

온톨로지 설계

‘온톨로지’란 정보화의 대상이 되는 세계를 전자적으로 표현할 수 있도록 구성한 데이터 기술 체계이다.¹⁾ 원래 온톨로지라는 말은 철학에서 ‘존재론’이라고 번역되는 용어로서 ‘존재에 대한 이해를 추구하는 학문’의 의미를 갖는 말이었다. 그러한 용어가 정보과학 분야에서 중요한 개념으로 등장하게 된 것은 인간이 세계를 이해하는 틀과 컴퓨터가 정보화 대상(콘텐츠)을 이해하는 틀 사이에 유사성이 있다고 보았기 때문이다. 그 틀은 바로 대상을 구성하는 요소들에 대응하는 개념들과 그 개념들 간의 연관 관계이다.²⁾

넓은 의미에서는 모든 정보화의 틀이 다 온톨로지일 수 있겠지만, 대상 자원을 ‘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.

다음의 예시를 통해 인문지식을 정보화하기 위한 온톨로지는 어떻게 구상하는 것을 좋을지 생각해 보기로 하자.

[예시] 명승 40호 ‘담양 소쇄원’ 안내판의 문화유산 설명문

1) 정보기술 분야에서 말하는 ‘온톨로지(ontology)’에 대한 가장 일반적인 정의는 그루버(Gruber, Thomas. 1959~)가 말한 ‘명시적 명세화의 방법에 의한 개념화’(explicit specification of a conceptualization)이다. (Gruber, ‘A Translation Approach to Portable Ontology Specifications’, *Knowledge Systems Laboratory Technical Report KSL 92-71*, Stanford University, 1992) 여기서 ‘개념화’(conceptualization)라는 것은 정보화하고자 하는 대상 세계를 일정한 체계 속에서 파악하는 것, 예를 들면 그 세계에 무엇이 있고, 그것은 어떤 속성을 품고 있으며, 그것들 사이의 관계는 무엇인가 하는 일정한 질문의 틀 속에서 대상 세계를 이해하는 방식이라고 할 수 있다. ‘명세화’(specification)란 대상 세계에 존재하는 개체, 속성, 관계 등을 일목요연한 목록으로 정리하는 것, 그리고 ‘명시적’(explicit)이라는 그 정리된 목록을 사람뿐 아니라 ‘컴퓨터가 읽을 수 있도록’(machine readable) 한다는 것이다.

2) 김현, 「한국 고전적 전산화의 발전 방향 - 고전 문집 지식 정보 시스템 개발 전략 -」, 『민족문화』 28 (2005)

담양 소쇄원(潭陽 瀟灑園)

명승 제40호
전라남도 담양군 남면 지곡리

소쇄원(瀟灑園)은 자연과 인공을 조화시킨 조선 중기의 대표적인 원림(園林)으로 우리나라 선비의 고고한 품성과 절의가 풍기는 아름다움이 있다.

양산보(梁山甫, 1503~1577)가 조성한 것으로 스승인 조광조(趙光祖)가 유배를 당하여 죽게 되자 출세의 뜻을 버리고 이곳에서 자연과 더불어 살았다. 소쇄원이라 한 것은 양산보의 호(號)인 소쇄옹(瀟灑翁)에서 비롯되었으며, 맑고 깨끗하다는 뜻이 담겨 있다.

오곡문(五曲門) 담장 밑으로 흐르는 맑은 계곡 물은 폭포가 되어 연못에 떨어지고, 계곡가 가까이에는 제월당(霽月堂: 비개인 하늘의 상쾌한 달이라는 뜻의 주인집)과 광풍각(光風閣: 비온 뒤에 해가 뜨며 부는 청량한 바람이란 뜻의 사랑방)이 들어서 있다.

소쇄원에는 영조 31년(1775) 당시 모습을 목판에 새긴 「소쇄원도(瀟灑園圖)」가 남아 있어 원형을 추정할 수 있다. 이곳은 많은 학자들이 모여들어 학문을 토론하고, 창작 활동을 벌인 선비정신의 산실이기도 하다. 지금의 소쇄원은 양산보의 5대손 양택지에 의해 보수된 모습이다.



