

역사용어 온톨로지 모형 적용 방안 연구*

- 한국근현대사 인물을 중심으로 -

A Study on the Model of History Ontology: A Focus on Korean Modern Historical Person

이 혜 원(Hyewon Lee)**

윤 소 영(So-Young Yoon)***

초 록

본 연구에서는 역사인물 중심의 온톨로지를 구축하기 위해 기존의 한국역사용어시소러스의 문제점을 분석하는 동시에 한국역사용어시소러스 및 국편의 다양한 정보검색 시스템을 활용하는 전문가들의 인터뷰를 실시하였다. 한국역사용어시소러스의 개념 표현 기준과 전문가들이 작성한 마인드맵을 분석하여 현재 시스템과 전문가간의 정보 표현의 차이를 확인하였다. 이를 바탕으로 역사정보 이용자들의 정보요구에 부합하고 잠재력 있는 정보검색시스템에 적용할 역사용어 온톨로지 모형을 설계하였다. 본 연구에서는 먼저 클래스, 속성을 정의하고, 인스턴스 입력시 고려해야 할 점들을 나열함으로써, 역사정보가 갖는 고유한 특성들을 정리하였다. 또한 온톨로지의 장점인 개념 확장을 활용하여 역사정보의 다양한 특성들을 조합하고 이를 통해 새로운 정보를 제공할 수 있는 가능성을 제시하였다.

ABSTRACT

The purpose of the study is to construct a History Ontology Model for historical person to analyse issues of Korean History Thesaurus and interview history specialists who are use information systems in National Institute of Korean History. This study verifies the difference between both descriptions through comparative analysis of term concept in Korean History Thesaurus and mind-map written by history major. Based on this, we build history ontology model to meet users' information needs and adapt to information retrieval system. First, to organize unique features of history, we define class and attribute and then enlisted considerations for instance input. The study suggests a possibility of new service through combination multiple features using concept extension that is a strength of ontology.

키워드: 역사 인물, 역사 온톨로지, 시소러스 기반 온톨로지, 지식-공학기반 온톨로지, 온톨로지 구축
Historical Person, History Ontology, Thesaurus based Ontology, Knowledge-Engineering
Methodology based Ontology, Ontology Construction

* 본 연구는 2010년도 국사편찬위원회 연구용역으로 수행되었음.

** 서울여자대학교 문헌정보학과 조교수(hwlee@swu.ac.kr) (제1저자)

*** 국사편찬위원회 사료연구위원(syoon@mest.go.kr) (교신저자)

논문접수일자 : 2011년 2월 18일 논문심사일자 : 2011년 2월 28일 게재확정일자 : 2011년 3월 11일

1. 서론

국사편찬위원회(이하 국편)에는 여러 정보 검색시스템이 오랜기간에 걸쳐 독립적으로 구축되어 왔으며 기관내 시스템간 정보공유와 활용이라는 측면에서 이러한 분산시스템들을 유기적으로 연결하기 위한 방안이 필요하게 되었다. 이러한 목적으로 한국역사용어시소러스를 구축하였으나 실제 활용에서는 그 역할을 충분히 수행하지 못하였다. 한국역사용어시소러스의 용어 추출과 선정은 중·고등학교 국사 교과서 및 역사 사전류, 국편에서 구축한 디지털 원전(原典) 자료 등을 대상으로 하였다. 현재 한국역사용어시소러스가 제공하는 용어수준과 표현구조는 전문 역사학자들을 주이용 대상으로 하는 제한점을 가지고 있다. 또한 그 주요 대상으로 삼고 있는 전문 역사학자들의 요구조차도 잘 반영하지 못하고 있는 상태라 할 수 있다.

본 연구에서는 이러한 문제점을 해결하기 위하여 기존의 한국역사용어시소러스를 분석하는 동시에 한국역사용어시소러스 및 국편의 다양한 정보검색 시스템을 활용하는 전문가들의 인터뷰를 실시하였다. 이를 바탕으로 이용자 정보 표현에 근접한 온톨로지 구축하고 이를 정보검색시스템에 적용하여 이용자 정보요구에 부합하는 서비스와 함께 시스템간 정보공유라는 목표를 달성하고자 하였다.

온톨로지를 구축하기 위해서는 시소러스와 같은 기존 도구를 발전시키는 방법과 대상 분야를 분석하여 초기 온톨로지를 구축한 뒤 반복적인 수정 과정을 거쳐 정련해 나가는 방법이 있다. 첫 번째 구축방법의 대표적인 예는 예

술 및 건축분야 시소러스(Art & Architecture Thesaurus, AAT)이다. 두 번째 구축방법은 지식-공학 방법론으로 Noy와 McGuinness(2000)가 제시한 것이다. 이 방법론에서는 시소러스와 같은 기존 도구의 활용 보다는 기존 도구의 참조 없이 도메인 분석에서부터 온톨로지를 구축해 나가는 것이다.

본 연구에서는 후자의 방법을 이용하여 온톨로지 모형을 설계하고자 하였다. 먼저 현재 이용하고 있는 시소러스의 문제점을 도출하고자 한국역사용어시소러스의 구조를 분석하고 주제영역이 유사한 기존의 역사 온톨로지 모형을 참고하여 온톨로지 구축을 위한 기본 모형을 도출하였다. 그 다음으로 실제 도메인에 적용할 수 있도록 역사전공자와의 인터뷰 결과 및 정보표현 예시를 통해 도출된 지식구조를 반영하여 온톨로지 모형을 수정해나가는 방법을 택하였다. 또한 역사인물 온톨로지의 확장을 통해 모의로 진행된 온톨로지 구축이 실제 시스템에 긍정적으로 적용될 가능성을 제안하였다.

2. 온톨로지 구축 방법 및 사례

2.1 온톨로지 구축 방법

온톨로지 구축은 기존의 지식을 이용한 시소러스 기반 구축 방법과 도메인 분석을 통한 개념 체계의 구축으로 나누어진다.

2.1.1 시소러스 기반 온톨로지 구축

시소러스는 특정 주제분야의 핵심 개념을 디스크립터로 표현하고, 디스크립터 간의 계층관

계와 연관관계를 표현해 주므로, 시소러스는 온톨로지를 구축하는데 좋은 바탕이 된다. 특히 예술 및 건축분야 시소러스(Art & Architecture Thesaurus, AAT)는 여러 예술품과 역사적인 건축물에 대한 디스크립터를 포함하고 있으며, 각각의 디스크립터에 적합한 패킷을 함께 제공하므로 문화예술 분야의 온톨로지 구축에도 적합하였다.

관련 연구로는 AAT의 디스크립터 중 엔틱 가구 부분을 온톨로지로 변환한 사례가 있다(Wielinga et al. 2001). 이 연구에서는 기존의 AAT가 시맨틱웹에 적합하도록 Protégé 온톨로지 구축 프로그램과 시각예술품에 대한 메타데이터인 Visual Resource Association(VRA) 버전 3.0을 이용하였다. 먼저 네 가지 유형의 디스크립터 그룹을 나누고, 각각은 엔틱 가구 클래스에 대한 슬롯의 범주 역할을 지시하였다. 예를 들면 제품 관련, 형태 관련, 기능 관련, 관리상의 디스크립터 등을 구분하고, 그 중 제품 관련 디스크립터는 생산자(제작자), 스타일/양식, 기법 등의 슬롯 역할과 연결되었다.

