# 울산대학교 ChatBot.

이태관 백동수 이진렬

©Saebyeol Yu.Saebyeol's PowerPoir

#### 1.목적과 수행방법

학과 생활을 하다보면 여러가지 질문이 생기기 마련이다.

질문을 할 수 있는 수단은 많다. 학과 홈페이지에 검색을 하거나 Q&A게시판에 질문을 하거나 또는 관련 부서에 전화 문의를 하면 된다.

그러나 검색을 통해 정보를 찾기 위해선 산발적으로 분포 되어 있는 정보들을 하나하나 탐색 해 가며 원하는 정보가 나올때까지 기약 없이 검색 해야한다.

이번 프로젝트를 통해 챗봇을 만들어서 산발적으로 분포 되어 있는 정보들을 하나로 통합하여 울산대학교 재학생들에게 편의성을 제공 하겠다.



# 목차 A table of contents.

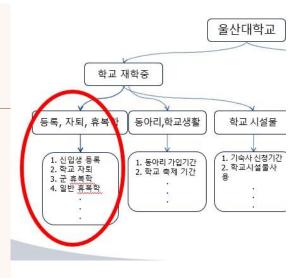
- 1 사용자 요구사항
- 2 시스템 요구사항
- 3 기능적 요구사항 비기능적 요구사항





#### 사용자 요구사항

- 1. 사용자는 각자가 궁금한 사항들을 질문 할 수 있다.
- 2. 사용자의 질문은 문장이되, 되도록 짧고 명확한 한 문장이 좋다.
- 3. 한시적으로 질문의 영역은 등록, 자퇴, 휴복학 관련에 한 한다.
- 4. 사용자는 질문을 하면 답변을 받을 수 있는데, 적절한 답변인지 아닌지 평가를 할 수 있다.



 $@Saebyeol\ Yu.Saebyeol's Power Point\\$ 

시스템의 전체적인 진행 시나리오는 다음과 같다

Part 1

#### 사용자 요구사항

# 시스템의 진행

1단계	2단계	3단계	4단계	5단계
질문	답변	피드백	1~3단계 반복	종료
사용자 → 시스템 사용자 ← 시스템		피드백때 사용자가 원하지 않는 답변임을 알려준다면 시스템은 다른 답변을 제시한다. 사용자가 원하는 답변을 제공 받음 또는 원하는 답변을 지속적으로 얻을 수 없을 때 시스템은 해당 질답을 종료한다.		
4	<b>-</b>			

©Saebyeol Yu.Saebyeol's PowerPoint

첫번째로 사용자가 시스템에게 질문을 한다.

두번째로 시스템은 사용자에게 답변을 한다.

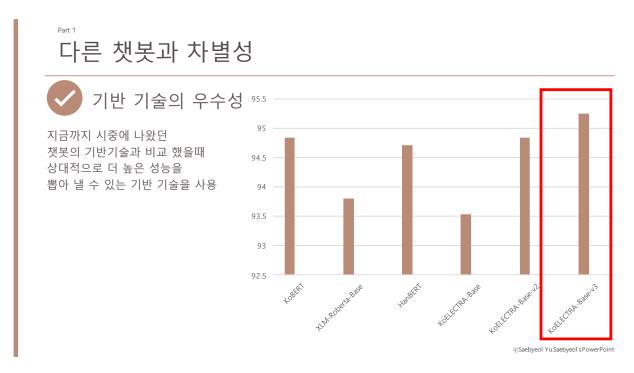
세번째로 사용자는 받은 답변이 만족스럽다면 바로 5단계로 뛰고 만족 스럽지 않다면 1~3단계를 반복한다.

Part 1

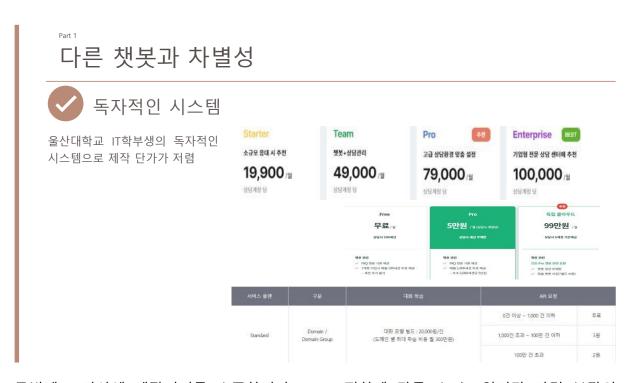
#### 다른 챗봇과 차별성

- 기반 기술의 우수성
- 독자적인 시스템
- 무궁무진한 성장 가능성 Q&A 데이터가 축적이 되면 될 수록 유지보수를 통해 더 좋은 성능 기대 가능

기존 챗봇과 본프로젝트에서 개발할 챗봇의 차별점을 살펴 보겠다.

