cafe.daum.net/gardenpro

<데이터 종류>

1. 정형 데이터: 항목이 뚜렷한 데이터

ex) 회원가입(이름, 주소, 전화번호) => Database(db) MS: oracle, mySQL

2. 반정형 데이터: html, xml, JSON

3. 비정형 데이터 (이미지, 동영상)

JSON

addr:구로구 벚꽃로

location: 30, 100

tel: 02-333-1234

request("www.naver.com") -> 네이버와 연동

xml문서 - 베타데이터

구로경찰서

<addr>구로구 벚꽃로</addr>

<location>30,100</location>

<tel>02-333-1234</tel>

<빅데이터 분석의 순서>

수집 => 전처리(다듬기..?) => 분석 => 시각화

1. 파일-공공데이터 (엑셀, csv) /
2. dbSQL(MySQL): SQL /
3. 크롤링(긁어오기=> 주로 xml 이나 json 형태로 수집 됨)

R프로그램 설치

RStudio 프로그램 설치

R과 파이썬을 프로그램을 부품조립식으로 함.

부품(기능을 가지고 있음)

테이블(회원정보) 테이블 형태여야 분석하기 좋음

이름 나이 연락처

김 100세 011

박 200 019

송 300 018

나이를 구하는 처리

입력값: 2000(year: 공간의 이름 🡺 변수)

나이를 계산하는 것은: 2019 – 2000 + 1

age = 2019 – year + 1 (=등호: 변수에 데이터를 넣는 역할, 대입, 할당)

CPU(계산)

RAM (저장공간) 내의 데이터를 구분하기 위해서

저장공간에 이름을 부여하여 구분

Year, age

부품 (라이브러리)

* R을 설치하면 기본적으로 라이브러리가 설치되어 있음.
* 추가할 때는 CRAN 저장소에서 다운로드 받아 설치

\*단축키: 컨트롤 + 시프트 + 1 (R 스크립트)

컨트롤 + 시프트 + 2 (R 콘솔)

컨트롤 + 시프트 + 4 (표준화면)

R 스크립트 실행(Run: 컨트롤 + Enter)

콘솔 클리어: 컨트롤 + L

< 데이터 >

가로: List 세로: Vector 🡺 데이터 프레임

리스트, 벡터는 변수들의 모음

<쇼핑몰>

회원관리 / 상품 / 게시판

MySQL – 회원, 상품, 게시판 => 특정한 회원이 특정한 상품 구매 후 게시판 작성

MyAge – yourAge => 자바 카멜 표기법

Mom\_age – dad\_age => 파이썬 1자 표기법

선생님 깃허브 -> bigdata201901

2000

숫자를 입력했을 때

기본값은 무조건 스트링(문자열)

* 숫자로 변환

공공데이터 포탈

* 데이터를 받는 방법 ( 2가지 )

1. 파일로 제공(엑셀, csv)
2. 검색어를 주면 찾아서 실시간으로 프로그램에 전달

(네트워크 프로그램 필요)

요청할 수 있는 키를 발급

Open API(함수, 방법)

Connect(<http://www.data.go.kr?key=>산”)

북한산 은평구 700 02-55-7894

Xml, json

Name: 북한산

Location: 은평구