

박사학위논문

상황별 재난 대비가 가능한
인공지능 경보방송 방법 연구

2023년

한 성 대 학 교 대 학 원

스마트융합건설팅학과

스마트융합제품전공

류 태 하

박 사 학 위 논 문
지도교수 김승천

상황별 재난 대비가 가능한 인공지능 경보방송 방법 연구

A Study on AI Alert Broadcasting Method for
Disaster Preparedness by Situation

2022년 12월 일

한 성 대 학 교 대 학 원

스마트융합컨설팅학과

스마트융합제품전공

류 태 하

박 사 학 위 논 문
지도교수 김승천

상황별 재난 대비가 가능한 인공지능 경보방송 방법 연구

A Study on AI Alert Broadcasting Method for
Disaster Preparedness by Situation

위 논문을 공학 박사학위 논문으로 제출함

2022년 12월 일

한 성 대 학 교 대 학 원

스마트융합건설팅학과

스마트융합제품전공

류 태 하

류태하의 공학 박사학위 논문을 인준함

2022년 12월 일

심사위원장 노광현 (인)

심 사 위 원 홍정완 (인)

심 사 위 원 임진호 (인)

심 사 위 원 임황용 (인)

심 사 위 원 김승천 (인)

국 문 초 록

상황별 재난 대비가 가능한 인공지능 경보방송 방법 연구

한 성 대 학 교 대 학 원
스 마 트 융 합 컨 설 팅 학 과
스 마 트 융 합 제 품 전 공
류 태 하

현대사회는 과학기술의 발전을 토대로 문명의 발전을 이루어왔다. 그러나 이로 인한 복잡성의 증가와 환경파괴와 같은 역기능으로 다양한 재난의 발생위험이 커지고 있으며, 그 종류도 매우 다양해져서 기존에는 존재하지 않았던 신종 재난들도 생겨나고 있다. 현행법상 재난을 크게 두 가지로 분류하는데 태풍, 홍수, 호우 등 자연 활동에서 유래한 자연재난과 이를 제외한 인간들의 과학 문명 발달과 함께 야기된 사회재난으로 구분한다. 그런데 자연재난과 사회재난 모두 문명의 발달과 함께 야기된 재난들이다. 자연재난의 경우 기후변화에 적극적으로 대응하지 않고 적응해 가면서 온실가스 감축에 실패하면서 나타난 결과라고 전문가들은 진단한다. 사회재난 또한 마찬가지로 인간의 문명 발달과 함께 발생하는 것으로 새로운 사회재난들이 계속 생겨나고 있다. 코로나19, 가축전염병, 미세먼지 문제 등은 물론이고 법률적으로 정의

되지는 않았지만 최근 이태원에서 발생한 대규모 압사 참사와 같은 재난들도 속출하고 있다. 이러한 재난은 통제할 필요가 있으며, 재난으로 인한 피해를 최소화하기 위한 제반 활동을 재난관리라고 한다. 재난을 4단계로 분류할 수 있는데 예방과 대비라는 사전 방재단계와 대응과 복구라는 사후 수습단계로 구분할 수 있다. 여기서 예방과 대비는 선제적 예방 활동을 통해 잠재적 위험을 회피하기 위한 활동이고, 대응과 복구는 발생한 피해 상황을 최소화하거나 피해자들과 지역사회를 정상화하기 위한 단계이다. 이러한 재난의 4단계는 독립적으로 구분되지 않고 서로 연계된 활동으로 볼 수 있다. 이러한 재난에 대응하기 위한 다양한 방법 중 본 연구에서는 재난대응 시스템 중 하나인 전관방송 시스템에 대비를 위한 방법을 추가하여 재난 대비를 위한 방법을 제안하였다. 전관방송 시스템이란 불특정 다수에게 일방향으로 음향 정보를 전달하는 안내방송시스템으로 일명 구내방송 장비라고도 한다. 평상시 배경음악이나 안내방송 송출 장비로 동작을 하다가 화재나 긴급상황 시 비상방송 시스템으로 전환되어 비상방송을 송출하는 장비이다.

본 연구의 근본적인 목적은 경제적이고 효과적인 재난 대비방법을 제안하는 것이다. 그 목적 달성을 위해 정부의 공공데이터 포털에서 제공하는 공공데이터와 재난 포털의 재난 관련 뉴스를 웹크롤링을 통해 가져와 정보의 전달력을 개선하기 위한 작업을 한다. 문장에 포함된 불필요한 내용과 수사 등을 제거하고, 긴 문장을 청자가 이해하기 쉬운 짧은 글로 요약하여 전관방송 시스템에 장착된 TTS를 통해 음성을 재생하여 송출하는 것이다. 구체적인 방법으로는 공공데이터 포털에서 가져온 데이터는 정형화된 문장을 통해 해당 수치나 상태 값을 바꾸면서 방송 콘텐츠로 활용 가능하다. 하지만, 재난 뉴스의 경우에는 딥러닝 기반 또는 통계 기반의 자연어 요약모델을 이용하여 문장을 요약한다. 일반적으로 좋은 성능을 보이는 언어 모델을 개발하거나 언어 모델에 미세조정하여 성능 향상을 위한 연구가 진행된다. 하지만, 본 논문에서는 성능이 좋은 여러 가지 모델들을 통해 요약된 문서를 시스템을 통해 실시간으로 평가하고, 가장 우수한 평가를 받은 요약문을 방송용으로 추천하는 방식이다. 다양한 자연어 처리용 언어 모델들이 좋은 평가를 받을 수도 있고 어떤 경우에는 그렇지 못할 수도 있다. 문제는 좋은 모델의 결과물이 항상

좋은 평가를 받을 수는 없다는 것이다. 이러한 결과물에 대한 평가를 위해서는 좋은 참조본(reference 또는 정답문서)이 있어야 하는데 이를 위해서 사람이 직접 참조본을 만들게 되면 실시간성이 떨어지고, 많은 인건비가 소요된다. 이에 대한 대안으로 각 언어 모델이 요약한 요약문을 평가를 위한 참조본으로 사용하여 상호 평가를 통해 결과를 얻는 것이다. ROUGE와 BLEU 점수를 평가지표로 실험하면서 확인한 사항은 항상 우수한 결과가 나타나는 모델은 없다는 것이다. 따라서 실시간 문서 요약 후 평가를 통해 요약문을 추천하는 본 연구에서는 어떤 원문이건 상관없이 최고의 언어 모델을 통해 요약한 문서를 얻을 수 있게 되었다. 이를 통해 사람들이 집중해서 듣고 쉽게 이해할 수 있는 재난 대비방송 콘텐츠로 활용할 수 있다.

본 연구에서 제안한 '실시간 재난데이터 수신 및 홍수, 가뭄, 미세먼지 등 각종 재난 뉴스 요약을 통한 경보 방식'을 전관방송 장비의 기능으로 채택한다면, 요약 문장을 활용한 방송이 재난을 회피하기 위한 좋은 도구로 활용될 수 있을 것으로 기대된다.

【주제어】 공공데이터, 재난관리, 인공지능, 자연어 처리, ROUGE

목 차

제 1 장 서 론	1
제 1 절 연구의 배경	1
제 2 절 연구의 필요성	9
제 3 절 연구의 목적 및 방법	13
1) 연구의 목적	13
2) 연구의 방법	14
제 2 장 관련 기술 및 선행연구 분석	15
제 1 절 재난관리 시스템 기술 현황	15
1) 인공지능을 활용한 재난안전기술	15
2) 산불 예방을 위한 첨단 기술 동향	20
3) 재난 예측 기술 개발 동향	22
제 2 절 재난방송 시스템 기술 현황	26
1) 미국의 재난방송 시스템 기술 동향	26
2) 일본의 재난관리 및 방송 시스템 현황	29
3) 국내 재난 예·경보 시스템 현황	32
4) 비상방송 시스템 연구 현황	36
제 3 절 공공데이터 관련 선행연구	40
1) 공공데이터의 개요	40
2) 민간 활용 사례 연구	47
제 4 절 자연어 처리 관련 선행연구	52
1) 자연어 처리의 개요	52
2) 문서 요약의 정의 및 동향	59
3) 문서 요약 모델	62

제 5 절 요약 및 시사점	73
1) 선행연구 분석 요약	73
2) 시사점	74
 제 3 장 시스템 설계	 75
제 1 절 연구 시스템 개요	75
제 2 절 시스템 구성	78
1) 공공데이터 수집	78
2) 텍스트 마이닝	82
3) 문서 요약	87
4) 평가 및 모델 추천	88
5) 방송 송출	96
 제 4 장 실험 결과 및 분석	 99
제 1 절 실험 개요	99
제 2 절 조건별 성능 비교	101
1) 뉴스 분야별 요약 결과 및 평가 결과	102
2) 참조본별 요약 결과 및 평가 결과	124
제 3 절 연구의 성과	130
1) 시스템 활용의 기대효과	130
 제 5 장 결론	 132
 참 고 문 헌	 134
 ABSTRACT	 146

표 목 차

[표 1-1] 재난의 종류	2
[표 2-1] 참여기관별 역할	26
[표 2-2] EAS 메시지 발령 내용(EEE) 코드 예시	27
[표 2-3] 공공데이터의 정의	40
[표 2-4] 공공부문(재난안전) 데이터 분석 및 활용 우수사례	44
[표 2-5] 국외 주요국 공공데이터 개방 및 활용 정책	46
[표 2-6] 범주별 NLP 응용	55
[표 2-7] 추출 요약 방법별 문장의 중요성 평가 기준	61
[표 3-1] 대기오염정보 API 개요 및 실험 조건	78
[표 3-2] 응답 메시지 명세	79
[표 3-3] 경고용 재난 상황 정의	85
[표 3-4] 평가 예시	95
[표 4-1] 모델별 실험 결과 1 (연합뉴스: 코로나19)	102
[표 4-2] 모델별 실험 결과 2 (연합뉴스: 코로나19)	105
[표 4-3] 모델별 실험 결과 3 (연합뉴스: 코로나19)	108
[표 4-4] 모델별 실험 결과 4 (KBS: 날씨)	111
[표 4-5] 모델별 실험 결과 5 (연합뉴스: 가뭄)	114
[표 4-6] 모델별 실험 결과 6 (KBS: 홍수)	117
[표 4-7] 모델별 실험 결과 7 (KBS: 미세먼지)	120
[표 4-8] 모델별 실험 결과 정리	122
[표 4-9] GPT 변경 실험 결과 1-1, 1-2 (연합뉴스: 코로나19)	124
[표 4-10] GPT 재요약 시 실험 결과 정리	128

그 립 목 차

[그림 1-1] 세계 재난 피해 규모	6
[그림 1-2] 재난 유형별 발생 비율	6
[그림 1-3] 자연재난 발생 현황	7
[그림 1-4] 최근 10년간 재원별 복구비 현황	8
[그림 1-5] 사회재난 발생 현황 (단위: 건)	8
[그림 1-6] 사회재난 재산 피해 현황	8
[그림 1-7] 재난의 4단계	9
[그림 1-8] 전관방송 시스템 기본 구성도	11
[그림 2-1] 인공신경망 기반 지진동 감지 모델	16
[그림 2-2] 딥러닝 기반 도시침수해석	17
[그림 2-3] 지역별 최대풍속 및 정전 가구수 예측 프로그램	18
[그림 2-4] 지붕 타입 식별 모델	19
[그림 2-5] ASTER 센서로 확인된 미국 그랑프리(Grand Prix) 산불	20
[그림 2-6] LoRa 네트워크 구성도	21
[그림 2-7] WIFERE 산불 확산 예측 지도	22
[그림 2-8] 건축물 화재 위험도 예측 결과	23
[그림 2-9] DART 4G System	24
[그림 2-10] IPAWS 구성도	28
[그림 2-11] NHK의 TOC 운영 체계	30
[그림 2-12] EWBS 체계	31
[그림 2-13] 재난문자방송 송출체계도	33
[그림 2-14] DMB 재난경보서비스 발령체계도	34
[그림 2-15] 재난방송온라인서비스 세부 구성도	35
[그림 2-16] 주소기반 비상방송 시스템	36
[그림 2-17] 인터넷 기반의 수신자 맞춤형 재난경보 전달시스템 시스템	37
[그림 2-18] 재난경보 방송의 재전송을 위한 지능형 이동 로봇들의 협업	38
[그림 2-19] 제안기술의 개념도(좌)과 방향탐지 앱(우)	39

[그림 2-20] 데이터 맵	42
[그림 2-21] 데이터 수집에서 활용까지의 과정	43
[그림 2-22] 서울 실시간 도시데이터 예시화면	45
[그림 2-23] 서울 생활이동 예시화면	46
[그림 2-24] 에어가드K 실행화면	48
[그림 2-25] 통합재난정보전달솔루션 구성도	49
[그림 2-26] 굿바이코로나 실행화면	50
[그림 2-27] enbrix 실행화면	51
[그림 2-28] 세계안전정보	51
[그림 2-29] NLP의 하위 분야	52
[그림 2-30] 심층 신경망(DNN) 개념도	54
[그림 2-31] 자연어 처리 과정	56
[그림 2-32] 추출 요약과 생성 요약 비교	60
[그림 2-33] TextRank 데이터 흐름	63
[그림 2-34] Copy Mechanism 모델	65
[그림 2-35] Transformer 모델	66
[그림 2-36] BERT 모델	68
[그림 2-37] GPT-3의 학습 및 출력	70
[그림 2-38] 텍스트 대 텍스트 프레임워크 다이어그램	71
[그림 3-1] 시스템 환경	75
[그림 3-2] 방송 결정 흐름도	76
[그림 3-3] HTML 태그와 애트리뷰트	83
[그림 3-4] KBS뉴스 웹 접근 화면	84
[그림 3-5] 웹 크롤러 구성도	85
[그림 3-6] 웹 스크래핑 순서도	86
[그림 3-7] 방송용 콘텐츠 생성	87
[그림 3-8] 문서 생성 모델의 활용 가능 평가 지표	89
[그림 3-9] BERTScore 절차도	93
[그림 3-10] 상호 평가 방법	95
[그림 3-11] 방송 시스템 연동	96

제 1 장 서 론

제 1 절 연구의 배경

인류는 과학기술의 발전을 바탕으로 문명을 발전시켜 왔으며, 최근에는 발전 속도에 가속도가 붙고 있다. 그러나 이로 인한 사회적 복잡성의 증가와 환경파괴와 같은 역기능 또한 증가하고 있어서 재난도 다양한 형태로 발생하고 있다. 전통적인 재난의 경우 기후에 의해 발생하는 태풍, 홍수, 한파, 폭염, 가뭄 등과 지진, 화산활동 등의 자연 활동에 의한 재난인 자연재난과 화재, 붕괴, 폭발, 환경오염 등의 사고를 사회적 재난으로 불렀었다. 과거에는 화재 등을 인적재난이라는 카테고리로 따로 구분하였지만 다양한 재난에 체계적으로 대응하기 위해서 2013년부터는 법률의 개정으로 사회재난에 포함하였다. 자연재난의 경우 자연현상에 의한 재난으로 명확한 구분이 가능하지만, 사회재난은 현대사회에서 COVID-19 등의 감염병, 가축전염병 및 미세먼지로 인한 피해 등과 같이 모든 종류의 재난을 포함하고 있으며 자연현상에 의한 재난을 제외한 인간들의 과학 문명의 발달과 함께 야기된 모든 경우의 재난들로 복잡하고 다양한 형태로 발생하고 있다.

대한민국 헌법 제34조 제6항에서 “국가는 재해를 예방하고 그 위험으로부터 국민을 보호하기 위하여 노력하여야 한다.”라고 명시하고 있다. 또한, 재난 및 안전관리 기본법(약칭:재난안전법) 제1조에서는 법의 제정 목적을 명시하고 있으며, 해당 법률 제3조1항에서는 재난에 대해서 다음과 같이 정의하고 있으며 [표 1-1]과 같이 구분하고 있다. “재난”이란 국민의 생명·신체·재산과 국가에 피해를 주거나 줄 수 있는 것으로서 다음 각 목의 것을 말한다.¹⁾

1) 재난 및 안전관리 기본법. 국가법령정보센터. (2022). <https://www.law.go.kr/법령/재난및안전관리기본법>

[표 1-1] 재난의 종류

구분	유형
가. 자연재난	태풍, 홍수, 호우(豪雨), 강풍, 풍랑, 해일(海溢), 대설, 한파, 낙뢰, 가뭄, 폭염, 지진, 황사(黃砂), 조류(藻類) 대발생, 조수(潮水), 화산활동, 소행성·유성체 등 자연우주물체의 추락·충돌, 그 밖에 이에 준하는 자연현상으로 인하여 발생하는 재해
나. 사회재난	화재·붕괴·폭발·교통사고(항공사고 및 해상사고를 포함한다)·화생방사고·환경오염사고 등으로 인하여 발생하는 대통령령으로 정하는 규모 이상의 피해와 국가핵심기반의 마비, 「감염병의 예방 및 관리에 관한 법률」에 따른 감염병 또는 「가축전염병예방법」에 따른 가축전염병의 확산, 「미세먼지 저감 및 관리에 관한 특별법」에 따른 미세먼지 등으로 인한 피해

또한, 현재 법률적으로 재난의 범주에 포함되지는 않았지만, 2022년 10월 29일 서울 이태원에서 발생한 대규모 압사 참사를 계기로 행정안전부는 주최자가 없는 축제에 대한 지자체의 안전관리 의무를 규정하는 ‘재난 및 안전관리 기본법’에 대해서 의원 발의안과 연계해 세부 규정으로 ‘다중밀집 인파사고 안전관리 지침’을 제정하기로 했다. 문화체육관광부는 공연장 등 내부에서의 유사 다중 밀집 인파사고를 예방하기 위해 공연장 재난대응 매뉴얼 등을 보완할 방침이다. 또 휴대전화 위치정보, 지능형 폐쇄회로(CC)TV, 드론 등 최신 과학기술을 이용해 실시간 다중 밀집도를 분석, 위험 예측 시스템을 구축할 계획이라고 밝혔다.²⁾

먼저, 특·광역시와 50만 이상 대도시를 중심으로, 지하철 역사 등에 다중 밀집 인파 사고 우려 시 사전경보, 지하철 환승역 등에 대한 밀집 시간대 예방 활동 추진, 대규모 밀집 행사에 대한 민관협력체계 구축 등을 지자체가 유관기관과 협력하여 즉시 검토하고, 시행하도록 안내하기로 하였다. 이태원 사고를 계기로 최근 국회에서 발의된 「재난안전법」 개정안(5건^{2022.11.2.기준})도 실제 현장에 적용될 수 있는 방향으로 개정될 수 있도록 검토하여 대응하기

2) 김수현. (2022년11월3일). 인파사고 안전관리지침 제정한다...공연장 재난대응 매뉴얼 보완(종합). 연합뉴스. <https://www.yna.co.kr/view/AKR20221103065151530?input=1195m>

로 했다.³⁾

그 외에도 국가적으로 중대한 재난이 발생했거나 재난이 예상되는 경우 주민의 인명과 재산 피해를 최소화하기 위한 사전적 신호 수단으로 정의된 민방위기본법에 따라 민방위 사태가 발생했을 때 국민들이 원활하게 대피하기 위해서 민방위기본법이 존재한다. 민방위기본법의 제1조에서 법령의 목적을 명시하고 있는데 그 내용은 다음과 같다. “전시·사변 또는 이에 준하는 비상사태나 국가적 재난으로부터 주민의 생명과 재산을 보호하기 위하여 민방위에 관한 기본적인 사항과 민방위대의 설치·조직·편성과 동원 등에 관한 사항을 규정함을 목적으로 한다.”⁴⁾

민방위의 정의로는 주민의 생명과 재산을 보호하기 위하여 정부의 지도하에 주민이 수행하여야 할 방공, 응급적인 방재·구조·복구 및 군사 작전상 필요한 노력 지원 등의 모든 자위적 활동이라고 설명하고 있다. 여기에는 재난 및 안전관리 기본법에 따른 재난사태 선포 또는 동법에 따른 특별재난지역 선포 등의 국가적 재난, 그 밖에 행정안전부 장관이 정하는 재난사태까지도 민방위라고 정의한다. 이에 따른 민방위기본법 제33조 3항에 의하면 경보단 말장비 설비와 관련하여 의무적 설치시설로 역, 터미널, 공항, 항만 등의 운수 시설과 3천㎡ 이상 복합쇼핑몰 등 대규모 점포, 7개 이상 상영관을 가진 영화관으로 지정되어 있는데 이러한 시설에서는 일반적인 재난 및 안전관리에 관한 법률에 더하여 민방위법에 따른 부분도 충족하도록 요구하고 있다.

해외에서는 재난을 어떻게 정의하는지 살펴보면, 유엔재해기구(United Nations International Strategy for Disaster Reduction)에서는 정상적인 생활 영위의 불가능성과 재난의 예측 불가능성에 초점을 맞추어 “갑작스럽게 발생하여 지역사회의 기본조직과 정상기능을 와해시키는 큰 규모의 사건으로서 그 영향을 받은 지역사회가 외부의 도움이 없이는 극복할 수 없고 생명과 재산, 사회 간접시설과 생활수단에 일상적인 능력으로 처리할 수 없는 피해를 일으키는 단일 또는 일련의 사건”⁵⁾이라고 정의하고 있다. 미국의 연방재난관

3) 행정안전부 보도자료. (2022년11월3일). 다중 밀집 인파사고 예방 안전관리 대책 관계장관회의 개최 결과.

4) 민방위기본법. 국가법령정보센터. (2022). <https://www.law.go.kr/법령/민방위기본법>

5) UNISDR(United Nations Office for Disaster Risk Reduction). (2017). <http://www.unisdr.org>

리청(Federal Emergency Management Agency, FEMA)에서는 “재난은 정부의 통상적인 관리절차나 자원으로는 대처할 수 없는 인적 및 물적 손상을 초래하는 사건을 말하고, 재난은 대개 돌연히 발생하지만 대처 과정에서 다수의 정부기관과 민간부문들의 즉각적이며 조정된 노력을 필요로 한다”고 정의하고 있다.

일본의 재해대책기본법 제2조 제1항에서 설명하는 재난의 정의는 다음과 같다. “재난은 태풍, 호우, 폭설, 홍수, 해일, 지진, 쓰나미(지진해일), 화산폭발, 그 밖의 이상한 자연현상 또는 대규모 화재, 폭발 기타의 원인에 의해 생기는 피해이다.”⁶⁾

이처럼 재난에 대한 정의가 물질문명의 발전과 함께 인적·사회적 요인에 의한 대형사고까지도 재난의 범주에 포함 시키고 있으며, 테러 및 전쟁으로 인한 재난까지도 포함하는 개념으로 정립되고 있다. 다시 말하면 근대 사회에서 현대사회로 발전하면서 재난의 관리대상 범위가 확대되고 있으나, 재난의 근본적인 개념적 요소는 변함이 없다고 할 수 있다. 재난은 피해 규모로서 일정 규모 이상의 피해, 사회적 관점에서 재난이 미치는 사회의 충격과 혼란 상태 등으로 구분하여 사고와는 구별되는 다양한 정의를 하고 있다.⁷⁾

신종 재난의 사례로는 부산 해운대 일대의 고층 빌딩들 사이에서 발생하는 빌딩풍에 의한 피해사례를 들 수 있다. 이는 공기가 기압이 높은 곳에서 낮은 곳으로 이동하면서 돌풍과 소용돌이 등을 일으키는 벤투리(Venturi) 효과에 의해서 건물의 유리창 등이 파손되는 것으로 기존에는 볼 수 없었던 신종 재난의 사례로 볼 수 있다. 이러한 사례들에 대응하기 위해서 “빌딩풍 재난법” 법안이 국회에 입법 상정된 상태이며, 부산시의회에서도 빌딩풍을 신종 재난으로 인식하여 이로 인한 시민의 생명과 재산을 보호하기 위하여 빌딩풍 예방과 빌딩풍으로 인한 피해 저감 및 사후관리 지원 등에 필요한 사항으로 규정하여 2021년 9월에 “부산시 빌딩풍 예방 및 피해 저감에 관한 조례”⁸⁾가

6) 진병문, 박재영, 안대명, 윤혜경. (2019). 『재난안전』. 서울: (주)도서출판 북스힐.

7) 한국방재협회 방재사전편찬위원회. (2010). 재난 disaster. 수록처: 『방재사전』.(305). 서울: R&D프레스.

8) 부산광역시 빌딩풍 예방 및 피해 저감에 관한 조례안. 부산광역시의회. (2021).

https://council.busan.go.kr/assem/Synap/skin/doc.html?fn=163184532320337.hwp&rs=/assem/upload_data/Synap

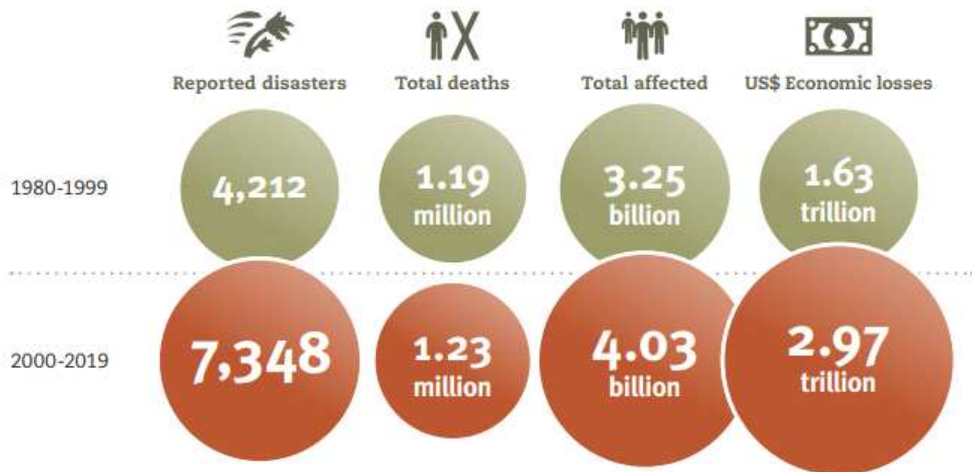
규정되어 예방 활동을 할 만큼 신중 재난들이 계속해서 발생하고 있다.

복합 재난의 해외 사례로 2011년 3월 발생한 후쿠시마 제1 원자력발전소 사고의 경우 일본 도호쿠 인근 태평양 연안에서 발생한 대지진으로 인해 초대형 쓰나미가 발생하였고 이로 인해 도호쿠 연안 지역에서 대규모 인적, 물적 피해가 발생하면서 특히 후쿠시마 원전 방사능 누출사고가 발생했었다. 이 사고의 경우 지진 및 쓰나미라는 자연재해와 이로 인한 원자력발전소의 전원 공급 중단이 발생하였고, 이러한 비상사태에 대비하기 위한 지하에 설치된 비상 디젤발전기가 침수로 인해 가동되지 않으면서 원자로 노심냉각장치의 동작 중지에도 의한 폭발 및 방사능 누출사고가 발생한 사례로서 이처럼 복합적으로 발생한 사례도 있다.

재난은 현대사회에서만 일어나는 것이 아니다. 조선왕조실록에 의하면 그 당시의 홍수, 산사태, 지진, 화재 등 다양한 재난과 관련된 기록들이 자세하게 기록되어 있다. 한 가지 사례로 인조실록 44권, 인조 21년(1643년) 6월 9일 첫 번째 기사에서는 “서울에 지진이 있었다. 경상도의 대구(大丘)·안동(安東)·김해(金海)·영덕(盈德) 등 고을에도 지진이 있어 연대(烟臺)와 성첩(城堞)이 많이 무너졌다. 울산부(蔚山府)에서는 땅이 갈라지고 물이 솟구쳐 나왔다. 전라도에도 지진이 있었다. 화순현(和順縣)에서는 부자(父子)가 벼락을 맞아 죽었고, 영광군(靈光郡)에서는 형제가 말을 타고 들에 나갔다가 말과 함께 일시에 벼락맞아 죽었다.”⁹⁾라고 피해 상황들을 기록하고 있다. 기록문화가 현대만큼 발달하지 않았던 시절의 기록이지만, 피해 상황들에 대해서는 비교적 상세하게 기록하고 있다. 이처럼 동서고금을 막론하고 재난은 항상 존재해 왔다.

재난 발생은 우리나라뿐만 아니라 전 세계적으로 더욱 빈번하게 발생하고 있으며, 도시화의 증가로 인구밀집도가 높아지면서 피해의 규모도 커지고 있다. 2020년10월 유엔 재난위험경감사무국(United Nations Office for Disaster Risk Reduction, UNDRR)은 2000~2019년 재해 보고서에서 지난 20년간 기후변화로 인한 자연재난이 엄청나게 증가했다고 보고했다.

9) 전국 각처에 지진이 있다. 조선왕조실록. https://sillok.history.go.kr/id/kpa_12106009_001



[그림 1-1] 세계 재난 피해 규모

보고서에 의하면 [그림 1-1]과 같이 7,348건의 자연재난이 발생해서 123만 명이 사망하고 40억 3천만 명이 피해를 봤으며 경제적 손실 규모는 약 2.97조 달러에 달했다. 이는 1980년부터 1999년 사이에 4,212건의 자연재난으로 119만 명이 사망하고 32억 5천만 명이 피해를 봤으며, 1.63조 달러의 경제적 피해가 발생한 것에 비하면 거의 1.7배 증가했다.

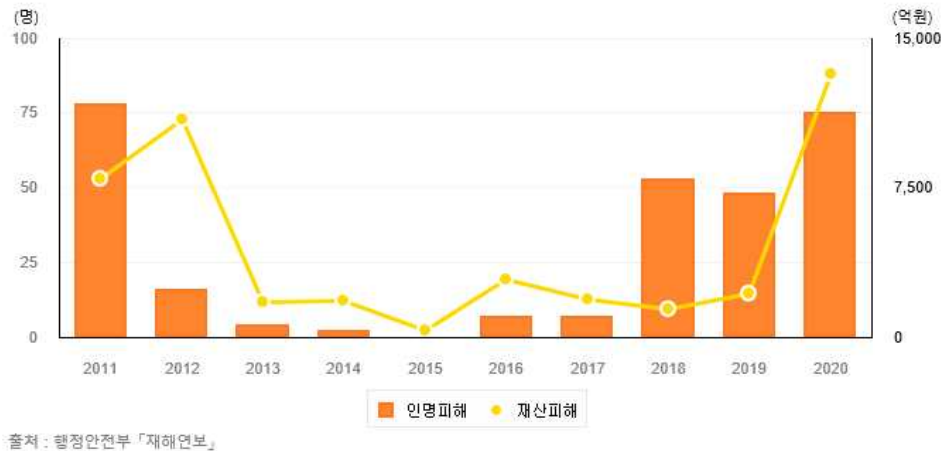


[그림 1-2] 재난 유형별 발생 비율

이러한 변화의 원인은 대부분 기후변화와 관련된 재난의 증가로 인한 것으로 지난 20년간 대홍수가 두 배 이상 발생했으며 [그림 1-2]와 같이 대부

분 홍수와 태풍에 의한 피해였다. 이러한 재해 발생의 원인은 기후변화를 막지 못하고 적응해 가면서 온실가스 감축에 실패하면서 나타난 결과라고 전문가들은 진단하고 있다.¹⁰⁾

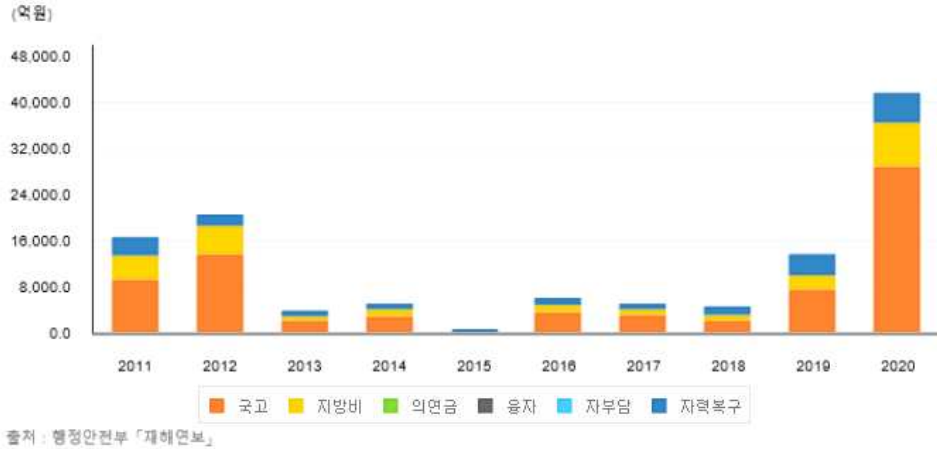
우리나라 자연재해의 경우 행정안전부의 재해연보를 통해 공개된 통계를 보면 [그림 1-3]과 같은 추세를 보이고 있으며,¹¹⁾ 사회재난도 [그림 1-5]와 같이 자연재난과 유사한 추세를 보이고 있다.



[그림 1-3] 자연재난 발생 현황

10) UNDRR. (2020). The human cost of disasters: an overview of the last 20 years (2000-2019). Centre for Research on the Epidemiology of Disasters UNDRR. <https://www.undrr.org/publication/human-cost-disasters-overview-last-20-years-2000-2019>

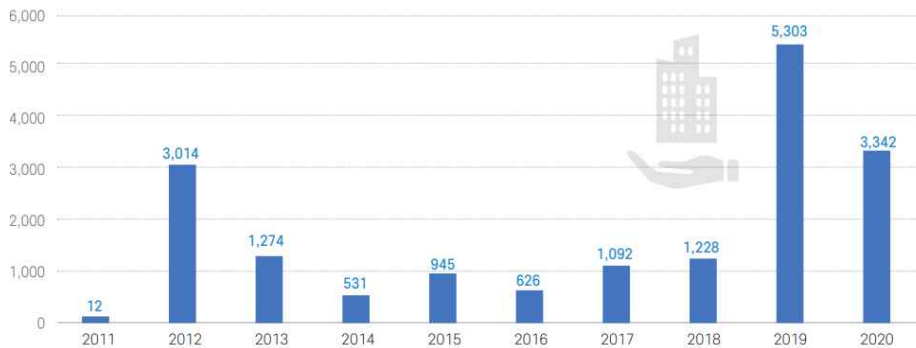
11) 기후변화와 에너지. e-나라지표. (2021). https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1628



[그림 1-4] 최근 10년간 재원별 복구비 현황



[그림 1-5] 사회재난 발생 현황 (단위: 건)



[그림 1-6] 사회재난 재산 피해 현황

제 2 절 연구의 필요성

이러한 재난을 통제하기 위해 재난관리를 시행하는데 이는 재난으로 인한 피해를 최소화하기 위한 제반 활동을 의미한다. 재난으로 인해 발생한 손실의 피해와 발생 정도는 불확실하기 때문에 ‘불확실성’을 제거하는 것이 재난관리의 핵심이다. 특히 오늘날에는 이러한 불확실성이 더욱더 높아지고 있다. 불확실성이란 현재의 위험이 재해로 이어질지 아닐지, 만약 재해로 이어진다면 언제, 어디서, 누구에게, 얼마나 큰 피해가 나타날지 예측할 수 없다는 것을 의미한다. 위험은 피해의 결과가 드러날 때까지 아무도 모른다는 것이다.

우리나라의 경우 재난 및 안전관리 기본법에서 재난관리를 예방, 대비, 대응, 복구의 4단계로 구분하고 있다.¹²⁾ 그 배경은 페탁(Petak, William J.)의 4단계 재난관리 모형을 기반으로 하였으며 내용을 크게 보면 재난이 발생하기 이전 단계인 예방, 대비와 발생 후의 단계인 대응, 복구로 구분할 수 있다.¹³⁾



[그림 1-7] 재난의 4단계

12) 재난관리, 네 가지 단계를 아시나요? 기상청블로그. (2016).

https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=kma_131&logNo=220594792032

13) Petak, William J. (1985). Emergency Management : A Challenge for Public Administration. *Public Administration Review*, Vol. 45, 3-7.

먼저 [그림 1-7]의 예방단계(Mitigation)는 재난의 부작용에 대한 완화를 말한다. UN의 자연재해 경감을 위한 국제적 전략기구인 UNISDR(United Nations International Strategy for Disaster Reduction)에 따르면 선제적 예방 활동을 통해 잠재적인 부작용을 완전히 회피할 수 있는 개념과 의도를 표현하고 있다. 두 번째 단계인 대비단계(Preparedness)는 재난 및 재해로 인한 피해를 대비하기 위한 재난 훈련들과 국민 인식 제고 캠페인 등이 포함된다. 세 번째 단계인 대응단계(Response)는 응급상황이나 재난이 발생한 이후에 해당하는 단계로서 이 단계에서는 인명피해를 비롯한 사회적, 경제적 피해를 줄이기 위한 활동들이 포함된다. 마지막 복구단계(Recovery)는 재난으로 인한 피해를 줄이기 위한 모든 활동을 수행하는 단계로서 재난으로 피해를 입은 피해자와 지역사회의 정상화를 위해 노력하는 단계를 의미한다. 이와 같이 재난의 4단계는 독립적으로 구분되지 않고 서로 연계된 활동이라고 볼 수 있다.

현대사회는 건축기술 및 ICT 기술의 발달로 다양한 요구를 충족하기 위해 건축물도 각기 다른 비정형 형태로 건설되고 있다. 복잡한 형태의 멀티플렉스 건물과 초고층 건물 등의 많은 랜드마크 건축물들이 건설되고 있다. 하지만 건물의 층고가 높아지고 구조가 복잡해짐에 따라 건물에서 발생할 수 있는 재난의 종류도 다양해지고 있다. 이에 대한 대응이나 대피에 있어서 재난 상황을 불특정 다수에게 전달하는 가장 경제적이고 효과적인 방법은 전관방송 시스템을 통한 대피 방송을 하는 것이다.

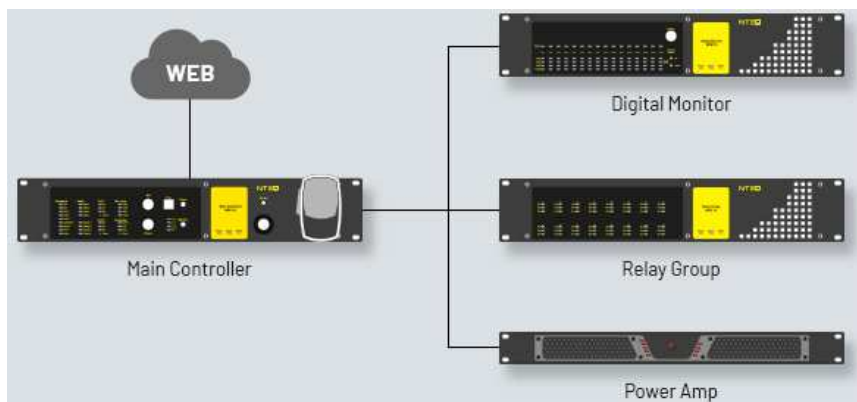
국민재난안전 포털에서 안내하는 사회재난 행동요령 중 화재 발생 시 불을 발견한 경우 제일 먼저 해야 하는 행동으로 “불이야!”라고 소리치거나 비상벨을 눌러 주변에 알리는 행동을 최우선으로 해야 한다고 안내하고 있다.¹⁴⁾ 비상벨 설비는 화재나 비상 상황 발생 시 경종으로 경보하는 설비로서 주로 방재실에서 전관방송 시스템과 연동되어 미리 지정된 지역에 비상방송도 송출하게 되어있다.

전관방송 시스템이란, 불특정 다수에게 일방적으로 음향 정보를 전달하는 안내방송 시스템이며, 다른 이름으로는 구내방송 시스템이라고도 하며 평상시에는 배경음악이나 안내방송에 사용된다. [그림 1-8]과 같은 구성으로 되어

14) 사회재난행동요령 화재. 국민재난안전포털.

<https://m.safekorea.go.kr/idsiSFK/neo/sfk/cs/contents/prevent/SDIJKM5116.html?menuSeq=127>

있으며 사용되는 곳은 업무용 빌딩, 공동주택, 근린생활시설, 공공기관, 학교 등 다양한 건축물로 평소에는 음향으로 홍보, 안내문 등을 전달하는 수단으로 활용되다가 화재나 긴급상황 시 비상방송 시스템으로 전환되어 시스템을 통해 기존 녹음된 음성이나 사이렌을 재생하여 화재 발생 사실을 통보하는 역할을 한다.¹⁵⁾



[그림 1-8] 전관방송 시스템 기본 구성도

상황을 파악하기 힘든 경종보다는 음성안내를 통하여 신속하고 효율적인 피난을 도와주어 생명과 안전을 지키는 장치로 사용된다. 하지만 이러한 방송 설비도 재난을 예방하는 시스템이 아니라 재난이 발생한 이후에야 동작한다. 이는 재난의 4단계 중 사후 단계인 3단계 대응단계에 해당하여 피해 발생을 피할 수는 없다.

재난의 경우 재난의 4단계 중 단계가 앞설수록 여러 장소에서 재난에 대비하고 그 영향 및 피해를 줄일 기회가 늘어난다. 현재까지의 시스템들은 대응 위주로 동작하는 시스템들이 주류였지만 발생 가능한 재난의 종류 및 장소를 예측하여 그러한 일이 발생할 때 피해를 최소화하는 활동이 필요하다. 이를 위해 공공데이터 및 재난 포털 등을 통해 데이터를 분석하고 경고문을 미리 인식하여 대비를 위한 안내를 해주는 시스템이 현대사회에는 더욱 필요하다.

15) Ryu, T. & Kim, S. (2020). Redundancy operation method for a distributed public address system. *Journal of the Korea Convergence Society*, 11(8), 47-53.

이태원 헬러윈 참사와 관련해서 향후 재발 방지 방안으로 질서유지를 위한 경찰 인력 확보나 행사 진행요원들의 일방통행 유도 등의 방법도 있지만, 인구가 밀집된 행사장이나 지역에 설치된 CCTV를 통해 수집된 영상정보를 딥러닝 기술로 분석하여 밀집도를 판단하고, 더불어 통신사의 기지국 데이터와 교통카드 승하차 인원 통계 등의 빅데이터를 참고하여 지하철이나 버스와 같은 대중교통 시스템이 해당 지역을 무정차나 우회 통과할 수 있게 하여 인파를 분산하는 방법을 사용할 수 있다.

또한, 행사장이나 지역에서 자연어 처리를 통한 경고 방송용 문구를 자동으로 생성하고 고출력 앰프+스피커 또는 IP 스피커를 CCTV와 함께 설치하여 경고 방송을 송출하면 위험 상황에 대한 경고나 신속한 전파가 되어 사고 예방에 도움 될 것으로 기대된다. 물론 이와 같은 일을 처리하기 위해서는 법 개정도 함께 추진되어야 한다.

제 3 절 연구의 목적 및 방법

1) 연구의 목적

재난은 종류에 따라서 피할 수 있는 재난과 피할 수 없는 재난이 있다. 자연재난과 같은 피할 수 없는 경우에는 피해를 최소화하는 방법을 찾아야 하고, 사회적 재난의 경우에는 예방과 대비에 집중하여야 한다. 본 연구는 근본적인 재난에 대한 대응보다는 최소한의 비용으로 피해를 최소화하기 위한 안내방송 방법에 관한 연구이다. 이를 위해 공공데이터와 웹상의 데이터를 분석하여 제공되는 정보를 기반으로 적정한 시기에 대비를 위한 문구를 실시간으로 자동 생성하여 안내하는 방법을 연구하여 제안하고자 한다. 기존의 비상방송 설비의 경우 재난의 4단계 중 발생 이후인 대응을 위한 장비였다면, 본 연구의 결과물은 재난 발생 이전 단계인 대비 방법에 대한 추가적 연구로 재난에 대한 선제적 대응 방안을 제공하는 것을 목적으로 한다.

재난이 우려되는 상황에 대해 미리 안내하는 방법에는 재난문자전광판을 통해 안내되는 시각정보와 전광방송 시스템의 스피커를 통해 안내되는 청각 정보, 그리고 TV, 스마트폰 등의 다양한 안내 채널이 존재한다. 본 연구에서는 새로운 하드웨어 장비를 추가하는 방식이 아닌 기존 전광방송 시스템을 이용하여 많은 추가 장비나 시간을 들이지 않으면서 효과적인 안내를 하는 것이 목적이지만, 재난문자전광판이나 스마트폰 앱 등을 통한 전달과 같은 방법으로 응용도 가능하다.

재난 뉴스를 불특정 다수에게 알리는 데 있어서 뉴스 기사의 길이가 너무 길면 글쓴이가 아닌 일반인의 시각에서 볼 때 내용을 이해하기 어렵기 때문에 뉴스 기사를 줄여서 정리할 필요가 있다. 정확한 정보의 전달과 설득력 있는 글을 위해서는 간결체가 적절하다.¹⁶⁾ 재난정보를 알리는 데 있어서 정보의 신뢰성이 확보되어야 하고, 안내하는 메시지의 품질도 중요하며, 상황별로 효과적인 안내 방법이 정의되어야 한다. 정보의 전달 시 TTS(Text to

16) 박선흥, 배은미, 이창환, 동아일보 미디어연구소 종합심의회. (2008). 수록처: 『신문 글 바꾸기: 올바른 기사 문장 작성법』. (10). 서울: LG상남언론재단.

speech)를 통해 음향정보로 전달하는 데 있어서 청자들이 집중하고 이해하기 쉬운 짧은 글로 요약할 필요가 있다.

2) 연구의 방법

본 연구는 하나의 논문으로 정리되었으나 연구의 대상은 다수가 존재한다. 연구목적의 달성을 위해서 수행한 연구의 종류 및 방법은 다음과 같다.

먼저, 선행연구들에 관한 분석으로 기존 재난관리 시스템 중 인공지능 기술을 활용한 재난 안전 기술들과 국내외 재난방송 시스템 그리고 비상방송 시스템의 기술 동향 및 기존 연구들이 놓친 문제점들이 무엇인지 살펴보았다. 그리고 공공데이터와 Open API를 활용한 재난대응과 관련된 내용으로 기존 연구사례를 분석하고 실제 공공데이터를 이용하여 재난에 대비가 가능한 시스템을 구현한 사례를 분석하였다. 그리고 다양한 자연어 처리 모델을 검토하여 실험에 사용할 모델들을 선정하였다.

다음은 연구한 내용으로 Open API를 통해 가져온 공공데이터와 더불어 재난과 관련된 재난 포털 등의 웹 문서로부터 웹 크롤링(Web Crawling)을 통한 텍스트 마이닝(Text Mining)을 진행하였다. 그래서 추출된 문서를 자연어 처리를 통해 안내하기 적절한 문장으로 요약 및 생성하는 방법을 연구하였다. 여기서 텍스트데이터의 수집 채널로는 국가 재난 포털 사이트 및 언론사(KBS, 연합뉴스)의 재난과 관련 기사가 대상이 된다.

마지막으로 요약된 문서의 품질에 대한 실시간 평가방법으로 기존과는 다른 새로운 방법을 연구하였다. 기존에는 평가를 위한 모범답안이 따로 필요하였지만 본 연구에서는 사람이 작성한 모범답안(참조본) 없이 평가하는 방법을 제안하였다. 그 외에 자연어 처리를 통해 생성된 안내방송 콘텐츠의 송출 여부를 자동으로 판단하고 실행하는 시스템 구현을 소개할 예정이다.

제 2 장 관련 기술 및 선행연구 분석

제 1 절 재난관리 시스템 기술 현황

본 절에서는 인공지능 및 IoT 등의 정보기술을 활용한 재난관리 시스템 및 국가별 재난방송 시스템의 현황, 그리고 다양한 형태로 연구 중인 대피방송 시스템들의 연구사례를 검토해 보았다. 연구자의 연구 내용이 인공지능 기술을 적용하여 재난에 미리 대비하는 방법에 관한 것이기 때문에 기존 연구와 중복되는 부분은 없는지 찾아보았다.

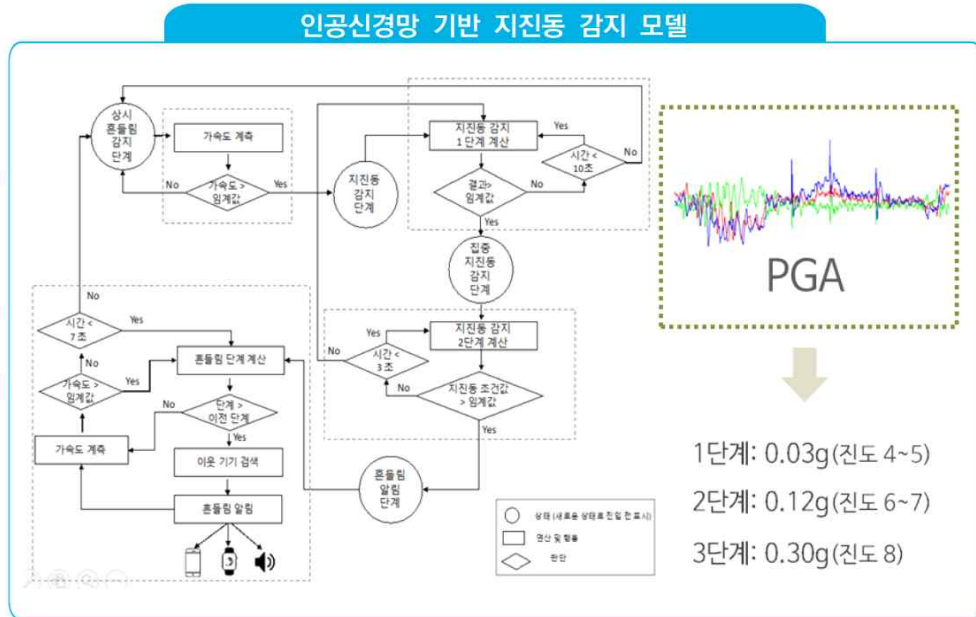
1) 인공지능을 활용한 재난안전기술¹⁷⁾

가) 국내 기술동향

(1) 인공신경망 모델을 활용한 지진동경보기 (국립재난안전연구원)

지진동을 알리고 맞춤형 정보를 적시에 제공하여 위해 개발되었으며, 저가형 가속도 센서(MEMS 센서)와 인공신경망(ANN) 모델을 활용하여 지진동 감지 모델 개발하였다. 규모 4.0 이상의 지진 발생 시 지진동 경보기가 지진동을 감지하고, 주변의 IoT 기기(스마트폰, TV, 스피커, 전기 및 가스 장치) 등과 연계하여 흔들림의 단계별 지진정보 및 대피 정보를 제공한다.

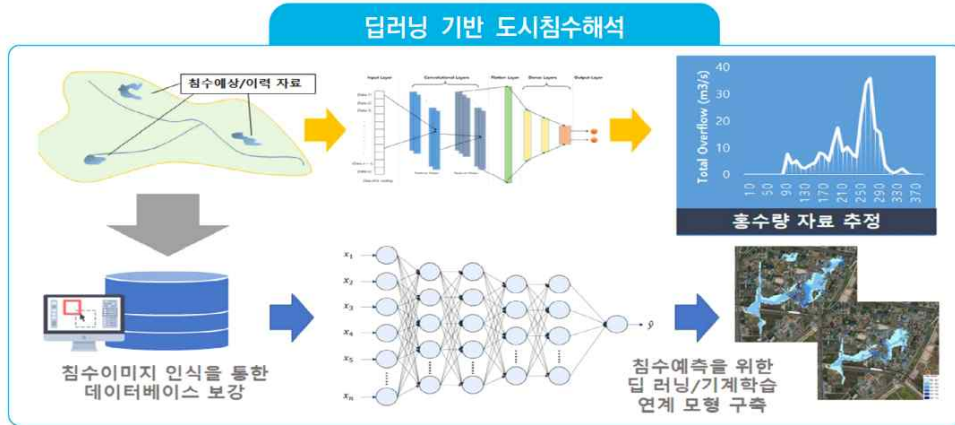
17) 국립재난안전연구원. (2022). 재난안전분야 국내외 인공지능 기술동향. 『재난안전연구데이터』, 2022(16).



[그림 2-1] 인공신경망 기반 지진동 감지 모델

(2) 딥러닝 기반 실시간 침수 예측 (국립재난안전연구원)

침수 예상지역에 대한 현행자료는 단순 이미지 및 침수심, 면적에 대한 정보만 제공하고 있어, 검증된 자료지향형 실시간 예측기법이 필요하다. 실시간으로 홍수 피해 및 범위를 예측하는 알고리즘 개발 및 초단기 강우예측 정보의 정도 향상을 위해 딥러닝 기술을 활용하여 홍수 발생 이전에 정확한 침수구역을 사전 예보한다.