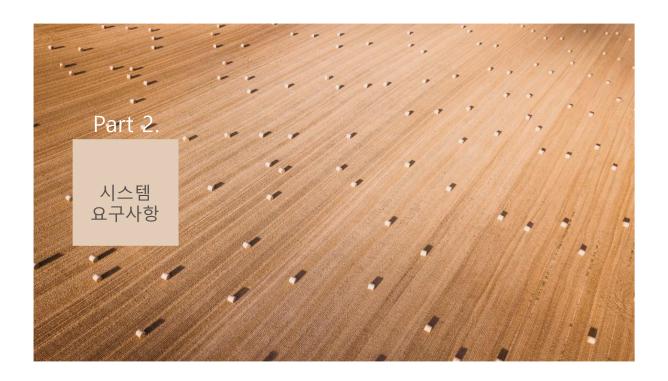


첫번째로 본 프로젝트에서 진행 하는 시스템의 기반 기술이 현재 공개적인 기반 기술 중에선 가장 신식이고 더 높은 성능을 가지고 있다.

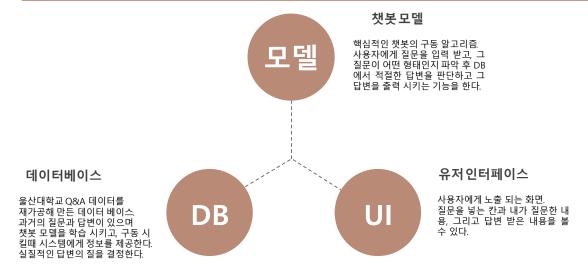


두번째로 타사에 제작의뢰를 수주하거나 API로 편하게 만들 수 는 있지만 가격 부담이

조금 있는 편이다. 그러나 프로젝트로 만들어진 독자적인 시스템으로 제작 단가가 저렴하다.



챗봇 시스템의 모듈



챗봇 시스템은 3개의 모듈들이 합쳐서 서비스 가능하게 한다.

챗봇 시스템의 모듈 결합 과정

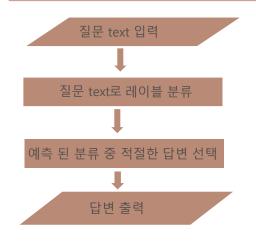


©Saebyeol Yu.Saebyeol's PowerPoint

모듈을 합치는 과정은 위 그림과 같다.

다음은 챗봇 딥러닝 모델 모듈의 Flow chart를 살펴 보겠다.

챗봇 딥러닝 모델 Flow Chart



#### Flow Chart

딥러닝모델의흐름도입니다 기본적으로단일레이블다중분류형태의 모델입니 다.

다음으로레이블분류하는 상세 알고리즘을살펴 보 겠습니다

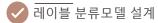
©Saebyeol Yu.Saebyeol's PowerPoint

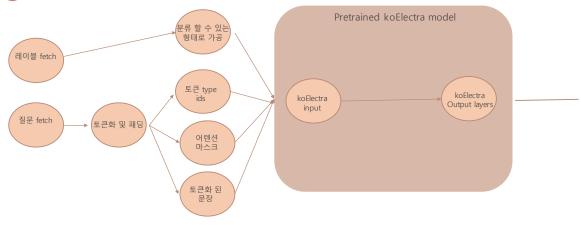
위 그림은 챗봇 모델의 전체적인 순서도 이다.

다음 그림에선 두번째 순서인 레이블 분류에서 상세 알고리즘을 살펴보겠다.

