

프로세스 마이닝과 리엔지니어링을 위한 제어경로 기반 프로세스 그룹 발견 프레임워크와 실험적 검증[☆]

Control-Path Driven Process-Group Discovery Framework and its Experimental Validation for Process Mining and Reengineering

응웬 탄 하이¹ 김 광 훈^{2*}
Thanh Hai Nguyen Kwanghoon Pio Kim

요 약

본 논문에서는 비즈니스 프로세스 모델의 생명주기관리를 지원하는 대표적인 지식발견기술인 프로세스 마이닝과 지식개선기술인 프로세스 리엔지니어링 접근방법을 기반으로 하는 새로운 유형의 프로세스 발견 프레임워크를 제안한다. 또한, 제안된 프레임워크를 기반으로 하는 프로세스 마이닝 시스템을 개발하고, 이를 통한 실험적 검증을 수행한다. 실험적 효과검증에 적용된 프로세스 실행 이벤트 로그를 특별히 프로세스 빅-로그(Process BIG-Logs)라고 정의하고, 분산 비즈니스 프로세스 관리 시스템의 로깅메커니즘과 연계된 조각-실행로그이력들을 클러스터링하는 전처리과정을 거친 마이닝의 임팩테이터셋으로 활용한다. 결과적으로, 본 논문에서는 구조적 정보제어넷기반 프로세스 마이닝 알고리즘인 -알고리즘을 개선한 제어경로기반 프로세스 그룹 발견 알고리즘과 프레임워크를 설계 및 구현하고, 구현된 시스템을 이용하여 제안한 알고리즘과 프레임워크의 정확성을 실험적으로 검증한다.

☞ 주제어 : 비즈니스 프로세스 모델, 정보제어넷, 분산 프로세스 관리 시스템, 프로세스 인텔리전스, 프로세스 제어경로 지식발견, 프로세스 마이닝, 프로세스 리엔지니어링

ABSTRACT

In this paper, we propose a new type of process discovery framework, which is named as control-path-driven process group discovery framework, to be used for process mining and process reengineering in supporting life-cycle management of business process models. In addition, we develop a process mining system based on the proposed framework and perform experimental verification through it. The process execution event logs applied to the experimental effectiveness and verification are specially defined as Process BIG-Logs, and we use it as the input datasets for the proposed discovery framework. As an eventual goal of this paper, we design and implement a control path-driven process group discovery algorithm and framework that is improved from the -algorithm, and we try to verify the functional correctness of the proposed algorithm and framework by using the implemented system with a BIG-Log dataset. Note that all the process mining algorithm, framework, and system developed in this paper are based on the structural information control net process modeling methodology.

☞ keyword : Business Process Models, Information Control Nets, Distributed Process Management Systems, Process Intelligence, Process Control-Path Knowledge Discovery, Process Mining, Process Reengineering

1. 서 론

분산 및 클라우드 컴퓨팅 환경을 기반으로 하는 비피엠 (BPM: Business Process Management) 중심 엔터프라이즈에서는 비즈니스 프로세스(이하 프로세스) 모델들의 효과적이고 효율적인 운용관리목표를 달성하기 위하여 탁월한 성능의 분산형 비피엠 시스템 뿐만 아니라 프로세스 인텔리전스 및 생명주기 관리체계를 구축할 필요가 있다. 특히, 프로세스 인텔리전스 관리체계의 핵심기술로서 최근에 많은 연구개발 관심이 집중되는 데이터 및 지식 발견 기술은 프로세스 마이닝 기술이며, 프로세스 생

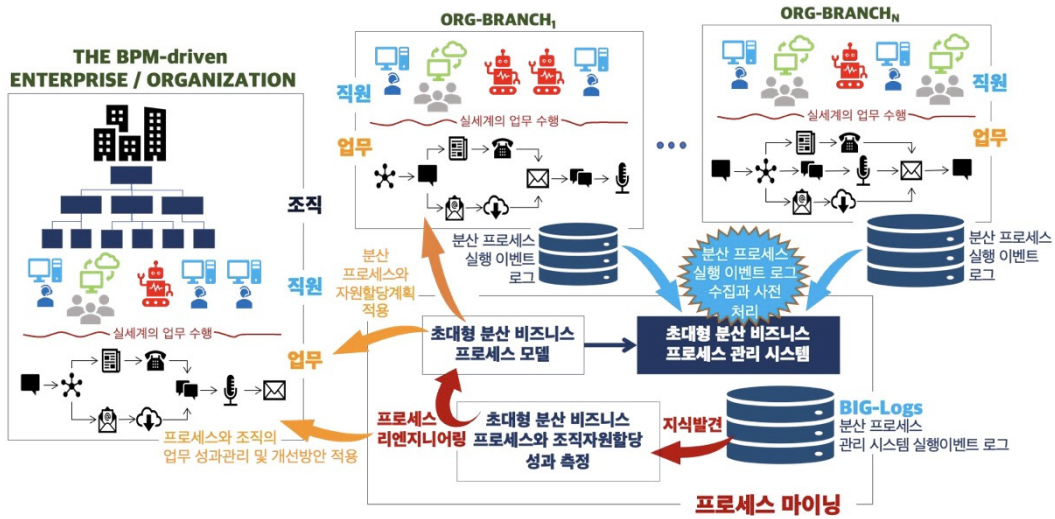
¹ Department of Computer Science, Graduate School, Kyonggi University, Suwon Gyeonggi, 16227, Republic of Korea

² Division of AI Computer Science and Engineering, Kyonggi University, Suwon Gyeonggi, 16227, Republic of Korea

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

[Received 18 August 2023, Reviewed 22 August 2023, Accepted 13 September 2023]

[☆] 본 연구는 한국연구재단 기초연구사업-중견연구자지원사업의 지원을 받아 수행되었음 (No. NRF-2022R1A2C2093002). 추가적으로, 본 연구는 2021년도 경기대학교 일반대학원 연구원장학생 장학금 지원에 의하여 수행되었음.



(그림 1) 비피엠 기반 엔터프라이즈의 프로세스 마이닝과 프로세스 리엔지니어링 개념도
 (Figure 1) A Situational View of the BPM-driven Enterprise with Process Mining and Reengineering

명주기 관리체계의 핵심기술은 프로세스 마이닝 및 지식 발견 기술을 기반으로 지속가능한 생명주기관리를 지원하는 프로세스 리엔지니어링 기술이다. 따라서, 본 논문에서는 비피엠 중심 엔터프라이즈의 핵심적인 컴퓨팅 인프라인 분산 프로세스 관리 시스템에 의해서 운영관리되는 일련의 초대형 분산 프로세스 모델들의 실행 로그로부터 제어흐름 중심의 지식을 발견하는 새로운 접근방법의 프로세스 마이닝 알고리즘과 프레임워크를 제안하고, 이렇게 제안된 새로운 유형의 접근방법을 이론적 기반으로 하는 소위 제어경로 기반 프로세스 마이닝 시스템을 개발함으로써 궁극적으로 비피엠 중심 엔터프라이즈를 구성하는 모든 프로세스 모델들의 생명주기를 효과적·효율적으로 관리할 수 있는 프로세스 마이닝 및 리엔지니어링 관리체계를 구축하고자 한다.

엔터프라이즈 비즈니스 프로세스 또는 워크플로우 모델과 그의 관리 시스템인 비피엠 기술의 등장과 발전이 거의 30여년 이상이 지속되면서 그 이후 많은 조직들이 엔터프라이즈 업무처리 프로세스 자동화 기술의 도입과 적용을 성공적으로 완성시켰고, 기술의 확산속도 역시 지속적으로 증가해 왔다. 또한, 분산 및 클라우드 컴퓨팅 환경의 확산과 더불어 비피엠 중심 엔터프라이즈 및 조직 역시 업무처리 프로세스 모델의 초대형화 및 분산화가 지속적으로 증가해 왔다. 결과적으로, 이와 같은 확산

일로의 비피엠 중심 엔터프라이즈에서 마침내 해당 업무 처리 프로세스 모델들에 대한 재설계와 리엔지니어링의 필요성과 자원할당정책의 재계획과 재배치의 긴급성이 대두되는 시점이 도래한 것이다. 즉 다시 말해서, 프로세스 실행 이벤트로그 및 이력데이터로부터의 지식발견과 재설계를 지원하는 프로세스 마이닝 기술과 프로세스 리엔지니어링 기술에 대한 연구개발의 시대가 시작된 것이다. 특히, 조직의 대형화와 컴퓨팅 환경의 분산화 그리고 그에 따른 업무처리 프로세스 모델들의 초대형화는 분산 및 클라우드 컴퓨팅 환경을 기반으로 하는 분산 프로세스 관리 시스템의 운용 관리체계와 프로세스 리엔지니어링 관리체계 역시 차원 높은 수준의 프로세스 마이닝 접근방법과 기술적 우수성을 요구하고 있다. 그림 1은 이러한 분산 및 초대형 비피엠 중심 엔터프라이즈 및 조직에서의 분산 프로세스 모델들과 그의 분산 실행 이벤트 로그 및 이력데이터 획득체계를 배경으로 하는 분산 프로세스 관리 시스템과 프로세스 마이닝 및 리엔지니어링 기술의 개념과 범위를 그림으로 나타낸 것이다. 결과적으로, 본 논문의 연구목표와 범위는 분산 프로세스 모델의 실행 이벤트 로그 및 이력데이터 수집과 전처리 기법 그리고 새로운 유형의 프로세스 지식발견 접근방법과 이를 위한 일련의 알고리즘 및 프레임워크를 제안하는데 있으며, 특별히 제안한 접근방법과 프레임워크의 개념적

우수성을 검증하기 위하여 일명 제어경로 기반 프로세스 마이닝 시스템을 개발하고 실험적 검증을 수행한다.

