Introduction to WordNet, HowNet, FrameNet and ConceptNet

Zi Lin

the Department of Chinese Language and Literature

August 31, 2017

WordNet

Begun in 1985 at Princeton University, WordNet is a semantic dictionary that was designed as a network, partly because representing words and concepts as an interrelated system seems to be consistent with evidence for the way speakers organize their mental lexicons. (G. A. Miller)

Construction
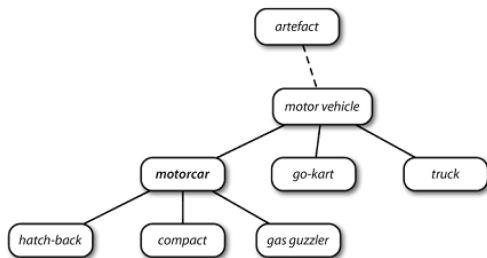
Unit - Synset

WordNet＇s design resembles that of a thesaurus in that its building block is a **synset** consisting of all the words that express a given concept. And WordNet contains almost 80,000 noun word forms organized into some 60,000 lexicalized concepts.

Relations - hyponymy, meronymy, and entailment

WordNet does much more than listing concepts in the form of synsets. The synsets are linked by means of a number of relations, including **hyponymy**, **meronymy**, and **entailment**.

Hyponymy

Organism: a living entity.

↑

Animal: an **organism** capable of voluntary movement and possessing sense organs and cells with noncellulose walls.

↑

Bird: a warm-blooded egg-laying **animal** having feathers and forelimbs modified as wings.

↑

Robin: a migratory **bird** that has a clear melodious song and a reddish breast with gray or black upper plumage.

Structures

{**robin**, redbreast}@→{**bird**}@→{**animal**, animal_being}@→{**organism**, life_form, living_thing}

Other Relations

```
>>> from nltk.corpus import wordnet as wn
>>> wn.synset('tree.n.01').part_meronyms()
[Synset('burl.n.02'), Synset('crown.n.07'), Synset('limb.n.02'), Synse
t('stump.n.01'), Synset('trunk.n.01')]
>>> wn.synset('eat.v.01').entailments()
[Synset('chew.v.01'), Synset('swallow.v.01')]
>>>
```

The "tennis problem"

Connecting racquet, ball, net and court game, or physician and hospital, is a challenge for electronic dictionaries that Roger Chaffin has called the "**tennis problem**."

WordNet cannot solve tennis problem

WordNet focuses on the **semantics of words and concepts** rather than on **semantics at the text or discourse level**, So WordNet contains no relations that indicate the words' shared membership in a topic of discourse.

Componential lexical semantics

The analysis of a words through structured sets of semantic features, or analyzing a word's meaning into semantic components.

Male = +[human] +[male] +[adult]; Female = +[human] -[male] +[adult]
Boy = +[human] +[male] -[adult]; Girl = +[human] -[male] -[adult]

WordNet does not apply componential semantics

It is assumed that the user already has the concept, and that meaning can be represented by any symbols that makes it possible to distinguish among them. Thus they concluded that it (componential semantics) was not the best theory for natural language processing.

HowNet

HowNet is an on-line common-sense knowledge base unveiling inter-conceptual relations and inter-attribute relations of concepts as connoting in lexicons of the Chinese and their English equivalents. (Zhendong Dong and Qiang Dong)
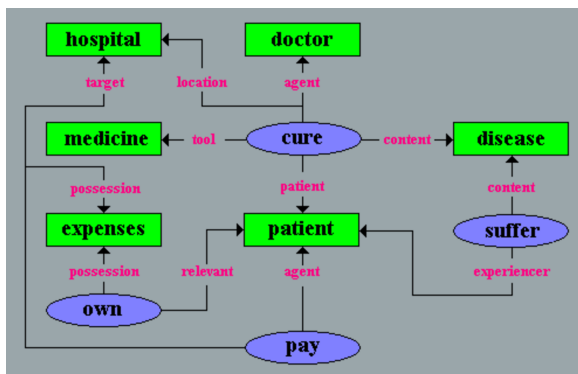
Sememe

A **sememe** refers to **the smallest basic semantic unit that cannot be reduced further**. We hypothesise that all concepts can be reduced to the relevant sememes. We deem further that there exist a close set of sememes, from which, composes an open set of concepts. If we can manage the close set of sememes to describe inter-concept relations as well as inter-attribute relations, an ideal knowledge base would be conceivable.

