

05 PJT

키워드 검색량 분석을 위한 데이터 수집

챕터의 포인트

- [도전] 키워드 검색량 분석을 위한 데이터 수집
- 제출

키워드 검색량 분석을 위한 데이터 수집

| 공통 요구사항

- 구글 검색 엔진을 활용하여 검색 결과에 따른 트렌드 분석 애플리케이션을 구현합니다.
 - 검색 결과 페이지의 “검색결과 개수” 를 활용합니다.
- Django 프로젝트의 이름은 `mypjt`, 앱 이름은 `trends` 로 지정합니다.
- `.gitignore` 파일을 추가하여 불필요한 파일 및 폴더는 제출하지 않도록 합니다.
- 명시된 요구사항 이외에는 자유롭게 작성해도 무관합니다.

| Model

- 정의할 모델 클래스 목록
 - Keyword
 - Trend

A. Keyword

- 정의할 모델 클래스의 이름은 Keyword 이며, 다음과 같은 정보를 저장합니다.

필드명	데이터 유형	역할
name	text	검색할 키워드명
created_at	Date	추가된 날짜

| B. Trend

- 정의할 모델 클래스의 이름은 Trend 이며, 다음과 같은 정보를 저장합니다.

필드명	데이터 유형	역할
name	text	검색을 수행한 키워드명
result	integer	검색 결과 수
search_period	text	검색 기간
created_at	Date	추가된 날짜

| URL

- trends 앱은 다음 URL 요청에 맞는 역할을 가집니다.

URL 패턴	역할
/trends/keyword/	분석을 원하는 키워드 입력 및 추가
/trends/keyword/<int:pk>	키워드 삭제
/trends/crawling/	크롤링 수행 및 결과 개수 출력
/trends/crawling/histogram/	크롤링 수행 및 결과 개수 막대 그래프로 출력
/trends/crawling/advanced/	지난 1년을 기준으로 크롤링 수행 및 결과 개수 막대 그래프 출력

| View

- trends 앱은 다음 역할을 가지는 view 함수를 가집니다.

View Method	역할
keyword	키워드 저장 및 keyword.html 렌더링
keyword_detail	키워드 삭제 및 keyword.html 로 리다이렉션
crawling	크롤링 수행 및 crawling.html 렌더링
crawling_histogram	크롤링 수행 후 수행 결과 막대 그래프 생성 및 crawling_histogram.html 렌더링
crawling_advanced	지난 1년을 기준으로 크롤링 수행 후 수행 결과 막대 그래프 생성 및 crawling_advanced.html 렌더링

| Templates

- 공유 템플릿 파일
 - A. base.html
- Trends 앱은 다음과 같은 템플릿 파일들을 가집니다
 - B. keyword.html
 - C. crawling.html
 - D. crawling_histogram.html
 - E. crawling_advaned.html

| A. base.html

- 공통 부모 템플릿
 - 모든 템플릿 파일은 base.html 을 상속받아 사용합니다.
 - 다른 파일 템플릿 경로로 이동할 수 있는 링크들을 출력합니다.
- 출력결과 예시

트렌드 분석 애플리케이션

[키워드 추가](#) | [크롤링](#) | [히스토그램](#) | [지난 1년 크롤링](#) |

| B. keyword.html

- 검색하고자 하는 키워드를 추가 및 삭제할 수 있도록 구성합니다.
- 생성하기 및 삭제하기 버튼을 통해, Keyword 테이블에 데이터를 저장 및 삭제하도록 구성합니다.
- 출력 결과 예시

트렌드 분석 애플리케이션

[키워드 추가](#) | [크롤링](#) | [히스토그램](#) | [지난 1년 크롤링](#) |

키워드 추가

키워드:

분석을 원하는 키워드 목록

1번째 키워드 - python

2번째 키워드 - Vue

| C. crawling.html

- Keyword 테이블에 저장된 키워드들을 활용하여 크롬 검색 결과 페이지 크롤링을 수행합니다.
- 페이지의 정보 중 “검색 결과 개수” 를 추출하여 Trend 테이블에 저장합니다.
 - 저장 시 검색 기간(search_period)을 “all” 로 저장합니다.
- 저장 시 이미 저장되어 있는 키워드라면, 새로 생성하지 않고 검색 결과 개수를 변경합니다.
- 출력 결과 예시

