

Keyword	LOD(Linked Open Data), 학술정보, 데이터, 시각화		
기술보유 기관	한국과학기술정보연구원 (KISTI)	기술판매형식	기술협력, 라이선스
연구 책임자	서 태 설	기술 완성단계(TRL)	시제품단계 (5단계)

기/술/개/요

문서로부터 식별된 키워드의 링크 정보를 생성하고, 그 링크 정보를 이용하여 각 키워드의 LOD가 인터링킹된 태그 클라우드를 화면에 표시하는 학술정보 태그 클라우드 기술

기존 기술의 문제점

- 연구자 학술지 정보 검색 시 개별 확인으로 인해 연구 내용 파악에 많은 시간 소요**
 - 관련연구 검색을 위해서는 다수의 학술지 및 학술내용을 일일이 검색하여 상세내용을 확인하여야 하므로 효율적으로 관련분야의 내용을 개괄적으로 파악할 수 있는 방법 모색
- 문자위주의 학술정보 제공으로 인해 직관적이고 효과적인 검색 제공 방법 모색 필요**
 - 종래 학술정보의 키워드를 시각화하여 표현할 수 있을 정도의 고품질 데이터를 확보하지 못하였고, 디비피디아(DBPedia)의 사전을 LOD(Linked Open Data)로 인터링킹할 때 평이한 문서를 단순 표현하여 정보전달력에 한계 존재
- 다양한 정보 수요자를 고려하여 학술지 정보 분석 범위 확대 필요**
 - 학술정보의 인용 분석 데이터 및 원문 XML을 기반으로 다양한 서비스가 개발되고 있으나, 다양한 분야에 적용될 수 있는 서비스 모델을 개선하고 향후 확대에 대한 기술개발 필요

기술 내용 및 차별성

기술 내용 차별성

정보문서의 태그 클라우드 기술을 통한 고품질의 시각화 표현 및 정보 전달의 질 획득

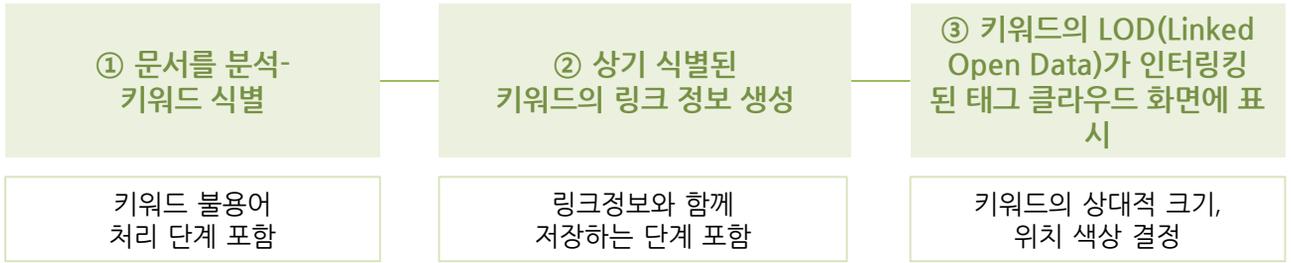
기술 내용

- XLM 본문 키워드 자동 식별, 링크노드 생성 후 태그 클라우드로 시각화된 화면상 표현
- 학술논문(DB, PDF, XML)으로부터 시각화 요소 추출 기능
- 추출된 키워드의 분석 기능(빈도수에 의한 중요도 계산)
- 키워드의 LOD 자동 태깅 및 DBPedia 인터링킹

