

Big Data Analytics for Animal hospital Recommendation System

Computer Science 14109343 Yura Baek
Computer Science 14109383 K. H. Choi

Contents

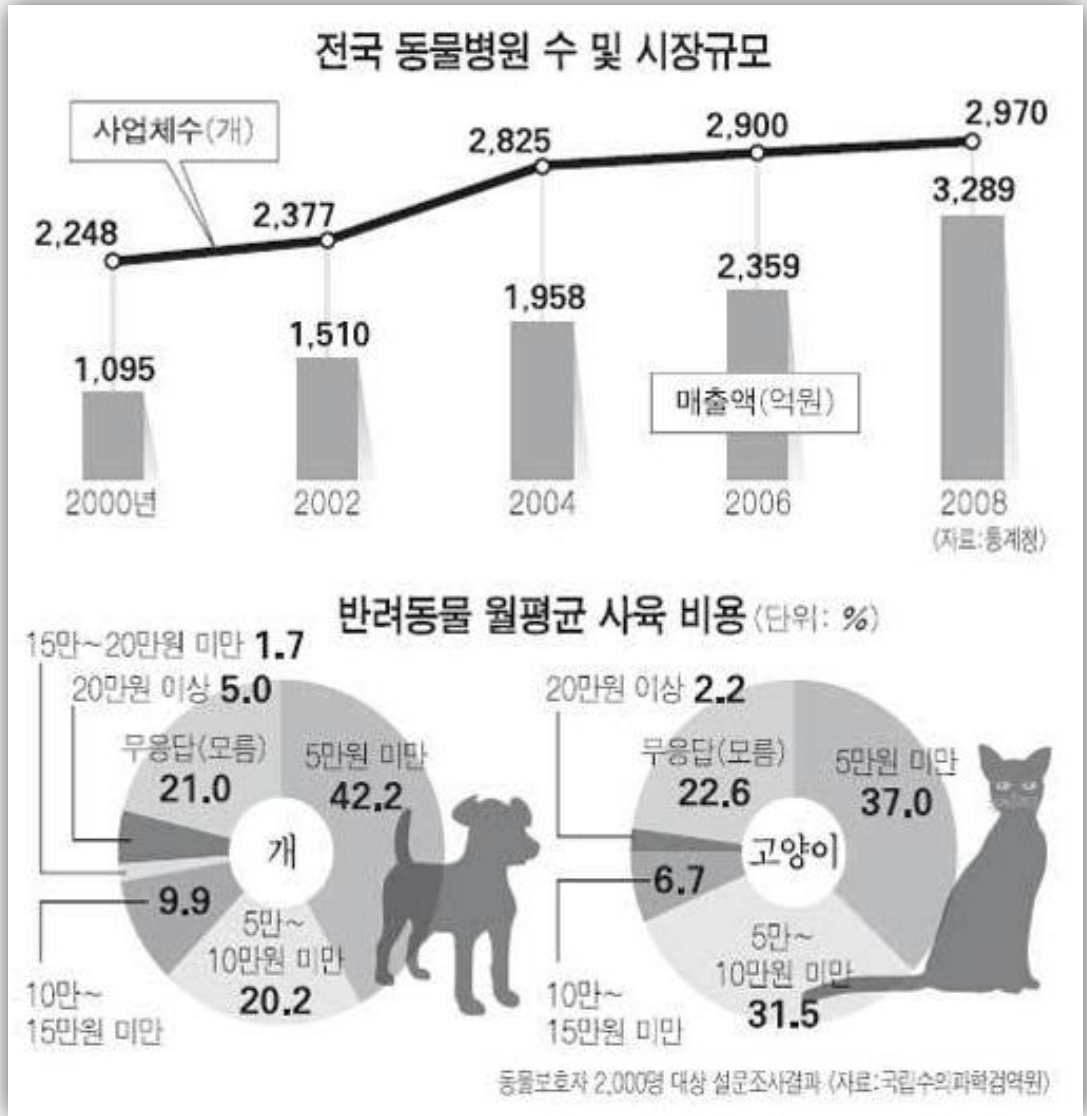
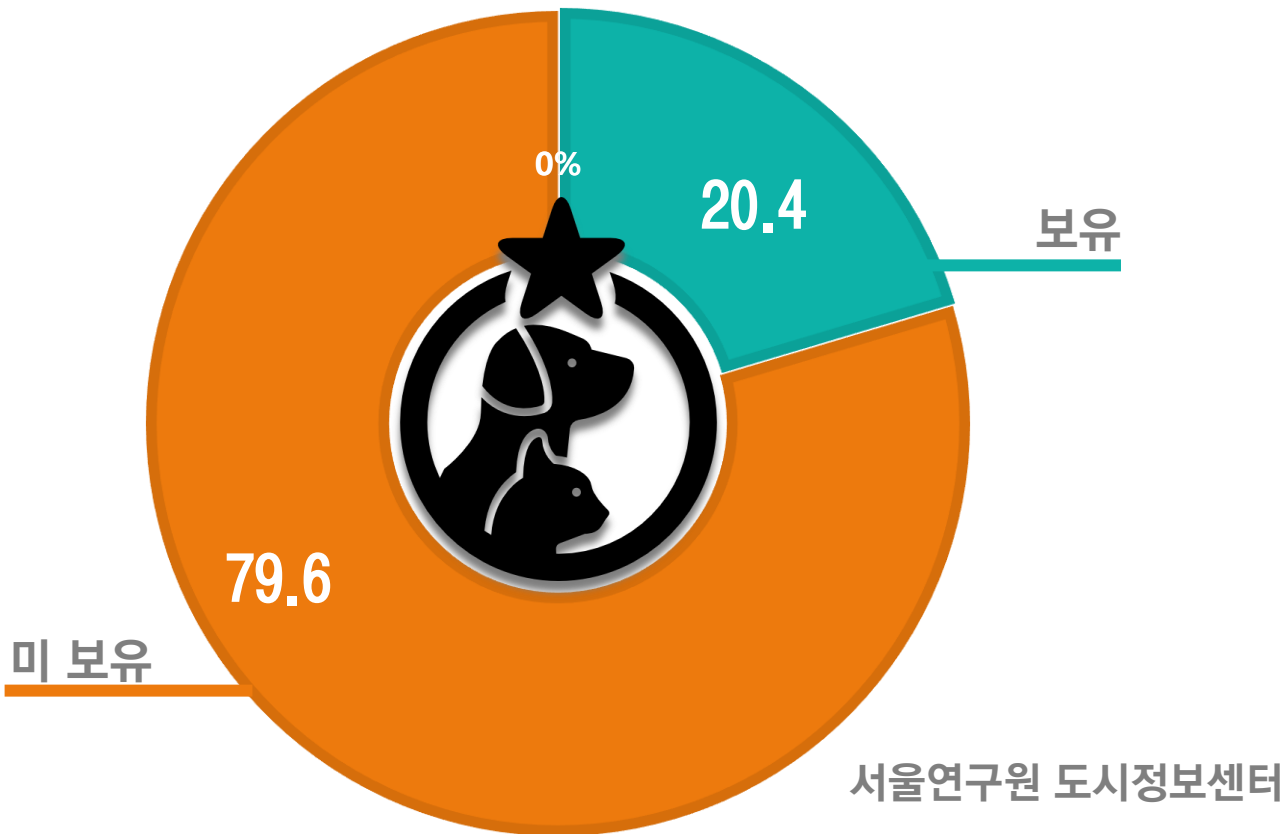
- Purpose of Project
- Project overview
- Goal & Requirements
- Development environment
- Modules
- Software architecture & Description
- Evaluation
- Further plans

PART1.

Purpose of Project

반려동물의 수 증가

서울시 반려동물 보유 여부 (단위: %)



Problem . 과잉진료비를 청구하는 동물병원

5배

NAVER 동물병원추천

통합검색 블로그 카페 지식iN 이미지 동영상 어학사전 뉴스 더보기 검색옵션

연관검색어 모임방 수원 동물병원 마크리스 동물병원 **믿을만한 과잉진료 동물병원** 신고

파워링크 '동물병원추천' 관련 광고입니다.

래미동물의료센터 www.rayvmc.co.kr
강남구 대치동, 24시간 야간응급실운영, 서울대 동물병원 출신 전공자의 분과별진료.

24시 N동물의료센터 www.namc.co.kr
노원본점 길음점, 24시응급진료, CT보유, 분과별2차진료, 특수동물, 고양이전문

로알동물메디컬센터 RoyalAmc.com
강마지수술전문, 각 분야 특화 전문수의사, 최신 의료장비, 365일 연중무휴동물병원

24시 지구촌 동물종합병원 map.naver.com/local/siteview.nhn?code=1...
구로구위치, 30명스텝 과별전문진료, 24시간응급진료, 슬개골탈구전문, 심장전문, 내시경

착한종합동물병원 홍대 IM www.impet.co.kr/
홍대입구역, 대학원UCdavis수료전문, 통합의료센터서비스, 최신시설, 착한가격

24시간진료 동물메디컬센터W www.animalw.co.kr
정확한진단, 검사시간단축, 최신형 CT도입, 반려동물 내, 외과 전문진료!

24시응급진료 베스트동물병원 jsbestamc.allthemedi.com
자신 도무변의 365일 24시간응급진료, 1인1과 전문의, 수술전문

동물병원 항목별 가격비교

항 목	최저가 및 병원 위치	최고가 및 병원 위치
초진료	3000원(서울 마포구)	1만5000원(경기도 안산, 서울 강북구)
전해질 검사	1만원(서울 관악구, 강동구, 경기도 수원시)	3만원(서울 강남구)
전혈구 검사	2만원(서울 관악구, 강동구)	3만원(서울 강남구, 대구 수성구)
혈압 측정	1만1000원(서울 구로구)	2만원(서울 강남구)
복부 초음파	4만원(울산 남구)	10만원(서울 강남구)
분변검사	5000원(서울 영등포구, 경기도 안산)	2만원(서울 관악구)
1일 입원비	2만2000원(서울 관악구)	4만3000원(서울 영등포구)
암컷 중성화수술	15만원(서울 관악구)	33만원(경기도 성남)
호흡마취	6만원(서울 양천구, 부산 남구)	11만원(서울 관악구, 경기도 안산)

(국민일보 기사에 실린 동물병원 항목별 가격비교)

9 한미정상

10 文대통령 유엔연설

2017.09.22 08:30 ~ 11:30 기준

청소년 인기검색어

네이버 웹

한국소비자원_반려동물 관련 소비실태 및 개선방안

Problem . 과잉진료비를 청구하는 동물병원

Home > 이코노미스트 > Cover Story

1422호 (2018.02.19) [44]

목차보기

기사 제보

편집장에게 한마디

[커지는 반려동물 의료비 부담] 수의사는 못 믿겠고 보험은 있으나 마나

함승민 기자 sham@joongang.co.kr

반려인 울리는 '깜깜이 진료비'...정부 '가격 공시제' 등 대안 검토

최근 연구에 따르면 단순히 의료비용만을 따졌을 때 국내 반려동물 진료비는 비싼 편이 아니다. 농림축산식품부가 발주한 '반려동물 산업 활성화를 위한 소비자 진료비 부담 완화 방안 연구 보고서'는 소득수준을 고려해도 미국·독일의 진료비에 비해 한국의 동물병원 비용이 높다고 할 수 없다고 지적했다. 다만 '동물병원 서비스의 질이 지불하는 비용만큼의 수준에 미치지 못해 고가라고 인식할 가능성이 있다'고 설명했다. 또 건강보험을 통해 본인부담금이 적게 드는 사람에 비해 동물은 의료비용 전부를 내야 하기 때문에 체감되는 부담이 더 크다는 지적도 있다.

