Journal of Convergence for Information Technology
Vol. 10. No. 1, pp. 139-148, 2020
e-ISSN 2586-4440
DOI : https://doi.org/10.22156/CS4SMB.2020.10.01.139

RDF 기반 온톨로지 검색 서비스 플랫폼 구현에 관한 연구

신유탁[1], 조재춘[2*]
[1](주)피씨엔 연구소 과장, [2]상명대학교 스마트정보통신공학과 교수

A Study on the Implementation of Ontology Retrieval Service Platform Based on RDF

Yutak Shin[1], Jaechoon Jo[2*]
[1]Manager, R&D center, PCN Co., Ltd.
[2]Professor, Dept. of Smart Information and Communication Engineering, Sangmyung University

요 약 인터넷과 컴퓨터 기술이 발전되면서 수많은 콘텐츠 속에서 문화와 역사, 전통 관련 자료를 효과적으로 검색하고 창작에 도움이 될 수 있는 전통문화 융·복합 지원이 필요하다. 본 연구는 이를 위한 RDF 기반 온톨로지 검색 서비스 플랫폼을 구축하고 사용성 및 타당성 검증을 수행하였다. 본 플랫폼은 트리플 검색, 키워드 간 관계 검색, 네트워크 그래프 검색, 스토리 탐색 및 관리, 큐레이션 관리 모듈로 구분된다. 이를 기반으로 데이터 간의 관계를 중심으로 검색 결과를 시각화하고 네트워크 그래프 검색과 스토리 검색을 통해 키워드 간의 관계를 쉽게 파악하고 키워드 간의 복잡한 관계 분석 및 검색 시간을 단축시킬 수 있다. 검증을 위해 자문 평가를 수행하였고 기존의 검색 기능보다 정보 간의 관계를 쉽게 파악하고 분석 및 검색 시간을 획기적으로 단축시킬 수 있는 향상된 검색이 가능하다는 평가를 받았다.

주제어 : 온톨로지, RDF, 트리플 데이터, 정보 검색, 스토리텔링, 큐레이션

Abstract As the internet and computer technology are developed, there is a need for service of traditional culture that can effectively search and create culture, history, and tradition-related materials in online contents. In this paper, we developed an RDF-based ontology retrieval service platform and verified usability and validity. This platform is divided into triple search, keyword search, network graph search, story search and management, curation management module. Based on this, the search results can be visualized based on the relationship between data, network graph search and story search can be used to easily understand the relationship between the keywords. An platform evaluation was conducted for verification, and it was evaluated that an intelligent search that can easily identify the relationship between information and shorten the analysis and search time than the existing search function.

Key Words : Ontology, RDF, Triple Data, Information Retrieval, Storytelling, Curation

*This research is supported by Ministry of Culture, Sport and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research&Development Program 2018. (No. R2017030045)
*Corresponding Author : Jaechoon Jo(jae@smu.ac.kr)
Received November 28, 2019 Revised December 30, 2019
Accepted January 20, 2020 Published January 28, 2020

1. 서론

특정 콘텐츠에 대한 정보를 탐색하기 위해서는 검색 서비스를 이용하며, 이러한 검색 서비스는 기본적으로 사용자로부터 검색 키워드를 입력받고, 입력받은 검색 키워드와 관련도가 높은 정보를 결과로 제공한다. 하지만 기존의 검색 서비스는 단순히 검색 키워드에 대한 일차원적이고 단편적인 정보만을 제공하였다. 예를 들어 기존의 검색 서비스는 '이순신'을 검색할 경우, 문장에 '이순신'이 포함된 모든 정보를 단순히 나열하는 수준에 불과하였다. 또한 '이순신'과 '원균'의 관계를 살펴보고자 할 때 사용자는 '이순신'과 '원균'을 각각 검색하고 검색 결과에 대한 단순 정보를 직접 취합하여야 했다. 따라서 검색 키워드 간의 관계를 파악하고자 할 경우 이를 쉽게 파악할 수 없어 많은 시간이 소요되는 문제가 있었다.

한편 현재 국내 검색 서비스는 전통문화 콘텐츠의 분류체계에 따른 연계 등이 부족하여 확장성이 미흡한 것으로 파악되고 있으며 체계적인 코드화를 통한 다양한 맞춤형 정보 제공 서비스가 요구되고 있다. 또한, 전통문화 연구자와 콘텐츠 창작자 간의 소통 채널 확대가 필요하다. 이를 위해 협업 인프라를 구축하고 다양한 분야에 활용도를 높이기 위해 전통문화 융·복합을 지원할 수 있는 플랫폼 구축이 필요하다.

이에 본 연구는 전통문화 데이터를 활용하여 RDF 기반 온톨로지 검색 서비스 및 플랫폼을 구축하였다. 또한, RDF 트리플 데이터를 이용하여 데이터 간의 관계를 중심으로 시각화하고, 복잡한 관계를 더 쉽고 빠르게 파악할 수 있도록 네트워크 그래프 검색과 스토리(Story) 검색을 제공하였다. 이를 통해 검색 키워드 또는 데이터 간의 관계를 시각화하여 정보를 쉽게 파악할 수 있도록 하여 분석시간을 획기적으로 단축할 수 있었고, 스토리 검색 및 편집을 기반으로 다양한 큐레이션 창작 및 공유가 가능하도록 하였다. 또한, 본 연구에서 제안하는 RDF 기반 온톨로지 검색 서비스 및 플랫폼의 평가를 시행하기 위해 전문가 자문 평가를 실시하였다. 이를 통해 관련 전문가, 시스템 개발자 등의 측면에서의 타당성 평가를 고려할 수 있도록 평가 기준과 요인을 도출하였다. 한편, 본 연구의 결과는 향후 다양한 데이터를 적용하여 전통문화 이외의 검색 서비스에서도 정보 간의 관계 및 분석 분야에서 다양하게 활용될 수 있을 것으로 보인다.

