Python을 이용한 웹 크롤러

지원자 : 남유진

크롤링을 이용한

게임가격비교 웹 사이트(진행 중)

“0o0”

웹 사이트 주소

http://106.10.53.210:8080/GameTest/all.jsp

OoO

사이트 소개

OoO는 각기 다른 게임사이트의 게임가격을 한데 모아 비교하여 소비자가 저렴한 가격에 게임을 구매할 수 있도록 정보를 제공하는 사이트입니다.

현재 스팀, 다이렉트 게임즈, 험블번들의 정보를 이용하고 있으며 더 추가할 예정입니다. 다양한 가격비교 사이트가 생겼지만 게임을 타겟으로 제공되는 사이트가 없어 이 문제를 해결하기 위해 OoO사이트를 제작했습니다.
가격을 모아 비교해줄 뿐만 아니라 사이트의 정가, 할인가격, 할인율로 자세한 정보도 제공하여 편의성을 높였으며 게임의 상세한 정보를 또한 모아 쉽게 볼 수 있도록 할 예정입니다.
OoO는 게임을 구매하기 앞서 꼭 필요한 웹 서비스가 될 것입니다.
현재도 개발 중에 있으며 광고를 추가하여 차후 상업화할 예정입니다.

크롤러 제작 과정

1 다이렉트게임즈 게임순위 확인

크롤링을 하려는 url
BeautifulSoup으로 파싱
배열을 선언하고

게임이름 수집

할인가격 수집

기존가격 수집
For문을 이용하여 데이터를
수집하고 배열에 저장합니다.

Print를 이용하여 나타냅니다.

180418 directg_실습.py - C:₩Users₩남유₩Desktop₩할인사이트프로젝트₩데이터베이스₩180418 directg_실습.py (3.7.0b2)

File Edit Format Run Options Window Help

```
import sys
import io

#-*-coding:utf8*-
from urllib.request import urlopen
from bs4 import BeautifulSoup

f=open('directg.txt','w', encoding='utf-8') #파일 쓸 준비

url=urlopen("https://directg.net/game/game_search.html?searchValue=")
soup=BeautifulSoup(url,"html.parser", from_encoding="utf8")
gameName=[]
salesPrice=[]
basePrice=[]

for link1 in soup.find_all(name="h2",attrs={"style":"font-size:14px"}):
    name=link1.find('a').text.strip('₩n₩n₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩n₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t')
    gameName.append(name)
    #print("게임이름 : " + str(gameName))

for link2 in soup.find_all(name="div",attrs={"class":"PricesalesPrice vm-display vm-price-value"}):
    sPrice=link2.find(name="span",attrs={"class":"PricesalesPrice"}).text.strip('₩₩ ₩n₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩t₩n₩t₩t₩t₩t₩t
    #print(count, "위:", str(titles)[str(titles).find('"')+1:str(titles).find('"')])
    #title=link2.find(name="span",attrs={"class":"PricesalesPrice"}).text
    salesPrice.append(sPrice)
    #print("할인가격 : " + str(sPrice))

for link3 in soup.find_all(name="div",attrs={"class":"PricebasePrice vm-display vm-price-value"}):
    bPrice=link3.find(name="span",attrs={"class":"PricebasePrice"}).text.strip('₩₩ ')
    basePrice.append(bPrice)
    #print("기존가격 : " + str(bPrice))

for i in range(0,20):
    #print(str(artistRank) + "위" + '₩n' + "제목 : " + str(titles[i]) + '₩n' + "아티스트 : " + str(artists[i]) +'₩n')
    print(str(i+1) + "위" + '₩n' + "게임이름 : " + str(gameName[i]) + '₩n' + "할인가격 : " + str(salesPrice[i]) +'₩n' + "기존
    f.write(str(i+1) + "위" + '₩n' + "게임이름 : " + str(gameName[i]) + '₩n' + "할인가격 : " + str(salesPrice[i]) +'₩n' + "기

f.close()
```