Ceusters 등(2005)의 연구에서는 국립암센터 시소러스(National Cancer Institute Thesaurus, NCIT)를 분석하여 온톨로지 설계에 활용하였다. NCIT는 암 연구에 이용되는 개념에 대한 정의와 코드를 제공하며, OWL 형식으로도 표현되어 있으므로 온톨로지 변환에 매우 적합하다고 볼 수 있겠다. 특정 분야의 시소러스를 기반으로 구축된 온톨로지는 기존의 지식체계를 잘 활용했다는 장점도 있지만 그 시소러스가 가지고 있는 고정된 규칙을 활용했다는 제한점 또한 포함하고 있다.

2.1.2 지식-공학 방법론

지식-공학 방법론이란 표현은 Noy와 Mc Guinness가 제시한 온톨로지 개발 방법론에서 사용한 것으로서 이 방법론에서는 시소러스와 같은 기존 도구의 활용 보다는 기존 도구의 참조 없이 도메인 분석에서부터 온톨로지를 구축해 나가는 과정을 설명하고 있다. 이 방법론에서는 와인을 사례로 하여 온톨로지 구축과정을 7단계로 구분하고 있다(Noy and McGuinness 2000).

Noy and McGuinness(2000)가 제시한 온톨로지 구축 단계는 온톨로지를 구축하고자 하는 대상 분야와 범위 설정을 통해 해당 온톨로지의 유형과 용도를 결정하는 단계로 시작되었다. 구체적인 단계에서는 대상 분야에 주요한 용어를 열거하고 이를 정리함으로써 클래스, 속성의 슬롯을 정의하였다. 마지막으로 클래스에 해당되는 개별 사례를 생성하였다. 그 중 두 번째 단계에서는 기존 온톨로지의 재활용을 고려하고 있지만 보통은 기존 온톨로지가 없다고 간주하고 새로운 개념 체계를 구축하였다.

지금까지 2.1에서는 온톨로지를 구축하는 방법론을 살펴보았다. 하나는 기존의 지식체계를 이용하는 것이었고, 두 번째는 새로운 지식체계를 생성하는데 초점을 맞추고 있다. 본 연구에서는 기존의 시소러스기반으로 하지 않고 그 반대로 기존의 정형화된 시소러스의 규칙을 좀 더 확장하기 위해 이용자의 의견이 반영된 새로운 지식체계를 구축하고자 하였다.

2.2 역사분야 온톨로지 구축 및 활용 사례

여기에서는 본 연구의 주제영역인 역사분야의 온톨로지에 대해 정리하였다.

2.2.1 FDR 역사 온톨로지

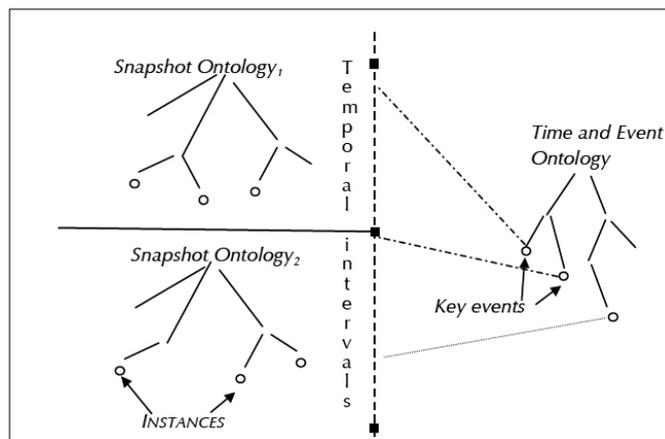
루즈벨트 도서관 프로젝트(FDR/Pearl Harbor project)에서는 프랭클린 루즈벨트 도서관(Franklin D. Roosevelt Presidential Library, FDR)에 소장된 문서를 효과적으로 검색하기 위해 FDR 역사 온톨로지(FDR historical ontology)를 개발하였다.

FDR에는 진주만 공격 이전의 10년 동안에 걸친 사건에 대한 문서가 소장되어 있다. 이에 이 프로젝트에서는 역사적인 문서는 시대에 따라 다르게 해석될 수 있으므로 하나의 고정된 온톨로지만을 사용할 수 없다고 판단하고, 다음과 같이 FDR 온톨로지를 설계하였다(Ide and Woolner 2007).

역사적인 문서를 다루기 위한 온톨로지는 특정 '사실'이 시간의 흐름에 따라 다르게 인식될 수 있는 문제점을 해결해야 한다. 하나의 온톨로지는 '그 당시의 맥락 정보(temporal contextualization)'를 표현하는 것이므로, 시간이 경과함에 따라 여러 버전의 온톨로지가 개발될 수

있다. 따라서 역사 온톨로지는 복수의 관점을 가진 온톨로지를 통합하는 과제도 안고 있다. Ide와 Woolner는 이를 각각 공시적인 관점(synchronic view)과 통시적인 관점(diachronic view)으로 구분하고 <그림 1>과 같이 표현했다. <그림 1>을 보면 공시적인 관점에서 개별 온톨로지가 각각 존재할 뿐 아니라, 이를 통시적인 관점에서 통합적으로 바라보는 또 다른 온톨로지도 존재하고 있다. 그리고 공시적인 온톨로지는 역사적인 사건에 따라 변화하게 되는데, 그 기준점이 되는 것을 '핵심 사건(Key events)'으로 표현하고 있다.

FDR 온톨로지는 역사 온톨로지의 특성으로서 시간과 사건의 개념을 본격적으로 다루었다는 특징이 있다. 그런데 공시적 관점의 개별 온톨로지가 시간의 흐름에 따라 변화할 때 인스턴스의 변경은 용이하지만 클래스의 변경은 어렵다는 단점이 있다. 특히 온톨로지 내의 개체를 정의할 때 한시적으로 존재하는 클래스인지 지속적으로 존재할 클래스인지를 구분하는 기



<그림 1> 역사 온톨로지의 개념 구조

(출처: Ide and Woolner 2007)

준이 명확하지 않다는 과제도 남아 있다.

그러나 하나의 온톨로지는 공시적인 관점에서 다루어야 하며 역사적인 사건들로 인해 온톨로지가 갱신되면 개별 온톨로지를 통합할 수 있는 통시적인 관점이 필요하다는 점은 향후 역사 온톨로지의 구축에 있어 지속적으로 반영해야 할 점이다.

