

320068-02

보안 과제(), 일반 과제(O) / 공개(O), 비공개() 발간등록번호(O)
가축질병대응기술개발사업 2021년도 최종보고서

발간등록번호

11-1543000-004111-01

재난형 동물질병 정보 시스템 구축
인공지능(AI)을 이용한

인공지능(AI)을 이용한 재난형 동물질병 정보 시스템 구축

2022. 7. 14.

2021

주관연구개발기관 / (주)노스스타컨설팅
협동연구개발기관 / (주)이지팜
협동연구개발기관 / 농림축산검역본부

농림축산식품부
농림식품기술기획평가원

농림축산식품부
(전문기관)농림식품기술기획평가원

제 출 문

농림축산식품부 장관 귀하

본 보고서를 "인공지능(AI)을 이용한 재난형 동물질병 정보 시스템 구축"(개발기간 : 2020. 4. 29 ~ 2021. 12. 31)과제의 최종보고서로 제출합니다.

2022. 7. 14.

주관연구개발기관명 : (주)노스스타컨설팅 (대표자) 이도원 (인)
협동연구개발기관명 : (주)이지팜 (대표자) 진교문 (인)
협동연구개발기관명 : 농림축산검역본부 (대표자) 박창신 (인)

주관연구책임자 : 박창신
협동연구책임자 : 김종현
협동연구책임자 : 윤하정



국가연구개발사업의 관리 등에 관한 규정 제18조에 따라 보고서 열람에 동의합니다.

< 요약 문 >

사업명		가축질병대응기술개발			총괄연구개발 식별번호 (해당 시 작성)			
내역사업명 (해당 시 작성)		사회 문제 해결형			연구개발과제번호		320068-2	
기술 분류	국가과학기술 표준분류	1순위 LB0710	70%	2순위 EE0108	30%	3순위 소분류 코드명	%	
	농림식품 과학기술분류	1순위 RB0201	70%	2순위 RB0203	30%	3순위 소분류 코드명	%	
총괄연구개발명 (해당 시 작성)								
연구개발과제명		인공지능(AI)을 이용한 재난형 동물질병 정보 시스템 구축						
전체 연구개발기간		2020. 04. 29 - 2021. 12 31(1년 9개월)						
총 연구개발비		총 778,000 천원 (정부지원연구개발비: 583,000 천원, 기관부담연구개발비 : 195,000 천원, 지방자치단체: 천원, 그 외 지원금: 천원)						
연구개발단계		기초[] 응용[] 개발[<input checked="" type="checkbox"/>] 기타(위 3가지에 해당되지 않는 경우)[]			기술성숙도 (해당 시 기재)		착수시점 기준() 종료시점 목표()	
연구개발과제 유형 (해당 시 작성)								
연구개발과제 특성 (해당 시 작성)								
연구개발 목표 및 내용		최종 목표		국가 재난형 동물질병에 대한 역학·방역 데이터를 통합하고 관리·분석하여 정보를 도출하는 인공지능 기반의 정보 시스템 구축				
		전체 내용		1) 국내·외 재난형 동물질병 관련 역학·방역 정보 수집 시스템 개발 - 콘텐츠 분류를 위한 동물질병 용어사전 개발 - 국내·외 방역당국의 발표자료 및 언론 동향 크롤링 - 역학조사보고서·백서·공문 및 발간물 수집·저장·관리 - 수집된 콘텐츠에 대한 파싱 기술 개발 2) 재난형 동물질병 관련 인공지능 콘텐츠 요약 및 기계독해 모델 개발 - 재난형 동물질병 콘텐츠에 대한 인공지능 요약모델 개발 - 재난형 동물질병 콘텐츠에 대한 인공지능 기계독해 모델 개발 3) 인공지능 기반 재난형 동물질병 정보 검색 시스템 개발 - 재난형 동물질병 정보 검색용 빅데이터 플랫폼 구축 - 인공지능 기반 재난형 동물질병 정보 검색 시스템 개발				
		1단계 (해당 시 작성)	목표	1) 재난형 동물질병 관련 정보 수집 시스템 개발 2) 재난형 동물질병 콘텐츠 요약모델 개발 3) 재난형 동물질병 정보 검색용 빅데이터 플랫폼 구축				
	내용	1) 재난형 동물질병 관련 정보 수집 시스템 개발 - 콘텐츠 분류를 위한 동물질병 용어사전 개발 - 국내·외 방역당국의 발표자료, 학술자료 및 언론 동향 등 크롤링 대상 사이트 선정						

			<ul style="list-style-type: none"> - 비 크롤링 역학조사분석보고서·백서·공문 및 발간물 선정 - 재난형 동물질병 콘텐츠 크롤링 서비스 개발 - 역학조사보고서·백서·공문 및 발간물 등록 서비스 개발 - 수집된 콘텐츠에 대한 파싱 기술 개발 <p>2) 재난형 동물질병 콘텐츠 요약모델 개발</p> <ul style="list-style-type: none"> - 재난형 동물질병 콘텐츠에 대한 인공지능 요약모델 개발 <p>3) 재난형 동물질병 정보 검색용 빅데이터 구축</p> <ul style="list-style-type: none"> - 콘텐츠 활용도를 높이기 위한 정보의 DB화 및 인덱싱 기술 개발 - 콘텐츠 활용을 위한 서비스(API) 기술 개발 - 정보 검색 시스템 프로토타입 개발
	2단계	목표	<p>1) 재난형 동물질병 관련 정보 수집 시스템 개발 고도화</p> <p>2) 재난형 동물질병 콘텐츠에 대한 인공지능 기계독해 모델 개발</p> <p>3) 인공지능 기반 재난형 동물질병 정보 검색 시스템 개발</p>
		내용	<p>1) 재난형 동물질병 관련 정보 수집 시스템 개발 고도화</p> <ul style="list-style-type: none"> - 동물질병 전파 모듈 개발 - 재난형 동물질병 콘텐츠 요약모델 업그레이드 <p>2) 재난형 동물질병 콘텐츠에 대한 인공지능 기계독해 모델 개발</p> <ul style="list-style-type: none"> - 재난형 동물질병 콘텐츠에 대한 기계독해 학습을 위한 학습데이터 개발 - 재난형 동물질병 콘텐츠에 대한 기계독해 모델 개발 <p>3) 인공지능 기반 재난형 동물질병 정보 검색 시스템 개발</p> <ul style="list-style-type: none"> - 사용자 편의성을 제공하는 UI/UX 기술 개발 - 사용자의 요청에 부합하는 정보를 제공하는 검색 기술 개발

연구개발성과	<p>1) 국내·외 재난형 동물질병 관련 역학·방역 정보 수집 시스템 개발</p> <p>2) 재난형 동물질병 관련 인공지능 콘텐츠 요약모델 개발</p> <p>3) 재난형 동물질병 관련 인공지능 기계독해 모델 개발</p> <p>4) 인공지능 기반 재난형 동물질병 정보 검색 시스템 개발</p>
--------	--

연구개발성과 활용계획 및 기대 효과	<p>1) 활용계획</p> <p>○ 연구개발성과의 활용분야 및 활용방안</p> <ul style="list-style-type: none"> - 수집·관리되는 역학조사서·백서·공문·발간물 등의 동물질병 비정형 콘텐츠에 대한 정보 분석에 활용 - 재난형 동물질병 역학·방역 콘텐츠를 활용한 인공지능 연구를 지원 - 재난형 동물질병 관련 인자들을 조회·분석하여 역학조사 및 방역 대책 등에 활용 - 위험요인을 다방면에서 조사·분석하여 효과적인 방역조치에 활용 - 콘텐츠의 범위를 추가·확장하여 국가방역체계 구축에 활용 - 주변국의 동물질병 발생정보 분석을 통한 경보체계로 활용 <p>○ 추가연구의 필요성</p> <ul style="list-style-type: none"> - 다양한 사용자의 검색 패턴에 따른 적합한 콘텐츠를 제공하기 위한 ‘기계독해 모델’ 고도화 수행 (주관기관 (주)노스스타컨설팅)
---------------------	--

- ▶ 검색 결과의 사용자 피드백(개선사항) 수집·분석 및 개선
 - ▶ 동물질병 질의 및 응답 데이터 추가 생성 및 기계독해 모델 고도화
 - 재난형 동물질병 콘텐츠 요약모델 고도화를 위한 ‘재난형 동물질병 말뭉치 연구 (협동기관 농림축산검역본부)’를 수행
 - 구축된 시스템을 시범 사용 및 개선점 도출(협동기관 농림축산검역본부)
 - ▶ 구축된 시스템의 효용성 평가 및 시스템 개선 사항 도출
 - ▶ 재난형 동물질병 콘텐츠 요약 서비스 고도화를 위한 ‘재난형 동물질병 말뭉치’ 개발
 - 사업화 추진방안
 - 주관기관 (주)노스스타컨설팅과 협동기관 (주)이지팜은 상호신뢰를 기반으로 공동사업을 추진
 - ▶ 구축된 ‘인공지능을 이용한 재난형 동물질병 정보 시스템’ 시범운영
 - ▶ 양 기관의 영업 네트워크를 활용하여 공동영업 추진
 - ▶ 인공지능협회, 축산협회 등이 주관하는 박람회 참여
 - ▶ 지속적인 시스템 고도화
 - 타 연구에 활용
 - 특화된 비정형 콘텐츠에 대한 수집·가공·저장 및 검색 연구에 활용
- 2) 기대효과
- 기술적 측면: 재난형 동물질병과 관련된 다양한 콘텐츠를 수집·가공·저장하는 빅데이터 플랫폼을 구축하고 인공지능을 활용하여 사용자의 요구에 적합한 콘텐츠를 검색·제공함으로써 동물질병에 대한 역학·방역 연구 발전에 기여
 - 경제·산업적 측면: 재난형 동물질병에 빠른 역학·방역 대응으로 방역 비용 절감 및 축산농가의 경제적 손실 최소화 기대
 - 사회적 측면 : 효율적인 질병발생위험 경감 및 신속한 대응으로 대국민 신뢰감 고취 및 불편 최소화

연구개발성과의 비공개여부 및 사유

연구개발성과의 등록·기타 건수	논문	특허	보고서 원문	연구 시설·장비	기술 요약 정보	소프트웨어	표준	생명자원		화합물	신품종	
								생명 정보	생물 자원		정보	실물
		1				3	1					
연구시설·장비 종합정보시스템 등록 현황	구입 기관	연구시설·장비명	규격 (모델명)	수량	구입 연월일	구입가격 (천원)	구입처 (전화)	비고 (설치장소)	ZEUS 등록번호			
국문핵심어 (5개 이내)	국가 재난형 동물질병		인공지능 기반 정보시스템		빅데이터		기계독해		정보전파로봇			
영문핵심어 (5개 이내)	Disaster-type animal disease		AI-based information system		Big Data		MRC		Information propagation robot			

< 목 차 >

1. 연구개발과제의 개요	12
가. 연구개발의 필요성 및 개요	12
(1) 연구개발의 필요성	12
(2) 연구개발의 개요	13
나. 기술 수준 및 시장 현황	15
(1) 국내 기술 수준 및 시장 현황	15
(2) 국외 기술 수준 및 시장 현황	20
2. 연구개발과제의 수행 과정 및 수행내용	26
가. 연구개발과제 목표	26
(1) 최종목표	26
(2) 세부목표	27
나. 1차년도 연구개발과제의 수행 과정 및 내용	28
(1) 국내외 재난형 동물질병 관련 역학·방역 정보 수집 시스템 개발	28
(가) 크롤링 수집 시스템 개발	28
(나) 재난형 동물질병 콘텐츠에 대한 인공지능 요약모델 개발	49
(2) 재난형 동물질병 정보 검색용 빅데이터 플랫폼 구축	50
(가) 콘텐츠 활용도를 높이기 위한 정보의 DB화 및 인텍싱 기술 개발	50
(나) 콘텐츠 활용을 위한 서비스(API) 기술 개발	56
(다) 정보 검색 시스템 프로토타입 개발	57
다. 2차년도 연구개발과제의 수행 과정 및 내용	62
(1) 재난형 동물질병 관련 역학·방역 정보 수집 시스템 개선	62
(가) 재난형 동물질병 콘텐츠에 대한 인공지능 요약모델 개선	62
(나) 주요 재난형 동물질병동향 전파기술 개발	65
(다) 한국어 콘텐츠 요약 개방형 API 서비스 개발	66
(2) 재난형 동물질병 콘텐츠에 대한 인공지능 기계독해 모델 개발	68
(가) 재난형 동물질병 콘텐츠에 대한 기계독해 학습 데이터 생성·검증	68
(나) 재난형 동물질병 콘텐츠에 대한 기계독해 모델 개발·검증	71
(3) 인공지능 기반 재난형 동물질병 정보 검색 시스템 개발	78
(가) 사용자에게 편의성을 제공하는 UX/UI 기술 개발	78
(나) 사용자의 요청에 적합한 정보를 제공하는 검색 기술 개발	85
(4) 인공지능 기반 재난형 동물질병 정보 검색 시스템 검증	94
(가) 인공지능 기반 재난형 동물질병 정보 검색과 일반 웹 검색의 비교	94
(나) 인공지능 기반 재난형 동물질병 정보 검색 시스템 검증	95
3. 연구개발과제의 수행 결과 및 목표 달성 정도	97
가. 연구수행 결과	97
(1) 정성적 연구개발성과	97
(2) 정량적 연구개발성과	100
(3) 세부 정량적 연구개발성과	100
(4) 계획하지 않은 성과 및 관련 분야 기여사항	101
나. 목표 달성 수준	101

4. 연구개발성과 및 관련 분야에 대한 기여 정도	104
5. 연구개발성과의 관리 및 활용 계획	105
참고문헌	107

< 표 차례 >

[표 1-1] 지적재산권 현황	20
[표 1-2] 원격 감시 현장 가축방역 통제시스템 개발 연구 보고서	21
[표 1-3] 경쟁기관 현황	24
[표 1-4] 지식재산권 현황	27
[표 2-1] END PRODUCT	27
[표 2-2] 세부목표	27
[표 2-3] 동물질병 사전 항목	29
[표 2-4] 크롤링 대상 사이트	32
[표 2-5] 수작업 관리대상	32

<그림 차례>

[그림 1-1] 인공지능 기반 정보 시스템 개념도	14
[그림 1-2] 기계독해 기반 질의응답 개념도	14
[그림 1-3] ELASTICSEARCH 개념도	15
[그림 1-4] 빅데이터 분석 플랫폼	16
[그림 1-5] 빅데이터 기반 동물질병 방역 의사결정 지원 체계	16
[그림 1-6] HPAI 권역화 시스템	17
[그림 1-7] 감염병 정보 수집 및 DB 구축	18
[그림 1-8] 감염병 자동수집 시스템	18
[그림 1-9] ELASTICSEARCH를 활용한 빅데이터 플랫폼 구축 사례(1)	19
[그림 1-10] ELASTICSEARCH를 활용한 빅데이터 플랫폼 구축 사례(2)	19
[그림 1-11] 시뮬레이션 설계 화면	21
[그림 1-12] 결과 표시 방식	22
[그림 1-13] 호주 가축전염병 전파모형 예시	22
[그림 1-14] 네덜란드 가축전염병 전파모형 예시	22
[그림 1-15] COMMUNITY FLU 2.0 입력 및 결과 표현 화면	23
[그림 1-16] STEM	23
[그림 1-17] INFLUSIM 입력 화면 및 데이터 표현 그래프	23
[그림 1-18] HEALTHMAP 사이트 화면	26
[그림 2-1] 목표 시스템 구성도	28
[그림 2-2] 1차년도 연구범위	29
[그림 2-3] 동물질병 사전 데이터	30
[그림 2-4] 동물질병 용어사전 모델링	30
[그림 2-5] 감염병 목록	30
[그림 2-6] 카테고리 목록 및 관리	30
[그림 2-7] 감염병 목록	31
[그림 2-8] 감염병 관리	31
[그림 2-9] 병원체 목록 및 관리	31
[그림 2-10] 선정 문서	33
[그림 2-11] 크롤링 엔진 데이터 처리 프로세스	34
[그림 2-12] 크롤링 서비스 개발 프로세스	34
[그림 2-13] 크롤링 서비스 CI/CD 프로세스	35
[그림 2-14] 크롤링 서비스 운영 프로세스	35
[그림 2-15] 크롤링 엔진 아키텍처 구성도	36
[그림 2-16] OKD 기반의 크롤링 엔진 아키텍처 구성도	36
[그림 2-17] 크롤링 엔진의 런타임 구성도	37
[그림 2-18] 크롤링 서비스의 런타임 구성도	37
[그림 2-19] 크롤링 서비스 데이터 모델링	38
[그림 2-20] 크롤링 배치프로그램 메뉴 구성도	38
[그림 2-21] 배치프로그램 목록	39
[그림 2-22] 배치프로그램 관리	39
[그림 2-23] 배치프로그램 실행 이력	40
[그림 2-24] 영문 문서의 한국어 번역 프로세스	40
[그림 2-25] 수집 문서 텍스트 변환 및 저장	41
[그림 2-26] 수집대상 여부 확인	42

[그림 2-27] 수집문서에 대한 후처리	42
[그림 2-28] 수집문서의 키워드, 요약 및 임베딩	43
[그림 2-29] 수집문서의 저장	43
[그림 2-30] 수집문서에 대한 저장 정보	44
[그림 2-31] 이미지 문서에서 텍스트 추출	44
[그림 2-32] 수집문서 관리를 위한 테이블 목록	45
[그림 2-33] 수집문서에 대한 관리 정보	45
[그림 2-34] 수집문서에 대한 파싱 결과 예	46
[그림 2-35] 수집문서 데이터 모델링	46
[그림 2-36] 문서 및 QA관리 메뉴	47
[그림 2-37] 수작업 등록문서 목록	47
[그림 2-38] 문서등록	47
[그림 2-39] 수집된 문서 목록	48
[그림 2-40] 문서정보 및 요약정보	48
[그림 2-41] 질의응답 목록	49
[그림 2-42] 문서에 대한 질의응답 관리	49
[그림 2-43] GENSIM 라이브러리를 활용한 콘텐츠 요약의 예	50
[그림 2-44] 동물질병 관련 문서의 요약 정보 제공	50
[그림 2-45] ELASTICSEARCH 설치 구성 및 실행 화면	51
[그림 2-46] 전통적 테이블	51
[그림 2-47] 테이블 데이터에서 한 줄씩 LIKE 검색	51
[그림 2-48] 역인덱스 구조 테이블	52
[그림 2-49] ELASTICSEARCH 인덱스 구조	52
[그림 2-50] TEXT 데이터 타입에 대한 ELASTIC사의 설명 자료	52
[그림 2-51] INDEX 매핑 정보	53
[그림 2-52] NORI 형태소 분석기	53
[그림 2-53] INDEX SETTING	53
[그림 2-54] FULL TEXT 검색 결과	54
[그림 2-55] VECTOR 검색 결과	54
[그림 2-56] 결과 화면 정보 의미	54
[그림 2-57] 서비스 구성도	55
[그림 2-58] BEATS 경량 수집기 설명	55
[그림 2-59] KIBANA 시각화 소개화면	56
[그림 2-60] ELASTICSEARCH NODE 모니터링	56
[그림 2-61] KIBANA DEV TOOL 예시	56
[그림 2-62] PYTHON CLIENT 설치 가이드의 일부	57
[그림 2-63] 문서 VECTOR 임베딩 PYTHON 샘플 코드	58
[그림 2-64] 적재된 INDEX 샘플	58
[그림 2-65] 문서 VECTOR 임베딩 및 적재 구조도	58
[그림 2-66] 질의어 VECTOR 임베딩 및 검색 절차도	59
[그림 2-67] 정보 검색 시스템 프로토타입 UI에 대한 의견 수렴 과정	60
[그림 2-68] 프로토타입 '질의' 기능에 대한 UI 구현 요구사항 및 예시	60
[그림 2-69] 프로토타입 '응답' 기능에 대한 UI 요구사항 및 예시	61
[그림 2-70] 2차년도 연구범위	62
[그림 2-71] 형태소 분석기의 문장구조 분석 예시	63
[그림 2-72] py-hanspell의 실시 예	63
[그림 2-73] 단절 텍스트 연결 알고리즘 적용 실시 예	63

[그림 2-74] 개선 전 후의 결과 비교	64
[그림 2-75] 텍스트 콘텐츠 요약 처리 시스템의 실행 화면	64
[그림 2-76] SMS 보내기 API 인터페이스 정의서	65
[그림 2-77] SMS 보낸 정보 받기 API 인터페이스 정의서	65
[그림 2-78] 문자메시지 연동 서비스 아키텍처	65
[그림 2-79] LINE SMS NOTIFY 보내기 API 인터페이스 정의서	66
[그림 2-80] SNS 연동 서비스 아키텍처	66
[그림 2-81] 요약 API의 아키텍처	67
[그림 2-82] 요약 API 인터페이스 정의서	67
[그림 2-83] KUBERNETES 서버에 배포 및 운영 상태의 요약 API	67
[그림 2-84] 요약 API의 실시 예	67
[그림 2-85] 문서별 생성된 질의응답 현황	68
[그림 2-86] PDF 원문과 추출된 텍스트 내용 제공	68
[그림 2-87] 질의응답 데이터 생성(답변 및 점검)	69
[그림 2-88] 데이터 생성자 및 관리자에게 제공된 오류 목록	69
[그림 2-89] 학습데이터 오류 코드	70
[그림 2-90] 오류코드별 학습데이터 오류발생 현황	70
[그림 2-91] 일자별 학습데이터 오류발생 현황	70
[그림 2-92] 데이터베이스에 저장된 동물질병 학습 데이터	71
[그림 2-93] 데이터베이스에 저장된 동물질병 질의응답 데이터 구조	71
[그림 2-94] 동물질병 정보 한국어 질의응답 처리 시스템 구성도	71
[그림 2-95] 동물질병 정보 기계독해 API 인터페이스 정의서	72
[그림 2-96] 동물질병 정보 한국어 질의응답 처리 시스템 API 검증	72
[그림 2-97] 동물질병 질의응답 API 서비스 구성도	73
[그림 2-98] 기계독해 API 인터페이스 정의서	73
[그림 2-99] 기계독해 API 내부 프로세스	74
[그림 2-100] KUBERNETES 서버에 배포 및 운영 상태의 기계독해 api	74
[그림 2-101] 기계독해 API 실행 예시	74
[그림 2-102] ELECTRA와 BERT 모델의 성능 비교	75
[그림 2-103] 한국어에 대한 자연어 오픈소스 모델의 성능 비교	75
[그림 2-104] 기계독해 모델의 학습 및 평가에 사용된 데이터	76
[그림 2-105] BASE 모델과 SMALL 모델 학습 파라메타	76
[그림 2-106] 기계독해 학습 GUI 서버의 사양 및 병렬처리 예	76
[그림 2-107] BASE 모델과 SMALL 모델의 성능 비교	77
[그림 2-108] BASE 모델과 SMALL 모델의 EPOCH별 성능 비교	77
[그림 2-109] 기존 기계독해의 평가 지표와 개발한 지표의 예	78
[그림 2-110] 로그인 화면 기획	78
[그림 2-111] 영역 정의	79
[그림 2-112] 질의 세부 설정 기획	80
[그림 2-113] 사용자 질의에 대한 응답 콘텐츠 기획	80
[그림 2-114] 추가 사용자 질의 UI 기획	81
[그림 2-115] 응답에 대한 관련 문서 보기 기획(1)	81
[그림 2-116] 응답에 대한 관련 문서 보기 기획(2)	82
[그림 2-117] 응답 별 관련 문서 구분 TAP 기획	82
[그림 2-118] 응답에 대한 평점 남기기 기획	83
[그림 2-119] 응답에 대한 평점 보기 기획	83
[그림 2-120] 로그인 화면 디자인	84

[그림 2-121] 응답에 대한 평점 남기기 기획	84
[그림 2-122] 단계별 진행 모식도	85
[그림 2-123] 적재 흐름도	85
[그림 2-124] MRC지능형 검색	86
[그림 2-125] 검색 프로토타이핑	86
[그림 2-126] 테이블 정의서 표지	87
[그림 2-127] 테이블 정의서 제·개정이력 및 참고사항	87
[그림 2-128] 토큰 정보 테이블	87
[그림 2-129] 회원 정보 테이블	88
[그림 2-130] 질의 정보 테이블	88
[그림 2-131] 응답(답변) 정보 테이블	89
[그림 2-132] 로그인 화면 디자인	89
[그림 2-133] 메인 화면	90
[그림 2-134] 질의 입력 화면	90
[그림 2-135] 질의 응답 화면	90
[그림 2-136] 답변 유형 선택	91
[그림 2-137] 정밀 검색 메시지	91
[그림 2-138] 간편 검색 및 정밀 검색 도움말	91
[그림 2-139] 질의 응답(답변) 요약 화면	91
[그림 2-140] 계속해서 검색하기 화면	92
[그림 2-141] 응답(답변) 리스트 화면	92
[그림 2-142] 원문 보기 화면	93
[그림 2-143] 답변 문서 내 키워드 검색 화면	93
[그림 2-144] 답변 문서 내 키워드 검색 결과 화면	93
[그림 2-145] 답변 문서 리스트 화면(SCORE 정렬)	94
[그림 2-146] 답변 문서 카테고리별 리스트 화면	94
[그림 2-147] 인공지능 기반 재난형 동물질병 정보검색	95
[그림 2-148] 구글 정보검색	95

1. 연구개발과제의 개요

가. 연구개발의 필요성 및 개요

(1) 연구개발의 필요성

- 국가 재난형 동물질병에 대한 역학·방역 데이터 통합 필요성
 - 2000년 67년 만에 국내에서 구제역 발생이 확인된 후, 고병원성조류인플루엔자, 아프리카돼지열병 등 많은 종류의 재난형 동물질병 발생이 확인되었으며, 역학·방역 등 업무 수행 과정에서 막대한 양의 데이터가 생성·축적되고 있음
 - 재난형 동물질병과 관련하여, 현재는 과거 발생 데이터(역학조사분석보고서 등) 및 유관기관 (가축방역지원본부, 생산자 단체, 관련 연구기관 등)의 발간물 등을 참고하여 위험도를 분석하고 향후 질병 관리에 대한 정보를 도출하고 있으나, 이 과정에서 수일~수주의 시간이 소모되고 있음
 - 매년 특별방역기간(10월~2월/3월)을 대비하여 방역정책 재정비 등을 위하여 역학적 정보 도출 등이 요구되고 있으나, 과거 데이터 소실이 많아 유의미한 정보분석이 어려우며 데이터 관리 문제는 향후 더 심각해질 것임
 - 재난형 동물질병 예방 및 발생 시 신속한 확산 차단을 위한 대응 체계 구축하려면 국가 재난형 동물질병에 대한 역학·방역 데이터의 통합이 필요함

- 국가 재난형 동물질병에 대한 인공지능 기반 정보 검색 시스템 필요성
 - 국가 재난형 동물질병에 대한 과거 자료 중 KAHIS 가축전염병 발생현황, 역학조사분석보고서, 백서, 공문 등이 접근 가능하며, 비정형·정형 데이터들이 섞여 있음
 - 출처마다 다른 형태로 다른 장소에 보관되고 있어 자료조사도 쉽지 않으며, 분류·분석에 상당한 시간을 소모하고 있음
 - 재난형 동물질병이 지속적으로 유입·발생하는 우리나라의 현실에서 데이터에 기반을 둔 과학적 방역정책에 대한 요구가 증가하고 있으며, 관련 데이터를 모두 통합하여 적시에 맞춤형 정보를 도출할 수 있는 시스템 구축이 시급함
 - 재난형 동물질병에 관하여 필요 시점에 필요한 국내외 데이터를 쉽게 검색하여 정보를 바로 현장에 적용함으로써 방역의 효율성을 도모할 수 있는 인공지능 기반 정보시스템 구축이 필요함

- 역학·방역 데이터 처리를 위한 빅데이터 플랫폼 필요성
 - 역학조사서, 백서, 공문, 발간물 등 다양하고 복잡한 데이터를 처리할 수 있는 빠르고 유연한 데이터 플랫폼 필요
 - 빠른 응답 속도와 안정적인 운영 및 확장성 보장 필요

☞ 따라서, 재난형 동물질병(아프리카돼지열병, 구제역, 고병원성조류인플루엔자 등) 관련 역학·방역 데이터를 통합하고 관리·분석하여 정보를 도출하는 인공지능 기반 정보 검색 시스템 개발이 필요

(2) 연구개발의 개요

○ 본 연구개발은

- 1) 국내외 재난형 동물질병 관련 정보 수집 및 관리 기술 개발
 - 2) 재난형 동물질병 관련 인공지능 훈련 및 분석 기술 개발
 - 3) 인공지능 기반 재난형 동물질병 정보 검색 시스템을 개발
- 을 통해 국가 재난형 동물질병에 대한 역학·방역 데이터를 통합하고 관리·분석·정보를 도출하는 인공지능 기반의 정보 시스템 구축을 목표로 함

해결방안 1 국내외 재난형 동물질병 관련 역학·방역 정보 수집 및 관리 기술 개발

본 국내외 재난형 동물질병 관련 역학·방역 정보 수집 및 관리 기술 개발 연구에서는

- 1) 국내외 방역당국의 발표 자료와 주요 동물질병에 대한 언론 동향 수집
 - 2) 재난형 동물질병 정보 검색용 빅데이터 플랫폼 구축
- 을 수행함

해결방안 2 재난형 동물질병 관련 인공지능 훈련 및 분석 기술 개발

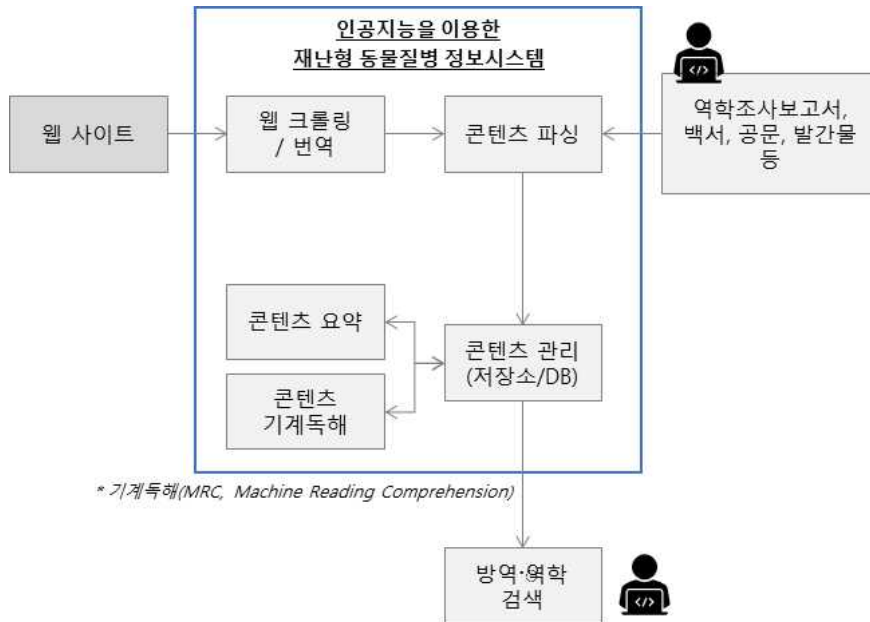
본 재난형 동물질병 관련 인공지능 훈련 및 분석 기술 개발 연구에서는

- 1) 국내외 재난형 동물질병 피해 사례에 대한 발생 요인 및 전파 요인 분석
 - 2) 인공지능 활용 콘텐츠 요약 및 기계독해(MRC, MACHINE READING COMPREHENSION) 모델 생성 기술 개발
- 을 수행함

해결방안 3 인공지능 기반 재난형 동물질병 정보 검색 시스템 개발

본 인공지능 기반 재난형 동물질병 정보 검색 시스템 개발 연구에서는

- 1) 사용자에게 편의성을 제공하는 UX/UI(사용자 인터페이스) 기술 개발
 - 2) 사용자의 요청에 적합한 정보를 제공하는 검색 기술 개발
- 을 수행함

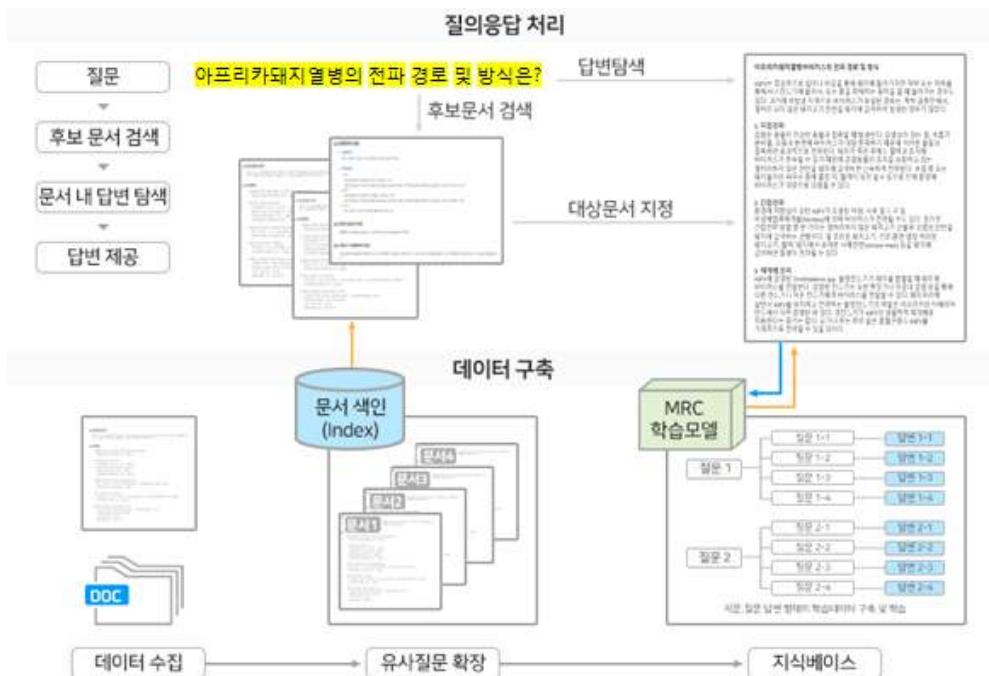


[그림 1-1] 인공지능 기반 정보 시스템 개념도

○ 본 연구에서 활용하는 기술의 개념을 살펴보면 다음과 같음

1) 인공지능 기반 정보 시스템

- 국가 재난형 동물질병에 대한 역학·방역 데이터를 통합하고 관리·분석하여 정보를 도출하고, 인공지능 기술을 활용하여 콘텐츠 요약 및 독해를 수행하여 사용자의 요청에 적합한 정보를 제공하는 정보 검색 시스템



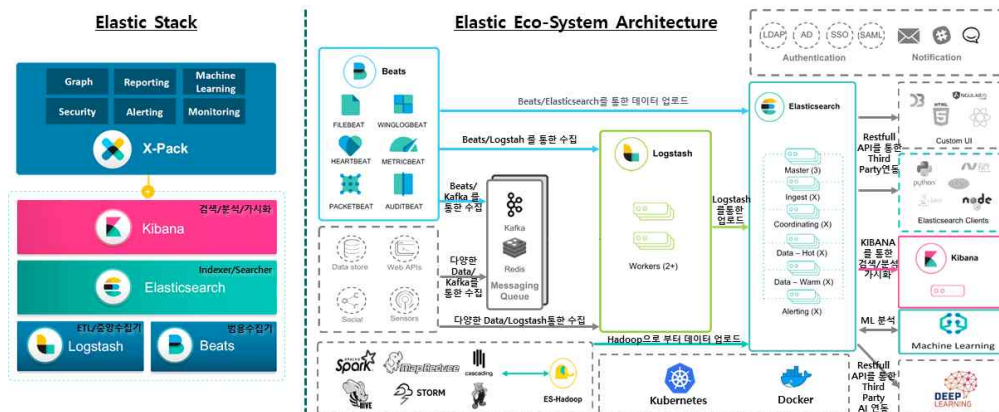
[그림 1-2] 기계독해 기반 질의응답 개념도

2) 기계독해 기반 질의응답(MACHINE READING COMPREHENSION QUESTION ANSWERING, MRCQA)

- 사람이 직접 지식 구축을 하지 않아도 기계가 문서를 읽고 질문에 대한 답변을 찾아 제시할 수 있도록 학습시키는 방식으로 대상 문서를 탐색하는 정보검색 방식과 문서에서 답변을 찾는 기계독해 방식을 결합하여 질문에 대한 답변을 제공

3) 검색 및 분석 엔진(ELASTICSEARCH)

- 텍스트, 숫자, 위치 기반 정보, 정형 및 비정형 데이터 등 모든 유형의 데이터를 위한 분산형 오픈 소스 검색 및 분석 엔진
- 다양한 환경으로부터의 정형, 반정형, 비정형 데이터에 대한 수집/정제/적재 및 서비스가 가능한 ALL-IN-ONE 빅데이터 플랫폼으로써, 필요시 다양한 ECHO-SYSTEM을 활용할 수 있는 빠르고 유연한 데이터 플랫폼



[그림 1-3] ELASTICSEARCH 개념도

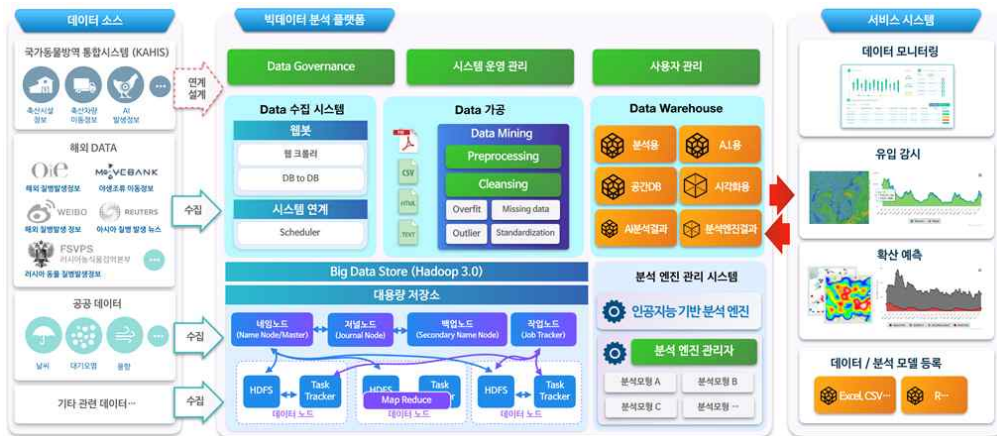
나. 기술 수준 및 시장 현황

(1) 국내 기술 수준 및 시장 현황

○ 기술현황

1 국가 재난형 동물질병 관련 역학·방역 정보 수집 및 관리 기술

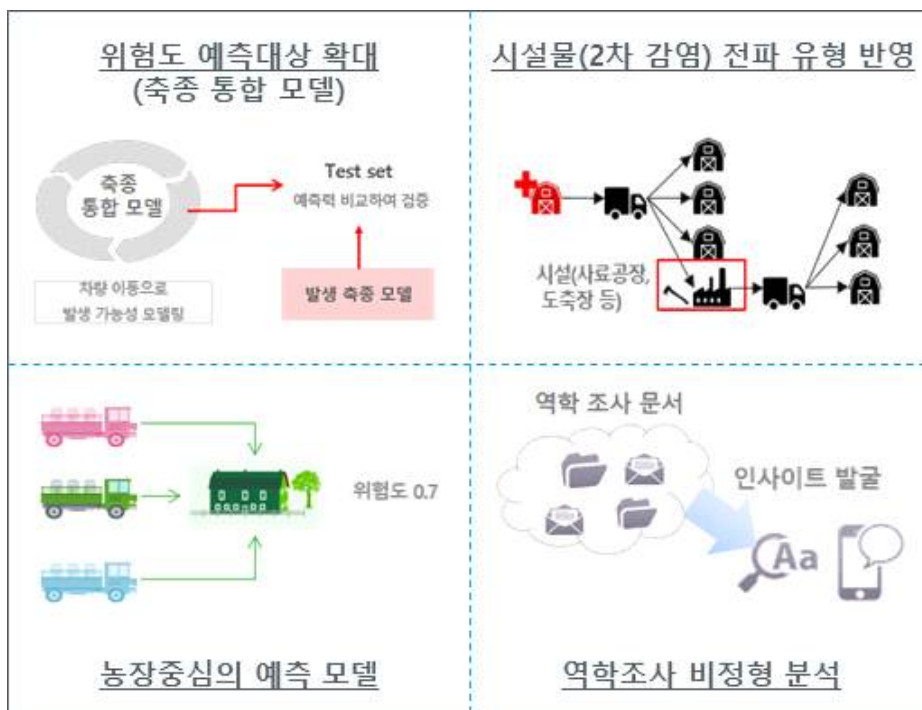
- 국가동물방역통합시스템(KOREA ANIMAL HEALTH INTEGRATED SYSTEM, KAHIS)
 - 농림축산검역본부에서는 동물질병 발생으로 인한 막대한 사회적 비용을 감소시키고자, 국내 가축 사육 및 동물질병 방역과 관련된 모든 자료를 통합하여 관리하는 국가동물방역통합시스템(KOREA ANIMAL HEALTH INTEGRATED SYSTEM, KAHIS)을 운영하고 있음
- 고병원성 조류인플루엔자 국외 데이터 수집
 - 과학기술정보통신부 연구과제 ‘인공지능기술을 활용한 고병원성 조류인플루엔자 국내 유입, 발생 조기 감지 및 확산 대응 시스템 개발’ 연구에서는 웹 크롤링 봇을 이용하여 국외 데이터를 실시간 수집하고, 수집한 데이터의 변형·교정·표준화하여 데이터 분석 목적에 맞춘 데이터베이스(DB)를 구성하고 있음. 이를 이용하여 WEB 기반 분석 지원 플랫폼을 구축하고 있음(2018~2020)



[그림 1-4] 빅데이터 분석 플랫폼

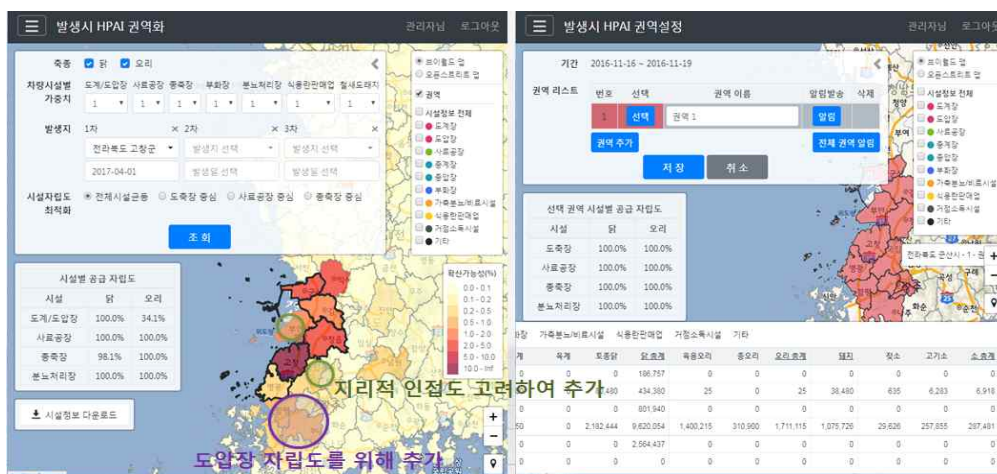
- 주관기관(주노스스타컨설팅)의 선행연구

- 미래창조과학부(현 과학기술정보통신부) 농업·ICT 창조비타민 연구과제에서 ‘재난형 동물질병 확산 대응 체계 구축’ 연구사업에 위탁기관으로 참여하여 농림축산검역본부에 ‘빅데이터 기반의 재난형 동물질병 방역조치 의사결정 지원 시스템 구축’을 수행함 (2017)
- 상기 시스템은 기계학습(MACHINE LEARNING) 등 빅데이터 분석 기법을 활용하여, 국내 축산 현황 및 관련 자료들이 포괄적으로 포함된 한국형 질병 발생위험도 평가 모델과 축산 차량 이동현황 조회, 비상 상황 시 관련자 위치 추적 모델, 수리역학 모델이 의사 결정 지원 시스템으로 활용되고 있음
- 인공지능망 모델을 활용한 분석 결과는 방역공무원이 결과를 해석하는데 기반이 되는 설명 자료가 부족하다는 한계가 존재함



[그림 1-5] 빅데이터 기반 동물질병 방역 의사결정 지원 체계

- 빅데이터 기반의 시뮬레이터 연구는 점차 확대되고 있으나, 데이터 수집 및 시뮬레이션의 개발은 해외 도구에 의존하는 상황임
 - 미래창조과학부(현 과학기술정보통신부) “한국형 감염병 확산시뮬레이션을 위한 대용량 공공데이터 활용방안” 연구에서는 오픈API를 활용하여 공공데이터를 기반으로 감염병 확산 시뮬레이션을 연구하였음. 확산 시뮬레이션 도구는 해외에서 이용되는 STEM, FLUSURGE 2.0, FLUID2.0 등을 이용함(2014)
 - 한국과학기술연구원(KIST)은 AI발생 초기, 바이러스가 어떤 패턴으로 확산되는지에 대한 분포 해석 모형과 행위자 기반 모형을 통하여 예측하는 모델을 개발하였음
 - KIST의 모델은 인구통계학적 정보 등을 활용한 분석으로, 현실적 축산정보를 활용하지 못하였고 왜 이런 결과가 나왔는지에 대한 추가적인 해석이 필요하다는 한계가 있음



[그림 1-6] HPAI 권역화 시스템

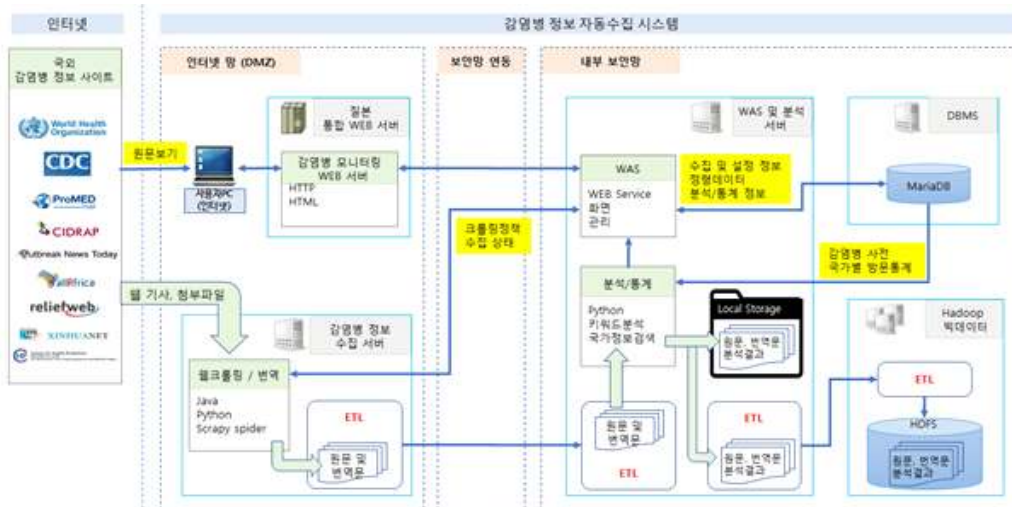
- 협동기관(주)이지팜) 선행연구
 - 농림축산식품부 “HPAI 발생 시 지역단위 권역화 방역체계 구축” 연구와 “구제역 확산 차단을 위한 권역화 방역체계 구축에 관한 연구” 연구에서 KAHIS, 축산 관련시설, 농장으로부터 제공된 자료를 권역화 관련 항목별로 정리하여 데이터베이스를 구축하고, 빅데이터 분석기법을 적용하여 핵심시설을 중심으로 클러스터 구조화하였음 (2018). 또한 농림축산검역본부 KAHIS 시스템의 농장 및 축산차량이동정보 데이터를 정제, 가공하는 시스템을 구축하여 권역 데이터를 현행화 하고 권역간 자립도와 지역간 네트워크 분석을 진행하였음(2020)
- SIMFLU(HTTP://LCBB.SNU.AC.KR/SIMFLU)
 - 인플루엔자 바이러스의 방향성 연구, 유전자 변이 패턴 예측 시뮬레이션 도구
 - 인플루엔자 바이러스는 유전자재조합으로 항원성이 다양한 변종 바이러스로 출현하기 때문에 예측이 어려워 이러한 특성을 확인하고자 2000년부터 2011년까지 세계 5만 6천 여 개의 인플루엔자 바이러스의 유전체서열들을 수집하여 시계열을 사용하여 변이패턴을 계산하는 SIMFLU시스템을 개발하였음
- 주관기관(주)노스스타컨설팅)의 선행연구
 - 방역연계범부처감염병연구개발사업단의 방역연계범부처감염병 연구에서 ‘감염병 예측을

위한 정보수집 엔진 및 국가 단위 감염병 유입위험지수 개발' 사업에 용역기관으로 참여 함 (2018)

- PROMED, GPHIN 등 영어권 37개, 비영어권 10개 사이트에서 181,833건의 감염병 정보를 취합하여 DB에 저장·관리 함 (2019.08.29.)