문제를 더 키우는 건 '깜깜이' 진료비다. 전문적인 동물의료에 대한 정보를 일반 소비자가 알기가 어렵고, 적정 기준이 없어 과잉 진료로 인한 부담이 커진다는 지적이다. 사람과 달리 진료비에 대한 기준이 마련되지 않아 병원마다 격차도 크다. 사단법인 소비자교육중앙회가 지난해 서울 및 6개 광역시에서 동물병원 25곳의 진료비 및 예방접종비 등을 조사한 결과에 따르면 병원비는 최대 6배, 예방접종비는 최대 8배까지 차이가 나는 것으로 나타났다. 반려인들이 '호갱님'으로 전락하기 쉬운 구조다.

부담을 줄일 수 있는 반려동물 전용보험, 일명 '펫보험'도 국내에서는 활성화하지 못했다. 보험 업계 조사에 따르면 한국의 펫보험 가입률은 0.1% 미만으로 추정된다. 영국(약 20%)·독일(15%)·미국(10%)·일본(2~3%)에 비해 낮다. 소비자들이 가입이 까다롭고 보장항목은 제한적인 펫보험을 외면하고 있기 때문이다. 현재 시중에는 삼성화재의 패밀리아리스 애견의료보험2와 현대해상화재보험의 하이펫 애견보험, 롯데손해보험의 롯데 마이펫보험 등 3종의 펫보험이 나와 있다. 모두 노령 반려동물은 가입이 되지 않고, 이마저도 순수 보장형으로 1년마다 갱신해야 한다. 예방접종·중성화·깃병 등 반려동물에게 자주 발생하는 질병은 보장항목에서 빠져 있는 경우가 많다. '펫보험을 드느니 적금을 든다'는 조소가 나오는 이유다.

Problem . 반려동물 소비자피해구제 현황

<표 3-2> 반려동물 관련 소비자피해구제 신청 현황

단위 : 건, (%)

구분	2010년	2011년	2012년	2013.6	계
애완견	100(76.9)	87(71.3)	131(77.1)	50(62.5)	368(73.3)
애완고양이	10(7.7)	18(14.8)	23(13.5)	8(10.0)	59(11.8)
동물병원	11(8.5)	8(6.6)	8(4.7)	6(7.5)	33(6.6)
애완동물용품	5(3.8)	6(4.9)	5(2.9)	13(16.3)	30(6.0)
애완동물사료	1(0.8)	1(0.8)	3(1.8)	3(3.8)	8(1.6)
기타	3(2.3)	2(1.6)	-	-	5(1.0)
계	130(100.0)	122(100.0)	170(100.0)	80(100.0)	402(100.0)
증가지수	100	93.8	130.8	123.1	

* 기타 : 2010년 3건 - 미용서비스, 2011년 2건 - 애견(훈련)학교

<표 4-34> 동물병원이나 미용서비스 이용 중 소비자불만·피해 이유(중복응답)
단위 : 명, (%)

구분	빈도	%
미용서비스 중 다치거나 후유증	126	(24.5)
다른 병원보다 진료비가 비쌌	88	(17.1)
과잉 검사 및 진료	79	(15.4)
진료내용·비용에 대한 사전설명 부족	73	(14.2)
수의사의 능력부족	65	(12.6)
미용서비스 요금 과다	58	(11.3)
동물치료 중 의료사고	15	(2.9)
기타*	10	(1.9)
계	514	(100.0)

* 기타는 검진기록 안줌, 진료비 기준 없음, 수의사의 성의부족, 수의사의 태도 불량, 질병 아는데 질병이라고 함, 각 1명 등

동물병원 추천에 대한 니즈

NAVER 동물병원추천

통합검색 블로그 카페 지식iN 이미지 동영상 어학사전 뉴스 더보기 검색옵션

연관검색어 ? 모임방 수원 동물병원 아크리스 동물병원 **믿을만한 과잉진료 동물병원** 신고 X

파워링크 '동물병원추천' 관련 광고입니다. [등록 안내 >](#)

래이동물의료센터 www.rayvmc.co.kr
강남구 대치동, 24시간 야간응급실운영, 서울대 동물병원 출신 전공자의 분과별 진료.

24시 N동물의료센터 www.namc.co.kr
노원본점 길음점, 24시응급진료, CT보유, 분과별 2차진료, 특수동물, 고양이전문

로알동물메디컬센터 RoyalAmc.com
강마지수술전문, 각 분야 특화 전문수의사, 최신 의료장비, 365일 연중무휴동물병원

24시 지구촌 동물종합병원 map.naver.com/local/siteview.nhn?code=1...
구로구위치, 30명스텝 과별전문진료, 24시간응급진료 슬개골탈구전문, 심장전문, 내시경

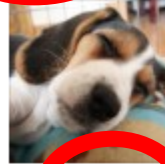
착한종합동물병원 흥대 IM www.impet.co.kr/
흥대입구역, 대학원UCDavis수료전문, 통합의료센터서비스, 최신시설, 착한가격

24시간진료 동물메디컬센터W www.animalw.co.kr
정확한진단, 검사시간단축, 최신형 CT도입, 반려동물 내, 외과 전문진료!

24시응급진료 베스트동물병원 jsbestamc.allthemedi.com
자신 도우며의 365일 24시간응급진료, 내과, 외과, 전문의, 수주 전문

블로그

카페



강북구
저도
(들) 평
매력

도봉구, 노원구 양심동물병원 추천부탁드려요!! 2017.09.03,
슈나 7년째 키우고 있구요~ 다니던 동물 병원이 있었는데 요즘들어 과잉진료가 너무 심한거같아 바꿔보려구
요ㅠㅠ 아가가 기운이 없는거 같아 데려갔었는데 특별한 병명은 말 안해주시고 계속 주사맞고...
대표 감사모-강마지분양, 유기견 애견분양, 개 애완... cafe.naver.com/dogpalza/1... +카페가입

[지식iN에 물어보기](#)

지식iN

Q 노원구 강마지 다리 수술 전문 동물병원 !!! 2005.07.23,

노원구에 살고있어요, 저희집강마지가 다리수술을해야할것같은데요, 다리수술전문으로하는 좋은병원 좀 추천해
주셨으면합니다. 중계동에있는 롯데마트 동물병원에서 미용을시켰...

A 한빛 동물병원 추천 합니다. 마들역과 상계역 중간지점에 있습니다. 오랜 경험도 있으시고 제가 알기로
아시는 동물병원 사이에서도 스승으로 통하시는 분입니다. Patellar...
개, 강마지 | 2

Q 노원구에 좋은 동물병원 추천해주세요~! 2004.07.23,

제가 노원구에 사는데요.. 강마지 예방접종을하는데 비싸더라고요.. 중계동 롯데마트에서 2차접종을 맞았는데
요.. 종합5종2만원 장염1만원 기관지1만원에 했고.....

A 롯데마트에도 동물병원이 있었나요? 진짜 너무하네요..... 있는 동물병원 다 돌아다녀봐도 다 8000원 받
아요;;;... 1층에 <고려동물병원>이라고 있어요 전 상계동이라...
노원구 | 1

Q 근처에 수의사님 친절하고 좋은 동물병원 있으면 추천해 주세요~ 2004.07.14,

좀 있으면 그쪽으로 미사를 가는데 창동 근처에 좋은 동물병원 추천해주세요... 돈 벌기에 급급해서 무뚝뚝하
고 불친절한 수의사는 너무 싫어요..

A 동물병원위드.. 수유점... 친절하고.. 꼼꼼하게 잘보고.. 다 좋은데.. 가격이 비싸죠.. ^^;; 우리아가는 한빛 동물
병원이라고.. 노원역에서 좀 내려가면있어요~ 한빛 추천이요~^^

미용, 건강관리 | 1

[지식iN 더보기 >](#)

PART2.

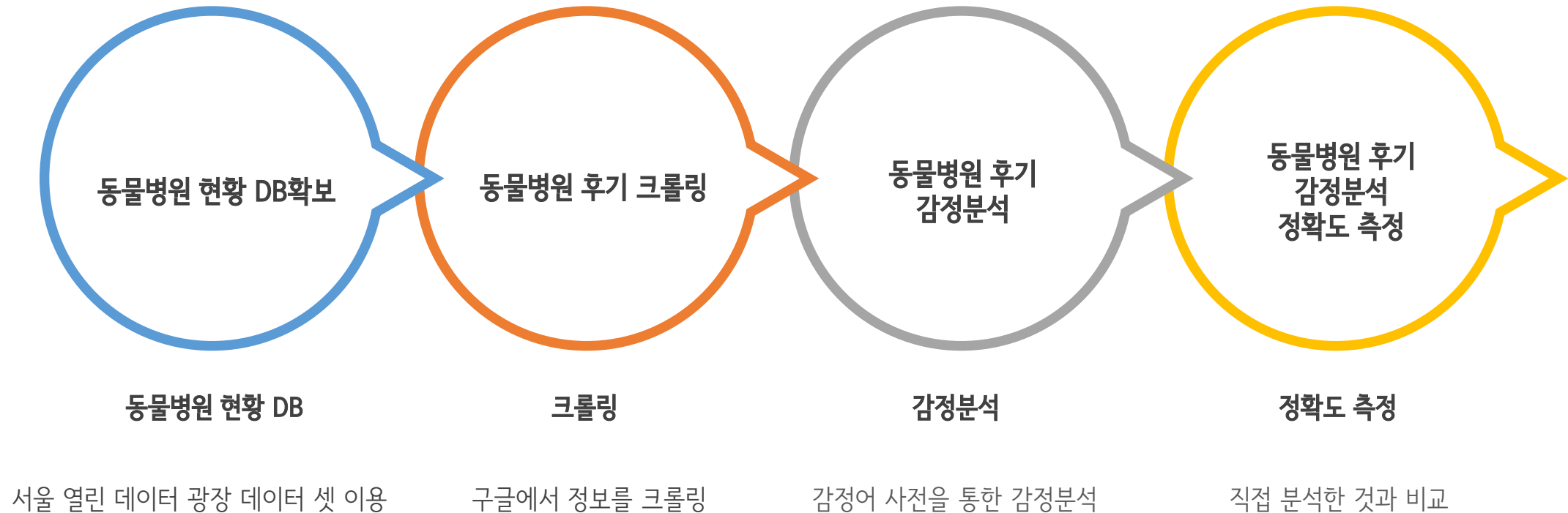
Project overview

동물병원 후기를 **크롤링** 하여,
감정분석을 토대로 동물병원 평가

PART3.

Goal & Requirements

Goal & Requirements



PART4.

