

# 축구선수 연관 단어

팀 명 정수환

팀 원 정수환

지도교수 이슬

멘토

## 개발 동기 및 목적

현재 대한민국 남자 축구 선수들이 해외 축구 리그에서 큰 활약을 하고 있다. 거기에 2022년 카타르 월드컵이 다가오면서 축구 선수들에 대한 관심이 많아지고 있다. 특히 월드컵 선발이 될 선수, 월드컵에서의 활약 등이 있다.



선수의 평가는 전문가들의 의견이 중요하지만 스포츠 경기는 어떻게 될지 모른다. 전문가들도 한 사람의 의견이다. 그래서 축구를 즐겨보는 사람들의 의견을 많이 모아 종합해보면 전문가보다 정확한 평가를 알 수 있을 것이라 예상하였다.



따라서 축구를 관심있게 보는 사람들은 경기 하이라이트에 댓글을 남겼을 것이고 각 선수에 대해 어떻게 평가하였는지 분석하면 전체적인 선수의 평가를 알 수 있을 것으로 본다.



## 개발 내용

공식 EPL 팀 Youtube 채널에 동영상 URL들을 crawl한다.

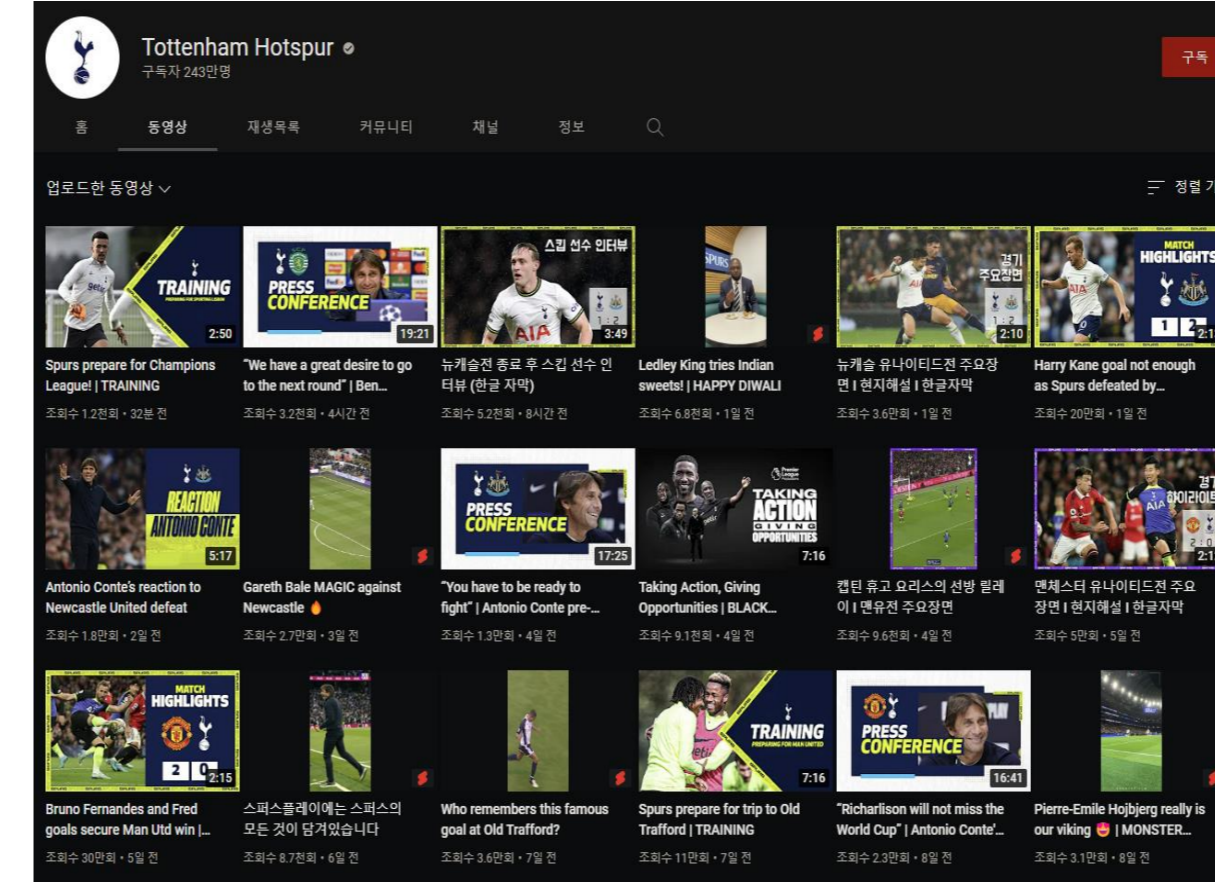
댓글이 많을 것으로 예상되는 하이라이트 영상 URL만 접속하여 댓글을 crawl한다.

