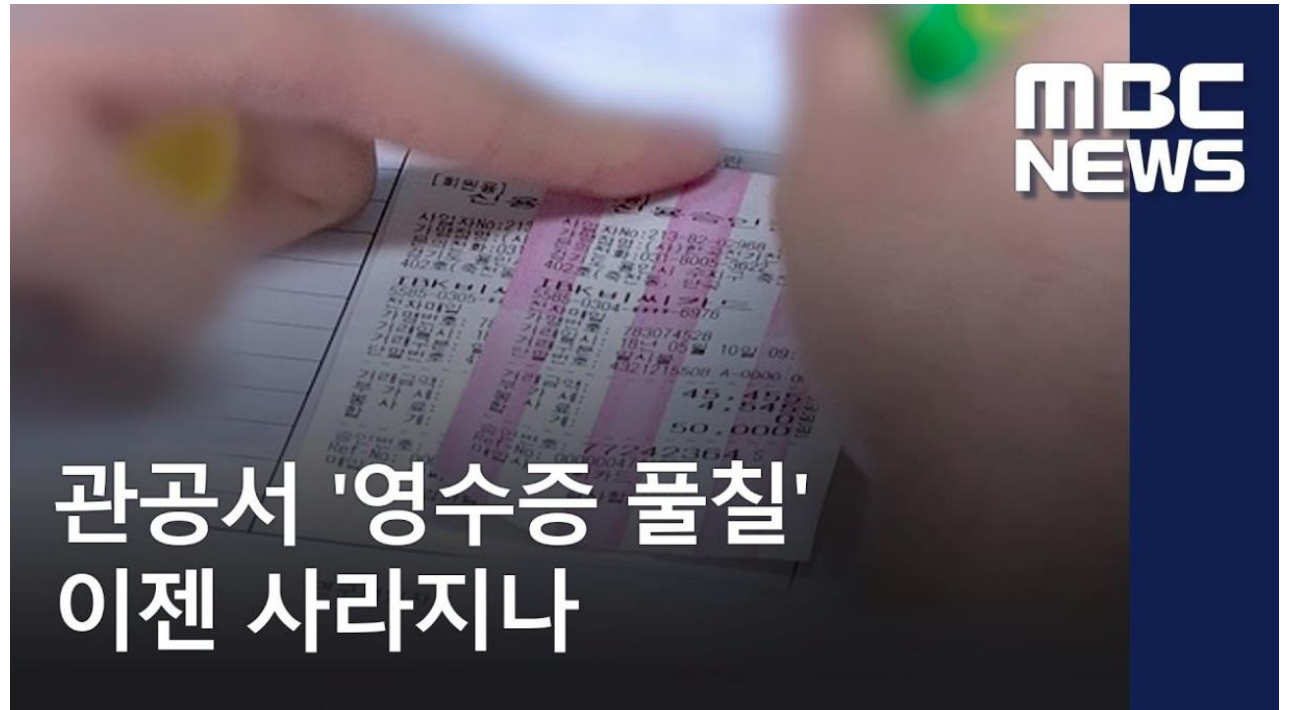


Word2vec을 활용한 야매추천시스템

싸이그램즈 김성근

심리학..그 머나먼 여정의 끝..

- 심리치료 공부
- 심리검사 개발
- 심리검사 영업
- 개인사업
- 대학원(MIS)
- 일반행정 회사원



풀칠만 하는 인생을 피하기 위해 스테디를 간간히 하고 있음

연구자 추천시스템의 시작

- 서로 어떤 연구를 하는지 몰라!
- 데이터가 있는데 분석을 못해!
- 분석은 하는데 데이터가 없어!
- 서로 가진 자원을 공유하고 함께 협업할 수 있음 좋겠다!



빅데이터 관련 연구자 세미나

연구자 추천시스템의 시작

- 특정 분야 연구를 하는 교수는 찾을 수 있을 것 같은데!!

아주 심플한 UX

 slack

&

아주 편리한 패키지

 python

Gensim Word2vec

연구자추천 시스템을 만들자!

연구자 추천시스템

- 관련성이 높은 무엇인가를 전달하는 것!
 - 나랑 비슷한 사용자들이 보는 콘텐츠를 추천
 - 나랑 비슷한 사용자들이 사는 물건을 추천

Congratulations! Movies we think **You** will ❤️

Add movies to your Queue, or **Rate** ones you've seen for even better suggestions.



