

# 주간 활동 보고서

## 목차

### 1. 활동시간 및 내용

- 스마트업 회의 및 활동 진행 시간
- 회의 및 활동 내용

### 2. 14주차 진행사항

- 14주차 진행사항

### 3. 추후 일정

- 앞으로 남은 기간 동안의 일정

## <주간 활동 보고서 14 주차> -2024.06.03(월) ~ 2024.06.09(일)-

### 1. 활동시간 및 내용

<스마트업 회의 및 활동 진행 시간>

-오프라인 회의 및 활동 시간

6 월 03 일(월) / 14 시 30 분~18 시 40 분 (4 시간 10 분) 오프라인

6 월 04 일(화) / 16 시~18 시 10 분 (2 시간 10 분) 오프라인

6 월 05 일(수) / 10 시~18 시 (8 시간) 오프라인

총 : 14 시간 20 분

>> [스마트업] 14 주차 회의 및 활동 내용

-물품 배송 수령

-라즈베리파이 터치스크린, 케이스 연결 및 조립  
터치스크린 작동 확인

-터치스크린 & 케이스 연결 후 작성한 최종 코드  
동작 확인

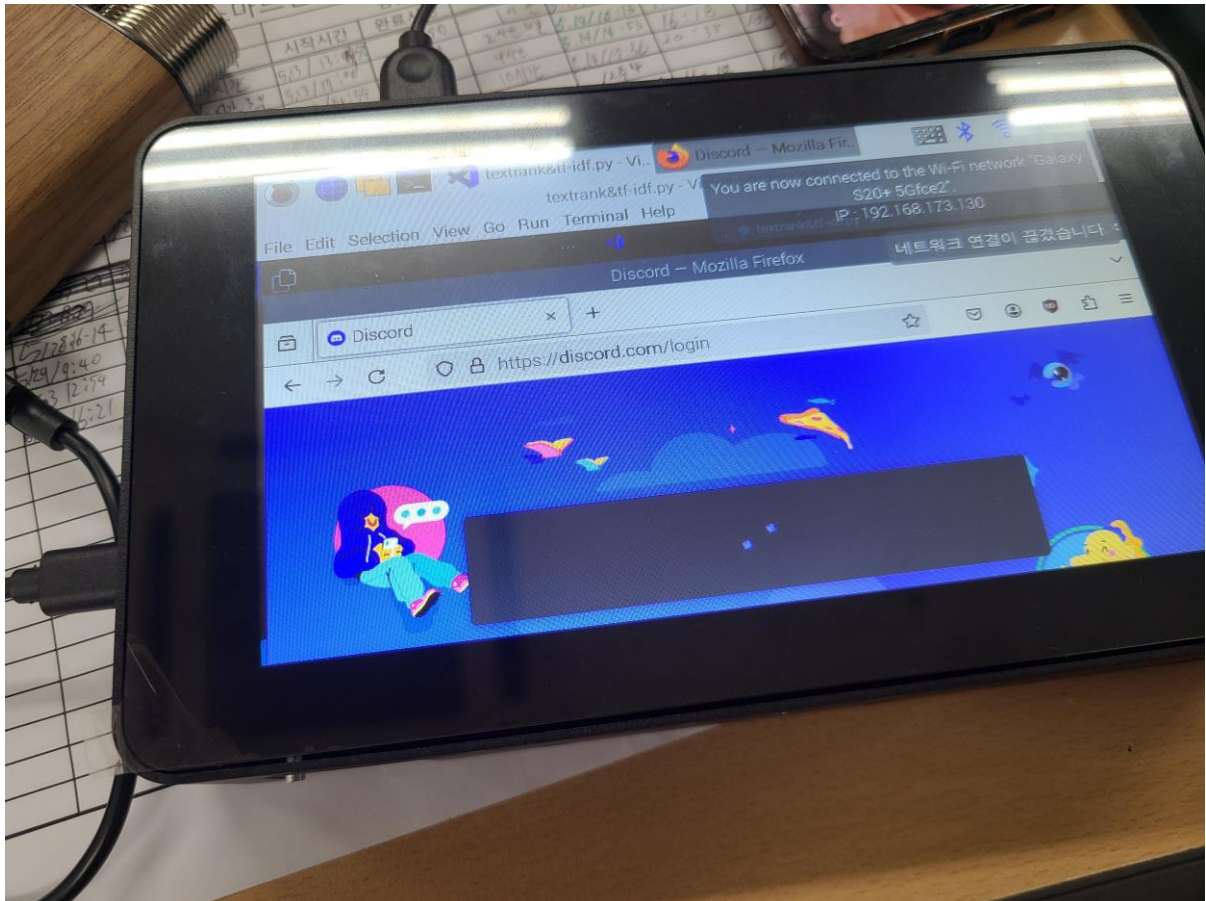
-3 개학과 (나노바이오,생명분자,컴퓨터공학)의  
학과 강의 30 분이상 강의를 요약알고리즘  
시스템을 활용해 STT 변환 및 요약을 진행 후,  
10 명의 평가단을 구성하여 평가 진행.