차별성

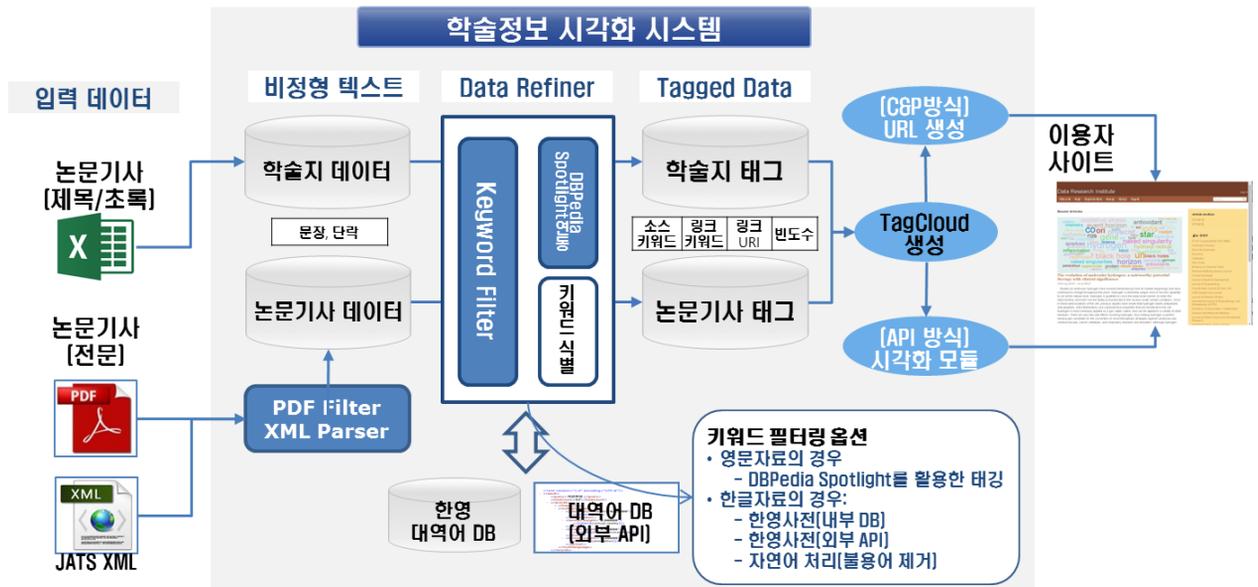
- LOD를 사용하여 독자, 저자, 도서관/출판사 관점의 요구 정보를 통해 얻어진 항목 추가 가능
- 종래기술은 독자 관점에서 정보를 제공하였기에 학술정보 내용 중심으로 제한적 항목 제공(경쟁기술 JATS(Journal and Article Tag Suits)의 경우 Journal Title, ISSN, Publish Name 등 3개 항목만 포함)
- XML 뿐만 아니라 PDF 문서로부터 문자를 추출, 보급형 학술정보 시각화 시스템 제공가능

주요기술구성



구현방법

- 문서로부터 식별된 키워드들의 링크 정보 생성 후 각 키워드의 LOD가 인터링킹 된 태그 클라우드를 화면에 표시
 - 학술논문을 분석하여 키워드들을 식별, 개발형 사전 시스템으로부터 식별된 키워드들에 대한 링크정보를 전달받아 저장, 링크정보 생성부는 태그단어, 논문 아이디, 섹션, 링크면, 링크 URI, 빈도수 등을 데이터베이스에 저장
 - LOD(Linked Open Data) 인터링킹(Interlinking)을 통해 태그클라우드 상의 단어 의미 해설
- XML 데이터로부터 직접 키워드를 식별, 통계 처리, 가중치 부여 등을 거쳐 시각화
 - 키워드의 시각화 요청 후 각 키워드의 빈도수 및 획득 사용자의 관심 키워드를 기반으로 키워드의 상대적 크기, 위치, 색상 등 결정하며 사용자의 관심키워드를 획득하여 이를 기반으로 시각화되어 표시
 - 구조화, 시각화 및 시각화 표현 기술을 활용한 LOD 연계 태그 클라우드를 사용하여 기존 태그 클라우드 대비 키워드 식별, 통계 처리, 가중치 부여 수준 향상



