

**올인원 패키지 Online.**

# 안녕하세요 크롤링 강의 이은찬입니다.

## **PART1** | 크롤링 기초

크롤링이란 무엇이고 어떻게 하나요?

## **PART2** | 정적 크롤링

빠르고 간단한 크롤링

## **PART3** | 동적 크롤링

셀레니움을 활용한 크롤링

# [크롤링 기초] http와 소켓

# HTTP?

## 1. 크롤링 기초



# HTTP?

## 1. 크롤링 기초

GET /index.html HTTP/1.1

User-Agent: Mozilla/4.0 (compatible; MSIE5.01; Windows NT)

Host: www.fastcampus.com

Accept-Language: ko-kr

Accept-Encoding: gzip, deflate

Connection: Keep-Alive

# HTTP?

## 1. 크롤링 기초

GET /index.html HTTP/1.1

User-Agent: Mozilla/4.0 (compatible; MSIE5.01; Windows NT)

Host: www.fastcampus.com

Accept-Language: ko-kr

Accept-Encoding: gzip, deflate

Connection: Keep-Alive

HTTP/1.1 200 OK

User-Agent: Mozilla/4.0 (compatible; MSIE5.01; Windows NT)

Host: www.fastcampus.com

Accept-Language: ko-kr

Accept-Encoding: gzip, deflate

Connection: Keep-Alive

Content-Type: text/xml; charset=utf-8

Content-Length: 1000

<html>hello</html>

# HTTP?

1.  
크롤링 기초

Request Line

Request-Line = Method SP Request-URI SP HTTP-Version CRLF

0개 이상의 헤더

빈 라인

바디

# HTTP?

1.  
크롤링 기초

Request Line

Request-Line = Method SP Request-URI SP HTTP-Version CRLF

0개 이상의 헤더

POST /upload HTTP/1.1

User-Agent: Mozilla/4.0 (compatible; MSIE5.01; Windows NT)

빈 라인

Host: www.fastcampus.com

Accept-Language: ko-kr

바디

Accept-Encoding: gzip, deflate

Connection: Keep-Alive

Content-Type: text/xml; charset=utf-8

Content-Length: 1000

Image body image body image body

# HTTP Methods

**1.**  
크롤링 기초

GET

HEAD

POST

PUT

DELETE

OPTIONS



# HTTP Example Login

1.  
크롤링 기초

GET /money HTTP/1.1

# HTTP Example Login

1.  
크롤링 기초

GET /money HTTP/1.1

HTTP/1.1 400 Unauthorized

# HTTP Example Login

1.  
크롤링 기초

100-199 정보전달용

200-299 성공

300-399 리다이렉트

400-499 클라이언트 에러

500-599 서버 에러

# HTTP Example Login

1.  
크롤링 기초

POST /login HTTP/1.1

User-Agent: Mozilla/4.0 (compatible; MSIE5.01; Windows NT)

id=apple&password=banana

HTTP/1.1 200 OK

Set-Cookie: JSessionId=abcd1234

# HTTP Example Login

1.  
크롤링 기초

GET /money HTTP/1.1

Cookie: JSessionId=abcd1234

HTTP/1.1 200 OK

money=15000

# HTTP vs. HTTPS?

1.  
크롤링 기초

POST /login HTTP/1.1

User-Agent: Mozilla/4.0 (compatible; MSIE5.01; Windows NT)

id=apple&password=banana

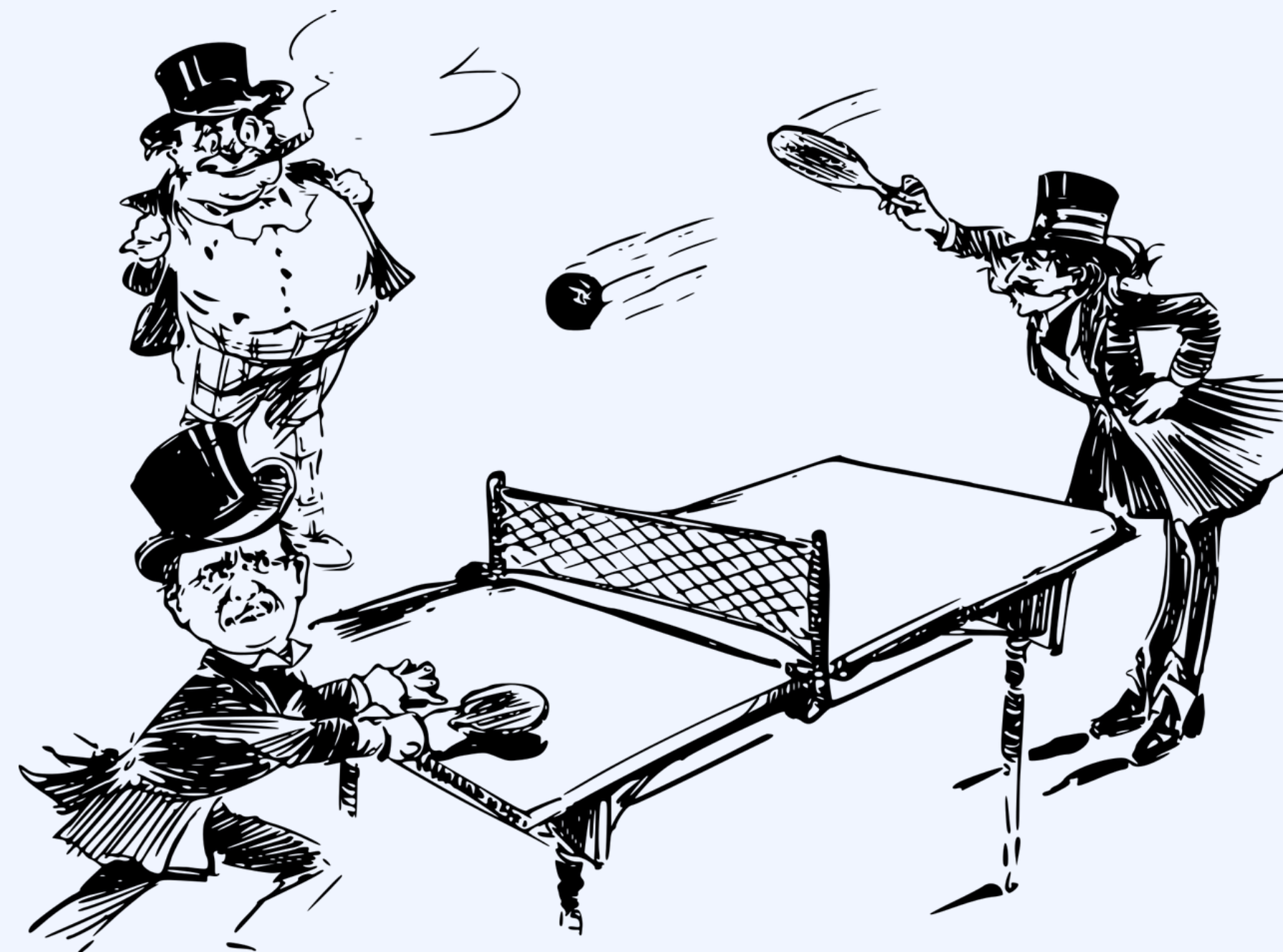
HTTP/1.1 200 OK

Set-Cookie: JSessionId=abcd1234

# HTTP vs. Socket

1.  
크롤링 기초

HTTP : 핑퐁



# HTTP vs. Socket

1.  
크롤링 기초

HTTP : 핑퐁

Socket : 자유로움..

