

4. TAXONOMY INDUCTION

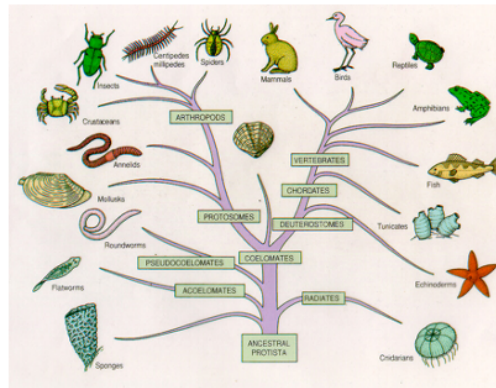
Taxonomy Induction

Information extraction

- Extract isolated facts from documents, e.g., *lion ISA animal*

Taxonomy induction

- Extract related facts from documents, e.g., classification of animals



Information extraction concerns the extraction of isolated facts, such as ISA relationships. Taxonomy induction aims at extracting related facts and organizing them in a structured knowledge basis, e.g. a hierarchical taxonomy. It is a special case of the more general ontology induction, which organizes knowledge using arbitrary relationships.

Use of Taxonomies

Hyponyms (subordinate terms) can inherit properties from hypernyms (more general terms)

- Due to transitivity of ISA, no need to learn inferred facts

No unique taxonomies

- Depending on the perspective and application different taxonomies may be useful:

A tiger and a puppy are both Mammals and hence belong close together in a typical taxonomy, but tiger is a WildAnimal (in the perspective of Animal-Function) and a JungleDweller (in the perspective of Habitat), while a puppy is a Pet (as function) and a HouseAnimal (as habitat), which would place them relatively far from one another

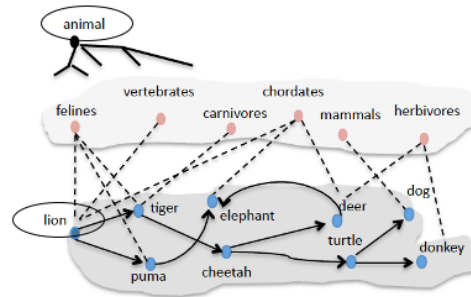
One of the advantages of taxonomy induction (and more generally ontology induction) is the possibility to perform inferences on the extracted knowledge. For example, in a ISA hierarchy, lower nodes in the hierarchy can inherit properties from higher nodes, thus these extra facts need not to be learnt separately.

One of the challenges in taxonomy induction is the fact that there is no notion of “correct” taxonomy. A taxonomy strongly depends in its intended use and on the specific perspective of the user on the domain. Integrating all possible perspectives into one single global taxonomy would not be feasible and useful, as the resulting taxonomy would potentially be too complex to be used.

Taxonomy Induction Task

Starting from a root concept and a basic concept

1. Learn relevant terms and their hypernym / hyponym relationships
2. Filter out erroneous terms and relations
3. Induce a taxonomy structure



©2019, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

Information Extraction- 4

We present in the following one specific approach to taxonomy induction, which was one of the first approaches proposed in the field. It starts from the assumption that one general concept (e.g. animals) and at least one basic concept of the taxonomy (e.g. lion) are provided as input. From this starting point the task is to identify more relevant concepts (represented as terms), establish the hypernym and hyponym relationships, filter out erroneous terms and relations and finally induce an overall taxonomy structure. The figure illustrates the process: at the bottom level first more related terms are identified. Then intermediate concepts, more abstract than the basic concepts are identified. From this data finally the taxonomy will be induced.

Learning Terms

Template approach

- Given a **root concept c** (e.g. animal) and a **seed s** (e.g. lion)
- Hyponym pattern: $P_i(c, s, X) = c$ such as s and X
- Hypernym pattern: $P_c(t_1, t_2, X) = X$ such as t_1 and t_2

The basic idea is to learn terms and relationships by querying a Web search engine. As patterns language template capturing hypernym relationships are used. This is analogous to the approach with Hearst patterns for extracting ISA relationships. However, as in a first phase one has to detect more relevant terms, so-called double anchored patterns are used, that relate one (known) term to another (unknown) term of the same class of concepts. One example of such a pattern would be “c such as s and X”, where c and s would be known and from the result the term at position X would be a new term from the same class of concepts as c. Using such patterns new terms can be harvested by recursively applying the pattern to the terms known so far.

