

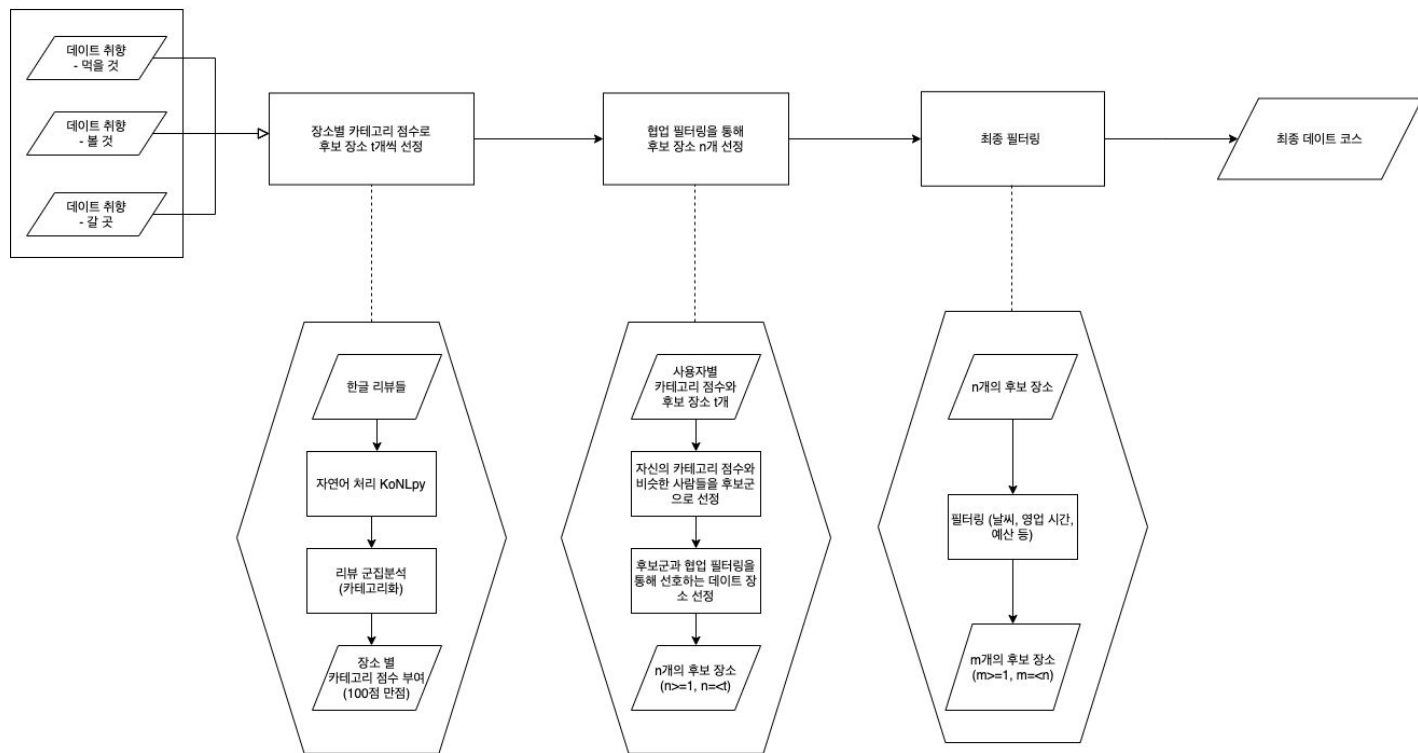
데이터분석캡스톤디자인

4주차 수행보고

Khupid 조

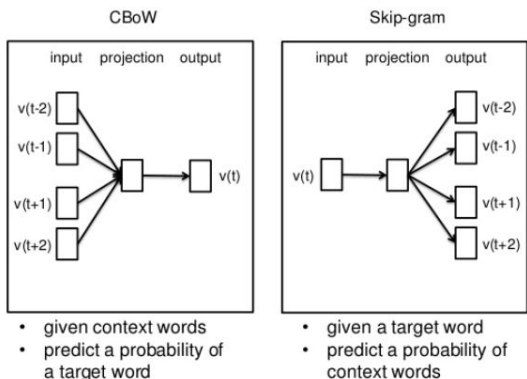
산업경영공학과 김동혁
관광학과 류연주
산업경영공학과 유정수

1. 데이터 필드 정의
2. 플로우 차트 제작 및 사용 기법 구체화
3. 수행 계획 개인별 구체화
4. 데이터 크롤링(각자 진행)



