

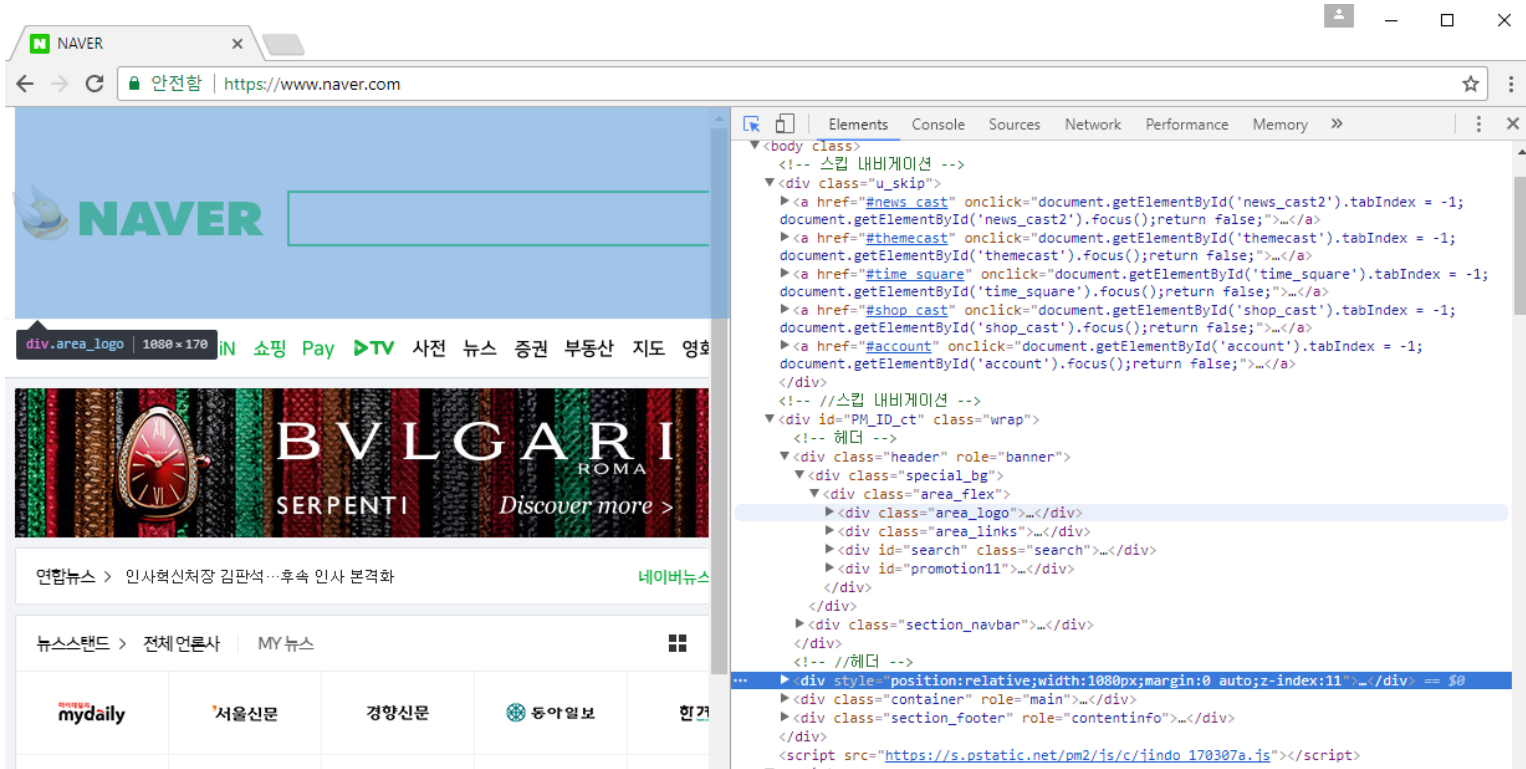
# Web Crawling - Basics

감성분석

박진수 교수

**Big Data Institute,  
Seoul National University**

# 웹 구조 이해하기



# 브라우저 설치

· 본 강의는 구글 크롬(google chrome) 브라우저를 기준으로 진행되니 크롬 브라우저가 없을 경우 컴퓨터에 이를 설치한다

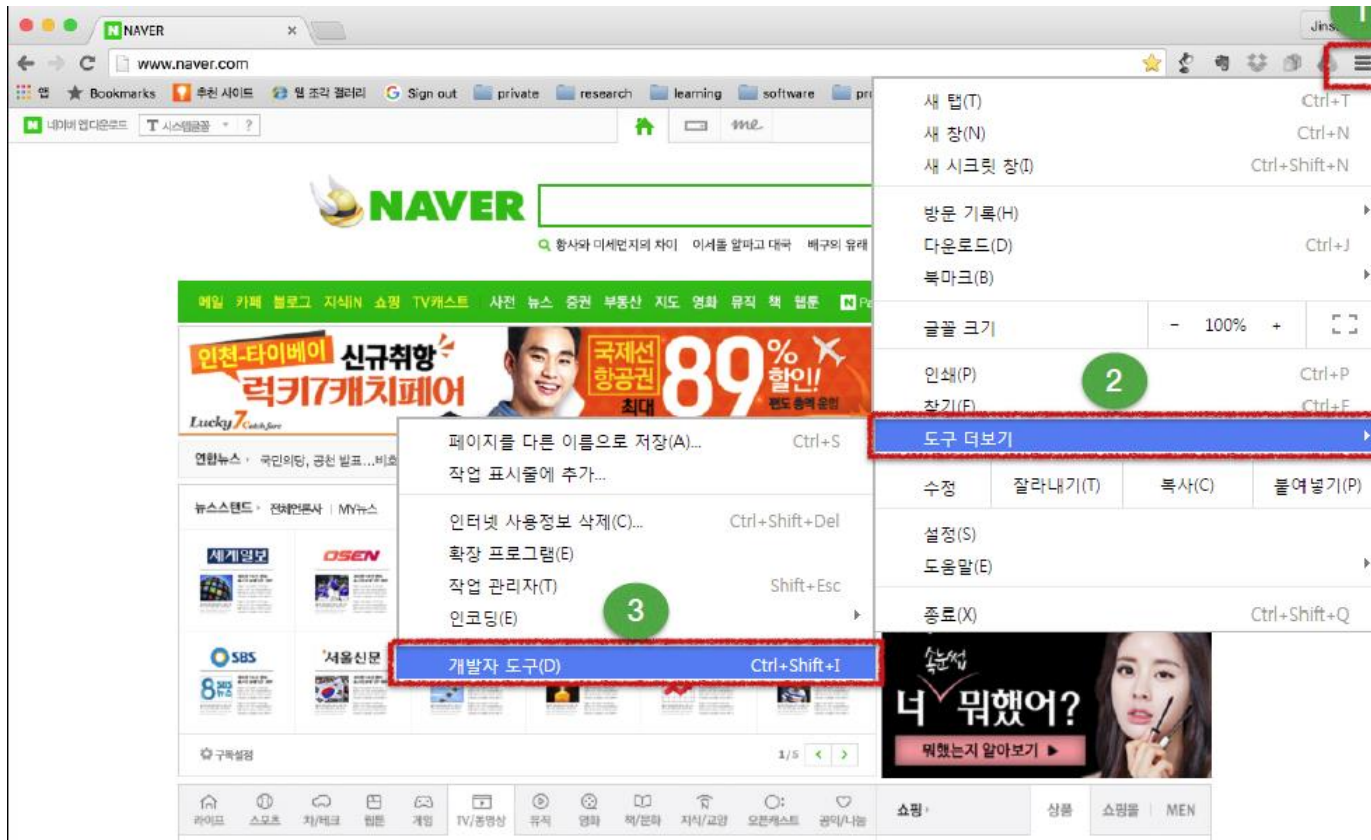
- 다운로드: <https://www.google.co.kr/chrome/browser/desktop/index.html>



# 개발자 도구(Developer Tools)

· 웹의 구조를 파악하기 위해서 크롬 브라우저의 '개발자 도구'를 활용한다

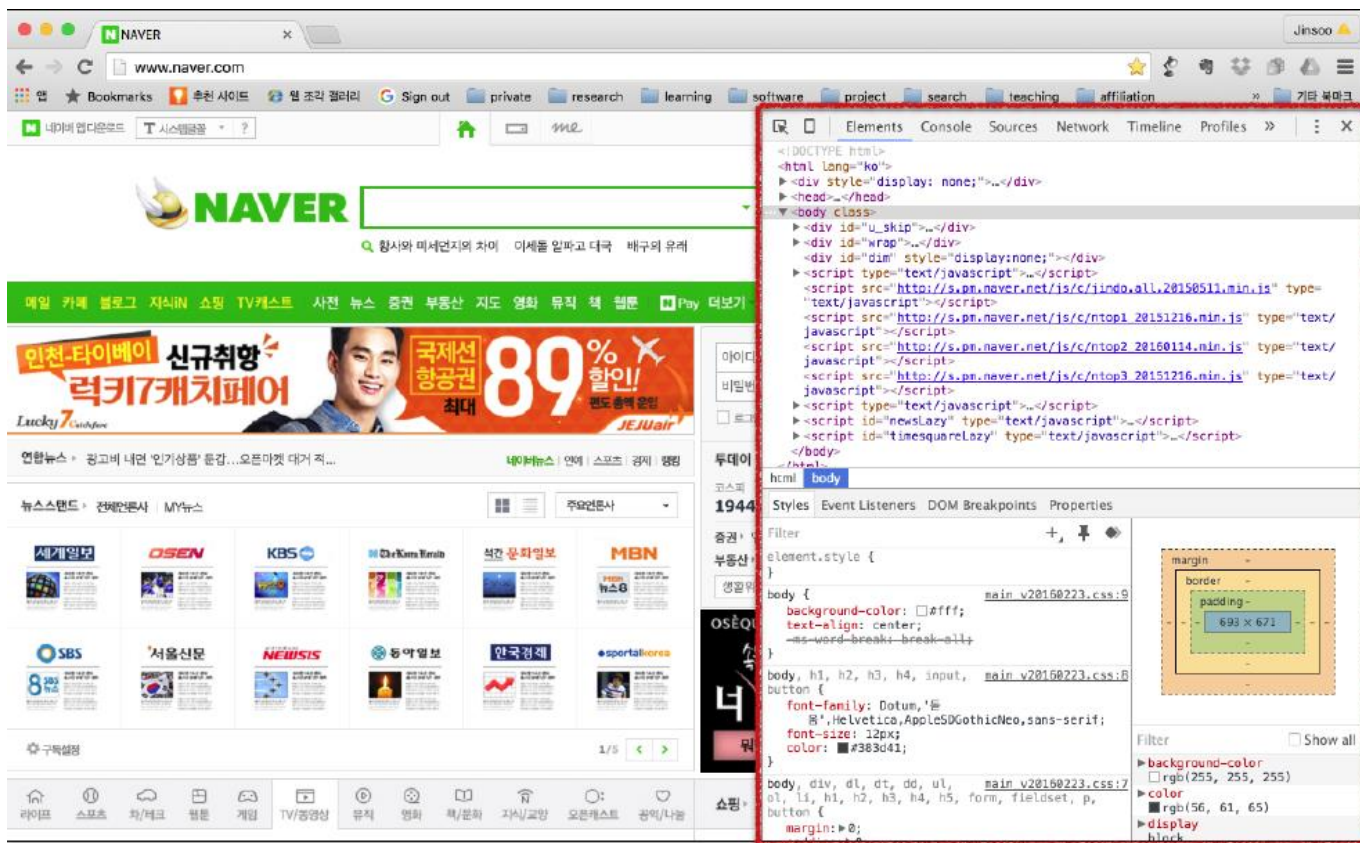
- 단축키: **Ctrl + Shift + I**



# 개발자 도구(Developer Tools)

· 웹의 구조를 파악하기 위해서 크롬 브라우저의 '개발자 도구'를 활용한다

- 단축키: **Ctrl + Shift + I**

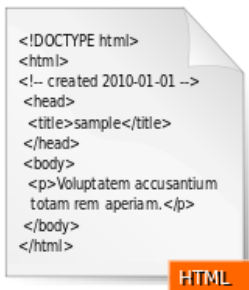


# HTML(HyperText Markup Language)

- 웹 페이지 제작을 위한 마크업 언어

- 마크업 언어(Markup Language): 태그, 속성 등을 이용하여 문서나 데이터의 구조를 정의하고 표현하는 언어
- HTML은 제목, 단락, 목록 등과 같은 본문을 위한 구조적 의미를 나타내는 것 뿐만 아니라 링크, 인용과 그 밖의 항목으로 구조적 문서를 만들 수 있는 방법을 제공
- HTML은 웹 페이지 콘텐츠 안의 꺾쇠 괄호(<>)에 둘러싸인 “태그”로 되어있는 HTML 요소 형태로 작성

```
<title> Hello HTML </title>  
<p> Hello World! </p>
```



