

Web Crawling 실습

비정형 데이터 분석을 통한 효율적인 의사결정

박진수 교수

**Big Data Institute,
Seoul National University**

실습 1 – 다음 사전

dmm 사전

백과사전 | 단어장 | 작은창 | 맞춤법검사기 *beta*

로그인 더보기 ▾

어학사전

맞춤법 검사기

curiosity



A가

Aa

가

あ

中

漢

...

실습 1-1. 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 its Korean meanings '1. 호기심' and '2. 큐리오시티' listed below it. The phonetic transcriptions for American and British English are also provided: '미국 [kjüeriósəti] <≫> 영국 [kjüeriósəti] <≫>'. To the right of the word is a blue circular button with a white plus sign. Below the word, there are three phrases: 'unquenchable curiosity' (억누를 수 없는 호기심), 'insatiable curiosity' (만족을 모르는 호기심), and 'satisfy curiosity' (호기심을 만족시키다).

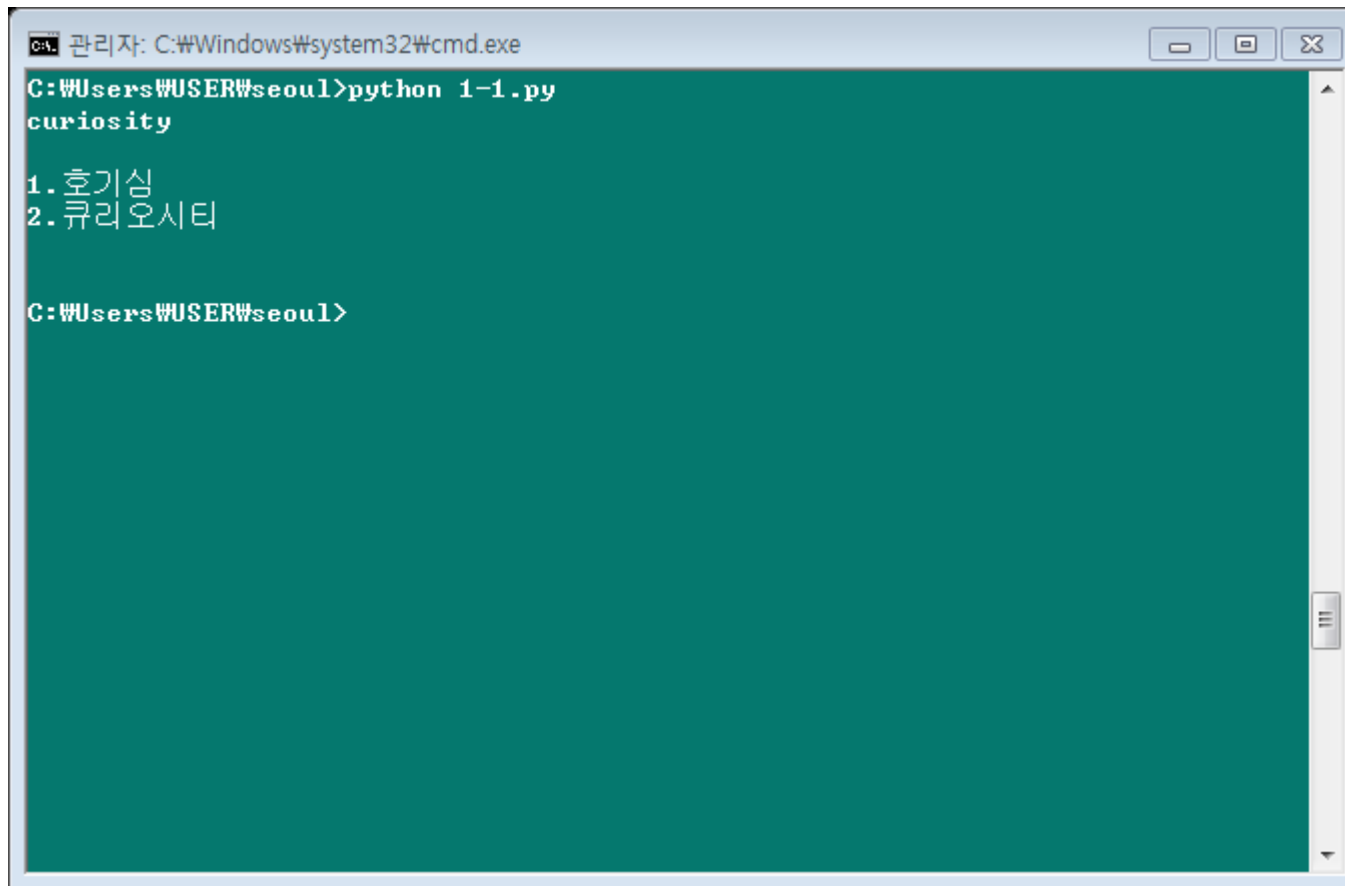
실습 1-1. 단어의 검색 결과 출력하기

· 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 1-1. 단어의 검색 결과 출력하기

- 출력 결과



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

1. 호기심
2. 큐리오시티

C:\Users\USER\seoul>
```

The image shows a Windows command prompt window titled "관리자: C:\Windows\system32\cmd.exe". The prompt is "C:\Users\USER\seoul>". The user has entered the command "python 1-1.py". The output of the script is "curiosity" followed by a blank line, then a numbered list: "1. 호기심" and "2. 큐리오시티". The prompt "C:\Users\USER\seoul>" is shown again at the bottom.

실습 1-2. 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

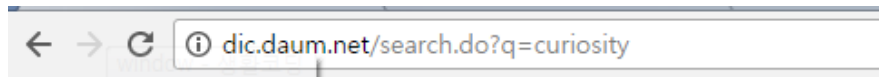
- 영어 단어의 의미를 출력한다

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 features the word 'curiosity' in large blue text, with a green box highlighting it. Below the word, it lists '1. 호기심' and '2. 큐리오시티'. The phonetic transcriptions are given as '미국 [kjūeriásəti] <≫> 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. 여러 단어의 검색 결과 출력하기

· TIPS

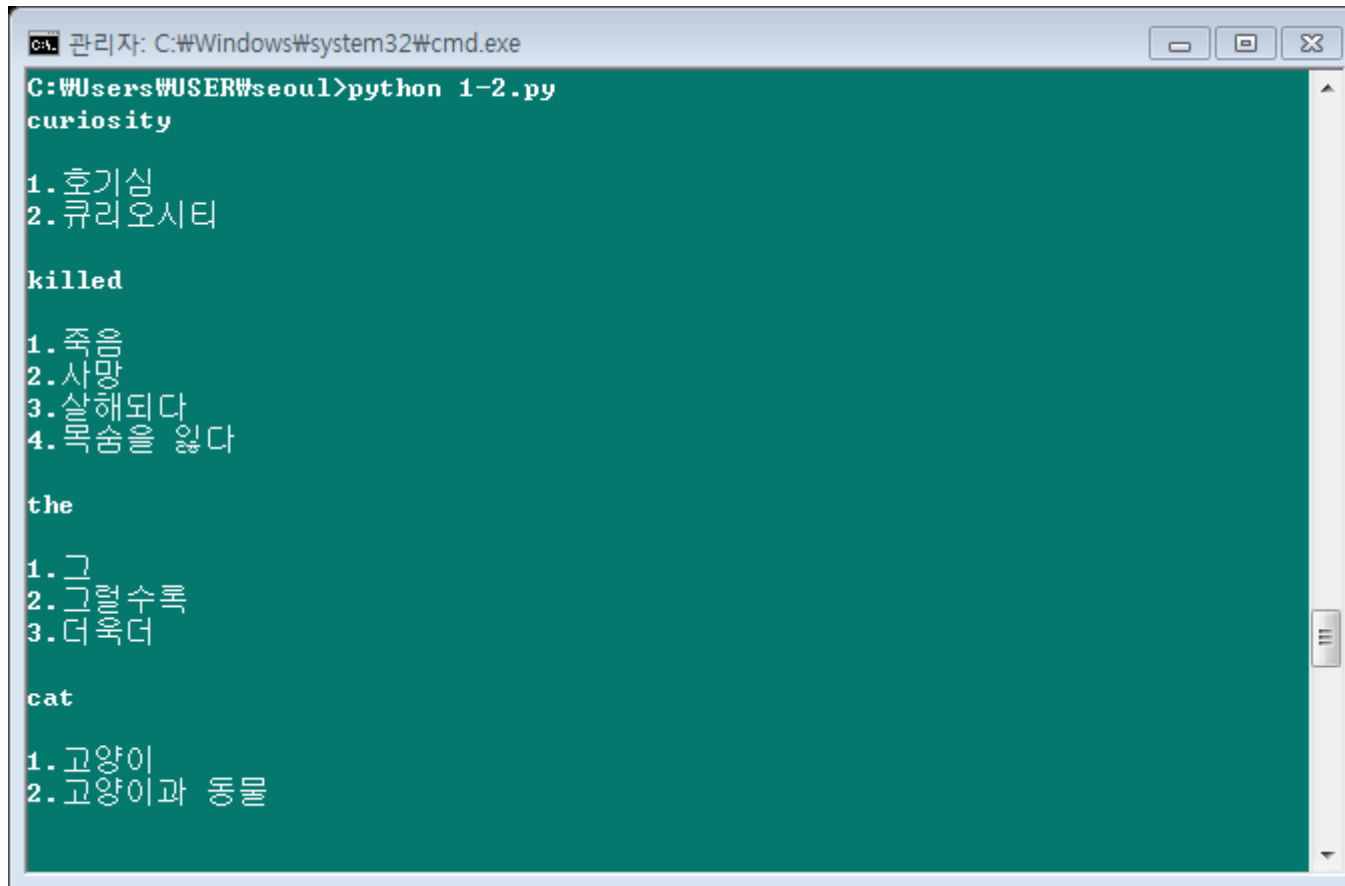
- 초기 화면에서 몇 개 단어를 검색해 보면서 URL 구조를 잘 생각해 본다



- for 문을 활용해 여러 단어를 돌아가면서 데이터를 출력한다

실습 1-2. 여러 단어의 검색 결과 출력하기

- 출력 결과



