

토픽맵 모델링을 위한 한국 관련 미국기록물의 주제분석 연구: 국내 주요 소장기관 중심으로

Topic Analysis of the United States' Historical Records about
Korea for Modeling of Topic Map: Focused on Major Archives
in Korea

권민정(Min Jeong Kwon)¹, 최상희(Sanghee Choi)²

E-mail: alsjwd8219@naver.com, shchoi@cu.ac.kr



¹ 제 1저자 대구가톨릭대학교 기록대학원
² 교신저자 대구가톨릭대학교 도서관학과 교수

논문접수 2022.04.26
최초심사 2022.05.02
게재확정 2022.05.23

ORCID

Min Jeong Kwon
https://orcid.org/0000-0002-3065-7965
Sanghee Choi
https://orcid.org/0000-0002-5864-8139

© 한국기록관리학회

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0/) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

본 논문은 권민정의 석사학위논문 「토픽맵 기반 한국 관련 미국기록물의 검색 개선방안에 관한 연구: 국내 주요 소장기관 중심으로」(2021)을 요약·수정함 것임.

초 록

이 연구에서는 웹사이트를 통해 한국 관련 해외기록물을 서비스하고 있는 기관인 국가기록원, 국립중앙도서관, 국사편찬위원회, 국방부 국방편찬연구소 4개의 기관을 연구 대상으로 선정하여 국내 기관에 분산되어 있는 서비스 되고 있는 기록물을 이용자들이 쉽게 기록물의 내용을 파악할 수 있도록 하고 통합하여 검색할 수 있는 방안을 제안하였다. 이를 위해 총 163,874건을 수집하여 이용자가 다각적으로 기록의 내용을 훑어볼 수 있도록 하는 방안으로 주제분석을 하였고 패싯과 토픽을 도출하였다. 도출된 패싯과 토픽을 기반으로 토픽맵을 구성하였다.

ABSTRACT

This study selected four institutions, namely the National Archives of Korea, the National Library of Korea, the National Institute of Korean History, and the Institute for Military History, that have collected the foreign records about Korea, providing web services for users. A total of 163,874 records were collected from these institutions to analyze the topics of the records, which resulted in a proposed topic map with 7 facets as top levels. This study suggests an integrated and visualized tool using a topic map to improve the searching service of foreign records about Korea from Korean institutions, especially the US records related to Korea.

Keywords: 토픽맵, 한국관련 해외기록물, 미국기록물, 통합검색, 패싯분석, 기록서비스
Topic Map, Foreign Records about Korea, US's Historical Records, Integrated Search, Facet Analysis, Archival Service

1. 서론

한국과 관련해서 해외에서 생산되고 있는 기록물은 한국전쟁, 군사정권 등 우리나라가 겪은 대표적인 사건들의 상황을 다각적으로 조명하고 연구할 수 있는 중요한 가치를 가지고 있는 기록물이다. 한국과 관련하여 해외에서 생산되고 있는 기록물인 ‘한국 관련 해외기록물’의 개념은 이를 연구한 다양한 선행연구에서 나타난 여러 정의를 종합 분석한 결과(김성겸, 2005; 마정춘, 2006; 이승휘, 2006; 박은봉, 2008; 조혜경, 2008; 김나정, 2016), 근·현대 전후 시기에 주로 생산되었으며, 해외에 출처를 두고 있고, 한국과 관련된 주제를 다루고 있는 기록물로 설명된다. 이 기록물들은 국내 기관에서 생산된 기록과 차별화하여 이용자에게 다양한 시사점을 파악할 수 있게 해준다. 한국과 관련되어 생산된 해외기록물은 역사적, 질적으로 높은 가치를 지니는 기록물로 여겨지며, 현재까지 한국사, 국방사 등 다양한 한국과 관련된 연구에 활용되고 있다. 이와 같은 이유로 현재 국가기록원, 국립중앙도서관 등 주요 국가 기관에서는 한국과 관련된 해외기록물을 수집하고 있으며 일부 기관에서는 이용자가 활용할 수 있도록 한국 관련 해외 생산 기록물을 서비스하고 있다. 그러나 각 기관별로 기록물의 수집범위도 다르고 기록물을 제공하는 방식도 상이하여 어떠한 기록물을 제공하는지 이용자가 해외기록물의 존재와 소재를 파악하기 어려우며, 각각 서비스 및 검색 방식이 달라 이용자가 쉽게 검색하기 힘들다는 문제점이 나타나고 있다. 또한 한국 관련 해외에서 생산된 기록물의 내용을 포괄적으로 훑어보거나 분산되어 있는 기록물을 통합하여 검색할 수 있는 환경도 구축되어 있지 못해 이용자가 해외 생산 기록물을 활용하는데 많은 장애가 발생하고 있다.

한국과 관련하여 기록물을 생산한 주요 국가는 미국, 중국, 러시아, 대만, 일본, 호주, 유럽에 소속된 국가들까지 다양하다. 한국 관련 기록물을 생산하는 국가들 중 특히 미국은 한국의 주요 수교국으로 한국과 관련된 다양한 기록물이 생산되는 국가이며 많은 양의 기록물을 생산하는 국가 중 하나이다. 미국은 한국전쟁 기록 등 우리나라의 역사기록물에 중요한 영향을 미친 국가이다. 이와 관련된 기록들은 휴전 이후 미국 기록물관리법에 의거하여 워싱턴연방기록보존소에서 미국 국립문서기록관리청(NARA)으로 이관되어 관리되었다. 1980년도를 기점으로 미국 국립문서기록관리청(NARA)에서 보존 중이었던 한국과 관련된 기록물이 정보공개법에 의하여 30년 이상 지난 문서들에 한해 공개되었다. 공개된 기록들 중 미국 군정기록, 6·25전쟁 관련 기록 등 한국의 근·현대사에 대한 정보가 해당하는 관련 기록이 대다수 포함되어 있었다. 해당 기록들은 역사적 사건 등의 증거로서 활용될 만한 가치가 크다는 판단하에 2000년대 초부터 국내의 기관에 의하여 적극적으로 수집되었다.

한국 관련 미국기록물은 한국 관련 해외기록물 중 주요 기록물에 해당하는 것으로 기록물의 생산 주체가 해외의 다양한 국가 중 ‘미국’에 출처를 두고 있는 기관에 의하여 생산된 기록물이다. 이 연구에서는 웹사이트를 통해 한국 관련 해외기록물을 서비스하고 있는 기관인 국가기록원, 국립중앙도서관, 국사편찬위원회, 국방부 국방편찬연구소 4개의 기관을 연구 대상으로 선정한 후 한국 관련 미국기록물 데이터를 수집하였다.

한국 관련 미국기록물을 수집하기 위하여 2021년 2월부터 5월까지 해당기관 웹사이트를 통해 이 기간 동안 서비스되고 있는 기록물을 조사한 후 기록물명이 포함된 목록을 수집하였다. 기관별 웹사이트를 통해 수집한 기록물 데이터는 총 163,874건이다. 이 연구에서는 위에서 기술한 것과 같이 국내 기관에서 수집된 한국 관련 해외기록물의 검색의 문제점을 개선하기 위하여 한국 관련 해외기록물을 생산하는 주요 국가인 미국에서 생산된 기록물을 중심으로 국내 기관에 분산되어 있는 기록물을 통합하여 검색하고 기록물의 내용을 파악할 수 있도록 하고자 하였다. 또한 키워드에 의존하는 현재 검색 방법을 개선하기 위하여 이용자가 다각적으로 기록의 내용을 훑어볼 수 있도록 하는 방안으로 주제분석을 하여 패시와 토픽을 도출하였고 도출된 패시와 토픽을 기반으로 토픽맵을 구성하였다.

2. 한국 관련 해외기록물의 특성 및 관련연구

한국 관련 해외기록물은 국내에서 중요한 기록물로서 인식되어 2000년대 전후부터 현재까지 20여 년이 넘는 기간 동안 수집이 이어져 오고 있다. 현재까지 많은 양의 한국 관련 해외기록물이 수집, 축적, 서비스되고 있지만 한국 관련 해외 기록물은 사전적으로 정확히 명시된 바가 없으며, 수집기관마다 각각 상이한 명칭인 해외기록물, 해외 한국관련 기록물, 해외 사료, 국외수집자료 등으로 명명되고 있다. 한국 관련 해외기록물의 정의와 특성은 다양하게 정의되어 오고 있었다. 주요 선행연구(김성겸, 2005; 마정춘, 2006; 이승휘, 2006; 박은봉, 2008; 조혜경, 2008; 김나정, 2016)를 통해 정의된 한국 관련 해외기록물은 연구자별로 조금씩 차이는 보이지만 대표적으로 다음 네 가지의 특성을 가진다.

첫째, 해외에 산재한 한국 관련 기록물은 한국의 근·현대사 관련 연구 및 영토분쟁, 피해자 보상, 군사 분야 등 다양한 방면에서 활용, 보존, 연구되어야 할 증거적 가치를 가진 1차 자료이다. 둘째, 한국 관련 해외기록물의 경우 원본은 해외에 출처를 두고 있으며 원소장처에 의하여 관리 및 보존되고 있다. 셋째, 한국 관련 해외 기록물은 일부 기증을 제외하고는 대부분 사본 및 영인본의 형태로 국내 수집기관에 의하여 수집, 관리, 서비스되고 있다. 넷째, 한국 관련 해외기록물은 기록의 대부분이 디지털 사본의 형태로 웹 자원의 특성을 가지고 있다.

한국관련 해외 기록물을 연구한 사례를 살펴보자면 우리나라의 유실된 역사를 복원하기 위한 방안으로 한국 관련 해외기록물을 수집 활용하고자 제안한 김성겸(2005)의 연구는 기록의 수집에서 기술까지 모든 과정을 하나의 프로세스로 구축함으로써 한국 관련 해외기록물을 서비스하는 기관이 실무에 적용할 수 있는 실질적인 해결방안을 제시하였다. 또한 해외에 소재하고 있는 한국과 관련된 역사기록의 수집과 지속적인 관리를 위한 방안으로 통합적인 분류체계와 기술요소의 표준을 제안하였다. 역사적인 분야에서 다방면으로 의미를 가지는 한국 관련 해외 기록물의 산재를 문제점으로 지적한 연구(박은봉, 2008)에서는 국립중앙도서관의 해외 한국 관련 기록의 수집, 정리 및 서비스 현황을 일본의 선진 사례와 비교하여 문제점을 지적하였다. 이 연구에서는 수집기관이 수집 기록물을 신속하게 정리한 후 공개할 수 있도록 국내 해외 수집기록물 데이터베이스 구축, 열람 및 활용의 활성화를 위한 원문 자료를 제공하는 것을 공동활용 방안으로 제시하였다.

한국 관련 해외 기록물의 관리 현황을 연구한 조혜경(2008)의 연구에서는 일제강점기와 관련하여 해외에 소재하고 있는 한국 관련 기록물이 수집되어 관리되는 방식이 일정한 분류방법과 기술규칙을 가지지 않고, 도서관에서 책을 관리하는 방식이 답습되고 있다고 언급하고 있다. 이로 인해 기록물의 분류와 기록물의 특성을 살린 다계층 기술이 불가하다는 점을 문제점으로 지적하였다. 이 문제점을 개선하기 위해서는 분류의 최상위 단계로 나타나는 컬렉션 개념이 필요하다고 지적하였다. 컬렉션을 생성한 후 그 결과가 이용자가 효율적으로 사용할 수 있도록 기술해야 하며, 자관에서 수집하는 기록물의 성격에 맞는 적절한 기술규칙을 만들어 수집된 기록물에 대한 통합적인 활용을 보장할 수 있어야 한다고 제안하였다. 마정춘(2006)은 규모가 큰 국가 예산의 투입과 정기적인 수집 활동에도 한국 관련 해외기록물을 수집하는 수집기관들은 관리에 치중하여 이용자의 효율적인 기록물 활용을 간과하고 있다는 점을 문제점으로 지적하였다. 이 연구에서는 기관에서 수집한 미 육군 역사기록물의 현황을 파악하여 미국 국립문서기록관리청(NARA)의 소장 현황과 비교하는 과정을 통해 수집기관이 중복 수집을 방지하고, 수집되는 기록의 성격에 맞는 정리가 이루어져야 한다고 제안하였다.

