

국회 입법활동 빅데이터 시각화를 통한 인사이트 발굴

최종 보고서

- 캡스톤 디자인 5조 -

201811724 박건현

201611883 문승욱

201611898 윤태균

201611899 이규석

201710002 권지희

- 목차 -

1. 주제 소개 및 진행 방식

1-1. 주제 소개

1-2. 기대효과

2. 과제 설명

2-1. 과제 내용 요약

2-2. 국회 입법활동 빅데이터 기초분석

2-3. 구체적 데이터 분석

2-4. 홈페이지 디자인 및 콘텐츠 내용

3. 결론

3-1. 기대효과 및 결론

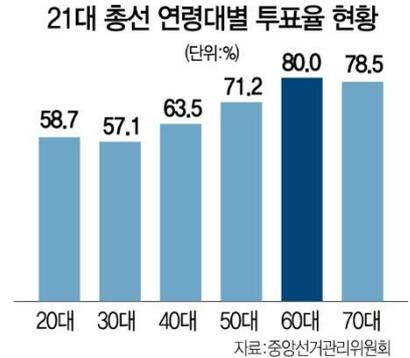
3-2. 한계점 및 시사점

4. 출처 및 참고자료

1. 주제

① 주제 소개

사회초년생으로서 정치에 대해 잘 이해하고 참여하여 더 나은 사회를 만들기 위해서는 민주정치에 참여도가 높아야 한다. 하지만 청년들의 정치활동 참여는 민주정치사회에서 매우 중요한 일이지만 정치에 대해 무지하고 관심이 없는 경우가 많다. 그리고 현재 나와 있는 정치 관련 자료들은 진입장벽이 높고 주관적 생각이 많이 들어간 경우가 많으며 시각화한 자료가 부족해 처음 정치에 관심을 가지는 이들이 쉽게 다가가기 어렵다. 실제로 투표 현황을 보면 20대의 투표 비율이 5-60대에 비해 확연히 적은 것을 알 수 있다.



그리고 다양한 정치 관련 사이트 중 정당별로 객관적인 자료를 제공하는 사이트는 찾기 어렵다. 특히, 유튜브나 보편적으로 많이 사용하는 SNS와 같은 곳에서도 작성자의 주관적인 성향이 들어간 경우가 많고 심지어 기사도 주관성이 들어간 경우가 많다. 객관적인 자료를 제대로 보지 못하고 주관적인 내용이 많이 들어간 자료를 보다 보면 정확한 정보를 알지 못해 정치와 관련된 정보를 제대로 이해할 수 없다. 또한, 정당에는 수많은 인물들이 있고 정당이 계속 변화하기 때문에 더욱 이해하기 어려운 부분도 있다. 그래서 우리는 정치와 관련된 객관적인 정보를 진입장벽이 낮은 방식으로 제공하여 정치에 무지한 사람들이 정치에 대한 개념을 잡을 수 있는 사이트를 개발하기로 하였다.

정치활동은 입법, 경제, 보건 등 우리의 삶에 많은 영향을 주고 국민들이 참여할 수 있는 분야도 다양하게 존재한다. 그 중 입법이라는 활동이 국민들의 삶에 밀접한 관계를 가지고 가장 중요한 활동이라고 할 수 있다. 또한, 입법 활동은 국민청원 등 일반인들도 쉽게 진입할 수 있는 방법이 있어 쉽게 마주할 수 있는 활동이다. 하지만 입법관련사이트들은 시각화 자료가 부족하고 정리가 되어있지 않아 한눈에 알아볼 수 없고 정당별로 비교한 내용은 존재하지 않는다. 이러한 이유 때문에 국회 입법 활동을 중심으로 전체적인 국회 활동, 정당 별 입법 활동 등을 분석하여 시각화와 구체화를 진행하였다. 이를 통해 청년들이 정치 관련 정보를 한 눈에 알아볼 수 있게 하는 인사이트를 발굴하기로 했다.



