

# POLITECNICO DI MILANO



**Corso di Laurea in Ingegneria Informatica  
Facoltà di Ingegneria dell'Informazione**

## **USO DI ONTOLOGIE PER L'ESTRAZIONE DI INFORMAZIONI STRUTTURATE DAI WIKI**

Miglierina Carlo – Matricola 672254

Relatore: Prof. Marco Colombetti

Correlatore: Ing. Davide Eynard

# INTRODUZIONE

## Scenario attuale:

- Web 2.0: condivisione di documenti e applicazioni per il web
- Creazioni di enciclopedie gratuite on-line
- Sviluppo di wiki usati per creare siti per condividere facilmente informazioni sul web



I vantaggi dei wiki:

- Facile creazione e modifica di pagine
- Editing collettivo e condivisione delle proprie conoscenze

Ma essi non hanno dati strutturati:

- Non è possibile fare ricerche avanzate
- Non è possibile l'utilizzo di tali informazioni da parte di elaboratori

Questa tesi cerca di fornire uno strumento per superare questi limiti

# SCOPO DELLA TESI

Questa tesi nasce con i seguenti obiettivi:

- Fornire uno strumento automatico per l'estrazione di dati dai wiki
- Costruzione di ontologie, così da avere dati strutturati
- Fornire uno strumento per l'interrogazione dell'ontologia
- Ampliamento di tale base di conoscenza tramite il reasoner di Jena

Come esempio di utilizzo, l'applicazione è stata usata per scaricare informazioni sui personaggi de "Il Signore degli Anelli".

# L'APPLICAZIONE

Prima parte:

Ricerca e estrazione dei dati dai wiki e salvataggio su file .txt per poi poter ricavare i dati in maniera più comoda.

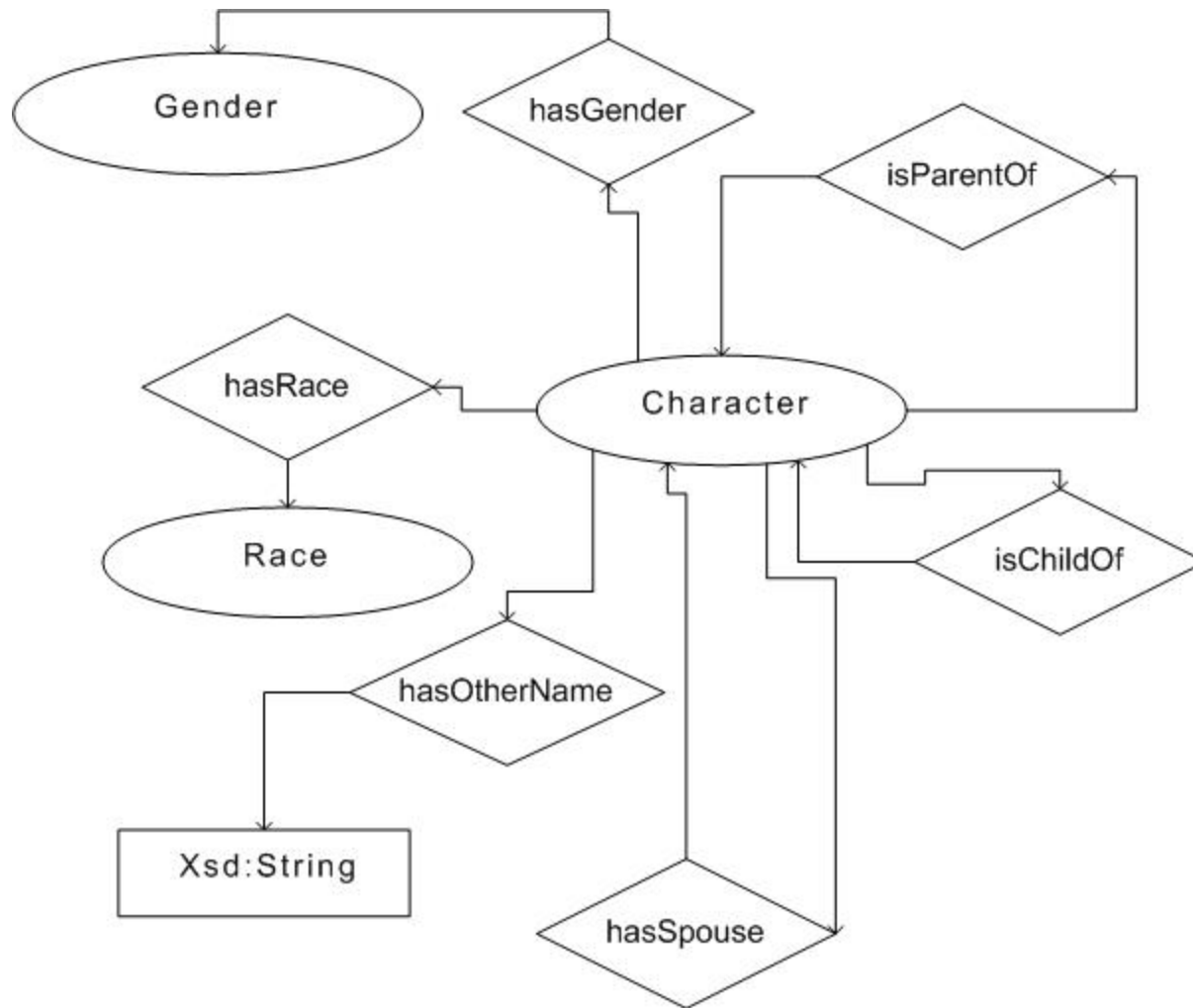
## Seconda parte:

