Uncommonsense: Informative Negative Knowledge about Everyday Concepts

Hiba Arnaout, Simon Razniewski, Gerhard Weikum, Jeff Pan

Outline

- **Motivation**
- **Existing methods**
- **Our method**
- **Experiments**
- **Conclusion and resources**

What is Commonsense Knowledge (CSK)?

Semi-structured knowledge about everyday concepts

- *athlete, laptop, …*

not to be confused with instances

- ~~*Cristiano Ronaldo*~~, ~~*Dell XPS*~~, …

What is Commonsense Knowledge (CSK)?

Semi-structured knowledge about everyday concepts

- *athlete, laptop, …*

not to be confused with instances

- ~~*Cristiano Ronaldo*~~, ~~*Dell XPS*~~, …

Long history in AI

- McCarthy (1959)
- Knowledge Bases: ConceptNet (1999) .., Ascent (2021)

Commonsense Knowledge Bases (CSKBs)

Store CSK in triple form (statements)

- **Positive:**
  `(lasagna, MadeOf, pasta)`

- **Negative:**
  `(lasagna, NotMadeOf, apple)`

Commonsense Knowledge Bases (CSKBs)

Store CSK in triple form (statements)

- **Positive:**
  `(lasagna, MadeOf, pasta)`

- **Negative:**
  `(lasagna, NotMadeOf, apple)`

Existing CSKBs focus on obtaining positives

  – ConceptNet < 2% negated statements

Baseline: Closed-world Assumption

Assume absent statements are false (=negative), but …

- **Missing positives**

  `(elephant, HasA, eye)` $\notin$ KB $\rightarrow$ `(elephant, NotHasA, eye)`

- **Trivial negatives**

  `(elephant, IsA, disease)` $\notin$ KB $\rightarrow$ `(elephant, NotIsA, disease)`

Baseline: Closed-world Assumption

Assume absent statements are false (=negative), but …

- **Missing positives**

  `(elephant, HasA, eye)` $\notin$ KB $\rightarrow$ `(elephant, NotHasA, eye)`

- **Trivial negatives**

  `(elephant, IsA, disease)` $\notin$ KB $\rightarrow$ `(elephant, NotIsA, disease)`

Many wrong or uninformative negatives!

Research Problem

Enrich CSKB concepts with informative negations.

(penguin, NotCapableOf, fly)

(lawyer, NotHasProperty, objective)

(cricket, NotIsA, olympic sport)

Outline

- ✓ **Motivation**
- ○ **Existing methods**
- ○ **Our method**
- ○ **Experiments**
- ○ **Conclusion and resources**

Related Work

- **Text-based**: Mining query logs [CIKM'19]

  why can't elephants.. → jump, use computers, hide in trees
  Low informativeness

Related Work

- **Text-based**: Mining query logs [CIKM'19]

  why can't elephants.. → jump, use computers, hide in trees

  Low informativeness

- **Corruption-based**: Generate negatives from positives [EMNLP'21]

  `(horse, IsA, expensive pet)` → `(horse rider, NotIsA, expensive pet)`

  Type-inconsistency

Related Work

- **Text-based**: Mining query logs [CIKM'19]

  why can't elephants.. → jump, use computers, hide in trees

  Low informativeness

- **Corruption-based**: Generate negatives from positives [EMNLP'21]

  `(horse, IsA, expensive pet)` → `(horse rider, NotIsA, expensive pet)`

  Type-inconsistency

- **LM-based**: Prompting LMs [ACL'20]

  Horses are not.. animal; they are mammals.

  Low accuracy

Outline

- ✓ **Motivation**
- ✓ **Existing methods**
- ○ **Our method**
- ○ **Experiments**
- ○ **Conclusion and resources**

Our method: *Uncommonsense*

Discover informative negations about target concepts
by exploiting positives about comparable concepts (e.g., type siblings).

