

Università degli Studi di Torino – polo Scienze della Natura
Laurea Magistrale in Informatica
Corso: Tecnologie del Linguaggio Naturale
Parte terza - professor Luigi Di Caro

Esercitazione 5

Per questa esercitazione si è scelto di procedere con l'implementazione di un sistema di *Open Information Extraction*.

Si è utilizzato un semplice dataset di 20 frasi prese dal web relative ad alcuni film e relativi personaggi.

Lo **scopo** è estrarre delle triple da ogni frase analizzata: ogni tripla è composta essenzialmente da Soggetto, Verbo, Oggetto, dove il verbo in questo caso è l'elemento che collega gli altri due argomenti.

Per prima cosa si è effettuato un parsing su ogni frase in modo da individuare le varie entità della frase, in particolare i PoS con cui poi andare a creare la tripla.

Per la loro creazione sono state implementate alcune semplici regole in modo tale da estrarre i concetti correttamente:

- *Trovare la dipendenza "ROOT"*: l'elemento avente come dipendenza ROOT è sempre il verbo, in quanto sono i verbi gli elementi cardine di ogni frase e periodo. Ricercando tale dipendenza quindi si ottiene con certezza il verbo, che è componente della tripla;
- *Trovare le dipendenze "subj/obj"*: in questo modo si identificano appunto gli altri due componenti della tripla.

Inoltre, si è prevista la gestione di nomi composti: se nelle dipendenze delle parole analizzate viene trovato l'attributo "*compound*", allora significa che è presente un nome composto; in questo caso si procede con l'accorpamento del soggetto e del lemma associato al compound in modo da avere all'interno della tripla il nome completo relativo al soggetto della frase.

Infine, grazie all'istanziamento della classe *OieTriple* si crea un oggetto a cui assegnare gli elementi estratti dal testo di nostro interesse, successivamente forniti in output.

Risultati

Nell'immagine seguente possiamo vedere i risultati prodotti per alcune frasi analizzate:

```
Frase 1 - Darth Vader has become one of the most iconic villain
Frase 2 - Darth Vader wears a black helmet in the first trilogy
Frase 3 - Darth Vader met his son on the Death Star
Frase 4 - Darth Vader uses a red lightsaber
Frase 5 - Luke Skywalker has been trained by Yoda in the swamp
```

```
Tripla 1 - ( Darth Vader , has become , villain )
Tripla 2 - ( Darth Vader , wears , helmet )
Tripla 3 - ( Darth Vader , met , son )
Tripla 4 - ( Darth Vader , uses , lightsaber )
Tripla 5 - ( Luke Skywalker , has been trained , Yoda )
```

```
Frase 12 - Ripley covers the role of warrant officer aboard the Nostromo
Frase 13 - Ellen Ripley has a daughter named Amanda
Frase 14 - Ellen Ripley fought the xenomorph and an android to survive
Frase 15 - Ellen Ripley defended herself from the xenomorph with a firethrower
Frase 16 - Xenomorph are coming out the walls
```

```
Tripla 12 - ( Ripley , covers , role )
Tripla 13 - ( Ellen Ripley , has , daughter )
Tripla 14 - ( Ellen Ripley , fought , xenomorph )
Tripla 15 - ( Ellen Ripley , defended , herself )
Tripla 16 - ( Xenomorph , coming , walls )
```