

Khrono System

도서 - 기사 일치율 계산 시스템

개요 (Overview)

카테고리



연도



Khrono System

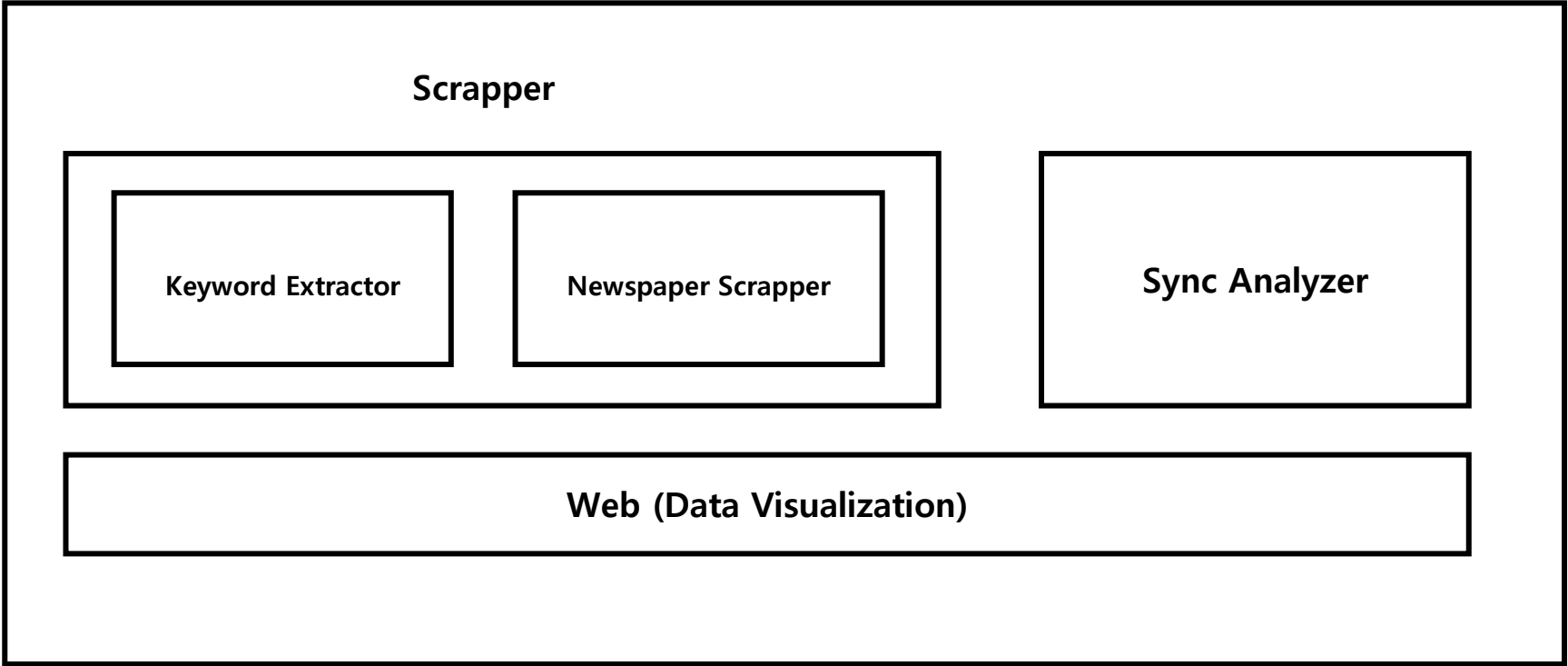


일치율 그래프



관련 뉴스

구성 요소 (System Design)





Component Keyword Extractor : 예시

```
PS C:\Users\KID\Desktop\Development\Project (Doggy)\BookDataScraper> python test.py
# [연도] [월_시작] [월_종료]
> 2023 01 10

# [SCIENCE / ECONOMY]
> SCIENCE

Data Scrappign Completed!
```