연구자 추천시스템

- 연구자 추천시스템
 - 특정 키워드와 관련성이 높은 연구자를 추천
 - 특정 교수와 관련성이 높은 연구자를 추천

우리학교에 풀칠 잘하는 사람 누가 있지?

풀칠

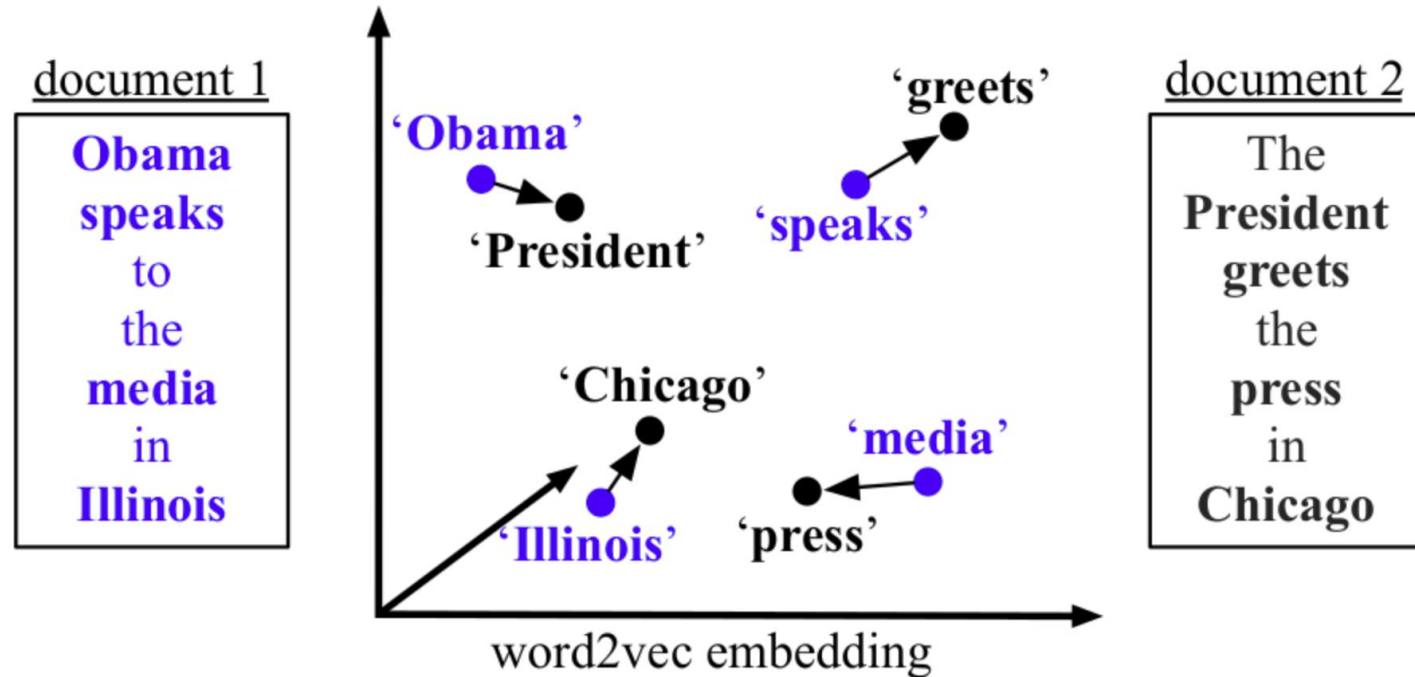


연구정보처
김성근
010-8956-495*

연구자 추천시스템

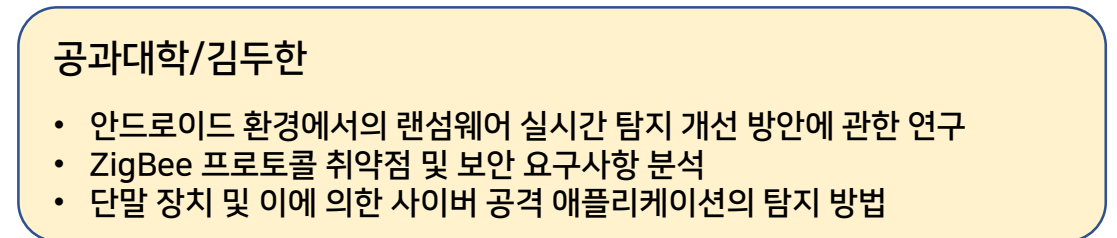
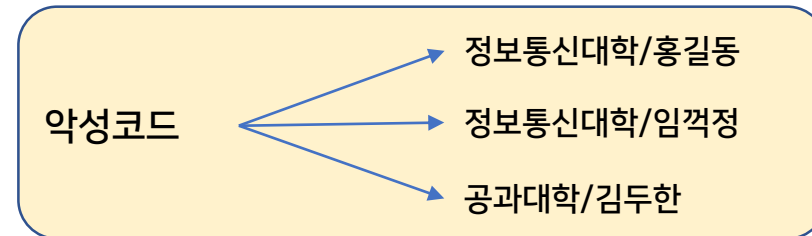