# HTML(HyperText Markup Language)

- HTML 요소: 가장 보편적인 형태는 세 가지 구성 요소를 가진다. “태그” 한 쌍, “속성”, 그리고 “콘텐츠”
  - 태그(tag): 요소를 구분할 수 있도록 하는 이름표. 일반적으로 시작 태그(<tag>)와 종료 태그(</tag>) 한 쌍으로 구성된다
  - 속성(attribute): 태그들의 특성을 부여
  - 콘텐츠(contents): 화면에 표시되는 문자와 그래픽 정보 → 크롤링을 할 때 프로그래머의 관심의 대상

```
<tag attribute_1 = "value_1", attribute_2 = "value_2", ...>  
    contents </tag>
```

# HTML(HyperText Markup Language)

- HTML 마크업 요소의 형태

- 구조적 마크업: 본문의 목적을 표현

- 제목(title), 문단(p), 표(table) 등

- 표현적 마크업: 기능에 관계없이 본문의 외관을 표현

- 굵게(b), 이탤릭(i) 등

- 하이퍼텍스트 마크업: 다른 문서와 연결시켜주는 문서의 부분

- 링크(a)

```
<a href = "http://ko.Wikipedia.org/"> 한국어 위키백과 </a>
```



# HTML(HyperText Markup Language)

## · 주요 HTML 요소

HTML 요소	기능	비고
 	줄바꾸기	빈 태그(끝 태그가 없다)
<p>	단락바꾸기	
<h>	제목(heading)	h1, h2 등으로 표현
<center>	가운데 정렬	
<font>	폰트를 바꿈	
<ul><li>...</li></ul>	순서없는 목록	기본: 까만 동그라미
<ol><li>...</li>...</ol>	순서있는 목록	기본: 숫자
<table></table>	표 만들기	
<tr></tr>	행	<table>...</table>...에 넣는다
<td></td>	열	<tr>...</tr> ...에 넣는다

# HTML(HyperText Markup Language)

## · 속성

- 대부분의 요소 속성들은 “이름-값” 형태이고 “=”으로 구분한다.

■ 요소의 시작 태그 내에서 요소의 이름 다음에 쓰인다

- 일반적인 속성

■ **id**: 요소에 대한 문서 전체의 고유 식별자를 제공

■ **class**: 유사한 요소를 분류하는 방법을 제공

■ **style**: 특정 요소에 표현적 속성을 제공

■ **title**: 요소에서 숨겨진 뜻을 설명하는 글을 첨부

```
<abbr id="anId" class="aClass" style="color:blue;" title="하이퍼텍스트 마크업 언어">  
HTML</abbr>
```

# HTML(HyperText Markup Language)

- Hello world 프로그램 in HTML

```
<!doctype html>
<html>
  <head>
    <title> Hello HTML </title>
  </head>
  <body>
    <p> Hello World! </p>
  </body>
</html>
```

- 참고 - More HTML & CSS: <https://www.codecademy.com/learn/web>

# 실습 0 – HTML 문서 생성하기



# 실습 1-1-1. HTML 문서 생성하기

- 아래 내용을 담은 HTML 문서를 생성하고 이를 열어 본다

```
<!DOCTYPE html>
<html>
<body>

<h1>My First Heading</h1>

<p>My first paragraph.</p>

</body>
</html>
```

## 실습 1-1-2. HTML 문서 생성하기

- 아래 내용을 담은 HTML 문서를 생성하고 이를 열어 본다

```
<!DOCTYPE html>
<html>
<body>

<a href="https://www.w3schools.com">This is a link</a>

</body>
</html>
```

## 실습 1-1-3. HTML 문서 생성하기

- 아래 내용을 담은 HTML 문서를 생성하고 이를 열어 본다

```
<!DOCTYPE html>
<html>
<body>

<p>This is a paragraph.</p>
<p>This is a paragraph.</p>
<p>This is a paragraph.</p>

</body>
</html>
```

## 실습 1-1-4. HTML 문서 생성하기

- 아래 내용을 담은 HTML 문서를 생성하고 이를 열어 본다

```
<!DOCTYPE html>
<html>
<body>

<p>I am normal</p>
<p style="color:red;">I am red</p>
<p style="color:blue;">I am blue</p>
<p style="font-size:36px;">I am big</p>

</body>
</html>
```



# 실습 2 – 다음 사전

**dmm** 사전

백과사전 | 단어장 | 작은창 | 맞춤법검사기 *beta*

로그인 더보기 ▾

## 어학사전

맞춤법 검사기

curiosity



A가

Aa

가

あ

中

漢

...

# 실습 1-2-1. 단어의 검색 결과 출력하기

· 다음 어학사전(<http://dic.daum.net/index.do?dic=all>)에 'curiosity' 단어를 검색하였을 때 출력 되는 화면에서 가장 상단의 결과를 출력한다

- 영어 단어의 의미(본 예제에서는 '1. 호기심 2. 큐리오시티')를 출력한다

The screenshot shows the Daum Dictionary search results for the word 'curiosity'. The search bar at the top contains 'curiosity'. Below the search bar, there are links for '관련' (related), 'curious', 'descendant', 'crumble', 'out of curiosity', 'intuition', 'courage', and 'out of'. The main content area is titled '영어사전' (English Dictionary). The word 'curiosity' is displayed in large blue text, with a green box highlighting it. Below the word, the meanings are listed: '1. 호기심' (curiosity) and '2. 큐리오시티' (curiosity). The phonetic transcriptions are also provided: '미국 [kjüeriásəti] <≫> 영국 [kjüeriósəti] <≫>'. Below the main entry, there are three related phrases: 'unquenchable curiosity' (억누를 수 없는 호기심), 'insatiable curiosity' (만족을 모르는 호기심), and 'satisfy curiosity' (호기심을 만족시키다).

# 실습 1-2-1. 단어의 검색 결과 출력하기

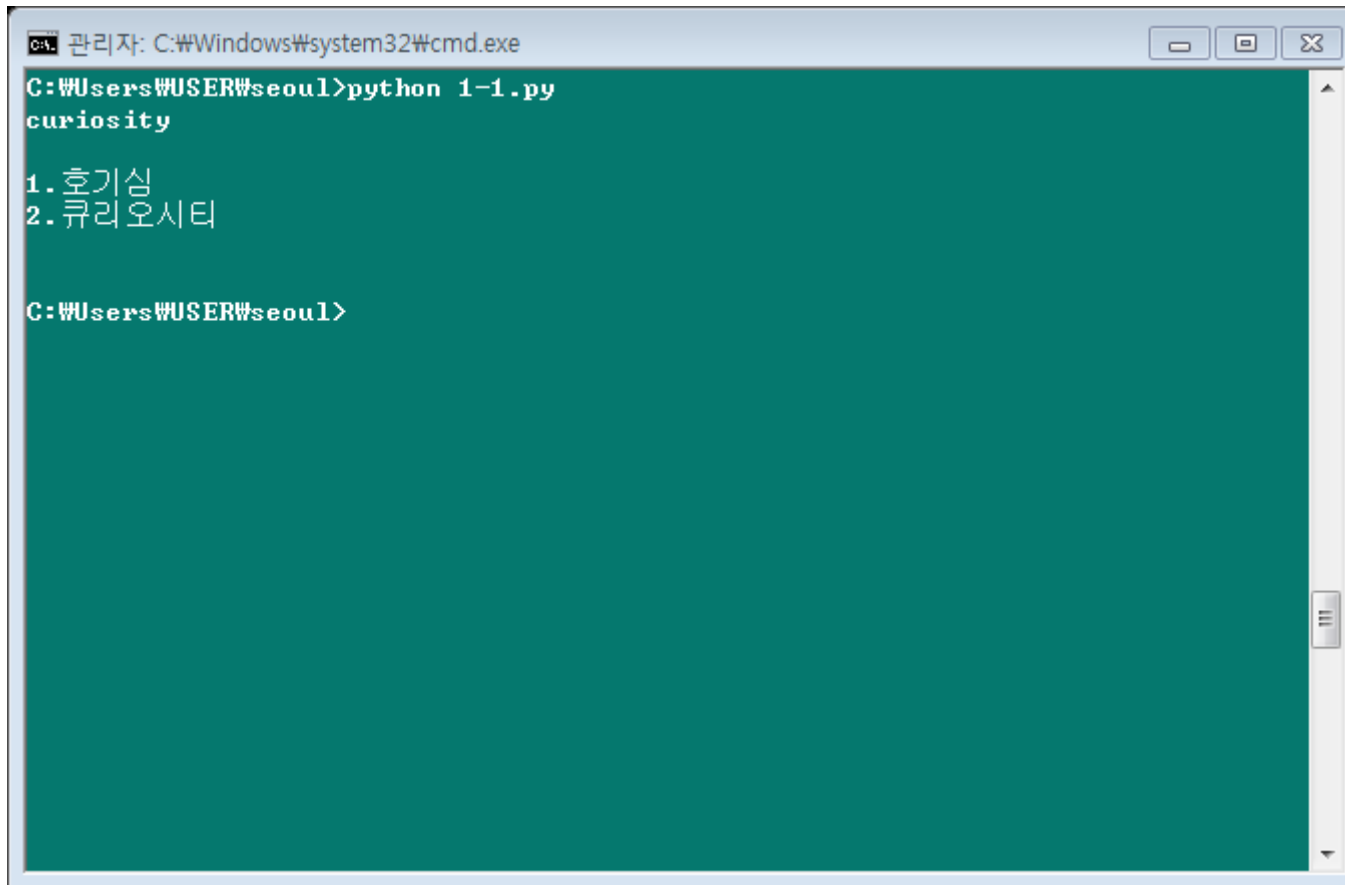
---

## · TIPS

- 가져오고자 하는 텍스트의 다른 HTML 구성요소와 구분되는 특성(CSS style)을 잘 생각해 본다
- 가장 상단에 있는 데이터임을 고려하면서 HTML 파싱 결과를 어떻게 슬라이싱(slicing)할 것인지 고민해 본다

## 실습 1-2-1. 단어의 검색 결과 출력하기

- 출력 결과



```
C:\> 관리자: C:\Windows\system32\cmd.exe
C:\Users\WUSER\seoul>python 1-1.py
curiosity

1. 호기심
2. 큐리오시티

C:\Users\WUSER\seoul>
```

The image shows a Windows command prompt window titled "관리자: C:\Windows\system32\cmd.exe". The prompt is "C:\Users\WUSER\seoul>". The user has entered the command "python 1-1.py". The output of the script is displayed on the next line: "curiosity". Below this, there is a numbered list: "1. 호기심" and "2. 큐리오시티". The prompt "C:\Users\WUSER\seoul>" is shown again at the bottom, indicating the command has finished executing.

