# SISRED Ontologia per una base documentale

17 febbraio 2022

Andrea Pruccoli, 0900057222

Web Semantico Corso di Laurea Magistrale in Ingegneria e Scienze Informatiche Università degli studi di Bologna

# Indice

1	Intr	oduzione	2
2	2 Modello		4
3	Defi	nizione delle entità	7
	3.1	Classi	7
	3.2	Object Properties	8
	3.3	Data Properties	10
4	Definizione delle regole		13
	4.1	Regole per le Classi	13
	4.2	Regole per le Object Properties	14
	4.3	Regole per le Data Properties	16
	4.4	Regole SWRL	18
5	5 Definizione delle interrogazioni		19
6	Con	clusioni	91

#### 1 Introduzione

Il World Wide Web nacque come strumento globale basato su ipertesti collegati tra loro per fronteggiare l'urgente necessità di difforndere informazioni tra i vari istituti di ricerca sparsi per il mondo.

Il contenuto fu progettato e pensato per essere letto e facilmente compreso da utenti umani, i quali sono in grado di comprendere il contesto delle informazioni e tramite le proprie conoscenze combinare i dati dalle monteplici fonti per giungere ai propri obiettivi, spesso attraverso dispendiose ricerche manuali.

Enormi quantità di dati sono state generate nei primi decenni di vita del WWW, e più aumentava questa quantità più cresceva il bisogno, la necessità, di riuscire a mettere ordine e regolamentizzare i collegamenti che si potevano creare tra i diversi documenti e fonti, assicurandone veridicità e correttezza.

Il Web Semantico rappresenta un'estensione del WWW, il quale aggiunge concetti, gerarchie e regole che permette non solo agli umani di poter navigare e comprendere il Web, ma di dare questa stessa possibilità anche alle macchine, ai calcolatori, con l'obiettivo di creare una collaborazione tra uomo e macchina per ottimizzare le operazioni di creazione e successiva ricerca dei documenti messi a disposizione di tutti.

Tutto ciò rende possibile ad agenti automatici di comprendere il contesto semantico di una fonte o documento e interpretare le varie relazioni esistenti tra le risorse, grazie ai meccanismi e concetti definiti. Non solo, sono anche in grado di verificare la consistenza della struttura di queste risorse, potendo anche formulare asserzioni sulle stesse.

Il progetto presentato in questo documento ha come obiettivo la definizione di un'ontologia facendo uso delle tecnologie legate al Web Semantico. Tale ontologia riguarderà un reale caso d'uso in ambito lavorativo inerente ad una base documentale, in particolare verranno identificate le principali entità che ne costituiscono il dominio evidenziando le relazioni e le proprietà che le caratterizzano.

### 2 Modello

L'ontologia è stata progettata basandosi sull'analisi di un'applicazione attualmente ancora in sviluppo presso l'azienda Maggioli SpA chiamata **Sistema Redazionale**.

L'applicazione è un backoffice utilizzato principalmente dai dipendenti dell'azienda con il ruolo di redattori, che ha come obiettivo alimentare diversi siti web inerenti la Pubblica Amministrazione gestiti da Maggioli Editore. Gli utenti si occupano di inserire e mantenere Documenti di diversa tipologia, ad esempio Normativa e Giurisprudenza, e Libri.

I Documenti vengono poi collegati a Prodotti (principalmente siti web), dando possibilità ai clienti di questi ultimi di consultare il materiale messo a disposizione.

Di seguito una descrizione delle principali entità modellate:

Documento (class) costituisce l'entità principale, rappresenta un documento consultabile da parte dei clienti su uno dei siti web di Maggioli Editore. Sono distinguibili nelle seguenti diverse tipologie: Dottrina (commenti di autori su una tematica), Giurisprudenza (sentenze), Normativa (singoli articoli o allegati di una legge), Prassi (circolari, delibere, pareri), Profor (procedure e formulari, è stata evoluta in raccoglitore di tutti i documenti non classificabili nelle altre tipologie), Provvedimento (non rappresentano un singolo documento ma una raccolta di tutti i documenti di normativa che compongono una legge) e Quesiti (quesiti provenienti da diverse fonti). Ha uno o più testi associati, ma solo una alla volta verrà considerata quella principale. C'è la possibilità di poter definire un legame tra i vari Documenti impostando la proprietà di rimando, in questo modo è facile poter consultare un altro Documento (anche di diversa tipologia) che viene citato in quello attuale. È inoltre possibile possibile estrarre riferimenti a Persone, luoghi o riferimenti normativi dai testi di un Documento.

