

엑소브레인 SW: 지능진화형 빅데이터 지식처리 인공지능 기술 개발

한국전자통신연구원 | 박상규*·김현기·장명길·류범모·임수중

21세기 데이터 신산업혁명의 촉발에 따른 빅데이터 분석 및 인공지능 고도화 SW 기술 확보가 국가 및 기업의 미래 경쟁력 좌우하는 지식산업혁명시대가 도래하였다. 본고에서는 차세대 지능형 플랫폼의 핵심 기술인 인간과 기계의 전문지식 소통을 위한 빅데이터 기반의 지식처리 인공지능 SW인 엑소브레인 SW 과제에 대해 소개한다. 또한 엑소브레인 SW 과제의 2차년도에 걸쳐 수행된 연구결과, 관련분야 시장 동향 및 미래 기술 전망에 대해 소개하고자 한다.

I. R&D 개요 및 필요성

엑소브레인(Exobrain)은 내 몸 바깥에 있는 인공 두뇌(外腦)라는 뜻으로, 기계가 자연어를 이해하고 지식을 자가학습하여 자연어로 기술된 사용자 질문에 대해 정답을 제공할 수 있는 자연어 질의응답(Natural Language Question Answering) SW이다. 엑소브레인 SW의 개발 목표는 다양한 IoT(Internet of Things) 산업 환경에서 전문가 수준의 질의응답 지식 서비스를 제공하는 것이다.

현재 정보의 80%는 자연어로 기술된 비정형 텍스트 빅데이터이며, 텍스트 빅데이터에서 '누가 먼저 많은 가치를 추출해 내느냐'가 기업과 국가의 성패를 좌우하는 지식산업혁명 시대가 도래하였다. 현재 선진국에서는 인공지능 SW 개발을 위해 국가 차원 혹은 글로벌 기업의 주도하에 대형·장기 프로젝트가 진행 중이다. 국가 주도의 프로젝트로는 미국 DARPA의 자연어 텍스트 의미이해를 위한 Machine Reading 프로젝트, 유럽의 인간 두뇌의 인지 형태로 지식을 처리

* 종신회원

† 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 SW 컴퓨팅산업원천기술개발사업의 일환으로 수행하였음. [10044797, 사용자와 의사소통을 하고 스마트기기 간 자율협업을 통한 지식공유 및 지능진화가 가능한 엑소브레인 SW 기술 개발]

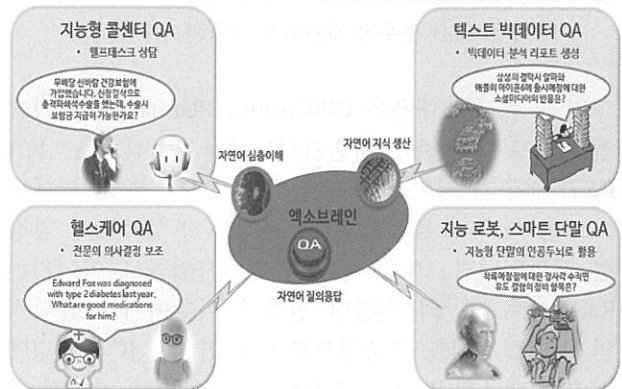


그림 1 엑소브레인 SW 개념



그림 2 연구 배경 및 개발 필요성

가능한 Human Brain 프로젝트(2013년부터 향후 10년간 추진), 일본에서는 2021년 동경대 입시 합격을 목표로 Today 로봇 프로젝트를 2011년부터 추진 중이다. 글로벌 기업의 경우에는 IBM의 Watson Deep QA, Google의 Knowledge Graph, MS의 Bing 검색엔진과 Apple의 Siri에 정답을 제공하는 울프럼 알파(Wolfram Alpha)가 대표적인 사례이다.

미래창조과학부는 First Mover형 SW 원천기술 확보를 위해 2013년부터 2022년까지 총 10년, 3단계로

수행되는 그랜드 챌린지 프로젝트인 엑소브레인 SW 기술개발 과제를 시작하였다. 본고에서는 지식 및 지능 처리 SW인 엑소브레인 과제의 핵심 연구내용 및 결과물에 대해서 살펴보고, 관련기술의 시장 동향 및 미래 기술 전망에 대해서 살펴보겠다.

2. R&D 주요내용 및 결과물

본 연구는 10년간 총 3단계에 걸쳐 수행되며, 최종 목표는 자연어를 이해하여 지식을 자가학습하며, 전문직종에 취업 가능 수준의 인간과 기계의 지식소통이 가능한 지식과 지능이 진화하는 SW 개발이다. 1단계(2013.5~2017.2) 연구목표는 일반지식 대상 분석형 엑소브레인 SW 기반기술 개발이며, 2단계(2017.5~2020.2) 연구목표는 전문지식(예: 의료, 법률, 금융 등) 대상 협업추론형 엑소브레인 SW 응용기술 개발이다. 마지막 3단계(2020.5~2023.2) 연구목표는 글로벌 전문 지식 대상 문제해결형 엑소브레인 SW를 개발하여 글로벌 시장에 진출하는 것이다. 차세대 인공지능 분야 세계 최고 기술선도를 위한 단계별 기술적 도적목표는 아래의 그림 3과 같다.

