

국사편찬위원회 한국근현대잡지자료 데이터(2024.03.27.)*

Data on Modern and Contemporary Korean Magazines by the National Institute of Korean History

김바로(Kim, Ba-ro)**

 0000-0002-4456-9616

목차

1. 서론
2. 한국근현대잡지자료 RAWDATA 취득
3. 한국근현대잡지자료 RAWDATA 탐색
 - 3.1 전체구조
 - 3.2 <Level1> 스키마
 - 3.3 <Level2> 스키마
 - 3.4 <Level3> 스키마
 - 3.5 기사본문(text) 스키마
 - 3.6 개선점
4. 결론

초록

본 논문은 국사편찬위원회에서 발행한 근현대잡지자료의 RAWDATA 를 취득하고 해당 데이터를 설계하고 구성을 탐색했다. 근현대잡지자료의 RAWDATA 를 얻기 위하여, 공공데이터 포털의 공공데이터 제공 신청과 문서 24 를 통한 데이터 제공 신청을 행하였다. 그 결과 2024 년 3 월 27 일 기준의 근현대잡지자료의 RAWDATA 를 취득하였다.

근현대잡지자료 RAWDATA 는 기본적으로 국사편찬위원회의 표준 XML 스키마(history.dtd)를 따르고 있다. <Level1>에서는 잡지 정보, <Level2>에서는 권호 정보, <Level3>에서는 개별기사 정보를 다루고 있다. 개별 기사의 본문은 <paragraph> 단위로 분리되어 있다. 문맥요소에는 index(객체명), emph(강조), pTitle(제목), name(저작자명), illustration(그림), tableGroup(표)이 존재한다. 다만, 현재 모든 잡지에 대한 본문 정보를 제공하고 있지 않기에 본문 텍스트가 있는 데이터에 대해서만 문맥요소가 제한적으로 존재한다.

근현대잡지자료 RAWDATA 는 근대 문인 언어 분석과 근대 문인 사회네트워크 분석에서 사용할 수 있을 뿐만 아니라, 나아가 근대 문학을 위한 형태소 분석기의 토대 데이터와 현대 한국어로의 번역 등 다양한 영역에서 사용할 수 있으리라 생각하며, 문학 연구자의 집단지성으로 한국근현대잡지자료의 RAWDATA 가 더욱 풍성해지기를 기대한다.

주제어: 한국근현대잡지자료, RAWDATA, 공공데이터, 데이터 리뷰, 디지털인문학

* 본 연구의 핵심인 근현대잡지자료 RAWDATA 를 편찬하고 제공해주신 국사편찬위원회에 감사의 말씀을 드립니다.

** 단독저자: 한국학중앙연구원, 인문정보학·인문지리학(인문정보학), 조교수, ddokbaro@gmail.com

Abstract

This paper acquires and explores the data schema of the RAWDATA of Korean Modern and Contemporary Magazine Materials from the National Institute of Korean History (NIKH). To acquire the RAWDATA, a request for public data provision was submitted through the Public Data Portal, and a request for data provision was submitted through Docu24. As of March 27, 2024, the RAWDATA of Korean Modern and Contemporary Magazine Materials was acquired.

The RAWDATA of Korean Modern and Contemporary Magazine Materials basically follows the NIKH standard XML schema (history.dtd). <Level1> deals with magazine information, <Level2> deals with volume information, and <Level3> deals with individual article information. The body of each article is divided into <paragraph> units. Contextual elements include index (object name), emph(emphasis), pTitle(title), name (author name), illustration(figure), and tableGroup(table), but they are currently only available for data that has body text, as not all magazines currently provide body text information. The RAWDATA of Korean Modern and Contemporary Magazine Materials can be used for analysis of modern literary language and modern literary social networks. It is also expected to be used in various fields such as the foundation data for morpheme analysis tools for modern literature and translation into modern Korean. We hope that the RAWDATA of Korean Modern and Contemporary Magazine Materials will become even richer through the collective intelligence of literary scholars.

Keywords: Korean Modern and Contemporary Magazine Materials, RAWDATA, Public Data, Data Review, Digital Humanities

1. 서론

본 논문은 국사편찬위원회의 한국근현대잡지자료의 RAWDATA를 취득하는 과정에서부터 시작하여 한국근현대잡지자료의 데이터 구조를 탐색하고 그 활용 가능성을 살펴보고자 한다. 국사편찬위원회의 한국근현대잡지자료는 한국사데이터베이스에서 제공하는 데이터베이스 중 하나이며, 근대 문학 연구, 특히 디지털인문학의 영역에서 진행되는 근대 문학 연구에서 자주 사용하고 있는 데이터베이스이다.¹

