AI융합비즈니스포럼

AI 융합전문가 4차시

데이터분석 I

2024.07.28..

현재 하는 일과 보유 자격증

소속	주요 업무
㈜제타데이터 대표이사	데이터 분석, 전략 및 컨설팅, 데이터 가치평가 ODA 컨설팅 (Official Development Assistance)
㈜지구파트너스 감사	창업보육, 투자, 기업·기술가치평가, 사업타당성 분석
㈜메타로직 컨설팅 수석	ISP 컨설팅 (Information Strategy Planning) ISMP 컨설팅 (Information System Master Plan)

◎ 자격증

1. 경영지도사31기 (인적자원, 2016)
2. 창업보육매니저 (BI협회, 2018)
3. 기업·기술가치평가사 (기업·기술가치평가협회, 2018)
4. 기업재난관리사 실무과정 (행정안전부, 2019)
5. 데이터분석 준전문가 ADsP (데이터산업진흥원 K-Data, 2021)
6. 빅데이터 분석기사 (과학기술정보통신부・통계청, 2021)
7. 국제공인컨설턴트 CMC (ICMCI, 2023)
8. 인공지능(AI) 활용마스터1급 (뉴미디어교육연구소, 2024)

데이터분석 관련 비즈니스

◎ 정보화전략계획수립(ISP) 컨설팅 수행

20.05~20.08. 창업진흥원
20.10~20.12. 한국연구재단
21.01~21.04. 소상공인시장진흥공단
21.11~22.06. 서울특별시
22.08~22.12. 경찰대학교
23.08~24.05 ODA (요르단 경찰청 PSD)

◎ 2022년 AI학습용데이터 구축사업 평가

1차 08 방송 콘텐츠 대화체 음성인식 데이터
09 방송 콘텐츠 한국어·영어 통번역 데이터
43 갑각류 종자생산 데이터
48 식생 탄소 포집량 식별 데이터
2차 74 축산 기자재(소, 돼지) 3D 데이터
75 소(한우, 젖소) 및 돼지 발정행동 데이터
3차 06 인공지능 신기술 선도(자유 공모)

◎ 데이터 가치평가 컨설팅

23.09~23.11 중소벤처기업진흥공단

발급번호:00KI-I83K-W6YQ-OA64-CG1Z

소프트웨어기술자 경력증명서

성명	홍용기			생년월일	1964.08.25
현 근무처	회사명			사업자등록번호	
	전화번호			업종	
	소재지				

근무경력	확인여부	근무기간	회사명	담당업무	부서/직위

기술자격	종목 및 등급	등록번호	취득일	발급기관
	빅데이터분석기사	BAE-002000023	2021.07.16	한국데이터산업진흥원
	ADsP(데이터 분석 준전문가)	ADsP-028000961	2021.04.09	한국데이터산업진흥원

학력	학교명	학과(전공)	수학기간	학위

교육	기간	과정	수료번호	교육기관

상훈	수여일	종류	상훈기관	근거

기술경력	확인여부	참여사업명	참여기간	발주자	소속사	직위	담당업무
	확인	국민 체감형 치안 안심 플랫폼 구축 정보화전략계획 사업	2022.08.04 ~ 2022.12.31	경찰대학교	(주)메타로직컨설팅	프리랜서	IT컨설팅 > 정보기술기획
	확인	서울시 차세대 지방세입정보시스템 통합구축 변화관리 컨설팅	2021.11.01 ~ 2022.06.30	서울특별시	(주)메타로직컨설팅	프리랜서	IT컨설팅 > 정보기술컨설팅
	확인	소상공인지원사업 디지털전환 정보화전략계획(ISP) 용역	2021.01.04 ~ 2021.05.03	소상공인시장진흥공단	(주)메타로직컨설팅	수석컨설턴트	IT컨설팅 > 정보기술컨설팅
	확인	한국연구재단 중장기 정보화전략계획(ISP) 수립	2020.10.05 ~ 2021.01.04	한국연구재단	(주)메타로직컨설팅	수석컨설턴트	IT컨설팅 > 정보기술컨설팅
	확인	창업기업확인시스템 구축을 위한 정보화전략계획(ISP) 수립	2020.05.18 ~ 2020.08.17	창업진흥원	(주)메타로직컨설팅	수석컨설턴트	IT컨설팅 > 정보기술컨설팅

「소프트웨어 진흥법」 제24조제3항 및 같은 법 시행규칙 제13조제3항에 따라 소프트웨어기술자의 경력 사항을 증명합니다.

2023년 01월 25일

강사 홍용기 컨설팅학 박사

데이터분석 관련 강의

- 데이터분석 및 실전 R코딩 (경영기술지도사회, 빅데이터 분석기사 자격증 취득 과정)
- 데이터분석 Python 심화과정 (서울 여성능력개발원 강동 여성인력개발센터 / 용산 여성인력개발센터)
- 파이썬 코딩을 통한 크롤링 자동화 인텐시브 과정 (경영지도사 및 컨설턴트)
- AI & ChatGPT 활용 및 데이터분석 컨설팅 방법론 (경영기술지도사회, 국제공인컨설턴트 CMC 양성과정)
- AI & 데이터분석 (매경아카데미, 동북아 ICT 포럼)

AI융합비즈니스포럼

강사 홍용기 컨설팅학 박사

책 쓰기 프로젝트

2018년 1월

2023년 5월

2024년 4월

2024년 6월

강사 홍용기 컨설팅학 박사

바로출판 POD 분야종합 - 교보

product.kyobobook.co.kr/pod/best?period=002&saleCmdtClstCode=33

교보문고 | eBook | sam | 핫트랙스 | 스토리 | 브랜드 더보기 | 회원가입 | 로그인 | 회원혜택 | 주문배송 | 매장안내 | 고객센터

KYOBO 교보문고

통합검색 | 140만 유튜버 핏블리 추천! 『근육연결도감』

리커버101 | 독서의이유 | 베스트 | 신상품 | 이벤트 | 사은품 | PICKS | CASTing | 교보ONLY | 할인혜택 | 출석체크 | 리커버:K 1~101

바로출판 POD

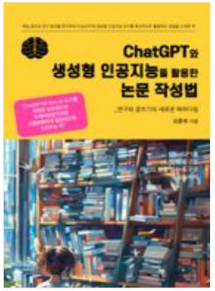
‹ POD 베스트

- 분야종합
- 소설
- 시/에세이
- 인문
- 가정/육아
- 요리
- 건강
- 취미/실용/스포츠
- 경제/경영
- 자기계발
- 정치/사회
- 역사/문화
- 종교
- 예술/대중문화
- 중/고등참고서
- 기술/공학
- 외국어

POD 베스트

2024년 06월 | 월간

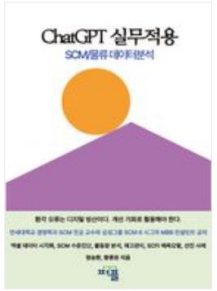
1
일반인을 위한 인공지능과 데이터 리터러시
홍용기 · 퍼플
15,000원



2
ChatGPT와 생성형 인공지능을 활용한 논문 작성법
신준석 · 퍼플
46,600원

3
개인정보 보호법 이해와 해설
강문석 · 퍼플
30,000원

4
음원 제작, 이건 알고 하자. 오디오 믹싱 & 마스터링
이대은 · 퍼플
26,900원



5
ChatGPT 실무적용 SCM/물류 데이터분석
정승환, 황종원 · 퍼플
35,000원

AD
초등 어린이 필독서!
이벤트
드래곤볼 GT 세트 30% 할인
쿠폰/혜택

What to do?

데이터 전문가 자격증

데이터분석 전문가 ADP (Advanced Data Professional)	필기 실기	국가 공인자격	K-Data
빅데이터 분석기사 BAE (Bigdata Analysis Engineer)	**필기 실기**	**국가 기술자격**	**과기정통부 & 통계청**
데이터분석 준전문가 ADsP (Advanced Data Semi Professional)	필기	국가 공인자격	K-Data


혹자는
인생 뭐 별거 있냐고 할테지만...
인생은 어쩌면
수많은 연결고리들의 집합일지도...
현재를 살면서 긴장을 늦출 수 없는 이유는
이 연결고리가 앞으로 어떻게 연결될지 지금은 알 수 없기 때문이다.
대체 불가능한 사람으로 살아남기 위해 해야 할 것

How to do?


