



UNIVERSITÀ DI PISA

Dipartimento di Filologia, Letteratura e Linguistica

Corso di Laurea Magistrale in Informatica Umanistica

**Annotazioni semantiche di pagine Web per creare dati strutturati:  
il caso Pundit**

Relatrice:

Valentina Lenzi Bartalesi

Correlatore:

Luca De Santis

Candidata:

Silvia Cuzzo

Anno Accademico 2020/2021

# INTRODUZIONE

Le annotazioni web rappresentano il punto di arrivo nel mondo virtuale di un bisogno umano che si protrae dalla nascita della scrittura. Le annotazioni sono importanti nell'ambito delle scienze umane e sociali (SSH), la loro creazione è stata possibile grazie al Semantic Web, il cosiddetto Web 3.0. Nel web semantico non si esprimono semplicemente testi, all'interno dei quali le informazioni si trovano nascoste e richiedono un umano per la loro comprensione ma, affermazioni ovvero informazioni non ambigue, che esprimono relazioni tra oggetti mediante un linguaggio formale. Le annotazioni sono informazioni semantiche associate a risorse web. Si usano per arricchire il contenuto informativo dei documenti e per esprimere in maniera formale, il significato di una porzione di testo di un documento. Alle tecnologie del Semantic Web è infatti dedicata la prima parte della tesi al fine di introdurre con chiarezza il contesto nativo delle annotazioni [2].

Le domande a cui il mio elaborato ha cercato di rispondere riguardano diversi aspetti della tematica presa in esame. Innanzitutto, il capire quando e come sono nate le annotazioni ripercorrendo per tanto la loro storia. Ho iniziato discutendo dei *marginalia* [10], comprendendone l'utilizzo passato tramite esempi di autori di diverse epoche, riferendomi in particolar modo agli studi della Jackson [19] [20]. Successivamente mi sono interessata di scoprire quali sono le funzionalità maggiormente utilizzate e quindi richieste dal lettore moderno sui propri dispositivi elettronici come per esempio gli ebook, tablet; servendomi di studi e sondaggi condotti a tal proposito [6]. Cambia quindi il luogo dedicato alle annotazioni, non più i margini del libro cartaceo ma, note virtuali. Ciò che non muta però è la loro richiesta e l'importanza che hanno per il lettore.

Infine, ho indagato la possibilità di realizzare le annotazioni sulle pagine Web [26]. Ho testato personalmente questa funzionalità nell'ambito del mio tirocinio svolto presso l'azienda Net Seven, durante il quale mi sono dedicata a studiare come poter migliorare il tool di annotazione Pundit [12].

Pundit ha diverse funzionalità, tra cui l'annotazione, il tagging, l'evidenziazione e consente anche di realizzare annotazioni semantiche in grado di creare dati strutturati.

Per contestualizzare il mio lavoro e supportare la tesi secondo cui i tool di annotazioni sono sempre più richiesti dagli utenti, ho realizzato uno stato dell'arte in cui ho passato in rassegna i più importanti tool di annotazione attualmente disponibili, confrontandoli tra loro.

Nel capitolo finale ho descritto in dettaglio il mio contributo nel migliorare il tool di annotazione Pundit attraverso le possibilità tecnologiche offerte del Semantic Web.

In particolare, mi sono occupata di effettuare test di conformità dell'endpoint di Pundit, realizzando uno script di valutazione automatica in Python; di estendere il W3C model per quanto concerne le annotazioni semantiche, studiando le specifiche e la documentazione per individuare un modo per serializzare le triple semantiche. Infine, mi sono occupata di come esprimere le informazioni relative all'utente tramite ontologie e di come serializzare le annotazioni che includono tag.

# INDICE

<b>INTRODUZIONE</b> .....	<b>1</b>
<b>CAPITOLO 1</b> .....	<b>6</b>
<b>1.1 Il Web: definizione e nascita</b> .....	<b>6</b>
<b>1.2 L'evoluzione del Web</b> .....	<b>7</b>
<b>1.3 I concetti chiave del Semantic Web</b> .....	<b>11</b>
1.3.1 Le tecnologie del Web .....	17
1.3.2 Le tecnologie del SW .....	19
1.3.3 RDF, RDFS E OWL .....	21
<b>1.4 Linked data: identificazione e scopo</b> .....	<b>31</b>
1.4.1 "RAW DATA NOW!" .....	34
1.4.2 Dai Linked Data ai Linked Open Data.....	37
1.4.3 Le cinque stelle dei LOD .....	38
<b>1.6 Il World wide web consortium</b> .....	<b>42</b>
<b>CAPITOLO 2</b> .....	<b>45</b>
<b>2.1 Marginalia: i precursori delle annotazioni moderne</b> .....	<b>45</b>
2.1.1 Forme e motivazioni dei marginalia .....	48
2.1.2 Esempi di marginalia nel tempo .....	49
<b>2.2 Le annotazioni sui supporti elettronici</b> .....	<b>51</b>
2.2.1 Cambiamenti di lettura e d'annotazione degli e-reader .....	51
2.2.2 Le abitudini dei lettori moderni .....	53
2.2.3 Considerazioni finali sullo studio in esame.....	58
<b>2.3 Le annotazioni Web</b> .....	<b>59</b>

<b>2.4 Il W3C Web Annotation .....</b>	<b>62</b>
2.4.1 L'architettura del Web Annotation .....	64
<b>CAPITOLO 3 .....</b>	<b>68</b>
<b>3.1 Contesto nativo di Pundit: l'azienda Net7 .....</b>	<b>68</b>
<b>3.2 Pundit: pin the web .....</b>	<b>70</b>
3.2.1 I servizi offerti .....	75
<b>3.3 Architettura software di Pundit: versione iniziale .....</b>	<b>77</b>
3.3.1 Cambiamenti necessari all'architettura pregressa .....	79
<b>3.4 Architettura software di Pundit: versione attuale.....</b>	<b>82</b>
3.4.1 Il back end di Pundit: W3C Search endpoint .....	83
3.4.2 Data layer .....	87
3.4.3 Pundit App.....	89
<b>3.5 Stato dell'arte degli attuali tool di annotazione.....</b>	<b>89</b>
3.5.1 Hypothes.is .....	91
3.5.2 B2NOTE.....	95
3.5.3 Weava.....	96
3.5.4 GetLiner: Search less, learn more .....	101
3.5.5 Notesalong.....	102
3.5.6 Remarq Lite.....	103
<b>3.6 Caratteristiche a confronto.....</b>	<b>104</b>
<b>CAPITOLO 4 .....</b>	<b>112</b>
<b>4.1 Inquadramento del contesto di tesi .....</b>	<b>112</b>
<b>4.1.2 Metodologie e strumenti.....</b>	<b>113</b>

<b>4.2 Test di conformità dell'endpoint di Pundit.....</b>	<b>113</b>
<b>4.3 Estensione W3C Model: Semantic Annotation .....</b>	<b>122</b>
<b>4.4 Estensione W3C Model: Altri dati e Social Feature.....</b>	<b>127</b>
<b>4.4.1 Tagging.....</b>	<b>128</b>
<b>CAPITOLO 5 .....</b>	<b>131</b>
<b>Conclusioni.....</b>	<b>131</b>
<b>APPENDICE .....</b>	<b>134</b>
<b>BIBLIOGRAFIA.....</b>	<b>136</b>
<b>SITOGRAFIA .....</b>	<b>140</b>

# CAPITOLO 1

## 1.1 Il Web: definizione e nascita

Il termine inglese Web vuol dire “ragnatela” ed è la forma abbreviata dell’espressione World Wide Web che, a sua volta, significa “ragnatela estesa in tutto il mondo”. Per tanto, “la parola indica, in informatica, tutto l’insieme delle pagine della rete telematica mondiale collegate tra loro attraverso nodi chiamati link e visualizzabili attraverso programmi destinati allo scopo (navigare nel w.; andare in w.). Tutti i siti web sono identificati da un indirizzo, che ne permette la rintracciabilità, costituito da una sequenza di caratteri preceduta dalla stringa www, che rappresenta la sigla dell’espressione world wide web.”<sup>1</sup>

La facile reperibilità delle informazioni sul web è resa possibile grazie alla presenza ed efficacia dei motori di ricerca e dei web browser in un modello di architettura definito *client-server*.

Bisogna intenderlo come uno spazio globale di informazioni in cui sono presenti documenti in formato HTML (*Hyper Text Markup Language*) chiamati pagine Web.

La pagina Web è un ipertesto digitale identificato da un IRI (*Internationalized Resource Identifier*), contenente testo, link che rinviano a loro volta ad altre pagine web, immagini, video, audio e molto altro.

La sua nascita si deve ad un’idea concepita da Tim Berners-Lee, informatico britannico, nel 1989 con l’obiettivo di condividere e diffondere documenti all’interno del suo gruppo di lavoro al CERN in Svizzera.

Quando presentò l’idea al suo supervisore, Mike Sendall, venne definita “vaga ma interessante”, a quel punto Berners-Lee decise di implementarla da solo, costruendo tutti i tool necessari.

---

<sup>1</sup> <https://www.treccani.it/vocabolario/web/>

La data di nascita del Web si fa comunemente risalire al 6 agosto 1991 quando venne pubblicato il primo sito Web ma, solo dopo 23 giorni, un primo utente esterno al CERN poté visitare la pagina. L'intuizione di Tim Berners-Lee fu innovativa e geniale al tempo stesso, tant'è che nel 1993 il CERN decise di rilasciare il codice sorgente al pubblico dominio sebbene soltanto dal 2000 si può parlare di una vera e propria era del Web.

## **1.2 L'evoluzione del Web**

*“Suppose I could program my computer to create a space in which anything could be linked to anything.”*

Sir. Tim Berners-Lee

La funzionalità del Web è garantita da un insieme di standard volti ad assicurare l'interoperabilità su vari livelli: il protocollo di trasporto TCP/IP stabilisce la comunicazione tra i processi presenti in rete, http e HTML provvedono invece al recupero e collegamento dei documenti ipertestuali.

Inizialmente il Web è stato utilizzato principalmente da scienziati e militari, ideato come un modo per visualizzare documenti ipertestuali statici mediante il linguaggio HTML. Le sue applicazioni commerciali erano del tutto inesistenti. La fruizione del servizio era soltanto per scopi lavorativi e destinati ad una cerchia limitata di persone.

## WEB 1.0 {HTML, PORTALI }



**Figura 1.1:** Web 1.0

In seguito, grazie ad una maggiore diffusione e ad una crescita del numero dei suoi fruitori, si ha la fase del Web 1.0, caratterizzata dalla presenza di pagine statiche e dalla creazione di contenuti grazie all'utilizzo di linguaggi come Perl e Java.

Si ebbe anche l'avvento di aziende come Yahoo! e Amazon intenzionate a creare siti web e contenuti da pubblicare online.

Un aspetto importante da sottolineare in questa fase è che l'utente rappresenta semplicemente un fruitore passivo dei contenuti messi a disposizione nella Rete.



**Figura 1.2:** Web 2.0

Dagli anni Novanta in poi, si ha la seconda fase, denominata Web 2.0, caratterizzata da un cambiamento radicale derivante dall'avvento del World Wide Web, grazie al quale l'utente non è più un semplice spettatore passivo ma, inizia ad essere attivo sia a livello di creazioni di contenuti sia al livello comunicativo attraverso forum, blog e dunque creando interazioni dirette con gli altri utenti. Quindi, da un web statico si passa ad un web dinamico in cui l'utente è il protagonista indiscusso.



**Figura 1.3:** Web 3.0

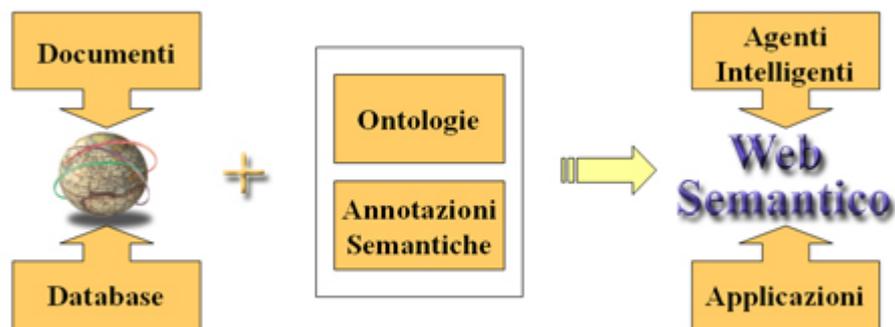
Ad oggi, si parla di Web 3.0, il cui obiettivo principale è quello di creare informazioni che mirino all'interpretabilità da parte delle macchine grazie all'utilizzo di metadati che ne specificano il contesto semantico in un formato comprensibile all'interrogazione, interpretazione e all'elaborazione automatica.

Il focus è la trasformazione del Web da *machine-representable* a *machine-understandable*, si passa da una connessione tra persone ad una connessione tra informazioni.

Il Web 3.0 non esprime semplicemente testi ma, affermazioni non ambigue che esprimono delle relazioni tra oggetti, risorse, esseri umani. Il Semantic Web non nasce con il presupposto di sostituire il web piuttosto di estenderlo con un linguaggio formale, lo scopo infatti rimane lo stesso: rendere la conoscenza disponibile a tutti, ovunque, in ogni momento e a basso costo ma con un incremento dell'automatizzazione.

Inoltre, i contenuti semantici dovrebbero cercare di mantenere alcune caratteristiche del Web, in particolare la sua immensa disponibilità di informazioni, ma al tempo stesso dovrebbero sfruttare il

potenziale della Rete al massimo, favorendo ad esempio lo scambio di informazioni tra le applicazioni, queste ultime reperiscono le informazioni e successivamente le elaborano in base al loro significato. Concettualmente, la creazione del SW si basa su due elementi fondamentali: le ontologie e i documenti annotati sul web consistenti di informazioni semantiche, in Figura 1.4 è riportato un grafico riassuntivo.<sup>2</sup>



**Figura 1.4:** Architettura del Semantic Web

### 1.3 I concetti chiave del Semantic Web

“Il semantic web è un’estensione dell’attuale Web, nella quale all’informazione viene dato un significato ben definito, permettendo così ai computer e alle persone di lavorare meglio in cooperazione”.<sup>3</sup>

---

<sup>2</sup>Lezoche, M. (2004). *Annotazione di Documenti nel Web Semantico Document annotation in the Semantic Web*, pp.196. hal-00656684

<sup>3</sup> Lee T.B., Hendler J., Lassila O. (2001), *A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, Scientific American: Feature Article: The Semantic Web

Il punto centrale del Semantic Web è che esso non si propone come un'alternativa al web ma, vuole essere una sua estensione in grado di aumentarne i benefici e le funzionalità.

Di seguito, sono riportate alcune limitazioni del web a cui il semantic web cerca di far fronte grazie ad un'organizzazione differente, successivamente esplicitata.

Uno degli usi principali del Web è sicuramente la ricerca dei documenti, la quale può avvenire in due modi differenti:

- L'uso dei motori di ricerca: l'utente digita nella barra di ricerca parole chiave relative all'argomento di interesse, bastano poche informazioni per ottenere in pochi millesimi di secondi una miriade di pagine web, questo è un chiaro vantaggio ma, ci sono alcuni svantaggi. I motori di ricerca innanzitutto non hanno accesso a tutto il contenuto del web, addirittura l'80% viene definito *hidden web* ovvero contenuto non indicizzato e quindi non reperibile.

Un'altra problematica legata a questo tipo di ricerca è la presenza sia di falsi positivi ovvero pagine web che contengono la nostra parola chiave ma, l'argomento non coincide con ciò che ricercavamo. Questa situazione è attribuibile a problemi di polisemia oppure di omonimia. Esistono poi i falsi negativi ovvero tra i risultati di ricerca non compaiono pagine web che trattano il nostro argomento. Questo accade perché la parola chiave da noi inserita non è presente in queste pagine, ma magari vi è un suo sinonimo, e per tanto il motore di ricerca non le restituisce come risultati;

- Un'altra metodologia di ricerca è quella ottenuta attraverso link ipertestuali. Se da una parte è facile intuire il contenuto della pagina a cui il link rimanda grazie al contesto oppure grazie alla frase o alla parola cliccabile, dall'altra il problema si pone nel momento in cui ci ritroviamo in una sorta di labirinto di link che rinviano ad altrettanti link, a quel punto ritrovare "il filo di Arianna" per poterne uscire o individuare l'informazione che vogliamo, può risultare complesso oltre che dispendioso da un punto di vista di tempo.

Un limite strettamente collegato all'azione precedente è che la ricerca restituisce documenti attraverso cui bisogna cercare ulteriormente ciò che vogliamo, manualmente dato che non viene restituita l'informazione in maniera diretta.

Inoltre, vi è una mancanza di integrazione di informazioni perché ciò che cerchiamo potrebbe presupporre la consultazione di più pagine web, non collegate tra loro.

A queste problematiche, il Semantic Web cerca di porre rimedio riconfigurando e ristrutturando i dati sul Web in una prospettiva tale per cui il loro significato sia comprensibile agli utenti ma anche ai programmi che li utilizzano.

Il focus è la diversa rappresentazione della conoscenza, l'aggettivo "semantico" infatti vuole sottolineare il voler dare un significato ai dati che saranno gestiti dai programmi, come viene espressamente definito nell'articolo "Semantic Web" di Tim Berners-Lee, Hendler e Lassila: "*The Semantic Web will enable machines to comprehend semantic documents and data, not human speech and writings.*"

Il primo step è quello di aggiungere informazioni che hanno lo scopo di esplicitare il significato dei contenuti, ciò è garantito grazie a documenti strutturati costituiti dai cosiddetti metadati, letteralmente "dati che parlano dei dati", che adeguati all'elaborazione automatica, restituiscono informazioni sui dati stessi tramite cui è possibile fare inferenze, quindi ricavare ulteriori informazioni a partire dai dati iniziali. Il termine inferenza, in logica, indica il processo attraverso cui una proposizione viene derivata, stabilendo la sua veridicità a partire da una proposizione iniziale assunta vera.

Il passaggio fondamentale è il seguente: da un **web di documenti** costituente una rete i cui nodi sono i documenti stessi collegati da link, si passa ad un **web dei dati**, in cui questi ultimi diventano nodi e i link stabiliscono le relazioni tra loro.

Per rendere tutto ciò concreto, sono stati utilizzati dei linguaggi di annotazione costruiti a partire da XML: RDF (*Resource Description Framework*), RDFS e OWL.

Una nozione importante all'interno dell'architettura del Semantic Web è quella di risorsa indicante tutto ciò che può essere definito da un'IRI.

Una risorsa è descrivibile tramite i cosiddetti vocabolari controllati che, assicurano un certo livello di controllo, standardizzazione e interoperabilità nel rappresentare la conoscenza. Il loro scopo è quello di organizzare l'informazione ma, anche di fornire una terminologia adatta per la categorizzazione.

Fondamentale sottolineare come i vocabolari controllati permettano l'uso soltanto di termini predefiniti e autorizzati dai designers, a differenza dei vocabolari che racchiudono il lessico dei linguaggi naturali, in cui non vige nessuna restrizione di tale tipo.

Esistono diversi tipi di vocabolari controllati per esprimere la conoscenza contenuta nei dati:

- **Tesauro:** vocabolario contenente sinonimi e contrari, si differenzia dal dizionario perché non definisce il significato delle parole e non le organizza in maniera alfabetica;
- **Tassonomia:** prevedono una classificazione dei concetti, può avere anche delle relazioni gerarchiche al suo interno;
- **KOS (Knowledge Organization System):** set di elementi utilizzati per descrivere, indicizzare oggetti e browsing collections. Sono una tipologia di vocabolari utilizzati soprattutto in aree scientifiche. Un esempio è SKOS, un vocabolario per condividere e collegare la conoscenza sul Web: molti KOS condividono una struttura simile e per tanto è utilizzata in diverse applicazioni, SKOS preleva la similarità e la rende esplicita in modo da far sì che i dati vengano utilizzati da applicazioni diverse;
- **Ontologia:** è un framework formale per rappresentare la conoscenza, codificata in modo tale da essere compresa da un computer. Definisce classi e relazioni di un dominio della realtà.

Si può definire come un' esplicita specificazione di una concettualizzazione: un'esplicita specificazione in quanto provvede a dare un **nome** per rappresentare le risorse; una **sintassi** per rappresentare le relazioni tra loro; una **semantica** per la conoscenza implicita. Per concettualizzazione invece si intende una visione semplificata e astratta del mondo.

Un'ontologia contiene: classi e proprietà che, rappresentano in forma astratta gli oggetti e le relazioni vigenti tra loro nonché un insieme di possibili relazioni che combinano classi e proprietà, ad esempio “is subclass of” crea una tassonomia della classe in questione; “subproperty of” crea, invece, una tassonomia della proprietà.

Quindi, un'ontologia definisce le classi e le proprietà le cui istanze occorrono in un network semantico, inoltre essa all'interno di una Knowledge Base rappresenta la TBox ovvero la sezione contenente classi, proprietà e assiomi; invece, per ABox ovvero l'Assertional Box intendiamo quella che contiene individui e asserzioni.

Le ontologie possono essere create in due modi:

- Manualmente: questo tipo di operazione richiede grande conoscenza e anche molto tempo per realizzarla;
- Semi Automatico: questo approccio permette di generare i concetti e le loro relazioni in un certo dominio in maniera automatica. Questo tipo di approccio può essere classificato in tre modi: **supervisionato** per cui c'è bisogno di una grande quantità di esempi di training; **metodi di NLP** che si fondano su regole che analizzano i pattern basati su categorie sintattiche; **metodi di clustering sintattico** attraverso cui il dataset viene diviso secondo un criterio e ad ogni cluster sono associati etichette, definendo in tal modo nuove tassonomie.

I limiti dell'approccio semi automatico sono diversi: la dipendenza dal linguaggio, dallo stile, inoltre la qualità delle ontologie risulta essere peggiore rispetto a quelle create da esperti della lingua.

Al di là della modalità di creazione delle ontologie, il loro contributo è importante per disambiguare il contesto grazie all'utilizzo di termini opportuni.

Ci sono stati due approcci diversi nei confronti delle ontologie che, hanno avuto una ripercussione anche nella visione riguardante il web semantico.

La prima visione concerne un approccio di tipo top-down, conosciuta anche come fase dell'**ontologia fondazionale**, la cui idea soggiacente era che ci fosse la possibilità di rappresentare la conoscenza a priori in maniera universale. Questo approccio ha inevitabilmente dei limiti legati alla natura mutevole e arbitraria del linguaggio naturale, in particolare: “nella difficoltà di raggiungere uno standard ontologico riconosciuto universalmente e definito a priori poiché il Web è un ambiente dinamico e in costante evoluzione e perché risulta difficile esprimere in un linguaggio formale la semantica naturale soggetta a diverse convenzioni sociali e individuali”.<sup>4</sup>

La seconda visione, proposta sempre da Tim Berners-Lee, è di tipo bottom up ovvero dal basso verso l'alto, puntando ad un coinvolgimento degli utenti e delle communities per la creazione delle ontologie. Non un approccio a priori ma, focalizzato sull'uso degli utenti quindi considerando la mutazione del linguaggio nel corso del tempo; inoltre, il padre del Web attribuisce un'importanza particolare ai linked data, modalità di pubblicazione di dati strutturati per collegarli tra loro (concetto che sarà approfondito nei prossimi paragrafi). Essi sono l'approdo ad un tipo di informazione flessibile, avvantaggiando il passaggio auspicato da un **web di documenti** ad un **web di dati** connessi tra loro.

---

<sup>4</sup>Castro, V. L. (2014). Web Semantico e Linked Open Data: best practices, prospettive e criticità. *Nuovi annali della Scuola speciale per archivisti e bibliotecari*, 28, 207.

### 1.3.1 Le tecnologie del Web

Il web è costituito sia da documenti in formato HTML sia da contenuti multimediali “il Web è uno dei principali servizi di Internet, che permette di navigare e usufruire di un insieme molto vasto di contenuti amatoriali e professionali (multimediali e non) collegati tra loro attraverso collegamenti (link), e di ulteriori servizi accessibili a tutti o ad una parte selezionata degli utenti di Internet”. gli oggetti da cui è costituito sono interpretabili dagli utenti, non dalle macchine per tanto non sono strutturati affinché queste ultime ne comprendano il significato ma, devono soltanto garantire la loro visualizzazione.

Esso ha ereditato i principi fondamentali di Internet che, sono tre: l'**interoperabilità** che, consiste nella compatibilità reciproca tra le specifiche dei linguaggi e dei protocolli; l'**evoluzione** garantita dai principi di progettazione come semplicità, modularità ed estensibilità ai fini di garantire un'apertura e un'efficienza del web nei confronti delle tecnologie emergenti come i diversi tipi di supporti: dispositivi mobile; tablet, smart tv. Infine, eredita il principio della **decentralizzazione**, l'architettura di internet non dovrebbe dipendere da nodi centrali, ma è di difficile attuazione sebbene soltanto in questo modo il Web potrebbe diffondersi su scala globale senza problematiche.

Le componenti principali del web possono essere riassunte come di seguito:

- HTML, *HyperText Markup Language*: è un linguaggio di marcatura per creare le pagine web, i tag vengono utilizzati per dare informazioni riguardanti la visualizzazione dei documenti;
- Link non tipizzati, ovvero collegamenti ipertestuali che demandano alle persone il compito di comprendere il significato delle relazioni;
- Web-API, *Application programming interface*: sono interfacce grafiche che permettono ai programmatori di sviluppare funzionalità aggiuntive o integrazioni con altre app;

- Database: è un contenitore di dati archiviati elettronicamente, costituiscono delle informazioni su una porzione del mondo reale. La modalità di rappresentazione dei dati, solitamente, è quella di una tabella contenente righe e colonne, questa struttura permette una maggiore efficienza di interrogazione attraverso le queries.

I dati possono essere visualizzati, gestiti, modificati mediante il linguaggio SQL (*Structured Query Language*).

L'archiviazione dei dati sottoforma tabellare non è una novità, basti pensare ai fogli di calcolo ma, ci sono delle differenze sostanziali tra le due tipologie: i fogli di calcolo sono stati pensati per un unico utente, per la manipolazione dei dati non è richiesto l'utilizzo di procedure complesse ma, cosa fondamentale la quantità delle informazioni è diversa. I database gestiscono grandissime quantità di dati, inoltre non sono predisposti all'utilizzo di un solo utente piuttosto di molteplici anche contemporaneamente che, eseguono queries mediante linguaggi di programmazioni complessi.

Sono stati progettati diversi tipi di database in funzione dello scopo finale: **database relazionali** in cui i dati sono rappresentati in formato tabellare, questo tipo di struttura è una delle più efficienti per accedere alle informazioni strutturate; i **database orientati agli oggetti**; i **database grafici** presentano i dati in termini di entità e di relazioni tra le stesse. Inoltre, ci sono database recenti nati grazie all'utilizzo dell'automazione come i **database Open Source**; i **database cloud** contengono dati strutturati e contenuti in una piattaforma cloud che, può essere pubblica, privata o ibrida. Un ultimo esempio, davvero interessante, è il **database** con funzionalità di **Self-Driving** ovvero autonomo, nel senso che utilizza tecniche di machine learning per automatizzare sicurezza, backup, aggiornamento e altre funzioni di routine.

Infine, indipendentemente dalla tipologia, oggi i database rappresentano per le aziende un modo per migliorare il processo decisionale e le stesse performance aziendali perché riescono a superare i tradizionali metodi di archiviazione, puntano a maneggiare enormi quantità di dati provenienti da molteplici sistemi, tutto ciò per una migliore gestione ed esecuzione delle attività.

### 1.3.2 Le tecnologie del SW

Come anticipato in precedenza, i contenuti del Web hanno un limite principale: la mancata possibilità di interpretazione da parte delle macchine, ciò viene superato, fornendo significato ai dati, dunque utilizzando le tecnologie del Semantic Web.

La strutturazione dell'informazione e la possibilità di interoperabilità tecnologica tra le applicazioni Web è garantita da XML, *Extensible Markup Language*. È un metalinguaggio di marcatura che non solo assegna regole sintattiche utili per la gestione della struttura dei documenti ma, fornisce anche una semantica ai tag, supportando le transazioni sul Web e gestendo lo scambio di informazioni tra database differenti. Il suo utilizzo semplifica la gestione di insiemi di documenti e la loro pubblicazione al livello internazionale, con il vantaggio di non essere dipendente né da una singola piattaforma né da un singolo linguaggio.

La sintassi XML utilizza tag di inizio e fine per marcare il campo informativo, il quale è denominato elemento , ad esempio `<importo>` e `</importo>`. L'elemento può essere arricchito da ulteriori informazioni come coppie nome/valore dette attributi, ad esempio `id = "ord001"`. È una sintassi molto semplice e intuitiva ma, è fondamentale che i tag siano inseriti correttamente, creando una corrispondenza tra tag di apertura e chiusura, inoltre possono esserci campi informativi nulli.

Un documento XML è decretato ben costituito, *well formed*, quando le regole di scrittura sono rispettate; invece, validato quando la sua struttura è coerente con quella definita nella DTD, *Document Type Definition*, la quale consente di definire in maniera non ambigua la struttura dei dati. Il codice DTD può essere presente sia all'interno del file XML oppure in un file con estensione .dtd collegato all'XML.

```

01 <?xml version="1.0"?>
02 <!DOCTYPE ordine [
03 <!ELEMENT ordine ( cliente, prodotto+ )>
04 <!ATTLIST ordine id ID #REQUIRED>
05 <!ELEMENT cliente EMPTY>
06 <!ATTLIST cliente db CDATA #REQUIRED>
07 <!ELEMENT prodotto ( importo )>
08 <!ATTLIST prodotto db CDATA #REQUIRED>
09 <!ELEMENT importo ( #PCDATA )>
10 ]>
11 <ordine id="ord001">
12   <cliente db="codcli123"/>
13   <prodotto db="prod345">
14     <importo>23.45</importo>
15   </prodotto>
16 </ordine>

```

**Figura 1.5:** Esempio di documento XML con DTD

La DTD ha diverse limitazioni derivanti dall'aver una sua sintassi, richiedendo quindi un suo editor, parser e processor, inoltre non possiede *datatype* ed è difficile estenderla. Da qui, la nascita di XML Schema Definition che, possiede le stesse funzioni di DTD ma, ha innumerevoli vantaggi come l'aggiungere informazioni semantiche, il rappresentare vincoli sui possibili valori, tipi complessi e relative gerarchie.

Un altro vantaggio dell'XML, questa volta rispetto all'HTML, è quello di fornire la possibilità di definire propri elementi di markup da parte dell'utente, per questo si parla di estensibilità, inoltre è proiettato verso la struttura del testo e verso il contenuto dei suoi elementi piuttosto che alla formattazione e visualizzazione.

La libertà derivante dal poter creare nuovi tag all'interno di un documento XML ha reso necessario uno stratagemma per identificare l'appartenenza dei tag ad uno stesso schema, per evitare problemi di omonimia: ad esempio, titolo potrebbe indicare sia il titolo di un libro ma anche un titolo professionale di una persona. Da qui, lo sviluppo dell'XML Namespaces che permette di utilizzare nomi già definiti in altri schemi in modo da evitare ambiguità e collocare il tag nella sfera semantica adeguata. I namespaces, infatti, sono nomi identificati da un unico URI, "se in un documento si decide di riutilizzare l'elemento Dublin Core, Creator: <http://purl.org/dc/term>, quando un elemento sarà

qualificato come creator esso dovrà essere interpretato secondo la definizione data nello schema di Dublin Core.”<sup>5</sup>

Infine, è importante sottolineare come ad oggi, XML ricopra un ruolo fondamentale nell’architettura del web ed è utilizzato per ogni nuovo linguaggio che voglia definire uno standard.

L’interoperabilità semantica è garantita grazie all’introduzione dei metadati, letteralmente “dati che parlano di sé” ovvero sono informazioni riguardanti una risorsa web identificata univocamente da un URI, *Uniform Resource Identifier*. Il punto focale è che i metadati permettono il passaggio da un web *machine-readable* ad uno *machine-understandable*; inoltre, in quanto dati possono essere memorizzati come tali all’interno di una risorsa, quest’ultima dunque può avere informazioni relative sia a se stessa che ad un’altra risorsa; inoltre possono anche essere descritti da altri metadati.

L’uso dei metadati necessita di convenzioni per la **semantica** ovvero per la descrizione delle risorse; la **sintassi** ovvero l’organizzazione sistematica degli elementi facilitando lo scambio tra applicazioni diverse e infine la **struttura** intesa come un vincolo sulla sintassi per una rappresentazione soddisfacente della semantica.

### 1.3.3 RDF, RDFS E OWL

I linguaggi standard del Semantic Web sono RDF, RDFS e OWL: **RDF**, *Resource Description Framework*, è un linguaggio, sviluppato dal W3C, utilizzato per esprimere informazioni riguardo alle risorse del Web, il suo scopo è quello di rendere le informazioni processabili dalle applicazioni ma, anche di permettere che ci sia uno scambio di informazioni tra le applicazioni senza perdita di significato.

La specifica dell’RDF consiste di due componenti: l’RDF Model and Syntax e l’RDF Schema.

---

<sup>5</sup>Fusetti, C. (2017). *Dati bibliografici e linked (open) data: prospettive e criticità* (Bachelor's thesis, Università Ca'Foscari Venezia).

Il primo fornisce un modello costituito da risorse, proprietà e valori. Qualunque cosa descritta dall’RDF è una risorsa, la quale può essere una pagina Web, una sua parte, una collezione di pagine web ma, anche la rappresentazione di un oggetto non reperibile su internet come un libro, un dipinto. Una proprietà è identificata da un nome e serve per esprimere relazioni che legano le risorse ai valori, questi ultimi possono essere sia a loro volta delle risorse sia un dato semplice come un literal. Ogni proprietà definisce un range di valori ammissibili, tipologie di risorse descrivibili e le relazioni con altre proprietà.

Questo modello si basa sull’idea degli statement nella forma di soggetto – predicato – oggetto, queste espressioni sono definite triple: il soggetto denota la risorsa, il predicato esplica un aspetto della risorsa ed esprime una relazione tra il soggetto e l’oggetto della tripla.

Esistono due tipologie di triple: le ground triple in cui il soggetto e il predicato sono rappresentati da un IRI, l’oggetto invece da un IRI o un letterale.

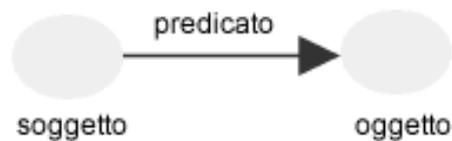
Le non-ground triple si differenziano dalle precedenti poiché il predicato è sempre rappresentato da un IRI, diversamente dal soggetto che può essere un IRI o un blank node; l’oggetto un IRI, un literal o un blank node.

I blank nodes sono degli oggetti astratti che, non identificano una risorsa specifica, quando sono presenti negli statement indicano l’esistenza di qualcosa senza esprimerla. Tutto ciò rende la conoscenza vaga e inevitabilmente ciò porta a costi computazionali elevati.

Gli statement sulle risorse possono essere rappresentati mediante un grafo orientato composto da nodi ed archi: le risorse sono rappresentate dai nodi, invece i predicati sono gli archi. Il grafo può essere considerato il modello concettuale dell’RDF.<sup>6</sup>

---

<sup>6</sup> Bartalesi Lenzi Valentina(2020/2021), *Corso universitario di Semantic Web*, Università di Pisa



**Figura 1.6:** Grafo orientato

RDF Model ha diversi formati che lo serializzano, ovvero diversi formati di file, per cui il modo in cui una tripla è codificata dipende dal formato. Esistono diversi tipi di serializzazione:

- XML/RDF: una sintassi basata sull'XML, primo formato standard per la serializzazione RDF;<sup>7</sup>
- Turtle: un formato *machine-readable* ma, anche *human friendly*;<sup>8</sup>
- N-Triples: un formato per memorizzare e trasmettere dati ed è un sottoinsieme di Turtle.

**L'RDF Schema**, *Resource Description Framework Schema*<sup>9</sup>, è un'estensione semantica dell'RDF, il cui obiettivo è quello di arricchire il Data model per una migliore rappresentazione della conoscenza; infatti, il primo non permette la validazione di un valore o la restrizione di un dominio di una proprietà. Una sua peculiarità rispetto sia all'XML Schema e DTD è che non vincola la struttura del documento, fornisce invece informazioni per la comprensione del documento stesso. Esso definisce un insieme di risorse RDF da usare per descrivere caratteristiche di altre proprietà e risorse RDF.

---

<sup>7</sup> Klein, M. (2001). XML, RDF, and relatives. *IEEE Intelligent Systems*, 16(2), 26-28.

<sup>8</sup> Beckett, D., Berners-Lee, T., Prud'hommeaux, E., & Carothers, G. (2014). RDF 1.1 Turtle. *World Wide Web Consortium*, 18-31.

<sup>9</sup> Brickley, D., Guha, R. V., & McBride, B. (2014). RDF Schema 1.1. *W3C recommendation*, 25, 2004-2014.