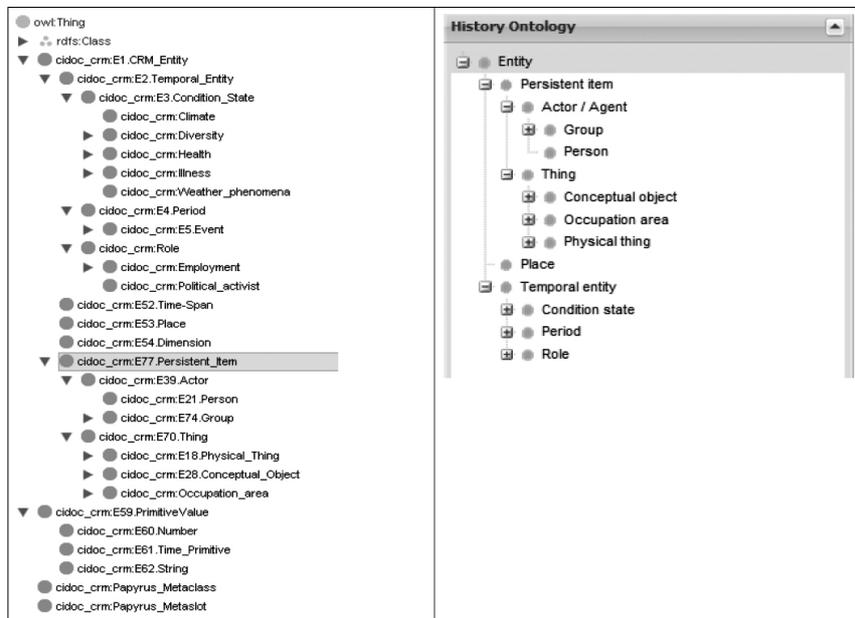
2.2.2 파피루스 역사 온톨로지

역사 분야에서는 CIDOC Conceptual Reference Model(이하 CRM이라고 함)을 바탕으로 개발한 파피루스 역사 온톨로지(Papyrus History Ontology)가 있다. 파피루스는 뉴스 아카이브에 속한 문화유산 컬렉션과 해당 컬렉션의 역사적인 속성을 연계하기 위해 개발된 것이다. 그리고 이 과정에서 다양한 매체로 분산된 컬렉션

의 상호운용성을 확보할 수 있는 개념들로 온톨로지가 선택되었다.

파피루스 역사 온톨로지의 기본구조는 CRM을 이용한다. CRM은 다양한 정보원으로 표현된 문화유산 정보의 교환을 지원하기 위한 것이므로, 분산된 정보원을 인터넷과 같이 보다 포괄적인 정보원으로 변환하는데 필요한 정의와 설명을 제공한다. 또한 CRM은 자원을 기술하는데 사용되는 특정 요소나 용어보다는 데이터 구조와 그 관계를 정의함으로써 의미론적인 상호운용성을 보장한다(ICOM/CIDOC 2008).

CRM과 파피루스 역사 온톨로지의 기본 클래스를 비교하면 <그림 2>와 같은데, CRM의 E2.Temporal_Entity 클래스와 E77.Persistent_Item 클래스가 파피루스 온톨로지에도 나타나고 있음을 알 수 있다. 특히 이 E2와 E77 클



<그림 2> CRM(왼쪽)과 파피루스 역사 온톨로지(오른쪽)의 클래스 비교
(출처: Kiyavitskaya 2010)

래스는 특정 시간 범위 동안 변화하는지의 여부에 따라 구분되는 특성을 가지고 있어 역사적인 개체를 다루는데 적합하다.

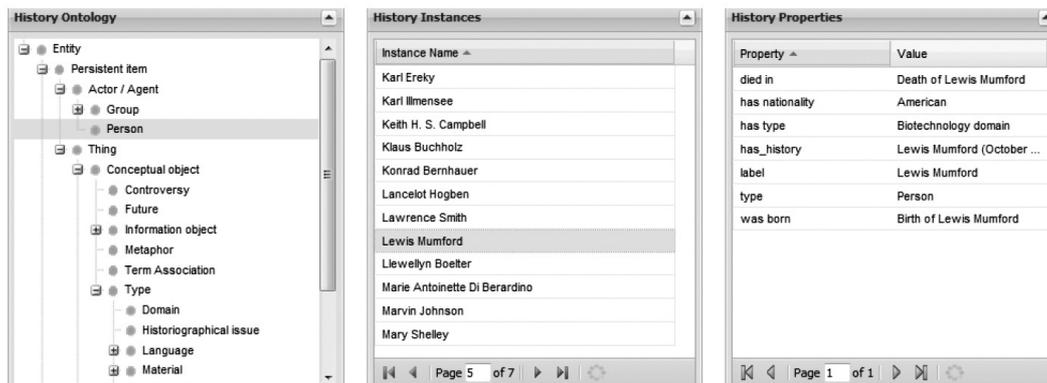
또한 파피루스 역사 온톨로지의 개체 중에서 인물(person)은 CRM의 E77 클래스와 동일하게 행위자(Actor/Agent) 클래스의 하위에 속해 있으며 개별 사례는 <그림 3>과 같다. <그림 3>의 왼쪽을 보면 여러 클래스 중 인물 클래스가 선택되어 있으며, 가운데에는 개별 사례에 해당하는 인물명이 열거되어 있다. 인물 중 Lewis Mumford라는 사람이 선택되었는데, 해당 인물의 속성과 속성값이 오른쪽에 제시되어 있다. 파피루스 온톨로지의 경우에는 인물 클래스의 속성으로 출생 및 사망 정보, 국적, 활동분야 등이 있다.

파피루스는 뉴스 분야의 개념을 구조화하기 위한 온톨로지로서 뉴스 아카이브 시스템의 일부가 된다. 파피루스의 전체 시스템 구조상 역사 온톨로지는 도메인 온톨로지의 일부로 사용된다. 즉, 다양한 매체로 구성된 뉴스 아카이브의 정보원을 효과적으로 검색하기 위해 온톨로

지를 구축하였으며, 온톨로지를 활용하기 위해 온톨로지 관리 도구가 함께 개발되었다.

이용자 측면에서 파피루스는 역사 연구자, 학생, 언론인, 문화유산 관리기관 등을 대상으로 개발되었는데, 역사 연구자의 관점에서는 다음과 같은 기대효과가 있다(Papyrus History Ontology [Online]). 첫째, 기존의 검색 모형에서는 검색 결과가 정적이며 질의를 수정할 수 있는 옵션이 거의 제공되지 않았기 때문에 관련 자료(related sources)를 검색하기가 어려웠지만, 파피루스는 더욱 다양한 맥락정보를 제공할 것이다. 예를 들면, 앞으로 자동차 운송을 연구하는 역사가는 철도나 항공 운송에 관한 뉴스도 검색할 수 있다. 둘째, 연구자가 뉴스를 이용하기 위해서 개별 신문을 각각 검색하거나 직접 기관에 방문해서 신문을 열람하지 않아도 웹을 통해 통합적으로 검색할 수 있다. 또한 연구자들은 검색 키워드에만 초점을 맞추지 않고, 보다 자연스러운 방식으로 정보를 검색할 수 있다.

