

Introduction

검색에 대한 해답 - Serlution 팀

Crawling

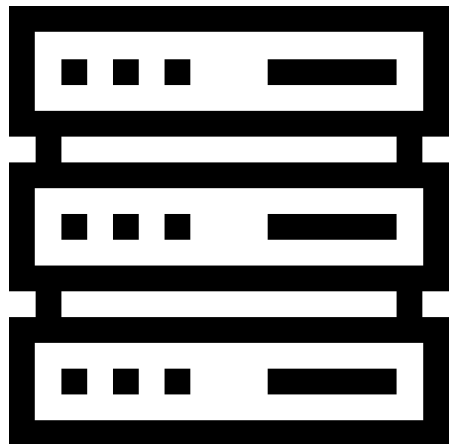


크롤러(프로그램)가 일정한 규칙으로 웹페이지를 브라우징
하는 것

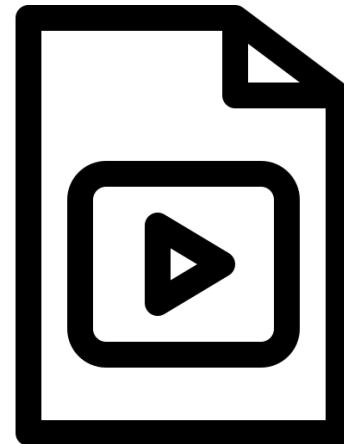
|| ?

- 스크래핑: 웹에서 원하는 데이터를 수집하는 작업

주요 대상



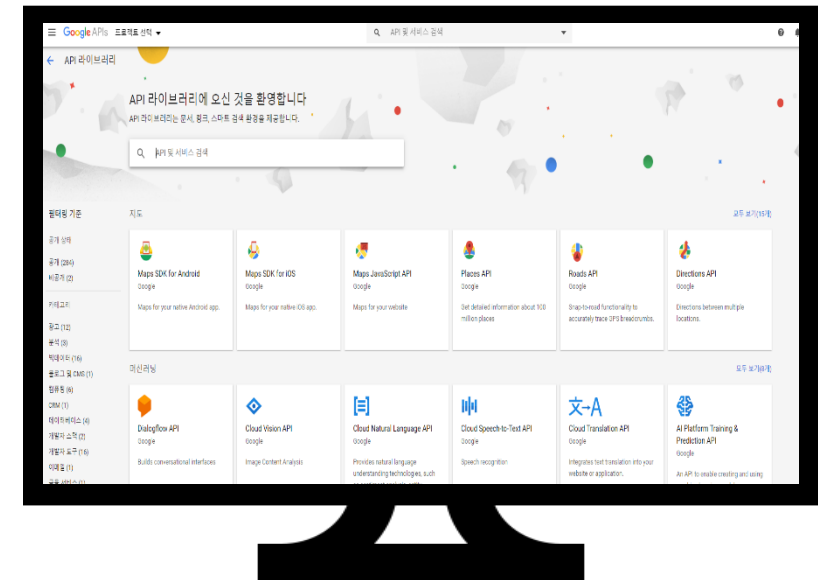
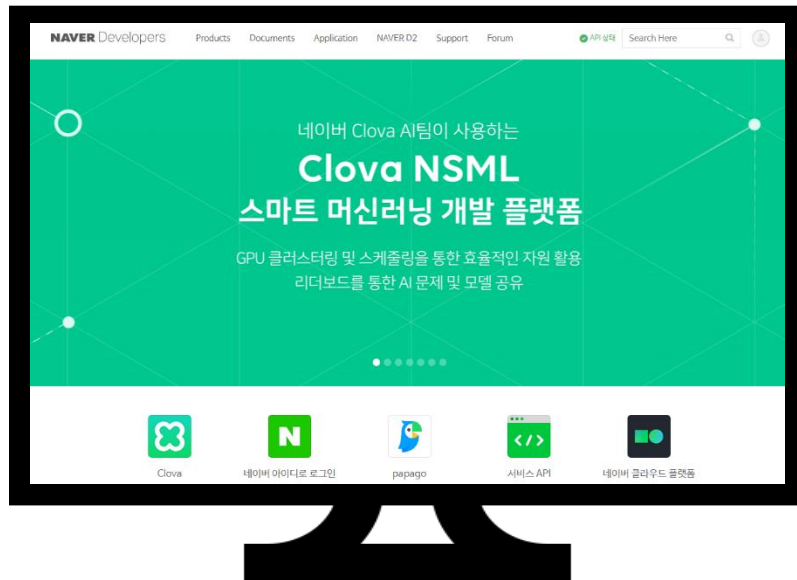
기업/조직이 생성하는 데이터
(뉴스, 제품정보, 미디어)



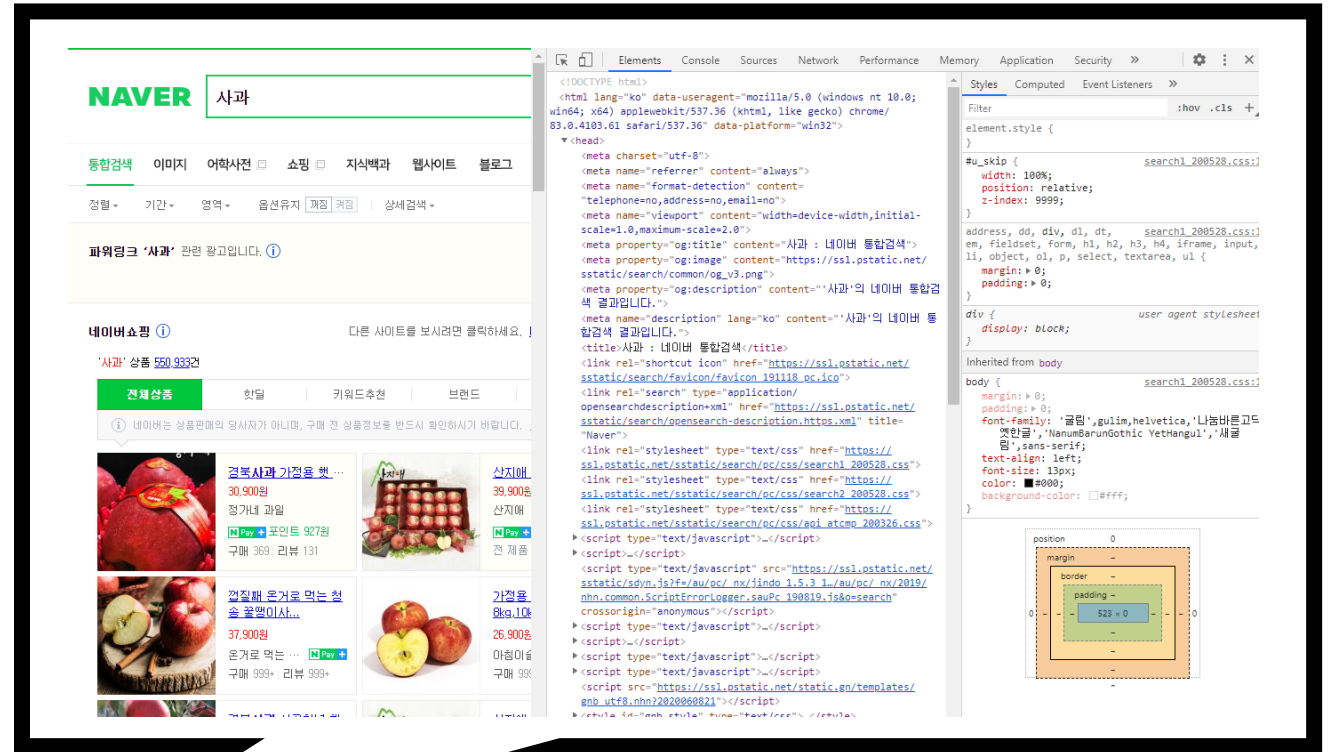
사용자가 생성하는 데이터
(조회수, 댓글, SNS)

