

저작 클러스터링 분석을 통한 FRBR의 목록 적용에 관한 연구*

A Study of FRBR Implementation to Catalog by Using Work Clustering

이미화(Mi-Hwa Lee)**

정연경(Yeon-Kyoung Chung)***

초 록

검색 자원의 증가와 멀티 버전 자료의 증대에 따라 목록의 기능 중 집중기능이 필요하게 되면서, FRBR(Functional Requirements for Bibliographic Records)이 제안되었으나, 기존 서지데이터를 FRBR로 변환시 불완전함과 FRBR을 구체화할 목록규칙의 부재 등 현행 목록데이터를 FRBR로 완벽히 구현하기에는 한계가 있다. FRBR을 온라인목록에 적용하기 위해서는 방대한 양의 기존 MARC 데이터를 FRBR 알고리즘으로 처리해야 하지만, 기입력된 MARC 데이터와 현행 목록규칙은 FRBR을 완벽히 지원하지 않기 때문에 FRBR로의 변환이 용이하지 않다. 따라서 목록규칙과 MARC 기술에 대한 지침이 필요하다. 또한 국내에 FRBR 시스템을 구축하기 위해서는 FRBR 알고리즘을 실제 목록시스템에 적용하여 문제점과 해결책을 모색해야 한다. 이에 본고는 FRBR을 목록에 적용하기 위해서는 FRBR을 위한 목록규칙과 MARC 데이터 기술 방안이 선행되어야 한다는 전제하에 FRBR을 위한 저작 표현형 알고리즘을 개발하고, 실제 도서관 시스템에 적용하여 저작의 집중도와 적용시의 문제점을 분석하고, FRBR의 저작 집중성을 높이기 위한 방안으로 목록규칙과 MARC 입력방안을 제안하고 이를 검증하였다.

ABSTRACT

The purposes of this study are to explore FRBR utilities such as work clustering and expression clustering and problems of application of the FRBR by developing work and expression clustering algorithm and implementing it into cataloging system, and to suggest new cataloging rules for FRBR and guideline of MARC description to improve FRBR work clustering. FRBR was suggested by necessitation of collocation function of bibliographic records according to increase of searching materials and multi-version materials, but FRBRization has some problems such as imperfect conversion of bibliographic records to FRBR records and inappropriateness of current cataloging rules for FRBR. Bibliographic records must be processed by FRBR algorithm to construct FRBRized system, but bibliographic records and current cataloging rules couldn't perfectly support FRBRization. Therefore cataloging rules and guidelines of MARC description for FRBR are needed. For constructing FRBRized cataloging system in Korea, it is needed to find problems and solution through FRBR practical application such as developing FRBR algorithm and applying it to cataloging records.

키워드: 서지레코드의 기능상의 요건, 목록규칙, 자원기술과 접근, 저작클러스터링
FRBR, cataloging rules, RDA, work clustering

* 본 연구는 이화여자대학교 대학원 박사학위논문을 요약한 것임.

** 한성대학교 홍보교류팀 과장(leemh@hansung.ac.kr) (제1저자)

*** 이화여자대학교 사회과학대학 문헌정보학전공 교수(ykchung@ewha.ac.kr) (공동저자)

▪ 논문접수일자: 2008년 8월 15일 ▪ 최초심사일자: 2008년 8월 18일 ▪ 게재확정일자: 2008년 8월 21일
▪ 情報管理學會誌, 25(3): 65-82, 2008. [DOI:10.3743/KOSIM.2008.25.3.065]

1. 서론

1.1 연구의 필요성

최근 검색 자원의 증가와 멀티 버전 자료의 증대에 따라 목록의 기능 중 집중기능이 필요하게 되면서, IFLA는 『서지레코드의 기능상의 요건(Functional Requirements for Bibliographic Records, FRBR)』을 제안하였다. FRBR은 이용자의 검색 행동에 따른 목록 대상 개체를 저작, 표현형, 구현형, 개별자료의 4개 계층으로 구분하고, 저작을 중심으로 자료를 집중시키고 자료 간의 관계에 중점을 둔다. 이러한 계층과 관계를 통해 FRBR은 이용자 측면에서 향상된 검색 기능을 제공할 수 있다.

국내에서는 기존 데이터를 FRBR로 변환하는 문제, FRBR을 구체화할 목록규칙의 부재 등 현행 목록데이터로 FRBR 저작 및 표현형의 집중을 구현하기 어렵기 때문에 FRBR이 크게 활성화되지 못하고 있다. FRBR을 온라인목록에 적용하기 위해서는 방대한 양의 기존 MARC 데이터를 FRBR 알고리즘으로 처리해야 하지만, 기입력된 MARC 데이터와 현행 목록규칙은 FRBR을 완벽히 지원하지 않는다. 따라서 FRBR 구현을 위한 목록규칙과 MARC 데이터 기술 방안이 마련되어야 한다.

이에 본 연구에서는 FRBR의 목록 적용을 위한 목록규칙과 MARC 기술 방안을 제안하고자 한다. 이를 위해 FRBR 저작 및 표현형 클러스터링 알고리즘을 설계하여 도서관 온라인목록에 적용하고, 저작 및 표현형 클러스터링 정도를 파악하고, 클러스터링이 낮은 문제점을 목록의 측면에서 분석하였다. 분석 결과

를 바탕으로 새로운 목록규칙을 제안하고, 실험 데이터의 구축과 재목록을 통해 FRBR을 위한 목록규칙을 검증하였다.

1.2 연구방법

저작 및 표현형 클러스터링 분석을 위한 알고리즘은 국내 대학도서관 중 전거데이터를 구축하는 기관이 많지 않은 점을 고려하고, FRBR 저작과 표현형의 속성에 해당하는 MARC 서브필드의 입력 정도를 분석해 설계하였다.

설계한 알고리즘을 한성대 도서관 총 304,861건의 레코드에 적용하여 총 저작 수, 저작당 구현형 수, 저작당 표현형 수로 FRBR의 저작 및 표현형 집중성을 살펴보았다. 또한 구현형을 많이 갖는 상위 저작을 주제별로 총 100개를 선정하고, 저작명으로 재검색하여 알고리즘으로 처리된 결과와 비교하여 누락 클러스터링과 오류 클러스터링을 조사하고, 누락 및 오류 레코드의 원인을 목록규칙의 측면에서 분석하였다.

마지막으로 FRBR을 목록에 적용하기 위한 목록규칙을 제안하고, 제안된 목록규칙의 적합성을 파악하기 위해 100개 저작 총 1,300개 레코드의 실험 데이터를 구축하고 서지레코드를 재목록하였다. 다시 알고리즘을 적용해 저작집중도가 재목록 전후 다른 점을 비교 검증하였다.

1.3 선행연구

FRBR을 서지레코드에 적용하기 위한 알고리즘과 구현 시스템에 대한 연구로, Hegna &

Murtomaa(2002)는 저작, 표현형, 구현형의 각 속성에 맞게 FinMARC과 NORMARC의 태그와 서브필드를 매핑하고 이를 BIBSYS에 구현해 FRBR의 적용 가능성을 증명하고 이용자 인터페이스를 향상시켰다. Hickey, O'Neill & Toves(2002)는 저작과 표현형으로 서지레코드를 그룹핑하는 알고리즘을 설계하여 실험하였다. Cho(2006)는 전자파일을 이용할 수 없는 국내 온라인목록 환경에서 FRBR을 구현하기 위한 알고리즘을 이론상으로 고안하였다. 김현희 등(2007)은 FRBR을 적용하기 위해 음악저작을 대상으로 음악자료 유형에 따라 표현형 식별을 위한 요소를 달리 적용한 알고리즘을 작성하였다.

