



KNOWLEDGE EXTRACTION FROM FICTION

USING GENERATIVE LLMS AND
MULTILINGUAL SEMANTIC PARSING



university of
 groningen

faculty of arts



European Research Council
Established by the European Commission

Federico Pianzola

text2story 2026, Delft

CULTURAL EVOLUTION



How do cultural traits spread, become successful, disappear?

TRANSMISSION OF STORY TRAITS



VARIANTS OF FICTIONAL CHARACTERS

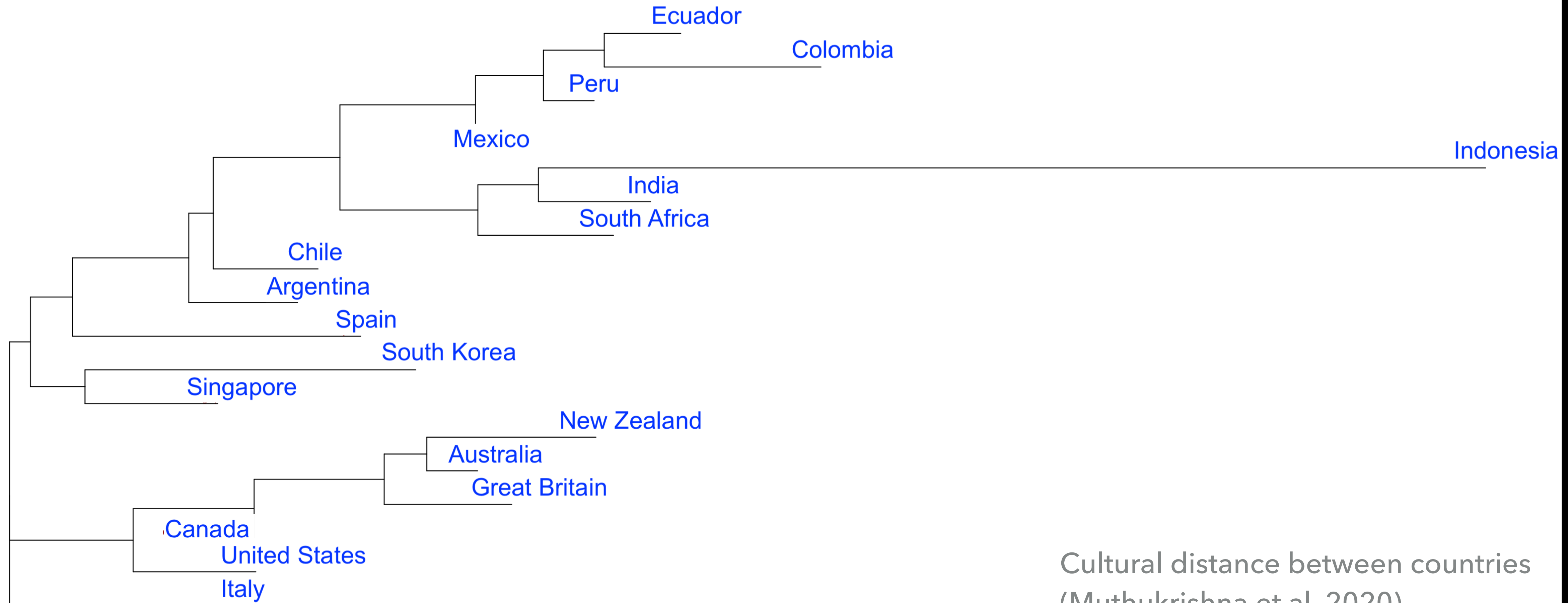


Zeus → ? ← Jupiter



CROSS-CULTURAL ANALYSIS OF FICTION

- ▶ 7 languages (EN, ES, IT, NL, KR, ID, CN)
- ▶ Various platforms (Wattpad, AO3, FFN...)



Cultural distance between countries
(Muthukrishna et al. 2020)

KNOWLEDGE EXTRACTION

(MULTILINGUAL) KE FROM FICTION

[...] a body was floating in the river [...]



(MULTILINGUAL) ABSTRACT MEANING REPRESENTATION (AMR)

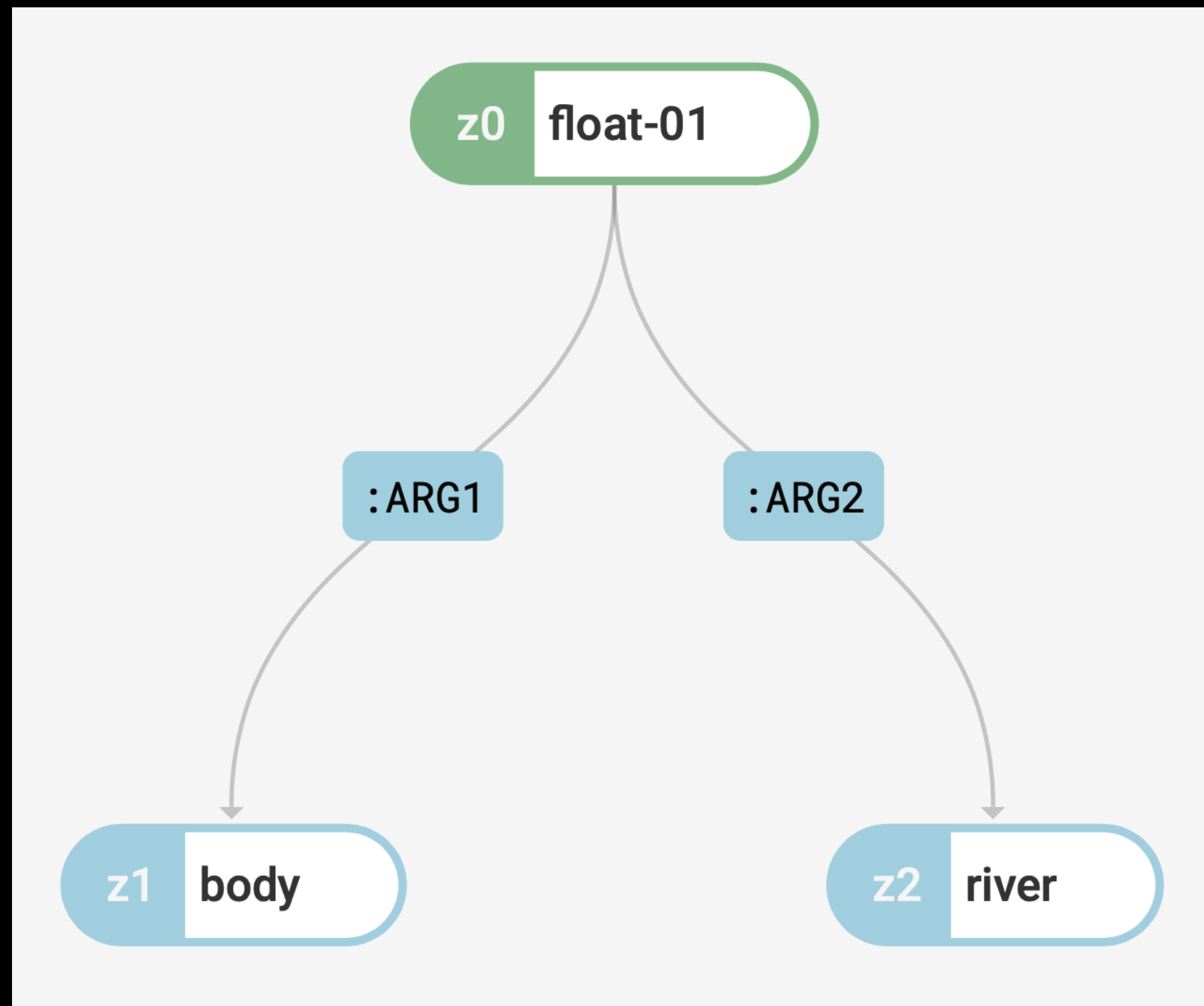
[...] a body was floating in the river [...]

