

ConceptNet

Il purpose di ConceptNet è diverso da quello degli altri strumenti che abbiamo già visto: “ConceptNet is a semantic network containing lots of things computers should know about the world, especially when understanding text written by people”.

ConceptNet è un database che memorizza e mette in relazione phrases al fine di rappresentare, per quanto possibile, brandelli di senso comune.

La **common sense knowleadge** è dunque la conoscenza del senso comune, normalmente interpretato dagli umani come “buon senso”.

Lo scopo di ConceptNet è quello di cablare informazioni come “to open a door, you must usually first turn the doorknob” o come “if you forget someone’s birthday, they may be unhappy with you”: il task non è semplice, staremo a vedere cosa si sono inventati i creatori di questa risorsa.

Cos'è common sense

È difficile dire formalmente cosa sia il common sense. Certamente possiamo affermare che il senso comune si applica agli aspetti spaziali, fisici, sociali, temporali e psicologici della vita di tutti i giorni.

La difficoltà nel definire il senso comune sta proprio nel fatto che non è mai espresso esplicitamente, ma si dà per scontato essere presente in ognuno di noi.

Secondo Minsky, il senso comune è composto di 30-60 milioni di informazioni che vengono continuamente utilizzate per porre analogie con ciò che ci accade intorno.

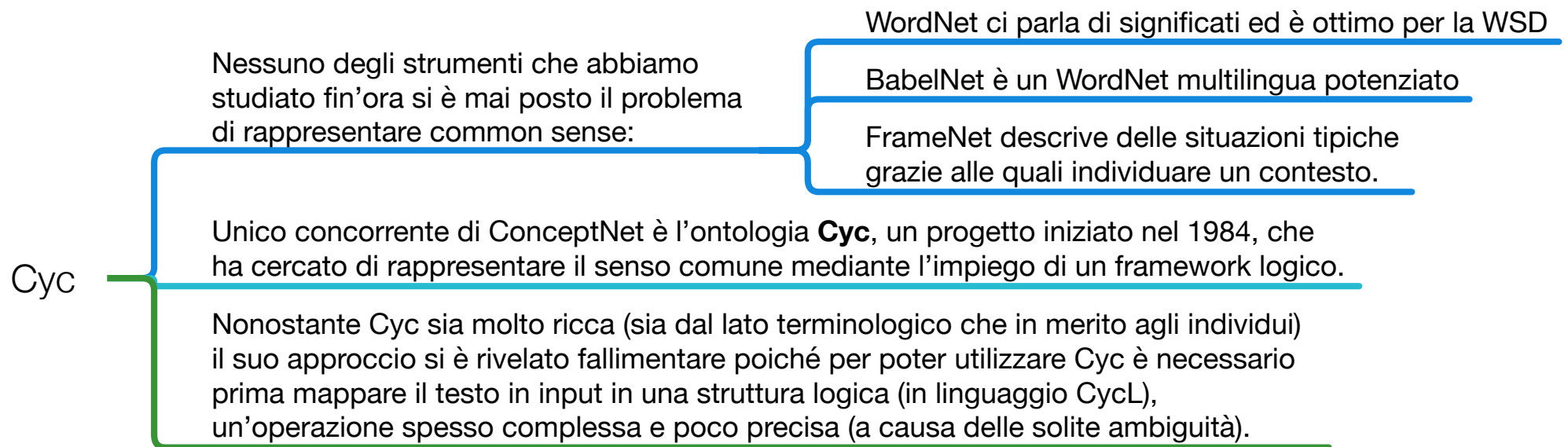
Per queste ragioni per comprendere appieno un testo abbiamo bisogno di una grande dose di commonsense (*che ogni persona possiede*)

Lo scopo di ConceptNet è di trovare un modo per fornire questo commonsense alle macchine

Es. se consideriamo il concetto “acqua” la conoscenza di commonsense legata a questo concetto è che:

L'acqua è tipicamente in uno stato liquido che è incolore, inodore, insapore

Quando le persone comunicano tra loro fanno riferimento ad un grande insieme di conoscenza comune di background che permette una comunicazione veloce ed efficace



La struttura interna di ConceptNet non è altro che un grafo costituito da 1,6 milioni di archi connettenti più di 300.000 nodi.

