

2019. 11. 20. ISS-00088

국민청원 데이터에 기반한 SW분야 이슈 분석

김정민 연구원 (jungmink26@spri.kr)
공영일 책임연구원 (kong01@spri.kr)

- 본 보고서는 「과학기술정보통신부 정보통신진흥기금」을 지원받아 제작한 것으로 과학기술정보통신부의 공식의견과 다를 수 있습니다.
- 본 보고서의 내용은 연구진의 개인 견해이며, 본 보고서와 관련한 의문사항 또는 수정·보완할 필요가 있는 경우에는 아래 연락처로 연락해 주시기 바랍니다.
 - 소프트웨어정책연구소 데이터분석·동향팀 김정민 연구원 (jungmink26@spri.kr)

《 요 약 문 》

국민청원은 2017년 정부의 출범 이후 줄곧 주요 사회이슈의 출발점으로 작용하며 대중 및 언론의 주요 관심사가 되어 왔다. 서비스 개시 이래 월 평균 17,000건의 청원이 꾸준히 등록되고 있으며 여론의 향방을 파악하는 근거로 활용되고 있다.

국민청원은 청원이 만료되는 30일간 개별 청원에 대한 국민들의 동의가 20만을 넘는 경우 청와대가 직접 답변을 하게 되어 있다. 그러나 이 답변 기준을 전문 영역인 소프트웨어 분야 청원이 넘기란 어렵기 때문에, 청원의 내용과 무관하게 조명되지 못하는 상황이다. 이런 연유에서 본고는 국민청원에서 소프트웨어 청원을 분류하고 이를 분석하여 주요 청원사례와 소프트웨어 기술에 대한 국민 인식을 정량화 하여 제시하였다.

첫 번째로 빅데이터 분석을 통해 2년간(2017.10.01.~2019.09.30.) 등록된 소프트웨어 분야 청원 2,050건을 6가지의 대주제(가상화폐, 개인정보, 게임산업, 교육, 기업지원, 생태계)로 분류하였다. 또한 유사도 측정 기법을 통해 각 주제에 가장 잘 부합하는 실제 청원을 탐색하고 이를 검토하여 관련 사례를 요약·정리하였다.

두 번째로 세간에 관심을 받는 소프트웨어 기술 총 4건(인공지능, 빅데이터, 블록체인, 가상현실)과 관련해 국민 관심의 우선순위를 측정하였다. 그 결과 인공지능 및 가상현실은 연구개발, 빅데이터는 교육, 블록체인은 규제에 관련한 국민의 관심이 높은 것으로 나타났다.

본 연구는 대규모 청원 정보를 목적에 맞게 분류 및 요약하여 표면적으로 드러나지 않는 소프트웨어 분야 청원을 종합해 보았다는데 의의가 있다. 이처럼 기존 SW분야 정책 개발 과정에서 상대적으로 간과되었던 일반 국민의 여론을 파악해봄으로써, 과학기술 분야의 국민 의견 수렴을 위한 맞춤형 창구의 필요성을 시사하였다.

《 Executive Summary 》

Since the inauguration of the government in 2017, National Petitions have been a major concern for the public and the media as a starting point for major social issues. Since the launch of the service, an average of 17,000 petitions have been registered on a monthly basis and have been used as a basis for identifying the direction of public opinion.

The National Petition is to be answered directly by the Blue House if the public's consent to an individual petition exceeds 200,000 for a 30-day period. However, it is difficult to achieve this criterion in the field of software, the area of expertise, so it is not illuminated regardless of the content of the petition. For this reason, this paper quantitatively presents the main cases of petitions and public perceptions of software technology by classifying and analyzing software petitions.

First, through the big data analysis, 2,050 petitions registered in the software field for two years (2017.10.01. ~ 2019.09.30.) Are subject to six major themes (virtual currency, personal information, game industry, education, corporate support, ecosystem). In addition, the similarity measurement technique was used to explore the actual petitions that best fit each subject, and related cases were summarized and organized.

Second, we measured the priorities of the public's attention with respect to four software technologies that are of great interest (artificial intelligence, big data, blockchain, virtual reality). As a result, AI and virtual reality showed high interest in R & D, big data in education, and blockchain in regulation.

The purpose of this study is to classify and summarize large-scale petition

information according to the purpose, and to synthesize the petitions that are not apparent on the surface. In this way, by understanding the relatively overlooked public opinion in the process of developing policy in the SW field, it suggested the necessity of environment for collecting public opinions in the science and technology field.

《 목 차 》

1. 연구 배경	1
2. 연구 방법	3
3. 소프트웨어 분야 청원 사례 분석	5
(1) 가상화폐	7
(2) 기업지원	8
(3) 보안	9
(4) 교육	10
(5) 게임산업	11
(6) 생태계	13
4. 소프트웨어 기술에 대한 국민의 인식	15
5. 요약 및 시사점	17
부록. 청원 데이터 분석 방법론	19

《 TABLE OF CONTENTS 》

1. Research Background	1
2. Research Method	3
2. Analysis of Petition Case on Software Area	5
(1) Crypto-currency	7
(2) Corporate Support	8
(3) Privacy	9
(4) Education	10
(5) Game Industry	11
(6) Ecosystem	13
3. People's Perception of Software Technology	15
4. Summary and Implications	17
Appendix. Petition Data Analysis Methodology	19

1. 연구 배경

□ 국민청원은 국가 사회 이슈 촉발의 출발점

- 국민청원은 ‘국민이 물으면 정부가 답한다’는 철학에 근거하여 2017년 8월부터 시행된 개방형 전자민원임
 - 개별 청원의 국민들의 동의가 기준선(200,000 동의)을 상회하는 청원에 대해 청와대가 직접 관련 담당자를 통해 답변¹⁾
- 월 평균 약 17,000건²⁾의 게시글이 꾸준히 등록되는 공개된 빅데이터로서 여론동향 파악 및 학술연구에 활발히 활용
 - 청원 게시글의 내용을 기사의 소재로 활발히 활용³⁾
 - 국민청원을 여론의 근거로 활용한 국문논문의 수⁴⁾가 해마다 증가하는 추세(2018년 48건 → 2019년 65건)
- 활발한 청원 개진이 이루어지는 반면, 청와대 답변 의무의 기준선이 높아 실제 참여 대비 답변율은 월평균 약 10% 미만⁵⁾
 - 첨예하게 대립하는 정치적·사회적인 주요 쟁점에 대한 찬반 여론이 절대 다수의 분야별 청원을 잠식해 버리는 현상이 발생

□ 청원 속 숨겨진 소프트웨어 민심

- 4차산업의 핵심인 소프트웨어는 대표적인 비주류 청원 주제로 표면적으로 드러나지 않는 다수 청원에 해당
 - 소프트웨어 분야 청원의 내용은 진입장벽이 존재하는 전문 영역으로의 특성을 가지고 있어 국민 전체가 공감하기 어려운 측면 존재

1) 답변형태는 청와대 홈페이지 내 답변 창구를 통해 월별 취합 후 동영상 형태로 제공하고 있음

2) 본고의 분석 기간(2017.10.01. ~ 2019.09.30., 2년) 동안 만료된 청원의 월별 평균치

3) 중앙·지역지 54곳을 대상으로 2019년 1월 ~10월 내 국민청원 게시글을 인용 또는 주제로 활용한 기사 수는 총 6,938건(한국언론진흥재단 ‘빅카인즈’ 플랫폼에 등록 된 언론사 전체를 기준함)

