SMART Task 3.0

Entity and Relation Linking using CLOCQ

Philipp Christmann, Rishiraj Saha Roy, Gerhard Weikum
Who plays Viserys in GRRM’s latest HBO series?

Paddy Considine
Who plays Viserys in GRRM’s latest HBO series?

Harry Lloyd

Entity Linking (EL)

“Viserys” \(\rightarrow\) Viserys III Targaryen
“GRRM” \(\rightarrow\) George R. R. Martin
“HBO” \(\rightarrow\) HBO

Link entity mentions in question to the KB entities
⇒ Fill SPARQL query slots
⇒ Initiate the search space (e.g. for graph processing)

Relation Linking (RL)

“plays” \(\rightarrow\) plays for team

Link relation mentions in question to KB relations
⇒ Fill SPARQL query slots
⇒ KB fact matching in search space

Green linking: correct
Red linking: wrong

Single error in EL/RL can lead to complete failure!
SMART Task 3.0

★ Relation Linking Task
★ Entity Linking Task
☆ Answer Type Prediction Task

Approach these tasks using CLOCQ

CLOCQ
★ Unsupervised linking tool
★ Parameters can be tuned
★ Publicly available ([clocq.mpi-inf.mpg.de](http://clocq.mpi-inf.mpg.de))

[Christmann et al., Beyond NED: Fast and Effective Search Space Reduction for Complex Question Answering over Knowledge Bases, WSDM 2022]
CLOCQ: Coherent Disambiguations

Approach both problems jointly
⇒ Mentions can lexically/semantically match with many KB items

“plays” ↦ plays for team, instrument, number of plays, character role, time played, ...

⇒ By considering question as a whole, actual meaning becomes clear
⇒ Linked entities and relations in KB should be coherent

[Christmann et al., Beyond NED: Fast and Effective Search Space Reduction for Complex Question Answering over Knowledge Bases, WSDM 2022]
CLOCQ: Coherent Disambiguations

Retrieve candidate disambiguations for all mentions (phrases detected via TagME NER)

Score candidates based on

- **Semantic coherence:** semantic similarity among candidates
  - HBO (network) - George R.R. Martin
  - Hollywood Bowl Orchestra (HBO) - George R.R. Martin

- **KB connectivity:** KB distance among candidates
  - HBO (network) - George R.R. Martin
  - Hollywood Bowl Orchestra (HBO) - George R.R. Martin

- **Question relatedness:** candidate vs. question
  - HBO (network) - “who plays Viserys in GRRM’s latest HBO series?”
  - Hollywood Bowl Orchestra (HBO) - “who plays Viserys in GRRM’s latest HBO series?”

- **Lexical matching:** candidate vs. mention
  - HBO (network) - “HBO”
  - Hollywood Bowl Orchestra (HBO) - “HBO”

Who plays Viserys in GRRM’s latest HBO series?

[Christmann et al., Beyond NED: Fast and Effective Search Space Reduction for Complex Question Answering over Knowledge Bases, WSDM 2022]
CLOCQ: Auto-k Setting

Who plays Viserys in GRRM’s latest HBO series?

Make up for potential errors

⇒ Mentions can be highly ambiguous

“Viserys” ↦ Viserys III Targaryen, Viserys I Targaryen, Viserys II Targaryen,…

⇒ Consider top-k results
⇒ k set automatically based on ambiguity of individual mention
⇒ Entropy of candidate distribution

[Christmann et al., Beyond NED: Fast and Effective Search Space Reduction for Complex Question Answering over Knowledge Bases, WSDM 2022]
CLOCQ: Potential Results

Who plays Viserys in GRRM’s latest HBO series?

<table>
<thead>
<tr>
<th>“Viserys”</th>
<th>KB item</th>
<th>Viserys III Targaryen</th>
<th>Viserys I Targaryen</th>
<th>Viserys II Targaryen</th>
<th>k = 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>“GRRM”</td>
<td>KB item</td>
<td>George R. R. Martin</td>
<td>k = 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“HBO”</td>
<td>KB item</td>
<td>HBO (network)</td>
<td>HBO (company)</td>
<td>k = 2</td>
<td></td>
</tr>
<tr>
<td>“plays”</td>
<td>KB item</td>
<td>character role</td>
<td>characters</td>
<td>instrument</td>
<td>k = 3</td>
</tr>
</tbody>
</table>

[Christmann et al., Beyond NED: Fast and Effective Search Space Reduction for Complex Question Answering over Knowledge Bases, WSDM 2022]
CLOCQ links ALL mentions
⇒ This helps sometimes, but is not always required

“latest” ⟷ latest start date, end time, latest release,…

“series” ⟷ TV series, part of the series, series (rank),…

Should not be linked for EL (or RL) task
⇒ Lead to noise
⇒ Decreased precision

⇒ Idea: Prune mentions that should not be disambiguated
⇒ Train seq2seq model to generate mentions in question
⇒ Consider only entities linked to such mentions (relaxed substring matching)
How to obtain training data for pruning module?