② 기대효과

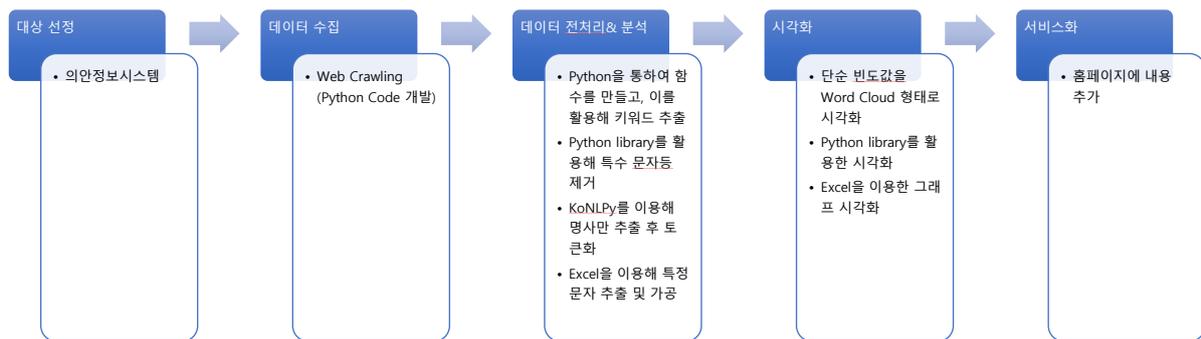
국회 활동에 대한 일반인의 관심을 제고할 수 있도록 관련 자료를 명확하게 보여주는 시각화를 진행한 후, 구체적 데이터 분석을 통해 수치적 자료 및 정당 별 법안 발의와 관련된 구체적 자료까지 한 눈에 볼 수 있는 인사이트를 발굴한다. 이로 하여, 청년들이 이 웹페이지를 이용하여 정당 별 입법 상황 및 특징을 쉽게 이해할 수 있도록 하여 정치에 대한 관심도를 높인다. 그리고

무궁무진한 정치관련 정보 중 객관적인 정보를 쉽게 살펴보며 기본적인 지식을 습득하게 한다. 그리고 많은 정보 중 믿을 수 있는 정보를 골라낼 수 있게 되고 스스로 정치 관련 정보를 잘 찾아볼 수 있는 방향성을 제시한다. 이로써, 자신의 정치에 관련된 의견을 논리적으로 생각할 수 있게 되고 스스로 합리적인 결정을 하도록 하여 정치활동에 좀 더 적극적인 참여를 할 수 있기를 기대한다.

2. 과제 설명

① 과제 요약 설명

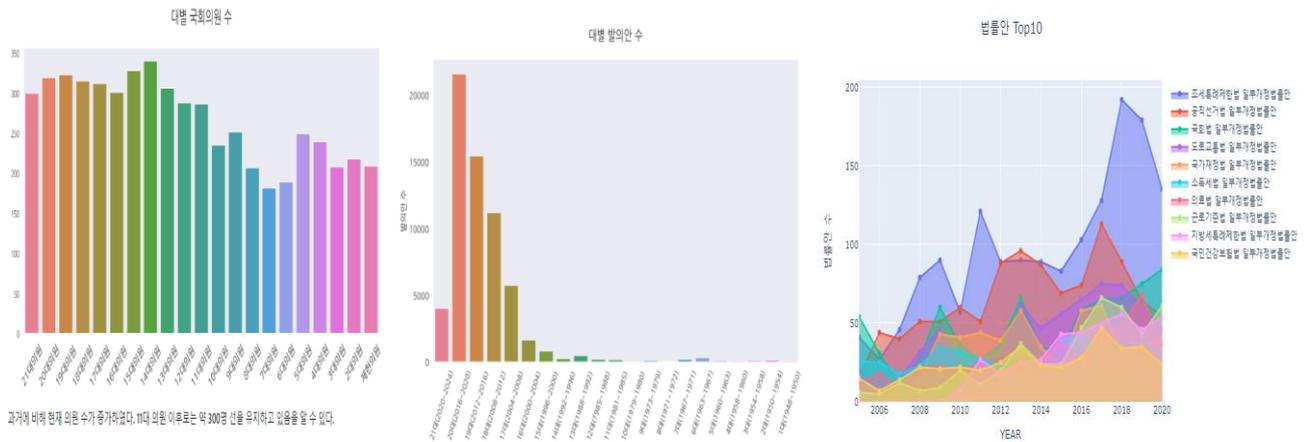
먼저, 개발 프로세스에 대해 설명하면 대상 선정 후 데이터 수집, 데이터 전처리 및 분석 그리고 시각화 하는 과정을 거치고 이를 활용하여 서비스화를 제공하는 순서로 이어진다. 데이터 수집에서 시작해 서비스화 하기까지의 전체적인 과정을 개발하여 데이터 엔지니어링, 데이터 에널리틱스 등의 분야에 대해 시스템경영공학부에서 배운 이와 관련된 지식들을 최대한 활용할 수 있도록 프로세스를 구성하였다.



