Building Reliable Activity Models Using Hierarchical Shrinkage and Mined Ontology

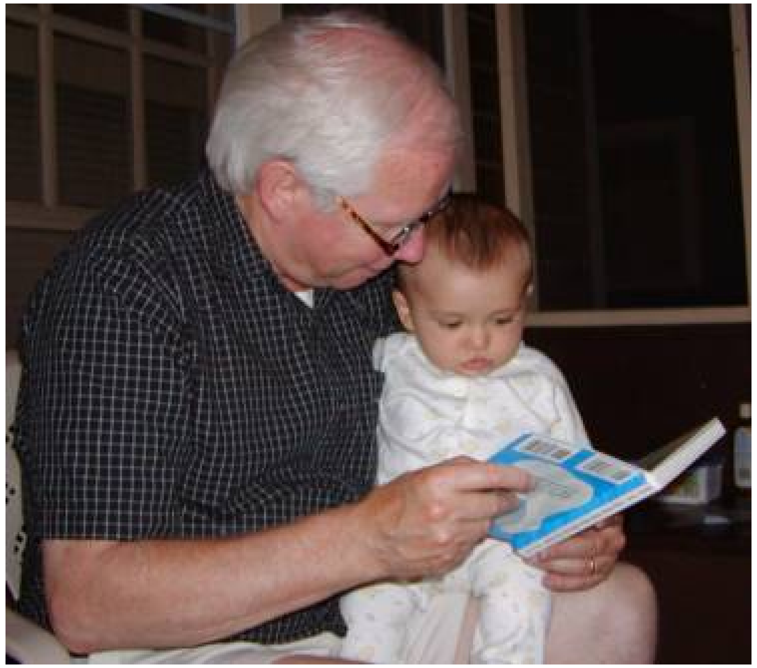
Emmanuel Munguia Tapia[1],
Tanzeem Choudhury and Matthai Philipose[2]
[1]Massachusetts Institute of Technology
[2]Intel Research Seattle

Activity Recognition applications

- Activity-aware actuation
- Proactive reminding
- Ubiquitous healthcare
- Embedded health assessment

Activities of Daily Living (ADLs)

Activity Class
Personal Appearance
Housework
Toileting
Washing up
Appliance Use
Taking care or an infant
Care of clothes and linen
Making a snack
Making a drink
Oral hygiene
... 23 classes, 1000s of activities



Classes of day-to-day activities that:

- Indicate cognitive well-being
- Indicate level of independence

e.g. Is elder still able to prepare a pasta?

What sensors could we use?

Activity inference based on object use

Dense sensing: attach sensors directly to objects in the environment.

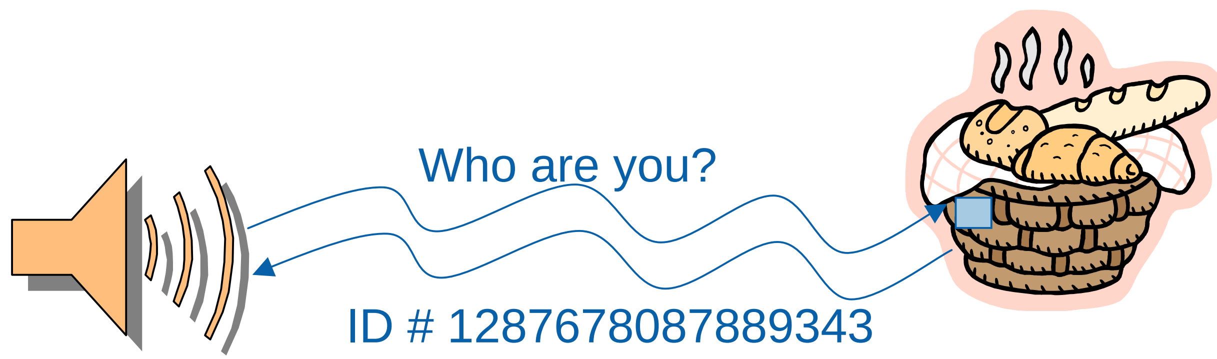
•Battery-free wireless stickers (RFIDs)

•Battery powered sensor nodes

Activity inference based on object use

Advantages

- Robust to environmental conditions
- High level information can be associated with the sensors



<ID = e3f000e13431, desc = “bread basket”, manufacturer = “…”, … >

What inference algorithms could we use?

Activity inference using Dynamic Bayesian networks (HMMs)

Advantages

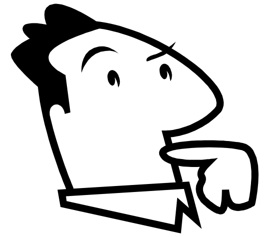
- Easy to incorporate common sense information
- Efficient inference algorithms

To create an activity model we need

- List of objects used while performing an activity
- Probability of using the objects

e.g $p(pot|cooking) = 0.7$

Techniques for constructing activity models: Hand definition



prepare pasta

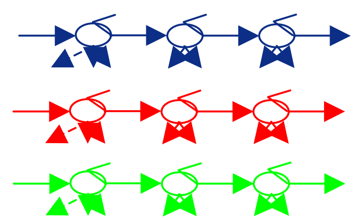
list of objects

```
pot
stove
spoon
spaghetti
```

probability of using objects

```
P(pot|boiling pasta)        0.2
P(stove|boiling pasta)      0.2
P(spoon|boiling pasta)      0.1
P(Spaghetti|boiling pasta) 0.5
```

models



Techniques for constructing activity models: Learn from data

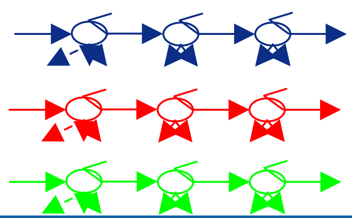
unlabeled data (objects used)

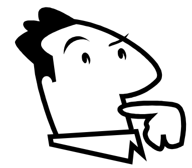
label

labeled data

learn

custom models

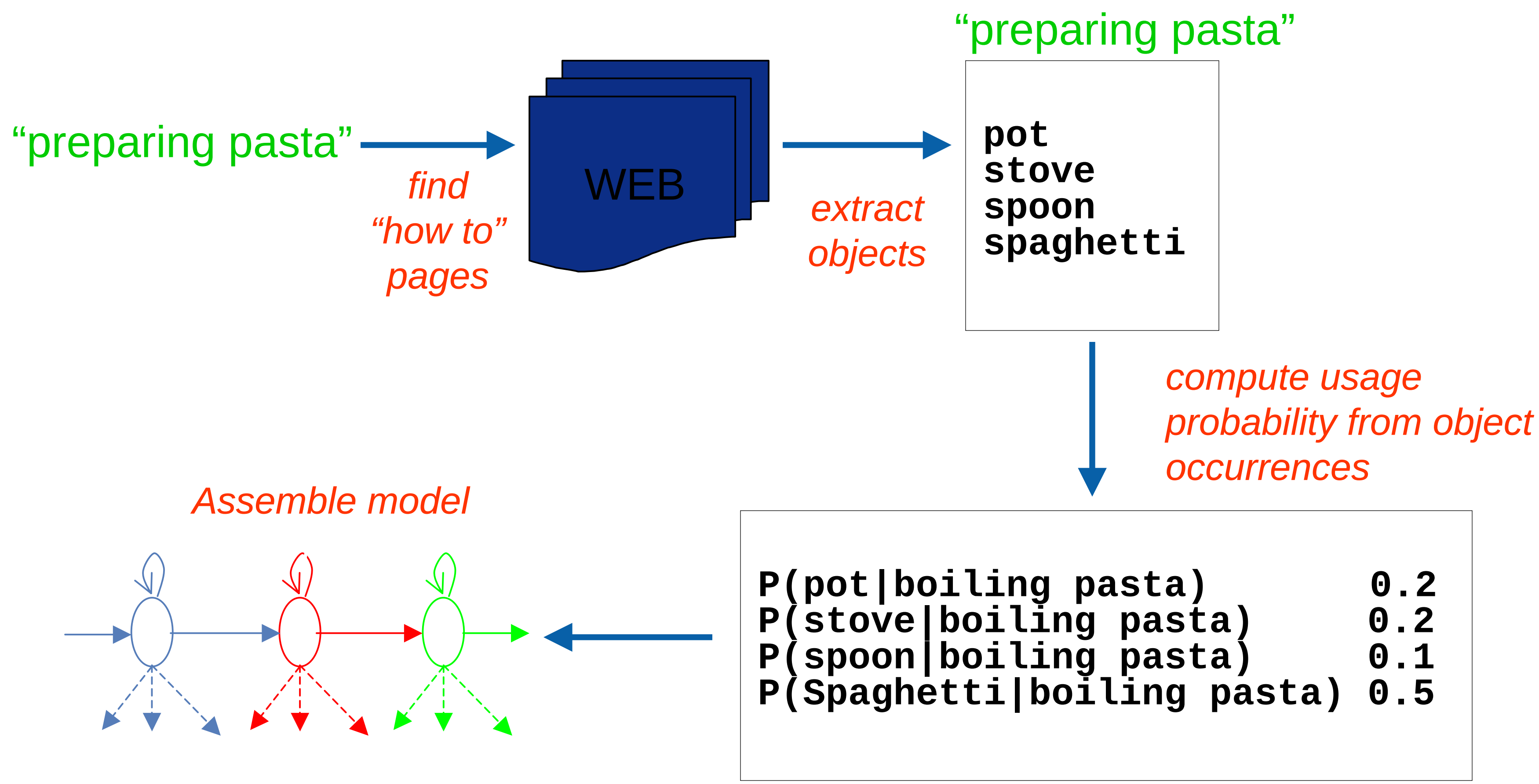




prepare pasta 20 times

Techniques for constructing activity models: Mine activity from the web

(Wyatt et. al. AAAI ‘05)



Judging the level of independence of an elder

(1)inappropriate probabilities (2) Missing objects

Preparing pasta

pot → kitchen range → spaghetti → spoon

Judging the level of independence of an elder

(1)inappropriate probabilities (2) Missing objects

Preparing pasta

pot → kitchen range → spaghetti → spoon

pan → stove → macaroni → fork

Dealing with Incompleteness

Exploit common sense information about objects that are functionally similar

pot → pan   spoon → fork

- automatically extract ontology from a hierarchically organized lexical system called WordNet

Adapt or improve model probabilities based on object similarity information

- apply statistical smoothing technique known as *shrinkage* to update model parameters

WordNet: Semantic relationships

Microwave#1: electromagnetic wave

Synset: microwave

Microwave#2: kitchen appliance that cooks food

Synset: microwave, microwave oven

WordNet: Semantic relationships

Microwave#1: electromagnetic wave

Synset: microwave

Hypernym tree

Electromagnetic radiation
- radiation
  - energy
    - Physical phenomenon
      - Natural phenomenon
        - Phenomenon

Microwave#2: kitchen appliance that cooks food

Synset: microwave, microwave oven

Hypernym tree

Kitchen appliance
- appliance
  - durables
    - Consumer goods
      - object
        - Entity

Hypernym: when a word sense is a superset of another

WordNet: Semantic relationships

How do we select the word sense automatically?

Hypernym tree	**Hypernym tree**
Electromagnetic radiation	Kitchen appliance
radiation	appliance
energy	durables
Physical phenomenon	Consumer goods
Natural phenomenon	object
Phenomenon	Entity

WordNet: Unique beginners for nouns

Thing, entity

Non-living thing, object

Living thing, organism

- ***Natural object***
- ***Artifact***
- ***Substance***
- ***food***

- ***Plant, flora***
- *Animal, fauna*
- *Person, human being*

WordNet: Semantic relationships

How do we select the word sense automatically?

Hypernym tree	**Hypernym tree**
Electromagnetic radiation	Kitchen appliance
radiation	appliance
energy	durables
Physical phenomenon	Consumer goods
Natural phenomenon	object
Phenomenon	Entity

Identify objects that are (1) nouns (2) subset of entity

WordNet: Semantic relationships

How do we select the word sense automatically?

~~**Hypernym tree**~~

~~Electromagnetic radiation~~
~~radiation~~
~~energy~~
~~Physical phenomenon~~
~~Natural phenomenon~~
~~Phenomenon~~

Hypernym tree

Kitchen appliance
appliance
durables
Consumer goods
object
Entity

Identify objects that are (1) nouns (2) subset of entity

Ontology extraction from WordNet

From the activity recipe mined for "preparing pasta", we have the list of objects used.

preparing pasta

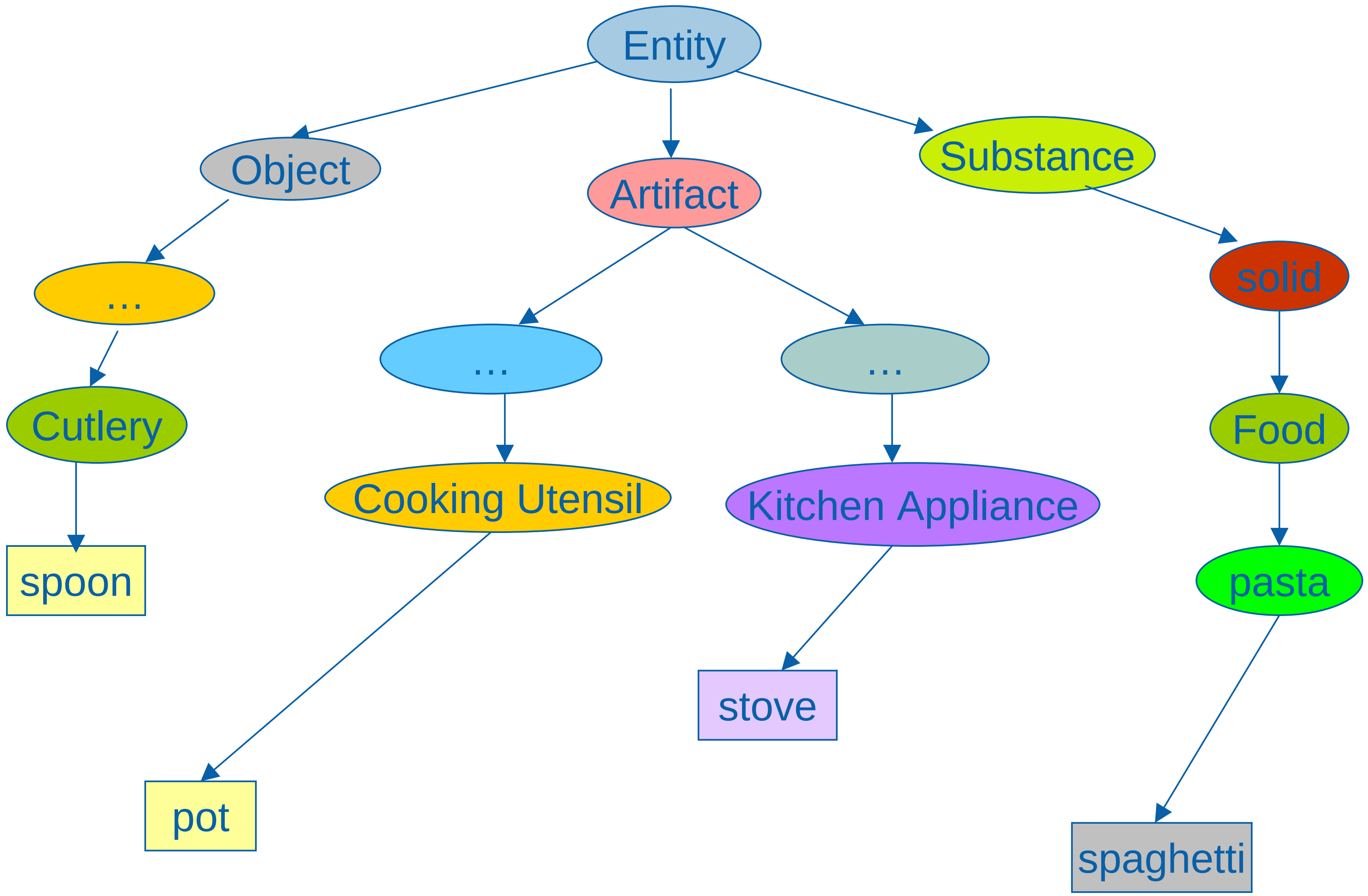
pot
stove
spoon
spaghetti

Ontology extraction from WordNet

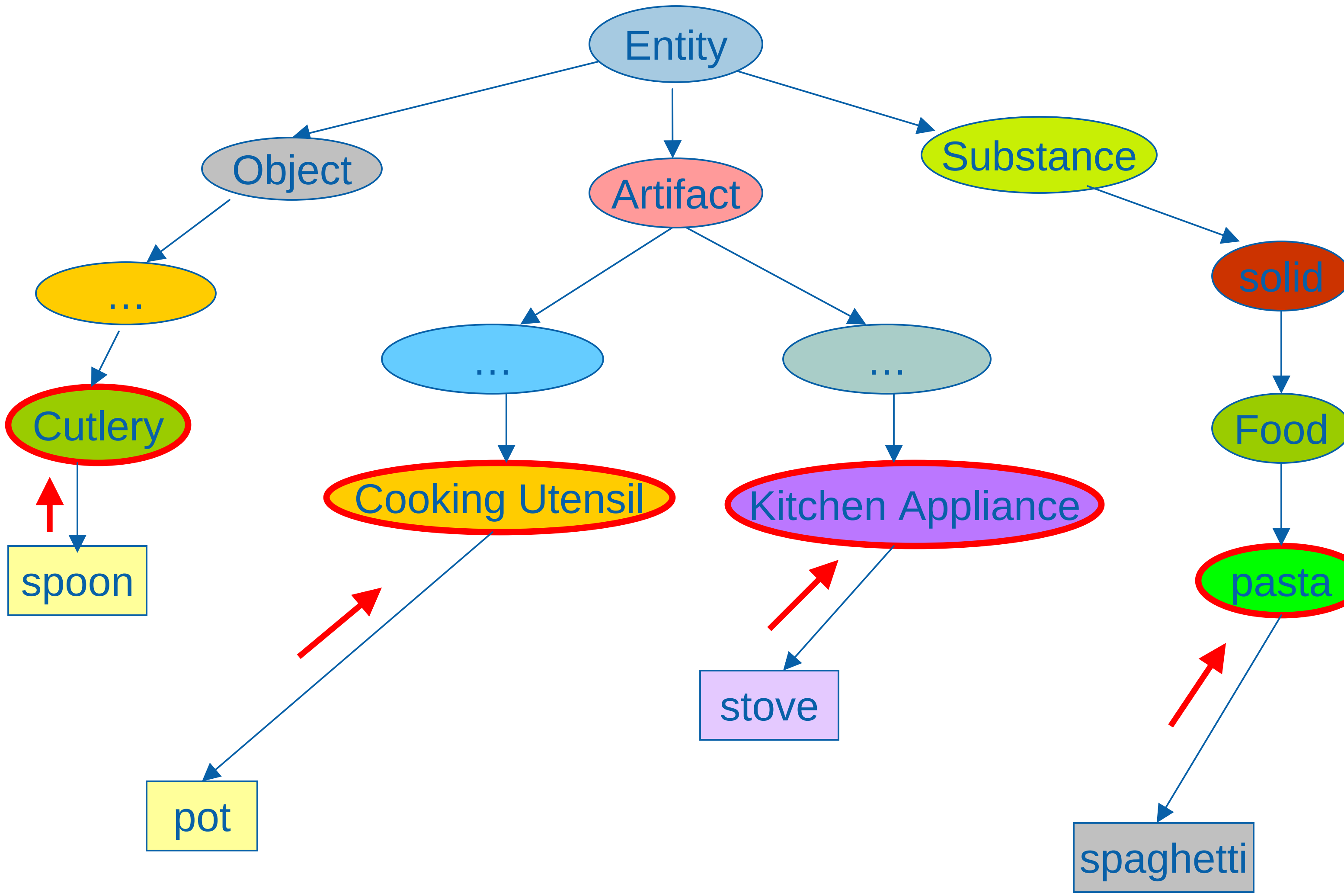
Finding the first hypernym tree that contain the concept entity for all the objects in our list we get

pot	**stove**	**spoon**	**pasta**
Cooking utensil	kitchen appliance	cutlery	food
Utensil	home appliance	tableware	solid
Implement	appliance	ware	substance
Instrumentality	durables	article	entity
Artifact	consumer goods	artifact	
Entity	commodity	object	
	artifact	entity	
	entity		

