

김현승, 「디지털 한국어문학 연구 방법론 2」, 성균관대학교 수업 자료, 2025.

온톨로지와 네트워크 그래프

시맨틱 데이터 편찬 도구

한국학중앙연구원 디지털인문학연구소

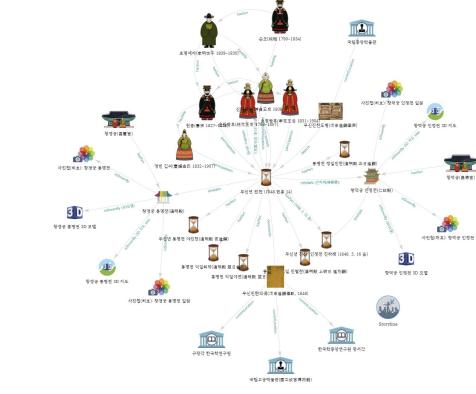
김현승

Intro

- 온톨로지에 대해 알아보기 전에, 온톨로지 설계가 필요한 이유에 대해 이야기해 보자.

디지털 큐레이션

- 디지털 환경에서 인문학 지식을 자기주도적으로 탐구하고 표현할 수 있는 공부 방법이다.
 - 전통적인 인문지식을 디지털 세계에서 소통할 수 있는 **데이터**로 전환하고, **의미 있는 맥락**을 형성한다.
 - 지식 정보 간의 유의미한 관계성을 탐구하며, 이를 통해 새로운 사실을 발견하기도 하고, 대상 세계에 대한 이해를 넓혀갈 수 있는 **융합적 지식 탐구 방법**이다.
 - 디지털 세계에서는 박물관, 연구실, 강의실, 기록 보관소 등의 공간이 분리될 필요가 없다.



2009년 미국에서 발표된 디지털인문학 선언 2.0

확장된 학술 활동으로서의 큐레이션

디지털 인문학자들은 큐레이션을 미래 인문학의 중심적인 특징으로 인식한다. 근대 대학이 큐레이션을 이차적 보조적 역할로 강등시키고 큐레이터를 박물관, 아카이브 및 도서관으로 추방하면서 학문에서 분리시켰던 반면, 디지털 인문학 혁명은 연구와 교육의 지형도를 근본적으로 재형성하도록 촉진한다. **디지털 인문학은 학자를 큐레이터로, 큐레이터를 학자로 재구성** 하고, 이로 인한 가능성과 수요의 확장을 통해 학문 활동에 다시 활기를 북돋으며 박물관, 도서관 및 아카이브의 학술적 사명을 재개하려 한다.

The Digital Humanities Manifesto 2.0, 2009. https://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf

시맨틱(Semantic)?

시맨틱(Semantic) : 의미론(意味論)

☞ 그리스어 σημαίνω '의미하다'

☞ *semantics*

(프랑스어: *sémantique*, 독일어: *Semantik*)

시맨틱(Semantic) = “의미 있는 맥락”

시맨틱 웹 Semantic Web

- 월드와이드웹(World Wide Web) 상에서 지식 정보(개체, individual)들이 서로서로 어떤 관계를 맺는지 명시적으로 드러내고, 이를 바탕으로 개체와 개체 사이의 관계를 추론하여 새로운 사실을 발견하거나 종합적으로 지식을 얻으려는 것이다.
- 미래의 월드와이드웹에 대한 기대와 이상을 표현한 용어이다.
- 월드와이드웹의 창시자 팀 버너즈 리(Tim Berners-Lee)가 1998년에 제안하였다.

