**데이터 수집 과정**

다음 querys에 해당하는 키워드를 바탕으로 검색했을 때 나오는 네이버 뉴스 기사에 대한 정보를 수집하였습니다.

* querys = ['조현병', '정신분열+"정신분열"','정신분열+"정신분열"+"사건"','조현병+"사건"','정신분열+"정신분열"+"범행"','조현병+"범행"', '정신분열+"정신분열"+"범죄"','조현병+"범죄"', '정신분열+"정신분열"+"살인"','조현병+"살인"', '정신분열+"정신분열"+"살해"','조현병+"살해"', '정신분열+"정신분열"+"폭행"','조현병+"폭행"', '정신분열+"정신분열"+"방화"','조현병+"방화"']

데이터 수집 과정은 다음과 같습니다.

1. 전체 검색 키워드 ['조현병', '정신분열+"정신분열"']를 키워드로 검색했을 때 나오는 기사를 수집
2. 특정 검색 키워드: ‘사건, 범행, 범죄, 살인, 살해, 폭행, 방화’의 특정 키워드를 추가하여 검색했을 때 검색된 기사를 수집
3. 1번에 포함되지만 2번에는 포함되지 않은 기사 중, 조현병 관련 범죄 사건과 관련이 있는 기사를 최종 데이터에 추가함.

**데이터 변수 설명**

|  |  |
| --- | --- |
| **변수** | **설명** |
| title | 기사 제목 |
| date | 기사 날짜 |
| content | 기사 내용 |
| company | 언론사(방송3사, 종편 및 뉴스전문채널 : SBS, KBS, MBC, 채널A, JTBC, MBN, TV조선, 연합뉴스, 연합뉴스TV, YTN) |
| url | 기사 링크 |
| comment | 기사 댓글 내용 |
| comment\_num | 기사 댓글 수 |
| is\_murder | 살인 키워드 여부('정신분열+"정신분열"+"살인"','조현병+"살인"', '정신분열+"정신분열"+"살해"','조현병+"살해"' |

**데이터 요약**

json파일로 저장하였으며 총 데이터 수는 3286개임.