

야담의 데이터, 야담으로부터의 데이터

: 한국 야담 데이터 모델의 구상

양승목* · 류인태**

1. 연구의 배경과 목적
2. 야담의 데이터, 그 층위와 성격
3. 야담 데이터 모델의 구상
4. 성과와 의의

Ⅰ 국문초록

근간된 『정보 한국 야담전집』은 조선후기 주요 야담집 20종을 대상으로 정본화 사업을 수행한 결과물이다. 본 연구는 이 성과를 잇는 후속 작업 중 하나로 ‘한국 야담 데이터 아카이브’ 구축을 목표로 하고 이를 수행하기 위한 구체적 방법을 모색해본 것이다. 조선후기 야담을 대상으로 한 데이터베이스 구현 사례는 아쉽게도 현재까지 없는 상황이다. 이를 고려할 때 『정보 한국 야담전집』의 원문을 플레인 텍스트(plain text) 데이터로 가공해 웹에 공개하는 것만으로도 그 효용이 적지 않을 것이다. 그러나 문자열(string) 기반의 데이터를 제공하는 방식은 야담에 담긴 당대의 다채로운 정보와 그 활물

* 제1저자. 동국대학교 한국문학연구소 전임연구원, zimwoo@naver.com

** 교신저자. 성균관대학교 국어국문학과 박사후연구원, ryuintae@gmail.com

적 성격을 온전히 드러내는 방안이라 하기 어렵다. 야담의 내용과 형식 그리고 야담을 매개로 수집 가능한 다양한 정보와 지식 및 그 문화적 의미를 체계적으로 정리·표현·공유하기 위해서는, 원문·번역문과 같은 단순 텍스트 정보만을 처리하는 문자열 기반의 데이터베이스가 아니라 복잡한 데이터의 군집 기록으로서 야담을 바라보고 유관 정보를 디지털 환경에서 고해상도로 펼쳐내 보일 수 있는 시맨틱 데이터베이스를 구축할 필요가 있다.

이 글은 시맨틱 기반 야담 데이터 아카이브를 구축하기 위한 기초 단계의 모색으로서 조선후기 야담 자료의 형식과 내용을 치밀하게 검토·분석하는 과정에서 진행한 개념적 데이터 모델링에 관한 내용을 담고 있다. 그 연장선상에서 ‘조선후기 야담 온톨로지’ 디자인의 기초 단계에 적용할만한 시맨틱 테크놀로지에 관해 자세히 언급하기보다 이 과정에서 이루어진 야담에 관한 고찰과 고민, 즉 인문학적 사유를 최대한 세밀하게 정리·공유하는 데 집중하였다. 이와 같은 접근은 디지털 환경에서 야담을 본격적으로 다루기 위한 기초 지식과 야담 데이터를 정교하게 처리하기 위한 논의를 점화함으로써 이 논제에 대한 여러 야담 연구자의 안목을 확대하고 유관 문제 의식을 활성화하기 위함이다. 이번 논의에서 인문학 자원을 대상으로 한 데이터 모델링 과정에서 보편적으로 활용되는 컴퓨터와 웹 기술에 관한 논급을 최소화하고, 가급적 야담을 다루는 연구자의 언어로 개념적 데이터 모델링의 과정을 설명하고자 한 이유이다.

주제어 ● 야담, 『정보 한국 야담전집』, 인문학 데이터 모델링, 시맨틱 데이터베이스, 디지털 인문학

1. 연구의 배경과 목적

2021년 8월 『정보 한국 야담전집』이 발간되었다.¹⁾ 조선후기 주요 야담집 20종을 대상으로 저본과 이본을 엄선하고 정밀한 교감 작업을 진행하여 각각의 정본을 구축한 결과물이다. 물론 그동안 전집을 표방하여 야담 자료의 취편을 시도한 예가 없지 않았다.²⁾ 그러나 기존의 성과가 거개 원자료를 영인해 묶은 것임에 반해, 이 책은 모든 자료를 활자화하고 구두점과 교감 내용을 표기한 원문을 제공하고 있다는 점에서 질적인 차이가 있다. 다루고 있는 대상 자료의 범위가 더욱 넓다는 것 또한 변별점으로 짚지 않을 수 없다. 요컨대 『정보 한국 야담전집』의 학적 의의는 조선후기 야담 문학의 전체적인 상을 구축하고 가독성과 활용성을 확보한 기준 텍스트를 마련했다는 데 있다 하겠다. 야담 연구의 새로운 활력을 불어넣을 소중한 자료적 자산을 확보한 것이다.

전통적인 인문학 연구의 프로세스로 보건대, 이 성과의 자연스러운 후속 작업은 크게 두 가지를 떠올릴 수 있다. 정보 원문에 기반을 둔 고품질의 번역물을 생산해서 연구자와 대중에게 제공하는 것, 그리고 이를 바탕으로 정밀하고 심층적인 연구를 진행하여 새로운 학적 통찰을 제시하는 것이다.³⁾

1) 정환국 책임교열, 『정보 한국 야담전집』(전10권), 보고서, 2021. 이 책은 한국학 분야 토대연구지원사업 <조선후기 야담집의 교감 및 정보화>의 최종결과물이며, 동국대학교 한국문학연구소의 자료총서로서 출간되었다.

2) 동국대학교 한국문학연구소 편, 『한국문헌선회전집』(전10권), 태학사, 1981; 박용식·소재영·大谷森繁 편, 『한국야담사화집성』(전5권), 국학자료원, 1995; 정명기, 『한국야담자료집성』(전24권), 계명문화사, 1995 등이 대표적인 성과이다. 이 이전에도 야담을 표제로 삼은 전집류 책이 있긴 하나, 그 용어의 정의와 자료의 범위가 조선후기 야담 문학의 경계를 넘어서 있는 경우가 일반이며 대개 자료적 정확성 내지 정밀성이 확보되지 않은 것인바 따로 명기하지 않는다.

3) 1970년대 이른바 『이조한문단편집』 3권이 출간된 이래 조선후기 야담에 대한 학적 관심은 꾸준히 이어져왔으며, 그에 따라 각종 야담집에 대한 번역본과 연구 성과 또한 적지 않게 제출되어 있다. 그럼에도 『정보 한국 야담전집』이 제공하는 방대하고도 정밀한 원문 자료는 다음 세대의 발전된 결과물을 생산하는 데 중요한 역할을 하

들은 작업의 내용이 서로 다르지만 결국 도서 또는 논문이라는 ‘지면’의 형태로 완성된다는 점에서⁴⁾ 학문적 성과의 창출 내지 확산 방법상 공히 레거시 시스템 속에서 이루어지는 것이라 하겠다. 이는 이대로 고유의 역할과 효용을 갖고 있고 여전히 활발하게 선택되는 방식이지만 그 한계 또한 분명함을 인정하지 않을 수 없다.

20세기 중엽에 발명된 이래 인류의 삶을 획기적으로 바꾸어온 컴퓨터, 또 그에 기초해 발전을 거듭하고 있는 디지털 테크놀로지는 인문학 연구의 장에서도 새로운 물결을 만들었다. 학술자료를 대상으로 한 데이터베이스는 그 가운데서도 가장 큰 축이다. 자료에 대한 접근, 보존, 활용 등 여러 면에서 과거에는 없던 편리함을 불러왔으며, 연구 절차와 환경 또한 대대적으로 바꾸어 놓았다. 특히 ‘검색’이라는 강력한 기능은 기존의 지면 방식이 갖고 있던 일방성과 비효율성을 극복하는 데 매우 큰 기여를 하였다. 그리고 그와 같은 흐름은 현재도 새로운 지평을 열고 있다. 문자열 기반의 검색 알고리즘을 뛰어 넘어 대상 자료에 담긴 수많은 정보와 지식 및 그들 사이의 다채로운 의미적 관계 등을 입체적으로 종합하고, 그에 접근·활용할 수 있게끔 하는 진보된 테크놀로지가 가능한 상황이다. 이른바 시맨틱 테크놀로지 기반의 데이터 아카이브가 그것이다.⁵⁾

본 연구는 ‘한국 야담 데이터 아카이브’ 구축을 『정보 한국 야담전집』의 후속 작업 가운데 하나로 설정하고, 이를 수행하기 위한 기초 단계의 방안을 구체적으로 모색해본 것이다. 지금껏 조선 후기 야담을 대상으로 한 데이

게 될 것이다. 단 보다 정확한 수준의 역주 작업, 참신한 시각에 바탕한 자료의 재편, 야담에 대한 폭넓고도 심층적인 분석 등이 필수적으로 요청된다.

