Completeness, Recall, and Negation in Open-World Knowledge Bases

Simon Razniewski, Hiba Arnaout, Shrestha Ghosh, Fabian Suchanek

1. Introduction & Foundations (Simon)
2. Predictive recall assessment (Fabian)
3. Counts from text and KB (Shrestha)
4. Negation (Hiba)
5. Wrap-up (Simon)
Open-world Assumption

Wikidata - Awards won by Stephen Hawking

42 awards (in KB)

Adams Prize
Pius XI Medal
Michelson–Morley Award
Royal Society Science Books Prize

Nobel Prize in Physics?
MAYBE!

Knowledge Base doesn’t know!
Problem

 Wikidata - Awards won by Stephen Hawking

42 awards (in KB)

Adams Prize — Albert Einstein Medal

Pius XI Medal — Oskar Klein Medal

Michelson–Morley Award — Hughes Medal

Royal Society Science Books Prize — Wolf Prize in Physics

Nobel Prize in Physics

Existing positive-only KBs are unaware of negation.
Solution: materialize ALL negations?

Wikidata - Awards won by Stephen Hawking

42 won awards, 30000 possible awards

Academy Award
Adams Prize
Più XI Medal
Michelson–Morley Award
Royal Society Science Books Prize
Nobel Prize in Physics
Albert Einstein Medal
NBA Most Valuable Player Award
Oskar Klein Medal
Hughes Medal
Wolf Prize in Physics

Challenge 1: Set of negative statements is quasi-infinite!
Challenge 2: Not every missing link is a negation
Do knowledge bases allow negatives?

- **ERDF knowledge bases**
  - Extended RDF to include ¬ and ∼
  - In OWA ¬awardReceived(Hawking, Nobel in Physics)

- **Count properties**

- **Object = No-value**

- **Negated properties**
  - DBpedia, never exceed alt
  - Knowlife, not caused by

- **Depreciated rank**
  - Flagging triples known to include errors

**Good news:** formally defined negative information

**Bad news:** KB Completion challenges, limited domains, no active collection

- Analyti et al., “Stable Model Theory for Extended RDF Ontologies”, ISWC’05
Research Problem

Identify *interesting* negative statements about *entities* in open-world KBs.

¬ (award, Nobel Prize in Physics)
¬ (award, Oscar)
¬ (headquarters, Silicon Valley)
Methodologies

★ Statistical Inferences

Textual Extractions

Language Models
Method: Peer-based Negation Inference

Given entity $e$ in KB:

- Gist:
  - select *highly related entities* (peers)
  - make *local CWA* (within peer group)
  - infer *expectations* about $e$ (*candidate* negations)
  - validate and score expectations

- Output: top *correct* and *salient* negations about $e$.

Arnaout, Razniewski, and Weikum, “Enriching KBs with interesting negative statements”, AKBC’20
What is a similar entity?

Class-based

- Stephen Hawking: Physicist
  *Albert Einstein, Marie Curie, Max Planck*

Jaccard-similarity

- predicate-object pairs shared by entities:
  Hawking AND Einstein = 423/750
  same profession, education, memberships, awards,..

Embedding-based similarity

- low-dimensional latent representations
  \[
  \text{cosine}_\text{sim}(\text{embeddings}(\text{Hawking}), \text{embeddings}(\text{Jane Wilde}))
  \]
Retrieve candidate negations

Arnaout, Razniewski, and Weikum, “Enriching KBs with interesting negative statements”, AKBC’20
Validate candidates

Challenge: correctness of inferred negations. Are they negatives or missing facts?

Retain candidate only in presence of other values (i.e., PCA)

(Hawking, award, {Copley Medal, …}) ⊨ ¬ (award, Nobel Prize in Physics)
(Hawking, hobby, ∅) ⊭ ¬ (sailing, reading)

Significantly boosts correctness of deductions: by 30%.

Arnaout, Razniewski, and Weikum, “Enriching KBs with interesting negative statements”, AKBC’20
Validate candidates

Challenge: correctness of inferred negations. Are they negatives or missing facts?

Retain candidate only in presence of other values (i.e., PCA)

(award, Nobel Prize in Physics) \not\models \neg (award, Nobel Prize in Physics)
(Hawking, hobby, ∅) \not\models \neg (sailing, reading)

Significantly boosts correctness of deductions: by 30%. 

