신문기사정보 패싯 택소노미 구축 방안*

- 예술 분야를 중심으로 -

Plan of Constructing Facet Taxanomies of Information on News Articles

- Focused on the area of Arts -

장 인 호(Inho Chang)**

〈목 차〉

초 록

신문기사를 주제 분야별로 나누고, 분야 내에서 각각의 범주들은 기본 패싯과 결합하는 패싯 택소노미 모형을 개발하였으며 구축 방안을 제시하고 패싯 택소노미를 구축하는 연구를 수행하였다. 패싯 택소노미는 신문기사를 주제 분야(정치, 경제 분야 등)로 나누고 범주(정치 분야의 경우, 정치일반, 행정, 사법 등) 및 하위 범주를 기본 패싯과 각각 결합한다. 하위 범주는 더욱 하위 구분할 수 있다. 택소노미는 범주 간의 계층 관계를 가질 수 있으며, 범주-패싯은 예를 들어, "예술"에 대해 '사람', '행위', '행사', '시간', '장소' 등과 결합한다. 그리고 예술의 하위 범주인 '미술', '음악', '무용' 등은 '예술'과 계층 관계를 이루어 추론과 브라우징에 활용할 수 있도록 구성하였다. 또한, 범주-패싯 결합은 기본 패싯순으로 계층 구조를 갖는다. 한편, 시험용 어휘 구축은 '예술 분야'를 대상으로 용어 145어를 본 연구에서 다루는 모든 구성요소를 포함하는 패싯 택소노미를 구축하고, 디스플레이를 예시하였다.

키워드: 신문기사정보, 기본 패싯, 패싯 택소노미, 디스플레이, 시험용 어휘 구축

ABSTRACT

Information on newspaper articles were categorized into different topics, and each categories within different topics were developed into a faceted taxonomies model which was combined with fundamental facets. After suggesting the plan to construct such a model, the research of actual faceted taxonomies were conducted. Faceted taxonomies divide information on news articles into different topics(such as politics, economies and others) and combine fundamental facets with categories(for example, politics can be sub-classified into general politics, administration, legal system, and others) and sub-categories. Each sub-categories can be further subdivided. In taxanomies, categories can have hierarchical relationships. Categories-Facets, for example, can be utilized to combine "arts" with "people", "action", "event", "time", "place" and others. And Sub-category of the classification of "arts" such as "art," "music," "dance" form hierarchical relationships with "arts" and, in turn, can be used for browsing and further inferences. Furthermore, combining category and facets results in hierarchical structure in order of fundamental facets. As for the pilot vocabulary construction, faceted taxonomies of 145 words from news paper articles on the topic of "arts" were constructed using all construction elements covered in this study.

Keywords: News articles, Fundamental facets, Faceted taxonomies, Display, Pilot vocabulary construction

* 이 논문은 2018년도 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2018S1A5A8030395)
** 대진대학교 공공인재대학 문헌정보학과 조교수(hoinchang@gmail.com)
•논문접수: 2019년 11월 20일 •최초심사: 2019년 11월 27일 •게재확정: 2019년 12월 18일
•한국도서관·정보학회지 50(4), 381-403, 2019. [http://dx.doi.org/10.16981/kliss.50.201912.381]

Ⅰ. 서론

1. 연구의 필요성 및 목적

신문은 불특정 다수를 대상으로 메시지를 전달하는 커뮤니케이션의 한 형태이다. 오늘날에는 신문의 내용이 종이뿐만 아니라 웹이나 모바일을 통해서도 전달되는 매체로서 매일 또는 주간, 월간 등 정기적으로 발행되고 있다. 또한, 내용이 시의성을 띠며, 텍스트 중심의 매체이다. 신문은 많은 정보를 전할 수 있으며, 기록성이 뛰어나고, 이용자들이 쉽고 접할 수 있는 편리한 매체이다(임영호 2013, 27). 신문은 배포지역(전국지, 지역지)이나 다루는 정보의 범위(종합지, 경제지, 스포츠지, 기타 산업 전문지), 대상 독자층에 따른 분류(대중지, 고급지) 등 매우 다양한 방법으로 분류할 수 있다(임영호 2013, 49). 이들이 쏟아내는 신문기사의 정보량은 가히 폭발적이며, 신문기사정보는 뉴스뿐만 아니라 문학작품, 공포된 법령, 붓글씨 등 다양한 정보를 전달하는 매체로서 우리의 실생활에 밀접한 정보원이다.

신문기사정보를 취급하는 정보관리는 대상으로 하는 정보 분야를 결정해야 하고, 정보량을 고려해야 하며, 해당 정보관리 시스템에 대응해야 한다(広木守雄 1981, 374). 신문기사는 종별 전체적으로 볼 때 삼라만상의 주제를 다루고 있으며, 정보의 깊이도 매우 다양하여 관리하기가 쉽지 않고(장인호 2013, 490), 용어의 사용도 정착되기 전까지는 표준화되지 않은 채 사용되며, 매체별로도 사용하는 용어가 서로 다르다. 이와 같이 신문기사정보는 다루는 범위가 넓고, 정보량이 많으며, 표준화되지 않는 용어가 여러 가지로 사용되기 때문에 제어 어휘(시소러스나 분류 체계 등)의 필요성이 그만큼 크다. 이에 따라 어떠한 다른 영역보다도 강력한 표준화된 제어 어휘의 구축 및 지침이 필요하며, 이에 대한 이론 및 실천의 연구가 필요하다.

신문업계는 이미 NewsML(IPTC Homepage)이라고 하는 메타데이터에 대한 표준이 적용되고 있고, 신문기사정보의 시소러스 및 분류 체계, 동의어 리스트 등이 각각의 매체별, 각 신문사별로 독자적인 도구가 사용되고 있으나 기본 패싯 기반의 분류 체계는 갖추고 있지 않다.

최근 빅 데이터 분석 시스템이나 뉴스 큐레이션 및 기사의 자동 요약 등에 관심이 폭발하고 있으며, 이들의 시스템과 신문기사를 활용하여 부가서비스를 창출하여 서비스를 확대하려는 시도가 있다. 그러나 많은 언론사나 부가서비스 업체들은 영세하거나 신규로 투자하기 어려운 상황에 있는 경우가 많다. 이에 제어 어휘 자원을 공유 및 재이용하려는 이들은 인프라로서의 제어 어휘를 크게 요구하고 있다.

본 연구는 그러한 필요성에 따라 신문기사정보의 패싯 택소노미 모형을 개발하고 구축 방안을 제시하고자 하였으며, '예술 분야'를 대상으로 실제의 구축 사례를 나타내고자 한다. 즉, 모형을 기반으로 주제 범주, 택소노미, 패싯을 구성요소로 하는 패싯 택소노미 구성 및 디스플레이 방안을 제시하고자한다. 또한, 예술 분야의 시험용 어휘를 구축하고 디스플레이를 예

시하였다. 본 연구는 자동 범주화나 개인화에 활용하는 등 부가서비스의 어휘자원의 인프라로 활용 될 수 있을 것으로 기대한다.

1.2 연구의 방법 및 절차

본 연구는 신문기사 정보관리를 위한 기본 패싯을 기반으로 하는 계층 구조를 가진 전조합 체계의 택소노미 구축 방안과 실제의 구축 사례를 제시하는 것이다. 이를 위해 ISO 25964-1의 제11절 패싯 분석과 ISO 25964-2의 제19절 택소노미를 중심으로 하는 국제표준, 그리고 MoTif의 시소러스 구축 방안(Ryan, 2014)과 Hedden의 택소노미 구축 방안(Hedden, 2016), Broughton의 패싯 기반 시소러스 구축 방안(Broughton, 2006) 등의 문헌조사를 통하여 패싯 택소노미의 새로운 모형을 개발하고 구축 방안과 시험용 어휘를 구축하였다.

이를 위해 다음과 같은 방법과 절차에 따라 연구를 수행하였다.

첫째, 문헌조사를 실시하였다. 시소러스 구축 기준의 국제표준(ISO 25964-1, -2) 및 연구 결과물(단행본, 논문 등)을 수집하고, 아울러 신문기사정보의 기존의 어휘집(분류 체계, 시소러스, 신문기사 장르 등), 그리고 기존의 기본 패싯을 조사・수집하였다.

둘째, 조사・수집된 문헌(제어 어휘 포함)을 분석하였다.

셋째, 신문기사정보를 위한 패싯 택소노미의 모형을 개발하였다.

넷째, 패싯 택소노미의 구축 방안을 제시하였다. 이 절차에서는 신문기사정보가 취급하는 주제 범주를 설정하고, 기본 패싯, 기호법, 디스플레이 방안 등을 제시하였다.

다섯째, 신문기사정보의 예술 분야에 대한 패싯 택소노미를 구축하였다. 이 절차에서는 기본 패싯을 바탕으로 패싯 분석에 의한 어휘를 조직화하고, 디스플레이를 나타내었다.

1.3 용어의 정의

택소노미라고 하는 용어는 매우 다양한 유형들의 어휘에서 이 이름을 가질 수 있기 때문에 너무 넓은 의미로 사용되거나 남용된다(ISO 25964-2, 60). 본 연구에서는 범주 간의 계층 구조를 가지며 범주-기본 패싯의 형태로 결합하여 이것 또한 기본 패싯별로 계층 구조를 가지는 제어 어휘를 패싯 택소노미라고 정의하였다.

Ⅱ. 이론적 배경

1. 택소노미

전형적인 택소노미는 계층 어휘로서 표현되고, 네트워크화된 환경에서 콘텐츠의 모든 유형의 분류나 범주화, 조직화, 브라우징, 내비게이션, 탐색 그리고 필터링에 사용된다. 일반적인 사용 사례는 예를 들어, 특히 웹 사이트, 인트라넷, 포털, 위키와 같은 전자 자원의 광범위한 집합을 통해 계층적 조직화와 브라우징에 의해, 내비게이션을 지원하는 것이다. 택소노미는 종종 웹 사이트의 메뉴를 제공하는데 사용된다. 탐색 성능을 가진 내비게이션의 기능을 보완하기 위해, 택소노미는 엔트리 텀(도입어)들로서, 은밀하게 운영하는 동의어들을 그리고 계층 구조에서 관련된 범주들 사이의 “도보라(See also)” 참조를 포함할 수 있다(ISO 25964-2, 59).

택소노미를 생각할 때, 계층 분류 체계가 일반적으로 떠오르는 개념일 것이다. 그래서 계층 구조로 구성된 택소노미는 특별히 계층 택소노미라고 한다. 이것은 시소러스에 가까운 개념이다. 계층 택소노미는 최상위어가 아닌 각 용어가 지정된 상위 및 최하위 용어가 아닌 경우 하나 이상의 하위어에 연결되는 일종의 제어 어휘이다. 계층 택소노미의 고전적인 예는 계층적 하향식 구조인 계, 문, 강, 목, 과, 속 및 종과 같은 생물 유기체의 린네식 택소노미이다. 계층 분류는 지역, 국가, 지방 및 도시와 마찬가지로 지형 공간 분류에서도 일반적이다. 계층 분류는 주로 일반 사물이나 개념에 사용되는 경향이 있지만 장소 이름, 제품 이름, 정부 기관 이름 또는 회사 부서 이름과 같이 자연스럽게 계층 구조에 속하는 고유 명사에도 사용된다(Hedden 2016, 5).

어떤 상황에서는 패싯 택소노미가 계층 택소노미와 다르게 간주될 수 있지만 패싯 택소노미는 상위 수준에서 특정 방식으로 구현되고 사용되는 계층 택소노미의 변형이다. 다른 계층 택소노미와 마찬가지로 패싯 택소노미는 최종 이용자가 위에서 아래로 시작하여 브라우징하기 위한 것이다. 패싯은 계층 구조와 유사하며 패싯 이름은 계층 구조에서 최상위 용어와 같다(Hedden 2016, 6).

택소노미는 비구조화된 지식을 공유하고 조직화하고 획득하기 위해 지식 자산을 분류할 수 있도록 하거나(Cheung et. al., 2005, 이종영 등, 2015, 이정민, 2016) 콘텐츠가 가진 정보를 효과적으로 색인하고 주석할 수 있다(최지수, 2014). 또한, 소프트웨어 봇을 연구하고, 설계하고, 분류하는 능력을 개선하기 위해 사용되거나(Lebeuf et. al., 2019), 시맨틱 웹 기술 등에서 데이터의 상호 연결 등을 위해 사용되기도 한다(Zong et. al., 2017).

Cheung 등(2005)은 비구조화된 지식을 공유하고 조직화하고 수집하기 위해 다차원 택소노미를 구성하여 지식 자산을 분류할 수 있도록 인공지능(AI)과 자연어 프로세스(NLP) 기술을 사용하여 다차원 택소노미, 시소러스 모델, 자동 분류 메커니즘, 지능 검색, 택소노미의 자가 유지(self-maintenance) 등 5가지 구성 요소를 기반으로 하는 시스템을 구현하였다.

이종영 등(2015)은 재난유형별 택소노미 분류 체계를 이용해서 관련 문서를 분류해내는 방법과 대규모의 문서집합으로부터 다항분포를 이루는 토픽을 자동으로 추출할 수 있는 LDA (Latent Dirichlet Allocation) 모델을 이용하여 다량의 뉴스 데이터에서 재해 관련 토

픽을 추출하고 주제 클러스터 레이블링 기법을 제시하였다. 재난유형별 택소노미는 법률 제12943호 재난 및 안전관리 기본법에 따라 크게 20개의 자연재난과 35개의 인적・사회적재난으로 구분하여 취급하였다.

이정민(2016)은 무용학의 지적 구조 분석 연구에서 주요 지식 정보로 나타난 단어를 가지고 사전을 구축하여 무용학 지식 분류 체계를 마련하였다. 그 결과 학문/연구, 문화예술, 춤/무용, 인물, 몸/신체, 표현 요소. 기관단체/장소, 시대/지역의 대 항목 8개, 중 항목 62개, 소 항목 3,103개 등 총 3,173의 키워드의 계층형 사전 형태로 무용학 지식을 분류하였고, 무용학 지식 분류 체계인 택소노미를 기반으로 정보 간의 관계를 의미망으로 시각화하였다.

최지수(2014)는 의학 다큐멘터리인 <생로병사의 비밀>을 중심으로 하여 해당 콘텐츠가 가진 정보를 효과적으로 인덱싱하고 주석할 수 있는 주석 시스템을 설계 및 구현하였다. 제안한 주석 모델에서는 이용자들의 정보 접근 의도와 콘텐츠에 특화된 정보를 고려하여 의학적인 택소노미를 정의하고, 이런 택소노미를 기반으로 한 주석을 하였다. 즉, 비디오 문장은 등장인물, 주제, 화면 내용에 맞추어서 여러 개의 비디오 문장 유형으로 분류하여 주석을 할 수 있도록 했다. 또한, 비디오 문단은 질병명 및 대상/개념 등에 해당하는 비디오 문단 유형과 키워드 주석을 할 수 있도록 하였다. 그리고 비디오 문단 유형 및 키워드 구조는 주석자에 의해 계속 추가/확장할 수 있도록 하였다.

Lebeuf 등(2019)은 소프트웨어 봇을 연구하고, 설계하고, 분류하는 능력을 개선하기 위해 소프트웨어 봇의 관측 가능한 특성과 행동을 논하는 제어 어휘를 제시하였다. 이것은 소프트웨어 봇에 대한 완전한 분류를 만들기 위해 결합할 수 있는 복수의 독립적인 관점(패싯)과 주체(소프트웨어 봇)를 분류할 수 있도록 하는 패싯 택소노미를 설계한 것이다.

Zong 등(Zong, et. al. 2017)은 링크드 데이터라고 알려진 데이터의 상호 연결에서 데이터의 공유와 재사용 가능성을 촉진하기 위해 서로 다른 데이터 출처와 다양한 도메인에서 온 이질적인 개체(entities)를 조직하는데 개념 택소노미를 사용하였다. 이 연구는 특정한 관점에서 분류하기 위한 택소노미를 구축하는 대신 개체 특성에 기초하여 다양한 관점에서 인스턴스를 분류하는 패싯 택소노미를 제안하였다.

2. 기본 패싯

패싯 분석은 주제의 패싯, 또는 주제 분야의 개념에 대해 측면으로 분리하여 구분의 광범위한 원칙을 적용함으로써 작동한다. 실용적인 의미에서, 이것은 주제의 용어가 일련의 범주로 정렬된다는 것을 의미한다. 일부 주제의 경우 특수한 추가 범주가 필요하며 해당 주제에 따라 만들어질 수 있다. 랑가나단은 다섯 가지 기본 범주가 있다고 이론화했다. 이미 인식된 '시간'과 '공간'(또는 장소) 범주, 그리고 '에너지' 범주라고 부르는 행동이나 활동을 표현한 개념, 물질(matter)이라고 불렀던 물리적인 물질(physical substances)인 개념들, '성격'이라고

불렀던 주제의 본질에 대한 범주를 말한다. 이 다섯 가지 범주는 복합 주제를 다룰 때 일반적으로 결합되는 순서를 반영하여 기본 패싯 공식인 Personality—Matter—Energy—Space—Time(PMEST)를 만들었다(Broughton 2006, 108). 이러한 범주는 대다수 주제에 대해 잘 작동하는 것으로 밝혀졌지만 모든 주제에 모두 적용되는 것은 아니다. 공통적인 특징이 있는 용어 그룹을 식별할 수 있는 경우 추가될 수 있다(Broughton 2006, 109). 그러나 최상위 수준에서는 대상, 행위, 물질, 에이전트, 시간, 장소 등의 기본 범주를 사용하는 것이 일반적이다(ISO 25964-1 2011, 68).

한편, 행위의 패싯을 하위 패싯으로 분석할 때 국제표준에서는 행위자가 어떤 대상에 영향을 받지 않을 때의 "열화" 또는 "숙성"과 같은 자동사적인 과정(process)과 행위자의 영향을 받는 "절단" 또는 "복구"와 같은 타동사적인 조작(operation)으로 나눌 수 있다고 기술하고 있다(ISO 25964-1 2011, 68).

패싯은 상위 온톨로지와 매우 유사한 면을 가지고 있다. 다만, 온톨로지에서는 클래스가 개체(individual)를 구성요소로 가지지만, 주제를 표현하는 시소러스에서는 그 하위 전개가 다르다. 기존의 국제규격(ISO 2788)에서는 언급이 없었으나, 새로운 국제규격(ISO 25964)에서는 기술하고 있다(홍기철 2017, 349).

기본 패싯은 시소러스 또는 분류 체계의 구축 시에 용어를 조직하는 데 사용될 수 있을 뿐만 아니라, 또한 포함된 용어들을 정의하는 데 제어 어휘에 추가되어진다. 예를 들어, 기본 패싯은 통합 의료 언어 시스템(UML)의 의미 네트워크에서 개념 카테고리 또는 유형으로서 사용된다(Aitchison et. al. 2000, 71).

패싯 분류의 현저한 특징은 동종의, 상호 배타적인 그룹을 생산하기 위해, 한 번에 단 하나만의 구분 원칙(또는 원리)을 사용하여, 범주로 또는 패싯으로 용어를 구분하는 것이다. 또 다른 특징은, 제한된 수의 범주 또는 패싯의 승인이다. 그것은 모든 개념에 표현될 수 있고 모든 주제 분야의 기초가 될 수 있다. 예를 들어, 개념 '페인팅'은 기본 패싯 행위의 구성원으로서 간주될 수 있고 '페인트'는 기본 패싯 재료의 표현(manifestation)으로서 간주될 수 있다(Aitchison et. al. 2000, 70).

랑가나단의 PMEST 기본 패싯에 기반하여, 심지영(2014)은 방송자료에 대한 지적 접근을 위한 뉴스 및 시사보도 내용 기술을 패싯 분석 기법에 적용하였다. 보도 장르의 내용적 요소와 형식적 구조를 반영하여 패싯의 구성요소를 추출하고, 기본 패싯을 보도 장르에 적합한 '누가', '언제', '무엇을', '어디서', '어떻게' 등을 생성하였다.

패싯 분석 기법을 활용하여 시소러스의 상하 개념에 적용하는 연구로 홍기철(2017)은 정보검색을 위한 건설 분야의 시소러스의 구축을 상정하여 패싯 분석 기법을 이용한 패싯 유형을 정립하고, 패싯 분석 기법에 따른 시소러스의 구축 방안을 제시하였다. 기본 패싯은 최상위 10개의 패싯(주체 및 수동체, 추상물, 인공물, 재료, 부품/구성요소, 속성, 공종, 매체, 프로세스, 시간, 공간)과 하위 20개의 패싯을 설립하고, 국제표준(ISO 25964-1)이 제시하고 있

는 구축 절차에 따른 방안을 제시하였다.

정연경(2013)은 한식 관련 정보를 개념화하고 조직할 수 있는 패싯 구조를 개발하였다. 공통 속성을 도출하기 위해 수집된 용어를 범주화하고, 기본 및 하위 패싯을 정의하였으며, 기본 열거 순서와 기호화, 패싯 간에 계층 구조를 결정하였다. 한식에 대한 16개의 기본 패싯과 85개의 하위 패싯으로 나누고 패싯의 인용 순서를 '음식 종류'를 중심으로 재료, 에너지, 공간, 시간 순으로 조합하도록 제안하였다.

또한, 시맨틱 웹 기술에도 활용하는 연구가 있다. 이은욱과 박희진(2018)은 이산가족 찾기 기록 검색을 위한 패싯 기반 온톨로지 모델을 제안하면서, 온톨로지 모델링을 위해 패싯 분석은 KBS 이산가족 찾기 기록과 국가기록원의 원문 내용 및 기술요소를 분석하였고, 이산가족 찾기 방송 기록과 사건 중심 기록을 분석하여 패싯을 도출하였다. 패싯은 5개의 상위클래스(KBS 이산가족기록, 인물, 출처, 원본기록, 사건)와 18개의 하위클래스로 구성하였다.

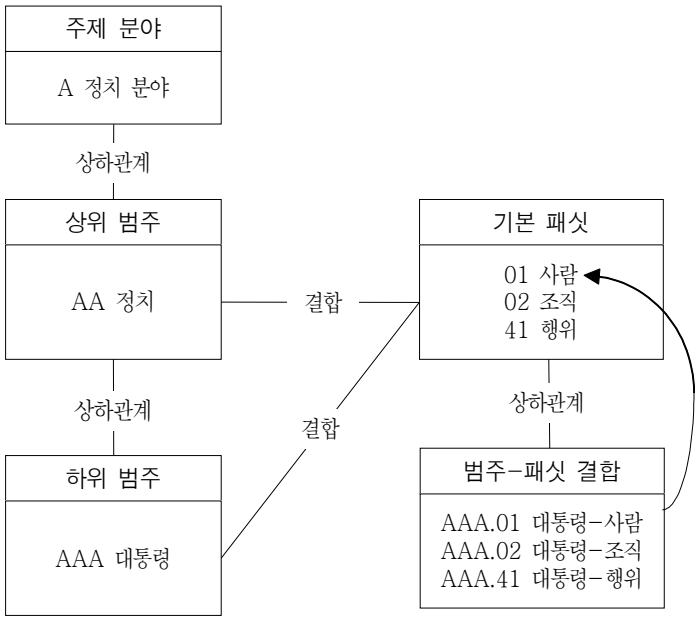
한편 특별한 영역을 한정하지 않고 기본 패싯을 도출하는 연구로 장인호(2018)는 기존의 일반 시소러스 내의 어휘를 패싯 분석하여 기본 패싯을 확립하기 위해 "캐나다 정부 핵심 주제 시소러스"의 2,260개 용어 전체를 애치슨(Aitchison) 등이 제시한 기본 패싯 16개를 이용하여 초기 분석하였고 최종 14개의 "정부 정보 자원 기본 패싯"을 확립하고 재분류하였다. 확립된 기본 패싯은 추상물, 법적 역할, 자연물, 인공물, 특성, 물질, 장르/형식, 사람, 단체, 과정, 조작, 이벤트, 장소, 시간이다.