2. 관련 연구

이 장에서는 분산 및 클라우드 컴퓨팅 환경 기반의 분산 프로세스 관리 시스템과 초대형 분산 프로세스 모델들에 대한 운용기간이 장기간 지속됨에 따라 수반되는 프로세스 실행 이벤트로그 빅데이터[6][7][8]로부터 제어 흐름 기반의 프로세스 모델 자체 뿐 만 아니라 조직의 자원관리 의사결정에 활용될 핵심적인 경영관리 지식을 발견·재발견하는 프로세스 마이닝 알고리즘[1][2][3][5][9]들과 프레임워크 및 시스템들에 대한 기존의 연구성과들을 조사분석하고, 본 논문의 핵심 연구주제인 제어경로 기반 프로세스 그룹 발견 알고리즘 및 프레임워크와의 개념적 차별성과 구체적 구현범위를 소개한다. 특히, K. Kim [10]의 연구성과인 초대형 프로세스 모델의 종적 또는 횡적 조각화 방법을 적용한 분산 프로세스 관리 엔진들로부터 실행이벤트로그들을 수집저장하고, 수집된 초대형 이벤트로그 빅데이터상의 프로세스 실행이벤트 트레이스들을 제어경로 특성에 따른 클러스터링과 그룹핑하며, 궁극적으로 이러한 제어경로 특성을 반영한 프로세스 마이닝 알고리즘과 프레임워크를 제안하는 것이 본 논문의 핵심 목표이자 연구범위이다.

프로세스 마이닝 기술의 핵심 구성요소인 초대형 프로세스 실행이력을 기록한 이벤트 로그 빅데이터는 분산 프로세스 엔진의 로깅 및 감사 컴포넌트[5]의 기능에 따른 프로세스 인스턴스 및 액티비티 워크아이템의 실행 이벤트 로그데이터 기록물이고, 각 이벤트 로그는 해당 비즈니스 프로세스 패키지-식별자, 프로세스 모델-식별자, 프로세스 인스턴스-식별자, 그리고 워크아이템의 액티비티-식별자를 우선-키로 하는 데이터레코드로 표현되며, 실행담당 수행자-식별자와 이벤트-식별자 및 타임스탬프 속성 등이 추가될 수 있다. 지금까지 프로세스 실행 이벤트 로그의 구조와 형식에 대한 여러 연구가 발표되었는데, M. Park, K. Kim[6]의 연구에서는 워크플로우 마이닝을 위한 XML 기반 워크플로우 실행 이벤트 로그 언어로서 XML기반의 XWELL을 제안하였고, 워크플로우 및 비즈니스 프로세스 기술의 국제 표준화기구인 WfMC*는 표준화 된 감사 및 로그 포맷에 관한 규격으로 BPAF**[7]를 발표한 바 있다. 특히, IEEE에서는 XES***[8]

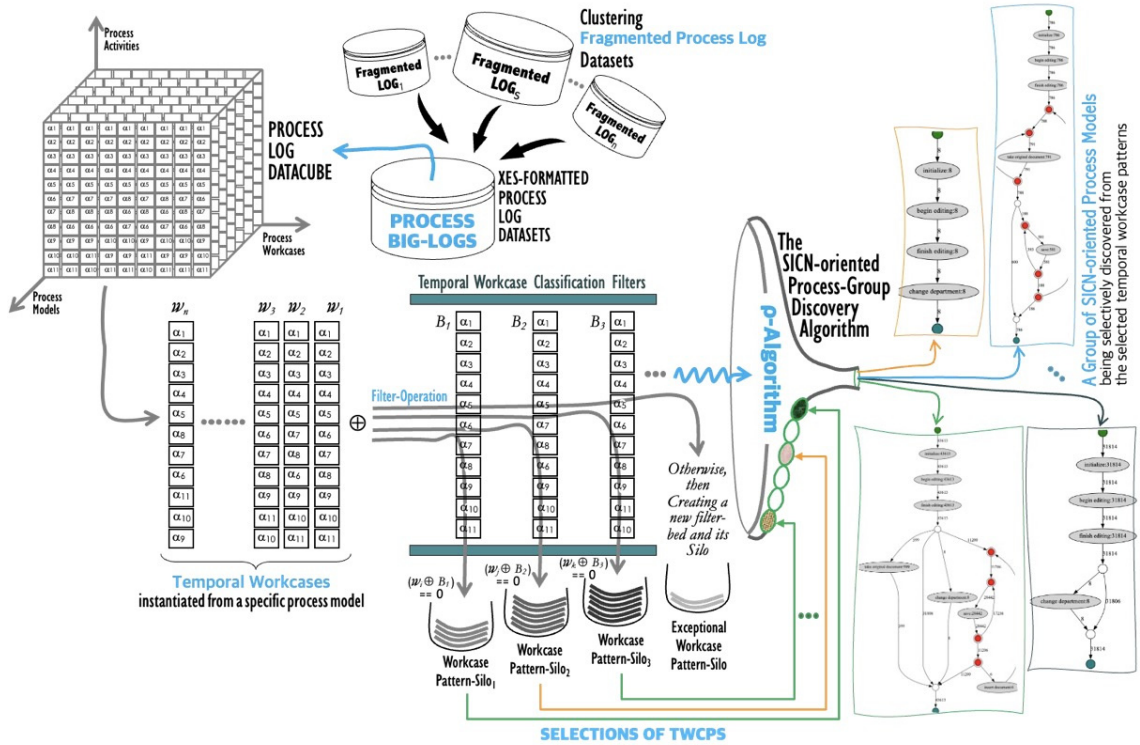
를 발표했으며, 이 XML기반의 이벤트 로그 표준 포맷의 목표는 정보 시스템 설계자에게 이벤트 로그 및 이벤트 스트림을 통해 시스템의 동작을 캡처할 수 있는 통합되고 확장 가능한 이벤트 로그 저장 국제표준포맷을 제공하는데 있으며, 본 논문에서 뿐 만 아니라 최근에 발표되는 대부분의 연구에서는 이 IEEE-XES 이벤트 로그 포맷을 기반으로 하는 실제 데이터세트를 대상으로 실험을 수행하고 있다. 또한, 현재까지 워크플로우 및 비즈니스 프로세스 실행 이벤트 로그 데이터세트를 제공하고 관리하는 공학·과학·경영분야의 대표적인 국제데이터저장소인 4TU.Centre for Research Data[4]에서도 이 국제표준포맷에 따른 데이터세트들을 제공하고 있다.

프로세스 마이닝 알고리즘과 분석 기법 측면에서의 연구성과는 본 논문에서 뿐 만 아니라 여러 연구사례가 수행되어 왔다. 특히, 프로세스 마이닝 알고리즘들과 관련된 대표적인 연구결과는 W.M.P. van der Aalst, A.J.M.M. Weijters[1]의 -알고리즘(-Algorithm)과 K. Kim, C.A. Ellis[3]의 -알고리즘(-Algorithm) 그리고 K.-S Kim, et al.[16]의 -알고리즘(-Algorithm) 등이 알려져 있는데, 본 논문에서 제안하는 제어경로 기반 프로세스 그룹 발견 알고리즘은 대표적인 정보제어넷 기반 프로세스 마이닝 알고리즘인 -알고리즘(-Algorithm)을 확장 적용한 연구결과이다, 또한, 제어흐름 측면의 지식 발견 접근 방법 이외에 업무수행자 측면 등 전반적인 프로세스 관련 지식 발견 기법들 중에서 대표적인 연구성과는 M. Park, et al.[14]의 워크플로우 기반 소셜네트워크 지식 발견 기법과 H. Ahn, et al[16]의 작업전이네트워크 지식 발견 기법 등 이다. -알고리즘은 페트넷 기반의 워크플로우 및 비즈니스 프로세스 모델의 실행이력 이벤트 로글로부터 프로세스를 재발견하는 알고리즘이고, -알고리즘과 -알고리즘은 정보제어넷 기반의 워크플로우 및 비즈니스 프로세스 모델의 실행이력 이벤트 로그로부터 제어흐름 지식을 발견하는 알고리즘이다. 특히, -알고리즘은 프로세스 모델링에 있어서의 기초적인 빌딩블록인 4가지 근본적인 프로세스 패턴, 즉 순차적, 선택적, 병렬적, 반복적 프로세스 패턴의 제어흐름 지식을 발견하는데 있어서 각 액티비티의 실행 횟수를 활용하는 접근방법으로서 본 논문의 제어경로에 따른 프로세스 마이닝 접근방법에 가장 적절히 확장적용할 수 있는 알고리즘이다. 본 논문에서 제안하는 제어경로 기반 프로세스 그룹

* WfMC: Workflow Management Coalition

** BPAF: Business Process Audit Format

*** XES: eXtensible Event Stream



(그림 2) 제어경로 기반 프로세스 그룹 발견 프레임워크
 (Figure 2) The Control-Path-driven Process Group Discovery Framework

발견 알고리즘과 프레임워크는 바로 이러한 정보제어넷 기반 프로세스 마이닝 알고리즘과 제어흐름 지식 발견 기법을 매우 시의적절하게 확장·적용한 연구성과이다.

