

HTTP 완벽가이드

by David Gourley, Brian Totty

1부 HTTP: 웹의 기초

1장 HTTP 개관

HTTP란?

전 세계의 웹브라우저, 서버, 웹 애플리케이션은 모두 HTTP를 통해 대화한다.

즉, HTTP는 현대 인터넷의 공용어이다.

* 인터넷? 컴퓨터로 연결하여 TCP/IP라는 통신 프로토콜을 이용해 정보를 주고받는 컴퓨터 네트워크

HTTP : HyperText Transfer Protocol

하이퍼텍스트 문서를 전송하기 위하여 사용되는 통신규약

‘현대’ 인터넷의 공용어, HTTP

HTTP는 1989년 팀 버너스 리(Tim Berners Lee)에 의하여 처음 설계되었다.

HTTP의 첫번째 버전은 인터넷을 통하여 가공되지 않은 데이터를 전송하기 위한 단순한 프로토콜이었으나, 데이터에 대한 전송과 요청·응답에 대한 수정 등 가공된 정보를 포함하는 프로토콜로 개선되었다.

팀 버너스 리 >>>>
W3C 창립자



근데 인터넷은 뭐지? HTTP보다 먼저 있었다.

인터넷은 1960~1970년대 미국 국방부 산하의 고등 연구국의 연구용 네트워크가 시초이다.

현재와 같이 TCP/IP 기반의 네트워크가 된 것은 1983년부터이며

위 ARPANET과 별도로 1986년 미국과학재단은 NSFnet을 만들었는데, 1980년대 말 ARPANET이 흡수통합되면서 세계 모든 곳을 연결하는 국제 통신망으로 발전하게 되었다.

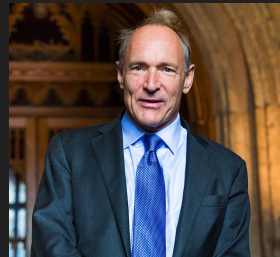
1989년 월드와이드웹이 등장하면서 네트워크 기술에 한 단계 진일보한 기술이 더해졌다. 컴퓨터의 종류가 다양해지면서 프로토콜을 재정비할 필요성이 부각되었다.

HTTP와 월드와이드웹의 관계

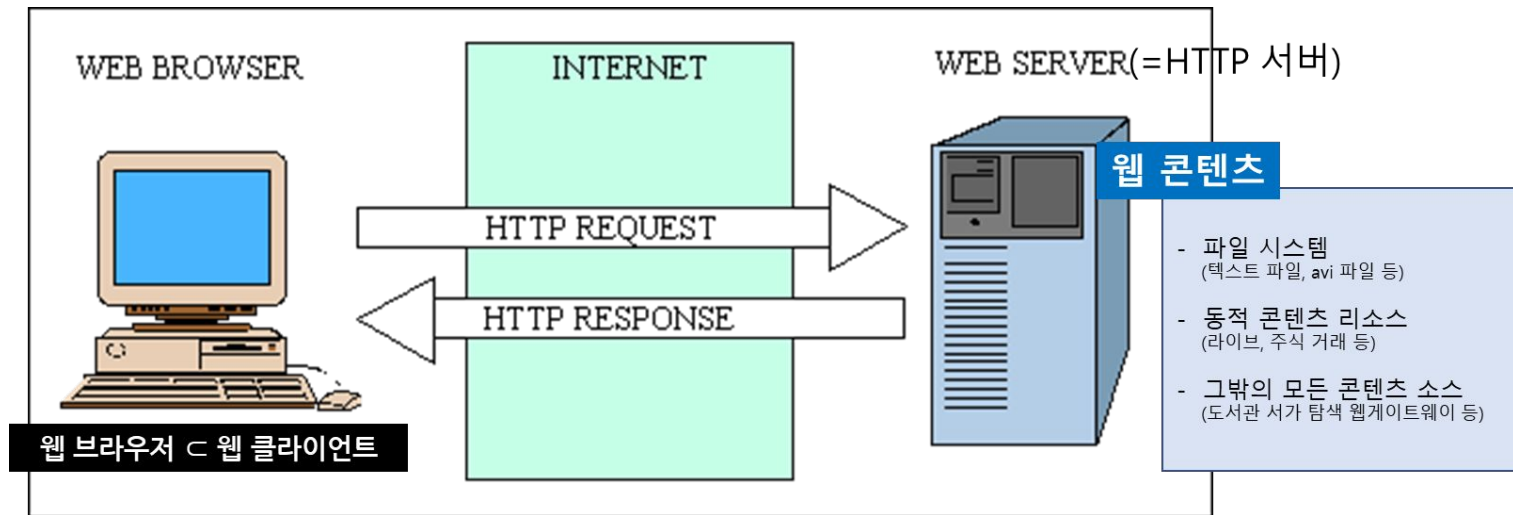
월드와이드웹(World Wide Web)은 세계 최초의 웹 브라우저이다. 1991년 팀 버너스리가 이 프로그램을 처음 소개하였다.

