Marco Ottina 795130

Sgaramella Davide 938632 A.A. 2019/2020

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 3

L’esercitazione proposta prevedeva l’implementazione della teoria di P. Hanks in cui si ha che il verbo è la “***radice del significato***”.

Si è deciso di usare una valenza pari a due in modo da considerare i due argomenti del verbo, in questo caso **transitivo**. Essenzialmente sono stati utilizzati due corpus di frasi generati da **Sketch Engine** relativi ad altrettanti verbi:

* To build;
* To cook.

Al fine di estrarre il filler di ogni frase viene prevista una fase di **pre-processing**: dapprima si è pulito il testo da eventuale punteggiatura, poi è stato effettuato il parsing delle frasi in modo da trovare i filler relativi alle frasi: soggetto e oggetto. Una volta trovati i possibili filler si è passati alla fase di disambiguazione degli stessi: per questo passaggio si è usato il metodo lesk() messo a disposizione dalla libreria Python “**nltk**”.

Dopo aver effettuato la fase di disambiguazione si è passati alla fase di ricerca dei super-sensi. In questo caso si è fatto ricorso al metodo lexname() che restituisce il super-senso della parola a cui viene applicato.

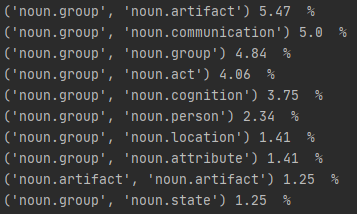
In questo modo sono state trovate le coppie di **semantic types**, successivamente fornite in output con le relative frequenze.

**Risultati**

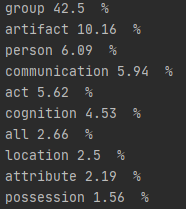
Di seguito sono mostrati i risultati ottenuti dalla sperimentazione (non tutti, solo i principali). Inoltre, si è anche previsto la generazione delle Word Cloud relative agli slot per ogni verbo analizzato.

***To build corpus***

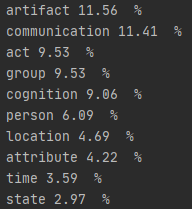
Top 10 semantic types:



Top 10 Filler slot 1:

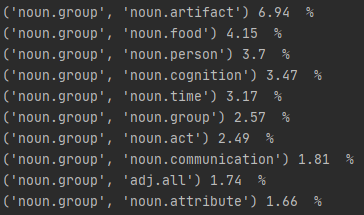


Top 10 Filler slot 2:

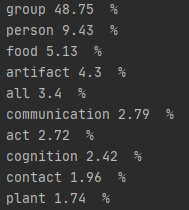


***To cook corpus***

Top 10 semantic types:



Top 10 Filler slot 1:



Top 10 Filler slot 2:

