기달리뷰
PROJECT

졸업작품3 4분반 1조

201835455 박윤재
201835514 장지아
201935120 임채윤

01

Overview

Project motivation and brief description of system structure and progress

02

IDEA

Implementation of developed review analyzer and maintenance

03

Demo Video

Watch the demo video

04

Marketing Plan

Strategy analysis and methods for future marketing

05

Cooperation and Roles

Evaluation and future plans for our project

01 Overview

01 Motivation

Overview

Time spent at home
increases due to COVID-19

Delivery services are widely spread

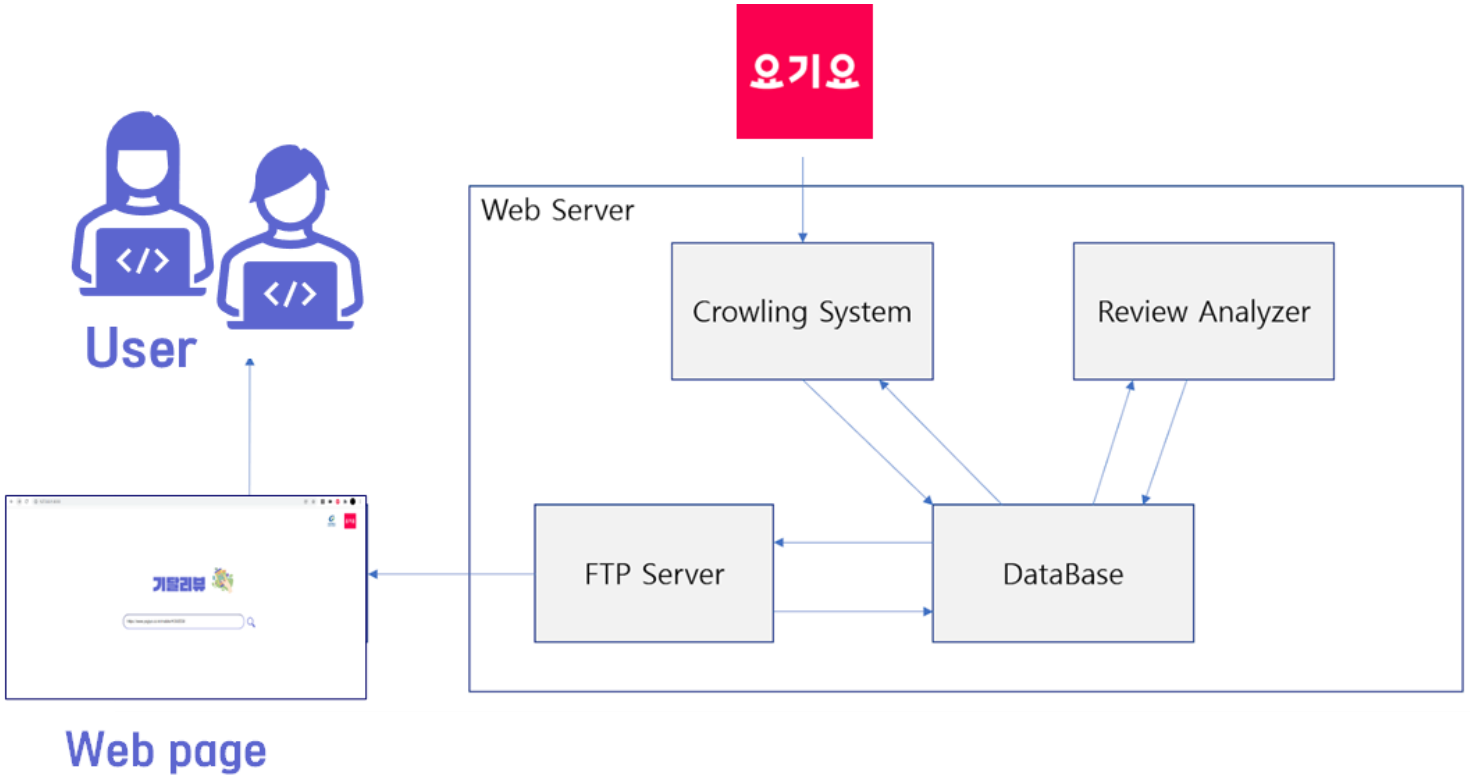
[식탁전쟁]① “별점 1개면 장사 끝”… 배달앱 전성시대의 그림자 ‘리뷰 갑질’

'배달앱 리뷰' 보고 주문? 당신은 치밀하게 낚였다

[주장] 코로나보다 더 무서운 고객리뷰... 배달 앱사, 자영업자 위해 방지책 마련해야

01
Overview

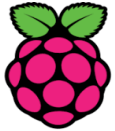
System structure



01 Progress

Overview

Server & Database



REAL VNC

Review analyzer

koelectra-review-finetune.bin

learn_data.db

word2vec.model

best_model.pkl

run.bat

Collecting data

Crawling automation system.py

total	taste	quantity	delivery	com_name	com_time	comment
필터	필터	필터	필터	필터	필터	필터
5	5	5	5	an**님	2021년 8월 24일	육아로 바쁜...리뷰진짜 안쓰는 1인입…
5	5	5	5	손님	2021년 8월 27일	맛있습니다. 많이시켰는데 시간내에 …
5	5	5	5	an**님	2021년 8월 27일	주말동안 쟁여놓고 먹으려고 주문했…
5	5	5	5	go**님	2021년 8월 29일	맛있습니다 깔끔하고
5	5	5	5	na**님	2021년 8월 30일	맛있어요! 마카롱 되게 괜찮아요
5	5	5	5	73**님	2021년 8월 30일	커피 잘받았습니다 전화받으신 사장…
5	5	5	5	fr**님	2021년 9월 2일	배달도빠르고 커피가 쓰지않고 맛있…
5	5	5	5	an**님	2021년 9월 16일	양버터는 처음 주문해봐요 커피랑 먹…
5	5	5	5	pi**님	2021년 9월 18일	맛있었어요! 디카페인커피도 괜찮았…
5	5	5	5	an**님	2021년 9월 19일	다른리뷰보고 디카페인 주문해봅니…