- 4) 실제 종이에 인간하는 것이든 전자책 또는 전자문서의 형태로 공표되는 것이든, ‘지면’이라는 매체적 특성은 본질적으로 다르지 않다.
- 5) 시맨틱 기반 데이터 아카이브의 개념과 구축 및 활용 방법론에 관해서는 다음 논지를 참고할 것. 김현·임영상·김바로, 『디지털 인문학 입문』, 한국외국어대학교 지식출판원, 2016, 151~214쪽; 류인태, 「데이터로 읽는 17세기 재지사족의 일상: 『지암일기(1692-1699)』 데이터베이스 편찬 연구」, 한국학중앙연구원 박사논문, 2019, 1~4쪽.

터베이스는 구현된 적이 없었던바, 『정보 한국 야담전집』의 내용을 텍스트 데이터로 가공해 웹에 개방하는 것만으로도 그 효용이 적지 않을 것이다. 그러나 그와 같은 접근은 당대와 긴밀하게 호흡하며 만들어진 야담의 다채로운 정보와 그 활물적 성격을 온전히 드러내는 방안이라 하기 어렵다. 야담의 내용과 형식, 야담으로부터 채집할 수 있는 정보와 지식을 체계적으로 정리·표현·공유하기 위해서는, 원문 및 번역문 텍스트를 웹에서 공유하는 문자열 기반의 데이터베이스가 아니라 야담을 다채로운 데이터의 근집체로 보고 그것을 디지털 환경에서 높은 해상도로 펼쳐내 드러낼 수 있는 시맨틱 테크놀로지 기반의 데이터 아카이브를 설계할 필요가 있다. 이 글은 그와 같은 시맨틱 기반 야담 데이터 아카이브를 구축하기 위한 기초 단계의 모색으로서, 조선후기 야담 자료의 형식과 내용을 치밀하게 검토·분석하는 과정에서 진행한 개념적 데이터 모델링에 관한 내용을 정리한 것이다. 그리고 ‘조선후기 야담 온톨로지’ 디자인의 기초 단계에 적용할만한 시맨틱 테크놀로지를 제시하기보다, 이 과정에서 이루어진 야담에 관한 고찰, 즉 인문학적 사유를 최대한 세밀하게 정리·공유하고자 하는 데 집중하였다. 디지털 환경에서 야담을 본격적으로 다루기 위한 기초 지식과 야담 데이터를 정교하게 처리하기 위한 논의를 점화함으로써, 이 논제에 대한 여러 야담 연구자의 안목을 확대하고 유관 문제의식을 활성화하는 데 목적을 두고 있기 때문이다.

2. 야담의 데이터, 그 층위와 성격

의미적 차원의 검색과 데이터 활용이 가능한 아카이브를 구축하기 위해서는 기초적으로 대상 자료의 특성을 정교하게 반영한 온톨로지 디자인이 요구된다.⁶⁾ 온톨로지 디자인은 데이터 모델링 과정의 일환으로서 대상 자료의 성격과 포집하고자 하는 데이터에 관한 깊은 이해가 선행되어야 하는

바, 먼저 조선후기 야담이 가진 자료적 특성과 그에 포함되어 있는 데이터의 존재 양상에 대해 살펴볼 필요가 있다.

한국 고전 자료를 디지털 환경에서 재현하는 일은 꽤 오래된 화두이다. 1995년 출시된 <조선왕조실록 CD-ROM>이 그 초기의 산물이라 할 것이요, 이후 웹 환경이 보편화됨으로써 <한국고전종합DB>를 포함한 다양한 목적·형식의 데이터베이스가 만들어져 운용되고 있으며, 현재도 새로운 시스템이 곳곳에서 기획·개발되고 있다. 그만큼 고전 자료를 대상으로 한 데이터 모델에 관한 고민과 실제 데이터베이스 구축에 있어서 적지 않은 경험과 결과물이 축적되어 있으며, 유관 논의와 지침 또한 곳곳에 산재하고 있다. 야담을 대상으로 한 데이터 모델링과 데이터베이스 구축에 관한 고민에 있어서 그와 같은 고전 자료 대상 DB 구축 결과물이 좋은 참고가 될 수는 있겠으나, 해당 시스템에 적용한 데이터 모델을 야담 아카이브에 직접 적용하기에는 그 한계 또한 분명하다.

유관 사례를 모두 짚기에는 지면의 한계가 있기에 저간의 사례를 대체적으로 살펴보자면, 대상 자원이 대부분 역사 기록물에 해당하는 것이 많고, 문학 자료를 직접적 대상으로 표방한 경우에 있어서도 데이터의 층위가 서지 정보나 시소러스 등을 정리하는 데 초점을 둔 결과물이 대부분이기 때문이다.⁷⁾ 그러나 야담은 사실과 허구를 넘나드는 ‘서사물’이라는 점을 고려

-
- 6) 인문학 연구에서의 온톨로지 활용 맥락과 그 구체적 설계 방법에 대해서는 류인태, 위의 글, 64~66쪽의 내용을 참고할 수 있다.
- 7) 국립중앙도서관 제공 <한국고문헌종합목록>, 고전번역원 제공 <한국고전종합DB>, 국사편찬위원회 제공 <한국사데이터베이스>, 한국학중앙연구원 제공 <한국고문서자료관>, 규장각한국학연구원 제공 <규장각 원문검색서비스>, 국학진흥원 제공 <일기류DB> 등이 그에 해당하는 대표적 유관 자원들이다. 문학 또는 역사 방면의 텍스트 자료를 대규모로 다룬다는 측면에서 그와 같은 데이터베이스 자원의 유용함은 분명하다. 한편으로 문자열 검색에 의존한 기존 정보 접근 방식을 리뉴얼하기 위한 방안 모색 또한 꾸준히 이루어지고 있는 상황이다. <한국고전종합DB>를 대상으로 한 최지연·조성덕·최동빈, 「데이터 정제의 관점에서 본 한국문집총간 XML 문서 검토」, 『한문학논집』 60집, 근역한문학회, 2021; 이병찬·민경주, 「한국고전종합DB의 한국문집총간분석 시각화 방안 연구」, 『민족문화』 57호, 2021 등의

해야 하며, 이로부터 정리·표현·공유해야 할 정보와 지식의 층위 또한 단순하지가 않다. 따라서 야담에 최적화된 데이터 모델을 새롭게 고민해야 하는 상황이다.

이와 관련해 보다 가까운 거리에서 참조할 만한 몇 가지 선행 연구가 있다. 먼저 이강욱 교수는 야담의 대중화를 위한 한 방안으로 데이터베이스 구축의 필요성을 강조한 바 있다. 그의 주장 가운데 눈여겨보아야 할 지점은 단순히 텍스트 기반의 데이터 구축이 아니라 야담 문학의 모티프, 인물형, 서사구조 등을 추출한 데이터베이스 구축을 역설했다는 것이다.⁸⁾ 야담을 활용한 스토리텔링과 문화콘텐츠 생산의 활성화를 위한 아이디어이지만, 야담이 가진 서사물로서의 성격을 주시하게 한다는 점에서 경청할 부분이 있다.

또 야담집의 색인 데이터베이스를 구상한 권기성·김동건의 논의도 주목할 만하다.⁹⁾ 이는 개별 야담집에 대한 번역서가 다수 나와 있고 부분적으로 색인이 첨부된 경우도 있지만 본격적인 의미의 통합 색인 시스템이 없다는 것에 문제의식을 갖고, 야담이 갖고 있는 다양한 정보를 표지하여 검색할 수 있는 데이터베이스 구축 방안을 검토한 것이다. 이 논의에서 가장 눈에 띄는 것은 야담 작품과 관련된 정보를 ‘작품정보’와 ‘이야기정보’로 나누어 체계화하고 있다는 점이다. 이 역시 야담의 내용 가운데 서사문학적 의미 자질의 데이터화 작업을 강조했다라는 점에서 특기할 만하다. 단 이 구상은 작품 정보를 메타데이터로 정리하여 카탈로그를 구성하는 방식을 취하고 있는바, 야담 텍스트를 직접적으로 다루는 형태가 아님을 유의할 필요가 있다.

연구가 그에 해당한다.

8) 이강욱, 「야담 연구의 대중화 방안」, 『어문학』 제115집, 한국어문화회, 2012, 303~308쪽.

9) 권기성·김동건, 「야담집 색인 데이터베이스의 구축방안 모색: 『기문총화(記聞叢話)』를 중심으로」, 『고전과 해석』 제22집, 고전문학한문학회, 2017.

야담을 다룬 것은 아니지만 고전의 원전 텍스트를 대상으로 그에 담지된 다양한 층위의 데이터를 체계적으로 분류·추출하는 모델링에 관한 고민을 구체적으로 전개하고 실제 그것을 결과물로 구현하였다는 점에서, 류인태의 연구 또한 시사하는 바가 크다. 특히 『지암일기(支菴日記)』의 데이터를, 자료의 구조에 해당하는 형식요소, 텍스트의 문맥에서 직접 언급되는 내용요소, 내용을 재구성한 의미요소 등 크게 세 영역으로 나누고 개체 간 관계를 정밀하게 정의하는 방식을 수립하였는데,¹⁰⁾ 이러한 접근은 야담(집)의 데이터 기술 체계를 고안하는 데 중요한 지침이 된다.