Input: target concept S

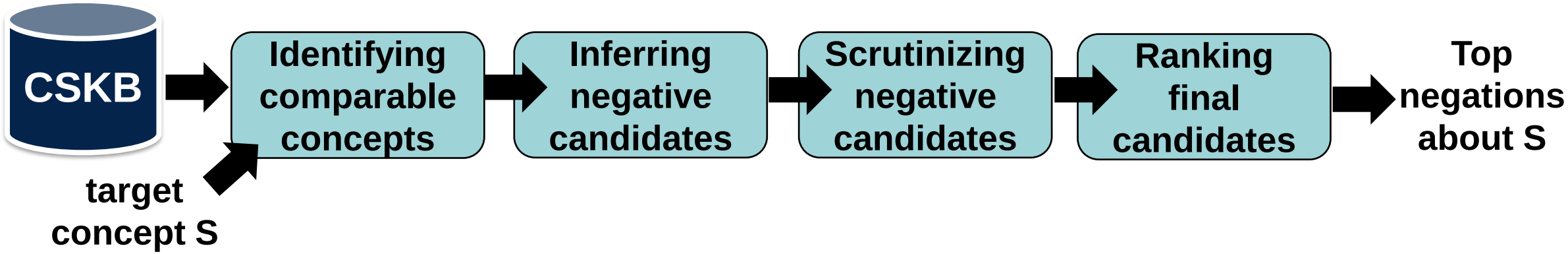
elephant

Output: top-negations about S

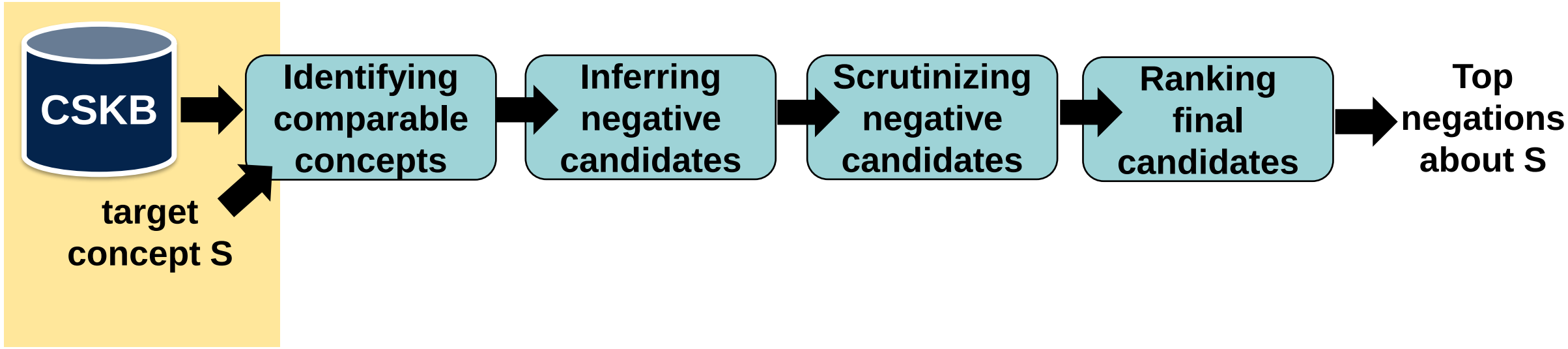
`(elephant, NotIsA, carnivore)`

unlike other land mammals, e.g., lion, tiger

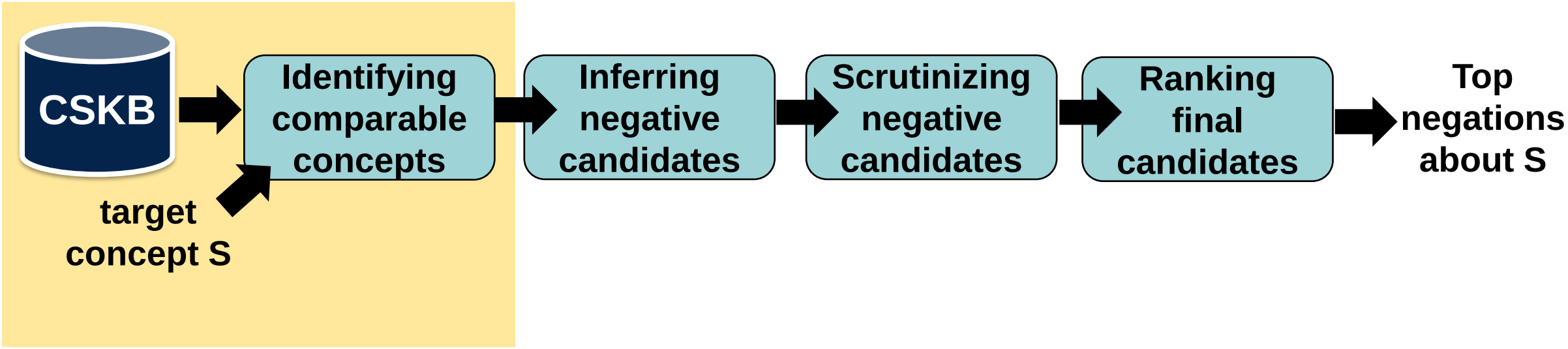
Uncommonsense: Overview



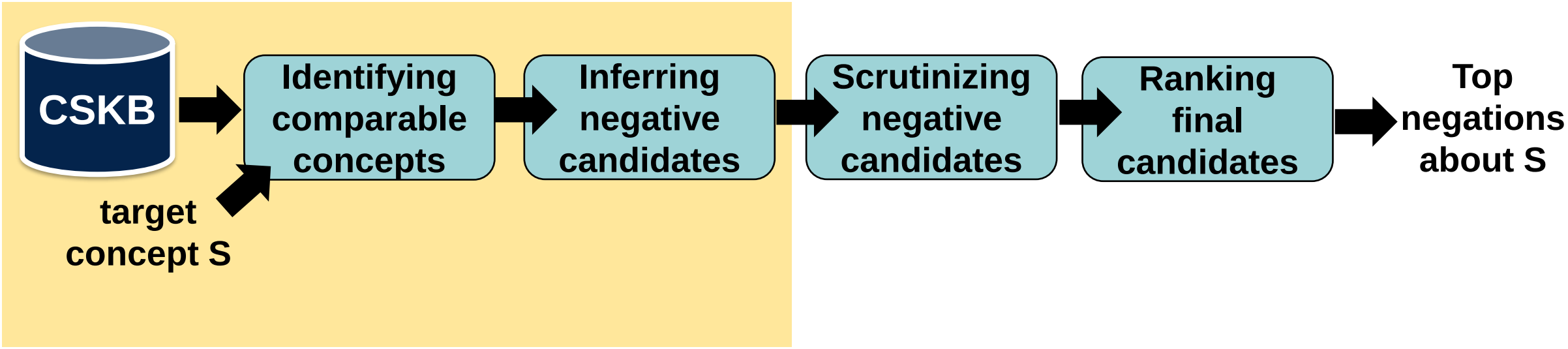