Finding Hyponyms

Recursively harvest new terms using a Web search engine

```
T = {s}; w(t) = 0
while T changes
  for all t in T: submit Pi(c, t, X) to search engine
    add to T all new terms tnew found in position X in a result
  w(t)++
```

Example

"animal such as lion and"

All Images Videos News Shopping More Settings Tools

6 results (0,56 seconds)

General Driving Tips - Safe Overlanding Tips | Avis Safari Rental
<https://www.avis.co.za/safari-rental/driving-tips/general-driving-tips> ▼
Undertaking repairs or doing vehicle extraction at night increases the risk exposure to opportunistic dangerous wild animal such as lion and hyena. The roads in ...

Digication e-Portfolio :: Natali Coronado Malena ePortfolio :: Child ...
https://hostos.digication.com/natali_coronado_malena_eporfolio/Child_Case_Study
He loves vegetables and his favorite vegetable is Zucchini, his favorite sport is basketball and he love the loud and fast animal such as Lion and Tiger. He loves ...

PHS 6 SCIENCE Mr.Gary
phs6ta1.blogspot.com/ ▼
Feb 22, 2017 - Predation is animal that is hunting other prey or animal such as lion and tiger . In prey such as zebra and pig.Example of predation is lion and ...

Download PDF - Springer Link
link.springer.com/content/pdf/10.1007%2Fs10739-007-9147-3.pdf
problem animal (such as lion and elephant) management.56 By 1945. Bigalke was insisting that "game preservation is the task of the scientist, i.e., the work of ...

Existing Term: lion

New Terms: hyena, tiger, elephant

This is an example of how terms related to an initial term “lion” would be found using double-anchored patterns.

Finding Hypernyms

$C = \{c\}$

for all t_1, t_2 in T with $w(t_i) > 0$:

 submit $P_c(t_1, t_2, X)$ to search engine

 add new term h found in position X to C

 add t_1 *ISA* h and t_2 *ISA* h to the hypernym relations H

$w(t_1, t_2, h)++$



Filtering

- rank concepts h by $\sum_{t_1 t_2} w(t_1, t_2, h)$
- keep top concepts

Once the search for terms is completed, hypernyms can be searched by using the same patterns, but now be putting a variable X in the place of the concept, and using the basic terms found so far. In this search only terms that have been leading to the discovery of other basic concepts are being used, which is expressed by the condition $w(t) > 0$.

Once the search for higher level concepts is completed, a filtering step is performed. The concepts found are ranked by the number of times they have been discovered starting from different term pairs and only the highest ranked concepts are retained.

Example

"such as lion and tiger"  

[All](#) [Images](#) [Videos](#) [News](#) [Shopping](#) [More](#) [Settings](#) [Tools](#)

About 10.600 results (0,30 seconds)

Language at the Speed of Sight: How We Read, Why So Many Can't, an...
<https://books.google.ch/books?isbn=0465080650>
Mark Seidenberg - 2017 - Science
Having accrued statistics about words such as LION and TIGER allows the listener (or, later, the reader) to infer much about the meaning of a new word such as ...

Were early hominids REALLY all that threatened by sabre toothed ...
<https://www.thenakedscientists.com/forum/index.php?topic=44769.0> ▼
Jul 16, 2012 - 3 posts - 3 authors
It is likely that early humans gave major predators such as lion and tiger ancestors a wide berth. It is just too great of a risk to ourselves to hunt ...

Were early hominids REALLY all that threatened by sabre toothed ...
<https://www.thenakedscientists.com/forum/index.php?topic=44769.0> ▼
Jul 16, 2012 - 3 posts - 3 authors
It is likely that early humans gave major predators such as lion and tiger ancestors a wide berth. It is just too great of a risk to ourselves to hunt ...

12 Animals With The Strangest Habit Of Sleeping - INVORMA
invorma.com > Family ▼
Oct 15, 2015 - This unusual habit of sleeping is also applied for the family of big cats such as lion and tiger. These cats are also popular as nocturnal hunters.