지금까지 역사분야라는 도메인에 관한 주제적인 접근에 도움이 되는 사례들을 정리하였다.



<그림 3> 파피루스 역사 온톨로지의 인물 클래스 사례

(출처: Kiyavitskaya 2010)

이를 통해 역사에서 중요하게 다루어야 하는 속성을 제시하고 특히 역사분야에서의 인물 표현에 대한 패킷들에 대해서 설명하였다.

3. 온톨로지 구축을 위한 이용자 요구 분석

본 연구가 역사분야의 온톨로지 구축에 대한 가능성을 알아보기 위한 것이므로, 연구가 시작되기 전에 역사와 관련된 한 부분을 중심으로 모형화하기로 결정하였다. 실제 온톨로지는 모든 개념들이 연결되면서 하나의 의미망을 이루기 때문에 역사분야에서 중요하게 다루어지는 부분을 한정하는 자체가 온톨로지를 구축하는데 방해요소로 작용하게 된다. 그러나 본 연구가 국편 정보검색시스템의 장기적인 발전 계획을 설계하는 기반 연구로 활용된다는 점을 감안하여 온톨로지 구축 범위를 한정하였다. 본 연구에서 온톨로지를 구축하고자 하는 부분은 '한국근현대사 인물'이었으며 그 이유는 다음과 같다.

첫째, 역사정보검색시스템의 이용자들이 검색결과에 가장 만족하지 못하는 부분이 인물 검색이라는 점을 감안하였다. 인물 검색은 동명이인, 다양한 표기법 등의 문제로, 하나의 용어로 관련 개념을 찾는 것이 어렵다.

둘째, 위와 같은 검색의 어려움에도 불구하고 역사분야에서 이용자들이 가장 많이 활용하는 용어가 인물이라는 점이다. 역사분야에서는 인물이라는 정보를 중심으로 많은 사건, 활동, 사상 등이 연계될 수 있다.

셋째, 우리나라 역사 중에서 인물들에 대한

정보가 가장 많이 남아 있는 시대가 한국근현대사이므로 좀 더 다양한 개념 정리가 요구되었다.

본 연구에서는 온톨로지를 구축하는 방법으로 기존 시소러스 활용이 아닌 초기 온톨로지를 구축한 뒤 반복적인 수정 과정을 통해 정련화하는 방식을 택하였다. 이러한 방식을 택한 이유는 다음과 같다. 실제 국편에서는 한국역사용어시소러스(<http://thesaurus.history.go.kr>)를 제공하고 있다. 한국역사용어시소러스는 시소러스가 갖는 특징, 즉 특정 주제분야의 핵심 개념을 디스크립터로 표현하고 디스크립터 간의 계층관계와 연관관계를 표현해 주고 있다. 또한 역사라는 하나의 도메인으로 볼 때, 한국역사용어시소러스는 잘 구조화된 지식체계임은 틀림없다. 그러나 현재 시소러스는 역사정보검색에서 이용자의 요구를 정확하게 처리하지도 못하며, 이용자가 이해하고 인식한 의미범주를 충분히 표현하지도 못하고 있다. 그러므로 본 연구에서는 한국역사용어시소러스와 역사전공자들의 지식체계간의 어떠한 차이가 있는지를 분석하고자 하였다. 이를 통해 기존의 한국역사용어시소러스의 문제점을 도출하는 동시에 역사분야의 고도화된 지식체계 발전 방안을 제시하였다.

3.1 이용자 요구 분석

이용자 요구 분석은 국편에 재직 중인 역사전공자들과 역사정보 시스템 운영자들을 중심으로 이루어졌다. <표 1>은 요구 분석을 위해 자료 제공 및 인터뷰에 응한 전문가 다섯 명의 전공/국편 업무 및 본 작업에서의 역할을 정리하였다.

〈표 1〉 인터뷰에 참여한 전문가 전공/업무 및 역할

전문가	전공 및 국편 업무	역 할
A, B, C	역사(근현대사) 역사정보화	• 개념에 대한 마인드맵 작성(인물 중심으로 개념 확장)
D	역사(근현대사) 역사정보시스템 구축 및 관리	• 개념에 대한 마인드맵 작성(인물 중심으로 개념 확장) • 한국역사용어시소러스 분석
E	정보학 역사정보시스템 구축 및 관리	• 개념을 시스템으로 표현하는 방식에 대한 설명 • 온톨로지에 대한 설명 • 한국역사용어시소러스 분석

역사분야 전문가들의 면담은 총 여섯 차례를 걸쳐서 집단 및 개인별로 이루어졌으며, 면담을 통해 아래와 같은 내용을 파악하였다.

- 근현대사 인물에 대한 마인드 맵¹⁾ 작성
- 한국역사용어시소러스나 국편에서 제공하는 검색시스템 문제점 제시
- 본 연구에서 구축한 온톨로지 모형 분석 및 확인
 - 실제 클래스, 속성, 인스턴스 입력

위의 전문가 인터뷰의 내용 중 비증있게 분석한 것은 한국역사용어시소러스의 문제점을 도출하는 것이었다. 이러한 과정을 통해 실제 연구자들이 가지고 있는 역사분야 인물 관련 개념과 한국역사용어시소러스에서 보여주고 있는 인물관련 정보가 어떠한 차이를 가지고 있는가를 분석하였고, 앞으로 역사정보시스템의 구축 방향을 제안할 수 있었다.

3.2 한국역사용어시소러스 분석

한국역사용어시소러스에서 제공되는 정보를

기준으로 한국역사용어시소러스의 문제점을 분석하였다. 이와 관련하여 예를 들어 설명하고자 한다.

해당 사이트의 검색창에 ‘안중근’을 입력하면, 안중근(安重根), 안중근사기(安重根史記), 안중근의사유묵(安重根義士遺墨), 안중근의 이토히로부미사살(安重根의伊藤博文射殺) 등과 같은 네 가지의 용어 검색결과가 제시된다. 여기에서 용어 검색결과 중의 하나인 안중근(安重根)을 클릭하면, 그 결과가 〈그림 4〉와 같다. 〈그림 4〉를 간략하게 정리하여 개념지도로 나타내면 다음의 〈그림 5〉와 같다. ‘안중근’이라는 개념은 사전정보(주요설명, 연도, 한국역사용어 분류 등과 같은 하위 항목으로 정리), 유의어 및 관련어 등으로 표현되었다.

〈그림 5〉와 전문가 A가 제시한 개념지도(그림 6 참조)를 비교해보면 다음과 같은 특징이 있다. 첫 번째는 〈그림 6〉에 비해 〈그림 5〉의 표현의 정밀성(granularity)이 떨어짐으로써 ‘안중근’이라는 개념이 가지는 다양한 특성들을 표현하지 못했다는 점이고, 두 번째는 한국역사용어시소러스 구조와 이를 사용하는 이용자의 개념구조에는 차이가 있다는 점이다.

1) 마인드맵 소프트웨어 Joerg Mueller's FreeMind v.0.8.1을 이용함.

