Unlocking Narratives

The Role of Knowledge Graphs and AI in Story Understanding

Pasquale Lisena
@pasqLisena

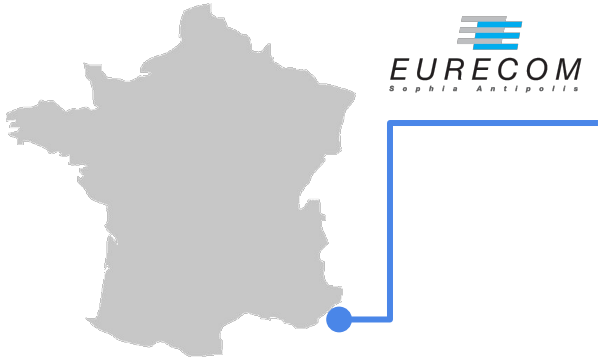
EURECOM
Sophia Antipolis

Who am I

Pasquale Lisena

http://pasqlisena.github.io/

@pasqLisena



We work on

- Semantic Web Technologies
- Knowledge modelling
- Information extraction
- Natural Language Processing
- Recommendation system

Credits

- Youssra Rebboud
- Mike de Kok
- Prof. Raphael Troncy

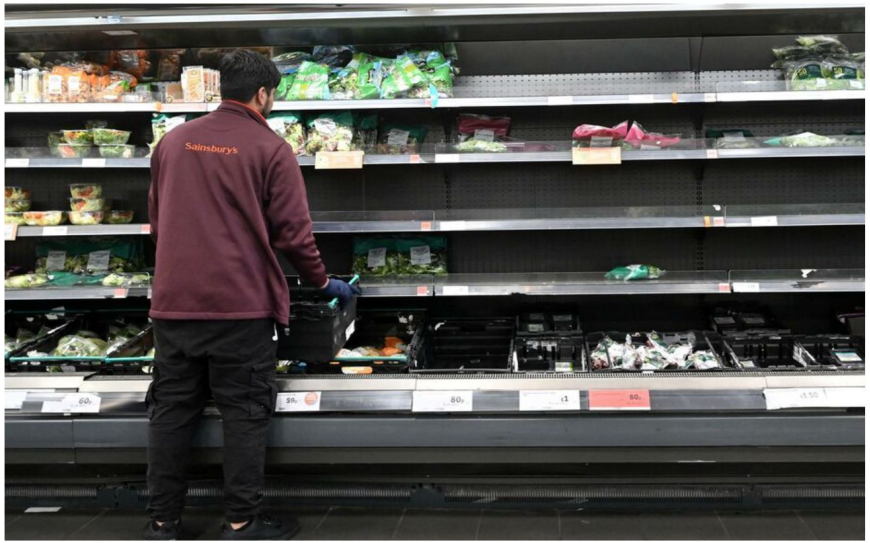
anr agence nationale de la recherche

kFLOW

International

Brexit : face aux pénuries, le Royaume-Uni accélère la formation de chauffeurs routiers

Les entreprises britanniques sont plombées depuis plusieurs mois par des problèmes d'approvisionnement, conséquence de la pandémie et du Brexit. Le gouvernement a annoncé vendredi l'accélération de la formation de chauffeurs poids lourds pour tenter d'y mettre fin.



Un employé réapprovisionne des étagères vides de laitue et de feuilles de salade à l'intérieur d'un supermarché Sainsbury's, à Londres, le 7 septembre