Language 체득
R
데이터분석 특화
기계어
문법 단순
비전공자 권장
Python
공학베이스
인간 친화적
문법 어려움
전공자 권장

2024년 데이터분석 자격증 시험일정

➔ K-data 데이터자격검정(https://www.dataq.or.kr/www/accept/schedule.do)

구분	회차		접수기간	수험표발급	시험일	사전점수공개 및 재검토 접수	결과발표	증빙서류 제출기간
빅데이터 분석기사	제8회	필기	3.4~8	3.22	4.6(토)	4.19~23	4.26	4.29~5.9
		실기	5.20~24	6.7	6.22(토)	7.5~9	7.12	-
	제9회	필기	8.5~9	8.23	9.7(토)	9.20~24	9.27	9.30~10.10
		실기	10.28~11.1	11.15	11.30(토)	12.13~17	12.20	-
데이터분석 전문가 ADP	제32회	필기	1.22~26	2.8	2.24(토)	3.15~19	3.22	-
		실기	3.22~29	4.12	4.27(토)	5.17~21	5.24	5.24~31
	제33회	필기	7.1~5	7.26	8.10(토)	8.30~9.3	9.6	-
		실기	9.9~13	9.27	10.12(토)	11.1~5	11.8	11.8~15
데이터분석 준전문가 ADsP	제40회	-	1.22~26	2.8	2.24(토)	3.15~19	3.22	-
	제41회	-	4.8~12	4.26	5.11(토)	5.31~6.4	6.7	-
	제42회	-	7.1~5	7.26	8.10(토)	8.30~9.3	9.6	-
	제43회	-	9.30~10.4	10.18	11.3(일)	11.22~26	11.29	-

기술의 가치

"가치를 아는 사람에게 기술이 가야 빛을 발한다."

WWW

가족끼리 암호를 정해 두셨나요?

Like every great presentation, I've divided my talk into three subjects. Steve Jobs -

Ⅰ. Data & Information

Ⅱ. Database & Schema

Ⅲ. World Wide Web & Crawling

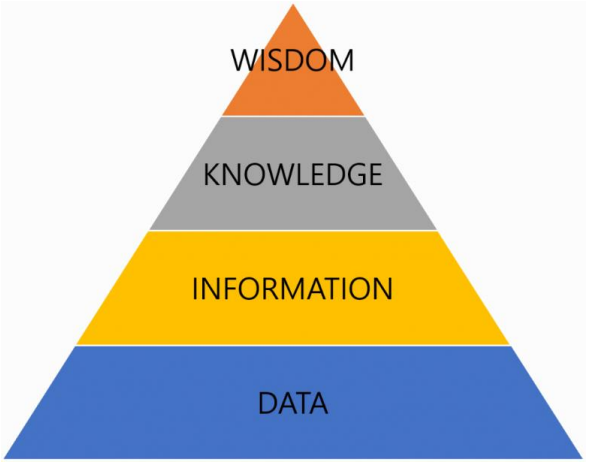
DATA

‘추론과 추정의 근거를 이루는 사실’

(A thing given or granted; something known or assumed as fact and made the basis of reasoning or calculation; an assumption of premiss from which inferences are drawn. OED, Vol. IV 264)

Data vs Information

피라미드 요소	설명
지혜 (wisdom)	● 근본 원리에 대한 깊은 이해를 바탕으로 도출되는 창의적 아이디어 ● 상황이나 맥락에 맞게 규칙을 적용하는 요소
지식 (knowledge)	● 다양한 정보를 구조화 하여 유의미한 정보로 분류하고 일반화 시킨 결과물 ● 정보에 기반해 찾아진 규칙
정보 (information)	● 여러가지 데이터 중에 사용자에게 필요한 데이터 ● 사용자의 필요에 따라 정제되거나 가공된 데이터를 정보라고 부름
데이터 (data)	● 객관적 사실로서 가공하기 전의 순수한 자료 ● 가공되지 않은 데이터는 그 자체로는 의미를 지니기 어려움



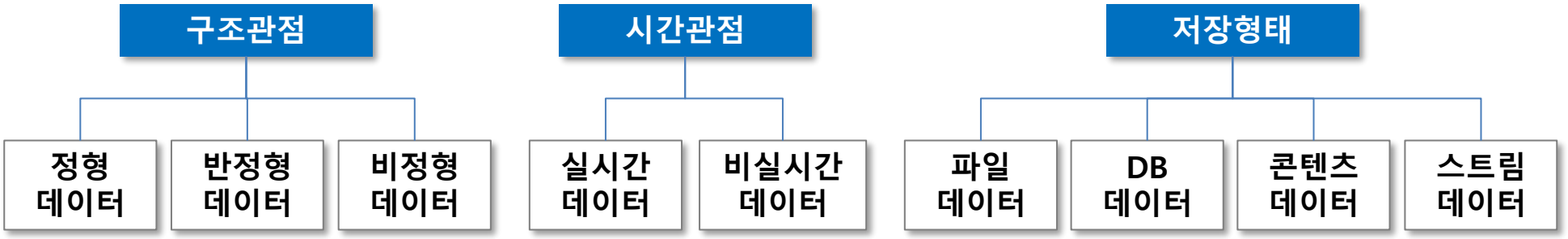
- ✓ **전국의 지역별로 조사한 매 시간별 기온 측정자료 ➔ 데이터**
- ✓ **어떤 목적에 의해 이 데이터를 기반으로 월별 또는 계절별 평균 기온을 산출 ➔ 정보**

ex. 어떤 지역의 온도에 따라 혹은 일교차에 따라 어떤 지역에서 어떤 과일을 재배하겠다는 등의 의사결정에 활용

‘데이터분석’이라고 하면…

어떠한 데이터를 어느 범위까지
분석해야 하는 것인가?

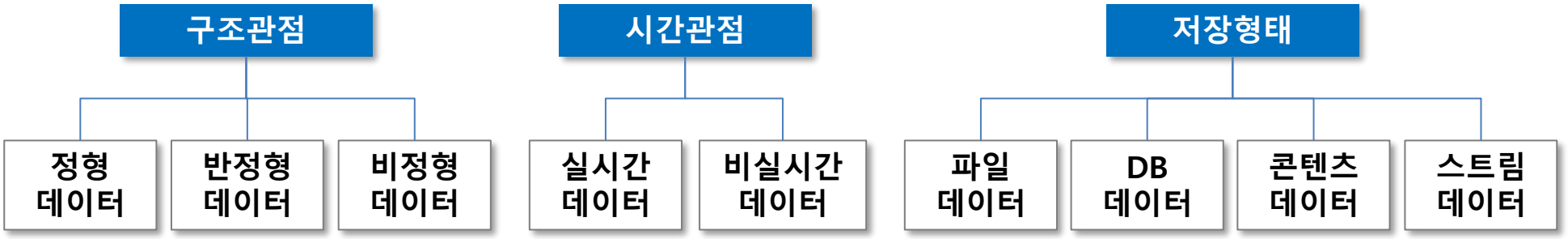
Type of Data



구조관점의 데이터 유형

유형	설명	종류
정형 데이터	정형화된 스키마(형태) 구조 기반으로 행과 열로 구성	관계형 데이터베이스, 스프레드 시트
반정형 데이터	스키마 구조 기반이나 값과 형식에서 일관성을 가지지 않는 데이터	XML, HTML, 로그데이터, JSON, 센서데이터
비정형 데이터	스키마 구조 형태를 가지지 않으며 고정된 필드에 저장되지 않는 데이터	SNS, 게시판, 텍스트, 이미지, 오디오/비디오

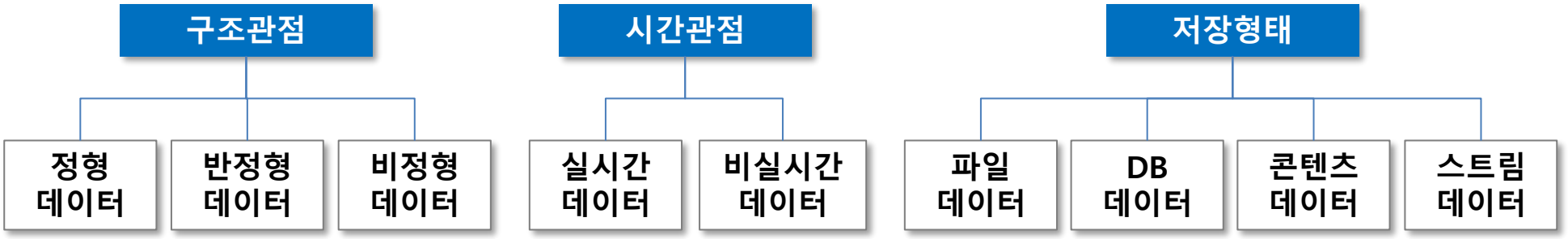
Type of Data



시간관점의 데이터 유형

유형	설명	종류
실시간 데이터	생성된 이후 수 초 ~ 수 분 이내에 처리되어야 의미가 있는 현재 데이터	센서 데이터, 시스템 로그, 보안 장비 로그, 알람, 네트워크 장비 로그
비실시간 데이터	이후에 처리되어야 의미가 있는 과거 데이터	통계, 웹 로그, 서비스 로그, 구매정보, 디지털 헬스케어 정보