2. 관련연구

2.1 시맨틱 웹

시맨틱 웹(Semantic web)은 인터넷(internet)과 같은 분산 환경에서 자원(웹 문서, 서비스, 각종 파일 등) 사이의 관계와 정보의 관계에 대한 의미 정보를 컴퓨터가 처리할 수 있도록 온톨로지(Ontology) 형태로 표현하고, 이를 컴퓨터가 처리하도록 하는 프레임워크 기술이다. 즉, 웹(Web)에 있는 정보의 의미를 컴퓨터가 이해할 수 있는 형태로 정의하여 사람의 직접적인 개입이 없어도 웹의 정보가 자동처리 가능하도록 하는 확장된 새로운 웹이다. 또한, 시맨틱 웹의 궁극적 목적은 웹의 정보를 컴퓨터가 빠르게 이해할 수 있도록 도와주는 기술과 표준을 개발하여 시맨틱 검색 또는 데이터 통합 등을 지원하는 것이다[1].

한편 인터넷이 보편화되고 디지털 콘텐츠가 날로 증가함에 따라 단순히 키워드 중심의 검색 방식에서 벗어나 보다 정교하고 표현되지 않은 사용자의 정보 검색 의도를 도출해낼 수 있는 시맨틱 검색을 필요로 하고 있다. 기존의 검색이 텍스트 또는 문서에서 키워드를 매칭하는 방식이라면, 객체(Objects) 검색 방식이 시맨틱 검색의 원리이다. 다시 말해 시맨틱 알고리즘이 적용된 검색 결과는 단순히 URL이 아닌 객체의 개념(Concepts), 관계(Relations), 속성(Properties), 개체(Instances)가 포함된 상대적으로 고도화된 정보를 제공한다는 것이다[2]. 즉, 시맨틱 검색은 사용자가 문서에서 해당 키워드가 의미하는 하나 또는 관계된 개념을 찾고자 할 때 유용하다[3]. 또한 키워드나 문서가 연계된 메타데이터가 적절하게 부여될 경우 시맨틱 검색 알고리즘은 보다 정교한 검색이 이루어질 수 있다. 다시 말해 시맨틱 웹에서 이러한 기능들을 지원하기 위해서는 웹 문서 내에 지식 표현을 위한 온톨로지와 지식 또는 개념 간의 관계를 설정하며 추론 규칙이 포함되어야 한다[4].

2.2 온톨로지

시맨틱 웹은 기존 웹이 확장된 온톨로지 기술에 기반하고 있다. 또한, 시맨틱 검색을 위해서는 문서나 텍스트의 의미를 정의할 수 있는 온톨로지가 필요하다. 온톨로지는 그리스어 Ontos(존재)와 Logos(단어)에서 유래한 것으로 존재하는 것의 기본적인 범주를 연구하는 학문이다. 원래 온톨로지는 철학에서 특정 분야에 사용되

는 용어의 집합 또는 개념을 표현하기 위해 사용되는 용어의 집합을 의미하였다. 그리고 개념 범주, 유형 등을 분류하는 분류체계와 동의어로 활용되었다. 한편 온톨로지는 90년대 초부터 자연어처리(Natural language processing), 지식공학(Knowledge engineering), 지식표현 Knowledge representation) 등 다양한 인공지능 분야의 주요 연구 소재가 되어 왔다. 그뿐만 아니라 최근에는 전자상거래, 지식경영, 지능형 정보 검색 분야에서도 활발하게 연구되고 있다.

온톨로지는 분야에 따라 다양하게 정의되는데 본 논문에서는 특정 분야에 사용되는 용어 간의 관계와 정의를 용어의 조합 규칙, 용어의 확장에 대한 관계를 정의한 것을 의미한다[5]. Gruber는 “공유된 개념화(Shared conceptualizatiuon)에 대한 정형화되고 명시적인 명세(Formal and explicit specification)”라고 정의하였으며[6], Rudi Studer 등도 온톨로지를 공유된 개념화의 형식적, 명시적 명세로 정의하고 있다. 여기서 ‘공유’는 “온톨로지가 합의된 지식(consensual knowledge)을 표현해야 한다는 것”을 의미한다[7]. ‘합의된 지식’이란 몇몇 개인이 임의로 정한 것이 아닌 관련된 구성원 모두의 동의를 얻어 수용된 개념으로서 이를 통해 각각의 개념 간의 관계를 표현한 지식을 말한다. ‘개념화’란 특정 분야 또는 특정 영역의 실세계와 관련된 개념을 나타내는 추상 모델을 말한다. ‘형식적’이란 온톨로지의 내용을 컴퓨터가 읽고 처리 가능한 형태로 표현해야 한다는 뜻이다. ‘명시적’이란 특정 분야 또는 영역을 모델링 (추상화)할 때 사용되는 개념들과 제약조건들을 명시적으로 정의해야 한다는 것을 의미한다[8]. 즉, 온톨로지는 개념 또는 지식을 표현하는 용어에 관한 정의와 용어 간의 상호관계를 명시함으로써 모델링 하고자 하는 실세계에 대한 체계적이고 논리적인 틀을 제공한다.

2.3 RDF

기존의 월드와이드웹(World Wide Web)과는 달리 시맨틱 웹(Semantic Web)상에서는 정보의 자원(Resource)들의 의미가 정의되어 있고, 이들 간의 의미적 연결을 지원한다. 기계도 이해할 수 있는 차세대 지능형 웹인 시맨틱 웹의 출현은 새로운 종류의 구조화된 데이터인 RDF(Resource Description Framework) 데이터가 출현하게 된 계기가 되었다[9]. RDF는 컴퓨터가 이해할 수 있도록 자원(Resource)에 대한 정보 또는 속성을 그래프 구조를 통해 기술하는 언어이다. 여기서 자원은 URI(Uniform Resource Identifier)를 부여할 수 있는 모든 개체, 개념을 의미한다. 다시 말해 자원이란 웹 페이지, 문서, 데이터베이스 혹은 다양한 식별 가능한 대상을 의미한다. 세상에 존재하는 모든 것, 사람이 만들어 놓은 모든 것, 개념적으로 생각할 수 있는 모든 것들이 자원이 될 수 있다. 설명(Description)은 자원을 상세하게 기술한 것을 의미하며 프레임워크 (Framework)는 틀 또는 표현방식 등으로 이해할 수 있다.