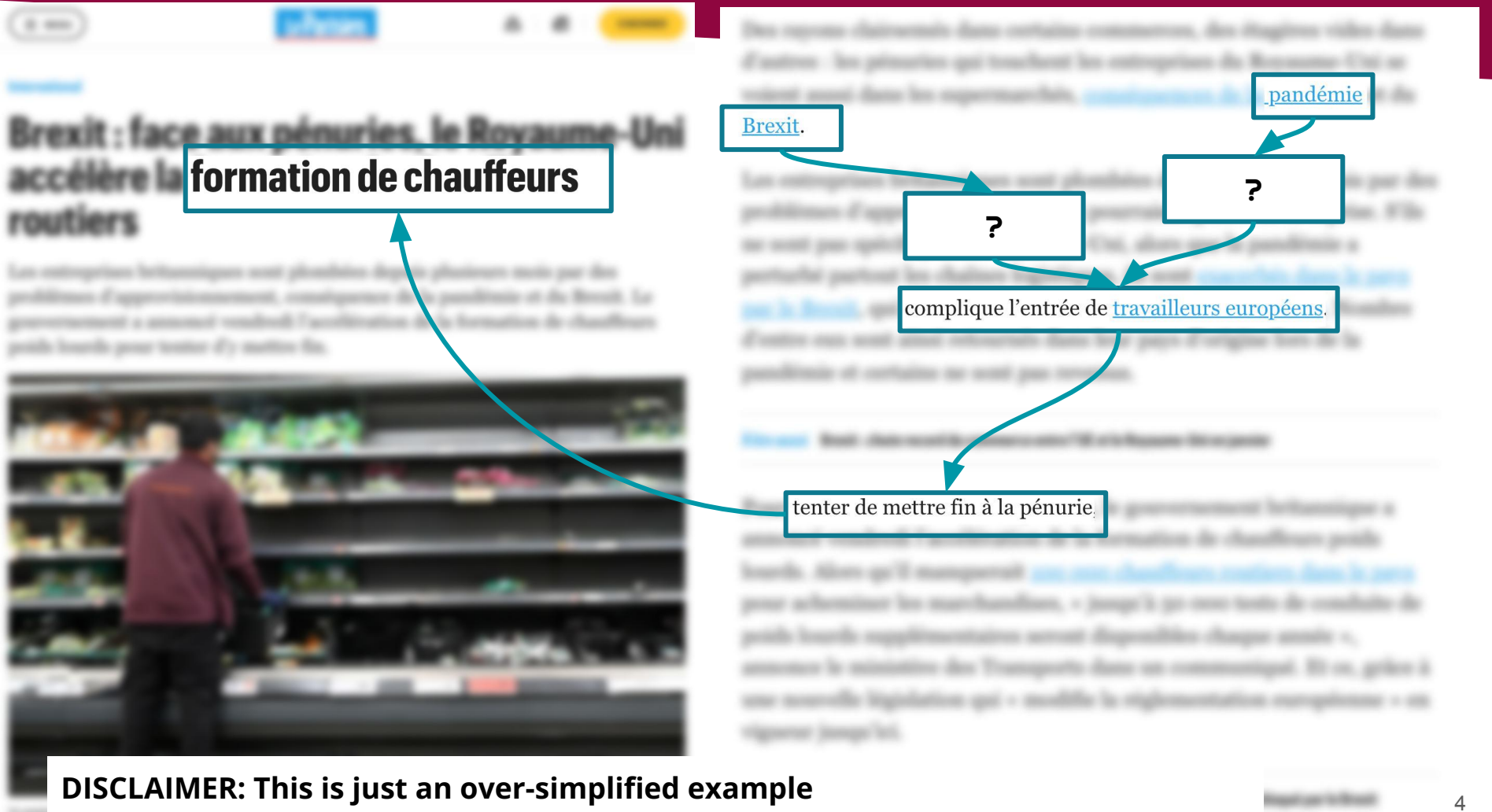
Des rayons clairsemés dans certains commerces, des étagères vides dans d'autres : les pénuries qui touchent les entreprises du Royaume-Uni se voient aussi dans les supermarchés, conséquences de la pandémie et du Brexit.

Les entreprises britanniques sont plombées depuis plusieurs mois par des problèmes d'approvisionnement qui pourraient peser sur la reprise. S'ils ne sont pas spécifiques au Royaume-Uni, alors que la pandémie a perturbé partout les chaînes logistiques, ils sont exacerbés dans le pays par le Brexit, qui complique l'entrée de travailleurs européens. Nombre d'entre eux sont ainsi retournés dans leur pays d'origine lors de la pandémie et certains ne sont pas revenus.

À lire aussi Brexit : chute record du commerce entre l'UE et le Royaume-Uni en janvier

Pour tenter de mettre fin à la pénurie, le gouvernement britannique a annoncé vendredi l'accélération de la formation de chauffeurs poids lourds. Alors qu'il manquerait 100 000 chauffeurs routiers dans le pays pour acheminer les marchandises, « jusqu'à 50 000 tests de conduite de poids lourds supplémentaires seront disponibles chaque année », annonce le ministère des Transports dans un communiqué. Et ce, grâce à une nouvelle législation qui « modifie la réglementation européenne » en vigueur jusqu'ici.

À lire aussi L'interminable attente d'un chauffeur routier britannique à Ouistreham, bloqué par le Brexit



DISCLAIMER: This is just an over-simplified example

Why it is important

TO UNDERSTAND STORIES

- **Interpret** the world
- **Connect** with the narrator
- Open the door to **discovery**
  (similarity, connected points, etc.)

