

토픽맵 기반의 기록정보 검색시스템 구축에 관한 연구*

권 창 호**

1. 서론
 - 1) 연구의 배경 및 목적
 - 2) 사례 및 선행연구
 - 3) 연구방법 및 한계
2. 기록정보 검색에서의 토픽맵 구조와 역할
 - 1) 기록정보 검색과 표현도구
 - 2) 시멘틱웹과 온톨로지의 연원
 - 3) 토픽맵의 개념 및 구조
 - 4) 토픽맵의 기록물검색시스템 구축의 의미
3. 토픽맵 기반의 기록물검색시스템 구축
 - 1) 온톨로지 모델링
 - 2) 토픽맵 구축
 - 3) 사용자 인터페이스
 - 4) 시스템 평가
4. 결론

* 이 연구는 명지대학교 기록과학대학원 석사학위논문을 수정, 보완한 것임.

** 동대문구정보화도서관 사서

[국문초록]

최근, 웹을 이용한 기록정보의 유통과 이용이 증가하고, 정보적 활용 가치가 제고되어 웹사이트를 이용한 기록정보서비스가 기록관의 중요업무로 부각되고 있다. 웹을 이용한 기록정보서비스의 핵심은 이용자가 원하는 기록정보의 검색을 용이하게 하는데 있다. 검색을 용이하게 하기 위해서는 검색시스템의 기본 메커니즘인 이용자질의와 기록정보표현의 매칭의 정확성이 요구된다. 이를 위해 기록정보 관리자들은 다양한 정보표현 도구를 이용하고 있지만, 이용자들은 여전히 정보검색 과정에서 어려움을 겪고 있다. 이를 개선하기 위해 본 연구에는 기록물의 기술정보 메타데이터를 중심으로 정보자원을 구조화하여 이용자 질의의 접근점을 확장하고, 의미있는 매칭을 통해 지식자원화된 검색결과값을 제공하기 위해 토픽맵 기반의 기록정보 검색시스템을 구축하고자 한다. 구축대상은 웹사이트를 이용하는 불특정 이용자이며, 구축범위는 국가기록포털의 기록자원 중 대통령 기록물로 선정하였다. 구축단계는 다음과 같다. 1)기록물의 기술정보 메타데이터를 중심으로 토픽맵 기반의 기록정보서비스를 위한 온톨로지 모델을 설계한다. 2)설계한 온톨로지 모델을 바탕으로 국가기록포털에서 추출한 정보자원목록을 에디터를 이용해 토픽맵으로 반입하여 검색시스템으로 구현한다. 3)구축된 검색시스템의 사용자 인터페이스에서 테스트질의를 통해 토픽맵기반 검색시스템의 특징을 확인하고 그 의미를 평가한다. 최종적으로, 의미적 추론에 의한 연관 네비게이션검색을 확인하고, 분산된 기록정보자원 간의 연계된 결과값을 통해 지식자원화의 가능성도 제시한다.

주제어: 토픽맵, 기록정보서비스, 검색시스템, 온톨로지, 시멘틱웹

1. 서론

1) 연구의 배경 및 목적

오늘날 정보기술의 급속한 발전은 기록관의 역할변화에 많은 영향을 끼치고 있다. 초창기 주로 일부 특권계층을 위해 국가적 차원에서 관리되고 보존되던 시기의 기록관은 보존과 보관 업무를 중심으로 운영되었다. 이러한 기록관 형태는 프랑스혁명을 통해 기록물에 대한 일반 국민의 접근의 기회가 보장된 이후에도 크게 변화되지 않았다. 이는 기록관 이용에 있어서의 시공간적 제약과 기록물의 정보적 활용가치에 대한 인식이 낮아 일반 국민들이 기록관 이용에 소극적이었기 때문이다. 하지만, 기록의 진본성과 설명책임성의 가치가 제고되고, 웹을 통한 기록정보의 유통과 요구가 증가하면서 웹을 이용한 기록정보서비스가 기록관의 중요업무로 부각되고 있다. 기록관에서 웹을 통해서 서비스를 제공하는 것의 가장 큰 부분을 차지하는 것은 이용자들이 인터넷을 이용해서 자신들의 소장기록을 검색하고 획득하는 것이 가능하게 하는 것이다.¹⁾ 하지만, 웹을 통한 기록정보서비스는 기록정보 관리자가 직접 이용자를 대면하여 정보를 해석해 줄 수 없고, 이용자의 질문이나 요구에 즉각적인 답변을 하지 못하므로 이용자는 기록정보 검색과정을 도움없이 직접 수행해야 한다. 웹사이트에서의 이용자 정보검색은 이용자인터넷 페이지 검색창에서 질의어를 입력하여 결과값을 추출하는 과정

1) 이희숙, 「인터넷을 통한 기록관의 서비스 활동에 관한 연구 - 교육.전시활동을 중심으로」, 명지대학교 기록과학대학원 기록관리학과 학위논문, 2005

을 의미하며, 검색시스템은 이용자 질의와 정보표현²⁾을 매칭하여 검색 결과값을 추출한다. 따라서, 검색의 정확률³⁾을 높이기 위해서는 기록정보서비스 기관 내부의 정보표현과 이용자질의의 매칭의 정확성이 전제되어야 한다. 하지만, 기록정보의 양이 폭발적으로 증가함에 따라, 이용자가 질의 시 정확한 단어, 혹은 어구를 선정하기가 쉽지 않고, 정보 제공자는 이용자 질의와 일치하도록 정보를 표현하기도 더욱 어려워져, 검색의 정확률은 향상되지 않고 있다. 이러한 문제는 국가기록포탈⁴⁾에서 진행된 이용자를 대상으로 한 요구사항 설문조사 결과(2007년 12월 6일)에도 그대로 반영되어 있다. 응답자의 과반수가 넘는 57.4%가 가장 먼저 기록물 통합/상세 검색의 개선을 요구하고 있는 실정이다. 기록정보 서비스기관은 이러한 단순 매칭에 따른 검색의 어려움을 해소하기 위해 표현의 범위를 전거제어, 시소러스 등으로 지속적으로 확장해 왔지만 가시적인 성과는 나타나지 않고 있다. 그 이유는, 기록자원이 담고 있는 정보가 색인어, 인벤토리, 전거제어, 시소러스 등의 정보표현으로 전부 표현되기 어렵고, 구축된 통제어휘⁵⁾가 원문헌의 전문(Full-Text)의 자연어집

-
- 2) 정보표현은 정보검색의 정확율을 향상시키기 위한 다양한 방안을 의미하며, 색인, 시소러스, 기술 등을 포함하는 정보학의 용어이다. 정보표현은, 문헌의 본질이 특화되고 표현될 수 있도록 문헌으로부터 어떤 요소(예, 키워드나 구)를 추출하거나 혹은 문헌에 용어(예, 디스크립터)나 주제명표목을 할당하는 것을 포함하는 개념이다.
 - 3) 정확률이란, 시스템에서 검색된 문헌 수 전체와 검색된 적합문헌 수의 비율이다. 이는 정보검색시스템의 분별능력, 즉 적합한 것과 부적합한 것을 분리하는 능력을 측정한다. 만일, 검색시스템에서 문헌 100건이 검색되었다면, 그 중 35건의 문헌이 적합한 결과값으로 판단된다면, 이 검색에 대한 정확률은 35%이다.
 - 4) <http://contents.archives.go.kr/next/content/viewMain.do>, 국가기록포탈홈페이지
 - 5) 정보표현과 검색을 위한 언어 중 자연어와 구분되는 의미로서, 어휘, 구분, 의미, 사용이 제한된 인공적 언어가 정보표현과 검색에 적용될 때, 통제어휘라고 부른다.

합의 극히 일부분에 불과하기 때문이다. 이처럼, 이용자 질의와 정보자원과의 매칭의 근원적인 한계를 극복하지 못하여, 최근에는 인공지능(AI)분야의 자연어처리 기술에서 유래한 시멘틱웹의 온톨로지가 새롭게 주목받고 있다.

시멘틱웹은 WWW(World Wide Web)의 창시자 팀 버너리스가 제안한 웹기술의 새로운 개념으로 컴퓨터가 이해할 수 있는 새로운 형태의 언어로 표현하여, 기계도 서로 의사소통이 가능하게 하는 지능형 의미기반 웹이다. 이를 위해서는 기계와 기계 사이에 정보를 공유하기 위해 상호간에 이해 가능한 용어로 개념을 정의하는 온톨로지 기술이 필요하다. 온톨로지는 인간이 객체에 의미를 부여하여 판단하는 것과 같이 컴퓨터도 객체에 의미관계를 부여하도록 모델링하는 것이다. 온톨로지를 구축하기 위한 표현 기술연구는 크게 웹의 표준을 담당하고 있는 W3C⁶⁾를 중심으로 한 RDF(Resource Description Framework)와 ISO를 중심으로 한 토픽맵(Topic Maps)기술로 나눌 수 있다. RDF 표현방식이 DB스키마부터 정보자원의 의미관계까지 세밀하게 표현해야되는 반면에 토픽맵 표현방식은 기존의 정보자원을 최대한 활용하여 시멘틱웹 구축을 자동화할 수 있다. 토픽맵은 주로 지식, 콘텐츠관리, 문헌 관리 등에서 의미 기반으로 자료를 연결하고 검색하기 위한 인간 중심의 온톨로지 구축을 위해 사용된다.⁷⁾ 토픽맵이 지금까지 가장 활발하게 응용된 분야는 콘텐츠들을 효과적으로 조직하기 위한 정보조직과 관련된 분야와 이용자들이 원하는 정보를 손쉽게 찾도록 하는 정보 네비게이션과 관련된 분야이다. 정보 네비게이션 측면에서 토픽맵은 모든 형태의

6) W3C은 World Wide Web Consortium의 약자로 웹 표준을 제정하고, 웹의 발전을 위해 1994년 창립된 인터넷 관련 국제 컨소시엄이다.

7) Graham, M, Topic Map technology the state of the art, XML 2000 Conference & Exposition, Washington, USA, 2000

색인을 작성하는데 가장 적합한 모델이라고 할 수 있으며, 지능적인 색인⁸⁾이 가능하다.⁹⁾ 이는 토픽맵이 온톨로지를 기반으로 하여, 하나의 주제 개념과 관련된 모든 정보들을 토픽을 매개로 연결하고, 서로 다른 토픽들을 다양한 유형의 관계를 활용하여 상세하게 표현함으로써 대용량정보를 정교하게 구조화한 색인을 기술할 수 있도록¹⁰⁾ 해주기 때문이다. 이와 같은 장점으로 토픽맵은 최근 전자상거래 분야, 통합전자정부, 정보자원의 지식 컨텐츠화 등 다양한 분야에서 활발히 연구되고 있다. 그 예로 「금융기관의 지식 관리 개선 방안 연구」(김현희, 2006), 「국가지식정보 온톨로지 표준개발」(신인철 외, 2006), 「텍스트 내용 지식기반의 철학 온톨로지 구축」(김정민 외, 2005)에 관한 연구 등이 있다. 하지만, 지금까지 연구되어 온 모든 분야와 긴밀하게 연관되어 있으며, 효율적인 검색이 요구되는 기록관리학 분야에서는 아직 이에 대한 연구가 전무한 실정이다. 이에 본 연구에서는 기록정보의 메타데이터를 중심으로 한 토픽맵기반의 검색시스템을 구축하여 이용자질의와 정보표현의 매칭을 향상하고, 기록정보를 자원화하여 이용자의 정보검색과정을 보다 용이하게 하고자 한다.