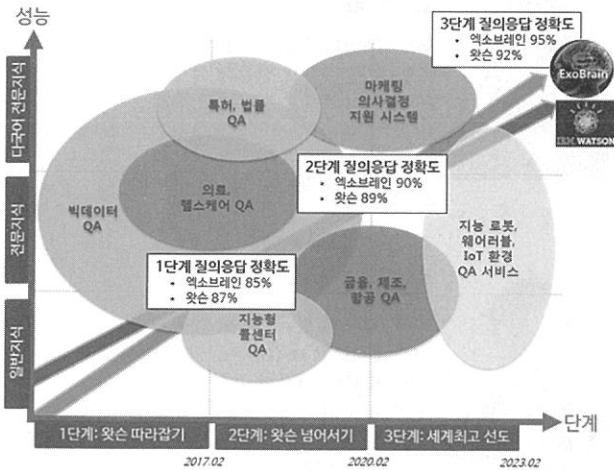


그림 3 단계별 연구목표

현재 추진중인 1단계는 세부과제 간 연구협력을 통한 최종목표 달성을 추진하는 병렬형 과제로 추진되고 있다. 한국전자통신연구원이 총괄과제 및 1세부과제를 주관하여 자연어 이해, 지속적 언어지식 학습, 자연어 질의응답 기술을 개발하고 있다. 2세부과제는 솔트룩스가 주관기관으로, 지식생산을 통한 지식베이스 구축을 추진하고 있다. 3세부과제는 인간모사형 자가학습 원천기술 개발을 위해 접근방법을 달리한 2개의 세부과제로 진행되고 있으며, 3-1세부과제는 기호적

접근방법으로 KAIST, 비기호적 접근방법으로 포항공대가 자가학습 원천기술을 개발 중이다. 4세부과제는 지식/기기의 협업을 위한 지능형 프레임워크 개발을 한국전자통신연구원이 추진 중이다.

엑소브레인 SW 개발을 위해 필요한 3대 핵심 기술인 인간모사형 지능 기술, 자율학습 기반 지식진화 기술, 문제해결형 협업 기술 확보를 위한 세부과제별 역할은 아래의 그림 4와 같다.

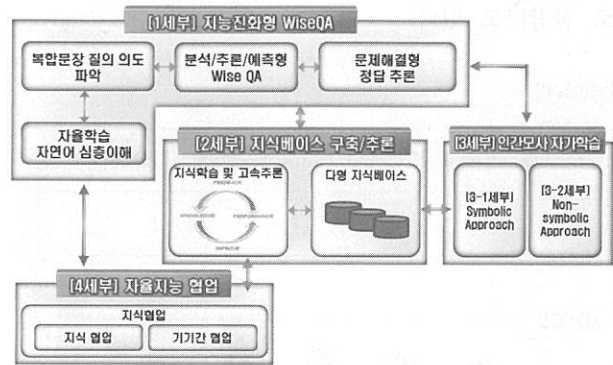


그림 4 세부과제 구성

2차년도 종료시점의 주요 연구성과로는 국내외 최고 수준의 자연어 이해기술을 개발하여 12개 기관에 기술보급을 추진하고 있으며, 단문단답형 자연어 질의응답 기술을 국내 최초로 개발하여 퀴즈대회 우승자 수준으로 성능 개선을 추진 중이다. 또한, 총 4,480만 트리플(Triple)의 아시아 최대 규모 지식베이스를 구축하여 확장을 추진 중이며, 국내 최대규모의 한국어 어휘지도를 구축하여 배포를 추진하고 있다. 또한 국내외 논문 게재 159건(SCI/E 19건 포함), IPR 확보를 위한 특허 출원 70건 및 국내외 표준기초서 15건을 제출하였다. 특히, 2015년도에 ITU-T에서 “Intelligent Question Answering Framework Service”를 국제표준 승인을 목표로 추진 중이다. 더불어 산/학/연 연대 과제 참여인력 300여명에 대한 인공지능 교육, 국내외 워크샵, 기술개발을 통한 인력양성 및 생태계 조성을 병행 중이다.

3. R&D 시장동향 및 전망

빅데이터 지식처리 인공지능 기술과 관련된 시장 동향은 IDC에 따르면 빅데이터 관련 분야는 연평균 40%의 성장률을 보이며 2015년 약 170억 달러로 예상되며, 인공지능 분야는 2013년 8억 달러에서 2015년 370억 달러로 증가되어 빅데이터 분야 시장을 능가할 것으로 예상된다. 이는 인공지능 기술이 빅데이터 관

런 전 분야에서 핵심 기술로 등장하게 될 것이기 때문으로 분석된다. 그림 5에 빅데이터 및 인공지능 기술의 시장 동향을 나타내었다.

