

# Mapping di Frame in WN Synsets

## 0. individuazione di un set di frame

Come prima operazione ciascuno deve individuare un insieme di frame (nel seguito riferito come **FrameSet**) su cui lavorare.

A tale fine utilizzare la funzione *getFrameSetForStudent(Cognome)*; nel caso il gruppo sia costituito da 2 o 3 componenti, utilizzare la funzione per trovare un FrameSet per ciascuno dei componenti del gruppo.

La funzione restituisce, dato un cognome in input, l'elenco di frame da elaborare. Gli studenti Mario Rossi e Marta Verdi eseguirebbero quindi due chiamate

```
getFrameSetForStudent('Rossi')  
getFrameSetForStudent('Verdi')
```

ottenendo in output, rispettivamente,

```
student: Rossi  
ID: 2562 frame: Manner_of_life  
ID: 1302 frame: Response  
ID: 1700 frame: Knot_creation_scenario  
ID: 2380 frame: Popularity  
ID: 1602 frame: Abundance
```

e

```
student: Verdi  
ID: 2481 frame: Erasing  
ID: 1790 frame: Means  
ID: 2916 frame: Distributed_abundance  
ID: 37 frame: Hearsay  
ID: 1816 frame: Removing_scenario
```

Questi saranno pertanto i frame utilizzati dai componenti del gruppo Rossi-Verdi.

## 1. Consegna

Per ogni frame nel FrameSet è necessario assegnare un WN synset ai seguenti elementi:

- **Frame name** (nel caso si tratti di una multiword expression, come per esempio 'Religious\_belief', disambiguare il termine principale, che in generale è il **sostantivo** se l'espressione è composta da NOUN+ADJ, e il **verbo** se l'espressione è composta da VERB+NOUN; in generale l'elemento fondamentale è individuato come il **reggente dell'espressione**).
- **Frame Elements (FEs)** del frame; e
- **Lexical Units (LUs)**.

I contesti di disambiguazione possono essere creati utilizzando le definizioni disponibili (sia quella del frame, sia quelle dei FEs), ottenendo  $Ctx(w)$ , il contesto per FN terms  $w$ .

Per quanto riguarda il contesto dei sensi presenti in WN è possibile selezionare glosse ed esempi dei sensi, e dei loro rispettivi iponimi e iperonimi, in modo da avere più informazione, ottenendo quindi il contesto di disambiguazione  $Ctx(s)$ .

## 2. Algoritmi di mapping

Il mapping può essere effettuato utilizzando (almeno) uno fra i due approcci descritti nel seguito.

- **Approccio a bag of words**, e scelta del senso che permette di massimizzare l'intersezione fra i contesti. In questo caso lo score è calcolato come

$$score(s, w) = |Ctx(s) \cap Ctx(w)| + 1$$

sarà selezionato il senso che massimizza lo  $score(s, w)$ .

- **Approccio grafico.** In questo caso si procede con la costruzione di un grafo che contiene tutti i synset associati ai termini in Framenet (FN)

$$FN = w(FEs) \cup w(LUs)$$

Ogni termine  $w$  appartenente ai FEs o alle LUs del frame viene mappato su un senso  $s$  di WordNet ( $WN$  nel seguito). Il mapping corrisponde all'argomento (su tutti i sensi  $s$  associati a  $w$  in  $WN$ ) che massimizza la probabilità condizionata di ottenere il senso  $s$  dal termine  $w$ :

$$\mu(w) = \arg \max_{s \in Senses_{WN}(w)} p(s|w) = \arg \max_s p(s, w)$$

può essere calcolato come  $p(s, w)$  con la formula

$$p(s, w) = \frac{score(s, w)}{\sum_{\substack{s' \in Senses_{WN}(w), \\ w' \in Words_{FN}(w)}} score(s', w')}.$$

Il contesto di disambiguazione dei termini  $w$  appartenenti a  $FN$ , riferito nella formula sottostante come  $Ctx(w)$ , è costruito utilizzando i termini (lemmatizzati e dopo filtraggio delle stopwords) presenti nelle descrizioni del frame e dei vari FEs.

Lo  $score(s, w)$  è calcolato sfruttando i synset associati ai termini da mappare e al loro contesto di disambiguazione. In particolare, si tratta di costruire il sottografo di  $WN$  contenente i synset presenti in tutti i path di lunghezza  $l \leq L$  (sperimentare per diversi  $L$  a partire da  $L=3$ ) fra i possibili sensi dei termini del contesto di disambiguazione e  $s$ ; la funzione che valuta l'importanza del senso  $s$  per il termine  $w$  calcola la seguente funzione:

$$score(s, w) = \sum_{cw \in Ctx(w)} \sum_{s' \in Senses_{WN}(cw)} \sum_{p \in paths_{WN}(s, s')} e^{-(len(p)-1)}$$

### 3. Valutazione dell'output del sistema

La valutazione dei risultati del mapping è fondamentale. A questo fine è necessario annotare con WN synset ID (ed eventualmente uno o due termini del synset) tutti gli elementi da mappare, e quindi

- **Frame name** (nel caso si tratti di una multiword expression, come per esempio 'Religious\_belief', cercare l'intera multiword su WordNet; se presente annotare con il relativo synset ID; diversamente disambiguare il termine principale, che in generale è il **sostantivo** se l'espressione è composta da NOUN+ADJ, e il **verbo** se l'espressione è composta da VERB+NOUN;
- **Frame Elements (FEs)** del frame; e
- **Lexical Units (LUs).**

Nel caso uno dei frame estratti non sia presente in WN sceglierne un altro rieseguendo il programma `getFrameSetForStudent()` per l'individuazione dei frame utilizzando l'iniziale minuscola del cognome dello studente. Per esempio, nel caso si verificasse che

```
ID: 1884 frame: Trendiness
```

non è presente in WordNet, ripetere l'esecuzione

```
getFrameSetForStudent('verdi')
```

ottenendo

```
student: verdi
ID: 1462 frame: Explaining_the_facts
ID: 7 frame: Motion
ID: 2051 frame: Emotions_by_possibility
ID: 1153 frame: Inclusion
ID: 1441 frame: Rising_to_a_challenge
```

Sarà da mappare il termine `explain` dal primo frame `Explaining_the_facts`.

La correttezza dell'output del sistema sviluppato è da calcolare in rapporto all'annotazione effettuata manualmente. Quindi l'annotazione costituisce un elemento molto importante nello svolgimento dell'esercitazione.

Il programma implementato dovrà quindi fornire anche la funzionalità di valutazione, che confronterà i synset restituiti in output dal sistema con quelli annotati a mano dalla studentessa o dallo studente; su questa base deve essere calcolata l'accuratezza del sistema, semplicemente come rapporto degli elementi corretti sul totale degli elementi.

### **Opzionale**

Confronto fra l'output dei due approcci descritti (bag-of-words e con grafo) .