2) 사례 및 선행연구

온톨로지와 토픽맵에 대한 연구는 해외 선진국에서 먼저 이

8) 지능적인 색인이라는 것은 개별 정보 컨텐츠 및 웹사이트에 기술되어 있는 현재의 색인이 제공하는 것 이상의 것을 정보이용자에게 제공함을 의미한다 (2007, 국가지식정보 온톨로지 표준 개발완료 보고)

9) 신인철 외, 「국가지식정보 온톨로지 표준개발」, 온톨로지연구소, 2006

10) '같은 책'

루어졌고, 현재 국제적으로 표준화 작업이 이루어지고 있다. 대표적인 토픽맵 적용 분야는 통합전자정부 구현 프로젝트로, 정부기관간의 정보공유와 의사소통을 위한 프로젝트가 진행되어 유럽과 미국에서 진행되고 있다.

유럽연합의 ADNOM(Administrative Nomenclature) 프로젝트¹¹⁾는 각 국가별 정부부서 간 정보 교환의 어려움을 해결하기 위해 추진되었다. 시멘틱웹을 이용해 다른 어플리케이션 및 용어간의 상호 운용성을 높이고, 토픽맵을 이용해 자원간의 매핑¹²⁾, 다중언어로 된 자원의 설계, 분산 환경 하에서 안정된 지식자원의 교환을 가능하게 한다. 네델란드의 ON-TOP 프로젝트는 현존하는 정보자원으로부터 토픽맵을 신속하게 생성하고, 기존의 기술기반을 이용하여 구축되고 효율적으로 관리된다는 사실을 보여준다. 또 하나의 대표적인 토픽맵 적용 분야는 지식관리시스템과 지식의 컨테츠화를 통한 지식포털사이트 구축이다. 노르웨이의 과학기술정보 포털¹³⁾과 소비자정보 포털¹⁴⁾, ITU교육프로젝트용 웹사이트¹⁵⁾ 미국의 국제청세금정보 포털¹⁶⁾은 인간적 사고흐름을 반영한 네비게이션 인터페이스 제공 및 지식정보간의 다양한 의미적 연관정보를 제공하는 토픽맵 기반의 지식정보제공 포털 사이트이다.

지식자원들을 상호 연관성에 따라 연결 및 조직하여 지식 구조를 기술할 수 있도록 토픽맵 모델을 기반으로 지식맵을 구축

11) <http://www.cenorm.be/cenorm/businessdomains/businessdomains/iss/activity/adnom.asp>

12) 매핑(mapping)은 하나의 값을 다른 값으로 또는 한 데이터 집합을 다른 데이터 집합으로 변환, 번역하거나 2개의 데이터 집합 사이에 1:1 대응 관계를 설정하는 것이다.

13) <http://www.forskning.no/>

14) <http://www.forbrukerportalen.no/>

15) <http://www.itu.no>

16) <http://www.missouribusiness.net/irs/taxmap/about.htm>

한 지식관리시스템으로 empolis의 토픽맵 기술¹⁷⁾, 더브레인 KM SI¹⁸⁾ 등이 있다. 토픽맵 어플리케이션을 이용하여 콘텐츠들의 연관관계 설정하고 의미적으로 구조화하여 콘텐츠를 효율적으로 관리하는 기능을 제공하는 프랑스 백과사전¹⁹⁾, 미국 에너지부 기밀문서분류 가이드 및 관리, Opensource ZTM, Tu-berlin iVS 등이 있다.²⁰⁾

국내 구축 사례는 철학 텍스트들의 내용분석에 의거한 디지털 지식 자원 구축을 위한 서울대학교 철학사상연구소의 철학 토픽맵²¹⁾, 한국생산기술연구원의 토픽맵 기반의 사출금형 기술 정보 통합 및 정보검색시스템²²⁾ 구축, 동대문구정보화도서관의 지식인 간의 관계구조를 토픽맵으로 구축한 지식인마을 시스템²³⁾ 등이 있다. 아직 실제 구축사례는 많지 않지만, 학문적 연구와 프로토타입은 다양한 분야에서 실시되고 있다. 남영준(2005)은 토픽맵의 기본요소인 토픽과 대상물, 연관관계 등을 활용하여 시소러스관리를 위한 구조화 방안을 제안하였다. 이현실 외(2005)는 FRBR의 Entity-relation과 토픽맵의 Topic-association의 개념적 동일성에 인한 토픽맵을 이용한 FRBR 모델구현 유효성을 주장하였다. 김현희 외(2004)는 금융기관의 효율적인 지식검색과 지식창출을 지식경영의 핵심요인으로 분석하고, 토픽맵 기반의 지식검색체제를 구축하기 위한 리포지토리 모형을 제안하였다. 김정민 외(2005)는 토픽맵 기반의 텍스트 온톨로지 방법

17) <http://www.empolis.com/en/technology>

18) <http://www.thebrain.com/#-43>

19) <http://www.quid.fr/>

20) 차세대 지식정보 공유 국제 컨퍼런스 2006 (Asian Topic Maps Summit) 자료집, 산업자원부 기술표준원

21) <http://plato.snu.ac.kr:8080/philvizmap/html/>

22) <http://eng,i-mfg.com/main/main.asp>

23) http://www.l4d.or.kr/dlsearch/new_ddl/homedata/topicmap.asp

론을 제시하였고, 실제로 철학 온톨로지에 기반한 철학 지식 포탈을 구현하였다.

3) 연구방법 및 한계

본 연구는 온톨로지 기술 중 정보자원을 의미기반의 연관관계를 바탕으로 기존 정보자원을 구조화하여 빠르고 정확한 검색이 가능한 토픽맵 기반의 기록물 검색시스템을 설계하고, 구축하였다. 연구의 진행방법은 다음과 같다.

첫째, 이론적 배경으로서 정보표현과 시멘틱웹, 온톨로지, 토픽맵의 특징과 연원에 대해 고찰한다.

둘째, 웹에서의 기록물의 기술정보 메타데이터를 중심으로 한 토픽맵 기반의 기록정보 서비스를 위한 온톨로지 모델을 설계한다.

셋째, 설계한 온톨로지 모델을 바탕으로 국가기록포탈에서 추출한 정보자원 목록을 에디터를 이용해 토픽맵으로 반입하여 기록정보 검색시스템을 구현한다.

넷째, 구축된 검색시스템의 사용자 인터페이스에서 테스트질의 통해 토픽맵 기반 기록물 검색시스템의 특징을 확인하고 구축의 의미를 평가한다.

본 연구의 한계로는 온톨로지 모델링의 대상범위가 제한적이었으며, 정보자원간의 의미관계를 단순화하여 기록정보의 지식자원화를 위한 의미적 연관관계의 구조화가 부족하다는 점을 들 수 있다.

2. 기록정보 검색에서의 토픽맵 구조와 역할

1) 기록정보 검색과 표현도구

정보는 형태와 상관없이 검색되기 위해서는 표현되어야 한다. 정보표현은, 문헌의 본질이 특화되고 표현될 수 있도록 문헌으로부터 어떤 요소(예, 키워드나 구)를 추출하거나 혹은 문헌에 용어(예, 디스크립터²⁴)나 주제명표목)을 할당하는 것을 포함²⁵하는 개념이다. 즉, 정보검색을 목적으로 정보 대체물이나 표현을 만들기 위한 다양한 방법을 의미하는 것으로, 정보검색의 기본 전제조건이다. 이는 정보자원의 원시정보²⁶는 그 자체를 대상으로 검색되지 않고, 검색되기 이전에 색인, 텍사노미, 메타데이터, 전거통제, 시소러스, 기술 등으로 표현되어야 하기 때문이다. 이러한 정보표현 과정을 통해 구조화된 기록정보자원들은 단순 키워드매칭의 부정확성을 보완해나가고 있다. 사실 기록정보는 업무의 맥락정보가 중요하고, 건(item)에서 시리즈에 이르기까지 여러 계층성을 가지고 있어, 타분야의 표현도구에 비해 기록물군이나 시리즈계층에 따라 이러한 표현도구로 더욱 구조화할 필요가 있다. 따라서, 기록물 정보표현에서의 색인 선정은 기록물의 특성인 계층성을 반영하여 하위 단위의 기록물에서 상위 컬렉션에 이르기까지의 모든 계층에서 추출한 용

24) 디스크립터는 우선어라고도 불리며 색인작성 및 검색 시 사용되는 시소러스에 등록된 용어로 문헌의 주제를 나타내는 용어나 기호이다.

25) Heting Chu, 「디지털 시대의 정보 표현과 검색」, 한국도서관협회, pp17

26) 원시정보(텍스트)는 정보표현의 과정을 거치지 않은 원형의 데이터를 의미하는 것이다. 정보검색은 원시정보를 대상으로 하는 것이 아니라 원시정보를 표현한 대체물(예, 색인, 기술)을 대상으로 한다.

어를 포함해야 한다. 기본적으로 기술정보와 함께 내용정보를 표현하는 핵심키워드를 포함해야 하며, 단지 기록물을 색인화하는 것에만 한정할 게 아니라, 기록물은 물론, 생산자에 대한 기술에서 추출한 용어까지도 포함해야 한다.²⁷⁾ 즉, 대용량이며 다변화, 다기능적인 기록정보의 특성상 전 계층의 기록정보 메타데이터를 색인화하여 검색의 접근점을 확장하는 것이 필요하다. 기록정보서비스에서 기록자원에 포함되어 함께 기술되는 메타데이터는 기록물건명, 기록물철명, 생산기관, 생산년도, 소장위치, 기록물유형, 기록물형태 등의 기술정보이다. 이러한 기술정보 메타데이터는 이용자의 기록물 검색의 중요한 접근점의 역할을 하며, 이용자 인터페이스에서 기록물건의 상세정보로서 매우 중요한 색인이 될 수 있다. 따라서, 본 연구에서는 기록물의 기술정보 메타데이터를 핵심 용어로 추출하여, 토픽클래스²⁸⁾로 변환하여 색인화하고, 메타데이터 상호간의 의미적인 연관 관계를 맺어 토픽맵으로 구축하였다. 결과적으로, 토픽클래스로 매핑된 기술정보 메타데이터들은 색인화되고, 자동 카테고리화되어, 의미적 추론 검색이 가능해진다.