FRBR을 온라인목록에 적용함에 따라 FRBR 유용성 검증에 대한 필요성이 커졌는데, Bennett, Lavoie & O'Neill(2003)은 저작의 집중 정도를 파악하기 위해 996건의 샘플 레코드를 대상으로 저작 클러스터링을 조사하였다. Hickey & O'Neill(2005)은 OCLC 저작 클러스터링 알고리즘을 설계해 적용한 결과 총 37,772,687개 저작 중 36,330,501개가 단일 구현형을 갖는 기초적인 저작임을 밝혀냈다. 김정현(2007)은 1,000건의 서지레코드를 표본으로 서지적 관계에 기반한 저작 유형을 분석한 결과 82.5%가 단일 구현형을 갖는 저작이었고, 16.9%는 2개 이상의 구현형을 갖는 저작이었으며, 저작당 구현형 수는 1.45개로 분석하였다.

이와 같이 온라인목록에 FRBR을 적용하기 위한 알고리즘을 개발하고, FRBR의 유용성을 분석하면서 데이터 질에 대한 문제점이 제기되고 있다.

2. FRBR의 개념

2.1 FRBR의 필요성

목록 환경의 변화와 검색 자원의 확장에 따라 목록 측면에서 이를 해결하기 위한 방안으로 FRBR의 필요성이 대두되었다. Svenonius(2000)는 목록 자동화와 분담목록의 확산 등 목록 환경의 급속한 변화로 목록 기능에 대한 제고와 함께 국제적인 표준, 세계적인 목록, 연계된 시스템이 필요하게 되면서 FRBR의 필요성이 강조되었다고 하였다.

온라인목록에 FRBR을 적용하여 얻는 장점을 Bennett, Lavoie & O'Neill(2003)은 서지 데이터베이스의 여러 관점을 지원하여 이용자의 다양한 요구를 수용하며, 둘째, 중복이라고 볼 수 있는 개별자료를 하나의 표목으로 융화시켜 목록 내에서 서지 개체의 계층성을 표현함으로써 검색을 향상시키고, 셋째, 신규 레코드 생성 시 작성할 거의 모든 정보를 선택할 수 있어 여러 서지레코드에서 정보를 결합해 목록의 생산성을 증대시킨다고 하였다. Tillett(2005)은 서지레코드를 찾고, 집중시키고, 저작과 관련된 저작 그룹을 디스플레이에서 제시하는 것이 FRBR의 장점이라고 하였다.

따라서 FRBR은 서지 개체의 계층과 관계를 통해 이용자의 검색을 향상시키고, 데이터의 중복입력을 방지하여 목록의 생산성을 증대시키는 장점을 갖는다.

2.2 저작 및 표현형 속성과 MARC

2.2.1 저작 속성과 MARC

FRBR은 개체 관계(Entity-Relation) 모델로, 개체는 제1그룹에 속하는 저작, 표현형, 구현형, 개별자료, 책임사항인 제2그룹에 속하는 개인과 단체명, 주제인 제3그룹에 포함되는 개념, 사물, 사건, 장소이다(IFLA 1998).

개체 중에서 특히 서지레코드를 그룹화하는 저작 개념이 중요하다. 저작은 지적 예술적 생산물로 추상적인 개념이고, 지시되는 실체가 아니다. 저작은 그룹핑의 개념으로서 일종의 저작 이름의 전거통제 기능을 하고 있다. 저작의 전거통제가 보다 효율적이기 위해서는 저작의 개념 정의와 저작의 범위가 중요한데 이는 저작의 정의에 따라 집중의 정도가 달라지기 때문이다. 저작의 범위가 너무 좁을 경우, 저작 집중도가 떨어져 FRBR의 유용성이 낮으며, 너무 포괄적으로 저작을 정의할 경우, 오히려 검색의 효율을 저하시킬 수도 있다.

IFLA(1998)는 저작의 개념을 목록 중심으로 동일한 서명과 저자를 갖는 것으로 좁게 정의하였다. 즉 동일 저작에 속하는 표현형에는 개정이나 갱신, 축약, 증보, 번역서, 개작이나 편곡, 더빙하거나 자막을 넣은 필름 등이지만, 지적이나 예술적 노력을 통해 저작의 내용을 독립적으로 개정한 경우에는 새로운 저작으로 취급하였다.

저작을 식별하는 속성은 저작의 표제, 저작의 형식, 저작일, 기타 식별 특성, 중간 예정일, 대상 이용자 등이며, 이는 MARC 서브필드와 대응될 수 있다. 예를 들어, 저작의 표제에 해당하는 속성은 통일서명 130 ▼a, 240 ▼a이고,

130, 240이 없을 경우 243, 245를 이용한다.

2.2.2 표현형 속성과 MARC

표현형은 저작의 실현물이며 표현형의 경계는 주로 내용 상의 변화이다. 저작을 표현하기 위해 사용된 지적 도구의 변화는 새로운 표현형으로 취급한다. 실질적으로 표현형이 나누어지는 기준은 자료와 이용자의 기대나 요구에 다르다(IFLA 1998).

표현형 속성에는 표현형 표제, 형식, 일자, 언어, 기타 특성, 확장, 개정, 내용 요약, 배경, 비평, 이용 제한 등이 해당된다. 표현형에서 형식의 속성은 MARC의 리더 06을, 언어는 008/35-37, 041 ▼a를, 기타 식별특성은 250 ▼a를 사용한다.

자료 유형에 따라 표현형의 속성이 변경되고, 표현형의 속성을 현재 목록 체계의 구성요소와 매핑시키기 어려워 프로그래밍을 통한 표현형 처리에 한계가 있기 때문에 표현형은 다른 개체에 비해 쉽게 구현되지 못하고 있다.

2.3 알고리즘 사례

2.3.1 OCLC 저작 세트 알고리즘

기 구축된 MARC 레코드를 수작업으로 FRBR로 구현하는 것은 거의 불가능하며, 프로그램을 통한 최대한의 자동화 알고리즘 설계의 필요성을 인식하여, OCLC에서는 WorldCat 데이터를 FRBR로 구현하기 위한 저작 세트 알고리즘을 설계하였다.

저작 세트 클러스터링을 위한 알고리즘은 NACO 전거파일에서 저자키와 저자/서명키를, 서지레코드에서는 레코드마다 서명 또는 저자/

서명키를 생성하였다. 이후 NACO에서 생산한 저자키, 저자/서명키와 서지레코드의 키를 비교하여 일치하면 전거의 대표형을 서지레코드의 대표 저작명으로 할당하였다. 할당된 대표키를 전체 서지레코드와 비교하여 저작명이 같으면 동일한 저작 세트로 취급하였다.

전거레코드에서 저자명이나 저자/서명의 대표형과 변형을 추출하기 위해 저자명은 100 ▼a, ▼b, ▼c, ▼d, ▼q에서 대표형을 만들고, 400 ▼a, ▼b, ▼c, ▼d, ▼q에서 이형 형태를 생성하였다. 서명은 100 ▼t, ▼n 130 ▼a에서 서명 대표형을 생성하고, 이형은 400 ▼t, ▼n에서 추출하였다.