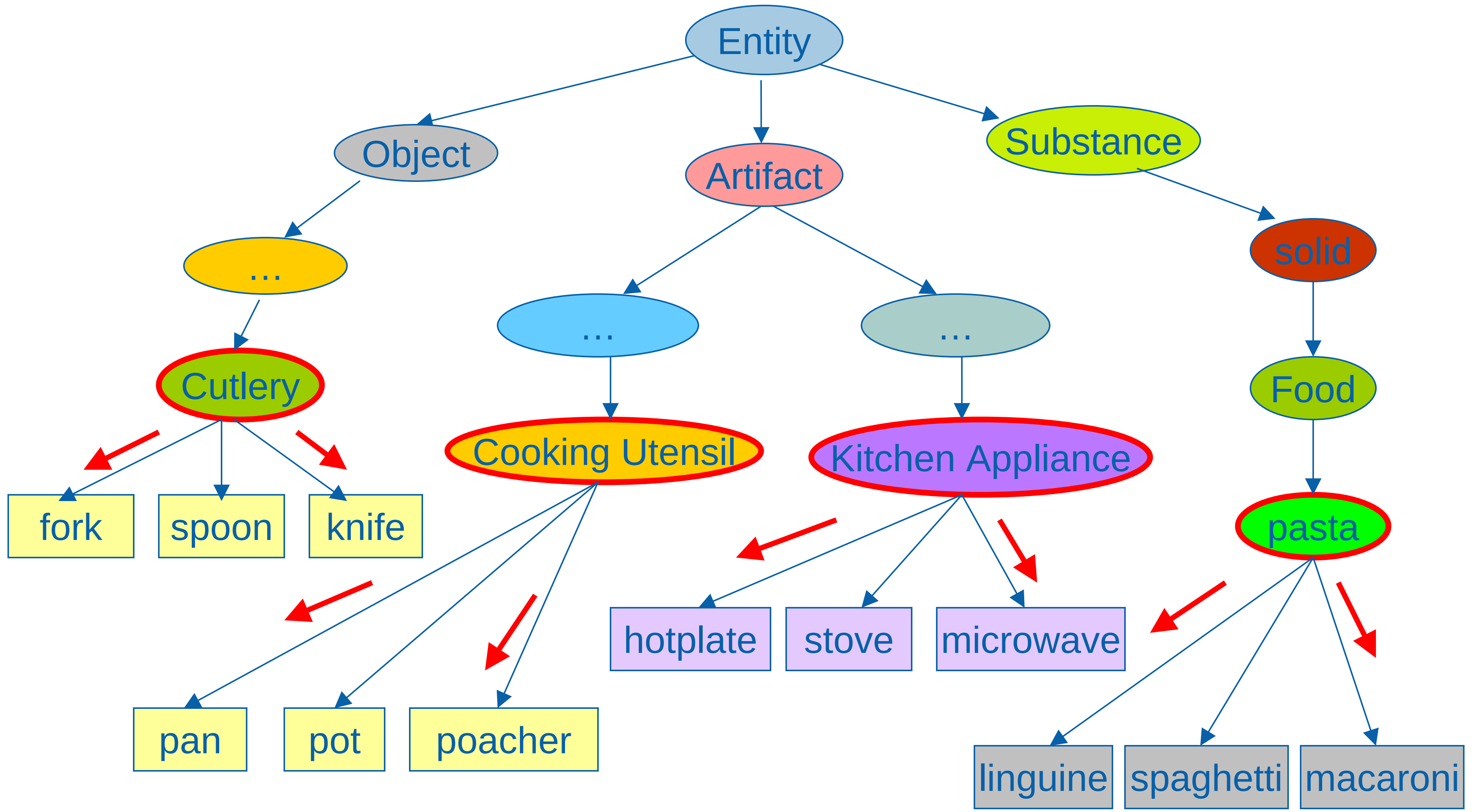
WordNet ontology generation



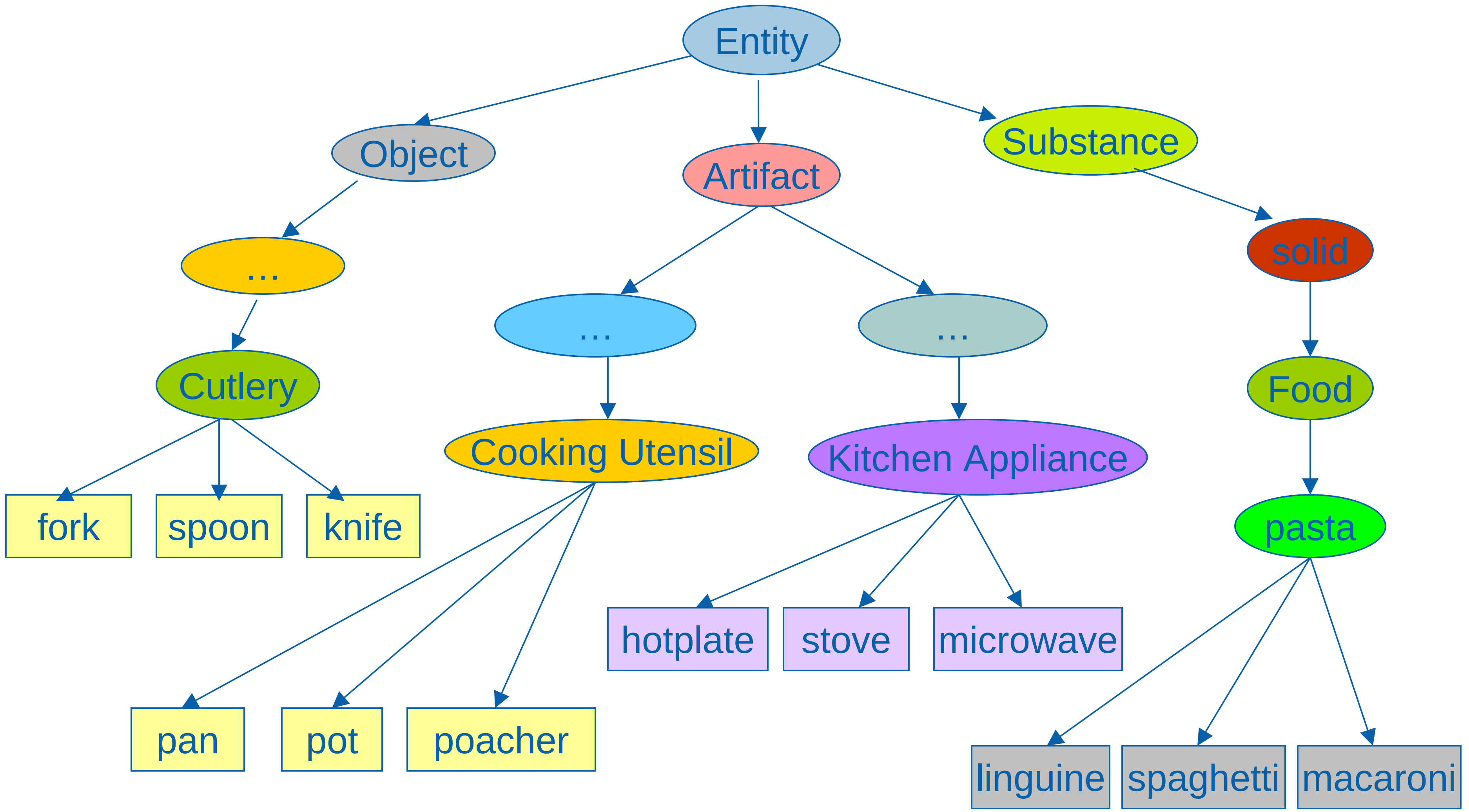
WordNet ontology expansion



WordNet ontology expansion



WordNet ontology



Ontology extraction from WordNet

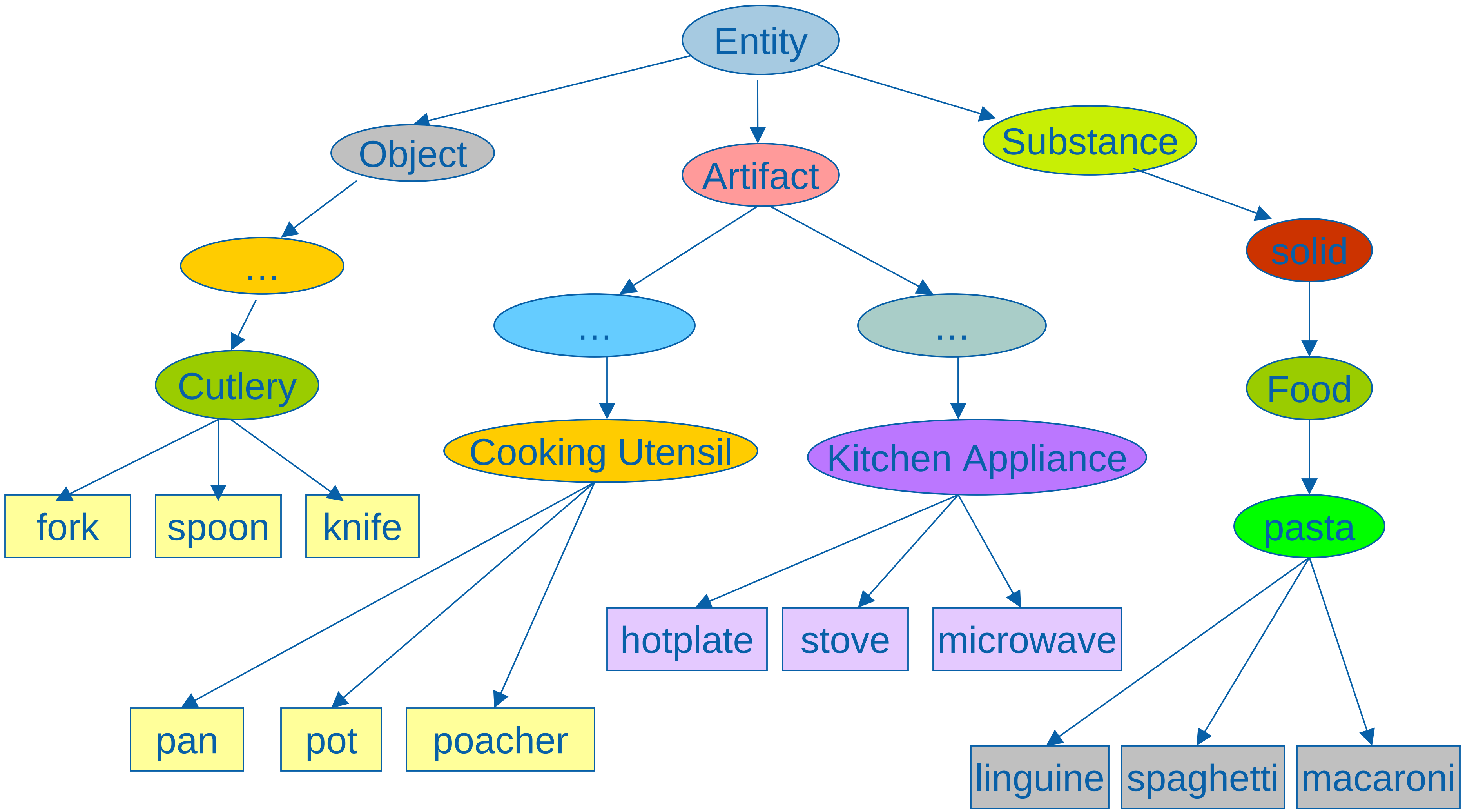
From the activity recipe mined for “preparing pasta”, we also have the probabilities of object use

preparing pasta

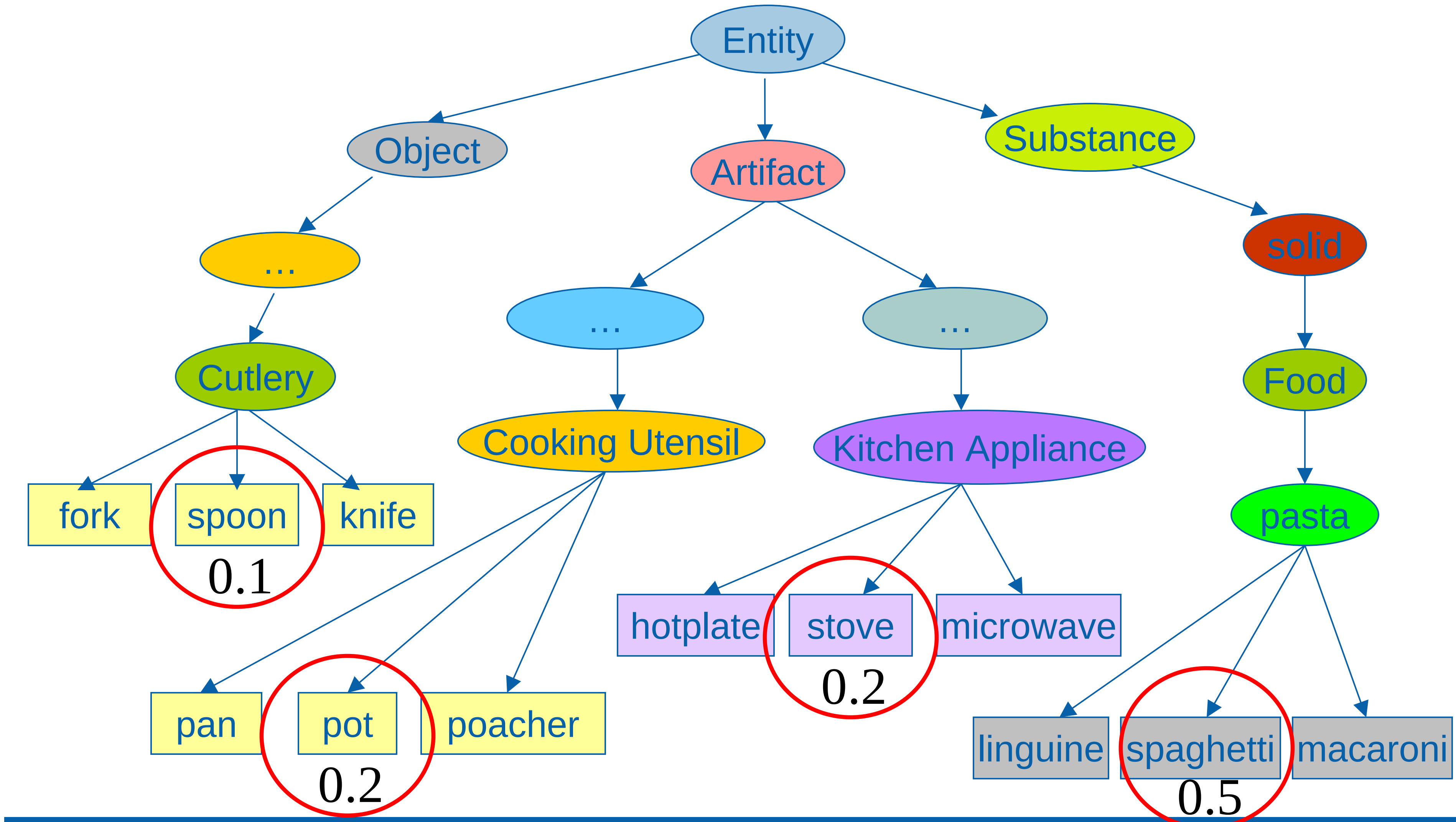
pot
stove
spoon
Spaghetti

P(pot|preparing pasta)	**0.2**
P(stove| preparing pasta)	**0.2**
P(spoon| preparing pasta)	**0.1**
P(Spaghetti| preparing pasta)	**0.5**

