Text Analysis Development proposal

TECHNOLOGY FOR HUMAN SAFETY

WINITECH

TEAM PROJECT



유사어,뭐에요?



매뉴얼을 확인하는 시간단축을 위한 유사어 추출의 중요성 대두







- 프로젝트 추진 개요
- 구축 범위
- 조직 구성도
 - 프로젝트 일정,WBS
 - 예상이슈 및 참고자료

[O]









프로젝트 추진 개요

- 구축 배경
- 공공기관의 재난 관련 솔류션 제공 기업으로서 <mark>재난관리 매뉴얼</mark>과 관계있음.
- 고객의 요청 사항 중 매뉴얼의 수직 비교에 대한 중요성 대두.

구축 목표

국가법령정보센터에서 소방,범죄,재난의 데이터 수집 후 웹 페이지 형식의 유사 항목 추출에 적합한 시스템 구축













프로젝트 구축범위

소스데이터

국가법령시스템 크롤링 데이터

구축범위

자연어 분석을 통한 유사도 분석 Python 프로그램

기대효과

업무 정확도 향상 시간 및 비용 절감 비지니스 성과 향상 사용자 만족도 향상













조직구성도

프로젝트 관리자(PM) 임도영 팀원 팀원 김도형 정 도 데이터 수집 및 전처리 자연어 처리, 모델 구축 웹페이지 구축







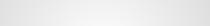






프로젝트 구성도





유사도분석 중심 우선 구축

(검색 문장을 입력하면 법의 3단 비교 기능 웹 구현 예정)













프로젝트 구성도

HTML 프레임 워크 : 3.9.15

예정 모델:

개발 툴:

Python 3.9.13

텍스트 유사도 분석 모델 01. 키워드 입력

기능:

02. 동일단어, 유사 단어, 관련 표현 파악가능

03. 유사 키워드 출력







WINITECH







프로젝트일정

06

	1주차 (~ 5. 4.)	2주차 (~ 5. 12.)	3주차 (~ 5. 19.)	4주차 (~ 5. 26.)	5주차 (~ 6. 2.)	6주차 (~ 6. 9.)	7주차 (~ 6. 16.)	담당자
분석 기획								
- 프로젝트 계획 수립								팀 전체
데이터 준비								
- 자료 수집								팀 전체
데이터 정제								
- 전처리								김도형
알고리즘 구현								
- 모델 구축								임도영, 정도
- 모델 평가 및 검증								정도
- 구현 페이지 구축								팀 전체
평가 및 전개								
프로젝트 평가 및 보고								













프로젝트일정

Work Break-down Structure																	5	릴(시	1간	단우	l)										
vvoik break-down Structure							3일						4일						8일						9일						
rocedures S	Steps	Tasks	담당자	작성일	산출물/비고	9	10	11	12	13	14 1	7	9 1	0 11	1 12	13	14	17	9	10	11	12	13	14	17	9	10	11	12	13 1	4 17
1.0.0. 착수 및 프로젝트 관리		\Box																													
1.	.1.0 프로젝트	관리					Т			П		\neg				1	Т											\neg		Т	Т
	1.1.	일간회의	WINITECH 팀 전체	매일	회의록, 데이터 정의서																									\top	\top
	1.2.	완료 최종 회의	WINITECH 팀 전체	6월 15일	ppt																									\top	\top
2.0.0. 계획/	너 작성																														
2.	1.0. 프로젝트																														\top
	2.1.1.	프로젝트 배경 및 시행목적 작성	임도영	5월 3일	프로젝트 계획서 ppt											1		1												\neg	\top
	2.1.2.	프로젝트 일정 정의	WINITECH 팀 전체	5월 3일																										\top	\top
	2.1.3.	예상 이슈 작성	WINITECH 팀 전체	5월 3일	프로젝트 계획서 ppt				-			\top						1												\top	\top
	2.1.4.	WBS 작성	정도,김도형	5월 8일	WBS							_						1												\top	\top
	2.1.5.	최종 계획서 작성	WINITECH 팀 전체	5월 8일	프로젝트 계획서 ppt, 계획서		t			t		1																		\top	\top
	터 수집 및 전																														
3.	1.0. 수집_분	V.										\top											(0.		10						一
	3.1.1.	데이터 수집	WINITECH 팀 전체	5월 8일	소방(정도),범죄(김도형),재난(임도영		1					\top	_	+	1	-															
	3.1.2.	데이터 전처리	김도형		자연어 처리 및 토큰화	_	\vdash					\top		+	-	1	1	1												_	_
.0.0. 모델링	Acres and Artista																														
4.	1.0. 모델링																													—	
	4.1.1.	머신러닝 기반 모델링	임도영, 정도		피클 파일 및 팡션		\vdash	1	<u> </u>	t		\top		\top		+	T	1						\vdash						+	+
	4.1.2.	딥러닝 기반 모델링	임도영, 정도		광 션	-	\vdash			\vdash		\top		\top		+		1						\vdash						\top	+
a a Fitic	_	모델 검증	정도		ру파일	\vdash	1		t	t		\dashv		+	1	1		t			<u> </u>		1	t	t					+	+
5.0.0. 디자인	7/1	TIOI.																													
5.	1.0. 메인 디기		WINDTEGLI EL 74 til		디자인 파일													1									-				
	5.1.1.	WEB메인 디자인 구성	WINITECH 팀 전체		니샤인 파일	_	t	1	1	\vdash		+	-	+	1	+	1	1						 						+	+
	5.1.2.	WEB메인 디자인 컨펌	WINITECH 팀 전체			_	+	\vdash	 	-		+		+	+	+	-	+	\vdash	+	1	\vdash	!	-	-			\rightarrow	-	+	+













프로젝트 요구사항 정의서

구분	서비스 (메뉴)	기능명	기능 설명	우선 순위
Common	Data	자료 수집	### COMPANIES OF THE PROPERTY	1
		전처리	자연어 처리, 토큰화	2
User	Model	유사도	1	2
		모델 검증	각 모델들의 유사도를 검정	3
	Web	검색	문서 내 찾고자 하는 단어를 검색	2
	0. West (#1,820)	유사도 시각화	기본법, 시행령, 시행규칙 내 검색 단어의 유사도를 시각화	3
	Graphs	유사도 시각화	문서간 유사도를 시각화하여 그래프로 나타냄	3











대응방안



80



프로젝트 예상이슈

프로젝트 진행 시 발생할 수 있는 이슈안 대응방안

예상이슈

- **수집한 자료 사용 가능 여부** (수시로 변화하는 법과 매뉴얼)
- 유사도 분석 시 낮은 정확도

중복 단어 및 문장 출력 위험성

데이터 사용여부 선확인 (수시로 변화하는 법과 매뉴얼)

전처리 과정 개선

코사인,자카드,유클리드 거리,맨허튼 등 다양한 유사도 측정 알고리즘 사용

단어 및 문장 임베딩,트랜스 포머기반 모델 등 다양한 특징의 벡터 사용

중복단어문장 과 유사도 기반 필터링

출력 결과의 다양성 확보















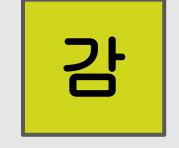






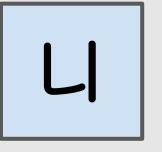
WINITECH_TP

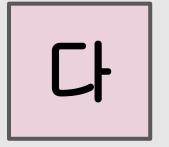










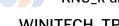












KNU_k-digital training님 외 2023명이 좋아합니다.

WINITECH_TP 위니텍 팀 프로젝트 계획서 발표를 마치겠습니다~! 끝~~