Uncommonsense: Overview



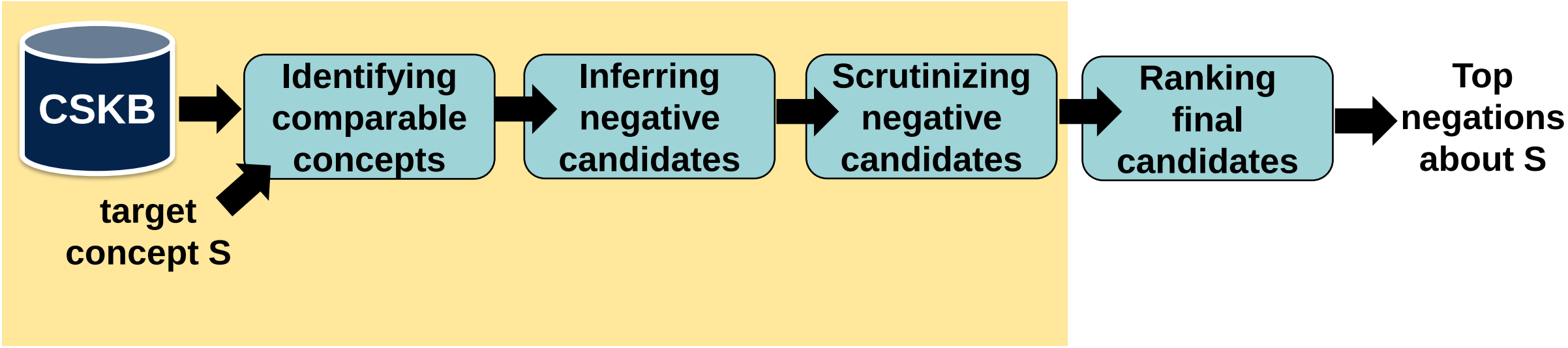
Uncommonsense: Overview



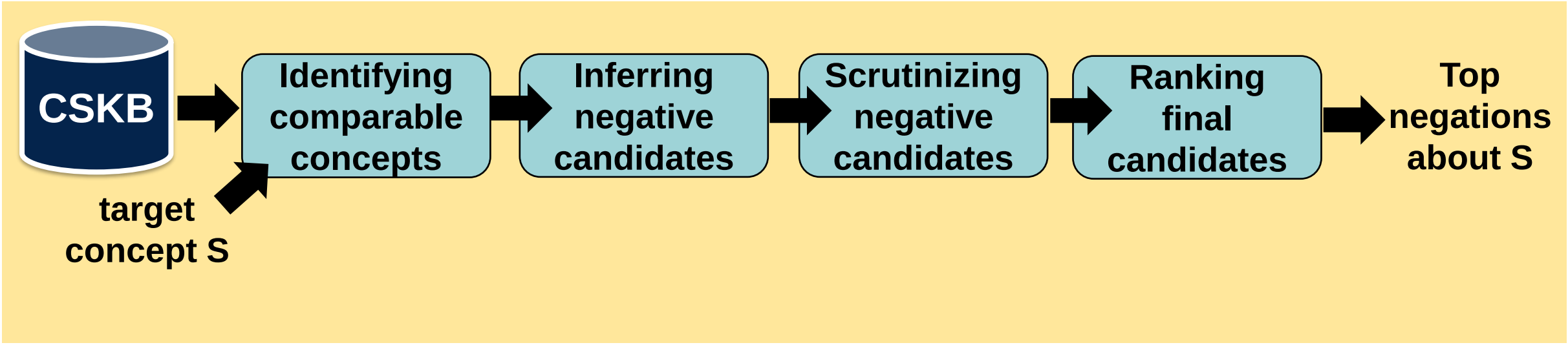
Uncommonsense: Overview



Uncommonsense: Overview



Uncommonsense: Overview



Step 1: Identify Comparable Concepts

Target S 

Step 1: Identify Comparable Concepts

Target S 

We need a similarity measure!