또한, 장인호(2019)는 ISO 25964-1 제11절 "패싯 분석"과 제5절의 "시소러스에 있어서의 개념 및 그들의 범위"를 분석하여, 제11절에 예시된 대상, 물질, 에이전트, 행위, 장소, 시간 등을 확장하는 연구를 수행하였다. 이를 위해 기존 온톨로지의 최상위 개념과 기존의 기본 범주들을 참조하여, 정신적 실체를 기본 범주에 명시적으로 추가하고, 기본 범주 확립을 위해 일부를 조정하였다. 최상위 범주를 독립 실체와 종속 실체로 이분하고 하위 구분으로 전자는 28범주, 후자는 2범주를 두었다.

Ⅲ. 패싯 택소노미 구축 방안

1. 모형 개발

신문기사정보를 주제 분야로 나누고, 주제 분야 내에서 범주들은 택소노미와 패싯 분류 체계를 형성하도록 하였다. 전조합 체계의 패싯 택소노미를 구성하는 모형을 개발하였다. 전체 모형도는 <그림 1>과 같다.



<그림 1> 전체 모형도

2 설계 원칙

가. 전체 구성

신문기사정보를 주제 분야(정치, 경제, 사회, 문화, 국제 등)와 범주(정치 분야의 경우, 정치일반, 행정, 사법, 외교, 군사 등)로 범주화한다. 그리고 각 하위 범주 내의 개념 또는 용어는 시소러스의 계층 관계에 상응하는 택소노미를 형성한다.

패싯 택소노미의 구성은 parent/child 관계 및 범주-패싯 결합 관계를 형성한다. 범주-기본 패싯 간의 관계는 예를 들어, "보험"에 대해 '보험—사람', '보험—조직', '보험—행위', '보험—행사', '보험—사건' 등으로 구성한다. 그리고 보험의 하위 범주의 '생명보험', '손해보험' 등은 보험과 계층 관계를 이루도록 설계하였다.

이와 같이 본 연구에서 패싯 택소노미는 parent/child 관계 및 범주-기본 패싯으로 결합하여 하위분류로 가지는 체계를 말한다. 택소노미라고 하는 용어의 사용이 매우 다양하기 때문에 택소노미가 패싯 분류 체계를 포함하는 의미로 패싯 택소노미라고 하는 용어를 선택하였다.

나. 패싯 택소노미의 확립 및 구성 방안

전조합의 유무, 다중 계층의 허용 여부, 기호법의 사용, 고유명사의 포함 여부 등을 고려하여 방안을 제시하였다. 패싯 택소노미의 확립 및 구성 방안은 다음과 같다.

① 전조합의 유무

패싯 택소노미는 범주와 기본 패싯을 조합하는 전형적으로 전조합시스템을 형성하도록 설계하였다. 예시는 다음과 같다.

예 : AA.01(정치—사람)

② 기호법의 사용

기호법을 사용하며, 주제 분야는 영문자 1자리수의 대문자, 최상위 범주는 영문자 2자릿수의 대문자로 하며, 더욱 하위 범주로 나눌 수 있다. 기본 패싯은 아라비아숫자로 구성하였으며, 이는 기호의 구분을 명확히 하고, 최종이용자가 축소 및 확대하여 사용할 수 있게 한 것이다. 또한, 확장성을 고려하여 사용하지 않는 영문자(특히, 모음)가 있다.

③ 다중계층의 허용 여부

패싯 택소노미에서는 허용하지 않는다.

④ 고유명사의 포함 여부

고유명사는 정보량이 많은 경우 채택한다. 단, 나머지는 전거파일에서 수용하는 것을 전제로 한다(본 연구에서는 전거파일에 대해서는 다루지 않는다.).

⑤ 지시 사항

각각의 범주는 지시사항에 의해 범주의 설명이나 범위 주기, 타 범주로의 안내 등을 지시할 수 있다.

⑥ 확장 방안

시소러스를 구축하고 상호 매핑하는 경우 더욱 상세한 택소노미를 구성하는 것을 상정하여 설계하였다. 예를 들어, 시소러스 우선어 '작곡가, '연주가, '지휘자' 등은 상위어 '음악가'를 가지고 택소노미 '음악—사람'에 매핑할 수 있다.

다. 패싯 택소노미의 구조화 방안

주제 범주를 먼저 구조화하고, 상위-하위의 각 범주의 택소노미에 패싯 기반의 분류를 형성한다. 기호법(notation)으로 주제 범주는 대문자 알파벳을 사용하고, 기본 패싯 기호는 아라비아 숫자를 사용하여 십진분류법의 한계를 극복할 수 있도록 설계하였다.

라. 디스플레이 및 내비게이션 방안

패싯 택소노미의 디스플레이는 인쇄형과 스크린형으로 나눌 수 있다. 인쇄형은 주제 범주 순과 기본 패싯 순으로 구성하였다. 패싯별 디스플레이에서의 순서는 기본 패싯의 순서에 따

른다. 즉, 행위주체(사람→조직→집단)→물리적 개체(시설→설치물→도구→자연물)→물질 및 재료(물질→재료)→산물(표현물→자연산물)→프로세스(행위→현상)→행사 및 상(행사→상)→사건사고(사건→사고)→장르 및 형식(장르→형식)→추상적 개체(개념→방법→학문→속성)→발생(시간→역사→장소) 순이다.

① 인쇄형

인쇄형 디스플레이는 범주순과 패싯순 디스플레이를 구성한다. 범주 순 디스플레이의 예시는 <표 1>과 같다.

<표 1> 범주순 디스플레이 예시

B 경제분야
BA 경제
BA.01 경제—사람
BA.02 경제—조직
BA.03 경제—집단
BA.11 경제—시설
BA.12 경제—설치물
BA.13 경제—도구
BA.21 경제—물질
BA.22 경제—재료
BA.31 경제—표현물
BA.32 경제—자연산물
BA.41 경제—행위
BA.42 경제—현상
BA.51 경제—행사
BA.52 경제—상
BA.61 경제—사건
BA.62 경제—사고
...
BB 재정
BB.01 재정—사람
BB.02 재정—조직
...
BBA 조세
BBA.01 조세—사람
BBA.02 조세—조직

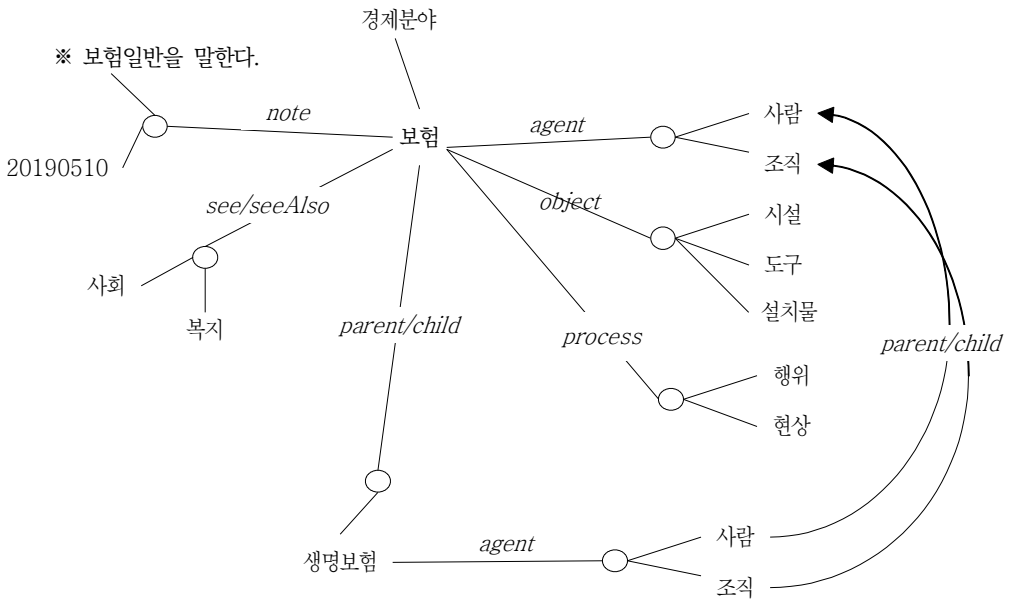
패싯순 디스플레이의 예시는 <표 2>와 같다.