## 실습 1-2-2. 여러 단어의 검색 결과 출력하기

· 다음 어학사전(<http://dic.daum.net/index.do?dic=all>)에 'curiosity', 'killed', 'the', 'cat' 4개 단어를 검색하였을 때 출력 되는 화면에서 가장 상단의 결과를 출력한다

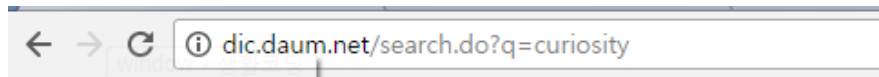
- 영어 단어의 의미를 출력한다

The screenshot shows the Daum Dictionary interface. At the top, there is a search bar with 'curiosity' entered. To the right of the search bar are links for '다른 사전' (Other dictionaries), '단어장' (Vocabulary), and '사전홈' (Dictionary home). Further right are '로그인' (Login) and '더보기' (More) buttons. Below the search bar, a horizontal list of related words is displayed: '관련 curious descendant crumble out of curiosity intuition courage out of'. The main content area is titled '영어사전' (English Dictionary). The word 'curiosity' is highlighted with a green box. Below it, the definitions '1. 호기심 2. 큐리오시티' are listed. Below the definitions, the pronunciations for American and British English are shown: '미국 [kjüeriósəti] <≫> 영국 [kjüeriósəti] <≫>'. Below this, three phrases are listed: 'unquenchable curiosity' (억누를 수 없는 호기심), 'insatiable curiosity' (만족을 모르는 호기심), and 'satisfy curiosity' (호기심을 만족시키다).

## 실습 1-2-2. 여러 단어의 검색 결과 출력하기

### · TIPS

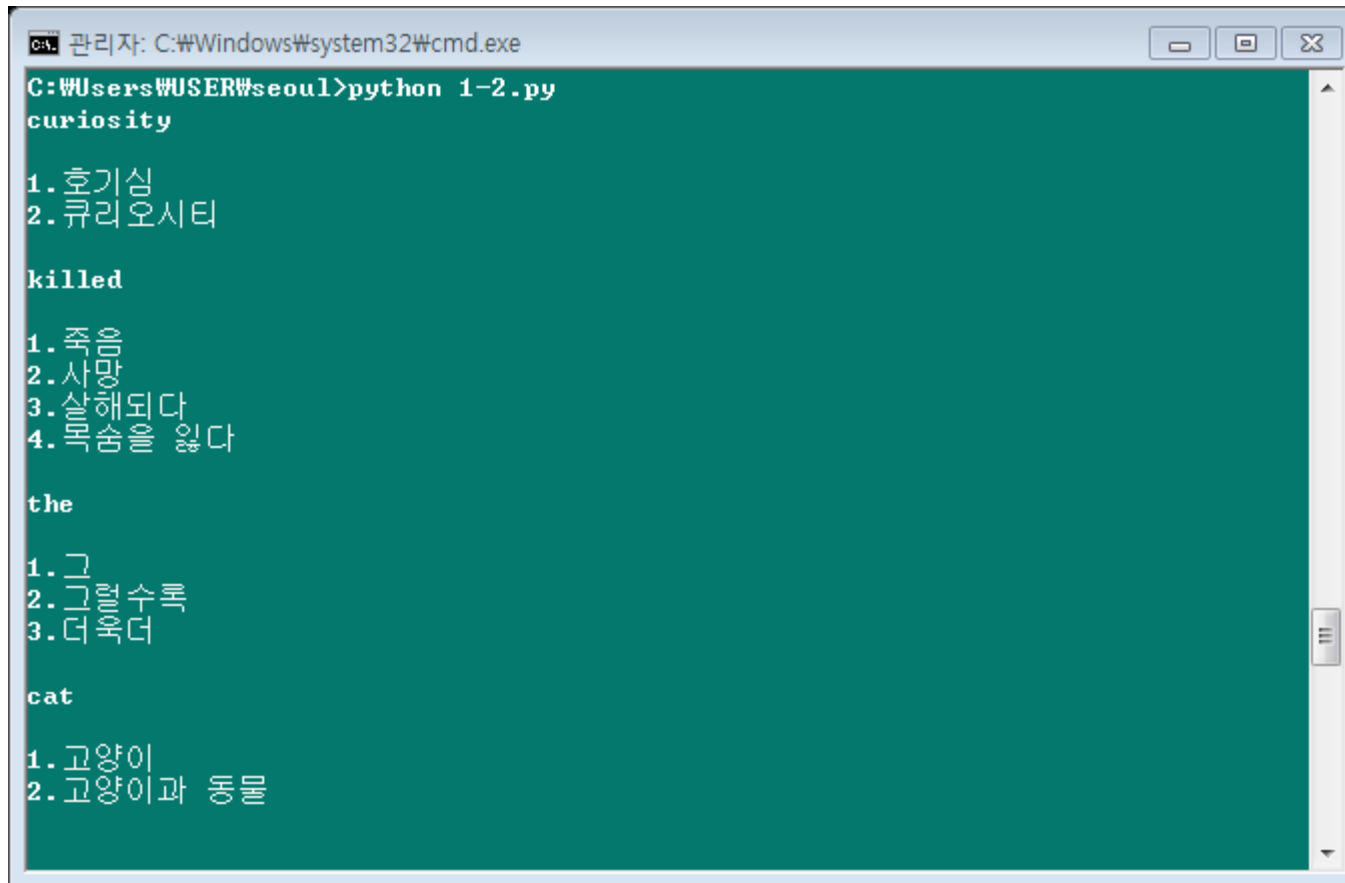
- 초기 화면에서 몇 개 단어를 검색해 보면서 URL 구조를 잘 생각해 본다



- for 문을 활용해 여러 단어를 돌아가면서 데이터를 출력한다

## 실습 1-2-2. 여러 단어의 검색 결과 출력하기

### · 출력 결과



```
C:\Users\WUSER\seoul>python 1-2.py
curiosity
1. 호기심
2. 큐리오시티

killed
1. 죽음
2. 사망
3. 살해되다
4. 목숨을 잃다

the
1. 그
2. 그럴수록
3. 더욱더

cat
1. 고양이
2. 고양이과 동물
```

# 실습 1-2-3. 여러 단어의 검색 결과 저장하기

· 다음 어학사전(<http://dic.daum.net/index.do?dic=all>)에 'curiosity', 'killed', 'the', 'cat' 4개 단어를 검색하였을 때 출력 되는 화면에서 가장 상단의 결과를 텍스트 파일에 저장한다

- 영어 단어의 의미를 저장한다

The screenshot shows the Daum Dictionary interface. At the top, there is a search bar with 'curiosity' entered. To the right of the search bar are links for '다른 사전' (Other dictionaries), '단어장' (Vocabulary), and '사전홈' (Dictionary home). Further right are '로그인' (Login) and '더보기' (More) buttons. Below the search bar, a horizontal list of related words is displayed: '관련 curious descendant crumble out of curiosity intuition courage out of'. The main content area features the word 'curiosity' in large blue text, with a green box highlighting it. Below the word, it lists '1. 호기심 2. 큐리오시티' and provides phonetic information: '미국 [kjūeriósəti] <≫> 영국 [kjūeriósəti] <≫>'. To the right of the word is a blue circular button with a white plus sign. Below this, three phrases are listed: 'unquenchable curiosity' (억누를 수 없는 호기심), 'insatiable curiosity' (만족을 모르는 호기심), and 'satisfy curiosity' (호기심을 만족시키다).