Di seguito alcune particolari caratteristiche delle sottotipologie:

- **Giurisprudenza** si possono definire testi riassuntivi per i Prodotti di cui fanno parte, chiamati Massime
- **Provvedimento** non hanno un Testo collegato poiché sono una collezione di documenti di Normativa
- **Normativa** non tutti i documenti di Normativa sono per forza associati ad un Provvedimento
- **Testo** rappresenta il testo scritto di un Documento, può derivare da un file di tipo PDF, RTF, MSWORD, JPEG o HTML
- **Prodotto** (class) rappresenta un sito web oppure una vecchia banca dati. Ad ogni Prodotto è associata una lista di documenti che sono messi a disposizione dei clienti da poter consultare
- **Libro** (class) rappresenta un testo di editoria disponibile in una delle aree tematiche del sito SistemaPA<sup>1</sup>, il testo del libro viene associato tramite PDF ed opzionalmente possono avere anche un'immagine di copertina associata
- **hasProdotto** (object property) proprietà che segna la presenza di un Documento all'interno di un Prodotto
- **hasTesto** (object property) proprietà che identifica il legame tra un Documento e uno o più entità di Testo
- **hasRimando** (object property) proprietà che permette di effettuare un'associazione tra due entità di Documento, rappresentando il riferimento che un Documento ha dell'altro

Il modello è stato realizzato utilizzando Protégé, editor per ontologie gratuito e opensource grazie alla collaborazione di una forte community da ambienti accademici, governativi e aziendali. Supporta RDF e OWL 2 ed è sviluppato in Java. Una delle funzionalità principali consiste in un classificatore in grado di convalidare la consistenza delle ontologie ed inferire nuove informazioni sulla base dell'analisi. Grazie al supporto di plugin è inoltre estensibile con funzionalità nuove e alternative.

<sup>1</sup>https://www.sistemapa.it

Si è inoltre deciso di integrare l'ontologia in esame con l'ontologia di terze parti **BIBO** (Bibliographic Ontology)<sup>2</sup> data la natura dell'argomento trattato. Infatti BIBO nasce con l'intento di esprimere citazioni e relazioni bibliografiche usando lo strumento RDF ed effettuare succesivamente delle interrogazioni su queste informazioni usando SPARQL. Ad oggi tale ontologia può essere utilizzata come base per altre ontologie che trattano lo stesso argomento, fornendo i concetti più comuni, riducendone la complessità.

<sup>&</sup>lt;sup>2</sup>https://bibliontology.com/

#### 3 Definizione delle entità

Di seguito si elencano le entità che compongono l'ontologia, integrate della libreria esterna foaf e l'ontologia esterna bibo.

#### 3.1 Classi

**Documento** elemento principale dell'ontologia, rappresenta un documento presente nella base documentale. Viene specializzato nelle tipologie *Dottrina*, *Giurisprudenza*, *Normativa*, *Prassi*, *Profor*, *Provvedimento*, *Quesiti*. È sottoclasse di *bibo:Document* 

**Provveimento** rappresenta una raccolta di documenti di tipo *Normativa* 

**Libro** rappresenta un libro presenta nella base documentale. È equivalente alla classe bibo:Book

**Prodotto** rappresenta uno dei siti dedicati ad una specifica tematica ancora attivi o una vecchia base dati

**Area Tematica** area dedicata ad una specificata tematica del sito SistemaPA<sup>1</sup>

**Testo** rappresenta lo scritto di un Documento, si specializza nelle tipologie HTML, PDF, RTF, MSWORD, JPEG

**Massima** rappresenta un estratto di un documento di tipo *Giurisprudenza* che verrà visualizzato su uno o più *Prodotti* 