-STT & 요약알고리즘 결합 코드 보완

## 2. 14주차 진행사항

- 14 주차 진행사항에 대한 정리 내용

6월 4일(화)에 물품 수령 후 팀원들과 터치스크린 및 케이스 연결을 진행하고 작동을 확인 했습니다.



요약알고리즘 시스템을 이용해 3 개  
학과(나노바이오,생명분자,컴퓨터공학)3 개 학과의 30 분 이상 강의를  
STT 로 변환 및 요약하고  
나노바이오 : 5 명  
생명분자 : 2 명  
컴퓨터공학 :3 명  
총, 10 명의 평가단을 구성하여 요약의 정확성 , 요약된 내용의 분량  
에 대한 평가를 진행했습니다.

요약된 최종 파일을 보내주고 요약의 정확성 , 요약된 내용의 분량에 대해

1. 매우 좋음

2. 좋음

3. 보통

4. 나쁨

5. 매우 나쁨

5 단계로 나누어 평가 결과를 취합했습니다.

<나노바이오>

|   | 요약 정확성 | 요약 분량 |
|---|--------|-------|
| 1 | 3      | 2     |
| 2 | 3      | 1     |
| 3 | 2      | 2     |
| 4 | 4      | 3     |
| 5 | 3      | 2     |

<생명분자>

|   | 요약 정확성 | 요약 분량 |
|---|--------|-------|
| 1 | 2      | 1     |
| 2 | 2      | 2     |

<컴퓨터공학>

|   | 요약 정확성 | 요약 분량 |
|---|--------|-------|
| 1 | 1      | 2     |
| 2 | 2      | 3     |
| 3 | 1      | 1     |

강의 요약된 파일들은 깃허브에 따로 업로드 하겠습니다.

저희가 결합한 요약 알고리즘이 원래 있던 텍스트 랭크에 비해서 강의 요약에 더 나은 성능을 가지고 있는가에 대해서 테스트가 필요하다고 생각하여 해당 테스트를 진행 했습니다.

저희는 강의를 요약 하는것이 목표이기 때문에 단순히 짧게 요약하는 것보다 핵심 내용을 최대한 많이 담는것을 목표로 했습니다.

우선 같은 내용의 글을 결합 알고리즘 , 텍스트랭크 각각에 넣어서 요약본을 도출해내고

해당 요약본을 ROUGE (Recall-Oriented Understudy for Gisting Evaluation) 라는 텍스트 요약 품질을 평가하는 지표 중 한곳에 넣고 그 결과를 비교해보았습니다.

우선 결합 알고리즘의 요약본입니다.

```
120
121
122 from tkinter import *
123 from tkinter.ttk import *
124
125 text = "인류 최초로 달 뒷면 샘플 채취를 목표로 발사된 중국의 달 탐사선 '창어(嫦娥·달의 여신 항아) 6호'가 2일 달 뒷면 착륙에 성공했다고 중국
126 textrank = TextRank(text)
127 for row in textrank.summarize(2): # 몇 줄로 요약할꺼야
128     print(row)
129     print('\n')
130     print('\n')
131 print(textrank.keywords())
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

PS C:\Python\textranktfidf> & C:/Users/hwanghae/AppData/Local/Microsoft/WindowsApps/PythonSoftwareFoundation.Python.3.11\_qbz5n2kfra8p0/python.exe c:/Python/textranktfidf/Untitled-1.py

인류 최초로 달 뒷면 샘플 채취를 목표로 발사된 중국의 달 탐사선 '창 어( 嫦娥· 달의 여신 항아) 6호' 가 2일 달 뒷면 착륙에 성공했다고

신화통신은 중국 국가 항 천국( 국가 우주국)을 인용, 창 어 6호가 이날 새벽 목표 지점인 달 뒷면 ' 남 극· 에이킨 분지 '에 착륙했다며 ' 인류 탐사선이 처음으로 달 뒷면에서 샘플 채취 임무, 즉 ' 두꺼비 궁전( 蟾宮· 달에 두꺼비가 산다는 전설에서 나온 말로 달을 가리킴 )에서 보물 찾기 '를 시작할 것' 이라고 전했다.

['착륙', '샘플', '채취', '이벤', '중국', '우주', '탐사선', '성공', '목표', '국가']

그리고 텍스트 랭크 요약본입니다.

