# Abilità informatiche

A.A. 2023/2024

03a - Internet e World Wide Web

### Sebastian Barzaghi

<u>sebastian.barzaghi2@unibo.it</u> <u>https://orcid.org/0000-0002-0799-1527</u>

Riassunto della lezione precedente (02b)

## Cos'è la Scienza Aperta?

Combinazione di vari movimenti e pratiche volti a rendere la conoscenza scientifica disponibile, accessibile e riutilizzabile per tutti

Creazione e condivisione di dati liberamente disponibili per l'accesso e il (ri)utilizzo, as open as possible, as closed as necessary



Melanie Imming, & Jon Tennant. (2018). Sticker open science: just science done right (ENG). Zenodo.

https://doi.org/10.5281/zenodo.128557

### Perché la Scienza Aperta (anche nelle Humanities)?



Riproducibilità e trasparenza Impatto sociale e coinvolgimento Utilità e finanziabilità

#### Forti incentivi:

- nuove prospettive di ricerca
- integrazione, pubblicazione e diffusione della ricerca
- impatto e visibilità
- allineamento alle politiche dell'UE (e quindi finanziamenti)

#### Lentezza e ostilità:

- valori e metodi contrastanti
- mancanza di finanziamenti
- problemi con copyright e licenze
- natura ibrida dei dati umanistici
- mancanza di (consapevolezza su) documentazione e standardizzazione
- mancanza di (cultura su) interoperabilità, riutilizzo e condivisione

### Principi FAIR, di nuovo



Migliorare e aumentare la condivisione dei dati di ricerca è vantaggioso

Quali regole di base dovrebbero essere date su come le persone condividono, quando e dove?

→ Sviluppo dei principi FAIR (*Findable, Accessible, Interoperable, Reusable*)

Sviluppati da FORCE 11 (un'organizzazione pan-disciplinare), questi principi forniscono una comprensione di base del valore che la condivisione dei dati può offrire e i requisiti di base per farlo



### Gestione dei dati, di nuovo

Rendere il processo di ricerca più efficiente

#### Pianificare

Capire i chi, cosa, quando, dove, come, quanto del progetto

#### Raccogliere

Quali dati? Quanti? Come li prendi? Come li organizzi?

#### Descrivere

Quali metadati? quali standard? Quale documentazione?

#### **Pubblicare**

Dove li pubblichi? Come? Sotto quali licenze?

#### Archiviare

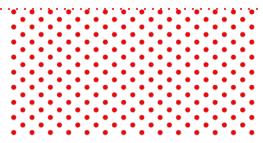
Come li conservi? Dove? Come li proteggi?



# Quiz 2



https://forms.gle/BUiK hK7no3nukuf46





# 3.1 Ipertesto

Definizione Storia Componenti Caratteristiche



Termine coniato da Nelson nel 1965, da *hyper* ("oltre") e *testo* 

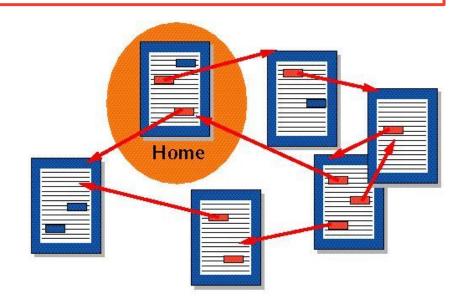
Ipotizza un sistema in grado di memorizzare documenti e consentire all'utente di costruire percorsi preferenziali e di navigare nella rete informativa costituita dall'insieme di documenti interconnessi By hypertext I mean <u>non-sequential</u> writing—text that <u>branches</u> and allows <u>choices</u> to the reader, best read at an <u>interactive</u> screen.

As popularly conceived, this is <u>a series of text chunks</u> connected by links which offer the reader different pathways.