위 예시의 글은 전라남도 담양군에 위치한 조선시대의 대표적인 정원인 소쇄원(瀟灑園)의 문화재 안내판 해설문이다. 이 문화유산과 관련된 다양한 지식의 실마리를 위의 해설문에서 찾고, 이것을 유형별로 정리하면 다음과 같은 키

워드 목록을 얻을 수 있을 것이다.

유형 범주		키워드
문화 유산	지정문화재	담양 소쇄원
	건조물	제월당, 광풍각, 오곡문
	기록물	소쇄원도
인물		양산보, 조광조, 양택지
장소(공간)		전라남도 담양군 남면 지곡리
연대(시간)		영조 31년(1775)
개념		원림(園林), 절의(節義)

위의 표에 정리한 어휘들은 담양 소쇄원 문화재 안내판 해설문에 적힌 내용의 뼈대를 이루는 문맥 요소(Contextual Element)들이다. 이 문맥 요소들은 문장 속의 단어로만 존재하는 것이 아니라, 소쇄원에 얽힌 역사, 그것의 문화적 의미에 관한 지식을 대표하는 용어라고 할 수 있다. 이것을 근거로, 담양 소쇄원 관련 지식의 관계도를 만들어 보기로 한다.

① 개체(individual) 탐색

온톨로지 설계를 위해 제일 먼저 할 일은 정보화 하고자 하는 지식 세계에 어떠한 지식 요소들이 있는지 탐색하고, 그 성격을 분석하는 것이다. 위의 예시에서 목록화한 개개의 요소들을 온톨로지 용어로는 ‘개체’라고 한다. 이것은 정보화하고자 하는 지식의 단위 요소이자, 관련 지식의 네트워크에서 관계의 접점이 될 노드(node)이다.

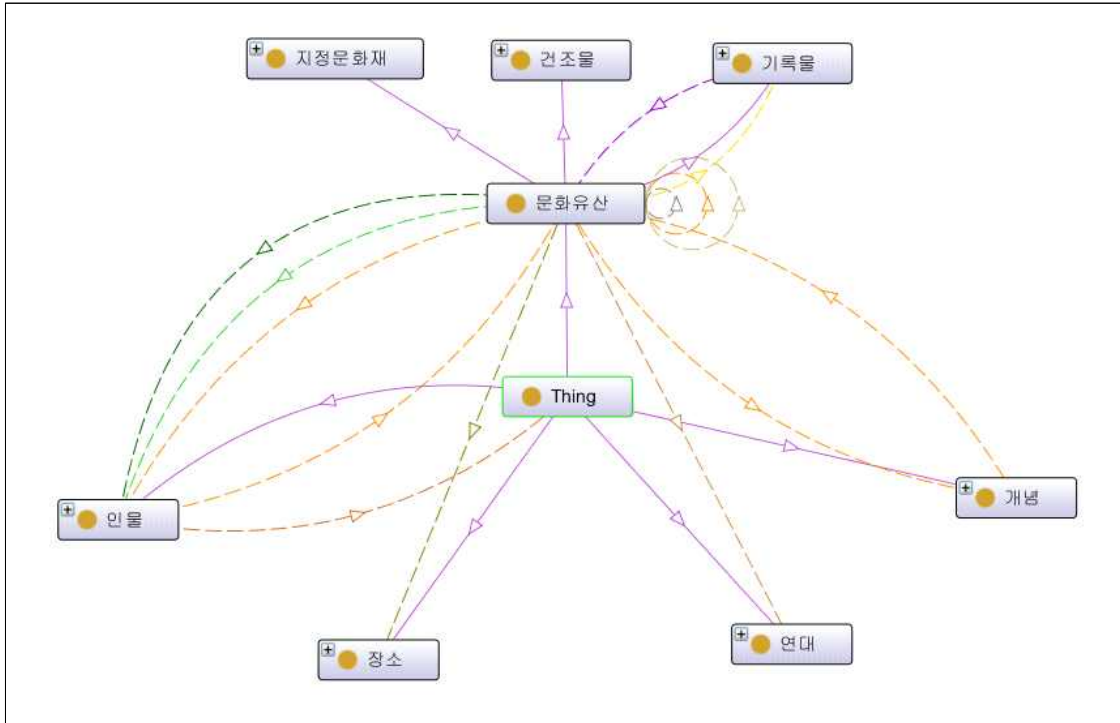
Individuals:



② 클래스(class) 설계

정보화의 대상 세계에 대체로 어떠한 성격의 문맥 요소가 존재하는지 파악 되었으면, 그 다음에 할 일은 그 요소들 가운데 서로 유사한 성격의 것들을 묶어줄 수 있는 범주를 정하는 것이다. 온톨로지 용어로는 이러한 범주를 클래스라고 한다. 이 클래스 설계는 대상 세계를 무차별적인 사실/사물의 나열로 보지 않고, 일정한 체계 속에서 이해하려는 노력이라고 할 수 있다. 따라서 어떠한 클래스를 정하느냐 하는 것은 연구자가 대상을 어떠한 시각에서 보고자 하느냐에 따라 달라진다. 여기서는 ‘담양 소재원’을 ‘문화유산’이라는 관점에서 보고, 그 문화유산의 역사적·문화적 문맥을 이해하는 데 도움이 되는 방향으로 지식의 체계를 만들어 보기로 한다. 일단, 위에서 나열한 12개의 문맥 요소를 분류할 수 있는 체계를 다음과 같이 설계하였다.

Class:



이 그림의 가운데에 그려진 ‘Thing’은 대상 세계의 모든 존재를 포괄하는 가상의 최상위 범주이다. 그 아래에 ‘문화유산’, ‘인물’, ‘장소’, ‘연대’, ‘개념’ 등 5가지 차상위 클래스를 두기로 하고, 다시 ‘문화유산’ 클래스 밑에 ‘지정문화재’, ‘건조물’, ‘기록물’ 등의 하위 클래스를 두었다. 문화유산 클래스를 이렇게 중층으로 구성한 것은 예시로 든 문화유산 사례에서 보듯이 ‘담양 소쇄원’이라는 지정문화재(명승 제40호) 안에 ‘광풍각’, ‘제월당’과 같은 부속 건조물도 있고, ‘소쇄원도’와 같은 기록물도 있기 때문에 디지털 정보세계가 그와 같은 체계를 반영할 수 있도록 하기 위해서였다. 문화유산의 범위를 넓히게 되면 또 다른 하위 요소도 추가될 수 있을 것이다. 유적지 경내에 있는 노거수(老巨樹), 묘정비(廟庭碑) 등도 고려 대상이 될 수 있다.

③ 속성(attribute) 설계

클래스 설계가 이루어졌으면, 각각의 클래스에 속하는 개체(individual)들이 어떤 속성을 갖는지를 살피고, 이 속성을 담을 수 있는 틀을 만들어야 한다. 예를 들어 ‘인물’ 클래스에 속하는 개체들에 대해서는 이름, 이름의 한자 표기,

별명(자, 호, 봉작호, 시호 등), 생몰년, 관직 등의 정보를 부여하는 것이다. 클래스 설계와 마찬가지로 속성 설계도 ‘이렇게 해야 한다’는 절대적인 기준은 없다. 대상 개체를 어떤 관점에서 보고 있으며, 그 정보를 어떠한 목적으로 사용할 것인가를 기준으로 삼으면 된다. 간단한 예시를 위해 여기서는 다음과 같이 최소한의 속성만을 부여하기로 한다.

Attributes:

Class	한자	유형	생년	몰년	위도	경도	서력	설명
지정문화재	○	○						○
건조물	○	○			○	○		○
기록물	○	○						○
인물	○	○	○	○				○
장소	○	○			○	○		○
연대	○	○					○	○
개념	○	○						○

④ 관계성(relation) 설계

각각의 클래스에 속하는 개체들이 서로 어떠한 의미적 연관 관계를 맺고 있는지 분석하여, 그 관계성을 표현할 수 있는 서술어를 정한다. 이 관계성 서술어는 주어와 목적어를 수반하여 RDF(Resource Description Framework) 문을 완성하게 된다. 이 때 주어 역할을 할 수 있는 개체들의 클래스를 ‘정의역’(定義域, domain)이라고 하고, 목적어 역할을 하는 개체들의 클래스를 ‘치역’(値域, range)라고 부른다. 담양 소재원과 관계가 있는 개체들을 기준으로 아래와 같은 17개의 관계 서술어를 정하였다.³⁾