국내기관의 한국 관련 해외기록물 서비스 현황을 조사한 김나정(2016)은 국내 수집기관들을 통해 이루어지는 한국 관련 해외 수집 기록물의 꾸준한 수집 활동에 비해 이용자 서비스가 적절하게 제공되고 있지 못한 점을 언급하면서 한국 관련 해외 수집 기록물의 활용단계와 이용자 서비스 영역에 중점을 두고 연구를 수행하였다. 특히 온라인 검색 및 열람 서비스 분야의 문제점을 지적하면서 이를 개선하기 위한 방안을 모색하고자 하였다. 그러나 문제점을 해결할 수 있는 구체적인 방안을 제시하고 있지 못한 한계점이 있다. 이와 같이 한국과 관련된

해외기록물에 대하여 여러 연구가 시도되었지만 한국 관련 해외기록물을 통합적으로 검색하는 방안을 모색하거나 기록물의 내용을 정리하여 이용자가 쉽게 파악할 수 있도록 제공하는 등 한국 관련 해외기록물 검색의 효율성을 개선하고자 하는 연구는 활발하게 수행되고 있지 못한 것으로 분석되어 이 연구에서는 한국 관련 해외기록물 중 미국 생산 기록물을 중심으로 이용자가 기록물의 주제를 쉽게 파악할 수 있도록 기록물의 키워드를 중심으로 주제분석을 하여 패킷과 하위 토픽을 도출하여 토픽맵으로 구성하여 제시하였다.

주제분석에 적용된 패킷과 관련된 연구를 살펴보자면 방송과 관련된 소재 중 정보이용 요구가 상대적으로 높게 나타난 시사보도 뉴스 프로그램을 대상으로 주제 표현력을 향상시키기 위한 방편으로써 패킷 분석한 사례가 있다(심지영, 2014). 이 연구는 방송 프로그램의 내용을 다면적인 관점으로 분석하여 다양한 주제 관계를 결합하는 효과적인 도구로 패킷 개념을 적용하였으며, 이로 인해 이용자가 방송 자료에 보다 다각적으로 접근할 수 있는 방안을 제시하였다. 기록물과 관련하여 수행된 패킷연구로 유네스코 기록물과 관련된 연구가 있는데 박도영(2020)은 유네스코가 소장하고 있는 기록 중 유네스코 가입 후부터 한국전쟁 그리고 전후 재건 및 개발시기까지 일련의 기간 동안에 생산된 기록물을 대상으로 그 기록물의 특성을 반영할 수 있고, 이용자의 효율적인 이용을 보장할 수 있는 적절한 분류체계로서 패킷 분류 체계를 제안하였다.

토픽맵과 관련하여 검색효율을 개선하고자 하는 연구는 특허, 사전이나 웹정보, 교육정보 등 다양한 콘텐츠를 대상으로 수행되었으나(고유미, 2005; 정경희, 김판구, 2006; 김소영 외, 2006; 이상철, 2012) 기록연구분야에서는 다양하게 시도되고 있지 못하였다. 최근 기록물과 관련하여 토픽맵 모델링을 위해 수행된 사례가 있는데 이 연구에서는 대학에서 사용되는 학사행정 기록물을 중심으로 대학 학사행정 업무에 중점을 둔 의미기반 검색을 하기 위하여 토픽맵 기반 검색 프레임워크를 설계하였으며 기록물 검색을 개선하는데 토픽맵을 적용하는 시도를 하였다(신지유, 2016).

3. 수집기관별 한국 관련 해외기록물 현황 및 분석

3.1 한국 관련 해외기록물 수집기관 현황

이 연구에서 조사대상으로 선정한 기관은 국가기록원, 국립중앙도서관, 국사편찬위원회, 국방부 군사편찬연구소 4곳이다. 이 기관들은 한국 관련 해외기록물을 수집하는 기관 중 웹으로 이용자가 기록에 접근할 수 있는 서비스를 운영하는 기관들이다. 4개 기관별로 기록물을 수집 방법, 제공하는 방식, 특성, 기록물 현황을 간략히 살펴본 후 공통점과 차이점에 대한 비교 및 분석을 하고 분석결과를 종합하여 한국 관련 미국기록물의 검색에 관한 문제점을 도출하였다.

3.1.1 국가기록원

국가기록원은 한국 관련 해외기록물을 ‘해외기록물’이라 칭하면서 해외에 산재한 한국 관련 기록물을 수집 및 관리하여 국내 기록의 보완, 한국의 근·현대사 연구 및 영토분쟁, 피해자에 대한 보상 등의 분야에 활용할 목적으로 기록을 수집하고 있다. 수집 방법은 해외에 소재한 기록관리기관과 MOU 체결, 기증, 구입, 사본 수집 등이다. 수집 대상으로 선정한 기록물은 외국의 정부, 단체, 개인, 기록관리기관 등에서 생산, 접수, 소장 중인 한국 관련 기록물이다. 국가적으로 보존할 가치가 높다고 판단되는 해외기록물로서 종이문서, 대장류, 시청각물, 각종 간행물, 도면, 지도, 카드, 구술자료, 기록의 사본, 디지털자료 등 다양한 형식의 기록을 수집한다. 국가기록원은 한국 관련 해외기록물을 수집하는 단위를 ‘점’으로 설정하였고, 2021년 현재까지 수집된 한국 관련 해외기록물은 총 2,715,785점이다.

3.1.2 국립중앙도서관

국립중앙도서관은 한국 관련 해외기록물을 디지털컬렉션 중 하나의 분야로 ‘한국 관련 기록물’이라고 지칭하면서 한국 관련 해외기록물을 수집하고 있다. 국립중앙도서관은 해외에 소재하고 있는 한국과 관련된 주요한 자료를 지속적으로 수집하여 연구자를 대상으로 기록물을 제공하고 있다. 2004년도부터 미국 국립문서기록관리청(NARA)에서 수집 및 보존하고 있는 미국 행정부가 생산한 공문서와 그 컬렉션 중 한국과 관련된 기록자료를 중심으로 수집하여 한국의 근·현대기 당시 한반도의 정세와 주변국과의 관계에 관한 연구를 위한 목적으로 1차 자료를 이용자에게 영인 제공하고 있다. 기록물 형식은 주로 문서를 스캔한 사본 또는 영인본이며, 주요 수집 분야는 일제강점기와 한국전쟁 전후기 관련 문서, 베트남 전쟁 참전 한국군 관련 문서 등 군사 분야 문서, 한국전쟁 시기에 노획 및 수집한 노획문서컬렉션 중 북한노획 문서 등이다. 수집된 기록물 양을 조사한 결과 2020년 말까지 기록의 수집 단위를 면으로 설정하여 총 3,135,036면을 수집한 것으로 나타났다.

3.1.3 국사편찬위원회

국사편찬위원회는 한국 관련 해외기록물을 전자자료관이라는 자체 웹사이트를 통해 ‘국외사료’라 칭하면서 수집된 기록물을 서비스하고 있다. 세계 각국에 산재하고 있는 한국사 관련 자료를 조사 및 수집하여 체계적인 정리를 통해 연구자와 일반 국민에게 제공하는 것을 목적으로 서비스를 운영하고 있다. 국외 사료는 미국, 일본, 중국 포함 14개의 국가를 대상으로 하고 있다. 국외 유관 기관과 협정, 국외자료조사위원회, 국외로의 출장, 자료의 교환 등의 방법을 통해 수집을 진행하고 있다. 이러한 방식으로 수집된 국외 사료를 분류, 정리, 등록 작업 후 인터넷을 통해 공개하고 자료집을 간행하는 방식을 취하고 있다. 수집된 기록물 유형은 문서, 마이크로필름(컷, 롤), 마이크로피시(장), CD 또는 DVD, 도서, 그림/지도/글씨(점), 사진, 테이프 및 비디오, 슬라이드 등이다. 수집된 기록물은 소장처, 주제별로 계층적으로 구조화하여 목록으로 제공하고 있다. 2001년부터 2017년까지 수집을 진행하였으며 총 9,076,538건이 수집되었다.

3.1.4 국방부 군사편찬연구소

국방부 직속 최고 군사연구기구인 군사편찬연구소는 한국 관련 해외기록물을 ‘국외수집자료’로 칭하면서 군사문제와 관련된 역사적인 자료의 제공으로 효율적인 군사 정책의 수립이 이루어지도록 하기 위해 관련 기록물을 수집하여 서비스하고 있다. 국방사와 전쟁사에 관한 연구 및 편찬, 군사작전과 관련된 자료를 포함하여 국내 및 해외의 관련 기관에서 소장하고 있는 각종 국방과 관련된 기록을 수집, 보존, 관리하고 있다. 수집 기간은 2019년부터 2021년까지이며, 총 2,061건(217,635매)이다.

한국 관련 해외기록물을 수집하여 웹으로 서비스하는 수집기관의 공통된 특성은 3가지로 다음과 같다.

첫째, 4개 수집기관 모두 해외에 출처를 두고 생산된 한국 관련 기록물을 수집하기 위하여 국가의 기관, 인력, 예산이 투입되고 있는 공적 기관들이다. 국가기록원은 행정안전부, 국립중앙도서관은 문화체육관광부, 국사편찬위원회는 교육부, 국방부 군사편찬위원회는 국방부에 소속된 기관이다.

둘째, 수집기관들은 각 수집기관별로 마련된 법적 근거에 의해 한국 관련 해외기록물을 수집하고 있다. 관련 법적 근거를 알아보면 국가기록원은 『공공기록물 관리에 관한 법률』 제46조(주요 기록정보 자료 등의 수집)와 동법 시행령 제84조(주요 기록정보 자료 수집 등), 국립중앙도서관은 『도서관법』 제20조의2 제6항 및 동법 시행령 제13조의2 제1항에 따른 수집대상 온라인 자료의 종류 및 형태에 관한 고시, 국사편찬위원회는 『사료의 수집·편찬 및 한국사의 보급 등에 관한 법률』, 국방부 군사편찬연구소는 국방부 군사편찬연구소령 제2조3에 근거하여 한국 관련 해외기록물을 수집하고 있다.

셋째, 이 수집기관들은 근·현대기 시기에 집중하여 기록물을 수집하고 있다. 공통적으로 나타나는 특성으로

네 기관이 수집한 한국 관련 미국기록물에는 한국전쟁과 관련된 내용의 기록이 주요 기록물인 것으로 조사되었다.

김나정(2016)의 연구를 참조하여 한국 관련 해외기록물을 수집하여 웹으로 서비스하는 수집기관의 차이점은 다음과 같이 분석되었다.

첫째, 한국 관련 해외기록물을 수집하는 지역의 측면에서 차이가 있다. 국가의 한정을 두지 않고 해외 전 지역을 대상으로 수집하는 기관은 국가기록원과 국사편찬위원회이며, 주로 미국을 대상으로 수집하는 기관은 국립중앙도서관과 국방부 군사편찬연구소이다.