트렌드 분석 애플리케이션

[키워드 추가](#) | [크롤링](#) | [히스토그램](#) | [지난 1년 크롤링](#) |

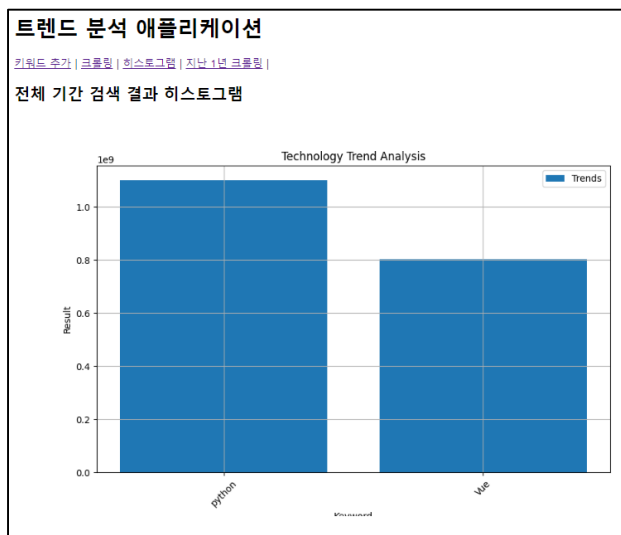
크롤링 기초 - 전체 기간 검색 결과

검색결과: python - 1100000000개 / 검색일자: 2023-05-24

검색결과: Vue - 802000000개 / 검색일자: 2023-05-24

| D. crawling_histogram.html

- 전체 기간 검색 결과를 이용하여 막대 그래프를 출력합니다.
- 크롤링을 다시 진행하지 않고, Trend 테이블에 저장된 데이터를 활용합니다.
- 출력 결과 예시



| E. crawling_advanced.html(1/3)

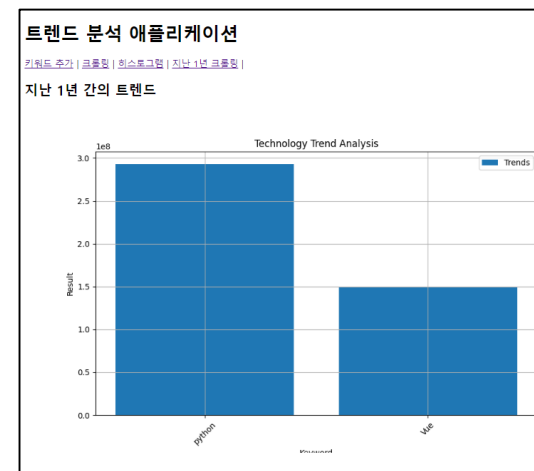
- 검색 결과 페이지 중 “지난 1년” 을 기준으로 필터링하여 크롤링을 수행합니다.
- [힌트] 크롬 페이지의 도구 - 검색 기간을 설정하며, URL 의 변화를 확인합니다.



google.com/search?q=python&source=Int&tbs=qdr:y&sa=X&ved=2ahUKEwiYks-u_Iz_AhXEC94KHV8QA84QpwV6BAgCEB0&biw=1920&bih=1007&dpr=1

| E. crawling_advanced.html(2/3)

- 분석한 URL 및 Keyword 테이블에 저장된 키워드들을 활용하여 크롤링을 수행합니다.
- 페이지의 정보 중 “검색 결과 개수” 를 추출하여 Trend 테이블에 저장합니다.
 - 저장 시 검색 기간(search_period)을 “year” 로 저장합니다.
- 저장 시 이미 저장되어 있는 키워드라면, 새로 생성하지 않고 검색 결과 개수를 변경합니다.
- 저장된 데이터를 활용하여 막대 그래프로 출력합니다.



E. crawling_advanced.html(3/3)

- 저장 및 출력 결과 예시
 - trends_trend 테이블

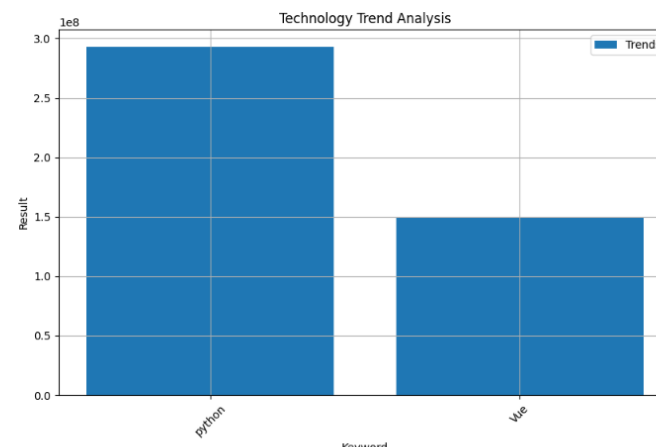
id	name	result	search_period	created_at
1	python	293000000	year	2023-05-24 02:42:29.318699
2	python	110000000	all	2023-05-24 02:42:46.766508
3	Vue	802000000	all	2023-05-24 02:46:09.656709
4	Vue	150000000	year	2023-05-24 02:46:28.243083

- 출력 결과 예시

트렌드 분석 애플리케이션

[키워드 추가](#) | [크롤링](#) | [히스토그램](#) | [지난 1년 크롤링](#)

지난 1년 간의 트렌드



제출

| 제출 시 주의사항

- 제출기한은 금일 18시까지입니다. 제출기한을 지켜 주시기 바랍니다.
- 반드시 README.md 파일에 단계별로 구현 과정 중 학습한 내용, 어려웠던 부분, 새로 배운 것들 및 느낀 점 등을 상세히 기록하여 제출합니다.
 - 단순히 완성된 코드만을 나열하지 않습니다.
- 위에 명시된 요구사항은 최소 조건이며, 추가 개발을 자유롭게 진행할 수 있습니다.
- <https://lab.ssafy.com/> 에 프로젝트를 생성하고 제출합니다.
 - 프로젝트 이름은 '프로젝트 번호 + pjt' 로 지정합니다. (ex. 05_pjt)
- 반드시 각 반 담당 교수님을 Maintainer 로 설정해야 합니다.