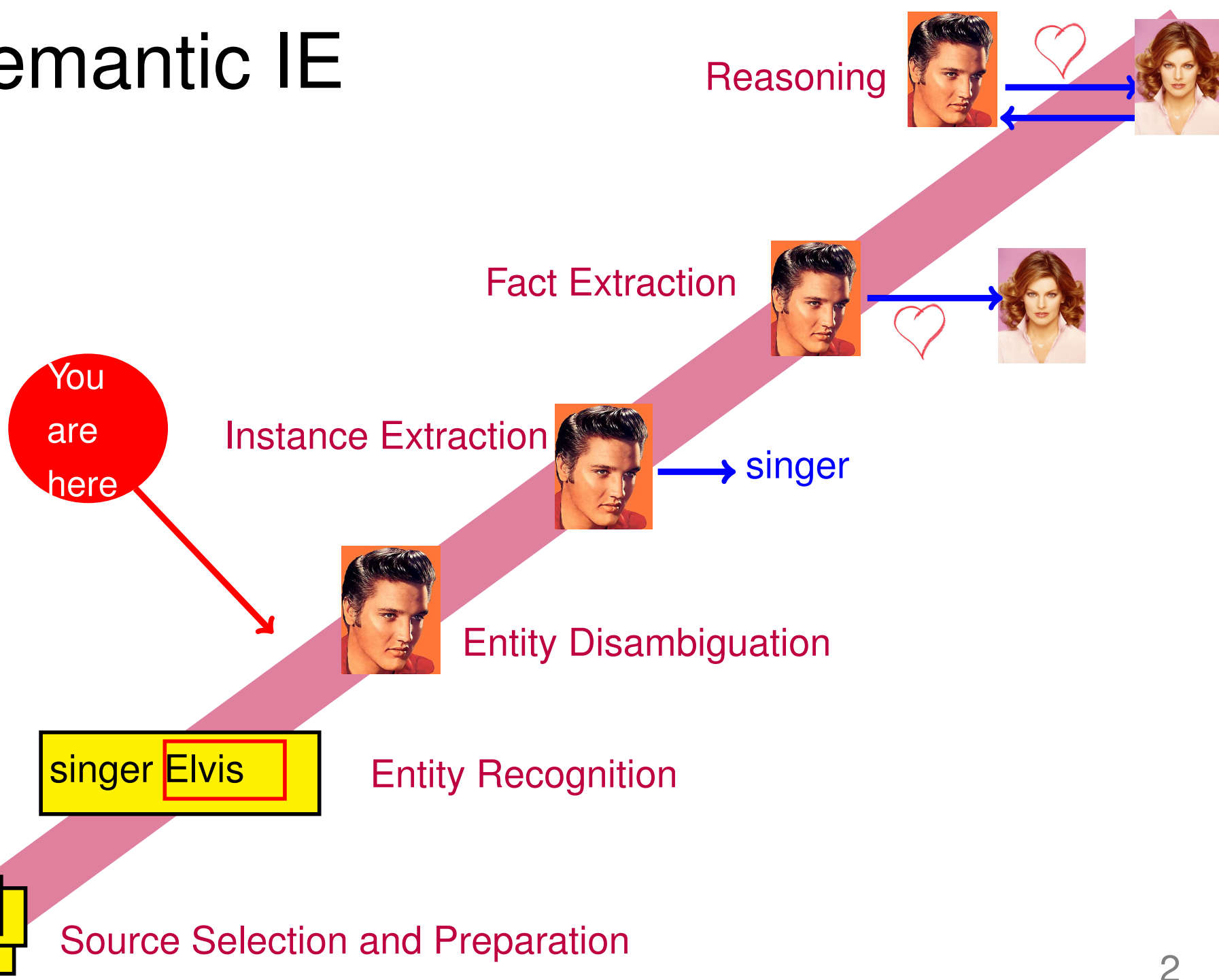


Evaluation

Fabian M. Suchanek

Semantic IE



Detect members of the Simpsons

in The Simpsons, Homer Simpson is the father of Bart Simpson and Lisa Simpson. The M above his ear is for Matt Groening.