[그림 1-7] 감염병 정보 수집 및 DB 구축

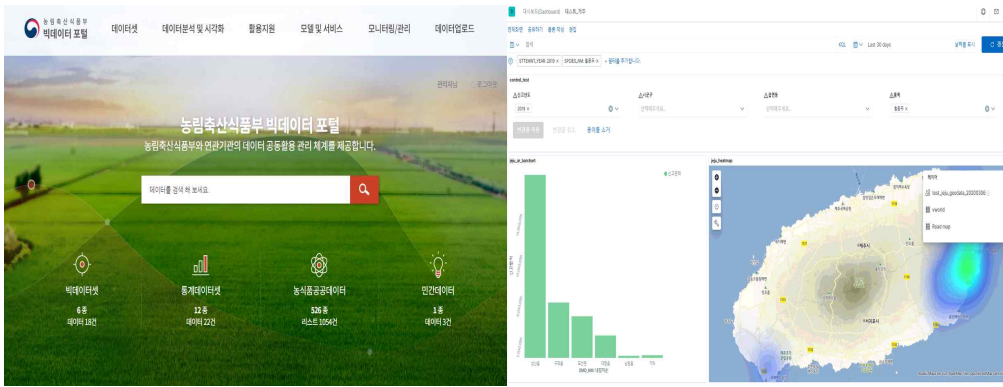


[그림 1-8] 감염병 자동수집 시스템

3 인공지능 기반 재난형 동물질병 정보 검색 시스템

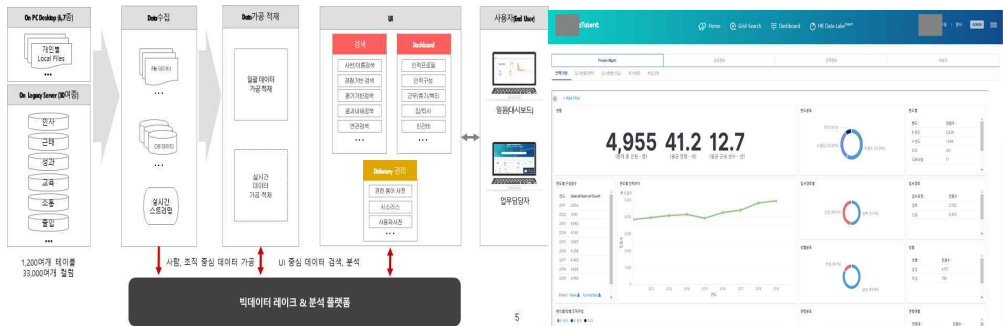
- 협동(주이지팜)의 선행연구

- 농림수산물교육문화정보원(농정원)-농식품 빅데이터 플랫폼을 엘라스틱서치 (ELASTIC-SEARCH)를 활용하여 구축하였고, 다양한 사용자를 위한 빅데이터셋을 구성하여 귀농 지원 서비스 모델, 쌀/월동작물 생산량 예측, 정책 검색 서비스 등 제공함 (2019)



[그림 1-9] ELASTICSEARCH를 활용한 빅데이터 플랫폼 구축 사례(1)

- 엘라스틱서치로 (주)SKT의 다양한 인사시스템들의 데이터를 통합하여 인사관리 빅데이터 플랫폼을 구축하여, 수집 및 검색 성능을 인정받았으며, 다양한 분석 도구 지원을 통해 인사이트 도출 및 업무 지원에 활용되고 있음(“2018)



[그림 1-10] ELASTICSEARCH를 활용한 빅데이터 플랫폼 구축 사례(2)

○ 시장현황 : 해당사항 없음

○ 경쟁기관현황 : 해당사항 없음

○ 지식재산권현황

- 동물 전염병의 확산 경로를 예측하거나, 추적하는 방법에 대한 특허는 존재하나, 이를 이용하여 타 정보시스템과 연계되어 있거나 방역 관련 시나리오에 대한 특허는 전무

고병원성 조류인플루엔자 확산 경로 예측 시스템 및 방법

<p>KEYWORD</p> <ul style="list-style-type: none"> 고병원성 조류인플루엔자 확산경로 예측 네트워크 송신부 순차패턴규칙 		<ul style="list-style-type: none"> 조명 램프 작동장치 및 감시용 카메라를 연결하여 양계장 내부에 설치, 일정 레벨의 조도 출력을 위한 타임테이블을 설정하여 통합제어 함 카메라를 통한 화재 및 조명 온/오프 여부 모니터링 함
--	--	--

가축 전염병 전파경로 추적 방법

<p>KEYWORD</p> <ul style="list-style-type: none"> 가축 전염병 방역 		<ul style="list-style-type: none"> 농가의 축산인, 축산차량, 방문자, 방문차량 등 소유한 사용자 단말기를 메인 서버에 등록 스마트폰 등과 같은 사용자 단말기로 시간대별 이동 경로 추적 전염병 발생 시 방역 원인 및 지역을 신속히 파악 및 소독
---	--	--

[표 1-1] 지적재산권 현황

(2) 국외 기술 수준 및 시장 현황

○ 기술현황

1 국가 재난형 동물질병 관련 역학·방역 정보 수집 및 관리 기술

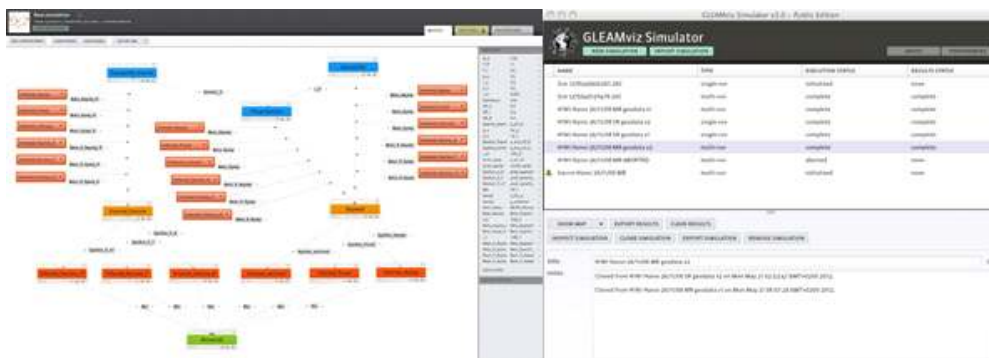
- GPS/GIS 기반 활동 감시 시스템, 실시간 차단 방역 시스템 등 가축 질병 확산방지를 위한 최신 융·복합기술 활용
 - GLEAMVIZ SIMULATOR : 범지구적 표현 범위를 가진 수리적 모델로 전 세계 항공 네트워크를 이용하여 도시·국가 간 전염병 전파 양상 표시
 - 소셜미디어 트위터를 기반으로 전 세계 조류인플루엔자 발병 현황을 크롤링하고, 이를 통하여 공식 리포트보다 더 빠르게 감지하는 시스템을 개발함 (YOUSEFINAGHANI 등, 2019)

개발기술	내용
(영국) 양 추적 GPS 기반 무리 행동 패턴 분석	양의 목에 부착된 GPS 송신기의 무선통신기술 - 양의 위치정보를 보여주는 GPS 단말기 기술 - 양 무리의 위치정보를 이용한 모델링 기술로 양의 무리 행동에 대한 패턴을 이용하여 방역 의사결정을 내릴 수 있도록 소프트웨어로 구현
(일본) 가축방역지도	이상가축 발생시 신속하게 해당 농장에 대한 상세목록 출력, 전염병 발생 의심이 되는 경우 청정성 검사지역, 이동제한구역, 반출제한구역 등 설정 - 동시에 지역 내 가축 사육농장이나 매물지, 집회장, 소독 포인트 등을 나열 등을 위한 GIS 시스템
(캐나다) BE SEE BE SAFE	농장 차단방역 및 질병 관리 시스템 - 질병 발생 시 위치 기반 정보 제공(GEO-AWARE) - 개별 농장의 실시간 차단 방역 강화 및 조치 가능 - 데이터 분석을 통한 질병 전파 예측 정보 제공 - 방문자 기록, 풍향 및 풍속 등 정보 분석
(미국) GLEAMVIZ	유행성 질병의 위험도를 분석하고 전파 모델을 개발하여 관련 정책 결정을 지원하는 데스크탑 어플리케이션 - 긴급 상황에 대한 조치계획 도출, 전염병 확산 예측, 국제적 전파에 대한 분석 등 주요 기능 탑재

[표 1-2] 원격 감시 현장 가축방역 통제시스템 개발 연구 보고서 (출처: 농림축산검역본부)

2 국가 재난형 동물질병 관련 인공지능 훈련 및 분석 기술

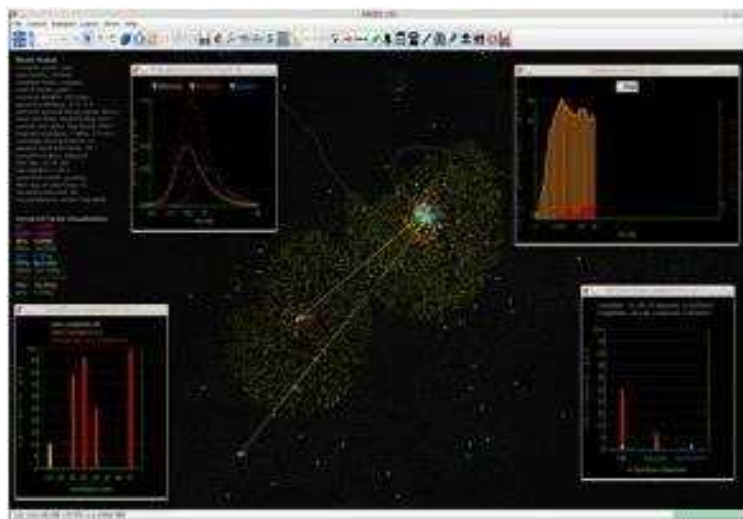
- 호주의 경우 가축전염병 위기대응 전략에 따라 시뮬레이션 기법을 활용하여 질병의 확산 양상 및 통제 전략의 효과성 및 경제성 분석을 수행하고 있으며, 확률적, 공간적 시뮬레이션 기법을 활용한 AUSTRALIAN ANIMAL DISEASE SPREAD(AADIS) 모델을 구축하여 AUSSPREAD라는 확산 모델로 발전시킴



[그림 1-11] 시뮬레이션 설계 화면

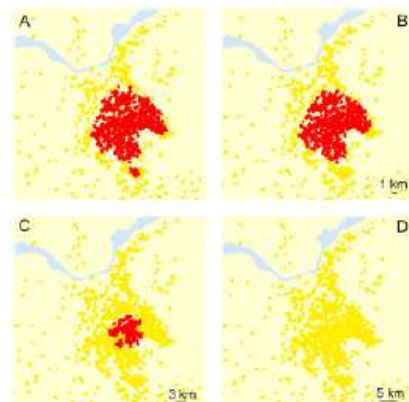


[그림 1-12] 결과 표시 방식



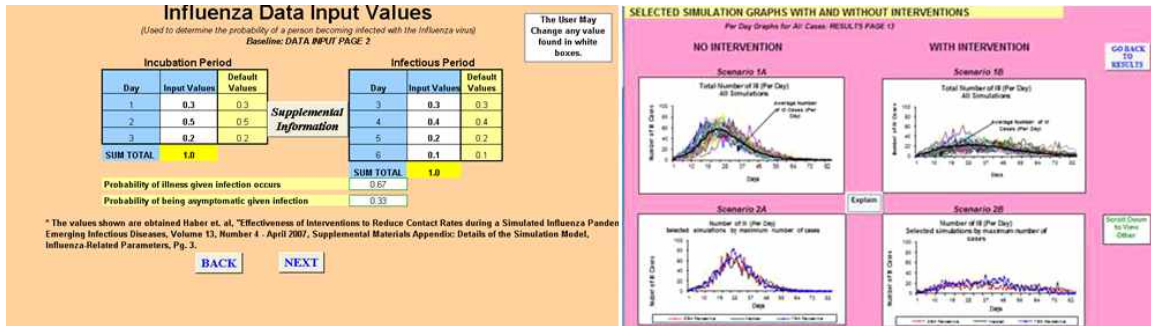
[그림 1-13] 호주 가축전염병 전파모형 예시

- BOENDERL 등(2007)은 네덜란드의 사례를 통해 조류인플루엔자, 구제역, 돼지열병 등 재난형 동물질병의 확산을 결정하는 변수에 대하여 수학적·확률적 기법을 활용한 연구에서 질병 전파와 관련하여 농장의 밀집도를 분석하고 이 모델을 활용하여 살처분 대상 농장의 지리적 범위 등에 따른 질병의 통제 효과 등을 예측 및 분석함



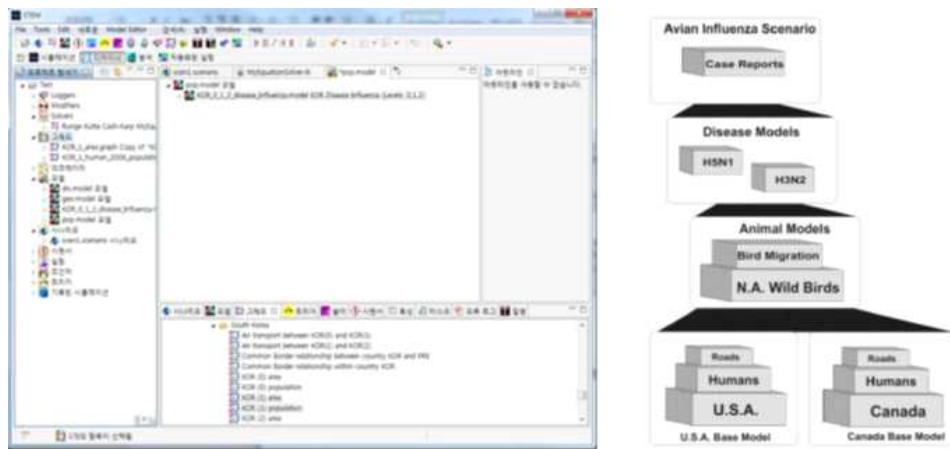
[그림 1-14] 네덜란드 가축전염병 전파모형 예시

- COMMUNITY FLU 2.0 : 미국 질병관리센터(CDC)에서 인플루엔자 확산을 예측하기 위해 개발하고, 수천세대의 가구로 구성된 표본 인구를 생성해서 수학적 모델을 적용하여 특정 커뮤니티에서 백신, 학교 휴업, 마스크 착용, 환자격리 등 다양한 변수에 따른 예측을 수행



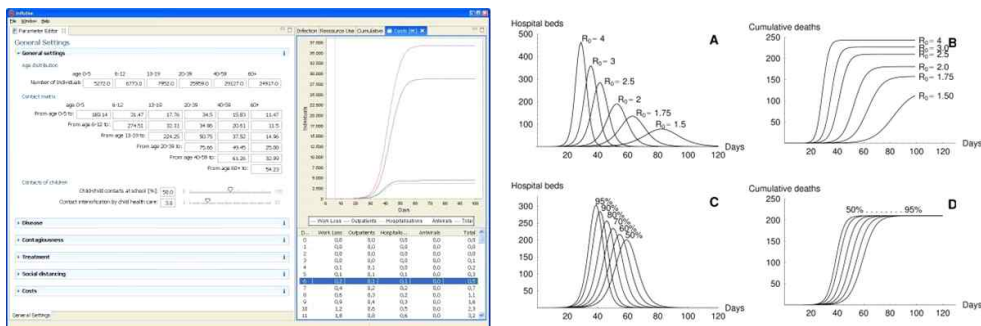
[그림 1-15] COMMUNITY FLU 2.0 입력 및 결과 표현 화면

- STEM : JAVA와 ECLIPSE 기반의 지리적 위치와 인구 집단에서의 상호작용을 통해 질병확산 위험도를 모델링할 수 있는 프로그램, 224개 국가의 지리적 정보 및 인구 데이터, 운송 시스템, 시뮬레이션 엔진, 질병 모델 등 다양한 플러그인으로 구성되어 있으며, 독립적으로 생성모델에 통합하여 사용 가능



[그림 1-16] STEM

- INFLUSIM : 인플루엔자 변수에 대한 값을 입력 받아 SEIR 모델을 수치적인 방법으로 계산하여 시간에 따른 변화를 그래프로 시각화(감염상태별로 경과 양상을 모델링)

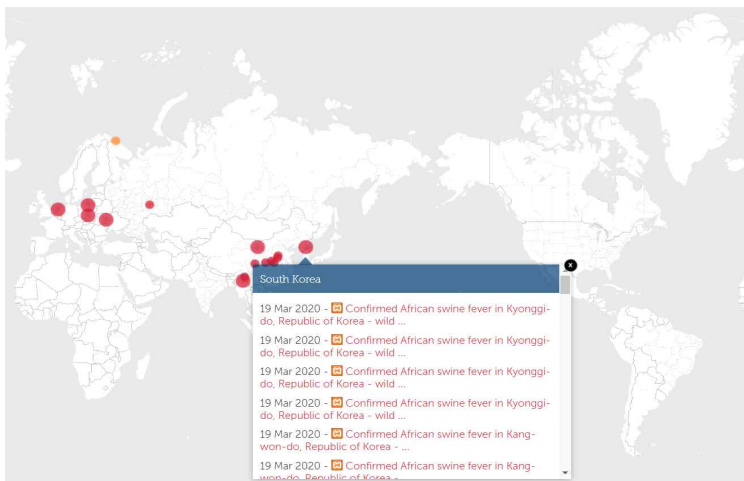


[그림 1-17] INFLUSIM 입력 화면 및 데이터 표현 그래프

3 인공지능 기반 재난형 동물질병 정보 검색 시스템

- 데이터 통합을 통한 질병 정보 시스템

- BLUEDOT : 65개의 언어로 발행되는 전 세계 뉴스, 항공권 판매 데이터, 인구통계 자료 등의 정보 등을 분석하여 바이어스의 확산을 예측
- HEALTHMAP : 구글 검색, 소셜 미디어, 블로그 게시물, 채팅 룸 데이터를 통합하여, 질병 발생 징후를 조기에 식별하고 위험도를 측정함



[그림 1-18] HEALTHMAP 사이트 화면

○ 시장 현황 : 해당사항 없음

○ 경쟁기관 현황

기술명	주요 현황	
조류인플루엔자 국내 유입 감시시스템 및 확산 대응시스템 개발	호주 AUSSPREAD	<ul style="list-style-type: none"> • 호주의 확률적, 공간적 시뮬레이션 모델로서 지역단위에서 구제역의 확산과 통제 정책 시뮬레이션
	미국, 캐나다 NAASDM	<ul style="list-style-type: none"> • 미국의 확률적·공간적·상태-전이 시뮬레이션 모델 • 전염성이 높은 질병에 대한 감수성 축군간 전파 및 방역조치 영향 시뮬레이션
	뉴질랜드 INTERSPREAD PLUS (IS+)	<ul style="list-style-type: none"> • 가축의 감염성 질병의 공간적, 확률적 시뮬레이션 모델
	영국 EXODIS-FMDTM	<ul style="list-style-type: none"> • 확률적, 공간적, 상태-전이 모델로서 축군간의 바이러스 전파를 시뮬레이션
	네덜란드 THE NETHERLANDS MODEL(NL)	<ul style="list-style-type: none"> • 네덜란드의 WAGENINGEN 대학의 수의연구소의 모델로 개체기반의 확률적, 공간적 시뮬레이션 모델 • 축군간 구제역의 전파 설명
	일본 가축방역지도	<ul style="list-style-type: none"> • GPS 기술을 이용한 위치정보 기반의 데이터를 활용하여 질병발생지 인근지역의 농장간 거리에 기초한 위험도를 평가하여 적절한 방역조치 수행

[표 1-3] 경쟁기관 현황

○ 지식재산권 현황

- 방역을 위한 시나리오 결정 특허는 존재하지만 단순한 데이터베이스에 기반 한 것으로 인공지능을 활용한 특허는 전무

DEVICE FOR SUPPORTING DECISION-MAKING FOR ANIMAL INFECTION DISEASE COUNTERMEASURE, METHOD FOR THE SAME, AND PROGRAM FOR THE SAME		
<p>KEYWORD</p> <ul style="list-style-type: none"> • ANIMAL DISEASE • SCENARIO 		<ul style="list-style-type: none"> • 동물 전염병에 대한 위험 정보를 시각화 • 조치 단계에서 의사결정을 위한 객관적인 인덱스를 제공 • 각 작업 단계에서 개체에 인덱스에 의해 결정된 것에 따라 결정된 작업 시나리오의 복수 후보를 결정 • 복수의 작업 시나리오에 따라 위험지수를 계산하여 손상 범위를 나타냄

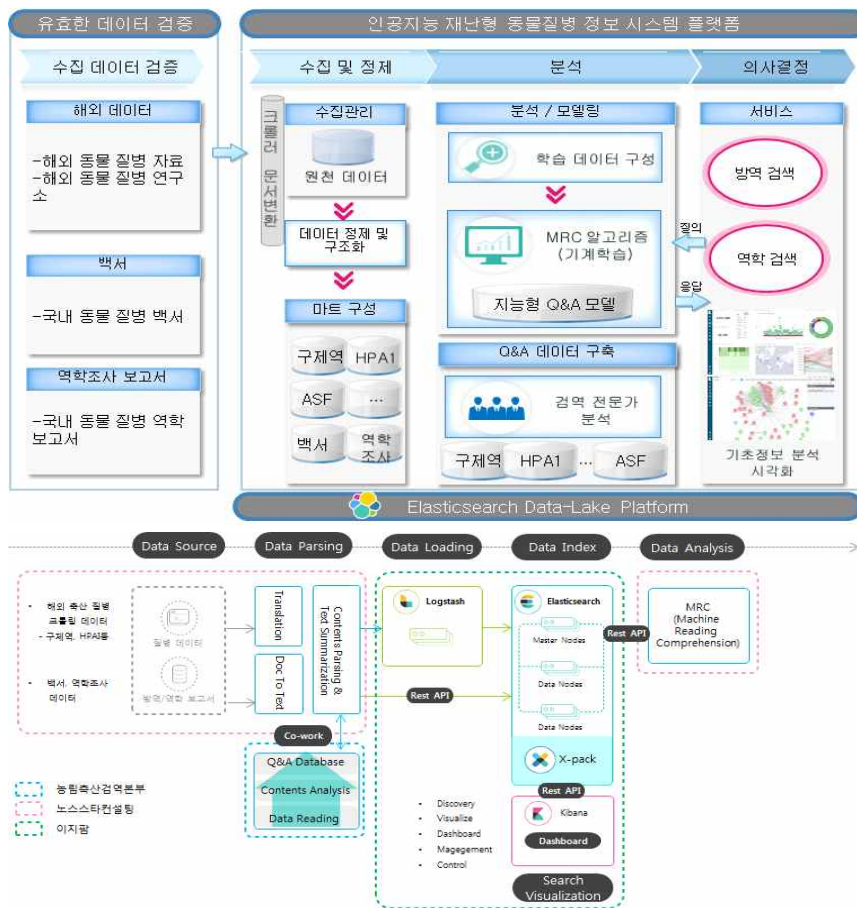
[표 1-4] 지식재산권 현황

2. 연구개발과제의 수행 과정 및 수행내용

가. 연구개발과제 목표

(1) 최종목표

- 국가 재난형 동물질병에 대한 역학·방역 데이터를 통합하고 관리·분석하여 정보를 도출하는 인공지능 기반의 정보 시스템 구축



[그림 2-1] 목표 시스템 구성도

- 본 과제는 기술 개발 목표를 중심으로 세 가지 핵심 기술 개발을 진행함
 - 1) 국내·외 재난형 동물질병 관련 역학·방역 정보 수집 시스템 개발
 - 2) 재난형 동물질병 관련 인공지능 콘텐츠 요약 및 기계독해 모델 개발
 - 3) 인공지능 기반 재난형 동물질병 정보 검색 시스템 개발

○ END PRODUCT

END PRODUCT	담당기관
<ul style="list-style-type: none"> ▪ 국내·외 재난형 동물질병 관련 역학·방역 정보 수집 시스템 	(주)노스스타컨설팅
<ul style="list-style-type: none"> ▪ 재난형 동물질병 관련 인공지능 콘텐츠 요약 모델 ▪ 재난형 동물질병 관련 인공지능 기계독해 모델 	(주)노스스타컨설팅
<ul style="list-style-type: none"> ▪ 인공지능 기반 재난형 동물질병 정보 검색 시스템 	(주)이지팜

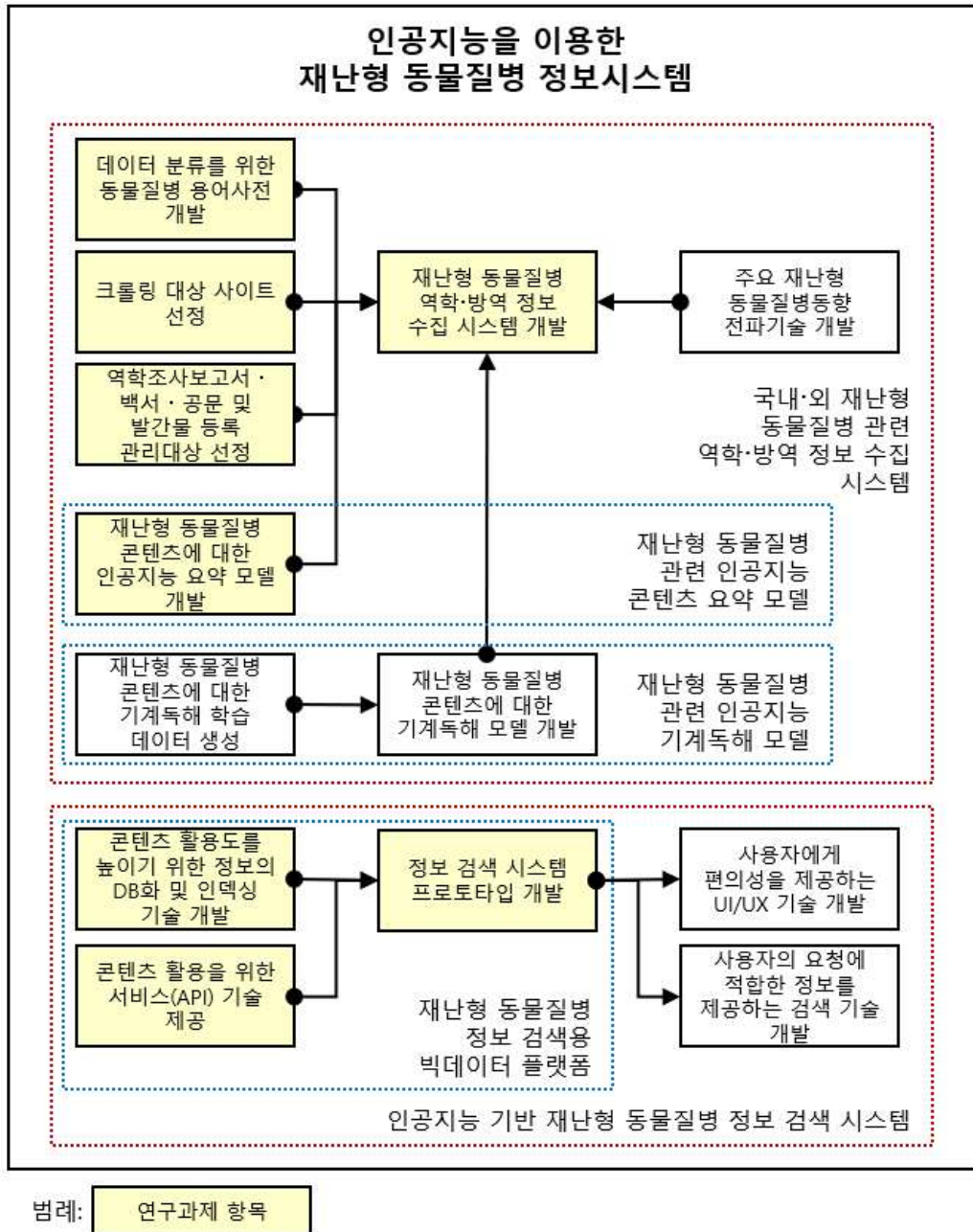
[표 2-1] END PRODUCT

(2) 세부목표

기관별 세부 목표	담당기관
1) 국내·외 웹사이트에 공개된 재난형 동물질병 정보 수집 시스템 개발 <ul style="list-style-type: none"> - 국내·외 방역당국의 발표자료, 학술자료, 언론 동향 크롤링 서비스 개발 - 역학조사보고서·백서·공문 및 발간물 등록 서비스 개발 - 수집된 콘텐츠에 대한 파싱 기술 개발 - 주요 재난형 동물질병동향 전과 기술 개발 2) 재난형 동물질병에 대한 인공지능 기술 모델 개발 <ul style="list-style-type: none"> - 재난형 동물질병 콘텐츠에 대한 인공지능 요약 모델 개발 - 재난형 동물질병 콘텐츠에 대한 인공지능 기계독해 모델 개발 	(주)노스스타컨설팅
1) 재난형 동물질병 정보 검색용 빅데이터 플랫폼 구축 <ul style="list-style-type: none"> - 콘텐츠 활용도를 높이기 위한 정보의 DB화 및 인덱싱 기술 개발 - 콘텐츠 활용을 위한 서비스(API) 기술 개발 - 정보 검색 시스템 프로토타입 개발 2) 인공지능 기반 재난형 동물질병 정보 검색 시스템 개발 <ul style="list-style-type: none"> - 사용자에게 편의성을 제공하는 UX/UI 기술 개발 - 사용자의 요청에 적합한 정보를 제공하는 검색 기술 개발 	(주)이지팜
1) 국내·외 재난형 동물질병 수집 대상 선정 및 인공지능 콘텐츠 요약 모델에 대한 검증 <ul style="list-style-type: none"> - 크롤링 대상 사이트 선정 및 데이터 분류를 위한 동물질병 용어사전 개발 - 역학조사보고서·백서·공문 및 발간물 검토·관리대상 선정·등록 - 재난형 동물질병 콘텐츠에 대한 인공지능 요약 결과에 대한 검증 2) 재난형 동물질병 콘텐츠 기계독해 검증 <ul style="list-style-type: none"> - 기계독해 학습을 위한 학습데이터 생성 - 기계독해 결과에 대한 검증 	농림축산 검역본부

[표 2-2] 세부목표

나. 1차년도 연구개발과제의 수행 과정 및 내용



[그림 2-2] 1차년도 연구범위

(1) 국내외 재난형 동물질병 관련 역학·방역 정보 수집 시스템 개발

(가) 크롤링 수집 시스템 개발

① 데이터 분류를 위한 동물질병 용어사전 개발

- 재난형 동물질병과 관련된 내용이 수집된 콘텐츠에 포함되어 있는지와 콘텐츠 분류를 위한 동물질병 용어사전을 정의함

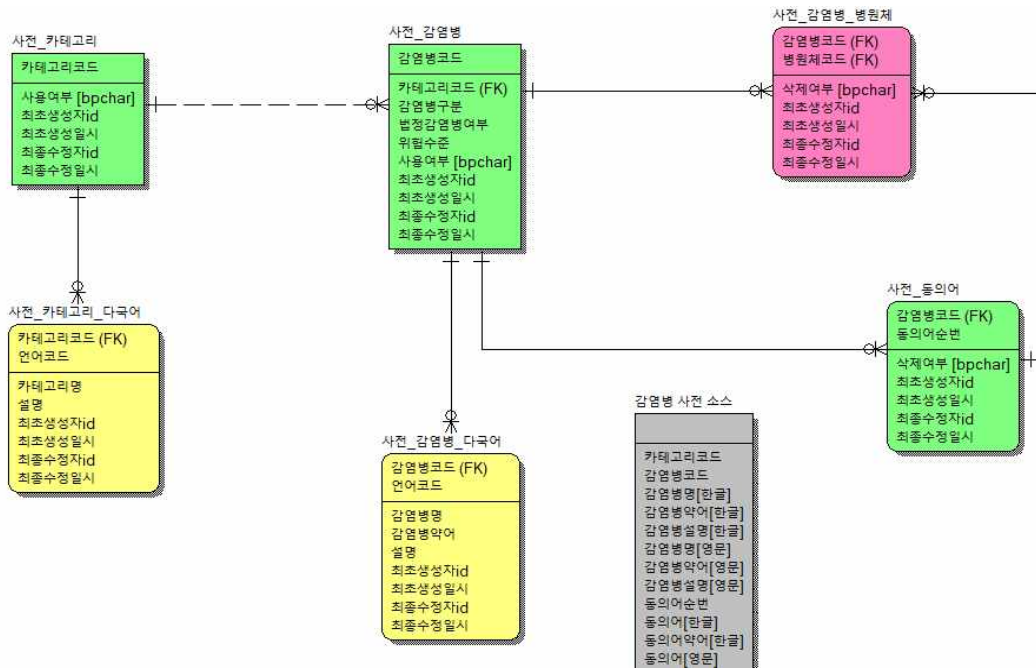
항목	설명	선정 건수
카테고리	크롤링 데이터 분류를 위한 카테고리 정의	2
감염병	관리대상인 동물질병 감염병과 속성에 대한 정의 (약어, 인수공통감염병, 법정감염병, 위험도 등)	69

[표 2-3] 동물질병 사전 항목

번호	감염병		감염병 동의어		병원체		비고
	한글명	영문명	한글약어	영문약어	한글명	영문명	
1	아프리카돼지열병	아프리카돼지열병/African swine fever; ASF/아프리카돼지열병	ASF	아프리카 돼지열병	아프리카돼지열병바이러스	African swine fever virus (ASFV)	ASFV
2	우역	우역(Rinderpest)은 감염된 소과우(우지축); 동물에서 고	Rinderpest	소 흑사병	Cattle plaque	Rinderpest virus (Morbillivirus, Paramyxoviridae)	우역코드 1115, 제1종 가축전염병
3	우역	우역(Rinderpest)은 감염된 소과우(우지축); 동물에서 고	Rinderpest	소흑사병	우역바이러스	Rinderpest virus (Morbillivirus, Paramyxoviridae)	우역코드 1115, 제1종 가축전염병
4	우역역	우역역은 마이코플라스마 mycoides subsp. mycoides small colony	CBPP		마이코플라스마	Mycoplasma mycoides subsp. Mycoides small colony	MimicC
5	구제역	구제역이란 돼지, 소, 양, 염소, 사슴 및 야생동물류 등이다	FMD	Foot and mouth disease	구제역 바이러스	Foot and Mouth Disease virus (FMDV)	FMDV
6	가성우역	가성우역은 감염된 산양 및 염양에서 구제, 구제내 목자	PPR	Peste des petits ruminants	가성우역바이러스	Peste des petits ruminants virus (Morbillivirus, Paramyxoviridae)	우역코드 0010, 제1종 가축전염병
7	블루타ング	블루타ング는 Reoviridae과, Orbivirus속에 속하는 블루타	BT		블루타ング바이러스	Bluetongue virus (Orbivirus, Reoviridae)	우역코드 0676, 제1종 가축전염병
8	리프트밸리열	리프트밸리열(Rift Valley fever)은 분자바이러스과(Rift Valley Fever)	RVF	리프트밸리열	리프트 밸리 열 바이러스	Rift Valley fever virus (Phlebotomus spread by an infected animal or mosquito (Phlebotomus, Phlebotomidae))	우역코드 0465, 제1종 가축전염병
9	umpykin병	umpykin병(Lumpy skin disease)은umpykin병 바이러스	LSD	umpykin병	umpykin병 바이러스	Lumpy skin disease virus (Capripoxvirus, Poxviridae)	우역코드 0446, 제1종 가축전염병
10	양주	양주(Scapripox)는 양과 염양의 감염을 위해 전신에 피부나 점막에	SP	Sheep Pox and Goat Pox	Clavie	Sheep pox and Goat pox virus (Capripoxvirus, Poxviridae)	우역코드 1047, 제1종 가축전염병
11	수포성구내염	수포성구내염은 황포바이러스와 백사포바이러스(vesicular stomatitis)	VS	수포성 구내염	수포성구내염바이러스	Vesicular stomatitis virus (Vesiculovirus, Rabdoviridae)	우역코드 3315, 제1종 가축전염병
12	아프리카마어	아프리카마어는 바이러스 감염에 의해 나타나는 말의 African Horse Sickness	AHS	아프리카 마어	아프리카마어바이러스	African horse sickness virus (Orbivirus, Reoviridae)	우역코드 1005, 제1종 가축전염병
13	돼지열병	돼지열병은 콜레라 바이러스(cholera virus)에 의한 Classical swine fever	CSF	돼지 열병	돼지열병바이러스	Classical swine fever virus (Pestivirus, Flaviviridae)	우역코드 0412, 제1종 가축전염병
14	돼지열병	돼지열병은 콜레라 바이러스(cholera virus)에 의한 Classical swine fever	CSF	돼지 열병	돼지열병바이러스	Classical swine fever virus (Pestivirus, Flaviviridae)	우역코드 0412, 제1종 가축전염병
15	돼지수포병	돼지수포병은 돼지 enterovirus 9형에 속하는 돼지 수포	SVD	돼지 수포병	돼지수포병바이러스	Swine vesicular disease virus (Enterovirus, Picornaviridae)	우역코드 0405, 제1종 가축전염병
16	뉴캐슬병	뉴캐슬병은 파라믹소바이러스과 뉴캐슬병바이러스에	ND	뉴캐슬병	뉴캐슬병바이러스	Newcastle disease virus (Orthomyxovirus, Paramyxoviridae)	우역코드 0333, 제1종 가축전염병
17	고병원성 조류인플루엔자	고병원성조류인플루엔자(Highly pathogenic avian influenza)	HPAI	고병원성조류인플루엔자	고병원성조류인플루엔자 바이러스	Highly pathogenic avian influenza virus (Influenza A virus)(RNA, Orthomyxoviridae)	우역코드 0111, 제1종 가축전염병
18	탄저	탄저(Bacillus anthracis)은 탄저균(Bacillus anthracis)	Anthrax		탄저균	Bacillus anthracis (Bacillus, Bacillaceae)	우역코드 1521, 제2종 가축전염병
19	기름저	기름저는 기름저균(Clostridium chauvoei)에 의해서 주	Blackleg		기름저균	Clostridium chauvoei (Clostridium, Clostridiaceae)	우역코드 0262, 제2종 가축전염병
20	브루셀라병	브루셀라병은 소의 브루셀라균의 감염으로 발생하는 인	BR	브루셀라병	브루셀라균	Brucella melitensis (Brucella, Brucellaceae)	우역코드 0663, 제2종 가축전염병

[그림 2-3] 동물질병 사전 데이터

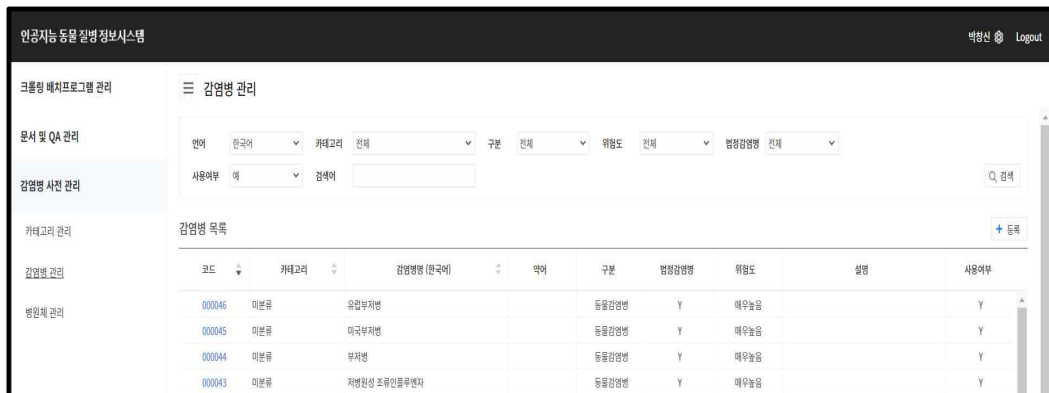
○ 동물질병 감염병에 대한 카테고리, 감염병 및 병원체를 관리하기 위한 데이터 모델 개발 및 데이터베이스(DB)화



[그림 2-4] 동물질병 용어사전 모델링

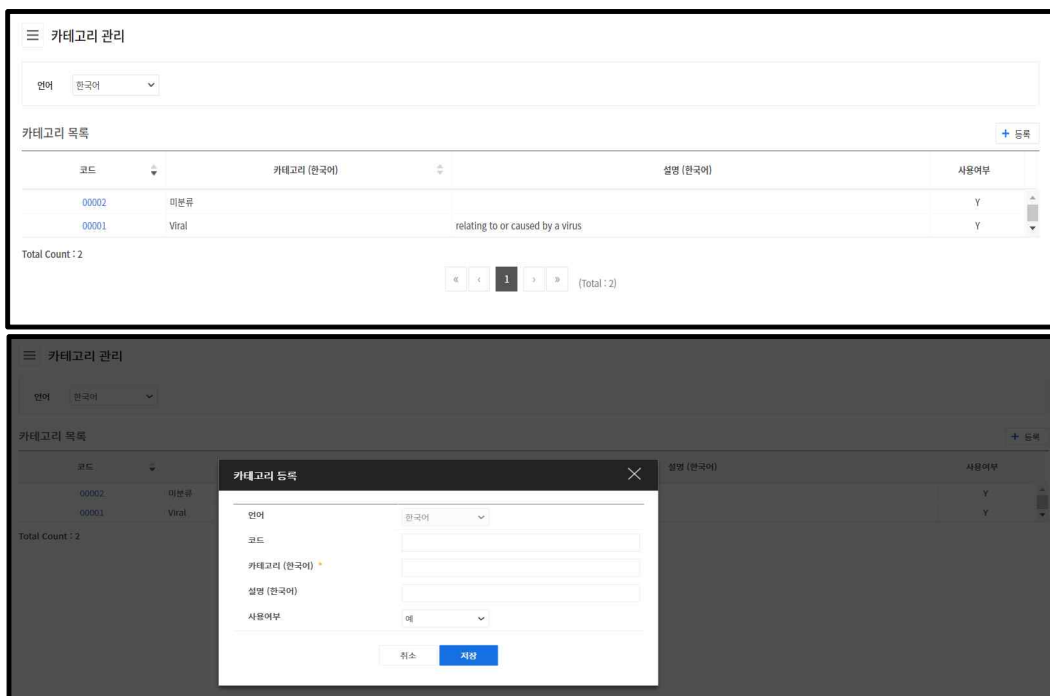
○ 동물질병 감염병 관리를 위한 카테고리, 감염병 및 병원체 목록 조회 및 등록·수정하는 ‘감염병 사전 관리’ 기능을 개발

- 감염병 사전 관리: 카테고리 관리, 감염병 관리, 병원체 관리 기능으로 구성



[그림 2-5] 감염병 목록

- 카테고리 관리: 등록된 카테고리 목록 제공 및 카테고리의 수정·등록 기능 제공



[그림 2-6] 카테고리 목록 및 관리

- 감염병 관리: 등록된 감염병 목록 제공 및 감염병의 수정·등록 기능 제공

The screenshot shows the 'Infectious Disease Management' (감염병 관리) interface. At the top, there are several dropdown menus for '언어' (Language), '카테고리' (Category), '구분' (Classification), '위험도' (Risk Level), and '병원감염병' (Hospital-acquired Disease). Below these is a search bar and a '+ 등록' (Register) button. The main area is a table titled '감염병 목록' (Infectious Disease List) with columns: 코드 (Code), 카테고리 (Category), 감염병명 (한국어) (Infectious Disease Name (Korean)), 약어 (Abbreviation), 구분 (Classification), 병원감염병 (Hospital-acquired Disease), 위험도 (Risk Level), 설명 (Description), and 사용자부 (User Department). The table contains 15 rows of data, including diseases like '유입부처병' (Imported disease), '미군부처병' (US military disease), '무저병' (Mujer disease), etc. At the bottom, it says 'Total Count : 45'.