[그림 2-2] 딥러닝 기반 도시침수해석

(3) 꽃가루 농도 위험지수 (기상청 날씨누리)

2017년 4월 1일부터 인공지능을 이용하여 봄철 꽃가루 알레르기 유발 위험도를 알려주는 서비스로 기상조건에 따른 꽃가루 알레르기 발생 가능 정도를 지수화했다. 과거 16년간 전국 10개 지점에서 관측된 꽃가루 관측 자료를 딥러닝 방식을 활용하여 위험 예측률을 15.9%에서 69.4%로 향상시켰다.

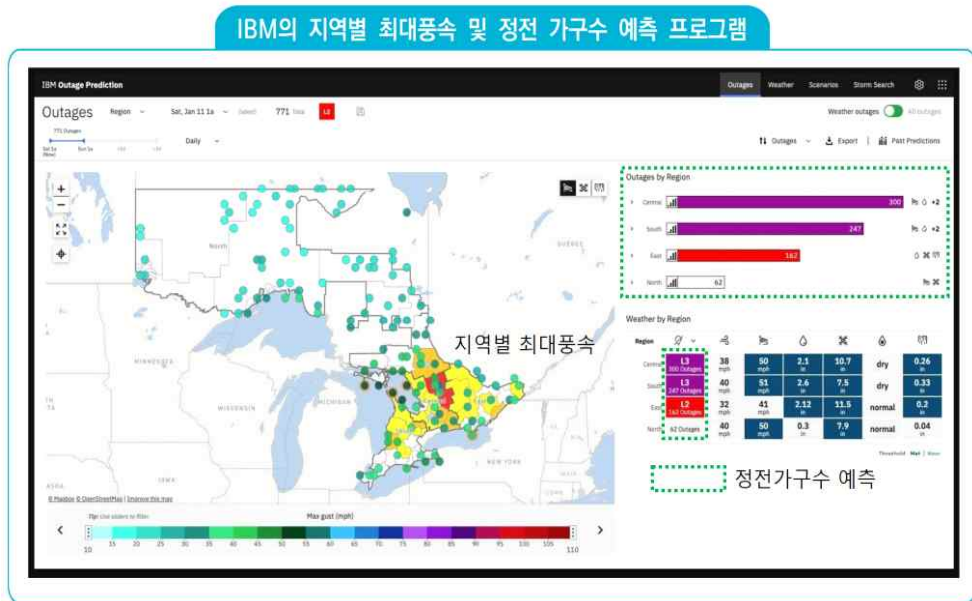
나) 국외 기술동향

국외의 경우, 민간에서 재난 안전 분야에 인공지능을 활용하고 있으며, 기상, 자연재해에 대한 예측이나 분석기술 개발이 활발하게 이루어지고 있다.

(1) Environmental Intelligence Suite : 정전예측 (IBM)

날씨로 인한 정전은 오랫동안 기업들에게 큰 문제로 인식되고 있다. 이에 IBM은 머신러닝 예측모델을 이용하여 최대 72시간 전에 날씨에 의한 정전을 예측함으로써 복구 작업을 최소화하고 가동중단을 최소화하는 방법을 제공하고 있다. Forecast Watch 날씨 데이터와 함께 AI, IoT 분석 기술을 결합하여

32종류의 다양한 폭풍 데이터를 학습한 AI 모델을 개발하였다. 날씨로 인한 정전의 신속한 예측과 즉각적인 대처를 통하여, 정전으로 발생하는 인명피해 및 기업 손실을 최소화할 수 있다. 해당 AI 모델을 활용하여 캐나다 뉴브런즈윅의 눈 폭풍을 예측하였으며, 정전 발생 가구 90% 이상을 24시간 만에 복구 완료시켰다.

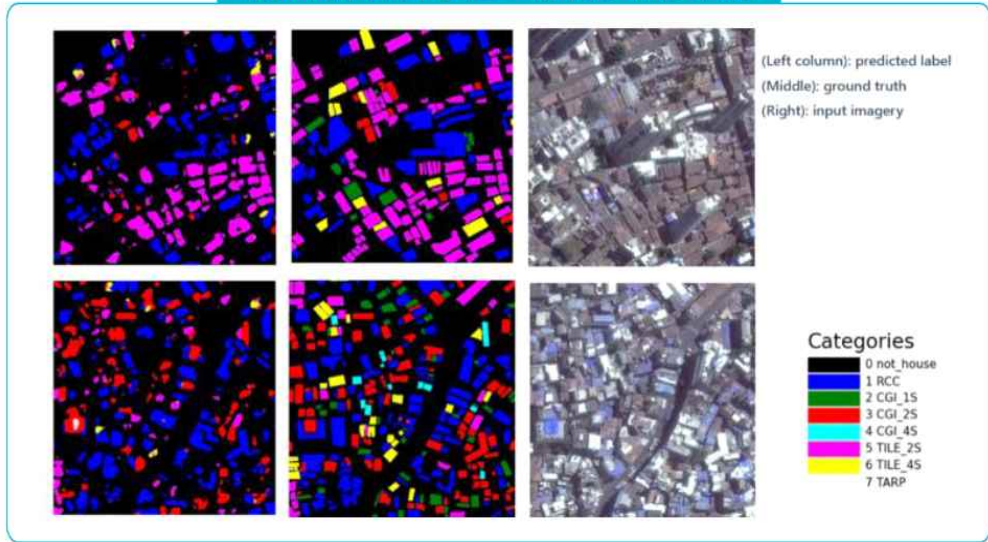


[그림 2-3] 지역별 최대풍속 및 정전 가구수 예측 프로그램

(2) India Sunny Lives : 사이클론&폭염위험 예측모델 (Microsoft)

심화되는 기후변화에 대응하여 도시 핵심 폭염 지역의 폭염 위험 예측 2 단계 모델을 개발하였다. 인도지역의 고해상도 위성 이미지, 주거 평가 및 건물평가데이터를 수집하여 AI모델을 개발하였다. 취약한 지붕을 식별할 수 있는 AI모델을 개발하여 고해상도 이미지와 매칭시켜 50,000개 이상의 주택에 수동으로 태그를 지정, 지붕을 7가지 범주로 나누었다. 또한, 도로망, 수역과 인접한 지역, 고도, 초목 등 다양한 레이어를 통해 모든 집에 대한 위험점수를 판단하는 기계학습 파이프라인을 생성하였다.

위성이미지와 90%의 정확도로 지붕타입을 식별할 수 있음



[그림 2-4] 지붕 타입 식별 모델

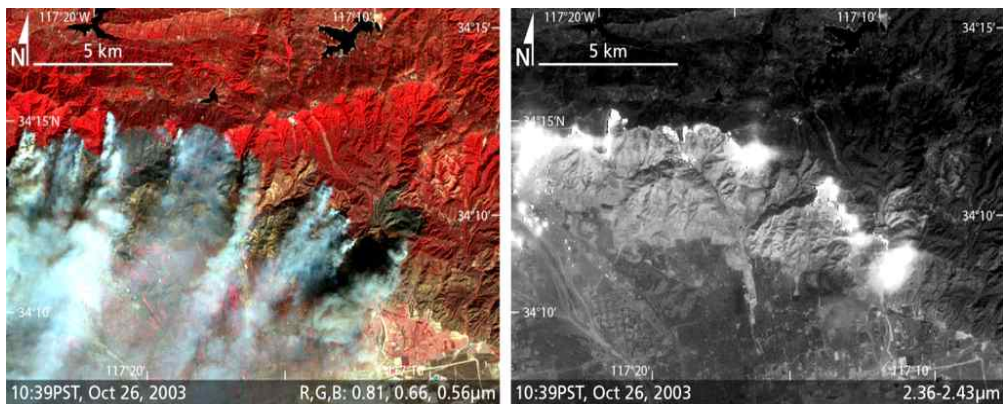
(3) Alibaba DAMO Academy : 물 인식 알고리즘 (Alibaba)

빗물로 급격히 불어난 하천물의 양과 수위를 원격으로 감지하고, 하천 주변 주택지의 대피를 계획할 수 있게 인공지능기술을 활용하여 홍수 피해를 최소화할 수 있다. 기존 위성 영상을 물, 경작지, 주택, 댐, 양식 설비 등을 AI 인식모델에 적용하여 분석하였다. 물의 양과 구름과 안개, 강우량 등 데이터를 종합적으로 분석하기 위해 지상 감시용 군용항공 레이더로 감측하는 기술인 SAR(Synthetic Aperture Radar)를 이용하여 원격 감지 이미지 생성 기술을 결합하였으며, 중점 경계 수위 지역에서 영상처리 수를 5배 늘리고, 영상분석속도를 100배 이상 높임으로써, 전이학습 기술을 적용해 SAR의 영상 분석에 AI 알고리즘을 접목하여 기능을 개선하였다.

2) 산불 예방을 위한 첨단 기술 동향¹⁸⁾

가) 인공위성을 활용한 산불 감지

미국 캘리포니아대학에서는 2013년부터 인공위성을 이용하여 산불을 감지하는 프로그램을 운영중에 있다. ‘지구 정지궤도 내 긴급 화재 추정장치(FUEGO: Fire Urgency Estimator in Geosynchronous Orbit, 스페인어로 불)’로 명명된 이 시스템은 산불로 발달할 수 있는 지점을 위성으로 확인한다. 만약 산불로 확대될 것으로 판단되면[그림 2-5], 경보를 울려 지상에 있는 소방대원에게 위치를 전송하여 산불이 크게 확산되기 전에 대응할 수 있도록 한다.¹⁹⁾

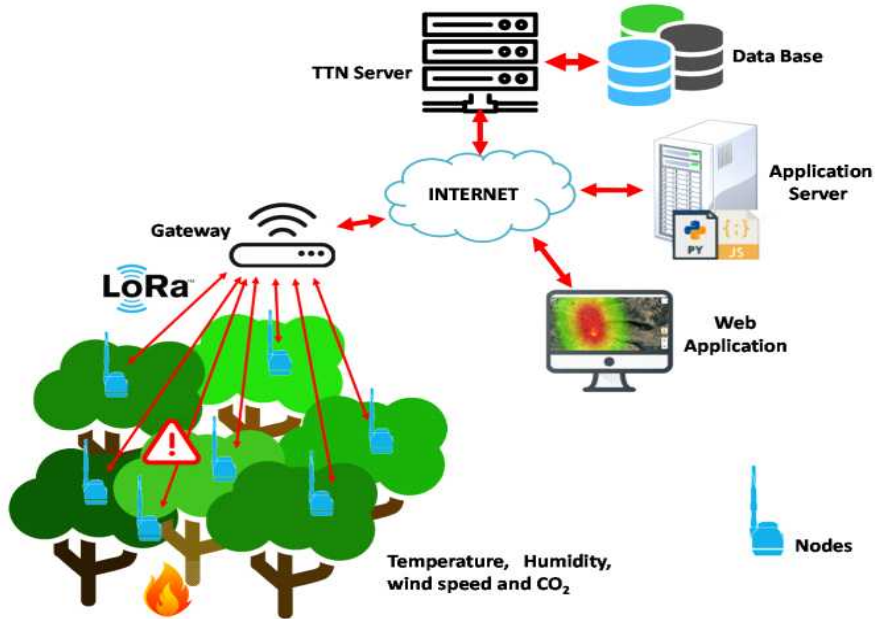


[그림 2-5] ASTER 센서로 확인된 미국 그랑프리(Grand Prix) 산불

18) 김태범, 조시범, 김다진솔. (2020). 산불 예방 및 대응을 위한 첨단기술. 『국립재난안전연구원 재난안전』, 22(4), 79-86.

19) Pennypacker, C. R., Jakubowski, M. K., Kelly, M., Lampton, M., Schmidt, C., Stephens, S., and Tripp, R., (2013). FUEGO.- Fire Urgency Estimator in Geosynchronous. *Remote Sens*, 5(10), 5173-5192.

나) IoT 기반의 센서를 활용한 산불 예방



[그림 2-6] LoRa 네트워크 구성도

산불이 시작될 때 가장 두드러지게 변화하는 환경요인은 바로 화재로 인해 특정 지점의 온도가 급속도로 올라가거나 주변 공기에 이산화탄소(CO₂) 발생량이 증가하는 것이다. 이러한 환경 변화를 사전에 감지한다면 산불로 발달하기 전에 진화할 수 있을 것이다. 이 같은 센서들을 상시 작동할 수 있도록 전력 공급을 계속한다는 것은 과거에는 비현실적이었다. 그러나 IoT 기술의 발달과 더불어 상시적 전력 공급이나 저전력으로 운용할 수 있는 각종 기술이 개발되면서 이제 산불 감지를 위한 계측망 구성이 가능해지고 있다[그림 2-6]. 특히, 저전력 광대역 네트워크(LPWAN: Low Power Wide Range Network) 중장거리 통신이 가능한 LoRa(Long Range)는 15km 이상의 통신 거리, 10년 이상의 배터리 수명 그리고 저렴한 모듈의 비용 등을 고려했을 때 산불 감지 시스템 구성을 위한 이상적인 시스템이라고 할 수 있다.²⁰⁾

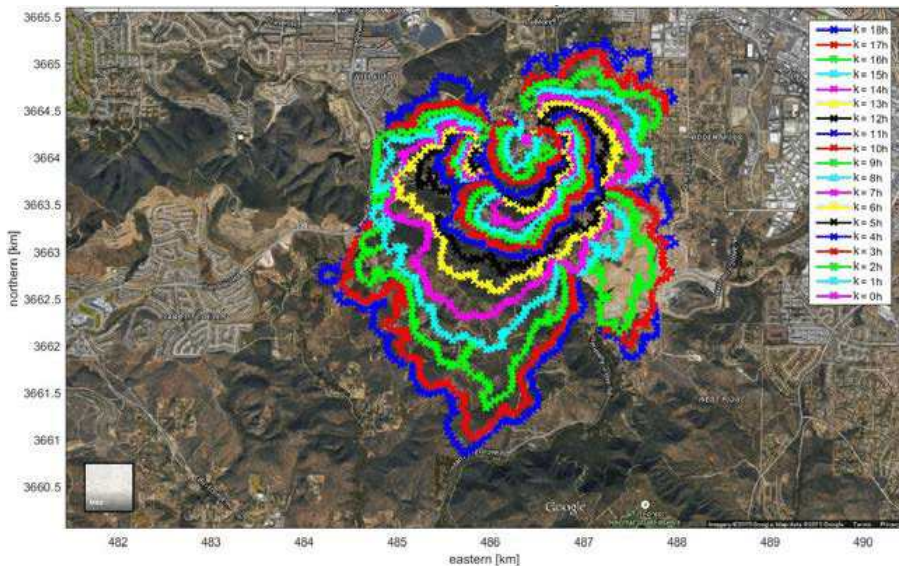
20) Sendra, S., García, L., Lloret, J., Bosch, I., and Vega-Rodríguez, R., (2020). LoRaWAN Network for Fire Monitoring in Rural Environments, *Electronics*, 9(3), 531.

3) 재난 예측 기술 개발 동향

가) 산불 확산 예측

캘리포니아를 비롯한 미국 일부 지역은 기후적 영향으로 대형산불이 빈번하게 발생하고 그 규모도 해가 갈수록 커지고 있다. 미국에서는 이러한 대규모 산불로 인한 인명 및 재산 피해를 줄이기 위해 산불 감지 및 예측 기술 개발을 위해 노력하고 있다.

WIFIRE 랩(샌디에이고 슈퍼컴퓨터센터, 캘리포니아대 샌디에이고캠퍼스, 쉐컴 연구소 등이 참여)에서는 빅데이터와 인공지능 기술을 활용하여 화재에 대한 예측 모델링, 실시간 화재 예측 및 화재 확산 시나리오 분석이 가능한 FireMap 프로그램을 개발하였다. WIFIRE의 화재 확산 예측모델은 널리 사용되는 화재 모델링 프로그램인 FarSite를 사용하고 있다. [그림 2-7]는 FarSite를 사용한 산불 확산 예측의 예시이다.²¹⁾



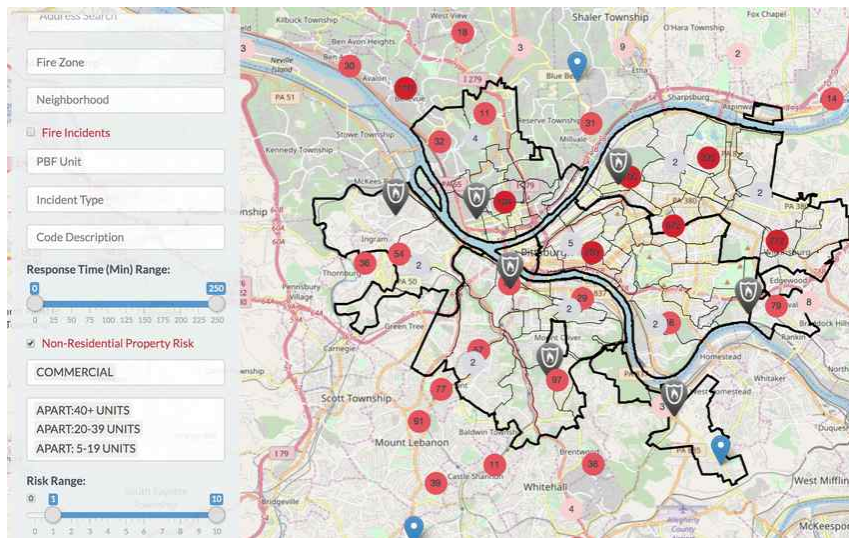
[그림 2-7] WIFERE 산불 확산 예측 지도

21) 박소영, 홍상기, 이강복. (2020). 재난예측 기술 개발 및 서비스 제공 동향. 『전자통신동향분석』, 35(1), 80-88.

나) 건축물 화재 위험도 예측

건축물 화재의 발생을 예측하기 위해 건축물과 관련된 다양한 요인이 화재 위험에 미치는 영향에 대한 연구도 진행되고 있다. 카네기멜론대학(Carnegie Mellon University)의 프로젝트의 일환인 ‘Metro21: Smart Cities Institute’은 건축물 화재 위험도 예측을 위한 모델을 연구하였다. 이 프로젝트에서는 예측 대상인 건축물에 대해 6개월 이내 화재 발생 가능성을 예측하는 모델을 개발하였으며, 예측 결과 도출된 화재 위험도는 1(낮은 위험)에서 10(높은 위험)까지의 숫자로 표현하였다.²²⁾

[그림 2-8]은 건물별 화재 위험 예측 결과를 조건에 따라서 조회한 결과를 시각화하여 보여준다.²³⁾ 이 모델은 현재 최신 화재 및 건물 데이터를 사용하여 화재 위험도(Risk score)를 업데이트하는 피츠버그 소방국(Pittsburgh Bureau of Fire)의 서버에 적용되었다.



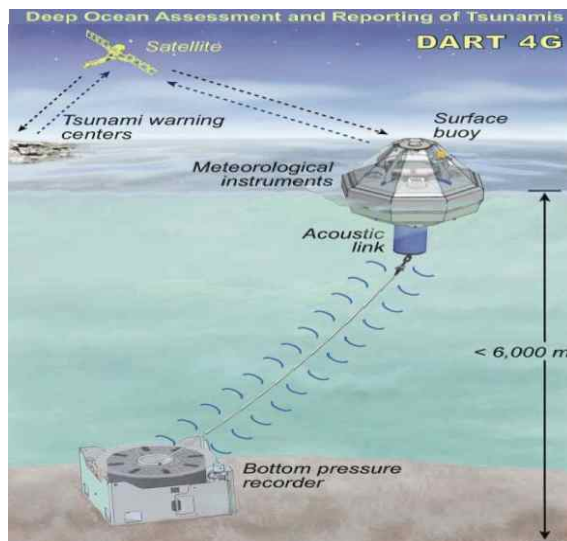
[그림 2-8] 건축물 화재 위험도 예측 결과

22) 박소영, 홍상기, 이강복, 상계논문.

23) Madaio, Michael A. (2018). Predictive Modeling of Building Fire Risk: Designing and Evaluating Predictive Models of Fire Risk to Prioritize Property Fire Inspections. Carnegie Mellon University, CC-BY-SA 3.0 IGO

다) 쓰나미 예측

태평양 쓰나미 경고센터(PTWC: Pacific Tsunami Warning Center)는 미국의 해양대기청(NOAA: National Oceanic and Atmospheric Administration)이 운영하는 지진해일 경고센터 가운데 하나로, 이곳은 미국 지진해일 경고 시스템(TWS: Tsunami Warning System) 프로그램의 일부로, 태평양 인접국과 그 밖의 나라에 대한 쓰나미 관련 정보를 제공하고 쓰나미 발생 시 경보를 발령하는 역할을 담당한다.²⁴⁾ NCTR(NOAA Center for Tsunami Research)에서는 쓰나미 경보를 위한 예측을 한다. 쓰나미 예측 목적의 해저 모니터링 장비로 DART 4G System이 있다. DART(Deepocean Assessment and Reporting of Tsunami) 시스템은 해저의 바닥 압력(Bottom pressure)을 측정하여 쓰나미를 예측하는 시스템이다([그림 2-9] 참고). DART 4G 시스템은 미국 오리건주 해안과 칠레 중부에 테스트를 위해 설치되어 있다.²⁵⁾



[그림 2-9] DART 4G System

24) 박소영, 홍상기, 이강복, 전계논문.

25) Acharya, R. (2018). Evolution of DART technology and development of fourth generation bouy system for deep ocean assessment and reporting of tsunami - (DART 4G). *Indian Journal of Geo Marine Sciences*, 47(3), 519-521.

라) 국내외 지진 조기 경보

지진 발생에 대한 예측은 거의 불가능하기 때문에 예측보다는 지진 발생 초기에 지역별로 지진의 규모를 예측하고 조기 경보시스템을 구축하는 것이 보다 효과적이라고 판단되어, 많은 국가에서 지진 조기 경보시스템 구축에 주력하고 있다. 미국지질조사국(USGS: United States Geological Survey)과 미국연방재난관리청(FEMA: Federal Emergency Management Agency) 등이 협력해서 지진 조기 경보시스템인 Shake Alert를 개발하였다. WEA(Wireless Emergency Alerts) 시스템 또는 MyShake 스마트폰 앱을 통해 지진정보 서비스를 제공하고 있다. 구글은 딥러닝 등 AI 기술을 활용한 여진 위치 예측모델을 연구해왔다.²⁶⁾ 일본에서는 세계 최초로 지진 조기 경보시스템인 UrEDAS를 개발하여 신칸센에 적용하였다. 우리나라는 2018년 11월부터 진동의 영향 정도를 지역별로 분류해 알려주는 '진도정보' 서비스를 제공하고 있다. 지진파 및 진동 관측 기준을 충족하는 규모 5.0 이상의 지진에 대해 경보가 발령된다.

마) 서울시 도시홍수예측시스템

서울시는 2014년 도림천 외 지역의 인명피해를 줄이기 위해 웹 기반 홍수 예측시스템을 구축하고, 내외수 연계 침수예측시스템 연구를 통해 내외수 연계 침수예보시스템을 구축했다. 서울시는 2021년 하수관로 성능관리시스템을 통해 침수를 예측할 수 있는 도림천 유역 SWMM 모델을 구축했다. 구축된 도림천 유역의 SWMM 모델을 도림천 침수예측시스템에 적용하면 도림천 유역의 침수예보 및 정보가 가능하다. 현재 서울시는 지속적인 도시홍수예측시스템 연구를 통해 30분 이상의 골든타임을 확보할 수 있는 내수침수를 고려한 도시홍수예측시스템 구축 기술을 확보했다.²⁷⁾

26) DeVries, Phoebe M. R., et al. (2018). Deep learning of aftershock patterns following large earthquakes. *Nature* 560, 632-634.

27) 황성환. (2022). 서울시 도시홍수예측시스템 구축 현황. 『국립재난안전연구원 재난안전』, 24(2), 32-35.

제 2 절 재난방송 시스템 기술 현황

1) 미국의 재난방송 시스템 기술 동향

미국의 재난방송 시스템은 1951년에 처음 시작되었다. 소련의 미국 침공을 전제로 공습경보시스템(CONELRAD: Control of Electromagnetic Radiation)이 도입되어 정부가 직접적이고 신속하게 긴급방송을 보낼 수 있게 되었다. 국가 비상사태가 발생하면 대통령의 지시에 따라 미국의 주요 통신 회사와 모든 라디오 방송국에서 지정된 비상 주파수를 통해 전쟁 발발을 국민들에게 알리게 된다. 그 이후로 CONELRAD는 긴급방송 시스템(EBS: Emergency Broadcasting System)으로 대체되었다.

EBS는 전쟁 상황뿐만 아니라 토네이도와 같은 자연재해나, 테러 상황까지 방송한다. 전국 단위 긴급방송의 경우 EAN(Emergency Action Notification)은 대통령 또는 대통령대리인의 명령으로 전송되며, EAN을 수신한 방송사는 즉시 정규방송을 중단하고 비상상황 내용을 고지할 의무가 있다.

[표 2-1] 참여기관별 역할

참여기관	역할
연 방 재 난 관 리 청 (FEMA)	EAS 개발, 운영, 관리, 수행 및 EAS 관련 지방정부 역할에 대한 계획 수립 및 하달.
연 방 통 신 위 원 회 (FCC)	EAS 기술과 운영 요구사항 및 EAS 재난 메시지 프로토콜 및 테스트 관련 제반 규정 정의(CFR(Code of Federal Regulations) Title 47 Part 11). ²⁸⁾
미국 기상청(NWS)	기상 정보 및 지진, 해일 등의 자연 재난정보 제공.

1997년 EBS는 통합재난관리를 위한 디지털화, 시스템 자동화 등의 고도화를 위해 EAS(Emergency Alert System)로 변경되었다. EAS 구축은 연방재난관리청(FEMA: Federal Emergency Management Agency), 미국 기상청

28) 조용성, 배병준, 김홍묵. (2018). 미국의 재난방송 기술 동향. 『전자통신동향분석』, 33(3), 51-58.

(NWS: National Weather Service), 연방통신위원회(FCC: Federal Communications Commission)가 공동으로 참여²⁹⁾하였으며 참여 기관별 역할은 [표 2-1]과 같다.

EAS의 가장 큰 특징은 디지털화된 EAS 재난 메시지를 정의이다. 여기에는 메시지를 보낸 주체, 발령 사유, 경보의 영향 지역, 메시지 유효기간, 메시지의 최초 발령 시간과 같은 정보가 포함되어 있다. FCC가 고시한 CFR Title 47 Part 11은 재난 상황을 전달하기 위해 방송국 간에 메시지를 교환할 목적으로 정의하고 있는 EAS 재난 메시지 프로토콜이다.³⁰⁾ EAS 메시지의 주요 발령 내용(EEE)은 [표 2-2]와 같다.

[표 2-2] EAS 메시지 발령 내용(EEE) 코드 예시

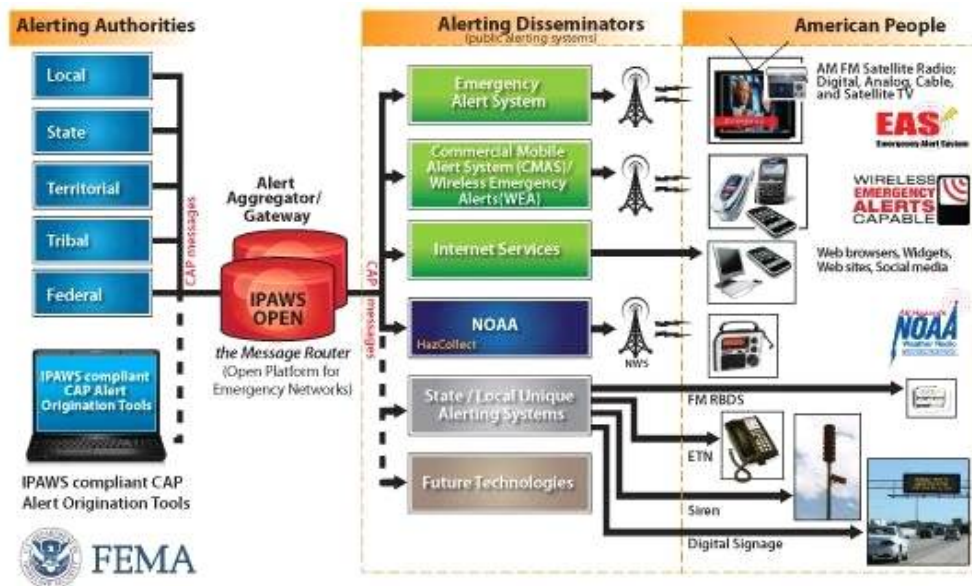
코드	내용
CAE	- Child Abduction Emergency (아동 유괴)
CDW	- Civil Danger Warning(대민피해 경보)
EQW	- Earthquake Warning(지진발생 경보)
EVI	- Evacuation Immediate(즉시 대피 경보)
FFW	- Flash Floow Warning(급박한 범람 경보)
HUW	- Hurricane Warning(허리케인 경보)
SPW	- Special Marine Warning(실내 대피 경보)
TOR	- Tornado Warning(토네이도 경보)

그러나 국가적 재난 상황에 효과적으로 대응하지 못한다는 비판에 대해 미국 정부는 2006년 모든 경보시스템을 통합하고 전국의 모든 매체를 통해 경보 메시지를 전달할 수 있는 통합 경보시스템(IPAWS: Integrated Public Alert and Warning Systems)을 구축했다. 지상파, 케이블, 위성 등 방송사를 통해 경보 메시지를 전송하는 EAS, 휴대폰과 무선단말기를 통해 전송되는 WEA(Wireless Emergency Alert), 미국해양기상청(NOAA: National Oceanic and Atmospheric Administration)의 긴급 재난방송 NWR(NOAA Weather Radio) 등이 IPAWS의 일부로 편입되었다.³¹⁾

29) 조용성, 배병준, 김홍목, 상계논문.

30) FCC CFR Title 47 Part 11. (2022, Nov 02). Emergency Alert System.
<https://www.ecfr.gov/current/title-47/chapter-I/subchapter-A/part-11>

IPAWS의 가장 큰 특징은 공통 경보 프로토콜(CAP: Common Alerting Protocol) 즉 모든 재난정보 발령기관의 공통 프로토콜을 사용한다는 점이다. 향후 새로운 전송 매체가 등장하더라도 CAP 메시지만 수신할 수 있으면 IPAWS에 쉽게 통합될 수 있다. 또한, 기존 시스템과 달리 텍스트 위주의 재난정보 대신 사진, 지도, 동영상 등 리치미디어 형태의 재난정보 제공이 가능하다.³²⁾ [그림 2-10]처럼 모든 재난정보 발령자는 CAP 프로토콜 기반의 재난정보를 전송하고, CAP 메시지를 수신하는 모든 전송시스템은 이를 자체 방식으로 변환하여 일반 대중에게 전송하게 된다.



[그림 2-10] IPAWS 구성도

31) Integrated Public Alert & Warning System. FEMA. (2022, Oct 25).
<https://www.fema.gov/emergency-managers/practitioners/integrated-public-alert-warning-system>

32) Common Alerting Protocol. (2021, Jan 6). FEMA.
<https://www.fema.gov/emergency-managers/practitioners/integrated-public-alert-warning-system/technology-developers/common-alerting-protocol>

2) 일본의 재난관리 및 방송 시스템 현황³³⁾

일본의 재난방송은 주관 방송사(NHK)와 지역의 민간방송사로 구분되어 운영되고 있다. NHK의 경우 법률에 의거해 재난방송으로서 중추적인 역할을 하고 있으며, 지역 민간방송사는 재난지역 주민들에게 맞춤형 방송 서비스를 제공하고 있다. 또한 재난 발생에 대비한 임시 재난방송사를 운영하여 방송이 국민 안전을 위한 역할을 충실히 수행하기 위해 노력하고 있다. 이처럼 일본의 경우 방송 주체를 분리하여 운영하고 있다.

가) 주관 방송사 -NHK

NHK는 1926년 도쿄도(東京都)에 설립된 공영방송국으로서 전국 54개 국에서 지상파 2개 채널, 위성 채널 2개, 라디오 3개 네트워크를 운영하고 있다. 일본 공영 방송인 NHK는 재난 상황에 대비하여 재난 대비 기본법(disaster countermeasures basic act) 제2조 및 대규모 지진 대책 특별조치법 제2조에 의거 재난방송 주관사로 지정되어 재난방송 송출에 관한 책임과 의무가 부과되어 있다.³⁴⁾

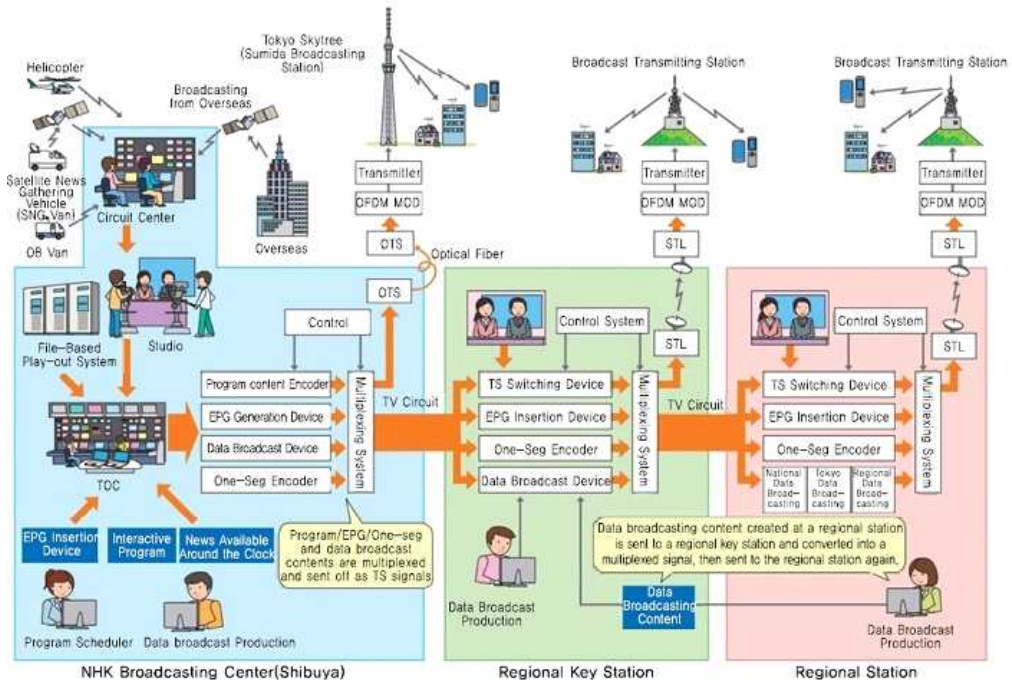
이에 따라 NHK는 재난 발생 시 재난방송을 실시하고 사전에 방재계획을 사전에 세워 총리에게 보고함으로써 재해 상황에 대비하도록 의무화하고 있다. 또한 방재 계획 및 사항을 국민에게 공포해야 할 의무가 있다.³⁵⁾ 신속·정확한 재난방송 전달을 위해 재난 관련 뉴스를 포함해 NHK 본사 및 지역 방송국의 모든 TV프로그램 및 라디오는 TOC(Technical Operation Center)를 거쳐 국민에게 방송하는 체제 [그림 2-11]로 되어있다. 평상시는 자동적인 스케줄 방송을 송출하지만, 뉴스센터를 통해 긴급 재난상황이 들어오면 각 담당별 방송지시담당자, 편성팀, 연락팀, 엔지니어팀이 수동으로 지상파·위성·라

33) 정상구, 박현호, 조정섭, 이용태. (2016). 일본의 재난관리 및 방송 시스템 현황. 『전자통신동향 분석』, 31(3), 112-121.

34) 목하균, (2010). 지상파 재난방송 서비스 현황. 『TTA Journal』, 131, 61-66.

35) 심홍진 외(정보통신정책연구원). (2014). 재난방송 제도 개선에 관한 연구. 『방송통신융합 정책 연구과제 최종보고서』.

디오를 조작해 긴급재난방송을 진행한다.



[그림 2-11] NHK의 TOC 운영 체계

이외에도 재난상황에 대한 정보를 신속하게 제공받을 수 있도록 정부 기관과 협력하고 있으며 지진, 태풍, 쓰나미 등의 재난경보가 발생하면 일본 기상청과 연결된 EWBS(Emergency Warning Broad-casting System)를 통해 재난 상황에 대한 정보를 제공한다. 수신된 정보는 화면과 자막으로 변환되어 즉시 속보로 단말기에 전송된다. EWBS는 NHK 기술연구소에서 표준화하여 분류하고 있으며 전국의 다른 모든 방송사는 표준 설계만 변경하여 [그림 2-12]와 같은 방식으로 EWBS 시스템을 구축하여 운영하고 있다.³⁶⁾

나) 지역방송사-TBC 동북방송

일본의 지역방송사 중 하나인 TBC 도호쿠(東北) 방송은 미야기현(宮城

36) 정상구, 박현호, 조정섭, 이용태, 전계논문.

県) 센다이시(仙台市)에 위치한 대표적 TV 및 라디오 방송사이다. TBC 도호쿠 방송과 같은 지역방송사는 J-ALERT, L-ALERT와 상시 연결되어 재난 상황에 대비하고 재난정보를 실시간으로 수신 및 수집한다. 또한 지자체의 재난대책본부와 소방청, 기상청 등 방재 기관과의 연계를 통해 재난정보를 전달 받아 지역 주민에게 재난 상황을 안내하는 역할을 수행하고 있다. 재난 발생으로 긴급상황이 발생하면 수집 및 제공된 정보를 바탕으로 TV, 라디오, 인터넷, 소셜 네트워크 등을 통해 독자적 판단하에 재난·재해 방송을 실시한다. 동일본 대지진 당시 재난 발생지역의 TBC 도호쿠 방송의 경우 24시간 긴급 TV 및 라디오를 통한 재난방송을 실시하였다.³⁷⁾



[그림 2-12] EWBS 체계

다) 임시 재해 방송사

임시 재해 방송사는 효고현(兵庫県) 남부에서 1995년 1월 17일 발생한 한신·아와지 대지진(阪神·淡路大地震) 당시 효고현이 재난·재해정보 전달을 목적으로 전용 방송국 개설을 요청하여 창설된 제도이다. 임시 재해 방송사는 일본 방송법 제8조에 규정된 ‘임시적이고 일시적인 목적(총무성령으로 정하는 것에 한한다.)에 대한 방송(임시 목적 방송)’ 중 방송법 시행 규칙 제7조 제2항 제2호에 규정하는 ‘폭풍, 호우, 홍수, 지진, 대형 화재, 기타에 의한 재해가 발생했을 경우 그 피해를 줄이기 위해 도움’을 목적으로 하는 방송을 수행할

37) 정상구, 박현호, 조정섭, 이용태, 전개논문.

방송국을 임시 재해 방송사로 지칭한다. 지방자치단체에서 피해 지역을 대상으로 피해주민들의 일상생활 안정기까지 운용되고 있다. 2011년 동일본 대지진 당시 이와테현(岩手県), 이바라키현(茨城県), 미야기현(宮城県), 후쿠시마현(福島県), 아키타현(秋田県)에서 임시 재해 방송사가 운영되었다.³⁸⁾

3) 국내 재난 예·경보 시스템 현황³⁹⁾

과거의 국내 재난정보시스템은 사람이 직접 통보하거나 근거리 사이렌을 울리는 아날로그 방식에서 통신과 방송을 이용한 디지털 방식이 도입되어 전파되고 있다. 현재 대표적인 재난 예·경보 시스템은 재난문자방송(CBS), DMB(디지털 멀티미디어 방송) 재난정보 서비스, 일반 TV 화면으로 전송되어 표시되는 재난온라인방송시스템(DITS) 등이 있다.

가) 재난문자방송(CBS, Cell Broadcasting Service)

재난문자방송은 재난 상황 발생으로 인해 인명 및 재산 피해를 방지하기 위해 CBS(Cell Broadcasting Service) 수신기능이 탑재된 휴대전화로 송출되는 재난문자방송을 말한다.⁴⁰⁾ 재난문자방송 서비스는 특정 지역의 모든 휴대전화 사용자에게 재난 상황을 전송하는 일대다(one-to-many) 서비스이다. SMS(Short Message Service)와 달리 네트워크 부하가 많이 걸리지 않고 신속하게 대량전송이 가능해 국가기관의 재난 상황 보고를 위한 주요 수단으로 활용하고 있다.⁴¹⁾ 관련된 법령으로는 재난 및 안전관리 기본법 제38조의 2(재난 예보·경보체계 구축·운영 등) 및 재난문자방송 기준 및 운영규정(행정안전부예규 제76호)에서 ‘전기통신사업자나 방송사업자, 휴대전화 또는 내비

38) 정상구, 박현호, 조정섭, 이용태, 전개논문.

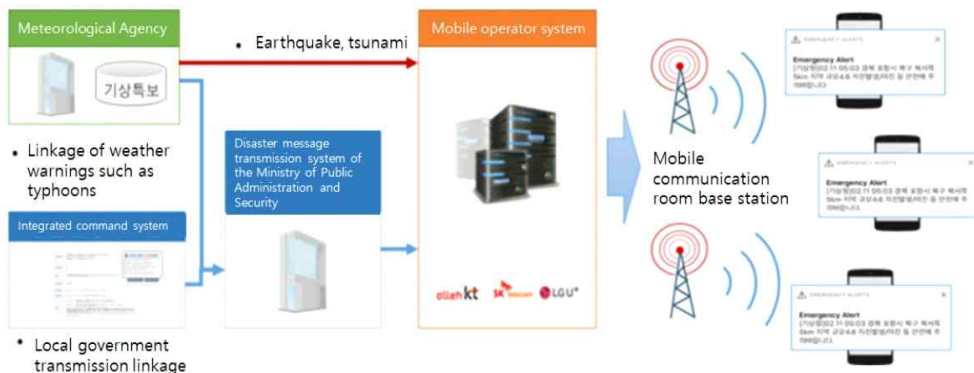
39) 황우석 & 표경수. (2022). 국내 재난 예경보 시스템 현황 및 통합 방안에 대한 분석 연구. 『방송공학회 논문지』, 27(1), 80-90.

40) National Disaster Management Research Institute, Disaster Response and Safety Support Technology Development for Vulnerable Groups. (2020). NDMI-PR-2020-02-01-01.

41) 변윤관, 이현지, 장석진, 최성중, 표경수. (2020). 재난문자 서비스에서 이미지 표출 방안. 『방송공학회논문지』, 25(3), 399-404.

게이션 제조업자는 재난의 예보·경보실시 사항이 사용자의 휴대전화 등의 수신기 화면에 반드시 표시될 수 있도록 소프트웨어나 기계적 장치를 갖추어야 한다.’고 정의하고 있다.

재난문자방송 발송에 대한 국내(TTA: Telecommunications Technology Association)·국제(3GPP: 3rd Generation Partnership Project) 표준이 제정되어 휴대전화로 재난문자방송 수신이 의무화되었다. 재난문자방송 송출은 태풍, 호우, 폭설 등의 기상특보와 연계하여 기상청에서 행정안전부의 재난문자 송신시스템으로 송출하거나 연계된 통합발령시스템에서 송출한다. 이후 재난 문자방송은 이동통신 3사의 시스템에서 해당 이동통신망 기지국으로 송출된다. 그 흐름은 [그림 2-13]과 같다.



[그림 2-13] 재난문자방송 송출체계도

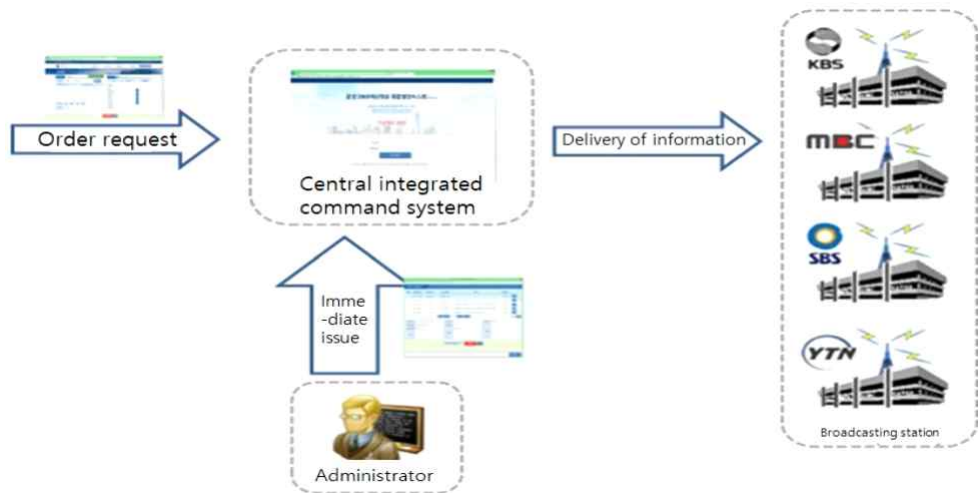
나) DMB 재난경보서비스(DMB Automatic Emergency Alert Service)

DMB(Digital Multimedia Broadcasting) 재난경보 서비스의 목적은 이동 중 TV, 라디오 및 데이터 방송을 복합적으로 수신이 가능한 방송이다. DMB 재난경보서비스는 방송 매체이기 때문에 제한된 전파 자원을 많은 사람들에게 효율적으로 전달할 수 있다. 마지막으로 방송과 통신의 융합으로 인해 DMB의 방송단말기와 통신기능의 융합으로 사용자별 양방향 주문형 데이터 방송이 가능하다.⁴²⁾

42) 최재웅. (2012). 뉴미디어를 활용한 재난방송 전달체계 개선 및 재난방송정책 연구. (석사학위,

DMB 재난정보서비스는 2010년 KBS를 시작으로 2018년에는 MBC, SBS, YTN과 함께 전국 서비스를 제공 중이다. 방송사업자 및 내비게이션 제조업자는 「재난안전법 제38조의 2」에 따라 재난 예·경보 실시사항이 DMB나 내비게이션 화면에 표시할 수 있도록 장치를 개발해야 하는 법적 근거를 따른다. 이때 신속한 재난 예보 및 경보 방송을 위해, 현재 시청 중인 방송을 끊김 없이 지상파 DMB의 데이터 채널을 이용하여 재난 관련 정보를 수신하는 것을 말한다.⁴³⁾

각종 지역 및 국가 재난이 발생했거나 발생 가능성이 있는 경우 통합발령 시스템에 출동 요청을 보낸다. 이때 DMB 통합 발령시스템을 통해 요청이 승인되면 승인된 발령정보는 지상파 3사 등의 방송국을 통해 지상파 DMB로 송출되는 시스템이다[그림 2-14].



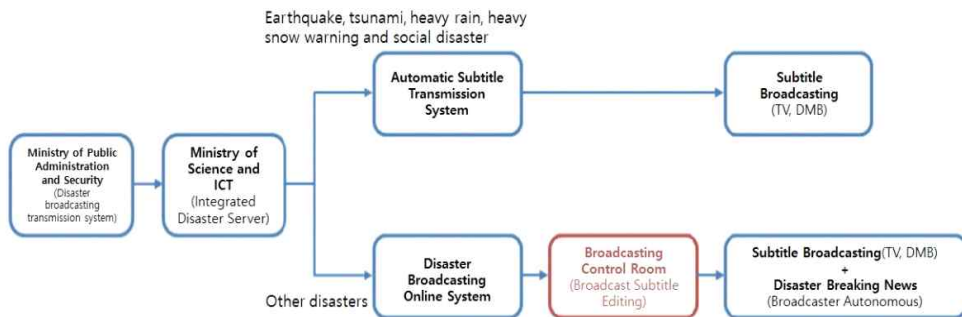
[그림 2-14] DMB 재난정보서비스 발령체계도

서울시립대학교 과학기술대학원).

43) Korea Information and Communication Technology Association (TTA). (2015). Terrestrial digital multimedia Fish Broadcasting (DMB): Disaster Alert Service. Information and Communication Organization Standard TTA.KO-07.046/R6, Korea.

다) 재난온라인방송시스템(DITS, Disaster Information Transform System)

재난온라인방송시스템은 재난 또는 재해가 발생하거나 발생 우려가 있을 때 「방송통신발전 기본법 제40조」에 따라 대피·구조·복구에 필요한 재난정보를 자막으로 TV에 송출하는 대국민 서비스이다. 재난의 종류에 따라 재난정보는 자막전송시스템 또는 온라인 재난방송시스템을 통해 전달된다. 재난온라인방송시스템은 지진·지진해일, 호우·대설경보 및 사회재난에 대해 「방송통신발전 기본법 제40조」 제2항에 따라 과학기술정보통신부 및 방송통신위원회에서 재난방송 요청 시, 방송사에서 자막작업을 거치지 않고 바로 TV 자막 송출이 가능하다. 긴급재난에 대비해 KBS, MBC, SBS, EBS, TV조선, YTN, MBN, JTBC, 채널A, 뉴스Y 등은 재난방송 요청 시 자막작업을 거치지 않고 바로 송출할 수 있는 자막 자동 송출시스템을 개발했다. 총 164개 방송국을 통해 TV 방송을 자막과 함께 시청하면서 재난 예보와 경보를 인지할 수 있는 시스템이다.⁴⁴⁾



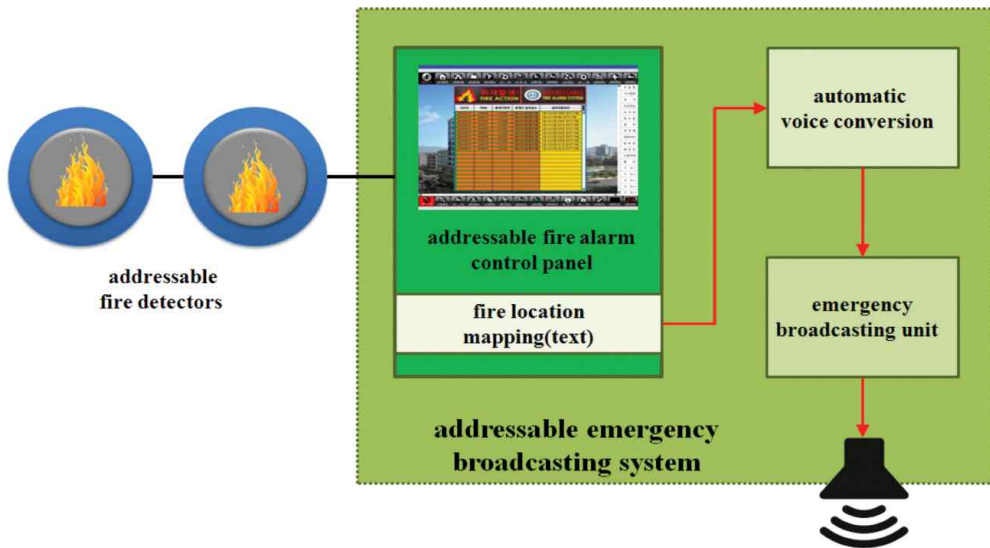
[그림 2-15] 재난방송온라인서비스 세부 구성도

44) 황우석 & 표경수, 전계논문.

4) 비상방송 시스템 연구 현황

가) 주소형 비상방송시스템에 관한 연구

피난 성능 향상을 위한 주소형 비상방송시스템의 필요성에 관한 연구⁴⁵⁾에서는 2015년 발생한 의정부 공동주택화재와 같이 화재 발생 시 초기 피난 실패로 인한 인명피해를 최소화하기 위해 피난 성능 향상을 위한 주소형 비상방송시스템의 필요성에 관한 연구이다. 특히 현대 건축물과 같이 복잡한 내부 구조물의 화재피해를 최소화하기 위해 정확한 위치를 거주자에게 알려 초기 피난 방향 오류를 방지하여 효율적인 대피가 가능한 적응형 비상방송 시스템에 관한 연구이다.

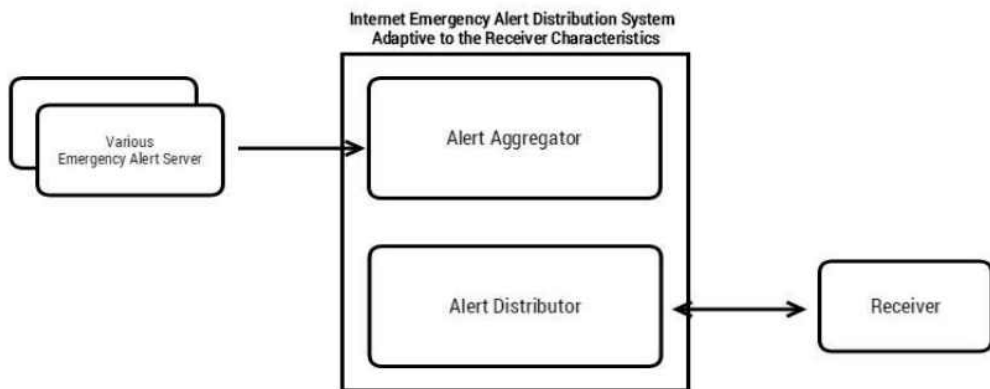


[그림 2-16] 주소기반 비상방송 시스템

45) 김시국, 이민용, 이춘하. (2015). 피난성능향상을 위한 주소형 비상방송시스템의 필요성에 관한 연구. 『한국화재소방학회 논문지』, 29(5), 14-22.