Type of Data

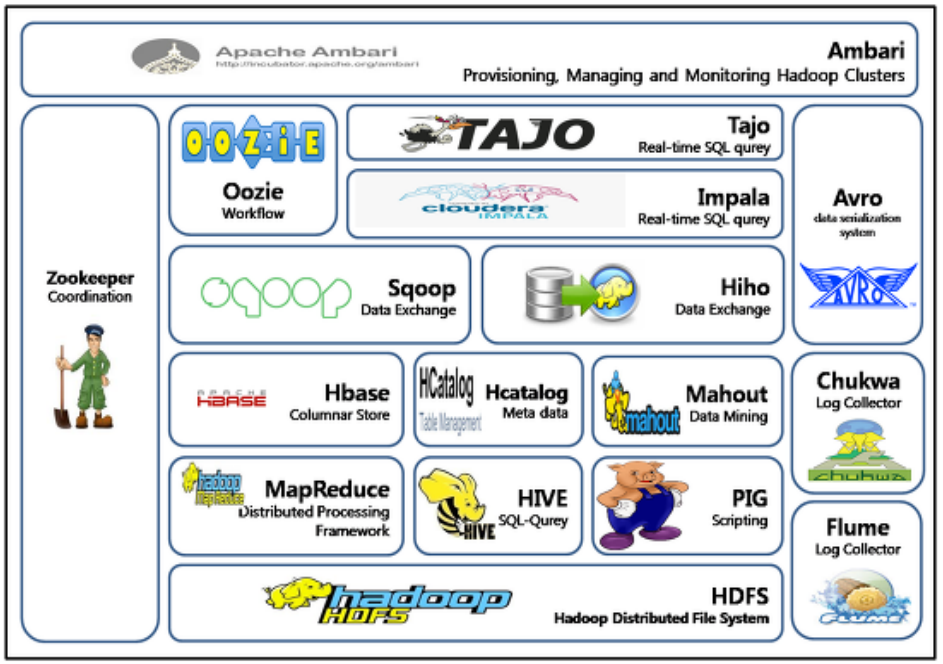


저장형태 관점의 데이터 유형

유형	설명
파일 데이터	시스템 로그, 텍스트, 스프레드 시트 등과 같이 파일 형식으로 파일 시스템에 저장되는 데이터
데이터베이스 데이터	관계형 데이터베이스, NoSQL, 인메모리 데이터베이스 등에 저장된 데이터
콘텐츠 데이터	텍스트, 이미지, 오디오, 비디오 등과 같이 개별적으로 데이터 객체로 구별 가능한 미디어 데이터
스트림 데이터	센서 데이터, HTTP 트랜잭션, 알람 등과 같이 네트워크를 통해 실시간으로 전송되는 데이터

Data Collection Methods and Technologies

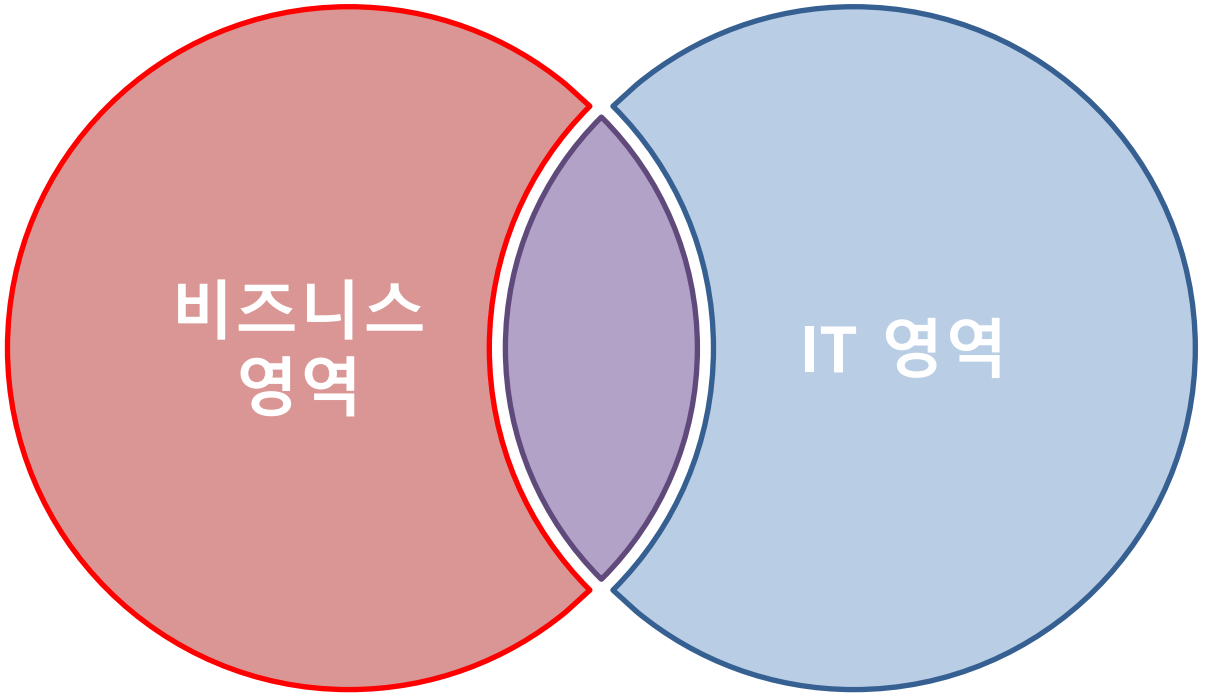
▪ ETL(Extract, Transform, Load) ▪ FTP(File Transfer Protocol) ▪ 스쿱(Sqoop) ▪ 스크래파이(Scrapy)	▪ 아파치 카프카(Apache Kafka) ▪ 플럼(Flume) ▪ 스크라이브(Scribe) ▪ 척와(Chukwa)	▪ CEP(Complex Event Processing) ▪ EAI(Enterprise Application Integration) ▪ CDC(Change Data Capture) ▪ ODS(Operational Data Store	▪ 크롤링(Crawling) ▪ RSS(Rich Site Summary) ▪ Open API ▪ 스트리밍(Streaming)



※출처: 하둡 에코시스템(Hadoop-Ecosystem)이란 https://butter-shower.tistory.com/73

Data Analyst Competency

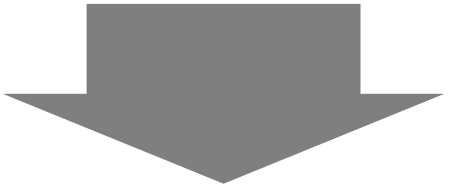
비즈니스
전문가



IT 전공자,
전문가

“강사/컨설턴트는 어느 영역까지 cover해야 할 것인가?”

강의/컨설팅 차원에서의 데이터 분석



데이터를 통한 문제 해결 vs 문제 해결을 위한 데이터

“업체가 보유한 데이터로도 충분하다.”

Data Analysis Framework

- 수십 기가바이트에 이르는 대용량 데이터 기반
- 데이터 마이닝 수행에 필요한 데이터 선별하여 샘플링 가능

- 데이터 분할 및 전처리
- 일관성이 없고 불완전한 데이터 정리
- 데이터를 분석하기 좋은 형태로 가공

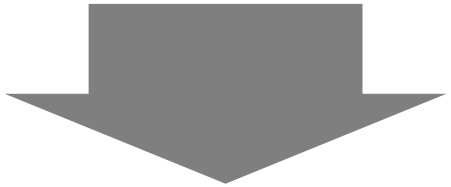
- 데이터 수치화 및 변수 간의 관계 관찰
- 시각화 활용

- 분석 모형을 활용한 모델 구축
- 모델의 적합성 평가

- 분석결과 보고
- 시각화 자료 첨부

“데이터분석 ➔ 데이터수집 + 실질적인 데이터 분석”

데이터분석 = 데이터수집 + 실질적인 데이터 분석



"어느 것이 더 어려운가?"

어느 것이 더 많은 노력과 시간이 들어가나?

Like every great presentation, I've divided my talk into three subjects. Steve Jobs -

I .

Data & Information

II .

Database & Schema

III .