[그림 2-7] 감염병 목록

The screenshot shows the 'Infectious Disease Management' (감염병 관리) interface for editing or viewing details. It has a '감염병 정보' (Infectious Disease Information) section with fields for '언어' (Language), '코드' (Code), '감염병명 (한국어)' (Infectious Disease Name (Korean)), '구분' (Classification), '위험도' (Risk Level), '카테고리' (Category), '약어 (한국어)' (Abbreviation (Korean)), '설명 (한국어)' (Description (Korean)), '병원감염병' (Hospital-acquired Disease), and '사용어부' (User Department). Below this are two sections: '동의어' (Synonyms) and '병원체' (Pathogen), each with a '+ 등록' (Register) button and a table for adding entries. At the bottom, it says 'Total Count : 0'.

[그림 2-8] 감염병 관리

- 병원체 관리: 등록된 병원체 목록 제공 및 병원체의 수정·등록 기능 제공

The screenshot shows the 'Pathogen Management' (병원체 관리) interface. The top part shows a search bar and a '+ 등록' (Register) button. Below is a table titled '병원체 목록' (Pathogen List) with columns: 코드 (Code), 병원체명 (한국어) (Pathogen Name (Korean)), 병원체약어 (한국어) (Pathogen Abbreviation (Korean)), and 사용자부 (User Department). The table is currently empty. Below the table, there is a modal window titled '병원체 등록' (Pathogen Registration) with fields for '언어' (Language), '코드' (Code), '병원체명 (한국어)' (Pathogen Name (Korean)), '병원체약어 (한국어)' (Pathogen Abbreviation (Korean)), and '사용어부' (User Department). At the bottom of the modal are '취소' (Cancel) and '저장' (Save) buttons.

[그림 2-9] 병원체 목록 및 관리

② 크롤링 대상 사이트 선정

- 동물질병 관련 정보 수집을 위한 언론 및 공공기관의 뉴스, 백서, 공고 사이트 선정
 - 검역본부의 방역·역학 전문가들이 업무를 수행하면서 활용하고 있는 콘텐츠를 대상으로 선정함.
 - 공공기관의 콘텐츠에 대한 분석과정을 통해서 유의미한 콘텐츠를 선별·등록함

분류	크롤링 대상	비고
공공기관	농림축산검역본부 도서관	공공기관이 보유한 콘텐츠 중에서 유의미한 콘텐츠를 선별
	농림축산검역본부	
	알림마당-보도/해명자료	
	농림축산검역본부 알림소식 보도자료	
해외언론	PROMED-MAIL	OIE-World Organization for Animal Health, WHO 정보 공유 채널
국내언론	농민신문 최신 기사	대한민국 유일한 축산계 언론 매체로 신뢰도 높음 - 수집된 콘텐츠를 대상으로 유의미성을 판정하여 선별 수집
	데일리넷 뉴스	
	돼지와사람 전체 기사	
	양돈타임즈 오늘의 뉴스	
	양돈타임즈 칼럼	
	양돈타임즈 해외양돈	
	양돈타임즈 현장	
	축산신문 전체 기사	

[표 2-4] 크롤링 대상 사이트

③ 역학조사보고서·백서·공문 및 발간물 등록 관리대상 선정

- 크롤링 서비스를 활용하여 수집할 수 없는 동물질병 관련문서(역학조사보고서·백서·공문 및 발간물)를 관리대상으로 선정

자료 유형	대상 문서	문서 수
기사	인터넷, 언론 기사	-
백서	가축전염병 백서 등	27
보고서	역학조사보고서 등	30
보도자료	기관 보도자료 등	5
법률	가축전염병 예방법 등	1
지침	가축전염병 역학조사 지침 등	10
매뉴얼	역학조사 및 방역 매뉴얼 등	12
공문	기관 공문서 등	-

[표 2-5] 수작업 관리대상

④ 재난형 동물질병 역학·방역 정보 수집 시스템 개발

- 크롤링 서비스는 사이트별로 독립적으로 개발되어 운영 관리되며, 개발 서버에서 개발 및 테스트 완료 후 운영 서버에 반영하는 이중화 시스템으로 운영 중에도 안정적으로 개발 수정이 가능한 시스템으로 구축

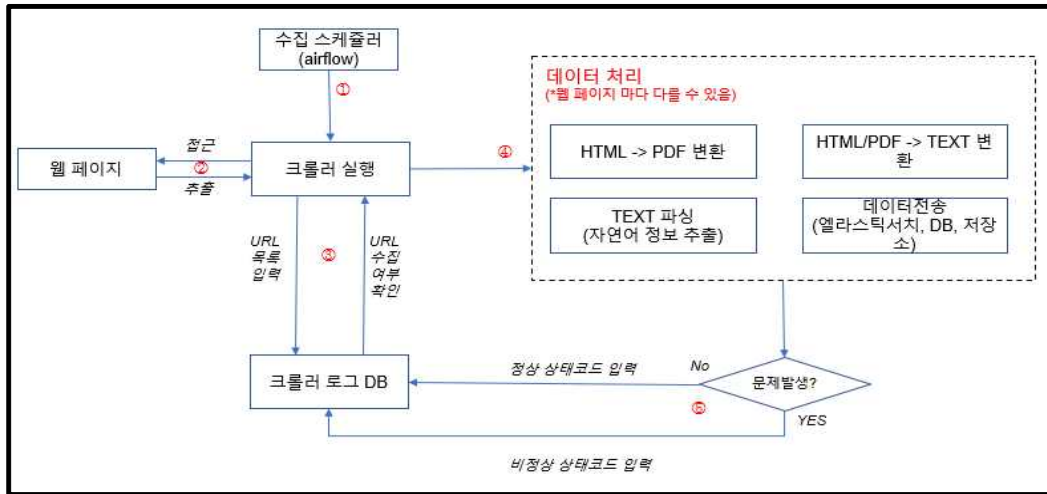
○ 크롤링 서비스 개발을 위한 엔진, 개발, CI/CD 및 운영관점의 프로세스 설계

- 크롤링 엔진 데이터 처리 프로세스

- ㉔ AIRFLOW 스케줄에 의해 콘텐츠 수집 사이트에 대한 크롤링 시작
- ㉕ 콘텐츠 수집 사이트로부터 동물질병 관련 콘텐츠 수집
- ㉖ 수집 콘텐츠에 대한 중복 수집 확인 (크롤러 로그 DB)
- ㉗ 수집 콘텐츠에 대한 전처리 후, 엘라스틱서치 및 DB에 저장
- ㉘ 수집 콘텐츠 처리 결과 저장 (크롤러 로그 DB)

v	제목	저자	발행처	발행일	페이지 수	PDF 유/무	URL	
1	o	가축 전염병 예방법	농림축산식품부	2019.07	337	유	http://www.law.go.kr/lsInFoP.do?isSeq=206651#0000	
2	2016년 가축방역사업 계획 및 실시요청	농림축산식품부	농림축산식품부	2016.01	160	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
3	2017년 가축방역사업 계획 및 실시요청	농림축산식품부	농림축산식품부	2017.01	168	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
4	2019년 가축방역 및 축산안전사업 계획 및 실시요청	농림축산식품부	농림축산식품부	2019.01	200	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
5	2020년 가축방역 사업 실시요청	농림축산식품부	농림축산식품부	2020.01	220	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
6	가축전염병 역학조사 지침	국립수의과학검역원	농림부 국립수의과학검역원	2013.11	210	유	http://ebook.qlia.go.kr/home/view.php?host=main&se=201311210	
7	o	2011년 가축전염병 역학조사 지침	농림수산검역검사본부	농림수산검역검사	2011.11	203	유	https://www.qlia.go.kr/view/webQiaCom.do?tid=9546&tyr
8	아프리카돼지열병 긴급발생지침(SOP)	농림축산식품부	농림축산식품부	2018.08	211	유	https://www.qlia.go.kr/view/webQiaCom.do?tid=46337&tyr	
9	아프리카돼지열병 긴급발생지침(SOP)	농림축산식품부	농림축산식품부	2018.07	224	유	https://www.qlia.go.kr/view/webQiaCom.do?tid=4732&tyr	
10	아성역학의 아프리카돼지열병 발생, 선행 및 차단방역	Vittorio Guberti, Sergei Khomenko	국립수의과학검역원	2020.03	78	무	http://www.nier.kr/board/article/view.do?brd_id=12	
11	o	아프리카돼지열병(ASF) 위험 재도출 여부조사	농림축산검역본부	농림축산검역본부	2019.12	529	농림축산검역본부 도서관	http://www.lib.qlia.go.kr
12	아프리카돼지열병(ASF) 위험 재도출 여부조사	농림축산검역본부	농림축산검역본부	2020.03	172	무	http://www.lib.qlia.go.kr	
13	아프리카돼지열병 예방 및 관리 지침	한국농수산식품유통공사	농림축산검역본부	2018.12	386	무	https://nrc.qlia.go.kr/SearchList.do	
14	o	조류인플루엔자 긴급발생지침	농림축산식품부	농림축산식품부	2018.12	239	유	https://nrc.qlia.go.kr/SearchList.do
15	2003~2004년 고병원성 조류인플루엔자 역학조사보고서	국립수의과학검역원	국립수의과학검역원	2004.11	386	유	https://nrc.qlia.go.kr/SearchList.do	
16	06~07 고병원성 조류인플루엔자 역학조사보고서	국립수의과학검역원	국립수의과학검역원	2007.06	387	유	http://dl.nanet.go.kr/SearchList.do	
17	08 고병원성 조류인플루엔자 역학조사보고서	국립수의과학검역원	국립수의과학검역원	2008.09	463	유	http://dl.nanet.go.kr/SearchList.do	
18	2010/2011 고병원성 조류인플루엔자 역학조사보고서	농림수산검역검사본부	농림수산검역본부	2011.12	302	유	http://dl.nanet.go.kr/SearchList.do	
19	2014~2016 고병원성 조류인플루엔자 역학조사분석보고서	농림축산검역본부	농림축산검역본부	2016.09	245	유	http://dl.nanet.go.kr/SearchList.do	
20	16~17 고병원성 조류인플루엔자 역학조사분석보고서	농림축산검역본부	농림축산검역본부	2017.12	234	유	http://dl.nanet.go.kr/SearchList.do	
21	17~18 고병원성 조류인플루엔자 역학조사분석보고서	농림축산검역본부	농림축산검역본부	2018.12	211	유	http://dl.nanet.go.kr/SearchList.do	
22	HPAI & FMD 역학조사 요청 실무매뉴얼 Handbook	농림축산검역본부 역학조사과	농림축산검역본부 역학조사과	2015	273	무	http://qlia.go.kr/view/webQiaCom.do?tid=38338&tidtype=2_3	
23	동물질병 역학조사 실무 매뉴얼(1) 고병원성조류인플루엔자	농림축산검역본부 역학조사과	농림축산검역본부 역학조사과	2015	253	유	http://qlia.go.kr/view/webQiaCom.do?tid=38338&tidtype=2_3	
24	2004 조류 인플루엔자 백신	보건복지부, 질병관리본부	보건복지부, 질병관리본부	2004.09	120	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
25	2007 고병원성 조류인플루엔자 백신	농림부, 국립수의과학검역원, 국립수의과학검역원	농림부, 국립수의과학검역원, 국립수의과학검역원	2007.11	598	유	http://dl.nanet.go.kr/SearchList.do	
26	2008 고병원성 조류인플루엔자 백신	경상북도	경상북도	2008.08	249	무	http://dl.nanet.go.kr/SearchList.do	
27	2008 고병원성 조류인플루엔자 백신	농림수산식품부, 국립수의과학검역원	농림수산식품부, 국립수의과학검역원	2008.11	880	유	http://dl.nanet.go.kr/SearchList.do	
28	2010/2011년 고병원성 조류인플루엔자 백신	농림수산식품부, 농림수산검역본부	농림수산식품부, 농림수산검역본부	2012.02	687	유	http://dl.nanet.go.kr/SearchList.do	
29	2011 고병원성 조류인플루엔자 백신	나주시	나주시	2011.10	297	무	http://dl.nanet.go.kr/SearchList.do	
30	2014년 경상북도 조류인플루엔자 구제역 백신	경상북도	경상북도	2014.01	629	유	http://dl.nanet.go.kr/SearchList.do	
31	2014 조류인플루엔자 백신	충청남도	충청남도	2014.12	555	무	http://dl.nanet.go.kr/SearchList.do	
32	(2014/15년 A백신(2014/15년 SI 발생, 확산 인원 및 재발률) 조류인플루엔자(A) 백신 중립 대책	한국농수산식품유통공사	농림축산검역본부	2015.06	614	유	http://www.priam.go.kr/homepage/entire/retrieveEfile.do	
33	2016년, SI 개발방안 및 행정확산을 위한 구제역, SI 방역대책	농림축산검역본부	농림축산검역본부	2016.02	109	무	http://dl.nanet.go.kr/SearchList.do	
34	주요국 HPAI 고병원성(조류인플루엔자) 정책조사 자료집 (1)	농림축산식품부	농림축산식품부	2017.04	570	무	http://dl.nanet.go.kr/SearchList.do	
35	주요국 HPAI 고병원성(조류인플루엔자) 정책조사 자료집 (2)	농림축산식품부	농림축산식품부	2017.04	968	무	http://dl.nanet.go.kr/SearchList.do	
36	SI, 구제역 발생방역 업무매뉴얼	농림축산검역본부	농림축산검역본부	2018.06	344	무	http://dl.nanet.go.kr/SearchList.do	
37	구제역 역학조사보고서	농림부, 국립수의과학검역원, 국립수의과학검역원	농림부, 국립수의과학검역원, 국립수의과학검역원	2001.09	316	유	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
38	2010 구제역 역학조사보고서	국립수의과학검역원 역학조사과	국립수의과학검역원 역학조사과	2010.10	278	유	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
39	2010~2011 구제역 역학조사보고서	농림수산검역검사본부 역학조사과	농림수산검역검사본부 역학조사과	2011.12	274	유	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
40	2000~2002년 구제역 역학조사분석보고서	농림수산식품부, 농림축산검역본부	농림축산검역본부	2003.02	331	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
41	2010 구제역 역학조사보고서	국립수의과학검역원 역학조사과	국립수의과학검역원 역학조사과	2010.01	278	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
42	2010~2011 구제역 역학조사분석보고서	농림수산식품부, 농림축산검역본부	농림축산검역본부	2011.12	291	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
43	2014~2015 구제역 역학조사분석보고서	농림수산식품부, 농림축산검역본부	농림축산검역본부	2015.12	174	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
44	2016~2017 구제역 역학조사분석보고서	농림수산식품부, 농림축산검역본부	농림축산검역본부	2017.12	79	유	http://www.lib.qlia.go.kr	
45	2018년 구제역 역학조사분석보고서	농림수산식품부, 농림축산검역본부	농림축산검역본부	2018.11	119	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
46	2019년 구제역 역학조사분석보고서	농림수산식품부, 농림축산검역본부	농림축산검역본부	2019.12	117	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
47	2014/2015년 구제역 Foot-and-mouth disease 역학조사분석	농림축산식품부	농림축산식품부	2015	175	유	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
48	2020 구제역 긴급통제지침 SOP	농림축산식품부	농림축산식품부	2020.05	352	유(함글파일)	http://www.qlia.go.kr/view/webQiaCom.do?tid=49743&tyr	
49	대한민국 구제역 관리방안	농림축산식품부	농림축산식품부	2010.08	121	무	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
50	동물질병 역학조사 실무 매뉴얼(1) 구제역(FMD)	농림축산검역본부	농림축산검역본부	2015	364	유	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
51	2014~2016 구제역 발생농장 역학조사 내역	가축위생연구소	가축위생연구소	2016.06	913	무	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
52	2001 구제역 백신	농림부	농림부	2001.01	393	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
53	구제역 백신	파주시	파주시	2001.12	434	무	http://www.lib.qlia.go.kr	
54	2002 구제역 백신	농림부	농림부	2003.03	472	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
55	구제역 백신	농림부	농림부	2003.02	968	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
56	2014~2016 구제역 백신	한국농촌경제연구원	한국농촌경제연구원	2016.05	499	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
57	o	2017 구제역 백신	농림축산식품부	농림축산식품부	2017.07	227	유	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c
58	2018 구제역 백신	농림축산식품부	농림축산식품부	2019.07	209	유	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
59	2019 구제역 백신	농림축산식품부	농림축산식품부	2019.07	309	유	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
60	구제역 중앙재난안전대책본부 운영 백서	행정안전부 중앙재난안전대책본부	행정안전부 재난대응정책센터	2011.12	923	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
61	2010 구제역 백신 Foot and Mouth Disease	경기도 축산위생연구소	경기도 축산위생연구소	2010.09	572	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
62	국립축산과학원 구제역 백신(2010-2011)	농촌진흥청 국립축산과학원	농촌진흥청 국립축산과학원	2011.12	273	무	http://www.lib.qlia.go.kr	
63	2010~2011 구제역 백신 Foot and Mouth Disease	경상남도	경상남도	2012.01	676	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
64	2011 GSND 구제역 방역백서 (1권)	경상남도	경상남도	2011.05	416	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
65	2011 GSND 구제역 방역백서 (2권)	경상남도	경상남도	2011.05	1129	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
66	2011 구제역 백신	경기도	경기도	2011.12	699	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
67	365월 가축위생 방역 및 중독 예방 구제역 SI 백신	안성시	안성시	2012.04	217	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
68	구제역 특별방역 대책 추진에 따른 구제역 방역백서	제주특별자치도	제주특별자치도	2011.06	584	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
69	돼지콜레라 백신	농림부, 국립수의과학검역원	농림부, 국립수의과학검역원	2005.02	650	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
70	2003년 돼지콜레라 발생에 대한 구상급 청구소송 백서	국립수의과학검역원	국립수의과학검역원	2008.10	268	유	http://dl.nanet.go.kr/SearchList.do	
71	돼지콜레라 긴급발생지침	농림부, 국립수의과학검역원	농림부, 국립수의과학검역원	2002.01	107	유	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
72	2002~2003년 돼지콜레라 역학조사보고서	농림부, 국립수의과학검역원	국립수의과학검역원	2003.12	222	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
73	제주도 돼지콜레라 백신(SI)을 통한 역학조사	농림부, 국립수의과학검역원	국립수의과학검역원	2005.12	149	유	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
74	2012년 역학조사보고서	농림수산검역검사본부 동물위생시험소	농림수산검역검사본부 동물위생시험소	2012.03	166	무	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
75	제주도 돼지콜레라 백신(SI)을 통한 역학조사	농림부, 국립수의과학검역원	국립수의과학검역원	2006.12	357	무	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
76	돼지콜레라 방역 지침서	농림부, 국립수의과학검역원	농림부, 국립수의과학검역원	1999.03	83	유	http://lib.qlia.go.kr/Search?tid=ea%b5%ac%ec%a0%9c	
77	o	아성조류(형세)에 의한 고병원성조류인플루엔자(HPAI) 발생	농림축산검역본부 역학조사과	농림축산검역본부 역학조사과	2017.05	327	유	http://lib.qlia.go.kr/SearchList.do
78	17/18 아성조류(형세)에 의한 고병원성조류인플루엔자(HPAI)	농림축산검역본부 역학조사과	농림축산검역본부 역학조사과	2018.05	138	유	http://lib.qlia.go.kr/SearchList.do	
79	18/19 아성조류(형세)에 의한 고병원성조류인플루엔자(HPAI)	농림축산검역본부 역학조사과	농림축산검역본부 역학조사과	2019.05	239	유	http://lib.qlia.go.kr/SearchList.do	
80	19~20년 아성조류(형세)에 의한 고병원성조류인플루엔자(HPAI)	농림축산검역본부 역학조사과	농림축산검역본부 역학조사과	2020.05	142	농림축산검역본부 도서관	http://www.lib.qlia.go.kr	
81	고병원성 조류인플루엔자(HPAI) 발생 예방을 위한 아성조류	농림축산검역본부	농림축산검역본부	2015.12	280	유	http://lib.qlia.go.kr/SearchList.do	

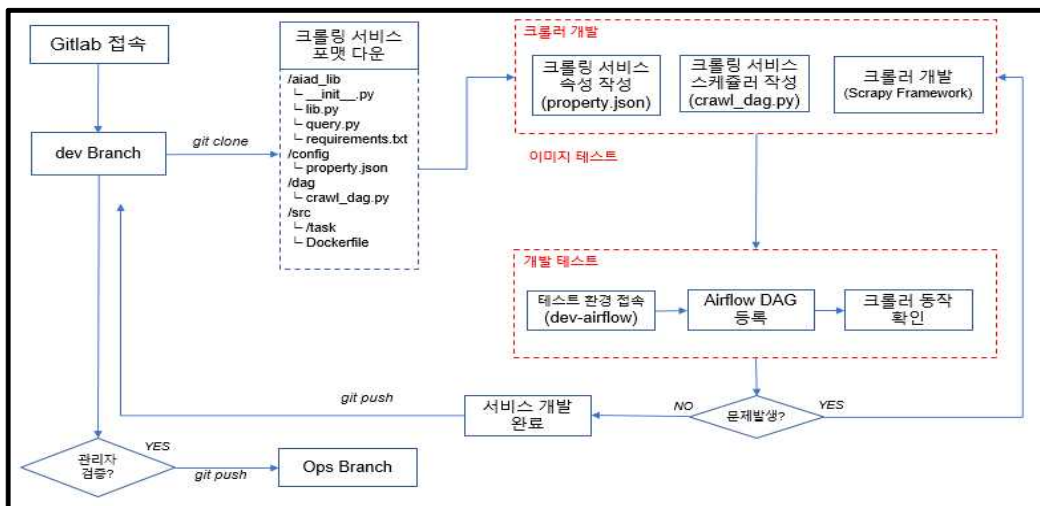
[그림 2-10] 선정 문서



[그림 2-11] 크롤링 엔진 데이터 처리 프로세스

- 크롤링 서비스 개발 프로세스

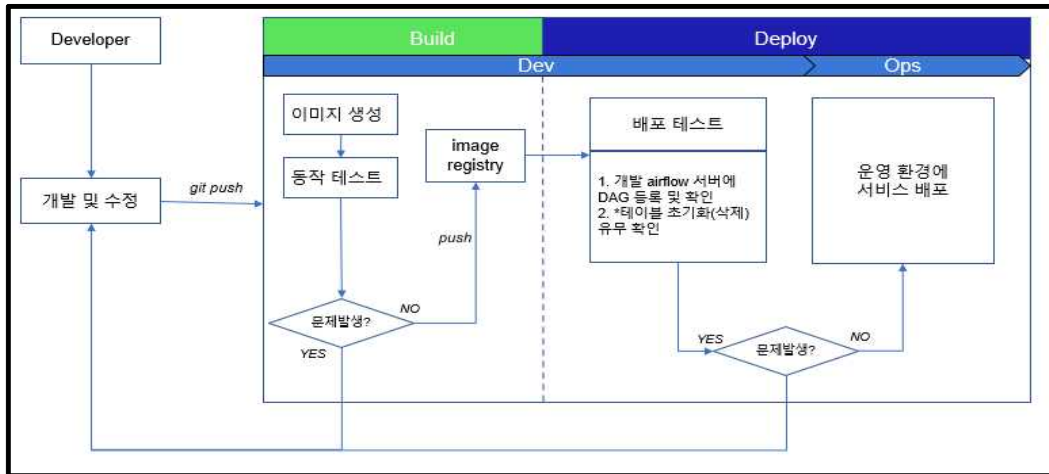
- ㉠ 신규 사이트 크롤링 개발을 위해 크롤링 서비스 템플릿 파일을 개발 환경에 복제
- ㉡ 크롤링 서비스 템플릿은 크롤링을 위한 수집, 변환, 저장의 기본 구성 포함
- ㉢ 사이트의 크롤링 대상 해당 페이지만의 특성, 수집 대상, 스케줄을 설정
- ㉣ 수집, 변환, 파싱, 추출, 저장 등의 데이터처리 워크플로 자동화
- ㉤ 크롤링 서비스 개발 완료 검증 후 운영에 반영



[그림 2-12] 크롤링 서비스 개발 프로세스

- 크롤링 서비스 CI/CD 프로세스

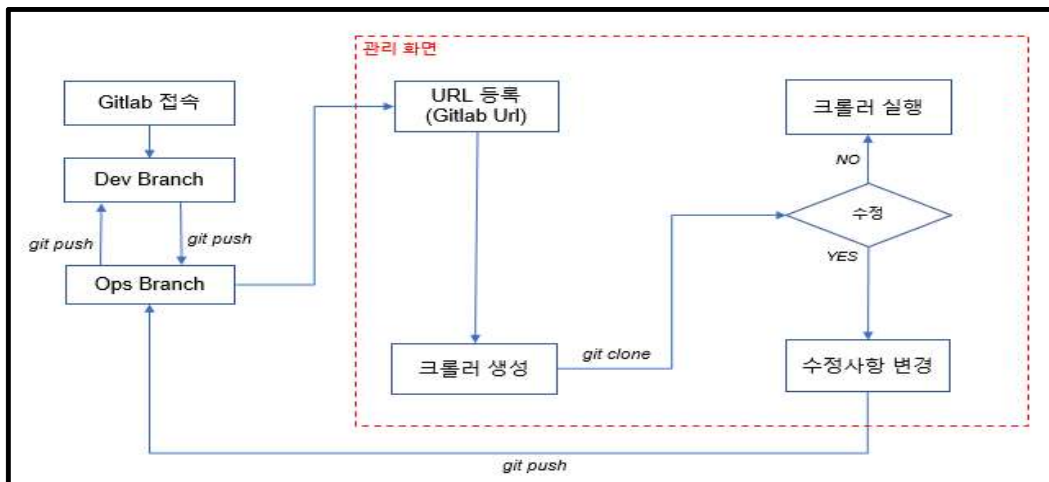
- ㉠ 크롤링 서비스는 수집, 변환, 번역, 파싱 및 정보 추출, 저장, 전송 각 기능별로 독립된 TASK로 구성되어 있으며, 각 TASK는 도커 컨테이너로 운영
- ㉡ 컨테이너로 구성된 각 TASK는 사이트별 크롤링 서비스에서 모듈 형식으로 재사용
- ㉢ 개발환경에서 작성된 크롤링서비스 모듈은 동작 검증 후 GIT REPOSITORY에 저장되고 필요한 시점에 운영 PUSH 하면 자동으로 서버에 등록 및 실행하는 배포 프로세스로 구성



[그림 2-13] 크롤링 서비스 CI/CD 프로세스

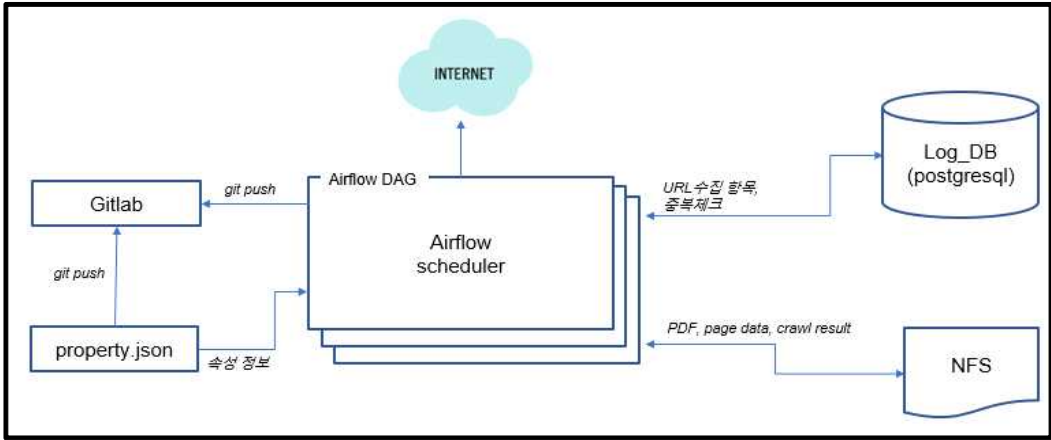
- 크롤링 서비스 운영 프로세스

- ㉠ 크롤링 서비스는 GIT CI/CD 서버, AIRFLOW WORKFLOW 서버, 관리화면 WEB 서버가 각각 독립적으로 운영
- ㉡ 실제 웹크롤링은 AIRFLOW WORKFLOW 서버에서 동작하며 스케줄링 및 로그 관리
- ㉢ 크롤링 서비스는 관리화면 WEB 서버에서 정보 조회/수정 및 실행 제어하면 GIT CI/CD 서버를 통해 실제 크롤링이 실행되는 AIRFLOW WORKFLOW 에 반영되어 서비스 수행



[그림 2-14] 크롤링 서비스 운영 프로세스

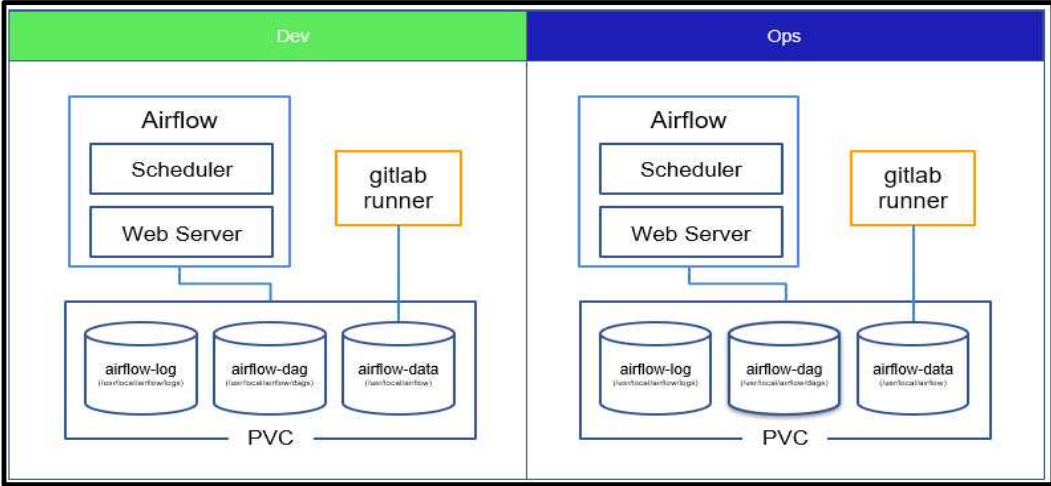
- 동물질병 관련 기사, 백서, 공고 등을 자동 수집하기 위한 크롤링 시스템의 아키텍처 설계
- ㉠ 크롤링 엔진은 GITLAB 서버, AIRFLOW 서버, POSTGRESQL 서버, NFS 서버로 구성
- ㉡ GITLAB 서버는 소스 버전 관리 및 CI/CD를 수행하며, 웹크롤링은 AIFLOW 서버에서 실행
- ㉢ AIRFLOW 는 정해진 스케줄로 웹크롤링을 수행하고 수집된 데이터는 NFS 서버에 저장
- ㉣ 웹크롤링 동작 로그 및 수집 URL 목록은 POSTGRESQL 에 저장하여 중복 수집 방지



[그림 2-15] 크롤링 엔진 아키텍처 구성도

- 크롤링 엔진의 아키텍처 구성도 (ON OKD)

- ㉠ 동물질병 관련 역학·방역 정보 수집 시스템의 모든 서버 구성은 쿠버네티스 컨테이너 운영환경인 OKD 플랫폼에서 구현
- ㉡ 웹크롤링이 실행되는 AIRFLOW 서버와 웹크롤링 SW의 CI/CD를 실행하는 GILAB은 쿠버네티스 환경의 PVC라는 저장소를 통해 크롤링 서비스가 연동
- ㉢ OKD 쿠버네티스 플랫폼을 이용해 크롤링 엔진은 동일한 구성으로 DEV(개발), OPS(운영) 운영 및 배포

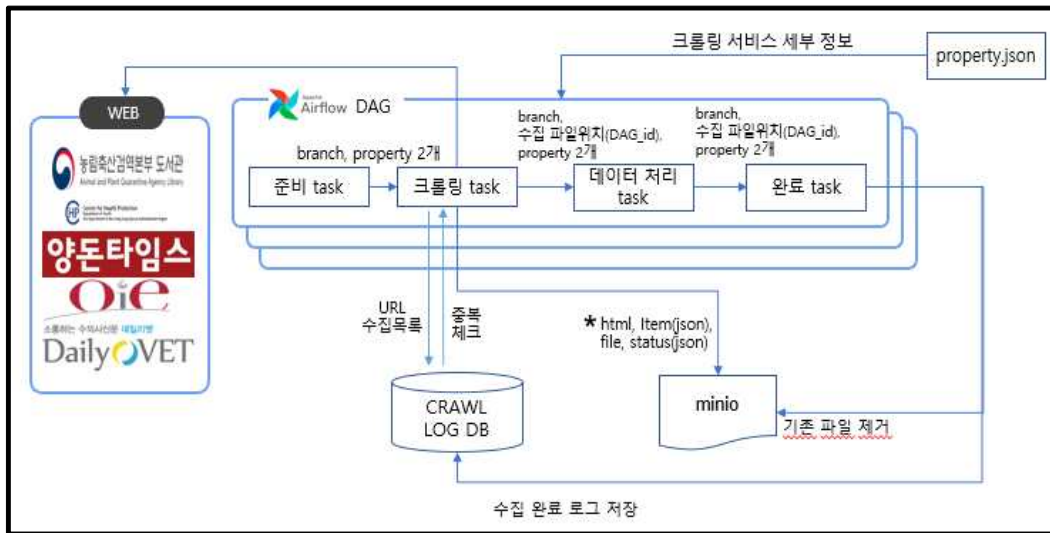


[그림 2-16] OKD 기반의 크롤링 엔진 아키텍처 구성도

- 크롤링 엔진의 런타임 구성도

- ㉠ 크롤링 엔진은 크롤링 수집 태스크 와 데이터처리(변환, 번역, 추출, 요약 등) 테스트들로 연동되는 WORKFLOW로 구성되며 AIFLOW 스케줄러에 의해 실행
- ㉡ 각 태스크들은 독립된 작업을 수행하며 언제나 재실행이 가능
- ㉢ 크롤링 엔진의 WORKFLOW는 크롤링 서비스의 설정 정보(PROPERTY.JSON)를 바탕으로 구되며 WEB SITE URL 정보 및 기사 수집 날짜 범위 등을 옵션으로 변경이 가능
- ㉣ WEB SITE의 기사 등 게시물에 대한 크롤링 로그를 DB에 저장하여 주기적으로 반복하여 최신 기사 및 과거 기사를 수집하며 중복 수집을 방지함

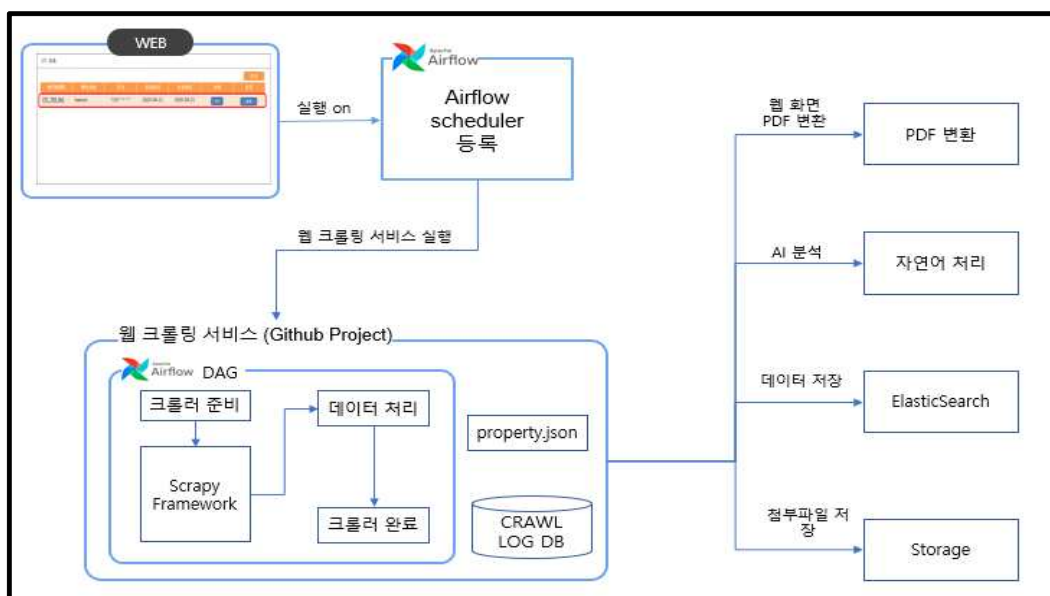
- ㉞ 모든 수집 문서, 기사와 데이터 처리를 통해 생성된 데이터는 MINIO라는 오브젝트 저장소 클러스터에 안전하게 분산 저장함



[그림 2-17] 크롤링 엔진의 런타임 구성도

- 크롤링 서비스의 런타임 구성도

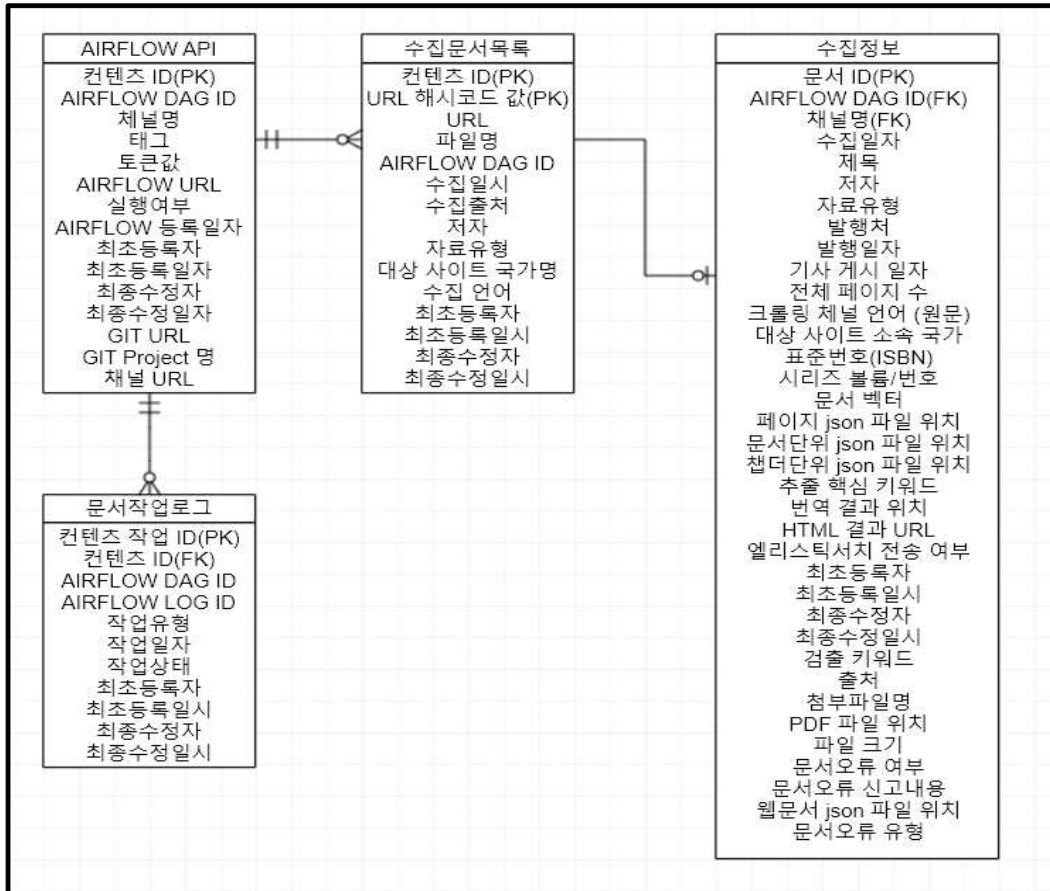
- ㉟ 크롤링 서비스는 웹관리 화면을 통해 크롤링 엔진의 동작 설정 및 제어를 하며 크롤링 엔진이 수행하는 크롤링 서비스의 결과인 수집 문서, 기사, 자연어 처리 결과는 오브젝트 스토리지에 영구 저장 및 빅데이터 저장소인 ELASTICSEARCH 에 전송 함
- ㊱ 크롤링 서비스는 AIRFLOW 서버에서 관리 운영되지만 AIRFLOW는 전문 엔지니어용 관리화면을 제공하는 관계로 일반인이 필요한 부분만 쉽게 관리할 수 있도록 WEB 화면을 별도로 제공



[그림 2-18] 크롤링 서비스의 런타임 구성도

○ GUI 관리 기반 동물질병 관련 웹사이트 크롤링 시스템 개발

- 크롤링 대상 사이트를 등록·관리하고 등록된 스케줄에 따라 콘텐츠를 조회·수집 및 전처리·저장하기 위한 GUI 관리 기반의 동물질병 관련 웹사이트 크롤링 시스템 개발
- GUI 관리 기반의 동물질병 관련 웹사이트 크롤링 시스템의 데이터 모델 설계



[그림 2-19] 크롤링 서비스 데이터 모델링

- 동물질병 감염병 콘텐츠 수집을 위한 대상 사이트를 등록·관리 및 크롤링 이력을 확인하는 기능 개발



[그림 2-20] 크롤링 배치프로그램 메뉴 구성도

채널명	태그	등록일자	크롤링 서비스 ID	실행
파일 업로드	기타	09/16/2020 09:00:00	aiad_crawler_web_upload	ON
축산신문 전체 기사	언론	01/01/2000 23:00:00	aiad_crawler_chuksannews_article	ON
양돈타임즈 현상	언론	01/01/2000 09:00:00	aiad_crawler_pigtimes_field	ON
양돈타임즈 해외양돈	언론	01/01/2000 09:00:00	aiad_crawler_pigtimes_overseas	ON
양돈타임즈 칼럼	언론	01/01/2000 09:00:00	aiad_crawler_pigtimes_column	ON
양돈타임즈 오늘의 뉴스	언론	01/01/2000 09:00:00	aiad_crawler_pigtimes_today	ON
돼지사람 전체 기사	언론	01/01/2000 09:00:00	aiad_crawler_pigpeople_article	ON
돼지사람 뉴스	언론	01/01/2000 09:00:00	aiad_crawler_dailyvet_news	ON
농민신문 최신 기사	언론	01/02/2000 06:00:00	aiad_crawler_nongmin_news	ON
농림축산검역본부 일일마당·보도/세방자료	공공기관	01/01/2000 09:00:00	aiad_crawler_qia_notice	ON
농림축산검역본부 도서관	공공기관	01/01/2000 09:00:00	aiad_crawler_qia_lib	ON
ProMed-Mail	해외	11/01/2020 11:00:00	aiad_crawler_promed_mail	OFF

[그림 2-21] 배치프로그램 목록

- 콘텐츠 수집 대상 사이트에 대한 목록 제공

- ▶ 배치프로그램은 각각의 콘텐츠 수집대상 사이트 별 크롤링서비스를 의미하며 수집 실행 제어 및 수집 실행 상태를 확인 가능
- ▶ 배치프로그램 관리는 크롤링 서비스의 URL 을 입력하면 해당 콘텐츠 수집대상 사이트의 크롤링 서비스 정보를 가져와 관리 및 제어가 가능

배치프로그램 소스 정보

채널명 * : 축산신문 전체 기사

채널 설명 * : 축산신문 전체 기사 크롤링서비스

채널 태그 * : 언론

채널 URL * : http://chuksannews.co.kr/news/article_list_all.html?sec_no=0

검색어

정부파일 수집여부 * : FALSE

1회 수집 제한 수 * : 20

크롤링 지연시간(초) * : 5

수집목록 초기화 * : FALSE

채널 국가 * : kr

채널 언어 * : ko

배치프로그램 설정 정보

배치 ID * : aiad_crawler_chuksannews_article

수집 시작날짜 * : 2000-01-01 14:00:00

배치 일정주기 * : 12

배치 일정단위 * : hours

배치 타임아웃 시간 * : 1

배치 타임아웃 단위 * : hours

실행 시 재실행 횟수 * : 1

재실행 지연시간 * : 10

재실행 지연시간 단위 * : minutes

알림 메일주소 * : minwoo.lee@northstar.co.kr

적용 실패 메일 전송 여부 * : TRUE

재실행 시 메일 전송 여부 * : FALSE

[그림 2-22] 배치프로그램 관리

- 콘텐츠 수집 대상 사이트 관리 (신규등록, 수정, 삭제)

- ▶ 크롤링 서비스의 URL을 통해 가져온 콘텐츠 수집 대상 사이트의 크롤링 서비스 정보를 확인
- ▶ 크롤링 서비스의 정보를 수정/저장하면 크롤링 엔진에 바로 반영

배치프로그램 동작 이력

ID: [] 작업ID: [] 이벤트: [v] 실행일자: [] [검색]

배치프로그램 동작 이력 목록

ID	시작시간	작업ID	세부작업ID	이벤트	실행일자	소유자	내역
60575	11/17/2020 14:34:03	aiad_crawler_plgpeople_a...	dag-task2	success	11/17/2020 14:01:37	airflow	
60574	11/17/2020 14:33:57	aiad_crawler_plgpeople_a...	dag-task2	cli_run	11/17/2020 14:01:37	root	["host_name": "aiadcrawl...
60573	11/17/2020 14:33:55	aiad_crawler_plgpeople_a...	dag-task2	running	11/17/2020 14:01:37	airflow	
60572	11/17/2020 14:33:54	aiad_crawler_plgpeople_a...	dag-task2	cli_run	11/17/2020 14:01:37	root	["host_name": "aiadcrawl...
60571	11/17/2020 14:33:40	aiad_crawler_plgpeople_a...	dag-save-task	success	11/17/2020 14:01:37	airflow	
60570	11/17/2020 14:33:31	aiad_crawler_qia_notice	dag-task2	success	11/17/2020 14:01:57	airflow	
60569	11/17/2020 14:33:27	aiad_crawler_plgpeople_a...	dag-save-task	cli_run	11/17/2020 14:01:37	root	["host_name": "aiadcrawl...
60568	11/17/2020 14:33:25	aiad_crawler_plgpeople_a...	dag-save-task	running	11/17/2020 14:01:37	airflow	
60567	11/17/2020 14:33:25	aiad_crawler_qia_notice	dag-task2	cli_run	11/17/2020 14:01:57	root	["host_name": "aiadcrawl...
60566	11/17/2020 14:33:24	aiad_crawler_plgpeople_a...	dag-save-task	cli_run	11/17/2020 14:01:37	root	["host_name": "aiadcrawl...
60565	11/17/2020 14:33:23	aiad_crawler_qia_notice	dag-task2	running	11/17/2020 14:01:57	airflow	
60564	11/17/2020 14:33:22	aiad_crawler_qia_notice	dag-task2	cli_run	11/17/2020 14:01:57	root	["host_name": "aiadcrawl...
60563	11/17/2020 14:33:14	aiad_crawler_qia_notice	dag-save-task	success	11/17/2020 14:01:57	airflow	
60562	11/17/2020 14:33:10	aiad_crawler_plgpeople_a...	dag-summarize-task	success	11/17/2020 14:01:37	airflow	
60561	11/17/2020 14:32:59	aiad_crawler_qia_notice	dag-save-task	cli_run	11/17/2020 14:01:57	root	["host_name": "aiadcrawl...