TO NARRATE STORIES

- **Preserve** them (Heritage)
- **Knowledge Transfer**
- **Memorisation**

THE ROLE OF COMPUTER SCIENCE	Media access Misinformation Education

Unlocking Narratives

with AI + Knowledge Graphs

Unlocking Event Relations

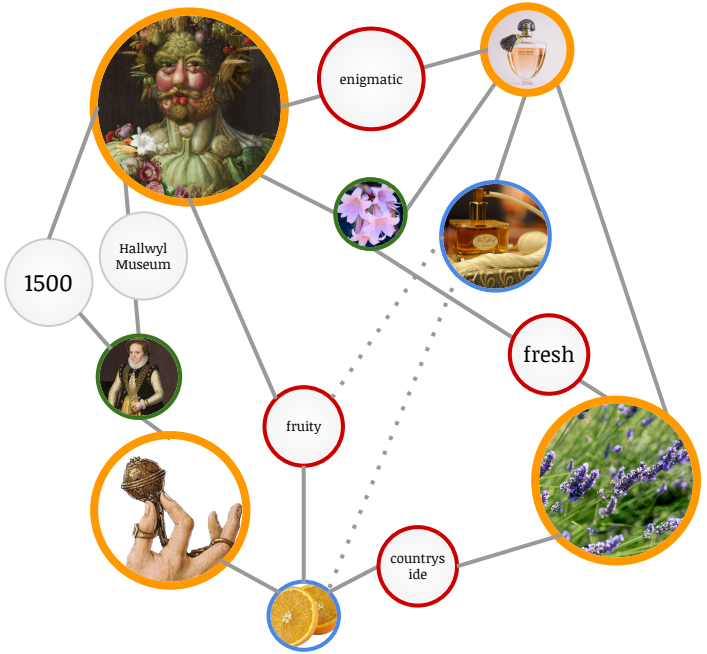
Unlocking Fact-checking

Unlocking Storytelling

What is a Knowledge Graph

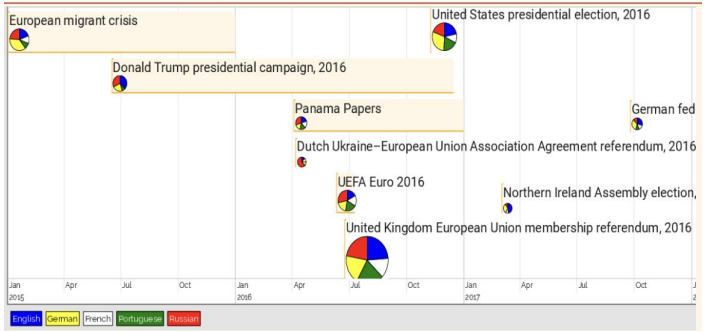
It is a specific kind of knowledge base which is:

- a **graph**
  connections between nodes are first-class citizens
- **semantic**
  the meaning of the connections are part of the data itself
- **smart**
  allows graph-computing techniques and algorithms
- **alive**
  easy to extend, access, reuse



source: https://hackernoon.com/wtf-is-a-knowledge-graph-a16603a1a25f

Unlocking Event Relations



Gottschalk, S., Demidova, E.: EventKG – the hub of event knowledge on the web and biographical timeline generation. *Semantic Web* 10, 1039–1070 (2019)

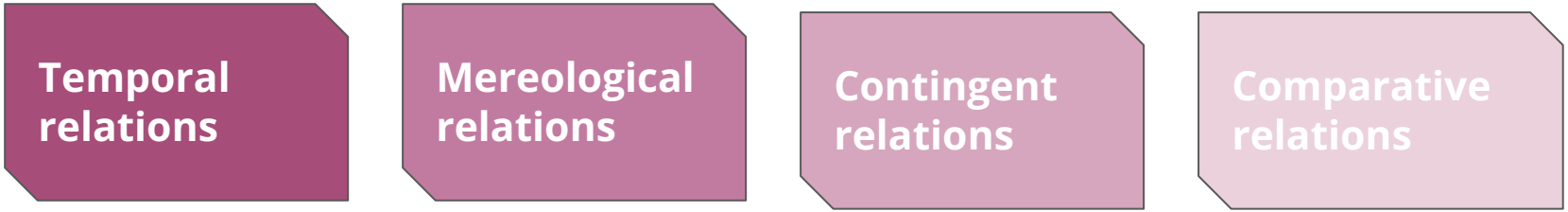
- OK to understand the **chronological order**

  BUT

- **Semantics** between relations are still **blurry**

Types of event relations

- **Event relation extraction** from **textual data** were vastly explored in the literature
- Four major types of event relations observed in **literature**:



Chronological order of two events {before, overlaps, during, etc}

Interaction between sub-events and super-events {sub_event}

{Causality, enabling, prevention, despite}.

{Competition, opposite, etc.}

What about the literature

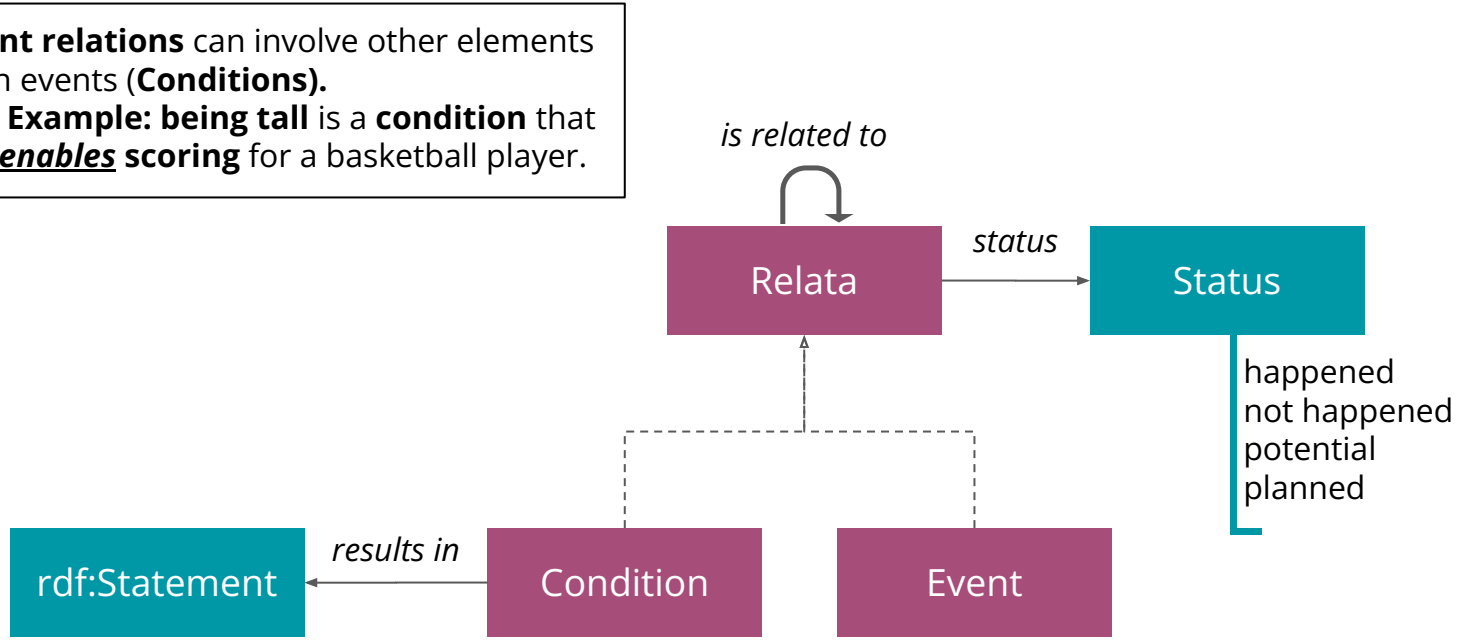
- **Temporal relations** (in ontologies and datasets) ✅
- Direct **causality** ✅