Arnaout, Razniewski, and Weikum, “Enriching KBs with interesting negative statements”, AKBC’20
Learn2Rank:

A. Obtain annotator judgments for statement interestingness [0..1]
   Is it interesting that Hawking never received a Nobel in Physics?
   .. is not the sibling of Maja Einstein?

B. Train supervised model: Linear Regression
   feature: peer frequency, entity and predicate authority, word embeddings

C. Rank assertions by descending scores

Arnaout, Razniewski, and Weikum, “Enriching KBs with interesting negative statements”, AKBC’20
Methodologies

Statistical Inferences

★ Textual Extractions

Language Models
Method: Mine Negations from User Query Logs

- Wisdom of the crowd:
  Search-engine **autocompletion** provides access to **salient user assertions**

- Probing with **negated prefixes**
  - Why didn’t <e>
  - Why hasn’t <e>
  - Why wasn’t <e>
  - …

---

Romero et al., “**Commonsense Properties from Query Logs and Question Answering Forums**”, CIKM’19
Arnaout, Razniewski, and Weikum, “**Enriching KBs with interesting negative statements**”, AKBC’20
Method: Mine Text Revisions

- Anti-knowledge base (AKB)
  KB of *common factual mistakes*
  Complement the positive-only KB

- Gist:
  Exploit entity/number swaps in Wikipedia update logs
  Score using web hits

Revision 505

_Einstein was born in Viennaa._

Revision 506

_Einstein was born in Ulm._

Karagiannis et al., “Mining an “Anti-Knowledge Base” from Wikipedia Updates with Applications to Fact Checking and Beyond”, PVLDB’19
Methodologies

Statistical Inferences

Textual Extractions

★ Language Models
Method: Unsupervised Discovery of Negatives in Commonsense KBs

- NegatER, steps:
  1) Fine-tuning BERT using **positives** commonsense knowledge.
  2) Generate corruptions of KB statements
     **plausible** candidate negatives by corrupting positives
  3) Estimate contradiction score
     over fine-tuned BERT

(horse, IsA, expensive pet)
(cat, IsA, expensive pet)
(goldfish, IsA, expensive pet)
(horse, IsA, expensive car)
### Overview of methods

**Research problem:** Given an entity, compile a list of **correct & salient** negations.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Method</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKBC’20</td>
<td>Peer-based</td>
<td>Salience</td>
<td>Correctness</td>
</tr>
<tr>
<td>Arnaout et al.</td>
<td></td>
<td>Coverage</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Canonicalization</td>
<td></td>
</tr>
<tr>
<td>EMNLP’21</td>
<td>Language Models</td>
<td>Salience</td>
<td>Correctness</td>
</tr>
<tr>
<td>Safavi et al.</td>
<td></td>
<td>Coverage</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Canonicalization</td>
<td></td>
</tr>
<tr>
<td>PVLDB’19</td>
<td>Anti-KB (mining revisions)</td>
<td>Correctness</td>
<td>Salience</td>
</tr>
<tr>
<td>Karagiannnis et al.</td>
<td></td>
<td></td>
<td>Coverage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Canonicalization</td>
<td></td>
</tr>
<tr>
<td>AKBC’20</td>
<td>Query-logs (pattern-based)</td>
<td>Salience</td>
<td>Coverage</td>
</tr>
<tr>
<td>Arnaout et al.</td>
<td></td>
<td>Correctness</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Canonicalization</td>
<td></td>
</tr>
</tbody>
</table>

**Next:** existing tools and datasets about useful negations.
Resources

**WikiNegata** *(online platform)*
Browse interesting negations about Wikidata entities

**Anti-KB** *(dataset)*
Ranked common factual mistakes from Wikipedia

**Explainable salient negations** *(dataset)*
"I did not know that!"
Order oriented peer-based negation inferences.

**ANION** *(dataset)*
Commonsense KB focusing on negated events

**Google Hotel Search** *(online platform)*
Hotel booking with negative features asserted
Resources

★ **WikiNegata** *(online platform)*
Browse interesting negations about Wikidata entities

**Anti-KB** *(dataset)*
Ranked common factual mistakes from Wikipedia

**Explainable salient negations** *(dataset)*
“I did not know that!”
Order oriented peer-based negation inferences.

**ANION** *(dataset)*
Commonsense KB focusing on negated events

**Google Hotel Search** *(online platform)*
Hotel booking with negative features asserted
- Peer-based method.
- Interesting negations about 500K entities.