한국근현대잡지자료는 대한제국기에 발행된 학회지 12종(『기호흥학회월보』, 『대동학회월보』, 『대조선독립협회회보』, 『대한유학생회학보』, 『대한자강회월보』, 『대한학회월보』, 『대한협회회보』, 『대한흥학보』, 『서북학회월보』, 『서우』, 『태극학보』, 『호남학보』)과 일제시기에 발행된 19종(『개벽』, 『대동아』, 『동광』, 『만국부인』, 『별건곤』, 『삼천리』, 『삼천리문학』의 7종 등)에 대해서는 원문 전체 텍스트를 제공하고 있으나², 아직은 많은 잡지들이 목차만 제공하고 있다.³

디지털인문학 연구를 위해서는 최대한 많은 양의 RAWDATA를 확보해야 한다. 안타깝게도 지금까지는 저작권의 문제로 인하여 개인 연구자들이 필요에 따라서 스크래핑을 통하여 소량의 데이터를 수집하여 연구를 진행해왔다.

이러한 스크래핑을 통한 연구 방식의 문제점은 첫째, 스크래핑을 수행하기 어려운 연구자들이 데이터 수집에 어려움을 겪어 연구의 기본을 다지기 어렵다는 점이다. 둘째로, 표면적인 텍스트 정보만을 수집하는 스크래핑으로는 한국근현대잡지자료의 본문 텍스트에서 식별한 인물, 단체, 관서, 지명, 기타 등의 의미 정보와 emph(강조), illustration(그림), tableGroup(표)의 형식 정보를 활용할 수가 없다는 것이다.

이에, 본 연구에서는 여러 연구자들의 보다 심도 깊은 연구를 위하여 한국근현대잡지자료의 RAWDATA를 취득하는 과정을 소개하고, 이를 사용하기 위한 준비로서 현 시점의 RAWDATA의 구조와 내용에 대해서 탐색하고자 한다. 또한 한국근현대잡지자료의 데이터를 활용할 수 있는 다양한 샘플 코드⁴를 제공한다.

2. 한국근현대잡지자료 RAWDATA 취득

연구자는 2024년 2월 15일 공공데이터 포털의 공공데이터 제공 신청을 통하여, 국사편찬위원회에 한국근현대잡지자료의 전체 RAWDATA를 요청했다.⁵ 2024년 3월 7일 국사편찬위원회는 저작권 소멸 시효(저자 사후 70년) 문제가 있기에 1910년 이전 간행 잡지 12종에 대해서만 RAWDATA 제공을 하였다.⁶ 이에 연구자는 2024년 3월 13일 문서²⁴⁷를 통해서, 개인 연구자 자격으로 국사편찬위원회에 직접 공문을 발송했다.⁸ 2024년 3월 15일 국사편찬위원회에서는 비영리적 학술연구로 이용 제한이 있는 전체 RAWDATA의 제공을 결정하였고⁹, 2024년 3월 27일¹⁰ 이메일로 RAWDATA 전체¹¹를 전송하였다.¹²