```
C:\Windows\system32\cmd.exe
C:\Users\USER>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 1-3. 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 transcriptions for American and British English: '미국 [kjūəriəsəti] <≫> 영국 [kjūəriəsəti] <≫>'. 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).

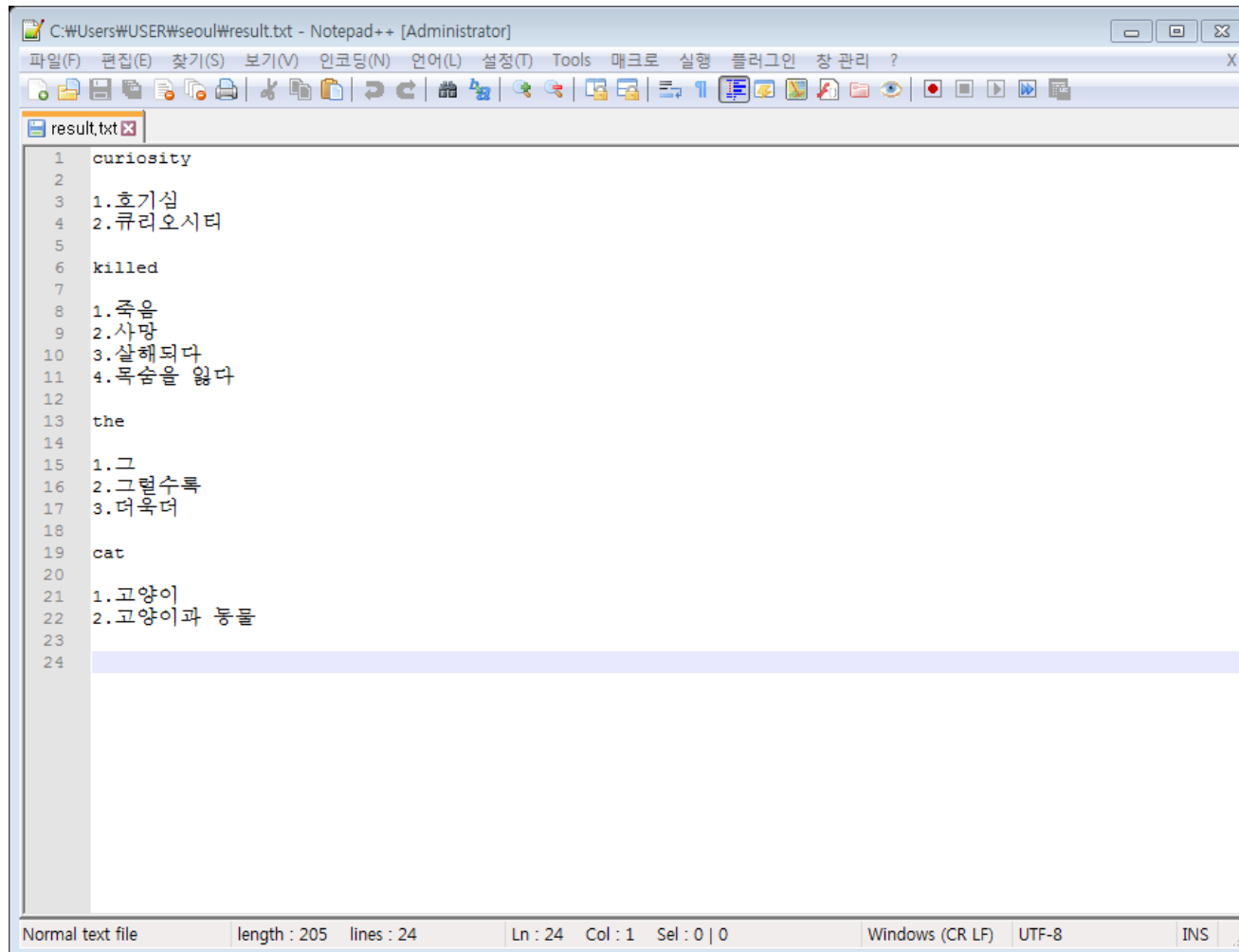
실습 1-3. 여러 단어의 검색 결과 저장하기

· TIPS

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

실습 1-3. 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 four different words, each with its English and Korean meanings. The results are as follows:

Word	Meanings
curiosity	1. 호기심 2. 큐리오시티
killed	1. 죽을 2. 사망 3. 살해되다 4. 목숨을 잃다
the	1. 그 2. 그럴수록 3. 더욱더
cat	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, INS.

실습 1-4. 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 with a green box. Below it, the definitions '1. 만족 2. 충족' are listed, followed by the phonetic transcriptions '미국 [sætɪsfækjən] <> 영국 [sætɪsfækjən] <>'. Below this, three related phrases are shown: 'self-satisfaction' with definitions '1. 자기 만족 2. 자만 3. 독선' and phonetic transcriptions '미국 <> 영국 <>'; 'express satisfaction' with the definition '만족감을 표시하다'; and 'demand satisfaction' with the definition '배상을 요구하다'. At the bottom of the page, there is a link '영어사전 더보기' (More English Dictionary).