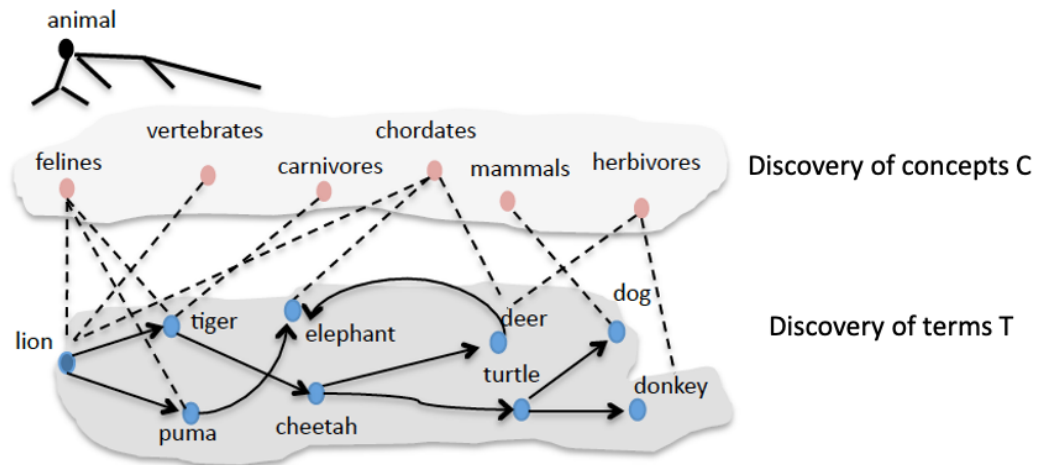
©2019, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

New Classes: predators, big cats,

But also: words, ...

Here a few examples of how higher level concepts related to the basic concepts “lion” and “tiger” can be found.

Example Result



©2019, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

Information Extraction- 10

As a result of the two previous step we obtain basic data on terms for basic concepts T and higher level concepts C, together with hypernym relationships indicated by dotted arrows.

Inducing Hypernym Graph

Many possible relationships among concepts and terms have likely not been discovered

For each pair t_1, t_2 in $T \cup C$

Construct query $q_1 = h(t_1, t_2)$ and $q_2 = h(t_2, t_1)$
with Hearst pattern $h(X, Y)$, e.g., $h(X, Y) = "X \text{ such as } Y"$

Submit query to search engine and count number of results

If $\#results(q_1) > \#results(q_2)$
then add $t_1 \text{ ISA } t_2$ to H
else add $t_2 \text{ ISA } t_1$ to H