**Organi** rappresenta un organo che emette un documento di tipo *Giurisprudenza*, *Normativa* o *Prassi* 

<sup>1</sup>https://www.sistemapa.it/

**Sedi** rappresenta la sede di un *Organo* a cui fa riferimento un *Documento*, si specializza nelle tipologie *Città*, *Ente*, *Regione* 

**Fonti** rappresenta la fonte di riferimento di un documento di tipo *Dottrina*, *Prassi*, *Provvedimento*, *Quesiti* 

**Tta** rappresenta la tipologia di un documento di tipo *Normativa* quando non fa parte di un *Provvedimento*, rappresenta contenuti aggiuntivi (allegati, tabelle, schede etc.)

**Tipo-Profor** rappresenta una ulteriore specializzazione di un documento di tipo *Profor* 

Autori rappresenta chi scrive il testo di un Documento. È sottoclasse di foaf:Person

**Editore** rappresenta chi cura la pubblicazione di un Libro. È sottoclasse di foaf:OrganiZation

**Utente** rappresenta un utilizzatore della piattaforma web *Sistema Redazionale*. È sottoclasse di *foaf:Person* 

Amministratore utente con pieni permessi, crea gli utenti di tipo *Redattore*Redattore utente dedicato alla redazione di documenti e alla creazione di legami che intercorrono tra essi

#### 3.2 Object Properties

**crea** definisce la proprietà di creazione dell'utenza di un utente. La proprietà inversa è creatoDa Redattore da parte di un utente Amministratore

**hasAreaTematica** associa un Documento o un Libro ad una specifica area del sito SistemaPA. La proprietà inversa è isAreaTematicaOf

**hasArticolo** definisce l'appartenenza di un documento di *Normativa* ad uno specifico *Provvedimento*. La proprietà inversa è hasProvvedimento

**hasAutore** attribuisce la creazione di un Documento ad un Autore. La proprietà inversa è isAutoreOf

**hasCittà** associa una *Città* ad un documento di *Giurisprudenza*. La proprietà inversa è isCittàOf

- **hasDocumento** associa ad un *Prodotto* un *Documento* presente nel relativo sito. La proprietà inversa è hasProdotto
- hasEstensore giudice che materialmetne redige una sentenza in un documento di Giurisprudenza, ha come range istanze di foaf:Person. La proprietà inversa è isEstensoreOf
- **hasFonte** associa una *Fonte* ad un documento di *Dottrina, Prassi, Provvedimento* o *Quesiti.* La proprietà inversa è *isFonteOf*
- hasImmagineCopertina . La proprietà inversa è isImmagineCopertinaOf
- **hasOrgano** associa un *Organo* ad un documento di *Giurisprudenza*, *Prassi* o *Provvedimento*. La proprietà inversa è isOrganoOf
- **hasSede** . La proprietà inversa è isSedeOf
- hasPresidente giudice a capo di un tribunale nella senteza di un documento di Giurisprudenza, ha come range istanze di foaf:Person. La proprietà inversa è isPresidenteOf
- hasRelatore giudice relatore in una senteza di un documento di Giurisprudenza, ha come range istanze di foaf:Person. La proprietà inversa è isRelatoreOf
- **hasRimando** permette di creare un collegamento fra due documenti che si citano a vicenda (o meno) nel testo. La proprietà inversa è *isRimandoOf*. È equivalente a *bibo:cites*
- **hasRiferimento** associa ad un *Testo* un'entità di tipo *foaf:Person*, *Sedi* o *Documento* che sono citati all'interno dello scritto e che sono stati estratti. La proprietà inversa è isRiferimentoOf
- **hasTesto** associa un *Testo* ad un *Documento* o *Libro*. La proprietà inversa è isTestoOf
- **hasTipoProfor** associa ad un documento di tipo Profor una ulteriore specializzazione. La proprietà inversa è isTipoProforOf
- **hasTta** associa una particolare tipologia anche un documento di *Normativa* quando non è incluso in un *Provvedimento*. La proprietà inversa è *isTtaOf*
- **hasMassima** rappresenta la relazione che intercorre tra un documento di *Giurispru*denza ed una relativa *Massima*. La proprietà inversa è isMassimaOf

