**사용전 환경설정**

1. 필요한 폴더 만들기
2. 설정 부분을 폴더 위치에 맞게 고치기
3. docoment 파일 다운로드 후 압축해제하기
4. 필요한 모듈 설치하기
5. **필요한 폴더 만들기 & 설정 부분을 폴더 위치에 맞게 고치기**

현재위치: medicine.py 파일이 있는 위치

현재위치에 아래의 파일들을 만든다.

의약품 정보 엑셀 파일 저장 폴더 : download

documentxxx.json 파일 저장 폴더 : document

dbxxx.json 파일 저장 폴더 : db

db.csv(dbxxx.json 파일들을 csv파일로 변환한 것)의 저장 폴더 : dbcsv

chart엑셀 파일 저장 폴더 : chart

chart 엑셀 파일을 csv로 변환한 파일 저장 폴더 : chart\_db

document의 xml파일 저장 폴더 : unzip

unzip/1954-2000 : 1954~2000의 document 파일들을 저장

unzip/2001-2005 : 2001~2005의 document 파일들을 저장

unzip/2006-2010 : 2006~2010의 document 파일들을 저장

unzip/2011-2015 : 2011~2015의 document 파일들을 저장

unzip/2016-2020 : 2016~2020의 document 파일들을 저장

허가년도별 XML 압축 파일 저장 폴더 : zipfiles

폴더 구조의 예시는 아래와 같다.

텍스트이(가) 표시된 사진

자동 생성된 설명

1. **설정 부분을 폴더 위치에 맞게 고치기**

“#설정”이라고 적힌 밑 부분의 excelfile\_path, documentfile\_path, dbfile\_path, csvfile\_path, chart\_excelfile\_path, chart\_db\_csvfile\_path, xmlfile\_common\_path의 경로를 현재 내 환경에 맞게 수정한다.

1. **docoment 파일 다운로드 후 압축해제하기**

아래 링크로 접속해서

<https://nedrug.mfds.go.kr/pbp/CCBGA01/getItem?totalPages=6&limit=10&page=2&&openDataInfoSeq=12>

허가년도별 XML 파일을 zipfiles 폴더에 다운받는다.

그리고 파이썬으로 아래 코드를 실행해 unzip/xxxx-xxxx 폴더 아래에 압축 해제한 파일들을 넣는다

아래 코드는 압축파일 1개당 1번씩 모두 실행해주어야 한다.

import zipfile

zip\_file\_path = ‘zipfiles/2016-2020.zip’

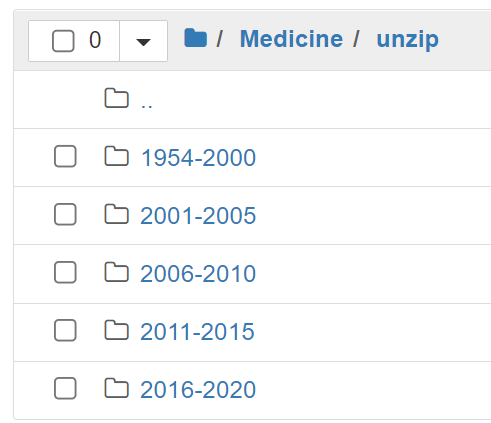
extract\_path = ‘unzip/2016-2020’

with zipfile.ZipFile(zip\_file\_path, 'r') as zip\_ref:

zip\_ref.extractall(extract\_path)

실행 완료하면 다운로드한 허가년도별 xml 압축 파일이 압축 해제가 되어 각 unzip/xxxx-xxxx폴더에 넣어져 있을 것이다.

폴더 구조의 예시는 아래와 같다.

테이블이(가) 표시된 사진

자동 생성된 설명

1. **필요한 모듈 설치하기**

아래 3개의 모듈을 설치한다.

!pip install xmltodict

!pip install pytz

!pip install requests

**사용 절차**

1. **최초 생성**

Db: 최초의 db를 생성하기 위해 cmd 창에서 medicine.py 파일이 위치한 곳으로 간 후 아래의 명령어를 실행한다.

medicine.py -a

Document: 최초의 document를 생성하기 위해 아래의 명령어를 실행한다.

medicine.py -e

1. **업데이트**

DB: 이후 db를 업데이트 시 아래의 명령어를 실행한다. db 업데이트는 오늘 날짜로부터 60일전~30일전의 업데이트 내용을 찾아서 업데이트 한다. -b 명령어는 허가날자를 기준으로 업데이트를 하고, -c 명렁어는 변경일자를 기준으로 업데이트 하고, -d는 취소날짜를 기준으로 업데이트 한다.