 AnalyzedKeywords.csv

 TotalBookData.csv

AnalyzedKeyword.csv

```
,YEAR,MONTH,KEYWORD
0,2023,1,과학
1,2023,1,세상
2,2023,1,생각
3,2023,1,인간
4,2023,1,우주
5,2023,1,지식
6,2023,1,지혜
7,2023,1,진화
8,2023,1,지구
9,2023,1,생명
10,2023,1,과학자
11,2023,1,콘서트
12,2023,1,뇌과학
13,2023,1,유전자
14,2023,1,전화
```

TotalBookData.csv

```
,ISBN,YEAR,MONTH,KEYWORD
0,9791160560510,2023,1,발자국
1,9791160560510,2023,1,인간
2,9791160560510,2023,1,세상
3,9791160560510,2023,1,생각
4,9791160560510,2023,1,강연
5,9791160560510,2023,1,통찰
6,9791160560510,2023,1,정재승
7,9791160560510,2023,1,과학
8,9791160560510,2023,1,지식
9,9791160560510,2023,1,지혜
10,9791160560510,2023,1,발자국
11,9791160560510,2023,1,인간
12,9791160560510,2023,1,세상
13,9791160560510,2023,1,생각
14,9791160560510,2023,1,강연
15,9791160560510,2023,1,통찰
16,9791160560510,2023,1,정재승
17,9791160560510,2023,1,과학
18,9791160560510,2023,1,지식
19,9791160560510,2023,1,지혜
```

Component Keyword Extractor

⚙️ 기능 : 공공 데이터와 교보문고에서 인기 도서와 키워드 데이터를 추출한다.

입력 : 데이터를 받을 날짜의 범위 (연도, 월), 카테고리 (경제 / 과학)

출력 : 키워드 데이터, 월별 인기 도서의 키워드 데이터

📄 입력 형식 : [받아올 연도], [시작_월; 범위의 시작], [종료_월; 범위의 끝], [카테고리; "ECONOMY" or "SCIENCE"]

📄 출력 형식 : AnalyzedKeywords.csv (월별 인기 키워드), TotalBookData.csv (월별 인기도서의 키워드)

```
PS C:\Users\KID\Desktop\Development\Project (Doggy)\BookDataScraper> python test.py
# [연도] [월_시작] [월_종료]
> 2023 01 10

# [SCIENCE / ECONOMY]
> SCIENCE

Data Scraapign Completed!
```

Component Keyword Extractor : 시연

```
PS C:\Users\KID\Desktop\Development\Project (Doggy)\BookDataScraper> python test.py
# [연도] [월_시작] [월_종료]
> 2023 01 10

# [SCIENCE / ECONOMY]
> SCIENCE

Data Scrappign Completed!


AnalyzedKeywords.csv
TotalBookData.csv
```

Component News Scraper : 예시

```
PS C:\Users\KID\Desktop\Development\Project (Doggy)\BookDataScraper> python test.py
# [연도] [월_시작] [월_종료]
> 2023 01 10

# [SCIENCE / ECONOMY]
> SCIENCE

News Scroppign Completed!
```

 **news_economy.csv**

News_economy.csv

,YEAR,MONTH,TITLE,CONTEXT
0,2023,11,'LG 29년만에 우승' 축포 왔지만...증권사를 목표주가 '시큰둥' 왜?, LG 트윈스가 2023 KBO리그 통합우승(정규시즌·한국시리즈 우승)을
19일 증권정보업체 Fn가이드에 따르면 지난달 18일부터 지난17일까지 한 달 동안에만 한국투자증권, 삼성증권, KB증권 등 17개 증권사가 LG에너
증권사들이 LG에너지솔루션에 대한 부정적인 평가를 연달아 내놓고 있는 건 최근 고객사들의 수요 전망이 부정적이기 때문이다. 최근 제너럴모터
이에 따라 지난 9월 8500억원 선이던 LG에너지솔루션의 4분기 영업이익의 컨센서스도 이날 기준 6600억원 대로 낮아졌다. 불과 두 달여만에 2000
LG화학도 지난 한 달 사이 12개 증권사의 목표주가 하향 러쉬가 이어졌다. LG에너지솔루션과 마찬가지로 전기차 수요 둔화가 주가의 발목을 잡은
이 기간 삼성증권, 메리츠증권 등 3곳은 LG전자의 목표주가가 낮춰 잡았다. 4분기에도 세트 수요 개선이 불확실한 점이 우려로 꼽히면서다. 경쟁
LG(1곳), LG이노텍(9곳), LG생활건강(13곳), LG디스플레이(4곳), LG유플러스(8곳), LG헬로비전(4곳)도 목표가 줄하향을 피해 가지 못했다.
과거 황제주 자리를 차지했던 LG생활건강의 목표가를 하향한 증권사도 13곳에서 나왔다. 중국 의존도가 절대적인 점이 주가를 짓누르고 있다는
광군제 마케팅도 축소되며 4분기에도 LG생활건강의 실적이 그리 좋지 못할 것이라는 관측까지 나온다.
박현진 신한투자증권 연구원은 "이외 기여가 가장 높은 화장품 실적 부진은 면세 부진과 중국 로컬 실적 성장 부재에 기인한다"며 "중국 내 이커"