서지레코드에서 저자명은 100, 110, 111 ▼a, ▼b, ▼c, ▼d, ▼q에서 추출하고, 서명은 130 ▼a, ▼n, ▼p, ▼s, 240 ▼a, ▼n에서 130, 240이 없는 경우 242, 245, 246, 247 ▼a에서 추출하였다.

저작 세트를 규명하는 키는 4가지 패턴으로 생성되었다. 첫째, 서지레코드가 통일서명을 갖는 경우 통일서명이 대표키가 되었다. 둘째, 저자(1XX)가 있는 경우에는 전거의 저자/서명과 비교하여 대표 형태를 할당하였다. 셋째, 1XX나 7XX는 없고 24X만을 갖는 레코드는 고유키를 갖도록 하기 위해 OCLC 관리번호와 결합시켰다. 넷째, 저자 기본기입이나 통일서명이 없는 경우 본표제와 관련 개별자료를 규명할 수 있는 저자필드로 7XX를 이용하되, 여러 저자를 결합하여 저작을 식별하였다.

2.3.2 LC 디스플레이 툴

LC 알고리즘은 디지털화된 형태의 자료 생산이 증대되면서 여러 버전의 자료를 저작, 표

현형, 구현형으로 효과적으로 디스플레이하기 위한 알고리즘이다.

검색된 저작을 디스플레이하기 위해 우선 저작 수준의 매칭을 실시하였다. 서명과 저자명이 일치하는 저작을 모으기 위해 저자명(1XX)과 서명(2XX)을 추출해 서지레코드를 상호 매칭하였고, 저자명과 서명이 같을 경우 동일 저작으로 우선 클러스터링하였다. 저작수준에서 서명은 130, 240, 243, 245를 저자명은 100, 110, 111을 사용하였다. 표현형 단계의 클러스터링은 동일 저작 내에서 표현형으로 클러스터링하는 단계로 Leader/06과 008/35-37을 이용하였다(Library of Congress 2004).

3. FRBR의 알고리즘 설계

3.1 알고리즘 설계기준

3.1.1 저작 및 표현형 속성 입력 통계 분석

FRBR 저작 클러스터링 알고리즘 설계를 위한 데이터 입력 정도를 파악하기 위해 한성대 도서관의 도서, 전자도서, 비디오 자료 전체를 대상으로 저작과 표현형 개체의 속성에 해당하는 MARC 데이터의 입력통계를 분석하였다.

저작의 서명 속성으로 USMARC의 130 ▼a는 118개(0.2%), 240 ▼a는 1,081개(1.4%), 243 ▼a는 2개(0%) 레코드에만 입력되었고, 245 ▼a는 100% 입력되었다. KORMARC에서는 130 ▼a, 240 ▼a에 입력한 경우는 거의 없으며, 245 ▼a는 모두 입력되었다. 특히 원서명을 기술하는 507은 ▼t가 21,907개(13.3%)가 입력되었으며, 이는 통일표제나 전거데이터

를 사용하지 않는 경우에 번역자료를 통제할 수 있는 도구이다.

저작의 저자명 속성으로 USMARC 100 ▼a는 56,941개(75.7%), 110 ▼a는 1,625개(2.2%), 111 ▼a는 336개(0.4%)가 입력되었다. KOR-MARC의 경우 100 ▼a는 105,668개(63.9%), 110 ▼a는 28,050개(17%), 111 ▼a는 89개(0.1%), 번역자료의 원저자를 기술하는 507 ▼a는 26,701개(16.2%)로 나타났다. 따라서 약 81%는 저작명으로 저자 + 서명을 키로 사용할 수 있고, 약 19%의 레코드는 저자 기본 기입을 사용하지 않기 때문에 서명만으로 또는 700 ▼a, 710 ▼a, 711 ▼a 부출저자명을 이용해 야 할 것으로 분석되었다.

저작의 형식을 기술하는 ▼k, 저작권 관련 일자를 기술하는 ▼f는 거의 입력되지 않았다. 따라서 현 목록데이터에서 저작의 식별을 위한 요소는 서명과 저자명이며, MARC 데이터에는 FRBR 구현을 위해 필요한 데이터가 완벽히 기술되지 않았다.

표현형의 속성으로 언어는 008의 35-37, 041 ▼a, ▼h, 130 ▼l, 240 ▼l, 개정은 250 ▼a이다. 통계 분석결과 008은 100% 입력되었으나 나머지는 입력이 낮았다. 개정을 기술하는 250 ▼a가 동서는 7,821개(4.7%), 양서는 9,567(12.7%)개 입력되었다.

서지레코드 간의 관계를 지시하는 데이터의 입력 정도도 매우 낮았다. FRBR의 가장 큰 장점은 자료 간의 관계 표현이지만 이를 현실적으로 구체화시키기 위해서는 입력 데이터가 매우 부족하였다.

3.1.2 정규화 및 매칭 규칙

130 통일서명이 있는 경우 서명만으로 색인을 작성하였다. 1XX가 있는 경우 서명 + 저자명으로 색인을 작성하였다. 1XX에 저자명이 없어 7XX 부출저자 사용 시 서명 + 저자명(2인)을 각각 색인 작성하고, 저자 2인 중 하나만 일치해도 동일 저작으로 취급하였다. 비디오 자료는 서명만으로 색인을 생성하였다.

서명 색인 작성 시 스페이스, 말줄임표 등 특수기호는 생략하였다. 양서의 정관사 및 관사는 생략하였고, and와 &는 동일한 것으로 취급하였다. 동서의 관계는 제외하고 색인하였다.

저자명은 성 이름 순으로 색인하였고, 서양 인명의 경우 성으로 비교한 후 성이 동일한 경우 이름과 비교하였다. 이름의 경우 약어 형태도 인정하여 William, Shakespeare와 William, S를 동일 저자로 처리하였다.

서명만 갖는 저작은 서명 색인에서 비교하였다. 서명과 저자를 갖는 저작은 서명/저자명 색인에서 비교하였다.

3.1.3 기타 특징

저작의 개념을 목록을 중심으로 동일한 서명과 저자를 갖는 것으로 정의하여 각색, 후속자료 등은 모두 다른 저작으로 취급하였다.

전거레코드를 사용하지 않아 동일 저자의 여러 이형 서명 등이 검색되지 않는 문제점을 해결하고자 국내 번역서에 대해서는 507 원서명을 색인으로 추출하였다.

합집 및 선집은 종합표제를 가진 경우는 종합표제를 저작명으로 처리하고, 종합표제가 없는 저작은 첫 번째 저작으로 저작 색인키를 생성하였다.