Ted Nelson, Literary Machines (1980)



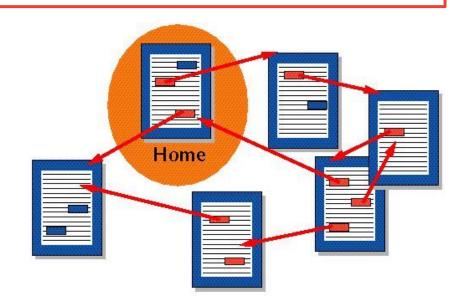
Un modo non lineare di presentazione dell'informazione tramite un insieme di unità informative (dette nodi), connesse tra loro da collegamenti (detti link)



Fonte: Robert Cailliau, CC BY-SA 3.0 <a href="https://creativecommons.org/licenses/by-sa/3.0">https://creativecommons.org/licenses/by-sa/3.0</a>, via Wikimedia Commons.

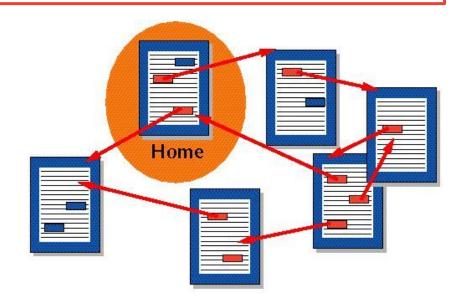
Un grafo orientato composto da

Un grafo orientato composto da un insieme di elementi (nodi) collegati tra loro da relazioni (archi) unidirezionali



Fonte: Robert Cailliau, CC BY-SA 3.0 <a href="https://creativecommons.org/licenses/by-sa/3.0">https://creativecommons.org/licenses/by-sa/3.0</a>, via Wikimedia Commons

- → diversi percorsi logici, ciascuno dotato di autonomia di significato
- → percorsi scelti a priori dal creatore del documento e a posteriori dall'utilizzatore, in base alla situazione o alle sue personali esigenze



Fonte: Robert Cailliau, CC BY-SA 3.0 <a href="https://creativecommons.org/licenses/by-sa/3.0">https://creativecommons.org/licenses/by-sa/3.0</a>, via Wikimedia Commons



# Elementi principali

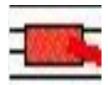
#### Nodo

unità minima di informazione, dotata di autonomia e completezza



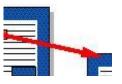
#### **Ancora**

frammento di un nodo da cui un link viene orientato verso un altro frammento di un nodo



#### Link

connessione tra una coppia di ancore



# Caratteristiche dell'ipertesto



- Granularità: decomponibile in parti più piccole e autosufficienti, dotate di senso, e utilizzabili
- Interattività: dà la possibilità al lettore di rapportarsi non-linearmente con esso
- Integrabilità: estendibile in maniera potenzialmente infinita
- Multimedialità: compresenza di molteplici mezzi di comunicazione differenti



#### Contesto storico



- Ipertestualità in quanto testualità con struttura reticolare e navigazione non lineare - è sempre esistita (es. note, indici, commenti, sperimentazioni letterarie)
- Concezione moderna (cioè digitale) di ipertesto: <u>Memex</u> ("Memory extender") di Vannevar Bush (anni '30), un dispositivo teorico impiantato in una struttura a scrivania tramite cui era possibile conservare, annotare, connettere, e leggere documenti registrati su microfilm
- Sistema di gestione collaborativa di documenti testuali chiamato <u>NLS</u> (oN-Line System) di Engelbart (1962)



#### Contesto storico

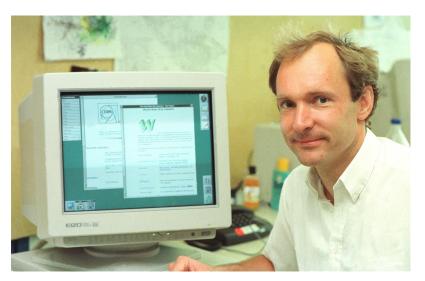


- Sistema ipertestuale <u>Xanadu</u> di Nelson (1967), una delle principali fonti di ispirazione per il World Wide Web
- Sistemi ipertestuali di prima generazione (ZOG, Xerox Alto, ecc.)
- Dopo il 1983 (anno della messa a disposizione massiva dei personal computer sul mercato): sistemi ipertestuali di seconda generazione (KMS, NoteCards, HyperCard, ecc.)
- 1987: prima conferenza internazionale sugli ipertesti