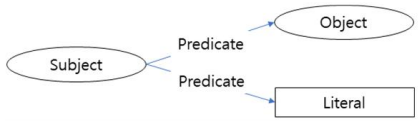

Fig. 1. RDF Data Model

Fig. 1과 같이 RDF는 기본적으로 주어(Subject), 술어(Predicate), 목적어(Object)의 트리플(Triple) 구조로 표현된다. 목적어에는 문자열(Literal)도 올 수 있다. 문자열은 통합 자원 식별자(Uniform Resource Identifier: URI)가 부여될 수 없는 문자열 자체를 말한다. 주어란 표현하고자 하는 데이터를 의미하며, 술어는 주어에 대해 기술하거나 주어와 목적어의 관계를 의미한다. 목적어란 술어에 대한 내용이나 값을 의미한다. Subject-Predicate-Object 트리플을 RDF 트리플이라고 정의하고 있고, RDF 데이터 모델에서 정보를 표현하는 최소 단위이다. 어떤 특정 도메인의 지식을 모델링할 때, 복잡한 지식도 다수의 RDF 트리플로 표현할 수 있어야 한다. 또한 각 내용에 대해서 URI를 통해 기술할 수 있다. 예를 들어 “Ora Lassila는 자원 http://www.w3.org /Home/Lassila의 작성자이다.”는 내용을 기술한다면 각 내용은 다음 Table 1과 같이 표현된다[10].

Table 1에서 표현된 모델은 다시 RDF 모델을 기술하는 트리플 그래프 모델로 표현할 수 있다. 그래프 모델에서 주어는 원형으로, 술어는 간선으로, 목적어는 URI일 경우 다원으로 표현되며 Literal일 경우 사각형으로 표현된다. Fig. 2는 Table 1의 내용을 RDF 트리플 그래프 모델로 나타내고 있다.

Table 1. RDF Triple representation

subject	http://sample.com/KIM
predicate	http://sample.com/term/creation
object	“Ontology”

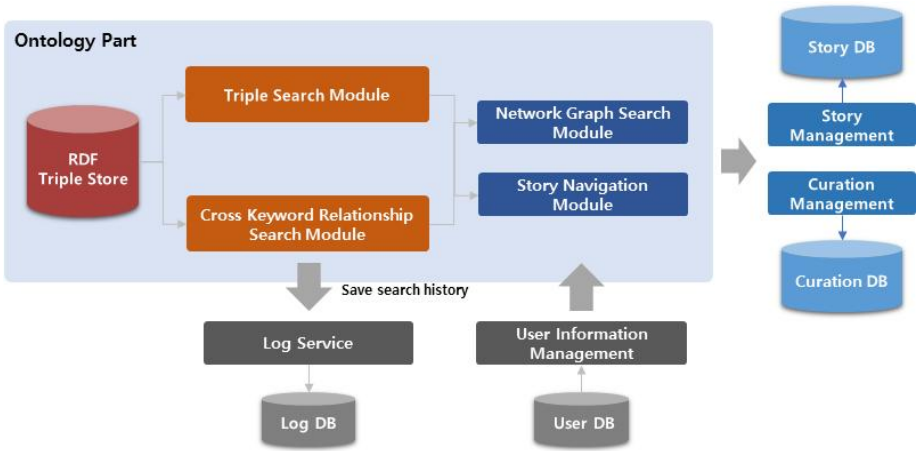

Fig. 3. Service Platform Configuration Diagram

RDF는 트리플을 연속적으로 기술함으로써 정보를 표현한다. 또한, 실제 데이터를 기술하는 것뿐만 아니라 데이터에서 사용되는 용어의 종류와 용어간의 관계까지 표현할 수 있다. 최근에는 SPARQL 질의 결과를 Daum Map API와 Highcharts API를 사용하여 웹으로 표현하거나[11], Record In Context(RIC) 개념모델과 온톨로지 검색을 구현하는 인터페이스를 제안하는 연구[12], RDF 표현기반의 사용자 중심 검색 시스템 구현[13], 문서와 토픽 키워드 간의 관계와 문서 간의 의미 관계를 자동으로 추출하는 토픽 모델링을 활용한 온톨로지 자동 구축[14] 등 다양한 연구가 아직까지 진행되고 있다.

Fig. 2. Graph representation for RDF models

3. RDF기반 온톨로지 검색 서비스

본 연구의 검색 서비스는 구축된 온톨로지의 RDF 데이터를 기반으로 데이터 간의 관계를 중심으로 시각화 하고, 복잡한 관계를 더 쉽고 빠르게 파악할 수 있도록 이야기(Story)를 탐색할 수 있는 네트워크 그래프를 활용한 지식 검색 서비스를 제공한다. 이를 통해 데이터 간의 관계를 시각화하여 관계를 쉽게 파악할 수 있고 분석시간을 획기적으로 단축할 수 있다. 해당 연구의 검색 서비스 플랫폼은 Fig. 3과 같이 구성되며, Table 2는 구현된 검색 서비스 플랫폼의 개발 정보와 최소 권장 사항을 나타낸다.

Table 2. Recommend System Requirements

Runtime Environment	JDK 1.7 or higher
Web container	Servlet 2.3 or higher
Operating system	Any OS (Windows, Unix, Linux, etc)
Memory	526MB or higher
Free Space	500MB or higher (depend on data volume)
Database	MongoDB, TDB(Apache Jena)

본 연구를 통해 구현된 검색 서비스는 Fig. 3과 같이 저장된 트리플 데이터를 기반으로 검색을 수행하는 온톨로지 영역과 검색 이력을 관리하는 로그 서비스, 사용자 정보를 관리하는 사용자 정보 관리, 검색 결과를 기반으로 형성되는 스토리를 관리하는 스토리 관리, 스토리를 기반으로 사용자에게 스토리 제작에 대한 큐레이션을 제공하는 큐레이션 관리로 구성된다.

온톨로지 영역은 트리플 데이터를 기반으로 사용자로부터 입력받은 검색 키워드의 결과를 네트워크 그래프로 결과 화면을 보여준다. 이를 위해서 트리플 데이터가 저장되는 트리플 스토어와 검색하고자 하는 키워드를 기반으로 트리플 스토어에서 검색을 수행하는 트리플 검색 모듈, 키워드 간의 관계를 검색하는 키워드 간 관계 검색 모듈, 네트워크 그래프 검색 모듈 및 스토리 탐색 모듈을 포함한다.

큐레이션 관리는 생성된 다수의 스토리를 병합하여 하나의 스토리를 만들 수 있는 기능으로 현재 저장되어 있는 스토리와 연결 가능한 노드가 있는 스토리를 추천하여 연결할 스토리를 사용자가 선택하게 하여 다양한 스토리 큐레이션 제작을 지원한다.