Website

02 IDEA

01 Review analyzer

02 Maintenance

03 Future Plan

02
IDEA

Review Analyzer – Similar Reviews

별점별 대표 리뷰

별점	리뷰	유사 리뷰	
★★★	맛은 있어요 그런데 왜 제가드린 주소가 아니고 다른데 일까요?일회용 숟가락젓가락은 왜 안주시나요? 정말실망했네요 배달하신 분께라도 죄송하다고하세요 여기저기 시간이 돈이신분인데... 이건아니라고봅니다	맛은 좋았어요! 다만 찬합에 있는 김발이,핫도그 사이드로 따로 시킨 김발이,핫도그 전부 안에가 차갑네요! 제대로 튀기지않으신 거 같네여ㅠㅠ 김발이,핫도그 못 먹고 그냥 버렸습니다그리고 국물떡볶이 쌀떡인지 밀떡인지 알 수 있나요?	가격대비 양은 생각보다 좀 적었지만..맛은 좋았습니다. 따뜻해서 맛이 더 있었습니다!!!
★★★★	맛잇엇어요^^전산오류로 밥이빠졋지만ㅎ	공수간 늘 맛있어요~~	맛있어요 ~~ 수고하세요 ~~
★★★★★	맛있게 잘 먹었습니다	맛있게 잘 먹었습니다~	맛있게 잘 먹었습니다!!!

02
IDEA

Review Analyzer – Similar Reviews

1. Vectorize the tokenized word through the Word2Vec technique.
2. Calculate the average of word vectors through the Document embedding technique.
3. Find cosine similarity between reviews.
4. Sort in order of high similarity.

02
IDEA

Review Analyzer – Star Rating Prediction Model

소담매운갈비찜&찜닭-호매실점

별점 4.8 ★★★★ 예측 별점 4.6 ★★★★

Average
of predicted star ratings

리뷰

빠른 배달^^맛도 완전 좋았어요

궁금해서 시켰는데 맛있어요~

아들 권유로 첨 먹어본 샹궈 넘맛있었경 담에 고오옥 다시배달 주문할 예정

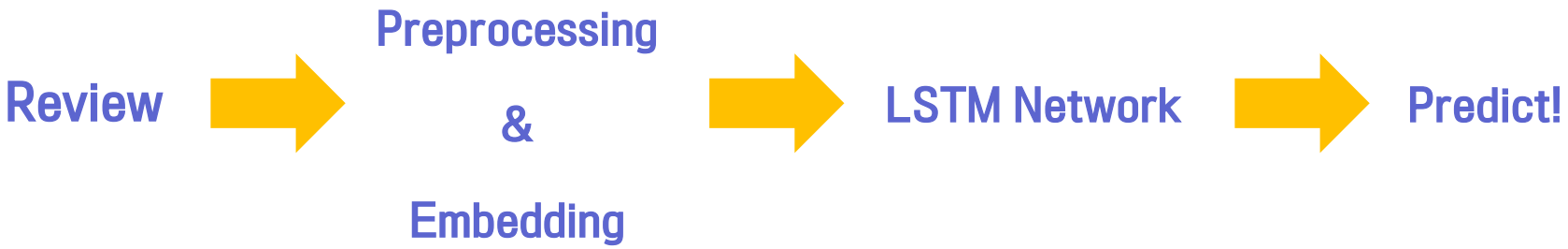
짧은 머리카락이 마라탕에서 나왔어요신경써주세요마라탕은 내용물 선택할 수 없는부분이 너무 불편하네요대체로 좀 밍밍하네요

02 IDEA

Review Analyzer – Star Rating Prediction Model

Total Crawled Reviews : 65,584

Total Dictionary words : 13,864



02
IDEA

Review Analyzer: Star Rating Prediction Model

Learning Reviews

- Restaurant web address is input from the user.
- Crawl restaurant information and reviews.
- Preprocess the review and load the learning model to derive predictive ratings.

02
IDEA

Review Analyzer: Review Event Prediction Model

리뷰	리뷰 이벤트 참여 가능성
맛없어요. 내 인생에 최악의 맛!	리뷰 이벤트 참여 가능성 낮음
모유수유중이라 안매운걸로 시켰는데 안짜고 안달고 양도많고 딱 맛잇어요 좋아요 !!	리뷰 이벤트 참여 가능성 낮음
너모 맛나요 신라면정도인데도 조금 맵기는 합니다마스크도 주셨어요	리뷰 이벤트 참여 가능성 낮음
맛있어요!매콤매콤하고 갑내 안나고 고기가 큼직큼직했어요양이 진짜 많아요ㅎㅎ배달은 좀 느렸지만 맛있게 잘 먹었습니다	리뷰 이벤트 참여 가능성 낮음
갈비찜만 시켜먹다가 삼겹살은 처음인데 엄청 맛있어요!!같이오는 떡이 레전드~~ 개꿀맛! 담엔 떡 추가 필수.. 김치찌개에 햄도 들어가고 양도 엄청 많고 맛있네요계란찜도 폭신폭신 맛있게 잘 먹었습니다!	리뷰 이벤트 참여 가능성 낮음
맛있게 잘 먹었습니다.늦은시간에 시켰는데 괜찮네요	리뷰 이벤트 참여 가능성 낮음
계란찜 귤 서비스까지 주시고 감사합니다..곱창전골은 기성품제품인듯해서 아쉽네요..	리뷰 이벤트 참여 가능성 높음
ㅎㅇ맛집저번에 먹었을때 떡이 레전드로 맛있어서 떡추가했어요ㅎㅎ갑자는 혹시나 안될까봐 걱정했는데 가능해서 너무 좋아요ㅠㅠ♡삼겹살이 큼직큼직하고 겉바속촉이라 맛있게 잘 먹었어요!!(버섯,마늘은 안먹어서 빼달라고 했어요)	리뷰 이벤트 참여 가능성 낮음

02 IDEA

Review Analyzer – Review Event Prediction Model

Participation in the review event

- Mark the result of predicting participation in the review event (high, low)
  - ➢ Implement filtering function so that you can only see the results of the selected participation.
  - ➢ Line Up function is implementation.

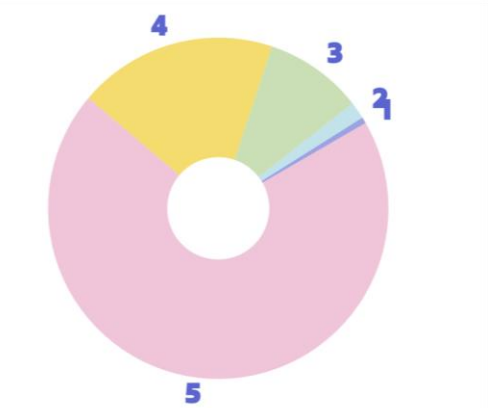
Review event prediction Model

- Collect 1,600 pieces of data each that participated in and did not participate in the review event.
  - ➢ It collects data by directly determining whether to participate in the review event.