Step 1: Identify Comparable Concepts

Target S 

We need a similarity measure!

Opt. A) Hypernymy relations: hamster, bat, lion.. (under mammal)

Lacks ranking

Opt. B) Latent representations: tiger, trunk, horse, circus, lion ..

Lacks type-consistenty

Step 1: Identify Comparable Concepts

Target S 

We need a similarity measure!

Opt. A) Hypernymy relations: hamster, bat, lion.. (under mammal)

Lacks ranking

Opt. B) Latent representations: tiger, trunk, horse, circus, lion ..

Lacks type-consistenty

Instead, a combination..

Step 1: Identify Comparable Concepts

Target S 

We need a similarity measure!

Opt. A) Hypernymy relations: hamster, bat, lion.. (under mammal)

Lacks ranking

Opt. B) Latent representations: tiger, trunk, horse, circus, lion ..

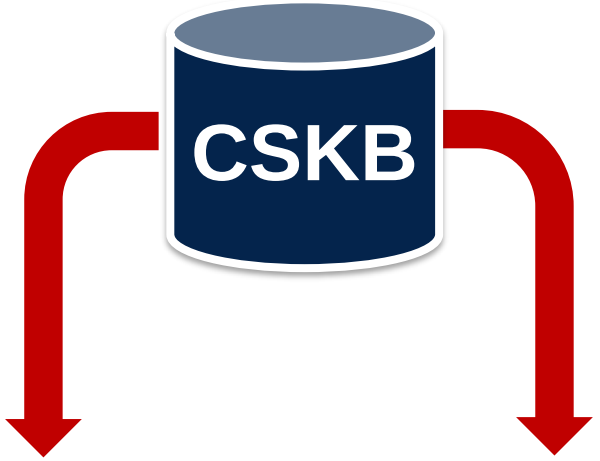
Lacks type-consistenty

Instead, a combination..

tiger, horse, lion, …bat, hamster,.. (k=3)

Step 2: Infer Negative Candidates

Step 2: Infer Negative Candidates



(horse, HasA, hoof)
(horse, CapableOf, eat grass)
(tiger, HasProperty, big animal)
(lion|tiger|horse, CapableOf, jump)

(elephant, HasProperty, large animal)

Step 2: Infer Negative Candidates

(horse, HasA, hoof)
(horse, CapableOf, eat grass)
(tiger, HasProperty, big animal)
(lion|tiger|horse, CapableOf, jump)

(elephant, HasProperty, large animal)

Step 2: Infer Negative Candidates

(horse, HasA, hoof)
(horse, CapableOf, eat grass)
(tiger, HasProperty, big animal)
(lion|tiger|horse, CapableOf, jump)

(elephant, HasProperty, large animal)

(elephant, NotHasA, hoof)
(elephant, NotCapableOf, eat grass)
(elephant, NotHasProperty, big animal)
(elephant, NotCapableOf, jump)

Step 3: Scrutinize Negative Candidates

(elephant, NotHasA, hoof)
(elephant, NotCapableOf, eat grass)
(elephant, NotHasProperty, big animal)
(elephant, NotCapableOf, jump)

Step 3: Scrutinize Negative Candidates

(elephant, NotHasA, hoof)
(elephant, NotCapableOf, eat grass)
(elephant, NotHasProperty, big animal)
(elephant, NotCapableOf, jump)

Remember.. Avoid false negatives

Step 3: Scrutinize Negative Candidates

(elephant, NotHasA, hoof)
(elephant, NotCapableOf, eat grass)
(elephant, NotHasProperty, big animal)
(elephant, NotCapableOf, jump)

Remember.. Avoid false negatives

Plausibility checks:

- **Internal (input KB):** semantic_sim(`big animal`, `large animal`) = *0.88*
- **External (LM):** [MASK] `eat grass` -> elephants

Step 3: Scrutinize Negative Candidates

(elephant, NotHasA, hoof)
(elephant, NotCapableOf, eat grass)
~~(elephant, NotHasProperty, big animal)~~
(elephant, NotCapableOf, jump)

Remember.. Avoid false negatives

Plausibility checks:

- **Internal (input KB):** semantic_sim(`big animal`, `large animal`) = *0.88*
- **External (LM):** [MASK] `eat grass` -> elephants

Step 3: Scrutinize Negative Candidates

(elephant, NotHasA, hoof)

~~(elephant, NotCapableOf, eat grass)~~

~~(elephant, NotHasProperty, big animal)~~

(elephant, NotCapableOf, jump)

Remember.. Avoid false negatives

Plausibility checks:

- **Internal (input KB):** semantic_sim(`big animal`, `large animal`) = *0.88*
- **External (LM):** [MASK] `eat grass` -> elephants

Step 4: Quantify Informativeness

(elephant, NotHasA, hoof)
(elephant, NotCapableOf, jump)

Step 4: Quantify Informativeness

(elephant, NotHasA, hoof)
(elephant, NotCapableOf, jump)

Full example 1352 candidates!