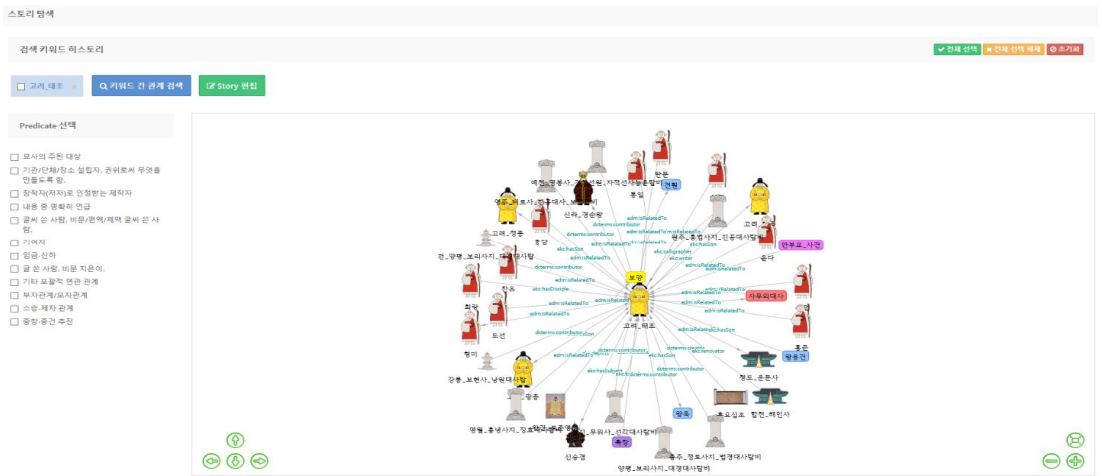

Fig. 5. Triple Data Search

3.1 트리플 스토어

트리플 스토어는 트리플 데이터의 저장 및 검색을 위한 데이터베이스로써, 의미 있는 문장들을 저장하는 그래프 데이터베이스이다. 트리플(triple) 구조는 주어와 술어 및 목적어로 구성된 구조를 의미하며, 트리플 데이터는 Fig. 4와 같이 주어와 목적어 그리고 주어와 목적어 간의 관계인 술어를 포함하는 문장의 형태이다.

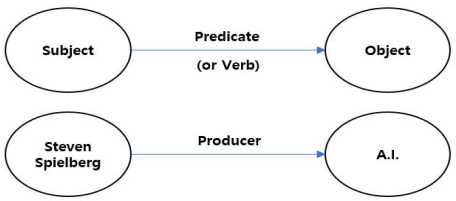

Fig. 4. RDF Triple Data Structure

예를 들어 트리플 데이터가 '스티븐 스필버그는 A.I.의 제작자이다.' 일 경우, 스티븐 스필버그는 주어이고 A.I.는 목적어이다. 제작자는 술어로서, 주어인 스티븐 스필버그와 목적어인 A.I.의 관계를 나타낸다. 여기서 주어인 스티븐 스필버그에 대한 목적어와 술어가 포함된 트리플 데이터는 여러 개가 존재할 수 있다. 예를 들어 '스티븐 스필버그는 A.I.의 제작자이다.' 외에도, '스티븐 스필버그는 레디 플레이 원의 감독이다.', '스티븐 스필버그는 레디 플레이 원의 제작자이다.', '스티븐 스필버그는 라이언 일병 구하기의 감독이다.' 등을 포함할 수 있다. 즉, Fig. 5와 같이 한 개의 주어가 다수의 목적어와 술어가 포함된 트리플 데이터를 가질 수 있다. 또한, 전술된 트리플 데이터는 관리자가 다수의 데이터를 트리플 스토어에 입력함으로써 목록화하여 데이터베이스화할 수 있다.

3.2 트리플 검색 모듈

트리플 검색 모듈은 트리플 스토어에서 트리플 데이터 목록을 조회한다. 사용자로부터 입력받은 검색 키워드를 주어로 하여 트리플 데이터 목록에서 트리플 데이터를 검색한다. 또한, 트리플 검색 모듈은 검색 키워드를 주어뿐 아니라 목적어와 술어에 대해서도 검색할 수 있다. 예를 들어 검색 키워드가 '스티븐 스필버그'일 경우, 트리플 검색 모듈은 '스티븐 스필버그는 A.I.의 제작자이다.'와 같이 주어가 스티븐 스필버그인 트리플 데이터와 '레디 플레이 원의 제작자는 스티븐 스필버그이다.'와 같이 목적어가 '스티븐 스필버그'인 트리플 데이터를 검색할 수 있다. 또한, 트리플 검색 모듈은 '더 라스트 건은 스티븐 스필버그의 영화이다.'와 같이 술어 자리에 '스티븐 스필버그'가 포함된 트리플 데이터도 유효 데이터로 조회할 수 있다. 즉, 트리플 검색 모듈은 사용자로부터 검색 키워드가 포함된 모든 트리플 데이터를 유효 데이터로 조회할 수 있다. 그리고 트리플 검색 모듈은 조회된 트리플 데이터 목록을 Fig. 5와 같이 노드(node)와 간선(edge) 형태로 형성한다. 간선은 술어에 해당하며 노드는 주어와 목적어에 해당한다. 노드는 네트워크 그래프 형태에서 주어와 목적어를 상징할 수 있는 이미지 등으로 표현되며 간선은 노드 사이를 연결하는 선으로 표현된다. 연결된 노드 사이의 관계인 술어는 텍스트로 표현된다.