EVENT:

(z0 / float-01

:ARG1 (z1 / body)

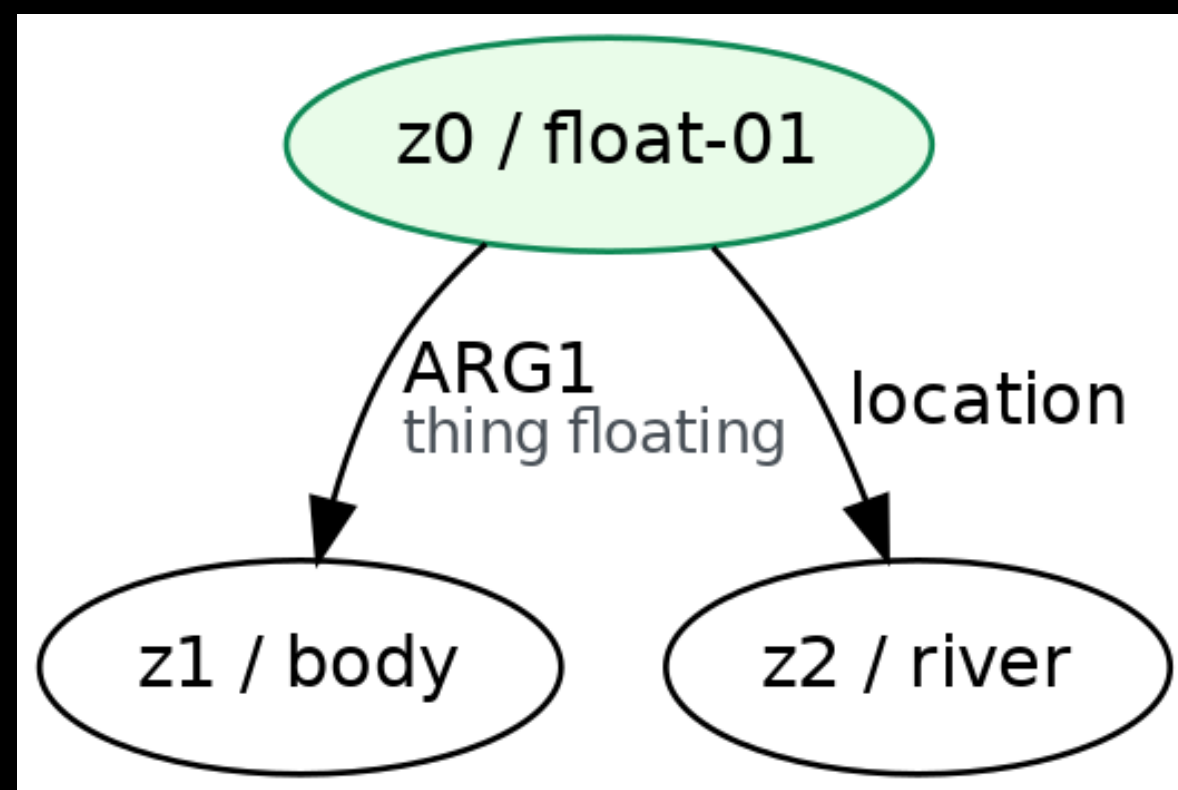
:ARG2 (z2 / river))