※ 실제 데이터에는 고유 ID를 추가

Component News Scraper

⚙️ 기능 : 네이버 뉴스에서 특정 카테고리의 뉴스를 수집한다.

입력 : 데이터를 받을 날짜의 범위 (연도, 월), 카테고리 (경제 / 과학)

출력 : 지정된 카테고리의 날짜 범위 내에서 수집된 뉴스

📄 입력 형식 : [받아올 연도], [시작_월; 범위의 시작], [종료_월; 범위의 끝], [카테고리; "ECONOMY" or "SCIENCE"]

📄 출력 형식 : news_[category_name].csv (지정된 카테고리의 날짜 범위 내에서 수집된 뉴스)

```
PS C:\Users\KID\Desktop\Development\Project (Doggy)\BookDataScraper> python test.py
# [연도] [월_시작] [월_종료]
> 2023 01 10

# [SCIENCE / ECONOMY]
> SCIENCE

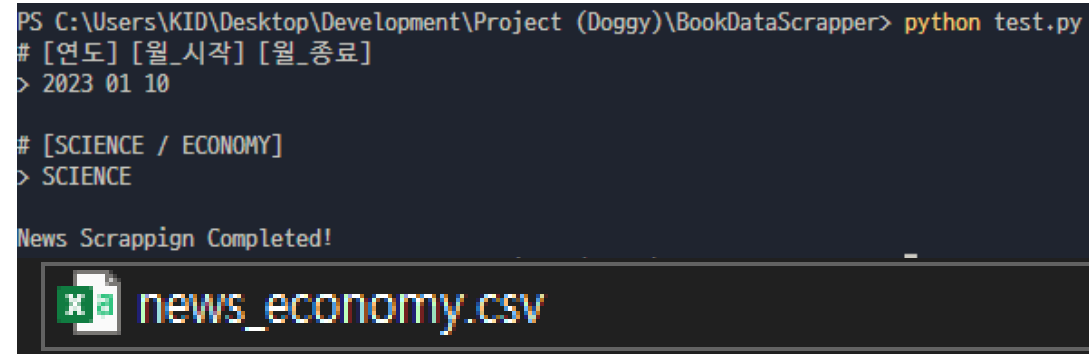
Data Scraapign Completed!
```


Component News Scraper : 시연

```
PS C:\Users\KID\Desktop\Development\Project (Doggy)\BookDataScraper> python test.py
# [연도] [월_시작] [월_종료]
> 2023 01 10

# [SCIENCE / ECONOMY]
> SCIENCE

News Scrappign Completed!
```



Component Sync Analyzer : 예시

```
PS C:\Users\KID\Desktop\Development\Project (Doggy)\BookDataScraper> python test.py
# [키워드 데이터 절대 주소]
> C:\Users\KID\Desktop\Development\Project (Doggy)\BookDataScraper\keywords.csv

# [기사 데이터 절대 주소]
> C:\Users\KID\Desktop\Development\Project (Doggy)\BookDataScraper\news.csv

Sync Analyzed!

sync_per_news.csv
monthly_sync.csv
```



sync_per_news.csv

	ID	PERCENTAGE
0	0005217187	0.32
1	0005217376	0.15

Monthly_sync.csv

	YEAR	MONTH	PERCENTAGE
0	2023	01	0.15
1	2023	02	0.32

Component Sync Analyzer

⚙️ 기능 : 키워드와 기사를 대조하여 일치율을 계산하여 이를 파일 형태로 출력한다.
입력 : (KeywordExtractor에서 출력된 AnalyzedKeywords.CSV), (News Scrapper에서 출력된 CSV 파일)
출력 : 월별 총 일치율, 기사별 일치율