좀 더 자세히 설명하면, 국회 입법 활동을 시계열 분석과 정당 별 분석에 대해 객관적인 자료를 만들어 전체적인 국회 입법 활동, 정당 별 입법 활동, 1인당 입법 활동 등을 분석한다. 그리고 코로나와 같은 시기별 이슈 이전과 이후의 입법 키워드, 정당 별 관심 키워드 등을 분석하여 웹 페이지를 제작한다. 이를 통해 청년들이 국회의 입법 활동 상황을 객관적으로 볼 수 있도록 한다. 분석을 위한 자료의 경우 의안정보시스템 홈페이지에서 웹크롤링을 통해 데이터를 수집한다. 그리고 Excel과 Python 라이브러리를 활용하여 국회 대수 별 발의 법안 수, 정당 별 발의안 수, 가결 법안 비율, 대통령 임기 중 법안 가결 비율 등의 자료를 분석한다. 그리고 특정 이슈 사건을 중심으로 입법 내용의 변화, 정당 별 관심있는 키워드 등의 분석을 하고 Word Cloud, Python Library, Excel을 통해 시각화를 한다. 이러한 자료를 이용하여 객관적인 정보를 한눈에 알아볼 수 있고 정당 별 구체적인 입법 활동을 보기 쉽게 정리한 웹페이지를 Adobe xd와 Wix를 활용해 만들어 서비스화 한다. 이를 통해 청년들이 국회 입법 활동에 대해 기본적인 지식을 습득하고 관심도를 높일 수 있도록 한다.

② 국회 입법활동 빅데이터 기초분석

국회 입법활동 데이터 중 시계열을 기준으로 기초적인 분석을 진행하였다. 데이터의 경우 경진 대회에서 활용되었던 데이터를 받아와 사용하였고 csv 형식으로 되어있는 데이터를 people, process, suggest 변수에 저장하고 각 데이터에 대한 shape를 확인해보니 (5605, 16), (68530, 28), (62658, 12)으로 되어있음을 확인할 수 있었다. people 변수에 데이터는 역대 국회의원 현황에 대한 정보를 담고 있고 process 변수에 데이터는 본회의 처리 안건에 대한 정보를 담고 있다. 그리고 suggest 변수에 데이터는 국회의원 발의 법의안에 대한 정보를 담고 있다. 이 데이터를 이용해 시계열 방식으로 기초분석을 진행하였다. 먼저, 역대 국회의원 현황에 대한 정보, 본회의 처리 안건에 대한 정보, 국회의원 발의 법의안에 대한 정보를 기초적으로 불러와서 나열하고 시계열 방식으로 국회의원 수, 발의된 법안 수, 그리고 발의안 Top10 시계열 시각화 분석을 진행하였다. 아래의 그래프가 시계열 분석의 결과이다.



③ 구체적 데이터 분석

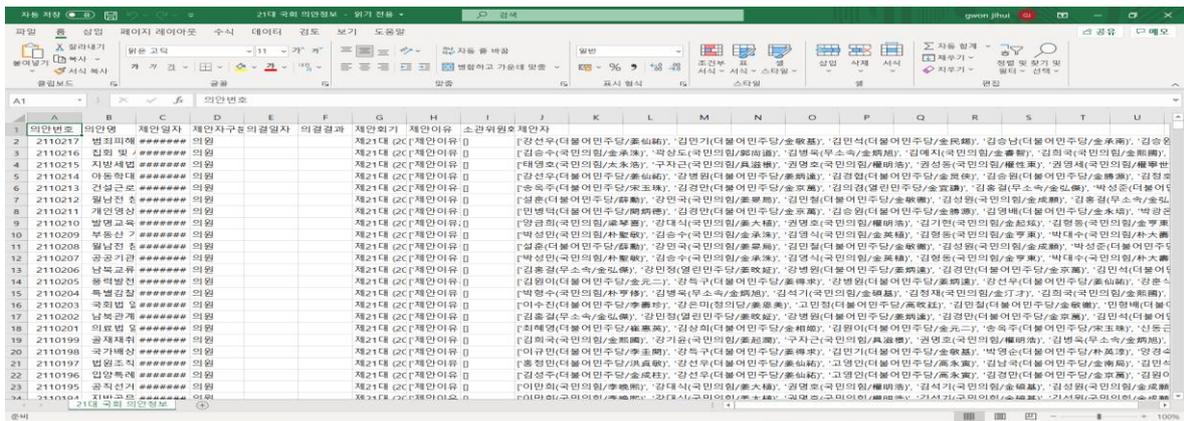
대부분의 작업의 경우 Python 라이브러리와 프레임워크를 활용하여 진행하였고, 추가적으로 필요한 부분이나 간단한 작업의 경우 직접 함수를 만들거나 Excel 소프트웨어를 활용하여 개발을 진행하였다. 수집 데이터 대상은 13대~21대의 기간 중 발의된 법률안을 대상으로 선정하였고 약 8만건의 데이터가 존재하였다. 이 중 분석에 활용된 표본집단 중 하나인 covid-19와 관련된 법률안의 경우 972건이 존재하였다. 이러한 데이터를 수집하기 위해 python 패키지인 BeautifulSoup을 활용하였고 동적인 상황에서의 크롤링을 위해 Selenium framework를 추가적으로 활용해 크롤링을 진행하였다. 법률안 자료를 수집하기 위해 의안정보시스템 (<http://likms.assembly.go.kr/bill/main.do>) 사이트를 이용하였다. 해당 사이트는 의안과 관련된 데이터를 자유롭게 공개하고 있어 데이터 수집 및 분석에 매우 용이하고 사이트의 정보에 대한 권한이 국회에 있어 신뢰성 또한 높아 크롤링할 사이트로 선정하게 되었다. 수집된 데이터의 경우 csv 포맷으로 저장하였고 79640건, covid-19에 관한 972건의 데이터가 확보되었음을 알 수 있다.