3.2 저작 및 표현형 클러스터링 알고리즘

3.2.1 저작 클러스터링

3.2.1.1 저작 대표키 생성 알고리즘

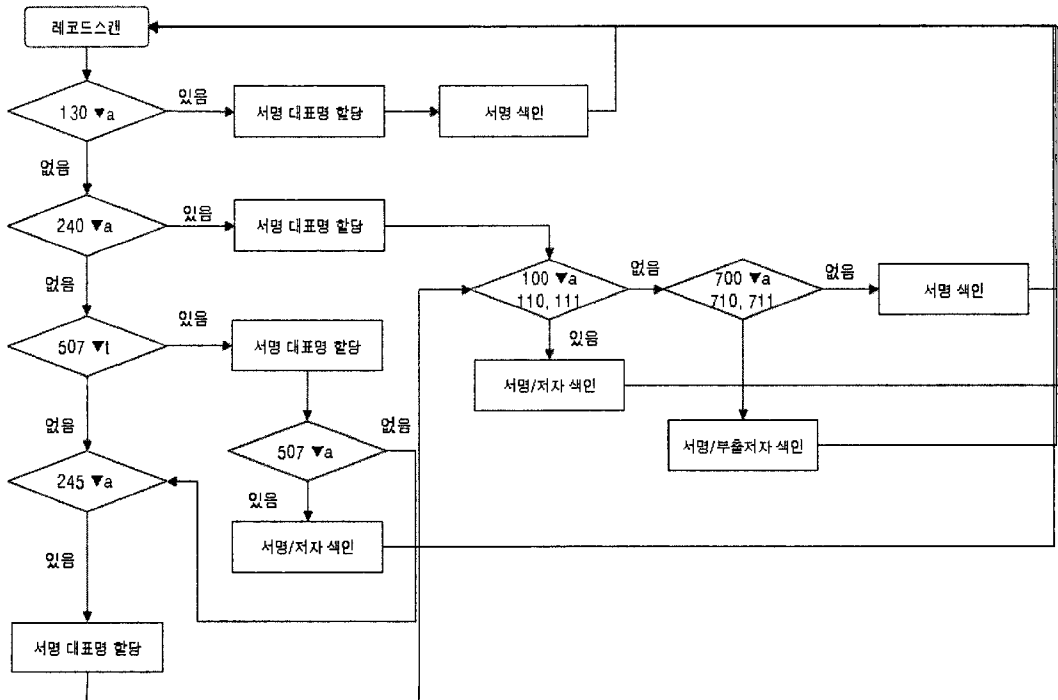
FRBR 구현을 위한 저작 클러스터링 알고리즘은 <그림 1>과 같다. 130 ▼a 통일서명을 갖는 자료는 통일서명으로 색인을 추출하고 종료하였다.

130 ▼a가 없는 경우 240 ▼a 값을 체크하고 이를 서명의 대표명으로 처리하였다. 저자명은 100 ▼a, 110 ▼a, ▼b, 111 ▼a를 추출하여 서명과 저자명을 조합한 서명/저자 색인을 생성하였다. 저자 기본기입이 없는 경우는 700 ▼a, 710 ▼a, ▼b, 711 ▼a 중에서 2개까지만 추출해 서명/부출저자 색인을 생성하였다.

240 ▼a가 없는 경우에는 507 ▼t를 서명으로 추출하고 507 ▼a에서 저자명을 추출해 색인키를 작성하였다.

507 ▼a나 ▼t가 없는 경우 245 ▼a에서 서명을 추출하고, 100 ▼a, 또는 110 ▼a, ▼b, 또는 111 ▼a를 추출하여 서명/저자 색인키를 생성하였다. 1XX가 없는 경우 700 ▼a, 710 ▼a, ▼b, 711 ▼a에서 서명/부출저자 색인을 생성하였다. 700 ▼a에서 색인을 작성할 경우 저자명을 2개까지 추출하되 서명이 같고 저자명이 한 개라도 같은 경우 동일 저작으로 취급하였다.

색인키 생성을 위한 MARC 서브필드로 <표 1>과 같이 서명의 경우 130, 240, 245 ▼a, 507 ▼t이며, 저자명은 1XX, 7XX, 507 ▼a를 사용하였다.



<그림 1> 저작 클러스터링 알고리즘 순서도

〈표 1〉 동서 / 양서 저작 색인 대상 서브필드

	동 서	양 서
서 명	130▼a 240▼a 245▼a 507▼t	130▼a 240▼a 245▼a
저 자	(개인) 100▼a 507▼a	(개인) 100▼a▼d
	(단체) 110▼a▼b 507▼a	(단체) 110▼a▼b
	(회의명) 111▼a	(회의명) 111▼a
	부출저자의 경우 700▼a 710▼a▼b 711▼a	부출저자의 경우 700▼a▼d 710▼a▼b 711▼a

알고리즘을 통해 서지레코드 마다 저작 대표 키를 생성하며, 추출되는 내용은 서명, 저자명, 해당하는 서지번호로 서지번호를 통해 MARC 레코드를 참조하였다.

[예] 저작 대표키: work_key(저작명) = 서지 번호

work_key(Little Prince/Saint) = 54930292
 work_key(Little Prince/Saint) = 20043435
 work_key(Little Prince/Saint) = 3455464

3.2.1.2 저작 세트 생성

서지레코드마다 생성된 대표 색인키를 하나 씩 모두 비교하여 동일 저작일 경우 같은 저작으로 묶고, 다른 경우는 새로운 저작으로 처리하였다.

통일서명을 갖는 자료는 서명 색인에서 비교하고, 서명/저자명을 갖는 자료는 서명과 저자명 색인에서 비교하였다. 둘 이상의 저자를 갖는 부출 저자와 서명색인의 경우 서명/저자명 색인에서 비교하였다.

비교 결과 저작 대표키가 동일한 경우 동일 저작에 해당하는 레코드의 서지번호를 나열한 저작 세트 리스트가 생성된다.

[예] 저작 세트키: work_set(저작명) = 서지번호

호.....

work_set(Little Prince) = 234453, 345465, 78977
 work_set(Little Prince/Saint) = 54930292, 20043435, 3455464

3.2.2 표현형 클러스터링

동일 저작이 만들어진 후 표현형 집중을 위한 색인을 생성하였다. 동일 저작 내 언어와 개정사항이 표현형 식별 정보가 된다. 언어를 구분하기 위해 008/35-37의 언어 코드를 중심으로 표현형 데이터를 추출하되, 해당 데이터가 없는 경우 041 ▼a에서 데이터를 추출하였다. 개정에 따른 표현형 구분을 위해서는 250 ▼a에서 숫자만을 추출하였다. 예를 들어, 『Little Prince/Saint』라는 저작 내에서 3개의 레코드 중 영어가 1개, 한국어 자료가 2개이며, 한국어 자료는 각각 1판과 2판일 경우 영어1판, 한국어 1판, 한국어2판으로 총 3개의 표현형 대표키가 생성되었다.

표현형 세트키는 동일 언어이면서 개정일 경우 해당하는 서지 번호가 생성되었다.

[예] 표현형 대표키: expression_key(저작명, 언어, 개정) = 서지번호
 expression_key(Little prince/Saint, ENG,

1) = 54930292
 expression_key(Little prince/Saint, KOR,
 1) = 20043435
 expression_key(Little prince/Saint, KOR,
 2) = 3455464

[예] 표현형 세트키: expression_set(저작명,
 언어, 개정) = 서지번호, 서지번호.....
 expression_set(Little prince/Saint, ENG,
 1) = 54930292
 expression_set(Little prince/Saint, KOR,
 1) = 20043435
 expression_set(Little prince/Saint, KOR,
 2) = 345546

4. FRBR의 온라인목록 적용

4.1 FRBR 저작 집중성

FRBR 저작 집중성은 저작을 중심으로 서지 레코드가 집중되는 정도에 대한 분석으로 총 저작의 수, 저작당 구현형 수, 저작당 표현형 수이다.