API 활용

데이터를 수집하고자 하는 사이트/회사에서 제공하는 API 활용



코드 작성



- API 제공 X
- 원하는 데이터 X

분석 방법(1)

키워드 빈도 분석

크롤링이란 웹사이트의 내용을 그대로 가져와서 필요한 데이터를 추출해 내는 것을 말합니다. **크롤링**이 웹사이트에 대하여 이루어지면 웹**크롤링**이라고 하고 소프트웨어 봇을 통상크롤러라고 부릅니다. 즉, **크롤링**은 데이터를 대량수집하는 기법으로써 대중에게 보여주게 됩니다. 모든 **크롤링** 행위가 위법이라고 단정지을 수는 없고 합법적인 **크롤링**과 불법적인 **크롤링**으로 구별지을 수는 있습니다. 합법적인 **크롤링**은 사이트 운영자에 의사에 반하지 않는 것이며, 반대로 **크롤링** 금지 표시를 무시하고 **크롤링**을 하였다면 이는 불법 **크롤링**입니다.

분석 방법(2)

키워드 중요도 분석

<코로나>

8일에도 코로나 확산세가 지속됐다고 **밝혔다**.
서울시는 8일 오후 6시 기준 오전 10시 대비 코로나19 확진자가 최소 17명이 발생했다고 **밝혔다**.
이날 서울 구로구는 가리봉동에 위치한 중국동포교회 쉼터가 지난 7일 리치웨이를 방문해 확진 판정을 받은 구 54번 확진자의 거주지라고 **밝혔다**.
구는 중국동포교회 거주자와 관계자 등 36명을 대상으로 긴급 전수 검체 결과를 실시했다고 **밝혔다**.
구는 해당 교회를 폐쇄조치하고 중국동포교회 신도 150여명의 명단도 확보해 전체 검체 검사를 할 예정이라고 **밝혔다**.

분석 방법(2)

키워드 중요도 분석

<코로나>

8일에도 코로나
서울시는 8일 오
나19 확진자가
이날 서울 구로
회 쉼터가 지난
을 받은 구 54
구는 중국동포교
상으로 긴급 전
다. 구는 해당
신도 150여명의
할 예정이라고

<삼성 경영권>

'불법 경영권 승계' 의혹을 받고 있는 이재용 삼성
전자 부회장의 영장실질심사가 8시간 반 만에 종료
라고 **밝혔다**. 굳은 표정으로 법원을 나온 이 부
회장은 취재진의 질문에 아무런 답을 하지 않은 채
서울구치소로 향했다고 **밝혔다**. 앞서 지난 4일
검찰은 이 부회장 등 삼성 전 현직 임원 3명에 대
해 구속영장을 청구했다고 **밝혔다**. 검찰은 이 부
회장이 삼성물산 합병 과정에서 경영권 승계를 목
표로 지시 기획한 것으로 의심한다고 **밝혔다**. 이
부회장이 혐의를 전면 부인하고 있고, 검찰은 추가
증거 인멸 우려를 들어 구속 수사의 필요성을 주장
한다고 **밝혔다**.

분석 방법(2)

키워드 중요도 분석

<코로나>

8일에도 코로나
서울시는 8일 오
나19 확진자가
이날 서울 구로
회 쉼터가 지난
을 받은 구 54
구는 중국동포교
상으로 긴급 전
다. 구는 해당
신도 150여명의
할 예정이라고