3. 한국근현대잡지자료 RAWDATA 탐색

3.1 전체구조

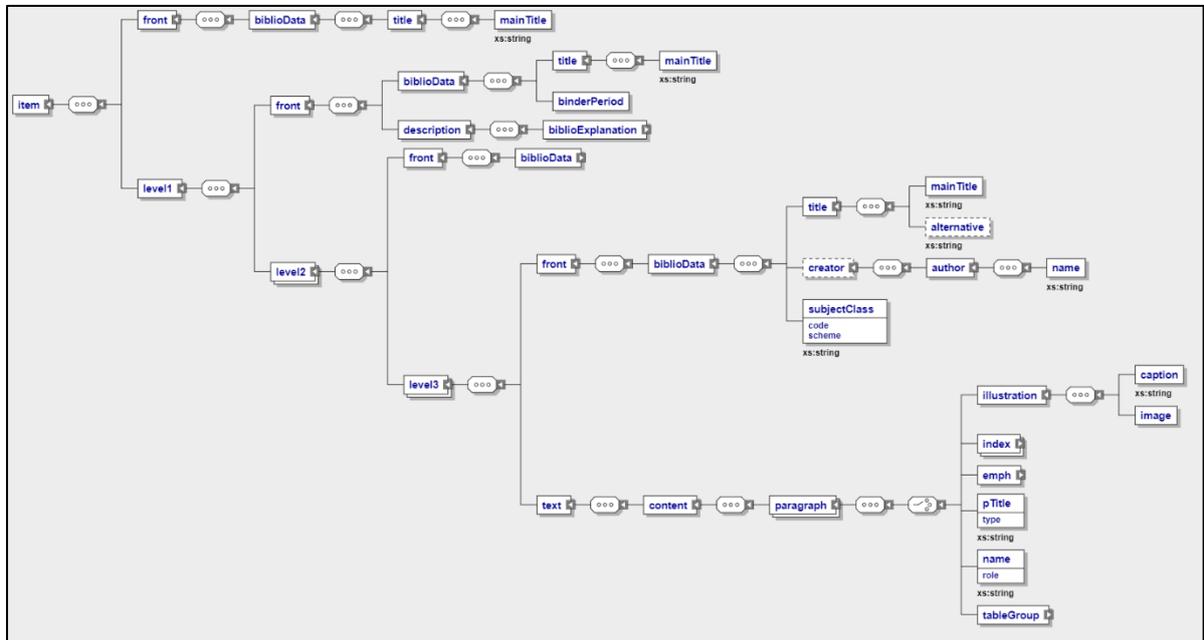


그림 1. 한국근현대잡지자료 데이터 스키마 개념도¹³

한국근현대잡지자료 RAWDATA는 기본적으로 국사편찬위원회 표준 XML 스키마(history.dtd)를 따르고 있다. 다만, 국사편찬위원회 표준 XML 스키마(history.dtd)는 디지털 조선왕조실록을 비롯하여 국사편찬위원회에서 제공하는 다양한 데이터베이스를 포괄하기에, 실제로는 한국근현대잡지자료에서는 history.dtd의 일부만을 사용하고 있다.

<Level1>에서는 잡지 정보를, <Level2>에서는 권호 정보를, <Level3>에서는 개별기사 정보를 다루고 있다. 한국근현대잡지자료의 ID 부여 규칙은 한국근현대잡지자료의 고유ID인 MA에 추가적인 일련번호를 부여하는 것이다. <Level1>의 잡지 정보에서는 “MA_001”, “MA_002” 형식으로, <Level2>의 권호 정보에서는 “ma_001_0010”, “ma_001_0020” 형식으로, <Level3>의 개별기사 정보에서는 “ma_001_0010_0010”, “ma_001_0010_0020” 형식으로 일련번호를 부여하고 있다.

3.2 <Level1> 스키마

<Level1>에서는 잡지이름(mainTitle), 창간일(binderPeriod@begin), 폐간일(binderPeriod@end), 해제정보(biblioExplanation) 등의 메타정보와 <Level2>의 권호들을 제공하고 있다.

```

<level1 id="ma_001">
  <front>
    <biblioData type="원문제공잡지">
      <title>
        <mainTitle>대한자강회월보</mainTitle>
      </title>
      <binderPeriod begin="1906-07-31" end="1907-07-25"/>
    </biblioData>
    <description>
      <biblioExplanation groupID="01" groupName="대한제국기 잡지 해제" id="ma_001_exp" name="대한자강회월보 大韓自強會月報">
        <content>
          <paragraph>대한자강회의 기관지. 편집인 겸 발행인은 김상범(1-4호), 이종준(5-9호), 심의성(10-12호), 현은(13호) 등이 차례로 역임했다. 1906년 7월 31일 창간하였고 매월 25일 월간으로 1907년 7월 25일까지 총 13호를 발행했다. 가격은 15錢으로서 서울과 인천, 개성, 해주, 철산 등에서 판매했다. 1976년 한국학문헌연구소 편집으로 아세아문화사에서 전 2권으로 영인되었다. 대한자강회월보는 1906년 4월에 창립된 대한자강회의 기관지였다. 대한자강회는 윤효정(尹孝定), 장지연(張志淵), 임진수(林珍洙), 김상범(金相範) 등 20여 명이, 1905년 창립된 헌정연구회(憲政研究會)를 확대시켜 만든 단체이다. 회장은 윤치호(尹致昊)였으며 그 밑에 부회장 1명, 평의원 20명, 간사원 20명, 일본인 고문 1명을 둔다고 규정하였다. 대한자강회는 헌정연구회의 주축인물들이 중심이 되어 창립되었으나 전자가 소수 지식인 단체였던데 비하여 전국적 대중조직(1907년 현재 전국 25개 지회)을 지향하였다는 점에서 본격적인 계몽운동의 시발점을 이루는 단체였다. 대한자강회에는 계몽운동 전반을 포괄하는 다양한 인물들이 참여하였으며, 다른 계몽운동단체, 학회 설립의 매개가 되기도 하였다. 대한자강회가 내세운 취지는 자강을 목표로 교육과 식산흥업을 주요 실천과제로 강조하였다. 이런 대한자강회의 취지는 이후 다른 계몽운동단체들도 거의 비슷하게 따랐으며 한말 계몽운동론의 핵심이 되었다. 대한자강회의 활동은 취지에서도 보이듯이 실력양성론의 입장에서 국권회복을 전망하는 것이었으며, 합법적 영역에서 계몽운동을 추구하는 것이었다. 그리고 이런 입장 때문에 황무지개척, 국채보상 문제 등에서는 일제의 침략에 대해 날카로운 비판을 가하면서도 의병활동에는 지극히 부정적인 양면성을 가졌다. 대한자강회월보는 대한자강회의 이런 실력양성론, 문명개화론적 입장을 선전하는 중요수단이었다. 구성을 보면 논설부, 교육부, 식산부, 정치부, 법률부, 역사지리부, 문예부, 기서, 회록, 내지회보, 외국정보, 회원동정, 회원명부, 광고로 되어있다. 구성에서 보이듯 상당한 짜임새를 가지고 있었으며 교육, 식산흥업 등에 관한 다양한 계몽논설이 꾸준히 게재되었고, 국문과 역사지리에 대한

