

Digital Transformation of Literary Collections in the Joseon Dynasty - Focused on Kang Hee-an's 'Yanghwasorok

조선시대 문집의 디지털 전환 -강희안의 「양화소록」을 중심으로-

이한나 (Hanna Lee), 명지대학교

본 연구는 조선시대 문집 텍스트의 디지털 전환을 통해, 문헌 안의 지식정보가 다른 문헌과 연결되어 탐색할 수 있도록 하는 데에 목적이 있다. 이를 위해 조선시대 문인 강희안(姜希顔, 1417~1465)의 문집 「양화소록(養花小錄)」을 대상으로 문헌을 기계가독적(machine readable) 형태인 XML 문서로 제작하고, 여기서 추출된 지식 정보를 시맨틱 데이터 베이스로 구축하여 대상 문헌과 유관 지식을 통합적인 자원으로 만들고자 한다.

「양화소록(養花小錄)」은 동생 강희맹(姜希孟, 1424~1483)이 편찬한 『진산세고(晉山世稿)』 권 4에 수록된 것이다. 진산세고는 한 집안의 글을 모은 세고(世稿) 가운데 가장 빠른 시기의 책으로, 강희맹이 형 강희안과 아버지 강석덕, 할아버지 강회백의 글을 모아 편찬한 문집이다.

양화소록은 총 16가지의 꽃, 나무, 그리고 괴석을 다루는데, 각 편은 해당 식물에 대한 중국 문인들의 언급, 중국 원예서에 등장한 이야기, 그리고 강희안이 직접 식생을 하며 얻은 지식과 감상이 기술되어 있다. 따라서 이 텍스트 안의 지식은 타 문헌으로, 그리고 타 영역으로 확장되고 연결될 수 있는 가능성을 내재하고 있다고 할 수 있다. 또한 단순히 문인의 인문학적 탐색에 그치는 것이 아니라 자연과학에 대한 실천적 탐구까지 담아내고 있다는 점이 특징적이다.

본 연구에서는 디지털 전환을 위한 첫 단계로 강희안의 양화소록에 등장하는 인물과 고사, 지명, 식생에 대한 지식정보, 문헌 등을 XML문서 형식으로 작성하고자 한다. XML로 구현된 하이퍼 텍스트는 그 텍스트 조각이 어떠한 특성을 가지고 있으며, 어떠한 의미맥락 속에 있는 것인지를 명시적으로 나타낼 수 있다.¹ 다시 말해 본문에서 언급되는 텍스트가 인물을 가리키는 것인지, 문헌을 가리키는 것인지, 그리고 그 문헌은 시인지, 편지인지 등 텍스트가 함의하는 바를 기계적으

¹ 김현, 김바로, 임영상, 『디지털인문학입문』, 2016, 124쪽.

로 처리할 수 있도록 서술하는 것이다.

「양화소록」 원문을 XML로 구현한 예시

다음 단계로 이렇게 작성된 기계 가독적인 데이터에서 지식 요소들을 추출하고 지식 사이의 관계를 보여줄 수 있는 RDF(Resource Description Framework) 형식의 시맨틱 데이터베이스²를 구축하고자 한다.

² 시맨틱 데이터란 개체에 대한 개별적 정보뿐 아니라 그것들 사이의 의미적 연관관계에 대한 정보를 담고 있는 데이터베이스를 말한다. 단순히 데이터를 저장하고 저장된 텍스트에 대한 문자열 검색 서비스를 제공하는 종래의 문헌정보 데이터베이스의 수준을 넘어서서, 텍스트 속의 문맥 요소를 식별하고, 이 요소와 유관하다고 정의된 정보를 함께 찾아서 그것들 사이의 연관 관계를 보여주는 기능의 발전된 데이터베이스이다. https://dh.aks.ac.kr/Edu/wiki/index.php/시맨틱_데이터베이스



이때 시맨틱 데이터에 포함되는 지식은 강희안의 「양화소록」에 텍스트로 언급되는 지식뿐만 아니라, 해당 지식과 연결 될 수 있는 문헌, 인물, 이야기, 장소 (경위도 정보), 이동 경로, 재배지, 식물 분류체계, 2 차 연구자료, 멀티미디어자료 등 양화소록의 지식 데이터를 이야기하는 데에 필요한 유관한 정보들을 모두 포



함하게 될 것이다.

본 연구를 통해 산출된 데이터 아카이브를 통해 이용자들은 양화소록이라는 텍스트를 명시적으로 파악할 수 있고, 다른 문헌과 다른 매체를 넘나들며 관련된 지식에 대한 탐색을 이어 나갈 수 있다. 또한 본 연구를 통해 산출된 '양화소록의

지식 요소들을 데이터로 서술하기 위한 명세서³인 온톨로지³는 디지털 전환을 앞둔 다른 문집 연구를 위한 하나의 모델로서 후속 연구에 기여할 수 있을 것이다.

³ 온톨로지란 정보화의 대상이 되는 세계를 전자적으로 표현할 수 있도록 구성한 데이터 기술 체계이다. 컴퓨터가 정보화 대상을 이해하는 틀이 넓은 의미로 온톨로지일 수 있겠지만, 대상자원을 '클래스 (class)'로 범주화하고, 각각의 클래스에 속하는 '개체(individuals)'들이 공통의 '속성(attribute)'을 갖도록 하고, 그 개체들이 다른 개체들과 맺는 '관계(relation)'를 명시적으로 기술하는 것이 가장 일반적인 온톨로지 설계 방법이라고 할 수 있다. 김현, 김바로, 임영상, 『디지털인문학입문』, 2016, 164쪽.