- Train the KoELECTRA model with the collected data.
- Implement a model that can determine participation in the review event.

Review Analyzer – Review Visualization

별점 분포



Star Rating Distribution

Word Cloud

Main Keyword

02 Maintenance

IDEA

gp**님 어제

★★★★★ | 맛 ★ 5 양 ★ 5 배달 ★ 5

배달도 빠르고 맛있어용너무

mi**님 2일 전

★★★★★ | 맛 ★ 5 양 ★ 5 배달 ★ 5

보쌈정식/1

맛있어요 서비스도 항상 챙겨주시고요

Gradual Learning

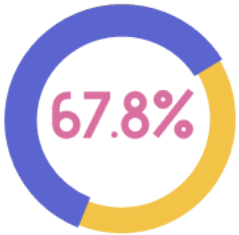
Reviews are **constantly updated**, so we implemented adding words to dictionaries **on a regular basis** and automating learning


```
def new_or_old_rev(dataf): #Determining if this is a new, untrained review
    sql_ = sqlite3.connect('database/Learned_review.db')
    try:
        learned_rev = pd.read_sql('SELECT * FROM Learned_review', sql_, index_col=None)
    except:
        learned_rev = pd.DataFrame()
    sql_.close()

    dataf = dataf[dataf.duplicated(keep='last') == False]

    not_inner=pd.concat([dataf,learned_rev])
    not_inner=pd.concat([dataf,not_inner[not_inner.duplicated(keep='last')]]) #To check for duplicates,
                                                                              # concat -> already learned if it is a duplicate

    not_inner = not_inner[not_inner.duplicated(keep=False)==False] #Filter only non-duplicate reviews

    print("{} reviews exist, {} reviews are new\n".format(len(dataf)-len(not_inner),len(not_inner)))
    return not_inner

def dtm(dataf, word_lib, rest):
    dataf['rest'] = rest
    dataf = new_or_old_rev(dataf) #Determining if this is a new, untrained review

    total_score = {}
    total_word = {}
    if len(dataf) == 0:  #If 0, there are no reviews to learn
        print("Already new data\n")
        return

    learning_data = []
    print("Start Preprocessing...")
    word,score,sent=kor_preprocessing.kor_preprocessing(dataf)
```

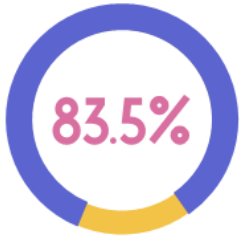
02 IDEA

Maintenance

New learning model



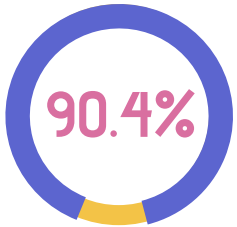
SVM



Margin of Error ±1

More Faster!

More Precisely!



LSTM Network

02 Maintenance

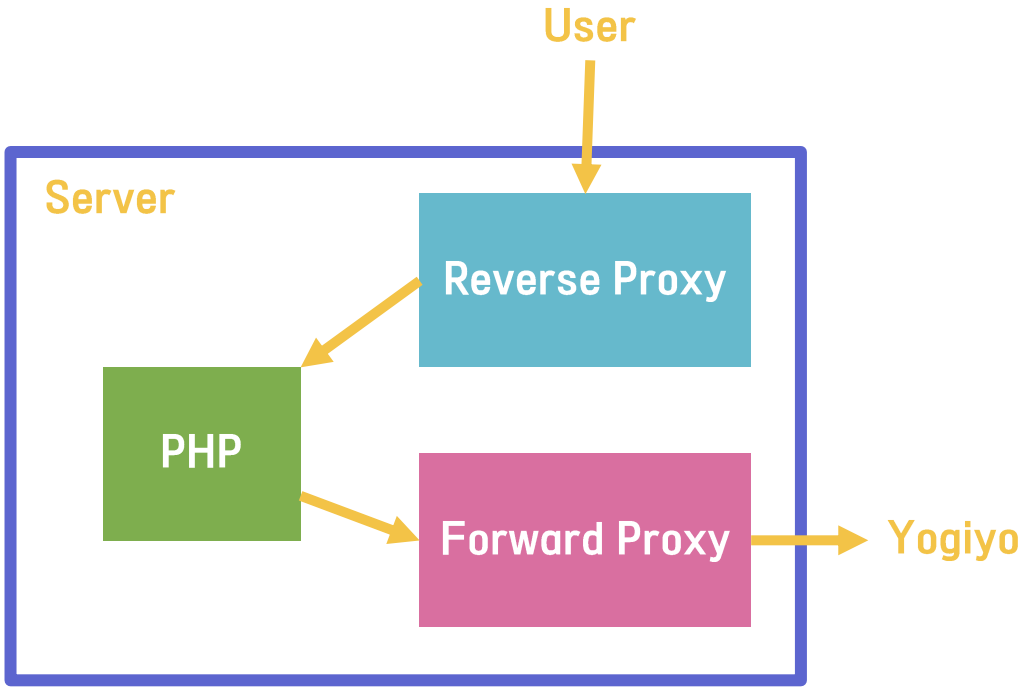
IDEA

Top Menu Bar

Top menu bar that does not move even when scrolling and is always **fixed** at the top of the screen

