**Università degli Studi di Torino**

Dipartimento di Informatica

Corsi di Laurea Magistrale in Informatica



**Relazione esercitazioni TLN**

**(prof. Di Caro)**

Salvatore Coluccia

a.a. 2020/2021

## Esercitazione 1

Questa esercitazione prevede il calcolo di similarità tra le definizioni date per 4 concetti mostrati a lezione.

L’obiettivo del laboratorio è calcolare in qualche modo l’agreement nelle definizioni degli studenti e capire se ci sono concetti per i quali si nota una netta differenza nelle definizioni.

L’algoritmo che ho utilizzato per calcolare queste percentuali è molto semplice e si divide in due macro-step:

1. Lemmatizzazione e rimozione delle stopword
2. Calcolo similarità tra tutte le definizioni di un dato concetto

Per la fase 1 ho utilizzato le funzioni di nltk che mette già a disposizione un elenco di stopwords inglesi che ho semplicemente rimosso dalle frasi che sono state precedentemente tokenizzate sempre utilizzando nltk.  
Sempre tramite la stessa libreria ho potuto procedere alla lemmatizzazione dei token trovati.

Per la fase 2 ho calcolato l’overlap medio dei lemmi delle definizioni di ogni concetto.

I risultati ottenuti sono i seguenti:

|  |  |  |
| --- | --- | --- |
|  | ASTRATTO | CONCRETO |
| GENERICO | 8.2% (Freedom) | 24.6% (Building) |
| SPECIFICO | 8.47% (Compassion) | 15.8% (Molecule) |

Dai risultati, per quanto chiaramente non possano essere dei dati scientificamente rilevanti dato il campione molto piccolo di esempi, si evince un agreement nettamente superiore per i concetti di tipo Concreto.  
Intuitivamente era in effetti quello che mi aspettavo in quanto descrivere un concetto concreto risulta generalmente più semplice di un concetto più astratto.

## Esercitazione 2

In questa esercitazione l’obiettivo è quello di identificare il concetto a partire da un insieme di definizioni.

Gli step che ho seguito per completare questo risultato sono i seguenti: