

# 토픽맵 기반의 고전문학 디지털 콘텐츠 온톨로지 설계

## A Design of Topic-map based Traditional literature's Digital Ontology

김동건\*, 정화영\*

Dong-Gun Kim\*, Hwa-Young Jeong\*

### 요 약

고전문학은 다양한 방법으로 대중화 접근을 시도하고 있었다. 이러한 예로서 디지털 아카이브 설계, 디지털 콘텐츠 설계 등을 들 수 있다. 그러나 이러한 노력에도 불구하고 고전문학은 대중화에 쉽지 않다는 한계를 갖는다. 이는 고전문학이 지니는 딱딱하고 타 영역에 비해 재미요소가 덜하기 때문일 수도 있다. 특히 고전문학의 자료나 계보를 찾는 것도 쉽지 않아 일반인들이 쉽게 고전문학을 찾고 이해하기 위한 환경 구축이 이루어지지 않았다. 본 연구는 고전문학을 디지털 콘텐츠화 하기 위하여 정보 프로파일을 이용한 온톨로지를 설계하고자 한다. 온톨로지의 각 요소들의 관계는 토픽맵을 이용하여 구축하였고, 토픽벡터를 이용하여 명세하였다.

### Abstract

Traditional culture's researcher has attempting to public use as a various method. Example is design of digital archive and digital contents. However, in spite of this effort, traditional culture's researcher has difficulty to public use. Because traditional culture is hard to understand, and less interest than the other area. Especially, traditional culture has not environment that user can searching and using the culture's information due to difficult to search the data and layers. We propose a design to make an ontology using information profile for digital contents of traditional culture. Also, we use topic-map for the factors of ontology's relation, and specify their relation using topic vector.

Key words : Ontology, Topic-map, Topic vector, Traditional literature's Digital Ontology

### I. 서 론

문학은 시대별 상황에 따라 많은 변화를 겪어왔고, 그 시대를 대변해 왔다. 문학은 시기별로 크게 현대 문학과 고전문학으로 나누는데, 특히 몇몇 장르의 고전문학은 장르에 따라 음악, 또는 표현 동작적 요소

와 결합된 복합적 텍스트로 생성되고 향수 되었다는 점에서 주로 언어로 표현된 현대문학의 텍스트와 변별된다. 고전문학은 현재의 시점에서 과거의 한 시점으로 돌아가서 성립된 것이기에 교육과 학습과정에서 현대문학과 다른 시대적 배경 등에 관해서 먼저 이해해야 할 점들이 많다 [1]. 고전문학은 오랜 역사

\* 경희대학교 후마니타스 칼리지(Humanitas College, KyungHee University)

· 제1저자 (First Author) : 김동건

· 교신저자 (Corresponding Author) : 정화영

· 투고일자 : 2012년 6월 15일

· 심사(수정)일자 : 2012년 6월 15일 (수정일자 : 2012년 8월 17일)

· 게재일자 : 2012년 8월 30일

적 전통과 다양한 삶의 모습을 담고 있는 보편적인 이야기이면서, 그것의 소재나 내용이 매우 다양하다. 또한 오랜 시간을 거쳐 전해내려 오면서 탄탄한 구성력을 갖추게 되었고, 오늘날 현대인들에게도 익숙하면서 쉽게 공감할 수 있다. 그래서 각종 문화콘텐츠의 창작 소재로 빈번히 활용되고 있다. 최근 고전문학을 활용하여 문화콘텐츠로 개발하는 사례들이 날이 갈수록 늘어나고 있다. 예컨대 드라마에서는 <쾌걸 춘향>, <쾌도 홍길동> 등이 고전소설을 현대적인 시각으로 각색하여 시청자들에게 색다른 모습을 보여 주었다. 마찬가지로 영화에서도 <춘향전>, <장화, 홍련>, <아랑> 등은 고전문학을 활용하되 패러디나 반전 기법 등 매체에 맞게 새롭게 각색하여 성공하였다. 그럼에도 고전문학과 문화콘텐츠의 연계방안에 대한 연구는 아직까지도 상당히 부족한 상황이다 [2]. 또한 고전문학의 자료 검색에서도 자료의 구성, 구체적인 설계 방법의 미비 등이 나타나고 있어 이에 대한 보완이 요구된다.