2) 시멘틱웹과 온톨로지의 연원

현재의 웹의 정보자원이 가지는 내용은 컴퓨터는 이해할 수 없고 인간만이 이해할 수 있는 수준이다. 이러한 한계는 인간과

27) Fredric. M. Miller, 한국국가기록연구원 역, 『아카이브와 메뉴스크립트의 정리와 기술』, 진리탐구, pp.166

28) 토픽맵에서의 토픽클래스는 타입을 의미하며, 하나의 클래스 카테고리에 여러 개의 인스턴스가 포함된다. 토픽클래스는 토픽들의 대표값으로 온톨로지 상에 존재하는 검색의 주제어로 사용된다.

컴퓨터가 상호 이해 가능할 수 있도록 정보자원들의 속성을 형식적으로 정의해 줌으로써 극복 할 수 있다. 형식적이라는 의미는 기계가 읽고 이해하여 처리함으로써 정보자원의 내용을 자동적으로 표현할 수 있다는 것이다. 시멘틱웹은 웹상의 어떠한 데이터도 의미적이고 기술적인 정보를 연관시키고, 이를 통해 인간과 기계가 상호 작용할 수 있는 환경을 만드는 것을 가능하게 한다.

의미적 관계를 통한 상호작용 환경이 구축되면, 정보 검색자는 원하는 정보자원을 의미적으로 검색하고, 활용할 수 있으며, 그 검색된 정보자원들은 자체로서의 정보가치 뿐 아니라 지식으로서 확장된 가치도 가지게 된다. 지능적이고 자동화된 시멘틱웹 환경을 구성하기 위한 핵심 기술요소는 온톨로지이다. 시멘틱웹에서 온톨로지는 의미 정보를 나타내기 위한 필요한 개념(concept)과 관계(relationship)을 정의한 정형적 명세(formal specification)이다.²⁹⁾ 여기서 ‘개념’이란 표현하고자 하는 대상을 추상화하고 단순화 시킨 것이며, ‘정형적 명세’란 규정된 용어들과 그들 사이의 관계를 컴퓨터가 이해할 수 있는 방법으로 표현하는 것이다. 이러한 표현을 가능하게 하는 것이 온톨로지이며, 용어의 집합으로서 그 용어들간의 정의뿐 아니라 연관관계와 간단한 추론규칙까지 기술해 주고 있기 때문에 시멘틱웹을 구현하기 위한 기반 기술이 된다. 온톨로지를 구축하는 온톨로지 언어에는 RDF, RDF(S), DAML+OIL, OWL, Topic Map 등이 있다.³⁰⁾ 토픽맵 이외의 온톨로지 언어들은 속성과 값으로 표현되므로,

29) Maedche, A., *Ontology Learning for the Semantic Web*, Kluwer Academic Publisher, 2002, pp.11-27

30) 김정민 외, 「K-BOX:토픽맵 기반의 온톨로지 관리 시스템」, 『정보과학회논문지』, 한국정보과학회, 2004

온톨로지 구축을 위해 변환과정을 거쳐야 하지만, 토픽맵은 정보자원의 위치정보를 이용하여 형태를 변환하지 않고도 정보자원들을 통합할 수 있으며, 서로 다른 토픽맵을 쉽게 병합할 수 있으므로 많은 양의 이종 정보자원들을 통합할 필요가 있을 때 매우 유용하다.³¹⁾

3) 토픽맵의 개념 및 구조

토픽맵은 정보 자원간의 상호 연관성을 도출, 연결 및 조직하여 효과적인 지식 구조를 기술할 수 있도록 하는데 목적이 있으며, 대용량의 구조화되지 않은 정보를 효율적으로 검색하기 위한 해결 방안으로 제안되었다. 토픽맵의 표준으로는 2001년 1월 HyTM으로 표현된 ISO/IEC³²⁾ 13250 Topic Maps가 최초로 제안되었고, 2001년 2월 XML 구문체계를 이용한 XML Topic Maps(XTM) 1.0이 Topicmaps.org에서 제안되었는데, 10개월 후 ISO/IEC 13250 (HyTM)과 XTM 1.0이 통합되어 ISO/IEC 13250:2002에서 HyTM, XTM 공동으로 존재한다.(ISO/IEC 2002) 현재, XML을 주요 토픽맵 구문으로 사용하고 있으며, XTM 구문은 거의 모든 토픽맵 도구가 지원하고 있다. 토픽맵은 대용량의 전자정보원의 지식관리 모델로서, 의미적으로 분류하고 구조화하여 원하는 정보에 접근할 수 있도록 하는 일종의 책의 색인과 같은 역할을 한다. 색인, 용어집, 시소러스 등이 모든 책이나 다른 종류의 자원에 내재적(implicit)으로 존재하는 정보구조를 매핑하는 방법이나, 토픽맵은

31) 함화진, 「 토픽맵 반자동 구축 도구의 설계 및 구현」, 이화여자대학교 과학기술대학원, 2005, p50

32) IEC는 the International Electrotechnical Commission의 약자로, 국제전자기술협의회를 지칭하며, ISO와 공동으로 토픽맵을 개발하였다.

기존 색인, 용어집, 시소러스 등의 제한적인 표현방법을 넘어 내재적 정보구조를 외재적으로 표현하는 온톨로지 개념을 도입한 것으로 정보와 지식의 분산관리를 지원하는 것을 목적으로 한다.³³⁾

토픽맵 모델의 계층은 [그림1]과 같이 정보층(index layer)과 지식층(content layer)으로 구성되어 정보와 지식을 구분, 연계하여 최적화된 방법으로 표현할 수 있으며 토픽, 어소시에이션, 어커런스 등의 기본요소로 구성된다. 지식층(content layer)은 기존의 정보자원 위에 구축하는 지식의 구조로서 특정주제를 나타내는 토픽과 주제들간의 연관관계를 나타내는 어소시에이션으로 구성된다. 정보층(index layer)은 정보자원을 나타내며, 이들 지식층과 정보층은 어커런스를 통해 상호 연결된다.

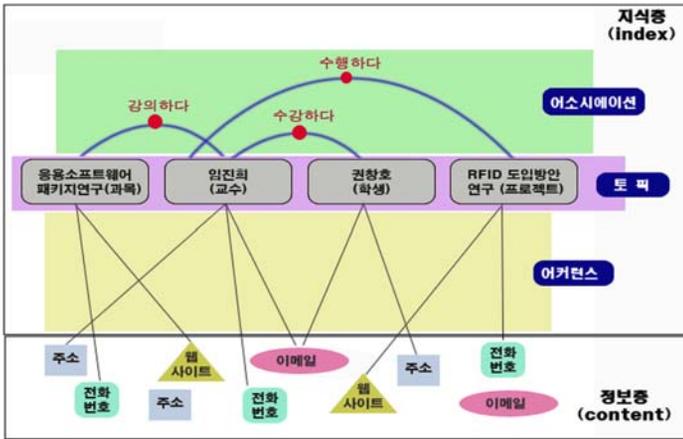


그림 1 토픽맵 모델의 계층과 요소

33) 오삼균, 박옥남, 「토픽맵-기반 관소리 검색시스템 구축 및 평가에 관한 연구」, 『한국도서관·정보학회지』, 한국도서관·정보학회지, pp79

토픽맵 모델의 핵심 요소는 토픽(Topic), 어커런스(Occurrence), 어소시에이션(Association)으로 볼 수 있다. 토픽은 주제 혹은 개념을 기계적으로 처리할 수 있도록 표현한 형태다. 사람, 개체, 개념, 의미 등 특수한 어떠한 것도 토픽이 될 수 있으며, 이것은 기존의 텍사노미³⁴⁾의 주제와 유사한 개념이라고 볼 수 있으며, 색인의 역할을 수행한다. 다양한 토픽들을 유사한 유형을 연결하고, 분류시켜 주기 위해서 토픽클래스를 사용하여, 유사 의미를 갖는 인스턴스(instance)³⁵⁾들을 묶는 역할을 하게 된다. 토픽클래스는 일종의 디렉토리 검색의 색인의 기능을 수행할 수도 있다. 토픽은 특정 도메인(개념이나 객체)의 지식을 나타내는 주제로서, 추상적이며 물리적으로 존재하는 어떤 것도 표현될 수 있다. 또한, 하나의 토픽에 여러개의 이름을 지정할 수 있어 시소러스의 동의어, 유의어, 관계어 설정보다 더욱 정교한 지능형색인의 구축이 가능하고, 이를 바탕으로 의미적 추론이 가능하게 된다. 학교라는 특정도메인의 토픽맵인 [그림1]에서 응용소프트웨어 패키지 연구, 임진희, 권창호, RFID 도입방안 연구는 토픽이며, 각 과목, 교수, 학생, 프로젝트라는 토픽클래스로 그룹핑된다.

어커런스는 특정 토픽에 종속되는 개념으로 해당 토픽에 대한 실제 내용이 담겨있는 정보자원의 주소(URI³⁶⁾로 표현된다)나 데이터 자체를 가리킨다. 토픽과 관련된 모든 데이터는 연결되

34) 텍사노미(Taxonomy)는 정보자원 대상들에 대한 분류의 의미하는 것으로, 관계형 네트워크보다는 트리형의 위계적 구조로 나타난다. 예를 들면, 웹사이트의 카테고리 구조나 사이트맵은 텍사노미(Taxonomy) 데이터라고 할 수 있다.

35) 인스턴스(instance)는 어떤 클래스로부터 만들어진 객체를 의미하며, 클래스로부터 객체를 만드는 과정을 인스턴스(instance)화라고 한다.

36) URI (Uniform Resource Identifier)는 인터넷에 있는 자원을 나타내는 유일한 주소이다. URI의 존재는 인터넷에서 요구되는 기본조건으로서 인터넷 프로토콜에 항상 붙어다니며, 하위개념으로 URL, URN이 있다.

고, 이런 데이터와 토픽은 어커런스를 통해 연결할 수 있다. 어커런스는 문서파일, 이미지 파일, 데이터베이스 내 특정레코드 형태가 된다. [그림 1]과 같이 권창호 토픽은 주소, 이메일 등 여러 형태의 정보자원과 연결될 수 있으며, 응용소프트웨어 패키지 연구 토픽도 웹 사이트, 전화번호 등의 정보자원과 연결될 수 있다. 이렇게 각 토픽에 연결된 전화번호, 이메일과 같은 정보자원들의 동일한 어커런스들은 하나의 어커런스 타입으로 그룹핑될 수 있어 효율적인 정보자원의 관리가 가능하다.