📄 입력 형식 : 각 CSV 파일이 존재하는 절대 경로
📄 출력 형식 : 2개의 CSV 파일 (월별 총 일치율, 기사별 일치율)

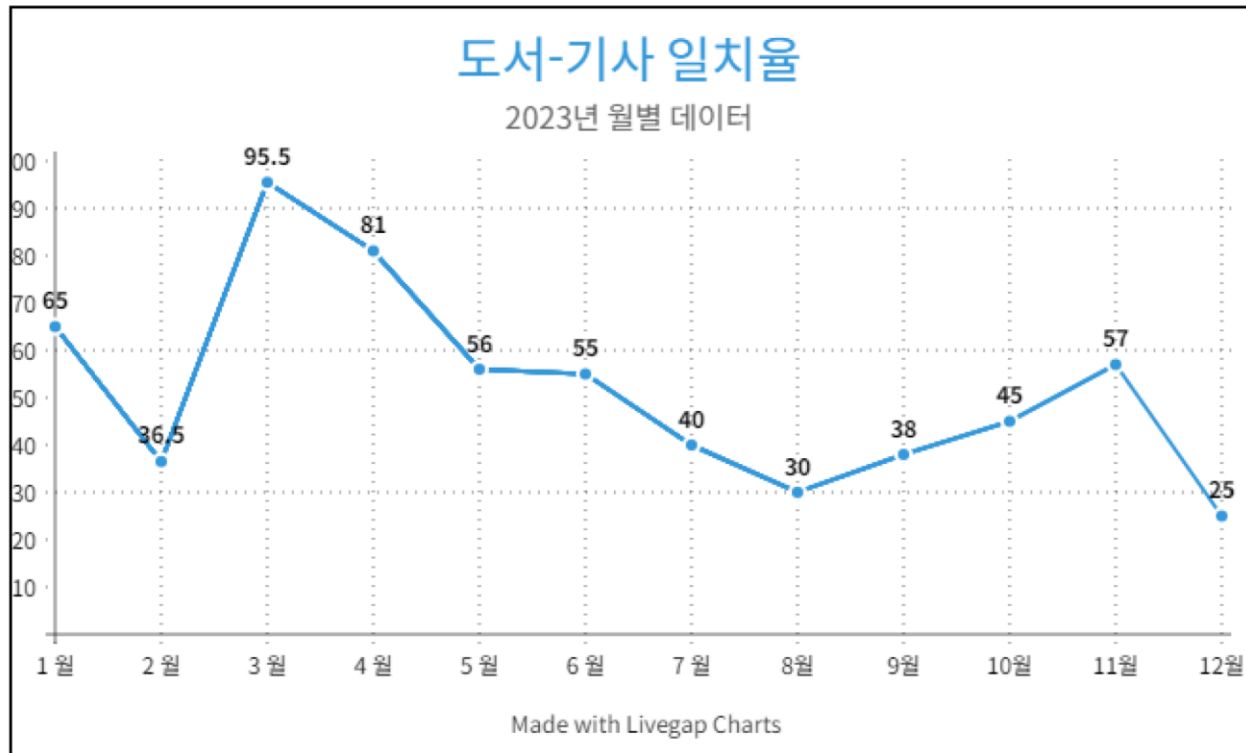
```
PS C:\Users\KID\Desktop\Development\Project (Doggy)\BookDataScrapper> python test.py
# [키워드 데이터 절대 주소]
> C:\Users\KID\Desktop\Development\Project (Doggy)\BookDataScrapper\keywords.csv

# [기사 데이터 절대 주소]
> C:\Users\KID\Desktop\Development\Project (Doggy)\BookDataScrapper\news.csv

Sync Analyzed!
```



2023년



NEWS

화이트 큐브 탈피한 대구 ...