그림 5와 같이 아직 인공지능 관련 시장은 시장 도입기 단계에 있는 것으로 전망되나, 인간의 지식노동을 전문가 수준으로 보조할 수 있는 자연어 질의응답 기술은 아래의 그림 6에서 Gartner, Techcast, McKinsey 등이 2020년 이후에 IT 역사 상 가장 혁신적 기술로 활용되어 고부가가치의 인공지능 시장이 형성될 것으로 전망하고 있다.

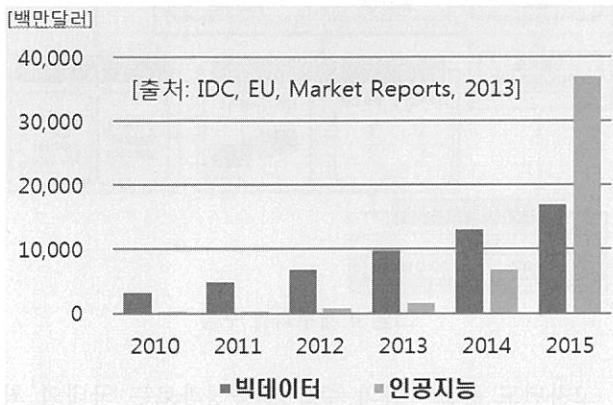


그림 5 빅데이터와 인공지능 시장 동향

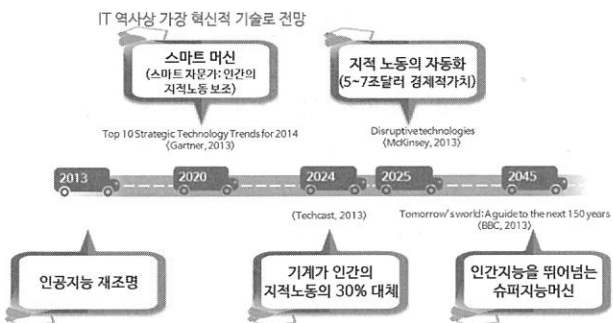


그림 6 지식처리 인공지능 미래 전망

엑소브레인 SW의 자연어 질의응답 기술은 인공지능 분야의 제일 어려운 난제이나, 성공 시 매우 수익성이 높은 SW분야의 미개척 분야이다. 국가 SW 경쟁력 확보를 위해 엑소브레인 과제를 성공적으로 수행하여 고위험, 고비용, 고난이도 차세대 인공지능 SW의 핵심 원천기술 확보하고자 한다. 이를 통해, 전문가 수준의 지식을 제공하는 지능형 의사결정 지원

SW, 스마트 기기 및 휴머노이드 로봇 등에 탑재되어 지식서비스를 제공하는 스마트 컨설턴트, 지식 및 지능 처리 신산업 창출, 글로벌 시장 진출 등을 통해 국가 SW산업의 미래 경쟁력을 확보하고자 추진 중이다.

약 력



박 상 규

1982 서울대학교 컴퓨터공학 학사
1984 KAIST 전산학 석사
1997 KAIST 전산학 박사
1987 한국전자통신연구원 자동통역인공지능연구센터장
2014~현재 한국전자통신연구원 SW기반기술원

구분부장

관심분야: 자연어 질의응답, 자동 통번역, 음성인식



김 현 기

1994 전북대 컴퓨터공학 학사
1996 전북대 컴퓨터공학 석사
2005 Univ. of Florida 전산학 박사
2011~현재 한국전자통신연구원 지식마인팅연구실장

관심분야: 자연어 이해, 텍스트 마이닝, 자연어 질

의응답



장 명 길

1988 부산대학교 계산통계학 학사
1990 부산대학교 전산학 석사
2002 충남대학교 정보과학 박사
1990~현재 한국전자통신연구원 지식마인팅연구실 책임연구원

관심분야: 자연어처리, 텍스트마이닝, 정보검색

및 질의응답



류 범 모

1995 경북대 컴퓨터공학 공학사
1997 POSTECH 컴퓨터공학 공학석사
2009 KAIST 전산학 공학박사
1997 한국전자통신연구원
2009~현재 한국전자통신연구원 지식마인팅연구실 선임연구원

관심분야: 정보검색, 자연어처리, 온톨로지, 질의응답 기술



임 수 종

1997 연세대학교 수학 학사
1999 연세대학교 컴퓨터과학 석사
2014 연세대학교 전산학 박사
2000~현재 한국전자통신연구원 책임연구원
관심분야: 자연어처리, 기계학습, 지식처리