



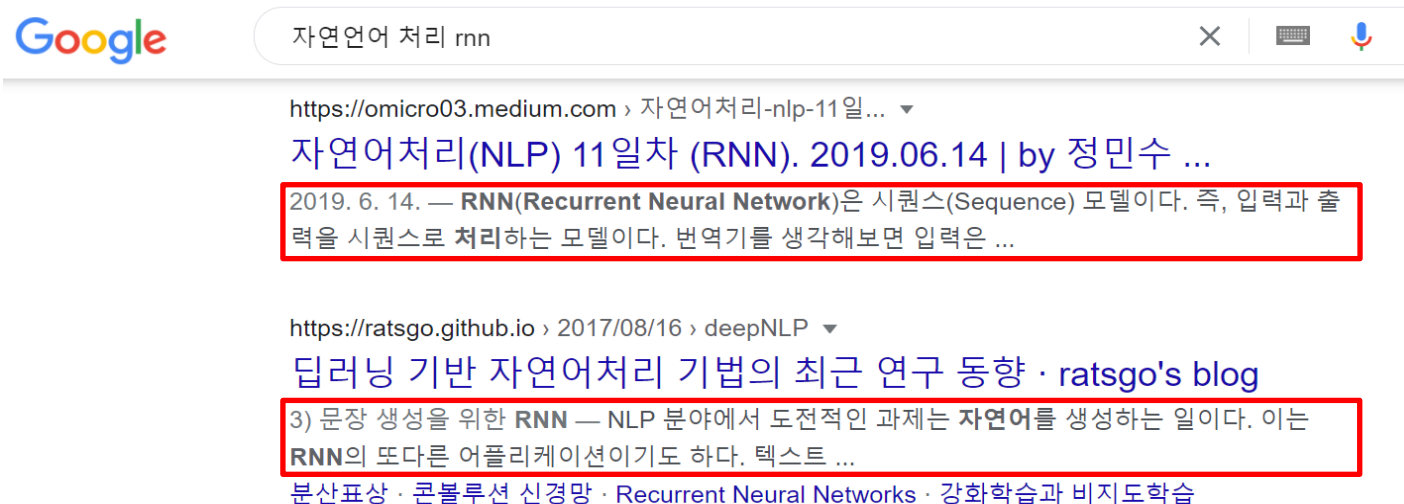
문서요약 시스템

4조(감사하는마음 → 최강팀)

김자은, 강용범, 손아름, 김경수

서론: 아이디어

- 일반적인 검색엔진
 - 검색 단어가 포함된 문장 1~2줄 출력
 - 따라서 해당 문서의 유효성 판단 없이 URL 클릭



아이디어

문서의 전반적인 내용을 한 눈에 파악할 수는 없을까?

서론: 프로젝트 소개

- **Trigram으로 요약**
 - 1개의 기사 scrawling
 - 해당 기사의 Trigram으로 문서 생성
- **RNN + POS 로 요약**
 - 1개의 기사 scrawling
 - 1개의 기사로만 학습
 - 해당 기사의 주요 단어 + 추가적인 품사정보 활용하여 자연스러운 문장을 기반으로 한 문서 생성

기대효과

전반적인 내용을 한 눈에 파악할 수 있도록, 기사 요약

서론: Quick Overview

===== Normal Sentence =====

하윤정의 의료세상 1억6400만 명이 감염되고 34만5000여 명이 사망한 코로나의 끝은 어디일까? 전문가들은 코로나19 종식은 불가능하다고 예측한다. 인류는 코로나와 공존하는 엔데믹 endemic 시대를 살아가는 지혜를 터득해야 한다. 마스크를 벗고 자유롭게 외출

백신 집단 면역 정부 세 때문
이상 국 발언 로 사망
수준 국민 연론 접 감염
한국 백신 언론 대하 중증 가능 도달 나타나
이상 반응

===== Generated by Trigram =====

백신 접종 다면 중증 환자 사망자 줄어들 방역 수준 조정 기대 코로나 19 자가 상대적으로 백신 접종 여행 발언 정부 근거 없다 반론면서 백신 접종 중요 이상 반응 자의 지원 지속 어야 다 집단 면역 가능성 60 이상에게 접종 결과 접종 부터 89 이상 감염 예방 효과
lenMaxSent : 64
sizeVocab : 418

===== Generated by RNN : Normal =====

백신 접종 후 사망 보도의 속도 경쟁으로 왜곡되다 일부 언론의 행태는 백신 접종률을 낮추는 결과로 나타나 수 있어 언론 스스로의 자율 규제도 필요하다. 코로나 19 자가 안전이 가기 때문이다. 이상 반응이 높아 이상하는 이상이고 정치인의 사망률
lenMaxSent : 65
sizeVocab : 418

