개인맞춤형 주거지역 추천 서비스

추천 알고리즘

팀 집탐험대

20190910 이현지

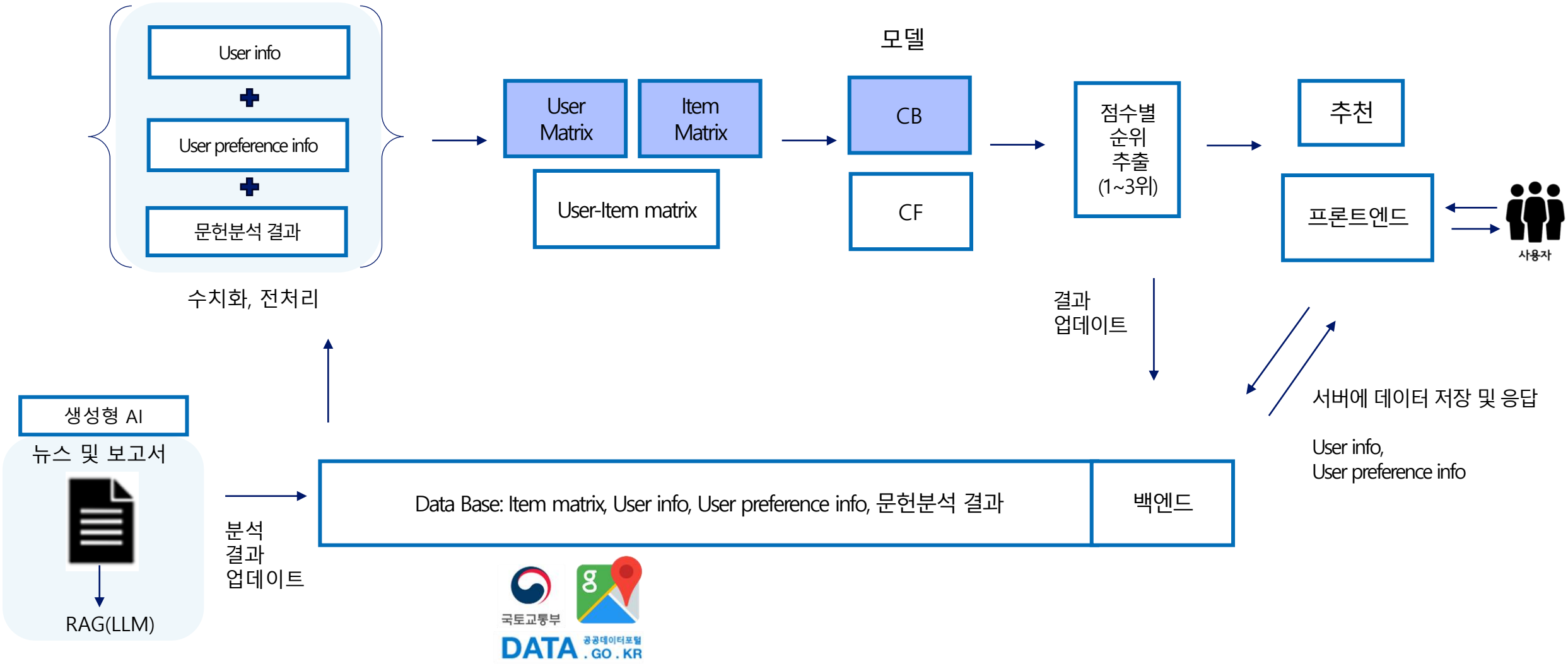
전체

1. 데이터

    세부 목차: 수집, 전처리

2. 시각화
3. 추천 알고리즘
4 . 웹디자인
5. 프론트엔드
6. 챗봇
7. 백엔드

추천 시스템: (1) 추천 시스템 구조



추천 시스템: (2) 설계 방법

콘텐츠 기반 모델 (CB)

✓ 모델 특징

- 사용자 정보, 아이템 정보를 활용하는 모델

✓ 모델의 강점

- 항목의 특성을 기반으로 분석
  - 콜드스타트, 특이 취향 유저에 대한 처리 가능
  - 대상 자체의 특성(feature)을 바탕으로 추천하기에 추천을 하는 근거를 설명할 수 있음
- 가장 원시적인 모델로 **직접적으로 특성에 대한 가중치 조절 가능**

데이터 수집

서울시 지역 정보 수집

- ✓ 추천 알고리즘의 item matrix에 사용될 데이터
- ✓ 항목
  - 42개 항목
  - 주로 기관 수 / 면적 조사
  - 안전: 범죄현황, cctv, 경찰서 등
  - 교육: 중학교, 고등학교, 특수학교 등
  - 의료, 교통, 생활시설, 환경, 기타
- ✓ 데이터 수집처
  - 공공데이터포털
  - 서울시 열린데이터 광장