La sua prima versione è stata presentata nel 1998 dal W3C, nel 2004 è diventato una raccomandazione prendendo il nome di RDF Vocabulary Description Language 1.0.

L’RDF Data Model permette di esprimere dichiarazioni sulle risorse attraverso proprietà e valori ma, è una rappresentazione limitante; da qui la necessità di creare ontologie definendo: classi e gerarchie delle stesse; proprietà con relative gerarchie; domini e range delle proprietà.

L’RDFS non fornisce un vocabolario di classi ad hoc per le applicazioni piuttosto i mezzi necessari per descriverle e per indicare quali classi e proprietà possano essere utilizzate insieme, si può affermare che l’RDFS assegna un sistema di tipi per l’RDF. Gli elementi dell’RDFS sono forniti sotto forma di un vocabolario RDF ovvero come un insieme di risorse RDF predefinite con il loro specifico significato. Le risorse sono associate ad un URI di riferimento, ad esempio “<http://www.w3.org/2000/01/rdf-schema#>, associato al prefisso QName rdfs.

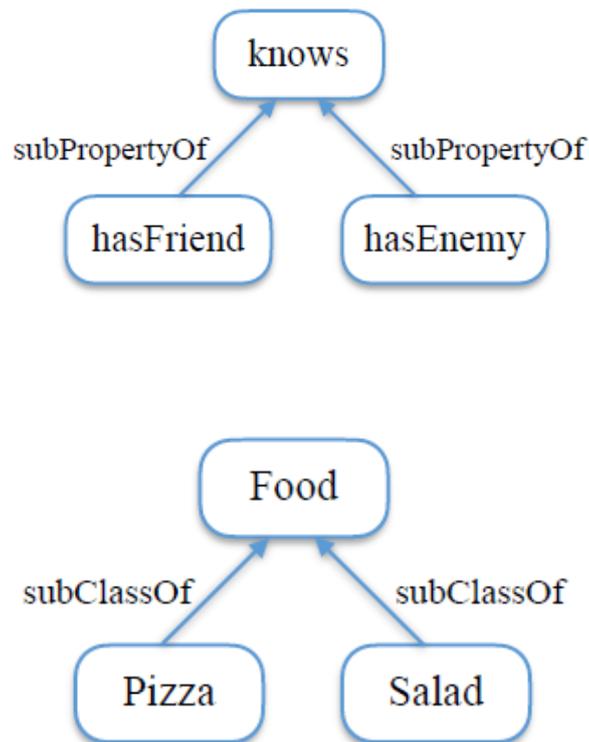
Le **classi**, le **proprietà** e gli **individui** rappresentano tipologie di risorse diverse: le prime rappresentano un insieme di risorse associando un concetto generico come Place, Person, Organization, quindi ad esempio le istanze formanti la classe Place, dove per istanze si fa riferimento alle risorse, rappresentano tutti i posti che possiamo rappresentare nel mondo. Gli individui, invece, sono risorse specifiche come un singolo individuo, luogo (Paola, Pisa...).

Il legame tra i due viene creato nel momento in cui utilizziamo la proprietà rdfs:type, enunciando dunque che un individuo è un’ istanza di una o più classi.

Le **proprietà** servono per connettere due risorse, ad esempio la proprietà :studiesAt permette di connettere un individuo della classe :Person con un individuo della classe :Organization. Inoltre, in riferimento alle proprietà parliamo di rdfs:domain che, rappresenta la classe di risorse che possono essere utilizzate come soggetto della proprietà, ad esempio scrivere “:studiesAt rdfs:domain :Person”; afferma che le istanze della classe Person possono essere il soggetto della proprietà :studiesAt.

Rdfs:range, invece, specifica quali risorse possono essere utilizzate come oggetto della proprietà, riprendendo l’esempio precedente :studiesAt rdfs:range :Organization, in questo caso si afferma che l’oggetto della proprietà “studiare presso” può essere un’istanza della classe Organization.

Sia le classi sia le proprietà possono avere delle gerarchie, rispettivamente `rdfs:subClassOf` e `rdfs:subproperties`.



**Figura 1.7:** Esempi grafici di `:subPropertyOf`, `:subClassOf`

Di seguito, alcuni esempi di costrutti dell'RDFS:

**rdfs:Class**, indica la classe delle risorse che sono classi RDF, permette di dichiarare una risorsa come una classe per altre risorse. Un esempio di `rdfs:Class` è `foaf:Person` nel vocabolario FOAF.

**rdf:type** è un'istanza di `rdf:Property` ed è usata per dichiarare che una risorsa è un'istanza di una classe. Una tripla del tipo “X `rdf:type` C” dichiara che C è un'istanza di `rdfs:Class` e X è un'istanza di C.

**rdfs:comment** è un'istanza di `rdf:Property` che può essere utilizzata per dare una descrizione di una risorsa in formato leggibile dall'essere umano, in modo da capire significato di classi e proprietà.

**rdfs:Literal** esplica la classe dei valori di tipo letterale come stringhe e interi. Essa è un'istanza di `rdfs:Class` e una sottoclasse di `rdfs:Resource`.

**OWL**, *Ontology Web Language*, è un linguaggio di markup, il cui sviluppo è stato voluto per migliorare e accrescere l'espressività di RDF e RDF Schema, aggiungendo ad esempio proprietà come la transitività, la riflessività, definizione di assiomi di cardinalità sulle classi, il punto importante è il bilanciamento tra espressività e decidibilità del linguaggio.

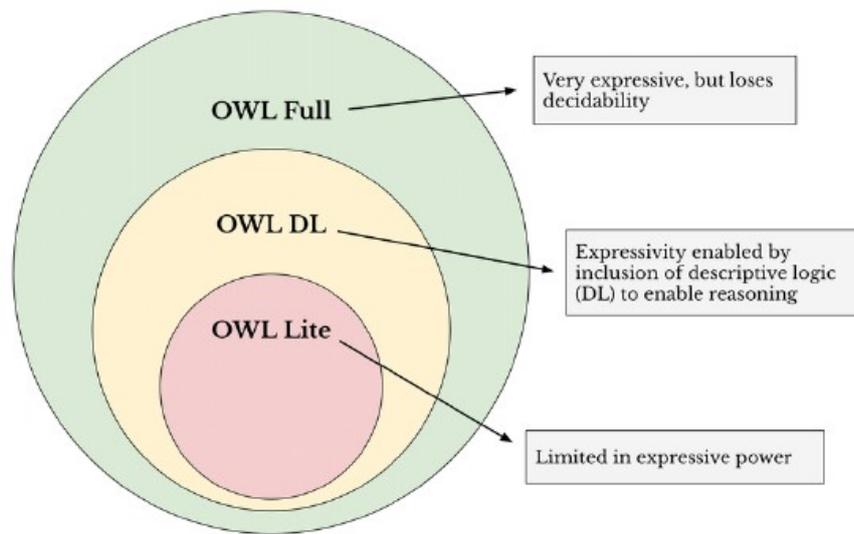
OWL è a tutti gli effetti un'estensione semantica di RDF, per tanto utilizza IRIs per i nomi; qualsiasi ontologia OWL fa riferimento ai datatypes RDF e XML Schema per i valori dei dati e per i datatypes, inoltre per ovviare al problema della decidibilità dell'ontologia ( un problema viene definito decidibile se esiste sempre un algoritmo che determina se esiste o meno una soluzione), OWL utilizza una logica descrittiva piuttosto che la logica del primo ordine di RDF.

L'Ontology Web Language ha una sua sintassi astratta che, può essere serializzata in diversi formati come RDF/XML, OWL/XML, Turtle.<sup>10</sup>

---

<sup>10</sup> Bartalesi Lenzi Valentina(2020/2021), *Corso universitario di Semantic Web*, Università di Pisa

# OWL Variants



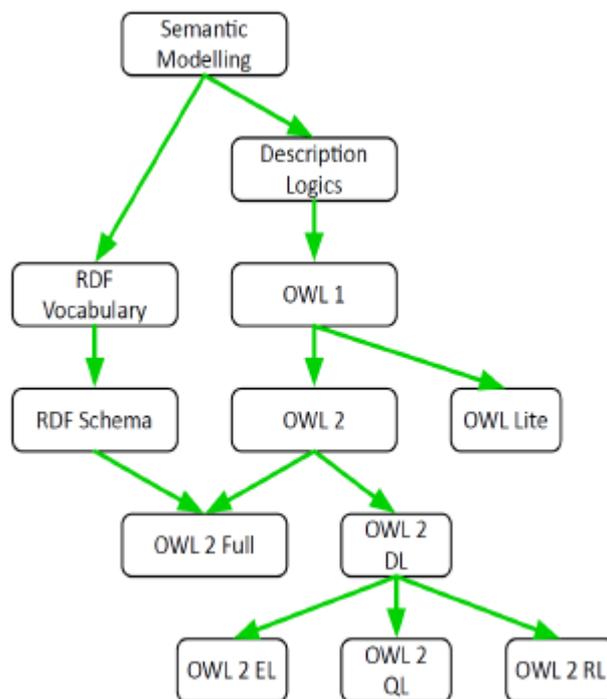
**Figura 1.8:** Varianti di OWL

Esistono diverse varianti di OWL:

- OWL DL: linguaggio realizzato mediante la codifica della sintassi astratta di OWL in una notazione concreta, è utilizzato dagli utenti che hanno bisogno della massima espressività ma, al tempo stesso necessitano della completezza computazionale e della decidibilità, ovvero le elaborazioni devono terminare in un tempo finito. L'aggettivo DL fa riferimento alla logica descrittiva, un campo di ricerca che ha determinato le basi formali di OWL;
- OWL Lite: è un sottoinsieme sintattico di OWL DL ed è utilizzato dagli utenti che vogliono rappresentare classificazioni gerarchiche e vincoli semplici;
- OWL Full: è un linguaggio risultante dall'estensione dell'RDF Schema con classi e proprietà necessarie per codificare la sintassi astratta di OWL in RDF. Viene utilizzato quando si vuole puntare alla massima espressività e libertà sintattica, ampliando il significato di vocabolari

predefiniti che appartengono sia all’RDF sia all’OWL ma, non dando garanzie computazionali.

Inoltre, OWL DL è definito *intractable* (un problema è definito *tractable* quando ha una soluzione in tempi polinomiali), per tale ragione è stato creato OWL 2 DL che presenta a sua volta tre profili: **OWL 2 EL**, molto espressivo ed è usato per ontologie con un gran numero di proprietà e classi ed è tractable; **OWL 2 QL** è utilizzato in applicazioni con un gran numero di dati; **OWL 2 RL** invece è utilizzato per quelle applicazioni che vogliono ottenere il massimo dell’espressività.



**Figura 1.9:** Profili e varianti di OWL 2

La famiglia OWL include una vasta gamma di varietà di linguaggi: si sviluppano lungo due direzioni, come è possibile vedere in figura, dividendosi a partire dai modelli semantici fino ad incontrarsi

all'OWL 2 Full. Il vantaggio di tutte queste varietà è che si crea un compromesso tra espressività e trattabilità.

OWL è in grado di creare ontologie definendo: classi e proprietà con relative gerarchie; assiomi sulle classi e sulle proprietà, ovvero affermazioni che devono essere vere. Inoltre, è in grado di costruire knowledge bases con individui e asserzioni su di essi.

Abilità di OWL è sicuramente quella di supportare il collegamento dei dati tra diverse sorgenti, la sua semantica di base permette di dedurre dati ulteriori rispetto ai dati forniti inizialmente. Ad esempio, la proprietà `:owlsameAs` serve per dichiarare che due individui sono uguali, anche se apparentemente possono non sembrarlo; in questo modo informazioni provenienti da diverse sorgenti ma, relative allo stesso individuo possono essere associate.

Di seguito, un esempio di inferenza: supponiamo di avere la proprietà `ex:haswritten`, avente come dominio `Person` e come range `LiteraryWork`, a partire da ciò il computer sarà in grado di fare inferenza. Tutti i soggetti che hanno il predicato `ex:haswritten` apparteranno alla classe persone e gli oggetti invece ad opere letterarie.

```
@prefix ex: <http://example.org/ontology#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.

Ex:hasWritten rdf:type rdf:Property;
               rdfs:domain ex:Person;
               rdfs:range ex:LiteraryWork;
               rdfs:subPropertyOf ex:HasAuthored.
```

**Figura 1.10:** Uso di OWL

Nel caso in cui avessimo come autore Herman Melville e come opera letteraria Moby Dick, il computer grazie all'URI della proprietà risalirà all'ontologia definita e di conseguenza eseguirà la seguente inferenza per cui Melville è una persona, Moby Dick un'opera letteraria. In questo esempio, dal momento che `ex:haswritten` è indicato come una sotto proprietà di `ex:hasAuthored`, eredita la stessa classe e valori.<sup>11</sup>

```
:HermanMelville a ex:Person.  
:MobyDick a ex:LiteraryWork.  
:HermanMelville ex:hasAuthored :MobyDick.
```

**Figura 1.11:** Esempio di inferenza ottenuto con OWL

OWL presenta delle classi che, forniscono un meccanismo di astrazione per raggruppare risorse che hanno caratteristiche simili e ad ogni classe sono associati degli individui che indicano le istanze della classe stessa. Esiste una classe generale chiamata `owl:Thing` che rappresenta la superclasse, racchiudente tutte le classi; invece, `owl:Nothing` è la classe che non ha nessuna istanza ed è la sottoclasse di tutte le classi.

In particolare, è utile accennare alcune caratteristiche che differenziano OWL dagli altri linguaggi come gli axioms, riportati di seguito:

- Class expressions, utilizzate per definire l'intensione della classe in maniera più specifica di RDFS. Si dividono in: **Set- theoretic class expressions** usate per combinare le classi come `unionOf`, `intersectionOf`, `complementOf`;

---

<sup>11</sup> Fusetti, C. (2017). *Dati bibliografici e linked (open) data: prospettive e criticità* (Bachelor's thesis, Università Ca'Foscari Venezia).

- Property restrictions: restringono i valori delle proprietà relative ad una classe, come ad esempio owl:minCardinality, in cui specifico che ogni membro di una classe può essere collegato almeno con tot individui di un'altra classe; owl:cardinality invece permette di definire il numero esatto di individui che devono legare due classi;
- Le Class expression axioms esprimono invece le relazioni tra le class expressions, ad esempio disjointWith per esprimere che gli individui di una classe A non sono uniti con quelli della classe B; oppure equivalentClass in cui viene espresso che gli individui appartenenti alla classe A sono presenti anche nella classe B e viceversa;
- Property Axioms: l'assioma inverseOf permette di definire l'inverso di una proprietà, invece le caratteristiche delle proprietà esprimono ad esempio transitività, simmetria, asimmetria e altro.<sup>12</sup>

## 1.4 Linked data: identificazione e scopo

Tim Berners-Lee durante una conferenza TED nel 2009 <sup>13</sup>ha introdotto per la prima volta il concetto di Linked Data (LD), i quali possono essere definiti come un insieme di pratiche e raccomandazioni per poter collegare dati strutturati, sono costituendo la base per lo sviluppo del Semantic Web.

I dati, legati e strutturati tra loro, creano un reticolo che grazie ai diversi collegamenti si accresce conducendo ad una maggiore divulgazione e fruizione delle informazioni.

Nel Semantic Web, i dati seguono il formalismo standard dell' RDF, di cui ho parlato in maniera approfondita nel paragrafo precedente.

Per creare un "dialogo", ovvero per favorire interoperabilità tra le fonti eterogenee pubblicate sul web, è possibile utilizzare il paradigma dei LD per cui le risorse vengono identificate da IRI, ogni ad

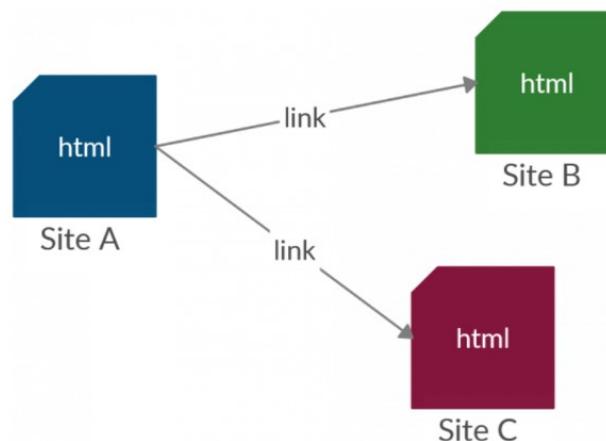
---

<sup>12</sup> McGuinness, D. L., & Van Harmelen, F. (2004). OWL web ontology language overview. *W3C recommendation*, 10(10), 2004.

ogni risorsa è associata una descrizione e vengono utilizzati vocabolari condivisi per rappresentare formalmente le relazioni semantiche tra le risorse.

Riassumendo, lo standard RDF permette l'interpretazione da parte delle macchine, gli URI invece permettono l'identificazione unica della risorsa in modo da evitare confusioni di interpretazioni e facilitare i collegamenti, le ontologie invece permettono di decidere le relazioni in base ai domini di appartenenza.

I documenti HTML che popolano il web sono collegati da link diretti e unidirezionali, ad esempio il sito A è collegato al sito B e C tramite due link, da questo collegamento non è possibile risalire ad un legame inverso o collegare ad esempio il sito B con il sito C.<sup>14</sup>



**Figura 1.12:** Web di documenti

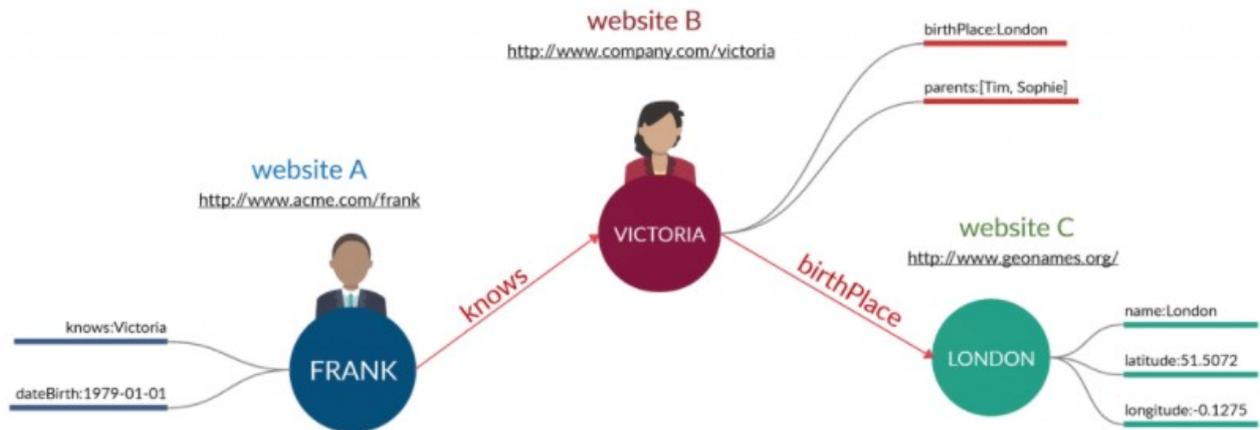
La differenza sostanziale del Semantic Web e dei Linked Data è quella di permettere di creare una rete di collegamenti, non unidirezionali né tantomeno soltanto diretti, sfruttando dati strutturati e regole di inferenza. Per inferenza si intende il processo attraverso cui da una proposizione accettata

---

<sup>14</sup> <https://fontistoriche.org/linked-data/>

come vera se ne deduce una seconda grazie al contenuto della prima; quindi, ricavare nuova conoscenza sono definita espressamente nel sistema.

Tramite un esempio, si può comprendere come i LD colleghino i dati ad altri dataset.



**Figura 1.13:** Esempio di Linked data

Abbiamo già parlato del fatto che RDF definisce le regole necessarie per trattare la struttura logica delle informazioni e le relazioni grazie alla logica dei predicati, creando le triple. Quest'ultime sono un modello sintagmatico costituito da: soggetto, predicato ed oggetto. Ciò può essere reso mediante un grafo etichettato ed orientato: i nodi rappresentano le risorse, gli archi le proprietà etichettate, i valori invece corrispondono a sequenze di caratteri.

Nell'esempio, la prima asserzione è Frank conosce Victoria: Frank è espresso da un URL, l'oggetto è un valore letterale e la proprietà che li lega è:knows.

Nella seconda asserzione, il soggetto è l'IRI che identifica Victoria, la proprietà è :birthPlace e l'oggetto è il letterale "London".

Da questo semplice esempio si può comprendere attraverso i LD si possano facilmente collegare dati provenienti da 3 siti diversi, permettendo di fare collegamenti ampliando i nostri dati, ad esempio

riferendosi alla Figura 1.13, i se anche Frank, come Victoria, è nato a Londra, si può di conseguenza consultare il website C (Geonames) e avere informazioni sulla suddetta città.

“L’atomizzazione della struttura dell’informazione esprime le caratteristiche del web dei dati; non si ha più un oggetto monolitico, bensì un insieme di singoli dati, delle particelle minime – atomi-riaggregabili con modalità e finalità differenti; ogni attributo dell’oggetto ha un valore in sé, partecipa alla sua natura, tramite relazioni parlanti, auto esplicanti. Le entità costituite dall’insieme degli atomi sono articolate in un insieme di dati strutturati, ciascuno in sé autonomo, combinabile logicamente con altri dati per produrre nuove entità [...]”.<sup>15</sup>

### 1.4.1 “RAW DATA NOW!”

Tim Berners-Lee durante una conferenza TED nel 2009 ha mostrato come l’uso dei dati “nudi e crudi” sia molto più vantaggioso rispetto al servirsi di documenti sul web: innanzitutto la ricerca è più immediata e si ottengono relazioni a catena che permettono di accedere ad un “enorme potenziale inesplorato”.

Durante la conferenza, egli mostra come ponendo la seguente domanda, mai fatta al motore di ricerca Google, “*What proteins are involved in signal trasduction and are related to pyramidal neurons?*”, esso restituisce 223.000 risultati senza che nessuno risponda davvero alla query posta.<sup>16</sup>

Successivamente mostra la differenza utilizzando un database in cui la risposta risulta immediata: nel processo sono coinvolte 32 proteine.

Berners-Lee ha definito i principi dei linked data ovvero un insieme di buone pratiche per pubblicare i dati nel web:

---

<sup>15</sup> Guerrini, M., & Possemato, T. (2012). Linked data: un nuovo alfabeto del web semantico. *Biblioteche oggi: Mensile di informazione aggiornamento dibattito*, 30(3), 7-15.

<sup>16</sup> <https://fontistoriche.org/linked-data/>

- Usare URIs come nomi delle cose: l'URI è un identificatore univoco usato per dare un nome a qualsiasi cosa: persone, oggetti, concetti. È possibile distinguere in questo modo risorse diverse oppure capire che una thing di un dataset è la stessa di un dataset diverso;
- Utilizzare HTTP URIs: il protocollo http provvede ad un meccanismo di ricerca e recupero delle risorse, quando le cose sono identificate da URI e trovate tramite il suddetto protocollo, è tutto più semplice, un'unione per facilitare l'acquisizione di informazioni;
- Fornire informazioni sulle risorse grazie agli URI deferenziabili, il loro scopo è raggiunto mediante l'URL della risorsa stessa. Bisogna distinguere tra l'URI del documento che contiene la risorsa e l'URI della risorsa stessa. L'URI deferenziabile può essere ottenuta utilizzando sia una rappresentazione HTML sia una in RDF. Per le interrogazioni di dataset diversi realizzati in RDF distribuiti sul Web viene utilizzato SPARQL, Simple Protocol And RDF Query Language;
- Per facilitare la scoperta di nuove informazioni, inserire link RDF ad altre URIs, si ha la creazione di un singolo grafo globale di dati interconnessi. Si crea una vera e propria rete in cui è possibile passare da un dataset all'altro, scoprendo di volta in volta informazioni e legami nuovi tra le risorse.

Un esempio di Linked Dataset è Wikidata, una base di conoscenza libera e aperta che può essere letta e modificata sia da esseri umani sia da macchine. Fornisce l'immagazzinamento centralizzato dei dati strutturati dei suoi progetti gemellati nell'ambito di Wikimedia, inoltre fornisce supporto anche ad altri servizi e siti. Il contenuto di Wikidata è disponibile sotto licenza libera, esportabile usando formati standard e può essere interconnesso ad altri insiemi di dati aperti.

È il più recente dei progetti gestiti dalla Wikimedia Foundation, fondazione no-profit che gestisce l'enciclopedia libera più famosa al mondo, Wikipedia, nato nell'ottobre del 2012. Il suo obiettivo è di raccogliere e strutturare i dati fondamentali delle voci e delle pagine degli oltre 800 progetti

Wikimedia in modo da poter essere letti, tradotti e utilizzati da chiunque in ciascuna delle 285 lingue ufficiali. Un progetto degno di menzione è sicuramente il progetto COVID-19.

Mentre il mondo combatte la **pandemia di COVID-19** in corso, Wikidata è stato estremamente utile nel raccogliere e modellare informazioni utili relative a **COVID-19**, da articoli di ricerca rilevanti e studi clinici a statistiche epidemiologiche di varie regioni, a lessemi per parole e frasi comunemente associate alla pandemia, e molto altro. Una gran mole di lavoro del **WikiProgetto COVID-19** va nella creazione e nel mantenimento di queste voci, rafforzando il ruolo di Wikidata come utile risorsa aperta nella pandemia.



**Figura 1.14:** Progetto COVID 19

Wikidata è una base di conoscenza secondaria, libera, collaborativa e multilingua: **secondaria** in quanto contiene non solo dati strutturati ma, anche le fonti da cui essi vengono tratti in modo da riflettere la diversità delle stesse nella descrizione di un determinato argomento e ribadire la centralità che la nozione di verificabilità dei progetti Wikimedia; **libera** in quanto i dati rilasciati sono di dominio pubblico, in modo da garantirne il più ampio riuso possibile (copia, modifica, distribuzione); **multilingua** ovvero i dati sono consultabili in 285 lingue ufficiali oltre ad essere in formati machine-readable; infine è **collaborativa** in quanto l’inserimento e la gestione dei dati è supervisionato dalla comunità di Wikidata, in particolare controlla che il materiale provenga da fonti affidabili attraverso i bot, programmi automatici.

## 1.4.2 Dai Linked Data ai Linked Open Data

I LD sono un insieme di dati interconnessi interpretabili dalle macchine, permettono di esprimere le informazioni, di condividerle tra differenti applicazioni e di renderle usufruibili da applicazioni diverse da quelle per cui sono state creati. Quando i dati sono distribuiti e usati in maniera gratuita, si parla di LOD (Linked Open Data).

Vi è una sottile differenza tra gli Open Data e i LOD: i primi sono dati soltanto utilizzabili da chiunque e distribuiti gratuitamente ma, non sono collegati ad altri dati tramite link e se lo fossero potrebbero non essere gratuiti per la distribuzione e il riutilizzo.

Quindi, i LOD sono sia Linked Data sia Open Data per tanto open sources e legati tra loro; la W3C community si impegna costantemente per aumentarne la presenza.

I benefici sono innumerevoli: facilità di integrazione tra i dati di formato diverso e di conseguenza la navigazione tra essi grazie ai collegamenti tramite link; aumento della conoscenza grazie all'inferenza che si ha con l'uso di dataset e fonti differenti.

Chiunque può pubblicare dati come Linked Data ma, ci sono dei criteri da rispettare, un riferimento possibile è *Methodological guidelines for publishing linked data* di Boris Villazòn- Terrazas che, possono essere riassunti in sette punti chiave:

- Identificare la fonte dei dati;
- Modellare il vocabolario sia utilizzando ontologie esistenti espresse in OWL o RDF sia creando nuove ontologie;
- Organizzare i dati in formato RDF con conseguente creazione di URI, fondamentali per omologare dati di fonti differenti;
- Pubblicare i dati;
- Controllare il formato dei dati creati per renderlo facilmente accessibile;

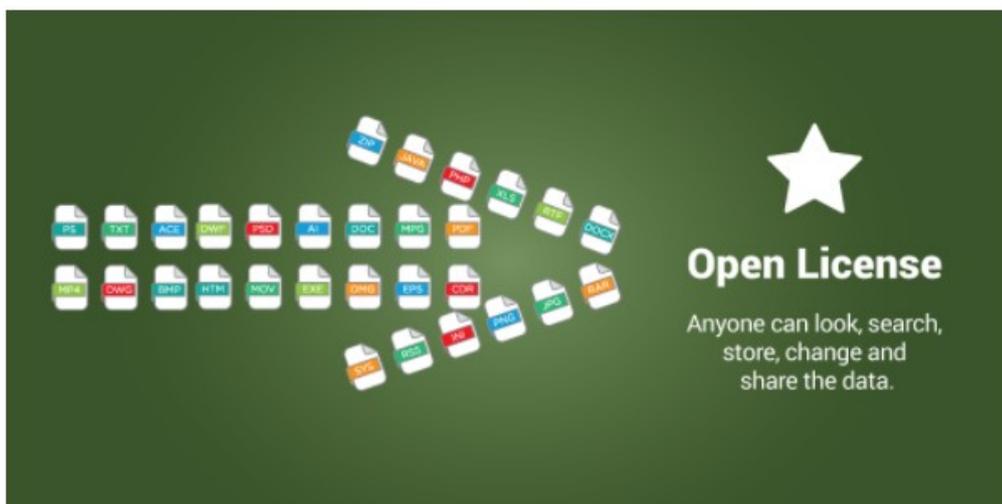
- Creare collegamenti tra i dataset e validarli;
- Rendere fruibili i dati, pubblicando il dataset sul CKAN Registry (Comprehensive Knowledge Archive Network), un registro per la pubblicazione di dati open in modo da renderli consultabili e riutilizzabili.<sup>17</sup>

### 1.4.3 Le cinque stelle dei LOD

Nel 2010, Tim Berners-Lee introdusse uno schema di classificazione per la pubblicazione dei Linked Open data, attribuendo una valutazione da una a cinque stelle, al fine di definire una sorta di standard di riferimento.

Di seguito, si analizza il significato di ciascuna stella con i relativi benefici per i dataset che ottengono più stelle.

#### One-Star Open Data



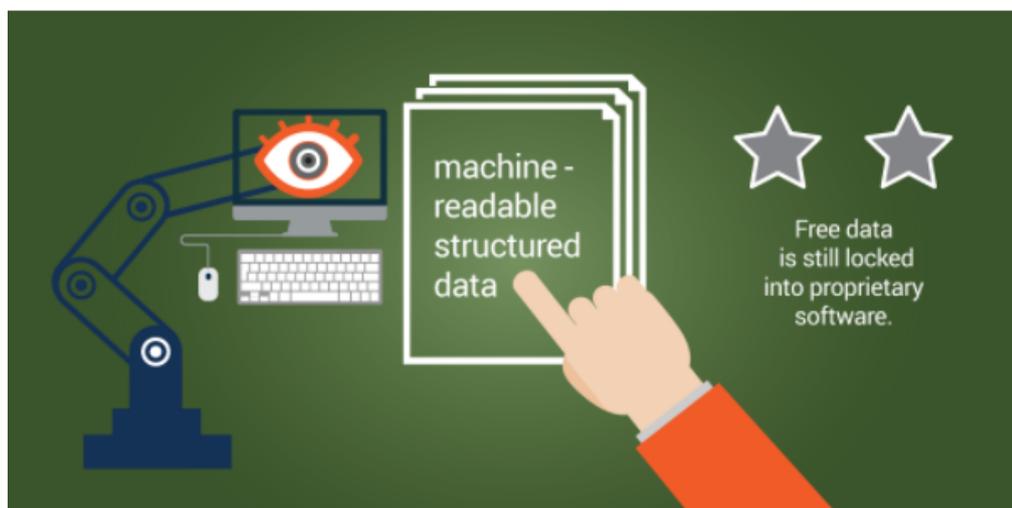
**Figura 1.15:** Una stella

<sup>17</sup> Guerrini, M., & Possemato, T. (2012). Linked data: un nuovo alfabeto del web semantico. *Biblioteche oggi: Mensile di informazione aggiornamento dibattito*, 30(3), 7-15.

L'attribuzione di una stella ad un dataset indica che i dati presenti sul Web hanno una licenza Open.

L'utente ha libertà di uso: può cercarli, memorizzarli, modificarli, condividerli.

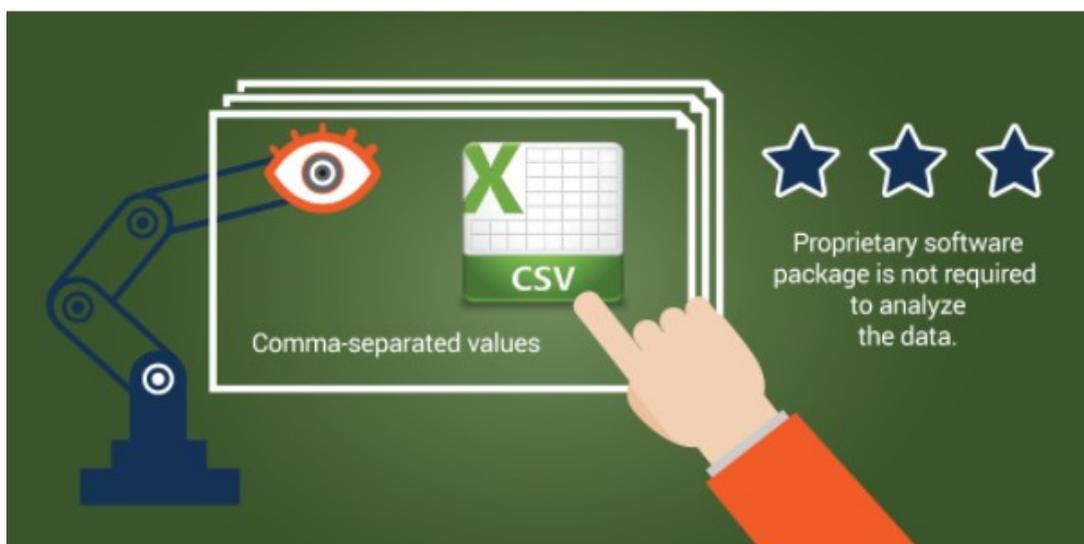
## Two-Star Open Data



**Figura 1.16:** Due stelle

La seconda stella viene attribuita agli Open Data che si presentano come dati strutturati interpretabili dalla macchina, ad esempio un'immagine di una tabella verrà convertita in un foglio Excel. Un vantaggio in più riguarda l'elaborazione dei dati con un software proprietario e la possibilità di esportarli in un altro formato, l'unico problema è che i dati sono ancora bloccati perché gli utenti dipendono dal software proprietario per scaricare i dati del documento.

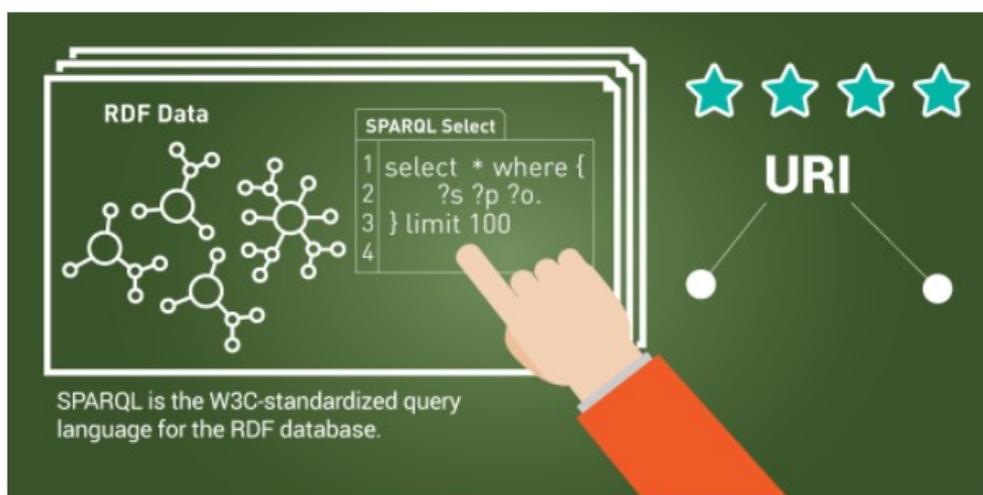
## Three-Star Open Data



**Figura 1.17:** Tre stelle

Il passaggio successivo riguarda l'ottenimento dei dati senza l'uso di un pacchetto software proprietario per analizzarli e/o scaricarli. Un esempio di ciò, è il formato CSV che memorizza dati tabulari in formato testo.