버너스리는 NeXT 컴퓨터를 세계 최초의 웹 서버로 사용하였다.

당시 버너스리는 CERN의 동료들과 함께 유럽입자물리연구소(CERN)로부터 지원을 받아 최초의 웹브라우저도 만들었다. 1990년 크리스마스에 버너스리와 동료들은 자신들의 컴퓨터에서 최초의 웹 브라우저를 통해 <http://info.cern.ch> 라고 등록한 웹서버에 접속했는데, 이렇게 완성된 최초의 성공적인 빌드가 월드와이드웹(WWW)이 탄생하는 순간이라고 평가되고 있다.



웹 트래픽 전달과정



HTTP 서버는 웹에서 전송되는 객체 각각에 **MIME**타입이라는 데이터 포맷 라벨을 붙인다.

웹 브라우저는 그 라벨을 통해 다룰 수 있는 객체인지 확인한 후 준비 및 실행한다.

§ MIME타입 전체목록은 627p~658p에서 확인가능

그래서 MIME 타입이 뭐지??

▼ 일반

요청 URL: `https://suji.hs.kr/main.php`

요청 메서드: GET

상태 코드:  200 OK

원격 주소: 220.78.188.171:443

리퍼러 정책: strict-origin-when-cross-origin

▼ 응답 헤더 파싱된 데이터 보기

HTTP/1.1 200 OK

Date: Fri, 08 Apr 2022 14:54:49 GMT

Server: goeia.go.kr

X-Frame-Options: SAMEORIGIN

P3P: CP="ALL IND DSP COR ADM CONo CUR CUSo IVAo IVDo PSA PSD TAI TELo OUR SAMo CNT COM INT NAV ONL PHY PRE PUR UNI"

Connection: close

Transfer-Encoding: chunked

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

URL : Uniform Resource Locator(통합자원지시자)



특정 서버의 한 리소스에 대한 구체적인 위치를 나타내고 있다.
오늘날 대부분의 URI는 URL이다.

아직 실험 중인 URN(Uniform Resource Name)

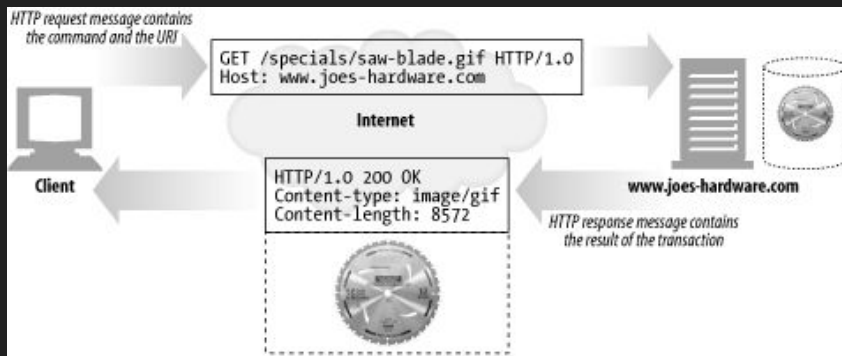
인터넷에 존재하는 각종 정보들의 유일한 이름을 표시하는 식별자이다.

URL 방식은 주소 위치에 대응되는 콘텐츠가 삭제되거나 더 이상 이용할 수 없을 경우, 검색 수단으로서의 기능을 상실하는 **문제점이 있다**.

URN에는 ISBN(국제표준도서번호), UCI(디지털콘텐츠식별체계), IPFS*(컴퓨터간 파일 시스템 ex. 토렌트) 등이 있다.

* 데이터의 내용을 변환한 해시값을 이용하여 전 세계 여러 컴퓨터에 분산 저장되어 있는 콘텐츠를 찾아서 데이터를 조각조각으로 잘게 나뉘서 빠른 속도로 가져온 후 하나로 합쳐서 보여주는 방식으로 작동한다.