<표 2> 패싯별 디스플레이 예시

행위주체(00)
・사람(01)
・・경제—사람(BA.01)
・・・금융—사람(BC.01)
・・・기업—사람(BD.01)
…
・・정치—사람(AA.01)
・・・국회—사람(AAC.01)
・・・대통령—사람(AAB.01)
・・・선거—사람(AAD.01)
・・・정당—사람(AAF.01)
…
・조직(02)
・・경제—조직(BA.02)
・・・금융—조직(BC.02)
・・・기업—조직(BD.02)
…
・・정치—조직(AA.02)
・・・국회—조직(AAC.02)
・・・대통령—조직(AAB.02)
・・・선거—조직(AAD.02)
・・・정당—조직(AAF.02)
…

② 스크린형 디스플레이

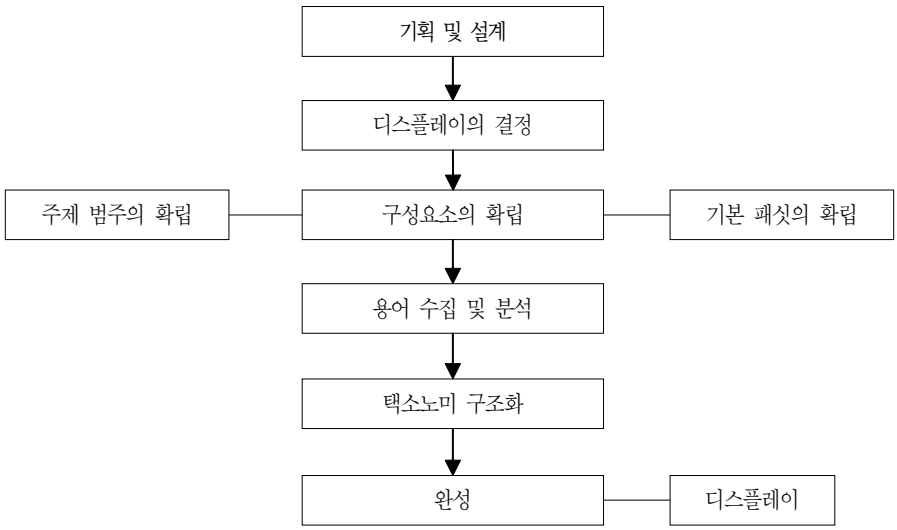
스크린형의 그래픽 디스플레이는 <그림 2>와 같다.



<그림 2> 그래픽 디스플레이 예시

3. 구축 절차 확립

Hedden(2016)과 Stewart(2008)의 택소노미 구축 절차 등을 분석하고 참조하여 패싯 택소노미 구축 절차를 확립하였다. 실제 구축 실험(예술 분야)을 실시하여 반복・검토하였다. 신문기사정보 패싯 택소노미의 구축절차는 <그림 3>과 같다.



〈그림 3〉 패싯 택소노미 구축 절차

4. 주제 분야의 확립

기존의 제어 어휘 중 분류 체계(예 : 뉴스코드 가이드북(한국 NewsML포럼 2008), IPTC Subject Codes(IPTC Homepage), 朝日記事データベース分類の手引き(朝日新聞社ニューメディア本部 1989) 및 시소러스(예 : 뉴스ML 시소러스(한국언론재단 2008), Washington Post Thesaurus(The Washington Post 1986), ニュース・シソーラス(広木守雄, 服部信司 2004), 日経シソーラス(日本経済新聞社 웹사이트) 등을 검토하여 구성하였다. 단, 본 제안에서 제시하는 택소노미를 고려하여 융합될 수 있도록 구성하였다. 확립된 주제범주는 <표 3>과 같다.

<표 3> 신문기사 주제 범주

기호	캡션	범위 및 설명	기호	캡션	범위 및 설명
A	**정치분야**		**F**	**사건사고**	범주이며, 기본 패싯과는 다르다.
AA	정치	정치일반, 대통령, 국회, 선거, 정당	FA	사건	범주이며, 기본 패싯과는 다르다.
AB	행정	행정, 지방행정 등	FB	사고	사고, 자연재해 등
AC	사법	검찰, 변호사, 법원, 경찰	**G**	**문화분야**	
AD	외교	외교, 각국간의 관계 등	GA	문화	문화일반
AF	군사	국방, 군사 등	GB	문화재	국보, 보물, 천연기념물 등
B	**경제분야**		GC	예술	문학, 미술, 음악 등
BA	경제	경제일반, 물가 등	GD	예능	예능, 연예 등
BB	재정	재정, 조세	GF	종교	종교일반, 각 종교 등
BC	금융	금융, 화폐, 은행, 증권 등	GG	스포츠	레저스포츠를 포함한다.
BD	기업	기업, 대기업, 중소기업 등	**J**	**교육분야**	
BF	보험	보험, 생명보험, 화재보험 등	JA	교육	교육일반, 유치원교육, 초등교육 등
BG	노동	노동, 고용, 노사문제 등	JB	학문	범주이며, 기본 패싯과는 다르다.
C	**산업분야**		**M**	**정보분야**	
CA	자원·에너지	자원, 에너지	MA	정보	정보산업은 'CH 정보산업'에서.
CB	농림수산	1차산업	MB	언론	신문, 방송 등을 다룬다.
CC	광공업	2차산업	MC	출판	출판, 도서, 연속간행물 등
CD	운수·교통	운수, 교통	**N**	**자연분야**	
CF	토목·건설	토목, 건설, 도시건설 등	NA	자연	우주는 'ND 천문'을 보라.
CG	상업	무역	NB	자연현상	기상 등 자연현상을 다룬다.
CH	정보산업	정보영역과 구분한다.	NC	생물	동물, 식물 등
D	**사회분야**		ND	천문	우주 관련 주제를 다룬다.
DA	사회	사회일반, 사회문제 등	**Z**	**국제분야**	
DB	사회복지	사회복지, 사회사업 등	ZA	국제	국제단체, 국제정치, 국제경제 등
DC	환경	환경공해	ZB	아시아	아시아단체, 아시아정치 등
DD	생활	의, 식, 주, 가정생활 등	ZC	중동	중동단체, 중동정치 등
DF	세대	유아, 아동, 청소년, 남성, 여성 등	ZD	아프리카	아프리카단체, 아프리카정치 등
DG	의료	의료, 건강, 질환 등	ZF	유럽	유럽단체, 유럽정치, 유럽경제 등
DH	행사	계절, 기념일 등. 패싯과 구별할 것.	ZG	아메리카	미주단체, 미주정치, 미주경제 등
DJ	레저	취미, 오락, 레저 등	ZH	오세아니아	오세아니아단체, 오세아니아정치 등
DK	관광	관광, 여행 등	ZJ	극지방	남극, 북극, 남극개발 등

5. 기본 패싯의 확립

기존의 상위 온톨로지(DOLCE(Masolo, C. et. al. 2003), BFO(Arp, R., Smith, B. & Spear, A. D., 2015), YAMATO(溝口理一郎, 2012))를 검토하고, 기존의 국내외 기본 패

싯(주제 분야의 전체를 대상으로 하는 Aitchison et. al.(2000)의 것뿐만 아니라, 건설 분야, 법률정보, 방송자료 등을 포함하여 도서나 논문 등의 문헌 등)을 조사·분석하여 임시 기본 패싯을 설정하고, 신문기사정보의 제어 어휘집(분류표, 시소러스 등)을 참조하여 확립하였다.

넓은 주제 분야 대상(정부 정보, 신문기사정보, 연속간행물 정보, 법령 정보 등)뿐만 아니라 전문정보를 대상으로 하는 기본 패싯 모두 행위주체, 행위, 이벤트, 시간적 발생, 장소, 물질, 대상(object) 등 주된 패싯은 대부분 일치하기 때문에 신문기사정보 및 일반성이 높은 패싯 이외의 것도 참조하여 설정하고 그 하위 패싯을 두도록 설계하였다. 기본 패싯은 2레벨까지로 나누었으며, 아라비아숫자를 사용하는 기호법을 선택하고, 활용에 있어서는 어느 단계까지를 선택하든 이용자에게 일임되도록 설계하였다. 도출된 기본 패싯은 <표 4>와 같다.