```

글들도 게재되었다. 특히 정치부분에서의 입헌군주제에 대한 논의와 법률부에서의 만국공법(萬國公法), 개인의 법적권리에 관한 글들은 진일보한 근대사상의 단면을 보여주는 것이었다. 그러나 대체적으로 근대화와 대중계몽의 필요성에 주목하면서도, 제국주의 침략은 정확히 인식하지 못하는 한계를 보였으며, 또한 그런 점에서 이후 계몽운동론의 전형이 되는 글들이 많이 게재되었다. 대한자강회월보는 창간 이래 매달 꾸준히 발간되었으나 1907년 8월 한일신협약(韓日新協約)에 반대하는 대한자강회 회원들의 활동과 관련하여 회가 강제해산됨에 따라 중단되었다. 대한자강회월보는 게재된 글들의 성격을 하나로 정리할 수는 없지만 가장 대표적이며 선구적인 계몽운동단체의 기관지라는 점에서 계몽운동세력의 전형적인 근대화론, 국권회복론을 검토할 수 있는 귀중한 자료라고 할 수 있다.</paragraph>

</content>

</biblioExplanation>

</description>

</front>

3.3 <Level2> 스키마

<Level2>에서는 권호제목(mainTitle), 발행자(publisher), 발행일(dateIssued) 등의 메타정보와 <Level3>의 기사들을 제공하고 있다.

```

<level2 id="ma_001_0010">
  <front>
    <biblioData>
      <title>
        <mainTitle    volume="제1호">대한자강회월보    제1호
      </mainTitle>
      </title>
      <publication>
        <publisher>대한자강회</publisher>
        <dateIssued  date="1906-07-31">1906년    7월    31일
      </dateIssued>
      </publication>
    </biblioData>
  </front>

```

3.4. <Level3> 스키마

<Level3>에서는 기사제목(mainTitle), 창작자(creator), 형태분류(subjectClass) 등의 메타정보와 기사본문(text)을 제공하고 있다.

```

<level3 id="ma_001_0010_0030">
  <front>
    <biblioData>
      <title>
        <mainTitle>本會會報</mainTitle>
      </title>
      <creator>
        <author>
          <name>尹孝定 編纂</name>
        </author>
      </creator>
      <subjectClass code="ma04" scheme="madoc">소식
</subjectClass>
    </biblioData>
  </front>

```

다만, 형태 분류는 본문텍스트를 제공하는 19종의 잡지에서만 제공하고 있으며, 총 16종으로 539,710건이 있다. 세부 수치는 다음과 같다.

표 1. 형태 분류 수량

분류	개수	분류	개수
논설	90510	대담·좌담	20800
소식	71305	설문	20759
소설	62707	문예평론	16732
문예기타	60483	기행문	13756
잡저	53372	사고·편집후기	11874
시	46265	세태비평	9327
회고·수기	26594	학술	8382
희곡·시나리오	26436	만화	408

3.5 기사본문(text) 스키마

기사본문(text)은 단락(paragraph) 단위로 분리되어 있으며, 문맥요소로 pTitle(제목), name(저작자명), index(객체명), emph(강조), illustration(그림), tableGroup(표)를 제공하고 있다. 다만, 기

사의 제목과 저작자명은 전체 잡지에서 제공을 하고 있으나, 객체명, 강조, 그림, 표는 본문전체 텍스트가 있는 19종의 잡지에서만 제공하고 있다.

index(객체명)은 총 131,337건으로, 세분화하여 이름 88,398건, 지명 29,594건, 단체 6,454건, 기타 5,448건, 관서 1,443건이 있다. 다음과 같은 방식으로 데이터에서 표현하며, 웹서비스에서는 별도의 처리를 하지 않고 있다.

```
<index type="이름">尹致昊</index>
<index type="지명">明洞</index>
<index type="단체">自強會</index>
<index type="기타">印度</index>
<index type="관서">法部</index>
```

emph(강조)는 문장의 강조 표시를 데이터화한 것으로 총 8,108건이 있으며, 그 중 밑줄(underline)이 5,714건, 굵은 표시(bold)가 2,394건이다. 다음과 같은 방식으로 데이터에서 표현하며, 웹서비스에서는 별도의 처리를 하지 않고 있다.

```
<emph align="left" type="underline">堯時君命漢時臣</emph>이 蓋譏之也로
다.</paragraph>
```

