



지역별 여행 정보 추천을 위한
여행 데이터 분석

노성진

개요

여행 데이터 분석

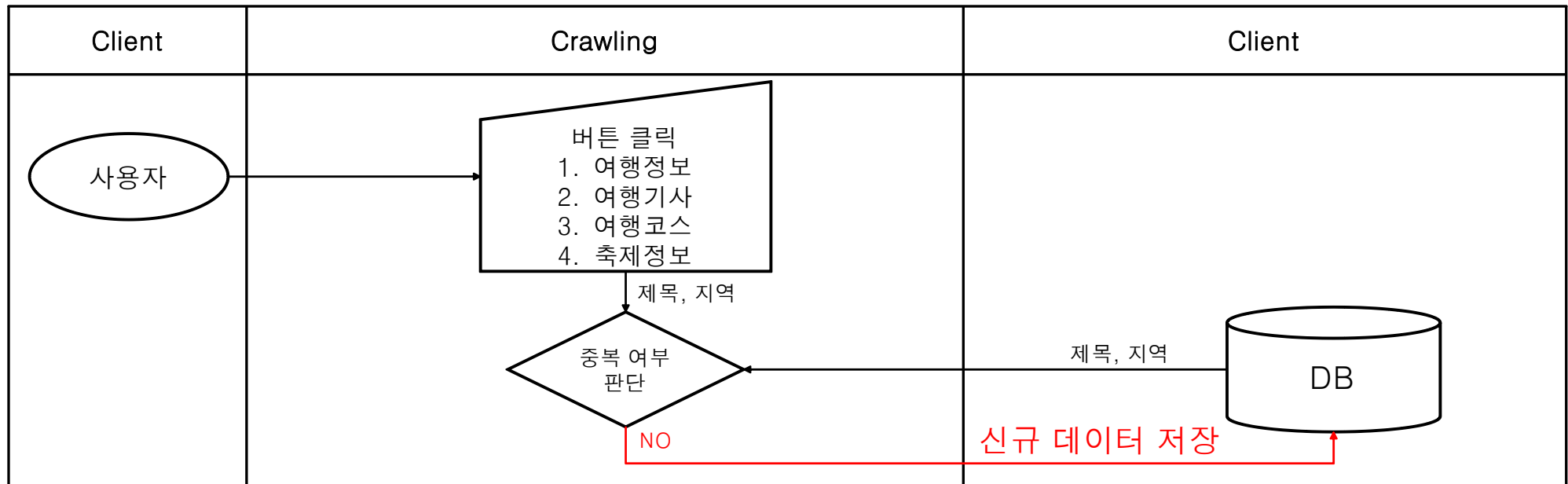
- 사이트 선정 이유
 - 지역별 로컬 여행 정보
 - 각 코스마다 정보와 인근 여행지
 - 날짜 별 축제 정보
- 기대 효과
 - 지역별 여행 정보와 코스 정보가 있어 여행 계획을 정할 때 용이
 - 휴가나 방학처럼 기간이 정해졌을 때 여행 계획에 축제 정보 검색 용이

수집 데이터

- 공통 수집 데이터
 - 제목
 - 지역
 - 조회수
 - 세부 내용
- 크롤러 항목별 수집 데이터
 - 여행 정보 : 안전 여행 지수
 - 여행 기사 : 수정일
 - 코스 정보 : 코스 총 거리
 - 축제 정보 : 축제 개최 날짜