한국 야담 데이터 모델은 이러한 요구와 착상, 또 방법론과 긴밀하게 대화하며 구상되어야 할 것이다. 이를 염두에 두고 야담의 데이터가 어떤 것들이 있는지, 이들을 어떻게 체계화할지, 또 이 과정에서 유념해야 할 사항은 무엇인지 등을 개괄적으로 점검해보기로 한다.

먼저 형식적 차원이다. 한국 야담 데이터 아카이브는 『정보 한국 야담전집』의 성과를 토대로 기획되는 것인바, 기본적으로 20종의 조선후기 야담집을 대상으로 하며¹¹⁾ 정보으로 확정된 원문 및 그에 상응되는 번역문을 기초로 한다. 단 정보를 각 야담집의 대표 텍스트로 삼되, 정보화를 위해 선정한 저본 및 이본의 텍스트와 이미지 또한 함께 열람할 수 있도록 시스템을 마련할 필요가 있다. 디지털 환경에서 구축되는 이점을 십분 활용하여 도서의 형태로 완성된 『정보 한국 야담전집』의 한계를 극복하기 위함이다. 요컨대 책에서는 교감 내용을 일정한 기준에 따라 선별적으로 수록할 수밖에 없었지만, 데이터 아카이브에서는 사용자의 요구에 따라 정보 텍스트와 각 이본 간의 전면적이고 직관적인 대조를 가능하도록 하는 것이다. 또 ‘야담집’의 하위항목으로 정보과 함께 주요 이본을 아울러 배치하는 것은 개별 야담 작품의 이본별 소재 여부, 나아가 다른 야담집에서의 출현 여부와 전

10) 이에 대해서는 류인태, 앞의 글, IV장을 참조할 것.

11) 참고로 현재 『정보 한국 야담전집』의 보유편 발간을 준비 중이며, 추후 보유편에 수록되는 자료 또한 데이터 아카이브의 대상이 될 것이다.

승관계를 간명하게 나타내기 위해 유리한 구조이다. 특히 이것은 책을 통해서 드러내지 못한 정보로, 데이터 아카이브에서 정밀하게 구축해야 할 데이터이다. 이 밖의 형식요소에 대한 구체적 내용은 다음 장에서 상술하기로 한다.

이제 내용의 차원에서 문맥에서 추출 가능한 데이터의 층위와 성격에 대해 살펴보자. 우선 야담에 담긴 내용을 대상으로 데이터 편찬을 시도할 때 가장 중요하게 고려해야 할 사항은 해당 데이터를 통해 서사문학으로서 야담의 특질과 그것의 재현 요소가 잘 드러나야 한다는 점이다. 소위 서사의 3요소라 일컬어지는 인물, 사건, 배경을 정밀하게 표지하면 되는 것 아닌가 싶지만, 그에 관한 내용은 좀처럼 간단치가 않다. 허구적이고 추상적인 요소가 야담에 다분한 까닭이다. 이를테면 인물의 경우, 실존인물뿐만 아니라 실존여부를 알 수 없는 인물, 또는 명백한 가공의 인물, 나아가 귀신, 동물, 이물 등 인간이 아니지만 등장인물로서 분명한 지분을 갖는 대상들까지 다양한 층위가 공존한다. 배경 또한 시간과 공간으로 간단히 구분하면 그만인 것처럼 보이지만, 이야기이기 때문에 면밀하게 따져보아야 할 지점들이 있다.

가령 시간적 배경의 경우 정확한 시점이 표기되기도 하나 ‘어릴 적’, ‘과거 공부를 할 때’, ‘어느 해 한겨울’, ‘한밤중’ 등과 같이 연월일시의 정규적 표현과는 다소 다른 층위의 시간 표현이 자주 나타난다. 공간적 배경도 마찬가지다. 위도와 경도 정보를 토대로 지도상에 점·선·면을 맵핑할 수 있는 구체적 장소가 있는 반면, ‘집’, ‘절’, ‘산속’, ‘해변’ 등과 같이 위치를 특정할 수 없거나 ‘명부(冥府)’, ‘용궁’처럼 비현실적 공간 또한 서사가 전개되는 공간으로서 빈번하게 제시된다. 이와 유사하게 사건도 실제 일어났던 역사적 사건이 작품 속에서 전개되는 경우와 등장인물의 행위로 인해 벌어지는 허구로서의 사건이 있으며, 이 두 가지는 서로 성격이 다른 것으로 구분해 접근할 필요가 있다.

이와 같은 요소들을 개개의 단위로 바라볼 경우 각각의 정보를 유의미한

것으로 수용하기 어려울 수도 있다. 그러나 20종 야담집에 수록된 수많은 이야기로부터 그와 같은 정보를 대량의 데이터로 체계화해 살펴볼 때 과연 어떤 지식을 발견할 수 있는가에 관해 누구도 쉽사리 장담할 수 없다. 시좌에 따라 이러한 정보의 조각들은 하나하나 중요한 의미 자질로 지목되고 활용될 잠재력을 갖고 있는 것이다. 정규적 맥락에서 수용되는 정보뿐만 아니라 그 외에 의미적 단위로 취급이 가능한 정보에 이르기까지 다양한 층위의 데이터를 얼마나 체계적으로 그리고 섬세하게 표지하는가에 따라 데이터 아카이브의 품질과 활용도가 결정되는바, 그러한 정보로서의 여러 가능성을 최대한 아우를 수 있게끔 데이터 모델을 정교하게 설계하는 것이 무척 중요하다. 여타 실기류(實記類) 자료를 대상으로 한 데이터 모델에서도 ‘인물’, ‘사건’, ‘시간’, ‘공간’ 등의 요소가 중요한 정보 항목으로 다루어져 왔으나, 그와 같은 기존 형식을 그대로 적용할 수 없는 이유가 바로 여기에 있다.

야담 속에 들어있는 정보와 지식의 단서는 이에 그치지 않는다. 물건이나 복식, 음식이나 의약품, 또 제도나 의식, 특정 책이나 시문·서화와 같은 작품들, 나아가 언급되고 있는 전고(典故) 등 데이터화 할 수 있는 수많은 정보가 산재해 있다. 서사 가운데 얼마나 중요한 역할과 기능을 하는 요소인가에 관한 문학적 관점의 평가나 중요도와 별개로, 어떠한 시각으로 접근하느냐에 따라 데이터의 가치와 활용에 관한 맥락은 가변적이다. 문학 연구에서 큰 가치는 없지만, 역사 연구에서 활용도가 높은 데이터가 있을 수 있고, 그 반대도 가능하다. 야담을 위한 데이터 모델은 예의 서사문학적 내용 요소뿐만 아니라 야담에 산재한 다양한 지식·정보의 질료들을 총체적으로 포집해 드러낼 수 있어야 하는 것이다.

한편 다른 문헌자료를 다룰 때와 달리 야담이기에 특별히 고려해야 할 사항이 있다. 바로 기이성 내지 비현실성을 띠고 있는 개체에 대한 처리이다. 주지하다시피 야담에는 이인(異人), 이계(異界), 이물(異物)을 비롯해 현실의 논리와 질서를 벗어나 있는 대상들이 다수 등장한다. 이러한 요소들은 야담의 서사세계를 흥미롭게 구성하는 중요한 질료라는 점에서 반드시 표

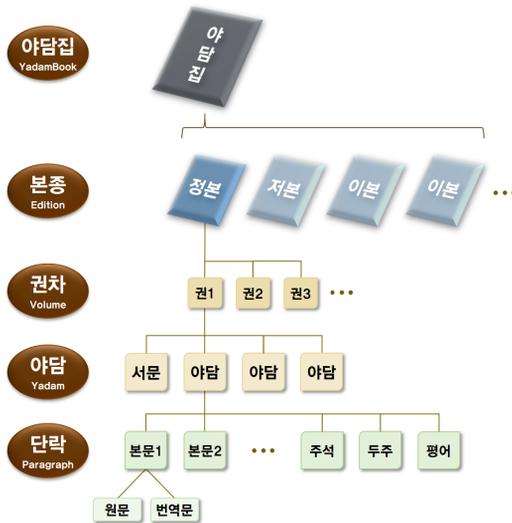
지되어야 하는 것들이다. 단 이에 대해서는 별도의 범주를 설정하기보다 기존 정보 항목을 최대한 활용하는 방식으로 접근하는 것이 최선이라 판단된다.