- **Other kind** of event impacting each others 🚫
- Difference between **cause, enabling** and **prevention**
- **Intentionality**

FARO Ontology

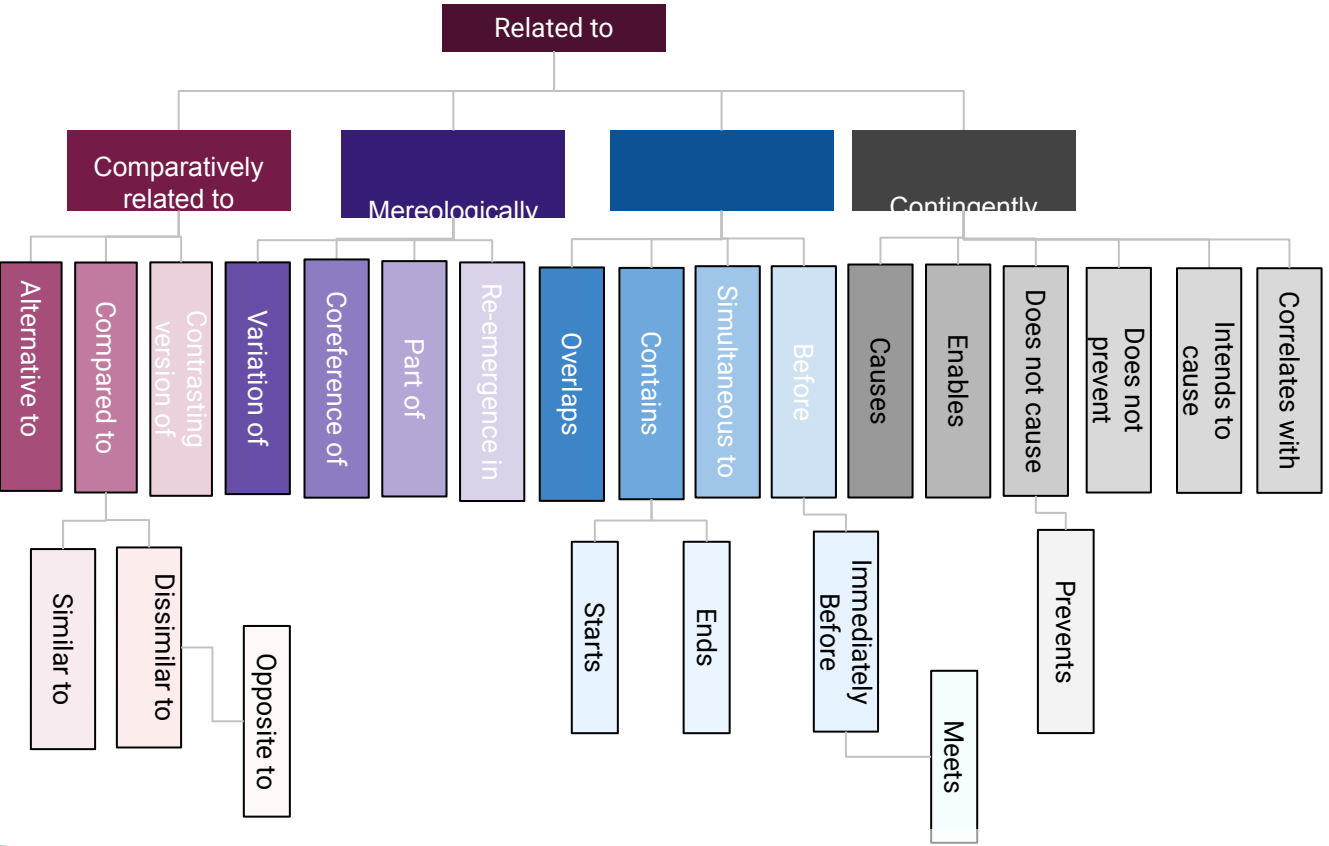
FARO: Representing Events

- **Event relations** can involve other elements than events (**Conditions).**
  - **Example: being tall** is a **condition** that ***enables*** **scoring** for a basketball player.