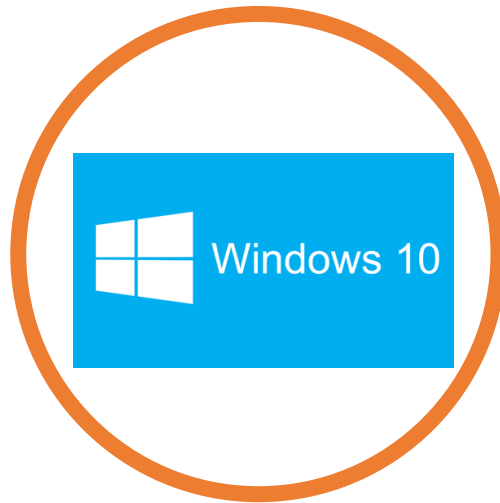
Development Environment

Development Environment



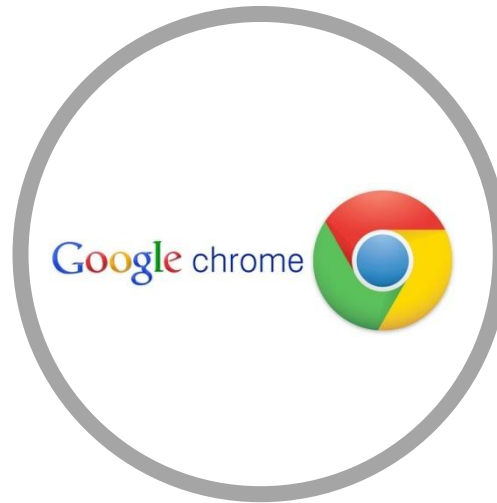
Programming Language

Python 3.7.0



OS

Windows 10 Home



Web Browser

Chrome 70.0.3538.77



IDE

PyCharm 2017.2.3

PART5.

Modules

List of Key Modules

- BeautifulSoup
- KoNLPy
- Os
- Selenium
- Sub process
- Sys

Beautiful Soup

- HTML과 XML 파싱에 사용되는 파이썬 라이브러리
- 추출된 HTML 페이지로부터 웹 스크래핑에 이용할 수 있게 파스트리를 생성

- Documentation:

<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

KoNLPy

- 한국어 정보처리를 위한 파이썬 패키지
- 기초적인 NLP(Natural Language Processing) 작업 수행을 지원
- NLP란?
텍스트에서 의미 있는 정보를 분석, 추출하고 이해하는 일련의 기술집합
- 이 프로젝트에서는 한국어 형태소 분석을 위해 사용
- Documentation:
konlpy.org/en/latest/

Matplotlib

- 자료를 차트나 플롯으로 시각화 하는 것을 지원하는 모듈
- 정형화된 차트나 플롯 이외에도 다양한 시각화 기능을 제공함

- Documentation:
<https://matplotlib.org/>

Selenium

- 일반적으로 웹앱을 테스트할 때 주로 사용하는 프레임 워크
- Selenium의 webdriver는 웹 브라우저를 제어할 수 있는 API를 제공

- Documentation:

<https://seleniumhq.github.io/selenium/docs/api/py/api.html>

Subprocess

- Sub process 모듈은 파이썬 프로그램 내에서 새로운 프로세스를 스폰하고 여기에 입출력 파이프를 연결하며 리턴코드를 획득할 수 있도록 하는 모듈로, 다른 언어로 만들어진 프로그램을 통합, 제어할 수 있게 함
 - 기존의 오래된 몇몇 모듈과 `os.system`, `os.spawn`과 같은 함수들을 대체하기 위해 만들어짐
-
- Documentation:
<https://docs.python.org/3/library/subprocess.html>

sys

- 인터프리터가 사용하거나 유지 관리 하는 일부 변수와 인터프리터가 강력하게 상호 작용하는 함수에 대한 액세스를 제공

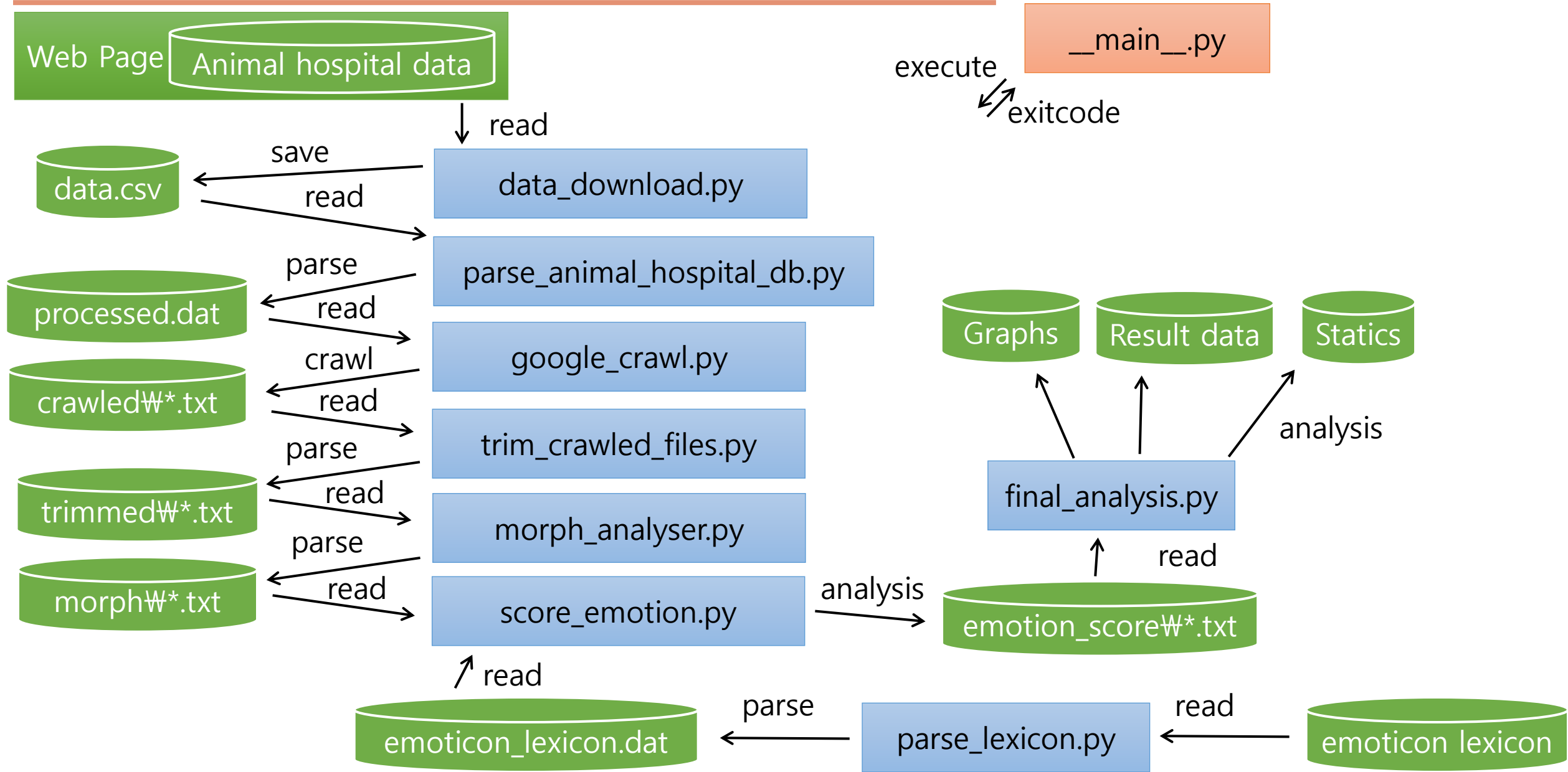
- Documentation:

<https://docs.python.org/3/library/sys.html>

PART6.

Software architecture & Description

Software Architecture



__main__.py

```
import subprocess
import sys

if __name__ == "__main__":
    subprocess.check_call([sys.executable, 'data_download.py'])
    subprocess.check_call([sys.executable, 'parse_animal_hospital_db.py'])
    subprocess.check_call([sys.executable, 'google_crawl.py'])
    subprocess.check_call([sys.executable, 'trim_crawled_files.py'])
    subprocess.check_call([sys.executable, 'morph_analyser.py'])
    subprocess.check_call([sys.executable, 'parse_lexicon.py'])
    subprocess.check_call([sys.executable, 'score_emotion.py'])
    subprocess.check_call([sys.executable, 'final_analysis.py'])
```

그림 - __main.py의 소스코드

- 프로젝트의 모든 서브 프로세스를 정해진 순서대로 작동시키고, 서브 프로세스의 exit code를 체크한다.
- 서브 프로세스 하나라도 exception을 throw할 경우 전체 프로세스가 중지된다.
- 프로젝트를 단 하나의 실행파일로 제어할 수 있게 한다.

Animal Hospital Information Data Extraction

필터선택							XLS	CSV	JSON
시군구코드	신고일자	업소명	영업상태코드	영업상태	사업장소재지(지번)	휴업시작일자	휴업종료일자	폐업일자	
3100000	19891117	공릉동물병원	0002	폐업	서울특별시 노원구 공릉동 59...			20070131	
3100000	19910820	노원동물병원	0002	폐업	서울특별시 노원구 상계동 16...			20101120	
3100000	19920707	한빛종합동물병원	0002	폐업	서울특별시 노원구 상계동 74...			20080501	
3100000	19941010	롯데동물병원	0000	정상	서울특별시 노원구 중계동 50...				
3100000	19950705	LK동물병원	0002	폐업	서울특별시 노원구 중계동 35...			20060121	
3100000	19951124	준동물병원	0002	폐업	서울특별시 노원구 상계동 14...			20080821	
3100000	19960607	리치동물병원	0002	폐업	서울특별시 노원구 공릉동 39...			20150701	
3100000	19980711	25시동물병원	0000	정상	서울특별시 노원구 중계동 36...				
3100000	19990521	은빛동물병원	0000	정상	서울특별시 노원구 상계동 11...				
3100000	19990820	은마로동물병원	0000	정상	서울특별시 노원구 중계동 17...				
3100000	19991007	그린동물병원	0000	정상	서울특별시 노원구 상계동 78...				
3100000	20000407	고려동물병원	0000	정상	서울특별시 노원구 상계동 79...				

< 1 / 1 page(s) > 65 count(s)

data_download.py

- 웹 페이지로부터 데이터 다운로드 받는 스크립트
- Selenium을 이용한 브라우저 자동화 제어과정
 1. 서울 열린 데이터 광장>서울시 노원구 동물병원 현황 접속
 2. Download 버튼 클릭 명령
 3. 데이터가 정상적으로 다운로드 되었는지 확인