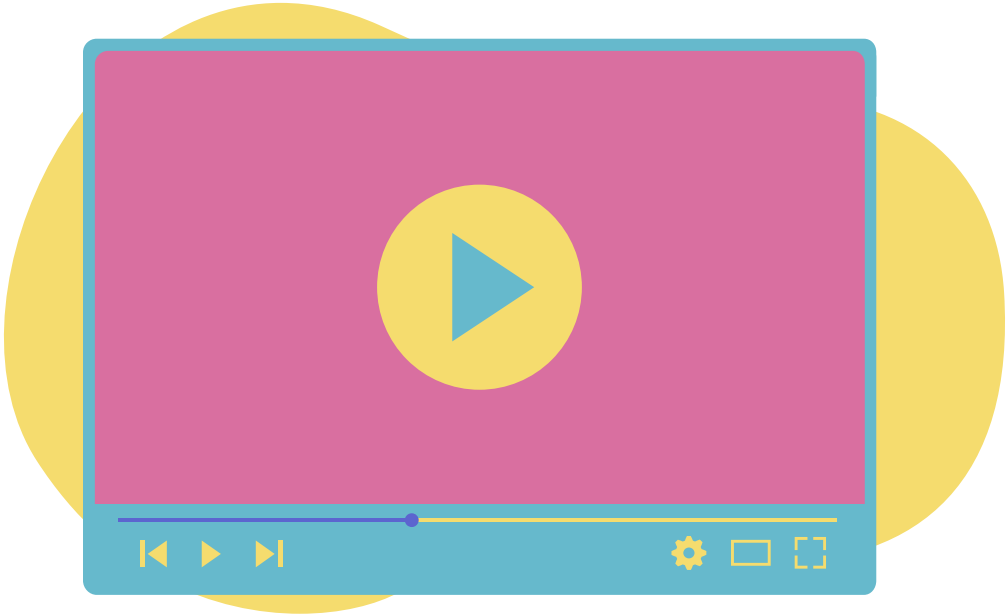
02 IDEA

Future Plan



Include Yogiyo webpage

Instead of directly entering the URL of the store, the Yogiyo homepage was included to the existing our homepage by building a Proxy server