둘째, 수집 범위의 차별성은 한국과 관련된 다양한 분야를 전반적으로 수집하는 기관과 특정 주제의 기록물을 수집하는 기관으로 나타난다. 포괄적으로 한국 전체에 대한 범위의 기록물을 수집하는 기관은 국가기록원, 국립중앙도서관, 국사편찬위원회이며, 군사 및 국방, 전쟁에 특화된 분야에 대한 기록을 수집하는 기관은 국방부 군사편찬연구소이다.

셋째, 수집 자료 유형 측면을 살펴보면 문서를 중심으로 기록을 수집하는 기관은 국립중앙도서관, 국방부 군사편찬연구소이고, 문서 외 다양한 자료 유형의 기록을 수집하는 기관은 국가기록원과 국사편찬위원회이다.

넷째, 기관별로 수집된 자료의 활용 분야를 알아보면 국가기록원은 영토 분쟁과 피해자 보상 등의 증거, 국립중앙도서관과 국사편찬위원회는 한반도 정세 및 주변국과의 연구, 국방부 군사편찬위원회는 한국 관련 군사 및 전쟁 분야의 연구를 위해 한국 관련 해외기록물을 서비스하고 있다.

<표 1> 한국 관련 해외기록물 수집기관별 특징

구분	국가기록원	국립중앙도서관	국사편찬위원회	국방부 군사편찬연구소*
수집지역	해외 전 지역	미국 위주	해외 전 지역	미국 위주
수집시기	근·현대	1945년 이후	조선시대~ 1980년대	1950 ~ 1975년대
수집주제	전 주제 영역 “외국의 정부, 단체, 개인, 기록관, 리기관 등이 생산, 접수, 소장 중인 대한민국 관련 기록물로서 국가적으로 보존가치가 높다고 인정되는 해외기록물”	한국 근현대사 “미국국립기록관리청(NARA)에서 보존중인 미 행정부 생산 공문서와 컬렉션(Collection) 중 한국 관련 기록자료”	한국사 전반 전 주제 영역	미국국립기록관리청(NARA)을 통해 수집한 한국 관련 군사사 및 전쟁사, 국방 관련 영역의 기록물
주요 수집주제	한국 근·현대사, 독립운동사, 북한관련기록, 한국전쟁관련기록, 한인관련일반기록 등	민사 및 군사 분야 문서, 제2차세계대전과 한국전쟁기에 노획, 수집한 노획문서 컬렉션 중 북한노획문서 등	근·현대 대외관계, 북한 관계, 재외동포, 독립운동, 일제 침략 관련, 한국전쟁, 한국 민주화운동, 경제정책 등	베트남 전쟁 및 6·25전쟁 관련 기록물
수집대상 자료유형	종이문서·대장류, 시청각물, 간행물, 도면, 지도, 카드, 구술자료, 기록 사본, 디지털 자료	문서류, 도서류	문서(이미지 포함), M/F, 마이크로피시, 도서, CD/DVD, 사진, 지도, 테이프, 슬라이드, 구술 자료	문서류
수집기록 활용분야 *	한국 근·현대사 연구 및 영토 분쟁, 피해자 보상 등	한국 근·현대기 한반도의 정세와 주변국과의 관계 연구	한국 근·현대기 한반도의 정세와 주변국과의 관계 관련, 한국사 연구 및 편찬	국방사, 전쟁사, 군사사, 국제분쟁사 등 한국 관련 군사 연구 및 편찬

출처: 김나정(2016), 한국 관련 해외 수집 기록물의 이용자 서비스 개선방안 연구, p. 33 <표 2> 기관별 수집 기준 및 범위

*: 김나정(2016)의 해외 기록물 수집 기준 및 범위를 정리한 표에 수집기관(국방부 군사편찬연구소)과 기관별 특징(수집기록 활용분야)을 반영하여 추가한 사항

각 기록 수집기관의 특성을 종합하여 요약적으로 제시한 것은 <표 1> 과 같다. 한국 관련 해외기록물을 수집해서 웹을 통하여 이용자에게 서비스하는 환경을 조사한 결과 몇 가지 문제점이 도출되었다.

첫째, 각각의 기관에서 한국 관련 해외기록물을 다양한 범위와 주제로 다량으로 수집, 서비스하고 있지만 각각의 사이트에서 분산되어 있고 목록을 일일이 보기 전에는 어떤 내용의 기록을 어떤 사이트에서 제공하는지 파악

하기 쉽지 않았다. 특히 한국 관련 해외기록물에 대한 사전지식이 없는 이용자의 경우 어떤 사이트에서 한국 관련 해외기록물을 제공하고 있는지 찾아내기도 쉽지 않은 현황이다.

둘째, 한국 관련 해외기록물은 다양한 주제로 구성된 기록물로서 보편적인 내용을 다루거나 일반적인 활동을 설명하는 기록이라기보다는 특정 주제나 시기, 사건 등에 해당하는 기록들이 많다. 따라서 이용자는 한국 관련 기록물의 전체 내용을 통합적으로 훑어보면서 제공되는 기록물의 범위를 파악할 필요성이 있다. 그러나 현재 제공되고 있는 서비스에서는 목록을 단편적으로 접근할 수 있도록 되어 있어 전체적으로 내용을 파악하기 어렵다.

셋째, 한국 관련 해외 기록물은 상세검색, 키워드 검색, 단순 목록 제공 등 수집기관마다 제각기 상이한 검색 방식과 면, 점, 건 등의 기록물 수집 단위를 취하고 있다. 상세한 키워드나 기록물의 단위에 대한 사전정보를 가지고 있지 않은 연구자 또는 연구자가 아닌 일반 이용자는 검색을 시작하는 것에서부터 어려움을 느낄 수 있다. 또한 이용자가 한국 관련 해외기록물에 대한 배경지식이 풍부하지 않다고 가정하였을 때, 적절한 키워드를 선택하지 못하면 찾고자 하는 기록물을 획득하는데 어려움이 따를 수 있다.

3.2 한국 관련 해외기록물의 데이터 수집 및 처리

한국 관련 해외기록물은 역사적, 증거적으로 중요한 가치를 지녔음에도 불구하고, 수집된 기록을 이용자가 활용할 때 소재와 내용을 파악하기 힘든 문제점이 있다. 특히 여러 나라에서 생산된 한국 관련 기록물을 수집하는 기관이 분산되어 있고 각 기관별로 수집하는 국가가 상이하여, 이용자는 생산 국가별로 한국 관련 기록물의 내용을 파악하기 어려웠다.

이 연구에서는 이와 같은 문제점을 개선하기 위하여 한국 관련 해외기록물 중 한국 관련 기록물 생산 주요 국가인 미국을 중심으로 기록물의 내용을 분석하여 이용자에게 통합 서비스하는 방안을 제시하고자 하였다.

미국에서 생산된 한국 관련 기록물을 수집하여 서비스하는 주요 기관은 국가기록원, 국립중앙도서관, 국사편찬위원회, 국방부 군사편찬연구소이다. 이 기관들은 웹을 기반으로 한국 관련 미국기록물을 서비스하고 있는 기관으로서 이용자가 해당 기록물을 이용할 수 있는 환경을 구축한 기관이다. 이 연구에서는 현재 이용자에게 한국 관련 미국기록물을 웹으로 서비스하는 기관을 연구대상으로 선정하여 데이터를 수집하였다. 선정된 기관의 웹사이트를 통해 제공되고 있는 한국 관련 미국기록물을 2021년 2월부터 5월까지 4개월 동안 수집한 결과는 총 163,874건으로 <표 2>와 같다.

<표 2> 수집기관별 한국 관련 미국기록물 데이터 수집 현황

구분	기록물 군	기록물 시리즈	기록물 철	기록물 건	총 데이터 수
국가기록원	-	-	2,251	67,094	69,345
국립중앙도서관	-	619	-	35,128	35,795
국사편찬위원회	-	1,055	55,648	-	56,799
국방부 군사편찬연구소	-	140	1,743	-	1,935
총 계					163,874

각 기관별로 한국 관련 미국기록물을 수집한 결과 기관별 기록물의 단위에 차이가 있었다. 국가기록원의 경우 기록물 군, 기록물 시리즈 단위를 사용하지 않고, 기록물 철, 기록물 건 단위를 사용하여 한국 관련 미국기록물을 제공하고 있었다. 국립중앙도서관의 경우 기록물 철을 제외한 기록물 군, 기록물 시리즈, 기록물 건 단위로 기록물을 제공하고 있었다. 국사편찬위원회와 국방부 군사편찬연구소의 경우 기록물 건을 제외한 기록물 군, 기록물 시리즈, 기록물 철 단위를 사용하고 있는 것으로 조사되었다.

기록물의 내용을 파악할 수 있는 기록물 단위를 선정하게 된 과정은 다음과 같다. 기관별로 차이를 보이고 있는 기록물의 단위를 기록물의 내용을 표현할 수 있도록 기술된 단위로 데이터를 수집하는 범위를 설정하고자 기록물의 단위를 기관별로 분석하였다. 기록물의 내용을 적절하게 기술한 단위는 기록물 시리즈 제목과 기록물 철 제목으로 나타났다. 그러나 국가기록원의 경우 다른 기관과 달리 기록물 시리즈 단위가 존재하지 않고 기록물 철과 건 단위로 제공하고 있었다. 국가기록원의 기록물철 제목의 경우 <미육군성 부관감실 - 제40보병사단 제143 야전포병대대 사령부 보고서, 1953>가 국립중앙도서관, 국사편찬위원회, 국방부 군사편찬연구소의 기록물 시리즈 제목에서는 각각 <Bureau of Far Eastern Affairs - General Records of the Office of the Country Director for Korea, 1952-66>, <Korean Armistice Implementation Records, 1951 - 1980 [Entries 14, 14B, 14D, 14E, 14F](General Records, 1951 - 1957 [Entry 14F])>, <SWNCC General Secretariat Records - Agenda, Feb. 1945-Mar.1946 / Subcommittee on the Far East General Records, 1945-48>로 나타나고 있었다. 또한 국가기록원의 기록물 건 제목의 경우 <장군들의 방문: Norman Gale 중장이 창경원 방문 당시 찍은 사진>은 국사편찬위원회, 국방부 군사편찬연구소의 기록물 철 제목에서는 각각 <미군 활동 기록 사진 (월남민이 거쳐야 할 과정들)>, <Photos of Pigeon Activities in Korea> 등으로 나타났다. 이와 같이 국가기록원의 기록물 철과 건 단위와 다른 기관의 기록물 시리즈와 철 단위를 비교한 결과 기록물 시리즈와 철 단위와 유사한 내용이 포함되어 있었다. 따라서 국가기록원의 경우 기록물 철과 건 단위를 기록물 시리즈와 철 단위로 상향 조정하였다.

국립중앙도서관은 다른 기관과 달리 기록물 철 단위가 존재하지 않고, 기록물 건 단위가 존재하였다. 국립중앙도서관의 기록물 건 제목의 경우 <G-3 Operations Journal, 22 Dec 1952>가 국사편찬위원회, 국방부 군사편찬연구소의 기록물 철 제목에서는 각각 <G-3 Journal, 20 Dec. 1951>, <3rd Infantry Div. Journals 1953.2 (G2-3)> 등으로 나타나고 있었다. 이처럼 국립중앙도서관의 기록물 건 단위와 다른 기관의 기록물 제목과 비교한 결과 기록물 철 단위와 유사한 내용을 표현하고 있었다. 따라서 국립중앙도서관의 경우 기록물 건 단위를 기록물 철 단위로 상향 조정하였다.