4.1.1 저작당 구현형

저작 및 표현형 클러스터링 알고리즘으로 단행본, 전자도서, 비디오 자료를 처리한 결과 전

체 304,861개의 서지레코드가 263,342개의 저작으로 클러스터링되어, 저작당 평균 1.2개의 구현형 레코드를 갖고 있었다. <표 2>와 같이 263,342개의 저작 중에서 2개 이상의 구현형 레코드를 갖는 저작은 15,985개로 전체 저작의 6%를 차지하며, 단일 구현형을 갖는 저작 수는 총 247,357개로 94%를 차지하여 단일 구현형을 갖는 저작이 대부분이었다. 2개 이상의 구현형을 갖는 저작에 포함된 총 구현형 레코드 수는 57,504개로 전체 서지레코드의 약 19%에 해당하였다. 나머지 81%는 저작이면서 1개의 구현형만을 갖는 단일 구현형 레코드였다.

분석 결과를 다른 연구와 비교하면 <표 3>과 같이 저작당 구현형 수는 높지 않으며, 2개 이상의 구현형을 갖는 저작의 수는 연구자의 연구방법과 저작의 범위에 따라 차이가 있었다.

저작당 구현형의 개수를 조사한 결과는 <표 4>와 같다. 구현형 1개만을 갖는 저작이 247,357개(94%)를 차지하고, 100개 이상의 구현형을 갖는 저작은 15개로 나타났다. 구현형이 많은 상위 1%의 저작은 4~5개 이상의 구현형을 포함하고 있었다.

분석 결과 『장길산/황석영』, 『삼국유사/일연』, 『아리랑/조정래』, 『토지/박경리』가 상위 저작이었다.

4.1.2 저작당 표현형

저작당 표현형 수를 조사한 결과 <표 5>와 같

<표 2> 저작 수와 구현형 수

	2개 이상 구현형 저작	1개 구현형 저작	총 수
저작	15,985 (6%)	247,357 (94%)	263,342 (100%)
구현형수	57,504 (19%)	247,357 (81%)	304,861 (100%)

〈표 3〉 저작 수 비교

연구자	한성대 (2008)	Hickey & O'Neill(2005)	Bennett, Lavoie & O'Neill(2003)	김정현 (2007b)
특징	전수조사-알고리즘처리	전수조사-알고리즘처리	샘플조사+알고리즘	샘플조사+수작업
2개이상 구현형 저작	15,985 (6%)		약 7,000,000 (22%)	16.9%
1개 구현형 저작	247,357 (94%)	36,330,501 (96%)	약 25,000,000 (78%)	82.5%
총저작수	263,342 (100%)	37,772,687 (100%)	32,000,000 (100%)	100.0%
저작당구현형수	1.2개	1.3개	1.5개	1.45개

〈표 4〉 저작당 구현형 수

저작당 구현형 수	저작 수	비율	누적비율
100개이상	15	0.01%	0.01%
99-50개	26	0.01%	0.02%
49-40개	31	0.01%	0.03%
39-30개	44	0.02%	0.05%
29-20개	116	0.04%	0.09%
19-10개	536	0.20%	0.29%
9개-5개	1,464	0.50%	0.79%
4-2개	13,753	5.22%	6.01%
1개	247,357	93.93%	100.00%
합계	263,342	100%	-

〈표 5〉 저작당 표현형 수

저작당 표현형	저작 수	비율	누적비율
1개	261,323	99.23%	99.23%
2개	1,752	0.67%	99.90%
3개	215	0.08%	99.98%
4개	30	0.01%	99.99%
5개	18	0.01%	100.00%
6개	3	0.00%	100.00%
7개	1	0.00%	100.00%
합계	263,342	100.00%	-

이 표현형 수는 1개에서 7개까지 있었으며, 표현형 수가 하나인 저작은 261,323개로 전체 저작의 99%를 차지하고, 2개 이상의 표현형을 갖는 저작은 2,019개로 적었다. 저작당 표현형 수는 평균 1개로 분석되었다.

4.2 저작 클러스터링 누락 및 오류 분석

오류 및 누락 레코드의 원인을 파악하기 위해 주제별로 구현형을 많이 갖는 상위 저작 총 100개를 선정하고, 이를 수작업으로 재검색하

여 시스템 처리에서 누락되었거나 오류로 처리된 구현형 수를 파악하였다. 또한 MARC 상세 내용 분석을 통해 클러스터링되지 않은 문제점을 목록 측면에서 분석하였다.

〈표 6〉과 같이 알고리즘 적용으로 동일 저작으로 클러스터링된 레코드는 1,068개지만, 수작업 검색으로 저작 클러스터링되어야 할 레코드는 1,309개로 44개 저작에 241개 레코드가 누락되었다. 241개를 주제별로 살펴보면 역사/지리, 문학 분야 자료가 가장 많았다. 오류 클러스터링 레코드 수는 6건으로, 대부분 저작 식별이 정확하지 않아서 발생한 문제였다.

표현형은 시스템 알고리즘으로 처리하였다. 분석된 100개 저작이 총 156개의 표현형을 포함하여 상위 저작의 저작당 표현형 수는 약 1.5개였다.

241개의 누락 레코드의 원인 분석 결과 〈표 7〉과 같이 MARC 데이터 필드의 해당 사항 미입력(34%) > 통일표제 미사용(22%) > MARC

데이터 필드 사용 오류(17.4%) > 저자 및 서명 변경(9.2%) > 역자·편지를 저자로 기술(7.5%) > 입력 오류(4.1%) > 기술과 접근에 대한 인식 부재(2.9%) 등의 순서로 나타났다.

MARC 필드의 미입력은 34%로 나타났다. 예를 들어, 번역서의 경우 원서명과 원저자명을 기술할 수 있는 507을 사용할 경우 원서명으로 집중될 수 있지만 필수 항목이 아니기 때문에 미입력되었거나, 서명만 기재한 경우, 저자명만 기재한 경우가 있었다.

통일표제를 사용하지 않아 동일 저작이 클러스터링되지 못한 경우는 22%였다. 예를 들어, 『백범일지』의 경우 245 ▼a에는 자료에 기술된 그대로 『백범일지 영인』, 『직해 백범일지』 등으로 기술했으나 이에 대한 통일표제가 없어 『백범일지』로 해당하는 자료를 집중시키지 못했다. 또한 네루다의 『세계사 편력』도 국내 번역시 『세계사 이야기』, 『역사이야기』 등으로 출간되었으나 통일표제가 없어 동일 저작으로

〈표 6〉 누락 및 오류 클러스터링 분석

분야	저작 수	시스템 처리 레코드수	수작업검색 레코드수	누락 저작 수/ 레코드수	오류 저작 수/ 레코드수	표현형수
000	7	46	66	3 / 20	0 / 0	17
100	11	95	125	8 / 30	0 / 0	15
200	11	63	82	3 / 19	2 / 4	12
300	10	238	246	4 / 8	0 / 0	10
400	10	55	57	2 / 2	0 / 0	22
500	11	66	80	4 / 14	0 / 0	14
600	10	83	90	2 / 7	0 / 0	11
700	10	62	69	4 / 7	1 / 1	27
800	10	198	253	6 / 55	1 / 1	15
900	10	162	241	8 / 79	0 / 0	13
총계	100	1,068	1,309	44 / 241	4 / 6	156

* 누락클러스터링: 해당 저작으로의 클러스터링이 되지 못함.