#### **isMassimaFor**

**redige** associa l'utente Redattore all'aggiunta nella base documentale di un Documento. La proprietà inversa è redattoDa

#### 3.3 Data Properties

**activated** se valorizzata a true, indica che l'utente della piattaforma Sistema Redazionale è attualmente attivo

alias viene utilizzato per assegnare un nome alternativo ad un Provvedimento

**anno** indica l'anno di pubblicazione di un *Libro* 

**articolo** rappresenta la numerazione all'interno di un *Provvedimento* per un documento di tipo *Normativa* 

autore rappresenta l'identità completa (nome e cognome) di un Autore

**cognome** cognome di un Autore, equivalente a foaf:familyName

**collana** rappresenta la collana editoriale di cui fa parte un *Libro* 

comma comma di un articolo di un documento di Normativa

data rappresenta la data di pubblicazione di un Documento

dataAbrogazione rappresenta la data di abrogazione di un documento di Normativa

**dataFonte** rappresenta la data della *Fonte* dal quale è stato pubblicato un documento di *Dottrina*, *Prassi*, *Provvedimento* o *Quesiti* 

**dataVigore** rappresenta la data di entrata in vigore del contenuto di un documento di *Normativa* 

**descrizione** rappresenta la descrizione generica di un *Documento*, *Libro* o *Prodotto* 

**descrizioneBreve** rappresenta un riassunto della descrizione di un Libro, equivalente a bibo:shortDescription

**edizione** rappresenta l'edizione con cui è stato stampato un *Libro* 

**email** email di un *Utente* della piattaforma *Sistema Redazionale* 

**fineEffetti** rappresenta la data di fine degli effetti descritti nel testo di un documento di *Normativa* 

firstName nome proprio di un Utente, equivalente a foaf:firstName

**formato** rappresenta sotto quale formato è pubblicato il libro, se Cartaceo, Digitale o iLibro

**formatoCartaceo** nel caso il libro sia pubblicato in formato cartaceo, questa proprietà rappresenta il formato di stampa. Ad oggi i formati presenti sono 15x21, 17x24, 19x27, 16, 7x24 e 21x29, 7

image\_url URL della copertina del libro

**inizioEffetti** rappresenta la data di inizio degli effetti descritti nel testo di un documento di *Normativa* 

**lastName** cognome di un *Utente*, equivalente a *foaf:lastName* 

**login** username di un *Utente*, equivalente a *foaf:nickname* 

mese mese in cui è stato pubblicato un Libro

**nome** rappresenta il nominativo di un'entità all'interno della piattaforma

numero numero della sentenza, della prassi o del provvedimento

**numeroFonte** rappresenta numero di pubblicazione di una *Fonte* 

**numeroTta** numero di riferimento del contenuto aggiuntivo

**oggetto** breve descrizione del contenuto di un *Documento* 

pagina sito sito di riferimento di un Libro

**principale** se valorizzato a true, rappresenta che il *Testo* di un *Documento* è quello valido attualmente

**protocollo** numero progressivo assegnato al documento una volta registrato sul registro di protocollo

sezione sezione del tribunale

sigla abbreviazione per un'entità di tipo Fonti, Organi, Tipo-Profor o Tta

sottotitolo di un Libro

**testiMultipli** se valorizzato a true, indica che il *Documento* ha più di un *Testo* associato

titolo del documento

**updated** rappresenta l'ultima data di aggiornamento di un *Documento* o *Libro* 

## 4 Definizione delle regole

Per arricchire l'ontologia sono state definite alcune regole sfruttando le potenzialità di OWL e di SWRL (Semantic Web Rule Language), grazie alle quali è stato possibile estendere l'espressività di RDFS.

Queste due tecnologie mettono a disposizione costrutti per favorire il ragionamento automatico ad opera di appositi strumenti, chiamati reasoner, che implementano i processi inferenziali e deduttivi con lo scopo di rispondere alle esigenze degli utenti.