3) 온톨로지의 어휘들은 컴퓨터 시스템이 인식할 수 있도록 일관성 있게 사용하면 되는 것이기 때문에 원천 어휘를 한국어로 하느냐 영어로 하느냐는 기술적으로 문제되지 않는다. 하지만 온톨로지 공유를 통해 소통하고자 하는 범위가 국제적이라면, 세계 여러 나라에서 이미 사용하고 있는 동의어의 사용이 권장된다. 이 책에서의 온톨로지 설계는 교육의 목적으로 하는 것이기 때문에 학생들이 이해하기 쉽도록 한국어 어휘를 위주로 하였다. 다만 관계성 서술어는 한국어 어휘를 쓸 경우 주어와 목적어의 식별이 모호해질 수 있기 때문에 편의상 영어식의 조어(造語)를 사용하기로 한다.

Relations:

Relation	Domain (A)	Range (B)	설명
created	문화유산	연대	A는 B에 만들어졌다
creator	문화유산	인물	A는 B가 만들었다
renovator	문화유산	인물	A는 B가 중건(중수)했다
depicts	기록물	문화유산	A는 B를 묘사한다
isDepictedIn	문화유산	기록물	A는 B에 묘사되었다
hasDescendant	인물	인물	A는 B를 후손으로 두었다
isDescendantOf	인물	인물	A는 B의 후손이다
hasDisciple	인물	인물	A는 B를 문인으로 두었다
isDiscipleOf	인물	인물	A는 B의 문인이다
hasPart	문화유산	문화유산	A는 B를 포함한다
isPartOf	문화유산	문화유산	A는 B의 일부이다
hasSon	인물	인물	A는 B의 아버지이다
isSonOf	인물	인물	A는 B의 아들이다
isEnshrinedIn	인물	문화유산	A는 B에서 제향된다
isFriendOf	인물	인물	A와 B는 친구이다
isLocatedIn	문화유산	장소	A는 B에 위치한다
isRelatedTo	개념, 인물, 문화유산	개념, 인물, 문화유산	A와 B는 관계가 있다

⑤ 개체(individual)에 대한 속성, 관계성 부여

지금까지 수행한 네 가지 과정이 온톨로지 설계에 해당하는 일이라고 한다면 이 다섯 번째 작업은 설계된 온톨로지에 따라 데이터베이스를 구축하는 일이다.⁴⁾ 문맥 요소로 발굴된 모든 개체를 해당 클래스에 귀속시킨 후, 각각의 개체에 고유한 속성을 부여하고, 개체와 개체 사이의 관계성을 지정한다.

§ Individual ‘담양 소쇄원’

* Class: 문화유산 > 지정문화재

* Attribute

한자: 潭陽 瀟灑園

유형: 명승

* Relation

4) 온톨로지 설계에 의해 만들어진 데이터베이스를 일반적인 관계형 데이터베이스(Relational Database, RDB)와 구분하여 ‘온톨로지 기반 데이터베이스’(Ontology Based Database, ODB)라고 부르기도 한다. 실험적인 ODB는 온톨로지 설계 도구를 이용하여 구현할 수 있다.

hasPart: 광풍각
 hasPart: 소쇄원도
 hasPart: 오곡문
 isDepictedIn: 소쇄원도
 creator: 양산보
 renovator: 양택지

§ Individual: '양산보'

* Class: 인물

* Attribute:

생년: 1503

몰년: 1557

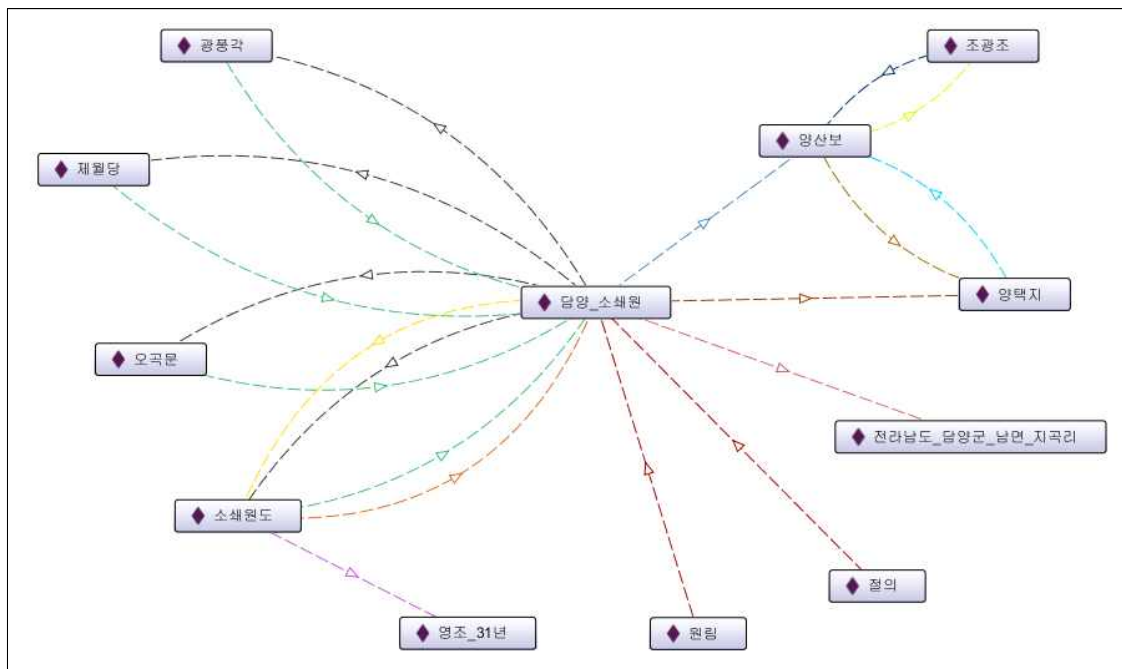
한자: 梁山甫

* Relation

hasDescendant: 양택지

isDiscipleOf: 조광조

12개 개체(individuals)에 개체에 대해 모두 이러한 작업을 수행하면 아래 그림과 같이 관계의 맥락을 명시적으로 드러내게 된다.



위의 그래프에 담겨 있는 정보는 다음과 같은 것들이다.

[문화유산]‘담양 소쇄원’은 [장소]‘전라남도 담양군 남면 지곡리’에 위치한다.

[문화유산]‘담양 소쇄원’은 [인물]‘양산보’가 세웠다.

[인물]‘양산보’는 [인물]‘조광조’의 제자이다.

[문화유산]‘담양 소쇄원’은 [개념]‘원림’/‘절의’와 관련이 있다.

[문화유산]‘담양 소쇄원’ 안에는 [건축물]‘광풍각’/‘제월당’/‘오곡문’이 있다.

[문화유산]‘담양 소쇄원’ 안에는 [기록물]‘소쇄원도’가 있다.

[기록물]‘소쇄원도’는 [문화유산]‘담양 소쇄원’을 묘사했다.