트랜잭션(Transaction: 거래)



애플리케이션은 하나의 작업을 수행하기 위해 여러 **HTTP** 트랜잭션을 수행한다. 따라서 시각적으로 풍부한 웹페이지를 가져올 때는 대량의 **HTTP** 트랜잭션을 수행하게 된다.

HTTP 트랜잭션은 **요청명령**과 **응답 결과**로 구성되어 있다.

메서드(Method) (3장에서 자세히 설명)

HTTP 메서드는 HTTP가 지원하는 요청 명령(메시지)을 의미한다.

모든 메시지는 한 개의 메서드를 갖는다.

[자주 쓰이는 메서드]

- **GET** : 서버에서 클라이언트로, 지정한 리소스를 보내라.
- **PUT** : 클라이언트에서 서버로 보낸 (요청)데이터를, 지정한 이름의 리소스로 저장하라(63p).
- **DELETE** : 지정한 리소스를 서버에서 삭제하라.
- **POST** : 클라이언트 데이터를 서버 게이트웨이 애플리케이션으로 보내라(63p).
- **HEAD** : 지정한 리소스에 대한 응답에서, HTTP 헤더 부분만 보내라.

상태 코드 (3장에서 자세히 설명)

모든 HTTP 응답 메시지는 상태 코드와 함께 반환된다.

[자주 쓰이는 상태 코드]

- 200 : OK. 문서가 바르게 반환되었다.
- 302 : 다시 보내라. 다른 곳에 가서 리소스를 가져가라.
- 404 : 없음. 리소스를 찾을 수 없다.

메시지 : 요청과 응답 메시지의 구조 (3장에서 자세히 설명)

▼ 요청 헤더 파싱된 데이터 보기

GET /main.php HTTP/1.1

시작줄: 무엇을 해야하는지 나타냄

Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9

Accept-Encoding: gzip, deflate, br

Accept-Language: ko-KR,ko;q=0.9,en-US;q=0.8,en;q=0.7

Cache-Control: max-age=0

Connection: keep-alive

Cookie: PHPSESSID=q4vhii7nknrad47uo2p5ij8ce7

Host: suji.hs.kr

Sec-Fetch-Dest: document

Sec-Fetch-Mode: navigate

▼ 응답 헤더 파싱된 데이터 보기

HTTP/1.1 200 OK

시작줄: 무슨 일이 일어났는지 나타냄

Date: Sun, 10 Apr 2022 15:54:49 GMT 일반 헤더(HTTP본문내용과 관련없음. 응답이 생성된 날짜 및 시간을 나타낸다.)

Server: goeia.go.kr

X-Frame-Options: SAMEORIGIN

P3P: CP="ALL IND DSP COR ADM CONo CUR CUSo IVAo IVDo PSA PSD TAI TELo OUR SAMo CNT COM INT NAV ONL PHY PRE PU

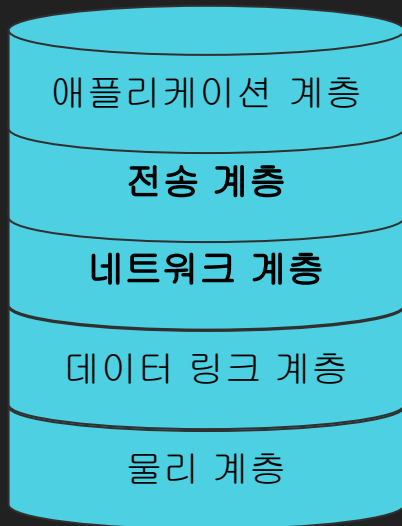
Connection: close

Transfer-Encoding: chunked

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

헤더

어떻게 메시지가 옮겨다니는 걸까?



- ❑ 애플리케이션 계층 : HTTP, FTP, DNS, SMTP ...
- ❑ 전송 계층 : TCP, UDP, DCCP ...
- ❑ 네트워크 계층 : IP ...
- ❑ 데이터 링크 계층 :
- ❑ 물리 계층 : 이더넷, wi-fi ...

HTTP는 네트워크통신의 세부사항을 TCP/IP에게 맡긴다.(신뢰성이 있기 때문!)

