01.

Evaluation of LLMs on Long-tail Entity Linking in Historical Documents

Authors:

Marta Boscariol, Luana Bulla, Lia Draetta, Beatrice Fiumanò, Emanuele Lenzi, Leonardo Piano

Workshop X-TAIL: eXtraction and eXploitation of long-TAIL Knowledge with LLMs and KGs

24TH INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT

02.

Long-tail entities

Entities that are infrequently mentioned or have limited representation in available KBs due to low popularity or specialised context

→ Focus on **historical**, **domain-specific** entities

28 Wikidata triples

Charles Hancock (Q16030597)
English organist and composer
(1852-1927)

'Liddle officiated as leader of the band, Mr. ***C. Hancock*** *presided at the harmonium, and Sir* ***George Elvey*** *conducted"**

George Job Elve (Q5541104)
English organist and composer
(1816-1893)

186 Wikidata triples

*Example taken from MHERCL Benchmark (Graciotti 2023)

03.

Entity linking

1. **Recognition** of entities within a text
2. **Disambiguation** of entities through a KB

Long-tail Entity linking

Detection and disambiguation of entities from niche domains has often proved to be challenging for EL tools

04.

Entity linking

But what about LLMs?

✓ extensive pre-training on large and diverse corpora

✓ deep contextual understanding

🔍 improved EL performances in long-tail scenarios

05.

Research questions

01. How does the most reliable state-of-the-art EL tool perform with long-tail entities?

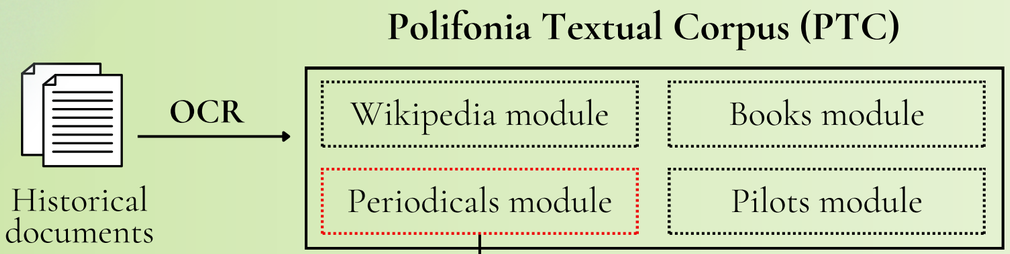
02. Are LLMs suitable for long-tail entity linking?

06.

Experimental setup

Benchmark

MHERCL v0.1.2 (**M**usical heritage **H**istorical named **E**ntities **R**ecognition, **C**lassification and **L**inking)

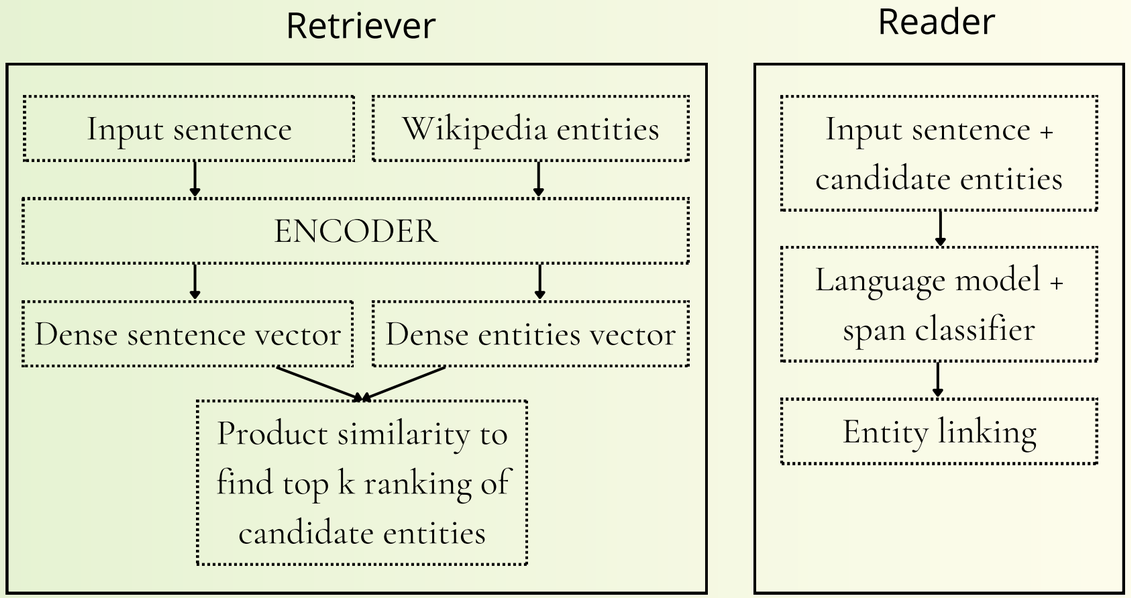


Language	Sentences	Unique NE	NE type
EN	928	966	59

Baseline

ReLiK (**Re**trieve, **Re**ad, **Link**), a state-of-the-art framework for Entity Linking and Relation Extraction

- Architecture: Retriever-Reader
- Performance: 86.4 F1 in-domain, outperforming other EL models



07.