Total Count : 60,575

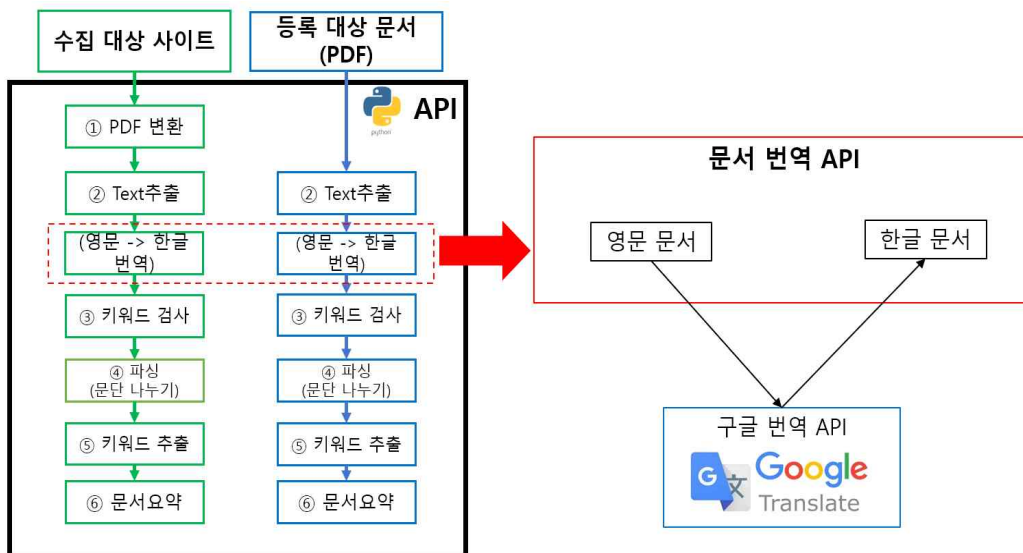
[그림 2-23] 배치프로그램 실행 이력

- 동물질병 관련 콘텐츠 크롤링 수집 실적

- ▶ 크롤링 엔진의 이벤트 로그를 통해 이상 여부를 확인하고 이상 이벤트 발생시 해당 이벤트의 링크를 통해 크롤링 서비스 서버로 연결하여 상세한 문제 내용 확인

○ 영문 수집 데이터의 한국어 번역 프로그램 개발

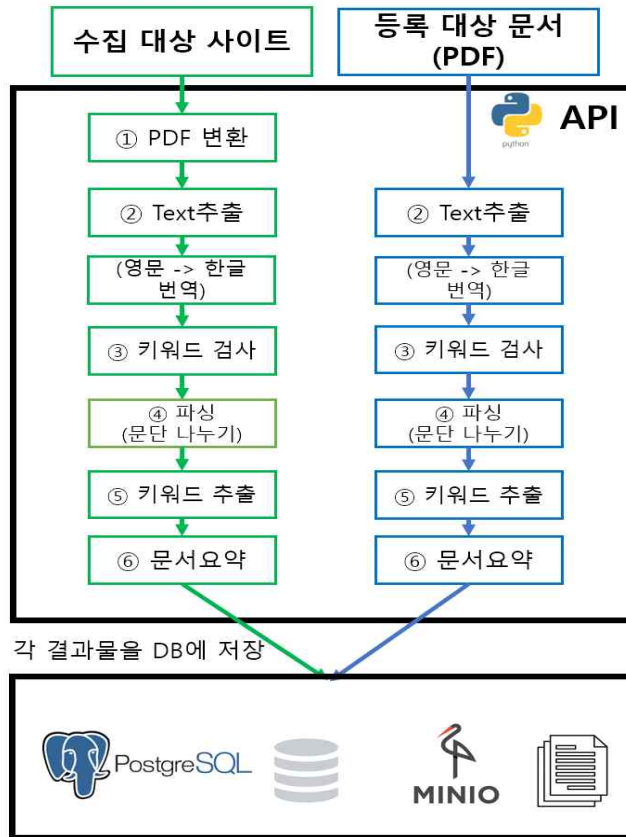
- 구글 번역 API를 적용한 SW를 개발하여 동물질병 정보 수집에서의 영문 문서에 대한 한글 문서화 및 정보 파싱



[그림 2-24] 영문 문서의 한국어 번역 프로세스

○ 백서 및 역학·방역 보고서 텍스트 변환 프로그램 개발

- 크롤링 수집 문서 및 백서, 역학방역 관련 보고서의 텍스트 변환, 키워드 추출 및 문서 요약



[그림 2-25] 수집 문서 텍스트 변환 및 저장

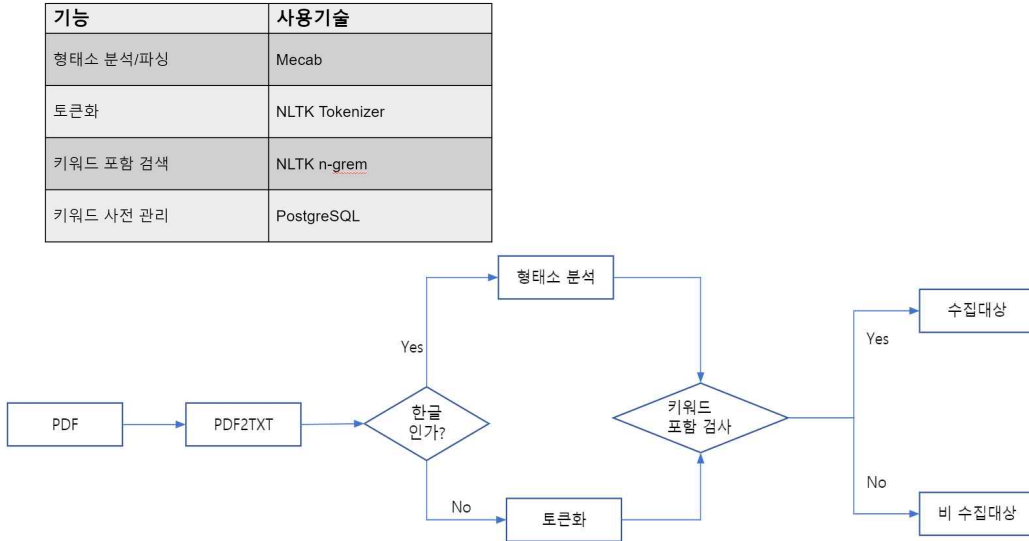
○ 수집된 콘텐츠에 대한 파싱 기술 개발

- 수집한 콘텐츠를 동물질병 용어사전 기반으로 전·후처리하여 가공이 쉽도록 DB화

㉔ 크롤링 및 사용자 등록으로 수집된 PDF 콘텐츠의 전처리 후 동물질병 용어사전 키워드 포함 여부 검사 수행으로 수집대상 여부 판단

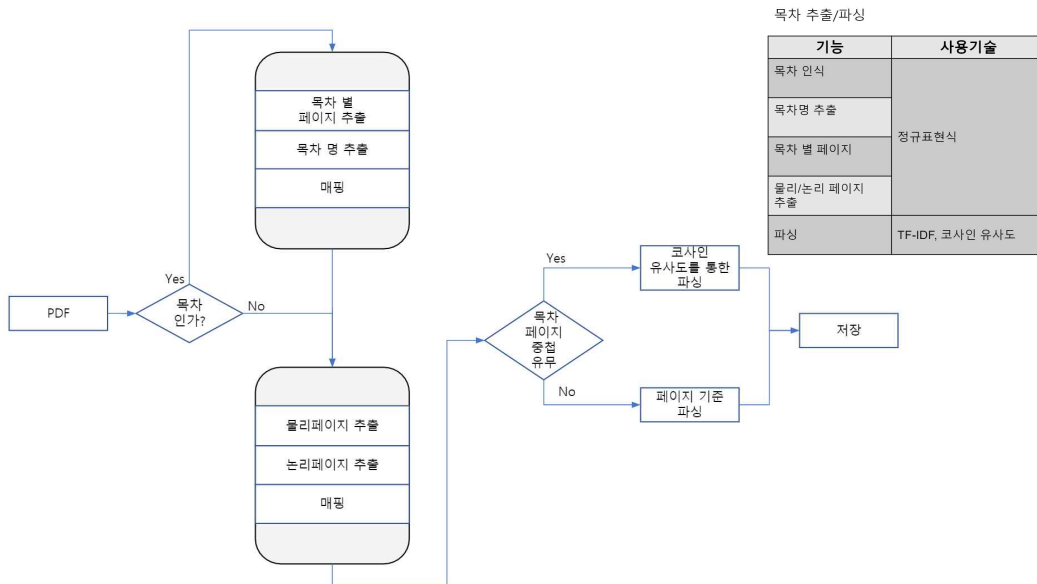
- ▶ 수집된 PDF 파일의 TEXT 변환/추출 과정 후 형태소 분석 및 토큰화 과정을 거쳐 동물질병 용어사전의 키워드 전체에 대한 포함 여부를 비교
- ▶ 동물질병 용어사전 키워드 포함여부에 따라 후처리 수행을 결정

수집 대상 여부 확인



[그림 2-26] 수집대상 여부 확인

- ㉞ 수집 대상 문서에 대한 후처리로는 문서내의 목차 추출, 핵심키워드 추출, 문서/챕터별 요약, 문서/챕터/페이지별 임베딩, 문서 정보로 구성하여 저장
- ▶ 문서에서 목차 페이지를 검출 후 목차 페이지의 텍스트를 파싱하여 목차명과 페이지를 맵핑

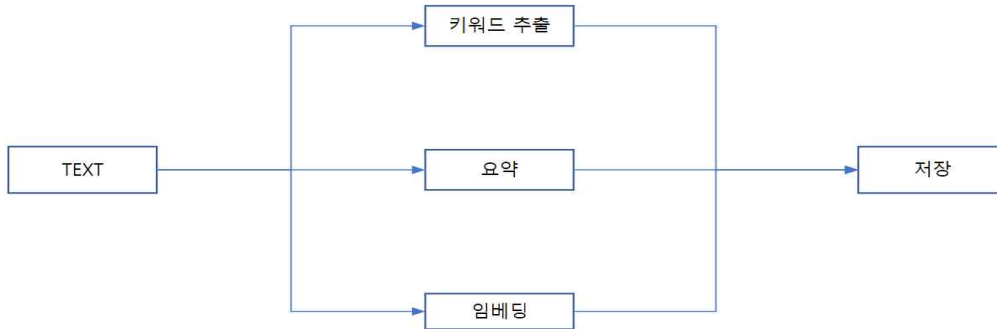


[그림 2-27] 수집문서에 대한 후처리

- ▶ 문서별, 챕터별, 페이지별로 정리된 텍스트는 각각의 단위로 키워드 추출, 요약, 임베딩 처리 후 해당 정보를 저장

PDF2TXT, 요약/키워드 추출, 임베딩

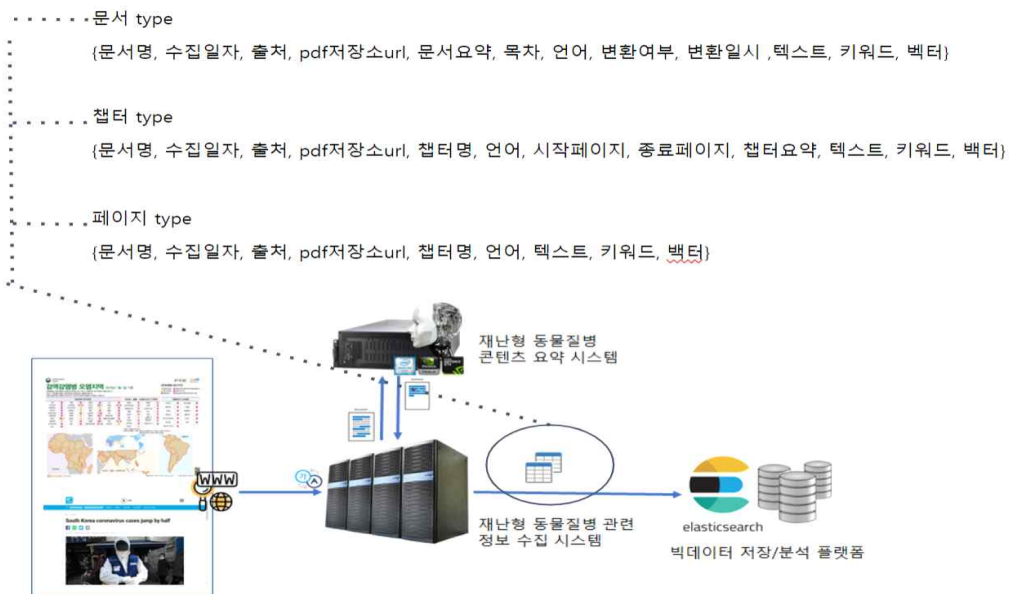
기능	사용기술
pdf2txt	pdftotree (python package)
pdf2Image	pillow (python package) pdf2image (python package)
image2txt	네이버 OCR API
키워드 추출	Kr-wordRank (Gensim)
요약	TextRank (Gensim)
임베딩	DOC2Vec



[그림 2-28] 수집문서의 키워드, 요약 및 임베딩

- ㉔ 수집 문서의 변환, 파싱, 요약 등의 처리결과는 빅데이터 처리를 위한 포맷으로 재구성하여 전송함으로써 빅데이터 DB화

- ▶ 수집 문서의 키워드 추출, 요약, 임베딩 등의 정보와 텍스트를 ELASTIC-SEARCH에 저장



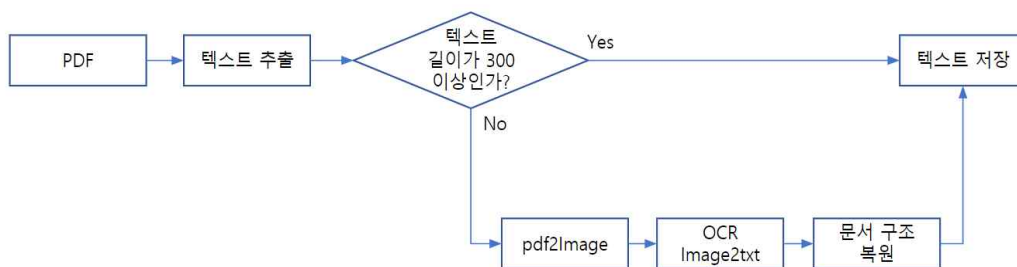
[그림 2-29] 수집문서의 저장

▶ 수집 문서의 문서, 챕터, 페이지별 텍스트, 키워드, 요약, 임베딩 등의 정보 구조

Seq	Filename	Key	Type	DateFormat	설명
1	doc_unit.json	doc_title_STR	String		PDF 문서명 (크롤링 결과)
		doc_article_name_STR	String		기사제목
		doc_author_STR	String		저자
		doc_pdf_url_STR	String		PDF 저장위치
		doc_id_STR	String		크롤링 문서 url 해시코드 값
		doc_type_STR	String		자료유형 (기사, 보고서, 백서 ...)
		publisher_STR	String		발행처
		published_date_DT	Datetime	yyyy-MM-dd HH:mm:ss	발행일자 default(1900-01-01)
		published_date_STR	String		발행일자
		article_date_DT	Datetime	yyyy-MM-dd HH:mm:ss	기사 게시 일자 default(1900-01-01)
		article_date_STR	String		기사 게시 일자
		doc_page_cnt_INT	Int		전체 페이지 수
		crawl_date_DT	Datetime	yyyy-MM-dd HH:mm:ss	수집일자 default(1900-01-01)
		channel_name_STR	String		크롤링 채널 명
		channel_language_STR	String		크롤링 채널 언어(원문)
		channel_nation_STR	String		대상 사이트 소속 국가
		isbn_STR	String		표준번호 (ISBN)
		doc_source_STR	String		출처
		file_size_STR	String		파일크기
		doc_summary_STR	String		문서 요약내용
		doc_content_LIST	List		문서 내용
		doc_trans_job_flag_BOOL	Boolean		번역여부
		doc_complete_date_DT	Datetime	yyyy-MM-dd HH:mm:ss	변환일시 default(1900-01-01)
		doc_text_STR	String		문서 텍스트 내용
trans_txt_url_STR	String		원문 위치		
extract_keyword_LIST	List		추출 핵심 키워드		
doc_vector_LIST	List		문서 벡터 값		
detection_keyword_LIST	List		검출 질명 키워드		
2	chapter_unit_XXX.json	doc_title_STR	String		PDF 문서명 (크롤링 결과)
		doc_pdf_url_STR	String		PDF 저장위치
		doc_id_STR	String		크롤링 문서 url 해시코드 값
		channel_name_STR	String		크롤링 채널 명
		chapter_name_STR	String		챕터명
		channel_language_STR	String		크롤링 채널 언어
		chapter_start_page_INT	String		챕터 시작 페이지
		chapter_end_page_INT	String		챕터 종료 페이지
		chapter_summary_STR	String		챕터 요약 내용
		chapter_text_STR	String		챕터 텍스트 내용
		chapter_vector_LIST	List		챕터 벡터 값
		extract_keyword_LIST	List		추출 핵심 키워드
3	page_unit_XXX.json	doc_title_STR	String		PDF 문서명 (크롤링 결과)
		doc_pdf_url_STR	String		PDF 저장위치
		doc_id_STR	String		크롤링 문서 url 해시코드 값
		channel_name_STR	String		크롤링 채널 명
		chapter_name_STR	String		챕터명
		channel_language_STR	String		크롤링 채널 언어
		page_text_STR	String		페이지 텍스트 내용
		extract_keyword_LIST	List		추출 핵심 키워드
page_vector_LIST	List		페이지 벡터 값		

[그림 2-30] 수집문서에 대한 저장 정보

- 수집·등록된 문서에서 텍스트 데이터를 추출하여 가공이 용이하도록 데이터베이스화
 - ㉔ 수집한 PDF 문서에는 활자 인쇄된 문서의 스캔 이미지로 PDF 문서화 되어 있는 페이지가 다수 존재, 이를 다시 텍스트로 추출하기 위해 PDF 페이지를 이미지로 변환하고 이를 OCR 기술을 사용하여 텍스트를 추출
 - ▶ PDF에서 기본 텍스트를 추출 후 텍스트가 기준 이하로 구성된 PDF는 이미지로 구성된 PDF로 판단, PDF를 페이지별로 이미지로 재구성하고 이를 OCR 로 텍스트 추출, 추출된 텍스트는 문서 순서로 재배열 과정을 거쳐 텍스트 문서로 저장



[그림 2-31] 이미지 문서에서 텍스트 추출

- ㉔ 수집·등록된 문서는 수집, 변환, 파싱, 추출, 요약, 임베딩 등의 모든 작업 및 결과를

DB화하여 관리, 이를 통해 문서별 검색 등의 재가공이 용이

- ▶ 수집 문서의 정보관리를 위한 DB 테이블 리스트로 동물질병용어사전 및 크롤링 서비스 관리, 사용자 등록문서 이력, 수집문서 파싱 결과 정보 테이블 등으로 구성

테이블명	Object ID	Owner	Tablespace	Row Count	Estimate	Partitions	Partition by	Extra Options	Comment
bdt_dict_category	19,587	aiad	pg_default	0		0			
bdt_dict_category_multi	19,593	aiad	pg_default	18		18			사전_카테고리_다국어
bdt_dict_disease	19,599	aiad	pg_default	69		69			
bdt_dict_disease_multi	19,605	aiad	pg_default	138		138			사전_감염병_다국어
bdt_dict_disease_organism	19,611	aiad	pg_default	0		0			
bdt_dict_organism	19,617	aiad	pg_default	0		0			
bdt_dict_organism_multi	19,623	aiad	pg_default	0		0			사전_병원체_다국어
bdt_dict_synonym	19,626	aiad	pg_default	51		51			
bdt_dict_synonym_multi	19,632	aiad	pg_default	69		69			사전_동의어_다국어
tb_airflow_api_job	19,646	aiad	pg_default	4		4			AIRFLOW API[API JOB]
tb_code	19,652	aiad	pg_default	72		72			코드
tb_code_group	19,658	aiad	pg_default	15		15			코드그룹
tb_collect_info	19,661	aiad	pg_default	258		258			수집정보[문서결과]
tb_collect_qa_info	19,668	aiad	pg_default	5		5			문서 QA 정보
tb_doc_list	19,674	aiad	pg_default	637		637			
tb_doc_upload_hist	19,680	aiad	pg_default	2		2			문서 업로드 이력 정보
tb_extract_keyword	19,686	aiad	pg_default	1,278		1,278			키워드_검출_결과
tb_job_log	19,692	aiad	pg_default	53,988		53,988			문서작업로그[문서이력]

[그림 2-32] 수집문서 관리를 위한 테이블 목록

- ▶ 수집문서 파싱 결과 정보 테이블로 문서에서 추출한 모든 정보를 저장

컬럼명	#	Data type	Length	Precision	Scale	Identity	Collation	Not Null	Default	Comment
doc_id	1	varchar	20	20			default	<input checked="" type="checkbox"/>		문서ID
dag_id	2	varchar	100	100			default	<input checked="" type="checkbox"/>		AIR FLOW DAG ID
channel_name	3	varchar	300	300			default	<input checked="" type="checkbox"/>		채널명
crawl_date	4	timestampz		35	6			<input checked="" type="checkbox"/>		수집일자
doc_title	5	varchar	100	100			default	<input checked="" type="checkbox"/>		제목
doc_author	6	varchar	100	100			default	<input type="checkbox"/>		저자
doc_type	7	varchar	20	20			default	<input checked="" type="checkbox"/>		자료유형
publisher	8	varchar	100	100			default	<input type="checkbox"/>		발행처
published_date	9	timestampz		35	6			<input type="checkbox"/>		발행일자
article_date	10	timestampz		35	6			<input type="checkbox"/>		기사 게시 일자
doc_page_cnt	11	int4		10				<input checked="" type="checkbox"/>		전체 페이지수
channel_language	12	varchar	10	10			default	<input checked="" type="checkbox"/>		크롤링 채널 언어 [원문]
channel_nation	13	varchar	10	10			default	<input checked="" type="checkbox"/>		대상 사이트 소속 국가
isbn	14	varchar	100	100			default	<input type="checkbox"/>		표준번호 [ISBN]
series_num	15	varchar	20	20			default	<input type="checkbox"/>		시리즈 불륨/번호
doc_vector	16	text					default	<input type="checkbox"/>		문서 벡터
pg_json_url	17	varchar	4,000	4,000			default	<input type="checkbox"/>		페이지 json 파일 위치
ch_json_url	18	varchar	4,000	4,000			default	<input type="checkbox"/>		챗터 단위 json 파일 위치
doc_json_url	19	varchar	4,000	4,000			default	<input type="checkbox"/>		문서 단위 json 파일 위치
extract_keyword	20	varchar	2,000	2,000			default	<input checked="" type="checkbox"/>		추출 핵심 키워드
trans_txt_url	21	varchar	200	200			default	<input type="checkbox"/>		번역 결과 위치
html_result	22	varchar	200	200			default	<input type="checkbox"/>		HTML 결과 URL
elastic_send_yn	23	varchar	1	1			default	<input checked="" type="checkbox"/>		엘라스틱 서치 전송 여부
create_id	24	varchar	35	35			default	<input checked="" type="checkbox"/>		등록자
create_date	25	timestampz		35	6			<input checked="" type="checkbox"/>		등록일자
update_id	26	varchar	35	35			default	<input checked="" type="checkbox"/>		수정자
update_date	27	timestampz		35	6			<input checked="" type="checkbox"/>		수정일자
detection_keyword	28	varchar	2,000	2,000			default	<input checked="" type="checkbox"/>		검출 키워드
doc_source	29	varchar	4,000	4,000			default	<input type="checkbox"/>		출처
article_name	30	varchar	200	200			default	<input checked="" type="checkbox"/>		첨부파일명
doc_pdf_url	31	varchar	4,000	4,000			default	<input type="checkbox"/>		PDF 파일 위치
file_size	32	varchar	100	100			default	<input type="checkbox"/>		파일크기
doc_error_yn	33	varchar	1	1			default	<input type="checkbox"/>	'N'::charac...	문서오류유무
doc_error_content	34	text					default	<input type="checkbox"/>		문서오류신고내용
web_doc_json_url	35	varchar	4,000	4,000			default	<input checked="" type="checkbox"/>		웹문서 json 파일 위치
doc_error_type	36	varchar	10	10			default	<input type="checkbox"/>		문서오류유형

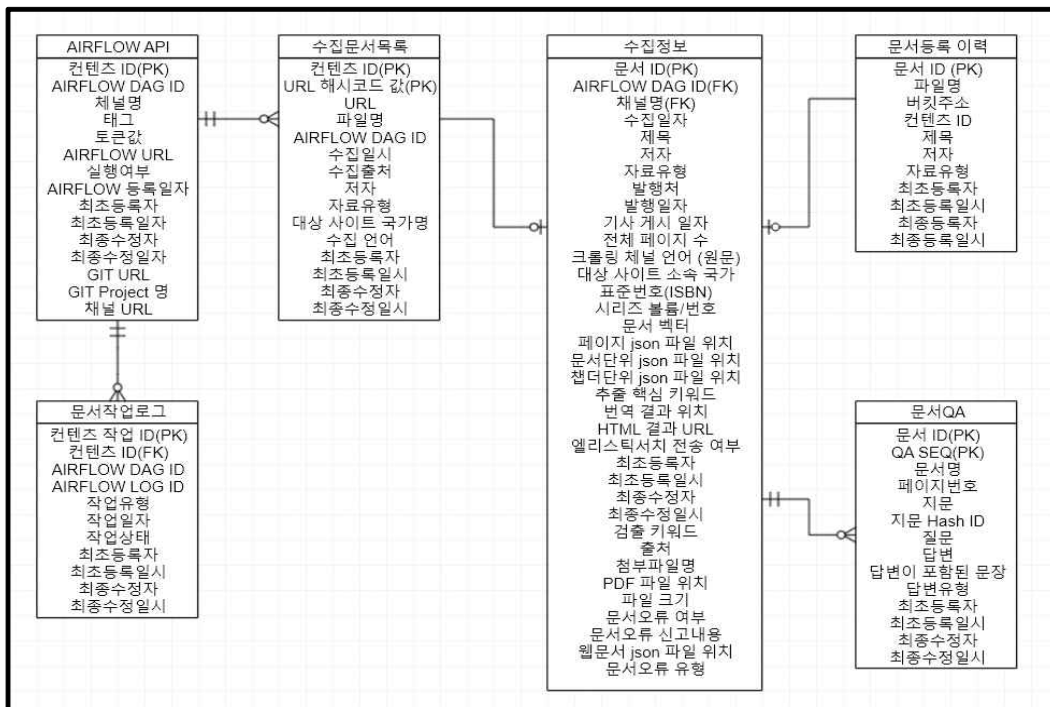
[그림 2-33] 수집문서에 대한 관리 정보

- ▶ 수집문서 파싱 결과 정보 테이블에 저장된 일부 데이터로 모든 문서는 고유의 HASH 값을 ID로 관리, 재수집 및 재처리 과정을 수행해도 해당 ID를 통해 정보 업데이트 가능

doc_id	dag_id	channel_name	crawl_date	doc_title	doc_author	doc_ty						
3b6e63163ea12e35e7b2	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:45:11	151022 구제역, HPAI 등 국가재난형 백신		보도자료						
b8d1efa03c1b19a1eb7e	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-11 03:09:45	2020-27 2020 한국 생물안전 학술	김민정	보도자료						
eeceb26b98b6f3a8c1de	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:45:26	151019 이젠 닭고기 살때도 동물복	백종민	보도자료						
1ebfb4fe19c6f9b783d2	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:46:06	151118 한중 구제역 공동대응을 위	백종민	보도자료						
73eb87f5d89b3c0d1cf3	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:45:56	151104 돼지유행성상사(PED) 발생	안영신	보도자료						
f020975f654535ae9e4	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:45:46	151111 농림축산검역본부 동물용	백종민	보도자료						
1da956655730b0ea7146	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:45:51	151113 '소해면상노출(BSE) 검사'	백종민	보도자료						
cadf33c5b4677f28079e	aiad_crawler_pigtimes_today	양돈타임즈 오늘의 뉴스	2020-11-11 05:28:43	ASF 남아, 멧돼지 CSF 전염	김현구	보도자료						
0093ca91b1fa650adec	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 19:06:40	150925 국내 가축방역기관의	말진	백종민	보도자료					
a790312a00ee6816da1d	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:33:20	151127 구제역 진단능력	제고를 통	백종민	보도자료					
fa11801a4dde523b2f2d	aiad_crawler_pigtimes_today	양돈타임즈 오늘의 뉴스	2020-11-11 05:28:23	경기강원 도축, 비육돈-오전	모든-오	김현구	보도자료					
6ae35d42a02b4fba8f8	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:33:15	151126 (백포시) 프랑스산	가금, 가	백종민	보도자료					
be01a15f366a5faf4d6	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:34:31	160201 농림축산검역본부장	취임	백종민	보도자료					
b482ba2d44e48076acd	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:34:05	160107 현장요구형 동물질병	표준	백종민	보도자료					
4e42934bb5c023f9902	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:34:11	160113 AI 검사 담당자	진단 역량	김현구	보도자료					
c09c1f97a19ba2525d9	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:34:26	160119 6년 연속 전국	원종계장 추	백종민	보도자료					
4e3e0c9e5f5089db648d	aiad_crawler_pigtimes_today	양돈타임즈 오늘의 뉴스	2020-11-11 05:28:53	ASF 없었다면 두수	늘었다	임정은	보도자료					
81c4f07cbe9ea39b7136	aiad_crawler_pigtimes_today	양돈타임즈 오늘의 뉴스	2020-11-11 05:28:38	생식회 위에 소득유 뿌리지	마세요	김현구	보도자료					
9c3477a0fa45fab3aa1	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:33:10	151126 농림축산검역본부,	질병진	백종민	보도자료					
fa87fb934dd9b7478103	aiad_crawler_pigtimes_today	양돈타임즈 오늘의 뉴스	2020-11-11 05:29:38	SOP 규정대로 재입식	허용하라	김현구	보도자료					
f897f69cd8b5918837d	aiad_crawler_pigtimes_field	양돈타임즈 현장	2020-11-11 07:06:14	[농장탐방] '양돈장요?'	천사 하의	김현구	보도자료					
1ad6fe77945a0e46c5a	aiad_crawler_pigtimes_field	양돈타임즈 현장	2020-11-11 07:06:39	[양돈현장] 하절기	파리 피해	예방 방	김현구	보도자료				
ee02596dfdb31eed99f4	aiad_crawler_pigtimes_today	양돈타임즈 오늘의 뉴스	2020-11-11 05:29:13	접경지역 14개 시군, 도축	전 ASF	김현구	보도자료					
dc27f9840dcd68b90ea	aiad_crawler_pigtimes_field	양돈타임즈 현장	2020-11-11 07:07:14	[농장탐방] '좋은	파트너 만나	양돈	매	양돈타임즈	보도자료			
c9fd6f8e6f7a6c7aa473	aiad_crawler_pigtimes_field	양돈타임즈 현장	2020-11-11 07:06:59	[농장탐방] 판스코	유전자센터, 한	단	양돈타임즈	보도자료				
ec813258b147f119e84e1	aiad_crawler_pigtimes_overseas	양돈타임즈 해외양돈	2020-11-11 04:39:27	역세 EU 돼지 값	더 떨어지	나	임정은	보도자료				
45211a1a05aa04eef882	aiad_crawler_pigtimes_overseas	양돈타임즈 해외양돈	2020-11-11 04:39:57	EU 내년 도축 수출	-10% 전망	임정은	보도자료					
4768da9eb67f1f56bdf3	aiad_crawler_pigtimes_overseas	양돈타임즈 해외양돈	2020-11-11 04:40:22	美 9월 돈가 상승세로	반전	임정은	보도자료					
f0f0d1dd900be88b676	aiad_crawler_pigtimes_today	양돈타임즈 오늘의 뉴스	2020-11-11 05:28:58	강원, 멧돼지	무제한 포획	나서	임정은	보도자료				
11bc605e958ce56bdaa2	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 19:06:48	150918 전남	나주 및	강진, AI	의식	백종민	보도자료			
f9951628ffdb7824c9	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 19:06:24	150903	고병원성조류인플루엔자	초	백종민	보도자료				
9e23e6f0de8449446494	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 19:06:28	150918 AI	위기단계 '주의'	발령	및	백종민	보도자료			
4270629ce0d431d54473	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 19:18:09	150818	위지 추적을 통한	야생	나	백종민	보도자료			
a10b1100a3714cbdb7c	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:46:26	151118	농림축산검역본부	서울지	오	백종민	보도자료			
35bf923ab5aa53c74b2	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:45:41	151031	(청와대발표 이후)	삼계탕,	백종민	보도자료				
9aded055b66ad01f3bdf	aiad_crawler_qia_notice	농림축산검역본부 알림마당-보도/해	2020-11-10 18:46:31	151120	한우고기, 홍콩	수출	일	백종민	보도자료			
7791df47c1182de9873	aiad_crawler_pigtimes_overseas	양돈타임즈 해외양돈	2020-11-11 04:39:22	독일 ASF	발생 첫	주	돈가	10% 하락	임정은	보도자료		
efed1d261cc8484bb2	aiad_crawler_pigtimes_overseas	양돈타임즈 해외양돈	2020-11-11 04:40:17	캄보디아 ASF	로	돼지	30% 감소	임정은	보도자료			
24c51a2d790501d49829	aiad_crawler_pigtimes_overseas	양돈타임즈 해외양돈	2020-11-11 04:40:12	뿔	양돈	코로나 ASF	돼지	'시름'	임정은	보도자료		
d09f4193ad03c3f1b5dd	aiad_crawler_pigtimes_overseas	양돈타임즈 해외양돈	2020-11-11 04:39:47	美	돈육	수출	늘다	두달	재	감소	임정은	보도자료

[그림 2-34] 수집문서에 대한 파싱 결과 예

○ 역학조사보고서·백서·공문 및 발간물 등록 서비스 개발



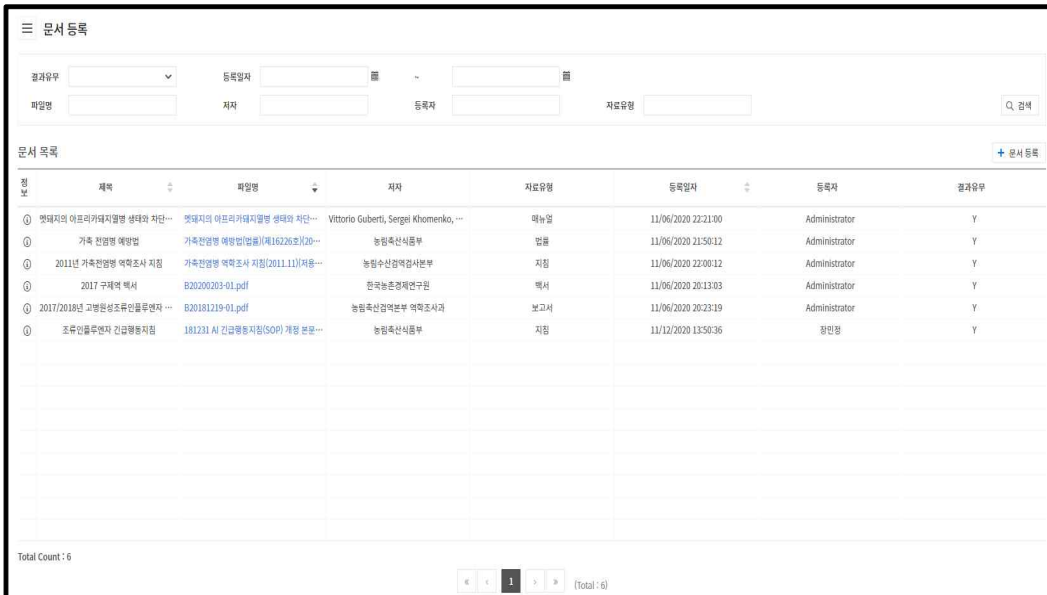
[그림 2-35] 수집문서 데이터 모델링

- 크롤링 서비스에 의해 수집된 콘텐츠와 수작업에 의해 등록된 문서 콘텐츠를 관리하기 위한 데이터 모델 설계
- 수집된 문서 및 QA 관리를 위한 기능 개발



[그림 2-36] 문서 및 QA관리 메뉴

- 수작업에 의해 등록된 문서 목록을 제공하는 기능 개발



[그림 2-37] 수작업 등록문서 목록

- 크롤러에 의해 수집할 수 없는 콘텐츠 등록하는 기능 개발



[그림 2-38] 문서등록

- 크롤러가 수집한 문서 및 수작업 등록된 문서의 목록을 제공하는 기능 개발

수집 문서 및 QA 등록

검색 필터: 검색어, 문서상태, 수집일자, 문서명, 채널명, 출처, 자유유형, 저자

수집 문서 목록 [문서상세화면 > QA 등록]

번호	문서명	채널명	자유유형	저자	출처	수집일자	검색명 키워드	문서상태
1	환경부가 말하지 않는 아열대성 돼지 A...	돼지위사람 전체 기사	보도자료	관리자	http://pigpeople.net/news/arti...	11/15/2020 07:33:49	아프리카돼지열병, 돼지열병	정상
2	행부 ASF가 발생 시 살처분에 대해 ...	돼지위사람 전체 기사	보도자료	관리자	http://pigpeople.net/news/arti...	11/12/2020 07:34:00		정상
3	돼지 동물결핵 상황을 자문 수집 분...	돼지위사람 전체 기사	보도자료	관리자	http://pigpeople.net/news/arti...	11/17/2020 05:17:01	돼지결핵, 조류인플루엔자, 아프리카...	정상
4	한탄강스마트림, '미래 친환경산업' ...	돼지위사람 전체 기사	보도자료	관리자	http://pigpeople.net/news/arti...	11/12/2020 19:34:10		정상
5	한탄강에 마두름과 살얼갈음 준 ...	돼지위사람 전체 기사	보도자료	관리자	http://pigpeople.net/news/arti...	11/15/2020 19:33:31	아프리카돼지열병, 돼지열병, 조류...	정상
6	피우 ASF 발병 교 집중재 투입, ...	돼지위사람 전체 기사	보도자료	관리자	http://pigpeople.net/news/arti...	11/16/2020 07:33:50	아프리카돼지열병, 돼지열병	정상
7	케이서이드, 업계 최초 구축형성공...	돼지위사람 전체 기사	보도자료	관리자	http://pigpeople.net/news/arti...	11/17/2020 05:33:32	구제역	정상
8	케이서이드, 'ASF' 백신 개발뿐만 마...	돼지위사람 전체 기사	보도자료	관리자	http://pigpeople.net/news/arti...	11/12/2020 19:34:11	아프리카돼지열병, 구제역, 돼지열병	정상
9	케이서이드 '2020년 하반기 우수단...	돼지위사람 전체 기사	보도자료	관리자	http://pigpeople.net/news/arti...	11/15/2020 19:33:31	구제역	정상
10	한반도에 ASF로 돼지 30% 감소	양돈타임즈 해외양돈	보도자료	임정은	http://www.pigtimes.co.kr/ne...	11/17/2020 03:24:04		정상
11	카일레그리부터, 방역공공 운영 5...	돼지위사람 전체 기사	보도자료	관리자	http://pigpeople.net/news/arti...	11/17/2020 07:34:02		정상
12	축산관련종사자 교육 온라인으로 전...	돼지위사람 전체 기사	보도자료	관리자	http://pigpeople.net/news/arti...	11/16/2020 07:33:51	아프리카돼지열병, 돼지열병	정상
13	전년 야생조류서 H5N8형 AI 환원 ...	데일리넷 뉴스	보도자료	윤상준	https://www.dailyet.co.kr/ne...	11/12/2020 19:32:48		정상
14	전년 야생조류서 H5N8형 AI 환원 ...	데일리넷 뉴스	보도자료	윤상준	https://www.dailyet.co.kr/ne...	11/15/2020 19:32:44	고병원성 조류인플루엔자, 조류인...	정상
15	집행당국에 멧돼지 울타리 200여곳...	돼지위사람 전체 기사	보도자료	관리자	http://pigpeople.net/news/arti...	11/15/2020 19:33:16	아프리카돼지열병, 돼지열병	정상

Total Count : 285

[그림 2-39] 수집된 문서 목록

- 수집·등록된 콘텐츠 본문과 콘텐츠 주요정보 그리고 AI에 의한 콘텐츠 요약정보를 제공하는 기능 개발

수집 문서 및 QA 등록

문서 정보: 제목, 채널명, 출처, 작성일자, 등록일자, 자유유형, 저자

콘텐츠 본문

콘텐츠 주요정보

콘텐츠 요약

[그림 2-40] 문서정보 및 요약정보

- 수집·등록된 콘텐츠에 대한 질의 및 응답 목록을 제공하는 기능 개발

번호	문서명	QA개수	자료유형	저자	출처	수입일자	검색용 키워드	문서상태
①	B20181219-01.pdf	2	보고서	농림축산검역본부 역학조사과	Administrator	11/11/2020 09:05:47	저병원성조류인플루엔자, 조류인플...	정상

[그림 2-41] 질의응답 목록

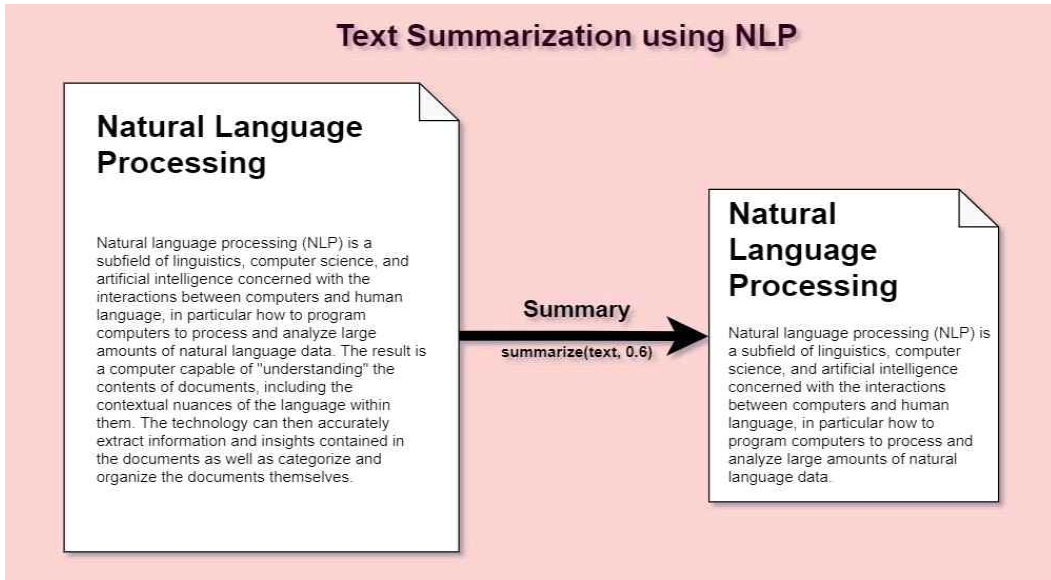
- 수집·등록된 콘텐츠에 대한 질의 및 응답은 신규등록·수정하는 기능 개발



[그림 2-42] 문서에 대한 질의응답 관리

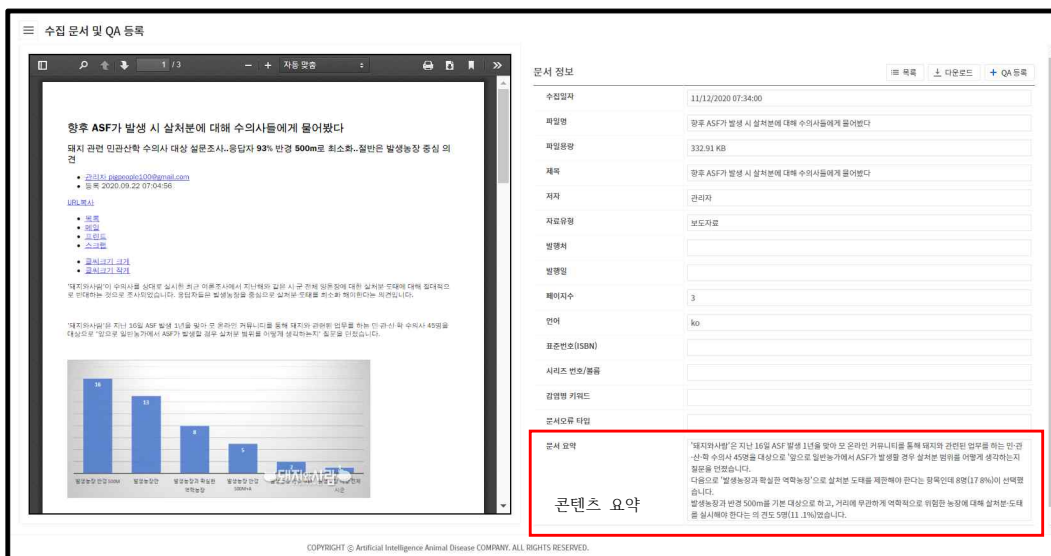
(나) 재난형 동물질병 콘텐츠에 대한 인공지능 요약모델 개발

- PYTHON TEXTRANK 알고리즘을 활용한 콘텐츠 요약 모델 개발 및 결과검증
 - 문서 집합을 요약하는 방법으로 키워드와 핵심 문장을 선택하는 EXTRACTIVE METHODS 방법 중 가장 널리 이용되는 방법 중 하나로 2004년에 제안된 TEXTRANK 적용
 - TEXTRANK는 WORD GRAPH나 SENTENCE GRAPH를 구축한 뒤, GRAPH RANKING 알고리즘인 PAGERANK를 이용하여 각각 키워드와 핵심 문장을 선택하여 주어진 문서 집합을 요약
 - GENSIM라이브러리의 요약 모듈을 활용해 콘텐츠 요약 모델 개발



[그림 2-43] GENSIM 라이브러리를 활용한 콘텐츠 요약의 예

- 동물질병 관련 문서의 적정 요약 길이를 출력할 수 있는 최적의 파라미터 선정하여 수집된 문서의 내용 요약 정보 제공



[그림 2-44] 동물질병 관련 문서의 요약 정보 제공

(2) 재난형 동물질병 정보 검색용 빅데이터 플랫폼 구축

(가) 콘텐츠 활용도를 높이기 위한 정보의 DB화 및 인덱싱 기술 개발

① 빠른 데이터 검색을 위한 역인덱스 구조로 INDEXING

- 역 인덱스의 데이터 구조를 기반으로 하는 오픈소스인 ELASTICSEARCH를 선정하고, [그림 2-45]와 같이 하나의 시스템에 SINGLE NODE로 구축함



[그림 2-45] ELASTICSEARCH 설치 구성 및 실행 화면

- ELASTICSEARCH를 구축하고 한글 형태소 분석기인 NORI를 설치하고 ANALYZER 구성을 마친 후 수집 대상인 문헌 자료에 대한 데이터 스키마를 생성하여 적재 및 검색 테스트 완료, ELASTICSEARCH 도입에 따른 데이터 구조변경 내용은 다음과 같음

ID	Text
doc1	The quick brown fox
doc2	The quick brown fox jumps over the lazy dog
doc3	The quick brown fox jumps over the quick dog
doc4	Brown fox brown dog
doc5	Lazy jumping dog

[그림 2-46] 전통적 테이블

- 일반적으로 오라클이나 MYSQL 같은 관계형 DB에서는 [그림 2-46]의 내용을 보이는 대로 테이블 구조로 저장. 만약에 이 테이블에서 TEXT 에 FOX가 포함된 행들을 가져온다고 하면 다음과 같이 TEXT 열을 한 줄씩 찾아 내려가면서 FOX가 있으면 가져오고 없으면 넘어가는 식으로 데이터를 가져오게 됨

ID	Text	Result
doc1	The quick brown fox	fox (O) → 선택
doc2	The quick brown fox jumps over the lazy dog	fox (O) → 선택
doc3	The quick brown fox jumps over the quick dog	fox (O) → 선택
doc4	Brown fox brown dog	fox (O) → 선택
doc5	Lazy jumping dog	fox (X) → 제외

[그림 2-47] 테이블 데이터에서 한 줄씩 LIKE 검색

- 전통적인 RDBMS 에서는 [그림 2-47]과 같이 LIKE 검색을 사용하기 때문에 데이터가 늘어날수록 검색해야 할 대상이 늘어나 시간도 오래 걸리고, ROW 안의 내용을 모두 읽어야 하기에 기본적으로 속도가 느림. ELASTICSEARCH는 데이터를 저장할 때 다음과 같이 역 인덱스(INVERTED INDEX)라는 구조를 만들어 저장하기 때문에 검색에 유리함

텀(Term)	ID	텀(Term)	ID
The	doc1, doc2, doc3	quick	doc1, doc2, doc3
brown	doc1, doc2, doc3, doc4	fox	doc1, doc2, doc3, doc4
jumps	doc2, doc3	over	doc2, doc3
the	doc2, doc3	lazy	doc2
dog	doc2, doc3, doc4, doc5	Brown	doc4
Lazy	doc5	jumping	doc5

[그림 2-48] 역인덱스 구조 테이블

- [그림 2-48]의 역 인덱스는 책의 맨 뒤에 있는 주요 키워드에 대한 내용이 몇 페이지에 있는지 볼 수 있는 찾아보기 페이지에 비유함. ELASTICSEARCH에서는 추출된 각 키워드를 텀(TERM) 이라고 함. 이렇게 역 인덱스가 있으면 FOX를 포함하고 있는 도큐먼트들의 ID를 바로 얻어올 수 있음

텀(Term)	ID
fox	doc1, doc2, doc3, doc4

[그림 2-49] ELASTICSEARCH 인덱스 구조

- 그림 2-49)와 같이 ELASTICSEARCH는 데이터가 늘어나도 찾아가야 할 행이 늘어나는 것이 아니라 역 인덱스가 가리키는 ID의 배열값이 추가되는 것 뿐이기 때문에 큰 속도의 저하 없이 빠른 속도로 검색이 가능함. 이런 역 인덱스를 데이터가 저장되는 과정에서 만들기 때문에 ELASTICSEARCH는 데이터를 입력할 때 저장이 아닌 색인을 한다고 표현하고, 테이블을 인덱스(INDEX)라고 부름

② 정형,반정형,비정형 대응을 위한 FULL TEXT INDEXING 기술 적용

- ELASTICSEARCH는 반정형, 비정형 대응을 위한 FULL TEXT INDEXING 기술을 기반으로 검색기능을 제공하기 위해 “TEXT” 데이터 타입을 제공하며, [그림 2-50]과 같은 옵션들을 제공하여 검색 결과의 품질을 높일 수 있도록 구현함

text

text 타입은 입력된 문자열을 텀 단위로 쪼개어 역 색인 (inverted index) 구조를 만듭니다. 보통은 풀텍스트 검색에 사용할 문자열 필드 들을 text 타입으로 지정합니다. text 필드에 설정 가능한 옵션들은 다음과 같은 것들이 있습니다.