TCP 커넥션 : HTTP 메시지 전송을 위해 필요한 연결

커넥션을 위해 필요한 두 가지(TCP의 준비물)

1. 서버 컴퓨터에 대한 IP 주소

=> URL을 통해 알아낸다.

i) URL에 IP 주소가 있는 경우 => OK

ii) URL에 IP주소가 없는 경우 =>DNS를 통해 호스트명을 IP로 변환

2. 그 서버에서 실행 중인 프로그램이 사용 중인 포트번호

i) URL에 포트번호가 있는 경우 => OK

ii) URL에 포트번호가 있는 경우 => 기본값 80이라고 가정

웹 애플리케이션

웹브라우저와 웹서버는 웹 애플리케이션이다.

그밖의 애플리케이션 중에서 또 중요한 것은? 알아야 하는 것은?

① 프락시

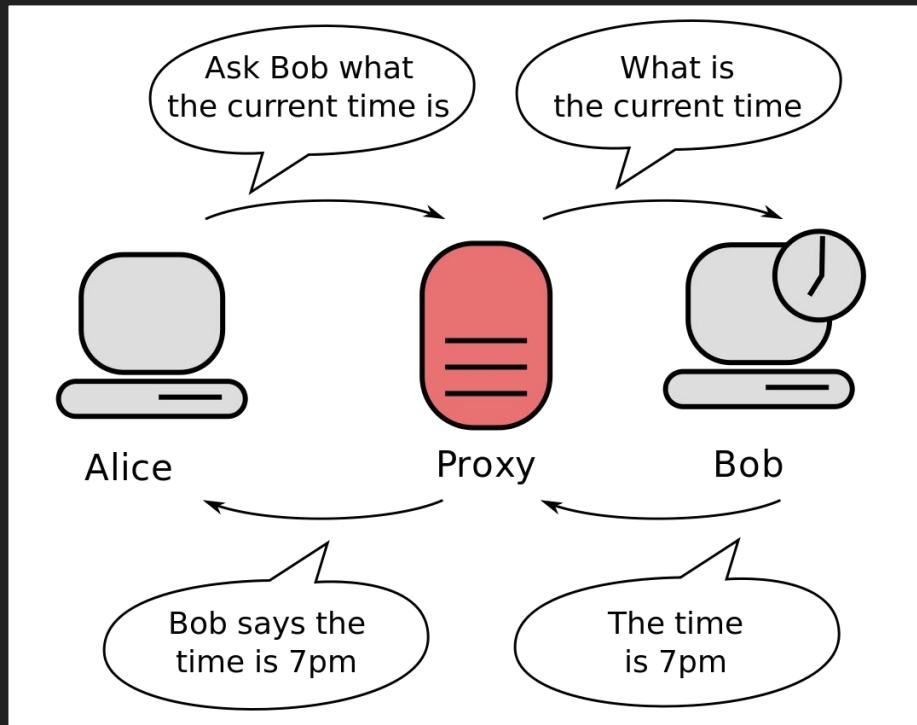
② 캐시

③ 게이트웨이

④ 터널

⑤ 에이전트

① Proxy : 클라이언트와 서버 사이에 위치한 HTTP 중개자



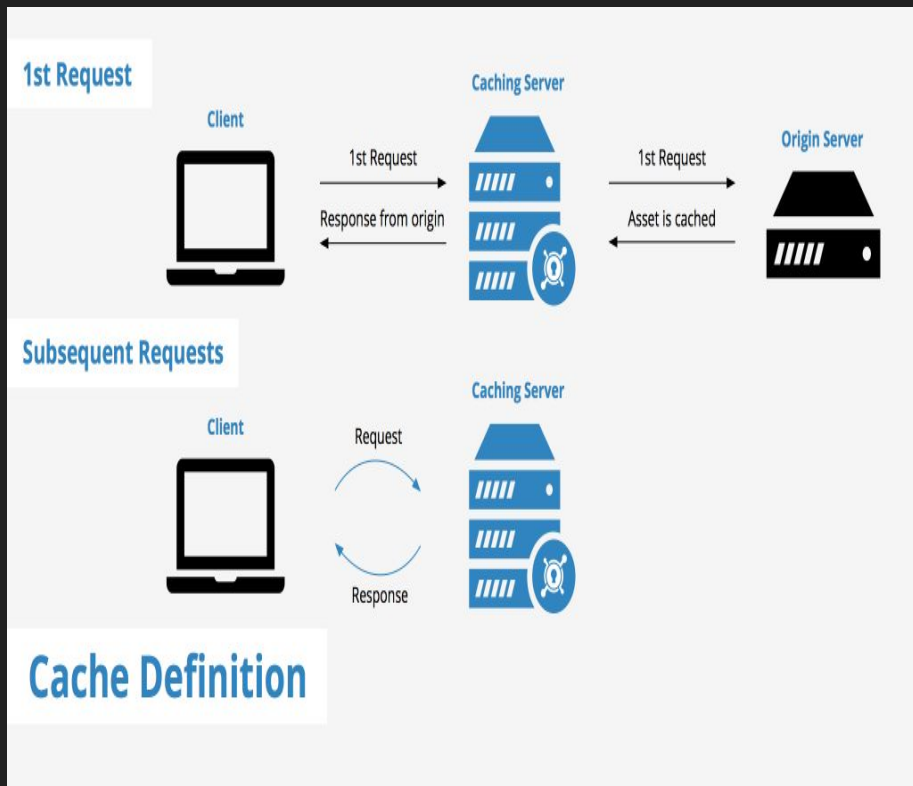
중개자이지만 클라이언트의 대리인 성격이다.