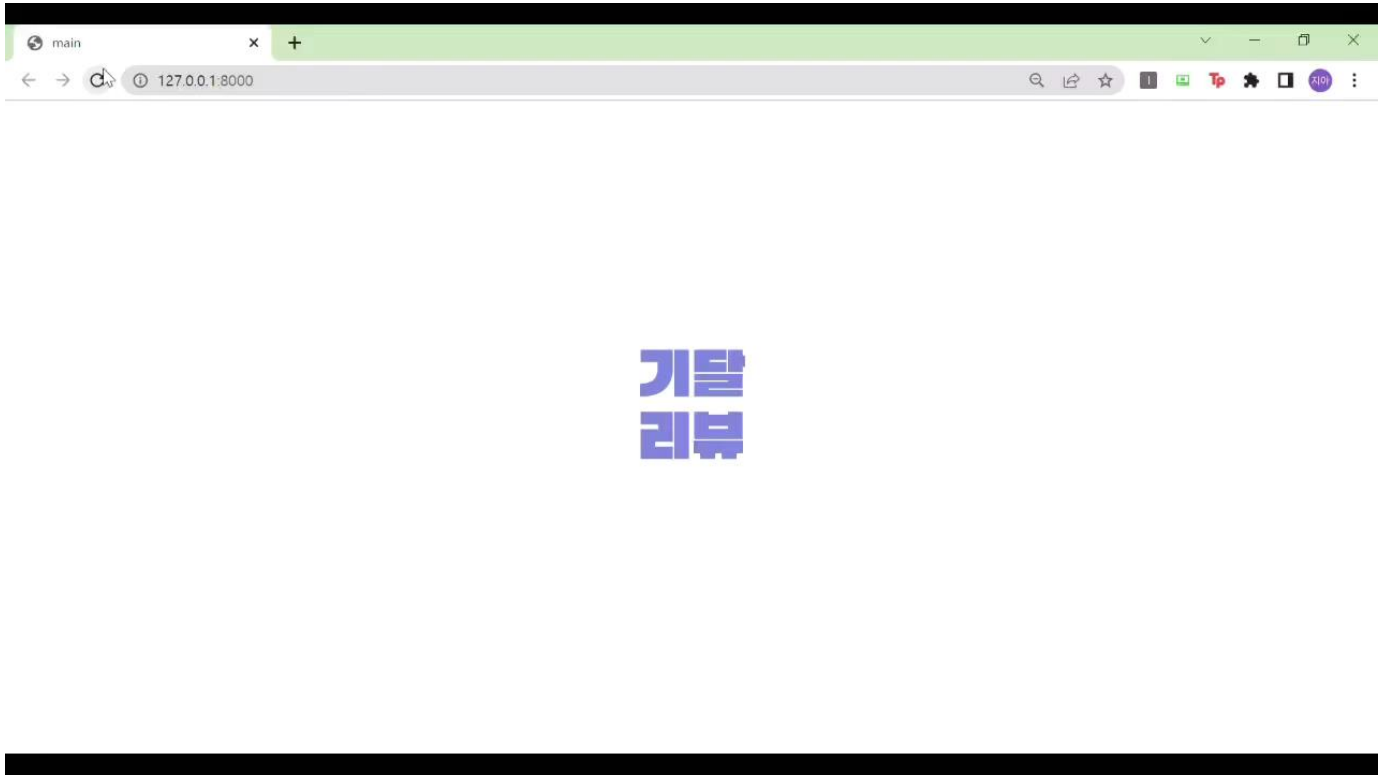


03
Demo Video

01 Watch Demo Video

03 Watch Demo Video

Demo Video



URL: https://www.youtube.com/watch?v=_8H-vUlWYPo

04
Marketing Plan

01 Market Analysis

02 Marketing Strategy

03 Marketing Method

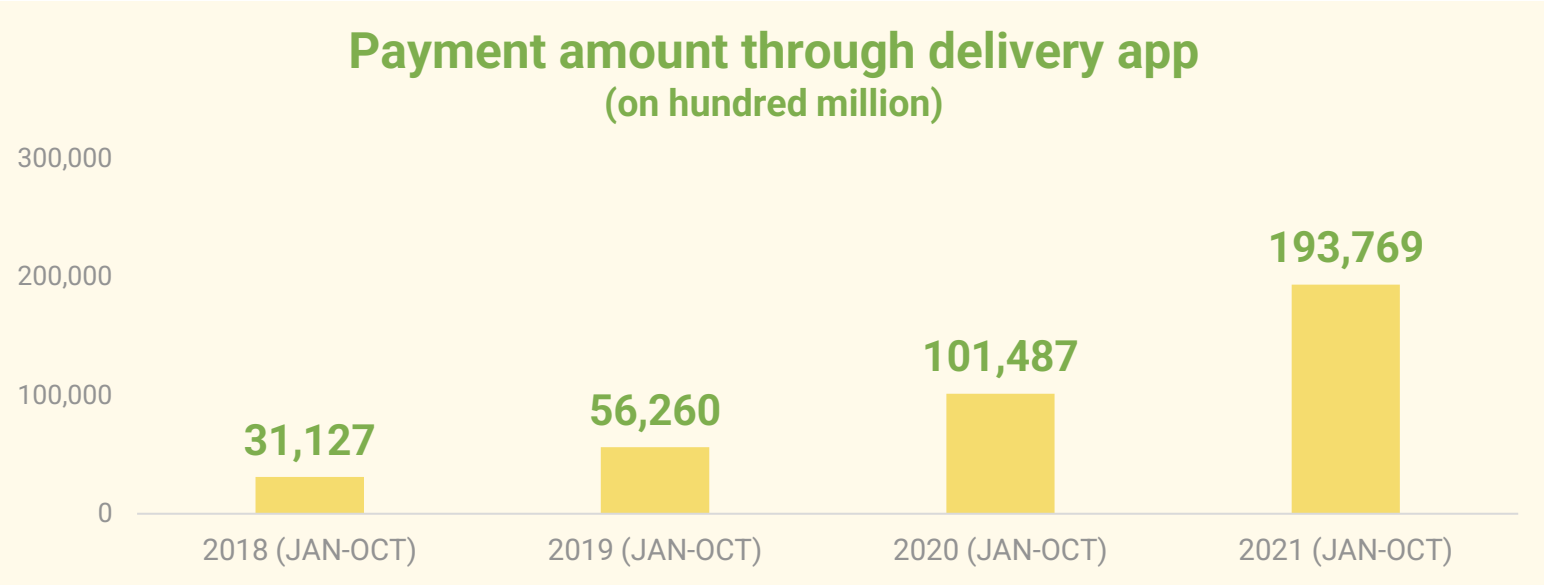
04 Marketing Plan

Market Analysis

90.9 %

$10T

As the delivery app market grows, our service target is delivery app users or delivery app companies.



04
Marketing
Plan

Marketing Strategy: SWOT

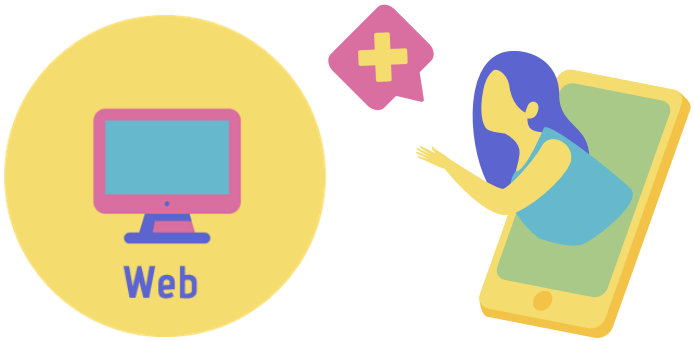
S

Strengths

Representative reviews by ratings

Keyword analysis

Filtering functions



Weakness

Time spent analyzing reviews

Currently available only on the Web

Inaccuracy of classification through AI

W

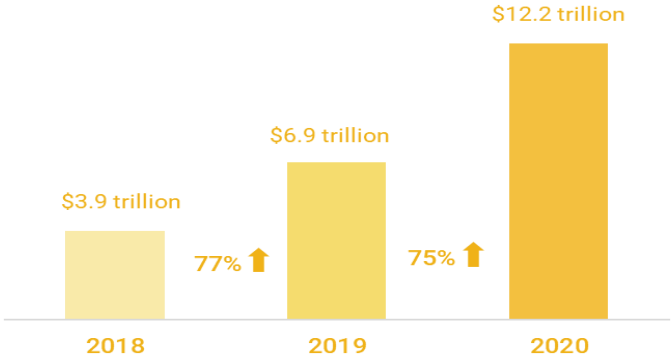
Marketing Strategy: SWOT

O

Opportunity

Reinforcement of consumption propensity to order through delivery

Expands Delivery transaction volume



Threat

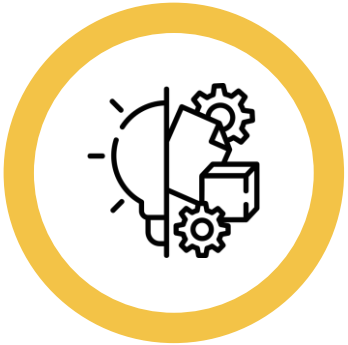
Active AI Review Analysis Research in Delivery App Market

Difficulty obtaining review data from application

T

04 Marketing Plan

Marketing Strategy: 4P



Product

- 기달리뷰 Project (Web)
- Provide store review analysis within Yogiyo app
  - Predictive score
  - Keyword analysis
  - AI review analysis

Place

- Directly partner with delivery apps
- Banner ads in delivery apps, Kakao Talk, etc

Price

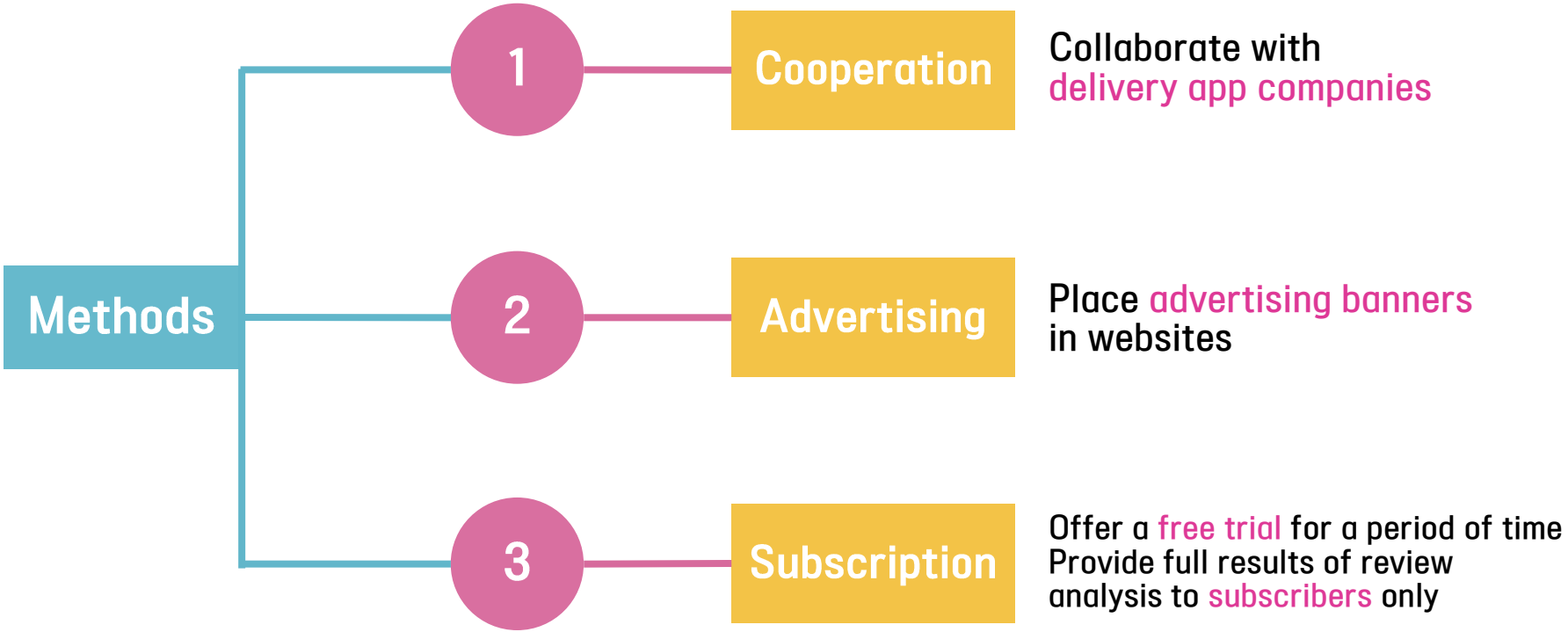
- Free Web Service
- Revenue generated by advertising

