KiSTi 지식리포트 2015. 10. 16(No.45)

학술분야 시맨틱 퍼블리싱 (Semantic Publishing) 연구 동향

현미환 · 이혜진 · 김혜선 · 장원홍

KiSTi 한국과학기술정보연구원
www.kisti.re.kr
Korea Institute of Science and Technology Information

차 례

| 요 약 |

1. 서 론 ····· 1

2. 연구동향 ····· 3

3. 주요 연구내용 ····· 5
   (1) 데이터 모델 및 온톨로지 개발 연구 ····· 5
   (2) 시맨틱 퍼블리싱 적용 및 효과 연구 ····· 6
   (3) 플랫폼 개발 연구 ····· 8
   (4) 시각화 연구 ····· 9

4. 시맨틱 퍼블리싱 관련 온톨로지 및 데이터 모델 ····· 11
   (1) DO(Document Ontology) ····· 12
   (2) SPAR(Semantic Publishing and Referencing Ontologies) ····· 13
   (3) ORB(Ontology of Rhetorical Blocks) ····· 15
   (4) EXPO Ontology ····· 15
   (5) SWAN(Semantic Web Application in Neuromedicine) 온톨로지 ····· 16
   (6) BIBO(Bibliographic Ontology Specification) ····· 17
   (7) 더블린 코어(Dublin Core) ····· 17
   (8) FOAF(Friend of a Friend) ····· 18
   (9) DOAF(Description of a Project) ····· 19
   (10) SKOS(Simple Knowledge Organization System) ····· 19
   (11) W3C Open Annotation Model ····· 19

5. 시맨틱 퍼블리싱 주요 기법 및 적용 ····· 12
(1) RDFa 생성 및 삽입 ····· 21
(2) 시맨틱 어노테이션(Semantic Annotation) ····· 23

6. 시맨틱 퍼블리싱 플랫폼 ····· 26
(1) Ontotext Dynamic Semantic Publishing Platform ····· 26
(2) Ambra Project ····· 29
(3) OSF(Open Semantic Framework) ····· 30
(4) LMF(Linked Media Framework) ····· 32
(5) redlink ····· 33
(6) Ximdex ····· 33

7. 시각화 ····· 35
(1) 학술정보 시각화 사례 ····· 35
(2) 학술정보 시각화 관련 툴 ····· 36

8. 활용사례 ····· 37
(1) Semantic Digital Library 관련 사례 ····· 37
(2) LOD 기반 사례 ····· 38

| 참고문헌 | ····· 41

표 차례

<표 1> 시맨틱 퍼블리싱 연구동향 ····· 3
<표 2> 주요 semantic enhancement 기법 ····· 7
<표 3> 학술정보 관련 온톨로지 구분 ····· 11
<표 4> SPAR 구성 온톨로지의 용도와 특징 ····· 41
<표 5> SKOS 어휘 ····· 91
<표 6> Open Annotation의 구성 요소 ····· 02
<표 7> 주요 Semantic Annotation 툴 및 플랫폼 ····· 42

그림 차례

[그림 1] 온톨로지를 활용한 학술 저작물의 기술 ····· 6
[그림 2] Semantic enhancement의 사례 ····· 8
[그림 3] Document Ontology 계층 구조 ····· 21
[그림 4] SPAR 구성 온톨로지와 상징 ····· 31
[그림 5] SPAR 온톨로지 아키텍처 ····· 41
[그림 6] EXPO 온톨로지의 개념적 위치 ····· 61
[그림 7] BIBO 온톨로지를 활용하여 기술된 프로시딩 정보의 예시 ····· 71
[그림 8] 더블린 코어 기술 예시 ····· 81
[그림 9] FOAF 기술 예시 ····· 81
[그림 10] Dublin Core와 FOAF를 활용한 RDFa 1.0 기술 예시 ····· 22
[그림 11] Semantic Annotation 기법의 분류 ····· 32
[그림 12] Ontotext GraphDB에 대한 추론 예시 ····· 72
[그림 13] Ontotext Scientific Publishing 절차에 관한 인포그래픽 ····· 82
[그림 14] Abmra 플랫폼 기반의 PLOS One 서비스 화면 예시 ····· 03
[그림 15] OSF 아키텍처 ····· 13
[그림 16] LMF 아키텍처 ····· 23
[그림 17] Ximdex 플랫폼 semantic enhancement 데모 화면 ····· 43
[그림 18] Imformation Science 분야에 대한 impact factor 시각화 예시 ····· 53
[그림 19] 마이크로소프트 Academic Search 시각화 화면 예시 ····· 63
[그림 20] Semantic Lancet 서비스 아키텍처 ····· 93
[그림 21] 생물의학 개체 간 관계의 시각화 예시 ····· 04

요 약

■ 시맨틱 웹(Semantic Web)의 활용

- 시맨틱 웹(Semantic Web)의 등장 이후 이를 실제적으로 구현하기 위한 기술이 발전함에 따라 지식관리, 미디어 관리 및 콘텐츠 관리 등 다양한 분야에 걸쳐 이를 적용하려는 연구가 활발하게 진행되어 왔음.

■ 학술분야에서 시맨틱 웹(Semantic Web) 기술의 활용

- 학술 출판에 있어서도 시맨틱 웹 기술을 적용하여 기존의 방식과는 크게 차별화가 될 만큼 풍부한 부가정보와 의미 정보를 제공할 수 있도록 하는 시맨틱 퍼블리싱(Semantic Publishing)에 관한 연구가 활발하게 진행되어 온 상태임.
- 시맨틱 퍼블리싱에 관한 연구는 크게 4가지 분야를 중심으로 진행되어 왔음.
  1) 데이터 모델과 온톨로지 설계
  2) 시맨틱 퍼블리싱 적용의 효과
  3) 시맨틱 퍼블리싱 플랫폼 개발
  4) 시각화 구현
- 이들 연구 분야 중 특히 시맨틱 퍼블리싱 전 과정을 효율적으로 관리할 수 있도록 지원하는 시맨틱 퍼블리싱 플랫폼과 연구와 편의성과 다른 차원의 이해를 제공하는 시각화에 관한 연구가 크게 현재 주목을 받고 있음.

1. 서 론

□ 시맨틱 퍼블리싱(Semantic Publishing)은 인터넷 기술의 발전과 함께 전자저널(Electronic Journal)의 출판 및 활용과 관련하여 2000년대 후반부터 현재에 이르기까지 꾸준하게 연구되고 있는 분야임.

□ Web 2.0과 Semantic Web이라는 개념의 등장과 함께 관련 표준과 기술 등이 활발하게 개발되기 시작하면서 이러한 기술들을 전자저널의 출판과 유통에 적용하고자 하는 연구가 활발해짐.

□ 이에 따라 시맨틱 퍼블리싱(Semantic Publishing)이라는 개념이 정립되고 계속적인 연구가 진행되고 있는 상태임. 연구에 따르면 시맨틱 퍼블리싱은 다음과 같이 정의하고 있음.

- 시맨틱 퍼블리싱은 저널간의 의미적인 상호연결을 가능하게 하고 저널 내의 데이터를 접근할 수 있게 하거나, 출간된 저널의 의미를 풍부하게 하기 위한 일련의 노력으로 정의함.
- 또한 적절한 메타데이터를 활용하여 출간된 정보의 검증가능성을 향상시키고 자동화된 발견과 요약을 가능하도록 하는 등의 자동화된 프로세싱이나 분석을 수월하게 해주는 것을 의미하기도 함.

□ 위키피디아에 따르면 시맨틱 퍼블리싱은 시맨틱 마크업이 포함된 문서 형태로 웹상에 정보를 발행하는 것을 의미하며, 이를 통해 컴퓨터(또는 에이전트)가 발행된 정보의 구조와 의미를 이해할

수 있는 방법을 제공하고, 발행된 정보가 보다 효율적으로 검색 및 통합될 수 있도록 하는 것을 뜻함.

□ 시맨틱 퍼블리싱과 관련하여 동적 시맨틱 퍼블리싱(Dynamic Semantic Publishing)이라는 개념이 소개된 바 있음. 동적 시맨틱 퍼블리싱이란 BBC에 의해 소개된 개념으로 자체적으로 구축한 퍼블리싱 프레임워크를 지칭함.

- BBC는 생성한 정보 콘텐츠에 대하여 분야별로 특화된 온톨로지를 활용하여 의미적으로 연관된 주석을 삽입하기 위해 자연어 처리기술을 사용함.
- 이렇게 상호 연계된 콘텐츠의 수집, 퍼블리싱, re-purposing을 자동화하기 위해 링크드 데이터 기술을 활용함으로써 BBC는 서비스에 이용자들에게 향상된 이용자 경험과 높은 수준의 이용자 참여를 제공함. 또한, 이를 통해 저널리스트들을 높은 수준의 콘텐츠 생성에 집중하면 되는 환경이 구축되었음.

2. 연구동향

□ 시맨틱 퍼블리싱은 정의된 바와 같이 시맨틱 기술을 활용하여 웹에 발행된 자원에 대한 '접근성의 향상', 그리고 에이전트나 검색엔진 등에 의한 자원의 '자동화 처리'를 핵심으로 하고 있음.

□ 시맨틱 퍼블리싱에 관한 연구는 크게 아래의 표와 같은 범주로 구분될 수 있음.

<표 1> 시맨틱 퍼블리싱 연구동향

구분	내용
시맨틱 퍼블리싱을 위한 데이터모델 및 온톨로지 개발 연구	체계화된 시맨틱 퍼블리싱을 지원하는 데이터모델 정의 및 온톨로지를 설계. Semantic enhancement 활용을 위한 표준 기반의 데이터 변환
시맨틱 퍼블리싱의 적용 및 효과 연구	저널 아티클, 논문을 대상으로 다양한 시맨틱 퍼블리싱 관련 기술과 기법을 실제로 적용해 보고 그 효과를 입증
플랫폼 개발 연구	저널 아티클, 논문에 대한 관리와 효율적인 시맨틱 퍼블리싱을 지원하는 플랫폼을 개발
시각화 연구	연구자와의 상호작용을 지원하는 데이터 시각화 기술

□ 시맨틱 퍼블리싱은 온라인 뉴스 및 미디어 분야에서 연구 및 적용이 이루어진 바 있으며, 대표적인 사례는 BBC임. BBC는 발행되는 뉴스기사에 대해 시맨틱 퍼블리싱 기술을 적용하여 기사의 품질과 연결성 등을 향상시키는데 가시적인 성과를 거둠.

• LOD(Linked Open Data) 광범위한 확산과 함께 최근 시맨틱 퍼블리

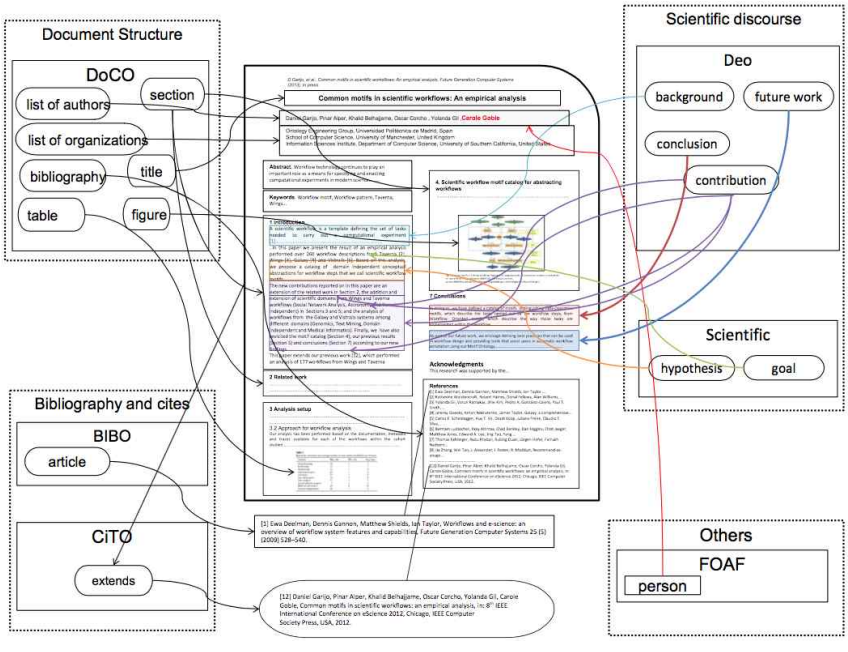
싱에 LOD를 활용하는 연구가 매우 활발하게 이루어지는 추세임

- 연구논문 및 아티클의 본문에 대한 Semantic enhancement를 통해 LOD 개체와 연결하거나 아티클로부터 LOD를 생성하는 것, 학술분야에 특화된 LOD를 구축하는 것 등 다양한 연구가 시도되고 있음
- Semantic enhancement를 위한 요소의 추출, 정제 그리고 연계 등의 작업을 효율적으로 수행할 수 있도록 하는 시맨틱 퍼블리싱 플랫폼(Semantic Publishing Platform)에 관한 연구도 활발하게 이루어지고 있는 추세임.