	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구	동대문구	동작구	마포구	서대문구	서초구	성동구	성북구	송파구	양천구	영등포구	용산구	은평구	종로구	중구	중랑구
범죄현황	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
cctv개수	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
경찰서	14	9	9	10	9	10	8	5	8	8	10	7	8	8	6	9	10	11	8	10	7	9	20	15	8
파출소	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
지구대	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
마트	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
미술관	12.0	2.0	2.0	3.0	4.0	3.0	1.0	NaN	6.0	4.0	5.0	1.0	5.0	6.0	10.0	6.0	13.0	10.0	NaN	3.0	12.0	3.0	60.0	20.0	NaN
공연장	36	5	5	2	3	14	9	5	4	9	3	3	38	9	32	10	13	12	5	7	14	1	171	26	2
영화관	10	3	3	6	2	6	3	2	5	2	1	2	8	5	3	2	2	4	2	3	3	2	6	9	3
체육시설	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
풋살	2.0	2.0	1.0	1.0	1.0	2.0	1.0	2.0	6.0	1.0	NaN	NaN	4.0	2.0	1.0	4.0	3.0	3.0	3.0	2.0	1.0	2.0	NaN	NaN	NaN
테니스	5	5	2	6	4	3	5	2	3	4	3	5	3	4	7	6	3	5	2	2	2	4	3	3	2
탁구	3.0	5.0	1.0	5.0	NaN	1.0	2.0	3.0	3.0	2.0	1.0	4.0	4.0	1.0	4.0	2.0	2.0	1.0	3.0	2.0	3.0	1.0	4.0	NaN	NaN
축구	5	5	2	6	4	3	5	2	3	4	3	5	3	4	7	6	3	5	2	2	2	4	3	3	2
족구	2.0	4.0	1.0	5.0	1.0	1.0	NaN	NaN	2.0	1.0	NaN	1.0	3.0	NaN	3.0	4.0	3.0	3.0	4.0	3.0	1.0	1.0	NaN	NaN	2.0
야구	NaN	2.0	NaN	NaN	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	1.0	NaN	2.0	NaN	1.0	3.0	1.0	NaN	NaN	1.0	NaN	2.0
스쿼시	2.0	NaN	1.0	NaN	1.0	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN	NaN	1.0	1.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
수영	8	7	5	3	4	5	8	4	5	4	6	5	6	3	8	6	4	6	7	6	5	3	3	4	3
배드민턴	8	10	5	12	9	4	8	5	10	7	7	8	9	9	7	11	8	11	11	7	4	5	11	3	7
배구	NaN	2.0	NaN	NaN	2.0	NaN	1.0	NaN	NaN	3.0	1.0	1.0	2.0	NaN	1.0	NaN	NaN	3.0	3.0	1.0	1.0	NaN	NaN	2.0	NaN
농구	4	9	1	7	3	2	5	3	5	3	3	4	9	6	10	3	3	10	8	10	5	3	6	3	4
지하철	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
시내버스	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
광역버스	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
기차	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
따릉이	171	122	61	189	79	86	114	76	136	76	84	74	120	81	148	92	82	218	112	155	91	98	102	88	93
공원	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
미세먼지	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
소음	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
풍수해	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
주택침수	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
병원	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
어린이집	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
유치원	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
초등학교	33	26	14	35	22	21	26	18	42	23	21	21	22	19	24	21	29	40	30	23	15	30	13	12	24
중학교	25	18	13	23	16	13	14	9	27	13	15	16	14	14	16	11	18	28	18	12	9	18	9	8	14
고등학교	22	14	7	23	17	9	14	6	25	10	11	7	9	7	12	9	13	19	15	9	10	18	15	12	10
특수학교	NaN	2.0	4.0	NaN	3.0	NaN	2.0	2.0	NaN	NaN	NaN	2.0	2.0	1.0	NaN	NaN	1.0	2.0	NaN	NaN	NaN	2.0	6.0	2.0	NaN
인구통계	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
집값	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
노인복지시설	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
장애인복지시설	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

데이터 전처리: (1) 지역 정보

Item matrix

✓ 과정

- Raw data의 도로명 주소에서 '구' 추출
- 서울시 자치구별 기관의 개수 count
- item matrix에 업데이트

	연번	전체기관명	도로명주소	구
0	1	경찰청 서울특별시경찰청 서울중부경찰서 을지지구대	서울특별시 중구 을지로 234	중구
1	2	경찰청 서울특별시경찰청 서울중부경찰서 광희지구대	서울특별시 중구 퇴계로 375-1	중구
2	3	경찰청 서울특별시경찰청 서울중부경찰서 약수지구대	서울특별시 중구 동호로 5길 15, 약수지구대	중구
3	4	경찰청 서울특별시경찰청 서울중부경찰서 신당파출소	서울특별시 중구 다산로 248 (신당동, 신당파출소)	중구
4	5	경찰청 서울특별시경찰청 서울중부경찰서 장충파출소	서울특별시 중구 동호로 261	중구

구	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구	동대문구	동작구	마포구	서대문구	서초구	성동구	성북구	송파구	양천구	영등포구	용산구	은평구	종로구	중구	중랑구
전체기관수	14	9	9	10	9	10	8	5	8	8	10	7	8	8	6	9	10	11	8	10	7	9	20	15	8

	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구	동대문구	동작구	마포구	서대문구	서초구	성동구	성북구	송파구	양천구	영등포구	용산구	은평구	종로구	중구	중랑구
범죄현황	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
cctv개수	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
경찰서	14	9	9	10	9	10	8	5	8	8	10	7	8	8	6	9	10	11	8	10	7	9	20	15	8

데이터 전처리: (1) 지역 정보

Item matrix

✓ 과정

- 기관의 개수와 관련된 정보: 자치구의 면적 대비 개수
- 정규화

	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구
경찰서	0.166040	0.504123	0.518537	0.922622	0.539030	0.305636	0.009182	0.888631	0.945022	0.517929
마트	0.413616	0.170858	0.122650	0.121031	0.743667	0.987296	0.528436	0.035151	0.381106	0.305310
미술관	0.000757	0.124618	0.925627	0.862807	0.695232	0.804820	0.070049	0.869554	0.753210	0.968609
공연장	0.655050	0.155790	0.482464	0.450374	0.052364	0.996938	0.015063	0.591317	0.172262	0.010102
테니스	0.520953	0.677668	0.193265	0.407466	0.317512	0.682270	0.443371	0.898093	0.205366	0.945218
축구	0.201792	0.790534	0.876274	0.529418	0.446671	0.372449	0.887418	0.755597	0.496506	0.592141
초등학교	0.317354	0.584853	0.697348	0.199177	0.507345	0.040719	0.212259	0.844177	0.530031	0.005382
공원	0.645463	0.421798	0.478593	0.336234	0.758430	0.146772	0.963641	0.516519	0.324349	0.641324
미세먼지	0.733647	0.394459	0.560116	0.164915	0.995805	0.865096	0.378240	0.802366	0.511513	0.386464
병원	0.824380	0.286702	0.503665	0.128179	0.045627	0.955536	0.427687	0.797893	0.026753	0.896670
지하철	0.797121	0.847622	0.362927	0.405662	0.504514	0.508467	0.883019	0.339916	0.070819	0.503527
따릉이	0.075610	0.193837	0.303140	0.924731	0.646911	0.058646	0.004158	0.769158	0.450794	0.799366
노인복지시설	0.829940	0.458442	0.998480	0.250926	0.310822	0.204735	0.947169	0.566732	0.276322	0.054862