#### Contesto storico

Fine anni Ottanta: l'ipertesto è estremamente popolare e testato → un fisico trentenne col compito di progettare un sistema di gestione dei documenti dell'istituto di ricerca per cui lavorava trae ispirazione da questa tecnologia, finendo per creare il più ampio, utilizzato e influente sistema ipertestuale di tutti i tempi



Fonte: https://home.cern/science/computing/birth-web



### 3.2 World Wide Web

Definizione
Client e server
Uniform Resource Locator
Protocollo HTTP
Hypertext Markup Language
Protocollo TCP/IP
Domain Name System

### Cos'è Internet?



Una rete di comunicazione tra dispositivi distribuiti a livello globale

- TCP
- IP
- DNS
- ..



Fonte: <a href="https://rubenverborgh.github.io/WebFundamentals/">https://rubenverborgh.github.io/WebFundamentals/</a>

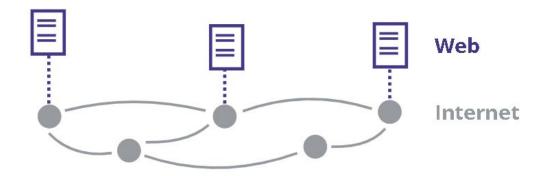
## Cos'è il World Wide Web?



Un sistema documentale ipertestuale distribuito su Internet



- HTTP
- HTML
- URL
- ...



Fonte: <a href="https://rubenverborgh.github.io/WebFundamentals/">https://rubenverborgh.github.io/WebFundamentals/</a>

### Cos'è il World Wide Web?



NON è Internet: è un servizio ospitato da Internet

Ideato da Cailliau e Berners-Lee al CERN di Ginevra verso la fine degli anni '80 come sistema di facilitazione delle comunicazioni via Internet fra ricercatori

1990: prima pagina Web

1993: rilasciato nel pubblico dominio

sistema di informazione e comunicazione globale per eccellenza dopo il 1993



## Cos'è il World Wide Web?



Il World Wide Web è costituito da un insieme di documenti ipertestuali (*pagine web*), ciascuno dei quali può essere collegato ad altri tramite link unidirezionali

Le pagine sono ospitate da nodi specializzati di Internet detti server web

Un utente può navigare nella struttura ipertestuale accedendo da un altro nodo, detto *client*, usando un apposito tipo di programma, detto *browser* 

# Terminologia di Internet

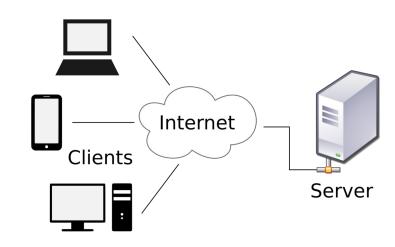
L'architettura di Internet è basata sull'interazione continua tra

#### Server

programmi (e i computer su cui operano) che *servono* dati e servizi

#### Client

programmi (e i computer su cui operano) che *fanno richiesta* di dati e servizi



Gnome-fs-client.svg: David Vignon, derivative work: Calimo, LGPL <a href="http://www.gnu.org/licenses/lgpl.html">http://www.gnu.org/licenses/lgpl.html</a>, via Wikimedia Commons

# Terminologia del Web

#### Uniform Resource Locator (URL)

schema di identificazione dei contenuti del Web e la singola stringa che agisce da identificatore univoco di una risorsa web

#### Hypertext Transfer Protocol (HTTP)

protocollo usato dai server e dai client per comunicare gli uni con gli altri e per il recupero della rappresentazione di una risorsa web tramite il suo URL