← 안중근(安重根) ▶ 영문

→ 대한제국기 교육자·의병장·의사, 황해도 해주 출신, 1909년 만주 하얼빈역에서 이토 히로부미를 사살.

연도 : 1879-1910

분류 : 인명>성격별>교육·연문인>교육자
 인명>성격별>사회운동가>의사
 인명>성격별>사회운동가>의병
 인명>시대별>대한제국기
 인명>지역별>황해도

★ 유의어 :

안중칠(安重七)

토마스(多默)

★ 관련어 :

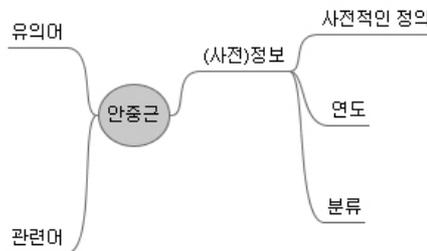
단지동맹(斷指同盟)
 동양평화론(東洋平和論)
 삼흥학교(三興學校)@안중근
 안명근(安明根)
 안중근사기(安重根史記)
 안중근의이토히로부미사살(安重根의伊藤博文射殺)

돈의학교(敦義學校)
 사죄단(謝罪團)
 안경근(安敬根)
 안명건(安命健)
 안중근의사유묵(安重根義士遺墨)
 조도선(曹道先)

〈그림 4〉 한국역사용어시소러스; ‘안중근’ 검색화면



〈그림 5〉 ‘안중근’에 대한 전문가 A의 간략 개념지도



〈그림 6〉 한국역사용어시소러스에서 제시된 ‘안중근’ 간략 개념지도

한국역사용어시소러스의 검색결과화면인 <그림 4>의 정보 중에서 사람이름이라고 간주될 수 있는 단어는 ‘안응칠’, ‘토마스’, ‘안명근’, ‘안정근’, ‘안병찬’, ‘조도선’ 등이 있다. 이때 한국역사용어시소러스의 이용자들은 이러한 이름들과 검색어인 ‘안중근’과 어떠한 관계를 가지고 있는가라는 점에 많은 관심을 가지고 있다.

이와 관련해서, <그림 7>을 참고할 필요가 있다. <그림 7>의 전문가 A의 개념지도는 ‘안중근’이라는 특성 중에서 관련 인물들과 그 관계를 정의하고 있다. <그림 7>은 ‘안중근’과 ‘안명근’은 친척의 관계이며, ‘안중근’의 아명이 ‘안응칠’인 것을 정의하였다. 물론 <그림 4>에서 ‘안응칠’과 ‘안명근’을 클릭하면 문장으로 통해 그 개념의 주요설명이 제시되고, 그 주요설명 안에서 ‘안중근’과 관련된 내용을 파악할 수 있다. 이는 잘 짜여진 구조로 정보를 연결하는 것이 아니라 각 개념들의 설명을 통해 이용자의 이성적인 판단으로 정보 연결을 인지하는 것뿐이다. 현 시소러스 구조에서는 이용자(인

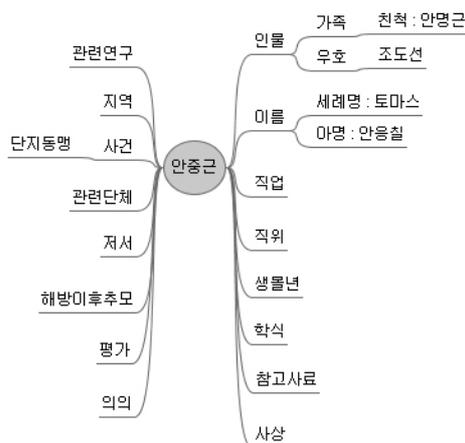
간)의 인지를 바탕으로 정보연결이 가능하기 때문에 시스템(기계)이 자동적으로 정보 처리를 할 수 없다는 제한점이 있다.

물론 현재 한국역사용어시소러스에서도 ‘유의어’와 ‘관련어’를 통해 해당 개념과 동일한 의미를 가지고 있는 용어와 해당 개념과 연결될 수 있는 용어들을 구별하고 있으나 용어들 간의 관계를 설명할 수 있는 의미 구조가 결여되어 있다.

종합적으로, 현 한국역사용어시소러스는 다음과 같은 문제점을 가지고 있다.

첫째, 역사를 전공하는 이용자들의 개념구조를 현 한국역사용어시소러스에서는 반영하지 못한다는 점이다. 일반적으로 역사분야의 전문가나 역사에 관심을 가지고 있는 이들은 단순한 개념을 찾는 것이 아니라 개념과 개념을 연결하는 의미를 찾고자 한다는 것이다.

둘째, 현 한국역사용어시소러스는 용어 표현 자체에 대한 정밀성이 떨어져서 개념의 다양한 특성을 표현하지 못한다는 점이다. 한국역사용



<그림 7> ‘안중근’에 대한 전문가 A 개념지도; 인물 및 이름부분을 확장

어시소리스는 하나의 용어에 대한 개념과 다른 용어와의 관계를 단순하게 유의어 및 관련어로만 표현하고 있다.

이와 같은 문제점을 개선하기 위해 4장에서는 역사정보를 이용하는 전문가들의 이용행태를 반영한 역사정보 온톨로지를 설계하였다.

4. 역사용어 온톨로지 설계 및 구축

본 연구에서는 역사를 중심으로 한국 근현대 시대의 인물 개념을 표현하는 ‘한국근현대사 인물(이하 역사인물이라고 함)’이라는 도메인을 범위로 설정하고 온톨로지를 설계하였다. 이와 같은 범위가 설정된 이유는 다음과 같다.

첫째, 국편에서 관리하는 정보의 범위는 역사라는 주제 분야를 총망라하고 있지만, 본 연구의 의도는 역사분야의 온톨로지 구축의 가능성을 실제 증명하는 것으로써 그 범위를 축소하여 정리하였다.

둘째, 역사를 표현하는 패킷은 너무 다양하다 못해 존재의 유무에 상관없이 모든 개체들에 대한 정의가 필요하므로 그 범위를 축소하고자 하였다. 본 연구에서는 우리나라를 중심으로 한 시대구분을 통해 범위를 축소하였다.

셋째, 역사를 바라볼 때 시대 구분을 통해 설명하기도 하지만 개별 개념으로 가장 많이 다루어지는 것은 인물, 조직, 사건, 장소 등이다. 시대를 한국의 근현대사로 한정했을 때 다양한 역사정보들을 연결할 중심의 개념이 필요했으며, 이에 대한 결정은 역사전공자들의 의견을 수용하였다. 논의 결과 ‘인물’이라는 개념이 가

장 포괄적인 정보를 가지고 있으며, 다양한 정보자원과도 연결됨을 알 수 있었다. 즉 ‘인물’은 명칭, 조직, 사건, 저작, 장소, 사상 등 다양한 패킷 분석이 이루어져야 하는 개념인 것이다.