3. 주요 연구내용

(1) 데이터 모델 및 온톨로지 개발 연구

- 시맨틱 웹 기술의 핵심은 온톨로지임. 철학적으로 온톨로지란 존재와 존재자의 본성을 연구하는 형이상학의 한 부분으로서 세상의 구성 요소에 대한 명확한 이해를 얻고자 하는 분야임. 정보 자원의 관리적 측면에서 온톨로지를 사람의 마음 속에 존재하는 내재적 생각이나 외재적 세계의 현상과 대상에 대하여 공유하는 개념을 컴퓨터가 이해할 수 있는 형식으로 명확하고 명시적으로 정의하고 규정하는 것임(고영만, 2006)
- RDF(Resource Description Framework)와 OWL(Web Ontology Language) 등 표준에 기반한 온톨로지는 시맨틱 퍼블리싱을 위한 기반 또는 체계를 제공함. 즉, 온톨로지를 기반으로 기술된 학술저작물과 그에 대한 서지 정보는 웹 환경에서의 자동화된 처리와 정보의 재사용성 보장과 같은 장점이 있음
- 학술 저작물에 최적화된 온톨로지 및 데이터 모델을 개발하기 위한 연구가 진행되어 왔으며, 이에 대한 성과로 다양한 종류의 온톨로지가 개발되어 웹상에 제공되고 있는 상태임
- 학술저작물의 구조, 서지 메타데이터 정보, 인용 정보 등의 기술을 위한 온톨로지가 있으며, 이러한 온톨로지 요소들의 적절한 조합 및 활용을 통해 하나의 학술저작물에 대한 상세한 기술이 가능함



출처(Ruiz-Iniesta, A., & Corcho, O., 2014)

[그림 1] 온톨로지를 활용한 학술 저작물의 기술

- 상기 그림은 아티클의 구조, 서지메타정보, 인용정보 등에 대해 다양한 온톨로지를 조합하여 기술하는 예시로 시맨틱 퍼블리싱에 활용할 수 있는 온톨로지가 이미 다수 개발되어 있는 상태임

(2) 시맨틱 퍼블리싱 적용 및 효과 연구

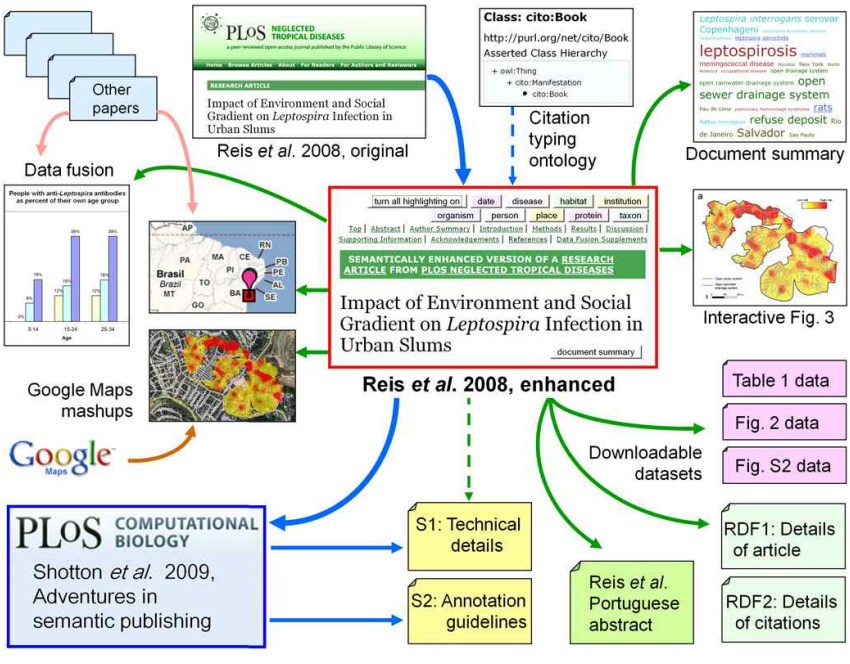
- 다양한 기법 또는 시맨틱 기술을 활용한 시맨틱 퍼블리싱이 가능함. XML, RDF, OWL 등과 같은 기술(Description) 언어를 활용하여 아티클 자체 또는 아티클에 대한 메타데이터를 기술하여 발행하는 것이 대표적임
- 특정분야의 아티클 일부를 선정하여 다양한 시맨틱 퍼블리싱 기법

을 적용한 후 그 효과를 검증하는 것이 대표적인 연구임. 이와 관련하여 Shotten(2009)는 PLoS Neglected Tropical Diseases 저널에 발행된 research article을 대상으로 시도한 semantic enhancement의 사례를 소개하고 있으며, 아래의 표와 같이 총 11개 항목의 시맨틱 퍼블리싱 방법을 제시한 바 있음

<표 2> 주요 semantic enhancement 기법

주요 Semantic Enhancement 기법	내용
DOI, 하이퍼링크 제공	연관정보 및 레퍼런스와 관련하여 클릭하여 확인가능한 DOI 링크 등의 하이퍼링크 제공
본문 텍스트에 대한 시맨틱 마크업 및 외부링크 제공	유형별 키워드 하이라이트를 통해 연구자들이 연구의 전체적인 맥락 빠른 시간 내에 파악할 수 있도록 함 주요 개념, 인물들에 대해 부가적인 정보를 제공하는 외부자원으로의 링크를 제공하여 풍부한 정보를 확보할 수 있도록 함
상호작용 가능한 그림 제공	연구자들이 삽입된 그림에 대한 데이터 또는 조건을 변경할 수 있도록 하여 변경에 대한 그래프 또는 그림의 변화를 볼 수 있도록 함
재배열 가능한 참고문헌 리스트 제공	연구자들이 선호하는 방식, 예를 들어 알파벳순, 저자순, 제목순, 발행년도순 등으로 참고문헌 목록을 배열할 수 있도록 함
아티클 요약정보 제공	연구에 대한 핵심요약, 태크 클라우드, 참고문헌 분석(인용횟수 등이 포함된 스프레드시트 등) 등이 포함된 아티클의 요약정보 제공
참고문헌 내용 툴팁 제공	참고문헌의 인용내용을 마우스 오버 시 툴팁 형태로 표시되도록 제공
태그 트리 제공	의미적으로 연결된 단어들을 태그 트리 형태로 시각화하여 제공
테이블, 그림에 대해 스프레드쉬트와 같은 형태로 다운로드 가능하도록 발행	연구자들이 데이터를 재검증할 수 있고 재활용하여 다른 연구에 활용할 수 있도록 데이터의 신뢰성과 재활용성을 증가
리서치 아티클과 구글맵 등을 활용하여 다양한 유형의 data fusion을 시도하여 발행	연구데이터 또는 Raw 데이터를 구글맵과 같은 공간,지각 플랫폼 등을 연계하여 보여줌으로써 데이터의 가독성과 가치를 증가
기계가독형 RDF 메타데이터 제공	검색엔진 또는 크롤러 등이 읽을 수 있도록 아티클과 참고문헌에 대한 RDF 메타데이터 제공
온톨로지 기반 참고문헌 기술(Description) 제공	CiTO(Citation Typing Ontology) 등의 온톨로지를 기반으로 참고문헌을 기술하여 제공

아래의 그림은 다양한 semantic enhancement 기법의 적용을 도식화한 것임. 주로 PDF 형태로 제공되고 있는 아티클과 비교하여 풍부한 시각화 자료, 연계정보 및 Raw 데이터 등을 제공하여 연구자의 이해를 높이고 재사용성을 크게 높일 수 있음



출처: (Shotten, 2009)

[그림 2] Semantic enhancement의 사례

(3) 플랫폼 개발 연구

- 시맨틱 퍼블리싱의 과정은 단순하지 않으며, 일반적인 정보 콘텐츠와는 달리 학술정보의 특성상 더 높은 수준의 신뢰성이 요구됨. 온톨로지 개발, Semantic enhancement 적용 등 시맨틱 퍼블리싱의 전체 과정이 정확하고 효율적으로 진행될 수 있도록 프로세스를 정립

하는 것이 요구됨.

- 시맨틱 퍼블리싱과 관련된 다양한 도구의 개발과 함께 시맨틱 퍼블리싱의 주요 절차 또는 전체과정을 효과적으로 관리할 수 있도록 하는 플랫폼의 개발도 활발하게 이루어지고 있음
- 오픈소스 라이선스를 따르는 플랫폼을 비롯하여 높은 품질의 서비스를 제공하는 상업적 플랫폼도 등장하고 있음. 최근 클라우드 컴퓨팅 기술의 발전에 따라 SaaS 방식인 플랫폼도 제공되고 있음
- 시맨틱 플랫폼의 특징은 RDF, OWL로 기술된 온톨로지의 지원, LD, JSON, JSON-LD와 같은 주요 데이터 포맷의 입출력 지원, 텍스트 분석/추출/분류 및 RDF 형식으로 기술된 콘텐츠 및 메타데이터 발행 기능 등을 특징으로 함
- ontotext, Ambra, Open Semantic Framework, redlink, Ximdex, meaning cloud 등이 주요 시맨틱 퍼블리싱 플랫폼의 범주에 포함될 수 있음

(4) 시각화 연구

- 연구데이터의 시각화에 관한 연구도 활발하게 이루어지고 있는 추세이며, 시맨틱 퍼블리싱과 관련한 시각화 연구는 1)웹표준 기술을 활용한 시각화 구현, 2)연구데이터의 재현과 재활용을 위한 시각화 방안 및 매시업과 3)interative 요소 적용을 통한 전달효과 극대화 등 다양한 방식으로 진행되고 있음
- HTML5의 광범위한 확산과 이에 기반한 다양한 시각화 기술이 개발되어 기존의 정적인 표나 이미지가 아닌 실시간으로 변경과 상호작용이 가능한 형태로 시각화 구현이 이루어지는 추세임
- 시각화 마크업의 경우 KML(Keyhole markup language), GXL(Graph eXchange Language), GML(Graph Modelling Language), DOT(graph

description language) 등이 있음

- 연구데이터 시각화와 관련하여 일반적으로 활용되는 툴(예를 들어 iPython notebook, Matlab, R Studio)에서의 시각화 지원 관련 연구도 활발하게 이루어지는 추세임

4. 시맨틱 퍼블리싱 관련 온톨로지 및 데이터 모델

- 학술분야 시맨틱 퍼블리싱을 위한 다양한 온톨로지가 개발되어 활용중임. 시맨틱 퍼블리싱 관련 온톨로지에 관한 연구 중 2014년에 출간된 Ruiz-Iniesta, A.와 Corcho, O.의 연구는 학술저널 및 아티클에 대한 정보를 기술하는 온톨로지에 대해 포괄적으로 기술하고 있음. 총 3개의 범주로 온톨로지의 범주를 구분하여 제시하고 있으며, 그 범주는 아래와 같음
- [Ruiz-Iniesta, A., & Corcho, O. (2014). A review of ontologies for describing scholarly and scientific documents. In Proceedings of 4th Workshop on Semantic Publishing (SePublica 2014), CEUR Workshop Proceedings. Aachen, Germany: CEURWS. org. OA at http://ceur-ws. org (Vol. 1155).]