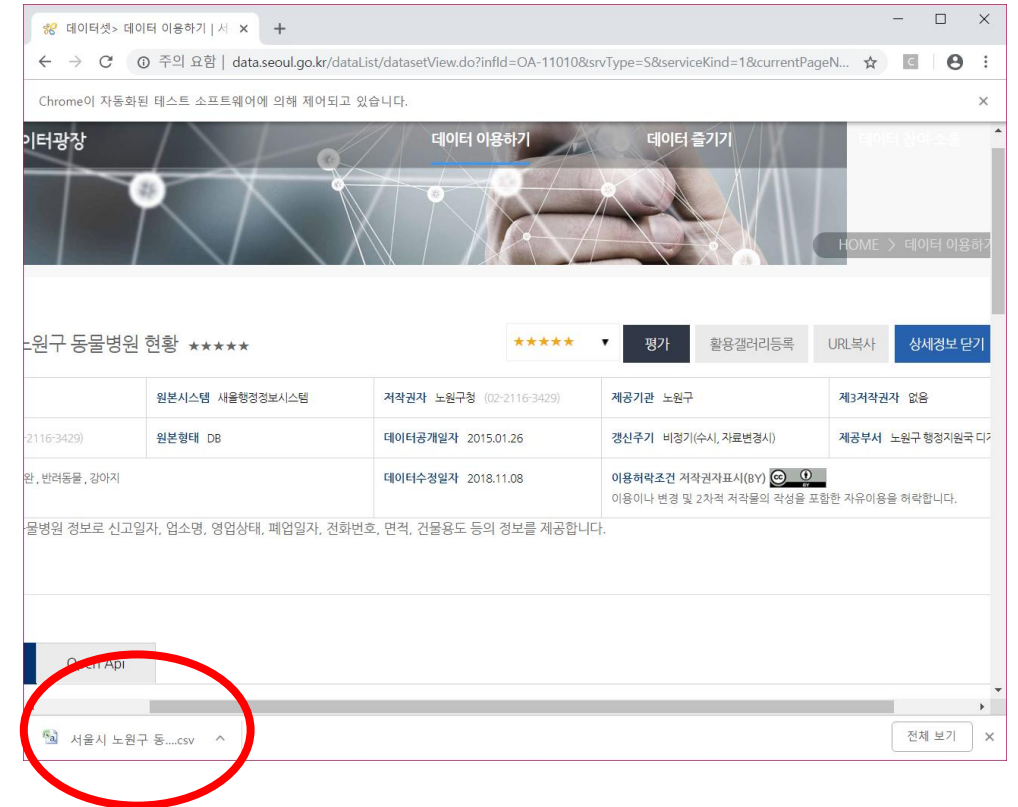


그림 - 프로그램 실행하여 .csv 파일을 자동으로 다운받는 모습

parse_animal_hospital_db.py

그림 - .csv 파일을 가공한 모습

```
서울시 노원구 동물병원 현황.csv - 메모장
파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)
"시군구코드","신고일자","업소명","영업상태코드","영업상태","사업장소재지(지번)","휴업시작일"
","","0","1","","","","3100000","19960607","리치동물병원","0002","폐업","서울특별시 노원구
00000","20010523","미래동물병원","0000","정상","서울특별시 노원구 공릉동 94번지 6호 13통 7
별시 노원구 상계동 651번지 상계주공11단지종합상가112호","","","20040212","","","","0.00","
41번지 152호","","","","971-7511","33.00","","0","1","","","3100000","2004091
","","","952-6426","23.10","","0","1","","","3100000","20060201","브라운동물병원"
""3100000","20080422","(주)한빛종합동물병원","0000","정상","서울특별시 노원구 상계동 749
","","","3100000","20120319","노원24시N동물의료센터","0002","폐업","서울특별시 노원
컬센터","0000","정상","서울특별시 노원구 공릉동 576번지 23호","","","","0.00","","
","","02-949-9075","","","","3100000","20180509","닥터스종합동물병원","0000
```



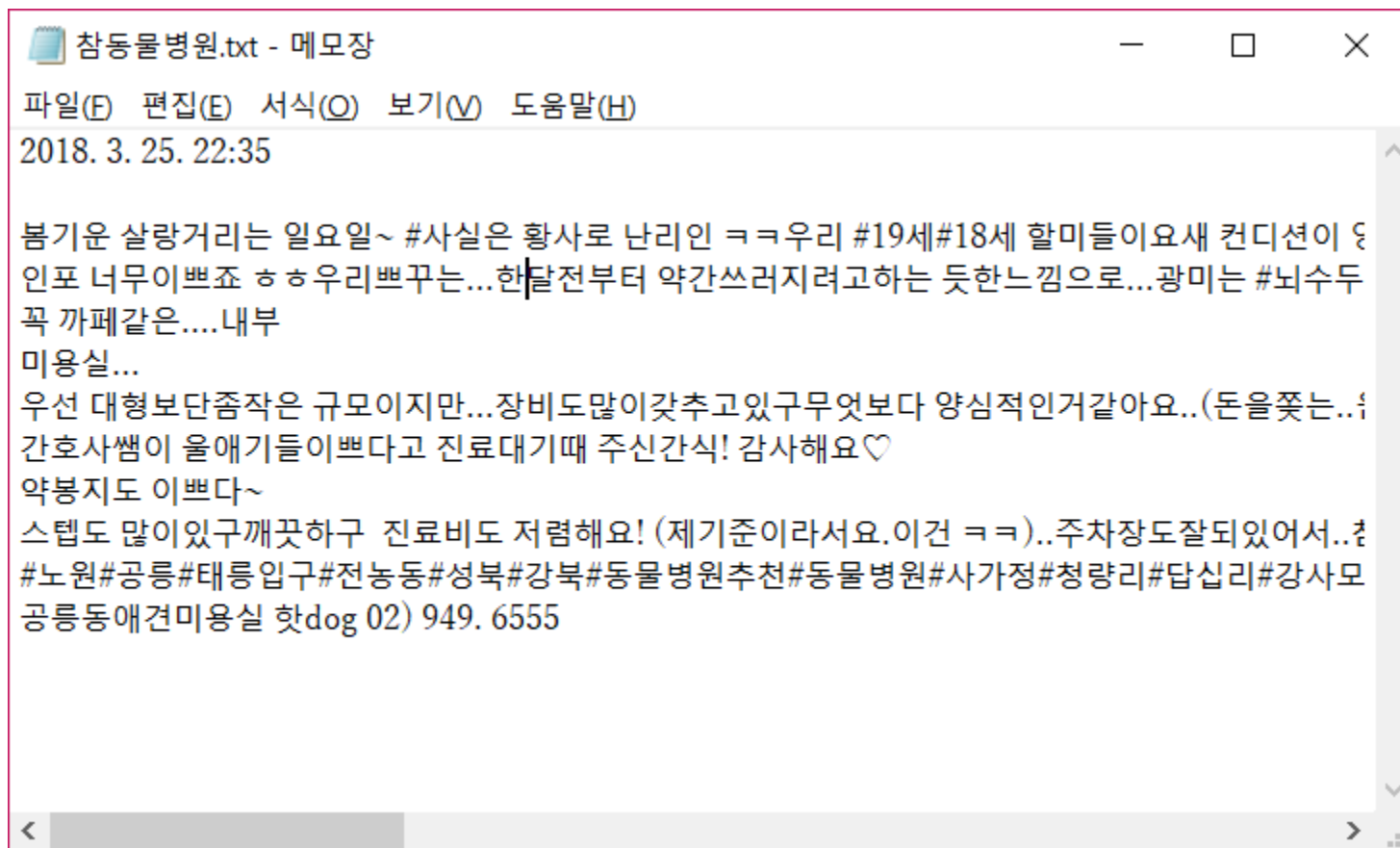
```
processed.dat - 메모장
파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)
"시군구코드":신고일자:업소명:영업상태코드:영업상태:사업장소재지(지번):휴업시작
3100000:19941010:롯데동물병원:0000:정상:서울특별시 노원구 중계동 503번지 시영
3100000:19980711:25시동물병원:0000:정상:서울특별시 노원구 중계동 363번지 대
3100000:19990521:은빛동물병원:0000:정상:서울특별시 노원구 상계동 1110번지 42
3100000:19990820:은마로동물병원:0000:정상:서울특별시 노원구 중계동 173번지 1
3100000:19991007:그린동물병원:0000:정상:서울특별시 노원구 상계동 78번지 15통
3100000:20000407:고려동물병원:0000:정상:서울특별시 노원구 상계동 79번지 145호
3100000:20010523:미래동물병원:0000:정상:서울특별시 노원구 공릉동 94번지 6호 1
3100000:20010613:우리들동물종합병원:0000:정상:서울특별시 노원구 중계동 512번
3100000:20020806:당고개동물병원:0000:정상:서울특별시 노원구 상계동 71번지 22
3100000:20020819:참사랑동물병원:0000:정상:서울특별시 노원구 중계동 366번지 10
3100000:20021230:참동물병원:0000:정상:서울특별시 노원구 상계동 2번지 1호 부영
3100000:20030130:한강동물병원:0000:정상:서울특별시 노원구 상계동 430번지 13호
3100000:20040401:스누피동물병원:0000:정상:서울특별시 노원구 공릉동 441번지 1
3100000:20050217:시마동물병원:0000:정상:서울특별시 노원구 상계동 749번지 5호
3100000:20050729:수 동물병원:0000:정상:서울특별시 노원구 상계동 1265번지 동방
```

- data_download.py에서 받은 .csv 파일에서 필요 없는 데이터를 제거하고,
다음 프로세스가 읽어 들이기 쉬운 형태로 데이터를 파싱 및 저장(processed.dat)
- 폐업한 동물병원의 후기는 크롤링할 이유가 없기 때문에 '폐업'을 포함한 동물병원 데이터는 제거.
- 기존의 큰 따옴표와 반점을 제거하고, 열을 :, 행을 \n로 구분한 .dat 파일로 저장

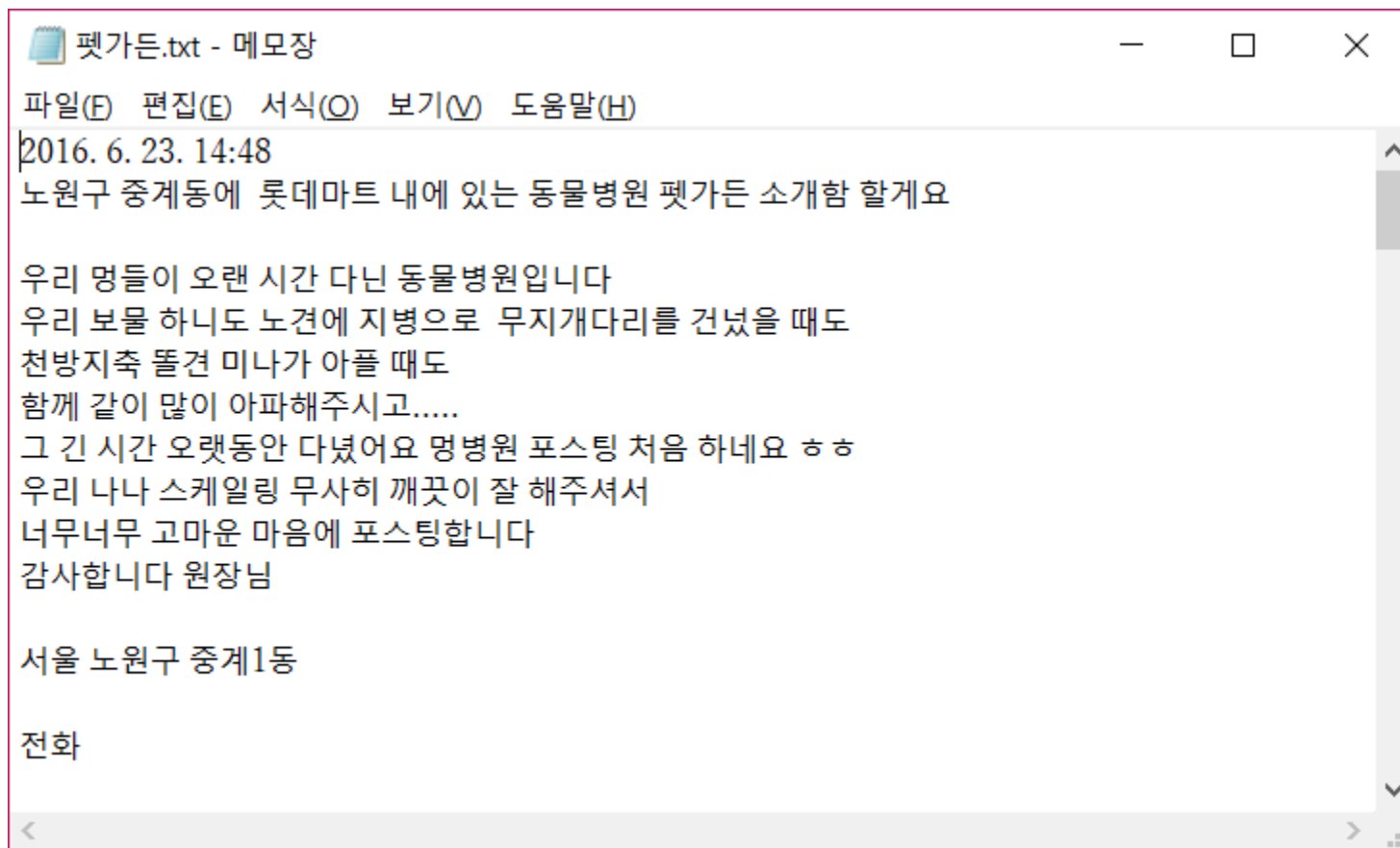
google_crawl.py

- 구글에서 크롤링하는 스크립트
- 과정
 1. processed.dat를 읽고 메모리에 리스트 자료 생성
 2. 구글 검색을 수월하게 하기 위한 몇 가지 키워드를 추가
for 리스트 자료:
 3. 구글 검색 후 그 페이지를 HTML로 추출 후 파싱
 4. 항목의 URL이 특정 키워드(blog, daum 등)를 포함하면 들어 감
 5. 그 페이지의 HTML 추출 후 파싱, 저장 및 URL 저장