웹서비스주소: http://db.history.go.kr/id/ma_001_0020_0070

illustration(그림)은 텍스트로 전달이 불가능한 이미지를 표현하기 위하여 사용하고 있으며, 총 1,508건이 있다. 다음과 같은 방식으로 데이터에서 표현하며, 웹서비스에서는 별도의 이미지 저장소에서 해당 이미지를 연결하는 방식을 취하고 있다.

```
<illustration>
<caption>系統圖</caption>
<image src="ku_ma_01_08_0021a" type="gif"/>
</illustration>
```

웹서비스주소: http://db.history.go.kr/id/ma_001_0080_0090

이미지URL:

https://db.history.go.kr/download.do?fileName=ku_ma_01_08_0021a.gif&levelId=ma_001_0080_0090

tableGroup(표)은 표 형태의 정보를 제공하기 위하여 사용하고 있으며, 총 109건이 있다. 다음과 같은 방식으로 데이터에서 표현하며, 웹서비스에서는 HTML의 표 출력 방식으로 제공하고 있다.

```
<paragraph>
```

```

<tableGroup>
  <caption>六大洲의 大小 方里로 擧야 表示</caption>
  <table>
    <tr>
      <td>漢字韓音</td>
      <td>英音韓文</td>
    </tr>
    <tr>
      <td>아세아주 <index type="지명">亞細亞</index>洲</td>
      <td>에시아</td>
    </tr>

—  중간 생략 —

    <tr>
      <td>남아미리카주 南<index type="지명">阿美利加洲
</index></td>
      <td>사우프아메리카</td>
    </tr>
  </table>
</tableGroup> 方里 (英里)
</paragraph>

웹서비스주소: http://db.history.go.kr/id/ma_001_0110_0160

```

3.6 개선점

<index>와 <name>은 고유객체이다. 이상적으로는 이에 고윳값(ID)를 부여하여 동명이인(同名異人)이나 동인이명(同人異名)과 같은 객체를 다르게 처리하여야 한다. 이를 통해서, 연구자들은 추가 작업 없이, 다양한 데이터에 접근할 수 있게 될 것이다. 현재 디지털 조선왕조실록에서는 <index>의 “ref” 속성을 통해서 고윳값을 부여하고 있으며 이는 연구 자료 수집을 용이하게 한다. 한국근현대잡지자료에도 객체의 고윳값이 적용된다면 신속한 연구 자료 수집에 도움이 될 것이다.

* 한국근현대잡지자료의 <index>와 <name>

<index type="이름">吳暹植</index>

<name>南嵩山人 張志淵</name>

* 디지털 조선왕조실록의 <index>

```
<paragraph align="center">○罷吏曹參議<index num="1088411_0" ref="M_0000934"
sort="K" type="이름">姜士弼</index>。 <index num="1088412_0" ref="M_0000934"
sort="K" type="이름">士弼</index>素無才德，而爲時輩所推，備歷臺閣，出身未十年，已陞承旨
。
<name role="필자">尹孝定 編纂</name>
```

웹서비스주소: 디지털 조선왕조실록. "선조수정실록 2권, 선조 1년 2월 1일辛巳 2번째기사".
http://sillok.history.go.kr/id/wnb_10102001_002

그런데 고깃값의 부여는 실제로 방대한 밀작업을 수반하며, 해당 시기의 연구자조차 익숙한 영역이 아니라면 온전히 고깃값을 부여하는 동일 식별 작업을 수행하기가 어려운 측면이 있다. 또한, 국사편찬위원회의 각고의 노력으로 한국근현대잡지자료가 지속적으로 구축되고 있지만, 아직 많은 잡지들의 본문이 플레인 텍스트(plain text)로 만들어지지 못하고 있는 실정이다. 이를 해결하는 방안으로 가장 간단한 방법은 국사편찬위원회의 관련 예산 증액을 통해서 사업을 확대 및 지속하는 것이다. 그런데 어떠한 사업이든 오타자부터 오식별까지 다양한 문제가 발생할 수밖에 없는 것도 사실이다.

현실적으로 제한된 예산과 인력에서 해결책으로 집단지성을 고민할 필요가 있다. 2000년대에 이미 전문가들로만 구축한 브리태니커 백과사전보다 위키백과의 정보의 깊이와 범위가 더 높은 수준이었으며,¹⁴ 2016년 9월부터 국립국어원에서 개통한 사용자 참여형 온라인 사전인 우리말샘¹⁵은 성공적으로 안착하여 다양한 분야에서 활발하게 활용되고 있다.¹⁶ 2024년부터는 국립중앙도서관에서 한국 위키미디어 협회와 업무 협약을 맺고, 국립중앙도서관이 소장한 디지털화 자료 일부는 위키미디어가 진행하는 '옛 한글 문헌 전자화 프로젝트'를 통해 텍스트 데이터로 재탄생할 예정이며, 2024년 9월부터는 AI-OCR 플랫폼을 통해 저작권이 해결된 디지털 원문을 이용자가 직접 인공지능(AI) 기술을 활용해 광학문자인식(OCR)으로 변환하고 텍스트로 활용할 수 있도록 할 예정이다.¹⁷

한국근현대잡지자료도 연구자들에게 다양한 편찬을 직접 수행할 수 있는 플랫폼을 제공하고, 그 편찬에 대한 공로를 인정할 수 있는 편찬자 명기 등의 방안을 모색한다면, 한국근현대잡지자료의 데이터가 보다 풍성해지고 심도 깊어질 것이다.