<표 3> 학술정보 관련 온톨로지 구분

구분	문서 구조 기술용 온톨로지	수사학적 요소 기술용 온톨로지	서지사항 및 인용 기술용 온톨로지
기술내용	section, paragraph 등 정보	introduction, result 등 정보	서지메타, 참고문헌 등 정보
주요 온톨로지	Document Ontology, DoCO	Ontology of Rhetorical Blocks	FaBiO, CiTO, BiRO, C4O, BIBO

- 시맨틱 퍼블리싱과 관련하여 가장 대표적인 온톨로지는 SPAR 온톨로지로 이를 구성하는 8개의 온톨로지들은 다양한 온톨로지에 결합되어 활용되고 있음

(1) DO(Document Ontology)

- http://www.cs.umd.edu/projects/plus/SHOE/onts/docmnt1.0.html
- 문서의 유형과 타입 및 구조를 기술하기 위한 온톨로지로 일반적인 모든 문서에 대한 기술을 지원하기 위해 개발됨
- 시맨틱 웹 관련연구 초기(2000년)에 정의된 온톨로지로 많은 요소들이 Dublin Core에 기반하고 있음
- Base 온톨로지와 General 온톨로지를 확장함
- Document 온톨로지의 텍사노미 구조는 아래의 그림과 같음

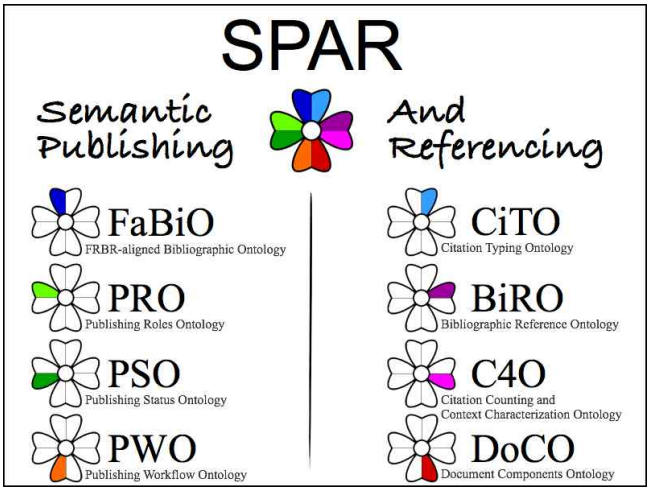
```
[gen.Event]
   Communication
      Document
         Abstract
         Comment
         Correspondence
            Discussion
            Email
            Letter
            Postcard
         Form
         Guideline
         Homepage
            OrganizationHomepage
            PersonalHomepage
         Index
         Lecture
         Manuscript
         Minutes
         Preprint
         Promotion
         Publication
            Advertisement
            Article
               BookArticle
               ConferencePaper
               JournalArticle
               WorkshopPaper
            Book
            Dictionary
            Editorial
            Manual
            Periodical
               Journal
               Magazine
               Newsletter
               Newspaper
            Proceedings
            Regulation
            Specification
            TechnicalReport
            Thesis
               DoctoralThesis
               MastersThesis
         Review
      PhoneCall
      Software
      Speech
[gen.Artifact]
   DocumentRepresentation
      PaperDocument
      ElectronicDocument
```

출처: http://www.cs.umd.edu/projects/plus/SHOE/onts/docmnt1.0.html

[그림 3] Document Ontology 계층 구조

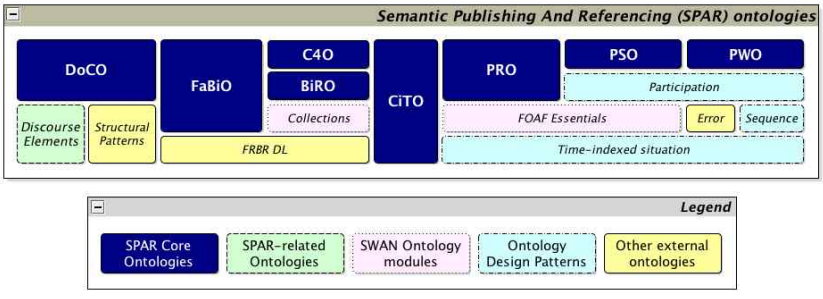
(2) SPAR(Semantic Publishing and Referencing Ontologies)

- SPAR는 시맨틱 퍼블리싱과 관련한 모든 측면의 기계가독형 RDF 메타데이터 기술을 지원하는 상호보완적 온톨로지의 집합(Suite)
- 총 8개의 온톨로지로 구성되어 있으며, 개별 온톨로지 또는 여러 온톨로지를 조합하여 메타데이터를 기술할 수 있음
- 각각의 온톨로지는 아래 그림과 같이 꽃모양의 다이어그램으로 상징화되어 있음
- 각 온톨로지는 OWL 2.0 DL로 정의되어 있음



[그림 4] SPAR 구성 온톨로지와 상징

- SPAR 온톨로지를 구성하는 8개의 온톨로지는 OWL2 DL로 기술된 3개의 보조 온톨로지인 Essential FRBR in OWL2 DL 온톨로지, DEO(the Discourse Elements Ontology) 온톨로지 그리고 Error 온톨로지를 활용하여 설계되어 있으며, 아래의 그림은 이러한 아키텍처를 도식화 한 것임



[그림 5] SPAR 온톨로지 아키텍처

- SPAR 온톨로지를 구성하는 각 온톨로지의 용도와 특징은 아래의 표와 같음

<표 4> SPAR 구성 온톨로지의 용도와 특징

온톨로지명	내용
FaBio (FRBR-aligned Bibliographic Ontology)	학술 저작물들에 대한 서지 레코드를 기록하고 발행하기 위한 온톨로지. 기술 대상이 되는 개체는 주로 단행본, 연속간행물, 저널 등과 같은 저작물이지만 데이터셋, 컴퓨터 알고리즘, 실험 프로토콜, 명세서, 어휘, 법률 레코드, 정부문서, 기술리포트, 비즈니스 리포트 등 기술대상 정보의 범주는 포괄적임. FRBR 데이터 모델(Work, Expression, Manifestation, Item)에 따라 구조화되어 있으며, 데이터 모델 확장을 위해 일부 프로퍼티가 추가됨.
CiTO (Citation Typing Ontology)	인용의 출처와 타입 등의 특징을 기술하기 위한 온톨로지로 직접(direct) 인용과 간접(indirect) 인용된 정보 등을 모두 기술함. 인용정보는 오브젝트 타입 프로퍼티인 cito:cites와 서브 프로퍼티로 기술하며, 1.6 버전에서 2.0 버전으로 업데이트 되면서 새로운 서브 프로퍼티가 추가되어 보다 정교한 기술이 가능해짐.
PRO (Publishing Roles Ontology)	출판 과정에서 저작자, 편집자, 리뷰어, 출판사, 사서, 회사, 법인 등과 같은 에이전트들의 역할과 특성을 기술하기 위한 온톨로지. 기술하고자 하는 서지적 개체와 관련하여 각 에이전트들의 역할 기술에 초점을 둠.
BiRO (Bibliographic Reference Ontology)	FaBio와 마찬가지로 FRBR 데이터 모델에 따라 구조화된 온톨로지로 서지 레코드(bibliographic records)와 서지 레퍼런스(bibliographic reference)를 정의하고 정의된 항목들을 묶어 컬랙션 또는 리스트 형태로 기술할 수 있도록 하는 FRBR 모델 기반의 온톨로지. 관련성을 가진 개별 서지 항목들을 묶을 수 있는 논리적 체계를 제공함.

온톨로지명	내용
PSO (Publishing Status Ontology)	출판상태, 이를테면 draft, submitted, under review, rejected, accepted for publication, proof, published, version of record, catalogued, archived 등과 같이 출판 상태를 기술하기 위한 온톨로지
C4O (Citation Counting and Context Characterization Ontology)	인용된 참고문헌에 대한 아티클 내에서의 인용횟수, 주요 서지정보 제공처인 Google scholar, Scopus, Web of Knowledge에서의 인용횟수를 기술하기 위한 온톨로지
PWO (Publishing Workflow Ontology)	문서의 출판과 관련한 워크플로우(작성됨, 리뷰중, XML 캡쳐, 페이지 디자인, 웹에 발행) 상의 단계를 기술하기 위한 온톨로지
DoCO (Document Components Ontology)	구조적 또는 수사학적 차원에서 문서의 구조를 기술하기 위한 온톨로지.Discourse Elements 온톨로지, Document Structural Patterns 온톨로지, SALT Rhetorical 온톨로지 등이 임포트(import)되어 있음

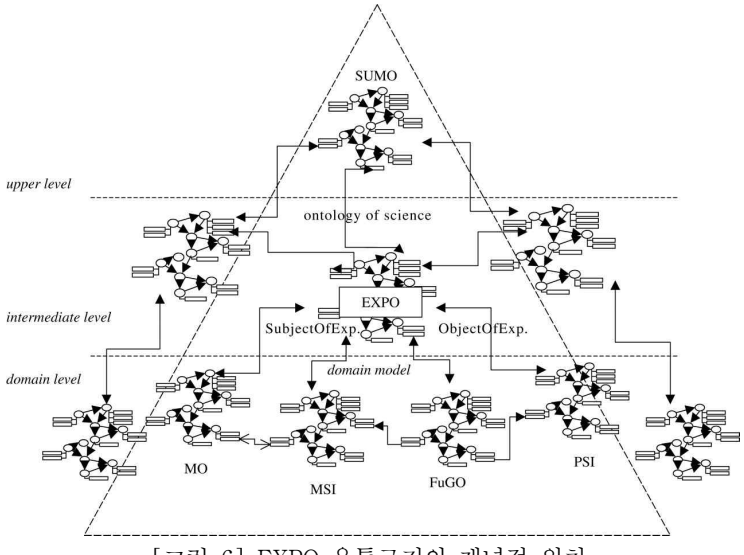
(3) ORB(Ontology of Rhetorical Blocks)

- 학술 저작물의 수사학적 구조를 기술하기 위한 온톨로지로 W3C HCLS(Semantic Web in Health Care and Life Sciences Interest Group)그룹에 속한 Scientific Discourse Task Force에 의해 설계됨
- 분야에 관계없이 과학분야 출판물들의 수사학적 구조를 형식화함 header, body, tail 세 부분으로 대상자원을 구조화

(4) EXPO Ontology

- 과학적 실험의 완벽한 기술을 목표로 구축된 온톨로지로 시맨틱 마크업을 생성하기 위한 218개의 개념을 정의하고 있음
- 과학적 실험의 디자인, 방법론, 실험결과 표현에 대한 일반적인 지식을 형식화하여 표현함
- OWL 형식으로 기술됨
- 대표적인 Upper 온톨로지인 SUMO 온톨로지를 기반으로 설계된 온톨

로지로 특정 도메인에 특화된 형태로 손쉽게 확장이 가능하며, 아래의 그림은 EXPO 온톨로지의 개념적 위치와 관계를 도식화 한 것임



[그림 6] EXPO 온톨로지의 개념적 위치

(5) SWAN(Semantic Web Application in Neuromedicine) 온톨로지

- 과학적 담론을 모델링하기 위한 온톨로지
- SWAN 프로젝트의 요구사항을 충족하기 위해 만들어진 온톨로지임. SWAN 프로젝트란 bio-medicine 분야의 과학적 담론에 대해 현실적이고 일반적이며 의미적으로 구조화된 프레임워크를 개발하기 위한 목적으로 진행된 프로젝트
- SWAN 온톨로지의 개발 목적은 비구조화된(unstructured) 또는 반구조화된(semi-structured) 디지털 과학정보에 따라 과학적 담론의 의미적 맥락이 생성, 저장, 접근 통합 및 교환될 수 있는 사회 기술적 생태계가 기능하도록 하는 것임

- 과학적 정보와 의견을 대해 풍부한 맥락을 체계적으로 정리할 수 있는 구조적 토대를 제공하고 OWL로 기술되어 있음
- SWAN 온톨로지는 3가지 타입의 모듈로 구성되어 있음 (basic, extension, distribution)

(6) BIBO(Bibliographic Ontology Specification)