4) 출처 : Google scholar

5) 2018년 기준, 국민청원 전체 청원 수 대비 공개답변과 관련 된 키워드가 포함 된 청원의 비중으로 6~10% 수준일 것으로 추정

- 일반 국민의 폭발적 관심을 얻은 가상화폐(2018년 1월) 및 카풀(2019년, 현재진행중) 이슈가 소프트웨어와 연관하다 볼 수 있으나, 실제 이에 부합하는 청원 다수는 소프트웨어와 관련성이 낮은 경향이 존재
- 소프트웨어 분야 청원의 동향을 한눈에 파악해 볼 수 있는 방법이 부재
- 청원 게시글은 총 17개 분야로 분류·운영되고 있으나 정보통신 및 소프트웨어 분야를 별도 분류하지 않기 때문에 손쉽게 동향을 파악하기 어려운 구조

〈표 2〉 국민청원 사이트의 게시글 분류 기준

행정	육아/교육	반려동물	교통/건축/국토
정치개혁	외교/통일/국방	미래	경제민주화
저출산/고령화대책	안전/환경	문화/예술/체육/언론	
일자리	성장동력	농산어촌	
인권/성평등	보건복지	기타	

* 출처 : 청와대 국민청원

□ 그러나 소프트웨어 정책에 대한 국민의 수요를 파악하는 측면에서,
소프트웨어 분야의 민의 파악은 중요한 과제

- 본 연구는 아래와 같은 두 가지 목적에 초점을 두었음
 - ① 전체 국민 청원 중 소프트웨어 분야 청원을 세부주제에 맞게 분류하고, 이를 분석하여 주요 청원 사례를 소개
 - ② 소프트웨어 신기술들에 관한 국민들의 관심사항을 정량화된 척도로 제시하고, 제시된 결과의 원인을 진단
- 위 결과 도출을 위해 소셜네트워크 분석(이하 SNA) 기법 및 그래프 이론을 주요 방법론으로 채택

2. 연구 방법

□ 분석 개요

- (기반 데이터) ‘17년 10월 ~ ‘19년 9월내 답변 기한이 만료된 국민 청원 게시글 총 431,842건
- (분석 목표) 자동화 분류 및 네트워크 분석 기법을 활용하여 SW분야 청원 게시글을 일목요연하게 요약하고 SW기술에 대한 국민 인식을 정량화

□ 분석 절차

1) 약 43만 건의 국민청원 게시글에서 SW분야의 청원을 추출

- (형태소 분석) 각각의 청원에 등장하는 단어의 형태소를 식별하고 각 문서를 명사형 키워드의 집합으로 변경
 - ※ 변경 예 : [데이터는 21세기의 원유] → [데이터, 21세기, 원유]
- (SW분야 청원 식별) IT어휘사전에 수록된 단어와 각각의 청원글 간의 유사도를 측정해 유사도가 높은 청원 2,050건을 SW분야 청원으로 판단
 - ※ Term Frequency-Inverse Document Frequency(TF-IDF)⁶⁾ 기법을 활용해 청원 게시글 각각을 단어의 벡터(vector)로 변환 후 유사도 측정에 활용
 - ※ 세부적인 유사도 측정 기준 및 판단 근거는 [부록] 참조

2) 그래프 클러스터링 기법을 활용해 SW분야 청원의 6개 주제 도출

- (키워드 분류) Modularity Maximization(MM) 알고리즘⁷⁾에 기반하여 SW청원에 등장한 키워드를 유사한 맥락에서 쓰인 키워드의 집합(이하 클러스터)으로 분류
- (주제별 주요 청원 사례 도출) SW분야 청원들과 각각의 클러스터에 속한 키워드 간 유사도를 측정하여 청원 주제별 주요 사례 도출
 - ※ 키워드 유사도 측정은 SW분야 청원 식별을 위해 활용 된 방법과 동일

6) 문서 내 출현 단어의 중요도를 측정하는 기법 중 하나로, 단어의 중요성을 판단하기 위해 개별 문서에서의 단어별 출현 빈도뿐만 아니라, 여러 문서에서 범용적으로 쓰이는 단어인지 여부를 함께 고려하는 특성을 가짐

7) 그래프 클러스터링 기법 중 하나로 분류 가능한 모든 클러스터링 결과 중 모듈성(Modularity)을 극대화할 수 있는 결과를 측정하는 알고리즘

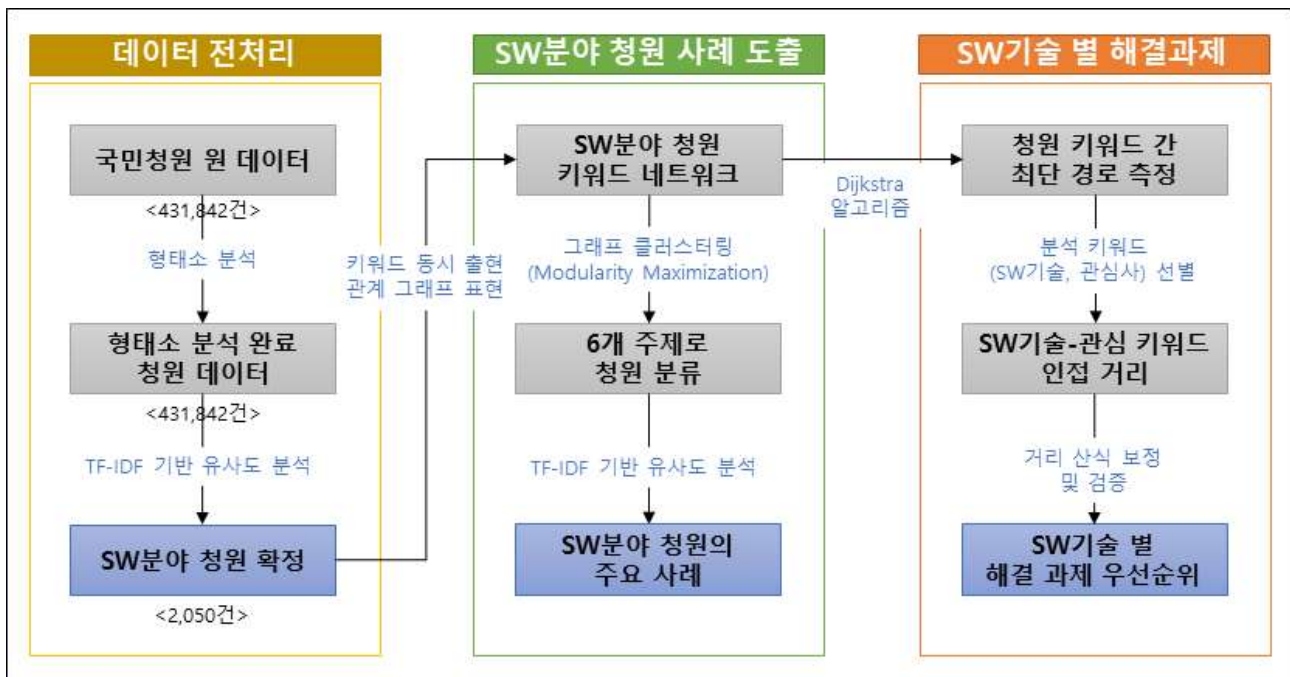
3) 주제별 주요 SW분야 청원 사례 검토 및 요약

- (실제 청원사례 검토) 청원 주제별 주요 사례를 직접 검토하여 유의미한 청원 사례를 발굴
 - ※ 검토한 청원은 2)번 절차를 통해 측정된 각 주제별 유사도가 높은 실제 청원 상위 50건을 대상으로 함
- (주제별 청원내용 요약) 상기 검토 결과를 토대로 유의미한 청원 내용 및 유사 안전을 다루는 다수 청원을 요약