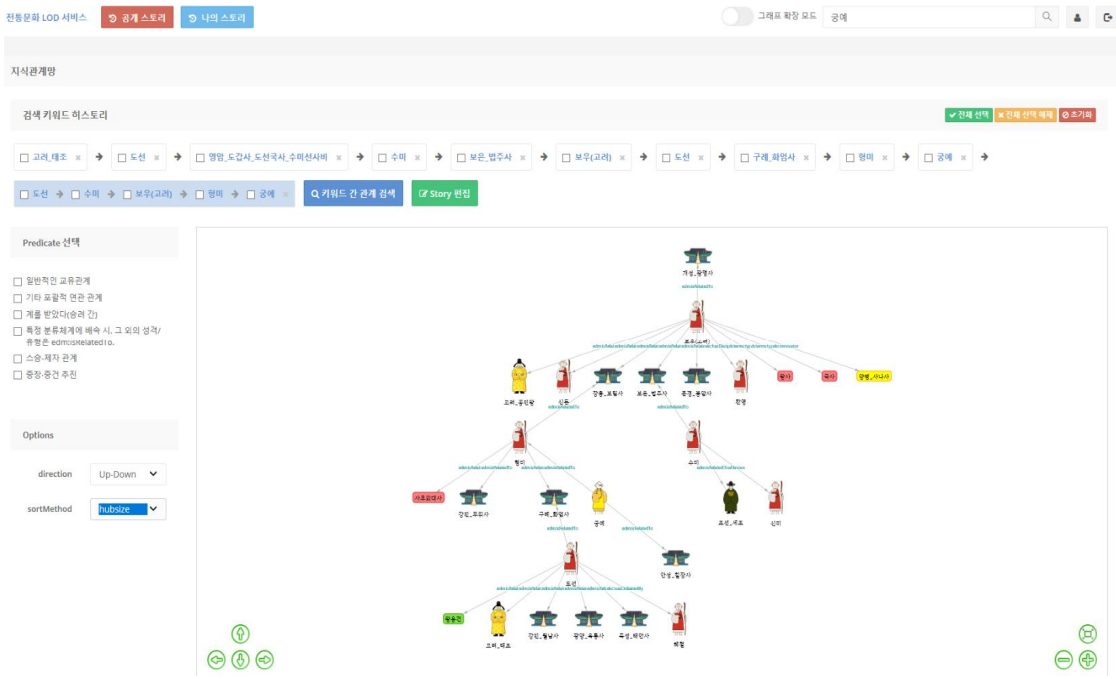

Fig. 6. Search for relationships among keywords and search for network graphs

3.3 키워드 간 관계 검색

트리플 키워드 간 관계 검색 모듈은 트리플 검색 모듈이 사용자로부터 두 개 이상의 검색 키워드를 입력받았을 경우 두 개 이상의 검색 키워드 간의 관계를 검색한다. 사용자로부터 검색 키워드를 입력받으면, 검색 키워드와 관련된 트리플 데이터 목록을 조회한 후 네트워크 그래프 검색 모듈을 통해 네트워크 그래프 형태로 표현한다. 또한, 트리플 검색 모듈이 사용자로부터 새로운 검색 키워드를 입력받으면, 트리플 검색 모듈은 관련된 트리플 데이터 목록을 네트워크 그래프 형태에 검색 결과를 추가하여 표현한다.

이때 Fig. 6와 같이 사용자로부터 입력받은 검색 키워드는 [검색키워드 히스토리]에 표시된다. [검색 키워드 히스토리]에는 그룹 형태로 구분되어 있다. 단일 검색 키워드는 단일 검색 키워드가 하나의 그룹 형태가 되며, 관계 검색을 위해 선택된 다수의 검색 키워드도 하나의 그룹 형태가 된다. 그룹의 형태가 된 검색 키워드는 추후 사용자가 재사용할 수 있다.

사용자는 [검색 키워드 히스토리]에서 관계를 조회하고자 하는 검색 키워드를 다중 선택한 후 [키워드 간 관계 검색] 버튼을 눌러 검색 키워드 간의 관계 검색을 할 수 있다. 예를 들어 트리플 검색 모듈이 사용자로부터 '이순신'과 '원균', 그리고 '거북선'을 검색 키워드로 입력받았을 경우, 우선 키워드 간 관계 검색 모듈은 '이순신'과 '원균' 각각에 대한 트리플 데이터 목록을 기반으로 주어와 목적어 노드(node) 목록과 술어 간선(edge) 목록을 생성한다. 이후 '이순신'과 '원균' 사이의 관계 간선 목록을 생성한다. 노드 목록과 간선 목록을 생성하기 전 중복은 제거된다. 이후 네트워크 그래프 검색 모듈은 생성된 관계 간선 목록을 이전에 생성된 '이순신'과 '원균' 각각에 대한 노드 목록 및 간선 목록과 함께 네트워크 그래프 형태로 표현한다. 키워드 간 관계 검색 모듈에 의해서 직접적인 관계가 없는 검색 키워드 중에서 의미 있는 관계를 찾아낼 수 있다. 또한, 키워드 간 관계 검색 모듈은 사용자로부터 검색 키워드의 우선순위(가중치)를 입력받을 수도 있다. 이 경우 키워드 간 관계 검색 모듈은 사용자가 선택하는 검색 키워드의 순서에 따라 가중치를 높게 주어 가중치가 높은 검색 키워드 순서대로 네트워크 그래프의 중심에 오도록 할 수 있으며, 가중치가 높은 검색 키워드와 관련된 트리플 데이터는 모두 표시하고 가중치가 낮을수록 검색 키워드와 관련이 낮은 트리플 데이터는 표시하지 않도록 할 수 있다.