- "analyzer" : "<애널리라이저명>" - 색인에 사용할 애널리라이저를 입력하며 디폴트로는 standard 애널리라이저를 사용합니다. 토큰나이저, 토큰필터들을 따로 지정할수가 없으며 필요하다면 사용자 정의 애널리라이저를 settings에 정의 해 두고 사용합니다.
- "search_analyzer" : "<애널리라이저명>" - 기본적으로 text 필드는 match 쿼리로 검색을 할 때 색인에 사용한 동일한 애널리라이저로 검색 쿼리를 분석합니다. search_analyzer 를 지정하면 검색시에는 색인에 사용한 애널리라이저가 아닌 다른 애널리라이저를 사용합니다. 보통 NGram 방식으로 색인을 했을 때는 지정 해 주는 것이 바람직합니다.
- "index" : <true | false> - 디폴트는 true 입니다. false로 설정하면 해당 필드는 역 색인을 만들지 않아 검색이 불가능하게 됩니다.
- "boost" : <숫자 값> - 디폴트는 1 입니다. 값이 1 보다 높으면 풀텍스트 검색 시 해당 필드 스코어 점수에 가중치를 부여합니다. 1보다 낮은 값을 입력하면 가중치가 내려갑니다.
- "fielddata" : <true | false> - 디폴트는 false 입니다. true로 설정하면 해당 필드의 색인된 텀 들을 가지고 집계(agggregation) 또는 정렬(sorting)이 가능합니다. 이 설정은 다이나믹 설정으로 이미 정의된 매핑에 true 또는 false로 다시 적용하는 것이 가능합니다.

[그림 2-50] TEXT 데이터 타입에 대한 ELASTIC사의 설명 자료

- 본 연구과제에서 아래와 같은 매핑(스키마)을 활용하여 FULL TEXT INDEXING을 수행하고 원활한 검색이 가능하도록 구현함

```

{"vector_test_mapped": {
  "mappings": {
    "properties": {
      "channel_language_STR": {
        "type": "text",
        "fields": {
          "keyword": {
            "type": "keyword"
          }
        }
      },
      "channel_name_STR": {
        "type": "text",
        "fields": {
          "keyword": {
            "type": "keyword"
          }
        }
      },
      "chapter_end_page_INT": {
        "type": "integer"
      },
      "chapter_name_STR": {
        "type": "text",
        "fields": {
          "keyword": {
            "type": "keyword"
          }
        }
      },
      "chapter_start_page_INT": {
        "type": "integer"
      },
      "chapter_summary_STR": {
        "type": "text"
      },
      "chapter_text_STR": {
        "type": "text"
      },
      "chapter_vector_LIST": {
        "type": "dense_vector",
        "dims": 300
      },
      "doc_id_STR": {
        "type": "text",
        "fields": {
          "keyword": {
            "type": "keyword"
          }
        }
      },
      "doc_pdf_url_STR": {
        "type": "text",
        "fields": {
          "keyword": {
            "type": "keyword"
          }
        }
      },
      "doc_title_STR": {
        "type": "text",
        "fields": {
          "keyword": {
            "type": "keyword"
          }
        }
      },
      "extract_keyword_LIST": {
        "type": "text",
        "fields": {
          "keyword": {
            "type": "keyword"
          }
        }
      }
    }
  }
}

```

[그림 2-51] INDEX 매핑 정보

③ 유연한 한글 처리를 위한 자체 한글형태소 분석기 NORI 탑재

- 모든 언어는 저마다의 고유한 어휘와 조사 등 기본형 등이 다름. 따라서 한글을 제대로 형태소로 나누고 저장하고 검색하기 위해서는 한글만의 형태소 분석기가 필요함. 본 과제에서는 한글의 형태소 분석기능을 활용하기 위해 NORI 형태소 분석기를 [그림 2-52]과 같이 구축하여 활용

```

[ezfarm@localhost elasticsearch-7.7.1]$ cd plugins
[ezfarm@localhost plugins]$ ll
drwxr-xr-x. 2 root root 149 6월 29 09:37 analysis-nori
[ezfarm@localhost plugins]$ cd analysis-nori/
[ezfarm@localhost analysis-nori]$ ll
drwxr-xr-x. 2 root root 149 6월 29 09:37
-rw-r--r--. 1 root root 11358 6월 29 09:37 LICENSE.txt
-rw-r--r--. 1 root root 35175 6월 29 09:37 NOTICE.txt
-rw-r--r--. 1 root root 13545 6월 29 09:37 analysis-nori-7.7.1.jar
-rw-r--r--. 1 root root 7487588 6월 29 09:37 lucene-analyzers-nori-8.5.1.jar
-rw-r--r--. 1 root root 1518 6월 29 09:37 plugin-descriptor.properties
[ezfarm@localhost analysis-nori]$

```

[그림 2-52] NORI 형태소 분석기

니다.

- 실제 데이터 인덱싱시 NORI 형태소 분석기가 활용되기 위해서는 아래 [그림 2-53]와 같이 인덱스 스키마에 특별한 설정이 필요함. 추가적으로 사용자 사전, 유의어 사전, 불용어 사전을 활용하여 검색 품질을 높일 수 있는 기능들을 제공함

```

{
  "vector_test_mapped": {
    "settings": {
      "index": {
        "number_of_shards": "2",
        "provided_name": "vector_test_mapped",
        "creation_date": "160228795446",
        "analysis": {
          "filter": {
            "stop_filter": {
              "type": "stop",
              "stopwords_path": "stopwords.txt"
            },
            "synonym_filter": {
              "type": "synonym_graph",
              "synonyms_path": "synonyms.txt"
            }
          }
        }
      }
    }
  },
  "npos_filter": {
    "type": "nori_part_of_speech",
    "stoppage": [
      "E",
      "IC",
      "IT",
      "MAG",
      "MM",
      "SP",
      "SSC",
      "SSO",
      "SC",
      "SE",
      "XFN",
      "XSA",
      "XSN",
      "XSV",
      "XJA",
      "XNA",
      "XVY"
    ]
  }
}

```

```

{
  "analyzer": {
    "default": {
      "filter": [
        "npos_filter",
        "synonym_filter",
        "stop_filter",
        "nori_readingform",
        "lowercase"
      ],
      "type": "custom",
      "tokenizer": "nori_user_dict_tokenizer"
    },
    "nori_user_dict_tokenizer": {
      "tokenizer": {
        "nori_user_dict_tokenizer": {
          "type": "nori_tokenizer",
          "user_dictionary": "userdict_ko.txt",
          "decompound_mode": "mixed"
        }
      },
      "number_of_replicas": "1",
      "uuid": "YAB1ZAVS1e1cThtSOFPig",
      "version": "7070199"
    }
  }
}

```

[그림 2-53] INDEX SETTING

- 실제 인덱싱된 문서 정보를 바탕으로 FULL TEXT 검색을 수행하면 아래 [그림 2-54]와 같은 결과를 얻을 수 있음

```

1 GET /_search
2 {
3   "query": {
4     "match_all": {}
5   }
6 }
7
8 GET /_search?pretty
9 {
10  "took": 10,
11  "timed_out": false,
12  "_source": true,
13  "hits": [
14    {
15      "_type": "nori_test",
16      "_id": "1",
17      "score": 1.0,
18      "source": {
19        "text": "안녕하세요! 반갑습니다."
20      }
21    },
22    {
23      "_type": "nori_test",
24      "_id": "2",
25      "score": 1.0,
26      "source": {
27        "text": "안녕하세요! 반갑습니다."
28      }
29    },
30    {
31      "_type": "nori_test",
32      "_id": "3",
33      "score": 1.0,
34      "source": {
35        "text": "안녕하세요! 반갑습니다."
36      }
37    },
38    {
39      "_type": "nori_test",
40      "_id": "4",
41      "score": 1.0,
42      "source": {
43        "text": "안녕하세요! 반갑습니다."
44      }
45    },
46    {
47      "_type": "nori_test",
48      "_id": "5",
49      "score": 1.0,
50      "source": {
51        "text": "안녕하세요! 반갑습니다."
52      }
53    },
54    {
55      "_type": "nori_test",
56      "_id": "6",
57      "score": 1.0,
58      "source": {
59        "text": "안녕하세요! 반갑습니다."
60      }
61    },
62    {
63      "_type": "nori_test",
64      "_id": "7",
65      "score": 1.0,
66      "source": {
67        "text": "안녕하세요! 반갑습니다."
68      }
69    },
70    {
71      "_type": "nori_test",
72      "_id": "8",
73      "score": 1.0,
74      "source": {
75        "text": "안녕하세요! 반갑습니다."
76      }
77    },
78    {
79      "_type": "nori_test",
80      "_id": "9",
81      "score": 1.0,
82      "source": {
83        "text": "안녕하세요! 반갑습니다."
84      }
85    },
86    {
87      "_type": "nori_test",
88      "_id": "10",
89      "score": 1.0,
90      "source": {
91        "text": "안녕하세요! 반갑습니다."
92      }
93    }
94  ]
95 }

```

[그림 2-54] FULL TEXT 검색 결과

④ REST API를 통한 데이터 쿼리 기능 제공

- 총 두 가지의 REST API 형태로, REST 요청 URI를 통해 검색 매개변수를 보내는 것과 REST 요청 본문으로 구분 제공
- 요청 본문 방식은 더 상세한 표현이 가능하고 또한 더 읽기 쉬운 JSON 형식으로 검색을 정의할 수도 있음. 아래 [그림 2-55]는 실제 VECTOR로 임베딩된 문서를 적재하고 검색한 예시임

```

22 GET vector_text_mapped_search
23 {
24   "size": 1000,
25   "source": ["doc_title_STX", "highlight",
26     "chapter_start_page_INT"],
27   "query": {
28     "script_score": {
29       "script": "
30         *
31         *
32         *source' : "' + doc['value'] + ' dotProduct(params
33         *_query_vector, 'chapter_vector_LIST)'; return value
34         *
35         *
36       "
37     }
38   }
39 }
40
41 {
42   "took": 24,
43   "timed_out": false,
44   "_shards": {
45     "total": 2,
46     "successful": 2,
47     "skipped": 0,
48     "failed": 0
49   },
50   "hits": {
51     "total": {
52       "value": 20,
53       "relation": "eq"
54     },
55     "max_score": 1000.00100,
56     "hits": [
57       {
58         "_index": "vector_text_mapped",
59         "_type": "_doc",
60         "_id": "w2_D0m412082-127gh",
61         "_score": 1000.00100,
62         "source": {
63           "chapter_start_page_INT": 79,
64           "doc_title_STX": "(2014-2015) 우제목 Foot-and-mouth disease 연구조사문헌보고서"
65         }
66       },
67       {
68         "_index": "vector_text_mapped",
69         "_type": "_doc",
70         "_id": "w2_D0m412082-127gh",
71         "_score": 1000.00100,
72         "source": {
73           "chapter_start_page_INT": 91,
74           "doc_title_STX": "(2014-2015) 우제목 Foot-and-mouth disease 연구조사문헌보고서"
75         }
76       },
77       {
78         "_index": "vector_text_mapped",
79         "_type": "_doc",
80         "_id": "w2_D0m412082-127gh",
81         "_score": 1000.00100,
82         "source": {
83           "chapter_start_page_INT": 115,
84           "doc_title_STX": "(2014-2015) 우제목 Foot-and-mouth disease 연구조사문헌보고서"
85         }
86       }
87     ]
88   }
89 }

```

[그림 2-55] VECTOR 검색 결과

○ 위 [그림 2-55]의 우측 결과 화면에 표기된 정보는 아래와 같은 의미를 가짐

```

took - Elasticsearch가 검색을 실행하는 데 걸린 시간(밀리초)
timed_out - 검색의 시간 초과 여부
_shards - 검색한 샤드 수 및 검색에 성공/실패한 샤드 수
hits - 검색 결과
hits.total - 검색 조건과 일치하는 문서의 총 개수
hits.hits - 검색 결과의 실제 배열(기본 설정은 처음 10개 문서)
hits.sort - 결과의 정렬 키(점수 기준 정렬일 경우 표시되지 않음)
hits.score 및 max_score - 문서 정확도를 score값으로 표기(높을수록 정확도가 높음)

```

[그림 2-56] 결과 화면 정보 의미

⑤ 높은 수집 성능을 위한 INGEST NODE 기술 적용

○ 초기 설계상으로는 오픈소스 ETL 툴인 LOGSTASH를 통한 데이터 수집 적재를 구축하려 계획하였으나, 연구 진행 과정에서의 분석 회의를 통해 DENSE VECTOR 적재시 문서 임베딩 모듈과 ELASTICSEARCH INGEST PIPELINE과 직접 통신하여 데이터 수집 적재를 진행함



[그림 2-57] 서비스 구성도

⑥ 다양한 수집기 BEATS 기술 제공

- 원격 수집의 경우 여러 가지 변수를 고려해야 하기 때문에 지속적으로 수집 항목과 리스크에 대한 논의 및 분석
- 최종적으로 오픈소스 데이터 수집기인 BEATS를 사용하여 원격지 데이터 수집 결정



[그림 2-58] BEATS 경량 수집기 설명

⑦ 데이터 시각화 및 관리 기능 제공

- 오픈소스 시각화 툴인 KIBANA를 구축하여 ELASTICSEARCH와 연동하여 데이터에 대한 시각화 및 전체 ELASTICSEARCH NODE에 대한 관리기능을 수행하도록 구축하였음



[그림 2-59] KIBANA 시각화 소개화면

- KIBANA에서 ELASTICSEARCH의 자원상황 및 동작에 대한 직관적인 모니터링을 실시간으로 제공함. [그림 2-60]는 실제 구축 운영 중인 시스템의 모니터링 화면임

Status	Nodes	Indices	JVM Heap	Total shards	Unassigned shards	Documents	Data
Yellow	1	12	2.4 GB / 16.0 GB	21	7	730	3.8 MB

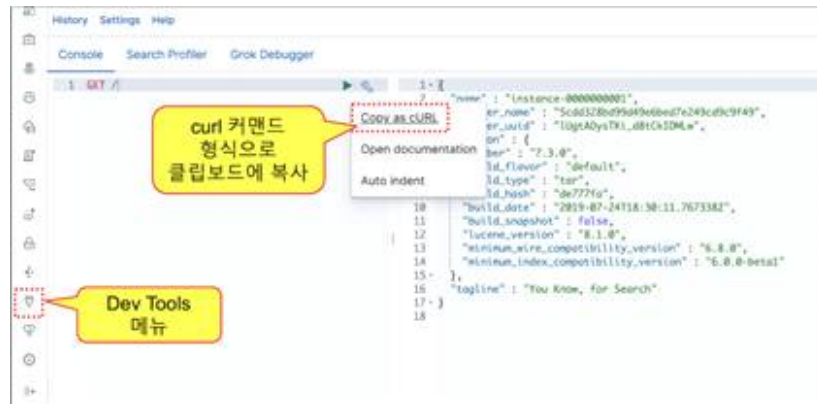
Name ↑	Status	Shards	CPU Usage	Load Average	JVM Heap	Disk Free Space
★ mrc-1 210.92.91.218:9300	Online	14	4% ↑ 4% max 0% min	0.15 ↑ 0.15 max 0.04 min	14% ↑ 14% max 12% min	465.2 GB ↓ 465.2 GB max 465.2 GB min

[그림 2-60] ELASTICSEARCH NODE 모니터링

- KIBANA에서 제공하는 DEV TOOL을 이용하여 검색 쿼리를 작성하고 실행하여 원하는 검색 결과를 얻을 수 있음

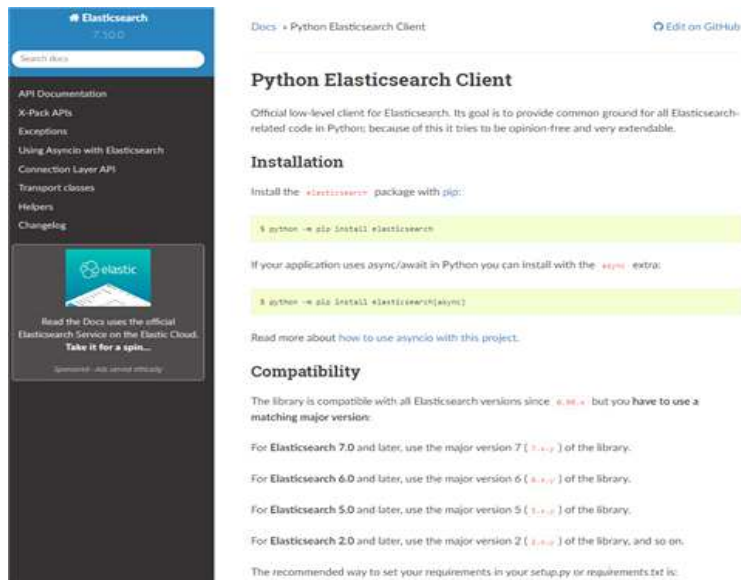
(나) 콘텐츠 활용을 위한 서비스(API) 기술 개발

① REST API를 통한 CRUD 처리기술 제공



[그림 2-61] KIBANA DEV TOOL 예시

- ELASTICSEARCH와 내외부간 데이터 통신에 REST API를 이용할 수 있게 구축하였으며, KIBANA의 DEV TOOL에서 직접 GET PUT POST DELETE API를 이용하여 컨트롤 가능



[그림 2-62] PYTHON CLIENT 설치 가이드의 일부

- 별도의 외부 프로그램에서 REST API를 이용하기 위해서는 LOW LEVEL CLIENT구축이 선행되어야 하는데 본 연구과제에서는 [그림 2-62]와 같이 PYTHON CLIENT를 구축하여 통신함

② 문헌자료 수집을 위한 문서변환 기술 제공

- 오픈소스 프로젝트인 APACHE TIKA(<https://tika.apache.org/>)를 활용하여 문서를 변환하는 것으로 설계 하여 주관기관인 노스스타컨설팅에 전달함

(다) 정보 검색 시스템 프로토타입 개발

① 해외 질병 데이터 및 백서/역학조사보고서 연계

- 모든 데이터/문서는 MRC 유사도 검색을 위해 VECTOR로 임베딩 되어 적재되기 때문에 적재되는 데이터마다 별도의 인터페이스가 존재하지 않기에, 문서변환 및 임베딩 모듈에서 REST API를 이용해 직접 적재를 하고 있음
- ELASTICSEARCH측에서는 적재될 INDEX의 MAPPING정보와 SETTINGS 정보를 데이터에 적합하고 생성하고 적재 시 데이터 검증을 진행함

```

1 from elasticsearch import Elasticsearch
2 from elasticsearch import helpers
3
4 import time
5 import json
6 import glob
7
8
9 class ES_Demans:
10     def __init__(self):
11         self.es = Elasticsearch(hosts=["http://localhost:9200"], timeout=30, max_retries=100, retry_on_timeout=True)
12         self.dense_idx = "dense_test"
13         # @용 vector dimension() 변경되면 필름 시간 조절 가능
14         self.vec_dims = 300
15
16     def dense_vector_es(self):
17         dense_s_body = {
18             "mappings": {
19                 "properties": {
20                     "doc_title_STR": {
21                         "type": "text"
22                     },
23                     "doc_pdf_url_STR": {
24                         "type": "text"
25                     },
26                     "doc_id_STR": {
27                         "type": "text"
28                     },
29                     "channel_name_STR": {
30                         "type": "text"
31                     },
32                     "chapter_name_STR": {
33                         "type": "text"
34                     },
35                     "channel_language_STR": {
36                         "type": "text"
37                     },
38                     "chapter_start_page_INT": {

```

[그림 2-63] 문서 VECTOR 임베딩 PYTHON 샘플 코드

health	status	index	uuid	pri	rep	docs.count	docs.deleted	store.size	pri.store.size	
2	yellow	open	vector_test_mapped	jVAB1ZAVS1icL7ht50FxIg	2	1	29	0	482.5kb	482.5kb
3	yellow	open	vector_test_mapped2	RVivsvaYSx6ZxzBrnT9CeQ	2	1	29	0	430.6kb	430.6kb
4	yellow	open	vector_test	pl_U8754QgicT5NXAhnaDw	1	1	29	0	467.4kb	467.4kb

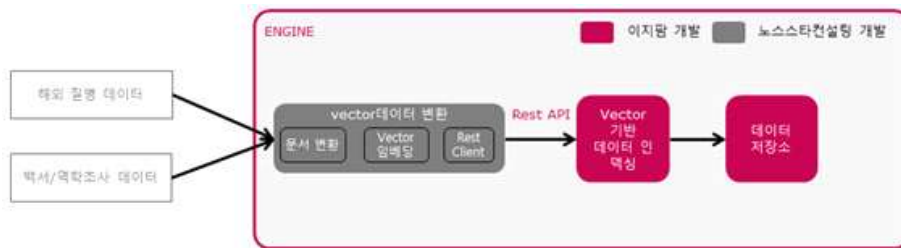
[그림 2-64] 적재된 INDEX 샘플

② Q&A DB 생성 및 관리

- Q&A DB는 곧 AI 학습의 기반이 되는 데이터이므로, 현재 양질의 데이터 적재를 위해 다양한 각도에서의 접근 및 면밀하고 반복적인 검토를 거치며 적재 중
- 최종적으로는 문서데이터와 마찬가지로 VECTOR로 임베딩된 데이터 구조 형태를 가지게됨

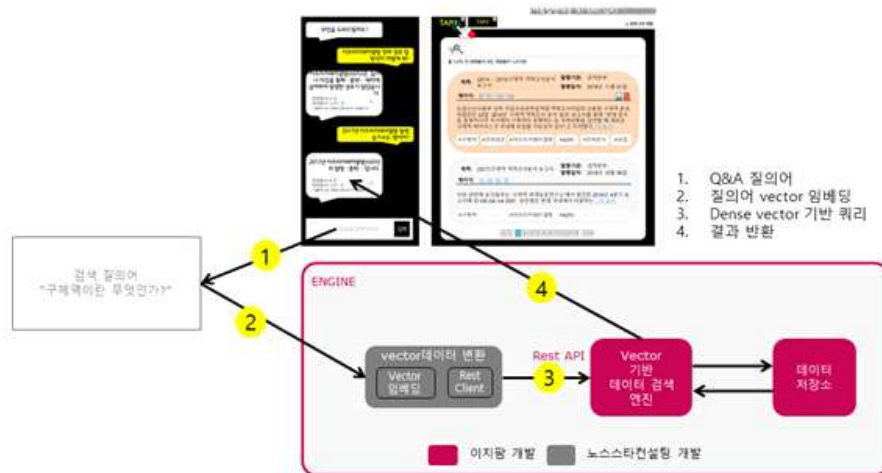
③ 기계독해 연계

- MRC와 ELASTICSEARCH간 연계는 REST API를 이용한 연계가 기본이며, 데이터 적재를 위한 연계부분과 질의어를 VECTOR 임베딩하는 두 부분으로 구분이 됨



[그림 2-65] 문서 VECTOR 임베딩 및 적재 구조도

- 첫 번째 데이터 적재의 경우 [그림 2-65]과 같이 해외 질병 데이터/백서/역학조사 문서에서 TEXT 추출을 위한 문서 변환과정을 거친 후 MRC의 알고리즘을 바탕으로 VECTOR 임베딩을 진행하고, 각각 적재 INDEX에 맞게 REST API를 호출하여 적재함



[그림 2-66] 질의어 VECTOR 임베딩 및 검색 절차도

- 두번째 검색의 경우 [그림 2-66]과 같이 UI에 질의어가 입력되면 해당 질의어를 VECTOR 임베딩하여 VECTOR 기반 검색 엔진으로 보내면 해당 쿼리에 대한 결과를 저장소로부터 반환받아 UI로 보내는 형태로 구현함

④ KIBANA를 통한 검색 기능 및 시각화

- KIBANA를 통해 검색 기능에 대한 검증 목적으로 DEV TOOL을 계획 하였고 개발단계에서 VECTOR 검색을 통해 결과가 정상적으로 반환되는 일련의 과정을 검증함.
- 자유로운 검색을 위해서는 DSL 쿼리 구조를 익혀야 하는 등의 문제로 적합하지 않다고 판단하여, 양질의 최종 결과물을 위해, 최종 UX/UI 기획의 비중을 확대하여, 좀 더 심도 있는 최종 결과물의 도출함을 우선하기로 함

⑤ 농림축산검역본부 시스템 연계

- 다양한 방면에 영향을 미칠 수 있는 민감한 부분으로, 각 부처 담당자와의 회의를 통해 연계 시스템 및 데이터에 대한 정의에 대한 지속적인 논의 중임.

○ 연계하여 추가적인 데이터 수집이 가능한 형태로 제반 사항에 대해 준비해 놓은 상태임

⑥ 사용자에게 편의성을 제공하는 UI/UX

○ 데이터 및 사용자 분석

- 수집된 데이터

- ▶ 해외 질병 데이터
- ▶ 백서·역학조사보고서
- ▶ Q&A 생성 데이터

를 분석하고, 사용자 면담을 통해 실증적이고 친근한 UI/UX 안을 수립함

○ 컨셉 기획 및 세부 기능 설정

- 사용자 면담을 통해 도출된 주된 요구사항은

- ▶ 질의/응답 란의 영역 분할
- ▶ 필터 기능 구현

이며, 이를 위해

- ▶ 사용자 친화적 UI 구성
- ▶ 데이터 출력 편의성 구현

의 단계적 세부 목표를 수립하여 화면 기획에 착수 함



[그림 2-67] 정보 검색 시스템 프로토타입 UI에 대한 의견 수립 과정

○ 사용자 친화적 UI 구성

- 가장 사용자에게 친근하게 다가갈 수 있는 보편적 UI를 채택하여, 본 연구 과제의 속성 및 형태에 맞게 마이그레이션하여 적용

- 참조 벤치마킹 사이트

- ▶ 마이크로소프트 (WWW.MICROSOFT.COM)
- ▶ 카카오 (WWW.KAKAOCORP.COM)
- ▶ 구글 (WWW.GOOGLE.COM)

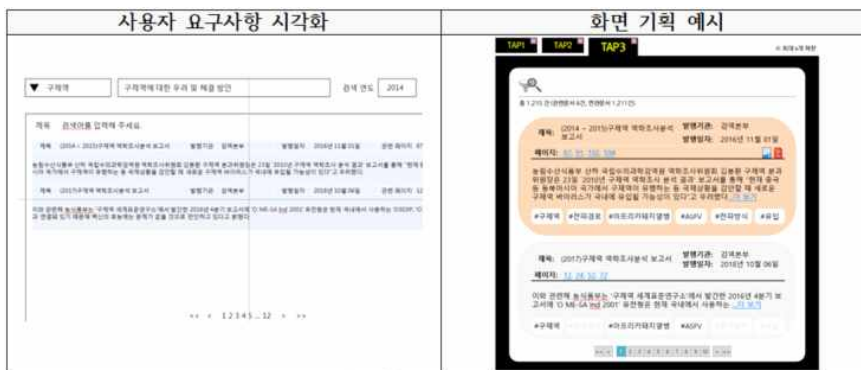
○ 질의: 메신저 형태의 UI 구성



[그림 2-68] 프로토타입 ‘질의’ 기능에 대한 UI 구현 요구사항 및 예시

- ▶ 참조1: 카카오톡에서 서비스하는 어플리케이션 카카오톡의 대화 기능
- ▶ 참조2: 마이크로소프트 공식 홈페이지 내 1:1 문의 기능
- ▶ 사용자 질의 시 해당 질의에 대한 AI 답변을 대화형태로 출력
- ▶ 해당 AI 답변의 근거가 되는 문서와 연관성이 있다고 판단되는 문서의 군집을 간략한 정보와 함께 아이콘화 하여 링크 제공
- ▶ 아이콘 클릭 시 관련/연관 문서 정보는 별도 영역에서 제공

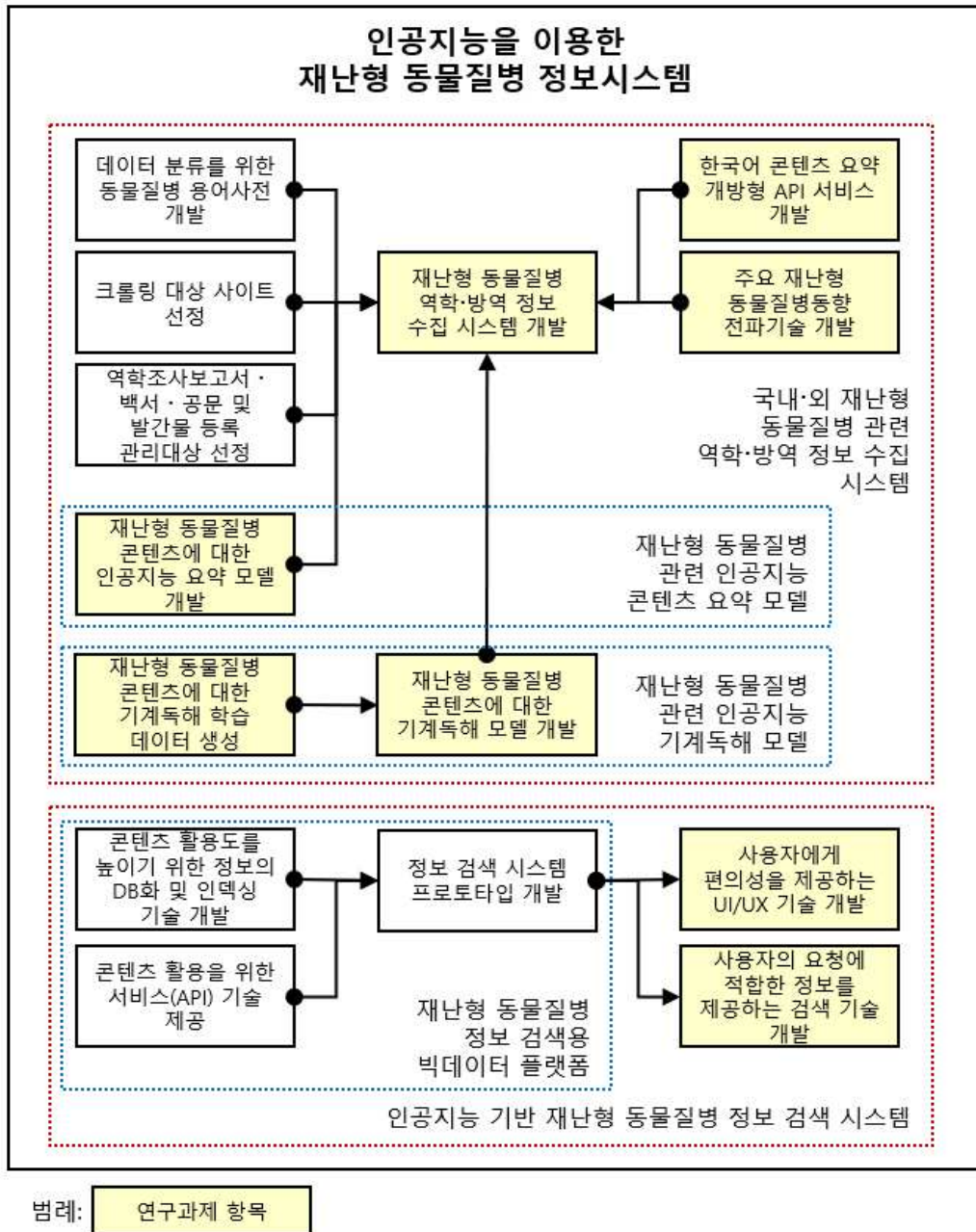
○ 응답: 웹 사이트 형 UI 구성



[그림 2-69] 프로토타입 ‘응답’ 기능에 대한 UI 요구사항 및 예시

- ▶ 참조1: 구글에서 서비스하는 브라우저 크롬
- ▶ 참조2: 구글에서 서비스하는 검색 엔진 UI 형태
- ▶ 답변에 대한 근거(관련) 문서와, 질의에 대한 연관 문서를 해당 영역에 별도 출력
- ▶ 전반적인 UI는 사용자 편의를 위해 구글에서 서비스하는 검색 엔진의 UI를 벤치마킹함

다. 2차년도 연구개발과제의 수행 과정 및 내용



[그림 2-70] 2차년도 연구범위

(1) 재난형 동물질병 관련 역학·방역 정보 수집 시스템 개선

(가) 재난형 동물질병 콘텐츠에 대한 인공지능 요약모델 개선

① 인공지능 요약모델 개선

- 수집문서에 대한 OCR 텍스트 추출 시 발생하는 문장의 분절 문제를 해결함으로써 문장의 의미를 살리는 콘텐츠 요약 모델 개선
 - 형태소 분석기(konply-kkma)를 사용해 문장구조를 분석하여 텍스트의 끊김이 발생 여부를 파악

2. "나는 밥을 먹는다" vs "하늘을 나는 자동차"

두 문장에서 "나는"에 집중해서 본다면, 각 분석기가 태깅할 때 단어의 의미와 주변부를 잘 살피는지 확인해볼 수 있습니다. 첫번째 문장에서 "나는"은 나/N + 는/J, 두번째 문장에서는 나(-ㄹ다)/V + 는/E 이 되는 것이 바람직합니다.

Hannanum	Kkma	Komorán	Mecab	Twitter
나 / N	나 / NP	나 / NP	나 / NP	나 / Noun
는 / J	는 / JX	는 / JX	는 / JX	는 / Josa
밥 / N	밥 / NNG	밥 / NNG	밥 / NNG	밥 / Noun
을 / J	을 / JKO	을 / JKO	을 / JKO	을 / Josa
먹 / P	먹 / VV	먹 / VV	먹 / VV	먹는 / Verb
는다 / E	는 / EPT	는다 / EC	는다 / EC	다 / Eomi

Hannanum	Kkma	Komorán	Mecab	Twitter
하늘 / N	하늘 / NNG	하늘 / NNG	하늘 / NNG	하늘 / Noun
을 / J	을 / JKO	을 / JKO	을 / JKO	을 / Josa
나 / N	날 / VV	나 / NP	나 / NP	나 / Noun
는 / J	는 / ETD	는 / JX	는 / JX	는 / Josa
자동차 / N	자동차 / NNG	자동차 / NNG	자동차 / NNG	자동차 / Noun

[그림 2-71] 형태소 분석기의 문장구조 분석 예시

- 띄어쓰기 교정 모델(py-hanspell)로 텍스트가 단절되어 줄바꿈 처리된 단어 또는 문장의 연결 여부를 판단. py-hanspell은 네이버 맞춤법 검사기를 이용한 파이썬 한글 맞춤법 검사 라이브러리

```
from hanspell import spell_checker

sent = "맞춤법 틀리면 외 안돼? 쓰고싶은대로쓰면돼지 "
spelled_sent = spell_checker.check(sent)

hanspell_sent = spelled_sent.checked
print(hanspell_sent)

맞춤법 틀리면 왜 안돼? 쓰고 싶은 대로 쓰면 돼지
```

[그림 2-72] py-hanspell의 실시 예

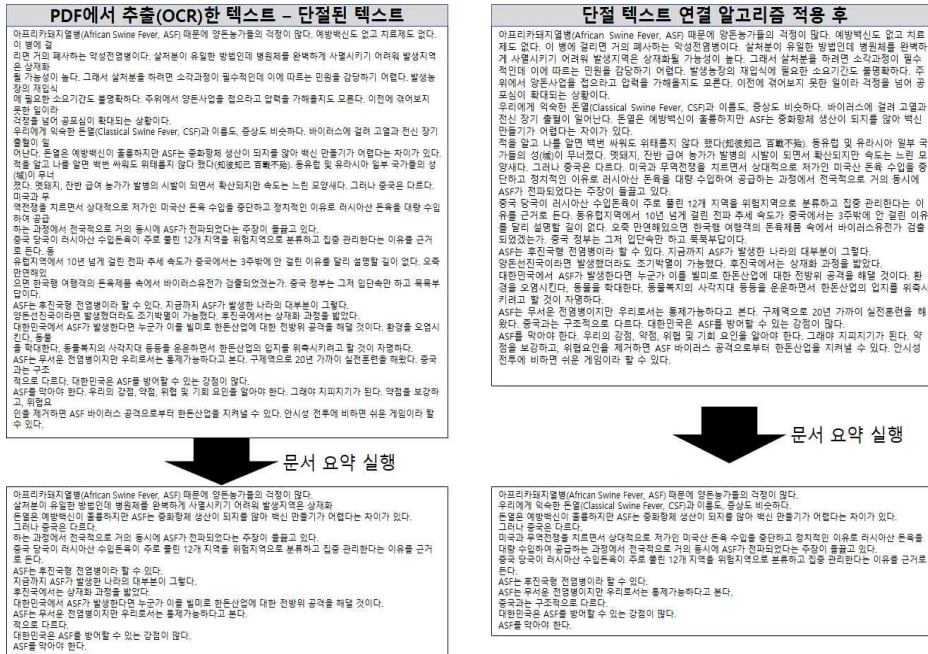
- 수집된 모든 PDF 문서 및 추후 수집될 문서도 해당 알고리즘을 적용

PDF에서 추출(OCR)한 텍스트 - 단절된 텍스트	단절 텍스트 연결 알고리즘 적용 후
<p>아프리카돼지열병(African Swine Fever, ASF) 때문에 양돈농가들의 걱정이 많다. 예방책도 없고 치료제도 없다. 이 병에 걸리면 거의 폐사하는 악성진염병이다. 살처분이 유일한 방법인데 병원을 폐쇄할 수밖에 없는 실정이다. 살처분이 어렵기 때문에 ASF가 발생하면 양돈농가들의 피해는 막대하다. ASF는 돼지에게만 감염되는 바이러스성 질환이다. 돼지에게 감염되면 ASF는 증후군을 일으키는 바이러스성 질환이다. 돼지에게 감염되면 ASF는 증후군을 일으키는 바이러스성 질환이다. 돼지에게 감염되면 ASF는 증후군을 일으키는 바이러스성 질환이다.</p>	<p>아프리카돼지열병(African Swine Fever, ASF) 때문에 양돈농가들의 걱정이 많다. 예방책도 없고 치료제도 없다. 이 병에 걸리면 거의 폐사하는 악성진염병이다. 살처분이 유일한 방법인데 병원을 폐쇄할 수밖에 없는 실정이다. 살처분이 어렵기 때문에 ASF가 발생하면 양돈농가들의 피해는 막대하다. ASF는 돼지에게만 감염되는 바이러스성 질환이다. 돼지에게 감염되면 ASF는 증후군을 일으키는 바이러스성 질환이다. 돼지에게 감염되면 ASF는 증후군을 일으키는 바이러스성 질환이다.</p>

[그림 2-73] 단절 텍스트 연결 알고리즘 적용 실시 예

○ 콘텐츠 문서에 대한 문서 및 챗터 단위의 요약 타당성을 검토한 결과, 챗터 요약은 필요하지 않다고 판단함.