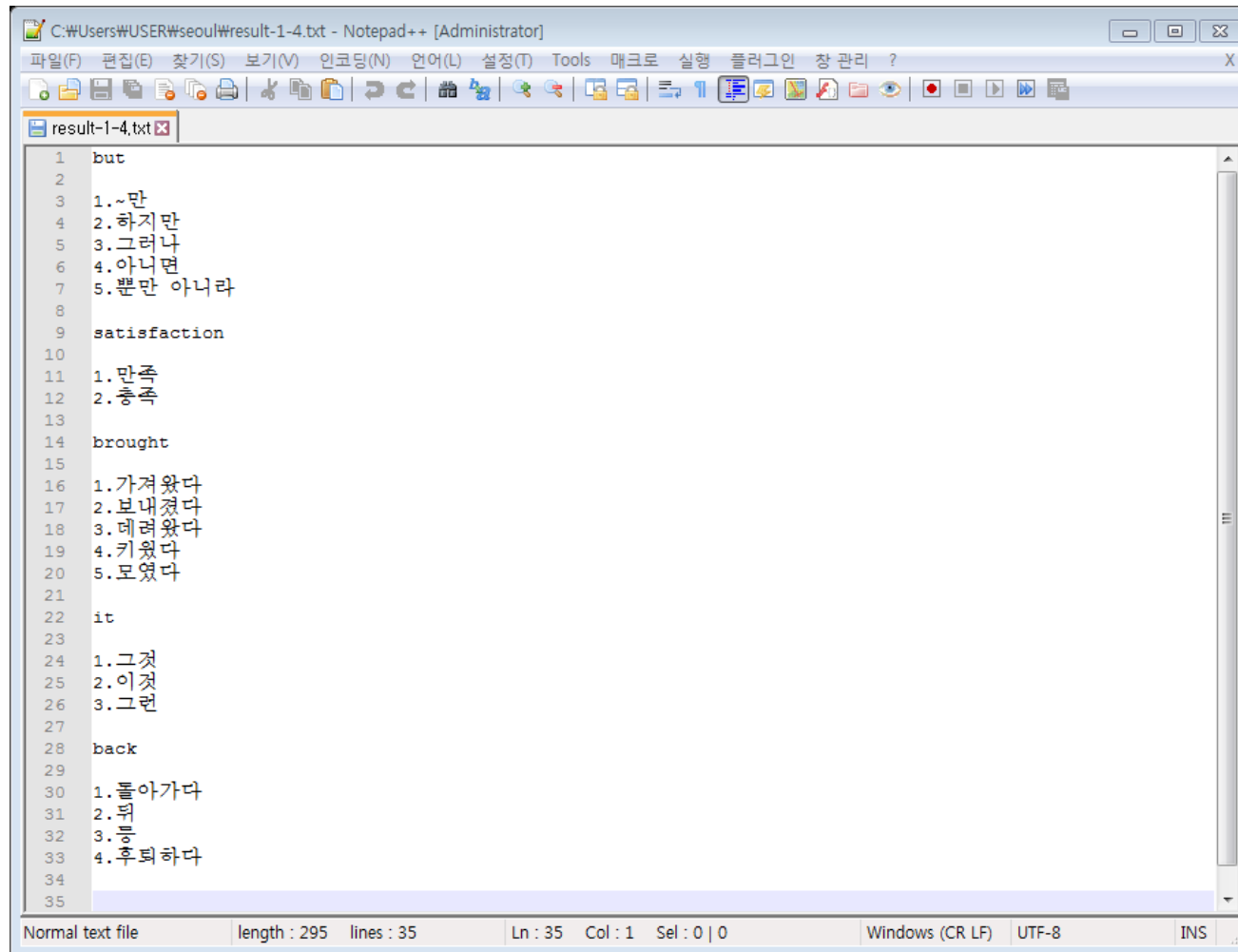
실습 1-4. 여러 단어를 불러와 검색 결과 저장하기

· TIPS

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

실습 1-4. 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 (F), Edit (E), Search (S), View (V), Encoding (N), Language (L), Settings (T), Tools, Macro, Run, Plugin, and Window (W). The toolbar contains icons for file operations, editing, and viewing. 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 shows: Normal text file, length : 295, lines : 35, Ln : 35, Col : 1, Sel : 0 | 0, Windows (CR LF), UTF-8, and INS.

실습 2 – Coinmarketcap.com

762 Currencies / 120 Assets / 4165 Markets

Market Cap: **\$114,344,076,944** / 24h Vol: **\$5,221,107,365** / BTC Dominance: **39.0%**

48

CryptoCurrency Market Capitalizations



5% Affiliate Program & Advanced Referral Tracking

Market Cap ▾ Trade Volume ▾ Trending ▾ Tools ▾

Search Currencies



All ▾

Currencies ▾

Assets ▾

USD ▾

Next 100 →

View All

#	Name	Market Cap	Price	Circulating Supply	Volume (24h)	% Change (24h)	Price Graph (7d)
1	Bitcoin	\$44,501,255,243	\$2713.09	16,402,425 BTC	\$1,758,920,000	3.59%	
2	Ethereum	\$32,993,789,011	\$356.07	92,660,703 ETH	\$979,084,000	-4.60%	

실습 2-1. 비트코인 관련 데이터 출력하기

· 코인마켓캡(<https://coinmarketcap.com/>) 메인 페이지에서 가장 상단에 있는 비트코인의 화폐 이름 (Name)과 시가총액(Market Cap), 개당 가격(Price), 그리고 24시간 볼륨(Volume)을 출력해 본다

- 한 줄에 하나씩 출력되도록 한다

CryptoCurrency Market Capitalizations							
<div>24/7 Support Team Dedicated Client Support</div> <div>Unique & Sophisticated User Interface</div>							
Market Cap ▾ Trade Volume ▾ Trending ▾ Tools ▾				Search Currencies		Q	
All ▾	Currencies ▾	Assets ▾	USD ▾	Next 100 →		View All	
#	Name	Market Cap	Price	Circulating Supply	Volume (24h)	% Change (24h)	Price Graph (7d)
1	Bitcoin	\$44,501,255,243	\$2713.09	16,402,425 BTC	\$1,758,920,000	3.59%	
2	Ethereum	\$32,993,789,011	\$356.07	92,660,703 ETH	\$979,084,000	-4.60%	
3	Ripple	\$12,329,543,959	\$0.322002	38,290,271,363 XRP *	\$762,189,000	10.12%	
4	Litecoin	\$2,434,948,352	\$47.15	51,646,757 LTC	\$575,602,000	-6.08%	
5	Ethereum Classic	\$2,018,384,427	\$21.75	92,799,284 ETC	\$153,872,000	-5.35%	
6	NEM	\$1,781,676,000	\$0.197964	8,999,999,999 XEM *	\$17,064,100	1.58%	
7	Dash	\$1,370,452,365	\$185.83	7,374,843 DASH	\$54,122,900	-7.68%	

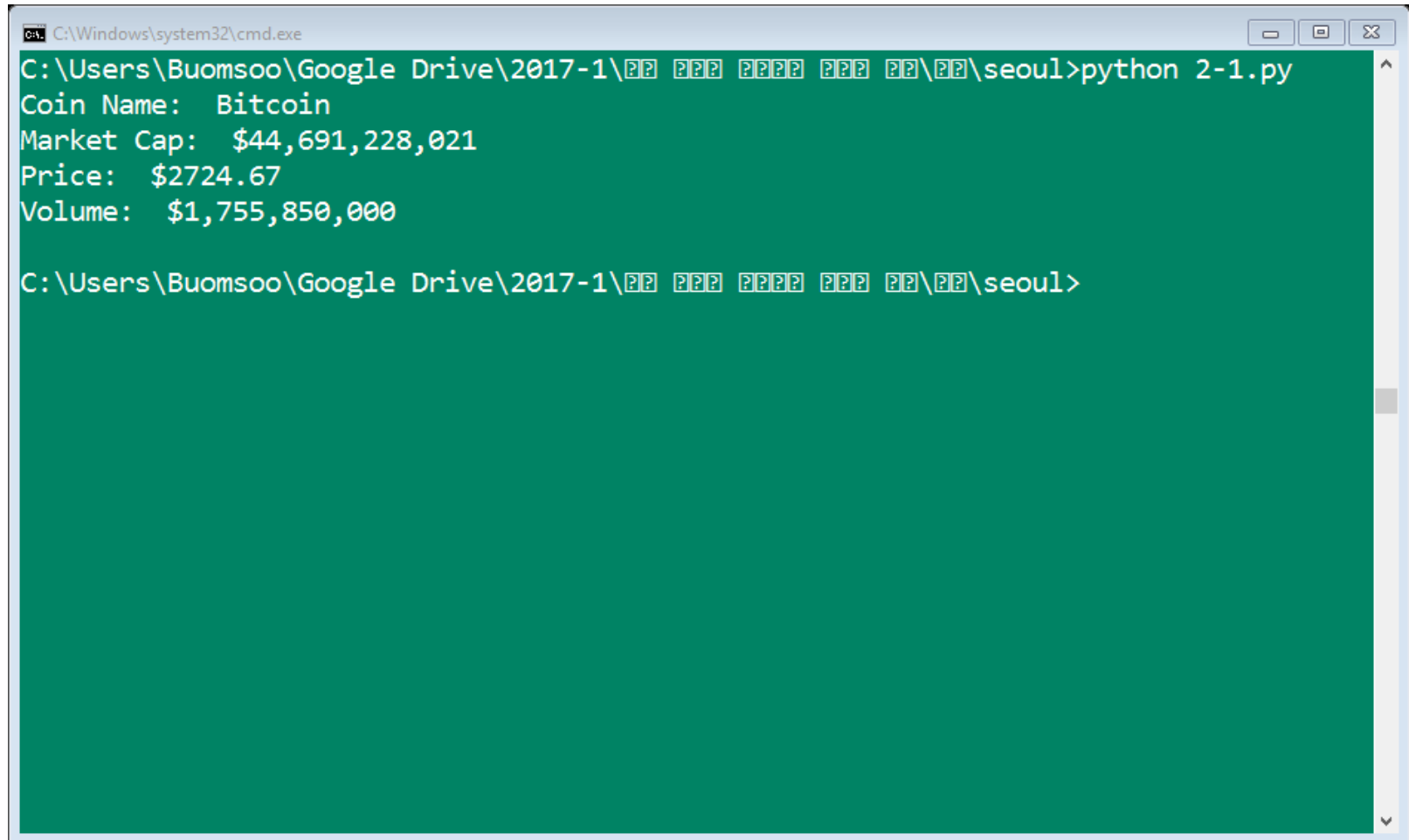
실습 2-1. 비트코인 관련 데이터 출력하기

- TIPS

- 분할 정복(Divide and Conquer)
- 우선 전체 코인 데이터가 포함되어 있는 테이블(table)을 찾는다
- 테이블 내에 비트코인 데이터가 있는 행(tr)을 찾는다