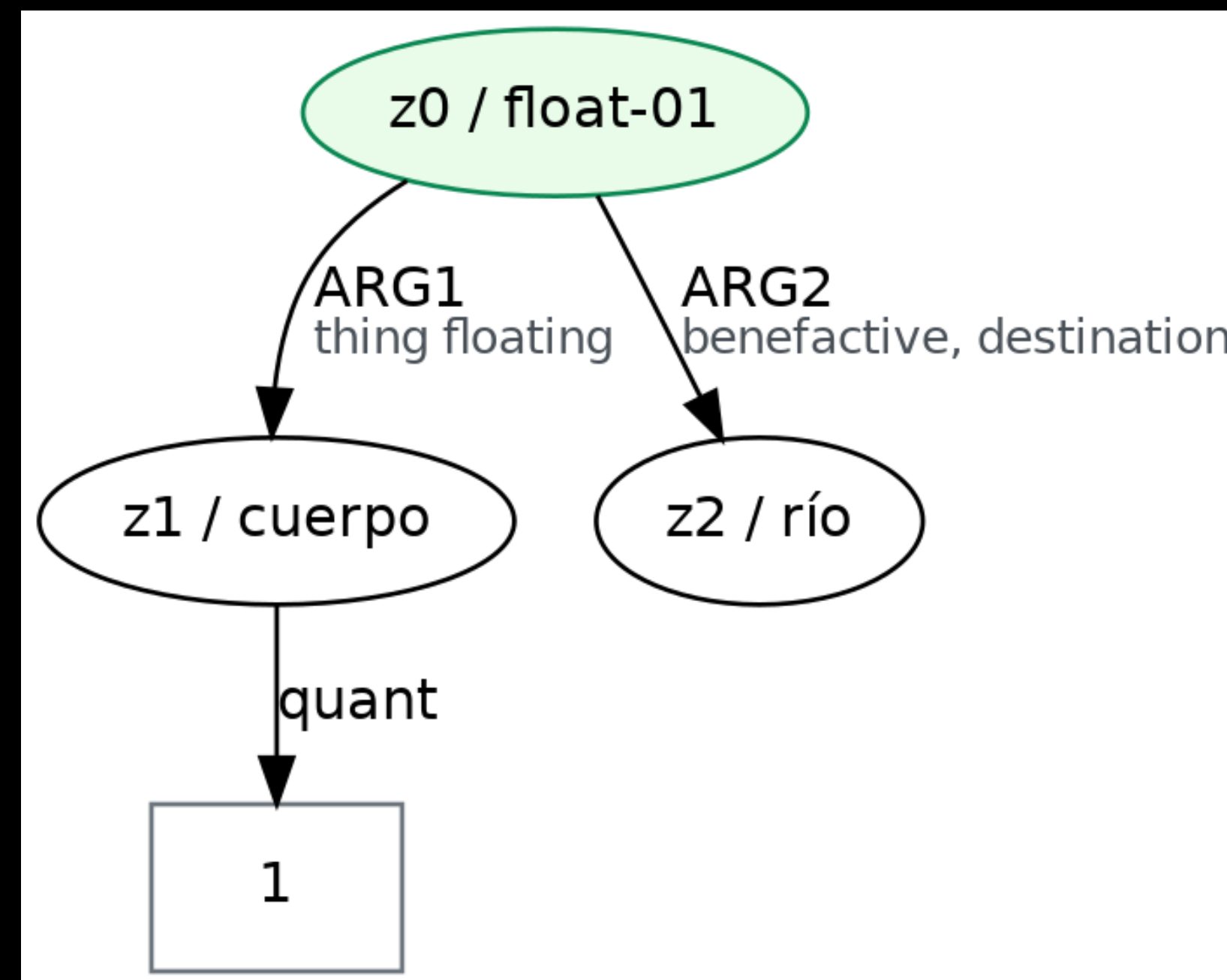
(MULTILINGUAL) ABSTRACT MEANING REPRESENTATION (AMR)

[...] a body was floating in the river [...]

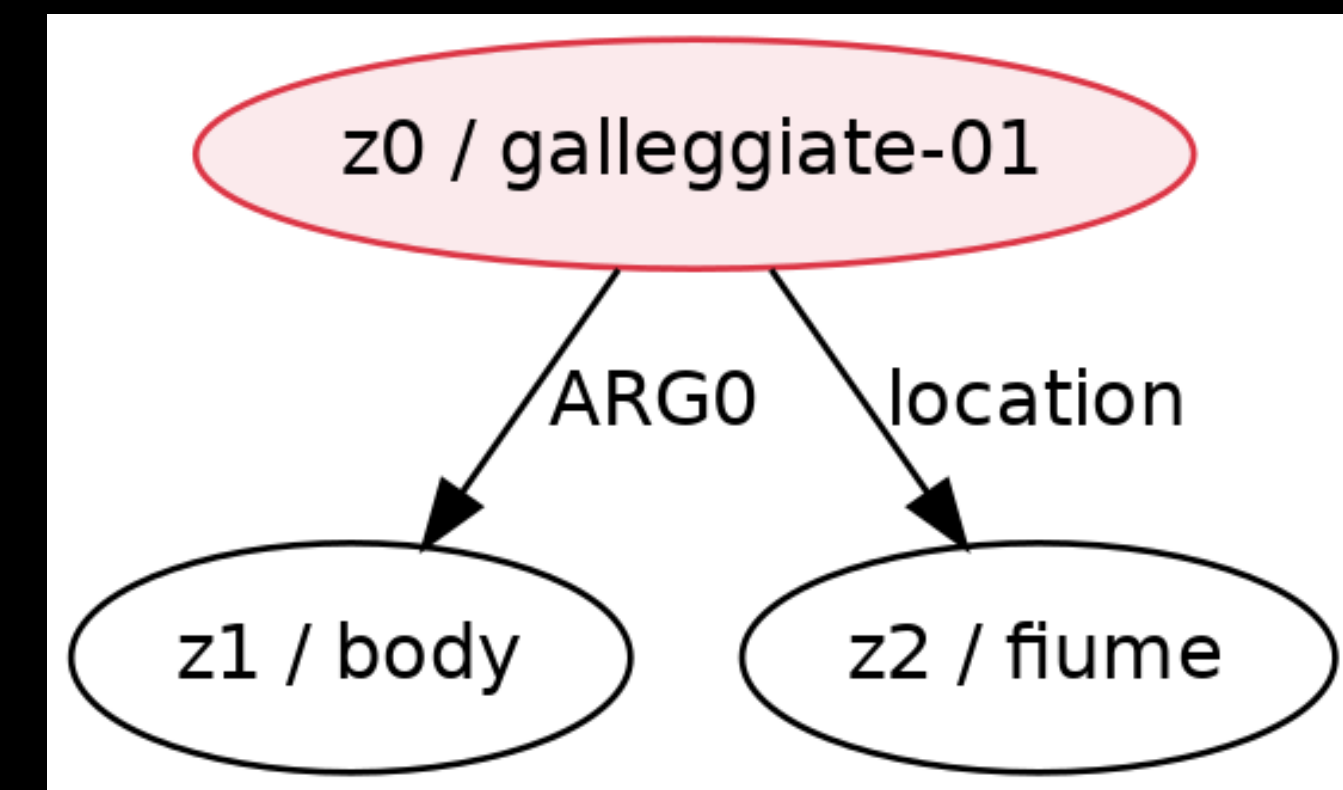
English



Spanish



Italian



KE FROM FICTION

[...] *Medusa* had killed *hundreds of men* with a single glance [...] She had *hundreds* hoarded in her own little corner of the island. [...] Even the pleasure of *turning men into stone* was beginning to bore her.