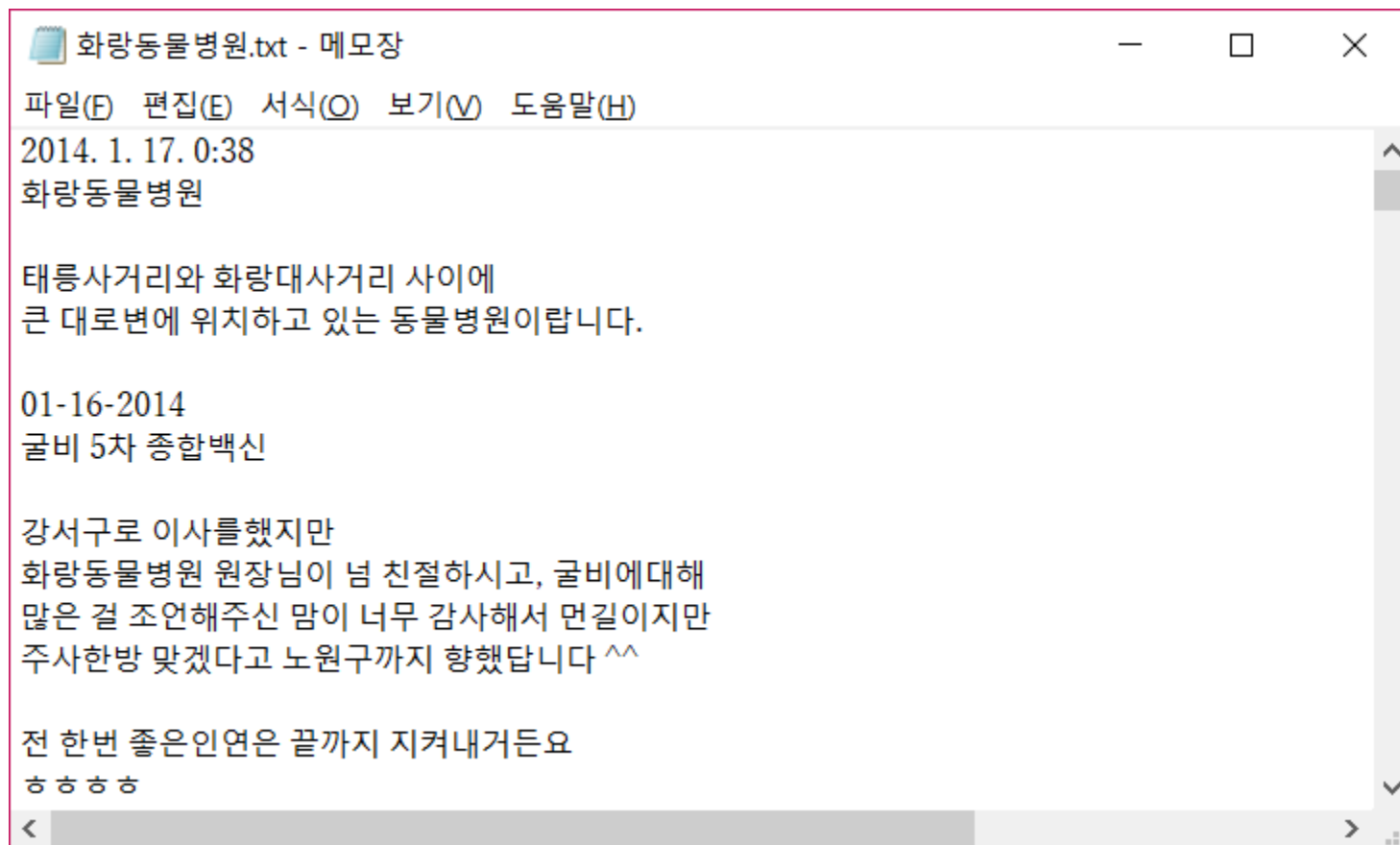
Result of google_crawl.py



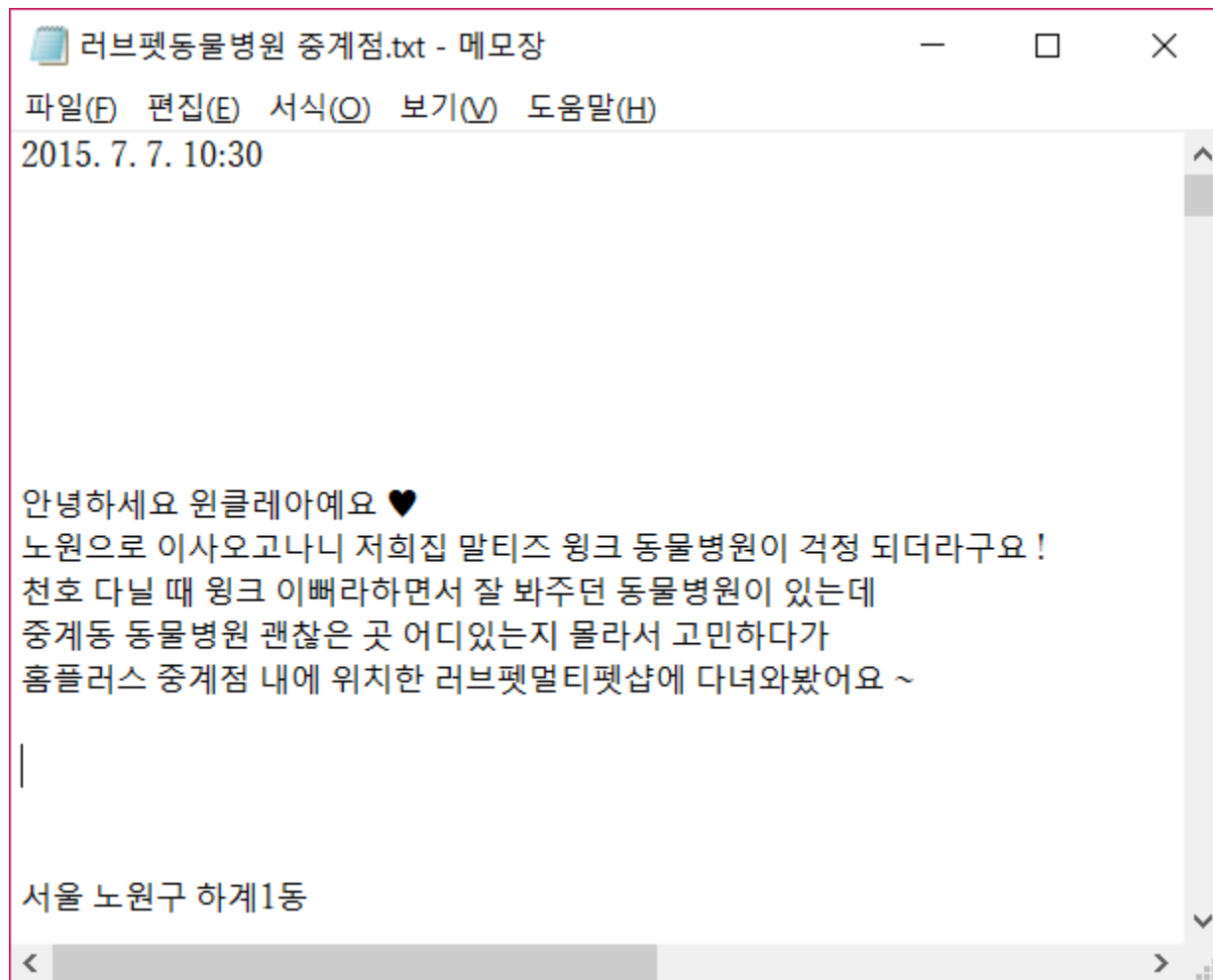
Result of google_crawl.py



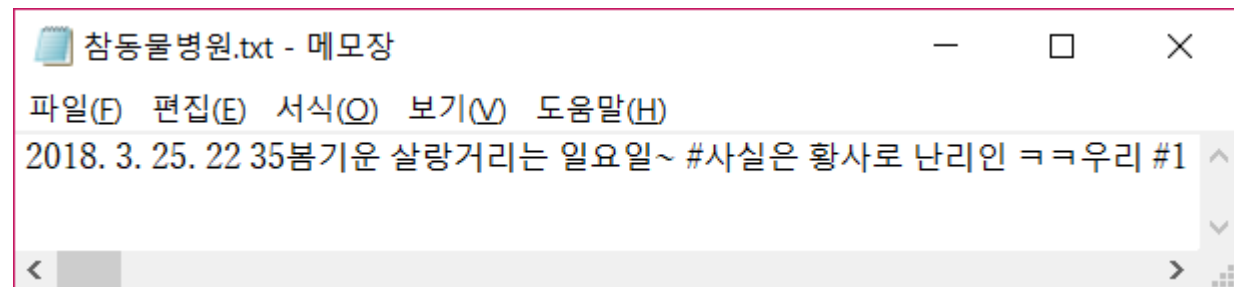
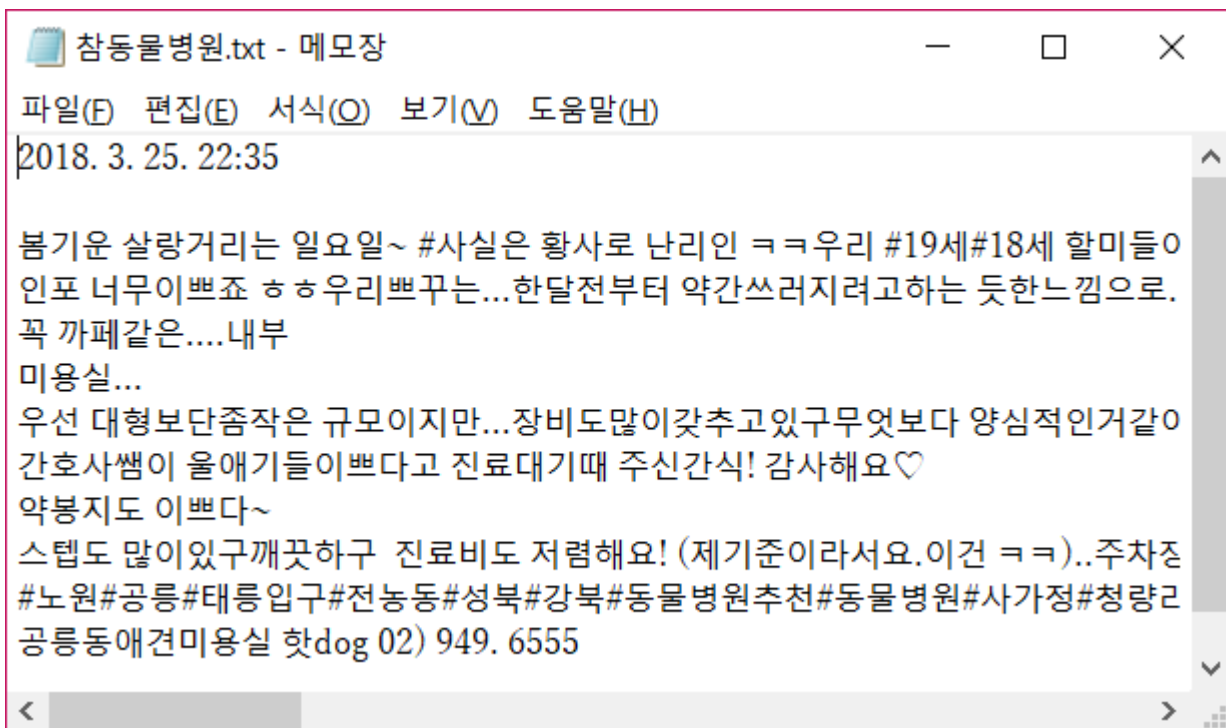
Result of google_crawl.py



