

# Semantica dei dati

È necessario affrontare anche l'ultimo problema della comunicazione: come posso comprendere il contenuto di un messaggio?

Concordando su una sintassi e semantica dei dati comune. È ancora un problema aperto.

## Web semantico

Il Web semantico si riferisce a un'estensione del Web che promuove formati di dati comuni per facilitare lo scambio di dati significativi tra le macchine.

Quando i motori di ricerca trovano e indicizzano i contenuti del web, la maggior parte dei dati presenti nelle pagine web non sono strutturati.

Le specifiche HTML da sole non definiscono un vocabolario condiviso che ci permetta di descrivere gli elementi di una pagina e le loro relazioni in modo standard e non ambiguo.

## Linked data

I Linked Data sono un insieme di best practice per la pubblicazione e la connessione di dati strutturati sul web, in modo che le risorse web possano essere interconnesse in modo da consentire ai computer di comprendere automaticamente il tipo e i dati di ciascuna risorsa. Questo è particolarmente interessante perché qualsiasi applicazione che comprenda il tipo di risorsa può raccogliere, elaborare e aggregare dati da fonti diverse in modo uniforme, indipendentemente da dove sono stati pubblicati.

## 10.1 JSON-LD

Il JSON spesso non è immediatamente chiaro dai soli dati. JSON può essere arricchito con metadati per migliorarne la chiarezza e l'usabilità.

- **Metadati** = informazioni aggiuntive che qualificano il contenuto  
Arricchire un JSON significa
- Associare i nomi (entità) agli URI, che sono identificatori unici riconoscibili in un contesto specifico.
- Associare i valori (letterali) a tipi che si riferiscono a un contesto conosciuto
- Collegare nomi e valori attraverso relazioni (proprietà) che si riferiscono anch'esse a un contesto noto.

Un contesto noto stabilisce una conoscenza condivisa: questo può assumere la forma di un **vocabolario**, di un'**ontologia** o di uno **schema**.

## JSON-LD

JSON-LD offre un modo semplice per aumentare semanticamente i documenti JSON aggiungendo informazioni sul contesto e collegamenti ipertestuali per descrivere la semantica dei diversi elementi di un oggetto JSON. È inteso principalmente come un modo per utilizzare i Linked Data in ambienti di programmazione basati sul Web, per costruire servizi Web interoperabili e per memorizzare i Linked Data in motori di archiviazione basati su JSON. JSON-LD estende il linguaggio JSON con diverse parole chiave rappresentate dai nomi speciali delle proprietà JSON che iniziano con il segno @

Keyword	Descrizione	Esempio
@context	URL che fa riferimento a un particolare schema	<a href="http://schema.org/Person">http://schema.org/Person</a> Imposta il contesto su schema.org per indicare che i dati utilizzano il vocabolario schema.org.

Keyword	Descrizione	Esempio
@id	Identificatore univoco (solitamente è un URI)	<a href="http://dbpedia.org/page/Mahatma_Gandhi">http://dbpedia.org/page/Mahatma_Gandhi</a>
@type	Un URL che riferenzia il tipo di un valore	<a href="http://www.w3.org/2001/XMLSchema">http://www.w3.org/2001/XMLSchema</a> "@type": "Organization" Definisce il tipo di entità descritta come Organizzazione

```
{
  "@context": "https://schema.org",
  "@type": "ItemList",
  "itemListElement": [
    {
      "@type": "Organization",
      "name": "Alfreds Futterkiste",
      "address": {
        "@type": "PostalAddress",
        "City": "Berlin",
        "Country": "Germany"
      }
    },
    {
      "@type": "Organization",
      "name": "Berglunds snabbköp",
      "address": {
        "@type": "PostalAddress",
        "City": "Luleå",
        "Country": "Sweden"
      }
    },
    ...
  ]
}
```

### 10.1.1 Shared context e shared knowledge

La conoscenza può essere rappresentata tramite i seguenti formati dati:

- **Vocabulary:** Un elenco di termini (entità), ciascuno con una definizione che include i loro tipi, che diventa un insieme di URI. I vocabolari definiscono l'insieme di concetti e relazioni utilizzati per descrivere e rappresentare un'area di interesse.
- **Graph:** Una rete in cui i termini (entità) sono collegati da proprietà, ciascuna con una definizione, che diventa un altro insieme di URI. I grafici rappresentano i dati come nodi (entità) e spigoli (relazioni), formando una rete che consente strutture di dati flessibili e interconnesse.
- **Ontology:** Specifica formale di un insieme di concetti all'interno di un dominio e delle relazioni tra questi concetti. Include entità, tipi, proprietà e un insieme di regole che definiscono la conoscenza di un dominio specifico, creando un quadro strutturato per la comprensione delle relazioni all'interno dei dati. Le ontologie supportano il ragionamento sulle entità del dominio e possono essere utilizzate per dedurre nuova conoscenza.

In realtà tutti possono creare una propria struttura personalizzata per rappresentare la conoscenza. Il problema è che questa struttura può essere applicata solo ad un determinato ecosistema.

Ecosistemi diversi dovrebbero condividere la stessa rappresentazione della conoscenza per diventare interoperabile. Questo specialmente in contesti come servizi di terze parti.

Un approccio comune è quello di utilizzare le **lightweight collaborative repositories**.

Esempi: Schema.org, Wikidata

## 10.1.2 Knowledge Graphs

Perché i grafi (della conoscenza)?

- Siamo circondati da entità che sono collegate da relazioni.
- Dobbiamo memorizzarle in qualche modo, ad esempio utilizzando un DB o un grafo.
- I grafi possono essere elaborati in modo efficiente e offrono un'astrazione conveniente

## 10.2 RDF: Resource Description Framework

RDF è un modello di dati per la rappresentazione dei dati sul Web  
basato su:

- **Triplette** - unità di base per organizzare le informazioni  
Statement: <subject> <predicate> <object>
- **Grafi diretti (etichettati)** - insiemi di triple
- **URI** - identificatori univoci di risorse

Tutto è risorsa.

**Prefisso `_:`** : Quando vedi `_:` come prefisso in RDF, si riferisce a un elemento anonimo o a una variabile, che non è identificato con un URI specifico, ma è comunque utilizzato all'interno del grafo RDF per rappresentare una risorsa anonima o un valore temporaneo.

```
@prefix ex: <http://example.org/> .  
ex:alice ex:hasFriend _:b1 .  
_:b1 ex:name "Bob" .
```

In questo esempio:

- `ex:alice` e `ex:hasFriend` sono URI pienamente qualificati.
- `_:b1` è una variabile anonima utilizzata per rappresentare un nodo senza un URI definito, ma che ha una relazione con un valore ("Bob").

Caratteristiche:

### Indipendenza

- Poiché i predicati sono risorse, qualsiasi organizzazione indipendente può inventarli.

#### Interscambio

- Poiché le proprietà RDF possono essere convertite in XML, è facile scambiarle.

#### Scalabilità

- Le proprietà RDF sono semplici triple <soggetto, predicato, oggetto>, quindi sono facili da gestire e da cercare.

#### Le proprietà sono risorse

- Ciò significa che possono avere le loro proprietà e possono essere trovate e manipolate come qualsiasi altra risorsa.

#### Soggetto e oggetto possono essere risorse

- Quindi, possono avere anche delle proprietà.