시맨틱 웹 Semantic Web



“현재의 웹에 명확한 의미의 정보를 부가하여
사람과 컴퓨터가 협업할 수 있도록 하는 것”

Timothy John Berners-Lee (1955~)

시맨틱 웹 Semantic Web

- ‘시맨틱 웹’의 이상적인 목표를 이루기 위한 첫 단계는 월드와이드웹 문서를 작성할 때 이것이 다른 것과 어떤 관계를 갖는지 표시해주는 것이다.
- Linked Data : 정보 개체들이 서로 관계가 있음을 알리는 데이터의 집합
- Linked Open Data, LOD : 링크드 데이터에 누구나 그 연결을 확장해 갈 수 있도록 하는 것

Tim Berners-Lee, “Linked Data”, W3C Design Issues (2006, 2009)

<https://www.w3.org/DesignIssues/LinkedData.html>

The screenshot shows a web page with the following details:

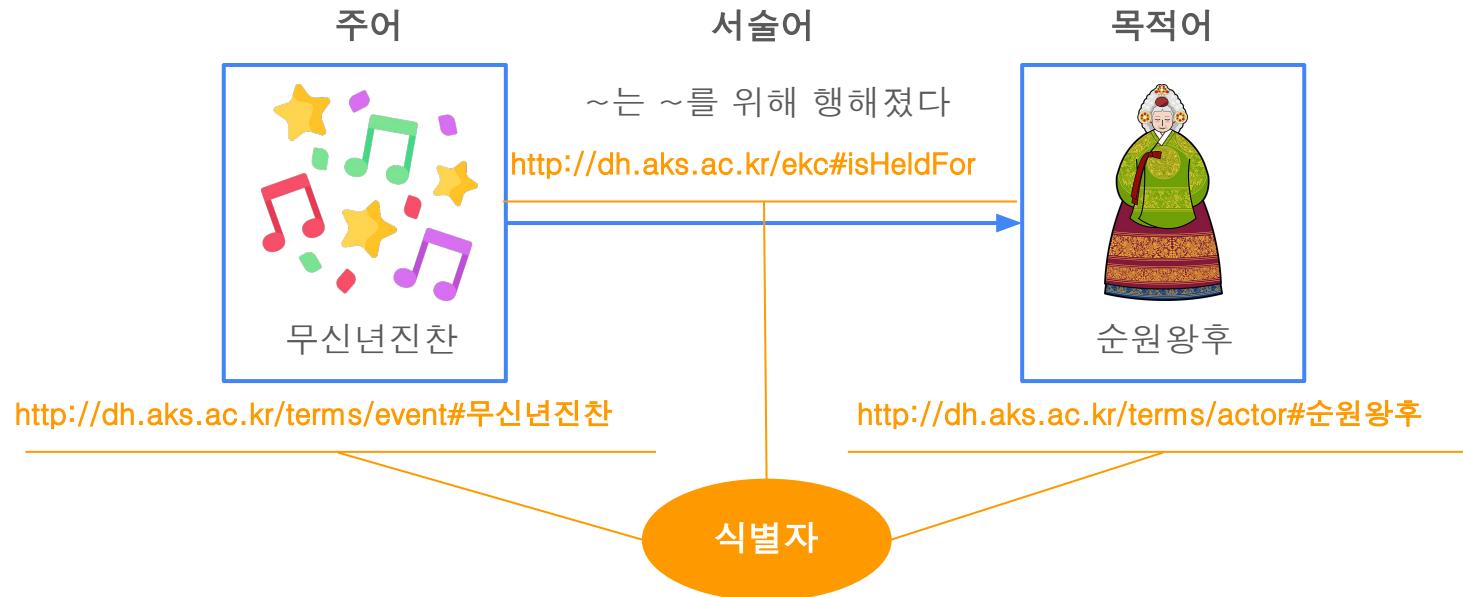
- Header:** Tim Berners-Lee
Date: 2006-07-27, last change: \$Date: 2009/06/18 18:24:33 \$
Status: personal view only. Editing status: imperfect but published.
Up to Design Issues
- Section:** Linked Data
- Text:** The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.
Like the web of hypertext, the web of data is constructed with documents on the web. However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they links between arbitrary things described by RDF. The user identify any kind of object or concept. But for HTML or RDF, the same expectations apply to make the web grow.
- List:** 1. Use URIs as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs, so that they can discover more things.
- Footnote:** Simple. In fact, though, a surprising amount of data isn't linked in 2006, because of problems with one or more of the steps. This article discusses solutions to these problems, details of implementation, and factors affecting choices about how you publish your data.

