

온톨로지와 텍스트 마이닝 기반 지능형 역사인물 검색 서비스[☆]

Ontology and Text Mining-based Advanced Historical People Finding Service

정도현* 황명권** 조민희*** 정한민**** 윤소영***** 김경선***** 김평*****
Do-Heon Jeong Myunggwon Hwang Minhee Cho Hanmin Jung Soyoung Yoon Kyungsun Kim Pyung Kim

요약

시맨틱 웹 기술은 특정 개체를 중심으로 의미적 연관 관계를 생성하고 연관 관계를 이용해서 다양한 지능형 정보 서비스를 구축하는데 활용되며, 텍스트 마이닝 기술은 비정형 데이터를 대상으로 의미 분석을 통해서 의미적 연관 관계를 생성하는데 활용될 수 있다. 본 연구에서는 역사인물을 중심으로 온톨로지 스키마, 인스턴스를 생성하는 가이드라인, 인스턴스 생성, 동명이인 해소를 위한 텍스트 마이닝, 추론을 활용한 지능화된 역사인물 검색서비스를 제안한다. 역사분야 전문가들이 생성한 역사적 사건, 기관, 인물 중심의 연관 관계와 국사편찬위원회에서 보유한 다양한 문헌들 간의 연계를 통해, 사용자들의 정보접근성을 향상시킴과 동시에 관계 정보에 기반한 새로운 역사인물 검색 서비스를 제안하였다. 새로운 역사인물 검색 서비스는 인물간의 소셜 네트워크를 사용하여 역사문헌에 나타난 동명이인을 해소함으로써 보다 정확한 검색서비스를 제공하는 것은 물론, 역사 인물 시소러스를 포함한 다양한 외부 정보와의 연계를 통해서 역사인물에 대한 고부가 정보를 제공하고 있다.

ABSTRACT

Semantic web is utilized to construct advanced information service by using semantic relationships between entities. Text mining can be applied to generate semantic relationships from unstructured data resources. In this study, ontology schema guideline, ontology instance generation, disambiguation of same name by text mining and advanced historical people finding service by reasoning have been proposed. Various relationships between historical event, organization, people, which are created by domain experts, are linked to literatures of National Institute of Korean History (NIKH). It improves the effectiveness of user access and proposes advanced people finding service based on relationships. In order to distinguish between people with the same name, we compares the structure and edge, nodes of personal social network. To provide additional information, external resources including thesaurus and web are linked to all of internal related resources as well.

☞ keyword : 온톨로지(Ontology), 텍스트 마이닝(Text Mining), 인물 검색 서비스(People Finding Service)

1. 서론

* 정 회 원 : 한국과학기술정보연구원 선임연구원
heon@kisti.re.kr

** 정 회 원 : 한국과학기술정보연구원 연구원
mgh@kisti.re.kr

*** 정 회 원 : 한국과학기술정보연구원 선임연구원
mini@kisti.re.kr

**** 정 회 원 : 한국과학기술정보연구원 책임연구원
jhm@kisti.re.kr

***** 정 회 원 : 국사편찬위원회 사료연구위원
syoon@mest.go.kr

***** 정 회 원 : (주)다이퀘스트 연구소장
kksun@diquest.com

***** 정 회 원 : 전주교육대학교 컴퓨터교육과 교수
pyung@jnu.ac.kr(교신저자)

[2012/09/04 투고 - 2012/09/19 심사 - 2012/10/08 심사완료]

☆ 본 논문은 2012년도 한국인터넷정보학회 하계학술발표대회 우수논문의 확장버전임.

기존의 인물정보 검색 서비스는 인물과 관련된 키워드를 검색어로 사용해서 키워드를 포함하고 있는 검색 결과를 제시해주는 서비스 또는 디렉토리 검색 형태로 특정 시대나 특정 단체를 중심으로 인물을 찾아주는 서비스들이 대부분이다. 그러나 인물에 대한 자세한 정보를 모를 경우 키워드 검색을 통해 방대한 문헌에서 원하는 정보를 찾는 것은 쉽지 않으며, 또한 인명의 경우 동명이인이 많기 때문에 원하는 인물을 중심으로 관련 문헌 정보를 찾는 것은 매우 어렵다.

시맨틱 웹 기술은 온톨로지를 지식 표현 모델로, 추론 엔진을 지능형 확장 도구로 사용해서 다양한 의미 기반 지능형 정보 서비스에 활용되고 있다[1]. 또한 역사정보를 이용하는 사용자의 요구는 다양해지고 있다[2]. 본 연구에서는 국사편찬위원회에서 보유하고 있는 다양한 문헌 정보들을 대상으로 사용자가 원하는 정보를 보다 효과적으로 검색·접근하기 위한 방법으로 시맨틱 웹 기술

기본 역사인물 검색 서비스 개발하였다. 한국 근현대사 인물을 중심으로 인물에 관한 상세 정보는 물론, 관련 인물·조직·사건 및 유관 정보를 제공하며, 사용자가 국사편찬위원회의 관련 문헌을 쉽게 접근할 수 있도록 서비스를 설계하였다.

2장에서는 시맨틱 웹 기술을 이용한 지능형 정보서비스와 동명이인 해소를 위한 기존 연구들을 살펴보고, 3장에서는 역사인물 검색 서비스 구성도와 온톨로지 스키마, 인스턴스 생성 방법, 인물 네트워크를 이용한 동명이인 해소 방법에 대해서 기술한다. 4장에서는 역사인물 검색 서비스의 사용자 화면을 중심으로 제공되는 서비스 기능에 대해서 기술하고 5장에서는 결론에 대해서 기술한다.

