

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Informatica di base 2023/2024

Cosa succede quando si clicca su un link: I protocolli di comunicazione

Ivan Heibi

Dipartimento di Filologia Classica e Italianistica (FICLIT)

Ivan.heibi2@unibo.it

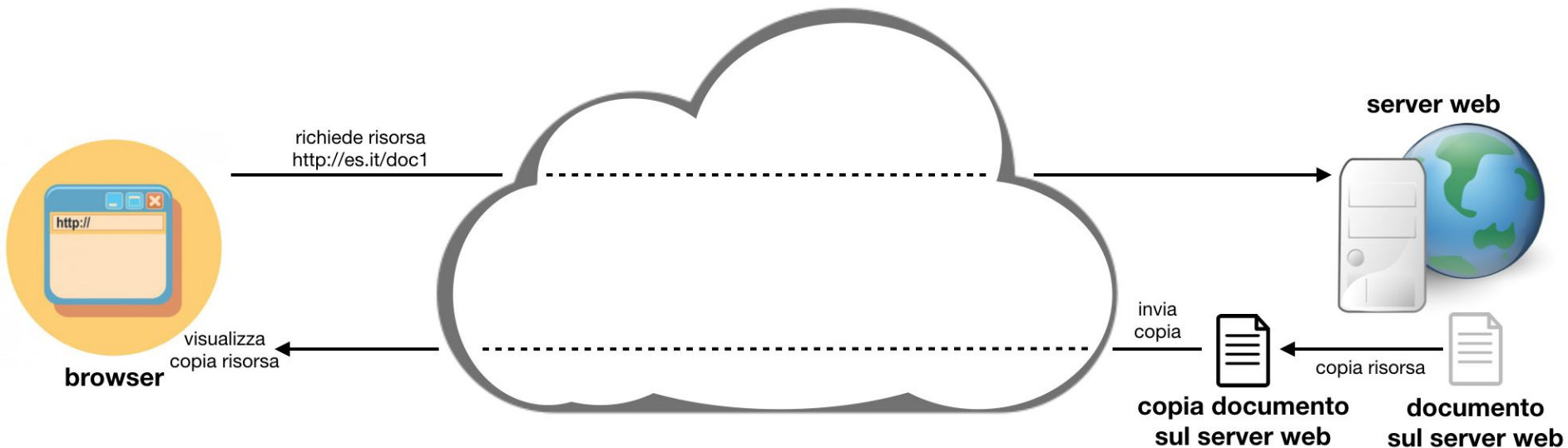
<https://www.unibo.it/sitoweb/ivan.heibi2>

Domanda

Cosa succede quando si clicca su un collegamento ipertestuale (o link) di una pagina Web?

Per rispondere pienamente a questa domanda, è necessario avere delle conoscenze di base relative sia sul come l'informazione viene trasmessa digitalmente, sia sulle componenti che compongono Internet e il Web

Riassunto di quello che abbiamo visto l'ultima volta



Protocollo di comunicazione

Un insieme di regole che due entità, parte di un sistema di comunicazione, devono seguire per scambiarsi informazioni

Cosa serve?

- Una **sintassi** da seguire per costruire i messaggi
- Delle regole interpretative del messaggio così da 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

Dove li usiamo

Nella vita di tutti i giorni quando vogliamo, per esempio, comunicare con qualcuno

Sintassi definita da regole, semantica derivata dall'apprendimento della lingua, mentre le altre componenti definite da regole sociali e/o esperienza

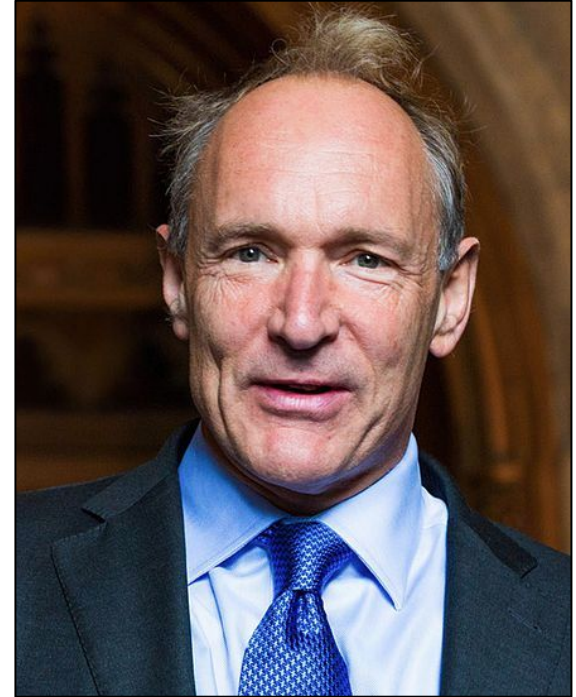


Protocollo HTTP

L'**Hypertext Transfer Protocol (HTTP)** è un protocollo di comunicazione inizialmente sviluppato da Tim Berners-Lee per facilitare l'implementazione del Web