본 연구는 고전문학이 가지는 특성을 최대한 나타내면서 이를 보다 효율적으로 검색 및 관리하기 위한 디지털 콘텐츠 온톨로지를 설계하고자 한다. 이를 위하여 온톨로지의 각 속성들 사이의 연관관계를 토픽맵을 이용하여 구성하였고, 그 관계는 토픽벡터를 이용하여 명세하였다. 본 논문의 구성은 제2장에서 관련연구를 기술하고, 제3장에서는 본 논문에서 제안하는 고전문학 디지털 콘텐츠를 위한 온톨로지를 설계하고 토픽맵을 이용하여 각 속성들의 관계를 명세하였다. 이러한 결과로 제4장에서 결론을 맺는다.

## II. 관련연구

### 2-1 고전문학의 디지털 콘텐츠

고전문학은 고전이라는 특수성과 문학이라는 보편성을 함께 갖고 있으며, 고전문학의 특수성은 그것을 생산하고 수용하던 주체, 그것의 토대가 되는 사회 문화, 그리고 문학사적 전통 등 맥락의 차원에서 접근할 때 정확한 이해와 정당한 평가가 가능하다 [3]. 이에 정보통신 기술의 발달로 인하여 고전문학

도 많은 변화를 가져오게 되었으며, 고전문학의 자료들은 디지털 콘텐츠화 되어 사용자에게 제공된다. 그러나 ‘디지털 시대의 고전문학’이라는 말은 마치 지금의 옛 문학 또는 현대의 고전문학이라고 하는 것처럼 형용모순 관계에 있다. 특히 고전문학이 가치 개념으로서 ‘고전(classic)’으로 평가되는 문학이 아니라, 시간 개념으로서 고대의 문학 또는 옛 문학이라는 뜻으로 쓰이기 때문이다. 또한 문학을 문학작품의 특성으로만 이해할 경우 이러한 모순은 더욱 심화된다. 왜냐하면 ‘디지털 시대의 고전문학’을 시간 개념으로 환원하면 ‘현대의 고대 문학작품’이라고 하는 것과 같기 때문이다 [4]. 이때 콘텐츠란 사전적 의미로는 내용이나 목차를 뜻하지만, 문화산업적 측면에서는 대개 매체에 담겨 전달되는 메시지나 내용, 이야기 등을 뜻한다. 그리고 문화콘텐츠란 기존의 대중문화를 산업적 관점으로 접근하면서 나타난 신조어로, 콘텐츠를 담은 그릇이자 다양하게 활용하는 도구들을 말한다. 여기에는 출판, 만화, 방송, 영화, 애니메이션, 게임, 캐릭터, 공연, 음반, 전시, 축제, 여행, 디지털콘텐츠 등 다양한 분야가 있고, 디지털 기술이 발전함에 따라 앞으로도 계속 새로운 매체들이 나타날 것으로 보인다. 이처럼 문화콘텐츠는 그 분야가 매우 광범위하여, 원소스 멀티유즈(one source multi-use)를 통해 부가가치를 극대화하는 특징을 가지고 있다 [5]. 이러한 장점으로 인하여 고전문학의 많은 작품들이 디지털 콘텐츠로 제작되어 일반인들에게 제공되고 있다. 그러나 항상 문제는 고전은 어렵고 난해하고 낯설다는 것이다. 또한 고전문학을 원천서사로 재가공하는 과정에서 고전문학의 서사적 정체성이나 가치를 어떻게 담아낼 것인가, 어디까지의 변용을 재창작으로 허용할 것인가의 문제가 있다 [6].

### 2-2 토픽맵

토픽맵(Topic maps)은 지식 표현 방법론으로서, 분산 환경 하에서 지식 구조를 정의하고, 정의된 구조와 지식 자원을 연계하는데 쓰이는 기술 표준이며, 정보자원의 구성, 추출, 네비게이션에 관련된 새로운 패러다임이라 할 수 있다 [7]. 토픽맵은 2001년 “ISO/IEC 13250 Topic Map”이라는 명칭으로 표준화

되었다. 제정된 XTM(XML Topic Map)의 핵심 엘리먼트는 "TAO of Topic Map"라고 불리는 토픽(Topic), 연관관계(Association), 어커런스(Occurrence)로써, 지식층은 상위계층으로 토픽과 토픽간의 연관관계로 구성된다 [8]. 토픽맵의 개념적 모델은 그림 1에 보인 바와 같이 정보와 콘텐츠 자원 영역인 정보층 (information layer)과 이와 연관된 메타정보와 지식구조 표현 영역인 지식층(knowledge layer)으로 구성된 이원화된 구조를 기반으로 하고 있다.