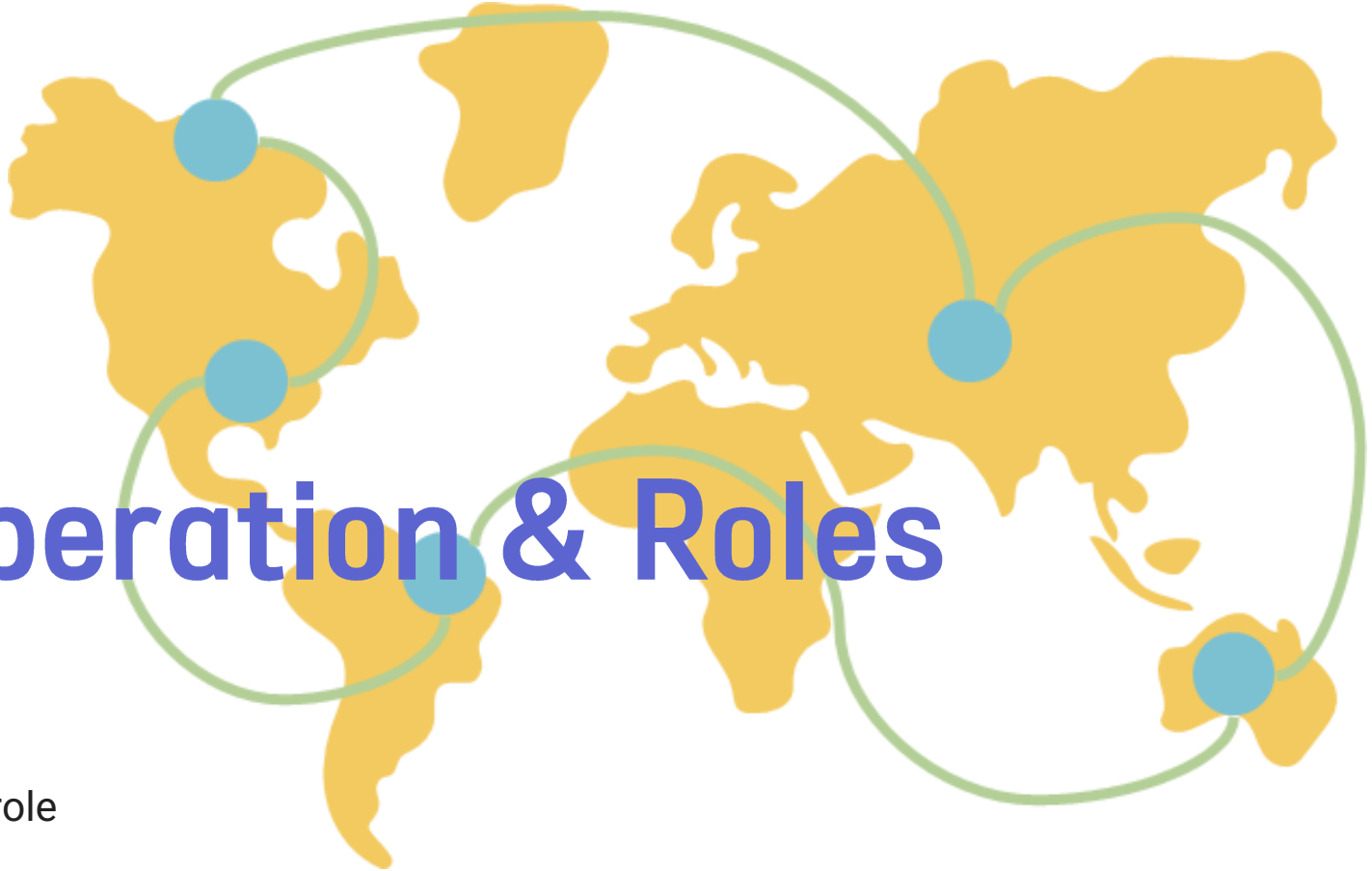
Promotion

- PPL via influencer
- Promote within the community
- Promote through a review event