- BIBO는 인용정보와 서지적 레퍼런스를 기술하기 위한 핵심 개념과 프로퍼티를 제공함
- 문서의 출처(저널, 단행본, 웹소스 등)를 바탕으로 문서의 형식을 식별하기 위한 클래스 개체를 정의하며, RDF에 따라 기술됨. 문서의 출처에 따라 분류할 때 유용하게 활용 가능한 온톨로지

```
<http://ex.net/x1> a bibo:Article ;
    dc:creator <http://examples.net/contributors/1> ;
    dc:title "Conference Paper" ;
    bibo:presentedAt <http://ex.net/conference> ;
    bibo:reproducedIn <http://ex.net/x2> .


<http://ex.net/conference> a bibo:Conference ;
    bibo:organizer <http://ex.net> ;
    event:place <http://dbpedia.org/resource/> ;
    event:time [ time:Interval [ timeline:beginsAtDateTime "2007-03-01"
; timeline:endsAtDateTime "2007-03-13"] ] .

<http://ex.net/x2> a bibo:Proceedings ;
    dc:title "Proceedings of the ABC Conference"@en ;
    dc:date "2000" ;
    dc:publisher <http://ex.net/z1> .
```

[그림 7] BIBO 온톨로지를 활용하여 기술된 프로시딩 정보의 예시

(7) 더블린 코어(Dublin Core)

- DCT(Dublin Core Metadata Term)는 전자자원에 대한 메타정보 기술을 위한 15개의 프로퍼티로 구성된 어휘

- HTML, XHTML, RDF/XML, Plain XML 등으로 기술됨
- 참고문헌이나 아티클의 내용 자체에 대해 기술하기에는 한계가 있음

```
@prefix dc: <http://purl.org/dc/elements/1.1/>
@prefix dcterms: <http://purl.org/dc/terms/>
@prefix dctype: <http://purl.org/dc/dcmitype/>

<http....> # url of the journal article
 dc:title "Common␣motifs␣in␣scientific␣workflows..." ;
 dcterms:dateSubmitted "2013-02-01"^^dcterms:W3CDTF  ;
 dcterms:dateAccepted "2013-09-05"^^dcterms:W3CDTF ;
 dcterms:format mime:application/pdf ;
 dcterms:license      <http://creativecommons.org/licenses/
    by/3.0/> ;
```

[그림 8] 더블린 코어 기술 예시

(8) FOAF(Friend of a Friend)

- 2000년 중반에 생성된 이후 꾸준히 업데이트 되고 있는 인물정보 기술을 위한 어휘 또는 온톨로지
- RDF 스키마로 기술됨
- 저작물에 대한 저자 정보를 기술. 저널 아티클의 저자정보 기술에 활용