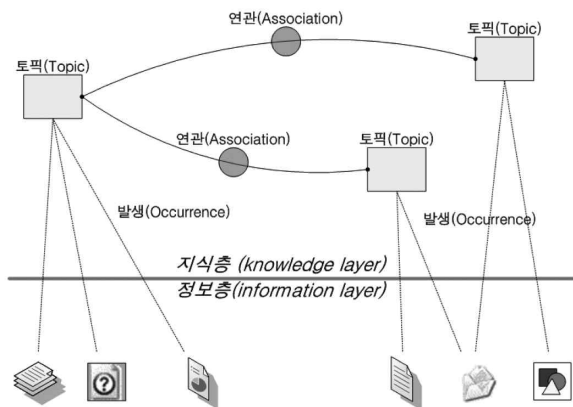


그림 1. 토픽맵 구성 요소  
Fig. 1. Construction factor of Topic-map

지식층은 주제(subject)를 표현하는 토픽(topic)과 이러한 주제 간의 연관 관계를 나타내는 연관 (association)으로 구성된다. 지식층에 토픽으로 표현된 주제와 정보 층의 정보 자원(information resource)과의 연관 관계는 발생(occurrence)으로 표현된다. 토픽은 주제(subject)를 표현하기 위한 단위이다. 존재 여부나 특정 성격에 관계없이 이에 대해 무엇이든 무슨 수단에 의해서든 서술될 수 있는 모든 대상이 주제가 될 수 있다. 토픽은 고유의 식별자를 통하여 구분되며 단일한 토픽타입(topic type)을 갖는다. 더불어 토픽은 다양한 명칭(name)들을 갖는 것이 가능하다. 발생(occurrence)은 토픽으로 표현된 주제와 연관된 정보를 갖고 있는, 디지털 콘텐츠와 같은 정보 자원과의 관계를 일컫는다. 발생은 발생타입(occurrence type)을 갖는다. 이러한 토픽맵의 핵심요소인 토픽 (Topic), 연관(Association), 발생(Occurrence)을 토픽맵의 TAO라 일컫는다 [9].

### 2-3 온톨로지

온톨로지라는 용어는 철학의 한 분야인 ‘존재론’에 기원하지만 인공지능(AI)과 지식표현(knowledge representation)에 대한 연구에 사용됨으로써 컴퓨터 과학 분야에서 널리 중요시되게 되었다 [10]. 온톨로지는 의미 정보를 나타내기 위해 필요한 개념 (concept)과 관계(relationship)를 정의한 정형적 명세 (formal specification)이다. 이는 Subject, Predicate, Object 또는 Resource, Property, Value와 같은 트리플 단위로 나타낼 수 있다 [11]. 기본적으로 온톨로지는 개념(concept)과 관계(relationship) 등의 구성요소로서 이루어지며, 자연어 또는 별도의 온톨로지 언어를 사용하여 표현된다. 정의에서와 같이 온톨로지는 특정한 관점에서 정의되거나 기술되어 관련 당사자들 간에 공유되는 어휘들로 이루어진다. 그러므로 온톨로지는 잘 정의되고 모호하지 않은 방법으로 용어의 의미를 측정 하는 도구로서 사용될 수 있다. 그러한 측면에서 온톨로지가 만들어지는 목적은 1)사람들 사이에 정확하고 모호하지 않은 의사소통을 돕고, 2)시스템들 간의 상호운용성을 달성하며, 3)시스템의 설계와 품질을 향상시키는 것 등의 3가지로 나타낼 수 있다. 온톨로지가 특정 도메인 내 혹은 이종 도메인 간의 의사소통과 상호운용성 달성의 목적으로 사용된다는 것은 하나의 공통적인 온톨로지가 존재하고 모든 참여자들이 이를 공유하여 사용하는 것을 의미한다. 현재 엔터프라이즈 모델링 분야, 생명 과학 분야, 의약분야 등과 같이 다양한 분야에서 해당 도메인을 위한 표준화된 온톨로지를 생성하고자 노력하고 있다 [10].