# 실습 1-2-3. 여러 단어의 검색 결과 저장하기

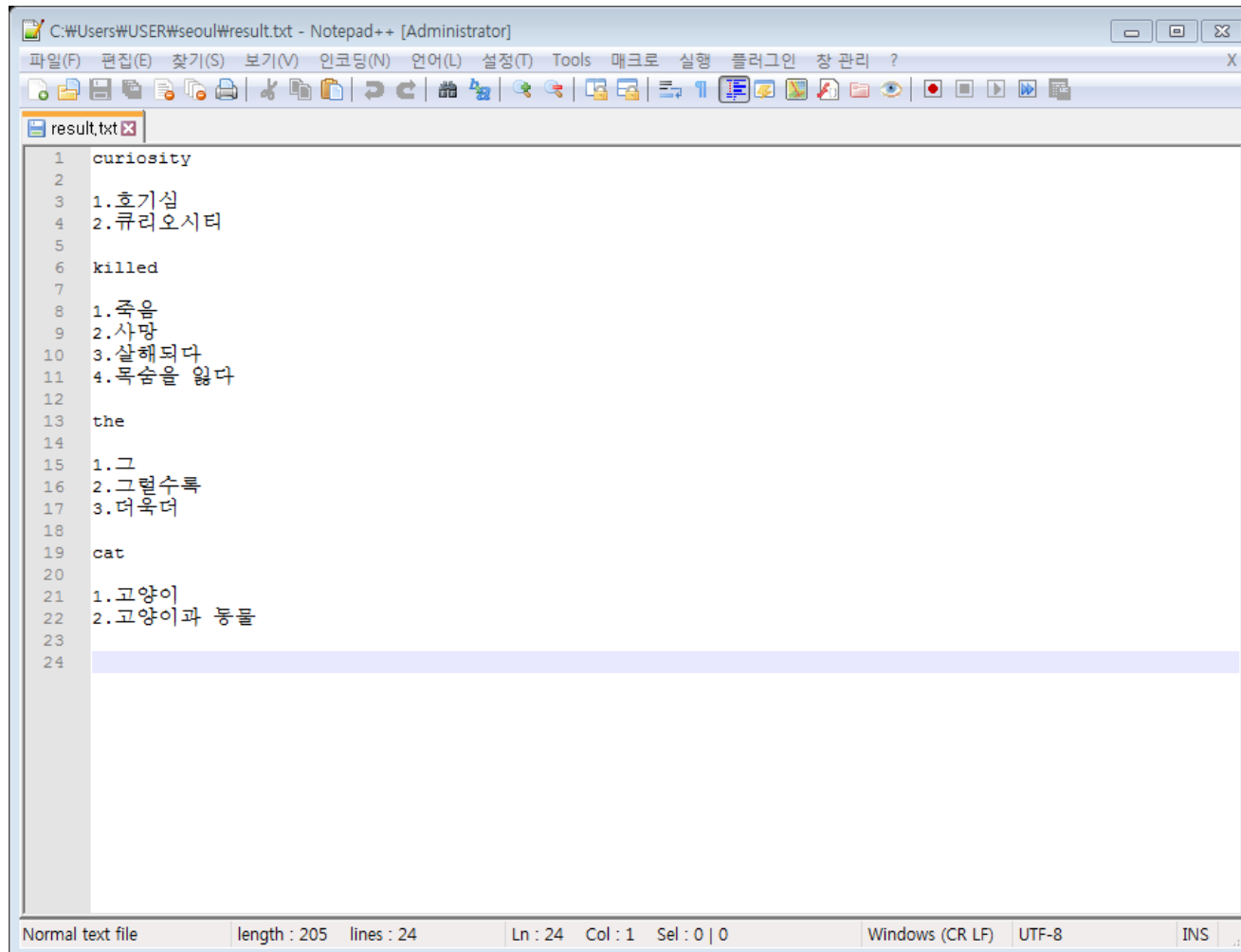
---

- TIPS

- 파이썬의 open 함수를 사용해 텍스트 파일로 저장한다
- 한글 저장을 위해 인코딩을 'utf-8'로 설정한다

# 실습 1-2-3. 여러 단어의 검색 결과 저장하기

## · 저장 결과



The screenshot shows a Notepad++ window titled "C:\Users\USER\seoul\result.txt - Notepad++ [Administrator]". The window contains a list of search results for the word "curiosity". The results are organized into four sections, each starting with a keyword followed by a list of related terms or translations. The keywords are "curiosity", "killed", "the", and "cat". The results for "curiosity" are: 1. 호기심, 2. 큐리오시티. The results for "killed" are: 1. 죽을, 2. 사망, 3. 살해되다, 4. 목숨을 잃다. The results for "the" are: 1. 그, 2. 그럴수록, 3. 더욱더. The results for "cat" are: 1. 고양이, 2. 고양이과 동물. The status bar at the bottom indicates "Normal text file", "length : 205", "lines : 24", "Ln : 24", "Col : 1", "Sel : 0 | 0", "Windows (CR LF)", "UTF-8", and "INS".

```
1 curiosity
2
3 1. 호기심
4 2. 큐리오시티
5
6 killed
7
8 1. 죽을
9 2. 사망
10 3. 살해되다
11 4. 목숨을 잃다
12
13 the
14
15 1. 그
16 2. 그럴수록
17 3. 더욱더
18
19 cat
20
21 1. 고양이
22 2. 고양이과 동물
23
24
```

Normal text file    length : 205    lines : 24    Ln : 24    Col : 1    Sel : 0 | 0    Windows (CR LF)    UTF-8    INS

## 실습 1-2-4. 여러 단어를 불러와 검색 결과 저장하기

· 다음 어학사전(<http://dic.daum.net/index.do?dic=all>)에 텍스트 파일(data.txt)에 있는 각 단어를 검색하였을 때 출력 되는 화면에서 가장 상단의 결과를 텍스트 파일에 저장한다

- 영어 단어의 의미를 저장한다

The screenshot shows the Daum Dictionary interface. At the top, there is a search bar with the word 'satisfaction' entered. To the right of the search bar are links for '다른 사전' (Other dictionaries), '단어장' (Vocabulary), and '사전홈' (Dictionary home). Further right are '로그인' (Login) and '더보기' (More) buttons. Below the search bar, a horizontal list of related words is displayed: '관련 customer satisfy survey satisfactory dissatisfaction satisfied index'. The main content area is titled '영어사전' (English Dictionary). The word 'satisfaction' is highlighted in a green box. Below it, the definitions '1. 만족 2. 충족' are listed, followed by the phonetic transcriptions '미국 [sætsɪfæksjən] <>' and '영국 [sætsɪfæksjən] <>'. Below this, three related phrases are listed: 'self-satisfaction' (1. 자기 만족 2. 자만 3. 독선), 'express satisfaction' (만족감을 표시하다), and 'demand satisfaction' (배상을 요구하다). At the bottom of the page, there is a link '영어사전 더보기' (More English Dictionary).

## 실습 1-2-4. 여러 단어를 불러와 검색 결과 저장하기

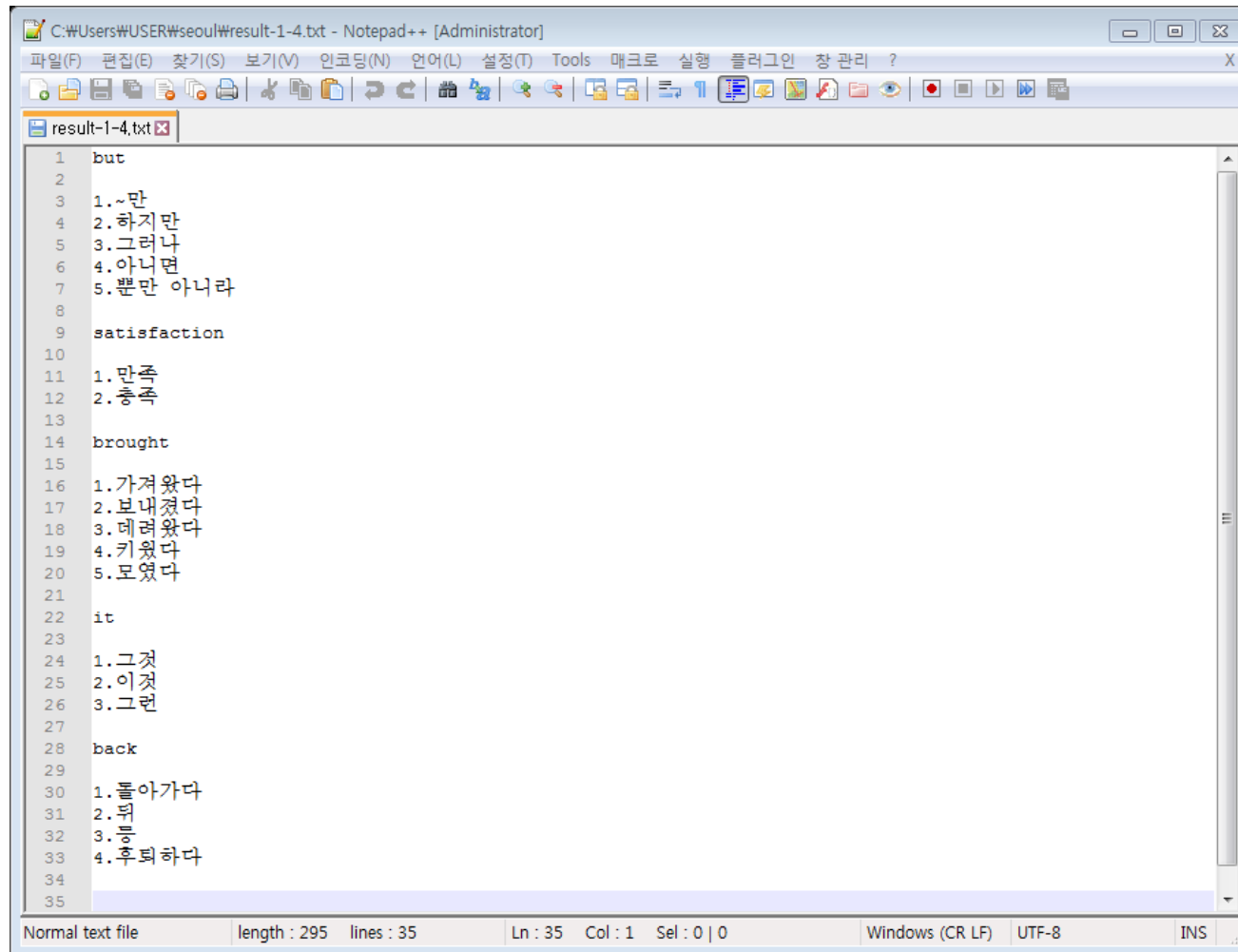
---

- TIPS

- 텍스트 파일의 내용을 불러올 때 `readlines` 함수를 사용한다
- 한글 저장을 위해 인코딩을 'utf-8'로 설정한다

# 실습 1-2-4. 여러 단어를 불러와 검색 결과 저장하기

## · 저장 결과



The screenshot shows a Notepad++ window titled "C:\Users\USER\seoul\result-1-4.txt - Notepad++ [Administrator]". The menu bar includes File (파일), Edit (편집), Search (찾기), View (보기), Encoding (인코딩), Language (언어), Settings (설정), Tools, Macro (매크로), Run (실행), Plugin (플러그인), and Window (창). The toolbar contains icons for file operations, editing, and searching. The text area displays the following content:

```
1 but
2
3 1.~만
4 2.하지만
5 3.그러나
6 4.아니면
7 5.뿐만 아니라
8
9 satisfaction
10
11 1.만족
12 2.충족
13
14 brought
15
16 1.가져왔다
17 2.보내쳤다
18 3.데려왔다
19 4.키웠다
20 5.모였다
21
22 it
23
24 1.그것
25 2.이것
26 3.그런
27
28 back
29
30 1.돌아가다
31 2.뒤
32 3.등
33 4.후퇴하다
34
35
```

The status bar at the bottom indicates "Normal text file", "length : 295", "lines : 35", "Ln : 35", "Col : 1", "Sel : 0 | 0", "Windows (CR LF)", "UTF-8", and "INS".

# 실습 3 – IMDb

AllIMDbPro | [Help](#)   

[Movies, TV & Showtimes](#) [Celebs, Events & Photos](#) [News & Community](#) [Watchlist](#) [Sign in with Facebook](#) [Other Sign in options](#)


[FULL CAST AND CREW](#) | [TRIVIA](#) | [USER REVIEWS](#) | [IMDbPro](#) | [MORE](#) [SHARE](#)



## Pulp Fiction (1994)

18 | 2h 34min | Crime, Drama | 10 September 1994 (South Korea)

 **8.9** /10  
1,426,334  [Rate This](#)



1:21 | [Trailer](#) | [7 VIDEOS](#) | [124 IMAGES](#)

### IMDb's Guide to Streaming



Check out our June Streaming Guide, find out if your favorite streaming show has been renewed or canceled, and more.

[Visit our Streaming Guide »](#)

# 실습 1-3-1. 영화 정보 출력하기

· IMDb 사이트에서 영화 펄프 픽션 페이지([http://www.imdb.com/title/tt0110912/?ref=mv\\_sr\\_1](http://www.imdb.com/title/tt0110912/?ref=mv_sr_1))에서 펄프 픽션 영화의 제목(title)과 감독(director), 그리고 출연 배우(cast)를 출력한다

FULL CAST AND CREW | TRIVIA | USER REVIEWS | IMDbPro | MORE | SHARE

+

**Pulp Fiction (1994)**

18 | 2h 34min | Crime, Drama | 10 September 1994 (South Korea)

★ 8.9 10

1,426,456

☆ Rate This



The lives of two mob hit men, a boxer, a gangster's wife, and a pair of diner bandits intertwine in four tales of violence and redemption.