2. 관련연구

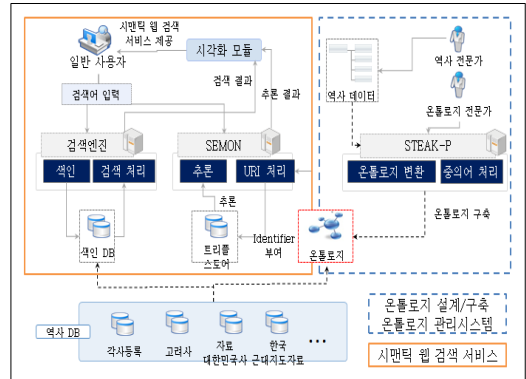
시맨틱 웹 기술을 이용한 지능형 정보 서비스 구축 연구와 동명이인 식별을 위한 연구를 중심으로 기존 연구들을 살펴보도록 하겠다.

2.1 시맨틱 웹 기술 기반 지능형 서비스

시맨틱 웹 기술은 국내의 IT분야, 전자정부, Health Care 등 다양한 분야에서 개체간의 관계를 명확하게 하거나 검색 성능의 개선, 데이터의 공유 및 재사용을 위해서 활발하게 사용되고 있다[1]. 스탠포드 대학에서는 다양한 바이오메디컬 자원들을 효과적이면서도 빠르게 검색하기 위해서 온톨로지 기반의 검색 서비스를 개발하였다[3]. KISTI에서도 과학기술 문헌을 대상으로 기술 생명주기 분석 기법을 이용한 기술기회 발굴을 위한 결정 지원 서비스를 개발함으로써 기술의 수준과 발전 단계, 발전 속도 등의 예측 기능을 제공하고 있다[4]. 국방기술품질원에서도 온톨로지 기반 무기정보 검색 서비스를 개발함으로써 무기의 제원과 분류를 이용한 무기 검색이 가능해졌다[5]. 이처럼 시맨틱 웹 기술은 정보 자원의 연계, 데이터 공유 및 재활용성을 높여주는 것[6]은 물론 정보 서비스의 검색 기능을 향상하는데 사용되고 있다.

2.2 동명이인 해소

동명이인 해소 문제는 실제 문헌에 나타난 동일 인명을 매칭하거나 동일 이름을 가지는 인물을 식별하는 문제로 세분화할 수 있다. 동명이인 해소를 위한 연구는 대



(그림 1) 역사인물 검색 시스템 구성도

부분 학술문헌의 서지정보에 나타난 저자명을 식별하는 연구들로서, 저자의 소속이나 이메일을 활용하는 연구[7]와 저자의 홈페이지를 활용하는 연구[8], 공저자를 이용한 연구[9], 서지정보에서 추출된 자질들이 저자 식별에 미치는 영향을 분석한 연구[10]가 있다. 온톨로지로 구축된 인스턴스들을 식별하기 위해서 온톨로지와 RDF 트리플을 사용해서 동일 인스턴스 여부를 확인하고 데이터를 병합하는 연구도 수행되었다[11].

3. 역사인물 검색 서비스 개발

역사인물 검색 시스템 구성과 데이터 생성 과정 및 동명이인 해소방법을 중심으로 역사인물 검색 서비스를 개발하였다.

3.1 시스템 구성

역사인물 검색 서비스는 (그림 1)과 같이 크게 역사 온톨로지 설계 및 인스턴스를 생성하고 관리하는 부분과 시맨틱 웹 서비스를 제공하는 부분으로 구성된다.

역사인물 검색 서비스는 사용자가 특정 인물을 검색하고 해당 인물을 중심으로 인물과 관련된 정보를 쉽게 찾는데 중점을 두고 인터페이스를 설계하였다. 또한 국사편찬위원회에서 보유하고 있는 다양한 문헌 정보를 대상으로 색인 정보를 생성하는 과정에서 동명이인이 해소된 인명정보와 연결함으로써 보다 정확한 정보 접근 경로를 제공하고 있다.

- 검색엔진: 사용자가 입력한 단어를 대상으로 불린

연산과 키워드 검색을 이용해서 검색 결과를 제공한다.

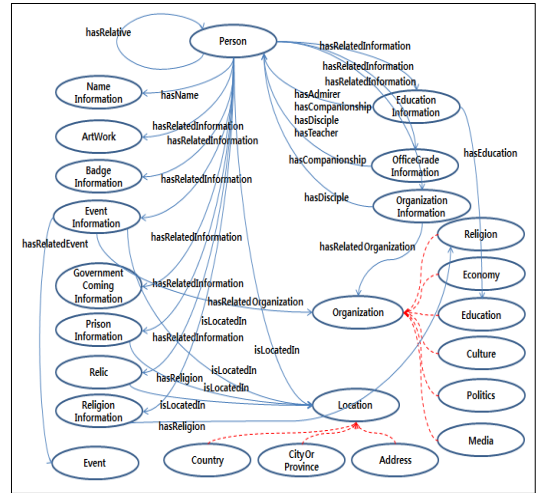
- **SEMON**: 시맨틱 웹 서비스를 제공하기 위한 서비스 프레임워크로 트리플 저장, URI 관리, 추론을 통한 검색 결과를 제공한다.
- **STEAK-P**: 네트워크 기반의 중의성 해소 모듈로서, 네트워크 구조 분석을 통해서 인명정보를 해소하는데 사용된다[12].

사용자는 인물 중심의 관련 네트워크나 트리뷰를 통해서 인물에 관한 상세 정보는 파악하는 것은 물론, 인물을 중심으로 관련 인물/조직/사건 및 관계 정보에 접근할 수 있다.