Result of google_crawl.py

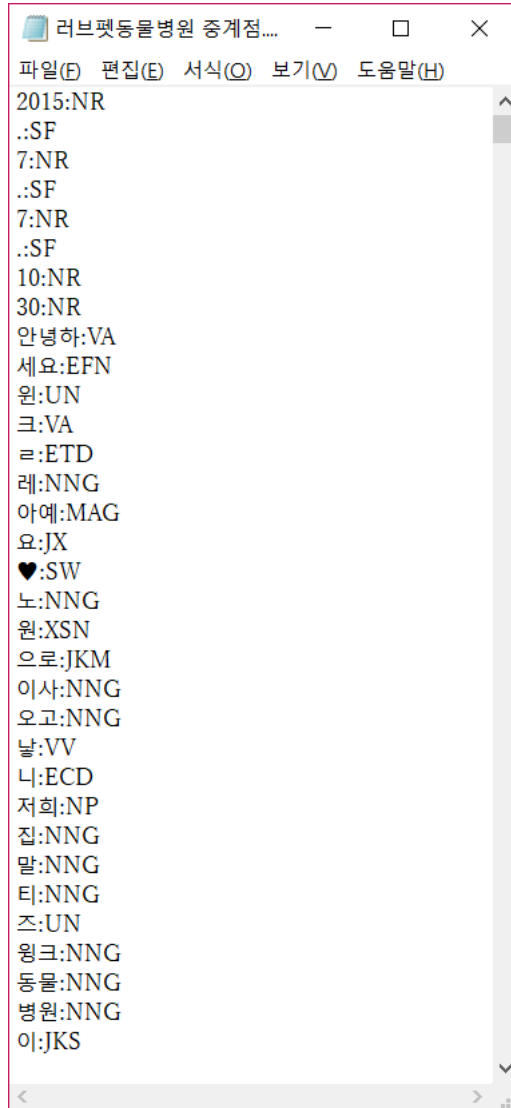


trim_crawled_files.py



- 다음 프로그램에서 처리하기 쉽도록 데이터를 가공
- 크롤링 된 파일들 사이사이의 'wn' 문자와 ':' 문자를 전부 제거

morph_analyser.py



러브펫동물병원 중계점....

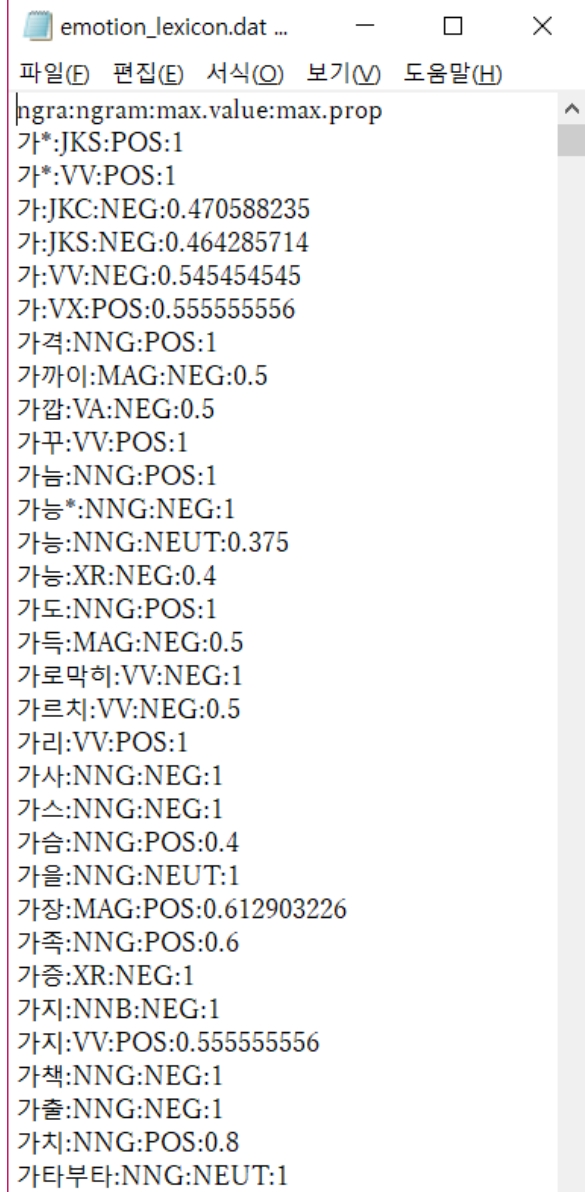
파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

2015:NR
.:SF
7:NR
.:SF
7:NR
.:SF
10:NR
30:NR
안녕하:VA
세요:EFN
원:UN
크:VA
≡:ETD
레:NNG
아예:MAG
요:JX
♥:SW
노:NNG
원:XSN
으로:JKM
이사:NNG
오고:NNG
날:VV
니:ECD
저희:NP
집:NNG
말:NNG
티:NNG
즈:UN
윙크:NNG
동물:NNG
병원:NNG
이:JKS

- 전 프로세스에서 가공한 데이터를 읽은 후 형태소로 분해
- (형태소, ':')\n 형식으로 저장
- 형태소 분해는 KoNLpy 모듈의 Kkma 클래스를 사용

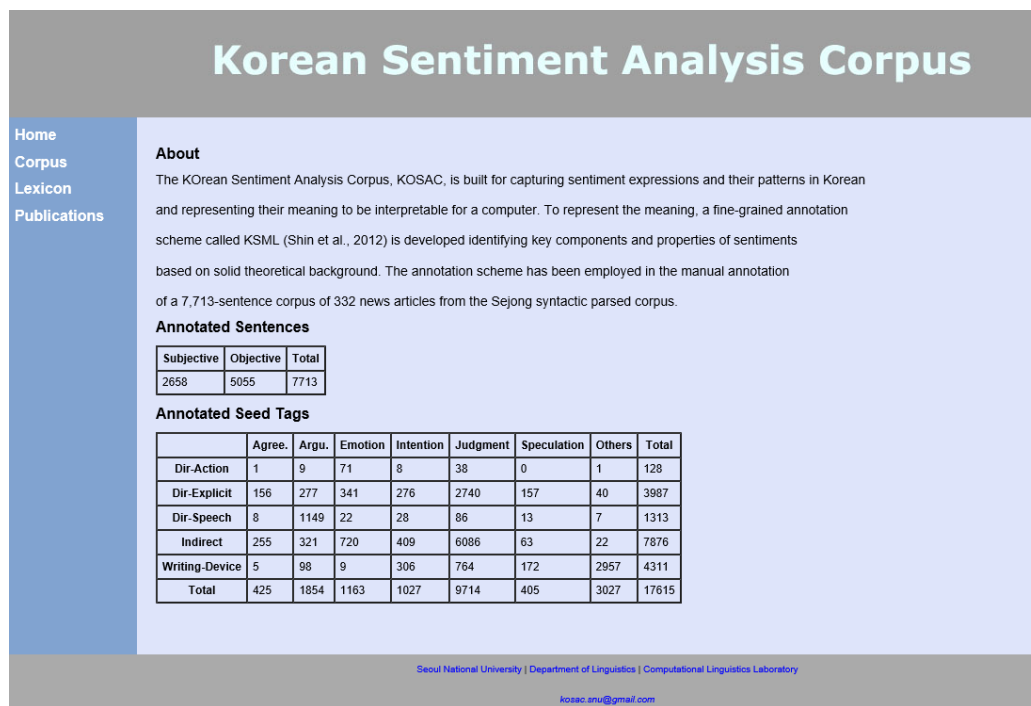
그림 - (형태소, ':')\n 형식

parse_lexicon.py



```
emotion_lexicon.dat ...
파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)
ngra:ngram:max.value:max.prop
가*:JKS:POS:1
가*:VV:POS:1
가:JKC:NEG:0.470588235
가:JKS:NEG:0.464285714
가:VV:NEG:0.545454545
가:VX:POS:0.555555556
가격:NNG:POS:1
가까이:MAG:NEG:0.5
가깝:VA:NEG:0.5
가꾸:VV:POS:1
가늠:NNG:POS:1
가늠*:NNG:NEG:1
가늠:NNG:NEUT:0.375
가늠:XR:NEG:0.4
가도:NNG:POS:1
가득:MAG:NEG:0.5
가로막히:VV:NEG:1
가르치:VV:NEG:0.5
가리:VV:POS:1
가사:NNG:NEG:1
가스:NNG:NEG:1
가슴:NNG:POS:0.4
가을:NNG:NEUT:1
가장:MAG:POS:0.612903226
가죽:NNG:POS:0.6
가증:XR:NEG:1
가지:NNB:NEG:1
가지:VV:POS:0.555555556
가책:NNG:NEG:1
가출:NNG:NEG:1
가치:NNG:POS:0.8
가타부타:NNG:NEUT:1
```

- 이미 구축된 감정사전 데이터에서 필요한 정보만 뽑아내고
감정 점수를 평가하기 좋은 형태의 데이터로 파싱
- “emotion_lexicon.dat”로 저장



Korean Sentiment Analysis Corpus

[Home](#)
[Corpus](#)
[Lexicon](#)
[Publications](#)

About

The Korean Sentiment Analysis Corpus, KOSAC, is built for capturing sentiment expressions and their patterns in Korean and representing their meaning to be interpretable for a computer. To represent the meaning, a fine-grained annotation scheme called KSML (Shin et al., 2012) is developed identifying key components and properties of sentiments based on solid theoretical background. The annotation scheme has been employed in the manual annotation of a 7,713-sentence corpus of 332 news articles from the Sejong syntactic parsed corpus.