medicine.py -b

medicine.py -c

medicine.py -d

Document: document 업데이트 시 아래의 명령어를 실행한다.

medicine.py -f

1. **CSV 파일로 변환**

DB: dbxxx.json 생성 후 csv 파일 형태로 바꾸기 위해서는 아래의 명령어를 실행한다.

medicine.py -g

Chart: chart 엑셀 파일 다운로드 후 csv 파일 형태로 바꾸기 위해서는 아래의 명령어를 실행한다.

medicine.py -i

medicine.py -j

1. **CSV 파일을 database에 넣기**

DB:

스키마명 : medicine

medicine 스키마를 생성하고 dbcsv폴더에 저장된 db.csv 파일을 import한다.

(mysql workbench -> medicine 스키마 오른쪽 클릭 -> Table Data Import Wizard -> Browse -> dbcsv폴더 -> db.csv파일선택)

테이블명 : medicine (의약품 정보를 갖고있는 테이블)

|  |  |  |  |
| --- | --- | --- | --- |
| 필드명 | Datatype | 특징 | 예시 |
| SEQ | INT | PK, NN, AI | 1 |
| ITEM\_SEQ | TEXT | Default : NULL | 195500002 |
| ITEM\_NAME | TEXT | Default : NULL | 종근당염산에페드린정 |
| BAR\_CODE | TEXT | Default : NULL | 8806433032005 |
| CANCEL\_DATE | TEXT | Default : NULL | 20200101 |
| CANCEL\_NAME | TEXT | Default : NULL | 유효기간만료 |
| CHANGE\_DATE | TEXT | Default : NULL | 20151217 |
| CHART | TEXT | Default : NULL | 본품은 백색의 정제다. |
| EDI\_CODE | TEXT | Default : NULL | 644902212,644902220 |
| EE\_DOC\_ID | TEXT | Default : NULL | https://nedrug.mfds.go.kr/pbp/cmn/  pdfDownload/195500002/EE |
| ENTP\_NAME | TEXT | Default : NULL | (주)종근당 |
| ETC\_OTC\_CODE | TEXT | Default : NULL | 전문의약품 |
| ITEM\_PERMIT\_DATE | TEXT | Default : NULL | 19550117 |
| MAKE\_MATERIAL\_FLAG | TEXT | Default : NULL | 완제의약품 |
| MATERIAL\_NAME | TEXT | Default : NULL | 1정 중 80밀리그램|염산에페드린|KP|25|밀리그램 |
| NARCOTIC\_KIND\_NAME | TEXT | Default : NULL | 한외마약 |
| NB\_DOC\_ID | TEXT | Default : NULL | https://nedrug.mfds.go.kr/pbp/cmn/  pdfDownload/195500002/NB |
| NEWDRUG\_CLASS\_NAME | TEXT | Default : NULL | 신약 |
| PACK\_UNIT | TEXT | Default : NULL | 12, 25, 50, 100, 500정 |
| PERMIT\_KIND\_NAME | TEXT | Default : NULL | 허가 |
| REEXAM\_DATE | TEXT | Default : NULL | 2020-03-19~2024-03-18 |
| REEXAM\_TARGET | TEXT | Default : NULL | 재심사대상(6년) |
| STORAGE\_METHOD | TEXT | Default : NULL | 기밀용기에 넣어 건냉암소에 보관. |
| UD\_DOC\_ID | TEXT | Default : NULL | https://nedrug.mfds.go.kr/pbp/cmn/  pdfDownload/195500002/UD |
| VALID\_TERM | TEXT | Default : NULL | 제조일로부터 36 개월 |

Chart:

스키마명:medicine

medicine 스키마에 chart\_db폴더에 저장된 chart\_yyyymmdd.csv 파일을 import한다.

(mysql workbench -> medicine 스키마 오른쪽 클릭 -> Table Data Import Wizard -> Browse -> chart\_db폴더 -> chart\_yyyymmdd.csv 파일선택)