3.2 데이터 생성 프로세스

역사인물 검색 서비스는 시맨틱 웹 표준을 준수하는 서비스 프레임워크를 기반으로 온톨로지 기반의 정보 표현 모델을 사용하고 있다. 온톨로지 데이터 생성을 위해서는 구축 대상 인물을 선정하는 작업이 가장 먼저 이루어진다. 역사 전문가들이 인물의 중요도, 인물에 대한 자료 등을 고려해서 총 1,769명(고대 360명, 고려 450명, 조선전기 422명, 조선후기 537명)을 선정하였으며, 데이터 생성 프로세스는 다음과 같은 과정으로 진행된다.

- ① 역사 온톨로지 스키마 설계: 역사 전문가들과 온톨로지 전문가의 협업을 통해서 개념과 객체, 데이터 타입 속성들을 도출한다.
- ② 인스턴스 생성 가이드라인 도출: 온톨로지 전문가가 인스턴스 표기를 위한 가이드라인을 도출하는 작업으로, 역사 전문가들이 가이드라인을 사용해서 온톨로지를 구성하는 속성과 관계들을 기술한다.
- ③ 역사 정보 템플릿 생성: 시대별 인물정보를 표현하기 위해서 역사 전문가들이 사용할 템플릿을 작성한다.
- ④ 역사 정보 템플릿 작성: 역사 전문가들이 각종 문헌 정보들을 활용해서 시대별 인물 정보를 템플릿에 채운다.
- ⑤ RDF 트리플 생성 및 관리: 온톨로지 인스턴스 에디터(OntoEdit)가 RDF 트리플을 생성하고 온톨로지 스키마를 기반으로 수정, 관리한다.
- ⑥ 동명이인 식별 모듈을 통해서 서로 다른 식별자 (URI)가 부여된 동일 인물 정보를 하나의 동일한



(그림 2) 온톨로지 스키마

URI로 변경하는 작업을 수행한다.

- ⑦ 인스턴스 생성: 서비스에 사용될 최종 RDF 트리플 정보를 생성한다.

3.3 온톨로지 스키마

역사인물 온톨로지는 한국 근대사를 중심으로 역사인물 검색 서비스에서 제공하고자 하는 인물 정보와 인물 관련 정보를 중심으로 개념을 도출하고 개념간의 관계와 데이터타입 속성을 중심으로 설계되었다. 역사 전문가들과 온톨로지 전문가들이 클래스와 데이터타입 속성, 객체 속성을 도출하는 작업을 공동으로 수행하였다.

역사인물 온톨로지는 인물이 기본적으로 가지는 속성(이름, 가족관계), 역사적물이 가지는 속성(관훈정보, 사건정보, 창작정보, 수형정보 등), 그리고 다양한 조직 활동과 학력 등을 상호 연계하였고, 연계된 정보를 기반으로 인물과 인물, 인물과 조직, 인물과 사건의 관점에서 다양한 연관 정보를 찾을 수 있도록 서비스에 활용되고 있다. (그림 2)는 온톨로지 스키마에서 클래스와 클래스간 객체 속성들의 상호 관계를 중심으로 보여주고 있으며, 점선은 클래스간 상위 관계(subClassOf) 관계를 표시하고 있다. 온톨로지는 24개의 클래스와 47개의 데이터타입 속성, 224개의 객체 속성으로 구성되며, 224개의 객체 속성은 인물과 인물간의 다양한 관계 정보 약 200개를 포함하는데, 부모, 형제, 조부모 등의 가족관계, 고모, 숙부, 백부 등의 인척관계 등으로 구성되어 있다.

(표 1) Person 클래스의 데이터타입 속성 예제

속성명	의미	Domain	Type
entryName	표제이름	Person	String
nickName	별명	Person	String
myohoName	묘호	Person	String
buName	부명	Person	String
birthDate	출생일	Person	Date
deathDate	사망일	Person	Date
badgeDate	수상일	Person	Date
officialRank	관직	Person	String
...

(표 1)은 역사인물 온톨로지를 구성하는 Person 클래스의 데이터타입 속성들의 일부로서 주로 인물의 명칭과 관련된 속성과 출생일, 사망일, 그리고 활동과 관련된 속성 등으로 구성되어 있다. 인물(Person) 클래스를 중심으로 다양한 클래스들이 객체 관계를 형성하고 있으며, 기관(Organization) 클래스는 종교(Religion), 경제(Economy), 교육(Education), 문화(Culture), 정치(Politics), 방송(Media) 클래스를 하위 클래스로, 지역(Location) 클래스는 국가(Country), 시도(CityOrProvince), 세부주소(Address)를 하위 클래스로 두고 있다.

3.4 인스턴스 생성 가이드라인

온톨로지 인스턴스 생성은 역사분야 전문가들에 의해서 수행되며, 이를 위해서 온톨로지의 구성 내용을 서식화된 템플릿으로 역사분야 전문가들에게 배포하였다. 역사분야 전문가들은 국사편찬위원회에서 보유하고 있는 역사관련 시소러스와 역사 관련 문헌 등을 사용해서 역사인물을 중심으로 시대별 인물 정보를 템플릿에 입력하면 된다. 템플릿은 온톨로지 스키마에 정의된 객체 속성과 데이터타입 속성들로 구성되며, 데이터 입력 방법을 표준화하기 위해서 입력을 위한 가이드라인을 작성해서 배포하였다.