나) 인터넷 기반의 수신자 맞춤형 재난경보 전달시스템 연구⁴⁶⁾

효과적인 정보전달은 재난 피해를 크게 줄일 수 있는데 정보전달 채널을 다양화하는 것이 정보전달 효과를 향상시킬 수 있는 방법 중 하나이다. 해당 논문에서는 기존 정보시스템을 분석하여 문제점을 보완한 새로운 정보시스템을 제안한다. 기존에 운영 중인 CBS (Cell Broadcasting System) 재난문자방송 서비스와 지상파 DMB 자동재난경보 시스템(이하 T-DMB 재난경보방송)은 콘텐츠의 제한, 수신 기기의 제한, 단방향 서비스라는 단점이 있다. 이러한 문제점을 해결하기 위해 RSS (Rich Site Summary)를 활용한 인터넷 기반의 수신자 맞춤형 재난경보 전달시스템을 제안하였다.

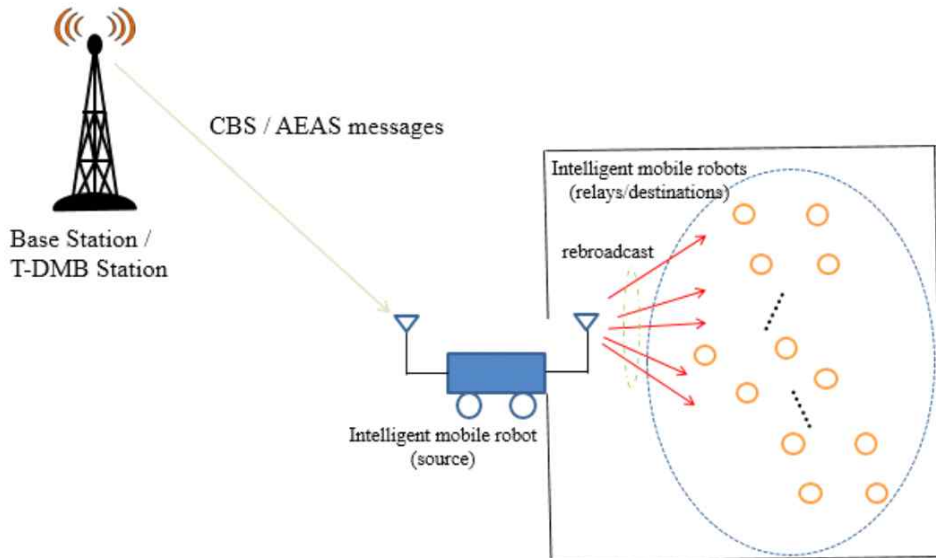


[그림 2-17] 인터넷 기반의 수신자 맞춤형 재난경보 전달시스템 시스템

46) 안소영, 전인찬, 김지희, 이용태, 최성중. (2015). 인터넷 기반의 수신자 맞춤형 재난경보 전달 시스템 연구. 『방송공학회논문지』, 20(4), 598-605.

다) 신뢰성 있는 재난경보 방송을 위한 지능형 이동 로봇의 협업⁴⁷⁾

이동통신 시스템과 T-DMB는 재난경보 방송을 위한 CBS와 AEAS (Automatic Emergency Alert Service) 기능이 정의하고 있다. 다만, 재난 발생 등으로 실내 공간의 통신 인프라가 불능일 경우 실내 거주자들은 재난경보를 수신하지 못할 수 있다. 해당 논문에서는 이러한 단점을 극복하기 위해 지능형 이동 로봇을 위한 협업 기술을 제안하였다. 재난경보 메시지를 성공적으로 수신한 지능형 이동 로봇들은 수신에 실패한 지능형 이동 로봇들에게 협력통신방식으로 재난 메시지를 재전송한다.

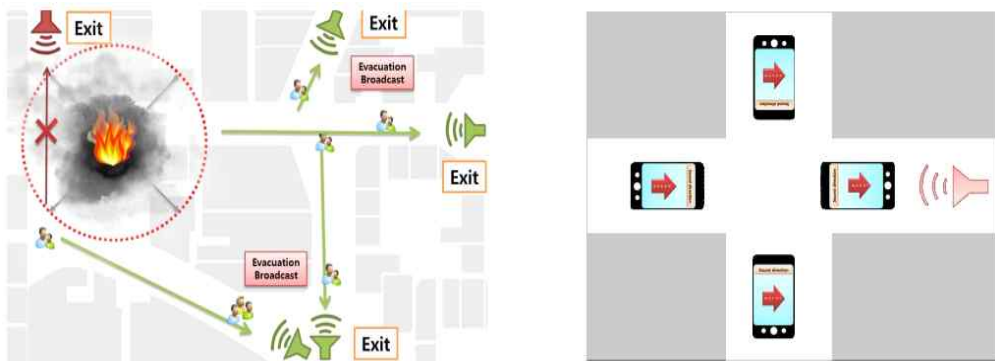


[그림 2-18] 재난경보 방송의 재전송을 위한 지능형 이동 로봇들의 협업

47) 장석진 & 이용태. (2019). 신뢰성 있는 재난경보 방송을 위한 지능형 이동 로봇의 협업 기법. 『방송공학회논문지』, 24(3), 396-401.

라) 스마트 디바이스를 활용한 가청, 비가청 신호 기반 피난방향 탐지⁴⁸⁾

해당 논문에서는 재난 발생 시 지향성 스피커와 스마트 기기를 활용한 피난 방향 탐지방법을 제안하였다. 연기, 소음 등으로 출구 방향조차 알 수 없는 재난 환경에서 지향성 스피커의 가청·비 가청 주파수 대역의 안내방송을 스마트 기기로 수신하여 피난 경로를 안내받을 수 있는 기법이다.



[그림 2-19] 제안기술의 개념도(좌)와 방향탐지 앱(우)

마) 비상방송설비의 음성명료도 및 재실자 인지성 개선방안 연구⁴⁹⁾

공동주택의 복잡한 평면 구조와 벽체 간 차음성능 향상으로 화재 상황 전달 및 인지가 어려워 피해가 증가하고 있다. 해당 연구에서는 국내 공동주택 평면 형상을 대상으로 한 소방시설 중 비상방송설비의 명료성에 대한 실태 분석을 수행하였다. 또한, 피난방송의 명료한 인지를 위한 음성 명료도 개선 방안을 시뮬레이션을 통해 분석하였다.

48) 현병천, 윤영욱, 박요한, 김영역. (2017). 스마트 디바이스를 활용한 가청, 비가청 신호 기반 피난방향 탐지 기법 연구. 『한국재난정보학회』, 13(1), 51-58.

49) 오소영, 조현민, 이영주, 이민주, 윤명오. (2018). 공동주택 내 비상방송설비의 음성명료도 실태 분석 및 재실자 인지성 개선방안 연구. 『한국화재소방학회』, 32(4), 60-68.

제 3 절 공공데이터 관련 선행연구

1) 공공데이터의 개요

가) 공공데이터의 정의 및 개요

공공데이터란 흔히 정부나 공공기관이 보유한 데이터로서 국민이 보다 쉽게 공유하고 활용할 수 있도록 공개된 데이터를 말한다. 이에 대하여 국가에서는 법률을 통해 제공 및 이용 활성화를 도모하고 있다. 2020년 12월 10일에 시행된 「공공데이터의 제공 및 이용 활성화에 관한 법률」(약칭: 공공데이터법) 제2조(정의)에 따르면 “공공기관”이란 국가기관, 지방자치단체 및 「지능 정보화 기본법」 제2조 제16호에 따른 공공기관을 말한다. “공공데이터”는 데이터베이스, 전자화된 파일 등 공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하는 광(光) 또는 전자적 방식으로 처리된 자료 또는 정보로서 다음 각 목의 어느 하나에 해당하는 것이라고 정의하고 있다.⁵⁰⁾ 이를 위해 공공기관이 준수해야 하는 관리 원칙과 기준이 존재하며, 이를 바탕으로 데이터를 제공하고 있다.

[표 2-3] 공공데이터의 정의

「공공데이터법」 제2조제2호	내용
가. 「전자정부법」 제2조제6호에 따른 행정정보	• 행정기관 등이 직무상 작성하거나 취득하여 관리하고있는 자료로서 전자적 방식으로 처리되어 부호, 문자, 음성, 음향, 영상 등으로 표현된 것
나. 「지능정보화 기본법」 3조제1호에 따른 정보 중 공공기관이 생산한 정보	• 특정 목적을 위하여 광(光) 또는 전자적 방식으로 처리되어 부호, 문자, 음성, 음향 및 영상 등으로 표현된 모든 종류의 자료 또는 지식 중 공공기관이 생산한 정보

50) 공공데이터의 제공 및 이용 활성화에 관한 법률.

<https://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률>

다. 「공공기록물 관리에 관한 법률 제20조제1항에 따른 전자기록물 중 대통령령으로 정하는 전자기록물	• 웹기록물 및 행정정보 데이터세트 등의 기록정보자료
라. 그 밖에 대통령령으로 정하는 자료 또는 정보	

정부는 2013년 7월에 공공데이터법을 제정하여 다양한 분야에서 공공데이터를 적극적으로 제공하고 있다. 특히, 2년마다 측정되는 OECD 공공데이터지수(Open-Useful-Reusable; OURdata Index)에서는 공공데이터 개방지수에서 3년 연속 1위를 차지하였다.⁵¹⁾ 이후 2016년 1월에 법이 일부 개정되었는데 개정이유는 '공공정보'의 개념을 개별 법률에서 공공데이터와 유사한 개념으로 규정하고 있어 혼동을 줄 수 있기 때문이다. 이는 공공기관이 민간과 유사하고 중복되는 서비스 개발을 금지함으로써 공공데이터 이용자의 권익을 보호하고 공공데이터를 활용한 기업가 정신을 도모하기 위함이다.⁵²⁾

공공데이터를 제공하는 가장 대표적인 기관으로 행정안전부의 '공공데이터포털(data.go.kr)'과 서울시의 '서울열린데이터광장(data.seoul.go.kr)'이 있다. 이곳들은 공공기관이 생성 또는 취득하여 관리하는 공공데이터를 제공하는 통합 창구이다. 특이점으로는 서울 자치구의 경우 공공데이터포털 보다 자체 데이터 포털인 열린데이터광장에 더 많은 데이터를 개방하고 있다.

나) 공공데이터포털

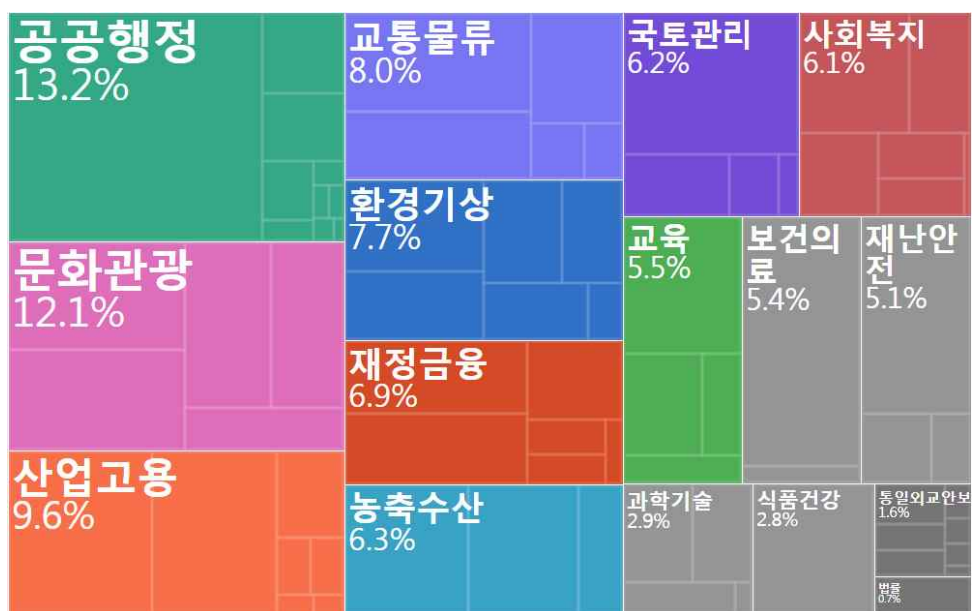
포털은 파일 데이터, 오픈 API, 시각화 등 다양한 방법을 제공하여 대중이 쉽고 편리하게 공개 데이터를 이용할 수 있고, 누구나 쉽고 편리한 검색을 통해 원하는 공개 데이터를 빠르고 정확하게 찾을 수 있다. 2022년 11월 현재 55,501건의 파일 데이터와 9,501건의 오픈 API 등 총 73,377건의 데이터가 등록되어 있다. 공공데이터 포털(data.go.kr)에서의 공공데이터 개방 방식은

51) 송채은 & 김학래. (2022). 지방자치단체의 공공데이터 개방 현황 분석과 개선방안. 『한국디지털컨텐츠학회』, 23(4), 697-705.

52) 김유리. (2016). 소프트웨어 교육용 공공개방데이터를 활용한 모바일 앱 구현에 관한 연구 (박사학위, 신라대학교 대학원).

파일 데이터 다운로드 방식과 오픈 API . CSV, XLS, XLSX, 를 통한 실시간 데이터 연계 방식을 이용하고 있다 파일 데이터는 HWP, PDF, XML, API REST XML, SOAP XML, XML Link, LOD, JSON, RDF 등의 다양한 형식으로 실시간 데이터 질의(request)/응답(response) 서비스를 제공하고 있다.⁵³⁾⁵⁴⁾

[그림 2-20]과 같이 공공행정을 비롯한 다양한 부문에 걸쳐 구성되어 있으며, 현재는 코로나19와 같은 보건의료 부문이나 미세먼지 같은 환경기상 정보, 그리고 저출산/고령화 같은 카테고리도 이슈화되고 있다.



[그림 2-20] 데이터 맵

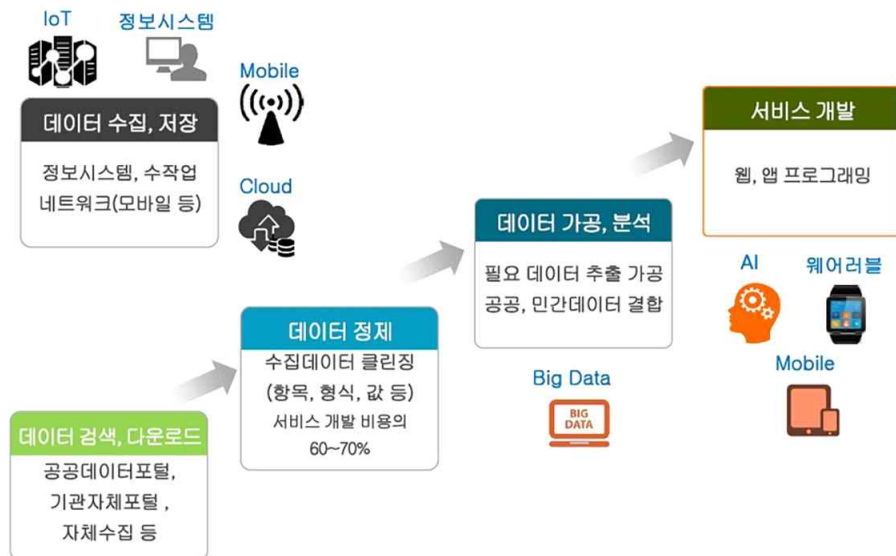
또한, 국가 중점데이터로 지정한 데이터에는 소방청의 119구급정보, 지자체의 IoT기술을 활용한 교통 및 환경정보, 한국수자원공사의 가뭄대응정보, 가뭄통합정보, 식품의약품안전처의 감염병 관련 정보 및 의약품 임상정보와 같은 재난과 관련된 데이터들이 상당수 중점데이터로 지정되어 있다. 데이터

53) 최성만. (2019). 정보시스템 측면에서 공공데이터 개방에 영향을 미치는 요인에 대한 연구: 제 공자와 이용자의 차이분석을 중심으로 (박사학위, 단국대학교 대학원).

54) 공공데이터포털. <https://www.data.go.kr>

의 수집에서 활용까지의 과정을 보면 [그림 2-21]처럼 공공에서 제공하는 자료를 이용하여 궁극적으로는 서비스를 개발하고 제공하는 것이다.

공공데이터 개방 이후 행정안전부는 데이터 분석을 중앙부처, 지자체, 공공기관 업무에 적용하고 구체적인 사례를 담은 ‘공공부문 데이터 분석·활용 우수사례집’을 발간했다. 과학적, 효율적 행정의 구체적 사례를 담은 사례로 행정안전부가 주관한 데이터 분석과제(2017~2020년)와 공공 빅데이터 분석 공모전, 빅데이터 우수사례 경진대회에 출품됐던 (2019~2020년) 데이터 분석 사례 중 실제 현장에서 활용한 우수사례를 중심으로 5대 분야(공공행정, 교육 복지, 재난안전, 보건의료, 산업고용)별로 나누어 수록되어 있다.⁵⁵⁾ 이 사례 중 재난안전과 관련된 내용들을 발췌하면 다음 [표 2-4]와 같다.



[그림 2-21] 데이터 수집에서 활용까지의 과정

55) 행정안전부. (2021). 공공부문 데이터 분석·활용 우수사례집 발간데이터 분석으로 공공업무를 개선한 18개의 우수사례 수록

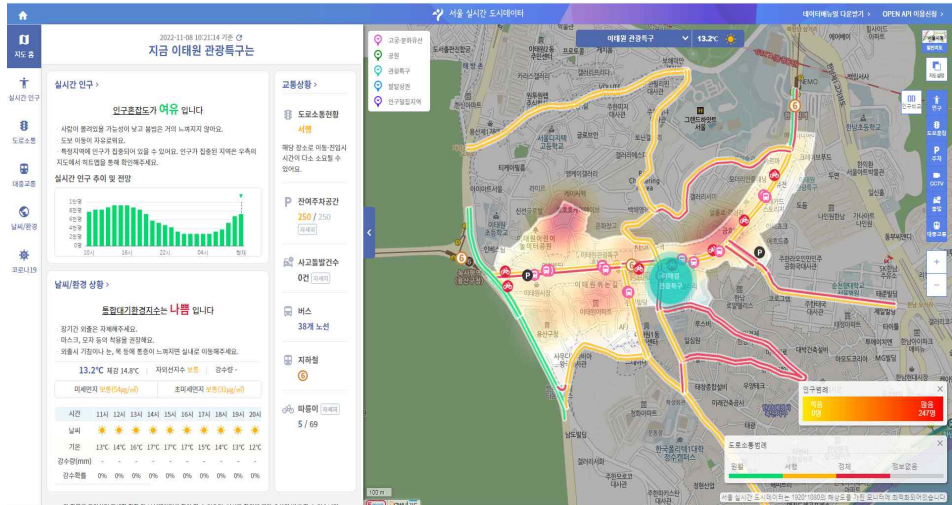
[표 2-4] 공공부문(재난안전) 데이터 분석 및 활용 우수사례

과제명	주요내용
빅데이터 기반 해양안전 위험지수 도출로 안전한 바다를!	• 해양안전사고 위험지역 예측을 위한 데이터 분석 - 과거 해양사고 데이터를 분석하여 지역별·시기별 해양안전 위험지수 도출 및 안전 시설물 우선 설치지역 선정에 활용
산불 예방과 피해 절감 솔루션 소방용수시설	• 골든타임 초기 대응을 위한 소방용수시설 취약지수 개발 - 화재 발생 현황, 소방용수시설, 건물정보 등을 결합·분석하여 위험 취약지수를 개발하고 소방용수시설 설치 우선지역 판별
영사민원시스템에 기반해 선제적으로 재외국민 보호	• 재외국민 사건사고 위험도 예측 분석 - 다양한 유형(국가, 연령, 비자유형, 해외송금 등)에 대한 빅데이터를 구축하고, 사건사고를 분석하여 상황별 재외국민의 위험도를 예측하고 선제적 안내에 활용
교통 빅데이터 로드맵으로 맞춤형 교통안전지원 모델 개발	• 사업용 자동차를 위한 교통안전관리 지원모델 개발 - 한국교통안전공단이 보유한 안전관리 데이터를 종합적으로 분석하여 운수회사별 안전등급 및 위험도 지표를 산출하고 안전 점검에 대한 관리체계 지원
생명 구하는 일등공신, 응급차량 출동을 책임지는 인공지능!	• 긴급차량 출동지연 요인 분석 - 소방차와 구급차의 과거 출동기록에 대한 데이터 분석으로 상습 지연구간을 파악하고 최단경로 알고리즘을 적용하여 긴급차량 출동시간 단축⁵⁶⁾

다) 서울 열린데이터광장

서울시 및 자치구, 그 산하기관에서 보유하고 있는 공공데이터를 공개한 것으로 연구, 관리, 서비스 제공 등 서울시 시정 활동 과정에서 수집된 다양한 데이터가 그 대상이다. 2022년 11월 기준 공공데이터 7,073건, 서비스 14,581건, AI학습데이터 15세트를 공개하고 있다. 공공데이터를 누구나 자유롭게 이용할 수 있고, 재사용이 가능하며, 데이터 사용에 문제가 발생하지 않도록 데이터에 대한 라이선스 정보도 함께 게시하고 있다.

56) 행정안전부, 상계서.



[그림 2-22] 서울 실시간 도시데이터 예시화면

특이점으로 서울시는 2022년 9월부터는 인공지능(AI) 전문기업 화이트스캔, KT와 함께 공공(교통, 환경)과 민간(통신)의 실시간 데이터를 융합한 ‘서울 실시간 도시데이터’[그림 2-22]를 개발하고 공개하였다. 서울 명소 50곳에 대한 분야별(실시간 인구, 도로소통, 대중교통, 날씨/환경, 코로나19) 실시간 정보들이 융합된 데이터로 실시간 인구예측 AI모델에서 제공하는 인구데이터를 기반으로 장소의 인구 추이와 예측을 통해 생활의 질 향상과 감염병 선제 방역 기타 위기대응체계 구축 등에 도움을 줄 것으로 전망되고 있다.⁵⁷⁾

그 외에도 KT와는 공공빅데이터와 통신데이터를 이용하여 추계한 서울의 특정 지역, 특정 시점에 존재하는 모든 인구데이터를 제공하는 ‘서울 생활인구’라는 서비스와 서울 내에서 이동하거나 서울 바깥에서 서울로 이동하는 통근, 통학, 여가, 쇼핑 등의 이동정보를 제공하는 ‘서울 생활이동 인구’[그림 2-23]정보를 제공하고, SKT와는 통신데이터 가명결합을 통해 추정한 서울 행정동단위 성, 연령별 1인가구와 서울시민의 생활특성 정보 등을 제공하는 서울 시민생활 데이터를 제공하고 있다.

57) 이상원. (2022년9월1일). 화이트스캔, 서울시·KT ‘실시간 도시데이터 서비스’ 개발. 전자신문. <https://www.etnews.com/20220901000247>



[그림 2-23] 서울 생활이동 예시화면

라) 주요국 공공데이터 정책

[표 2-5] 국외 주요국 공공데이터 개방 및 활용 정책

국가	주요내용
미국	정보자유법('96)을 근간으로 공공데이터의 이용 및 재배포를 규정함. 국정운영 기조에 따라 정부기관 전체의 개방 가능한 정보를 제공하는 연방 데이터 공개 사이트(data.gov)구현 및 운영('09.5)
EU	‘공공부문 정보의 재활용에 관한 지침’제정('03.11)을 통해 상업적 이용을 위한 법적 근거 마련.
영국	‘공공정보 재이용규칙’제정('05.6) 및 공공정보 공개사이트(data.gov.uk) 구현('10.1)으로 공공데이터를 통한 부가가치 창출로 10파운드의 시장 규모 창출 전망
호주	‘열린 정부 산업 촉구 보고서('09)’를 통해 공공정보 활용의 사회경제적 중요성을 인식하고 공공기관 연계서비스(data.australia.gov.au) 구현('10)

(자료 : 행정자치부, 한국정보화진흥원, 2013)⁵⁸⁾

공공데이터의 개방과 활용은 세계적인 흐름이며, 각국은 새로운 패러다임을 수용하게 되었다. 주요 선진국들은 이미 공공데이터와 공공정보가 사회적

58) 김유리, 전제논문.

인 가치나 경제적인 가치에서 높은 위치에 있음을 인식하고 공공정보 개방과 민간이용을 위한 다양한 정책을 추진하고 있다. [표 2-7]는 주요 해외 공공정보 공개 및 활용 정책을 정리한 것이다.

2) 민간 활용 사례 연구

공공데이터의 개방으로 이용 가능한 서비스의 유형은 점점 더 다양해지고 있고 실제 공공데이터를 이용한 서비스도 점점 확산되고 있다. 그 중 재난안전과 관련된 데이터를 민간에서 이용한 실제 서비스 사례를 찾아보았다. 주로 스마트폰의 앱 또는 웹 서비스 형태로 정보를 제공하고 있었다.

가) 대기오염 관련 앱 서비스 사례

현재 서울시 대기질측정소에서 측정된 데이터를 기반으로 지역별 미세먼지 측정농도를 서비스하고 있으며, 다양한 입력 인자를 기반으로 모델링한 결과 등을 기초로 미세먼지 현황 및 예보 서비스를 제공하고 있다. 미세먼지 농도는 행정동별 농도별로 스마트폰 앱이나 네이버 또는 아래 사이트를 통해 현지에서 제공하고 있다. 미세먼지 환경기준을 초과한 지역과 보통 수준을 보이는 지역은 서울의 다양한 지역에서 다르게 나타날 것이며, 이 시공간 데이터는 앱을 통해 위치정보와 연계된 서비스로 제공되고 있다. 미세먼지에 대한 환경기준 및 측정지점에 따른 미세먼지의 공간적 분포를 초과하여 환경정보 서비스를 받고자 하는 수요자의 위치와 가장 가까운 측정소의 미세먼지 농도를 제공한다. 이에 대한 기상청, 에어코리아, 서울시의 미세먼지 측정값을 비교하여 대기질 정보를 파악하는데 환경정보 수요자는 다양한 앱을 활용하고 있다.⁵⁹⁾

59) 김원석. (2019). *공공데이터를 활용한 생활환경 정보서비스 제공 앱 개발* (석사학위, 한남대학교 대학원).



[그림 2-24] 에어가드K 실행화면

에어가드K(AirGuard K)는 기상정보 제공 업체 케이웨더(K-Weather)에서 제공하는 앱이다. 기상정보와 대기오염 예보, 미세먼지 정보를 동시에 확인하기에 편리하다. 이는 한국환경공단에서 제공하는 실시간 대기환경 정보와 케이웨더에서 제공하는 대기오염 예보를 함께 보여준다. 또한 케이웨더의 자체 예보를 기반으로 대기오염 예보, 생활 습관 예보, 대기 정보를 볼 수 있다. 케이웨더는 서울 12개, 경기도 10개 등 전국 85개 측정기를 통해 미세먼지 정보를 제공하고 있다.⁶⁰⁾

나) 통합재난정보전달솔루션

(주)에스에이텍이 제공하는 통합재난정보전달솔루션은 관내 지역정보를 기반으로 지역 맞춤형 재난정보를 제공하여, 주민의 안전 확보와 적절한 대응을 유도한다. 또한 유관기관 재난정보시스템 연계 및 관내 기상측정 장비 연계를

60) 오원석. (2017). [IT후마니타스] 미세먼지 걱정되면 스마트폰에 물어봐요. *중앙일보*.
<https://www.joongang.co.kr/article/21470477#home>

통해 신속한 상황판단이 가능하며, TTS, 실시간 음성방송, 녹음방송 등의 다양한 동보방송 기능을 통해 관내 각종 행정, 생활정보 등을 그룹화하여 맞춤형 서비스를 제공한다.

이 시스템은 기상청에서 제공하는 기상특보, 일기예보, 지진정보 등의 정보를 Open API를 통해 가져오고, 한국환경공단으로부터 미세먼지, 초미세먼지, 오존 등의 대기질 정보를 가져와서 기상특보 대상지역, 미세먼지 기준치 초과 등일 때 자동으로 안내방송을 송출한다.⁶¹⁾

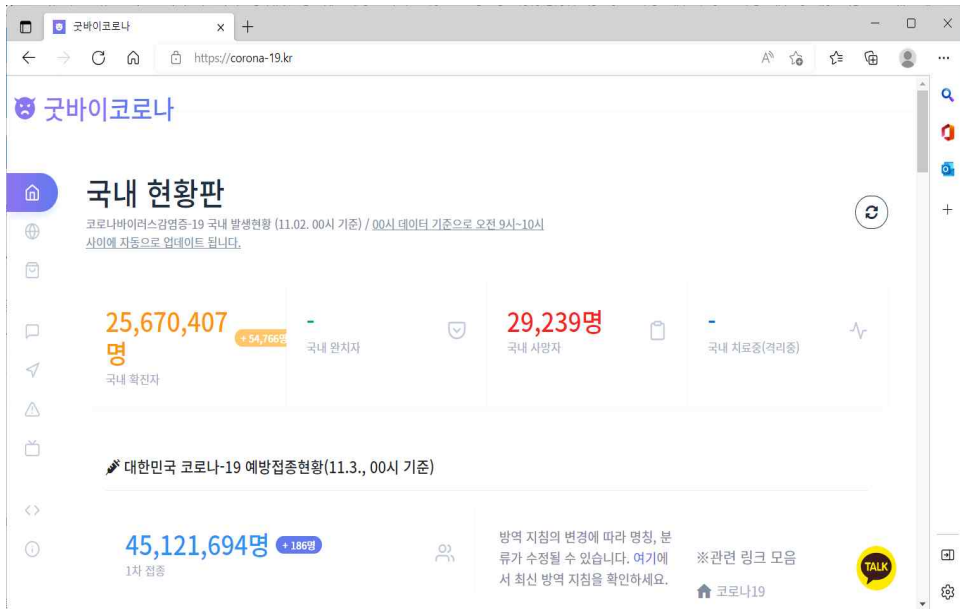


[그림 2-25] 통합재난정보전달솔루션 구성도

61) 에스에이텍 통합재난정보전달솔루션. <http://www.safealarm.co.kr/satech/solution01.asp>

다) 굿바이코로나

굿바이코로나는 코로나바이러스 감염증의 국내 발생 현황을 매일 업데이트하여 보여준다. 행정안전부와 건강보험심사평가원의 Open API를 활용하여 매일 아침 9시경에 업데이트 해준다.⁶²⁾

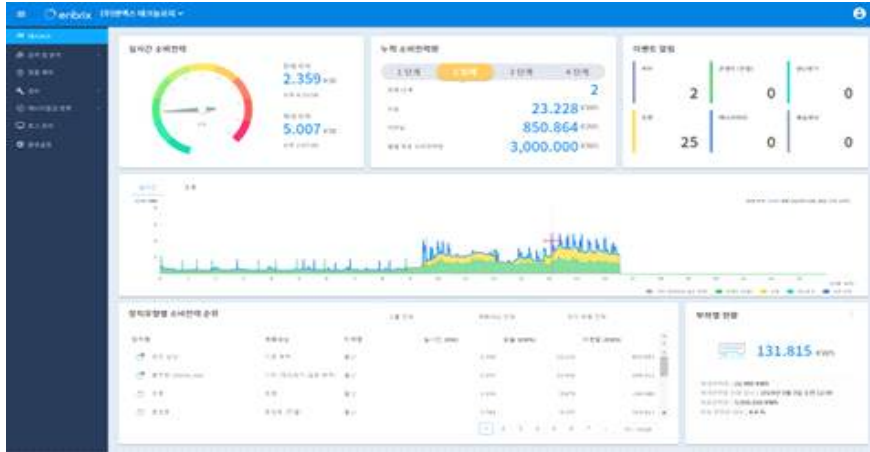


[그림 2-26] 굿바이코로나 실행화면

라) 스마트 에너지 시스템 솔루션

(주)엔엑스테크놀로지의 엔브릭스(enbrix)는 다양한 공공데이터를 활용하여 건물의 전력 사용량을 실시간으로 관리 및 통제하여 에너지 소비를 절감하는 스마트 에너지 시스템 플랫폼으로, 다양한 건물의 환경에 맞게 구축 가능하며 또한 상황 맞춤형 운영이 가능하다. 한국전력공사, 행정안전부, 기상청, 한국환경공단 등에서 제공하는 Open API를 활용하여 건물 내에서 필요한 에너지 절감을 위한 정보를 제공한다.⁶³⁾

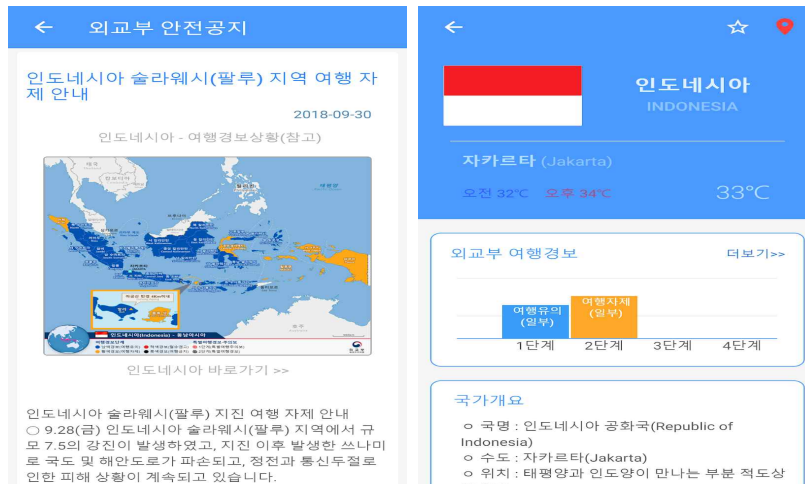
62) 굿바이코로나. <https://corona-19.kr/>



[그림 2-27] enbrix 실행화면

마) 세계안전정보

세계안전정보는 세계 각국에 대한 안전과 관련된 외교부의 공지사항이나 돌발 상황 대비용 매뉴얼 등을 제공하여 안전한 여행을 위해 다양한 국가 안전 정보를 제공하는 앱으로 Open API를 통하여 외교부의 공공데이터(국가별 공지사항 목록조회, 국가별 안정정보)를 활용하여 실시간 공지를 제공한다.



[그림 2-28] 세계안전정보

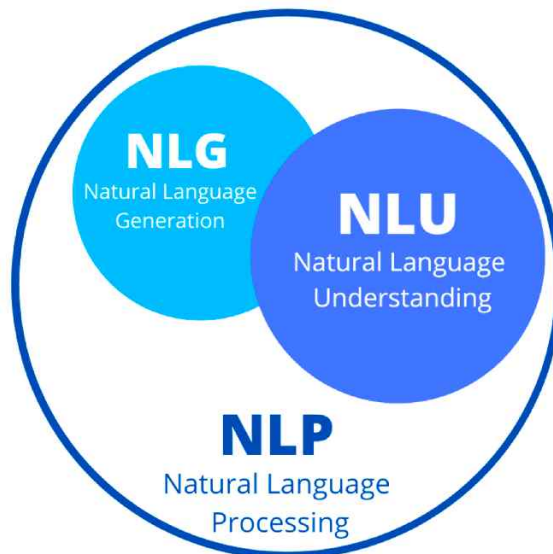
63) 엔브릭스(enbrix). <http://enbrix.net/kr/main/main.php>

제 4 절 자연어 처리 관련 선행연구

1) 자연어 처리의 개요

가) 자연어 처리의 정의

자연어(natural language)란 사람들이 일상에서 사용하는 언어를 말하며 사람들 사이에서 이루어지는 대화는 일반적으로 자연어가 사용된다.⁶⁴⁾ 하지만 사람과 컴퓨터 사이의 커뮤니케이션에서는 컴퓨터에서 요구하는 형식에 맞추어 문장을 작성해 주어야 한다. 컴퓨터에게 원하는 일을 시키기 위해서 프로그래밍을 할 경우에도 프로그래밍 언어로 명령을 작성해야 하며 작성 시 언어의 문법 등도 정확하게 지켜주어야만 원하는 결과를 얻을 수 있다. 자연어 처리(NLP: natural language processing)는 우리가 일상생활에서 사용하는 언어의 의미를 분석하여 컴퓨터가 처리할 수 있도록 하는 과정이다.



[그림 2-29] NLP의 하위 분야

64) 임희석 외. (2020). 『자연어처리바이블-핵심이론·응용시스템·딥러닝』. 서울: 휴먼싸이언스.

그 처리의 종류에는 [그림 2-29]과 같이 자연어 이해(NLU: natural language understanding)와 자연어 생성(NLG: natural language generation)이 있다. NLU와 NLG는 입력과 출력이 서로 반대이기 때문에 처리 방법이 다르고 연구의 방향도 서로 다르다.

자연어 처리가 최근 많은 주목을 받아 다양한 분야에서 활용되고 있지만 그 인기에 비해서 처리 난이도는 상당히 높다. 자연어 처리가 어려운 이유는 딥러닝에 대한 이해는 물론 인간 언어에 대한 이해도 필요하기 때문에 접근하기 어려운 분야이다. 또한, 언어의 종류와 형태가 다양하기 때문에 처리가 매우 어렵다. 예를 들어 영어는 명확한 띄어쓰기가 있지만, 중국어나 일본어에는 띄어쓰기가 없기 때문에 단어 단위의 임베딩이 어렵다. 또한, 자연어 처리에 사용되는 용어들에 대한 이해도 필요하다.⁶⁵⁾

나) 자연어 처리 연구의 패러다임

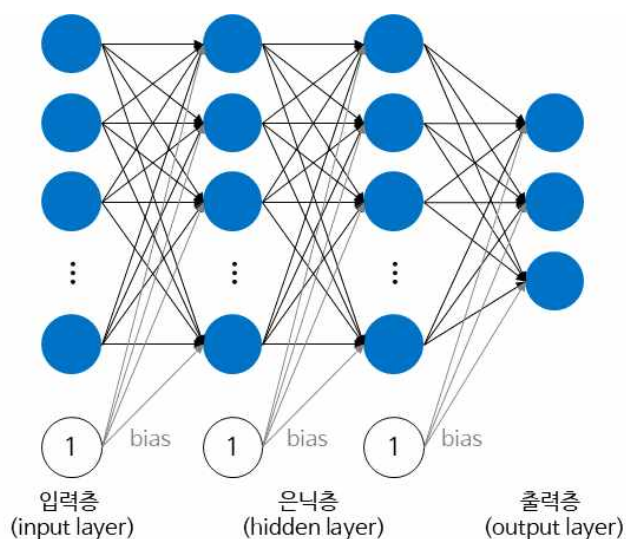
'규칙 기반' 언어처리는 사전에 등록된 문법 규칙을 사용하여 자연어를 처리하는 방법으로 프로그래밍 방식과 유사하다. 그 기원은 1954년 조지타운 대학교와 IBM이 공동 개발한 러시아어-영어 번역기이다. 이 모델의 가장 큰 문제는 미리 언어의 규칙을 스스로 구축해야 한다는 것이다. 언어를 구성하는 수많은 규칙이 있고 항상 예외가 있기 때문이다. 특히 한국어와 같이 복잡한 언어의 경우 룰 설정이 너무 많아 정확도가 떨어진다.

'통계 기반' 언어처리는 규칙 기반 자연어 처리의 한계를 극복하기 위해 제안된 방법으로 수학적 개념인 '조건부 확률'에 기반한 자연어 처리 모델이다. 이 방법은 언어에 일정한 규칙이 있으면 규칙과 관련된 단어(형태소) 간에 통계적으로 유의한 상관관계가 나타난다는 생각에서 시작되었다. 컴퓨터 성능이 향상되면서 많은 문장을 빠르게 분석할 수 있게 되었고, 이 처리 방법에서 결과가 나타나기 시작했다. 인간의 개입 정도는 줄어들었지만 여전히 번거로움이 있어서 이러한 문제를 해결하기 위해 딥러닝이 채택되었다.

'딥러닝 기반'을 설명하기 전에, 일반적으로 '알고리즘'이라고 하는 문제는

65) 서지영. (2021). 『딥러닝 텐서플로 교과서』. 서울: (주)도서출판 길벗.

어떤 상황에 어떻게 대응하고 어떤 값을 계산해야 하는지에 대해 사전에 정의하여 처리하는 방식이다. 그러나 언어의 경우 입출력에 대한 경우의 수가 너무 많아 언어처리에는 사용이 거의 불가능하다. 머신러닝에서는 프로그램을 작성한 후 바로 사용하는 대신 입력 데이터 예제를 알고리즘에 지속적으로 입력하고 내부 처리에 사용되는 가중치를 지속적으로 업데이트한다. 이 과정을 학습이라고 하며, 학습 후에 실제로 사용할 수 있는 알고리즘을 '모델'이라고 한다. '인공신경망'은 머신러닝의 한 형태로, 인간 두뇌의 신경계와 유사한 입력 데이터를 처리하고, 연산 가중치를 결정하고, 다음 단계로 연결하는 알고리즘의 형태를 따서 명명됐다.



[그림 2-30] 심층 신경망(DNN) 개념도

딥러닝이란, 머신러닝 중 신경망 구조에서 뉴런의 계층(layer)의 수를 늘려 성능을 높이는 것을 의미한다. 딥러닝의 가장 큰 장점 중 하나는 단순한 통계적 분석을 넘어선 결과를 제공할 수 있다는 점이며, 딥러닝 기반의 자연어 처리가 시작된 이후 문장이나 단락의 전체적인 맥락을 파악하는 자연어 처리 모델의 생성이 가능해졌다. 모델을 설계할 때 문장 전체에 고르게 연결된 인공 신경망 요소를 생성하면 이 변수에는 학습 후 문장 전체에 대한 정보가 일부 포함되어 있다고 할 수 있다. 이처럼 딥러닝 모델은 목적에 따라 다양하

게 설계되고 있으며, 자연어 처리의 다양한 하위 분야에서 높은 정확도를 보여주고 있다.

다) 실제 응용사례⁶⁶⁾

[표 2-6] 범주별 NLP 응용

검색	웹	문서	자동 완성
편집	철자	문법	문체
대화	챗봇	가상 비서	일정 관리
저술	색인	용어집	차례(목차)
이메일	스팸 필터	분류	우선순위
텍스트 마이닝	요약	지식 추출	의료 진단
법	법리 추론	판례 검색	소환장 분류
뉴스	사건 검출	사실관계 점검	표제 작성
출처 파악	표절 검출	법의학 문서 분석	문체 조언
정서 분석	분위기 감시	상품평 선별	고객 지원
행동 예측	금융·재무	선거 예측	마케팅
문예 창작	영화 대본	시	가사

검색엔진과 스마트폰의 키 입력 시 볼 수 있는 자동 완성 기능도 NLP를 이용해서 사용자의 의도에 가까운 단어들을 제안해준다. 또한, 현재 문서를 작성하고 있는 문서편집기에도 철자 교정 기능과 문법 점검 기능이 제공되고 있으며, AP(미국 연합통신)은 NLP ‘로봇기자’를 이용해서 금융 기사와 스포츠 경기 결과 기사를 작성한다.⁶⁷⁾ 이메일 프로그램들에는 NLP 스팸 필터가 장착되어 있어 사용자들에게 편의를 제공하고 있다. 2016년 미국 대선에 관한 트윗의 약 20%는 챗봇에 의해 작성된 것으로 추정된다.⁶⁸⁾⁶⁹⁾

66) 홉슨 레인, 하네스 막스 하프케, 콜 하워드. (2020). *자연어처리 인 액션*. (류광, 역). 파주: 제이펍. (원본 출판 2019).

67) The Verge, (2015). AP's 'robot journalists' are writing their own stories now, <http://www.theverge.com/2015/1/29/7939067/ap-journalism-automation-robots-financial-reporting>

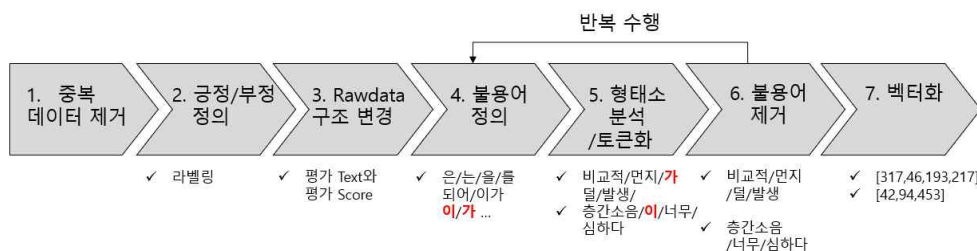
68) New York Times, automated Pro-Trump Bots Overwhelmed Pro-Clinton Messages,

이전의 경우보다 본격적으로 NLP 시스템을 사용하는 사례를 들면 인터넷 포털 등에 올릴 장문의 영화평이나 상품평을 NLP로 작성하는 것이 가능하다. 실제 극장에 가본 적도, 해당 제품을 사용해 본 적이 없더라도 자동화된 NLP 파이프라인이 생성된 사용 후기가 많이 있다. 특히, 고객 서비스와 관련된 서비스를 제공하는 데 사용하면 값비싼 인건비를 절감할 수 있고 상담원들이 겪는 스트레스도 줄이는 효과가 있다.

라) 자연어 처리의 절차

이처럼 머신러닝과 딥러닝의 발전으로 자연어 처리의 성능과 정확도가 크게 향상되었다. 요즘은 자연어 처리와 관련하여 딥러닝을 적용하지 않은 연구를 찾기 어려울 정도로 많이 발전했다. 딥러닝을 이용한 자연어 처리는 모델에 따라 세부적으로 다르지만 대부분 다음과 같은 단계를 거친다.

첫째, 인간의 언어인 자연어가 입력 텍스트에 들어온다. 이때 인간의 언어가 다양한 것처럼 처리 방식도 조금씩 다르다. 둘째, 입력 텍스트의 전처리가 필요하다. 셋째, 전처리 된 단어를 삽입한다. 즉, 단어를 벡터로 변환한다. 마지막으로 컴퓨터가 이해할 수 있는 모델을 사용하여 데이터에 대한 분류 및 예측을 수행한다. 이때 데이터의 흐름에 따라 분류와 예측의 결과가 달라진다.



[그림 2-31] 자연어 처리 과정

Researchers Say. (2016).

<https://www.nytimes.com/2016/11/18/technology/automated-pro-trump-bots-overwhelmed-pro-clinton-messages-researchers-say.html>

69) MIT Technology Review, (2016). How the Bot-y Politic Influenced This Election, <https://www.technologyreview.com/s/602817/how-the-bot-y-politic-influenced-this-election>

마) 전처리

머신러닝이나 딥러닝에서는 텍스트 자체를 기능으로 사용할 수 없다. 텍스트 데이터의 전처리가 필요하다. 이 경우 전처리를 위해 토큰화, 불용어 제거, 어간 추출 등의 작업이 필요하다.

(1) 결측치(Missing Value) 확인

데이터셋에서 결측치가 몇 개인지 확인하는 것은 매우 중요한 일이다. 결측치는 주어진 데이터셋에서 데이터가 없는(NaN) 것이다. 처리 방법으로는 임의의 값(예:0)으로 채워 모든 결측치를 채우는 방법이 있고, 결측치 전후의 값으로 결측치를 채우는 방법이 있다. 누락된 값의 양쪽에 있는 값의 중앙값을 찾아 누락된 값을 채우는 보간법(補間法, interpolation)을 사용하여 채우는 방법도 있고, 그렇지 않으면 결측치가 있는 행을 삭제하는 방법도 사용된다. 하나의 셀이라도 NaN이 있으면 해당 행을 삭제하거나, 한 행의 모든 셀에 NaN이 있으면 전체 행을 삭제하는 등 다양한 방법으로 처리가 가능하다.

(2) 토큰화(Tokenization)

토큰화(tokenization)는 주어진 텍스트를 컴퓨터가 인식하기 쉽게 단어/문자 단위로 절단하는 것을 의미한다. 토큰의 단위는 보통 의미를 가지는 최소 의미 단위로 선택된다. 단위가 단어(word)이면 단어 토큰화(Word Tokenization)라고 하고, 문장인 경우 문장 토큰화(Sentence Tokenization)라고 한다. 예를 들어 'Time is an illusion. Lunchtime double so!'라는 문장이 있을 때 토큰화를 진행하면 각각의 단어인 'Time', 'is', 'an', 'illusion', 'Lunchtime', 'double', 'so'로 분리된다. 토큰화는 NLTK 라이브러리를 주로 사용하고, 한글 토큰화는 KoNLPy 라이브러리를 사용한다. 영어의 경우에는 띄어쓰기(whitespace)를 기준으로 토큰화를 진행하지만 한국어의 경우에는 토큰화에 어려움이 있다.

한국어 토큰화가 어려운 이유는 띄어쓰기 단위가 ‘단어’가 아닌 ‘어절’이기 때문이다. 그 근본적인 이유인 언어의 특성이 교착어이기 때문인데 한국어 토큰화에서는 뜻을 가진 가장 작은 말의 단위인 형태소(morpheme)란 개념이 있어서 이에 대한 분리 작업도 필요하기 때문이다. 예를 들면 ‘나는 학교에 간다’라는 문장이 있다면, 띄어쓰기 단위의 토큰화를 진행하면 [‘나는’, ‘학교에’, ‘간다’]라는 결과가 나올 것이다. 이에 대해 형태소 분석을 하게 되면 자립형태소인 ‘나’, ‘학교’와 의존형태소인 ‘-는’, ‘-에’, ‘가-’, ‘-다’로 분리할 수 있다. 이처럼 한국어는 어절 토큰화가 아니라 형태소 토큰화를 수행해야 한다.

(3) 어간 추출(Stemming) 및 표제어 추출(Lemmatization)

어간 추출과 표제어 추출은 서로 다른 단어들이지만, 단어의 원형이 같다면 일반화하여 문서 내의 단어의 종류를 줄이는 작업이다. 표제어(lemma, 標題語)는 한글로 ‘기본 사전형 단어’의 의미를 가지고 있다. 예를 들면 am, are, is는 서로 다른 단어이지만 원형은 be인 것을 알 수 있다. 이때 표제어는 be라고 말할 수 있다. 어간(stem, 語幹)추출은 형태학적 분석을 단순화한 것으로 정해진 규칙으로만 단어의 어미를 자르는 추측작업으로 볼 수 있다.

(4) 불용어 제거(Stop-word Removal)

불용어(stop-word)는 데이터에 자주 등장하지만 분석에서 의미를 부여하기 어려운 단어들을 의미한다. 예를 들어 ‘a’, ‘the’ 등의 관사나 조사, 접미사 같은 단어들은 모든 구문(phrase)에 많이 등장하지만 없어도 의미 파악에 문제가 되지 않는 단어들이다. 불용어는 자연어 처리에서 효율성을 저하시키고 처리 시간을 지연시키는 큰 단점이 있기 때문에 반드시 제거할 필요가 있다. 이전 단계에서 만든 토큰화를 통해 한국어의 불용어를 제거하는 방법이 있지만, 사용자가 직접 불필요하다고 생각되는 명사와 형용사에 대해서 불용어 사전에 등록하고 삭제할 수도 있다.

2) 문서 요약의 정의 및 동향

문서 요약이란 원본 문서의 의미는 그대로 유지하면서 주요한 정보가 담긴 원본 문서보다 짧은 문서나 문장으로 출력하는 것을 말한다.⁷⁰⁾ 문서 요약은 I(interpretation), T(transformation), G(generation) 세 가지 기본 과정으로 정의할 수 있으며, I는 주어진 문서(source text)를 해석(interpretation)하여 컴퓨터가 이해할 수 있도록 표현(representation)하는 것이다. 이 과정에서는 문장 표현(sentence representation)을 통해 더 큰 문서 단위로의 표현(document representation)도 포함된다. T는 문서 표현(source representation)을 요약문으로 표현(summary text representation)될 수 있도록 변형/가공(transformation)하는 것이다. G는 요약문에 대한 표현(summary representation)으로부터 최종 요약문(summary text)로 생성하는 것이다.⁷¹⁾

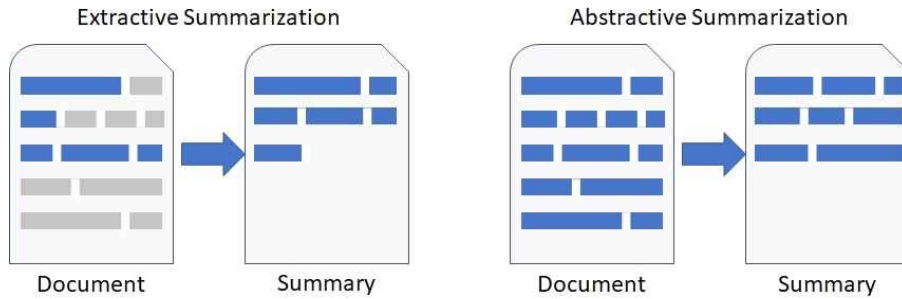
문서 요약은 크게 추출 요약(Extractive Summarization)과 생성 요약(Abstractive Summarization)의 두 가지 방법으로 나뉜다. 추출 요약⁷²⁾은 원본 문서에 있는 문장들 중에서 문서의 주제를 담고 있는 주요한 문장 몇 개를 선택해서 출력해 요약문으로 사용하는 방법이고, 생성 요약⁷³⁾은 문서 전체의 의미를 이해하고 이에 맞는 문장을 생성하여 요약문으로 사용하는 방법이다.