Step 4: Quantify Informativeness

(elephant, NotHasA, hoof)
(elephant, NotCapableOf, jump)

Full example 1352 candidates!

$$\textbf{score}((\text{elephant, NotHasA, hoof}), \{\text{lion, tiger, horse}\}) = \frac{|\{\text{horse}\}|}{|\{\text{lion, tiger, horse}\}|} = 0.33$$

$$\textbf{score}((\text{elephant, NotCapableOf, jump}), \{\text{lion, tiger, horse}\}) = 1.0$$

Outline

- ✓ **Motivation**
- ✓ **Existing methods**
- ✓ **Our method**
- ○ **Experiments**
- ○ **Conclusion and resources**

Setup

- **CSKB**

  Ascent++ (total = 2 million statements)

  200 target concepts (*randomly sampled*)

- **Metrics**

  plausibility: crowd (*is the statement truly negative?*)

  informativeness: crowd (*is the statement interesting?*)

  recall: ground-truth (Conceptnet's 14k negated statements)

Baselines

1. **Closed-world baseline**
   absent = negative, *no ranking*

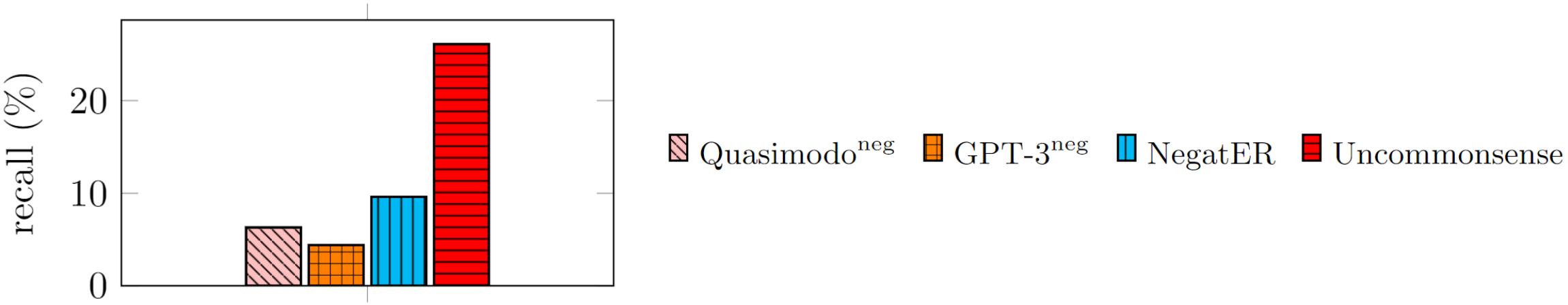
2. **GPT-3^{neg}**
   prompt with target concept & negative relations, *ranking = probabilities*

3. **Quasimodo^{neg}**
   Quasimodo KB's negations, *ranking = confidence*

4. **NegatER**
   corruption-based, *ranking = LM's classification scores*

Results

Method	False Negatives	Informativeness	Negations for *elephant*
CW-baseline	<u>0.07</u>	0.07	(elephant, can't practice law)
Quasimodo^{neg}	0.61	0.32	(elephant, can't survive)
GPT-3^{neg}	0.63	0.30	(elephant, isn't found in the zoo)
NegatER	0.26	0.29	(elephant, isn't interested)
Uncommonsense	0.25	<u>0.50</u>	(elephant, isn't carnivore)



Use Cases

1. **Knowledge Base Completion**
   LM-based classification w/ ***strong*** negative samples.
   Accuracy +4%.

2. **Multiple-choice Question Answering**
   Automatic elimination of ***unlikely*** answers.
   Helpful eliminations +18%.

Outline

- ✓ **Motivation**
- ✓ **Existing methods**
- ✓ **Our method**
- ✓ **Experiments**
- ○ **Conclusion and resources**

Conclusion

- Take-away message:
  - Current CSKBs focus on positive knowledge
  - Explicit negation boost the usability of CSKBS, e.g., QA
  - **Uncommonsense** method produces informative and accurate negations

Conclusion

- Take-away message:
  - Current CSKBs focus on positive knowledge
  - Explicit negation boost the usability of CSKBS, e.g., QA
  - **Uncommonsense** method produces informative and accurate negations