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<#DG>
    a foaf:Person ;
    foaf:name "Daniel␣Garijo" ;
    foaf:mbox <mailto:dgarijo@fi.upm.es> ;
    foaf:knows [
        a foaf:Person ;
        foaf:name "Oscar␣Corcho"
    ] .
```

[그림 9] FOAF 기술 예시

(9) DOAF(Description of a Project)

- 소프트웨어 프로젝트 정보를 기술하기 위한 어휘
- RDF 스키마 또는 XML로 기술됨

(10) SKOS(Simple Knowledge Organization System)

- SKOS는 시소러스, 분류체계, 텍사노미, 색인 목록을 비롯하여 다양한 유형의 구조화된 어휘를 표현할 수 있도록 디자인된 W3C 권고사함으로 링크드 데이터와 관련하여 다양한 어휘들을 손쉽게 발행하고 활용할 수 있도록 하기 위함임
- 일반적인 시소러스의 구성 요소들을 표현하기에 적합한 클래스와 프로퍼티를 정의하고 있으며, SKOS를 구성하는 주요 요소들은 아래의 표와 같음

<표 5> SKOS 어휘

Concepts	Labels & Notation	Documentation	Semantic Relations	Mapping Properties	Collections
Concept	prefLabel	note	broader	broadMatch	Collection
Concept Scheme	altLabel	changeNote	narrower	narrowMatch	ordered Collection
inScheme	hiddenLabel	definition	related	relatedMatch	member
hasTop Concept	notation	editorialNote	broader Transitive	closeMatch	memberList
topConceptOf		example	narrower Transitive	exactMatch	
		historyNote	semantic Relation	mapping Relation	
		scopeNote			

(11) W3C Open Annotation Model

- 시맨틱 퍼블리싱과 관련하여 가장 큰 문제는 발행 주체와 플랫폼에 대한 종속의 문제임. 즉, 시맨틱 퍼블리싱을 통해 의미적으로 풍부

해진 아티클과 아티클에 대한 메타데이터가 발행주체와 서비스 플랫폼 간에 공유되지 못한다는 것임

- 이러한 문제를 해결하기 위해 W3C는 Web Annotation Working Group을 구성함. 웹으로 접근 가능하고 공유가능한 정보간의 관계 정보를 생성할 수 있는 구조화된 접근방법을 정의하는 것을 목적으로 함. 이를 통해 annotation ecosystem의 구성이 가능해지며, 다양한 발행 주체의 아티클에 대한 annotation 정보를 공유하고 재활용하는 것이 가능해짐.
- Open Annotation Data Model은 월드와이드웹의 아키텍처를 따르는 방법론을 활용하여 관련된 정보자원과 annotation 간의 연계정보 생성을 위한 상호운용성 있는 프레임워크를 기술함.
- Open Annotation Data Model은 OWL 온톨로지로 RDF/XML, Turtle, JSON_LD 및 XHTML+RDFa 등의 형태로 기술됨
- 플랫폼간의 손쉽게 공유될 수 있다는 것을 특징으로 하며, 복잡한 요구사항을 충족할 수 있을 만큼 풍부한 표현력을 제공함과 동시에 일반적인 용도에서도 활용하기에 충분할 만큼 단순함
- 아래의 표와 같은 요소들로 구성됨

<표 6> Open Annotation의 구성 요소

구분	구성요소
Classes	Annotation Choice Composite CssStyle DataPositionSelector FragmentSelector HttpRequestState List Motivation Selector SemanticTag SpecificResource State Style SvgSelector Tag TextPositionSelector TextQuoteSelector TimeState
Object Properties	annotatedBy cachedSource default equivalentTo hasBody hasScope hasSelector hasSource hasState hasTarget item motivatedBy serializedBy styledBy
Data Properties	annotatedAt styleClass end exact prefix serializedAt start suffix when
Named Individuals	motivationScheme oa:bookmarking oa:classifying oa:commenting oa:describing oa:editing oa:highlighting oa:identifying oa:linking oa:moderating oa:questioning oa:replying oa:tagging

5. 시맨틱 퍼블리싱 주요 기법 및 적용

- 온톨로지를 기반으로 한 시맨틱 퍼블리싱이 대표적인 연구 경향이나 앞서 살펴본 바와 같이 다양한 기법들을 활용한 시맨틱 퍼블리싱이 가능함

(1) RDFa 생성 및 삽입

- RDFa는 W3C 권고사항으로 HTML 페이지에 기계가 읽을 수 있는 형태의 구조화된 메타데이터를 삽입하여 이용자 또는 검색엔진이 콘텐츠를 효과적 처리할 수 있도록 함. 검색엔진은 HTML 사이에 삽입되어 기술된 RDFa 메타데이터를 처리하여 풍부한 검색결과를 이용자에게 제공 가능함
- RDFa 1.0의 경우 XHTML을 위해서만 기술되었으나 RDFa 1.1은 XHTML, HTML5 및 XML 기반의 문서들도 고려하여 기술되어 있음
- 아래의 그림은 XHTML과 문서에 RDFa 1.0 형식에 따라 기술된 메타데이터의 예시로 Dublin Core와 FOAF를 활용하여 인물과 인물간의 관계정보를 기술하고 있음

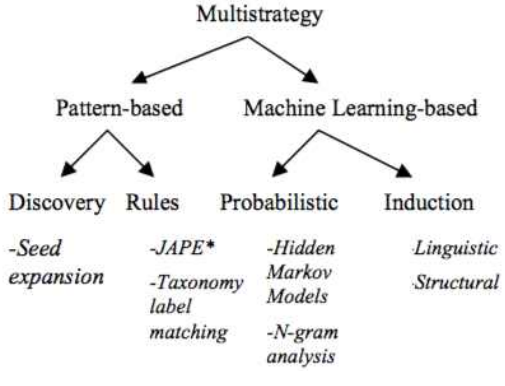
```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
    "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    version="XHTML+RDFa 1.0" xml:lang="en">
  <head>
    <title>John's Home Page</title>
    <base href="http://example.org/john-d/" />
    <meta property="dc:creator" content="Jonathan Doe" />
    <link rel="foaf:primaryTopic" href="http://example.org/john-d/#me" />
  </head>
  <body about="http://example.org/john-d/#me">
    <h1>John's Home Page</h1>
    <p>My name is <span property="foaf:nick">John D</span> and I like
      <a href="http://www.neubauten.org/" rel="foaf:interest"
        xml:lang="de">Einstürzende Neubauten</a>.
    </p>
    <p>
      My <span rel="foaf:interest" resource="urn:ISBN:0752820907">favorite
      book is the inspiring <span about="urn:ISBN:0752820907"><cite
      property="dc:title">Weaving the Web</cite> by
      <span property="dc:creator">Tim Berners-Lee</span></span></span>.
    </p>
  </body>
</html>
```

[그림 10] Dublin Core와 FOAF를 활용한 RDFa 1.0 기술 예시

- RDFa 생성툴 및 다양한 에디터를 활용하여 RDFa 문서를 생성 가능하며, 언어별로 구현된 라이브러리를 활용한 개발을 통해 자동으로 생성되어 HTML에 삽입되도록 구현도 가능함
- XHTML, XML 형식으로 기술된 학술정보와 아티클에 RDFa 메타정보를 삽입하여 검색엔진에 의한 처리와 노출 효과를 극대화 할 수 있음
- 최근 콘텐츠 관리 시스템(CMS)들은 자동으로 RDFa가 생성되어 삽입되도록 지원하는 추세임. 예를 들어 Wordpress, Drupal과 같은 CMS들은 모듈 설치와 활성화를 통해 RDFa 생성을 지원함

(2) 시맨틱 어노테이션(Semantic Annotation)

- 시맨틱 어노테이션은 정보의 검색과 상호운용성 향상을 위해 문서의 내용에서 개념과 개념 사이의 관계정보를 식별하여 부가적으로 의미있는 정보를 제공하기 위한 기법임. 즉, 온톨로지를 기반으로 한 정보 개체와 같이 구조적으로 정의된 정보를 활용하여 웹문서, PDF 등의 내용에 기술되어 있는 개념을 식별하고 맵핑하여 이용자나 풍부한 정보를 제공받을 수 있도록 하고 검색엔진의 색인을 향상시킴
- 2000년대 중반부터 다양한 시맨틱 어노테이션 기법들을 적용한 툴, 시스템 또는 플랫폼 등에 관한 연구가 활발하게 진행었으며, 주요 어노테이션 기법은 아래의 그림과 같이 pattern-based 와 machine learning-based 방식으로 구분됨



출처: Reeve, L., & Han, H. (2005, March). Survey of semantic annotation platforms. In Proceedings of the 2005 ACM symposium on Applied computing (pp. 1634-1638). ACM.

[그림 11] Semantic Annotation 기법의 분류

- 위와 같은 어노테이션 기법을 활용하여 구현된 주요 툴 및 플랫폼은 아래의 표와 같음

<표 7> 주요 Semantic Annotation 툴 및 플랫폼

Tool	Dynamicity of Annotation	Annotation Storage	Level of Automation	Method Type	Scalable/Customizable
Aero DAML	Dynamic	RDF knowledge base	Automatic	Rule based	Scalable with diverse ontologies.
AKTive Media	Dynamic/ Static	Annotation Server	Automatic/ Manual	Wrapper induction	Difficult to scale despite having diverse techniques available to reduce effort of annotating.
Armadillo	Dynamic	RDF knowledge store	Automatic	Pattern Matching	Needs seeds to scale. Uses a cyclical annotating process with different techniques.
Cerno	Dynamic	External database	(Semi-) Automatic	Wrapper induction	Can adapt to new domains.
CREAM	Dynamic/ Static	Embedded in webpage	Automatic/ Manual	Rule based/ wrappers	Difficult to scale but well suited for highly structured web documents.
Drupal	Dynamic	Embedded in webpage	Automatic	Rule based	Supports diverse ontologies and adapts easily to new domains.
GoNTogle	Dynamic/ Static	Ontology Server	Automatic/ Manual	Wrapper induction	Automatic annotation requires learning method of past annotations. Supports diverse document formats.
KIM	Dynamic	RDF knowledge base	Automatic	Rule based	Ontology pre-populated with large number of instances.
KnowItAll	Dynamic	RDF knowledge base	Automatic	Rule based	Automatic generated domain specific extraction rules.
KnowWe	Dynamic	RDF knowledge base	Manual/ (Semi-) Automatic	Rule based	Annotations are made via a special markup. Each article corresponds to a separate knowledge base containing all the annotated relations and knowledge markups