이상의 내용을 바탕으로 야담 데이터 모델을 설계하는 몇 가지 원칙을 정리해 제시한다. 첫째, 야담 데이터 모델은 각 야담집의 정보를 대표 텍스트로 하되 주요 이본의 원전 이미지와 텍스트를 함께 제공한다. 이를 통해 디지털 환경에서 1차 자료의 방대한 정보에 접근할 수 있는 데이터베이스를 마련한다. 둘째, 야담의 서사문학적 특성에 기초한 내용적 자질과 다양한 지식·정보의 단서들을 섬세하게 드러낼 수 있는 체계여야 한다. 문자열 기반의 데이터 처리를 뛰어넘어 지식 기반의 의미적 네트워크로서 야담 세계를 총체적으로 구현하기 위함이다. 셋째, 객관성을 확보할 수 있는 데이터를 표지하는 데 중점을 둔다. 이를테면 개별 작품의 주제와 같이 독자의 주관적 판단에 의존할 수밖에 없는 내용은 데이터 편찬 작업을 혼란스럽게 할 뿐 아니라 결과물의 정교함에 오히려 해가 될 수 있다. 데이터 처리 프로세스를 섬세하게 기획하되 그 결과물로서 데이터 아카이브를 활용해 ‘야담’에 관한 해석을 어떻게 전개할 것인가에 관한 문제는 사용자의 몫이다.

3. 야담 데이터 모델의 구상

앞 장에서 수행한 예비적 고찰을 바탕으로, 야담 자료를 대상으로 한 시맨틱 데이터 아카이브 구현의 기초 단계로서, 개념적 데이터 모델링 과정에서 의 문제의식을 아래에 정리하였다.

1) 야담 자료의 체계와 구성 및 형식을 중심으로



〈그림1〉 야담 텍스트의 체계도

위의 〈그림1〉은 야담 데이터를 디지털 환경에서 섬세하게 재현하기 위해 야담을 중심으로 구성된 다양한 층위의 텍스트를 체계적으로 나타낸 것이다. 여러 층위의 텍스트 가운데 최상위 범주는 ‘야담집’으로, 『정보 한국 야담전집』에 수록된 20종의 조선후기 야담집이 여기에 해당한다. 참고할만한 지점이 있으니, 여기서 야담집이란 추상적 개념으로서 정의된 것이다. 실물로서의 책 곧 정보, 저본, 이본 등은 야담집을 구성하는 하위항으로 규정할 수 있으며, 그것들은 ‘본종(Edition)’이라는 범주에 속한다. 이하로 ‘권차(Volume)’ 구분을 두고, 각 권 아래에 개별 작품으로서 ‘야담’이 배속된다. 물론 이러한 구조는 정보뿐만 아니라 여타 이본들에도 동일하게 적용된다. 야담은 다시 내용의 층위에 따라 본문, 주석, 두주, 평어 등으로 성격을 구분한 뒤 각각에 원문 및 번역문 텍스트를 품는다. 본문의 경우 작품의 편목을 고려해 의미단락에 따라 임의적으로 분절하기로 한다. 가독성은 물론 원문과

번역문을 쉽게 대응시켜 열람할 수 있다는 이점을 확보하기 위함이다. 또한 평어의 경우 야담집의 성격을 타진하는 주요한 자질이기도 한바, 따로 구분 지어 줌으로써 연구적 활용도를 기할 수 있다.

텍스트 차원에서 적시할 필요가 있는 정보는 속성(attribute)을 설계해 그에 대한 값을 제공하는 방식으로 나타낸다. 위의 체계도에서 드러나는 다양한 텍스트 층위 가운데 ‘야담집’, ‘본종’, ‘야담’ 세 정보 범주는 상대적으로 내포한 정보가 많기 때문에, 그것을 다루기 위한 속성 설계가 중요하다. 아래에 나열한 표는 각기 ‘야담집’, ‘본종’, ‘야담’에 해당하는 정보를 정리하기 위한 속성을 디자인한 것으로서 클래스별 속성 항목 예시라 할 수 있다.

[표 1] ‘야담집(YadamBook)’ 범주의 속성 항목 예시

식별자 (id)	YD014
대표명 (title)	청구야담
한자명 (titleC)	靑邱野談
영문명 (titleE)	Cheong-gu-yadam
편찬자 (writer)	미상
제작시기 (time)	19세기 중엽
서지구성 (volume)	10권 10책
수록회수 (content)	290화
저본 (mainEd)	버클리대본
이본 (othEd)	교토대본, 동양문고본, 국립중앙도서관본, 성균관대본, 가람문고본, 동경대본, 청구야설본

[표 2] ‘본종(Edition)’ 범주의 속성 항목 예시

식별자 (id)	YD014_ed04
범주 (catagory)	이본
대표명 (name)	국립중앙도서관본
원문표제 (titleC)	靑邱野談
서지구성 (volume)	6권 6책

크기 (size)	31.5 × 20.2 cm
수록화수 (content)	181화
소재 (location)	한국 국립중앙도서관
청구기호 (classNum)	한古朝48-190

[표 3] '야담(Yadam)' 범주의 속성 항목 예시

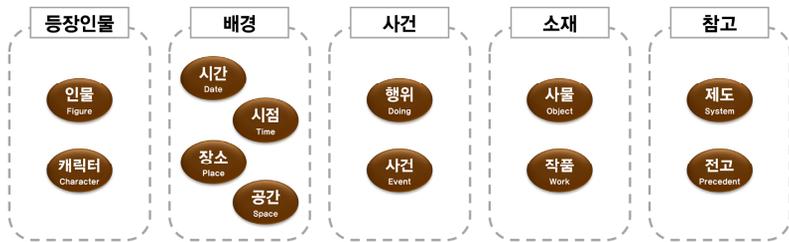
식별자 (id)	CG1007
제목 (title)	유 동지가 바다를 표류했다가 단구를 발견함
원문제목 (titleC)	識丹邱劉郎漂海
영문제목 (titleE)*	Mr.Yu found a Dan-gu while drifting in the sea
권차 (volumeOrd)	10권
순서 (contentOrd)	7화
주요인물 (figure)	유 동지, 단구노옹
공간배경 (place)	강원도 고성, 바다(동해), 단구
시간배경 (time)	
관련사건 (event)	
주요화소 (motif)	표류, 이계

관련해서 특기해둘 사항이 몇 가지 있다. 첫 번째, '야담집' 개체의 속성 값은 정보를 기준으로 작성하며, 이에 따라 '본종' 개체의 속성값은 저본과 이본만을 대상으로 한다. 두 번째, '야담' 개체의 속성 가운데 주요화소 항목은 내용적 요소에 속하지만 텍스트에 직접 마크업(Markup)하는 것이 용이하지 않은바, 개별 작품의 정보를 담는 단일 속성으로서 처리한다. 이에 대해서는 기존 야담 연구의 성과를 적극 수용하되 야담 자료 전반을 대상으로 다양한 모티프를 추출하여 풀(pool)을 갖추어줄 필요가 있다. 예컨대 치부, 결연, 보은, 복수, 급제, 풍수, 예지, 도적, 귀신, 이인, 이계, 송사, 표류, 횡재 등의 풀 가운데 해당 속성 값으로 적용할 수 있는 화소를 모두 선택해서 입력하는 방식이다.¹²⁾ 해당 데이터는 추후 데이터베이스 고도화 작업에서 화

12) 복수선택이 가능하다는 것으로, 권기성·김동건이 앞의 글에서 '가변적 분류체계'

소 클래스의 개체를 독립적으로 구축하기 위한 기초 자료로 활용한다. 세 번째는 ‘야담’의 속성 항목 가운데 영문제목을 둔 데 대한 부언이다. 한국 야담 데이터 아카이브는 일차적으로 한문 원문과 한글 번역문으로 구축하 되, 장기적으로는 영문 번역문을 추가하여 영어권 연구자나 대중이 용이하게 접근할 수 있는 웹 서비스를 지향한다. 영문제목은 이 단계를 염두에 둔 것으로 추후 활성화되어야 할 항목이며, 다음 절에서 살펴볼 속성들에 영문 명 항목이 있는 것 또한 같은 차원에서 기재되어 있는 것임을 밝혀둔다.

2) 문맥을 통해 확인되는 다채로운 정보를 중심으로



야담에 산재한 다양한 정보와 지식을 체계적 데이터로 편찬하기 위해 앞서 예비적 고찰을 진행하였으며, 그로부터 위의 다섯 가지 대분류 하에 12 개의 클래스를 디자인하고, 해당 클래스에 배속된 개체들의 구체적인 정보를 표지할 수 있도록 적합한 속성을 설계하였다. 클래스를 디자인한 근거를 차례대로 서술하고, 정의한 속성의 내용을 테이블 형식으로 정리해 제시하는 가운데 필요한 부분에는 추가 설명을 부기하였다.