70) Jones, K. Sparck. (1999). Automatic summarizing: factors and directions. Advances in automatic text summarization. 1-12.

71) 임희석 외, 전게서.

72) 윤재민, 정유진, 이종혁. (2004). 육하원칙 활성화도를 이용한 신문기사 자동추출요약. 『정보과학회논문지』, 31(4), 505-515.

73) Chopra, S., Auli, M., Rush, A. M. (2016). Abstractive sentence summarization with attentive recurrent neural networks. Proceedings of NAACL-HLT16. 93-98.



[그림 2-32] 추출 요약과 생성 요약 비교

추출 요약의 경우 원본 문서의 모든 문장 중에서 중요하지 않은 문장을 제외하고 문서의 주제가 담긴 주요한 문장들만 선택하여 추출하므로 생성 요약을 할 때 기존 문서 전체가 아닌 추출 요약을 통해 추출된 문장들을 입력으로 사용하면 더 좋은 요약문을 생성할 수 있다.⁷⁴⁾ 하지만, 추출 요약에서는 가독성과 특정 단어의 반복문제 처리에 있어서 단점이 있다.⁷⁵⁾ 가독성 부분은 추출 요약이 가지고 있는 최대 단점 중 하나이다. 문서 내의 단어들을 가지고 추출을 통해 요약하기 때문에 요약문을 구성할 때 사람이 만든 것처럼 자연스럽지 못하여 글에 대한 가독성이 떨어지는 단점이 있다.

생성 요약은 추출 요약보다 기술적 난이도가 높지만 최근 심층 신경망(Deep Neural Network)을 통한 학습(Deep Learning)에 대한 연구가 활발해지면서 그동안 시도하기 어려웠던 품질이 높은 문장을 생성하는 자연어 생성(NLG: natural language generation) 부분에서 두각을 나타내고 있다. 이러한 흐름에 따라서 딥러닝 기반의 생성 요약이 더 주목을 받으면서 딥러닝 기반 추출 요약보다 상대적으로 더 많은 연구가 시도되고 있다.

가) 추출 요약(Extractive Summarization)

추출 요약은 원본 문서를 구성하는 문장 중 원본 문서의 주제를 포함하는

74) 전재원, 황현선, 이창기. (2019). 추출 요약과 생성 요약을 결합한 2단계 문서 요약. 『한국정보과학회 학술발표논문집』, 2019(6), 581-583.

75) 장동현 & 맹성현. (1997). 자동요약시스템. 『정보과학회』, 15(10), 42-49.

여러 문장을 선택하여 추출하고, 추출된 문장을 요약 문장으로 사용하는 방법을 말한다. 원본 문서의 주제를 포함하는 문장을 추출하기 위해 다양한 방법이 연구되었지만 핵심 원칙은 텍스트의 문장을 중요도에 따라 순위를 매기는 것이다. 문장의 중요성은 그것이 문서의 전체 의미에 얼마나 기여하는지에 따라 할당될 수 있다. 순위를 매긴 후 가장 높은 순위의 문장을 선택하고 요약을 선택할 수 있다. [표 2-7]은 문장의 중요성을 평가하는 기준이다.

[표 2-7] 추출 요약 방법별 문장의 중요성 평가 기준

평가 기준	내용
단어 출현 빈도 기반	명사 등 문장에서 중요한 역할을 하는 단어가 몇 개나 나오는지에 따라 각 문장의 가중치를 계산하여 추출하는 방법이다. ⁷⁶⁾
문장 간 유사도를 이용하는 방법	주어진 문장이 문서의 다른 문장과 얼마나 유사한지를 측정한다. 문장 유사도는 단어, 위치, 문장 순서 등의 정보를 기반으로 코사인 유사도와 같은 유사도 함수를 사용하여 문장 간의 유사도를 계산한다. ⁷⁷⁾
그래프를 이용하는 방법	각 문장은 하나의 노드로 표현되며, 문장에서 공통 단어가 등장하는 등 연결 관계가 있는 문장을 연결 그래프로 나타낸다. ⁷⁸⁾
딥러닝을 이용하는 방법	순환 신경망(RNN)이나 합성곱 신경망(CNN)과 같은 딥러닝 모델을 통해 입력 문장에 대한 문장 임베딩을 계산하고 각 문장에 대한 가중치를 계산하여 추출하는 방법이다. ⁷⁹⁾

나) 생성 요약(Abstractive Summarization)

76) McKeown, K., Robin, J., and Kukich, K. (1995). Generating Concise Natural Language Summaries. *Information Processing & Management*, 31(5), 703-733.

77) Mallett, D., Elding, J., Nascimento, M. A. (2004). Information-Content Based Sentence Extraction for Text Summarization. *IEEE International Conference on Information Technology: Coding and Computing*, 2, 214-218

78) 이일주 & 김민구. (2006). 그래프 분할을 이용한 문장 클러스터링 기반 문자요약. 『정보처리학회논문지』, 13-B(2). 149-154.

79) Kedzie, C., McKeown, K., Hal Daume III. (2018). Content Selection in Deep Learning Models of Summarization. 1818-1828.

추상 요약이라고도 하는 생성 요약 모델은 원본 문서의 문장들을 입력받아 원본 문서의 내용을 이해하고 이해한 내용을 바탕으로 새로운 문장을 간결하게 재작성하는 방법이다.⁸⁰⁾ 이처럼 원래 문서의 의미를 이해한 후 요약을 만들기 위해서는 복잡한 알고리즘이 필요하고 이를 구현하기가 훨씬 어렵기 때문에 생성 요약은 보통 신경망 모델과 같이 진보된 기술로 이뤄진다.

생성 요약은 입력 문서를 가져와서 해당 테스트에 대한 생성 요약을 실시하는 것으로 입력 텍스트는 입력 단어나 문자들의 시퀀스로 간주할 수 있으며, 요약은 출력 시퀀스로 간주할 수 있다. 따라서 문서 요약은 하나의 시퀀스를 다른 시퀀스로 변환하는 시퀀스 투 시퀀스(sequence-to-sequence) 문제로 볼 수 있다. 하지만 문서 요약 작업은 입력 문장의 길이가 상당히 길기 때문에 Long Distance Dependency 문제가 발생하기도 한다. 이를 해결하기 위해 주로 어텐션 방법론(attention mechanism)을 함께 사용한다. 또한 문서 요약의 경우 입력 시퀀스에 존재하는 단어가 출력 시퀀스에도 자주 등장하기 때문에 복사 방법론(copy mechanism)을 함께 사용하기도 한다.⁸¹⁾

3) 문서 요약 모델

가) TextRank⁸²⁾

TextRank는 비지도 학습 기반이며 2004년 제안된 이후 추출 문서 요약에 널리 사용되는 알고리즘이다.⁸³⁾ 이는 최근까지 가장 잘 알려진 알고리즘 중 하나인 PageRank⁸⁴⁾에 기반을 두고 있다. PageRank는 1998년 제안된 알

80) 전재원, 황현선, 이창기. (2019). 추출 요약과 생성 요약을 결합한 2단계 문서 요약. 『한국정보과학회 학술발표논문집』, 2019(6), 581-583.

81) 전재원. (2019). *딥러닝 기반의 2단계 한국어 문서 요약* (석사학위, 강원대학교 대학원).

82) Mihalcea, R. & Tarau. P. (2004). TextRank: Bringing Order into Text. *Association for Computational Linguistics*. 404-411.

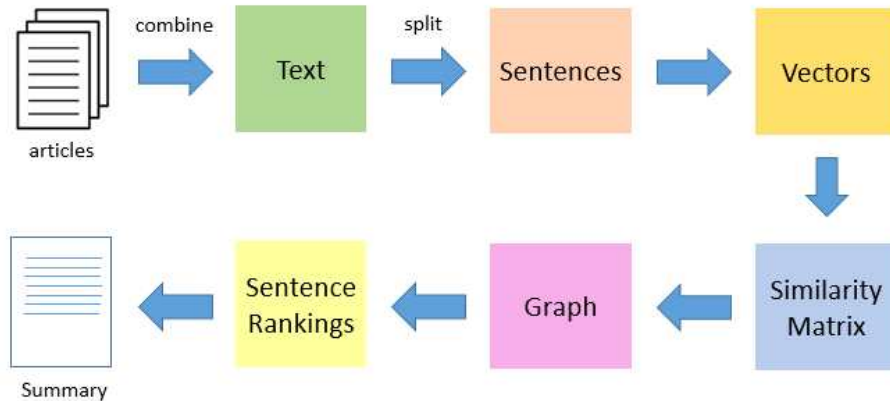
83) 쇼흠 고시 & 드와이트 거닝. (2020). *예제로 배우는 자연어 처리 기초* (김창엽, 최민환, 역). 서울: 에이콘출판(주). (원본 출판 2019).

84) Page, L., Brin, S., Motwani, R., Winograd, T. (1998). The PageRank Citation Ranking: Bringing Order to the Web. Stanford University.

고리즘으로 Google이 검색 결과를 정렬하는 데 사용한 최초의 검색 정렬 알고리즘으로 특정 페이지를 참조하는 다른 페이지의 총 수를 기반으로 페이지 순위를 지정하는 원칙에 따라 작동한다. 마찬가지로 TextRank 알고리즘은 텍스트에서 특정 단어가 다른 문장과 얼마나 관계를 맺고 있는지 계산하는 방식의 그래프 기반의 랭킹 모델(graph-based ranking model)이다. 어휘 및 문장간의 관계에서 중요성을 계산하여 중요한 키워드나 요약문을 생성해 낸다.⁸⁵⁾ 그 공식은 다음과 같다.

$$TR(V_i) = (1-d) + d^* \sum_{V_j \in \text{In}(V_i)} \frac{w_{ji}}{\sum_{V_k \in \text{Out}(V_j)} w_{jk}} TR(V_j) \quad \text{수식(2-1)}$$

여기서 $TR(V_i)$ 는 문장 또는 단어 V_i 에 대한 TextRank 값을 의미하고 w_{ij} 는 문장 또는 단어 i와 j 사이의 가중치, d는 댐핑 팩터(damping factor)로써 PageRank에서 웹 서핑을 하는 사람이 해당 페이지를 만족하지 못하고 다른 페이지로 이동하는 확률이다. 이 값을 TextRank에서도 그대로 사용한다.⁸⁶⁾ TextRank 알고리즘은 [그림 2-33]과 같이 동작한다.



[그림 2-33] TextRank 데이터 흐름

85) Mihalcea, op. cit., pp.404-411.

86) 강종렬, 남지성, 박지나, 김웅섭. (2019). 키워드 요약의 세 가지 방법론 비교. 『한국정보처리학회 2019년 추계학술대회 논문집』, 26(2), 852-854.

참고로, 젠심(Gensim) NLP 라이브러리는 자연어를 벡터로 변환하는데 필요한 대부분 기능을 제공해주는 라이브러리로 Word2Vec 등을 포함하고 있으며 문서 요약기도 포함돼 있다. 따라서 직접 구현한 문서 요약기를 쓰는 대신 이미 공개된 것을 프로젝트에 활용할 수 있으며 많은 기능을 가진 NLP 라이브러리이다. 젠심의 문서 요약기는 gensim.summarization 패키지에 있으며, 주요 메소드로는 summarize가 있으며 단어 개수, 문서의 백분율 등을 반환하기 위한 여러 가지 옵션을 제공한다. 기본 구현은 이미 앞서 설명한 TextRank 알고리즘으로 되어있다.

나) WordRank

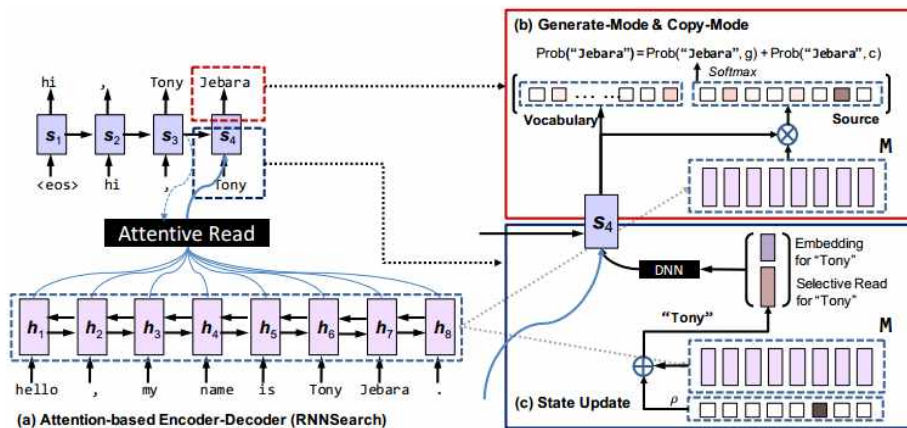
WordRank는 PageRank나 HTTS 같은 그래프 기반 랭킹 모델(graph-based ranking model)로서 HTTS algorithm을 사용한다. 중국어와 일본어에서 효과적이며 효율적인 성능을 보인 WordRank의 경우 한국어 단어 추출에 있어서 언어의 특징으로 인해 한계가 있었다. KR-WordRank⁸⁷⁾의 경우 추출 대상은 ‘의미를 지니며 실제로 자주 사용되는 독립적인 단어 혹은 복합형태소’이다. KR-WorkRank는 다음과 같은 장점이 있다. 첫째, 주어진 문서 집합에 포함된 주제 유형이 적을수록, 즉 문서 집합의 내용이 덜 이질적일수록 문서 내에서 단어와 키워드를 더 잘 추출할 수 있다. 둘째, 실제로 단어로 사용되는 표현을 추출하기 때문에 신조어나 오타의 영향을 덜 받는다. 마지막으로, 사전에 학습 데이터를 등록하지 않고도 단어를 인식할 수 있다.

WordRank와 마찬가지로 학습 데이터를 사용하지 않고 키워드를 추출하는 TextRank 역시 상호 강화 관계 방법을 사용하여 그래프에서 각 노드의 중요도인 authority를 계산한 뒤, 이를 단어/키워드 가능성으로 사용한다. 두 알고리즘의 차이점은 TextRank는 단어가 인식되는 상황에서 각 단어 간의 관계를 그래프로 표현한 후 키워드를 추출하는 반면, WordRank는 주어진 문서 집합에서 단어와 키워드를 추출한다는 점이다.

87) 김현중, 조성준, 강필성. (2014). KR-WordRank: WordRank를 개선한 비지도학습 기반 한국어 단어 추출 방법. 『대한산업공학회지』, 40(1), 18-33.

다) 복사 방법론(Copy Mechanism)

이 방법은 주로 sequence-to-sequence 모델의 디코딩 과정에서 단어를 출력할 때 사전에 없는 단어에 대한 문제(OOV: out-of-vocabulary)와 고유 명사들에 대한 출력 확률이 낮아지는 문제를 해결하기 위해 고안되었다. 출력에 필요한 어휘를 입력 열에서 찾아 출력 없이 복사(copy)하여 사용하는 방법이다.⁸⁸⁾ 사전에 없는 단어를 출력하려 할 경우 <UNK>와 같은 대체 토큰이 출력될 수도 있다. 이때 입력 시퀀스의 각 단어 임베딩마다 카피 어텐션(copy attention) 값을 따로 계산하고, 입력 시퀀스의 단어가 출력에 적합한 단어일 경우 카피 어텐션 값을 계산하여 출력 시퀀스로 출력한다.⁸⁹⁾



[그림 2-34] Copy Mechanism 모델

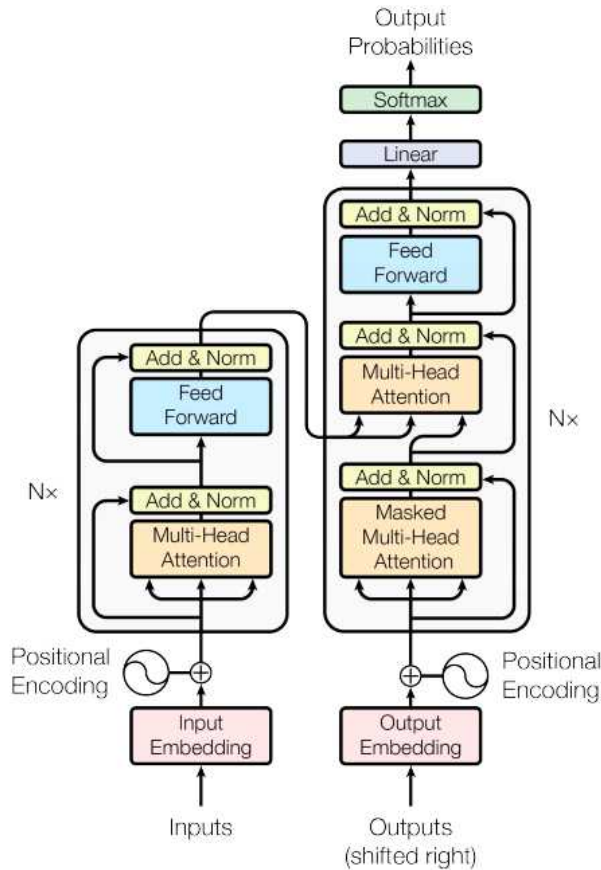
라) Transformer

Transformer는 기존의 sequence-to-sequence 모델에서 사용된 순환 신경망(RNN: Recurrent Neural Network) 레이어나 합성곱 신경망(CNN: Convolutional Neural Network) 레이어를 사용하지 않고 오직 어텐션

88) Gu, J., et al. (2016). Incorporating copying mechanism in sequence-to-sequence learning. arXiv preprint arXiv:1603.06393.

89) Lin, C. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. In Text Summarization Branches Out, p74-81.

(attention)만으로 구현한 인코더-디코더 모델이다.



[그림 2-35] Transformer 모델

중요한 특징은 데이터 처리 과정이 병렬로 수행되므로 학습시간을 단축할 수 있다. 인코더에 대한 입력은 먼저 하나의 특정 단어를 인코딩하기 위해 인코더에 대한 입력의 다른 단어의 의미를 참조하도록 구성된 멀티헤드 어텐션 (multi-head attention) 블록을 통과한다. 멀티헤드 어텐션에서는 헤드 수만큼 셀프 어텐션(self attention) 연산이 수행된다. 셀프 어텐션이란 입력 단어 간의 어텐션, 즉 그들 사이의 어텐션이다. 입력이 셀프 어텐션 층을 통과하고 출력은 피드 포워드(feed-forward) 신경망으로 다시 들어가게 된다. 동일한 피드 포워드 신경망이 각 위치의 단어에 독립적으로 적용돼 출력된다.⁹⁰⁾

마) BERT

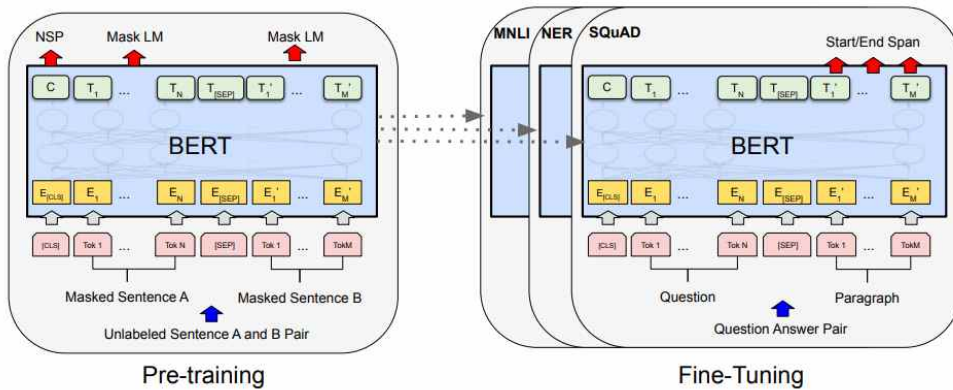
BERT(Bidirectional Encoder Representations form Transformers)의 약자로서 GPT와 유사한 트랜스포머(Transformer) 기반의 전이학습 모델이다. 트랜스포머의 구조를 거의 그대로 사용하고 있지만, 디코더(decoder) 부분을 제거하고 엔코더(encoder) 부분만 사용하는 특징이 있다. 트랜스포머처럼 인코더-디코더 구조를 이용해서 입력(input)이 있을 때 특정한 출력(output)을 생성하는 작업을 모델이라기보다는 대규모 말뭉치(corpus) 학습을 통해 단어의 분산 표현에 유리한 모델이라 볼 수 있다. 잘 학습된 단어 분산 표현을 통해 다양한 새로운 작업에 맞게 미세조정(fine-tuning)해서 사용할 수 있는 특징이 있다.⁹¹⁾

단어 임베딩 방법으로 자주 활용되는 Word2Vec 모델과의 차이점은 BERT는 문맥(context)을 고려한 모델이라는 점이다. 즉, BERT는 Word2Vec와는 달리 문장 형태와 위치에 따라 동일한 단어에 대해 다른 임베딩 벡터값을 가지므로 단어의 모호성 문제를 해결할 수 있다. 예를 들어 “bank account”와 “bank of the river”에서는 Word2Vec 모델에서는 벡터 값이 동일한 반면 BERT에서는 문장에 따라 임베딩 벡터 값이 다르다.⁹²⁾

90) Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017). Attention is all you need. *In Advances in Neural Information Processing Systems*, 6000–6010.

91) Devlin, J., Chang, M., Lee, K., Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

92) 황상흠 & 김도현. (2020). 한국어 기술문서 분석을 위한 BERT 기반의 분류모델. 『한국전자거래학회지』, 25(1), 203–214.



[그림 2-36] BERT 모델

바) RoBERTa

RoBERTa(A Robustly Optimized BERT Pretraining Approach)⁹³⁾는 이 름에서 알 수 있듯 BERT에서 파생된 모델로 기존 BERT에 학습방법을 추가 하여 성능을 개선한 모델이다. 모델의 큰 구조적인 변화 없이 학습방법을 부 분적으로 수정하여 모델의 성능을 크게 향상시킬 수 있음을 보여준다. 연구팀 은 다양한 실험을 통해 BERT가 여전히 과소훈련(undertrain) 되어있고, 사전 훈련(pre-training) 과정에서 4가지 기법을 적용하면 성능을 향상할 수 있다.

BERT와 비교하여 배치의 크기를 증가시켜 더 많은 학습 데이터로 더 오래, 더 크게 학습을 수행한다. 사전 훈련(pre-training) 시 NSP(Next Sentence Prediction) 태스크를 생략하고 MLM(Masked Language Model) 태스크만 학습에 사용했다. BERT는 짧은 입력 문장들을 사용하는 다운스트림(downstream) 작업에 대한 사전훈련 시, 0.1의 확률로 최대 길이보다 짧은 길이의 데이터를 사용한다. 이 로직을 제거하고 최대 길이의 데이터로만 학습을 진행한다. 마지막으로 마스킹(Masking) 시, BERT의 정적 마스킹(static masking)을 대신하여 입력을 만들 때마다 마스킹하는 동적 마스킹(dynamic masking) 기법을 적용한다.

93) Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv.

사) GPT-3

GPT(Generative Pre-trained Transformer 3)⁹⁴⁾는 트랜스포머의 디코더 부분을 이용해서 만든 모델로 2018년 6월 11일 OpenAI에서 발표하였다. BERT보다 먼저 발표된 것으로, GPT(Generative Pre-trained Transformer)를 우리말로 직역하면 생성적 사전학습 트랜스포머이다. 여기서 생성적이란 변수들의 관계를 밝히는데 사용되는 통계적 모델을 의미한다. GPT는 주어진 단어를 기반으로 다음에 오는 단어를 맞추는 방식으로 사전학습을 진행한다.

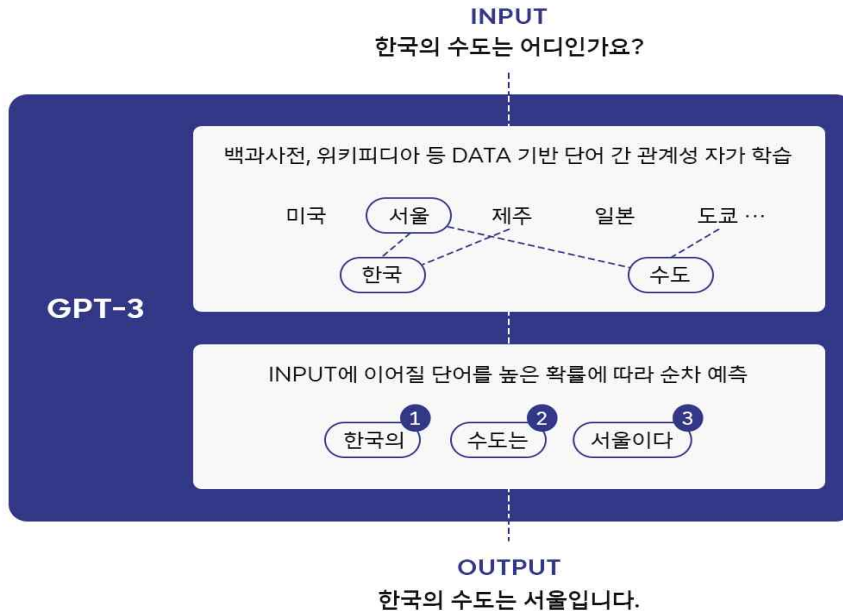
학습 진행 시 문장에서 토큰별로 어텐션 가중치를 이용하여 가장 근접한 다음 문장을 찾아내는 방식으로 학습하고 마지막 문장의 경우 추가 예측할 내용이 없기 때문에 가중치를 '0'으로 준다. 그래서 주어진 토큰이 아닌 미래의 토큰으로부터 문맥 정보를 수집하지 못하도록 해주는 셀프 어텐션(self-attention) 기법을 마스크드 셀프 어텐션(masked self-attention)이라고 한다.

이처럼 GPT에서 마스크드 셀프 어텐션을 사용함으로써 BERT와 큰 차이점을 가지게 된다. BERT는 한 위치에 해당하는 토큰을 예측하는 MLM(Masked Language Model)을 실행할 때 필요한 정보를 앞뒤 양방향으로부터 토큰들로부터 가져온다. BERT의 B는 양방향(bidirectional)을 의미한다. 반면 GPT는 예측할 토큰 뒤에 나오는 토큰으로부터 문맥 정보를 가져올 수 없는 단방향(unidirectional)의 성격을 지니고 있다.⁹⁵⁾

GPT 언어 모델은 자연어 처리의 기반이 되는 조건부 확률 예측 도구로서 사람처럼 문장 자체를 이해하는 것은 아니다. 하지만 글을 스스로 쓸 수 있는 이유는 어마어마한 학습량 때문이다. Open AI의 GPT-1은 1억1,700만개의 매개변수로 학습을 했고, GPT-2는 기존의 다양한 NLP태스크를 Question Answering 태스크로 통합했다. 또 모델 크기에 따라 약 1억 2,400만 개에서 15억 개로 10배 정도 파라미터를 늘린 모델이다. 그리고 GPT-3는 GPT-2를 발전시킨 모델로 구조적 변화는 없이 더 많은 데이터를 사용하여 셀프 어텐션 레이어를 쌓아 1,750억 개의 파라미터를 가지고 높은 성능을 달성했다.

94) Brown, T. B, et al. (2020). Language Models are Few-Shot Learners, *arXiv preprint arXiv:2005.14165*.

95) 이진기. (2022). 『안녕,트랜스포머』. 서울: 에이콘출판주식회사.



[그림 2-37] GPT-3의 학습 및 출력

KoGPT는 한국어에 특화된 GPT 모델이다. [그림 2-37]는 GPT-3의 학습 예로서, 단어들 간의 연관성을 파악하면서 스스로 학습을 통해 한국, 미국, 서울, 제주, 수도, 도시 등에서 한국, 수도와는 서울이 통계 확률적으로 가깝다는 것을 알아내는 것이다. GPT-3의 대답 과정은 “한국의 수도는 어디인가?”에 이어질 단어로 확률이 가장 높은 ‘한국의’이고, 이어서 나올 후보 단어는 ‘수도는’, 마지막으로 ‘서울입니다.’를 순차적으로 찾아내는 것이라고 할 수 있다. 이런 지능은 백과사전 등의 글들을 사전 학습하여 얻어낸 결과이다. GPT-3는 이런 단어들의 관계를 한국, 서울, 수도의 3차원이 아닌 거대 규모의 차원으로 계산하기 때문에, 짧은 문장 하나가 아니라, 기사나 짧은 에세이 정도 분량의 문장으로 단어들을 만들어 낼 수 있는 것이다.⁹⁶⁾

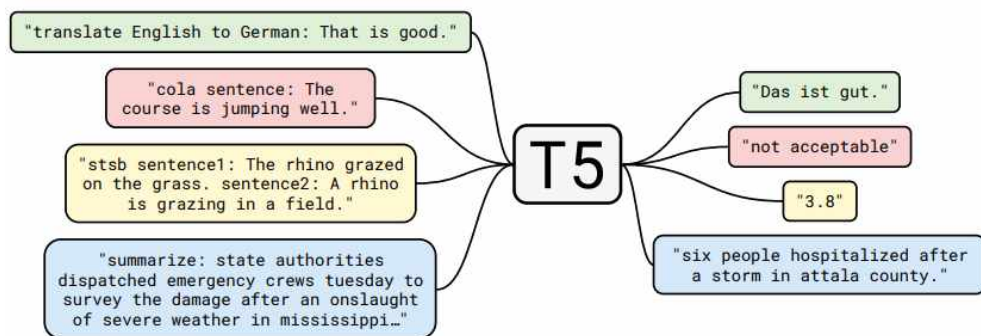
GPT가 문장을 요약하는 방식은 앞서 설명한 방식으로 문서의 전체적인 맥락을 이해한 후 그 내용을 표현하기 적합한 간결한 문장을 작성하는 방식으로 요약된 새로운 문서를 생성하는 것이다.

96) 에이닷에 적용된 거대언어모델 GPT-3가 무엇일까?, SK텔레콤. (2022년5월26일). <https://news.sktelecom.com/178491>

아) T5(Text-To-Text Transfer Transformer)⁹⁷⁾

최근 NLP 기술의 동향을 살펴보면 전이학습(transfer learning)이 주축을 이루고 있다. 전이학습은 레이블 되지 않은 풍부한 텍스트데이터에서 자기 지도 태스크(self-supervised task)로 미리 학습하여 모델을 만든다. 그리고 희소하게 레이블이 지정된 텍스트데이터로 모델을 파인튜닝 한다. NLP 기반의 전이학습 방법으로 GPT, ULMFiT, ELMo, BERT, XLNet, RoBERTa, ALBERT, Reformer, MT-DNN 등이 지속적으로 나오고 있는데 이 분야에서는 어떤 개선 사항이 중요하고, 얼마나 효과가 있는지에 대한 평가하기가 힘들다. 기존 BERT 기반의 모델들은 태스크별 출력 형태가 다르게 되어있었다. 그리고 기존 데이터셋에는 품질 이슈나 데이터의 다양성에 대한 이슈가 존재한다.

전이학습 방법들의 효과에 대한 대규모의 실증적인 조사를 하고, 이 결과를 반영하여 T5(Text-To-Text Transfer Transformer)라는 새로운 모델을 만들었다. T5는 BERT 기반의 모델들과 달리 통합된 텍스트 입력-텍스트 출력으로 NLP 태스크를 처리한다. 이러한 특성으로 어떤 NLP 태스크에 대해서도 같은 모델, loss 함수, 그리고 하이퍼 파라미터를 쓸 수 있다.



[그림 2-38] 텍스트 대 텍스트 프레임워크 다이어그램

97) Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P. J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, 21(2020), 1-67.

T5를 한국어 버전으로 학습시킨 모델로는 KETI(Korea Electronics Technology Institute)에서 공개한 KE-T5⁹⁸⁾와 와이즈넷에서 공개한 KoT5가 있다. KE-T5는 구글의 언어 모델 알고리즘을 T5를 기반으로 구축한 언어처리 모델이다. 한국어와 영어처리에 모두 활용 가능한 자가지도 학습 방식 이중 언어처리를 할 수 있는 특징이 있고, 언어의 의미 인식 특성과 표현 특성을 모두 포함하는 범용의 언어 모델로서 소수의 학습 데이터만으로 다양한 언어처리가 가능하다.

KoT5는 와이즈넷에서 개발한 모델이며 구글의 T5 모델 구조를 그대로 사용하며, 한국어 데이터를 학습한 모델로 각 레이어마다 셀프 어텐션(self attention), 멀티헤드 어텐션(multi-head attention), 피드 포워드(feed forward) 구조를 가지고 있다.

98) 신사임, 김산, 서현태. (2021). KE-T5: 한국어-영어 대용량 텍스트를 활용한 이중언어 사전학습기반 대형 언어모델 구축. 『한국정보과학회 언어공학연구회: 학술대회논문집』, 2021.10a, 419-422.

제 5 절 요약 및 시사점

1) 선행연구 분석 요약

재난과 관련된 연구는 이미 오래전부터 진행되어온 과업이었음을 확인할 수 있었다. 과거 자연재해에 대한 연구 중심에서 사회재난과 복합적인 신규 재난들에 대해서 재난의 4단계라는 개념을 만들고 이에 대해 대응을 하고 있었다. 또한 재난 상황을 전파하는 방법에 대한 연구에서 벗어나 재난을 예측하는 방법에 대한 다양한 연구도 진행되고 있었다. 인공위성과 GPS를 이용한 정보획득과 인공지능 기술을 활용한 예측 기술들이 연구되었다.

또한, 다중에게 재난 상황을 전파하는 방식에 있어서도 다양한 매체를 통하여 위험성을 알리고, 발생한 재난에 대한 인식율을 높여서 피해를 최소화하려는 노력을 하고 있었다.

데이터의 중요성을 체감하고 있는 현대사회에서는 이를 활용한 이윤추구나 공적인 목적의 서비스를 위해서 공공데이터를 사용하고 있었다. 하지만 사회가 복잡해지고 다양해지면서 필요로 하는 데이터의 종류도 더욱 다양해지고 있다. 하지만 이러한 부분들에 대해서도 공공이나 민간에서도 적극적으로 데이터를 공개하고 있다.

또한 인공지능 기술의 발전으로 인하여 인간의 언어를 이해하고 생성해내는 기술인 자연어 처리도 발전하고 있다. 앞서 설명한 공공데이터 개방과 관련하여 공공데이터 포털에서는 자연어 처리에서 사용 가능한 텍스트 데이터셋도 공개하고 있다. 이는 자연어 처리에서 굉장히 중요한 양질의 학습 데이터를 국가에서 지원하는 것으로 이는 개발비용과 시간을 단축시키는 중요한 사업으로 인공지능 발전에 큰 도움이 되고 있다.

과거의 자연어 처리 중 요약의 연구 방향은 추출 요약이 중심이었다면 최근의 문서 요약의 연구 방향은 생성 요약으로 귀결된다. 구글의 BERT, Open AI의 GPT-3 모두 트랜스포머를 기반으로 하는 최신 모델들로 좋은 성능을 나타내고 있다. 이러한 배경으로 트랜스포머 계열의 언어 모델들에 대한 연구가 활발히 진행되고 있다.

2) 시사점

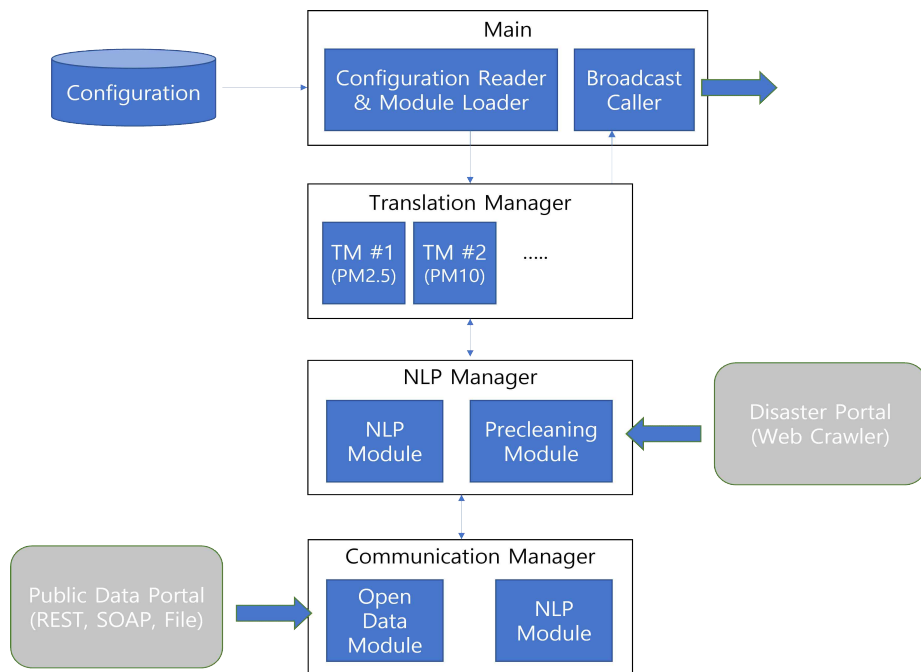
앞서 살펴본 재난관리, 재난방송, 공공데이터 활용, 그리고 자연어 처리 등의 연구들이 나름대로 해당 분야에서는 의미있는 연구들이 진행되고 있음을 확인할 수 있었다. 하지만 이러한 기술들이 융합되어서 재난 안전을 위한 새로운 시스템으로 발전하지는 못하고 있는 것으로 보인다. 그 이유로 재난과 관련된 분야의 법률적인 문제, 즉 재난안전법, 소방법 등으로 인한 제한점도 있을 수 있고, 그 외에도 이해관계와 관련한 문제 등 다양한 문제가 있을 수 있지만, 그 외에 과학기술인이 할 수 있고 해야 할 일로는 이러한 기술들을 융합하여 더 나은 결과를 얻는 것이다.

본 연구에서는 이러한 연구방법들을 융합하고 결과물의 품질에 대한 평가 방법을 개선하였다. 그래서 보다 효율적인 안내 방법에 대한 연구를 수행하는 것이 의미 있다고 판단하였다.

제 3 장 시스템 설계

제 1 절 연구 시스템 개요

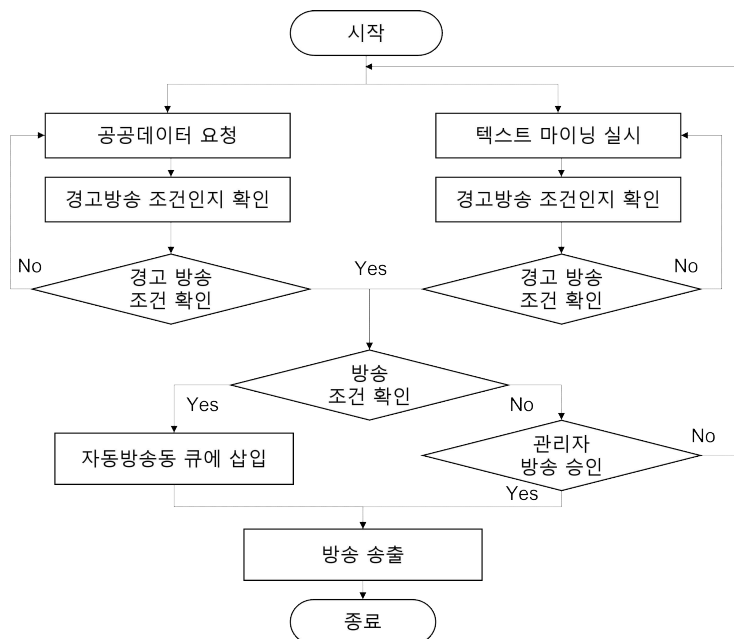
연구시스템은 크게 공공데이터 포털로부터 OpenAPI를 이용하여 실시간 공공데이터를 가져와서 방송용 데이터로 가공하는 Communication Manager와 국가 재난 포털 및 언론사에서 운영하는 재난 포털 웹사이트를 통해 배포된 재난 관련 기사 등을 웹 크롤링을 통해 수집 후 텍스트마이닝과 자연어 처리를 통해 문서를 요약 정리하는 NLP(Natural Language Processing) Manager로 구성된다. 이렇게 취득한 데이터는 운영시스템에서 사용자에게 의해 정의된 경고 방송 송출 요건에 해당되는 경우 방송 장비에서 미리 정의된 REST API를 통해 송출요청을 하게 되는데 여기까지가 연구의 범위이다.



[그림 3-1] 시스템 환경

연구의 범위에 속한 시스템은 실시간 재난정보를 알려주는 시스템과는 차이가 있다. 재난 대비정보를 공공데이터와 재난 포털에서 제공하는 재난 대비정보를 비상방송설비에서 사용할 수 있는 정보로 재가공하는 것이다. 비상방송설비(전관방송장비)는 재난안전관리법상의 재난 안내 방송의무는 없다. 다만 소방청 고시 화재안전기준에 의한 화재방송은 일정 기준에 의해 방송되어야 한다는 기준(NFSC 202)이 있다. 그러므로 제공하는 데이터 또한 실시간 데이터는 없으며, 공공데이터 포털이나 재난 포털에 공개된 데이터를 5분 간격으로 확인하고 대비를 위한 경고 방송용 콘텐츠인지 판단하여 운영한다.

데이터수신 서버의 구성은 크게 NLP(Natural Language Processing) Manager, CM(Communication Manager), TM(Translation Manager), Main Processor로 구성된다. 그중 CM은 REST(REpresentational State Transfer) API를 사용하여 공공데이터를 가져오고 NLP를 통하여 자연어 처리가 된 데이터를 가져오는 역할을 한다. 다음으로 TM은 데이터 정규화를 통해 가져온 데이터를 시스템에서 분석이 가능한 형태로 변환한다. 마지막으로 메인에서는 불러온 데이터가 방송 조건과 일치하는지 판단한다.



[그림 3-2] 방송 결정 흐름도

메인을 구성하는 방송 발신자는 데이터 분석 처리를 수행하고 방송이 필요한지 판단하고 방송 조건이 충족되면 전관방송 시스템에 방송을 요청한다. 여기서 가장 큰 특징은 웹 크롤링을 통해 가져온 뉴스가 시·군·구 단위의 광범위한 지역에 대한 정보라면, 공공데이터는 장비가 설치된 동단위로 좀 더 구체적인 현장이라는 차이점이 있다. 이 경우, 판단에 도움이 되도록 실시간 공공데이터를 통해 획득한 정보를 참고하여, 아래의 흐름도와 같이 조건을 만족하면 지정된 시간에 알림 요청을 통해 방송을 송출한다.

예를 들어, 재난 포털의 뉴스에서 기상특보가 발령되더라도 특정 지역(읍면동 등)의 공공데이터에서는 미세먼지가 기준 이하로 포착 될 경우가 있다. 이 경우 방송할 필요는 없지만, 공공데이터에서도 미세먼지 농도도 높게 측정되면 경고 방송을 위한 조건으로 판단한다.

반대로, 특정 지역의 공공데이터에서 미세먼지 농도가 일시적으로 높게 나와도 재난 포털에서 미세먼지에 대한 경보가 없었다면, 이는 자동으로 방송되지 않고 운영자 시스템 관리 화면에 확인 윈도우가 표시되도록 하여 운영자가 방송 여부를 결정할 수 있다. 운영자는 팝업 방식의 화면을 통해 방송 여부를 결정한다. 이 경우, 방송 방식이 자동 또는 수동으로 결정되면 방송을 위한 텍스트와 방송에 필요한 정보가 이더넷을 통해 전관방송 시스템으로 전송된다. 전관방송 시스템은 자체 TTS 모듈을 사용해 이 텍스트를 음성으로 재생해서 전관방송에서 사용할 음원으로 활용한다.

전관방송 시스템에서는 문자 및 데이터를 수신한 직후 또는 일정 시간 후에 재난정보 방송 우선순위에 따라 방송을 송출하여 불특정 다수의 건물 입주자에게 대비를 위한 정보를 제공한다. 따라서 미세먼지 등으로 인한 경제적, 사회적 비용이 발생하지 않도록 사전에 안내하는 것이 가능하다.

제 2 절 시스템 구성

1) 공공데이터 수집

인터넷을 통해 재난 혹은 대기오염 등의 공공데이터를 가져오기 위해 공공데이터 포털을 이용하여 국가의 각 기관에서 제공하는 데이터를 가져올 수 있는데 재난과 관련해서는 행정안전부에서 재난 안전 정보를 제공하고 있으며 갱신주기가 1년 정도로 긴 통계데이터 위주로 제공되고 있다. 실시간 재난정보의 경우 주로 날씨나 대기오염과 같은 종류의 공공데이터가 제공되는데 본 연구의 목적을 달성하기 위해서는 실시간으로 제공되는 공공데이터인 대기오염 정보를 수집데이터로 정하였다.

공공데이터를 받아오는 데이터수신 서버는 공공데이터 포털에서 제공하는 재난정보를 가져오기 위해 전관방송 시스템의 상위 노드에 설치된다. 데이터 수신 시 [그림 3-1]의 Communication Manager(이하 CA)로부터 데이터를 받아오는데 CA는 공기질과 관련된 공공데이터를 환경부 산하기관인 한국환경공단에서 운영하는 에어코리아의 대기오염정보 조회 서비스를 이용하였다. 데이터 제공방식으로 REST, SOAP 및 파일 메소드가 있는데, 본 연구에서는 Open API 중 하나인 REST 방식을 이용하여 정보를 수집하여 데이터 세트에서 분석에 필요한 미세먼지 관련 데이터를 서버 데이터베이스에 저장하였다. 그 외에 공공데이터의 실험 조건은 다음 [표 3-1]과 같다.

[표 3-1] 대기오염정보 API 개요 및 실험 조건

API명	대기오염정보 조회 서비스
API 설명	각 측정소별 대기오염정보를 조회하기 위한 서비스로 기간별, 시도별 대기오염 정보와 통합대기환경지수 나쁨 이상 측정소 내역, 대기질(미세먼지/오존) 예보 통보 내역 등을 조회 ⁹⁹⁾
서비스 인증/권한	[O] ServiceKey [] 인증서 (GPKI/NPKI) [] Basic (ID/PW) [] 없음
메시지 레벨 암호화	[] 전자서명 [] 암호화 [O] 없음
전송 레벨 암호화	[] SSL [O] 없음
인터페이스 표준	[] SOAP 1.2

	[O] REST (GET) [] RSS 1.0 [] RSS 2.0 [] Atom 1.0 [] 기타												
교환 데이터 표준 (중복선택가능)	[O] XML [O] JSON [] MIME [] MTOM												
메시지 교환유형	[O] Request-Response [] Publish-Subscribe [] Fire-and-Forgot [] Notification												
대기오염정보 데이터 생성(갱신)주기	<table border="1"> <thead> <tr> <th>상세기능명</th><th>데이터 생성주기</th></tr> </thead> <tbody> <tr> <td>측정소별 실시간 측정정보 조회</td><td>매시 15분 내외</td></tr> <tr> <td>통합대기환경지수 나쁨이상 측정소 목록조회</td><td>매시 15분 내외</td></tr> <tr> <td>시도별 실시간 측정정보 조회</td><td>매시 15분 내외</td></tr> <tr> <td>대기질 예보통보 조회</td><td>매일 4회(5,11,17,23시) 각 시별 10분 내외</td></tr> <tr> <td>초미세먼지 주간예보 조회</td><td>1일 1회, 17시30분 내외</td></tr> </tbody> </table>	상세기능명	데이터 생성주기	측정소별 실시간 측정정보 조회	매시 15분 내외	통합대기환경지수 나쁨이상 측정소 목록조회	매시 15분 내외	시도별 실시간 측정정보 조회	매시 15분 내외	대기질 예보통보 조회	매일 4회(5,11,17,23시) 각 시별 10분 내외	초미세먼지 주간예보 조회	1일 1회, 17시30분 내외
상세기능명	데이터 생성주기												
측정소별 실시간 측정정보 조회	매시 15분 내외												
통합대기환경지수 나쁨이상 측정소 목록조회	매시 15분 내외												
시도별 실시간 측정정보 조회	매시 15분 내외												
대기질 예보통보 조회	매일 4회(5,11,17,23시) 각 시별 10분 내외												
초미세먼지 주간예보 조회	1일 1회, 17시30분 내외												

에어코리아의 OpenAPI를 이용하여 SO₂, CO, O₃, NO₂(ppm), PM₁₀, PM_{2.5}($\mu\text{g}/\text{m}^3$) 등의 대기환경 오염정보를 가져온다. 이러한 정보들을 정의된 주기마다 수집한다.¹⁰⁰⁾

실제 사용한 대기오염 정보와 관련된 공공데이터 API 사용의 경우 각 측정소별 대기오염 정보를 조회하기 위한 서비스로 기간별, 시도별 대기오염 정보와 통합대기 환경지수 나쁨 이상 측정소 내역, 대기질(미세먼지/오존) 예보통보 내역 등을 조회할 수 있다. 또한, 요청 시 결과물의 리턴 타입을 XML 또는 JSON 중 선택할 수 있고 페이지당 결과 수, 측정소명, 기간 등을 지정하여 요청할 수 있으며 이에 대한 응답 메시지는 [표 3-2]¹⁰¹⁾와 같다.

[표 3-2] 응답 메시지 명세

항목명(영문)	항목명(국문)	항목 크기	항목 구분	샘플데이터	항목설명
resultCode	결과코드	2	1	00	결과코드
resultMsg	결과메세지	50	1	NORMAL	결과메세지

99) OpenAPI - 공공데이터포털. <https://www.data.go.kr>

100) 류태하 & 김승천. (2022). 공공데이터와 IoT 센싱 데이터를 활용한 경보방송 방법에 관한 연구. 『한국인터넷방송통신학회 논문지』, 22(1), 21-27.