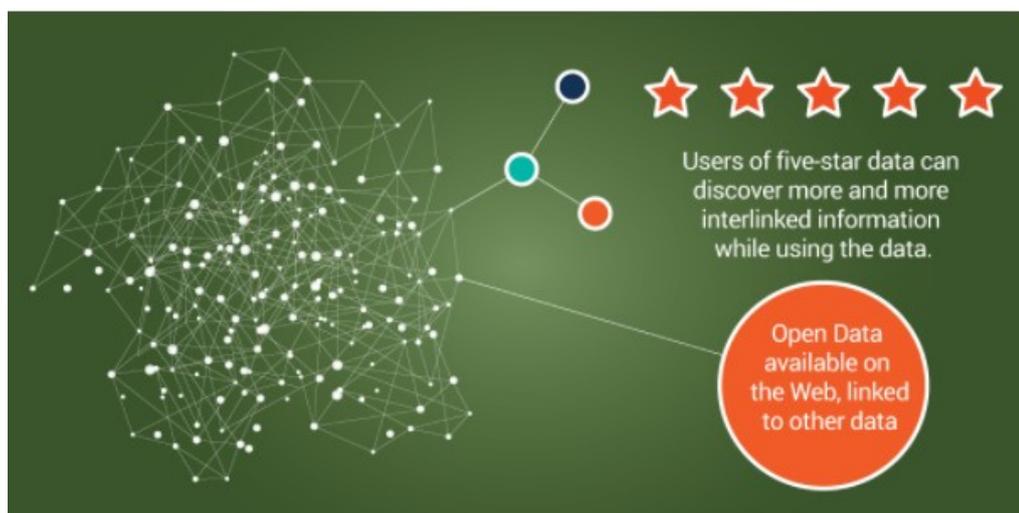
## Four-Star Open Data



**Figura 1.18:** Quattro stelle

La quarta stella viene attribuita ai dati che utilizzano gli standard del W3C: RDF che, permette di realizzare database di grafi semantici in modo da gestire dati interconnessi e mapparli, secondo un concetto fondamentale ovvero gli URI, che sono ID univoci di una risorsa. SPARQL, invece, è il linguaggio di interrogazione standard per un database RDF.

## Five-Star Linked Open Data



**Figura 1.19:** Cinque stelle

La quinta ed ultima stella viene attribuita ai Linked Open Data che sono collegati ad altri dati per fornire quante più informazioni possibili, il database RDF è in grado di mappare collegamenti a fonti di dati aperti come DBPedia, GeoNames e altri. È tutto a vantaggio sia degli editori sia dei consumatori che possono scoprire il contesto e nuovi collegamenti rispetto ai dati iniziali che hanno consultato.<sup>18</sup>

<sup>18</sup> <https://www.ontotext.com/knowledgehub/fundamentals/five-star-linked-open-data/>

## 1.6 Il World wide web consortium

“Il W3C è la sigla di World wide web consortium, comunità internazionale che si occupa della definizione di standard web aperti per promuovere l’accessibilità e la compatibilità delle tecnologie in rete (tra gli standard, per es., HTML, XHTML, XML, RDF, CSS, PNG, SVG e SOAP)<sup>19</sup>

Dunque, il W3C è preposto all’emanazione degli standard adottati sul Web, guidandone anche lo sviluppo e l’interoperabilità. Tim Berners-Lee fondò quest’organizzazione nel 1994 a Ginevra, successivamente venne spostata negli USA presso il laboratorio di Computer Science dell’istituto di tecnologia del Massachusetts.

Il W3C è diviso in diversi gruppi, l’iter affinché una specifica diventi uno standard è il seguente: quando un gruppo riesce a raggiungere il consenso su delle specifiche, viene pubblicato un documento definito last call, se non sono necessarie modifiche, si passa allo stato di “raccomandazione”, W3C Recommendation, che si invita a seguire per gli sviluppatori del web.

Nella pagina ufficiale del W3C, oltre ad essere esplicitata la loro missione, sono descritti i 7 principi che si pongono come obiettivi da raggiungere e che, quindi, guidano il loro agire.

**Accesso Universale:** il W3C ha come scopo quello di rendere i benefici derivanti dal Web come le nuove forme di comunicazione e di condivisione della conoscenza, accessibili a tutti indistintamente dall’hardware, software, linguaggio nativo, cultura, abilità fisica/mentale. A testimonianza di ciò, il W3C si impegna in progetti come “Attività di indipendenza dall’ Hardware, Iniziativa per l’accessibilità del Web, Attività di browser vocali”.<sup>20</sup>

---

<sup>19</sup> [https://www.treccani.it/enciclopedia/w3c\\_%28Lessico-del-XXI-Secolo%29/](https://www.treccani.it/enciclopedia/w3c_%28Lessico-del-XXI-Secolo%29/)

<sup>20</sup> <https://www.w3.org/Voice/> , <https://www.w3.org/WAI/>

**Web Semantico:** un altro obiettivo è quello di affinare e sviluppare terminologie tali per cui i computer siano in grado di interpretare e scambiare dati, in vista del superamento dei problemi relativi all'interconnessione con altri servizi, aiutando le persone a trovare in maniera semplice, veloce e precisa ciò a cui ambiscono che siano informazioni sanitarie, recensioni di film, ristoranti. Infatti, il W3C si è impegnato nell'implementazione di linguaggi come RDF, XML per costruire la base del web semantico.

**Fiducia:** il web è diventato sempre più un ambiente collaborativo in cui l'utente non è soltanto uno spettatore passivo piuttosto un fruitore attivo, affinché sia consapevole e responsabile del contenuto che immette in rete, è necessario creare un Web basato sulla fiducia per migliorare e potenziare la confidenza delle persone con lo strumento. Obiettivo pertinente con progetti come XML signatures, sistemi di annotazione, gruppi di creatori ecc.

**Interoperabilità:** in principio, i programmi funzionavano soltanto con applicazioni dello stesso produttore. Oggi questo approccio è improponibile perché gli utenti vogliono visualizzare i contenuti usando applicativi differenti come sintetizzatori vocali, navigatori e altro. Il W3C è un'organizzazione indipendente dai produttori quindi punta a creare e promuovere linguaggi e protocolli open source per favorire l'interoperabilità, inoltre mira anche alla promozione del consenso tra i produttori e di favorire le discussioni pubbliche.

**Evolvibilità:** un ulteriore sforzo e obiettivo è quello di rendere quanto più possibile il Web flessibile e aperto a sviluppi futuri, dal momento che è un ambiente in continuo divenire, per agevolare questa idea, i principi di semplicità, modularità, compatibilità sono la stella polare di ogni progetto.

**Decentralizzazione:** è un obiettivo essenziale per il Web moderno perché un sistema centralizzato rende ogni comunicazione complessa perché deve essere approvata da un'autorità centrale portando

inevitabilmente problemi legati alla congestione del traffico di dati e alla vulnerabilità dell'architettura di internet.

**Cooler Multimedia:** il W3C, come già accennato, mira ad una maggiore partecipazione e consapevolezza degli utenti, questo include anche la possibilità di accrescere la loro creatività mediante file multimediali (immagini, audio, video). Attraverso le segnalazioni degli utenti si impegna a creare il cosiddetto *Cooler Web* con linguaggi come Scalable Vector Graphics.

## CAPITOLO 2

### 2.1 Marginalia: i precursori delle annotazioni moderne

Marginalia è un termine coniato nel 1832 da Samuel Taylor Coleridge<sup>21</sup> e si riferisce a tutte le modifiche che il lettore apporta al testo scritto, includendo le note al margine, le sottolineature e le evidenziazioni. Questa pratica precede il libro stampato estendendosi fino ad includere i commenti dei monaci medievali negli scriptorium<sup>22</sup>.

I marginalia provvedono sia a fornire una prospettiva unica riguardo ciò che pensa il lettore durante il processo di comprensione del testo sia a soddisfare molteplici scopi: aiutare a ricordare, comunicare qualcosa, trascrivere spunti di riflessione.

Sono stati scritti diversi articoli scientifici e libri riguardo i marginalia di personaggi storici, in particolare Heather Jackson, professoressa di Inglese all'Università di Toronto, è stata una pioniera nello studio dei marginalia di cui ha analizzato la storia, la forma che possono assumere nonché la psicologia che li soggiace.<sup>23</sup>

Inizialmente la Jackson si è concentrata sugli scritti di Coleridge, poeta britannico e scrittore prolifico di marginalia, la sua fama di annotatore era tale che i suoi amici gli consegnavano libri affinché lui li arricchisse con note e opinioni al riguardo.<sup>24</sup>

“Marginalia: Readers Writing in Books” (2001) è un libro di Jackson basato sullo studio di 2000 libri annotati da lettori famosi e anonimi degli ultimi tre secoli. Oltre ad occuparsi di Coleridge, la professoressa espande il suo studio includendo i commenti del poeta romantico britannico John Keats

---

<sup>21</sup> <https://www.studenti.it/samuel-taylor-coleridge-vita-libri-poesie.html>

<sup>22</sup> Jackson, H.J. (2001). *Marginalia: Readers writing in books*. New Haven, CT: Yale University Press.

<sup>23</sup> <https://yalebooks.yale.edu/book/9780300097207/marginalia>

<sup>24</sup> Wagstaff, K. L. (2012). *The evolution of marginalia. Unpublished paper. http://www.wkiri.com/slis/wagstaff-libr200-marginalia-1col.pdf [accessed 24 April 2016].*

riguardo opere di John Milton<sup>25</sup>; i copiosi commenti marginali di Horace Walpole<sup>26</sup>, le note del critico letterario del ventesimo secolo Frye<sup>27</sup> riguardo opere di John Bunyan, predicatore e teologo inglese del '600, e opere di Marianne Moore, poetessa e scrittrice statunitense del '900. La Jackson si occupa anche di analizzare scrittori anonimi di marginalia, le cui note riguardano sia richieste di restituzione di libri sia conversazioni scritte tra amici che condividevano gli stessi libri.

Gli annotatori anonimi presi in considerazione appartengono ad un periodo che abbraccia tre secoli, dal 1700 al 200. Le motivazioni di questo anonimato possono essere attribuite al fatto che storicamente i bibliotecari hanno avuto poca tolleranza per le note nei libri, non vedendo alcuna differenza tra queste e altre forme di deturpazione, ad esempio i bibliotecari monastici scrivevano maledizioni e altre ingiurie nei libri per dissuadere i potenziali abusatori.

In *“The Medieval Professional Reader at work”* sono analizzati i marginalia di chierici e scrivani medievali per comprendere il pensiero del lettore tardo medievale inglese, un annotatore è conosciuto come “Red Ink Anno-tator” quasi come fosse un misterioso e spavaldo bandito.<sup>28</sup>

Un altro esempio di annotatore anonimo è presentato dalla Jackson in seguito all' esaminare una copia del libro scritto da Boswell, scrittore scozzese del XVIII secolo, sulla vita di Samuel Jackson.<sup>29</sup>

Questa copia presenta tantissime note a margine fornendo una lettura contemporanea dell'opera di partenza ma, nessuno ha mai determinato con certezza chi si nascondesse dietro la firma di “Scriblerus”, come si riferisce l'annotatore altrimenti anonimo. La cosa certa è che Scribleus conosceva personalmente i contemporanei di Johnson perché egli prova piacere a correggere i fatti di Boswell e confutare le sue opinioni.<sup>30</sup>

---

<sup>25</sup> [https://it.wikipedia.org/wiki/John\\_Milton](https://it.wikipedia.org/wiki/John_Milton)

<sup>26</sup> [https://it.wikipedia.org/wiki/Horace\\_Walpole](https://it.wikipedia.org/wiki/Horace_Walpole)

<sup>27</sup> [https://it.wikipedia.org/wiki/Northrop\\_Frye](https://it.wikipedia.org/wiki/Northrop_Frye)

<sup>28</sup> Kerby-Fulton, K., & Hilmo, M. (Eds.). (2001). *The Medieval Professional Reader at Work: Evidence from Manuscripts of Chaucer, Langland, Kempe, and Gower* (Vol. 85). Victoria, BC: University of Victoria.

<sup>29</sup> [https://it.wikipedia.org/wiki/Samuel\\_Johnson](https://it.wikipedia.org/wiki/Samuel_Johnson)

[https://www.researchgate.net/publication/271296191\\_Jackson\\_H\\_J\\_Marginalia\\_Readers\\_Writing\\_in\\_Books\\_New\\_Haven\\_Conn\\_and\\_London\\_Yale\\_Univ\\_Pr\\_2001\\_324p\\_2795\\_alk\\_paper\\_ISBN\\_0300088167\\_LC\\_00-043721](https://www.researchgate.net/publication/271296191_Jackson_H_J_Marginalia_Readers_Writing_in_Books_New_Haven_Conn_and_London_Yale_Univ_Pr_2001_324p_2795_alk_paper_ISBN_0300088167_LC_00-043721)

Uno dei motivi principali che spinge la studiosa ad occuparsi dello sviluppo di questa pratica nel corso del tempo è quello di rivelare la relazione tra le pratiche di pubblicazione e di lettura ma, anche di analizzare i cambiamenti che influenzano il pensiero degli scrittori nel corso della stesura.

Il suo studio si conclude con l'invito a considerare maggiormente i marginalia, al fine di studiare come maggiore accuratezza e profondità le opere di un determinato autore.<sup>31</sup>

Heather Jackson è anche autrice di “*Working with Paradata, Marginalia and Fieldnotes*”, opera del 2017, il cui settimo capitolo è dedicato ai marginalia di John Adams, secondo Presidente degli Stati Uniti d’America (4 marzo 1797 – 4 marzo 1801). Adams ha creato dei marginalia con l’intento di fornire alle generazioni future una migliore e più completa visione delle sue prospettive politiche.

In quel periodo, i libri erano molto costosi ed era pratica comune scambiarli, pertanto le note fatte a margine da lettori non erano destinate a restare private ma, concepite per essere lette da quante più persone possibile.<sup>32</sup>

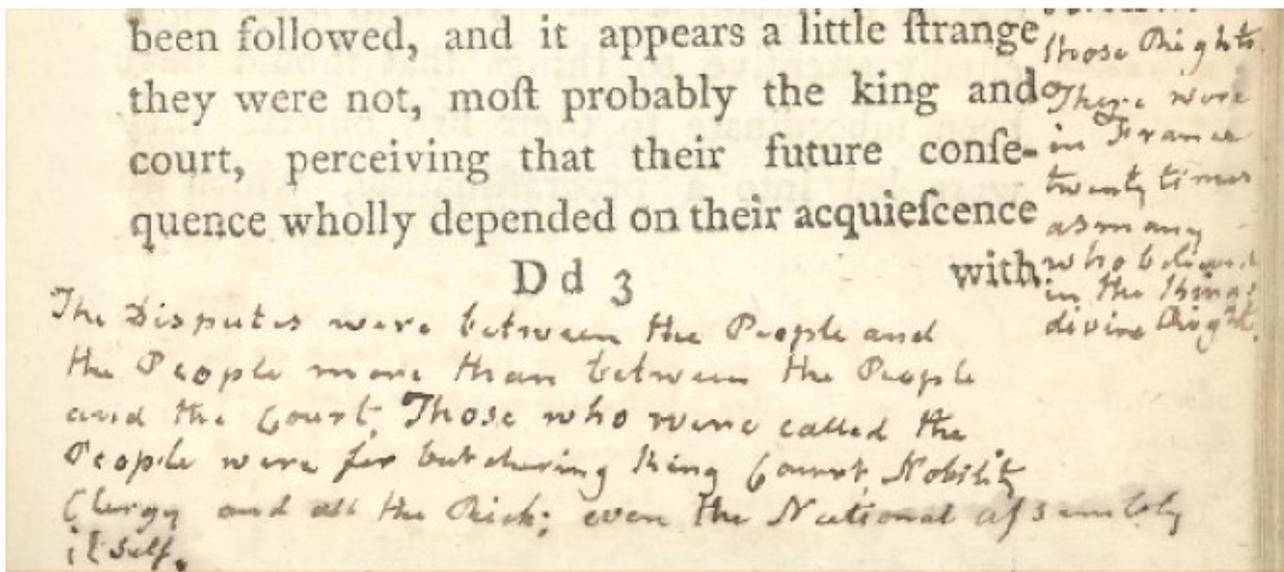
Nella Figura 2.1 è riportato un esempio di annotazione di Adams del libro “*An historical and moral view of the origin and progress of the French Revolution*” di Mary Wollstonecraft, filosofa e scrittrice britannica, considerata fondatrice del femminismo liberale.<sup>33</sup>

---

<sup>31</sup> <https://prabook.com/web/heather.jackson/3777666>

<sup>32</sup> Jackson, H. J. (2017). John Adams’s marginalia: then and now. In *Working with Paradata, Marginalia and Fieldnotes*. Edward Elgar Publishing.

<sup>33</sup> [https://it.wikipedia.org/wiki/Mary\\_Wollstonecraft](https://it.wikipedia.org/wiki/Mary_Wollstonecraft)



**Figura 2.1:** Marginalia scritti da John Adams a pagina 405 del libro di Mary Wollstonecraft's *"An historical and moral view of the origin and progress of the French Revolution"*

### 2.1.1 Forme e motivazioni dei marginalia

In passato, la forma, il tipo di marginalia erano legati al formato e allo spazio messo a disposizione ai margini del libro fisico. Inizialmente, il lettore utilizzava il frontespizio e la pagina finale per scrivere un suo personale riassunto del libro o un indice per appuntare le pagine per lui più rilevanti. Per quanto concerne il corpo del testo, il lettore era solito sottolineare ed evidenziare le parole, le frasi più importanti, le note ai margini del libro, invece, dipendevano dallo spazio a disposizione e racchiudevano un insieme di commenti, scarabocchi, definizioni, segni di varia natura e abbreviazioni.<sup>34</sup>

Per quanto riguarda invece i motivi che hanno spinto le persone a scrivere nei loro libri, in primo luogo, la studiosa Jackson, nella sua opera del 2001 *"Marginalia: Readers writing in books"*, racconta

<sup>34</sup> Wagstaff, K. L. (2012). The evolution of marginalia. *Unpublished paper*. <http://www.wkiri.com/slis/wagstaff-libr200-marginalia-1col.pdf> [accessed 24 April 2016].

che uno dei motivi principali fosse legato all'espressione di proprietà del libro quindi si scriveva il proprio nome o la data di acquisto, ma anche maledizioni per evitarne il furto. A tal proposito, Drogin, autore e illustratore statunitense del '900, nel suo libro "Anathema!" ha raccolto una collezione di maledizioni contenute nei libri medievali. Ciò risulta spiegabile considerando che nel medioevo i libri rappresentavano un oggetto di lusso, la produzione di una copia richiedeva anche mesi o anni, per tanto il furto o la perdita di un libro rappresentava una grave disgrazia. Per spaventare i possibili ladri, venivano redatte delle vere e proprie maledizioni come riporta Drogin in un esempio di uno scriba del 1172 "Se qualcuno porta via questo libro, che muoia; che sia fritto in una padella; che il male e la febbre lo colgano; che sia spezzato sulla ruota e impiccato. Amen."<sup>35</sup>

In secondo luogo, in relazione all'uso dei marginalia esiste un beneficio pedagogico riservato agli studenti o chiunque cerchi di imparare da un testo. L'arte di sottolineare, commentare mentre si legge aiuta il processo di apprendimento.

In terzo luogo, nel XIX e XX secolo la lettura era un'attività socialmente condivisa per tanto i commenti marginali erano indirizzati solitamente a qualcuno. A volte il destinatario, scriveva a sua volta dei commenti e poi restituiva il libro al proprietario. Coleridge, ad esempio, scriveva commenti diversi per copie di uno stesso libro a seconda che fossero per uso personale, commento critico o pubblicazione.<sup>36</sup>

## 2.1.2 Esempi di marginalia nel tempo

Si può affermare che la relazione tra testo e marginalia è antica quanto la scrittura, numerosi esempi possono essere ritrovati nella storia: papiri di autori greci, quali Pindaro e Platone, nei quali vi sono marginalia dovuti alla rilettura, a pensieri successivi che finiscono per corredare l'opera stessa. Anche

---

<sup>35</sup> Drogin, M. (1983). *Anathema!* Totowa, NJ: Allanheld & Schram

<sup>36</sup> Wagstaff, K. L. (2012). The evolution of marginalia. *Unpublished paper*. <http://www.wkiri.com/slis/wagstaff-libr200-marginalia-1col.pdf> [accessed 24 April 2016].

nel mondo latino, diversi autori ci hanno lasciato testimonianze sull'uso delle annotazioni marginali. Seneca, filosofo, drammaturgo e politico romano esponente dello stoicismo di età imperiale, nelle *Epistulae morales ad Lucilium*, raccolta di 124 lettere indirizzate alla crescita individuale dell'amico Lucilio,<sup>37</sup> consiglia vivamente di non sprecare tempo in letture prolisse ma, di farsi guidare dalle sue *notae*, redatte per indirizzare l'amico ad un miglior approccio testuale indicandogli i passaggi cruciali per l'apprendimento.

Un altro esempio, Plinio il giovane e Tacito si scambiavano reciprocamente le opere prima della pubblicazione per poi restituirle con annotazioni di varia natura.<sup>38</sup>

Il margine del testo è dunque lo spazio fisico dedicato ad esprimere il dialogo con il testo da parte del lettore e dell'autore, di conseguenza le forme e le varianti che le annotazioni possono assumere sono infinite.

Un ultimo esempio che, merita di essere citato, riguarda i Canti Leopardiani di cui possediamo gli autografi con diverse correzioni e varianti. Quelli relativi alle canzoni del 1818-23, presentano sui margini note filologico-linguistiche ai testi. Nel caso di Giacomo Leopardi, poeta italiano dell'Ottocento, le note hanno una loro specificazione tematica (civile, filosofica) e si caratterizzano per la loro qualità. Il suo scopo è quello di dare informazioni aggiuntive riguardo alle sue scelte linguistiche, in particolare motivando voci e costrutti in base all'uso documentato nella tradizione letteraria italiana dal Trecento al Seicento nonché riferimenti ai classici latini e greci. In altri casi, l'autore vuole specificare il significato di alcuni termini, creando un duplice dialogo sia verso se stesso sia verso i lettori, "a questo proposito la nota ai vv. 135-6 di All'Italia «Chiudere i moribondi lumi non è ridondanza perché si possono chiudere gli occhi anche e altro che per morte»"<sup>39</sup>.

---

<sup>37</sup>[https://it.wikipedia.org/wiki/Lucio\\_Anneo\\_Seneca#Le\\_Epistole\\_a\\_Lucilio:\\_la\\_lettera\\_filosofica\\_come\\_genere\\_letterario](https://it.wikipedia.org/wiki/Lucio_Anneo_Seneca#Le_Epistole_a_Lucilio:_la_lettera_filosofica_come_genere_letterario)

<sup>38</sup> Campanelli, M. (2002). Scrivere in margine, leggere il margine: frammenti di una storia controversa. *Scrivere in margine, leggere il margine: frammenti di una storia controversa*, 851-939.

<sup>39</sup> Blasucci, L. (1991). CANZONI LEOPARDIANE: dalle note autografe alle Annotazioni del 1824. *Belfagor*, 46(3), 261- 274.

## 2.2 Le annotazioni sui supporti elettronici

L'avvento del digitale non solo ha comportato dei cambiamenti nel modo di leggere e creare appunti ma ha determinato anche un passaggio terminologico, da marginalia ad annotazioni perché sui supporti elettronici non esiste più un margine sui cui scrivere.

In particolare, le annotazioni sui supporti elettronici, gli e-book<sup>40</sup>, sebbene non posseggano caratteristiche standard, comunemente permettono di evidenziare il testo, inserire segnalibri e aggiungere commenti in parti specifiche del testo. Un grande vantaggio per i lettori digitali, definiti e-reader, è che le note possono essere ricercate, filtrate, aggregate, copiate, incollate e condivise.<sup>41</sup>

Ad esempio, l'ebook prodotto da Amazon, il Kindle, memorizza tutte le annotazioni del lettore sui suoi server e mostra il numero di persone che hanno evidenziato certi passaggi.<sup>42</sup>

### 2.2.1 Cambiamenti di lettura e d'annotazione degli e-reader

I supporti elettronici hanno rivoluzionato il modo di relazionarsi al testo scritto, studi recenti<sup>43</sup> hanno dimostrato che il processo di lettura è diverso per i libri elettronici rispetto a quelli cartacei. I primi vengono “usati” piuttosto che letti ovvero le persone sono più orientate a cercare determinate informazioni scorrendo il testo rispetto a condurre una lettura lineare. Inoltre, anche le abitudini di annotazione differiscono. Ciò è stato evidenziato da uno studio condotto da Catherine C. Marshall, professoressa presso il dipartimento di Computer Science e Ingegneria dell'Università del Texas, e

---

<sup>40</sup> <https://it.wikipedia.org/wiki/E-book>

<sup>41</sup> Wolfe, J. (2002). Annotation technologies: A software and research review. *Computers and Composition*, 19(4), 471-497.

<sup>42</sup> Alter, A. (2012, July). Your e-book is reading you. Retrieved November 18, 2012, from <http://online.wsj.com/article/SB10001424052702304870304577490950051438304.html>

<sup>43</sup> Staiger, J. (2012). How e-books are used. *Reference & User Services Quarterly*, 51 (4), 355 -

AJ Bernheim Brush, ricercatrice nel campo dell'automazione presso Microsoft. In questo studio<sup>44</sup> del 2004 per comprendere le differenze che sussistono tra i diversi tipi di annotazioni, sono stati presi come campione undici studenti, ai quali è stato chiesto di leggere e annotare prima una copia stampata e poi una versione elettronica dello stesso libro. I risultati hanno riportato che gli studenti hanno scritto solo un quarto delle annotazioni cartacee nella versione elettronica (379 versus 1535). Le ragioni che le due studiose hanno attribuito a questa discrepanza numerica sono varie: un maggiore sforzo per aggiungere le annotazioni elettroniche, la preferenza di condividere i propri pensieri in posti personali rispetto a supporti che possono renderle pubbliche. Inoltre, è stata scoperta anche una differenza qualitativa ovvero le annotazioni elettroniche consistono soprattutto in contenuti semantici caratterizzati da un uso colto della grammatica, in contrasto con le sottolineature, evidenziazioni, abbreviazioni dei supporti cartacei.

I libri stampati continuano a rivestire un ruolo importante soprattutto nella sfera psicologica e intima dei lettori ma, se gli e-book non potranno mai sostituire del tutto i libri stampati, la loro creazione, pubblicazione e consumo continueranno sempre di più a svilupparsi.<sup>45</sup>

Un punto in comune tra i due supporti è l'interazione sociale che ne deriva, in passato gli amici si scambiavano libri annotandoli, oggi invece la condivisione si ha tramite internet. Gli utenti del web vogliono avere la possibilità di esprimersi e commentare, ricercano uno spazio per farlo nei social network, blog, nei tool forniti per le annotazioni.<sup>46</sup>

---

<sup>44</sup> Marshall, C. C., & Brush, A. B. (2004, June). Exploring the relationship between personal and public annotations. In *Proceedings of the 4th ACM/IEEE-CS joint conference on digital libraries* (pp. 349-357).

<sup>45</sup> Braun, L. W. (2011). Now is the time: E-books, teens, and libraries. *Young Adult Library Services*, 9 (4), 27–30.

<sup>46</sup> Wagstaff, K. L. (2012). The evolution of marginalia. *Unpublished paper*. <http://www.wkiri.com/slis/wagstaff-libr200-marginalia-1col.pdf> [accessed 24 April 2016].

## 2.2.2 Le abitudini dei lettori moderni

Nel recente articolo “*Marginalia in the digital age: Are digital reading devices meeting the needs of today's readers?*”<sup>47</sup> è riportato un sondaggio sulle nuove attitudini e comportamenti dei lettori riguardo ai marginalia sia sui supporti cartacei che elettronici. Ciò che è emerso è che le persone desiderano avere la possibilità di annotare sui loro device elettronici ma, al tempo stesso, reputano l’annotazione come difficile perché complicata da attuare o richiedente troppo tempo e ciò scoraggia inevitabilmente il lettore ad annotare.

La possibilità di annotare un testo viene resa disponibile in un dispositivo o in un tool soltanto se c’è una richiesta adeguata al riguardo; pertanto, si può considerare come un campo in sviluppo, soggetto a continui miglioramenti sotto ogni aspetto.

I dati, per dimostrare lo studio sopra citato riguardo le abitudini dei lettori, sono stati raccolti nel 2013 attraverso un sondaggio online, pubblicizzato da siti web come Wired.com, rivista statunitense, attraverso cui era possibile accedere al link (oggi non più attivo)<sup>48</sup>. Il campione preso in esame consiste di 510 lettori maggiorenni ma, esso non può essere ritenuto rappresentativo dell’intera popolazione ma, permette di essere rappresentativo della popolazione di lettori attivi e consumatori di dispositivi di lettura digitale. Il 72% degli intervistati erano degli Stati Uniti, il 13% del Regno Unito, 6% di altri Paesi europei, il 4% del Canada, 2% Australia e 3% di altri Paesi.

Inoltre, il 59% degli intervistati erano donne e il 41% uomini, dato coerente con le osservazioni passate che hanno messo in evidenza come le donne tendano a leggere maggiormente rispetto agli uomini.<sup>49</sup> Lo scopo di questo studio è evidenziare sia la necessità delle annotazioni per i lettori sia

---

<sup>47</sup> Bold, M. R., & Wagstaff, K. L. (2017). *Marginalia in the digital age: Are digital reading devices meeting the needs of today's readers?*. *Library & Information Science Research*.

<sup>48</sup> Liu, J. H. (2013, March 21). *Do you write in the margins?* Wired Retrieved from <http://archive.wired.com/geekdad/2013/03/do-you-write-in-the-margins/>

<sup>49</sup> Tepper, S. J. (2000). *Fiction reading in America: Explaining the gender gap*. *Poetics*, 27(4), 255–275.

l'importanza di dotare gli strumenti di lettura futuri di questa possibilità. Le domande a cui il sondaggio vuole rispondere sono:

- RQ1. Quant'è importante la possibilità di annotare libri per i lettori?
- RQ2. Le abitudini dei lettori riguardo le annotazioni sono diverse tra i supporti elettronici e i libri stampati?
- RQ3. I supporti di lettura digitali offrono il giusto supporto per le annotazioni?

Lo studio si basa su delle ipotesi formulate come conseguenza alle domande precedenti:

- H0. Annotare per piacere e per lavoro sono ugualmente comuni (ipotesi nulla);
- H1. Le pratiche di annotazioni si differenziano tra supporti elettronici e cartacei;
- H2. I lettori sono più propensi ad annotare libri quando leggono per lavoro o motivi scolastici piuttosto che per piacere.

Ogni ipotesi è stata testata usando uno specifico test di verifica d'ipotesi, in statistica questi test sono utilizzati per verificare la bontà di un'ipotesi intesa come un'affermazione che ha come oggetto accadimenti nel mondo reale che, si presta ad essere confermata o smentita dai dati osservati sperimentalmente.<sup>50</sup> Per la valutazione dell'ipotesi è stato utilizzato il **test z**, un test statistico di tipo parametrico che ha lo scopo di verificare se il valore medio di una distribuzione si discosta significativamente da un certo valore di riferimento.<sup>51</sup>

Nei test di verifica d'ipotesi è inoltre importante considerare il valore p, comunemente detto p-value, definito come segue dall'Enciclopedia Treccani della matematica<sup>52</sup>: “**p-value** o valore p, in statistica,

---

<sup>50</sup> [https://it.wikipedia.org/wiki/Test\\_di\\_verifica\\_d'ipotesi](https://it.wikipedia.org/wiki/Test_di_verifica_d'ipotesi)

<sup>51</sup> <https://datascience.eu/it/matematica-e-statistica/z-test-t-test-somiglianze-e-differenze/>

<sup>52</sup> [https://www.treccani.it/enciclopedia/p-value\\_%28Enciclopedia-della-Matematica%29/](https://www.treccani.it/enciclopedia/p-value_%28Enciclopedia-della-Matematica%29/)

valore sotto il quale i dati ricavati da un test statistico conducono al rifiuto dell'ipotesi nulla. Nella pratica, anziché fissare in anticipo il livello di significatività  $\alpha$ , che corrisponde alla probabilità di commettere un errore di prima specie (cioè di rifiutare l'ipotesi nulla mentre essa è vera), spesso si determina direttamente il p-value, se esso è molto piccolo (generalmente minore di 0,05), si rifiuta l'ipotesi nulla".

Marginalia practices for printed books.

Practice	Reading for				Difference significant? ( $p = 0.001$ )
	Pleasure		Work or education		
	N	%	N	%	
Write in margins	290	57	408	80	Y
Write inscription	315	62	179	35	Y
Write name	369	72	358	70	N
Highlight passage	264	52	403	79	Y
Dog-ear page	339	66	313	61	N
Write sticky note	280	55	413	81	Y
Blank sticky note	345	68	436	85	Y

Note: Percentages sum to more than 100% because respondents ( $n = 510$ ) were allowed to select more than one choice.

**Figura 2.2:** Marginalia nei libri stampati

La Figura 2.2 mostra le pratiche di annotazione nei libri stampati (scrivere ai margini, dediche, scrivere il proprio nome, sottolineare dei passaggi, piegare l'angolo della pagina, l'uso di note adesive e note bianche) e la differenziazione tra lettura di piacere e di lavoro/formazione. Infine, sono riportati i valori del p value, la cui risposta risulta essere binaria (Yes o Not). Ai fini di una comprensione corretta dei dati, la somma delle percentuali non è 100 perché i partecipanti hanno potuto optare per più opzioni.

Analizzando la tabella si evince che quando si legge per piacere, il 72% degli intervistati preferisce annotare il libro con il proprio nome, supportando l'idea della Jackson, precedentemente citata, per

cui i segni di proprietà sono quelli più comuni<sup>53</sup>. Di questi intervistati, il 68% preferisce scrivere su note adesive bianche, il 57% scrive al margine del testo.

Invece, quando si legge per lavoro o studio, 85%, degli intervistati usa note adesive bianche, l'81% scrive su note adesive e l'80% al margine.

Dai risultati del **test z** si evince un significativo aumento di marginalia (creare note marginali, evidenziazioni e la scrittura su note adesive e bianche) quando le persone leggono per lavoro o istruzione rispetto a quando leggono per piacere. Questi risultati confermano l'ipotesi H2.

La pratica di creare una dedica ha mostrato la tendenza opposta: i lettori sono significativamente più propensi a farlo quando leggono per piacere rispetto a quando leggono per lavoro o istruzione ( $p = 0,001$ ). Nessuna differenza significativa è emersa riguardo allo scrivere il proprio nome nel libro o al piegarne le pagine.

Al contrario, per gli e-book, sono state osservate poche differenze nel comportamento di annotazione quando si legge per intenti diversi, come riportato in Figura 2.3.

Innanzitutto, un numero molto inferiore di intervistati ha annotato gli e-book rispetto ai libri stampati; l'attività più comune risulta essere l'uso di segnalibri digitali (51% quando è per lettura di piacere e il 48% per lavoro o istruzione). Evidenziare e aggiungere commenti risultano attività meno comuni ma, ancora una volta, sono state rilevate più frequentemente quando si leggeva per lavoro o istruzione piuttosto che per piacere. Interessante notare come per il 93% degli intervistati, la ragione principale dell'annotare fosse per un "aiuto alla memoria" (87%), seguito da "lavorare sulle idee/analisi critica", (67%), "fare domande" (53%), "esprimere le proprie idee e pensieri", (52%), e "scrivere la proprietà dell'oggetto" (51%).

---

<sup>53</sup> Jackson, H.J. (2001). *Marginalia: Readers writing in books*. New Haven, CT: Yale University Press.

Marginalia practices for e-books.

Practice	Reading for				Difference significant? ( $p = 0.001$ )
	Pleasure		Work or education		
	N	%	N	%	
Type comment	122	24	180	35	N
Type inscription	44	9	47	9	N
Highlight passage	193	38	239	47	N
Add bookmark	260	51	247	48	N

Note: Percentages sum to more than 100% because respondents ( $n = 510$ ) were allowed to select more than one choice.

**Figura 2.3:** Marginalia negli e-book

Come riportato nelle Figura 2.2 e 2.3, gli intervistati sono più propensi a scrivere in un libro letto per lavoro o studio piuttosto che per piacere. Questo risultato supporta l'ipotesi H2 (le annotazioni sono più probabili quando si legge per lavoro o studio che per piacere).

Per i libri stampati, l'uso di note marginali, di iscrizioni, evidenziazioni e note adesive è statisticamente più comune quando si legge per lavoro piuttosto che per piacere; dunque, l'ipotesi nulla che sono ugualmente comuni è respinta (z-test,  $p = 0,001$ ).

Per i libri elettronici, le differenze sono molto più piccole; l'unico tipo di annotazione che è significativamente più comune quando si legge per lavoro è risultato l'uso di commenti digitati, che possiedono un livello di significatività basso (z-test,  $p = 0.05$ ).

Ci potrebbero essere molte ragioni per le differenze riscontrate quando si tratta di attività diverse, di seguito alcune risposte fornite dagli intervistati a cui è stato chiesto il perché di una maggiore annotazione in libri per fini lavorativi. Un intervistato ha risposto: "I libri di fiction sono in qualche modo più sacri per me!"; un altro intervistato: "Io non mi faccio scrupoli a scrivere sui libri per scopi educativi, l'ho usato come tecnica di studio. Ma per i libri d'intrattenimento, mi piace mantenerli il più incontaminati possibile, senza note[...]".