- Costruzione dell'ontologia, cioè si definiscono le classi e le proprietà
- Popolamento dell'ontologia dopo aver scaricato i dati con gli individui, si applicano le proprietà
- Esecuzione di Query
- Applicazione regole

Per l'esempio di applicazione, sono stati utilizzati i dati presenti sui wiki:

- <http://tolkiengateway.net>
- <http://lotr.wikia.com>

# L'ONTOLOGIA CREATA




Una volta creata l'ontologia è possibile fare delle query utilizzando SPARQL.

Esse sono state poi effettuate usando le funzioni messe a disposizione da Jena.

Un primo esempio è la ricerca dei padri

```
PREFIX fo:<D:/Documenti/Poli/LotrOntology.owl#>  
SELECT DISTINCT ?x  
WHERE { ?x fo:isParentOf ?y .  
        ?x fo:hasGender fo:male.}
```





Il risultato della query è stato il seguente:

Denethor\_II  
Elrond

Si sono trovati solo due individui in quanto ci sono poche informazioni sulla proprietà isParentOf e sul genere, ma c'è la possibilità di colmare almeno parzialmente tali lacune

Per la ricerca di tutti i genitori, si è fatto uso anche della proprietà isChildOf, in questo modo è stato possibile trovare tutti i genitori

```
PREFIX fo:<D:/Documenti/Poli/LotrOntology.owl#>
```

```
SELECT DISTINCT ?x
```

```
WHERE {
```

```
    {?x fo:isParentOf ?y .} union
```

```
    { ?z fo:isChildOf ?x.}
```

```
}
```

I personaggi ritrovati sono stati i seguenti:

Elrond

Drogo\_Baggins

Belladonna\_Tuc

Primula\_Brandibuck

Finduilas\_of\_Dol\_Amroth

Bungo\_Baggins

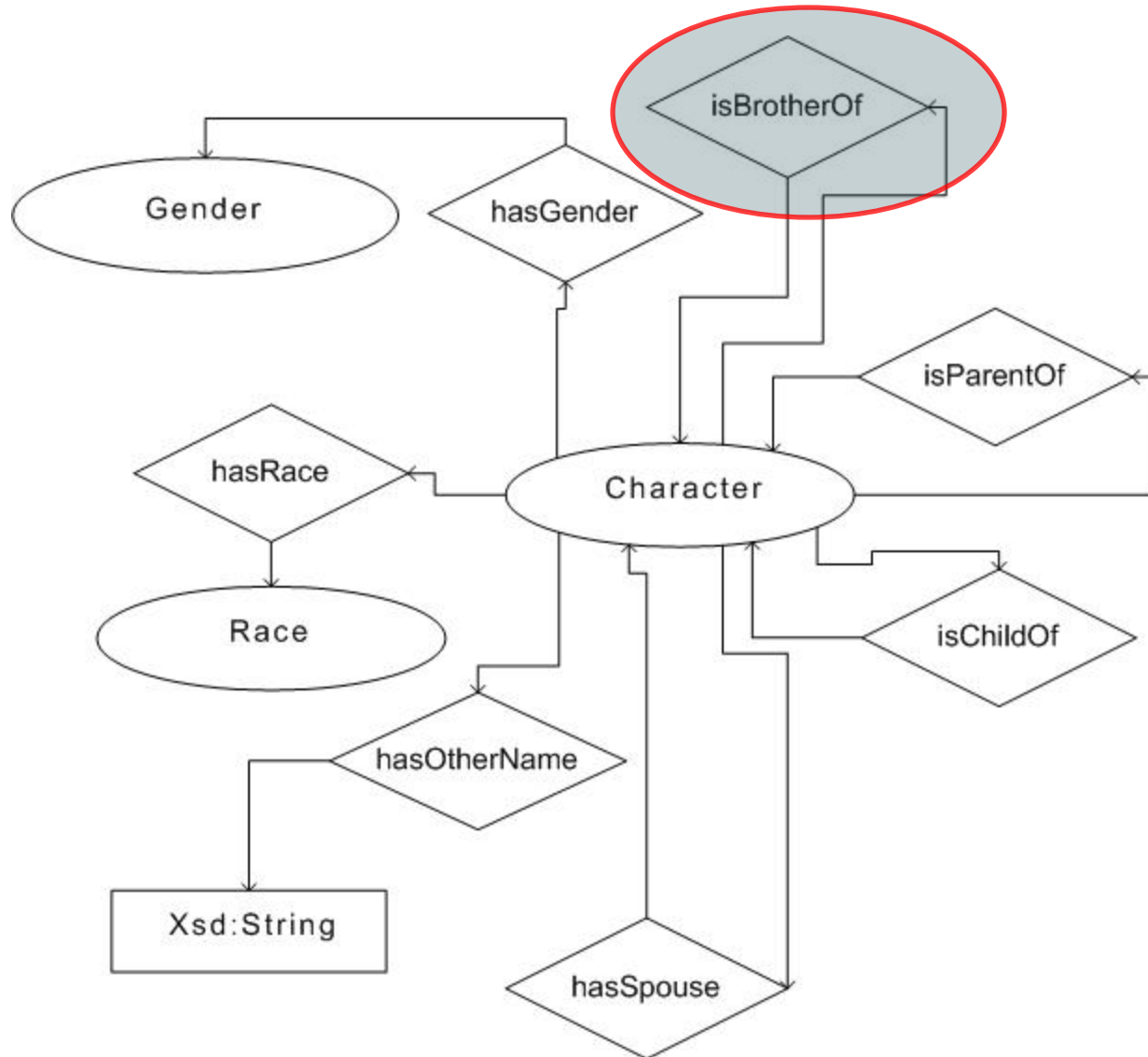
Denethor\_II

# LE REGOLE

In seguito sono state applicate le seguenti regole, con le quali è stato possibile colmare alcune lacune e aggiungere nuove informazioni:

- `?x lotr:isChildOf ?y > ?y lotr:isParentOf ?x`
- `?x lotr:isParentOf ?y and ?y lotr:hasRace ?z and > ?x lotr:hasRace ?z`
- `?x lotr:isParentOf ?y and ?x lotr:hasRace ?z and > ?y lotr:hasRace ?z`
- `?x lotr:isParentOf ?y and ?x lotr:isParentOf ?z and ?y owl:differentFrom ?z > ?y lotr:isBrotherOf ?z`

# L'ONTOLOGIA INFERITA



# RISULTATI DELLE REGOLE

- Sono aumentate le informazioni sui personaggi.
- È stata inferita nuova conoscenza

Applicando la query usando solo la proprietà isParentOf il risultato è stato:

Elrond

Drogo\_Baggins

Belladonna\_Tuc

Primula\_Brandibuck

Finduilas\_of\_Dol\_Amroth

Bungo\_Baggins

Denethor\_II

Purtroppo non è stato possibile restringere la ricerca ai soli padri, in quanto fra i personaggi trovati solo di Erlond e Denethor si sa essere maschi.

Inoltre è stato possibile ricercare i fratelli con la query:

```
PREFIX fo:<D:/Documenti/Poli/LotrOntology.owl#>  
SELECT DISTINCT ?x ?y  
WHERE { ?x fo:isBrotherOf ?y .}
```

I fratelli presenti nell'ontologia sono

- Arwen, Elladan e Elrohir
- Faramir e Boromir

# CONCLUSIONI

- Grazie all'applicazione creata è possibile estrarre dati dai wiki in maniera automatica
- Si possono fare ricerche avanzate su tali dati sfruttando l'ontologia
- È possibile ampliare le conoscenze grazie all'applicazione di regole



# SVILUPPI FUTURI

- Prendere informazioni da altri wiki per completare le informazioni
- Possibilità di estendere l'ontologia per raccogliere informazioni di altri libri o film
- Estensione o modifica dell'ontologia in modo che si possa raccogliere informazioni di qualsiasi tipo o argomento