(표 2)는 인스턴스 생성 가이드라인의 예제로써, 인스턴스 생성 가이드라인은 이름, 주소, 시기, 가족관계, 전통교육, 출사, 관직 등 온톨로지 스키마에 기록된 모든 항목에 대해서 인스턴스를 기술하는 과정에서 발생할 수 있는 역사 전문가들의 이표기를 통일화하기 위해서 활용된다. 역사 전문가들은 인스턴스 생성 가이드라인을 참고하여 엑셀 파일로 작성된 역사 정보 템플릿을 채우기

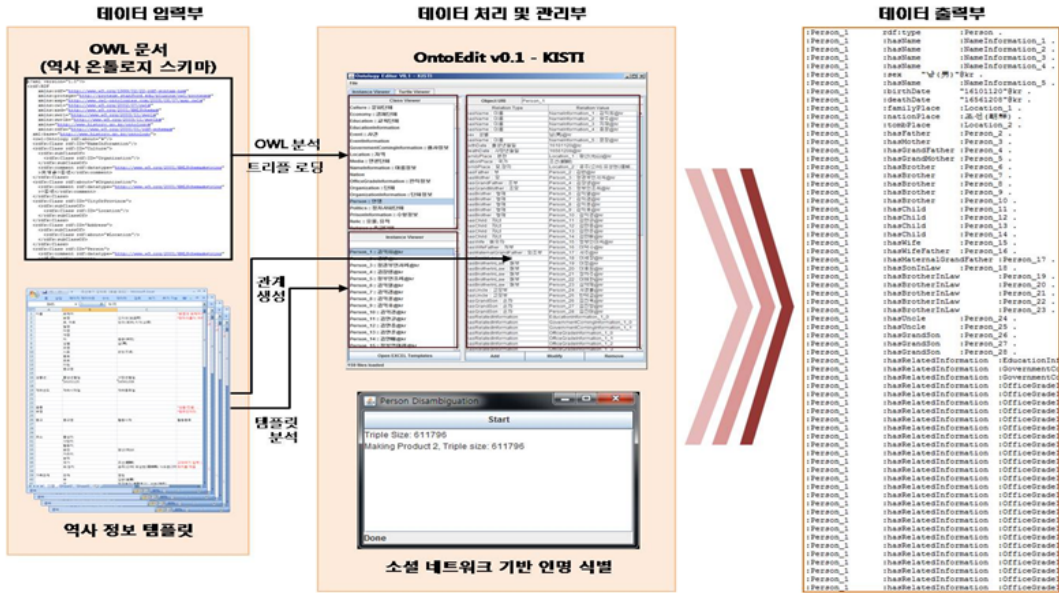
(표 2) 인스턴스 생성 가이드라인 예제

항목	기술규칙
이름	<ol style="list-style-type: none"> 이표기가 여러개인 경우 “:(콜론)”으로 연결 승려의 경우, 속성(俗姓)만 나와 있는 경우 ‘본관+성씨’로 ‘이칭’란에 표기 본명보다 일반적으로 잘 알려진 이름이 있을 경우 표제어로 사용 왕의 경우 표제어에는 시호나 묘호를 사용 고대의 경우 삼국사기의 왕명을, 고려는 고려사의 왕명 표기, 조선의 경우 묘호를 사용
주소	<ol style="list-style-type: none"> 현재 주소로 바꾸지 않고 자료상 등장하는 지명을 그대로 사용 귀화인의 경우 원래 국적을 출생지에 기입 본관은 한국역대인물종합정보의 것으로 기입
시기	<ol style="list-style-type: none"> 연도는 “YYYYMMDD”로 작성, 기원전일 경우에는 “-YYYYMMDD”로, 윤달일 경우에만 “YYYYMMDD윤” 이라고 작성함. 연월일 미상일 경우 00으로 표기 시기는 왕대를 의미하며 한자로 표기하지 않음 생몰년에 이설이 있는 경우 행을 바꾸어 복수개 기입
가족관계	<ol style="list-style-type: none"> 표제어를 기준으로 가족관계 표기 다른 표기가 있을 경우 “:(콜론)”으로 병기 여성의 경우 표제어에 해당하는 명칭을 알 수 없을 경우 ‘본관+스페이스+성씨’로 표기 여러 표기법으로 등장하는 고대 인물의 경우는 ‘삼국사기’를 제1기준, ‘삼국유사’를 제2기준으로 표기
...	...

때문에 온톨로지 구조를 알지 못해도 작업하는데 문제가 발생하지 않는 것은 물론, 엑셀 파일을 사용해서 쉽게 인스턴스 생성에 필요한 정보를 채울 수 있다.

3.5 인스턴스 생성

인스턴스 생성은 (그림 3)과 같이 역사 온톨로지 스키마와 역사 전문가들이 생성한 역사 정보 템플릿을 입력으로 받아서 온톨로지 인스턴스 에디터(OntoEdit)가 RDF 트리플을 관리하고 생성하는 1단계, 소셜 네트워크 기반 인명 식별 모듈을 통해서 서비스에 사용될 최종 RDF 트리플을 생성하는 2단계로 구분할 수 있다. 1단계에서는 역사 정보 템플릿에 나타난 모든 인물들이 동일인인지 여부를 판단하지 않고 서로 다른 식별자(URI)를 부여하기 때문에 서로 다른 템플릿에서 나타난 동일 인물에 대한 정보를 병합하는 것은 불가능하다. 2단계에서는 소셜



(그림 3) 인스턴스 생성 프로세스(13)

네트워크 기반의 인명 식별 결과를 통해서 동일인에 부여된 서로 다른 URI를 병합하는 작업이 진행된다.