World Wide Web & Crawling

일반적으로 엑셀 프로그램을 사용하는 방식

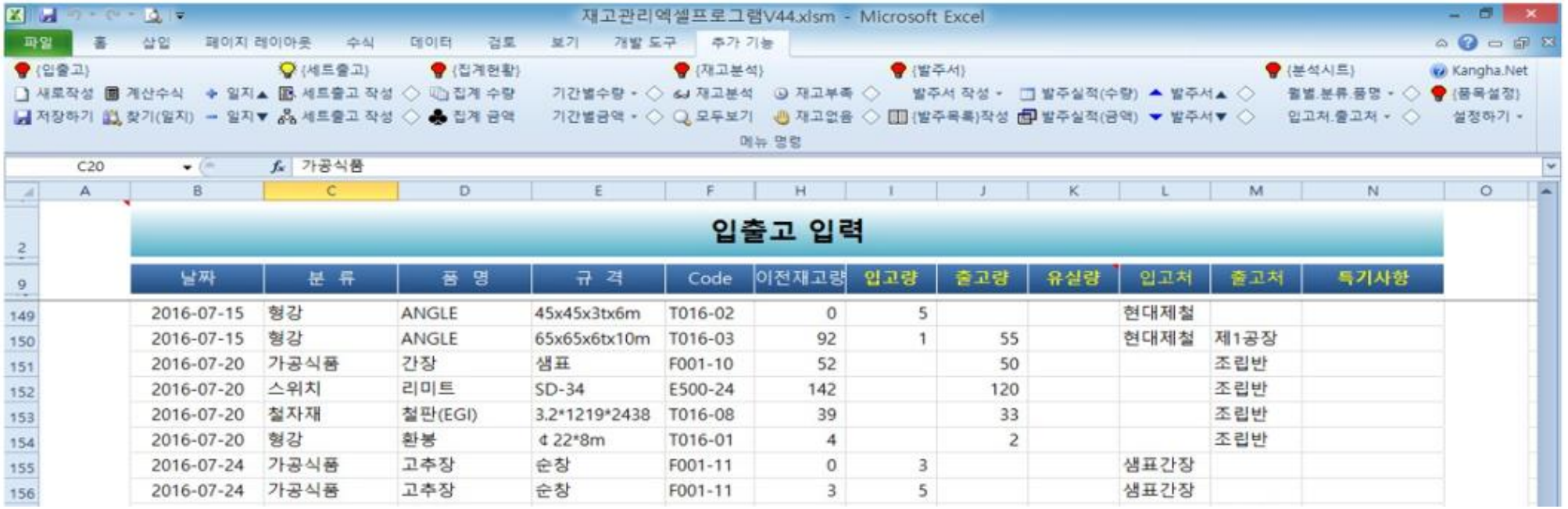
재고관리엑셀프로그램V44.xlsm - Microsoft Excel

입출고 입력

날짜	분 류	품 명	규 격	Code	이전재고량	입고량	출고량	유실량	입고처	출고처	특기사항
2016-07-15	형강	ANGLE	45x45x3tx6m	T016-02	0	5			현대제철		
2016-07-15	형강	ANGLE	65x65x6tx10m	T016-03	92	1	55		현대제철	제1공장	
2016-07-20	가공식품	간장	샘표	F001-10	52		50			조립반	
2016-07-20	스위치	리미트	SD-34	E500-24	142		120			조립반	
2016-07-20	철자재	철판(EGI)	3.2*1219*2438	T016-08	39		33			조립반	
2016-07-20	형강	환봉	¢ 22*8m	T016-01	4		2			조립반	
2016-07-24	가공식품	고추장	순창	F001-11	0	3			샘표간장		
2016-07-24	가공식품	고추장	순창	F001-11	3	5			샘표간장		
2016-07-24	가공식품	고추장	순창	F001-11	8		5			제2공장	
2016-07-24	센서	말굽	MMG-RE	E300-04	2		3			제2공장	
2016-07-24	식품	김치	담근김치	F222-01	13	8			청정원		
2016-07-31	스위치	푸쉬버튼	SD-37	E500-27	79		5			금형반	
2016-07-31	철자재	파이프	20*40*1.4*8m	S103-01	38	2			포스코		
2016-07-31	철자재	파이프	30*30*1.4*6m	S103-02	43		11			제1공장	
2016-08-26	센서	온도	DHS-23	E300-05	34	5			포스코		
2016-08-28	가공식품	간장	샘표	F001-10	2		1			제3공장	
2016-08-29	기타	페인트	암적색	T001-01	76		3			제3공장	

출처: https://www.kangha.net/bbs/board.php?bo_table=excel01&wr_id=3750&sca=%EC%9E%AC%EA%B3%A0%EA%B4%80%EB%A6%AC

Primary key, Foreign key



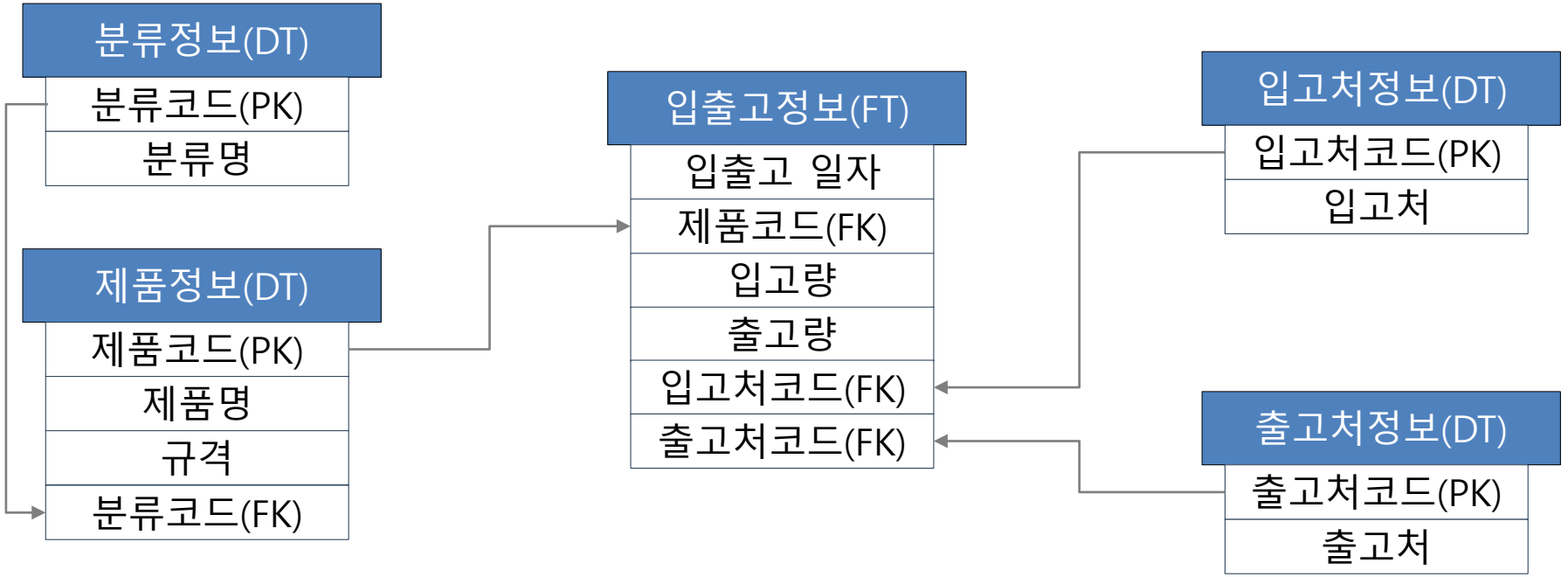
날짜	분 류	품 명	규 격	Code	이전재고량	입고량	출고량	유실량	입고처	출고처	특기사항
2016-07-15	형강	ANGLE	45x45x3tx6m	T016-02	0	5			현대제철		
2016-07-15	형강	ANGLE	65x65x6tx10m	T016-03	92	1	55		현대제철	제1공장	
2016-07-20	가공식품	간장	샘표	F001-10	52		50			조립반	
2016-07-20	스위치	리미트	SD-34	E500-24	142		120			조립반	
2016-07-20	철자재	철판(EGI)	3.2*1219*2438	T016-08	39		33			조립반	
2016-07-20	형강	환봉	¢22*8m	T016-01	4		2			조립반	
2016-07-24	가공식품	고추장	순창	F001-11	0	3			샘표간장		
2016-07-24	가공식품	고추장	순창	F001-11	3	5			샘표간장		

기본 키(**Primary key**)는 특정 행을 유일하게 식별할 수 있는 속성 (ex. 주민등록번호)

- ✓ Unique한 값으로 중복이 허용되지 않음 ➔ 테이블 생성 시 반드시 하나 이상의 필드/열을 지정
- ✓ Null값 포함이 불가 (Non-null)
- ✓ 가장 세분성(Granulairty)이 높아야 함

데이터 모델링

Dimension 테이블과 Fact 테이블은
특정한 기본키(primary key)와 외래키(foreign key)를 기준으로 상호 연결되며
DT:FT는 일대다(1:n)의 관계(cardinality, 선택도)를 갖게 됨



데이터 모델링

분류코드	분류명
1	형강
2	가공식품
3	스위치
4	철자재
5	센서
6	식품
7	기타

제품코드	제품코드2	제품명	규격	분류코드
1	T016-02	ANGLE	45	1
2	T016-03	ANGLE	65	1
3	F001-10	간장		2
4	E500-24	리미트		3
5	T016-08	철판		4
6	T016-01	환봉		1
7	F001-11	고추장		2
8	E300-04	말굽		5
9	F222-01	김치		6
10	E500-27	푸쉬버튼		3
11	S103-01	파이프	20	4
12	S103-02	파이프	40	4
13	E300-05	온도		5
14	T001-01	페인트		7

입출고일자	제품코드	입고량	출고량	입고처코드	출고처코드
2024-07-15	1	5		1	
2024-07-16	2	1		1	
2024-07-16	2		55		1
2024-07-17	3		50		2
2024-07-18	4		120		2
2024-07-19	5		33		2
2024-07-20	6		2		2
2024-07-21	7	3		2	
2024-07-22	7	5		2	
2024-07-23	7		5		3
2024-07-24	8		3		3
2024-07-25	9	8		3	
2024-07-26	10		5		4
2024-07-27	11	2		4	
2024-07-28	12		11		1
2024-07-29	13	5		4	
2024-07-30	3		1		5
2024-07-31	14		3		5