-
- Death event
 - Victim? *hundreds of men*
 - Perpetrator? *Medusa*
 - Mode of demise? *petrification*

KE FROM FICTION

#stories	#death events	#victims	#perpetrators	#modes of demise (all)	#modes of demise (unique)
59	234	218	160	168	63

Mode of Demise top 10

Mode of Demise	Count
slaying	13
stabbing	12
being eaten	11
petrification	10
being shot	9

Mode of Demise	Count
drowning	8
snake bite	7
war	7
tearing to pieces	6
sacrifice	6

KE WITH LLMS – MARCH 2025

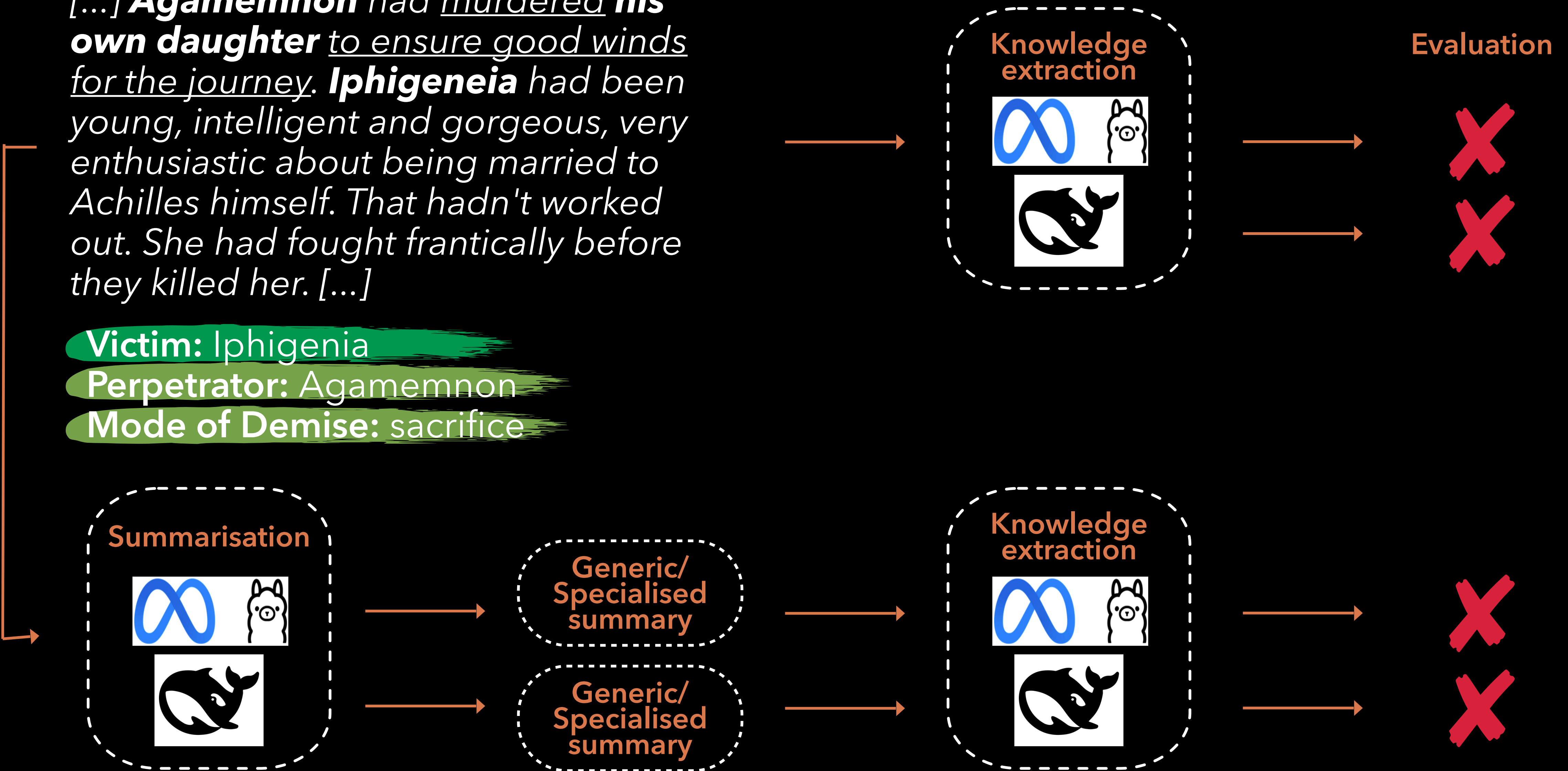
KE FROM FICTION WITH LLMs (70B)

[...] **Agamemnon** had murdered his own daughter to ensure good winds for the journey. **Iphigeneia** had been young, intelligent and gorgeous, very enthusiastic about being married to Achilles himself. That hadn't worked out. She had fought frantically before they killed her. [...]

Victim: Iphigenia

Perpetrator: Agamemnon

Mode of Demise: sacrifice



KE FROM FICTION WITH LLMS (70B)

Input	Victim			Perpetrator			Mode of Demise		
	P	R	F1	P	R	F1	P	R	F1
Generic Summary	<u>0.85</u>	0.27	<u>0.41</u>	<u>0.73</u>	<u>0.20</u>	<u>0.32</u>	<u>0.56</u>	0.17	0.26
Specialised Summary	0.87	0.43	0.58	0.76	0.33	0.46	0.66	0.22	0.33
Story	0.79	<u>0.35</u>	<u>0.49</u>	0.53	0.19	0.28	<u>0.56</u>	0.26	0.35

Table 4: Fine-grained evaluation metrics (Precision, Recall, F1) for Victim, Perpetrator, and Mode of Demise.



Knowledge extraction from Specialised Summary is best
....but NOT for the Mode of Demise



...followed by knowledge extraction from Story



But overall, **F1 <60%**

KE FROM FICTION WITH HUMANS

this mission was radio silence until The Avengers they were flying back home.

