

Costruzione di ConceptNet

In questa sezione andremo a studiare come è stata costruito ConceptNet.

Il processo adottato al fine di popolare la risorsa è automatico e consta fondamentalmente in due fasi:

1. Extraction phase: applicazione di un insieme di extraction rules sulle frasi semi strutturate del corpus OMCS.

2. Relaxation phase: applicazione di una serie di relaxation procedures per ottimizzare la connettività della rete semantica.

Extraction phase

Sono circa cinquanta le extraction rules adottate per mappare le frasi di OCMS in relazioni binarie cablate poi in ConceptNet.

Le frasi poste agli utenti erano semi-strutturate (e.g., The effect of ____ is ____), e questo ha reso la fase di estrazione relativamente semplice.

Es. *The effect of [falling off a bike] is [you get hurt]*

Nelle frasi per le quali non è stato possibile estrarre alcun tipo di relazione viene usata la relazione generica *ConceptuallyRelatedTo*.

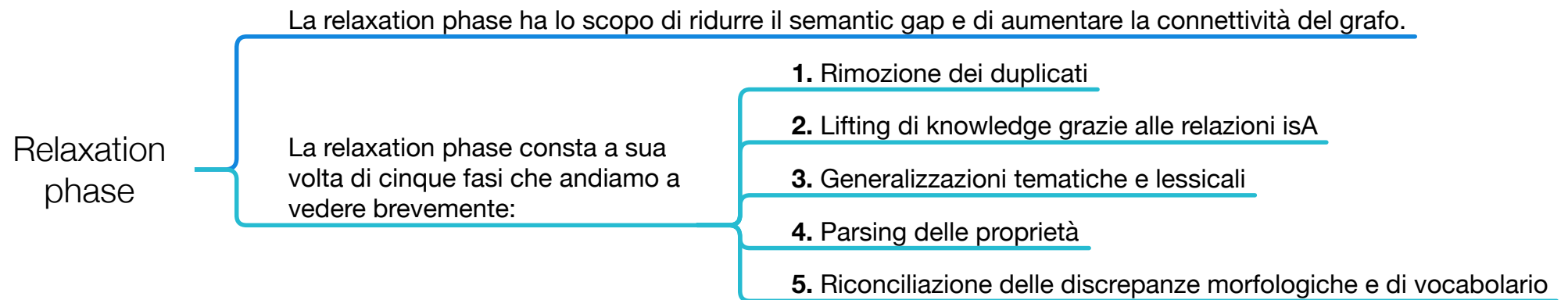
Il risultato di questa fase consiste dunque in una serie di nodi collegati fra loro e strutturalmente omogenei: ogni nodo è un frammento in lingua inglese composto dalla combinazione di quattro costruzioni sintattiche (non tutte necessariamente presenti):

verbi (e.g., buy, not eat, drive)

locuzioni nominali (e.g., red car, laptop computer)

locuzioni preposizionali (e.g., in restaurant, at work)

locuzioni aggettivali (e.g., very sour, red).



Rimozione duplicati

La prima operazione che viene effettuata consiste nella rimozione dei duplicati dato che alcuni fatti comuni vengono tendenzialmente asseriti molteplici volte.

Mentre si compie questa operazione viene computato per ogni relazione il metadato *frequency* che indica quante volte un certo statement è stato asserito (*come avevamo detto nella MM "Storia la seconda versione"*)

Lifting knowledge

Lifting di knowledge grazie alle relazioni isA

Il secondo passo della relaxation phase consiste nel lifting di informazioni mediante l'uso della relazione isA: a partire dai nodi figli le relazioni di isA sono sfruttate (con l'ausilio di euristiche) per far portare conoscenza più in alto nella struttura.

```
[ (IsA 'apple' 'fruit');  
  (IsA 'banana' 'fruit');  
  (IsA 'peach' 'fruit') ]  
  
AND  
  
[ (PropertyOf 'apple' 'sweet');  
  (PropertyOf 'banana' 'sweet');  
  (PropertyOf 'peach' 'sweet') ]  
  
IMPLIES  
  
  (PropertyOf 'fruit' 'sweet')
```