04 Marketing Plan

Marketing Method





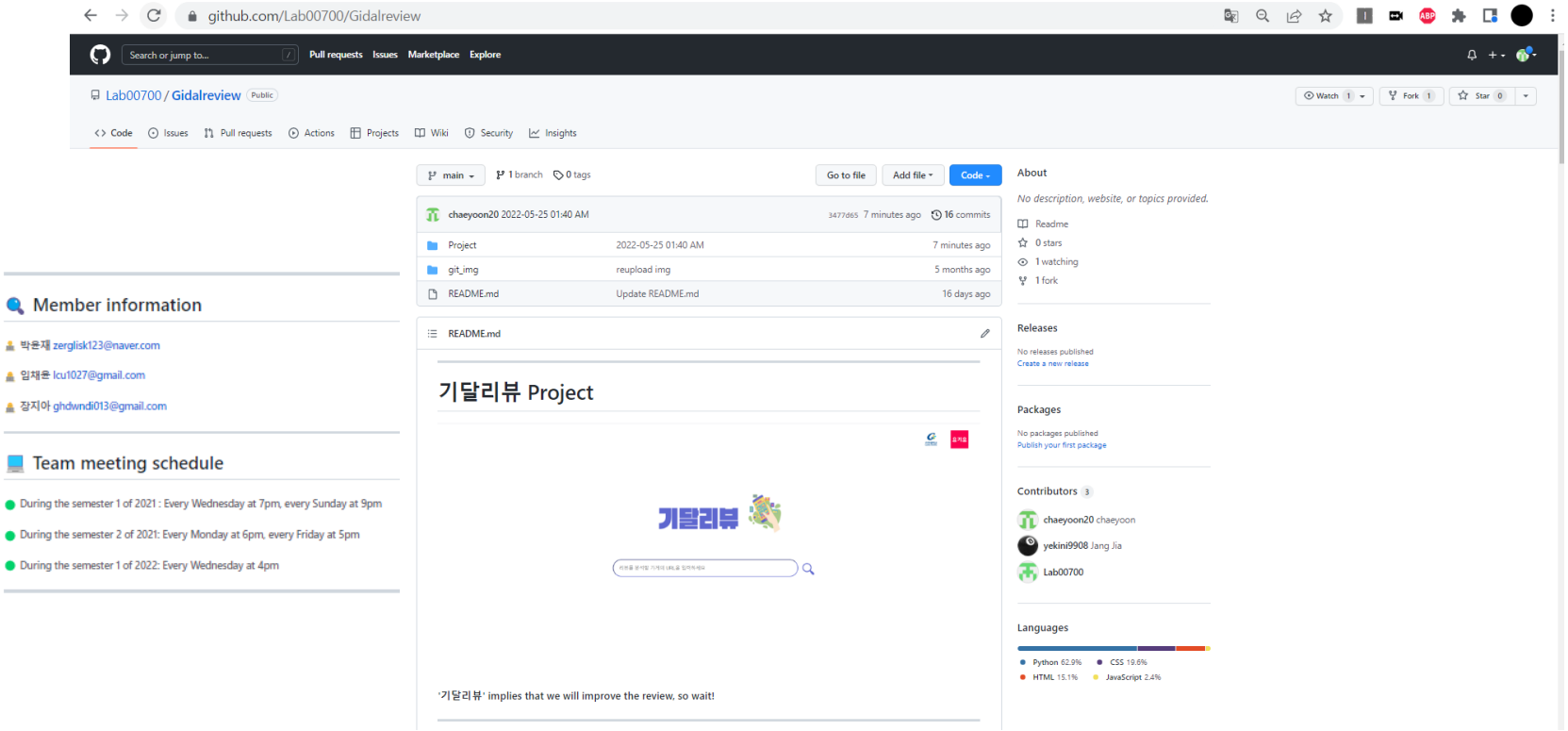
05
Cooperation & Roles

01 GitHub

02 Member role

05 Cooperation and Roles

GitHub



URL: https://github.com/Lab00700/Gidalreview

05
Cooperation and Roles

GitHub: Progress

Progress plan

2021 Semester 1

Week 3-4

🔸 Prior Studies
Comparison and analysis of delivery apps (배달의민족, 요기요, 쿠팡이츠, 배달통, etc)
Analysis Reviews

🔸 Review numerous reviews, select review criteria, reconfirm reviews based on criteria

- Low star rating without explanation
- Low star rating with positive reviews
- Low star rating for unfounded reasons
- Store slander, insult review
- Low star rating for food from other stores
- Use a review agency

🔸 Related papers, Model & Library
Research papers related to NLP(Natural Language Processing), emotional analysis, and review data
Selenium, BeautifulSoup, PyKoSpacing, KoNLPy, Tesseract-OCR, KorBERT, etc

Week 5-7

Inspect numerous reviews and select review criteria

- Yogiyo has 4 star rating criteria: taste, quantity, delivery, and recommended
- Thinking about points to become a special study, not just a review analysis

Week 9-11

Crawling small data for the Yogiyo website. Use Beautiful Soup 4.

Crawling Target Data

[About the store]

Name, total star rating (whether recommended, taste, quantity, delivery), review event notice (CEO notification)

[Review]

Star ratings for each review (recommended, taste, quantity, delivery), review content, review creation date, order menu

Week 12

Spacing preprocessing

- Use PyKoSpacing to preprocess spaces
- Converting non-spaced Korean sentences into spacing sentences

Spelling preprocessing

1. Pre-processing the entire sentence

- Analyze the grammar of sentences using KcBERT
- Correct grammar and typos throughout the sentence if there are errors in grammar

2. Calculate the frequency of keyword appearance and preprocess if it is less than a certain frequency

- Use KoNLPy to divide sentences by morpheme and calculate frequency by keyword
- Divide into upper and lower levels according to frequency and preprocess for lower keywords

Week 13-15

🔸 Implement for event participation review classification

1. Classify as text
   Categorize sentences by morpheme using KoNLPy
   Explore if keywords that match the review event item exist
   Categorize reviews that are suspected of being eventful, such as services, events, etc
2. Categorize as an image Use Tesseract to perform OCR processing that replaces letters in images with text
   Analyze replaced text to classify eventuality reviews

🔸 Re-evaluate event review star rating

1. High star rating, negative reviews Comparison of review sentiment analysis results using KorBERT
   Identify the negative and positive characteristics of the categorized words and reconstruct the star rating
2. High star rating, positive reviews Comparison of review sentiment analysis results using KorBERT
   Review Reconfiguration Verification Procedure

Summer vacation

We investigated the database construction and implemented a crawling automation system.

05
Cooperation and Roles

GitHub: Progress

2021 Semester 2

Week 1-2

Research the databases and servers you want to use and plan for a semester
Database candidates: Mongo DB, SQLite
Discuss how to build a server. The plan is as follows

1. Build your own server
2. Rent a server

Week 3-4

Data collection and classification criteria selection

- Collect data by categorizing them into categories that Yogiyo side categorizes
- 620-2 Gachon University, Bokjeong-dong, Sujeong-gu, Seongnam-si, Gyeonggi-do
- Selected as a restaurant for review events

Categorization of review data by category

- Categorize reviews into 8 categories
- Create folders, files for each category
- Repeat crawl by category type
- Separate data frames are divided into several tables and created as a single DB file

Week 5-6

Building a Server with Raspberry Pi

- The goal is to automate crawling on the server itself.
- Specify the crawling folder as the ftp server folder to allow users to receive files through the ftp server for easier retrieval of crawled data from the server
- Enable concurrent operations with RealVNC for smooth operation

Week 9-10

Improvement and maintenance of crawling automation system after studying web crawling and data preprocessing techniques

Week 11

Finalize data crawling on Yogiyo website and end data preprocessing.
Project ideas such as participation in review events and classification of non-participation reviews are starting to be realized in earnest.