시스템 구조 설계

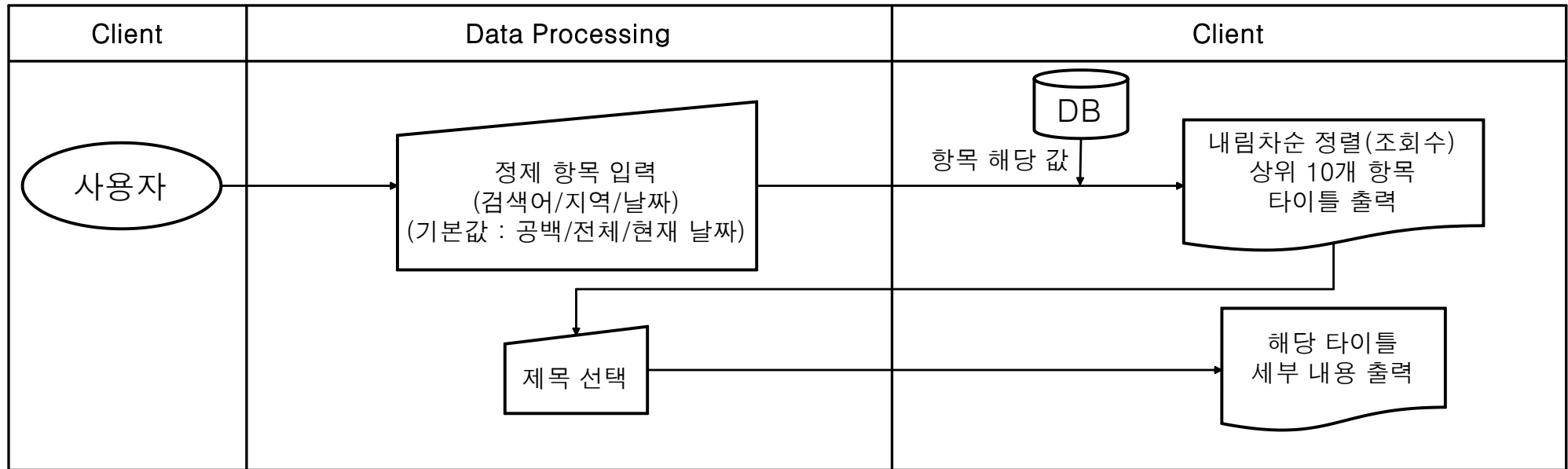
데이터 수집



대한민국 구석구석 사이트의 데이터를 크롤링합니다.
버튼에 따라 여행정보, 여행기사, 코스, 축제 데이터를 수집합니다.

시스템 구조 설계

데이터 가공



검색어 / 지역 / 날짜 항목을 선택하고 DB에서 해당하는 데이터를 추출해 출력합니다.
상위 10개 항목의 타이틀을 출력하고 타이틀을 선택하면 해당 타이틀의 세부 내용을 출력합니다.

테이블 정의서

여행정보

테이블 한글명		여행정보	테이블 영문명		koreatrip_info
칼럼 한글명	칼럼 영문명	데이터 타입	길이	Not Null	PRIMARY KEY
제목	title	VARCHAR	100		Y
지역	location	VARCHAR	15		Y
조회수	click_cnt	VARCHAR	10	Y	
안전지수	safe_cnt	VARCHAR	10	Y	
세부내용	detail_text	VARCHAR	10000	Y	

여행기사

테이블 한글명		여행기사	테이블 영문명		koreatrip_article
칼럼 한글명	칼럼 영문명	데이터 타입	길이	Not Null	PRIMARY KEY
제목	title	VARCHAR	100		Y
지역	location	VARCHAR	15		Y
조회수	click_cnt	VARCHAR	10	Y	
수정일	date	VARCHAR	100	Y	
세부내용	detail_text	VARCHAR	10000	Y	

여행코스

테이블 한글명		여행코스	테이블 영문명		koreatrip_course
칼럼 한글명	칼럼 영문명	데이터 타입	길이	Not Null	PRIMARY KEY
제목	title	VARCHAR	100		Y
지역	location	VARCHAR	15		Y
조회수	click_cnt	VARCHAR	10	Y	
코스총거리	total_distance	VARCHAR	100	Y	
코스정보	course_info	VARCHAR	10000	Y	

축제정보

테이블 한글명		축제정보	테이블 영문명		koreatrip_festival
칼럼 한글명	칼럼 영문명	데이터 타입	길이	Not Null	PRIMARY KEY
제목	title	VARCHAR	100		Y
지역	location	VARCHAR	15		Y
조회수	click_cnt	VARCHAR	10	Y	
개최날짜	date	VARCHAR	100	Y	
세부내용	detail_text	VARCHAR	10000	Y	

윈도우 폼 (PyQt5)