클라이언트의 모든 HTTP 요청을 받아 요청을 수정한 뒤(=필터링) 서버에 전달한다.

보안 목적으로 사용. ex) 회사 컴퓨터로 다운로드 받을 때 바이러스 검출하는 것. 증권사이트 차단하는 것.

유해콘텐츠 차단과 같은 필터링 기능을 하기도 한다.

② Cache : 많이 찾는 웹페이지를 클라이언트 가까이 보관하는 HTTP 참고

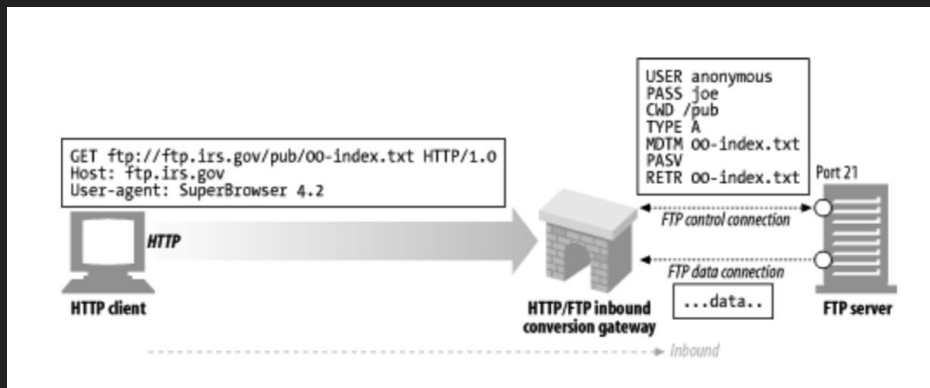


사전적 의미 : 은닉처

웹캐시와 캐시 프락시는 자주 찾는 것의 사본을 저장하는 특별한 종류의 프락시 서버이다.

Origin 서버에서 가져올 때보다 신속하고 효율적으로 동작한다.

③ Gateway : 다른 애플리케이션과 연결된 특별한 웹 서버



다른 서버들의 중개자로 동작하는 특별한 서버이다.

주로 **HTTP**트래픽을 다른 프로토콜로 변환하기 위해 사용된다.

HTTP도 알아채지 못할 만큼 진짜 **FTP** 서버인 것처럼 처리한다.

④ Tunnel : HTTP 통신을 그대로 전달하기만 하는 프락시



HTTP가 아닌 데이터를 그대로 전송해주기 위해 사용된다.

방화벽에서 웹 트래픽(HTTP데이터)만 허용하는 경우,

터널을 이용하면 통과시킬 수 있게 된다.

⑤ 에이전트 : 자동화된 HTTP 요청을 만드는 준지능적 웹클라이언트



사용자를 위해 HTTP 요청을 만들어주는 클라이언트 프로그램이다.

웹 브라우저도 에이전트에 속한다.

- 웹 로봇 : 사람과의 상호작용 없이 연속된 웹 트랜잭션들을 자동으로 수행하는 소프트웨어 프로그램(ex. 검색엔진 로봇, 주식 그래프 로봇, 가격 비교 로봇)
- 웹 로봇은 크롤러, 스파이더, 웜, 봇, 스크래퍼, 지능에이전트 등으로 불린다.