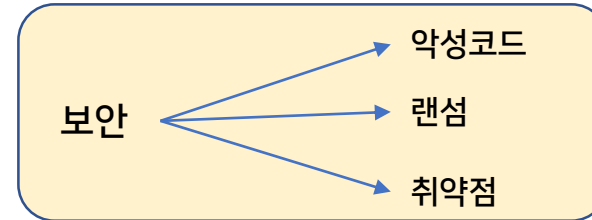
- Word2vec

- 인공지능망을 활용해 특정 단어 주변에서 자주 출현하는 단어들 끼리 서로 비슷한 수치를 갖게함



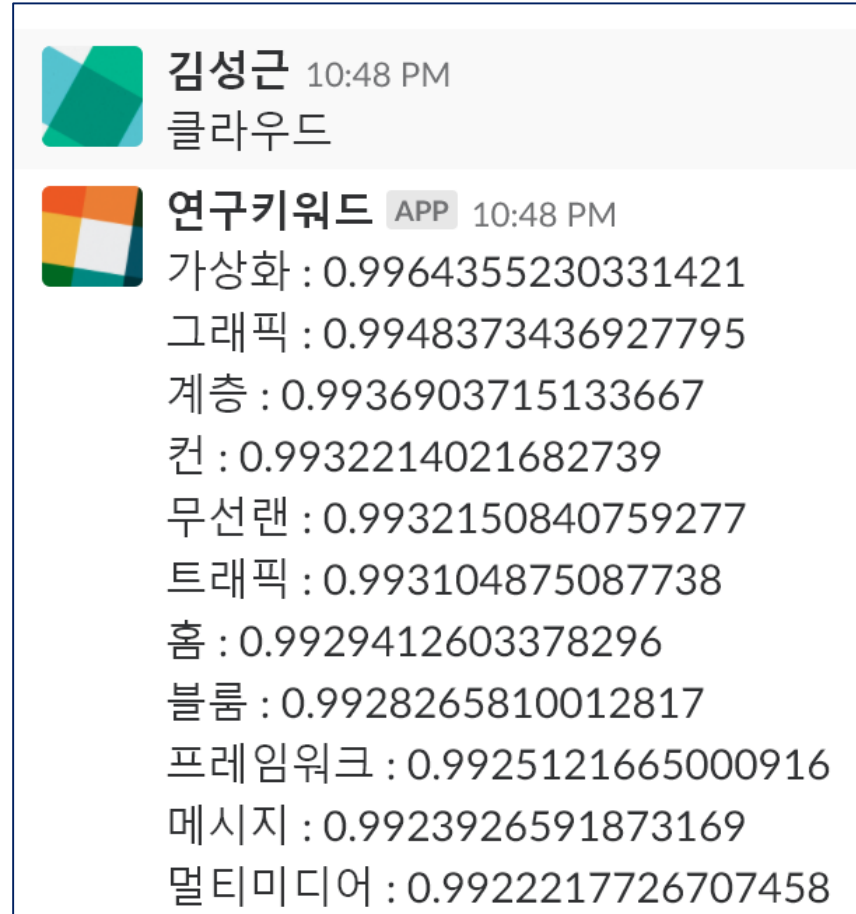
연구자추천 시스템 - 기획

- 1. 특정 분야의 단어를 넣으면
 - 관련된 연구 키워드를 알려주고
- 2. 연구 키워드를 입력하면
 - 관련된 연구자들을 알려주고
- 3. 연구자 이름을 입력하면
 - 그 연구자의 연구목록을 보여주자



연구자추천 시스템 - 흐름

- 1. 특정 분야의 단어를 넣으면
 - 관련된 연구 키워드를 알려주고
- 2. 연구 키워드를 입력하면
 - 관련된 연구자들을 알려주고
- 3. 연구자 이름을 입력하면
 - 그 연구자의 연구목록을 보여주자



김성근 10:48 PM
클라우드

연구키워드 APP 10:48 PM
가상화 : 0.9964355230331421
그래픽 : 0.9948373436927795
계층 : 0.9936903715133667
컨 : 0.9932214021682739
무선랜 : 0.9932150840759277
트래픽 : 0.993104875087738
홈 : 0.9929412603378296
블룸 : 0.9928265810012817
프레임워크 : 0.9925121665000916
메시지 : 0.9923926591873169
멀티미디어 : 0.9922217726707458

연구자추천 시스템 - 흐름

- 1. 특정 분야의 단어를 넣으면
 - 관련된 연구 키워드를 알려주고
- 2. 연구 키워드를 입력하면
 - 관련된 연구자들을 알려주고
- 3. 연구자 이름을 입력하면
 - 그 연구자의 연구목록을 보여주자



김성근 6:53 PM

컴퓨팅



연구자탐색기 APP 6:53 PM

정보통신대학/ 김성근

정보통신대학/ 김성근

정보통신대학/ 김성근

정보통신대학/ 김성근

정보통신대학/ 김성근

정보통신대학/ 김성근

정보통신대학/ 김성근

정보통신대학/ 김성근

연구자추천 시스템 - 흐름

- 1. 특정 분야의 단어를 넣으면
 - 관련된 연구 키워드를 알려주고
- 2. 연구 키워드를 입력하면
 - 관련된 연구자들을 알려주고
- 3.1 연구자 이름을 입력하면
 - 그 연구자의 연구목록을 보여주자



김성근 5:47 PM

경영대학/강주영



연구이력 APP 5:47 PM

2012:Issues of Applying Lightweight Semantic Web Ser
2012:엔터프라이즈 컴퓨팅 도입의 우선순위에 관한 연
2013:유료 모바일 애플리케이션 수용 저항 요인에 관한
2013:공간정보사업의 중복사업 검색을 위한 의미기반
2013:세부사례의 공유 및 교환을 위한 시맨틱 사례기반
2013:Context-Aware Services Framework Based on Se
Integration of Context
2013:A Conceptual Framework of Lightweight Semanti
2014:정보시스템 성공모형을 기반으로 한 모바일월렛
2014:Factors Influencing the Adoption of Enterprise CI
2014:Factors influencing electronic commerce adoption
2014:Development of Demand Forecasting Model for M