Main Page – 크롤링 탭

Main Page

KOREATRIP

크롤링 데이터 추천

옵션 ☒ 여행정보 ☐ 여행기사 ☐ 여행코스 ☐ 축제정보

제목

지역

조회수

세부 내용

추가 정보 양호 : 0 ~ 50%
보통 : 51 ~ 70%
주의 : 71 ~ 90%
경계 : 91 ~ 100%

확인 취소

분류	여행정보, 여행기사, 여행코스, 축제정보
공통 수집 데이터	제목, 지역, 조회수, 세부 내용
추가 정보	<ul style="list-style-type: none"> 여행정보 <ul style="list-style-type: none"> 안전 여행 지수(코로나) 여행기사 <ul style="list-style-type: none"> 최종 수정일 여행코스 <ul style="list-style-type: none"> 코스 총 거리 축제정보 <ul style="list-style-type: none"> 축제 진행 날짜

윈도우 폼 (PyQt5)

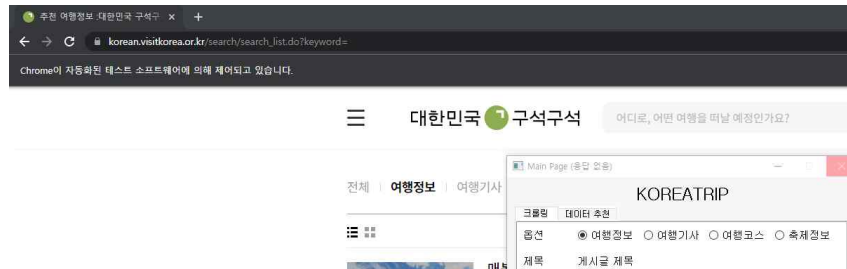
Main Page – 데이터 추천 탭

The screenshot shows a PyQt5 window titled 'Main Page' with a header 'KOREATRIP'. It features a tab interface with '크롤링' and '데이터 추천'. The '데이터 추천' tab is active, showing search options for travel information, travel articles, travel courses, and festivals. The '여행정보' option is selected. Search criteria include a search term, a region (set to '전체'), and a date (set to '2022-05-22'). There are also buttons for selecting time intervals (1, 3, or 6 months). A calendar widget is displayed to the right, showing the month of May 2022 with the 12th highlighted. The main form has '확인' (Confirm) and '취소' (Cancel) buttons at the bottom.

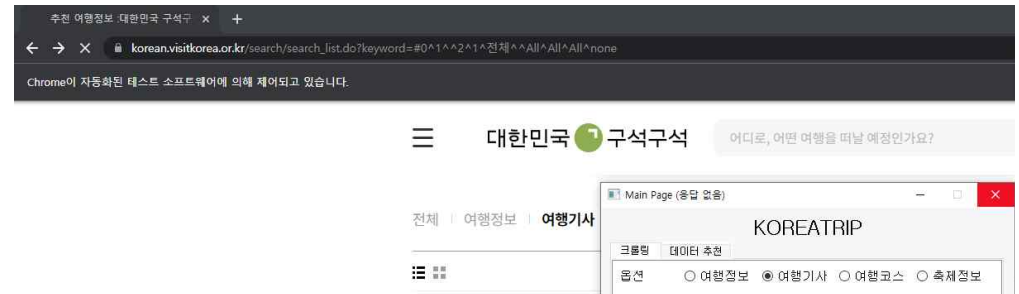
분류	여행정보, 여행기사, 여행코스, 축제정보
검색어 입력	<ul style="list-style-type: none"> 축제정보 항목 사용 불가 <ul style="list-style-type: none"> 기본값 : 공백
지역 입력	<ul style="list-style-type: none"> 모든 항목 사용 가능 <ul style="list-style-type: none"> 기본값 : 전체 (전국)
날짜 입력	<ul style="list-style-type: none"> 축제정보 항목만 사용 가능 <ul style="list-style-type: none"> 시작일 기본값 : 현재 날짜 종료일 기본값 : 공백 날짜 칸 클릭시 달력 생성 <ul style="list-style-type: none"> 달력 - 확인 : 선택 날짜로 입력 달력 - 취소 : 현재 날짜로 입력
데이터 출력	<ul style="list-style-type: none"> 최대 상위 10개 항목 출력 타이틀 선택시 세부 내용 출력 상위 타이틀 및 세부 내용 출력 창은 모달(modal) 형태로 구성

윈도우 폼 (PyQt5)