Item matrix

데이터 전처리: (1) 지역 정보

Item matrix

	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구	동대문구	동작구	마포구	서대문구	서초구	성동구	성북구	송파구	양천구	영등포구	용산구	은평구	종로구	중구	중랑구
범죄현황	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
cctv개수	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
경찰서	14	9	9	10	9	10	8	5	8	8	10	7	8	8	6	9	10	11	8	10	7	9	20	15	8
파출소	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
지구대	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
마트	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
미술관	12.0	2.0	2.0	3.0	4.0	3.0	1.0	NaN	6.0	4.0	5.0	1.0	5.0	6.0	10.0	6.0	13.0	10.0	NaN	3.0	12.0	3.0	60.0	20.0	NaN
공연장	36	5	5	2	3	14	9	5	4	9	3	3	38	9	32	10	13	12	5	7	14	1	171	26	2
영화관	10	3	3	6	2	6	3	2	5	2	1	2	8	5	3	2	2	4	2	3	3	2	6	9	3
체육시설	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
풋살	2.0	2.0	1.0	1.0	1.0	2.0	1.0	2.0	6.0	1.0	NaN	NaN	4.0	2.0	1.0	4.0	3.0	3.0	3.0	2.0	1.0	2.0	NaN	NaN	NaN
테니스	5	5	2	6	4	3	5	2	3	4	3	5	3	4	7	6	3	5	2	2	2	4	3	3	2
탁구	3.0	5.0	1.0	5.0	NaN	1.0	2.0	3.0	3.0	2.0	1.0	4.0	4.0	1.0	4.0	2.0	2.0	1.0	3.0	2.0	3.0	1.0	4.0	NaN	NaN
축구	5	5	2	6	4	3	5	2	3	4	3	5	3	4	7	6	3	5	2	2	2	4	3	3	2
족구	2.0	4.0	1.0	5.0	1.0	1.0	NaN	NaN	2.0	1.0	NaN	1.0	3.0	NaN	3.0	4.0	3.0	3.0	4.0	3.0	1.0	1.0	NaN	NaN	2.0
야구	NaN	2.0	NaN	NaN	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	1.0	NaN	2.0	NaN	1.0	3.0	1.0	NaN	NaN	1.0	NaN	2.0
스쿼시	2.0	NaN	1.0	NaN	1.0	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN	NaN	1.0	1.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
수영	8	7	5	3	4	5	8	4	5	4	6	5	6	3	8	6	4	6	7	6	5	3	3	4	3
배드민턴	8	10	5	12	9	4	8	5	10	7	7	8	9	9	7	11	8	11	11	7	4	5	11	3	7
배구	NaN	2.0	NaN	NaN	2.0	NaN	1.0	NaN	NaN	3.0	1.0	1.0	2.0	NaN	1.0	NaN	NaN	3.0	3.0	1.0	1.0	NaN	NaN	2.0	NaN
농구	4	9	1	7	3	2	5	3	5	3	3	4	9	6	10	3	3	10	8	10	5	3	6	3	4
지하철	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
시내버스	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
광역버스	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
기차	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
따릉이	171	122	61	189	79	86	114	76	136	76	84	74	120	81	148	92	82	218	112	155	91	98	102	88	93
공원	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
미세먼지	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
소음	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
풍수해	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
주택침수	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
병원	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
어린이집	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
유치원	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
초등학교	33	26	14	35	22	21	26	18	42	23	21	21	22	19	24	21	29	40	30	23	15	30	13	12	24
중학교	25	18	13	23	16	13	14	9	27	13	15	16	14	14	16	11	18	28	18	12	9	18	9	8	14
고등학교	22	14	7	23	17	9	14	6	25	10	11	7	9	7	12	9	13	19	15	9	10	18	15	12	10
특수학교	NaN	2.0	4.0	NaN	3.0	NaN	2.0	2.0	NaN	NaN	NaN	2.0	2.0	1.0	NaN	NaN	1.0	2.0	NaN	NaN	NaN	2.0	6.0	2.0	NaN
인구통계	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
집값	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
노인복지시설	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
장애인복지시설	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

데이터 전처리: (2) 사용자 정보

User matrix

- ✓ User matrix(표3)는 유저 선호도 정보를 담음 (**항목별 가중치**)
- ✓ 과정
  - 사용자 설문 결과표(표1)와 사용자가 지정한 우선순위표(표2), 문헌분석결과를 바탕으로 생성
  - 표1과 표2는 설문(프론트엔드에서 진행) 결과를 표로 저장(백엔드)한 후 불러와서 사용

	사용자	성별	연령	가족형태	혼인여부	자녀	취미생활	교통	공원	기타	복지시설	성향
0	A	여성	20대	4인	미혼	없음	테니스	지하철,버스	매우 중요	없음	없음	활발
1	B	남성	30대	2인	기혼	없음	축구	지하철	보통	없음	없음	활발
2	C	남성	40대	4인	기혼	초등학생1명	공연	버스	보통	없음	노인복지	조용
3	D	여성	50대	3인	기혼	대학생1명	미술관	자차	중요	없음	없음	조용

표1. User Info (사용자 설문 결과)

	사용자/우선순위	안전	생활시설	교육	의료	환경	교통	기타
0	A	1	3	5	2	6	4	7
1	B	2	3	6	1	7	4	5
2	C	4	2	7	1	5	3	6
3	D	1	4	3	5	2	6	7