- 자료수집 사이트: 의안정보시스템



- 수집 데이터

이름	수집한 날짜	유형	크기
13대 국회 의안정보.csv	2021-05-20 오후 9:11	Microsoft Excel ...	4,561KB
14대 국회 의안정보.csv	2021-05-20 오후 7:56	Microsoft Excel ...	4,275KB
15대 국회 의안정보.csv	2021-05-20 오후 7:16	Microsoft Excel ...	7,443KB
16대 국회 의안정보.csv	2021-05-20 오후 5:54	Microsoft Excel ...	7,687KB
17대 국회 의안정보.csv	2021-05-20 오후 4:15	Microsoft Excel ...	21,656KB
18대 국회 의안정보.csv	2021-05-20 오후 12:09	Microsoft Excel ...	32,083KB
19대 국회 의안정보.csv	2021-05-20 오전 6:06	Microsoft Excel ...	41,035KB
20대 국회 의안정보.csv	2021-05-18 오후 1:41	Microsoft Excel ...	14,339KB
20대 국회 의안정보2.csv	2021-05-19 오후 12:55	Microsoft Excel ...	38,162KB
21대 국회 의안정보.csv	2021-05-19 오후 10:57	Microsoft Excel ...	23,513KB



데이터를 확보한 후 특수기호나 접속사 같은 불필요한 요소들의 전처리 과정을 수행하였다. 전처리를 위해 Python 라이브러리인 pandas 와 numpy 를 활용하였고 추가적으로 자연어 처리를 위해 KoNLpy 그리고 Excel software 를 활용하였다. Python 라이브러리를 통해 전처리 된 데이터 중 코로나와 관련된 키워드가 의안정보에 존재하는 데이터만을 추출하여 csv 포맷으로 저장하였다. 그리고 Excel 을 활용해서 대표 제안자가 포함된 정당만을 추출하여 새로운 column 에 저장하고 이를 대표정당별로 그룹화하여 데이터를 분류하였다.

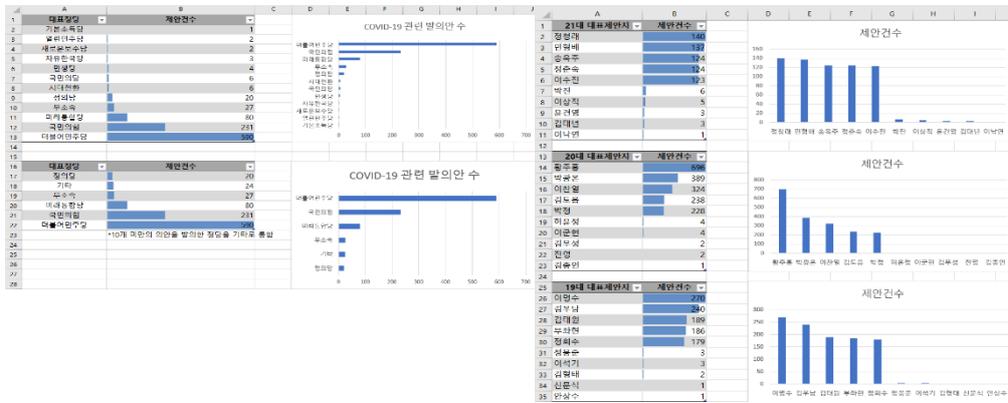


이러한 전처리 과정을 거쳐 얻은 데이터를 가지고 분석과 시각화를 진행하였다. 시각화를 위해 Python 라이브러리인 seaborn, matplotlib 이 활용되었고 추가적으로 Excel software 를 이용하였으며, 워드 클라우드 시각화를 위해 python 워드 클라우드 라이브러리를 이용했다.