#### Hypertext Markup Language (HTML)

linguaggio di marcatura di cui sono composte le pagine web e che permette la rappresentazione di una risorsa web

### Internet + Web = ?



- rappresentare risorse tramite documenti ipertestuali ( → Hypertext Markup Language, o HTML);
- mettere a disposizione i suddetti documenti ipertestuali ( → attraverso server Web);
- identificarli mediante l'utilizzo di un opportuno identificativo ( → Uniform Resource Locator, o URL) e richiederli mediante l'utilizzo di uno specifico protocollo di comunicazione ( → Hypertext Transfer Protocol, o HTTP);
- visualizzarli su un computer ( → attraverso l'uso di un client, come un browser)

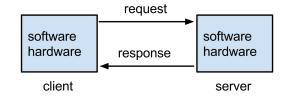
# 3.2.0 Client e Server

Definizione Elementi Esempi

#### Modello Client - Server



Architettura di rete in cui un agente computer o software - (il client) accede ai servizi o alle risorse di un altro agente - computer o software -(il server) attraverso una rete



Fonte: Lubaochuan, CC BY-SA 4.0

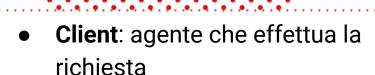
 $<\! https://creative commons.org/licenses/by-sa/4.0>, via$ 

Wikimedia Commons.

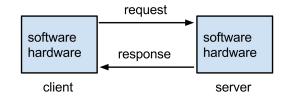
https://upload.wikimedia.org/wikipedia/commons/6/6

9/Client-server\_model.svq

### Modello Client - Server



 Server: agente che dovrebbe avere informazioni su una risorsa e che risponde alla richiesta

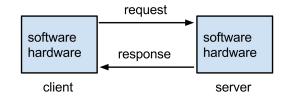


Fonte: Lubaochuan, CC BY-SA 4.0 <a href="https://creativecommons.org/licenses/by-sa/4.0">https://creativecommons.org/licenses/by-sa/4.0</a>, via Wikimedia Commons. <a href="https://upload.wikimedia.org/wikipedia/commons/6/6">https://upload.wikimedia.org/wikipedia/commons/6/6</a>

#### Modello Client - Server



- Messaggio di richiesta: il messaggio mandato dal client al server in cui vengono chieste informazioni riguardo una specifica risorsa indicata da un URL
- Messaggio di risposta: il messaggio che il server restituisce al client, che può essere sia positivo sia negativo



Fonte: Lubaochuan, CC BY-SA 4.0

<a href="https://creativecommons.org/licenses/by-sa/4.0">https://creativecommons.org/licenses/by-sa/4.0</a>, via

Wikimedia Commons.

 $\underline{https://upload.wikimedia.org/wikipedia/commons/6/6}$ 

9/Client-server\_model.svq



#### Cos'è un client?



Un computer o un software che accede a un servizio fornito da un server

Consente agli utenti di interagire facilmente con un servizio o una risorsa remota (es. sito Web, un server di posta elettronica o un sistema di archiviazione cloud)



Fonte: Raysonho @ Open Grid Scheduler / Grid Engine, Public domain, via Wikimedia Commons.

https://upload.wikimedia.org/wikipedia/commons/4/40/PCExpanion.jpg

### Cos'è un client?

Il computer che state usando

Browser (Firefox, Chrome, Edge, Opera, DuckDuckGo, Tor, ecc.)

Applicazioni Web, desktop e mobile (Netflix, Dropbox, Trello, Gmail, Excel, Canva, ecc.)