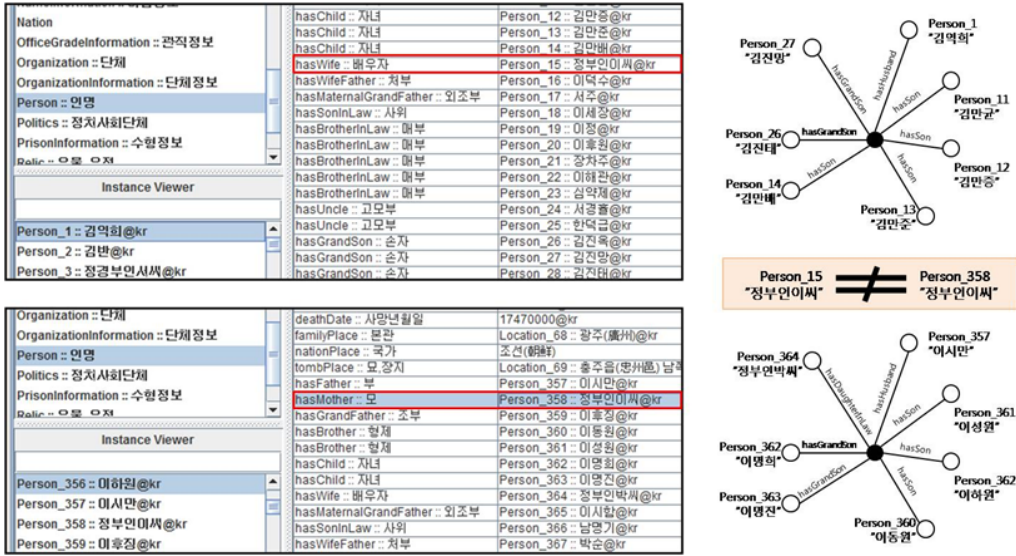
OntoEdit는 온톨로지 스키마 로딩 모듈, 템플릿 분석 모듈, 1차 동명이인 처리 모듈, RDF(S)/OWL 트리플 생성 모듈로 구분되며, 각 모듈은 다음과 같은 기능을 수행한다.

- 온톨로지 스키마 로딩 모듈: 역사 정보 클래스들의 정의 및 관계를 Protege를 이용하여 정의하였다. 분석된 템플릿의 내용들을 온톨로지로 변환하기 위해 온톨로지 스키마를 로딩한다.
- 템플릿 분석 모듈: 역사 전문가에 의해 작성된 역사 정보 템플릿은 엑셀로 작성되며, 역사 인물 한명을 중심으로 다양한 관계가 작성된다. 각각의 템플릿 내용을 분석하여 온톨로지 스키마와 연결한다.
- 1차 동명이인 처리 모듈: 하나의 템플릿 파일에는 여러 사람이 기술되며, 경우에 따라 동일 이름이 출현한다. 동일 파일에 기술된 동일 이름의 경우는 동일인으로 결정한다.
- RDF(S)/OWL 트리플 형성 모듈: 고대 360명, 고려 450명, 조선 전기 422명, 조선 후기 537명에 대해 각각 시기별로 로딩된 전체 트리플을 생성한다.

OntoEdit에서 형성한 트리플의 경우, 다른 파일에 존재하는 동일 이름에 대한 동명이인 처리는 이루어지지 않은 상태이다. 동명이인에 대한 처리는 WSD (Word Sense Disambiguation)의 한 부분이며, 본 연구에서는 네트워크 방식[14]을 응용하여 접근하였다. 동명이인이 발생하는 경우는 다음과 같다.

- 경우 1: 한 템플릿의 관계 인물의 이름이 다른 템플릿의 이름과 동일한 경우
- 경우 2: 한 템플릿의 관계 인물의 이름이 다른 템플릿의 관계 인물의 이름과 동일한 경우

트리플 내에서, 동일 이름이 발견 될 경우는, 각 이름의 인물은 서로 다른 사람으로 판단되어 있다. 하지만, 위의 경우와 같이 동일 인물이 다수 존재하며, 특히 유명한 인물일수록 그 빈도는 많아진다. 이에 대해 동일 인물 여부를 판단하기 위해, 같은 이름으로 파악된 사람의 소셜 네트워크를 형성한다. (그림 4)는 “정부인이씨”에 대한 예를 보이고 있으며, 실제 역사 정보 속에는 “이하원”의 “모” 관계 그리고 “김익희”의 “배우자” 관계로 발견된다. 이에 대해 동일 인물 여부를 판단하기 위해 각 이름과 연계 있는 소셜 네트워크를 서로 비교하여 일치하는



(그림 4) 인물 네트워크 기반 동명이인 해소

:NameInformation_16	rdftype	:NameInformation .
:NameInformation_16	:entryName	"거질미왕"@kr .
:NameInformation_16	:entryName	"居叱彌王"@cn .
:NameInformation_17	rdftype	:NameInformation .
:NameInformation_17	:realName	"김거질미"@kr .
:NameInformation_17	:realName	"金居叱彌"@cn .
:NameInformation_18	rdftype	:NameInformation .
:NameInformation_18	:realName	"김금률"@kr .
:NameInformation_18	:realName	"金今勿"@cn .
:Location_3	rdftype	:Location .
:Location_3	:locationName	"김해(金海)"@kr .
:Person_14	rdftype	:Person .
:Person_14	:hasName	:NameInformation_19 .
:Person_14	:hasName	:NameInformation_8421 .
:NameInformation_19	rdftype	:NameInformation .
:NameInformation_19	:realName	"마품왕"@kr .
:NameInformation_19	:realName	"麻品王"@cn .