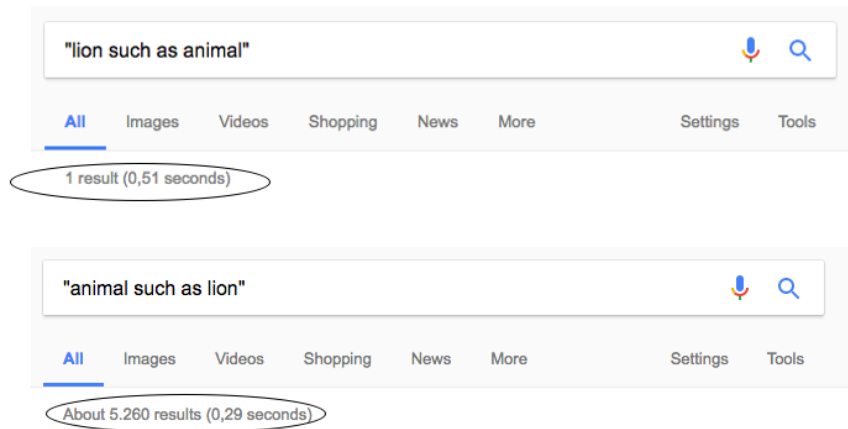
Result: A directed hypernym graph H

©2019, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

Information Extraction- 11

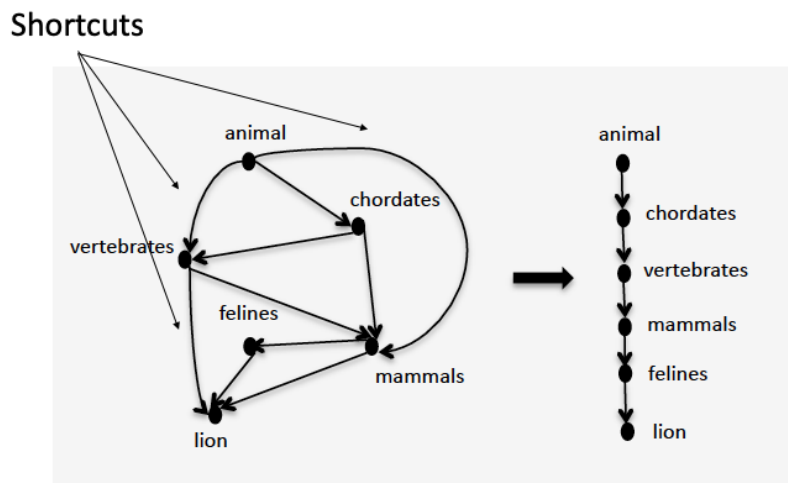
In the steps performed so far, many possible relationships among higher level concepts and basic concepts and higher level concepts may not have been discovered. For discovering those again queries against a search engine are performed, testing the two possible directions of a hypernym relationship for each pair of terms. The alternative that produces more results in the search is considered as being the correct one. As templates we can use any of the Hearst patterns, such as "X such as Y", "X are Y that", "X including Y", "X like Y", "such X as Y" etc

Example



This example shows how powerful this method to test a direction of a relationship is.

Example



Graph H: may contain redundant paths

©2019, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

Information Extraction- 13

After adding all possible relationships to the set of concepts, we obtain a graph that contains many redundant paths, including many of the induced transitive relationships. The method would, for example, find that animal ISA chordate, and chordate is a vertebrate, but also that animal is a vertebrate. In the last step such redundant relationships are removed.

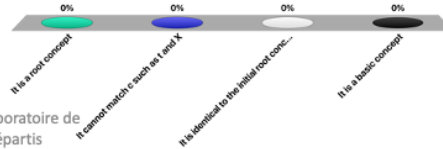
Cleaning the Hypernym Graph

1. Determine all **basic concepts**
 - Not hypernym of another concept
2. Determine all **root concepts**
 - Have no hypernyms
3. For each basic concepts - root concept pair:
 - Select all hypernym paths that connect them
4. Choose the longest hypernym paths for the final taxonomy

For cleaning up the taxonomy graph, first the basic and root concepts are identified, which are the ones that are not hypernym of another concept, resp. have no hypernym. Then for every possible pair of a root concept and a basic concept all paths between the two are extracted, and only the longest one (or the longest ones) is retained.

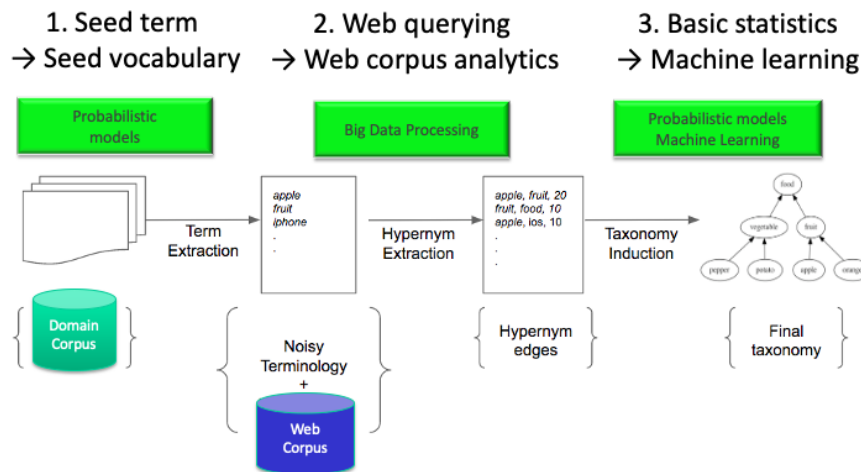
If t has no Hypernym ..