10 Peter Parker was sitting at lunch with Ned and MJ when

Two-Way | friend (Peter Parker to Ned)
Two-Way | Potential | friend (Peter Parker to Ned)
Two-Way | friend (Ned to MJ)

Peter Parker Emotion/mental state
he suddenly felt an overwhelming wave of worry;

Peter Parker Emotion/mental state Peter Parker
he felt his chest constrict; he looked around rapidly, but
couldn't find anything wrong

KE FROM FICTION WITH HUMANS

- ▶ Annotation of **characters** and **narrative events** for 7 languages (Bahasa Indonesia, Chinese, Dutch, English, Italian, Korean, Spanish)
- ▶ ~100k words per language
- ▶ Use annotations to evaluate and improve LMs for KE

CHARACTER FEATURES ANNOTATION

Property Type	Description	Examples	Stability
Biography	Stable life facts	Age, occupation, nationality, education, superpowers	Stable
Physical	Outer, perceptible characteristics	Height, hair color, clothing, facial features, smell	Stable/Transitory
Personality	Inner, stable traits	Shy, ambitious, kind, arrogant	Stable
Emotion/ mental state	Inner, temporary states	Happy, anxious, confused, relieved, embarrassed	Transitory

MULTILINGUAL SEMANTIC PARSING

KE WITH TEXT 2 AMR 2 KG (FRED)

Sirius was a 20 years old man, with long, brown hair and a smart face.

- ▶ Sirius
- ▶ 20 years old
- ▶ man
- ▶ long, brown hair
- ▶ smart face.

KE WITH TEXT 2 AMR 2 KG (FRED)

Sirius was a 20 years old man, with long, brown hair and a smart face.

```
query_gender = ""
```

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

```
[...]
```

```
SELECT ?character ?name ?gender
```

```
WHERE {
```

```
  # Find instances whose type is a subclass of Person (e.g., Man, Woman)
```

```
  ?character rdf:type ?characterType .
```

```
  ?characterType rdfs:subClassOf dul:Person .
```

```
  ?character rdf:type ?gender .
```

```
  VALUES ?gender { fred:Man fred:Woman }
```

```
  OPTIONAL { ?character schema:name ?name }
```

```
}
```

```
""
```

KE WITH TEXT 2 AMR 2 KG (FRED)

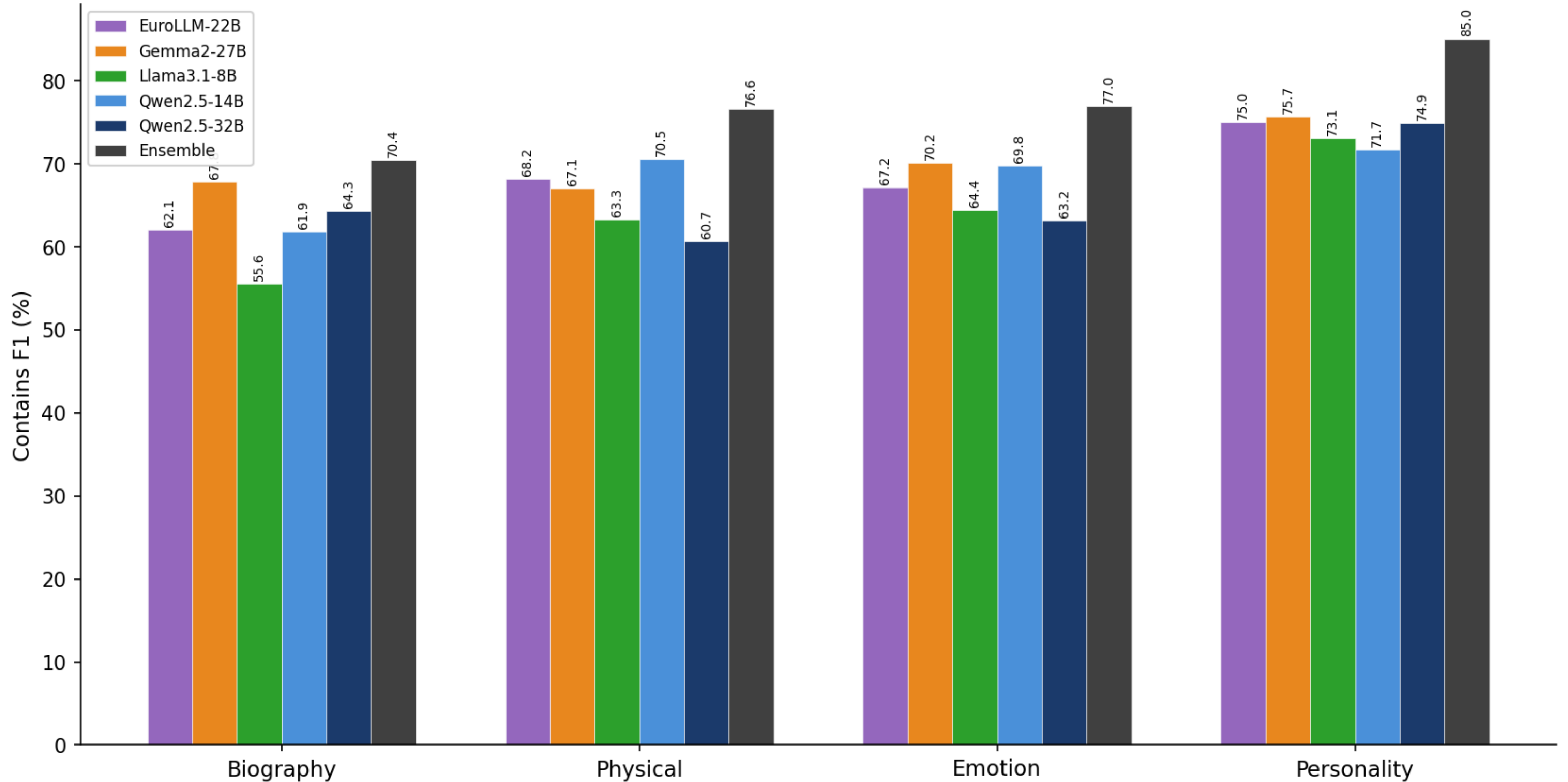
Problems:

- ▶ long parsing time (9 seconds, for "Sirius" example sentence)
- ▶ inaccuracy of multilingual parser
- ▶ KE constrained by predefined queries

KE WITH LLMS – MARCH 2026

KE WITH LLMS (ITALIAN)

Best Contains F1 per model per property



KE WITH LLMS

emotion_prompt_v1_fewshot_5_similarity

Model	EM F1	Cont F1	BS F1	Avg Time
ENSEMBLE	23.6%	77.0%	83.3%	0.00s
gemma2-27b-awq	21.4%	70.0%	62.7%	1.89s
qwen2.5-14b-awq	18.1%	69.8%	61.8%	1.17s
eurollm	16.1%	65.6%	57.7%	1.27s
llama3.1-8b-awq	15.8%	63.5%	56.2%	0.66s
qwen2.5-32b-awq	13.5%	62.8%	55.5%	2.52s

DYNAMIC FEW-SHOT LEARNING

- ▶ For each inference instance, the **K most similar gold examples** are injected into the prompt. Few-shot examples are drawn from the same property type only.
- ▶ **Similarity strategy**: examples are selected by cosine similarity on sentence embeddings. Rankings are precomputed once per property type and cached for all K values.