표2. User preference info
(사용자가 지정한 우선순위)

	경찰서	마트	미술관	공연장	테니스	축구	초등학교	공원	미세먼지	병원	지하철	버스	노인복지시설
A	10	6	1	1	6	1	4	10	3	7	11	11	2
B	7	7	1	1	1	6	3	8	2	8	12	7	4
C	7	7	1	6	1	1	8	7	4	8	6	11	8
D	8	5	6	1	1	1	6	12	8	6	3	3	2

표3. User matrix

추천 시스템: (3) CB 모델

콘텐츠 기반 모델(CB)

✓ 내적

	경찰서	마트	미술관	공연장	테니스	축구	초등학교	공원	미세먼지	병원	지하철	버스	노인복지시설
A	10	6	1	1	6	1	4	10	3	7	11	11	2
B	7	7	1	1	1	6	3	8	2	8	12	7	4
C	6	7	1	6	1	1	8	7	4	8	6	11	8
D	8	5	6	1	1	1	6	12	8	6	3	3	2

User matrix

	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구
경찰서	0.166040	0.504123	0.518537	0.922622	0.539030	0.305636	0.009182	0.888631	0.945022	0.517929
마트	0.413616	0.170858	0.122650	0.121031	0.743667	0.987296	0.528436	0.035151	0.381106	0.305310
미술관	0.000757	0.124618	0.925627	0.862807	0.695232	0.804820	0.070049	0.869554	0.753210	0.968609
공연장	0.655050	0.155790	0.482464	0.450374	0.052364	0.996938	0.015063	0.591317	0.172262	0.010102
테니스	0.520953	0.677668	0.193265	0.407466	0.317512	0.682270	0.443371	0.898093	0.205366	0.945218
축구	0.201792	0.790534	0.876274	0.529418	0.446671	0.372449	0.887418	0.755597	0.496506	0.592141
초등학교	0.317354	0.584853	0.697348	0.199177	0.507345	0.040719	0.212259	0.844177	0.530031	0.005382
공원	0.645463	0.421798	0.478593	0.336234	0.758430	0.146772	0.963641	0.516519	0.324349	0.641324
미세먼지	0.733647	0.394459	0.560116	0.164915	0.995805	0.865096	0.378240	0.802366	0.511513	0.386464
병원	0.824380	0.286702	0.503665	0.128179	0.045627	0.955536	0.427687	0.797893	0.026753	0.896670
지하철	0.797121	0.847622	0.362927	0.405662	0.504514	0.508467	0.883019	0.339916	0.070819	0.503527
따릉이	0.075610	0.193837	0.303140	0.924731	0.646911	0.058646	0.004158	0.769158	0.450794	0.799366
노인복지시설	0.829940	0.458442	0.998480	0.250926	0.310822	0.204735	0.947169	0.566732	0.276322	0.054862

Item matrix

	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구
A	31.525291	40.147040	46.375280	25.018446	26.245367	47.813658	34.462801	44.625601	46.227770	42.904204
B	30.535635	35.860424	39.424981	24.476356	24.415362	40.641647	30.857115	40.718282	41.600804	40.894238
C	34.218555	37.673022	44.729631	24.398890	28.858180	46.085683	35.468856	41.550753	48.605106	46.087203
D	22.557123	27.982963	34.262560	23.662304	24.757876	39.852693	32.774256	35.071538	41.685719	37.477234

User-item matrix

```
# 사용자 C의 점수가 가장 높은 구 3개 찾기
user_C_scores = user_item_matrix3.loc['C']
top_3_gu = user_C_scores.nlargest(3).index

print("사용자 C에 대한 가장 높은 점수를 가지는 구 3개:")
print(top_3_gu)
```

```
사용자 C에 대한 가장 높은 점수를 가지는 구 3개:
Index(['노원구', '도봉구', '광진구'], dtype='object')
```

생성형 AI - 가중치

실험 : 가중치 관련

가중치 관련 실험

✓ A 방법

- **주관적 가중치만을** 사용 : User info + User preference info
- 자녀: 초등학생 1명 → '초등학교'에 가중치 부여

✓ B 방법

- **객관적 가중치 추가**: 사용자 설문 결과 + 사용자가 지정한 우선순위 + **객관적 가중치**
- 생성형 AI를 통한 뉴스 및 보고서 분석 결과를 통해 설정한 기준
- 자녀: 초등학생 1명 → ' 초등학교'에 가중치 부여 + **'안전'에 가중치 부여**

* 사용자 기본 정보(성별, 연령, 가족형태, 혼인여부, 자녀)에 대한 분석

사용자 C

남성
40대
4인 가족
기혼
자녀: 초등학생 1명

취미생활: 공연관람
교통: 버스
공원: 중요도 – 보통
기타 – 없음
복지시설: 노인복지시설
성향: 조용

가중치

가중치 설정 방식

✓ CB 모델에서는 가중치 설정이 매우 중요

✓ 설정 방식: 가중치 합

	경찰서	마트	미술관	공연장	테니스	축구	초등학교	공원	미세먼지	병원	지하철	버스	노인복지시설
A	8+2	6	1	1	1+5	1	4	6+4	3	7	5+1+5	5+1+5	2
B	7	6+1	1	1	1	1+5	3	5+3	2	8	5+2+5	5+2	4
C	5+2	7	1	1+5	1	1	1+2+5	4+3	4	8	6	6+5	3+5
D	8	5	1+5	1	1	1	6	7+1+4	7+1	4+2	3	3	2

① 주관적: User preference info

- 사용자가 홈페이지 내 설문을 통해 입력한 우선순위 정보 (1~8점)

② 주관적: User info 사용자의 설문 결과를 바탕으로 1~5점

- 사용자의 설문 결과를 바탕으로 (1~5점)

