

Department of Software Convergence

Data Analysis Capstone Design Project   
Weekly Progress Report

| Project Title | 딥러닝을 활용한 노래 가사 분석 및 추천 시스템 |
| --- | --- |
| Student / ID | 김은비 / 2019100858  유창현 / 2018102121  이은경 / 2019100896  이인석 / 2018110659 |
| Reporting Week | 2022-04-08 ~ 2022-04-14 |
| Project Manager of Reporting Week | 김은비 |
| Faculty Supervisor | 유창현 |

# **Tasks Outlined in Previous Weekly Progress Report** (Provide detailed information on the tasks to be completed in this week)

* Word2Vec을 활용한 7개 감정별 감성사전 구축

# **Progress Made in Reporting Week** (Provide detailed information on the progress that you made in the reporting week. Limit your write-up to no more than two page)

> 세부감정 도출에 참고한 두가지 모델

* Russell(1980) 모델(4가지)

1. 신이 난, 흥분한, 놀라운, 아주 기뻐하는, 행복한
2. 기쁜, 만족스러운, 흡족한, 고요한, 차분한, 여유있는
3. 불안해 하는, 두려워 하는, 긴장한, 화난, 괴로워 하는, 짜증이 난
4. 비참한, 우울한, 슬픈, 따분한, 의기소침한

* Tellegen-Watson Clark 모델(8가지)

1. 놀란
2. 흥분된, 즐거운
3. 행복, 유쾌한
4. 편안한,평온한
5. 무관심한, 노곤한
6. 권태로운, 피곤한
7. 실증난, 슬픔
8. 두려운, 초조한

⇒ 7가지 감정 도출 및 유사도 측정 기준 키워드 도출

1. 놀란 -> 놀라(다)
2. 두려움(무서운) -> 두렵(다), 무섭(다)
3. 행복(즐거운, 설렘) -> 행복, 즐기(다), 설레(다)
4. 슬픔(우울한, 괴로운, 비참한) -> 슬픔, 슬프(다), 우울
5. 분노(미워하는, 짜증이 난) -> 미움, 미워하(다)
6. 무관심(실증난, 권태로운, 피곤한) -> 무관심, 질리(다), 피곤
7. 편안함(여유있는) -> 편안, 여유