연구자추천 시스템 - 흐름

- 3.2 연구자 이름을 입력하면
 - 그 연구자와 관련성이 높은 연구 키워드를 주자



김성근 8:05 PM

정보통신대학/오성근

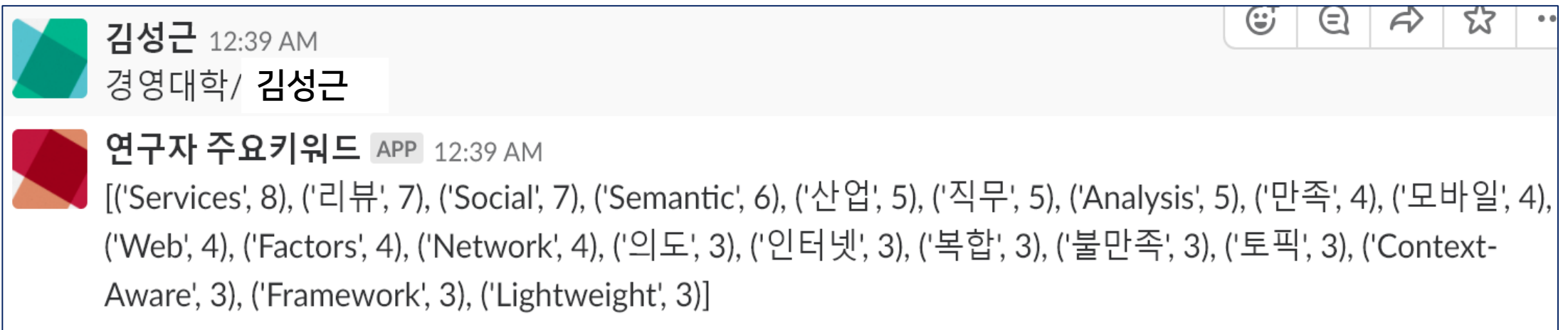


연구자 키워드 탐색 APP 8:05 PM

메모리,폰,가드,로봇,복구,셀룰러,랜덤,성하,이동,라우터,추적,단말기
기,플래시,coordinates,송수,스팅,레이더,분산,다중,크레인,프레임울

연구자추천 시스템 - 흐름

- 3.3 연구자 이름을 입력하면
 - 최근 5년간 그 연구자의 연구 제목에 가장 많이 출현한 단어를 주자;;



The screenshot shows a chat interface with two messages. The first message is from '김성근' (Kim Seong-geun) at 12:39 AM, with a profile picture of a green and blue geometric shape. The second message is from '연구자 주요키워드 APP' (Researcher Keyword APP) at 12:39 AM, with a profile picture of a red and orange geometric shape. The message content is a list of keywords with their frequencies in parentheses.

김성근 12:39 AM
경영대학/ 김성근

연구자 주요키워드 APP 12:39 AM
[('Services', 8), ('리뷰', 7), ('Social', 7), ('Semantic', 6), ('산업', 5), ('직무', 5), ('Analysis', 5), ('만족', 4), ('모바일', 4), ('Web', 4), ('Factors', 4), ('Network', 4), ('의도', 3), ('인터넷', 3), ('복합', 3), ('불만족', 3), ('토픽', 3), ('Context-Aware', 3), ('Framework', 3), ('Lightweight', 3)]

연구자추천 시스템 - 흐름

- 4. 연구자 이름을 입력하면
 - 그 연구자의 연락처를 줘서 협업을 할 수 있게!! 마무리!



김성근 11:00 PM

김성근



교원연락처 APP 11:00 PM

김성근 - 경영대학 - e-비즈니스학과 - 교수 - 2910 - 010- 0000-0000 jykang@ajou.ac.kr

연구자추천 시스템 - 데이터

- A대학 10년치 논문데이터

```
In [1]: import pandas as pd  
raw_data = pd.read_excel('연구실적_180602.xlsx', sheet_name='Sheet1')
```

```
In [2]: raw_data.head(2)
```

Out[2]:

	depatment_2	name	title
0	정보통신대학	김성근	연구윤리와 한국의 나노과학기술
1	정보통신대학	김성근	자동요금징수시스템(ETCS) 표준화 연구(주파수 방식을 중심으로)

연구자추천 시스템 - 데이터

- 문제
 - 연구자를 식별할 고유 값이 없음
- 해결
 - 소속+이름으로 연구자를 식별할 수 있는 고유 값을 만듦

Out[11]:

	name_2	title
0	정보통신대학/ 김성근	연구윤리와 한국의 나노과학기술

연구자추천 시스템 - 데이터 전처리

- 문제

- 영어, 한글이 혼재되어 있어서, 자연어 처리를 어찌할까..난감

	name_2	title
0	정보통신대학/ 김성근	연구윤리와 한국의 나노과학기술
1	정보통신대학/ 김성근	자동요금징수시스템(ETCS) 표준화 연구(주파수 방식을 중심으로)
5	공과대학/ 김성근	Derivation of Gaussian Distribution in Physics...
6	공과대학/ 김성근	Special Relativity Says That Energy is a Vect...