3. 프로세스 그룹 발견 프레임워크

본 논문의 핵심 연구내용은 분산 컴퓨팅 환경을 기반으로 하는 종적·횡적·혼합적 조각화 기법[10]에 따른 분산 프로세스 조각들의 실행이력 이벤트 로그데이터들을 모두 수집하여 저장한 “프로세스 빅-로그” 빅데이터 저장소로부터 제어흐름 중심의 프로세스 지식을 마이닝하는 프로세스 발견 알고리즘과 이를 이용하여 제어경로에 따른 일련의 프로세스 그룹을 발견하는 프레임워크를 제안하고 구현하는 것이다. 특히, 프로세스 빅-로그는 초대형 분산 프로세스 모델의 실행이력 이벤트로그 빅데이터로서 초대형 여부를 판단하는 정량적 기본 속성을 만족한다고 가정하는데, 그 기본 속성은 조직 내의 프로세스

패키지 및 모델의 수, 각 프로세스모델을 구성하는 액티버티의 수, 각 프로세스 모델의 인스턴스의 수, 각 프로세스 모델의 실행에 참여하는 수행자의 수 그리고 각 프로세스 모델의 실행 시에 발생하는 이벤트의 수 등이다. 본 장에서는 그림 2에 도식화한 것과 같이 제안하는 제어경로 기반 프로세스 그룹 발견 프레임워크를 구성하는 일련의 알고리즘을 비롯한 각 구성요소에 대한 정형적 정의와 세부 포맷을 자세히 기술하고자 한다.

3.1 프로세스 빅-로그의 정형적 정의

분산 프로세스 모델의 인스턴스가 실행될 때 비퍼एम (BPM: Business Process Management) 엔진의 로깅 함수는 각 액티버티-워크아이템의 실행 이벤트를 로그 저장소에 기록하게 되는데, 본 논문에서는 각 이벤트로그의 저장 포맷으로서 IEEE-XES 표준 이벤트 스트림 태그 언어의 저장 포맷을 기반으로 한다고 가정한다. 그림 3에서 나타난 바와 같이, 프로세스 빅-로그를 구성하는 각 분산 프

로세스-조각 인스턴스들과 그들의 액티비티-워크아이템들의 실행 이벤트 로그에 대한 정형적 정의는 다음과 같다.

[정의 1] 프로세스 인스턴스의 액티비티-워크아이템 실행 이벤트 로그: $we = (, pc, wf, wc, ac, p^*, t, s)$

- = 액티비티-워크아이템 이벤트 로그 식별자
- pc = 프로세스 패키지 식별자
- wf = 프로세스 모델 식별자
- wc = 프로세스 인스턴스 식별자
- p^* = 수행자 식별자
- t = 타임스탬프 (timestamp)
- s = 액티비티-워크아이템의 실행 상태 (state): *ready, assigned, reserved, running, completed, cancelled*

동일한 프로세스 인스턴스 식별자를 갖는 액티비티-워크아이템 실행 이벤트 로그들은 결과적으로 해당 프로세스 인스턴스의 실행이벤트 트레이스 (Trace)를 구성하게 되는데, 이를 프로세스 인스턴스 이벤트 트레이스라고 정의하며, 프로세스 마이닝 기법의 기본적인 처리단위가 된다. 분산 프로세스-조각 인스턴스 이벤트 트레이스에 대한 정형적 정의는 다음과 같다.

[정의 2] 프로세스-조각 인스턴스 이벤트 트레이스: $FTWC(ac) = (we_1, \dots, we_n)$, ac = 프로세스-조각 인스턴스 식별자,

$$\{ we_i \mid we_i.wc = ac \wedge we_i.t \leq we_j.t \wedge we_i.pc = we_j.pc \wedge we_i.wf = we_j.wf \wedge we_i.wc = we_j.wc \wedge i < j \wedge 1 \leq i, j \leq n \}$$

궁극적으로, 각 이벤트 로그의 속성인 타임스탬프와 액티비티-워크아이템의 실행상태 속성을 고려하여 프로세스-조각 인스턴스의 액티비티-워크아이템 이벤트 로그를 타임스탬프 시간 순서대로 정렬시킨 이벤트 로그 시퀀스를 프로세스-조각 인스턴스 이벤트 트레이스라고 정의할 수 있다. 여기서 의미있는 시간 순서는 다음과 같은 타임스탬프의 유형에 따라 정의될 수 있다.

- 예정 시간 : 작업 항목의 상태가 준비 (READY)에서 할당 (ASSIGNED)로 변경될 때 취해진다.
 $wet.s \Rightarrow (t = we.t \wedge s = we.s \wedge s = 'assigned')$
- 할당 시간 : 작업 항목의 상태가 준비에서 예약 (RESERVED)로 변경될 때 취해진다.

$$wet.e \Rightarrow (t = we.t \wedge s = we.s \wedge s = 'assigned')$$

- 시작 시간 : 작업 항목의 상태가 예약에서 수행 (RUNNING)로 변경될 때 취해진다.

$$wet.u \Rightarrow (t = we.t \wedge u = we.s \wedge u = 'running')$$

- 완료 시간 : 작업 항목의 상태가 수행에서 완료 (COMPLETED)로 변경될 때 취해진다.

$$wet.o \Rightarrow (t = we.t \wedge o = we.s \wedge o = 'assigned')$$

앞서 정의한 프로세스 빅-로그에 대한 정형적 정의는 이론적 기반으로 하는 프로세스-조각 인스턴스 실행이력 이벤트 로그로부터 일련의 프로세스 그룹을 발견하는 알고리즘을 고안하기 위해서는 프로세스 마이닝 알고리즘의 입력데이터를 구성하는 기본적인 처리단위인 프로세스-조각 인스턴스 실행이력 이벤트 트레이스들에 대한 정형적 모델을 프로세스-조각 인스턴스의 시간적 워크케이스 모델이라고 정의하여 활용할 필요가 있다. 다음의 [정의 3]은 그림 3에서 나타난 시간적 워크케이스 모델에 대한 정형적 정의를 나타낸 것이다.

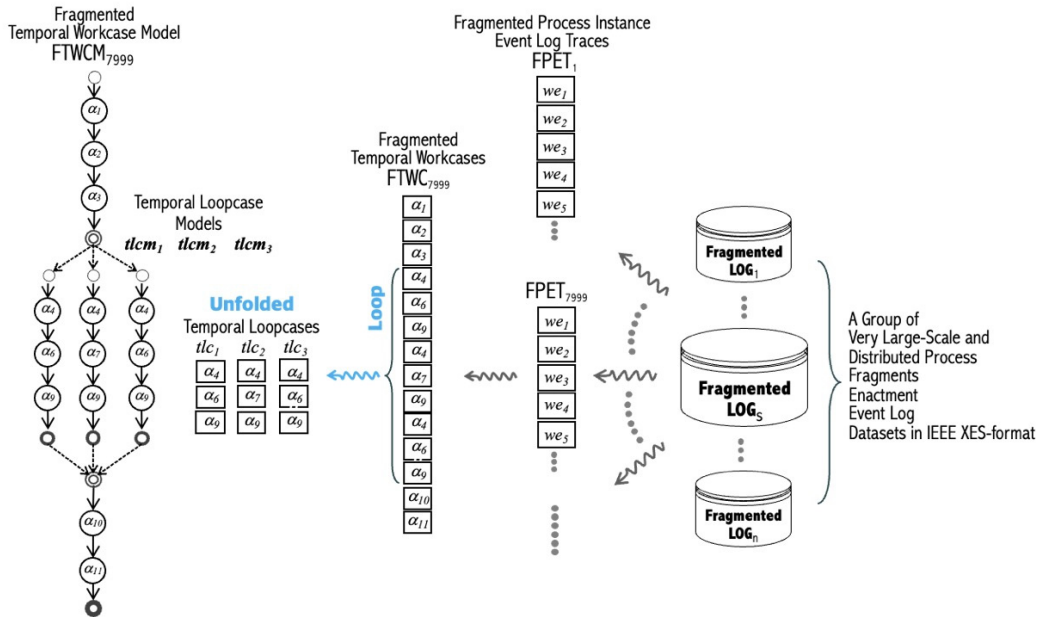
[정의 3] 프로세스-조각 인스턴스의 시간적 워크케이스 모델, FTWC (c): 특정 워크플로우 인스턴스 wc 에 해당하는 시간적 워크케이스 모델은 이벤트 로그의 타임스탬프 유형에 따라 4가지 유형의 시간적 워크케이스 모델이 정의될 수 있다.

- $\in \{ s, e, u, o \}$

$$\bullet FTWC (wc) = (we_{\alpha_1}^{\tau[\phi]}, \dots, we_{\alpha_m}^{\tau[\phi]}):$$

$$\{ we_{\alpha}^{\tau[\phi]} \mid \alpha = we.ac \wedge \tau = we.t \wedge \phi \in \{s, e, u, o\} \wedge we_{\alpha}.wc = c \wedge (we_{\alpha_i}^{\tau_i} < we_{\alpha_j}^{\tau_j})^{10} \wedge \tau_i < \tau_j \wedge i < j \wedge 1 \leq i, j \leq m \},$$

결과적으로, 각 이벤트 트레이스를 수학적 그래프 모델로 정의한 시간적 워크케이스 모델에 대한 정형적 정의와 그래픽적 정의는 각각 [정의 3]와 (그림 3)에서 나타내었다. 즉, 타임스탬프를 중심으로 시간적으로 정렬된 프로세스-조각 인스턴스의 액티비티-워크아이템 이벤트 시퀀스에 대한 정형적 모델과 그래픽적 모델을 시간적 워크케이스 모델(Temporal Workcase Model)로 정의한 것이다. 또한, (그림 3)에서 나타내었듯이 하나의 이벤트 시퀀스에는 반복제어흐름과 연계된 액티비티-워크아이템들은 별도의 루핑-이벤트 트레이스들을 포함할 수 있는데, 이들은 시간적 루핑케이스 모델(Temporal Loopcase



(그림 3) 프로세스-조각 인스턴스 실행이벤트 트레이스와 시간적 워크케이스 모델의 정형적 개념
 (Figure 3) The Formal Concepts of Fragmented Process Instance Execution Event Traces and Models

Model)로 정의될 수 있다. 이와 같은 시간적 워크케이스 모델과 시간적 루프케이스 모델은 예정 시간, 할당 시간, 시작 시간 및 완료 시간과 같은 타임스탬프의 유형에 따라 시간적 워크케이스 모델 뿐만 아니라 시간적 루프케이스 모델은 다음과 같이 4 가지 유형으로 구분될 수 있다.