Crawler (programmi che navigano e prendono informazioni dal Web)

Sensori ed altri strumenti "incorporati" (embedded) (es. termostati intelligenti, dispositivi di domotica, smartwatch, veicoli)

#### Cos'è un server?



Computer o software che fornisce servizi o risorse ai client, attraverso una rete

Progettato per eseguire specifiche funzioni di elaborazione e gestione delle risorse

→ Es. un server Web ospita un sito Web e lo serve ai client che richiedono di accedere e visualizzare il sito



Victor Grigas, CC BY-SA 3.0 <a href="https://creativecommons.org/licenses/by-sa/3.0">https://creativecommons.org/licenses/by-sa/3.0</a>, via Wikimedia Commons.

https://upload.wikimedia.org/wikipedia/commons/6/69/ Wikimedia\_Foundation\_Servers-8055\_35.jpg

### Interazione Client - Server: esempio del browser



- → Avviata tipicamente da una richiesta dell'utente, tramite digitazione nella barra degli indirizzi del browser o per attivazione di un link
- → Il browser (client) interpreta il comando come la richiesta da parte dell'utente della pagina web corrispondente all'indirizzo specificato
- → Il client codifica la richiesta della pagina secondo determinate specifiche e inoltra la richiesta al server, sfruttando opportunamente l'infrastruttura di rete e i relativi servizi



### Interazione Client - Server: esempio del browser



- → La richiesta viene dunque instradata e raggiunge il server, che la codifica, la interpreta e cerca di soddisfarla
- ightarrow Se il server approva la richiesta del client, manda al client una risposta "200 OK", genera una copia del file corrispondente alla richiesta e la spedisce al client
- → Arrivata a destinazione, il client interpreta la copia e la visualizza per l'utente



# 3.2.1 URL

Definizione Struttura Esempi

## Cos'è un Uniform Resource Locator?

Meccanismo usato dai browser per recuperare risorse pubblicate sul Web

Nome (o più propriamente, indirizzo) della pagina web a cui un link punta e che permette di accedervi cliccandoci sopra

Una stringa di testo che *identifica* (~codice fiscale) e *localizza* (~indirizzo di casa) una risorsa web

https://it.wikipedia.org/wiki/Amore e Psiche

### Cos'è un Uniform Resource Locator?

→ Ogni volta che, da un proprio dispositivo (un computer, uno smartphone, etc.), si clicca su un link, il dispositivo stesso recupera una copia della risorsa a cui l'URL si riferisce, per poi visualizzarla sullo schermo del dispositivo

→ In teoria un URL valido punta ad un'unica risorsa, la quale può essere una pagina HTML, un documento CSS, un'immagine, ecc.; in pratica, esistono delle eccezioni, come un URL che punta ad una risorsa che non esiste più o che è stata spostata





Possiamo pensare ad un URL come se fosse una sorta di codice postale





lo **schema** (**scheme**) rappresenta il *protocollo* (l'insieme di regole) per accedere alla risorsa



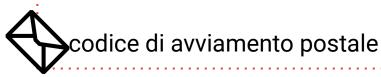


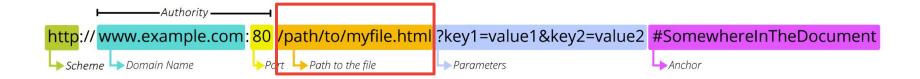
il **dominio (domain name)** identifica il tipo di entità che possiede il sito web, il servizio o la risorsa specifica a cui si accede





la porta (port) è il numero della porta (punto di contatto virtuale) del server





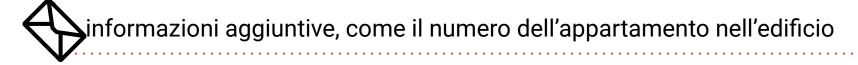
il percorso (path) rappresenta il percorso della risorsa nel server



l'edificio dove il pacco deve essere inviato

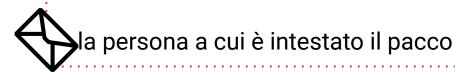


i *parametri (parameters)* rappresentano informazioni operative sulla risorsa (es. filtri in una ricerca a faccette)





l'ancora (anchor) è l'identificativo di una sezione specifica nella risorsa



# 3.2.2 HTTP

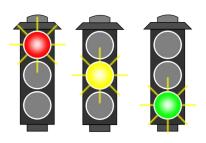
Definizione Elementi Esempi

# Cos'è un protocollo?

Per comunicare, gli elementi all'interno delle reti devono seguire **regole comuni** 