이러한 과정을 통해 기록물 단위를 조정한 수집데이터 결과는 <표 3>과 같다.

<표 3> 수집기관별 한국 관련 미국기록물의 수집 데이터 조정 현황

구분	기록물 시리즈	기록물 철	총 데이터 수
국가기록원	2,251	67,094	69,345
국립중앙도서관	619	35,128	35,795
국사편찬위원회	1,055	55,648	56,799
국방부 군사편찬연구소	140	1,743	1,935
총 계			163,874

한국 관련 해외기록물에서 기록물 시리즈와 기록물 철에서 추출된 단어를 대상으로 불용어 등 내용분석 시 무의미한 용어들을 삭제한 후 빈도 수가 높은 고빈도어를 중심으로 데이터를 정련하였다. 기록물 시리즈에서는 총 4,731개의 단어가 산출되었고 이 중 불용어를 처리한 주제 키워드는 3,604개이다. 기록물 철에서는 총 74,645개의 단어가 산출되었고 이 중 불용어를 처리한 단어 중 출현 빈도 5 이상의 12,525개를 주제 키워드로 선정하였다. 외국어 자료와 한글 자료가 분산됨을 방지하기 위하여 먼저 영어와 한글로 주요 단어를 추출한 뒤 단어 단위에서 영어를 한국어로 번역한 결과로 전거통제를 실시하였다. 전거통제는 국가명 및 지역명, 기관명, 동음이의어, 기록의 유형, 외국어 등 총 5가지 분야에서 수행되었다.

전거통제를 한 키워드를 대상으로 기록물 시리즈와 기록물 철의 주제어 합계를 통합한 16,129개의 키워드에 대한 내용분석을 실시하였다. 정련한 주제 키워드를 분석한 결과 동의어와 유사어, 축약어 형태로 다양하게 동일

개념이 분산되어 있는 것으로 나타나 2차로 동의어와 유사어, 축약어에 대한 전거통제를 시행하였다.

4. 한국 관련 미국기록물의 내용분석 및 개선방안

4.1 한국 관련 미국기록물의 패킷 분석

패킷 분석은 기록물에서 산출된 키워드를 내용에 따라 코딩하면서 귀납적 방식을 통해 범주화하는 단계로 수행되었다. 이 연구에서는 진행되는 패킷 분석은 Vickery의 패킷 분석 7단계(La Barre, 2010)를 적용하여 패킷 분석 단계 중 1단계부터 4단계까지 적용하였다. 적용한 결과는 <표 4>와 같다.

<표 4> Vickery 패킷 분석에 근거한 한국 관련 미국기록물 패킷 분석 과정

Vickery 패킷 분석 7단계		한국 관련 미국기록물 패킷 분석 과정	
패킷 분석 과정	1단계: 주제 영역 설정	패킷 주제 영역 설정	한국 관련 미국기록물 163,874건에서 도출된 키워드 분석 및 내용분석 및 코드화 키워드 범주화
	2단계: 패킷 범위 설정	패킷 정의	각 패킷별 범위 설정 7개 패킷 도출
패킷, 패킷분류체계 분석	3단계: 패킷의 확장 및 구조화	패킷별 토픽 구조화	군대, 기관, 기록물, 인물, 주제, 장소 및 지명, 시대 7개의 패킷 도출
	4단계: 패킷 범위 주기 생성	중간토픽, 세부토픽의 정의	도출된 패킷별 정의, 중간토픽, 하위토픽으로 분류되는 토픽 예시 설명
	5단계: 패킷 배열	토픽맵 생성	각 패킷별 토픽별 시각화

패킷 주제 영역 설정 단계에서는 한국 관련 미국기록물 163,874건에서 도출되었던 키워드와 내용분석, 코드화 그리고 키워드 범주화를 실시하였다. 전처리를 거쳐 산출된 키워드를 대상으로 내용에 따라 코딩을 실시하여 주요 개념을 도출하였고 개념사전을 생성하였다. 생성된 개념사전을 기반으로 총 7개의 패킷(기관, 군대, 기록물, 인물, 주제, 장소 및 지명, 시대)을 도출하였다. 도출된 패킷에 할당된 키워드를 분석하여 각 패킷의 주요 하위 개념으로 구조화하였다. 도출된 패킷과 하위토픽의 검증은 2번에 걸쳐 실시되었으며 1차 검증은 문헌정보학 전공자 2인에 의해 수행되었고 2차 검증은 대학도서관 사서 2인을 통해 검수를 받은 후 지적된 패킷을 보완하였다. 검증과정에서 생성된 패킷이 다양한 접근점을 제공하는데 있어 차별성이 있는지에 대하여 논의되었고 군대 패킷이 기관과 분리되는 것에 대하여 검증자 전원이 분리되는 것으로 동일의견을 제시하였다. 또한 일부 하위 토픽의 분류가 재편성되었다. 각 패킷별 토픽은 중간토픽, 세부토픽으로 계층화되었으며 이 과정을 통해 생성된 개념사전과 계층관계를 바탕으로 ATLAS, Ti를 활용하여 토픽맵을 생성하였다.

패킷 정의 단계에서 나타난 패킷들은 총 7개로 각각 군대, 기관, 기록물, 인물, 주제, 장소 및 지명, 시대이며 각 패킷에 할당된 키워드 사례는 <표 5>와 같다. 기관패킷은 한국 관련 미국기록물에서 나타나는 의사결정이나 실행에 참여하는 조직이나 집단을 의미한다. 군대패킷은 국가에서 조직한 일정한 조직 편제를 가진 군인 집단을 의미하는 것으로 이 연구에서는 군인 집단과 군인 집단의 활동 및 연관 장소, 역사, 기록물 등을 군인 패킷이라고 정의하였다. 군대는 기관의 유형으로 볼 수 있지만 한국 관련 미국기록물 내에서 군대와 관련된 키워드의 수와 내용이 많다는 점에서 핵심내용이라고 판단하여 다른 유형의 패킷과 분리하여 독립 패킷으로 처리하였다. 기록물

패킷은 한국 관련 미국기록물에서 나타나는 기록물의 생산되는 주기, 기록물의 성격, 외적인 특성 등을 반영하여 간행물, 기념물, 기록, 보고서, 비망록, 사진 및 도면 등을 포함하는 패킷이며 인물패킷은 인물과 관련된 국적, 신분, 직업 등과 인물의 행동을 유형화하여 생성된 패킷이다. 인물 패킷에 범주화된 사례로는 국가원수, 한국인, 외국인, 이민자, 변호사, 통역가, 번역가, 언론인 등이 해당된다. 주제패킷은 일반적으로 대화 또는 연구 등에서 중심이 될 수 있는 문제를 의미하는 패킷이며 장소 패킷은 인물이 일정 행위를 수행하는 공간과 자연적으로 생겨난 장소 등이 함께 포함되도록 규정하였다. 시대패킷은 특정 사건명을 통해 시대를 구분하는 것이 아니라 한국 관련 미국기록물 상에서 나타난 연도를 사용하여 패킷을 도출하였다. 시대 패킷은 계층적 관계가 다양하지 않기 때문에 토픽맵으로는 발전시키지는 않았으나 한국 관련 미국기록물의 통합 검색 시 필터로 활용하고자 패킷으로 도출하였다. 즉, 세부 연도에 대한 접근점을 제공하기 보다는 시대로 통합하여 접근점을 제공하나 시대를 단독으로 접근점을 제공하기에는 도출된 시대의 수가 적었다. 향후 세부 연도를 분석하여 시대보다는 세밀한 단위로 접근점을 제공하는 방안을 모색하는 것도 필요하다.

〈표 5〉 패킷 분석에 적용된 키워드 사례 및 키워드 건수

패킷명	분석 키워드 사례	해당 키워드 건수
기관	감사부, 외자청, 고문단, 공단, 기술국, 교통부, 국방부, 도서관, 국회, 농림부, 상공회의소, 방송국, 연구소, 국무부, 외교단체, 이사회, 병원, 정당, 조합, 성당, 정부 등	19,201
군대	미군, 유엔군, 육·해·공군, 감찰부, 병사, 장교, 기지, 38선, 사령부, 사단, 부대, 작전, 총격, 무기, 훈련 등	99,930
기록물	신문, 관보, 각서, 합의서, 개요서, 각종 보고서, 설계도, 포스터, 사진첩, 영화, 인터뷰, 명부, 견적서, 유인물, 진단, 교재, 각종 파일, 메일, 필름, 기념물 등	70,226
인물	미국인, 한국인, 민간인, 외국인, 이민자, 입국자, 국가원수 등	17,437
주제	감사, 거래, 기금, 각종 화학물질, 항공기, 철도, 각종 전자기계, 통신기계, 법령, 범죄, 재판, 의료기구, 의약품, 광복, 재난, 전쟁, 건설, 건축, 장비, 출판, 친선, 회담, 조약, 귀화, 여행, 이민, 순방, 입당, 민주주의, 가톨릭, 불교, 예배, 통역, 기술학, 정치학, 고용, 복지, 회의 등	29,786
장소 및 지명	덕수궁, 검문소, 광산, 논, 상점, 귀화국, 기숙사, 미국, 태국, 일본, 한국, 뉴욕, 나고야, 서울 등	18,507
시대	1756, 1897, 1910, 1945 ... 2008	67,396
총계		322,483

4.2 한국 관련 미국기록물의 패킷별 중간토픽 및 세부토픽 분석

4.2.1 기관 패킷 내 토픽분석

한국 관련 미국기록물의 기관 패킷 내용분석을 통해 코딩한 결과 총 508개의 용어가 기관 패킷에 할당되었다. 이를 다시 세부개념으로 구조화 한 결과 중간토픽은 21개, 세부토픽은 74개로 도출되었다. 키워드 수를 기준으로 상위 4개 중간토픽을 차례로 살펴보면 ‘경제’, ‘산업’, ‘외교’, ‘법’으로 나타났는데 중간토픽의 총 키워드 합계 중 과반수 이상을 차지하고 있었다. 그 중 주목할 만한 점은 ‘산업’과 관련된 기관이 특히 농업, 상공업, 제조업에 기록물이 편중되어 있다는 것이다. 이것은 한국 관련 미국기록물이 생산되고 활용되는 당시 시대적 배경이 주로 생산과 관련된 산업에 집중되어 있다는 것으로 해석된다.

세부토픽을 분석한 결과 기록물이 많은 상위 4개 세부토픽을 중간토픽과 연관하여 차례로 살펴보면 ‘경제’의 ‘원조’, ‘외교’의 ‘단체’, ‘산업’의 ‘농업’, ‘외교’의 ‘재외공관’으로 나타난다. 다양한 기관 유형이 나타남에도 위 4가지 세부토픽의 빈도 수가 높게 산출된 것은 한국 관련 미국기록물과 관련된 기관이 경제 원조와 외교에 집중되어 있기 때문이며 따라서 한국 관련 기록물의 내용과 출처를 파악하는데 경제 원조 관련 기관과 농업 관련 기관, 외교 기관은 중요한 요소가 될 것이다.

4.2.2 군대 패킷 내 토픽분석

한국 관련 미국기록물의 군대 패킷에 대한 내용분석을 통해 코딩한 결과 총 319개의 용어가 군대 패킷에 할당되었다. 이를 다시 세부개념으로 구조화 한 결과 중간토픽은 9개, 세부토픽은 35개로 도출되었다. 토픽별로 상세한 내용은 아래와 같다.