토픽맵의 또 하나의 핵심 요소인 어소시에이션(association)은 토픽과 토픽을 연결시켜 주는 관계를 표현하는 것이다.³⁷⁾ 토픽과 토픽의 관계를 정의하여 명확하게 상호관계를 설정할 수 있게 된다. 어소시에이션은 토픽 상호간의 관계 설정의 제한이 없으므로, 어떠한 토픽 간에도 가능하며, 시소러스의 제한적 연계와는 달리 의미적 관계의 설정도 유연하며 확장이 가능하다. 이는 토픽맵이 관계 설정에 있어, 자연어 형식으로 표현가능하고, 설정의 제한이 없이 자유롭기 때문이다. [그림 1]과 같이 어소시에이션 관계를 통해 응용소프트웨어 패키지 연구 과목, 임진희 교수, 권창호 학생, RFID 도입방안 연구 프로젝트의 각각의 토픽들은 역할을 수행하고 있다. 임진희 교수는 응용소프트웨어 패키지 연구를 강의하고, 권창호 학생은 응용소프트웨어 패키지 연구를 수강하며, 임진희 교수는 RFID 도입방안에 대한 연구 프로젝트를 수행하고 있다. 이와같이 토픽간의 ‘강의하다’, ‘진행하다’, ‘수강하다’와 같은 역할(Role)을 어소시에이션으로 표현하고, 동일한 유형의 어소시에이션들은 동일한 타입으로 그룹핑할 수 있다. 즉, 어소시에이션 타입은 토픽간의 관계 정의이며, 이때 관계가 형성된 토픽 클래스를 어소시에이션 타입의 역할

37) Garxhol, Lars Marius, Living with topic maps and rdf, 2003

이라고 한다. 지금까지 설명한 토픽맵 요소를 대통령기록물이라는 도메인에 적용한 개념도는 [그림2]와 같다.

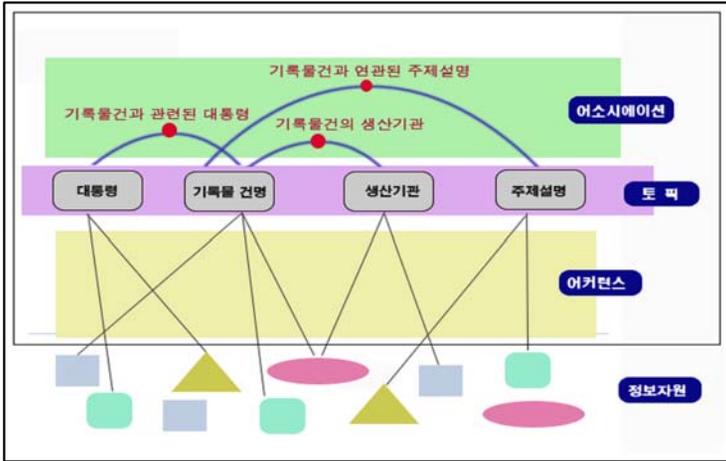


그림 2 대통령 도메인의 토픽맵의 개념도

[그림2]에서 대통령, 기록물건명, 생산기관, 주제설명은 토픽 클래스이며 각각의 토픽 클래스들은 어커런스라는 속성값을 가지며, 상호간에 어소시에이션을 통해 연관관계가 생성되어 있다. [그림2]에서는 대통령은 기록물건명과 ‘기록물건과 관련된 대통령’이라는 어소시에이션 관계이며, 기록물건명은 주제설명과 ‘기록물건과 연관된 주제설명’이라는 어소시에이션 관계를 가지므로, 대통령은 주제설명과 추론 연관되는 구조를 보여주고 있다. 이와 같이 개별 토픽 클래스들은 상호 연관관계를 가지므로, 질의시 검색의 접근점이 확장되며, 의미적으로 구조화되어 네비게이션 검색이 가능하게 된다.

4) 토픽맵의 기록물 검색시스템 구축의 의미

기록물은 업무의 범위가 다양하며, 맥락정보가 중요하며, 계층성을 가지고 있어, 이용자의 검색질의어 선정의 범위가 방대하다. 따라서, 타 분야의 표현도구에 비해 기록물군이나 시리즈 계층에 따라 정보표현을 더욱 구조화하여 이용자의 접근점을 확장할 필요가 있다. 따라서, 기록물의 정보표현은 하위 단위의 기록물에서 상위 컬렉션에 이르는 전 계층에서 추출한 모든 용어와 기술정보, 내용정보, 맥락정보를 포괄하여 특정 주제나 사건의 도메인별로 지식화 할 필요가 있다. 이러한 지식화 작업은 계층에 따라 전 범위에 흩어져 있는 기록정보의 특성상 수작업 시 많은 인력과 시간이 소요되며, 지속적인 유지관리가 쉽지 않다. 따라서, 기록물이 가진 계층성과 맥락성을 포괄하여 정보자원들을 상호 연관성에 따라 유기적으로 연결 및 조직하는 구조화 작업에 토픽맵이 대안이 될 수 있다. 토픽맵의 도입으로 전 계층의 정보자원이 유기적인 검색 접근점이 되어 의미적 추론과 매핑이 가능하므로, 의미적 자원 우선검색, 구문검색, 정보자원 간의 네비게이션 검색이 가능하다. 또한, 제한이 없는 연관관계 설정과 복수토픽 네임칭 처리로 효율적인 지식정보 관리가 가능하며, 분류스키마, DB, 메타데이터를 색인화하여 관련자원간의 그룹핑을 통해 효율적인 지식 자원화도 가능하게 한다. 또한, 구축의 측면에서 토픽맵 시스템은 기존의 단위시스템과 정보자원의 형태와 구조에 영향을 주지 않고 상층부에 통합시스템으로 구축할 수 있어, 단위기관별로 다른 단위시스템과 통합서비스하는 데에 시간과 비용이 많이 들지 않는다는 장점이 있다. 하지만, 지식의 전달과 이해에 탁월한 효용성이 있는 그래픽 네비게이션은 기록정보의 추론된 데이터의 결과값의 수

픽맵을 구축한다. 에디터로 구축된 토픽맵은 토픽맵 프레임워크를 통해 웹 네비게이션 검색시스템으로 구축한다.

1) 온톨로지 모델링

본 장에서는 토픽맵 온톨로지를 모델링하는 개발 방법 및 절차를 알아보고, 실제 모델링 과정을 다룬다. 온톨로지를 모델링하는데 있어 가장 중요한 것은 구축의 목적과 범위를 선정하는 단계이다. 온톨로지 모델 설계에 따른 토픽요소 생성 이후에 목적과 범위의 오류를 발견하여 재작업하는 일이 없도록 신중을 기해야 한다. 토픽맵 온톨로지의 구축과정은 [표1]과 같다.

가. 목적 및 범위 선정	<ul style="list-style-type: none"> - 목적 정의 - 범위 설정 - 요구사항 분석
나. 용어수집 및 추출	<ul style="list-style-type: none"> - 자료수집 및 분석 - 주제어 추출 - 목록 작성(클래스, 인스턴스(instance), 어커런스 항목 분류) - 온톨로지기반 검색을 위한 구성요소 도출
다. 온톨로지 모델 설계	<ul style="list-style-type: none"> - 클래스 정의(토픽클래스 생성) - 클래스별 인스턴스(instance) 정의 (클래스별 자료분류) - 클래스 및 인스턴스(instance) 간 관계정의(어소시에이션 타입 생성) - 클래스별 - 클래스별 어커런스 정의 (어커런스 타입 생성) - PSI³⁸⁾ 정의

표 1 온톨로지 모델링 구축 절차

38) PSI(Public Sector Information)는 주제 식별자로서, 사용하고자 하는 온톨로지(어휘, 주제어, 토픽 등)가 의미하는 내용을 명확히 하기 위하여 필요하다. PSI는 동형이의어(homonym), 이형동이의어(synonymy) 등으로 인한 언어의 모호성을 해소하고, 의미의 호환과 정의를 명확하게 한다

먼저 목적 및 범위선정 단계에서는 목적정의/ 범위설정/ 요구사항 분석을 수행해야 한다. 목적은 명확하게 설정하고, 설계에 필요한 기초자료를 광범위하게 분석하고, 핵심개념을 파악하여 추구하고자 하는 방향을 설정한다. 범위설정은 어떤 시각으로 온톨로지를 작성할지와 얼마나 정밀하게 작성할 것인가를 시간과 비용을 분석하여 결정하는 문제이다. 온톨로지를 구축하는 설계·개발 어플리케이션은 관리자를 대상으로 하는지, 이용자를 대상으로 하는지에 따라 설계의 범위는 달라진다. 범위설정의 결정은 기존 도메인의 구조화 정도가 중요하다. 기존 도메인이 텍사노미나 시소러스가 정교하게 구축되어 있는 경우에는 대상범위를 최소화하여 시간과 비용을 절감할 수 있다.

요구사항분석은 목표와 범위를 결정한 후 토픽맵 온톨로지 설계가 어플리케이션 구축 목적의 요구사항과 이용자의 요구사항을 정확히 반영할 수 있도록 기존의 데이터 소스를 분석하고, 정확히 이해하는 과정이다.

용어수집 및 추출 단계는 데이터 분석을 통해 수집된 정보자원 목록을 정제하여 핵심적인 주제와 개념, 용어들을 도출하는 과정이다. 도출된 용어의 목록에서 토픽클래스, 인스턴스(instance)를 구분하고, 온톨로지 기반 검색을 위한 사용자 검색질의어를 도출한다.

이상의 목적 및 범위선정, 용어수집 및 추출 등의 기초작업을 통해 추출한 목록을 기반으로 실제 온톨로지 모델링을 시작한다. 온톨로지 모델 설계단계는 토픽 클래스, 클래스별 인스턴스(instance), 클래스 및 인스턴스(instance)간의 어소시에이션, 클래스별 어커런스 PSI를 정의하고, 컴퓨터가 처리할 수 있도록 온톨로지 디자인을 구현하는 단계이다. 온톨로지 디자인을 구현하는 방법은 크게 다음 세 가지가 있다. 텍스트편집기, XML/SGML 편

집기를 이용한 XTM³⁹⁾ 코드 문서를 작성하는 방식과 XTM으로 자동변환될 수 있는 코드의 작성 방식, 에디터(editor)나 컨버터(convertor)를 이용한 방식이 있다. 텍스트 편집기나 코드 작성을 이용한 방법은 정확한 코딩작업과 구현기술이 필요하여 시간과 비용이 많이 소요되는 반면, 에디터나 컨버터를 이용한 소프트웨어방식은 시간과 비용이 절감된다.