Lixto	Dynamic/ Static	Real time	(Semi-) Automatic	Wrapper induction	It scans Web pages continuously and extracts relevant information.

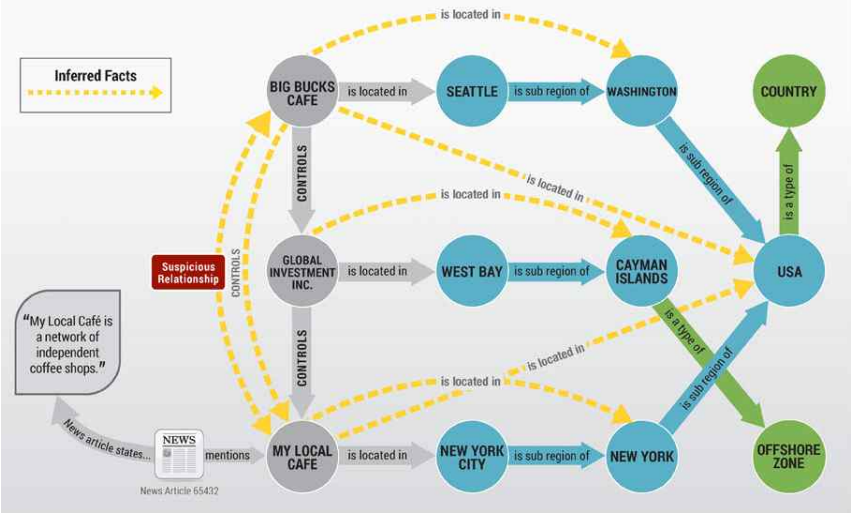
Tool	Dynamicity of Annotation	Annotation Storage	Level of Automation	Method Type	Scalable/Customizable
Magpie	Dynamic/ Static	Real time	Automatic	Wrapper induction	Works within web browser. User can choose the ontology
MnM	Dynamic/ Static	Embedded in webpage	(Semi-) Automatic	Wrapper induction	Learns extraction rules from training corpus. Offers API for connection to ontology servers.
Open Calais	Static	Embedded in webpage	Automatic	Rule based	Uses NLP and machine-learning techniques. Free but with daily limit request.
PANKOW	Dynamic	RDF knowledge base	Automatic	Pattern Matching	Based on counting Google hits of instantiated linguistic patterns.
Semantic Miedia Wiki	Dynamic/ Static	Embedded in webpage	Manual/ (Semi-) Automatic	Rule based	Annotations are made via a special markup. Ontologies can be imported. Every article corresponds to exactly one ontological element.
Semantic Wikipedia	Dynamic/ Static	Embedded in webpage	Automatic	Pattern Matching	Domain must have a structured description.
SemTag	Dynamic	RDF knowledge base	Automatic	Rule based	Performs structural analysis. Can access database records and metadata.
Thresher	Dynamic/ Static	Real time	Automatic/ Manual	Wrapper induction	Aimed to Web pages with similar content.
Zemanta	Dynamic/ Static	Real time	(Semi-) Automatic	Rule based	Uses Machine-learning techniques. Targeted for unstructured documents. Algorithm for natural language and semantic processing is proprietary.

출처: Oliveira, P., & Rocha, J. (2013, April). Semantic annotation tools survey. In Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on (pp. 301-307). IEEE.

6. 시맨틱 퍼블리싱 플랫폼

(1) Ontotext Dynamic Semantic Publishing Platform

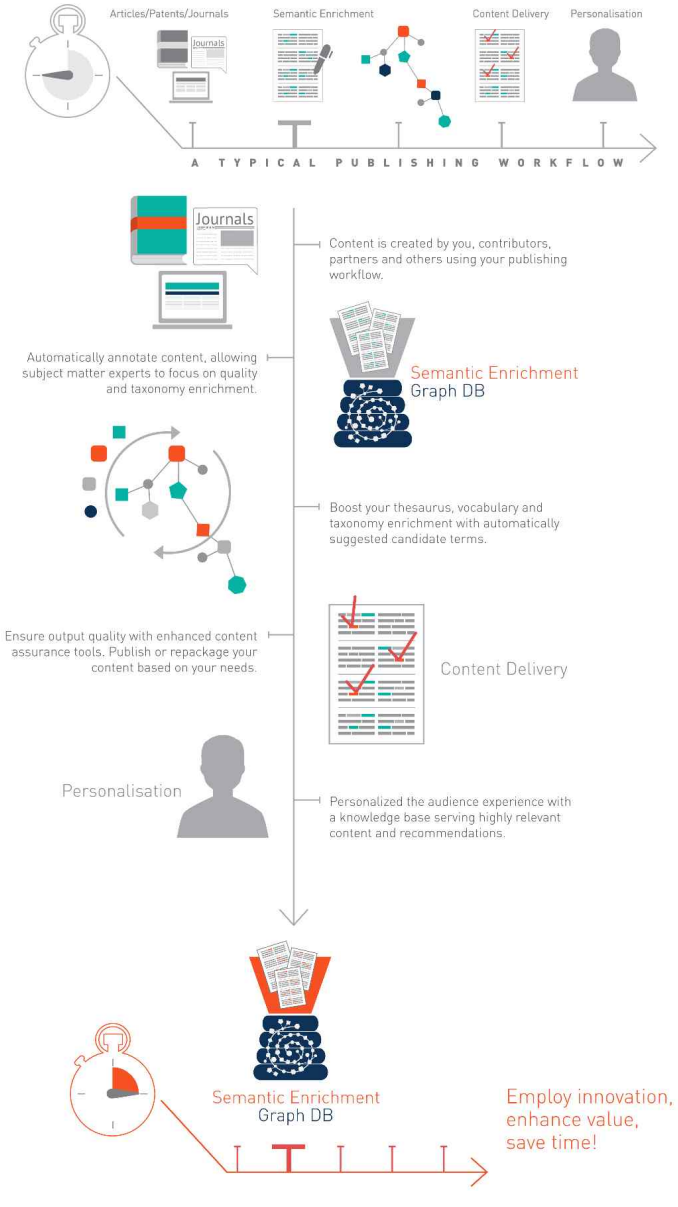
- 데이터베이스, 머신러닝 알고리즘, API와 관련도구들로 구성된 플랫폼임. 뉴스분야, 교육분야, 과학분야, 학술분야 및 기타 출판분야 등 광범위하게 활용 가능한 솔루션을 제공함
- 텍스트분석, 개체 식별, 개념추출과 분류를 위한 머신러닝 기술 적용
- 시맨틱 인덱스, 통합된 개체 프로파일, 링크드 데이터와 이용자 행동 프로파일 저장을 위한 데이터베이스 제공(GraphDB)
- 텍스트 분석, 모델 트레이닝, 검색, 추천, 컨텐츠 관리, 컨셉 프로파일을 위한 API 제공
- 문맥에 따른 저작(contextual authoring), 콘텐츠 향상 모니터링(enrichment monitoring), 큐레이션과 품질 보장, 탬블릿 정의를 위한 UI 구축을 위한 도구
- RDF 데이터 저장 및 프로세싱에 최적화된 GraphDB 저장소를 제공하며 수십억개의 RDF 트리플 관리가 가능함. 아래의 그림 예시와 같이 추론 엔진을 통해 GraphDB 저장소의 저장된 트리플에서 개체 간의 관계와 같은 정보 추론이 가능함



출처: ontotext.com

[그림 12] Ontotext GraphDB에 대한 추론 예시

- 처리가능한 데이터의 규모와 기능명세에 따라 Lite/Standard/Enterprise 버전을 제공하며, 클라우드 기반 옵션도 제공함
- 풀텍스트 대상의 텍스트 마이닝과 Semantic enhancement를 지원하는 S4(Self-Service Semantic Suite) 제공
- 아래의 그림은 Ontotext 기반의 과학적 출판 절차를 도식화한 것임. 생성된 저작물에 대해 자동화된 semantic enhancement가 순차적으로 진행됨

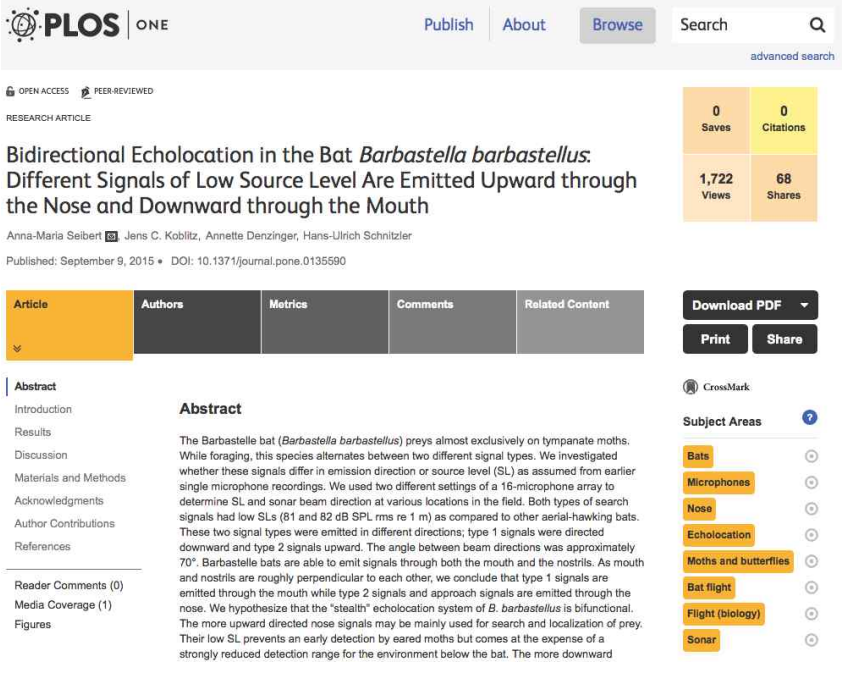


출처: ontotext.com

[그림 13] Ontotext Scientific Publishing 절차에 관한 인포그래픽

(2) Ambra Project

- 오픈 엑세스 저널을 RDF와 함께 발행 가능한 플랫폼으로 PLOS ONE/Medicine/Biology 등에서 활용중임
- 하나의 플랫폼으로 다수의 저널들에 대한 관리 지원
- 이용자 프로파일 및 토의와 같은 커뮤니티 기능 지원
- 아티클과 관련하여 외부 웹서비스로부터의 TrackBack과 Pingback 저장 및 검색 기능 지원
- 일자별, 주제 카테고리별로 신규 아티클에 대한 Atom 피드 발행 지원
- 아티클 또는 저널 단위의 노트와 코멘트에 대한 Atom 피드 발행 지원
- 아티클 인용정보, PDF 및 XML 파일 다운로드 기능 지원
- 관련된 아티클과 저널 이슈에 대한 링크정보 제공
- 주제카테고리 및 발행일 기준으로 아티클 탐색 지원
- 고급 검색기능 제공
- 콘텐츠 퍼블리싱 관리, 이용자 관리, 코멘트 관리를 위한 관리자 포털 지원
- 볼륨, 이슈단위 컬렉션 생성 기능 지원
- 웹 콘텐츠와 디지털 개체의 빠른 엑세스를 위한 캐싱기능 지원



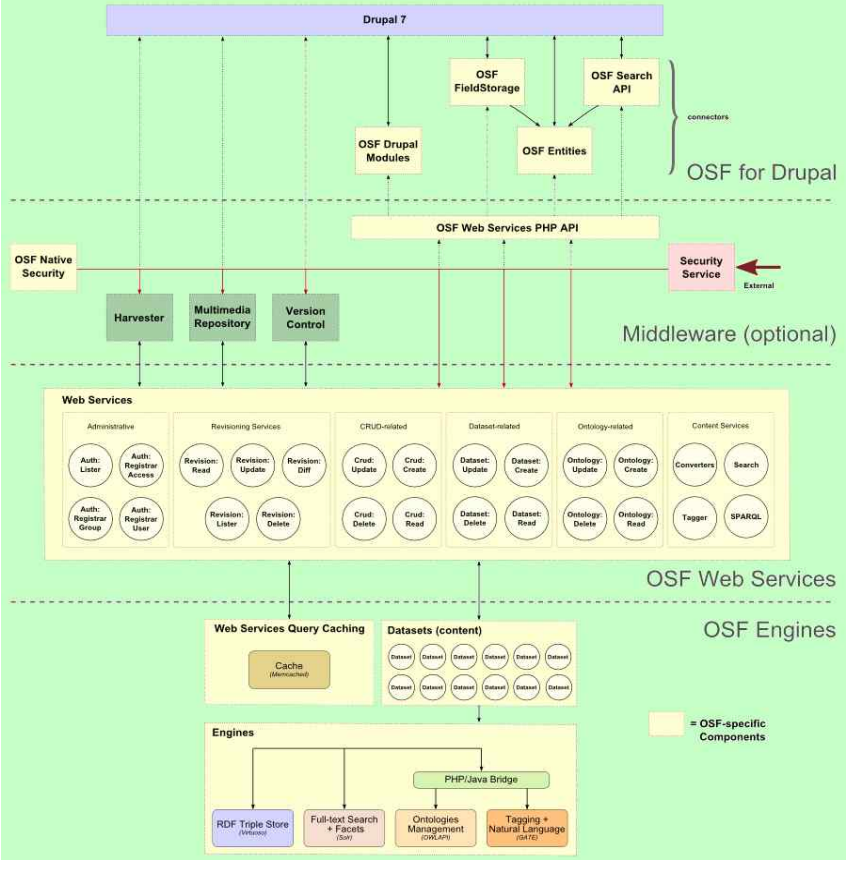
출처: http://journals.plos.org/

[그림 14] Abmra 플랫폼 기반의 PLOS One 서비스 화면 예시

(3) OSF(Open Semantic Framework)

- 오픈소스 시맨틱 기술을 활용한 소프트웨어 스택으로 시맨틱 기술에 기반한 지식 관리 플랫폼임
- 데이터 통합, 지식관리, 전사적인 시맨틱 검색, 분산데이터 엑세스, 비정형 데이터 관리 등을 특징으로 하며, Semantic Publishing 플랫폼으로 활용 가능
- 아파치 2 라이선스에 따라 활용 가능함
- 다양한 온톨로지의 반입을 지원하면 저장되는 모든 콘텐츠는 RDF 데이터 모델에 따라 저장됨

- OSF는 아래의 그림과 같이 다양한 오픈소스 소프트웨어가 결합되어 다계층 모델로 구성되어 있음



출처: http://wiki.opensemanticframework.org

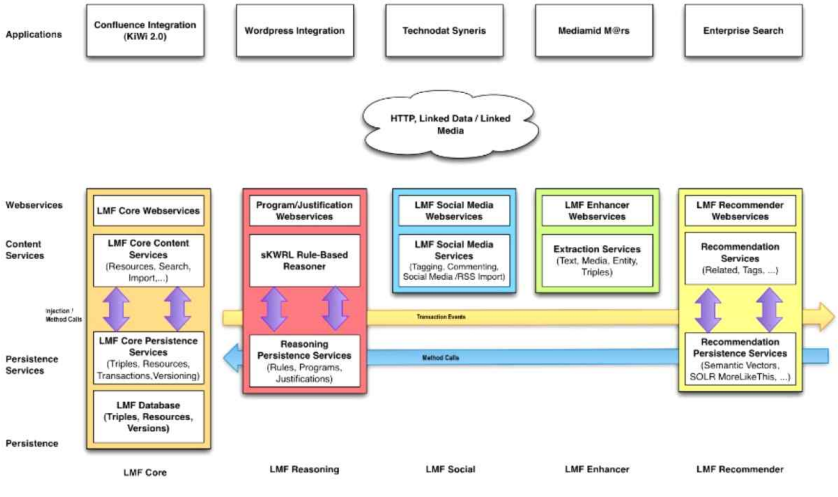
[그림 15] OSF 아키텍처

- 엔터프라이즈 환경에서 광범위하게 활용되고 있는 Drupal CMS와 결합되어 있으며, Virtuoso를 통한 RDF와 OWL 관리 및 Solr을 통한 검색 환경을 지원함

- 약 6년간의 개발기간을 통해 꾸준히 버전 업그레이드 되고 있으며, 손쉬운 설치를 위한 인스톨러를 제공
- 클라우드 환경 지원을 위한 Amazon EC2 AMI 이미지 제공

(4) LMF(Linked Media Framework)

- Linked Data 발행, 관리 및 시맨틱 검색 등을 지원하는 플랫폼으로 semantic enhancement, 시맨틱 검색 및 효율적인 트리플 데이터 관리를 지원함
- Linked Data 저장소, 검색엔진 및 개발 프레임워크 등을 모두 포함
- Apache 재단의 Marmotta, Stanbol 및 Solr 등 주요 오픈소스 소프트웨어가 결합되어 있음
- Linked Data 관리, 인터링킹 및 Linked Data 기반의 Semantic enhancement 기능 제공



출처: http://www.slideshare.net/wastl/linked-media-and-data-using-apache-marmotta

[그림 16] LMF 아키텍처

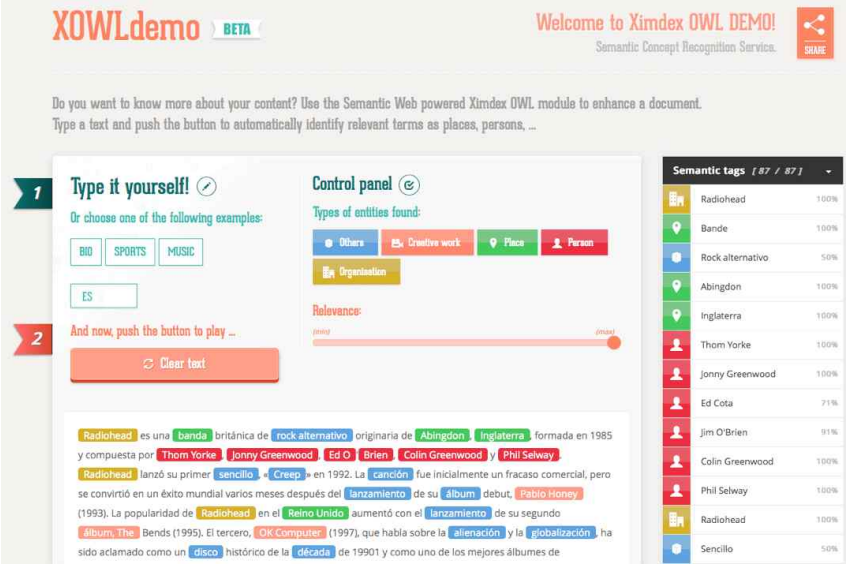
- 링크드 데이터 관리, semantic enhancement, 시맨틱 검색을 손쉽게 구현할 수 있다는 강점이 있음
- LOD 클라우드에 대한 빠른 접근을 위해 데이터셋에 대한 캐싱 기능을 제공
- 링크드 데이터의 추가, 수정, 삭제 제공 및 Apache Stanbol 모듈을 통해 LOD를 활용한 In-text semantic enhancement 기능을 제공

(5) redlink

- http://redlink.co
- 텍스트 분석, 분류, semantic enhancement를 지원하는 클라우드 기반 플랫폼
- Linked Data 발행, 텍스트 인터링킹 등을 제공하는 시맨틱 플랫폼 제공
- 아티클, 뉴스 등 풀텍스트에 대한 분석, 주요 키워드 자동 인터링킹 등 제공
- Apache marmotta, stanbol, solr 등에 기반한 기술로 플랫폼 구현. 클라우드 기반의 상업적 서비스를 제공함

(6) Ximdex

- technologies: PHP, J2EE, .NET, XML/XSLT, JS, RDF, JSON, RoR 등의 기술을 활용하여 콘텐츠, 데이터 및 웹 어플리케이션의 생성과 처리를 지원하는 시맨틱 콘텐츠 관리 시스템(Semantic Content Management System)
- 유연성, 확장성 및 보안성을 제공
- github를 통해 오픈소스로 제공
- In-text에 대한 semantic enhancement 지원



출처: http://demo.ximdex.com/xowl/

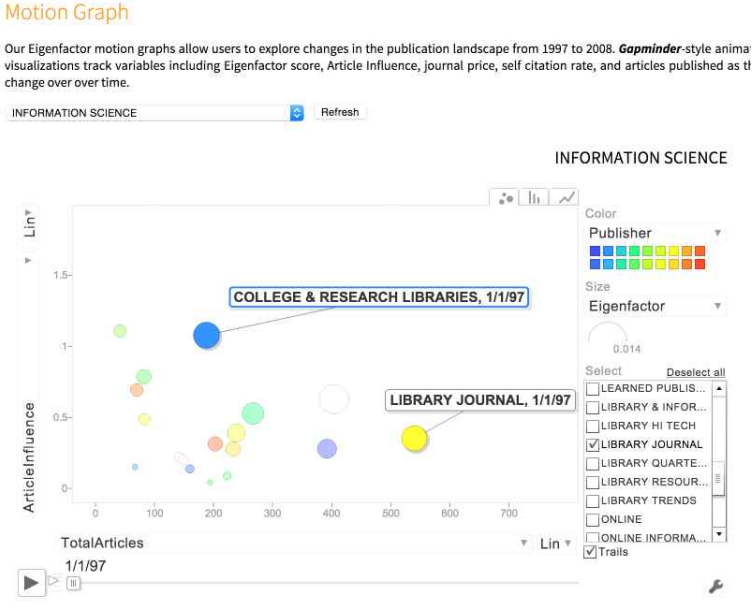
[그림 17] Ximdex 플랫폼 semantic enhancement 데모 화면

- 콘텐츠에 독립적인 구조로 모든 유형의 파일 포맷에 대응 및 자유로운 구조 설계 지원
- XML 자동변환 지원 및 손쉬운 편집을 위한 시각 인터페이스 제공
- object-action UI를 통한 이용자/권한/워크플로우의 시각적 편집 지원
- 플랫폼 중립적이고 유연한 설계로 다양한 어플리케이션 서버 환경에 대응하며 .NET, PHP, ASP, J2EE, XML/XSLT, flash, XHTML, HTML5 등 다양한 언어 환경을 지원함
- 모바일 환경에 대응

7. 시각화

(1) 학술정보 시각화 사례

- Eigenfactor
- 분야별 저널의 impact factor를 동적인 그래프로 시각화하여 구현한 웹 서비스로 학술출판의 경향을 손쉽게 파악할 수 있도록 함
- 1997년부터 2008년까지 impact factor가 가장 높은 저널들을 시각화하며, 다양한 조건들을 조합하여 시각화 결과를 조회 가능함



[그림 18] Information Science 분야에 대한 impact factor 시각화 예시

- Microsoft Academic Search ‘Visual Explorer’
- 학자, 연구자 및 학생 등이 어떻게 학술 콘텐츠와 연구자 또는 연구기관들을 찾는지에 대해 살펴보기 위한 목적으로 그래프 시각화를 시도함
- 학술저작물, 주제, 저자, 인용 등의 관계를 시각화하여 디스플레이함
- Silverlight 기술 기반으로 구현됨

[그림 19] 마이크로소프트 Academic Search 시각화 화면 예시

(2) 학술정보 시각화 관련 툴

8. 사례조사

(1) Semantic Digital Library 관련 사례

- Semantic Digital Library는 에이전트의 자동화된 처리를 포함하여 메타데이터의 포괄적인 접근이 가능하도록 시맨틱 웹 기술을 적용한 Digital Library 솔루션, 플랫폼 등을 의미함. 도서관, 박물관, 아카이브 등이 보유하고 있는 디지털 컬렉션에 대한 의미적 주석 또는 메타데이터를 추가하려는 다양한 시도가 있었음
- JeromeDL의 사례: Semantic Digital Library 영역에서의 한 사례로 이용자들이 도서, 아티클 등 정보자원에 의미적 주석을 추가할 수 있도록 지원함. 시맨틱웹 기술과 소셜 네트워킹 기술을 활용하였으며, 이용자들이 생성한 모든 주석은 RDF포맷으로 저장되도록 하여 다른 어플리케이션을 통해서도 활용가능하도록 함.