〈표 4〉 신문기사 기본 패싯

최상위 패싯	서브패싯	설명 및 범위
00-행위주체	01-사람	직업과 직위도 포함하며, 수동체도 포함한다.
	02-조직	기관단체. 하나처럼 작용하는 사람들의 집합을 말한다. 기관단체를 일컫는다.
	03-집단	단체, 그룹
10-물리적 개체	11-시설	행위주체, 설치물, 각종 도구를 구성요소로 하는 개념이다.
	12-설치물	건물 등 지상에 설치된 상징물을 제외한 각종 설치물을 일컫는다. 상징물은 '31 표현물'을 사용한다.
	13-도구	지상에 설치되지 않은 도구, 장비, 설비 등을 일컫는다.
	14-자연물	행위주체가 만들지 않는 것을 말한다.
20-물질 및 재료	21-물질	형태를 갖지 않는 물리적 개체를 말한다.
	22-재료	건축재료, 박물관자료 등을 말한다.
30-산물	31-표현물	미술품, 음악작품 등을 말한다.
	32-자연산물	곡식, 계란, 우유 등을 말한다.
40-행위 및 현상	41-행위	참여자에 의해 행해지는 프로세스. 활동, 행정, 정책을 포함한다.
	42-현상	참여자가 없는 프로세스.
50-행사 및 상	51-행사	수상식 이외의 각종 행사를 다룬다.
	52-상	수상식 외에 일반적으로 해당 상에 대한 기사도 다룬다.
60-사건사고	61-사건	행위주체에 의한 의도를 가진 사건을 말한다. 역사적 사건을 포함한다.
	62-사고	교통사고, 자연재해 등 의도하지 않은 사고를 말한다.
70-장르 및 형식	71-장르	예술 등
	72-형식	매체, 법령, 관습 등 표현매체를 포함한다.
80-추상적 개체	81-개념	머릿속의 개념을 말한다.
	82-방법	'OO하는 방법'을 말한다.
	83-학문	인문과학, 사회과학, 자연과학 등 학문을 말한다. 이론을 포함한다.
	84-속성	생물학적 특성, 위치적 특성, 종교적 특성 등을 포함한다.
90-발생	91-시간	계절
	92-역사	역사 시대 구분이나 역사에 대한 것을 말한다. 역사적 사건은 '61 사건'을 사용한다.
	93-장소	지역, 경관, 풍경 등을 포함한다.

Ⅳ. 패싯 택소노미 구축 사례 : 예술 분야

1. 용어 수집

예술 분야의 용어 수집은 기존의 신문기사 분류표를 중심으로 분류 기호에 대한 캡션(caption)에서 1차로 추출하였으며, 더욱 보완하기 위해 신문기사 시소러스 및 실제의 신문에서 개념의 레벨을 고려하여 선정하였다. 수집된 용어는 145어이며, 용어 리스트는 <표 5>와 같다.

〈표 5〉 수집된 용어 리스트

가곡	동시	미술전람회	서예가	영화사(映畵社)	지공예
가짜미술품	동양화	미술평론가	서예전	영화상	촬영
건축	만화가	미술평론가협회	서체	영화세트	칠기공예
건축가	모델	미술품	서커스단	영화제	타악기
건축가	목공예	미술품가격	성악가	영화촬영	탁본
건축전	무용가	미술품거래	소설	영화촬영장	탤런트
고미술상	무용단	미술품경매	수중촬영	예능단체	판화
고미술품	문학관	미술품복원	스케치	예술단	포스터제작
고미술품	문학상	미술품복제	시(문학)	예술사	표절시비
고악기	문학작품	미술품수입	시인	예술품	피아노
골동품	물감	미술행사/활동	실내악단	예술품도난	피아노조율
공예	미술	미술행정	아이스발레단	예술품화재	한국미술사
공예가	미술감상	미술협회	악기	유리공예	한국화
공예전	미술계	발레단	악기조율	음악당	합창단
공예품	미술공모전	방송작가	악단	음악사	현악기
공중촬영	미술관	배우	야외촬영	음악이론	화가
교향악단	미술단체	벽화	연극감상	음악저작권	화랑
국악기	미술사	보석공예	연극사	작가	화랑협회
국악인	미술상(美術商)	비보이	연주법	작품변조	화방
극장	미술상(美術賞)	사진	염직공예	작품위조	회화
금속공예	미술세미나	사진작가	영화감상	장식공예	회화도구
금지곡	미술심포지엄	사진전	영화관	전각	
도안	미술이론	상업미술	영화배급	조각	
도예가	미술인	서양화	영화법규	조각가	
도자기공예	미술재료	서예	영화사(映畵史)	조각품	

2. 어휘 분석

수집된 용어는 기본 패싯 별로 그룹화(클러스터링)하였다. 예시를 <표 6>에 열거하였다. 이를 바탕으로 기본 패싯을 정제하였으며, 실제 주제 분야에 따라서는 사용되지 않는 패싯도 있을 수 있다.

〈표 6〉 어휘 분석 예시

행위주체 사람 : 미술인(화가, 모델, 만화가, 건축가, 도예가, 사진작가, 미술평론가, 공예가, 조각가, 미술상(美術商), 고미술상), 배우, 건축가, 무용가, 작가, 방송작가, 시인, 탤런트, 국악인, 비보이, 서예가, 성악가 조직 : 미술단체(기관/단체), 미술협회, 화랑협회, 미술평론가협회, 미술계, 영화사(映畵社) 집단 : 악단, 교향악단, 실내악단, 예능단체, 합창단, 서커스단, 발레단, 무용단, 예술단, 아이스발레단 물리적 개체 시설 : 미술관, 화랑, 화방, 음악당, 문학관, 극장, 영화관 설치물 : 영화세트, 도구 : 회화도구, 악기, 현악기, 피아노, 타악기, 국악기, 고악기 물질 및 재료 물질 : 물감 재료 : 미술재료 산물 표현물 : 예술품, 조각품, 가곡, 고미술품, 공예품, 문학작품, 소설, 시(문학), 동시, 벽화 행위 및 현상 행위 : 미술행정, 미술품복제, 미술품복원, 미술품거래, 미술품수입, 포스터제작, 영화배급, 촬영, 영화촬영, 수중촬영, 공중촬영, 야외촬영 행사 및 상 행사 : 미술행사/활동, 미술세미나/심포지엄, 미술품경매, 미술공모전, 미술전람회, 건축전, 공예전, 서예전, 사진전, 영화제 상 : 미술상(美術賞), 문학상, 영화상
사건사고 사건 : 예술품도난, 작품위조, 작품변조, 표절시비 사고 : 예술품화재 장르 및 형식 장르 : 미술, 판화, 스케치, 건축, 조각, 공예, 사진, 상업미술, 도안, 목공예, 금속공예, 보석공예, 유리공예, 지공예, 칠기공예, 염직공예, 장식공예, 도자기공예, 서예, 미술품(고미술품, 골동품, 회화, 서양화, 동양화, 한국화 형식 : 영화법규 추상적 개체 개념 : 미술감상, 연극감상, 영화감상, 음악저작권, 금지곡 방법 : 연주법, 피아노조율, 악기조율 등 학문 : 음악이론, 미술이론 속성 : 미술품가격, 서체, 탁본, 전각, 가짜미술품 발생 역사 : 예술사, 미술사, 한국미술사, 음악사, 연극사, 영화사(映畵史) 장소 : 영화촬영장

3. 택소노미의 구조화

기존 신문기사 분류표 등에서 열거된 예술 분야의 분류 항목을 기반으로 다음과 같이 범주화하여 구조화하고, 각각의 패싯과 결합하였다. <표 7>에 예술 분야의 구조화된 범주를 나타내었다.