실습 2-1. 비트코인 관련 데이터 출력하기

- 출력 결과



```
C:\Windows\system32\cmd.exe
C:\Users\Buomsoo\Google Drive\2017-1\?? ??? ??? ???? ??? ??\??\seoul>python 2-1.py
Coin Name:  Bitcoin
Market Cap:  $44,691,228,021
Price:  $2724.67
Volume:  $1,755,850,000










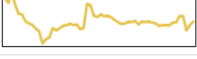


C:\Users\Buomsoo\Google Drive\2017-1\?? ??? ??? ???? ??? ??\??\seoul>
```

The screenshot shows a Windows command prompt window with a green background. The title bar indicates the path C:\Windows\system32\cmd.exe. The prompt is at C:\Users\Buomsoo\Google Drive\2017-1\?? ??? ??? ???? ??? ??\??\seoul>. The user has executed the command python 2-1.py, which has produced the following output: Coin Name: Bitcoin, Market Cap: \$44,691,228,021, Price: \$2724.67, and Volume: \$1,755,850,000. The prompt is now ready for the next command.

실습 2-2. 상위 100개 코인 관련 데이터 저장하기

· 코인마켓캡(<https://coinmarketcap.com/>) 메인 페이지에 있는 상위 100개 코인의 화폐 이름(Name)과 시가총액(Market Cap), 개당 가격(Price), 그리고 24시간 볼륨(Volume)을 텍스트 파일로 저장한다

- 한 줄에 코인 하나의 데이터를 저장하고, 각 행에서는 탭으로 구분한다

CryptoCurrency Market Capitalizations							
<div>24/7 Support Team Dedicated Client Support</div> <div>Unique & Sophisticated User Interface</div>							
Market Cap ▾ Trade Volume ▾ Trending ▾ Tools ▾				Search Currencies		Q	
All ▾	Currencies ▾	Assets ▾	USD ▾	Next 100 →		View All	
▲#	Name	Market Cap	Price	Circulating Supply	Volume (24h)	% Change (24h)	Price Graph (7d)
1	 Bitcoin	\$44,501,255,243	\$2713.09	16,402,425 BTC	\$1,758,920,000	3.59%	
2	 Ethereum	\$32,993,789,011	\$356.07	92,660,703 ETH	\$979,084,000	-4.60%	
3	 Ripple	\$12,329,543,959	\$0.322002	38,290,271,363 XRP *	\$762,189,000	10.12%	
4	 Litecoin	\$2,434,948,352	\$47.15	51,646,757 LTC	\$575,602,000	-6.08%	
5	 Ethereum Classic	\$2,018,384,427	\$21.75	92,799,284 ETC	\$153,872,000	-5.35%	
6	 NEM	\$1,781,676,000	\$0.197964	8,999,999,999 XEM *	\$17,064,100	1.58%	
7	 Dash	\$1,370,452,365	\$185.83	7,374,843 DASH	\$54,122,900	-7.68%	