1 다이렉트게임즈 게임순위 확인 실습

크롤링 결과

게임 순위를 지정하고
게임의 이름, 할인가격, 기존가격을
표시하는 크롤러를 만들었습니다.

```
1위
게임이름 : 토탈워: 삼국
할인가격 : 53,800
기존가격 : 59,800

2위
게임이름 : 시드 마이어의 문명 VI - 몰려드는 폭풍
할인가격 : 41,900
기존가격 : 45,000

3위
게임이름 : 에이스 컴뱃 7: 스카이즈 언노운
할인가격 : 51,500
기존가격 : 54,800

4위
게임이름 : 에이스 컴뱃 7: 스카이즈 언노운 디지털 (
할인가격 : 71,500
기존가격 : 75,800

5위
게임이름 : 슈퍼 픽셀 레이서즈
할인가격 : 14,800
기존가격 : 16,500

6위
게임이름 : 바이오하자드 RE:2 [레지던트 이블 2 ]
할인가격 : 55,000
기존가격 : 59,000

7위
게임이름 : 바이오하자드 RE:2 디럭스 에디션 [레지던
할인가격 : 62,000
기존가격 : 69,000
```

2 스팀 게임순위 확인 실습

steamTest.py - C:₩Users₩남유₩Desktop₩할인사이트프로젝트₩데이터베이스₩steam...

File Edit Format Run Options Window Help

```
def spider(max_pages):
    conn = pymysql.connect(host='localhost',port=3306,
        user='root',
        password='1234',
        db='game',
        charset='utf8mb4')

    #f = open("C:/Users/cho/Desktop/새파일.txt", 'w',encoding='UTF8')
    page=1
    while page< max_pages:
        url = 'http://store.steampowered.com/search/?l=koreana&page='+str(page)
        response = requests.get(url)
 #응답 html코드를 text로 변환
        html = response.text

#응답받은 html코드를 BeautifulSoup에 사용하기 위하여 인스턴스 지정
        soup = BeautifulSoup(html, 'html.parser')

        info = soup.find("div",{"id":"search_result_container"})
#원하는 태그 지정해서 출력 span[class=title]
        for link in info.select('a[class="search_result_row ds_collapse_flag "]'):
            for img in link.select('div[class="col search_capsule"]'):
                print("사진:"+str(img.find('img')['src']))
                image=str(img.find('img')['src'])
                image=image.split("/")[5]
                print("넘버:"+image)
            for title in link.select('span[class=title]'):
                print("영어 제목: "+title.text )
                #f.write("영어 제목: "+title.text+"₩n")
            for day in link.select('div[class="col search_released responsive_secondro
                print("날짜: "+day.text)
                #f.write("날짜: "+day.text+"₩n")
                #할인시 search_price discounted responsive_secondrow 이거 사용
            for count in link.select('div[class="col search_price discounted responsiv
                print("가격: "+count.text.strip())
                #f.write("가격: "+count.text.strip()+"₩n")
            for count in link.select('div[class="col search_price responsive_secondrow
                print("가격: "+count.text.strip())
                #f.write("가격: "+count.text.strip()+"₩n")
            with conn.cursor() as cursor:
                sql = 'select exists(select enname from steam where enname = "'+title
                cursor.execute(sql)
                count=cursor.fetchall()
                count=count[0][0]
                print(count)
                if(count==0):
                    print(title.text+" 게임이 없다.")
                    #없으니까 추가 insert
                #다음으로는 가격확인 가격이 다르면 가격을 변경 update

                #    sql = 'INSERT INTO steam (enname, day, price) VALUES (%s, %s,%s)'
                #    cursor.execute(sql, (title.text, day.text,count.text.strip()))
                conn.commit()


        page +=1
        #print(cursor.lastrowid)
        print(page)
    #f.close()
    conn.close()

#실제 실행되는 곳
start_time = time.time()
spider(2)

print("start_time", start_time) #출력해보면, 시간형식이 사람이 읽기 힘든 일련번호형식입
print("--- %s seconds ---" %(time.time() - start_time))
```

< 페이징 추가

페이징을 추가하여
전체정보를 수집합니다.