③ 객관적: 문헌 분석 결과를 바탕으로

- 성별, 연령, 가족형태, 혼인여부, 자녀와 관련된 기본 정보 사용 (성별+연령 / 가족형태 / 혼인여부 / 자녀)
- 사용자의 기본 정보를 바탕으로 자동으로 가중치가 부여되도록 (사용자의 설문 결과와는 별개로)
- **생성형 AI를 활용한 자료 분석**: 여러 데이터 (기사, 국토교통부 보고서 등)를 PDF 형태로 수집

실험 : 가중치 관련

가중치 관련 실험

- ✓ A 방법
- ✓ B 방법: 생성형 AI의 분석 결과를 바탕으로

	경찰서	마트	미술관	공연장	테니스	축구	초등학교	공원	미세먼지	병원	지하철	버스	노인복지시설
A	10	6	1	1	6	1	4	10	3	7	11	11	2
B	6	7	1	1	1	6	3	8	2	8	12	7	4
C	5	7	1	6	1	1	8	7	4	8	6	11	8
D	8	5	6	1	1	1	6	12	8	6	3	3	2

	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구
A	31.525291	40.147040	46.375280	25.018446	26.245367	47.813658	34.462801	44.625601	46.227770	42.904204
B	29.947611	35.098558	38.619373	24.400999	24.234897	39.703618	30.065751	40.196848	41.382887	40.268812
C	33.630531	36.911157	43.924023	24.323533	28.677715	45.147654	34.677492	41.029320	48.387188	45.461778
D	22.557123	27.982963	34.262560	23.662304	24.757876	39.852693	32.774256	35.071538	41.685719	37.477234

```
print("사용자 C에 대한 가장 높은 점수를 가지는 구 3개:")
print(top_3_gu)
```

```
사용자 C에 대한 가장 높은 점수를 가지는 구 3개:
Index(['노원구', '도봉구', '광진구'], dtype='object')
```

A방법: 주관적 가중치

	경찰서	마트	미술관	공연장	테니스	축구	초등학교	공원	미세먼지	병원	지하철	버스	노인복지시설
A	10	6	1	1	6	1	4	10	3	7	11	11	2
B	7	7	1	1	1	6	3	8	2	8	12	7	4
C	7	7	1	6	1	1	8	7	4	8	6	11	8
D	8	5	6	1	1	1	6	12	8	6	3	3	2

	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구
A	31.525291	40.147040	46.375280	25.018446	26.245367	47.813658	34.462801	44.625601	46.227770	42.904204
B	30.535635	35.860424	39.424981	24.476356	24.415362	40.641647	30.857115	40.718282	41.600804	40.894238
C	34.806579	38.434888	45.535239	24.474247	29.038645	47.023712	36.260220	42.072187	48.823023	46.712629
D	22.557123	27.982963	34.262560	23.662304	24.757876	39.852693	32.774256	35.071538	41.685719	37.477234

```
print("사용자 C에 대한 가장 높은 점수를 가지는 구 3개:")
print(top_3_gu)
```

```
사용자 C에 대한 가장 높은 점수를 가지는 구 3개:
Index(['노원구', '광진구', '도봉구'], dtype='object')
```

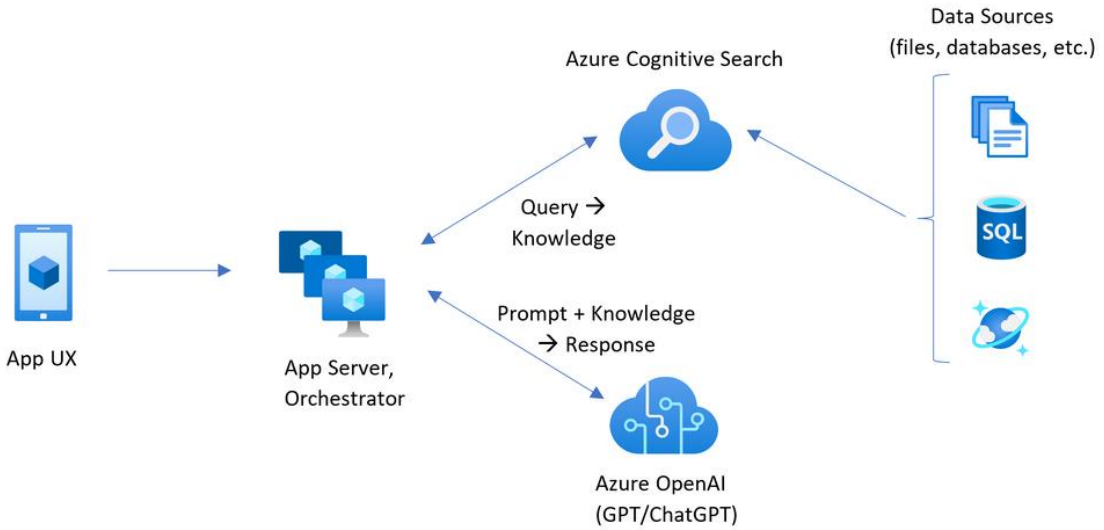
B방법: 주관적 + 객관적 가중치

생성형 AI 제작

구조

✓ 생성형 AI

- ChatGPT-like experiences over your own data using the Retrieval Augmented Generation(RAG) Pattern
- Tools
  - Azure Developer CLI, Python 3.10, Node.js 14+, Git, Powershell 7+ (pwsh)
  - Azure OpenAI Service: to access the ChatGPT model (gpt-35-turbo)
  - Azure Cognitive Search: for data indexing and retrieval