4.1 역사인물 온톨로지 설계

4.1.1 온톨로지 클래스 및 속성 설계

본 장에서는 역사용어 클래스와 해당 클래스의 데이터타입과 오브젝트타입 속성을 구분하여 설계하였다. 클래스가 가지는 속성은 크게 데이터타입과 오브젝트타입 속성으로 나뉘고, 데이터타입속성은 클래스가 지닌 고유한 특성을 말하며 이는 클래스에 대한 설명을 일정한 규칙 안에서 제시할 수 있는 기반을 제공한다. 오브젝트타입 속성은 클래스간의 관계를 확장하여 표현하고 관계를 설정하기 위한 규칙을 제공한다.

본 연구에서는 전문가들과의 인터뷰를 통해 역사 인물에 대한 개념을 구체화하고 클래스 정의를 실시하였다. 클래스의 깊이(depth)는 3 단계로 의미를 구체화하여 1, 2, 3 계층의 클래스를 정의하였으며, 데이터타입 속성의 깊이는 2 단계로 분리하여 1, 2 계층의 데이터타입 속성을 정의하였다. 최종적으로는 인물, 인적정보, 조직, 교육기관, 정부기관, 사건, 활동, 저작, 사상 등 9개의 최상위 클래스를 작성하였고, 클래스 2는 18개, 클래스 3은 14개로 제시하였다. 그중 클래스 1에 대한 설명과 그와 관련된 데이터타입 속성을 제시함으로써, 클래스 1의 항목들에 대한 정의를 다음과 같이 정리할 수 있다.

- 인물: 한 개인에 대한 정체성(identity)을

확인하는 클래스

- 식별번호, 이름(한글명, 한자명, 영문명), 일본어표기 등
- 인적정보: 정체성이 확인된 개인에 대한 일반적인 정보를 확인하는 클래스
 - 성별, 거주_장소, 개인_연락처, 종교 등
- 조직: 근현대사에서 확인된 기관에 대한 클래스(단, 교육 및 정부기관은 제외됨)
 - 조직_한글명, 조직_별칭, 설립인 등
- 교육기관: 근현대사에서 확인된 교육기관에 대한 클래스
 - 교육기관_한글명, 교육기관_별칭 등
- 정부기관: 근현대사에서 확인된 정부기관에 대한 클래스
 - 정부기관_한글명, 정부기관_별칭 등
- 사건: 근현대사의 주된 역사적인 사건에 대한 클래스
 - 사건_명칭, 사건_당시개념, 사건_주제 등
- 활동: 근현대사의 주된 역사적인 활동에 대한 클래스(조서, 인터뷰, 구술, 성명 등을 포함)
 - 활동_명칭, 활동_주도인물, 활동_주도_조직 등
- 저작: 특정 인물이 책임을 가지고 생산한 지적산물에 대한 클래스
 - 저작_명칭, 공개_유형 등
- 사상: 특정 인물의 사건, 활동, 저작에 바탕이 되는 신념에 대한 클래스
 - 사상_명칭, 사상_주도자, 사상_주도_조직 등

위에서 정의된 클래스의 특성들을 정리한 데이터타입 속성은 속성 1계층은 66개, 속성 2계

층은 8개로 작성되었다. 또한 73개의 오브젝트 타입을 제안하였으며, 이는 4.2.2장에서 제시될 역사인물 온톨로지의 확장에도 많은 영향을 주었다.

4.1.2 인스턴스 입력시 유의점

온톨로지 구축시 인스턴스 입력에서 고려해야 할 점들은 다음과 같다. 개념 정의도 중요하지만 개념에 실제 적용되는 인스턴스의 범위를 확인하는 것도 중요하다.

역사분야에서는 이름의 변화나 사건 및 활동 상황의 구별이 어려우며 특히 실제 인스턴스를 입력할 때 관련된 클래스 및 속성을 파악하는 것이 어려울 수 있다. 이와 관련하여 아래와 같은 특수한 경우들을 정리하였으며, 앞으로 실제 시스템이 구축되었을 때 더 많은 추가 사항들이 발생할 수 있다.

■ 명칭이 변한 경우

- 인물
 - 개명으로만 표현: 시간구분 필요함
 - > 인물>이름>별칭>개명 클래스에 추가
- 조직, 교육기관, 정부기관
 - [조직/교육기관]이름만 약간 변경: 이칭으로 표현: 시간구분 필요
 - > 조직>조직_별칭>조직_이칭 클래스에 추가
 - [정부기관]이름이 바뀌면 무조건 새롭게 입력
 - > 정부기관 클래스에 추가
 - 전후관계: '명칭변경_특성' 속성: '전후'로 표현
 - > 오브젝트타입속성 중 '조직/교육기관/정부기관_명칭바뀌다'로 연결
 - 관련기관

- > 오브젝트타입속성 중 '조직/교육기관/정부 기관간의_관련이있다'로 연결
- 통합: '명칭변경_특성' 속성: '합당, 통합' 등 하나로 표현
- > 오브젝트타입속성 중 '조직/교육기관/정부 기관 명칭바뀌다'로 연결
- 분리: '명칭변경_특성' 속성: '분리'로 표현
- > 오브젝트타입속성 중 '조직/교육기관/정부 기관 명칭바뀌다'로 연결

■ 활동 클래스관련

- 활동>특화된활동>인터뷰/구술: 신문기사 안에서 왈~, 답~에 적용됨

지금까지 역사인물 온톨로지 클래스, 속성 설계 및 인스턴스 입력시 주의해야 할 점들을 정의해보았으며, 이를 기반으로 4.2에서는 온톨로지 구축 틀을 이용하여 역사인물 온톨로지를 구축하였다.

■ 인물 클래스관련

- 인물>이름>호: 아호를 포함
- 인물>이름>별칭: 일반적으로 통용되는 것

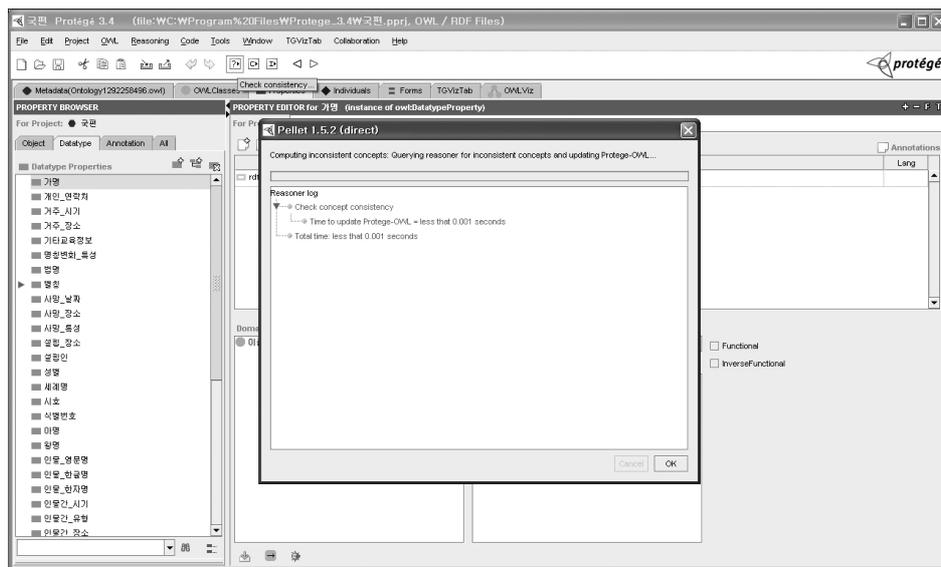
■ 사건 클래스관련

- 사건>당시개념: 육하원칙으로 진술: 현 시소러스 참고

4.2 역사인물 온톨로지 구축 및 확장

4.2.1 역사인물 온톨로지 구축

본 연구에서는 Protégé 3.4.4(<http://protege.stanford.edu/>)를 통해 역사인물 온톨로지를 구축하였으며, 또한 컨스트럭터와 공리에 대한 규칙은 Pellet 1.5.2 reasoner를 통해 일관성과 신뢰성을 확인하였다(그림 8 참조).