4) 국민이 생각하는 SW기술별 관심 사안 우선순위 정량화

- (키워드 간 최단경로 측정) SW분야 청원에 등장하는 키워드 간 최단 경로를 측정
 - ※ 최단경로 측정 기법으로는 다익스트라(dijkstra) 알고리즘⁸⁾을 활용
- (SW기술-관심 키워드 간 거리 산출) 전문가 합의를 통해 SW기술 키워드 및 국민의 관심사안인 키워드를 선별하고 키워드 간 상대적인 연관성을 측정하여 비교

[그림 3] 국민청원 데이터 분석 절차



8) 정점(vertex)과 간선(edge)로 이루어진 관계그래프망에서 특정 정점에서 정점으로 도달 가능한 모든 경로 중 가장 짧은 경로를 탐색하는 알고리즘

□ 활용 도구

- Python 개발환경을 바탕으로 오픈소스를 도입해 데이터 분석의 기반이 되는 전처리 및 네트워크 분석 등을 수행

〈표 3〉 분석 활용 도구

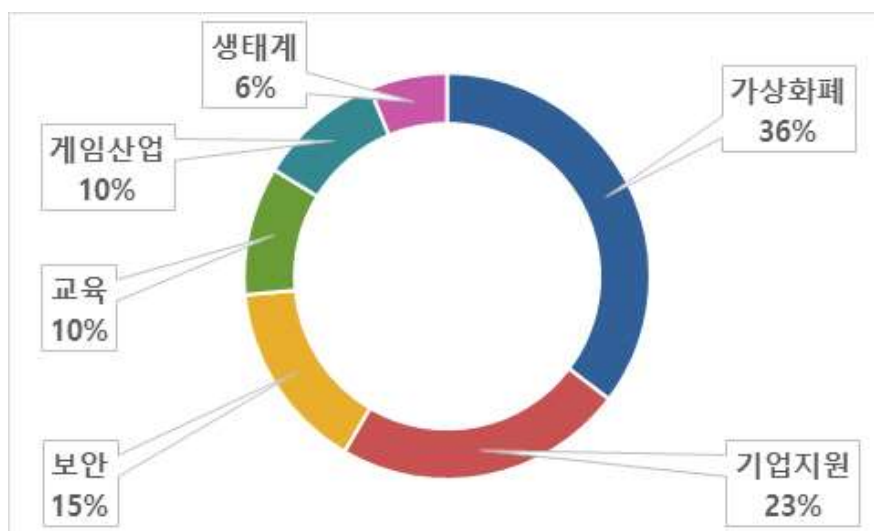
구분	명칭	비고
국민청원 수집	- lovit-petition_scraper	- 오픈소스
어휘 사전 구축	- TTA 정보통신용어사전	- 해당 사전을 활용하여 전문가 검토를 통해 SW분야 어휘 선별
형태소 분석	- Konlpy 0.5.1 - customized-konlpy 0.0.64	- Konlpy에 내장된 트위터 형태소 분석기에 기반하여 C-konlpy의 라이브러리 활용
분석·시각화	- Gephi 0.9.2 - Python 3.7	

3. 소프트웨어 분야 청원 사례 분석

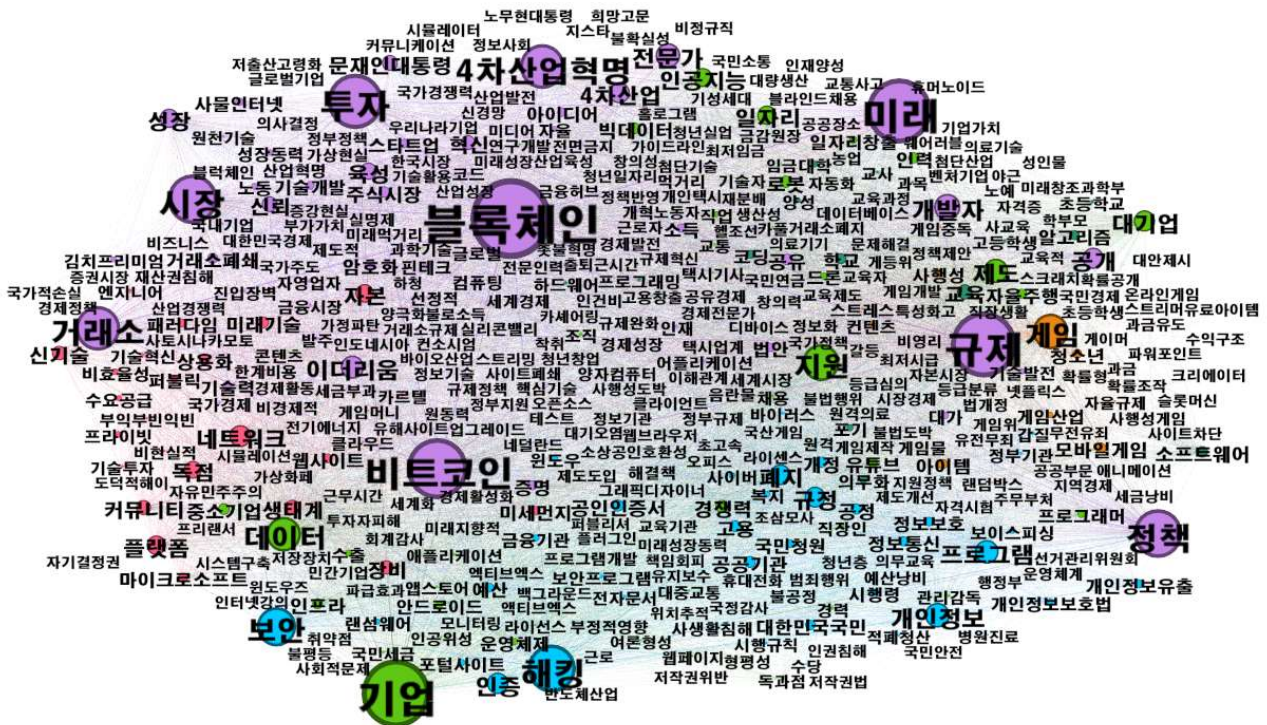
가. 소프트웨어 청원은 6개의 대주제로 구분

- 청원 전체에서 가장 규모가 큰 주제는 사회적인 파급효과가 크게 작용한 가상화폐 이슈(36%)였으며, 기업의 지원과 관련한 주제가 2위(23%), 보안, 교육, 게임산업, 생태계 관련 주제가 그 뒤를 따르는 것으로 나타남