** 오류클러스터링: 해당 저작으로의 클러스터링이 옳지 않음.

〈표 7〉 누락 레코드의 원인 분석

누락 원인		누락 레코드	비율
목록규칙 적용 오류	저자 및 서명 변경	22	9.2%
	역자, 편자를 저자로 기술	18	7.5%
	통일표제 미사용	53	22.0%
	합집 자료 처리 미비	2	0.8%
	본표제 선정에러	1	0.4%
	기술, 접근 인식 미흡	7	2.9%
MARC 사용오류	MARC 필드 미입력	82	34.0%
	MARC 필드 사용오류	42	17.4%
전거데이터 미구축에 따른 문제		4	1.7%
입력 오류(스펠링 오류 포함)		10	4.1%
계		241	100%

모이지 않았다.

MARC 데이터 필드 사용 오류는 17.4%였다. 예를 들어, 서명 다음에 MARC 서브필드 없이 부서명이나 권차를 입력하여 시스템 처리로는 저작을 식별할 수 없었다.

저자나 서명이 변경되어 클러스터링이 수행되지 않은 경우가 9.2%를 차지하였는데, 특히 개정판 저자 변경에 따른 누락이 많았다. 예를 들어, 『Applied Linear Statistical Models』 1판부터 4판까지는 Neter가, 5판은 Kutner로 기본 기입 저자가 변경되어 동일 저작으로 집중되지 않았다.

역자나 편자를 저자로 기술한 경우는 7.5%로 나타났다. 예를 들어, 역자를 저자로 기술한 『삼국지』는 하나의 저작으로 집중되지 못했다.

4.1% 레코드는 입력 오류(스펠링 오류)로 저작으로 집중되지 않았다.

기술과 접근에 대한 인식의 부재로 예러가 발생한 경우가 2.9%였다. 접근과 기술의 측면에서 100, 700 등을 접근점으로 일관성있게 기술하지 못해 저작으로 집중되지 않았다. 그외

에도전거데이터, 합집자료의 처리 미비 등도 클러스터링 누락의 원인이었다.

따라서 FRBR 적용을 위해서 목록규칙과 규칙에 따른 목록 적용 지침과 MARC 기술 방안이 필요하다.

4.3 FRBR 적용시 목록규칙의 문제

4.3.1 저작 식별

통일표제나 기본표목을 사용하지 않을 경우, 저작을 집중시키기 어려웠다. 예를 들어, 성경의 경우, 통일표제가 사용되지 않아 개별적인 서명으로 저작 식별기가 생성되어 저작으로의 집중이 높지 않았다.

KCR4에서는 통일표제나 기본표목은 불필요한 기능으로 보고, 전거에서 처리하도록 하였다. AACR2에서는 통일표제와 기본표목을 모두 허용하며, 기술하도록 지침을 두었다. AACR3이라 할 수 있는 RDA 초안에서는 저작의 집중 및 통일을 위해 저작과 표현형을 위한 우선 접근점(preferred access points)을 사용하도록 하였다.

통일표제는 동일한 저작의 둘 또는 그 이상의 구현형들이 서로 다른 본표제를 가지고 있을 때 그와 같은 구현형들에 대한 저록을 함께 모아주는 표제로 전자데이터를 구축하지 않는 상황에서 저작의 통제를 위해 필요한 기능이며, 또한 전거가 구축된 상황에서도 정확한 저작 식별을 위해서는 통일표제와 기본표목의 기능을 고려해야 한다.

4.3.2 개정자료 식별

목록규칙에서 동일 저작 내 개정판에 대해서 개정판들이 서로 다른 표제를 가진 경우라도 통일표제를 사용하지 않고, 상호 부출로서 연계시키도록 하였다. 이에 따라 개정자료의 서명이나 저자명이 변경된 경우 개정 전 저작과 집중될 수 없었다.

4.3.3 번역자료 식별

목록규칙에서 번역자료는 통일표제를 기술하여 원서명과 번역서명을 집중시키도록 하였다. 번역자료의 통일을 위해 통일표제를 사용해야 하지만 분석된 서지데이터의 경우 통일표제를 사용하지 않고, 원서명 주기(507 ▼a 원저자명, ▼t 원서명)에만 기술한 경우가 많았다. 하지만 원서명주기는 필수요소가 아니기 때문에 입력에 한계가 있었으며 원저작으로의 집중이 용이하지 않았다.

4.3.4 합집저작 식별

종합표제가 있는 경우는 종합표제로 저작명을 색인하지만, 포함된 개별저작으로 저작키를 식별하기 위해서는 목차정보나 7XX에 분출된 개별 저작으로 색인해야 한다. 하지만, 현재 대

부분의 MARC 데이터에는 합집이나 선집 등의 모음 저작을 지시하는 정보가 없어 이를 처리하기 어려웠다. 또한 레코드에 포함된 모든 부출 및 분출 정보와 목차를 색인하여 매칭할 경우는 오히려 저작 클러스터링의 오류가 커질 수 있다. 따라서 합집 선집의 모음 저작에 포함된 개별 저작으로 저작키를 생성하기 위해 모음 저작임을 지시하는 목록 기술이 필요하다.

4.3.5 표현형 식별

표현형을 위한 식별정보는 주로 개정정보와 번역정보이다. 개정 정보는 250 ▼a에 기술되고 있으나 모호한 사항이 많아 표현형을 정확하게 기술할 수 없어, 표현형을 위한 기술항목이 별도로 필요하다. 예를 들어, '2nd edition', '3rd edition' 이외에도 'New Edition', 'International Edition'과 같은 정보를 기술하기 위해서는 개정에 대한 별도 기술이 필요하다. 따라서 표현형을 체계적으로 구분하기 위해서는 이를 정확히 구분할 수 있는 기술요소의 입력이 중요하다.

4.3.6 자료의 내용과 매체

FRBR에서는 자료의 내용과 유형에 따라 표현형 속성을 달리하고, 이용자에게 자료 유형에 따른 필터링을 제공하기 때문에 자료의 내용과 매체에 대한 범주가 매우 중요하다. 하지만, GMD(General Material Designation)와 SMD(Specific Material Designation)를 혼동하여 기술하고 있었다. 이는 GMD와 SMD의 리스트 내에 표현형, 구현형이 섞여 있으며 개념이 중복이 되어 판단하기 어렵기 때문이었다. 둘째, 기존의 GMD나 SMD 입력정보로 해당하는 자료를 식별해 정확한 표현형 속성을 찾

기 어려웠다.

이에 대해 Tom Delsey는 GMD 용어가 자원의 물리적 포맷과 계층, 매체 형태가 혼합되어 있어 GMD에는 내용적인 측면, SMD는 물리적 매체로 둘 것을 제안하였다(Tillet 2002 재인용). RDA 초안에서는 자료의 내용과 매체를 구분하고, 내용 유형은 GMD로 형식은 SMD로 기술하였다. 특히 내용 측면에서 지도 자료를 지도 데이터셋, 지도 이미지, 지도 동영상 이미지 등으로 세부적으로 구분하였고, 컴퓨터 파일도 컴퓨터 데이터셋, 컴퓨터 프로그램을 구분하였다. 동영상에서도 일반 동영상과 3차원 동영상을 구분하고, 음악자료는 악보, 연주곡, 일반 사운드, 강연으로 구분하여 일반음악과 새소리 등의 소리, 각종 강연을 각각 구분하여 기술할 수 있도록 하였다. 매체의 대분류는 오디오, 컴퓨터, 마이크로폼, 현미경, 영상, 스테레오, 비디오, 중개 장치 사용하지 않는 것으로 나누어 매체의 유형별로 대구분이 명확하게 이루어졌다.