(그림 5) 역사인물 RDF 트리플 예제

문서ID/	인물명 / 인물URI/	시대별/	시대
an_131	김상덕 Person_a2817	근현대	근현대 ↓
an_138	김영두 Person_a217	근현대	근현대 ↓
an_139	김영두 Person_a217	근현대	근현대 ↓
an_1310	김영두 Person_a217	근현대	근현대 ↓
an_1311	김영두 Person_a217	근현대	근현대 ↓
an_1313	김영두 Person_a217	근현대	근현대 ↓
an_1315	김영두 Person_a217	근현대	근현대 ↓
an_1324	박상훈 Person_a847	근현대	근현대 ↓
an_1329	박상훈 Person_a847	근현대	근현대 ↓
an_1330	박상훈 Person_a847	근현대	근현대 ↓
an_1331	박상훈 Person_a847	근현대	근현대 ↓
an_1332	박상훈 Person_a847	근현대	근현대 ↓
an_1343	김상덕 Person_a2817	근현대	근현대 ↓
an_1343	김상덕 Person_a4686	근현대	근현대 ↓
an_1358	김익진 Person_a477	근현대	근현대 ↓
an_1365	김익진 Person_a477	근현대	근현대 ↓

(그림 6) 역사 문헌별 인물 URI 정보

관계속성이 존재한다면 동일인물로 판단한다. (그림 4)의 경우는 같은 이름을 가진 두 사람이 서로 다른 인물로 판단된 예를 보여주고 있다.

(그림 5)는 역사인물에 대한 RDF 트리플 예제로서 동명이인 문제가 해결된 최종 RDF 트리플은 시맨틱 웹 프레임워크를 통해 트리플 저장소에 저장된 후 서비스에 사용된다.

(그림 6)은 역사문헌에 포함된 인명의 중의성을 해소하여 생성된 정보로 각 문헌에 나타난 인물에 대해 구축된 네트워크 기반의 동명이인 해소 기법을 적용하여 온톨로지 기반의 고유 ID인 URI를 부여하였다. 이 정보는 각 문헌에 인물의 메타정보태그 형태로 추가되며 인물정보와 메타정보태그를 결합하여 색인정보로 사용하여 검색 시스템에서 발생할 수 있는 각 문헌에 나타난 동명이인 문제를 해소하였다.



(그림 7) 검색어 자동 완성

4. 역사인물 검색 서비스 기능

시맨틱 웹 기술 기반의 지능형 역사인물 검색 서비스는 국사편찬위원회가 보유한 다양한 역사 문헌들을 대상으로 의미기반 검색 기능을 제공함으로써, 사용자는 원하는 정보에 빠르고 정확하게 접근할 수 있다. 사용자가 입력한 인물을 중심으로 관련 역사 인물, 사건, 기관 간의 연관 관계를 제시하기 위해서 검색 엔진과 시맨틱 웹 서비스 프레임 워크를 사용한다. 검색어 자동 완성, 시대별·가나다별 인물 검색, 동시대 인물 검색, 동명이인 구분 검색, 트리와 방사형 네트워크를 이용한 검색 등 다양한 기능들이 서비스에 포함되어 있다.

4.1 검색어 자동 완성

검색어를 입력 시 검색어에 대한 자동완성 기능을 제공하며, (그림 7)과 같이 동명이인이 있는 경우 인물 정보를 통해 특정 인물을 선택해서 검색을 수행할 수 있다. 인물의 사진은 위키피디아 및 웹에서 제공하는 사진을 연결하여 보여준다.

4.2 시대별·가나다별 검색

사용자가 삼국시대, 고려시대, 조선시대 등 특정 시대나 연도/기간을 선택한 경우, (그림 8)과 같이 해당 시기나 기간에 활동한 인물 목록이 제공되는 것은 물론, 가나다별 정렬된 이름을 기준으로 인물 목록을 확인할 수 있다. 또한 사용자는 가나다별 검색을 선택해서 이름별로 정렬된 인물 목록을 확인해서 검색할 수 있다. 인물에 대한 간략한 정보를 제공함으로써 사용자가 특정 인물을 선택하는 것을 도와준다.



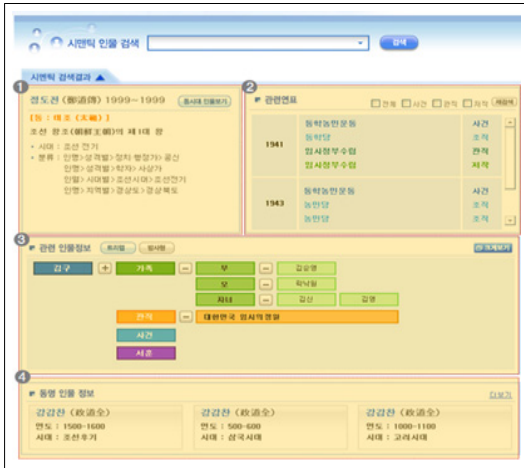
(그림 8) 시대별·연도별 인물 검색



(그림 9) 동명이인 구분 검색

4.3 동명이인 구분 검색

기존 키워드 검색 방식에서는 동명이인에 대한 처리 기술이 적용되지 않았기 때문에 사용자가 검색한 인물이 동명이인일 경우 다수의 검색결과에 대해 인물들을 하나씩 문헌 내용을 파악하면서 사용자가 직접 부적절한 정보를 필터링해야만 한다. 예를 들면 독립운동가 '김구'를 검색하려고 '김구'로 입력해서 검색하는 경우 고려시대 1명, 조선시대 2명의 인물이 문헌에 나타난다. 본 연구에서는 그림 9와 같이 동명이인 해소 모듈을 통해서 역사 문헌에 나타난 이름들을 구분해서 제공함으로써 사용자는 3명의 '김구'와 관련된 결과중에서 본인이 원하는 '김구' 관련 정보만을 검색 결과로 얻을 수 있다.