지금까지 토픽맵 온톨로지의 일반 구축방법을 알아보았고, 이를 바탕으로 기록정보 서비스를 위한 토픽맵 온톨로지를 [표1]의 순서대로 구축하였다. 첫 번째 시스템구축의 목적은 토픽맵 기반의 기록물검색시스템 구축의 타당성을 알아보는 데 있다. 상술하면, 토픽맵의 장점을 확인하고 기록정보의 지식자원화의 가능성을 확인하고, 이를 바탕으로 토픽맵기반 기록물검색시스템 도입에 대한 기초적인 근거자료로 삼고자 하는 데 있다. 목적에 따른 온톨로지 구축의 범위는 이용자를 대상으로 한 기록정보서비스 웹사이트의 기록검색 어플리케이션 개발에 초점을 맞추었고, 정밀도의 정도는 시간과 비용을 최소화 하기위해 이미 활용 가능한 소스가 존재하고, 구조화되어 있는 정보자원을 활용하였다. 이를 위해, 구축범위는 국내의 대표적인 기록정보서비스 웹사이트인 국가기록포털의 기록자원의 종류별 분류에서 역사적·상징적 의미를 지닌 대통령기록물로 선정하였다. 온톨로지 설계의 범위는 국가기록포털에서 대통령기록물과 이미 구조화되어 있는 주제설명과 참고자료를 대상으로 온톨로지를 자동구축하였다. 본 연구의 온톨로지 설계에 있어서의 이용자요구사항과 어플리케이션 개발의 요구사항은 동일하다. 따라

39) XTM(XML Topic Maps)은 ISO/IEC 13250:2002에서 HyTM과 함께 토픽맵 표준으로 지정되었으며, 현재 XTM 구문은 거의 모든 토픽맵 도구가 지원하고 있어, 주요 토픽맵 구문으로 사용되고 있다.

서, 요구사항분석은 현재 국가기록포탈에서 진행하고 있는 이용자 설문 결과로 대체하였다. 이용자 설문 질문인 “현재 국가기록포탈에서 가장 개선이 필요하다고 생각되는 부분은?”에 대한 이용자 답변은 ‘기록물통합/상세검색’이 57.5%로 제일 많았고, 그 다음으로, ‘토픽, 컬렉션 등 콘텐츠의 내용 및 구성’이 17.7%를 차지하였다. 이를 통해 이용자들은 우선 기록물의 검색의 정확성 향상을 요구하고 있으며, 국가기록포탈의 기록정보컨텐츠화 추진에도 불구하고 현재 서비스되고 있는 콘텐츠보다 더욱 정교하게 구조화된 지식컨텐츠를 요구하고 있음을 알 수 있다. 이러한, 이용자들의 요구를 반영하여, 기존 정보자원의 지능형 색인화를 통해 검색의 정확률을 향상시키고, 의미적 연관관계를 활용하여 정보자원의 의미관계 네비게이션이 가능하도록 구현하여 기록정보를 지식자원화하는 온톨로지 모델링 구축에 중점을 두었다.

용어수집 및 추출의 단계는 기존 시스템 및 데이터에서 온톨로지 설계의 대상인 대통령 관련기록물에 관한 주요용어와 개념을 추출하고, 클래스와 인스턴스(instance)로 분류하는 과정이다. 기록정보서비스에서 이용자의 주요 검색 질의의 대상이자 이용자서비스의 기본단위인 기록물건명과 건명의 기술정보인 메타데이터, 주제설명, 관련대통령을 의미적으로 모델링하기 위해, 다음과 같은 목록을 추출하였다.

가. 기술정보 메타데이터	기록물건명, 기록물철명, 생산기관, 생산년도, 소장위치, 주제설명, 기록물유형, 관련대통령, 기록물형태, 공개구분, 대통령, 재임기간
나. 주제설명	주제설명, 근거, 참고자료, 관련주제 내용, 참고자료, 관련주제
다. 관련대통령	대통령, 재임기간, 공화국

표 2 추출된 용어 목록

가. 기술정보 메타데이터는 실제 기록물 검색의 중요한 접근점으로 전자기록 자원에 일부분으로 포함되어 있을 수도 있고, 별개로 독립적인 저장소에 기술되어 있을 수도 있다. 기록정보 서비스에서 기록자원에 포함되어 있으며, 함께 기술되는 메타데이터는 기록물건명, 기록물철명, 생산기관, 생산년도, 소장위치, 기록물유형, 기록물형태 등의 기술정보이다. 이러한 기술정보 메타데이터는 이용자의 기록물 검색의 중요한 접근점의 역할을 하며, 이용자인터페이스 기록물건의 상세정보로서 매우 중요하다. 따라서, 본 연구에서는 기록물의 기술정보 메타데이터를 핵심 용어로 추출하였으며, 기술정보 메타데이터 중심의 기록물 온톨로지를 모델링하기 위해서 기록정보서비스에서 제공하고 있는 기술정보들을 용어 목록으로 추출하였다.

나. 주제설명은 국가기록포탈에서 각 분야별로 전문 연구자들이 참여하여 콘텐츠 주제를 선정하고 주제를 계층 구조화하였으며 각 주제에 대한 배경지식을 집필한⁴⁰⁾ 기술이다. 주제 설명은 기록물건의 배경지식과 맥락을 표현하는 기술이므로, 온

40) <http://contents.archives.go.kr/next/intro/viewIntroduction.do> 국가기록포탈 안내페이지

톨로지 모델링을 통해 기록물건의 기술정보 메타데이터와 의미적 연관관계를 맺기 위해 주요 용어로 추출하였다. 주제설명과 함께 주제설명의 메타데이터인 근거, 내용, 참고자료, 관련주제도 용어 목록으로 추출하였다.

다. 관련 대통령은 국가기록포탈의 대통령 컬렉션에서 역대 대통령을 주요용어로 추출하였고, 온톨로지 모델링 설계 과정에서 대통령과 기록물의 생산년도의 추론을 위해 재임기간과 공화국을 추출하였다. 이는 이용자의 검색 시 대통령의 정확한 재임기간과 공화국에 대한 질의어를 예상하고, 기록정보자원과 함께 연관정보자원을 검색 결과에 포함하여 대통령과 관련된 보다 많은 정보를 전달하고 지식화하기 위해서 도출하였다.

추출된 용어 목록 중 토픽 클래스는 [표3]과 같다. 일반적이고, 범용적인 의미를 가지고, 어소시에이션 관계 설정이 필요한 경우 토픽클래스로 선정한다. 이용자의 기록물검색의 기본단위인 기록물건명과 기록물건명의 메타데이터인 기술정보, 주제설명, 관련대통령은 모두 어소시에이션 관계를 가지며, 범용적인 카테고리로서 이용될 수 있다.

토픽클래스	기록물 건명, 기록물 철명, 생산기관, 생산년도, 소장위치, 기록물유형, 기록물형태, 관련 대통령, 주제설명
-------	--

표 3 토픽 클래스

목적 및 범위선정, 용어수집 및 추출 등의 기초작업을 통해 추출한 목록과 토픽클래스를 기반으로 대통령관련 기록물 도메인의 온톨로지 모델링을 시작한다. 모델링은 상용화된 에디터⁴¹⁾를 이용한 방식으로 구축하였다. 도출된 주요 개념의 목록 중

용어수집 및 추출 단계에서 선정된 [표3]의 토픽클래스 이외의 토픽들은 의미적인 연관관계 설정을 위해 어커런스, 어소시에이션 항목으로 분류하였다. 대통령관련 기록물 도메인의 의미적 구조화를 위해 모든 토픽클래스는 어소시에이션 관계를 갖는다. 어소시에이션 타입은 [그림4]와 같이 기록물건명과 대통령 토픽의 어소시에이션은 ‘기록물건과 관련된 대통령’ 어소시에이션으로 설정하였으며, 기록물건명과 주제설명의 어소시에이션은 ‘기록물건과 연관된 주제설명’으로 설정하였다.

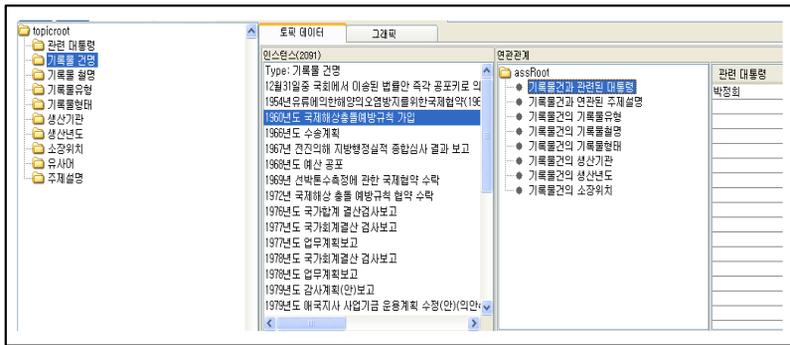


그림 4 기록물건과 관련대통령의 연관관계 예

[그림5]는 [그림4]와 같이 기록물건명과 관련된 토픽클래스의 연관관계를 그래픽 네비게이션으로 표현한 것이다. 토픽맵 구축은 [그림5]와 같이 토픽간의 연관관계를 그래픽으로 보여줄 경우에 범위가 작은 도메인의 의미적 연관관계를 시각적으로 보여주어 효율적으로 지식구조의 전달과 이해가 가능하다는 장점이 있다.

41) 국내 최초로 상용화된 토픽맵 온톨로지 모델링 에디터(editor)인 PRO-TM EDITOR를 사용하였다.

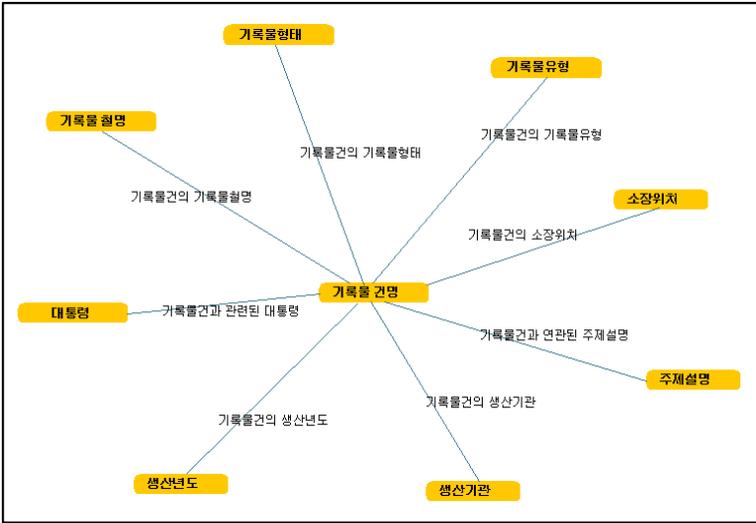


그림 5 기록물건명의 연관관계 그래픽 네비게이션

이용자 질의와 정보검색 서비스의 기본 단위인 기록물건은 공개여부와 정보공개신청을 어커런스로 갖는다. 원하는 기록물건을 검색한 이용자는 기록물공개여부를 확인하고, 공개가능한 기록물건을 웹상으로 정보공개신청을 하고, 이메일이나 우편으로 기록물을 전송받게 된다. 대통령 토픽 클래스는 모든 토픽 클래스와 어소시에이션 관계를 통해 연관관계를 가지며, 재임 기간과 공화국을 어커런스로 갖는다. 최종적으로, 기록물건과 대통령의 배경지식과 맥락정보를 제공하기 위한 주제설명 토픽 클래스는 관련주제를 어소시에이션 관계로 설정하고, 내용, 근거, 참고자료를 어커런스로 선정하였다.