- A. It is a root concept
- B. It cannot match c such as t and X
- C. It is identical to the initial root concept
- D. It is a basic concept



©2019, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

Taxonomy Induction: 7 years later



Gupta, Amit, et al. "Taxonomy Induction using Hypernym Subsequences.", CIKM 2017

©2019, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

Information Extraction- 16

We will present now a more advanced technique for automated taxonomy induction that we recently developed. It takes advantage of different advances in the available tools, but also incorporates some very fundamental ideas that have not been taken into account in the literature and largely help to improve the quality of results. The main ideas are the following:

1. Instead of starting from a clean predefined set of seed terms, the method starts from a document corpus and uses keyphrase extraction to generate an initial vocabulary. This vocabulary may be noisy, but the method will help to clean it up while generating the taxonomy.
2. Instead of querying the Web, the method uses a Hypernym database that has been generating from analyzing a Web corpus.
3. Instead of performing simple statistics, the method uses various machine learning techniques for inducing the taxonomy.

Hypernym Extraction: WebIsADB

WebIsADB

This is a demo intended to show partial results.

Instance:

prefix lemma suffix

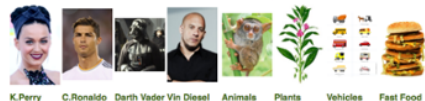
Class:

prefix lemma suffix

Tuple Frequency:

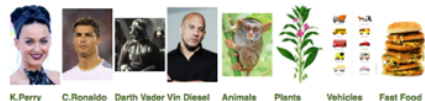
min max

Examples by instance



K. Perry C. Ronaldo Darth Vader Vin Diesel Animals Plants Vehicles Fast Food

Examples by class



K. Perry C. Ronaldo Darth Vader Vin Diesel Animals Plants Vehicles Fast Food

Found 30989 matches on WebIsADatabase:						
	PreTerm	Term	PostTerm	PrecClass	Class	PostClass
1		apple			company	5,536
2		apple			fruit	3,556
3		apple			apple	2,119
4		apple			vegetable	928
5		apple		tech	company	619
6		apple			brand	483
7		apple		hardware	company	480
8		apple			manufacturer	428
9		apple		technology	company	427
10		apple			company	383
11		apple			in the world	370
12		apple			food	353

<http://webdatacommons.org/isadb/>

Seitner, J., Bizer, C., Eckert, K., Faralli, S., Meusel, R., Paulheim, H. and Ponzetto, S.P., A Large DataBase of Hypernymy Relations Extracted from the Web. In *LREC* 2016

We have already seen WebIsADB that is used in this method for determining potential hypernym relationships.

Example: Hypernyms for “apple”

Candidate hypernym	Occurrence frequency
company	5536
fruit	3898
apple	2119
vegetable	928
orange	797
tech company	619
brand	463
hardware company	460
technology company	427
food	370

Note that hypernyms occur at difference level of abstraction
e.g. apple – fruit – food

WebISDB provides statistics of how often a term is in a possible hypernym relationship with another term. The example shows that the same term can be related to different terms at different levels of abstraction.

Key Observations

Current approaches in the literature make two main assumptions

1. The vocabulary is free of noise
 - Requires manual cleaning step before taxonomy induction
2. They assess taxonomy quality by estimating the probability of correctness of individual hypernym relationships
 - There is evidence that this works not well for more general terms

In particular the second assumption is important. Considering only individual hypernym relationships for assessing correctness means that contextual knowledge, i.e., the other relationships the terms have, is not considered and thus important information is lost.

Semantic Drift in Generalization

TopEdge	2	blintz→goody
	3	blintz→goody→thing
	4	blintz→goody→ulead→editor
	5	blintz→goody→ulead→social networking → networking→part
	6	blintz→goody→ulead→editor→storyliner→role
	<hr/>	
	2	oat→food
	3	oat→crop→thing
	4	oat→crop→total loss→partial loss→loss
	5	oat→cereal grain→grain→balanced diet→diet→factor
	6	oat→cereal→industry→field of life→other carrier→carrier

Considering only individual hypernym relationships frequently results in semantic drift, in particular at the higher levels of the taxonomy.

Key Ideas

1. Allow noisy vocabulary, but clean the resulting taxonomy after induction
 - After taxonomy induction more information is available to identify noisy terms
2. Estimate the probability of correctness of a complete path of hypernym relationships!
 - More contextual information is exploited in assessing the correctness of hypernym relationships

Based on the two key observations, two key ideas are applied in order to achieve better performance in taxonomy induction. Both ideas are based on the approach of exploiting more contextual information when performing the taxonomy induction task.