<표 7> 예술 분야의 범주 구조화

기호법	분야명	패싯과의 결합
G	문화분야	문화 예술 분야를 말한다.
GA	문화	문화일반
GB	문화재	국보, 보물, 천연기념물 등
GC	예술	문학, 미술, 음악 등
GCA	문학	문학—사람, 문학—조직 …
GCAA	시	시—사람, 시—조직 …
GCAB	소설	소설—사람, 소설—조직 …
GCAC	수필	수필—사람, 수필—조직 …
GCAD	희곡	희곡—사람, 희곡—조직 …
GCB	미술	미술—사람, 미술—조직 …
GCBA	회화	회화—사람, 회화—조직 …
GCBB	건축	건축—사람, 건축—조직 …
GCBC	공예	공예—사람, 공예—조직 …
GCBD	서예	서예—사람, 서예—조직 …
GCBF	사진	사진—사람, 사진—조직 …
GCBG	상업미술	상업미술—사람 …
GCC	음악	음악—사람, 음악—조직 …
GCCA	국악	국악—사람, 국악—조직 …
GCD	뮤지컬/오페라	뮤지컬/오페라—사람 …
GCF	연극	연극—사람, 연극—조직 …
GCG	영화	영화—사람, 영화—조직 …
GCH	무용	무용—사람, 무용—조직 …
GD	예능	예능, 연예 등
GF	종교	종교일반, 각 종교 등
GG	스포츠	레저스포츠를 포함한다.

4. 디스플레이

시험용 구축에서의 패싯 택소노미의 디스플레이는 범주순과 패싯순 디스플레이를 예시하였다.

가. 범주순 디스플레이

범주순 디스플레이는 기호법에 따라 배열하였으며, 예시는 <표 8>과 같다.

〈표 8〉 구축된 범주순 디스플레이 예시

[G 문화]
GC 예술
GC.01 예술—사람
GC.02 예술—조직
GC.03 예술—집단
GC.11 예술—시설
GC.12 예술—설치물
GC.13 예술—도구
GC.21 예술—물질
GC.22 예술—재료
GC.31 예술—표현물
GC.41 예술—행위
GC.42 예술—현상
GC.51 예술—행사
GC.51 예술—상
...
GCC 음악
GCC.01 음악—사람
GCC.02 음악—조직
GCC.03 음악—집단
GCC.11 음악—시설
...

나. 패싯순 디스플레이

패싯순 디스플레이는 기본 패싯에 따라 '범주-패싯'이 계층 구조를 가지도록 하였으며, 예시를 <표 9>에 나타내었다.

〈표 9〉 구축된 패싯순 디스플레이 예시

행위주체(00)
·사람(01)
··예술—사람(GC.01)
···문학—사람(GCA.01)
····소설—사람(GCAB.01)
····수필—사람(GCAC.01)
····시—사람(GCAA.01)
····희곡—사람(GCAD.01)
···미술—사람(GCB.01)
···음악—사람(GCC.01)
...
·조직(02)
··예술—조직(GC.02)
···문학—조직(GCA.02)
····소설—조직(GCAB.02)
····수필—조직(GCAC.02)
····시—조직(GCAA.02)
····희곡—조직(GCAD.02)
···미술—조직(GCB.02)
···음악—조직(GCC.02)
...

5. 구축 결과 및 색인 사례

구축 방안에서 확립한 주제 분야는 대분야 10개, 중분야 52개이며, 기본 패싯은 상위 패싯 10개, 하위 패싯 26개였다. 시험 구축한 범주의 총 수는 18개이고, 패싯 결합 총 범주 수는 540이었으며, 디스플레이 유형은 범주 순 및 패싯 순의 예시를 나타내었다.

신문기사정보를 실제 색인하는 경우, 예를 들어, 부산국제영화제에 대한 기사와 전주국제영화제를 다루고 있는 신문기사는 “영화—행사”로 색인될 것이다. 각각의 행사명이나 신문기사에서 사용하고 있는 용어는 자연어(free-text)로 탐색이 가능하며, 시소러스 등의 제어 어휘와 상호운용성을 확보하는 경우에는 “영화—행사”의 하위 개념으로 매핑될 수 있다.

Ⅴ. 결론

패싯 기반의 분류 체계에 대한 연구는 종종 있지만, 추론 및 내비게이션을 위한 동일한 패싯으로 개념계층을 형성하는 택소노미를 조직하고, 주제 범주 등과 다차원으로 모형을 개발한 사례는 찾기 어렵다. 본 연구는 신문기사 정보를 대상으로 다음과 같은 연구를 실시하였다.

첫째, 택소노미의 계층 관계를 추론 및 내비게이션을 고려하여 조직하고, 해당 택소노미 범주에 기본 패싯을 조합하여 하위분류를 구성하는 모형을 개발하였다. 패싯 택소노미는 신문기사정보를 상위 분야(정치, 경제, 사회, 문화, 국제 등)와 범주(정치 범주의 경우, 정치일반, 행정, 사법, 외교, 군사 등)로 조직화하고 각 하위 범주 내의 개념 또는 용어는 기본 패싯과 결합하여 패싯 택소노미를 형성한다.

둘째, 택소노미의 구성은 범주 간의 계층 관계를 가질 수 있으며, 범주-패싯 결합은 예를 들어, “예술”에 대해 ‘사람’, ‘조직’, ‘시설’, ‘행위’, ‘행사’, ‘시간’, ‘장소’ 등과 결합한다. 그리고 예술의 하위 범주 ‘미술’, ‘음악’, ‘무용’ 등은 ‘예술’과 계층 관계를 이루어 추론과 브라우징에 활용할 수 있도록 구성하였다. 또한, 범주-패싯 결합도 기본 패싯순으로 계층 구조를 가지도록 하였다.

이와 같이 본 연구에서 패싯 택소노미는 계층 관계 및 기본 패싯으로 범주를 하위분류로 가지는 체계를 말한다. 택소노미라고 하는 용어의 사용이 매우 다양하기 때문에 택소노미가 패싯 분류 체계를 포함하는 의미로 패싯 택소노미라는 용어를 선택하였다.

셋째, 패싯 택소노미 구축 방안은 설계 원칙을 제시하고, 구축 절차를 확립하였으며, 주제 분야, 기본 패싯 등을 확립하였다. 설계 원칙은 패싯 택소노미의 전체 구성, 구성 방안, 구조화 방안, 디스플레이 등을 제시하였다.

넷째, 시험용 구축은 예술 범주에 대해 용어 145어를 대상으로 전 구성요소를 포함하는 패싯 택소노미를 구축하고, 디스플레이를 예시하였다. 시험 구축한 범주의 총 수는 18이며,

패싯 결합 총 범주 수는 540이었고, 디스플레이 유형은 범주 순 및 패싯 순으로 예시를 나타내었다.

이와 같은 연구는 기존의 십진 분류 체계 등 전조합 시스템에서는 기대할 수 없는 동일 패싯을 가진 계층 구조를 이용하는 추론에 사용할 수 있으며, 후조합 시스템의 대표적인 시소러스 등과 향후 상호운용성을 통해 더욱 상세한 탐색이 가능하게 할 수 있다. 신문기사를 빅데이터화 한 빅카인즈(Big Kinds, https://www.bigkinds.or.kr/) 시스템에서는 시소러스를 내장하고 있으며 본 연구의 결과인 패싯 택소노미와 상호운용성을 확보하면 효율성이 제고될 것이다.

현재 신문기사를 활용하여 새로운 사업을 창출하는 시도가 많다. 그들을 위해 공유 또는 재이용할 수 있는 어휘자원 인프라를 제공하고, 자동 범주화(자동분류), 뉴스 큐레이션 및 신문기사정보서비스의 개인화를 촉진하는데 기초자료로 활용될 수 있을 것이다. 특히, 넓은 분야에 걸쳐 형성되는 주제 범위(특히, 법령 정보, 연속간행물기사 정보, 정부 생산 정보 등)의 모형 개발 및 구축 지침을 작성하는 데에 참조가 될 수 있을 것으로 기대한다.