## III. 토픽맵 기반의 고전문학 온톨로지 설계

### 3-1 고전문학 온톨로지

많은 연구영역에서 온톨로지를 이용하고 있다. 김지영, 김윤희 [11]의 연구에서는 그리드 자원선택에 대한 방법에서 사용하였고, 정현숙, 신영숙 [12]의 연구에서는 강의계획서에 온톨로지를 이용하였다. 또

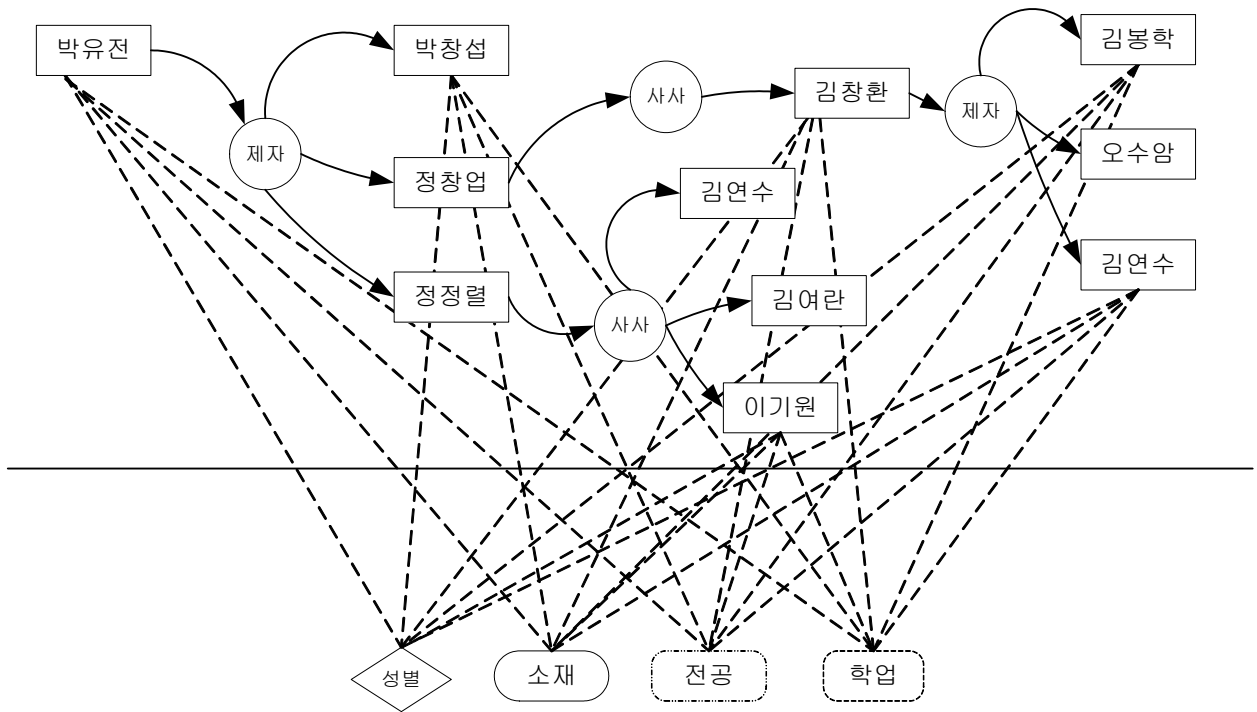


그림 3. 판소리 명창의 토픽맵  
Fig. 3. Topic-map for Pansori's actors/actress

한 나방현, 권창희 [13]의 연구에서는 역사정보 서비스에 온톨로지를 이용하였다. 본 연구에서는 고전문학의 계보에 온톨로지를 활용하였다. 고전문학의 자료는 방대한 용량에 범위가 매우 넓어 모든 정보를 온톨로지표로 표현하고 구성하기에는 한계가 있다. 이에 본 연구에서는 판소리의 명창에 대한 계보를 온톨로지표로 구성하였다. 표 1은 판소리 명창을 나타내는 정보 프로파일이다.

표 1. 판소리 명창 정보 프로파일  
Table 1 Information profile for Pansori actors/actress

명칭	설명
이름	판소리 명창의 이름
성별	남. 여
활동위치	명창의 소재
관계	명창과 명창 사이의 관계
영향	명창에 끼친 영향
전공	명창의 전공 영역

특히 관계는 명창과 다른 명창과의 스승-제자, 동문 등의 관계를 나타내며, 영향은 명창에게 영향을 받거나 준 다른 명창 또는 스승을 들 수 있다. 이를 통해 명창의 관계를 고려한 온톨로지는 그림2와 같다.

