L'esercitazione prevede i seguenti passaggi:

1. Caricamento dei dati content-to-form (presente su Moodle);
2. Preprocessing (si veda esercitazione precedente, a vostra scelta);
3. Utilizzo di WordNet come sense inventory, per inferire il concetto descritto dalle diverse definizioni;
4. Definire ed implementare un algoritmo (efficace ma anche efficiente) di esplorazione dei sensi di WordNet, usando concetti di similarità (tra gloss e definizioni, esempi d'uso, rappresentazioni vettoriali, etc.);

- Suggerimento A: sfruttare principi del genus-differentia;
- Suggerimento B: sfruttare tassonomia WordNet nell'esplorazione;
- Suggerimento C: pensare a meccanismi di backtracking.

2.2 Svolgimento

Inizialmente viene caricato il file contenente le definizioni relative ai termini. Vengono poi create otto liste ciascuna relativa alle definizioni di un particolare concetto. Per ogni set di definizioni viene applicato un pre-processing (tokenizzazione, rimozione stopwords e punteggiatura, lemmatizzazione) e vengono calcolati i termini più comuni attraverso la funzione `getCommonWords()`. La funzione ritorna una lista ordinata in base alla frequenza dei termini presenti.

Dopo aver calcolato i termini più frequenti nelle definizioni di un termine, vengono calcolati i sensi associati ai 10 termini più frequenti. Questo viene fatto perchè si ipotizza che i termini più frequenti siano quelli più rilevanti nella definizione del concetto.
La ricerca dei synset si basa sul meccanismo Genus-differentia, secondo il quale la descrizione di un concetto è composta da due parti:

- Genus, include il concetto da definire in una tassonomia, ovvero prevede di descrivere il concetto attraverso un suo iperonimo;
- Differentia, porzione della definizione che differenzia il concetto dal genus (discriminante).

In base a questo principio, la ricerca dei synset associati alle parole viene fatta su iperonimi e iponimi. Nel metodo `getConcept()` per ogni parola (delle 10 più frequenti), viene inizialmente calcolata la lista di synset associati ad essa e viene calcolato l'overlap tra il contesto del synset (definizione ed esempi) e le definizioni del concetto (`intersection()`). Per ogni synset, la tupla (*synset, overlap*) viene aggiunta alla lista che sarà poi restituita dal metodo `getConcept()`. Lo stesso procedimento viene applicato agli iperonimi ed agli iponimi della parola in esame, i cui synset trovati andranno ad arricchire la lista *overlaps_list*. Alla fine, le tuple all'interno della suddetta lista vengono ordinate in maniera decrescente a seconda del loro valore di overlap e vengono restituite solo le pirme 10.

I risultati sono elencati nella tabella più in basso. Come si può notare, solo in un caso l'algoritmo mappa in maniera corretta il synset alle definizioni (`'greed.n.01'`) mentre in 5 casi viene individuato il contesto corretto (`'right.n.01', 'responsiveness.n.02', 'section.n.03', 'central_heating.n.01', 'pin.n.09'`). Nel caso dei termini *"food"* e *"vehicle"* il synset trovato non è corretto. Riporto i 10 migliori synset individuati dall'algoritmo proprio per i suddetti termini:

food
[*Synset('parasite.n.01'), Synset('embryo.n.02'), Synset('omnivore.n.02'), Synset('reservoir.n.04'), Synset('food.n.01'), Synset('organism.n.01'), Synset('herbivore.n.01'), Synset('predator.n.02'), Synset('reseed.v.02'), Synset('material.n.01')*]

Notiamo che food è presente all'interno di questa lista.

vehicle
[*Synset('play.v.24'), Synset('air_transportation_system.n.01'), Synset('highway_system.n.01'), Synset('airlift.n.01'), Synset('transport.v.02'), Synset('move.v.02'), Synset('move.v.02'), Synset('move.v.02'), Synset('move.v.02'), Synset('transport.v.02')]*

Tabella dei risultati

Correct term	Synset	Synset definition
justice	`Synset('right.n.01')`	an abstract idea of that which is due to a person or governmental body by law or tradition or nature; ; - Eleanor Roosevelt.
patience	`Synset('responsiveness.n.02')`	the quality of being responsive; reacting quickly; as a quality of people, it involves responding with emotion to people and events.
greed	`Synset('greed.n.01')`	excessive desire to acquire or possess more (especially more material wealth) than one needs or deserves.
politics	`Synset('section.n.03')`	a distinct region or subdivision of a territorial or political area or community or group of people.
food	`Synset(' parasite.n.01')`	an animal or plant that lives in or on a host (another animal or plant); it obtains nourishment from the host without benefiting or killing the host.
radiator	`Synset('central_heating.n.01')`	a heating system in which air or water is heated at a central furnace and sent through the building via vents or pipes and radiators.
vehicle	`Synset('play.v.24 ')`	cause to move or operate freely within a bounded space.
screw	`Synset('pin.n.09')`	a small slender (often pointed) piece of wood or metal used to support or fasten or attach things.