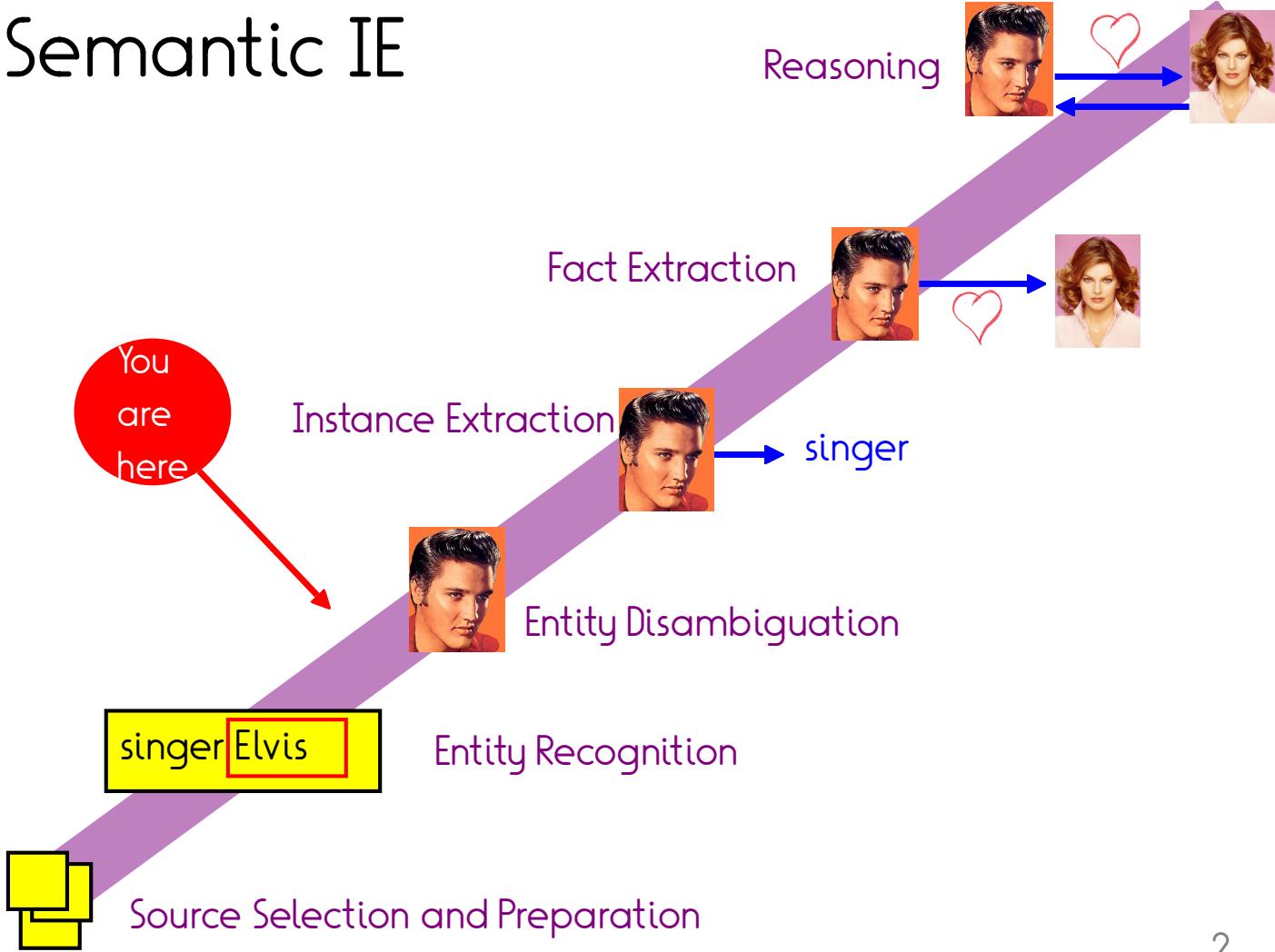


Disambiguation

Nada Mimouni

Based on slides by:
Fabian M. Suchanek

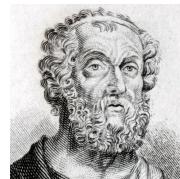
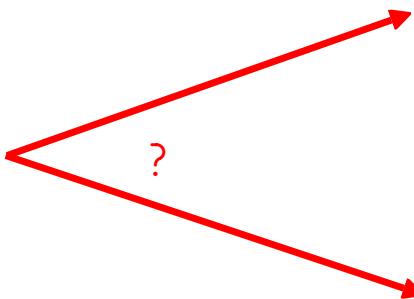
Semantic IE



Def: Disambiguation

Given an ambiguous name in a corpus and its meanings, **disambiguation** is the task of determining the intended meaning.