[그림 4] SW관련 청원 내 6개 주제 비중



[그림 5] SW분야 국민청원 네트워크



주1) 동일 주제에 해당하는 키워드는 동일 색상으로 나타나 있음

주2) 원의 크기는 국민청원 네트워크 전체에서 해당 키워드가 얼마나 빈번히 언급되는지를 의미함

- 각 주제별 키워드는 근접 중앙성(Closeness Centrality)⁹⁾이 높은 상위 5개 어휘이며 해당 수치가 높을수록 해당 군집(Cluster)에 해당하는 청원 게시글에서 등장 빈도가 높은 키워드로 판단 가능

〈표 4〉 SW분야 국민청원 클러스터링 결과 및 주제별 핵심 키워드

가상화폐		보안		게임산업		교육		기업지원		생태계	
키워드	인접 중앙성	키워드	인접 중앙성	키워드	인접 중앙성	키워드	인접 중앙성	키워드	인접 중앙성	키워드	인접 중앙성
블록체인	84.9%	개인정보	85.9%	게임	91.7%	교육	87.2%	기업	80.5%	독점	92.9%
규제	80.2%	해킹	83.8%	아이템	73.3%	학교	77.3%	지원	69.6%	자본	92.9%
비트코인	78.6%	보안	81.7%	청소년	68.8%	소프트웨어	75.6%	데이터	69.1%	플랫폼	92.9%
투자	77.5%	프로그램	79.8%	모바일 게임	67.7%	코딩	73.9%	경쟁력	66.5%	네트워크	89.7%
미래	77.5%	인증	76.1%	사행성	67.7%	프로그래밍	69.4%	일자리	64.4%	생태계	89.7%

9) 클러스터내 주요한 키워드를 의미하는 지표 중 하나로, 연결 그래프에서 정점과 다른 모든 정점 사이의 최단경로길이의 역수로 계산

나. 6대 주제별 청원 사례¹⁰⁾

(1) 가상 화폐

“가상화폐를 안하는 대다수의 사람들을 지키기 위해선 규제가 반드시 필요하고, 결단을 내리는게 더 큰 피해를 막을 수 있다고 생각합니다.”

“가상화폐 공개 (ICO)는 완전히 불법화 돼 있습니다. 국민들의 안전을 최우선으로 하는 것은 좋으나, 너무 험한 규제는 오히려 독이 될 수 있다 생각합니다.”

“블록체인과 가상화폐는 분리될 수 없고 ICO는 미래 산업 준비에 필요한 혁신적이고 새로운 창업 자금 조달 창구이다. 다만 조세회피, 자금은닉, 불법 자금 거래 등에 대한 문제를 잘 해결 할 수 있으면 된다.”

□ (요약) 개인 투자자의 가상화폐 규제 문제와 블록체인 기술 및 산업 발전에 대한 우려가 공통적으로 나타남

① 규제의 필요성 강조

■ 비정상적 투기는 사회적 부작용 야기

- 블록체인 기술의 발전을 고려하기 이전에 가상화폐로 인한 비정상적인 투기 현상은 사회적 부작용을 야기할 것이라는 의견 존재
- 보편화 된 기술은 모두 적절한 규제 테두리 내에서 사회적 비용을 발생시키지 않는 적정선을 준수하고 있으므로, 가상화폐도 예외가 될 수 없음을 주장

② 규제로 인한 산업발전 저해에 관한 우려

■ 블록체인 기술과 가상화폐의 연관성

- 블록체인 기술은 가상화폐 개념과 별도로 분리할 수 없음
- 가상화폐 공개(ICO)를 통해 자체발행 화폐에 대한 가치 인정과 투자유치가 가능해야만 블록체인 산업이 발전할 수 있다고 제언

10) 본 장에서 언급되는 청원 사례는 주제별 실제 청원 사례들 중, 유의미한 청원으로 판단되는 경우에만 한하여 내용을 소개 및 요약한 것으로, 본고에서 다룬 청원 사례가 SW분야 청원 전체를 대표하는 것이 아님에 유의

(2) 기업 지원

“벤처업계는 CVC가 설립돼 대기업 자금이 벤처 생태계로 흘러들어야 혁신 기업 창업이 활성화 된다고 강조해 왔습니다. 신생 벤처 기업 창업이 활성화 된 미국은 대기업이 CVC를 운영하고 있습니다. 구글은 CVC를 통해 로봇, 무인항공기, 자율주행 분야 등에 투자하고 있습니다.”

“공공 데이터 신청 절차와 제공절차가 복잡하고, 신청부터 제공까지 기간이 최대 20일까지 소요되어, 일반인이나 기업들이 발 빠르게 필요한 데이터를 신청하고 제공받기에 무리가 있습니다.”

□ (요약) 기업 육성 및 운영을 위한 환경 조성 및 공공데이터 활용 개선 등으로 구성

① 기업 투자 여건 조성

■ 기업형 벤처캐피탈(CVC)¹¹⁾ 설립

- 대기업 자금이 벤처 생태계로 유입되기 위해서 현행 벤처지주회사¹²⁾가 아닌 기업형 벤처캐피탈 설립이 필요하다는 주장
- 미 대기업인 Google, GE, Intel 등의 경우 기업 자체 CVC운영을 통해 4차 산업 관련 기술 분야에 투자 중

[관련 해설]

- 금산분리 원칙으로 인해 국내는 CVC 도입이 제한되나, 2018년 8월 공정거래위원회 주도로 벤처지주회사 설립 문턱을 낮추는 형태의 대안이 마련됨
- 벤처기업은 여전히 CVC를 선호하고 있으나, 대기업의 벤처 투자가 미진한 근본적 원인이 벤처지주회사에 있는지 관해서는 다양한 시각 존재

② 공공데이터 활용 개선

■ 공공데이터 신청 기간 단축

- 공공 데이터를 신청할 시 데이터 제공까지의 행정소요기간이 급한 활용이 필요한 기업에게 불편한 요소라 지적
- Open API로만 제공받을 수 있는 데이터의 경우, 프로그래밍 능력이 없어도 제공 받을 수 있도록 제공 형태를 다양화 할 필요가 있음

11) 대기업이 출자한 벤처캐피탈(VC)로 금전적 지원 이외 모기업의 인프라를 활용할 수 있는 이점을 가지기 때문에 벤처업계가 선호하는 경향이 있음

12) 벤처기업 지분 소유 비중이 전체의 절반을 넘는 지주회사로, 금산분리로 인해 CVC를 운영할 수 없는 국내 환경에서 유망 벤처기업을 지원하여 대기업의 인수를 유도하기 위한 대안으로 도입된 제도

(3) 보안

“주택의 보안은 거주자에게 책임이 있듯이 공공기관의 보안과 기업들의 보안은 누가 책임지는 겁니까? 본인들 스스로 책임져야 하는 거 아닙니까. 근대 현실은 본인들의 보안유지를 사용자에게 떠넘기고 있습니다.”

“개인정보가 너무나도 쉽게 유출되는 것 같습니다. 이를 악용하여 돈벌이로 생각하는 몰상식한 인간들에 대한 수사 및 처벌을 강화하여 주세요.”

“내가 개인정보를 제공 동의한 기업과 그 기업에서 어떻게 사용되고 관리되는지 확인 가능했으면 좋겠습니다. 내가 가입한 사이트를 한곳에서 조회 가능한 방식으로요.”