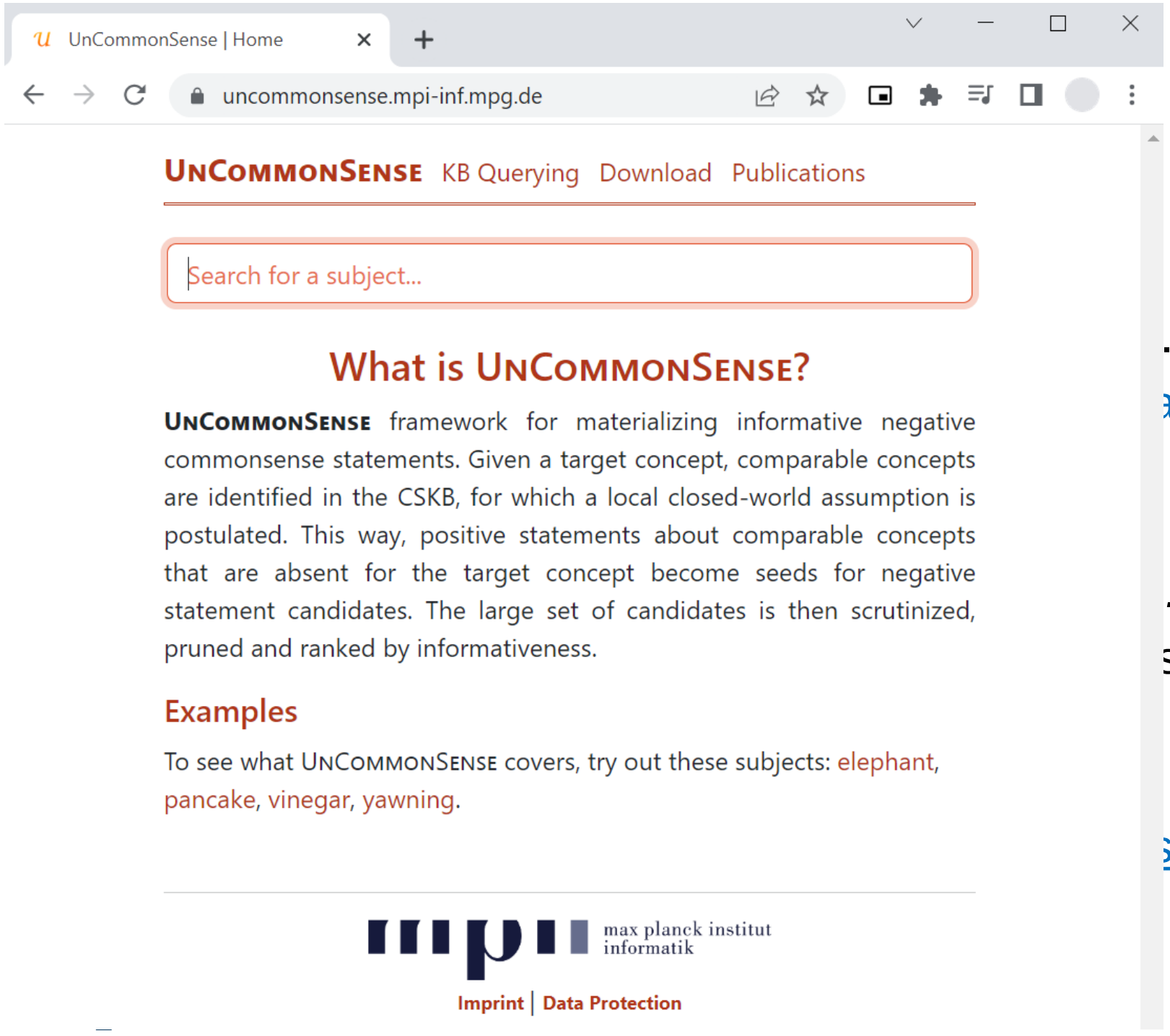
- More in the paper:
  - Insights into specific topics *(food, animal, social,..)*
  - Provenance generation (+32% in informativeness in negative trivia)

Conclusion

- Take-away message:
  - Current CSKBs focus on positive knowledge
  - Explicit negation boost the usability of CSKBS, e.g., QA
  - **Uncommonsense** method produces informative and accurate negations

- More in the paper:
  - Insights into specific topics *(food, animal, social,..)*
  - Provenance generation (+32% in informativeness in negative trivia)

- Resources:



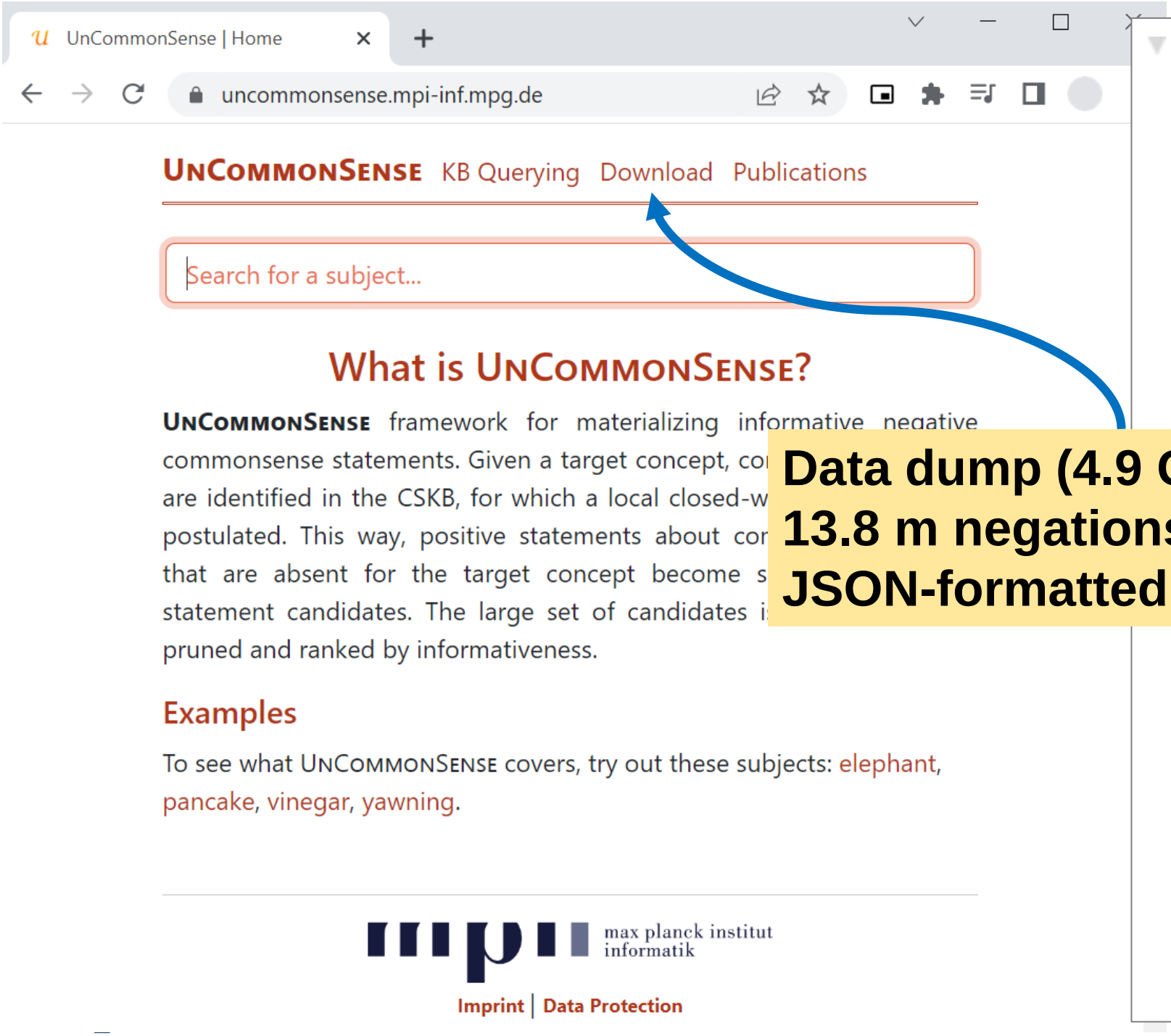
.g., QA

and accurate negations

.)

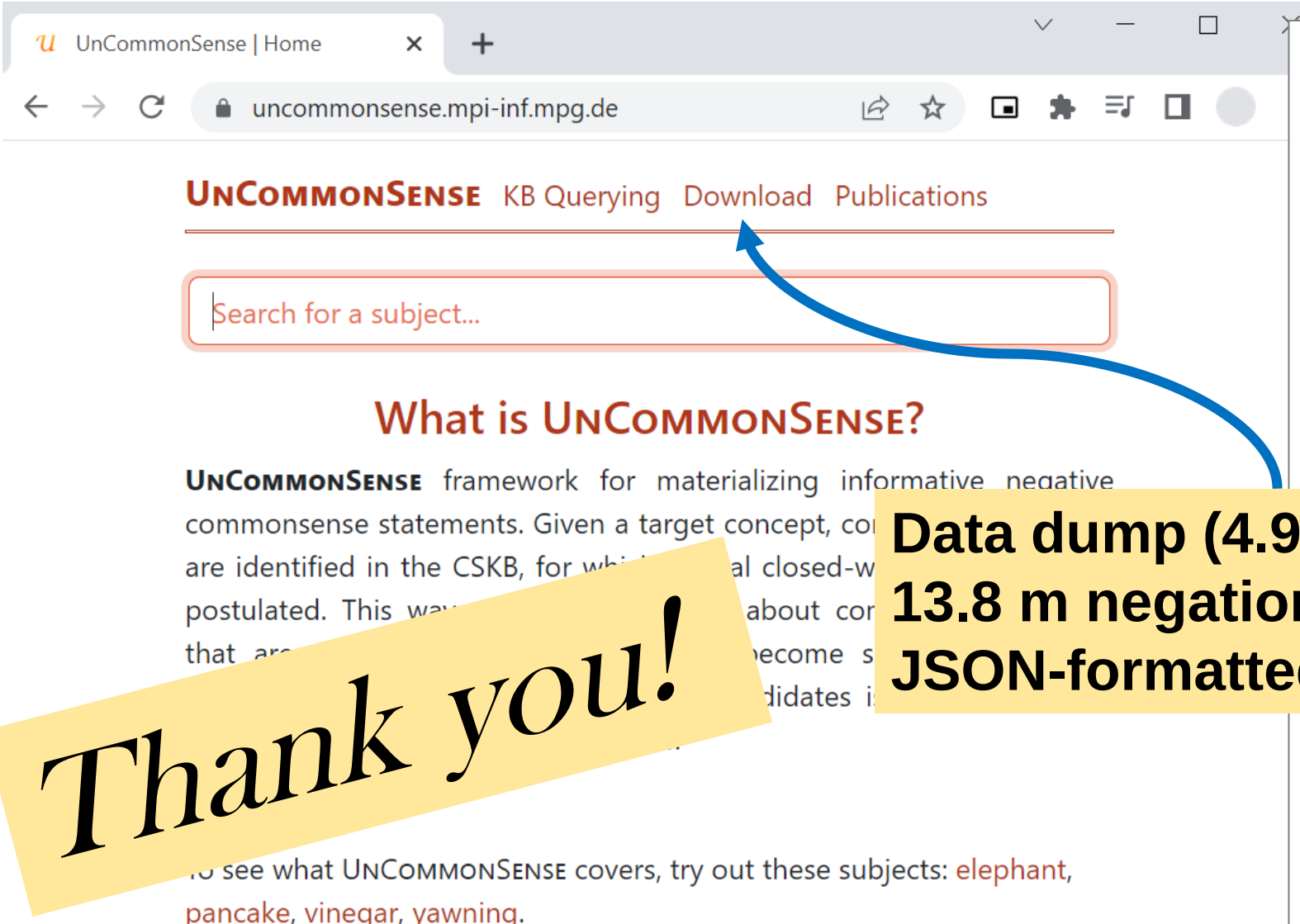
s in negative trivia)