2) 토픽맵 구축

토픽맵 온톨로지 디자인을 완성한 후, 이를 토대로 온톨로지의 콘텐츠를 구성하게 될 인스턴스(instance) 데이터들을 반입하여 토픽맵을 구축한다. 인스턴스(instance) 데이터를 생성하는 방법은 수작업으로 입력하는 방식과 기존 데이터 소스로 자동 생성하는 방식이 있다. 수동생성은 데이터의 양이 적거나, 자동변환할 수 있는 데이터 소스가 없을 시에 이용된다. 하지만, 일정 규모 이상의 데이터를 수동으로 구축할 경우, 막대한 시간과 인건비가 소요된다. 따라서, 인스턴스(instance)를 자동으로 생성시키는 것이 효율적이며, 자동생성은 데이터의 양, 작성될 토픽맵 온톨로지의 복잡성, 구조화의 정도, 데이터 유형 등을 고려하여 구축하여야 한다.

토픽맵 자동구축은 온톨로지 과정에서 기존 시스템과 자료를 분석하여 도출된 정보자원들을 기초자료로 사용한다. 기존 시스템의 주제분류, 텍사노미, 시소러스, 메타데이터 등의 표현정보와 분류스키마, 데이터베이스 구조 등을 이용할 수 있다. 표현정보와 분류스키마 요소는 토픽맵의 요소들로 쉽게 매핑할 수 있다. 분류스키마의 계층성의 관계는 토픽맵의 어소시에이션 관계로 매핑할 수 있으며, 각 클래스는 토픽으로 매핑되며, 분류된 정보 자원들은 토픽타입의 어커런스로 매핑할 수 있다. 가장 많이 이용되는 DB구조인 관계형 데이터베이스는 테이블을 토픽클래스로, 컬럼을 토픽 인스턴스로, 열을 이름 또는 어커런스로 매핑할 수 있다. 정보자원의 일부분인 메타데이터들은 토픽맵 구축 시 인스턴스로 생성되며, 메타데이터명(기록물의 경우 기록물건명)의 경우는 토픽의 이름으로 변환될 수 있다.

온톨로지 개념 모델링 설계에 따라 에디터로 토픽 클래스, 어

커런스 타입, 어소시에이션 타입 요소를 생성하고 각 요소에 인스턴스(instance) 데이터를 반입한다. 생성한 요소에 반입할 기록 정보자원은 국가기록포탈에서 MS-OFFICE 파일로 추출하였으며, 반입할 데이터 목록의 일부는 [표4]와 같다.

가. 기술 정보 메타 데이터 목록 예	A	B	C	D	E	F	G	H	I	J	
	기록물 건명	기록물 설명	생성기관	생성년도	소장위치	주제설명	기록물유형	관련대통령기록물형태	공개구분		
나. 주제 설명 및 관련대통령	2	중앙학도 훈장단	중앙학도 훈장단	대한통번사	1976	대통령기록	중앙학도 훈	대한통번사 기록물	방정희	문서	공개가능
	3	15.15학생자살	15.15학생자살	대한통번사	1980	대통령기록	15.15학생자살	대한통번사 기록물	이승만	문서	공개가능
	4	15.15학생자살	15.15학생자살	대한통번사	1980	대통령기록	15.15학생자살	대한통번사 기록물	이승만	문서	공개가능
	5	15.15학생자살	15.15학생자살	대한통번사	1980	대통령기록	15.15학생자살	대한통번사 기록물	이승만	문서	공개가능
	6	3.15학생자살	3.15학생자살	대한통번사	1981	대통령기록	3.15학생자살	대한통번사 기록물	이승만	문서	공개가능
	7	3.15학생자살	3.15학생자살	대한통번사	1980	대통령기록	3.15학생자살	대한통번사 기록물	이승만	문서	공개가능
	8	3.15학생자살	3.15학생자살	대한통번사	1980	대통령기록	3.15학생자살	대한통번사 기록물	이승만	문서	공개가능
	9	특별감찰부	특별감찰부	대한통번사	1981	대통령기록	특별감찰부	대한통번사 기록물	이승만	문서	공개가능
	10	15.10여기서말	15.10여기서말	대한통번사	1949	대통령기록	15.10여기서말	대한통번사 기록물	이승만	문서	공개가능
	11	15.10여기서말	15.10여기서말	대한통번사	1949	대통령기록	15.10여기서말	대한통번사 기록물	이승만	문서	공개가능
	12	15.10여기서말	15.10여기서말	대한통번사	1949	대통령기록	15.10여기서말	대한통번사 기록물	이승만	문서	공개가능
	13	15.10여기서말	15.10여기서말	대한통번사	1949	대통령기록	15.10여기서말	대한통번사 기록물	이승만	문서	공개가능
	14	15.16이전	15.16이전	대한통번사	1982	대통령기록	15.16이전	대한통번사 기록물	이승만	문서	공개가능
	15	15.16이전	15.16이전	대한통번사	1982	대통령기록	15.16이전	대한통번사 기록물	이승만	문서	공개가능
	16	15.16이전	15.16이전	대한통번사	1982	대통령기록	15.16이전	대한통번사 기록물	이승만	문서	공개가능
	17	15.16이전	15.16이전	대한통번사	1981	대통령기록	15.16이전	대한통번사 기록물	이승만	문서	공개가능
	18	말기(1981.05.14말기)	말기(1981.05.14말기)	대한통번사	1981	대통령기록	말기(1981.05.14말기)	대한통번사 기록물	이승만	문서	공개가능

주제 설명	근거	내용	참고자료	관련주제	대통령	제정기간	공약구
중앙학도 훈장단	대한민국	학원의 사정권 인정과 반	공공부 《문공401사》 1988		이승만	1948~1992	1.공약구
15.15학생자살	대한민국	1980년 3월 15일 일어난 학생자살	에스케이뉴스 2012년 11월 15일		이승만	1948~1992	2.공약구
15.10여기서말	대한민국	1981년 2월 16일 새벽, 특별감찰부(법무부사도 16)의 관리관	방정희 1984-1992		이승만	1948~1992	3.공약구
특별감찰부	대한민국	1981년 2월 16일과 17일	대한통번사 1981년 2월 16일		이승만	1948~1992	4.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	5.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	6.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	7.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	8.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	9.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	10.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	11.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	12.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	13.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	14.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	15.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	16.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	17.공약구
3.15학생자살	대한민국	1981년 3월 15일 일어난	대한통번사 1981년 3월 15일		이승만	1948~1992	18.공약구

표 4 추출된 용어 목록

추출된 정보자원을 이용해 토픽맵을 구축하는 방법들은 시간 및 비용과 의미론적 구조화 사이에서 트레이드 오프(trade-off)⁴²⁾를 갖는다. 수동으로 구축하는 방법은 토픽맵을 의미론적으로 잘 구조화 할 수 있어, 가장 정확한 검색결과값을 도출할 수 있지만, 시간과 비용이 많이 소요되며, 자동 구축 방법은 시간과 비용이 현저히 절약되는 데에 반해 토픽맵의 의미적 구조화가 어렵다.⁴³⁾

42) 트레이드 오프(Trade Off)는, 두 목표가 양립할 수 없는 경우로, 어느 한 쪽을 위해 다른 한 쪽을 희생해야 한다.
 43) 함화진, 「토픽맵 반자동 구축도구의 설계 및 구현」, 이화여자대학교 과학

따라서, 이론적으로 반자동구축 방법이 시간·비용과 의미론적 구조화를 잘 절충할 수 있지만, 본 연구는 기록물 대상범위를 국가기록포탈의 대통령관련 기록으로 한정하였고, 국가기록포탈의 대통령관련 기록은 대통령컬렉션, 관련주제 설명, 기술정보로 구조화되어 있으므로 자동구축 방법으로 구축하였다. [표4] 같이 국가기록포탈에서 MS-OFFICE 파일로 추출된 목록을 토픽맵 에디터로 일괄 반입하여 인스턴스(instance) 데이터를 생성하는 자동구축 방식으로 진행하였다. 에디터상에서 인스턴스(instance) 데이터를 반입하여 생성된 토픽클래스와 인스턴스(instance)의 관계 중 대통령클래스와 인스턴스의 그래픽 네비게이션은 [그림6]과 같다.

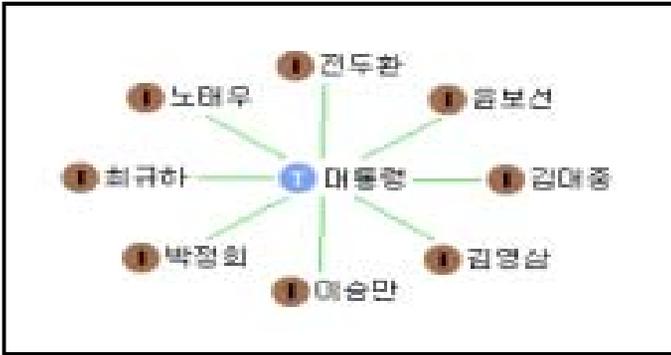


그림 6 토픽클래스와 인스턴스(instance) 관계 그래픽네비게이션

전체 토픽클래스, 어소시에이션 타입, 어커런스 타입에 인스턴스(instance)를 생성하고 데이터를 반입한 에디터상의 토픽맵 구조는 [그림7]과 같다.

기술대학원, 2005



그림 7 에디터의 토픽맵 구조 예시

[그림7]과 같이 토픽맵 에디터의 구조는 토픽클래스, 토픽인스턴스(instance), 어소시에이션, 관련기록물건, 어커런스, 콘텐츠로 구성되어 있다. [그림7]을 보면 주제설명 토픽클래스의 수많은 인스턴스(instance) 중 “3.15 부정선거”는 ‘부정선거 관련자 처벌법’ 등의 기록물건과 ‘기록물건과 연관된 주제설명’이라는 어소시에이션 관계를 가지고 있음을 알 수 있다. 또한, “3.15 부정선거”는 ‘근거’, ‘내용’, ‘참조자료’ 어커런스 타입 중 <제2공화국 의원내각제하의 정부위기를 중심으로>, <법무부사> 등의 Contents(정보자원)와 ‘참조자료’ 어커런스로 연결되어 있음을 알 수 있다. [그림7]과 같이 에디터의 각 토픽클래스는 각각의 인스턴스 값을 가지고 있으며, 기록물건명 토픽클래스(타입)의 인스턴스(instance)로 생성된 파일의 일부는 [표5]와 같다.