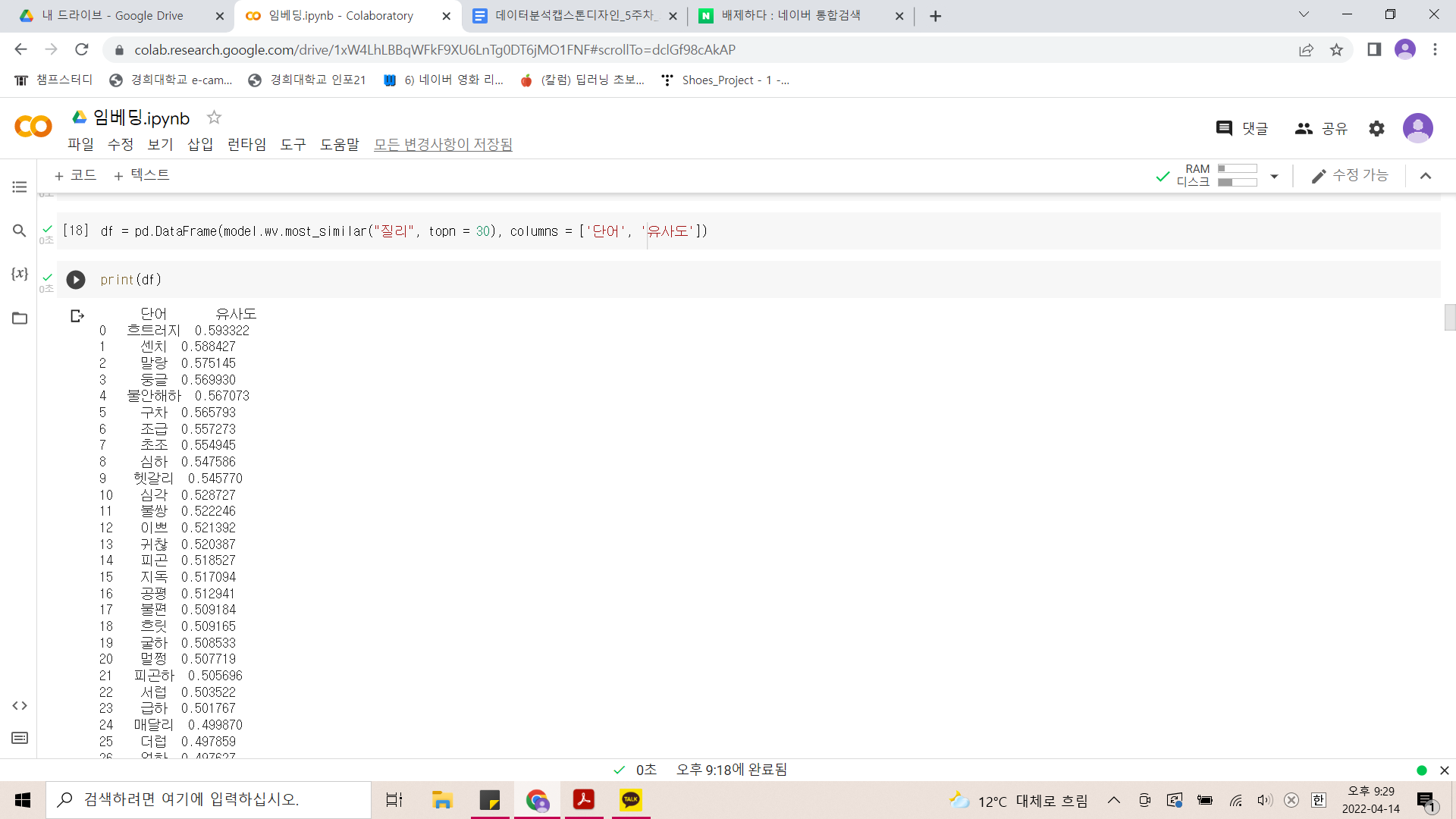
* Word2Vec 모델 파라미터 결정

1. CBOW vs. skip-gram -> CBOW(문맥(주변단어)를 통해 중심단어를 예측)
2. size = 100
3. window = 5
4. iter = 10
5. min\_count = 3
6. sg = 0