2 스팀 게임순위 확인 실습

크롤링 결과

다양한 정보 제공을 위해
사진링크, 게임 넘버, 이름(영어),
출시날짜, 기존가격 등 수집범위를
넓혔습니다.

3 MYSQL에서 로그인과 데이터 베이스 생성

로그인 >

game 데이터 베이스 > 생성

1.데이터베이스생성.py - C:₩Users₩남유₩Desktop₩할인사이.

File Edit Format Run Options Window Help

```
#데이터베이스 생성
import pymysql.cursors

conn = pymysql.connect(host='localhost',port=3306,
        user='root',
        password='1234',
        charset='utf8mb4')

try:
    with conn.cursor() as cursor:
        sql = 'CREATE DATABASE game'
        cursor.execute(sql)
    conn.commit()
finally:
    conn.close()
```

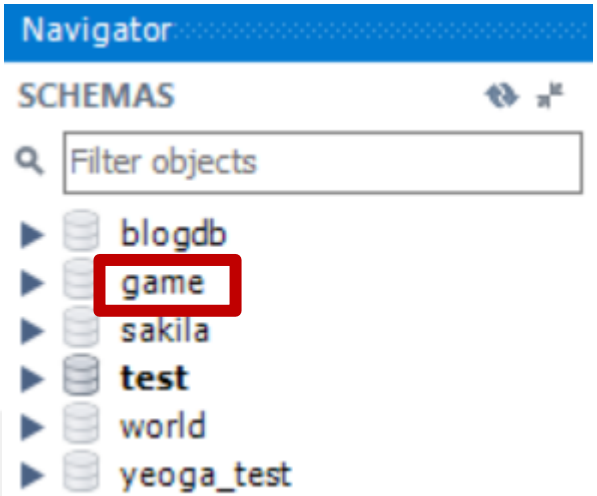
실행 결과

Python 3.7.0b2 Shell

File Edit Shell Debug Options Window Help

```
Python 3.7.0b2 (v3.7.0b2:b0ef5c979b, Feb 28 2018, 02:24:20) [MSC v.1912 64 bit (
AMD64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
======== RESTART: C:₩Users₩남유₩Desktop₩할인사이트프로젝트₩데이터베이스₩1.데이터
베이스생성.py ========
>>>
```

결과



4 테이블 생성

2.테이블생성.py - C:₩Users₩남유₩Desktop₩할인사이트프로젝트₩데이터베...

File Edit Format Run Options Window Help

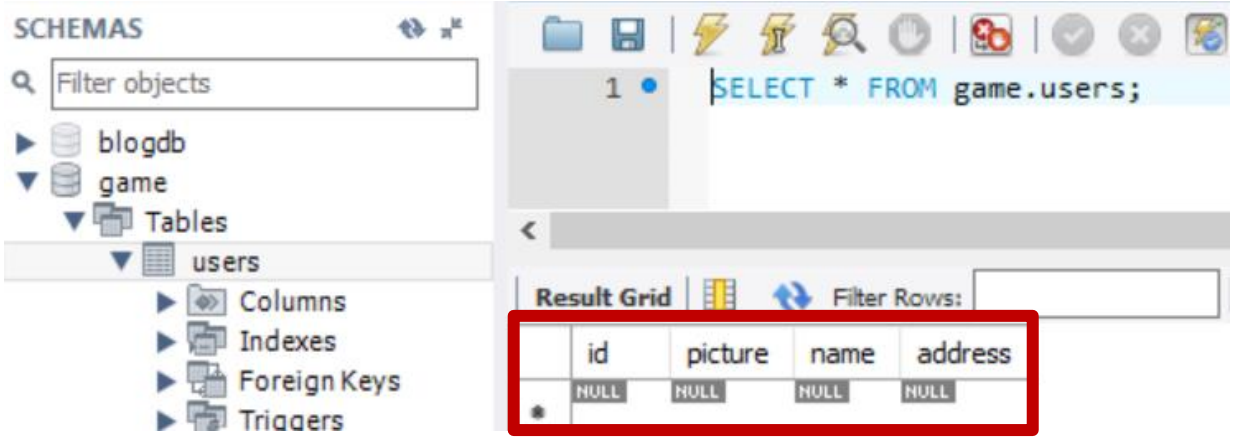
```
#테이블 생성
import pymysql.cursors

conn = pymysql.connect(host='localhost',
        user='root',
        password='1234',
        db='game',
        charset='utf8mb4')

try:
    with conn.cursor() as cursor:
        sql = '''
            CREATE TABLE users (
                id int(11) NOT NULL AUTO_INCREMENT PRIMARY KEY,
                picture varchar(255) NOT NULL,
                name varchar(255) NOT NULL,
                address varchar(255) NOT NULL
            ) ENGINE=InnoDB DEFAULT CHARSET=utf8
...
        cursor.execute(sql)
    conn.commit()
finally:
    conn.close()
```