Infine, gli intervistati hanno espresso preferenze riguardo a quali strumenti vorrebbero fossero disponibili nei loro supporti elettronici. I seguenti strumenti sono in ordine di popolarità.<sup>54</sup>

1. Segnalibri;
2. Ricerca del testo;
3. Evidenziazione di passaggi;
4. Definizioni/accesso al dizionario;
5. Possibilità di copiare le selezioni per un uso successivo;
6. Note marginali.

### **2.2.3 Considerazioni finali sullo studio in esame**

L'uso di marginalia è una pratica che continua a dividere le opinioni dei lettori e anche se le persone leggono sempre più spesso e-book, specialmente per l'istruzione o il lavoro, il libro stampato rimane un oggetto a cui sono affezionati.

Dallo studio è emerso che l'arte di annotare è una pratica trasferitasi nella sfera digitale, specialmente per lavoro. Molte persone percepiscono i libri stampati, in particolare quelli letti per piacere, come "sacri" e ritengono quindi debbano rimanere uguali a se stessi, senza annotazioni. Al contrario, ritengono che annotare i libri elettronici sia accettabile e vantaggioso.

Un numero inferiore di intervistati si è impegnato in annotazioni digitali rispetto al numero di coloro che hanno scritto sui libri stampati. Questa apparente contraddizione potrebbe essere spiegata dalla relativa difficoltà di annotare negli attuali dispositivi di lettura elettronica, per tale motivo è importante considerare come queste pratiche possano e debbano essere migliorate.

---

<sup>54</sup> Bold, M. R., & Wagstaff, K. L. (2017). Marginalia in the digital age: Are digital reading devices meeting the needs of today's readers?. *Library & Information Science Research*, 39(1), 16-22.

Molti degli intervistati hanno dichiarato che desidererebbero avere la possibilità di inserire segnalibri, evidenziare passaggi e fare note a margine nei loro supporti elettronici ma che spesso queste opzioni non erano supportate o risultavano pratiche difficili.

La ricerca suggerisce che rendendo un'esperienza di e-reading più simile alla lettura di un libro fisico si possa incrementare l'attività di annotazione.<sup>55</sup> Infine, anche se molti degli intervistati hanno usato le annotazioni come note private, altri l'hanno trovata un'esperienza sociale.

La pratica dei marginalia, sia in stampa che in digitale è stata sempre un'esperienza legata al sociale, fornendo un modo per connettersi ai lettori passati e futuri, con la differenza che i lettori di libri stampati non potranno mai incontrare e parlare con gli annotatori anonimi o defunti dei loro libri, i lettori digitali, invece, hanno l'opportunità di confrontarsi con altri lettori attraverso i loro e-reader e le app di lettura.

## **2.3 Le annotazioni Web**

Dopo aver fatto un excursus storico sull'uso dei marginalia, discusso la loro evoluzione e il passaggio su supporti elettronici, in questa sezione mi focalizzerò sulle annotazioni web. Queste rappresentano un tentativo di ricreare ed estendere la possibilità di annotare testi le utilizzando Internet. L'idea è quella che chiunque dovrebbe avere la possibilità di annotare ovunque qualsiasi cosa, che sia una pagina web, un video, un'immagine, un flusso audio, o dati in forma grezza o visualizzata. Le annotazioni sul web, infatti, possono essere collegate, condivise tra servizi, cercate e memorizzate ovunque l'autore desideri<sup>56</sup>. Un'ulteriore evoluzione step nel campo delle annotazioni web è rappresentata dalle annotazioni web di tipo semantico, nelle quali le informazioni che si intendono esprimere devono appartenere ad un certo dominio di interesse; pertanto, va scelta l'ontologia più

---

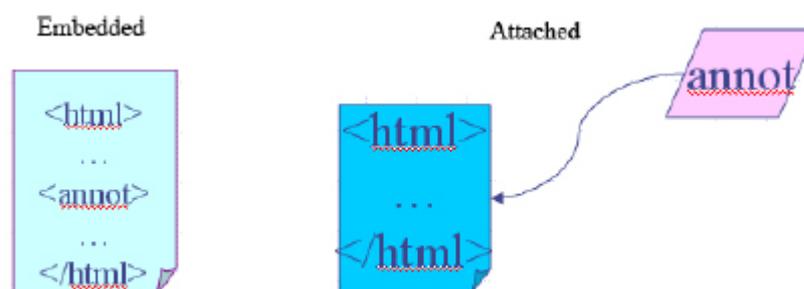
<sup>55</sup> Liesaputra, V., & Witten, I. H. (2012). Realistic electronic books. *International Journal of Human-Computer Studies*, 70, 588–610.

<sup>56</sup> <https://www.w3.org/annotation/>

affine al dominio da rappresentare. Un'ontologia è un framework formale per modellare la conoscenza che fornisce classi e proprietà che nominano e descrivono le risorse e le relazioni tra essi vigenti in un determinato dominio del discorso.<sup>57</sup>

Tecnicamente, come ampiamente illustrato nel primo capitolo di questa tesi, il Semantic Web rappresenta formalmente la conoscenza in forma di triple (soggetto- predicato – oggetto).

Da un punto di vista tecnico, le annotazioni semantiche sono concepite come metadati, pertanto si configurano come informazioni supplementari del documento e possono essere memorizzate localmente, su uno o più server oppure su server pubblici, disponibili a chiunque. Un'annotazione può essere, *embedded*, cioè inserita all'interno del documento annotato oppure *attached*, ovvero memorizzata altrove e collegata al documento tramite un link, un esempio visivo è mostrato nella Figura 2.4



**Figura 2.4:** Esempi di locazione di un'annotazione

In particolare, le annotazioni semantiche possono riguardare vari tipi di risorse: documenti, come frammenti di testo, pagine HTML, immagini; Web Service in cui viene descritto il compito di un

---

<sup>57</sup> Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology?. In Handbook on ontologies (pp. 1-17). Springer, Berlin, Heidelberg.

servizio, riportandone i parametri di input e output; strutture dati ovvero processi che coinvolgono lo scambio di informazioni tra applicazioni software che cooperano tra loro.

La classificazione delle annotazioni può variare in base al livello analizzato: dalla formalità del linguaggio usato, al destinatario dell'annotazione, al tipo di risorsa annotata, alla modalità adoperata. L'annotazione si definisce informale se il destinatario è un umano ed espressa in linguaggio naturale, nel caso in cui l'annotazione serva per esprimere il significato di una risorsa per un agente software deve essere espressa in un linguaggio formale come RDF, RDFS o OWL. In Figura 2.5 viene presentato un esempio di annotazione formale utilizzando RDF, nella successiva, Figura 2.6, un esempio informale.

Paolo Atzeni è un professore dell'università Roma TRE.

**Figura 2.5:** Esempio di annotazione informale in linguaggio naturale

```
<Professor rdf:about=""file:///C:/.../PagineHtml/PaoloAtzeni.htm#PaoloAtzeni">  
<first_Name> Paolo </first_Name>  
<last_Name> Atzeni </last_Name>  
<has_affiliation rdf:resource=""file:///C:/.../PaoloAtzeni.htm#UniRomaTRE"/>  
</Professor>
```

**Figura 2.6:** Esempio di annotazione formale

Infine, ci sono diversi livelli di restrizione sul linguaggio: *Absence* quando il linguaggio è libero da vincoli, *Advised* in cui si ha la libertà di scegliere se utilizzare un glossario, un'ontologia o il

linguaggio naturale; *Mandatory* in cui si deve utilizzare obbligatoriamente un'ontologia di riferimento ed è definita Annotazione Ontology-based.<sup>58</sup>

## 2.4 Il W3C Web Annotation

Il W3C Web Annotation Working Group fa parte del Publishing@W3C,<sup>59</sup> attività sotto la quale sono organizzati i gruppi legati all'editoria del W3C ovvero due gruppi di lavoro, due gruppi comunitari e un gruppo di affari. I gruppi di lavoro si occupano dello sviluppo e della supervisione degli standard formali del W3C, chiamati Raccomandazioni.

Tecnici e membri regolari del W3C si occupano di redirigere documenti dettagliati che descrivono il lavoro preso in carica da un determinato gruppo e la tempistica del suo compimento.

Il Web Annotation Group<sup>60</sup> non si interessa soltanto della pubblicazione dei dati ma, ha anche un'attenzione particolare per il loro contenuto. L'intento è sviluppare un set di specifiche per la creazione di una proposta di Web Annotation Architecture,<sup>61</sup> che serva a descrivere i servizi web necessari e i componenti fondamentali per un'annotazione. Il fine ultimo del Web Annotation Group è quello di creare uno schema annotativo che sia condivisibile, interoperabile e distribuibile.

Il W3C si è imposto questo obiettivo poiché annotare è un'attività molto diffusa in molti domini, Per esempio, la capacità di annotare su supporti elettronici è cruciale per l'avanzamento dell'e-learning,<sup>62</sup> metodologia e tecnica di apprendimento che si è rivelata indispensabile durante la pandemia mondiale di Covid-19.<sup>63</sup>

---

<sup>58</sup> Lezoche, M. (2004). *Annotazione di Documenti nel Web Semantico Document annotation in the Semantic Web*, pp.196. hal-00656684

<sup>59</sup> <https://www.w3.org/publishing/>

<sup>60</sup> <https://www.w3.org/annotation/>

<sup>61</sup> <https://www.w3.org/annotation/diagrams/annotation-architecture.svg>

<sup>62</sup> <https://www.treccani.it/enciclopedia/e-learning/>

<sup>63</sup> [https://www.ilsole24ore.com/art/verso-modello-ibrido-apprendimento-ADYBERs?refresh\\_ce=1](https://www.ilsole24ore.com/art/verso-modello-ibrido-apprendimento-ADYBERs?refresh_ce=1)

Ad oggi, non esiste un approccio strutturato e adeguato a creare annotazioni sul Web, il che rende l'attuale processo inutilmente complesso e dispendioso al livello di tempo.

Il focus che si prefigge il suddetto Working Group è quello di creare un approccio aperto per le annotazioni, realizzare un ecosistema in cui gli utenti abbiano accesso alle loro annotazioni, possano condividerle e archivarle.

Per realizzare ciò, sono stati individuati alcuni elementi ritenuti essenziali, ovvero:

- Un Abstract Data Model: un modello e un formato strutturato per permettere alle annotazioni di essere condivise e riutilizzate attraverso diverse piattaforme hardware e software;
- Un vocabolario: per definire/descrivere il Data Model;
- Un formato di serializzazione: che permetta di tradurre il modello astratto di dati in formati standard come JSON/JSON-LD/HTML;
- Una http API: una specifica API per creare, modificare, manipolare le annotazioni tramite http;
- Una Client Side API: un'interfaccia di script per facilitare la creazione di sistemi di annotazioni di browser;
- Un metodo di Robust Link Anchoring: cioè un meccanismo per poter selezionare una parte di testo o di media da annotare, garantendo flessibilità.

Il Web Annotation Working Group si è interessato di stilare delle Raccomandazioni per le Web Annotation sottoforma di tre documenti:<sup>64</sup> il Web Annotation Data Model per descrivere la struttura del modello e il formato utile, JSON, affinché le annotazioni possano essere condivise e riutilizzate su diverse piattaforme software e hardware;<sup>65</sup> Web Annotation Vocabulary che specifichi un insieme di classi RDF, predicati, entità, termini raccomandati da altre ontologie e fornisce il contesto JSON-

---

<sup>64</sup> <https://www.w3.org/blog/news/archives/6156>

<sup>65</sup> <https://www.w3.org/TR/annotation-model/>

LD e le definizioni di profilo necessarie per usare la serializzazione JSON in un contesto di Linked Data.<sup>66</sup> Infine, il Web Annotation Protocol che descriva i meccanismi attraverso cui si creano e gestiscono le annotazioni.<sup>67</sup>

L' Open Annotation Core Data Model,<sup>68</sup> invece, serve per realizzare un framework interoperabile creando associazioni tra le risorse, usando una metodologia conforme all'architettura del WWW. In questo modo, le annotazioni pubbliche possono essere condivise tra piattaforme mantenendo la ricchezza espressiva e la facilità dell'uso in ogni contesto.<sup>69</sup>

### 2.4.1 L'architettura del Web Annotation

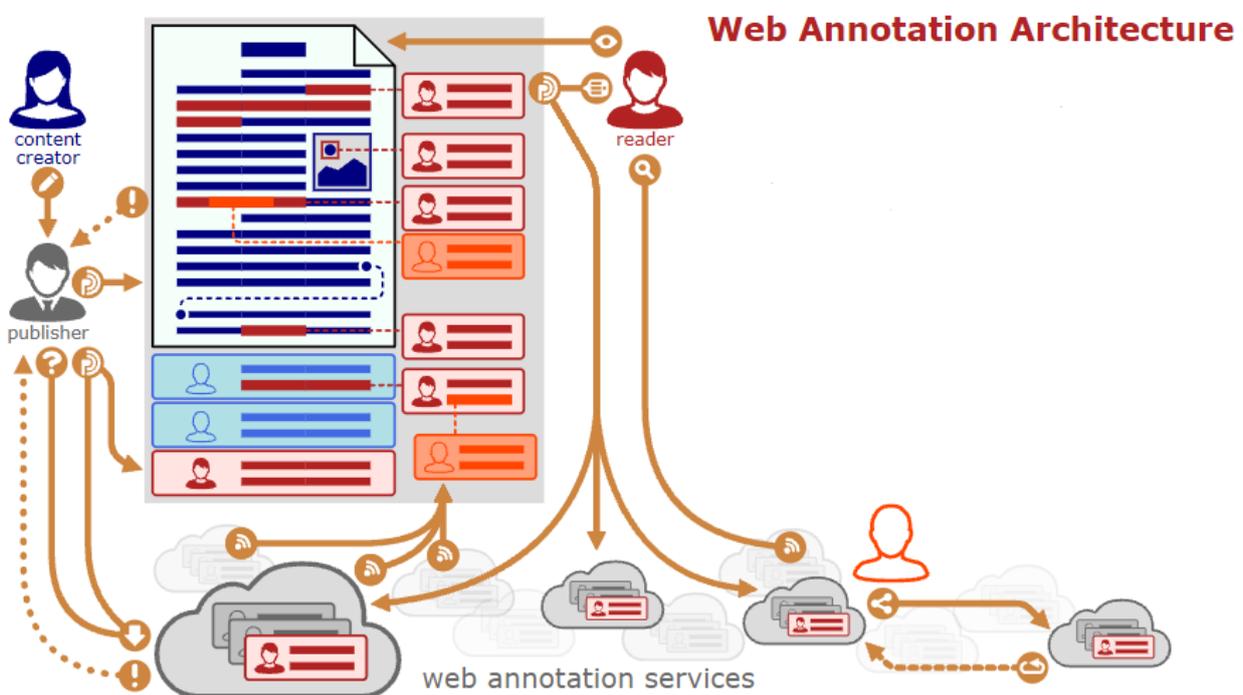


Figura 2.7: Architettura del Web Annotation

<sup>66</sup> <https://www.w3.org/TR/annotation-vocab/>

<sup>67</sup> <https://www.w3.org/TR/annotation-protocol/>

<sup>68</sup> <http://www.openannotation.org/spec/core/>

<sup>69</sup> <http://www.openannotation.org/spec/core/>

La Web Annotation Architecture è il modello ideale di annotazione proposto dal W3C Web Annotation Group, caratterizzato da una struttura decentralizzata e aperta.

Il W3C ha reso disponibile un diagramma interattivo<sup>70</sup> per fornire una visione più dettagliata della loro idea di architettura del Web Annotation. La Figura 2.7 propone il diagramma in cui sono mostrati i protagonisti fondamentali di questo processo e il loro scopo.

Il Content Creator crea il testo originale, le immagini ed altre risorse, ciò può avvenire in maniera collaborativa o singola.

Il Publisher è una figura che può coincidere con quella del Content Creator, il suo compito è distribuire il contenuto delle annotazioni sia su supporti elettronici come gli e-book, sia sul Web.

Il Content, ovvero il contenuto vero e proprio, pubblicato nella sua forma finale, può essere identificato da un IRI oppure da altri tipi di localizzatori come un ISBN, *International Standard Book Number*, un numero che identifica a livello internazionale e in modo univoco il contenuto di un volume pubblicato da un determinato editore. Nel contenuto possono esserci due tipologie di annotazione: note a piè di pagina e commenti.

Le note a piè di pagina possono essere ancorate ad una specifica sezione del contenuto, possono avere la stessa provenienza del contenuto principale oppure no.

I commenti dei lettori, invece, spesso non hanno la possibilità di legarsi ad una sezione specifica, anche se di recente sono nati sistemi di commento che permettono di creare una nota da parte del lettore in un punto esatto del paragrafo o della frase in questione. Funzionalità disponibili nei pdf di Adobe<sup>71</sup> oppure su Dropbox vi è la possibilità di aggiungere un commento in un punto specifico di un file video o audio.<sup>72</sup>

I commenti, se di alta qualità, sono importanti ai fini del miglioramento del contenuto perché i feedback, le correzioni, le informazioni aggiuntive possono generare dibattiti produttivi coinvolgendo

---

<sup>70</sup> <https://www.w3.org/annotation/diagrams/annotation-architecture.svg>

<sup>71</sup> <https://helpx.adobe.com/it/acrobat/using/commenting-pdfs.html>

<sup>72</sup> <https://help.dropbox.com/it-it/files-folders/share/comment-timestamp>

altri lettori. Il problema maggiore dei commenti riguarda il fatto che non sempre gli editori sono in grado di gestirli, di verificarli prima della pubblicazione.

Un agente fondamentale all'interno dell'architettura è sicuramente il Reader che può agire in diversi modi: in maniera passiva, leggendo e analizzando il contenuto del testo; in maniera attiva provvedendo a commentare, correggere, sottolineare sia il testo sia un contenuto multimediale come un'immagine, una posizione su una mappa interattiva, un video.

Per quanto riguarda la pubblicazione del contenuto, il lettore non deve necessariamente pubblicarlo sul sito di appartenenza ma, può avvalersi sia di servizi di annotazioni di terze parti sia di servizi offerti sul proprio dispositivo. Questa decentralizzazione è il punto di forza delle Web Annotation, in quanto il lettore che crea i contenuti può decidere come e dove pubblicare ciò che scrive.

Il Publisher riceve una notifica per ogni annotazione pubblica fatta sul sito, questo meccanismo informativo avviene in due modi: un evento lato client che informa la pagina degli URL di pubblicazione delle annotazioni; una notifica lato server inviata all'indirizzo di contatto noto dell'editore. Oltre alla ricezione della notifica, l'editore può richiedere maggiori informazioni sull'annotazione, incluse le credenziali dell'autore, la licenza di pubblicazione e può anche recuperare l'annotazione stessa.

Uno dei tanti vantaggi delle annotazioni web è che offrono un modo alternativo di scoprire nuovi contenuti, nuove discussioni su di essi, nuovi individui o comunità. I servizi di annotazione possono permettere la visualizzazione singola di un'annotazione nel suo contesto nativo o all'interno di una collezione, ciò avviene grazie a servizi come Reddit, Twitter o Facebook.

I lettori possono fare ricerche sfaccettate che abbracciano più siti annotati, filtrandole ad esempio per argomenti specifici o hashtag che sarebbero difficili da fare con sistemi di commenti relegati ad un singolo e specifico sito.

Una volta che un lettore scopre un argomento, una comunità o un servizio interessante, può iscriversi a quel determinato *topic* e visualizzare le annotazioni quando visita un URL per lui rilevante.

Le annotazioni sono caricate da uno o più servizi, visualizzate nella pagina e ancorate dinamicamente alla specifica sezione del contenuto, mostrando l'annotazione nel contesto. Attualmente, questa funzionalità è disponibile soltanto mediante estensioni JavaScript del browser ma l'obiettivo finale è quello di creare un supporto nativo del browser.

Inoltre, il servizio mira alla condivisione, ovvero quando un utente trova un'annotazione che gli piace può condividerla con altri mediante un permalink, ripubblicarla annotandola con il proprio contenuto.

Il vantaggio del Data model dell'Open Annotation è che fornisce uno standard dell'annotazione tale per cui un'annotazione web può essere condivisa sullo stesso servizio di annotazione o su un diverso servizio, garantendo in tal modo l'interoperabilità.

Un'attenzione particolare è data al creatore del contenuto originale a cui si può sempre risalire poiché ogni annotazione include la provenienza, l'IRI originale e una "catena di attribuzioni" che mostra tutti coloro che l'hanno condiviso.<sup>73</sup>

---

<sup>73</sup> <https://www.w3.org/annotation/diagrams/annotation-architecture.svg>

## CAPITOLO 3

### 3.1 Contesto nativo di Pundit: l'azienda Net7

Net7<sup>74</sup> è un'azienda pisana che è attiva da vent'anni nel settore della progettazione di soluzioni per il web, per l'industria 4.0,<sup>75</sup> l'e-commerce,<sup>76</sup> per le Digital Humanities e l'Open Science, si occupa di UX/UI design,<sup>77</sup> campagne di comunicazione web e tradizionale. Inoltre, è da sempre impegnata in progetti europei, in collaborazione con aziende ed enti pubblici.

Tra i suoi settori di interesse vi è il Semantic Web (SW), evoluzione del tradizionale concetto di Web, che permette di creare con maggiore facilità applicazioni che utilizzano e integrano dati provenienti da fonti eterogenee. Il SW è utilizzato in diversi contesti sia di ricerca, per esempio nel mondo delle Digital Humanities,<sup>78</sup> o in contesti commerciali e aziendali. In quest'ultimo ambito, le tecnologie semantiche cercano di facilitare l'accesso ai dati attraverso una visione logica di "alto livello", ovvero indipendentemente dal modo in cui questi sono memorizzati nei database, utilizzando tecniche di estrazione automatica delle informazioni per individuare concetti nei testi oppure favorire operazioni di classificazione di documenti.<sup>79</sup>

Nel 2011 l'azienda Net7 ha ideato **Pundit**, un tool per realizzare annotazioni web. A questo tool l'azienda lavora ancora, costantemente, per migliorarlo, integrandolo in diversi progetti al fine di sviluppare a pieno il suo potenziale.

A dimostrazione dell'impegno costante dell'azienda verso un approccio semantico nella gestione dei dati, è opportuno citare anche altri progetti in cui ha collaborato. Di seguito, si riporta una breve

---

<sup>74</sup> <https://www.netseven.it/chi-siamo/>

<sup>75</sup> [https://it.wikipedia.org/wiki/Industria\\_4.0](https://it.wikipedia.org/wiki/Industria_4.0)

<sup>76</sup> [https://it.wikipedia.org/wiki/Commercio\\_elettronico](https://it.wikipedia.org/wiki/Commercio_elettronico)

<sup>77</sup> <https://www.italiaonline.it/risorse/ux-e-ui-cosa-sono-differenze-e-importanza-2726>

<sup>78</sup> <https://www.netseven.it/servizi/digital-humanities/>

<sup>79</sup> <https://www.netseven.it/servizi/open-data/>

panoramica. **SemLib** (SEMantic tools for digital LIBraries) è un progetto europeo il cui obiettivo è stato lo sviluppo di un sistema di annotazione semantica di oggetti digitali e pagine web.<sup>80</sup>

L'**EuropeanaSounds**,<sup>81</sup> progetto europeo, il cui obiettivo è stato rendere disponibili oltre 540.000 oggetti audiovisivi digitali. Le registrazioni sonore selezionate per il progetto comprendono l'intera varietà del patrimonio culturale europeo: musica classica e popolare, narrazioni; effetti sonori e suoni ambientali; lingue, dialetti e memorie orali. Queste collezioni riflettono le diverse culture e storie europee degli ultimi 130 anni. In questo caso, Net7 ha contribuito all'arricchimento semantico delle registrazioni sonore e dei video tramite Pundit.

**Muruca**<sup>82</sup> è un framework realizzato da Net7<sup>83</sup> volto alla costruzione di biblioteche digitali per pubblicare edizioni di testi e di immagini, corredate da apparati critici<sup>84</sup> realizzati grazie alla sezione di annotazione semantica di Pundit. È una soluzione flessibile e accessibile che permette di passare dal testo scritto al contenuto multimediale. Muruca è compatibile con gli standard di catalogazione internazionali e si rivolge principalmente a gruppi di ricerca, musei o fondazioni culturali che vogliono inserire online i propri contenuti culturali.

Infine, merita di essere annoverato un recente progetto, ancora in corso, chiamato **Triple**,<sup>85</sup> acronimo di *Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration*, finanziato dalla Commissione Europea nell'ambito di Horizon 2020.<sup>86</sup> Triple mira ad aiutare la ricerca europea in questo ambito ad ottenere maggiore visibilità, ad essere più efficiente, a migliorare il suo utilizzo e ad aumentare il suo impatto sulla società. Il progetto vorrebbe creare nuovi modi di condurre, connettere e scoprire la ricerca; promuovere la diversità culturale sostenendo applicazioni scientifiche, industriali e sociali delle scienze SSH.<sup>87</sup>

---

<sup>80</sup> <https://www.netseven.it/project/semlib/>

<sup>81</sup> <https://www.netseven.it/project/europeana-sounds/>

<sup>82</sup> <https://www.muruca.org/>

<sup>83</sup> <https://www.netseven.it/project/muruca/>

<sup>84</sup> [https://it.wikipedia.org/wiki/Apparato\\_critico](https://it.wikipedia.org/wiki/Apparato_critico)

<sup>85</sup> <https://project.gotriple.eu/>

<sup>86</sup> <https://ec.europa.eu/programmes/horizon2020/what-horizon-2020>

<sup>87</sup> <https://operas.hypotheses.org/2644>

Gli scopi che il progetto europeo si prefigge sono messi in pratica grazie ad un'innovativa piattaforma, GoTriple che sarà un servizio di OPERAS,<sup>88</sup> infrastruttura europea di ricerca per lo sviluppo di una comunicazione accademica aperta nell'ambito delle scienze umane e sociali (SSH). La piattaforma punta alla scoperta multilinguistica delle SSH in nove lingue europee (croato, inglese, francese, tedesco, greco, italiano, polacco, portoghese, spagnolo), fornendo un unico punto di accesso per utilizzare pubblicazioni, dati di ricerca, descrizioni di progetti e profili di ricercatori che sono attualmente sparsi in archivi locali e specifici.<sup>89</sup>

Tra i partner di questo progetto vi è Net7 che provvede a dare il suo contributo alla piattaforma GoTriple mediante l'integrazione al suo interno del tool di annotazione Pundit.

Lo scopo è quello di fornire agli utenti di GoTriple un servizio a supporto delle loro attività quotidiane di ricerca, grazie alla possibilità di creare annotazioni di varia natura (highlights, commenti, tag ma anche annotazioni semantiche) su contenuti web.<sup>90</sup>

### 3.2 Pundit: pin the web

**Pundit**<sup>91</sup> è un tool di annotazione web, appartenente ai servizi per l'umanistica digitale, utilizzato liberamente da tutti come servizio cloud<sup>92</sup> e gestito dall'azienda Net7.

Esso permette di evidenziare e/o commentare passaggi di pagine web, creando di conseguenza annotazioni che sono raccolte in notebook. Ogni annotazione è associata ad un notebook e ogni notebook è associato ad un utente.

---

<sup>88</sup><https://operas.hypotheses.org/>

<sup>89</sup> <https://project.gotriple.eu/gotriple-platform/>

<sup>90</sup> De Santis Luca, (2021). Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration, Report on the Open Annotation Tool.

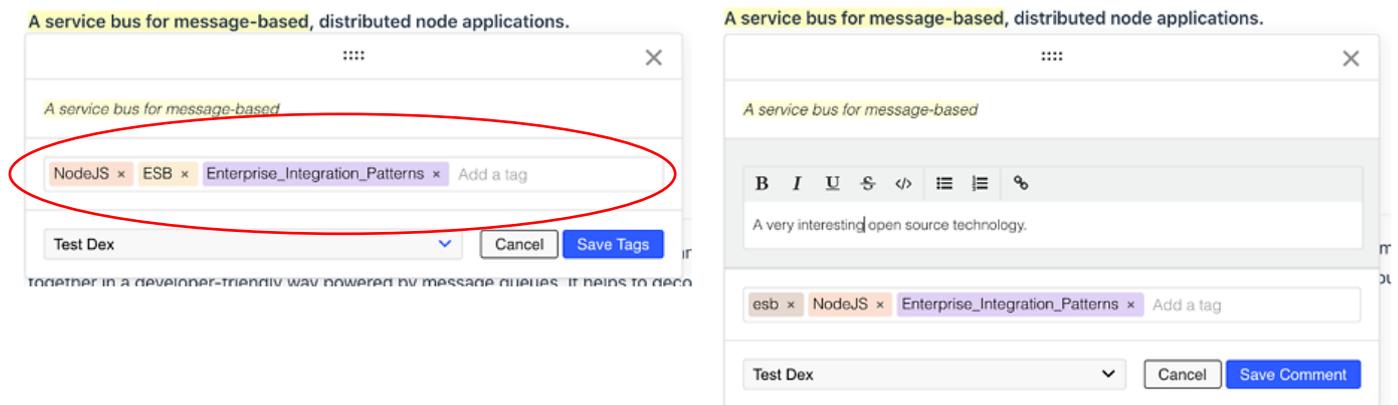
<sup>91</sup> <https://thepund.it/>

<sup>92</sup> [https://it.wikipedia.org/wiki/Cloud\\_computing](https://it.wikipedia.org/wiki/Cloud_computing)

Gli utenti possono avere uno o più notebook, possono decidere se renderli pubblici oppure salvarli in forma privata. Tra le nuove funzionalità in sviluppo, c'è anche la possibilità di condividere i propri notebook con altri utenti.<sup>93</sup>

Un'altra azione possibile riguarda la creazione di annotazioni per etichettare un testo selezionato o un'intera pagina web, dando vita ad una categoria personale in forma libera chiamata **tag** e costituita da una stringa alfanumerica di massimo 128 caratteri. È un modo agile per classificare le loro annotazioni, da questo punto di vista può essere considerata come la forma più “semplice” di annotazione semantica fornita. Gli utenti possono aggiungere tag ai loro commenti e alle loro annotazioni semantiche.

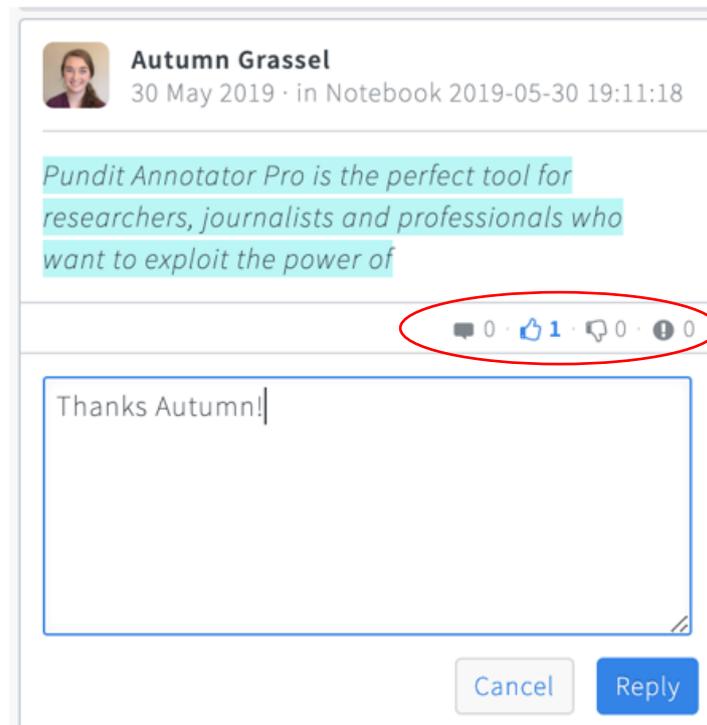
Inoltre, nel momento in cui l'utente attiva Pundit, tutti i suoi tag vengono automaticamente caricati e, attraverso il completamento automatico, suggeriti durante l'operazione di tagging; un esempio di ciò è riportato nella Figura 3.1.



**Figura 3.1:** Esempio di tagging

<sup>93</sup> De Santis Luca, (2021). Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration, Report on the Open Annotation Tool.

Il tool fornisce anche la possibilità di utilizzare le tipiche **social features** come: reply, like, dislike e il poter segnalare un commento ritenuto inopportuno all'amministratore di Pundit. In Figura 3.2 è possibile osservare, al di sotto di un'annotazione, quattro icone che permettono rispettivamente di: creare un commento, mettere like (pollice in su), dislike (pollice in giù) e fare una segnalazione.



**Figura 3.2:** Social Features

Pundit ha una caratteristica distintiva rappresentata dal supporto di annotazioni semantiche, con le quali è possibile creare un'annotazione sottoforma di tripla RDF,<sup>94</sup> composta da soggetto/predicato/oggetto. Nello specifico:

- Il soggetto è rappresentato da un frammento di testo o dall'intero documento web;

<sup>94</sup> [https://en.wikipedia.org/wiki/Semantic\\_triple](https://en.wikipedia.org/wiki/Semantic_triple)

- Il predicato definisce una relazione formale tra il soggetto e l'oggetto, secondo regole specificate in una determinata ontologia del Semantic Web;
- L'oggetto può essere: un letterale (ad esempio una stringa), un IRI che identifica una risorsa esterna o un'entità definita in un'ontologia.

Questa funzionalità può rivelarsi estremamente utile per professionisti che hanno la necessità di creare "note" in modo formale e strutturato.

Di seguito è riportato un esempio chiarificatore di annotazione semantica realizzabile in Pundit.

The screenshot shows a web interface for creating semantic annotations. At the top, the text "Mexican poetry" is highlighted in yellow. Below it, a section titled "Semantic relations" contains two rows of input fields. The first row has a dropdown menu set to "identifies", a text input field containing the URL "https://en.wikipedia.org/wiki/Mexican\_literature", a "URI" label, and a vertical ellipsis icon. The second row has a dropdown menu set to "describes", a text input field containing the string "I poeti della frontiera con gli Stati Uniti", a "LITERAL" label, and a vertical ellipsis icon. Below these rows is a text input field with the placeholder "Add a tag". At the bottom of the interface, there is a "Test Dex" dropdown menu, a "Cancel" button, and a "Save" button.

**Figura 3.3:** Esempio di annotazione semantica

Nella Figura 3.3 , sono presenti due annotazioni semantiche aventi come **soggetto**, il frammento di testo "Mexican poetry" per entrambe le triple; due **predicati**, "identifies" e "describes" e come **oggetto**, rispettivamente un'entità di Wikipedia e un letterale.

Il vantaggio derivante dal poter creare un'annotazione di questo tipo è quello di avere i dati formalmente definiti e comprensibili dalla macchina, offrendo migliori possibilità di ricerca, riutilizzo e sfruttamento delle annotazioni effettuate dagli utenti.

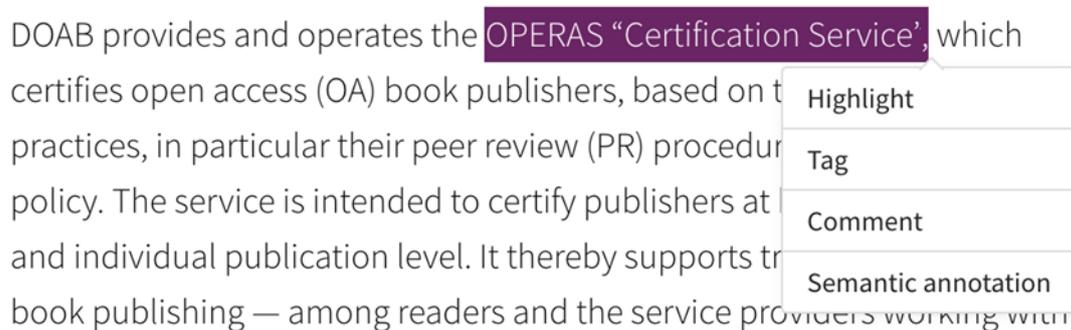
Una definizione semplice ma efficace di annotazione semantica è proposta in “*Semantic Annotation, Indexing, and Retrieval, Web Semantics: Science, Services and Agents on the World Wide Web*”<sup>95</sup> in cui si afferma che "l'annotazione semantica consiste nell'assegnare alle entità del testo collegamenti alle loro descrizioni semantiche”.

I punti chiave per fornire "descrizioni semantiche" efficaci sono, in primo luogo, un'attenta definizione delle ontologie, ovvero il vocabolario di termini, classi, predicati e proprietà secondo cui le annotazioni semantiche vengono eseguite dagli utenti.

In secondo luogo, un uso per quanto possibile di ontologie standard, il cui significato è ben noto e accettato, indispensabile per permettere il riutilizzo delle annotazioni.