DYNAMIC FEW-SHOT LEARNING

gemma2-27b-awq

Setting	EM F1	Cont F1	BS F1	Avg Time
emotion_prompt_v1_fewshot_4_similarity	20.2%	70.2%	62.6%	1.80s
emotion_prompt_v1_fewshot_5_similarity	21.4%	70.0%	62.7%	1.89s
emotion_prompt_v1_fewshot_3_similarity	20.6%	69.6%	62.3%	1.70s
emotion_prompt_v1_fewshot_2_similarity	19.8%	69.4%	61.8%	1.62s
emotion_prompt_v1_fewshot_5_random	20.3%	68.4%	61.1%	1.81s
emotion_prompt_v1_fewshot_1_similarity	20.1%	67.7%	60.1%	1.53s
emotion_prompt_v1_fewshot_2_random	19.4%	67.4%	59.9%	1.60s
emotion_prompt_v1_fewshot_4_random	20.9%	67.3%	60.0%	1.74s
emotion_prompt_v1_fewshot_3_random	20.4%	67.2%	60.0%	1.67s
emotion_prompt_v1_fewshot_1_random	18.3%	66.2%	58.9%	1.52s
emotion_prompt_v1_zeroshot	19.1%	64.1%	57.1%	1.60s

DO MORE EXAMPLES ALWAYS IMPROVE RESULTS?

Contains F1 by K (similarity strategy), averaged across models.

K	biography	physical	emotion	personality	avg
0	46.9%	51.7%	58.4%	61.5%	54.6%
1	55.2%	60.8%	63.1%	68.2%	61.8%
2	58.1%	63.4%	64.3%	71.5%	64.3%
3	59.5%	64.4%	65.4%	72.3%	65.4%
4	61.2%	65.0%	66.5%	72.6%	66.3%
5	62.1%	65.9%	66.3%	73.6%	67.0%

CAN WE DO KE ONLY WITH LLMS?

CAN WE DO KE ONLY WITH LLMS?

NO, GROUND KNOWLEDGE IS NEEDED

KNOWLEDGE GRAPH

GOLEM-UI Search all content METADATA VIEW FANDOM VIEW POTTERVERSE FANDOM VIEW LITERARY QUALITY AND READABILITY VIEW FEEDBACK INFO INSTRUCTIONS EN

Metadata view **Results: 18710 Fanfiction stories**

Narrow down by:

Language

- en [16804]
- de [890]
- zh [243]
- ru [215]
- es [1041]

Fandom

- Harry Potter - J. K. Rowling [9805]

EXPORT

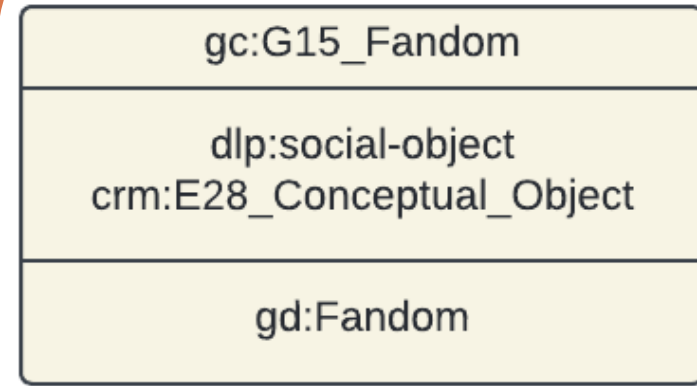
Rows per page 10 6231-6240 of 18710

URI	Title	Word Count	Kudos Count	Comment Count	Language	Publication date	Fandom
https://data.golemlab.eu/entity/W00017327	i've been a bad girl, daddy	1551	592	4	en	2018-07-14	One Direction (Band)
https://data.golemlab.eu/entity/W00017328	Jikook AU #2 (ABO)	5012	490	19	en	2018-07-12	BTS
https://data.golemlab.eu/entity/W00017329	sweet honey past your lips	2016	1085	32	en	2018-07-12	Yuri!!! on Ice (Anime)
https://data.golemlab.eu/entity/W00017330	Potential	727	93	4	en	2018-07-12	One Direction (Band)
https://data.golemlab.eu/entity/W00017331	The Care and Feeding of a Pack's Omega	8690	537	31	en	2018-07-11	My Hero Academia