Homer eats
a doughnut.



This year, for our 175th anniversary, we have looked over copies of *The Economist* Past to see how what the paper covered has changed over the decades. The results (more of which are available online at economist.com/pastpeople) were not as clear-cut as we had hoped. Some of this is down to the limitations of current software, some to our past style. Captains of industry like Andrew Carnegie or Henry Ford proved hard to separate from their namesake libraries, institutions and car companies. And we usually referred to Napoleon III of France, for example, simply as "The Emperor"—the same way we referred to the tsars or kaisers of their days, not to mention, more recently, large penguins. As we became more

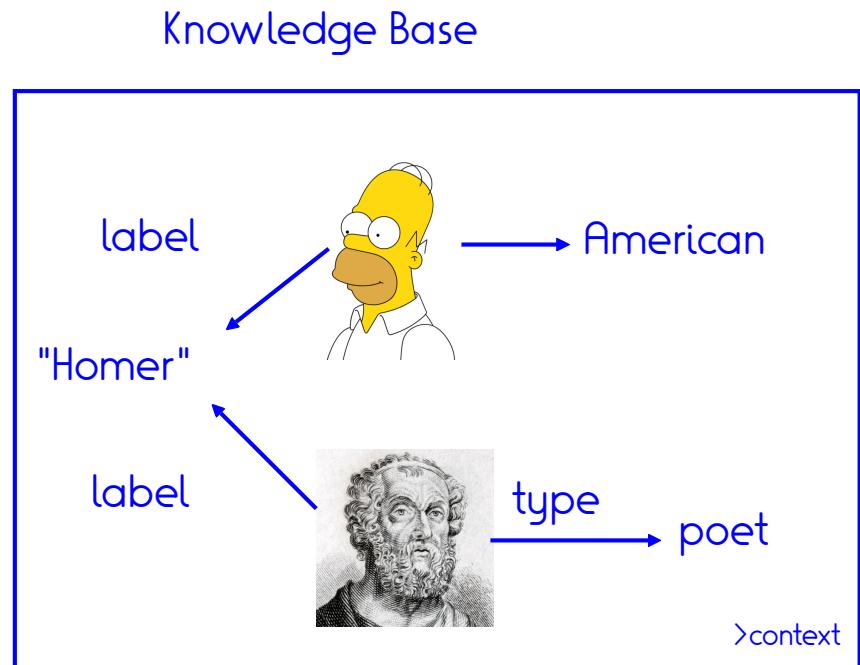
[The Economist, 2018-12-22]

Disambiguation Setting

Usually Named Entity Recognition (NER) runs first, and the goal is to map the names to entities in a Knowledge Base (KB).

NER'ed
corpus

Homer eats
a doughnut.



Def: Context of a word

The context of a word in a corpus is the multi-set of the words in its vicinity without the stopwords.

(The definition may vary depending on the application)

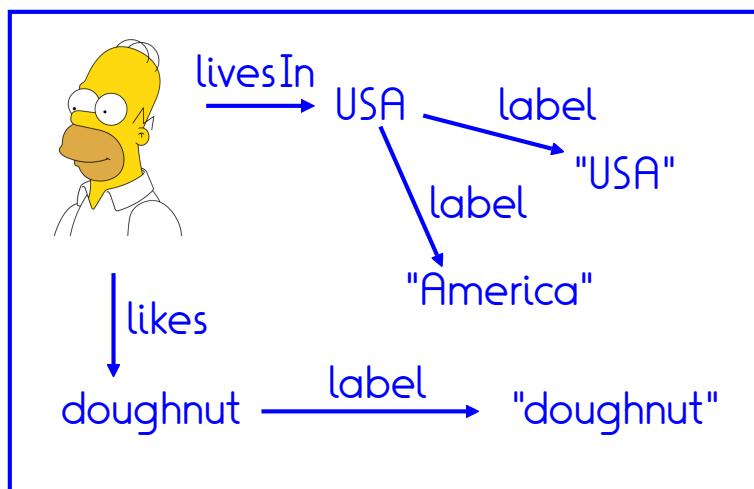
Homer eats
a doughnut.

Context of "Homer":
{eats, doughnut}

Def: Context of an entity

The context of an entity in a KB is the set of all labels of all entities in its vicinity.