Infine, un sapiente utilizzo dei Linked Data, ad esempio come oggetti della tripla semantica. Essi riguardano la pubblicazione di dati strutturati sul web e dataset come Geonames,<sup>96</sup> Wikidata,<sup>97</sup>, DBPedia<sup>98</sup> che forniscono sia enormi insiemi di entità a cui si può fare riferimento nel processo di annotazione, sia famiglie di ontologie standard.

In Figura 3.4, sono riassunte le operazioni possibili, finora discusse, utilizzando il tool.



**Figura 3.4:** Azioni possibili con Pundit

<sup>95</sup> Kiryakov A. et al., *Semantic Annotation, Indexing, and Retrieval, Web Semantics: Science, Services and Agents on the World Wide Web*, Volume 2, Issue 1, pp. 49-79, 2004

<sup>96</sup> <http://www.geonames.org/>

<sup>97</sup> [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)

<sup>98</sup> <https://www.dbpedia.org/>

### 3.2.1 I servizi offerti

I servizi principali offerti da Pundit possono essere divisi in due sezioni: la prima è quella Front-End Services per gli utenti, costituita dall'**Annotator**, prima componente, strumento utilizzato per creare le annotazioni e implementato come un'estensione di Google Chrome.

Una seconda componente è l'**App Pundit** tramite cui bisogna registrarsi, inserendo i propri dati, per poter usufruire del servizio.

Infine, la terza componente è rappresentata dal **FeedThePundit**, servizio che sarà presto messo a disposizione per coloro che non hanno intenzione di installare l'estensione Chrome o di utilizzare il suddetto browser. In tal modo, gli utenti di Pundit possono utilizzarlo su qualsiasi browser e su qualsiasi device.

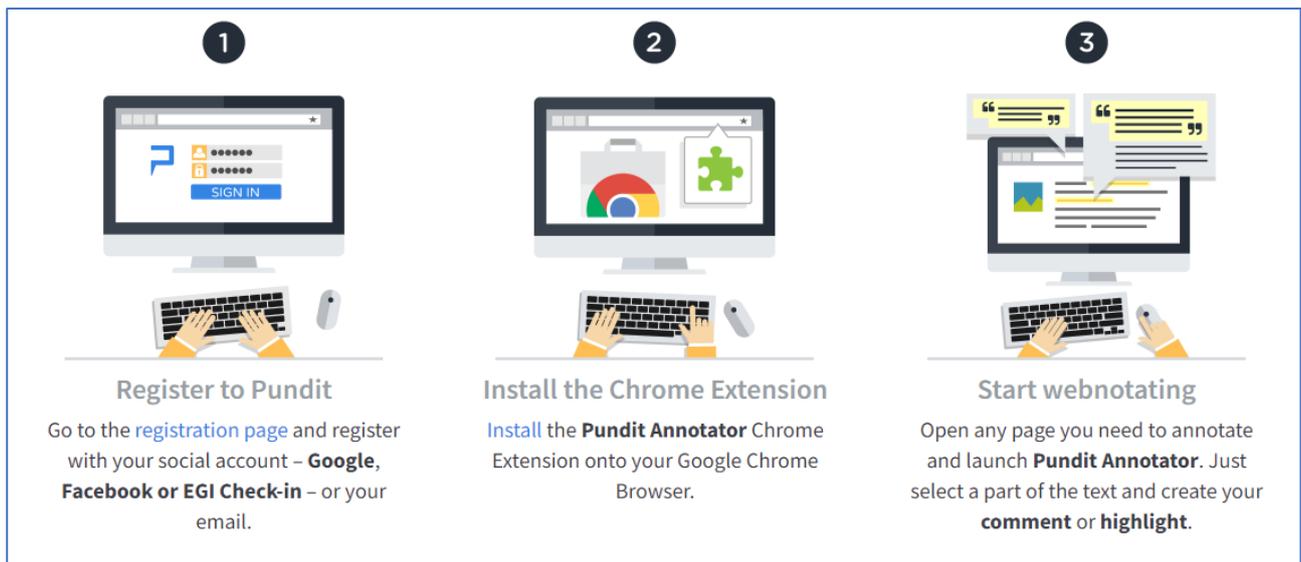
La seconda sezione è di tipo Back-End Service, riservata agli sviluppatori, e dispone pubblicamente di una serie di API,<sup>99</sup> che permettono lo sviluppo di possibili scenari di integrazione personalizzati con il servizio.<sup>100</sup>

In termini pratici, l'utente che ha intenzione di utilizzare questo tool deve effettuare tre semplici passaggi, mostrati in Figura 3.5.

---

<sup>99</sup> [https://it.wikipedia.org/wiki/Application\\_programming\\_interface](https://it.wikipedia.org/wiki/Application_programming_interface)

<sup>100</sup> De Santis Luca, (2021). Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration, Report on the Open Annotation Tool.



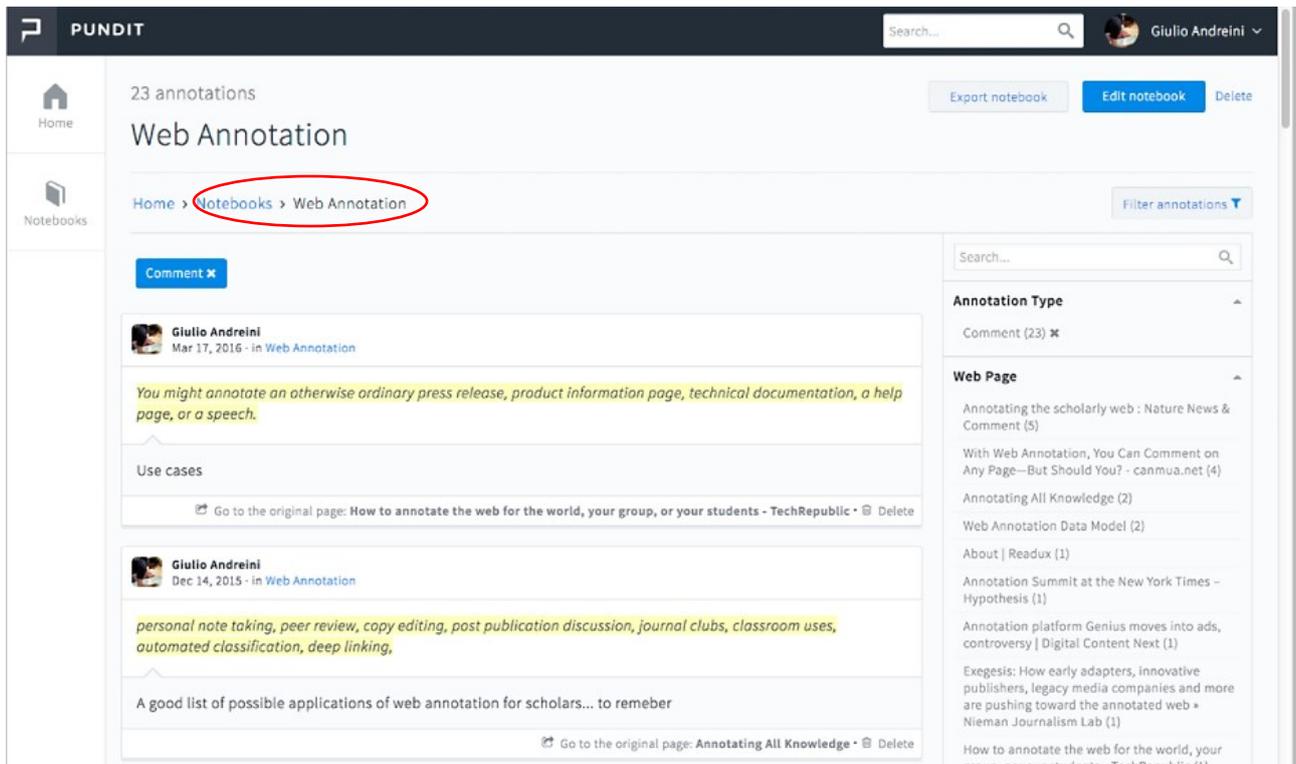
**Figura 3.5:** Passaggi per utilizzare Pundit

Innanzitutto, bisogna registrarsi alla pagina di riferimento<sup>101</sup> tramite un account Google, un social network o un'e-mail, successivamente bisogna installare la Pundit Annotator Chrome extension nel proprio browser<sup>102</sup> e infine iniziare ad utilizzare il tool come si preferisce.

Il tool oltre a gestire la registrazione e l'autenticazione degli utenti, fornisce anche una dashboard web dove è possibile vedere, cercare ed esportare tutte le annotazioni che gli utenti hanno realizzato sul web e che sono memorizzate nei loro notebook. Un aspetto positivo è che i notebook possono essere esportati in formati esterni così da poterli riutilizzare come si preferisce, ad esempio docx, JSON-LD. In Figura 3.6 è presentato un esempio di dashboard in cui sono visibili i diversi notebook con le relative annotazioni.

<sup>101</sup> <https://app.thepund.it/register>

<sup>102</sup> <https://chrome.google.com/webstore/detail/pundit-annotator/eilojhemeemnpfmailfeiggpgilblaco>



**Figura 3.6:** Pundit Dashboard

### 3.3 Architettura software di Pundit: versione iniziale

Pundit è un tool in continua evoluzione, molta strada è stata fatta dal suo primo rilascio risalente al 2011.<sup>103</sup> La sua iniziale architettura presentava diversi problemi che rendevano difficile e non soddisfacente l’esperienza degli utenti. In primis, l’Annotator era basato sull’obsoleta tecnologia AngularJS;<sup>104</sup> inoltre, spesso non funzionava, parzialmente o totalmente, sui siti web moderni.

Un esempio di come la precedente versione di Pundit si comportasse male su alcuni siti web è illustrato nella seguente Figura 3.7 che mostra il comportamento errato del servizio sul sito “Isidore.science”.<sup>105</sup>

<sup>103</sup> Morbidoni C. et al., *Pundit: Semantically structured annotations for web contents and digital libraries*, Proceedings of the 2nd International Workshop on Semantic Digital Archives, 2012

<sup>104</sup> [https://www.w3schools.com/angular/angular\\_intro.asp](https://www.w3schools.com/angular/angular_intro.asp)

<sup>105</sup> <https://isidore.science/>



**Figura 3.7:** Esempio di funzionamento errato di Pundit

Inoltre, il back-end del servizio, l'Annotation Server, era basato su un'architettura monolitica<sup>106</sup> sviluppata in Java che era diventata molto difficile da mantenere ed inefficiente in termini di prestazioni. In particolare, il protocollo usato dall'Annotator per comunicare con le API era davvero ridondante, con molte duplicazioni e dati non necessari.

Anche le strategie di gestione dei dati, basate su un Triple Store RDF<sup>107</sup> (GraphDB di Ontotext<sup>108</sup>) in cui erano memorizzate le annotazioni e i dati dei notebook, si sono rivelate non ottimali. Era utilizzato un modello semantico ispirato, ma non perfettamente conforme, al W3C Web Annotation Data Model.<sup>109</sup> La scelta di utilizzare un database a grafi RDF non aveva senso, perché nessuna delle funzionalità specifiche che questo tipo di tecnologia garantisce (ad esempio la navigazione a grafi attraverso i dati, l'inferenza/ragionamento) veniva realmente utilizzata in Pundit.

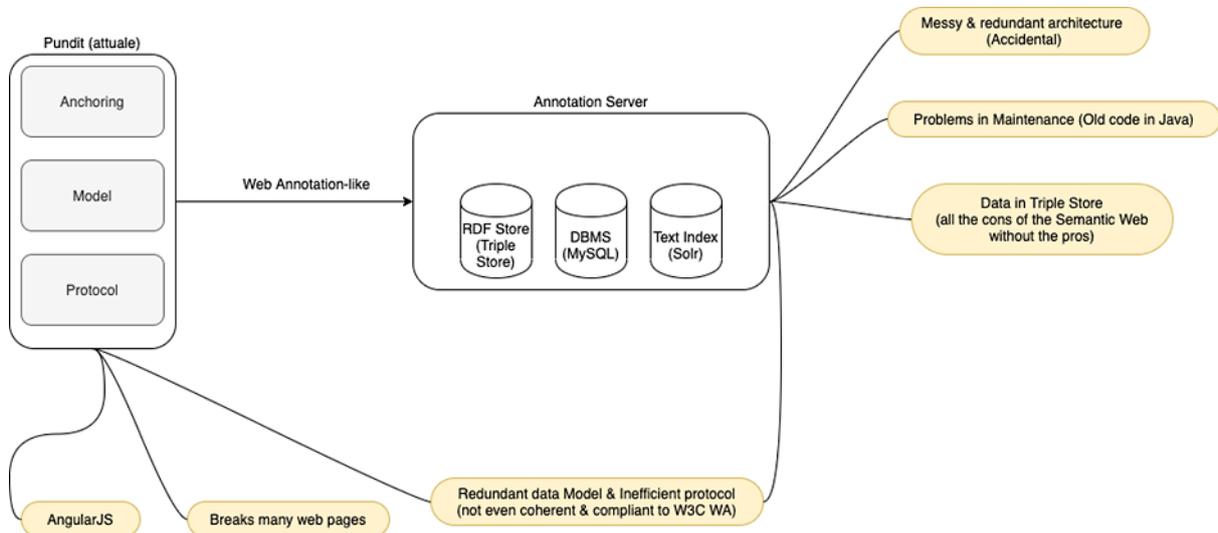
<sup>106</sup> <https://docs.microsoft.com/it-it/dotnet/architecture/containerized-lifecycle/design-develop-containerized-apps/monolithic-applications>

<sup>107</sup> <https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf-triplestore/>

<sup>108</sup> <https://www.ontotext.com/products/graphdb/>

<sup>109</sup> <https://www.w3.org/TR/annotation-model/>

Nella Figura 3.8, è mostrata l'architettura iniziale del tool di annotazione con alcuni dei maggiori problemi come: l'uso di AngularJs, la ridondanza del Data Model e l'inefficienza del protocollo perché non coerente allo standard W3C. Per quanto concerne l'Annotation Server invece i problemi erano relativi all'uso del Triple Store per i dati, ad un'architettura ridondante e confusionaria.



**Figure 3.8:** Una sintesi dei problemi della precedente versione

### 3.3.1 Cambiamenti necessari all'architettura pregressa

L'architettura di Pundit ha subito, negli ultimi anni, un'ampia evoluzione, in particolare, il codice dell'Annotator è stato completamente riscritto e non è stato più basato su AngularJS ma sulla sua evoluzione Angular.<sup>110</sup> Per quanto concerne la logica precedente dell'anchoring, cioè il meccanismo con cui le annotazioni possono essere attaccate correttamente alla parte di una pagina HTML, non era efficiente e resiliente ai cambiamenti della struttura della pagina. Per esempio, un nuovo progetto

<sup>110</sup> <https://angular.io/>

grafico poteva causare un fallimento nel riconoscimento del frammento di testo dove l'annotazione doveva essere attaccata.

Per tale motivo, il modulo di ancoraggio è stato completamente riscritto utilizzando il codice open source dello strumento di annotazione web Hypothes.is.<sup>111</sup>

L'aggiornamento più importante è stato fatto sul back-end di Pundit, cioè sull'Annotation Server e sulle sue API. **L'Annotation Server** è un insieme di servizi, implementati come applicazioni Node.js<sup>112</sup> ma, distribuite su Amazon AWS<sup>113</sup> come applicazioni Lambda,<sup>114</sup> che condividono una parte del loro codice ed espongono le loro funzionalità tramite API.

Il **Data layer** è stato aggiornato in quanto, nella vecchia versione, i dati erano memorizzati in un Triple Store RDF, non completamente conforme, al W3C Web Annotation Data Model.

Al fine di rendere le ricerche più efficienti, i dati delle annotazioni sono stati anche copiati nel motore di ricerca testuale Apache Solr.<sup>115</sup>

Inoltre, per i dati degli utenti (profilo, informazioni di autenticazione, preferenze) è stato utilizzato, come nella precedente versione, il database relazionale MariaDB,<sup>116</sup> pienamente compatibile con il più noto MySQL<sup>117</sup>. Invece, i dati relativi alle annotazioni sono memorizzati su Elasticsearch.<sup>118</sup>

Questo refactoring generale, sia sul lato Annotator che sul lato Annotation Server, ha portato a ridisegnare anche il protocollo di comunicazione tra di loro, ora basato su API REST<sup>119</sup>, le quali forniscono un modo flessibile e leggero per integrare le applicazioni e rappresentano il metodo più comune per connettere i componenti nelle architetture di microservizi,<sup>120</sup> e scambiandosi dati in formato JSON.<sup>121</sup>

---

<sup>111</sup> <https://web.hypothes.is/>

<sup>112</sup> <https://it.wikipedia.org/wiki/Node.js>, <https://nodejs.org/it/>

<sup>113</sup> <https://aws.amazon.com/it/>

<sup>114</sup> [https://docs.aws.amazon.com/it\\_it/lambda/latest/dg/deploying-lambda-apps.html](https://docs.aws.amazon.com/it_it/lambda/latest/dg/deploying-lambda-apps.html)

<sup>115</sup> <https://solr.apache.org/>

<sup>116</sup> <https://mariadb.org/>

<sup>117</sup> <https://www.mysql.com/it/>

<sup>118</sup> <https://www.elastic.co/>

<sup>119</sup> <https://www.ibm.com/it-it/cloud/learn/rest-apis>

<sup>120</sup> <https://aws.amazon.com/it/microservices/>

<sup>121</sup> <https://www.json.org/json-it.html>

È stata inoltre implementata una API di ricerca pubblica con il duplice intento di:

- Fornire un modo intuitivo e moderno per gli sviluppatori di recuperare le annotazioni utilizzando il linguaggio di query GraphQL<sup>122</sup>;
- Recuperare i dati in formato JSON-LD,<sup>123</sup> totalmente conforme al W3C Web Annotation Data Model.<sup>124</sup>

Infine, anche la **Pundit App** ha subito una trasformazione, è infatti stata implementata in PHP<sup>125</sup> utilizzando il framework Laravel.<sup>126</sup> Dal punto di vista del design utente, invece, la sua interfaccia non è cambiata molto dalla versione precedente.

Lo scopo principale dell'app è duplice: da un lato gestisce la registrazione e l'autenticazione degli utenti; dall'altro fornisce agli utenti di Pundit la dashboard web dove possono vedere, cercare ed esportare tutte le annotazioni che hanno realizzato sul web.<sup>127</sup>

---

<sup>122</sup> <https://graphql.org/>

<sup>123</sup> <https://json-ld.org/>

<sup>124</sup> <https://www.w3.org/TR/annotation-model/>

<sup>125</sup> <https://it.wikipedia.org/wiki/PHP>

<sup>126</sup> <https://laravel.com/>

<sup>127</sup> De Santis Luca, (2021). Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration, Report on the Open Annotation Tool.

### 3.4 Architettura software di Pundit: versione attuale

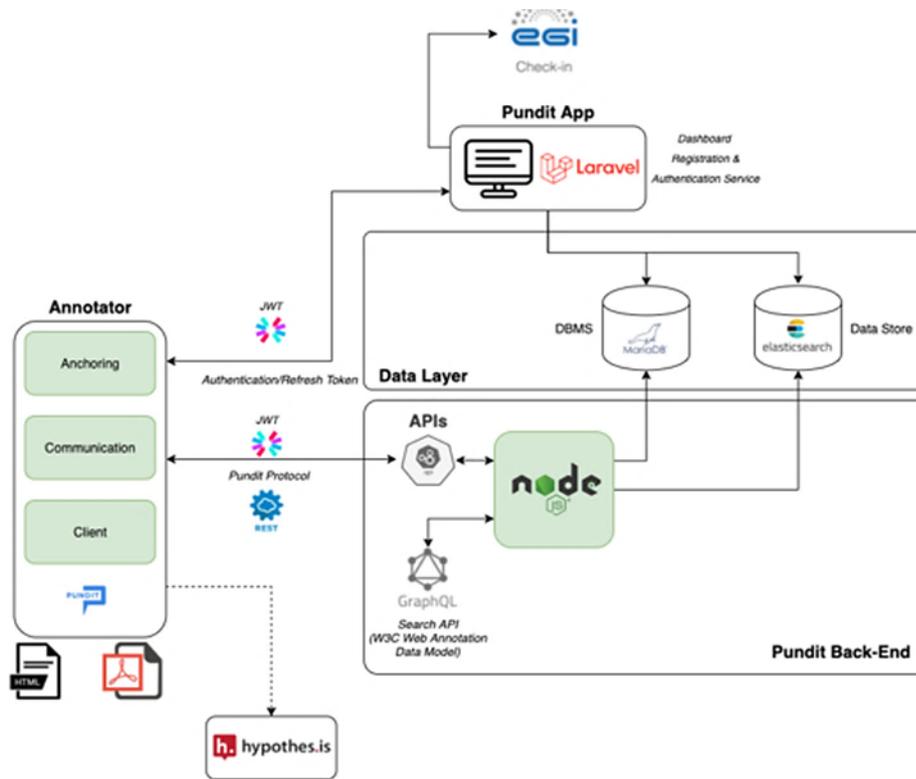


Figura 3.9: Architettura software aggiornata

Nella Figura 3.9 è mostrata la versione aggiornata dell'architettura software di Pundit.

l'**Annotator**, ovvero un'applicazione web, può essere usata sia come estensione Chrome sia incorporata in qualsiasi sito web.

L'Annotator si articola in tre parti:

1) l'**Anchoring** si occupa della corrispondenza tra le annotazioni ed evidenziazioni degli utenti e gli elementi strutturali di una pagina web, questo modulo è progettato sulla logica di Hypothes.is, altro tool di annotazione.

2) Il modulo **Communication** si occupa di gestire le richieste web tra il client di Pundit e il server<sup>128</sup> di annotazione.

3) Il modulo **Client** invece è quello con cui gli utenti interagiscono direttamente, è composto da una barra laterale che appare sul lato destro della pagina web nel momento in cui si effettua il login in Pundit tramite cui l'utente può consultare sia le sue annotazioni realizzate in precedenza sia quelle pubbliche di altri utenti.

### 3.4.1 Il back end di Pundit: W3C Search endpoint

Per quanto concerne il back end di Pundit, l'Annotation server, oltre alle caratteristiche sopra citate, propone un endpoint flessibile per cercare con semplicità le annotazioni generate con Pundit. L'endpoint è ad accesso pubblico ed è compatibile con GraphQL,<sup>129</sup> un query language per le API, che rispetta il modello di dati della specifica W3C Web Annotation.

Ad oggi, il W3C Web Annotation Protocol non fornisce uno standard per specificare come le annotazioni di una pagina identificata da un URL possano essere recuperate.

Per arginare questa problematica, si è pensato ad un'API flessibile implementata con GraphQL, il quale permette di selezionare gli elementi da restituire nella risposta di ricerca.

Inoltre, è stato deciso di restituire i dati in un formato compatibile con il W3C Annotation Data Model<sup>130</sup> ovvero sottoforma di entità Annotation<sup>131</sup> contenute in una sequenza di Annotation Pages.<sup>132</sup>

Il formato è JSON-LD, formato di serializzazione RDF che è stato scelto dal gruppo di ricerca del W3C per la raccomandazione ufficiale del Web Annotation Data Model.

---

<sup>128</sup> [https://it.wikipedia.org/wiki/Sistema\\_client/server](https://it.wikipedia.org/wiki/Sistema_client/server)

<sup>129</sup> <https://graphql.org/>

<sup>130</sup> <https://www.w3.org/TR/annotation-model/#introduction>

<sup>131</sup> <https://www.w3.org/TR/annotation-model/#annotations>

<sup>132</sup> <https://www.w3.org/TR/annotation-model/#annotation-page>

Il Search Endpoint<sup>133</sup> è costruito come parte dell'Annotation Server, usando le stesse tecnologie con l'aggiunta della libreria GraphQL.js per validare ed eseguire query GraphQL. Se le richieste inviate dall'utente sono espresse in un formato valido, secondo lo schema GraphQL, vengono tradotte in query Elasticsearch. In questo modo, l'API recupera le annotazioni sfruttando la stessa logica di ricerca utilizzata nelle API dell'Annotation Server e le formatta in una pagina di annotazioni web JSON-LD.

L'endpoint di ricerca può essere utilizzato interattivamente con un client GraphQL (come Altair GraphQL Client<sup>134</sup>), che permette all'utente di accedere e navigare facilmente la documentazione relativa alla realizzazione di una richiesta (query GraphQL) e sui formati di risposta (JSON-LD Web Annotation Data). Oppure può essere usato in modo non interattivo, inviando richieste compatibili con GraphQL su HTTP(S), definite richieste POST.<sup>135</sup>

Gli utenti possono opzionalmente specificare i seguenti filtri nella ricerca:

- **annotationId**: per cercare solo le annotazioni con un id specifico;
- **target\_source**: per cercare solo le annotazioni fatte sull'IRI specificato;
- **motivation**: per cercare solo le annotazioni con una specifica motivazione, come indicato nella specifica W3C Web Annotation Data Model. I valori validi per Pundit sono: commenting, highlighting, linking e tagging;
- **collectionIds**: per cercare solo annotazioni contenute in una o più collezioni di annotazioni, in Pundit le collezioni corrispondono ai notebook;
- **dateFrom, dateTo**: per cercare solo le annotazioni che soddisfano la condizione di data specificata;
- **size**: per limitare il numero di annotazioni restituite.

---

<sup>133</sup> [https://it.wikipedia.org/wiki/Communication\\_endpoint](https://it.wikipedia.org/wiki/Communication_endpoint)

<sup>134</sup> <https://altair.sirmuel.design/>

<sup>135</sup> [https://www.w3schools.com/tags/ref\\_httpmethods.asp](https://www.w3schools.com/tags/ref_httpmethods.asp)

Di seguito, nella Figura 3.10, un esempio di una query GraphQL valida e della corrispondente risposta restituita dall'endpoint W3C Search.

### Request:

```
POST https://api.thepund.it/w3c/annotation?page=1
```

```
{
  search {
    id
    type
    next
    items {
      id
      type
      created
      creator
      motivation
      generator
      body {
        type
        value
      }
      target {
        source
        selector {
          type
          ... on TextQuoteSelector {
            suffix
            prefix
            exact
          }
          ... on TextPositionSelector {
            start
            end
          }
          ... on RangeSelector {
            startSelector {
              type
              value
            }
            endSelector {
              type
              value
            }
          }
        }
      }
    }
  }
  ...
}
```

### Response:

```
{
  "@context": "http://www.w3.org/ns/anno.jsonld",
  "id": "urn:urn-5:OGU2ZGM5MTY2Nz1mMjM4ZDQ5OGV1NDA2ZTBkOWZkZTc=:1",
  "type": "AnnotationPage",
  "next": "urn:urn-5:OGU2ZGM5MTY2Nz1mMjM4ZDQ5OGV1NDA2ZTBkOWZkZTc=:2",
}
```



```

    "type": "TextPositionSelector",
    "start": 19771,
    "end": 19885
  },
  {
    "type": "TextQuoteSelector",
    "suffix": "10) Of course, a cosmopolitan, c",
    "prefix": " and I don't think it should -- ",
    "exact": "wouldn't the nation and the world benefit more? I think so. "
  }
]
...
}

```

**Figura 3.10:** Request e response di una query GraphQL

### 3.4.2 Data layer

Il **Pundit data layer** usa un approccio di persistenza poliglotta<sup>136</sup> in cui sono utilizzate diverse tecnologie di archiviazione dei dati. In particolare, i dati delle annotazioni sono memorizzati su ElasticSearch, mentre i dati degli utenti sono conservati in un database relazionale, MariaDB.

Per entrambi vengono utilizzati i servizi AWS, Amazon OpenSearch Service<sup>137</sup> e Amazon Relational Database Service<sup>138</sup>. I dati in ElasticSearch sono i seguenti :

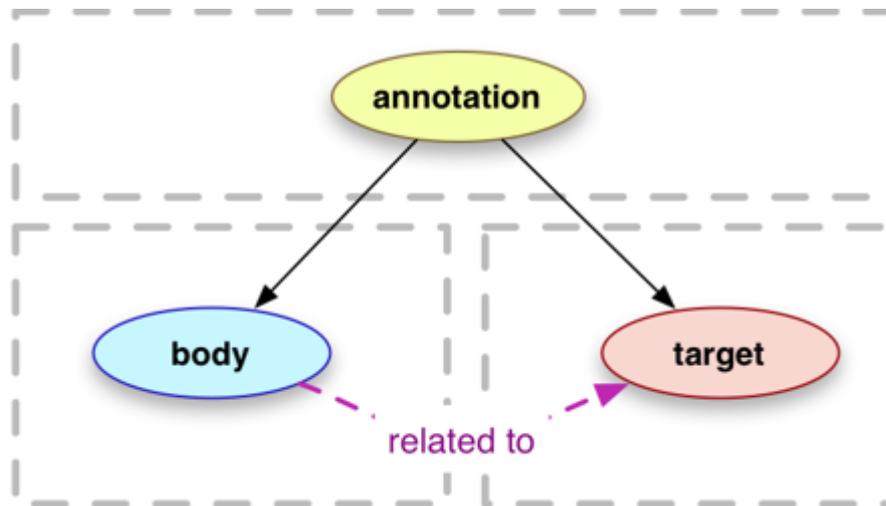
- **User:** dati dell'utente necessari da mostrare in un'annotazione, includendo il nome completo e l'URL della miniatura che rappresenta l'avatar dell'utente;
- **Notebook:** ogni utente ha un notebook e tra gli sviluppi futuri è pianificata la possibilità per gli utenti di condividere i taccuini con altri sia in sola lettura sia in scrittura;
- **Annotation:** contiene tutte le annotazioni, ognuna delle quali ha un creatore (userID) ed è contenuta in un solo quaderno (notebookID). La struttura del documento, mostrata in Figura 3.11, è stata implementata rispettando le direttive del W3C Web Annotation Data Model per

<sup>136</sup> Fowler M., *Polyglot Persistence*, 2011, <https://martinfowler.com/bliki/PolyglotPersistence.html>

<sup>137</sup> <https://aws.amazon.com/it/opensearch-service/the-elk-stack/what-is-opensearch/>

<sup>138</sup> <https://docs.aws.amazon.com/rds/index.html>

cui un'annotazione deve essere divisa in tre parti: l'annotation consistente in un insieme di metadati, il body che contiene il vero contenuto dell'annotazione, il target che identifica l'oggetto che è stato annotato, il quale può essere il frammento di un testo oppure l'intero documento.



**Figura 3.11:** Struttura di un'annotazione secondo il W3C

- **Tag:** usato per creare free-text categories scelte dall'utente, per semplificare il processo sono stati inseriti dei suggerimenti di completamento automatico dei vari tag usati in precedenza;
- **Social features:** interazioni sociali realizzate su una specifica annotazione come like, dislike.

Invece, le entità presenti nel database relazionale open source MariaDB sono:

- **Profile information:** descrive l'utente in termini di "first name, last name, email, short bio, location, website, linkedin, twitter". Il valore "public profile" è un booleano, in caso fosse TRUE è per indicare che si tratta di un profilo pubblico in cui tutti i dati sono consultabili e visualizzabili, FALSE nel caso opposto;

- **Roles and permissions:** per il momento sono permessi due ruoli, basic e admin;
- **Data used for authentication:** nel caso in cui l'utente fosse registrato su Pundit, ci si riferisce alle sue credenziali salvate, altrimenti viene utilizzata un'identità esterna all'estensione, come Google e Facebook;
- **User preferences:** le preferenze dell'utente possono cambiare frequentemente, per esempio l'ID del notebook per annotare.

### 3.4.3 Pundit App

La componente **Pundit App** ha due scopi principali: implementare la dashboard utilizzata dagli utenti per accedere alle annotazioni e gestire l'autenticazione degli utenti mediante un token JWT.<sup>139</sup>

Il suo codice è abbastanza complesso e diviso in diverse componenti che, non saranno approfondite in queste sede, ciò che conta è il suo utilizzo: quando l'utente accede alla dashboard può vedere i suoi notebook, cliccando su uno di essi riesce ad accedere alle annotazioni precedentemente realizzate. L'app Pundit ha anche un servizio di ricerca che, grazie ad un sistema di filtri riesce ad interrogare e recuperare le annotazioni dell'utente invocando le API dell'Annotation Server.

## 3.5 Stato dell'arte degli attuali tool di annotazione

In questa sezione mi occuperò di analizzare alcuni strumenti di annotazione web confrontandoli tra loro in base a determinati parametri, in seguito discussi.

Prima di entrare nel dettaglio dei singoli tool, è necessario sottolineare come essi siano stati oggetto di discussione del white paper<sup>140</sup> “*OPERAS Tools Research and Development White Paper*” redatto

---

<sup>139</sup> [https://en.wikipedia.org/wiki/JSON\\_Web\\_Token](https://en.wikipedia.org/wiki/JSON_Web_Token)

<sup>140</sup> [https://it.wikipedia.org/wiki/Libro\\_bianco#:~:text=Con%20il%20termine%20libro%20bianco,argomento%20o%20settore%20di%20attivit%C3%A0.](https://it.wikipedia.org/wiki/Libro_bianco#:~:text=Con%20il%20termine%20libro%20bianco,argomento%20o%20settore%20di%20attivit%C3%A0.)

a luglio del 2021.<sup>141</sup> OPERAS, acronimo di “*Open scholarly communication in the european research area for social sciences and humanities*”, è un'infrastruttura di ricerca europea, i cui obiettivi sono quelli di rendere l'Open Science<sup>142</sup> una realtà per la ricerca delle scienze umane e sociali (SSH) e di realizzare un sistema di comunicazione accademica in cui la conoscenza delle SSH vada a beneficio di ricercatori, accademici, studenti e più in generale dell'intera società europea e mondiale, senza barriere.

Il panorama europeo della comunicazione accademica delle SSH è attualmente disomogeneo, frammentato e non abbastanza organizzato per essere efficiente, soprattutto per affrontare la sfida della transizione all'Open Science. Ciò è dovuto a diversi fattori, tra cui le risorse ridotte, il loro sottofinanziamento e la mancanza di sostenibilità in questo settore. La natura delle discipline SSH aggiunge anche sfide specifiche come la diversità delle lingue di pubblicazione, il radicamento in diversi contesti culturali e la necessità di forme specifiche di comunicazione accademica (monografie, edizioni critiche e bibliografie).

OPERAS fornisce alla comunità di ricerca il mattone mancante di cui ha bisogno per creare e diffondere in modo semplice ed efficiente le SSH in tutta Europa, sbloccando le risorse di comunicazione accademica e permettendo all'intero settore di reinventarsi nel nuovo paradigma dell'Open Science.<sup>143</sup>

Il white paper sopracitato è frutto del lavoro dell'OPERAS Special Interest Group (SIG) ed è una versione aggiornata di un precedente white paper del 2018.<sup>144</sup>

Gli obiettivi del SIG sono: fornire un'analisi del campo di interesse, identificare le tendenze emergenti ed elencare le aree per potenziali miglioramenti e sviluppi. Dal 2018, vari studi e iniziative hanno

---

<sup>141</sup> <https://www.operas-eu.org/special-interest-group-living-book/operas-tools-research-and-development-white-paper-june-2021/>

<sup>142</sup> <https://library.isti.cnr.it/index.php/it/supporto-alla-ricerca/open-science>

<sup>143</sup> <https://www.operas-eu.org/about/operas-in-a-nutshell/>

<sup>144</sup> <https://operas.hypotheses.org/aboutoperas/working-groups/tools-wg>

confermato la necessità sia di coordinare gli sviluppi degli strumenti, sia di fornire una guida agli utenti.<sup>145</sup>

Di seguito, mi occuperò di analizzare e confrontare diversi tool:

- Hypothes.is;
- Pundit;
- B2NOTE;
- Weava;
- GetLiner;
- Notesalong;
- Remarq Lite.

### 3.5.1 Hypothes.is

L'Hypothes.is Project<sup>146</sup> è un'organizzazione non-profit, nata dall'idea di poter creare una conversazione sul tutto il web mediante il loro tool di annotazione.