Facebook FastText

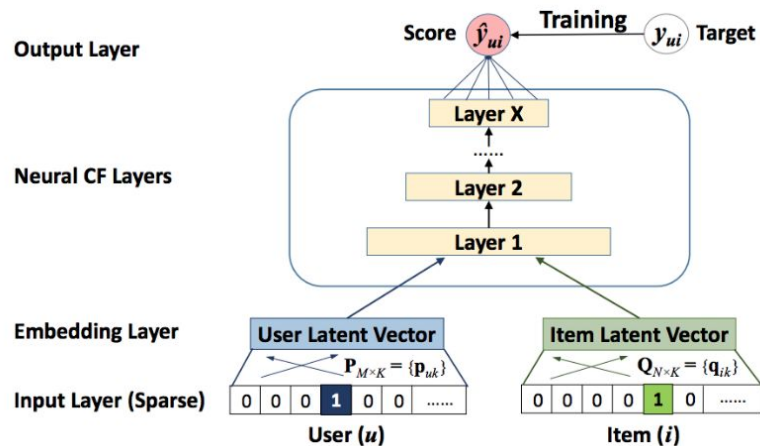
2 models in word2vec



wikipedia 데이터 기반으로 학습된 전세계 294개 언어로 된 pre-trained fastText를 제공하고 있음

기존에 자주 쓰이던 word2vec 기법은 형태학적인 단어의 유사도를 구별하지 못함
ex) teach, teacher, teaching

Neural Collaborative filtering



Collaborative filtering(협업 필터링) 기법을 Deep Neural Network를 이용해 학습하여 후보군을 추출한다.



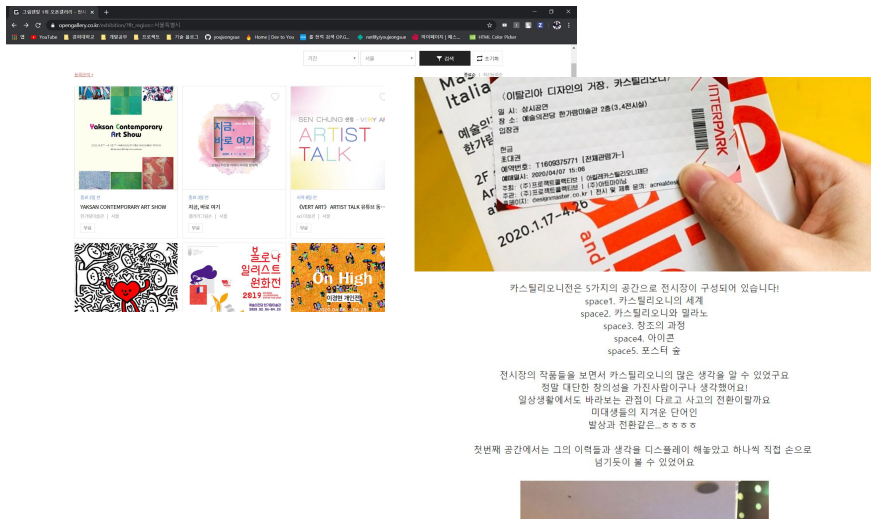
서울 구별 100개 장소씩, 총 2400개의 장소 + 장소별 리뷰 수집 중

추가적인 플랫폼 이용으로
데이터 보완 및 보충

순서	변수명	변수 설명	자료형
1	name	장소명	str
2	address	주소	str
3	main_mn	메인메뉴	str
4	opng_tm	영업시간	str
5	rating	총 평점	float
6	rvw_cnt	리뷰 수	int
7	tags	리뷰 태그	str