===== Generated by RNN : Ending =====

백신 접종 후 집단 면역 달성을 위하여 사용하 수 있는 정책은 백신의 접종률을 높이는 길뿐이다. 코로나 19 감염 지원 도록 소급 적용도 가능하다. 이상 반응 감염 예방 효과가 확인하였고 치명률이다. 집단 면역 다 발언 환자 코에서 나로 여행길 막
lenMaxSent : 64
sizeVocab : 446

===== Generated by RNN : POS =====

백신 접종 후 사망 보도의 속도 경쟁으로 왜곡되다 일부 언론의 행태는 백신 접종률을 낮추는 결과로 나타나 수 있어 언론 스스로의 자율 규제도 필요하다. 이상 반응으로 사망하다 코로나를 접종하고 중증하어야 접종하였을 때 이득이 이상하다.

본론: (1) Trigram 시도

원문 (Cleaning)

사이언스카페 모더나 노바백스 화이자 백신 접종후 중화항체 많이 유도하고 예방효과 커 우리나라와 mou 체결한 노바백스 모더나 워싱턴열린 한미 백신 기업 파트너십 행사 에 노바백스 모더나 백신이 전시돼 있다.

전처리 (형태소분석 (어근추출) + String化)

사이언스 카페 모 더나 노바 백스 화 이자 백신 접종 후 중화 항체 많이 유도 하 고 예방 효과 커 어 우리나라 와 mou 체결 하 는 노바 백스 모 더나 워싱턴열리 는 한미 백신 기업 파트너십 행사 에 노바 백스 모 더나 백신 이 전시 되 어 있 다 .

Trigram

'백터 전달 통하', '보건 기구 who', '보다 중화 항체', '보다 지만 예방', '보다 항체 생성', '보세 최고 기자', '보여주 면역 대리', '분석 결과 백신', '분석 에서 항체', '붉은색 단백질 호흡기', '비교 어야 s다', '비드 알트 교수', '비레 코로나 백신', '

사전

['걸치 비접' : ['종자'], '결과 백신' : ['접종'], '결합 바이러스' : ['인체'], '결합 시키' : ['감염', '침투'], '결합 중화' : ['항체'], '경험 먼저' : ['결합'], '고르 주입' : ['s다'], '공격 느라' : ['면역력'], '과정 시간' : ['물론'], '과학 전문' : ['기자'], '과학자 러시아' : ['백신'].....]

Seed(생성한 사전을 Value의 개수로 정렬)

[('예방 효과', ['50', '60', 'who', '가장', '다고', '다는', '면역', '비교', '비레', '상관', '세로축', '시드니', '우리나라', '입증']), ('중화 항체', ['가장', '많이', '생성', '양은', '유도', '코로나', '형성']).....]

본론: (1) Generated by Trigram

예방 효과 세로축 mrna 1273 노바 백신 nvx cov 2373 이자
bnt 162 러시아 스푸트니크 rad 26 rad 백신 완치 보다 항체
생성 시키 인하 코로나 예방 효과 50 이상 으로 정하 중화
항체 코로나 예방 효과 who 기준 50 코로나 백신 체량 가로축
완치 대비 코로나 예방 효과 다는 의미 이자 백신 대규모 임상
시험 시간 비용 줄이'

결과

부자연스럽지만 나름 의미 있는 결과. 개선 HOW?

본론: (2) RNN 시도

- Seed 설정 - Trigram에서 많이 사용된 단어 2개

Dictionary Key("예방 효과")
에 Value 개수 많은 순으로 정
렬

```
[('예방 효과', ['50', '60', 'who', '가장', '다고',  
'다는', '면역', '비교', '비례', '상관', '세로축',  
'시드니', '우리나라', '입증']), ('중화 항체',  
['가장', '많이', '생성', '양은', '유도', '코로나',  
'형성']), ('코로나 백신', ['개발', '분석', '인체',  
'체량', '효능']), ('이자 백신', ['대규모',  
'스파이크', '인체', '접종']), ('노바 백신', ['nvx',  
'더나', '이자'])]
```

본론: (2) Generated by RNN: Normal

- RNN 결과

예방 효과는 who 기준 이 효과 50를 넘었다. 중화
항체를 더 많이 유도하고 코로나 예방 효과도 크었
다. 코로나 백신이 나오았다. 이자 백신 이 지만 았
다. 노바 백신 화이자 백신은 후 중화 항체 많이 유도
하고 예방 효과 커어