가. 기록물건명 토픽클래스의 인스턴스로 생성된 파일 예	<ul style="list-style-type: none"> • 국가보안법중개정법률안(시안)설명서 • 신국가보안법개정에관한건의 • 국가 보안법 • 국제형사경찰기구가입 • 국제형사경찰기구대한민국국가사무국운영요강중개정 예규(내무부예규제358호) • 반공법 • 반공법 시행령 공포의 건 • 반공법 시행령 중 개정의 건 공포의 건
나. 주제설명 토픽클래스의 인스턴스로 생성된 파일 예	<ul style="list-style-type: none"> • 중앙학도호국단 • 3·15 부정선거 • 5.10선거 • 거창양민학살사건 • 건국준비위원회 • 국가보안법위반사건 • 국민방위군사건 • 명동성당 사건 • 남북적십자회담 • 국가보안법제정

표 5 각 토픽클래스의 인스턴스(instance)로 생성된 파일 예시

위 예시와 같이 각 요소들의 인스턴스(instance)를 생성하여 구축된 토픽맵의 각 요소와 인스턴스들은 에디터 상에서 추가, 수정, 삭제가 가능하여 지속적인 데이터 관리가 가능하다. 또한, 에디터 인터페이스에서도 토픽맵의 구조를 그래픽 네비게이션으로 변환할 수 있으며, 그래픽은 사용자인터페이스에 에디터와 동일하게 표현된다.

3) 사용자 인터페이스

에디터로 구축된 토픽맵은 토픽맵 프레임워크⁴⁴⁾를 통해 웹 네비게이션 검색시스템으로 구축된다. 기록물은 웹으로 직접 서비스되는 경우 보다 웹에서 이용자가 검색한 결과값의 기술 정보를 확인 후 공개기록물에 한하여, 정보공개청구를 통해 기록물건의 내용정보를 이메일이나 우편, 팩스를 통해 서비스받는 경우가 일반적이다. 따라서, 기록물 토픽맵의 사용자인터페이스는 여타의 검색시스템과는 달리 내용정보를 서비스해야 할 인터페이스에서 기록물 건의 공개여부와 정보공개신청하기로 구현하였다. [그림8]은 토픽맵 프레임워크를 통해 구축한 사용자 인터페이스에서 기록물을 검색한 상세결과 페이지이다.

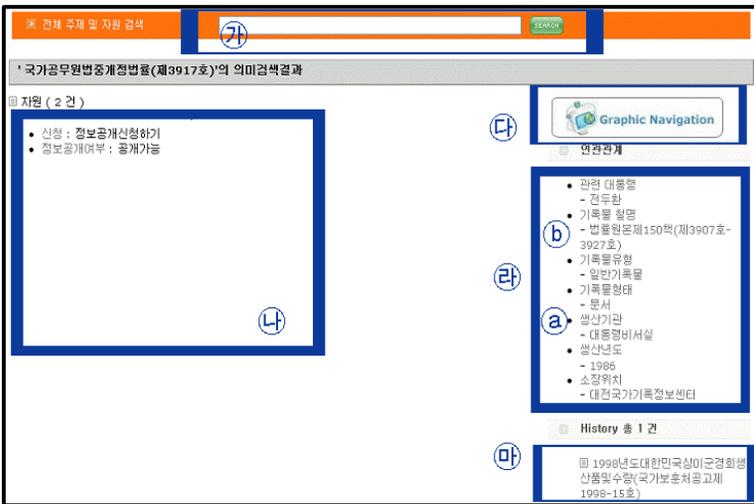


그림 8 이용자인터페이스의 검색결과 상세페이지

44) 본 연구에서는 에디터로 작성된 토픽맵 파일을 웹으로 검색, 네비게이션 할 수 있도록 하는 솔루션 중 Pro-TM web Framework를 사용하였다.

[그림 8]의 ㉑는 이용자인터페이스의 검색창이며, ㉒는 그래픽 네비게이션 링크 버튼으로 클릭시 의미관계를 그래픽 네비게이션으로 보여준다. ㉓는 검색한 기록물건의 어커런스 부분으로, 정보공개여부와 정보공개신청하기로 구성된다. 만일, 토픽맵 구축 시 기록물건이 원문텍스트나 URL을 메타데이터로 포함하고 있다면, ㉔의 어커런스 부분에 추가할 수 있다. ㉕는 기록물건과 연관된 토픽 (원정보자원의 기술정보메타데이터)으로 구성되어 있다. 각 토픽클래스와 토픽 인스턴스들은 디렉토리화 되어 있어 해당 토픽클래스와 인스턴스들은 클릭 시 연관정보를 보여준다. 즉, ㉖생산기관이란 토픽클래스를 클릭하면 토픽맵 상의 모든 생산기관명을 보여준다. ㉗“법률원본 제150책”이라는 기록물칭명 인스턴스를 클릭하면 기록칭명과 연관된 모든 기록물건명 보여준다. ㉘는 이용자가 검색한 질의어를 순서대로 보여주는 History 부분이다.

4) 시스템 평가

본 단락에서는 2.3장에서 언급한 온톨로지를 기반으로 한 의미적 연관관계의 구조화와 의미추론적 측면에서의 토픽맵의 특징인 의미적 자원 우선검색, 자료간의 네비게이션 검색, 구문검색, 통합지식검색의 특징을 구축한 기록물검색시스템을 질의 테스트하여 추출된 결과값을 통해 평가해 본다. 실제 기록물검색시스템의 평가를 위해 “군사혁명위원회와 관련된 기록”이란 구문으로 질의해 보았다. 군사혁명위원회로 질의한 검색결과 페이지의 구성은 [그림9]와 같다.



그림 9 군사혁명위원회의 검색결과 페이지

①은 군사혁명위원회에 관한 내용과 참고자료 보여주고 있으며, 군사혁명위원회 주제설명의 어커런스이다. ②는 군사혁명위원회와 관련된 주제설명(토픽클래스) “군사혁명위원회”, “제5차 개헌” (토픽 인스턴스)을 보여주고 있으며, 어소시에이션으로 연관된 자원이다. ③은 군사혁명위원회와 관련된 관련주제(토픽클래스)인 제“5차 개헌” 토픽 인스턴스를 보여주고 있다. ④는 군사혁명위원회와 관련된 기록물건명 8건 중 “군사혁명위원회 포고 제4호 폐지에 관한 법률(제1244호)”이다. [그림 10]은 [그림 9]의 ④“군사혁명위원회 포고 제4호 폐지에 관한 법률(제1244호)”을 클릭한 결과값이다.



그림 10 ④“군사혁명위원회 포고 제4호 폐지에 관한 법률(제1244호)”을 클릭한 결과값

[그림10]은 “군사혁명위원회 포고 제4호 폐지에 관한 법률(제1244호)”에 관한 8건의 연관관계 토픽과 2건의 어커티스(자원)로 구성되어 있다. ㉠은 “군사혁명위원회 포고 제4호 폐지에 관한 법률(제1244호)”기록물건의 정보공개여부와 정보공개신청의 어커티스에 해당한다. ㉡은 기록물건과 연관된 토픽 (원정보자원의 기술정보 메타데이터)으로 구성되어 있다. 각 토픽클래스와 토픽 인스턴스들은 디렉토리화 되어 있어 해당 토픽클래스와 인스턴스들은 클릭 시 연관정보를 보여준다. ㉢은 해당 기록물건을 생산한 생산기관 토픽클래스와 생산기관 인스턴스에 해당한다. [그림11]은 [그림10]의 ㉢생산기관 토픽클래스의 인스턴스인 감사원사무처 심의실을 클릭한 결과값이다.



그림 11 ㉢감사원사무처 심의실을 클릭한 결과값

[그림11]의 ㉣는 감사원사무처 심의실에서 생산한 기록물 건명으로, 동일 기관에서 생산한 기록물건이 그룹핑되어 ‘감사원사무처 심의실’이라는 기술정보 메타데이터가 토픽으로서 새로운 검색의 접근점 역할을 하며, 이용자 인터페이스에서 대통령 관련 기록물 도메인의 중요한 색인이 될 수 있음을 보여준다.

지금까지 이용자인터페이스에서 ‘군사혁명위원회와 관련된 기록물 건명’으로 질의한 검색결과에서 연관관계를 통해 ‘감사원사무처 심의실’에서 생산한 기록에 이르기까지의 네비게이션 검

색을 실행해 보았다. 이러한 네비게이션 검색은 토픽 상호간의 제한이 없는 연관관계를 기반으로 하여 의미적으로 추론하는 토픽맵의 특징으로 테스트질의를 통해 확인할 수 있었다. 본 연구에서는 토픽맵의 네비게이션 검색의 특징을 기록물 검색시스템에 효과적으로 반영하기 위해 기록물 검색질의에 중요한 접근점이 되는 기술정보 메타데이터와 관련 정보자원을 핵심용어로 추출하여 토픽맵으로 모델링한 시스템을 구축하였다. 그 결과, 토픽클래스로 매핑된 기술정보 메타데이터들은 색인화되고, 자동카테고리화되어 추론검색이 가능하였다. 이를 기존의 키워드매칭 기반의 기록물검색 비교하여 도식화하면 [그림12]와 같다.

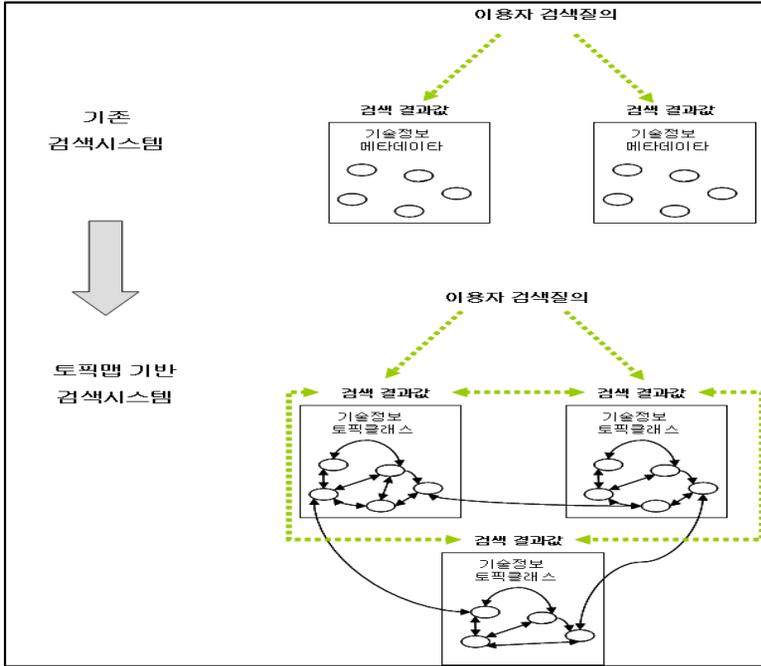


그림 12 기존 키워드 검색시스템과 구현된 토픽맵기반 검색시스템의 비교 개념도

[그림12]에서 기존의 키워드매칭 검색시스템에서는 이용자가 작성한 질의어에 따른 검색 결과값의 기술정보 메타데이터 요소들은 새로운 검색결과값을 보여줄 수 있는 색인으로서 접근점의 역할을 할 수 없다. 따라서, 이용자 질의시에 추출된 검색 결과값 상세페이지에서 또 다른 검색 결과값을 얻기 위해서는 재질의의 절차를 거쳐야 한다. 하지만, 토픽맵 기반으로 구축된 검색시스템은 검색 결과값의 기술정보 메타데이터들은 정보탐색자의 최종적인 검색 결과값도 되며, 새로운 검색을 위한 접근점이 될 수도 있다. 즉, 검색결과 페이지의 모든 자원들이 새로운 검색을 위한 색인의 역할을 수행하여 이용자가 결과값에 만족하지 못해 검색창에서 재질의를 반복하는 번거로움을 해소해 줄 수 있다.

