Enriching KBs with *interesting negative statements*
Hiba Arnaout, Simon Razniewski, and Gerhard Weikum
## Awards of Stephen Hawking

### Wikidata

<table>
<thead>
<tr>
<th>Award</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albert Einstein Medal</td>
</tr>
<tr>
<td>Wolf Prize in Physics</td>
</tr>
<tr>
<td>Copley Medal</td>
</tr>
<tr>
<td>Presidential Medal of Freedom</td>
</tr>
<tr>
<td>Naylor Prize and Lectureship</td>
</tr>
<tr>
<td>Eddington Medal</td>
</tr>
<tr>
<td>Michelson–Morley Award</td>
</tr>
<tr>
<td>Fellow of the Royal Society</td>
</tr>
<tr>
<td>Order of the British Empire</td>
</tr>
</tbody>
</table>

.............

42 awards in total.
## Awards of Stephen Hawking

**Wikidata**

- Albert Einstein Medal
- Wolf Prize in Physics
- Copley Medal
- Presidential Medal of Freedom
- Naylor Prize and Lectureship
- Eddington Medal
- Michelson–Morley Award
- Fellow of the Royal Society
- Order of the British Empire

42 awards in total.

One **salient** award that he has NOT won …

**The Nobel Prize in Physics!**
Problem: Existing positive-only KBs are unaware of salient negation.
Problem: Existing positive-only KBs are unaware of salient negation.

Our proposal.

‘Explicitly adding salient negative statements to KBs.’
Problem: Existing positive-only KBs are unaware of salient negation.

Our proposal.

‘Explicitly adding salient negative statements to KBs.’

¬(Hawking; award; Nobel Prize in Physics)
¬(Turkey; capital; Istanbul)
¬(Iceland; member of; EU)
Problem: Existing positive-only KBs are unaware of salient negation.

Our proposal.
‘Explicitly adding salient negative statements to KBs.’

¬(Hawking; award; Nobel Prize in Physics)
¬(Turkey; capital; Istanbul)
¬(Iceland; member of; EU)

But how are we going to identify these salient negations?
Problem: Existing positive-only KBs are unaware of salient negation.

Our proposal.

‘Explicitly adding salient negative statements to KBs.’

\[\neg (\text{Hawking; award; Nobel Prize in Physics})\]
\[\neg (\text{Turkey; capital; Istanbul})\]
\[\neg (\text{Iceland; member of; EU})\]

But how are we going to identify these salient negations?

KBs necessarily operate under OWA..

…so what is not in there is not necessarily false.
Problem: Existing positive-only KBs are unaware of salient negation.

Our proposal.

‘Explicitly adding salient negative statements to KBs.’

\[-(\text{Hawking}; \text{award}; \text{Nobel Prize in Physics})\]
\[-(\text{Turkey}; \text{capital}; \text{Istanbul})\]
\[-(\text{Iceland}; \text{member of}; \text{EU})\]

But how are we going to identify these salient negations?

KBs necessarily operate under OWA..

…so what is not in there is not necessarily false.

Our proposal:

peer-based statistical inference + local CWA + learning to rank.
Given a knowledge base KB, and an entity e:

- Gist: we select *highly related entities (peers)* to e, that set *expectations* about e, where the negation of these expectations are potentially *salient*.

- Implicit assumption: within a group of *peers*, we are assuming *local CWA*.

If KB does NOT list..
Nobel in Physics as an award won by Hawking
BUT.. *list it for his peers*..
it is assumed to be *false* for Hawking
(and not merely a missing statement)
Peer-based statistical inference - steps

Input: KB, and e
KB = Wikidata, e = Stephen Hawking

Selecting highly related entities: entity embeddings, structured facets, graph-base measures...

Measure for people -> Occupations(Hawking) = physicist.

Collecting peers of e

Inferring negative candidates

Top-k interesting negations about e
1. ¬ (award; Nobel in Physics)
2. ¬ (citizen; U.S.A.)
3. ¬ (employer; University of Zurich)

Further scoring using a set of features: Property frequency, pivoting...

Learning to rank
1. **Intrinsic**: Ability to rank negations by interestingness;
   Stephen Hawking: \(\neg\ (\text{award; Nobel in Physics})\), \(\neg\ (\text{citizen; U.S.})\),
   \(\neg\ (\text{citizen; Egypt})\) \(\neg\ (\text{actedIn; Titanic})\).

2. **Extrinsic 1**: General entity summarization of only positive statements vs a mix of positive and negative statements.

3. **Extrinsic 2**: Decision making on hotel booking using pos features vs a mix.

4. **Extrinsic 3**: Question answering.
Entity summarization

Setup.
Mixed Wikidata entities.

Task.
Which set contain more interesting information about Hawking?

A
(native language; English)
(child; Lucy Hawking)
(award; Wolf Prize in Physics)
(occupation; astronomer)
(employer; Gonville and Caius College)

B
¬ (award; Nobel Prize in Physics)
(child; Lucy Hawking)
(award; Wolf Prize in Physics)
(occupation; astronomer)
¬ (citizen; U.S.A.)

Results.
72% (mix pos & neg);
16% (pos only);
12% (either or neither).
Hotel booking

Setup.
Booking.com hotel listings.

Task.
Which set of features is more helpful for you to make a decision about staying in this hotel?

Results.
63% (mix pos % neg);
21% (pos only);
16% (either or neither).
Negations are **useful** for entity summarization, decision making, and question answering.

We propose a method for automatically discovering **salient** negations about entities in KBs: peer-based statistical inference.

*More in the paper.*
- More experiments on interestingness and correctness of our inferred negative statements.
- A **second methodology** on automatically extracting salient negations from text – Query-logs.
- First **datasets** on interesting negations from Wikidata – publicly available. [tinyurl.com/yb5dtfqt]
Thank you!

harnaout@mpi-inf.mpg.de