How to Find Sememe

- The Chinese characters (including simple word) is a close set that can be exploited to express both simple and complex concepts, as well as the inter-concept and inter-attribute connections.
- The set of sememe is established on meticulous examination of about **6000 Chinese characters**. Finally, the process arrived at a set of around **2,800 sememes** we are now using in Hownet.

Structure of Sememes

```
entity|实体
  thing|万物 [#time|时间,#space|空间]
    physical|物质 [!appearance|外观]
      animate|生物 [*alive|活着,!age|年龄,*die|死,*metabolize|代谢]
        AnimalHuman|动物 [!sex|性别,*AlterLocation|变空间位置,*StateMental|精神状态]
          human|人 [!name|姓名,!wisdom|智慧,!ability|能力,!occupation|职位,*act|行动]
            humanized|拟人 [fake|伪]
          animal|兽 [^*GetKnowledge|认知]
            beast|走兽 [^*GetKnowledge|认知]
            livestock|牲畜 [$foster|饲养,~$consume|摄取,~?edible|食物]
            bird|禽 [*fly|飞,~$consume|摄取,~?edible|食物]
            InsectWorm|虫 [~undesired|莠]
            fish|鱼 [*swim|游,#waters|水域,~$consume|摄取,~?edible|食物]
      plant|植物 [^*SelfMove|自移,^*StateMental|精神状态]
        crop|庄稼 [$planting|栽植,?material|材料,#edible|食物]
        tree|树 [#wood|木,?material|材料]
```

Representation of Concepts

- 打鼾: {MakeSound| 发声: cause = {ill| 病态}, time = {sleep| 睡}}
- 浴衣: {clothing| 衣物: {PutOn| 穿戴: instrument = { ~ }, location = {part| 部件: PartPosition = {body| 身}, whole = {human| 人}}, TimeAfter = {swim| 游} {wash| 洗涤: PartOfTouch = {part| 部件:PartPosition = {body| 身}, whole = {human| 人}}}}}

FrameNet

The FrameNet Project is building a lexical database of English that is both human - and machine-readable, based on annotating examples of how words are used in **actual texts**.
(https://framenet.icsi.berkeley.edu)

Semantic Frame

Basic Idea: the meanings of most words can best be understood on the basis of a **semantic frame**, a description of a type of event, relation, or entity and the participants in it.

Example: Cooking

- the person who doing the cook (Cook)
- the food that is to be cooked (Food)
- something to hold the food while cooking (Container)
- a source of heat (Heating_instrument)

Semantic Frame

FrameNet is mainly composed of three parts - **Definition**, **FEs** and **Frame-frame Relations**.

Example: **Hit_target**

Definition:

An Agent causes the ballistic motion of a projectile and its subsequent impact on a Target.
It looks like he HIT the bull's-eye with his first arrow!

If he SHOT her, he is fully culpable whether he meant to or not.

FEs:

Core:

Agent []   The individual that directs the Projectile such that it impacts the Target.
Semantic Type: Sentient
Target []   The entity which is hit by a Projectile directed by the Agent.

Non-Core:

[Instrument [ins]]   The entity, most often ammunition or a firearm, which the Agent manipulates to hit the Target.
Semantic Type:
Physical_entity
Manner [man]   Any description of the intentional act which is not covered by more specific FEs, including secondary
Semantic Type: Manner   effects (quietly, loudly), and general descriptions comparing events (the same way). In addition, it may
                                      indicate salient characteristics of an Agent that also affect the action (presumptuously, coldly, deliberately,
                                      eagerly, carefully).

Example: **Hit_target**

Means [mea]
Semantic Type: State_of_affairs
The action performed by the Agent that accomplishes hitting the Target.

Place [pla]
Semantic Type: Locative_relation
The location where the event happens.

Purpose [pur]
Semantic Type: State_of_affairs
The state-of-affairs that the Agent hopes to bring about by hitting the Target

Subregion [sub]
Semantic Type: Locative_relation
The part of the Target hit by the Agent.
He SHOT me in the arm!

Time [tim]
Semantic Type: Time
The time at which the Agent hits the Target.