□ (요약) 정부차원에서 추진 중인 공인인증서 폐지와 공공기관 사이트의 ActiveX 및 플러그인 제거, 개인정보 보호법의 강화 등으로 구성

① 공인인증서 · ActiveX 폐기

■ 보안을 위한 불편함 및 비용을 사용자에게 전가

- 공공기관 · 은행 홈페이지 공인인증서 및 ActiveX가 사용자의 불편함을 야기
- 보안의 주체는 사용자가 아님에도, 기업 편의에 맞춰진 ActiveX 설치 요구와 사용자가 비용을 부담해야만 하는 유료 공인인증서에 대한 이용자들의 반감

■ 본인인증 수단의 보편성 제고

- 스마트폰 사용에 제약이 있는 연령층(저연령층, 노년층)이 손쉽게 활용 가능한 본인인증 수단이 부족
- 타인 명의의 스마트폰을 활용하는 경우 대체가능한 인증수단이 유료 공인인증서에 국한되는 상황이 발생

② 개인정보보호법 강화 및 시행

■ 개인정보 유출기업의 제재 강화

- 국민의 개인정보를 기업에서 무분별하게 유통 또는 남용하는 사안에 대해 실효성 있는 대책이 필요
- 개인정보 유출 피해로 인한 국민-기업간 소송에서 국민이 불리하다고 인식
- 기업의 각종 보안시설 정비 및 관리인원을 적정수준 배치하는 등 개인정보 유출에 대한 강력한 대비 및 처벌의 필요성을 강조

③ 개인정보 활용 동의처 관리의 편의성 제고

■ 민간기업·공공 통합 개인정보 활용 현황 조회 플랫폼 요구

- 공적기관 및 은행에서 개인정보를 열람할 시 국민이 그 사실을 즉시 통보받을 수 있는 시스템 도입이 필요
- 개인정보를 제공 동의한 기업과 기업의 개인정보 관리 현황을 한 곳에서 조회 가능한 홈페이지 필요

(4) 교육

“4차 산업혁명으로 인공지능 및 로봇에게 밀려 갈 곳을 잃는 대한민국의 국민이 아닌 전 세계적으로 우수함을 떨치는 민족으로 다시 한 번 거듭났으면 합니다.”

“코딩교육은 장점만이 아닌 단점도 많이 포함하고 있습니다. 기존 시수에 1시간 추가되는 것이라 보니 여느 과목과 다를 바 없이 암기과목이 된다는 것입니다. 학생들에게는 그저 과목 1개가 더 늘어나는 꼴이다 보니 여러 학부모들도 크게 실망하고 있습니다.”

“담론으로만 4차 산업을 이야기할 것이 아니라 구체적으로 산업과 연계된 전공교육을 통해 4차 산업 혁명을 대비해야 한다는 생각이 많이 듭니다.”

“정보 소외 계층이 세상의 변화에 더 이상 뒤처지지 않도록 이들을 대상으로 한 지원이 간절히 필요합니다. 각종 디지털 기기 사용법을 익히고 첨단 기술을 체험할 기회를 제공하는, 강화된 정보 교육을 지역 문화센터 등에서 실시해 주십시오.”

□ (요약) 신기술로 인한 일자리 위협, 코딩 교육의 문제점, 4차 산업 관련 체계적 교육 수요 등으로 구성

① 미래 일자리에 대한 대책

■ 로봇·인공지능에 의한 양극화 우려

- 신기술이 기존 일자리를 대체하여 빈곤층의 생계에 위협으로 작용할 것
- 4차 산업혁명으로 인해 직업을 잃을 것으로 예상하며 살아남기 위해 무엇을 해야 하는지 고민

■ 선제적인 신기술의 역할 부여가 필요

- 미래 인간의 노동 역할과 기술로서 대체될 역할에 관하여 합리적인 선에서 국민의 불안감을 해소할 수 있는 법적 준비에 대한 요구

② 4차 산업 관련 교육 개선

■ 청소년 코딩교육 의무화

- 현재 시행중인 코딩교육은 일주일에 1시간정도의 시수로 운영되어 암기과목으로 변질될 우려를 제기
- 기존 교과에 시수에 추가되는 부담요소로 비춰지고 있음
- 아직까지 와이파이 등 기초적인 인프라 구축이 안 된 학교가 많음

■ 실효성 있는 4차 산업 학과 개설

- 4차 산업과 관련 된 학과가 대학에 거의 없는 것으로 보이며, 4차 산업이 중요하다라는 주장이 거대 담론으로만 비춰져 체감할 수가 없다는 의견 제기
- 산업과 직접적으로 연계되어 취업 가능한 4차 산업 관련 학과의 개설이 필요하며 이에 따른 전공교육 필요

■ 저소득층·중장년층의 기술 소외 대책

- 각종 디지털 기기 사용법 및 첨단기술 체험의 기회 등을 지역 문화센터를 통해 실시 확대해주는 것이 필요함

(5) 게임 산업

“랜덤박스란 원하는 아이템을 사기 위해 확률적으로 더 많은 돈을 지출하는 구조를 가지고 있습니다. 부디 사행성 게임에 대한 철저한 규제를 통해 기업들이 올바른 방향으로 나아갈 수 있도록 감시해 주십시오.”

“저는 인디게임 개발자입니다. 최근에 게임 심의에 대한 문제인식이 있어서 법 개정도 이루어지고 법 규정도 많이 완화되었습니다. 그중 가장 큰 성과가 '자율심의'입니다. 기업이 정부에서 허가를 받으면 자체적으로 게임을 심의 할 수 있습니다. 하지만 저 같은 소규모 개발자들은 그런 인프라를 마련한다는 것은 말도 안 되기 때문에 게임등급위원회에서 심의를 받아야 합니다. 높은 심의 비용과 여러 가지 서류절차 등 그 번거로움이 말로 다할 수 없습니다.”

“모바일 첫다운제로 확대는 해서는 안 됩니다. 우리나라 청소년들의 작은 행복마저 빼앗아 버리는 아주 잔인한 짓이라고요. 우리나라 대부분의 청소년들이 학원을 끝내고 오면 10시입니다. 몸 깨끗하게 씻는 데에 30-40분을 소요, 이제 학원숙제와 학교 수행평가를 포함한 과제물 준비까지 마치면 이미 밤 12시가 넘어갑니다.”

□ (요약) 사행성을 조장하는 게임 콘텐츠 규제 및 게임 내 아이템의 가치 보존, 게임 심의제도의 완화 등으로 구성

① 확률형 아이템¹³⁾ 규제

■ 게임내 확률형 아이템은 합법적 도박

- 청소년이 즐길 수 있는 12세 이상 등급판정을 받은 게임에서 확률형 아이템을 도입하는 것은 자제력이 부족한 미성년자를 고려하지 않는 처사
- 확률적으로 획득 가능한 과금 상품을 만드는 것은 복권을 판매하는 것과 동일

■ 모바일 게임의 결제 제한

- 게임의 평균 생존주기가 짧아 과금의 결과에 관한 충분한 가치 보존이 불가능
- 게임 결제액의 제한을 두어 불가피한 피해 사례 및 무분별한 과금을 제한할 필요

② 게임등급 심의 규정 완화

■ 소규모 게임 개발사에게 불리한 국내 심의 요건

- 게임을 이용하게 할 목적으로 배포하는 자는 반드시 게임물관리위원회(이하 계등위)의 심의를 받게 되어있음

[관련 해설]