Basato su metodi di comunicazione di **richiesta-risposta**

Esteso negli anni successivi – ad esempio con il protocollo **HTTPS**: un'estensione di HTTP che implementa un protocollo di comunicazione più sicuro sul Web, utile quando è necessario trasmettere dati di autenticazione come nomi utenti e password



Entità coinvolte

Client: l'agente che fa la richiesta, ad esempio un browser

Server: l'agente che dovrebbe avere informazioni su una risorsa, ad esempio un server Web, e che **risponde** alla richiesta

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

Come iniziare una comunicazione

Le tipologie di richieste che possono essere effettuate da un client vengono definite grazie all'uso di uno specifico **metodo**, ovvero l'operazione di richiesta relativa all'URL coinvolto nella comunicazione

Questa operazione può essere di **diversa natura**, come una semplice richiesta di informazioni, ad azioni più operative che determinano l'esecuzione di una specifica azione del server

Non tutte le operazioni a disposizione, ovviamente, sono permesse – dipendono esclusivamente dalle capacità del server e dai permessi concessi al client

Metodi HTTP

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

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

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

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

Classificazione metodi HTTP

Safe: viene usato solo per recuperare delle informazioni dal server web, **senza cambiarne lo stato**, ad esempio aggiungendo nuovi dati (ad esempio GET)

Idempotente: molteplici richieste effettuate allo stesso URL hanno lo **stesso effetto**, sul server web, che effettuare una **sola richiesta** (ad esempio GET e PUT)

GET è preponderante nel Web, visto che viene usato per richiedere una qualsiasi risorsa, mentre gli altri metodi non sempre sono liberamente utilizzabili da tutti i client – sta sempre al server decidere quali di questi metodi possano essere usati nella comunicazione

Messaggio HTTP – Header

i metadati relativi alla comunicazione – come il nome di client usato per effettuare la richiesta, il nome del server web che dovrebbe avere a disposizione la risorsa, etc.

Messaggio HTTP



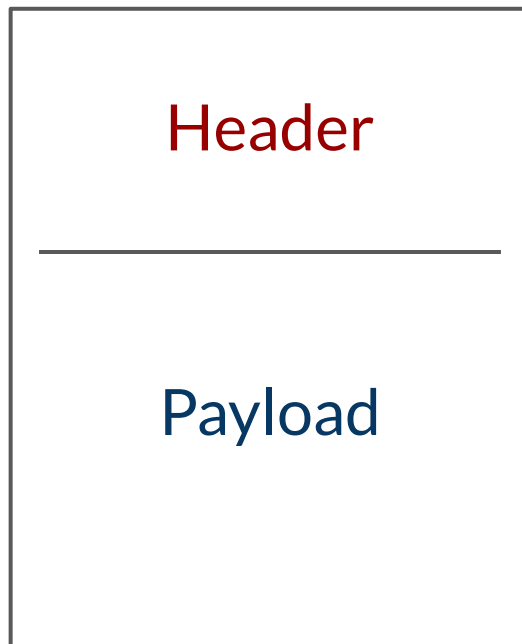
The diagram consists of a large rectangular box with a thin black border. Inside the box, the word "Header" is written in a large, dark red font. A horizontal line is drawn across the box, positioned below the word "Header", separating the header area from the rest of the message body.

Header

Messaggio HTTP – Payload

i metadati relativi alla comunicazione – come il nome di client usato per effettuare la richiesta, il nome del server web che dovrebbe avere a disposizione la risorsa, etc.