군대는 그 조직만의 규율과 질서를 가지고 조직된 군인의 집단으로 주로 기관으로서 정의된다. 이로 인해 군대가 기관의 유형 중 하나라고 여겨질 수 있으나 이 연구에서 군대는 기관적 특성을 반영하여 토픽을 생성한 것이 아니라 군대와 관련된 다양한 측면을 패킷의 토픽으로 설정하였다. 군대의 중간토픽 중 군대에서 특화되어 있는 중간토픽은 ‘군중’, ‘전략’, ‘편제’이지만 ‘국가’, ‘기관’, ‘기록물’, ‘역사’, ‘인물’, ‘장소’, ‘전략’과 같은 일반적 요소가 군대패킷을 구성하는 주요 하위요소로도 나타나 군대패킷으로 할당하여 군대를 대주제로 하여 기록물이 모일 수 있도록 처리하였다.

키워드 합계를 기준으로 상위 4개의 중간토픽을 차례로 살펴보면 ‘편제’, ‘인물’, ‘전략’, ‘군사기록물’로 나타난다. ‘편제’, ‘인물’, ‘전략’, ‘군사기록물’의 키워드 합계가 높게 나타나는 것은 ‘군중’, ‘기관’, ‘장소’ 등에 비해 다양한 세부토픽이 포함되어 있기 때문이다. 세부토픽 중 키워드 합계를 기준으로 상위 4개 세부토픽을 차례로 살펴보면 중간토픽 ‘편제’의 ‘부대’, 중간토픽 ‘인물’의 ‘지휘관’, 중간토픽 ‘편제’의 ‘사령부’, 중간토픽 ‘인물’의 ‘병사’로 나타난다. 이 4가지 세부토픽의 빈도 수가 높게 산출된 것은 군대에 해당되는 조직과 구성원이 한 국가가 아닌 다양한 국가의 군대가 존재하고 그 군대의 조직, 구성원이 모두 포함시켜 패킷을 구성했기 때문이다.

4.2.3 기록물 패킷 내 토픽분석

한국 관련 미국기록물의 기록물 패킷에 대한 내용분석을 통해 코딩한 결과 총 447개의 용어가 기록물 패킷에 할당되었다. 이를 다시 세부개념으로 구조화 한 결과 중간토픽은 10개, 세부토픽은 46개로 도출되었다. 토픽별로 상세한 내용은 아래와 같다.

기록물 패킷 내 형태적 측면에서 기록물을 구분해보면 먼저 문서의 형태로 간행물로, 기록류, 보고서류, 비망록류, 일반문서류가 포함된다. 이 경우 기록류, 보고서류, 일반문서류는 유사한 유형이라고 여길 수 있지만 이 연구에서 보고서류에 각종 보고서를 모두 포함시켰고, 기록류는 보고서류를 제외한 기관 또는 개인이 생산하는 문서로 정의하였다. 그리고 일반문서류에는 보고서류와 기록류를 제외한 그 외 형식의 기록물로 공적으로 활용되는 목적의 문서보다는 일상생활에서 주로 사용될 목적으로 생산된 목록, 증명서, 출판물 등이 다수 포함되어 있다. 이와 같은 차이점으로 기록물 유형은 기록류, 보고서류, 일반문서류는 구분되어 정의된다.

비망록류는 보통 어떤 사실을 기억하기 위해 적어 놓은 골자 또는 문서 등을 의미하는데 이 중 편지, 서한 등 서신이 개인이 생산한 문서로 판단되지만, 한국 관련 미국기록물 내에 포함된 서신은 기관에서 생산한 서신의 종류도 포함되어 있어 주목할 만한 중간토픽이다. 문서 유형이 아닌 유형으로 생산된 기록물은 사진 및 도면류, 영상 및 음성류, 전자문서류, 필름류, 기념물류가 포함된다. 중간토픽 중 세부토픽이 해당되지 않는 것은 필름류이다. 기록물이 많은 상위 4개의 중간토픽은 차례로 ‘기록류’, ‘비망록류’, ‘일반문서류’, ‘간행물류’로 나타났다. 세부토픽을 구조화 한 결과 기록물이 많은 상위 4개의 세부토픽은 ‘보고서류’의 ‘행정’, ‘비망록류’의 ‘서신’, ‘간행물류’의 ‘잡지’, ‘일반문서류’의 ‘출판물’이다.

4.2.4 인물 패킷 내 토픽분석

한국 관련 미국기록물의 인물 패킷에 대한 내용분석을 통해 코딩한 결과 총 323개의 용어가 인물 패킷에 할당되었다. 이를 다시 세부개념으로 구조화 한 결과 중간토픽은 3개, 세부토픽은 41개로 도출되었다. 주요 중간토픽으로 도출된 ‘국적’이나 ‘직종’은 일반적으로 인물의 유형을 구분하는데 종종 사용되는 개념이다. 하지만 한국 관련 미국기록물에서 나타나는 인물의 유형 중 눈여겨보아야 할 인물을 구분하는 토픽은 ‘이출민’이다.

이출민의 경우 일상 생활에서 보편적으로 사용되는 개념은 아니지만 한국 관련 미국기록물 특성 상 이민, 입국, 출국 등의 활동을 하는 인물의 유형이 나타나므로 의미있는 중간토픽이다. 다음으로 ‘민간인’ 또는 ‘외국인’과 같이 군대와 외교와 관련된 주제가 인물과 관련된 토픽으로 도출되었다. 이는 한국전쟁과 미국이민 등의 기록물이 주요 기록물로 나타나는 한국 관련 미국기록물의 특성이 반영된 것으로 분석된다. 주요 세부토픽으로 도출된 것은 ‘직종’의 ‘행정’, ‘국적’의 ‘한국인’, ‘직종’의 ‘법’, ‘직종’의 ‘경제’이다. 직종 중 가장 높은 수치를 보이는 행정은 다른 분야의 직종과 달리 행정업무를 실시하는 다양한 기관과 관련된 인물이 기록물의 내용에 포함되어 있기 때문이다.

4.2.5 주제 패킷 내 토픽분석

한국 관련 미국기록물의 주제 패킷에 대한 내용분석을 통해 코딩한 결과 총 593개의 용어가 주제 패킷에 할당되었다. 이를 다시 세부개념으로 구조화 한 결과 중간토픽은 21개, 세부토픽은 105개로 도출되었다. 토픽별로 상세한 내용은 아래와 같다. 주제 패킷 내 중간토픽은 주제와 관련된 활동 또는 활동에 수반되는 기술, 수단, 지식 등이 포함된다. 대부분의 중간토픽은 일반적으로 주제를 나타낼 때 보편적으로 사용하는 개념이지만 ‘사건’과 ‘이출’은 주목해야 할 주제 개념이다. 한국 관련 미국기록물은 단순히 한국과 관련된 기록물이 아니라 ‘전쟁’, ‘전투’, ‘정치적 사건’ 등 역사적으로 기록되어야 할 만한 순간들이 다수 포함되어 있다. ‘이출’ 또한 한국 관련 미국기록물이 ‘입출국’, ‘이민’, ‘귀화’, ‘여행’ 등의 활동과 관련 기록물이 많은 부분을 차지하고 있다는 것을 보여주는 중간토픽이다.

4.2.6 장소 및 지명 패킷 내 토픽분석

한국 관련 미국기록물의 장소 및 지명 패킷 내용분석을 통해 코딩한 결과 총 323개의 용어가 장소 및 지명 패킷에 할당되었다. 이를 다시 세부개념으로 구조화한 결과 중간토픽은 2개, 세부토픽은 16개로 도출되었다. 토픽별로 상세한 내용은 아래와 같다. 장소 및 지명 중간토픽 다른 패킷의 중간토픽들 중 가장 적은 양의 중간토픽 수를 가진다. 장소 패킷에서는 특별한 기능을 하는 의미보다 주제와 관련된 활동이나 인물이 생활하는 환경 등이 포함되어 있다. 지명은 정의와 같이 마을, 지방, 산천, 지역 등을 이르는 명칭으로 사용되지만 이 연구에서는 한국 관련 해외기록물에 나타난 국가명과 지역명을 포함하는 것이 특징이다. 장소보다는 지명이 훨씬 많은 수치가 나타나고 있는데 장소는 일정한 기능을 수행하는 기관과 분리되어 일부가 포함되었고, 국가명은 다른 패킷에 분리되지 않고 온전히 모두 포함되었기 때문에 차이를 보이는 것이다.

4.3 한국 관련 미국기록물의 주제 검색을 위한 토픽맵

이 연구에서는 키워드 검색의 한계점을 개선하기 위하여 한국 관련 미국기록물의 내용을 직관적으로 파악할 수 있도록 분석된 패킷과 토픽을 시각화하였다. 특히 한국 관련 해외기록물을 활용하는 이용자는 기록물에 대한 사전지식이 충분하지 않을 경우 검색에 적절한 키워드를 선정하는데 어려움이 있을 수 있다. 따라서 이 연구에서는 이용자에게 각 기관에 분산되어 있는 한국 관련 미국기록물을 통합하여 내용을 분석한 후 시각화하여 제공함으로써 검색서비스를 개선하고자 하였다.

시각화는 한국 관련 미국기록물의 패킷 분석으로 생성된 개념사전을 기반으로 구조화된 하위 토픽을 기반으로 수행되었다. 패킷과 패킷별로 도출된 중간토픽, 세부토픽으로 계층화된 결과를 바탕으로 ATLAS. Ti를 활용하여 토픽맵을 생성하였다.

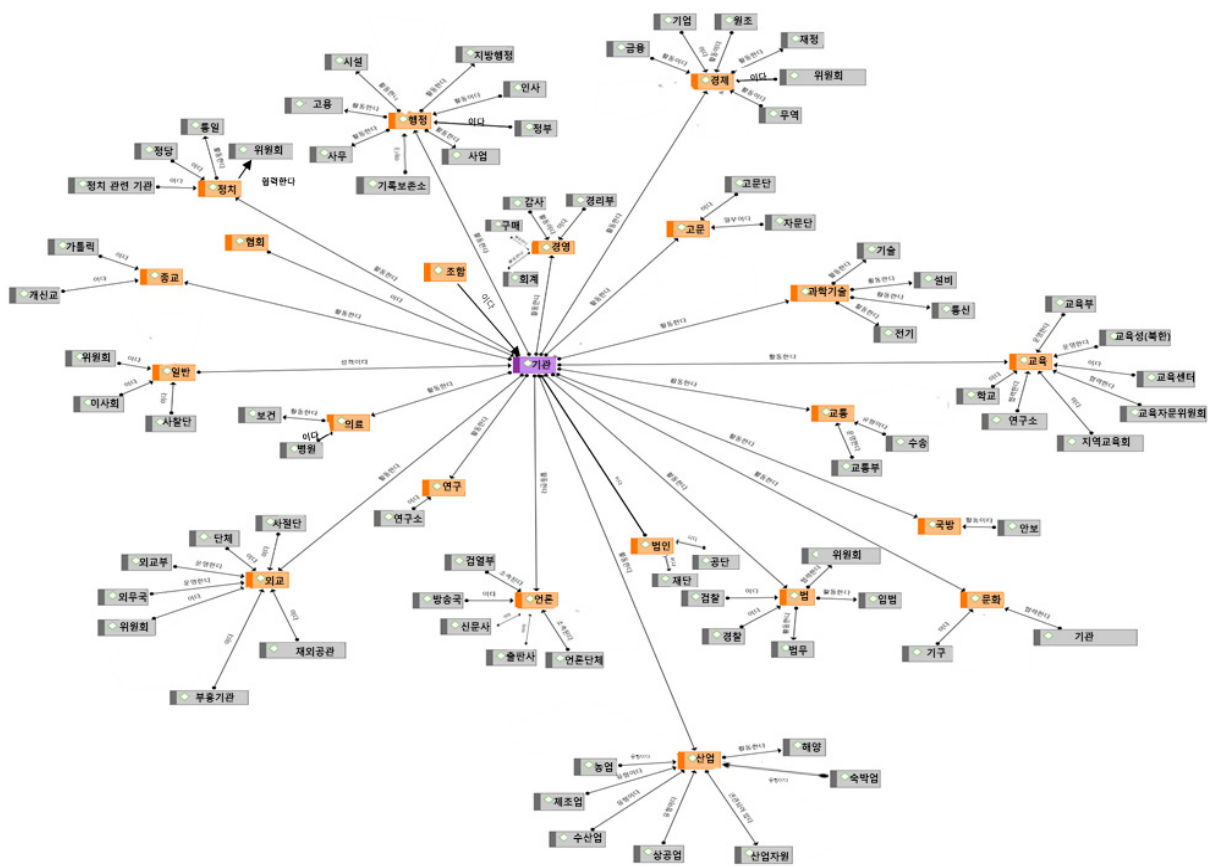
한국 관련 미국기록물 토픽맵은 기관, 군대, 기록물, 인물, 주제, 장소 및 지명 총 6가지 패킷별로 구축되었다. 각 패킷은 토픽맵에서 최상위 토픽으로 설정하였고 앞 단계에서 생성되었던 중간토픽은 토픽맵 상에서 하위 토픽