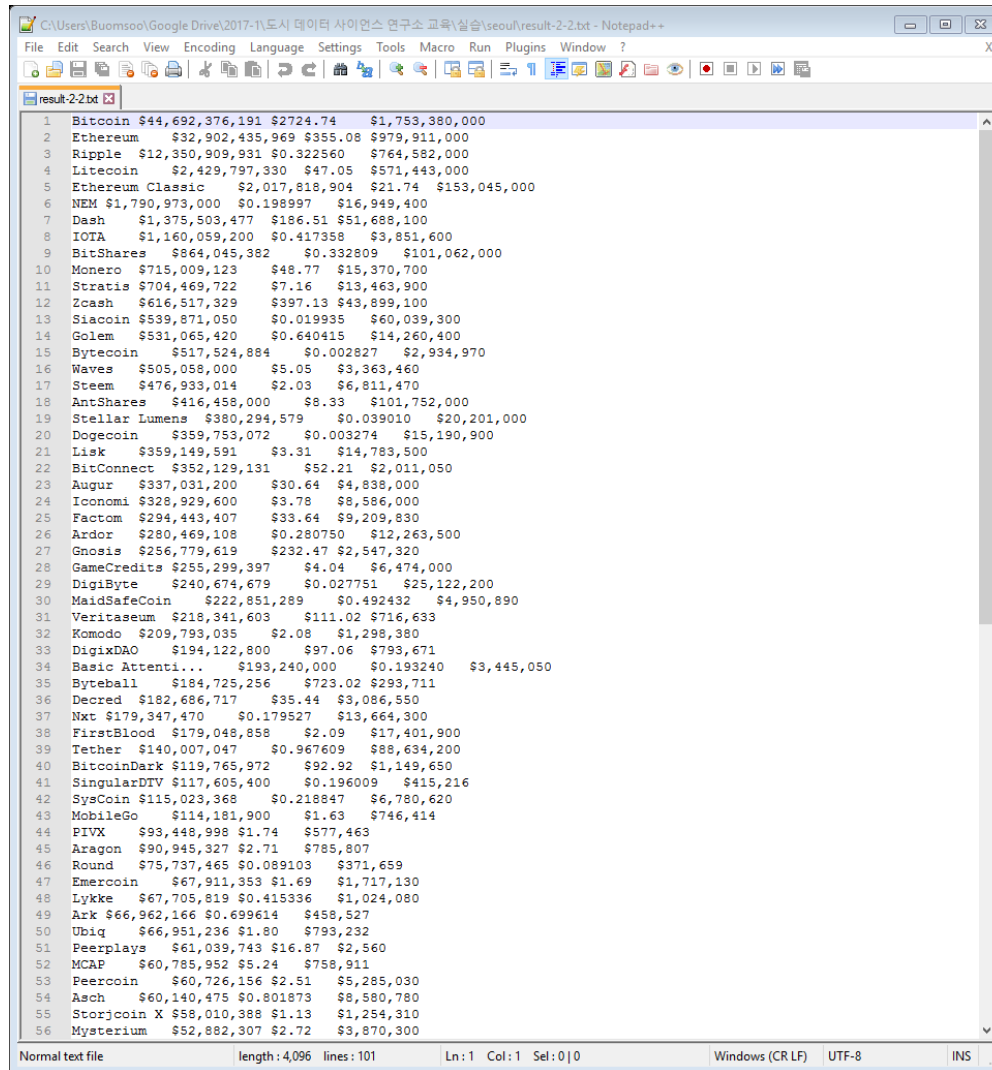
실습 2-2. 상위 100개 코인 관련 데이터 저장하기

· TIPS

- 실습 2-1에서 활용했던 코드를 조금만 바꾸면 쉽게 풀 수 있다
- for문을 활용해 텍스트 파일에 코인 하나씩 데이터를 write한다

실습 2-2. 상위 100개 코인 관련 데이터 저장하기

· 저장 결과




```
1 Bitcoin $44,692,376,191 $2724.74 $1,753,380,000
2 Ethereum $32,902,435,969 $355.08 $979,911,000
3 Ripple $12,350,909,931 $0.322560 $764,582,000
4 Litecoin $2,429,797,330 $47.05 $571,443,000
5 Ethereum Classic $2,017,818,904 $21.74 $153,045,000
6 NEM $1,790,973,000 $0.198997 $16,949,400
7 Dash $1,375,503,477 $186.51 $51,688,100
8 IOTA $1,160,059,200 $0.417358 $3,851,600
9 BitShares $864,045,382 $0.332809 $101,062,000
10 Monero $715,009,123 $48.77 $15,370,700
11 Stratis $704,469,722 $7.16 $13,463,900
12 Zcash $616,517,329 $397.13 $43,899,100
13 Siacoin $539,871,050 $0.019935 $60,039,300
14 Golem $531,065,420 $0.640415 $14,260,400
15 Bytecoin $517,524,884 $0.002827 $2,934,970
16 Waves $505,058,000 $5.05 $3,363,460
17 Steem $476,933,014 $2.03 $6,811,470
18 AntShares $416,458,000 $8.33 $101,752,000
19 Stellar Lumens $380,294,579 $0.039010 $20,201,000
20 Dogecoin $359,753,072 $0.003274 $15,190,900
21 Lisk $359,149,591 $3.31 $14,783,500
22 BitConnect $352,129,131 $52.21 $2,011,050
23 Augur $337,031,200 $30.64 $4,838,000
24 Iconomi $328,929,600 $3.78 $8,586,000
25 Factom $294,443,407 $33.64 $9,209,830
26 Ardor $280,469,108 $0.280750 $12,263,500
27 Gnosis $256,779,619 $232.47 $2,547,320
28 GameCredits $255,299,397 $4.04 $6,474,000
29 DigiByte $240,674,679 $0.027751 $25,122,200
30 MaidSafeCoin $222,851,289 $0.492432 $4,950,890
31 Veritaseum $218,341,603 $111.02 $716,633
32 Komodo $209,793,035 $2.08 $1,298,380
33 DigixDAO $194,122,800 $97.06 $793,671
34 Basic Attention... $193,240,000 $0.193240 $3,445,050
35 Byteball $184,725,256 $723.02 $293,711
36 Decred $182,686,717 $35.44 $3,086,550
37 Nxt $179,347,470 $0.179527 $13,664,300
38 FirstBlood $179,048,858 $2.09 $17,401,900
39 Tether $140,007,047 $0.967609 $88,634,200
40 BitcoinDark $119,765,972 $92.92 $1,149,650
41 SingularDTV $117,605,400 $0.196009 $415,216
42 SysCoin $115,023,368 $0.218847 $6,780,620
43 MobileGo $114,181,900 $1.63 $746,414
44 FIVX $93,448,998 $1.74 $577,463
45 Aragon $90,945,327 $2.71 $785,807
46 Round $75,737,465 $0.089103 $371,659
47 Emercoin $67,911,353 $1.69 $1,717,130
48 Lykke $67,705,819 $0.415336 $1,024,080
49 Ark $66,962,166 $0.699614 $458,527
50 Ubiq $66,951,236 $1.80 $793,232
51 Peerplays $61,039,743 $16.87 $2,560
52 MCAP $60,785,952 $5.24 $758,911
53 Peercoin $60,726,156 $2.51 $5,285,030
54 Asch $60,140,475 $0.801873 $8,580,780
55 Storjcoin X $58,010,388 $1.13 $1,254,310
56 Mysterium $52,882,307 $2.72 $3,870,300
```

실습 2-3. 특정 조건을 만족하는 코인 관련 데이터 저장하기

· 상위 100개의 코인 중 24시간 볼륨(Volume)이 \$1,000,000 이상이고 개당 가격이 \$50 이상인 코인들의 데이터만 저장한다

- 한 줄에 코인 하나의 데이터를 저장하고, 각 행에서는 탭으로 구분한다

CryptoCurrency Market Capitalizations							
<div><div>24/7 Support Team Dedicated Client Support</div><div>Unique & Sophisticated User Interface</div></div>							
Market Cap ▾ Trade Volume ▾ Trending ▾ Tools ▾				Search Currencies		Q	
All ▾	Currencies ▾	Assets ▾	USD ▾	Next 100 →		View All	
▲#	Name	Market Cap	Price	Circulating Supply	Volume (24h)	% Change (24h)	Price Graph (7d)
1	 Bitcoin	\$44,501,255,243	\$2713.09	16,402,425 BTC	\$1,758,920,000	3.59%	
2	 Ethereum	\$32,993,789,011	\$356.07	92,660,703 ETH	\$979,084,000	-4.60%	
3	 Ripple	\$12,329,543,959	\$0.322002	38,290,271,363 XRP *	\$762,189,000	10.12%	
4	 Litecoin	\$2,434,948,352	\$47.15	51,646,757 LTC	\$575,602,000	-6.08%	
5	 Ethereum Classic	\$2,018,384,427	\$21.75	92,799,284 ETC	\$153,872,000	-5.35%	
6	 NEM	\$1,781,676,000	\$0.197964	8,999,999,999 XEM *	\$17,064,100	1.58%	
7	 Dash	\$1,370,452,365	\$185.83	7,374,843 DASH	\$54,122,900	-7.68%	

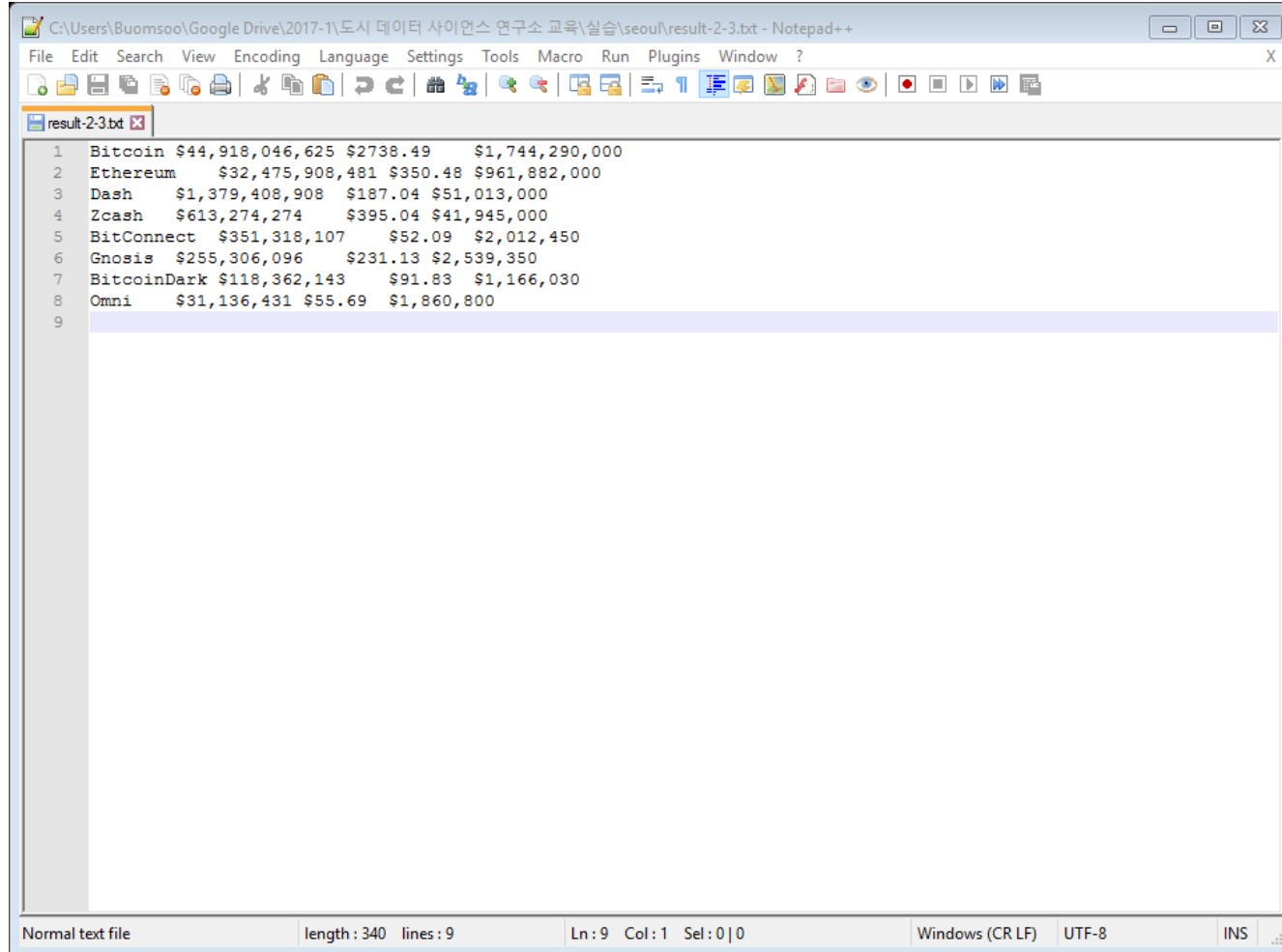
실습 2-3. 특정 조건을 만족하는 코인 관련 데이터 저장하기

· TIPS

- 각 코인의 개당 가격과 볼륨에서 replace 함수를 사용해 '\$', ',' 등 기호들을 제거한다
- 가격과 볼륨을 float 데이터 타입으로 변환한다
- If문을 활용해 조건을 만족하는 코인의 데이터만 저장한다