Frame-frame Relations:

Inherits from: Intentionally_affect
Is Inherited by:
Perspective on:
Is Perspectivized in:
Uses: Cause_impact, Hit_or_miss, Shoot_projectiles, Shooting_scenario
Is Used by:
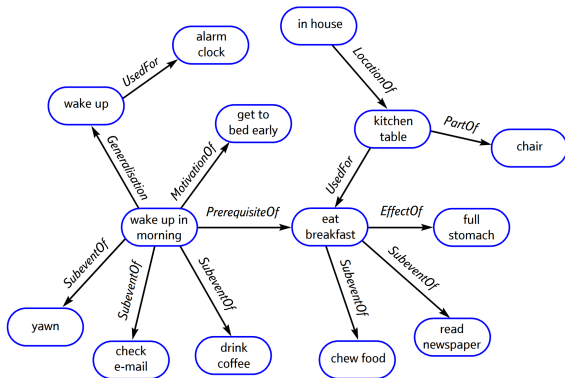Subframe of:
Has Subframe(s):
Precedes:
Is Preceded by:
Is Inchoative of:
Is Causative of:
See also:

ConceptNet

ConceptNet is a freely-available large-scale **commonsense knowledge base** with an integrated NLP tool-kit that supports many practical textual-reasoning tasks over real-world documents.

Commonsense Knowledge

Of the different sorts of semantic knowledge that are researched, arguably the most general and widely applicable kind is knowledge about the everyday world that is possessed by all people.

Example

- A lemon is sour.
- To open a door, you must usually first turn the doorknob.
- If you forget someone's birthday, they maybe unhappy with you.

Construction

- ConceptNet is generated automatically from the English sentences of the Open Mind Common Sense (OMCS) corpus.
- OMCS also turns to the general public for help. Over 14,000 web contributors who logged in entered sentences in a fill-in-the-blank fashion, we amassed over 700,000 English sentences of commonsense.
- By applying NLP and extraction rules to the semi-structured OMCS sentences, **300,000 concepts** and **1.6 million binary-relational assertions** are extracted to form ConceptNet＇s semantic network knowledge base.
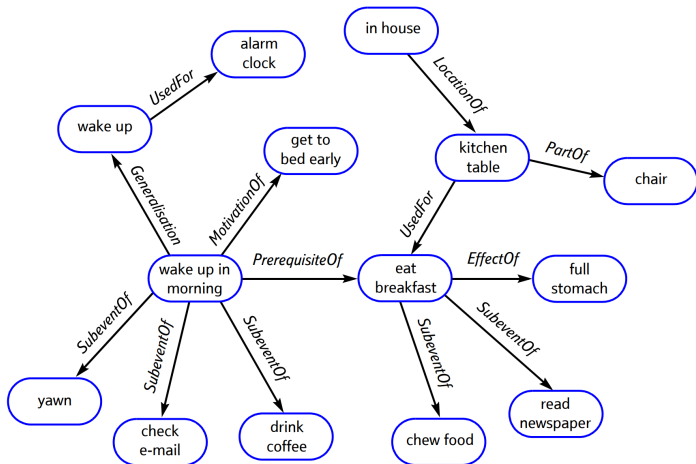
Difference between ConceptNet and WordNet

- WordNet is optimized for **lexical categorization** and **word-similarity determination**, while ConceptNet is optimized for **making practical context-based inferences over real-world texts**.
- ConceptNet extends WordNet's notion of a node in the semantic network from purely lexical items (words and simple phrases with atomic meaning) to include **higher-order compound concepts**, which compose an action verb with one or two direct or indirect arguments (e.g. 'buy food', 'drive to store').

Difference between ConceptNet and WordNet

- ConceptNet extends semantic relations from the triplet of synonym, is-a, and part-of, to a present repertoire of **twenty semantic relations**.

example

- Location Of (A,B): **Books** are in the **library**.
- Used For (A,B): Forks are used for eating.
- Subevent Of (A,B): After **waking up in morning**, he **checked his email**.

References

Fellbaum Christiane. (1998).
WordNet: An Electronic Lexical Database.
Cambridge, Mass: MIT Press.

Zhendong Dong, Qiang Dong, Changling Hao. (2007).
Theoretical Findings of HowNet.
Journal of Chinese Information Processing.

Baker F Collin, Charles J Fillmore, John B. Lowe. (1998).
The Berkeley FrameNet Project.
Proceedings of the 17th International Conference on Computational Linguistics

Liu H, Singh P. (2004).
ConceptNet - a practical commonsense reasoning toolkit.
BT Technology Journal

YUAN Yulin, LI Qiang. (2014).
How to Use Qualia Structure to Solve "Tennis Problem"?
Journal of Chinese Information Processing.

Thank You