- 심사를 통해 자체등급분류사업자로 지정된 기업의 경우 자율심의가 가능
- 스마트폰 OS를 독점하는 애플, 구글이 자율심의가 가능해짐에 따라 모바일 게임의 심의규제는 일정수준 해소되었으나, PC기반 게임을 개발하는 경우 여전히 문제로 지적되는 상황임

- 게임물의 심의 비용이 높고 복잡해 자본이 부족한 인디 게임 제작사에게는 심의 비용이 부담으로 작용된다는 의견 존재
- 인디 게임 제작사가 모바일게임 개발을 선호하는 주요 원인 중 하나로 지적

■ 비영리 목적의 게임물 사전심의는 부담

- 비영리 목적으로 제작 된 게임의 경우도 사전 심의의 대상으로 개인 제작자 또는 소규모 제작사가 포트폴리오를 구성하기 어렵다는 지적이 있음

- 「게임산업 진흥에 관한 법률 시행령」 개정안을 통해 영리를 목적으로 하지 않는 비영리 게임에 대한 등급분류 심의가 면제(2019.09.03. 시행)

13) 인게임 환경에서 구매 가능한 과금을 통해 획득 가능한 상품으로, 게임 내 상품 가치가 확률적으로 변경되는 형태의 아이템(예 : 랜덤박스)

③ 기타의견

■ 모바일 게임 셧다운제에 대한 반대 의견

- (요약) 청소년의 모바일 게임 이용을 특정 시간 이후 강제적으로 차단하는 제도에 대한 반감

◦ 해당사안에 대해 여성가족부는 모바일 게임 셧다운제를 검토하지 않는 것으로 답변¹⁴⁾

■ 여성혐오 또는 소수자혐오를 조장하는 게임의 규제 기준 강화

- (요약) 여성을 남성 권력의 상징으로 표현하는 선정적 광고를 통해 인기를 구가하는 특정 모바일 게임이 성행하고 있어 이에 대한 우려 표명
- 관련 선정적 광고 행위는 법적 제재로 개선된 것으로 비춰지나, 게임물 자체에 대한 반감 여론은 여전히 존재

(6) 생태계

“미래 기술 투자 많이 해서 성장 동력으로 삼아주세요. 과학 기술 투자로 미국처럼 성장할 수 있겠어요. …(중략)… 스타트업 창업 센터도 20개로 늘려 주세요. 미래기술을 이용해서 일자리 문제도 해결 가능해요.”

“VR은 하드웨어보다 사실 콘텐츠, 소프트웨어가 매우 중요한 변수입니다. 소위 킬러 앱으로 불리우는 이런 소프트웨어들이 플랫폼을 견인하게 되는데 이런 킬러앱을 개발하기 위해서는 엄청난 비용과 기술력이 필요합니다.”

“우리나라도 뒤쳐지지 않게 양자컴퓨터 개발에 능력 있는 개발자들에게 안정적인 투자 지원 해주어 미래 산업에 큰 발판이 될 수 있도록 도와주세요!!! 늦지 않았습니다. 우리나라 유망한 개발자들이 외국으로 유입 되지 않도록 든든한 지원 바랍니다.”

□ (요약) 국내 신기술 시장 활성화에 대한 염원과 VR·양자컴퓨팅 등의 기술에 대한 투자 필요성 등으로 구성

① 신기술 시장

14) “모바일 게임 셧다운제 적용 검토 안해”, 여성가족부(2019.10.24.)

■ 신기술 시장 활성화를 위한 바람

- 기술 관련 규제 완화를 통해 미래 기술을 국가의 성장 동력으로 삼아야함
- 스타트업 창업 센터 확충이 필요

■ VR하드웨어 → 소프트웨어로 투자 전략 선회

- 외산 VR기기에 기반한 독점 플랫폼이 시장을 잠식할 가능성이 높다는 의견 존재
- VR과 관련 된 투자를 소프트웨어에 집중하여 어플리케이션 경쟁력을 키워야 한다는 주장

■ 양자컴퓨터 R&D 예산 확대

- 장기적 안목에서 양자컴퓨터와 관련 된 R&D 예산을 확대하는 것은 당장 성과가 나오지 않아도 국가발전을 위해 국민이 납득 할 수 있을 것

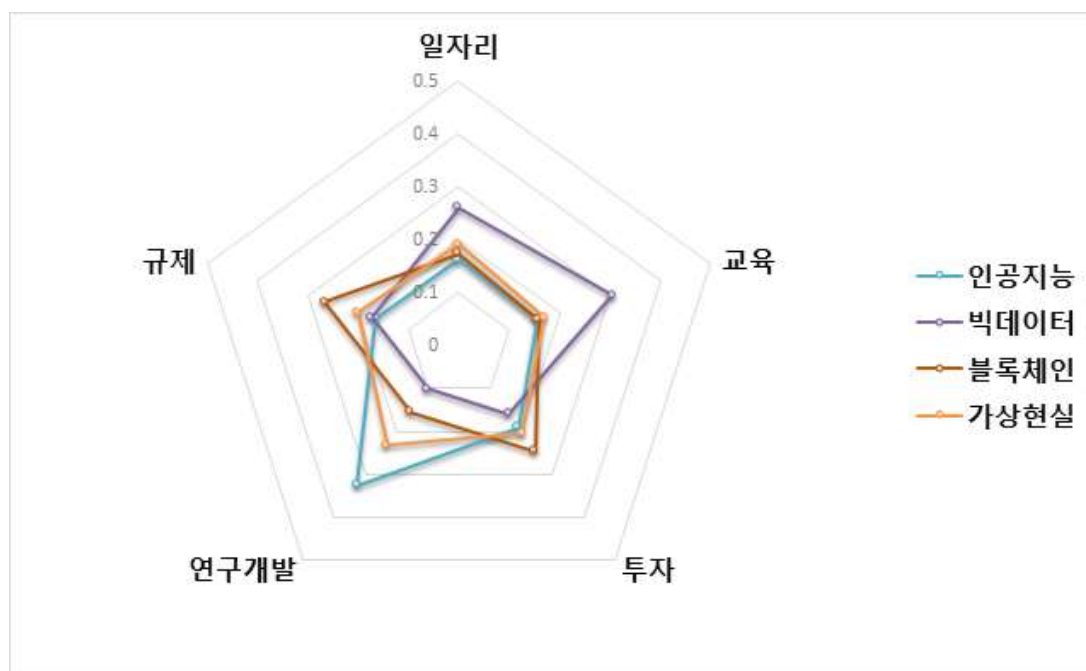
4. 소프트웨어 기술에 대한 국민의 인식

□ 키워드 간 인접 거리(Distance) 분석

- (분석 목적) 청원에서 드러나는 소프트웨어 기술에 대한 관심사안을 상대적 연관성으로 표현해 국민 관점의 정책적 우선순위 도출
- (분석 방법) 키워드 네트워크의 최단경로(Shortest Path) 계산에 기반해 기술 키워드와 관심 키워드 사이의 상대적 연관성을 계산
 - 기술 키워드 : 인공지능, 데이터, 블록체인, 가상현실
 - ※ SW기술 중 고려되었던 클라우드, IoT의 경우 국민청원에서 이와 관련된 언급 빈도가 낮아 대상에서 제외
 - 관심 키워드 : 일자리, 교육, 투자, 규제, 연구개발
 - ※ 청원 플랫폼의 게시글 특성과 및 정책의 범주를 고려해 전문가 합의를 통해 키워드 도출
 - 세부적인 가중치 적용 및 산식은 [부록] 참조