Hypothes.is si basa su determinati principi guida: il software deve essere Open source,<sup>147</sup> deve poter funzionare ovunque, il suo fine deve essere il bene pubblico consistente nell' ampliare la comprensione umana; neutrale quindi senza favorire posizioni ideologiche o politiche, avere un approccio bottom up, dal basso verso l'altro, in quanto l'elemento essenziale è rappresentato dalla comunità di utenti; deve essere internazionale, trasparente, verificabile; utilizzabile in diversi contesti e formati.<sup>148</sup>

---

<sup>145</sup> <https://www.operas-eu.org/special-interest-groups/>

<sup>146</sup> <https://web.hypothes.is/>

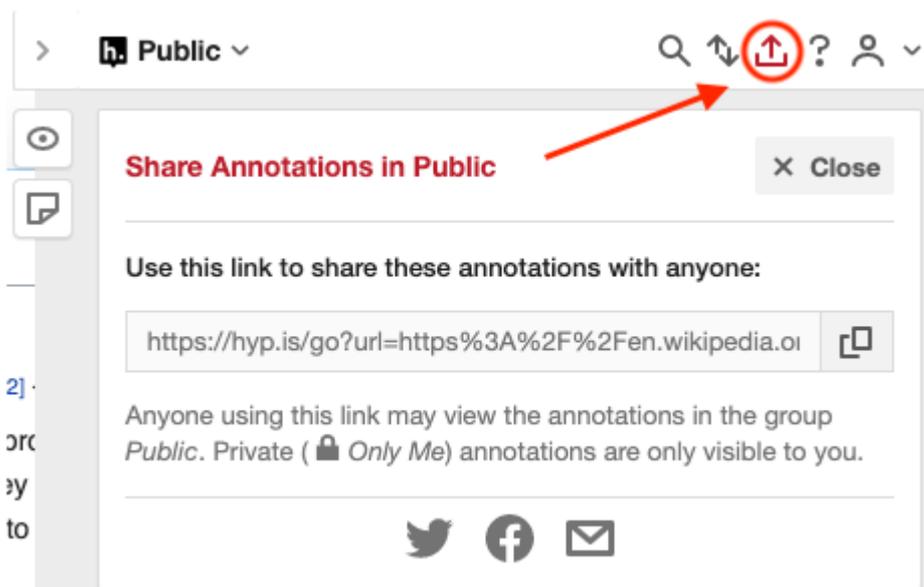
<sup>147</sup> <http://www.gnu.org/philosophy/free-sw.html>

<sup>148</sup> <https://web.hypothes.is/principles/>

Il tool può essere utilizzato in due modalità differenti: innanzitutto bisogna creare un account<sup>149</sup> per poter gestire e memorizzare le proprie annotazioni, successivamente è possibile utilizzare Hypothes.is sia come estensione Chrome sia aggiungendolo come bookmarklet alla propria barra dei segnalibri, in modo da poterlo utilizzare su diversi browser.<sup>150</sup>

Un'altra modalità di utilizzo del tool è senza l'ausilio di un'estensione browser, grazie al servizio proxy<sup>151</sup> di Hypothesis, che permette di usare il tool in circostanze o ambienti in cui altrimenti sarebbe impossibile, ad esempio quando si vuole annotare una pagina o un documento in un browser che non usa o non può usare l'estensione o il bookmarklet, ad esempio da computer pubblici, computer per i quali non si hanno privilegi di amministratore ecc.

Per utilizzare il tool senza l'estensione browser, è necessario copiare il link del documento o annotazione di nostro interesse all'interno dell'app del tool. Nella Figura 3.12 è riportata l'icona per poter compiere quest'azione.



**Figura 3.12:** Icona per condividere annotazioni

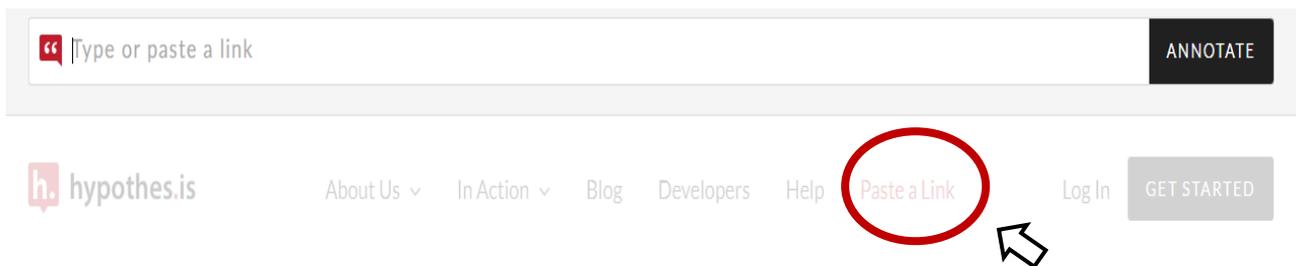
<sup>149</sup> <https://web.hypothes.is/start/>

<sup>150</sup> <https://www.teachology.ca/knowledgebase/what-is-hypothes-is-and-how-can-it-be-used-to-annotate-the-web-for-learning/>

<sup>151</sup> <https://it.wikipedia.org/wiki/Proxy>

Quando il link viene aperto in un browser, Hypothes.is verifica se l'estensione browser o il bookmarklet siano installati, in caso contrario, il tool funzionerà mediante il proxy service.<sup>152</sup>

Un altro modo per annotare è mediante l'opzione "Paste a Link" presente nella barra del menu della pagina iniziale,<sup>153</sup> riportata nella Figura 3.13, oppure si può visitare direttamente il sito dedicato per testare il server proxy.<sup>154</sup>



**Figura 3.13:** Esempio di utilizzo di Hypothes.is dalla home page

Il tool ha diverse funzioni base come il poter selezionare una parte di testo per evidenziarla e annotarla in maniera pubblica o privata è possibile replicare o condividere qualsiasi annotazione; creare annotazioni di gruppo e infine cercare le proprie note o quelle di profili pubblici.<sup>155</sup>

Altre caratteristiche avanzate riguardano il creare gruppi privati, pensati per classi di studenti, gruppi di lavoro; aggiungere tag liberi, il che significa che qualsiasi cosa che viene aggiunta come tag lo diventa.

<sup>152</sup> <https://web.hypothes.is/help/what-is-the-via-proxy/>

<sup>153</sup> <https://web.hypothes.is/>

<sup>154</sup> <https://via.hypothes.is>

<sup>155</sup> <https://web.hypothes.is/>

Infine, è possibile filtrare le annotazioni pubbliche in base ad una determinata disciplina o soggetto di interesse.<sup>156</sup>

Hypothes.is è impiegato per fornire supporto alle annotazioni in diversi settori:

1) Educativo, permettendo a studenti ed educatori di analizzare testi digitali in modo collaborativo, siano essi pagine web standard, PDF e documenti caricati su servizi cloud come Google Drive.

Inoltre, è offerta una guida su come utilizzare il tool sia per gli studenti sia per gli insegnanti.<sup>157</sup>

2) In ambito giornalistico, ad esempio Climate Feedback<sup>158</sup> utilizza Hypothes.is in modo particolarmente efficace. Esso è un gruppo di quasi 200 scienziati del clima che, utilizzano l'annotazione web per contrassegnare articoli sul cambiamento climatico, aggiungendo informazioni contestuali ed evidenziando le imprecisioni fattuali e i ragionamenti errati. Climate Feedback fornisce un forum critico per i veri esperti all'interno della comunità scientifica per verificare le affermazioni sulla scienza e discutere le inevitabili controversie che sorgono su questo argomento politicamente rilevante.<sup>159</sup>

3) Un altro ambito in cui il tool è utilizzato è quello della ricerca,<sup>160</sup> progetti che utilizzano il tool appartengono al campo delle scienze umane, “JHUP and Hypothesis Partner to Enable Annotation on Modernism/modernity”;<sup>161</sup> delle scienze sociali, “Increasing Openness in Qualitative Research With QDR and Cambridge University Press”<sup>162</sup> e STEM, acronimo per “Science, Technology, Engineering and Mathematics”, termine utilizzato per indicare le discipline scientifico-tecnologiche, con il progetto “*Who says neuroscientists don't need more brains? Annotation with SciBot*”.<sup>163</sup>

---

<sup>156</sup><https://www.teachology.ca/knowledgebase/what-is-hypothes-is-and-how-can-it-be-used-to-annotate-the-web-for-learning/>

<sup>157</sup> <https://web.hypothes.is/education/>

<sup>158</sup> <https://climatefeedback.org/>

<sup>159</sup> <https://web.hypothes.is/journalism/>

<sup>160</sup> <https://web.hypothes.is/research/>

<sup>161</sup> <https://web.hypothes.is/blog/jhup-and-hypothesis-partner-to-enable-annotation-on-modernism-modernity/>

<sup>162</sup> <https://web.hypothes.is/blog/ati/>

<sup>163</sup> <https://web.hypothes.is/blog/annotation-with-sciobot/>

### 3.5.2 B2NOTE

Il primo tool ad essere esaminato è B2NOTE, servizio di annotazione di EUDAT CDI che permette di creare, cercare e gestire facilmente le annotazioni. B2NOTE supporta tre tipologie di annotazione:

- **Semantic tag:** l'uso di una parola chiave inclusa nell'ontologia di riferimento, che al momento quella di riferimento è Bioportal, un vocabolario biomedico;<sup>164</sup>
- **Free-text keyboard:** utilizzata quando non è possibile trovare un termine semantico specifico nelle repository;
- **Comment:** un'annotazione più ampia e completa.

Le annotazioni sono create e memorizzate in un formato leggibile dal computer utilizzando il modello W3C Web Annotation, inoltre è permesso agli utenti di esportare le annotazioni in diversi formati, ad esempio in JSON-LD.

L'utente deve registrarsi per creare e memorizzare le annotazioni, un'ulteriore registrazione può essere fatta ad un servizio terze parti, B2SHARE,<sup>165</sup> che ha abilitato la funzionalità di annotazione con B2NOTE. La versione B2NOTE integrata con B2SHARE è un servizio di annotazione crowdsourcing, il che significa che tutte le annotazioni sono disponibili pubblicamente. Per preservare l'anonimato, l'utente può registrarsi e creare uno pseudonimo.<sup>166</sup>

In Figura 3.14, è riportato un esempio di come sia possibile creare delle annotazioni con B2NOTE su un file presente all'interno del servizio, cliccando sul pulsante "Annotate in B2NOTE", comparirà il widget e si potrà iniziare ad annotare.<sup>167</sup>

---

<sup>164</sup> <https://bioportal.bioontology.org/>

<sup>165</sup> <https://eudat.eu/services/userdoc/b2share-usage>

<sup>166</sup> <https://marketplace.eosc-portal.eu/services/b2note>

<sup>167</sup> <https://b2note.eudat.eu/>

## RDA Foundation Governance Document

by [Research Data Alliance Council](#); [RDAZ](#);

Dec 21, 2016

**Abstract:** A document describing the high-level structures of the Research Data Alliance Foundation. This document is separate from the procedures and processes.

**Keywords:** [Research Data Alliance](#); [RDA](#); [Governance](#); [Foundation](#); [RDA Policy](#);

**PID:** [11304/619eda56-100f-43fo-9e72-g8a22792eb25](#)

[Copy](#)

The screenshot shows a file management interface. On the left, there is a 'Files' section with a table containing one file entry: 'rdafile' with a size of '9B'. The file name 'rdafile' is circled in red. Below the file name, there is a 'Checksum' field with the value 'md5:c8afdb36c52cf4727836669819e69222' and a 'PID' field with the value '11304/619eda56-100f-43fo-9e72-g8a22792eb25'. A 'Copy' button is next to the PID. Below the PID, there is an 'Annotate in BzNote' button, which is also circled in red. On the right, there is a 'Basic meta' sidebar with options: 'Open Access', 'License', 'Contact Email', and 'Resource Type'.

**Figura 3.14:** Annotazione con B2NOTE

### 3.5.3 Weava

Weava<sup>168</sup> è un tool online di ricerca accademica, che permette a studenti, professori e ricercatori di svolgere la propria attività in maniera più efficiente. Può essere utilizzato installando l'estensione Chrome<sup>169</sup> ma vi è anche la possibilità di utilizzarlo sui dispositivi mobile con sistema operativo iOS.<sup>170</sup> Ci sono due abbonamenti possibili, uno gratuito e uno premium con funzionalità aggiuntive, come per esempio la possibilità di esportare immagini o avere una memoria di archiviazione illimitata.<sup>171</sup>

<sup>168</sup> <https://www.weavatools.com/>

<sup>169</sup> <https://chrome.google.com/webstore/detail/weava-highlighter-pdf-web/cbnaodkpfinfiiipjblikofhlhlcickei>

<sup>170</sup> <https://apps.apple.com/us/app/weava-app/id1451423065>

<sup>171</sup> <https://www.weavatools.com/premium/>

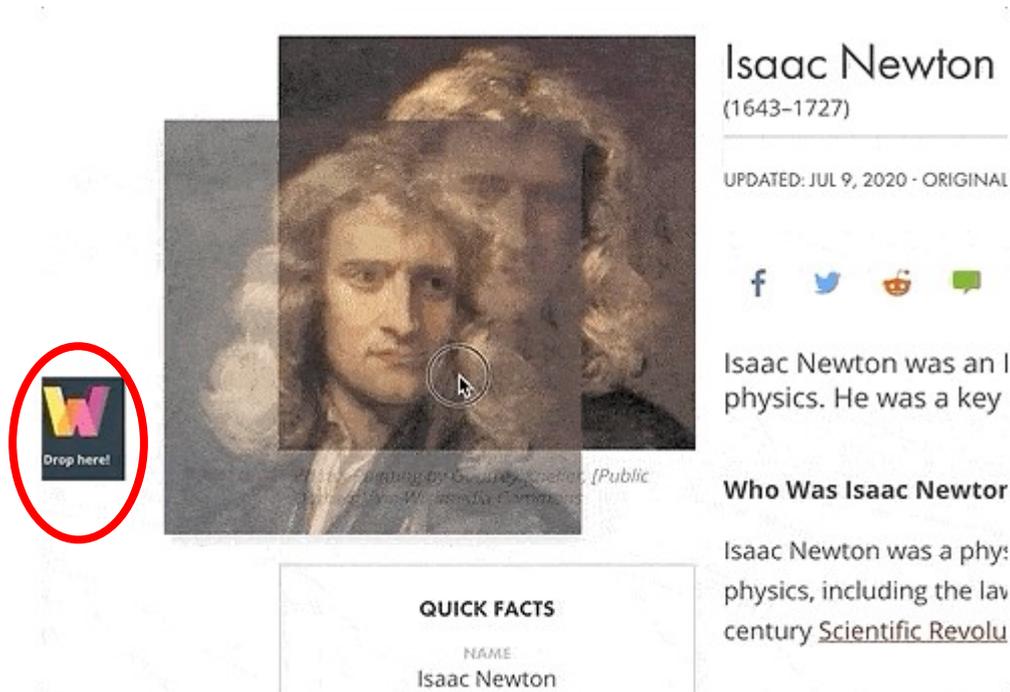
L'attività centrale del tool, a cui tutto il resto è collegato, è l'evidenziazione. . Essa consente di evidenziare direttamente sulle pagine web e PDF online ma, questi ultimi possono essere anche locali sul proprio laptop o caricati sul tool.

Gli scopi possono essere differenti, si possono evidenziare parole difficili da voler cercare in seguito; citazioni e materiali per la ricerca o la scrittura di un saggio. Un modo utile per rendere il processo di ricerca più facile è quello di utilizzare colori differenti.

Con la versione gratuita di Weava, è possibile utilizzare cinque colori predefiniti, con la versione Premium non ci sono limiti di colore.

Tutte le evidenziazioni sono salvate su Weava e si sincronizzano con il cloud in modo da essere accessibili ovunque. Sono visibili in cartelle a cui si può accedere sia tramite la barra laterale, sidebar, sia tramite la dashboard Weava.

Una caratteristica innovativa, offerta agli utenti premium, è la capacità di salvare le immagini su qualsiasi sito web trascinandole in una piccola casella su cui compare la scritta "Drop here!" , ciò è mostrato nella Figura 3.15, riportata di seguito.



**Figura 3.15:** Esempio di salvataggio di un'immagine

Alcune limitazioni che Weava presenta per l'azione dell'evidenziazione riguardano ad esempio il non essere in grado di supportarle su Google Drive o sui file di Google Classroom<sup>172</sup> oppure il non funzionare in alcuni siti web che hanno un codice conflittuale con l'evidenziazione. Sfortunatamente, poiché questi problemi sono legati all'implementazione delle funzionalità rispettivamente di Google o di alcuni siti web, non c'è una soluzione.<sup>173</sup>

Oltre ad evidenziare le informazioni presenti sulle fonti digitali, un'altra azione importante è quella di prendere appunti per raccogliere i propri pensieri e idee su un dato argomento. È possibile annotare le proprie evidenziazioni, cliccando e scrivendo nella casella di testo "Prendi nota".

La sottolineatura grigia, visibile in Figura 3.16, serve per indicare che in quel punto è presente un'annotazione; ciò rende il suo reperimento più diretto e intuitivo, senza la necessità di cliccare su un'evidenziazione per controllarla oppure navigare la barra laterale o la dashboard.<sup>174</sup>

<sup>172</sup> [https://it.wikipedia.org/wiki/Google\\_Classroom](https://it.wikipedia.org/wiki/Google_Classroom)

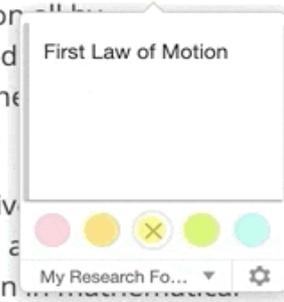
<sup>173</sup> <https://www.weavatools.com/highlight-with-weava/>

<sup>174</sup> <https://www.weavatools.com/annotate-with-weava/>

The **First Law of Motion** states, "A body at rest will remain at rest, and a body in motion will remain in motion unless it is acted upon by an external force." This simply means that things cannot start, stop, or change direction themselves. It takes some force acting on them from the outside to make a change. This property of massive bodies to resist changes in their motion is sometimes called *inertia*.

The **Second Law of Motion** describes what happens to a massive body when it is acted upon by an external force. It states, "The force acting on a body is equal to the mass of that object times its acceleration." This is written in mathematical form as  $F = ma$ , where  $F$  is force,  $m$  is mass, and  $a$  is acceleration. The bold letters indicate that force and acceleration are *vector* quantities, which means they have both magnitude and direction. The force can be a single force, or it can be the vector sum of more than one force, which is the net force after all the forces are combined.

When a constant force acts on a massive body, it causes it to accelerate, i.e., to change its velocity, at a constant rate. In the simplest case, a force applied to an



**Figura 3.16:** Esempio di annotazione

I modi in cui le note possono essere organizzate sono molti e diversi ed è possibile farlo mediante la barra laterale e la dashboard Weava. Quest'ultima possiede funzioni aggiuntive rispetto alla prima, come la creazione di citazioni, l'esportazione del proprio lavoro, e altro ancora.

La Figura 3.17 esplica le differenze tra le due modalità organizzative. Le funzionalità che la dashboard supporta rispetto al sidebar sono: selezionare tutte le evidenziazioni, la possibilità di creare citazioni<sup>175</sup>, esportare evidenziazioni e note in diversi formati come Word, Excel, .csv, .txt.; modificare l'URL d'origine del sito web/PDF; possibilità di creare documenti; raggruppare le evidenziazioni tramite colore, accettare inviti di collaborazioni di cartelle e poter ricevere notifiche.<sup>176</sup>

<sup>175</sup> <https://www.weavatools.com/how-to-create-citations-with-weava/>

<sup>176</sup> <https://www.weavatools.com/organize-with-weava/>

	The Weava Sidebar	The Weava Dashboard
Add, edit and delete notes	✓	✓
Copy highlight/note	✓ (individual)	✓ (also multiple)
Search	✓	✓
Highlight color filter	✓	✓ (+ also image filter)
Upload PDF	✓	✓
Show/hide highlights filter	✓	✓
Invite collaborators	✓	✓
Add, edit and delete folders	✓	✓
Create sub-folders	✓	✓
Move highlights between folders	✓	✓
Add, edit and delete color labels	✓	✓
Order by "old to new/new to old"	✓	✓
Drag folders into other folders	✓	✓
Open website/PDF	✓ (directly to source)	✓ (inside dashboard iframe)
"Select all" highlights	✗	✓
Create citations	✗	✓
Export highlights/notes	✗	✓
Edit source URL of website/PDF	✗	✓
Create Weava documents	✗	✓
Group highlights by color	✗	✓
Accept folder invitations	✗	✓
Weava notifications	✗	✓

**Figura 3.17:** Differenze tra Weava sidebar e dashboard

Infine, Weava è anche un tool collaborativo grazie alla possibilità di creare cartelle condivise. Si può collaborare per fini educativi, per esempio gli insegnanti possono condividere con i loro studenti una cartella con le letture del semestre, oppure possono precompilare i testi con evidenziazioni e note per facilitarne la lettura e la comprensione.

Si può collaborare anche per fini lavorativi, si possono creare infatti cartelle condivise contenenti materiali, link, evidenziazioni e annotazioni utili al gruppo di lavoro. I supervisor possono poi leggere, modificare e fornire un feedback sui rispettivi siti web.<sup>177</sup>

### 3.5.4 GetLiner: Search less, learn more

Liner è un servizio di annotazione disponibile sia come estensione browser, in particolare per Chrome,<sup>178</sup> Edge,<sup>179</sup> Safari,<sup>180</sup> Opera<sup>181</sup> e Whale,<sup>182</sup> sia come mobile App per Android<sup>183</sup> o iOS.<sup>184</sup>

Esistono tre piani per utilizzarlo: Basic; Premium e Pro, gli ultimi due offrono ulteriori servizi oltre quelli di base e soprattutto senza limitazioni( il Basic ha limitazioni per quanto concerne i commenti, massimo tre per pagina web, un solo colore per l'evidenziazione e massimo 15 per pagina).<sup>185</sup>

Liner è uno strumento che consente di evidenziare e commentare frammenti di testo su pagine Web, su PDF e su Youtube. L'utente ha la possibilità di sincronizzare le sue annotazioni su un numero illimitato di dispositivi, di non perdere le annotazioni riaggiornando la pagina e di aggiungere tag sulle evidenziazioni per categorizzarle.<sup>186</sup>

Liner provvede a fornire anche un Search assistant che permette di migliorare la ricerca su Google rendendo l'esperienza utente semplificata in quanto si trova più velocemente ciò che si cerca. Ciò avviene grazie a tre funzionalità messe a disposizione: (i) Liner crede nelle scelte della sua comunità pertanto è possibile leggere soltanto le pagine consultate da altri utenti di tale servizio, eseguendo dunque una ricerca più mirata; (ii) è possibile avere un'anteprima dei punti salienti del contenuto di

---

<sup>177</sup> <https://www.weavatools.com/collaborate-in-weava/>

<sup>178</sup> [chrome.google.com/webstore/detail/liner-searchfasterhighl/bmhcbmnbenmcecpmpepghooflbehcack](https://chrome.google.com/webstore/detail/liner-searchfasterhighl/bmhcbmnbenmcecpmpepghooflbehcack)

<sup>179</sup> [microsoftedge.microsoft.com/addons/detail/linersearchfasterh/kmhegedbanhfbnoboomeafpmojfdlp](https://microsoftedge.microsoft.com/addons/detail/linersearchfasterh/kmhegedbanhfbnoboomeafpmojfdlp)

<sup>180</sup> [apps.apple.com/kr/app/liner-web-pdf-highlighter/id1463530746?mt=12](https://apps.apple.com/kr/app/liner-web-pdf-highlighter/id1463530746?mt=12)

<sup>181</sup> [support.getliner.com/hc/en-us/articles/115000862772-How-to-install-LINER-for-Opera-](https://support.getliner.com/hc/en-us/articles/115000862772-How-to-install-LINER-for-Opera-)

<sup>182</sup> [store.whale.naver.com/detail/ooelpmkcpjkmoffkdgefbejfgfpafhic](https://store.whale.naver.com/detail/ooelpmkcpjkmoffkdgefbejfgfpafhic)

<sup>183</sup> <https://play.google.com/store/apps/details?id=com.getliner.liner>

<sup>184</sup> <https://apps.apple.com/app/appname/id955395198?amp%3Bl=1>

<sup>185</sup> <https://getliner.com/upgrade>

<sup>186</sup> <https://getliner.com/highlighter>

una pagina web così da capire in pochi secondi se è ciò di cui si ha bisogno; (iii) infine è possibile visualizzare sezioni chiave della pagina evidenziate dagli utenti.

Inoltre, Liner fornisce dei suggerimenti all'utente quando si evidenzia una sezione in modo tale da mostrare contenuti di interesse che potrebbero essere sfuggiti.<sup>187</sup>

### 3.5.5 Notesalong

Notesalong è un tool di annotazione, può essere installato come estensione Chrome. Presenta diverse funzionalità: la possibilità di evidenziare e annotare pagine web e PDF in sei colori differenti; la possibilità di condividere sia tutte le annotazioni che una specifica; permette di usare tag, emoticon e colori per personalizzare ciò che si scrive. È, inoltre, possibile organizzare le proprie note in cartelle di argomento differente (scienza, filosofia) e filtrarle per una ricerca più veloce delle stesse.<sup>188</sup>

In Figura 3.18 è riportato un esempio del box che compare quando si crea un'annotazione.

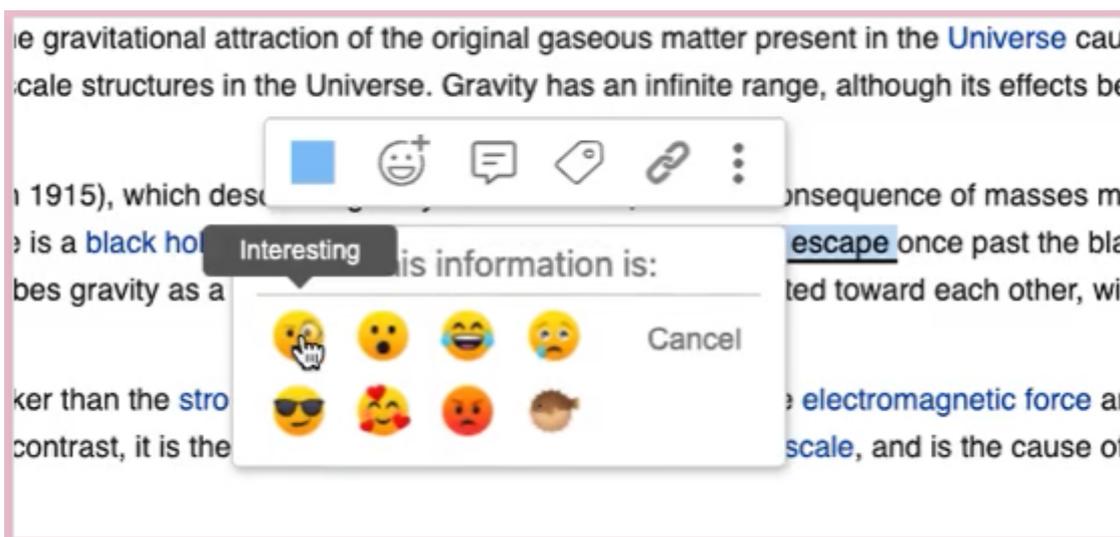


Figura 3.18: Box di annotazione

<sup>187</sup> <https://getliner.com/en>

<sup>188</sup> <https://notesalong.com/>

### 3.5.6 Remarq Lite

Remarq è una rete di collaborazione accademica sviluppata dall'azienda RedLink,<sup>189</sup> progettata per ricercatori e utenti qualificati. Di recente, RedLink ha rilasciato un'estensione Chrome, per facilitare l'annotazione, la collaborazione e la connessione attraverso il Web, in particolare per aiutare gli studenti e gli educatori ad essere più efficienti nelle loro attività.

L'estensione è Remarq Lite,<sup>190</sup> che consente agli utenti di integrare le note e le evidenziazioni da qualsiasi fonte online nel loro profilo Remarq.

Il plugin è gratuito, utilizzabile con Chrome, invece in versione segnalibro anche in Firefox, Microsoft Edge e Safari, la realizzazione di un plugin completamente integrato per questi browser è in fase di sviluppo.

Remarq Lite è stato progettato specificamente per supportare gli utenti nel mondo dell'istruzione, permettendo:

- Agli educatori di creare gruppi privati favorendo la collaborazione tra studenti;
- Agli studenti di creare gruppi privati per realizzare progetti.

Remarq Lite consente di evidenziare, prendere note private o pubbliche durante la navigazione; condividere contenuti e partecipare a sondaggi o recensioni in diversi campi di interesse. Inoltre, è possibile collaborare con i propri colleghi discutendo di articoli, passaggi specifici di un testo o di una tematica in particolare. Tutto ciò può avvenire tramite la creazione di gruppi privati in cui solo le persone che sono invitate possono partecipare; gruppi protetti ovvero gruppi visibili per cui è necessaria una richiesta per diventarne membri che deve essere accettata dall'amministratore.

---

<sup>189</sup> <https://www.redlink.com/mission-and-purpose/>

<sup>190</sup> <https://remarqable.com/web/remarq-lite.html>

Infine, ci sono i gruppi aperti in cui tutti possono prendere parte senza previa autorizzazione<sup>191</sup>.

### 3.6 Caratteristiche a confronto

Dopo aver analizzato singolarmente e nel dettaglio i tool presi in esame nella sezione “*Annex III Analytical table of annotation tools*” del white paper “*OPERAS Tools Research and Development White Paper*” redatto a luglio 2021,<sup>192</sup> in questo paragrafo saranno messi a confronto per avere una panoramica migliore delle loro differenze e somiglianze.

	Hypothes.is	Pundit	B2NOTE	Weava	GetLiner	Notesalong	Remarq Lite
web site	<a href="https://web.hypothes.is/">https://web.hypothes.is/</a>	<a href="https://the.pund.it">https://the.pund.it</a>	<a href="https://b2note.eudat.eu/">https://b2note.eudat.eu/</a>	<a href="https://www.weavatools.com/">https://www.weavatools.com/</a>	<a href="https://getliner.com/">https://getliner.com/</a>	<a href="https://notesalong.com/">https://notesalong.com/</a>	<a href="https://remarkable.com/web/index.html">https://remarkable.com/web/index.html</a>
type	web annotation tool	web annotation tool	Not a generic web annotation tool, since it is integrated with the B2Share service. It has been included here because theoretically the service can also be used via a widget on other web sites. It is also a EOSC-HUB service and it is listed in	web annotation tool	web annotation tool	web annotation tool	Web annotation tool plus platform for sharing and discovery users and annotated content.

<sup>191</sup> <https://remarkable.com/web/index.html#!#groups>

<sup>192</sup> <https://www.operas-eu.org/special-interest-group-living-book/operas-tools-research-and-development-white-paper-june-2021/>

			the EOSC marketplace				
<b>Annotator Functionalities</b>							
Highlighting	Yes	Yes	No	Yes	Yes	Yes	Yes
Different highlight colors	No (still planned?)	No in the current version Yes in the new one in the making	No	Yes	Yes	Yes	No
Commenting	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Tagging	Yes. Free text tags can be associated to an annotation	No in the current version Yes in the new one in the making	Yes An annotation can be a free text tag.	Yes, tags can be associated to a highlight color and organized in folders.	Yes	Yes. also support for emoticon	Yes (private for the single user)
Annotating a text fragment	Yes	Yes	No	Yes	Yes	Yes	Yes
Annotating a whole web page/ resource	Yes (they are called "Page notes")	Yes	Yes, but only if published in B2Share (resource description page plus attachments ) or if the content provider integrates B2Note in the web site	Yes	Yes	No	No
Semantic annotations (RDF triples in annotations)	No	Yes	Yes, but limited: only link to vocabularies (not known the predicate used).	No	No	No	No
Personal annotation	Yes	Yes (via	Yes	Yes	yes	yes	Yes

		Private notebook)					
Groups (Annotation Container)	Yes	Yes (Notebooks)	No	Yes	no	no	No but it's possible to publish them in public groups
Public discussion	Yes	Yes	No	No	no	no	Yes, for public groups. Remark Lite does not support public comments. This is possible only for those journals and publications that integrate the full version of Remark.
Share an annotation	Yes	No in the current version. Yes in the new one in the making	No	Yes	yes	yes	No
PDF support	Yes	No in the current version Yes in the new one in the making	No	Yes	Yes	yes	Yes
Mobile support	Limited: possible through the proxy service via.hypothes.is. Decent on a tablet, quite impossible to use on a	No	Yes when embedded in the B2Share service, since the latter supports	Yes but only through a dedicated iOS App	Yes but only through a specific App and a dedicated web browser	Limited: annotations can be viewed on a mobile device through the inweb.notes along.com	Apparently not this version (Remark Lite) but the commercial edition, that can be

	mobile phone.		responsive design.			proxy service	embedded on publishers' sites, supports responsive design.
Search	Yes	Yes. "Filter annotations" option	Yes	Yes	yes (tag and color filters)	yes yes (tag, color filters and folders)	Yes (people, groups, annotations, articles)

**Service characteristics**

Exporting the annotations	No apparently (only via coding/APIs)	Yes those of a Notebook via file (XSLX, DOCX, ODT, JSON-LD/W3C Web Annotation Data Model) or ready to use API endpoint but only of entire notebooks (no filtering)	Yes, those of a B2Share resource, via file (JSON-LD, RDF/XML, RDF/Turtle)	Yes (word, txt, csv, xls)	yes (word, onenote, evernote, txt)	no	No apparently
---------------------------	--------------------------------------	--	---	---------------------------	------------------------------------	----	---------------

Works everywhere	Yes	Yes	No, only web documents published through B2Share	Yes (browser extension)	yes (browser extension)	no (chrome extension)	Only for personal notes
------------------	-----	-----	--	-------------------------	-------------------------	-----------------------	-------------------------

**Ecosystem & maintainer**

W3C standard - data model (W3C WA)	they declared "Yes" in the previous version of the SIG Tools document. Actually the	Partial support in the current version Full support for the Web	Yes, when exporting the annotations of a resource	not specified	not specified	not specified	In the old OPERAS document it was stated as "Claimed" but no
------------------------------------	---	---	---	---------------	---------------	---------------	--

	public API doesn't seem compliant to W3C WA. See: for example <a href="https://hypothes.is/api/search">https://hypothes.is/api/search</a> and <a href="https://www.w3.org/TR/annotation-model/">https://www.w3.org/TR/annotation-model/</a>	Annotation Data Model in the new one for some public APIs					evidence can be found.
W3C standard - protocol	No. Declared in progress in the previous version of the document but no evidence about that in their web site	No.	No	not specified	not specified	not specified	In the old OPERAS document it was stated as "No" but no evidence can be found.
Open source	Yes	Yes	Yes <a href="https://github.com/EUDAT-B2NOTE/b2note">https://github.com/EUDAT-B2NOTE/b2note</a>	No	No	No	No

**Tabella 3.1:** Tool a confronto

La Tabella 3.1 confronta Hypothes.is, Pundit, B2NOTE, Weava, GetLiner, Notesalong e Remarq Lite a partire da tre macrocategorie: **Annotator Functionalities, Service characteristics ed Ecosystem & maintainer.**

Inizialmente sono riportati i rispettivi siti web e specificata la tipologia: sono tutti tool di annotazione sebbene **B2NOTE** sia anche un servizio di EOSC-HUB<sup>193</sup> e sia integrato con la piattaforma

<sup>193</sup> <https://www.eosc-hub.eu/about-us>

B2SHARE<sup>194</sup> in cui può essere utilizzato come widget. Lo stesso vale per **Remarq Lite** che appartiene ad una piattaforma di condivisione, Remarq.<sup>195</sup>

Per quanto concerne le funzionalità del tool, sono stati presi in considerazione i seguenti campi:

- Evidenziazione;
- Uso di differenti colori;
- Commentare;
- Taggare;
- Annotare frammenti di testo o intere pagine web;
- Possibilità di realizzare annotazioni semantiche;
- Annotazioni personali/pubbliche/ di gruppo;
- Possibilità di condividere l'annotazione;
- Possibilità di supporto in formato PDF e mobile;
- Filtraggio.

La possibilità di evidenziazione è una caratteristica di tutti i tool, tranne di **B2NOTE**; invece, la possibilità di utilizzare diversi colori è implementata in Weava, GetLiner e Notesalong.

Il commentare e il taggare sono permessi da tutti i tool, **Pundit** ha implementato il tagging nell'ultima versione; inoltre, **Weava** organizza i tag per colore e cartelle, **Notesalong** invece è l'unico a supportare anche le emoticon nella funzionalità del tagging.

Per l'annotazione di frammenti, soltanto il tool **B2NOTE** non lo permette, invece a non supportare l'annotazione di intere pagine web sono **Notesalong** e **Remarq Lite**.

---

<sup>194</sup> <https://sp.eudat.eu/catalog/resources/709c053d-3b3f-4c54-8ef4-6efeea387816>

<sup>195</sup> <https://remarqable.com/web/index.html>

Per l'annotazione semantica, soltanto **Pundit** è in grado di dare agli utenti questa possibilità; **B2NOTE** lo permette ma, in maniera limitata.

Le annotazioni personali sono garantite da tutti i tool, quelle di gruppo invece soltanto da **Hypothes.is**, **Pundit**, **Weava**. **Remarq Lite** invece permette soltanto di pubblicare l'annotazione di un utente in un gruppo pubblico.

Le discussioni pubbliche sono garantite da **Hypothes.is** e **Pundit**; in **Remarq Lite** solo se i gruppi sono pubblici; invece, i commenti completamente pubblici sono concessi per i siti di pubblicazioni e riviste giornalistiche che integrano l'intera versione di **Remarq**.

Infine, in ambito annotativo, la condivisione dell'annotazione è concessa da tutti tranne che da **B2NOTE** e **Remarq Lite**, attualmente **Pundit** la sta rilasciando.

La possibilità di supportare la versione PDF delle annotazioni è garantita da tutti i tool tranne che da **Pundit**, in fase di rilascio, e **B2NOTE**.