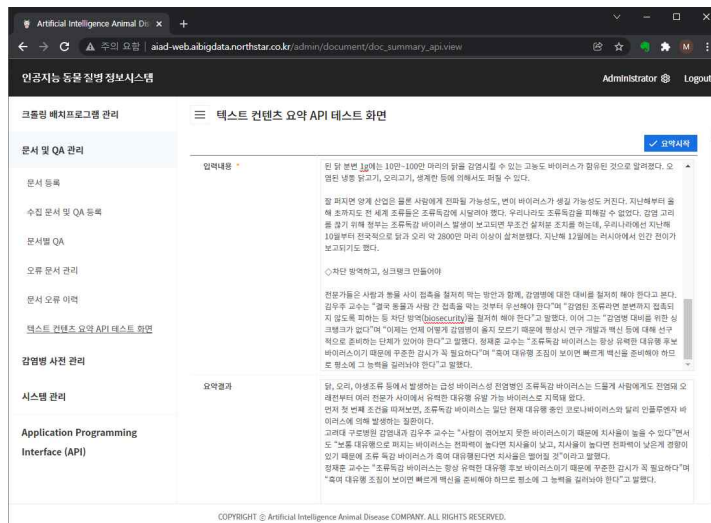
- 챗터가 존재하는 백서의 경우 내용이 많아 문서의 차레가 최적 요약이 되며, 웹 기사의 경우 챗터가 존재하지 않는 단일 문서로 챗터 단위의 요약 프로세스 삭제
- 기계독해 서비스 운영 시 사용자 질문에 대한 후보 문서를 찾는 과정에서도 요약된 텍스트를 활용하지 않고 원문을 활용
- 단절 텍스트 연결 알고리즘을 적용한 문서에 대해 문서 단위의 요약을 재 실행하여 요약의 품질 향상



[그림 2-74] 개선 전 후의 결과 비교

② 한국어 콘텐츠 요약 처리 시스템 구축

○ 동물질병 관련 역학·방역 정보 수집 시스템에 텍스트 콘텐츠 요약 처리 시스템 추가 개발



[그림 2-75] 텍스트 콘텐츠 요약 처리 시스템의 실행 화면

(나) 주요 재난형 동물질병동향 전파기술 개발

① 문자메시지 연동 서비스를 이용한 질병동향 전송 서비스 개발

○ NAVER CLOUD PLATFORM 문자 발송 API를 사용하여 문자메시지 연동 서비스를 이용한 질병동향 전송 서비스 구축

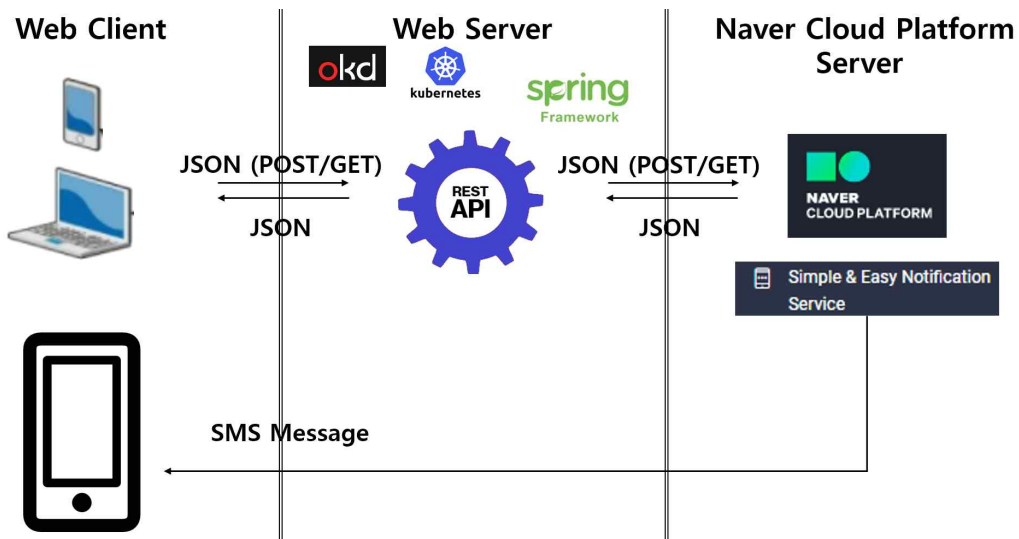
- SMS 보내기 API 주소 : <http://api.aibigdata.northstar.co.kr/api/sendsms.do>
- SMS 보낸 정보 받기 API 주소 : <http://api.aibigdata.northstar.co.kr/api/getsms.do>

ID	IF005										
I/F NAME	SMS 보내기 API										
Request Method	POST										
Parameter Type	RequestBody										
Inbound Information					Outbound Information						
header	Auth-Token : 로그인 인증 토큰 값					header	Response				
Request	JSON					Response	JSON				
필드명	설명	타입	길이	Nullable	비고	필드명	설명	타입	길이	Nullable	비고
subject	제목	varchar	50	N	byte	result	결과	varchar	100	N	T: 정상, F: 실패
message	내용	varchar	1000	N	byte	errorCode	에러코드	varchar	100	Y	
incomingNumberList	수신번호목록	varchar	11	N	전화번호목록 incomingNumberList : ["010 1112222", "010 1112222", "010 1112224"]	message	메시지	varchar	1000	Y	
						HttpStatus	HTTP 상태	varchar	100	N	
						token	로그인 인증 토큰	varchar	1000	N	없음
						data	결과 데이터				
						data.smsSendRequestId	SMS요청ID	varchar		Y	
						data.smsSendRequestDate	SMS요청일자	varchar		Y	

[그림 2-76] SMS 보내기 API 인터페이스 정의서

ID	IF007										
I/F NAME	SMS 받기 API										
Request Method	POST										
Parameter Type	RequestBody										
Inbound Information					Outbound Information						
header	Auth-Token : 로그인 인증 토큰 값					header	Response				
Request	JSON					Response	JSON				
필드명	설명	타입	길이	Nullable	비고	필드명	설명	타입	길이	Nullable	비고
requestId	SMS 요청ID	varchar	200	N		result	결과	varchar	100	N	T: 정상, F: 실패
						errorCode	에러코드	varchar	100	Y	
						message	메시지	varchar	1000	Y	
						HttpStatus	HTTP 상태	varchar	100	N	
						token	로그인 인증 토큰	varchar	1000	N	없음
						data	결과 데이터	varchar			* 응답데이터는 추후 업데이트로 변경하겠습니다
						data.messageList	메시지목록	varchar			{ "messageList": ["...", "status": ""] }
						data.messageListId	수신번호	varchar			
						data.messageList.status	상태	varchar			READY: 대기 PROCESSING: 처리 중 COMPLETED: 처리 완료
						data.messageList.staticMessage	상태에 따라	varchar			
						data.messageList.completeTime	완료일시	varchar			

[그림 2-77] SMS 보낸 정보 받기 API 인터페이스 정의서



[그림 2-78] 문자메시지 연동 서비스 아키텍처

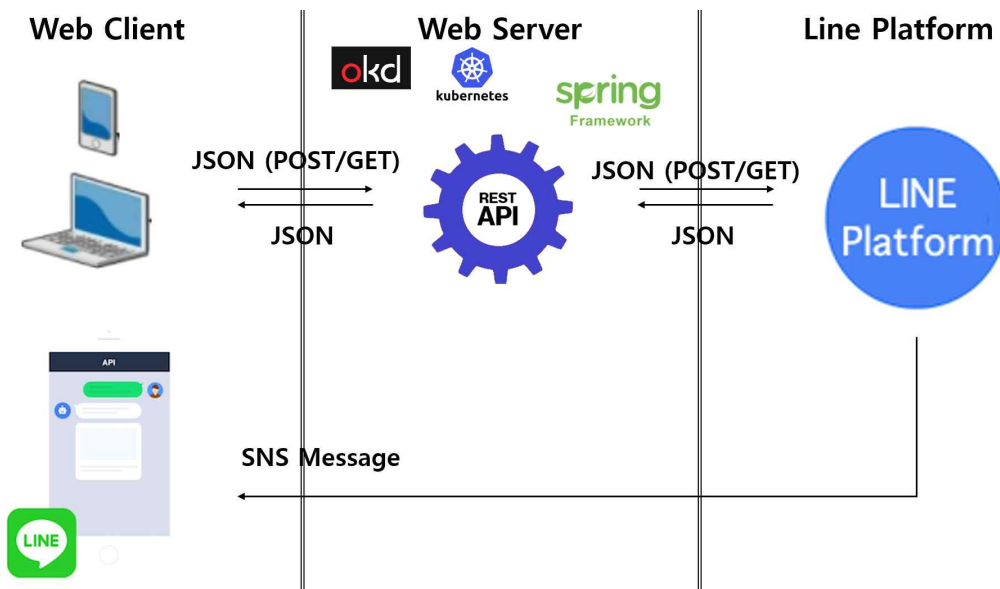
② SNS 연동 서비스를 이용한 질병동향 전송 서비스 개발

○ LINE SNS 메시지 보내기 API를 사용하여 SNS 연동 서비스를 이용한 질병동향 전송 서비스 구축

- LINE SNS NOTIFY 보내기 API 주소: <http://api.aibigdata.northstar.co.kr/api/sendlinenotify.do>

ID	IF008				
V/F NAME	Line SNS notify 보내기 API				
Request Method	POST				
Parameter Type	RequestBody				
Inbound Information			Outbound Information		
header	Auth-Token: 로그인 인증 토큰 값				
Request	JSON				
header	Response				
Response	JSON				
필드명	설명	타입	길이	Nullable	비고
message	메시지	varchar	1000	N	
result	결과	varchar	100	N	T: 정상, F: 실패
errorCode	에러코드	varchar	100	Y	
message	에러메시지	varchar	1000	Y	
httpStatus	HTTP 상태	varchar	100	N	
token	로그인 인증 토큰	varchar	1000	N	없음
data	응답 데이터	varchar			* 응답 데이터는 후속 업데이트로 변경하겠습니다
data.status	상태	varchar			
data.message	메시지	varchar			

[그림 2-79] LINE SMS NOTIFY 보내기 API 인터페이스 정의서



[그림 2-80] SNS 연동 서비스 아키텍처

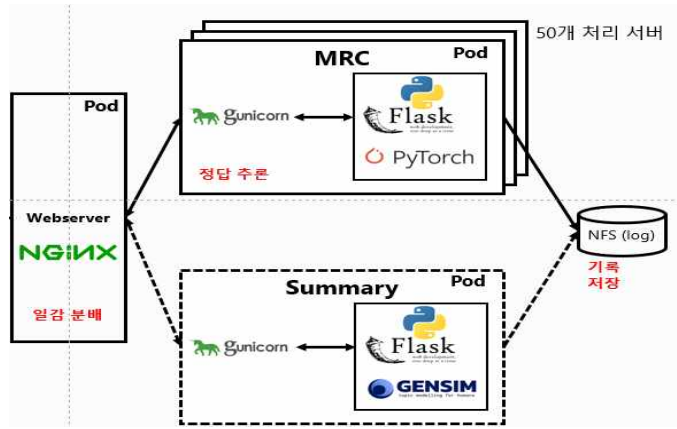
(다) 한국어 콘텐츠 요약 개방형 API 서비스 개발

○ 한국어 콘텐츠 요약 개방형 API 서비스 시스템 구축

- PYTHON GENSIM 요약 패키지, 웹 프레임워크인 FLASK를 사용해 콘텐츠 요약 엔진 개발

○ 개방형 API 서비스의 API 인터페이스 정의

- 콘텐츠 요약 API 주소 : <http://api.aibigdata.northstar.co.kr/api/nlp/summary.do>



[그림 2-81] 요약 API의 아키텍처

ID	IF3036					Inbound Information					Outbound Information						
V/F NAME	콘텐츠 요약 API					Request					Response						
Request Method	POST					header					header						
Parameter Type	RequestBody					application/json					application/json						
header	JSON					JSON					JSON						
Request																	
필드명	설명	타입	길이	Null여부	비고	필드명	설명	타입	길이	Null여부	비고	필드명	설명	타입	길이	Null여부	비고
content	요약할 내용	varchar	100000	N	300자 이상	statusCode	상태코드	varchar	100	Y		statusCode	상태코드	varchar	100	Y	
contentCode	요약 할 내용의 코드 (ex. hash)	varchar	100	N	sha256 hash	message	메시지	varchar	1000	Y		message	메시지	varchar	1000	Y	
						data	결과 데이터					data	요약 결과	varchar	100000	Y	
						data.summary	요약 요약	varchar	100000	Y		data.summary	요약 요약	varchar	100000	Y	
						data.contentCode	입력된 할 내용의 코드(ex. hash)	varchar	100	Y		data.contentCode	입력된 할 내용의 코드(ex. hash)	varchar	100	Y	

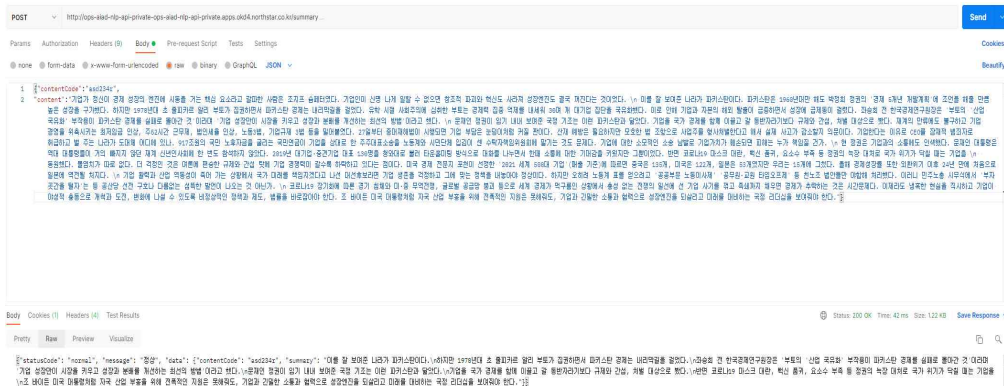
[그림 2-82] 요약 API 인터페이스 정의서

- 동물질병에 국한되지 않고 모든 분야에 대한 콘텐츠 요약 기능 개발
 - 콘텐츠 요약 엔진에서 수행되는 모든 로그를 NFS서버에서 관리하는 실시간 로그 기능 개발
 - KUBERNETES 기반의 콘텐츠 요약 엔진 및 NGINX 서버 구축 운영



[그림 2-83] KUBERNETES 서버에 배포 및 운영 상태의 요약 API

- 콘텐츠 요약 API 인터페이스 테스트 및 동작 검증



[그림 2-84] 요약 API의 실시 예

(2) 재난형 동물질병 콘텐츠에 대한 인공지능 기계독해 모델 개발

(가) 재난형 동물질병 콘텐츠에 대한 기계독해 학습 데이터 생성·검증

- 크롤링을 통해 수집된 동물질병 관련 뉴스 기사, 백서, 보도사항 등의 자료에서 역학 및 방역 지식에 대한 질의응답 데이터를 본 과제에서 개발한 WEB 시스템을 통해 문서별 질의응답 데이터를 손쉽게 제작 및 관리

QA 목록

번호	문서명	QA개수	자료유형	저자	출처	분석일자	검열명 키워드	문서상태	사용여부
①	통계리서 ASF 첫 발생	4	보도자료	양한은	http://www.pigtimes.co.kr/news/...	2021.10.07 09:12:26	돼지질병, 아프리카돼지열병, 아프리카...	정상	사용
①	조류인플루엔자 긴급행동지침	250	지침	농림축산식품부		2021.10.07 09:12:26	구제, 비치, 구제책, 우해제, 주역리, 뉴...	오류	사용
①	제주 구제역 형제 미묘 농가, 도축도 끝...	2	보도자료	관리자	http://pigpeople.net/news/article...	2021.10.07 09:12:26	구제역	오류	사용
①	인명당 국제우편물을 통한 수입금지 조...	2	보도자료	이혜영	http://www.qia.go.kr/newsweb/qia...	2021.10.07 09:12:26	구제역, 돼지질병, 아프리카돼지열병, ...	정상	사용
①	육류종제 시행관련 지침서	46	지침	농촌진흥청 국립축산과학원		2021.10.07 09:12:26	구제, 농림축산, Atrophic rhinitis, Ja...	오류	사용
①	유리팩토리 사육지침서(보통 실용오리)	24	지침	농촌진흥청 국립축산과학원		2021.10.07 09:12:26	구제, 구제책, 마백병, 돼지인두, 돼지열...	오류	사용
①	유리팩토리의 사양관리 지침서(보통 실용오리)	82	지침	농촌진흥청 국립축산과학원		2021.10.07 09:12:26	구제, 구제책, 마백병, 돼지인두, 돼지열...	오류	사용
①	외국인근로자용 영문판 매뉴얼	20	매뉴얼	농림축산식품부		2021.10.07 09:12:26	구제역, 돼지우역형성상, 돼지전염병위...	오류	사용
①	양돈포류(말세)에 의한 구형발생조류...	70	보고서	한국환경생태연구소, 농림축산식품부		2021.10.07 09:12:26	구제, 농림축산, 조류인플루엔자, 오리...	오류	사용
①	영문판 돼지 ASF-8002년 넘어서	2	보도자료	김민규	http://www.pigtimes.co.kr/news/...	2021.11.07 21:35:32	기동지, 마백병, 돼지질병, 아프리카...	정상	사용
①	아프리카돼지열병(ASF) 역학조사원 매...	302	보고서	농림축산식품부 역학조사과		2021.10.07 09:12:26	구제, 유행, 천치, 구제책, 돼지질병, 브...	오류	사용
①	아프리카돼지열병(ASF) 방역 추진상황	16	보도자료	구제역방제과	https://www.mafra.go.kr/bbs/mnf/...	2021.11.08 00:57:02	돼지질병, 아프리카돼지열병, African...	오류	사용
①	아프리카돼지열병 긴급행동지침(SOP)	402	지침	농림축산식품부		2021.10.07 09:12:26	구제, 우해, 광견병, 구제책, 우해제, 돼...	오류	사용
①	아프리카돼지열병 Q&A	10	가이드	농림축산식품부 동물위생안전과		2021.10.07 09:12:26	구제책, 농림축산, 농림축산, 돼지질병, ...	오류	사용
①	돼지독위 아프리카돼지열병 생애와 차...	128	매뉴얼	Vittorio Gaburri, Sergei Khomenko...	Administrator	2021.10.07 09:12:26	구제, 비치, 구제책, 기동지, 마백병, 돼...	오류	사용

Total Count : 39

[그림 2-85] 문서별 생성된 질의응답 현황

- 수집된 자료의 PDF 원문과 PDF에서 추출된 내용을 텍스트화하여 제공함으로써, 질의응답 데이터 생성자의 작업 속도 향상

The figure shows two side-by-side screenshots. The left screenshot is a PDF viewer displaying a document with Korean text. The right screenshot is a web browser showing the same document's content extracted into a structured HTML format. The extracted content includes a title, a list of bullet points, and a table of contents.

[그림 2-86] PDF 원문과 추출된 텍스트 내용 제공

- 질의응답 데이터 작업자의 실수 방지를 위해 답변 및 답변문장이 질문에 존재하는지를 확인 및 데이터의 품질을 점검
- 기계독해 학습 데이터 사용 가능 여부를 확인하는 과정을 AIRFLOW로 자동화하여 매주 오류 목록과 내용을 데이터 생성자 및 관리자에게 아래와 같이 전달

페이지 번호 *	6
답변유형 *	대상
지문(View) <input type="button" value="지문편집하기"/>	II 병인체 1. 바이러스의 특성 및 항원성 조류인플루엔자 바이러스(AIV)는 Orthomyxoviridae 과(과, Family), A형 Influenza virus 속(屬, Genus)으로 분류되며, negative sense RNA 바이러스로서 다른 종류의 RNA 바이러스와는 달리 서로 다른 8개의 RNA 분절(segment)로 구성되어 있다. 이들은 혈구응집소(hemagglutinin, HA)와 뉴라미니다제(neuraminidase, NA)와 표면항원 유전자인 M, NP, PB2 등 6개의 내부유전자(internal gene)로 나뉘어진다. 병원성은 주로 HA 유전자와 관련이 있으며 HA 단백질 분절부위에 특정한 병원성 관련 유전자 배열을 나타내면 고병원성으로 간주되고 있다. 감염속주 특이성과 관련이 가장 많은 유전자는 HA 및 NA 유전자이다 다른 내부 유전자들도 복합적으로 관련되어 있다. 인플루엔자바이러스의 혈청형은 크게 3종(A, B, C형)으로 분류되며, 그 중 B형과 C형은 사람에게 감염되고, A형 바이러스는 사람을 비롯하여 닭·칠면조·아생오리·돼지·말·잉크·꿀개 등 다양한 종류의 척추동물에 감염된다. A형 A 바이러스는 다양한 아형(subtype)이 있는데 바이러스 표면에 존재하는 혈구응집소의 특성에 따라 H1부터 H16까지 16종이 있으며, 뉴라미니다제라는 효소가 나타내는 표면 단백질의 특성에 따라 N1부터 N9까지 9종의 아형으로 구분된다. 따라서 H형과 N형을 조합할 경우, A형 인플루엔자 바이러스는 이론적으로 총 144종(=16×9)의 아형이 존재하게 된다. 속주의 종류에 따라서 감염될 수 있는 AI 바이러스의 아형에는 상당한 차이가 있다. 예를 들어, 조류에는 144종의 아형 모두가 감염될 수 있지만 조류의 종속에 따라 감염성과 질병 발현여부는 각기 다르다. 야생조류 중 특히 오리, 도요새 등의 물새류는 감염성이 높은 편이어서 AI 바이러스에 감염되면 임상중상이 없이도 상당량의 바이러스를 체외로 배출한다. 돼지에서는 H1 및 H3형이 주로 감염되며, 사람에게 감염되는 H 아형은 H1, H2, H3에 국한되었지만 근래에 들어 H5, H7 및 H9형으로 확대되고 있는 추세에 있다. 또한 동일한 아형의 바이러스라 할지라도 그 바이러스의 유전자 배열에 따
질문(22) *	A형 바이러스가 감염될 수 있는 개체는?
답변(12) *	다양한 종류의 척추동물
답변문장(124)	인플루엔자바이러스의 혈청형은 크게 3종(A, B, C형)으로 분류되며, 그 중 B형과 C형은 사람에게 감염되고, A형 바이러스는 사람을 비롯하여 닭·칠면조·아생오리·돼지·말·잉크·꿀개 등 다양한 종류의 척추동물에 감염된다.

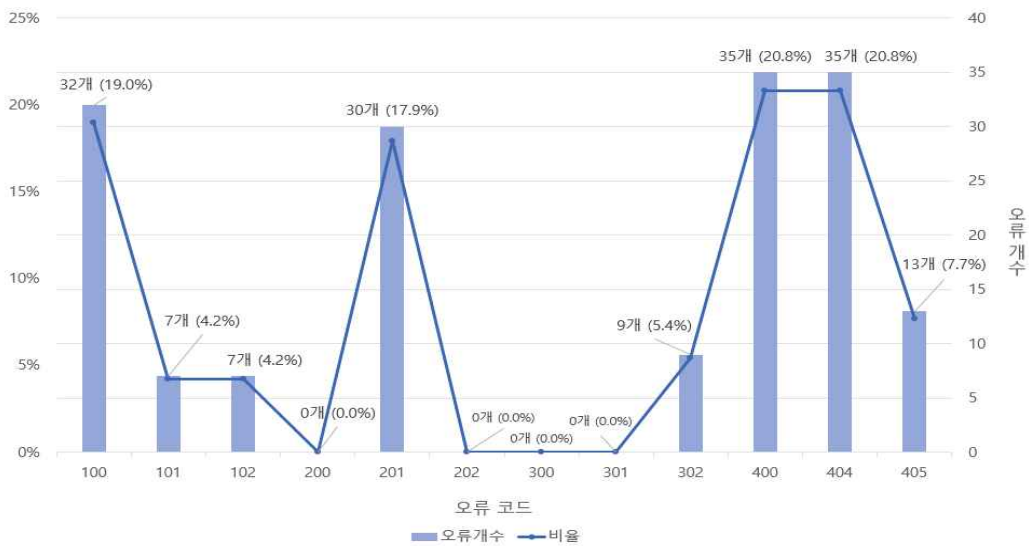
[그림 2-87] 질의응답 데이터 생성(답변 및 점검)

A	B	C	D	E	F	G	H	I	J	K	L	M	N
번호	정답일자	문서이름	QA생성지	QA생성일	QA수정지	QA수정일	시행일자	에코코드	제러메시지	지문	질문	답변	답변 근거
0										제 4장 구제역 미발상 시 표준행동요령			
										1. 일시 표준행동요령 1. 방역기관 및 우체국 농장은 구제역 발생의 사전예방을 위한 조치 강구 가. 구제역의 예방과 조기발견 및 신속한 신고를 위한 체계 유지 나. 농식품부 질병관리본부(지체·법외(단체) 등 방역기관은 구제역 발생 시 질병관리 당국을 위해 긴급한 협조체계를 구축하여 신속한 대비태세를 갖추어야 하며, 조동태세를 위한 각 기관별 방역기구 구성과 운영요령 숙달이 필요 다. 우체국농장 소유지 및 종사자는 구제역의 중요도를 이해하고 실천할 수 있도록 관련 규정을 준수하고 자진방역을 적극 실천 2. 기관 및 단체별 조치사항 가. 농림축산식품부 1) 구제역 방역대책 수립시달 가) 전국 방역기관의 비상 방역태세 유지 나) 가장방역준비 실시(연 1회 이상) 다) 질병본부, 방역지원본부, 농협양돈제주, 생신자협회 등에서 방역교육 토목 조치 2) 방역대책 상황 평가 및 대책 마련 가) 농식품부지자체 구제역 방역대책 실시 결과 평가 * 우체국 관련 종사자(방역실력 및 방역대책에 대한 평가 및 개선점 발굴 나) 구제역 예방방역 7개 중 5개 이상 및 개선점 발굴 3) 구제역 예방을 위하여 정기 백신접종 소요되는 예방안 수급계획 마련 제 4장 구제역 미발상 시 표준행동요령			
1	2021-08-15	긴급행동지침(SCP)	yj6669289	0422:50.4	yj6669289	0422:50.4	47	201	3	1. 일시 표준행동요령 1. 방역기관 및 우체국 농장은 구제역 발생의 사전예방을 위한 조치 강구 가. 구제역의 예방과 조기발견 및 신속한 신고를 위한 체계 유지 나. 농식품부 질병관리본부(지체·법외(단체) 등 방역기관은 구제역 발생 시 질병관리 당국을 위해 긴급한 협조체계를 구축하여 신속한 대비태세를 갖추어야 하며, 조동태세를 위한 각 기관별 방역기구 구성과 운영요령 숙달이 필요 다. 우체국농장 소유지 및 종사자는 구제역의 중요도를 이해하고 실천할 수 있도록 관련 규정을 준수하고 자진방역을 적극 실천 2. 기관 및 단체별 조치사항 가. 농림축산식품부 1) 구제역 방역대책 수립시달 가) 전국 방역기관의 비상 방역태세 유지 나) 가장방역준비 실시(연 1회 이상) 다) 질병본부, 방역지원본부, 농협양돈제주, 생신자협회 등에서 방역교육 토목 조치 2) 방역대책 상황 평가 및 대책 마련 가) 농식품부지자체 구제역 방역대책 실시 결과 평가 * 우체국 관련 종사자(방역실력 및 방역대책에 대한 평가 및 개선점 발굴 나) 구제역 예방방역 7개 중 5개 이상 및 개선점 발굴	구제역 의심가축 발견 시 신고해야 할 기관 전화번호는 무엇인가요?	1589-9060	나) 농림축산검역본부(1589-9060)
										1. 일시 표준행동요령 1. 방역기관 및 우체국 농장은 구제역 발생의 사전예방을 위한 조치 강구 가. 구제역의 예방과 조기발견 및 신속한 신고를 위한 체계 유지 나. 농식품부 질병관리본부(지체·법외(단체) 등 방역기관은 구제역 발생 시 질병관리 당국을 위해 긴급한 협조체계를 구축하여 신속한 대비태세를 갖추어야 하며, 조동태세를 위한 각 기관별 방역기구 구성과 운영요령 숙달이 필요 다. 우체국농장 소유지 및 종사자는 구제역의 중요도를 이해하고 실천할 수 있도록 관련 규정을 준수하고 자진방역을 적극 실천 2. 기관 및 단체별 조치사항 가. 농림축산식품부 1) 구제역 방역대책 수립시달 가) 전국 방역기관의 비상 방역태세 유지 나) 가장방역준비 실시(연 1회 이상) 다) 질병본부, 방역지원본부, 농협양돈제주, 생신자협회 등에서 방역교육 토목 조치 2) 방역대책 상황 평가 및 대책 마련 가) 농식품부지자체 구제역 방역대책 실시 결과 평가 * 우체국 관련 종사자(방역실력 및 방역대책에 대한 평가 및 개선점 발굴 나) 구제역 예방방역 7개 중 5개 이상 및 개선점 발굴	우체국 농장의 종사 업무에 있는 발판소독소의 소독액은 얼마마다 관리되어야 하나요?	소독액은 2~3일에 한번씩 교체하며 유기를 오염 시키지 않아야 함	종사 인규에는 발판소독소가 실시되어야 하며 소독액은 2~3일에 한번씩 교체하며 유기를 오염 시키지 않아야 함

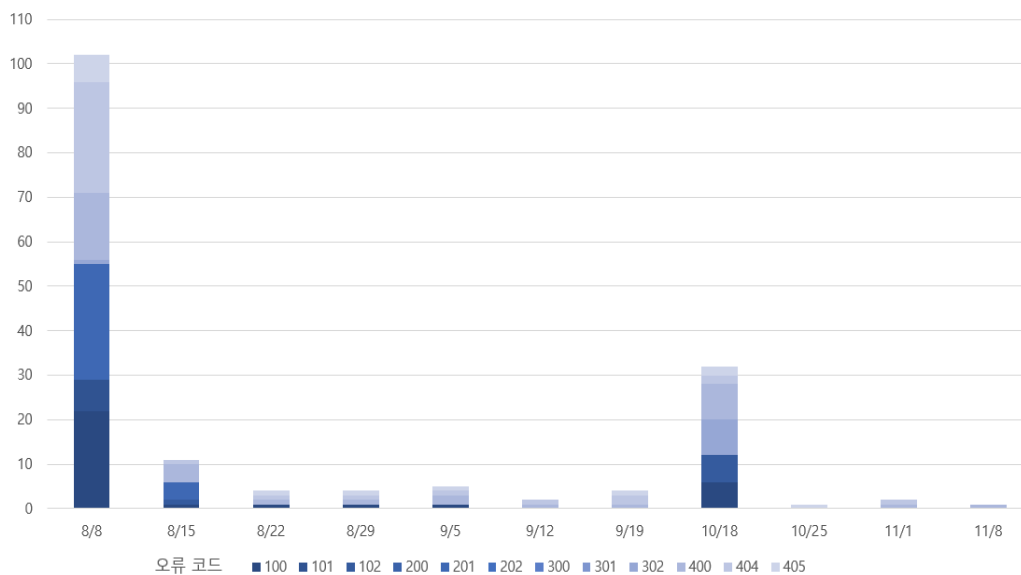
[그림 2-88] 데이터 생성자 및 관리자에게 제공된 오류 목록

분류	오류코드	오류메시지
정답	100	정답이 지문의 내용과 정확하게 일치하지 않음
	101	정답이 너무 길니다(200자 이상)
	102	정답이 너무 짧습니다(1자 이하)
지문	200	지문이 올바르지 않습니다
	201	지문이 너무 길니다(3000자이상)
	202	지문이 너무 짧습니다(100자 이하)
질문	300	질문이 올바르지 않습니다
	301	질문이 너무 길니다(200자 이상)
	302	질문이 너무 짧습니다(5자 이하)
정답근거	400	정답근거가 지문의 내용과 정확하게 일치하지 않음
	404	정답 근거에 정답이 포함되어 있지 않습니다
	405	중복된 정답근거가 지문에 존재합니다.

[그림 2-89] 학습데이터 오류 코드



[그림 2-90] 오류코드별 학습데이터 오류발생 현황



[그림 2-91] 일자별 학습데이터 오류발생 현황

○ 검수 완료된 질의응답 데이터의 DB 저장

seq	doc_id	qa_seq	page_num	doc_contents	question	answer	doc_answer	answer_type	create_id
1	654951184a01cbda4d	1	24	2.2. 육체구조 파악을 통한 이상징후 식별 이상징후 육체구조 파악하고 난 후 동물상태를 최소승물다각형법 (Minimum Convex Polygon Method)	동물과 분석을 국제적으로 통용되는 방법인 최소승물다각형법	50	manhia38		
2	654951184a01cbda4d	15	30	나. 이상징후를 통한 HPAI 바이러스 유입 가능성에 대한 이상징후를 통한 HPAI 바이러스 유입 가능성에 대한 Bayesian 기법	한국감염병국립수의과 국립수의과학검역원 등 국내 기관에	10	manhia38		
3	654951184a01cbda4d	21	12	0. 바이러스의 유역 A항 바이러스는 사육을 비축하며, B 바이러스는 주로 이월개 출몰 수 있으나?	주로 직접 접촉에 의해 전파	50	manhia38		
4	672143d27e4530543	63	43	2. 임상증상 발현과 감염률 추정(2018년 국제역인 3.1) 대한이론 무엇인가?	미생물 및 바이러스학의 이론	10	y6669299		
5	af1bb8e36a158a6d9	54	31	계양역학조사 및 분석은 개요(1) 중앙역학조사(1) 역학조사 실시, 역학조사 협조에 관련된 근거는 어떤 법?	가축전염병예방법 제13조(역학조사)	10	roren9797		
6	672143d27e4530543	65	44	- 물시에 농장주 이동이 일어나면 사육을 관리하였으나 경기도 일대에서 2차 이상 발생된 농장은 누가 관리하였는지 외국인 근로자가 농장을 관리하고, 축주는 2차 농장을	○ 경기 일대서 발생한 ○2차 농장 관리농장(2차 발생)	10	y6669299		
7	654951184a01cbda4d	14	30	나. 이상징후를 통한 HPAI 바이러스 유입 가능성에 대한 이상징후를 통한 HPAI 바이러스 유입 가능성에 대한 Andrus 방법	저온 냉동 보관된 GS 95는 동부 동물을 기온도프지	10	manhia38		
8	940416389776899644a	112	72	4. 오리 사육 특성(1) 오리 사육 특성(1) 오리 사육 특성(1) 오리 사육 특성(1) 오리 사육 특성(1)	오리는 국내에서 사육	10	manhia38		
9	654951184a01cbda4d	24	60	3.3. 가금농장의 이상징후를 접근성 분석하기 위해 가금 농장 이상징후가 가금 농장 이상징후가 가금 농장 이상징후가 가금 농장 이상징후가	수컷이 평균 94.8% 이상 출현한 농장	30	manhia38		
10	940416389776899644a	42	26	3.3. 가금농장의 이상징후를 접근성 분석하기 위해 가금 농장 이상징후가 가금 농장 이상징후가 가금 농장 이상징후가 가금 농장 이상징후가	수컷이 평균 94.8% 이상 출현한 농장	30	manhia38		
11	940416389776899644a	46	28	2. 우리나라 고병원성조류인플루엔자 발생현황(1) 국내 HPAI 발생 시 분리를 위한 바이러스의 특성 및 역학사항	농업과학기술원(NAHRI)에서	10	roren9797		
12	654951184a01cbda4d	23	60	3.3. 가금농장의 이상징후를 접근성 분석하기 위해 가금 농장 이상징후가 가금 농장 이상징후가 가금 농장 이상징후가 가금 농장 이상징후가	수컷이 평균 94.8% 이상 출현한 농장	30	manhia38		
13	af1bb8e36a158a6d9	51	50	3. 임상증상 분석결과(축주) 가금(H5N6) 발생 농장에서 유출된 농장(2) 국내 H5N6 발생 농장에서 유출된 농장(2) 국내 H5N6 발생 농장에서 유출된 농장(2)	신경성 또는 유전적 동원(23.2%)	10	roren9797		
14	940416389776899644a	47	28	3. 우리나라 고병원성조류인플루엔자 발생현황(1) 국내 HPAI 발생 시 분리를 위한 바이러스의 특성 및 역학사항	농업과학기술원(NAHRI)에서	10	roren9797		
15	940416389776899644a	14	14	2. 조류인플루엔자 바이러스가 인플루엔자 바이러스 A형 인플루엔자 바이러스는 인플루엔자 중 몇 가지?	인플루엔자 A형 인플루엔자 B형 인플루엔자 C형 인플루엔자 D형 인플루엔자	10	manhia38		
16	940416389776899644a	55	35	1. 역학조사 목적(1) 역학조사 목적(1) 역학조사 목적(1) 역학조사 목적(1) 역학조사 목적(1)	역학조사 목적(1) 역학조사 목적(1) 역학조사 목적(1) 역학조사 목적(1) 역학조사 목적(1)	10	manhia38		
17	940416389776899644a	112	72	4. 오리 사육 특성(1) 오리 사육 특성(1) 오리 사육 특성(1) 오리 사육 특성(1) 오리 사육 특성(1)	오리는 국내에서 사육	10	manhia38		
18	654951184a01cbda4d	21	60	3.3. 가금농장의 이상징후를 접근성 분석하기 위해 가금 농장 이상징후가 가금 농장 이상징후가 가금 농장 이상징후가 가금 농장 이상징후가	수컷이 평균 94.8% 이상 출현한 농장	30	manhia38		
19	940416389776899644a	11	18	5. 2014년 국제역인 2010/2011년 발병 이후 관련 국제적 정황(1) 국제적 정황(1) 국제적 정황(1) 국제적 정황(1) 국제적 정황(1)	국제적 정황(1) 국제적 정황(1) 국제적 정황(1) 국제적 정황(1) 국제적 정황(1)	10	y6669299		
20	654951184a01cbda4d	49	81	3. 임상증상 분석결과(축주) 가금(H5N6) 발생 농장에서 유출된 농장(2) 국내 H5N6 발생 농장에서 유출된 농장(2) 국내 H5N6 발생 농장에서 유출된 농장(2)	신경성 또는 유전적 동원(23.2%)	10	roren9797		
21	654951184a01cbda4d	22	60	3.3. 가금농장의 이상징후를 접근성 분석하기 위해 가금 농장 이상징후가 가금 농장 이상징후가 가금 농장 이상징후가 가금 농장 이상징후가	수컷이 평균 94.8% 이상 출현한 농장	30	manhia38		

[그림 2-92] 데이터베이스에 저장된 동물질병 학습 데이터

○ 최종 10,218건의 QA셋으로 동물질병 기계독해 데이터셋(ADQuAD) 생성

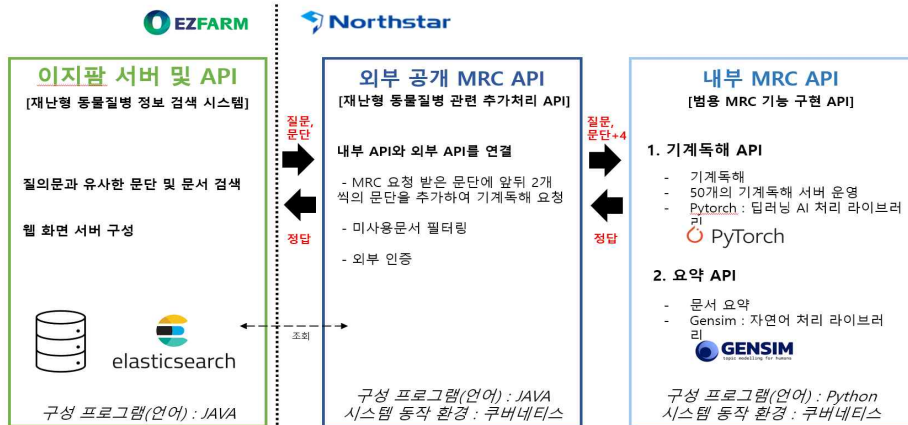
컬럼명	#	Data type	Length	Precision	Scale	Identity	Not Null	Comment	Collation
doc_id	1	varchar	20	20			[X]	문서ID	default
qa_seq	2	numeric					[X]	QA SEQ	
page_num	3	varchar	10	10			[X]	페이지번호	default
doc_contents	4	text					[X]	지문	default
question	5	text					[X]	질문	default
answer	6	text					[X]	답변	default
doc_answer	7	text					[]	답변이 포함된 문장	default
answer_type	8	varchar	10	10			[X]	답변유형	default
create_id	9	varchar	35	35			[X]	등록자	default
create_date	10	timestampz			35	6	[X]	등록일자	
update_id	11	varchar	35	35			[X]	수정자	default
update_date	12	timestampz			35	6	[X]	수정일자	
doc_name	13	varchar	200	200			[X]	문서(파일)명	default
doc_hash_id	14	varchar	200	200			[]	지문 Hash ID	default

[그림 2-93] 데이터베이스에 저장된 동물질병 질의응답 데이터 구조

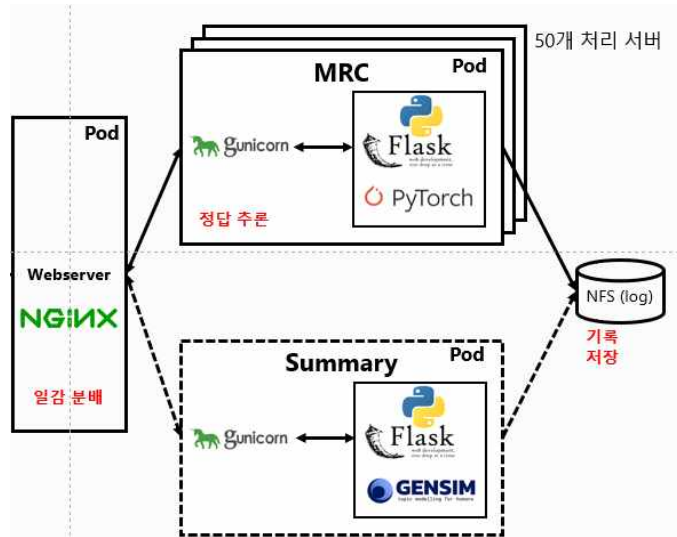
(나) 재난형 동물질병 콘텐츠에 대한 기계독해 모델 개발·검증

① 인공지능을 활용한 콘텐츠 독해모델 생성 기술을 위한 질의응답 시스템 구축

○ 동물질병에 대한 역학, 방역 지식에 대한 한국어 질의응답 처리 시스템 구축을 위해 범용 MRC 서비스 API 시스템과 재난형 동물질병 질의응답용 MRC API를 분리하여 시스템 구축



[그림 2-94] 동물질병 정보 한국어 질의응답 처리 시스템 구성도



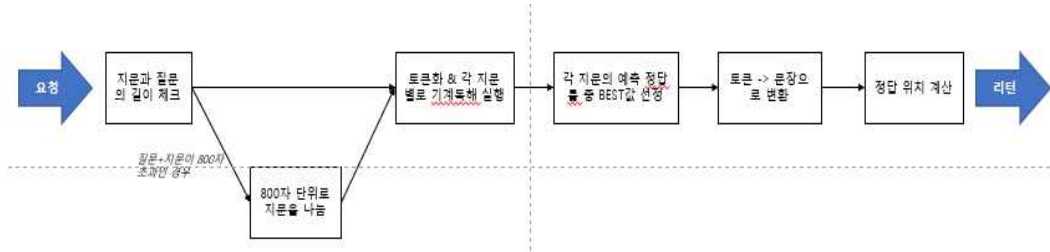
[그림 2-97] 동물질병 질의응답 API 서비스 구성도

- PYTORCH의 딥러닝 프레임워크, 학습된 KoELECTRA MRC모델, 웹 프레임 워크인 FLASK를 사용해 기계독해 엔진 개발
- API 사용자가 질문, 지문을 입력하면 질문에 대한 정답을 지문에서 찾아 정답, 정답위치, 정답확률을 제공을 제공하며, API 인터페이스 정의

ID	JF004										
I/F NAME	기계독해 API										
Request Method	POST										
Parameter Type	RequestBody										
Inbound Information					Outbound Information						
header	Auth-Token : 로그인 인증 토큰 값					header	JSON				
Request	JSON					Response	JSON				
필드명	설명	타입	길이	Null여부	비고	필드명	설명	타입	길이	Null여부	비고
docId	문서ID	varchar	156	N		result	결과	varchar	100	N	T:정상, F:실패
sentenceIndex	문장Index	varchar		N	1	errorCode	에러코드	varchar	100	Y	
sentence	문장	varchar	10000	N	우리나라에서 처음으로 발생한.....	message	에러메시지	varchar	1000	Y	
question	질문	varchar		N	우리나라 구제역 발생지는?	httpStatus	HTTP 상태	varchar	100	N	
questionHash	질문해시	varchar		N	sha256 hash	token	로그인 인증 토큰	varchar	1000	N	없음
similarity	유사도	varchar		N	80	data	결과 데이터				
						data.docId	문서ID	varchar	150	N	
						data.sentenceIndex	문장Index	varchar		N	1
						data.questionHash	질문해시	varchar		N	sha256 hash
						data.answer	정답	varchar		Y	
						data.start	정답 시작 위치	varchar	3	Y	1
						data.end	정답 끝 위치	varchar	3	Y	2
						data.probability	정답확률	varchar		Y	34.25
						data.docUseYn	문서사용여부	varchar	1	N	Y/N
						data.pageNumber	페이지번호	varchar	10	Y	10
data.end가 data.sentenceIndex에 문자열 길이보다 클 경우, 다음 문단계 정답이 연결된 것입니다. data.end은 두 문단을 합했을 때 정답이 온 위치입니다.											
docId, sentenceIndex, sentence는 엘라스틱 서치의 aiaid_paragraph인덱스의 doc_id_STR, paragraph_number_INT, paragraph_text_STR의 정보형 필명											

[그림 2-98] 기계독해 API 인터페이스 정의서

- 1개의 기계독해 요청에 대한 처리 속도는 약 0.5초이며, 다수의 요청을 신속하게 처리할 수 있도록 50개의 서버를 운영
- 동물질병에 대한 기계독해에 특화되어 있지만, 일반 상식 분야에 대한 기계독해도 가능하도록 모델을 학습 및 적용하여 기계독해 엔진의 범용성을 향상
- KoELECTRA 모델은 최대 512 토큰(단어)밖에 처리하지 못하는 단점을 보완하여 토큰의 수의 제한 없이 정확하게 기계독해를 처리할 수 있는 알고리즘을 개발



[그림 2-99] 기계독해 API 내부 프로세스

- 기계독해 엔진을 S3와 연동해, 최신 기계독해 모델을 실시간 서비스할 수 있도록 구축
- 기계독해 엔진에서 수행되는 모든 로그를 NFS서버에서 관리함으로써 실시간 로그 파악 가능
- 쿠버네티스를 사용함으로써 다수의 기계독해 요청에도 손쉽게 서버를 증설할 수 있어 트래픽 관리가 용이
- 기계독해 엔진 및 NGINX는 KUBERNETES에 배포하여 운영

P aiad-nlp-api-private-mrc-854466897c-wwtq8	NS ops-aiad-nlp-api-private	Running	1/1	RS aiad-nlp-api-private-mrc-854466897c
P aiad-nlp-api-private-mrc-854466897c-x4r22	NS ops-aiad-nlp-api-private	Running	1/1	RS aiad-nlp-api-private-mrc-854466897c
P aiad-nlp-api-private-mrc-854466897c-xlsmm	NS ops-aiad-nlp-api-private	Running	1/1	RS aiad-nlp-api-private-mrc-854466897c
P aiad-nlp-api-private-mrc-854466897c-z8rcw	NS ops-aiad-nlp-api-private	Running	1/1	RS aiad-nlp-api-private-mrc-854466897c
P aiad-nlp-api-private-mrc-854466897c-zmct9	NS ops-aiad-nlp-api-private	Running	1/1	RS aiad-nlp-api-private-mrc-854466897c
P aiad-nlp-api-private-summary-77b6cdd4bf-j8c8s	NS ops-aiad-nlp-api-private	Running	1/1	RS aiad-nlp-api-private-summary-77b6cdd4bf
P nginx-644b5fc7b6-rrgvd	NS ops-aiad-nlp-api-private	Running	1/1	RS nginx-644b5fc7b6
P ops-aiad-nlp-api-private-runner-gitlab-runner-65d55f599c-9p4fk	NS ops-aiad-nlp-api-private	Running	1/1	RS ops-aiad-nlp-api-private-runner-gitlab-runner-65d55f599c