### 4.1 Regole per le Classi

- Editor: creatoDa exactly 1 Amministratore
- Testo: is Testo Of exactly 1 Documento
- Libro: editor max 1 Editore
- $\bullet$  Libro:  $equivalent To \ bibo: Book$
- Documento: hasTesto min 1 Testo
- Dottrina: hasFonte max 1 Fonti
- Giurisprudenza: hasCittà max 1 Città
- Giurisprudenza: hasEstensore max 1 Estensore
- Giurisprudenza: hasOrgano max 1 Org Giurisprudenza
- Giurisprudenza: hasPresidente max 1 Presidente
- Giurisprudenza: hasRelatore max 1 Relatore

- Giurisprudenza: hasSede max 1 Sedi
- Normativa: hasProvvedimento max 1 Provvedimento
- Normativa: hasTta max 1 Tta
- Prassi: hasFonte max 1 Fonti
- Prassi: hasOrgano max 1 Org Prassi
- Prassi: hasSede max 1 Sedi
- Prassi: hasTta max 1 Tta
- Profor: has Tipo Profor exactly 1 Tipo-Profor
- Provvedimento: hasFonte exactly 1 Fonti
- Provvedimento: hasOrgano exactly 1 Org\_Normativa
- Provvedimento: hasSede exactly 1 Sedi
- ullet Quesiti: hasFonte exactly 1 Fonti

### 4.2 Regole per le Object Properties

- crea: inverseOf creatoDa
- creatoDa: inverseOf crea
- hasAreaTematica: **inverseOf** isAreaTematicaOf
- isAreaTematicaOf: inverseOf hasAreaTematica
- hasArticolo: inverseOf hasProvvedimento
- hasProvvedimento: **inverseOf** hasArticolo
- hasCittà: **inverseOf** isCittàOf
- isCittàOf: inverseOf hasCittà
- hasDocumento: **inverseOf** hasProdotto

- hasProdotto: **inverseOf** hasDocumento
- hasEstensore: **inverseOf** isEstensoreOf
- isEstensoreOf: **inverseOf** hasEstensore
- hasFonte: **inverseOf** isFonteOf
- isFonteOf: **inverseOf** hasFonte
- hasMassima: **inverseOf** isMassimaOf
- isMassimaOf: **inverseOf** hasMassima
- hasOrgano: **inverseOf** isOrganoOf
- isOrganoOf: **inverseOf** hasOrgano
- hasPresidente: **inverseOf** isPresidenteOf
- isPresidenteOf: **inverseOf** hasPresidente
- hasProvvedimento: inverseOf hasArticolo
- hasArticolo: inverseOf hasProvvedimento
- hasRelatore: **inverseOf** isRelatoreOf
- isRelatoreOf: **inverseOf** hasRelatore
- hasRimando: **inverseOf** isRimandoOf
- hasRimando: equivalentTo bibo:cites
- isRimandoOf: **inverseOf** hasRimando
- isRimandoOf: **equivalentTo** bibo:citedBy
- hasSede: **inverseOf** isSedeOf
- isSedeOf: **inverseOf** hasSede
- hasTesto: **inverseOf** isTestoOf
- isTestoOf: **inverseOf** hasTesto