Pixelartzen.de

Detect members of the Simpsons

in The Simpsons, Homer Simpson is the father of Bart Simpson and Lisa Simpson. The M above his ear is for Matt Groening.

1. `[A-Z][a-z]+ Simpson`



Prologarten.de

Detect members of the Simpsons

in The Simpsons, Homer Simpson is the father of Bart Simpson and Lisa Simpson. The M above his ear is for Matt Groening.

1. `[A-Z][a-z]+ Simpson`

4 matches (1 wrong)



Prologarten.de

Detect members of the Simpsons

in The Simpsons, Homer Simpson is the father of Bart Simpson and Lisa Simpson. The M above his ear is for Matt Groening.

1. $[A-Z][a-z]^+$ Simpson

4 matches (1 wrong)

2. $[A-Z][a-z]^+ [A-Z][a-z]^+$



Pixelart.de

Detect members of the Simpsons

in The Simpsons, Homer Simpson is the father of Bart Simpson and Lisa Simpson. The M above his ear is for Matt Groening.

1. $[A-Z][a-z]^+$ Simpson

4 matches (1 wrong)

2. $[A-Z][a-z]^+ [A-Z][a-z]^+$

5 matches (2 wrong)



Pixelartzen.de

Detect members of the Simpsons

in The Simpsons, Homer Simpson is the father of Bart Simpson and Lisa Simpson. The M above his ear is for Matt Groening.

1. $[A-Z][a-z]^+$ Simpson

4 matches (1 wrong)

2. $[A-Z][a-z]^+ [A-Z][a-z]^+$

5 matches (2 wrong)

3. Homer Simpson



Detect members of the Simpsons

in The Simpsons, Homer Simpson is the father of Bart Simpson and Lisa Simpson. The M above his ear is for Matt Groening.

1. $[A-Z][a-z]^+$ Simpson

4 matches (1 wrong)

2. $[A-Z][a-z]^+ [A-Z][a-z]^+$

5 matches (2 wrong)

3. Homer Simpson

1 match



Def: Gold Standard

The gold standard (also: ground truth) for an IE task is the set of desired results of the task on a given corpus.

Task: Detect Simpson members

Corpus:

in The Simpsons, Homer Simpson is the
father of Bart Simpson and Lisa Simpson.
The M above his ear is for Matt Groening.

Gold Standard:

{Homer Simpson, Bart Simpson, Lisa Simpson}

Def: Precision

The precision of an IE algorithm is the ratio of its outputs that are in the respective gold standard.

$$prec = \frac{|Output \cap GStandard|}{|Output|}$$

Output: {Homer, Bart, Groening}

G.Standard: {Homer, Bart, Lisa, Marge}

Def: Precision

The precision of an IE algorithm is the ratio of its outputs that are in the respective gold standard.

$$prec = \frac{|Output \cap GStandard|}{|Output|}$$

Output: {Homer, Bart, Groening}



G.Standard: {Homer, Bart, Lisa, Marge}

Def: Precision

The precision of an IE algorithm is the ratio of its outputs that are in the respective gold standard.

$$prec = \frac{|Output \cap GStandard|}{|Output|}$$

Output: {Homer, Bart, Groening}
 ✓ ✓

G.Standard: {Homer, Bart, Lisa, Marge}

Def: Precision

The precision of an IE algorithm is the ratio of its outputs that are in the respective gold standard.

$$prec = \frac{|Output \cap GStandard|}{|Output|}$$

Output: {Homer, Bart, Groening}
 ✓ ✓ ✗

G.Standard: {Homer, Bart, Lisa, Marge}

Def: Precision

The precision of an IE algorithm is the ratio of its outputs that are in the respective gold standard.

$$prec = \frac{|Output \cap GStandard|}{|Output|}$$

Output: {Homer, Bart, Groening}
 ✓ ✓ ✗

G.Standard: {Homer, Bart, Lisa, Marge}

=> Precision: $2/3 = 66\%$

Def: Recall

The recall of an IE algorithm is the ratio of the gold standard that is output.