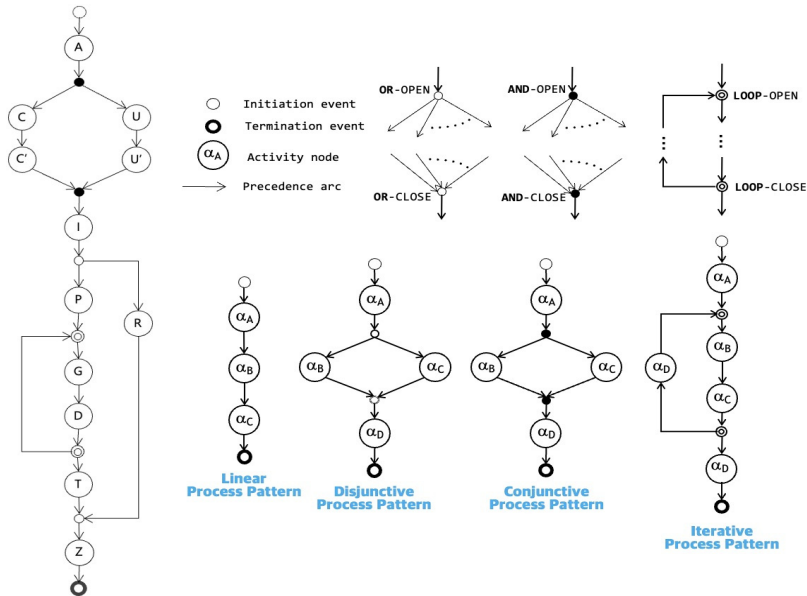
- 예정시간 (= s) 기반 시간적 워크케이스 모델
- 할당시간 (= e) 기반 시간적 워크케이스 모델
- 시작시간 (= u) 기반 시간적 워크케이스 모델
- 종료시간 (= o) 기반 시간적 워크케이스 모델

3.2 프로세스 제어경로와 시간적 워크케이스 패턴

비즈니스 프로세스 모델의 기본구조에 대한 정형적 정의는 액티버티 집합과 천이조건 집합을 대상으로 하는 두 개의 함수, 즉 해당 프로세스를 구성하는 액티버티들의 선·후행 액티버티 집합들을 정의하는 함수와 각 액티버티들의 선·후행천이조건 집합들을 정의하는 함수로 정의되는데, 본 논문에서는 대표적인 비즈니스 프로세스 모델링 방법론인 정보제어넷(Information Control Nets) 그

래프 모델을 이론적 기반으로 하며, 프로세스 모델을 구성하는 액티버티들 간의 선·후행 제어흐름은 그림 4에서 나타난 바와 같이 네 가지 유형의 기본적인 제어흐름 유형들, 즉 순차형, 선택형, 병렬형, 반복형 제어흐름 유형들을 기초단위의 빌딩블록으로 하여 프로세스 모델의 제어흐름을 모델링하는 것으로 가정한다. 특히, 정보제어넷 기반 프로세스 모델에서의 제어흐름 유형이 갖는 전형적인 속성은 바로 구조적 제어흐름 속성을 만족해야 한다는 점이다. 구조적 제어흐름 속성은, 그림 4에서 알 수 있듯이, 제어흐름 게이트웨이-노드들이 반드시 Split-Join의 순서쌍 속성(Matched Pairing Property)과 다중 Split-Join 순서쌍의 적절한 네스팅 속성(Proper Nesting Property)을 만족시켜야 하는 필수조건이다. 그림 4의 왼쪽부분에 나타난 하나의 구조적 정보제어넷 기반 프로세스 모델은 이러한 구조적 제어흐름 속성을 만족하는 제어흐름 빌딩블록들과 이들을 적용한 모델링 결과를 나타낸 것이다.

본 논문에서 제안하는 프로세스 그룹 발견 알고리즘과 프레임워크에서의 핵심 개념인 제어경로는 구조적 정보제어넷 기반 프로세스 모델의 시작이벤트로부터 출발해서 종료이벤트에 이르기까지의 도달가능한 단위업무



(그림 4) 정보제어넷 기반 프로세스 모델의 기본적인 제어흐름 유형
(Figure 4) Primitive Process Patterns of the SICN-oriented Process Model

액티비티들 간의 모든 선·후행 실행순서의 경로들을 의미하며, 해당 프로세스 모델의 인스턴스 실행 이벤트 트레이스, 즉 시간적 워크케이스들의 군집화 또는 클러스터링을 가능하게 하는 개념이다. 예를 들어, 그림 4에서 나타난 정보제어넷 기반 프로세스 모델로부터 다음과 같은 제어경로 유형들을 기반으로 시간적 워크케이스들을 군집화시킬 수가 있다.

- 제어경로 유형-1: $A(CC' \bullet UU')IRZ$
 - 가능한 실행이벤트 트레이스의 시간적 워크케이스 패턴 (Temporal Workcases Patterns):
 $ACC'UU'IRZ / AUU'CC'IRZ /$
 $ACUC'U'IRZ / AUCU'C'IRZ /$
 $ACUU'C'IRZ / AUCC'U'IRZ$
- 제어경로 유형-2: $A(CC' \bullet UU')IP[GD]TZ$
 - 가능한 실행이벤트 트레이스의 시간적 워크케이스 패턴 (Temporal Workcases Patterns):
 $A(CC' \bullet UU')IPGDITZ /$
 $A(CC' \bullet UU')IPGDGDTZ /$
 $A(CC' \bullet UU')IPGDGDGDTZ /$
 ...

이상과 같이 예제의 정보제어넷 기반 프로세스 모델에서 두 가지의 제어경로 유형, 즉 제어경로 유형-1과 제어경로 유형-2로부터 가능한 프로세스 인스턴스 실행 이벤트 트레이스들을 군집화할 수 있는 시간적 워크케이스 패턴들을 보여준 것이다. 특히, 제어경로 유형-2에서는 제어경로 유형-1의 병렬형-빌딩블록에 이어진 반복형-빌딩블록의 반복 회수에 따른 시간적 워크케이스 패턴들을 보여 주고 있다. 결과적으로, 본 논문의 궁극적 목표는 이와 같은 개념의 제어경로 유형들 뿐 만 아니라 각 제어경로 유형과 연관된 시간적 워크케이스 패턴들을 중심으로 일련의 프로세스 그룹을 프로세스 빅-로그 데이터셋으로부터 발견할 수 있는 일명 제어경로 기반 프로세스 그룹 발견 알고리즘과 프레임워크 그리고 시스템을 개발하는데 있다.

그림 2에서 나타난 제어경로 기반 프로세스 그룹 발견 프레임워크의 중간부분에는 제어경로 유형을 기반으로 하는 각 시간적 워크케이스 패턴에 따른 프로세스 인스턴스 실행 이벤트 트레이스들의 군집화 알고리즘을 개념적으로 도식화하고 있다. 이 군집화 알고리즘의 핵심 개념은 시간적 워크케이스 분류 여과장치 (Temporal Work-case Classification Filter-Bed)의 구현을 위한 접근방

Algorithm 1 The Temporal Workcase Class-Filtering Algorithm

Require: A Dataset of Process Enactment Logs in XES-Format ($EL.xes$) in XES-Format

Ensure: A Dictionary as Silos of Temporal Workcase Class ($DTWC$)

```

1:  $DTWC = \{\}$ ;
2: openfile  $EL.xes$ ;
3: while ( $!EOF$  in  $EL.xes$ ) do
4:    $line \leftarrow currentLine$ ; ▷ Reading line by line in the process enactment log in XES-Format
5:    $listTraceID = [ ]$ ;
6:   if (" $< trace >$ "  $\in line$ ) then
7:      $traceid \leftarrow Get Trace ID from line$ ;
8:      $traceclass = [ ]$ ; ▷ Constructing a list of WorkcaseClass Silos
9:   else
10:    if (" $< event >$ "  $\in line$ ) then
11:       $activity \leftarrow \emptyset$ ;
12:    else
13:      if (" $< /event >$ "  $\in line$ ) then
14:         $activity \leftarrow Get activity from line$ ;
15:         $traceclass.append(activity)$ ;
16:      else
17:        if (" $< /trace >$ "  $\in line$ ) then
18:          if ( $traceclass \notin DTWC$ ) then ▷ If traceclass not in DTWC then create a new filter and its Silo
19:             $DTWC[traceclass] = 1$ ;
20:          else
21:             $DTWC[traceclass] += 1$ ; ▷ If traceclass in DTWC then adding into the existing Silo
22:          end if
23:        end if
24:      end if
25:    end if
26:  end if
27:  seek to the next line;
28: end while
29: closefile;

```

법 또는 알고리즘에 있다. 첫 번째 접근방법은 프레임워크의 그림으로 나타내었듯이 분류 알고리즘을 설계 및 구현하는 방법으로서, 각 시간적 워크케이스의 액티비티-워크아이템 이벤트 시퀀스를 2진수의 비트-시퀀스로 변형시킨 후, 새로운 비트-시퀀스가 발견될 때마다 이를 새로운 분류 필터로 채택하면서 분류 여과장치를 구축해나가는 접근방법이다. 비트-시퀀스로 변형된 시간적 워크케이스의 이벤트 로그 트레이스들과 분류 필터들과의 배타적-OR (Exclusive-OR) 비트-오퍼레이션을 수행하여 그 결과가 "0"이면 해당 분류 필터, 즉 시간적 워크케이스 패턴에 속하는 시간적 워크케이스로 분류시킴으로써 프로세스 빅-로그에 저장된 어느 특정 프로세스 모델의 실행이력으로 저장된 모든 시간적 워크케이스(인스턴스) 실행 이벤트 로그 트레이스)들을 분류시킬 수 있다.