101) 에어코리아 대기오염정보 조회 서비스 기술문서, v1.0, (2020).
<https://www.data.go.kr/data/15073861/openapi.do>

항목명(영문)	항목명(국문)	항목 크기	항목 구분	샘플데이터	항목설명
				SERVICE.	
numOfRows	한 페이지 결과 수	4	1	100	한 페이지 결과 수
pageNo	페이지 번호	4	1	1	페이지 번호
totalCount	전체 결과 수	4	1	60	전체 결과 수
items	목록	-	0..n	-	목록
dataTime	측정일	20	1	2020-11-25 13:00	오염도측정 연-월-일 시간: 분
mangName	측정망 정보	10	1	도시대기	측정망 정보 (도시대기, 도로변대기, 국가배경 농도, 교외대기, 항만)
so2Value	아황산가스 농도	10	1	0.007	아황산가스 농도 (단위 : ppm)
coValue	일산화탄소 농도	10	1	0.4	일산화탄소 농도 (단위 : ppm)
o3Value	오존 농도	10	1	0.043	오존 농도 (단위 : ppm)
no2Value	이산화질소 농도	10	1	0.024	이산화질소 농도 (단위 : ppm)
pm10Value	미세먼지(PM10) 농 도	10	1	73	미세먼지(PM10) 농도 (단위 : $\mu\text{g}/\text{m}^3$)
pm10Value24	미세먼지(PM10) 24시간예측이동농도	10	0	55	미세먼지(PM10) 24시간예측이동농도 (단위 : $\mu\text{g}/\text{m}^3$)
pm25Value	미세먼지(PM2.5) 농도	10	0	44	미세먼지(PM2.5) 농도 (단위 : $\mu\text{g}/\text{m}^3$)
pm25Value24	미세먼지(PM2.5) 24시간예측이동농도	10	0	31	미세먼지(PM2.5) 24시간예측이동농도 (단위 : $\mu\text{g}/\text{m}^3$)
khaiValue	통합대기환경수치	10	1	75	통합대기환경수치
khaiGrade	통합대기환경지수	10	1	2	통합대기환경지수
so2Grade	아황산가스 지수	10	1	1	아황산가스 지수
coGrade	일산화탄소 지수	10	1	1	일산화탄소 지수
o3Grade	오존 지수	10	1	2	오존 지수
no2Grade	이산화질소 지수	10	1	1	이산화질소 지수
pm10Grade	미세먼지(PM10) 24시간 등급	10	1	2	미세먼지(PM10) 24시간 등급자료
pm25Grade	미세먼지(PM2.5) 24시간 등급	10	0	2	미세먼지(PM2.5) 24시간 등급자료

항목명(영문)	항목명(국문)	항목 크기	항목 구분	샘플데이터	항목설명
pm10Grade1h	미세먼지(PM10) 1시간 등급	10	0	2	미세먼지(PM10) 1시간 등급자료
pm25Grade1h	미세먼지(PM2.5) 1시간 등급	10	0	2	미세먼지(PM2.5) 1시간 등급자료
so2Flag	아황산가스 플래그	10	1	점검및교정	측정자료 상태정보(점검및교정, 장비점검,자료이상,통신장애)
coFlag	일산화탄소 플래그	10	1	점검및교정	측정자료 상태정보(점검및교정, 장비점검,자료이상,통신장애)
o3Flag	오존 플래그	10	1	점검및교정	측정자료 상태정보(점검및교정, 장비점검,자료이상,통신장애)
no2Flag	이산화질소 플래그	10	1	점검및교정	측정자료 상태정보(점검및교정, 장비점검,자료이상,통신장애)
pm10Flag	미세먼지(PM10) 플 래그	10	1	점검및교정	측정자료 상태정보(점검및교정, 장비점검,자료이상,통신장애)
pm25Flag	미세먼지(PM2.5) 플래그	10	0	자료이상	측정자료 상태정보(점검및교정, 장비점검,자료이상,통신장애) ¹⁰²⁾

또 다른 공공데이터로 기상청의 기상특보 조회 서비스를 통하여 기상특보를 가져올 수가 있는데 받아 올 수 있는 내용으로는 기상특보목록, 기상특보 통보문, 기상정보목록, 기상정보문, 기상속보목록, 기상속보, 기상예비특보목록, 기상예비특보, 특보코드, 특보현황 정보 등을 수신할 수 있다. 이 또한 OpenAPI를 통해 정보를 요청할 때 요청 메시지에 수신할 데이터의 종류를 지정하여 가져올 수 있다.

기상청의 공공데이터 중 기상특보 외에도 지진정보, 태풍정보, 영향 예보(폭염, 한파), 생활기상지수(동파, 자외선, 대기확산, 체감온도) 등 다양한 서비스가 있으며 데이터의 갱신주기 또한 수시로 갱신되는 기상이나 태풍 등으로부터 1시간, 3시간 단위의 생활기상지수 그리고 하루에 한 번 갱신되는 영향 예보 등 다양하다. 다양한 서비스를 일정 주기로 불러오는데 따른 부담이 있기 때문에 서비스를 위해서는 기상특보를 통해 강풍, 풍랑, 호우, 대설, 건조, 폭풍해일, 한파, 태풍, 황사, 폭염 등 10개의 항목에 대해서 대비용 콘텐츠를 생성하는 것으로 범위를 제한하였다.

102) OpenAPI - 공공데이터포털. <https://www.data.go.kr>

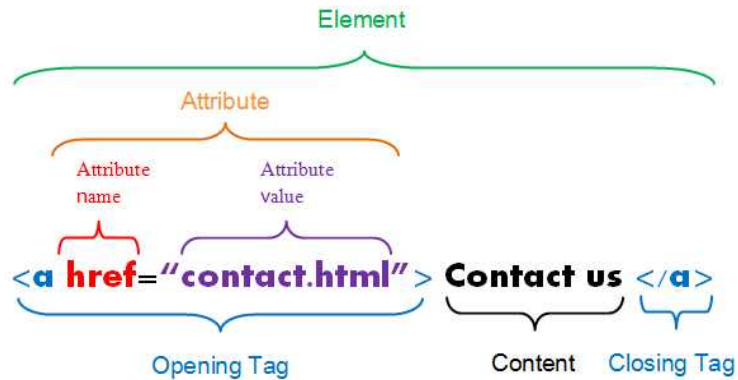
2) 텍스트 마이닝

방송 조건에 대한 신뢰성을 추가로 확보하기 위해서 공공데이터를 이용한 재난 데이터수신에 더해서 웹 크롤링을 통해 재난 포털 사이트로부터 재난과 관련된 속보 등의 정보를 가공하여 제공하는 방법을 추가하였다. 텍스트 마이닝은 자연어 처리 기술을 사용하여 텍스트를 정형화된 데이터로 변환하고, 머신러닝 기법을 적용하여 우리가 관심 있는 이벤트를 예측하는 방법론으로 본 연구에서는 연합뉴스 재난 포털(<https://www.yna.co.kr/safe/index>) 및 KBS 뉴스 홈페이지의 재난·환경 카테고리에 있는 뉴스 중 재난과 관련된 최신 기사를 그 대상으로 하였다.

이번 항에서 설명하고자 하는 것은 [그림 3-1]에 있는 NLP Manger이다. 웹 페이지로부터 텍스트 데이터를 수집하고 정보를 추출하는데 있어서 대부분의 웹사이트에서 사용하고 있는 HTML(Hypertext Markup Language)에 대한 처리가 필요하였다. HTML은 이름에서 알 수 있듯이 마크업 언어이다. 태그 등을 이용하여 문서나 데이터의 구조를 지정하는 언어 중 하나이다. 원래 태그는 텍스트와 별도로 원고에 교정 코드와 주석을 표현하기 위한 것이었지만 점차 문서의 구조를 표현하는 용도로 확대되었다.¹⁰³⁾ 웹 페이지의 이러한 특징 때문에 텍스트데이터를 추출하려면 HTML의 구성을 이해해야 하는데 기본적인 구조는 [그림 3-3]과 같다.¹⁰⁴⁾

103) Coombs, James H., Renear, Allen H., DeRose, Steven J. (1987). Markup system and the future of scholarly text processing. Communications of the ACM, 30(11), 933-047.

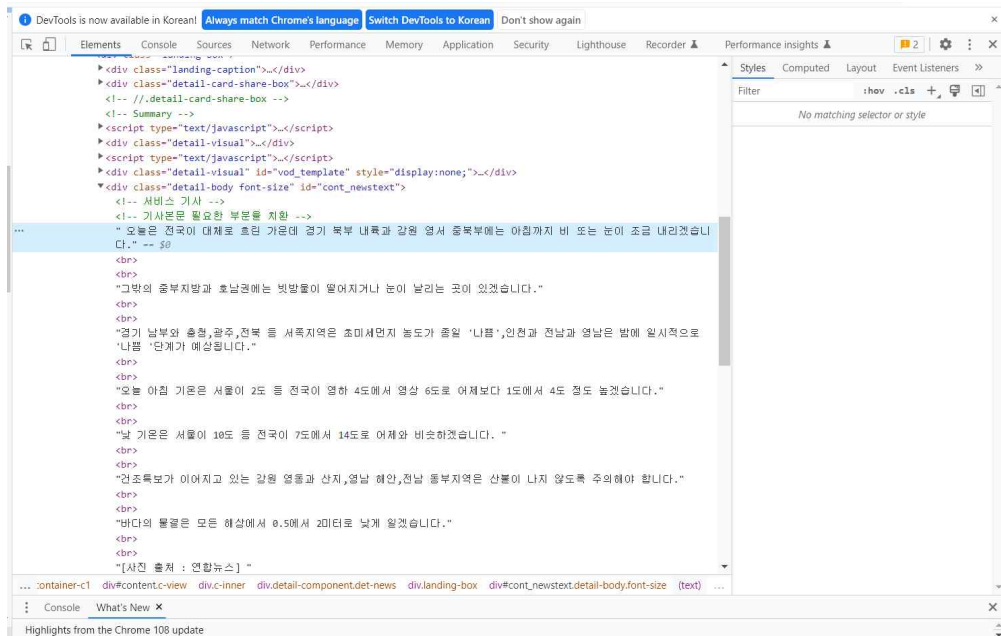
104) Jiang, Yue. (2020). Web Scraping Intro to Data Science.
<https://www2.stat.duke.edu/courses/Spring20/sta199.002/slides/web-scraping.html#4>



[그림 3-3] HTML 태그와 애트리뷰트

HTML 파일로부터 각종 태그들을 처리하고 필요한 정보를 추출하기 위한 방법으로 전통적인 HTTP(HyperText Transfer Protocol)의 구조인 클라이언트/서버 모델을 이용하여 클라이언트인 브라우저에서 연결을 열고 메시지를 HTTP서버(Nginx, Apache 등)에 보내고, 서버는 응답 후 연결을 닫는 구조를 활용한 전통적인 Socket 방식이 있다. 이는 파이선에서 HTTP 요청을 수행하기 위해서 TCP 소켓을 열고 수동으로 HTTP 요청을 보내는 방법이다. 이 외에 urllib3 & LXML 라이브러리를 이용하는 방법이 있는데 urllib 및 urllib2 라이브러리는 파이선에서 기본으로 제공하는 라이브러리이지만 urllib3는 개선된 버전으로 따로 설치를 해주어야 한다. urllib3는 HTTP 요청으로 대부분의 원하는 작업을 수행할 수 있는 고급 패키지로서 간결한 코드로 다양한 작업을 수행할 수 있는 장점이 있다. 다음으로는 Requests & BeautifulSoup 패키지가 있는데 이는 가장 대중적인 패키지로서 HTTP의 POST 요청을 수행하고, 쿠키 처리 및 매개변수를 쉽게 질의할 수 있는 장점이 있다.¹⁰⁵⁾

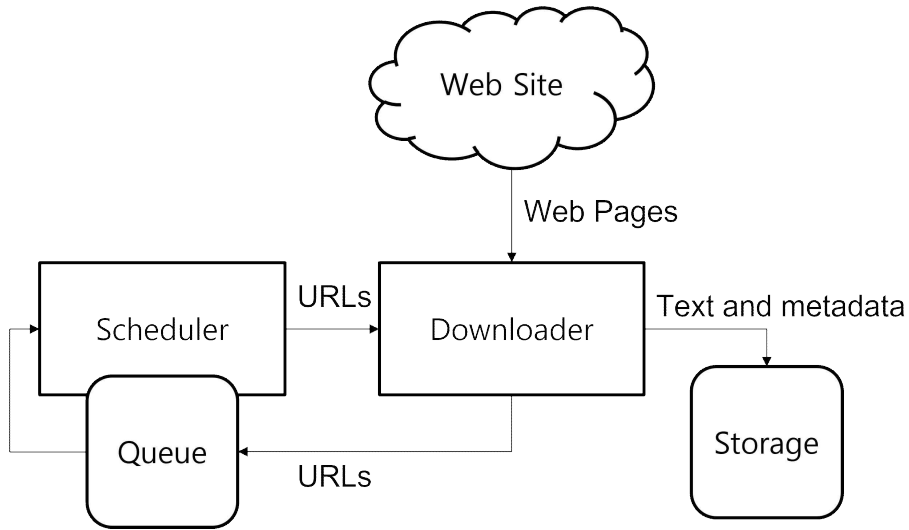
105) Web Scraping With Python: Everything You Need To Know. (2022).
<https://www.scrapingbee.com/blog/web-scraping-101-with-python/>



[그림 3-4] KBS뉴스 웹 접근 화면

본 연구에서는 웹 스크래핑 능력이 차지하는 비중이 크지는 않기 때문에 웹 스크래핑 방법으로 비교적 가벼운 BeautifulSoup & Request 라이브러리를 이용해서 웹 스크래퍼를 구성하고, 이를 실행하여 분석할 텍스트를 다운로드 하였다. BeautifulSoup은 해당 정보를 텍스트 파일로 변환하여 스토리지(Storage)에 텍스트 파일 형태로 저장한다.

이러한 방법으로 수집하여 저장한 파일을 문서분류를 통하여 재난의 종류별로 코로나19, 기상특보, 홍수, 태풍, 산사태, 폭염, 대기오염 등의 분류를 통하여 재난 종류를 구분하게 된다. 이때 미리 설정된 재난 유형에 대해서만 문서를 취득하여 문서 요약(Summarization)을 통해 중요한 내용을 추출하여 요약문을 작성한다.



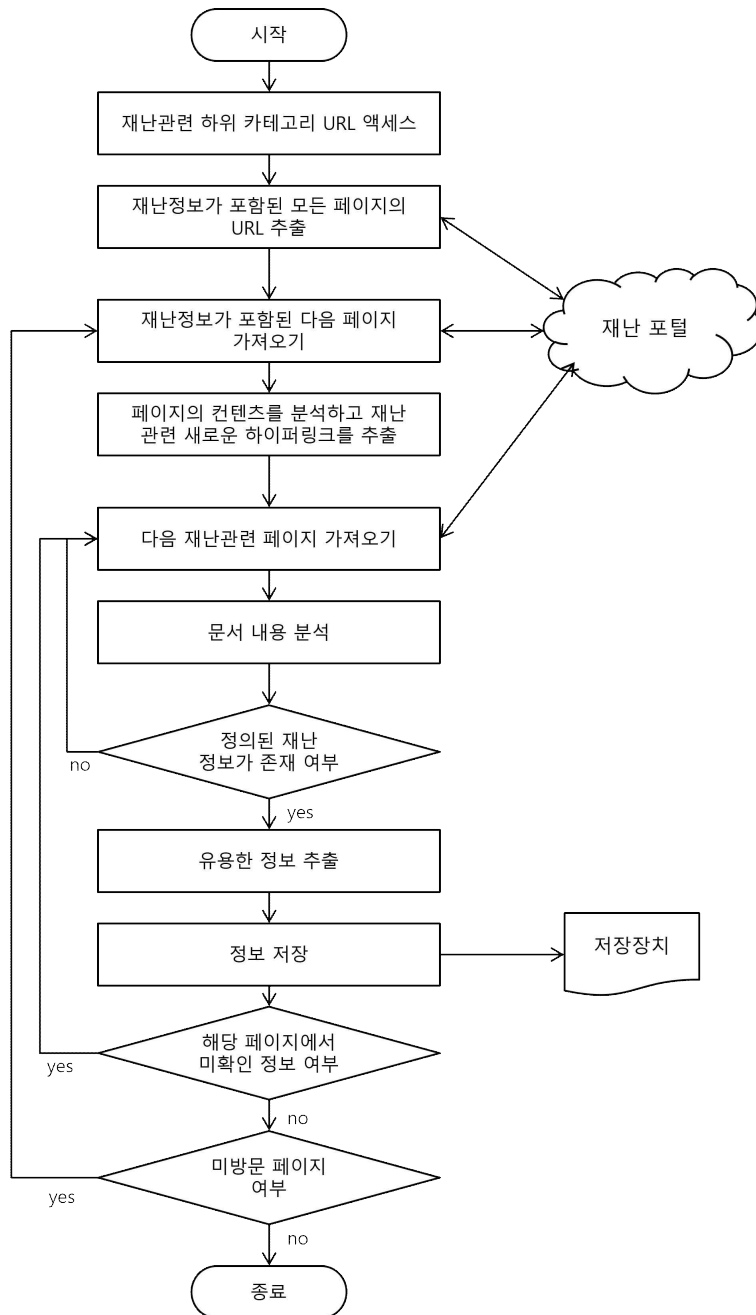
[그림 3-5] 웹 크롤러 구성도

추출된 데이터 중 이를 대비용 안내 문장으로 사용하기 위해서 문서의 내용이 어떠한 카테고리에 속해있는지에 따라서 대상이 결정된다. 앞선 항인 공공데이터 수집과 텍스트 마이닝 결과 다음의 재난에 해당할 경우 문장을 요약하거나 안내하는 절차를 거치게 한다. 재난 포털에서 제공하는 뉴스 중 지진의 경우 중요한 뉴스이지만 실시간 대피가 이뤄져야 하는 재난의 특성상 대비가 불가능하기 때문에 경고 대상에서 제외한다. 또한 재난 뉴스의 경우 재난의 종류로 정의되지 않은 뉴스들도 많이 있다. 이에 대해서는 점진적으로 리스트에 추가하거나 삭제하는 등의 유연한 운영이 필요하다.

[표 3-3] 경고용 재난 상황 정의

구분	자연재난	비고
공공데이터	태풍, 호우, 강풍, 풍랑, 폭풍해일, 지진 해일, 폭염, 대설, 한파, 건조, 황사, 화산, 감염병	자연재난의 황사주의보는 '미세먼지경보'로 대체 (2017년1월13일)
재난 포털	태풍, 호우, 홍수, 산사태, 폭염, 대설, 대기오염, 산불, 전력, 방사선, 감염병	지진은 제외

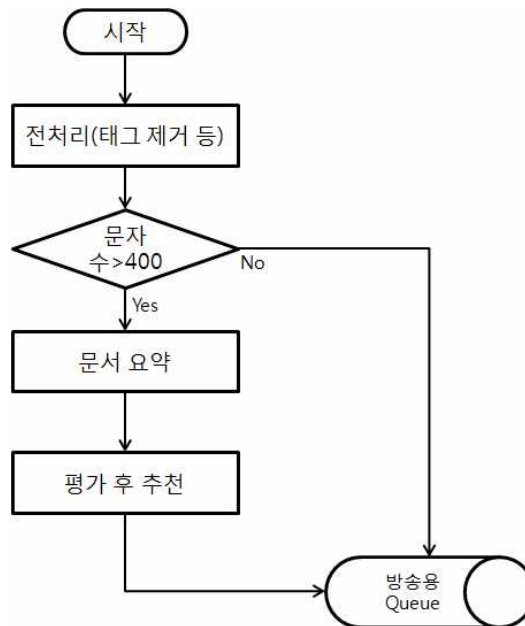
문서의 취득 부분인 웹 스크래퍼를 상세히 명시하면 다음과 같다.



[그림 3-6] 웹 스크래핑 순서도

3) 문서 요약

먼저 재난안내 기사 중 방송을 위해 지정해 둔 재난의 종류에 따라서 해당 문서라고 판단이 되면 문서의 길이를 먼저 확인한다. 그 이유는 웹 크롤링을 통해 획득한 문서의 경우 뉴스의 내용으로 문서의 길이가 너무 길면 방송용으로 적합하지 않을 수 있다. 서울 송파구의 L모 아파트의 안내방송 멘트 중 최근 100여 건의 문자 길이를 확인 결과 평균 총 문자의 길이는 한국어 기준 400자 정도이고, 공백과 개행문자 등을 제외한 길이는 약 320자 정도였다. 이를 TTS를 이용하여 재생할 경우 최종 방송용으로 재생되는 시간이 1분을 약간 넘는 수준이었다. 그래서 웹 크롤링을 통해 획득한 문서에 대하여 [그림 3-7]와 같이 그 길이를 400자 기준으로 두고 그 이하이면 뉴스 내용을 그대로 송출하고 400자를 넘게 되면 문서 요약을 통하여 방송하기 적합한 수준으로 내용을 줄일 필요가 있다.



[그림 3-7] 방송용 콘텐츠 생성

문서 요약이란 문서의 길이가 너무 긴 경우 일정한 분량으로 내용을 요약할 필요가 있는데 이를 위해서 문서의 의미를 분석하여 그 의미를 벗어나지 않은 상태로 문서를 재구성하는 것이다.¹⁰⁶⁾ 본 논문에서 문서 요약의 경우 딥러닝 기법중 하나인 트랜스포머(Transformer) 기반의 요약모델인 BERT와 GPT-3 계열의 한국어 사전학습 언어 모델들, 그리고 그래프 기반 랭킹 모델(TextRank, WordRank)을 이용하여 요약한 결과물 중 가장 품질이 뛰어난 요약문서를 실시간으로 선정하기 위해 우선 5가지 모델을 이용하여 문서를 요약한다.

4) 평가 및 모델 추천

앞서 진행된 연구 프로세스에서 시스템을 이용한 객관적인 평가가 가능한 항목은 문서 요약과 관련된 부분에 대한 평가이다. 이는 트랜스포머 기반 요약모델을 사용하여 생성된 요약문서 및 비지도 학습 모델인 그래프 기반 랭킹 모델에 대한 품질을 평가한다. KoBART, KoGPT 등의 트랜스포머 기반 요약모델은 요약 응답(참조본) 없이 원래 텍스트 입력만 수신하여 요약문을 자동으로 생성한다. 즉, 입력 문서는 벡터로 구조화되고 이미 학습된 컨텍스트 정보를 사용하여 구조화된 문서 벡터에 해당하는 요약문을 생성한다. 그래프 기반 랭킹 모델은 구글의 검색엔진 기반의 링크 순서 결정용으로 개발된 PageRank를 기반 기술로 하는 TextRank 등의 모델을 이용하여 문장을 요약한다.

이렇게 요약된 문장 품질은 인간이 작성한 요약 응답(참조본)과 정량적인 비교를 통해 측정할 수 있다. 평가도구(Metric)를 이용한 평가에 있어서 일반적으로 사용되는 분류로는 MT: 기계번역(Machine Translation), IC: 이미지 캡션(Image Caption), SR: 음성인식(Speech Recognition), SUM: 문서 요약(Summarization), DG: 문서생성(Document Generation), QG: 질문생성(Question Generation), RG: 질의응답 생성(Dialog Response Generation)이

106) Jones, K. Sparck. (1999). Automatic summarizing: factors and directions. Advances in automatic text summarization. 1-12.

있다.

본 연구에서는 평가지표로서 주목받기 시작한 BERTScore나 MoverScore와 같은 기계학습을 거친 평가 기준들 대신, 학습이 필요 없고 현재 가장 널리 사용되는 평가 기준인 ROUGE와 BLEU를 이용하여 평가하였다.¹⁰⁷⁾ ROUGE 점수를 이용하여 평가하는 이유는 [그림 3-8]처럼 문서 요약과 생성에 특화된 평가지표이기 때문이며, BLUE 점수를 참고로 하는 이유는 요약 또한 기계번역과 마찬가지로 문장을 생성하는 방식이기 때문에 ROUGE와 마찬가지로 평가에 활용할 수 있다. 학습하지 않음으로 인해 문장의 정확성이나 유창성 부분에서 부족한 것은 사실이나, 반대로 학습이 불필요하다는 사실이라는 장점은 학습에 필요한 시간을 아낄 수 있다는 점이다. 만약 평가에 필요한 문장의 수가 늘어난다면 훨씬 많은 시간의 차이가 있을 수 있다.¹⁰⁸⁾ 두 가지 지표에 대한 설명은 다음 항에서 자세히 설명하고 있다.

	Metric	Property	MT	IC	SR	SUM	DG	QG	RG
n-gram overlap	BLEU	<i>n</i> -gram precision	✓	✓			✓	✓	✓
	NIST	<i>n</i> -gram precision	✓						
	F-SCORE	precision and recall	✓	✓	✓	✓	✓	✓	✓
	WER	% of insert,delete,replace			✓				
	ROUGE	<i>n</i> -gram recall				✓	✓		
	METEOR	<i>n</i> -gram w/ synonym matching	✓	✓			✓		
	HLEPOR	unigrams harmonic mean	✓						
	RIBES	unigrams harmonic mean							
	CIDER	<i>tf-idf</i> weighted <i>n</i> -gram similarity		✓					
distance-based	EDIT DIST.	cosine similarity	✓	✓	✓	✓	✓	✓	
	TER	translation edit rate	✓						
	WMD	earth mover distance on words		✓		✓			
	SMD	earth mover distance on sentences		✓	✓	✓			
content overlap	PYRAMID				✓				
	SPICE	scene graph similarity		✓					
	SPIDER	scene graph similarity		✓					

[그림 3-8] 문서 생성 모델의 활용 가능 평가 지표

가) ROUGE

ROUGE(Recall-Oriented Understudy for Gisting Evaluation) SCORE란

107) Celikyilmaz, A., Clark, C., Gao, J. (2021). Evaluation of Text Generation: A Survey. arXiv:2006.14799v2

108) 비립. (2021). *한국어 문장 단위 환원 현상 연구* (박사학위, 연세대학교 언어정보협동과정).

텍스트 자동요약, 기계번역 등 자연어 생성모델의 성능을 평가하기 위한 지표로 사람이 미리 만들어 놓은 참조 요약본(reference)과 요약생성 모델이 생성한 요약본(summary)을 비교해서 성능 점수를 계산하는 방법으로 ROUGE-N, ROUGE-L, ROUGE-W, ROUGE-S 등 다양한 지표가 있다.¹⁰⁹⁾ 각각 지표별로 재현율(recall)과 정밀도(precision) 둘 다 구하는 것을 기반으로 하는 F1-Score로 측정이 가능하다. 재현율과 정밀도는 아주 단순한 방법을 통해 구할 수 있다.

ROUGE-N은 n-gram 방식으로 문자열의 선두에서부터 한 문자씩 읽어서 n문자 단위로 일치 여부를 확인하는 방법이다. ROUGE-1은 unigram, ROUGE-2는 bigram, ROUGE-3는 trigram으로 숫자 1, 2, 3은 연속된 단어의 개수를 의미한다.

$$ROUGE-N = \frac{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count(gram_n)} \quad \text{수식(3-1)}$$

ROUGE-L은 Longest Common Subsequence로 두 문장에서 공통으로 나타나는 단어 개수를 순서에 상관없이 측정하는 것으로 생성된 문장의 예시와 정답문장이 완전히 일치하지 않아도 좋은 점수가 나올 수도 있다.

ROUGE-W는 Weighted Longest Common Subsequence로 두 문자에서 공통으로 나타나는 단어 개수에 대해 연속적인 일치 여부를 고려하여 가중치를 주어 값을 측정한다.

ROUGE-S는 Skip-bigram Co-Occurrence Statistics로 skip-gram 방식과 같이, 최대 2칸(bigram) 내에 위치하는 단어 쌍의 재현율을 계산하는 방식으로 skip-gram의 특성상 이어지지 않아도 되므로 상대적으로 거리에 영향을 덜 받는 방식이다.

ROUGE 점수에서 얻을 수 있는 값을 재현율(recall), 정밀도(precision), F1인데 이는 다음과 같은 특성을 가진다. 재현율은 참조 요약본을 구성하는

109) Lin, C. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. *In Text Summarization Branches Out*, p74-81.

단어 중 몇 개가 요약본의 결과와 중복되는가를 확인하는 것으로 우선적으로 필요한 정보들이 다 담겨 있는지를 확인한다. 정밀도는 모델이 생성 또는 추출한 단어 중 몇 개가 참조 요약본과 겹치는가를 확인하는 것으로 요약된 문장에 필요한 정보만을 얼마나 담고 있는지를 확인하는 것이다. F1은 재현율과 정밀도의 조화 평균을 사용하여 모델이 생성 또는 추출한 단어와 요약본 간의 점수를 계산하는 방법이다.¹¹⁰⁾

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad \text{수식(3-2)}$$

나) BLEU

BLEU(Bi-Lingual Evaluation Understudy)¹¹¹⁾는 기계번역을 한 결과와 사람이 번역한 결과와 유사도를 측정하는 성능측정 방법이다. 기계번역을 위한 평가지표로 알려져 있지만 요약에 대한 평가지표로 사용이 가능하다.

BLEU에서 정확도의 기본 개념은 후보 문장(요약본)을 구성하고 있는 단어가 참조 문장(참조본)에 자주 등장할수록 정확도가 높다는 것으로, 이러한 측정 방법을 단어 1개당으로 계산했다는 의미로 유니(uni)를 사용하여 유니그램 정확도(unigram precision)이라고 한다.¹¹²⁾ BLEU는 문장의 단어 순서를 고려하기 위해 유니그램에서 n-gram으로 확장하여 보정 정밀도를 계산한다.¹¹³⁾

n-gram을 통해 생성된 후보(candidate)문장과 참조본(reference) 문장의 일치 여부를 확인하는 방법으로 여기에 같은 단어가 연속으로 나올 때 높은

110) 김의순 & 임희석. (2021). 사전학습 기반의 법률문서 요약방법 비교연구, 『ACK 2021 학술발표대회 논문집』, 28(2). 614-617.

111) Papineni, K., Roukos, S., Ward, T., Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL)*, 311-318.

112) 이영배. (2021). 딥러닝 기반 자동작곡에서 마디 임베딩을 이용한 곡의 생성 및 정량적 평가 방법에 대한 연구 (석사학위, 한성대학교 미래융합컨설팅학과).

113) 이주현. (2021). 딥러닝 기반 대화 문장의 보완대체의사소통 상징 시퀀스 변환 (석사학위, 성신여자대학교).

점수가 나오는 것을 보정(clipping)하고 문장의 길이가 길어질수록 점수가 높아지는 것을 막기 위해 방지(Brevity Penalty)를 적용하였다.

보정이 반영된 Cnt_{clip} 은 생성된 문장 내에서 n-gram이 발생한 횟수 Cnt_{pred} 와 생성된 문장 내에서 n-gram 이 발생한 횟수 중 작은 값을 사용한다.

$$Cnt_{clip}(ngram) = \min(Cnt_{pred}(ngram), Cnt_{ref}(ngram)) \quad \text{수식(3-3)}$$

보정을 고려하여 수정된 n-gram precision은 다음과 같이 계산한다. 여기서 $test$ 는 테스트 세트로 생성한 문장과 원본 문장 쌍의 집합이다.

$$P_n = \frac{\sum_{(pred, ref) \in test} \sum_{ngram \in pred} Cnt_{clip}(ngram)}{\sum_{(pred, ref) \in test} \sum_{ngram \in pred} Cnt_{pred}(ngram)} \quad \text{수식(3-4)}$$

n-gram의 precision P_n 은 문장의 길이가 짧을수록 높은 값을 가지는 특징이 있다. 따라서 이를 보정하는 패널티(Brevity Penalty, BP)를 계산해 BLEU Score에 곱해준다. 여기서 $\leq n_{pred}$ 와 $\leq n_{ref}$ 는 각각 생성한 문장과 참조본의 길이를 의미한다.

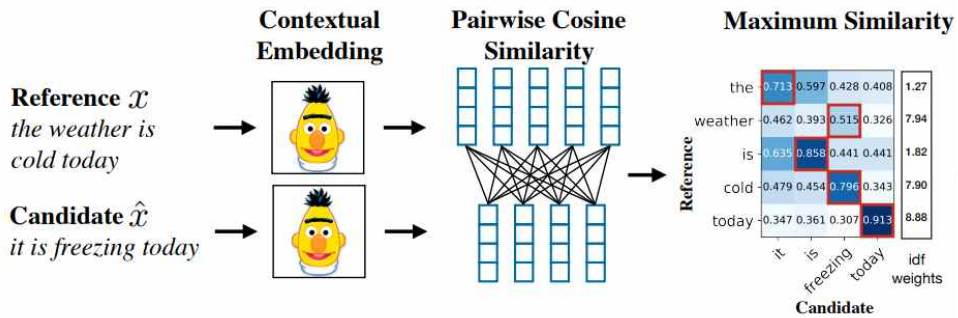
$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases} \quad \text{수식(3-5)}$$

$$BLEU = BP \cdot \exp\left(\sum_{n=0}^N \frac{1}{N} \log P_n\right) \quad \text{수식(3-6)}$$

N은 최대 gram의 수, P_n 은 위에서 정의된 수정된 n-gram precision이다. BLEU 점수는 0~1 사이의 값을 가지며 1에 가까울수록 후보문장이 참조 문장과 유사함을 의미한다. 일반적으로 점수가 0.3~0.4 이상일 때 모델이 양호한 품질을 가지고 있다고 볼 수 있다.

다) BERTScore¹¹⁴⁾ & MoverScore¹¹⁵⁾

BERTScore는 참조 문장(참조본)의 각 토큰과 함께 후보 문장(요약본)의 각 토큰에 대한 유사성 점수를 계산한다. BERTScore는 컨텍스트 임베딩을 사용하여 토큰 유사성을 계산한다. 이 방법을 통해 BERTScore는 의미적으로 유사하지만 표면적으로는 일치하지 않는 문장에 대해 보다 정확한 평가를 수행할 수 있다.



[그림 3-9] BERTScore 절차도

BERTScore의 평가 절차는 [그림 3-9]에 나와 있다. BERTScore는 컨텍스트 임베딩을 사용하여 토큰을 나타내고 코사인 유사성을 사용하여 참조 문장과 후보문장 간의 일치를 계산한다. BERTScore는 매칭 유사성 점수를 최대화하기 위해 그리디 매칭(greedy matching)을 사용하여 각 토큰이 다른 문장에서 가장 유사한 토큰과 매칭된다. 전체 점수는 재현율(recall)을 계산하기 위해 후보의 토큰에 대한 참조의 각 토큰과 정밀도(precision)를 계산하기 위해 참조의 토큰에 후보의 각 토큰을 일치시킨다. 의미를 표현하기 위한 컨텍스트 임베딩 기능으로 인해 BERTScore는 문장의 의미를 비교할 수 있다.

그러나 BERTScore가 번역된 텍스트의 품질을 정확하게 반영한다고 말할

114) Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., Artzi, Y. (2019). BERTScore: Evaluating Text Generation with BERT. *arXiv:1904.09675*.

115) Zhao, W., Peyrard, M., Liu, F., Gao, Y., Meyer, C. M., Eger, S. (2019). MoverScore: Text Generation Evaluating with Contextualized Embeddings and Earth Mover Distance. *arXiv:1909.02622*.

수는 없다. BERTScore는 컨텍스트 임베딩을 사용하기 때문에 결과는 컨텍스트 임베딩의 기능에 따라 달라진다. BERTScore는 BLEU보다 사람의 평가와 상관관계가 높지만 기계번역의 성능 평가는 궁극적으로 사람의 평가에 의해 결정된다.¹¹⁶⁾

MoverScore는 텍스트 생성 작업(예: 기계번역, 텍스트 요약, 이미지 캡션, 질문 응답 등)을 위한 의미(semantic) 기반 평가 메트릭으로, 시스템 및 참조 텍스트는 다자연어-추론에서 미세 조정된 문맥화된 단어 임베딩으로 인코딩된다. 그런 다음 Earth Mover Distance를 활용하여 세트를 비교하여 의미론적 거리를 계산한 값으로 평가한다.

라) 상호 평가

문서 요약에 사용된 언어 요약모델은 한국어에 적합하게 만들어진 모델들로 추출 요약 모델인 TextRank, KrWordRank 그리고 트랜스포머(Transformer) 계열이면서 추출 요약 모델인 RoBERTa와 생성 요약 모델인 KoBART 그리고 KoGPT를 이용한 생성 요약을 사용하였다. 그리고 요약된 문서의 평가를 위해서는 요약 참조본이 반드시 필요한데 이를 매번 사람이 작성할 수 없기 때문에 각 언어 모델에서 생성된 문서들을 참조본(Reference)으로 하여 모델들을 평가하였다. 콘텐츠의 종류에 따라서 요약된 문장의 ROUGE 점수가 높게 나올 때도 있고 상대적으로 낮게 나오는 경우도 있다. 그래서 본 연구에서는 하나의 대상(원본문서)에 대해서 앞서 언급한 모델들로 요약을 한 후 서로 참조본으로 지정하여 점수를 산출하는 방식을 택하였다.

한 가지 예시로 [표 3-4]와 같이 A라는 언어 모델과 B라는 언어 모델을 비교할 경우 먼저 A모델의 경우 B, C, D, E 모델의 결과를 참조본으로 하여 ROUGE 점수와 BLEU 점수를 산출해 낸 후 평균값을 확인한다. B모델의 경우도 마찬가지로 결과를 산출하여 점수가 가장 높게 나온 모델에서 요약한 기사에 대한 요약본을 추천한다. 표에서 B모델의 평가점수가 A모델보다 높게 나왔기 때문에 A모델에서 생성한 결과물을 최종적으로 추천하게 된다.

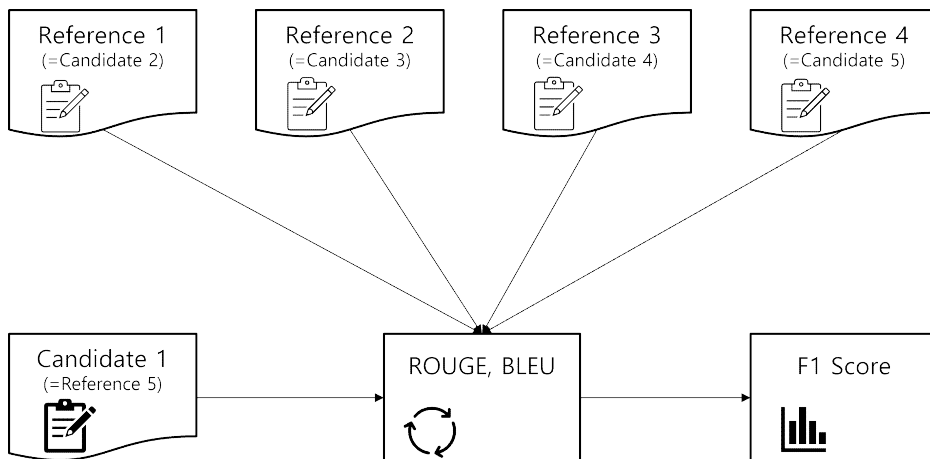
116) 왕규현. (2021). *Paraphrasing training data for neural machine translation*. (석사학위, 고려대학교 대학원).

[표 3-4] 평가 예시

언어 모델	참조본	ROUGE			BLEU
		1	2	L	
A					
	B	0.2	0.1	0.2	0.4
	C	0.3	0.2	0.3	0.4
	D	0.2	0.1	0.2	0.4
	E	0.3	0.2	0.3	0.4
	평균	0.25	0.15	0.25	0.4

언어 모델	참조본	ROUGE			BLUE
		1	2	L	
B	A	0.4	0.3	0.4	0.5
	C	0.3	0.2	0.3	0.4
	D	0.4	0.3	0.4	0.5
	E	0.3	0.2	0.3	0.4
	평균	0.35	0.25	0.35	0.45

[그림 3-10]에서 각 모델들이 평가를 위해 참조하는 방법을 설명하고 있다. A모델에서 생성된 요약문(Candidate 1)에 대한 평가를 위해서 B, C, D, E모델에서 생성된 요약문(Candidate 2~5)을 참조본(Reference 1~4)으로 비교 평가를 진행하였다. 나머지 모델에 대해서도 동일한 방법으로 진행한다.



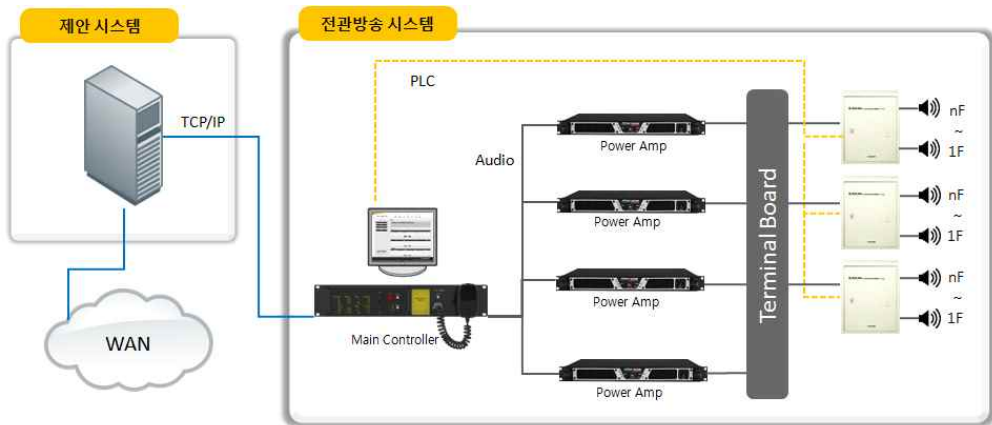
[그림 3-10] 상호 평가 방법

5) 방송 송출

연구시스템에서 생성된 재난안내용 방송 콘텐츠는 최종적으로 전관방송 시스템에 전달되고 전관방송시스템에서는 이를 TTS를 통해 음성으로 전환한 후 방송으로 송출하게 된다.

전관방송 시스템은 화재 발생 등의 비상상황이 발생했을 때 자동 또는 수동으로 비상경보를 확성기를 통해 알려주는 비상방송설비이다. 건축법 시행령에 의해 연면적 3,500㎡ 이상인 건물, 지하층을 제외한 층수가 11층 이상인 곳 또는 지하층의 층수가 3개 층 이상인 곳에 의무적으로 설치해야 한다. 설치 대상으로는 공동주택(아파트, 연립, 다세대), 주상복합건물, 빌딩형 아파트 등 구내전송선로설비를 갖춘 건축물에 설치하고 실내외의 공지사항, 일반방송, 비상통제, 원격방송, 일반안내, 음악방송 등을 필요로 하는 건축물에 설치하게 되어있다.

[그림 3-11]의 전관방송 시스템 부분을 보면 크게 입력부, 제어부, 증폭부, 출력부로 나눌 수 있다.



[그림 3-11] 방송 시스템 연동

가) 입력부

음성이나 음원 등의 아날로그적인 진폭을 전기적인 신호로 변경하는 장치

로서 주장치(Main Controller)의 전단에 위치하며 마이크로폰(Microphone), CD/MP3플레이어, 튜너(Tuner) 등이 있다.

나) 제어부

입력부의 음성이나 특정 음향 등의 아날로그적인 진폭을 전기적인 신호로 가공 및 제어하는 일련의 장치를 총괄하는 부분으로 음량, 음질 등의 음향조절과 시스템연동 등을 담당한다. 또한 입력 신호에 대한 출력을 선택하여 절체해주는 장치로서 구성형태는 매트릭스(Matrix) 형태로 구성할 수 있으며 Web이나 스마트폰 앱을 통하여 설비를 자유롭게 제어할 수 있다.

평상시 배경음악이나 안내방송을 위해 운영을 하지만 소방화재 수신반과 연동하여 화재 발생 시 비상경보 외의 방송을 차단하는 구조로 되어있으며, 화재 정보 수신 후 10초 이내에 일정 수준의 음량으로 화재 발생 및 피난에 대한 정보가 자동으로 송출되어야 하는 조건이 있다.

다) 증폭부

제어부를 거쳐 가동 및 제어된 신호를 이용하여 전송하기 위해 음향조정 및 신호의 증폭을 담당하는 장치가 증폭기(Power Amplifier)이며 증폭기는 출력부와 전기적인 상관관계를 고려하여야 한다. 입력부의 출력 및 제어부의 입출력 신호는 1V의 약한 출력이 나오는데 이를 선로를 통해 출력부로 전달하기 위해서는 큰 출력으로 바꾸어야 한다. 이때 증폭기는 저 잡음 및 고출력 회로가 적용된 제품을 적용하여 신호 대 잡음비(S/N : Signal-to-Noise Ratio)비가 우수하여야 하며, 온도특성에 따라 그 특성이 달라짐에 따라 과입력 또는 과부하를 감지하여 출력을 제한할 수 있는 기능이 반영되어야 한다.¹¹⁷⁾

117) 표준공법-전관방송설비. <http://pubicaddresssystem.blogspot.com/2017/03/2.html>

라) 출력부

입력부-제어부-증폭부를 거친 전기신호를 다시 공기의 진동으로 환원하여 출력하는 최종 단계로서 스피커(Speaker)가 해당 역할을 담당한다. 스피커는 입력부의 마이크로폰과 반대의 기능으로 마이크로폰이 음원의 진동에너지를 전기에너지로 바꿔준다면, 스피커는 증폭기에서 증폭된 전기신호를 사람의 청각을 통해 들을 수 있는 진동에너지로 바꾸어준다. 실내의 경우 실링(Ceiling)스피커, 컬럼(Column)스피커, 월(Wall)스피커가 사용되며, 옥외의 대출력이 필요한 곳에서는 혼(Horn)스피커가 많이 사용된다.

이러한 순서대로 비상방송설비로 운영되는 전관방송시스템에서 연구 결과물인 시스템을 연동하여 동일한 순서로 방송을 할 수 있다. 즉 재난정보를 생성한 경우 인터넷을 통하여 REST API로 전관방송 시스템의 제어부인 메인 컨트롤러에 텍스트데이터를 전달하게 되면, 방송이 가능한지 먼저 확인하게 된다. 만약 다른 방송이 진행 중이라면 우선순위에 의해 대기하거나 기존 방송을 끊고 송출할 수도 있다. 만약, 방송이 가능하다면 전달받은 텍스트를 전관방송 장비의 주장치에 내장된 TTS(Text to Speech)를 통해 음성으로 변환하여 음향 증폭기로 보내게 되고 음향 증폭기에서는 스피커를 통해 청중들에게 음향정보가 전달된다.

제 4 장 실험 결과 및 분석

제 1 절 실험 개요

본 장에서는 논문에서 연구하고 제안한 시스템에서 평가가 가능한 언어 요약과 관련된 성능 위주로 실험 결과를 기록하였다. 비지도 학습 기반의 언어 요약 모델인 TextRank, KrWordRank 등과 기계학습 기반의 언어 요약 모델인 BERT 및 GPT 등 Transformer 계열의 사전학습 모델을 사용하여 요약 성능 검증을 실시하였으며, 앞서 언급한 모델들을 이용하여 생성 요약과 추출 요약을 모두 실시하였다.¹¹⁸⁾

하지만 사전학습 기반 언어 모델들은 공통적으로 학습을 위한 데이터셋이 필수적인데 데이터셋에 대한 품질도 상당히 중요하다. 본 연구에서는 이러한 학습을 위한 부분에 많은 리소스를 할애하기 어려웠기 때문에 사전 학습된 언어 모델들을 이용하되 이에 대한 파인튜닝을 통해 성능을 개선하면서 실험을 진행하였다. 또한 모델들이 한국어에 대한 자연어 처리가 기본적으로 제공되어야 하기 때문에 그에 적합한 모델들을 선정하여 진행하였다.

생성 요약 모델은 카카오브레인에서 제공하는 자연어 태스크에 대응 가능한 통합형태의 자연어 처리 프레임워크인 PORORO(Platform Of neuRAl mOdelS for natuRal language prOcessing)를 사용하였으며 이는 허깅페이스(HuggingFace)와 유사한 플랫폼으로 한국어 태스크에 최적화된 장점이 있다. 라이브러리 내부의 KoBART 모델을 사용하여 생성추출을 하였고, 이 모델의 학습에 사용한 데이터는 데이콘(DACON) 문서추출경진대회 데이터¹¹⁹⁾와 AI HUB에서 공개한 AI 학습용 데이터를 이용하였다. 그리고 추출 요약을 위해서는 페이스북(현 메타플랫폼스)에서 개발한 모델인 RoBERTa 모델을 이용하

118) 남원경, 이지수, 장백철. (2022). 사전 학습된 BERT 기반 모델을 이용한 구어체 텍스트 요약. 『한국컴퓨터정보학회논문지』, 27(8), 41-47.

119) 한국어 문서 추출 요약 AI 경진대회. DACON.
<https://dacon.io/competitions/official/235671/data/>

여 앞서 설명한 말뭉치(corpus)를 훈련시켜 사용하였다. 마지막으로 사용한 KoGPT의 경우 카카오브레인에서 제공하는 GPT-3기반의 한국어 모델이다. 카카오의 라이언(Ryan) 데이터셋을 이용하여 학습했다. 파인튜닝을 할 때 Temperature 파라미터를 조정했는데 그 범위는 0~1 사이의 값을 가지며 수치가 0에 가까울수록 GPT-3가 생성하는 단어가 결정성(deterministic)과 반복성(repetitive)을 보이게 된다. 후보 단어 중에서 출력될 확률이 높은 단어가 선택될 확률이 높아지고 정형화된 문장이 생성된다. 반대로 수치가 1에 가까워지면 많은 후보 단어들의 출력 확률이 비슷해져서 예상하기 힘든 내용이 생성된다. 뉴스임에도 불구하고 심한 경우 전혀 다른 가짜 뉴스를 생성할 수 있는 문제가 있어서 결정성을 염두에 두고 파라미터를 조정하였다.

언어처리 모델들은 선행연구에서 설명했듯이 한국어 처리에서는 결과가 만족스러운 성능을 보여주지 못한다. 이는 학습을 위한 데이터의 양이 충분하지 못했기 때문이다. 구글 BERT의 경우 한국어를 포함한 다국어 모델이 있지만 학습 데이터에서 한국어 비중이 상대적으로 낮아 한국어와 관련된 여러 가지 다운스트림 테스트에서 성능이 낮은 편이다. KoBERT는 한국어 데이터의 정확도를 높이기 위해 한국어 위키피디아에서 약 500만 개의 문장과 5,400만 개의 단어를 학습시켰고, 많은 한국어 다운스트림에서 좋은 성과를 보였다. 또한 SK텔레콤의 경우 사내 업무의 효율을 높이기 위해서 상담 챗봇, 법무 및 특허 등록 지원용 AI 검색 서비스, 내부 마케팅 자료에 정확한 답변을 추출하는 기계 독해 기술에도 사용하였다.¹²⁰⁾

또 다른 모델인 RoBERTa(Robustly Optimized BERT pre-training Approach)는 BERT의 파생모델이며 연구원들은 BERT가 충분히 학습되지 않았음을 확인하고, BERT 모델을 사전 학습시킬 때 사용할 수 있는 몇 가지 방법을 제안하였다. RoBERTa는 기본적으로 BERT와 동일하며, 사전학습 시 하이퍼 파라미터(hyper parameter)를 조정하여 성능을 높이는 방법으로 성능을 개선한 모델이다. 실험은 문서 요약 학습 데이터를 이용한 모델 학습, 학습된 모델을 이용한 평가데이터의 요약생성, 요약된 문장과 평가데이터의 요약내용에 대해서 시스템을 통한 평가를 거쳐서 점수를 기록하였다.

120) 수다르산 라비찬다란. (2021). *구글 BERT의 정석*. (전희원, 정승환, 김형준, 역). 서울: 한빛 미디어(주). (원본 출판 2021).

제 2 절 조건별 성능 비교

다음의 표들은 각 모델이 만들어낸 요약의 결과에 대해 ROUGE 점수를 이용하여 실험 결과의 우열을 판단하는 기준으로 삼았다. ROUGE의 경우 평가를 위해 사람이 직접 참조 요약본(정답요약본: reference)을 따로 준비하지 않고 각각의 모델들에서 생성된 요약본을 참조 요약본으로 사용하였다. 추가적으로 BLEU 점수도 살펴보았는데 참조본을 4개씩 가지면서 1, 2, 3, 4-gram에 대한 가중치는 편의상 모두 동일하게 0.25씩 할당하여 실험을 진행하였다.

일반적으로 언어 모델의 결과를 평가하기 위해서 전문가들이 수동으로 축약을 하고 이에 대해 평가자들이 정보전달력(informativeness) 점수와 가독성(readability) 점수를 평가하고 점수를 평균하여 이용한다. 하지만 이러한 방법에 있어서 평가자 간의 일치도에서 차이가 있었다.¹²¹⁾ 본 연구에서는 실시간으로 생성된 재난 관련 뉴스 기사를 요약해야 하기 때문에 사람이 수동으로 작성한 참조본을 사용할 수가 없었다. 그래서 전문가들 대신 서로 다른 요약 모델에서 생성한 요약본을 요약 참조본으로 사용하여 평가하였으며, 평가 지수는 ROUGE 1, 2, L의 항목 중 재현율과 정밀도의 조화 평균인 F1 점수와 또 다른 평가지표인 BLEU 점수를 평가 점수로 사용하였다. 점수가 1에 가까울수록 참조본과 유사하기 때문에 요약 성능이 우수하다고 평가할 수 있다. 그래서 두 가지 평가지표에 대해서 점수가 가장 높은 모델부터 순차적으로 점수를 5~1점씩 할당하였으며, 두 가지 점수의 합이 동일한 경우에는 순위의 편차가 크지 않은 모델을 추천하게 하였다. 예를 들어 ROUGE 점수 랭킹 2위와 BLEU 점수 랭킹 5위인 모델A의 결과와 ROUGE 점수 3위와 BLEU 점수 2위인 모델B가 있다면 순위를 편차가 작은 모델B의 결과물을 추천한다.

실험에서 사용한 재난상황은 코로나19, 날씨, 가뭄, 홍수, 미세먼지 5가지에 대해서만 실험을 진행하였으며, 연합뉴스 재난포털과 KBS뉴스의 재난.환경 범주로 한정하여 실험하였으며, 상황별 각각 5개씩의 데이터를 사용하여

121) 이경호. (2020). *추출 기반 문서 요약과 문장 축약 기술을 이용한 한국어 문서 요약에 관한 연구* (박사학위, 충남대학교 대학원).