(그림 10) 인물 정보 상세보기



(그림 11) 동시대 인물 검색

4.4 인물 정보 상세보기

사용자가 선택한 특정 인물을 중심으로 관련 정보를 보여주는 (그림 10)은 다음과 같은 정보들을 포함하고 있다.

- ① 인물에 대한 요약 정보: 요약 정보에는 선택된 인물의 간단한 설명과 시대, 분류 정보, 동시대의 인물들을 보기를 통해서 (그림 11)과 같이 동시대 인물을 볼 수 있는 기능을 제공한다.
- ② 인물 연표 정보: 사건, 관직, 저작 별로 연표를 나누어서 볼 수 있도록 항목을 선택할 수 있는 체크 박스를 제공하고 기본 값은 전체로 설정됨.
- ③ 인물 관련 정보: 가족, 서훈, 사건, 관직, 저작 별로 구분해서 볼 수 있도록 방사형 네트워크 또는 트리형 구조로 볼 수 있도록 지원하며, (그림 11)과 같이 트리형을 확장



(그림 12) 트리보기

해서 검색 조건을 입력하는 화면으로 활용할 수 있음. ④ 사용자가 검색창에서 입력한 인물명과 동명이인이 있을 경우: 동명이인에 대한 인물정보를 제공하고 특정 인물을 선택하면 해당 인물에 대한 화면으로 페이지 이동하는 기능을 제공한다.

(그림 12)에서는 인물 관련 정보를 트리보기 모두 펼쳐보기한 형태로써, 트리에 나타난 특정 인물을 선택한 경우 ‘관련자료’, ‘중심검색’ 메뉴가 제시되며, ‘관련자료’를 선택하면 하단의 검색 결과에 검색어에 대한 정보와 선택한 속성 정보를 같이 제시한다. 예를 들면 검색어 ‘김구’에 대한 제시된 관련 인물중 ‘안중근’의 ‘관련자료’ 선택시 하단의 검색결과로 ‘김구+안중근’에 대한 정보를 제시한다. ‘중심검색’을 선택하면 사용자가 선택한 관련 속성으로 검색화면이 변경된다. 검색어 ‘김구’에 대한 제시된 관련 인물중 ‘안중근’의 ‘중심검색’을 선택하면 안중근에 대한 관련네트워크, 관련정보, 검색결과를 제시한다.

5. 결 론

시맨틱 웹 기술 기반 역사인물 검색 서비스는 온톨로지를 이용해서 인물을 중심으로 관련 정보를 제공함으로써, 특정 인물을 찾아가기 위한 다양한 접근 방법을 제공한다. 이를 위해서 역사 전문가와 온톨로지 전문가가 역사인물을 중심으로 온톨로지 스키마를 설계하고 인스턴스를 생성하는 작업을 수행하였다. 역사인물 온톨로지는 국사편찬위원회에서 보유하고 있는 다양한 문헌과 외부 자원들을 이용해서 인물간의 관계, 사건과 조직, 문화 유물과 유적 정보를 포함하고 있다. 역사인물 인스턴스는 역사 전문가들이 역사정보 템플릿을 사용해서 고대, 고

려, 조선전후기 역사인물 1,769명을 대상으로 생성하였으며, 역사 전문가들의 표기 방법을 통일하기 위해서 인스턴스 생성 가이드라인도 작성해서 활용하였다. 또한 인적 네트워크를 기반으로 동명이인을 해소한 후 국사편찬위원회에서 보유하고 있는 다양한 문헌 정보와 연계함으로써 검색의 정확성을 향상시킬 수 있었다.

역사인물 검색 서비스는 시맨틱 웹 서비스 프레임워크를 사용해서 검색의 편의성과 정확도를 개선하는데 목표를 두었다. 사용자는 검색어 자동 추천 기능을 통해 검색 대상 인물을 선택할 수도 있고, 시대별, 가나다별 인물 목록을 통해서 특정 인물을 선택하는 것은 물론, 선택된 인물을 중심으로 다양한 관련 정보를 네트워크 보기나 트리보기를 통해서 확인할 수 있다. 또한 특정 인물을 중심으로 ‘관련검색’이나 ‘중심검색’을 통해서 재검색을 보다 쉽게 할 수 있도록 검색 서비스를 개발하였다. 본 연구에서는 한국 근현대사의 역사 인물을 중심으로 온톨로지가 구축되어 서비스에 사용되었고, 향후에는 한국 역사 전반에 걸쳐서 확대할 계획이며, 온톨로지와 연계된 역사 문헌의 대상도 점차 확대할 계획이다.