이러한 방식을 이용하여 국회 대수 별, 정당 별, 개인 별로 분석을 진행하였고, 이러한 분석을 통해 객관적인 수치로 데이터를 파악하여 입법 활동의 특징에 대해 정리하였다. 특히 정당에 대한 이해를 높이기 위해 정당 별 발의안 수, 가결 비율 등의 분석을 통해 정당 별 특징을 이해할 수 있는 자료를 수집하는데 집중했다. 분석 결과를 아래에 첨부한다.

- 대표 정당, 인물별 발의안 수 시각화 (Excel 이용)

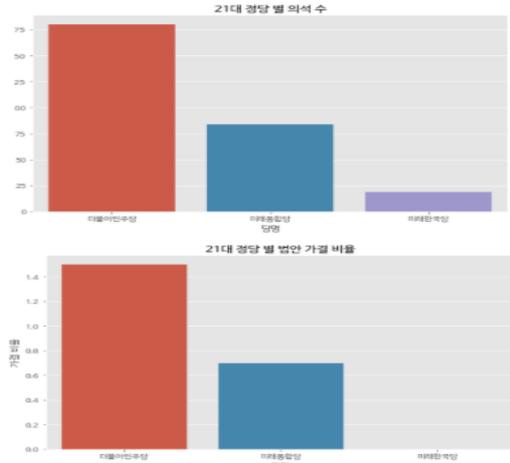
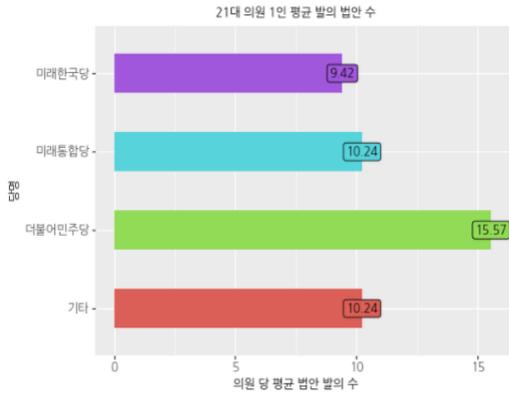


- 대수 및 정당 별 발의안 수 비율 (Python 라이브러리 이용)



- 정당 별 1인 평균 발의 법안 수

- 정당 별 의석 수 및 법안 가결 비율



이러한 분석 이후 키워드를 활용한 응용 분석을 진행하였다. 키워드는 최근 가장 이슈가 되었던 코로나로 선정하였고 코로나와 관련된 법안들을 모두 선정해 정당별, 시기별로 분석을 진행하였다. 시기는 코로나가 눈에 띄게 확산되었던 1 차 대유행, 2 차 대유행, 3 차 대유행 전후 그리고 백신 접종 전후로 코로나 관련 법안을 분석하였다. 응용 분석 결과는 텍스트 마이닝을 사용하여 좀 더 시각적으로 정보를 얻을 수 있도록 하였다. 분석 후 텍스트 마이닝한 결과를 아래에 첨부하였다.

- 정당 별 코로나 텍스트 마이닝



더불어 민주당



미래 통합당



국민의 힘



정의당



무소속



소수 정당

- 시기 별 코로나 텍스트 마이닝

1차 대유행 (20.3.3): 신천지 사건 발생



신천지 이전



신천지 이후

2차 대유행 (20.8.27)



2차 유행 이전



2차 유행 이후

3차 대유행 (20.11.26)



3차 유행 이전



3차 유행 이후

백신 접종 (21.02.06)