Y. Rebboud, P. Lisena, R. Troncy. Beyond Causality: Representing Event Relations in Knowledge Graphs. In Knowledge Engineering and Knowledge Management (EKAW), Bolzano, Italy, 2022,

FARO: Relations

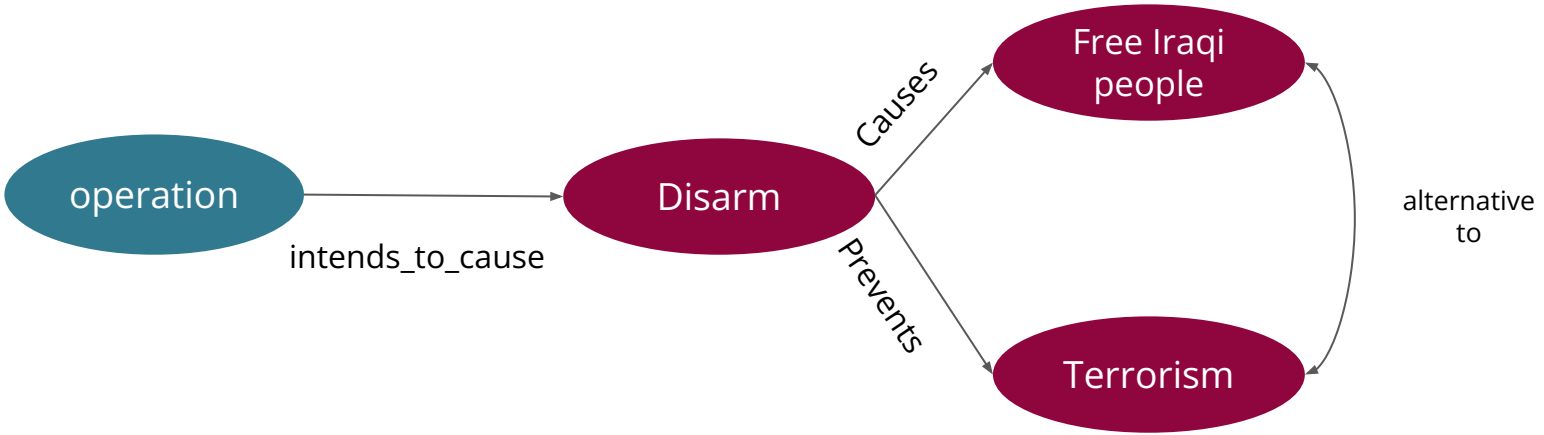


- **FARO** : developed to be as much **complete** as possible.
- **Harmonizes** other data models
- **Enable reasoning**
  - **Hierarchical structure** of properties
  - **Logic constraints** (owl properties)

Y. Rebboud, P. Lisena, R. Troncy. Beyond Causality: Representing Event Relations in Knowledge Graphs. In Knowledge Engineering and Knowledge Management (EKAW), Bolzano, Italy, 2022,

Example

“As US claimed, the **intent** of the military operation was to disarm Iraq of weapons of mass distruction, to **end** support for terrorism and free iraqi people”



Disclaimer: this is the representation of the statement from the text, without judgement whether it is true or false.

Problem

- **Not existing dataset** with precise event relations
- **Our first attempt** resulted in **small** and **unbalanced** dataset.

- **Two data augmentation strategies**
  a. With GenAI
  b. With Common Sense

Target relations:
- Enabling
- Intention
- Prevention
- Direct causality

Prompt Based Data Augmentation with LLMs

Prompt(ERx) = definition(Event) + definition(ERx) + request(ER) + examples(ERx)