La versione mobile è attiva in maniera limitata in **Hypothes.is**, relegata principalmente ai tablet; in **B2NOTE** quando il servizio è integrato con B2SHARE, in **Weava** solo per i dispositivi con sistema operativo iOS; in **GetLiner** solo tramite app e web browser apposito e in uso limitato anche in **Noteaslong**. **Remarq Lite** lo renderà possibile nella versione commerciale utilizzabile su siti di pubblicazione. L'unico a non averla è **Pundit** ma è in progetto la sua implementazione.

L'ultima caratteristica presa in esame in questa sezione è la ricerca delle annotazioni, possibile con tutti i tool ma, in modo diverso. Ad esempio, **GetLiner** la realizza attraverso i filtri del colore e dei tag, invece **Remarq Lite** lo fa attraverso le persone, gruppi, annotazioni e articoli.

L'altra macrocategoria riguarda le caratteristiche del servizio, in particolare l'esportazione delle annotazioni e il poter lavorare ovunque nel senso che il tool non deve essere necessariamente implementato sul sito che si sta visitando, concetto ben espresso sulla pagina web di **Hypothes.is**.<sup>196</sup>

---

<sup>196</sup> <https://web.hypothes.is/>

La possibilità di esportare le annotazioni è realizzata da **Pundit**, in cui le annotazioni possono assumere diverse estensioni come docx, JSON-LD; da **B2NOTE** che consente l'esportazione in JSON-LD, RDF/XML, RDF/TURTLE; in **Weava** che esporta in word, txt, csv e xls e in **GetLiner** che implementa l'esportazione in word, onenote, txt.

**Notesalong** e **Remarq Lite** non lo permettono.

Infine, la possibilità di usare il tool ovunque è garantita da **Hypothes.is**, **Pundit**, **Weava**, **GetLiner** e **Remarq Lite** ma solo per le note personali.

L'ultima macrocategoria riguarda l'ecosistema in cui sono inseriti, in particolare si analizza se sono fedeli agli standard del W3C data model,<sup>197</sup> W3C protocol<sup>198</sup> e se sono Open source.

Per quanto riguarda l'essere fedeli al W3C data model, **Hypothes.is** non lo è, **Pundit** lo è nella nuova versione, **B2NOTE** invece solo quando le annotazioni sono esportate, per gli altri tool non è specificato.

Il W3C protocol invece non è rispettato pienamente da nessun tool. Infine, solo **Hypothes.is**, **Pundit** e **B2NOTE** sono tool di tipo Open source.<sup>199</sup>

---

<sup>197</sup> <https://www.w3.org/TR/annotation-model/>

<sup>198</sup> <https://www.w3.org/TR/annotation-protocol/>

<sup>199</sup> <https://www.operas-eu.org/special-interest-group-living-book/operas-tools-research-and-development-white-paper-june-2021/#Tools-Annexes>

## CAPITOLO 4

### 4.1 Inquadramento del contesto di tesi

A partire dall'attività svolta nell'ambito del mio tirocinio curriculare della durata di tre mesi, presso l'azienda pisana Net7<sup>200</sup>, ho iniziato a studiare le tecnologie del Semantic Web , focalizzandomi in particolare sullo sviluppo dello strumento di annotazione, Pundit.<sup>201</sup>

Net7, infatti, è stata impegnata in un aggiornamento del suo tool nell'ambito del progetto europeo di ricerca TRIPLE.<sup>202</sup> In particolare, ha sviluppato un endpoint ad accesso pubblico compatibile con GraphQL,<sup>203</sup> un query language per le API, che rispetta il modello di dati della specifica W3C Web Annotation. Il tipo di endpoint risulta molto flessibile ed efficace per cercare con semplicità le annotazioni generate con Pundit dal momento che il vantaggio risiede nel fatto che quando si invia una query GraphQL alla propria API, si ottiene esattamente ciò di cui si ha bisogno, non dati superflui. Le app che usano GraphQL, come Pundit, sono veloci e stabili perché sono loro a controllare i dati che ricevono, non il server. L'output restituito in questo caso è conforme allo standard del W3C Web Annotation, nel formato di serializzazione JSON-LD.<sup>204</sup>

Lo scopo del tirocinio è stato duplice: da un lato verificare la conformità dell'endpoint rispetto allo standard del W3C, attraverso un test automatico, dall'altro proporre un'estensione dell'output dell'endpoint secondo ontologie web standard per gestire altre informazioni come le annotazioni semantiche, che, non sono descritte esplicitamente dallo standard Web Annotation ma che Pundit gestisce.

---

<sup>200</sup> <https://www.netseven.it/chi-siamo/>

<sup>201</sup> <https://thepund.it/>

<sup>202</sup> <https://project.gotriple.eu/>

<sup>203</sup> <https://graphql.org/>

<sup>204</sup> <https://json-ld.org/>

## 4.1.2 Metodologie e strumenti

Gli strumenti che ho utilizzato a supporto delle mie attività sono i seguenti:

- Python per lo sviluppo dello script del test automatico;
- La teoria del Semantic Web;
- Lo studio approfondito delle specifiche del W3C Web Annotation: Data Model e Protocol;
- GraphQL: studio della teoria e della sintassi per questo protocollo di interrogazione dati;
- GoogleDrive, Colab, Visual Studio Code;
- Altair, un tool per interrogare endpoint con query GraphQL<sup>205</sup>;
- l'estensione di Pundit per comprendere il suo funzionamento.

## 4.2 Test di conformità dell'endpoint di Pundit

Un endpoint è un punto di accesso ad un server<sup>206</sup> che segue un protocollo, ovvero una serie di regole che definiscono il formato delle richieste del client e delle risposte inviate dal server.<sup>207</sup>

L'endpoint del World Wide Web Consortium (W3C)<sup>208</sup> segue diverse specifiche ovvero il W3C Annotation Protocol<sup>209</sup> per definire il formato delle richieste e il tipo di dato da restituire; il W3C Annotation Data Model <sup>210</sup>per definire il formato dei dati. Le risposte sono in JSON-LD, un formato Linked Data simile al JSON ma i cui campi sono ben definiti all'interno di ontologie.<sup>211</sup>

Un esempio esplicativo è mostrato in Figura 4.1.

---

<sup>205</sup> <https://altair.sirmuel.design/>

<sup>206</sup> <https://it.wikipedia.org/wiki/Server>

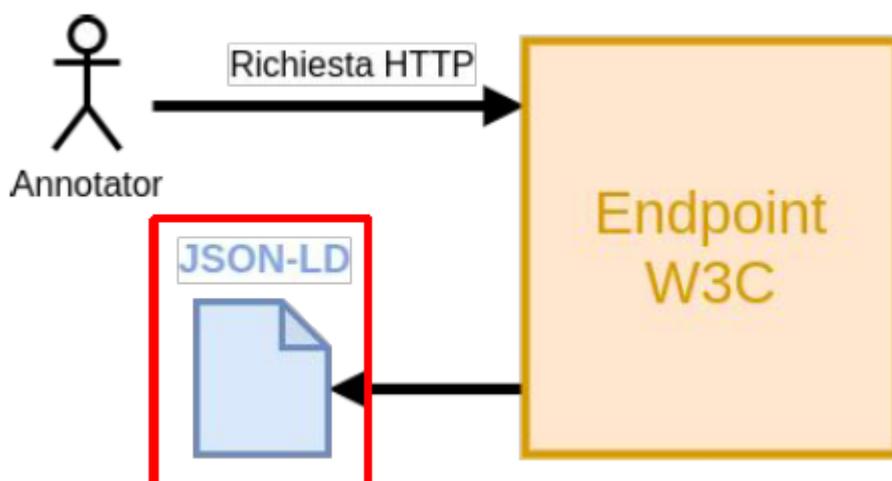
<sup>207</sup> [https://it.wikipedia.org/wiki/Communication\\_endpoint](https://it.wikipedia.org/wiki/Communication_endpoint)

<sup>208</sup> <https://www.w3.org/>

<sup>209</sup> <https://www.w3.org/TR/annotation-protocol/>

<sup>210</sup> <https://www.w3.org/TR/annotation-model/>

<sup>211</sup> <https://json-ld.org/>



**Figura 4.1:** Richiesta e risposta dell'endpoint W3C

Come detto in precedenza, la Net7 ha implementato un endpoint GraphQL per una API di ricerca che punta a restituire l'output di Pundit nel formato di serializzazione JSON-LD. Il Web Annotation Protocol manca di una API di ricerca e quello ideato da Net7 è quindi una soluzione originale.

Il mio compito è stato quello di testare la compatibilità dell'endpoint con i sottoinsiemi delle specifiche del W3C Annotation che l'API di ricerca di Net7 implementa.

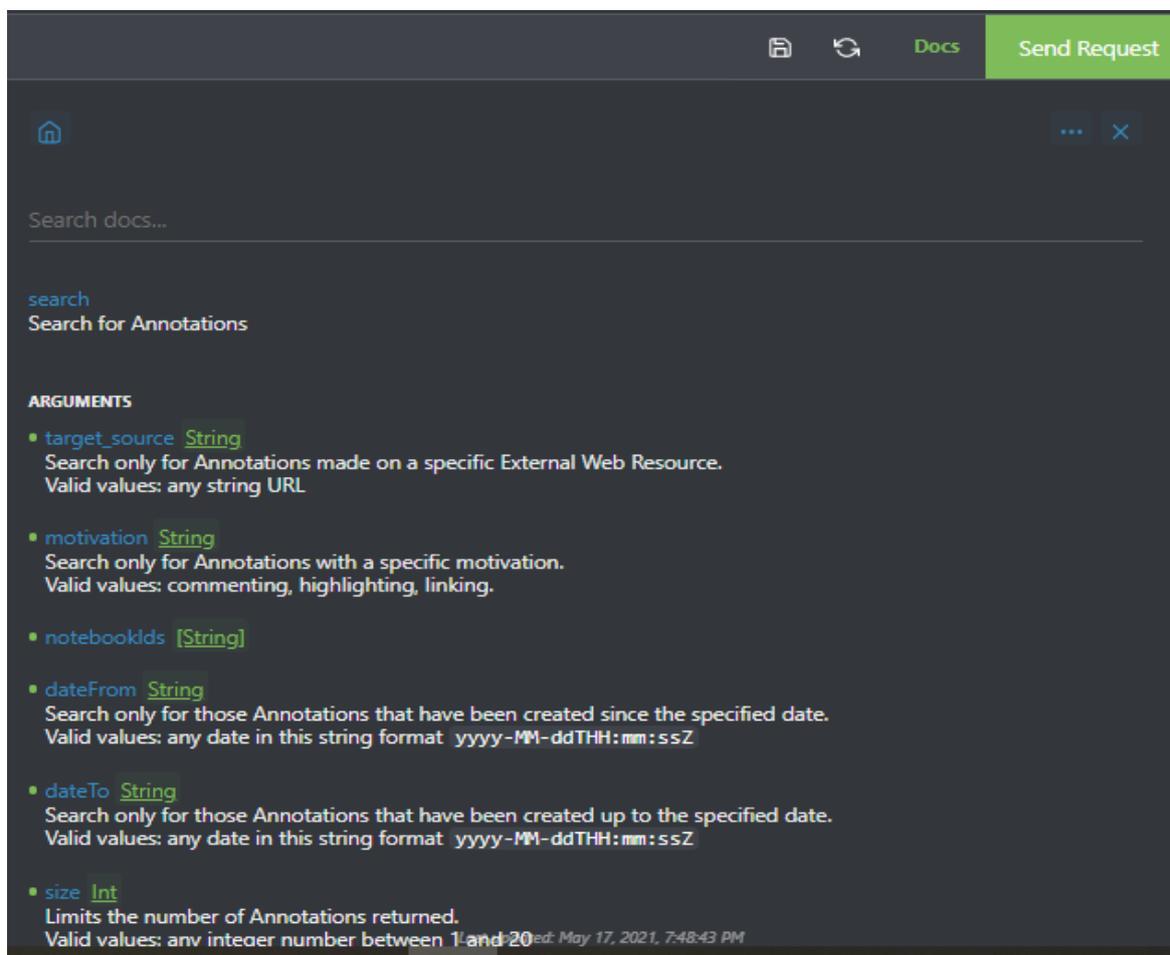
La prima parte di questa fase ha riguardato uno studio teorico del W3C e delle specifiche sopracitate, in modo che potessi avere una base solida per capire ed affrontare ciò che avrei dovuto realizzare successivamente. Ho studiato la teoria e la sintassi di GraphQL, linguaggio di interrogazione che permette di dare una descrizione completa e comprensibile dei dati nella propria API, fornendo ai clienti la possibilità di chiedere esattamente ciò che vogliono estrapolare dal database.

Le app che usano questo linguaggio sono veloci e stabili perché controllano i dati che ricevono, non il server. Inoltre, con una singola query è possibile ottenere contemporaneamente più risorse dal momento che le query GraphQL non accedono solo alle proprietà di una risorsa ma anche ai riferimenti ad essa collegati.

Dopo aver letto la documentazione, ho inizialmente utilizzato un endpoint di esempio<sup>212</sup> per prendere dimestichezza con le query GraphQL.

Successivamente, ho utilizzato l'estensione Chrome Altair<sup>213</sup>, un tool per interrogare l'endpoint<sup>214</sup> fornitomi con query GraphQL. Altair ha un'interfaccia molto intuitiva, nella sezione Docs è presente la documentazione per porre queries in maniera corretta: sono specificati i campi con i relativi tipi e anche alcuni sottocampi necessari. Con la guida del mio tutor, sono riuscita a capire con facilità come porre queries e ad analizzare la risposta che di volta in volta mi veniva fornita.

Di seguito, nella Figura 4.2 è riportata la documentazione in cui sono presenti i campi, una loro breve descrizione e i relativi tipi.



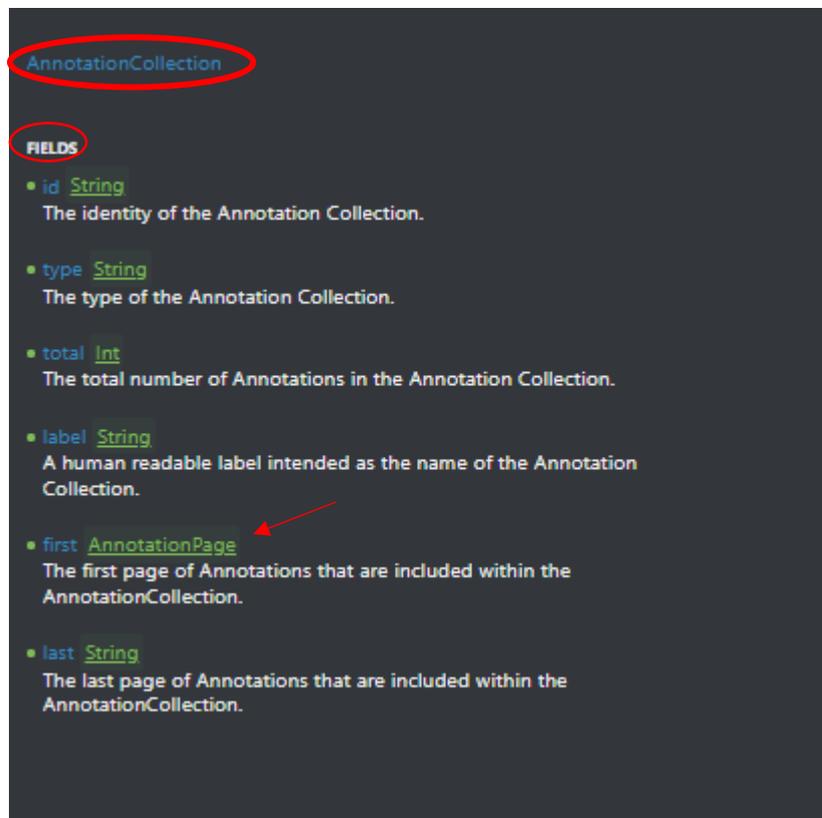
**Figura 4.2:** Documentazione in Altair

<sup>212</sup> <https://lucasconstantino.github.io/graphiql-online/>

<sup>213</sup> <https://chrome.google.com/webstore/detail/altair-graphql-client/flnheeellpciglgpaodhkhmapeljopja?hl=en>

<sup>214</sup> <https://aw1qltfpxi.execute-api.eu-south-1.amazonaws.com/w3c/annotation>

Invece, in Figura 4.3 è riportato un esempio di sottocampo, “Annotation Collection”, appartenente al campo “partOf”. Cliccando il sottocampo, compare la sua descrizione con i relativi campi ed eventualmente altri sottocampi necessari, come “AnnotationPage” per il campo “first”.



**Figura 4.3:** Esempio di sottocampo “Annotation Collection”

Una volta compreso il funzionamento, ho scritto diverse queries su Altair, cambiando di volta in volta le richieste, per ottenere risposte differenti. Di seguito, in Figura 4.4, riporto un esempio di query con associata la sua risposta.

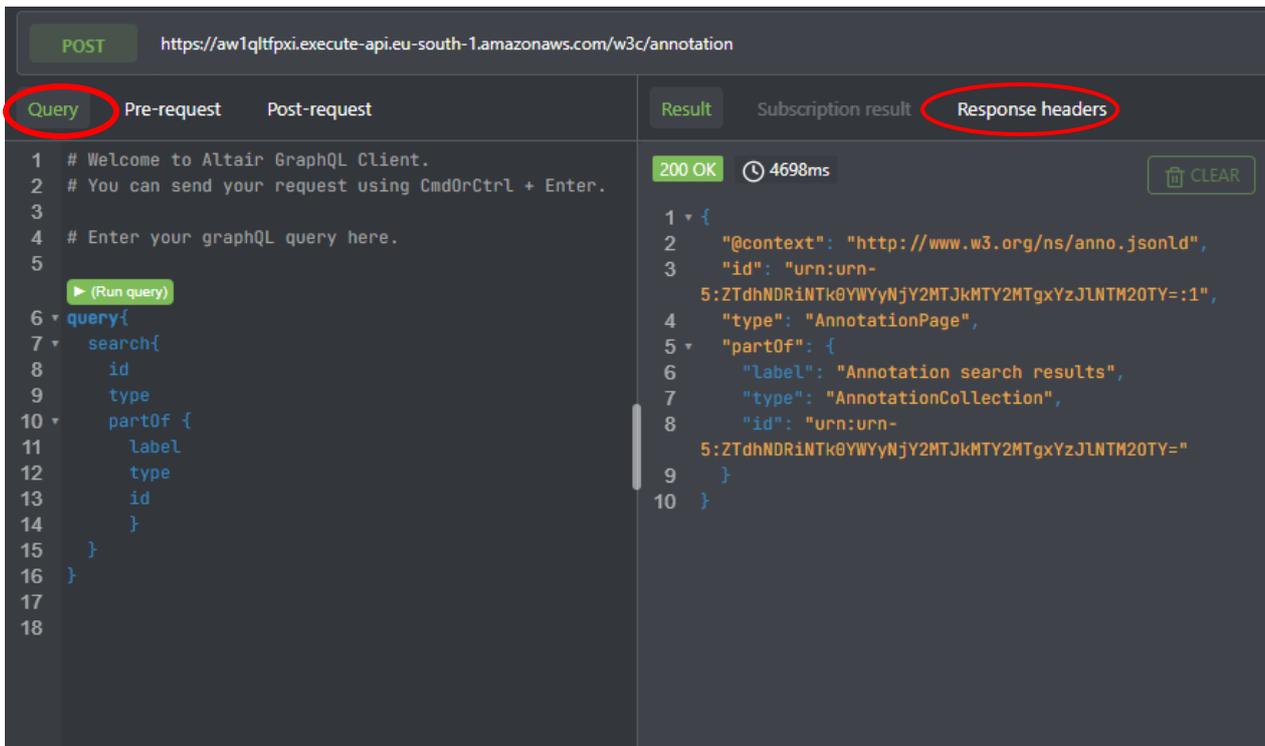


Figura 4.4: Query e relativa risposta in Altair

Inizialmente ho testato la validità degli output rispetto agli standard W3C su un tool online,<sup>215</sup> come mostrato in Figura 4.5, in cui dovevo inserire la *response* di Altair e successivamente cliccare sul tasto “check JSON”, per poi vagliare gli eventuali errori restituiti.

Fill the textarea below with JSON output from your annotation client implementation that supports the following criteria:

Web Annotations:

- MUST include certain properties (keys)
- MUST NOT have both body and bodyValue keys simultaneously
- MUST satisfy model constraints on values of required and any optional Annotation keys used

Note: Implementation of optional Annotation keys (features), optional constraints on key values, and optional keys and constraints on Agents involved in an Annotation checked by other tests.

Check JSON

Figura 4.5: Box di inserimento dell’output

<sup>215</sup> <http://w3c-test.org/annotation-model/annotations/annotationMusts-manual.html>

Dopo aver fatto qualche prova per capire il funzionamento e l'organizzazione della pagina, ho realizzato uno script di valutazione automatica in Python su Colab, abbreviazione di Colaboratory, permette di scrivere ed eseguire codice Python direttamente nel proprio browser.<sup>216</sup> Lo script è riportato in Appendice.

Per l'implementazione dello script, ho utilizzato la libreria Python Selenium<sup>217</sup>, come è specificato anche nel README del test sull'Annotation Model.<sup>218</sup> Questa libreria permette di simulare la navigazione da browser da codice, ovvero tramite codice si esegue la query di interesse, l'output viene inserito automaticamente nel riquadro presente in Figura 4.5. Successivamente si esegue e il *submit* del tasto "Check JSON" presente sulla pagine del tool online<sup>219</sup> e si estrae l'output generato dalla pagina in formato JSONLD, inserendolo in una cartella di Google Drive.<sup>220</sup>

Il test di validazione funziona in questo modo:

1. Passo manualmente allo script di validazione l'URL dell'endpoint W3C;
2. Faccio partire lo script, facendo una POST request<sup>221</sup>;
3. Faccio la print del numero totale dei test, del numero di pass e di fail (se sono tutti pass stampa "Test superato" altrimenti "Test fallito") e del tipo di fail che ci sono stati;
4. Dai risultati estrapolo solo i fail;
5. Inserisco in un dizionario: la query, l'URL dell'endpoint e i fail, li trascrivo in un file JSONLD.

---

<sup>216</sup> [https://colab.research.google.com/?hl=it#scrollTo=5fCEDCU\\_qrC0](https://colab.research.google.com/?hl=it#scrollTo=5fCEDCU_qrC0)

<sup>217</sup> <https://www.selenium.dev/about/>

<sup>218</sup> <http://w3c-test.org/annotation-model/README.md>

<sup>219</sup> <http://w3c-test.org/annotation-model/annotations/annotationMusts-manual.html>

<sup>220</sup> [https://it.wikipedia.org/wiki/Google\\_Drive](https://it.wikipedia.org/wiki/Google_Drive)

<sup>221</sup> [https://www.w3schools.com/tags/ref\\_httpmethods.asp](https://www.w3schools.com/tags/ref_httpmethods.asp)

Nello script ho utilizzato i Colab Forms<sup>222</sup>, moduli per inserire facilmente dei parametri direttamente all'interno codice. Nel mio caso, i parametri necessari sono: URL dell'endpoint, query e nome del file JSONLD che di volta in volta veniva creato e inserito in un'apposita cartella Google Drive, precedentemente collegata a Colab. Il vantaggio dei Colab Forms è che quando si cambia il valore in un modulo, il valore corrispondente nel codice cambierà, questo mi permetteva di inserire rapidamente le queries e ottenere i relativi output.

Di seguito, un output JSONLD salvato nella mia cartella Drive, in cui sono mostrati i parametri che mi interessavano:

```
{
  "query": "{search{id,type,partOf{first{id}, label} next, items {id,type,
creator, motivation}}}",
  "endpoint_url": "https://aw1qltfxi.execute-api.eu-south-
1.amazonaws.com/w3c/annotation",
  "results": [
    {
      "Result": "Fail",
      "Test Name": "1:3 Implements **Annotation_type_key** and
'**Annotation**' is **a value of type** - [model
3.1] (https://www.w3.org/TR/annotation-model/#annotations)",
      "Message": "assert_true: ERROR: Annotation is missing type key or
'Annotation' is not a value of type.; expected true got false\n
at
JSONtest.<anonymous> (http://w3c-test.org/annotation-
model/scripts/JSONtest.js:545:17)\n
at Test.step (http://w3c-
test.org/resources/testharness.js:2095:25)\n
at test (http://w3c-
test.org/resources/testharness.js:566:30)\n
at JSONtest.<anonymous>
(http://w3c-test.org/annotation-model/scripts/JSONtest.js:533:13)\n
at
Array.forEach (<anonymous>)\nAsserts run"
    },
    {
      "Result": "Fail",
      "Test Name": "1:4 Implements **target_key** - [model
3.1] (https://www.w3.org/TR/annotation-model/#annotations)",
      "Message": "assert_true: Error: Annotation is missing target key.;
expected true got false\n
at JSONtest.<anonymous> (http://w3c-
test.org/annotation-model/scripts/JSONtest.js:545:17)\n
at Test.step
(http://w3c-test.org/resources/testharness.js:2095:25)\n
at test
(http://w3c-test.org/resources/testharness.js:566:30)\n
at
JSONtest.<anonymous> (http://w3c-test.org/annotation-
model/scripts/JSONtest.js:533:13)\n
at Array.forEach
(<anonymous>)\nAsserts run"
    },
    {
      "Result": "Fail",
```

<sup>222</sup> <https://colab.research.google.com/notebooks/forms.ipynb>

```

    "Test Name": "1:5 Each **_target_** is one of a **string of format uri**, an **External Web Resource**, a **Choice** or **Set**, or a **Specific Resource** - [model 3.2](https://www.w3.org/TR/annotation-model/#bodies-and-targets), [model 4](https://www.w3.org/TR/annotation-model/#specific-resources)",
    "Message": "assert_true: ERROR: One or more Targets of the Annotation is not one of a string of format uri, an External Web Resource, a Choice or Set, a Specific Resource.; expected true got false\n    at JSONtest.<anonymous> (http://w3c-test.org/annotation-model/scripts/JSONtest.js:545:17)\n    at Test.step (http://w3c-test.org/resources/testharness.js:2095:25)\n    at test (http://w3c-test.org/resources/testharness.js:566:30)\n    at JSONtest.<anonymous> (http://w3c-test.org/annotation-model/scripts/JSONtest.js:533:13)\n    at Array.forEach (<anonymous>)\nAsserts run"
  }
]
}

```

**Figura 4.6:** Output JSONLD

Lo scopo di questa mia attività era quello di analizzare i diversi errori e capire tramite la documentazione come risolverli. Mi sono resa conto di un problema ovvero ponendo queries differenti, gli errori che il tool di validazione mi restituiva erano sempre gli stessi, evidenziati in grigio scuro in Figura 4.6.

Discutendone con il tutor e il gruppo Pundit, siamo giunti alla conclusione che fosse un problema del tool di validazione, non adatto al nostro scopo. Per tale motivo, sono stata incaricata di scrivere una mail alla comunità del W3C<sup>223</sup> per chiedere se esistesse un tool di validazione automatica per l'intero data model e non per le singole annotazioni.

Purtroppo, la mail è rimasta senza risposta ma abbiamo iniziato a ragionare su un possibile cambiamento: sostituire l'Annotation Containers<sup>224</sup> con l'Annotation Collection<sup>225</sup>. L'Annotation Container è un elemento del Web Annotation Protocol,<sup>226</sup> ed è allo stesso tempo sia un Container, un

<sup>223</sup> <https://lists.w3.org/Archives/Public/public-openannotation/>

<sup>224</sup> <https://www.w3.org/TR/annotation-model/#annotation-collection>

<sup>225</sup> <https://www.w3.org/TR/annotation-protocol/#annotation-containers>

<sup>226</sup> <https://www.w3.org/TR/annotation-protocol/ed>

servizio per la gestione delle annotazioni, sia un Ordered Collection, ovvero una lista ordinata di annotazioni.

L'Annotation Collection invece è un elemento del Web Annotation Data Model<sup>227</sup> ed è una collezione di annotazioni. Essa mantiene informazioni su se stessa, inclusa la creazione o le informazioni descrittive volte alla comprensione della collezione e anche i riferimenti ad almeno la prima pagina di annotazioni. Iniziando con la prima annotazione nella prima pagina, e attraversando le pagine fino all'ultima annotazione dell'ultima pagina, tutte le annotazioni nella collezione saranno state scoperte. Questa sostituzione porta a rinunciare completamente ad essere conformi al Data Protocol e abbracciare solo la compatibilità con le specifiche del Data Model.

Per questo motivo ho confrontato la documentazione relativa all'Annotation Collection e all'Annotation Containers, specifica del WA Protocol, giungendo alla conclusione che era possibile e funzionale questo cambiamento.

La prima implementazione dell'endpoint Pundit restituiva dati serializzati secondo le seguenti regole:

- la query restituiva risultati sotto forma di Annotation Page<sup>228</sup>;
- la prima query era serializzata come contenitore di Web Annotation, includendo anche la prima pagina serializzata come Annotation Page. Le richieste successive restituivano, invece, soltanto una Annotation Page (non un contenitore).

Dunque, vi è una differenza nel formato dei risultati tra la prima e le chiamate successive.

La decisione presa, dopo un'attenta analisi, è stata quella di creare un endpoint che restituisce sempre ed esclusivamente Annotation Page, rimuovendo qualsiasi dipendenza dal protocollo WA,

---

<sup>227</sup> <https://www.w3.org/TR/annotation-model/>

<sup>228</sup> <https://www.w3.org/TR/annotation-model/#annotation-page>

rappresentata dall'Annotation Container. La conformità dell'endpoint riguarda pertanto solo il Data Model.<sup>229</sup>

### 4.3 Estensione W3C Model: Semantic Annotation

La seconda fase del mio tirocinio ha riguardato l'estensione dell'annotazione semantica. Nella specifica W3C Data model non è prevista una relazione esplicita che associ il soggetto di un'annotazione ad un oggetto. Il mio obiettivo è stato quello di studiare una possibile serializzazione delle annotazioni semantiche in un formato conforme al Web Annotation Data model, ricercando sia tra le ontologie sviluppate dal W3C sia tra ontologie esterne.

Un'annotazione semantica è considerata come un insieme di risorse collegate, tipicamente comprendente un body, ovvero l'oggetto della tripla semantica, un target rappresentante il soggetto, il "related to" sarebbe il predicato che serve per esprimere una relazione tra le altre due entità mediante ontologie e la natura di questa relazione cambia a seconda dell'intenzione dell'annotazione. Questa prospettiva del W3C si traduce in un modello base diviso in tre parti, rappresentato in Figura 4.7.<sup>230</sup>

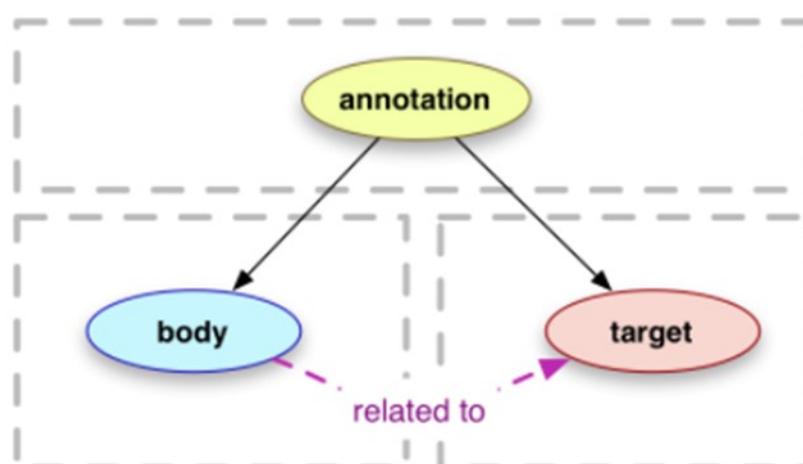


Figura 4.7: Modello di annotazione

<sup>229</sup> <https://docs.google.com/document/d/10CmEJgv3urdgzlBcEgL8wFLQrSy8dCQ3pBagJERyxiE/edit#>

<sup>230</sup> <https://www.w3.org/TR/annotation-model/#bodies-and-targets>

Le annotazioni possono essere raccolte in una lista ordinata, chiamata Collection, in modo da poter avere sia informazioni sulle singole annotazioni sia sull'intero insieme.

Il modello Collection è diviso in due sezioni: l' Annotation Collection che gestisce la lista e le Annotation Pages che elencano le annotazioni, membri della Collection.

Per essere conformi al WA data model, lo standard prevede che il modello Collection abbia la proprietà @context, presente una o più volte ma, è obbligatorio che una sia <http://www.w3.org/ns/anno.jsonld>.

Ai fini della serializzazione del predicato, per poter avere la possibilità di ampliare la gamma di scelte inserendo altre ontologie, una soluzione plausibile potrebbe essere quella di inserire oltre al @context obbligatorio, un altro @context che faccia riferimento al documento in cui sono presenti le estensioni aggiuntive. Le annotazioni possono far parte contemporaneamente di collezioni multiple proprio mediante la presenza di una o più proprietà @context.<sup>231</sup> Nella Figura 4.8 è riportato un esempio.

```
{
  "@context": [
    "http://www.w3.org/ns/anno.jsonld",
    "http://www.w3.org/ns/ldp.jsonld"
  ],
  "id": "http://example.org/annotations/?iris=1",
  "type": ["BasicContainer", "AnnotationCollection"],
  "total": 42023,
  "modified": "2016-07-20T12:00:00Z",
  "label": "A Container for Web Annotations",
  "first": "http://example.org/annotations/?iris=1&page=0",
  "last": "http://example.org/annotations/?iris=1&page=42"
}
```

**Figura 4.8:** Esempio di @context multiplo

---

<sup>231</sup> <https://www.w3.org/TR/annotation-model/#annotation-collection>

Per quanto riguarda il **target**, rappresentante il soggetto della tripla, può essere sia un'intera pagina web sia un suo frammento. La sua serializzazione può far riferimento al modello WA per la serializzazione delle External Web Resources,<sup>232</sup> un esempio è riportato in Figura 4.9, oppure può essere utilizzata l'istanza "linking" della classe Motivation che, permette di esplicitare la relazione tra una risorsa e il target,<sup>233</sup> mostrato in Figura 4.10.

```
{
  "@context": "http://www.w3.org/ns/anno.jsonld",
  "id": "http://example.org/anno2",
  "type": "Annotation",
  "body": {
    "id": "http://example.org/analysis1.mp3",
    "format": "audio/mpeg",
    "language": "fr"
  },
  "target": {
    "id": "http://example.gov/patent1.pdf",
    "format": "application/pdf",
    "language": ["en", "ar"],
    "textDirection": "ltr",
    "processingLanguage": "en"
  }
}
```

**Figura 4.9:** Serializzazione del soggetto come una External Web Resource

---

<sup>232</sup> <https://www.w3.org/TR/annotation-model/#external-web-resources>

<sup>233</sup> <https://www.w3.org/TR/annotation-model/#motivation-and-purpose>

```

{
  "@context": "http://www.w3.org/ns/anno.jsonld",
  "id": "http://example.org/anno15",
  "type": "Annotation",
  "motivation": "linking",
  "body": [
    {
      "type": "TextualBody",
      "value": "readme",
      "purpose": "tagging"
    },
    {
      "type": "TextualBody",
      "value": "A good description of the topic that bears further investigation",
      "purpose": "describing"
    }
  ],
  "target": "http://example.com/page1"
}

```

**Figura 4.10:** Serializzazione del soggetto con Motivation

Per quanto riguarda l’**oggetto** della tripla, esso va inserito all’interno del body e il suo valore può essere di diversi tipi. Lo standard permette di inserire sia target sia body multipli.

Nel caso in cui l’oggetto fosse un literal, esso può essere espresso mediante il Textual Body, soluzione preferibile, o tramite lo String Body.<sup>234</sup> Quest’ultimo è in grado di esprimere stringhe di testo ma il suo uso ha diverse restrizioni tra cui: deve essere una singola “xsd:string” e il tipo di dati non deve essere espresso nella serializzazione; deve essere interpretata come se fosse il valore della proprietà value di un Textual Body, e molte altre. Lo standard enuncia che se una qualsiasi delle restrizioni non è rispettata, allora la costruzione Textual Body deve essere usata al posto del String Body, un esempio è mostrato in Figura 4.11.