□ 분석 결과

[그림 4] SW기술-관심 키워드 간 상대적 연관성(그림)



〈표 5〉 SW기술-관심 키워드 간 상대적 연관성(데이터)

구분	인공지능	빅데이터	블록체인	가상현실
일자리	16.2%	26.1%	17.4%	19.2%
교육	15.6%	30.5%	15.8%	17.1%
투자	19.0%	16.0%	24.6%	20.5%
연구개발	32.6%	10.1%	15.5%	23.2%
규제	16.5%	17.3%	26.6%	20.1%

- (인공지능) 인공지능의 경우 연구개발(32.6%)과 투자(19.0%)에 국민 관심이 높은 것으로 나타나 4차산업의 핵심기술로서 인지하고 있음을 알 수 있음
 - 4차산업의 핵심기술로서, 국내 대기업의 투자 및 정부의 인공지능 관련 정책 예고 등에 영향을 받은 결과로 해석됨
 - 높은 관심에도 불구하고, 연구개발에 관한 척도가 가장 높은 것은 현재 기술 수준이 범용적으로 활용되기에는 성숙하지 않다고 판단하기 때문으로 분석
- (빅데이터) 빅데이터는 교육(30.5%)과 일자리(26.1%)에 대한 관심이 가장 높은 것으로 나타나 실제 일자리와 연계 가능한 기술로 인지하고 있음을 시사
 - 본 결과를 통해 국민의 빅데이터 교육에 대한 수요와 일자리 증가의 기대감이 반영된 것이라 판단 가능함
 - 규제 관련 척도는 개인정보를 포함한 민감 데이터에 관한 찬반여론에 주요한 영향을 받는 것으로 추정
- (블록체인) 블록체인은 규제(26.6%)와 투자(24.6%)가 가장 높게 나타났으며 이와 같은 결과는 ICO 규제 이슈와 블록체인 기술 투자 측면의 영향으로 분석됨
 - ICO는 블록체인 기술 성장 및 투자와 관련된 청원에서 빈번히 언급되는 사안으로, 블록체인 기업 육성을 위해 고려해야할 우선적인 이슈가 될 가능성이 높음
- (가상현실) 연구개발(23.2%)과 투자(20.5%)가 높게 나타났으나 이외 관심

사안도 비교적 균형 있는 분포를 띄어, 관련 기술에 대해 뚜렷한 정책적 관심 사안이 형성되지 않았다 판단됨

- 연구개발과 투자에 관한 연관성이 상대적으로 높은 원인으로는 국내 관점에서 아직 국민이 체감 할 만큼 충분한 기술 수준과 어플리케이션 생태계가 구축되지 않은 상황이기 때문이라 추정
- 가상현실 분야는 원천 기술의 고도화에 집중하면서도 시장형성, 인재 양성, 법제도 정비 등 전반적인 환경 조성이 병행되어야 할 것임을 시사

5. 요약 및 시사점

□ (청원 사례) 소프트웨어 분야 국민청원은 크게 6가지 주제로 구분 가능하며, 주제별 아래와 같은 요구사항이 있는 것으로 파악됨

- (가상화폐) 규제(사회적 부작용)와 투자(블록체인 기술) 관점의 충돌
- (기업지원) 벤처 투자 환경 개선 및 공공데이터 접근성 강화
- (보안) 개인정보의 보안강화와 본인인증의 편의성 강화
- (교육) 미래일자리 불안감해소 및 4차 산업 관련 기술에 대한 교육 강화
- (게임산업) 확률형 아이템 강력 규제와 게임물 등급 심의 완화
- (생태계) 신기술 투자 확대와 기업하기 좋은 여건 조성

□ (SW기술-관심 키워드간 상대적 연관성) 국민들의 주요 관심사가 SW기술별로 서로 다르게 나타남을 알 수 있음

- 인공지능과 가상현실은 연구개발, 빅데이터는 교육, 블록체인은 규제에 대한 국민의 관심이 가장 높게 측정됨
- 본 결과는 정부가 SW기술에 대한 일자리, 교육, 연구개발 등의 정책 중 무엇에 집중하는게 효과적인지 선택하는데 도움을 줄 수 있을 것이며, 정책의 만족도를 높이기 위한 방안 마련에 있어 유용할 것으로 기대

□ (시사점) 일반 여론에 대응하는 SW분야 정책 의견수렴 환경이 필요

- SW분야는 학계와 산업계 의견 수렴을 위한 공청회, 업계간담회 등이 활성화 된 반면, 일반 여론을 파악하기 위한 환경은 상대적으로 부족
- 국민신문고로 대표되는 전자민원 제도를 활용해 여론을 파악해 볼 수 있겠으나, 민원 절차가 비교적 까다로워 일반 국민의 적극적 참여를 유도하기 어려운 한계가 있음

※ 국민신문고 민원 신청은 ①본인 인증, ②기본 정보 기입, ③유사 민원 확인절차 등을 거쳐야 하는 번거로움이 존재하나, 관련 부처의 직접적인 답변을 기대할 수 있는 장점이 있음

- 향후 국민을 대상으로 SW이슈를 쉽게 설명하고 의견을 수렴하기 위한 정책적 고려 필요

□ (기타 제안) 청원 답변 기준선(30일 간 20만명 동의)에 미달하는 청원들을 부처 의견수렴에 효과적으로 활용하기 위해 다음과 같은 사안 고려 가능

- 국민청원 등록 시 부처별 선택이 가능하도록 개선함으로써 정책 개발 및 여론 수렴 필요시 관련 사항들에 대해 부처가 검토하기 용이한 여건 조성
- 인공지능 기술을 활용해 중복 청원 및 이상 청원에 대한 식별을 자동화함으로써, 검토를 위한 시간적 비용을 줄이고 양질의 청원을 발굴

[부 록] 청원 데이터 분석 방법론

(1) 국민청원 데이터 수집

- 오픈소스 활용을 통한 분석 대상 데이터 웹 크롤링(Web Crawling)
- 등록된 국민청원 전수(435,652건, 10월 8일 기준) 확보 후 상기 분석 기간 내 만료 된 청원(431,842건, 중복제외) 추출
 - 분석 기간 : 2017.10.01. ~ 2019.09.30.(2년 간)
 - 수집 방법 : 개별 청원을 json 형태로 추출 후 → 파워 쿼리를 활용하여 병합

(2) 형태소 분석

- (한글 형태소 분석) 수집 데이터 중 본문 내용을 형태소로 분리
 - 형태소 분석기 : 고유명사 및 연어 수동 편집기능 등을 제공하는 라이브러리인 Customized Konlpy를 활용(Twitter 형태소 분석기 기반)
- (어휘 사전 구축) 한글 형태소 분석의 불완전성 보완과 SW분야 어휘 식별을 위해 고유 명사, 연어, 합성어 처리 용도의 어휘사전 구축
 - IT용어 사전(총 1,840건) : TTA 정보통신용어사전내 등장 어휘 10,196개 중 SW분야에 해당하는 어휘 선별
 - 고유명사 사전(총 2,481건) : 고유명사로서 형태소 분석 과정에서 인지하지 못하는 청원 데이터 내 출현 어휘
 - 합성어 사전(총 5,339건) : 관용적으로 띄어쓰기와 붙여쓰기를 혼용하는 합성어휘를 반복적인 테스트를 통해 식별
 - Stopwords 사전(총 895건) : 형태소 분석 결과 불용 조사 및 어휘를 식별 후 형태소 분석 전 제거
 - Replace_words 사전(총 16건) : 빈번하게 일어나는 오타 및 영문과 한글 표기를 혼용하는 언어에 대한 용어 통일