<삼성>

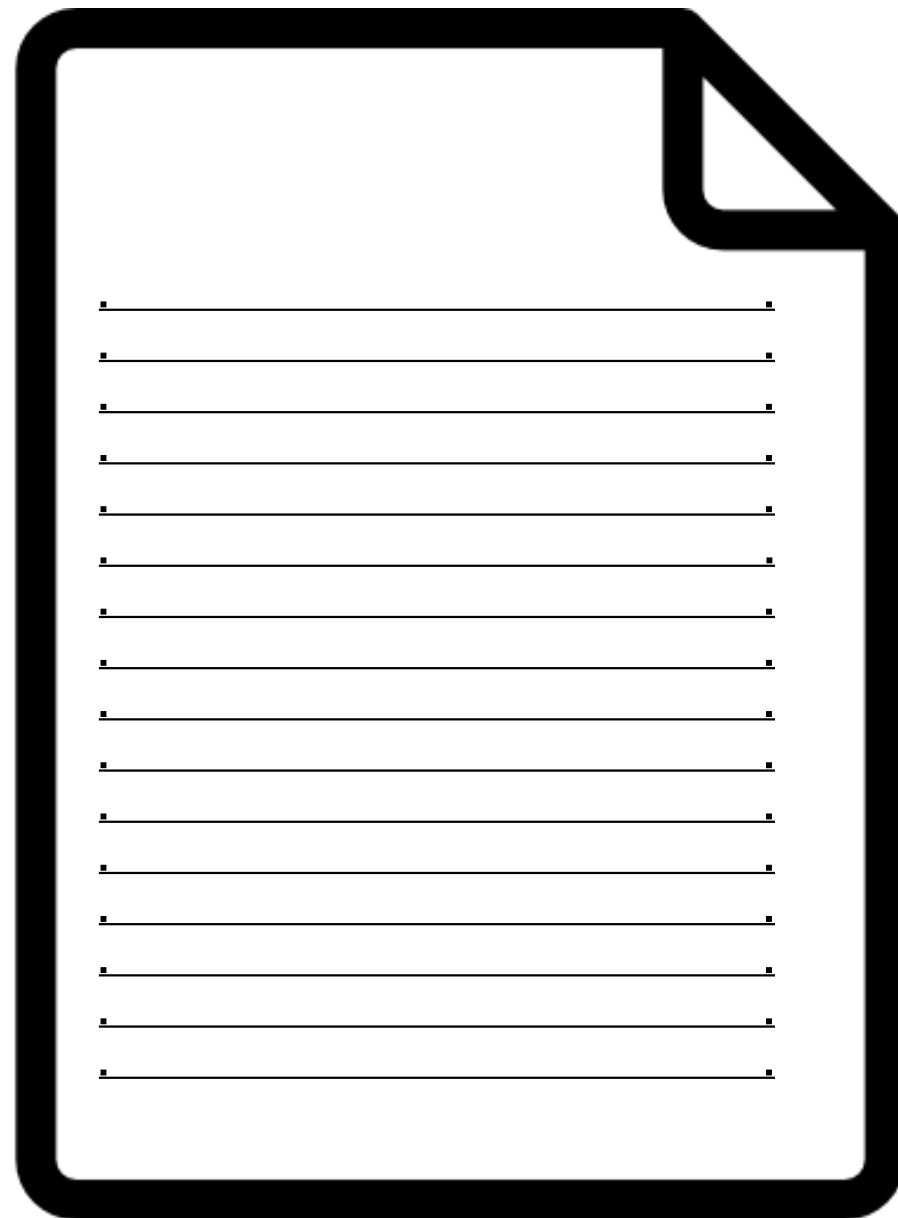
'불법 경영권 승
전자 부회장의
라고 **밝혔다**
회장은 취재진
서울구치소로
검찰은 이 부회
해 구속영장을
회장이 삼성물
표로 지시 기획
부회장이 혐의
증거 인멸 우려
한다고 **밝혔다**

<승용차 판매>

세계 최대 자동차 시장인 중국에서 5월 신차 판매
량은 전년 동월 대비 1.9% 늘어난 163만9000대
를 기록했다고 8일 **밝혔다**. 매체에 따르면 이날
5월 신차 판매 대수가 전월보다는 12.3% 크게 늘
어났다고 **밝혔다**. 승용차에는 SUV와 미니밴
등을 포함한다. 미니밴을 제외한 협의의 승용차 판
매는 전년 같은 달보다 1.8% 증가했다고 **밝혔다**.
4월에 비해선 12.6% 증대했다고 밝혔다. 내역을
보면 일반 승용차가 작년 동월에 비해 4.8% 줄어
들었다고 **밝혔다**. 반면 SUV는 75만9000대,
미니밴이 4.9% 증가한 3만대로 집계됐다고 **밝**
혔다.

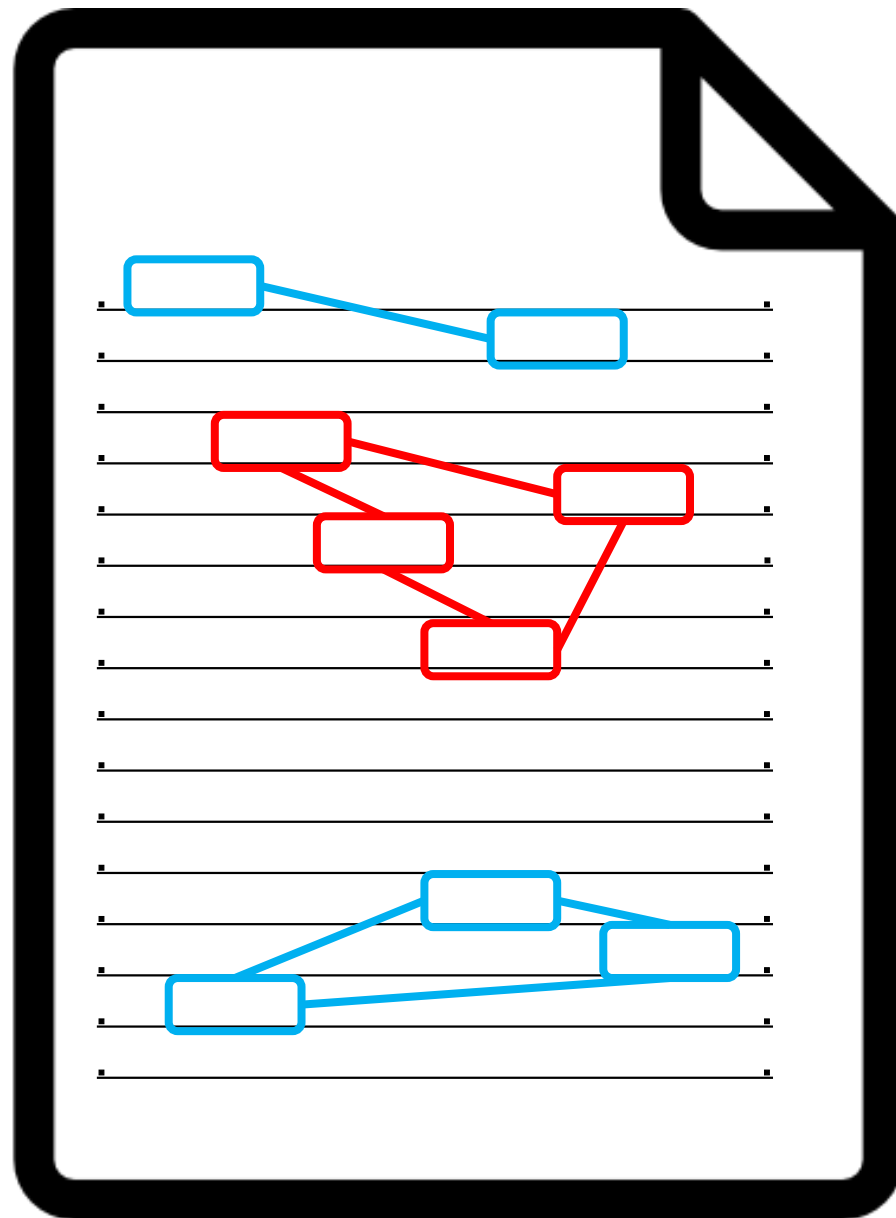
분석 방법(3)