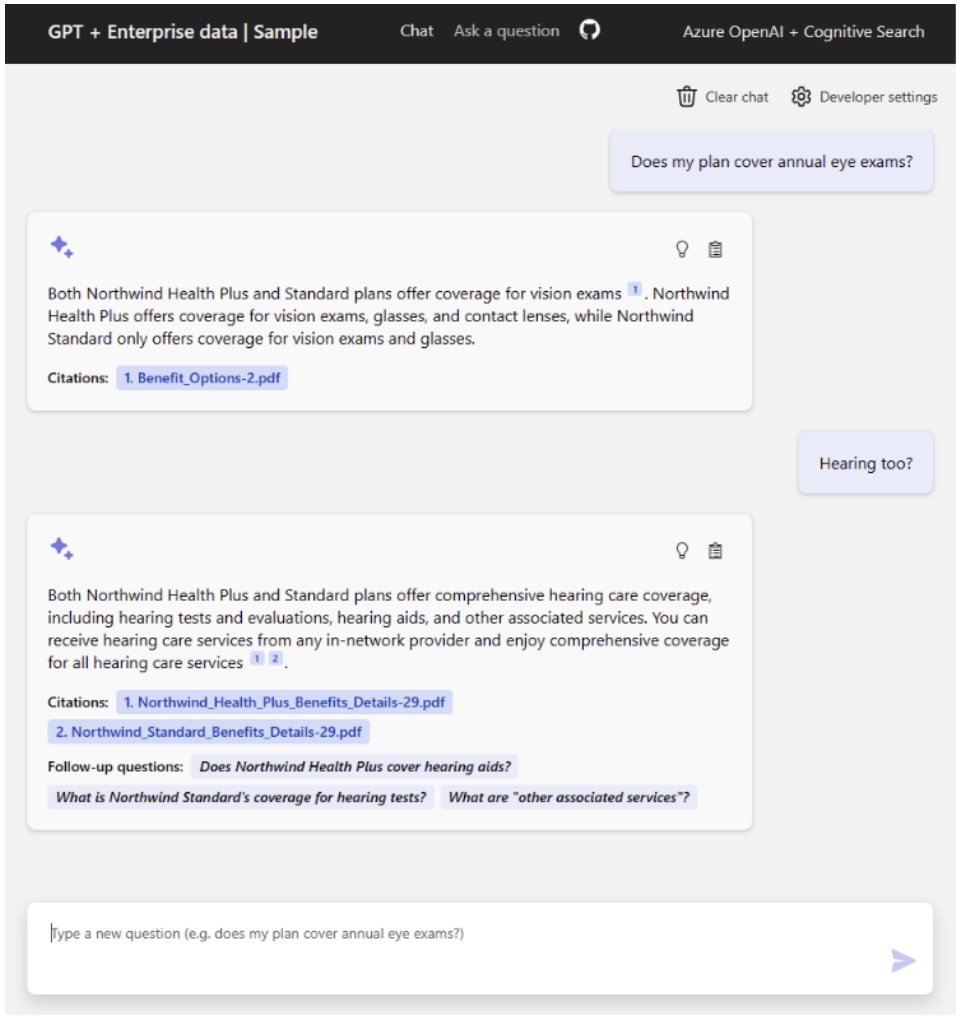
생성형 AI 제작

생성형 AI

✓ 실행 과정

- 뉴스 기사, 보고서 등 다양한 정보를 크롤링하여 얻은 pdf 파일을 생성형 AI에 upload한다.
- 기본정보(연령, 나이 등)에 따라 가장 중요하게 생각하는 요소가 무엇인지 묻는다.
- 제작한 생성형AI는 입력 받은 pdf 파일들을 RAG(LLP) 모델을 활용해 분석 결과를 제공해준다.

- 질문: 20대 여성이 중요하게 생각하는 요소는?
- 질문: 기혼, 초등학생 자녀가 있는 사람이 중요하게 생각하는 요소는?
- 생성형AI의 분석 결과를 바탕으로 가중치를 설정한다.



추후 계획

계획

- ✓ 협력필터링(CF) 모델 개발
  - 샘플 데이터 수집: 20대 여성 대상으로 설문조사 실시, 사용자 정보와 선호 지역 3곳
  - 모델 개발
- ✓ 프론트엔드, 백엔드와 통신
  - 백엔드에서 데이터 받아오기, 추천 시스템의 추천 결과를 전송하기
  - Node.js의 child process, python shell 혹은 Flask 사용
- ✓ 가중치 스케일링

부록 및 참고문헌

가중치 부여 예시

User matirx 생성

✓ 수집한 3가지 정보를 바탕으로 생성

- 가중치 합
- User preference info: 1위부터 7위까지
  - (1~8점. 기본적으로 1점이 주어지고 순위에 따라 점수가 추가되는 형태)
- User info: 1~5점으로 부여. 매우 중요 / 중요 / 보통 / 중요하지 않음 / 전혀 중요하지 않음
- 문헌 분석 결과: 사용자 기본 정보에 대해 1~2점(1, 2위)으로 부여. (성별+연령 / 가족형태 / 혼인여부 + 자녀)
- 가중치 스케일링 필요

사용자 C

남성
40대
4인 가족
기혼
자녀: 초등학생 1명

취미생활: 공연관람
교통: 버스
공원: 중요도 – 보통
기타 – 없음
복지시설: 노인복지시설
성향: 조용

	경찰서	마트	미술관	공연장	테니스	축구	초등학교	공원	미세먼지	병원	지하철	버스	노인복지시설
User preference info	4위(안전) 5	2위(생활시설) 7	X 1	X 1	X 1	X 1	7위(교육) 2	5위(환경) 4	5위(환경) 4	1위(의료) 8	3위(교통) 6	3위(교통) 6	6위(기타) 3
User info				취미생활 5			자녀- 초등학생 1명 5	중요도-보통 3				교통-버스 5	노인복지시설 5
문헌분석 결과	자녀 - 1위 2						자녀 - 2위 1						
총합	7	7	1	6	1	1	8	7	4	8	6	11	8

CF 모델

협력필터링(CF) 모델

✓ A 방법 : 임의로 생성한 sample data를 활용

- 사용자의 기본 정보(임의로 생성), rating(근거 없음)
- 이를 바탕으로 지식 기반 모델을 돌려 그 결과를 sample data로 활용해서 협력필터링 모델 학습에 활용
- 문제점: 실제 선호도가 반영되지 않음. 근거 없는 무작위의 데이터였기 때문에 정확도가 낮음