노란 글자는 사용자사
전으로 극복가능
이자 → 화이자

빨간색은 연구 필요

밑줄은 문법적으로 극복가능
나오+았+다 → 나왔다

본론: (3) Generated by RNN: Ending

- Ending
 - 새 문장 선택시 Seed는 Trigram에서 많이 사용 단어
 - 주로 문장의 처음에 사용되는 배제하는 데 의미

예방 효과 와 아 벡터 가 네 카 아 아 스트라 제 네 카 오
아 안 세 L 백신 보다 중화 항체 를 더 많이 유도 하 고
코로나 예방 효과 도 크 었 다 . 중화 항체 를 mrna L
있 는 것 이 밝혀지 었 다 . 코로나 백신 은 가 바이러스
를

본론: (4) Generated by RNN: Ending + Reset

- 학습

- 전처리시 종결어미(".", "?", "!")는 유지
- kkma.sentences() 활용
 - 종결어미 + 자체 판단
- 최대 문장길이를 학습, 따라서 입력 문장의 길이를 학습시킨 문장 길이 이내로 유지 필요

- 예측

- 생성되는 문장이 학습문장의 길이를 초과

예방 효과 와 백신 은 백신 이 코로나 완치 자의 20 에 불
과 가 은 바이러스 를 미국 는 유전자 하 인체 이 코로나
예방 효과를 가장 크 다 하 었 다 . 중화 항체 가
스파이크 더 이 처 백신 의 백신 은 독성 을 없애 는
바이러스 자체 를 주입'

본론: (5) Generated by RNN: Ending + Reset + Pos

• 품사정보 활용

- 학습/예측에 품사정보 활용
- 활용 예
 - 예방효과 : 명사+명사
 - 예방만이 : 명사+[보조사]+조사
 - 예방하는 : 명사+접(미)사+어미
 - 갑니다 : 용언+선어말어미+어말어미
 -
- 명사+명사(기존에는 무조건 출현 빈도 수)
- 명사+ 조사
- 형용사 + 명사
- 용언 + 관형사형 어미 등
- 다양한 seed 선택 방법 확장 가능

[' 사이언스-NNG 카페-NNG 모-NNG
더나-NNP 노바-NNG 백스-UN 화-NNG
이자-NNG 백신-NNG 접종-NNG 후-
NNG 중화-NNG 항체-NNG 많이-MAG
유도-NNG 하-XSV 고-ECE 예방-NNG
효과-NNG 커-VV 어-ECS 우리나라-NNG
와-JC mou-OL 체결-NNG 하-XSV L-
ETD 노바-NNG 백스-UN 모-NNG 더나-
NNP 워싱턴-NNG
.....열리-VV L-
ETD 한미-NNG 백신-NNG 기업-NNG
파트너십-NNG 행사-NNG 에-JKM 노바-
NNG 백스-UN 모-NNG 더나-NNP 백신-
NNG 이-JKS 전시-NNG 되-XSV 어-ECS
있-VXV 다-EFN .-SF',]

본론: (5) Generated by RNN: Ending + Reset + Pos

- 좋은 Seed

[('예방-nng 효과-nng', ['50-nr', '60-nr', 'who-ol', '가장-mag', '다고-efn', '다는-etd', '면역-nng', '비교-nng', '비례-nng', '상관-nng', '세로촉-nng', '시드니-nng', '우리나라-nng', '입증-nng']),
(('중화-nng 항체-nng', ['가장-mag', '많이-mag', '생성-nng', '양은-nng', '유도-nng', '코로나-nng', '형성-nng']),
(('코로나-nng 백신-nng', ['개발-nng', '분석-nng', '인체-nng', '체량-nng', '효능-nng']),
(('이자-nng 백신-nng', ['대규모-nng', '스파이크-nng', '인체-nng', '접종-nng']),
(('노바-nng 백신-un', ['nvx-ol', '더나-nnp', '이자-nng']),
(('백신-nng 접종-nng', ['에서-jkm', '중화-nng', '체량-nng

- 결과

'예방 효과 와 비례 하 었 다 . 중화 항체 를 더 많이 유도 하
고 코로나 예방 효과 도 면역 라고 가 있 는 것 으로 밝혀지 었
다 . 코로나 백신 의 혈 중 항 체량 가로촉 예방 효과 와 비례
하 었 다 . 이자 백신 의 mrna 가 많이 예방 효과 를 입증 하 '

- **모델 설계**

- Hidden Layer 추가
- Conv1D 등
- Node 최적화 등

- **모델 추가**

- LSTM
- Attention Mechanism 등

- **다른 형태소 분석기**

- 현재) 꼬꼬마 형태소 활용

- **그 외**

- 향상된 Seed Selector
- 향상된 Generator
- Context를 고려한 모델
(context free -> context sensitive)

- **효율적인 검색 포털**
 - URL을 클릭하기 이전에, 전반적인 내용 파악 가능하도록
- **각종 기사 및 보고서 요약 등**
 - 전반적인 내용을 한 눈에 파악할 수 있도록



감사합니다