(The definition may vary depending on the application)



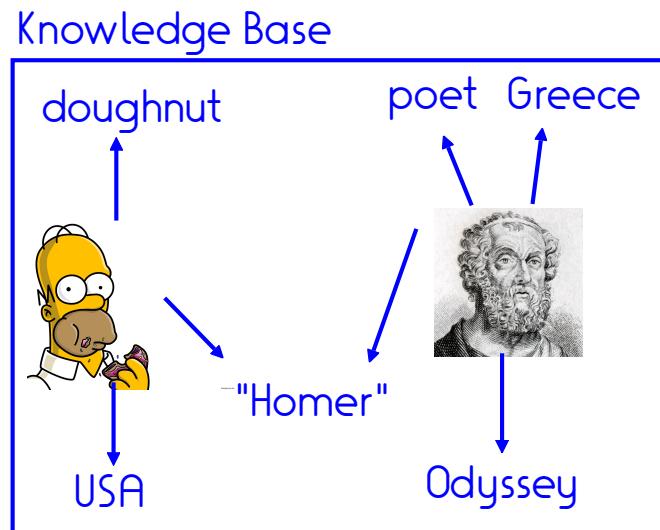
Context of Homer:
{doughnut, USA,
America}

Def: Context-based disambiguation

Context-based disambiguation (also: bag of words disambiguation)
maps a name in a corpus to the entity in the KB whose context
has the highest overlap to the context of the name.

(The definition may vary depending on the application)

For USA Today, Homer is among the top 25 most influential people of the past 25 years.

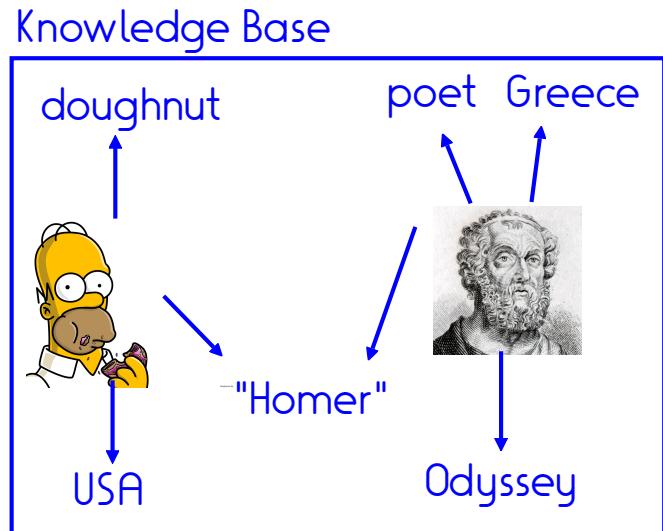


Def: Context-based disambiguation

Context-based disambiguation (also: bag of words disambiguation)
maps a name in a corpus to the entity in the KB whose context
has the highest overlap to the context of the name.

For USA Today, Homer is among the top 25 most influential people of the past 25 years.

Context of "Homer" in corpus:
{USA, Today, top, influential,
people, past, years}

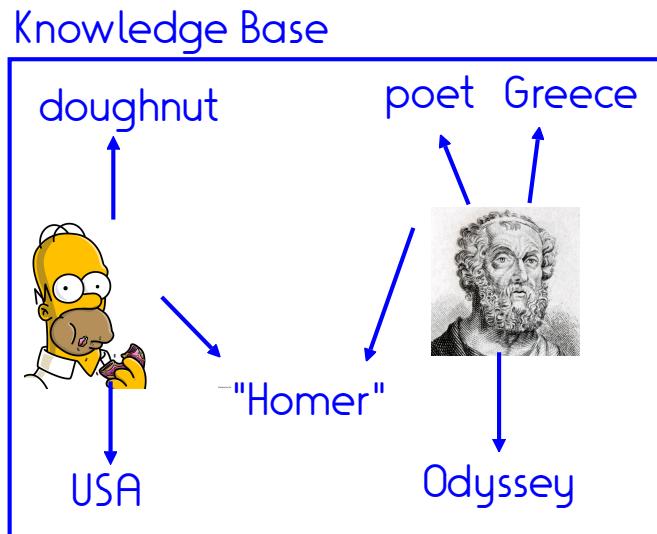


Def: Context-based disambiguation

Context-based disambiguation (also: bag of words disambiguation)
maps a name in a corpus to the entity in the KB whose context
has the highest overlap to the context of the name.

For USA Today, Homer is among the top 25 most influential people of the past 25 years.