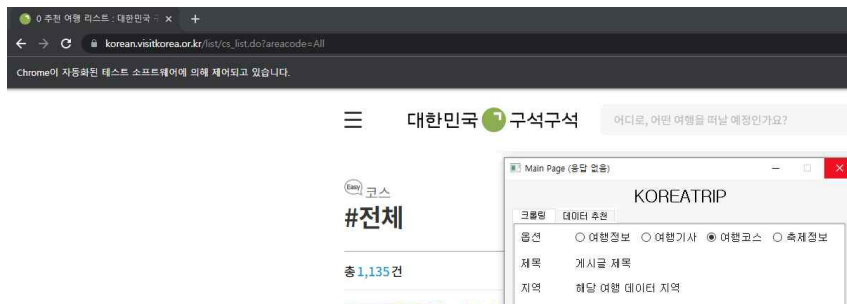
크롤링



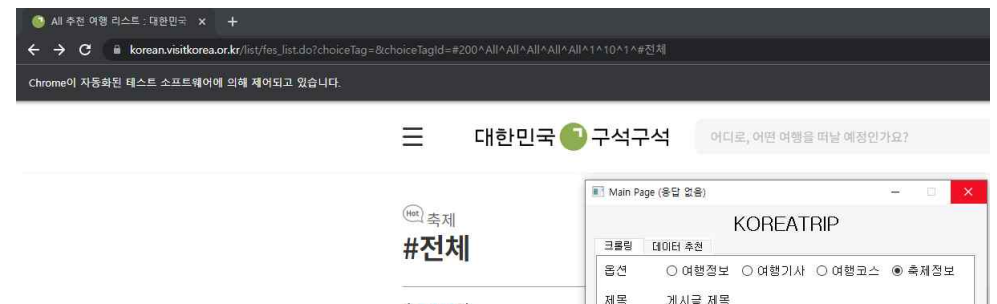
여행정보



여행기사



여행코스



축제정보

윈도우 폼 (PyQt5)

데이터 추천 - 여행정보

The image shows three windows of a PyQt5 application titled 'KOREATRIP'.
1. Main Page: A window with a title bar 'KOREATRIP'. It has a sidebar with '여행정보' (Travel Info) selected. The main area shows a list of travel packages. The first package, '오록도 스카이워크', is selected and highlighted in blue. Below the list are input fields for '검색어' (Search) and '지역' (Location), and a '날짜' (Date) field set to '2022-06-22'. At the bottom are '확인' (Confirm) and '취소' (Cancel) buttons.
2. Print_Info: A window titled 'Print_Info' showing the details of the selected package. It displays '검색어: None, 지역: 전체, 총 데이터 수: 29904개'. The list shows '오록도 스카이워크' as the first item. At the bottom are '저장' (Save) and '뒤로' (Back) buttons.
3. Detail_Info: A window titled 'Detail_Info' showing the detailed description of the '오록도 스카이워크' package. It includes fields for '제목' (Title), '지역' (Location), '조회수' (Views), and '안전지수' (Safety Index). The '세부 내용' (Detailed Content) field contains a long text description of the package. At the bottom are '확인' (Confirm) and '뒤로' (Back) buttons.

default 값을 적용한 경우

The image shows three windows of a PyQt5 application titled 'KOREATRIP'.
1. Main Page: A window with a title bar 'KOREATRIP'. It has a sidebar with '여행정보' (Travel Info) selected. The main area shows a list of travel packages. The first package, '서울시립미술관', is selected and highlighted in blue. Below the list are input fields for '검색어' (Search) and '지역' (Location), and a '날짜' (Date) field set to '2022-06-22'. At the bottom are '확인' (Confirm) and '취소' (Cancel) buttons.
2. Print_Info: A window titled 'Print_Info' showing the details of the selected package. It displays '검색어: 서울, 지역: 서울, 총 데이터 수: 296개'. The list shows '서울시립미술관' as the first item. At the bottom are '저장' (Save) and '뒤로' (Back) buttons.
3. Detail_Info: A window titled 'Detail_Info' showing the detailed description of the '서울시립미술관' package. It includes fields for '제목' (Title), '지역' (Location), '조회수' (Views), and '안전지수' (Safety Index). The '세부 내용' (Detailed Content) field contains a long text description of the package. At the bottom are '확인' (Confirm) and '뒤로' (Back) buttons.

입력값을 적용한 경우

윈도우 폼 (PyQt5)

데이터 추천 - 여행기사

Main Page

KOREATRIP

여행길

데이터 추천

옵션

☐ 여행정보

☒ 여행기사

☐ 여행코스

☐ 축제정보

검색어

지역

전체

날짜

2022-05-22

~

검색

검색

검색

확인

취소

Print_Info

검색어 : None, 지역 : 전체, 총 데이터 수 : 503개

1

1 기록에 가볼 만한 경북 문화 산행여행 1박2일 코스

2 [한산여행] 장주형 드라마 속 여간 여다국내 드라마 촬영 명소

3 서울 근교, 가볍게 쏙바람 쐬기 좋은 파주 당일치기 여행

4 책트리의 마을, 옛 문트의 전곡, 서천 한교마을

5 가장 볼나스 여행을 이끄는 '트로이카' V자형 출렁다리, 놀러향부리길, 선암지...

6 [국내 트레킹 추천]제주 불하길 걷기 여행

7 유복이 여행 1번지! 당일 걷기 여행, 당일 가볼 만한 곳, 당일 가볼 만한 곳

8 같은 곳엔 이제 NO!자연과 예술이 함께하는 씨다른 여수 여행

9 [한산여행] 청량길 가득한 숲을 거닐어요, 한국 가볼 만 한 숲 5곳

10 남쪽 다- 가는 포항 일교, 신과 바다를 100% 즐길 수 있는 포항 가볼만한곳

검색

확인

뒤로

Detail_Info

제목 가볼배 가볼 만한 경북 문화 산행여행 1박2일 코스

지역 경북 봉화군

조회수 23600

수정일 2020.10.20

세부 내용

[1]북 국립백두대간수목원 ~ 위치 : 경북 봉화군 춘양면 춘양로 1501 / 054-679-1000~x80 ~ 입장료 : 성인 5000원 / 청소년 4000원 / 어린이 3000원 ~ 관람시간 : 09:00~18:00 (하루가) / 09:00~17:00 (물고기) / 매주 월요일, 1월 2일, 설날 및 추석 당일 휴무 / 관람 종료 : 1시간 전 입장 마감. * 북 문헌역 ~ 위치 : 경북 봉화군 춘천면 문헌길 49, *x80~2-2강구역 ~ 위치 : 경북 봉화군 봉화읍 해저리 725 / 010-9169-5578~x80 ~ 숙박료 : 2인실 5만원부터 수~5인실 15만원(까지), *x80~2인실(별채형상(마을) ~ 위치 : 경북 봉화군 봉화읍 출재길 44 / 054-674-0963 ~ 입장료 : 무료~x80 ~ 관람시간 : 09:30~17:00(한이 일찍을 때에는 바로 앞 송림박물관 직원에게 투어)x80 ~ 기타 정보 :x80 초록을 훼손하거나 잡초를 뽑아가는 행위 불가, 청양정 정자 출입 불가. *x80봉화목재문화체험장~ 위치 : 경북 봉화군 봉안면 구갈로 151 / 054-674-3363~x80~ 관람시간 : 09:00~18:00 (월요일 휴무)x80~ 입장료 : 무료 (목재체험은 유료.)]

default 값을 적용한 경우

Main Page

KOREATRIP

여행길

데이터 추천

옵션

☐ 여행정보

☒ 여행기사

☐ 여행코스

☐ 축제정보

검색어

지역

서울

날짜

2022-05-22

~

검색

검색

검색

확인

취소

Print_Info

검색어 : 서울, 지역 : 서울, 총 데이터 수 : 13개

1

1 다 같이 풀자 서울 한 바퀴- 405번 버스 여행

2 브라운 마이 잉여 라이다! 정순이로 반할 서울 잉여적 상송 스왑 5

3 신인한 감각으로 재탄생하다. 리사이클링 서울

4 건축과 역사 따라 걷는 서울 서소문 순례길

5 수고한 나에게 주는 선물(호안소와 함께하는 서울 나들이

6 율리듬오브코리아 홍보영상 #서울 #사랑가Feel the Rhythm of Korea - SEOUL1

7 율리듬오브코리아 홍보영상 #서울 #사랑가Feel the Rhythm of Korea - SEOUL2

8 가볼배 일요면 덕 풍운서울의 도심 속 숲길 4

9 가볼배 만볼 수 있는서울의 숨은 명소 장소

10 서울 상영보통 교육문화센터 [2020 신인관람 13선]