**Director:** [Quentin Tarantino](#)

**Writers:** [Quentin Tarantino](#) (stories), [Roger Avary](#) (stories) | [1 more credit](#) »

**Stars:** [John Travolta](#), [Uma Thurman](#), [Samuel L. Jackson](#) | [See full cast & crew](#) »

94 Metascore

From metacritic.com

2,308 user | 220 critic

Popularity

154 (▲ 19)

**Top Rated Movies #7** | **Won 1 Oscar.** Another 60 wins & 68 nominations. [See more awards](#) »

Cast Edit

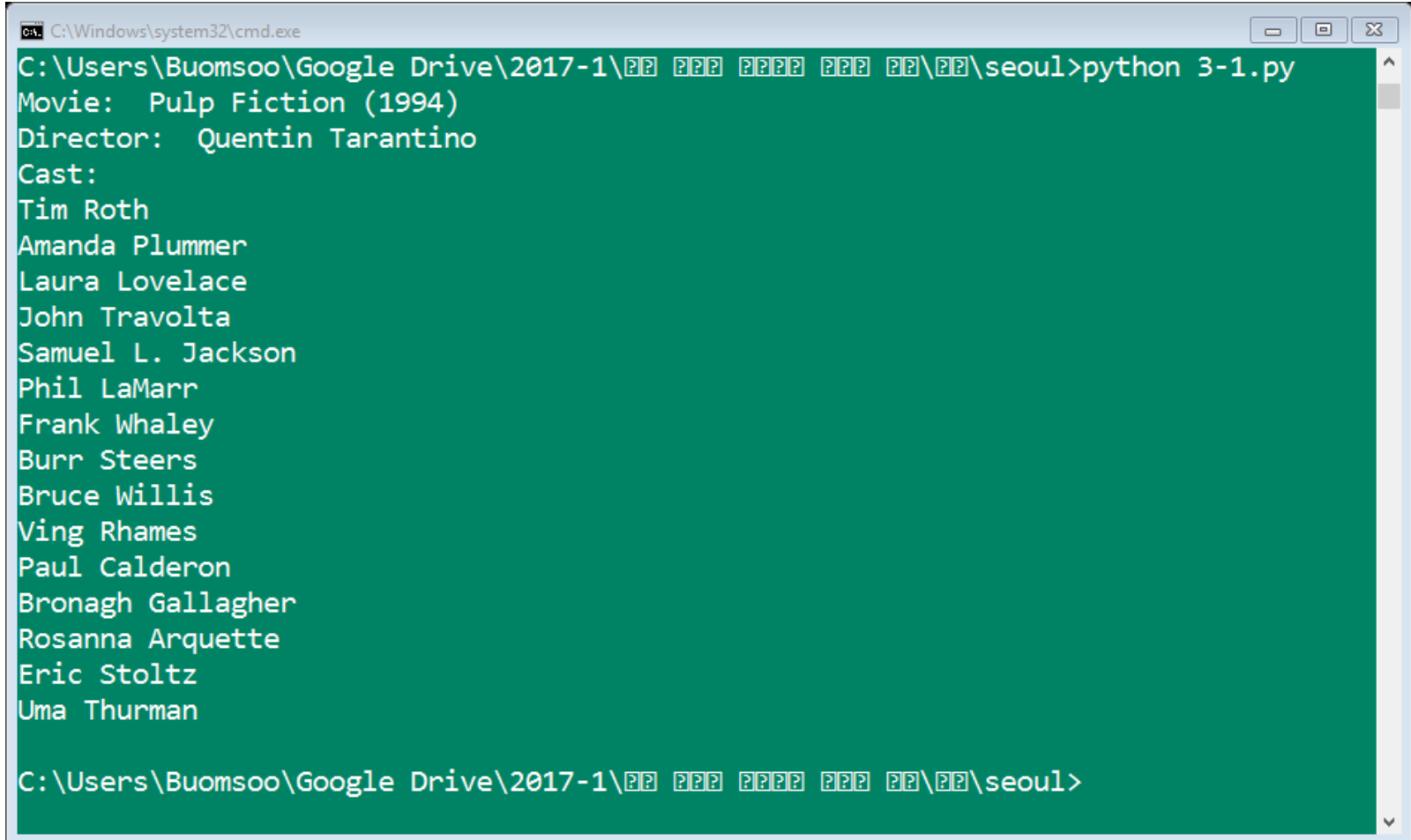
Cast overview, first billed only:

	<a href="#">Tim Roth</a>	...	<a href="#">Pumpkin</a>
	<a href="#">Amanda Plummer</a>	...	<a href="#">Honey Bunny</a>
	<a href="#">Laura Lovelace</a>	...	<a href="#">Waitress</a>
	<a href="#">John Travolta</a>	...	<a href="#">Vincent Vega</a>
	<a href="#">Samuel L. Jackson</a>	...	<a href="#">Jules Winnfield</a>
	<a href="#">Phil LaMarr</a>	...	<a href="#">Marvin</a>
	<a href="#">Frank Whaley</a>	...	<a href="#">Brett</a>
	<a href="#">Burr Steers</a>	...	<a href="#">Roger</a>
	<a href="#">Bruce Willis</a>	...	<a href="#">Butch Coolidge</a>
	<a href="#">Ving Rhames</a>	...	<a href="#">Marsellus Wallace</a>
	<a href="#">Paul Calderon</a>	...	<a href="#">Paul</a>
	<a href="#">Bronagh Gallagher</a>	...	<a href="#">Trudi</a>
	<a href="#">Rosanna Arquette</a>	...	<a href="#">Jody</a>
	<a href="#">Eric Stoltz</a>	...	<a href="#">Lance</a>
	<a href="#">Uma Thurman</a>	...	<a href="#">Mia Wallace</a>

[See full cast](#) »

## 실습 1-3-1. 영화 정보 출력하기

- 출력 결과



```
C:\Windows\system32\cmd.exe
C:\Users\Buomsoo\Google Drive\2017-1\?? ??? ???? ??? ??\??\seoul>python 3-1.py
Movie:  Pulp Fiction (1994)
Director:  Quentin Tarantino
Cast:
Tim Roth
Amanda Plummer
Laura Lovelace
John Travolta
Samuel L. Jackson
Phil LaMarr
Frank Whaley
Burr Steers
Bruce Willis
Ving Rhames
Paul Calderon
Bronagh Gallagher
Rosanna Arquette
Eric Stoltz
Uma Thurman

C:\Users\Buomsoo\Google Drive\2017-1\?? ??? ???? ??? ??\??\seoul>
```



## 실습 1-3-2. 여러 영화 정보 저장하기

· 영화 펄프 픽션(Pulp Fiction), 포레스트 검프(Forrest Gump), 그리고 다크 나이트(The Dark Knight)의 제목(title)과 감독 이름(director)를 텍스트 파일에 저장한다

- 영화 하나당 한 줄씩 출력되도록 한다

The image shows the IMDb page for the movie Pulp Fiction (1994). The title "Pulp Fiction (1994)" is highlighted with a green box. Below the title, the director "Quentin Tarantino" is also highlighted with a green box. The page includes a video player showing a trailer, a synopsis, and various ratings and awards.

FULL CAST AND CREW | TRIVIA | USER REVIEWS | IMDbPro | MORE | SHARE

+ **Pulp Fiction (1994)** ★ 8.9 <sup>10</sup> 1,426,456 ☆ Rate This

18 | 2h 34min | Crime, Drama | 10 September 1994 (South Korea)

**PULP FICTION**  
WINNER BEST PICTURE - ACADEMY AWARDS (1995)  
10+  
JOHN TRAVOLTA  
SAMUEL L. JACKSON  
UMA THURMAN  
BRITNEY SPOFFORD  
TIM ROTH  
AMANDA PLAMACK  
NADIA D. VICENTINI  
TOMMY LIAISON  
ERIC STROUT  
ROSARINA ARNETTE  
CHRISTOPHER MALKIN  
BY BRUCE WILKES

1:21 | Trailer | 7 VIDEOS | 124 IMAGES

The lives of two mob hit men, a boxer, a gangster's wife, and a pair of diner bandits intertwine in four tales of violence and redemption.

**Director:** [Quentin Tarantino](#)

**Writers:** [Quentin Tarantino](#) (stories), [Roger Avary](#) (stories) | 1 more credit »

**Stars:** [John Travolta](#), [Uma Thurman](#), [Samuel L. Jackson](#) | See full cast & crew »

94 Metascore From metacritic.com | Reviews 2,308 user | 220 critic | Popularity 154 (▲ 19)

**Top Rated Movies #7** | Won 1 Oscar. Another 60 wins & 68 nominations. See more awards »

## 실습 1-3-2. 여러 영화 정보 저장하기

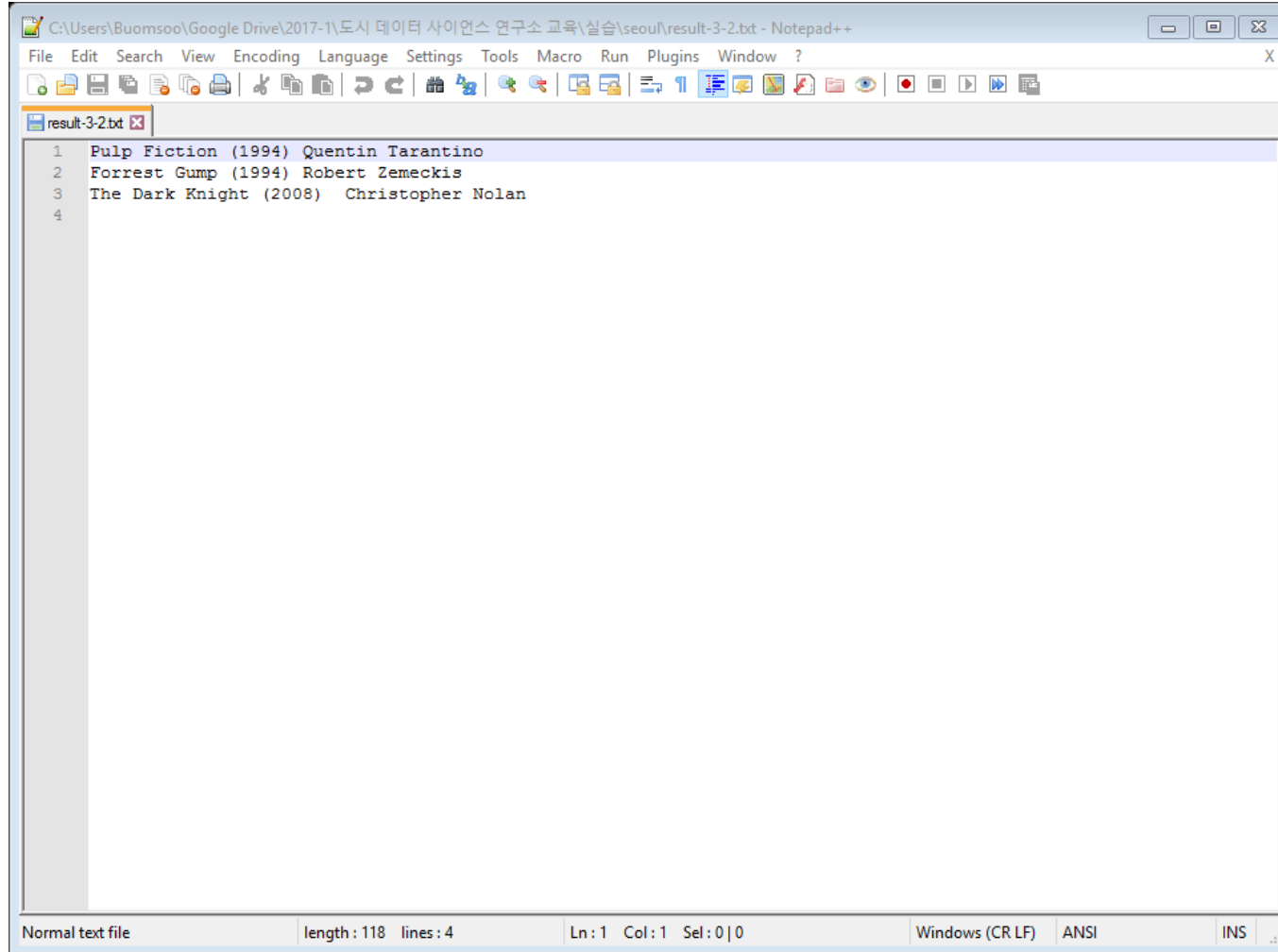
### · TIPS

- 영화가 달라짐에 따라 각 영화 정보의 URL이 어떻게 변하는지 잘 살펴본다



## 실습 1-3-2. 여러 영화 정보 저장하기

· 저장 결과



```
C:\Users\Buomsoo\Google Drive\2017-1\도시 데이터 사이언스 연구소 교육\실습\seoul\result-3-2.txt - Notepad++  
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?  
result-3-2.txt  
1 Pulp Fiction (1994) Quentin Tarantino  
2 Forrest Gump (1994) Robert Zemeckis  
3 The Dark Knight (2008) Christopher Nolan  
4  
Normal text file length: 118 lines: 4 Ln: 1 Col: 1 Sel: 0 | 0 Windows (CR LF) ANSI INS
```