<그림 8> Pellet 1.5.2 reasoner 확인 화면

본 연구에서 Protégé로 구축한 온톨로지 파일(OWL)의 일부는 아래와 같다.

아래에서 제시한 OWL 문서의 주요 내용은 인물과 교육기관 클래스를 연결하는 오브젝트 타입 '인물_교육기관_소속되다'를 표현한 것이

다. '인물_교육기관_소속되다'는 '교육기관_소속_시기' 및 '교육기관_소속_유형' 등의 데이터타입속성을 함께 포함하여 상황에 대한 정보를 좀 더 정확하게 표현할 수 있었다.

```

<?xml version="1.0"?>
<rdf:RDF
  xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns="http://www.owl-ontologies.com/Ontology1292258496.owl#"
  xml:base="http://www.owl-ontologies.com/Ontology1292258496.owl">
  <owl:Ontology rdf:about=""/>
  .....
  <owl:ObjectProperty rdf:ID="인물_교육기관_소속되다">
    <rdfs:range>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <rdfs:Class rdf:about="#교육기관"/>
          <owl:Restriction>
            <owl:someValuesFrom rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
            <owl:onProperty>
              <owl:DatatypeProperty rdf:ID="교육기관_소속_시기"/>
            </owl:onProperty>
          </owl:Restriction>
          <owl:Restriction>
            <owl:someValuesFrom rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
            <owl:onProperty>
              <owl:DatatypeProperty rdf:ID="교육기관_소속_유형"/>
            </owl:onProperty>
          </owl:Restriction>
        </owl:unionOf>
      </owl:Class>
    </rdfs:range>
    <rdfs:domain rdf:resource="#인물"/>
  </owl:ObjectProperty>
  .....

```

4.2.2 역사인물 온톨로지 확장

앞서 4.2.1에서 기술한 바와 같이 본 연구에서는 전문가들의 의견이 반영된 역사인물 온톨로지를 구축하였다. 본 연구에서 제시한 역사인물 온톨로지는 국편에서 앞으로 개선하고자 하는 고도화된 역사정보시스템의 기반 모형으로 활용될 수 있으며, 국편의 한국역사용어시소러스를 발전시킨 모델로 독립적인 하나의 정보시스템으로 발전할 수 있다.

위와 같은 역사인물 온톨로지가 실제 시스템에서 활용될 수 있도록 여러 가지의 시나리오를 작성해보면 다음과 같다.

• 기본적인 확장

역사인물 온톨로지는 특정 인물에 대한 개념이 다양하게 표현되며, 표현의 관계를 함께 제공한다. 이는 검색의 효율성을 높이는 동시에 시각적으로 정보를 표현할 수 있는 시스템을 제공한다. 예를 들어 특정 인물에 대한 가족관계, 다양한 이름의 표현 등이 하나의 개념도로 표현될 때 이용자는 검색 질문에 대한 확장 및 구체화 기회를 얻을 수 있으며, 이런 다양한 표현 자체만으로도 이용자의 요구에 부합할 수 있을 것이다.

• 동명이인 처리

역사분야에서의 동명이인 처리가 불완전한 경우에는 정보검색의 정확성이 현저히 낮아진다. 동명이인의 처리는 한 인물을 중심으로 다양한 정보를 조합하여 활용하면 간단히 해결될 수 있다. 즉 특정 인물의 출생연도, 한자명, 직업 및 출신학교들을 함께 제공하는 것이 바람직하다.

• 특정 인물에 대한 사상적인 변화

역사분야에서의 특정 인물의 사상적인 변화는 매우 중요한 정보로 활용된다. 특히 한국근현대사에서는 사상의 변화에 따라서 개인이 가담한 사건 및 활동에도 많은 영향을 주게 된다. 사상적인 변화는 인물과 연결된 조직, 사건, 활동, 저작과 관련된 정보를 시간순으로 배열하면 인물의 사상적인 변화를 감지할 수 있다.

• 인물들 간의 관계 파악

역사분야를 연구하는 전문가들이나 역사에 관심을 가지고 있는 일반적인 이용자들에게 가치 있는 정보 중 하나는 특정 인물들 간의 관계 유형을 파악하는 것이다. 역사정보 이용자들은 특정 인물들이 동지/우호관계 또는 적대관계였는지를 파악하고자 하며, 또한 관련된 사건, 활동, 조직, 사상 등이 함께 제시되기를 바란다. 동지/우호 및 적대관계를 규정하는 정보를 제공하기 위해서는 사건, 활동, 조직, 사상 간의 관계를 규정하고, 이러한 관계 규정안에서 비슷한 시기에 위치한 특정한 인물들을 배열하면 된다.

• 저작을 중심으로 한 저작자인가 주제인가를 파악

역사정보를 검색할 때 이용자들은 인물과 저작 간의 관계를 좀 더 정확하게 파악하고자 하였다. 특정 저작을 기준으로 하여, 특정 인물에 의해 생산된 것인지 아니면 그 저작이 특정 인물에 대한 것인지를 확인할 수 있어야 한다. 첫 번째 경우에는 오브젝트타입 속성을 이용하여 특정 인물과 특정 저작을 '집필하다'라는 속성으로 연결할 수 있을 것이다. 또한 특정 인물의 호, 별칭, 한자표기 등을 이용하여 질의어를 확

장할 수 있는 정보검색 환경을 제공하고 동시에 정확한 정보검색결과를 확보할 수 있다.

• 인물-저작-사상 간의 관계 파악

특정 인물은 특정 사상의 영향을 받아서 특정 저작을 생산하는 것이다. 이와 같은 관계를 정보검색 환경에서 제공할 수 있다면 이용자의 만족도는 매우 높아질 것이다. 오브젝트타입 속성인 '기반사상이다'를 통해 저작과 인물, 조직(교육/정부기관 포함) 등을 연결한 후 이를 확장하여 사상 클래스와 다시 연결하였다. 두 번째 연결은 오브젝트타입 속성의 특성을 제한하는 데이터타입 속성을 정의하고, 그 속성의 값을 사상 클래스와 연결시켰다.