La relazione isA è usata per apprendere che un *fruit* è *sweet* dato che ogni figlio di fruit è sweet (sarà compito dell'euristica stabilire se tale implicazione è sensata o meno)

Generalizzazioni tematiche e lessicali

Il terzo passo della relaxation phase
consta nella generalizzazione
tematica e del lessico:

Ogni nodo (i.e., ossia ogni locuzione inglese) viene
generalizzato in senso tematico e lessicale creando
quindi una serie di nodi “sinonimi”, collegati dal loro
nodo di origine alla relazione *SuperThematicKLine*

(SuperThematicKLine 'buy food' 'buy')

(SuperThematicKLine 'purchase food' 'buy')

Es.

Parsing delle proprietà

Il quarto step consiste nell'andare ad analizzare quei noi che contengono *noun phrases* ed *aggettivi*

La presenza di aggettivi nelle locuzioni ci permette di aggiungere relazioni del tipo *PropertyOf* ampliando così la conoscenza del sistema

```
[(IsA 'apple' 'red round object');  
 (IsA 'apple' 'red fruit')]
```

IMPLIES

```
(PropertyOf 'apple' 'red')
```

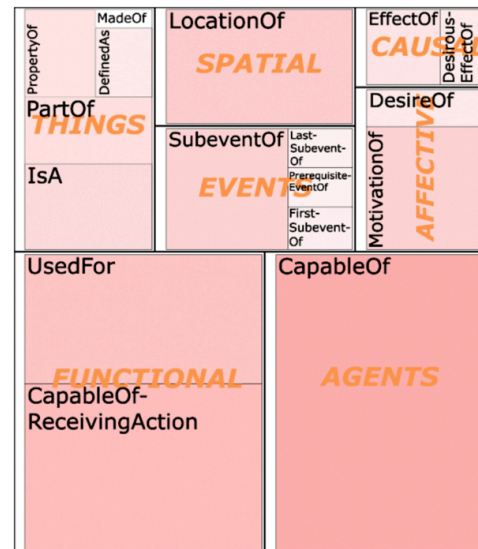
Es.

Dato che nel sistema è presente lo statement *apple isA red fruit*, il concetto *apple* è arricchito dalla relazione *PropertyOf* verso *red*.

Riconciliazione
discrepanze

Riconciliazione delle discrepanze morfologiche e di vocabolario

Al termine della creazione della struttura, ConceptNet conta 1,6 milioni di asserzioni (delle quali 1.25 sono k-lines) in un network semantico di più di 300.000 nodi.



La dimensione dei quadrati indica la quantità di relazioni di quel tipo

Le relazioni semantiche adottate sono 20 e ne vediamo una schematizzazione in Figura

Figure

Relationship	Number of associations	% of associations
RELATEDTO	1,449,431	51.25%
FORMOF	273,560	09.67%
ISA	247,387	08.75%
SYNONYM	237,772	08.41%
HASCONTEXT	177,677	06.28%
DERIVEDFROM	116,243	04.11%
USEDFOR	42,443	01.50%
SIMILARTO	29,480	01.04%
ATLOCATION	28,960	01.02%
CAPABLEOF	26,354	00.93%
HASSUBEVENT	25,896	00.92%
HASPREREQUISITE	23,493	00.83%
ETYMOLOGICALLYRELATEDTO	20,723	00.73%
ANTONYM	19,967	00.71%
CAUSES	17,088	00.60%
HASPROPERTY	13,553	00.48%
PARTOF	12,795	00.45%
MOTIVATEDBYGOAL	9,807	00.35%
RECEIVESACTION	8,383	00.30%
HASA	7,735	00.27%

Distribuzione dei valori all'interno di ConceptNet 5.5.0 (sono mostrate solo le 20 relazioni più importanti)