4.3.7 책임사항 기술

AACR2에서는 번역자료의 경우 원저작자를 책임저자로 보고 번역 저자는 부출시키도록 지시하고 있으나 이것이 올바르게 적용되지 않아 저작으로 집중되지 않았다. 예를 들어, 『삼국지』의 경우 대부분 나관중을 원작으로 하고 있어, 저자명 기술 시 100에 나관중, 700에 역자를 부출해야 하지만, 일부 자료는 100에 역자를 기술하고 원저작자명은 기술하지 않아 저작의 집중이 떨어지는 문제가 있었다.

4.3.8 파트 자료

일부 자료는 부서명이나 부제, 또는 파트번

호와 파트명이 포함되어야만 저작이 구분되는 경우가 있었다. 그러나 저작을 구분하기 위해 파트 서명이나 부서명을 저작명에 포함시킬 경우 오히려 저작 클러스터링이 낮아질 수 있었다. 따라서 저작으로의 집중을 위해 245 ▼a로만 저작을 식별하는 것도 바람직하다.

5. FRBR을 위한 목록규칙

5.1 목록규칙 및 MARC 기술방안

FRBR 적용을 위한 목록 규칙과 MARC 기술 방안을 다음과 같이 제안한다.

5.1.1 저작 식별 정보의 기술

- 통일표제와 기본표목을 기술하여 저작 식별을 용이하게 한다.
- 저작명은 개인, 단체의 책임사항 우선 접근점과 우선 서명이다.
- 저자명이 없는 저작은 기본저록 통일표제를, 저자명이 있으면서 통일표제가 필요한 자료는 저자표제 통일표제로 기술한다.

저작의 정확한 식별을 위해 통일표제와 기본표목이 필요하며, 특히 전거를 사용하지 않는 기관에서는 필수적이다. 저자명이 없는 경우, 통일표제로 130 ▼a를 사용하고, 저자명이 있는 경우, 240 ▼a를 사용한다.

- 저자명이 불확실한 경우 서명만으로 기술한다.
- 화상자료는 통일표제를 사용하여 저자없이 저작명만으로 우선 서명을 기술한다.

- 2인 이상의 공저자를 갖는 저작물인 경우, 서명으로 우선 서명을 기술하고, 저자는 모두 부출로 기술하되, 각 저자별로 역할어를 기술한다.
- 편집물은 편집장과 편집자료의 서명으로 기술하며, 편집자가 저자가 아닌 경우에는 저자와 서명으로 저작명을 작성하고, 저자를 파악할 수 없는 경우 서명만으로 기술한다.
- 파트 자료의 경우, 파트 서명을 기술하며, 저작 클러스터링에서는 전체 저작을 나타내는 저자명과 서명으로 저작명을 생성한다.
- 각색 저작은 원본과 구분하기 위해 각색자를 기본 책임사항으로 선정한다.

예를 들면, 박경리의 『토지』 원작과 만화를 식별하기 위해서 각색 저작의 경우 박경리로 저자를 기술하지 않고, 각색의 주요 책임저자를 기술한다.

[예]

100 1 ▼a박경리,▼d1926-
245 20 ▼a(만화) 토지 / ▼d박경리 원작; ▼e오세영 각색·그림.

[재목록]

100 1 ▼a오세영
245 20 ▼a(만화) 토지: ▼b박경리 원작

- 개정자료는 이전 자료의 서명과 저자명을 통일표제와 저자 기본기입으로 기술하고 개정자료의 서명과 저자명은 부출시킨다.
- 축약자료의 경우, 원작의 서명과 저자명을

통일표제와 저자 기본기입으로 기술하고 축약자료의 서명과 저자명은 부출시킨다.

- 번역자료의 경우, 원작의 서명과 저자명을 통일표제와 저자 기본기입으로 기술하고, 번역자료의 서명과 저자명은 부출시킨다.

5.1.2 표현형 식별 정보의 기술

- 개정자료의 경우 표현형 식별을 위해 개정 정보를 기술한다. 개정에 대한 사항은 개정 판차를 숫자로 기술하고, 'new edition' 과 같은 사항은 개정으로 취급하지 않는다. 만일, 번역자료인 경우 원본의 개정에 따라 개정정보를 기술한다.
- 축약자료는 축약정보를 기술한다.
- 번역자료의 상세기술을 위해 기술하는 자료의 언어를 기술한다.

개정사항의 기술 시에는 130, 240, 245 ▼s 버전정보에 ▼s[revision 2nd] ▼s[revision 5th] 등으로 기술할 수 있다. 축약자료의 기술 시에는 ▼s[abridged]로 기술한다. 번역자료의 상세기술을 위해 MARC에서는 표제 다음에 ▼l을 이용해 ▼l[English] ▼l[Korean] ▼l[Japan] 등으로 기술한다.

5.1.3 합집 기술

- 합집의 처리는 종합표제를 사용하는 경우와 개별 저작을 사용하는 경우로 나뉘며, 접근점은 포함된 개별 저작으로 식별한다. 따라서 합집을 표시하기 위해 [Collected works] [Selections]를 기술한 후에 포함된 개별 저작으로 저작명을 작성하도록 분출한다.

5.1.4 자료의 내용/매체 기술

- 자료의 내용, 매체, 상세 매체의 리스트는 RDA의 구분을 기준으로 하여, 내용(contents), 매체(media), 상세 매체(carrier)를 기술한다. 내용은 리더 06에, 매체유형은 MARC의 130, 24X ▼h에 해당하며, 상세 매체 유형은 SMD로 007/01로 매핑된다고 볼 수 있다.

5.2 실험데이터의 목록규칙 검증

5.2.1 검증방법

제안한 목록규칙으로 재목록하기 위해 실험 데이터를 별도로 구축하고 관련 모듈을 설치하였다. 실험 데이터베이스에 중복된 서지데이터 및 에러로 미반입된 데이터 9건을 제외한 총 1,300개의 레코드를 반입하였다.

제안한 목록규칙에 따라 저작과 표현형 식별 정보를 추가하여 1,300건의 데이터를 모두 재목록하였다.

재목록이 완료된 후 알고리즘을 일부 수정하여 실험 데이터에 적용한 후 이전 목록의 클러스터링과 비교하였다.

비교한 결과 100개 저작에 1,068개 레코드가 클러스터링되었던 이전 결과와 비교해 106개의 저작에 1,300개의 구형형 레코드가 모두 클러스터링되었다. 저작 수는 106개로 증가되었으나 이는 오류 클러스터링을 해결했기 때문이었고, 누락 레코드가 모두 적절하게 집중되었다.

5.2.2 목록규칙의 적용

누락된 클러스터링을 해결하기 위해 재목록시에 통일표제와 저자 기본기입을 사용하였다.