실습 2-3. 특정 조건을 만족하는 코인 관련 데이터 저장하기


· 저장 결과



```
1 Bitcoin $44,918,046,625 $2738.49 $1,744,290,000
2 Ethereum $32,475,908,481 $350.48 $961,882,000
3 Dash $1,379,408,908 $187.04 $51,013,000
4 Zcash $613,274,274 $395.04 $41,945,000
5 BitConnect $351,318,107 $52.09 $2,012,450
6 Gnosis $255,306,096 $231.13 $2,539,350
7 BitcoinDark $118,362,143 $91.83 $1,166,030
8 Omni $31,136,431 $55.69 $1,860,800
9
```

Normal text file length : 340 lines : 9 Ln: 9 Col: 1 Sel: 0 | 0 Windows (CR LF) UTF-8 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 3-1. 영화 정보 출력하기

- IMDb 사이트에서 영화 펄프 픽션 페이지(http://www.imdb.com/title/tt0110912/?ref=nm_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

1,426,456

☆ Rate This

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

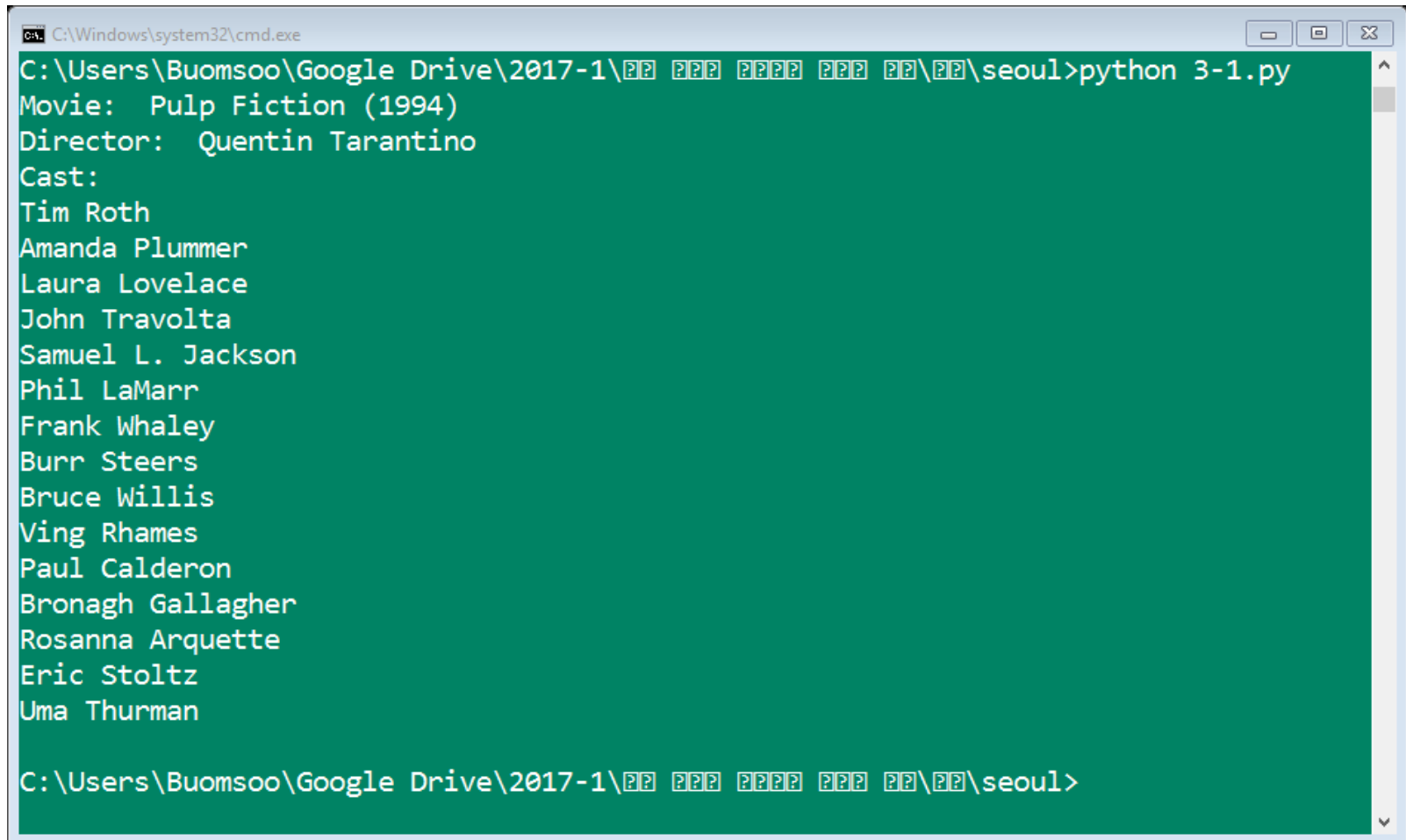
Popularity
154 (▲ 19)

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

Cast			Edit
Cast overview, first billed only:			
	Tim Roth	...	Pumpkin
	Amanda Plummer	...	Honey Bunny
	Laura Lovelace	...	Waitress
	John Travolta	...	Vincent Vega
	Samuel L. Jackson	...	Jules Winnfield
	Phil LaMarr	...	Marvin
	Frank Whaley	...	Brett
	Burr Steers	...	Roger
	Bruce Willis	...	Butch Coolidge
	Ving Rhames	...	Marsellus Wallace
	Paul Calderon	...	Paul
	Bronagh Gallagher	...	Trudi
	Rosanna Arquette	...	Jody
	Eric Stoltz	...	Lance
	Uma Thurman	...	Mia Wallace

실습 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 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 movie poster, a video player showing a scene with John Travolta, and various ratings and reviews.

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

Pulp Fiction (1994) ★ 8.9 ¹⁰ 1,426,456 Rate This

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

PULP FICTION
WINNER-BEST PICTURE-1994 CANNES FILM FESTIVAL
10+
JOHN TRAVOLTA
SAMUEL L. JACKSON
UMA THURMAN
BRUCE CAMPBELL
TIM ROTH
AMANDA PLAMACK
NINA GOLD
ERIC STROPE
ROSARINA ARNETTE
CHRISTOPHER MALKIN
BY BRUCE WILLIS