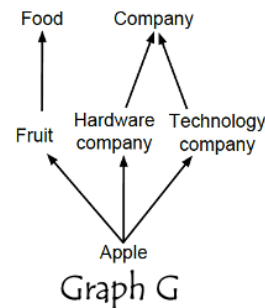
Approach

Find a DAG of generalizations

- Starting from one seed term in the vocabulary, e.g., "apple"
- Only one path for each hypernym of the seed term

Candidate hypernym	Freq.
company	5536
fruit	3898
apple	2119
vegetable	928
orange	797
tech company	619
brand	463
hardware company	460
technology company	427
food	370

Evidence E



Graph G

©2019, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

Information Extraction- 22

The basic approach of the method is to induce for each term in the vocabulary a DAG that consists of hypernym paths from the term to a root concept. In the approach it is assumed that for every hypernym of the term only a single path is generated. Each of the paths can correspond to a different sense of the term.

Probabilistic Model

Ideally: given the evidence E , find the most probable Graph G
 $\operatorname{argmax}_G P(G/E)$ (not feasible)

Aproximation: $\operatorname{argmax}_G P(G/E) = \operatorname{argmax}_G \prod_i P(E/S_i) \times P(S_i)$
where S_i subsequence from seed term to a root

An independence assumption

- But weaker than assuming all hypernym edges are independent of each other!

Need to estimate $P(E/S_i)$ and $P(S_i)$

©2019, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

Information Extraction- 23

Ideally, we would find the most likely graph structure, given the evidence E provided in the hypernym database. Since it is not feasible to explore the complete space of all possible graphs, we will apply an independence assumption. We assume that the different paths from a leaf to the root of a hypernym graph are independent of each other (which is of course not correct, since the paths are in general overlapping). But in this way we need no more to estimate the probability of a graph, but only those of path. Note that this is still a weaker independence assumption than the one being made when assuming that all hypernym edges are independent of each other.

Solution Strategy

1. Estimation of probabilities of subsequences
2. Search strategy for subsequences
3. Optimizing the resulting DAG

Probability Estimation: $P(S)$

For $S = t \rightarrow h_1 \rightarrow h_2 \rightarrow h_3 \rightarrow \dots \rightarrow h_n$

$$P(S) = P(t, h_1) \times P(h_1, h_2) \times \dots \times P(h_{n-1}, h_n)$$

$P(a, b)$ - edge probability

Edge Probability

$$P(a, b) \propto \exp(\mathbf{w} \cdot \mathbf{f}(a, b))$$

Edge features $\mathbf{f}(a, b)$

- Normalized count, $n(a, b) = \text{freq}(a, b) / \max_c (\text{freq}(a, c))$
- Normalized difference
- String-based features (prefix, suffix, substring, length)
- Generality based features

Weights \mathbf{w} obtained from a classifier trained on a manually annotated set of edges

Probability Estimation: $P(E|S)$

For $S = t \rightarrow h_1 \rightarrow h_2 \rightarrow h_3 \rightarrow \dots \rightarrow h_n$

$$P(E|S) = \sum_j P(E_j|S)$$

$$P(E_j|S) \propto \max(\text{sim}(E_j, h_i)) \text{ for all } h_i$$

$$\text{sim}(a, b) = \max(P(a, b), P(b, a))$$

Why maximum (and not, e.g., sum)?