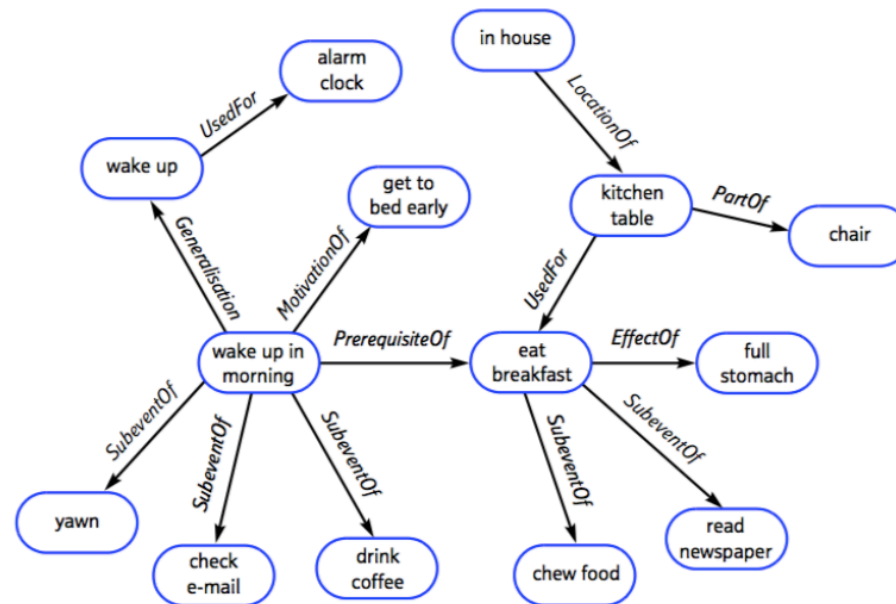
I nodi sono frammenti in lingua inglese

Gli archi dispongono di un tipo (una fra le venti relazioni semantiche possibili che un arco può assumere).

Nota:

I nodi di ConceptNet sono anche chiamati concetti; una terminologia pericolosa considerando il fatto che si tratta di frammenti di testo e dunque di qualcosa di molto diverso rispetto ai concetti che abbiamo trattato finora.

ConceptNet
caratteristiche
fondamentali



Piccolo estratto di ConceptNet

Concetti composti

Abbiamo già detto che i nodi di ConceptNet sono stringhe di testo (i.e., concetti composti), tendenzialmente costituite da un verbo di azione corredato di uno o due argomenti indiretti (e.g., buy food, drive to store).

Questo tipo di rappresentazione è evidentemente molto agevole per rappresentare un'infinità di concetti del mondo reale, ma in un certo senso ci riporta alle primissime pagine di questo volume (un salto bello lungo!): siamo tornati alle stringhe, perdendo così la semantica delle parole (il significato). **ConceptNet non distingue fra i significati delle parole.**

20 relazioni semantiche

ConceptNet estende il repertorio di WordNet delle relazioni: dalle sole sinonimie, antonimie, meronimie, e iponimie si ottiene un ricco insieme di venti relazioni semantiche fra cui:

EffectOf(causality), SubeventOf(event hierarchy), CapableOf(agent's ability), PropertyOf, LocationOf e MotivationOf(affect)

Nota:

Molte applicazioni che ragionano sul contesto al massimo richiedono di distinguere le relazioni tra:

relazioni affettive

relazioni temporali

relazioni spaziali

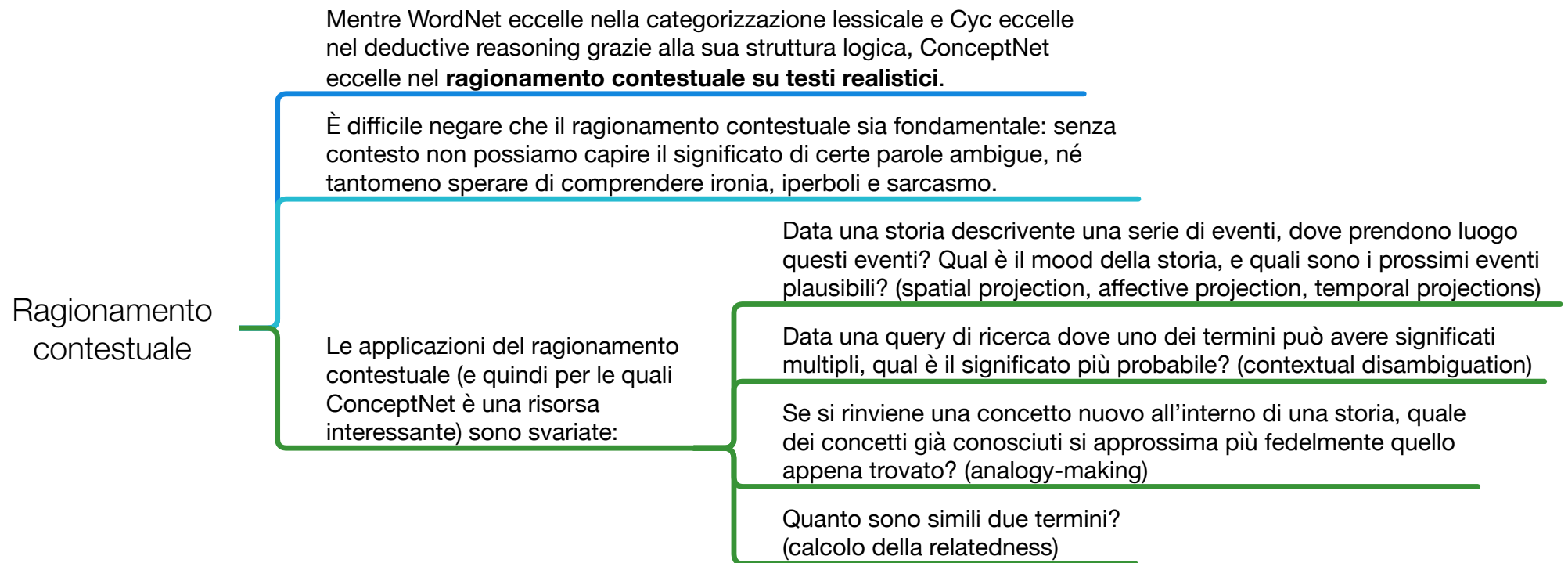
Conoscenza rivedibile

Rispetto a WordNet la conoscenza di ConceptNet è più informale, ritrattabile e naturalmente valutabile.

Es. la tassonomia formale di WordNet ci permette di rappresentare che dog isA canine, che a sua volta isA carnivore che a sua volta isA mammal, mentre non è possibile rappresentare la pratica e orientata associazione member-to-set dog isA pet (i.e., animale domestico).

ConceptNet possiede molta conoscenza rivedibile, ossia è possibile esprimere qualcosa di vero in molti casi ma non sempre e.g., EffectOf("fall off bicycle", "get hurt").

D'altronde le eccezioni ed i casi particolari accadono nella vita di tutti i giorni, dunque è importante avere questa sorta di ragionamento per default.



Ragionamento
contestuale
pt2

In merito al secondo punto dobbiamo fare una precisazione: giacché i nodi di ConceptNet sono semplicemente stringhe, effettuare WSD è tendenzialmente impossibile giacché i significati non sono cablati all'interno di ConceptNet (non vi sono sense-id); inoltre gli archi uscenti da un nodo sono eterogenei rispetto al significato del nodo stesso, ossia, archi per diversi significati del frammento di testo del nodo sono tutti trattati allo stesso modo.

Ciò che è possibile fare mediante ConceptNet, però, è andare ad esplorare le adiacenze di un certo nodo che matcha con parte di un testo fornito in input e sfruttare le relazioni trovate per dedurre informazioni in merito alla query. I frammenti di testo hanno tendenzialmente senso univoco e dunque non esiste alcun problema di ambiguità.

Es. *play guitar* non è ambiguo rispetto a *play tennis*, mentre soltanto *play* lo sarebbe. Pertanto il problema di ambiguità riguarda tendenzialmente termini singoli e non locuzioni.

Se presente, l'ambiguità riguarda i termini collegati da relazioni e non le relazioni stesse, che sono invece salde seppure non dotate di un id.