텍스트 마이닝



분석 방법(3)

텍스트 마이닝



분석 방법(3)

텍스트 마이닝

노조파업
소상공인들 구조조정
고착화 강성귀족노조
포장업계 기업투자 감세 정규직화
위기관리 퇴직자 국제정세 재벌개혁
대내외적 디플레이션 장기화 민생안정
25대책 경제상황 실업난 역차별 자유화
인구감소 관료주의 고용불안 경영안정
경제성장 청년층 실업 사회문제 무한경쟁
노동시장 양극화 가계부채 부익부
불황 비관론 양극화 부실기업 노동계
취업난 빈부격차 상시화 동일노동
주거안정 인구증가 고령화 청년세대
구직자들 미증유 인구절벽 경제체제
사회전반 대응전략 대안적 대책마련
소득주도성장론 금융연구원
마이너스금리
하우스푸어

대표적인 예시로 **연관어 분석**이 있음

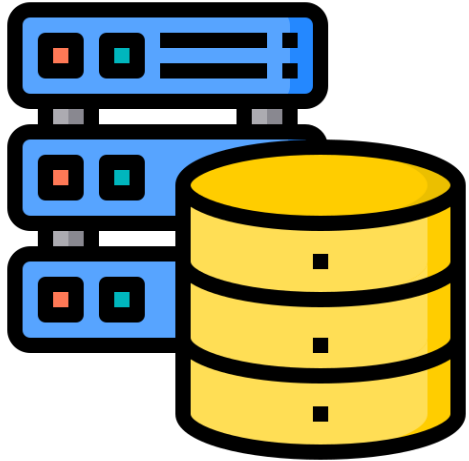
관련기술

Searution 팀

DataBase



DataBase



여러 사람이 공유하여 사용할 목적으로 체계화해 통합, 관리하는 **데이터의 집합**



RDB

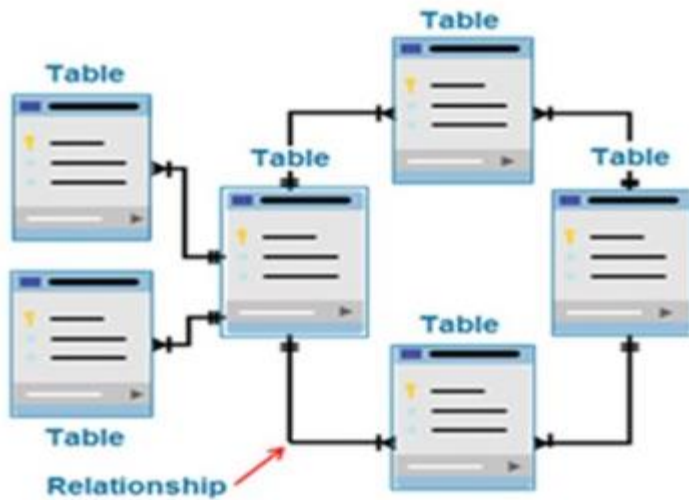


NoSQL



Etc. 네트워크형, 계층형

RDBMS



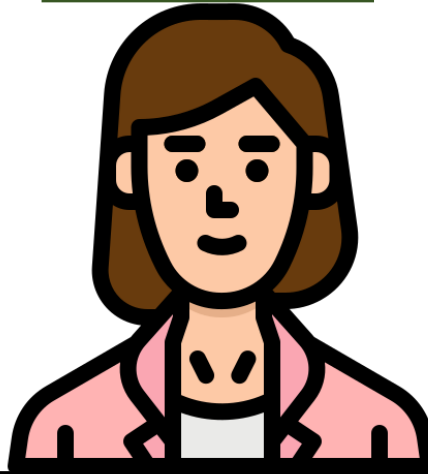
관계형 데이터베이스 관리 시스템,
RDBMS는 정해져 있는 데이터 스키마에 따라 데이터베이스 **테이블에**
저장하며,
테이블간 연결(= 관계)을 통해 데이터를 관리함
따라서 RDBMS는 데이터 관리를 효율적으로 하기 위해 **구조화가 중요**
함