입고처코드	입고처
1	현대제철
2	샘표간장
3	청정원
4	포스코

출고처코드	출고처
1	제1공장
2	조립반
3	제2공장
4	금형반
5	제3공장

※ 참고: Star 스키마와 Snowflake 스키마 https://www.integrate.io/ko/blog/snowflake-schemas-vs-star-schemas-what-are-they-and-how-are-they-different-ko/

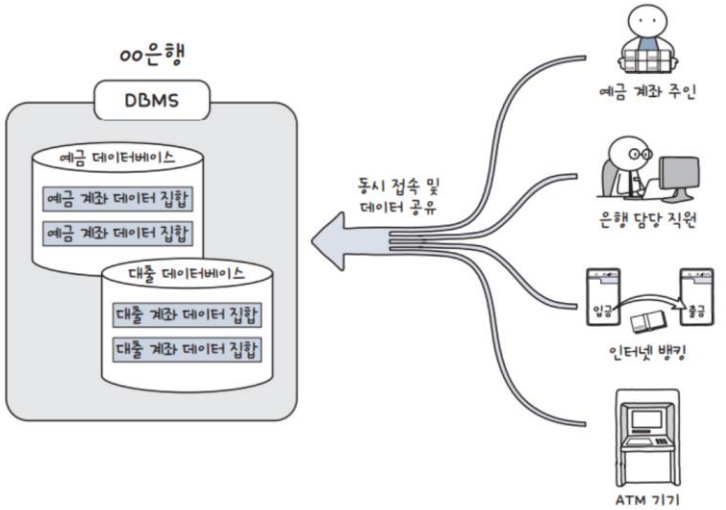
데이터베이스

- 데이터베이스(database, DB): 여러 사람이 공유하여 사용할 목적으로 체계화해 통합, 관리하는 데이터의 집합이다. (위키백과)

- DBMS(Database Management System): 데이터베이스를 운용하고 관리하는 소프트웨어(예: MySQL, Oracle, SQL 서버, PostgreSQL, MariaDB)

- SQL(Structured Query Language): 구조화된 질의 언어라는 뜻으로 관계형 데이터베이스에서 사용되는 언어

- 데이터베이스는 1개 이상의 테이블로 이루어지며, 테이블은 fact 테이블과 dimension 테이블로 구분됨



※그림 출처: 혼공러들의 스터디 공간 https://hongonghanbitcoky%EB%8D%B0%EC%9D%B4%ED%84%B0%EB%B2%A0%EC%9D%B4%EC%8A%A4%EC%9D%B4%ED%95%B4%ED%95%98%EA%B8%B0-databaseob-dbms-sql%EC%9D%98-%EA%B0%9C%EB%85%90/

데이터베이스

```
Microsoft Windows [Version 10.0.22631.3155]
(c) Microsoft Corporation. All rights reserved.

C:\Users\821033669010>cd "C:\program files\MariaDB 10.8\bin

C:\Program Files\MariaDB 10.8\bin>mysql.exe -u root -p
Enter password: *******
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MariaDB connection id is 6
Server version: 10.8.5-MariaDB mariadb.org binary distribution

Copyright (c) 2000, 2018, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input
statement.

MariaDB [(none)]> SHOW databases;
+--------------------+
| Database           |
+--------------------+
| financial_terms    |
| information_schema |
| investar           |
| kotra              |
| mysql              |
| opentutorials      |
| performance_schema |
| scraping           |
| sys                |
| wikipedia          |
| workbench          |
+--------------------+
11 rows in set (0.002 sec)
```

```
MariaDB [(none)]> USE opentutorials;
Database changed
MariaDB [opentutorials]> SHOW tables;
+------------------------+
| Tables_in_opentutorials |
+------------------------+
| author                 |
| topic                  |
| topic_backup           |
+------------------------+
3 rows in set (0.010 sec)

MariaDB [opentutorials]> DESC topic;
+-------------+-------------+------+-----+---------+----------------+
| Field       | Type        | Null | Key | Default | Extra          |
+-------------+-------------+------+-----+---------+----------------+
| id          | int(11)     | NO   | PRI | NULL    | auto_increment |
| title       | varchar(100) | NO   |     | NULL    |                |
| description | text        | YES  |     | NULL    |                |
| created     | datetime    | NO   |     | NULL    |                |
| author_id   | int(11)     | NO   |     | NULL    |                |
+-------------+-------------+------+-----+---------+----------------+
5 rows in set (0.018 sec)

MariaDB [opentutorials]> SELECT * FROM topic;
+----+---------------+-----------------------+---------------------+-----------+
| id | title         | description           | created             | author_id |
+----+---------------+-----------------------+---------------------+-----------+
|  1 | MySQL Database | MySQL is Relational ... | 2023-12-19 10:03:32 |         1 |
|  2 | Oracle        | Oracle is ...         | 2023-12-19 10:09:53 |         1 |
|  5 | MongoDB       | MongoDB is ...        | 2023-12-19 11:29:44 |         1 |
+----+---------------+-----------------------+---------------------+-----------+
3 rows in set (0.011 sec)
```

Like every great presentation, I've divided my talk into three subjects. Steve Jobs -

I .

Data & Information

II .

Database & Schema

III .

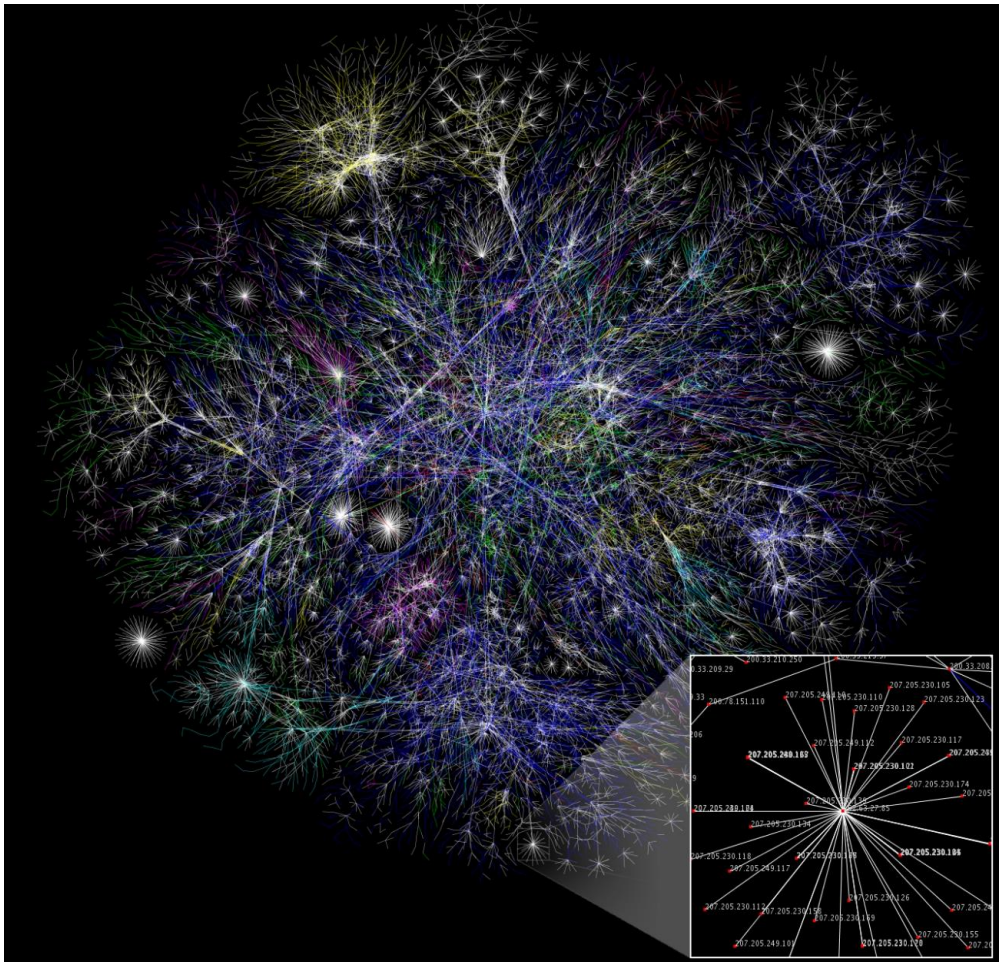
World Wide Web & Crawling

Internet

인터넷은 인간이 발명해 놓고도 이해하지 못하는 최초의 발명품이며, 역사상 최대 규모의 무정부주의에 대한 실험이다.

The Internet is the first thing that humanity has built that humanity doesn't understand, the largest experiment in anarchy that we have ever had.

- Eric Emerson Schmidt



※라우터를 통해 연결된 인터넷을 시각화한 그림(위키백과)

Internet

인터넷(Internet)은
인터넷 프로토콜 스위트(TCP/IP)를 기반으로 하여 전 세계적으로
연결되어 있는 컴퓨터 네트워크 통신망을 일컫는 말이다.
그야말로 인류의 역사상 전례 없는 거대한 정보의 바다인 셈이다.