Ontology: setting probabilities

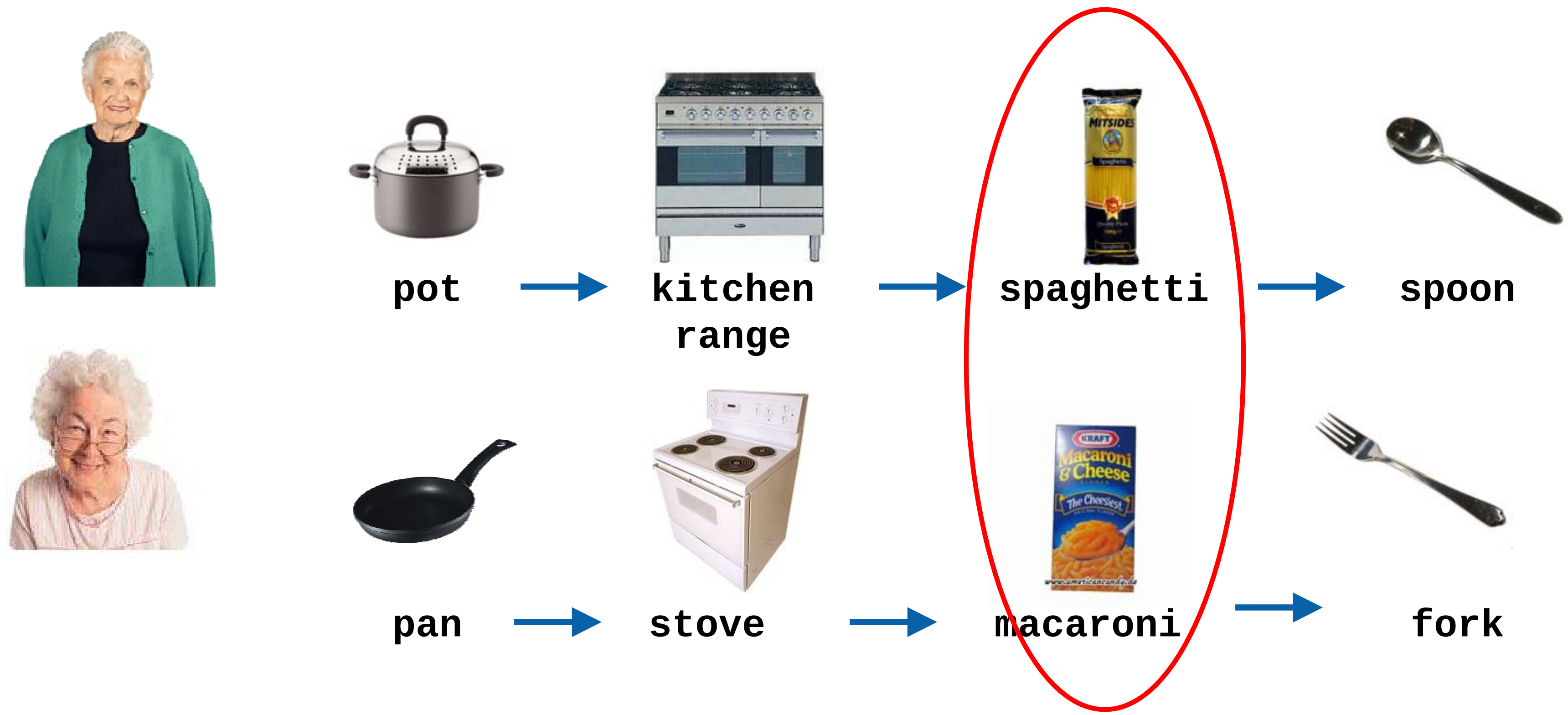


Ontology: setting probabilities

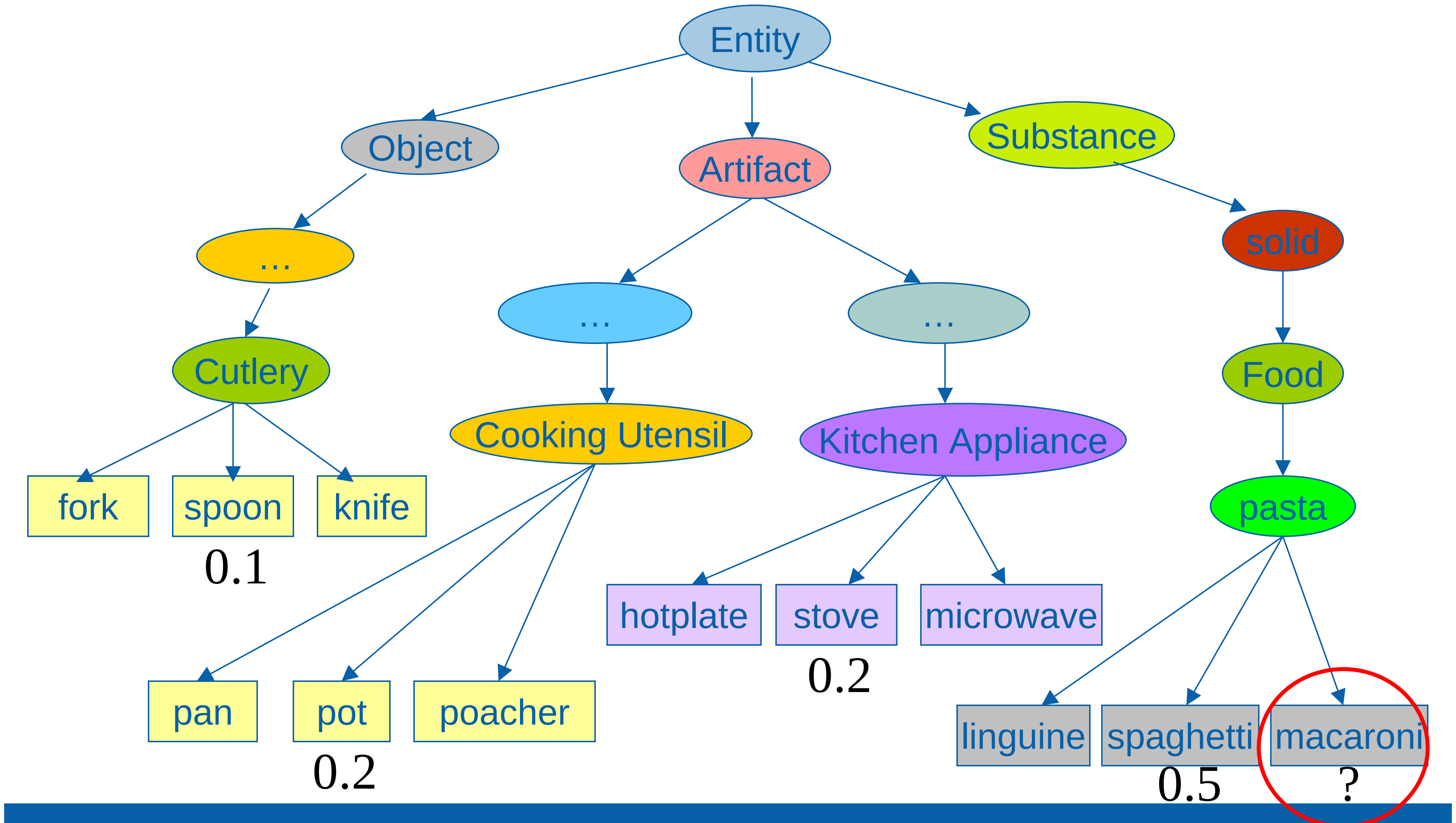


What if we use macaroni instead?

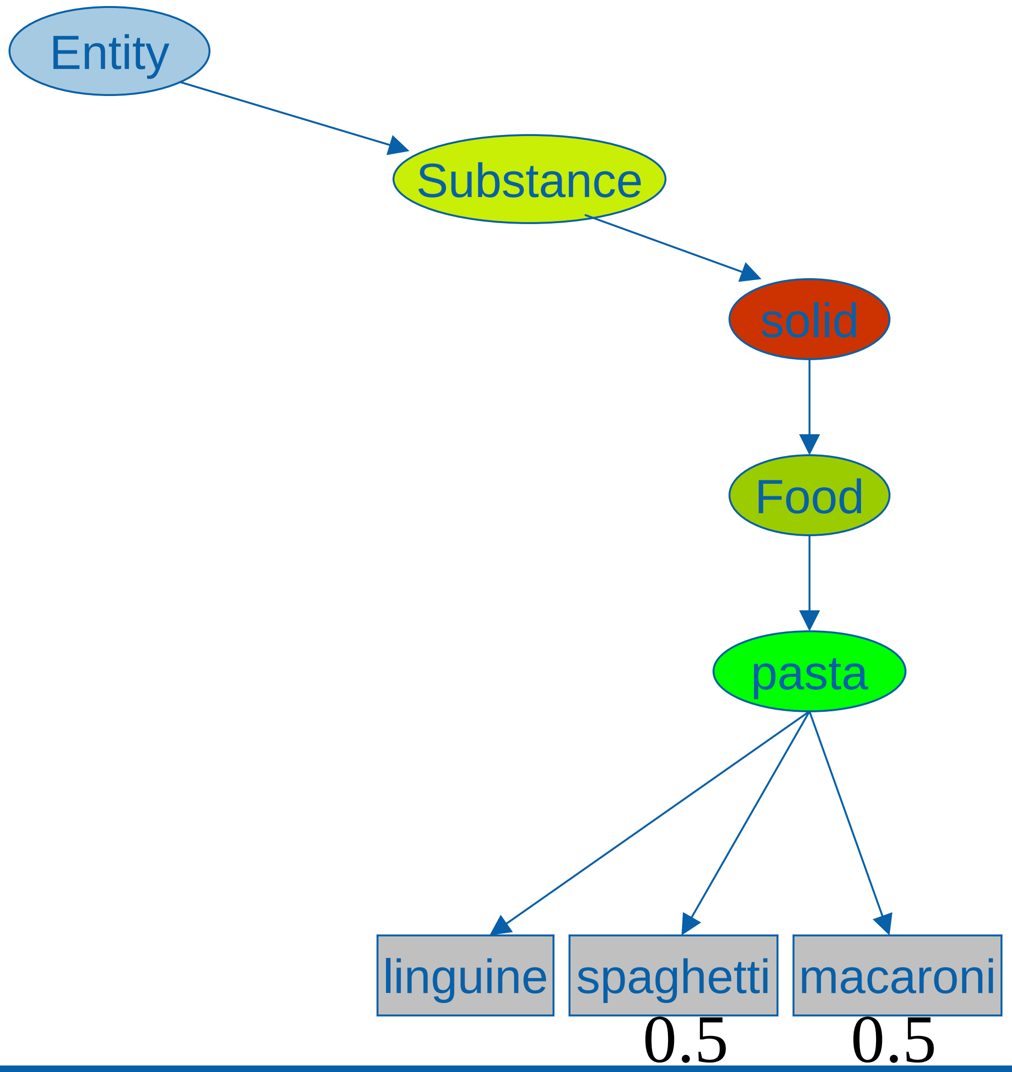
Preparing pasta



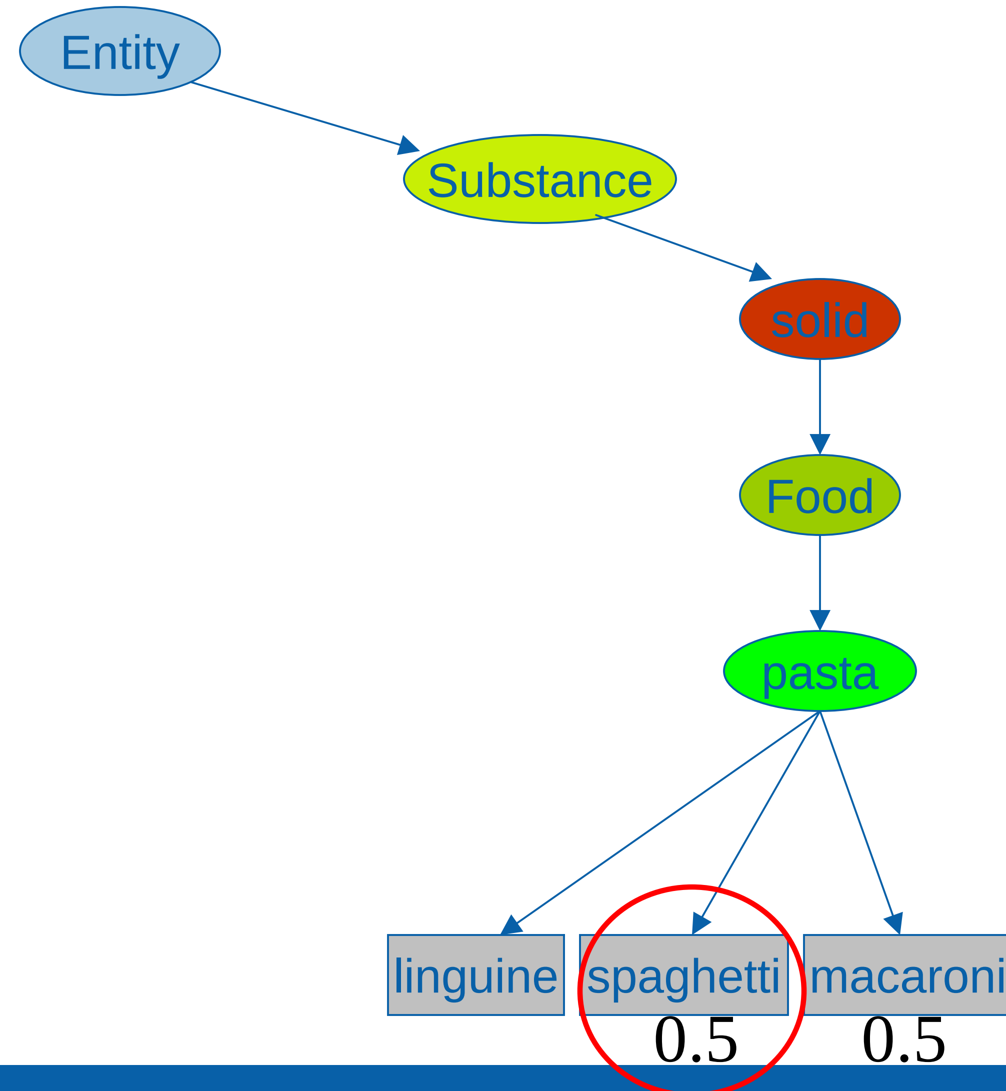
How do we propagate probabilities?



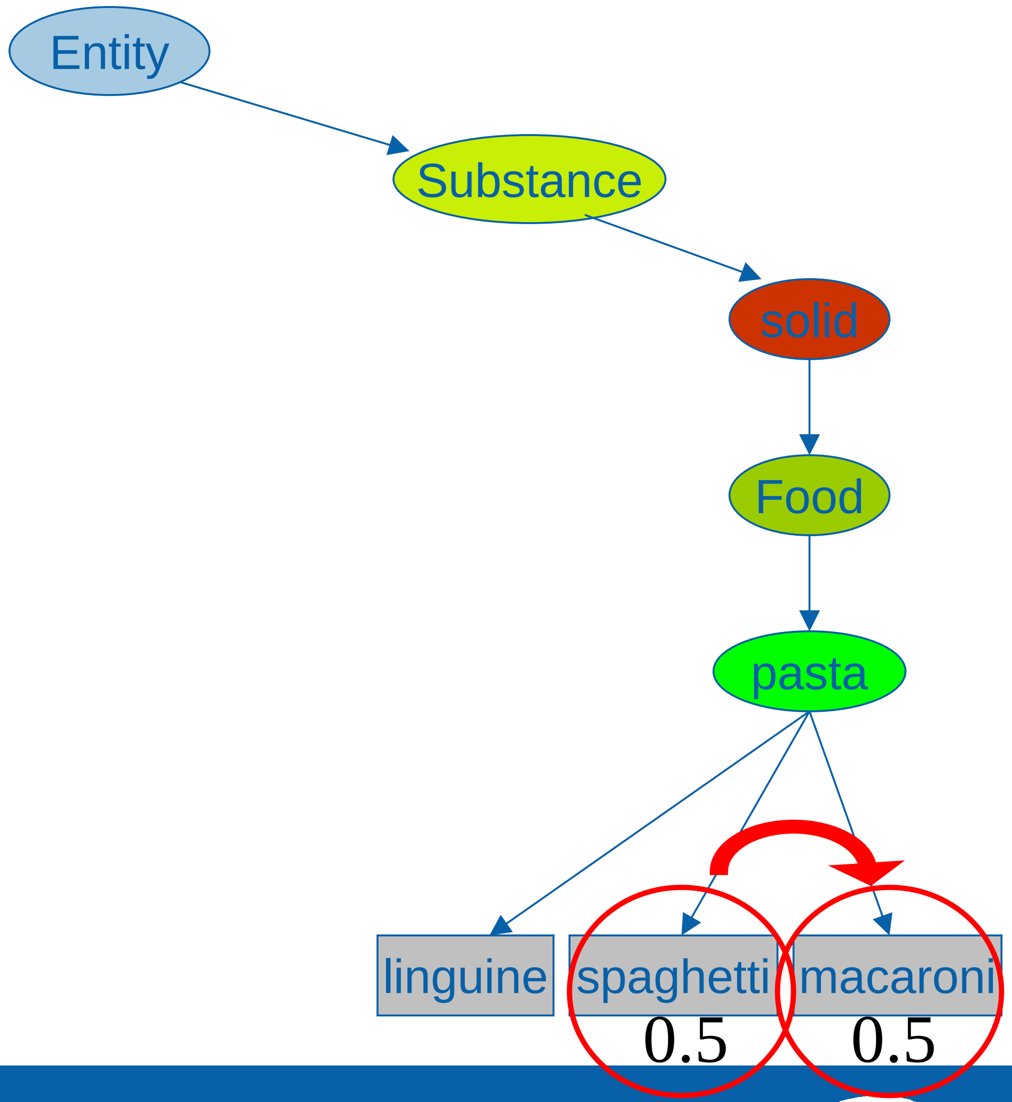
Copy probability from sibling node to sibling node