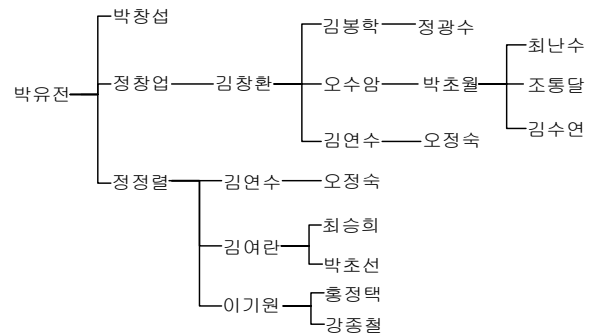


그림 2. 명창을 중심으로 한 정보 온톨로지  
Fig. 2. Information ontology for Pansori's actor/actress

3-2 고전문학, 판소리 명창 정보의 토픽맵

표 1과 그림 2를 기반으로 판소리 명창 정보 토픽맵은 그림 3과 같이 구현된다. 이를 기반으로 각 토픽간의 관계를 토픽 벡터 값을 이용하여 명세하면 아래의 식 1과 같이 표현된다.

$$T = \sum_{i=1}^n PT_i = 1 \tag{1}$$

이때 PT는 명창에 대한 토픽이 된다. 즉 토픽 벡터

값 T는 각 토픽에 대한 합이 되며 그 전체의 합은 1로 산정한다. 각 토픽에 대한 관계는 토픽내의 벡터 요소 값의 조합이 되며, 각 요소 값들은 식 2와 같이 구성된다. 즉  $P_g$ 는 명창 토픽의 성별,  $P_l$ 은 명창 토픽의 활동위치,  $P_r$ 은 관계,  $P_i$ 는 영향,  $P_m$ 은 명창의 전공을 나타낸다.

$$PT_i : PT_i = (P_g, P_l, P_r, P_i, P_m), \quad (2)$$

$$\exists_i \in \{1, 2, \dots, n\},$$

$$n = \text{Total count of } T$$

이때의 활동위치 원소값들을 통해 활동위치 토픽은 식 3과 같이 구성된다.

$$P_l \in P_{l_1}, P_{l_2}, \dots, P_{l_n} \quad (3)$$

$$P_l = \prod_{i=1}^n P_{l_i}$$

즉 명창의 활동위치 값들 중 해당 토픽에 맞는 원소 값이 선택되어지며, 이러한 방식은  $P_g, P_l, P_r, P_i, P_m$ 모두 같은 방식으로 산출된다. 이러한 연구방법의 결과는 다음과 같은 장점을 준다.

- 고전문학의 디지털 콘텐츠 화를 위한 기존의 방법인 디지털 아카이브 설계방법은 각 콘텐츠 주제들 간의 연관관계만을 명세하지만, 제안된 방법은 각 콘텐츠를 토픽으로 구분하고 토픽과 토픽간의 관계를 명세함으로써 콘텐츠가 가지는 의미를 명확히 명세할 수 있었다.
- 토픽들에 대한 정보 온톨로지를 구축함으로써 디지털아카이브가 표현하지 못하는 콘텐츠의 연관성을 명세할 수 있었다.
- 토픽맵 모델을 통하여 명창들의 상관관계를 명세함으로써 디지털 아카이브에서 나타내지 못하였던 의미기반의 검색, 주제 검색 및 관리 등에 보다 효율적으로 구축할 수 있었다.

#### IV. 결 론

본 연구에서는 토픽맵을 기반으로 한 고전문학 디

지털 온톨로지를 설계하였다. 대상 고전문학은 고전문학 영역에서도 판소리의 명창 정보를 기반으로 하였으며, 명창 정보는 판소리 명창정보 프로파일로 구성하였다. 이를 통하여 각 명창 정보들의 토픽을 구성하고, 각 토픽들의 관계를 명세하였다. 특히 명창 토픽들은 해당 토픽과 다른 토픽들과의 스승-제자, 수업-사사등의 관계를 가진다. 판소리 명창에서 이러한 정보가 가지는 의미는 매우 크기 때문에 각 관계를 고려한 토픽 모델을 선정하였으며, 각 요소 값들은 해당 요소 값들 중에서 선택되어지고, 이러한 요소 값들의 전체 합 (토픽벡터)은 1 (백분율을 1로 환산)로 산출되었다. 본 연구가 가지는 의미는 토픽들의 상관관계를 명세하고 이를 구체적으로 분석함에 따라 향후 고전문학 분야의 판소리 명창 정보를 디지털 콘텐츠 화 할 때 매우 유용한 데이터베이스 기초 자료가 된다.