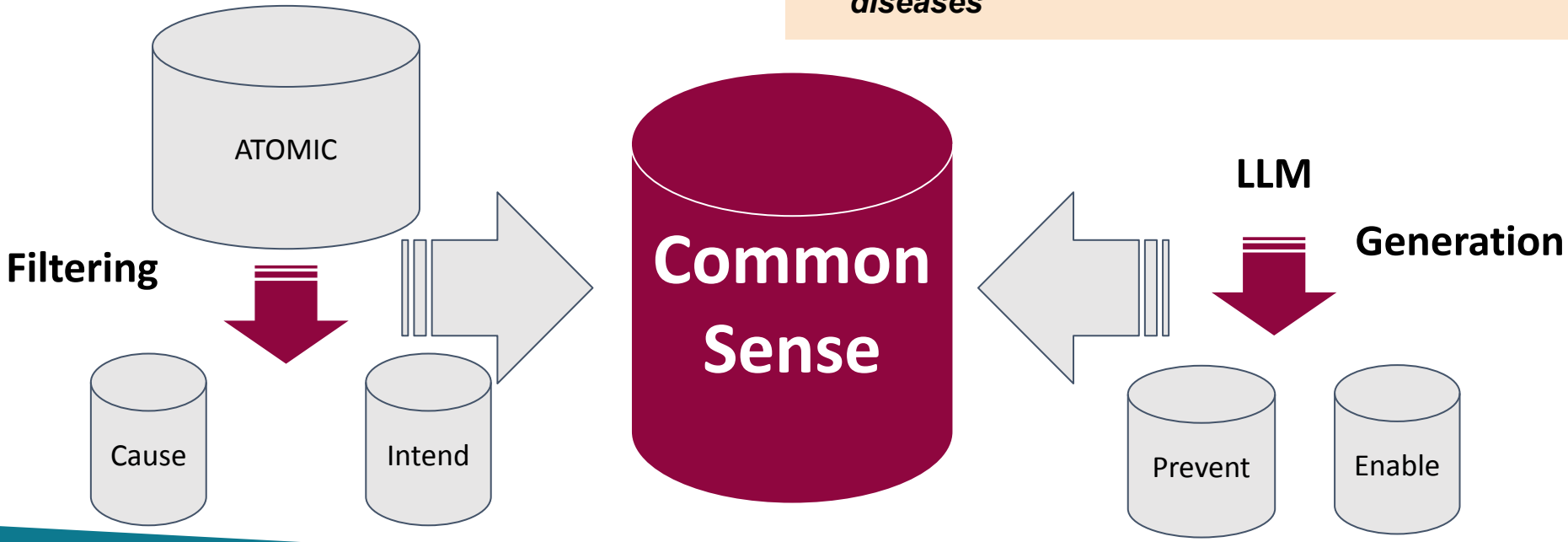
New dataset size: **2,000+ sentences**
Performance increment (F1):
Relation Classification **+27%**
Event Extraction: **+11%**

Y. Rebboud, P. Lisena, R. Troncy. Prompt-based Data Augmentation for Semantically-precise Event Relation Classification. In Semantic Methods for Events and Stories Workshop (SEMMES), ESWC 2023, Heraklion, Greece, May 2023

Common Sense Data Augmentation

Example of Common Sense Data

exercising *regularly prevents* ***cardiovascular diseases***



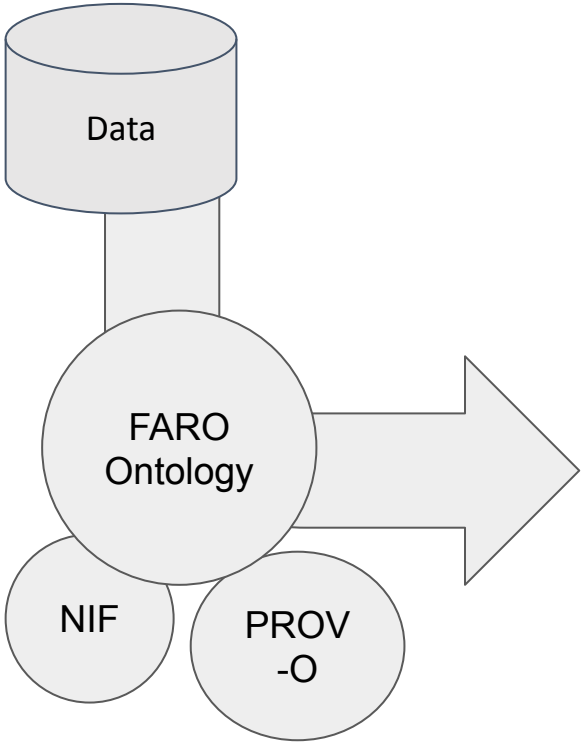
Final Dataset

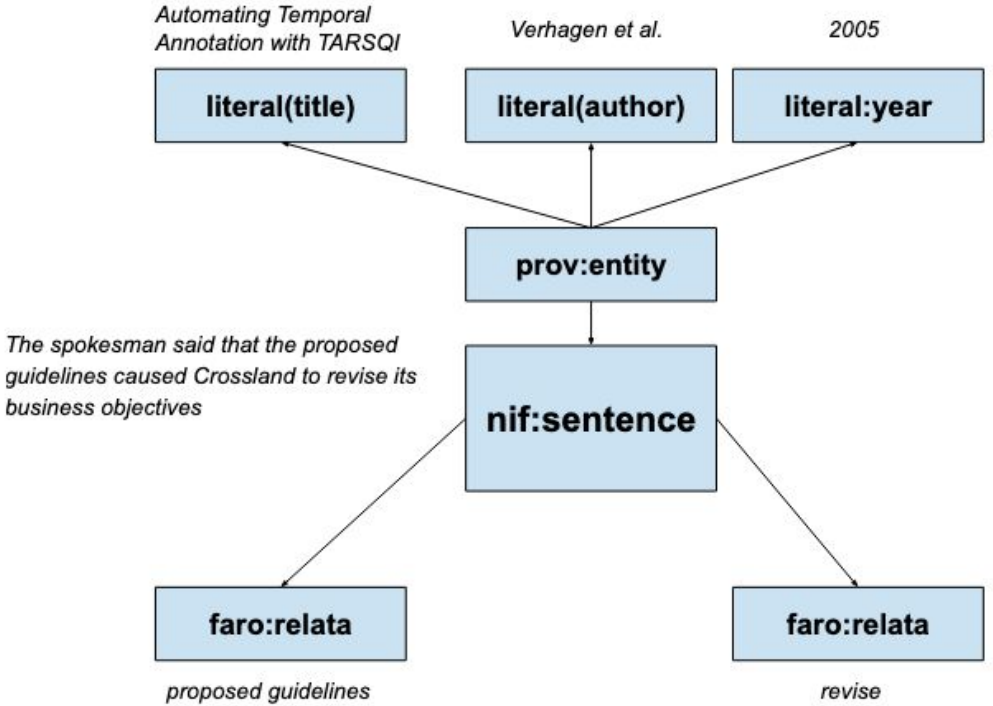
TOTAL	Cause	Enable	Prevent	Intend	No relation
6792	3520	814	948	944	566