검색

확인

뒤로

Detail_Info

제목 다 같이 풀자 서울 한 바퀴- 405번 버스 여행

지역 서울

조회수 5400

수정일 2020.01.22

세부 내용

[1]위치 : 서울시 중구 세종대로 99문의 : 02-771-9952입장료 : 개인(만 25세~64세) 1000원, 단체(10인 이상) 800원, 만 24세 이하 및 만 65세 이상 내국인 무료관람시간 : 9:00 ~ 21:00 (입장 마감 20:1), 정기 휴관일 : 매주 월요일~ 405번 버스 타고 가는 밤방'시작관람'에서 하차 후 도보 3분', '위치 : 서울시 중구 경마로 432입장료 : 무료개방시간 : 상시 개방 (24시간)~ 405번 버스 타고 가는 밤방'시작관람'에서 하차 후 도보 5분', '위치 : 서울시 용산구 신촌로 95-9번 405번 버스 타고 가는 밤방'보신(여름과 일교에서 하차 후 도보 2분', '위치 : 서울시 용산구 이태원로 200문의 : 02-2014-7800입장료 : 무료, 현대카드 미스지자 입장 가능 (바로 앞 현대카드 뮤직 라이브러리와 별개)관람시간 : 화~토 12:00 ~ 21:00, 일요일 및 공휴일 12:00 ~ 18:00, 매주 월요일 휴무, 순/숙적 연휴 휴무~ 405번 버스 타고 가는 밤방'한남동 세마로길고에서 하차 후 도보 1분']

입력값을 적용한 경우

윈도우 폼 (PyQt5)

데이터 추천 - 여행코스

The screenshot shows three windows of the KOREATRIP application. The Main Page window has the '여행코스' (Travel Course) tab selected. The Print_Info window shows a list of 10 travel courses. The Detail_Info window shows the details for the first course, '여수 도심길 가로질러 오동도까지 걸어라'.

Main Page

KOREATRIP

크롤링 데이터 추천

옵션 ☐ 여행정보 ☐ 여행가사 ☒ 여행코스 ☐ 축제정보

검색어

지역

날짜 ~

Print_Info

검색어 : None, 지역 : 전체, 총 데이터 수 : 1029개

1

- 1 여수 도심길 가로질러 오동도까지 걸어라
- 2 양화보다 더 평화로운 여행, 부산
- 3 빙으로 빙나는 강주의 야간투어 코스
- 4 (강릉+고성) 동강 따라 옛 마차, 7만 국도 드라이브
- 5 제주도 세계자연유산 탐방 3 코스
- 6 바다와 호수가 만났어온 멋진 드라이브 코스 1
- 7 부산의 자연생태 체험 코스
- 8 화산섬 제주와 속살을 만나다
- 9 낮이 좋아? 밤이 좋아? 서울의 주말, 야경 여행코스
- 10 (영광+고성) 숲과 바다를 낚아드는 힐링 로드

Detail_Info

제목 여수 도심길 가로질러 오동도까지 걸어라

지역 전라남도 여주시

조회수 28000

코스총거리 5.78km

세부 내용

코스1 : 여수 협성캐이볼카
주소 : 전라남도 여주시 출신읍 출신로 3600-1
코스 인근 여행지
충화물다방 | 모이린 | 남만캐페 |

코스2 : 흥산공원
주소 : 전라남도 여주시 출신읍 무두리 산 1
코스 인근 여행지
남만캐페 | 충화물다방 | 모이린 |

코스3 : 아쿠아플라넷 여수
주소 : 전라남도 여주시 오동도로 61-11
코스 인근 여행지
남만캐페 | 충화물다방 | 모이린 |

코스4 : 한려해상국립공원 (오동도)
주소 : 전라남도 여주시 오동도로 222
코스 인근 여행지
충화물다방 | 모이린 | 남만캐페 |

default 값을 적용한 경우

The screenshot shows three windows of the KOREATRIP application. The Main Page window has the '여행코스' (Travel Course) tab selected. The Print_Info window shows a list of 10 travel courses. The Detail_Info window shows the details for the first course, '낮이 좋아? 밤이 좋아? 서울의 주말, 야경 여행코스'.