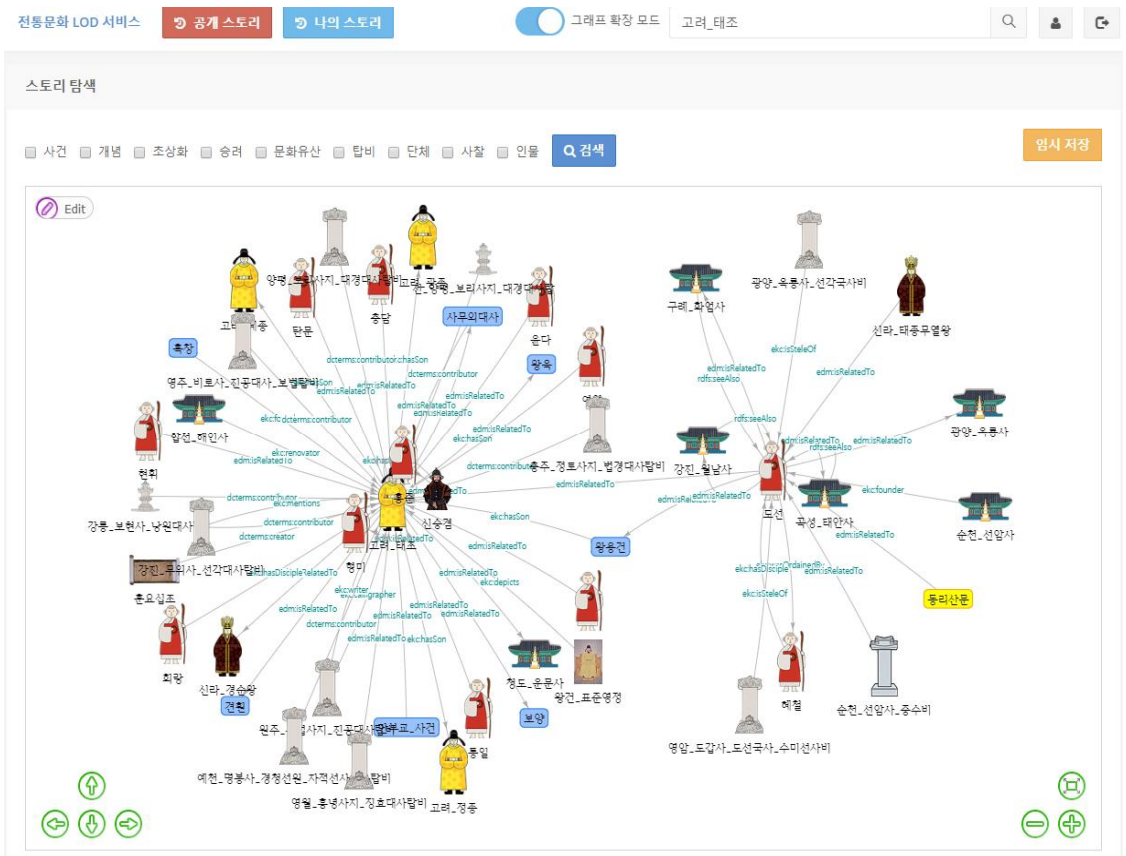

Fig. 7. Extended Network Graphs

3.4 네트워크 그래프 검색 및 스토리 탐색

네트워크 그래프 검색 모듈은 트리플 검색 모듈로부터 생성된 노드(node) 목록과 간선(edge) 목록 또는 키워드 간 관계 검색 모듈로부터 생성된 관계 간선 목록을 네트워크 그래프 형태로 표현한다. 이는 검색 키워드 입력창을 통해 사용자로부터 검색 키워드를 입력받으면 트리플 검색 모듈이 해당 검색 키워드에 대한 트리플 데이터 목록을 트리플 스토어에서 조회하고, 네트워크 그래프 검색 모듈은 조회된 트리플 데이터로부터 생성된 노드 목록과 간선 목록을 기반으로 네트워크 그래프를 표현한다. 물론 사용자로부터 다중 검색 키워드를 입력 받고 이들의 관계 조회 명령을 입력받았을 경우, 네트워크 그래프 검색 모듈은 노드 목록과 간선 목록 및 관계 간선 목록을 기반으로 네트워크 그래프를 표현한다.

스토리 데이터는 스토리 탐색 모듈은 검색 키워드에 대해서 사용자가 스토리를 만들 수 있도록 지원한다. 이를 위해 네트워크 그래프에서 노드와 간선 및 관계 정보 중 의미 있는 정보만을 남기고 노드를 확장 검색함으로써 스토리를 탐색한다. 또한, 스토리 탐색을 위해 네트워크 그래프 형태로 구현된 노드와 간선 중 하나를 추가하거나 수정, 삭제할 수 있다. 즉, 스토리 탐색 모듈은 트리플 검색 모듈로부터 생성된 노드 목록과 간선 목록, 키워드 간 관계 검색 모듈로부터 생성된 관계 간선 목록을 Fig. 7과 같이 확장된 네트워크 그래프 형태로 표현할 수 있다.

Fig. 7과 같이 [그래프 확장 모드]는 스토리 탐색 시 기존 검색 키워드를 중심으로 네트워크 그래프를 표현할지 새로운 검색 키워드를 중심으로 네트워크 그래프를 표현할지를 결정한다. [그래프 확장 모드]는 사용자가 온(on)/오프(off)를 선택할 수 있다.

검색 키워드에 대한 네트워크 그래프를 표현할 때, 특정 노드를 더블 클릭할 경우 새로운 검색 키워드를 중심으로 네트워크 그래프를 표현할 수 있다. 또한, 검

색 키워드에 대한 트리플 데이터 중 술어만 중복을 제거하여 필터링할 수 있는 옵션을 제공한다. 특정 술어를 선택하면 해당 술어와 관계된 노드를 중심으로 필터링하여 네트워크 그래프를 재구성할 수 있다. 예를 들어 인물인 경우 직계친족, 형제, 부녀관계, 모녀관계, 창작자(저자)로 인정받는 제작자, 묘사의 주된 대상 등을 포함할 수 있다.

또한, 네트워크 그래프 이동 버튼을 이용하여 네트워크 그래프를 확대하거나 축소할 수 있다. 여기서 네트워크 그래프 확대/축소 기능뿐 아니라 자동 크기 조절 기능도 구비 할 수 있으며, 이 경우 자동 크기 조절 기능은 네트워크 그래프 영역 크기에 맞게 자동으로 크기를 조절하는 기능을 수행할 수 있다.

스토리 탐색 모듈은 검색된 네트워크 그래프에 대한 스토리를 두 가지 방식으로 저장할 수 있다. 첫 번째는 네트워크 그래프를 스냅샷(Snapshot) 형태로 스토리 데이터베이스에 저장하는 [임시 저장]과, 두 번째는 스냅샷 형태로 임시 저장된 정보에 스토리 메타 정보를 입력하여 스토리 데이터베이스에 저장하는 [Story 저장] 방식이 있다.

3.5 스토리 관리 및 큐레이션 관리

스토리 관리 모듈은 사용자가 작성한 스토리 정보를 저장하고 관리한다. 이를 위해서 스토리 관리 모듈은 탐색과 추가, 수정 및 삭제를 통하여 노드(node) 목록과 간선(edge) 목록 또는 노드 목록과 간선 목록 및 관계 간선 목록으로 만들어진 스토리 스냅샷(Snapshot) 정보와 스토리 메타 정보를 스토리 데이터베이스(DB : Data Base)에 저장하여 관리한다. 여기서 스토리 스냅샷 정보는 노드 목록과 간선 목록이 되고 기본 키(Primary key)는 사용자 아이디(user_id)와 스냅샷 아이디(snapshot_id)를 가지는 복합키로 구성된다. 또한, 스토리 메타 정보는 스토리의 제목, 시놉시스, 공개 여부, 만든 사람과 같은 정보를 포함하며, 공개 여부를 공개로 설정할 경우 다른 사용자들도 저장된 스토리 정보를 볼 수 있다. 또한, 다른 사용자가 만든 스토리를 바탕으로 새로운 스토리를 생성할 수도 있으며, 이 경우 바탕이 된 스토리가 레퍼런스(참조) 스토리가 된다.