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 3-2. 여러 영화 정보 저장하기

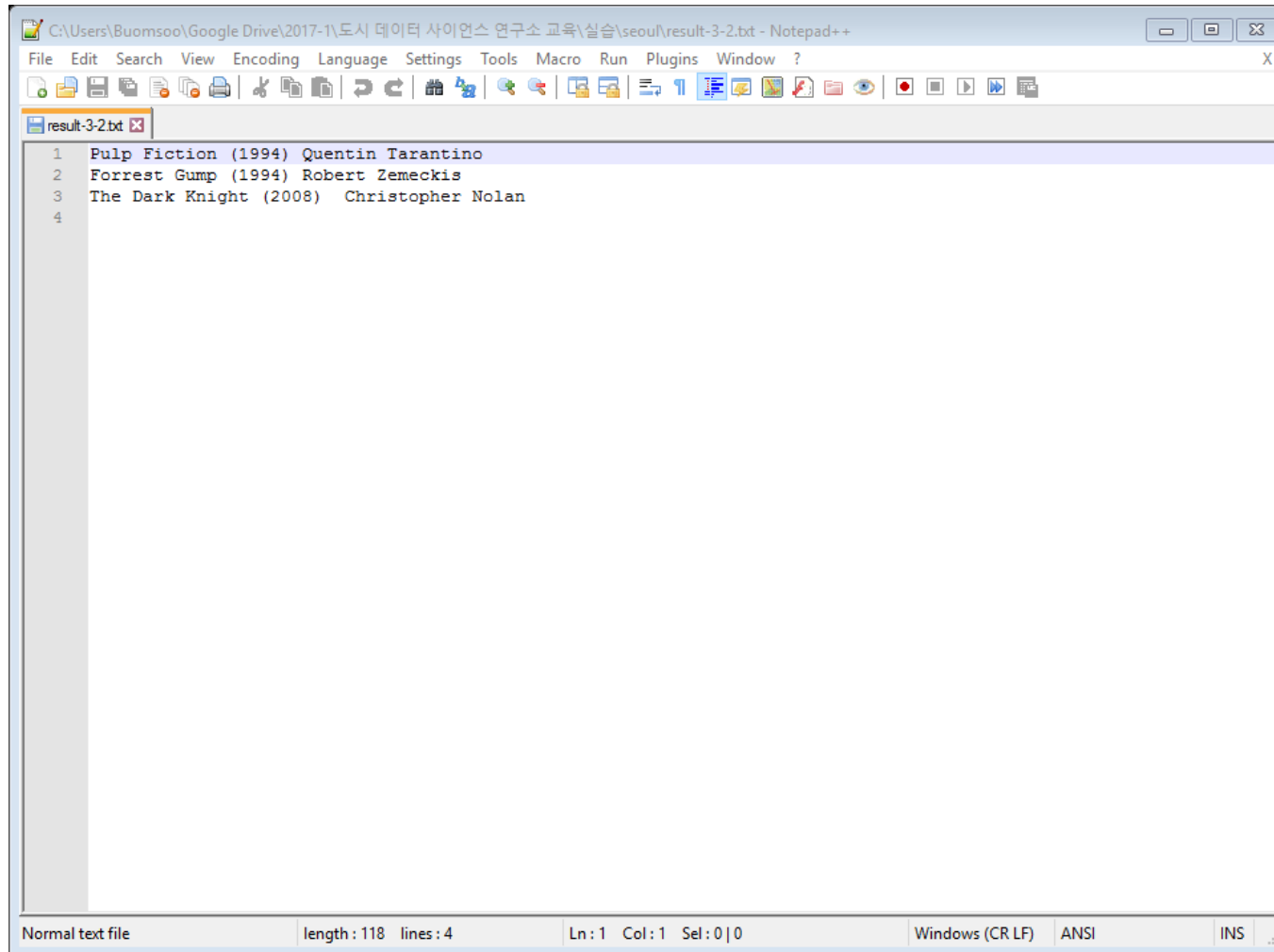
· TIPS

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



실습 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 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 for the movie 'The Dark Knight' (2008). The page is titled 'Reviews & Ratings for The Dark Knight'. On the left sidebar, there are links for 'Own the rights?', 'Buy it at Amazon', 'More at IMDb Pro', 'Add to Watchlist', and 'Update Data'. Below these are 'Quicklinks' for reviews, 'Top Links' for trailers, cast, trivia, and official sites, 'Overview' for main details, 'Awards & Reviews' for user reviews, and 'Plot & Quotes' for plot summary and synopsis. The main content area shows the movie poster, the title 'The Dark Knight', and a 'Write review' button. Below this is a filter dropdown set to 'Best' and a 'Hide Spoilers' checkbox. The page indicates 'Page 1 of 492' and 'Index 4913 reviews in total'. A section titled '3117 out of 4092 people found the following review useful:' features a user profile for 'straightblaster' from the United States, dated '9 July 2008'. The review text is highlighted with a green border and contains the following content:

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.

At the bottom of the review, there is a poll asking 'Was the above review useful to you?' with 'Yes' and 'No' buttons. Below the poll, it states '2196 out of 3207 people found the following review useful:'.

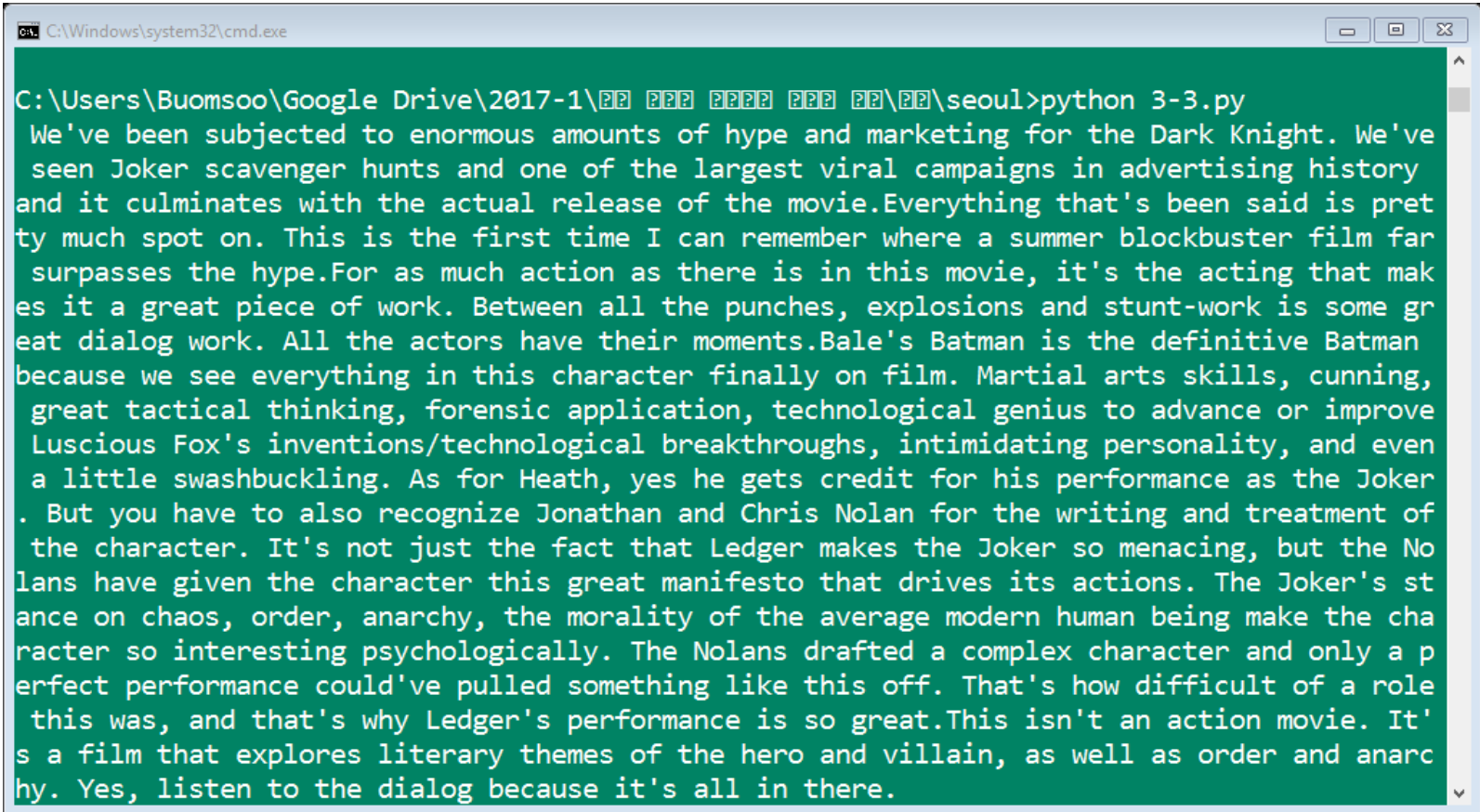
실습 3-3. 영화 리뷰 출력하기

· 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 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 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](#)

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

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](#)

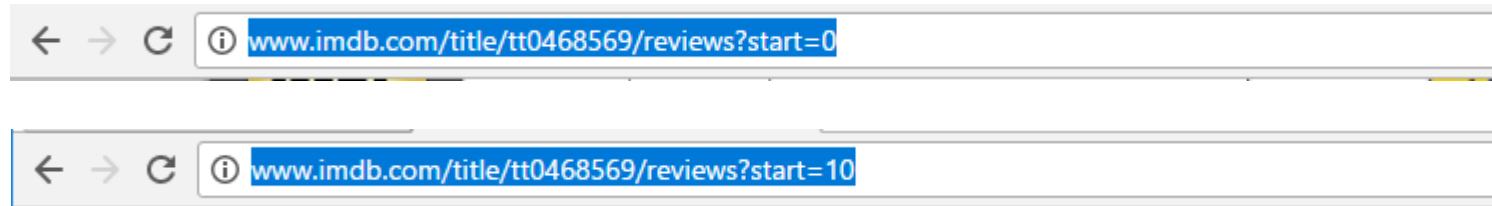
2196 out of 3207 people found the following review useful:

Film Surpasses "Elegance" in every aspect!!!