(1) 인물·캐릭터

서사문학에서 등장인물을 분류하는 기준은 몇 가지가 있다. 주동·반동,

중심-주변 등을 그 예로 떠올릴 수 있을 것이다. 그러나 이것은 서사에서 차지하는 역할이나 비중, 중요도 등에 의거한 것으로 경우에 따라 주관적 판단이 개입할 소지가 있으며, 작품을 해석하는 시각에 따라 변동될 수 있는 가능성도 갖고 있다. 또 이 기준 자체가 야담에 등장하는 다종다양한 인물의 층위를 긴요하게 나누는 자질이 되지 못하거니와, 이를 데이터로 편찬함에 있어 그다지 이점이 있다고 판단되지도 않는다.

야담에 등장하는 인물은 대단히 복잡한 존재양상을 갖고 있다. 실존인물 뿐만 아니라 가공인물도 허다하고, 이름이 명확히 제시되는 경우도 있고 그렇지 않은 경우도 있으며, 때로는 특정 인물이 산 사람으로 나오다가 사망 사건이나 서사 전개 과정의 특수성으로 인해 귀신으로 등장하는 경우도 있다. 뿐만 아니라 보통 사람과는 다른 기이한 능력을 가진 인물도 있고, 겉으로는 인간의 모습을 하고 있지만 실제 사람과는 질적으로 다른 존재들도 있으며, 때로는 인간이 아닌 동물이나 괴물이 일정한 서사적 비중을 갖고 등장하는 경우도 있다. 이와 같이 다종다양한 인물군상을 효과적으로 정리할 수 있는 모델을 고민해야 하는데, 이를 위해 고려해 볼 수 있는 방안이 ‘인물’과 ‘캐릭터’를 나누어 표지하는 것이다.

‘인물’은 독립된 인격체로서 존재하는 여러 대상을 집적하는 클래스이다. 이를 판정하는 기준은 해당 인물을 특정하는 호칭이 있는가의 여부이다. 성명자호를 비롯해 직칭이나 별칭 등 여러 형태가 있을 수 있지만, 그것이 해당 인물군을 범칭하는 것이 아니라 작품 속에서 한 인물을 한정해서 칭하고 그 대상이 서사 속에서 유의미한 비중을 갖고 있다면 인물로서 포집할 수 있는 것이다. 반면 ‘캐릭터’는 특정 인물이 아니라 어떠한 인물군이나 계층 또는 인물이 아닌 대상을 표지하기 위한 클래스이다. 인물 클래스의 대표명은 해당 인물에 대한 이름이 되지만, 캐릭터의 경우 작품에서 이름이 호명되지 않는바 이름을 대표명 속성의 값으로 상정하기는 어렵다. 예컨대 노비, 유모, 기녀, 상인, 무변, 술사, 원귀, 여귀, 신선, 호랑이 등이 캐릭터의 각 개체가 대표명이 되는 것이다. 이를 통해 서사 속에서 비중이 적고 특정정도

없지만 야담 속에 흩어져 있는 다양한 인물군상을 데이터로 수습할 수 있다.

‘인물’과 ‘캐릭터’는 서로 설정한 목적이 다른바 간혹 한 대상에게 중복으로 마크업이 이루어질 수도 있다. 마크업의 중복은 가급적 피하는 것이 좋겠으나, 이 경우는 야담 속 등장인물을 한층 입체적인 데이터로 재현하기 위한 방편으로서 중복을 허용하는 것이 합리적이라는 판단이다.

[표 4] ‘인물(Figure)’ 범주의 속성 항목 예시

식별자 (id)	FG0000	영문명 (nameE)	
대표명 (name)		한자명 (nameC)	
자 (nick1)		자 ^한 재 (nick1C)	
호 (nick2)		호 ^한 재 (nick2C)	
시호·묘호 (nick3)		시호·묘호 ^한 재 (nick3C)	
본관 (clan)		본관 ^한 재 (clan)	
생년 (birth)		몰년 (death)	
급제년도 (pass)		주요관력 (post)	
주요활동 (career)		주요저작 (work)	
성별 (sex)	남 / 여	실존성 (facticity)	+ / -
국적 (nation)	한국 / 중국 / 일본 / 기타	존재성 (existence)	현실 / 비현실
신분 (position)	왕실 / 문반 / 무반 / 중인 / 양인 / 천인 / 궁인 / 승려		

[표 5] ‘캐릭터(Character)’ 범주의 속성 항목 예시

범주 (catagory)	사람 (human)	귀신 (ghost)	별인 (ano_human)	동물 (animal)
식별자 (id)	CH0000	영문명 (nameE)		
대표명 (name)		한자명 (nameC)		
이칭 (oth_name)		실존성 (facticity)	+ / -	
성별 (sex)		존재성 (existence)	현실 / 비현실	

위는 ‘인물’ 및 ‘캐릭터’ 클래스의 속성을 쉽게 확인할 수 있도록 테이블 형식으로 정리한 것이다. 속성값을 많이 채울수록 데이터의 정교함이나 그

풍부함의 정도가 더욱 높아질 것이나, 모든 속성 항목에 대응하는 값을 찾기란 쉽지 않을 것이다. 속성 가운데 특기할만한 요소가 있으니, ‘실존성(facticity)’과 ‘존재성(existence)’이다.

실존성은 ‘인물’의 경우 실존인물로 확인되는지의 여부를 표기하는 것이요, ‘캐릭터’의 경우 현실세계에 실존하는 존재인지를 표기하는 것이다. 존재성은 존재의 성격이 현실적인지 비현실적인지를 기입하는 것이다. 둘은 언뜻 비슷해 보여 실존성이 있는 개체라면 현실적 존재로, 혹은 반대로 으레 짝이 이어질 듯하지만, 야담에 등장하는 인물과 캐릭터는 두 속성이 교차하는 대상도 더러 있다. 일례로 ‘김치(金織)’는 실존인물이 틀림없지만 야담 속에서 그려지는 모습은 대단히 비현실적이다. 실존성과 존재성 두 속성은 양자의 조합을 통해 야담 속 인물 또는 캐릭터의 다면적인 특성을 효율적으로 나타내기 위한 정보로 이해할 수 있다.

(2) 시간·시점 / 장소·공간

서사요소 가운데 배경 정보를 드러내기 위한 클래스다. ‘시간’ 클래스는 기본적으로 연도 혹은 날짜 정보를 표지하기 위한 것이다. ‘시점’이라는 클래스를 독립적으로 디자인한 것은 야담에서 시간적 배경을 제시하는 방식이 절대적 시계열로 보여주지만 하는 것이 아니기 때문이다. 앞서 2장에서 언급한대로 ‘어렸을 적’이나 ‘과거공부를 하고 있을 때’ 또는 ‘한겨울’이나 ‘한밤중’처럼 개인적 차원 또는 추상적 맥락에서 언급된 예를 자주 발견할 수 있다. 이 또한 서사물에서 시간적 배경을 제시하는 특유의 방식이라 할 것이며, 이로부터 해석의 실마리를 찾기도 하는바 간과할 수 없는 정보에 해당한다. 이에 대한 해법으로 ‘주기(cycle)’ 개념을 범주화하여 속성에 도입하는 방식을 고려하였다. 예컨대 속성값으로서 ‘생애주기(life)’는 유년기, 소년기, 청년기, 장년기, 노년기 등 생애 기준의 시간적 표현에 대응한 값으로, ‘연주기(annual)’는 봄, 여름, 가을, 겨울과 같은 사계절 또는 24절기의 표현에 대응한 값으로, ‘일주기(diurnal)’는 새벽, 아침, 점심, 저녁, 밤 등 하루 기

준의 시간적 표현에 대응한 값으로, ‘사회주기(social)’는 미급제시, 임관시, 유배시, 은퇴시 등 개인의 사회 활동 주기에 대응한 값으로 삼을 수 있다. 이와 같은 방식의 접근은 연, 월, 일과 같은 정규적 시간 개념과는 다른 형식으로 야담에 기록된 다채로운 시간적 관념까지도 데이터화하고자 하는 의도로부터 기인하는 것이다. 아래는 ‘시간’과 ‘시점’ 클래스의 속성을 쉽게 확인할 수 있도록 테이블 형식으로 정리한 것이다.