Es. se abbiamo una proprietà il cui significato è noto, e.g., *CapableOf*, possiamo conoscere la relazione che intercorre fra coppie come (lion, hunt zebra), (football player, pass ball) o (airplane, fly). Queste relazioni possono a loro volta essere utilizzate per estrarre ulteriori informazioni (posto che lion, hunt zebra, ecc., non siano polisemici: anche in quel caso, però, si potrebbe tentare di filtrare i risultati di ConceptNet utilizzando l'overlap con glosse e descrizioni di WN).

Insomma, ConceptNet è impreciso dal punto di vista semantico, ma la verità è che che nessuno strumento o tecnica è sufficiente a domare la complessità del linguaggio naturale, ma tutti possono essere utili se usati con le dovute accortezze.

Associazioni e rappresentazione nel NL

Le ragioni per cui ConceptNet è adatto agli usi che abbiamo appena elencato sono i suoi sforzi nella conoscenza associazionale e l'uso di una rappresentazione diretta in linguaggio naturale.

Del milione e mezzo di fatti che interlacciano i concetti in ConceptNet, approssimativamente un milione e un quarto (i.e., 1.25m) sono dedicati all'esplicitazione di connessioni generiche fra concetti (e.g., ConceptuallyRelatedTo, ThematicKLine etc.).

Questo tipo di conoscenza viene detta **k-lines** da Minsky, il quale sostiene che associazioni di questo genere siano i meccanismi primari adottati dagli umani per processare i contesti e sfruttare la memoria.

È infatti questo tipo di conoscenza a migliorare la connettività e la copertura della rete semantica implementata da ConceptNet, rendendo concetti esterni più probabilmente rimappabili su qualche nodo del sistema.

A differenza dei simboli logici (adottati ad esempio da Cyc), parole diverse possono disporre di significati uguali e giacché i nodi di ConceptNet sono frammenti testuali, è possibile generare delle gerarchie lessicali alla WordNet per rendere i significati di un nodo più flessibili (e quindi aumentare la possibilità di matchare dall'esterno un nodo).

Es. i nodi *buy food* e *purchase groceries* possono essere riconciliati riconoscendo che buy e purchase sono in qualche modo sinonimi e che groceries è un'istanza di food.

Storia Il corpus base

Mentre WordNet e Cyc derivano principalmente da un lavoro manuale, ConceptNet è generato automaticamente da frasi Inglesi derivate dal corpus OMCS (i.e., Open Mind Common Sens)

OMCS:

Grazie al contributo 14.000 persone che hanno inserito frasi come (e.g., The effect of eating food is...; A knife is used for...) sono state collezionate nel corpus OMCS più di 700.000 frasi inglesi di commonsense

L'**OMCS** è un progetto collaborativo aperto nel 2000 che si rivolge al pubblico al fine di collezionare frasi di senso comune. Grazie ai più di 14.000 contributors, OCMS è arrivata a contare più di 700.000 frasi di common sense in soli tre anni. Le frasi sono state poste ai contributors secondo una modalità *fill-the-blank*, e.g., The effect of ____ is _____. Chiaramente le frasi da compilare che sono state fornite erano relative a molti domini diversi, così da ampliare la copertura di ConceptNet.

Es. di frasi:

The effect of eating food is...; A knife is used for...

Storia la seconda versione

Nella versione 2 di ConceptNet sono stati aggiunti dei sistemi di weighting della conoscenza. Ogni asserzione binaria è stata valutata basandosi su:

Quante volte è stata affermata in OCMS.

Quanto è facile inferirla direttamente da altri fatti in ConceptNet.

Sempre nella seconda versione sono stati aggiunte delle *extraction rules* che permettono di estrarre ulteriore conoscenza a partire dalle frasi di OCMS. Affronteremo in dettaglio l'argomento fra qualche paragrafo, intanto facciamoci venire l'acquolina in bocca con questo esempio:

Es. data la frase *A lime is a sour (i.e., apro) fruit* possiamo ovviamente estrarre l'informazione basilare `IsA(lime, fruit)` ma grazie all'uso di *extraction rules* più sofisticate possiamo anche inferire `PropertyOf(lime, sour)`.

Anche le *generalizzazioni* vengono inferite:

Es. se la maggior parte dei frutti hanno la proprietà `sweet` allora tale proprietà viene liftata (i.e., passata) alla classe padre come: `PropertyOf(fruit, sweet)`.

Nota:

Ad oggi, ConceptNet è anche una risorsa multilingua.