Experimental setup

LLM-based Entity linking

Entity Linking Prompt

```
You are a powerful Entity Linking system.
Given a sentence, identify the key entities and output their exact labels as found on
the corresponding Wikipedia pages. Generate a structured JSON output, formatted as
[{"Entities":{"text entity span": "Wikipedia page title"}].

Here there are some examples:
Sentence:"of Rameau was represented in 1735, it was a balletopera Les Indes galantes."
Output:  [{"Entities":{"Rameau":"Jean-Philippe Rameau","Les Indes galantes":"Les Indes
galantes"}]
```

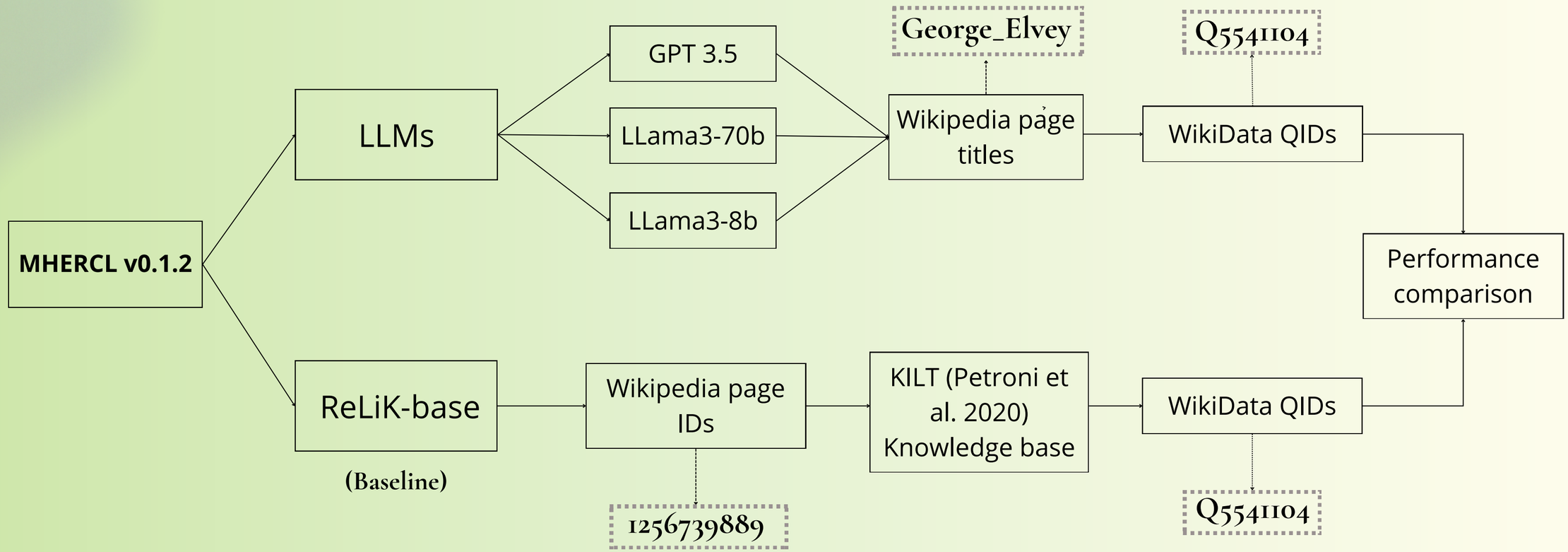
Output

Entity recognition + disambiguation

Example (one-shot approach)

08.

Methodology



09.

Results

Model	Precision(%)	Recall(%)	F1(%)
Relik	**72.8**	45.7	**56.1**
GPT 3.5	48.6	58.8	53.2
Llama3-70b	47.3	**60.3**	53
Llama3-8b	34.9	40.1	37.3

Table 1. Comparison between LLMs and Relik

Relik

✓ Higher accuracy (precision 72.8%)

✗ Smaller number of retrieved entities (recall 45%)

LLMs

✓ Except for LLama3-8b, LLMs correctly linked a **higher number of entities**

✓ LLama3 -70b reached a recall score of 60.3%

✗ Low precision: LLMs tended to over-generate fictional entities, **raising the number of false positives.**

10.

Results

How does performance change with **varying entity popularity**?