MariaDB

ORACLE®



몬티 와이드니어스



My



Maria

MariaDB

DML



select, insert, delete, update 등



데이터 조작어: 데이터를 조회하거나 변형(삽입, 수정, 삭제)을 하는 명령어

DDL



create, drop, alter, rename 등



데이터 정의어: 테이블과 같은 데이터구조를 정의하는데 사용되는 명령어

DCL



grant, revoke



데이터 제어어: 데이터베이스에 접근하고 객체들을 사용하도록 권한을 주고 회수하는 명령어

DB 처리기술

Hadoop

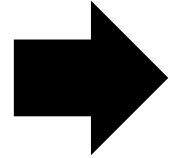


빅데이터 인프라 기술 중에 하나로 분산처리를
통해 수많은 데이터를 저장하고 처리하는 기술

› 맵-리듀스

› 분산형 파일 시스템

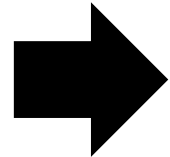
Map-Reduce



100

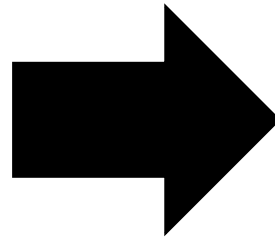


100



100

Map-Reduce



Map-Reduce

Map

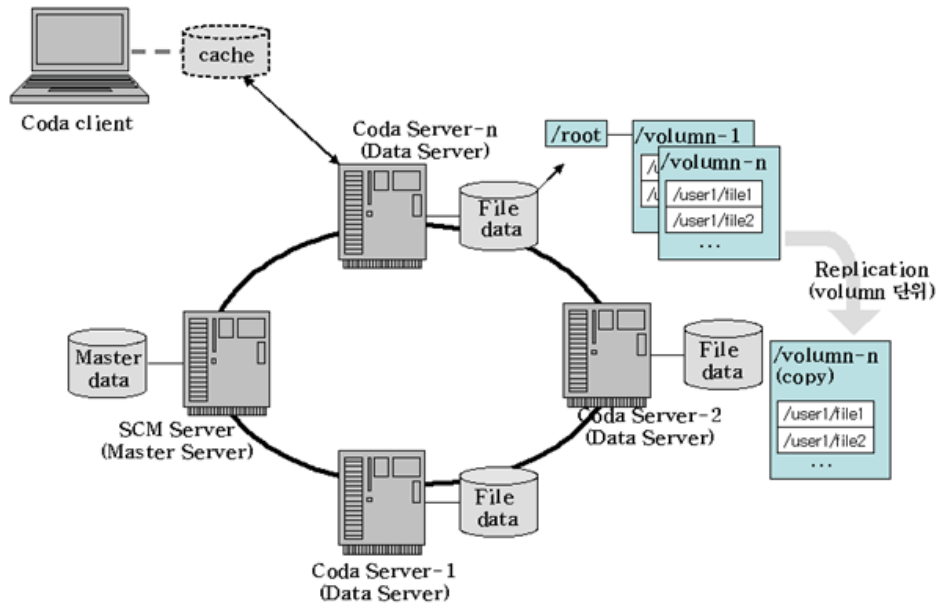
흩어져 있는 데이터를 **key, value**로 데이터를 묶는 단계

Reduce

Map단계의 **key**를 중심으로 필터링 및 정렬하는 단계

map, reduce 함수를 통해서 구현할 수 있음.

분산형 파일 시스템



여러 대의 컴퓨터가 데이터를 처리하는 시스템.

하지만 데이터를 분산 처리하되 3개의 카피를 나누어 저장해두기 때문에 **한** 컴퓨터에서 고장이 발생하거나 느려지더라도 데이터에 접근할 수 있다.

web



Framework

Frame + work

“틀”

“일”

➔ “틀 위에서 일한다.”

소프트웨어 어플리케이션이나 솔루션의 개발을 수월하게 하기 위해 소프트웨어의 구체적 기능들에 해당하는 부분의 설계와 구현을 재사용 가능하도록 협업화된 형태로 제공하는 소프트웨어 환경

Example

언어



프레임워크



Spring



자바 플랫폼을 위한 오픈소스 애플리케이션 프레임워크로서 엔터프라이즈급 애플리케이션을 개발하기 위한 모든 기능을 종합적으로 제공하는 경량화된 솔루션

“Spring Framework는 경량 컨테이너로 자바 객체를 담고 직접 관리함.

객체의 생성 및 소멸 그리고 라이프 사이클을 관리하며

언제든 Spring 컨테이너로 부터 필요한 객체를 가져와 사용할 수 있음.

이는 Spring이 IOC(?) 기반의 Framework임을 의미함.”

출처: 구글링 된 어느 블로그

주요개념: 의존성

1. 의존성

```
public class StartGame {  
    public static void main(String[] args)  
    {  
        new MainFrame();  
    }  
}
```

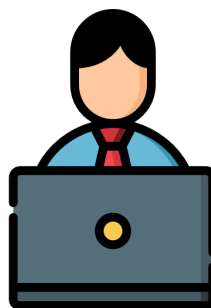
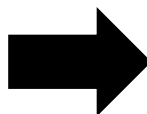
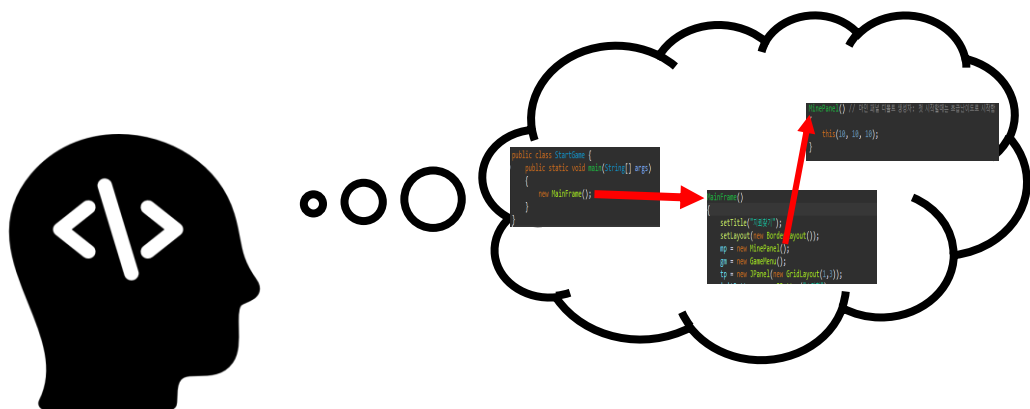
```
MainFrame()  
{  
    setTitle("지뢰찾기");  
    setLayout(new BorderLayout());  
    mp = new MinePanel();  
    gm = new GameMenu();  
    tp = new JPanel(new GridLayout(1,3));  
}
```

```
MinePanel() // 마인 패널 디폴트 생성자: 첫 시작할때는 조금난이도로 시작할  
{  
    this(10, 10, 10);  
}
```

클래스 내에서 다른 클래스의 객체를 생성하고 해당 객체내의 메서드를 호출하는 것
예) *StartGame*은 *MainFrame*에 의존성이 있다

주요개념: IOC (제어의 역전)