각 하이라이트 영상에 대한 댓글을 모은 뒤, 키워드(검색하고 싶은 선수)가 있는 문장을 가져와 stopword를 제외하고 word count를 한다.

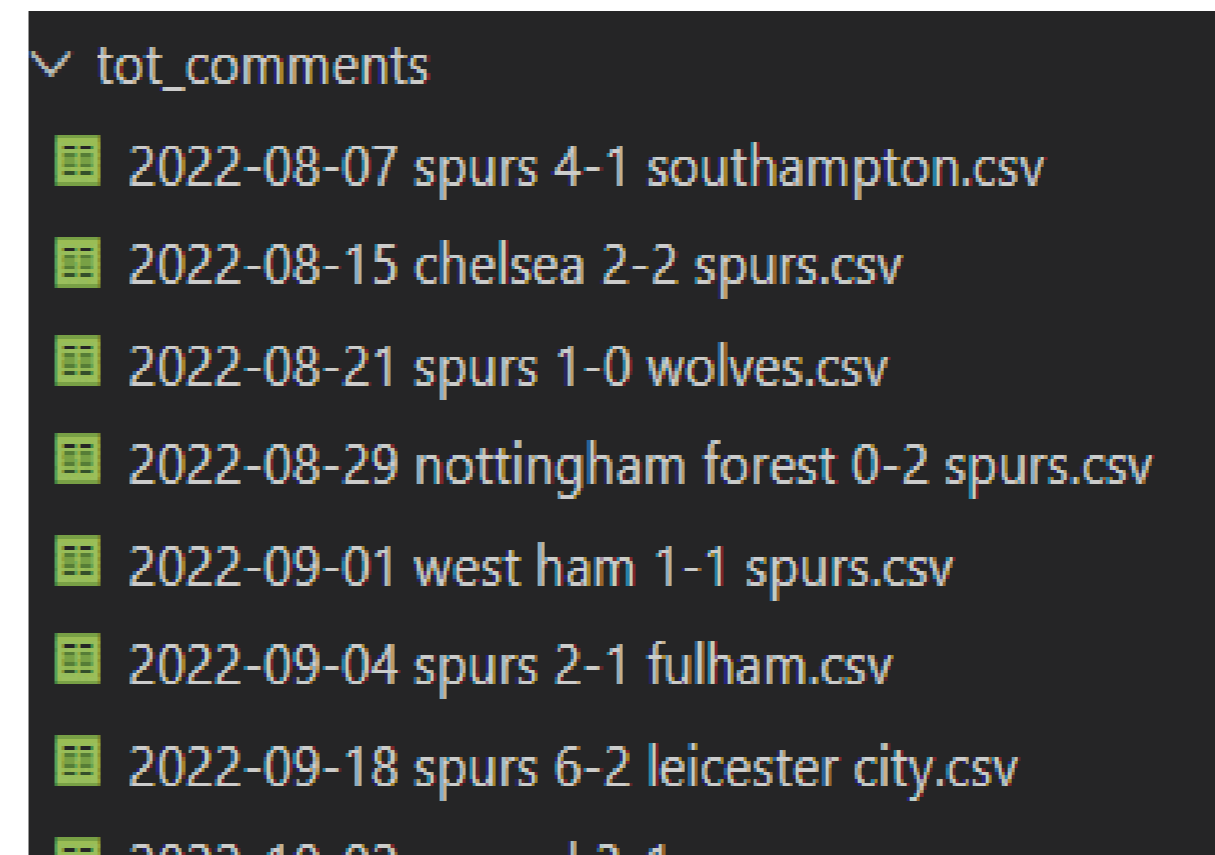
그 단어들 중 키워드를 포함한 frequent pair를 구하여 저장한다.