4. 결론

본 논문은 국사편찬위원회 근현대잡지자료 RAWDATA를 취득하고, 데이터의 구조와 세부 내용에 대해서 탐색을 수행했다. 한국근현대잡지자료 RAWDATA는 다양한 층위의 근대 문학 연구에 활용될 수 있는 자료이다. 디지털인문학의 분석 방법론을 활용하여, 근대 문학 잡지별, 문인별, 형태분류별 등의 다양한 방법으로 언어 분석을 수행할 수 있으리라 기대한다.

또한, 창작자와 본문에 등장하는 주요 객체(인물, 단체, 판서, 지명, 기타)를 활용하여, 근대 문

학자 사회네트워크를 살펴볼 수도 있다고 본다. 그뿐만이 아니라, KIWI¹⁸의 Subword Tokenizer¹⁹를 활용하여, 근대 문학 텍스트에서 활용할 수 있는 형태소 분석기를 만들 수도 있으며, 근대 문학 텍스트를 현대 한국어로 전환하는 인공지능 번역기까지도 개발이 가능할 것으로 보인다. 무엇보다 RAWDATA 접근을 통하여, 기계가독형데이터(machine readable data)의 가능성을 인지하고, 문학자를 위한, 문학자에 의한, 문학자의 문학 RAWDATA 설계와 구축이 보다 활성화될 수 있다고 믿는다.

¹ 이재연(2016). “`생활`과 `태도` - 기계가 읽은 『개벽』과 『조선문단』의 작품 비평어와 비평가 -”. <개념과 소통>. 5-52. <https://www.riss.kr/link?id=A102687767>; 이재연(2016). “키워드와 네트워크: 토픽 모델링으로 본 『개벽』의 주제 지도 분석”. <상허학보>. 277-334. <https://www.riss.kr/link?id=A103645649>; 최지명(2018). “기계학습을 이용한 역사 텍스트의 저자판별: 1920년대 개벽 잡지의 논설 텍스트”. <언어와 정보>. 91-122. <https://www.riss.kr/link?id=A105121622>; 전성규, 김병준(2019). “디지털인문학 방법론을 통한 『서북학회월보』와 『태극학보』의 담론적 상관관계 연구”. <개념과 소통>. 141-188. <https://www.riss.kr/link?id=A106279126>; 李載然, 鄭瑜炅(2020). “국문학 내 문학사회학과 멀리서 읽기 - 새로운 검열연구를 위한 길마중”. <大東文化研究>. 295-337. <https://www.riss.kr/link?id=A107071534>; 전성규(2020). “근대계몽기 잡지 자연어 전처리 작업에서 문체 분류와 형태분석 방안에 대한 모색”. <개념과 소통>. 33-72. <https://www.riss.kr/link?id=A106928580>; 조은경, 드프레메리웨인(2021). “『동광』의 언어적 근대성 -데이터과학적 분석-”. <언어와 문화>. 359-382. <https://www.riss.kr/link?id=A107846076>; 조은경, 드프레메리웨인(2021). “『별건곤』의 텍스트 근대성: 데이터과학적 언어분석”. <문화와 융합>. 35-54. <https://www.riss.kr/link?id=A107889002>; 전성규(2023). “한국 근대 잡지의 계량적 연구 방법에 대한 논의: 코퍼스 구축 및 데이터 분석의 사례를 중심으로”. <민족문화사연구>. 159-210. <https://www.riss.kr/link?id=A108736463>

² 현재 데이터 기준은 “ma_001.xml”부터 “ma_019”의 데이터이다.

³ 국사편찬위원회. “한국근현대잡지자료 소개”. https://db.history.go.kr/introduction/modern/intro_ma.html

⁴ 김바로. “Github KoreanModernMagazine”. <https://github.com/ddokbaro/KoreanModernMagazine>; 김바로. “한국근현대잡지자료_데이터탐색.ipynb”. https://colab.research.google.com/drive/1bK1zoo_ezIxx7fOPE1tvo4o75eH6-BXk; 본 코드는 Colab 환경에서 작성되었으며, Python 3.10.12, Pandas 2.0.3, BeautifulSoup 4.12.3를 활용했다.