기존의 프로그래밍 방식

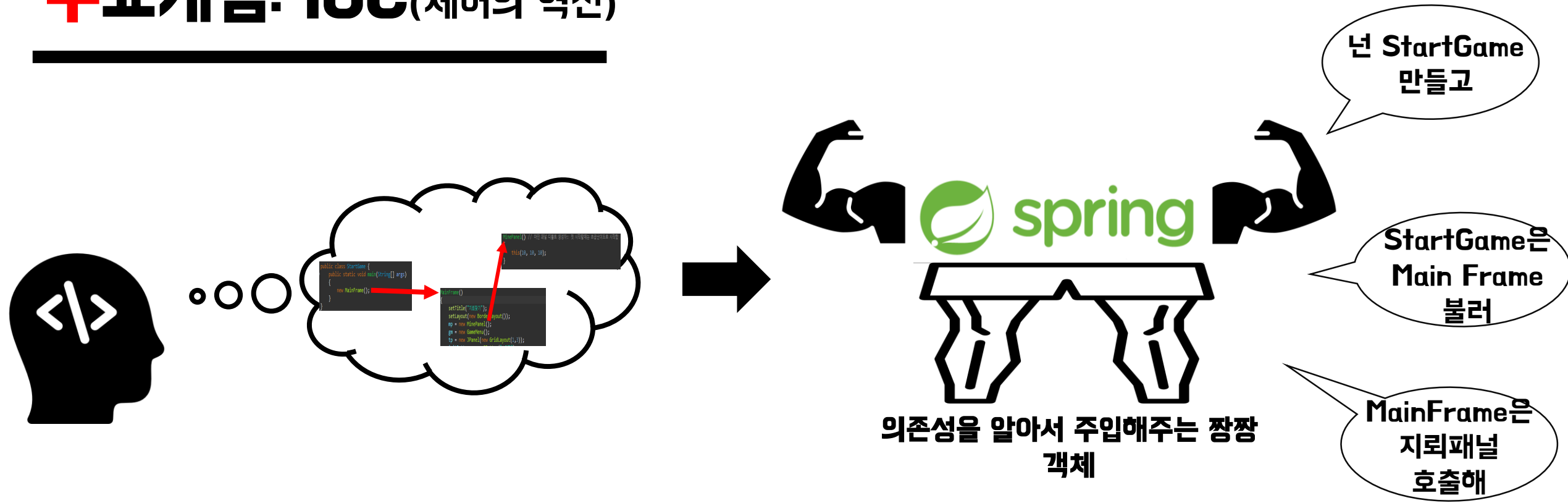


보자.. StartGame의 메인에서 MainFrame을 호출하고..

MainFrame에서는 지뢰패널을 호출하고 게임을 시작해야겠네..

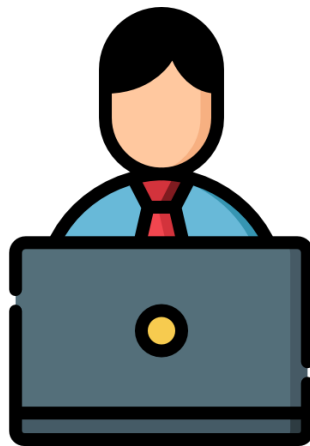
지뢰패널은 버튼객체를 갖고 있으면 될 것 같구...

주요개념: IOC(제어의 역전)



IOC란 기존 사용자가 모든 작업을 제어하던 것을 특별한 객체에 모든 것을 위임하여 객체의 생성부터 생명주기 등 모든 객체에 대한 제어권이 넘어 간 것을 IOC. 즉, 제어의 역전 이라고 함

Spring의 단점



- ➡ **버전 관리**가 힘들다.(버전 별로 따로 해줘야 함).
- ➡ 개발자들이 개발할 때 과거에 성공적으로 사용했던 버전 및 기본 설정만을 가져다 쓰다 보니 **프로그램 모델이 경직**되어 있다. 따라서 **새롭고 효율적인 변화에 대한 도전이 부족**하다.
- ➡ **프로젝트 환경 설정에 긴 과정과 시간이 할애**된다. 따라서 관련 분야의 전문가가 필요하다.
- ➡ **전문가 의존도가 높다.**

SpringBoot

기존 스프링의 단점을 보완하고자 만든 프로젝트,
스프링 부트는 스프링 프레임 워크에 속하는 서브 프로젝트이다.

장점

자동설점

Example



```
<dependency>
  <groupId>org.springframework</groupId>
  <artifactId>spring-webmvc</artifactId>
  <version>4.2.2.RELEASE</version>
</dependency>
<dependency>
  <groupId>com.fasterxml.jackson.core</groupId>
  <artifactId>jackson-databind</artifactId>
  <version>2.5.3</version>
</dependency>
<dependency>
  <groupId>org.hibernate</groupId>
  <artifactId>hibernate-validator</artifactId>
  <version>5.0.2.Final</version>
</dependency>
<dependency>
  <groupId>log4j</groupId>
  <artifactId>log4j</artifactId>
  <version>1.2.17</version>
</dependency>
```



Spring Boot

```
<dependency>
  <groupId>org.springframework.boot</groupId>
  <artifactId>spring-boot-starter-web</artifactId>
</dependency>
```