참고문헌

심지영. 2014. 패싯분석 기법을 적용한 방송자료의 내용 구조화에 관한 연구: 시사보도 뉴스 프로그램을 대상으로.『정보관리학회지』, 31(3): 313-329.

이은욱, 박희진. 2018. 이산가족 찾기 기록 패싯 기반 온톨로지 모델 설계에 관한 연구.『한국기록관리학회지』, 18(4): 231-257.

이정민. 2016.『무용학의 지적 구조 분석 연구 : 텍스트 마이닝 기반의 빅데이터 분석을 중심으로』. 박사학위논문, 성균관대학교 일반대학원 예술학협동과정 예술학 전공.

이종영 등. 2015. 재난유형 택소노미 분류체계와 주제 클러스터 레이블링 분석기법을 통한 재난보도탐색 시스템.『한국정보과학회 학술발표논문집』, 2015(6): 1609-1611.

임영호, 2013.『신문원론』 제3판, 서울: 한나래 아카데미.

장인호. 2013.『뉴스 코어 시소러스』의 구축 및 활용 방안에 관한 연구.『한국도서관정보학회지』, 44(3): 489-512.

장인호. 2018. 일반 시소러스의 어휘분석을 통한 기본 패싯 확립에 관한 연구.『인문사회 21』, 9(6): 1059-1070.

장인호. 2019. 시소러스 국제표준 기반 기본 범주의 확장에 관한 연구.『한국도서관정보학회지』, 50(1): 273-291.

정연경. 2013. 한식 정보 조직을 위한 패싯 구조화에 관한 연구.『한국문헌정보학회지』, 47(1): 15-37.

최지수. 2014.『의학 다큐멘터리의 효과적인 주제별 검색을 위한 택소노미 기반 동영상 주석 시스템 설계 및 구현 : <생로병사의 비밀>을 중심으로』. 석사학위논문, 서강대학교 대학원 컴퓨터공학과.

한국 NewsML포럼. 2008.『뉴스코드 가이드북』. 서울 : 한국언론재단.

한국언론재단. 2008.『뉴스ML 시소러스』(상)(하). 서울 : 한국언론재단.

홍기철. 2017. 패싯 분석 기법을 활용한 건설 시소러스 구축 방안에 관한 연구.『한국도서관정보학회지』, 48(1): 345-371.

広木守雄, 服部信司. 2004. ニュース・シソーラス : 新聞・放送ニュース検索のための主題14000語, 第4版, 東京: 日外アソシエツ.

広木守雄. 1981. ニュース・シソーラス : 新聞情報管理のための用語集, 第2版. 東京: 中日新聞本社.

溝口理一郎. 2012.『知の科学 : オントロジー工学の理論と実践』. 東京: オーム社.

日本経済新聞社. 日経シソーラス. http://t21.nikkei.co.jp/public/help/contract/price/20/thesaurus/index_AA.html [cited 2019. 10. 10].

朝日新聞社ニューメディア本部. 1989. 朝日記事データベース分類の手引き. 東京 : 朝日新聞社ニューメディア本部.

Aitchison, J., Gilchrst, A., & Bawden, D. 2000. *Thesaurus construction and use: a practical manual.* 4th ed. London: Aslib imi.

Arp, R., Smith, B. & Spear, A. D. 2015. *Building Ontologies with Basic Formal Ontology.* Cambridge, Massachusetts: The MIT Press.

Big Kinds, https://www.bigkinds.or.kr/

Broughton, V. 2006. *Essential thesaurus construction.* 1st ed. London: Facet Publishing.

Cheung, C.F., Lee, W.B. & Wang, Y. 2005. "A multi-facet taxonomy system with applications in unstructured knowledge management." *JOURNAL OF KNOWLEDGE MANAGEMENT,* 9(6): 76-91.

Hedden, H. 2016. *The Accidental Taxonomist.* New Jersey: Information Today, Inc..

IPTC Home page. <http://cv.iptc.org/newscodes/subjectcode/> [cited 2019. 10. 10].

ISO 25964-1. 2011. *Information and documentation - Thesauri and Interoperability with other vocabularies Part 1: Thesauri for information retrieval.* Switzerland: ISO.

ISO 25964-2. 2013. *Information and documentation - Thesauri and Interoperability with other vocabularies Part 2: Interoperability with other vocabularies.* Switzerland: ISO.

Lebeuf, C. et. al., 2019. *Defining and Classifying Software Bots: A Faceted Taxonomy.* 2019 IEEE/ACM 1st International Workshop on Bots in Software Engineering

(BotSE).

Masolo, C. et. al. 2003. WonderWeb Deliverable D18 Ontology Library(final), IST Project 2001-33052 WonderWeb: Ontology Infrastructure for the Semantic Web. http://www.loa.stc.cnr.it/Papers/D18.pdf. [cited 2019. 10. 1].

Ryan, C. 2014. *Thesaurus construction guidelines: An introduction to thesauri and guidelines on their construction.* Dublin: Royal Irish Academy and National Library of Ireland.

Stewart D. L. 2008. *Building Enterprise Taxonomies.* Mokita Press.

The Washington Post. 1986. *The Washington Post Thesaurus.* Washington : The Washington company.

Zong, Nansu,. Kim, Hong-Gee & Nam, Sejin. 2017. "Constructing faceted taxonomy for heterogeneous entities based on object properties in linked data." *Data & Knowledge Engineering,* 112: 79-93.

국한문 참고문헌의 영문 표기
(English translation / Romanization of reference originally written in Korean)

Chang, Inho. 2013. "A Study on the Establishment and Applications of the "News Core Thesaurus"." *Journal of the Korean Library and Information Science Society,* 44(3): 489-512.

Chang, Inho. 2018. "A Study on the Establishment of Fundamental Facets Using Vocabulary Analysis with General Thesaurus." *The Journal of Humanities and Social Sciences,* 21, 9(6-2): 1059-1070.

Chang, Inho. 2019. "A Study on the Expansion of Fundamental Categories Based on Thesaurus Internation Standards." *Journal of the Korean Library and Information Science Society,* 50(1): 273-291.

Choi, Jisoo. 2014. *Design and implementation of video annotation system for effective search by subject of medical documentary using taxonomies : focused on kbs <mysteries of the human body>.* Graduate thesis, Sogang University.

Chung, Yeon-Kyoung. 2013. "A Study on Structure of a Faceted Classification for Organizing Korean Food Information." *Journal of the Korean Library and Information Science Society,* 47(1): 15-37.

Hiroki Morio. *News Thesaurus.* Chunichishinbunsha, 2nd ed., 1981.

Hiroki, Morio, *Hatori, Shinji. News Thesaurus.* Tokyo: Nichigai Associates (publisher),

2004.

Hong, Ki-Churl 2017. "A Study on Building Method of the Construction Industry Thesaurus Using Facet Analysis Method." *Journal of the Korean Library and Information Science Society,* 48(1): 345-371.

Korea NewsML Forum. 2008. *News Code Guidebook.* Seoul: Korea Press Foundation.

Korea Press Foundation. 2008. *News Articles Thesaurus.* Seoul: Korea Press Foundation.

Lee, Eun-Uk and Park, Heejin. 2018. "A Study on a Facet-Based Ontology Design for Archival Records Finding Dispersed Families." *Journal of Korean Society of Archives and Records Management.* 18(4): 231-257.

Lee, Jongyoung et. al. 2015. "Explorable System of Disaster Broadcasting Using Taxonomy Classification for Disaster Types and Analytical Methods of Topic Cluster Labeling." *Korean Institute of Information Scientists and Engineers Korea Computer Congress,* 2015(6) 1609-1611.

Lee, Junmin. 2016. Intellectual Structure Analysis in Korean Dance Studies: focused on Text Mining based Big Data Analytics. Graduate Thesis, Sungkyunkwan University.

Lee, Yung-Ho. 2005. *Principles of Newspaper.* Seoul: Hanarae.

Mizoguchi, R. 2012. *Science of Intelligence : Theory and Practice of Ontology Engineering.* Tokyo: Ohmsha.

Shim, Jiyoung. 2014. "A Faceted Classification Analysis of TV content: Using News and Current Affairs Programs." *Korean Society for Information Management,* 31(3): 313-329.