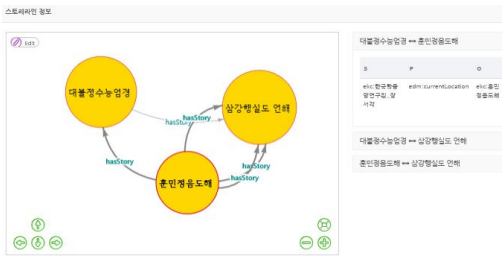

Fig. 8. Story created by a bundle of episodes

큐레이션 관리 모듈은 생성된 다수의 스토리를 병합하여 하나의 통합 스토리로 만들 수 있는 기능을 제공한다. 예를 들어 Fig. 8과 같이 저장된 각각의 스토리를 하나의 네트워크 그래프로 연결함으로써 통합된 하나의 스토리로 생성할 수 있다. 이를 위해서 현재의 스토리와 연결 가능한 노드가 있는 스토리를 추천하여 연결할 스토리를 사용자가 선택할 수 있도록 하고 각각의 스토리는 연결 고리 즉, 서로 연결되는 노드가 있어야 한다. 통합된 스토리는 큐레이션 데이터베이스에 저장하여 추가와 수정 및 삭제하는 관리 기능을 제공한다. 이를 위해 큐레이션 관리 모듈은 통합된 스토리를 큐레이션 데이터베이스에 저장하여 추가와 수정 및 삭제하여 관리하는 큐레이션 관리 모듈과 통합된 스토리가 저장되는 큐레이션 데이터베이스를 포함한다.

4. 평가 및 결과

본 연구에서 제안하는 RDF 기반 온톨로지 기반 검색 서비스의 평가를 위해 전문가 자문 평가를 시행하였다. 이를 통해 서비스에 대한 타당성 및 사용성 검증을 입증할 수 있었다. 또한, 전문가 자문 평가를 통해 서비스 기능, 서비스 내용, 설계, 화면 UI/UX 등의 사용성 평가뿐만 아니라 전통문화 관련 전문가, 시스템 개발자 등의 측면에서의 타당성 평가를 고려할 수 있도록 평가 기준과 요인을 도출하였다.

본 연구의 도출된 평가 기준과 요인을 통해 사용성 및 타당성 검증을 위한 평가 영역은 다음과 같다. 첫째, 전통문화 융·복합 지원을 위한 RDF 트리플 데이터를 활용한 온톨로지 검색 서비스는 일반적인 동작 방식과 기대에 맞게 기능들이 작동되어야 한다. 둘째, 단순한 키워드 중심의 검색 결과를 전달하는 것을 넘어서 사용자에게 의미 있는 정보를 효과적으로 전달하고 전통문화와 관련하여 융·복합 지원이 가능해야 한다. 따라서

각 정보 간의 관계를 제공하고 유의미하게 관련된 정보를 지능적으로 제공하고 사용자가 개별적으로 관리할 수 있도록 기능적인 면을 제공해야 한다. 셋째, 기술적 안정성이 제공되어야 한다. 사용하기 편리한 UI/UX와 지능적인 검색 결과를 제공할 수 있는 플랫폼이라도 기술적인 기능이 제대로 구현되지 않으면 사용성에 상당한 지장을 초래할 수 있다. 즉, UI/UX 영역, 검색 기술 영역, 플랫폼 안정성 영역, 3가지 영역으로 평가 기준을 설정하였다[15].

앞서 도출한 평가 기준을 토대로 개발된 RDF 기반 온톨로지 검색 서비스의 검증을 위한 전문가 자문 평가를 시행하였다. 전문가 자문 평가 위원은 전통문화 관련 창작 경험이 있는 전문가 3명, 전통문화 관련 연구 경험이 있는 전문가 2명, 시스템 개발 전문가 2명, 지능형 검색 및 컴퓨터 공학 분야 전문가 2명으로 총 9명으로 구성되었으며 2차례에 걸쳐 평가가 시행되었다. 본 평가에서는 해당 분야에 5년 이상의 직무 또는 연구 경력을 가지고 있는 자를 전문가로 정의하였다.

전문가 자문 평가 결과, UI/UX 영역에서는 사용자 의지에 따라 쉽게 페이지 이동이 가능하고 검색, 수정, 저장, 필터링과 같은 기능 동작에 대한 통제가 가능하게 하는 사용자 통제성이 적절하여 플랫폼 사용에 불편없이 효과적으로 검색을 수행할 수 있다고 평가하였다. 또한, 화면 배치, 아이콘의 일반성, 스타일과 텍스트 가시성 등 시각적, 기능적, 인지적 요소가 적절하게 설계되었다고 평가하였다.

검색 기술 영역에서는 해당 연구의 검색이 기존 방식과 차이점을 보이면서 네트워크 그래프로 검색 결과를 확인할 수 있어서 좋다고 평가하였다. 그리고 스토리와 큐레이션을 제작할 수 있고 공유할 수 있다는 점에서 전통문화 콘텐츠 향유에 대한 흐름을 알 수 있는 틀을 마련한 것이 플랫폼의 가장 큰 장점이라고 평가하였다. 또한, 창작 과정에서 수집한 자료를 바탕으로 평소 그려보던 인물 관계망을 네트워크 그래프 표현하여 더 정확하고 더 넓은 연결고리를 제공해줘서 좋고 교육용으로도 충분한 가치를 가지는 플랫폼으로 활용될 수 있다고 평가하였다.

플랫폼 안정성 영역에서는 페이지 이동 간 응답시간, 검색 결과 응답시간, 버튼 또는 메뉴 기능의 일치, 다양한 웹 브라우저 호환, 기능/시스템 오류에 대해 안정성이 높다고 평가하였다. 다만 정보나 기능에 대한 안내 기능이 제공되지 않아 시스템 오류나 기능에 대한 안내 기능이 추가되어야 한다는 의견이 있었다.