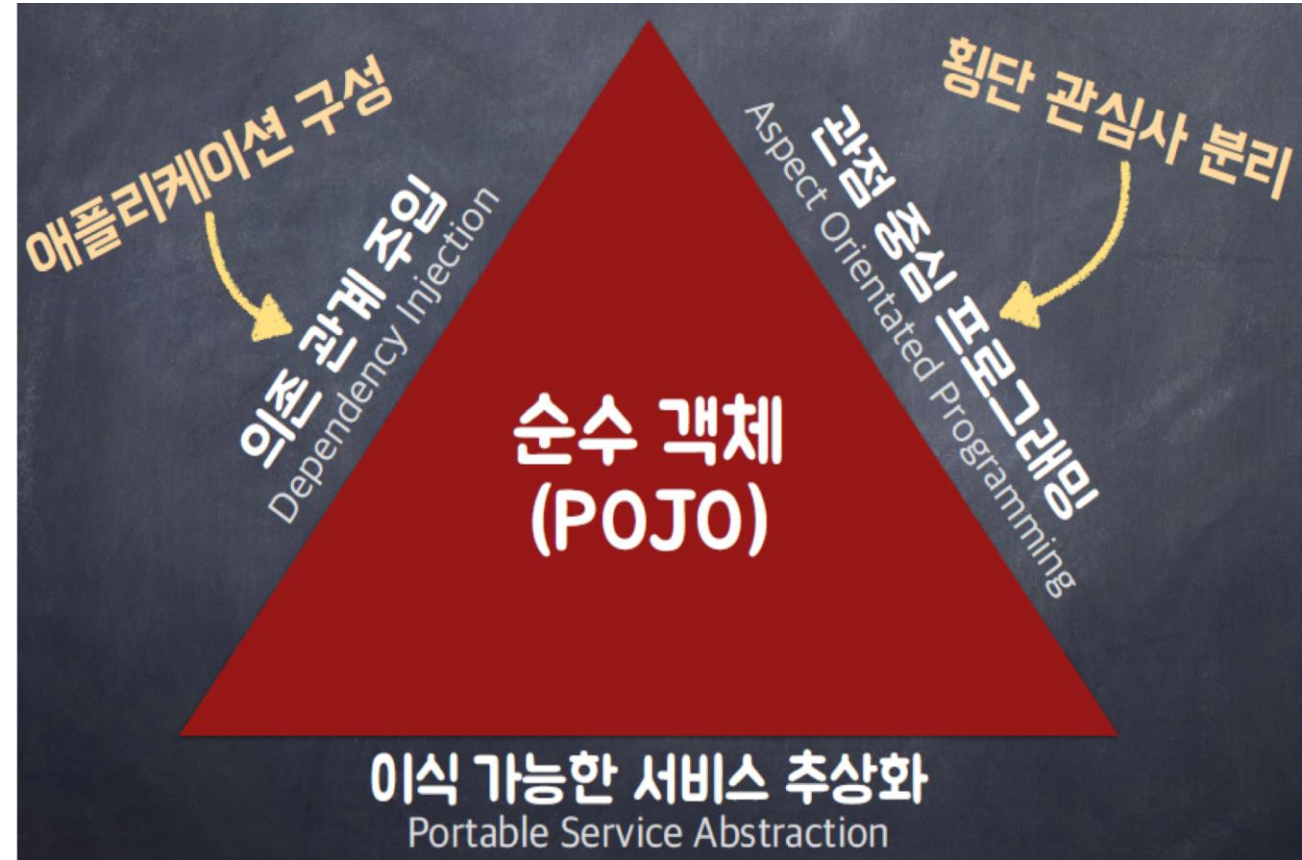
Spring
server

3 가지의 중요개념

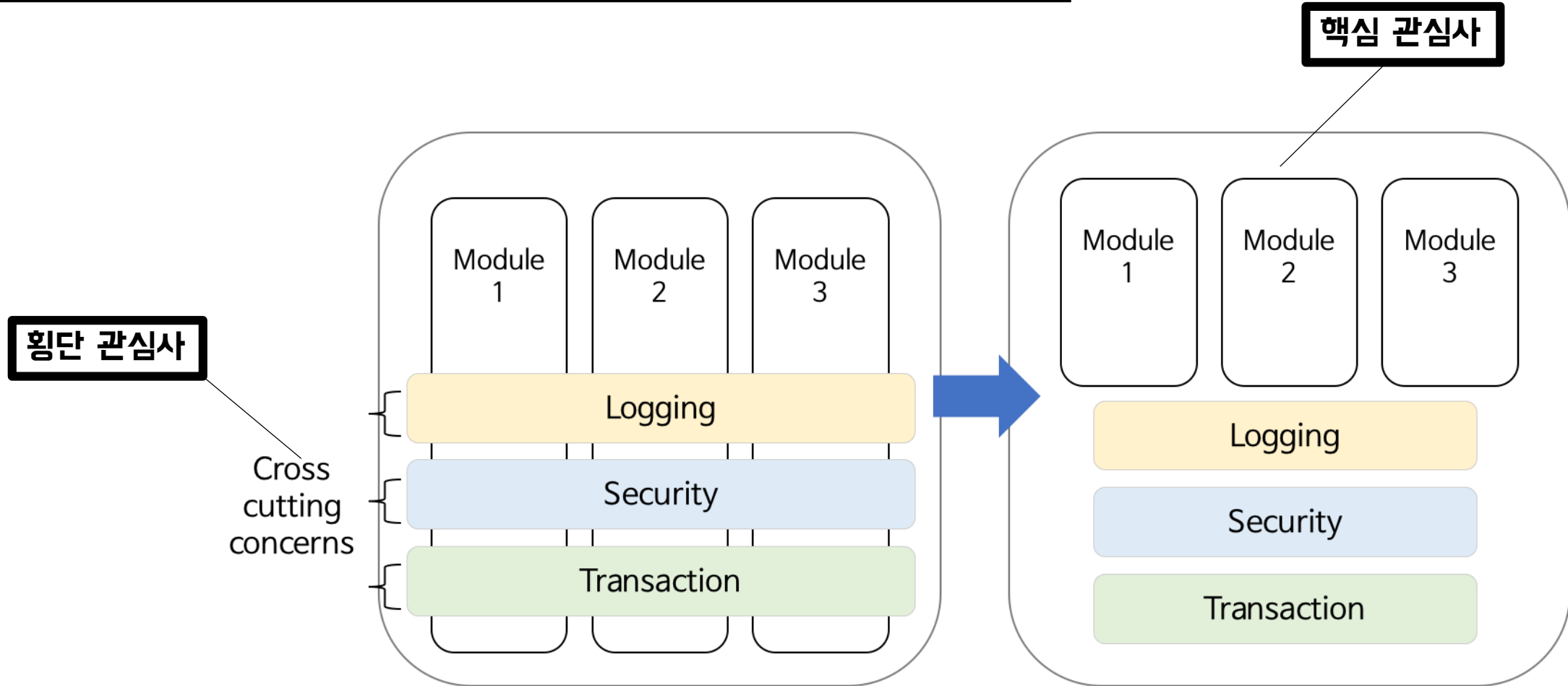
1. IoC(제어의 역전) / 의존성(DI)