총 25세트의 요약과 100회의 ROUGE 평가, 25회의 BLEU 평가를 진행하였다. 이 실험의 목적이 특정한 뉴스에 대해 특정 요약 모델을 이용해야 한다는 것을 증명하기 위한 실험이 아니기 때문에 횡수는 중요하지 않다. 실험 대상을 다섯 가지 뉴스로 선정한 이유는 2022년 뉴스 기준으로 글자 수 300자 이상의 기사가 나오는 뉴스 중 다섯 가지를 임의로 선정하였다.

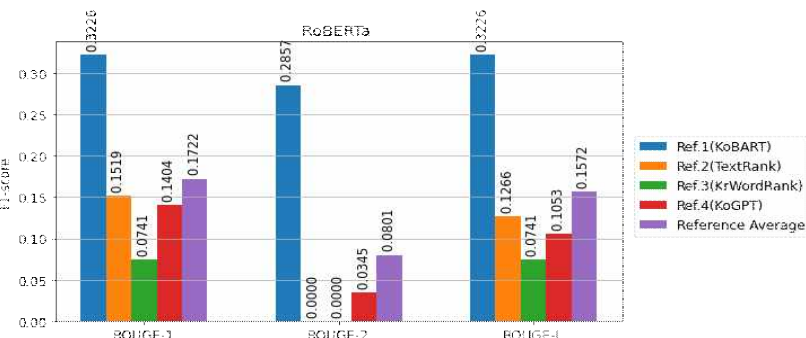
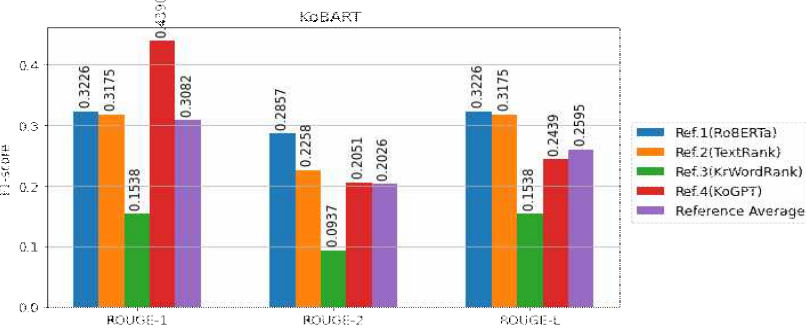
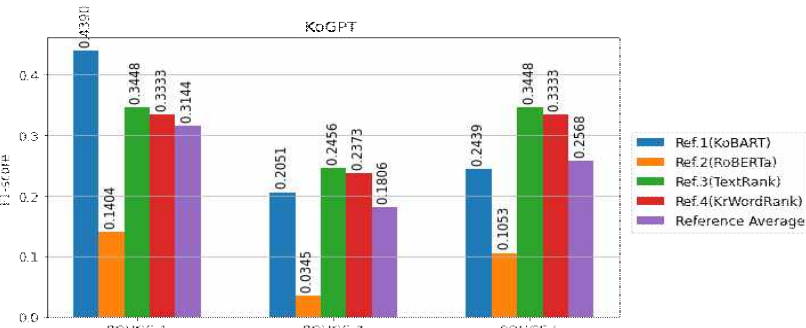
1) 뉴스 분야별 요약 결과 및 평가 결과

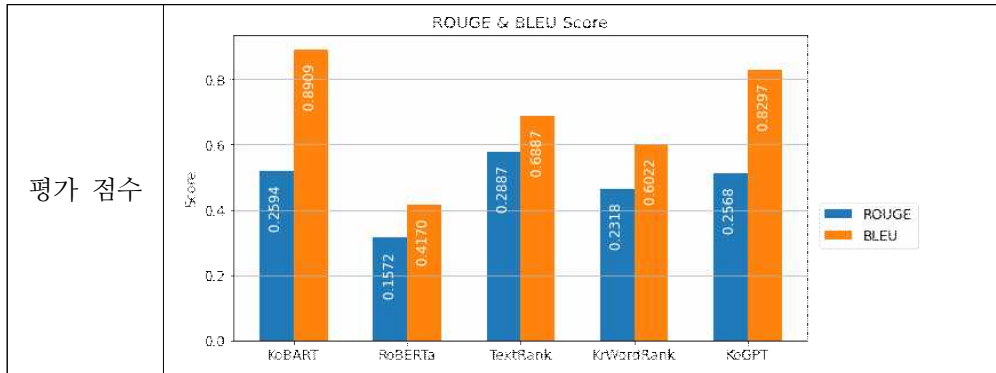
첫 번째 실험으로 다양한 재난의 종류별 뉴스 기사를 요약함에 있어서 특별히 우수한 성능을 내는 모델이 있는지 확인하였으며 아래의 [표 4-1]~[표 4-7]와 같은 결과를 확인할 수 있었다. 주로 정형화된 형태의 글인 코로나19 뉴스나 일기예보의 경우 다른 날짜의 다른 뉴스들로 추가 실험을 하였다.

[표 4-1] 모델별 실험 결과 1 (연합뉴스: 코로나19)

구분	문서
원문	코로나19 7차 유행이 진행 중인 가운데 18일 신규 확진자 수가 4만 명대로 내려왔다. 중앙방역대책본부는 이날 0시 기준 코로나19 확진자가 4만9천418명 늘어 누적 2천646만2천319명이 됐다고 밝혔다. 신규 확진자 수는 전날(5만5천437명)보다 6천19명 적다. 1주일 전인 지난 11일(5만4천519명)보다도 5천101명 줄었는데, 한동안 이어졌던 유행 확산세에서 이례적이다. 1주일 전 대비 확진자수는 지난 주의 경우 1만명 안팎 증가하는 추세였지만, 지난 16일 4만119명, 17일 80명 증가하는 데 그쳤고 이날 감소세로 돌아섰다. 2주일 전인 4일(4만3천449명)과 비교해도 5천969명 늘어 증가폭이 크지 않다. 지난 12일부터 이날까지 1주일간 하루 신규 확진자 수는 5만4천328명→4만8천465명→2만3천765명→7만2천883명→6만6천587명→5만5천437명→4만9천418명으로, 일 평균 5만2천983명이다. 이날 신규 확진자 중 해외유입 사례는 66명으로 전날(39명)보다 27명 많다. 국내 지역감염 사례는 4만9천352명이다. 지역별 확진자 수(해외유입 포함)는 경기 1만4천63명, 서울 9천741명, 인천 2천865명, 경북 2천600명, 부산 2천546명, 경남 2천453명, 대구 2천179명, 충남 2천19명, 강원 1천716명, 전북 1천658명, 충북 1천644명, 대전 1천509명, 광주 1천387명, 전남 1천338명, 울산 873명, 세종 405명, 제주 356명이다. 재원 중인 위중증 환자 수는 전날(380명)보다 15명 줄어든 365명으로, 지난 4일 이후 2주째 300명 이상을 유지하고 있다. 전날 코로나19로 사망한 환자는 63명으로 직전일(67명)보다 4명 줄었

	다. 누적 사망자 수는 2만9천925명이다. 정부는 이번 겨울철 재유행이 본격화하면서 백신 접종률 제고에 힘쓰고 있다. 방역 당국은 지난달 11일 60세 이상 고령자 등 고위험군을 중심으로 동절기 추가접종을 시작한 뒤 지난 7일 18세 이상으로 확대한 바 있다. 이날 0시 기준 동절기 추가접종 백신 접종률은 인구 대비 60세 이상이 12.4%, 전 연령대(18세 이상) 4.3%에 그치고 있다. 접종률은 접종 대상자 대비로는 60세 이상이 16.0%, 전연령대 5.4%다. 정부는 다음달 18일까지 집중 접종기간을 운영하면서 60세 이상 고령층의 50%와 감염 취약시설 거주·이용자 및 종사자의 60% 접종을 목표치로 제시한 바 있다.¹²²⁾																								
요약모델	자동요약 문서																								
TextRank	중앙방역대책본부는 이날 0시 기준 코로나19 확진자가 4만9천418명 늘어 누적 2천646만2천319명이 됐다고 밝혔다 1주일 전인 지난 11일(5만4천519명)보다도 5천101명 줄었는데, 한동안 이어졌던 유행 확산세에서 이례적이다 이날 0시 기준 동절기 추가접종 백신 접종률은 인구 대비 60세 이상이 12.4%, 전 연령대(18세 이상) 4.3%에 그치고 있다 <table><thead><tr><th>Metric</th><th>Ref. 1(KoBART)</th><th>Ref. 2(RoBERTa)</th><th>Ref. 3(KrWordRank)</th><th>Ref. 4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.3175</td><td>0.1519</td><td>0.4146</td><td>0.3448</td><td>0.3072</td></tr><tr><td>ROUGE-2</td><td>0.2258</td><td>0.0000</td><td>0.2683</td><td>0.2456</td><td>0.1849</td></tr><tr><td>ROUGE-L</td><td>0.3175</td><td>0.1266</td><td>0.3659</td><td>0.3448</td><td>0.2887</td></tr></tbody></table>	Metric	Ref. 1(KoBART)	Ref. 2(RoBERTa)	Ref. 3(KrWordRank)	Ref. 4(KoGPT)	Reference Average	ROUGE-1	0.3175	0.1519	0.4146	0.3448	0.3072	ROUGE-2	0.2258	0.0000	0.2683	0.2456	0.1849	ROUGE-L	0.3175	0.1266	0.3659	0.3448	0.2887
Metric	Ref. 1(KoBART)	Ref. 2(RoBERTa)	Ref. 3(KrWordRank)	Ref. 4(KoGPT)	Reference Average																				
ROUGE-1	0.3175	0.1519	0.4146	0.3448	0.3072																				
ROUGE-2	0.2258	0.0000	0.2683	0.2456	0.1849																				
ROUGE-L	0.3175	0.1266	0.3659	0.3448	0.2887																				
KrWordRank	방역 당국은 지난달 11일 60세 이상 고령자 등 고위험군을 중심으로 동절기 추가접종을 시작한 뒤 지난 7일 18세 이상으로 확대한 바 있다. 중앙방역대책본부는 이날 0시 기준 코로나19 확진자가 4만9천418명 늘어 누적 2천646만2천319명이 됐다고 밝혔다. 정부는 이번 겨울철 재유행이 본격화하면서 백신 접종률 제고에 힘쓰고 있다. <table><thead><tr><th>Metric</th><th>Ref. 1(KoBART)</th><th>Ref. 2(RoBERTa)</th><th>Ref. 3(TextRank)</th><th>Ref. 4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.1538</td><td>0.0741</td><td>0.4146</td><td>0.3333</td><td>0.2440</td></tr><tr><td>ROUGE-2</td><td>0.0937</td><td>0.0000</td><td>0.2683</td><td>0.2373</td><td>0.1498</td></tr><tr><td>ROUGE-L</td><td>0.1538</td><td>0.0741</td><td>0.3659</td><td>0.3333</td><td>0.2318</td></tr></tbody></table>	Metric	Ref. 1(KoBART)	Ref. 2(RoBERTa)	Ref. 3(TextRank)	Ref. 4(KoGPT)	Reference Average	ROUGE-1	0.1538	0.0741	0.4146	0.3333	0.2440	ROUGE-2	0.0937	0.0000	0.2683	0.2373	0.1498	ROUGE-L	0.1538	0.0741	0.3659	0.3333	0.2318
Metric	Ref. 1(KoBART)	Ref. 2(RoBERTa)	Ref. 3(TextRank)	Ref. 4(KoGPT)	Reference Average																				
ROUGE-1	0.1538	0.0741	0.4146	0.3333	0.2440																				
ROUGE-2	0.0937	0.0000	0.2683	0.2373	0.1498																				
ROUGE-L	0.1538	0.0741	0.3659	0.3333	0.2318																				

RoBERTa	코로나19 7차 유행이 진행 중인 가운데 18일 신규 확진자 수가 4만 명대로 내려왔다. 1주일 전 대비 확진자수는 지난주의 경우 1만명 안팎 증가하는 추세였지만, 지난 16일 4만119명, 17일 80명 증가하는 데 그쳤고 이날 감소세로 돌아섰다. 이날 신규 확진자 중 해외유입 사례는 66명으로 전날(39명)보다 27명 많다.  <table><caption>RoBERTa F1-scores</caption><thead><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(TextRank)</th><th>Ref.3(KrWordRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.3226</td><td>0.1519</td><td>0.0741</td><td>0.1404</td><td>0.1722</td></tr><tr><td>ROUGE-2</td><td>0.2857</td><td>0.0000</td><td>0.0000</td><td>0.0345</td><td>0.0801</td></tr><tr><td>ROUGE-L</td><td>0.3226</td><td>0.1266</td><td>0.0741</td><td>0.1053</td><td>0.1572</td></tr></tbody></table>	Metric	Ref.1(KoBART)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.3226	0.1519	0.0741	0.1404	0.1722	ROUGE-2	0.2857	0.0000	0.0000	0.0345	0.0801	ROUGE-L	0.3226	0.1266	0.0741	0.1053	0.1572
Metric	Ref.1(KoBART)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.3226	0.1519	0.0741	0.1404	0.1722																				
ROUGE-2	0.2857	0.0000	0.0000	0.0345	0.0801																				
ROUGE-L	0.3226	0.1266	0.0741	0.1053	0.1572																				
KoBART	코로나19 7차 유행이 진행 중인 가운데 18일 신규 확진자 수가 전날보다 6천19명 적은 4만9천418명 늘어 누적 2천646만2천319명이 되었으며 한동안 이어졌던 유행 확산세에서 이례적이다.  <table><caption>KoBART F1-scores</caption><thead><tr><th>Metric</th><th>Ref.1(RoBERTa)</th><th>Ref.2(TextRank)</th><th>Ref.3(KrWordRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.3226</td><td>0.3175</td><td>0.1538</td><td>0.4390</td><td>0.3082</td></tr><tr><td>ROUGE-2</td><td>0.2857</td><td>0.2258</td><td>0.0937</td><td>0.2051</td><td>0.2026</td></tr><tr><td>ROUGE-L</td><td>0.3226</td><td>0.3175</td><td>0.1538</td><td>0.2439</td><td>0.2595</td></tr></tbody></table>	Metric	Ref.1(RoBERTa)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.3226	0.3175	0.1538	0.4390	0.3082	ROUGE-2	0.2857	0.2258	0.0937	0.2051	0.2026	ROUGE-L	0.3226	0.3175	0.1538	0.2439	0.2595
Metric	Ref.1(RoBERTa)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.3226	0.3175	0.1538	0.4390	0.3082																				
ROUGE-2	0.2857	0.2258	0.0937	0.2051	0.2026																				
ROUGE-L	0.3226	0.3175	0.1538	0.2439	0.2595																				
KoGPT	중앙방역대책본부는 18일 0시 기준 코로나19 확진자가 4만9천418명 늘어 누적 2천646만2천319명이 됐으며, 신규 확진자 수는 전날(5만5천437명)보다 6천19명 줄었다고 밝혔다.  <table><caption>KoGPT F1-scores</caption><thead><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(RoBERTa)</th><th>Ref.3(TextRank)</th><th>Ref.4(KrWordRank)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.4390</td><td>0.1404</td><td>0.3448</td><td>0.3333</td><td>0.3144</td></tr><tr><td>ROUGE-2</td><td>0.2051</td><td>0.0345</td><td>0.2456</td><td>0.2373</td><td>0.1806</td></tr><tr><td>ROUGE-L</td><td>0.2439</td><td>0.1053</td><td>0.3448</td><td>0.3333</td><td>0.2568</td></tr></tbody></table>	Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KrWordRank)	Reference Average	ROUGE-1	0.4390	0.1404	0.3448	0.3333	0.3144	ROUGE-2	0.2051	0.0345	0.2456	0.2373	0.1806	ROUGE-L	0.2439	0.1053	0.3448	0.3333	0.2568
Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KrWordRank)	Reference Average																				
ROUGE-1	0.4390	0.1404	0.3448	0.3333	0.3144																				
ROUGE-2	0.2051	0.0345	0.2456	0.2373	0.1806																				
ROUGE-L	0.2439	0.1053	0.3448	0.3333	0.2568																				

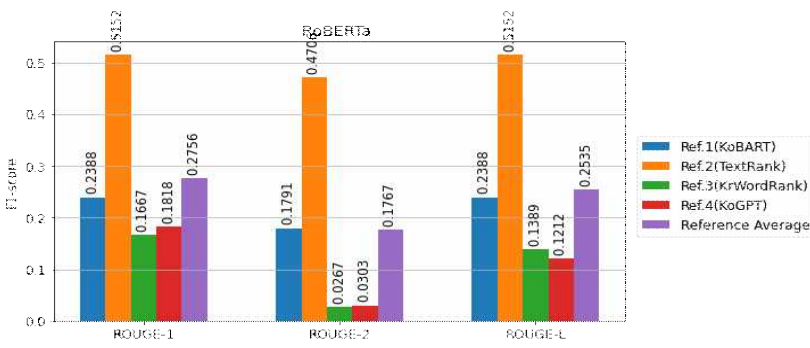
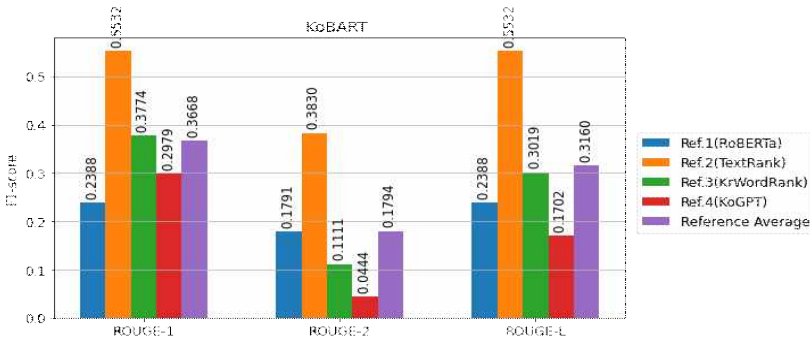
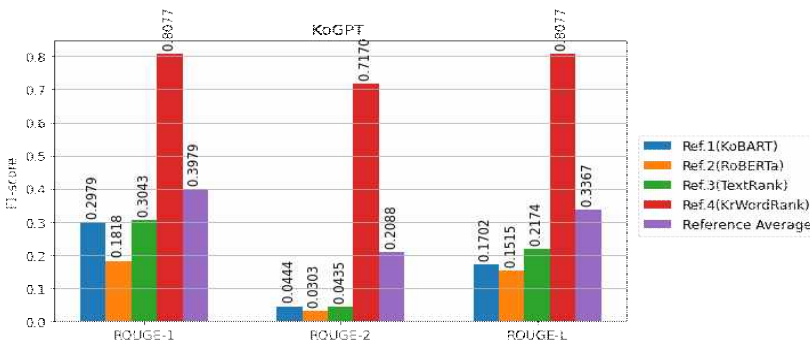


[표 4-2] 모델별 실험 결과 2 (연합뉴스: 코로나19)

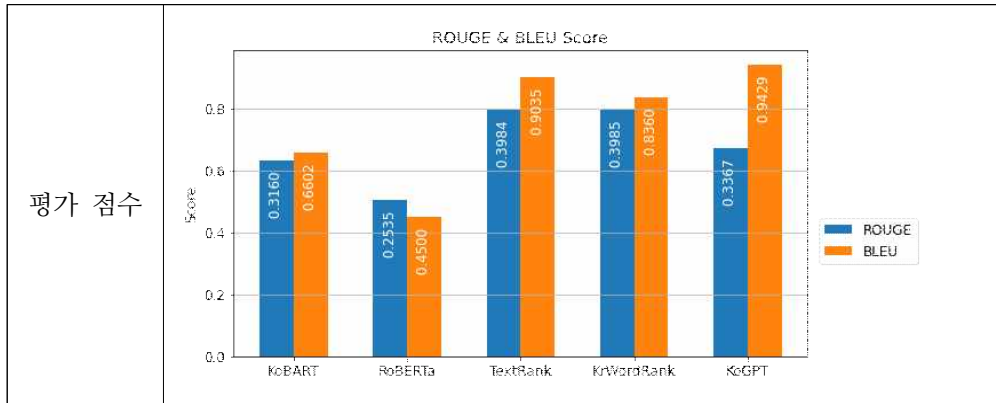
구분	문서
원문	코로나19 신규 확진자가 4만 명대로 집계돼 사흘째 전주보다 줄었다. 증가세가 주춤한 모습이지만, 위중증 환자는 가파르게 늘며 두달새 최다를 기록했다. 코로나19 누적 사망자는 이날로 3만 명을 넘어섰다. 중앙방역대책본부는 20일 0시 기준 코로나19 확진자가 4만6천11명 늘어 누적 2천655만8천765명이 됐다고 밝혔다. 신규 확진자 수는 전날(5만589명)보다는 4천578명, 1주일 전인 13일(4만8천442명)보다 2천431명 각각 줄었다. 1주일 전보다 신규 확진자 수가 줄어든 것은 이날까지 사흘째다. 18일 5천101명, 19일 3천739명 각각 전주 대비 신규 확진자 수가 감소했다. 이날 신규 확진자 수를 2주일 전인 지난 6일(3만6천665명)과 비교하면 9천346명 많다. 지난 14일부터 이날까지 1주일간 하루 신규 확진자 수는 2만3천751명→7만2천866명→6만6천569명→5만5천424명→4만9천418명→5만589명→4만6천11명으로, 일평균 5만2천90명이다. 최근 유행 추세와 관련해 방역 당국은 "(신규 확진자 수가) 정체 및 일시적 소폭 감소 경향을 보이고 있다"며 "하지만 당분간 증가의 방향을 유지하며 증감을 반복할 것으로 전망한다"고 예상했다. 이날 신규 확진자 중 해외유입 사례는 57명으로 전날(50명)보다 7명 많다. 국내 지역감염 사례는 4만5천954명이다. 지역별 확진자 수(해외유입 포함)는 경기 1만2천388명, 서울 8천804명, 인천 2천792명, 부산 2천623명, 경남 2천486명, 경북 2천439명, 충남 1천955명, 대구 1천771명, 전북 1천699명, 충북 1천604명, 광주 1천475명, 대전 1천443명, 강원 1천350명, 전남 1천277명, 울산 1천77명, 세종 432명, 제주 354명, 검역 42명이다. 신규 확진자 수가 전주보다 줄고 있긴 하지만 재원 중 위중증 환자 수나 사망자 수는 높은 수준이다. 위중증 환자 수는 전날(419명)보다 32명 증가한 451명으로, 9월 21일(494명) 이후 두달 사이 가장 많았다. 위중증 환자 수는 지난 14일 이후 1주일간 하루 평균 407명

122) 김병규. (2022년11월18일). 신규확진 4만9418명, 1주전보다 5천여명↓…증가세 주춤(종합). 연합뉴스. <https://www.yna.co.kr/view/AKR20221118043900530?section=coronavirus/news>

	으로, 직전 1주일의 357명보다 50명 많다. 코로나19 누적 사망자 수는 3만31명을 기록하며 3만 명을 넘어섰다. 전날 코로나19로 사망한 환자는 41명으로 직전일(65명)보다 24명 줄었다. 사흘 연속 60명대를 기록하다가 40명대로 내려왔다. 인구 10만명당 사망자수는 10월 4주 0.3명, 11월 1주 0.44명, 11월 2주 0.51명으로 증가 추세가 계속되고 있다. ¹²³⁾																								
요약모델	자동요약 문서																								
TextRank	코로나19 신규 확진자가 4만 명대로 집계돼 사흘째 전주보다 줄었다 코로나19 누적 사망자는 이날로 3만 명을 넘어섰다 1주일 전보다 신규 확진자 수가 줄어든 것은 이날까지 사흘째다 <table><thead><tr><th></th><th>ROUGE-1</th><th>ROUGE-2</th><th>ROUGE-L</th></tr></thead><tbody><tr><td>Ref.1(KoBART)</td><td>0.5532</td><td>0.3830</td><td>0.5532</td></tr><tr><td>Ref.2(RoBERTa)</td><td>0.5152</td><td>0.4706</td><td>0.5152</td></tr><tr><td>Ref.3(KrWordRank)</td><td>0.3846</td><td>0.1818</td><td>0.3077</td></tr><tr><td>Ref.4(KoGPT)</td><td>0.3043</td><td>0.0435</td><td>0.2174</td></tr><tr><td>Reference Average</td><td>0.4393</td><td>0.2697</td><td>0.3984</td></tr></tbody></table>		ROUGE-1	ROUGE-2	ROUGE-L	Ref.1(KoBART)	0.5532	0.3830	0.5532	Ref.2(RoBERTa)	0.5152	0.4706	0.5152	Ref.3(KrWordRank)	0.3846	0.1818	0.3077	Ref.4(KoGPT)	0.3043	0.0435	0.2174	Reference Average	0.4393	0.2697	0.3984
	ROUGE-1	ROUGE-2	ROUGE-L																						
Ref.1(KoBART)	0.5532	0.3830	0.5532																						
Ref.2(RoBERTa)	0.5152	0.4706	0.5152																						
Ref.3(KrWordRank)	0.3846	0.1818	0.3077																						
Ref.4(KoGPT)	0.3043	0.0435	0.2174																						
Reference Average	0.4393	0.2697	0.3984																						
KrWordRank	신규 확진자 수는 전날(5만589명)보다는 4천578명, 1주일 전인 13일(4만8천442명)보다 2천431명 각각 줄었다.코로나19 누적 사망자 수는 3만31명을 기록하며 3만 명을 넘어섰다.중앙방역대책본부는 20일 0시 기준 코로나19 확진자가 4만6천11명 늘어 누적 2천655만8천765명이 됐다고 밝혔다. <table><thead><tr><th></th><th>ROUGE-1</th><th>ROUGE-2</th><th>ROUGE-L</th></tr></thead><tbody><tr><td>Ref.1(KoBART)</td><td>0.3774</td><td>0.1111</td><td>0.3396</td></tr><tr><td>Ref.2(RoBERTa)</td><td>0.1667</td><td>0.0267</td><td>0.1389</td></tr><tr><td>Ref.3(TextRank)</td><td>0.3846</td><td>0.1818</td><td>0.3077</td></tr><tr><td>Ref.4(KoGPT)</td><td>0.8077</td><td>0.7176</td><td>0.8077</td></tr><tr><td>Reference Average</td><td>0.4341</td><td>0.2591</td><td>0.3985</td></tr></tbody></table>		ROUGE-1	ROUGE-2	ROUGE-L	Ref.1(KoBART)	0.3774	0.1111	0.3396	Ref.2(RoBERTa)	0.1667	0.0267	0.1389	Ref.3(TextRank)	0.3846	0.1818	0.3077	Ref.4(KoGPT)	0.8077	0.7176	0.8077	Reference Average	0.4341	0.2591	0.3985
	ROUGE-1	ROUGE-2	ROUGE-L																						
Ref.1(KoBART)	0.3774	0.1111	0.3396																						
Ref.2(RoBERTa)	0.1667	0.0267	0.1389																						
Ref.3(TextRank)	0.3846	0.1818	0.3077																						
Ref.4(KoGPT)	0.8077	0.7176	0.8077																						
Reference Average	0.4341	0.2591	0.3985																						
RoBERTa	코로나19 신규 확진자가 4만 명대로 집계돼 사흘째 전주보다 줄었다. 1주일 전보다 신규 확진자 수가 줄어든 것은 이날까지 사흘째다. 최근 유행 추세와 관련해 방역 당국은 "(신규 확진자 수가) 정체 및 일시적 소폭 감소 경향을 보이고 있다"며 "하지만 당분간 증가의 방향을 유지하며 증감을 반복할 것으로 전망한다"고 예상했다.																								

	 Bar chart showing F1-score for ROUGE-1, ROUGE-2, and ROUGE-L for Ref.1(KoBERT), Ref.2(TextRank), Ref.3(KrWordRank), Ref.4(KoGPT), and Reference Average. The chart is titled 'KoBERT'. <table><tr><th>Metric</th><th>Ref.1(KoBERT)</th><th>Ref.2(TextRank)</th><th>Ref.3(KrWordRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr><tr><td>ROUGE-1</td><td>0.2388</td><td>0.5152</td><td>0.1667</td><td>0.1818</td><td>0.2756</td></tr><tr><td>ROUGE-2</td><td>0.1791</td><td>0.4708</td><td>0.0267</td><td>0.0303</td><td>0.1767</td></tr><tr><td>ROUGE-L</td><td>0.2388</td><td>0.5152</td><td>0.1389</td><td>0.1212</td><td>0.2535</td></tr></table>	Metric	Ref.1(KoBERT)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.2388	0.5152	0.1667	0.1818	0.2756	ROUGE-2	0.1791	0.4708	0.0267	0.0303	0.1767	ROUGE-L	0.2388	0.5152	0.1389	0.1212	0.2535
Metric	Ref.1(KoBERT)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.2388	0.5152	0.1667	0.1818	0.2756																				
ROUGE-2	0.1791	0.4708	0.0267	0.0303	0.1767																				
ROUGE-L	0.2388	0.5152	0.1389	0.1212	0.2535																				
KoBERT	코로나19 신규 확진자가 0시 기준 4만 명대로 집계돼 사흘째 전주보다 줄었으며 위중증 환자는 가파르게 늘어 두달 새 최다를 기록하였으나 누적 사망자는 3만 명을 넘어섰다.  Bar chart showing F1-score for ROUGE-1, ROUGE-2, and ROUGE-L for Ref.1(RoBERTa), Ref.2(TextRank), Ref.3(KrWordRank), Ref.4(KoGPT), and Reference Average. The chart is titled 'KoBERT'. <table><tr><th>Metric</th><th>Ref.1(RoBERTa)</th><th>Ref.2(TextRank)</th><th>Ref.3(KrWordRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr><tr><td>ROUGE-1</td><td>0.2388</td><td>0.5532</td><td>0.3774</td><td>0.2979</td><td>0.3668</td></tr><tr><td>ROUGE-2</td><td>0.1791</td><td>0.3830</td><td>0.1111</td><td>0.0444</td><td>0.1794</td></tr><tr><td>ROUGE-L</td><td>0.2388</td><td>0.5532</td><td>0.3019</td><td>0.1702</td><td>0.3160</td></tr></table>	Metric	Ref.1(RoBERTa)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.2388	0.5532	0.3774	0.2979	0.3668	ROUGE-2	0.1791	0.3830	0.1111	0.0444	0.1794	ROUGE-L	0.2388	0.5532	0.3019	0.1702	0.3160
Metric	Ref.1(RoBERTa)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.2388	0.5532	0.3774	0.2979	0.3668																				
ROUGE-2	0.1791	0.3830	0.1111	0.0444	0.1794																				
ROUGE-L	0.2388	0.5532	0.3019	0.1702	0.3160																				
KoGPT	중앙방역대책본부는 20일 0시 기준 코로나19 확진자가 4만6천11명 늘어 누적 2천655만8천765명이 되었다고 밝히며, 신규 확진자 수는 전날(5만589명)보다는 4천578명, 1주일 전인 13일(4만8천442명)보다 2천431명 각각 줄었다.  Bar chart showing F1-score for ROUGE-1, ROUGE-2, and ROUGE-L for Ref.1(KoBERT), Ref.2(RoBERTa), Ref.3(TextRank), Ref.4(KrWordRank), and Reference Average. The chart is titled 'KoGPT'. <table><tr><th>Metric</th><th>Ref.1(KoBERT)</th><th>Ref.2(RoBERTa)</th><th>Ref.3(TextRank)</th><th>Ref.4(KrWordRank)</th><th>Reference Average</th></tr><tr><td>ROUGE-1</td><td>0.2979</td><td>0.1818</td><td>0.3043</td><td>0.8077</td><td>0.3979</td></tr><tr><td>ROUGE-2</td><td>0.0444</td><td>0.0303</td><td>0.0435</td><td>0.7170</td><td>0.2088</td></tr><tr><td>ROUGE-L</td><td>0.1702</td><td>0.1515</td><td>0.2174</td><td>0.8077</td><td>0.3367</td></tr></table>	Metric	Ref.1(KoBERT)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KrWordRank)	Reference Average	ROUGE-1	0.2979	0.1818	0.3043	0.8077	0.3979	ROUGE-2	0.0444	0.0303	0.0435	0.7170	0.2088	ROUGE-L	0.1702	0.1515	0.2174	0.8077	0.3367
Metric	Ref.1(KoBERT)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KrWordRank)	Reference Average																				
ROUGE-1	0.2979	0.1818	0.3043	0.8077	0.3979																				
ROUGE-2	0.0444	0.0303	0.0435	0.7170	0.2088																				
ROUGE-L	0.1702	0.1515	0.2174	0.8077	0.3367																				

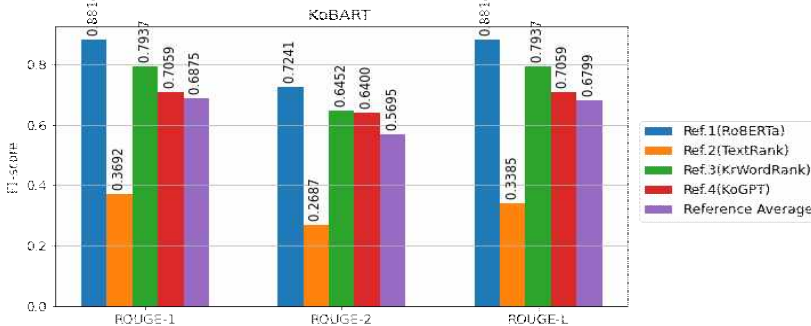
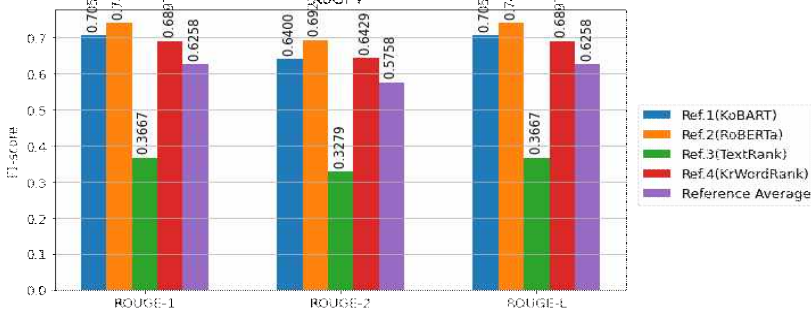
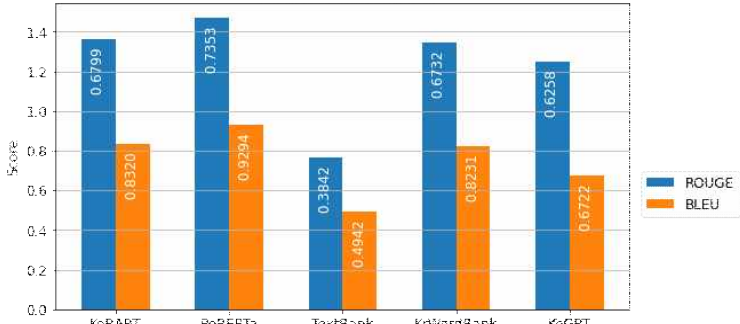
123) 김병규. (2022년11월20일). 4만6011명 확진, 3일째 전주보다 줄어...당국 "일시적 감소"(종합2보). 연합뉴스. <https://www.yna.co.kr/view/AKR20221120015851530?section=coronavirus/news>



[표 4-3] 모델별 실험 결과 3 (연합뉴스: 코로나19)

구분	문서
원문	경기도는 19일 0시 기준으로 집계된 18일 하루 도내 코로나19 확진자가 1만4천345명이라고 밝혔다. 전날인 17일 1만4천69명보다 276명 늘었지만, 한 주 전 같은 금요일이었던 지난 11일 1만5천283명보다는 938명 줄었다. 이날까지 누적 확진자는 715만8천272명이다. 사망자의 경우 9명이 발생해 누적 사망자가 7천506명이 됐다. 시·군별 확진자를 보면 수원시가 1천385명으로 가장 많았고 이어 용인시(1천209명), 고양시(1천158명), 화성시(1천67명), 성남시(964명), 부천시(816명) 등의 순이었다. 가평군(42명)·연천군(63명)·동두천시(68명)·과천시(72명)·여주시(74명) 등 5개 시·군은 100명 미만으로 나왔다. 도내 코로나19 전담 병상 가동률은 39.4%로 전날(40.6%)보다 1.2%포인트 낮아졌고, 이 가운데 중증 환자 병상 가동률도 30.3%로 전날(33.6%)보다 3.3%포인트 내려갔다. 재택치료 중인 확진자는 9만147명으로 사흘째 9만명 초반대를 유지했다.¹²⁴⁾
요약모델	자동요약 문서
TextRank	경기도는 19일 0시 기준으로 집계된 18일 하루 도내 코로나19 확진자가 1만4천345명이라고 밝혔다 이날까지 누적 확진자는 715만8천272명이다 도내 코로나19 전담 병상 가동률은 39.4%로 전날(40.6%)보다 1.2%포인트 낮아졌고, 이 가운데 중증 환자 병상 가동률도 30.3%로 전날(33.6%)보다 3.3%포인트 내려갔다

	<table><caption>TextRank F1-score Data</caption><thead><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(RoBERTa)</th><th>Ref.3(KrWordRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.3692</td><td>0.4706</td><td>0.3611</td><td>0.3667</td><td>0.3919</td></tr><tr><td>ROUGE-2</td><td>0.2687</td><td>0.4058</td><td>0.3014</td><td>0.3279</td><td>0.3260</td></tr><tr><td>ROUGE-L</td><td>0.3385</td><td>0.4706</td><td>0.3611</td><td>0.3667</td><td>0.3842</td></tr></tbody></table>	Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.3692	0.4706	0.3611	0.3667	0.3919	ROUGE-2	0.2687	0.4058	0.3014	0.3279	0.3260	ROUGE-L	0.3385	0.4706	0.3611	0.3667	0.3842
Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.3692	0.4706	0.3611	0.3667	0.3919																				
ROUGE-2	0.2687	0.4058	0.3014	0.3279	0.3260																				
ROUGE-L	0.3385	0.4706	0.3611	0.3667	0.3842																				
KrWordRank	경기도는 19일 0시 기준으로 집계된 18일 하루 도내 코로나19 확진자가 1만4천345명이라고 밝혔다.전날인 17일 1만4천69명보다 276명 늘었지만, 한 주 전 같은 금요일이었던 지난 11일 1만5천283명보다는 938명 줄었다.사망자의 경우 9명이 발생해 누적 사망자가 7천506명이 됐다. <table><caption>KrWordRank F1-score Data</caption><thead><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(RoBERTa)</th><th>Ref.3(TextRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.7957</td><td>0.8435</td><td>0.3611</td><td>0.6897</td><td>0.6732</td></tr><tr><td>ROUGE-2</td><td>0.6452</td><td>0.8111</td><td>0.3014</td><td>0.6429</td><td>0.6005</td></tr><tr><td>ROUGE-L</td><td>0.7957</td><td>0.8435</td><td>0.3611</td><td>0.6897</td><td>0.6732</td></tr></tbody></table>	Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.7957	0.8435	0.3611	0.6897	0.6732	ROUGE-2	0.6452	0.8111	0.3014	0.6429	0.6005	ROUGE-L	0.7957	0.8435	0.3611	0.6897	0.6732
Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.7957	0.8435	0.3611	0.6897	0.6732																				
ROUGE-2	0.6452	0.8111	0.3014	0.6429	0.6005																				
ROUGE-L	0.7957	0.8435	0.3611	0.6897	0.6732																				
RoBERTa	경기도는 19일 0시 기준으로 집계된 18일 하루 도내 코로나19 확진자가 1만4천345명이라고 밝혔다. 전날인 17일 1만4천69명보다 276명 늘었지만, 한 주 전 같은 금요일이었던 지난 11일 1만5천283명보다는 938명 줄었다. 이날까지 누적 확진자는 715만8천272명이다. <table><caption>RoBERTa F1-score Data</caption><thead><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(TextRank)</th><th>Ref.3(KrWordRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.8814</td><td>0.4706</td><td>0.8435</td><td>0.7407</td><td>0.7353</td></tr><tr><td>ROUGE-2</td><td>0.7241</td><td>0.4058</td><td>0.8111</td><td>0.6923</td><td>0.6587</td></tr><tr><td>ROUGE-L</td><td>0.8814</td><td>0.4706</td><td>0.8435</td><td>0.7407</td><td>0.7353</td></tr></tbody></table>	Metric	Ref.1(KoBART)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.8814	0.4706	0.8435	0.7407	0.7353	ROUGE-2	0.7241	0.4058	0.8111	0.6923	0.6587	ROUGE-L	0.8814	0.4706	0.8435	0.7407	0.7353
Metric	Ref.1(KoBART)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.8814	0.4706	0.8435	0.7407	0.7353																				
ROUGE-2	0.7241	0.4058	0.8111	0.6923	0.6587																				
ROUGE-L	0.8814	0.4706	0.8435	0.7407	0.7353																				

KoBART	경기도는 19일 0시 기준으로 집계된 18일 하루 도내 코로나19 확진자는 전날인 17일 1만4천69명보다 276명 늘었지만, 누적 확진자는 715만8천272명으로 한 주 전 같은 금요일이었던 지난 11일 1만5천283명보다는 938명 줄었다고 밝혔다.  <table><caption>KoBART F1 Scores</caption><thead><tr><th>Metric</th><th>Ref.1(RoBERTa)</th><th>Ref.2(TextRank)</th><th>Ref.3(KrWordRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.8814</td><td>0.3692</td><td>0.7937</td><td>0.7059</td><td>0.6875</td></tr><tr><td>ROUGE-2</td><td>0.7241</td><td>0.2687</td><td>0.6452</td><td>0.6400</td><td>0.5695</td></tr><tr><td>ROUGE-L</td><td>0.8814</td><td>0.3385</td><td>0.7937</td><td>0.7059</td><td>0.6799</td></tr></tbody></table>	Metric	Ref.1(RoBERTa)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.8814	0.3692	0.7937	0.7059	0.6875	ROUGE-2	0.7241	0.2687	0.6452	0.6400	0.5695	ROUGE-L	0.8814	0.3385	0.7937	0.7059	0.6799
Metric	Ref.1(RoBERTa)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.8814	0.3692	0.7937	0.7059	0.6875																				
ROUGE-2	0.7241	0.2687	0.6452	0.6400	0.5695																				
ROUGE-L	0.8814	0.3385	0.7937	0.7059	0.6799																				
KoGPT	경기도는 19일 0시 기준으로 집계된 18일 하루 도내 코로나19 확진자가 1만4천345명이라고 밝혔으며, 한 주 전 같은 금요일이었던 지난 11일 1만5천283명보다는 938명 감소한 수치이다.  <table><caption>KoGPT F1 Scores</caption><thead><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(RoBERTa)</th><th>Ref.3(TextRank)</th><th>Ref.4(KrWordRank)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.7059</td><td>0.7417</td><td>0.3667</td><td>0.6887</td><td>0.6258</td></tr><tr><td>ROUGE-2</td><td>0.6400</td><td>0.6937</td><td>0.3279</td><td>0.6429</td><td>0.5758</td></tr><tr><td>ROUGE-L</td><td>0.7059</td><td>0.7407</td><td>0.3667</td><td>0.6887</td><td>0.6258</td></tr></tbody></table>	Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KrWordRank)	Reference Average	ROUGE-1	0.7059	0.7417	0.3667	0.6887	0.6258	ROUGE-2	0.6400	0.6937	0.3279	0.6429	0.5758	ROUGE-L	0.7059	0.7407	0.3667	0.6887	0.6258
Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KrWordRank)	Reference Average																				
ROUGE-1	0.7059	0.7417	0.3667	0.6887	0.6258																				
ROUGE-2	0.6400	0.6937	0.3279	0.6429	0.5758																				
ROUGE-L	0.7059	0.7407	0.3667	0.6887	0.6258																				
평가 점수	 <table><caption>ROUGE & BLEU Scores</caption><thead><tr><th>Model</th><th>ROUGE</th><th>BLEU</th></tr></thead><tbody><tr><td>KoBART</td><td>0.6799</td><td>0.8320</td></tr><tr><td>RoBERTa</td><td>0.7353</td><td>0.9294</td></tr><tr><td>TextRank</td><td>0.3842</td><td>0.4942</td></tr><tr><td>KrWordRank</td><td>0.6732</td><td>0.8231</td></tr><tr><td>KoGPT</td><td>0.6258</td><td>0.6722</td></tr></tbody></table>	Model	ROUGE	BLEU	KoBART	0.6799	0.8320	RoBERTa	0.7353	0.9294	TextRank	0.3842	0.4942	KrWordRank	0.6732	0.8231	KoGPT	0.6258	0.6722						
Model	ROUGE	BLEU																							
KoBART	0.6799	0.8320																							
RoBERTa	0.7353	0.9294																							
TextRank	0.3842	0.4942																							
KrWordRank	0.6732	0.8231																							
KoGPT	0.6258	0.6722																							

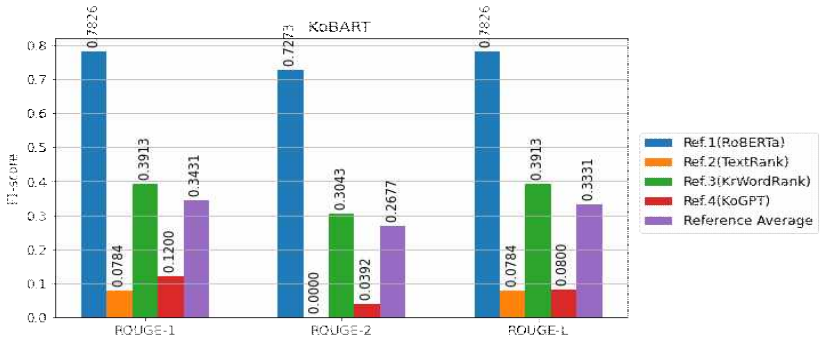
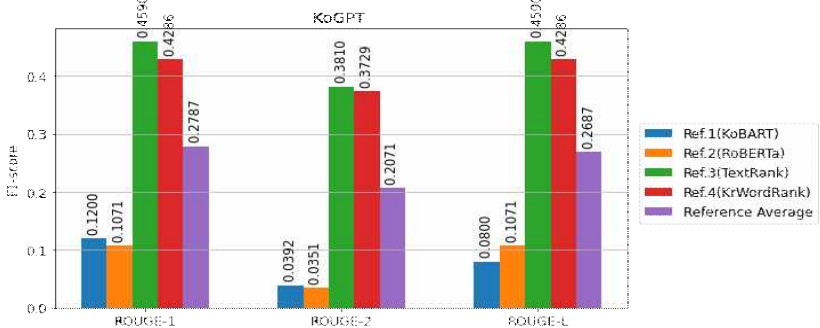
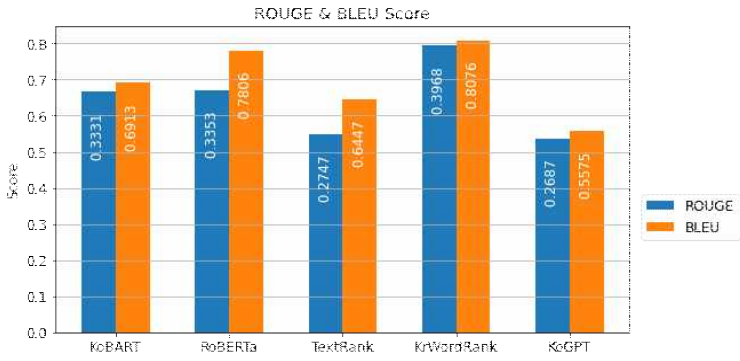
124) 최찬홍. (2022년11월19일). 경기 1만4345명 신규 확진...1주 전보다 938명 감소. 연합뉴스. <https://www.yna.co.kr/view/AKR20221119020300061?section=coronavirus/news>

[표 4-1]~[표 4-3]은 코로나 19와 관련된 매일의 상황에 대한 뉴스를 요약한 것이다. 앞서 설명한 기준에 의해 높은 점수로 추천된 모델의 결과물과 가장 낮은 점수를 보이는 모델이 동일한 유형의 뉴스에서도 계속 달라짐을 확인할 수 있다. 첫 번째 시험에서는 KoBART가 ROUGE 및 BLUE 점수에서 높은 결과로 추천이 되었고, RoBERTa가 가장 나쁜 결과가 나왔었다. 하지만 두 번째 시험에서는 KoBART가 RoBERTa와 함께 좋지 않은 결과가 나왔으며, 나머지가 비슷하게 나왔다. 세 번째 시험에서는 계속해서 양호한 성적을 보였던 TextRank가 좋지 않은 결과를 보이고 기존에 좋지 않은 결과를 보인 RoBERTa가 높은 점수로 추천되었다.

