcher and developer at Aztek Engineering, American Educational Products Inc., and IBM in USA, as well as at Electronics and Telecommunications Research Institute (ETRI) in South Korea. In present, he is a vice-chair of the BPM Korea Forum. He has been in charge of a country-chair (Korea) and ERC vice-chair of the Workflow Management Coalition. He has also been on the editorial board of the journal of KSII, and the committee member of the several conferences and workshops. His research interests include groupware, workflow systems, BPM, CSCW, collaboration theory, Grid/P2P distributed systems, process warehousing and mining, workflow-supported social networks discovery and analysis, process-aware information systems, data intensive workflows, and process-driven Internet of Things.