본 연구에서 구축한 토픽맵 기반의 기록물검색시스템은 기존에 구조화되어 있는 정보자원과 기술정보 메타데이터를 주요 토픽맵 요소로 추출하여 토픽맵 에디터로 자동 구축하여, 기록물의 특징을 반영한 의미있는 결과값을 추출할 수 있었다. 이를 통해 구조화되어 있는 시스템을 이용할 경우, 많은 시간과 비용을 투자하지 않고, 기존 시스템의 정보자원의 형태와 구조를 그대로 변환하여 의미기반의 토픽맵 구축이 가능하다는 점을 확인할 수 있었다.

4. 결론

본 연구에서는 웹사이트를 통한 기록물서비스와 검색의 주요 대상이 되는 기록물건의 기술정보 메타데이터를 중심으로 토픽맵 기반의 기록물검색시스템을 설계, 구축하고, 그 의미를 평가

하였다. 검색시스템의 이용 대상은 생산기관 내부의 이용자가 아닌 웹사이트를 이용한 불특정 다수의 이용자를 대상으로 하였으며, 시스템구축을 통한 세부 목적은 기록물건의 기술정보 메타데이터의 의미적 연관관계를 표현하여 이용자질의와의 매칭의 정확도를 향상시키고, 검색된 기록자원의 관련데이터와 기술정보 메타데이터와의 연관관계를 바탕으로 한 검색 접근점 확대와 추론에 의한 연관 네비게이션 검색을 통해 대통령 관련 도메인 기록의 지식자원화의 가능성을 확인하였다.

시스템구축을 위한 온톨로지 설계과정에서는 기존의 국가기록포탈의 대통령관련 기록정보자원에서 주요 용어목록을 추출하고 토픽, 어커런스, 어소시에이션 요소를 생성하였다. 이러한, 온톨로지 모델링 설계과정에서 기록물건의 모든 기술정보메타데이터와 관련 데이터들을 토픽 클래스로 설정하고 어소시에이션 관계를 설정하여 네비게이션 검색을 유도하였다. 온톨로지 모델링 설계 후 에디터를 이용해 각각의 토픽, 어커런스, 어소시에이션 타입 요소에 인스턴스(instance) 데이터를 에디터로 자동반입하여 토픽맵을 구축하였다. 온톨로지 설계 이후에 토픽맵 데이터 반입은 자동구축 방식을 이용하여 추가적인 의미적 연관관계 설정은 하지 않았으나, 기존 콘텐츠를 최대한 구조화하여 모델링하여 시간과 비용을 절감할 수 있었다. 구축한 토픽맵은 프레임워크 솔루션을 이용해 웹 검색시스템과 사용자인터페이스를 구축하였다. 구축한 토픽맵을 테스트 질의하여 검색 결과값을 바탕으로 네비게이션 검색을 확인할 수 있었고, 지식자원으로의 확장 가능성도 확인하였다. 즉, 대통령 기록이라는 도메인의 모든 토픽클래스(기록물 건명, 기록물 철명, 생산기관, 생산년도, 소장위치, 기록물유형, 기록물형태, 관련대통령, 주제설명)와 관련 정보자원이 제한이 없는 연관관계 설정으로 색인

의 역할을 수행하므로 효율적인 도메인의 지식자원화에 대한 가능성도 확인하였다.

이상의 결과는 웹을 통한 기록정보서비스에 이용자의 검색 접근점을 확장하고, 지식자원화된 검색결과를 제공하기 위해 토픽맵의 적용을 고려할 수 있는 기초자료로서의 의의와 그 타당성에 대한 기초적인 근거자료가 될 수 있을 것이다. 이번 연구를 보완하여 향후에 특정 기록물 도메인 전체를 의미적으로 정교하게 구조화한 온톨로지 모델링 설계과정에 대한 연구와 이를 바탕으로 한 토픽맵 구축으로 기록물의 지식자원화를 기대해본다.

참 고 문 헌

【논문】

- 강소연(2002), 『기록물 검색도구 전산화를 위한 EAD DTD에 관한 연구』, 명지대학교 기록과학대학원 기록관리학과
- 강혜선(2005), 『중앙행정기관 공통업무 시소러스 구축에 관한 연구』, 이화여자대학교 정책과학대학원
- 김정민 외(2005), 『텍스트 내용 지식 기반의 철학 온톨로지 구축』, 정보과학회
- 김정민 외(2004), 『K-BOX:토픽맵 기반의 온톨로지 관리 시스템』, 한국정보과학회
- 김진욱(1998), 『정보검색 기법 연구』, 한국건설기술연구원 학술정보
- 김현희(2006), 『금융기관의 지식 관리 개선 방안 연구』, 한국문헌정보학회, 한국문헌정보학회지 제40권 2호, pp.103-123
- 고유미 외(2005), 『의미네비게이션을 지원하는 온톨로지 기반 한의학 논문 검색시스템 설계 연구』, 한국한의학연구원 논문집, 한국한의학연구원
- 고영만, 시소러스기반 온톨로지에 관한 연구, 정보관리학회지, 5권, 2006
- 남영준(2005), 『토픽맵을 이용한 시소러스의 구조화 연구』, 한국정보관리학회, 정보관리학회지 제22권 제3호, pp.37-53
- 설문원, 천주권(2005) 전자기록철의 구조와 관리방안 - 영국 ERMS 표준을 중심으로, 한국기록관리학회지, 22(2):229-251
- 신인철 외(2006), 『국가지식정보 온톨로지 표준개발』, 온톨로지연구소

- 오삼균 외(2005), 『토픽맵-기반 판소리 검색시스템 구축 및 평가에 관한 연구』, 한국도서관 정보학회지
- 임수연 외, 『온톨로지내의 계층관계를 이용한 문서검색』, 한국정보과학회 학술발표논문집, 한국정보과학회
- 이정희 외(2007), 『대학 전자기록물을 위한 온톨로지 기반 검색시스템 설계 및 구현』, 한국정보관리학회, 정보관리학회지, pp.343~362
- 이정희 외(2007), 『대학 전자기록물을 위한 온톨로지 기반 검색시스템 설계 및 구현』, 정보관리학회지, 제24권
- 이현실 외(2005), 『Topic Maps를 이용한 MARC데이터의 FRBR모델 구현에 관한 연구』, 한국정보관리학회, 정보관리학회지 제22권 3호, p.289-306
- 전현주 외(2005), 『시멘틱웹 기반 와인 지식 검색을 위한 웹 서비스 설계』, 이화여대 정보과학 대학원
- 정호영, 김정민, 정준원, 김형주(2003), 『XTM 기반의 지식맵』, 데이터베이스연구회지
- 한성국 외, 시소러스를 활용한 온톨로지 구축방안 연구, 한국비블리아학회지 17권 제1호, 2006
- 함화진(2005), 『토픽맵 반자동 구축도구의 설계 및 구현』, 이화여자대학교 과학기술대학원
- Maedche, A(2002), 『Ontology Learning for the Semantic Web』, Kluwer Academic Publisher
- Steve Pepper, 『The tao of topic maps』
- Steve Pepper(2004), 『Topic Maps : The Tutorial』 2000-2004 Ontopia AS.
- ISO/ICE 2002. ISO/IEC 13250: Topic Map. ISO/IEC.
- garxhol, Lars Marius, Livings with Topic Maps and rdf, 2003
- Cutter, Chjarles, Ammi, 『Rules for a printed dictionary catalogue』,

【웹사이트】

국가기록포탈 <http://contents.archives.go.kr/next/content/viewMain.do>
<http://www.w3.org/>
<http://www.techquila.com/tmtab.html>
<http://www.topicmaps.org/xtm/1.0/>
ADNOM프로젝트 <http://www.cenorm.be/cenorm/businessdomains/iss-activity/adnom.asp>
노르웨이의 과학기술정보 포탈 <http://www.forskning.no/>
소비자정보 포탈 <http://www.forbrukerportalen.no/>
TU교육프로젝트용 웹사이트 <http://www.itu.no>
국세청세금정보포탈 <http://www.missouribusiness.net/irs/taxmap/about.htm>
empolis의 토픽맵 기술 <http://www.empolis.com/en/technology>
더브레인 KMS <http://www.thebrain.com/#-43>
프랑스 백과사전 <http://www.quid.fr/>
한국생산기술연구원의 토픽맵 기반의 사출금형 기술정보 통합
및 정보검색시스템 <http://plato.snu.ac.kr:8080/philvizmap/html/>
동대문구정보화도서관의 지식인마을 시스템
http://www.l4d.or.kr/dlsearch/new_ddl/homedata/topicmap.asp
서울대학교 철학사상연구소의 철학토픽맵
<http://plato.snu.ac.kr:8080/philvizmap/html/>

【단행본】

Fredric. M. Miller(2002), 아카이브와 메뉴스크립트의 정리와 기술,
서울: 진리탐구
Heting Chu(2007), 디지털시대의 정보 표현과 검색, 서울: 한국도

서관협회

노영희(2005), 『개념 기반 정보검색 기법』, 경기: 한국학술정보

【기타자료】

산업자원부 기술표준원, 『차세대 지식정보 공유 국제 컨퍼런스
2006 (Asian Topic Maps Summit) 자료집, 2006

ABSTRACT

Construction of Record Retrieval System based on Topic Map

Kwon, Chang-Ho

Recently, distribution of record via web and coefficient of utilization are increase. so, Archival information service using website becomes essential part of record center. The main point of archival information service by website is making record information retrieval easy. It has need of matching user's request and representation of record resources correctly to making archival information retrieval easy. Archivist and record manager have used various information representation tools from taxonomy to recent thesaurus, still, the accuracy of information retrieval has not solved. This study constructed record retrieval system based on Topic Map by modeling record resources which focusing on description metadata of the records to improve this problem.

The target user of the system is general web users and its range is limited to the president related sources in the National Archives Portal Service. The procedure is as follows; 1) Design an ontology model for archival information service based on topic map which focusing on description metadata of the records. 2) Build practical record retrieval system with topic map that received information source list, which extracted from the National Archives Portal Service, by editor. 3) Check and assess features of record retrieval system based on topic map through user interface.

Through the practice, relevance navigation to other record sources by semantic inference of description metadata is confirmed. And also, records could be built up as knowledge with result of scattered archival sources.

Key words: Topic Map, Record Information Service, Record Retrieval System, Ontology, Semantic Web