- (명사형 형태소 추출) 형태소 분석 결과 중 명사형 어휘를 제외한 기타 어휘 제거

(3) SW분야 청원 데이터 식별

- (자동화 식별) TF-IDF 기법을 통해 국민청원 문서(Document) 각각을 벡터화한 후, IT용어 사전에 수록된 어휘와의 유사성을 측정해 점수화(Scoring)
 - 문서 유사도 점수 계산법 : 개별 문서 벡터에서 IT용어 사전 수록 어휘에 해당하는 벡터값을 별도 추출하여 합산
 - ※ 코사인 유사도(Cosine Similarity)를 통해 도출된 결과에 비하여 나은 성능을 보이는 것으로 판단하여 채택
 - SW분야 청원 선별을 위한 임계값(Threshold) 설정 : 문서 전체의 유사도 점수를 내림차순으로 누적합산 후 3분위(상위 25%)에 해당하는 문서를 SW분야 청원으로 판단
 - ※ 해당 기준은 누적 합계 분위별 결과 검토를 거쳐 전문가 합의에 의해 결정
- (식별된 데이터) 총 431,842건의 국민청원 중 2,050건의 SW분야 청원 도출

(4) 소셜네트워크 분석(SNA)을 위한 데이터 형태 변환

- (키워드 연관성 데이터 생성) SNA를 위해 필요한 청원 내 등장 키워드 관계를 표현한 데이터 셋 추출
 - 변환 예시 : 청원 게시글에 「빅데이터」, 「규제철폐」, 「산업」 총 3가지 키워드가 등장할 시, 키워드 연관성을 아래와 같이 표현
[빅데이터-규제철폐, 규제철폐-산업, 빅데이터-산업]
 - 해당 분석에서 각 등장 키워드는 그래프의 정점(vertex)이며, 키워드 간 관계는 간선(edge)으로 정의
- (키워드간 관계 가중치 설정) 청원 각각을 위와 같이 변환 후, 동일 키워드 관계가 등장하는 빈도에 따라 키워드 간 관계의 강도(weight)를 설정
 - 예시 : 5개의 청원글에서 모두 [빅데이터-규제철폐]란 키워드 관계가 등장한다면, [빅데이터-규제철폐]의 연결 강도를 5로 판단

(5) 청원 키워드 클러스터링

- (키워드 클러스터링) SNA에서 범용적으로 활용하는 알고리즘 중, 잘 알려진 알고리즘인 ‘Modularity Maximization’ 메소드¹⁵⁾를 통해 빈번히 함께 등장하는 키워드간의 군집을 도출
- (주제 도출) 클러스터링을 통해 도출 된 각각의 클러스터를 검토하여 각 클러스터가 어떤 청원 주제의 집합을 의미하는지 추정
- (주제별 대표 청원 소개) 클러스터별 키워드 분포와 등장 키워드의 유사성이 높은 상위 50건의 실제 청원을 추출하고 관련 결과를 검토하여 내용 요약
 - 유사도 측정 지표 : (3)번 과정의 문서 유사도 측정 방법을 적용

(6) SW기술-관심 키워드 상대적 연관성 측정

- (측정 목표) SW기술 키워드(예 : AI, 빅데이터, 가상현실, 블록체인)와 관심 키워드(예 : 규제, 투자, 일자리 등)간 인접 척도를 계산
 - 관련 측정을 위해서는 키워드 간 최단거리뿐만 아니라, SW기술 키워드가 동일 청원에 2개 이상 동시 등장하는 경우를 고려해 인접척도를 보정하는 방안이 함께 강구되어야 함
 - 인접 척도 계산을 위해 고려한 지표는 아래 3가지임
 - ① 측정 대상 키워드간 최단거리 : 기본 인접 척도
 - ② SW기술 키워드간 최단거리 : SW기술 키워드가 동일 청원글에 등장하는 정도를 측정
 - ③ 측정대상 외 SW기술과 측정대상인 관심 키워드간 최단거리 : ②의 결과에 기반해 측정대상의 인접척도를 보정하기 위한 수치

15) 그래프 클러스터링 알고리즘 중 하나로, 분류 된 클러스터 간의 낮은 응집도(Cohesion)과 클러스터내 높은 커플링(Coupling)을 만족하는 최적화 된 클러스터 결과를 탐색

○ SW기술-관심 키워드 간 거리 계산식

$$\left[sp(t, i) \times \sum_{w_1}^{W(t \notin w)} \frac{1}{sp(t, w_n) \cdot sp(w_n, i)} \right]^{-1}$$

- sp = 최단경로(Shortest Path)¹⁶⁾
- t = 측정 대상 SW기술
- i = 관심 키워드
- W = t 를 제외한 SW기술의 집합

- ①번 지표를 기본 수치로 하여, ②번 지표의 수치만큼 ③번 지표에 반비례하여 값을 보정
- 최단경로는 값이 낮을수록 긍정적 지표이므로, 마지막에 역수를 취함

○ SW기술-관심 키워드 간 상대적 연관성

- 위 계산식에 의거 도출된 기술별 관심 키워드 간 거리정보를 백분율로 환산하여 상대적인 연관성 비중 측정

16) 그래프에서 최단경로란, 특정 정점에서 정점까지 도달하기 위해 거쳐야 할 간선의 최솟값을 의미. 최단 경로는 1개 이상일 수 있음.

[참고문헌]

1. 국내문헌

- 이동현, 이호, 김정민. (2016). 그래프 이론 및 네트워크 모델을 이용한 지식경영연구 논문 트렌드 분석.
- 이한동, & 김종배. (2017). 문서 유사도 기법을 활용한 이슈 키워드 추출 방법-인터넷 뉴스 기사를 대상으로.
- 김진수, 황종선, & 조영임. (1993). 무향 그래프 공간에서의 모든 정점간의 최단경로 알고리즘에 관한 연구.

2. 국외문헌

- Okamoto, K., Chen, W., & Li, X. Y. (2008.6). Ranking of closeness centrality for large-scale social networks.
- Newman, M. E. (2004). Fast algorithm for detecting community structure in networks.
- Good, B. H., De Montjoye, Y. A., & Clauset, A. (2010). Performance of modularity maximization.
- Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries.
- Achananuparp, P., Hu, X., & Shen, X. (2008, September). The evaluation of sentence similarity measures.

3. 기타

- 존 스콧, “소셜 네트워크 분석”, 커뮤니케이션북스, 2012.
- 청와대 국민청원 홈페이지. <https://www1.president.go.kr/petitions>
- 한국언론재단 빅카인즈. <https://www.bigkinds.or.kr/>
- 구글 학술검색. <https://scholar.google.co.kr/>
- 한국정보통신기술협회 정보통신용어사전. <http://terms.tta.or.kr/dictionary/>

주 의

1. 이 보고서는 소프트웨어정책연구소에서 수행한 연구보고서입니다.
2. 이 보고서의 내용을 발표할 때에는 반드시 소프트웨어정책연구소에서 수행한 연구결과임을 밝혀야 합니다.