테이블명 : chart (의약품 특징 정보를 갖고 있는 테이블)

|  |  |  |  |
| --- | --- | --- | --- |
| 필드명 | Datatype | 특징 | 예시 |
| ITEM\_SEQ | TEXT | Default : NULL | 200808876 |
| ITEM\_NAME | TEXT | Default : NULL | 가스디알정50밀리그램(디메크로틴산마그네슘) |
| ENTP\_SEQ | TEXT | Default : NULL | 19540006 |
| ENTP\_NAME | TEXT | Default : NULL | 일동제약(주) |
| CHART | TEXT | Default : NULL | 녹색의 원형 필름코팅정 |
| ITEM\_IMAGE | TEXT | Default : NULL | https://nedrug.mfds.go.kr/pbp/cmn/itemImageDownload/147426403087300104 |
| PRINT\_FRONT | TEXT | Default : NULL | IDG |
| PRINT\_BACK | TEXT | Default : NULL | LT |
| DRUG\_SHAPE | TEXT | Default : NULL | 원형 |
| COLOR\_CLASS1 | TEXT | Default : NULL | 연두 |
| COLOR\_CLASS2 | TEXT | Default : NULL | 하양 |
| LINE\_FRONT | TEXT | Default : NULL | - |
| LINE\_BACK | TEXT | Default : NULL |  |
| LENG\_LONG | TEXT | Default : NULL | 7.6 |
| LENG\_SHORT |  |  | 7.6 |
| THICK | TEXT | Default : NULL | 3.6 |
| IMG\_REGIST\_TS | TEXT | Default : NULL | 20100326 |
| CLASS\_NO | TEXT | Default : NULL | 20211 |
| CLASS\_NAME | TEXT | Default : NULL | 기타의 소화기관용약 |
| ETC\_OTC\_CODE | TEXT | Default : NULL | 전문의약품 |
| ITEM\_PERMIT\_DATE | TEXT | Default : NULL | 20080820 |
| FORM\_CODE\_NAME | TEXT | Default : NULL | 당의정 |
| MARK\_CODE\_FRONT\_ANAL | TEXT | Default : NULL | P,d |
| MARK\_CODE\_BACK\_ANAL | TEXT | Default : NULL | 필름코팅정 |
| MARK\_CODE\_FRONT\_IMG | TEXT | Default : NULL | https://nedrug.mfds.go.kr/pbp/cmn/itemImageDownload/147938640332900133 |
| MARK\_CODE\_BACK\_IMG | TEXT | Default : NULL | https://nedrug.mfds.go.kr/pbp/cmn/itemImageDownload/147938640332900082 |
| MARK\_CODE\_FRONT | TEXT | Default : NULL | r0049 |
| MARK\_CODE\_BACK | TEXT | Default : NULL | r0027 |
| CHANGE\_DATE | TEXT | Default : NULL | 20210422 |

**기능 요약**

|  |  |  |
| --- | --- | --- |
| 번호 | 명령어 | 기능 |
| 1 | -a | 최초 전체 db 생성 |
| 2 | -b | 업데이트 추가 & 새로운 허가분에 대한 document 저장 |
| 3 | -c | 업데이트 변경일자 |
| 4 | -d | 업데이트 취소일자 |
| 5 | -e | 최초 전체 document 생성 |
| 6 | -f | 엑셀파일과 비교했을 때 document에 누락된 item\_seq를 찾아내서 api를 요청한 후 documnet 생성 |
| 7 | -g | db.json을 csv파일로 바꿔서 저장 |
| 8 | -i | chart 엑셀 파일 다운로드 |
| 9 | -j | chart excel 파일 csv로 바꿔서 저장 |