Insieme di regole e di messaggi che governano la comunicazione tra due entità

La definizione di ogni protocollo consiste nel fornire un insieme di regole non ambigue, definendo i messaggi che possono essere scambiati tra entità, il loro significato e le azioni da intraprendere in ogni situazione (~ semafori)



# Cos'è un protocollo?



Deve essere espresso in un particolare linguaggio, un insieme di segnali e/o simboli comprensibili alle entità al fine di avviare e gestire la comunicazione

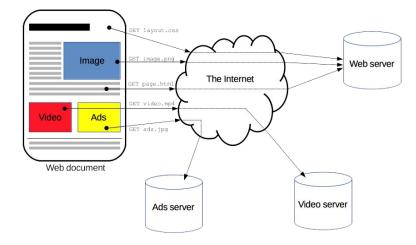
- Una sintassi da seguire per costruire i messaggi
- Delle regole interpretative del messaggio, per definire la semantica dei messaggi
- Dei meccanismi per **sincronizzare** la comunicazione
- Dei meccanismi per correggere e/o gestire eventuali errori che possono intercorrere nello scambio dei messaggi

# Cos'è il protocollo HTTP?



Per comunicare, gli elementi all'interno delle reti devono seguire regole comuni

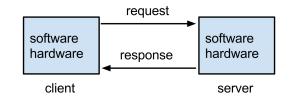
Protocollo di comunicazione utilizzato per il trasferimento di dati tra un client e un server su una rete



Fonte:

# Cos'è il protocollo HTTP?

HTTP standardizza come client e server comunicano tramite il **metodo richiesta-risposta**, in cui il client invia una **richiesta** al server e il server risponde con una **risposta** contenente i dati richiesti



Fonte: Lubaochuan, CC BY-SA 4.0

<a href="https://creativecommons.org/licenses/by-sa/4.0">https://creativecommons.org/licenses/by-sa/4.0</a>, via

Wikimedia Commons.

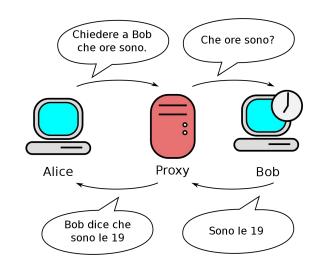
https://upload.wikimedia.org/wikipedia/commons/6/6

9/Client-server\_model.svq

### Cos'è il protocollo HTTP?



Tra il client e il server ci sono numerosi intermediari, collettivamente chiamati *proxy*, che eseguono diverse operazioni per migliorare le prestazioni e la sicurezza (gateway, cache, ecc.)



Fonte: H2g2bob, CC0, via Wikimedia Commons. https://upload.wikimedia.org/wikipedia/commons/4/4f /Proxy\_concept\_it.svg

# Messaggio HTTP



Permette lo scambio di dati tra un server e un client

Esistono due tipi di messaggi HTTP:

- **richiesta**: messaggio mandato dal client per richiedere una determinata azione da parte del server
- risposta: messaggio mandato dal server come risposta alla richiesta del client



# Messaggio HTTP



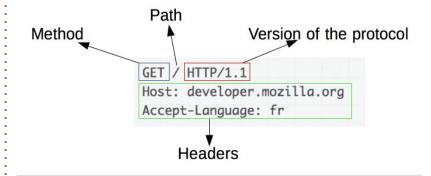
- una testa composta da:
  - una riga iniziale che descrive il messaggio
  - un insieme opzionale di *intestazioni* (o *header*) che specificano la richiesta o descrivono il corpo del messaggio
- una riga vuota che separa le informazioni dal corpo del messaggio
- un corpo (o payload) opzionale contenente dati o contenuti associati al messaggio



