

Passo 3 aggiungere relazioni semantiche

Questo passo consiste nel creare relazioni semantiche fra Babel synset ed è indubbiamente il più semplice dei tre, oltre che essere il meno innovativo.

Per creare relazioni fra Babel synset si collezionano le relazioni fra concetti di WordNet (i.e. se c'era una relazione fra il senso s_ϕ^i ed il senso p_ϕ^i , allora vi sarà la stessa relazione fra il Babel synset che contiene s_ϕ^i ed il Babel synset che contiene p_ϕ^i) e di Wikipedia (i.e. internal links) e si vanno a pesare utilizzando una misura di relatività basata sul **coefficiente di Dice**

Più precisamente, per quantificare la potenza di una relazione fra Babel synsets si procede come segue:

1. Data una relazione semantica fra due sensi di WordNet s' e s'' andiamo a collezionare i sinonimi dei due sensi, le loro glosse e le glosse degli elementi contenuti nei synset a loro direttamente connessi, creando così due bag-of-words S e S' .

Le stop words vengono rimosse e le parole rimanenti vengono lemmatizzate se necessario.

2. A questo punto, il grado di associazione fra i due synset viene calcolato mediante il coefficiente di Dice:

$$\frac{2 \times |S \cap S'|}{|S| + |S'|}$$

$play_n^1$ {drama, **dramatic play**, **work**, **performance**, **dramatic work**, genre, **dramatic composition**, **television**, **actor**, **stage**, act, **subdivision**, **opera**, **ballet**, dramatic, **perform**, theater, morality play, **movie**, allegorical, satyr play, chorus, burlesque, role, stage direction, horrific, nature, macabre, playwright, dramatist, playlet};
 act_n^3 {**subdivision**, play, **opera**, **ballet**, concert dance, music, story, representation, theatrical, **perform**, trained, dancer, overture, sing, interlude, **dramatic play**, **work**, **performance**, **dramatic work**, **actor**, intend, **stage**, scene, **dramatic composition**, **television**, **movie**}.

Esempio di uso del coeff. di Dice

Confrontiamo i bag-of-words di $play_n^1$ e di act_n^3

Il coefficiente di Dice dà:
 $(2 \times 13) / (31 + 26) = 0.46$

Abbiamo quindi una forte correlazione infatti, i due insiemi contengono rispettivamente 31 e 26 termini e 13 di questi sono in comune