

tecnologie web

Matteo

January 24, 2026

1 uri

URI (uniform resource identifier): stringa di caratteri che identifica in modo univoco una risorsa su internet esistono due tipi di URI:

- URL (uniform resource locator): indica l'indirizzo di una risorsa su internet
- URN (uniform resource name): identifica una risorsa in modo univoco senza fornire il suo indirizzo

gli uRL sono piu immediati da usare ma piu soggetti a cambiamenti gli uRN sono piu stabili ma meno intuitivi

Nella visione moderna, la distinzione tra locator e name è secondaria rispetto al concetto di schemi: ogni URI appartiene ad uno schema (la parte della stringa che precede i due punti).

gli uri devono essere:

- trascrivibili
- Fornire identificazione, non interazione
- essere organizzati gerarchicamente

{URI=schema://authority/path?query#fragment}

schema = protocollo tcp usato.

l'authority è una parte specifica dell'URI che identifica chi controlla o ospita la risorsa.

La parte path è la parte identificativa della risorsa all'interno dello spazio di nomi identificato dallo schema e (se esistente) dalla authority.

query: fornisce parametri aggiuntivi per identificare o interagire con la risorsa.

fragment: identifica una parte specifica della risorsa o un punto di interesse all'interno della risorsa stessa.

i caratteri degli uri possono essere:

- caratteri riservati: hanno un significato speciale all'interno dell'URI
- caratteri non riservati: possono essere usati liberamente
- caratteri escaped: usati per rappresentare caratteri speciali o non consentiti

Una **route** è un'associazione della parte path di un URI ad una risorsa gestita o restituita da un server web.

esiste managed route: il server associa ogni URI ad una risorsa o attraverso il file system locale (risorse statiche) oppure generate attraverso una computazione (risorse dinamiche)

File-system route: il server associa la radice della parte path ad una directory del file system locale e ogni filename valido all'interno di quella directory genera un URI corretto e funzionante.

un uri assoluto contiene tutte le informazioni necessarie per localizzare una risorsa, mentre un uri relativo(uri reference) fornisce un percorso in relazione a un altro URI di base.

risoluzione degli URI relativi: il processo di conversione di un URI relativo in un URI assoluto utilizzando un URI di base come riferimento.

esistono diversi schemi

http e https di gran lunga i più usati schema file: per accedere a file locali schema data: per incorporare dati direttamente all'interno di un URI ftp: per trasferire file tra computer su una rete

2 codifica caratteri

esistenza di tante lingue diverse crea difficoltà

per rappresentare i caratteri servono regole: ordine, contiguità, raggruppamento in gruppi logici

gli **shift** servono per cambiare il set di caratteri in uso, i codici liberi sono errori di trasmissione, poi ci sono i codici di controllo.

ascii definisce 128 caratteri, usa 8 bit per rappresentare ogni carattere, il primo bit è di controllo, per questo sono 128 e non 256.

ci sono inoltre 33 caratteri di controllo non stampabili.

code page: estensioni di ASCII per supportare caratteri aggiuntivi. così vengono aggiunte le altre lingue.

negli anni 90 due commissioni diverse hanno cercato di creare uno standard universale per la codifica dei caratteri: iso/iec e unicode, ora sono stati unificati.

unicode contiene ora tutti i caratteri di tutti gli alfabeti, si cerca di essere efficienti e che ogni carattere abbia uno scopo.

inoltre quando possibile si cerca di usare caratteri preesistenti per evitare problemi di compatibilità.

iso ha 2 schemi di codifica: ucs 2 e ucs 4 ucs 2 usa 2 byte per carattere, ucs 4 usa 4 byte per carattere. visto che ucs 4 ha 4 miliardi di combinazioni, fu inventato utf-8: usa da 1 a 4 byte per carattere, e è compatibile con ascii.

i sistemi big endian e little endian gestiscono l'ordine dei byte in modo diverso. quindi per evitare ambiguità si usa il BOM(byte order mark) all'inizio del file. Zero-Width Non-Break Space (ZWNBSP), un carattere che può essere usato in qualunque contesto di whitespace (cioè ovunque tranne in mezzo alle parole) senza modificare il significato dei testi, questo indica l'ordine dei byte.

3 markup

i dati esistono come valori isolati(25), come coppie chiave valore o etichette(eta:25), e come record, o raccolte di etichette(nome:matteo, eta:25, ecc...) poi li puoi rappresentare in molti modi: alberi, tabelle, liste ecc....

Testi sono molto soggettivi La teoria del markup descrive i testi come un albero ordinato ed etichettato di elementi e nodi di testo.

HTML:Attuale formato di punta per i documenti di testo sul Web Il testo è organizzato in elementi racchiusi all'interno di tag. C'è un elenco di tag che costituisce il vocabolario HTML. Utilizza linguaggi aggiuntivi (ad esempio, CSS e Javascript per tipografia, layout e interattività sofisticati

XML piu incentrato alla struttura dei dati rispetto ad HTML Permette di definire nuovi tag e strutture di dati personalizzate.

esiste il formato binario per rappresentare i dati in modo piu efficiente, non è leggibile dall'uomo. il formato leggibile si invece.

il parsing e il processo di conversione di un documento in un albero di oggetti in memoria.

Il markup interno inserisce istruzioni di presentazione all'interno del testo, in mezzo alle parole. Il markup esterno prevede due blocchi di informazioni: il contenuto e il markup, separati e collegati da indirezione.

esistono anche diversi scopi di markup:

- presentazione: come visualizzare il testo
- puntazione: struttura logica del testo
- procedurale: istruzione per gli effetti.
- descrittivo: informazioni sui dati a livello strutturale
- referenziale: collegamenti tra parti del testo o tra testi diversi
- metamarkup: informazioni sul markup stesso

3.1 linguaggi di markup

TROFF/NROFF: uno dei primi linguaggi di markup, usato per la stampa di documenti tecnici.

TEX: complesso ma permette macro. poi vi si ricava LateX, più semplice.

linguaggi wiki: usato per creare pagine web in modo semplice.

markdown: linguaggio di markup leggero, facile da leggere e scrivere, usato per documentazione e blog.

4 semantic web

semantic web: struttura comune che consente di condividere e riutilizzare dati tra applicazioni, aziende e comunità attraverso link: permette di creare collegamenti semantic web stack: illustra l'architettura del web semantico diverse classi sono organizzate in modo taxonomico attraverso un modello chiamato RDF.

RDF(resource description framework):serve a fare affermazioni(statement) sulle risorse nella forma di triple(soggetto-predicato-oggetto)

un garfo RDF e un insieme di triple RDFle risorse sono rappresentate come nodi.
il modello a triple e semplice e minimalista. il modello RDF inoltre e modulare:

- la gestione delle informazioni puo essere parallelizzata
- informazioni parziali sono comunque valide

svantaggi:

- il modello di dati RDF e costituito da elementi di dati piccoli e frammentati, quindi un database di medie dimensioni risulta in miliardi di triple
- limitazione delle relazioni n-arie, non ci sono modi semplici di descriverli
- limitata possibilita di attribuire informazioni alle triple stesse

reificazione: prendo una tripla e gli do un identificativo per farla diventare parte di un alòtra tripla

microformati: sono ad esempio embedding di triple RDF all'interno di ambienti ospiti
importante l'aspetto della serializzazione. manca il reasoning e la generazione di nuove informazioni

con RDF si possono introdurre le inferenze attraverso altre cllassi e proprieta
il modello tabellare e il migliore che abbiamo nel complesso, ma e inefficiente per le relazioni M-N il modello ad albero ha senso per le relazioni di tipo 1-N meno mer M-N