# 실습 1-3-3. 영화 리뷰 출력하기

- 영화 다크 나이트(The Dark Knight)의 리뷰 페이지에 들어가 첫 번째 리뷰 내용을 출력해 보자

The screenshot shows the IMDb website interface. At the top, there's a search bar and navigation links. The main content area is titled "Reviews & Ratings for The Dark Knight". It includes a "Write review" button and a filter set to "Best". The page number is "Page 1 of 492". A review by "straightblaster" from the United States, dated 9 July 2008, is highlighted with a green border. The review text is: "We've been subjected to enormous amounts of hype and marketing for the Dark Knight. We've seen Joker scavenger hunts and one of the largest viral campaigns in advertising history and it culminates with the actual release of the movie. Everything that's been said is pretty much spot on. This is the first time I can remember where a summer blockbuster film far surpasses the hype. For as much action as there is in this movie, it's the acting that makes it a great piece of work. Between all the punches, explosions and stunt-work is some great dialog work. All the actors have their moments. Bale's Batman is the definitive Batman because we see everything in this character finally on film. Martial arts skills, cunning, great tactical thinking, forensic application, technological genius to advance or improve Luscious Fox's inventions/technological breakthroughs, intimidating personality, and even a little swashbuckling. As for Heath, yes he gets credit for his performance as the Joker. But you have to also recognize Jonathan and Chris Nolan for the writing and treatment of the character. It's not just the fact that Ledger makes the Joker so menacing, but the Nolans have given the character this great manifesto that drives its actions. The Joker's stance on chaos, order, anarchy, the morality of the average modern human being make the character so interesting psychologically. The Nolans drafted a complex character and only a perfect performance could've pulled something like this off. That's how difficult of a role this was, and that's why Ledger's performance is so great. This isn't an action movie. It's a film that explores literary themes of the hero and villain, as well as order and anarchy. Yes, listen to the dialog because it's all in there." Below the review, there's a "Was the above review useful to you?" section with "Yes" and "No" buttons. The page also includes a sidebar with "Quicklinks", "Top Links", "Overview", "Awards & Reviews", and "Plot & Quotes".

IMDb > The Dark Knight (2008) > Reviews & Ratings - IMDb

Reviews & Ratings for  
**The Dark Knight** [More at IMDbPro »](#)

[Write review](#)

Filter:  Hide Spoilers: ☐

Page 1 of 492: [\[1\]](#) [\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#) [\[6\]](#) [\[7\]](#) [\[8\]](#) [\[9\]](#) [\[10\]](#) [\[11\]](#) ▶

[Index](#) 4913 reviews in total

3117 out of 4092 people found the following review useful:

**Film surpasses the hype**  
★★★★★  
Author: [straightblaster](#) from United States  
9 July 2008

We've been subjected to enormous amounts of hype and marketing for the Dark Knight. We've seen Joker scavenger hunts and one of the largest viral campaigns in advertising history and it culminates with the actual release of the movie.

Everything that's been said is pretty much spot on. This is the first time I can remember where a summer blockbuster film far surpasses the hype.

For as much action as there is in this movie, it's the acting that makes it a great piece of work. Between all the punches, explosions and stunt-work is some great dialog work. All the actors have their moments.

Bale's Batman is the definitive Batman because we see everything in this character finally on film. Martial arts skills, cunning, great tactical thinking, forensic application, technological genius to advance or improve Luscious Fox's inventions/technological breakthroughs, intimidating personality, and even a little swashbuckling.

As for Heath, yes he gets credit for his performance as the Joker. But you have to also recognize Jonathan and Chris Nolan for the writing and treatment of the character. It's not just the fact that Ledger makes the Joker so menacing, but the Nolans have given the character this great manifesto that drives its actions. The Joker's stance on chaos, order, anarchy, the morality of the average modern human being make the character so interesting psychologically. The Nolans drafted a complex character and only a perfect performance could've pulled something like this off. That's how difficult of a role this was, and that's why Ledger's performance is so great.

This isn't an action movie. It's a film that explores literary themes of the hero and villain, as well as order and anarchy. Yes, listen to the dialog because it's all in there.

Was the above review useful to you?

2196 out of 3207 people found the following review useful:

**Film surpasses "Batman" in every aspect!!!**

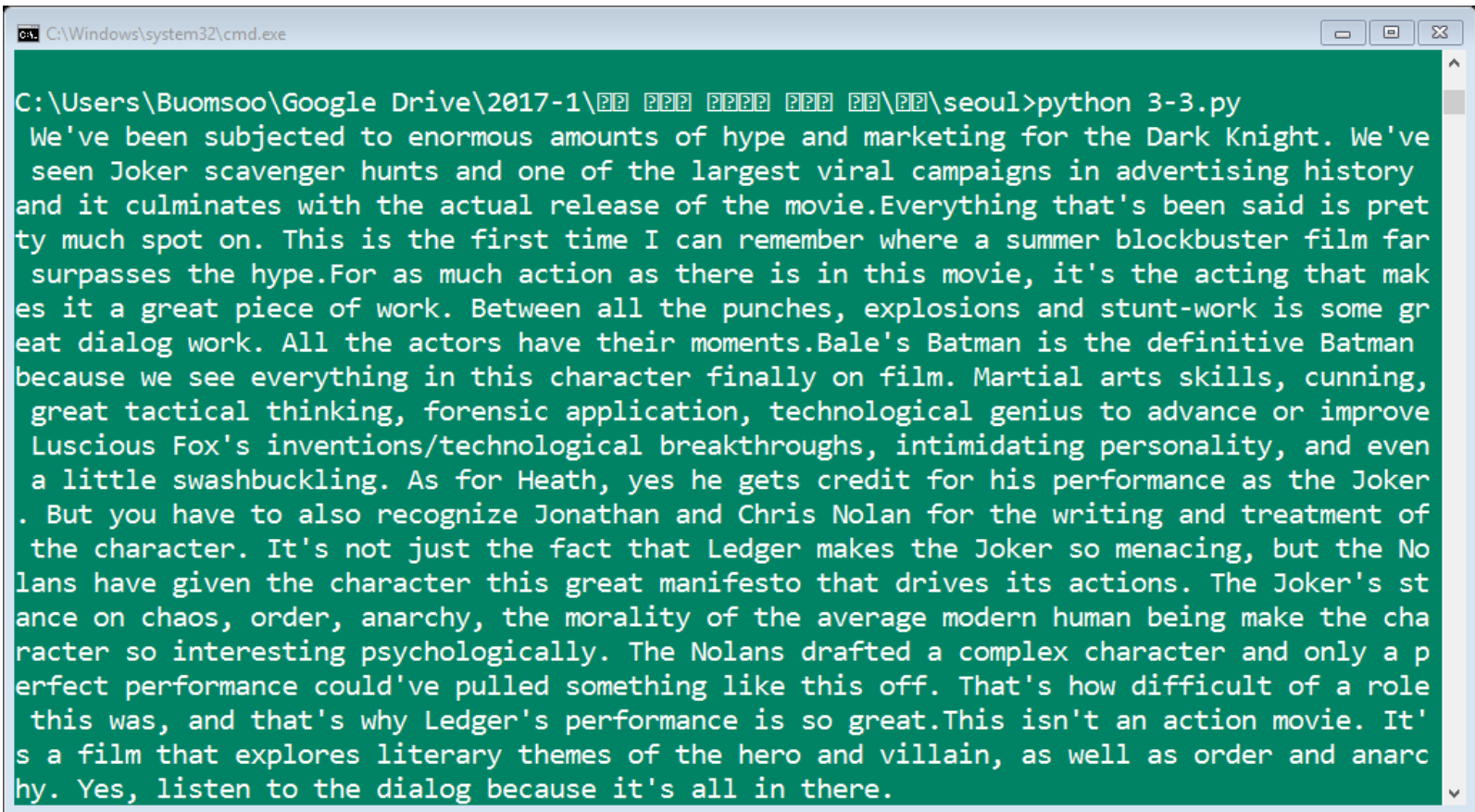
## 실습 1-3-3. 영화 리뷰 출력하기

### · TIPS

- Table의 구성요소 중 'p' 태그를 가진 요소를 모두 찾는다
- 'p' 태그 중 내용에 'This review may contain spoilers' 및 'Add another review'가 포함된 태그를 제외하고 출력한다
- if문과 break를 활용해 첫 번째 리뷰만 출력한다
- replace함수를 이용해 불필요한 공백을 제거한다('\r', '\n' 등)

## 실습 1-3-3. 영화 리뷰 출력하기

- 출력 결과



```
C:\Windows\system32\cmd.exe

C:\Users\Buomsoo\Google Drive\2017-1\?? ??? ??? ???? ??? ??\seoul>python 3-3.py
We've been subjected to enormous amounts of hype and marketing for the Dark Knight. We've
seen Joker scavenger hunts and one of the largest viral campaigns in advertising history
and it culminates with the actual release of the movie. Everything that's been said is pret
ty much spot on. This is the first time I can remember where a summer blockbuster film far
surpasses the hype. For as much action as there is in this movie, it's the acting that mak
es it a great piece of work. Between all the punches, explosions and stunt-work is some gr
eat dialog work. All the actors have their moments. Bale's Batman is the definitive Batman
because we see everything in this character finally on film. Martial arts skills, cunning,
great tactical thinking, forensic application, technological genius to advance or improve
Luscious Fox's inventions/technological breakthroughs, intimidating personality, and even
a little swashbuckling. As for Heath, yes he gets credit for his performance as the Joker
. But you have to also recognize Jonathan and Chris Nolan for the writing and treatment of
the character. It's not just the fact that Ledger makes the Joker so menacing, but the No
lans have given the character this great manifesto that drives its actions. The Joker's st
ance on chaos, order, anarchy, the morality of the average modern human being make the cha
racter so interesting psychologically. The Nolans drafted a complex character and only a p
erfect performance could've pulled something like this off. That's how difficult of a role
this was, and that's why Ledger's performance is so great. This isn't an action movie. It'
s a film that explores literary themes of the hero and villain, as well as order and anarc
hy. Yes, listen to the dialog because it's all in there.
```

# 실습 1-3-4. 영화 리뷰 저장하기

· 영화 다크 나이트(The Dark Knight)의 첫 번째 부터 10번째 페이지까지의 리뷰를 텍스트 파일에 저장한다

- 총 100개의 리뷰가 저장되어야 한다

IMDb > The Dark Knight (2008) > Reviews & Ratings - IMDb

Find Movies, TV shows, Celebrities and more... All

IMDbPro Help

Movies, TV & Showtimes Celebs, Events & Photos News & Community Watchlist Sign in with Facebook Other Sign in options

Reviews & Ratings for **The Dark Knight** [More at IMDbPro »](#)

[Write review](#)

Filter: **Best** Hide Spoilers: ☐

Page 1 of 492: [\[1\]](#) [\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#) [\[6\]](#) [\[7\]](#) [\[8\]](#) [\[9\]](#) [\[10\]](#) [\[11\]](#) [▶](#)

[Index](#) 4913 reviews in total

3117 out of 4092 people found the following review useful:

**Film surpasses the hype**  
★★★★★  
Author: [straightblaster](#) from United States  
9 July 2008

We've been subjected to enormous amounts of hype and marketing for the Dark Knight. We've seen Joker scavenger hunts and one of the largest viral campaigns in advertising history and it culminates with the actual release of the movie.