연말정산, 미리 준비하세요 ...

범어도서관, '도서관 바깥 ...

Component Web

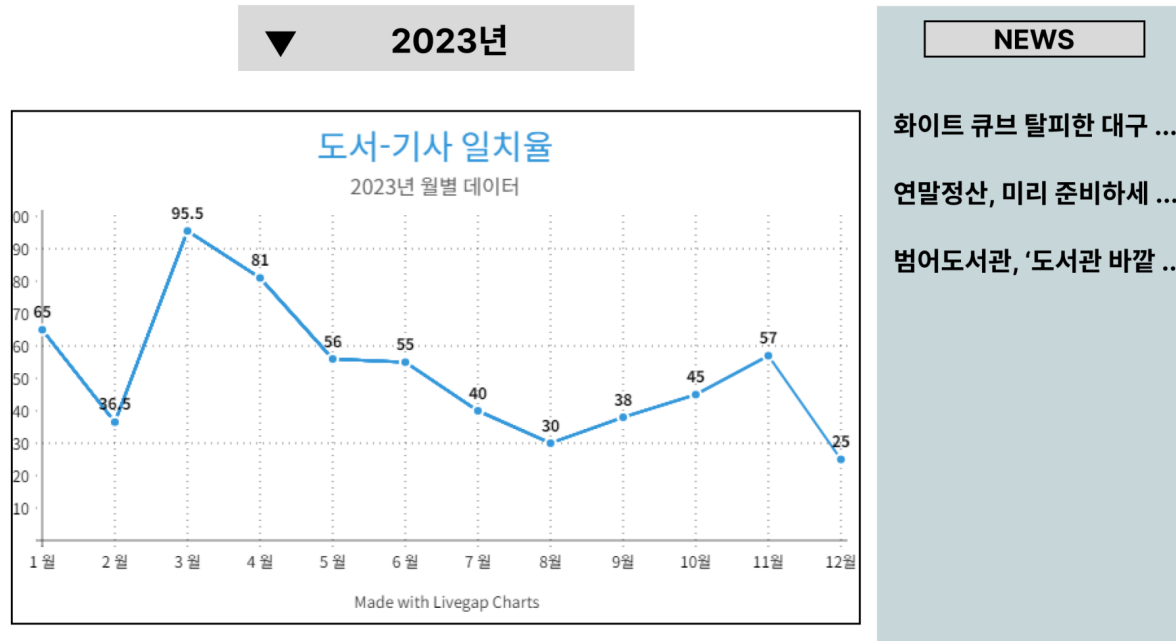
⚙️ 기능 : 월별 일치율 데이터를 보여주고 일치율이 높은 관련 뉴스를 보여준다.

입력 : 연도 (제시된 범위 중)

출력 : 월별 일치율, 일치율이 높은 관련 뉴스

📋 입력 형식 : 리스트에 제시된 연도 중 하나를 선택

📋 출력 형식 : 월별 일치율 그래프, 사이드 부분에 리스트 형태로 뉴스를 제공



변경된 사항, 발생한 문제

발생한 문제

1. 중심적인 지역 언론이 적고 상당히 지엽적인 정보를 포함하고 있다.
2. 카테고리의 제한이 없으면 데이터의 방향성이 중구난방이다.
3. 하루에 발생하는 뉴스의 양은 너무 많다, 어디까지 수집할 것인가?

변경된 점

1. 지역 단위가 아니라 전국 단위로 변경 (의미있는 결과를 기대)
2. 카테고리를 경제와 과학으로 제한 (명확한 데이터의 방향성)
3. 네이버 뉴스에서 특정 카테고리의 모든 뉴스 수집 (정확성을 위해서)
4. 더 나은 시스템을 위해서 시스템 재구조화

앞으로의 구현 계획

1. 수집된 데이터를 기반으로 일치율을 계산할 컴포넌트 제작
2. 수집된 일치율 데이터를 통해서 웹 사이트 구현하기