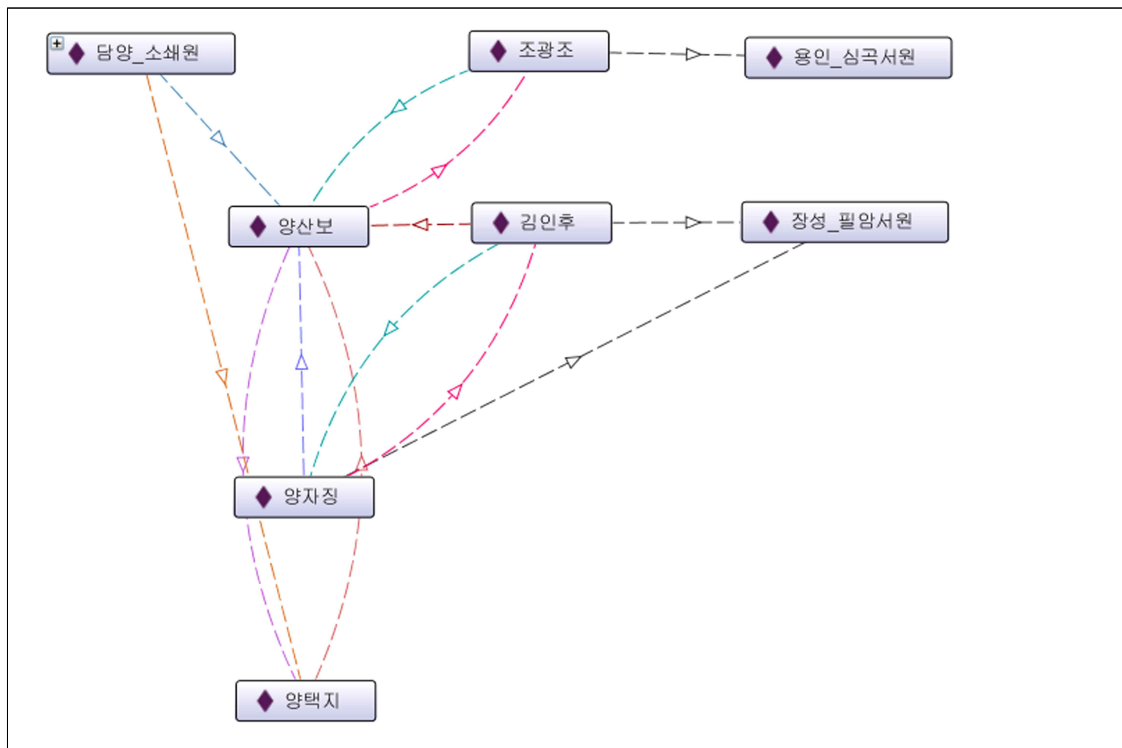
[기록물]‘소쇄원도’는 [연대]‘영조 31년’에 제작되었다.

[인물]‘양택지’는 [인물]‘양산보’의 후손이다.

[인물]‘양택지’는 [문화유산]‘담양 소쇄원’을 중수했다.

이 데이터는 이것의 원시 데이터라고 할 수 있는 안내판 해설문 속의 정보를 컴퓨터가 읽고 해석할 수 있는 형태로 가공한 것이라고 할 수 있다. 이와 같은 기계가독형(Machine-Readable) 데이터를 만들어서 얻는 이점은 무엇일까? ‘담양 소쇄원’이라는 하나의 문화유산 정보만 디지털화한 상황에서는 온톨로지 기반 데이터베이스의 장점을 드러내기 힘들다. 하지만 이러한 정보가 수천 건, 수만 건 있다고 가정해 보면, 다음과 같은 효과를 쉽게 기대할 수 있다.

여러 가지 기대 효과 가운데 가장 우선시되는 것은 별개의 문화유산 해설문 속에 담겨 있는 정보들이 서로서로 관련성을 좋아 광대한 의미의 연결망을 만들 수 있을 것이라는 점이다. ‘용인 심곡서원’에는 양산보의 스승 조광조가 제향되어 있다. ‘장성 필암서원’에 제향된 김인후는 양산보의 친구이고, 김인후와 함께 제향된 그의 제자 양자징은 양산보의 아들이다. 이 세 곳의 문화유산들은 공통의 관련 인물을 매개로 다음 그림과 같은 관계망을 형성한다.



이와 같이 서로서로 의미의 연결고리를 맺고 있는 지식의 네트워크 속에서는 그 관계를 추론하여 새로운 지식을 얻는 것도 가능하다. 우리가 이 예시를 위해 만든 데이터 안에서도 ‘소쇄원을 만든 사람의 스승은 누구인가?’, ‘그를 제향한 서원은 어디인가?’라는 질문을 던지면, 컴퓨터는 그 질문에 대해 정확한 답변을 제공한다. 이 온톨로지를 공유하는 더 많은 데이터가 만들어졌을 경우, 컴퓨터는 보다 유용한 지식 탐구의 동반자 역할을 할 수 있을 것이다.