- Consider $S = \text{apple} \rightarrow \text{fruit} \rightarrow \text{food} \rightarrow \text{substance} \rightarrow \text{matter} \rightarrow \text{entity}$
- For the term $E_j = \text{fruit}$, the hypernyms *matter* and *entity* occurring in the sequence should not hurt, even when unrelated

Intuitive Interpretation

$\Pr(S)$ promotes subsequences, which consist of individual edges with a larger probability of hypernymy

$\Pr(E|S)$ promotes subsequences, which contain a larger number of candidate hypernyms from E

Finding Optimal Subsequences

Basic algorithm

1. Iterate over all candidate hypernyms h of t
2. For each h perform a depth-limited beam search over the space of possible subsequences by recursively exploring the candidate hypernyms of h
3. For each h choose the subsequence S with the highest score
4. Choose the top- b candidates based on sequence scores

Problem:

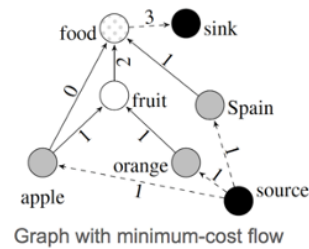
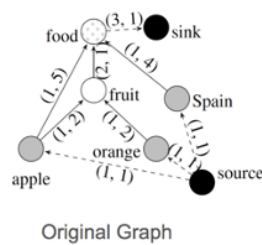
for term t , find best subsequence with first edge $t \rightarrow h$

Optimizing the Resulting DAG G

Objective: prune the DAG composed of the set of subsequences constructed (noise removal)

Flow Network Optimization

- demand controls how many seed terms should be discarded
- Cost related to quality of paths



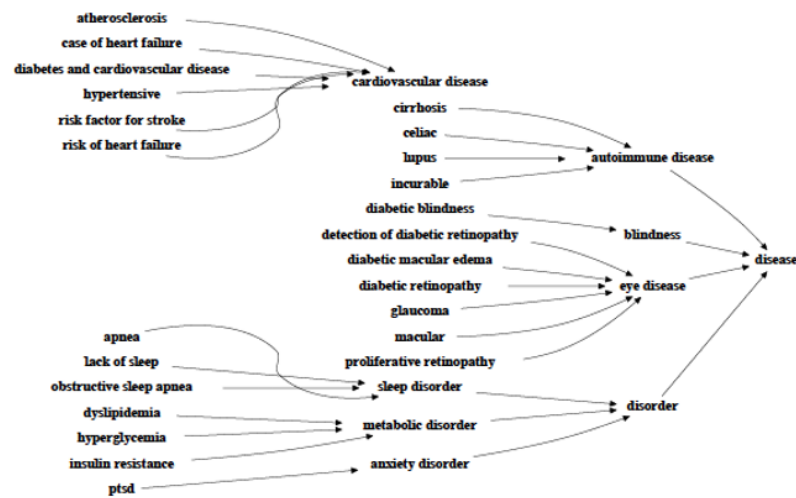
Results

TopEdge	2	blintz→goody
	3	blintz→goody→thing
	4	blintz→goody→ulead→editor
	5	blintz→goody→ulead→social networking →networking→part
	6	blintz→goody→ulead→editor→storyliner→role
	2	oat→food
	3	oat→crop→thing
	4	oat→crop→total loss→partial loss→loss
	5	oat→cereal grain→grain→balanced diet→diet→factor
	6	oat→cereal→industry→field of life→other carrier→carrier
	2	blintz→homemade jewish food→food
	3	blintz→homemade jewish food→food→supply
SubSeq	4	blintz→thin pancake→pastry→snack food→food
	5	blintz→homemade jewish food→food→supply→necessity→thing
	6	blintz→homemade jewish food→food→supply→keyword→beta test→test
	2	oat→cereal grain→grain
	3	oat→cereal grain→grain→supply
	4	oat→cereal grain→grain→complex carbohydrate→carbohydrate
	5	oat→cereal grain→grain→complex carbohydrate→carbohydrate→essential nutrient→nutrient
	6	oat→cereal grain→grain→supply→keyword→beta test→test

©2019, Kari Aberer, EPFL-IL, Laboratoire de systèmes d'informations repartis

201901100011 14:00:00 2.2

Example: Resulting Taxonomies



References

Course material based on

- Kozareva, Zornitsa, and Eduard Hovy. "A semi-supervised method to learn and construct taxonomies using the web." *Proceedings of the 2010 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 2010.
- Gupta, A., Lebrecht, R., Harkous, H., & Aberer, K. (2017, November). Taxonomy Induction using Hypernym Subsequences. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 1329-1338). ACM.
- Faralli, Stefano, et al. "A Web application to search a large repository of taxonomic relations from the Web." *CEUR workshop proceedings*. Vol. 1690. RWTH, 2016.