$$rec = \frac{|Output \cap GStandard|}{|GStandard|}$$

Output: {Homer, Bart, Groening}

G.Standard: {Homer, Bart, Lisa, Marge}

Def: Recall

The recall of an IE algorithm is the ratio of the gold standard that is output.

$$rec = \frac{|Output \cap GStandard|}{|GStandard|}$$

Output: {Homer, Bart, Groening}

G.Standard: {Homer, Bart, Lisa, Marge}



Def: Recall

The recall of an IE algorithm is the ratio of the gold standard that is output.

$$rec = \frac{|Output \cap GStandard|}{|GStandard|}$$

Output: {Homer, Bart, Groening}

G.Standard: {Homer, Bart, Lisa, Marge}



Def: Recall

The recall of an IE algorithm is the ratio of the gold standard that is output.

$$rec = \frac{|Output \cap GStandard|}{|GStandard|}$$

Output: {Homer, Bart, Groening}

G.Standard: {Homer, Bart, Lisa, Marge}



Def: Recall

The recall of an IE algorithm is the ratio of the gold standard that is output.

$$rec = \frac{|Output \cap GStandard|}{|GStandard|}$$

Output: {Homer, Bart, Groening}

G.Standard: {Homer, Bart, Lisa, Marge}

✓

✓

✗

✗

Def: Recall

The recall of an IE algorithm is the ratio of the gold standard that is output.

$$rec = \frac{|Output \cap GStandard|}{|GStandard|}$$

Output: {Homer, Bart, Groening}

G.Standard: {Homer, Bart, Lisa, Marge}

✓ ✓ ✗ ✗

=> Recall: $2/4 = 50\%$

Example: Precision & Recall

Gold Standard: {Homer Simpson, Bart Simpson, Lisa Simpson}

Algorithm 1: [A-Z][a-z]+ Simpson

Output: {The Simpson, Homer Simpson,
Bart Simpson, Lisa Simpson}

Precision: $3 / 4 = 75\%$

Recall: $3/3=100\%$

Example: Precision & Recall

Gold Standard: {Homer Simpson, Bart Simpson, Lisa Simpson}

Algorithm 2: [A-Z][a-z]+ [A-Z][a-z]+

Output: {The Simpson, Homer Simpson,
Bart Simpson, Lisa Simpson, Matt Groening}

Precision: $3/5=60\%$

Recall: $3/3=100\%$

Example: Precision & Recall

Gold Standard: {Homer Simpson, Bart Simpson, Lisa Simpson}

Algorithm 3: "Homer Simpson"

Output: {Homer Simpson}

Precision: $1/1=100\%$

Recall: $1/3=33\%$

Precision & Recall Trade-Off

Algorithm 1: [A-Z][a-z]+ Simpson

Finds all Simpsons, but also one bad name.

High Recall, Low Precision

Precision: $3/4=75\%$

Recall: $3/3=100\%$

Algorithm 3: Homer Simpson

Finds only one Simpson, but this one is correct.

High Precision, Low Recall

Task: Precision & Recall

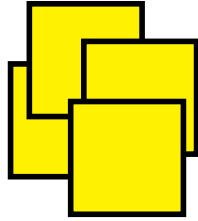
What is the algorithm output, the gold standard, the precision and the recall in the following cases?

1. Nostradamus predicts a trip to the moon for every century from the 15th to the 20th incl.
2. The weather forecast predicts that the next 3 days will be sunny. It does not say anything about the 2 days that follow. In reality, it is sunny during all 5 days.
3. On Elvis Radio TM, 90% of the songs are by Elvis. An algorithm learns to detect Elvis songs. Out of 100 songs on Elvis Radio, the algorithm says that 20 are by Elvis (and says nothing about the other 80). Out of these 20 songs, 15 were by Elvis and 5 were not.
4. How can you improve the algorithm?