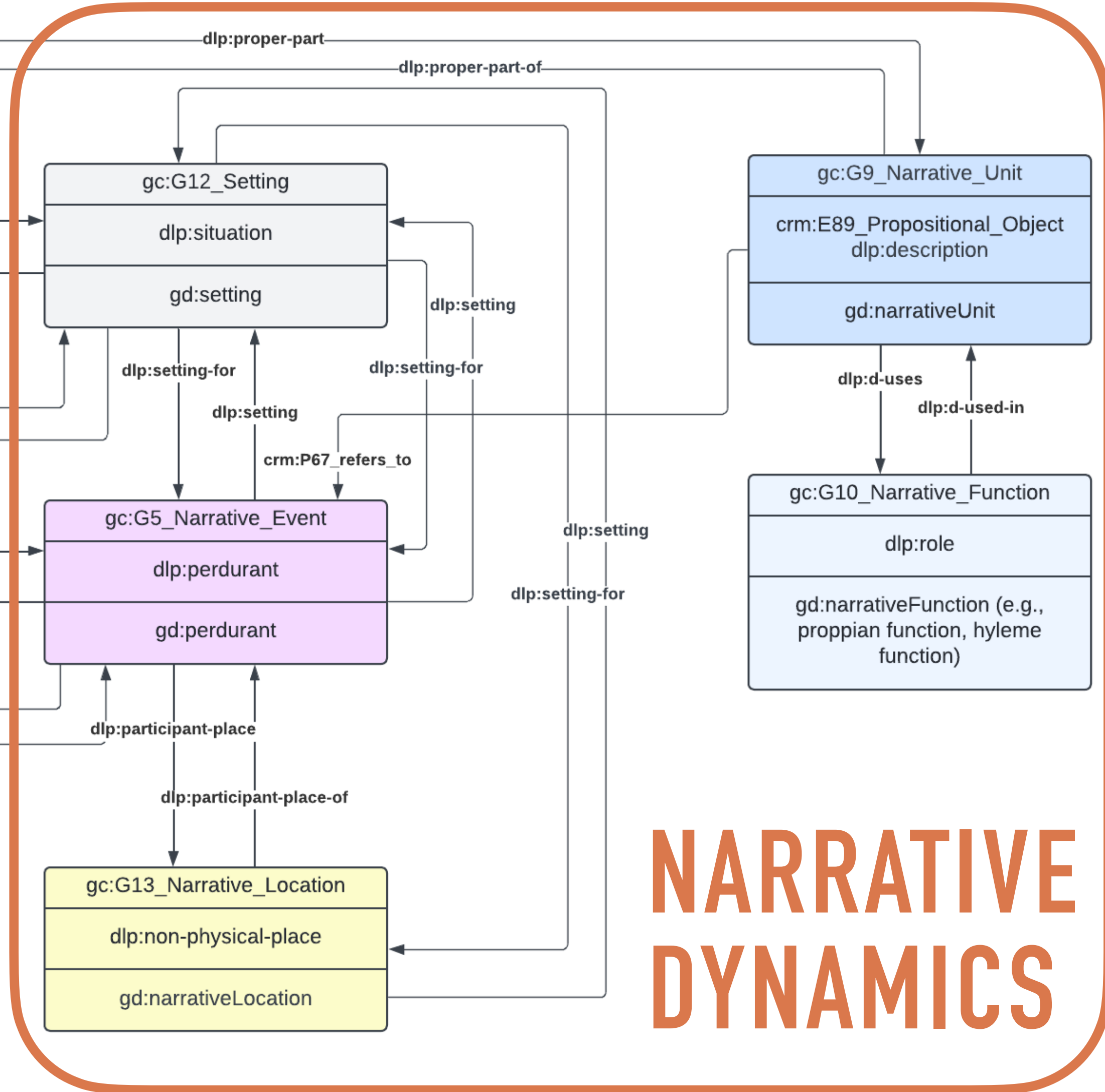
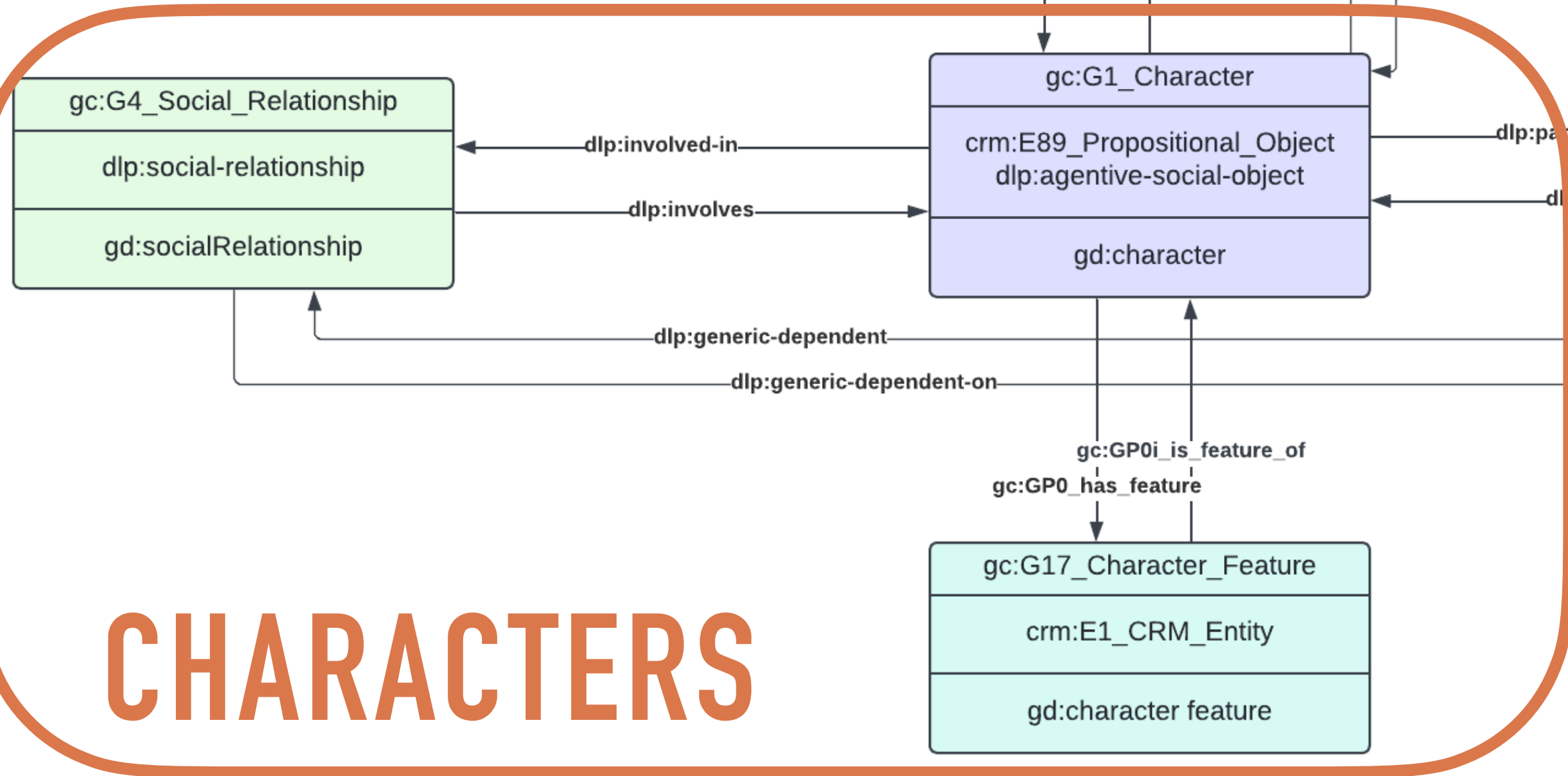
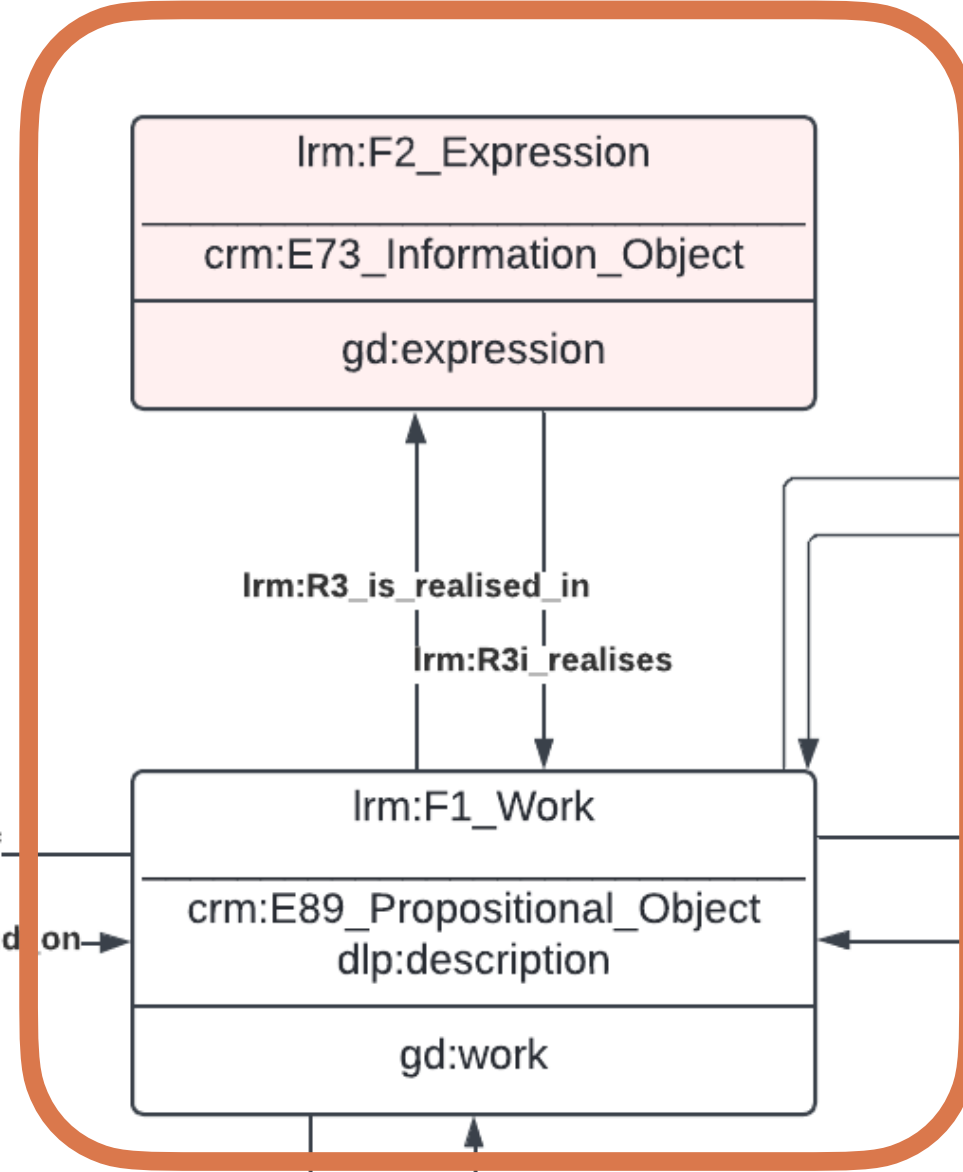
Publicly available at: <http://graph.golemlab.eu:8890/sparql/> (SPARQL endpoint) and <http://search.golemlab.eu:3006> (UI prototype)