두 번째 접근방법은 한층 더 시스템적이고 초대형의 분산 프로세스 모델과 그의 프로세스 빅-로그에 적합한 접근방법으로서 대용량 분산데이터의 분류처리 플랫폼인 맵리듀스(MapReduce) 프레임워크를 활용하는 것이다.

본 논문의 저자들이 속한 연구그룹에서는 초대형 분산 프로세스 마이닝의 전처리 기법으로서 맵리듀스 기반 프로세스 빅-로그 클러스터링 기법[17]을 성공적으로 개발한 바 있다. 즉, 구글에서 대용량 데이터 처리를 분산 병렬 컴퓨팅에서 처리하기 위한 목적으로 개발된 소프트웨어 프레임워크인 맵리듀스 플랫폼은 일반적으로 사용되는 Map과 Reduce라는 함수 기반으로 구동된다. 맵리듀스의 입력은 각 레코드가 (Key, Value)의 쌍으로 구성되는 일련의 리스트형태를 취하는데, 전처리 알고리즘의 입력으로 프로세스 빅-로그를 구성하는 시간적 워크케이스를 Value(TWC-value)로 하고, 그의 파일 내의 라인번호 필드를 Key(TWC-key)로 하는 일련의 리스트들을 조직할 수 있고, 결과적으로 Map() 함수는 이 (TWC-key, TWC-value) 쌍의 리스트를 읽어 다시 (pattern-key, count-value) 형태의 중간 결과를 Emit() 함수를 통해 출력한다. 이 중간 결과들은 pattern-key를 기준으로 동일값으로 군집화 또는 클러스터화시킨 (pattern-key, count-values[]) 쌍의 리스트 형태로 구성되고, Reduce() 함

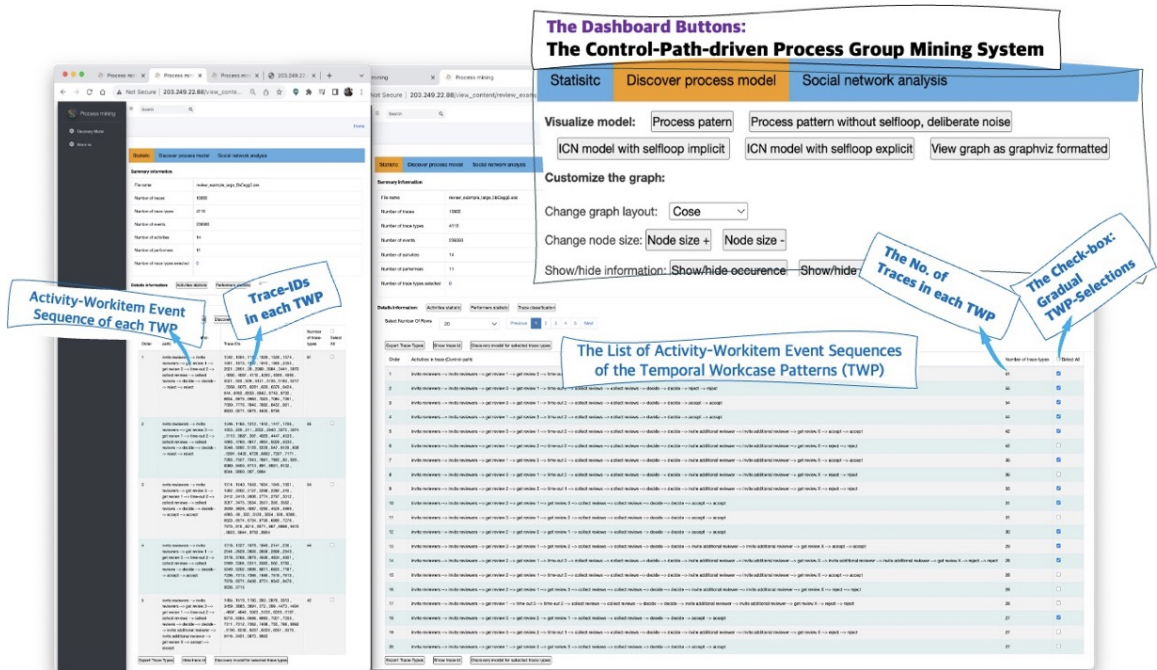
수는 각 **pattern-key** 값이 해당되는 (**pattern-key**, **pattern-values**) 쌍의 리스트들에 대해 집계연산을 수행하고, 최종 결과, 즉 각 패턴 (시간적 워크케이스 패턴)과 그의 발생 총합인 (**pattern p**, **count sum_p**)을 Emit() 함수를 통해 출력하게 된다. 결과적으로, 이 맵리듀스 플랫폼을 적용한 분류-필터를 구현 및 검증한 바 있다.

정리하자면, 이상과 같은 두 가지 접근방법으로 프로세스 제어경로 기반의 시간적 워크케이스 패턴 분류 여과장치와 관련된 알고리즘과 전처리 기법을 성공적으로 개발하고 그 기능적 정확성을 검증한 바 있다. 본 논문에서는 첫 번째 접근방법을 기반으로 하는 시간적 워크케이스 분류-필터 알고리즘을 Algorithm 1에서와 같이 설계하였으며, 이를 적용한 제어경로 기반 프로세스 그룹 마이닝 시스템을 성공적으로 개발하였다.

3.3 프로세스 그룹 발견 알고리즘

프로세스 그룹 발견 알고리즘의 이론적 기초는 정보 제어넷 기반 프로세스 모델[10]과 일명 -알고리즘[16]으

로 알려진 대표적인 구조적 정보제어넷 기반 프로세스 마이닝 알고리즘에 있다. 즉, 본 논문에서는 -알고리즘의 기능확장으로 통한 새로운 개념의 제어경로 기반 프로세스 그룹 발견 알고리즘을 제안한다. 그림 2에서 나타난 프레임워크의 오른쪽 부분에서 깔때기-모양의 알고리즘 형상그림을 통해 개념적인 구현방안을 도식화하였듯이, 구조적 정보제어넷 기반 프로세스 그룹 발견 알고리즘은 앞 절에서 정의한 제어경로 유형에 따른 분류-필터-알고리즘을 적용하여 시간적 워크케이스 패턴들과 그에 속한 실행이벤트 트레이스들을 해당 프로세스 빅로그 데이터세트로부터 발견할 수 있으며, 이렇게 발견된 시간적 워크케이스 패턴들 중에서 임의적 또는 분석적 선택과정을 통해 하나 또는 그 이상의 마이닝 대상 패턴들로 선택하고 그에 속한 시간적 워크케이스 (실행 이벤트 로그 트레이스)들로 구성되는 입력데이터세트로부터 하나의 구조적 정보제어넷 기반 프로세스 모델을 발견한다. 이러한 제어경로를 기반으로 하는 마이닝 대상 시간적 워크케이스 패턴들의 선택과정을 반복할 때마다 하



(그림 5) 제어경로 기반 프로세스 그룹 마이닝 시스템의 기능구조와 대시보드

(Figure 5) Dashboard and Functional Components of the Control-Path driven Process Group Mining System

나의 구조적 정보제어넷 기반 프로세스 모델을 발견하게 되며, 이러한 반복적인 마이닝 과정들로 발견되는 프로세스 모델들로 해당 프로세스 박-로그 데이터세트에 해당하는 프로세스 그룹을 구성하는 것이다.

결과적으로, **Algorithm 2**는 기존의 -알고리즘을 개선한 제어경로 기반 프로세스 그룹 발견 알고리즘의 설계 결과를 나타낸 것이며, 이 알고리즘을 적용한 프로세스 그룹 마이닝 시스템을 성공적으로 개발 완료하였다. **Algorithm 2**에서 나타내었듯이, 입력데이터세트는 모든 프로세스 인스턴스의 액티비티-워크아이템 실행 이벤트 시퀀스들로부터 구축되는 일명 프로세스 패턴 그래프(각 액티비티의 실행 회 수 포함)와 선택된 시간적 워크케이스 분류-필터와 그에 속한 트레이스들로 구성되며, 이로부터 발견되는 출력결과는 구조적 정보제어넷 기반 프로세스 모델임을 알 수 있다. 이러한 분류-필터 선택과정과 함께 하는 마이닝 과정을 반복함으로써 발견해내는 일련의 프로세스 모델들을 모아서 프로세스 그룹을 구성하는 것이다.

다음의 장에서는 **Algorithm 2**를 적용하여 구현된 프로세스 그룹 마이닝 시스템을 활용하여 제안된 접근방법의 개념적 합리성과 개발된 시스템의 기능적 정확성을 검증

하고자 한다.

4. 실험적 검증

앞서 기술한 제어경로 기반 프로세스 그룹 발견 프레임워크와 관련 알고리즘을 웹기반의 프로세스 마이닝 시스템으로 구현하였으며, 4TU.Centre for Research에서 제공한 **BPI Challenges 2018[4]**의 공개 데이터세트의 하나인 고객-적요 프로세스 모델(**Customer Summary Process Model**)의 실행 이벤트 로그 데이터세트를 이용하여 제안된 접근방법의 개념적 우수성과 구현시스템의 기능적 정확성에 대한 실험적 검증을 수행하였다.