**세부 모듈 정리**

|  |  |
| --- | --- |
| 모듈명 | **설명** |
| download\_excel\_file | 의약품 정보 엑셀 파일을 다운로드 한다.  다운로드 위치 : https://nedrug.mfds.go.kr/pbp/CCBGA01/  getItem?totalPages=6&limit=10&page=2&&openDataInfoSeq=12 |
| load\_file | 의약품 정보 엑셀 파일 내용을 읽어서 dataframe형태로 반환한다. |
| get\_list\_from\_db | Dbxxx.json 파일을 저장 또는 업데이트 할 내용의 리스트(out\_json)를 반환한다. |
| make\_db\_json\_list | db 리스트(out\_json)를 생성하고 반환한다.  ex)  {"195500002": {"item\_seq": "195500002", "item\_name": "종근당염산에페드린정", "entp\_name": "(주)종근당", "chart": "본품은 백색의 정제다.", "item\_permit\_date": "19550117", "edi\_code": "", "permit\_kind\_name": "허가", "cancel\_name": "유효기간만료", "cancel\_date": "20200101", "etc\_otc\_code": "전문의약품", "main\_item\_ingr": "[M040420]염산에페드린", "atc": "R03CA02"}, ... } |
| get\_data\_dbjson | Dbxxx.json에 넣을 형태에 맞게 데이터를 생성한고 반환한다. |
| get\_file\_name | item\_seq를 받으면 해당 아이템이 속해야 하는 파일 이름을 반환한다.  ex) 195500002 -> 1955\_00001-00100 |
| save\_or\_update | 생성된 리스트(out\_json)를 통해 파일 저장 또는 업데이트 |
| get\_list\_from\_document | Document의 xml 파일로부터 documentxxx.json 파일을 저장할 내용의 리스트(out\_json)를 반환한다. |
| get\_content\_from\_document | document의 xml파일을 읽어 beautifulsoup형태로 반환 |
| get\_list\_from\_api | documentxxx.json에 저장되어야 할 아이템 중 document의 xml파일이 존재하지 않아 api로부터 xml 파일을 구해야 하는 경우이다. api로부터 누락된 아이템의 documentxxx.json에 저장될 리스트(out\_json)를 생성한다.  Api 정보 : https://www.data.go.kr/data/15057662/openapi.do |
| get\_item\_seq\_list\_from\_out\_json | Out\_json에서 item\_seq의 리스트를 반환한다. |
| push\_into\_out\_json | Out\_json에 데이터 집어넣어 out\_json을 반환한다. |
| xml\_to\_json | Beautifulsoup의 xml을 json형태로 바꾸어 반환한다. |
| get\_missing\_seq | Document파일의 아이템 중 빠진 아이템의 item\_seq리스트를 생성해 반환한다. |
| Json\_to\_pd | Json 파일을 dataframe object로 바꿔서 반환한다 |
| Write\_csv | 컬럼명 리스트, 데이터 리스트를 파라미터로 입력받으면 csv파일을 생성한다. |
| download\_chart\_excelfile\_by\_api | Chart엑셀 파일 다운로드 |
| Load\_chart | Chart 엑셀 파일을 dataframe으로 반환한다. |
| Chart\_data\_update | Chart의 데이터 가공한다. |
| Chart\_column\_update | Chart의 컬럼값을 영어로 바꿔주어 반환한다. |

|  |  |  |
| --- | --- | --- |
| **모듈명** | **파라미터** | **리턴값** |
| download\_excel\_file\_by\_api | 1.오늘날짜  2.엑셀파일 저장할 위치 | 없음 |
| load\_file | 1.로드할 파일의 날짜  2.엑셀파일 저장 위치 | 엑셀 파일을 dataframe으로 반환한 내용 |
| get\_list\_from\_db | 1.save/update  2.dataframe  3.업데이트 할 기준 시작날짜  4.업데이트 할 기준 종료날짜  5.허가일자/변경일자/취소일자 | 저장 또는 업데이트할 내용이 담긴 리스트(out\_json) |
| Make\_db\_json\_list | dataframe | out\_json |
| get\_data\_dbjson | 1.dataframe의 현재 아이템의 인덱스  2.dataframe | data |
| Get\_file\_name | item\_seq(아이템 시퀀스)  ex) 195500002 | 파일명  ex) 1955\_00001-00100 |
| Save\_or\_update | 1.out\_json  2.업데이트 기준  (허가일자/변경일자/취소일자)  3.파일을 저장할 위치 | 없음 |
| get\_list\_from\_document | document 파일이 있는 폴더 위치 | out\_json |
| get\_content\_from\_document | Document xml 파일이 있는 위치 | xml파일을 beautifulsoup 형태로 변환한 것 |
| get\_list\_from\_api | Document 내용에 추가해야할 item\_seq의 리스트 | out\_json |
| get\_item\_seq\_list\_from\_out\_json | Out\_json | Out\_json에 포함되어 있는 item\_seq\_list |
| push\_into\_out\_json | 1.해당 데이터가 속해 있는 파일명  2.out\_json  3.item\_seq  4.data | Out\_json |
| Xml\_to\_json | Beatifulsoup 형태의 document 내용 | data |
| Get\_missing\_seq | 1.dataframe  2.document 파일이 있는 폴더 위치 | Document 파일에 없는 item\_seq의 리스트 |
| Json\_to\_pd | Dataframe object로 바꿀 json 파일 위치 | dataframe |
| Write\_csv | 1.컬럼명 리스트  2.데이터 리스트  3.시작을 체크하는 변수(true) | 없음 |
| Download\_chat\_excelfile\_by\_api | 1.오늘 날짜  2.chart엑셀 파일을 저장할 위치 | 없음 |
| load\_chart | 1.오늘 날짜  2.chart 엑셀 파일 위치 | dataframe |
| chart\_data\_update | Dataframe | Dataframe |
| Chart\_column\_update | dataframe | datframe |