Messaggio HTTP



eventuali dati (o contenuto) da accompagnare al messaggio – può essere vuoto nel caso in cui si richiede una pagina, per esempio

Messaggi di risposta

Tutte le risposte, oltre che a contenere metadati nell'header e, talvolta, dati nel payload, specificano anche un **codice di stato** (sempre nell'header)

Un codice di stato è un **numero di tre cifre**, la cui prima (da 1 a 5) definisce la classe di risposta:

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

Server web

È un computer “speciale” che **esegue un software specifico** che permette di ricevere, gestire, e soddisfare richieste HTTP provenienti da un client, ad esempio un browser

Il **nome del server web** è incluso nell'URL che identifica la risorsa di cui si vuole ottenere informazioni, nella parte host, come introdotto nella lezione precedente precedente (ad esempio it.wikipedia.org)

La richiesta HTTP da mandare al server è molto più strutturata rispetto a quello che abbiamo visto la scorsa volta con il semplice URL



Informazioni principali di una richiesta

Quando un client (ad esempio, un browser) richiede informazioni su una specifica risorsa (ad esempio una pagina web) utilizzando l'URL relativo, ad esempio

```
http://it.wikipedia.org/wiki/Uniform_Resource_Locator
```

la richiesta che viene davvero effettuata ha la seguente forma:

Metodo	Protocollo e versione
HTTP percorso	versione
GET /wiki/Uniform_Resource_Locator	HTTP/1.1
HOST: it.wikipedia.org	
Nome del server web da contattare	

Cosa succede lato server

Una volta che il server web riceve una richiesta fatta come indicato in precedenza, **cerca localmente informazioni del documento** indicate nel percorso della richiesta

Se trova la risorsa, ne fa una **copia** che viene impacchettata nel payload in un nuovo messaggio HTTP di risposta e la invia al mittente originale, specificando “200 OK” come stato

Se non trova la risorsa, associa il **codice di errore relativo** nello stato e, opzionalmente, potrebbe anche specificare nel payload un qualche messaggio per specificare meglio l'errore o, in generale, la situazione occorsa

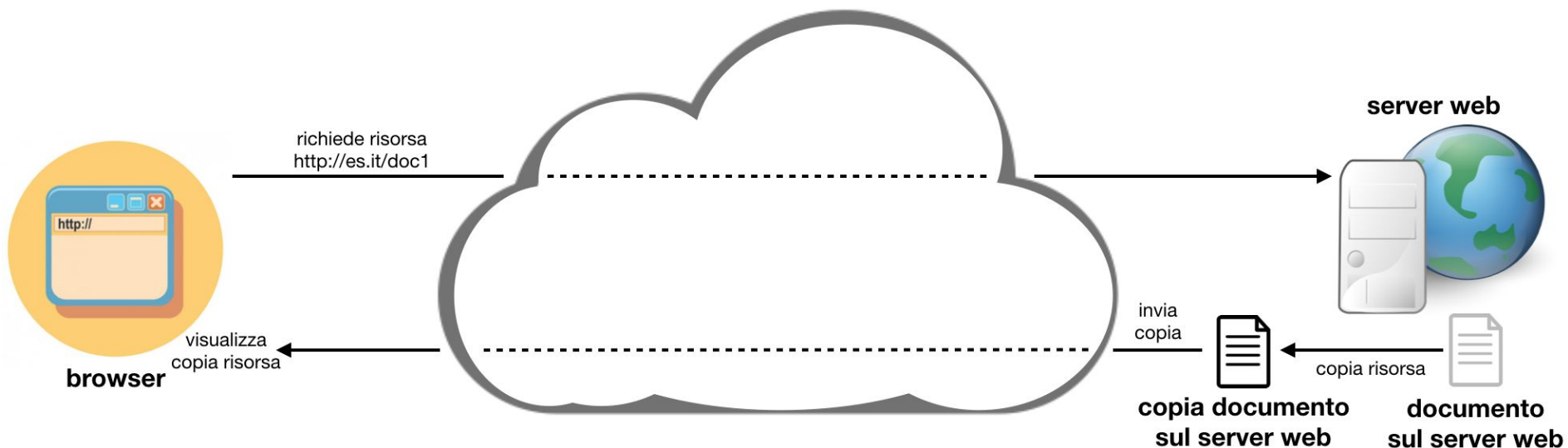
Indirizzo del server web

Seppur noi, utilizzatori di client come i browser, cerchiamo di accedere ad una risorsa a disposizione in un certo server web **usando il nome** di quest'ultimo (ad esempio `it.wikipedia.org`), in realtà il server web non è direttamente raggiungibile usando il suo nome ma attraverso un **indirizzo specifico**

Un qualunque server web è identificato (semplificando, per ora) in modo univoco da una sequenza di quattro numeri separati da punti, ad esempio `130.136.130.1`, ove ogni numero può essere un valore che va da 0 a 255

Questo numero si chiama **indirizzo IP**, dove "IP" è l'acronimo di "Internet Protocol" – che discuteremo nel dettaglio nella prossima lezione

Riassunto di quello che abbiamo visto oggi – prima



Abbiamo nuovi dettagli da integrare ...

Riassunto di quello che abbiamo visto oggi – dopo