Annotated Sentences

Subjective	Objective	Total
2658	5055	7713

Annotated Seed Tags

	Agree	Argu.	Emotion	Intention	Judgment	Speculation	Others	Total
Dir-Action	1	9	71	8	38	0	1	128
Dir-Explicit	156	277	341	276	2740	157	40	3987
Dir-Speech	8	1149	22	28	86	13	7	1313
Indirect	255	321	720	409	6086	63	22	7876
Writing-Device	5	98	9	306	764	172	2957	4311
Total	425	1854	1163	1027	9714	405	3027	17615

Seoul National University | Department of Linguistics | Computational Linguistics Laboratory
kosac.anu@gmail.com

그림 - 파싱된 데이터(emoticon_lexicon.dat)

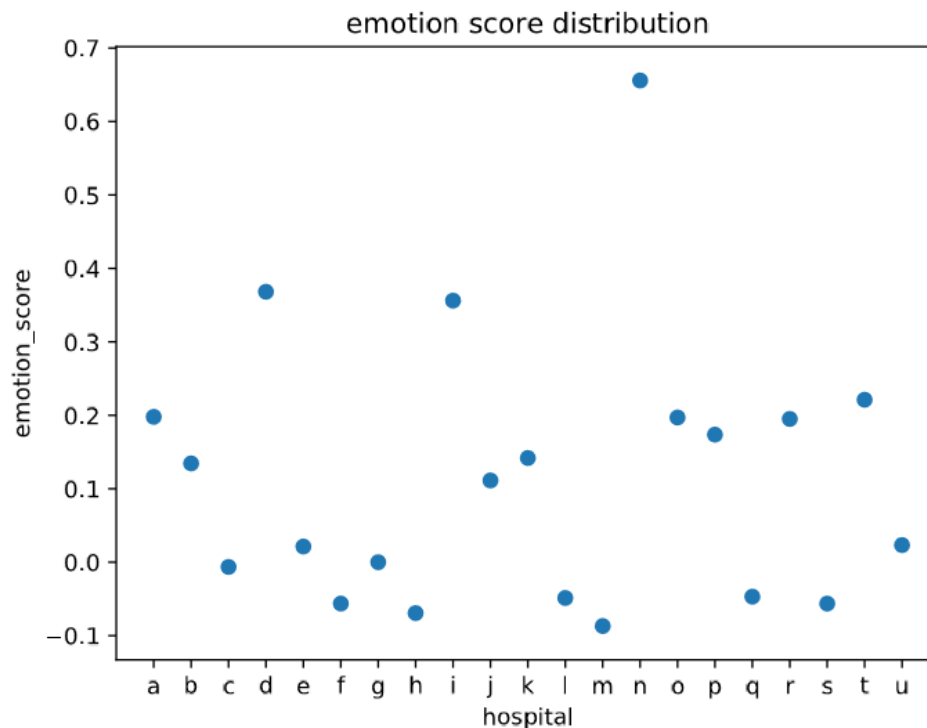
그림 - 감정사전 KOSAC

score_emoticon.py

- 감성 점수를 평가하는 스크립트
- 과정
 1. 감정사전 데이터를 메모리에 파이썬 딕셔너리 자료형으로 저장한다.
 2. 평가받을 데이터를 단어 단위로 딕셔너리 자료형의 key 값([단어, 형태소 종류] 튜플)에 넣어 딕셔너리 자료형의 value 값(감정 점수)를 구한다.
 3. 한 병원에 대한 감정 점수를 모두 더하고 총 단어의 수만큼 나누어 준다.
 - > 그 병원의 감정에 대한 평가
 4. 저장

final_analysis.py

```
statics.txt - 메모장
파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)
Max value: 0.655847273
Min value: -0.08690004458163263
Mid value: 0.1345009172754822
Average value: 0.11554964973400393
Variance: 0.032486318139209706
```



- 각 병원에 대해 평점을 매긴 데이터를 한 emotion_score.txt으로 저장.
- 노원구에 있는 동물병원에 대한 통계수치를 statics.txt에 저장
- 그 후 감정 점수 분포 그래프를 생성하여 저장

```
emotion_score.txt
파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)
(주)한빛종합동물병원.txt: 0.1979221435236770
25시동물병원.txt: 0.1345009172754822
LK 동물병원.txt: No data
고려동물병원.txt: No data
그린동물병원.txt: -0.006415427587078653
닥터스종합동물병원.txt: No data
당고개동물병원.txt: 0.3682415134285715
러브펫동물병원 중계점.txt: 0.0213144711205128
롯데동물병원.txt: -0.0563546490639098
메인동물메디컬센터.txt: No data
미래동물병원.txt: No data
브라운동물병원.txt: No data
서정동물병원.txt: 0.0
수 동물병원.txt: No data
수락 동물병원.txt: -0.06925054287749997
스누피동물병원.txt: No data
스마일플러스동물병원.txt: 0.3561379354285714
스마트동물병원.txt: 0.11128305442253514
시마동물병원.txt: 0.14188524334938257
아란종합동물병원.txt: No data
엔젤동물병원.txt: No data
우리들동물종합병원.txt: -0.04869295120215628
우솔동물병원.txt: -0.08690004458163263
웰니스 동물병원.txt: No data
유림동물병원.txt: No data
은마로동물병원.txt: No data
은빛동물병원.txt: No data
이앤박동물병원.txt: 0.655847273
주식회사 노원24시동물의료센터.txt: 0.19702271
중계이오동물병원.txt: 0.17365709020572925
참동물병원.txt: -0.046864827123076915
참사랑동물병원.txt: No data
쿨펫동물병원.txt: No data
```

PART7.

Evaluation

Method of Evaluation

- 감정 점수가 평가된 글 중 몇 가지 표본을 고른다.
- 표본의 URL에 들어가서 직접 글을 읽어본 후 평가된 감정 점수가 적절한지 주관적으로 판단한다.
- 이 보고서에서는 표본을 각각 감정 점수의 최솟값, 최댓값, 평균에 근접한 값을 갖는 페이지로 하였다.

1. 최대 평점 – 이앤박 동물병원(1)

도치 일상

이앤박 동물병원 다녀왔어요/노원구 상계동

안녕하세요~ 빅뱅소녀 입니다!!

오늘은 며칠 전에 쫘이 구충하느라 동물병원에 다녀왔습니다! 도치들도 보는 동물병원인 이앤박 동물병원입니다.

선생님이 들어오시고

쫘이 외관상태를 보기 위해서

선생님이 꺼내려고 하는데...

저희는 거의 보지 못해서 놀랐어요.

막 딸꾹질같이 꺽꺽 거리면서

손이 닿으면 펄쩍 뛰더라고요~

그래서 장갑을 가지고 오시면서 겨우 꺼냈어요.

그런데 헐~ 꺼내니 완전 순진순진 모드!

무게를 재봤더니 360g정도 나왔네요~

생각보다 적게 나가네요?

혹시 도치들 평균 몸무게 아시는 분 있나요?

암튼 본론으로 넘어가서...

내시경 같은거로 등을 보시더니 톱밥이랑

가루가 많이묻어있다고 베딩을 바꾸라고 하시네요ㅠㅠ 미안해~

이제 구충할차례!!!

주사기에다가 약을 넣고

찔러넣는게 아니라 흡수할 수 있게

등가시에 부분마다 몇방울씩 넣어주는 모습이

엄청~~신기했어요!!

그래서 겨우 안아서 한번 보시더니

아주 순하고 예쁘다고 훈녀라는 칭찬 들었네요^^

역시 우리 쫘이는 저를 닮...ㅋ아서 ㄱ...그런가?

요즘 미모에 물이오르고 있는 우리 쫘이 양!

요런식으로 가시에 묻혀 주시더라고요.

쫘이는 가끔씩 움츠리고 가만히 있었어요.

여러분들도 한번 가보세요!

강추강추~~

이상 빅뱅소녀의 포스팅 이였습니다~

전화번호는 02-939-6898입니다~

1. 최대 평점 – 이안박 동물병원(2)

- 블로그 글에 대해 사람이 읽고 평가했을 때 긍정적
- 즉, 프로그램이 적절하게 평가하였음(평점 약 0.65 - 중상 정도의 긍정)
- 다만, 이 병원에 대해 평가한 글이 하나 밖에 없어 신뢰도가 떨어짐.
-> 즉, 더 많은 데이터를 구할 방안이 필요

2. 최소 평점 – 우솔 동물병원(1)

[석계역 동물병원] 우솔동물병원 _ 강아지 백내장 검사 비용, 설명 듣고왔어요

나가자는말에 눈뚱그래져서 뛰어나온 와와

그래놓고 눈바닥은 밟지도 못하는 가련한 치와와같으니라구

와와는 어릴때부터 눈동자에 하얀끼가 조금 보였어요.

문제 없다고는 생각했는데 주변에 나이많은 댕댕이 키우시는 분들이, 이거 백내장일수도있다고

진행 많이안된 백내장이라면 일찍 검사하고 안약 처방받으면 거의 안오다시피 늦춰진다고 경험을

말해주길래 혹시나해서 검사받으러갔어요

오픈 10:00

클로즈 8:30

시츄한마리도 어느샌가부터 생김

백내장 검사 비용은 22000원 이었어요
생각보다 넘나 저렴한것?

화요일은 문은열려있는데 의사선생님 진료 휴무예요.

화요일에 갔다가 안계셔서 다시 수요일에 방문했어요.

백내장의 위험이 있다해도 이 검사한번으로 계속 약을 처방받을수있기도하고

백내장이 아니더라도 기분좋게 안심하고 집에갈수있는 비용 ☆

석계역 우솔 동물병원은 석계역 2번 출구 앞에있어요.

여기 동물병원은 제가 와와를 데려온게 2011년 3월쯤음인데 그때 탈장수술 바로 하면서부터 다니기 시작한 곳이에요.

아무래도 집앞이라 가깝기도하고

이것저것 검사기록이 남아있어서 떼데안가고 여기가는편이에요.

한동안 이사가있는동안 돌곳이에 대학동물병원인가 가서 눈탱이 엄청 맞음^^^^.....

항상 이안에 어슬렁어슬렁 다니는 냥냥이

고양이는 볼때마다 너무 신비롭게 생긴거같지만

난 고양이는 평생 안키울거같아유

나랑 습성이 너무 안맞아

그러면서 백내장 1기부터 말기까지 증상을 설명해줬는데

2기에서 확연히 나타나는 증상이

저기 그려진건처럼 Y자로 수정체가 갈라지는 거래요.

참에 이해가 안됐는데

사진보니 이해완료!

여튼 눈 건강해서 기분이 좋게 집왔어요 :)

2. 최소 평점 – 우술 동물병원(2)

[석계 동물병원] 석계역 우술동물병원

곰양의 집에서는 슈나우저를 기릅니다.
3대 지랄견이라고도 하는데 비글에 이어.... 키우기 힘들다는 슈
슈나우저 중에서도 더욱 유별난 개 우리집 슈나는
동물병원에서도 유명합니다.

병원가기전부터 도살장에 끌려가는 소처럼
동네 떠나가라 울어댑니다.
마치
내가 지금 병원으로 가고있으니 우리 주인좀 말려달라고
외치는것 처럼 말이죠

병원에서도 우리 '나무'(애견이름)가 도착하기 5분전에
이미 인지를 하고있더군요
왕왕 짖으면서 병원으로 가니까요 -_-
동네 창피 합니다.