4.1 프로세스 그룹 마이닝 시스템

본 논문에서 제안한 그림 2의 제어경로 기반 프로세스 그룹 발견 프레임워크와 관련 알고리즘들을 기반으로 새로운 유형의 웹기반 프로세스 그룹 마이닝 시스템을 개발 완료하였으며, 그림 5는 개발된 시스템의 대시보드와 기능적 구조를 캡처-윈도우들을 통해 보여준 것이다. 웹기반 프로세스 그룹 마이닝 시스템의 사용자 인터페이스 윈도우는 두 개의 영역으로 구분되어 있는데, 윗부분영

Algorithm 2 The Revised ρ -Algorithm: the SICN-oriented Process Group Discovery Algorithm

Require: A Weighted Process Pattern Graph, $G = (V, E)$ and Appended WorkcasesClass-Silos, $aWCS$

Ensure: A Structured Information Control Net Graph, G^{SICN}

```

1: for (  $\forall traceclass \in aWCS$  ) do
2:   for (  $\forall activity \in traceclass$  ) do
3:     if (  $activity \notin V$  ) then
4:        $V.add(activity)$ ;                                ▶ Adding an activity into Weighted Process Pattern Graph, G
5:     end if
6:     if ( (  $activity, nextactivity$  )  $\notin E$  ) then
7:        $E[(activity, nextactivity)] = aWCS[traceclass]$ ; ▶ Adding an edge of (activity, nextactivity) in the traceclass into G
8:     else
9:        $E[(activity, nextactivity)] += aWCS[traceclass]$ ; ▶ Adding a weight (occurrence) to the edge
10:    end if
11:  end for
12: end for
13:  $G.removeDeliberateNoiseEdge()$ ; (*)
14: for (  $\forall vertex \in G$  ) do
15:   if (  $vertex.edgeOutGoing > 2$  ) then
16:      $G^{opengate} \leftarrow ProcessForTheOpenGate(G, vertex)$ ; ▶ Discovering and building three types (XOR/AND/LOOP) of
OPEN-gateway nodes by applying the decision-making rules with  $\rho$ -operations.
17:   end if
18:   if (  $vertex.edgeInComing > 2$  ) then
19:      $G^{closegate} \leftarrow ProcessForTheCloseGate(G, vertex)$ ; ▶ Discovering and building three types (XOR/AND/LOOP) of
CLOSE-gateway nodes by applying the decision-making rules with  $\rho$ -operations.
20:   end if
21: end for

```

역은 프로세스 그룹 발견 프레임워크를 구성하는 단계적 기능함수들과 함께 가시화 측면의 기능함수들과 연계된 대시보드이며, 아랫부분영역은 시간적 워크케이스 패턴들과 그에 속한 프로세스 인스턴스 실행 이벤트 시퀀스(트레이스)들의 가시화 영역과 프로세스 발견에 적용될 시간적 워크케이스 패턴들의 선택여부를 표시하는 클릭-박스 영역으로 구성되어 있다.

특히, 토글-스위치 버튼(show/hide)으로 각 시간적 워크케이스 패턴에 속한 트레이스들의 식별자들을 나타내는 컬럼-필드의 가시화 여부를 선택할 수 있도록 구성되어 있다. 왼쪽의 캡처-윈도우에서는 트레이스 식별자들을 보여준상태이며, 오른쪽 캡처-윈도우에서는 트레이스 식별자들을 감춘상태이다. 또한, 오른쪽 캡처-윈도우에는 프로세스 발견에 적용할 시간적 워크케이스 패턴들의 선택여부를 결정하는 클릭-박스를 보여주고 있다. 결과적으로, 시각적 워크케이스 패턴들을 나타내는 리스트를 구성하는 필드들은 패턴-식별자, 액티비티-워크아이템 실행 시퀀스, 패턴에 속한 트레이스의 개 수, 프로세스 발견에 적용될 패턴의 선택여부 클릭-박스이다.

4.2 실험적 검증과 결과의 활용

앞 절에서 개발한 제어경로 기반 프로세스 그룹 마이닝 시스템의 개념적 우수성과 기능적 정확성을 검증하기 위하여 4TU.Centre for Research에서 제공한 BPI Challenges 2018[4]의 공개 데이터셋의 하나인 고객-적요 프로세스 모델(Customer Summary Process Model)의 실행이력 이벤트 로그 데이터셋에 적용하는 실험적 적용 사례연구를 수행하였다. 이 데이터 세트에는 총 43,808개의 프로세스 인스턴스 실행에 따른 이벤트 트레이스를 기록저장하고 있으며, 기록된 모든 프로세스 인스턴스들의 실행이력로그 트레이스들에는 해당 액티비티-워크아이템의 실행 이벤트 로그가 포함되어 있다. 참고로, 고객-적요 프로세스 모델의 실행이력 프로세스 빅-로그 데이터셋으로부터 발견한 정량적 통계수치는 다음과 같다.

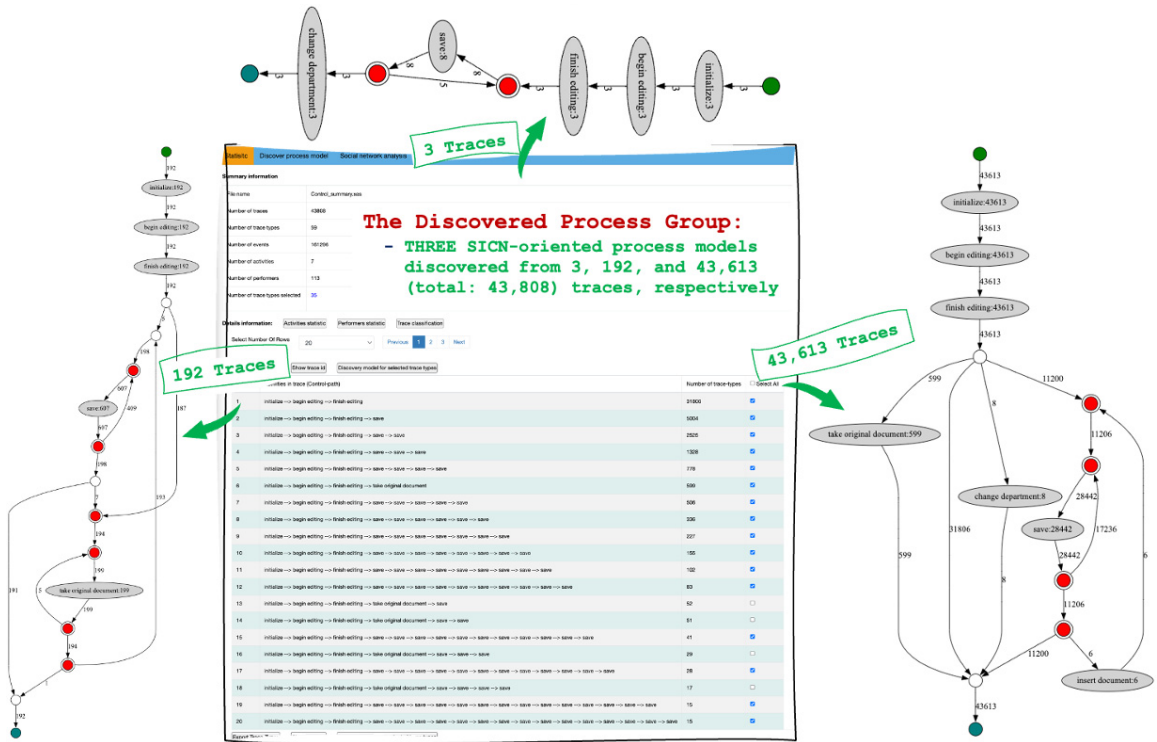
- 프로세스 인스턴스 실행 트레이스 수: 43,808개
- 단위업무 액티비티 수: 7개
- 액티비티-워크아이템 이벤트 수: 161,296개
- 업무수행자 수: 113명
- 시간적 워크케이스 패턴 수: 59개

그림 6은 이러한 데이터셋에 적용한 제어경로 기반

프로세스 그룹 마이닝 시스템의 실험 상황과 그의 실험 결과인 총 59개의 시간적 워크케이스 패턴들과 각 패턴에 속한 트레이스들의 개수 및 식별자 리스트 그리고 최종적으로 제어경로 선택을 통해 발견한 세 개의 프로세스 모델들(각각 3개, 192개, 43613개의 트레이스들로부터 발견됨)로 구성된 프로세스 그룹의 발견결과를 가시화하고 있다. 특히, 입력데이터셋으로부터 오직 하나의 프로세스 모델을 발견하는 기존의 프로세스 마이닝 시스템들의 실행결과와는 달리 본 논문에서 제안 및 구현한 프로세스 마이닝 시스템에서는 제어경로를 기반으로 하는 시간적 워크케이스 패턴들 중에서 프로세스 모델의 발견에 적용할 대상 패턴들을 선택하는 수단을 제공함으로써 입력데이터셋으로부터 다중의 프로세스 모델들로 구성되는 프로세스 그룹을 발견할 수 있는 새로운 프로세스 마이닝 접근방법을 구현한 것이다. 이 실험적 검증을 통해 제어경로 기반 프로세스 그룹 발견 접근방법의 실현가능성과 구현 시스템의 기능적 정확성을 증명할 수 있었다.