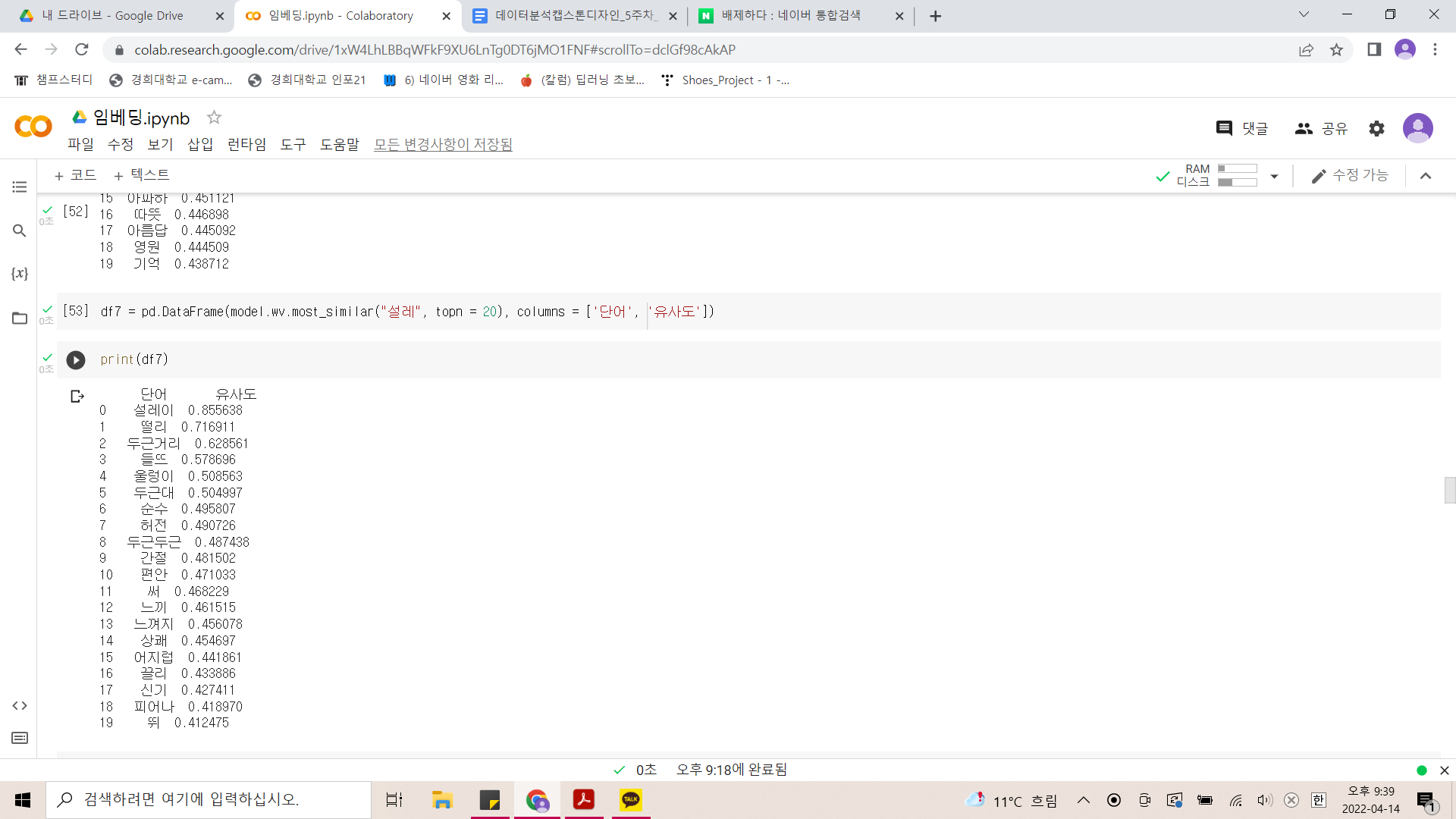
각 감정별로 결정된 키워드를 기준으로 유사도가 가장 높은 단어들부터 감성사전에 반영하기로 결정하였다. 0.5를 임계값으로 하여 0.5이상의 유사도를 갖는 단어만 반영하고, 2개 이상의 감정에 중복해서 등장하는 단어는 아예 배제하였다.



[Word2Vec 모델 학습]



[무관심 감정에 키워드인 ‘질리(다)’와 유사도가 가장 높은 단어]



[행복 감정에 키워드인 ‘설레’’와 유사도가 가장 높은 단어]

# **Difficulties Encountered in Reporting Week** (Provide detailed information on the difficulties and issues that you encountered in the reporting week. Limit your write-up to no more than one page)

* 각 감정을 대표하는 키워드를 무엇으로 설정할지에 대한 고민

우리가 결정한 7가지의 감정이 여러 감정을 대표하는 한 감정을 선정한 것이기 때문에 만약 행복이라는 감정이라면 그 안에 설렘, 즐거움 등이 포함된다고 할 수 있다. 그런데 ‘행복’이라는 키워드로 유사도를 계산했을 때 이런 여러 감정들을 포함할 수 있는 단어들이 도출될 수 있을지 고민이 되었다. 그래서 직접 학습된 word2vec모델을 통해 여러 키워드를 통해 도출된 단어들을 확인하고 가장 대표성이 높다고 생각되는 키워드를 선정하여 구축하기로 결정했다.

* word2vec 모델의 파라미터 결정

word2vec 모델의 파라미터를 결정하기 위해서 반복횟수, min\_count, window 등을 여러 번 바꾸어 시도해보면서 도출된 단어들이 어느 정도 각 감정에 대입되는 단어들로 구성되어 있는지 확인해 보았고, 가장 적합하다고 생각되는 파라미터 값을 팀원들끼리 논의하여 최종적으로 결정하였다.

# **Tasks to Be Completed in Next Week** (Outline the tasks to be completed in the following week)

* 구축한 감정사전을 바탕으로 라벨링 작업 진행
* 감정 분류 모델 결정 및 학습 시작