L'esercitazione prevede i seguenti passaggi:

1. Caricamento dei dati sulle definizioni (file *definizioni.xls* o documento Google presente su Moodle);
2. Preprocessing (su frequenza minima dei termini, stemming, etc. a vostra scelta);
3. Calcolo similarità tra definizioni (cardinalità dell'intersezione dei termini normalizzata su lunghezza minima tra le due, o varianti a scelta);
4. Aggregazione sulle due dimensioni (concretezza / specificità come da schema in basso);
5. Interpretazione dei risultati e scrittura di un piccolo report (da inserire nel vostro portfolio per l'esame).

	Astratto	**Concreto**
Generico	%	%
Specifico	%	%

1.2 Svolgimento

Le definizioni presenti nel file *definizioni.csv* riguardano i seguenti termini:

- building (concreto generico);
- molecule (concreto specifico);
- freedom (astratto generico);
- compassion (astratto specifico).

Dopo aver letto il file, viene applicato un pre-processing ad ogni definizione, attraverso le seguenti operazioni:

- tokenizzazione;
- rimozione stopwords;
- rimozione punteggiatura;
- stemming.

Alla fine del processo di preprocessing viene ottenuto un insieme che contiene tutte le parole utilizzate in ogni definizione, per ogni concetto. La similarità delle definizioni per ogni concetto viene quindi calcolata utilizzando le occorrenze delle parole all'interno dell'insieme definizione: una volta calcolate le suddette occorrenze viene preso il valore massimo, (riguardante la parola che si ripete di più) e viene diviso per il numero di parole presenti all'interno dell'insieme, ottenendo così un valore di similarità.

Il risultato ottenuto è il seguente:

	Astratto	**Concreto**
Generico	11%	19%
Specifico	13%	11%

Si può notare che le definizioni riguardanti i termini astratti abbiano meno termini in comune rispetto a quelli concreti. In più la differenza tra astratto e concreto la si vede anche tra concetto generico e specifico, infatti se per il concetto concreto è più facile avere similarità tra definizioni se l'oggetto da definire è generico, nel caso di quelli astratti, è più facile se il concetto da definire rappresenti qualcosa di specifico.