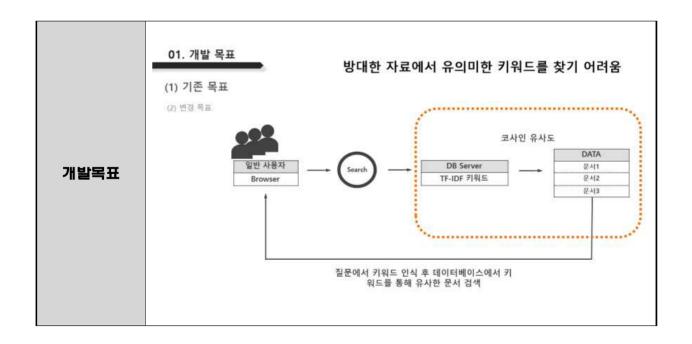
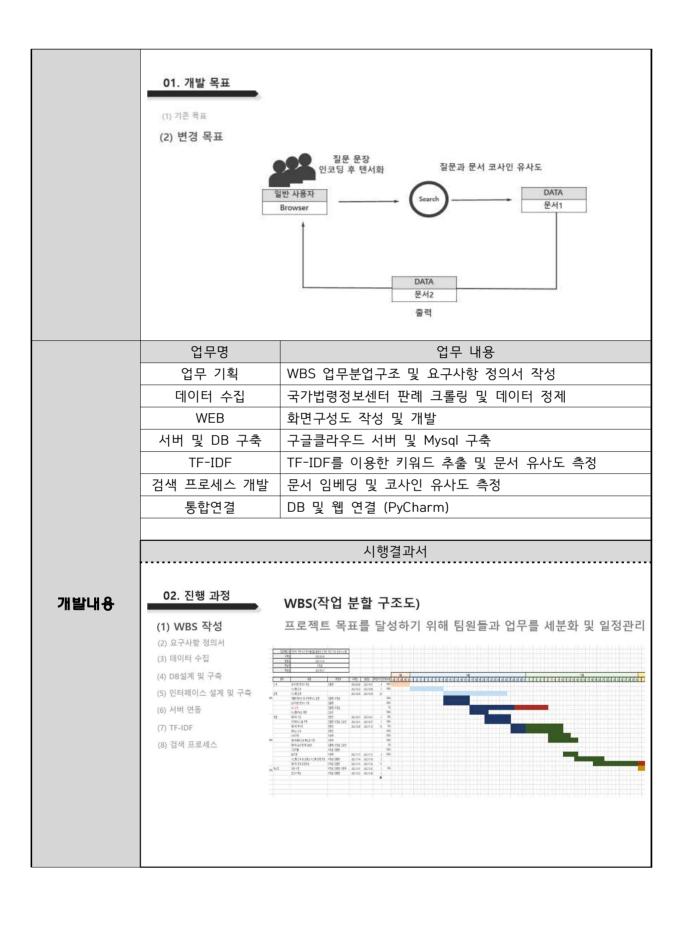
# Projet (프로젝트)

프로젝트명	자연어 기반 R과 텐서플로를 활용한 온라인 챗봇 지식 검색 시스템				
개발기간	2022.09.24. ~ 2022.12.08				
참여인원	4명				
담당업무	서버 및 데이터베이스 구축, 데이터 수집(크롤링), 데이터 전처리				
개발환경	Toad for Mysql(sql), google cloud platform(웹서버,DB서버), Python(Colab)				
사용도구	Pandas, Numpy, tensorflow, sql, ssh				

	진행과정 요약
	1)프로젝트를 달성하기 위해 wbs(작업 분할 구조도)로 팀원의 활동과 업무를 세
	분화하여 프로젝트 역할 및 작업관리
프로젝트	2)요구사항 분석을 통해 필요한 기능 분석
_포크= 요약	3)검색 시스템을 만들 데이터 크롤링
<b>≖</b> ≒	4)데이터베이스에 데이터 저장
	5)사용자 인터페이스(UI) 설계 및 제작
	6)검색 프로세스 개발
	7)질문과 문서의 코사인유사도를 이용하여 문서추출





#### 02. 진행 과정

#### (1) WBS 작성

#### (2) 요구사항 정의서

- (3) 데이터 수집
- (4) DB설계 및 구축
- (5) 인터페이스 설계 및 구축
- (6) 서버 연동
- (7) TF-IDF
- (8) 검색 프로세스