<sup>234</sup> <https://www.w3.org/TR/annotation-model/#embedded-textual-body>

```

{
  "@context": "http://www.w3.org/ns/anno.jsonld",
  "id": "http://example.org/anno7",
  "type": "Annotation",
  "body": {
    "type": "TextualBody",
    "value": "Comment text",
    "format": "text/plain"
  },
  "target": "http://example.org/target1"
}

```

**Figura 4.11:** Textual Body

Se l'**oggetto** invece è un IRI o un frammento di pagina web, potrà essere serializzato come una External Web Resource inserendolo nel body, un esempio è mostrato nella Figura 4.12 in cui è possibile vedere come nel body sia inserito un frammento della pagina, in particolare un audio.

```

{
  "@context": "http://www.w3.org/ns/anno.jsonld",
  "id": "http://example.org/anno2",
  "type": "Annotation",
  "body": {
    "id": "http://example.org/analysis1.mp3",
    "format": "audio/mpeg",
    "language": "fr"
  },
  "target": {
    "id": "http://example.gov/patent1.pdf",
    "format": "application/pdf",
    "language": ["en", "ar"],
    "textDirection": "ltr",
    "processingLanguage": "en"
  }
}

```

**Figura 4.12:** Serializzazione dell'oggetto

## 4.4 Estensione W3C Model: Altri dati e Social Feature

In questa Sezione , riporto lo studio fatto sulle informazioni relative all'utente e al modo in cui esprimerle tramite ontologie.

Le informazioni presenti nel database riguardanti l'utente sono le seguenti:

- givenName;
- familyName;
- name;
- mbox;
- thumb;
- shortBio;
- location;
- website;
- linkedin;
- twitter;
- publicProfile;
- created (annotation);
- changed (annotation).

I campi definiti dall'Annotation Model espressi mediante l'uso di diverse ontologie (foaf<sup>235</sup>, skos<sup>236</sup>, schema<sup>237</sup>) sono visibili nel seguente JSONLD: <https://www.w3.org/ns/anno.jsonld>.

Dopo un'attenta analisi, ho proposto questa organizzazione:

- givenName: *foaf:givenName*, [http://xmlns.com/foaf/spec/#term\\_givenName](http://xmlns.com/foaf/spec/#term_givenName)
- familyName: *foaf:familyName*, [http://xmlns.com/foaf/spec/#term\\_familyName](http://xmlns.com/foaf/spec/#term_familyName)
- name: *foaf:name*, [http://xmlns.com/foaf/spec/#term\\_name](http://xmlns.com/foaf/spec/#term_name)
- mbox: *foaf:mbox*, [http://xmlns.com/foaf/spec/#term\\_mbox](http://xmlns.com/foaf/spec/#term_mbox)
- thumb: *foaf:thumbnail*, [http://xmlns.com/foaf/spec/#term\\_thumbnail](http://xmlns.com/foaf/spec/#term_thumbnail)
- shortBio: *foaf:PersonalProfileDocument*, [http://xmlns.com/foaf/spec/#term\\_PersonalProfileDocument](http://xmlns.com/foaf/spec/#term_PersonalProfileDocument)
- location: *homeLocation*, <https://schema.org/homeLocation>
- website: *foaf:weblog*, [http://xmlns.com/foaf/spec/#term\\_weblog](http://xmlns.com/foaf/spec/#term_weblog)
- linkedin: <https://schema.org/sameAs>
- twitter: <https://schema.org/sameAs>

---

<sup>235</sup> <http://xmlns.com/foaf/0.1/>

<sup>236</sup> <http://www.w3.org/2004/02/skos/core#>

<sup>237</sup> <http://schema.org/>

- `publicProfile`: classe *Profile*, <https://www.w3.org/TR/activitystreams-vocabulary/#dfn-profile>
- `created:dcterms:created`, <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#http://purl.org/dc/terms/created>
- `changed:dcterms:modified`, <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#http://purl.org/dc/terms/modified>

Questa mia proposta ha messo in luce alcuni problemi. Il Public Profile, per esempio, era stato mappato male in quanto il suo valore doveva essere un booleano. Infatti, solo nel caso in cui il valore fosse TRUE doveva restituire le informazioni personali dell'utente. Inoltre, anche i valori relativi "created e "changed" non dovevano essere restituiti dall'endpoint di ricerca, riferendosi relativamente alla data di creazione e modifica dell'annotazione.<sup>238</sup>

Inoltre, abbiamo ragionato sul fatto che potesse essere utile introdurre l'affiliazione in modo tale da capire anche verso cosa fosse orientato l'uso del tool, ad esempio nel caso in cui si trattasse di uno studente le sue ricerche saranno rivolte verso articoli e letture attinenti al suo percorso di laurea.

Per quanto riguarda la formalizzazione dei profili social degli utenti, ho proposto di utilizzare la proprietà *sameAs* di [schema.org](https://schema.org)<sup>239</sup> anziché `owl:sameAs`<sup>240</sup>. Tra i vantaggi c'è la possibilità di inserire i dati in una lista oppure singolarmente, inoltre, in questo modo, possono essere utilizzati sia i markup per il JSON-LD sia implementare direttamente gli [schema.org](https://schema.org) markup.<sup>241</sup>

#### 4.4.1 Tagging

L'ultima parte della mia ricerca è focalizzata su una possibile soluzione di serializzazione per le annotazioni che includono tag.

---

<sup>238</sup> <https://www.w3.org/TR/annotation-model/#lifecycle-information>

<sup>239</sup> <https://schema.org/Person>

<sup>240</sup> <https://graphdb.ontotext.com/documentation/standard/sameAs-background-information.html>

<sup>241</sup> <https://thomas.vanhoutte.be/miniblog/schema-org-social-media-websites/>, <https://rich-snippets.io/social-profile-links/>

Il Tagging è un'istanza della classe Motivation che esprime il motivo per cui l'utente intende associare un tag al Target.<sup>242</sup>

Ci sono due tipi di relazioni attraverso le quali un tag può essere associato ad un'annotazione:

- **motivation**: esprime la relazione tra l'annotazione e la classe Motivation. Per ogni annotazione deve esserci almeno una motivation;
- **purpose**: esprime il motivo della presenza del Textual Body all'interno dell'annotazione. Ci possono essere zero o più purpose associati ad esso.

La differenza principale tra i due tipi di relazione è che la prima è legata solo all'annotazione, la seconda anche al Textual Body.

La mia scelta è stata orientata verso la seconda relazione poiché include anche caratteristiche obbligatorie del Textual body ovvero:

- **id**: proprietà che rappresenta un IRI;
- **type**: rappresenta il tipo di risorsa, il cui body deve avere la classe Textual body. È possibile aggiungere altre risorse come le Specific Resource<sup>243</sup> per indicare risorse web esterne con i propri IRI;
- **value**: proprietà che esprime il contenuto del Textual body sotto forma di caratteri.

In Figura 4.13 è riportato un esempio di serializzazione del tagging con purpose.

---

<sup>242</sup> <https://www.w3.org/TR/annotation-model/#motivation-and-purpose>

<sup>243</sup> <https://www.w3.org/TR/annotation-model/#specific-resources>

```

{
  "@context": "http://www.w3.org/ns/anno.jsonld",
  "id": "http://example.org/anno15",
  "type": "Annotation",
  "motivation": "highlighting",
  "body":
  {
    "type": "TextualBody",
    "value": "readme",
    "purpose": "tagging"
  },
  "target": "http://example.com/page1"
}

```

**Figura 4.13:** Serializzazione del tagging con purpose

Un vantaggio dell'usare il purpose riguarda proprio la possibilità di inserire risorse esterne nelle annotazioni semantiche. Un esempio dell'uso della Specific Resource è riportato in Figura 4.14.

```

{
  "@context": "http://www.w3.org/ns/anno.jsonld",
  "id": "http://example.org/anno18",
  "type": "Annotation",
  "body": {
    "type": "SpecificResource",
    "purpose": "tagging",
    "source": "http://example.org/city1"
  },
  "target": {
    "id": "http://example.org/photo1",
    "type": "Image"
  }
}

```

**Figura 4.14:** Uso di Specific Resource

## CAPITOLO 5

### Conclusioni

Nell'ampio contesto del Semantic Web, il focus del mio lavoro si è incentrato sull'importanza, uso ed evoluzione delle annotazioni web.

Prima di introdurre il concetto fondamentale, ho discusso della nascita e sviluppo del Web fino al Semantic Web, sottolineando come da un **web di documenti** rappresentante una rete in cui i nodi sono i documenti stessi collegati da link, si passa ad un **web di dati**, in cui questi ultimi diventano nodi e i link stabiliscono le relazioni tra loro. Lo scopo, però, rimane lo stesso: rendere la conoscenza disponibile a tutti, ovunque, in ogni momento e a basso costo ma, con un incremento dell'automatizzazione grazie all'introduzione di un linguaggio formale.

In particolare, ho affrontato i linguaggi costitutivi del SW ovvero RDF, RDFS e OWL, analizzandoli nel dettaglio e fornendo esempi esplicativi.

Successivamente ho introdotto il W3C, sigla di World wide web consortium, comunità internazionale che si occupa della definizione di standard web aperti per promuovere l'accessibilità e la compatibilità delle tecnologie in rete. Esso è preposto all'emanazione degli standard adottati sul Web, guidandone anche lo sviluppo e l'interoperabilità. Ho studiato alcune specifiche realizzate e fatto riferimento ai suoi documenti e standard per tutto il mio lavoro di tesi e di tirocinio.

Il cuore della mia tesi è stato sulle annotazioni, le domande che mi sono poste sono state molteplici: innanzitutto mi sono chiesta cosa fossero, le forme che potessero assumere, la loro utilità e perché sono così importanti per l'essere umano tanto da sviluppare tool e sistemi per renderle disponibili anche sul web.

Per rispondere a queste domande, sono tornata indietro nel tempo, excursus storici che mi hanno permesso di capire come le annotazioni ai margini del libro cartaceo siano nate con la scrittura.

Gli usi si sono differenziati nel corso del tempo: per memoria, per appuntare pensieri e riflessioni come nel caso di scrittori, invece, nel medioevo per scrivere maledizioni affinché non venissero rubati i libri, un bene di lusso da preservare e custodire.

L'uso delle annotazioni si è protratto nel tempo, fino ad arrivare ai giorni nostri, ricerche e sondaggi hanno dimostrato che oggi le annotazioni vengono realizzate soprattutto su libri impiegati per studio e lavoro, non sui libri di lettura che vengono percepiti come qualcosa di sacro.

Inoltre, con lo sviluppo di supporti elettronici come gli e-book, il lettore ha mostrato interesse nell'avere la possibilità di sottolineare, appuntare. Ciò che si è evinto da ricerche, studi e sondaggi, è che le annotazioni rappresentano un vero e proprio bisogno umano durante la lettura.

Da qui, nasce la necessità di fornire gli strumenti di annotazione anche sul web grazie alla creazione di tool specifici che, in modo semplice e intuitivo permettono di realizzarle. Per concludere il cerchio e avvalorare anche la mia tesi sull'importanza delle annotazioni, mi sono occupata di esaminare diversi tool, analizzandone le differenze e i punti in comune, mi sono focalizzata su uno di essi, Pundit, oggetto del mio tirocinio di cui ho discusso i suoi dettagli e funzionalità. Il mio lavoro di tesi, infatti, è nato dall'interesse nutrito verso il Semantic Web che ho deciso di approfondire svolgendo un tirocinio ad esso legato.

L'esperienza vissuta alla Net7 mi ha permesso di imparare molto, di capire ancora di più quanto il mondo umanistico e informatico siano profondamente legati, quanto il loro connubio sia necessario per soddisfare le richieste degli utenti moderni che vogliono poter utilizzare strumenti digitali avendo a disposizione le stesse funzionalità di quelli analogici.

Le annotazioni web rappresentano un bisogno moderno che il Semantic web riesce a soddisfare grazie all'utilizzo di tool intuitivi, semplici ed efficienti.

Pundit è un esempio di tool di annotazione, ma non solo, è un esempio di come gli interessi delle persone evolvano rapidamente e di come la tecnologia è chiamata a soddisfarli.

Mi sono interessata di campi nuovi e interessanti che hanno riguardato l'avanzamento del tool, sono stata parte integrante di un gruppo che mi ha permesso di approfondire le mie conoscenze, impararne di nuove e anche di esporre le mie idee.

Il tool non è ancora arrivato allo stadio finale, ha ancora altre funzionalità da aggiornare e offrire ai propri utenti. Sviluppi futuri riguardano l'introduzione di nuove funzionalità di Pundit come la finalizzazione dell'implementazione delle Annotazioni Semantiche che offrirà la possibilità di far scegliere agli utenti tra più ontologie, la definizione e la gestione di ontologie personali.

L'annotazione di documenti PDF e non solo di pagine HTM, insieme alla possibilità di condividere i propri notebook con altri utenti in modalità di sola lettura o anche scrittura.

Sarà poi riattivato il servizio FeedThePundit Proxy per permettere l'uso di Pundit con browser diversi da Google Chrome e anche con dispositivi mobili.

Vi è anche l'intenzione di consentire l'interoperabilità con le annotazioni fatte con l'annotatore web Hypothes.is: questo permetterà agli utenti di Pundit di vedere e interagire con le annotazioni fatte con questo strumento di annotazione web molto popolare, oltre a quelle create con Pundit.

## APPENDICE

```
!pip install selenium
!apt-get update # to update ubuntu to correctly run apt install
!apt install chromium-chromedriver
# !cp /usr/lib/chromium-browser/chromedriver /usr/bin

from selenium import webdriver
from selenium.webdriver.common.keys import Keys

from selenium import webdriver
import sys
import requests
import time

from google.colab import drive
drive.mount('/content/drive')

chrome_options = webdriver.ChromeOptions()
chrome_options.add_argument('--headless')
chrome_options.add_argument('--no-sandbox')
chrome_options.add_argument('--disable-dev-shm-usage')
browser = webdriver.Chrome(options=chrome_options)

browser.get('http://w3c-test.org/annotation-model/annotations/annotationMusts-
manual.html')
time.sleep(4)

# Utilizzo i Colab Forms
w3c_endpoint_url = 'https://aw1qltfpxi.execute-api.eu-south-
1.amazonaws.com/w3c/annotation' #@param {type:"string"}
query = "{search{id,type,partOf{total,label,first{id,type,next},last}, next, items
{body{type,value}target{source}}}" #@param {type:"string"}
data = {"query":query}
nome_report = 'report_6' #@param {"type": "string"}

# Effettuo una Post request
jsonld_response = requests.post(w3c_endpoint_url, data = data)
print (jsonld_response.text)

text_area = browser.find_element_by_id('annotation-input')

text_area.send_keys(jsonld_response.text)
time.sleep(4)

button = browser.find_element_by_id('annotation-run')
```

```

button.click()
time.sleep(4)

# Ottengo il numero dei test
testNumbers = print(browser.find_element_by_xpath('//*[@id="summary"]/p').text)

# Controllo il numero dei pass e dei fail
passes = print(browser.find_element_by_xpath('//*[@id="summary"]/div/div[1]').text)
fails = (browser.find_element_by_xpath('//*[@id="summary"]/div/div[2]/label').text)
if fails[0] != 0 :
    print (fails),
    print ("Test fallito")
else:
    print ("Test superato")

# Inserisco in una lista di dizionari i valori che estrapolo
tests = []
for i, a in enumerate(browser.find_elements_by_tag_name('td')):
    tests.append(a.text)

results = [{"Result": tests[i*5], "Test Name": tests[i*5+1], "Message":
tests[i*5+2]} for i in range(54)]

# Dai risultati estrapolo soltanto i Fails che inserisco in una lista
fails = [r for r in results if r['Result'] == 'Fail']

# Inserisco in un dizionario la query, la url dell'endpoint e i fails ottenuti
report = {
"query": query,
"endpoint_url": 'https://aw1qltfxi.execute-api.eu-south-
1.amazonaws.com/w3c/annotation',
"results": fails
}

#Scrivo in un file Json il dizionario
import json
with open ('./drive/MyDrive/Report/report_6.json', "w") as output:
    json.dump(report,output, indent = 2)

```

## BIBLIOGRAFIA

- [1] Alter, A. (2012, July). Your e-book is reading you. Retrieved November 18, 2012, from <http://online.wsj.com/article/SB10001424052702304870304577490950051438304.html>
- [2] Bartalesi Lenzi Valentina(2020/2021), *Corso universitario di Semantic Web*, Università di Pisa
- [3] Beckett, D., Berners-Lee, T., Prud'hommeaux, E., & Carothers, G. (2014). RDF 1.1 Turtle. *World Wide Web Consortium*, 18-31.
- [4] Bergamaschi, S., Martino, L., & Fergnani (2002), A. *Annotazione Semantica di Portali Web*
- [5] Blasucci, L. (1991). CANZONI LEOPARDIANE: dalle note autografe alle Annotazioni del 1824. *Belfagor*, 46(3), 261- 274
- [6] Bold, M. R., & Wagstaff, K. L. (2017). Marginalia in the digital age: Are digital reading devices meeting the needs of today's readers?. *Library & Information Science Research*, 39(1), 16-22
- [7] Bouquet P., Ferrario R. (2003), *Il semantic web*, SWIF – ISSN 1126–4780
- [8] Braun, L. W. (2011). Now is the time: E-books, teens, and libraries. *Young Adult Library Services*, 9 (4), 27–30
- [9] Brickley, D., Guha, R. V., & McBride, B. (2014). RDF Schema 1.1. *W3C recommendation*, 25, 2004-2014
- [10] Campanelli, M. (2002). Scrivere in margine, leggere il margine: frammenti di una storia controversa. *Scrivere in margine, leggere il margine: frammenti di una storia controversa*, 851-939
- [11] Castro, V. L. (2014). Web Semantico e Linked Open Data: best practices, prospettive e criticità. *Nuovi annali della Scuola speciale per archivisti e bibliotecari*, 28, 207.

- [12] De Santis Luca, (2021). Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration, Report on the Open Annotation Tool
- [13] Drogin, M. (1983). *Anathema!* Totowa, NJ: Allanheld & Schram
- [14] Fowler M., *Polyglot Persistence*, 2011
- [15] Fusetti, C. (2017). *Dati bibliografici e linked (open) data: prospettive e criticità* (Bachelor's thesis, Università Ca'Foscari Venezia).
- [16] Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology? In Handbook on ontologies (pp. 1-17). Springer, Berlin, Heidelberg
- [17] Guerrini, M., & Possemato, T. (2012). Linked data: un nuovo alfabeto del web semantico. *Biblioteche oggi: Mensile di informazione aggiornamento dibattito*, 30(3), 7-15.
- [18] Istituto italiano edizioni Atlas, *Definizione del tipo di documento (DTD)*
- [19] Jackson, H. J. (2017). John Adams's marginalia: then and now. In *Working with Paradata, Marginalia and Fieldnotes*. Edward Elgar Publishing
- [20] Jackson, H.J. (2001). *Marginalia: Readers writing in books*. New Haven, CT: Yale University Press
- [21] Jones, M., Campbell, B., & Mortimore, C. (2015). JSON Web Token (JWT) profile for OAuth 2.0 client authentication and authorization Grants. *May-2015*. {Online}. Available: <https://tools.ietf.org/html/rfc7523>
- [22] Kerby-Fulton, K., & Hilmo, M. (Eds.). (2001). *The Medieval Professional Reader at Work: Evidence from Manuscripts of Chaucer, Langland, Kempe, and Gower* (Vol. 85). Victoria, BC: University of Victoria
- [23] Kiryakov A. et al., *Semantic Annotation, Indexing, and Retrieval*, Web Semantics: Science, Services and Agents on the World Wide Web, Volume 2, Issue 1, pp. 49-79, 2004

- [24] Klein, M. (2001). XML, RDF, and relatives. *IEEE Intelligent Systems*, 16(2), 26-28.
- [25] Lee T.B., Hendler J., Lassila O. (2001), *A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, Scientific American: Feature Article: The Semantic Web
- [26] Lezoche, M. (2004). *Annotazione di Documenti nel Web Semantico Document annotation in the Semantic Web*, pp.196. hal-00656684
- [27] Liesaputra, V., & Witten, I. H. (2012). Realistic electronic books. *International Journal of Human-Computer Studies*, 70, 588–610
- [28] Liu, J. H. (2013, March 21). Do you write in the margins? Wired Retrieved from <http://archive.wired.com/geekdad/2013/03/do-you-write-in-the-margins/>
- [29] Lo Castro, “Web Semantico e Linked Open Data:best practices, prospettive e criticità
- [30] Marshall, C. C., & Brush, A. B. (2004, June). Exploring the relationship between personal and public annotations. In *Proceedings of the 4th ACM/IEEE-CS joint conference on digital libraries* (pp. 349-357)
- [31] McGuinness, D. L., & Van Harmelen, F. (2004). OWL web ontology language overview. *W3C recommendation*, 10(10), 2004.
- [32] Morbidoni C. et al., *Pundit: Semantically structured annotations for web contents and digital libraries*, Proceedings of the 2nd International Workshop on Semantic Digital Archives, 2012
- [33] Moro M.(2009), *WEB SEMANTICO: lo stato dell'arte*, Facoltà di Ingegneria Corso di Laurea in Ingegneria Informatica
- [34] Ong, S. P., Cholia, S., Jain, A., Brafman, M., Gunter, D., Ceder, G., & Persson, K. A. (2015). The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on Representational State Transfer (REST) principles. *Computational Materials Science*,97, 209-215

- [35] Parodi M., Ferrara A.(2002), XML, *Semantic web e rappresentazione della conoscenza*, Mondo digitale n.3
- [36] Pescarmona, F., Corno, F., Farinetti, L., & Squillero, G. (2003), *Ontologie integrate da lessici per indicizzazione e ricerca nel Semantic Web*, Facoltà di Ingegneria dell'Informazione Corso di Laurea in Ingegneria Informatica
- [37] Rossi L.C (2003), *Finalità e metodi della filologia*, ICoN-Italian Culture on the Net Services, 9 (4), 27–30
- [38] Signore O. (2003), *Strutturare la conoscenza:XML, RDF, Semantic Web*
- [39] Staiger, J. (2012). How e-books are used. *Reference & User Services Quarterly*, 51 (4), 355 -36
- [40] Tepper, S. J. (2000). Fiction reading in America: Explaining the gender gap. *Poetics*, 27(4), 255–275
- [41] Wagstaff, K. L. (2012). The evolution of marginalia. *Unpublished paper*. <http://www.wkiri.com/slis/wagstaff-libr200-marginalia-1col.pdf> [accessed 24 April 2016]
- [42] Weston P. G., Ridi R., Fusetti C. (2016), *Dati bibliografici e linked (open) data Prospettive e criticità*, Corso di Laurea magistrale(ordinamento ex D.M. 270/2004) in Storia e gestione del patrimonio archivistico e bibliografico
- [43] Wolfe, J. (2002). Annotation technologies: A software and research review. *Computers and Composition*, 19(4), 471-497
- [44] Wood, L., Le Hors, A., Apparao, V., Byrne, S., Champion, M., Isaacs, S., ... & Wilson, C. (1998). Document object model (dom) level 1 specification. *W3C recommendation*, 1

## SITOGRAFIA

[45] <https://altair.sirmuel.design/>

[46] <https://angular.io/>

[47] <https://app.thepund.it/register>

[48] <https://apps.apple.com/app/appname/id955395198?amp%3Bl=1>

[49] <https://apps.apple.com/kr/app/liner-web-pdf-highlighter/id1463530746?mt=12>

[50] <https://apps.apple.com/us/app/weava-app/id1451423065>

[51] <https://aw1qltfpxi.execute-api.eu-south-1.amazonaws.com/w3c/annotation>

[52] <https://aws.amazon.com/it/>

[53] <https://aws.amazon.com/it/microservices/>

[54] <https://aws.amazon.com/it/opensearch-service/the-elk-stack/what-is-opensearch/>

[55] <https://b2note.eudat.eu/>

[56] <https://bioportal.bioontology.org/>

[57] <https://chrome.google.com/webstore/detail/altair-graphql-client/flnheeellpciglpaodhkhmapeljopja?hl=en>

[58] <https://chrome.google.com/webstore/detail/liner-searchfasterhighl/bmhcbmnbenmcecpmpepghooflbehcack>

[59] <https://chrome.google.com/webstore/detail/pundit-annotator/eilojhemeemnpfmailfeiggpgilblaco>

[60] <https://chrome.google.com/webstore/detail/weava-highlighter-pdf-web/cbnaodkpfinfipjblikofhlhlcickei>

[61] <https://climatefeedback.org/>

[62] <https://colab.research.google.com/notebooks/forms.ipynb>

[63] <https://datascience.eu/it/matematica-e-statistica/z-test-t-test-somiglianze-e-differenze/>

[64] [https://docs.aws.amazon.com/it\\_it/lambda/latest/dg/deploying-lambda-apps.html](https://docs.aws.amazon.com/it_it/lambda/latest/dg/deploying-lambda-apps.html)

[65] <https://docs.aws.amazon.com/rds/index.html>

[66]  
<https://docs.google.com/document/d/10CmEJgv3urdgzIBcEgL8wFLQrSy8dCQ3pBagJERyxiE/edit#>

[67] <https://docs.microsoft.com/it-it/dotnet/architecture/containerized-lifecycle/design-develop-containerized-apps/monolithic-applications>

[68] <https://ec.europa.eu/programmes/horizon2020/what-horizon-2020>

[69] [https://en.wikipedia.org/wiki/JSON\\_Web\\_Token](https://en.wikipedia.org/wiki/JSON_Web_Token)

[70] [https://en.wikipedia.org/wiki/Semantic\\_triple](https://en.wikipedia.org/wiki/Semantic_triple)

[71] <https://eudat.eu/services/userdoc/b2share-usage>

[72] <https://fontistoriche.org/linked-data/>

[73] <https://getliner.com/en>

[74] <https://getliner.com/highlighter>

[75] <https://getliner.com/upgrade>

[76] <https://graphdb.ontotext.com/documentation/standard/sameAs-background-information.html>

[77] <https://graphql.org/>

[78] <https://help.dropbox.com/it-it/files-folders/share/comment-timestamp>

[79] <https://helpx.adobe.com/it/acrobat/using/commenting-pdfs.html>

[80] <https://isidore.science/>

[81] [https://it.wikipedia.org/wiki/Apparato\\_critico](https://it.wikipedia.org/wiki/Apparato_critico)

[82] [https://it.wikipedia.org/wiki/Application\\_programming\\_interface](https://it.wikipedia.org/wiki/Application_programming_interface)

[83] [https://it.wikipedia.org/wiki/Cloud\\_computing](https://it.wikipedia.org/wiki/Cloud_computing)

[84] [https://it.wikipedia.org/wiki/Commercio\\_elettronico](https://it.wikipedia.org/wiki/Commercio_elettronico)

[85] [https://it.wikipedia.org/wiki/Communication\\_endpoint](https://it.wikipedia.org/wiki/Communication_endpoint)

[86] <https://it.wikipedia.org/wiki/E-book>

[87] [https://it.wikipedia.org/wiki/Google\\_Classroom](https://it.wikipedia.org/wiki/Google_Classroom)

[88] [https://it.wikipedia.org/wiki/Horace\\_Walpole](https://it.wikipedia.org/wiki/Horace_Walpole)

[89] [https://it.wikipedia.org/wiki/Industria\\_4.0](https://it.wikipedia.org/wiki/Industria_4.0)

[90] [https://it.wikipedia.org/wiki/John\\_Milton](https://it.wikipedia.org/wiki/John_Milton)

[91]  
[https://it.wikipedia.org/wiki/Libro\\_bianco#:~:text=Con%20il%20termine%20libro%20bianco,argomento%20o%20settore%20di%20attivit%C3%A0](https://it.wikipedia.org/wiki/Libro_bianco#:~:text=Con%20il%20termine%20libro%20bianco,argomento%20o%20settore%20di%20attivit%C3%A0)

[92]  
[https://it.wikipedia.org/wiki/Lucio\\_Anneo\\_Seneca#Le\\_Epistole\\_a\\_Lucilio:\\_la\\_lettera\\_filosofica\\_come\\_genere\\_letterario](https://it.wikipedia.org/wiki/Lucio_Anneo_Seneca#Le_Epistole_a_Lucilio:_la_lettera_filosofica_come_genere_letterario)

[92] [https://it.wikipedia.org/wiki/Mary\\_Wollstonecraft](https://it.wikipedia.org/wiki/Mary_Wollstonecraft)

[93] <https://it.wikipedia.org/wiki/Node.js>, <https://nodejs.org/it/>

[94] [https://it.wikipedia.org/wiki/Northrop\\_Frye](https://it.wikipedia.org/wiki/Northrop_Frye)

[95] <https://it.wikipedia.org/wiki/PHP>

[96] <https://it.wikipedia.org/wiki/Proxy>

[97] [https://it.wikipedia.org/wiki/Samuel\\_Johnson](https://it.wikipedia.org/wiki/Samuel_Johnson)

[98] <https://it.wikipedia.org/wiki/Server>

[99] [https://it.wikipedia.org/wiki/Sistema\\_client/server](https://it.wikipedia.org/wiki/Sistema_client/server)

[100] [https://it.wikipedia.org/wiki/Test\\_di\\_verifica\\_d'ipotesi](https://it.wikipedia.org/wiki/Test_di_verifica_d'ipotesi)

[101] <https://json-ld.org/>

[102] <https://laravel.com/>

[103] <https://library.isti.cnr.it/index.php/it/supporto-alla-ricerca/open-science>

[104] <https://lists.w3.org/Archives/Public/public-openannotation/>

[105] <https://lucasconstantino.github.io/graphiql-online/>

[106] <https://mariadb.org/>

[107] <https://marketplace.eosc-portal.eu/services/b2note>

[108] <https://microsoftedge.microsoft.com/addons/detail/liner-search-fasterh/kmhegedbanhfblnoboomoeafpmojfdlp>

[109] <https://notesalong.com/>

[110] <https://operas.hypotheses.org/>

[111] <https://operas.hypotheses.org/aboutoperas/working-groups/tools-wg>

[112] <https://play.google.com/store/apps/details?id=com.getliner.liner>

[113] <https://prabook.com/web/heather.jackson/3777666>

[114] <https://project.gotriple.eu/>

[115] <https://project.gotriple.eu/gotriple-platform/>

[116] <https://remarqable.com/web/index.html>

[117] <https://remarqable.com/web/index.html#!#groups>

[118] <https://remarqable.com/web/remarq-lite.html>

[119] <https://schema.org/Person>

[120] <https://solr.apache.org/>

[121] <https://sp.eudat.eu/catalog/resources/709c053d-3b3f-4c54-8ef4-6efeea387816>

[122] <https://store.whale.naver.com/detail/ooelpmkcpjkmoffkdgefbejfgfpafhic>

[123] <https://support.getliner.com/hc/en-us/articles/115000862772-How-to-install-LINER-for-Opera->

[124] <https://thepund.it/>

[125] <https://thomas.vanhoutte.be/miniblog/schema-org-social-media-websites/>, <https://rich-snippets.io/social-profile-links/>

[126] <https://via.hypothes.is>

[127] <https://web.hypothes.is/>

[128] <https://web.hypothes.is/blog/annotation-with-scibot/>

[129] <https://web.hypothes.is/blog/ati/>

[130] <https://web.hypothes.is/blog/jhup-and-hypothesis-partner-to-enable-annotation-on-modernism-modernity/>

[131] <https://web.hypothes.is/education/>

[132] <https://web.hypothes.is/help/what-is-the-via-proxy/>

[133] <https://web.hypothes.is/journalism/>

[134] <https://web.hypothes.is/principles/>

[135] <https://web.hypothes.is/research/>

[136] <https://web.hypothes.is/start/>

[137] <http://w3c-test.org/annotation-model/annotations/annotationMusts-manual.html>

[138] <http://w3c-test.org/annotation-model/README.md>

[139] <https://www.dbpedia.org/>

[140] <https://www.elastic.co/>

[141] <https://www.eosc-hub.eu/about-us>

[142] <http://www.geonames.org/>

[143] <http://www.gnu.org/philosophy/free-sw.html>

[144] <https://www.ibm.com/it-it/cloud/learn/rest-apis>

[145] [https://www.ilsole24ore.com/art/verso-modello-ibrido-apprendimento-ADYBERs?refresh\\_ce=1](https://www.ilsole24ore.com/art/verso-modello-ibrido-apprendimento-ADYBERs?refresh_ce=1)

[146] <https://www.italiaonline.it/risorse/ux-e-ui-cosa-sono-differenze-e-importanza-2726>

[147] <https://www.json.org/json-it.html>

[148] <https://www.muruca.org/>

[149] <https://www.mysql.com/it/>

[150] <https://www.netseven.it/chi-siamo/>

[151] <https://www.netseven.it/project/europeana-sounds/>

[152] <https://www.netseven.it/project/muruca/>

[153] <https://www.netseven.it/project/semlib/>

[154] <https://www.netseven.it/servizi/digital-humanities/>

[155] <https://www.netseven.it/servizi/open-data/>

[156] <https://www.ontotext.com/knowledgehub/fundamentals/five-star-linked-open-data/>

[157] <https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf-triplestore/>

[158] <https://www.ontotext.com/products/graphdb/>

[159] <http://www.openannotation.org/spec/core/>

[160] <https://www.operas-eu.org/about/operas-in-a-nutshell/>

[161] <https://www.operas-eu.org/special-interest-group-living-book/operas-tools-research-and-development-white-paper-june-2021/>

[162] <https://www.operas-eu.org/special-interest-group-living-book/operas-tools-research-and-development-white-paper-june-2021/#Tools-Annexes>

[163] <https://www.operas-eu.org/special-interest-groups/>

[164] <https://www.redlink.com/mission-and-purpose/>

[165] [https://www.researchgate.net/publication/271296191\\_Jackson\\_H\\_J\\_Marginalia\\_Readers\\_Writing\\_in\\_Books\\_New\\_Haven\\_Conn\\_and\\_London\\_Yale\\_Univ\\_Pr\\_2001\\_324p\\_2795\\_alk\\_paper\\_ISBN\\_0300088167\\_LC\\_00-043721](https://www.researchgate.net/publication/271296191_Jackson_H_J_Marginalia_Readers_Writing_in_Books_New_Haven_Conn_and_London_Yale_Univ_Pr_2001_324p_2795_alk_paper_ISBN_0300088167_LC_00-043721)

[166] <http://schema.org/>

[167] <https://www.selenium.dev/about/>

[168] <https://www.studenti.it/samuel-taylor-coleridge-vita-libri-poesie.html>

[169] <https://www.teachology.ca/knowledgebase/what-is-hypothesis-and-how-can-it-be-used-to-annotate-the-web-for-learning/>

[170] <https://www.teachology.ca/knowledgebase/what-is-hypothesis-and-how-can-it-be-used-to-annotate-the-web-for-learning/>

[171] <https://www.treccani.it/enciclopedia/e-learning/>

[172] [https://www.treccani.it/enciclopedia/p-value\\_%28Enciclopedia-della-Matematica%29/](https://www.treccani.it/enciclopedia/p-value_%28Enciclopedia-della-Matematica%29/)

[173] [https://www.treccani.it/enciclopedia/w3c\\_%28Lessico-del-XXI-Secolo%29/](https://www.treccani.it/enciclopedia/w3c_%28Lessico-del-XXI-Secolo%29/)

[174] <https://www.treccani.it/vocabolario/web/>

[175] <https://www.w3.org/>

[176] <http://www.w3.org/2004/02/skos/core#>

[177] <https://www.w3.org/annotation/>

[178] <https://www.w3.org/annotation/diagrams/annotation-architecture.svg>

[179] <https://www.w3.org/blog/news/archives/6156>

[180] <https://www.w3.org/publishing/>

[181] <https://www.w3.org/TR/annotation-model/>

[182] <https://www.w3.org/TR/annotation-model/#annotation-collection>

[183] <https://www.w3.org/TR/annotation-model/#annotation-page>

[184] <https://www.w3.org/TR/annotation-model/#annotations>

[185] <https://www.w3.org/TR/annotation-model/#bodies-and-targets>

[186] <https://www.w3.org/TR/annotation-model/#embedded-textual-body>

[187] <https://www.w3.org/TR/annotation-model/#external-web-resources>

[188] <https://www.w3.org/TR/annotation-model/#introduction>

[189] <https://www.w3.org/TR/annotation-model/#lifecycle-information>

[190] <https://www.w3.org/TR/annotation-model/#motivation-and-purpose>

[191] <https://www.w3.org/TR/annotation-model/#specific-resources>

[192] <https://www.w3.org/TR/annotation-protocol/>

[193] <https://www.w3.org/TR/annotation-protocol/#annotation-containers>

[194] <https://www.w3.org/TR/annotation-vocab/>

[195] <https://www.w3.org/Voice/>

[196] <https://www.w3.org/WAI/>

[197] [https://www.w3schools.com/angular/angular\\_intro.asp](https://www.w3schools.com/angular/angular_intro.asp)

[198] [https://www.w3schools.com/tags/ref\\_httpmethods.asp](https://www.w3schools.com/tags/ref_httpmethods.asp)

[199] <https://www.weavertools.com/>

[200] <https://www.weavertools.com/annotate-with-weava/>

[201] <https://www.weavertools.com/collaborate-in-weava/>

[202] <https://www.weavertools.com/highlight-with-weava/>

[203] <https://www.weavertools.com/how-to-create-citations-with-weava/>

[204] <https://www.weavertools.com/organize-with-weava/>

[205] <https://www.weavertools.com/premium/>

[206] [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)

[207] <http://xmlns.com/foaf/0.1/>

[208] <https://yalebooks.yale.edu/book/9780300097207/marginalia>