으로, 세부토픽은 토픽맵 상에서 어커런스로 생성하였다. 각 토픽별로 생성된 토픽맵은 아래와 같다. 제시된 토픽맵은 상위 토픽, 하위 토픽, 어커런스별로 색깔을 구분해 표현하였으므로 단위별로 제시된 결과를 효율적으로 확인할 수 있다.

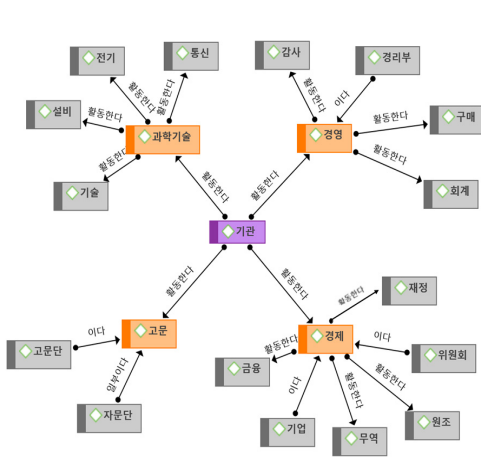
한국 관련 미국기록물의 패킷과 토픽 간 관계를 계층적으로 나타내기 위해 다양한 관계 속성 즉, 관계자를 활용하였는데 가장 많이 활용된 관계자는 ‘...의 일부이다’, ‘...로 구성되다’, ‘포함되다’, ‘...와 관련이 있다’, ‘...에 위치한다’, ‘...와 교류하다’, ‘원인이 되다’ 등이 기존에 온톨로지의 속성을 표현하기 위해 사용된 관계자이다. 이를 기반으로 패킷과 토픽간의 관계를 설명하되 자체적으로 추가 생성하여 사용한 관계자는 ‘생산된다’, ‘소비한다’, ‘운영한다’, ‘운행된다’, ‘조직된다’, ‘처벌된다’, ‘활동한다’, ‘활용한다’ 등이 있다. 추가 생성된 관계자는 패킷과 토픽, 토픽과 토픽이 내포한 주제와 내용을 분석하여 도출되었다. 각 패킷별 토픽맵과 패킷별 통합 토픽맵 내용은 아래와 같다.

4.3.1 기관 토픽맵

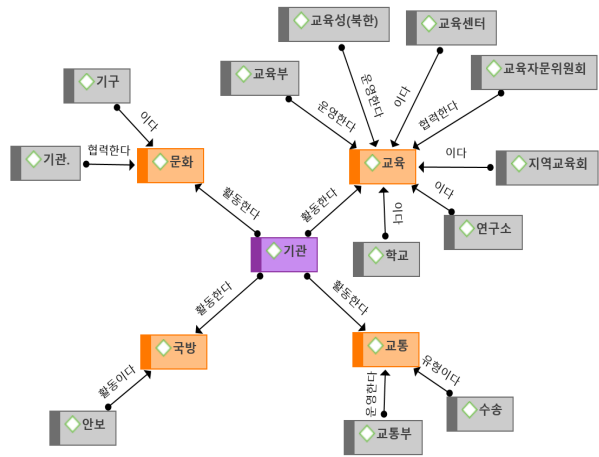
<그림 1>은 한국 관련 미국기록물의 기관 패킷을 토픽맵으로 시각화한 결과이다. 가장 중앙에 위치한 기관 토픽은 경영, 경제, 고문, 공단, 과학기술, 문화, 법, 언론, 의료 등 총 21개의 중간토픽과 감사, 구매, 금융, 무역, 원조, 병원, 재외공관, 위원회, 협회 등 74개의 세부토픽을 가진다. 기관 토픽맵 상에서 관계의 속성은 주로 ‘...이다’, ‘활동한다’, ‘협력한다’ 등으로 표현되었다. 기관패킷의 토픽맵의 세부 내용을 표현한 분할도는 <그림 2> ~ <그림 6>과 같다.



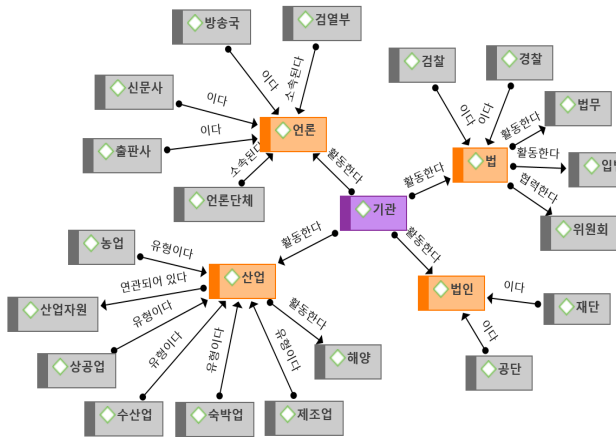
<그림 1> 기관 토픽맵



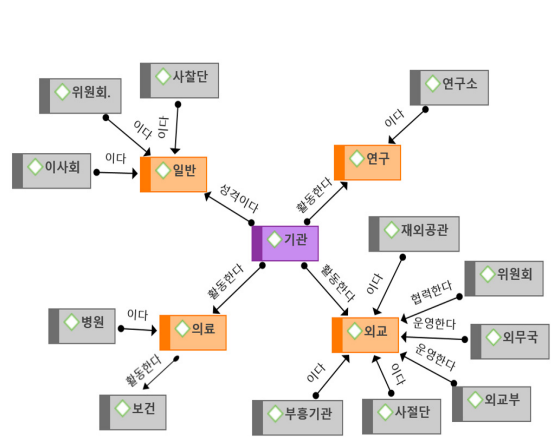
〈그림 2〉 기관 패시 토픽맵 분할도 1
- 경영, 경제, 고문, 과학기술



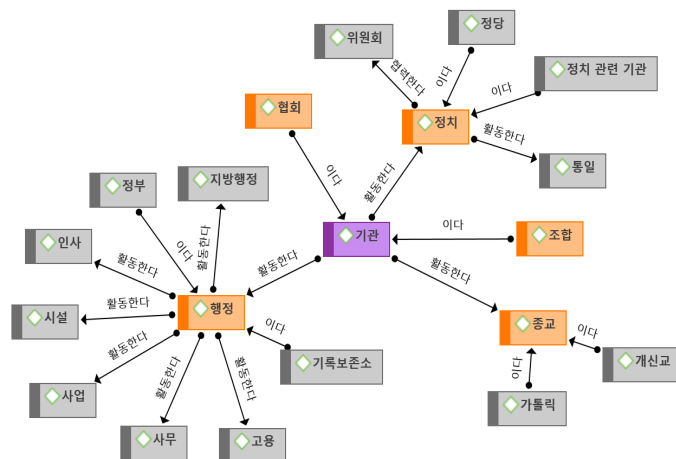
〈그림 3〉 기관 패시 토픽맵 분할도 2
- 교육, 교통, 국방, 문화



〈그림 4〉 기관 패시 토픽맵 분할도 3
- 법, 법인, 언론, 산업



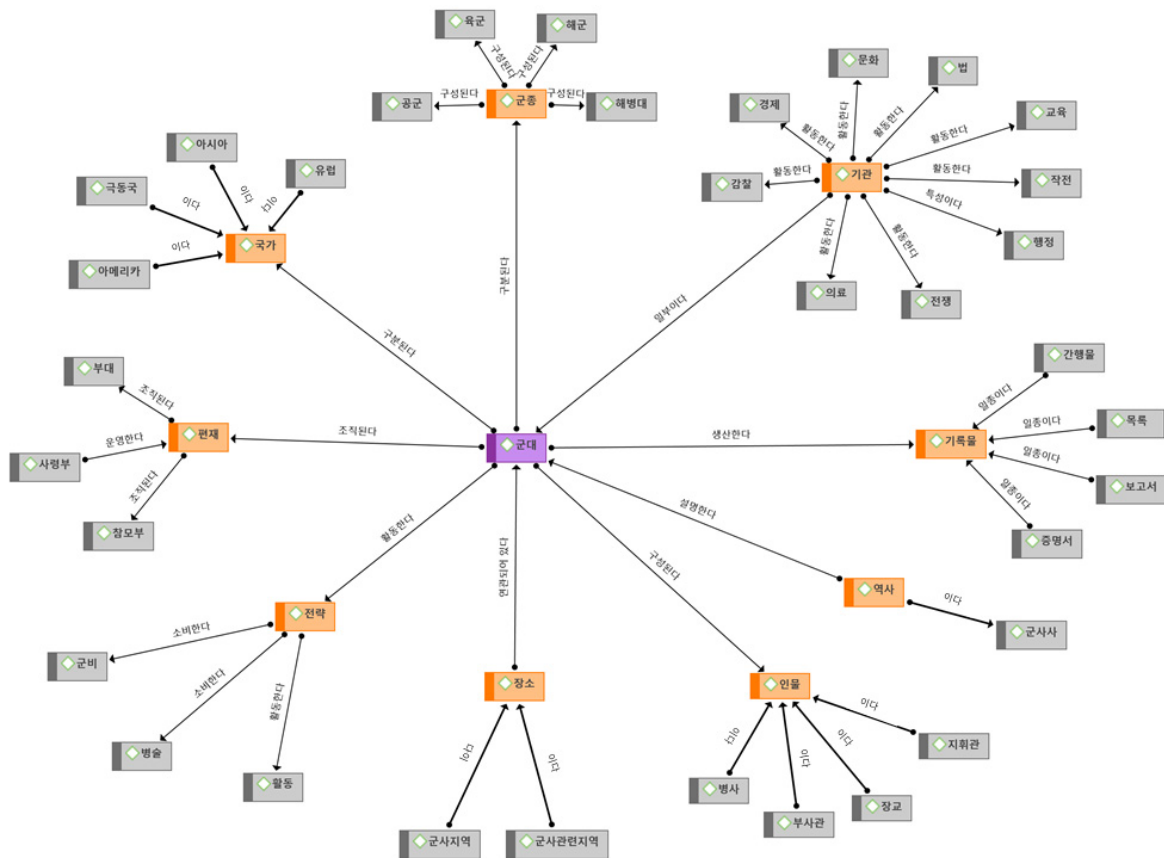
〈그림 5〉 기관 패시 토픽맵 분할도 4
- 연구, 외교, 의료, 일반



〈그림 6〉 기관 패시 토픽맵 분할도 5
- 정치, 조합, 종교, 행정, 협회

4.3.2 군대 토픽맵

<그림 7>은 한국 관련 미국기록물의 군대 패킷을 토픽맵으로 시각화한 결과이다. 가장 중앙에 위치한 군대 토픽은 국가, 군종, 기관, 인물, 전략, 편제 등 총 9개의 중간토픽과 아메리카, 육군, 해군, 감찰, 병사, 장교, 병술, 사령부 등 35개의 세부토픽을 가진다. 군대 토픽맵에서 관계의 속성은 주로 ‘...이다’, ‘구성된다’, ‘조직된다’, ‘활동한다’ 등으로 표현하여 토픽 간의 상호연관성을 시각화하였다.