로그인 >

테이블 > 생성

데이터베이스와 테이블을 생성하여
수집한 데이터를 가공할 수 있도록 저장합니다.

결과



5 게임사이트 ALL페이지 JSP

앞서 소개한 페이지의 코드입니다.

```jsp
<%@ page language="java" contentType="text/html; charset=UTF-8" pageEncoding="UTF-8"%>
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
<%request.setCharacterEncoding("UTF-8");
response.setContentType("text/html; charset=UTF-8");%>
<title>0o0-스팀할인정보</title>
</head>
<body>
    <jsp:include page="sub/header.jsp"></jsp:include>
    <jsp:include page="sub/menu.jsp"></jsp:include>
    <jsp:include page="sub/price.jsp"></jsp:include>
    <jsp:include page="sub/footer.jsp"></jsp:include>
</body>
</html>
```

6 게임사이트 PRICE 페이지 JSP

ALL페이지에서 가격을 표시하는 PRICE부분입니다.

데이터를 저장한 MYSQL과 연결하여 정보를 가져와
가격 정보 테이블에 해당하는 값을 넣었습니다.

```
int pagenumber;
if(request.getParameter("page")==null){
   pagenumber=1;
}else{
   pagenumber=Integer.parseInt(request.getParameter("page"));
}

  saledao dao=new saledao();
  //dao.dbConn();    db 연결 확인 작업

   ArrayList<saledto> list=dao.saleList(pagenumber);

   for(saledto dto:list){


  <tr>
          <td><img alt="" src="<%=dto.getImage()%>"> </td>
          <td><%=dto.getDirectgenname() %></td>
          <td><%=dto.getDay() %></td>
          <td><%=dto.getSteamprice() %></td>
          <td><a href="https://store.steampowered.com/app/<%=dto.getSteampagenumber() %> " target="_blank"><%=dto.getSteamprice() %>원</a></td>
          <td><a href="https://directg.net/game/game_page.html?product_code=<%=dto.getDirectgpagenumber() %>" target="_blank"><%=dto.getDirectgprice() %>
          <td>준비중...</td>

  </tr>
```

결과 가격비교 페이지 제작

PC버전

게임 가격

	게임이름	출시	정가	스팀	다이렉트	험블번들
	Elven Magic	2019-06-03	4,400원	4,400원	판매 안함	준비중...
	Computer Mechanic Simulator 2019	2019-02-06	알수없음	알수없음	판매 안함	준비중...
	My Fox Sister|	2019-02-04	무료	무료	판매 안함	준비중...
	WarGround	2019-02-02	무료	무료	판매 안함	준비중...
	Another Bad Day in the Future	2019-02-02	20,500원	20,500원	판매 안함	준비중...

모바일 버전

게임 가격

Elven Magic
2019-06-03
4,400원
스팀: 4,400원
다이렉트: 판매 안함
험블번들: 준비중...

Computer Mechanic Simulator
2019-02-06
알수없음
스팀: 알수없음
다이렉트: 판매 안함
험블번들: 준비중...

My Fox Sister|
2019-02-04
무료
스팀: 무료
다이렉트: 판매 안함
험블번들: 준비중...

다양한 디바이스를 사용하는 게임유저들의 접근성을 고려하여 반응형 웹으로 제작하였습니다.