[표 4-4] 모델별 실험 결과 4 (KBS: 날씨)

구분	문서
원문	오늘은 전국에 가끔 구름이 많이 끼겠습니다. 밤에는 수도권과 충청, 광주 등 서쪽 지역의 초미세먼지 농도가 '나쁨' 수준을 보이겠습니다. 강원 산지에는 오후에 강한 바람이 불겠습니다. 대학수학능력시험일인 내일은 전국이 맑은 가운데 큰 추위는 없겠습니다. 새벽부터 아침 사이에는 내륙을 중심으로 짙은 안개가 끼겠습니다. 아침 기온은 서울이 6도 등 전국이 0도에서 9도로 오늘과 비슷하거나 조금 낮겠습니다. 다만, 경기 동부와 강원 내륙·산지, 경북 북부 내륙은 아침 기온이 0도보다 낮은 곳이 있겠습니다. 낮 최고 기온은 서울이 16도, 광주는 19도 등 전국이 14도에서 19도로 오늘보다 1도에서 3도 정도 높겠습니다. 내륙 지역은 일교차가 15도 안팎까지 벌어지겠습니다. 인천과 경기 남부, 충청지역은 내일 초미세먼지 농도가 '나쁨' 수준까지 높아지겠습니다. 내일 바다의 물결은 제주도 남쪽 바깥 먼바다에서 2미터 안팎으로 비교적 높게 일겠습니다.¹²⁵⁾
요약모델	자동요약 문서
TextRank	오늘은 전국에 가끔 구름이 많이 끼겠습니다 아침 기온은 서울이 6도 등 전국이 0도에서 9도로 오늘과 비슷하거나 조금 낮겠습니다 다만, 경기 동부와 강원 내륙·산지, 경북 북부 내륙은 아침 기온이 0도보다 낮은 곳이 있겠습니다

	<table><caption>TextRank F1-score Data</caption><thead><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(RoBERTa)</th><th>Ref.3(KrWordRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.0784</td><td>0.1053</td><td>0.4561</td><td>0.4590</td><td>0.2747</td></tr><tr><td>ROUGE-2</td><td>0.0000</td><td>0.0000</td><td>0.3793</td><td>0.3810</td><td>0.1901</td></tr><tr><td>ROUGE-L</td><td>0.0784</td><td>0.1053</td><td>0.4561</td><td>0.4590</td><td>0.2747</td></tr></tbody></table>	Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.0784	0.1053	0.4561	0.4590	0.2747	ROUGE-2	0.0000	0.0000	0.3793	0.3810	0.1901	ROUGE-L	0.0784	0.1053	0.4561	0.4590	0.2747
Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.0784	0.1053	0.4561	0.4590	0.2747																				
ROUGE-2	0.0000	0.0000	0.3793	0.3810	0.1901																				
ROUGE-L	0.0784	0.1053	0.4561	0.4590	0.2747																				
KrWordRank	아침 기온은 서울이 6도 등 전국이 0도에서 9도로 오늘과 비슷하거나 조금 낮겠습니다.새벽부터 아침 사이에는 내륙을 중심으로 길은 안개가 끼겠습니다.대학수학능력시험일인 내일은 전국이 맑은 가운데 큰 추위는 없겠습니다. <table><caption>KrWordRank F1-score Data</caption><thead><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(RoBERTa)</th><th>Ref.3(TextRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.3913</td><td>0.3462</td><td>0.4561</td><td>0.4286</td><td>0.4055</td></tr><tr><td>ROUGE-2</td><td>0.3043</td><td>0.2692</td><td>0.3793</td><td>0.3729</td><td>0.3314</td></tr><tr><td>ROUGE-L</td><td>0.3913</td><td>0.3462</td><td>0.4211</td><td>0.4286</td><td>0.3968</td></tr></tbody></table>	Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.3913	0.3462	0.4561	0.4286	0.4055	ROUGE-2	0.3043	0.2692	0.3793	0.3729	0.3314	ROUGE-L	0.3913	0.3462	0.4211	0.4286	0.3968
Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.3913	0.3462	0.4561	0.4286	0.4055																				
ROUGE-2	0.3043	0.2692	0.3793	0.3729	0.3314																				
ROUGE-L	0.3913	0.3462	0.4211	0.4286	0.3968																				
RoBERTa	밤에는 수도권과 충청, 광주 등 서쪽 지역의 초미세먼지 농도가 '나쁨' 수준을 보이겠습니다. 강원 산지에는 오후에 강한 바람이 불겠습니다. 대학수학능력시험일인 내일은 전국이 맑은 가운데 큰 추위는 없겠습니다. <table><caption>RoBERTa F1-score Data</caption><thead><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(TextRank)</th><th>Ref.3(KrWordRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.7826</td><td>0.1053</td><td>0.3462</td><td>0.1071</td><td>0.3353</td></tr><tr><td>ROUGE-2</td><td>0.7273</td><td>0.0000</td><td>0.2692</td><td>0.0351</td><td>0.2579</td></tr><tr><td>ROUGE-L</td><td>0.7826</td><td>0.1053</td><td>0.3462</td><td>0.1071</td><td>0.3353</td></tr></tbody></table>	Metric	Ref.1(KoBART)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.7826	0.1053	0.3462	0.1071	0.3353	ROUGE-2	0.7273	0.0000	0.2692	0.0351	0.2579	ROUGE-L	0.7826	0.1053	0.3462	0.1071	0.3353
Metric	Ref.1(KoBART)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.7826	0.1053	0.3462	0.1071	0.3353																				
ROUGE-2	0.7273	0.0000	0.2692	0.0351	0.2579																				
ROUGE-L	0.7826	0.1053	0.3462	0.1071	0.3353																				

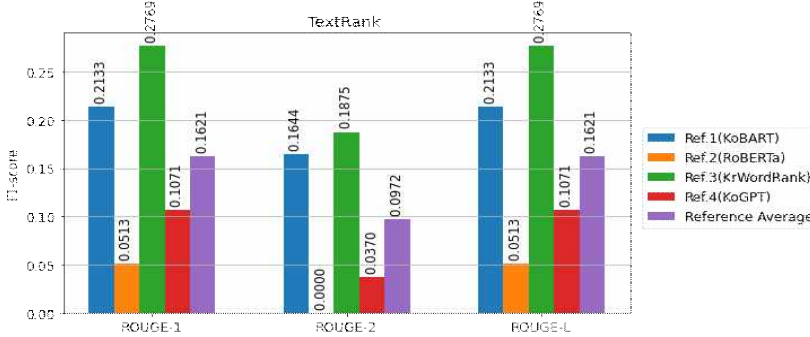
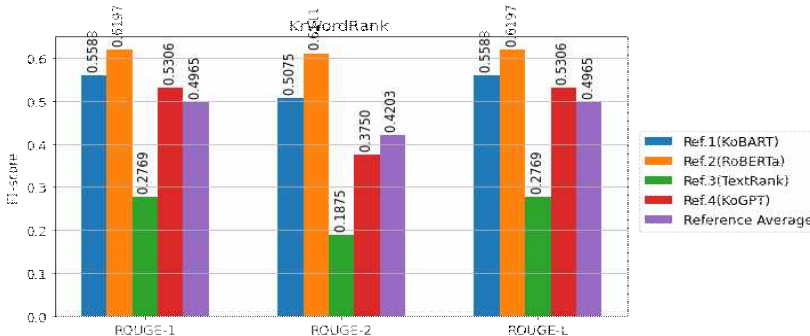
KoBART	대학에는 수도권과 충청, 광주 등 서쪽 지역의 초미세먼지 농도가 '나쁨' 수준을 보이고, 대학수학능력시험일인 내일은 전국이 맑은 가운데 큰 추위는 없겠습니다.  <table><caption>KoBART F1-score Data</caption><thead><tr><th>Metric</th><th>Ref. 1 (RoBERTa)</th><th>Ref. 2 (TextRank)</th><th>Ref. 3 (KrWordRank)</th><th>Ref. 4 (KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.7826</td><td>0.0784</td><td>0.3913</td><td>0.1200</td><td>0.3431</td></tr><tr><td>ROUGE-2</td><td>0.7273</td><td>0.0000</td><td>0.3043</td><td>0.0392</td><td>0.2677</td></tr><tr><td>ROUGE-L</td><td>0.7826</td><td>0.0784</td><td>0.3913</td><td>0.0800</td><td>0.3331</td></tr></tbody></table>	Metric	Ref. 1 (RoBERTa)	Ref. 2 (TextRank)	Ref. 3 (KrWordRank)	Ref. 4 (KoGPT)	Reference Average	ROUGE-1	0.7826	0.0784	0.3913	0.1200	0.3431	ROUGE-2	0.7273	0.0000	0.3043	0.0392	0.2677	ROUGE-L	0.7826	0.0784	0.3913	0.0800	0.3331
Metric	Ref. 1 (RoBERTa)	Ref. 2 (TextRank)	Ref. 3 (KrWordRank)	Ref. 4 (KoGPT)	Reference Average																				
ROUGE-1	0.7826	0.0784	0.3913	0.1200	0.3431																				
ROUGE-2	0.7273	0.0000	0.3043	0.0392	0.2677																				
ROUGE-L	0.7826	0.0784	0.3913	0.0800	0.3331																				
KoGPT	주말인 내일은 전국이 가끔 구름이 많이 끼겠고, 아침 기온은 서울이 6도 등 전국이 0도에서 9도로 오늘과 비슷하거나 조금 낮겠고, 낮 최고 기온은 서울이 16도, 광주는 19도 등 전국이 14도에서 19도로 오늘보다 1도에서 3도 정도 높겠다.  <table><caption>KoGPT F1-score Data</caption><thead><tr><th>Metric</th><th>Ref. 1 (KoBART)</th><th>Ref. 2 (RoBERTa)</th><th>Ref. 3 (TextRank)</th><th>Ref. 4 (KrWordRank)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.1200</td><td>0.1071</td><td>0.4590</td><td>0.4286</td><td>0.2787</td></tr><tr><td>ROUGE-2</td><td>0.0392</td><td>0.0351</td><td>0.3810</td><td>0.3729</td><td>0.2071</td></tr><tr><td>ROUGE-L</td><td>0.0800</td><td>0.1071</td><td>0.4590</td><td>0.4286</td><td>0.2687</td></tr></tbody></table>	Metric	Ref. 1 (KoBART)	Ref. 2 (RoBERTa)	Ref. 3 (TextRank)	Ref. 4 (KrWordRank)	Reference Average	ROUGE-1	0.1200	0.1071	0.4590	0.4286	0.2787	ROUGE-2	0.0392	0.0351	0.3810	0.3729	0.2071	ROUGE-L	0.0800	0.1071	0.4590	0.4286	0.2687
Metric	Ref. 1 (KoBART)	Ref. 2 (RoBERTa)	Ref. 3 (TextRank)	Ref. 4 (KrWordRank)	Reference Average																				
ROUGE-1	0.1200	0.1071	0.4590	0.4286	0.2787																				
ROUGE-2	0.0392	0.0351	0.3810	0.3729	0.2071																				
ROUGE-L	0.0800	0.1071	0.4590	0.4286	0.2687																				
평가 점수	 <table><caption>ROUGE & BLEU Score Data</caption><thead><tr><th>Model</th><th>ROUGE</th><th>BLEU</th></tr></thead><tbody><tr><td>KoBART</td><td>0.3331</td><td>0.6913</td></tr><tr><td>RoBERTa</td><td>0.3353</td><td>0.7806</td></tr><tr><td>TextRank</td><td>0.2747</td><td>0.6447</td></tr><tr><td>KrWordRank</td><td>0.3968</td><td>0.8076</td></tr><tr><td>KoGPT</td><td>0.2687</td><td>0.5575</td></tr></tbody></table>	Model	ROUGE	BLEU	KoBART	0.3331	0.6913	RoBERTa	0.3353	0.7806	TextRank	0.2747	0.6447	KrWordRank	0.3968	0.8076	KoGPT	0.2687	0.5575						
Model	ROUGE	BLEU																							
KoBART	0.3331	0.6913																							
RoBERTa	0.3353	0.7806																							
TextRank	0.2747	0.6447																							
KrWordRank	0.3968	0.8076																							
KoGPT	0.2687	0.5575																							

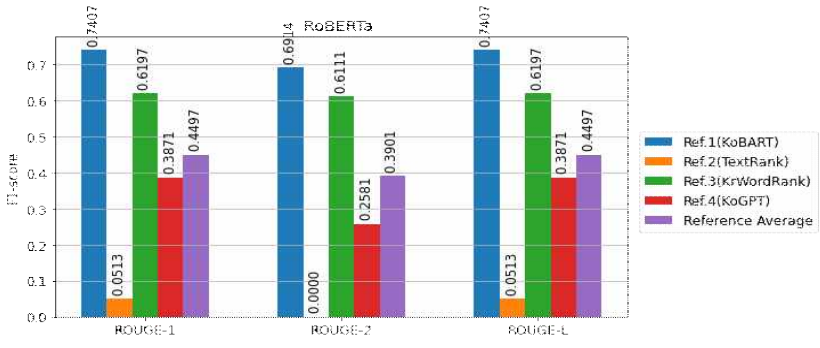
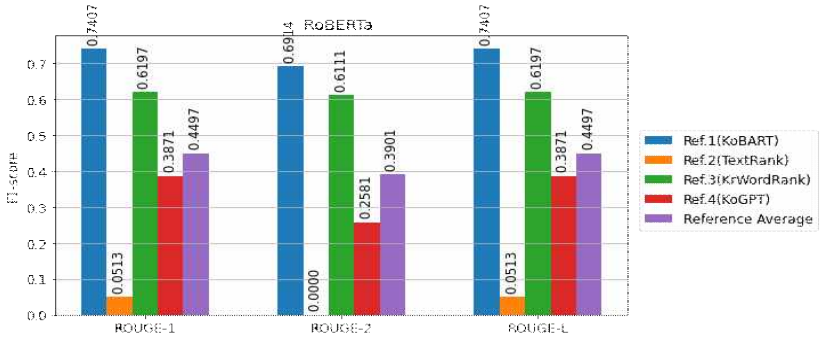
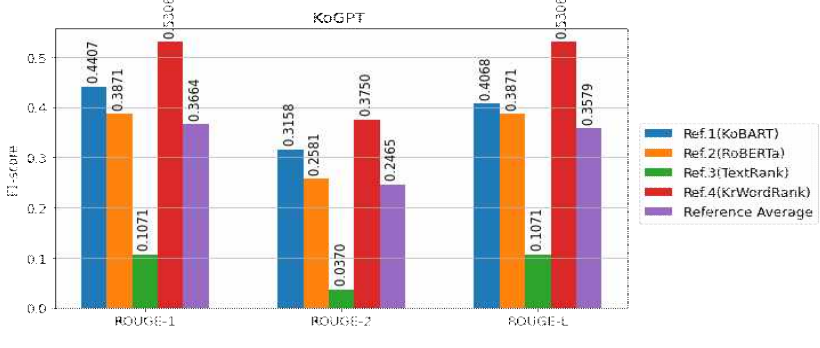
[표 4-4]는 날씨와 관련된 뉴스를 다양한 언어 모델을 이용하여 요약하였다. 여기서는 첫 번째 예제를 이용한 실험에서 KrWordRank를 이용한 요약 결과물에서 가장 좋은 결과가 나왔다. 두 번째 실험에서는 KoGPT에서 생성 요약한 결과물이 가장 높은 점수를 나타내어서 추천되었으며, 세 번째 실험에서는 앞선 실험에서 추천된 KoGPT가 가장 점수가 좋지 못했고 나머지 모델을 통한 요약 문서에서는 유사한 점수 분포를 보였다. 이처럼 특정한 뉴스에 대해서 어떠한 모델이 우수하다고 평가하기는 어려웠다. 추가로 약간 다른 유형의 재난 관련 뉴스들을 요약해 보았다. [표 4-5]~[표 4-7]는 가뭄, 홍수, 미세먼지와 관련된 뉴스를 요약하면서 어떤 추세를 보이는지 확인해 보았다.

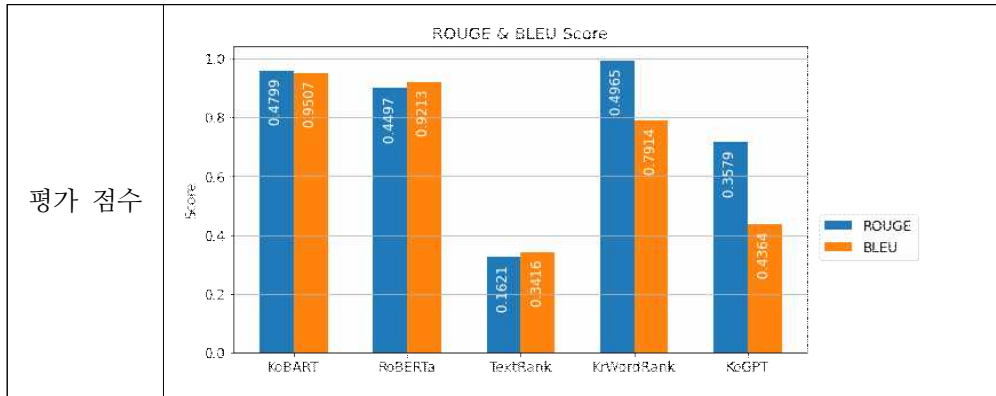
[표 4-5] 모델별 실험 결과 5 (연합뉴스: 가뭄)

구분	문서
원문	전남 완도 노화·소안·금일지역 수원지 저수율이 10%에도 미치지 못하면서 지역민들이 최악의 가뭄을 겪고 있다. 생활용수는 물론 먹을 물도 바닥이 나면서 단수 지역이 확대되고 있다. 10일 완도군에 따르면 노화 넓도의 경우 수원지 저수율(6%)이 증가하지 않아 5월부터 현재까지 1일 급수, 6일 단수를 시행하고 있다. 현재 2천300명의 주민이 거주하는 소안면은 미라제 저수율이 8%에 불과해 1일부터 2일 급수, 5일 단수를 하고 있다. 인구 3천650명이 사는 금일읍도 사정이 심각해, 지난 7일부터 2일 급수 4일 단수를 시행했다. 갈수기에 접어들면서 계속해서 비가 내리지 않을 시 제한 급수 지역은 늘어날 것으로 보인다. 완도군은 이달부터 금일, 노화, 고금, 약산, 소안 보길 등을 대상으로 9월 사용량 대비 10% 절감을 목표로 절수 운동을 대대적으로 펼치고 있다. 12월부터는 제한 급수 지역에 병물을 공급할 계획이다. 가뭄 장기화에 대비한 단계별 대응에도 나섰다. 1단계는 이달 말까지로 급수선과 대형 관정을 통한 하천수 펌핑으로 수원지 용수를 보충한다. 마을별 소형 관정, 하천수, 간이 상수도를 정비하고 물탱크를 설치해 생활용수를 공급한다. 2단계는 이달 중순 시작한다. 금일, 소안, 노화, 고금의 수원지 준설을 시작하고 마을 샘물 급수선 설치, 마을 단위 중형 관정 개발 사업을 추진한다. 환경부 시범 사업인 해수 담수화 플랜트를 통해 소안 미라제에는 하루 300t을 공급한다. 3단계는 내년 1월부터 시행할 예정으로 대형 철부선과 차량을 이용해 광역 상수도를 공급할 계획이다. 군 관계자는 10일 "제한 급수로 많은 불편함을 겪고 있는 도주민들을 위해 장·단기 계획을 세워 급수 공급 가능 일을 최대한 늘릴 수 있도록 최선을 다하겠다"고 밝혔다.¹²⁵⁾

125) 김세현. (2022년11월16일). 수능일 ‘큰 일교차’ 주의…내일도 수도권·충청 미세먼지 ‘나쁨’. KBS뉴스. <https://news.kbs.co.kr/news/view.do?ncd=5602787>

요약모델	자동요약 문서																								
TextRank	10일 완도군에 따르면 노화 읍도의 경우 수원지 저수율(6%)이 증가하지 않아 5월부터 현재까지 1일 급수, 6일 단수를 시행하고 있다 12월부터는 제한 급수 지역에 병물을 공급할 계획이다 마을별 소형 관정, 하천수, 간이 상수도를 정비하고 물탱크를 설치해 생활용수를 공급한다  <table><caption>TextRank F1-score Data</caption><thead><tr><th>Reference</th><th>ROUGE-1</th><th>ROUGE-2</th><th>ROUGE-L</th></tr></thead><tbody><tr><td>Ref. 1(KoBART)</td><td>0.2133</td><td>0.1644</td><td>0.2133</td></tr><tr><td>Ref. 2(RoBERTa)</td><td>0.0513</td><td>0.0000</td><td>0.0513</td></tr><tr><td>Ref. 3(KrWordRank)</td><td>0.2769</td><td>0.1875</td><td>0.2769</td></tr><tr><td>Ref. 4(KoGPT)</td><td>0.1071</td><td>0.0370</td><td>0.1071</td></tr><tr><td>Reference Average</td><td>0.1621</td><td>0.0972</td><td>0.1621</td></tr></tbody></table>	Reference	ROUGE-1	ROUGE-2	ROUGE-L	Ref. 1(KoBART)	0.2133	0.1644	0.2133	Ref. 2(RoBERTa)	0.0513	0.0000	0.0513	Ref. 3(KrWordRank)	0.2769	0.1875	0.2769	Ref. 4(KoGPT)	0.1071	0.0370	0.1071	Reference Average	0.1621	0.0972	0.1621
Reference	ROUGE-1	ROUGE-2	ROUGE-L																						
Ref. 1(KoBART)	0.2133	0.1644	0.2133																						
Ref. 2(RoBERTa)	0.0513	0.0000	0.0513																						
Ref. 3(KrWordRank)	0.2769	0.1875	0.2769																						
Ref. 4(KoGPT)	0.1071	0.0370	0.1071																						
Reference Average	0.1621	0.0972	0.1621																						
KrWordRank	12월부터는 제한 급수 지역에 병물을 공급할 계획이다.전남 완도 노화·소안·금일지역 수원지 저수율이 10%에도 미치지 못하면서 지역민들이 최악의 가뭄을 겪고 있다.생활용수는 물론 먹을 물도 바닥이 나면서 단수 지역이 확대되고 있다.  <table><caption>KrWordRank F1-score Data</caption><thead><tr><th>Reference</th><th>ROUGE-1</th><th>ROUGE-2</th><th>ROUGE-L</th></tr></thead><tbody><tr><td>Ref. 1(KoBART)</td><td>0.5588</td><td>0.5075</td><td>0.5588</td></tr><tr><td>Ref. 2(RoBERTa)</td><td>0.6197</td><td>0.6197</td><td>0.6197</td></tr><tr><td>Ref. 3(TextRank)</td><td>0.2769</td><td>0.1875</td><td>0.2769</td></tr><tr><td>Ref. 4(KoGPT)</td><td>0.5306</td><td>0.3750</td><td>0.5306</td></tr><tr><td>Reference Average</td><td>0.4965</td><td>0.4203</td><td>0.4965</td></tr></tbody></table>	Reference	ROUGE-1	ROUGE-2	ROUGE-L	Ref. 1(KoBART)	0.5588	0.5075	0.5588	Ref. 2(RoBERTa)	0.6197	0.6197	0.6197	Ref. 3(TextRank)	0.2769	0.1875	0.2769	Ref. 4(KoGPT)	0.5306	0.3750	0.5306	Reference Average	0.4965	0.4203	0.4965
Reference	ROUGE-1	ROUGE-2	ROUGE-L																						
Ref. 1(KoBART)	0.5588	0.5075	0.5588																						
Ref. 2(RoBERTa)	0.6197	0.6197	0.6197																						
Ref. 3(TextRank)	0.2769	0.1875	0.2769																						
Ref. 4(KoGPT)	0.5306	0.3750	0.5306																						
Reference Average	0.4965	0.4203	0.4965																						
RoBERTa	전남 완도 노화·소안·금일지역 수원지 저수율이 10%에도 미치지 못하면서 지역민들이 최악의 가뭄을 겪고 있다. 생활용수는 물론 먹을 물도 바닥이 나면서 단수 지역이 확대되고 있다. 완도군은 이달부터 금일, 노화, 고금, 약산, 소안 보길 등을 대상으로 9월 사용량 대비 10% 절감을 목표로 절수 운동을 대대적으로 펼치고 있다.																								

	 RoBERTa <table><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(TextRank)</th><th>Ref.3(KrWordRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr><tr><td>ROUGE-1</td><td>0.7407</td><td>0.0513</td><td>0.6197</td><td>0.3871</td><td>0.4497</td></tr><tr><td>ROUGE-2</td><td>0.6914</td><td>0.0000</td><td>0.6111</td><td>0.2581</td><td>0.3901</td></tr><tr><td>ROUGE-L</td><td>0.7407</td><td>0.0513</td><td>0.6197</td><td>0.3871</td><td>0.4497</td></tr></table>	Metric	Ref.1(KoBART)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.7407	0.0513	0.6197	0.3871	0.4497	ROUGE-2	0.6914	0.0000	0.6111	0.2581	0.3901	ROUGE-L	0.7407	0.0513	0.6197	0.3871	0.4497
Metric	Ref.1(KoBART)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.7407	0.0513	0.6197	0.3871	0.4497																				
ROUGE-2	0.6914	0.0000	0.6111	0.2581	0.3901																				
ROUGE-L	0.7407	0.0513	0.6197	0.3871	0.4497																				
KoBART	전남 완도 노화·소안·금일지역 수원지 저수율이 10%에도 미치지 못하면서 지역민들이 최악의 가뭄을 겪고 있는 가운데 금일, 노화, 고금, 약산, 소안 보길 등을 대상으로 9월 사용량 대비 10% 절감을 목표로 절수 운동을 대대적으로 펼치고 12월부터는 제한 급수 지역에 병물을 공급할 계획이다.  RoBERTa <table><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(TextRank)</th><th>Ref.3(KrWordRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr><tr><td>ROUGE-1</td><td>0.7407</td><td>0.0513</td><td>0.6197</td><td>0.3871</td><td>0.4497</td></tr><tr><td>ROUGE-2</td><td>0.6914</td><td>0.0000</td><td>0.6111</td><td>0.2581</td><td>0.3901</td></tr><tr><td>ROUGE-L</td><td>0.7407</td><td>0.0513</td><td>0.6197</td><td>0.3871</td><td>0.4497</td></tr></table>	Metric	Ref.1(KoBART)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.7407	0.0513	0.6197	0.3871	0.4497	ROUGE-2	0.6914	0.0000	0.6111	0.2581	0.3901	ROUGE-L	0.7407	0.0513	0.6197	0.3871	0.4497
Metric	Ref.1(KoBART)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.7407	0.0513	0.6197	0.3871	0.4497																				
ROUGE-2	0.6914	0.0000	0.6111	0.2581	0.3901																				
ROUGE-L	0.7407	0.0513	0.6197	0.3871	0.4497																				
KoGPT	완도 노화, 소안, 금일의 수원지 저수율이 10%에도 미치지 못하면서 지역민들이 최악의 가뭄을 겪고 있어 제한 급수 지역이 늘어날 것으로 보인다.  KoGPT <table><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(RoBERTa)</th><th>Ref.3(TextRank)</th><th>Ref.4(KrWordRank)</th><th>Reference Average</th></tr><tr><td>ROUGE-1</td><td>0.4407</td><td>0.3871</td><td>0.1071</td><td>0.5308</td><td>0.3664</td></tr><tr><td>ROUGE-2</td><td>0.3158</td><td>0.2581</td><td>0.0370</td><td>0.3750</td><td>0.2465</td></tr><tr><td>ROUGE-L</td><td>0.4068</td><td>0.3871</td><td>0.1071</td><td>0.5308</td><td>0.3579</td></tr></table>	Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KrWordRank)	Reference Average	ROUGE-1	0.4407	0.3871	0.1071	0.5308	0.3664	ROUGE-2	0.3158	0.2581	0.0370	0.3750	0.2465	ROUGE-L	0.4068	0.3871	0.1071	0.5308	0.3579
Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KrWordRank)	Reference Average																				
ROUGE-1	0.4407	0.3871	0.1071	0.5308	0.3664																				
ROUGE-2	0.3158	0.2581	0.0370	0.3750	0.2465																				
ROUGE-L	0.4068	0.3871	0.1071	0.5308	0.3579																				

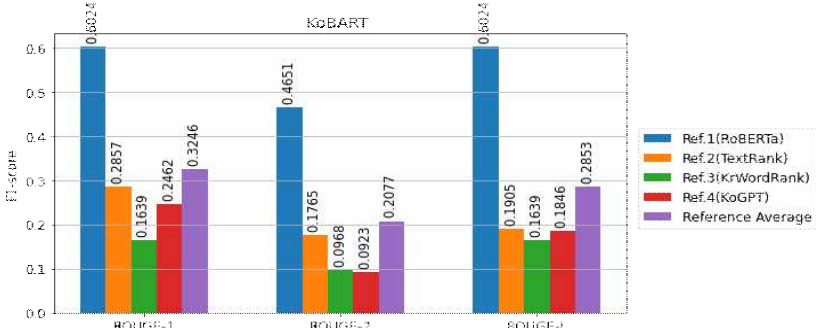
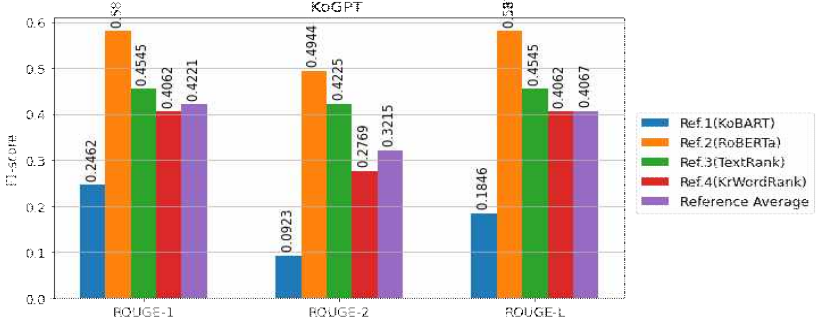
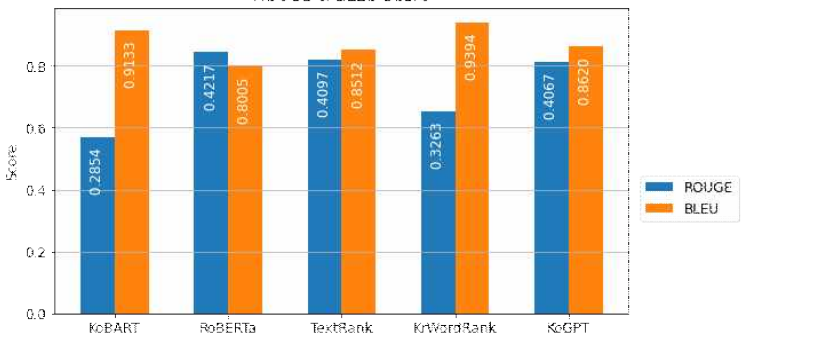


[표 4-6] 모델별 실험 결과 6 (KBS: 홍수)

구분	문서
원문	형산강 유역에 많은 비가 내리면서 형산강의 경주시 강동대교 지점에 홍수주의보가 내려졌습니다. 낙동강홍수통제소는 6일 5시 10분에 형산강 유역 경주시 강동대교(경상북도 경주시 강동면 오금리 1046-1 (강동대교)) 지점에 홍수주의보를 발령한다고 밝혔습니다. 낙동강홍수통제소는 경주에서 하천 수위 상승과 범람에 따른 홍수 피해가 우려된다며 주민들에게 하천변에 접근하지 말고 안전한 곳에 머무는 등 홍수 피해가 없도록 유의해 달라고 당부했습니다. 경주시 강동대교 지점의 수위는 5시 현재 3.75m(수위표 기준)를 기록한 뒤 계속 상승 중이며, 100분 뒤인 6시 40분에는 홍수주의보 기준인 6.55m까지 상승할 것으로 예측됐습니다. 홍수주의보는 최대 홍수량의 50%까지 수위가 높아질 때 내려지며, 천변 둔치와 자전거도로 등 시설물이 침수될 수 있습니다. 경주시 강동대교 지점의 홍수주의보 수위는 6.50m, 홍수경보 수위는 7.50m입니다. 또, 하천 제방이 버틸 수 있는 한계를 의미하는 계획홍수위는 9.73m입니다. 현재 한강 홍천군(홍천교) 지점, 한탄강 포천시(영평교) 지점, 낙동강 산청군(경호교) 지점, 형산강 경주시(강동대교) 지점, 포항시(형산교) 지점 등 5개 지점에 홍수주의보가 발효 중입니다.¹²⁶⁾
요약모델	자동요약 문서
TextRank	형산강 유역에 많은 비가 내리면서 형산강의 경주시 강동대교 지점에 홍수주의보가 내려졌습니다 낙동강홍수통제소는 6일 5시 10분에 형산강 유역 경주시 강동대교(경상북도 경주시 강동면 오금리 1046-1 (강동대교)) 지점에 홍수주의보를 발령한다고 밝혔습니다 경주시 강동대교 지점의 홍수주의보 수위는 6.50m, 홍수경보 수위는 7.50m입니다

126) 조근영. (2022년11월10일). '먹을 물도 바닥'...완도 최악 가뭄, 저수율 최저. 연합뉴스. <https://www.yna.co.kr/view/AKR20221110081400054?section=safe/news>

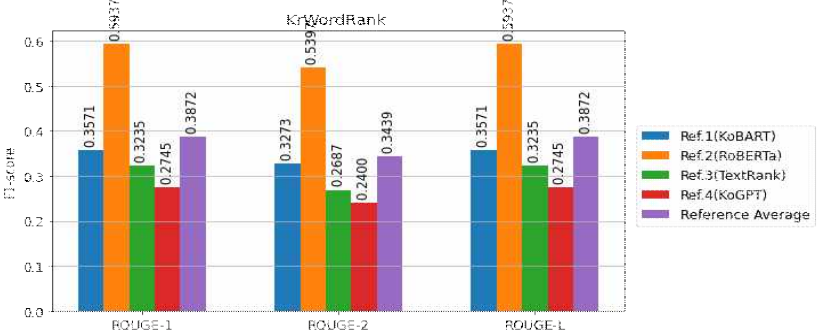
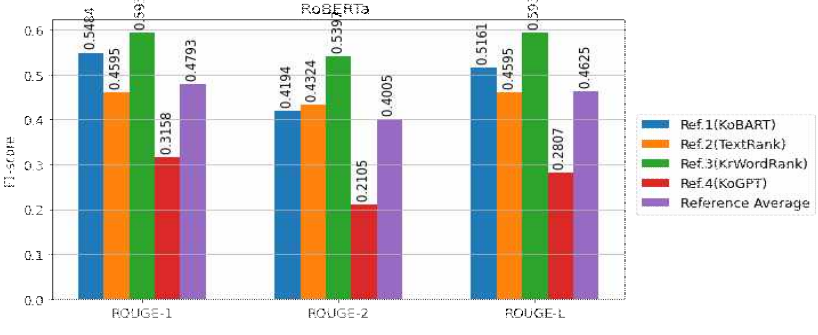
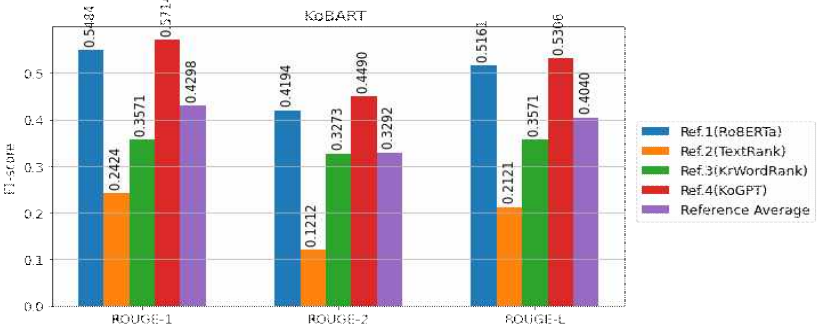
	<table><caption>TextRank F1-score Data</caption><thead><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(RoBERTa)</th><th>Ref.3(KrWordRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.2857</td><td>0.3810</td><td>0.6129</td><td>0.4545</td><td>0.4335</td></tr><tr><td>ROUGE-2</td><td>0.1765</td><td>0.3478</td><td>0.5588</td><td>0.4225</td><td>0.3764</td></tr><tr><td>ROUGE-L</td><td>0.1905</td><td>0.3810</td><td>0.6129</td><td>0.4545</td><td>0.4097</td></tr></tbody></table>	Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.2857	0.3810	0.6129	0.4545	0.4335	ROUGE-2	0.1765	0.3478	0.5588	0.4225	0.3764	ROUGE-L	0.1905	0.3810	0.6129	0.4545	0.4097
Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.2857	0.3810	0.6129	0.4545	0.4335																				
ROUGE-2	0.1765	0.3478	0.5588	0.4225	0.3764																				
ROUGE-L	0.1905	0.3810	0.6129	0.4545	0.4097																				
KrWordRank	경주시 강동대교 지점의 홍수주의보 수위는 6.50m, 홍수경보 수위는 7.50m입니다.또, 하천 제방이 버틸 수 있는 한계를 의미하는 계획홍수위는 9.73m입니다.형산강 유역에 많은 비가 내리면서 형산강의 경주시 강동대교 지점에 홍수주의보가 내려졌습니다. <table><caption>KrWordRank F1-score Data</caption><thead><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(RoBERTa)</th><th>Ref.3(TextRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.1639</td><td>0.1220</td><td>0.6129</td><td>0.4062</td><td>0.3263</td></tr><tr><td>ROUGE-2</td><td>0.0968</td><td>0.0000</td><td>0.5588</td><td>0.2769</td><td>0.2331</td></tr><tr><td>ROUGE-L</td><td>0.1639</td><td>0.1220</td><td>0.6129</td><td>0.4062</td><td>0.3263</td></tr></tbody></table>	Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.1639	0.1220	0.6129	0.4062	0.3263	ROUGE-2	0.0968	0.0000	0.5588	0.2769	0.2331	ROUGE-L	0.1639	0.1220	0.6129	0.4062	0.3263
Metric	Ref.1(KoBART)	Ref.2(RoBERTa)	Ref.3(TextRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.1639	0.1220	0.6129	0.4062	0.3263																				
ROUGE-2	0.0968	0.0000	0.5588	0.2769	0.2331																				
ROUGE-L	0.1639	0.1220	0.6129	0.4062	0.3263																				
RoBERTa	낙동강홍수통제소는 6일 5시 10분에 형산강 유역 경주시 강동대교 (경상북도 경주시 강동면 오금리 1046-1 (강동대교)) 지점에 홍수주의보를 발령한다고 밝혔습니다. 낙동강홍수통제소는 경주에서 하천 수위 상승과 범람에 따른 홍수 피해가 우려된다며 주민들에게 하천변에 접근하지 말고 안전한 곳에 머무는 등 홍수 피해가 없도록 유의해 달라고 당부했습니다. 홍수주의보는 최대 홍수량의 50%까지 수위가 높아질 때 내려지며, 천변 둔치와 자전거도로 등 시설물이 침수될 수 있습니다. <table><caption>RoBERTa F1-score Data</caption><thead><tr><th>Metric</th><th>Ref.1(KoBART)</th><th>Ref.2(TextRank)</th><th>Ref.3(KrWordRank)</th><th>Ref.4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.6024</td><td>0.3810</td><td>0.1220</td><td>0.5814</td><td>0.4217</td></tr><tr><td>ROUGE-2</td><td>0.4651</td><td>0.3478</td><td>0.0000</td><td>0.4944</td><td>0.3268</td></tr><tr><td>ROUGE-L</td><td>0.6024</td><td>0.3810</td><td>0.1220</td><td>0.5814</td><td>0.4217</td></tr></tbody></table>	Metric	Ref.1(KoBART)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average	ROUGE-1	0.6024	0.3810	0.1220	0.5814	0.4217	ROUGE-2	0.4651	0.3478	0.0000	0.4944	0.3268	ROUGE-L	0.6024	0.3810	0.1220	0.5814	0.4217
Metric	Ref.1(KoBART)	Ref.2(TextRank)	Ref.3(KrWordRank)	Ref.4(KoGPT)	Reference Average																				
ROUGE-1	0.6024	0.3810	0.1220	0.5814	0.4217																				
ROUGE-2	0.4651	0.3478	0.0000	0.4944	0.3268																				
ROUGE-L	0.6024	0.3810	0.1220	0.5814	0.4217																				

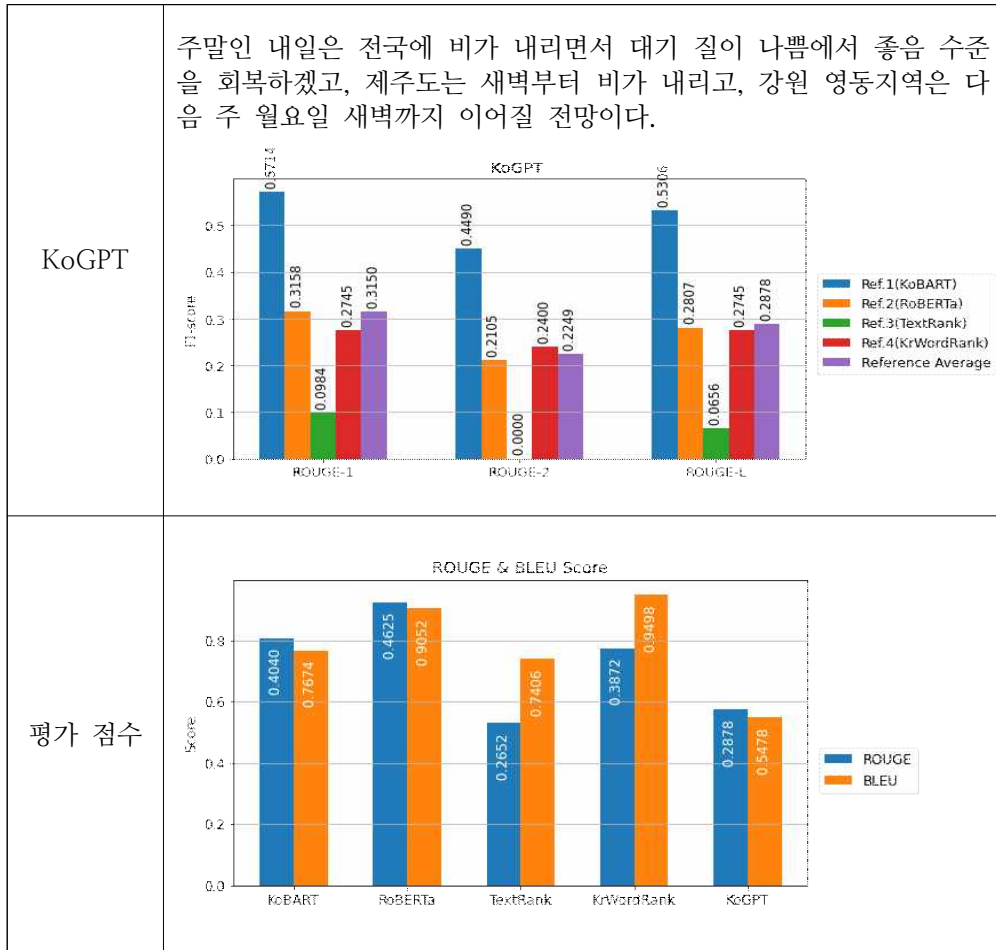
	낙동강 6일 5시 10분에 낙동강홍수통제소는 하천 수위 상승과 범람에 따른 홍수 피해가 우려되어 형산강의 경주시 강동대교 지점에 홍수주의보를 발령하며 주민들에게 하천변에 접근하지 말고 안전한 곳에 머무는 등 홍수 피해가 없도록 유의해 달라고 당부했다.																								
KoBART	 <table><caption>KoBART F1-score Data</caption><thead><tr><th>Reference</th><th>ROUGE-1</th><th>ROUGE-2</th><th>ROUGE-L</th></tr></thead><tbody><tr><td>Ref.1(RoBERTa)</td><td>0.6024</td><td>0.4651</td><td>0.6024</td></tr><tr><td>Ref.2(TextRank)</td><td>0.2857</td><td>0.1765</td><td>0.1905</td></tr><tr><td>Ref.3(KrWordRank)</td><td>0.1639</td><td>0.0968</td><td>0.1639</td></tr><tr><td>Ref.4(KoGPT)</td><td>0.2462</td><td>0.0923</td><td>0.1846</td></tr><tr><td>Reference Average</td><td>0.3246</td><td>0.2077</td><td>0.2853</td></tr></tbody></table>	Reference	ROUGE-1	ROUGE-2	ROUGE-L	Ref.1(RoBERTa)	0.6024	0.4651	0.6024	Ref.2(TextRank)	0.2857	0.1765	0.1905	Ref.3(KrWordRank)	0.1639	0.0968	0.1639	Ref.4(KoGPT)	0.2462	0.0923	0.1846	Reference Average	0.3246	0.2077	0.2853
Reference	ROUGE-1	ROUGE-2	ROUGE-L																						
Ref.1(RoBERTa)	0.6024	0.4651	0.6024																						
Ref.2(TextRank)	0.2857	0.1765	0.1905																						
Ref.3(KrWordRank)	0.1639	0.0968	0.1639																						
Ref.4(KoGPT)	0.2462	0.0923	0.1846																						
Reference Average	0.3246	0.2077	0.2853																						
KoGPT	낙동강홍수통제소는 6일 5시 10분에 형산강 유역 경주시 강동대교 (경상북도 경주시 강동면 오금리 1046-1 (강동대교)) 지점에 홍수주의보를 발령한다고 밝혔으며, 홍수주의보는 최대 홍수량의 50%까지 수위가 높아질 때 내려지며, 하천 제방이 버틸 수 있는 한계를 의미하는 계획홍수위는 9.73m입니다.  <table><caption>KoGPT F1-score Data</caption><thead><tr><th>Reference</th><th>ROUGE-1</th><th>ROUGE-2</th><th>ROUGE-L</th></tr></thead><tbody><tr><td>Ref.1(KoBART)</td><td>0.2462</td><td>0.0923</td><td>0.1846</td></tr><tr><td>Ref.2(RoBERTa)</td><td>0.5814</td><td>0.4944</td><td>0.5814</td></tr><tr><td>Ref.3(TextRank)</td><td>0.4545</td><td>0.4225</td><td>0.4545</td></tr><tr><td>Ref.4(KrWordRank)</td><td>0.4062</td><td>0.2769</td><td>0.4062</td></tr><tr><td>Reference Average</td><td>0.4221</td><td>0.3215</td><td>0.4067</td></tr></tbody></table>	Reference	ROUGE-1	ROUGE-2	ROUGE-L	Ref.1(KoBART)	0.2462	0.0923	0.1846	Ref.2(RoBERTa)	0.5814	0.4944	0.5814	Ref.3(TextRank)	0.4545	0.4225	0.4545	Ref.4(KrWordRank)	0.4062	0.2769	0.4062	Reference Average	0.4221	0.3215	0.4067
Reference	ROUGE-1	ROUGE-2	ROUGE-L																						
Ref.1(KoBART)	0.2462	0.0923	0.1846																						
Ref.2(RoBERTa)	0.5814	0.4944	0.5814																						
Ref.3(TextRank)	0.4545	0.4225	0.4545																						
Ref.4(KrWordRank)	0.4062	0.2769	0.4062																						
Reference Average	0.4221	0.3215	0.4067																						
평가 점수	 <table><caption>ROUGE & BLEU Score Data</caption><thead><tr><th>Model</th><th>ROUGE</th><th>BLEU</th></tr></thead><tbody><tr><td>KoBART</td><td>0.2854</td><td>0.9133</td></tr><tr><td>RoBERTa</td><td>0.4217</td><td>0.8005</td></tr><tr><td>TextRank</td><td>0.4097</td><td>0.8512</td></tr><tr><td>KrWordRank</td><td>0.3263</td><td>0.9394</td></tr><tr><td>KoGPT</td><td>0.4067</td><td>0.8620</td></tr></tbody></table>	Model	ROUGE	BLEU	KoBART	0.2854	0.9133	RoBERTa	0.4217	0.8005	TextRank	0.4097	0.8512	KrWordRank	0.3263	0.9394	KoGPT	0.4067	0.8620						
Model	ROUGE	BLEU																							
KoBART	0.2854	0.9133																							
RoBERTa	0.4217	0.8005																							
TextRank	0.4097	0.8512																							
KrWordRank	0.3263	0.9394																							
KoGPT	0.4067	0.8620																							

127) 임재성. (2022년9월6일). 형산강 경주시 강동대교 지점 홍수주의보. KBS뉴스.
<https://news.kbs.co.kr/news/view.do?ncd=5549853>

[표 4-7] 모델별 실험 결과 7 (KBS: 미세먼지)

구분	문서																								
원문	오늘 중부지방은 대체로 맑겠고, 남부지방에는 구름이 많이 끼겠습니다. 오늘도 대기 정체가 이어지면서 수도권과 강원 영서, 충청, 전북, 대구와 경북지역의 초미세먼지 농도가 ‘나쁨’ 수준을 보이고 있습니다. 주말인 내일은 전국에 비가 내리면서 대기 질이 ‘보통’에서 ‘좋음’ 단계를 회복하겠습니다. 비는 내일 새벽, 제주도를 시작으로 오전에는 남해안, 오후에는 전국으로 확대하겠습니다. 이번 비는 일요일인 모레 오전에 대부분 그치겠지만, 강원 영동지역은 다음 주 월요일 새벽까지 이어지겠습니다. 주말 동안 예상 강수량은 경기 북부와 강원 영서 북부, 제주 동남부와 산지에 20~60mm, 나머지 지역은 5~40mm 등입니다. 내일 아침 기온은 서울이 12도 등 전국이 7도에서 15도로 오늘과 비슷하거나 조금 높겠습니다. 낮 최고기온은 서울이 20도 등 전국이 19도에서 24도로 오늘과 비슷하거나 조금 낮겠습니다. 내일 바다의 물결은 제주 부근 해상과 모든 해상의 먼바다에서 최고 4m로 높게 일겠습니다.¹²⁸⁾																								
요약모델	자동요약 문서																								
TextRank	오늘도 대기 정체가 이어지면서 수도권과 강원 영서, 충청, 전북, 대구와 경북지역의 초미세먼지 농도가 ‘나쁨’ 수준을 보이고 있습니다. 비는 내일 새벽, 제주도를 시작으로 오전에는 남해안, 오후에는 전국으로 확대하겠습니다. 내일 아침 기온은 서울이 12도 등 전국이 7도에서 15도로 오늘과 비슷하거나 조금 높겠습니다. Figure 1: F1-score of TextRank model <table><thead><tr><th>Metric</th><th>Ref. 1(KoBART)</th><th>Ref. 2(RoBERTa)</th><th>Ref. 3(KrWordRank)</th><th>Ref. 4(KoGPT)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.2424</td><td>0.4595</td><td>0.3235</td><td>0.0984</td><td>0.2809</td></tr><tr><td>ROUGE-2</td><td>0.1212</td><td>0.4333</td><td>0.2687</td><td>0.0000</td><td>0.2056</td></tr><tr><td>ROUGE-L</td><td>0.2121</td><td>0.4595</td><td>0.3235</td><td>0.0656</td><td>0.2652</td></tr></tbody></table>	Metric	Ref. 1(KoBART)	Ref. 2(RoBERTa)	Ref. 3(KrWordRank)	Ref. 4(KoGPT)	Reference Average	ROUGE-1	0.2424	0.4595	0.3235	0.0984	0.2809	ROUGE-2	0.1212	0.4333	0.2687	0.0000	0.2056	ROUGE-L	0.2121	0.4595	0.3235	0.0656	0.2652
Metric	Ref. 1(KoBART)	Ref. 2(RoBERTa)	Ref. 3(KrWordRank)	Ref. 4(KoGPT)	Reference Average																				
ROUGE-1	0.2424	0.4595	0.3235	0.0984	0.2809																				
ROUGE-2	0.1212	0.4333	0.2687	0.0000	0.2056																				
ROUGE-L	0.2121	0.4595	0.3235	0.0656	0.2652																				
KrWordRank	오늘 중부지방은 대체로 맑겠고, 남부지방에는 구름이 많이 끼겠습니다. 주말인 내일은 전국에 비가 내리면서 대기 질이 ‘보통’에서 ‘좋음’ 단계를 회복하겠습니다. 비는 내일 새벽, 제주도를 시작으로 오전에는 남해안, 오후에는 전국으로 확대하겠습니다.																								

	 Bar chart showing F1-score for KrWordRank across ROUGE-1, ROUGE-2, and ROUGE-L metrics. The chart compares four references: Ref.1(KoBART), Ref.2(RoBERTa), Ref.3(TextRank), Ref.4(KoGPT), and a Reference Average. The values are: ROUGE-1 (0.3571, 0.5937, 0.3235, 0.2745, 0.3872), ROUGE-2 (0.3273, 0.5397, 0.2687, 0.2400, 0.3439), ROUGE-L (0.3571, 0.5937, 0.3235, 0.2745, 0.3872).
RoBERTa	오늘 중부지방은 대체로 맑겠고, 남부지방에는 구름이 많이 끼겠습니다. 오늘도 대기 정체가 이어지면서 수도권과 강원 영서, 충청, 전북, 대구와 경북지역의 초미세먼지 농도가 ‘나쁨’ 수준을 보이고 있습니다. 주말인 내일은 전국에 비가 내리면서 대기 질이 ‘보통’에서 ‘좋음’ 단계를 회복하겠습니다.  Bar chart showing F1-score for RoBERTa across ROUGE-1, ROUGE-2, and ROUGE-L metrics. The chart compares four references: Ref.1(KoBART), Ref.2(TextRank), Ref.3(KrWordRank), Ref.4(KoGPT), and a Reference Average. The values are: ROUGE-1 (0.5484, 0.4595, 0.5937, 0.3158, 0.4793), ROUGE-2 (0.4194, 0.4324, 0.5397, 0.2105, 0.4005), ROUGE-L (0.5161, 0.4595, 0.5937, 0.2807, 0.4625).
KoBART	대기 정체가 이어지면서 초미세먼지 농도가 '나쁨' 수준을 보이고 있는 가운데 주말인 내일은 전국에 비가 내리면서 대기 질이 ‘보통’에서 ‘좋음’ 단계를 회복하겠고 강원 영동지역은 다음 주 월요일 새벽까지 이어질 것입니다.  Bar chart showing F1-score for KoBART across ROUGE-1, ROUGE-2, and ROUGE-L metrics. The chart compares four references: Ref.1(RoBERTa), Ref.2(TextRank), Ref.3(KrWordRank), Ref.4(KoGPT), and a Reference Average. The values are: ROUGE-1 (0.5484, 0.2424, 0.3571, 0.5714, 0.4298), ROUGE-2 (0.4194, 0.1212, 0.3273, 0.4490, 0.3292), ROUGE-L (0.5161, 0.2121, 0.3571, 0.5386, 0.4040).



[표 4-5]~[표 4-7]에서 가뭄, 홍수, 미세먼지와 관련된 뉴스 요약시 추천되는 모델의 유형을 확인해 보았다. 이 실험에서는 KoBART, KrWordRank, RoBERTa 등 다양한 모델의 요약 결과가 추천되었다. 이처럼 어떤 특별한 모델이 재난 관련된 뉴스 요약에 특화가 되었다고 추론하기가 어려웠다.

