

Ask Company

부경대 WAP 파이썬 컨퍼런스 (2019년 5월)

파이썬으로 할 수 있는 모든 것들

여러분의 파이썬/장고 페이스메이커가 되겠습니다.

이진석

- Python Holic 개발자 : 버전 2.4 (2004년) ~
- Django Holic 개발자 : 버전 1.0 (2008년) ~
- Ask Company 개발/운영자
 - 국내 최대 파이썬/장고 Q&A 그룹 : Ask Django 운영 (회원 3,000여명)
 - 국내 유일 파이썬/장고 온라인 VOD 서비스 개발/운영 by 파이썬/장고
- Microsoft MVP Award 수상
 - Python / Azure 클라우드 분야 (2016년/2017년/2018년)
- 다수 기업/학교에 파이썬/장고 교육 및 컨설팅
 - LG전자, 삼성전자, SK텔링크, 이노션 월드와이드, 부산대학교 등 다수
- (전) 서울대학교 외래교수 (2016년/2017년)
 - 경영대학교, 벤처창업웹프로그래밍 (파이썬/장고 강의)



<https://askcompany.kr>
me@askcompany.kr

다양한 강의 이력 → <https://www.askcompany.kr/pages/about/>

Ask Company

Ask Django Q&A 페이스북 그룹

대전을 비롯한 다양한 지역에서 정기적인 유료/무료 세미나/강의

온라인 파이썬 관련 VOD 교육 서비스

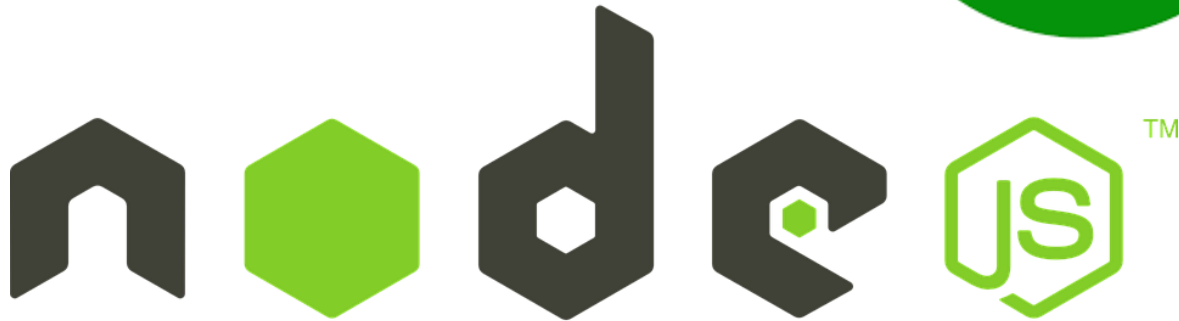
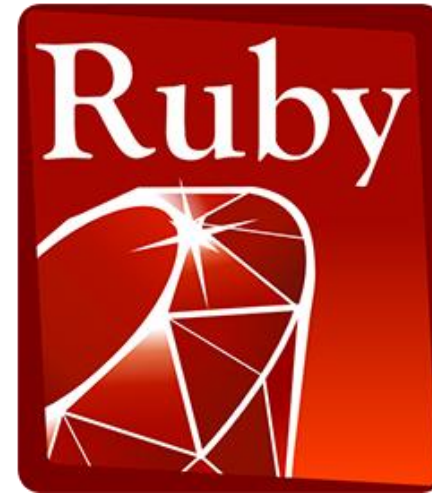
파이썬/장고 컨설팅

질문은 자유롭게 해주세요.

질문은 참는 것이 아닙니다. 😊



수많은 프로그래밍 언어



파이썬 유저층

전문
소프트웨어
개발자

데이터분석가

(Pandas/NumPy/Matplotlib 등)
→ EDA (탐색적 자료 분석) 툴

사무, 업무 자동화

(Pandas, 엑셀 연동, 웹개발 등)
→ 업무에 따라 다양한 시나리오
(데이터 수집/분석 자동화, 엑셀 파일
분석/통계 자동화 등)

머신러닝 개발자

(Tensorflow, Keras, PyTorch 등)

마케터

(크롤링, 애플리케이션 UI자동화,
데이터분석 통계 등)

코드 자동화가 필요한 모든 분들

(주식 자동화, 티켓 자동예매 등)

파이썬으로 할 수 있는 것들

데이터 분석 by Pandas 등

머신러닝/인공지능 by Tensorflow, keras, pytorch 등

Azure/AWS/Google 인공지능 API 활용 by requests 등

다양한 API 활용 by requests 등

엑셀 UI 자동화 by xlwings 등

엑셀 파일 자동화 by openpyxl, xlrd 등

웹서비스 개발, 챗봇 개발 by Django, flask 등

윈도우 애플리케이션 개발 by PyQt/PySide, IronPython, tkinter 등

공장 자동화 애플리케이션 개발 by ctypes, PyQt/PySide 등

그리고, 여러분이 **상상**할 수 있는 거의 모든 것들

필요한 프로그램이 있다면, 누군가가 만들어주기를 기다리기보다, 직접 만들어쓰실 수 있어요.