[그림 2-100] KUBERNETES 서버에 배포 및 운영 상태의 기계독해 API

```

POST http://ops-aiad-nlp-api-private-ops-aiad-nlp-api-private.apps.ok4.norlabs.co.ke/req
Content-Type: application/json
{
  "question": "구독자 배려나스 여행은 속 날 누가 존재나스?",
  "sentence": "123",
  "id": "123"
}

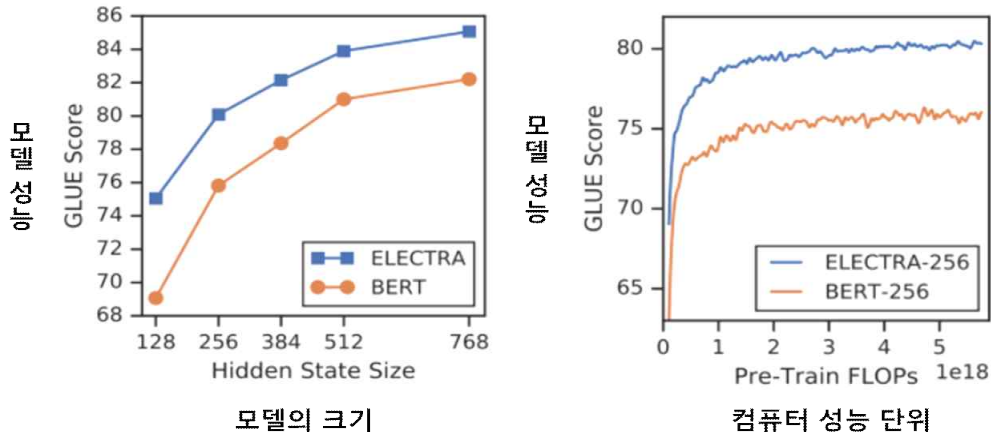
{
  "message": "성공",
  "data": {
    "question": "123",
    "answer": "구독자",
    "start": 999,
    "end": 999,
    "processed": 99.73
  }
}

```

[그림 2-101] 기계독해 API 실행 예시

③ 한국어 딥러닝 언어모델 학습 시스템 구축

< ELECTRA vs. BERT 성능 비교 >



[그림 2-102] ELECTRA와 BERT 모델의 성능 비교

- GOOGLE에서 개발한 자연어 처리 모델인 ELECTRA를 한국어로 학습시킨 KoELECTRA 모델을 사용해 기계독해 모델 개발
- KoELECTRA는 자연어 처리 모델에서 일반적으로 사용되는 BERT보다 학습속도가 빠르고 성능이 좋아 비교적 적은 비용으로 모델을 개발할 수 있음
- KoELECTRA는 한국어 자연어 오픈소스 모델 중 우수한 성능을 보이며, 약 34GB의 말뭉치로 학습시켜 풍부한 한국어를 처리할 수 있음

모델명	KorQuAD1.0	
	EM	F1
KoBERT (Base Size)	52.81	80.27
multilingual-BERT (Base Size)	70.42	90.25
KoELECTRA-Small-v3	82.11	91.13
HanBERT (Base Size)	78.74	92.02
KoELECTRA-Base-v3	84.83	93.45
enliple PLM (Large Size)	85.61	93.89
ETRI KorBert	86.4	94.18

[그림 2-103] 한국어에 대한 자연어 오픈소스 모델의 성능 비교

- 한국어 딥러닝 언어모델에 동물질병에 대한 역학, 방역 지식 학습 수행. 학습 데이터로 동물질병 QA 데이터셋을 사용
- 동물질병 QA데이터셋(ADQuAD)의 KorQUAD 1.0, AIHUB의 뉴스와 도서자료 QA데이터 셋을 기계독해 학습 모델로 사용
 - 4개의 데이터셋에서 각 500개는 TEST 데이터로 사용해 총 TEST 데이터 2000개를 모델의 성능을 평가하는 데이터로 활용, 그 외 데이터는 TRAIN 데이터로 모델 학습에 사용

데이터 셋		데이터 수			데이터 비율		
		train	test	합계	train	test	합계
내부 데이터	ADQuAD (10.12일 기준)	2,725	500	3,225	84%	16%	100%
외부 데이터	KorQuAD	57,061	500	57,561	99%	1%	100%
	Aihub 뉴스	147,086	500	147,586	100%	0%	100%
	Aihub 도서관	544,384	500	544,884	100%	0%	100%
	합계	748,531	1,500	750,031	100%	0%	100%
합계		751,256	2,000	753,256	99.7%	0.3%	100.0%

[그림 2-104] 기계독해 모델의 학습 및 평가에 사용된 데이터

- 모델별 학습, 테스트, 평가 과정을 통해 적정 학습파라미터 선정

학습파라미터	값	학습파라미터	값
Model Size	Base	Model Size	Small
Batch Size	64	Batch Size	64
Epochs	20	Epochs	40
learning rate	5E-05	learning rate	5E-05

[그림 2-105] BASE 모델과 SMALL 모델 학습 파라메타

- 기계독해 학습 서버는 RTX6000 그래픽카드 3개로 병렬로 학습하여, 학습 시간 단축

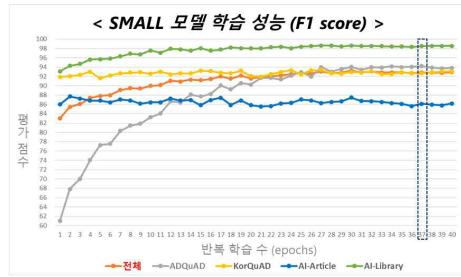
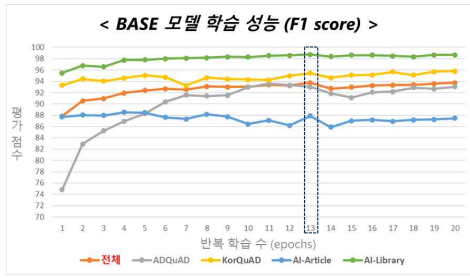
```

NVIDIA-SMI 440.64 Driver Version: 440.64 CUDA Version: 10.2
-----
GPU Name Persistence-M Bus-Id Disp.A | Volatile Uncorr. ECC
Fan Temp Perf Pwr:Usage/Cap | Memory-Usage | GPU-Util Compute M.
-----+-----+-----
0 Quadro RTX 6000 On 00000000:00:06.0 Off | 20504MiB / 22698MiB | 100% Default 0
N/A 54C P0 88W / 250W
1 Quadro RTX 6000 On 00000000:00:07.0 Off | 19148MiB / 22698MiB | 98% Default 0
N/A 53C P0 82W / 250W
2 Quadro RTX 6000 On 00000000:00:08.0 Off | 17444MiB / 22698MiB | 63% Default 0
N/A 52C P0 243W / 250W
-----
Processes:
GPU PID Type Process name GPU Memory Usage
-----+-----+-----
0 7596 C /home/ubuntu/anaconda3/bin/python 20493MiB
1 7596 C /home/ubuntu/anaconda3/bin/python 19137MiB
2 7596 C /home/ubuntu/anaconda3/bin/python 17433MiB

```

[그림 2-106] 기계독해 학습 GPU 서버의 사양 및 병렬처리 예

- 기본 BASE모델과 모델의 크기가 작지만 학습속도와 추론 속도가 빠른 SMALL모델도 학습. BASE와 SMALL모델의 성능을 비교한 결과는 아래와 같음
 - 기계독해의 정답 추론 성능은 BASE모델이 우수하지만, BASE모델을 서비스 운영 모델로 사용시, 고사양의 GPU서버가 필요. 반면, SMALL모델은 GPU서버 없이 서비스운영이 가능한 장점이 있어 SMALL모델로 기계독해 API를 제작



평가 기준		EM	F1	Answer Contain
전체		82.4	93.76	96.25
내부 데이터	ADQuAD	85	93	97
	KorQuAD	85.5	95.44	97.6
외부 데이터	Aihub 기사	61	87.87	91.9
	Aihub 도서	95.9	98.73	98.6

평가 기준		EM	F1	Answer Contain
전체		80.8	92.87	95.75
내부 데이터	ADQuAD	86	94.21	97
	KorQuAD	83.8	92.67	95.2
외부 데이터	Aihub 기사	56	86.11	92.8
	Aihub 도서	95.4	98.5	98

[그림 2-107] BASE 모델과 SMALL 모델의 성능 비교

< BASE >		Epochs																				
Score		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
ALL (2,000)																						
EM		69.7	73.9	76.10	77.95	79.25	80	80.5	80.75	81	80.95	81.45	81.85	82.4	82.85	83.4	82	82.10	82.10	82.95	82.80	
F1		87.81	90.34	90.95	91.95	92.40	92.85	93.56	93.95	94.29	94.30	94.37	94.27	94.76	94.70	94.76	94.55	94.52	94.57	94.93	94.93	
Answer Contain		93.9	94.65	95.65	96.65	96.1	96	95.9	96	96.45	96.3	96.4	96.2	96.4	96.25	96.4	96.15	96.25	96.45	96.5	96.55	
ADQuAD (500)																						
EM		45.5	56.6	53.4	59.2	73.6	78.2	82	81.3	83.4	85.6	86.3	86.4	86	84	84.6	85.8	86.4	87	87.5	87.4	
F1		74.40	82.93	85.23	86.95	93.30	93.39	91.55	91.51	92.36	93.59	93.31	93.00	93.07	91.10	92.07	92.10	92.57	92.59	93.02	93.02	
Answer Contain		87.5	91	93.9	94.4	95.9	96	96.4	96.6	97.6	97.5	97.9	97.6	97	96	97	96.8	97	97.4	97.4	97.4	
KorQuAD (500)																						
EM		85	87	86.2	86.4	86.8	86.8	86.4	86.4	86.4	86.5	86	86	86.9	86.9	86.2	86.6	87.6	86.3	87.4	87.6	
F1		93.23	94.45	94.02	94.54	95.05	94.70	95.24	94.64	94.59	94.31	94.22	94.92	95.44	94.64	95.10	95.14	95.57	95.09	95.72	95.79	
Answer Contain		95.6	96.6	97.6	97.2	97	96.6	96.8	96.8	97	97.2	97	96.2	97.4	97.6	97	97.4	97	97.2	97	97.4	97.4
Article (500)																						
EM		54.5	59.9	62.8	62.6	61.8	60.6	60.4	60.4	60	57.4	57.5	58.0	61	58.2	59.6	60.8	59	56	60.2	60.2	
F1		87.69	88.92	87.94	88.55	88.43	87.85	87.55	88.74	87.75	86.44	87.11	86.18	87.87	88.84	87.80	88.19	86.92	87.16	87.27	87.45	
Answer Contain		94.2	93.2	95	95	95.8	95.2	95.4	95	95	95.4	95	92.4	97.8	97.8	92.4	92.2	92.6	92.8	92.8	92.8	
Library (500)																						
EM		63.5	61.6	61	59.6	64.2	64.6	64.6	65	65	64.5	64.4	64.4	65.8	64.4	64.2	65.6	65.70	65.3	66	66	
F1		95.45	94.75	94.51	97.74	97.80	98.10	98.18	98.18	98.33	98.25	98.55	98.60	98.74	98.39	98.64	98.62	98.44	98.35	98.67	98.67	
Answer Contain		95.5	97.0	97.2	97.6	97.8	98.2	98	98.4	98	98.2	98.9	98.2	98.6	98.6	98.6	98.6	98.5	98.5	98.5	98.5	
< SMALL >		Epochs																				
Score		1	2	3	7	8	11	16	17	19	24	26	27	28	29	30	36	37	38	39	40	
ALL (2,000)																						
EM		63.6	66.7	67.7	72.25	72.56	72.6	77	79.1	78.35	80.42	80	81.1	80.55	80.7	81.15	80.85	80.6	80.7	80.6	80.56	
F1		88.99	89.47	89.95	90.65	89.95	90.17	91.46	92.92	90.3	92.46	92.68	92.73	92.63	93.47	92.71	92.87	92.94	92.79	92.96	92.96	
Answer Contain		90.15	91.1	91.1	93.1	94.2	94.78	94.5	95.58	95.65	96	95	95.85	95.7	95.6	95.8	95.75	95.68	95.5	95.5	95.2	
ADQuAD (500)																						
EM		70.4	77.6	41	56.4	55.4	65.2	71.4	75.4	77.6	83.6	83.2	87.6	85.5	85	86.4	87.5	88	87.5	87.4	87.6	
F1		86.05	87.54	70.36	80.31	81.45	84.05	89.24	89.07	90.55	92.16	91.99	94.02	93.34	93.59	94.31	94.97	94.21	93.92	93.75	93.92	
Answer Contain		77.6	80	87	89.36	92.8	93.4	94.8	95.9	95.6	97.4	95.9	97.4	95.3	97.2	97.4	97.6	97	97	97.4	97.2	
KorQuAD (500)																						
EM		82.2	83.2	83.4	83.4	84.2	85	84.2	87	84.1	86	84.2	84	84	84	84.4	83.5	83.2	83.9	83.6	84	
F1		91.90	92.38	92.34	92.65	92.89	93.05	93.14	92.93	93.24	93.36	93.39	93.70	93.70	93.85	93.97	93.87	93.80	93.69	93.66	93.66	
Answer Contain		94.4	94.4	95.6	95.2	96	95.4	95.4	95.5	95	95.5	95.4	95.2	95.5	95.2	94	95	95.2	95	95	95	
Article (500)																						
EM		58.4	58.6	59.2	57.4	58.9	57.2	60.6	59.8	57.2	59.4	57.2	57.4	56	57.4	55	55.9	56	56	55.9	56.4	
F1		85.09	87.71	87.25	87.04	86.87	86.49	89.32	87.4	86.64	86.58	86.84	88.80	88.54	86.66	87.47	85.65	86.11	85.90	85.81	85.82	
Answer Contain		62.2	66	65.2	64.2	65.2	66.6	66	66.4	67.8	67.8	68.6	68.6	68.6	68.6	68.6	68.6	68.6	68.6	68.6	68.6	
Library (500)																						
EM		85.4	87.2	88.2	91.6	90	88	88.4	94.2	94.2	94.8	95.4	95.4	95.4	95.4	95.4	95.2	95.4	95.4	95.4	95.4	
F1		98.09	98.27	94.59	94.39	94.39	97.11	97.59	97.79	98.07	98.06	98.49	98.03	98.30	98.45	98.33	98.37	98.50	98.56	98.59	98.56	
Answer Contain		96.1	97	95.9	97.6	97	97.2	99.4	99.4	98.1	98	98	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	

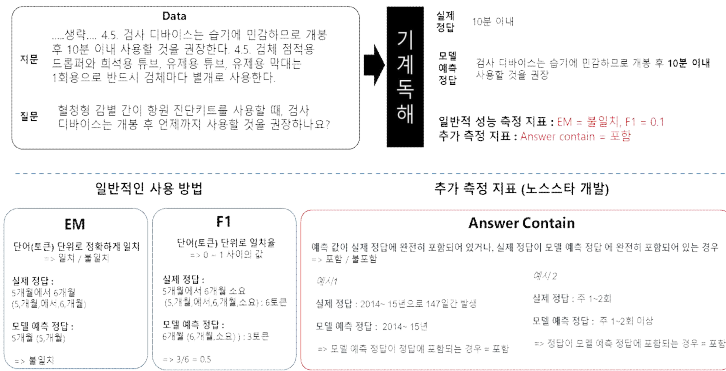
< KorQuAD 데이터셋 모델과의 성능 비교 >

모델명	KorQuAD 1.0	
	EM	F1
KoBERT (Base Size)	52.81	80.21
multilingual-BERT (Base Size)	70.42	90.21
KoELECTRA-Small-v3	82.11	91.11
HamBERT (Base Size)	78.74	92.01
KoELECTRA-Base-v3	84.83	93.41
emilep FLN (Large Size)	85.61	93.81
ETRI KoBERT	86.4	94.11

KorQuAD 1.0 온라인 Leaderboard 등록모델

[그림 2-108] BASE 모델과 SMALL 모델의 EPOCH별 성능 비교

- 각 EPOCH별 세부 성능은 아래와 같음. 동등한 비교를 다른 모델들과 할 수 없으나, KORQUAD1.0리더보드에 기록된 성능을 본 연구과제에서 개발한 모델과 비교했을 때, 비슷한 수준을 보임
- 기계독해 성능 테스트로 일반적으로 사용되는 지표인 EM, F1외에 추가 평가 지표 개발, EM은 기계독해 모델이 예측한 정답이 실제 정답과 정확하게 일치하는지 판단하는 것이고, F1은 예측정답과 실제정답이 일치하는 정도를 비율로 판단하는 것이다. 하지만, 기계독해 운영시에는 정확하게 정답을 찾는 것 뿐만 아니라 정답이 포함된 문장을 찾는 것이 사용자에게 도움을 줄 수 있다. 따라서 정답이 포함된 문장을 얼마나 잘 예측하는지에 대해 평가하는 ANSWER CONTAIN의 평가 지표를 개발하고, 모델 성능 평가에 사용. 정확한 실시 예는 아래와 같음



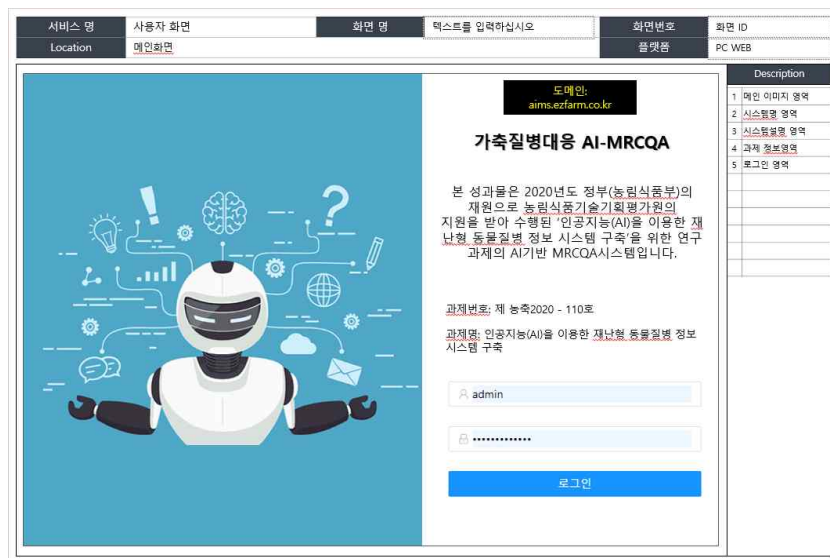
[그림 2-109] 기존 기계독해의 평가 지표와 개발한 지표의 예

(3) 인공지능 기반 재난형 동물질병 정보 검색 시스템 개발

(가) 사용자에게 편의성을 제공하는 UX/UI 기술 개발

1차년도에 수행한 데이터 분석과 사용자 요구사항 수렴을 기반으로 최종적인 필요 기능을 정의한 후, 이를 반영하여 최적의 UX/UI 기술을 도입한 화면정의서를 도출하고, 디자인 및 퍼블리싱을 진행함

① 로그인 및 전체 컨셉 기획



[그림 2-110] 로그인 화면 기획

○ 로그인 화면

- 가장 사용자에게 친근하게 다가갈 수 있는 보편적 UI를 채택
- 대국민이 아닌 폐쇄적 특성을 고려하여, 메인화면을 생략하고 로그인 화면으로 대체
- 사용자 범위에 맞게 회원 가입을 생략 (DB입력식)
- 로그인 화면 좌측에 시스템을 대표할 수 있는 이미지 영역을 할당
- 로그인 화면 우측에 연구 과제의 상세 정보를 나열하여 시스템의 정체성을 표명

○ 시스템 컨셉 (영역 정의)

- 기획->화면정의->디자인->퍼블리싱->개발

각 단계로 진행되는 과정에서 업무적 혼선을 방지하기 위해 사전에 각 기능과 영역을 정의함

- ▶ 질의 응답 채팅 영역: 질의와 응답에 대한 히스토리를 대화형으로 구현
- ▶ 질의 영역: 조건 설정과 입력 창으로 나누어 대화형으로 구현
- ▶ 질의응답 내용 요약 영역: 다양한 질의와 응답이 오고갔을 경우를 고려하여, 해당 응답이 어떠한 질의로 인해 호출됐는지 히스토리 영역을 구현
- ▶ 답변 문서 영역: 사용자 질의에 대한 응답의 근거가 되는 관련 문서 호출 영역
관련 문서에 대한 기본 정보(문서명/발행일 등) 영역과 다운로드 기능 및 원문 영역을 구분하여 직관적 UI를 구현
- ▶ 결과 내 검색 영역: 관련 문서 내 '키워드'검색을 위한 검색 기능 구현



[그림 2-111] 영역 정의

- ▶ 답변 목록 영역: 다양한 관련 문서를 카테고리 별로 구분하여, MRC SCORE 순으로 나열함으로써 사용자에게 정보를 제공

② 질의 영역

○ 질의 설정

- 대화형 UI 컨셉에 충실하게 구현
- 다양한 문서(PDF, 기사, 고지자료 등)에서의 검색을 필요로 하는 시스템의 특성을 고려하여 답변을 유형을 사용자가 설정할 수 있도록 구현하고, 시간의 흐름에 따라 결과값이 달라지는 검색의 특성에 맞게 답변 문서의 기간 설정 기능을 추가 구현하여, 검색의 폭을 좁혀 조금 더 정확한 답을 찾을 수 있도록 편의성을 제공
- 단, 사용자 편의를 위해 선택은 자유로 두며, 조건 미 선택 시 전체를 대상으로 검색)



[그림 2-112] 질의 세부 설정 기획

○ 질의 응답

- 사용자 질의에 대한 답을 응답 형태로 제공
- 해당 응답에 대한 근거가 되는 관련 문서 상세 보기 기능 제공

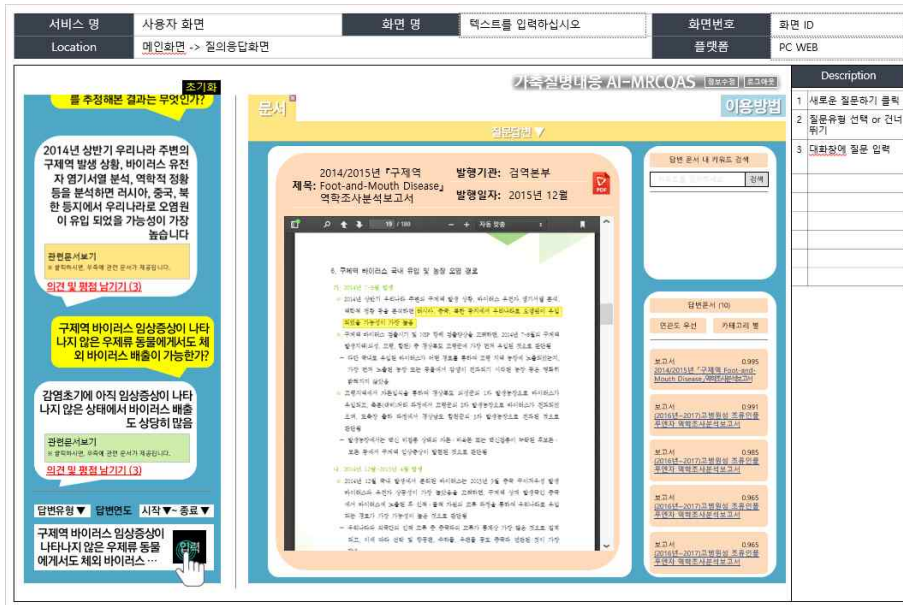


[그림 2-113] 사용자 질의에 대한 응답 콘텐츠 기획

- 관련문서 보기의 경우 응답에 대한 근거 외에도 MRC 점수를 기준으로 상위 문서를 제공함으로써 사용자가 다각도로 질문에 대한 문서를 분석할 수 있도록 구현

○ 추가 질의

- 추가 질의시 대화형 UI 컨셉에 맞게 답변과 관련 문서를 묶음하여 제공



[그림 2-114] 추가 사용자 질의 UI 기획

③ 응답에 대한 관련 문서 영역

○ 관련 문서 보기

- 질의 영역에서의 응답이 도출된 근거 문서에 대한 상세 정보를 제공
- 또한, 해당 응답에 대한 문서 외에도 유사도가 높은 문서들에 대한 정보를 추가 제공함으로써, 사용자가 다각도로 질의에 대한 문서를 분석할 수 있도록 구현



[그림 2-115] 응답에 대한 관련 문서 보기 기획(1)

- ▶ 답변 문서 영역: 사용자 질의에 대한 응답의 근거가 되는 관련 문서정보를 호출 관련 문서에 대한 기본 정보(문서명/발행일 등) 영역과 문서 원문 다운로드 기능 및 원문 영역을 구분하여 직관적 UI를 구현
- ▶ 결과 내 검색 영역: 관련 문서 내 '키워드'검색을 위한 검색 기능 구현
- ▶ 답변 목록 영역: 다양한 관련 문서를 카테고리 별로 구분하여, MRC SCORE 순으로 나열함으로써 사용자에게 정보를 제공

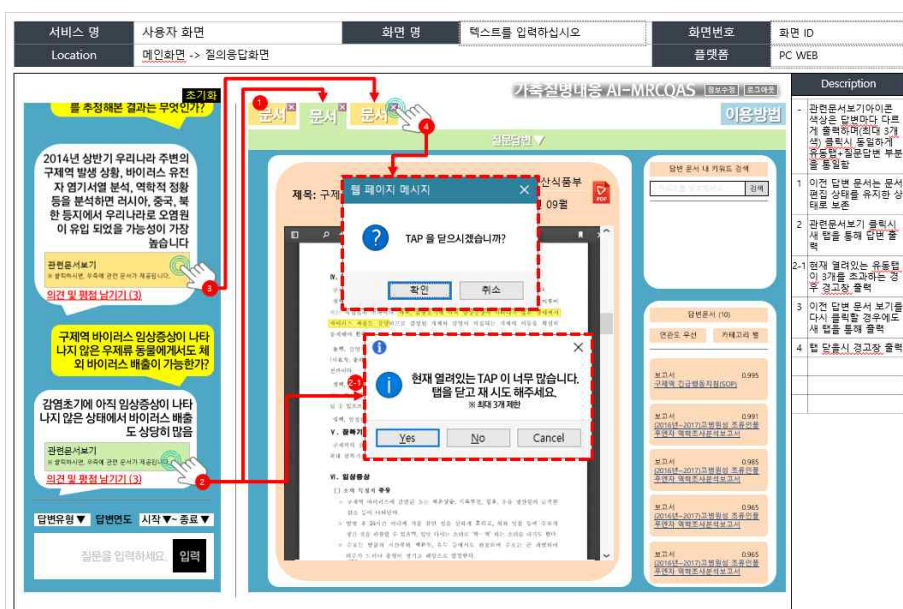


[그림 2-116] 응답에 대한 관련 문서 보기 기획(2)

- ▶ 질의응답 내용 요약 영역: 다양한 질의와 응답이 오고갔을 경우를 고려하여, 해당 응답이 어떠한 질의로 인해 호출됐는지 히스토리 영역을 구현
- ▶ 원문 보기 기능: 문서가 아닌 보도자료가 응답 근거인 경우 관련 링크를 새 창으로 띄움

○ TAP기능

- 사용자의 단일 질의로 원하는 모든 정보를 제공받을 수 없는 경우가 다분함으로, 각 질의마다 관련 문서 보기 기능을 TAP 형태로 구현하여, 이 전 질문에 대한 응답 히스토리 대조는 물론 각 응답에 대한 관련 문서를 비교·분석할 수 있도록 사용자 편의를 제공
- 단, 시스템의 물리적 수용치를 고려하여 TAP은 최대 3개로 제한함

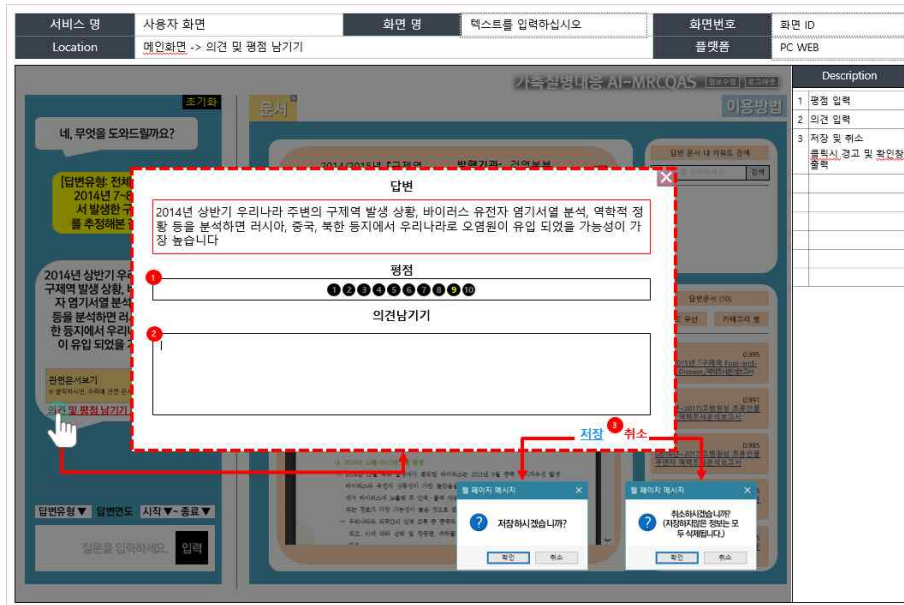


[그림 2-117] 응답 별 관련 문서 구분 TAP 기획

④ 기타

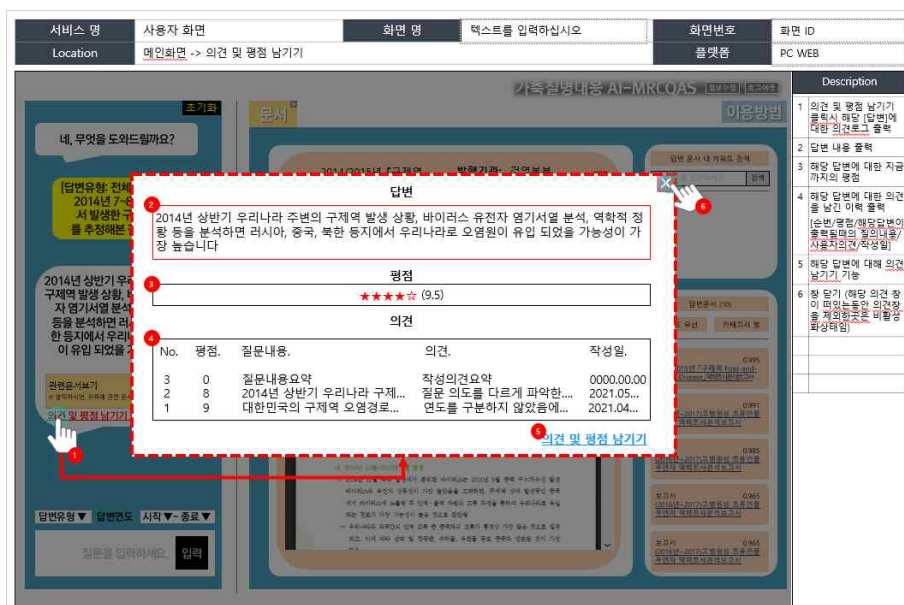
○ 평점 남기기

- 장기적인 관점에서의 질의 응답 신뢰도를 제고하기 위해, 각 응답에 대한 평가 기능을 구현
- 수집된 피드백을 수립하여, 개선 방안을 도출할 수 있는 기반을 마련



[그림 2-118] 응답에 대한 평점 남기기 기획

- ▶ 응답 내용 요약: 사용자가 평가하고자 하는 대상 응답 요약
- ▶ 평점: 10/10을 기준으로 해당 응답에 대한 점수 평가
- ▶ 의견남기기: 응답에 대한 피드백 남기기 기능



[그림 2-119] 응답에 대한 평점 보기 기획

- ▶ 응답 내용 요약: 사용자가 평가하고자 하는 대상 응답 요약
- ▶ 평점: 해당 응답에 점수를 준 사용자들의 평균 점수를 시각/수치화하여 출력
- ▶ 의견: 사용자가 해당 응답을 도출한 질의가 서로 다를 수 있으므로, 응답 신뢰도 제고를 위한 의견 수렴 및 정보 제공 차원에서 질의 내용을 별도 표기하고, 응답에 대한 점수와 의견 및 작성일을 표기

⑤ 디자인 및 퍼블리싱

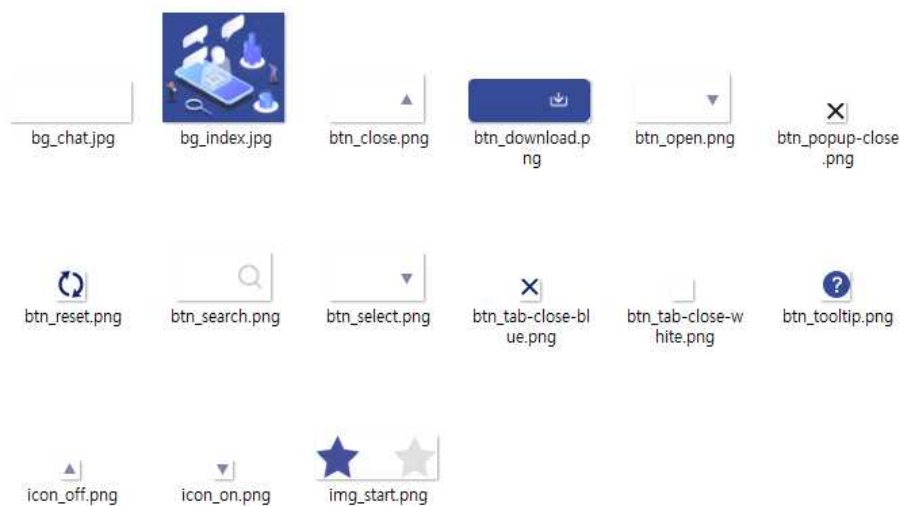
○ 디자인 컨셉

- 시스템의 AI 챗봇을 형상화하여 메인 이미지를 구성
- 정부 기관에서 가장 많이 활용되는 남색(#354993)을 메인 컬러로 선정함으로써 텍스트 시연성을 확보하고, 사용자로 하여금 친숙함을 유도



[그림 2-120] 로그인 화면 디자인

- ▶ 연구 과제 정보를 기존 로그인 창 화면 상단에서 대표 이미지 하단으로 변경

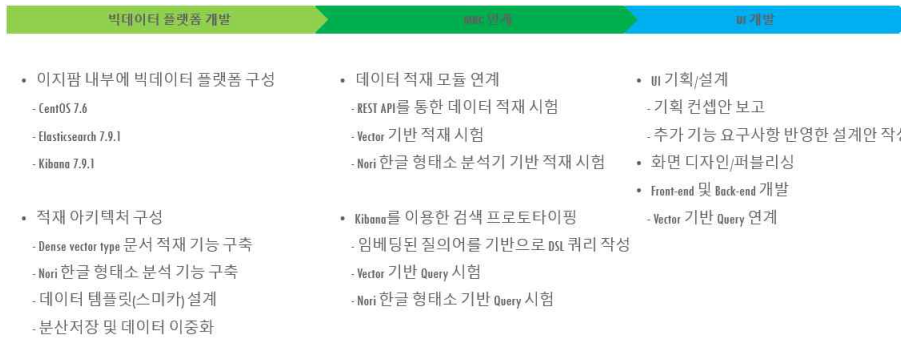


[그림 2-121] 응답에 대한 평점 남기기 기획

- ▶ 전체적으로 가볍지 않고, 최대한 직관적인 아이콘을 제작하여 활용

(나) 사용자의 요청에 적합한 정보를 제공하는 검색 기술 개발

1차년도부터 수행한 데이터 분석과 사용자 요구사항 수렴을 기반으로 정의된 화면정의서를 기준으로 디자인 및 퍼블리싱을 완료하고, 실제 개발 환경에 맞춰 미세 조정해 나감으로써 시행착오를 줄이고 검수 및 테스트 기간을 늘려 완성도 높은 시스템을 개발함



[그림 2-122] 단계별 진행 모식도

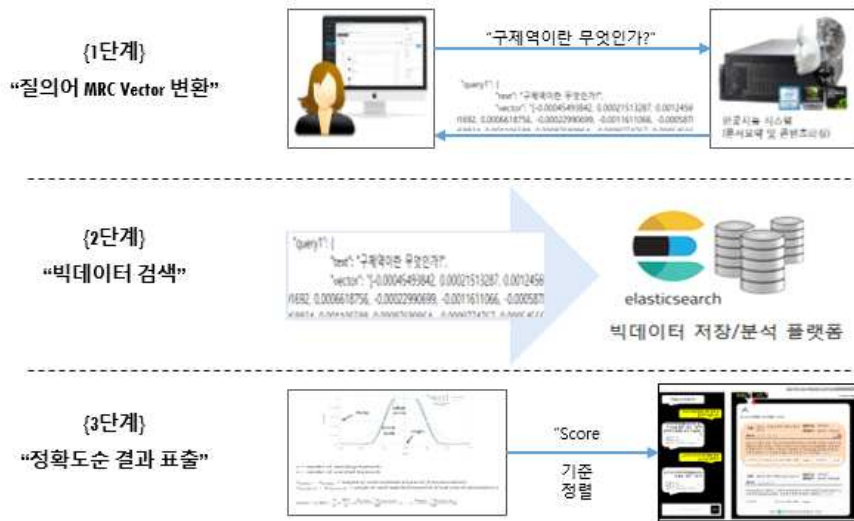
① 시스템 구조(구조도, 흐름도)

○ 서비스 흐름

- 크게 적재와 검색으로 나뉘어지며, UI/ES/MRC 실시간 연결을 기반으로 함



[그림 2-123] 적재 흐름도

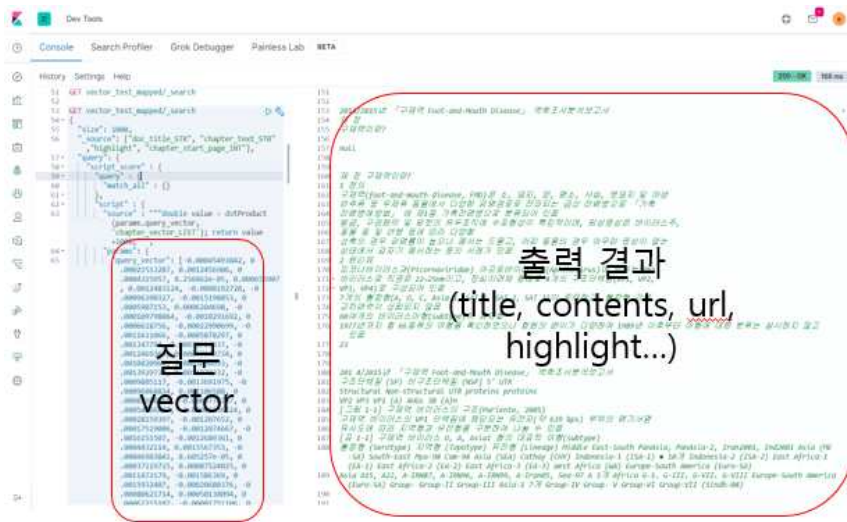


[그림 2-124] MRC지능형 검색

- ▶ 재난형 동물질병 정보 수집 시스템을 통해 추출 및 분류되고 MRC를 통해 VECTOR 변환된 데이터를 빅데이터 플랫폼인 ELASTICSEARCH에 주기적으로 적재하는 역할을 수행
- ▶ 적재된 데이터를 대상으로 사용자 질의(UI)에 대한 MRC지능형 검색 및 한글 플랫폼 검색을 수행

○ 검색 프로토타이핑

- KIBANA DEV TOOL 검색



[그림 2-125] 검색 프로토타이핑


- ▶ DSL QUERY문을 작성하고 질의 결과를 확인할 수 있는 개발자 인터페이스를 이용한 프로토타이핑 수행

② 시스템 DATABASE

○ 테이블 정의서

- 시스템 구동을 위한 테이블(DB)정의서 작성

- ▶ 시스템 이관 및 활용을 고려하여 표준화된 행정표준용어정의서를 기준하여 작성함

	인공지능(AI)을 이용한 재난형 동물질병 정보 시스템 구축 - 테이블 정의서 -	문서번호	D02_01.테이블정의서
		개정번호	Ver.1.2
	I. 검토 및 승인	쪽 번호	12쪽 중 2쪽

검토 및 승인

과 제 명	인공지능(AI)을 이용한 재난형 동물질병 정보 시스템 구축		
작 업 명	개발관리		
단 개 명	설계		
문 서 명	테이블정의서		
문서번호	D02_01.테이블정의서	개정번호	Ver.1.1
		페이지수	10

관리구분	<input checked="" type="checkbox"/> 관리본	<input type="checkbox"/> 비관리본
------	---	-------------------------------

구분	소속	부서	직위	성명	서명	일자
승인	(주)이지팜	피그컬처사업본부	책임	노형수		
작성	(주)이지팜	피그컬처사업본부	선임	여해원		

[그림 2-126] 테이블 정의서 표지

	인공지능(AI)을 이용한 재난형 동물질병 정보 시스템 구축 - 테이블 정의서 -	문서번호	D02_01.테이블정의서
		개정번호	Ver.1.2
	II. 제·개정이력 및 참고사항	쪽 번호	12쪽 중 3쪽

제·개정이력

번호	작성일	버전	페이지	차이	제·개정내용	작성자
1	2020. 10. 28.	0.1	-	-	최초 작성	여해원
2	2021. 02. 20.	0.2	-	-	데이터 항목 정의	여해원
3	2021. 03. 08.	1.0	-	-	데이터 타입 수정 등 최종 개정 및 배포	여해원
4	2021. 11. 02.	1.1	-	-	데이터 항목 수정	여해원
4	2021. 11. 08.	1.2	-	-	데이터 키 수정	여해원


참고사항

○ 본 문서는 다음의 문서를 참고하여 작성함.