How not to design an IE algorithm

Task: Find Simpson pets

Corpus:



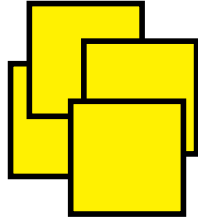
Algorithm: `Regex "Snowball I*"`



How not to design an IE algorithm

Task: Find Simpson pets

Corpus:



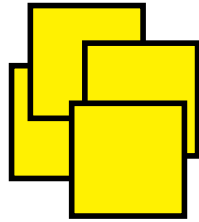
Algorithm: `Regex "Snowball I*"`

Output: `{Snowball I, Snowball II}`

How not to design an IE algorithm

Task: Find Simpson pets

Corpus:



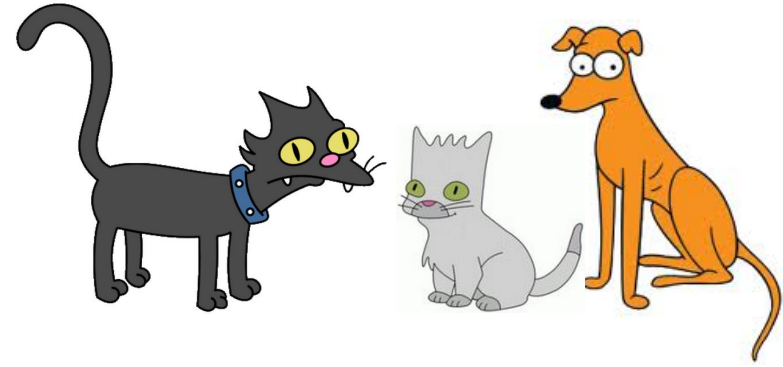
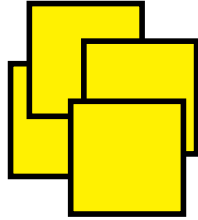
Algorithm: Regex: "Snowball (I—V)*"



How not to design an IE algorithm

Task: Find Simpson pets

Corpus:



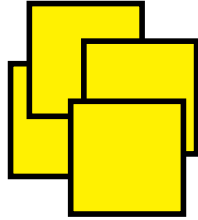
Algorithm: Regex: "Snowball (I—V)*"

Output: {Snowball I,Snowball II,Snowball IV}

How not to design an IE algorithm

Task: Find Simpson pets

Corpus:



Algorithm: Regex: "Snowball (I—V)*"

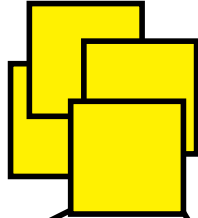
Output: {Snowball I,Snowball II,Snowball IV}

Is this algorithm good?

How to design an IE algorithm

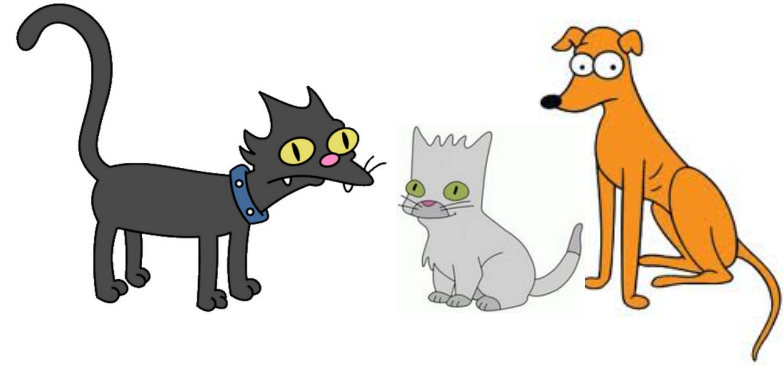
Task: Find Simpson pets

Corpus:



Take only a sample
of the corpus

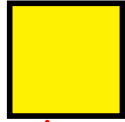
Lisa decides to play music on her saxophone for Coltrane, but the noise frightens him and he commits suicide. As Gil swerves to avoid hitting Snowball V, his car hits a tree and bursts into flames. Since the cat is unhurt, Lisa takes it as a sign of good luck and adopts her. [...]



How to design an IE algorithm

Task: Find Simpson pets

Corpus:



Consider only
the sample corpus.

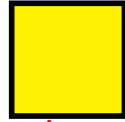


How to design an IE algorithm

Task: Find Simpson pets



Corpus:



Consider only
the sample corpus.