또한, 발견된 프로세스 그룹의 활용방안 측면의 우수성을 검증하고 한다. 본 논문에서 제안한 제어경로 기반 프로세스 그룹 발견 알고리즘 및 프레임워크와 이를 구현한 프로세스 마이닝 시스템의 핵심 목표는 프로세스 마이닝 기술의 궁극적 활용분야인 프로세스 리엔지니어링 또는 프로세스 재설계를 통한 프로세스 모델의 생명주기관리를 체계적이고 효율적·효과적으로 지원하는데 있다. 따라서, 그림 6에서 보여준 BPI Challenge 2018 데이터셋인 고객-적요 데이터셋으로부터 발견한 프로세스 그룹에 대한 프로세스 리엔지니어링 측면의 활용방안을 설명함으로써 본 논문에서 제안한 새로운 형식의 프로세스 마이닝 접근방법과 구현시스템의 우수성을 확인할 수 있다. 참고적으로, 비즈니스 프로세스 모델을 설계 및 정의하는데 있어서 안정적이고 합리적인 모델을 만들어내는 것이 중요한 성공요인이다. 즉, 제어흐름의 구조적 속성을 만족하도록 프로세스 모델을 설계하는 것이 바람직한 모델링 접근방법이며, 여기서 구조적 속성은 기본적인 제어흐름-빌딩블록인 XOR/AND/LOOP 빌딩블록에서의 Split-Join의 순서쌍구조 속성과 다중 Split-Join 순서쌍의 적절한 네스팅구조 속성이다.

그림 7은 고객-적요 데이터셋으로부터 발견한 세 개의 프로세스 모델로 구성되는 프로세스 그룹을 나타낸 것이다. 그림에서 볼 수 있듯이, 프로세스 그룹을 구성하는 각각의 구조적 정보제어넷기반 고객-적요 프로세스 모델은 이벤트(시작/끝) 액티비티와 단위업무 액티비티 그리고 구조적 속성인 Split-Join의 순서쌍구조 속성과 다중

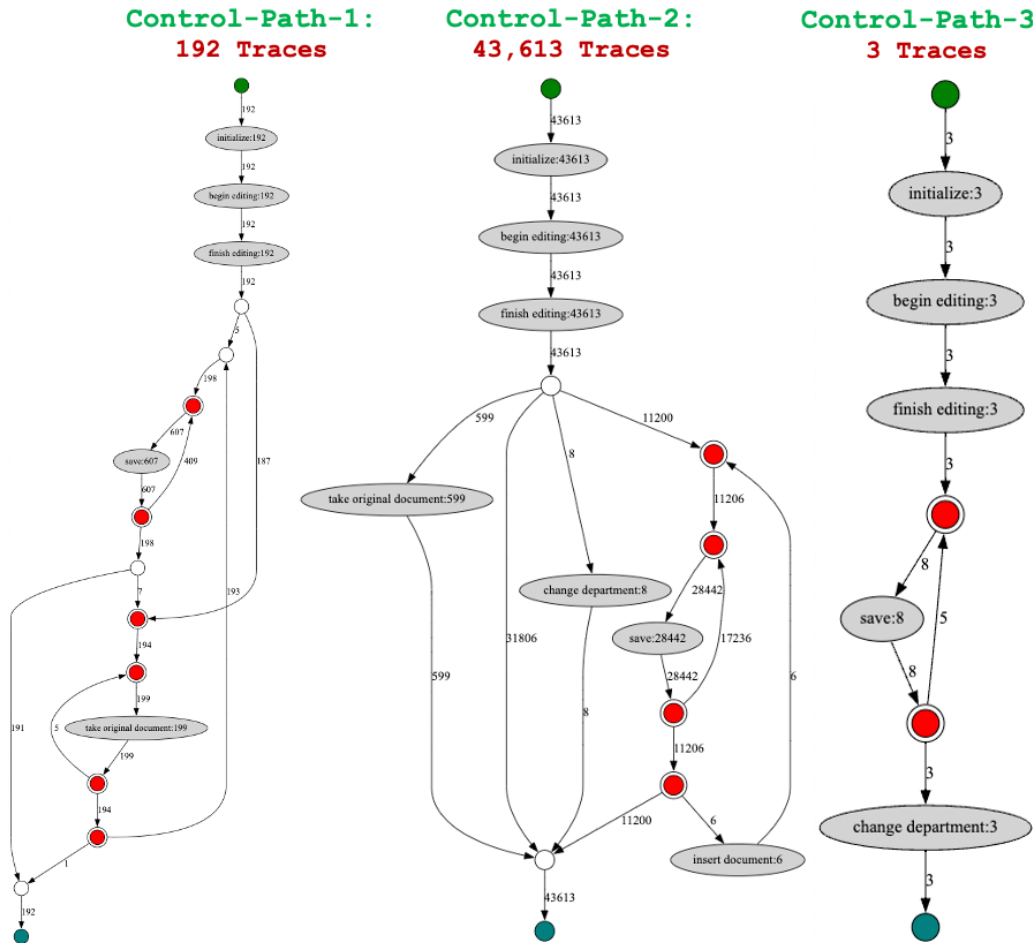


(그림 6) 제어경로 기반 프로세스 그룹 마이닝 시스템의 프로세스 그룹 발견 결과
 (Figure 6) The Discovered Process Group by the Control-Path driven Process Group Mining System

Split-Join 순서쌍의 적절한 네스팅구조 속성을 만족하는 반복적 게이트웨이(loop-split/join gateway) 액티비티와 선택적(exclusive-OR split/join) 게이트웨이 액티비티로 구성되는 완벽한 구조적 정보제어넷 기반 프로세스 모델임을 알 수 있다. 해당 프로세스 그룹을 구성하는 세 가지 프로세스 모델들은 모두 구조적 정보제어넷 기반 프로세스 모델이며, 원래의 프로세스 모델의 제어흐름 구조를 정확히 알 수 없지만 이렇게 발견된 프로세스 그룹을 기반으로 매우 적절하고 합리적인 제어경로를 설계할 수 있도록 프로세스 리엔지니어링 및 재설계에 반영될 수 있을 것으로 판단된다.

상세하게 분석하자면, 제어경로-2로부터 파생된 시간적 워크케이스 패턴들이 총 59개의 패턴들 중에서 35개이며, 해당되는 시간적 워크케이스(실행이벤트로그 트레이스) 개수는 전체 43,808개 중에서 43,613개인 99.56%를 차지하고 있다. 해당 객체-적요 프로세스 모델을 실행하면서 생성된 대부분의 프로세스 인스턴스들은 바로 제어경로-2에 따른 완벽한 구조적 형태의 프로세스 모델로부

터 파생된 것이므로 프로세스 재설계 시에 반드시 핵심적 제어흐름 구조로 반영되어야 하는 프로세스 모델임에 틀림없다. 제어경로-1은 22개의 시간적 워크케이스 패턴과 이에 속한 192개의 트레이스들이 연계되어 있고, 이 역시 완벽한 구조적 형태의 프로세스 모델이다. 하지만 제어경로-1에 따른 총 22개의 시간적 워크케이스 패턴들 중에서 대부분인 13개의 패턴들이 단일 개수의 트레이스를 포함하고 있고, 50이상의 트레이스를 포함하는 패턴의 개수는 2개(각각 52개와 51개)에 불과하다. 따라서, 제어경로-1에 대한 프로세스 재설계 시의 반영여부는 예외상황의 프로세스 인스턴스들로 고려하는 것이 바람직할 것으로 판단된다. 마지막으로, 제어경로-3로부터 파생된 시간적 워크케이스 패턴의 개수는 단지 2개이며, 그에 속한 트레이스들의 개수 역시 3개에 불과하므로 이 제어경로로부터 발견된 프로세스 모델에 대한 프로세스 재설계 시의 반영여부는 의미 있는 주요 고려사항이 아님을 예상할 수 있다.



(그림 7) 고객-적요 데이터셋의 프로세스 그룹 발견 결과
 (Figure 7) The Discovered Process Group from the Customer-Summary Dataset

결과적으로, 본 논문에서 제안한 제어경로 기반 프로세스 그룹 발견 프레임워크와 구현한 프로세스 마이닝 시스템은 실험적 검증을 통해 증명한 바와 같이 개념적 우수성과 기능적 정확성이 입증되었을 뿐 만 아니라 기존의 전통적인 프로세스 마이닝 접근방법들과는 전혀 새로운 차원의 프로세스 마이닝 접근방법임을 강조하고자 한다. 특히, 발견된 프로세스 그룹 개념은 프로세스 리엔지니어링과 재설계에 따른 효과적·효율적 프로세스 생명주기관리방안으로 활용될 수 있는 매우 적합하고 합리적인 접근방법임에 틀림없다.

5. 결 론

본 논문에서는 초대형 분산 비즈니스 프로세스 모델의 실행이력 이벤트 로그로부터 제어흐름 구조 측면에서 합리적이고 구조적 속성을 만족하는 프로세스 모델을 발견하기 위한 새로운 형식의 프로세스 발견 및 마이닝 접근방법을 제안하였다. 특히, 초대형 분산 비즈니스 프로세스 모델의 실행이력 이벤트 로그가 갖는 초대형적 속성, 즉 조직 내의 프로세스 모델의 수, 각 프로세스 모델을 구성하는 액티버티의 수 그리고 각 프로세스 모델의 인스턴스 수 측면에서 빅 데이터의 5V 속성을 만족하고 있어, 이를 특별히 프로세스 빅-로그로 정의하였으며, 이

로부터 시간적 워크케이스 패턴에 따른 인스턴스 실행이 벤트로그 트레이스 분류-필터 알고리즘과 제어경로 기반 시간적 워크케이스 패턴의 선택에 따른 프로세스 그룹 발견 알고리즘으로 구성되는 프로세스 그룹 발견 프레임워크를 제안하였다. 결과적으로, 제안된 프레임워크의 개념적 우수성과 구현된 시스템의 기능적 정확성을 검증하기 위하여 잘 알려진 프로세스 박-로그 데이터셋인 고객-적요 프로세스 데이터셋을 이용한 실험적 검증을 수행하였으며, 이러한 실험적 검증의 최종결과와 함께 프로세스 리엔지니어링 및 재설계에 따른 프로세스 생명주기관리방안으로의 활용방안을 제시하였다.