Relation	Sentence pattern
LocatedNear	You are likely to find <i>NP</i> near <i>NP</i> .
DefinedAs	<i>NP</i> is defined as <i>NP</i> .
SymbolOf	<i>NP</i> represents <i>NP</i> .
ReceivesAction	<i>NP</i> can be <i>VP</i> .
HasPrerequisite	<i>NP VP</i> requires <i>NP VP</i> .
MotivatedByGoal	You would <i>VP</i> because you want <i>VP</i> .
CausesDesire	<i>NP</i> would make you want to <i>VP</i> .
MadeOf	<i>NP</i> is made of <i>NP</i> .
HasSubevent	One of the things you do when you <i>VP</i> is <i>NP VP</i> .
HasLastSubevent	The last thing you do when you <i>VP</i> is <i>NP VP</i> .
IsA	<i>NP</i> is a kind of <i>NP</i> .
UsedFor	<i>NP</i> is used for <i>VP</i> .
HasA	<i>NP</i> has <i>NP</i> .
CapableOf	<i>NP</i> can <i>VP</i> .
Desires	<i>NP</i> wants to <i>VP</i> .
CreatedBy	You make <i>NP</i> by <i>VP</i> .
PartOf	<i>NP</i> is part of <i>NP</i> .
Causes	The effect of <i>VP</i> is <i>NP VP</i> .
HasFirstSubevent	The first thing you do when you <i>VP</i> is <i>NP VP</i> .
AtLocation	Somewhere <i>NP</i> can be is <i>NP</i> .
HasProperty	<i>NP</i> is <i>AP</i> .

K-LINES (1.25 million assertions) (ConceptuallyRelatedTo 'bad breath' 'mint' 'f=4;i=0;') (ThematicKLine 'wedding dress' 'veil' 'f=9;i=0;') (SuperThematicKLine 'western civilisation' 'civilisation' 'f=0;i=12;')
THINGS (52 000 assertions) (IsA 'horse' 'mammal' 'f=17;i=3;') (PropertyOf 'fire' 'dangerous' 'f=17;i=1;') (PartOf 'butterfly' 'wing' 'f=5;i=1;') (MadeOf 'bacon' 'pig' 'f=3;i=0;') (DefinedAs 'meat' 'flesh of animal' 'f=2;i=1;')
AGENTS (104 000 assertions) (CapableOf 'dentist' 'pull tooth' 'f=4;i=0;')
EVENTS (38 000 assertions) (PrerequisiteEventOf 'read letter' 'open envelope' 'f=2;i=0;') (FirstSubeventOf 'start fire' 'light match' 'f=2;i=3;') (SubeventOf 'play sport' 'score goal' 'f=2;i=0;') (LastSubeventOf 'attend classical concert' 'applaud' 'f=2;i=1;')
SPATIAL (36 000 assertions) (LocationOf 'army' 'in war' 'f=3;i=0;')
CAUSAL (17 000 assertions) (EffectOf 'view video' 'entertainment' 'f=2;i=0;') (DesirousEffectOf 'sweat' 'take shower' 'f=3;i=1;')
FUNCTIONAL (115 000 assertions) (UsedFor 'fireplace' 'burn wood' 'f=1;i=2;') (CapableOfReceivingAction 'drink' 'serve' 'f=0;i=14;')
AFFECTIVE (34 000 assertions) (MotivationOf 'play game' 'compete' 'f=3;i=0;') (DesireOf 'person' 'not be depressed' 'f=2;i=0;')

41

Conteggio dei nodi di ConceptNet divisi per tipologia semantica (in grassetto). Sotto ogni tipologia c'è qualche esempio correlato degli indici f ed i che rappresentano rispettivamente il numero di volte che il fatto è stato asserito in OMCS ed il numero di volte che è invece stato inferito da ConceptNet stesso.

Figure In Figura possiamo osservare qualche dato sui nodi