흔히 웹(WEB)이라고 줄여 부르는
월드 와이드 웹(World Wide Web; WWW)만 생각하기 쉽지만
인터넷은 월드 와이드 웹, 전자 메일, 파일 공유(토렌트, eMule 등),
웹캠, 동영상 스트리밍, 온라인 게임, VoIP, 모바일 앱 등
다양한 서비스들을 포함한다.

※출처: 나무위키(https://namu.wiki/인터넷)

WWW의 탄생

1989년 3월, CERN(유럽 입자 물리 연구소)의
소프트웨어 공학자 팀 버너스리는 CERN에서 인사 재배치 등으로
기존에 수행했던 실험 결과를 비롯한 각종 문서들이 유실되는 비율이
높은 것을 보고 이를 줄이기 위해
Information System: A Proposal을 제안하였다.

또한 여러 연구기관에 흩어져 있는 문서들을 체계화하여
전 세계의 대학 및 연구소들끼리 정보를 신속하게 교환할 수 있도록
해야 한다고 판단하여 문서 뿐만 아니라 소리, 동영상 등을 망라하는
데이터베이스를 구축하고 이를 전문 열람 소프트웨어로 열람하는
방식을 생각해 냈다.

인터넷과 WWW

워낙 WWW가 대세이기에 WWW를 인터넷으로 착각하는 경우가 많지만, 웹은 TCP/IP 기반 물리적 통신망인 인터넷을 활용한 서비스로 인터넷의 하위 개념으로 볼 수 있다.

위키백과에 따르면 WWW은 다음 세 가지의 기능으로 요약할 수 있음

첫 번째, 통일된 웹 자원의 위치 지정 방법 ➔ 예를 들면 URL(Uniform Resource Locator)

두 번째, 웹의 자원 이름에 접근하는 프로토콜(protocol) ➔ 예를 들면 HTTP

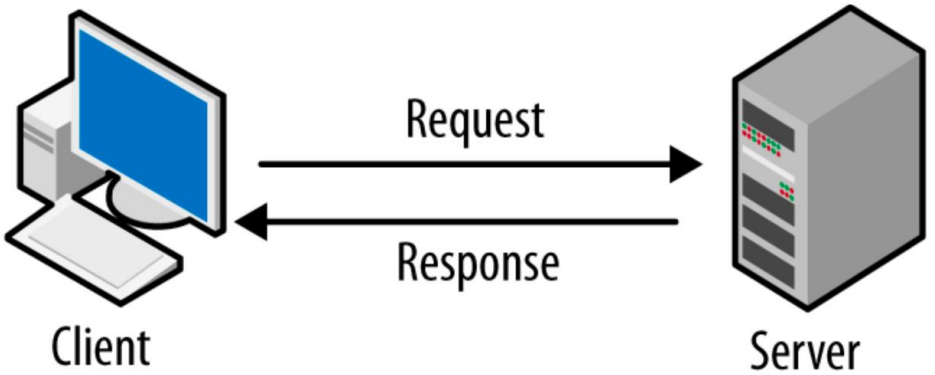
세 번째, 자원들 사이를 쉽게 항해할 수 있는 언어 ➔ 예를 들면 HTML

※ 세계 최초의 웹사이트: https://info.cern.ch/

HTTP

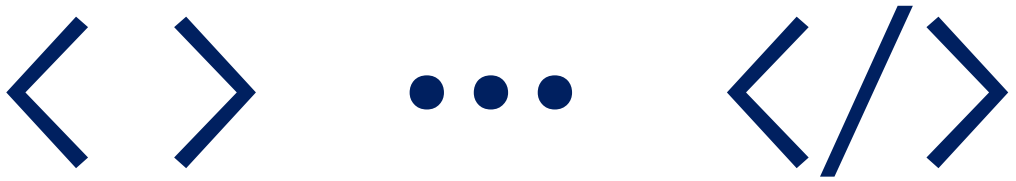
Hyper Text Transfer Protocol은 W3 상에서 정보를 주고받을 수 있는 프로토콜(규약)

- HTTP는 클라이언트와 서버 사이에 이루어지는 요청/응답 (request/response) 프로토콜(규약)
- 클라이언트인 웹 브라우저가 HTTP를 통하여 서버로부터 웹페이지(HTML)나 그림 정보를 요청하면, 서버는 이 요청에 응답하여 필요한 정보를 해당 사용자에게 전달
- 이 정보가 모니터와 같은 출력 장치를 통해 사용자에게 나타나는 것서 흔히 볼 수 있는 htm이나 html 확장자가 바로 이 언어로 작성된 문서



※그림 출처: https://velog.io/@seosu2000/Client-Server란 무엇인가

HTML



https://namu.wiki/w/HTML

HTML

웹사이트의 모습을 기술하기 위한 마크업 언어

- 프로그래밍 언어가 아니라 마크업 정보를 표현하는 마크업 언어로 문서의 내용 이외의 문서의 구조나 서식 같은 것을 포함
- HTML의 ML이 마크업 언어라는 뜻으로 웹사이트에서 흔히 볼 수 있는 htm이나 html 확장자가 바로 이 언어로 작성된 문서

```
<!DOCTYPE html>
<html>
    <head>
        <meta charset="utf-8">
    </head>
    <body>
        Hello, world!
    </body>
</html>
```