128) 신방실. (2022년11월11일). 오늘도 초미세먼지 ‘나쁨’…주말 전국에 비. *KBS뉴스*.
<https://news.kbs.co.kr/news/view.do?ncd=5599418>

[표 4-8] 모델별 실험 결과 정리

실험	모델	ROUGE-L		BLEU		추천
		F1점수	순위	F1점수	순위	
1	KoBART	0.2595	2	0.8909	1	○
	RoBERTa	0.1572	5	0.4170	5	
	TextRank	0.2887	1	0.6887	3	
	KrWordRank	0.2318	4	0.6022	4	
	KoGPT	0.2568	3	0.8297	2	
2	KoBART	0.3160	4	0.6602	4	
	RoBERTa	0.2535	5	0.4500	5	
	TextRank	0.3984	2	0.9035	2	○
	KrWordRank	0.3985	1	0.8360	3	
	KoGPT	0.3367	3	0.9429	1	
3	KoBART	0.6799	2	0.8320	2	
	RoBERTa	0.7353	1	0.9294	1	○
	TextRank	0.3842	5	0.4942	5	
	KrWordRank	0.6732	3	0.8231	3	
	KoGPT	0.6258	4	0.6722	4	
4	KoBART	0.3331	3	0.6913	3	
	RoBERTa	0.3353	2	0.7806	2	
	TextRank	0.2747	4	0.6447	4	
	KrWordRank	0.3968	1	0.8076	1	○
	KoGPT	0.2687	5	0.5575	5	
5	KoBART	0.4799	2	0.9507	1	○
	RoBERTa	0.4797	3	0.9213	2	
	TextRank	0.1626	5	0.3416	5	
	KrWordRank	0.4965	1	0.7914	3	
	KoGPT	0.3579	4	0.4364	4	
6	KoBART	0.2854	5	0.9133	2	
	RoBERTa	0.4217	1	0.8005	5	
	TextRank	0.4097	2	0.8512	4	
	KrWordRank	0.3262	4	0.9394	1	○
	KoGPT	0.4067	3	0.8620	3	
7	KoBART	0.4040	2	0.7674	3	
	RoBERTa	0.4625	1	0.9052	2	○
	TextRank	0.2652	5	0.7406	4	
	KrWordRank	0.3872	3	0.9498	1	
	KoGPT	0.2878	4	0.5478	5	

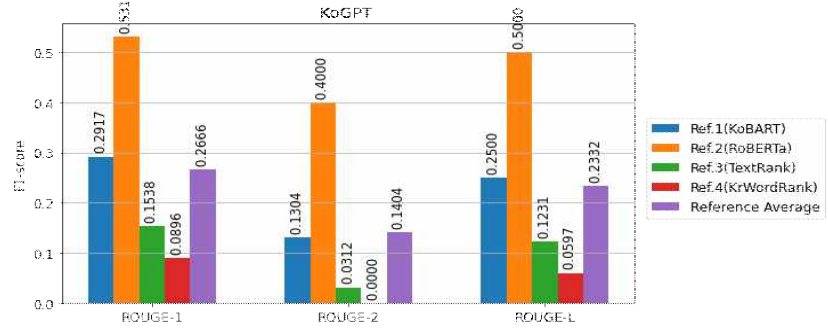
2) 참조본별 요약 결과 및 평가 결과

KoGPT 모델의 경우 생성 요약으로 요약을 진행할 때마다 다른 요약 결과가 나왔다. 그래서 다르게 생성된 요약 결과물을 바꾸어가며 참조본으로 이용했을 때 나머지 모델에 미치는 영향을 확인하기 위한 실험을 추가로 진행하였다. 원본 데이터는 실험 1, 2에서 진행했던 코로나19와 날씨 뉴스로 진행했으며 각각 2번씩 요약생성을 재실험하여 총 6번의 실험을 했으며 결과는 다음과 같이 나타났다.

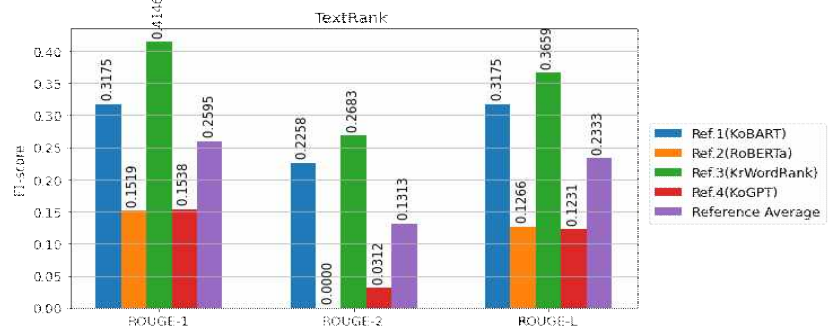
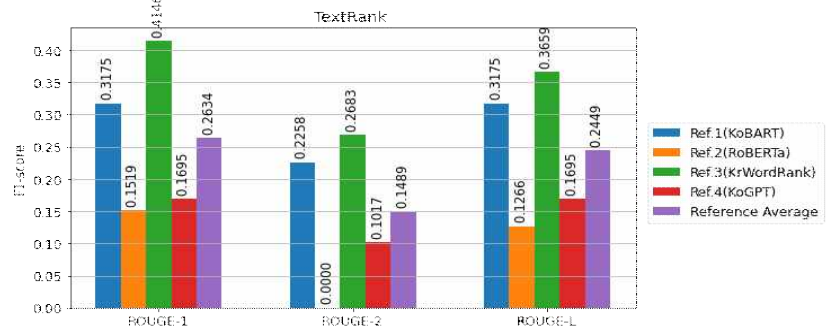
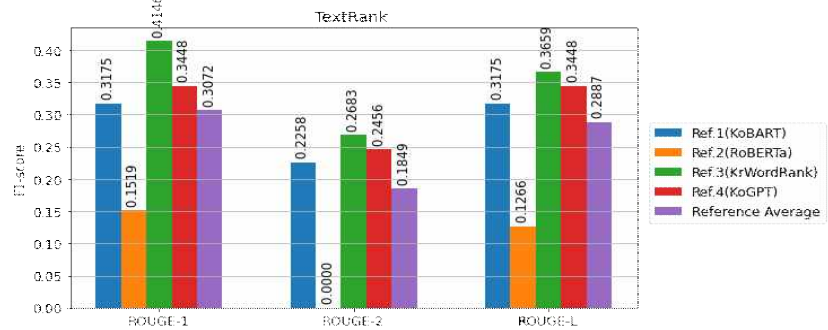
[표 4-9] GPT 변경 실험 결과 1-1, 1-2 (연합뉴스: 코로나19)

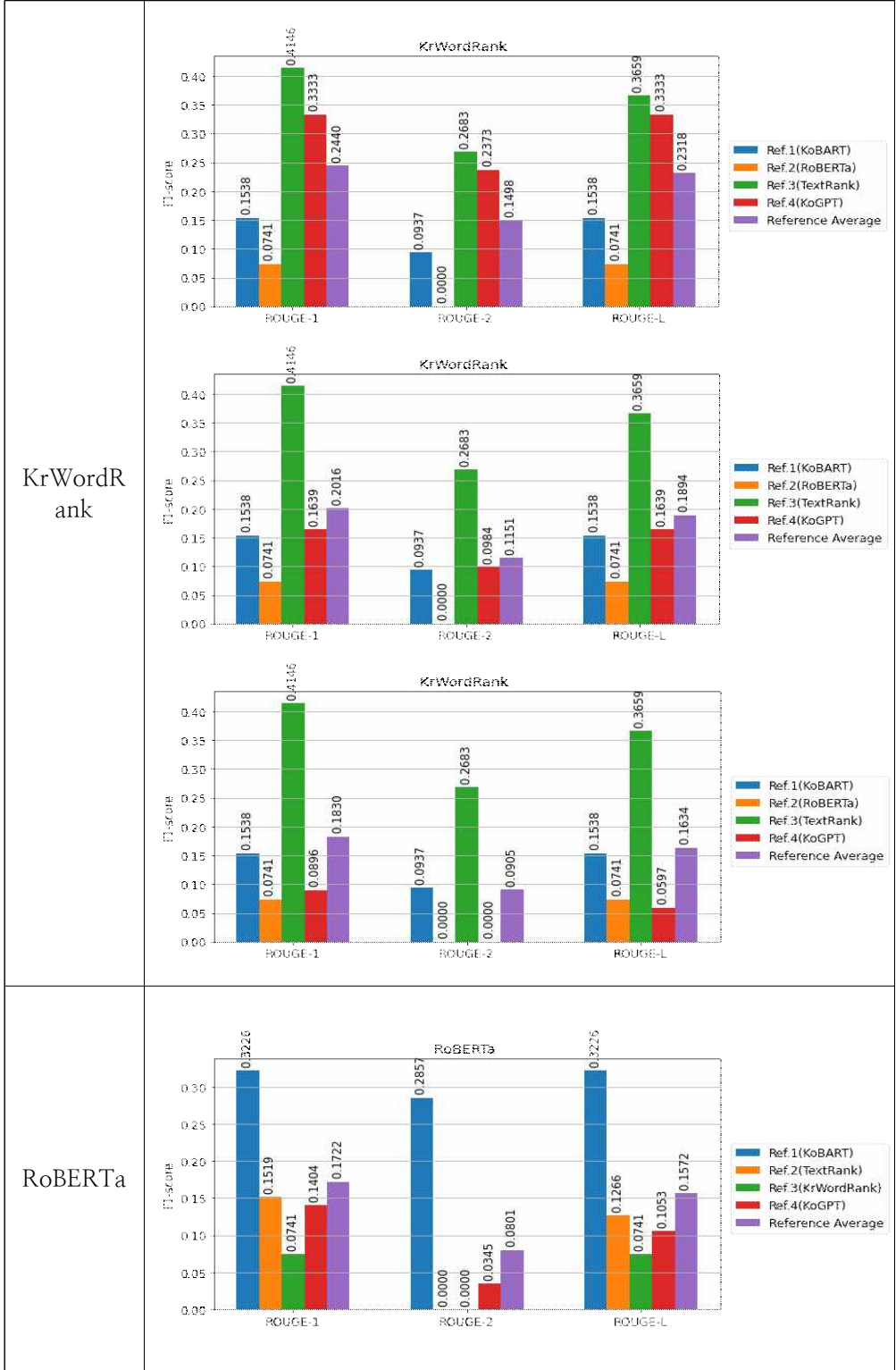
요약모델	자동요약 문서																																																
	중앙방역대책본부는 18일 0시 기준 코로나19 확진자가 4만9천418명 늘어 누적 2천646만2천319명이 됐으며, 신규 확진자 수는 전날(5만5천437명)보다 6천19명 줄었다고 밝혔다. KoGPT <table><thead><tr><th></th><th>Ref. 1(KoBART)</th><th>Ref. 2(RoBERTa)</th><th>Ref. 3(TextRank)</th><th>Ref. 4(KrWordRank)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.4330</td><td>0.1404</td><td>0.3448</td><td>0.3333</td><td>0.3144</td></tr><tr><td>ROUGE-2</td><td>0.2051</td><td>0.0345</td><td>0.2456</td><td>0.2373</td><td>0.1806</td></tr><tr><td>ROUGE-L</td><td>0.2439</td><td>0.1053</td><td>0.3448</td><td>0.3333</td><td>0.2568</td></tr></tbody></table> KoGPT 18일 신규 확진자 수가 4만명대로 내려왔는데, 이는 코로나19 7차 유행이 진행 중인 가운데 18일 신규 확진자 수가 4만9천418명 늘어 누적 2천646만2천319명이 되었다고 발표했다. KoGPT <table><thead><tr><th></th><th>Ref. 1(KoBART)</th><th>Ref. 2(RoBERTa)</th><th>Ref. 3(TextRank)</th><th>Ref. 4(KrWordRank)</th><th>Reference Average</th></tr></thead><tbody><tr><td>ROUGE-1</td><td>0.6667</td><td>0.3793</td><td>0.1695</td><td>0.1639</td><td>0.3448</td></tr><tr><td>ROUGE-2</td><td>0.5854</td><td>0.3333</td><td>0.1017</td><td>0.0984</td><td>0.2797</td></tr><tr><td>ROUGE-L</td><td>0.6667</td><td>0.3448</td><td>0.1695</td><td>0.1639</td><td>0.3362</td></tr></tbody></table>		Ref. 1(KoBART)	Ref. 2(RoBERTa)	Ref. 3(TextRank)	Ref. 4(KrWordRank)	Reference Average	ROUGE-1	0.4330	0.1404	0.3448	0.3333	0.3144	ROUGE-2	0.2051	0.0345	0.2456	0.2373	0.1806	ROUGE-L	0.2439	0.1053	0.3448	0.3333	0.2568		Ref. 1(KoBART)	Ref. 2(RoBERTa)	Ref. 3(TextRank)	Ref. 4(KrWordRank)	Reference Average	ROUGE-1	0.6667	0.3793	0.1695	0.1639	0.3448	ROUGE-2	0.5854	0.3333	0.1017	0.0984	0.2797	ROUGE-L	0.6667	0.3448	0.1695	0.1639	0.3362
	Ref. 1(KoBART)	Ref. 2(RoBERTa)	Ref. 3(TextRank)	Ref. 4(KrWordRank)	Reference Average																																												
ROUGE-1	0.4330	0.1404	0.3448	0.3333	0.3144																																												
ROUGE-2	0.2051	0.0345	0.2456	0.2373	0.1806																																												
ROUGE-L	0.2439	0.1053	0.3448	0.3333	0.2568																																												
	Ref. 1(KoBART)	Ref. 2(RoBERTa)	Ref. 3(TextRank)	Ref. 4(KrWordRank)	Reference Average																																												
ROUGE-1	0.6667	0.3793	0.1695	0.1639	0.3448																																												
ROUGE-2	0.5854	0.3333	0.1017	0.0984	0.2797																																												
ROUGE-L	0.6667	0.3448	0.1695	0.1639	0.3362																																												

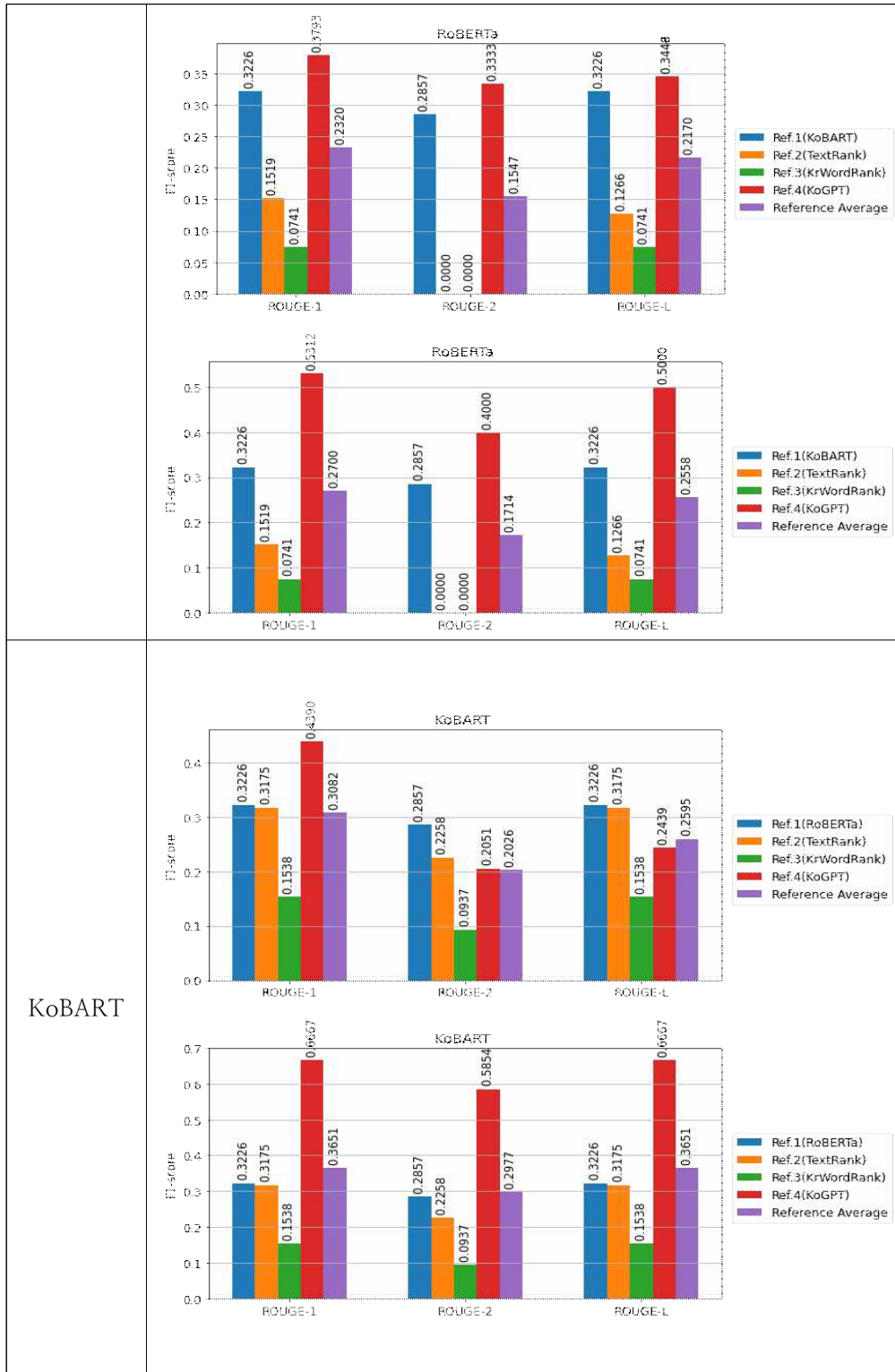
18일 코로나19 신규 확진자 수가 4만명대로 감소하였는데, 이는 최근 유행 확산세에서 이례적으로 감소한 것으로, 지난 16일 4만119명, 17일 80명 증가하는 데 그쳤고 이날 감소세로 돌아섰다.

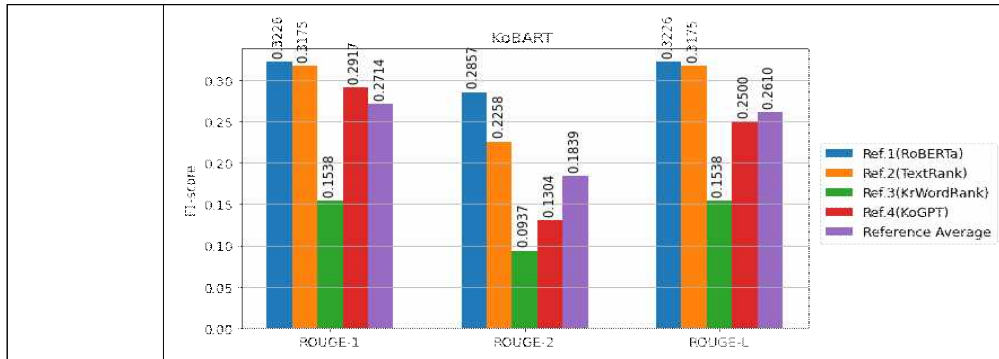


TextRank









[표 4-10] GPT 재요약 시 실험 결과 정리

실험	모델	ROUGE-L		BLEU		추천
		F1점수	순위	F1점수	순위	
1	KoBERT	0.2595	2	0.8909	1	O
	RoBERTa	0.1572	5	0.4170	5	
	TextRank	0.2887	1	0.6887	3	
	KrWordRank	0.2318	4	0.6022	4	
	KoGPT	0.2568	3	0.8297	2	
1-1	KoBERT	0.3651	1	0.8733	1	O
	RoBERTa	0.2170	4	0.4717	5	
	TextRank	0.2449	3	0.6790	3	
	KrWordRank	0.1894	5	0.5989	4	
	KoGPT	0.3362	2	0.8286	2	
1-2	KoBERT	0.2610	1	0.8411	1	O
	RoBERTa	0.2558	2	0.6070	4	
	TextRank	0.2333	3	0.6780	3	
	KrWordRank	0.1634	5	0.6023	5	
	KoGPT	0.2332	4	0.8033	2	
7	KoBERT	0.3331	3	0.6913	3	O
	RoBERTa	0.3353	2	0.7806	2	
	TextRank	0.2747	4	0.6447	4	
	KrWordRank	0.3968	1	0.8076	1	
	KoGPT	0.2687	5	0.5575	5	
7-1	KoBERT	0.3331	3	0.6913	3	O
	RoBERTa	0.3353	2	0.7806	2	
	TextRank	0.2747	4	0.6447	4	
	KrWordRank	0.3968	1	0.8076	1	
	KoGPT	0.2687	5	0.5575	5	
7-2	KoBERT	0.3708	2	0.6913	3	

	RoBERTa	0.3689	3	0.7806	2	
	TextRank	0.2473	5	0.5727	4	
	KrWordRank	0.4276	1	0.8094	1	○
	KoGPT	0.3433	4	0.5702	5	

KoGPT 모델의 경우 KoBART처럼 요약에 특화되지 않은 생성 요약 모델이며 설정 매개변수(parameter)에 따라 예상치 못한 형태로 요약된다. [표 4-6] 홍수 기사에서 매개변수 Temperature를 1.0으로 설정했을 때, ‘대구지방환경청’이나 ‘형산강 홍수통제소’와 같이 원문에는 존재하는 않는 내용으로 문장이 생성되는 경우도 있었다. 이는 정확한 사실에 근거한 뉴스의 요약에서는 적합하지가 않다. 0.5로 설정한 경우에는 요약된 내용이 사실에 기반을 두고 있지만, 내용을 예측하기 어렵게 다양한 결과가 나왔었다. 마지막으로 0에 가까운 0.1로 설정했을 경우에는 비교적 안정적으로 문장을 생성함을 알 수 있다. 하지만 ROUGE 점수는 그리 높지가 않았다. 값에 따라 요약 내용이 크게 달라지는 경우도 있었고, 동일한 결과를 보이는 경우도 있었다. 값이 커질수록 변동성이 커지고 작을수록 안정적인 결과를 나타내었다.

KoGPT 모델을 통한 문서 요약 시에는 5회 실행 후 점수평가를 자동으로 진행한 후 가장 우수한 결과물을 추천하게 하였다. 실험에서는 요약모델(KoBART, RoBERTa, TextRank, KrWordRank)에서 생성한 요약문을 참조본으로 하여 KoGPT 모델의 요약문을 평가하고, 또다시 요약문을 만들고 평가한 후 가장 우수한 점수가 나오는 문장으로 나머지 모델들의 요약문을 평가하는 방식으로 진행하였다.

제 3 절 연구의 성과

1) 시스템 활용의 기대효과

다섯 가지 언어 모델들(TextRank, KrWordRank, RoBERTa, KoBART, KoGPT)로 요약한 문서를 ROUGE 및 BLEU 점수 기준으로 평가하였을 때 가장 우수한 값을 가지는 모델이 데이터에 따라서 계속 달라지는 것을 확인할 수 있다. 그래서 특정 모델이 문서 요약에 우수하다기보다 모델별로 입력 데이터에 따라 순위가 다른 결과가 나오는 것을 확인할 수 있다. 이는 일반적으로 트랜스포머 기반의 모델에서 우수한 성능이 나타난다고 알려진 바와는 다른 결과이다.

F1 점수를 기준으로 보면 각 언어 모델에 대한 참조 요약본에 따라 상당히 상이한 점수가 분포하는 것으로 보이지만, 요약 방식에 따라 서로 영향을 주고 있음을 확인할 수 있다. 추출 요약모델(TextRank, KrWordRank)들 간의 평가에서는 서로 높은 점수를 보이고 있으며, 생성 요약모델(KoBART, KoGPT) 간에도 서로 높은 점수를 보임을 확인할 수 있다.

실험의 일부 예시에서 볼 수 있듯이 ROUGE 점수는 생성 요약보다는 추출 요약에서 더 나은 점수를 나타내고 있다. 그 이유는 생성모델의 경우 동일한 내용을 이해하고 문장을 생성하기 때문에 매개변수의 값에 따라 문서 요약에서 본문에는 없는 새로운 단어로 대체될 수 있기 때문이다. 이 경우 내용은 충실하였지만 평가 점수는 높지 않은 경우가 많았다. 예를 들면, “저수율(6%)이 증가하지 않았다”라고 요약이 되었는데 참조본에는 “저수율이 10%에도 미치지 못했다”라고 되어있다면 “저수율” 외에는 중복되는 단어(체언, 용언)가 없기 때문에 의미상으로는 유사성이 높아도 ROUGE 점수에서는 결과가 좋게 나올 수가 없다는 한계는 있다.

평가방법에 있어서 기존 연구사례와 비교하면 그 가치를 확인할 수 있다. 요약의 품질을 평가하기 위해서는 자동요약문과 참조본의 유사도를 계산하는 내용평가 방법이 가장 대표적인 평가방법이다. 기존 방식에서 참조본은 이상적인 요약본을 의미하며, 사람이 작성한 매뉴얼 요약본이 참조본으로 사용되

어 왔다. 구체적으로는 다양한 알고리즘을 통해 기계가 자동으로 생성한 요약문과 사람이 수작업으로 요약한 표준 문서를 비교하여 두 문서에서 정확히 일치하는 문장의 수를 세고 이를 통해 상대적인 유사성을 평가하는 방식이다.¹²⁹⁾¹³⁰⁾

하지만 연구를 통해 제시한 방식으로 평가하게 되면 5개의 모델이 요약한 요약문 중 1개의 후보문과 4개의 참조본이 자동으로 생성되어 사람이 참조본을 만들어서 평가할 필요가 없게 된다. 그리고 ROUGE나 BLEU 점수로 평가하는 이유는 사람이 평가하기까지 걸리는 시간과 높은 비용을 들일 필요가 없다는 장점 때문이다. 이처럼 다양한 모델을 이용하여 문장이 요약될 때마다 즉시 평가하고 평가점수가 높은 결과물을 선정해서 경고 방송용 콘텐츠를 선정하여 그 결과물을 전관방송시스템이나 알림 문자 등을 통해 전파한다면 전달력 향상에 큰 도움이 될 것으로 기대된다.

129) 고은정. (2018). *텍스트 분석을 통한 문서 요약 품질의 자동 평가 방법론*. (석사학위, 국민대학교 비즈니스IT전문대학원).

130) Radev, D., Jing, H., Budzikowska, M. (2004). Centroid-Based Summarization of Multiple Documents, *Information Processing & Management*, Vol.40, 919~938.

제 5 장 결론

본 연구에서는 복잡해진 사회에서 직면할 수 있는 다양한 재난 상황에 대비하여 미리 선제적 대응이 가능하도록 사전에 안내하는 방법에 관한 연구를 수행하였다. 재난에 대비하는 방법에는 다양한 방법이 존재한다. 그 방법으로 방송 장비의 성능 향상보다는 다가올 재난을 미리 감지하고 사람들이 듣고 이해하기 쉽게 정보를 요약하는 기술에 관심을 갖고 연구를 시작하였다. 국가에서 제공하는 실시간 공공데이터와 국가가 지정한 재난방송 주관사에서 배포하는 뉴스 기사를 자동으로 가져와서 재난 대비 방송을 할 수 있게 자동으로 문장을 요약해 주는 방법에 관한 연구를 진행하였다.

대부분 산업에서도 마찬가지로이지만 기술의 진화 방향은 인공지능 기술의 도움을 받는 방향으로 흘러가고 있는데 이는 비용과 성능 면에서도 우수하기 때문이다. 이는 재난 대비를 위한 경보 방송 시스템의 구성에서도 마찬가지이다. 기존의 전관방송 시스템은 재난이나 비상상황 시에 외부의 소방장비나 지진감지장치 같은 외부 장비의 요청에 의해 동작하는 시스템으로 재난에 능동적으로 대비할 수 있는 방법은 없었다. 재난에 대비하기 위한 정보를 실시간으로 가져오고 이를 안내하기 좋은 문장으로 만들어서 방송시스템의 TTS를 통해 알아듣기 쉽게 안내할 수 있는 방법을 고민하였다.

먼저 이를 위해 공공데이터로부터 실시간으로 가져올 수 있는 날씨와 미세먼지 정보를 이용하여 미리 시스템에서 정의해둔 수치 이상의 상태가 감지되어 경보가 필요한 경우 안내 문구 프리셋을 이용하여 적절한 안내문을 만들어서 방송을 해주는 것이다. 이때는 안내 문구의 내용이 복잡할 이유가 없기 때문에 상태의 좋고 나쁨 수준으로만 알려주면 된다.

다음으로, 국가지정 재난방송사의 재난 포털에서 재난과 관련된 뉴스들을 가져와서 방송에 적합한 콘텐츠로 재작성하는 부분이다. 재난 포털로부터 가져온 콘텐츠가 현지점에서 해당 지역에 관련된 뉴스인지 텍스트 마이닝을 통하여 분석한 후 방송용 콘텐츠로 적합한지 판단하였다. 이후 전관방송 시스템을 통해 방송을 송출하기 위해서 해당 기사를 방송하기 적당한 분량으로 재

작성하는 절차가 필요한데 여기서 다양한 방법으로 문서 요약을 진행하였다. 현시점에서 가장 대중적인 문서 요약모델들과 성능이 우수하다고 평가되고 있는 딥러닝 기반의 모델들을 사용하였다.

다음으로 언어 모델에 의해 요약된 문서를 실시간 평가하여 방송 콘텐츠를 추천하는 방법에 관한 연구를 진행했다. 그런데 기존의 요약문 평가방법에서는 평가를 위해서 사람이 만든 요약 참조(정답)본을 가지고 ROUGE 등의 평가방법을 통하여 요약된 문서의 품질을 평가하였는데 이를 요약 전문가들 대신 다양한 모델에서 생성한 요약본을 이용하여 실시간으로 평가하는 방법을 적용한 것이다. 여기서 사람이 개입하여 평가하게 된다면 실시간성이 떨어질 뿐만 아니라 인적 자원을 유지하기 위해 많은 비용을 지불해야 하는 문제가 발생한다. 본 연구에서는 사람이 개입할 필요가 없으며 시간과 비용 측면에서 훨씬 유리하다.

마지막으로 전관방송시스템은 오래전부터 존재했던 전통적인 시스템으로 전관방송의 기술개발 방향도 지금까지는 각종 센서 장비를 개발하여 추가하거나 음질의 개선을 위해 지향성 스피커를 이용한다거나 하드웨어적인 부분에서 개선하려는 움직임이 있었다. 간혹 AI 기술을 이용한 연구가 진행된 사례가 있었지만, 대피경로를 안내하는 등의 재난대응에 관한 연구만 존재했다. 하지만 재난 대비단계에서 안내하는 문구의 품질 개선을 위한 연구는 처음이다. 또한, 인공지능 기술 중 자연어 처리(NLP)를 적용하여 듣기 어려운 재난 뉴스를 간략히 요약 안내하여 재난대응이 가능해졌다는 것에 의의가 있다. 이를 통해 재난의 4단계 중 기존에는 대응단계에 대해서만 사용이 가능했던 비상방송 장비가 사전 단계인 대비단계까지 가능하여 재난 대비 장비로 활용의 폭을 넓히는 효과가 있게 되었다.

추후 추가 연구가 진행된다면 새로운 언어 모델을 사용하거나 더 좋은 데이터셋을 통한 개선된 모델을 이용한 요약 등도 있지만 요약문의 평가방법에 관한 연구가 중요할 것으로 생각된다. 이러한 연구가 더 진행된다면 자연어 처리 기술의 발전이 재난대비기술의 발전으로 이어지게 될 것으로 기대된다.

참 고 문 헌

1. 국내문헌

강종렬, 남지성, 박지나, 김웅섭. (2019). 키워드 요약의 세 가지 방법론 비교.

『한국정보처리학회 2019년 추계학술대회 논문집』, 26(2), 852-854.

고은정. (2018). 텍스트 분석을 통한 문서 요약 품질의 자동 평가 방법론. (석사학위, 국민대학교 비즈니스IT전문대학원).

공공데이터의 제공 및 이용 활성화에 관한 법률. <https://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률>

공공데이터포털. <https://www.data.go.kr>

국립재난안전연구원. (2022). 재난안전분야 국내·외 인공지능 기술동향. 『재난안전연구데이터』, 2022(16).

굿바이코로나. <https://corona-19.kr>

기후변화와 에너지. e-나라지표. (2021).

https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1628

김병규. (2022년11월18일). 신규확진 4만9418명, 전주보다 5천여명↓…증가세 주춤(종합). 연합뉴스.

<https://www.yna.co.kr/view/AKR20221118043900530?section=coronavirus/news>

김병규. (2022년11월20일). 4만6011명 확진, 3일째 전주보다 줄어…당국 "일시적 감소"(종합2보). 연합뉴스.

<https://www.yna.co.kr/view/AKR20221120015851530?section=coronavirus/news>

김세현. (2022년11월16일). 수능일 ‘큰 일교차’ 주의…내일도 수도권·충청 미

세먼지 ‘나쁨’. KBS뉴스.

<https://news.kbs.co.kr/news/view.do?ncd=5602787>

김수현. (2022년11월3일). 인파사고 안전관리지침 제정한다...공연장 재난대응 매뉴얼 보완(종합). 연합뉴스.

<https://www.yna.co.kr/view/AKR20221103065151530?input=1195m>

김시국, 이민용, 이춘하. (2015). 피난성능향상을 위한 주소형 비상방송시스템의 필요성에 관한 연구. 『한국화재소방학회 논문지』, 29(5), 14-22.

김원석. (2019). 공공데이터를 활용한 생활환경 정보서비스 제공 앱 개발 (석사학위, 한남대학교 대학원).

김유리. (2016). 소프트웨어 교육용 공공개방데이터를 활용한 모바일 앱 구현에 관한 연구 (박사학위, 신라대학교 대학원).

김의순 & 임희석. (2021). 사전학습 기반의 법률문서 요약방법 비교연구, 『ACK 2021 학술발표대회 논문집』, 28(2). 614-617.

김태범, 조시범, 김다진솔. (2020). 산불 예방 및 대응을 위한 첨단기술. 『국립재난안전연구원 재난안전』, 22(4), 79-86.

김현중, 조성준, 강필성. (2014). KR-WordRank: WordRank를 개선한 비지도학습 기반 한국어 단어 추출 방법. 『대한산업공학회지』, 40(1), 18-33.

남원경, 이지수, 장백철. (2022). 사전 학습된 BERT 기반 모델을 이용한 구어체 텍스트 요약. 『한국컴퓨터정보학회논문지』, 27(8), 41-47.

류태하 & 김승천. (2022). 공공데이터와 IoT 센싱 데이터를 활용한 경보방송 방법에 관한 연구. 『한국인터넷방송통신학회 논문지』, 22(1), 21-27.

목하균. (2010). 지상파 재난방송 서비스 현황. 『TTA Journal』, 131, 61-66.

민방위기본법. 국가법령정보센터. (2022). <https://www.law.go.kr/법령/민방위기본법>

- 박선흥, 배은미, 이창환, 동아일보 미디어연구소 종합심의팀. (2008). 수록처: 『신문 글 바꾸기:올바른 기사 문장 작성법』. (10). 서울: LG상남언론재단.
- 박소영, 홍상기, 이강복. (2020). 재난예측 기술 개발 및 서비스 제공 동향. 『전자통신동향분석』, 35(1), 80-88.
- 변윤관, 이현지, 장석진, 최성중, 표경수. (2020). 재난문자 서비스에서 이미지 표출 방안. 『방송공학회논문지』, 25(3), 399-404.
- 부산광역시 빌딩풍 예방 및 피해 저감에 관한 조례안. 부산광역시의회. (2021).
https://council.busan.go.kr/assem/Synap/skin/doc.html?fn=163184532320337.hwp&rs=/assem/upload_data/Synap
- 비립. (2021). 한국어 문장 단위 환원 현상 연구 (박사학위, 연세대학교 언어정보협동과정).
- 사회재난행동요령 화재. 국민재난안전포털.
<https://m.safekorea.go.kr/idsiSFK/neo/sfk/cs/contents/prevent/SDIJKM5116.html?menuSeq=127>
- 서지영. (2021). 『딥러닝 텐서플로 교과서』. 서울: (주)도서출판 길벗.
- 송채은 & 김학래. (2022). 지방자치단체의 공공데이터 개방 현황 분석과 개선방안. 『한국디지털컨텐츠학회』, 23(4), 697-705.
- 쇼홈 고시 & 드와이트 거닝. (2020). 예제로 배우는 자연어 처리 기초, (김창엽, 최민환, 역). 서울: 에이콘출판(주). (원본 출판 2019).
- 수다르산 라비찬다란. (2021). 구글 BERT의 정석. (전희원, 정승환, 김형준, 역). 서울: 한빛미디어(주). (원본 출판 2021).
- 신방실. (2022년11월11일). 오늘도 초미세먼지 ‘나쁨’...주말 전국에 비. KBS 뉴스. <https://news.kbs.co.kr/news/view.do?ncd=5599418>
- 신방실. (2022년11월13일). 비구름 지나고 기온 ‘뚝’...강원 산지 5cm 눈.

- KBS뉴스. <https://news.kbs.co.kr/news/view.do?ncd=5600264>
- 신사임, 김산, 서현태. (2021). KE-T5: 한국어-영어 대용량 텍스트를 활용한 이중언어 사전학습기반 대형 언어모델 구축. 『한국정보과학회 언어공학연구회: 학술대회논문집』, 2021.10a, 419-422.
- 심홍진 외(정보통신정책연구원). (2014). 재난방송 제도 개선에 관한 연구. 『방송통신융합 정책연구과제 최종보고서』.
- 안소영, 전인찬, 김지희, 이용태, 최성종. (2015). 인터넷 기반의 수신자 맞춤형 재난경보 전달시스템 연구. 『방송공학회논문지』, 20(4), 598-605.
- 에스에이텍 통합재난정보전달솔루션.
<http://www.safealarm.co.kr/satech/solution01.asp>
- 에어코리아 대기오염정보 조회 서비스 기술문서, v1.0, (2020).
<https://www.data.go.kr/data/15073861/openapi.do>
- 에이닷에 적용된 거대언어모델 GPT-3가 무엇일까?, SK텔레콤. (2022년5월 26일). <https://news.sktelecom.com/178491>
- 엔브릭스(enbrix). <http://enbrix.net/kr/main/main.php>
- 오소영, 조현민, 이영주, 이민주, 윤명오. (2018). 공동주택 내 비상방송설비의 음성명료도 실태 분석 및 재실자 인지성 개선방안 연구. 『한국화재소방학회』, 32(4), 60-68.
- 오원석. (2017). [IT후마니타스] 미세먼지 걱정되면 스마트폰에 물어봐요. 중앙일보. <https://www.joongang.co.kr/article/21470477#home>
- 왕규현. (2021). Paraphrasing training data for neural machine translation. (석사학위, 고려대학교 대학원).
- 윤재민, 정유진, 이종혁. (2004). 육하원칙 활성화도를 이용한 신문기사 자동 추출요약. 『정보과학회논문지』, 31(4), 505-515.
- 이경호. (2020). 추출 기반 문서 요약과 문장 축약 기술을 이용한 한국어 문서 요약에 관한 연구 (박사학위, 충남대학교 대학원).

- 이상원. (2022년9월1일). 화이트스캔, 서울시·KT ‘실시간 도시데이터 서비스’ 개발. 전자신문. <https://www.etnews.com/20220901000247>
- 이영배. (2021). 딥러닝 기반 자동작곡에서 마디 임베딩을 이용한 곡의 생성 및 정량적 평가 방법에 대한 연구 (석사학위, 한성대학교 미래융합컨설팅학과).
- 이일주 & 김민구. (2006). 그래프 분할을 이용한 문장 클러스터링 기반 문자 요약. 『정보처리학회논문지』, 13-B(2). 149-154.
- 이주현. (2021). 딥러닝 기반 대화 문장의 보완대체의사소통 상징 시퀀스 변환 (석사학위, 성신여자대학교).
- 이진기. (2022). 『안녕,트랜스포머』. 서울: 에이콘출판주식회사.
- 임재성. (2022년9월6일). 형산강 경주시 강동대교 지점 홍수주의보. KBS뉴스. <https://news.kbs.co.kr/news/view.do?ncd=5549853>
- 임희석. (2020). 『자연어처리바이블-핵심이론·응용시스템·딥러닝』. 서울: 휴먼싸이언스.
- 장동현 & 맹성현. (1997). 자동요약시스템. 『정보과학회』, 15(10), 42-49.
- 장석진 & 이용태. (2019). 신뢰성 있는 재난경보 방송을 위한 지능형 이동 로봇의 협업 기법. 『방송공학회논문지』, 24(3), 396-401.
- 재난 및 안전관리 기본법. 국가법령정보센터.(2022). <https://www.law.go.kr/법령/재난및안전관리기본법>
- 재난관리, 네 가지 단계를 아시나요? 기상청블로그. (2016). https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=kma_131&logNo=220594792032
- 전국 각처에 지진이 있다. 조선왕조실록. https://sillok.history.go.kr/id/kpa_12106009_001
- 전재원, 황현선, 이창기. (2019). 추출 요약과 생성 요약을 결합한 2단계 문서 요약. 『한국정보과학회 학술발표논문집』, 2019(6), 581-583.

- 전재원. (2019). 딥러닝 기반의 2단계 한국어 문서 요약 (석사학위, 강원대학교 대학원).
- 정상구, 박현호, 조정섭, 이용태. (2016). 일본의 재난관리 및 방송 시스템 현황. 『전자통신동향분석』, 31(3), 112-121.
- 조근영. (2022년11월10일). '먹을 물도 바닥'...완도 최악 가뭄, 저수율 최저. 연합뉴스.
<https://www.yna.co.kr/view/AKR20221110081400054?section=safe/news>
- 조용성, 배병준, 김홍묵. (2018). 미국의 재난방송 기술 동향. 『전자통신동향분석』, 33(3), 51-58.
- 진병문, 박재영, 안대명, 윤혜경. (2019). 『재난안전』. 서울: (주)도서출판 북스힐.
- 최성만. (2019). 정보시스템 측면에서 공공데이터 개방에 영향을 미치는 요인에 대한 연구: 제공자와 이용자의 차이분석을 중심으로 (박사학위, 단국대학교 대학원).
- 최재웅. (2012). 뉴미디어를 활용한 재난방송 전달체계 개선 및 재난방송정책 연구. (석사학위, 서울시립대학교 과학기술대학원).
- 최진호. (2022년11월9일). 중부지방 밤새 짙은 안개...내일 제주도에 가끔 비. KBS뉴스. <https://news.kbs.co.kr/news/view.do?ncd=5597465>
- 최찬홍. (2022년11월19일). 경기 1만4345명 신규 확진...1주 전보다 938명 감소. 연합뉴스.
<https://www.yna.co.kr/view/AKR20221119020300061?section=coronavirus/news>
- 표준공법-전관방송설비.
<http://pubicaddresssystem.blogspot.com/2017/03/2.html>
- 한국방재협회 방재사전편찬위원회. (2010). 재난 disaster. 수록처: 『방재사전

』.(305). 서울: R&D프레스.

한국어 문서 추출 요약 AI 경진대회. DAICON.

<https://dacon.io/competitions/official/235671/data/>

행정안전부 보도자료. (2022년11월3일). 다중 밀집 인파사고 예방 안전관리 대책 관계장관회의 개최 결과.

행정안전부. (2021). 공공부문 데이터 분석활용 우수사례집 발간데이터 분석으로 공공업무를 개선한 18개의 우수사례 수록.

현병천, 윤영욱, 박요한, 김영익. (2017). 스마트 디바이스를 활용한 가청, 비가청 신호 기반 피난방향 탐지 기법 연구. 『한국재난정보학회』, 13(1), 51-58.

황상흠 & 김도현. (2020). 한국어 기술문서 분석을 위한 BERT 기반의 분류 모델. 『한국전자거래학회지』, 25(1), 203-214.

황성환. (2022). 서울시 도시홍수예측시스템 구축 현황. 『국립재난안전연구원 재난안전』, 24(2), 32-35.

황우석 & 표경수. (2022). 국내 재난 예경보 시스템 현황 및 통합 방안에 대한 분석 연구. 『방송공학회 논문지』, 27(1), 80-90.

홉슨 레인, 하네스 막스 하프케, 콜 하워드. (2020). 자연어처리 인 액션, (류광, 역). 파주: (주)제이펍. (원본 출판 2019)

OpenAPI - 공공데이터포털. <https://www.data.go.kr>

2. 국외문헌

- Acharya, R. (2018). Evolution of DART technology and development of fourth generation bouy system for deep ocean assessment and reporting of tsunami – (DART 4G). *Indian Journal of Geo Marine Sciences*. 47(3), 519–521.
- Brown, T. B, et al. (2020). Language Models are Few-Shot Learners, arXiv preprint arXiv:2005.14165.
- Celikyilmaz, A., Clark, C., Gao, J. (2021). Evaluation of Text Generation: A Survey. arXiv:2006.14799v2
- Chopra, S., Auli, M., Rush, A. M. (2016). Abstractive sentence summarization with attentive recurrent neural networks. *Proceedings of NAACL-HLT16*. 93–98.
- Common Alerting Protocol. (2021, Jan 6). FEMA.
<https://www.fema.gov/emergency-managers/practitioners/integrated-public-alert-warning-system/technology-developers/common-alerting-protocol>
- Coombs, James H., Renear, Allen H., DeRose, Steven J. (1987). Markup system and the future of scholarly text processing. *Communications of the ACM*, 30(11), 933–947.
- Devlin, J., Chang, M., Lee, K., Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.
- DeVries, Phoebe M. R., et al. (2018). Deep learning of aftershock patterns following large earthquakes. *Nature* 560, 632–634.
- FCC CFR Title 47 Part 11. (2022, Nov 02). Emergency Alert System.

<https://www.ecfr.gov/current/title-47/chapter-I/subchapter-A/part-11>

Gu, J., et al. (2016). Incorporating copying mechanism in sequence-to-sequence learning. arXiv preprint arXiv:1603.06393.

Integrated Public Alert & Warning System. FEMA. (2022, Oct 25).

<https://www.fema.gov/emergency-managers/practitioners/integrated-public-alert-warning-system>

Jiang, Yue. (2020). Web Scraping Intro to Data Science.

<https://www2.stat.duke.edu/courses/Spring20/sta199.002/slides/web-scraping.html#4>

Jones, K. Sparck. (1999). Automatic summarizing: factors and directions. Advances in automatic text summarization. 1-12.

Kedzie, C., McKeown, K., Hal Daume III. (2018). Content Selection in Deep Learning Models of Summarization. 1818-1828.

Korea Information and Communication Technology Association (TTA). (2015). Terrestrial digital multimedia Broadcast (DMB): Disaster Alert Service. Information and Communication Organization Standard TTAK.KO-07.046/R6, Korea.

Lin, C. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. In Text Summarization Branches Out, p74-81.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv.

Madaio, Michael A. (2018). Predictive Modeling of Building Fire Risk: Designing and Evaluating Predictive Models of Fire Risk to Prioritize Property Fire Inspections. Carnegie Mellon University,

- Mallett, D., Elding, J., Nascimento, M. A. (2004). Information-Content Based Sentence Extraction for Text Summarization. IEEE International Conference on Information Technology: Coding and Computing, 2, 214-218
- McKeown, K., Robin, J., and Kukich, K. (1995). Generating Concise Natural Language Summaries. Information Processing & Management, 31(5), 703-733.
- Mihalcea, R. & Tarau. P. (2004). TextRank: Bringing Order into Text. Association for Computational Linguistics. 404-411.
- MIT Technology Review, (2016). How the Bot-y Politic Influenced This Election,
<https://www.technologyreview.com/s/602817/how-the-bot-y-politic-influenced-this-election>
- National Disaster Management Research Institute, Disaster Response and Safety Support Technology Development for Vulnerable Groups. (2020). NDMI-PR-2020-02-01-01.
- New York Times, automated Pro-Trump Bots Overwhelmed Pro-Clinton Messages, Researchers Say. (2016).
<https://www.nytimes.com/2016/11/18/technology/automated-pro-trump-bots-overwhelmed-pro-clinton-messages-researchers-say.html>
- Page, L., Brin, S., Motwani, R., Winograd, T. (1998). The PageRank Citation Ranking: Bringing Order to the Web. Stanford University.
- Papineni, K., Roukos, S., Ward, T., Zhu, W.-J. (2002). BLEU: A method

- for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL), 311–318.
- Pennypacker, C. R., Jakubowski, M. K., Kelly, M., Lampton, M., Schmidt, C., Stephens, S., and Tripp, R., (2013). FUEGO.– Fire Urgency Estimator in Geosynchronous. Remote Sens, 5(10), 5173–5192.
- Petak, William J. (1985). Emergency Management : A Challenge for Public Administration. Public Administration Review, Vol. 45, 3–7.
- Radev, D., Jing, H., Budzikowska, M. (2004). Centroid-Based Summarization of Multiple Documents, Information Processing & Management, Vol.40, 919~938.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P. J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research, 21(2020), 1–67.
- Ryu, T. & Kim, S. (2020). Redundancy operation method for a distributed public address system. Journal of the Korea Convergence Society, 11(8), 47–53.
- Sendra, S., García, L., Lloret, J., Bosch, I., and Vega-Rodríguez, R., (2020). LoRaWAN Network for Fire Monitoring in Rural Environments, Electronics, 9(3), 531.
- The Verge, (2015). AP's 'robot journalists' are writing their own stories now, <http://www.theverge.com/2015/1/29/7939067/ap-journalism-automation-robots-financial-reporting>

- UNDRR. (2020). The human cost of disasters: an overview of the last 20 years (2000–2019). Centre for Research on the Epidemiology of Disasters UNDRR.
<https://www.undrr.org/publication/human-cost-disasters-overview-last-20-years-2000-2019>
- UNISDR(United Nations Office for Disaster Risk Reduction). (2017).
<http://www.unisdr.orgw>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems, 6000–6010.
- Web Scraping With Python: Everything You Need To Know. (2022).
<https://www.scrapingbee.com/blog/web-scraping-101-with-python/>
- Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., Artzi, Y. (2019). BERTScore: Evaluating Text Generation with BERT. arXiv:1904.09675.
- Zhao, W., Peyrard, M., Liu, F., Gao, Y., Meyer, C. M., Eger, S. (2019). MoverScore: Text Generation Evaluating with Contextualized Embeddings and Earth Mover Distance. arXiv:1909.02622.

ABSTRACT

A Study on AI Alert Broadcasting Method for Disaster Preparedness by Situation

Ryu, Tae-Ha

Major in Smart Convergence Product

Dept. of Smart Convergence Consulting

The Graduate School

Hansung University

Modern society has achieved the development of civilization based on the development of science and technology. However, due to adverse effects such as increase in complexity and environmental destruction, the risk of various disasters is increasing, and the types of disasters are becoming very diverse, resulting in new disasters that did not exist before. Under the current law, disasters are largely divided into natural disasters caused by natural activities such as typhoons, floods, and heavy rains, and social disasters caused by the development of human science and civilization. However, both natural disasters and social disasters are disasters caused by the development of civilization. In the case of natural disasters, experts have diagnosed that they are the result of failing to

reduce greenhouse gas emissions while adapting without actively responding to climate change. Social disasters also occur along with the development of human civilization, and new social disasters continue to arise. In addition to COVID-19, livestock epidemics, and fine dust issues, disasters that are not legally defined, such as the recent large-scale stampede in Itaewon, continue to emerge. These disasters need to be controlled, and various activities to minimize damage caused by disasters are called disaster management. Disasters can be classified into four stages: the pre-disaster prevention stage, which is prevention and preparation, and the post-disaster management stage, which is response and recovery. Here, prevention and preparation are activities to avoid potential risks through preemptive preventive activities, and response and recovery are steps to minimize the damage situation that has occurred or to normalize victims and communities. These 4 stages of disaster can be viewed as activities that are linked to each other without being distinguished independently. Among various methods to respond to such disasters, this study proposed a method for disaster preparedness by adding a method for preparation to the public address system, one of the disaster response systems. A public address system is an information broadcasting system that transmits sound information in one direction to an unspecified number of people, and is also called premises broadcasting equipment. It is a device that operates as background music or information broadcasting transmission equipment in normal times, but is converted into an emergency broadcasting system and transmits emergency broadcasting in case of fire or emergency.

The fundamental purpose of this study is to propose an economical and effective disaster preparedness method. To achieve that goal, public data provided by the government's public data portal and disaster-related news from the disaster portal are retrieved through web crawling to

improve the delivery of information. The unnecessary content and rhetoric included in the sentence are removed, and the long sentence is summarized into a short sentence that is easy for listeners to understand, and the speech is reproduced and transmitted through the TTS installed in the public address system. In a specific way, the data brought from the public data portal can be used as broadcast content while changing the corresponding numerical value or status value through a standardized sentence. However, in the case of disaster news, sentences are summarized using a deep learning-based or statistical-based natural language summary model. In general, research is conducted to improve performance by developing a language model that shows good performance or fine-tuning the language model. However, in this paper, documents summarized through various models with good performance are evaluated in real time through the system, and the summary with the best evaluation is recommended for broadcasting. It is a method of recommending summaries that received excellent reviews for broadcasting. Various language models for natural language processing may or may not receive good evaluations depending on the case. The thing is that the output of a good model can't always get a good evaluation. In order to evaluate these results, there must be a good reference (reference or answer document). For this, if a person makes a reference directly, real-time performance is poor and labor costs are high. As an alternative to this, the summary sentence summarized by each language model is used as a reference for evaluation and the result is obtained through mutual evaluation. What was confirmed while experimenting using the ROUGE and BLEU scores as evaluation indicators is that there is no model that always shows excellent results. Therefore, in this study, which produces real-time recommendation through evaluation after real-time document summarization, it is possible to obtain a sentence summarized

by the best language model regardless of any original sentence. Through this, it can be used as disaster preparedness broadcasting content that people can concentrate on and understand easily.

If the 'method of receiving real-time disaster data and summarizing and alerting various disaster news such as flood, drought, and fine dust' proposed in this study is adopted as a function of public address system, broadcasting using summarized sentences is expected to be used as a good tool to avoid disasters.

【keyword】 Public Data, Disaster Management, Artificial Intelligence, Natural Language Processing, ROUGE Score