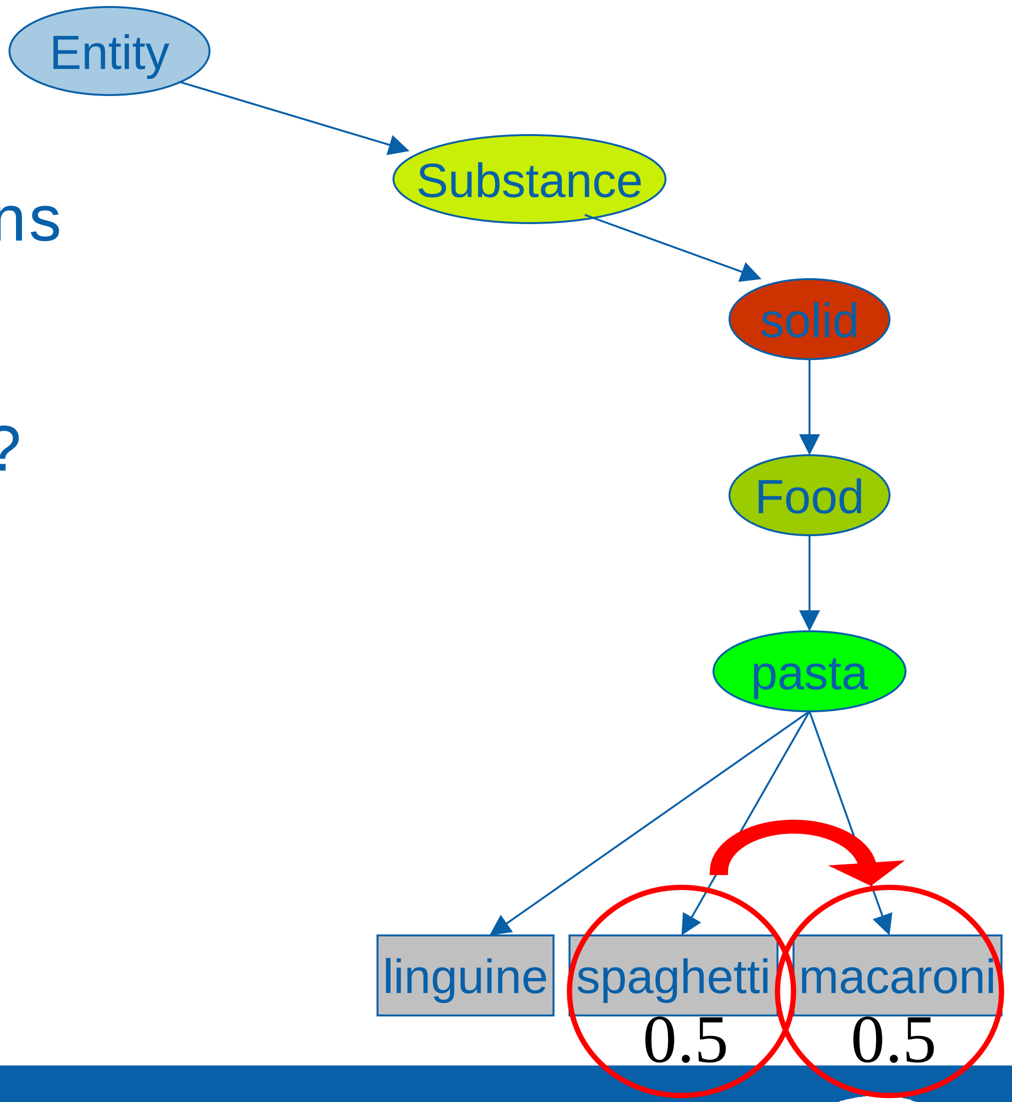
Copy probability from sibling node to sibling node



Copy probability from sibling node to sibling node



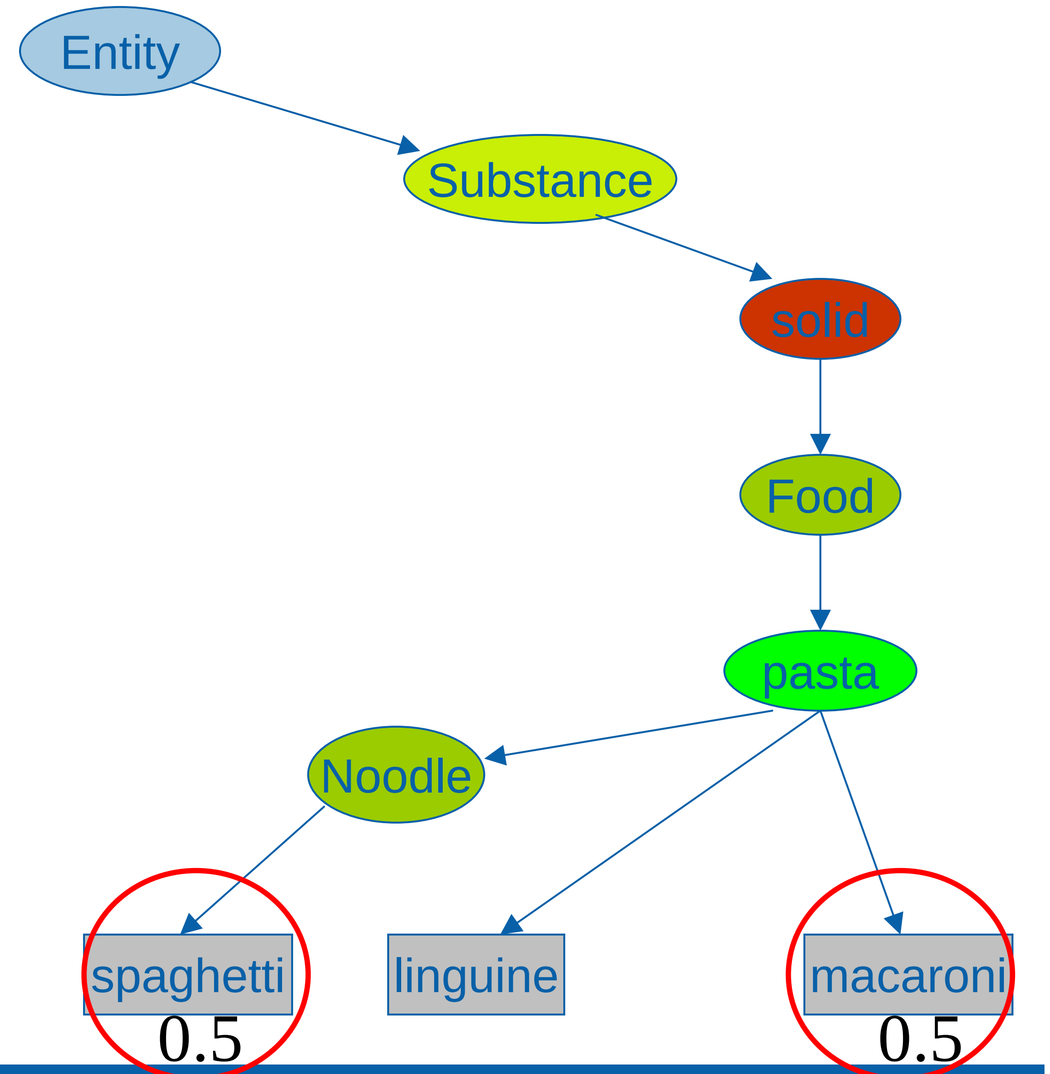
Copy probability from sibling node to sibling node



Problem: What happens if the probability for spaghetti is inappropriate or bad?

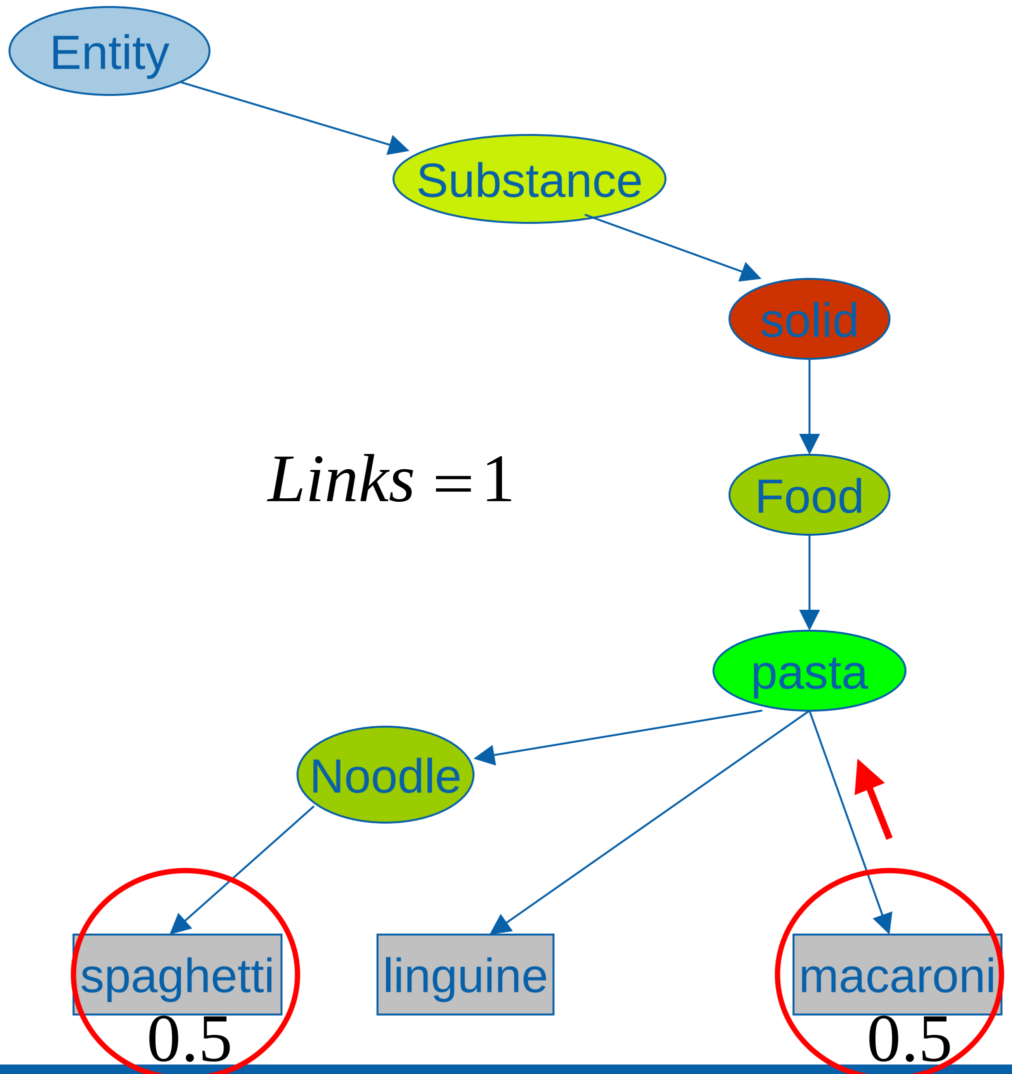
Count number of links from node to node

Problem: What happens if spaghetti is not an immediate sibling of macaroni?



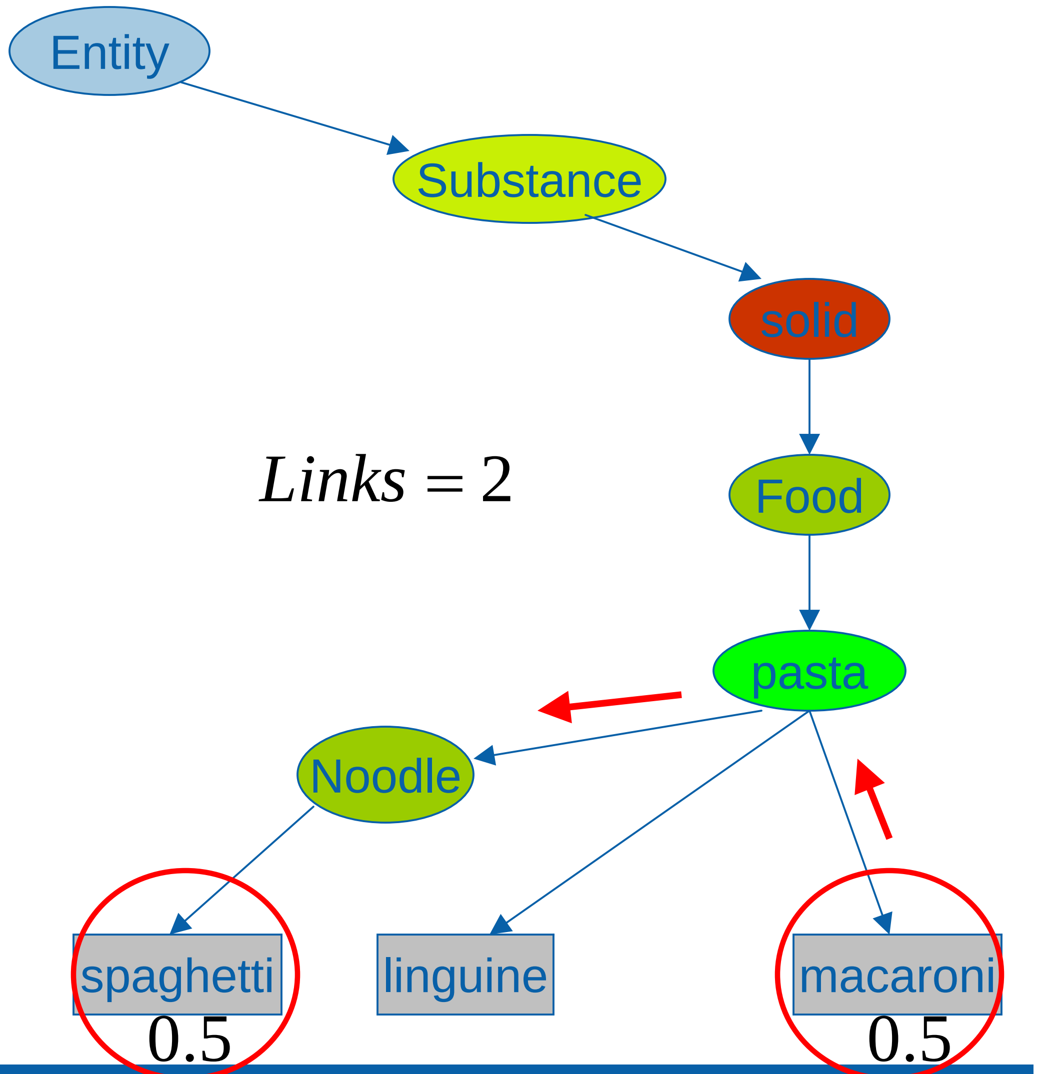
Count number of links from node to node

Problem: What happens if spaghetti is not an immediate sibling of macaroni?



Count number of links from node to node

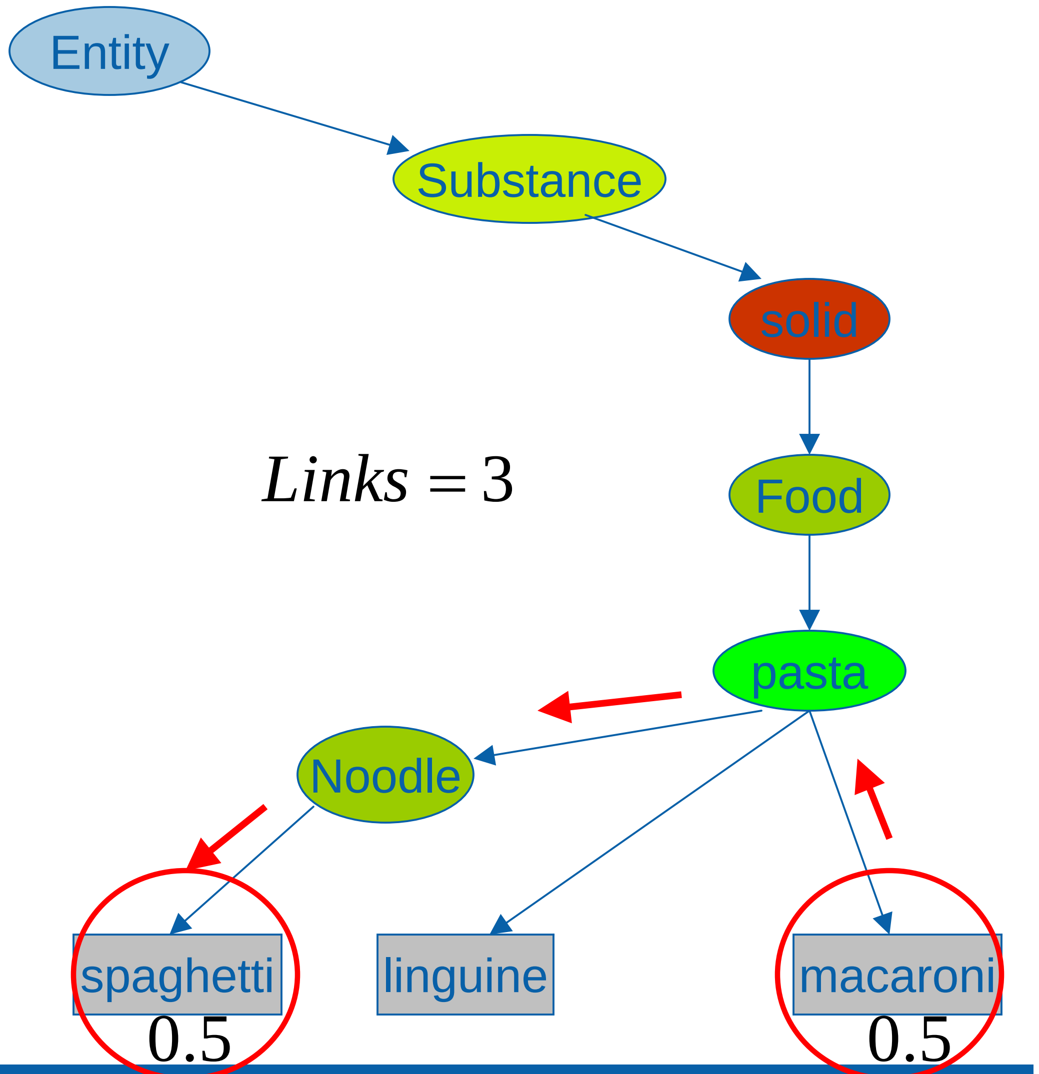
Problem: What happens if spaghetti is not an immediate sibling of macaroni?



$$Links = 2$$

Count number of links from node to node

Problem: What happens if spaghetti is not an immediate sibling of macaroni?