HTML은 '정보전달' 이 주목적

디자인 요소

UI

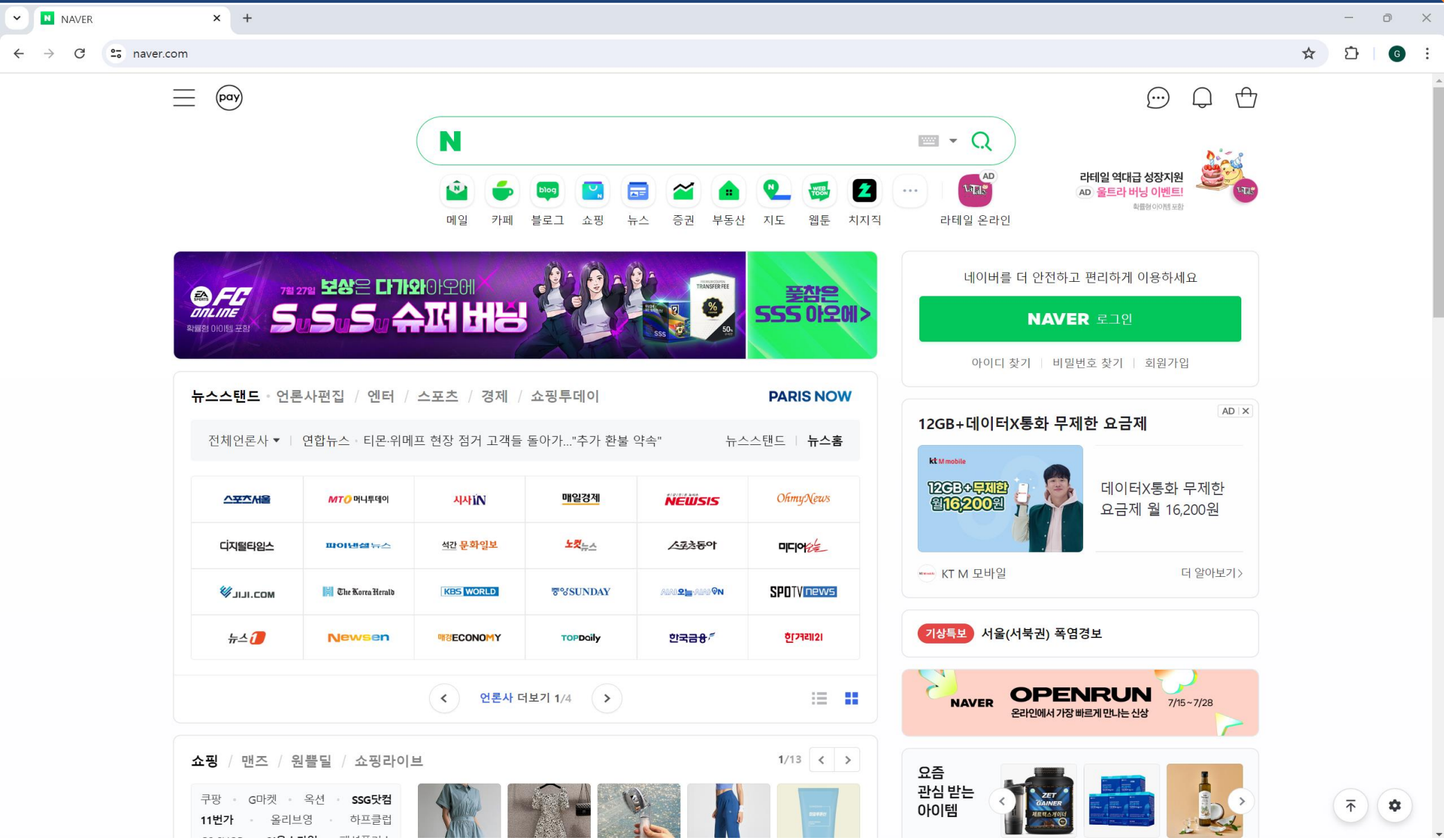
HTML은 '정보전달'에 충실

디자인 요소

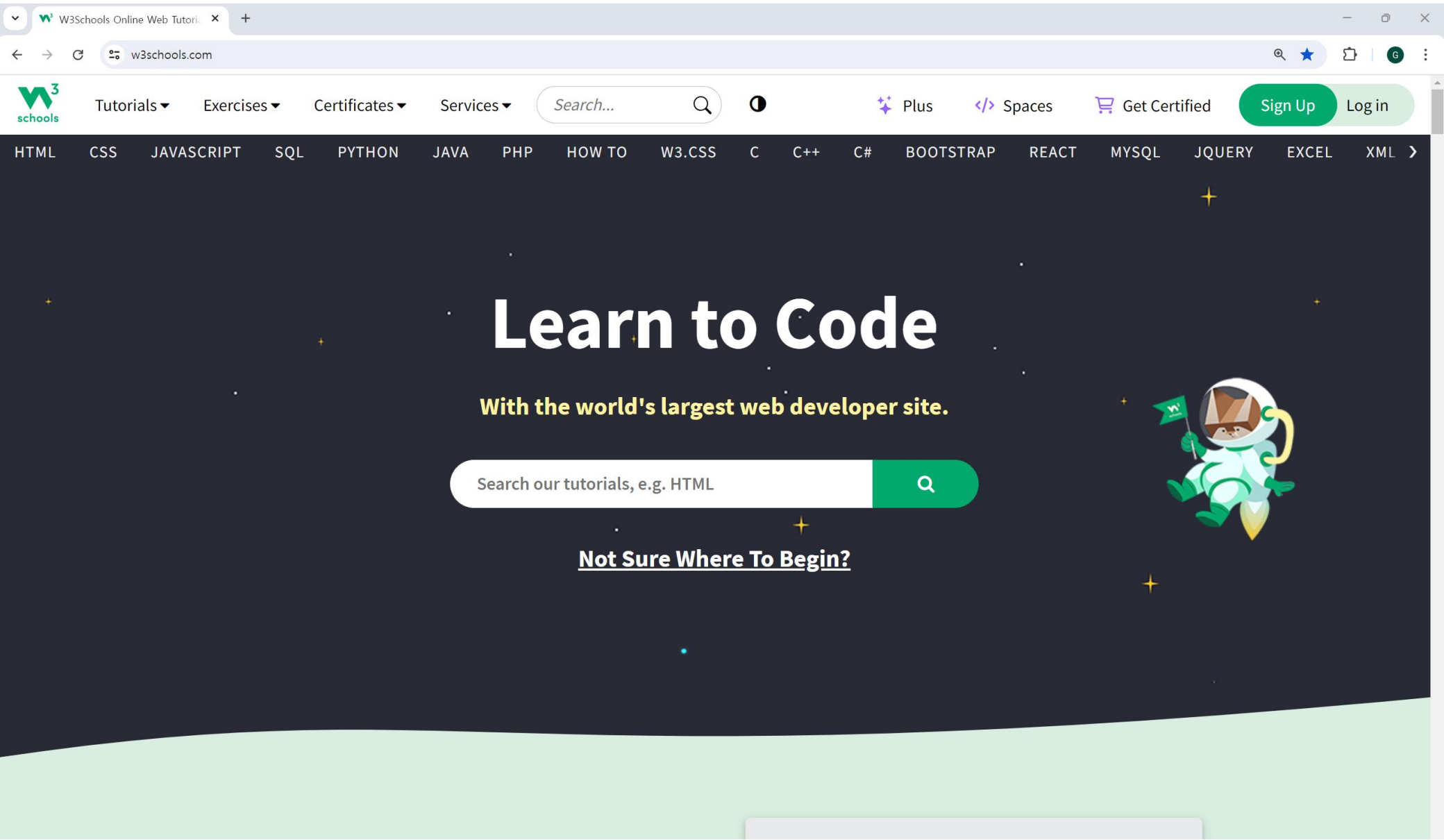
CSS (Cascading Style Sheet)

UI

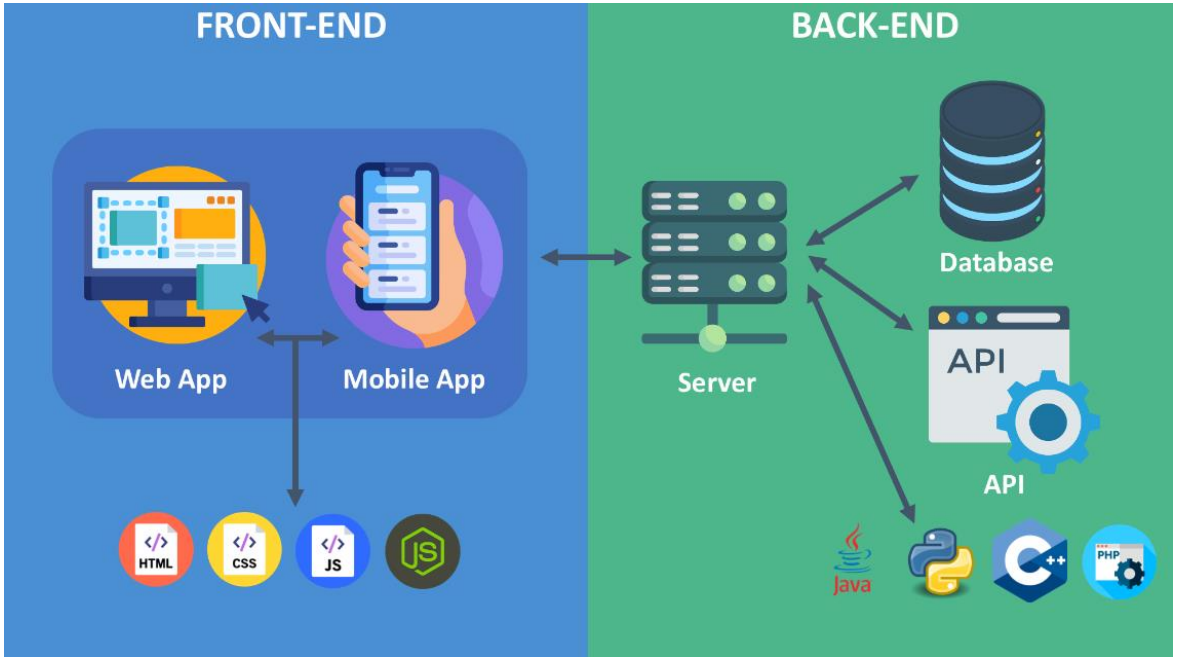
JavaScript


NAVER
naver.com
pay
메일
카페
블로그
쇼핑
뉴스
증권
부동산
지도
웹툰
치지직
라테일 온라인
라테일 역대급 성장지원
울트라 버닝 이벤트!
7월 27일 보상은 다가와 아오에
SSS 슈퍼 버닝
풀참은 SSS 아오에>
네이버를 더 안전하고 편리하게 이용하세요
NAVER 로그인
아이디 찾기 | 비밀번호 찾기 | 회원가입
뉴스스탠드 · 언론사편집 / 엔터 / 스포츠 / 경제 / 쇼핑투데이
PARIS NOW
전체언론사 | 연합뉴스 · 티몬·위메프 현장 점거 고객들 돌아가..."추가 환불 약속"
뉴스스탠드 | 뉴스홈
언론사 더보기 1/4
12GB+데이터X통화 무제한 요금제
데이터X통화 무제한 요금제 월 16,200원
KT M 모바일
더 알아보기>
기상특보 서울(서북권) 폭염경보
NAVER OPENRUN
온라인에서 가장 빠르게 만나는 신상
7/15~7/28
쇼핑 / 맨즈 / 원쁠딜 / 쇼핑라이브
1/13
쿠팡 · G마켓 · 옥션 · SSG닷컴
11번가 · 올리브영 · 하프클럽
요즘 관심 받는 아이템

WWW


W3Schools Online Web Tutori
w3schools.com
schools
Tutorials
Exercises
Certificates
Services
Search...
Plus
Spaces
Get Certified
Sign Up
Log in
HTML CSS JAVASCRIPT SQL PYTHON JAVA PHP HOW TO W3.CSS C C++ C# BOOTSTRAP REACT MYSQL JQUERY EXCEL XML
Learn to Code
With the world's largest web developer site.
Search our tutorials, e.g. HTML
Not Sure Where To Begin?