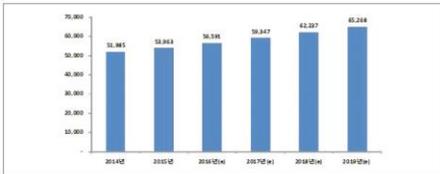
기술 동향

- **국내** : 정부 주도 혹은 글로벌 기업의 주도로 지식/지능처리 소프트웨어 관련 연구 진행 중
- 국외 기업들과의 기술격차를 줄이기 위해 정부 주도로 장기간의 국책사업이 시작 단계에 있음
- **국외** : 인공지능, 시맨틱 웹, 빅데이터, LOD 기술을 선도하고 있음
- 네이처(Nature) : 미국의 전문 과학잡지인 '네이처(Nature)'가 미국의 최대 백과사전인 브리태니커와 온라인 백과사전인 '위키 피디아'의 오류 분석을 통하여 인터넷을 통한 정보의 정확성을 검증
- IBM : DeepQA 프로젝트를 통해 Watson 질의응답시스템을 2006년부터 개발
- Google : 인공지능 관련 벤처기업인 딥마인드테크놀로지(Deep Mind Technologies)를 인수하여, 기계 학습과 언어 프로세스를 검색 엔진에 통합
- **일본** : 2011년부터 Fujitsu 연구소와 NII(National Institute for Information)가 공동으로 2021년 동경대학교 입시 합격이 가능한 수준의 인공지능 시스템 개발을 위해 Todai 프로젝트를 진행

시장 동향

- **학술정보 서비스 시장 규모 성장, 데이터 서비스 분야의 전반적인 성장으로 적용가능 범위 확대**
- 국내 데이터 서비스 시장 전망은 연평균 성장률 6.3%로 2019년까지 약 8천억 원대 전망
- 학술분야는 2015년 추정치가 522억 원 정도
- **디지털정보콘텐츠 시장에서는 비즈니스와 산업 정보를 네트워크를 통한 서비스 제공과 이용 증가에 따라 시장규모 성장**
- 향후 AI, 빅데이터, 맞춤형 정보제공, 온라인 기반의 정보제공 방식의 변화 등이 성장을 견인할 것으로 전망

<디지털정보콘텐츠 시장규모 및 전망>



※ 출처: 2016 국내 디지털 콘텐츠 실태조사, 2016, 舊 미래창조과학부

<국내 데이터 서비스 시장 전망>

(단위: 억원)

구 분	2014년	2015년(예)	2016년(예)	2017년(예)	2018년(예)	2019년(예)	CAGR 2014~2019년
데이터 서비스	58,633	62,323	63,944	68,297	73,458	79,528	6.3%

※ 출처: 2016 데이터산업 백서, 2017, 한국데이터진흥원

기술활용분야 및 권리현황

기술활용분야

기술 수요처	적용처
학술 정보 검색 제공자	논문, 논문기사 시각화 자료 제공
인터넷 검색 업체	실시간 순위 집계 서비스
홍보 수단	빅데이터 통계를 이용한 문화재, 행정, 인물 정보 등의 홍보수단으로 이용

권리현황

- 국내 등록특허 3건

발명의 명칭	특허번호	비고
학술논문 XML 폴텍스트 생성을 위한 자동화 입력 시스템 및 방법	2012-0133622	등록

추가기술정보

기술분류	정보/통신>소프트웨어/ 임베디드S/W
관련과제 정보	(사업명) 2017년 기술실용화사업 (과제명) LOD 연계 학술정보 시각화 시스템 실용화 (수행기간) 2016.06~2016.12 (수행기관) 한국과학기술정보연구원

시장전망
학술정보 서비스 시장 및 전문정보제공 분야를 포함한 디지털정보콘텐츠 시장은 꾸준히 5% 이상 성장

기술문의
서태설 책임연구원(정보서비스실)
02-3299-6290
tsseo@kisti.re.kr

기술문의
한만호 실장(기획부/성과확산실)
042-869-0945
mhh7535@kisti.re.kr