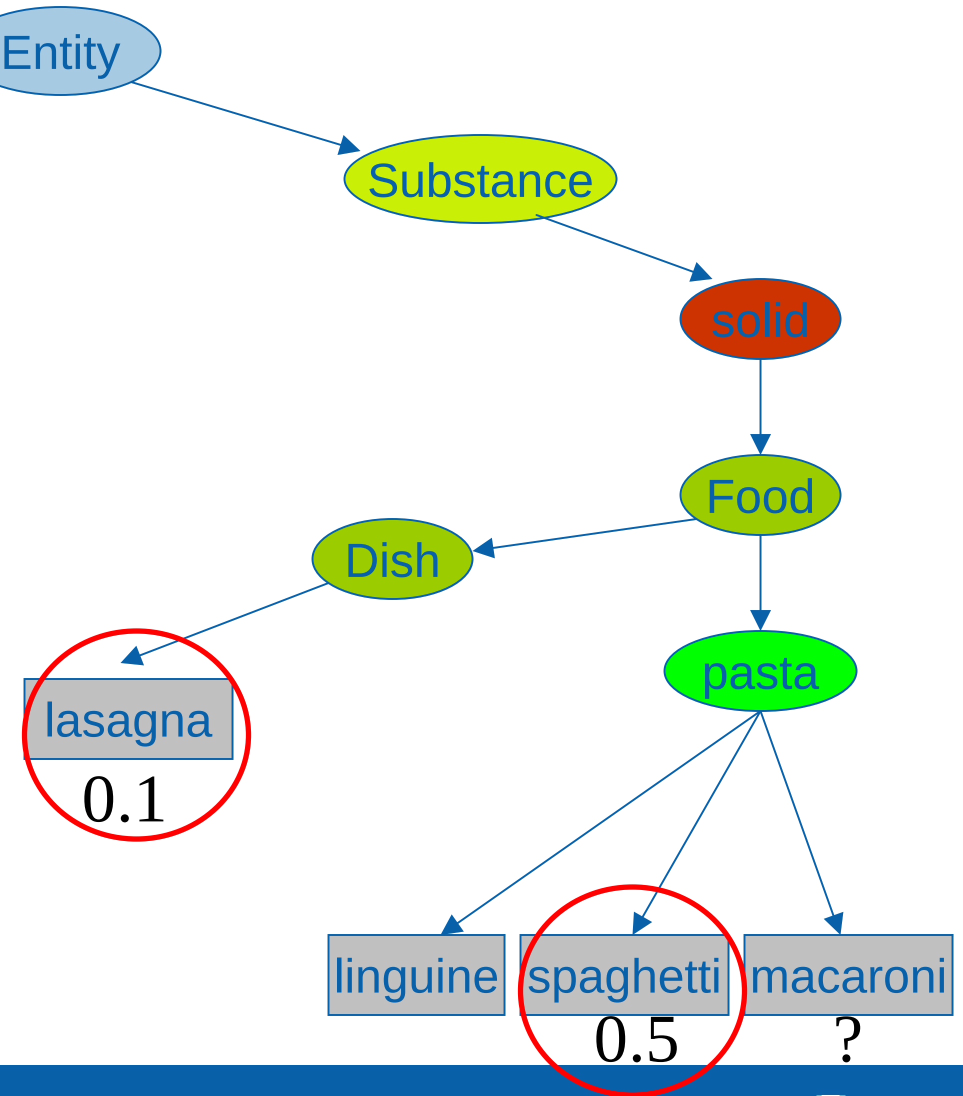


$$Links = 3$$

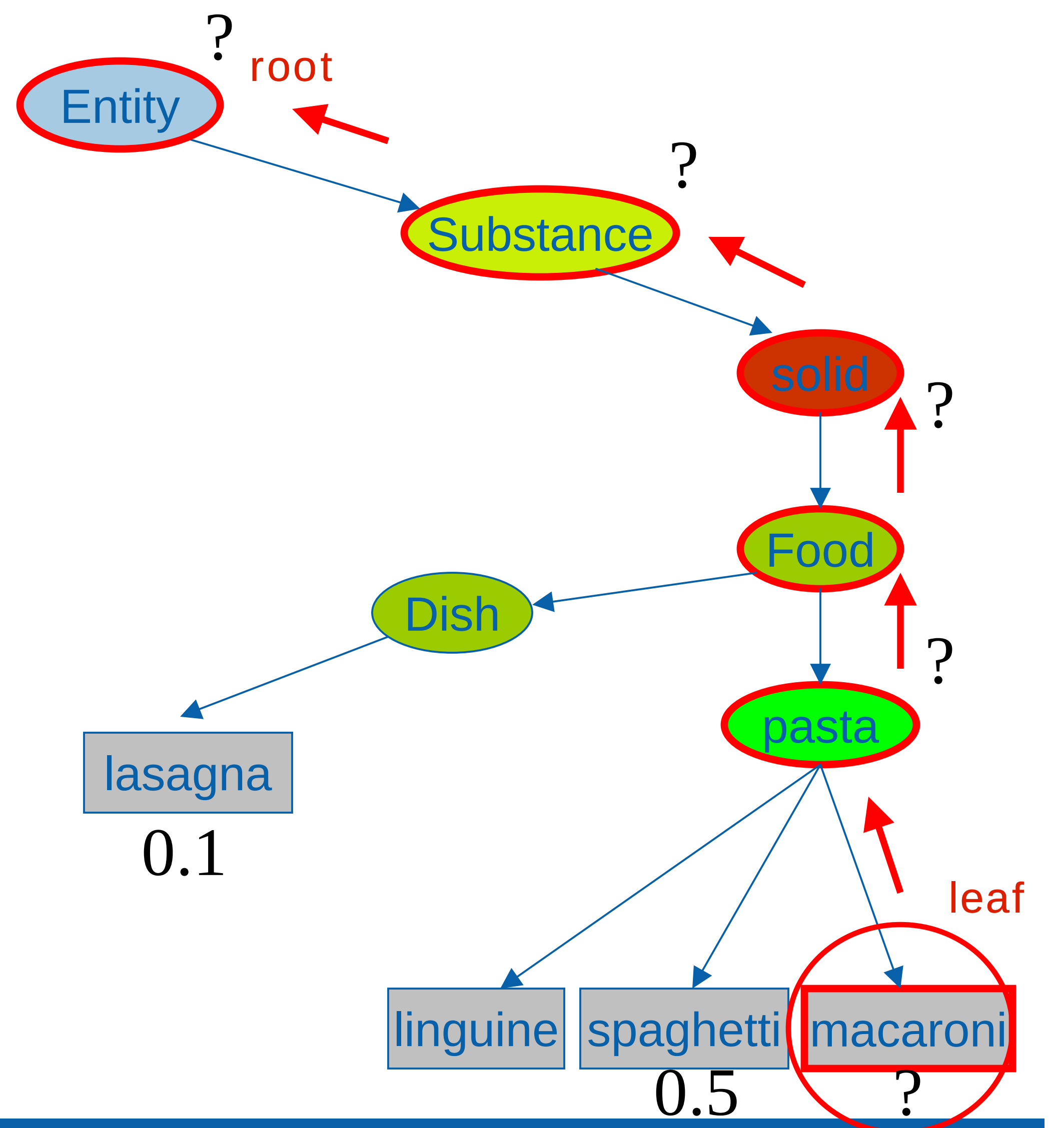
Taking additional information into account

What happens if we also know that somebody preparing pasta can use lasagna?

None of the two previous techniques can take advantage of this extra information



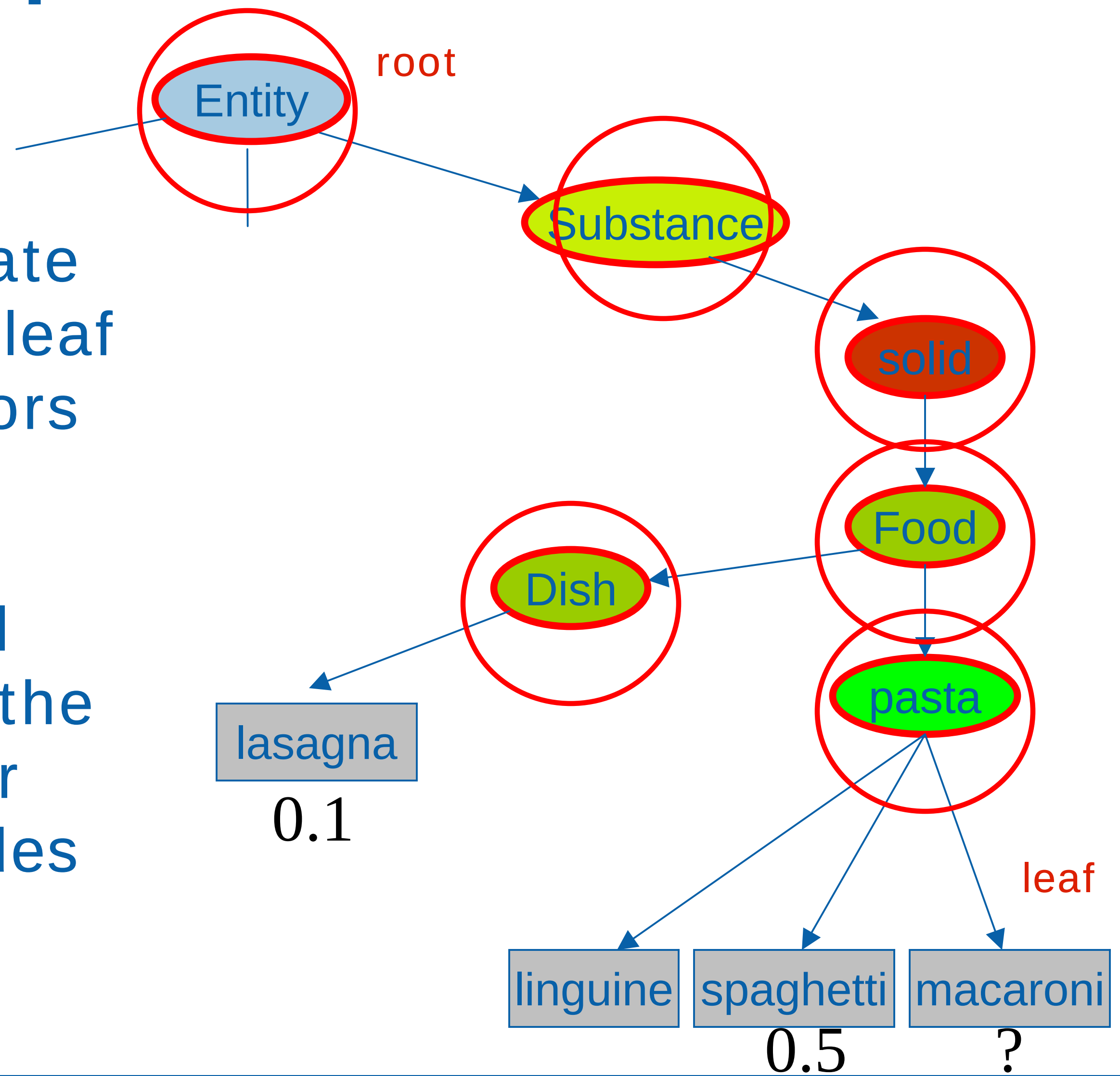
Shrinkage: Key idea



Idea: find or improve the parameter estimate of leaf nodes by linearly interpolating the estimates of its ancestor nodes

How do we compute estimates for internal nodes?