A white ceramic mug is shown on the right, featuring the text "LINKED OPEN DATA" and "IS YOUR DATA 5 *?". Below the text are several yellow stars and the words "On the web", "Machine-readable", "Non-proprietary", "RDF standards", and "Linked RDF".

RDF, Resource Description Framework

- 개체 사이의 관계성을 표시해 주는 방법이다.
- 월드와이드웹 세계에서 데이터를 기술하는 표준 형식으로 사용 중이다.
- 개체에 식별자를 부여하고, 두 개체 사이의 관계성을 **약속된 서술어**로 기술하는 방식이다.
- 고유한 식별자가 부여된 개체와 개체 사이의 관계, 그리고 개체와 속성의 관계가 사람을 비롯하여 컴퓨터도 인식할 수 있는 표준적인 서술어로 기술된다면, 월드와이드웹은 단순한 지식의 조각을 찾는 수준이 아니라 더 나아가 의미적 연관 관계를 맺는 지식 정보들을 연결해 주는 역할까지 할 수 있게 될 것이다.

RDF, Resource Description Framework



시맨틱 웹 구현 핵심 요소



의미가 통하는 시맨틱 웹을 만들기 위해 데이터 기술문의 어휘에 관해 일정한 약속을 만들어 사용해야 한다. 정보 기술 세계에서는 이런 약속을 ‘온톨로지’라고 부른다.

김현 외, 『디지털 인문학 입문』, 한국외국어대학교 지식출판원, 2016, 163쪽.

온톨로지 Ontology

- 온톨로지란, 정보화의 대상이 되는 세계를 전자적으로 표현할 수 있도록 구성된 데이터 기술 체계이다.
- 철학 용어 ‘존재론’ : 사물의 존재 의미를 논의하는 철학적인 연구 영역을 뜻하는 말이다.

☞ 정보과학 분야에서도 중요한 개념으로 등장

“인간이 세계를 이해하는 틀과 컴퓨터가 정보화 대상을 이해하는 틀 사이에 유사성이 있다고 본다. 그 틀은 대상을 구성하는 요소들에 대응하는 개념들과 그 개념들 간의 연관 관계이다.”

김현 외, 『디지털 인문학 입문』, 한국외국어대학교 지식출판원, 2016, 164쪽.

온톨로지

文 A 34개 언어 ▾

문서 토론

읽기 편집 역사 보기 도구 ▾

위키백과, 우리 모두의 백과사전.



이 문서는 전산학과 컴퓨터 과학에서의 온톨로지에 관한 것입니다. 철학의 존재론에 대해서는 [존재론](#) 문서를 참고하십시오.



이 문서에는 여러 문제가 있습니다. [문서를 편집하여 수정하시거나](#) [토론 문서에서 의견을 나눠주세요.](#) [펼치기]

온톨로지(Ontology)란 사람들이 세상에 대하여 보고 듣고 느끼고 생각하는 것에 대하여 서로 간의 토론을 통하여 합의를 이룬 바를, 개념적이고 컴퓨터에서 다룰 수 있는 형태로 표현한 모델로, 개념의 타입이나 사용상의 제약조건들을 명시적으로 정의한 기술이다. 온톨로지는 일종의 지식표현(knowledge representation)으로, 컴퓨터는 온톨로지로 표현된 개념을 이해하고 지식처리를 할 수 있게 된다. 프로그램과 인간이 지식을 공유하는데 도움을 주기 위한 온톨로지는, 정보시스템의 대상이 되는 자원의 개념을 명확하게 정의하고 상세하게 기술하여 보다 정확한 정보를 찾을 수 있도록 하는데 목적이 있다. 온톨로지는 시맨틱 웹을 구현할 수 있는 도구로서, 지식개념을 의미적으로 연결할 수 있는 도구로서 RDF, OWL, SWRL 등의 언어를 이용해 표현한다.

온톨로지는 일단 합의된 지식을 나타내므로 어느 개인에게 국한되는 것이 아니라 그룹 구성원이 모두 동의하는 개념이다. 그리고 프로그램이 이해할 수 있어야 하므로 여러 가지 정형화가 존재한다.