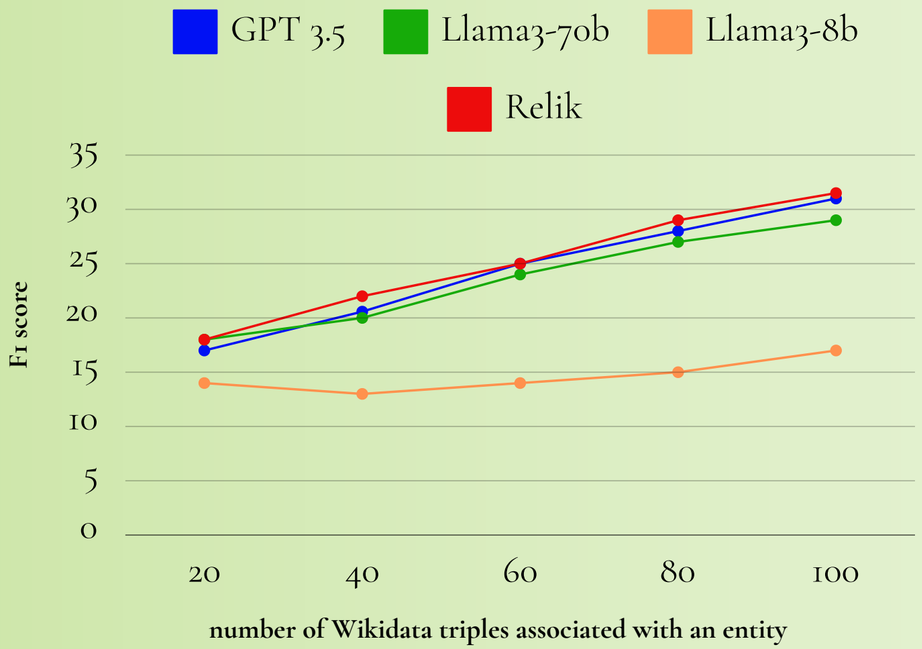
How to define long-tail entities?

We measured the number of **Wikidata triples** associated with each entity

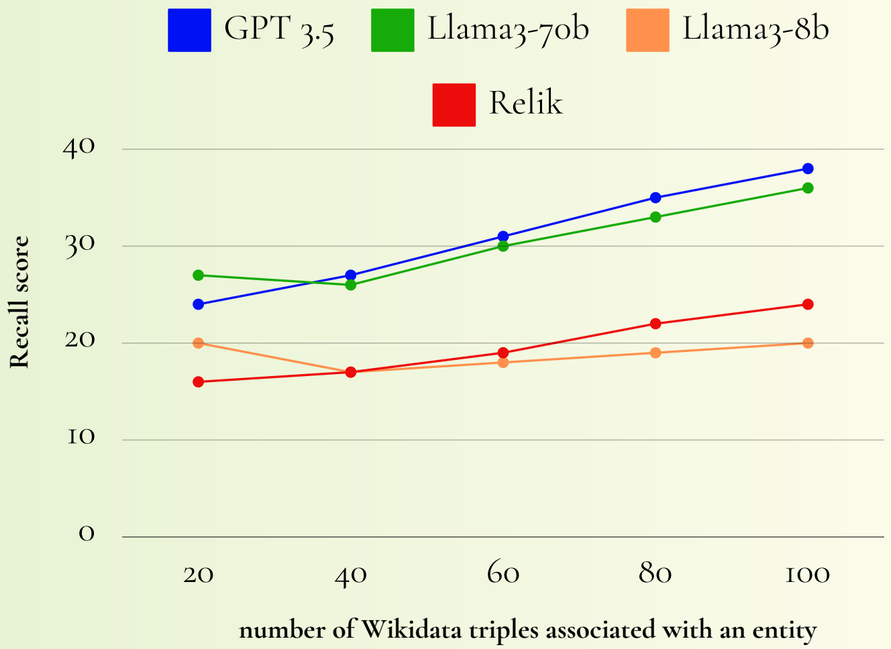
Charles Hancock → 28 WikiData triples

Ludwig van Beethoven → 1419 WikiData triples

F1 Score Trends with varying entity frequency threshold



Recall score Trends with varying entity frequency threshold



II.

Qualitative evaluation

OCR noise with little context

‘Mr. Mocre is the adaptor of words to this composition, which is a tirana, arranged by Mr. Bishop’

Thomas Moore (Q315346)

- ✓ Human annotators
- ✗ Relik
- ✗ GPT 3.5
- ✗ Llama3-70b
- ✗ Llama3-8b

OCR noise with context

‘One man may lived, who ean read the heart, and whose power was not: based upon, his own experience but if so, we may well call William Shakspeare superhuman,THenee it was that whiffe i m Rossint’s ‘Barber of Seville,’ ar Cimarosa’s ‘Seeret Marriage’’

Gioachino Rossini (Q9726)

The Secret Marriage (Q428319)