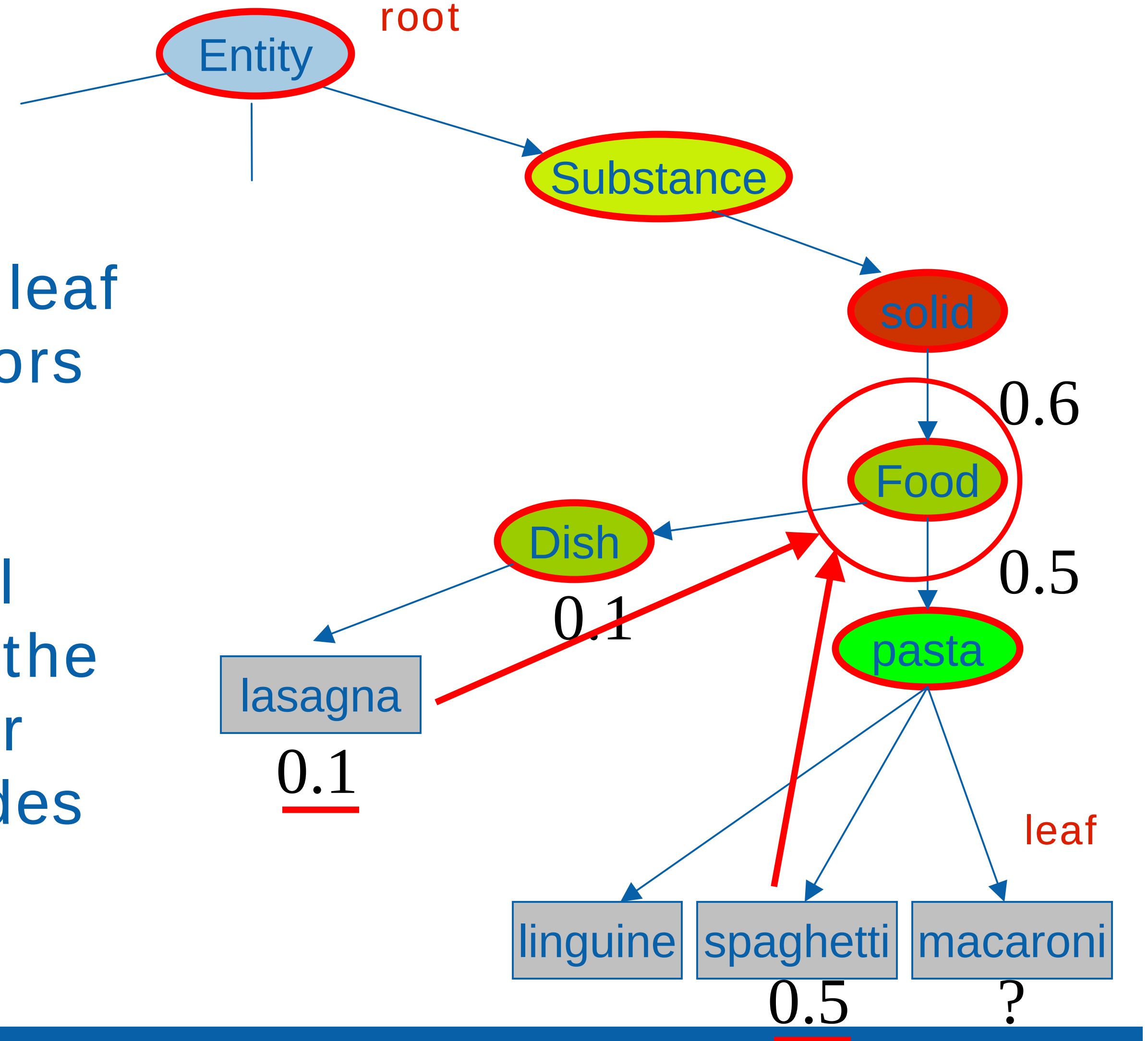
Shrinkage: Step two



Answer: Propagate information from leaf nodes to ancestors

For all internal nodes, compute the sum of all their children leaf nodes

Shrinkage: Step two

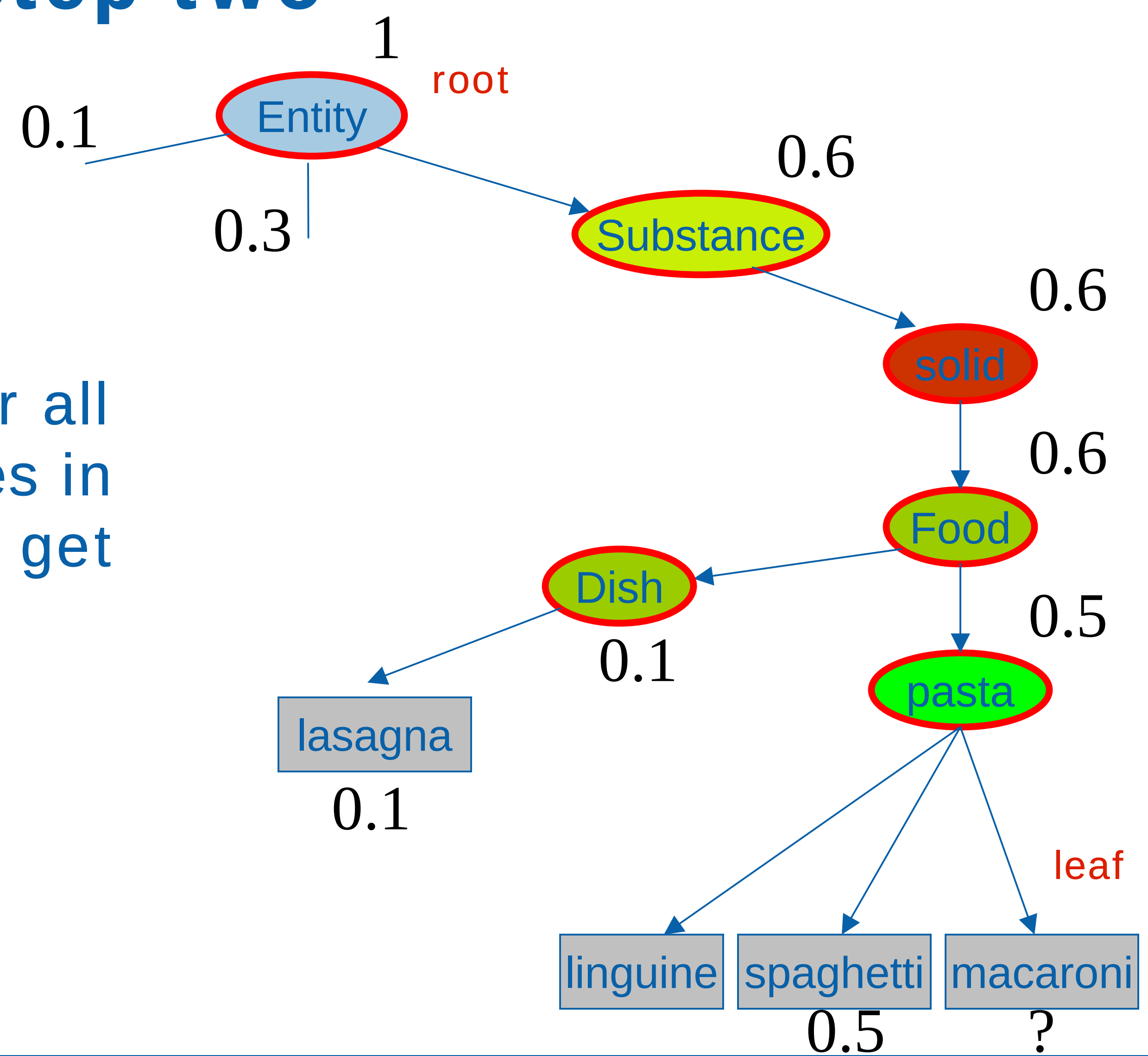


Propagate information from leaf nodes to ancestors

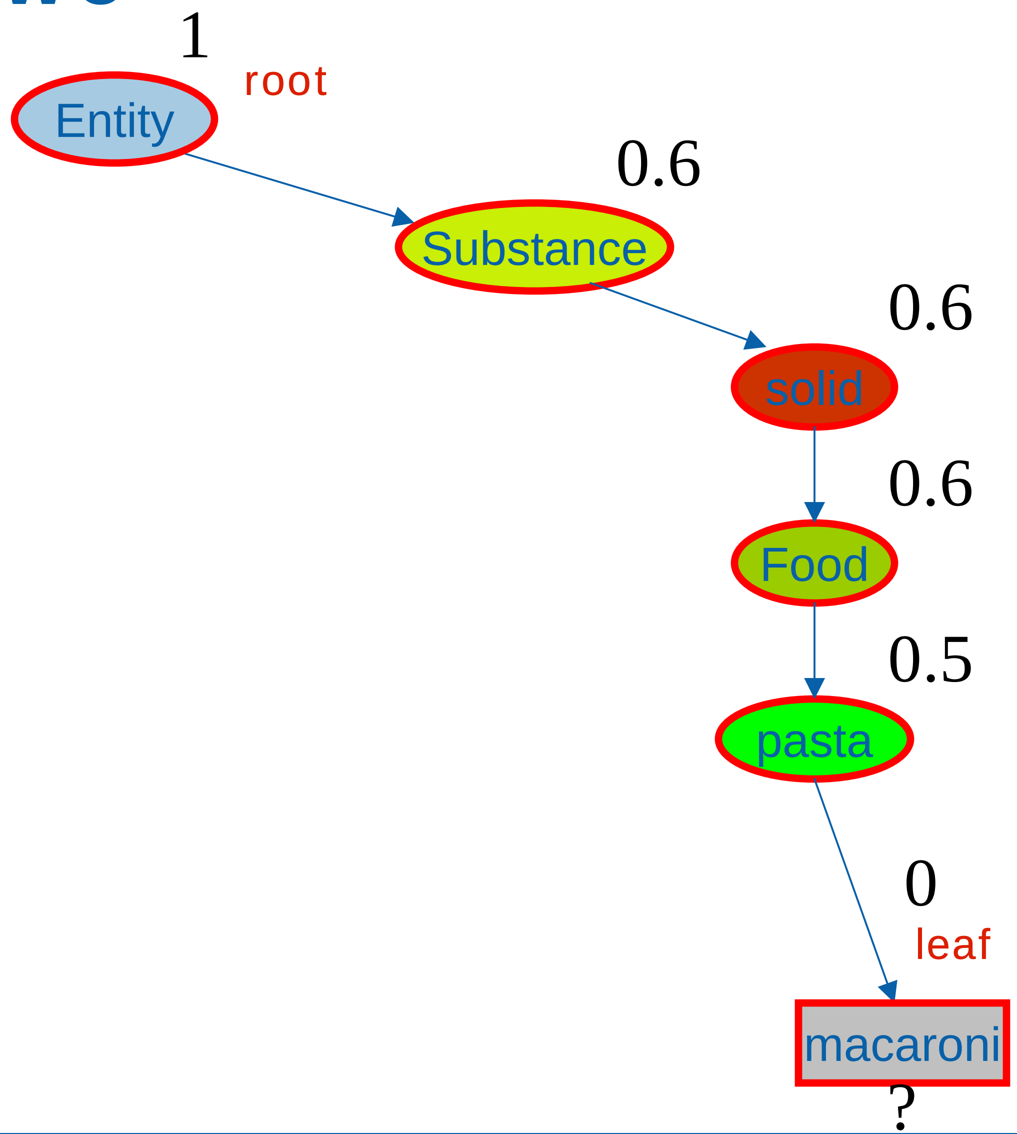
For all internal nodes, compute the sum of all their children leaf nodes