- BRICKS 프로젝트의 사례: BRICKS(Building Resources for Integrated Cultural Knowledge Services) 프로젝트는 현존하는 디지털 자원들을 공유된 저장소에 통합하고자 함. 메타데이터의 조직, 관리에 OWL-DL을 활용함
- DOMEO 프로젝트의 사례: DOMEO는 GWT(Google Web Toolkit)와 자바스크립트를 활용하여 개발된 웹 컴포넌트로 이용자들이 수작업(manually) 또는 반자동화된 방식(semi-automatically)으로 비구조화된(unstructured) 의미적 주석이나 일부 구조화된(semi-structured) 주석을 생성할 수 있도록 하고 이렇게 생성된 의미적 주석을 권한에 따라 공유할 수 있도록 지원함

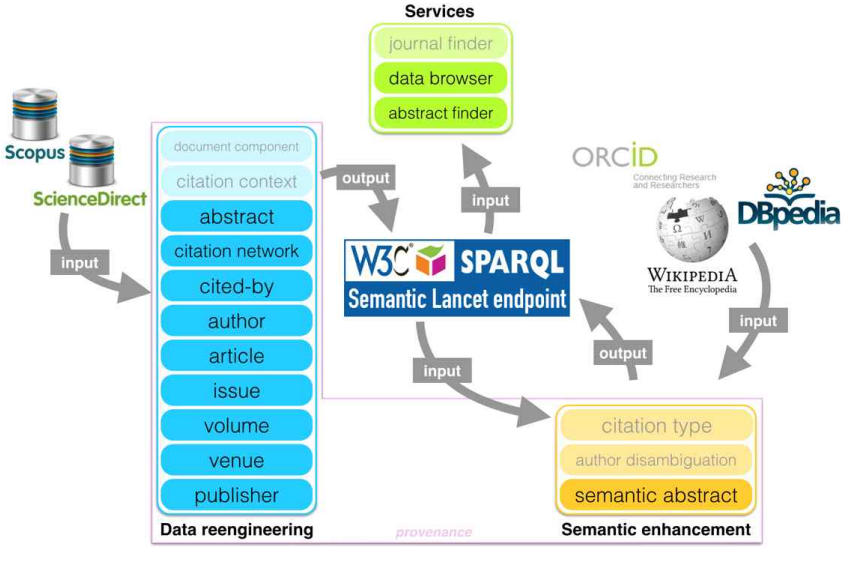
- Living Document의 사례: The Living Document makes use of the documents as an interface to the Web of Data(WoD), a self-descriptive document fully interoperable within the Web.

(2) LOD 기반 사례

가. LOD 발행사례

- NPG Linked Data Platform의 사례: 1845년 이후 Nature지에 발행된 페이퍼에 대한 정보를 약 4억개의 트리플 데이터로 변환하여 발행한 시도임. 데이터는 Dublin Core, FOAF, PRISM, BiBO 등을 활용하여 기술되어 있음. 꽤 높은 품질의 데이터이나 초록 전문이 포함되어 있지 않고 분류의 기준이 엉성하다는 문제를 포함하고 있음. 하지만 페이퍼 외적인 데이터, 이를테면 CrossRef와의 연결정보 등을 포함하고 있다는 것은 강점임
- JISC OpenCitation corpus의 사례: Open Access PubMed Central에 발행된 페이퍼에 대한 데이터를 무료로 제공함. corpus는 340만건 페이퍼에 대한 정보를 제공하며, SPAR 온톨로지, PRISM 및 Dublin Core를 활용하여 기술되어 있음. 특징적인 점은 PRO(http://purl.org/spar/pro)온톨로지 활용하여 역할(role) 정보와 원출처(authorship) 정보 등을 기술하고 제공하고 있는 것임
- BioTea 프로젝트의 사례: PubMed Central를 통해 제공되는 생물의학 관련 페이퍼들을 RDF로 발행하는 것을 목표로 함. 2400개 저널에 수록된 약 27만건의 페이퍼에 대한 정보를 기술함. BiBO, Dublin Core, FOAF등을 활용하여 기술되어 있음. 다양한 툴들을 활용하여 자동화된 워크플로우를 통해 RDF 데이터를 생성했다는 것, 그리고 DoCO 온톨로지를 활용하여 섹션, 단락 등에 대한 정보를 제공하는 것을 특징으로 함

- Semantic Lancet 프로젝트의 사례: 이 프로젝트는 학술적 출판물에 대해 풍부한 시맨틱 데이터를 제공할 수 있도록 하고, 이를 기반으로 다양한 서비스를 제공하는 것을 목표로 함. 첫째, SPAR 온톨로지에 따라 기술된 RDF 데이터를 생성하고, 둘째, 트리플스토어와 SPARQL 엔드포인트를 통해 RDF 데이터를 공개하여 제공하고 이 데이터에 기반하여 구축된 서비스를 제공하는 것임. 서비스 모델은 아래 그림과 같음. 1) Scopus와 ScienceDirect의 raw 데이터를 SPAR 온톨로지를 바탕으로 RDF로 변환하여 트리플스토어에 저장(data reengineering), 2) DBpedia, Wikipedia와 같은 다양한 소스의 정보를 활용하여 변환된 RDF의 품질을 향상시킴(semantic enhancement), 3) 이러한 절차를 거쳐 생성된 데이터를 활용한 웹서비스를 구축하여 이용자들에게 제공(service)

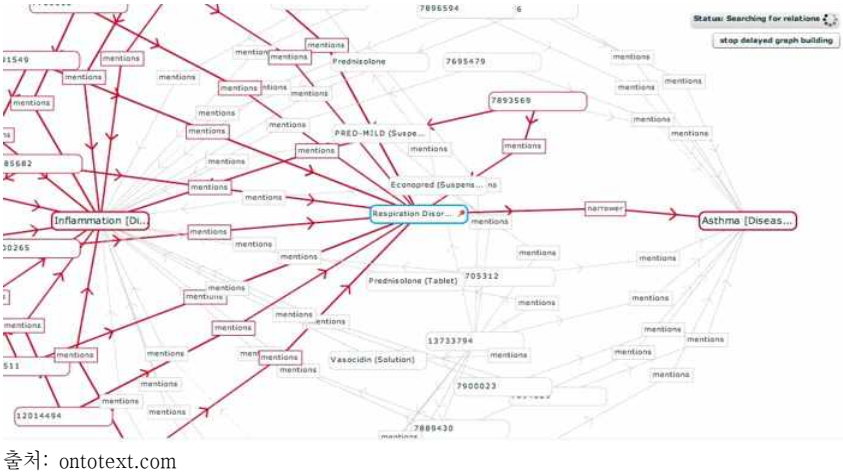


[그림 20] Semantic Lancet 서비스 아키텍처

- EBI RDF Platform 사례: EBI(European Bioinformatics Institute)는 유럽에서 가장 큰 규모의 생물정보공학 정보를 제공하는 기관임. (The EBI RDF platform: linked open data for the life sciences 참고하여 추가 기술 진행)

나. LOD 활용 사례

- AstraZeneca 사례: AstraZeneca는 글로벌 제약회사로 생물의학 개체 사이의 직간접적인 관계정보의 명확한 식별을 통해 초기 가설검증 절차를 향상시키고자 함. 이를 위해 25개에 이르는 데이터 소스를 통합하고, 관계정보 식별을 위해 링크드 데이터(Linked Life Data) 활용을 시도하였음. 링크드 데이터 통합, 관리 및 semantic enhancement를 지원하는 시맨틱 퍼블리싱 플랫폼을 활용하여 개체 식별 및 개체 간 관계정보 파악할 수 있도록 구현함. 결과적으로 연구자가 생물의학 개체 사이의 관계를 효과적으로 파악할 수 있는 환경이 구축됨. 시스템 구축에 Ontotext 플랫폼을 적용함



출처: ontotext.com

[그림 21] 생물의학 개체 간 관계의 시각화 예시

<참고문헌>

[1] Aberer, K., Boyarsky, A., Cudré-Mauroux, P., Demartini, G., & Ruchayskiy, O. (2011). ScienceWISE: a Web-based Interactive Semantic Platform for scientific collaboration. ISWC (Demonstration Track). Retrieved from https://diuf.unifr.ch/main/xi/sites/diuf.unifr.ch.main.xi/files/ScienceWISEdemopaperforISWC11-v7.pdf

[2] Bagnacani, A., Ciancarini, P., Di Iorio, A., Giovanni, A., Nuzzolese, S. P., & Vitali, F. (n.d.). Towards a Linked Open Dataset for Scholarly Publishing: Semantic Lancet Project. Semantic-Web-Journal.net. Retrieved from http://www.semantic-web-journal.net/system/files/swj744.pdf

[3] Bäumer, F. S., Geierhos, M., Gim, J., Jung, H., & Jeong, D.-H. (2014). Linked Open Data System for Scientific Data Sets. Retrieved from http://www.researchgate.net/profile/Michaela_Geierhos/publication/276287910_Linked_Open_Data_System_for_Scientific_Data_Sets/links/5555e49708ae6943a872bffd.pdf

[4] Castro, L. J. G., McLaughlin, C., & Garcia, A. (2013). Biotea: RDFizing PubMed Central in support for the paper as an interface to the Web of Data. Biomedical Semantics, 4(Suppl 1), S5. Retrieved from http://www.biomedcentral.com/content/pdf/2041-1480-4-S1-S5.pdf

[5] Ciccarese, P., Ocana, M., & Clark, T. (2012). Open semantic annotation of scientific publications using DOMEO. Journal of Biomedical Semantics, 3 Suppl 1, S1.

http://doi.org/10.1186/2041-1480-3-S1-S1

[6] Ciccarese, P., Wu, E., Wong, G., Ocana, M., Kinoshita, J., Ruttenberg, A., & Clark, T. (2008). The SWAN biomedical discourse ontology. Journal of Biomedical Informatics, 41(5), 739-751. http://doi.org/10.1016/j.jbi.2008.04.010

[7] Clark, T. (2014). Next Generation Scientific Publishing and the Web of Data. Semantic Web J. Retrieved from http://semantic-web-journal.org/system/files/swj670.pdf

[8] Damjanovic, V., Kurz, T., Westenthaler, R., Behrendt, W., Gruber, A., & Schaffert, S. (2011). Semantic enhancement: the key to massive and heterogeneous data pools. In Proceeding of the 20th international IEEE ERK (electrotechnical and computer science) conference. 147.91.128.71. Retrieved from http://147.91.128.71/sites/default/files/damjanovi__(semantic_enhancement_)p.pdf

[9] De Carvalho Moura, A. M., Porto, F., Vidal, V., Magalhães, R. P., Maia, M., Poltosi, M., & Palazzi, D. (2015). A semantic integration approach to publish and retrieve ecological data. International Journal of Web Information Systems, 11(1), 87-119. http://doi.org/10.1108/IJWIS-08-2014-0028

[10] De Waard, A. (2010). From Proteins to Fairytales: Directions in Semantic Publishing. Intelligent Systems, IEEE, 25(2), 83-88. http://doi.org/10.1109/MIS.2010.49

[11] Di Iorio, A., Lange, C., Dimou, A., & Vahdati, S. (2015). Semantic Publishing Challenge-Assessing the Quality of Scientific Output by Information Extraction and Interlinking. arXiv Preprint arXiv:1508.06206. Retrieved from http://arxiv.org/abs/1508.06206

[12] Di Iorio, A., Nuzzolese, A. G., Peroni, S., Shotton, D., & Vitali, F. (2014). Describing bibliographic references in RDF. In Proceedings of 4th Workshop on Semantic Publishing (SePublica 2014). http://ceur-ws. org (Vol. 1155). mayor2.dia.fi.upm.es. Retrieved from http://mayor2.dia.fi.upm.es/oeg-upm/files/eswc2014/Workshops/SePublica2014/SePublica_paper_5.pdf

[13] Garcia-Castro, A., Labarga, A., Garcia, L., Giraldo, O., Montaña, C., & Bateman, J. A. (2010). Semantic Web and Social Web heading towards Living Documents in the Life Sciences. Web Semantics: Science, Services and Agents on the World Wide Web, 8(2-3), 155-162. http://doi.org/10.1016/j.websem.2010.03.006

[14] Gradmann, S. (2014). From containers to content to context: The changing role of libraries in eScience and eScholarship. Journal of Documentation, 70(2), 241-260. http://doi.org/10.1108/JD-05-2013-0058

[15] Hare, J., Newman, D., Peters, W., Greenwood, M., & Eggink, J. (2014). SemanticNews: Enriching publishing of news stories. Retrieved from http://eprints.soton.ac.uk/366832/