끝으로, 본 논문에서 제안한 프로세스 마이닝 접근방법과 시스템을 기반으로 하는 향후의 연구주제와 내용으로서 최근에 각광받는 인공지능-딥러닝 기술을 기반으로 하는 예측적 프로세스 모니터링 기술과 예측적 프로세스 모델링 기술과의 연계를 추진할 예정이다.

참고문헌(References)

- [1] W. M. P. van der Aalst and A. J. M. M. Weijters, "Process mining: a research agenda," *Journal of Computers in Industry*, Vol. 53, Issue 3, 2004.
<https://doi.org/10.1016/j.compind.2003.10.001>
- [2] Kyoungsook Kim, et al., "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>
- [3] Kim, Kwanghoon and Ellis, Clarence A., "X-Algorithm: Structured Workflow Process Mining Through Amalgamating Temporal Workcases," *The Proceedings of PAKDD2007, Advances in Knowledge Discovery and Data Mining, Lecture Notes in Artificial Intelligence*, Vol. 4426, pp. 119-130, 2007.
https://doi.org/10.1007/978-3-540-71701-0_14
- [4] BPI Challenge 2012, 2013, 2014, 2015, 2016, 2017, 2018, 4TU.Centre for Research Data,
<https://data.4tu.nl/repository/collection:event-logs-real>.
- [5] Kim, Kwanghoon, "A XML-Based Workflow Event Logging Mechanism for Workflow Mining," *The Proceedings of Advanced Web and Network Technologies, and Applications, APWeb 2006*, pp. 132-136, 2006.
https://doi.org/10.1007/11610496_17
- [6] Minjae Park and Kwanghoon Kim, "XWELL: A XML-Based Workflow Event Logging Mechanism and Language for Workflow Mining Systems," *Lecture Notes in Computer Science*, Vol. 4707, pp. 900 - 909, 2007.
https://doi.org/10.1007/978-3-540-74484-9_76
- [7] Michael zur Muehlen and Keith D. Swenson, "BPAF: A Standard for the Interchange of Process Analytics Data," *Lecture Notes in Business Information Processing*, Vol. 66, pp. 170 - 181, 2011.
https://doi.org/10.1007/978-3-642-20511-8_15
- [8] IEEE, "IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams," *IEEE 1849-2016*, 2016.
<https://doi.org/10.1109/IEEESTD.2016.7740858>.
- [9] Kyoungsook Kim, Young-Koo Lee, Hyun Ahn, Kwanghoon Pio Kim "An Experimental Mining and Analytics for Discovering Proportional Process Patterns from Workflow Enactment Event Logs," *Proceedings of the International Conference on Big Data Technologies and Applications*, Exeter, England, Great Britain, Sept. 4rd - 5th, 2018.
<https://doi.org/10.1007/s11276-018-01899-z>
- [10] Kwanghoon Kim, "A Model-Driven Workflow Fragmentation Framework for Collaborative Workflow Architectures and Systems," *Journal of Network and Computer Applications*, Volume 35, Issue 1, pp. 97-110, 2012.
<https://doi.org/10.1016/j.jnca.2011.03.029>
- [11] K. Lee, Y. Lee, H. Choi, Y. F. Chung and B. Moon, "Parallel Data Processing with MapReduce: A Survey," *SIGMOD Record*, Vol. 40, No. 4, pp. 11-20, 2011.
<https://doi.org/10.1145/2094114.2094118>
- [12] C. Goncalves, L. Assuncao, j. C. Cunha, "Flexible MapReduce Workflows for Cloud Data Analytics," *International Journal of Grid and High Performance Computing*, Vol. 5, No. 4, pp. 17, 2013.
<https://doi.org/10.4018/ijghpc.2013100104>
- [13] Kim KH., Ahn HJ., "An EJB-Based Very Large Scale Workflow System and Its Performance Measurement," In: Fan W., Wu Z., Yang J. (eds) *Advances in*

- Web-Age Information Management. WAIM 2005, Lecture Notes in Computer Science, Vol. 3739. pp. 526-535, Springer, Berlin, Heidelberg, 2005.
https://doi.org/10.1007/11563952_46
- [14] Minjae Park, Hyun, Ahn, and Kwanghoon Pio Kim, "Workflow-supported social networks: Discovery, analyses, and system," *Journal of Network and Computer Applications*, Vol. 75, pp. 355-373, 2016.
<https://doi.org/10.1016/j.jnca.2016.08.014>
- [15] 이경하, 박원주, 조기성, 류원, "대규모 데이터 분석을 위한 MapReduce 기술의 연구 동향," *전자통신동향분석*, 제28권 제6호, pp. 156-166, 2013.
- [16] Kyoungsook Kim, Dinh-Lam Pham, and Kwanghoon Pio Kim, "Algorithm: A SICN-oriented Process Mining Framework," *IEEE Access*, Vol. 9, pp.139851-139875, 2021.
<https://doi.org/10.1109/ACCESS.2021.3119011>
- [17] Min-Hyuck Jin, Kwanghoon Pio Kim, "A MapReduce-Based Workflow BIG-Log Clustering Technique," *Journal of Internet Computing and Services*, Vol. 20, No. 1, pp. 87-96, 2019.
<https://doi.org/10.7472/jksii.2019.20.1.87>
- [18] Kyoungsook Kim, Dinh-Lam Pham, Young-In Park, and Kwanghoon Pio Kim, "Experimental verification and validation of the SICN-oriented process mining algorithm and system," *Journal of King Saud University - Computer and Information Sciences*, Vol. 34, Issue. 10, pp. 9793-9813, 2022.
<https://doi.org/10.1016/j.jksuci.2021.12.013>
- [19] Kwanghoon Pio Kim, "Experiment Analyses of Temporal Activity-Sequencing Anomalies in Process Mining," *Applied Sciences*, Vol. 13, Issue. 5, pp. 3143-3157, 2023.
<https://doi.org/10.3390/app13053143>
- [20] Kwanghoon Pio Kim, "Functional Integration with Process Mining and Process Analyzing for Structural and Behavioral Properness Validation of Discovered Processes from Event Log Datasets," *Applied Sciences*, Vol. 10, No. 4, pp. 1493-1518, 2020.
<https://doi.org/10.3390/app10041493>

◎ 저 자 소개 ◎



응웬 탄 하이 (Thanh Hai Nguyen)

2008년 Pedagogy of Informatics, Thai Nguyen University of Education (이학사)

2012년 Computer Science, Thai Nguyen University (이학석사)

2008년~현재 IT Engineering, Thai Nguyen University, Lecturer

2021년~현재 경기대학교 일반대학원 컴퓨터학과 박사과정

관심분야: 프로세스 마이닝, 예측적 프로세스 모니터링, 비즈니스 프로세스 모델링 방법론, 비즈니스 프로세스 관리 시스템, 그래픽 신경망 학습 이론, 프로세스 지식 발견 인텔리전스, 영상기반 사회무질서 레벨링과 위험측정 방법론, 영상 검색 서비스 시스템, 영상 관제 빅데이터 플랫폼, 영상기반 빅데이터 페브리 플랫폼, 영상기반 빅데이터 레이크 플랫폼, 데이터-집약 디지털 트윈 플랫폼, 데이터-집약 스마트공장 플랫폼, 영상기반 범죄예방·예측 서비스 플랫폼

E-mail : nguyenthanhhai@kgu.ac.kr



김 광 훈 (Kwanghoon Pio Kim)

1984년 경기대학교 전자계산학과(이학사)

1986년 중앙대학교 일반대학원 전자계산학과(이학석사)

1994년 University of Colorado Boulder, Department of Computer Science (MS)

1998년 University of Colorado Boulder, Department of Computer Science (PhD)

1986년~1991년 한국전자통신연구원 TDX전전자교환기개발단 연구원

2012년~2013년, 2018년~2021년 경기대학교 전산정보원 원장

2003년 경기대학교 소상학술상 수상 / 2009년 경기대학교 우수학술상 수상

2006년 대한민국 e비즈니스대상 - 산업자원부 장관상 수상

2013년 한국인터넷정보학회 학술상 수상

1998년~현재 경기대학교 AI컴퓨터공학부 / 데이터·프로세스공학연구실 교수

2006년~현재 WiMC ERC Vice-Chair, Country(South Korea) Chair

2007년~현재 경기대학교 콘텐츠융합소프트웨어연구소(한국연구재단 이공분야 대학중점연구소) 소장

2020년~현재 경기대학교 범죄예방능동빅데이터연구소(경기대학교 대표연구소) 소장

2023년~현재 한국인터넷정보학회 회장

관심분야: 비즈니스 프로세스 모델링 방법론, 비즈니스 프로세스 관리 시스템, 프로세스 지식 발견 인텔리전스, 프로세스 마이닝, 예측적 프로세스 모니터링, 예측적 프로세스 모델링, 영상 맥락화 기술, 맥락기반 영상 검색 시스템, 영상 관제 플랫폼, 영상 관제 빅데이터 플랫폼, 영상기반 빅데이터 페브리 플랫폼, 영상기반 빅데이터 레이크 플랫폼, 데이터-집약 디지털 트윈 플랫폼, 데이터-집약 스마트공장 플랫폼, 영상기반 범죄예방·예측 서비스 플랫폼

E-mail : kwang@kgu.ac.kr