본 장에서는 역사인물 온톨로지의 확장을 통해 모의로 진행된 온톨로지 구축이 실제 시스템에 긍정적으로 적용될 가능성을 제안하였다. 위와 같은 역사인물 온톨로지 확장은 전문가 인터뷰를 통해 그들이 가장 원하는 정보검색 유형을 선택하여 제안한 것이다. 그러므로 앞으로 더 많은 검색 패턴을 대비한 온톨로지 확장이 요구된다.

5. 결론

본 연구는 국편에서 활용할 수 있는 온톨로지 모형을 구축하고자 하였다. 특히 한국근현

대사 인물중심이라는 범위를 설정하고 역사인물 온톨로지를 모의로 구축하였다.

역사인물 온톨로지를 구축하기 위해 기존의 한국역사용어시소러스를 분석한 후 그 문제점을 개선하는 방향으로 본 연구는 진행되었다. 동시에 한국역사용어시소러스 및 국편의 다양한 정보검색 시스템을 활용하는 전문가들의 인터뷰를 실시하였다. 한국역사용어시소러스의 개념 표현 기준과 전문가들이 작성한 마인드맵을 분석하여 현재 시스템과 전문가간의 정보 표현의 차이를 확인하였다. 현재 시스템은 시소러스라는 점을 감안할 때 개념간의 관계표현이 다양하지 않았으며 그 설명 또한 빈약하였다.

본 연구에서는 역사정보 이용자들의 정보요구에 부합하고 잠재력있는 정보관리시스템을 구축하기 위한 기반이 되는 역사용어 온톨로지를 구축하였다. 먼저 클래스, 속성을 정의하고, 인스턴스 입력시 고려해야 할 점들을 나열함으로써 역사정보가 갖는 고유한 특성들을 정리하였다. 또한 온톨로지의 장점인 개념 확장을 활용하여 역사정보의 다양한 특성들을 조합하고 이를 통해 새로운 정보를 제공할 수 있는 가능성을 제시하였다.

본 연구는 역사라는 분야 그리고 근현대사 인물로 범위를 제한하여 온톨로지를 구축하였다. 이 과정은 본래 온톨로지가 추구하는 지식의 연결을 위한 하나의 모형을 제안하는 것으로 지속적인 발전 및 새로운 시도가 요구된다.

참 고 문 헌

- 류주희. 2005. 문화콘텐츠 기반으로서의 한국 역사용어 시소러스. 『인문콘텐츠』, 4: 199-211.
- 박지영. 2004. 『다차원 시소러스 구축에 관한 실험적 연구』. 석사학위논문. 연세대학교, 문헌정보학과.
- 이혜원. 2008. 학교도서관과 MLA 연계를 위한 교육 온톨로지 모형 구축에 관한 연구. 『한국비블리아학회지』, 19(1): 1-18. 한국역사용어시소러스. [cited 2010.10.15]. <<http://thesaurus.history.go.kr/>>.
- Aitchison, J. and Clarke, S. D. 2004. "The Thesaurus: A Historical Viewpoint, with a Look to the Future." *Cataloging & Classification Quarterly*, 37(3): 5-21.
- Broughton, Vanda. 2004. *Essential classification*. London: Facet.
- Broughton, Vanda. 2006. "The need for a faceted classification as the basis of all methods of information retrieval." *Aslib Proceedings*, 58: 49-72.
- Joerg Mueller's FreeMind. [cited 2010.9.2]. <<http://freemind.sourceforge.net>>.
- Noy, Natalya F. and Deborah L. McGuinness, 2000. *Ontology Development 101: A Guide to Creating Your First Ontology*. [cited 2010.11.8]. <http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html>.
- Gilchrist, A. 2003. Thesauri, taxonomies and ontologies: an etymological note. *Journal of Documentation*, 59(1): 7-18.
- Wielinga, B. J., A. Th. Schreiber, J. Wielmaker and J. A. C. Sandberg. 2001. From thesaurus to ontology. Proceedings of the 1st international conference on Knowledge capture. 194-201.
- Hernández, Francisca, Luis Rodrigo, Jesús Contreras and Francesco Carbone. 2008. Building A Cultural Heritage Ontology For Cantabria. 2008 Annual Conference of CIDOC, Athens, September 15-18, 2008. [cited 2010.12.4]. <<http://cidoc.mediahost.org/archive/cidoc2008/Documents/papers/drfile.2008-06-28.pdf>>.
- ICOM/CIDOC. CIDOC CRM. [cited 2010.12.1]. <<http://cidoc.ics.forth.gr/>>.
- Ide, N. and D. Woolner, 2007. Historical Ontologies. In Ahmad, K, Brewster, C., and Stevenson, M.(eds.), *Words and Intelligence II: Essays in Honor of Yorick Wilks*, Springer, 137-152.
- La Barre, K. 2010. Facet analysis. In B. Cronin ed. *Annual Review of Information Science and Technology*. Volume 44: 243-284. Medford, NJ: Information Today.

- Kakali, Constantia, Irene Lourdi, Thomais Stasinopoulou, Lina Bountouri, Christos Papatheodorou, Martin Doerr and Manolis Gergatsoulis. 2007. Integrating Dublin Core metadata for cultural heritage collections using ontologies. [cited 2010.11.15].
<<http://eprints.rclis.org/12522/1/11.pdf>>.
- Kiyavitskaya, Nadzeya. 2010. Documentation on Papyrus Ontologies. [cited 2010.10.2].
<<http://www.ict-papyrus.eu/files/Documentation%20on%20Papyrus%20Ontologies.pdf>>.
- Noy, Natalya F. and Deborah L. McGuinness. 2000. *Ontology Development 101: A Guide to Creating Your First Ontology*. [cited 2010.11.30].
<<http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html>>.
- Ranganathan, S. R. 1967. *Prolegomena to Library Classification*. New York: Asia Publishing House.
- Papyrus History Ontology. [cited 2010.10.1].
<<http://www.ict-papyrus.eu/>>.
- Protégé. Stanford Biomedical Informatics. [cited 2010.3.11].
<<http://protege.stanford.edu/>>.
- Smith, Michael K., Chris Welty and Deborah L. McGuinness. 2004. *OWL Web Ontology Language Guide*, W3C Recommendation. [cited 2010.10.17].
<<http://www.w3.org/TR/2004/REC-owl-guide-20040210/>>.
- Gruber, Thomas R. 1993. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2): 199-220. Available at:
<<http://tomgruber.org/writing/ontology-kaj-1993.pdf>>.
- Ceusters, Werner, Barry Smith and Louis Goldberg. 2005. A Terminological and Ontological Analysis of the NCI Thesaurus. preprint version of paper in *Methods of Information in Medicine*, 44: 498-507. [cited 2010.11.22].
<<http://ontology.buffalo.edu/medo/NCIT.pdf>>.
- World Wide Web Consortium. *Web Ontology (WebONT) Working Group Charter*. [cited 2010.12.4].
<<http://www.w3.org/2001/sw/WebOnt/charter>>.