#### 프로젝트 개발의 단계에서 요구할 사항을 정리





#### 02. 진행 과정

- (1) WBS 작성
- (2) 요구사항 정의서

### (3) 데이터 수집

- (4) DB설계 및 구축
- (5) 인터페이스 설계 및 구축
- (6) 서버 연동
- (7) TF-IDF
- (8) 검색 프로세스

# 국가법령정보센터 판례데이터 크롤링 후 데이터 정제

#### 국가법령정보센터 (law.go.kr)





#### 02. 진행 과정

#### (1) WBS 작성

- (2) 요구사항 정의서
- (3) 데이터 수집

#### (4) DB설계 및 구축

- (5) 인터페이스 설계 및 구축
- (6) 서버 연동
- (7) TF-IDF
- (8) 검색 프로세스

# DB설계 후 MySQL로 DB구축

		2014221		The same of the sa	
41.64	Appen pain	998 30001818 9399 30001878	101 100	(日本) (日本) (日本) (日本) (日本) (日本) (日本) (日本)	
0.6	\$40,000			**************************************	
100	- Partition of the last of the	principal descriptions	3 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -	SALE PROPERTY.	
1 459	4.7	付かり 本管 (特別) ウギ	142	A Parameter Control of the Control o	
	从意刊起	AUMBER MIST MOL	11.57	- Control of the Cont	
Otal	75.8	HARCAGICION:		A STATE OF THE PARTY OF THE PAR	
1	100	34419482,901		- 175 - 175	
1 000	17	MARCHARDORN.		The state of the s	
E State Code	800.85	049CH40200:		TOP TO STATE OF THE PARTY OF TH	
6   tigs, colo	20236	- All Marie	-		6 00 mm
		T49 00001919			
11147	ANDIER, SHIS	0.000 HE FE	1910 1128	The second second control of the second seco	
0.4	SACHUR.	1		The state of the s	
	IN NOT THE	7.7		Comment of the Commen	
10 100	17	STATE OF THE PARTY OF	T#D	The second design to the second state of the second state of the second	
411,500	5500	MARCHINE (SIX)		Committee of the property of the second seco	TOTAL AT A SEC. TAL.
T 199,30%	2241	W85WW0596		The state of the s	
1. gra, item	教育円金	SARCHIRGISHS		The second of th	
E. papered	1110-1-1	PROTOGRAM:		THE RESIDENCE OF THE PARTY OF T	
6 g4,19974	1114	\$6\$(448)390			rt break income
6 January Street	1827	9550,0000000			
			111111	The second secon	
111(6)	HECOD DATA	A ER PROCESS	995 1989	The state of the s	-
84	SAT OF SHIP	0.00m 3000 HB /B	2018		
	OWN TO CHICK				to be before the
198	4.9	THE PER POST TO	140	Terretario mandrida Querra, incomprenda que la comprenda france de la comprenda del comprenda de la comprenda	
Arrest Ac.	E-12 CAG	nidCount(100)	- 74		min pri de la constitución de la
T permet per	5345 586	maki (maki) kati		The second of the last of the second of the	Al-resident
1   pretent nor	FIGUR CAS.	VARCANES (RED.		NI SOLDERSON	





(6) 서버 연동

(3) 데이터 수집

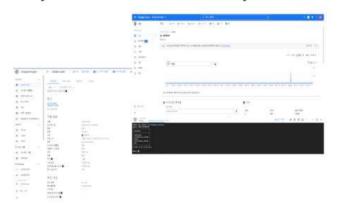
- (7) TF-IDF
- (8) 검색 프로세스



# 02. 진행 과정

# Toad for MySQL 프로그램을 구글 클라우드 MySQL 서버연결

- (1) WBS 작성
- (2) 요구사항 정의서
- (3) 데이터 수집
- (4) DB설계 및 구축
- (5) 인터페이스 섬계 및 구축
- (6) 서버 연동
- (7) TF-IDF
- (8) 검색 프로세스



# 02. 진행 과정

- (1) WBS 작성
- (2) 요구사항 정의서
- (3) 데이터 수집
- (4) DB설계 및 구축
- (5) 인터페이스 설계 및 구축
- (6) 서버 연동
- (7) TF-IDF
- (8) 검색 프로세스

# TF-IDF는 단어의 빈도와 역 문서 빈도를 사용하여 각 단어들마 다 중요한 정도를 가중치로 주는 방법



-4		
WAS	GE MT	-117,78.00
82	21.24	135.85
\$C.	1915	-71.3942
4	出海	-530.951
20	발행	-201.785
	det rap.	-292.50
952	2131	-39.5094
#1.	es w	<366.671
	문제	19.798.139
HEC:	411 E.I	-20.7944
110	28	+170 mt.
65	86.53	-452.621
18.	가장	15,9421
W47	04.90	150,8721
19.	8118	1952,831
10	m.m	16.6351
110	田様	FD4.661
10	20 M	388.764
944	91 K)	45333.53
201	大雅.	-104.000
25	母原	+817.835
12	29-89-46	-3,46574
22	3010年	:-0.09315
20.	49.50	-0.69311
#	-9.81	-94.9612
to:	世が世	170,6310
MC.	77.80	-9.59(834
20	神妙林	-3:37254
29.	2) 包	-2.7725
Mr.	型麻	76.919
11	07 W	11,25.46
NE.	<b>*</b> U	K12.608
NO.	0/10	138.2%
34	입지	-61.6901
MIC)	40	+10142.1
20	料品	420.74
an.	22.44	-258,54
44	<b>相外解</b>	-69.3141

	(1) WBS 작성 (2) 요구사항 정의서 (3) 대이터 수집 (4) DB설계 및 구축 (5) 인터페이스 설계 및 구축 (6) 서버 연몽 (7) TF-IDF (8) 검색 프로세스 (8) 검색 프로세스	Intification of American Statement of Americ			
	프로젝트 실적	보완사항			
결 과	검색 프로세스 개발 1) 문서 임베딩 2) 질문 문장을 인코딩 후 텐서화 3) 저장한 임베딩 데이터와 코사인 유 사도 측정 4) 가장 높은 유사도를 가진 문장 추출	1) 데이터베이스 연결을 통해 데이터 등록 시간 최소화 2) 웹 연결을 통해 챗봇 UI 시각적으로 구현 3) 유사도 측정뿐만 아니라 다양한 특징 추출을 통해 검색 프로세스 개선			
	o 역량강화 데이터 전처리 기법, DB 구축 등 다양한 실무기술 습득				
기대효과	ㅇ 취업연계       데이터가 과다한 현재 사회에서 데이터를 다루는 기술을 익힘으로써 데이터 및       아케텍쳐 관련 회사 취업에 긍정적인 영향을 줄 수 있음				