연구자추천 시스템 - 데이터 전처리

- 문제

- 영어, 한글이 혼재되어 있어서, 형태소 분석하기 난감함
- Langdetect 패키지로 언어 파악 후 한글, 영어 각각 형태소 분석

```
In [8]: from langdetect import detect
```

	name_2	title	language
0	정보통신대학/ 김성근	연구윤리와 한국의 나노과학기술	ko
1	정보통신대학/ 김성근	자동요금징수시스템(ETCS) 표준화 연구(주파수 방식을 중심으로)	ko
4	공과대학/ 김성근	Special Relativity, and Gravity and Coulomb Force	en
5	공과대학/ 김성근	Derivation of Gaussian Distribution in Physics...	en

연구자추천 시스템 - 데이터 전처리

- 문제

- 연구분야 Stopwords set이 없음...열심히 만듦

```
Out[40]: Counter({'연구윤리': 6,  
                 '한국': 356,  
                 '나노과학': 1,  
                 '기술': 526,  
                 '자동': 120,  
                 '요금': 14,  
                 '징수': 11,  
                 '시스템': 1678,  
                 '표준화': 30,  
                 '연구': 3314,  
                 '주파수': 134,  
                 '방식': 187,
```

	index	0	stopwor
24	및	4009	s
9	연구	3314	s
84	방법	2619	s
43	분석	1693	s
7	시스템	1678	s
60	기반	1550	s
176	대한	1155	s
842	장치	1141	s
178	개발	1086	s
14	용	803	s
493	평가	801	s
26	기법	758	s

연구자추천 시스템 - 데이터 전처리

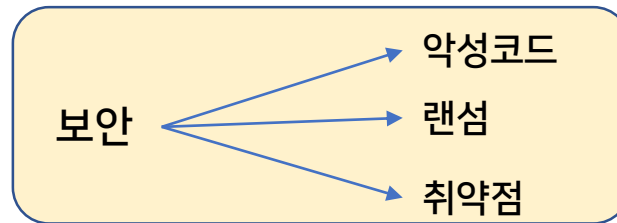
- 깔끔하게 word2vec 인풋데이터 만듦

```
In [168]: clean_title
```

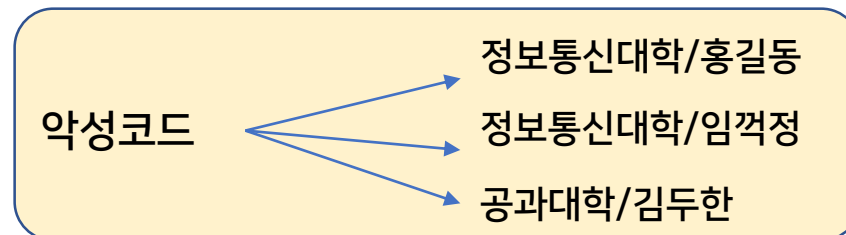
```
Out[168]: [['연구윤리', '나노과학'],  
           ['자동', '요금', '징수', '표준화', '주파수'],  
           ['기지국', '대역', '배열', '패치', '안테나', '설계'],  
           ['시뮬레이터', '인터넷', '공격', '표현', '맵핑'],
```

연구자추천 시스템 - w2v 모델

- 1. 특정 분야의 단어를 넣으면
 - 관련된 연구 키워드를 알려주고



- 2. 연구 키워드를 입력하면
 - 관련된 연구자들을 알려주고



연구자추천 시스템 - w2v 모델

- 키워드 -> 관련 연구 키워드
 - most_similar() 로 똑딱

```
In [107]: import gensim

num_features = 300
min_word_count = 1
num_workers = 1
context = 5
downsampling = 1e-3
```

```
model = gensim.models.Word2Vec(total_title, workers=num_workers,
                               size=num_features, min_count = min_word_count,
                               window = context, sample = downsampling)
```

```
model_name = "stopwords_only_title"
model.save(model_name)
```

```
In [112]: model.most_similar("머신", topn=30)
```

```
Out[112]: [('항재밍', 0.9985334277153015),
            ('네', 0.9980396628379822),
            ('기밀', 0.9976990222930908),
            ('안내', 0.9971907138824463),
            ('목적', 0.9970749616622925),
            ('매커니즘', 0.9969769716262817),
            ('렌더링', 0.9968122243881226),
```

연구자추천 시스템 - w2v 모델

- 키워드 -> 관련 연구자
 - 논문제목에 연구자이름을 끼워넣고 w2v모델 만듦

['워크플로우', '기반', '공과대학/ 김성근', '프로세스', '구현'],
['가전제품', '조립', '공과대학/ 김성근', '향상', '방안', '연구'],

- most_similar() 결과물에서 정규표현식으로 이름만 필터링

이것이 야매 추천 시스템!!!