<그림 7> 군대 토픽맵

4.3.3 기록물 토픽맵

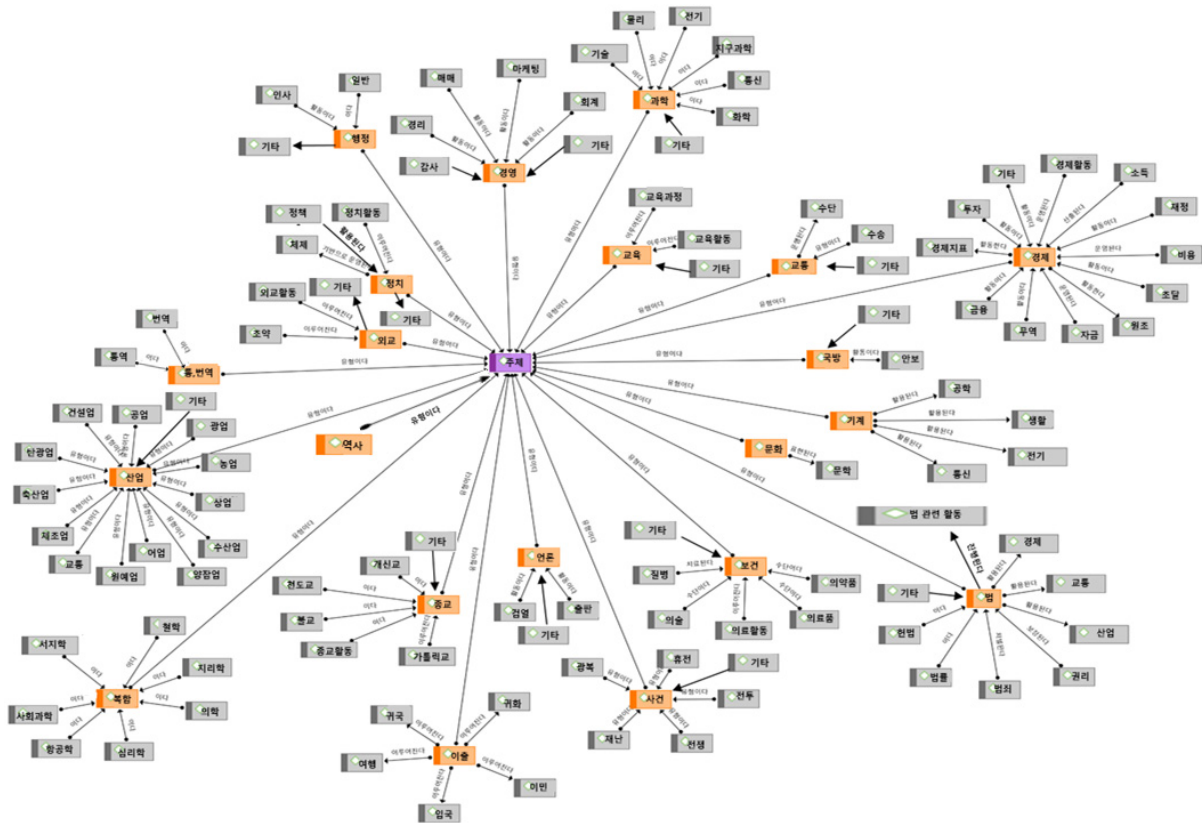
<그림 8>은 한국 관련 미국기록물의 기록물 패킷을 토픽맵으로 시각화 한 결과이다. 가장 중앙에 위치한 기록물 토픽은 간행물류, 기념물류, 보고서류, 비망록류, 영상 및 음성류 등 총 10개의 중간토픽과 신문, 훈장, 각서, 각종보고서, 서신, 사진 도면, 영상, 음성 등 46개의 세부토픽을 가진다. 기록물 토픽맵에서 관계의 속성은 ‘...이다’, ‘생산된다’, ‘유형이다’, ‘활용한다’ 등으로 표현하여 토픽 간 관계를 설명하였다.

4.3.4 인물 토픽맵

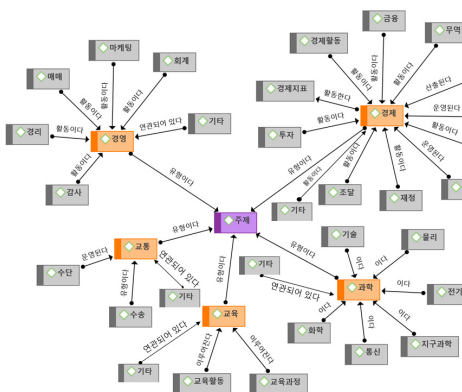
<그림 9>는 한국 관련 미국기록물의 인물 패킷을 토픽맵으로 시각화 한 결과이다. 가장 중앙에 위치한 인물 토픽은 국적, 이출민 직종 총 3개의 중간토픽과 미국인, 한국인, 민간인, 이민자, 고문, 국가원수 등 41개의 세부토픽으로 표현되었다. 기록물 토픽맵에서 관계의 속성은 ‘...이다’, ‘구분된다’ 등으로 표현되었다.

4.3.5 주제 토픽맵

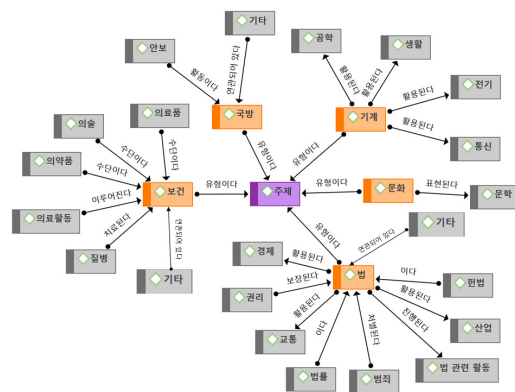
한국 관련 미국기록물의 주제 패시를 토픽맵으로 시각화한 것은 <그림 10>과 같다. 가장 중앙에 위치한 주제 토픽은 경영, 경제, 교통, 기계, 보건, 산업, 언론, 이출, 행정 등 총 21개의 중간토픽이며 감사, 무역, 재정, 조달, 전기, 교육과정, 광복, 통·번역 등 105개의 세부토픽으로 구조화되었다. 주제 토픽맵에서 관계의 속성은 ‘...이다’, ‘운영된다’, ‘유형이다’, ‘이루어진다’, ‘활동이다’ 등으로 표현되었다. 주제 패시의 토픽맵의 세부 내용을 표현한 분할도는 <그림 11>부터 <그림 14>까지와 같다.



<그림 10> 주제 토픽맵



<그림 11> 주제 패시 토픽맵 분할도 1
- 경영, 경제, 과학, 교육, 교통

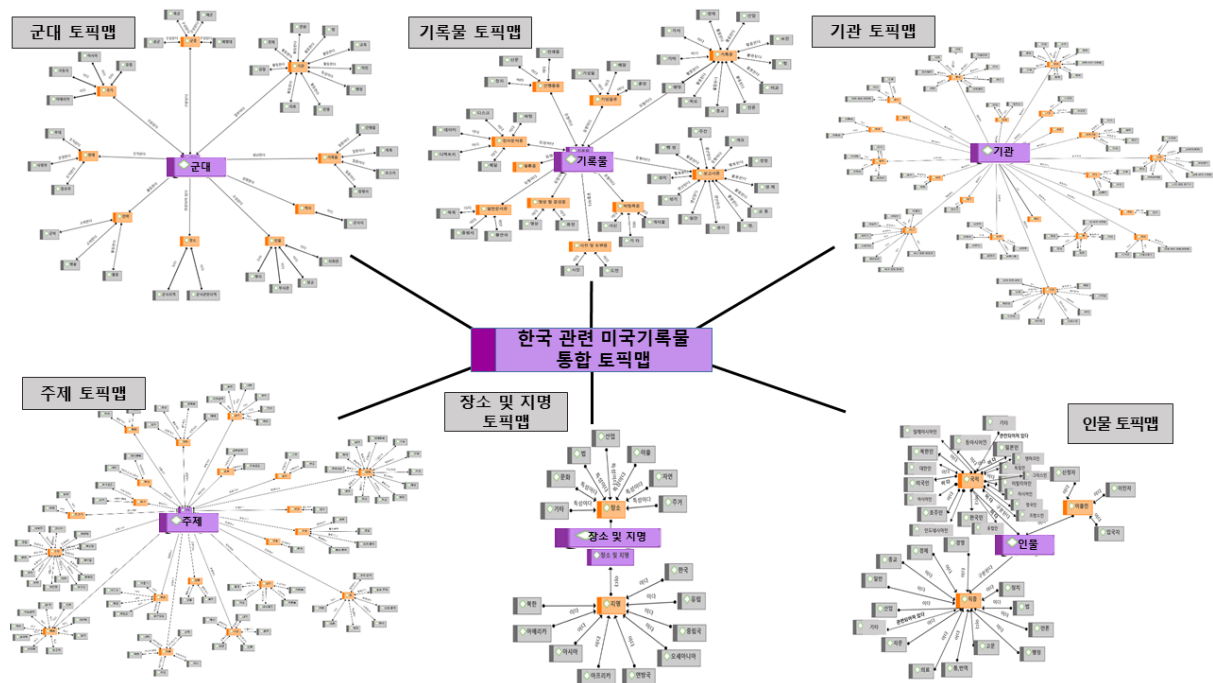


<그림 12> 주제 패시 토픽맵 분할도 2
- 국방, 기계, 문화, 법, 보건

충분히 도출되지 못하였다. 시대 패킷에는 총 4개의 범주가 도출되었는데 국사편찬위원회 내 한국사데이터베이스의 ‘시대별 열람’의 시대 구분인 통사, 고대, 고려시대, 조선시대, 대한제국, 일제강점기, 대한민국 중 ‘조선시대’, ‘대한제국’, ‘일제강점기’, ‘대한민국’을 차용하여 해당 시대별로 특정 연도를 범주화하였다. 따라서 총 4개의 하위 범주가 생성되어 토픽맵으로 시각화하기 보다는 검색 토픽맵으로 검색된 결과를 시대로 제한하는 필터로 적용한 것이 적절하다고 판단되어 토픽맵에서는 제외하였다.

통합토픽맵을 활용하여 한국 관련 미국기록물을 시각화한 후 이용자에게 제공하게 될 경우 나타날 개선 효과는 다음과 같다.