Gold Standard:

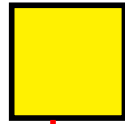
{Coltrane, Snowball I, ...}

Manually make
a gold standard

How to design an IE algorithm

Task: Find Simpson pets

Corpus:



Algorithm

Gold Standard:
{Coltrane, Snowball I, ...}

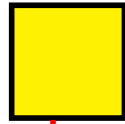


How to design an IE algorithm

Task: Find Simpson pets



Corpus:



Algorithm

Gold Standard:

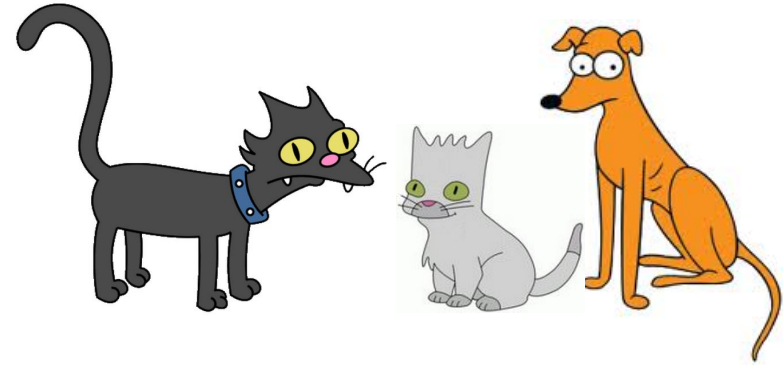
{Coltrane, Snowball I, ...}

Output:

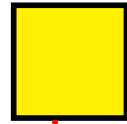
{...}

How to design an IE algorithm

Task: Find Simpson pets

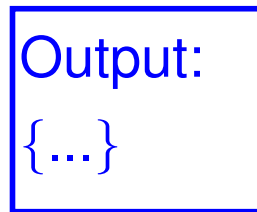


Corpus:



Gold Standard:
{Coltrane, Snowball I, ...}

Algorithm



Evaluator

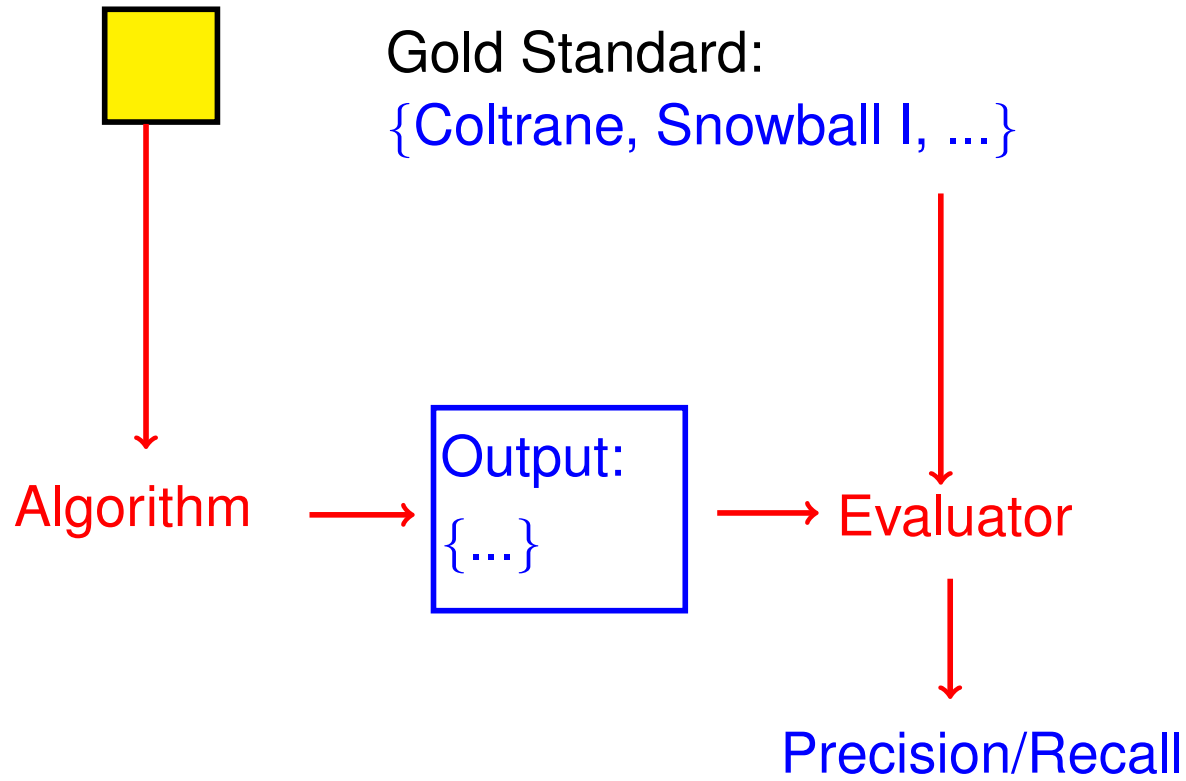


How to design an IE algorithm

Task: Find Simpson pets

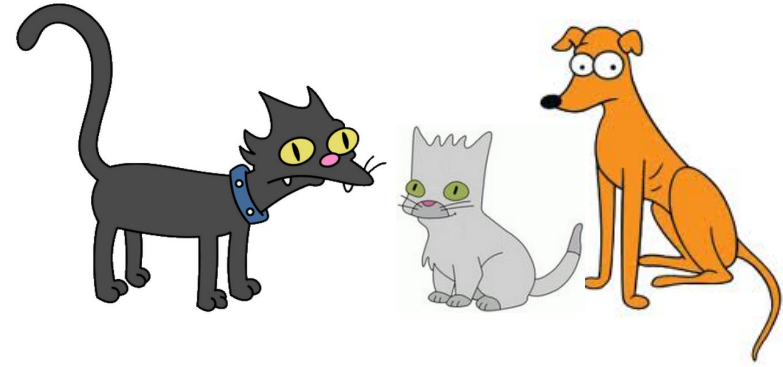


Corpus:

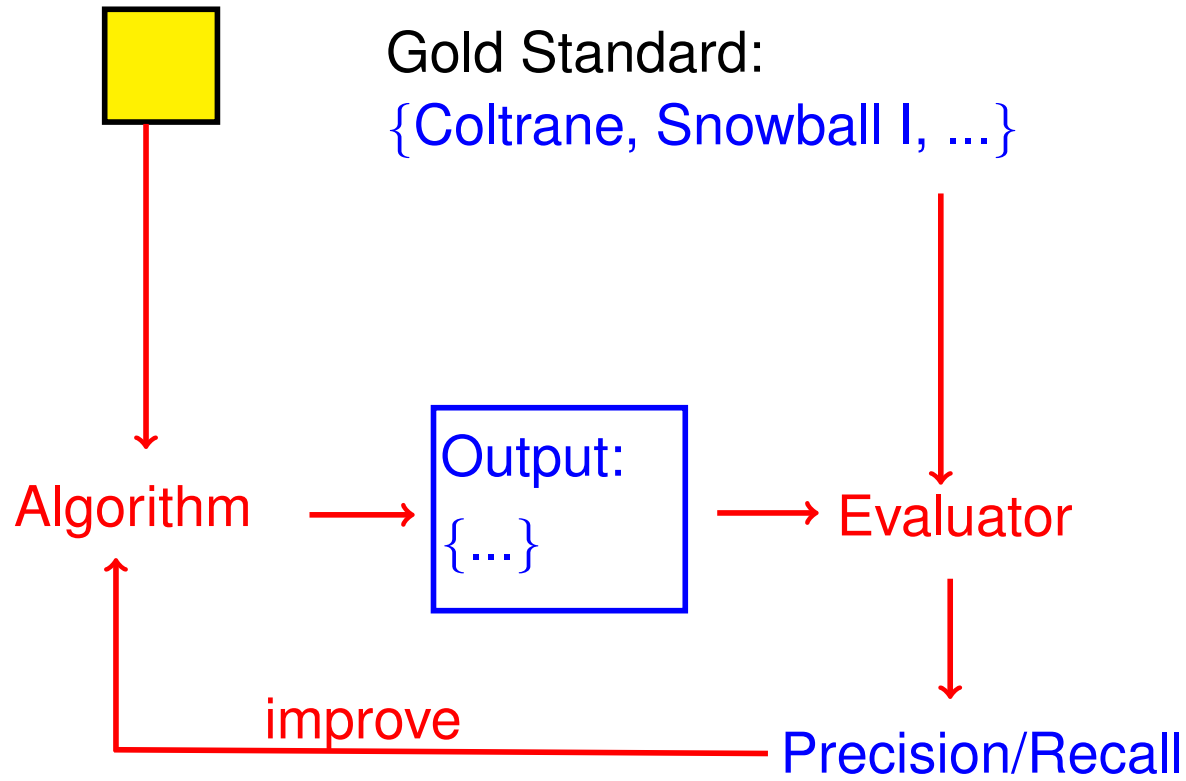


How to design an IE algorithm

Task: Find Simpson pets



Corpus:



Evaluation on a Sample

Corpus:

1: ...A...B...
2: ...C....
3: ..D.....E...
4:H.....
5: ..I...J...K...

Algorithm:

{1: A, Z
2: C
3: D, E, K
4: L,
5: I, K, X}

Sample:

3: ..D.....E...
4:H.....

Gold Standard:
{D, E, H}

Sample:

{
3: D, E, K
4: L
}

Precision: 2/4

Recall: 2/3



A, B, etc. can be entities, but also facts

Evaluation with 1 target per doc

Corpus:

A: ...A'...
B: ...B'...
C: ...C'...
D: ...D'...
E: ...E'...



Sample:

C: ...C'...
D: ...D'...
E: ...E'...



Gold Standard

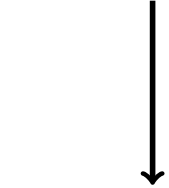
on sample:

{C:C', D:D', E:E'}



Precision: 1/2

Recall: 1/3



Algorithm:

{A: A'
B: X
C: Z
D: D'
}



Sample output:

{
C: Z,
D: D'
}



Evaluation with 1 output per input

Corpus:

A: ...A'...
B: ...B'...
C: ...C'...
D: ...D'...
E: ...E'...



Sample:

C: ...C'...
D: ...D'...
E: ...E'...



Gold Standard

on sample:

{C:C', D:D', E:E'}



Precision: 1/3

Recall: 1/3

Algorithm:

{A: A'
B: X
C: Z
D: D'
E: K }

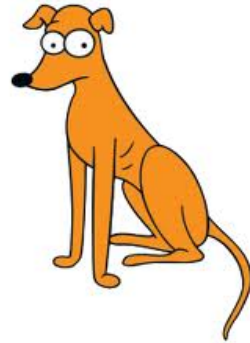


Sample output:

{
C: Z,
D: D',
E: K
}

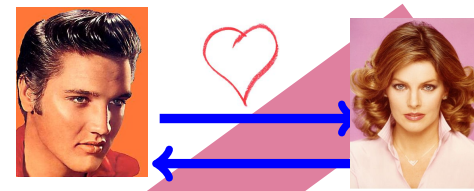


If the algorithm produces
one output per input, prec=rec.

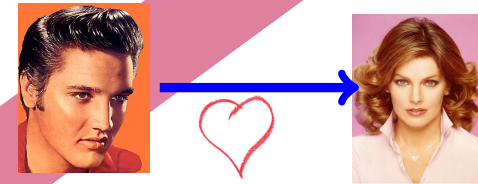


Semantic IE

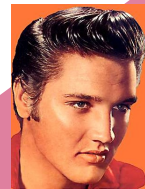
Reasoning



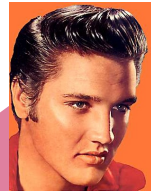
Fact Extraction



Instance Extraction



Entity Disambiguation



Entity Recognition



->named-entity-annotation

->disambiguation

->instance-extraction

Source Selection and Preparation