Context of "Homer" in corpus:
{USA, Today, top, influential,
people, past, years}



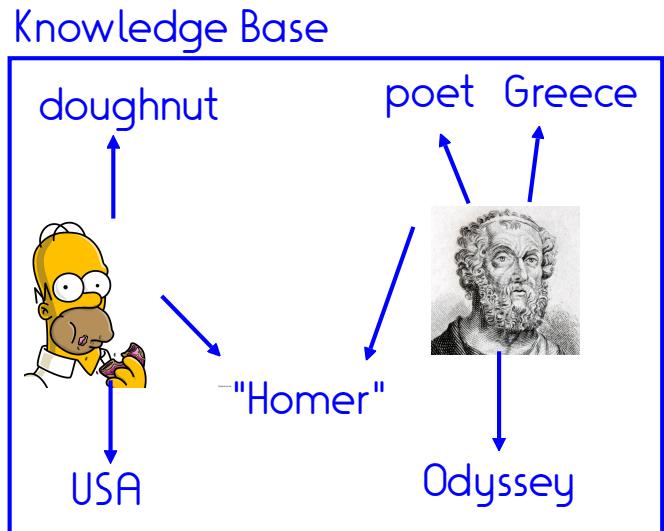
{USA, doughnut} {poet, Geece, 0.}

Def: Context-based disambiguation

Context-based disambiguation (also: bag of words disambiguation)
maps a name in a corpus to the entity in the KB whose context
has the highest overlap to the context of the name.

For USA Today, Homer is among the top 25 most influential people of the past 25 years.

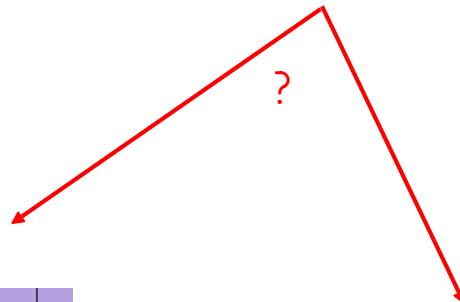
Context of "Homer" in corpus:
{USA, Today, top, influential,
people, past, years}



highest overlap -> Homer Simpson wins

What if there is little context?

This is very important for the Simpsons.

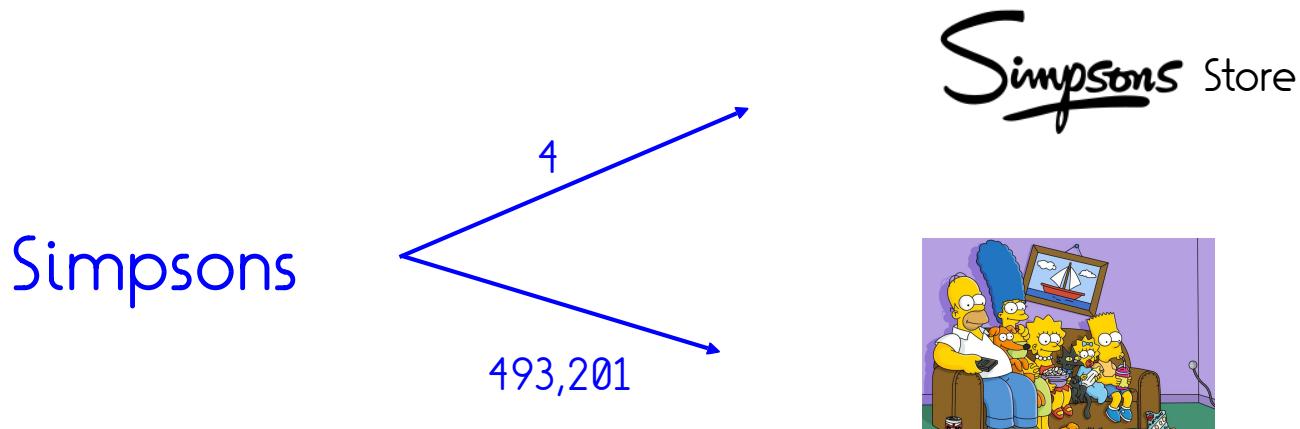


Simpsons

The Robert Simpson
Department Store.
Defunct since 1990.