[표 6] ‘시간(Date)’ 범주의 속성 항목 예시

연 (year)		[양력연 (year_s)]	
월 (month)		[양력월 (month_s)]	
일 (oth_name)		[양력일 (day_s)]	
절기 (division)		재위연차 (era)	

[표 7] ‘시점(Time)’ 범주의 속성 항목 예시

범주 (catagory)	생애주기 (life)	연주기 (annual)	일주기 (diurnal)	사회주기 (social)
식별자 (id)	TM0000	영문명 (nameE)		
대표명 (name)		한자명 (nameC)		

공간적 배경의 경우 크게 두 개념으로 층위를 나누어 볼 수 있다. 지도상에 해당 지점을 표시할 수 있는 ‘장소(place)’로서의 개념이 있을 것이고, 구체적인 위치가 특정되지 않거나 애초 그것이 불가능한 상상의 세계를 포괄하는 ‘공간(space)’의 개념을 상정할 수 있을 것이다. 둘은 존재 양태와 속성이 서로 다르며, 데이터를 정리하는 목적이나 맥락도 상이하다. 예컨대 ‘장소’ 데이터의 경우 전자지도 형식으로 데이터를 시각적으로 출력해서 나타내는 방안에 관한 고민을 주로 하게 되겠지만, ‘공간’ 클래스로 포집되는 데이터는 야담의 서사공간으로서 문학적 해석의 단서로 활용할 방안에 집중할 가능성이 크다. 따라서 이 둘을 별개의 범주로 나누고, 아래와 같이 각각의 속성을 설계하였다.

[표 8] '장소(Place)' 범주의 속성 항목 예시

범주 (category)	행정단위 (district)	도로 (road)	자연 (nature)	건물 (building)
하위유형 (subtype)	지방도 대도호부 도호부 목군현	육로 수로 해로 고개 거리	산 하천 호수 평야 바다 섬	궁성문보 가옥 누정 관청 학교 역참 사찰
식별자 (id)	PL0000		영문명 (nameE)	
대표명 (name)			한자명 (nameC)	
이칭 (oth_name)			형태 (form)	점 / 선 / 면
경도 (long)			위도 (lat)	

장소 클래스의 개체는 전술한대로 위경도 좌표 정보를 기반으로 지도상에 맵핑할 수 있는 정보를 대상으로 하되, 작품 내에서 분명하게 특정되었으나 현재 그 위치를 확인할 수 없는 개체는 위경도 정보를 비워두는 방식으로 정리한다. 대표명은 현재 널리 사용되는 명칭을 쓰고, 야담에 서술된 표기나 조선시대에 사용된 명칭 및 별칭 등은 이칭 속성으로 입력한다.

[표 9] '공간(Space)' 범주의 속성 항목 예시

범주 (category)	일상 (living)	공공 (public)	특수 (special)	자연 (nature)	별계 (oth_world)
식별자 (id)	SP0000		영문명 (nameE)		
대표명 (name)			한자명 (nameC)		
이칭 (oth_name)			존재성 (existence)	현실 / 비현실	
* 대표명 예시	집방 마루 대문 행랑 사당 마당 논밭 등	우물 시장 주막 기루 약포 성문 절 암자 사당 마을 고을 관아 역 감옥 등	산소 명당 묘혈 비경 복지 전장 미로 등	산숲 들샘 시내 강 바다 섬 동굴 등	천계 명부 용궁 백옥경 엄부주 단구 백화국 등

공간 클래스의 개체는 위치를 특정할 수 없으나 야담 내에서 서사가 펼쳐지는 공간으로서 제시된 곳을 포집한다. 크게 일상 공간, 공공 공간, 특수

공간, 자연 공간, 별계 공간 등으로 범주화하여 정리할 것이며, 각각의 성격은 위에 제시한 대표명 예시를 참조하도록 한다. 공간에는 위치정보 대신 ‘존재성(existence)’이라는 속성 항목을 설정하였다. 상상의 공간을 비롯해 현실 공간으로 서술되어 있다 하더라도 허구성이 농후한 공간 등은 별도로 드러낼 수 있어야 하기 때문이다.

(3) 행위·사건

서사 요소 가운데 사건 정보를 데이터화하기 위한 클래스이다. 앞서 간단히 지적한 바와 같이 야담에서 사건은 크게 두 층위에서 볼 필요가 있다. 먼저 작품 속 서사 내에서 인물들에 의해 일어나는 사건이 있을 테고, 따로 역사적 사건으로서 야담의 서사와 접촉하게 되는 것이 있을 것이다. 야담이 서사물인 이상 전자가 다채롭게 있음은 말할 것이 없지만, 후자의 정보 또한 적지 않게 산포되어 있는 것 또한 사실이다. 둘은 그 성격이 상이하기에 각각 독립된 클래스로 설정하였다.

[표 10] ‘행위(Doing)’ 범주의 속성 항목 예시

범주 (catagory)	사고 (accident)	학업 (study)	생업 (job)	이동 (move)	술수 (divine)	개인 (personal)
식별자 (id)	DO0000		영문명 (nameE)			
대표명 (name)			한자명 (nameC)			
* 대표명 예시	질병 상해 사망 살해 호환 포함 등	독서 입학 응시 급제 낙방 폐업 사가독서 등	경작 수렵 채취 어로 제조 매문 품팔이 상거래 등	방문 유람 상경 부임 사행 항해 표류 피난 이상이동 등	예언 예지 천문 풍수 점복 택일 관상 등	사랑 이별 기호 포폄 서신 회상 보은 복수 음주 등

‘행위’ 클래스는 야담 속 인물이 행하거나 당하는 일을 표지하기 위한 것이다. 이 클래스는 개체의 양이 중요한 것이 아니라 야담 텍스트에서 드러나는 다양한 행위를 포착하고 체계적으로 정리하는 것이 핵심이다. 그렇기

때문에 야담에서 드러나는 인물들의 다양한 행위의 층위와 양태를 고려해 위와 같이 ‘범주’를 속성화하고 그에 대한 구체적인 값으로 여섯 가지 요소를 반영하였다. 그 외의 뚜렷한 속성은 없지만, ‘인물’을 비롯한 여타 클래스 개체들과의 ‘관계 맺음’ 속에서 유의미한 지점이 드러날 것이다. 첨부한 대표명 예시는 그 실재를 가늠해볼 수 있는 단서이다. 그 가운데 이채로운 정보가 하나 있다. ‘이상이동’이라는 항목이 그것인데, 이는 이계로의 이동이나 신술(神術)에 의한 순간이동 등 비현실적인 이동을 표지하기 위한 것이다.

[표 11] ‘사건(Event)’ 범주의 속성 항목 예시

범주 (catagory)	정치 (politics)		재난 (disaster)
하위유형 (subtype)	즉위 폐위 정쟁 반정 역모 민란		전쟁 자연재해 역병
식별자 (id)	EV0000	영문명 (nameE)	
대표명 (name)		한자명 (nameC)	
이칭 (oth_name)		발생시기 (time)	
관련인물 (person)		발생장소 (place)	
내용 (summary)			

‘사건’ 클래스는 실제 현실에서 벌어진 사건들을 표지하기 위한 것이다. 정치적 사건과 재난 사건 두 가지로 범주화하고 아래에 구체적인 유형을 배치해 둔 것은 해당 사건의 성격을 간명하게 나타내기 위함이다. 다른 클래스와 달리 ‘사건’에는 사건의 경과를 약술할 수 있도록 ‘내용(summary)’이라는 속성을 설정하였다. 단편적인 정보항만으로 사건에 관한 정보를 충실하게 전달하기에는 한계가 있기 때문인데, 이는 추후 기획에 따라 외부의 사전류 데이터베이스와 연동시키는 방식으로 대체할 수도 있다.

(4) 사물·작품

‘사물’과 ‘작품’은 소재의 차원에서 야담에 산재해 있는 다대한 정보 조각

들을 모으기 위해 설정한 클래스이다. 이들의 실체는 아래 테이블 형식으로 정리한 속성 항목을 살펴보는 것이 빠르다.

[표 12] '사물(Object)' 범주의 속성 항목 예시

범주 (catagory)	복식 (fashion)	식품 (food)	의약품 (medicine)	도구 (disaster)	탈것 (vehicle)	재화 (valuable)	기타 물건 (item)
식별자 (id)	OB0000		영문명 (nameE)				
대표명 (name)			한자명 (nameC)				
이칭 (oth_name)			존재성 (existence)		현실 / 비현실		

'사물'에 대해서는 위와 같이 7개의 범주를 설정하였다. 몇 가지에 대해 부연하자면, 복식은 의복을 비롯하여 장신구나 신 등 착용품 전반을 포괄하며, 도구는 가구나 집기, 문구, 농기구 등 일상에서 사용하는 각종 용품들을 아우른다. 탈것은 가마, 수레, 말 등이 개체가 될 것이며, 재화는 돈이나 값어치 있는 보물 등을 구분하기 위한 범주이다. 사물의 속성 가운데 존재성이 들어가 있는 것은 비현실적인 소재를 특기하기 위함으로, 이를 통해 상상과 욕망이 만들어낸 흥미로운 물건들을 드러낼 수 있다.