✓ B 방법

- 사전에 수집한 설문조사 결과를 바탕으로 학습시킴
- 20대 여성에 한정하여 설문 진행: user info, user preference info와 선호하는 지역 3군데 수집
- 20대에 한정하더라도 유의미한 결과를 제공할 수 있도록 하기 위함
- 졸업 전시 시 수집되는 데이터를 실시간으로 업데이트하여 활용
- 보다 신뢰도 있는 결과 제공

- 행렬 데이터 형태를 이용
- 지식기반모델을 사용하며 전시회장에서 사용자의 데이터를 행렬 형태로 수집해둔 것을 바로 적용할 수 있어 용이

CF 모델

협력필터링(CF) 모델

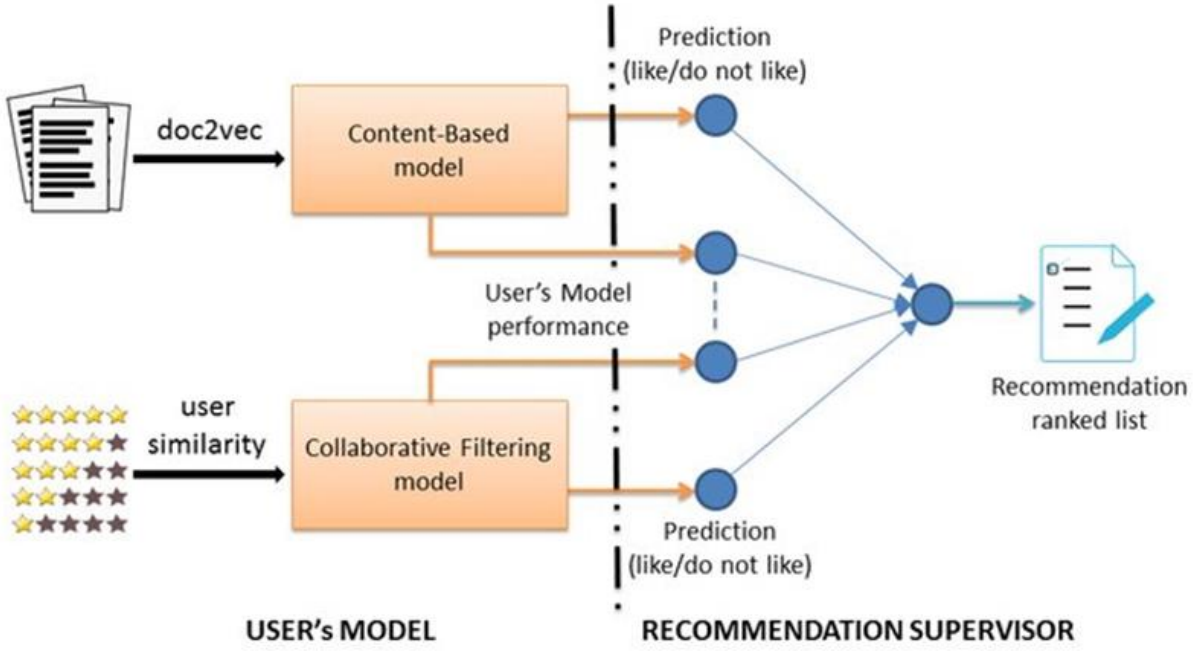
✓ A 방법에서 사용했던 샘플데이터 형태

유저	연령	소득	공원	취미	의료시설	지하철	가격 (최대)			도봉구	강남구	송파구	서초구	동작구
A	25	2500	중요	수영	1	중요	3000			3.5	5	5	4.2	1
B	30	3200	중요	등산	2	중요	2000			4	4.2	4	3.5	2
C	28	2300	3개 이상	배구	3	3호선	8000			4	2	5	3	2
D	65	5000	5개 이상	축구	0	환승역필요	15000			1	2	4	5	3
E	21	1200	2개 이상	테니스	5	중요	2100			5	2	2	4	5
F	45	8600	불필요	수영	중요함	중요	4500			2	4	3	5	1
G	53	9000	2개 이상	등산	5	불필요	8600			8	2	4	5	3
H	26	5000	1개 이상	테니스	3	중요	2400			1	2	5	3	5

CB + CF 모델

✓ Recommendation Supervisor

- 추천 결과를 검증하는 과정으로, CB 모델과 CF 모델의 추천 결과를 비교하여 융합한 순위 목록을 작성
- 이 과정을 통해 최종적으로 작성된 순위 목록을 바탕으로 사용자에게 추천 결과를 제공
- 피드포워드 신경망(Feedforward Neural Network, FNN)이 활용될 수 있음

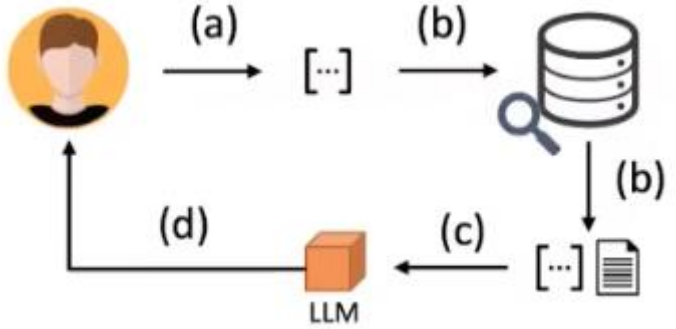


생성형 AI 제작

Retrieval Augmented Generation Pattern(RAG)

✓ 장점

- Knowledge base의 어떠한 부분을 참고하였는지 알기 쉬움
- 정보 검색 task에서 좋은 성능을 보여줌
- 신뢰할 수 있는 자료 수집