참 고 문 헌

- [1] <http://www.w3.org/2001/sw/sweo/public/UseCases/>
- [2] 전정현, 이지현, “온라인 역사정보서비스에 관한 연구”, 정보관리학회지, 제27권, 제3호, pp.67-82, 2010
- [3] C. Jonquet, P. LePendu, S. M. Falconer, A. Coulet, N. F. Noy, M. A. Musen, and N. H. Shah, “NCBO resource index: ontology-based search and mining of biomedical resources”, *Web Semantics*, vol.9, no.3, pp.316-324, 2011
- [4] 이미경, 정한민, 김평, 성원경, “연구개발 전략 수립 지원을 위한 테크놀로지 인텔리전스 서비스”, 한국정보과학회지, 제17권, 제5호, pp.337-341, 2011.
- [5] 최중환, 박정호, 서동민, 이승우, 정한민, 김평, “시맨틱 웹 기반 지능형 무기정보 검색 서비스”, 한국인터넷정보학회 하계학술발표대회 논문집 제13권, 제1호, pp.83-84, 2012.
- [6] L. Ding, T. Lebo, J. S. Erickson, D. DiFranzo, G. T. Williams, X. Li, J. Michaelis, A. Graves, J. Zheng, Z. Shanguan, J. Flores, D. L. McGuinness, J. A. Hendler, “TWC LOGD: A portal for linked open government data ecosystems”, *Web Semantics*, vol.9, no.3, pp.325-333, 2011.
- [7] A. Culotta, P. Kanani, R. Hall, M. Wick, A. McCallum, “Author disambiguation using error-driven machine learning with a ranking loss function”, *IWeb-2007*, 2007.
- [8] N. Aswani, K. Bontcheva, H. Cunningham, “Mining information for instance unification”, *ISWC-2006*, pp.329-342, 2006
- [9] H. Han, C. L. Giles, H. Zha, C. Li, and K. Tsioutsoulis, “Two supervised learning approaches for name disambiguation in author citations”, *JCDL-2004*, 2004.
- [10] 강인수, 이승우, 정한민, 김평, 구희관, 이미경, 성원경, 박동인, “저자 식별을 위한 자질 비교”, 한국콘텐츠학회논문지, 제8권, 제2호, pp.41-47, 2008.
- [11] T. Kim, P. Kim, S. Lee, H. Jung, and W.K. Sung, “OntoURIRResolver: URI Resolution and Recommendation Service Using LOD”, *proceedings on U- and E-Service Science and Technology(UNESST 2011)*, CCIS 264, pp.245-250, 2011.
- [12] D.H. Jeong, M. Hwang, and W.K. Sung, “Generating Knowledge Map for Acronym-Expansion Recognition”, *Proceedings on U- and E-Service Science and Technology(UNESST 2011)*, CCIS 264, pp.287-293, 2011
- [13] 황명권, 정도현, 조민희, 정한민, 김평, 윤소영, 한궁희, “역사정보 온톨로지 구축에 대하여”, 한국인터넷정보학회 하계학술발표대회 논문집 제13권, 제1호, pp.87-88, 2012.
- [14] M. Hwang, C. Choi, and P. Kim, “Automatic Enrichment of Semantic Relation Network and its Application to Word Sense Disambiguation”, *IEEE Transaction to Knowledge and Data Engineering*, Vol.23, No.6, pp.845-858, 2011.

◎ 저 자 소 개 ◎



정 도 현 (Do-Heon Jeong)

1997년 연세대학교 문헌정보학과
2003년 연세대학교 대학원 문헌정보학과(석사)
2011년 연세대학교 대학원 문헌정보학과(박사수료)
2003년~현재 한국과학기술정보연구원(KISTI) 선임연구원
관심분야 : 시맨틱 웹, 텍스트 마이닝, 정보 분석
E-mail : heon@kisti.re.kr



황 명 권(Myunggwon Hwang)

2004년 조선대학교 컴퓨터공학부(공학사)
2006년 조선대학교 대학원 전자계산학과(이학석사)
2011년 조선대학교 대학원 컴퓨터공학과(공학박사)
2011년~현재 한국과학기술정보연구원(KISTI) 연구원
관심분야 : 시맨틱 웹, 데이터 마이닝, 의미 식별
E-mail : mgh@kisti.re.kr



조 민 희 (Minhee Cho)

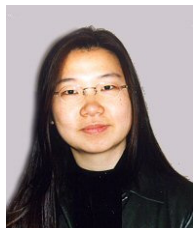
2003년 연세대학교 전산학과(이학사)
2005년 연세대학교 전산학과(석사)
2005년~현재 한국과학기술정보연구원(KISTI) 선임연구원
관심분야 : 자연어처리, 텍스트마이닝
E-mail : mini@kisti.re.kr



정 한 민 (Hanmin Jung)

1994년 POSTECH 전자계산학과(공학석사)
1994년~2000년 ETRI 선임연구원
2000년~2004년 (주)다이퀘스트 연구소장/CTO
2003년 POSTECH 컴퓨터공학(공학박사)
2004년~현재 한국과학기술정보연구원(KISTI) 책임연구원
관심분야 : 시맨틱 웹, HCI, 자연어처리
E-mail : jhm@kisti.re.kr

● 저 자 소 개 ●



윤 소 영 (Soyoung Yoon)

1994년 숙명여자대학교 문헌정보학과(학사)
1996년 숙명여자대학교 대학원 문헌정보학과(석사)
2005년 연세대학교 대학원 문헌정보학과(박사)
2005년~현재: 국사편찬위원회 사료연구위원
관심분야 : 정보검색, 정보조직, 정보기술
E-mail : syoon@mest.go.kr



김 경 선 (Kyungsun Kim)

1997년 서강대학교 전자계산학과 (공학사)
1999년 서강대학교 대학원 컴퓨터공학과 (공학석사)
2006년 서강대학교 대학원 컴퓨터공학과 (공학박사)
2006년~2008년 삼성전자통신연구소
2008년~현재: (주)다이퀘스트 기술연구소 소장
관심분야 : 자연어처리, 정보검색, 시맨틱 웹
E-mail : kksun@diquest.com



김 평 (Pyung Kim)

1997년 충남대학교 컴퓨터과학과(이학사)
1999년 충남대학교 대학원 컴퓨터과학과(이학석사)
2004년 충남대학교 대학원 컴퓨터과학과(이학박사)
2004년~2012년 한국과학기술정보연구원 선임연구원
2012년~현재 전주교육대학교 컴퓨터교육과 교수
관심분야 : 자연어처리, 시맨틱 웹, 정보교육
E-mail : pyung@jnue.kr