Shrinkage: Step two

Repeating for all internal nodes in ontology, we get



Shrinkage: Step two

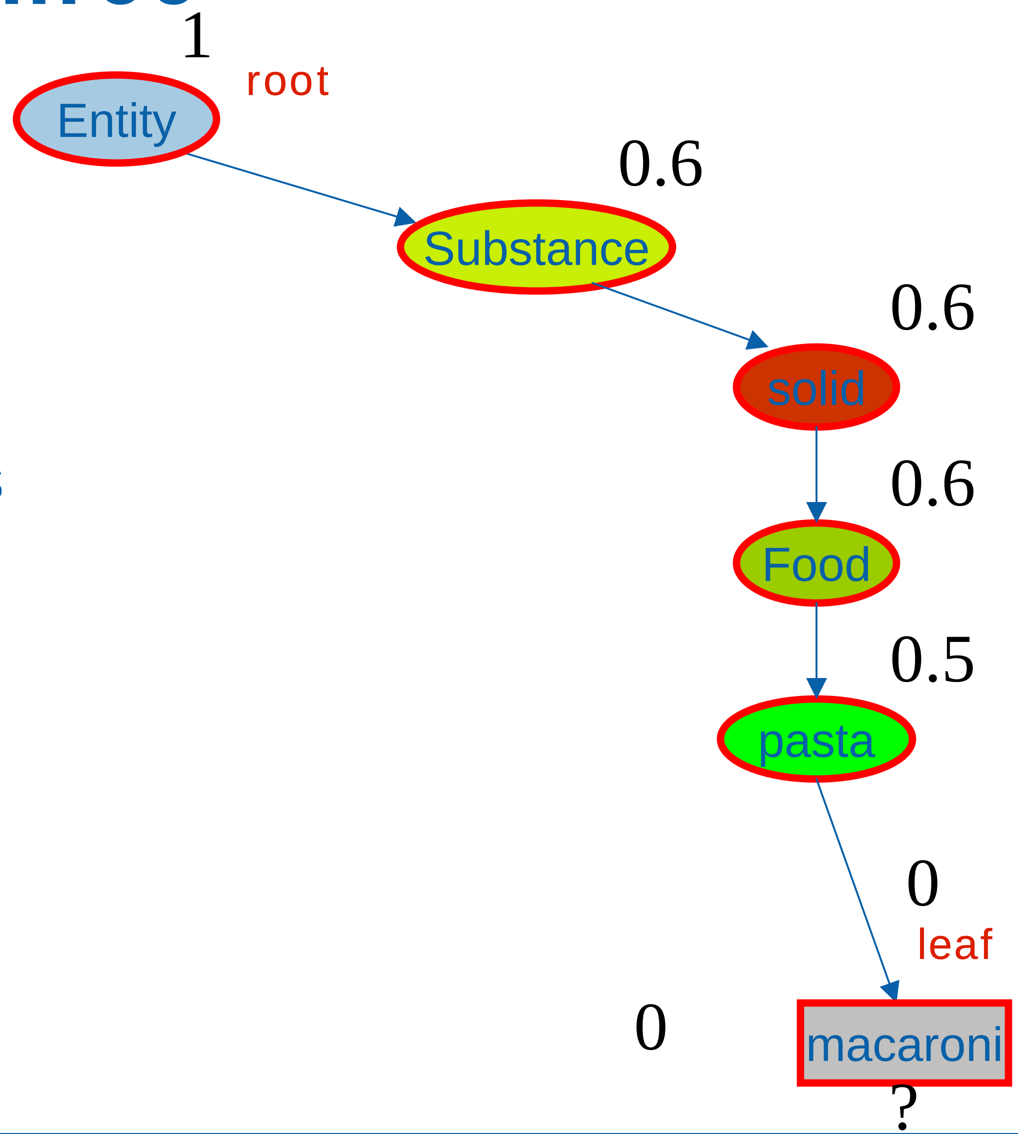


Now, we just focus on the nodes from the leaf to the root

$$\tilde{p}(macaroni) = ?$$

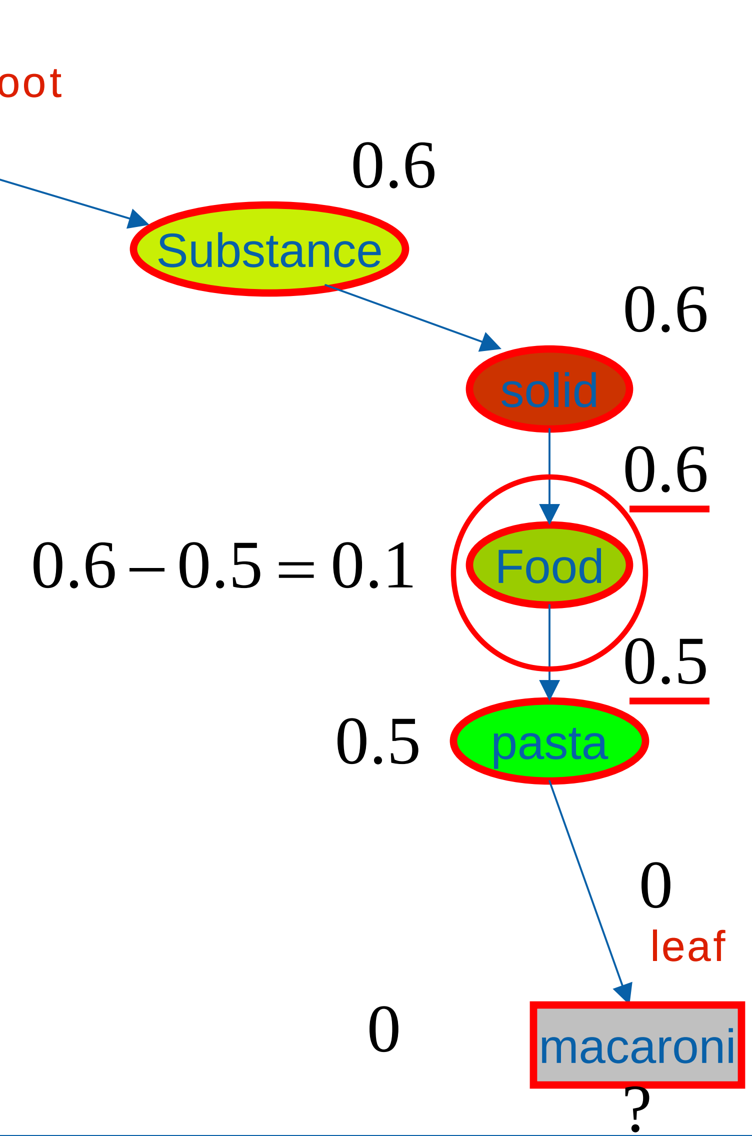
Shrinkage: Step three

Reduce dependencies by subtract counts of consecutive nodes

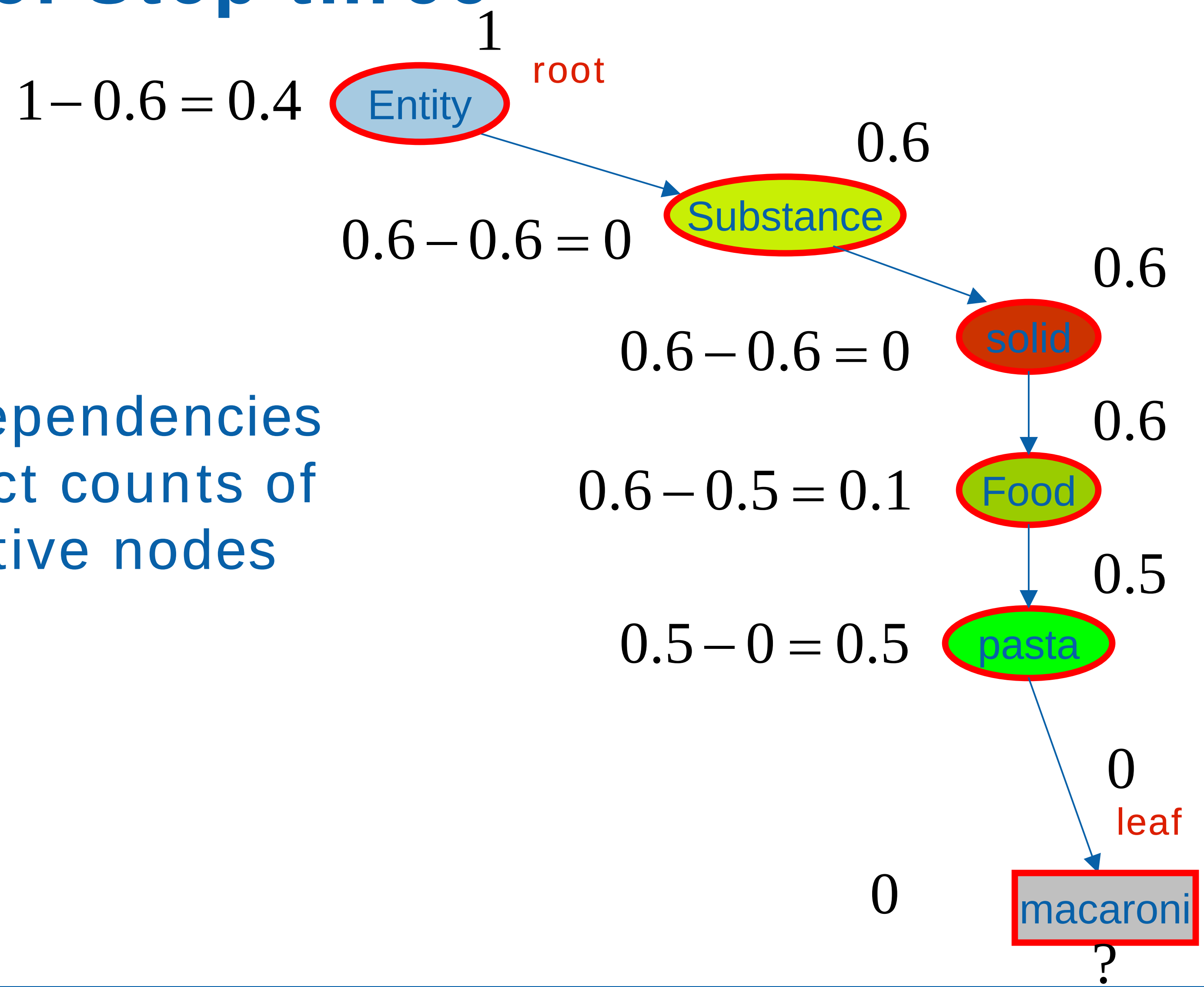


Shrinkage: Step three

Reduce dependencies by subtract counts of consecutive nodes

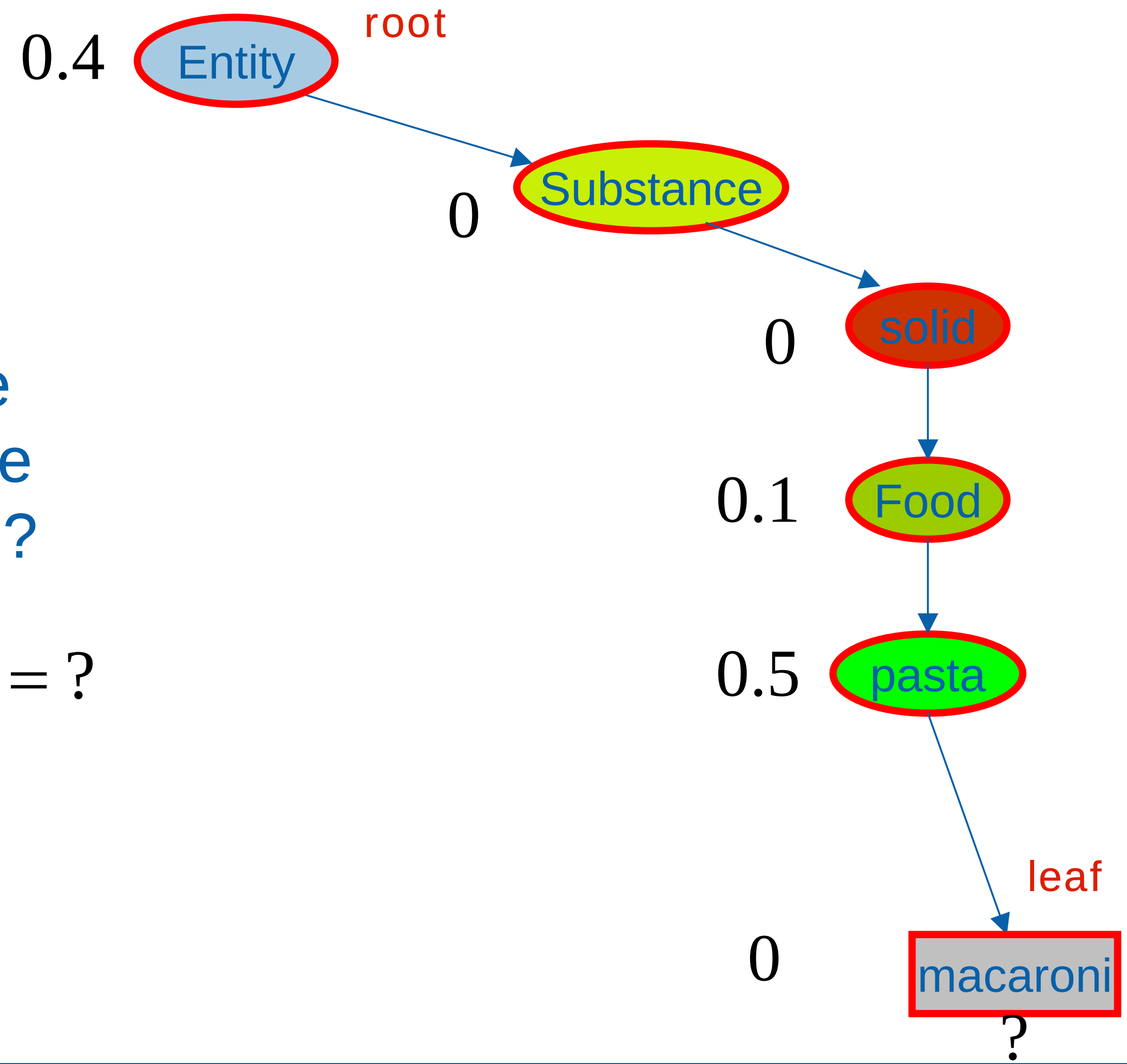


Shrinkage: Step three



Reduce dependencies by subtract counts of consecutive nodes

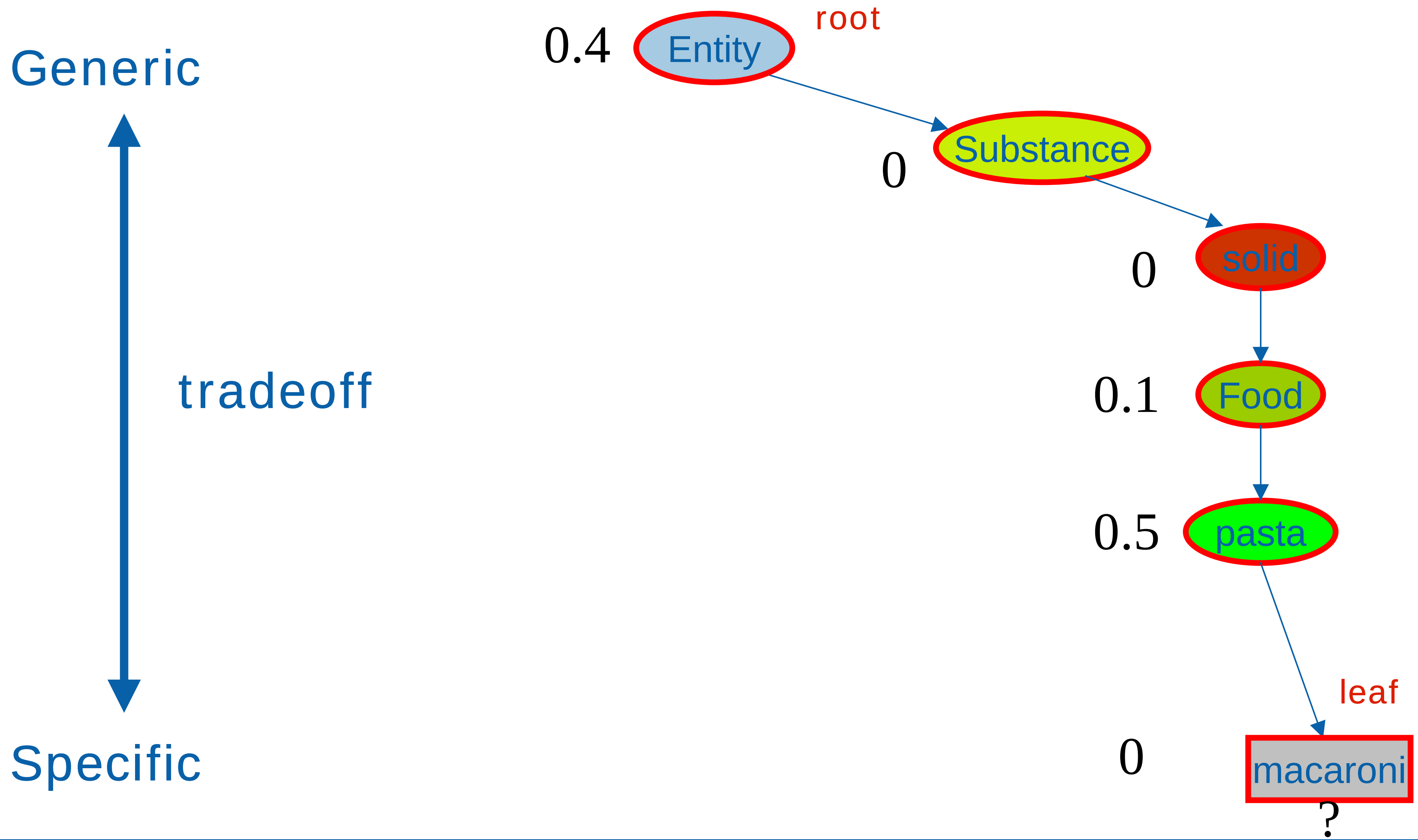
Shrinkage: Information tradeoff



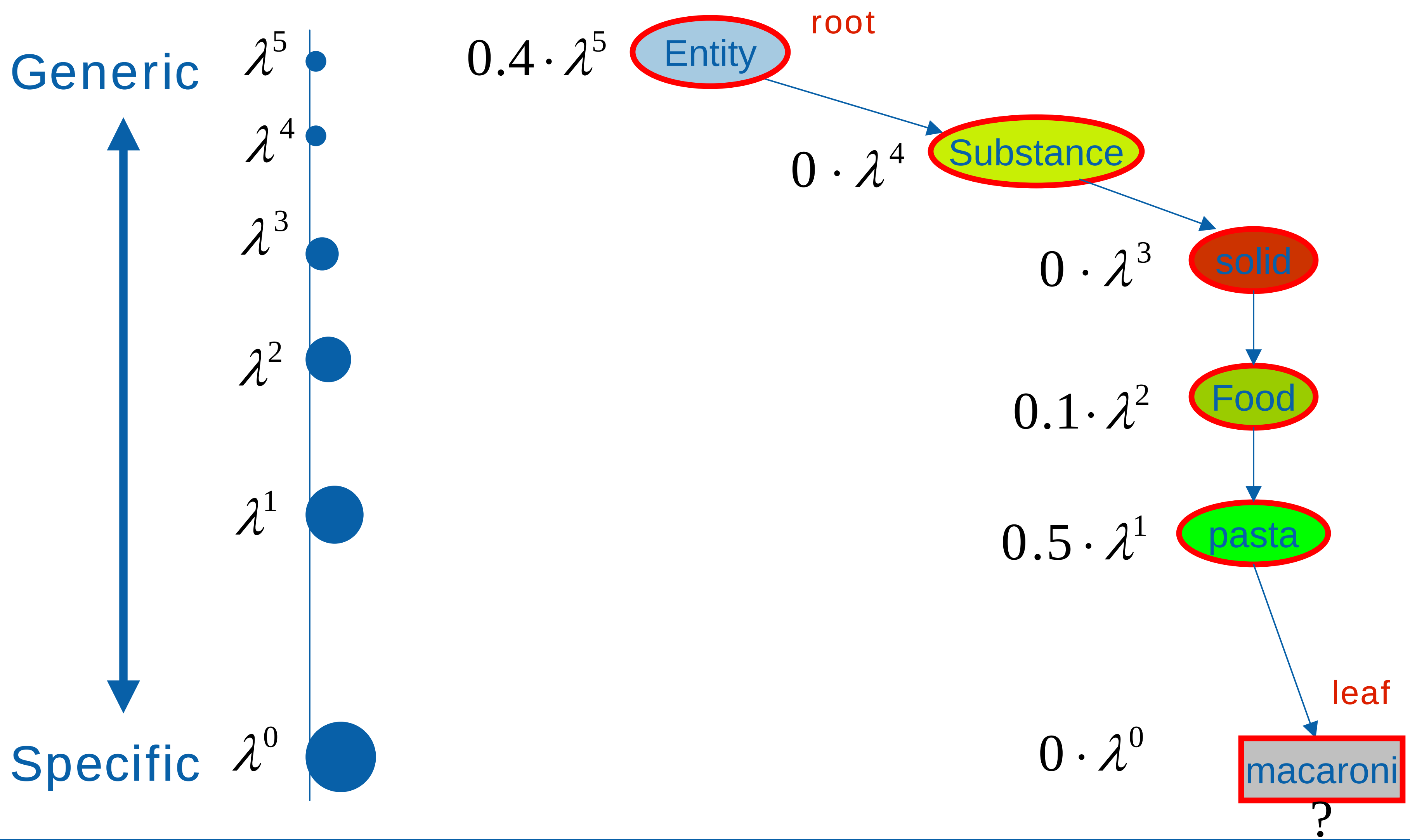
How do we combine the information?

$$\tilde{p}(macaroni) = ?$$

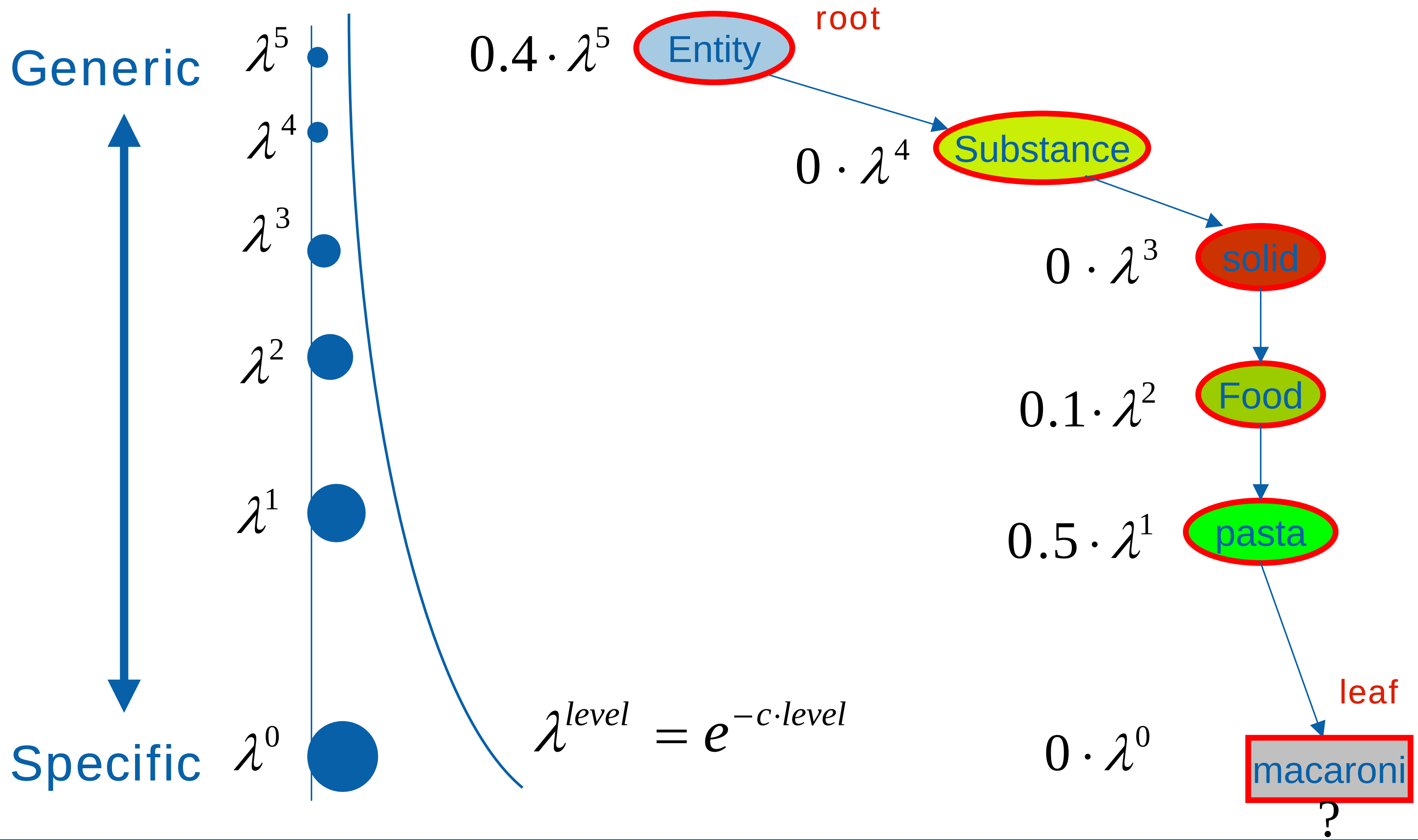
Shrinkage: Information tradeoff



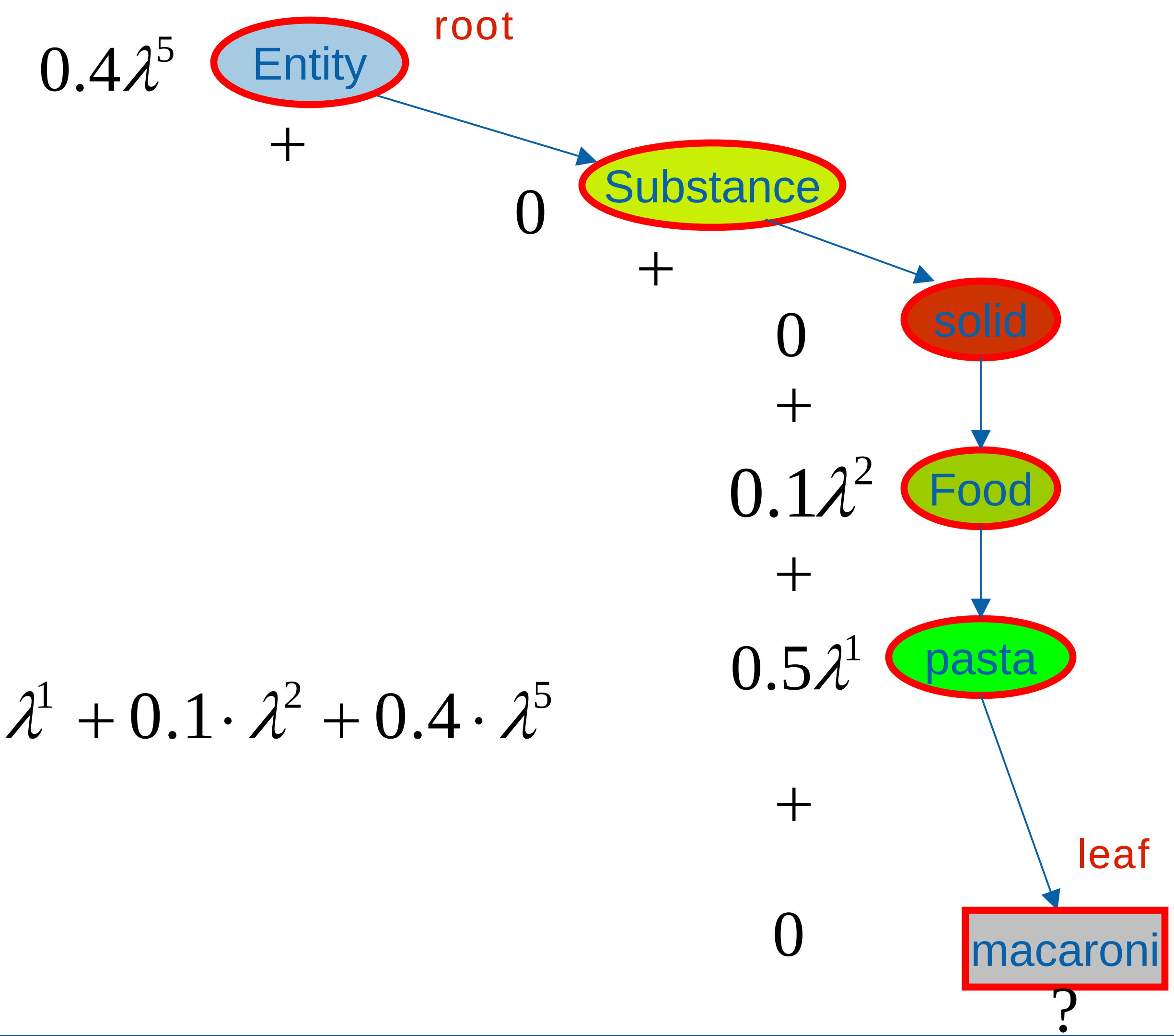
Shrinkage: Information tradeoff



Shrinkage: Information tradeoff



Shrinkage: Information tradeoff



$$\tilde{p}(macaroni) = 0.5 \cdot \lambda^1 + 0.1 \cdot \lambda^2 + 0.4 \cdot \lambda^5$$

Advantages of shrinkage

1. Create improved probability estimates for the leaf nodes (objects) of the ontology

The effect of this improvement is a reduction in the number of training examples required to achieve a desired accuracy

2. Compute probabilities for objects not present in the models

The effect is robustness when objects not present in the activity models are used while performing an activity

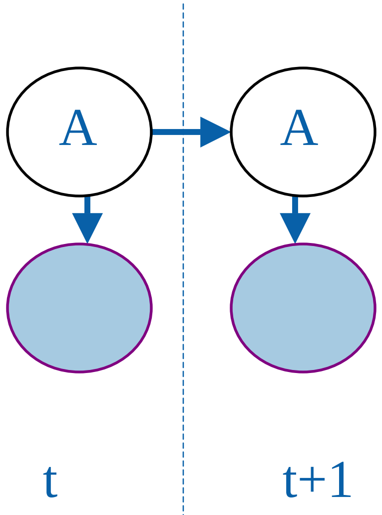
Performance on data collected from multiple individuals

- installed **108 RFIDs in real home**
- **9 subjects**
- 126 examples of 26 activities
- Using RFID glove reader
- **Web mined activity models web**

Ontology: generated from 108 tagged objects and objects present in web mined activity models.