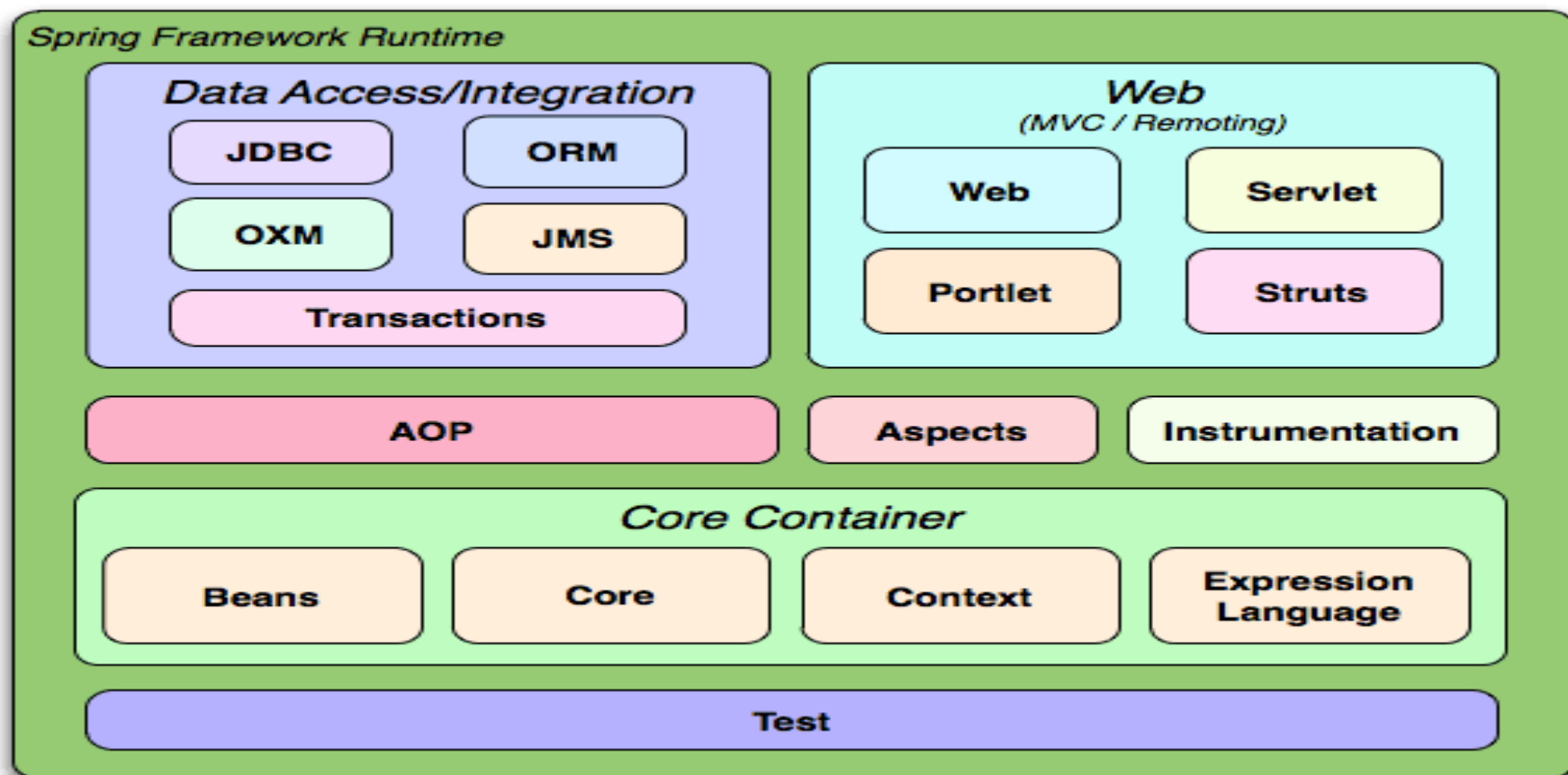
2. 서비스의 추상화

3. AoP(관점 지향 프로그래밍)

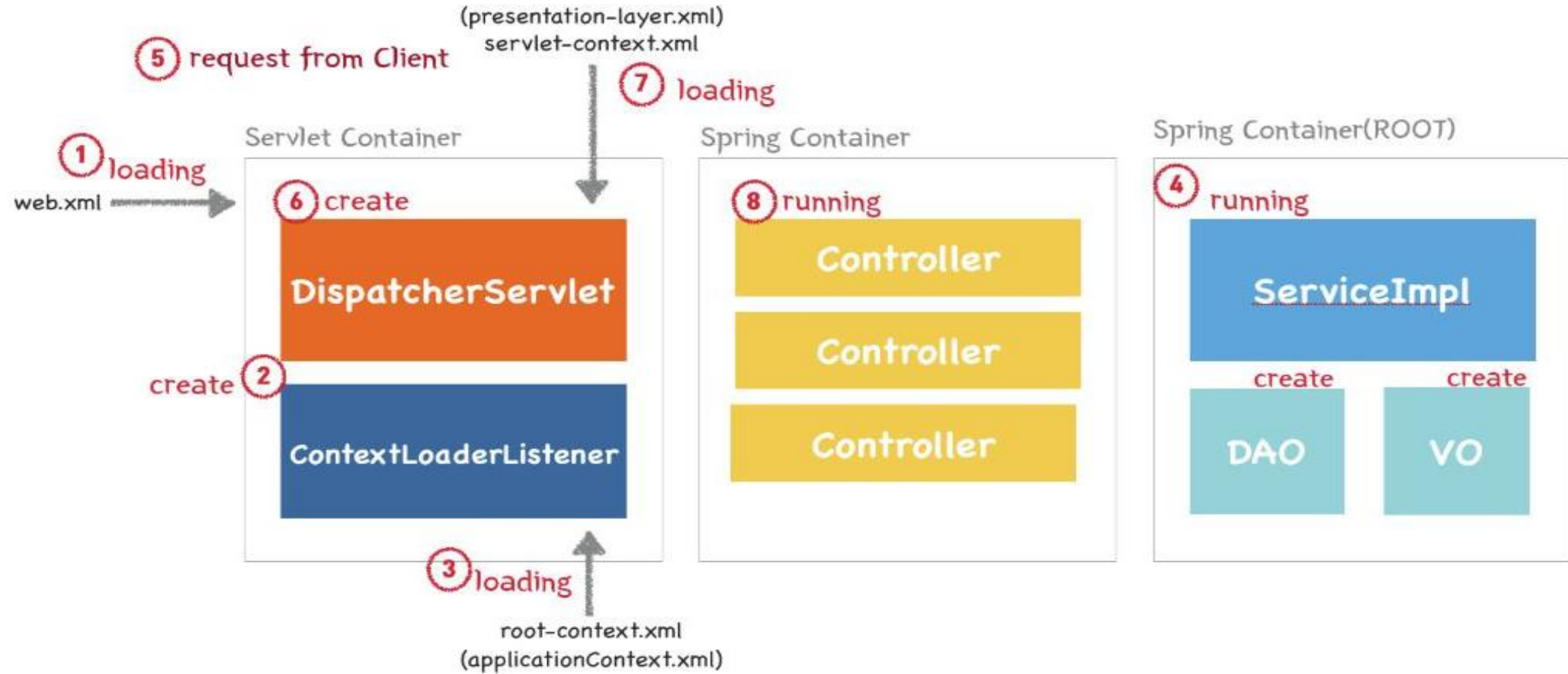


Aspect Orientated Programming





작동 원리



Spring Framework

끝

보충 설명이 필요하시면 말로 하겠습니다.