향후 연구과제로는 각 토픽 값들이 명창 정보에 제한되지 않고 보다 넓은 영역의 정보를 포괄하여 전체 고전문학 영역의 디지털 콘텐츠 화 할 때 필요한 기본 자료가 생성될 수 있도록 확장하여야 한다.

#### 참 고 문 헌

- [1] 서종문, 고전문학 교육의 단위와 비중, 통합 문제, *국어교육연구* 제46집, 2010.
- [2] 윤종선, 고전문학과 문화콘텐츠 교육방법론 연구, 한 국비평문학회, *비평문학*, 제35호 2010. p. 301-332
- [3] 최광석, 고전문학 교육의 진단과 방법론적 설계, 한국어문학회, *어문학*, 제105집, 2009. p.251-279
- [4] 임재해, 디지털 시대의 고전문학과 구비문학 재인식, 국어국문학회, *국어국문학*, 제143권, 2006. p. 33-72
- [5] 윤종선, 문화콘텐츠로서 고전문학의 연구 현황과 전망, 한국어문학회, *어문학*, 제103집, 2009, p. 167-196
- [6] 함복희, 고전문학의 문화콘텐츠화, 중앙어문학회, *어문론집*, 제46집, 2011, p.91-120
- [7] 박정우, 이상훈, 사용자 프로파일을 이용한 개인화된 토픽맵 랭킹 알고리즘, *정보과학회논문지: 소프트웨어 및 응용* 제35권 제8호 2008.
- [8] 이시화, 이만형, 황대훈, 태그 기반 토픽맵 생성 시

스텝의 설계 및 구현, *한국멀티미디어학회 논문지* 제13권 제5호, 2010. p.730-739

- [9] 이정수, 채희권, 김광수, 김철한, 온톨로지의 상호 운용성을 위한 온톨로지 아키텍처에 관한 연구, 한국전자거래학회, *한국전자거래학회지*, 제13권 제1호, 2008, p.1-20
- [10] 김현주, 최증민, 온톨로지 생성과 공유를 위한 시맨틱 웹 기반 위키 시스템, *정보과학회논문지: 소프트웨어 및 응용* 제33권 제8호, 2006,
- [11] 김지영, 김윤희, 온톨로지 기반 그리드 자원 선택 기법, *한국정보기술학회논문지* 제7권 제4호, 2009. p.1-9
- [12] 정현숙, 신영숙, 온톨로지 기반 강의계획서 관리 시스템 설계 및 구현, *한국정보기술학회논문지* 제6권 제5호, 2008, p.108-114.
- [13] 나방현, 권창희, 온톨로지 기반 역사정보서비스 설계, *한국향행학회논문지*, 제14권 제2호, 2010, p.143-150

## 김 동 건



1995년 2월 : 경희대학교 국어국문학과 (문학사)

1997년 2월 : 경희대학교 국어국문학과 (문학석사)

2001년 2월 : 경희대학교 국어국문학과 (문학박사)

2011년 ~ 현재 : 경희대학교 후마니타스

칼리지 부교수

관심분야 : 고전소설, 판소리, 고전정보학

## 정 화 영 (鄭華泳)



1994년 2월 : 경희대학교 전자계산 공학과(공학석사)

2004년 8월 : 경희대학교 전자계산 공학과(공학박사)

1994년 3월~1998년 12월 : (주)아주 시스템 기술연구소 전임연구원

2000년 3월~2005년 2월 : 예원예술 대학교 게임영상학부/정보경영학부 조교수

2005년 4월~2011년 2월 : 경희대학교 교양학부 조교수

2011년 3월~현재 : 경희대학교 후마니타스 칼리지 조교수

관심분야 : 소프트웨어 공학, 웹 공학, 웹 서비스