

학습 정리

팀	빅나물	구성원	이지현, 조영현
---	-----	-----	----------

일정	발제자	주제
3일차(5/29)	조영현	웹 크롤링을 위한 개발 환경 설정

주요 내용 요약

<웹 크롤링 기본 용어>

- **웹 스크래핑**
 - 실시간 개념 존재 가능
 - 필요로 하는 특정 정보 추출
 - 가공 및 저장해서 비교 분석 자료로 활용
 - 주로 CMS(Content Management System; 웹 상 정보 이용해서 사용자에게 제공하는 서비스) 어플리케이션의 원천 기술
 - 프로토콜, 브라우저 방식 존재
- **웹 크롤링**
 - 웹 스크래핑의 정식 명칭
 - 일정한, 주기적으로 웹 에서 정보 추출 및 저장
 - 주로 Crawler(크롤러)가 수행
 - 넓게 보면 스크래핑 범주에 포함
 - 최신 정보 유지(웹 크롤러) - 검색엔진
- **머신러닝**
인간의 학습 능력을 컴퓨터 또는 기계로 구현하는 것.
기본개념
(1) 기본적으로 알고리즘을 통해 데이터를 분석 및 학습
-> 컴퓨터로 학습 내용을 기반으로 판단, 추세, 예측
(2) 대량의 데이터와 알고리즘을 통해 학습시키는 것에 목표
(3) 문자, 음성, 얼굴, 지문, 게임, 의료, 로봇 등 다양한 분야에서 사용
- **딥러닝**
머신러닝에서는 학습에 있어 인간의 개입이 존재.
but, 딥러닝은 인간이 개입X. 신경망을 통해 스스로 학습 -> 그러므로 사용하는 데이터 표본의 양이 훨씬 많아야 함. 시스템 성능도 훨씬 좋아야 함.
GPU 연산을 하므로.(CPU X)
ex) 알파고 CPU 1202, GPU 176, 서버급 300대
인공지능>머신러닝>딥러닝
- **머신러닝 프로세스**
데이터 수집 -> 데이터 가공 -> 데이터 학습 -> 학습방법 선택 -> 매개변수 조정 -> 모델학습 -> 정확도 평가 -> 만족하다면 성공, 아니면 매개변수 조정 단계로 돌아감.

<개발환경 설정>

아나콘다

기본개념

- python 기반의 데이터 분석에 필요한 오픈소스를 모아놓은 개발 플랫폼
- 수준 높은 패키지 관리자를 통해 파이썬 효율성 극대화해서 사용. 필요한 모듈 관리
- 가상환경 관리자를 통한 프로젝트별 개발환경 구성. 프로젝트 관리
-> 충돌을 막을 수 있음.

기본 명령어

```
conda --version #현재 버전 확인
conda update conda #최신 버전 나오면 업데이트
conda --list #설치된 환경 목록(?)
conda info --envs #가상환경 목록
conda create --name test1 python=3.4 #가상환경 설치. 설치 시 python도 같이
옆에 적어 설치해주는 것이 좋음. 그 외 필요한 패키지 미리 옆에 나열하여 설치할
수 있음.
#ex)
conda create -n test1 python=3.4 simplejson pillow
# 풀네임인 경우 --, 약어의 경우 -

activate test1 #가상환경 활성화
conda list #가상환경에 기본적으로 설정된 목록이 나옴.
pip install simplejson #새로운 패키지 설치
conda list # simplejson 설치된 것을 볼 수 있음.

pip install --ignore-installed simplejson #기존에 simplejson이 설치되어있어도
무시하고 설치하겠다는 의미.
conda deactivate #가상환경 종료

conda remove --name test1 --all #가상환경 삭제

conda clean --all #아나콘다 느려짐 방지 위해 캐시 데이터 지우는 명령어
```

- atom
 - firt -> settings -> install
 - script, autocomplete-python 설치
 - packages에서 확인 가능
 - atom editor에서 파일 실행 단축키

ctrl+shift+b

참고 사이트

인프런

파이썬 입문 및 웹 크롤링을 활용한 다양한 자동화 어플리케이션 제작하기

<https://www.inflearn.com/course/python-%ED%8C%8C%EC%9D%B4%EC%8D%AC-%EC%9B%B9-%EB%8D%B0%EC%9D%B4%ED%84%B0-%ED%81%AC%EB%A1%A4%EB%A7%81/dashboard>