⁵ 김바로. “공공데이터신청 상세 페이지”. https://github.com/ddokbaro/KoreanModernMagazine/blob/main/00_문건모음_공공데이터/20240215_공공데이터제공신청_상세.png

⁶ 공공데이터포털. “교육부 국사편찬위원회_한국사데이터베이스 정보_근현대잡지자료_대한제국시기_20240304”. <https://www.data.go.kr/data/15127033/fileData.do>

⁷ 행정안전부. “문서 24”. <https://docu.gdoc.go.kr/>

⁸ 김바로. “20240313_국사편찬위원회_학술연구를 위한 한국근현대잡지자료 RAWDATA 신청”. https://github.com/ddokbaro/KoreanModernMagazine/blob/main/00_문건모음_공공데이터/20240313_국사편찬위원회_학술연구를 위한 한국근현대잡지자료 RAWDATA 신청.pdf

⁹ 국사편찬위원회. “20240315_공문_학술연구를 위한 한국근현대잡지자료 RAWDATA 신청 회신”. https://github.com/ddokbaro/KoreanModernMagazine/blob/main/00_문건모음_공공데이터/20240315_공문_학술연구를 위한 한국근현대잡지자료 RAWDATA 신청 회신.pdf

¹⁰ 제공 결정과 실제 수령에 간극이 존재하는 것은 신규 구축 데이터를 포함하여 제공을 하였기 때문이다.

¹¹ 국사편찬위원회. “20240327_제공데이터_한국근현대잡지자료(ma) RAW DATA_DTD+XML(2023 구축 반영)”. [https://github.com/ddokbaro/KoreanModernMagazine/blob/main/00_문건모음_공공데이터/20240327_제공데이터_한국근현대잡지자료\(ma\) RAW DATA_DTD+XML\(2023_구축_반영\).zip](https://github.com/ddokbaro/KoreanModernMagazine/blob/main/00_문건모음_공공데이터/20240327_제공데이터_한국근현대잡지자료(ma) RAW DATA_DTD+XML(2023_구축_반영).zip)

¹² 공공데이터 신청 방법과 제논의는 다음 논문에 있다. 김바로(2022). “<공공데이터법>과 인문데이터-공공기관 보유 인문데이터 공개 신청 사례를 중심으로”. <한국고전연구>. 167-192. <https://www.riss.kr/link?id=A108145939>

¹³ 본 개념도를 마이크로소프트 비주얼 스튜디오 2019의 “스키마 만들기” 기능을 통하여 “ma_001.xml”의 내용을 XSD 파일로 생성한 이후에, <http://www.xml-tools.net/>을 통해서 작성했다.

¹⁴ 위키백과. “위키백과의 신뢰도”. https://ko.wikipedia.org/wiki/위키백과의_신뢰도

¹⁵ 국립국어원, “우리말샘”. <http://opendict.korean.go.kr>

¹⁶ 뉴스시스. “배꼽인사·아웃렛·얼음땡 등 500 개 어휘 ‘표준국어대사전’ 추가 등록”. 2023-10-23. https://www.chosun.com/culture-life/culture_general/2023/10/23/3NWTI6NDYJCFTEL4LJZ2PYKOGQ/

¹⁷ 뉴스 1. “"디지털 자료, 텍스트 데이터 활용 물꼬 트여"...'옛 한글 문헌 전자화 프로젝트'”. 2024-03-21. <https://www.news1.kr/articles/?5357660>

¹⁸ Minchul Lee (2022). “bab2min/Kiwi: v0.14.0 (v0.14.0)”. Zenodo. <https://doi.org/10.5281/zenodo.7041425>; 이민철. “Kiwi : 지능형 한국어 형태소 분석기(Korean Intelligent Word Identifier)”. <https://github.com/bab2min/Kiwi>

¹⁹ 이민철. “Module kiwipiemy.sw_tokenizer”. https://bab2min.github.io/kiwipiemy/v0.17.1/kr/sw_tokenizer.html

참고문헌

- 연구데이터

국사편찬위원회. “20240327_제공데이터_한국근현대잡지자료(ma) RAW DATA_DTD+XML(2023 구축 반영)”.
[https://github.com/ddokbaro/KoreanModernMagazine/blob/main/00_문건모음_공공데이터/20240327_제공데이터_한국근현대잡지자료\(ma\) RAW DATA_DTD+XML\(2023 구축 반영\).zip](https://github.com/ddokbaro/KoreanModernMagazine/blob/main/00_문건모음_공공데이터/20240327_제공데이터_한국근현대잡지자료(ma) RAW DATA_DTD+XML(2023 구축 반영).zip)

김바로. “Github KoreanModernMagazine”. <https://github.com/ddokbaro/KoreanModernMagazine>

