

La terza esercitazione prevede:

1. Come prima operazione ciascuno deve individuare un insieme di frame (nel seguito riferito come `FrameSet`) su cui deve lavorare, attraverso la funzione `getFrameSetForStudent(cognome)`. La funzione restituisce, dato un cognome in input, l'elenco di frame da elaborare.
2. Per ogni frame nel `FrameSet` è necessario assegnare un WN synset ai seguenti elementi:
 - *Frame name* (nel caso si tratti di una multiword expression (es. "Religious_belief") disambiguare il termine principale, che in generale è il sostantivo se l'espressione è composta da NOUN+ADJ, o il verbo se l'espressione è composta da VERB+NOUN;
 - *Frame Elements* (FEs) del frame;
 - *Lexical Units* (LUs).

I contesti di disambiguazione possono essere creati utilizzando le definizioni disponibili (sia quella del frame, sia quelle dei FEs), ottenendo `Ctx(frame)`. 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(synset)`.

Il mapping può essere effettuato utilizzando (almeno) uno fra i due approcci:

- Approccio bag of words (scelto per questa esercitazione);
 - Approccio grafico.
3. L'ultima parte prevede la valutazione dell'output del sistema.
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 dallo studente; su questa base deve essere calcolata l'accuratezza del sistema, semplicemente come rapporto dei corretti sul totale.

1.2 Svolgimento

Il `FrameSet` individuato in base al cognome è il seguente:

student: Pellegrino

ID: 1178	frame: Interior_Profile_Relation
ID: 189	frame: Quantified_Mass
ID: 2113	frame: History
ID: 1017	frame: Noise_Makers
ID: 2972	frame: Transition_to_a_quality

Per ogni frame presente nel `FrameSet` il metodo `getFrameSynsets(frame id)` restituisce 3 liste contenenti i synset associati automaticamente al nome, ad ogni *element* e ad ogni *lexical unit* (se presente) del frame. Il metodo, con un approccio *bag of words*, dati il contesto del frame (calcolato come nelle disposizioni della consegna dal metodo `getFrameData(frame id)`) ed il contesto di ogni synset associato alle parole sulle quali mappare un senso (calcolato con il metodo `getSynsetContext(synset)`) usa l'algoritmo di lesk per restituire il senso più probabile.

La funzione `getFrameSynset()` ritorna dunque:

- Il miglior synset determinato per il frame name (`getNameSynset(frame name, frame context)`);
- Il miglior synset per ogni element del frame (`getElementSynset(frame name, frame context)`);
- Il miglior synset per ogni lexical unit del frame (`getLuSynset(frame name, frame context)`).

Mapping dei Synset sul Frame name

La funzione `getElementSynsets(frame name, frame context)` prende in ingresso il contesto ed il nome del frame.

Nel caso in cui il frame name sia una multiword expression, cioè contiene il carattere “_”, è necessario disambiguare il termine principale. Per fare questo si calcola il part-of-speech del frame name diviso in token e il termine principale corrisponderà al verbo o al nome. Per ogni synset associato al frame name viene calcolato l’overlap tra il contesto del frame e quello del synset. Il contesto del synset (calcolato tramite `getSynsetContext(synset)`) è composto dai seguenti elementi:

- La definizione del synset;
- Gli esempi d’uso del synset;
- Le definizioni e gli esempi d’uso degli iperonimi del synset;
- Le definizioni e gli esempi d’uso degli iponimi del synset.

Il senso più probabile è quello con overlap maggiore.

Mapping dei Synset sui Frame elements

La funzione `getElementsSynsets(frame elements, frame context)` prende in ingresso il contesto e la lista di elementi del frame. Per ogni frame element, vengono determinati i synset associati e si applica l’algoritmo di lesk. Il synset che ottiene l’overlap maggiore verrà scelto per essere mappato come senso più probabile per quell’elemento.

Il metodo restituisce una lista contenente tutti i synset relativi ai frame element calcolati.

Mapping dei Synset sulle Lexical units

La funzione `getLuSynset(frame lexical units, frame context)` prende in ingresso il contesto del frame e la lista di lexical units (se presenti). Per ogni lexical unit, vengono determinati i synset associati e si applica l’algoritmo di lesk. Il synset che ottiene l’overlap maggiore verrà scelto per essere mappato come senso più probabile per quella unità lessicale.

Il metodo restituisce una lista contenente tutti i synset relativi alle lexical units del frame calcolati.

Le liste ottenute vengono poi utilizzate per costruire un dizionario per ogni frame, con campi: *name*, *synsetName*, *elements*, *synsetElement*, *Lexical units*, *synsetLu*.

—

La seconda parte dell’esercitazione prevede la valutazione dell’annotazione manuale rispetto ai risultati ottenuti. L’annotazione manuale prevede il mapping tra il frame e il synset più appropriato sulla base dell’elemento del frame (frame name, frame element e lexical unit) e viene effettuata andando ad utilizzare wordnet sul web (<http://wordnetweb.princeton.edu/perl/webwn>), confrontando le definizioni di ogni parte del frame da mappare e scegliendo il senso che rappresenta di più quello indicato sul frame. Ogni frame mappato viene rappresentato come un dizionario con campi: *name*, *synsetName*, *elements*, *synsetElement*, *Lexical units*, *synsetLu*. L’accuratezza viene calcolata con una semplice formula:

$$(\text{associazioni corrette} / \text{numero totale di elementi}) * 100$$

Ed i risultati sono i seguenti:

The accuracy for the frame 1 is : 57.14
The accuracy for the frame 2 is : 60.94
The accuracy for the frame 3 is : 22.22
The accuracy for the frame 4 is : 52.63
The accuracy for the frame 5 is : 50.0