웹개발은 크게 기본, 프론트엔드, 백엔드 등 3가지 영역으로 구분



기본

HTML CSS JavaScript Git/GitHub

프론트엔드

라이브러리
jQuery, D3.js, Bootstrap 등

프레임워크
React, Angular, Vue 등

백엔드

서버언어
Python, Java, PHP 등

프레임워크
Node.js, Spring, Django 등

※그림 출처: https://velog.io/@xenxxn/01, https://1.pearlvely.com/5

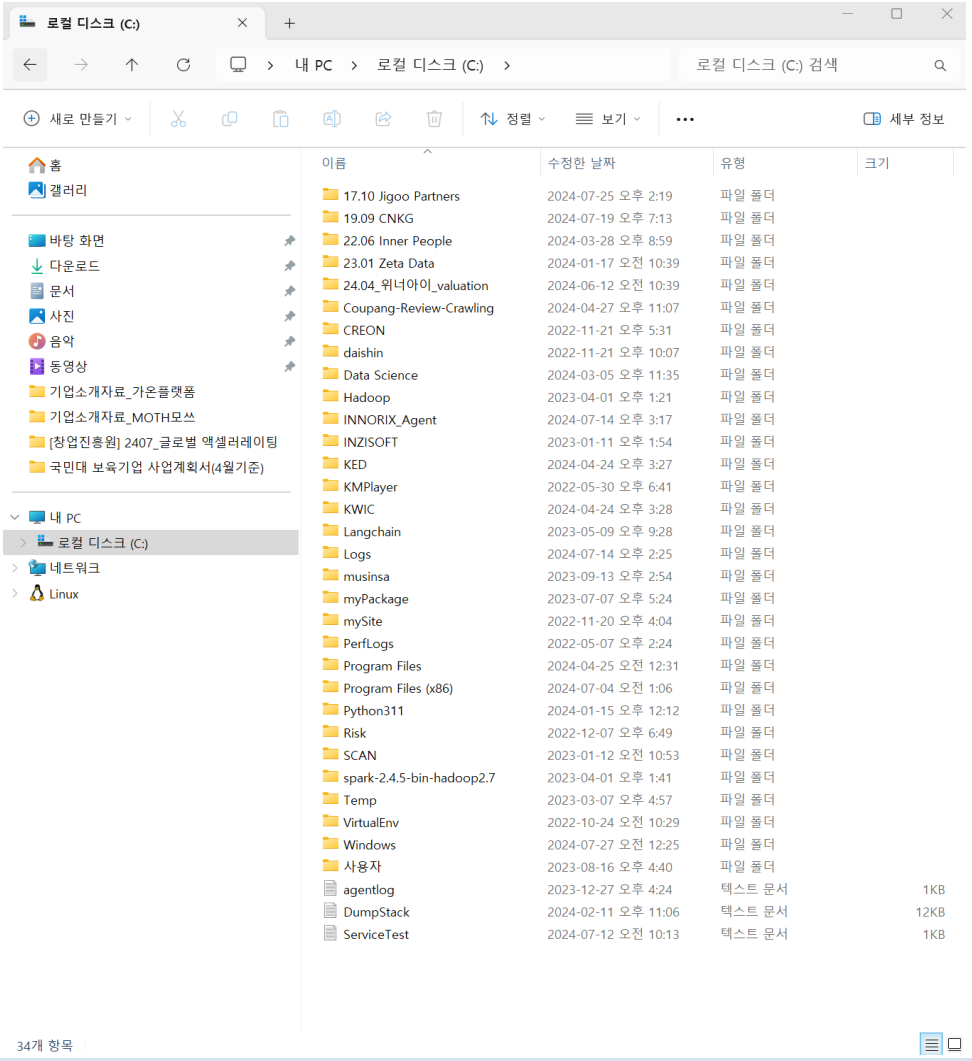
검색 엔진 최적화

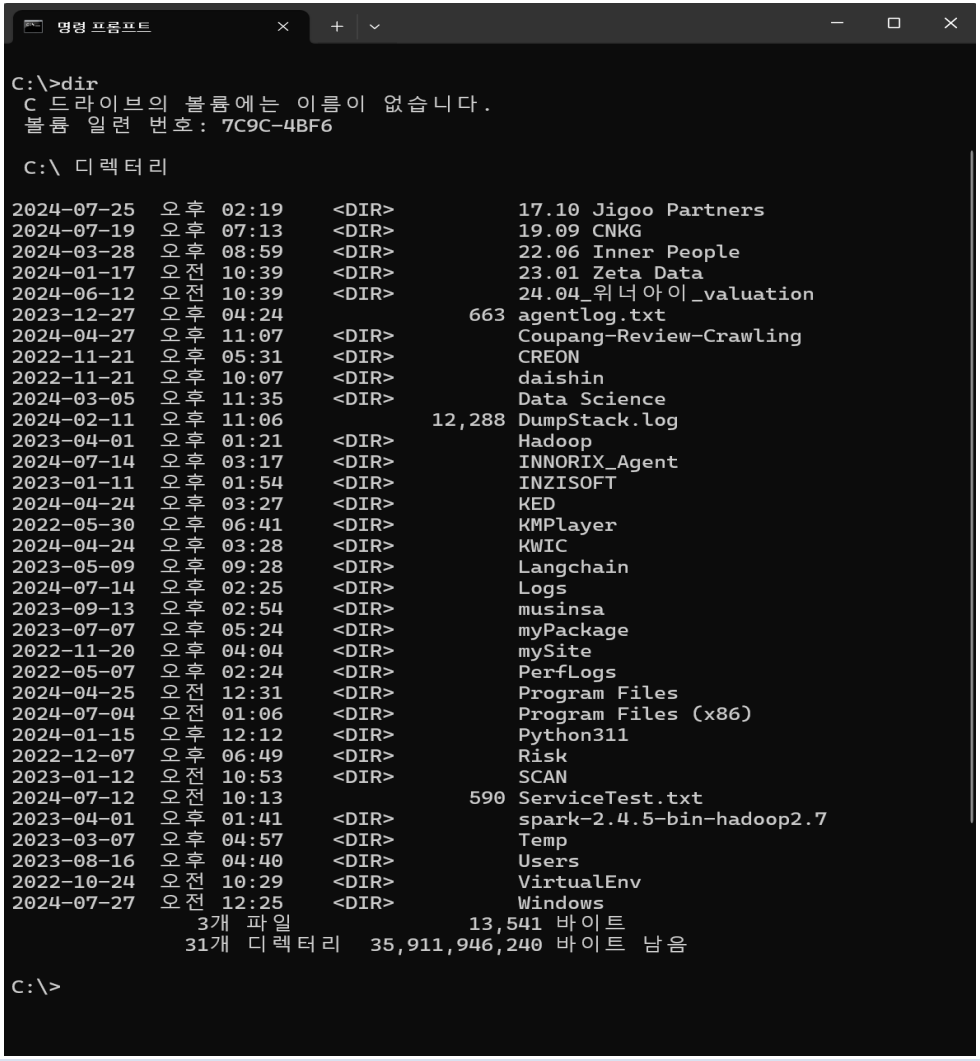
검색엔진최적화(Search Engine Optimization)는 검색엔진으로부터 웹사이트나 웹페이지에 대한 트래픽의 품질과 양을 개선하는 과정

- 웹 페이지 검색엔진이 자료를 수집하고 순위를 매기는 방식에 맞게 웹 페이지를 구성해서 검색 결과의 상위에 나올 수 있게 함
- 웹 페이지와 관련된 검색어로 검색한 검색 결과 상위에 나오게 된다면 방문 트래픽이 늘어나기 때문에 효과적인 인터넷 마케팅 방법 중의 하나이며 비용처리 없는 마케팅이라고 할 수 있음
- 기본적인 작업 방식은 특정한 검색어를 웹 페이지에 적절하게 배치하고 다른 웹 페이지에서 링크가 많이 연결되도록 하는 것
- 구글 등장 이후 검색 엔진들이 콘텐츠의 신뢰도를 파악하는 기초적인 지표로 다른 웹사이트에 얼마나 인용되었나를 사용하기 때문에 타 사이트에 인용되는 횟수를 늘리는 방향으로 최적화함

Web Crawling... 시작하기 전에

DOS를 아십니까?





웹 데이터 가져오기

단 3줄로 가능한 쿠팡 상품 댓글 크롤링

1. 인용 사이트 : https://github.com/JaehyoJJAng/Coupang-Review-Crawling
2. GitHub 회원가입 (https://github.com/)
   ➔ 참조 사이트 : https://velog.io/@noyohanx/GIT-Github-%EA%B0%80%EC%9E%85%ED%95%98%EA%B8%B0
3. Python 설치
   ➔ (https://dotiromoook.tistory.com/32)
4. 명령 프롬프트 또는 "cmd" 실행 ➔ DOS 화면 뜸
5. $ git clone https://github.com/JaehyoJJAng/Coupang-Review-Crawling.git
6. $ cd Coupang-Review-Crawling
7. $ pip install -r ./requirements.txt

반드시 참조

실행1 $ python main.py

실행2 복사된 url 주소 붙여넣기 ➔ Enter ➔ 페이지 수 지정 ➔ Enter

“꾸준한 연습과 반복을 권합니다.”

홍용기 컨설팅학박사
010-3366-9010 / 123biz@naver.com