Week 12

After determining the detailed function of the review analyzer, each member is responsible for implementing it.

1. Review Event Prediction Model - KoELECTRA (장지아)
2. Predict Star ratings - TF-IDF (박윤재)
3. Similar Review - Word2Vec, Cosine Similarity (임채윤)

Week 13

- Evaluate performance after model training.
- Implement review event participation sorting and filtering.
- Visualize functions according to their respective roles.

Week 14-15

The results of the review analysis through the model will be displayed on a web page using the long-range framework.
After implementing a web page through HTML and css languages, connect it to the model.

2022 Semester 1

We did documentation work to distribute to users, make marketing plans, and continue to supplement the system through feedback after actual distribution.

[Maintenance]

- Reviews are constantly updated, so we implemented adding words to dictionaries on a regular basis and automating learning. - Add Top menu bar and Paging function.

- Top menu bar that does not move even when scrolling and is always fixed at the top of the screen.
- Paging function to the AI Review section so that we could see 5 reviews on each page.

05
Cooperation and Roles

GitHub: Wiki

Home

chaeyoon edited this page on 28 Mar · 44 revisions

기달리뷰 Project

'기달리뷰' implies that we will improve the review, so wait!

Brief description

This is a web service for review analysis implemented with Django, a Python web framework.
When you enter the URL of the restaurant you want to analyze reviews, the review data is crawled in real time by the automated crawling system. The crawled data derives review analysis results through various libraries and trained deep learning-based models.

Contents

Overview

chaeyoon edited this page on 25 Mar · 6 revisions

Project ov

This project is

As the untact c
The number of
ordering has als
All delivery app
The opinion tha

In a delivery ap
cause problems

Therefore, we p
apps.

Objective

Based on the id
credible reviews

Expected

Due to the syst
terrorism exper
through credibl

Key features

chaeyoon edited this page on 25 Mar · 1 revision

Crawling automation system

The crawling au
restaurants, rev
according to pr

- yogiyo rev
- Total 5000

Crawling Target
[Store Informat

- Name
- Total Horo
- Review Eve

[Review]

- Star Rating
- Review Co
- Order men
- Date

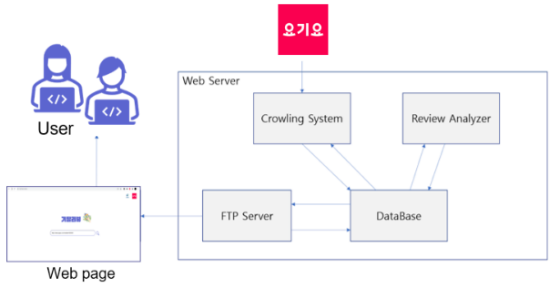
Review anal

The review ana
produced word
results through

- [Finding Si

System structure

chaeyoon edited this page on 25 Mar · 1 revision



It's the overall structure of the "기달리뷰" system.

The system consists of a crawling automation system, a web server, a database, a review analyzer, and a web page.

The web server is a Linux-based web server built with Raspberry Pi, and we have built an online public development environment for developers using Real VNC.

FTP servers built on web servers help you analyze reviews and access data to provide users.

The user enters the URL of the company that wants to analyze the review through the web browser.

The web page data of the URL is crawled by the crawling automation system, and the review is analyzed through a review analyzer, and the user can check it through the website produced.

05 Cooperation and Roles

GitHub: Wiki

Used library

chaeyoon edited this page 3 days ago · 4 revisions

⚒️ Crawling Automation System

Selenium

```
pip install selenium
```

https://www.selenium.dev/

BeautifulSoup

```
pip install beautifulsoup4
```

https://www.crummy.com/software/BeautifulSoup/bs4/doc/

⚒️ Data preprocessing

pykospacing

```
pip install git+https://github.com/haven-jeon/PyKoSpacing.git
```

https://github.com/haven-jeon/PyKoSpacing

How to Use

chaeyoon edited this page 2 minutes ago · 6 revisions

Access to a Domain Address

Click here to use 기달리뷰

🤦 http://gidalreview.kro.kr/

Make use of our project

Download our project

```
git clone https://github.com/Lab00700/Gidalreview.git
```

Crawling Automation System

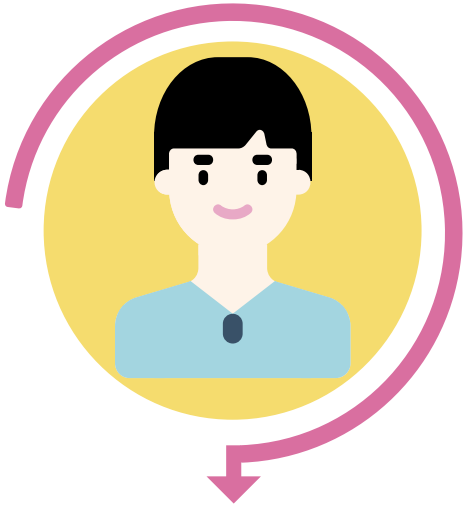
기달리뷰 automatically crawls the review of the URL entered by the user. The file written for crawling is as follows.

```
CrawlingAutomationSystem.py
```

Install the chrome driver for automated crawling.

05
Cooperation and Roles

Member role

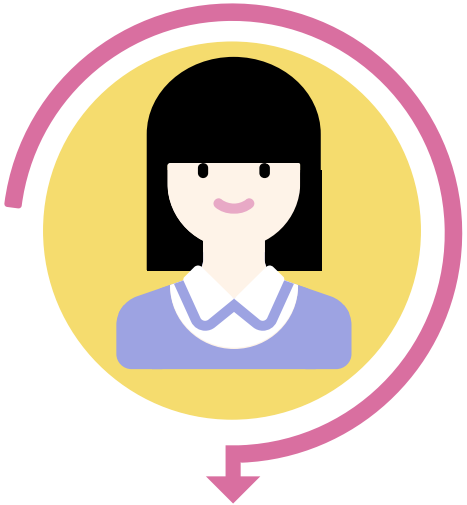


박윤재

- Implement Crawling Automation System
- Build server and databases
- Create a data dictionary
- Implement a Star Ratings Prediction Model

장지아

- Data classification and preprocessing
- Implement Review Event Prediction Model using KoEELECTRA
- Build a website through the Django framework



임채윤

- Implement a Representative Review and Similar Review Model
- Visualize star distribution, key keywords, word cloud
- Build a website through the Django framework

Thanks!

졸업작품3 4분반 1조

201835455 박윤재
201835514 장지아
201935120 임채윤