정보과학

일반 관점

접근 · 아키텍처 · 행동 · 관리 · 검색 · 추구 · 사회 · 지식 조직화 · 온톨로지 · 철학 · 과학기술학 · 분류 체계

관련 분야 및 하위 분야

계량서지학 · 분급 · 검열 · 도서 분류 · 기억 장치 · 문화연구 · 데이터 모델링 · 인포메틱스 · 정보기술 · 지적 자유 · 지식 재산권 · 문헌정보학 · 기억 · 보존 · 프라이버시 · 양자정보과학

V · T · E

[IT용어 아하!] 온톨로지(Ontology)

입력 2007.03.14. 오전 6:11 기사원문



공감



댓글



가가



▶



온톨로지(Ontology)는 원래 사물의 존재 의미를 논의하는 철학적인 연구 영역을 뜻하는 말입니다. 온톨로지라는 단어 자체부터 실제라는 의미의 그리스어 'onto'와 학문 또는 강연 등의 의미를 갖는 'logia'의 합성어로부터 유래됐습니다.

그러나 최근의 시맨틱 웹, 지식공학, 인공지능, 자연어처리 등 정보기술 분야에서의 온톨로지는 각각의 지식(혹은 단어, 개념)이 전체 지식체계 중에서 어디에 위치하는지를 밝히는 연구 분야를 의미하는데요. 즉, 어떤 단어와 단어 사이의 상관관계를 보다 빠르고 편하게 검색할 수 있도록 돋는 연구 분야를 의미합니다.

미국 스탠포드 대학 인공지능 전문가인 탐 그루버는 인공지능분야에 있어서의 온톨로지는 컴퓨터로 하여금 인간의 지식체계를 갖도록 개념화시키는 작업이라 설명하고 있습니다.

< 모바일로 보는 디지털타임스 3553+NATE/magic®/ez-i >

온톨로지 Ontology



클래스 (Class) : 다양한 사물(대상)들을 그룹화하는 범주(유형) 또는 카테고리

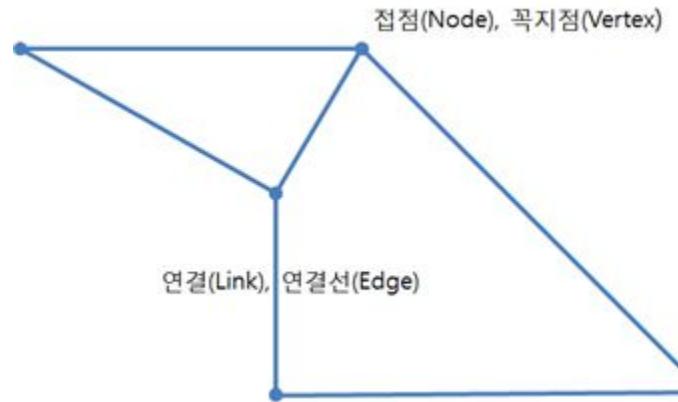
개체 (Individuals) : 클래스에 속하는 개개의 사물(대상)

속성 (Attribute) : 하나의 클래스(범주, 유형)를 설명하는 정보들

관계 (Relation) : 개체와 개체 사이의 의미적 연관관계

네트워크 그래프

- 네트워크 그래프 : 개체와 개체 사이의 관계를 그래프로 보여 주는 데이터의 시각화 기술 중 하나이다.
- 구성 요소 : 개체의 목록과 각 개체 사이를 연결하는 관계성의 목록이 필요하다. 개체가 연결되어 만나는 점집을 노드(Node)라 하고, 그 연결선을 링크(Link)라 한다.



네트워크 그래프

- 네트워크 그래프 : 개체와 개체 사이의 관계를 그래프로 보여 주는 데이터의 시각화 기술 중 하나이다.
- 구성 요소 : 개체의 목록과 각 개체 사이를 연결하는 관계성의 목록이 필요하다. 개체가 연결되어 만나는 점집을 노드(Node)라 하고, 그 연결선을 링크(Link)라 한다.
- 데이터 시각화 필요성 : 우리가 이미 알고 있는 것을 남들에게 보여 주기 위해서 만은 아니다. 이를 통해 더욱 정확한 데이터를 만들 수 있고, 그 속에서 미처 몰랐던 새로운 사실을 발견할 수도 있다.