[16] Jupp, S., Malone, J., Bolleman, J., Brandizi, M., Davies, M., Garcia, L., … Jenkinson, A. M. (2014). The EBI RDF platform: linked open data for the life sciences. Bioinformatics , 30(9), 1338-1339. http://doi.org/10.1093/bioinformatics/btt765

[17] Kurz, T., Güntner, G., Damjanovic, V., Schaffert, S., & Fernandez, M. (2014). Semantic enhancement for media asset management systems. Multimedia Tools and Applications, 70(2), 949-975. Retrieved from

http://link.springer.com/article/10.1007/s11042-012-1197-7

[18] Ma, Z. (2015). Generating Visualizations From RDF Graphs. Retrieved from http://courses.cecs.anu.edu.au/courses/CSPROJECTS/15S1/Reports/Zhuo_Ma_Report.pdf

[19] Mirizzi, R., Ragone, A., Di Noia, T., & Di Sciascio, E. (2010). Semantic Tag Cloud Generation via DBpedia. In E-Commerce and Web Technologies (pp. 36–48). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-15208-5_4

[20] Nicolay, R. (2014, December 9). Semantic Enhancement of Lecture Material. arXiv [cs.CY]. Retrieved from http://arxiv.org/abs/1412.2901

[21] Oliveira, P., & Rocha, J. (2013). Semantic annotation tools survey. In Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on (pp. 301–307). ieeexplore.ieee.org. http://doi.org/10.1109/CIDM.2013.6597251

[22] Peroni, S. (2012, May 10). Semantic Publishing: issues, solutions and new trends in scholarly publishing within the Semantic Web era. alma. Retrieved from http://amsdottorato.unibo.it/4766/

[23] Peroni, S., Tomasi, F., Vitali, F., & Zingoni, J. (2013). Semantic Lenses as Exploration Method for Scholarly Articles. In Bridging Between Cultural Heritage Institutions (pp. 118–129). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-54347-0_13

[24] Reeve, L., & Han, H. (2005). Survey of Semantic Annotation Platforms. In Proceedings of the 2005 ACM Symposium on Applied Computing (pp. 1634–1638). New York, NY, USA: ACM. http://doi.org/10.1145/1066677.1067049

[25] Ruiz-Iniesta, A., & Corcho, O. (2014). A review of ontologies for describing scholarly and scientific documents. In Proceedings of 4th Workshop on Semantic Publishing (SePublica 2014), CEUR Workshop Proceedings. Aachen, Germany: CEURWS. org. OA at http://ceur-ws. org (Vol. 1155). Retrieved from http://www.researchgate.net/profile/Oscar_Corcho/publication/263495382_A_review_of_ontologies_for_describing_scholarly_and_scientific_document/links/53d8bc880cf2a19eee836c97.pdf

[26] Seringhaus, M. R., & Gerstein, M. B. (2007). Publishing perishing? Towards tomorrow's information architecture. BMC Bioinformatics, 8, 17. http://doi.org/10.1186/1471-2105-8-17

[27] Shotton, D., Portwin, K., Klyne, G., & Miles, A. (2009). Adventures in semantic publishing: exemplar semantic enhancements of a research article. PLoS Computational Biology, 5(4), e1000361. http://doi.org/10.1371/journal.pcbi.1000361

[28] Shotton, D. (2009). Semantic publishing: the coming revolution in scientific journal publishing. Learned Publishing: Journal of the Association of Learned and Professional Society Publishers, 22(2), 85–94. http://doi.org/10.1087/2009202

[29] Soldatova, L. N., & King, R. D. (2006). An ontology of scientific experiments. Journal of the Royal Society, Interface / the Royal Society, 3(11), 795–803. http://doi.org/10.1098/rsif.2006.0134

[30] Tanweer Ali, D., & P. Abeles, T. (2014). The fate of academic publishing and academia in a semantic environment. On the Horizon. Retrieved from http://www.emeraldinsight.com/doi/abs/10.1108/OTH-06-2014-0022

[31] Thanos, C. (2014). The Future of Digital Scholarship. Procedia

Computer Science, 38, 22-27.
http://doi.org/10.1016/j.procs.2014.10.005

● 저 자

현미환	KISTI 정보서비스실 선임연구원 mhhyun@kisti.re.kr
이혜진	KISTI 정보융합연구실 선임연구원 hyejin@kisti.re.kr
김혜선	KISTI 정보서비스실 책임연구원 hskim@kisti.re.kr
장원홍	LG 상남도서관 janwonhong@gmail.com

KISTI 지식리포트 [No.45]

학술분야 시맨틱 퍼블리싱
(Semantic Publishing) 연구 동향

인 쇄	2015년 00월 00일
발 행	2015년 00월 00일
펴낸곳	한국과학기술정보연구원
펴낸이	한선화
편집인	김혜선
주 소	서울시 동대문구 회기로 66 고객센터(ARS) 080-969-4114
등 록	1991. 2. 12, 제5-258호
ISBN	000-00-000-0000-0 00000
인쇄처	승림디엔씨

※ 본 연구의 내용은 본 연구원의 공식적인 견해가 아닌 참여 연구원들의 의견임을 밝혀둔다.

□ KISTI 지식리포트 발행 목록

호	서명	저자	발간일
1	학술지 수집 정책 수립을 위한 국내 현황 분석	이재윤, 김혜선, 이혜진	2009.06.11
2	국내 과학기술지식의 글로벌 확산 전략	서태설, 최현규	2009.09.04
3	국가 과학기술 진흥을 위한 KISTI의 전략적 정보자원 개발방안	황혜경, 최호남, 윤희윤	2009.09.21
4	학술논문 오픈 액세스를 위한 공공접근정책 방향	서태설, 허 선, 노경란	2009.10.23
5	과학데이터의 공유와 활용	이상환, 심원식	2009.11.10
6	E.infrastructure기반 국가 R&D 정보서비스의 지능화 방안	송인석, 오세홍	2009.12.04
7	한국과학기술인용색인서비스(KSCI)의 현황 및 발전 전략	최선희, 이재윤	2010.01.29
8	학술정보센터의 새로운 서비스 모델 : 오픈 액세스 출판	노경란, 이혜진	2010.02.11
9	과학기술정보의 아카이빙 체제 구축	황혜경, 이선희, 최호남, 서혜란	2010.02.16
10	세계 주요 과학기술 정보기관의 최근 동향	이상환, 노경란, 김혜선, 황혜경, 정은경	2010.03.02
11	과학기술 정보자료 보존관리 : 현황분석 및 미래예측	이선희, 황혜경, 류범종, 윤희윤, 김석영	2010.06.30
12	리포지터리 사례분석 및 시사점 도출	이상기, 정영미	2010.10.07
13	디지털 콘텐츠 유통을 위한 저작권 쟁점 분석	유수현	2010.10.07
14	이공계 대학 교수의 과학기술정보 이용 현황	김환민, 김재훈	2010.12.02
15	계량서지적 분석용 공개 소프트웨어 활용 방안	최선희, 김희정, 이재윤	2011.01.24
16	2010년도 한국 과학자의 SCI 논문 계량분석	김완종, 노경란, 박민수, 최현규	2011.04.06
17	국내 과학기술정보 이용실태 조사 분석	박민수, 이상환, 최현규, 정정수	2011.04.08
18	국내 과학기술정보 이용자 니즈 및 형태 연구	박민수, 이상환, 최현규, 정정수	2011.04.08
19	학술지 유통환경 변화와 국내 학술지의 국제화	서태설, 김규환, 최현규	2011.06.07

호	서명	저자	발간일
20	우리나라 SCI급 논문의 영향력 분석 : NCR for Korea 1981-2010을 기준으로	김완종, 노경란, 최현규, 박민수	2011.08.17
21	연구자들의 소셜 미디어 이용	노경란, 최현규	2011.10.19
22	연구자를 위한 소셜 미디어 활용 가이드	노경란, 최현규	2011.10.19
23	서비스 사이언스 기반 과학기술 콘텐츠 서비스 방안	김지영, 신기정, 황혜경, 조부연	2011.10.26
24	효율적 연구를 위한 소셜 미디어 활용	노경란, 유수현, 최현규	2011.11.09
25	SEO(검색엔진최적화)를 통한 검색순위 올리기 전략	현미환, 이태석, 문영수, 권정혁	2011.11.11
26	연구자 협업지원형 정보서비스 사례연구	이혜진, 현미환, 김혜선, 박민수, 최현규	2011.12.01
27	과학기술 R&D 라이프사이클 연구 : 생명공학 및 나노분야를 중심으로	김혜선, 권나현, 정은경, 이정연, 최현규	2011.11.30
28	웹사이트의 사용성 개선을 위한 단계별 전략	현미환, 박민수, 이태석, 최현규	2011.12.05
29	국내 과학기술 연구자의 소셜 미디어 활용 현황	현미환, 이혜진, 김혜선, 박민수, 최현규	2011.12.08
30	이용현황 분석을 통한 학술정보 활용지표 개발	이혜진, 유수현, 김혜선, 이재윤	2012.01.31
31	FTA와 개정 저작권법이 정보유통에 미치는 영향	유수현, 이대희	2012.02.21
32	고품질 정보서비스 지원을 위한 컴퓨팅자원 인프라 구축	정영임, 신용주, 한성근, 김재훈, 김정환, 최호남	2012.03.30
33	모바일 앱(App) 개발을 위한 특화 기술 분석	현미환, 신수미, 김혜선	2012.08.10
34	한국과학기술정보연구원(KISTI) 원문제공서비스 현황 분석	이선희, 김지영, 문영수, 신기정	2012.12.31
35	정보서비스 공공기관 검색엔진 활용 만족도 조사	이태석, 신수미, 유수현, 정용일, 이은정	2013.07.31

호	서명	저자	발간일
36	국내외 오픈액세스 기술 및 서비스 동향	이은지, 현미환, 김완종, 김혜선	2013.11.29
37	학술정보의 공개 및 공유를 위한 법적 쟁점	유수현, 김완종, 김혜선, 최진원	2013.12.06
38	정보 공유를 위한 국내외 OpenAPI 서비스 동향	현미환, 신수미, 김혜선	2013.12.10
39	클라우드 기반 정보서비스 동향 및 사례	신수미, 현미환, 김혜선	2013.12.13
40	도서관 분야의 링크드 데이터 구축 동향	현미환, 김완종, 이혜진, 유수현, 김혜선, 박진호	2014.01.07
41	NDSL 연구자 정보검색 유형 분석	이태석, 신수미, 유수현, 이선희, 이은정, 김혜선	2014.05.30
42	개방형 데이터(Open Data) 평가를 위한 오픈데이터 측정 지표 현황 분석	현미환, 이혜진, 김혜선, 박진호	2014.12.01
43	국내외 공공데이터 제공 및 활용 사례 연구	이혜진, 현미환, 김혜선	2014.12.08
44	학술 출판을 위한 온톨로지 어휘	서태설, 김 평, 김혜선, 조성남	2015.10.16