Arnaout et al., “Wikinegata: A Knowledge Base with Interesting Negative Statements”, VLDB’21
Resources

**Wiki**negata *(online platform)*
Browse interesting negations about Wikidata entities

★ **Anti-KB** *(dataset)*
Ranked common factual mistakes from Wikipedia

**Explainable salient negations** *(dataset)*  "I did not know that!"
Order oriented peer-based negation inferences.

**ANION** *(dataset)*
Commonsense KB focusing on negated events

**Google Hotel Search** *(online platform)*
Hotel booking with negative features asserted
Anti-knowledge base

- Wikipedia update log.
- **116k** likely mistakes where entities or numbers were corrected.

Penicillin was discovered in 1928 by Scottish scientist Alexander Baldwin.

**Confidence score** = 0.619

Karagiannis et al., “Mining an “Anti-Knowledge Base” from Wikipedia Updates with Applications to Fact Checking and Beyond”, PVLDB’19
Resources

WikiNegata *(online platform)*
Browse interesting negations about Wikidata entities

Anti-KB *(dataset)*
Ranked common factual mistakes from Wikipedia

★ Explainable salient negations *(dataset)* “I did not know that!”
Order oriented peer-based negation inferences.

ANION *(dataset)*
Commonsense KB focusing on negated events

Google Hotel Search *(online platform)*
Hotel booking with negative features asserted
• Dataset with 12.5M salient negations
• Inferred using the peer-based method (with ordered peers).

<table>
<thead>
<tr>
<th>Entity</th>
<th>Negation</th>
<th>Contextualized Verbalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abraham Lincoln</td>
<td>¬(death cause, natural)</td>
<td>Unlike the previous 17 U.S. presidents.</td>
</tr>
<tr>
<td>Jeff Bezos</td>
<td>¬(occupation, politician)</td>
<td>Unlike the previous 17 of 21 Time Person of the Year winners.</td>
</tr>
<tr>
<td>Angela Merkel</td>
<td>¬(gender, male)</td>
<td>Unlike the previous 6 Chairmen of the CDU.</td>
</tr>
</tbody>
</table>

Arnaout et al., “Negative Statements Considered Useful”, JWS’21
**Resources**

WikiNegata *(online platform)*
Browse interesting negations about Wikidata entities

**Anti-KB (dataset)**
Ranked common factual mistakes from Wikipedia

**Explainable salient negations (dataset)** *“I did not know that!”*
Order oriented peer-based negation inferences.

**★ ANION (dataset)**
Commonsense KB focusing on negated events

**Google Hotel Search (online platform)**
Hotel booking with negative features asserted
• Commonsense knowledge graph
  624K if-then rules

https://github.com/liweijiang/anion

Jiang et al., “I’m Not Mad: Commonsense Implications of Negation and Contradiction”, NAACL’21
Resources

**Wiki**negata *(online platform)*
Browse interesting negations about Wikidata entities

**Anti-KB (dataset)*
Ranked common factual mistakes from Wikipedia

**Explainable salient negations (dataset)*
“I did not know that!”
Order oriented peer-based negation inferences.

**ANION (dataset)*
Commonsense KB focusing on negated events

**Google Hotel Search (online platform)*
Hotel booking with negative features asserted
Data crawled from:
- Hotel websites
- Third-party services
- User reviews

https://www.google.com/travel/hotels/
Takeaway: negation

- KBs lack meaningful negative knowledge
- Interest in the *explicit addition* of negation to OWKBs
- **Applications:**
  - Commercial decision making (e.g., hotel booking)
  - General-domain QA systems (e.g., is Switzerland a member of the EU?)
- **Methodologies:** statistical inferences, text extractions, language models
- **KB Challenges:**
  - Class hierarchies
    - profession not `producer` but `film-producer`
  - Modelling issues
    - field not `Information Technology` but `Informatics`
  - Maintenance
    - \(\neg\) (Elon Musk, award, Time’s Person of the Year) valid negation *up until 2021*