Given trace, infer activities

Activities modeled by an HMM



model#1

Pr(object | activity)

- Single state per activity (26)
- Uniform prior
- Self-transition for smoothing
- Uniform inter-state transition

$$\Pr(a_i) = \frac{1}{26} = 0.04$$

$$\Pr(a_i \mid a_i) = T_{ii} = 0.8$$

$$\Pr(a_i \mid a_j) = T_{ij} = \frac{1-0.8}{n-1} = \frac{0.2}{25} = 0.01$$

$$P(o \mid a) = \text{mined from web}$$

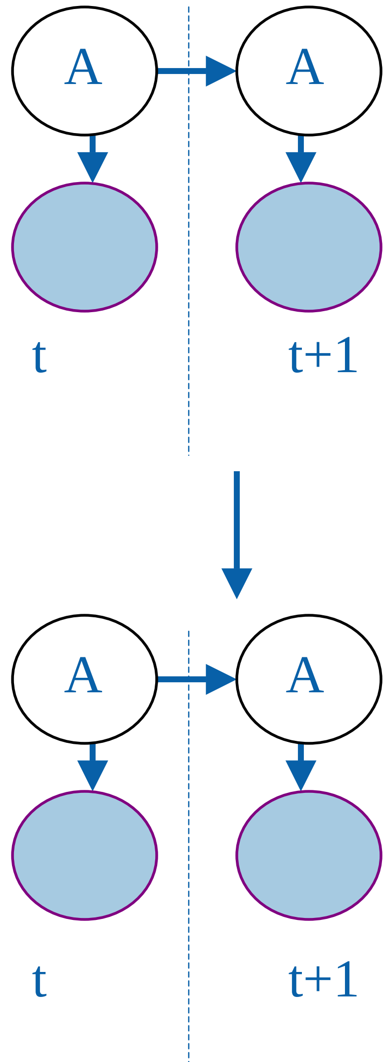
T

	A1	A2	…	A26
A1	0.8	0.01	…	0.01
A2	0.01	0.8	…	0.01
⋮				
A26	0.01	0.01	…	0.8

Activities modeled by an HMM

model#1

model#2

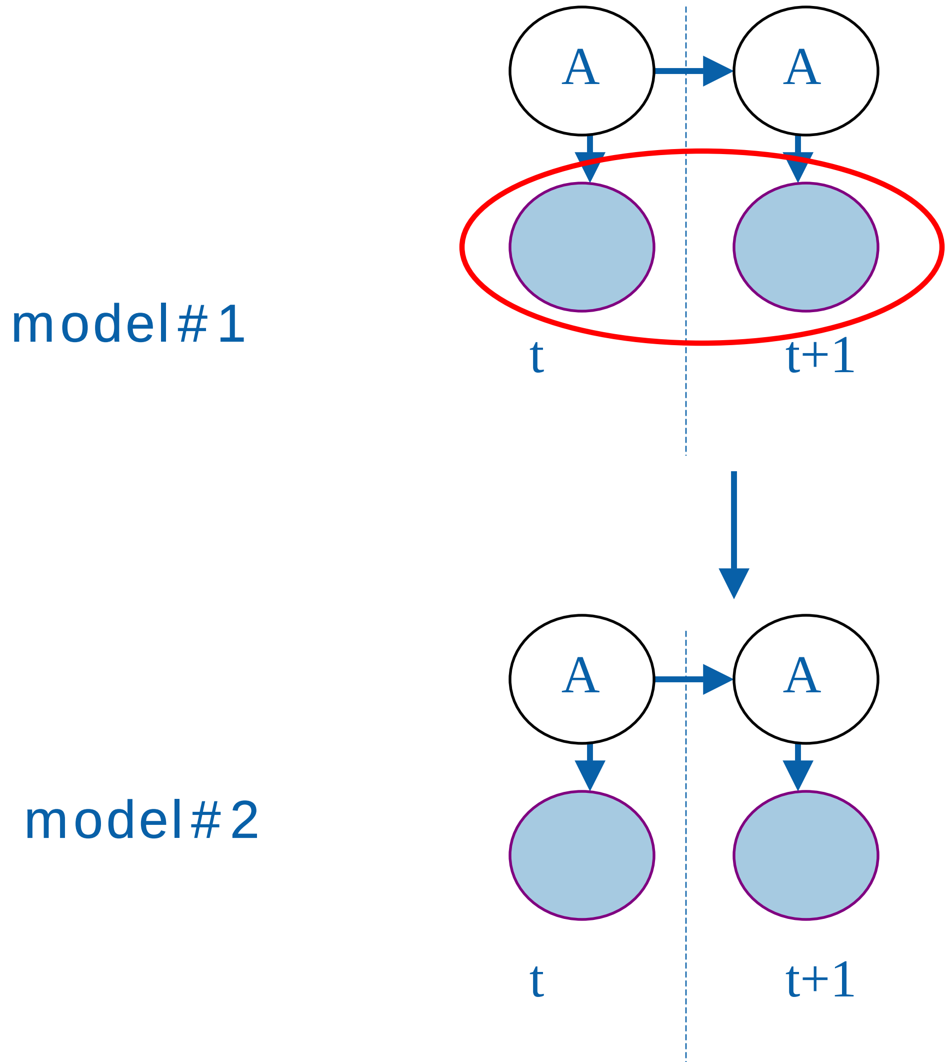


Shrinkage over observation matrix using

$$H1 = \lambda^{level} = e^{-c \cdot level}$$

$$H2 = \lambda^{level} = 1/c^{level}$$

Activities modeled by an HMM



Shrinkage over observation matrix using

$$H1 = \lambda^{level} = e^{-c \cdot level}$$

$$H2 = \lambda^{level} = 1/c^{level}$$

Experiment 1: Improvement of overall accuracy

Assemble all the activity examples in a single sequence

Infer the most likely state (activity) sequence using Viterbi decoding

Compute total accuracy

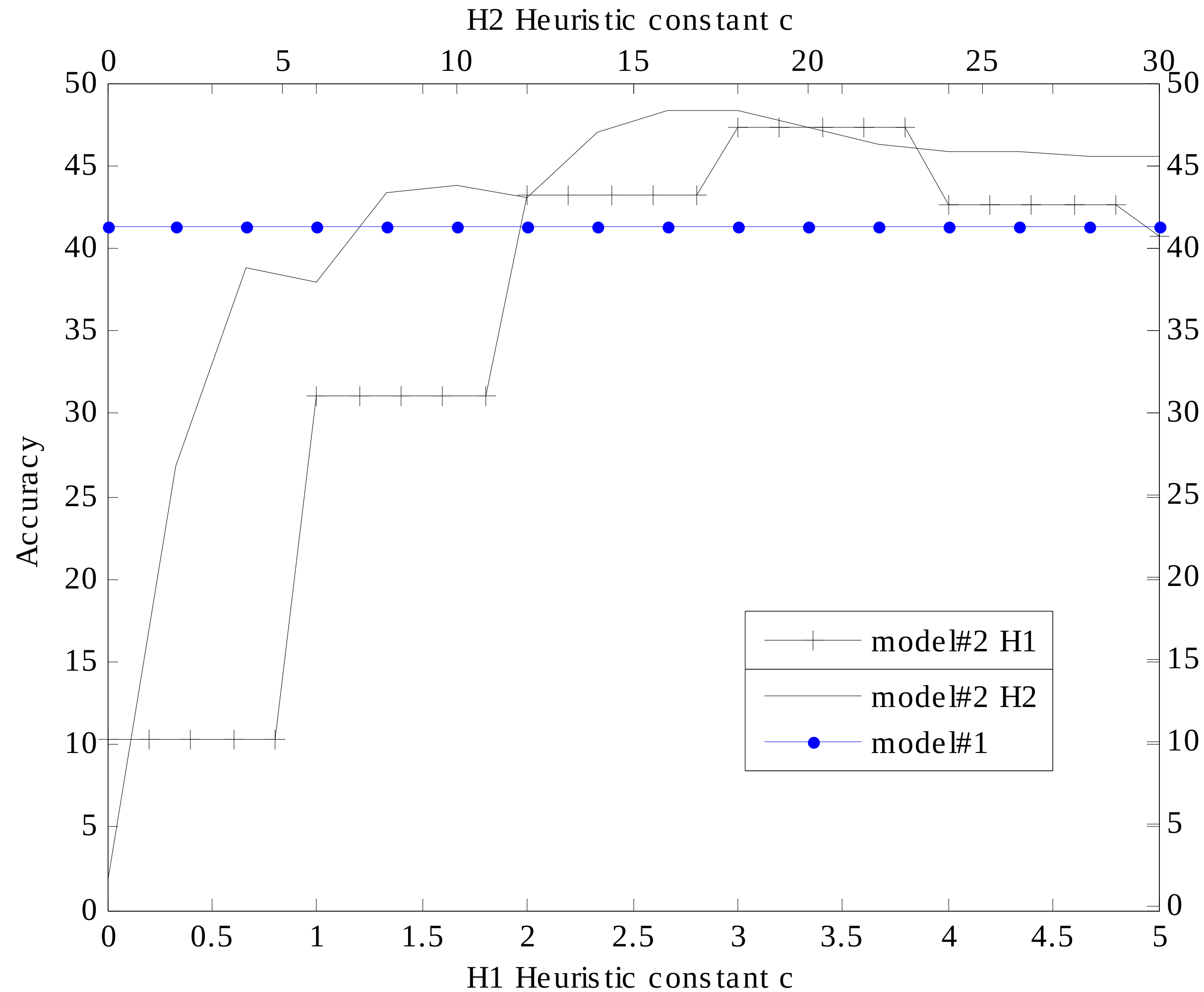
Shrinkage: Learning the weights

No shrinkage: 42%

With shrinkage: 48.35%

Improvement: 15.11%

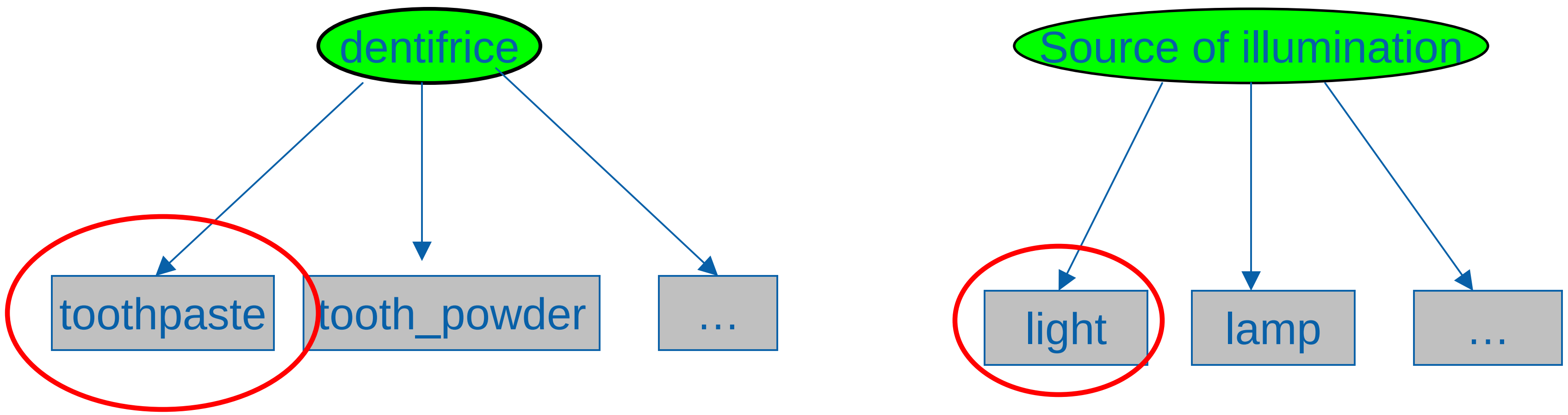
Bootstrapping by learning using 126 sensor traces: 19.2%



Experiment 2: Robustness to unseen observations

Replace *m*% of the observations in the activity examples by observations of one of their randomly selected sibling nodes in the ontology

Brushing teeth	Original:	light	toothpaste	floss	light
	Replaced:	light	tooth_powder	floss	lamp



Results: Robustness to unseen observations

When replacing 100%

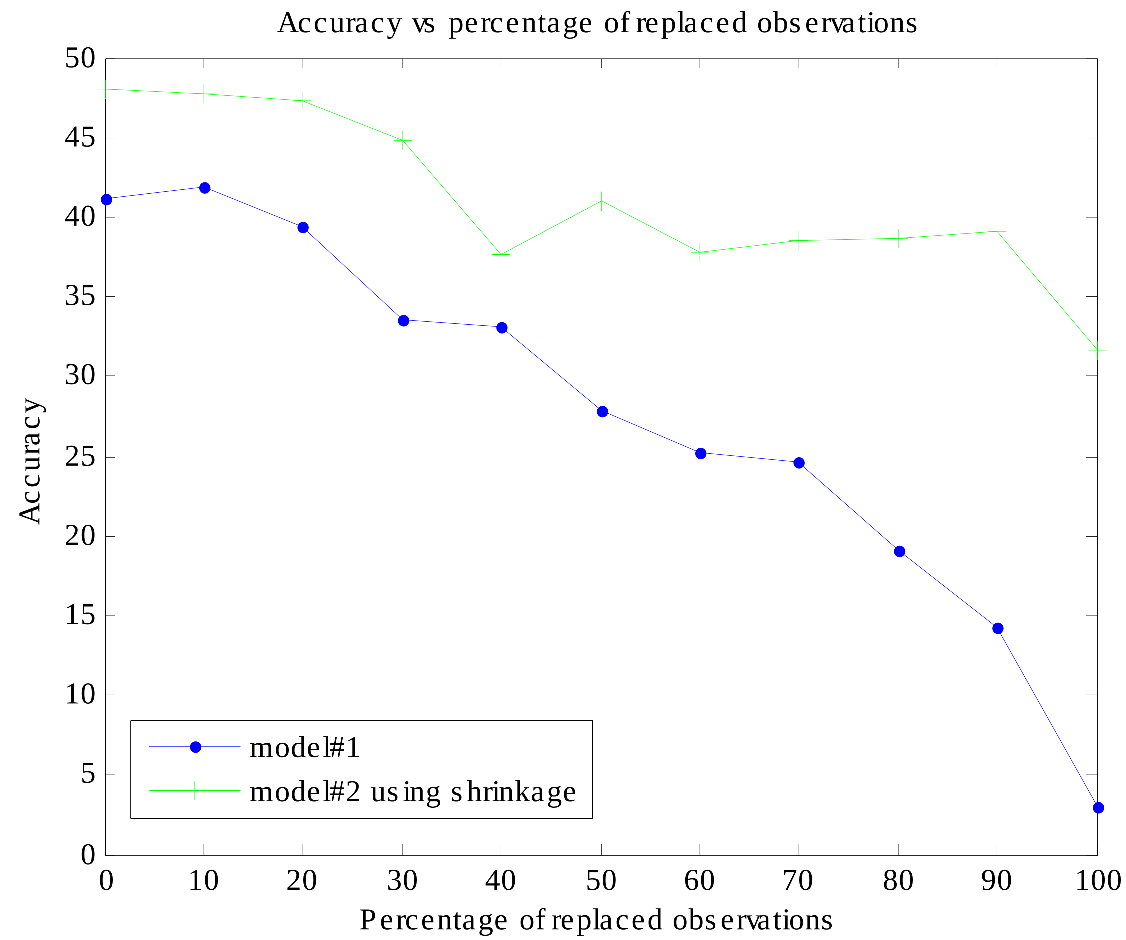
No shrinkage: Randomly guessing

3.8%

Drops 91.6%

Shrinkage:

Drops 33%



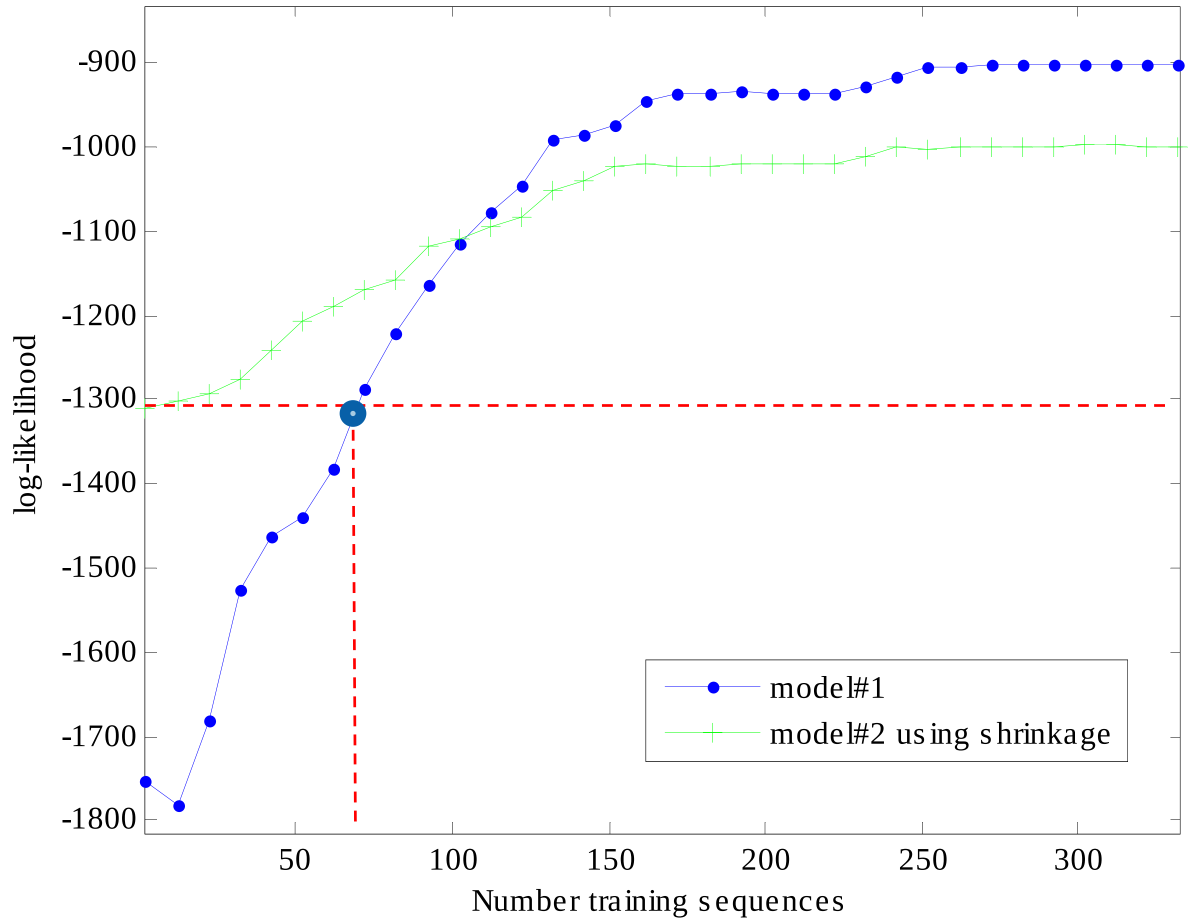
Effect of limited training data by simulation

Simulations were performed to investigate the impact of shrinkage with limited training data

The ontology

- was generated from a list of 815 objects
- consists of 4188 nodes
- 815 leaf nodes
- has a maximum depth of 14.

Experiment: Effect of limited training data



Conclusion

- Previous work demonstrated that it is possible to mine useful models of *arbitrary* day-to-day activities from the web
- Here we show that it is possible to deal with model incompleteness by incorporating common sense knowledge
- we compute probabilities for objects not originally present in the models
- We can improve the probability estimates
- we can learn higher quality models with less amount of training data
- Towards a completely unsupervised approach to learning activity models

Thank you!