Run CLOCQ on train set: obtain (mention, entity) pairs for each instance

<table>
<thead>
<tr>
<th>CLOCQ outputs</th>
<th>Gold entities</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Viserys” ➔ Viserys III Targaryen, Viserys I Targaryen,…</td>
<td>Viserys I Targaryen</td>
</tr>
<tr>
<td>“GRRM” ➔ George R. R. Martin</td>
<td>George R. R. Martin</td>
</tr>
<tr>
<td>“latest” ➔ latest start date, end time, latest release,…</td>
<td>HBO (network)</td>
</tr>
<tr>
<td>“HBO” ➔ HBO (network), HBO (company),…</td>
<td>HBO (network)</td>
</tr>
<tr>
<td>“series” ➔ TV series, part of the series, series (rank),…</td>
<td></td>
</tr>
</tbody>
</table>

Training instance

**Input:** Who plays Viserys in GRRM’s latest HBO series?

**Output:** Viserys|GRRM|HBO
Experimental Setup

Knowledge Base
⇒ Wikidata – 94M entities, 3k predicates

Metrics
⇒ Precision – how many predictions are correct?
⇒ Recall – how many of the relevant entities/relations are found?
⇒ F1 – harmonic mean between precision and recall
SMART 3.0: Entity Linking Task

Training instances: 49,987
Test instances: 13,794

DC-3 is operated by which airline?

Douglas DC-3

Methods
★ CLOCQ (k=1)
★ CLOCQ (k=AUTO)
★ CLOCQ (k=1) + Pruning
★ CLOCQ (k=AUTO) + Pruning

airline
SMART 3.0: Relation Linking Task

<table>
<thead>
<tr>
<th>Training instances</th>
<th>Test instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>24,112</td>
<td>6,029</td>
</tr>
</tbody>
</table>

**Methods**

- ★ CLOCQ – top-1 relation per mention
- ★ CLOCQ – all relations per mention

**Question:** Where was the actress from Frasier born?

**Answer:** Place of birth

**Cast member**
Entity Linking

If recall is of utmost interest, go with auto-\(k\)

- CLOCQ (\(k=1\))
- CLOCQ (\(k=\text{AUTO}\))
- CLOCQ (\(k=1\)) + Pruning
- CLOCQ (\(k=\text{AUTO}\)) + Pruning
Entity Linking

![Bar chart showing precision, recall, and F1 scores for different entity linking methods.](chart)

- **CLOCQ (k=1)**
- **CLOCQ (k=AUTO)**
- **CLOCQ (k=1) + Pruning**
- **CLOCQ (k=AUTO) + Pruning**

**Pruning module helps to filter noise**
Entity Linking

Combination with auto-k helps to boost recall

- CLOCQ (k=1)
- CLOCQ (k=AUTO)
- CLOCQ (k=1) + Pruning
- CLOCQ (k=AUTO) + Pruning

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLOCQ (k=1)</td>
<td>0.28</td>
<td>0.71</td>
<td>0.41</td>
</tr>
<tr>
<td>CLOCQ (k=AUTO)</td>
<td>0.15</td>
<td>0.77</td>
<td>0.40</td>
</tr>
<tr>
<td>CLOCQ (k=1) + Pruning</td>
<td>0.73</td>
<td>0.87</td>
<td>0.50</td>
</tr>
<tr>
<td>CLOCQ (k=AUTO) + Pruning</td>
<td>0.24</td>
<td>0.83</td>
<td></td>
</tr>
</tbody>
</table>
Relation Linking

Precision for top-1 relation per mention better than using all

Precision for CLOCQ - top-1 mention is 0.38, for CLOCQ - all mention is 0.27.

Recall for CLOCQ - top-1 mention is 0.37, for CLOCQ - all mention is 0.41.

F1 score for CLOCQ - top-1 mention is 0.35, for CLOCQ - all mention is 0.28.
Relation Linking

Difference in recall not that large

Precision: 0.38, 0.27
Recall: 0.37, 0.41
F1: 0.35, 0.28

CLOCQ - top-1
CLOCQ - all
EL – Anecdotal Examples

What was **Toby Wright**’s profession?

“Toby Wright” ↦ Toby Wright (football player)

CLOCQ (k=1)

“Toby Wright” ↦ Toby Wright (football player)

CLOCQ (k=AUTO)

What was **Toby Wright**’s profession?

“Toby Wright” ↦ Toby Wright (football player)

CLOCQ (k=1)

“Toby Wright” ↦ Toby Wright (football player)

CLOCQ (k=AUTO)

“middlesbrough” ↦ Middlesbrough F.C. (football club)

CLOCQ (k=1)

“middlesbrough” ↦ Middlesbrough F.C. (football club)

CLOCQ (k=AUTO)

⇒ **Top-k** helps to resolve ambiguity

⇒ **Preferrable** if remaining QA system can deal with multiple linkings
RL – Anecdotal Examples

- Which **child** of John Adams **died on** February 23, 1848?

  - **child** \(\leftrightarrow\) **child**
  - **died on** \(\leftrightarrow\) **date of death**

  - CLOCQ – top-1

- What is the **point in time** that Nicolaus Cusanus **was made cardinal** by the Holy Roman Church?

  - **point in time** \(\leftrightarrow\) **point in time**
  - **point in time** \(\rightarrow\) **position held**
  - **point in time** \(\rightarrow\) **located in time zone**

  - CLOCQ – top-1

  - CLOCQ – all

\[\Rightarrow\] CLOCQ can **identify multiple** relations in a question

\[\Rightarrow\] Some questions may require **deeper understanding**
Conclusion

CLOCQ
★ Top-k linkings for all mentions (entities, relation, types, concepts)
★ k set automatically for each individual mention
★ Neural pruning reduces noise in post-hoc manner

SMART Task 3.0
★ Recall of CLOCQ for entity linking task well above 0.8
★ Trade-off between recall and precision: best setting highly depends on QA system
★ CLOCQ is a suitable system for both tasks
☆ Integrating neural components in CLOCQ algorithm could improve results

⇒ SMART tasks can help to identify, understand, and eliminate failure cases of QA modules

clocq.mpi-inf.mpg.de

Thank you!