Flask로 만든 웹페이지에서 사용자가 키워드를 검색하면 키워드가 있는 frequent pair가 뜨게 된다.

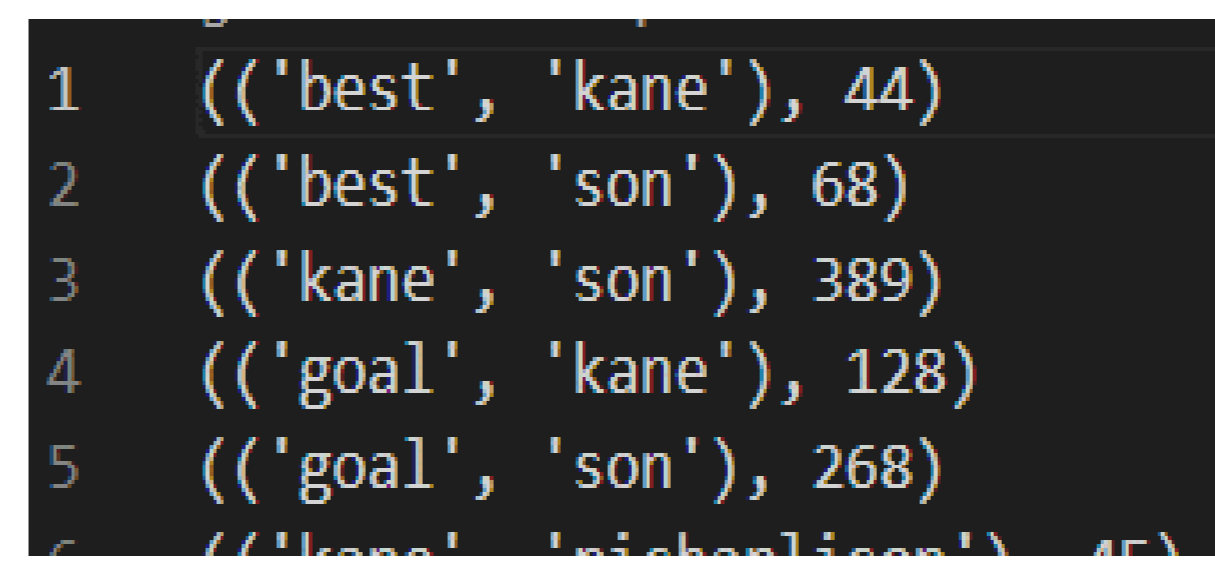
## 주요기술



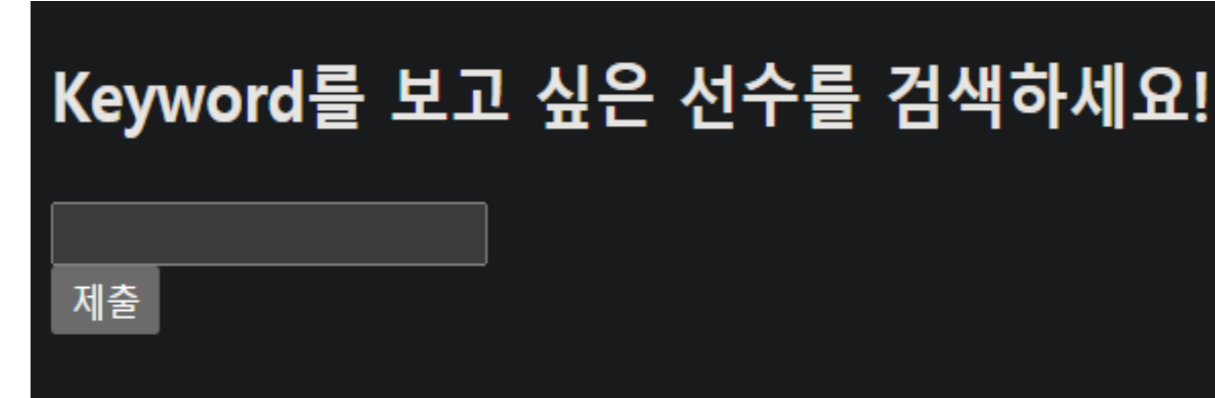
EPL 팀 공식 Youtube 채널 URL crawling



URL 중 경기 하이라이트 영상 댓글 crawling



댓글 word count 및 stopword 제거 후 Frequent pair 구하기



Flask로 웹페이지 구현

## 결과 및 분석

선수의 별명, 선수의 대한 형용사들, 관련 선수들 등 연관 단어를 가지고 사람들이 저 선수에 대해 어떻게 생각하는지 알 수 있다.

임의의 stopword가 필요하고 추후에 많은 경기 데이터를 모으고 비를 발전시키고 선수마다 이름 대신에 불리는 별명이 있기 때문에 그것도 포함하여 처리를 해주면 조금 더 선수와 관련된 단어를 많이 정확히 찾을 수 있을 것으로 예상된다.

kane선수의 Keyword.

```
[ 'son가 2050번', 'goal가 611번', 'harry가 607번', 'good가 283번', 'best가 263번', 'sonny가 230번', 'penalty가 206번', 'great가 173번', 'duo가 143번', 'better가 136번', 'top가 121번', 'striker가 120번', 'harry가 108번', 'assist가 106번', 'son가 97번', 'trophy가 89번', 'moura가 87번', 'kick가 86번', 'lucas가 86번', 'performance가 84번', 'heung가 84번', 'kulu가 81번', 'romero가 71번', 'goal가 70번', 'kulusevski가 69번', 'score가 66번', 'grealish가 65번', 'selfish가 