Information Extraction Topics WiSe 2023/24

Kathy Hämmerl
haemmerl@cis.lmu.de

CIS, LMU

08 November 2023
1 Fact Extraction and Verification
2 Canonicalisation of Open Knowledge Bases
3 Knowledge Graph Reasoning
Table of Contents

1 Fact Extraction and Verification

2 Canonicalisation of Open Knowledge Bases

3 Knowledge Graph Reasoning
Misinformation and disinformation are widespread problems.
• Misinformation and disinformation are widespread problems
• Fact Extraction and Verification is the task of finding claims in text and checking for evidence for or against the claim(s)
• Misinformation and disinformation are widespread problems
• Fact Extraction and Verification is the task of finding claims in text and checking for evidence for or against the claim(s)
• “FEVER” dataset and workshop focusing especially on the verification part by finding relevant evidence
Fact Extraction and Verification


Table of Contents

1 Fact Extraction and Verification

2 Canonicalisation of Open Knowledge Bases

3 Knowledge Graph Reasoning
• Open knowledge bases can be extracted from web text, e.g., as triples of (noun phrase, verb phrase, noun phrase)
Open knowledge bases can be extracted from web text, e.g., as triples of (noun phrase, verb phrase, noun phrase).

Multiple NPs can mean the same entity; multiple VPs can mean the same relation.
Canonicalisation of Open Knowledge Bases

- Open knowledge bases can be extracted from web text, e.g., as triples of (noun phrase, verb phrase, noun phrase)
- Multiple NPs can mean the same entity; multiple VPs can mean the same relation
- Without addressing this, the database becomes redundant, very hard to search
• Open knowledge bases can be extracted from web text, e.g., as triples of (noun phrase, verb phrase, noun phrase)
• Multiple NPs can mean the same entity; multiple VPs can mean the same relation
• Without addressing this, the database becomes redundant, very hard to search
• Canonicalisation: Recognise synonyms, cluster them together
Open knowledge bases can be extracted from web text, e.g., as triples of (noun phrase, verb phrase, noun phrase).

Multiple NPs can mean the same entity; multiple VPs can mean the same relation.

Without addressing this, the database becomes redundant, very hard to search.

Canonicalisation: Recognise synonyms, cluster them together.

Example

"Obama was born in Honolulu" == "Barack Obama’s birthplace is Honolulu"
• Luis Galárraga, Geremy Heitz, Kevin Murphy, and Fabian M. Suchanek. 2014. Canonicalizing Open Knowledge Bases. In *CIKM ’14*. ACM.


Table of Contents

1 Fact Extraction and Verification

2 Canonicalisation of Open Knowledge Bases

3 Knowledge Graph Reasoning
Knowledge bases are most useful if we can apply them to, e.g., question answering.
Knowledge bases are most useful if we can apply them to, e.g., question answering.

That is, if we can *extract more information* from the relations that we already have.
• Knowledge bases are most useful if we can apply them to, e.g., question answering
• That is, if we can *extract more information* from the relations that we already have
• This is *related* to the canonicalisation problem
• Knowledge bases are most useful if we can apply them to, e.g., question answering
• That is, if we can *extract more information* from the relations that we already have
• This is *related* to the canonicalisation problem
• This topic is a bit more complex—try not to get too lost in technical detail but focus more on goals, successes and limitations