Def: Disambiguation Prior

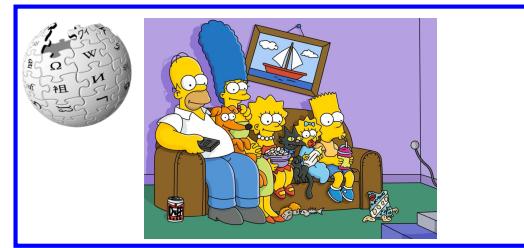
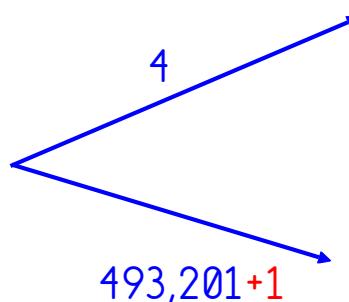
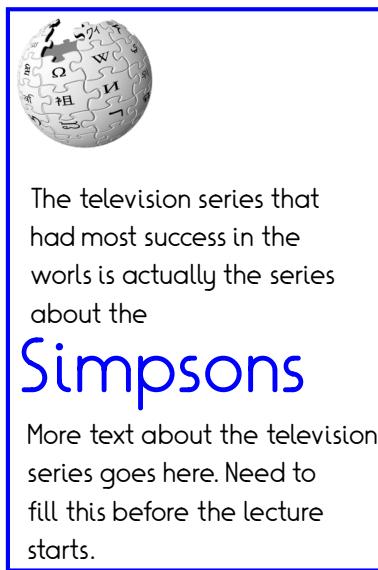
A **disambiguation prior** is a mapping from names to their meanings, weighted by the number of times that the name refers to the meaning in a reference corpus.



Def: Disambiguation Prior

A **disambiguation prior** is a mapping from names to their meanings, weighted by the number of times that the name refers to the meaning in a reference corpus.

It can be computed, e.g., from Wikipedia:



Coherence

Bart and Homer accidentally launch a rocket into the Springfield church, causing Lisa to leave Christianity.

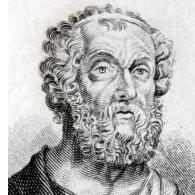
?



?



?



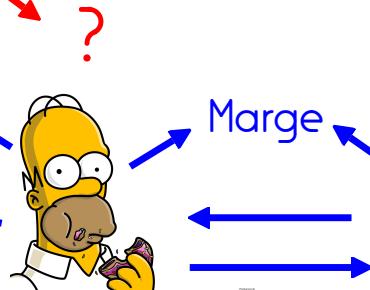
Coherence

Bart and Homer accidentally launch a rocket into the Springfield church, causing Lisa to leave Christianity.

?



Simpson
dumb



Marge



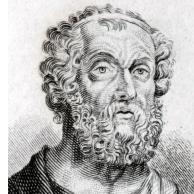
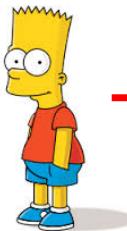
?



Def: Coherence Criterion

The **Coherence Criterion** postulates that entities that are mentioned in one document should be related in the KB.

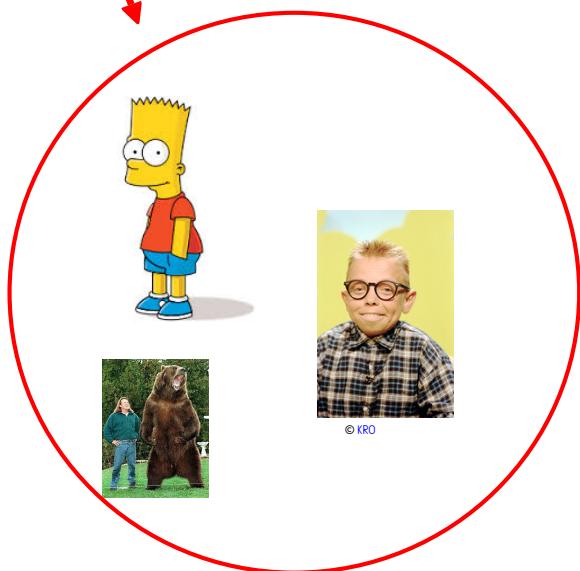
Bart and Homer accidentally launch a rocket into the Springfield church, causing Lisa to leave Christianity.