파이썬으로 할 수 없는 것들

Android 앱

iOS 앱

물론 시도하시는 고마우신 분들도 계십니다. → beeware.org



오늘 살펴볼 주제들

1. 파이썬 크롤링 Demo
2. 파이썬 데이터분석 Demo
3. 파이썬/엑셀 자동화 Demo
4. 파이썬/장고를 활용한 웹서비스 + Telegram봇 Demo

파이썬 크롤링 Demo

크롤링

- 사람이 웹페이지에 접속해서 정보를 읽어들이듯이
- 인터넷 상에 흩어져있는 자료들을 사람 대신에 프로그램을 통해 서핑하며 수집&가공
- 이때 프로그램 구성에 따라 서핑능력의 차이가 발생.
 - 대표적으로 자바스크립트 혹은 ActiveX 처리 여부

"Web scraping is a computer software technique of extracting information from websites."

Ask Company VOD에서의 크롤링 구분

초급편

requests/BeautifulSoup4 라이브러리 활용

- HTTP 프로토콜 이해
- 다양한 크롤링 방식에 대한 이해
- 크롤링한 데이터를 저장하는 다양한 방법
- 파일을 다운받고 저장하는 방법
- 텔레그램 봇에 크롤링 기능 붙이기
- Demo → 네이버 맞춤법 검사, 네이버 통합검색, 네이버 영화차트 5개, 멜론/벅스 차트, 네이버 카페글 검색/수집, 클리앙 알뜰구매 게시글/댓글 수집, 네이버 쇼핑 검색결과 수집, 네이버 쇼핑 특정 상품의 리뷰 및 평점 수집
- Pandas를 활용한 크롤링과의 비교
- 로컬 머신에서의 크롤링 스케줄링
(→ 고급편에서 클라우드에서의 스케줄링과 FaaS를 다룹니다.)
- 맛보기) requests와 selenium 간의 비교

중급편

파이썬
Selenium 활용

고급편

Scrapy와
클라우드를
활용한 크롤링
자동화 및
스케줄링

초고급편

ActiveX
사이트 크롤링

Pandas 기초 (데이터 분석 라이브러리)

DataFrame/Series 자료구조를 통해, 보다 편리한 데이터 처리

파이썬 크롤링 Demo

네이버 실시간 검색어

네이버 블로그 검색

네이버 영화 차트

멜론 차트

중고나라를 비롯한 네이버 카페글 검색 및 수집

클리앙 알뜰구매 게시판/댓글

네이버 쇼핑 상품 검색

네이버 쇼핑 상품 평점 및 리뷰 수집

법원 경매 정보

➔ 자동화의 영역으로 확대 가능

파이썬/엑셀 자동화 Demo

다양한 파이썬 엑셀 라이브러리

openpyxl : <https://openpyxl.readthedocs.io>

엑셀 2010 파일포맷 read/write 지원

pandas의 read_excel 함수에서 내부적으로 사용

pyexcel : <https://github.com/pyexcel/pyexcel>

csv, ods, xls,xlsx, xlsxm 등의 파일들에 대해 하나의 API로 접근 지원

xlsxwriter : <https://xlsxwriter.readthedocs.io/>

xlsx 포맷에 대한 생성을 풍성하게 지원

xlrd : <https://github.com/python-excel/xlrd>

xlwt : <https://github.com/python-excel/xlwt>

엑셀 97/2000/XP/2003 포맷에 대한 읽기/쓰기를 지원

xlwings : <https://www.xlwings.org/>

엑셀 프로그램 자동화 라이브러리

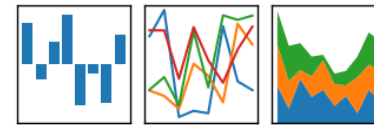
유일하게 엑셀 프로그램에 의존적

파이썬의 강력함/유연함으로
엑셀에 날개를 달아주세요.



pandas

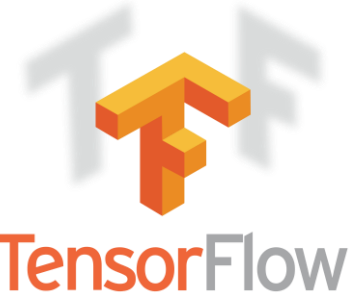
$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Notebook



Requests
http for humans



xlwings 라이브러리

엑셀 프로그램에 대한 자동화 라이브러리

엑셀의 익숙함과 파이썬의 강력함의 콜라보

다른 엑셀 라이브러리들은 엑셀 프로그램과의 연동이 아니라, 엑셀 파일 포맷을 지원하는 형태

엑셀을 띄워놓고, 파이썬을 통한 값 가져오기/변경 지원

엑셀의 매크로 기능을 파이썬으로 구현 지원

VBA 대체 가능

설치 : Anaconda Python에 포함

셸> pip install xlwings 혹은 conda install xlwings

윈도우/맥 지원

xlwings 지원 엑셀 파일 만들기

xlwings addin install

xlwings quickstart demo (모듈명으로 사용되기에 변수명 네이밍룰. 인코딩 이슈로 경로명에 한글 X)


cd demo

start demo.xlsm

필요한 엑셀 설정

옵션 → 보안 센터 → 보안 센터 설정 → 매크로 설정 → 개발자 매크로 설정 → VBA 프로젝트 개체 모델에 안전하게 액세스할 수 있음. 옵션 체크

xlwings.conf 활성화 → Interpreter 설정


자동 저장 ☐ 


demo.xlsm - 저장됨

파일홈삽입그리기페이지 레이아웃수식데이터검토보기도움말Team

xlwings

Interpreter:
PYTHONPATH:

 Import Functions

UDF Modules:
☒ Debug UDFs
 Restart UDF Server

Log File:
☐ RunPython: Use UDF Server
Version: 0.15.4

Python

User Defined Functions (UDFs)

Advanced

D16



	A	B	C	D	E	F	G	H
1	Interpreter	C:\Anaconda3\pythonw.exe						
2	PYTHONPATH							
3	UDF Modules							
4	Debug UDFs	FALSE						
5	Log File							
6	Use UDF Server	FALSE						
7								
8								
9	Instructions:							
10	Rename this sheet to "xlwings.conf" to activate it. When activated, these settings override the global settings in							
11	You can delete the rows you don't need. The settings have to start in cell A1.							
12	If you want to rely on the settings in the addin, then you can delete this sheet.							
13								

Sheet1

xlwings.conf



텔레그램 봇 Demo, 파이썬/장고를 활용한 웹서비스

파이썬을 활용한 텔레그램 봇 기본 코드

- 필요한 라이브러리 설치
 - 명령 프롬프트> `pip install python-telegram-bot requests beautifulsoup4 jinja2`
- BotFather를 통해 봇을 생성하고, Token 획득
 - 봇 생성 후에, `/setprivacy` 명령을 통해 Disable해주셔야, 그룹방에서 봇이 메시지를 받을 수 있습니다.
- 획득한 Token을 코드에 적용
 - 소스코드 편집기를 통해 수정
 - 소스코드 편집기가 없다면 Visual Studio Code 추천 : <https://code.visualstudio.com>
- 다음 명령으로 서버 실행
 - 명령 프롬프트> `python telegram_server_using_polling.py`

<https://gist.github.com/allieus/32471ceee2571b56e9f1fb7793b36968>

다양한 개선안

- 코드 모듈화
- 말을 좀 더 지능적으로 알아듣기
 - 정규표현식, 머신러닝, Google Dialogflow 서비스 API
- 텍스트 응답 외에 Video/Audio 응답
- 현재의 Polling 방식 코드를 Webhook 방식으로 변경
 - 웹 애플리케이션 개발 기술을 활용 → Django
 - 24시간 가용한 상태로 개발 가능하며, 보다 다양한 기능을 구현 가능
 - 배포는 Heroku를 사용하신다면, 무료 범위로 가능 → Azure/AWS/Google로도 저비용으로 운영 가능

웹프레임워크가 왜 필요한가요?

우선 웹서비스가 왜 필요한가요?

서버의 역할

모든 서비스의 근간 !!! 어떤 서비스이든 웹서비스는 당연히 잘 해야하는 분야. 덜 회자된다고 해서 지나간 유행이 아니다. 서버없이 머신러닝만 한다고 해서 서비스가 되겠는가??? 앱만 한다고 해서 서비스가 되겠는가???

만들 수 있는 것

웹서비스, 앱 서버, 챗봇 서비스 등등

웹서비스를 만들때마다 반복되는 것들을 표준화해서 묶어놓음.

거의 모든 언어마다 웹프레임워크가 존재.

다양한 파이썬 웹프레임워크

Django : The Web framework for perfectionists with deadlines.

백엔드 개발에 필요한 거의 모든 기능을 제공

Flask : a **micro** framework for Python based on Werkzeug.

백엔드 개발에 필요한 일부분의 기능을 제공

ORM으로서 SQLAlchemy를 주로 사용

Pyramid : a **small**, fast, down-to-earth

Bottle : a fast and **simple micro** framework for small web-applications.

Django

Lawrence Journal-World 신문사에서 2003년부터 개발, 2005년에 세상에 공개

2008년에 1.0 릴리즈 → 최신 2.2.X

기타리스트 Django Reinhardt 이름을 딴 작명

파이썬의 인기와 더불어 국내외에 장고를 쓰는 곳이 많습니다.

Instagram, Spotify, Disqus, Pinterest, Eventbrite, 요기요,뱅크샐러드, 하이퍼커넥트, 한국은행, On Demand Korea 등

django

Django의 강점

1. Python 생태계 !!!

여러분의 크롤링, 자동화, 머신러닝 코드와 같은 언어

2. 건강한 커뮤니티

3. 풀스택 웹프레임워크

4. 10년동안 충분히 성숙

Q&A : 파이썬에 대해 궁금하신 부분은???



인생은 짧습니다.
파이썬/장고를 쓰세요.

여러분의 파이썬/장고 페이스메이커가 되겠습니다.

많이 도와주세요. ;)