

Da Triple ad Ontologia

Davide Camino

Giugno 2024

Abstract

In questi appunti mostriamo alcuni esperimenti fatti in merito alla traduzione di triple del tipo (*soggetto*, *verbo*, *complemento*) in un'ontologia. Lo scopo di questa traduzione è strutturare i dati in modo formale per poter poi manipolarli attraverso strumenti sofisticati, sia sviluppati ad hoc da noi sia sviluppati da altri come i potenti reasoner che fanno inferenza sulle ontologie.

1 Dati di partenza

Per tradurre i dati grezzi utilizzeremo CDuce, un linguaggio di programmazione sviluppato appositamente per manipolare documenti XML. Per questa ragione le triple che andremo a tradurre dovranno essere salvate in formato XML, presentiamo un breve esempio del documento che andremo a tradurre per esemplificarne la struttura:

Listing 1: Dati Grezzi

```
1  <?xml version="1.0"?>
2  <information>
3    <info>
4      <subject>Leonardo da Vinci</subject>
5      <verb>painting</verb>
6      <complement>the Mona Lisa</complement>
7      <sex>M</sex>
8    </info>
9    <info>
10     <subject>Jane Austen</subject>
11     <verb>wrote</verb>
12     <complement>Pride and Prejudice</complement>
13     <sex>F</sex>
14   </info>
15 </information>
```

Da questo codice possiamo vedere che i dati di partenza costituiscono semplicemente un set di informazioni o asserzioni senza alcuna struttura.

Notiamo inoltre l'aggiunta di un attributo che specifica il sesso della persona, facciamo questo perché uno dei nostri obiettivi è quello di valutare i bias di rappresentazione, siamo quindi interessati a verificare se i soggetti di un certo sesso vengono associati a certi verbi più o meno frequentemente dei soggetti del sesso opposto.

2 Ontologia di arrivo

Descriviamo brevemente come sarà fatta l'ontologia che vorremmo ottenere.

2.1 Struttura

L'ontologia che vogliamo creare per dare una struttura formale ai dati grezzi sarà costituita, come ogni ontologia da classi, individui e relazioni.

2.1.1 Classi

L'ontologia che vogliamo ottenere deve rappresentare 2 classi fondamentali:

- Persone: la classe che contiene i soggetti delle nostre triple, rappresenta il concetto di persona e tutte le caratteristiche che derivano dall'appartenere alla specie Homo Sapiens Sapiens;
- Complementi: questa classe rappresenta gli oggetti su cui sono svolte le azioni. è una classe estremamente eterogenea dato che non conterrà solamente oggetti fisici, ma anche astratti quali idee, concetti e pensieri

2.1.2 Individui

Gli individui apparterranno a una delle due classi definite precedentemente, vogliamo fare in modo che se due triple di partenza parlano dello stesso soggetto o dello stesso oggetto questo sia rappresentato da un solo individuo anche nell'ontologia che creiamo. Ad esempio:

- Crick, Franklin e Watson scoprirono la struttura del DNA, da queste 3 affermazioni vorremmo ottenere 3 individui di classe Persona e un solo individuo di classe Complemento;
- Albert Einstein scoprì l'effetto fotoelettrico e ideò la teoria della relatività, da queste 2 triple vorremmo ricavare 2 complementi ma un solo soggetto.

2.1.3 Relazioni

Per ogni tripla vogliamo aggiungere una relazione alla nostra ontologia che colleghi il soggetto al complemento mediante la specifica azione; Ovviamente vorremmo collegare più soggetti allo stesso complemento (anche mediante azioni differenti) e collegare più complementi allo stesso soggetto.

2.2 Formalismo

Esistono diversi modi per esprimere struttura e contenuto di un'ontologia, alcuni formalismi sono più espressivi di altri e pagano questo aumento di espressività con una complessità maggiore sia per quanto riguarda la struttura stessa dell'ontologia che per ciò che riguarda gli algoritmi che faranno inferenza e/o reasoning sui dati contenuti nell'ontologia.

Come già detto nella sezione precedente il linguaggio di programmazione che usiamo è specifico per la manipolazione di documenti XML; le ontologie che creeremo quindi avranno la struttura di documenti XML e in particolare in questa analisi valutiamo vantaggi e svantaggi di 2 formati: XM