>random-walk

Implementation

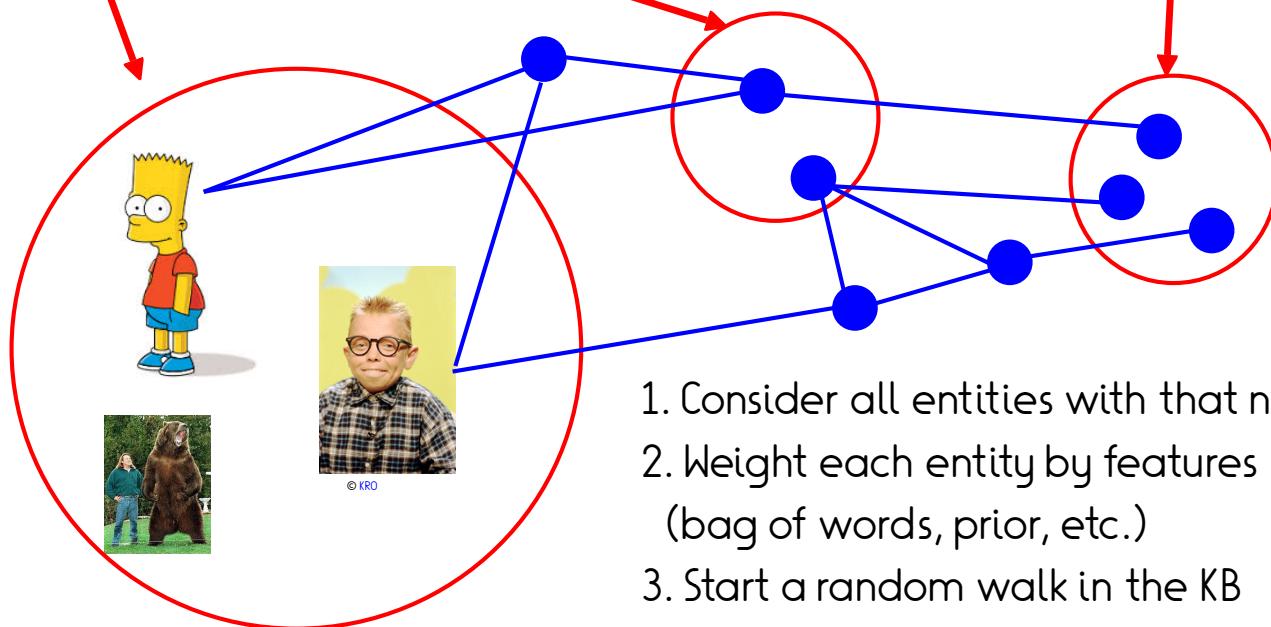
Bart and Homer accidentally launch a rocket into the Springfield church, causing Lisa to leave Christianity.



1. Consider all entities with that name
2. Weight each entity by features
(bag of words, prior, etc.)

Implementation

Bart and Homer accidentally launch a rocket into the Springfield church, causing Lisa to leave Christianity.



Example: Disambiguation by AIDA

AIDA is a system for the disambiguation of entity names, based on YAGO.



Try it out

Example: Disambiguation by AIDA

Disambiguation Method:

[prior](#) [prior+sim](#) [prior+sim+coherence](#)

Parameters: (defaults should be OK)

Prior-Similarity-Coherence balancing ratio:

prior VS. sim. balance = [0.4](#)

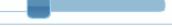
(prior+sim.) VS. coh. balance [0.6](#)



Ambiguity degree [7](#)



Coherence robustness test threshold: [0.9](#)



Entities Type Filters:

Enter the types her

Mention Extraction:

[Stanford NER](#) [Manual](#)

You can manually tag the mentions by putting them between [[and]].

HTML Tables are automatically disambiguated in the manual mode.



Lisa, Bart, and Homer all love the
mother of the house, Marge.

Input Type: TEXT Overall runtime: 43s, 78ms

[Types list](#)

[Types tag cloud](#)

[Focused Ty](#)

[Lisa Simpson] [Lisa](#), [Bart Simpson] [Bart](#), and Homer all love the mother of the house, [Marge Simpson] [Marge](#).

Summary: Disambiguation

We saw 3 indicators for disambiguation:

1. Context

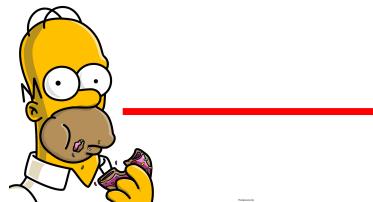
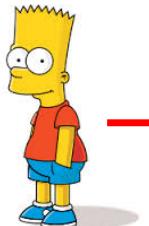
Homer eats a doughnut.

2. Disambiguation prior



> Simpsons

3. Coherence



References

Mohamed Amir Yosef, Johannes Hoffart, Ilaria Bordino, Marc Spaniol, Gerhard Weikum:
“AIDA: An Online Tool for Accurate Disambiguation”, VLDB 2011

->instance-extraction