실습 3-4. 영화 리뷰 저장하기

· TIPS

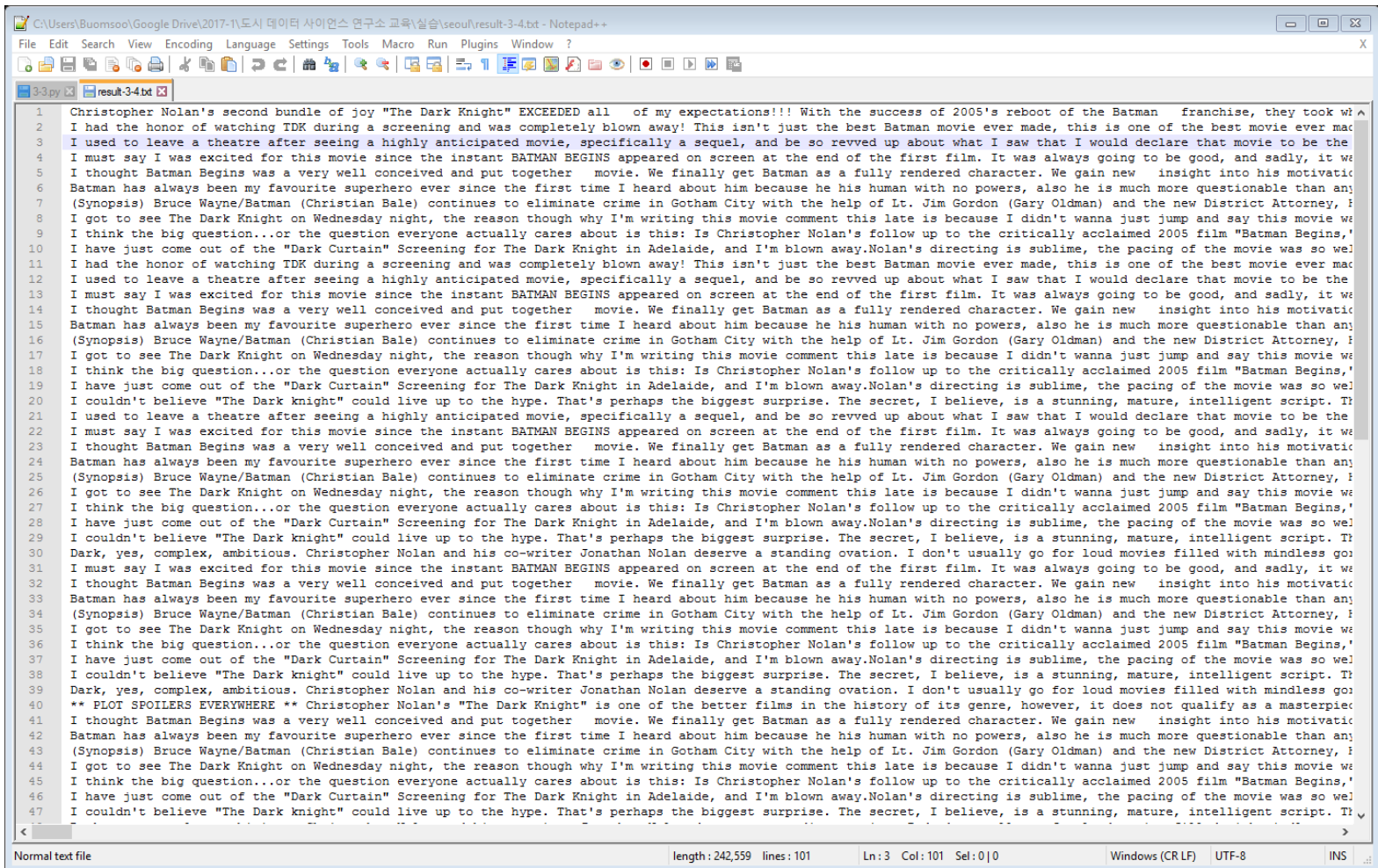
- 영화 리뷰 페이지의 페이지 번호를 바꾸어 가면서 URL 구조를 세심히 살펴보자



- 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 3-4. 영화 리뷰 저장하기

· 저장 결과



The screenshot shows a Notepad++ window with the title bar "C:\Users\Buomsoo\Google Drive\2017-1\토시 데이터 사이언스 연구소 교육\실습\seoul\result-3-4.txt - Notepad++". The menu bar includes File, Edit, Search, View, Encoding, Language, Settings, Tools, Macro, Run, Plugins, and Window. The toolbar contains various icons for file operations and editing. The text area displays a movie review for "The Dark Knight" by Christopher Nolan, with line numbers 1 through 47 on the left margin. The review text is repeated multiple times, including a section titled "** PLOT SPOILERS EVERYWHERE **". The status bar at the bottom shows "Normal text file", "length: 242,559 lines: 101", "Ln: 3 Col: 101 Sel: 0 | 0", "Windows (CR LF)", "UTF-8", and "INS".

```
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 motivatic
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 any
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 motivatic
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 any
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 motivatic
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 any
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 motivatic
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 any
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 motivatic
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 any
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 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](#)

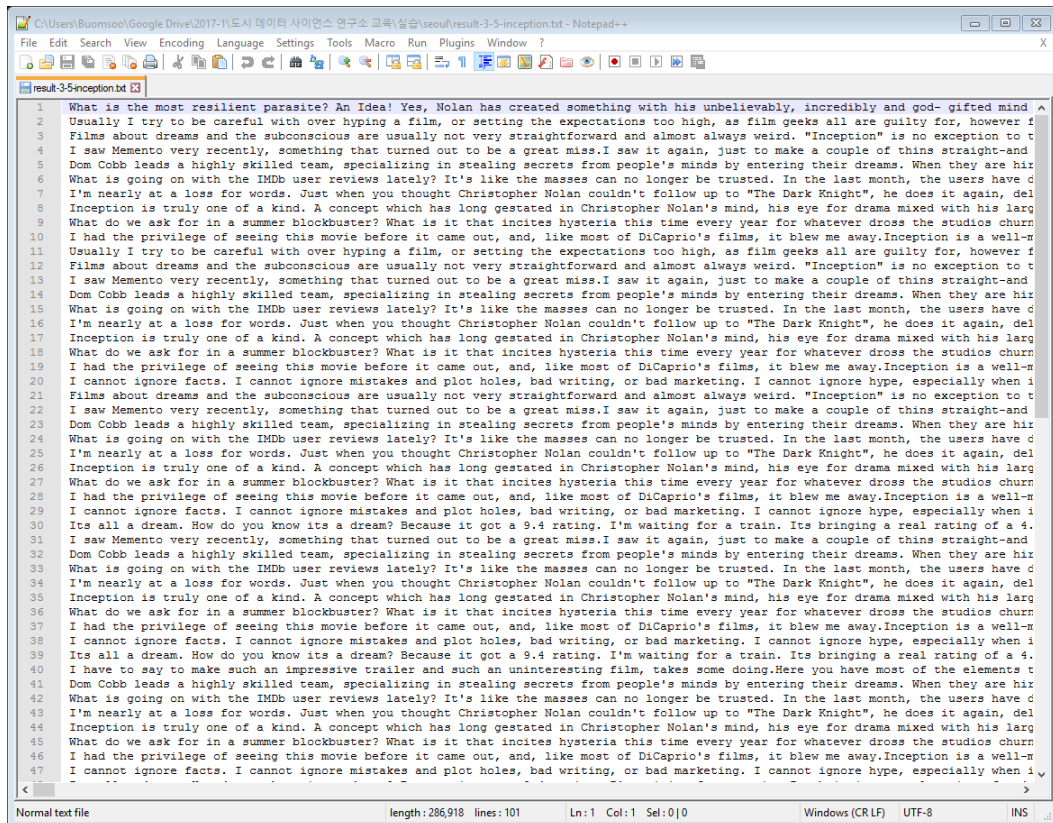
[Own the rights?](#)
[Buy it at Amazon](#)
[More at IMDb Pro](#)
[Add to Watchlist](#)
[Update Data](#)

Quicklinks
reviews ▼




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

실습 3-5. 영화 리뷰 저장하기(2)

· 저장 결과



The screenshot shows a Notepad++ window with the file path "C:\Users\Buomsoo\Google Drive\2017-1\도시 데이터 사이언스 연구소 교육\실습\result-3-5-inception.txt". The text is a movie review for "Inception" by Christopher Nolan, discussing its themes of dreams and reality, and comparing it to other films like "Memento" and "The Dark Knight". The review is repeated twice. The status bar at the bottom indicates "Normal text file", "length: 286,918 lines: 101", "Ln: 1 Col: 1 Sel: 0 | 0", "Windows (CR LF)", "UTF-8", and "INS".

 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