se.mpi-inf.mpg.de)



Data dump (4.9 GB)
13.8 m negations
JSON-formatted


```
{
    "subject": "gorilla",
    "predicate": "HasProperty",
    "object": "territorial",
    "tail_phrase": "be territorial",
    "score": 0.23,
    "strict_siblings": [
        {
            "wild animal": [
                "tiger",
                "lion",
                "monkey",
                "chimpanzee"
            ]
        },
        {
            "species": [
                "wombat",
                "tarsier",
                "gibbon"
            ]
        }
    ]
}
```



Thank you!

Data dump (4.9 GB)
13.8 m negations
JSON-formatted


```
{
    "subject": "gorilla",
    "predicate": "HasProperty",
    "object": "territorial",
    "tail_phrase": "be territorial",
    "score": 0.23,
    "strict_siblings": [
        {
            "wild animal": [
                "tiger",
                "lion",
                "monkey",
                "chimpanzee"
            ]
        },
        {
            "species": [
                "wombat",
                "tarsier",
                "gibbon"
            ]
        }
    ]
}
```

Additional slides

Multiple-choice QA dataset

Table 5: Example of MCQA through elimination process (eliminated choice and correct choice).

Concept = hand, **Query** = What is a hand?
Eliminator = NegatER **A.** foot (-) **B.** feet (-) **C.** digestive organ (-) **D.** body part (-) **E.** help (-)
Eliminator = UNCOMMONSENSE **A.** foot (¬ foot) **B.** feet (¬ foot) **C.** digestive organ (¬ digestive system) **D.** body part (-) **E.** help (-)

COMMONSENSEQA: A Question Answering Challenge Targeting Commonsense Knowledge

Alon Talmor[*,1,2] **Jonathan Herzig**[*,1] **Nicholas Lourie**[2] **Jonathan Berant**[1,2]

[1]School of Computer Science, Tel-Aviv University
[2]Allen Institute for Artificial Intelligence
{alontalmor@mail, jonathan.herzig@cs, joberant@cs}.tau.ac.il, nicholasl@allenai.org

Baseline: Closed-world Assumption

Missing positives: `(elephant, HasA, eye)` $\notin$ KB $\rightarrow$ `(elephant, NotHasA, eye)`!

Many Trivial (nonsensical) negatives:

	Elephant	Newspaper	Cake	..
Atlocation, oven	✗	✗	✓	✗
IsA, herbivore	✓	✗	✗	✗
MadeOf, paper	✗	✓	✗	✗
IsA, disease	✗	✗	✗	✗
..	✗	✗	✗	✗

Elephant in Ascent KB: 1000 positives, 832000 negatives!

Provenance Generation

Target Concept	Negation
muffin	¬(is runny) unlike other *breakfast item*, e.g., *syrup, yogurt*
gorilla	¬(is territorial) unlike other *wild animal*, e.g., *tiger, lion, monkey, chimpanzee*
vinegar	¬(has iron) unlike other *ingredient*, e.g., *fennel, celery, fenugreek* and *acidic food*, e.g., *tomato*
ear	¬(is muscular) unlike other *body part*, e.g., *shoulder, loin, neck*

Ablation Study

Configuration	False Negatives	Informativeness
w/o comparable concepts	0.19	0.26
w/o quality checks	0.28	0.22
w/o plausibility checks	0.49	0.38
w/o ranking	0.39	0.29
complete configuration	*0.25*	*0.50*