	name	address	main_mn	opng_tm	rating	rvw_cnt	tags
0	중앙해장	서울특별시 강남구 대치동 996-16	양선지해장국	{'매일': '24시간 영업', '일요일': '자정 12시 - 오후 10시 (월요일...}	4.3	49	저녁식사, 점심식사, 술모임, 식사모임, 가족외식, 회식, 데이트, 아침식사, 혼밥, 차모임, 기념일...
1	브루클린더버거조인트	서울특별시 강남구 신사동 540-11	더 치즈버거	{'매일': '24시간 영업'}	4.2	46	저녁식사, 점심식사, 데이트, 혼밥, 아침식사, 식사모임, 간식, 브런치, 여자끼리, 남자끼리, 가족...
2	농민백암순대본점	서울시 강남구 대치동 896-33 (알앤지타운 1F)	국밥	{'월-금': '오전 11시 - 오후 9시 (11:10분 시작)', '토요일': '...	4.5	36	점심식사, 저녁식사, 혼밥, 술모임, 데이트, 가족외식, 혼술, 식사모임, 아침식사, 회식, 실버푸드...
3	쥬즈	서울특별시 강남구 신사동 540-4	소릉포	{'매일': '오전 11시 30분 - 오후 9시 (Break time 15:00~1...	4.1	38	저녁식사, 점심식사, 데이트, 식사모임, 혼밥, 가족외식, 여자끼리, 기념일, 남자끼리, 친구들, 친...
4	다운타우너청담점	서울특별시 강남구 신사동 666-25	아보카도버거	{'매일': '오전 11시 30분 - 오후 9시 30분 (Last Order 20:...	4.5	38	저녁식사, 점심식사, 데이트, 혼밥, 간식, 식사모임, 여자끼리, 가족외식, 아침식사, 기념일, 남자...

정수 크롤링 - 전시(오픈갤러리)



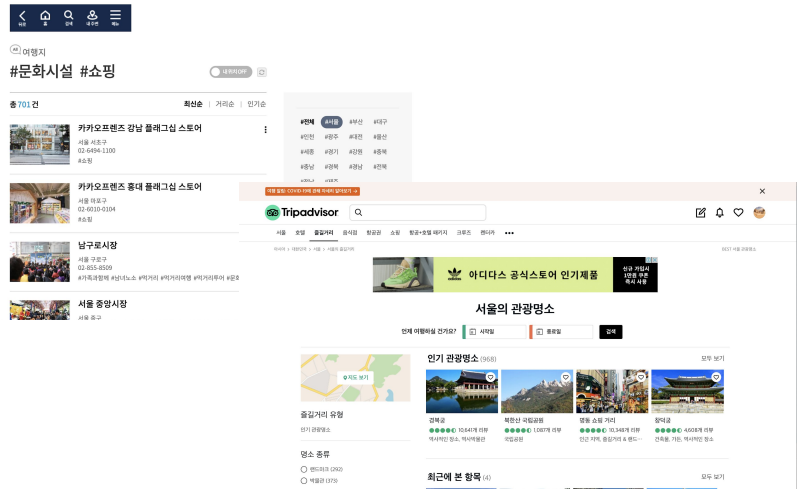
현재 **서울 지역**에서 진행 중인
12개 전시에 대한 정보 수집

각 전시에 대한 네이버 블로그 &
 카페 리뷰 수집 중

순서	변수명	변수 설명	자료형
1	title	전시 이름	str
2	artists	작가	str
3	location	위치 정보	location
4	open_at	오픈 시간	date
5	close_at	마감 시간	date
6	day	영업일	str
7	site	전시 정보 사이트	str
8	explain	전시 설명	str
9	image	대표 이미지 주소	str
10	price	관람료	int

연주 크롤링 - 갈곳

순서	변수명	변수 설명	자료형	수집방법
1	name	장소 이름	str	
2	location	소재지	category	
3	category	시설 카테고리	category	
4	open_at	개점 시간	date	
5	close_at	폐점 시간	date	
6	price	입장료 혹은 이용료	int	
7	score	리뷰 평점	int	
8	tags	리뷰 태그들	str	
9				
10				



result				
	name	location	category	tags
0	봉은사(서울)	강남구	1.0	#가족여행#관광지#교과서속여행#당일치기여행#봉은사(서울)#사찰&산사
1	최순우 옛집	성북구	1.0	#관광지#근대한국#기념관#아이와함께#역사#역사문화재#역사속#역사속으로
2	덕수궁	종로	1.0	#15_16한국관광100선#17_18한국관광100선#19_20한국관광100선#가족여행
3	국립민속박물관&국립민속박물관 어린이박물관	종로구	1.0	#관광지
4	창경궁	종로구	1.0	#15_16한국관광100선#17_18한국관광100선#19_20한국관광100선#가족여행
...
1937	방이동 생태경관 보전지역	송파구	8.0	#관광지#방이동생태경관보전지역#생태#수도권#자연
1938	한강	성동구	8.0	#관광지#자연
1939	아차산	광진구	8.0	#관광지#등산#산#수도권#아차산#자연#전망좋은곳
1940	반포 서래섬	서초구	8.0	#관광지#자연
1941	관악산	관악구	8.0	#관광지#관악산#등산#수도권#자연#힐링