네트워크 그래프

- 네트워크 그래프 : **개체와 개체 사이의 관계**를 그래프로 보여 주는 데이터의 시각화 기술 중 하나이다.
- 구성 요소 : 개체 **관계 중심의 시맨틱 데이터**의 목록이 필요하다.
개체가 연결되어 만나는 점(집합)을 노드(Node)라고 하고, 그 연결선을 링크(Link)라고 부른다.
- 데이터 시각화 필요성 : 우리가 이미 알고 있는 것을 남들에게 보여 주기 위해서만은 아니다. 이를 통해 더욱 정확한 데이터를 만들 수 있고, 그 속에서 미처 몰랐던 새로운 사실을 발견할 수도 있다.

네트워크 그래프

- 네트워크 그래프 : 개체와 개체 사이의 관계를 **그래프**로 보여 주는 데이터의 시각화 기술 중 하나이다.
- 구성 요소 : 개체가 연결되어 있는 **정점**을 노드(Node)라고 하고, 그 연결선을 링크(Link)라고 부른다.
흥미에 따른 자유로운 탐색의 목록이 필요하다.
- 데이터 시각화 필요성 : 우리가 이미 알고 있는 것을 남들에게 보여 주기 위해서만은 아니다. 이를 통해 더욱 정확한 데이터를 만들 수 있고, 그 속에서 미처 몰랐던 새로운 사실을 발견할 수도 있다.

네트워크 그래프

- 네트워크 그래프 : 개체와 개체 사이의 관계를 그래프로 보여 주는 데이터의 시각화 기술 중 하나이다.
데이터들이 어떻게 연결되어 있는지 시각적으로 확인 가능
- 구성 요소 : 개체가 연결되어 만나는 점집을 노드(Node)라 하고, 그 연결선을 링크(Link)라고 부른다.
- 데이터 시각화 필요성 : 우리가 이미 알고 있는 것을 남들에게 보여 주기 위해서만은 아니다. 이를 통해 더욱 정확한 데이터를 만들 수 있고, 그 속에서 미처 몰랐던 새로운 사실을 발견할 수도 있다.

네트워크 그래프

- 네트워크 그래프 : 개체와 개체 사이의 관계를 그래프로 보여 주는 데이터의 시각화 기술 중 하나이다.
- 구성 요소 **개체의 목록**과 각 개체 사이를 연결하는 **관계성의 목록**이 필요하다.
개체가 연결되어 만나는 점집을 노드(Node)라 하고, 그 연결선을 링크(Link)
라고 부른다.
- **온톨로지 설계를 기반으로 하여 데이터 추가 및 유지가 가능하고,**
방대한 데이터의 구조를 시각화할 수 있다.
볼랐던 새로운 사실을 발견할 수도 있다.

네트워크 그래프

- 네트워크 그래프 : 개체와 개체 사이의 관계를 그래프로 보여 주는 데이터의 시각화 기술 중 하나이다.
- 구성 요소 : 개체의 목록과 각 개체 사이를 연결하는 관계성의 목록이 필요하다. 개체가 연결되어 만나는 점집을 노드(Node)라 하고, 그 연결선을 링크(Link)라고 부른다.
- 데이터 시각화 필요성 : 우리가 이미 알고 있는 것을 남들에게 보여 주기 위해서 만은 아니다. 이를 통해 더욱 정확한 데이터를 만들 수 있고, 그 속에서 미처 몰랐던 새로운 사실을 발견할 수도 있다.

네트워크 그래프

네트워크 그래프의 시각화 종류는 다양하며, 여러 분야에서 데이터의 정보와 그 규모를 분석하는 데 사용하지만, 우리는 데이터가 서로간에 어떤 관계를 맺고 있는지 그 의미와 구조를 이해하고, 각 데이터가 개념에 적합하게 연결되었는지 검증하며, 혹시 누락된 정보가 있지는 않은지, 외롭게 고립된 개체나 반대로 과도하게 집중된 개체가 발생했는지 등을 점검하기 위한 **시맨틱 데이터 편찬 도구로 활용** 한다.

