

Relazione esercizio content2Form

L'esercitazione 2 ha affrontato il problema della ricerca onomasiologica. Questo task consiste nell'individuare un concetto a partire dalla sua definizione. Al pari della scrittura delle definizioni, la ricerca onomasiologica rappresenta un problema difficile da affrontare. Visto che l'obiettivo che persegue tale task è sostanzialmente l'opposto della scrittura di definizioni, abbiamo studiato la ricerca onomasiologica introducendo il meccanismo del **Genus – differentia**, con il quale si vuole rappresentare un concetto partendo dal genus che solitamente è un iperonimo del concetto, e individuando una serie di caratteristiche distintive del concetto(differentia) tali da differenziarlo da qualsiasi altro nella gerarchia nella quale esso viene inquadrato. Il meccanismo appena introdotto può essere usato al contrario cercando nella definizione un qualche elemento che possa fungere da genus e che ci rimanda ad una gerarchia di concetti nella quale, si spera, possiamo collocare il concetto da individuare.

Le risorse per questa esercitazione sono le stesse usate nella precedente: un file **"definizioni.csv"** contenente un insieme di definizioni associate a 4 termini (Brick, Person, Revenge, Emotion), aggregati sulle dimensioni di concretezza/specificità. Per individuare il genus di ogni concetto, le definizioni sono state preprocessate rimuovendo spazi multipli, segni di punteggiatura e stopwords, creando così un bag of words. In genere, il genus rappresenta un termine in comune ad un insieme di definizioni, poichè l'individuazione del concetto nella sua tassonomia risulta un modo naturale per costruire definizioni. Partendo da questa osservazione si è ritenuto di individuare il genus nella parola di contenuto che ricorre maggiormente nelle definizioni associate ad un certo concetto.

Il file **"main.py"** contiene l'entry point del programma. Le operazioni appena descritte sono realizzate dalla funzione **"most_frequent"**, dichiarata nel file **"onomasiologic_research.py"**. Recuperate le definizioni e individuato il genus in ognuna, il programma tenta la ricerca del concetto più probabile avvalendosi dei synset di wordNET. Il genus colloca il concetto in una tassonomia in cui egli risulta un iperonimo, ne consegue che il concetto da ricercare è un potenziale iponimo del genus. In genere non si va molto in profondo nella generalizzazione del concetto, nella sua definizione, per cui la ricerca può protrarsi solo sul primo livello di iponimi del genus. La funzione **"best_score"** esegue una disambiguazione del synset più probabile sulla base di un punteggio calcolato come sovrapposizione lessicale dei 2 contesti, da una parte quello del synset e dall'altra quello del concetto. Il contesto del synset è calcolato come in bag of words di tutti i termini nelle definizioni dei possibili synset associati al genus e dei loro iponimi diretti, mentre il contesto del concetto è dato banalmente dal bag of words di tutte le sue definizioni. I synset vengono ordinati per valore di score decrescente ed infine vengono restituiti i primi 5 con il punteggio più alto.

Risultati

Come si può notare dai seguenti risultati, la ricerca onomasiologica risulta essere un task non proprio semplice. Fra i possibili risultati, una certa attenzione merita il concetto person, che con i suoi synset, almeno da un punto di vista del campo semantico restituito, ci avviciniamo leggermente. Per quanto riguarda i concetti di Brick e Revenge, il metodo del termine più frequente si è rivelato fallimentare, dato che ha portato a nominare genus parole come “used” e “someone” che non sono per nulla informative ai fini della costruzione dei loro significati.

```
Onomasiologic research for the concept: Emotion
Genus found: feel
```

```
Synsets found:
- Synset('touch.n.12')
- Synset('cool_off.v.03')
- Synset('die.v.05')
- Synset('glow.v.05')
- Synset('suffer.v.03')
```

```
Onomasiologic research for the concept: Person
Genus found: human
```

```
Synsets found:
- Synset('homo_sapiens.n.01')
- Synset('world.n.08')
- Synset('homo_erectus.n.01')
- Synset('homo_habilis.n.01')
- Synset('rhodesian_man.n.01')
```

```
Onomasiologic research for the concept: Revenge
Genus found: someone
```

```
Synsets found:
- Synset('adult.n.01')
- Synset('worker.n.01')
- Synset('expert.n.01')
- Synset('anomaly.n.02')
- Synset('bad_person.n.01')
```

```
Onomasiologic research for the concept: Brick
Genus found: used
```

```
Synsets found:
- Synset('exploit.v.01')
- Synset('exploit.v.02')
- Synset('put.v.04')
- Synset('address.v.05')
- Synset('cannibalize.v.02')
```