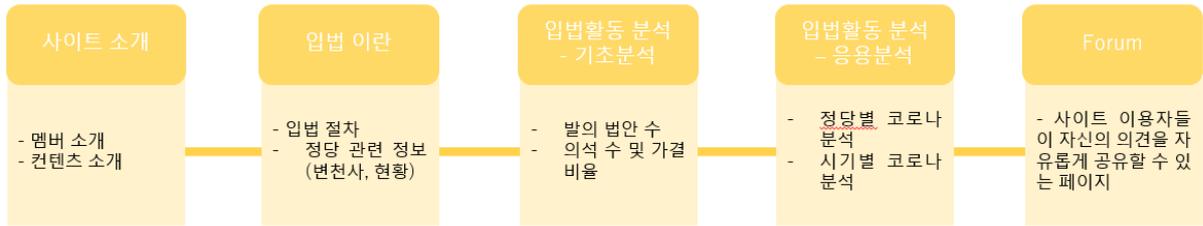
백신 접종 이전



백신 접종 이후

④ 홈페이지 디자인 및 콘텐츠 내용

홈페이지는 Adobe XD 를 사용하여 디자인을 하고 Wix 를 통해 홈페이지로 만들었다. 위의 분석 결과를 토대로 내용을 구체화하고 조사하여 웹페이지를 구성하였다. 홈페이지의 구성은 크게 사이트 소개, 입법 활동에 대한 정보, 입법 활동 기초 분석, 응용 분석, 토론광장으로 나누어진다. 사이트 소개에는 멤버 소개, 콘텐츠 소개 등의 내용이 들어간다. 입법 활동에 대한 정보는 '입법 이란' 이라는 카테고리를 통해 전달한다. 여기에는 입법 활동의 절차와 기초적인 용어 설명이 들어가고 정당의 변천사를 소개한다. 그리고 21 대 국회 현황에 대해 알려주어 현재 국회의 상황을 알 수 있게 해준다. 그리고 입법 활동 분석 관련 페이지는 기초 분석과 응용 분석으로 나뉘는데 기초분석은 전체적인 국회의원 수, 정당 별 의석 수, 가결 비율 등에 대한 분석 결과와 이러한 데이터가 나온 이유에 대해 설명해준다. 이 때, 페이지 구성 순서를 전체 국회, 정당 별, 1 인당으로 진행하여 전체적인 내용에서 점차 구체적인 내용으로 볼 수 있도록 배치하였다. 그리고 응용 분석에는 시기별 코로나 키워드 분석, 정당 별 코로나 키워드 분석의 내용을 이용하여 자료를 제공한다. 그리고 Forum 이라는 페이지를 통해 사이트 이용자들이 자유롭게 의견을 공유할 수 있는 페이지를 제작한다. 홈페이지는 정치에 무지한 사람들을 위해 자료를 제공하는 사이트이기 때문에 잡지형식으로 구성하였다.



분석한 데이터를 토대로 결과를 분석하고 자료를 조사하여 콘텐츠를 구성하였다. 이제부터 구체적으로 페이지별 콘텐츠 내용을 소개할 것이다.

‘입법 이란’이라는 카테고리에는 인터넷에서 수집한 입법활동과 관련된 기초적인 정보를 제공한다. 입법 절차에 관한 내용과 정당 당명의 변천사, 그리고 2021 년 정당 현황에 대한 내용이 있다. 이 페이지를 통해 국회 입법 활동과 정치를 이해하기 위한 기본적인 정당에 대한 정보를 제공하여 사용자들이 우리의 콘텐츠를 잘 이해할 수 있도록 도움을 준다.

‘입법 활동 분석 - 기초 분석’ 카테고리에서는 국회 대수 별 발의 법안 수, 국회의원 1 인당 평균 발의 법안 수, 대수 별 정당 의석 수 및 가결 법안 비율 분석의 내용을 제공하여 입법 활동에 대해 국회가 얼마나 잘 활동하고 있는지를 보여준다. 특히, 대수 별, 정당 별, 개인별로 정리하여 큰 범위에서 작은 범위까지 아우르는 정보를 제공해준다. 내용을 간단히 추려서 말하면 국회 대수 별 발의 법안 수와 관련된 데이터 분석 결과에서는 17 대를 기점으로 급증하는 모습을 보이는 이유를 설명하였다. 급증하게 된 이유는 의원 발의 수가 공천의 기준이 된 것, 공동 발의 최소인원 기준 완화 등 때문이다. 그리고 정당 별 법안 가결 비율에서 여당의 발의 법안 가결 비율이 높는데 여당과 야당의 의석 수 차이가 법안 가결에 영향을 준다는 점을 알 수 있다. 여당과 제 1 야당의 의석 수 차이가 적으면 견제력이 커져 여당의 법안 가결율이 낮아지고, 의석 수 차이가 크면 견제력이 작아져 가결율이 높아지는 경향이 있다. 그리고 국회의원 1 인당 평균 발의 법안 수에 대한 내용을 통해 어떤 의원이 가장 입법활동에 활발히 참여하는지에 대한 정보를 얻을 수 있었다.

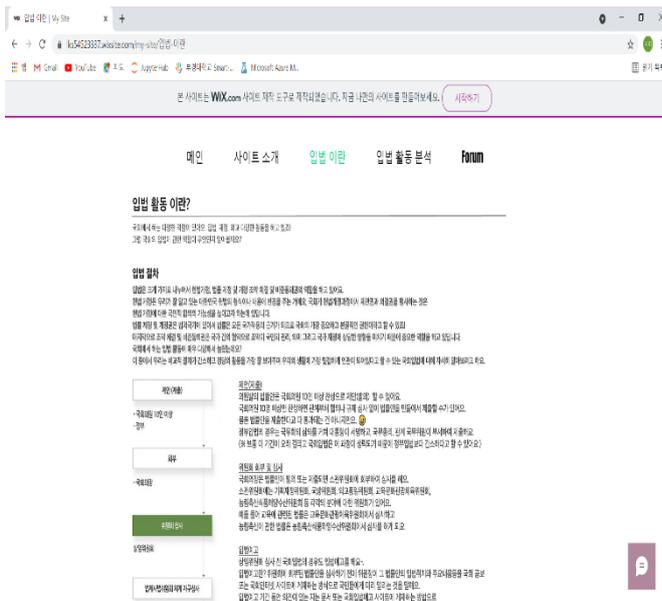
‘입법 활동 분석 - 응용 분석’ 카테고리에서는 코로나를 통해 정당별, 시기별로 텍스트 마이닝한 결과에 대해 분석하였다. 이를 통해 정당 별 차이점, 코로나 관련 의안에서 중요하게 생각하는 키워드를 알 수 있었다. 무소속과 소수 정당은 의석 수가 매우 적어 비교적 정확한 분석 결과를 나타내지 못하였지만 여당과 제 1 여당에 대한 차이점을 알 수 있었다. 여당에서는 강화, 대응, 보호 등 긍정적인 키워드가 많이 출현하였는데 야당에서는 어려움, 부담, 악화와 같은 부정적인 키워드가 많이 나타난 것을 볼 수 있다. 시기별 분석으로는 국회 입법 활동을 하는 의원들이 얼마나 상황에 민감하게 반응하고 의안에 적용을 하는 지 알 수 있었다. 신천지 사건 발생인 1 차 대유행, 2 차 대유행, 3 차 대유행, 백신 접종을 기준으로 분석을 하였다. 각 시기마다 뚜렷한 차이점을 볼 수 있었다 특히, 대유행 발생 이후 이에 대한 대응방안과 일맥상통 한 키워드들이 많은 비중을 차지 한 걸 알 수 있었고 코로나 발생 이후 1 년이 지난 시기부터는 온라인, 플랫폼 등의 키워드가 등장하면서 비대면 방식의 프로세스가 강조되고 있음을 파악할 수 있었다.

'Forum'카테고리에서는 Q&A, 게시물 등을 사용자들이 직접 올려 지식을 공유하고 의견을 공유하는 페이지를 만들었다. 이러한 카테고리에서 사용자들이 원하는 정보에 대한 분석을 진행하여 더욱 다양한 콘텐츠를 제공하는 사이트로 개발하려 한다.

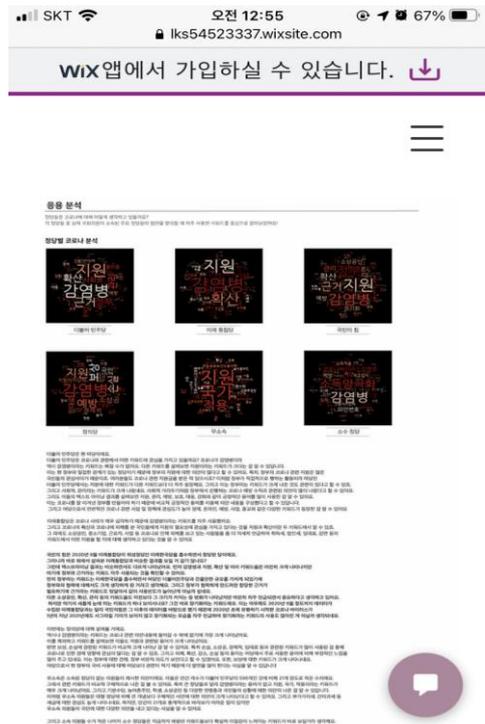
홈페이지의 자세한 내용은 아래의 홈페이지 링크를 통해 제공한다.