```
26
27 # 테스트를 위한 입력 한글 텍스트
28 text = "인류 최초로 달 뒷면 샘플 채취를 목표로 발사된 중국의 달 탐사선 '창어(嫦娥·달의 여신 항아) 6호'가 2일 달 뒷면 착륙에 성공했다고 중국 신화통
29
30 summary = textrank(text)
31 print(summary)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

PS C:\Python\textranktfidf> & C:/Users/hwanghae/AppData/Local/Microsoft/WindowsApps/PythonSoftwareFoundation.Python.3.11\_qbz5n2kfra8p0/python.exe c:/Python/textranktfidf/only\_textrank.py  
인류 최초로 달 뒷면 샘플 채취를 목표로 발사된 중국의 달 탐사선 '창어(嫦娥·달의 여신 항아) 6호'가 2일 달 뒷면 착륙에 성공했다고 중국 신화통신이 보도했다. 신화통신은 중국 국가항천국(국가우주국)을 인용, 창어 6호가 이날 새벽 목표 지점인 달 뒷면 '남극-에이킨 분지'에 착륙했다며 "인류 탐사선이 처음으로 달 뒷면에서 샘플 채취 임무, 즉 '두꺼비 궁전(蟾宮·달에 두꺼비가 산다는 전설에서 나온 말로 달을 가리킴)에서 보물찾기'를 시작할 것"이라고 전했다

각각의 요약본을 로그 점수로 평가했습니다. 저희가 사용한 로그 점수는 ROUGE-1, ROUGE-2, ROUGE-L, 그리고 ROUGE-Lsum으로 이루어져 있고 각 점수는 정밀도(Precision), 재현율(Recall), 그리고 F1 점수로 구성됩니다.

## 1.ROUGE-1

Precision (정밀도): 예측된 요약에서 참조 요약과 일치하는 1-그램(단어)의 비율을 의미합니다.

Recall (재현율): 참조 요약에서 예측된 요약과 일치하는 1-그램(단어)의 비율을 의미합니다.

F1 Score: 정밀도와 재현율의 조화 평균으로, 두 값의 균형을 나타냅니다.

## 2.ROUGE-2

Precision (정밀도): 예측된 요약에서 참조 요약과 일치하는 2-그램(연속된 두 단어)의 비율을 의미합니다.

Recall (재현율): 참조 요약에서 예측된 요약과 일치하는 2-그램의 비율을 의미합니다.

F1 Score: 정밀도와 재현율의 조화 평균입니다.

### 3.ROUGE-L

Precision (정밀도): 예측된 요약에서 참조 요약과 일치하는 가장 긴 공통 부분 문자열(Longest Common Subsequence, LCS)의 비율을 의미합니다.

Recall (재현율): 참조 요약에서 예측된 요약과 일치하는 가장 긴 공통 부분 문자열의 비율을 의미합니다.

F1 Score: 정밀도와 재현율의 조화 평균입니다.

### 4.ROUGE-Ls (ROUGE-Lsum)

Precision (정밀도): 문서 요약의 문장별로 가장 긴 공통 부분 문자열의 정밀도를 측정합니다.

Recall (재현율): 문서 요약의 문장별로 가장 긴 공통 부분 문자열의 재현율을 측정합니다.

F1 Score: 정밀도와 재현율의 조화 평균입니다.

해당 기준으로 평가 되고 먼저 결합 요약의 점수입니다.





요약 정도를 따지는 정밀도는 비슷하지만 포괄적으로 전체적인 내용을 잘 포함하고 있는지를 따지는 재현율이 결합 알고리즘이 더 큰폭으로 수치가 높아 강의 내용에서 중요한 내용을 잘 빼먹지 않고 요약 할것으로 기대됩니다.

### 3. 추후 일정

– 앞으로 남은 기간 동안의 일정입니다.

[15 주차]

➤ 6월14일(금) 캡스톤디자인 최종 발표