석계역 우술동물병원에 가면 요 녀석이 있습니다. 우술동물병원에는 고양이 1마리, 개1마리가 있는데 우술동물병원 원장님 사진 도촬 (죄송해용~~~)
저는 개인적으로 이 냥이 녀석이 매력돈아요~

참 도도합니다.

우술동물병원에서 진료받으려고 기다리는데 저 위에서 저를 쳐다보고 있어

흡짓! -_-+

석계역 우술동물병원 입구 입니다.
굉장히 작죠? 인테리어 보다는 원장님의 정직한 실력과 친절함이
지속적으로 이용하게 만드는 것 같습니다.

개 바리깡 살때도 원장님은 인터넷에서 주문하시라고
여기서 사시면 비싸다고 정직! 아..매력 돈아요 석계역 우술동물병원 위 원장님!

'나무'가 계속 몸을 굽길래 병원에 데려갔더니 역시나 귓병이 하..
사람먹는거 계속 주냐고... 묻길래.....

" 안줘요, 저희는 주는거 없는데 몰래몰래 뭘 훔쳐먹는거 같긴 해요 -_-;"

이녀석이 성견이 되더니 밥상위고 김치냉장고 위고 올라가서
마구마구 훔쳐먹다가 걸리면 엄마한테 뒤통이맞고 결국
몸이 가려워서 굶다가 우술동물병원으로 데려옵니다 ㅠ.ㅠ

우술동물병원에 적금 붓는것도 아니고
그만!!!훔쳐먹으란 말이다!!!

진료하는 내내 어찌나 어리광을 피우던지 자기 죽는다고 소리소리를 지르고
우술동물병원 위원장님께 참 민망할정도로 버둥거리는 바람에..
뭐라 드릴 말씀이 ...

결국 병원 도착해서도 안들어 가려고 난리치지만 억지로 끌고 들어갑니다.
'나무'가 우리집에 온 첫날부터 3년 반 시간동안 진료를 받고 있는
석계역에 위치한 '우술동물병원' 입니다.

석계역 2번 출구에 위치해있고, 작은 동물병원입니다.
주말에는 늘 사람과 애견으로 북적이고, 진료 혹은 사료를 사기위해
지나가다 사람들이 많이 들릅니다.

원장님이 워낙 친절하셔서 그런듯 싶네요

진료후 결국 약을 타서 집에왔습니다.
ㄷ 집에 돌아올때는 소리도 안지르고 입 꼭 다물고 집으로 오더라고요
간식도 그만줘야겠습니다.

우술동물병원 덕에 나무는 늘 건강하고 털에 윤기가 찰찰흐르지만
자주 병원에 가는 덕에;;; 민망해 죽겠음~~ ㅠ

관동, 장위동 근처에 동물병원 다 다녀봤지만 우술동물병원 원장님 만큼
친절하고 정직하게 진료하는 분 아직 보지 못했네요~

늘 감사하게 생각하고있어요~~~
나무가 발버둥 쳐도 그저 미소로만 대답하시니 ^^

우술 동물병원원 석계역 2번출구에 있습니다~
노원구 월계동 26-13

2. 최소 평점 – 우솔 동물병원(3)

- 최소 평점이지만, -0.11로 거의 중립에 가까운 수치임. 약한 부정.
- 직접 읽어본 결과, 감정사전으로 단어에 대해 통계처리를 하여, 부정적인 단어가 많았음.
하지만, 동물병원에 대한 글인 만큼 동물이 아프다거나 동물이 무서워한다는 글 자체를 부정적으로 봄.
-> 즉, 전체적인 문맥을 파악해야 하는 Machine Learning기법이 필요하다는 것을 알게 됨.
- 감정사전으로만 동물병원 후기를 분석한다면, 단어에 대한 몇 가지 지표를 설정을 해주는 방안이 있음.
예를 들면, 동물병원에 대해 부정적인 단어로는 비싸다, 불 친절하다, 효과가 없다로 미리 설정 해둠.
긍정적인 단어로는 친절, 저렴 등이 있을 것임.
이런 단어를 한정적으로 긍정, 부정을 나누어 감정사전을 새로 구축할 필요성을 느낌

3. 평균 평점 – 스마트 동물병원(1)

 <p>스마트동물병원(노원) @nwsmart1 · 7월 18일 연봉도 쎄데 차 하나 사면 되지... naver.me/5c</p> <p>🗨️ ↻️ ❤️</p>	 <p>스마트동물병원(노원) @nwsmart1 · 2016년 9월 12일 노원 스마트 동물병원 명절 진료시간 안내 blog.naver.com/nwsm</p> <p>🗨️ ↻️ ❤️</p>	 <p>스마트동물병원(노원) @nwsmart1 · 2016년 8월 15일 [길냥이]아기 고양이의 가족을 찾습니다._노원 스마트 동물병원 blog.naver.com/nwsmart/220788...</p> <p>🗨️ ↻️ ❤️</p>
 <p>스마트동물병원(노원) @nwsmart1 · 4월 4일 80억 루킹3진... naver.me/5oldl4NB</p> <p>🗨️ ↻️ ❤️</p>	 <p>스마트동물병원(노원) @nwsmart1 · 2016년 8월 17일 푸딩이의 식도이물 치료기 _ 강아지 내시경 _ 노원구 스마트 동물 잘못 삼켰어요 blog.naver.com/nwsmart/220790...</p> <p>🗨️ ↻️ ❤️</p>	 <p>스마트동물병원(노원) @nwsmart1 · 2016년 7월 15일 노원 동물병원_ 셀프 반려동물 이름표! _스마트동물병원노원점 blog.naver.com/nwsmart/220763...</p> <p>🗨️ ↻️ ❤️</p>
 <p>스마트동물병원(노원) @nwsmart1 · 4월 3일 성육이 맘 속으로 욕한사람 반성하자 일단 나부터 naver.me/GOOWI44F</p> <p>🗨️ ↻️ ❤️</p>	 <p>스마트동물병원(노원) @nwsmart1 · 2016년 8월 17일 치와와 찌개의 슬개골탈구 수술 후기_ 강아지 슬개골탈구 4기 _ 원구 스마트동물병원 blog.naver.com/nwsmart/220790...</p> <p>🗨️ ↻️ ❤️</p>	 <p>스마트동물병원(노원) @nwsmart1 · 2016년 7월 12일 노원 동물병원_ 개냥이 연이와 상자 ♡_노원스마트동물병원 blog.naver.com/nwsmart/220760...</p> <p>🗨️ ↻️ ❤️</p>
 <p>스마트동물병원(노원) @nwsmart1 · 2017년 켈크라이 ππππ naver.me/Fxiuhnvk</p> <p>🗨️ ↻️ ❤️</p>	 <p>스마트동물병원(노원) @nwsmart1 · 2016년 8월 17일 송이의 각막궤양 치료기 _ 강아지 각막궤양 수술 _ 안과 수술 _ 노 병원 blog.naver.com/nwsmart/220789...</p> <p>🗨️ ↻️ ❤️</p>	 <p>스마트동물병원(노원) @nwsmart1 · 2016년 7월 6일 노원 동물병원_ 고양이+박스=환상!! _노원 스마트동물병원 blog.naver.com/nwsmart/220756...</p> <p>🗨️ ↻️ ❤️</p>
 <p>스마트동물병원(노원) @nwsmart1 · 2016년 말티즈 포포의 강아지 슬개골탈구 수술 후기</p> <p>🗨️ ↻️ ❤️</p>	 <p>스마트동물병원(노원) @nwsmart1 · 2016년 8월 17일 둘리의 발가락 종괴 치료기 _ 발에 혹이 났어요_ 노원구 스마트동 blog.naver.com/nwsmart/220789...</p> <p>🗨️ ↻️ ❤️</p>	 <p>스마트동물병원(노원) @nwsmart1 · 2016년 7월 4일 노원 스마트동물병원 타니의 잠복고환 증성화 수술 / 잠복고환수술 / 노원구 증성 화수술 blog.naver.com/nwsmart/220753...</p> <p>🗨️ ↻️ ❤️</p>
 <p>스마트동물병원(노원) @nwsmart1 · 2016년 애견미용 ♥ blog.naver.com/nwsmart/22083...</p> <p>🗨️ ↻️ ❤️</p>	 <p>스마트동물병원(노원) @nwsmart1 · 2016년 8월 17일 젤리의 눈 치료기 _ 고양이 안내 출혈 _ 눈 안이 빨개요 _ 스마트동 blog.naver.com/nwsmart/220789...</p> <p>🗨️ ↻️ ❤️</p>	 <p>스마트동물병원(노원) @nwsmart1 · 2016년 7월 1일 빠져나올 수 없는 고양이 마약 간식.. 연이 중독되다..._노원스마트동물병원 blog.naver.com/nwsmart/220751...</p> <p>🗨️ ↻️ ❤️</p>

3. 평균 평점 – 스마트 동물병원(2)

스마트 일상

노원 동물병원_ 고양이+박스=환상!! _노원 스마트동물병원

스마트동물병원 노원점
2016. 7. 7. 15:36

안녕하세요!

일상지기가 돌아왔습니다!



이번에 저희 스마트 동물병원 노원점에 물품이 잔뜩 들어왔어요~

이쁜 머리핀도 들어오고 다양한 목줄과 신발 방석 등등

이 많은 제품들이 들어오면서 큰~박스가 생겼는데요!

박스.. 박스 하면... 고양이!!!!!!!!!!!!!!

고양이가 좋아하는 박스!!

고양이하면 연이!!!

해서..

연이에게 큰 박스를 선물해주기로 했습니다!

이왕 큰 박스가 생겼는데 그냥 줄 수는 없지요!

열심히 이리 뚫고 저리 뚫고 구멍 뿡뿡뿡!

뚫어서 연이한테 속- 줘봤어요!



역시 우리 연이야..

바로 쓱~! 들어가서 푹망푹망 큰 눈으로 주변을 살피는군요!

자.. 이제 본격적으로 한번 놀아볼까..



허술한 박스지만 연이가 너무너무 좋아하는 걸 보니 제가 다 기분이 좋네요



3.평균 평점 – 스마트 동물병원(3)

- 프로그램에서 평가한 이 동물병원의 감정 점수는 약 0.11정도(약한 긍정)이다.
- 그러나, 직접 읽어본 결과 트위터 글은 특성 상, 문장이 짧아 의미 있는 정보를 찾기 어려웠고, 때문에 사람이 점수를 평가하기에 문제가 있었다.
 - > 데이터 수집을 할 때, 고려해야할 사항에 추가
- 또, 위 글 모두 동물병원에서 작성한 광고성 글임.
 - > 광고를 구별할 수 있도록 장치를 마련해야함.

PART8.

Further plans

프로젝트 개선 방향

- 주어가 병원 혹은 의사, 원장님, 동물병원인 문장일 경우에만 감정 분석을 하도록 함
- 부정어와 긍정어를 글의 목적(동물병원 평가)에 맞게 한정해야 함
 - 슬프다, 걱정되다 등의 부정어 -> 부정어에서 제외
 - 친절하다, 비싸다 등 동물병원 평가에 중요한 단어 -> 가중치 크게
- 문맥을 포함하는 다차원적 분석 도입
 - > 형태소 단위가 아닌 문장 단위로 분석, Machine Learning 기법 도입
- 더 많은 후기 데이터를 구할 방안 마련
- 광고성 글이 아닌 신뢰성 있는 후기 데이터를 구할 방안 마련.
- 코드 리팩터링(코드 개선 및 함수, 클래스화 필요성)