[표 13] '작품(Work)' 범주의 속성 항목 예시

범주 (catagory)	문헌 (book)	시문 (literature)	실용문 (practical)	서화 (painting)
하위유형 (subtype)	경전 철학서 역사서 필기 개인문집 등	시 산문 소설 희곡 등	상소문 계문 공문 비문 서간 등	글씨 그림 필첩 화첩 등
식별자 (id)	WK0000		영문명 (nameE)	
대표명 (name)			한자명 (nameC)	
이칭 (oth_name)			작자 (writer)	
언어 (language)	한문 / 한글		제작시기 (time)	
포함여부 (contain)	+ / -		국적 (nation)	
			한국/중국/일본/기타	

‘작품’ 클래스는 야담에서 발견되는 서적류나 문예작품 등을 데이터화하고자 설정한 것으로, 크게 문헌(책), 시문(개별 문학작품), 실용문, 서화 등으로 범주화된다. 이들은 야담 내에서 간단히 언급되는 경우가 일반적이지만, 간혹 해당 작품이 야담 내에 실려 있는 경우도 있고 그 중에는 야담 자료에서만 볼 수 있는 작품도 있다. 작품 원형의 야담 자료 수록 여부에 관한 정보는 매우 중요하게 취급될 수 있는 데이터에 해당하며, 속성 가운데 ‘포함여부(contain)’ 항목을 둔 것은 이를 표지하기 위함이다.

(5) 제도·전고

한편 제도와 전고 또한 야담에서 추출할 수 있는 흥미로운 데이터다. ‘제도’ 클래스에 포집되는 데이터의 경우 서사문학으로서 접근할 때는 그다지 긴요한 정보가 아니지만, 야담을 통해 당대 사회의 제도나 의식(儀式), 관습, 풍속 등을 재구하고자 한다면 대단히 유용한 단서가 될 수 있다. 특히 야담에서 기술되는 방식이나 장면, 어떤 용어로 그 개념을 표현하고 있는지 등은 섬세한 추론의 재료가 된다. 제도와 관련된 데이터는 기구, 관직, 시험, 사법, 행정, 의식, 풍속 등 일곱 가지로 범주화하여 수집하기로 한다.

[표 14] ‘제도(System)’ 범주의 속성 항목 예시

범주 (category)	기구 (government)	관직 (post)	시험 (exam)	사법 (justice)	행정 (admin)	의식 (ceremony)	풍속 (custom)
하위유형 (subtype)	중앙관서 지방관아 향촌기구	당상관 당하관 외관직 파견직 명예직	소과 대과 잡과 성균관시	옥사 송사 처벌 사면	인사 정책 명령 구휼	국가의례 사회의례 관혼상제	관습 놀이
식별자 (id)	SY0000		영문명 (nameE)				
대표명 (name)			한자명 (nameC)				
이칭 (oth_name)							

* 대표명 예시	조정 홍문관 내명부 도호부 유향소 등	영의정 이조정랑 경기도관찰사 암행어사 참봉 좌수 등	진사시 생원시 식년시 알성시 춘당대시 승제 국제 등	친국 심문 판결 수감 유배 해배 산송 신원 등	임명 좌천 대동법 금주령 윤음 금벌 환국 등	즉위 국상 나례 계회 삼일유가 면신례 등	금줄 씨름 소싸움 등

전근대 문헌 자료에서 전고(典故)가 차지하는 위상에 대해서는 다시 말할 것이 없다. 야담에서 간취되는 전고는 중국에서 유래한 전통적인 고사뿐만 아니라 우리나라의 문학 전통에서 산생된 것, 또 서사문학에서 특유하게 형성된 것까지 실로 다채로운 지평을 갖고 있다. 뿐만 아니라 우리나라 속담도 적잖이 찾을 수 있는데, 이들을 지식의 질료로서 체계 있게 집적할 필요가 있다.

[표 15] '전고(Precedent)' 범주의 속성 항목 예시

범주 (catagory)	전고 (precedent)		속담 (proverb)
식별자 (id)	PR0000	영문명 (nameE)	
대표명 (name)		한자명 (nameC)	
이칭 (oth_name)		관련성어 (idiom)	
언어 (language)	한문 / 한글	관련인물 (figure)	
국적 (nation)	한국 / 중국	관련작품 (work)	
내용 (summary)			

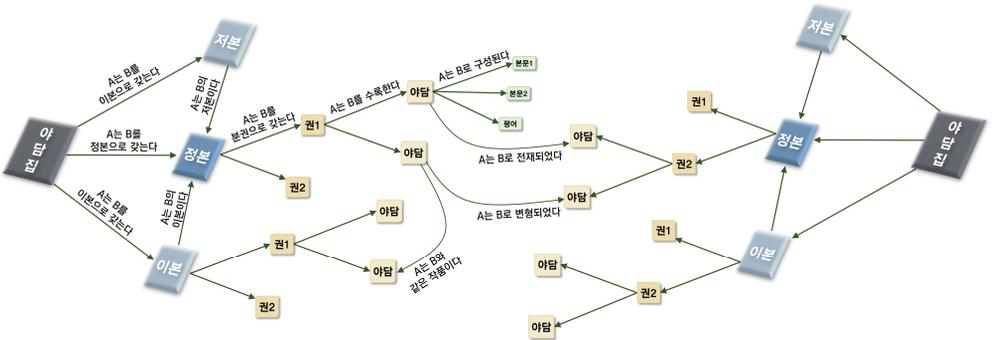
‘제도’ 및 ‘전고’의 클래스로 구성되는 데이터는 일종의 사전(glossary)으로서 기능할 수 있음에 주목할 필요가 있다. 이는 야담을 이해하는 것에서 나아가 야담으로부터 다른 고전문학 자료를 보다 풍부하게 읽을 수 있는 지식 체계를 수확하는 과정이라는 점에서 더욱 특기해 둘 만하다.

3) 개체 간 관계를 중심으로

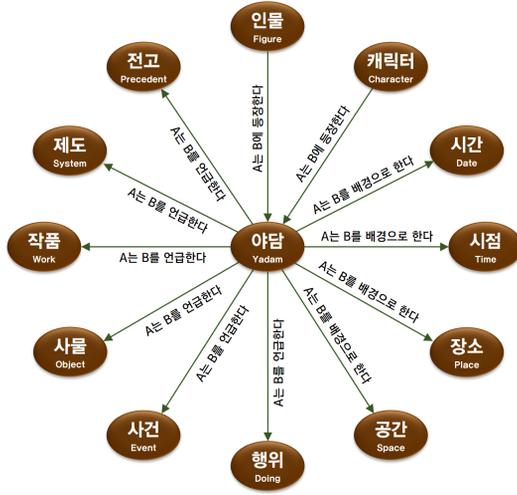
시맨틱 테크놀로지에 기반을 둔 데이터 아카이브의 특장은 개체 사이의

관계를 나타낼 수 있다는 것이다. 이를 통해 방대한 양의 데이터는 각자의 클래스 속에서만 존재하는 것이 아니라 다른 클래스의 유관 데이터들과 활발하게 접속하며 거대한 지식·정보의 군집체가 만들어진다. 본고에서 구상하는 보다 진보된 야담 데이터 모델이란 궁극적으로 이러한 데이터의 활발성을 구현하는 것이다. 이를 위해서는 개체 간의 관계에 대한 설계 구상이 있어야 한다.

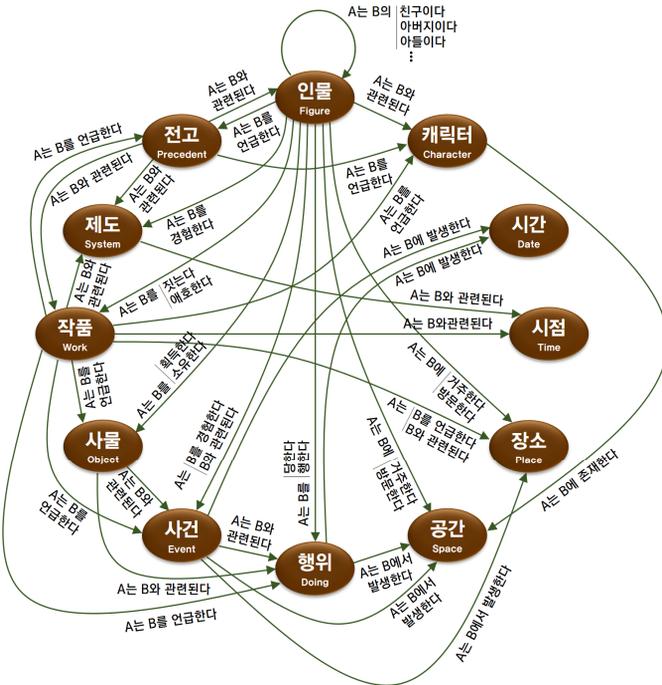
관계(relation) 데이터 설계의 골자를 직관적으로 이해할 수 있게끔 RDF와 같은 Triple(svo) 형식의 다이어그램을 그려 나타내보았다. 도식을 해석하는 법은 화살표의 시작에 있는 항목을 주어(Source)로, 가리키고 있는 항목을 목적어(Target)로 대입해 읽는 것이다. 단 논리 구조를 쉽게 전달하기 위해 구체적인 개체를 노드로 적용하지 않고 클래스명을 노드로 배치하였기에, 클래스에 배속되는 구체적인 개체가 실제 노드에 포지셔닝 된다는 것을 전제로 할 필요가 있다. 가독성을 고려해 세 층위로 나누어 제시한다.