Everything that's been said is pretty much spot on. This is the first time I can remember where a summer blockbuster film far surpasses the hype.

For as much action as there is in this movie, it's the acting that makes it a great piece of work. Between all the punches, explosions and stunt-work is some great dialog work. All the actors have their moments.

Bale's Batman is the definitive Batman because we see everything in this character finally on film. Martial arts skills, cunning, great tactical thinking, forensic application, technological genius to advance or improve Luscious Fox's inventions/technological breakthroughs, intimidating personality, and even a little swashbuckling.

As for Heath, yes he gets credit for his performance as the Joker. But you have to also recognize Jonathan and Chris Nolan for the writing and treatment of the character. It's not just the fact that Ledger makes the Joker so menacing, but the Nolans have given the character this great manifesto that drives its actions. The Joker's stance on chaos, order, anarchy, the morality of the average modern human being make the character so interesting psychologically. The Nolans drafted a complex character and only a perfect performance could've pulled something like this off. That's how difficult of a role this was, and that's why Ledger's performance is so great.

This isn't an action movie. It's a film that explores literary themes of the hero and villain, as well as order and anarchy. Yes, listen to the dialog because it's all in there.

Was the above review useful to you? ☒ Yes ☐ No

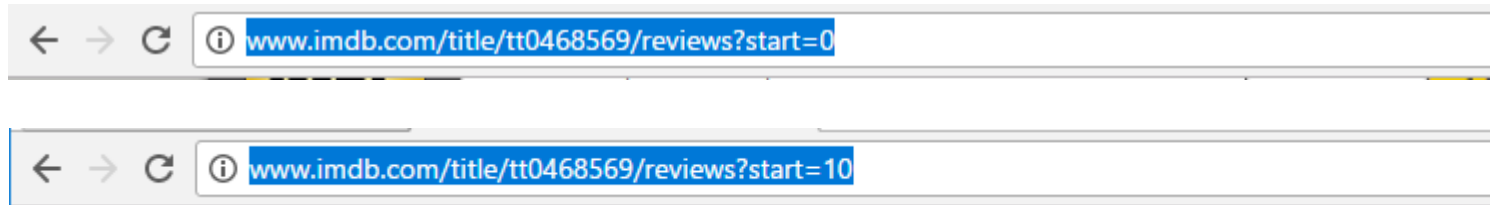
2196 out of 3207 people found the following review useful:

**Film surpasses the hype**  
★★★★★  
Author: [straightblaster](#) from United States  
9 July 2008

## 실습 1-3-4. 영화 리뷰 저장하기

### · TIPS

- 영화 리뷰 페이지의 페이지 번호를 바꾸어 가면서 URL 구조를 세심히 살펴보자

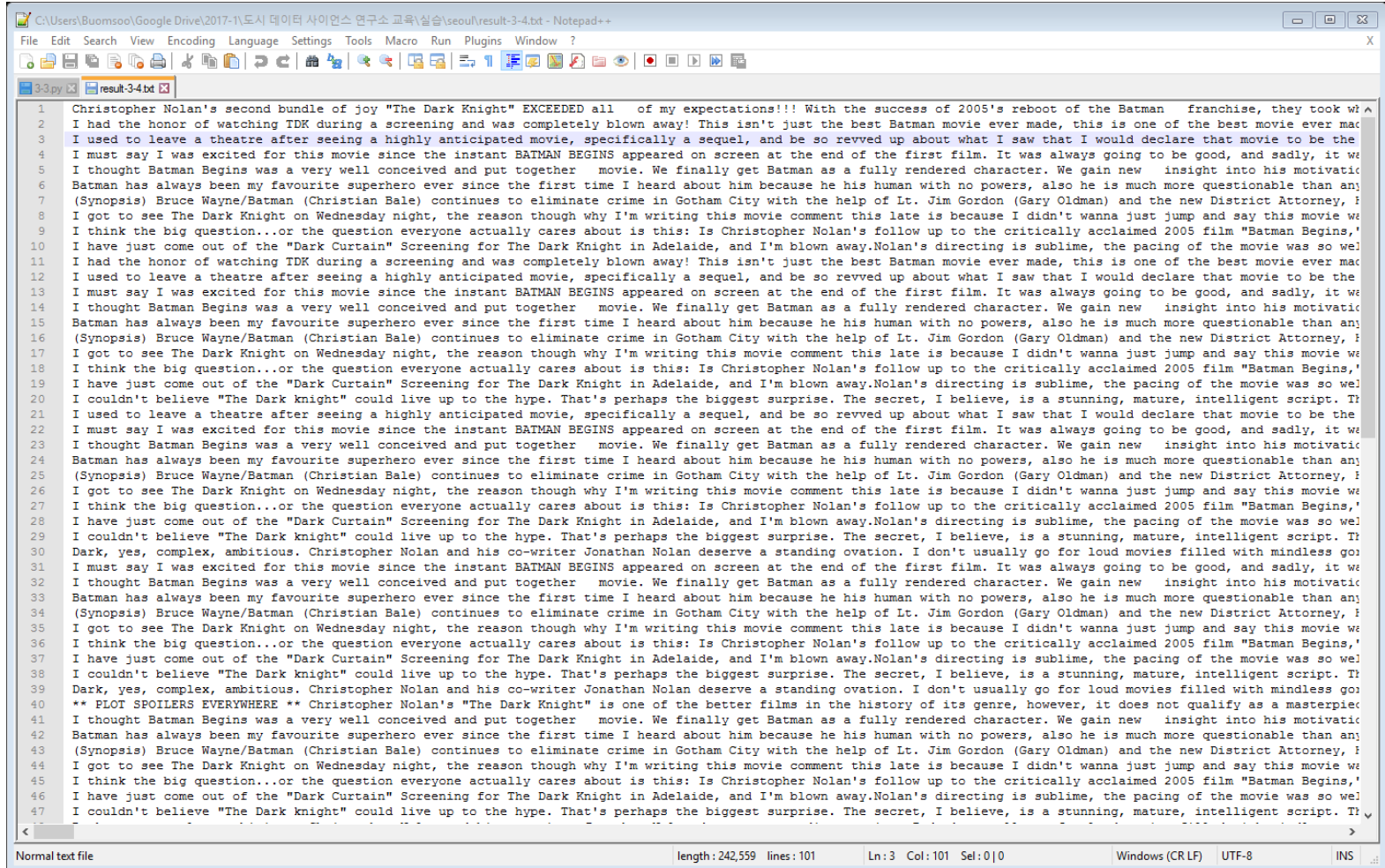


- While 문을 활용해 영화 리뷰를 페이지마다 모두 저장한다(while문 마지막에 구문 탈출 조건 명시)
- replace함수를 이용해 불필요한 공백을 제거한다('\r', '\n' 등)
- 파일을 저장할 때 인코딩(encoding) 설정을 꼭 '**utf-8**'으로 한다



# 실습 1-3-4. 영화 리뷰 저장하기

## · 저장 결과



```
C:\Users\Buomsoo\Google Drive\2017-1\토시 데이터 사이언스 연구소 교육\실습\seoul\result-3-4.txt - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
3-3.py result-3-4.txt
1 Christopher Nolan's second bundle of joy "The Dark Knight" EXCEEDED all of my expectations!!! With the success of 2005's reboot of the Batman franchise, they took w
2 I had the honor of watching TDK during a screening and was completely blown away! This isn't just the best Batman movie ever made, this is one of the best movie ever mac
3 I used to leave a theatre after seeing a highly anticipated movie, specifically a sequel, and be so revved up about what I saw that I would declare that movie to be the
4 I must say I was excited for this movie since the instant BATMAN BEGINS appeared on screen at the end of the first film. It was always going to be good, and sadly, it wa
5 I thought Batman Begins was a very well conceived and put together movie. We finally get Batman as a fully rendered character. We gain new insight into his motivati
6 Batman has always been my favourite superhero ever since the first time I heard about him because he his human with no powers, also he is much more questionable than an
7 (Synopsis) Bruce Wayne/Batman (Christian Bale) continues to eliminate crime in Gotham City with the help of Lt. Jim Gordon (Gary Oldman) and the new District Attorney, B
8 I got to see The Dark Knight on Wednesday night, the reason though why I'm writing this movie comment this late is because I didn't wanna just jump and say this movie wa
9 I think the big question...or the question everyone actually cares about is this: Is Christopher Nolan's follow up to the critically acclaimed 2005 film "Batman Begins,"
10 I have just come out of the "Dark Curtain" Screening for The Dark Knight in Adelaide, and I'm blown away.Nolan's directing is sublime, the pacing of the movie was so wel
11 I had the honor of watching TDK during a screening and was completely blown away! This isn't just the best Batman movie ever made, this is one of the best movie ever mac
12 I used to leave a theatre after seeing a highly anticipated movie, specifically a sequel, and be so revved up about what I saw that I would declare that movie to be the
13 I must say I was excited for this movie since the instant BATMAN BEGINS appeared on screen at the end of the first film. It was always going to be good, and sadly, it wa
14 I thought Batman Begins was a very well conceived and put together movie. We finally get Batman as a fully rendered character. We gain new insight into his motivati
15 Batman has always been my favourite superhero ever since the first time I heard about him because he his human with no powers, also he is much more questionable than an
16 (Synopsis) Bruce Wayne/Batman (Christian Bale) continues to eliminate crime in Gotham City with the help of Lt. Jim Gordon (Gary Oldman) and the new District Attorney, B
17 I got to see The Dark Knight on Wednesday night, the reason though why I'm writing this movie comment this late is because I didn't wanna just jump and say this movie wa
18 I think the big question...or the question everyone actually cares about is this: Is Christopher Nolan's follow up to the critically acclaimed 2005 film "Batman Begins,"
19 I have just come out of the "Dark Curtain" Screening for The Dark Knight in Adelaide, and I'm blown away.Nolan's directing is sublime, the pacing of the movie was so wel
20 I couldn't believe "The Dark knight" could live up to the hype. That's perhaps the biggest surprise. The secret, I believe, is a stunning, mature, intelligent script. Th
21 I used to leave a theatre after seeing a highly anticipated movie, specifically a sequel, and be so revved up about what I saw that I would declare that movie to be the
22 I must say I was excited for this movie since the instant BATMAN BEGINS appeared on screen at the end of the first film. It was always going to be good, and sadly, it wa
23 I thought Batman Begins was a very well conceived and put together movie. We finally get Batman as a fully rendered character. We gain new insight into his motivati
24 Batman has always been my favourite superhero ever since the first time I heard about him because he his human with no powers, also he is much more questionable than an
25 (Synopsis) Bruce Wayne/Batman (Christian Bale) continues to eliminate crime in Gotham City with the help of Lt. Jim Gordon (Gary Oldman) and the new District Attorney, B
26 I got to see The Dark Knight on Wednesday night, the reason though why I'm writing this movie comment this late is because I didn't wanna just jump and say this movie wa
27 I think the big question...or the question everyone actually cares about is this: Is Christopher Nolan's follow up to the critically acclaimed 2005 film "Batman Begins,"
28 I have just come out of the "Dark Curtain" Screening for The Dark Knight in Adelaide, and I'm blown away.Nolan's directing is sublime, the pacing of the movie was so wel
29 I couldn't believe "The Dark knight" could live up to the hype. That's perhaps the biggest surprise. The secret, I believe, is a stunning, mature, intelligent script. Th
30 Dark, yes, complex, ambitious. Christopher Nolan and his co-writer Jonathan Nolan deserve a standing ovation. I don't usually go for loud movies filled with mindless go
31 I must say I was excited for this movie since the instant BATMAN BEGINS appeared on screen at the end of the first film. It was always going to be good, and sadly, it wa
32 I thought Batman Begins was a very well conceived and put together movie. We finally get Batman as a fully rendered character. We gain new insight into his motivati
33 Batman has always been my favourite superhero ever since the first time I heard about him because he his human with no powers, also he is much more questionable than an
34 (Synopsis) Bruce Wayne/Batman (Christian Bale) continues to eliminate crime in Gotham City with the help of Lt. Jim Gordon (Gary Oldman) and the new District Attorney, B
35 I got to see The Dark Knight on Wednesday night, the reason though why I'm writing this movie comment this late is because I didn't wanna just jump and say this movie wa
36 I think the big question...or the question everyone actually cares about is this: Is Christopher Nolan's follow up to the critically acclaimed 2005 film "Batman Begins,"
37 I have just come out of the "Dark Curtain" Screening for The Dark Knight in Adelaide, and I'm blown away.Nolan's directing is sublime, the pacing of the movie was so wel
38 I couldn't believe "The Dark knight" could live up to the hype. That's perhaps the biggest surprise. The secret, I believe, is a stunning, mature, intelligent script. Th
39 Dark, yes, complex, ambitious. Christopher Nolan and his co-writer Jonathan Nolan deserve a standing ovation. I don't usually go for loud movies filled with mindless go
40 ** PLOT SPOILERS EVERYWHERE ** Christopher Nolan's "The Dark Knight" is one of the better films in the history of its genre, however, it does not qualify as a masterpiec
41 I thought Batman Begins was a very well conceived and put together movie. We finally get Batman as a fully rendered character. We gain new insight into his motivati
42 Batman has always been my favourite superhero ever since the first time I heard about him because he his human with no powers, also he is much more questionable than an
43 (Synopsis) Bruce Wayne/Batman (Christian Bale) continues to eliminate crime in Gotham City with the help of Lt. Jim Gordon (Gary Oldman) and the new District Attorney, B
44 I got to see The Dark Knight on Wednesday night, the reason though why I'm writing this movie comment this late is because I didn't wanna just jump and say this movie wa
45 I think the big question...or the question everyone actually cares about is this: Is Christopher Nolan's follow up to the critically acclaimed 2005 film "Batman Begins,"
46 I have just come out of the "Dark Curtain" Screening for The Dark Knight in Adelaide, and I'm blown away.Nolan's directing is sublime, the pacing of the movie was so wel
47 I couldn't believe "The Dark knight" could live up to the hype. That's perhaps the biggest surprise. The secret, I believe, is a stunning, mature, intelligent script. Th
Normal text file length: 242,559 lines: 101 Ln: 3 Col: 101 Sel: 0 | 0 Windows (CR LF) UTF-8 INS
```

