

- 8.16 회의결과

데이터	회의결과	데이터 전처리 및 활용 계획
1. 진료과목(전공과목)	딥러닝을 이용하여 치과 전문과목 대표 10개로 다중 Classification 참조모델 : https://wikidocs.net/22933	1. 각 의사 약력에서 학회, 협회 등 단체를 10개의 대표과목과 연결 2. 전문의 / 인정의도 마찬가지로 연결, 단 전문의 / 인정의 / 정회원은 구분되어야함
2. 학회회원 여부		
3. 전문의 (보건복지부)		
4. 의사별 리뷰	감성분석을 통한 알림(경고창)이나 키워드로 활용. 단 시간상 추후 활용	모두닥 데이터 + 구글, 네이버, 다음 리뷰와 평가를 평점과 연결 시킨 후, 각 병원마다 적용. 또는 각 병원마다 평균평점 제시
5. 해당병원의 리뷰,평가		
6 논문 데이터	대표 10개 과목과 관련된 전문용어 그룹을 각각 생성 후, 각 그룹을 축으로 하여 의사별로 논문은 어느 그룹에 속하는지 표시 : 이후 위 딥러닝 모델과 병합하여 전문성으로 표시.	논문을 모든 전문의에 연결 후 각 대표 10개 과목과 관련된 단어그룹을 생성해야 함 (토픽 모델링 활용). 이후 수업 때 배운 모델 응용
7. 각 병원의 개업일자	경력을 Normalization 후에 6번 논문데이터를 활용한 전문성에 곱하는 것으로 계획	공공데이터를 병합
8. 면허취득 일자	구할 수 있나..? > 개인정보라 구하기 힘들	
9. 지식인 답변내용	증상-질병 매칭글에서 증상 키워드 추출 (domain 지식이 필요할 것으로 예상) 후 10개 과목 라벨링	각 답변 데이터를 10개의 대표과목으로 라벨링 (토픽 모델링 활용 또는 각 의사의 전문과목 활용 또는 domain지식 활용)
10. 관련기사/컬럼	유명인 혹은 인지도가 높은 의사만 혹은 활발한 의사만 많은 경우가 있어 전문성으로는 활용하기 어려움.	숫자나 목록을 표시해주는 정도로 활용

치과 전문과목 10

과목명	주요 진료 내용
구강악안면외과	구강암, 주걱턱 성형, 양악수술, 구순구개열, 턱관절 질환, 사랑니 발치, 임플란트
치과보철과	손상·상실된 치아 대체(임플란트, 의치)
치과교정과	치열교정
소아치과	소아청소년 치과 질환 치료관리·교정
치주과	치은염·치주염 등 치주 질환, 잇몸뼈 손상, 임플란트
치아보존과	치아 손상, 치수·치근단 질환, 치아 심미치료
구강내과	구강점막 질환, 턱관절 장애, 만성구강안면통증, 코골이·수면무호흡증
구강악안면방사선과	각종 치과 질환의 영상검사 및 진단
구강병리과	치과 질환의 세포·조직 검사 및 진단
예방치과	구강 질환 발병 전 예방적 처치

- 약력 이후 진행계획

서동욱 : 진료과목 10개에 관련된 각각의 데이터 수집

김은하 : 의사 약력 정규화

서주영 : 논문 의사 이름 정규화

박도건 : 논문 의사 이름 정규화

송환의 : 지식인, 증상-질병 매칭글에서 증상 키워드 추출 => 10개 과목 라벨링

- 의사를 나타낼 수 있는 데이터

데이터	활용방안	수집계획
1. 진료과목(전공과목)	각 진료과목에 연관된 내용을 활용 ★	노가다를 뛰면서 약력 긁어오는중
2. 학회회원여부 (전문의/인정의 포함)	각 학회나 전문의 별로 중점적으로 다루는 키워드나 내용을 활용 ★	노가다를 뛰면서 약력 긁어오는중 + 각학회
3. 관련기사/컬럼	각 문서에서의 키워드나 토픽정도 활용	BoilerPipe활용 계획, 하지만 데이터가 별로 없음
4. 의사별 리뷰	키워드 정도로 활용	거의 없거나 모두닥 데이터 정도
5. 해당병원의 리뷰,평가	키워드나 토픽 활용 ★	네이버, 구글 각병원 리뷰나 카페 리뷰 활용, 모두닥 등의 리뷰 활용
6 논문 데이터	Abstract를 이용한 키워드나 토픽 활용	약력을 통해 연결하려하지만 될지는 의문
7. 각 병원의 개업일자	오래된 병원일수록 경험이 많고 폐업하지 않았다는 뜻이므로 신뢰도를 높이는 자료로 활용 가능	공공데이터에 존재
8. 면허취득 일자	구할 수 있나..?	
9. 지식인 답변내용	의사별 리뷰처럼 키워드로 활용?	

- 요약

많은 의사에게서 구할 수 있는 것 : 1, 2, 5, 7

현재 모두닥에서 연결한 의사 1621명, 각 의사별로 리뷰가 있다는 가정을하면 4번도 활용 가능.

적은 의사에게서만 희박하게 구할 수 있을 것으로 예상되는 것 : 3, 6, 9

+ 모두닥에서 docinfo가 있는 의사만 연결을 해놓은 것이라서 docinfo가 없지만 병원-의사 연결 된 의사는 의사별 리뷰 데이터(4)를 활용 가능

- 남은 의문점

결국 대부분 의사는 1, 2, 5, 7 데이터만으로 특성을 나타내야 하는데 차이점이 너무 적어 보이는데...?

최대한 3, 4, 5, 6, 9를 적용해서 승부를 봐야한다는 것인데 어떻게..?

1. 리뷰, 평가 데이터는 별점을 활용해서 감성분석을 통해 보여준다.
2. 각 학회, 전문의, 경력에 대한 다중분류모델을 만들어서 답러닝 시킨 후, 증상 입력시 분류가 나오게 한다.

- 해야하는 것

1. 노가다를 뛰어서 각 의사별 진료과목(전공과목), 학회 여부를 긁어오기
 2. 노가다를 뛰어서 각 의사별 전문내용관련 컬럼/기사(?) 긁기
-
3. 각 의사별 기사/컬럼 크롤링해서 키워드로 활용
 4. 의사별 리뷰 활용해서 키워드나 감성분석
 5. 각 병원 별 리뷰, 평가 활용해서 토픽 or 데이터로 활용
 6. 논문 데이터 의사별로 연결
 7. 각 '학회'를 분석 및 자연어처리해서 활용
 8. 각 병원의 개업일자가 오래되면 가산점으로 활용 or 이용자에게 보여줌