Main Page

KOREATRIP

크롤링 데이터 추천

옵션 ☐ 여행정보 ☐ 여행가사 ☒ 여행코스 ☐ 축제정보

검색어

지역

날짜 ~

Print_Info

검색어 : 서울, 지역 : 서울, 총 데이터 수 : 19개

1

- 1 낮이 좋아? 밤이 좋아? 서울의 주말, 야경 여행코스
- 2 서울시티투어 버스를 이용하면 서울여행이 쉬워진다
- 3 서울의 랜드마크를 만나다
- 4 서울의 거리에 핀 세계의 맛과 멋, 서울맛집 탐방
- 5 서울의 상수동 이색골목 탐방
- 6 서울의 아름다운 밤을 바라보다
- 7 서울의 순교성지 약현상당 탐방 코스
- 8 북서울 꿈의 숲과 불광산의 올라 야경을 보다
- 9 명동에서 이태원까지 서울의 밤나들
- 10 서울 양천구 서서울호수공원 코스

Detail_Info

제목 낮이 좋아? 밤이 좋아? 서울의 주말, 야경 여행코스

지역 서울

조회수 13100

코스총거리 7.39km

세부 내용

코스1 : 광화문
주소 : 서울특별시 종로구 사직로 161
코스 인근 여행지
국립고궁박물관 | 명상힐후 조난지 | 동십자각 |

코스2 : 충계현
주소 : 서울특별시 종로구 창신동
코스 인근 여행지
동화상가 | 덕빌스파 웰대문현대시티아울렛점 | 서울
탐방시절 |

코스3 : 남대문시장
주소 : 서울특별시 중구 남대문시장4길 21
코스 인근 여행지
남대문 종합상가 | 남대문 중앙상가 | 남대문 역사사인
상가 |

코스4 : 남산서울타워
주소 : 서울특별시 용산구 남산공원길 105
코스 인근 여행지

입력값을 적용한 경우

윈도우 폼(PyQt5)

데이터 추천 - 축제정보

버튼(1개월, 3개월, 6개월) : 현재 날짜 기준 -(1, 3, 6)개월, +(1, 3, 6)개월로 시작일, 종료일을 입력합니다.

The screenshot shows the KOREATRIP application with the '축제정보' (Festival Information) tab selected. The '지역' (Region) is set to '전체' (All). The '날짜' (Date) is set to '2022-05-12'. The '기간' (Period) is set to '1개월' (1 month). The '축제정보' (Festival Information) table is empty. The 'Print_Info' window shows the default values for the festival information form.

구분	내용
1	대한 불꽃축제
2	28회 서울디지털페어 [갑사육하전]
3	여보아일랑드 불빛동화축제
4	[경제다교육제]전주 남강유등축제
5	2022년 청학궁 달빛기행
6	송암오르보정마드축제
7	부여서문연꽃축제
8	단산불빛마루
9	2022 서울강의축제
10	피스트가든 불꽃제 - 불빛이 흐르는 정월 2021

default 값을 적용한 경우

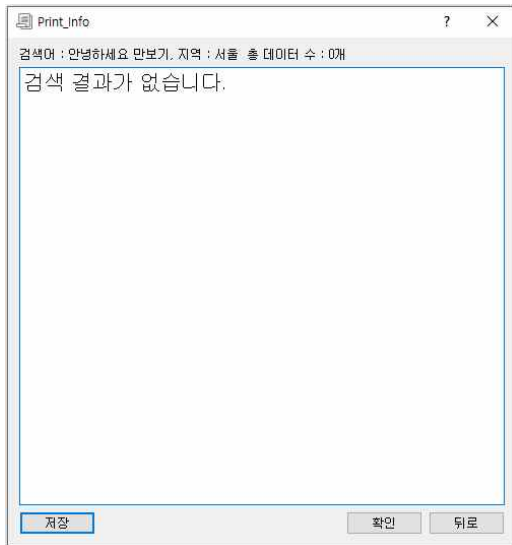
The screenshot shows the KOREATRIP application with the '축제정보' (Festival Information) tab selected. The '지역' (Region) is set to '서울' (Seoul). The '날짜' (Date) is set to '2022-04-12'. The '기간' (Period) is set to '3개월' (3 months). The '축제정보' (Festival Information) table is populated with the following data:

구분	내용
1	28회 서울디지털페어 [갑사육하전]
2	2022년 청학궁 달빛기행
3	2022 서울강의축제
4	서울국제도시전
5	궁중문화축제
6	2022년 경복궁 생각방
7	연등회
8	2022 대한민국 국제요리&제과 경연대회
9	청학궁 야간 상사관람
10	2022 종로대제

입력값을 적용한 경우

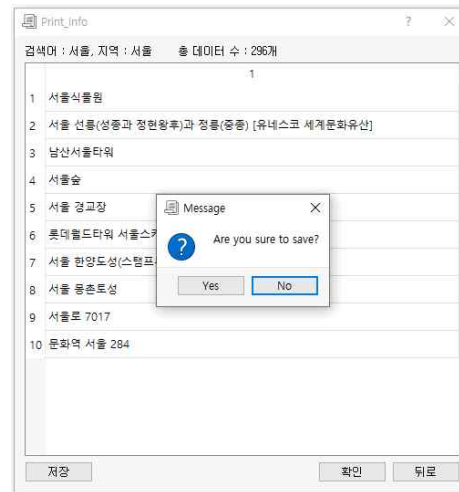
추가 기능

검색 결과 없음

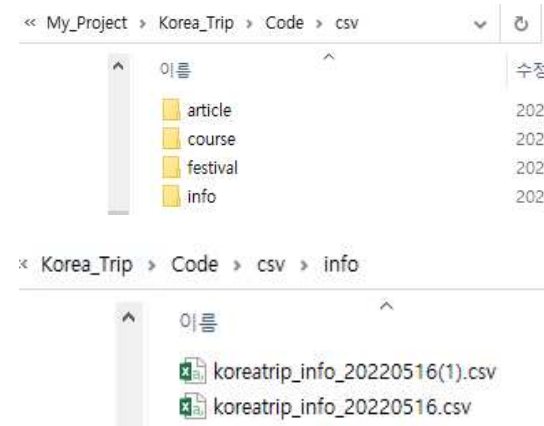


검색 결과가 없을 경우
해당 문구를 띄웁니다.

파일 저장



디렉터리 구조



여행정보 : info
여행기사 : article
여행코스 : course
축제정보 : festival

- 저장 버튼을 클릭할 경우 yes/no 메시지를 띄우고 yes 클릭시 저장합니다.
- 파일명은 'DB table명_현재날짜' 형식으로 상대경로를 통해 저장합니다.
- 중복된 파일이 있을 경우 숫자를 붙여 저장합니다.
- 상위 10개 항목이 아닌 정제된 총 데이터를 저장합니다.

후기

크롤러

크롤러 프로세서를 만들면서 HTML 구조를 이해하고 사용 능력을 올릴 수 있었습니다. 크롤링을 하면서 가장 많이 겪은 에러는 Stale Error와 Timeout Error 였습니다. 컴퓨터 사양과 서버 문제로 생기는 경우도 있고 특정 게시글 HTML 구조가 다른 경우가 있어 상황에 맞게 해결하면서 분석 능력과 해결 능력을 키울 수 있었습니다. 그리고 글쓴이 마다 쓰는 방식이 틀려 어떤 유형이 유효한 데이터를 많이 가지고 있는지 분석하면서 새로운 경험도 할 수 있었습니다.

GUI

PyQt5를 사용해 윈도우 폼을 구축하면서 Layout과 Widget 사용법을 배워 상황에 맞게 배치했습니다. Layout을 먼저 설계하고 Widget을 추가하면서 프로세스 구조를 익힐 수 있었습니다. 평소에 사용하던 다른 인터페이스를 참고하여 사용자가 사용하기 편리하게 만들었고 다른 기능도 생각해 볼 수 있는 시간이었습니다.

감사합니다

노성진
shtjdwlsvs@naver.com