

Esercitazione 1

Come da consegna, sono state analizzati quattro insiemi definizioni di altrettanti concetti ed è stato assegnato un valore a ciascun insieme che rappresenti il grado di qualità, similarità e coerenza delle stesse.

Nel particolare, per ogni insieme di definizioni si sono valutate tutte le possibili coppie di definizioni utilizzando una funzione di similarità. La valutazione di un insieme di definizioni è un numero reale compreso tra 0 ed 1 che illustra la proporzione, rispetto alla cardinalità dell'insieme di coppie formato dall'insieme considerato, delle coppie di definizioni che hanno una similarità superiore ad una data soglia.

Le definizioni sono state pre-processate tramite l'eliminazione delle *stop words* ed il processo di lemmatizzazione delle parole rimanenti, così da eliminare delle parole prive di contenuto, o che ne apportino in quantità irrisorie, e facilitare il confronto tra due definizioni.

La soglia è stata fissata arbitrariamente a 0.5 e la funzione di similarità è la *funzione di sovrapposizione (overlap)*, che in sintesi relaziona la cardinalità dell'intersezione quella dell'insieme avente cardinalità minore.

Considerando che la valutazione di un insieme di definizioni è una proporzione rispetto al totale, si riporta tale valutazione come valore espresso in termini percentuali ed arrotondato ad una singola cifra decimale per una migliore leggibilità.

Di seguito, sono riportati i valori nel caso in cui la soglia sia considerata esclusiva:

- concreto-generico - building: 11.7%
- concreto-specifico - molecule: 7.6%
- astratto-generico - freedom: 2.3%
- astratto-specifico - compassion: 1.8%

Di seguito, sono riportati i valori nel caso in cui la soglia sia considerata inclusiva:

- concreto-generico - building: 29.2%
- concreto-specifico - molecule: 17.0%
- astratto-generico - freedom: 8.8%
- astratto-specifico - compassion: 4.1%

Analizzando manualmente le definizioni fornite ho potuto constatare che alcune di esse, dopo l'operazione di pre-processamento, sia sono state ridotte ad insiemi di pochi termini, ossia meno di quattro, od anche un singolo termine sia sono composti da termini molto diversi tra di loro. Considerando che la funzione di valutazione della similarità è basata sulla cardinalità degli insiemi valutati, una differenza anche minima, ossia di un singolo elemento, tra due insiemi risulta in una valutazione sensibilmente differente rispetto ad una valutazione tra due insiemi privi di tale differenza.