항목	문서명	버전	내 용
참조문	항정표준용어정의서-Ver.1.0	1.0	2010년 항정표준용어사전을 기준으로 작성된 문서

[그림 2-127] 테이블 정의서 제·개정이력 및 참고사항

- ▶ 시스템에서 MRC 접속을 위한 토큰 정보 테이블
- ▶ 시스템 과부하를 막기 위해 시스템 자체에서 일정 기간동안 토큰을 부여받아 저장하고, 이를 사용자에게 동일하게 부여하는 방식을 채택


	인공지능(AI)을 이용한 재난형 동물질병 정보 시스템 구축 - 테이블 정의서 -	문서번호	D02_01.테이블정의서
		개정번호	Ver.1.2
	III. 본문	쪽 번호	12쪽 중 6쪽

1.1 토큰 정보

테이블ID	REFRESH_TOKENS	테이블명	토큰 정보	업무영역				
정의	MRC 시스템 접속 토큰 정보							
Row길이		초기조건		발생단위				
인덱스 정의								
인덱스명		인덱스유형	UNIQUE	결합명				
순번	결합ID	결합명	PK	FK	데이터타입	N/N	DEFAULT	범위 및 Value Check조건
1	USER_ID	유저 아이디	○		VARCHAR(20)	N		
2	REFRESH_TOKEN	재발송 토큰			VARCHAR(1000)	N		

[그림 2-128] 토큰 정보 테이블

- ▶ 사용자 정보 테이블
- ▶ 시스템 활용 범위를 고려하여, 최소한의 정보만을 입력함
- ▶ 향후 사용자층 및 사용 범위 확대를 고려하여 개인 정보 수집 영역을 미리 확보해둠
(본 연구과제 종료 시점에서는 수집하지 않음)


	인공지능(AI)을 이용한 재난형 동물질병 정보 시스템 구축 - 테이블정의서 - Ⅲ. 본문		문서번호	D02_01 테이블정의서
			개정번호	Ver.1.2
			쪽 번호	12쪽 중 7쪽

1.2 회원 정보

테이블ID	USER	테이블명	회원 정보	업무영역				
정의	시스템 가입 회원 정보							
Row길이		초기값		발생연/년				
인덱스 정의								
인덱스명	인덱스유형	UNIQUE	컬럼명					
순번	컬럼ID	컬럼명	PK	FK	데이터타입	N/N	DEFAULT	범위 및 Value Check조건
1	USER_SEQ	회원 번호	○		INT(100)	N		
2	USER_ID	회원 ID			VARCHAR(300)	N		
3	PASSWORD	비밀번호			VARCHAR(1000)	N		
4	NAME	회원명			VARCHAR(300)	Y		
5	EMAIL	이메일			VARCHAR(100)	Y		
6	PHONE	전화번호			VARCHAR(11)	Y		
7	CRE_DT	생성일			TIMESTAMP	Y		
8	LUP_DT	수정일			TIMESTAMP	Y		
9								
10								
11								
12								

[그림 2-129] 회원 정보 테이블

- ▶ 사용자 질의에 대한 정보 테이블
- ▶ 시스템 개선(정확도 제고 등) 및 평점 기능을 위한 질의 정보 수집

	인공지능(AI)을 이용한 재난형 동물질병 정보 시스템 구축 - 테이블정의서 - Ⅲ. 본문		문서번호	D02_01 테이블정의서
			개정번호	Ver.1.2
			쪽 번호	12쪽 중 8쪽

1.3 질의 정보

테이블ID	QUESTION	테이블명	질의 정보	업무영역				
정의	시스템 질의 정보							
Row길이		초기값		발생연/년				
인덱스 정의								
인덱스명	인덱스유형	UNIQUE	컬럼명					
순번	컬럼ID	컬럼명	PK	FK	데이터타입	N/N	DEFAULT	범위 및 Value Check조건
1	QST_SEQ	질문 번호	○		INT(11)	N		
2	ANS_SEQ	답변 번호			INT(11)	Y		
3	USER_ID	사용자 ID			VARCHAR(300)	Y		
4	QST_CONTENT	질문 내용			VARCHAR(5000)	Y		
5	QST_CRINION	질문 의견			VARCHAR(5000)	Y		
6	QST_ANS_RATE	점수			DOUBLE	Y		
7	CRE_DT	생성일			TIMESTAMP	Y		
8	LUP_DT	수정일			TIMESTAMP	Y		
9								
10								
11								

[그림 2-130] 질의 정보 테이블

- ▶ 사용자 질의에 대한 응답(답변) 정보 테이블
- ▶ 시스템 개선(정확도 제고 등) 및 평점 기능을 위한 질의 정보 수집
- ▶ 관련 문서 연계를 위한 문서/문장 ID 정의

1.3 질의 정보

테이블ID	QUESTION	테이블명	질의 정보	업무영역				
정의	시스템 질의 정보							
Row길이		조기건수		발생간수년				
인덱스 정의								
인덱스명		인덱스유형	UNIQUE	컬럼명				
순번	컬럼ID	컬럼명	PK	FK	데이터타입	N/N	DEFAULT	범위 및 Value Check조건
1	QST_SEQ	질문 번호	○		INT(11)	N		
2	ANS_SEQ	답변 번호			INT(11)	Y		
3	USER_ID	사용자 ID			VARCHAR(300)	Y		
4	QST_CONTENT	질문 내용			VARCHAR(5000)	Y		
5	QST_ORINION	질문 의견			VARCHAR(5000)	Y		
6	QST_ANS_RATE	점수			DOUBLE	Y		
7	CRE_DT	생성일			TIMESTAMP	Y		
8	UPD_DT	수정일			TIMESTAMP	Y		
9								
10								
11								

[그림 2-131] 응답(답변) 정보 테이블

③ 시스템 전체 요약

○ 로그인 화면

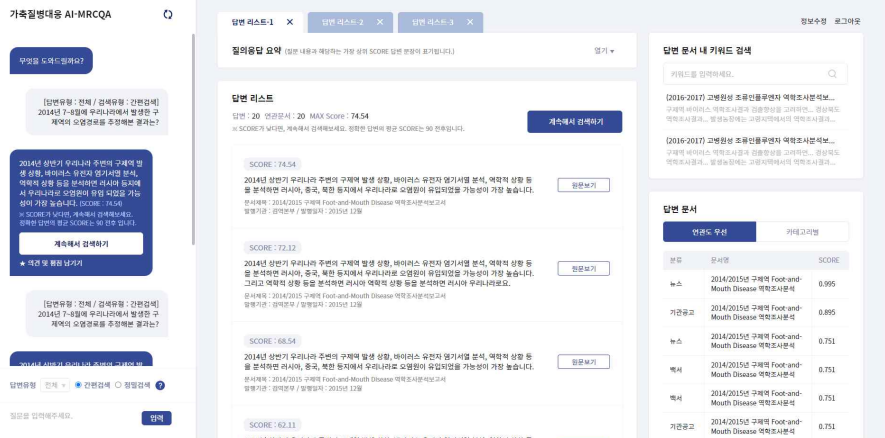
- 정부 기관에서 가장 많이 활용되는 남색(#354993)을 메인 컬러로 선정함으로써 텍스트 시연성을 확보하고, 사용자 하여금 친숙함을 유도



[그림 2-132] 로그인 화면 디자인

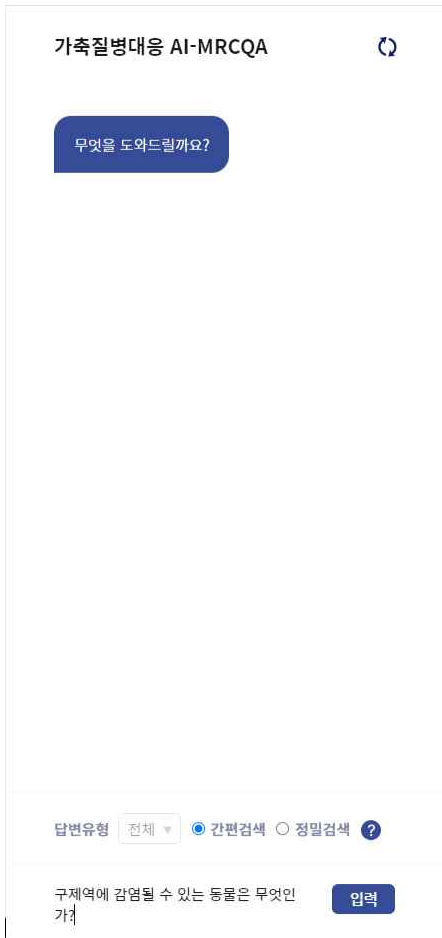
○ 메인 화면

- 질의/응답영역, 질의요약 영역, 관련문서 영역, 결과내 검색 영역, 문서별 카테고리 영역 총 5가지 영역으로 콘텐츠를 박스형태로 명확하게 구분하여 손쉬운 UI를 구현
- 각 영역의 세부 설정을 최대한 줄이고, 콘텐츠 내 명령어 역시 의미가 불분명한 아이콘을 지양하고 최대한 텍스트 및 수치 등 정보 전달 위주의 직관적인 UX를 구현
 - ▶ 질의 응답 채팅 영역: 질의와 응답에 대한 히스토리를 대화형으로 구현
 - ▶ 질의 영역: 조건 설정과 입력 창으로 나누어 대화형으로 구현



[그림 2-133] 메인 화면

- ▶ 질의 응답 내용 요약 영역: 다양한 질의와 응답이 오고갔을 경우를 고려하여, 해당 응답이 어떠한 질의로 인해 호출됐는지 히스토리 영역을 구현
- ▶ 관련 문서 영역: 사용자 질의에 대한 응답의 근거가 되는 관련 문서 호출 영역 관련 문서에 대한 기본 정보(문서명/발행일 등) 영역과 다운로드 기능 및 원문 영역을 구분하여 직관적 UI를 구현
- ▶ 결과 내 검색 영역: 관련 문서 내 '키워드'검색을 위한 검색 기능 구현
- ▶ 답변 문서 영역: 다양한 관련 문서를 카테고리 별로 구분하여, MRC SCORE 순으로 나열함으로써 사용자에게 정보를 제공



[그림 2-134] 질의 입력 화면



[그림 2-135] 질의 응답 화면

④ 세부기능 설명

○ 질의 응답 (채팅) 영역

- 질의 응답

- ▶ 사용자 질의와 시스템 응답(답변)을 대화 챗봇 형태로 구현함으로써, 직관적인 UI를 제공함과 동시에 질의 응답에 대한 히스토리를 확인하며 업무 효율을 높일 수 있도록 구현함
 - * 질의: 대화 형태로 질의
 - * 응답: 응답 문장/응답 문장의 SCORE/관련 문서 보기/의견 및 평점 남기기

- 질의 조건 설정

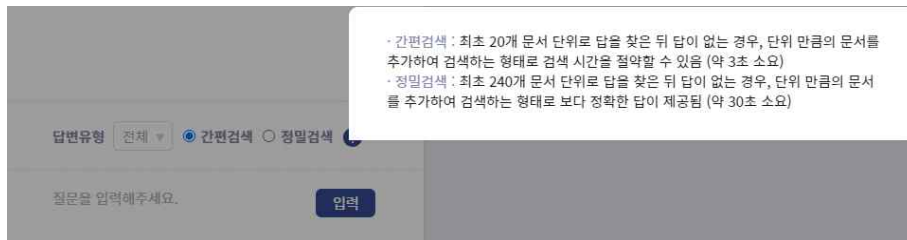
- ▶ 답변 문장을 찾을 문서의 유형을 선택할 수 있도록 제공함으로써, 정확도를 확보
- ▶ 검색 유형을 간편/정밀로 나눠서 사용자의 목적에 맞게 선택할 수 있도록 구현함
 - * 답변 유형: 답변을 찾을 문서의 유형을 선택 (예: 백서 선택시 백서에서만 답변 검색)
 - * 간편 검색: 20개 문서 단위로 단계적 검색함으로써 검색 시간을 절약
 - * 정밀 검색: 240개 문서 단위로 단계적 검색함으로써 보다 정확한 답변 확률을 확보



[그림 2-136] 답변 유형 선택



[그림 2-137] 정밀 검색 메시지

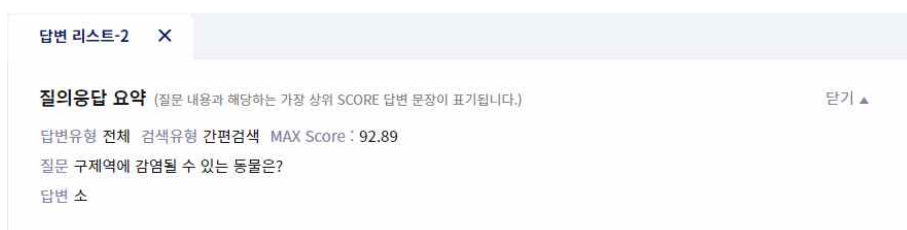


[그림 2-138] 간편 검색 및 정밀 검색 도움말

○ 관련 문서 영역

- 질의 응답 요약

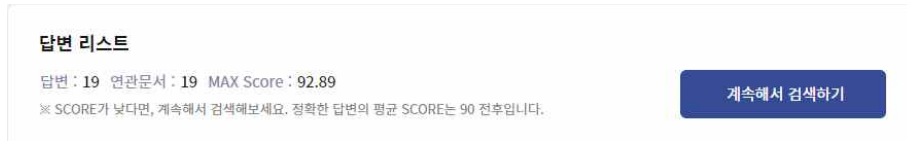
- ▶ 해당 응답(답변)의 사용자 질의 및 설정 등을 요약하여 정보를 제공
 - * 사용자 질의 정보 제공
 - * 사용자 질의 시 설정한 조건 정보 제공
 - * 사용자 질의에 대한 응답(답변) 제공
 - * 사용자 질의에 대한 응답(답변)의 SCORE 제공
 - * 질의 응답 요약 페이지는 열기/닫기 기능을 제공함



[그림 2-139] 질의 응답(답변) 요약 화면

- 계속해서 검색하기

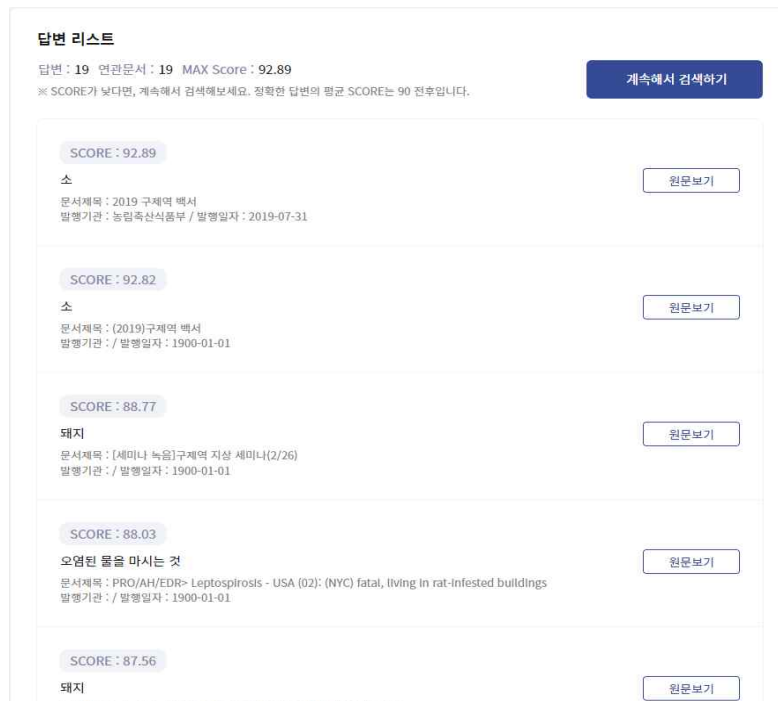
- ▶ 동일한 사용자 질의에 대한 추가적인 검색 기능
 - * 해당 사용자 질의를 더 많은 문서를 대상으로 추가적인 검색 수행
 - * MAX SCORE 갱신 시 기존 응답(답변) 갱신



[그림 2-140] 계속해서 검색하기 화면

- 답변 리스트

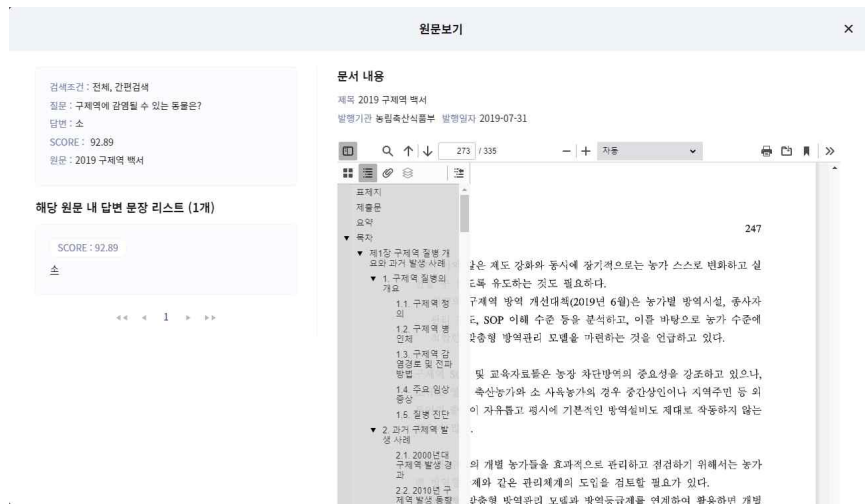
- ▶ 사용자 질의에 대한 응답(답변)의 근거가 되는 문서 리스트 제공
- ▶ 응답(답변)의 정확도를 확보하기 위해, 추가 검색 기능을 제공함으로써, 초기 사용자 질의시 설정한 검색 조건(간편/정밀검색)보다 더 많은 문서를 대상으로 추가 서치 기능을 제공
- ▶ 답변 리스트에 제공된 문장 리스트의 원문 보기 기능 구현
 - * 해당 사용자 질의에 대해 몇 개의 응답(답변)이 검색되었는지 제공
 - * 해당 사용자 질의에 대해 몇 개의 문서가 검색되었는지 제공
 - * 해당 사용자 질의에 대해 최대 SCORE 수치를 제공
 - * 동일한 사용자 질의에 대해 추가적인 검색 기능을 제공
 - * 각 응답(답변)의 SCORE, 응답(답변)문장, 근거 문서 제목, 근거 문서의 발행 기관 및 발행일자 등을 요약하여 제공
 - * 응답(답변)의 근거 문서 상세 보기 기능을 제공



[그림 2-141] 응답(답변) 리스트 화면

- 원문 보기

- ▶ 응답(답변) 문장의 근거가 되는 문서 원문 보기 기능
 - * 요약: 검색조건, 질의, 응답(답변), 응답(답변)의 SCORE, 원문명 정보 제공
 - * 해당 원문 내 답변 문장 리스트: 해당 원문 내에 사용자 질의에 대한 응답(답변)이 추가적으로 있을 경우 리스트로 보여줌으로써, 하나의 원문에서 추출된 응답(답변)문장을 손 쉽게 확인할 수 있음
 - * 문서 내용: 문서명, 발행기관, 발행일자 등에 대한 정보를 제공하고, 원문 내 해당 응답(답변) 문장의 위치로 이동하며 필요시 원문 문서를 다운로드 할 수 있도록 다운로드 기능을 제공함



[그림 2-142] 원문 보기 화면

○ 결과 내 검색 영역

- 사용자 질의에 대한 응답(답변) 문장이 속한 문서들 내에서의 키워드 검색 기능을 제공
 - ▶ 사용자 질의에 대한 응답(답변) 문서 내에 키워드 검색 기능 제공
 - ▶ 해당 키워드가 포함된 문서 제목 표기
 - ▶ 해당 키워드가 포함된 문장 표기
 - ▶ 해당 키워드는 볼트처리하여 한 눈에 알아 볼 수 있도록 구현



[그림 2-143]

답변 문서 내 키워드 검색 화면



[그림 2-144]

답변 문서 내 키워드 검색 결과 화면

○ 답변 문서 영역

- 사용자 질의에 대한 관련 문서 전체에 대한 리스트 제공 기능
 - ▶ 사용자 질의에 대한 관련 문서 리스트 전체 제공
 - ▶ 해당 문서 리스트는 연관도(SCORE)/카테고리별로 구분하여 정보 제공
 - ▶ 관련 문서 선택 시 원문 보기 메뉴를 팝업으로 출력함

분류	문서명	MAXS core
뉴스	PRO/AH/EDR> Chronic wasting disease - North America (05): USA (OH) deer	95.63
기관 광고	주요 가축질병 및 식물검역 현장문제 해결형 우수 연구성과 도출	95.28
백서	(2019년)구제역 역학조사분석보고서. 2019.12.	95.01
백서	(2019년)구제역 역학조사분석보고서. 2019.12.	95.01
기관 광고	광견병 등 악성전염병 첨단진단법 개발	94.99
백서	2019년 구제역 역학조사 분석보고서	94.89
뉴스	PRO/AH/EDR> Rabies (34): Americas (USA) various animals, human exp	94.52
뉴스	[참고] ASF에 대한 임상 예방 및 제어 기술지집	91.95
기관 광고	인천항 국제우편물을 통한 수입금지 축산물 적발사례 증가	91.47
기관 광고	아프리카돼지열병 조기 검출 가능한 항체 진단기술 개발 성공	91.36

[그림 2-145]

답변 문서 리스트 화면(SCORE정렬)

분류	문서명	MAXS core
백서	(2019년)구제역 역학조사분석보고서. 2019.12.	95.01
백서	(2019년)구제역 역학조사분석보고서. 2019.12.	95.01
백서	2019년 구제역 역학조사 분석보고서	94.89

[그림 2-146]

답변 문서 카테고리별 리스트 화면

(4) 인공지능 기반 재난형 동물질병 정보 검색 시스템 검증

(가) 인공지능 기반 재난형 동물질병 정보검색과 일반 웹 검색의 비교

○ 인공지능 기반 재난형 동물질병 정보검색

- 검색 대상: 방역업무 전문가에 의해 검증되고, 방역업무에 적합하도록 가공·정제된 신뢰성 높은 콘텐츠
- 검색 알고리즘: 기계독해(MRC)는 인공지능 알고리즘이 스스로 분석하고 질문에 최적화된 답안을 찾아내는 기술이며, 사람이 텍스트를 읽고 질문에 답변을 추론하듯이 인공지능이 문장 속에서 의미를 찾고 답변을 제시함으로 유의미한 콘텐츠를 제시함

○ 일반 웹 검색

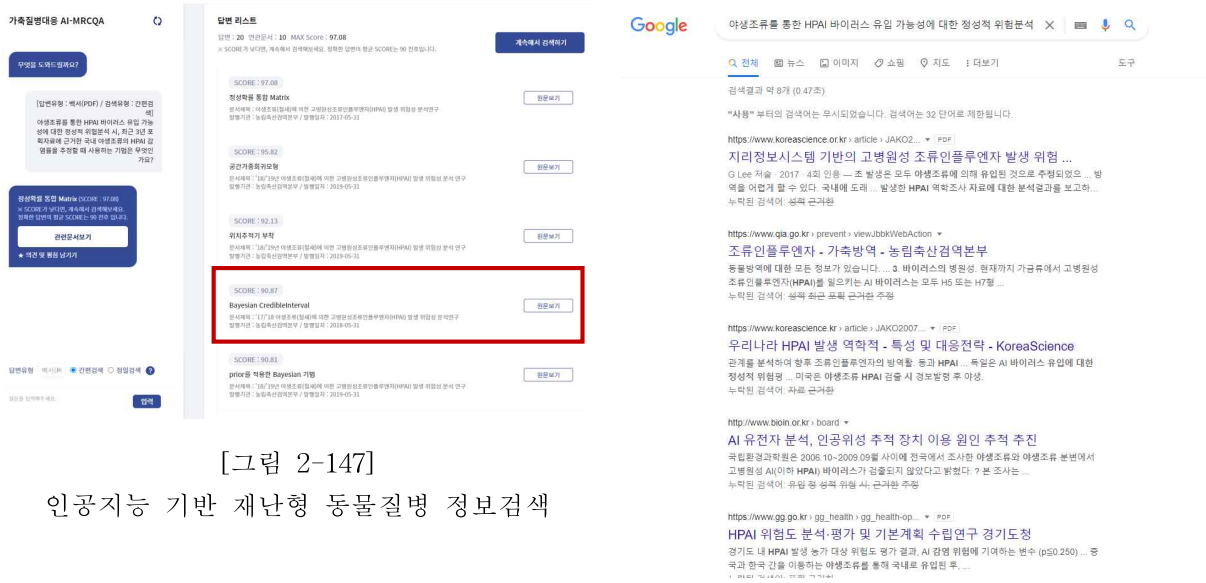
- 검색 대상: 인터넷에 공개된 비 검증된 콘텐츠

- 검색 알고리즘

- ▶ 키워드 또는 문장 포함 여부에 따른 문서 목록만을 제시하여, 사용자가 해당 문서를 직접 읽어 정보를 찾아 제공
- ▶ 최근에는 인공지능을 이용한 검색 서비스를 제공하고 있지만, 검증되지 않은 콘텐츠를 대상으로 검색하여 제공함으로 유의미한 콘텐츠를 찾는 데 많은 시간이 소요되거나 찾을 수 없음

○ 검색 사례

- 질문: 야생조류를 통한 HPAI 바이러스 유입 가능성에 대한 정성적 위험분석 시, 최근 3년 포획자료에 근거한 국내 야생조류의 HPAI 감염률을 추정할 때 사용하는 기법은 무엇인가요?
- 정답: Bayesian 기법
- 인공지능 기반 재난형 동물질병 정보 시스템은 사용자가 원하는 답을 4번째 콘텐츠(Socre 90.87)로 제공하나, 구글은 관련없는 정보만을 제공 함.



[그림 2-147]

인공지능 기반 재난형 동물질병 정보검색

[그림 2-148] 구글 정보검색

(나) 인공지능 기반 재난형 동물질병 정보 검색 시스템 검증

○ 콘텐츠의 신뢰성 확보

- 동물질병 방역·역학 전문가에 의해 검증된 사이트에서 콘텐츠를 수집하고 개발된 동물질병 사전을 활용하여 무의미한 콘텐츠 선별
- 동물질병 방역·역학 전문가의 검토·분석을 통해 유의미한 콘텐츠로 판별된 공공기관의 역학조사서·백서·공문·발간물 수집

○ 사용자 질의에 대한 콘텐츠 검색·제공

- 수집된 콘텐츠를 대상으로 사용자 질의에 대한 답변을 포함한 양질의 콘텐츠 검색·제공
- 검색 결과 품질 향상을 위한 동물질병 말뭉치 연구 및 시범 사용을 통한 개선점 도출 및 고도화 필요

○ 동물질병 방역·역학 업무 개선

- 동물질병에 대한 분석을 위하여 다양한 소스로부터 정보를 검색·분석·정제하는 과정을 수행하지 않음으로 업무 효율성 증대
- 인터넷을 활용하여 구축된 시스템에 접근·활용함으로써 공간적 제약 제거

3. 연구개발과제의 수행 결과 및 목표 달성 정도

가. 연구수행 결과

(1) 정성적 연구개발성과

구분 (연도)	세부과제명	세부연구목표	연구개발 수행내용	연구결과
1차 년도 (2020)	인공지능(AI)을 이용한 재난형 동물질병 정보 시스템 구축	1) 국내·외 방역당국의 발표자료, 학술자료, 언론 동향 크롤링 서비스 개발	(1) 동물질병 용어사전 정의 및 DB화	- 동물질병 감염병 카테고리, 감염병 및 병원체 데이터 모델 개발
			(2) 크롤링 대상 사이트 선정 및 크롤링 서비스 개발, 콘텐츠 수집	- 크롤링 시스템 개발 및 동물질병 콘텐츠 수집
		2) 역학조사보고서·백서· 공문 및 발간물 등록 서비스 개발	(1) 역학조사보고서·백서· 공문 및 발간물에 대한 관리대상 선정	- 동물질병 감염병에 대한 관리대상
			(2) 문서등록 서비스 개발 및 문서등록	- 동물질병 감염병 문서등록 시스템 개발
		3) 수집된 콘텐츠에 대한 파싱 기술 개발	(1) 수집한 콘텐츠를 동물질병 용어사전 기반으로 전·후처리하여 가공이 용이 하도록 DB화	- 동물질병 관련 사이트에서 수집한 정보를 DB에 저장·관리
			(2) 수집·등록된 문서에서 텍스트 데이터를 추출하여 가공이 용이 하도록 DB화	- PDF 또는 이미지 문서에서 텍스트 추출 및 DB에 저장·관리
		4) 재난형 동물질병 콘텐츠에 대한 인공지능 요약 모델 개발	(1) Python TextRank 알고리즘을 활용한 콘텐츠 요약 모델 개발 및 결과검증	- 동물질병 콘텐츠에 대한 요약 모델 및 서비스 개발
		5) 콘텐츠 활용도를 높이기 위한 정보의 DB화 및 인덱싱 기술 개발	(1) 빠른 데이터 검색을 위한 역인덱스 구조로 Indexing	- 하나의 Elasticsearch 노드 구축
			(2) 정형,반정형,비정형 대응을 위한 Full Text Indexing 기술 적용	- Elasticsearch의 형태소 분석기능 구축
			(3) 유연한 한글 처리를 위한 자체 한글형태소 분석기 Nori 탑재	- Nori Plug-in 설치 및 분석기 적용 구축
			(4) Rest API를 통한 데이터 쿼리 기능 제공	- Elasticsearch의 low level client API for Python 적용 구축
			(5) 높은 수집 성능을 위한 Ingest Node 기술	- 분석변환 모듈과 ingest Pipeline간 직접 통신기능 구축
			(6) 다양한 수집기 Beats 기술 제공	- 문서변환 모듈과 ingest Pipeline간 직접 통신 기능 구축으로 변경

			구축함
		(7) 데이터 시각화 및 관리 기능 제공	- Kibana 설치 및 Elasticsearch와 연동 구축
	6) 콘텐츠 활용을 위한 서비스(API) 기술 개발	(1) Rest API를 통한 CRUD 처리기술 제공	- vector 데이터 처리를 위한 dense vector 기술 구축
		(2) 문헌자료 수집을 위한 문서변환 기술 제공	- 별도의 문서 변환기술 적용
	7) 정보 검색 시스템 프로토타입 개발	(1) 해외 질병 데이터 및 백서/역학조사보고서 연계	- Index mapping/settings 생성 및 데이터 적재 테스트 완료
		(2) Q&A DB 생성 및 관리	- Q&A 데이터 생성 및 연계 대기
		(3) 기계독해 연계	- vector 데이터 처리를 위한 dense vector 기술 구축
		(4) Kibana를 통한 검색 기능 및 시각화	- Kibana의 dev tool을 통한 데이터 검색 및 결과 확인 완료
		(5) 농림축산검역본부 시스템 연계	- 연계 협의 예정
		(6) 사용자에게 편의성을 제공하는 UI/UX	- UI/UX 기획
2차 년도 (2021)	인공지능(AI)을 이용한 재난형 동물질병 정보 시스템 구축	1) 주요 재난형 동물질병동향 전파기술 개발	(1) 문자메시지 연동 서비스를 이용한 질병동향 전송 서비스 구축 - Naver Cloud 문자메시지 서비스를 이용한 질병동향 전송 서비스 API 개발
		(2) SNS 연동 서비스를 이용한 질병동향 전송 서비스 구축	- Line SNS 서비스를 이용한 질병동향 전송 서비스 API 개발
		(3) 한국어 콘텐츠 요약 개방형 API 서비스 개발	- 한국어 콘텐츠 요약 시스템 및 서비스 구현
	2) 재난형 동물질병 콘텐츠에 대한 인공지능 요약 모델 개선	(1) 형태소 분석기와 띄어쓰기 교정 모델을 활용한 단절 텍스트 연결 알고리즘 개발	- 인공지능 요약 모델 결과 개선
		(2) 콘텐츠 문서에 대한 문서 및 챕터 단위의 요약 필요성 검토	- 본 연구과제의 목적에 부합하지 않는다고 판단
	3) 재난형 동물질병 콘텐츠에 대한 인공지능 기계독해 모델 개발	(1) 동물질병에 대한 역학·방역 지식 학습용 QA 데이터셋 구축	- 최종 10,218건의 QA셋으로 동물질병 기계독해 데이터셋 (ADQuAD) 생성하여 데이터베이스저작권 등록
		(2) 인공지능을 활용한 콘텐츠 독해모델 생성 기술을 위한 질의응답 시스템 구축	- 범용 MRC 서비스 API 시스템과 재난형 동물질병 질의응답용 MRC API를 분리하여 시스템 구축
		(3) 인공지능을 활용한 콘텐츠 독해모델 생성 기술을 위한 개방형 API 서비스 개발	- 재난형 동물질병 질의응답용 MRC 서비스 API 개발 및 인터페이스 사양서 작성

	(4) 한국어 딥러닝 언어모델 학습 시스템 구축	<ul style="list-style-type: none"> - KoELECTRA 모델을 적용하여 연구목적에 적합한 기계독해 모델 개발 및 한국어 딥러닝 언어모델 학습 시스템 구축 - 재난형 동물질병에 대한 기계독해 모델 생성
4) 기계독해 학습을 위한 학습 데이터 생성	(1) 학습 데이터 생성 및 전문가 검증	<ul style="list-style-type: none"> - 재난형 동물질병 콘텐츠에 대한 기계독해 학습 데이터 생성·검증
	(2) 학습 데이터 등록·수정·관리를 위한 시스템 개발	
	(3) 학습 데이터로 사용 가능한지를 점검하는 서비스 개발	
5) 사용자에게 편의성을 제공하는 UX/UI 기술 개발	(1) 고급 분석 기능 및 쿼리 결과에 대한 scoring 조정 기술 제공	<ul style="list-style-type: none"> - Elastic Search를 통해 고급 분석 기능 및 쿼리 결과 scoring 조정 기술을 제공하고 UI에 scoring 결과를 제공 완료
	(2) 전용 UX/UI 제공	<ul style="list-style-type: none"> - 전용 UI/UX를 구현하였으며 카카오톡 같은 대화형 UI와 질의 응답 및 문서를 볼 수 있는 UI/UX 구현 완료
6) 사용자의 요청에 적합한 정보를 제공하는 검색 기술 개발	(1) 사용자의 요청에 적합한 정보를 제공하는 검색 기술 개발	<ul style="list-style-type: none"> - Elastic Search 기반 데이터 정교화 및 다양한 서비스를 위한 마트 구성 및 Ngram, Edge-Ngram 등 토큰나이징 기술 적용 - Elastic Search 기반 Logstash ETL을 통한 유연한 마트 구성 기능 제공 및 높은 QPS 처리를 위한 Coordination node 기술 적용 완료 - 사용자의 요청에 적합한 정보를 제공하는 검색 기술을 적용하여 인공지능 기반 재난형 동물질병 정보 검색 시스템 구축 완료

(2) 정량적 연구개발성과

(단위 : 건수, 백만원, 명)

성과 목표	사업화지표										연구기반지표									
	지식 재산권				기술 실시 (이전)		사업화				기술 인증	학술성과			교 육 지 도	인 력 양 성	정 책 활 용· 홍 보		기 타 (타 연 구 활 용 등)	
	특 허 출 원	특 허 등 록	품 종 등 록	S M A R T 평 준 등 급	건 수	기 술 료	제 품 화	매 출 액	수 출 액	고 용 창 출		투 자 유 치	논 문				학 술 발 표	정 책 활 용		홍 보 전 시
													S C I	비 S C I						
단위	건	건	건	건	건	백만원	건	백만원	백만원	명	백만원	건	건	건	명	건	건			
가중치	20						30			30							20			
최종 목표	1						3			5							1			
1차 년도	목표						1			3										
	실적						1			8										
2차 년도	목표	1					2			2							1			
	실적	1					3			8			1					2		
달성률 (%)	100						133			320							100			

※제품화 : 프로그램등록, 데이터베이스등록

(3) 세부 정량적 연구개발성과

[과학적 성과]

국내 및 국제 학술회의 발표

번호	회의 명칭	발표자	발표 일시	장소	국명
1	2021 한국수리생물학회 동계 학술대회	윤하정	2021.12.17	제주시 소노벨제주리조트	대한민국

[기술적 성과]

지식재산권(특허, 실용신안, 의장, 디자인, 상표, 규격, 신품종, 프로그램)

번호	지식재산권 등 명칭 (건별 각각 기재)	국명	출원				등록			기여율	활용 여부
			출원인	출원일	출원 번호	등록 번호	등록인	등록일	등록 번호		
1	문단탐색 및 문단선택 기반의 동물질병 정보검색을 위한 기계독해 시스템 및 방법	대한민국	(주)노스스타컨설팅	2021.12.29	10-2021-0191124					100%	

□ 저작권(소프트웨어, 서적 등)

번호	저작권명	창작일	저작자명	등록일	등록 번호	저작권자명	기여율
1	재난형 동물질병 정보 수집 시스템	2020.11.16	(주)노스스타 컨설팅	2020.11.18	C-2020-042879	(주)노스스타 컨설팅	65%
						농림축산검역본부	35%
2	가축질병대응 에이아이-엠알씨큐아이	2021.11.30	(주)이지팜	2021.12.13	C-2021-054497	(주)이지팜	100%
3	동물질병정보 기계독해 API(에이피아이) 시스템 소프트웨어	2021.12.07	(주)노스스타 컨설팅	2021.12.16	C-2021-054497	(주)노스스타 컨설팅	100%
4	동물질병 질문 및 응답 데이터 세트	2021.12.20	(주)노스스타 컨설팅	2021.12.28	D-2021-000162	(주)노스스타 컨설팅	50%
						농림축산검역본부	50%

□ 고용 창출

순번	사업화명	사업화 업체	고용창출 인원(명)		합계
			2020년	2021년	
1	인공지능(AI)을 이용한 재난형 동물질병 정보 시스템 구축	(주)노스스타컨설팅	5	4	9
2		(주)이지팜	2	2	4
3		농림축산검역본부	1	2	3
합계			8	8	16

[사회적 성과]

□ 설계 기준/설명서(시방서)/지침/안내서에 반영

번호	구분	활용 구분 (신규/개선)	설계 기준/설명서/지침/안내서 명칭	반영일	반영 내용
1	표준기술활용 기술서	신규	기계독해 학습을 위한 동물질병 질의-응답 데이터 세트 구성 기법	2022.02	‘기계독해 학습을 위한 동물질병 질의-응답 데이터 세트 구성 기법’을 표준기술로 활용하기 위한 세부사항을 표현한 설명서

□ 홍보 실적

번호	홍보 유형	매체명	제목	홍보일
1	기고	농림축산검역본부 정보지 아담드리 37호	기계독해(Machine Reading Comprehension, MRC)엔진에 대하여	2021.01.14
2	기고	농림축산검역본부 정보지 아담드리 40호	4차산업혁명 시대의 기계독해(MRC) 시스템 활용 사례	2021.10.27

(4) 계획하지 않은 성과 및 관련 분야 기여사항

- 데이터베이스제작저작권리 등록: 동물질병 질문 및 응답 데이터 세트
- 학술대회 발표: 2021 한국수리생물학회 동계 학술대회
 - DEVELOPMENT OF ANIMAL DISEASE INFORMATION SYSTEM BASED ON ARTIFICIAL INTELLIGENCE
- 홍보 실적: 농림축산검역본부 정보지 아람드리 37호, 40호 기고
 - 기계독해(Machine Reading Comprehension, MRC)엔진에 대하여
 - 4차산업혁명 시대의 기계독해(MRC) 시스템 활용 사례

나. 목표 달성 수준

추진 목표	달성 내용	달성도 (%)
○국내·외 방역당국의 발표자료, 학술자료, 언론 동향 크롤링 서비스 개발	- 동물질병 용어사전 정의 및 데이터베이스(DB) 화 - Keyword 사전 기반 Site별 동물질병 기사 탐색 및 자동수집 크롤링 프로그램 개발 - 영문 수집 데이터의 한국어 번역 SW 개발 - GUI 관리 기반 동물질병 관련 웹사이트 크롤링 시스템 개발	100%
○역학조사보고서·백서·공문 및 발간물 등록 서비스 개발	- 역학조사보고서·백서·공문 및 발간물 수집 대상 선정 및 문서등록 서비스 개발 - 백서, 역학·방역 관련 보고서 텍스트 변환	100%
○수집된 콘텐츠에 대한 파싱 기술 개발	- 수집한 콘텐츠를 동물질병 용어사전 기반으로 전·후처리하여 가공이 용이 하도록 데이터베이스(DB) 화 - 수집·등록된 문서에서 텍스트 데이터를 추출하여 가공이 용이 하도록 데이터베이스(DB) 화	100%
○주요 재난형 동물질병동향 전파 기술 개발	- 문자메시지 연동 서비스를 이용한 질병동향 전송 서비스 구축 - SNS 연동 서비스를 이용한 질병동향 전송 서비스 구축	100%
○재난형 동물질병 콘텐츠에 대한 인공지능 요약 모델 개발	- Python TextRank 알고리즘을 활용한 콘텐츠 요약 모델 개발 및 결과검증 - 한국어 콘텐츠 요약 처리 시스템 구축 - 한국어 콘텐츠 요약 개방형 API 서비스 개발 - 재난형 동물질병 콘텐츠에 대한 인공지능 요약 모델 개선	100%
○재난형 동물질병 콘텐츠에 대한 인공지능 기계독해 모델 개발	- 동물질병에 대한 역학·방역 지식 학습용 QA 데이터셋 구축 - 인공지능을 활용한 콘텐츠 독해모델 생성 기술을 위한 질의응답 시스템 구축 - 인공지능을 활용한 콘텐츠 독해모델 생성 기술을 위한 개방형 API 서비스 개발 - 한국어 딥러닝 언어모델 학습 시스템 구축 - 저작권 등록(동물질병정보 기계독해 API(에이피아이) 시스템 소프트웨어)	100%
○기계독해 학습을 위한 학습 데이터 생성	- 재난형 동물질병 학습 데이터 10,219건 생성 - 데이터베이스제작제작자 권리등록(동물질병 질문 및 응답 데이터 세트 (ADQuAD))	100%
○콘텐츠의 활용도를 높이기 위한 정보의 DB화 및 인덱싱 기술 개발	- 역인덱스 구조 및 Full Text Indexing 기술 적용, 자체 한글형태소 분석기 Nori 탑재, RestAPI를 통한 데이터	100%

	쿼리 기능 개발, Ingest Node 기술 개발/ 다양한 수집기 Beats 기술 개발, 데이터 시각화 및 관리 기능 개발	
○콘텐츠 활용을 위한 서비스 기술(API) 개발	- Rest API를 통한 CRUD 처리기술 개발, 문헌자료 수집을 위한 문서변환 기술 개발	100%
○정보검색시스템 프로토타입 개발	- 해외 질병 데이터 및 백서·보고서 연계, Q&A DB 생성/관리 및 기계독해 연계	100%
○사용자에게 편의성을 제공하는 UX/UI 기획	- 데이터 분석 및 사용자 요구 수렴 사항을 기반으로 기능 및 화면정의와 디자인 및 퍼블리싱 수행	100%
○사용자의 요청에 적합한 정보를 제공하는 검색기술 개발	- 수행 문서를 바탕으로 사용자 요청에 적합한 정보를 제공하는 검색 시스템 [가축질병대응 AI-MRCQA] 개발 - 저작권 등록(가축질병대응 에이아이-엠알씨큐에이)	100%
○크롤링 대상 사이트 선정 및 데이터 분류를 위한 동물질병 용어사전 개발	- 카테고리 2개, 감염병 69개의 사전 항목 정의 - 재난형 동물질병 관련 역학·방역 정보 수집 시스템에 용어 등록	100%
○역학조사보고서·백서·공문 및 발간물 검토·관리대상 선정·등록	- 크롤링 서비스를 활용하여 수집할 수 없는 동물질병 관련 문서 선정 (백서 27종을 포함한 85종 선정) - 선정된 문서등록 (~포함한 ~종 등록)	100%
○재난형 동물질병 콘텐츠에 대한 인공지능 요약 결과에 대한 검증	- 인공지능 요약 결과에 대한 검증 수행	100%
○기계독해 학습을 위한 학습 데이터 생성 지원 및 검증	- 재난형 동물질병 학습 데이터 생성을 위한 조언 및 생성된 학습 데이터에 대한 검증 수행 - 데이터베이스제작제작자 공동 권리등록(동물질병 질문 및 응답 데이터 세트 (ADQuAD))	100%
○기계독해 결과에 대한 검증	- 기계독해 결과에 대한 검증 수행	100%

4. 연구개발성과 및 관련 분야에 대한 기여 정도

- 재난형 동물질병 콘텐츠 기계독해(MRC)를 위한 학습데이터 생성 표준 절차 도출
 - 특화된 콘텐츠 기반의 기계독해 연구자를 위한 안내서로 활용
 - ‘동물질병 질문 및 응답 데이터 세트(ADQuAD)’ 제작으로 재난형 동물질병에 대한 역학·방역 연구 활동 및 관련 산업발전에 기여

- 인공지능을 이용한 재난형 동물질병 정보 시스템 개발
 - 재난형 동물질병과 관련된 다양한 콘텐츠를 수집·가공·저장하는 빅데이터 플랫폼을 구축하고 인공지능을 활용하여 사용자의 요구에 적합한 콘텐츠를 검색·제공함으로써 재난형 동물질병에 대한 역학·방역 연구 발전에 기여
 - 재난형 동물질병에 빠른 역학·방역 대응으로 방역 비용 절감 및 축산농가의 경제적 손실 최소화 기대
 - 효율적인 질병발생 위험 경감 및 신속한 대응으로 대국민 신뢰감 고취 및 불편 최소화
 - 농림축산검역본부 전자도서관, 축산신문 등 관련 산업의 콘텐츠 검색 서비스의 고도화 기대

5. 연구개발성과의 관리 및 활용 계획

○ 연구개발성과의 활용분야 및 활용방안

- 수집·관리되는 역학조사서·백서·공문·발간물 등의 동물질병 비정형 콘텐츠에 대한 정보 분석에 활용
- 재난형 동물질병 역학·방역 콘텐츠를 활용한 인공지능 연구를 지원
- 재난형 동물질병 관련 인자들을 조화·분석하여 역학조사 및 방역 대책 등에 활용
- 위험요인을 다방면에서 조사·분석하여 효과적인 방역조치에 활용
- 콘텐츠의 범위를 추가·확장하여 국가방역체계 구축에 활용
- 주변국의 동물질병 발생정보 분석을 통한 경보체계로 활용
- 재난형 동물질병 관련 연구에서 활용할 수 있도록 개방을 원칙으로 하고, 사용자의 요청을 받아 내부 검토 후 접근권한을 부여함

○ 지속적인 재난형 동물질병 정보 수집

- 크롤링 대상으로 선정된 사이트에서 재난형 동물질병 콘텐츠를 지속적 수집
- 신규 유의미한 사이트에 대한 발굴 및 등록
- 농림축산검역본부에서 시범운영을 통하여 유의미한 역학조사보고서·백서·공문 및 발간물을 등록

○ 추가연구의 필요성

- 다양한 사용자의 검색 패턴에 따른 적합한 콘텐츠를 제공하기 위한 ‘기계독해 모델’ 고도화 수행 (주관기관 (주)노스스타컨설팅)
 - ▶ 검색 결과의 사용자 피드백(개선사항) 수집·분석 및 개선
 - ▶ 동물질병 질의 및 응답 데이터 추가 생성 및 기계독해 모델 고도화
- 재난형 동물질병 콘텐츠 요약모델 고도화를 위한 ‘재난형 동물질병 말뭉치 연구 (협동기관 농림축산검역본부)’를 수행
- 구축된 시스템을 시범 사용 및 개선점 도출(협동기관 농림축산검역본부)
 - ▶ 구축된 시스템의 효용성 평가 및 시스템 개선 사항 도출
 - ▶ 재난형 동물질병 콘텐츠 요약 서비스 고도화를 위한 ‘재난형 동물질병 말뭉치’ 개발
 - ▶ 임상수의사들의 현장 정보를 빠르게 수집·활용할 수 있도록 개선

○ 사업화 추진방안

- 주관기관 (주)노스스타컨설팅과 협동기관 (주)이지팜은 상호신뢰를 기반으로 공동사업을 추진
 - ▶ 구축된 ‘인공지능을 이용한 재난형 동물질병 정보 시스템’ 시범운영
 - ▶ 양 기관의 영업 네트워크를 활용하여 공동영업 추진
 - ▶ 인공지능협회, 축산협회 등이 주관하는 박람회 참여
 - ▶ 지속적인 시스템 고도화 (검역본부의 말뭉치 개발 결과 반영)
- 본 과제 수행을 통해 개발·확보된 기술을 활용한 사업 추진
 - ▶ 재난형 동물질병뿐만 아니라 일반 동물질병에 대한 정보 동물질병 빅데이터 플랫폼

구축

- ▶ 인공지능을 활용한 콘텐츠 요약 및 기계독해에 대한 검색 시스템 구축

○ 타 연구에 활용

- 특화된 비정형 콘텐츠에 대한 수집·가공·저장 및 검색 연구에 활용

○ 연구개발성과의 활용 계획

구분(정량 및 정성적 성과 항목)		연구개발 종료 후 5년 이내				
		2022	2023	2024	2025	2026
국외논문	SCIE					
	비SCIE					
	계					
국내논문	SCIE					
	비SCIE					
	계					
특허출원	국내					
	국외					
	계					
특허등록	국내		1			
	국외					
	계		1			
인력양성	학사	1	2	1	1	1
	석사					
	박사					
	계	1	2	1	1	1
사업화	상품출시		1			
	*기술이전		1	1	1	1
	**공정개발		1	1	1	1
제품개발	시제품개발					
성과홍보			1		1	
포상 및 수상실적						
정성적 성과 주요 내용						

* 기술이전: 통상실시권 및 타 연구 활용

** 공정개발: 프로그램 고도화

참고 문헌

- [1] Park, Jangwon. (2020). KoELECTRA: Pretrained ELECTRA Model for Korean. GitHub repository. <https://github.com/monologg/KoELECTRA>
- [2] Clark, K., Luong, M. T., Le, Q. V., & Manning, C. D. (2020). Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555.
- [3] 민재욱, 박진우, 조유정, & 이봉건. (2020). BERT 를 이용한 한국어 특허상당 기계독해. 한국정보처리학회지, 4(9), 145-152.
- [4] 임승영, 김명지, & 이주열. (2018). KorQuAD: 기계독해를 위한 한국어 질의응답 데이터셋. 한국정보과학회 학술발표논문집, 539-541.
- [5] 김영민, 임승영, 이현정, 박소윤, & 김명지. (2020). KorQuAD 2.0: 웹문서 기계독해를 위한 한국어 질의응답 데이터셋. 정보과학회논문지, 47(6), 577-586.
- [6] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- [7] 이태민, 박기남, 박정배, 정영희, 채정민, & 임희석. (2020). 기계 독해를 이용한 COVID-19 뉴스 도메인의 한국어 질의응답 챗봇. 한글 및 한국어 정보처리 학술대회 논문집, 32

주 의

1. 이 보고서는 농림축산식품부에서 시행한 가축질병대응기술개발사업의 연구보고서입니다.
2. 이 보고서 내용을 발표하는 때에는 반드시 농림축산식품부에서 시행한 가축질병대응기술개발사업의 연구결과임을 밝혀야 합니다.
3. 국가과학기술 기밀 유지에 필요한 내용은 대외적으로 발표 또는 공개하여서는 안 됩니다.