ONTOLOGY

FANDOM



WORK



CHARACTERS

NARRATIVE DYNAMICS

GOLEM TEAM

- ▶ Andreas van Cranenburgh
- ▶ Franziska Pannach
- ▶ Arianna Graciotti
- ▶ Xiaoyan Yang
- ▶ Noa Visser Solissa
- ▶ Xiaoyu Zhou
- ▶ Luca Scotti
- ▶ Luotong Cheng
- ▶ 20 annotators and data curators

REFERENCES

- ▶ Graciotti, A., Pannach F., Presutti V., & Pianzola F. (2025). Llamas Don't Understand Fiction: Application and Evaluation of Large Language Models for Knowledge Extraction from Short Stories in English. *Computational Humanities Research 2025*. 4-32. <https://doi.org/10.63744/iCGYNUN0uUAe>
- ▶ Pianzola, F. Cheng, L., Pannach, F., Yang, X., & Scotti, L. (2025). The GOLEM Ontology for Narrative and Fiction. *Humanities*, 14(10), 193; <https://doi.org/10.3390/h14100193>
- ▶ Yang, X., & Pianzola, F. (2025). Fans Reconstruct Heroes: Modeling Fictional Characters in Participatory Culture. *Semantic Web - Interoperability, Usability, Applicability*. <https://www.semantic-web-journal.net/content/fans-reconstruct-heroes-modeling-fictional-characters-participatory-culture>
- ▶ Scotti, L., Pianzola, F., & Pannach, F. (2025). Grounding the development of an ontology for narrative and fiction. *Semantic Web - Interoperability, Usability, Applicability*. <https://www.semantic-web-journal.net/content/grounding-development-ontology-narrative-and-fiction>
- ▶ Pannach, F., Cheng, L., & Pianzola, F. (2024). The GOLEM-Knowledge Graph and Search Interface: Perspectives into Narrative and Fiction. *CHR 2024: Computational Humanities Research Conference, December 4-6, 2024, Aarhus, Denmark*. CEUR Workshop Proceedings 3834:462-471. <https://ceur-ws.org/Vol-3834/paper80.pdf>
- ▶ Pannach, F., Yang, X., Visser Solissa, N., Yu, Z., van Cranenburgh, A., van der Ree, M., & Pianzola, F. (2024). The GOLEM Triple Store: A Graph-Based Representation of Narrative and Fiction. *Semantic Methods for Events and Stories (SEMMES) Workshop*. https://anr-kflow.github.io/semmes/papers2024/SEMMES_2024_paper_3.pdf

THANK YOU!

f.pianzola@rug.nl

golemlab.eu

github.com/GOLEM-lab