첫째, 이 연구에서 제시된 바와 같이 각 기관별로 제공하고 있는 기록물을 통합하여 토픽맵을 제공하게 될 경우 이용자는 각 기관별로 분산되어져 있는 한국 관련 해외기록물의 소재를 한 번에 파악할 수 있다. 이로 인해 한국 관련 해외기록물에 대한 이해가 높아지고 기록물에 대한 접근이 용이하게 되면서 기록을 활용한 다양한 연구와 활동이 활성화될 것이다.



<그림 16> 한국 관련 미국 기록물 통합 토픽맵

둘째, 한국 관련 미국 기록물 내용을 분석한 것을 기반으로 주요 주제어를 도출하여 이용자에게 토픽을 제공하기 때문에 기록물에 대한 이해를 높이고 기록물에 대한 접근을 쉽게 할 수 있을 것이다. 특히 한국 관련 미국 기록물과 같이 주제가 특화되어 있거나 쉽게 해당 내용을 연상할 수 없는 경우에는 기록의 내용을 파악할 수 있는 검색도구를 제공하는 것이 필요하다. 따라서 이 연구에서 제안한 바와 같이 기록물의 내용을 분석하여 주제 토픽으로 제공하여 훑어볼 수 있게 하게 될 경우 이용자는 검색대상이 되는 기록물의 내용을 쉽게 파악하여 검색에 적용할 수 있을 것이다.

셋째, 검색대상의 내용을 시각화함으로써 기존에 존재했던 키워드 검색 방식에서 나타난 검색의 한계점이 보완될 수 있을 것이다. 또한 토픽맵으로 발전시키지 않은 시대 패킷을 필터로 활용함으로써 나머지 6개의 토픽과 결합하여 검색결과를 조합하는 형태로 사용할 수 있다. 시대 패킷은 주제성을 크게 띠지 않는 연도를 특정 시대로

유추할 수 있도록 표현하기 위해 조선시대, 대한제국, 일제강점기, 대한민국으로 구분하여 도출하여 검색 필터로 활용하였다. 이로 인해 이용자는 특정 사건 또는 특정 연도로 검색하지 않고도 한국 관련 미국기록물에 접근하는 것이 용이해진다.

5. 결론

한국 관련 미국기록물은 국내 다수 기관에 의해 수집, 관리, 서비스되고 있지만 이용자는 각 기관 웹사이트 등을 직접 접속해 검색한 결과를 각각 비교해 보아야 하는 상황으로 한국 관련 해외기록물은 어느 정도의 양이 어떤 내용의 기록물이 존재하는지에 대한 일차적인 현황 파악조차 어렵고, 검색 방식 면에서도 많은 제한이 있다.

이러한 문제점을 개선시키고자 이 연구에서는 이용자에게 한국 관련 미국기록물의 통합 검색을 지원하기 위해 한국 관련 미국기록물을 웹으로 제공하는 국내 주요 기관들을 중심으로 하여 기록물에 대한 패킷 분석을 진행하여 하위 토픽을 도출하였고, 이를 시각화하여 토픽맵을 생성하였다. 연구에서 수행된 주요 내용은 다음과 같다.

첫째, 한국 관련 해외기록물의 기록물 시리즈 단위의 데이터를 분석하여 도출된 패킷은 기관, 군대, 기록물, 인물, 주제, 장소 및 지명, 시대 총 7개이다. 한국 관련 미국기록물은 다양한 전쟁이 발발했을 때 또는 그 전후의 발생한 다양한 사건과 관련된 기록이 전체 수집된 기록물의 주요 영역을 차지하였다. 이러한 특성을 토대로 한국 관련 미국기록물을 표현하는 주요 패킷으로 군대를 독립적으로 설정하게 되었다. 군대는 다른 패킷에 포함될 수 있는 유사성이 존재하지만 일반 기관이나 단체에 동일하게 분류하는 것에는 군대만의 조직적인 특색이 강하고, 양적인 측면에서도 상당한 기록의 양을 보유하고 있어 기관 및 단체, 주제 패킷과 분리하여 패킷을 도출하는 것이 의미가 있는 것으로 분석되었다.

둘째, 기관 패킷 분석 결과 기관은 일정 규모를 가진 조직이 행하는 업무와 관련된 기능에 의해 기록물이 산출되었고, 해당하는 기록물을 살펴본 결과 한국 관련 미국기록물이 생산 당시 전쟁 등을 겪는 과정에서 각종 원조나 지원에 해당하는 기록물이 상당수를 차지한 점에서 주목할 만한 사항이다.

셋째, 인물의 유형 패킷에서는 대분류로 국적, 이출민, 직종이 도출되었는데 이 중 이출민은 한국 관련 미국 기록물에서 특징적으로 나타난 하위분류이다. 이출민의 경우 입국자, 출국자, 이민자 등 다른 인물 유형에 해당될 수 없는 특성으로 독립적인 패킷으로 신규 생성하게 되었다.

넷째, 통합 토픽맵으로 분산된 자료를 통합하여 검색할 수 있게 하였고, 내용을 직관적으로 파악할 수 있게 시각화하였다. 또한 각 패킷별로 기록물의 양이나 토픽은 균등하지 않으나 이용자는 수집된 한국 관련 미국기록물에 해당되는 내용과 주제를 전체적으로 훑어볼 수 있게 하였다.

이 연구에서 제시한 주요 패킷과 하위토픽은 한국 관련 미국기록물을 통합하여 서비스하게 될 경우 디렉토리 서비스의 기반이 될 수 있으며 통합 검색의 기능 요소로서도 활용될 수 있다. 이에 이 연구에서 제시한 패킷과 토픽을 기반으로 검색에 적용하여 검색의 효과성을 평가할 수 있는 후속연구가 필요하다. 또한 이 연구의 한계점은 다양한 국가에서 생산한 한국 관련 해외기록물 중 한국 관련 미국기록물에 대한 분석만을 실시하였으므로 한국 관련 미국기록물 외에 다른 주요 국가에 대한 후속연구가 수행되어야 할 것이다.

참고문헌

- 고유미 (2005). 토픽맵 기반의 특허정보 서비스를 위한 시스템 구축에 관한 연구. 석사학위논문, 숙명여자대학교대학원 문헌정보학과.
- 김나정 (2016). 한국 관련 해외 수집 기록물의 이용자 서비스 개선방안 연구. 석사학위논문, 한국외국어대학교 대학원 정보·기록관리학과.
- 김성겸 (2005). 해외소재 한국관련 역사기록의 체계적인 관리방안 연구. 석사학위논문, 명지대학교 기록관리대학원 기록관리학과.
- 김소영, 임순범, 우성호, 최윤철 (2006). 토픽맵을 이용한 3D 건축물의 구조모델링 기법 연구. 한국정보처리학회 학술대회논문집, 13(2), 223-226.
- 마정춘 (2006). 해외소재 한국관련 역사기록물 현황과 정리에 관한 연구: 6.25전쟁기 미 육군 역사기록물을 중심으로. 석사학위논문, 중앙대학교 대학원 기록관리학과.
- 박도영 (2020). 유네스코 소장 한국 관련 수집 기록물의 패킷 분류 체계 연구: 1950...1979년 기록을 중심으로. 석사학위논문, 숙명여자대학교 대학원 문헌정보학과 기록관리학과.
- 박은봉 (2008). 한국관련 해외기록의 공동활용 현황과 과제. 한국기록관리학회 학술발표논문집 2008, 189-200.
- 신지유 (2016). 대학 학사행정 전자기록물을 위한 토픽맵 기반 검색 프레임워크 설계. 석사학위논문, 동의대학교 대학원 문헌정보·사학과.
- 심지영 (2014). 패킷분석 기법을 적용한 방송자료의 내용 구조화에 관한 연구: 시사보도 뉴스 프로그램을 대상으로. 정보관리학회지, 31(3), 313-329. <https://doi.org/10.3743/KOSIM.2014.31.3.313>
- 이상철 (2012). 스마트폰에서 활용 가능한 온라인 사전의 사용성 평가에 관한 연구: 토픽맵 활용을 중심으로. 석사학위논문, 충남대학교 대학원 문헌정보학과 문헌정보학전공.
- 이승휘 (2006). 해외소재 한국관련 수집자료의 정보공유를 위한 활용 방안: 2006년 국립중앙도서관 열린정책세미나 자료집, 2, 25-46.
- 정경희, 김판구 (2006). 토픽맵 기반 개인별 학습 콘텐츠 탐색 네비게이터 구조 설계. 한국정보처리학회 학술대회논문집, 13(2), 23-26.
- 조혜경 (2008). 일제강점기 해외소재 한국관련 기록의 통합활용에 관한 연구: 컬렉션 기술을 중심으로. 석사학위논문, 명지대학교 기록과학대학원 기록관리학과 기록관리학전공.
- La Barre, K. (2010). Facet analysis. *Annual Review of Information Science and Technology*, 44(1), 2010, 243-284. <https://doi.org/10.1002/aris.2010.1440440113>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Jeung, Kyounghui & Kim, Pankoo (2006). Design of the personalized searching navigator of learning contents based on the topic maps. *Proceedings of the Korea Information Processing Society Conference*, 13(2), 23-26.
- Jo, Hye Kyoung (2008). A Study for Integrated Use of Overseas Record on Korea During the Japanese Rule - Focus on Collection and Description. Master's Thesis, The Graduate School of Archival Science, Myongji University.
- Kim, Na Jung (2016). A Study on Improving Archival Information Services of Overseas Korea-related Manuscripts. Master's Thesis, Department of Information and Archival Science, The Graduate School of Hankuk University of Foreign Studies.
- Kim, Soyong, Lim, Soon-bum, Woo, Sungho, & Choy, Yoonchul (2006). Structure modeling techniques for the 3D architecture using topic maps. *Proceedings of the Korea Information Processing Society Conference*, 13(2), 223-226.
- Kim, Sung Kyum (2005). Management and Retention for Collection of Korea Historical Data from International Source. Master's Thesis, The Graduate School of Archival Science, Myongji University.
- Ko, You Mi (2005). A Study on the Implementation of Patent Information Service System Based on the Antibody Subject

- Using XTM. Master's Thesis, Department of Library & Information Science Graduate School of Sookmyung Women's University.
- Lee, Sang-Chul (2012). Usability Testing for Two On-line Dictionaries on Smart-phone Device - for Using Visualizing Topic-map. Master's Thesis, Department of Library & Information Science, Graduate School Chungnam National University.
- Lee, Seung-Hwi (2006). Sharing of Overseas Record on Korea. Proceeding of Open Policy Seminar. National Library of Korea. 2, 25-46.
- Ma Jeong Chun (2006). Research on Korea related archive located abroad Status and Arrangement. Master's Thesis, Dept. of Records and Archives Management, Graduate School of Chung-Ang University.
- Park, Do Young (2020). A Study on the Faceted Classification Scheme for the Korea-related Records(1950~1979) Collected from UNESCO Archive. Master's Thesis, Department of Library and Information Science, The Graduate School Sookmyung Women's University.
- Park, Eun Bong (2008). Issues and current status of integrated use for overseas record on Korea. Proceedings of Korean Society of Archives & Records Management 2008, 189-200.
- Shim, Jiyoung (2014). A faceted classification analysis of TV content: using news and current affairs programs. Journal of the Korean Society for Information Management, 31(3), 313-329. <https://doi.org/10.3743/KOSIM.2014.31.3.313>
- Shin, Ji-yu (2016). The Design of a Topic Maps-based Retrieval Framework for Electronic Records Related to the Academic Administration of Universities. Master's Thesis, Department of Library and Information Science & History, Graduate School, Dong Eui University.