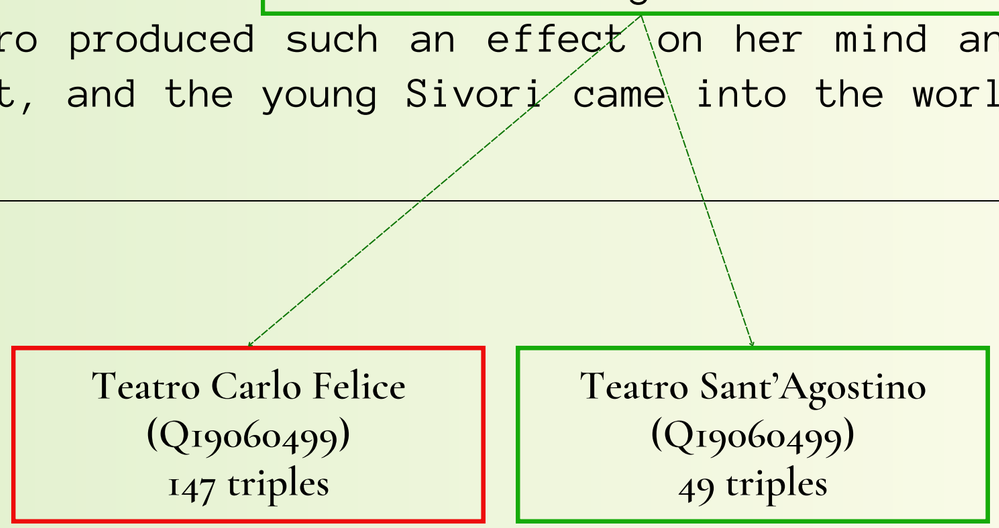
- ✓ GPT 3.5
- ✓ Llama3-70b
- ✗ Relik
- ✗ Llama3-8b

Qualitative evaluation

Niche entities

‘We may mention as a remarkable circumstance that, on the evening of the night on which he was born, his mother, notwithstanding the delicacy of her situation, was induced to go to a concert given by Paganini at the Teatro Santo Augustino in Genoa, when the performance of the great Maestro produced such an effect on her mind and nerves as to precipitate her accouchement, and the young Sivori came into the world somewhat before his time.’

- ✓ Relik
- ✗ GPT 3.5
- ✗ Llama3-70b
- ✗ Lamma3-8b



13.

Conclusion

RQ1: How well does a reliable state-of-the-art EL tool perform in long-tail scenarios?

Our exploratory study reveals that:

- The baseline model's performance tends to decline as entity popularity decreases
- ReLiK consistently achieved the highest precision and F1 scores compared to the LLMs
- However, Relik shows lower recall wrt LLMs

RQ2: Are LLMs suitable for long-tail entity linking?

Our exploratory study shows that:

- LLMs achieved higher recall, recovering a greater number of entities in a long-tail, domain-specific scenario
- However, they exhibited lower precision, often retrieving non-relevant entities alongside the correct ones

→

Long-tail entity linking is still an open challenge!

Future work

→ Further explore limitations and potential of LLMs
→ Identify approaches to increase accuracy

Explore In-Context Learning (ICL) approaches to produce more accurate outputs

Leverage Knowledge Injection methods to augment LLMs' knowledge and their contextual understanding

Thank you!

Questions?

Marta Boscariol, University of Turin - marta.boscariol@unito.it

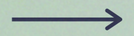
Luana Bulla, University of Catania - luana.bulla@phd.unict.it

Lia Draetta, University of Turin - lia.draetta@unito.it

Beatrice Fiumanò, University of Bologna - beatrice.fiumano@unibo.it

Emanuele Lenzi, University of Pisa - emanuele.lenzi@isti.cnr.it

Leonardo Piano, University of Cagliari - leonardo.piano@unica.it



References

A. Graciotti, Knowledge extraction from multilingual and historical texts for advanced question answering, in: C. d'Amato, J. Z. Pan (Eds.), Proceedings of the Doctoral Consortium at ISWC 2023 co-located with 22nd International Semantic Web Conference (ISWC 2023), Athens, Greece, November 7, 2023, volume 3678 of CEUR Workshop Proceedings, 2023.

R. Orlando, P.-L. Huguet-Cabot, E. Barba, R. Navigli, Relik: Retrieve and link, fast and accurate entity linking and relation extraction on an academic budget, arXiv preprint arXiv:2408.00103 (2024).

Llama Team, AI @ Meta, The llama 3 herd of models, 2024. URL: https://arxiv.org/abs/2407.21783. arXiv:2407.21783.

F. Petroni, A. Piktus, A. Fan, P. Lewis, M. Yazdani, N. De Cao, J. Thorne, Y. Jernite, V. Karpukhin, J. Maillard, et al., Kilt: a benchmark for knowledge intensive language tasks, arXiv preprint arXiv:2009.02252 (2020).