• hasTipoProfor: **inverseOf** isTipoProforOf

• isTipoProforOf: **inverseOf** hasTipoProfor

• hasTta: **inverseOf** isTtaOf

• isTtaOf: inverseOf hasTta

• redattoDa: **inverseOf** redige

• redige: **inverseOf** redattoDa

#### 4.3 Regole per le Data Properties

• activated: activated exactly 1 xsd:boolean

ullet alias: alias  $\max 1$  rdfs:Literal

• anno: anno max 1 xsd:unsignedInt

• articolo: articolo max 1 xsd:unsignedInt

• autore: autore max 1 rdfs:Literal

• bis: bis max 1 rdfs:Literal

• cognome: cognome max 1 rdfs:Literal

• cognome: cognome equivalentTo foaf:familyName

• collana: collana max 1 rdfs:Literal

• comma: comma max 1 xsd:unsignedInt

• data: data max 1 xsd:dateTime

• dataAbrogazione: dataAbrogazione max 1 xsd:dateTime

• dataFonte: dataFonte max 1 xsd:dateTime

• dataVigore: dataVigore max 1 xsd:dateTime

• descrizione: descrizione max 1 rdfs:Literal

- descrizioneBreve: descrizioneBreve max 1 rdfs:Literal
- descrizioneBreve: descrizioneBreve equivalentTo bibo:shortDescription
- edizione: edizione max 1 rdfs:Literal
- email: email exactly 1 rdfs:Literal
- alias: alias max 1 rdfs:Literal
- fineEffetti: fineEffetti max 1 xsd:dateTime
- firstName: firstName max 1 rdfs:Literal
- firstName: firstName equivalentTo foaf:firstName
- formato: formato max 1 {"Cartaceo", "Digitale", "iLibro"}
- formatoCartaceo: formatoCartaceo max 1 {"17x24", "29x24"}
- inizioEffetti: inizioEffetti max 1 xsd:dateTime
- lastName: lastName max 1 rdfs:Literal
- login: login max 1 rdfs:Literal
- login: login equivalentTo foaf:nick
- mese: mese max 1 xsd:unsignedInt
- nome: nome exactly 1 rdfs:Literal
- numero: numero **max 1** xsd:unsignedInt
- numeroFonte: numeroFonte **max 1** xsd:unsignedInt
- numeroTta: numeroTta max 1 xsd:unsignedInt
- oggetto: oggetto exactly 1 rdfs:Literal
- pagina sito: pagina sito max 1 rdfs:Literal
- previgenti: previgenti exactly 1 xsd:boolean
- principale: principale exactly 1 xsd:boolean

- protocollo: protocollo max 1 xsd:unsignedInt
- sezione: sezione max 1 rdfs:Literal
- sigla: sigla exactly 1 rdfs:Literal
- sottotitolo: sottotitolo max 1 rdfs:Literal
- testiMultipli: testiMultipli max 1 xsd:boolean
- titolo: titolo max 1 rdfs:Literal

#### 4.4 Regole SWRL

Di seguito sono riportate le regole SWRL definite nell'ontologia.

S1: Se un documento ha più testi, per il testo che non è segnato come *principale* a true viene impostato come *principale* a false

```
sisred:hasTesto(?doc, ?t2) ^ sisred:principale(?t2, true) ^
sisred:Testo(?t1) ^ sisred:Testo(?t2) ^ sisred:Documento(?doc) ^
sisred:hasTesto(?doc, ?t1) ^ owl:differentFrom(?t1, ?t2) ->
sisred:principale(?t1, false)
```

S2: Se un documento ha più testi, viene impostato testiMultipli a true

S3: I documenti di tipo *Normativa* hanno relazione di appartanenza ad elementi di tipo *Prodotto* a cui appartiene il relativo *Provvedimento* 

## 5 Definizione delle interrogazioni

In questo capitolo vengono presentate alcuni esempi di interrogazione che possono essere effettuate sull'ontologia. Le interrogazioni sono scritte nel linguaggio SPARQL che permette di interrogare fonti distribuite sul web agendo su dati rappresentati mediante RDF e OWL.

Il seguente è l'elenco dei prefissi utilizzati nelle interrogazioni presentate

1: Elenco dei prodotti rappresentati col loro nome e il numero di documenti che contengono

2: Elenco dei documenti che hanno rimandi ad altri documenti, elencando anche i documenti che sono rimando di quest'ultimi

2: Recupero di nome, editore, collana, edizione ed elenco delle aree tematiche del sito SistemaPA in cui è presente il libro con ISBN 9788891611888

#### 6 Conclusioni

Il caso di studio presentato ha permesso di approfondire, sia dal punto di vista teorico ma soprattutto pratico, i concetti principali legati al Web Semantico. Durante lo sviluppo dell'ontologia è stato infatti possibile lavorare con le varie tecnologie e gli strumenti collegati a questa estensione del Web, compreso il software open source Protégé, tool che si è rivelato molto utile nel ridurre i tempi di sviluppo del progetto.

L'ontologia sviluppata potrebbe essere inoltre considerata un punto di partenza per l'implementazione vera e propria di una Knowledge Base per la base documentale descritta in questo documento, argomento che è stato proposto diverse volte nel corso del tempo ma che non ha mai avuto reale modo di essere esaminato.