연구자추천 시스템 - w2v 모델

- 키워드 -> 관련 연구자
 - most_similar() 결과물
 - 정규표현식으로 이름만 필터링

```
In [253]: model = gensim.models.Word2Vec.load('nametitle_raw_win_15')

import re
def find_person(word):
    try:
        tuple = model.wv.most_similar(word, topn=200)
        members = []
        p = re.compile('^([가-힣]+)/[가-힣]+$')
        for i in tuple:
            person = p.findall(i[0])
            if len(person)>0:
                members.append(person)
        return(members)
    except:
        return('없는 단어 입니다')
```

```
In [263]: find_person('gan')
```

```
Out[263]: [['대학원/안광준'],
            ['대학원/윤호섭'],
            ['대학원/이민수'],
```


연구자추천 시스템 - Slack 연동



Start here

Building Slack apps

Recent updates

Best practices

App blueprints

App features

Internal integrations

Incoming webhooks

Slash commands

Bot users

Dialogs

Shared Channels

Enterprise Grid

Legacy custom integrations

Workspace apps

Messages

Introduction

Basic formatting

Attaching content

Building Slack apps

Customize functionality for your own workspace or build a beautiful bot to share with the world. Provide your ingenious integrations with a suitably configurable container. Build a Slack app.



- [App capabilities](#)
- [Creating apps](#)
- [Building internal integrations](#) for your workspace
- [Installing apps](#) (with and without OAuth)
- [Distributing apps to other workspaces](#)
- [How users interface with your app](#)
- [About Slack's app directory](#)
- [Caring for and managing your Slack app record](#)

What your apps do is up to you — maybe it performs one distinct task incredibly well or perhaps it orchestrates a technological concert between individual Slack features and your internal accounting service.

First-timer or seasoned veteran? Often it's best to just get started and learn along the way.

Create a Slack app

Manage your apps

Create a Slack App



Interested in the next generation of apps?

We're improving app development and distribution. Join the API Preview period for workspace tokens and the Permissions API.

App Name

e.g. Super Service

Don't worry; you'll be able to change this later.

Development Slack Workspace

Development Slack Workspace

Your app belongs to this workspace—leaving this workspace will remove your ability to manage this app. Unfortunately, this can't be changed later.

By creating a Web API Application, you agree to the [Slack API Terms of Service](#).

Cancel

Create App

연구자추천 시스템 - Slack 연동

App Credentials

These credentials allow your app to access the Slack API. They are secret. Please don't share your app credentials with anyone, include them in public code repositories, or store them in insecure ways.

App ID

AAJSS4LRJ

Date of App Creation

May 6, 2018

Client ID

334098964660.358910156868

Client Secret

.....

Show

Regenerate


You'll need to send this secret along with your client ID when making your [oauth.access](#) request.


Verification Token

XxuJRp6aX1xTbxdjw1zuLjF6

Regenerate

For interactive messages and events, use this token to verify that requests are actually coming from Slack. Slash commands and interactive messages will both use this verification token.

 run.py

 slackbot_settings.py

```
API_TOKEN = "xoxb-334098964660-378937926466-J1ysgAOX2KlcbMV8GpH62irc"
```

연구자추천 시스템 - Slack 연동

- 슬랙봇 패키지로 간단히!

```
7 from slackbot.bot import Bot
8 from slackbot.bot import respond_to
-
27 @respond_to(".+")
28 def help(message):
29     answer = find_person(message.body["text"])
30     answer = "WWn".join(sum(answer, []))
31     answer = answer.replace("WWn", "Wn")
32     message.send(answer)
33
34 def main():
35     bot = Bot()
36     bot.run()
37
38 if __name__ == "__main__":
39     main()
```

```
model = gensim.models.Word2Vec.load('nametitle_raw_win_15')

import re
def find_person(word):
    try:
        tuple = model.wv.most_similar(word, topn=200)
        members = []
        p = re.compile('^[가-힣]+/[가-힣]+$')
        for i in tuple:
            person = p.findall(i[0])
            if len(person)>0:
                members.append(person)
        return(members)
    except:
        return('없는 단어 입니다')
```

끝!