〈그림2〉 야담 텍스트 체계에서 개체 간의 관계도



〈그림3〉 야담과 내용 요소 클래스의 개체 간 관계



〈그림4〉 내용 요소 클래스의 개체 간 관계

4. 성과와 의의

이번 논의는 『정보 한국 야담전집』을 자료적 자원으로 삼아 한국 야담 데이터 아카이브를 구현하기 위한 구체적인 데이터 모델을 구상해본 것이다. 기실 이 구상은 실제 결과물 구축이라는 목표를 고려할 때, 온톨로지를 정확하게 표현할 수 있는 RDF나 OWL과 같은 언어로 그 실재를 기술하는 방식이 타당하다. 그럼에도 불구하고 이 글은 서두에서 선언한 바와 같이 온톨로지 기술에 관한 서술 형식보다는 가급적 ‘야담’을 직접 다루는 연구자의 언어로 개념적 데이터 모델링 과정을 전달하고자 하였다. 이유는 다름 아니다. 조선후기 야담 자료를 고품질의 데이터 아카이브로 구축하기 위한 실질적이고 본격적인 논의를 유관 분야의 연구자들 사이에서 점화하고자 하는 것이 본 연구의 목적이기 때문이다. 여기서 소개한 데이터 모델은 하나의 프로토타입으로, 여러 고민과 시행착오를 거쳤음에도 여전히 보완해야 할 부분이 적지 않다. 이것을 보다 정교하게 만들어가는 작업은 여러 야담 연구자의 지식과 안목이 활발하게 교류되는 속에서 양분을 얻을 수 있을 것이다. 한국 야담 데이터 아카이브를 위한 완성된 온톨로지의 실체는 그와 같은 보완 과정을 거쳐 추후에 다시 발표할 것을 기약한다.

이와 별개로 이번 논의의 성과와 의의를 정리하지 않을 수 없다. 가장 큰 소득은 조선후기 야담 자료 전반을 대상으로 그에 산포된 다대한 정보와 지식을 데이터로 취집하여 체계적으로 편찬할 수 있는 구체적인 모델을 처음으로 확보했다는 것이다. 이는 야담을 지면에서 디지털 환경으로 옮기는 프로젝트의 실질적인 방법론을 제시한 것이자, 그 과정이 지향해야 할 방향을 선언한 것이기도 하다. 이와 함께 본고를 통해 소개하는 모델이 고전 ‘서사문학’ 자료를 겨냥한 것이라는 점 또한 강조될 자격이 있다. 이미 지적했다시피 고전 자료의 디지털화 작업이 곳곳에서 이루어지고 있지만 대부분의 경우 ‘사실정보’를 취합하기 위한 역사기록물류에 초점이 맞추어져 있으며, 문학 자료를 대상으로 하는 경우라 하더라도 그 문학적 성격을 충분히

드러낼 수 있는 접근과 설계는 활발하게 고민되지 못하고 있는 형편이다. 야담에 등장하는 인물을 ‘인물’과 ‘캐릭터’로 분리해 정리하고자 한 것, ‘시간’을 연월일과 같은 정규적 시간과 상대적 주기 개념으로 구분해서 접근한 것, 물리적 위치로서의 ‘장소’와 서사가 전개되는 ‘공간’을 나누어 이해한 것, 등장인물의 행위를 독립적인 클래스로 설정한 것 등의 시도는 이후 서사물 특유의 속성과 문법에 입각한 또 다른 데이터 모델을 고민하는 데 있어서 새로운 시각과 착상을 제공할 수 있을 것이다.

아울러 이번 논의는 디지털 인문학 연구의 본령을 새삼 확인케 한다는 점에서도 의의가 있다. 야담에 최적화된 데이터 모델을 고안하는 것은 야담의 자료적 특성과 내용, 의미에 대한 심도 있는 이해가 없이는 불가능에 가까우며, 그 이해의 깊이야말로 데이터 모델의 품질을 결정하는 중요한 인자라 할 것이다. 데이터의 관점에서 대상을 해체, 조립하며 분석하는 일련의 과정은 그 자체로 또 다른 차원에서 대상을 심층적으로 연구하는 방법론임을 보고를 통해 체감할 수 있길 기대한다.

■ 참고문헌

□ 자료

- 이우성·임형택 편역, 『이조한문단편집 상』, 일조각, 1973.
_____, 『이조한문단편집 중』, 일조각, 1978.
_____, 『이조한문단편집 하』, 일조각, 1978.
- 동국대학교 한국문학연구소 편, 『한국문헌실화전집』(전10권), 태학사, 1981.
- 박용식·소재영·大谷森繁 편, 『한국야담사화집성』(전5권), 국학자료원, 1995.
- 정명기, 『한국야담자료집성』(전24권), 계명문화사, 1995.
- 정환국 책임교열, 『정보 한국 야담전집』(전10권), 보고서, 2021.

□ 논저

- 김현·임영상·김바로, 『디지털 인문학 입문』, 한국외국어대학교 지식출판원, 2016.
- 권기성·김동건, 「야담집 색인 데이터베이스의 구축방안 모색: 『기문총화(記聞叢話)』를 중심으로」, 『고전과 해석』 제22집, 고전문학한문학회, 2017, 93~126쪽.(UCI: G704-SER000010566.2017.22..005)
- 류인태, 「데이터로 읽는 17세기 재지사족의 일상: 『지암일기(1692-1699)』 데이터베이스 편찬 연구」, 한국학중앙연구원 박사논문, 2019.
- 이강옥, 「야담 연구의 대중화 방안」, 『어문학』 제115집, 한국어문학회, 2012, 289~319쪽.(UCI: G704-000112.2012..115.015)
- 최지연·조성덕·최동빈, 「데이터 정제의 관점에서 본 한국문집총간 XML 문서 검토」, 『한문학논집』 60집, 근역한문학회, 2021, 157~194쪽.
- 이병찬·민경주, 「한국고전종합DB의 한국문집총간분석 시각화 방안 연구」, 『민족문화』 57호, 2021, 5~32쪽.(DOI: 10.15752/itkc.57.202103.5)

Abstract

Data of Yadam, Data from Yadam: Designing of Korea Yadam data model

Yang, Seung-Mok · Ryu, In-Tae

Korea Yadam Collection (officially certified version), which was just published, is the outcome of Research for an authentic version of 20 famous tales from the late Joseon Dynasty. As one of the follow-up tasks to this achievement, this research attempts to create a “Korea Yadam (野談) Data Archive” and investigates specific techniques for doing so. Unfortunately, no database implementation case for Yadam of the late Joseon dynasty has been found. In light of this, just converting the original text of the *Korea Yadam Collection* (officially certified version) to plain text data and distributing it on the web would be of considerable use. However, it is impossible to conclude that presenting string-based data is a sufficient means for properly revealing the colorful and vivid information included in Yadam at the time. It is vital to look at yadams as a record of clusters of complex facts in order to systematically organize, communicate, and share the content and form of sagas, varied information and knowledge that may be obtained through sagas, and their cultural connotations. It is required to develop a semantic database that can display relevant information in high resolution in a digital environment, rather to a string-based database that processes just simple text information such as original texts and translated texts.

This article provides the conceptual data modeling carried out in the process of thoroughly reviewing and analyzing the format and content of the

late Joseon Dynasty Yadam data archive as a search for the basic steps to establish such a semantic-based Yadam data archive. Rather than concentrating on the semantic technology that can be used to the fundamental stage of the creation of ‘Yadam in the Late Joseon Dynasty Ontology’ as an extension, we focused on organizing and sharing as much material as possible about the yadam considerations made during this process. Such an approach aims to start a serious discussion about sophisticated processing of Yadam data in a digital context. It also aims to widen the perspectives of diverse academics on the subject and raise awareness of connected issues. The objective of this paper is to focus on the humanities discussion of Yadam rather than the description of computer and online technologies often utilized in the data modeling process for humanities resources.

Keyword ● Yadam (野談), *Korea Yadam Collection (officially certified version)*, Humanities data modeling, Semantic database, Digital humanities

■ 논문투고일 : 2022. 3. 20 ■ 심사완료일 : 2022. 4. 16 ■ 게재확정일 : 2022. 4. 16