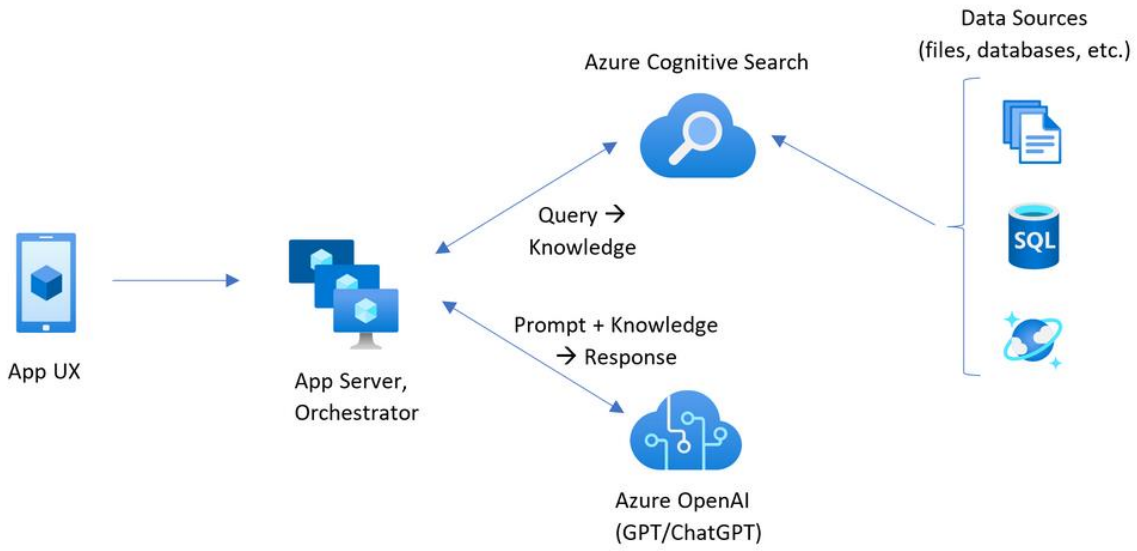


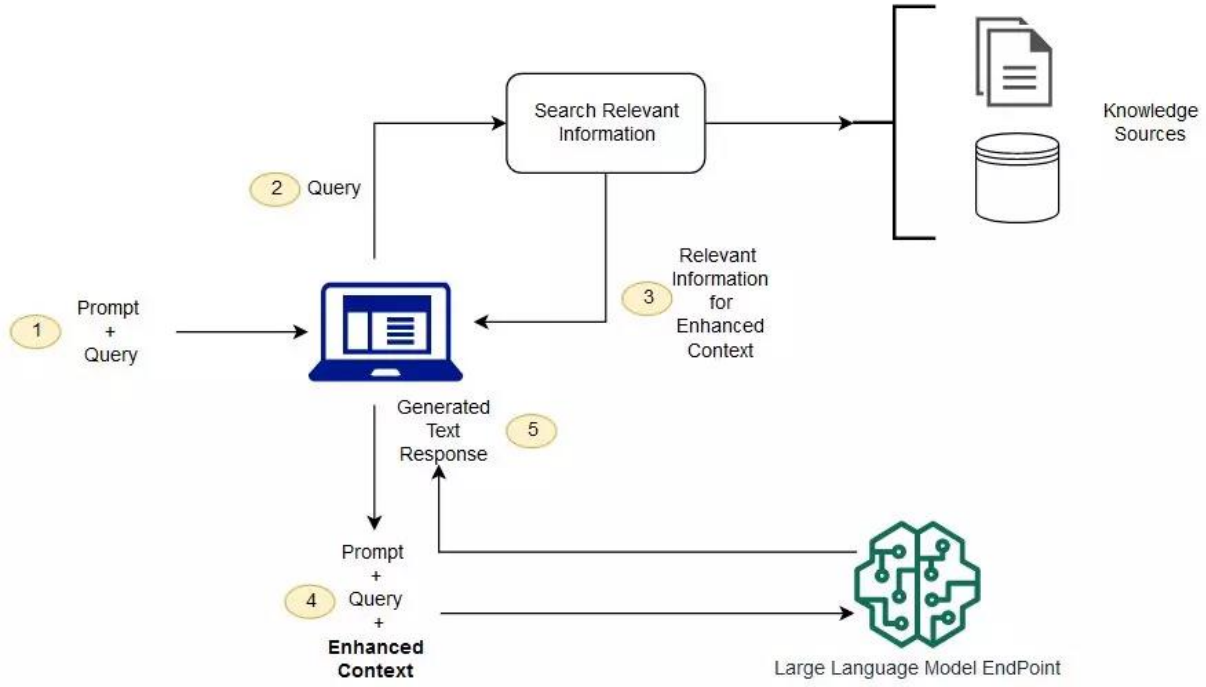
✓ 과정

- 자체 데이터 수집: 원본 파일에서 문서 데이터 추출, 문서를 청크 단위로 잘라 벡터 형태로 임베딩, 벡터DB에 저장
- 문서 쿼리:
  - 질문을 벡터로 임베딩 -> 가장 일치하는 청크를 찾음 (벡터DB에서) -> 프롬프트 작성(질문과 찾은 유사한 청크)하여 LLM에 전달 -> LLM의 응답을 사용자에게 전달

생성형 AI 제작

Retrieval Augmented Generation Pattern(RAG)





참고 문헌

[1] 차루아가르왈.박희원 외 옮김.『 추천 시스템』. 에이콘, 2022.

[2] Felfernig, Alexander & Burke, Robin. (2008). Constraint-based recommender systems: Technologies and research issues. ACM International Conference Proceeding Series. 3. 10.1145/1409540.1409544.

[3] Gabriele Sottocornola, Fabio Stella, Markus Zanker, and Francesco Canonaco. 2017. Towards a deep learning model for hybrid recommendation. In Proceedings of the International Conference on Web Intelligence (WI '17). Association for Computing Machinery, New York, NY, USA, 1260–1264. https://doi.org/10.1145/3106426.3110321

[4] https://github.com/Azure-Samples/azure-search-openai-demo/