김바로. “한국근현대잡지자료_데이터탐색.ipynb”.
https://colab.research.google.com/drive/1bK1zoo_ezIxx7fOPE1tvo4o75eH6-BXk

- 논문

김바로(2022). “<공공데이터법>과 인문데이터-공공기관 보유 인문데이터 공개 신청 사례를 중심으로”. <한국고전연구>. 167-192. <https://www.riss.kr/link?id=A108145939>

이재연(2016). “‘생활’과 ‘태도’ - 기계가 읽은 『개벽』과 『조선문단』의 작품 비평어와 비평가 -”. <개념과 소통>. 5-52. <https://www.riss.kr/link?id=A102687767>

이재연(2016). “키워드와 네트워크: 토픽 모델링으로 본 『개벽』의 주제 지도 분석”. <상허학보>. 277-334. <https://www.riss.kr/link?id=A103645649>

李載然, 鄭瑜炅(2020). “국문학 내 문학사회학과 멀리서 읽기 - 새로운 검열연구를 위한 길마중”. <大東文化研究>. 295-337. <https://www.riss.kr/link?id=A107071534>

전성규, 김병준(2019). “디지털인문학 방법론을 통한 『서북학회월보』와 『태극학보』의 담론적 상관관계 연구”. <개념과 소통>. 141-188. <https://www.riss.kr/link?id=A106279126>

전성규(2020). “근대계몽기 잡지 자연어 전처리 작업에서 문체 분류와 형태분석 방안에 대한 모색”. <개념과 소통>. 33-72. <https://www.riss.kr/link?id=A106928580>

전성규(2023). “한국 근대 잡지의 계량적 연구 방법에 대한 논의: 코퍼스 구축 및 데이터 분석의 사례를 중심으로”. <민족문학사연구>. 159-210. <https://www.riss.kr/link?id=A108736463>

조은경, 드프레메리웨인(2021). “『동광』의 언어적 근대성 -데이터과학적 분석-”. <언어와 문화>. 359-382. <https://www.riss.kr/link?id=A107846076>

Data Paper

조은경, 드프레메리웨인(2021). “『별건곤』의 텍스트 근대성: 데이터과학적 언어분석”. <문화와 융합>. 35-54.
<https://www.riss.kr/link?id=A107889002>

최지명(2018). “기계학습을 이용한 역사 텍스트의 저자판별: 1920년대 개벽 잡지의 논설 텍스트”. <언어와 정보>. 91-122. <https://www.riss.kr/link?id=A105121622>

Minchul Lee(2022). “bab2min/Kiwi: v0.14.0 (v0.14.0)”. Zenodo. <https://doi.org/10.5281/zenodo.7041425>

- 신문

뉴스 1. “"디지털 자료, 텍스트 데이터 활용 물꼬 트여"...'옛 한글 문헌 전자화 프로젝트'”. 2024-03-21.
<https://www.news1.kr/articles/?5357660>

뉴스스. “배꼽인사·아웃렛·얼음뎡 등 500개 어휘 ‘표준국어대사전’ 추가 등록”. 2023-10-23.
https://www.chosun.com/culture-life/culture_general/2023/10/23/3NWTI6NDYJCFTEL4LJZ2PYKOGQ/

- 웹사이트

공공데이터포털. “교육부 국사편찬위원회_한국사데이터베이스 정보_근현대잡지자료_대한제국시기_20240304”. <https://www.data.go.kr/data/15127033/fileData.do>

국립국어원, “우리말샘”. <http://opendict.korean.go.kr>

국사편찬위원회. “20240315_공문_학술연구를 위한 한국근현대잡지자료 RAWDATA 신청 회신”.
https://github.com/ddokbaro/KoreanModernMagazine/blob/main/00_문건모음_공공데이터/20240315_공문_학술연구를_위한_한국근현대잡지자료_RAWDATA_신청_회신.pdf

국사편찬위원회. “한국근현대잡지자료 소개”. https://db.history.go.kr/introduction/modern/intro_ma.html

김바로. “20240313_국사편찬위원회_학술연구를 위한 한국근현대잡지자료 RAWDATA 신청”.
https://github.com/ddokbaro/KoreanModernMagazine/blob/main/00_문건모음_공공데이터/20240313_국사편찬위원회_학술연구를_위한_한국근현대잡지자료_RAWDATA_신청.pdf

김바로. “공공데이터신청 상세 페이지”.
https://github.com/ddokbaro/KoreanModernMagazine/blob/main/00_문건모음_공공데이터/20240215_공공데이터제공신청_상세.png

이민철. “Kiwi: 지능형 한국어 형태소 분석기(Korean Intelligent Word Identifier)”. <https://github.com/bab2min/Kiwi>

Data Paper

이민철. “Module kiwipiemy.sw_tokenizer”. https://bab2min.github.io/kiwipiemy/v0.17.1/kr/sw_tokenizer.html