Part 1

#### 챗봇 딥러닝 모델 Flow Chart



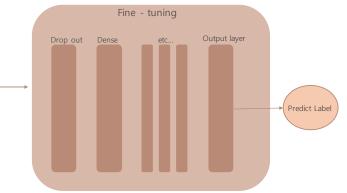


©Saebyeol Yu.Saebyeol's PowerPoint

Part 1

#### 챗봇 딥러닝 모델 Flow Chart

레이블 분류모델 설계



#### koElectra를 Fine tune

챗봇모델은koBer의 진화형인koElectra finetuning 해서만들었다

단일레이블다중 분류모델이기때문에Cros£ntropy Los를 사용한다

그러나본 프로젝트에선학습 데이터 불균형이있었 기에 FocalLos를 사용하였다

Focal Loss 참조:

https://arxiv.org/pdf/1708.02002.pdf

# DB 전처리

# Data 출처

- 울산대학교 UWINS Q&A 광장
- 직접 크롤링 해서 얻었다.



©Saebyeol Yu.Saebyeol's PowerPoint

데이터 크롤링은 python 언어의 BeautifulSoup과 Selenium 라이브러리를 혼합하여 진행 하였다.

다음은 전처리 하는 과정에서 분석된 3가지의 문제와 해결방안을 설명 하겠다.



• 한 질문에 여러 개의 질문과 여러 개의 답변이 모호하게 분리 되어 있는 경우

Case 1

■ 한 질문에 여러 개의 질문과 여러 개의 답변이 명확하게 분리 되어 있는 경우

Case 2

Case 3

• 한 질문에 여러 개의 질문이 있으나 답변이 부분적으로 되어 있는 경우

■ UwinsData중 학습의정확도를향상시키는데 방해가되는요소들을분석하고해결하였다

#### DB 전처리



©Saebyeol Yu.Saebyeol's PowerPoint

Part 1

# DB 전처리



#### DB 전처리



©Saebyeol Yu.Saebyeol's PowerPoint

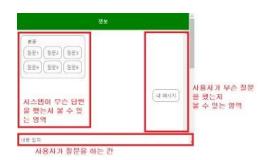
다음은 인터페이스 요구사항을 살펴 보겠다.

Part 1

#### UI - 인터페이스 요구사항

메인 화면을 구성하는 기능은 다음과 같다

- 1. 사용자가 질문을 입력 할 수 있는 칸
- 2. 사용자가 무슨 질문을 했는지 볼 수 있는 영역
- 3. 시스템이 무슨 답변을 했는지 볼 수 있는 영역







#### 기능적 요구사항

- 사용자에게 각자가 궁금한 사항들을 질문 입력 할수 있는 영역을 제공 해야한다.
- 2. 시스템은 사용자가 질문을 입력 하면 적절한 대답을 출력 해야한다.
- 3. 사용자가 원하는 대답을 받지 못했을때 시스템은 다 른 답변 또는 솔루션을 제시해야한다
- 4. 사용자는 질문 하는 것이 육체적으로 심리적으로 불 편하지 않아야 한다

# 비기능적 요구사항

프로덕트 요구사항 - 유용성

- 1. 사용자에게 질문 입력 완료를 Enter 키 또는 마우스 클릭으로 표현 할 수 있게 제공 해야한다
- 2. UI의 채팅박스는 수직적으로 밑으로 계속 쌓는다. 이때 화면의 세로 길이는 증가한 다
- 3. 사용자의 입력 또는 시스템의 출력 시 자 동으로 스크롤바가 제일 밑으로 이동되 게 한다.



# 비기능적 요구사항

프로덕트 요구사항 - 신뢰성

1. 대답 가능한 영역을 축소 하더라도 최대한 답변에 대한 정확도를 높여야한다



# 비기능적 요구사항

프로덕트 요구사항 - 효율성

1. 사용자가 질문 입력 후 답변 받는 데 까지는 2sec 내외의 시간으로 한다.



# 비기능적 요구사항

조직 요구사항 - 배포

- 1. Django를 이용하여 웹페이지를 만든 후 배포 및 서비스 한다
- 2. 아직 미정



# 비기능적 요구사항

#### 조직 요구사항 - 구현

- 1. Python을 메인 언어로 하여 구현 한다
- 2. 딥러닝 모델은 pytorch로 작성한다
- 3. Data 크롤링은 BeautifulSoup과 Selenium을 적절히 혼합해서 한다
- 4. Python 개발 플랫폼은 Pycharm – professional 버전을 사용한다



#### 비기능적 요구사항

조직 요구사항 - 윤리

1. 무분별한 학습으로 윤리적으로 어긋난 대답을 피하기 위해서 일반 대화 채팅은 아예 제거하고 오로지 울산대학교 관련 한 답변만 기능한다.