57번', 'one가 56번', 'gil가 55번', 'dier가 55번', 'defender가 52번', 'record가 51번', 'perisic가 48번', 'lloris가 48번', 'dele가 47번', 'levy가 47번', 'richarlison가 45번', 'happy가 45번', 'defense가 44번', 'amazing가 44번', 'coys가 43번', 'champion가 43번', 'shot가 43번', 'need가 43번', 'level가 42번', 'messi가 40번', 'point가 40번', 'scorer가 40번', 'pep가 39번', 'nothing가 39번', 'attack가 39번', 'day가 38번', 'midfield가 37번', 'line가 36번', 'legend가 36번', 'please가 36번', 'salah가 36번', 'win가 35번', 'wink가 35번', 'important가 35번', 'moment가 35번', 'half가 35번', 'nice가 34번', 'bad가 34번', 'show가 34번', 'strong가 34번', 'passing가 34번', 'bergwijn가 33번', 'role가 33번', 'problem가 33번', 'transfer가 33번', 'header가 33번', 'chelsea가 32번', 'combination가 32번', 'ndombele가 32번', 'offsides가 32번', 'pitch가 31번', 'long가 31번', 'pass가 31번', 'trio가 31번', 'hope가 31번', 'fantastic가 31번', 'doherty가 31번', 'money가 30번', 'haaland가 30번', 'england가 30번', 'perfect가 30번', 'bench가 30번', 'minute가 30번', 'history가 30번' ]
```

등장하였습니다.