



백엔드

💡 해야할 것

🔄 CI/CD

기본 세팅

- 도커
- Nginx
 - CertBot
 - 경로 분기
- DB 들,,

Logging

- 컨테이너 내부 로그들
- Mattermost 연동

젠킨스

크롤링 서버 (Flask)

- 스케줄러 (주기적으로)

클러스터와 API 서버 통신

📰 데이터 수집 및 전처리

크롤링 or 수집

- 유튜브

Python 빠른 시작 | YouTube Data API | Google Developers
<https://developers.google.com/youtube/v3/quickstart/python?hl=ko>

- 뉴스 기사
- 트위터

<https://github.com/twintproject/twint>

Analyzing Tweets with NLP in minutes with Spark, Optimus and
 Social media has been gold for studying the way people communicate and behave, in this article I'll show you the easiest way of analyzing...
<https://towardsdatascience.com/analyzing-tweets-with-nlp-in-minutes-spark-optimus-and-twint-a0c96084995f>

- 나무위키
 - 아이돌 정보 자동으로

▼ 데이터 스키마 정의

▼ 뉴스

- 수집
 - : 아래처럼 만들어서 던져준다

```
{
  제목: 삼성 sds...,
  본문(전문): ~~,
  아이돌이름(번호):
  작성일:
  언론사:
```

링크:

마지막 수정일자:

작성자:

수집 날짜:

}

- 분석

list → dataframe → Spark 한 행씩 분산처리

1. 일단.. 문장별로 쪼개는거 해서 긍부정지수 산출
2. 해당 문장 쪼개서 단어별 카운팅

{

제목: 삼성 sdss...,

본문(앞에 짧게 일부만): ~~~,

아이돌이름(번호): 1,

작성날짜

언론사

긍부정지수: #

?문장수?:

키워드 종류(타입10): {5: [], 4: [], ...} # 워드카운트

}

▼ 트위터

```
Index(['conversation_id', 'created_at', 'date', 'day', 'hashtags', 'hour', 'id', 'link', 'location', 'name', 'near', 'nlikes', 'nreplies', 'nretweets', 'place', 'profile_image_url', 'quote_url', 'retweet', 'search', 'timezone', 'tweet', 'user_id', 'user_id_str', 'username'], dtype='object')
```

{

제목:

본문(전문):

글링크?

영상 있으면 링크:

수집 날짜: 및

좋아요수:

댓글수:

리트윗수:

view수:

태그리스트:

작성?수정?시간:

}

▼ 유튜브

▼ 데이터

Videos | YouTube Data API | Google Developers

<https://developers.google.com/youtube/v3/docs/videos?hl=ko>

```
{
  "kind": "youtube#video",
  "etag": etag,
  "id": string,
  "snippet": {
    "publishedAt": datetime,
    "channelId": string,
    "title": string,
    "description": string,
    "thumbnails": {
      (key): {
        "url": string,
        "width": unsigned integer,
        "height": unsigned integer
      }
    },
    "channelTitle": string,
    "tags": [
      string
    ],
    "categoryId": string
  },
  "contentDetails": {
    "duration": string,
    "dimension": string,
    "definition": string,
    "caption": string,
    "licensedContent": boolean,
    "regionRestriction": {
      "allowed": [
        string
      ],
      "blocked": [
        string
      ]
    }
  },
  "contentRating": {
    "mpaaRating": string,
    "tvpgRating": string,
    "bbfcRating": string,
    "chvrsRating": string,
    "eirinRating": string,
    "cbfcRating": string,
    "fmocRating": string,
    "icaaRating": string,
    "acbRating": string,
    "oflcRating": string,
    "fskRating": string,
    "kmrBRating": string,
    "djctqRating": string,
    "russiaRating": string,
    "rtcRating": string,
    "ytRating": string
  }
},
  "status": {
    "uploadStatus": string,
    "failureReason": string,
    "rejectionReason": string,
    "privacyStatus": string,
    "license": string,
  }
```

```

        "embeddable": boolean,
        "publicStatsViewable": boolean
    },
    "statistics": {
        "viewCount": unsigned long,
        "likeCount": unsigned long,
        "dislikeCount": unsigned long,
        "favoriteCount": unsigned long,
        "commentCount": unsigned long
    },
    "player": {
        "embedHtml": string
    },
    "topicDetails": {
        "topicIds": [
            string
        ],
        "relevantTopicIds": [
            string
        ]
    },
    "recordingDetails": {
        "locationDescription": string,
        "location": {
            "latitude": double,
            "longitude": double,
            "altitude": double
        },
        "recordingDate": datetime
    },
    "fileDetails": {
        "fileName": string,
        "fileSize": unsigned long,
        "fileType": string,
        "container": string,
        "videoStreams": [
            {
                "widthPixels": unsigned integer,
                "heightPixels": unsigned integer,
                "frameRateFps": double,
                "aspectRatio": double,
                "codec": string,
                "bitrateBps": unsigned long,
                "rotation": string,
                "vendor": string
            }
        ],
        "audioStreams": [
            {
                "channelCount": unsigned integer,
                "codec": string,
                "bitrateBps": unsigned long,
                "vendor": string
            }
        ],
        "durationMs": unsigned long,
        "bitrateBps": unsigned long,
        "recordingLocation": {
            "latitude": double,
            "longitude": double,
            "altitude": double
        },
        "creationTime": string
    },
    "processingDetails": {
        "processingStatus": string,
        "processingProgress": {
            "partsTotal": unsigned long,
            "partsProcessed": unsigned long,
            "timeLeftMs": unsigned long
        },
        "processingFailureReason": string,
        "fileDetailsAvailability": string,
        "processingIssuesAvailability": string,
        "tagSuggestionsAvailability": string,
        "editorSuggestionsAvailability": string,
        "thumbnailsAvailability": string
    }
}

```

```

},
"suggestions": {
  "processingErrors": [
    string
  ],
  "processingWarnings": [
    string
  ],
  "processingHints": [
    string
  ],
  "tagSuggestions": [
    {
      "tag": string,
      "categoryRestricts": [
        string
      ]
    }
  ],
  "editorSuggestions": [
    string
  ]
}
}

```

API 개발 (Spring)

회원 (소셜 로그인)

- 카카오
- 스프링 시큐리티 (프론트와 협업)

기본 세팅

- 프로젝트 생성
- DB 들과 연결
- Swagger3

데이터 분석

감성분석(긍부정)

- 유튜브
- 트위터
- 뉴스 기사

워드 카운트(Word Count)

- 뉴스 기사

각종 수치 종합 계산

- 일단 데이터 처리가 되어야...

분석 결과 Spring에 보내기

- 어떤 형식으로 보내줘야 하지?

추가적인 모델 서치,,

할 일 목록

Aa 제목	≡ 귀걸	≡ 주희	≡ 현민
<u>Untitled</u>			
<u>Untitled</u>			
<u>Untitled</u>			

참고 자료

정리

AWS 접속하기

링크