# 실습 1-3-5. 영화 리뷰 저장하기(2)

· 자신이 좋아하는 영화를 3개 이상 골라 3-4의 코드를 조금씩 바꾸어 영화 리뷰를 각각 다른 텍스트 파일에 저장해 본다

- 리뷰를 영화 별로 최소한 100개 이상 수집한다

IMDb Find Movies, TV shows, Celebrities and more... All Q IMDbPro Help f t i

Movies, TV & Showtimes Celebs, Events & Photos News & Community Watchlist Sign in with Facebook Other Sign in options

IMDb > The Godfather: Part II (1974) > Reviews & Ratings - IMDb

Reviews & Ratings for  
**The Godfather: Part II** [More at IMDbPro »](#)

[Write review](#)

Filter: **Best** Hide Spoilers: ☐

Page 1 of 68: [\[1\]](#) [\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#) [\[6\]](#) [\[7\]](#) [\[8\]](#) [\[9\]](#) [\[10\]](#) [\[11\]](#) ▶

[Index](#) 679 reviews in total

---

405 out of 514 people found the following review useful:

**To call it a sequel is a travesty**  
★★★★★  
Author: [taimur74](#)  
9 May 2001

This movie is way to be good to be labelled a sequel to The Godfather . Rather it is more of a companion piece to the original and the two perfectly compliment each other . IT is both a sequel and prequel showing the rise of the young vito and moral decline of Micheal . Both characters are brought to life with uncanny ability by Robert DeNiro and Al Pacino . To say that these two are good actors is like saying that a nuclear bomb makes a loud noise and in this movie they prove why they are at the top of their respective crafts .

Al Pacino is the standout in the ensemble cast and its amazing how his eyes have changed from the first part . They are now cold , ruthless and unemotional and betray the price which Micheal Corleone has paid for power .

Watch this movie and learn why it is the greatest gangster film of all time.

Was the above review useful to you? [Yes](#) [No](#)

[Overview](#)




**Quicklinks**  
reviews ▼

**Top Links**  
- trailers and videos  
- full cast and crew  
- trivia  
- official sites  
- memorable quotes

# 실습 1-3-5. 영화 리뷰 저장하기(2)

## · 저장 결과

```
1 What is the most resilient parasite? An Idea! Yes, Nolan has created something with his unbelievably, incredibly and god- gifted mind
2 Usually I try to be careful with over hyping a film, or setting the expectations too high, as film geeks all are guilty for, however f
3 Films about dreams and the subconscious are usually not very straightforward and almost always weird. "Inception" is no exception to t
4 I saw Memento very recently, something that turned out to be a great miss.I saw it again, just to make a couple of thins straight-and
5 Dom Cobb leads a highly skilled team, specializing in stealing secrets from people's minds by entering their dreams. When they are hir
6 What is going on with the IMDB user reviews lately? It's like the masses can no longer be trusted. In the last month, the users have d
7 I'm nearly at a loss for words. Just when you thought Christopher Nolan couldn't follow up to "The Dark Knight", he does it again, del
8 Inception is truly one of a kind. A concept which has long gestated in Christopher Nolan's mind, his eye for drama mixed with his larg
9 What do we ask for in a summer blockbuster? What is it that incites hysteria this time every year for whatever dross the studios churn
10 I had the privilege of seeing this movie before it came out, and, like most of DiCaprio's films, it blew me away.Inception is a well-m
11 I cannot ignore facts. I cannot ignore mistakes and plot holes, bad writing, or bad marketing. I cannot ignore hype, especially when i
12 Films about dreams and the subconscious are usually not very straightforward and almost always weird. "Inception" is no exception to t
13 I saw Memento very recently, something that turned out to be a great miss.I saw it again, just to make a couple of thins straight-and
14 Dom Cobb leads a highly skilled team, specializing in stealing secrets from people's minds by entering their dreams. When they are hir
15 What is going on with the IMDB user reviews lately? It's like the masses can no longer be trusted. In the last month, the users have d
16 I'm nearly at a loss for words. Just when you thought Christopher Nolan couldn't follow up to "The Dark Knight", he does it again, del
17 Inception is truly one of a kind. A concept which has long gestated in Christopher Nolan's mind, his eye for drama mixed with his larg
18 What do we ask for in a summer blockbuster? What is it that incites hysteria this time every year for whatever dross the studios churn
19 I had the privilege of seeing this movie before it came out, and, like most of DiCaprio's films, it blew me away.Inception is a well-m
20 I cannot ignore facts. I cannot ignore mistakes and plot holes, bad writing, or bad marketing. I cannot ignore hype, especially when i
21 I saw Memento very recently, something that turned out to be a great miss.I saw it again, just to make a couple of thins straight-and
22 Dom Cobb leads a highly skilled team, specializing in stealing secrets from people's minds by entering their dreams. When they are hir
23 What is going on with the IMDB user reviews lately? It's like the masses can no longer be trusted. In the last month, the users have d
24 I'm nearly at a loss for words. Just when you thought Christopher Nolan couldn't follow up to "The Dark Knight", he does it again, del
25 Inception is truly one of a kind. A concept which has long gestated in Christopher Nolan's mind, his eye for drama mixed with his larg
26 What do we ask for in a summer blockbuster? What is it that incites hysteria this time every year for whatever dross the studios churn
27 I had the privilege of seeing this movie before it came out, and, like most of DiCaprio's films, it blew me away.Inception is a well-m
28 I cannot ignore facts. I cannot ignore mistakes and plot holes, bad writing, or bad marketing. I cannot ignore hype, especially when i
29 Its all a dream. How do you know its a dream? Because it got a 9.4 rating. I'm waiting for a train. Its bringing a real rating of a 4.
30 I saw Memento very recently, something that turned out to be a great miss.I saw it again, just to make a couple of thins straight-and
31 Dom Cobb leads a highly skilled team, specializing in stealing secrets from people's minds by entering their dreams. When they are hir
32 What is going on with the IMDB user reviews lately? It's like the masses can no longer be trusted. In the last month, the users have d
33 I'm nearly at a loss for words. Just when you thought Christopher Nolan couldn't follow up to "The Dark Knight", he does it again, del
34 Inception is truly one of a kind. A concept which has long gestated in Christopher Nolan's mind, his eye for drama mixed with his larg
35 What do we ask for in a summer blockbuster? What is it that incites hysteria this time every year for whatever dross the studios churn
36 I had the privilege of seeing this movie before it came out, and, like most of DiCaprio's films, it blew me away.Inception is a well-m
37 I cannot ignore facts. I cannot ignore mistakes and plot holes, bad writing, or bad marketing. I cannot ignore hype, especially when i
38 Its all a dream. How do you know its a dream? Because it got a 9.4 rating. I'm waiting for a train. Its bringing a real rating of a 4.
39 I have to say to make such an impressive trailer and such an uninteresting film, takes some doing.Here you have most of the elements t
40 Dom Cobb leads a highly skilled team, specializing in stealing secrets from people's minds by entering their dreams. When they are hir
41 What is going on with the IMDB user reviews lately? It's like the masses can no longer be trusted. In the last month, the users have d
42 I'm nearly at a loss for words. Just when you thought Christopher Nolan couldn't follow up to "The Dark Knight", he does it again, del
43 Inception is truly one of a kind. A concept which has long gestated in Christopher Nolan's mind, his eye for drama mixed with his larg
44 What do we ask for in a summer blockbuster? What is it that incites hysteria this time every year for whatever dross the studios churn
45 I had the privilege of seeing this movie before it came out, and, like most of DiCaprio's films, it blew me away.Inception is a well-m
46 I cannot ignore facts. I cannot ignore mistakes and plot holes, bad writing, or bad marketing. I cannot ignore hype, especially when i
```

 result-3-5-inception	21/06/2017 17:40	Text Document	281 KB
 result-3-5-old_boy	21/06/2017 17:42	Text Document	181 KB
 result-3-5-whiplash	21/06/2017 17:41	Text Document	216 KB