5. 결론

국내에는 한국문화의 역사와 전통을 연구하는 기관이 제작하고 공유하는 콘텐츠가 방대하지만, 정작 창작자들은 콘텐츠의 소재가 어디에 있으며 어떻게 찾아야 하는지 모르는 경우가 많다. 따라서 창작에 도움이 될 만한 자료를 효과적으로 검색하거나 창작할 수 있는 플랫폼 지원이 요구된다.

본 연구는 한국 문화와 역사, 전통 관련 융·복합 자료를 효과적으로 검색하고 창작에 도움이 될 수 있는 전통 문화 융·복합 지원을 위한 RDF 기반 온톨로지 검색 서비스 플랫폼을 구축하고 사용성 및 타당성 검증을 수행하였다. 본 서비스는 트리플 검색, 키워드 간 관계 검색, 네트워크 그래프 검색, 스토리 탐색 및 관리, 큐레이션 관리를 지원하며, RDF 트리플 데이터를 기반으로 데이터 간의 관계를 중심으로 검색 결과를 시각화하고 네트워크 그래프 검색과 스토리 검색을 통해 키워드 간의 관계를 쉽게 파악하고 관계 확장을 할 수 있다. 이를 통해 키워드 간의 복잡한 관계 분석 및 검색 시간을 획기적으로 단축할 수 있다.

개발된 서비스의 검증은 전문가 자문 평가를 통해 사용성 및 타당성 평가를 수행하였다. 총 3가지로 UI/UX 영역, 검색 기술 영역, 플랫폼 안정성 영역으로 구분하여 평가하였으며, UI/UX가 적절하게 설계되었고 트리플 검색, 키워드 간 관계 검색, 네트워크 그래프 검색, 스토리 검색을 통해 기존 검색 기능보다 정보 간의 관계를 쉽게 파악하고 분석 및 검색 시간을 단축시킬 수 있는 지능적인 검색이 가능하고 플랫폼 안정성이 우수하다는 평가를 받았다. 다만 본 타당성 평가는 9명의 전문가를 통해 수행되었으며 본 연구의 결과가 플랫폼 구축에 초점이 맞추어져 있어 서비스 제공 방식, 기술적 지원, 다수의 사용자로부터 플랫폼 효과성 및 신뢰성을 분석하는 후속 연구가 필요하다. 또한, 본 연구에서는 전통문화 데이터만을 가지고 연구를 실시하였다. 향후 전통문화 데이터 이외의 다양한 데이터를 적용하는 확장성에 대한 연구로 후속연구가 필요할 것으로 사료된다.

REFERENCES

[1] Y. M. Choi. (2003). Overview and Research Trends of the Semantic Web. *Communications of the Korean Institute of Information Scientists and Engineers, 21(3),* 4-10.

[2] R. Guha, R. McCool & E. Miller. (2003). Semantic Search. *WWW 2003 Conference. May 20-24, ACM Press.*

[3] C. Richa et al. (2004). A Hybrid Approach for Searching in the Semantic Web. *WWW 2004 Conference. New York. USA.*

[4] S. Decker, P. Mitra & S. Melnik. (2000) Framework for the Semantic Web: an RDF tutorial. *IEEE Internet Computing, 4(6),* 68-73.

[5] N. Robert et al. (1991). Enabling Technology for Knowledge Sharing. *AI Magazine.*

[6] T. Gruber. (1993). A translation approach to portable ontologies, *Knowledge Acquisition, 5(2),* 199-220.

[7] R. Studer, V. R. Benjamins & D. Fensel. (1998). Knowledge Engineering : Principles and Methods. *Data & Knowledge Engineering, 25(2),* 184-185.

[8] J. H. Lee. (2003). Ontology Languages for the Semantic Web. *Communications of the Korean Institute of Information Scientists and Engineers, 21(3),* 18-27.

[9] O. Lassila & R. Swick. (1999). Resource Description Framework (RDF) Model and Syntax Specification. *W3C Recommendation.*

[10] O. Lassila. (1998). WEB METADATA: A Matter of Semantics. *IEEE INTERNET COMPUTING, 2(4),* 30-37.

[11] H. S. Seok & Y. J. Lee. (2019). Ontology-based IoT Context Information Modeling and Semantic-based IoT Mashup Services Implementation. *The Journal of the Korea institute of electronic communication sciences, 14(4),* 671-678.

[12] Y. B. Lee & H. Y. Rieh. (2017). A Suggestion of Interface for Ontology-Based Record Retrieval System. *Journal of Korean Society of Archives and Records Management, 17(1),* 217-244.

[13] C. S. Kim, J. W. Lee & H. K. Jung. (2015). A Study on Semantic Web based User Oriented Retrieval System. *Journal of the Korea Institute of Information and Communication Engineering, 19(4),* 871-876.

[14] H. Jeong. (2015). A Study on Ontology and Topic Modeling-based Multi-dimensional Knowledge Map Services. *Journal of Intelligence and Information Systems, 21(4),* 79-92.

[15] H. D. Song, C. Lim, Y. Lee & Y. Lee. (2009). Development and Implementation of a Digital Textbook Platform Usability Assessment Instrument. *Journal of Educational Technology, 25(4),* 125-155.

신 유 탁(Yutak Shin) [정회원]

· 2012년 2월 : 단국대학교 경영학과 (경영학사)
· 2015년 8월 : 연세대학교 정보대학원 디지털경영 (정보시스템학석사)
· 2017년 7월 : ㈜클라루스코리아 기술연구소 선임연구원
· 2017년 8월 ~ 현재 : ㈜피씨엔 연구소 과장
· 관심분야 : 인공지능, 빅데이터
· E-Mail : ytshin@pcninc.co.kr

조 재 춘(Jaechoon Jo) [정회원]

· 2010년 2월 : 제주대학교 컴퓨터교육과 (이학사)
· 2012년 2월 : 고려대학교 컴퓨터교육과 (이학석사)
· 2018년 2월 : 고려대학교 컴퓨터과 (공학박사)
· 2018년 3월 ~ 2019년 2월 : 고려대학교 연구교수
· 2019년 3월 ~ 현재 : 상명대학교 공과대학 조교수
· 관심분야 : 컴퓨터교육, 자연어처리, 인공지능
· E-Mail : jae@smu.ac.kr