#### Richiesta HTTP

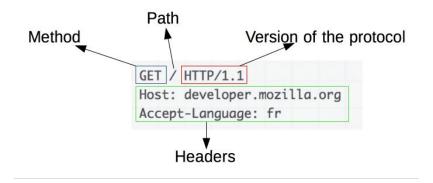


- testa:
  - riga di richiesta: metodo, URL, versione di HTTP
  - intestazioni: nome del client, lingua
- un'eventuale riga vuota
- un eventuale corpo



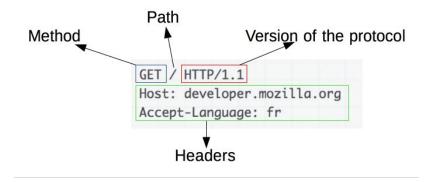
Fonte:

**GET** (il più usato nel web) permette di richiedere informazioni sulla risorsa definita dall'URL specificato – per esempio come conseguenza di un click su un collegamento ipertestuale



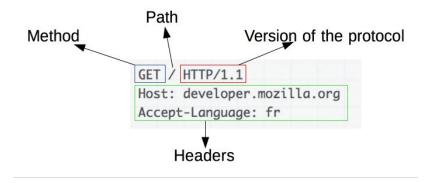
Fonte:

**POST** permette di creare sul server web contattato la risorsa specificata dall'URL utilizzato nella richiesta, e di associare informazioni incluse nella richiesta



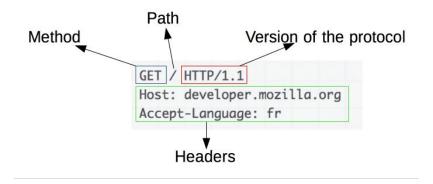
Fonte:

**DELETE** permette di rimuovere dal server web contattato tutte le informazioni relative alla risorsa specificata nella richiesta tramite l'URL



Fonte:

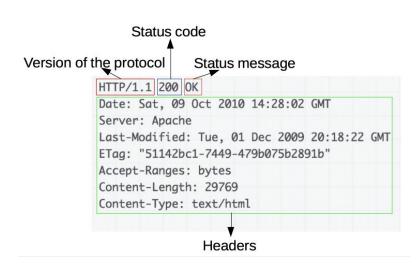
**PUT** permette di specificare informazioni aggiuntive, incluse nella richiesta, ad una risorsa esistente che già risiede sul server



Fonte:

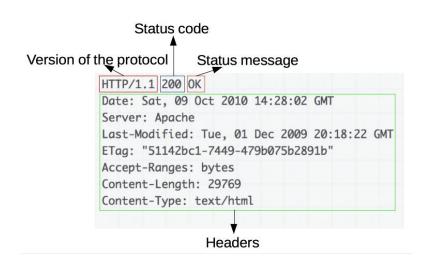
### Risposta HTTP

Tutte le risposte, oltre che a contenere metadati nell'header (e talvolta dati) nel corpo, specificano anche un codice di stato (un numero di tre cifre che specifica la classe di risposta



### Risposta HTTP

- 1XX: Risposta informativa (molto raro, di solito non usato);
- 2XX: Successo (es. 200 OK);
- 3XX: Redirezione (es. 303 See other);
- 4XX: Errore del client (es. 403 Forbidden, 404 Not Found);
- **5XX**: Errore del server (es. 503 Service Unavailable



Fonte: <a href="https://developer.mozilla.org/en-US/docs/Web/HTTP/Overvie">https://developer.mozilla.org/en-US/docs/Web/HTTP/Overvie</a> <a href="https://developer.mozilla.org/en-US/docs/Web/HTTP/Overvie">https://developer.mozilla.org/en-US/docs/Web/HTTP/Overvie</a> <a href="https://docs/web/HTTP/Overvie">w</a>

# Abilità informatiche

A.A. 2023/2024

03a - Fine

#### Sebastian Barzaghi

sebastian.barzaghi2@unibo.it
https://orcid.org/0000-0002-0799-1527