개정자료의 경우 개정 전 서명을 통일표제에 기술하고, 245에는 변경된 서명을 기술하였는데 예를 들면, 『Applied Linear Statistical Models』는 원저자인 Neter, John으로 저자를 기술하고 개정판의 저자를 부출하였다.

번역자료의 경우 통일표제를 기술하여 『오만과편견』, 『모비딕』 등의 번역서가 원저작명으로 모두 클러스터링되었다.

합집 등의 모음 저작의 경우 [collected works]를 통일표제로 기술함으로 시스템에서 포함된 개별 저작으로 저작을 생성하도록 하여, 7XX분출 저자와 서명이 저작명이 되었다. 이에 따라 『논어』, 『맹자』합집과 『삼국사기』와 『삼국유사』합집은 각각 개별 저작으로 클러스터링되었다.

각색 자료인 만화 『토지』는 저자 기본기입에 각색자를 기술하여 원본과 각색저작을 구분하였다. 예를 들어, 만화 『토지』는 각색자 오세영을 책임 저자로 기술하여 각각 다른 저작으로 클러스터링되도록 하였다.

자료의 내용과 매체 입력시 리더 06, 008, 007은 이미 고정된 기호가 사용되고 있어 실험 데이터 구축에서는 표제 다음에 ▼h에 자료의 내용 - 매체 - 상세 매체를 함께 기술하였다.

기술방식은 <표 8>과 같이 도서의 경우 내용 유형은 text, 매체 유형은 unmediated, 매체 타입은 volume으로 기술하였다. 온라인 동영상 자료의 경우 내용유형은 moving image이고, 매체유형은 computer, 상세 매체 유형은 online resource으로 기술하였다. 비디오의 경우, 내용 유형은 moving image, 매체는 video, 상세 매체는 videodisc로 구분하였다. 따라서 각각 245 ▼h[text-unmediated-volume], ▼h[moving image-video-videodisc]로 기술하였다.

〈표 8〉 자료유형별 내용 / 매체 기술 방식

구분	도서	온라인 동영상	비디오	지도
Content type (Leader 06)	text	moving image	moving image	cartographic image
Media type (24X▼h 007/00)	unmediated	computer	video	unmediated
Carrier type (SMD 007/01)	volume	online resource	videodisc	unmediated
Extent (300 ▼a)	24 pages	1 online resource	1 videodisc	1 map on 2 sheets

출처: JSC. 2007a. RDA: Resource Description and Access Part A - Constituency Review of March 2007 Draft of Chapter 3. <<http://www.collectionscanada.gc.ca/jsc/docs/5rda-parta-ch3rev.pdf>> 재구성.

6. 결론

지금까지 저작 및 표현형 클러스터링 알고리즘을 설계하여 FRBR의 저작 집중성의 정도와 저작 집중성이 낮은 문제점을 분석하고, 이를 높이기 위한 방안으로 목록규칙과 MARC 기술방안에 대한 지침을 마련하고 검증하였다.

본 연구를 통해 분석된 연구 내용을 종합하면 다음과 같다. 첫째, 한성대학교 도서관 온라인목록에 저작 및 표현형 클러스터링 알고리즘을 적용해 FRBR 저작 집중성을 분석한 결과 304,861개의 레코드가 263,342개 저작으로 집중되었고, 저작당 평균 1.2개의 구현형 레코드를 포함하였다.

둘째, FRBR의 저작 클러스터링이 효과적으로 수행되지 못한 원인으로 데이터 품질의 문제는 목록규칙 적용 시의 오류, MARC 사용 오류, 입력 오류의 문제였다.

셋째, FRBR에 맞는 데이터 구축을 위해 목

록규칙 및 기술방안을 제안하고 이를 검증한 결과 106개 저작에 1,300개 레코드 모두 클러스터링이 수행되었다.

제안한 목록규칙 적용 전후의 저작 클러스터링의 차이는 결국 데이터 구축이 FRBR 구현에 가장 중요한 부분임을 보여주었다. MARC 레코드가 일관성 있게 작성되었다면, FRBR의 저작 및 표현형의 집중도가 높을 수 있음을 보여주었다.

본 연구를 통해 제안한 FRBR을 위한 목록규칙은 FRBR을 보다 구체적이며 실질적으로 적용할 수 있도록 할 것이며, KCR4의 FRBR을 위한 목록규칙 개정에도 도움이 될 것이다. 또한 전거데이터가 없는 상태에서 알고리즘을 고안하였으나, 전거를 구축한 기관에서도 저작대표키 생성 시에 전거와 매칭하는 모듈만 삽입한다면 FRBR 알고리즘으로 사용할 수 있을 것이다.

참 고 문 헌

- 국립중앙도서관. 2007. 『표준 메타데이터 모델 및 semantic language 연구개발』. 서울: 국립중앙도서관.
- 김정현. 2007. 한국어 서지레코드에 있어 FRBR 모형의 유용성에 관한 연구. 『한국문헌정보학회지』, 41(4): 295-314.
- 김현희, 유영준, 박서은. 2007. FRBR 모형의 KORMARC 데이터베이스로의 적용가능성에 대한 실험적 연구. 『한국도서관정보학회지』, 38(2): 185-202.
- Bennett, Rick, Brian F. Lavoie & Edward T. O'Neill. 2003. "The concept of a work in WorldCat: an application of FRBR." *Library Collections, Acquisitions & Technical Services*, 27: 45-59.
- Cho, J. 2006. "A study on the application method of the Functional Requirements for Bibliographic Records(FRBR) to the online public access catalog (OPAC) in Korean libraries." *Library Collections Acquisitions, & Technical Services*, 30: 202-213.
- Hegna, K. and E. Murtomaa. 2002. *Data mining MARC to Find: FRBR?*, [online]. [cited 2004.3.10]. <<http://folk.uio.no/knuthe/dok/frbr/datamining.pdf>>.
- Hickey, T. B. and E. T. O'Neill. 2005. "FRBRizing OCLC's WorldCat." *Cataloging & Classification Quarterly*, 39(3/4): 239-251.
- Hickey, T. B., E. T. O'Neill, and J. Toves. 2002. "Experiments with the IFLA Functional Requirements for Bibliographic Records(FRBR)." *D-Lib Magazine*, 8(9). [online]. [cited by 2007.06.05]. <<http://www.dlib.org/dlib/september02/hickey/09hickey.html>>.
- IFLA. 1998. 『서지레코드의 기능상의 요건』. 김태수 역. 서울: 국립중앙도서관.
- JSC. 2007. RDA: Resource Description and Access Part A - Constituency Review of March 2007 Draft of Chapter 3. [online]. [cited 2007.10.3]. <<http://www.collectionscanada.gc.ca/jsc/docs/5rda-parta-ch3rev.pdf>>.
- Network Development and MARC Standards Office, Library of Congress. 2004. *FRBR Display Tool Version 2.0* [online]. [cited 2007.7.12]. <<http://www.loc.gov/marc/marc-functional-analysis/tool.html>>.
- Svenonius, E. 2000. *The Intellectual Foundation of Information Organization*. Boston: MIT Press.
- Tillett, B. B. 2002. General Material Designations. [online]. [cited 2005.7.12]. <<http://www.collectionscanada.gc.ca/jsc/docs/gmd.pdf>>.
- Tillett, B. B. 2005. "FRBR and cataloging for the future." *Cataloging & Classification Quarterly*, 39(3/4): 197-205.