▶ <https://lks54523337.wixsite.com/my-site>

- PC 화면



- 핸드폰 화면



3. 결론

① 기대효과 및 결론

이 과제를 통해 데이터 분석, 웹 크롤링, 시각화 및 텍스트마이닝 기술에 대한 이해도를 높일 수 있었고 홈페이지 제작 시 고려해야할 점, 콘텐츠 구성 방법 등에 대한 내용을 배울 수 있었다. 그리고 우리가 제작한 홈페이지의 수요층은 정치에 무지한 사람들로 설정하였다. 기초적인 내용을 조사하고 정리하여 객관적인 내용만을 제공하기 위해 노력하였다. 국회 입법 활동에 대한 자료를 명확하게 보여주는 시각화를 진행하여 정치에 대한 진입장벽을 낮춰주는 인사이트를 제공하였다. 이로 인해, 정치에 대한 이해도를 높이고 자세한 내용은 사용자들이 직접 정당 홈페이지, 국회 사이트를 검색해보며 자신의 정치에 관련된 의견을 옳은 정보를 토대로 내세울 수 있도록 도와준다. 분석을 하며 입법 활동에 정당이 큰 영향을 준다는 사실을 알 수 있었다. 특히 소수정당일 경우에는 진행이 어려운 경우가 많아 보였다. 한 인물의 공약이나 인품을 보고 투표를 하는 것도 중요하겠지만 이 사람이 어떤 정당에 속해 있는지를 확인하는 것도 매우 중요하다는 것을

알 수 있다. 그 이유는 입법이라는 중요한 활동을 할 때는 당론이 정해지기 때문에 정당이 중요하다. 총선이라는 게 대통령의 권력을 좌지우지하기도 하고 견제력에 차이를 줄 수 있기 때문에 입법 활동에 참여하는 국회의원 선거도 매우 중요하다고 할 수 있다. 또한, 코로나와 같은 키워드 분석을 통해 정당 및 의원안의 구체적 내용을 요약적으로 볼 수 있었다. 국회에서 입법 활동을 할 때 시기에 맞춰 잘 안건을 내세우는지 정당별로 어느 분야에 관심이 높은 지 알 수 있었다. 이러한 활동을 통해 국회에서 입법 활동을 얼마나 열심히 하고 있는지, 의미 있는 발의안을 내고 국민을 위해 일하고 있는지 생각해볼 수 있는 계기가 되었다.

② 한계점 및 시사점

정당 별로 관심이 높은 분야, 내세우는 안건의 특징에 대해 확실히 비교 분석하여 차이점을 찾으려 했는데 각 정당은 모든 분야에 관심을 가지고 있었으며 분석 결과에서 생각보다 큰 차이가 없었다. 또한, 웹 크롤링 과정이 생각보다 오래 걸려서 코로나 이외의 키워드에 대한 분석을 하지 못하는 한계가 있었다. 이후에 사이트가 활성화되고 토론 콘텐츠에서 이용자들의 관심사, 요청사항 등을 취합해 검색기능을 추가한 형식으로 사이트를 개발해 나갈 예정이다. 이후, 검색 기능을 추가하여 경제 분야, 보건 분야, 사회적 이슈 분야 등의 파트로 나누어 더욱 다양한 정보를 제공할 것이다.

4. 출처 및 참고자료

- 정부입법지원센터(<https://www.lawmaking.go.kr/>)
- 국회입법예고(<https://pal.assembly.go.kr/main/mainView.do>)
- 대한민국국회(<https://www.assembly.go.kr/assm/userMain/main.do>)
- 의안정보시스템 (<http://likms.assembly.go.kr/bill/PetitionStat.do>)
- 국회 신뢰 제고 방안 연구 보고서 (국회미래연구원, 2019)
- 19대 국회 입법현황 및 향후 입법 방향성 분석 연구 (차현숙, 강현철, 2016)

- <https://dacon.io/competitions/official/235679/data/>
- <https://news.mt.co.kr/mtview.php?no=2017010915417694297>
- <https://www.icouncil.kr/news/articleView.html?idxno=44341>
- <http://newslabit.hankyung.com/article/201804195943G>
- <https://www.peoplepower21.org/Politics/537048>
- <https://www.etoday.co.kr/news/view/1984051>