Three subtasks

Subtask	**Best performing model**	**F1 Score**	**LLM (GPT4o)**
Relation Detection *Is this sentence including a relation?*	RoBERTa-based end-to-end classifier	0.98	0.59
Relation Classification *Which relation type is in this sentence?*	RoBERTa-based end-to-end classifier	0.78	0.54
Event Extraction *What are the text token involved?*	REBEL end-to-end	0.70	0.45

Knowledge Graph



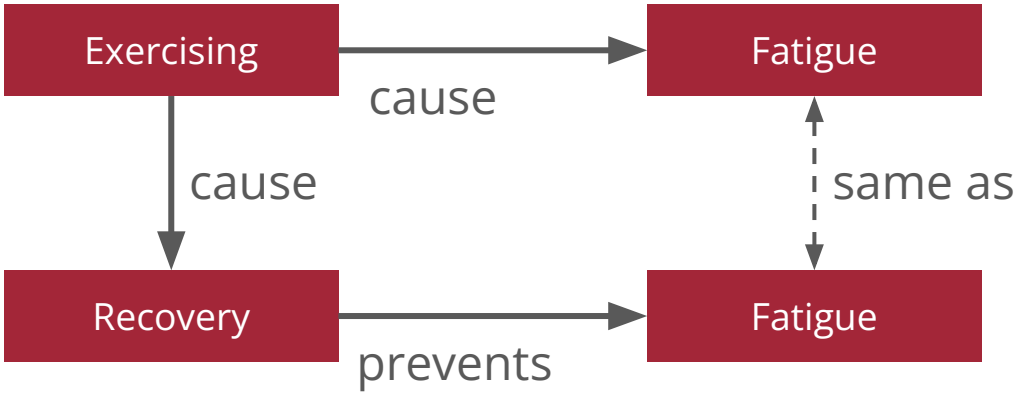


Unlocking Fact-Checking

How can event relations support fact-checking

Claim: *Exercising* daily **causes** muscle *fatigue* over time.

Evidence: Research shows that daily low-intensity *exercise* **activates** *recovery mechanisms* in the body, **preventing** the onset of chronic muscle *fatigue* and improving overall stamina instead.



Challenges

- Where to find rules?
- How to find connections between claims and evidences?

Where to find rules?

prevents[op]

IRI: http://purl.org/faro/prevents

Connect a Relata entity with the event for which is the cause of not happening.

Example: the *strike* was sufficient to block the *changement in working conditions*.

has characteristics: asymmetric

has super-properties

does not cause [op]

has range

Event [c]

Sub-cases

- Logical Alignment
- Logical Misalignment
- Causal loops
- Cherry-picking scenarios

How to find connections between claims and evidences?

SIMILARITY	Cosine similarity between SentenceBERT embeddings + threshold
POLARITY	DistilBERT for sentiment analysis

Advantages

- Logic check
- Complementary to existing methods
- Interpretability

Limitations

- Applies only if there are event relations
- Depends on the performances of the extraction
- Known issues: double negation
- Not yet ready to check in a full dataset

Unlocking Storytelling

Narrative Graphs

- Captures entities and **interconnected** links.
- Enabling an understanding of the relationships between events and **facilitating storytelling**
- They covers information about the **4W**:
  - What (event)
  - Who (actor)
  - Where (place)
  - When (time)
- **Lack** of more **semantically rich** event relations

Build a semantically rich Narrative Graph

Starting Point: ASRAEL KG

- Contains news articles with links to Wikidata events
- Extract the **4W** information from Wikidata for each event article
  - Follow the *owl:sameAs* (Wikidata link to event)

Use Event Relation Extraction (REBEL) for:

- Precise Event Spans
- Semantically Precise relations

Use Event Coreference resolution (EECEP) for:

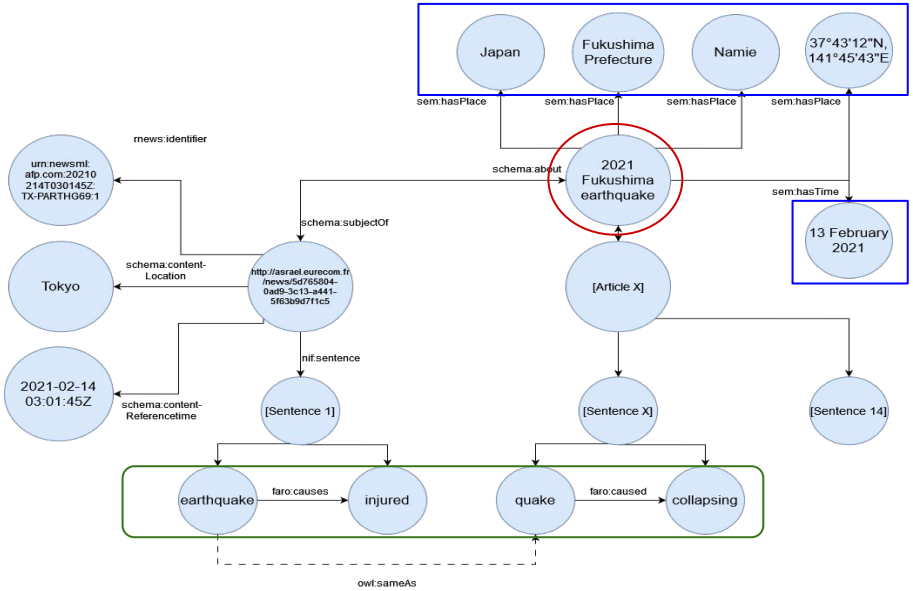
- Merge the same events when appear in different sentences/articles

Rudnik, C., Ehrhart, T., Ferret, O., Teyssou, D., Troncy, R., Tannier, X.: Searching news articles using an event knowledge graph leveraged by wikidata. In: Companion Proceedings of The Web Conference, 2019.

Information Selection + Text Generation

- A SPARQL query has been used to extract the essential nodes for the a given article.
  - Select the **Date, location, actor** of the article.
  - Select the **mentions** (**events**) from the **sentences** of the article

- This query prioritizes the selection of entities with higher frequencies of incoming edges.
- We finetune a **JointGT** (based on T5) on our KG to generate text.



M. de Kok,, Rebboud Y., Lisena P., Troncy R. & Tiddi, I. From Nodes to Narratives: A Knowledge Graph-based Storytelling Approach. In Seventh International Workshop on Narrative Extraction from Texts (Text2Story), colocated with ECIR 2024, March 24th, 2024, Glasgow, UK

Findings

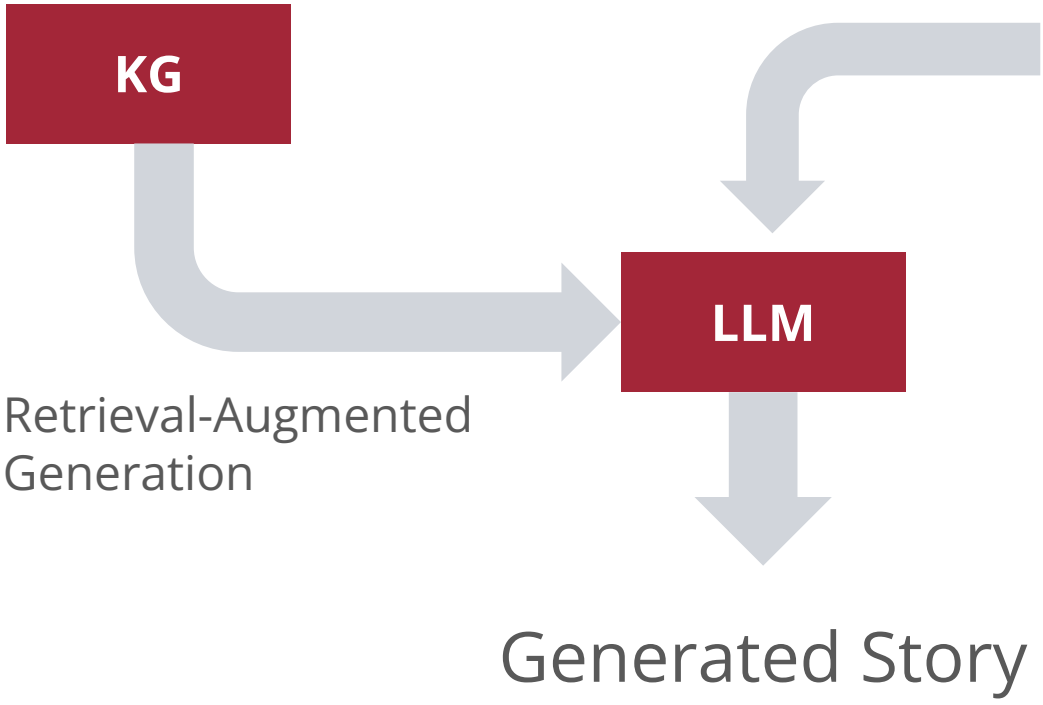
Metric	Base JointGT	Finetuned JointGT
BLEU	0.6529	0.6101
METEOR	0.4681	0.4409
ROUGE	0.7535	0.7260

Task	Fluency			Adequacy		
	Win %	Lose %	Tie %	Win %	Lose %	Tie %
Manually annotated article	33.3	16.7	50.0	**58.3**	8.3	33.3

Triples	Label	Base JointGT	Finetuned JointGT
(Demand, cause, benefited)	The company benefited from continued strong demand and higher selling prices for titanium dioxide, a white pigment used in paints, paper and plastics.	benefited is the cause of the demand	The company said it benefited from the strong demand for its products and services from a growing number of customers.

What’s next

Interactive storytelling



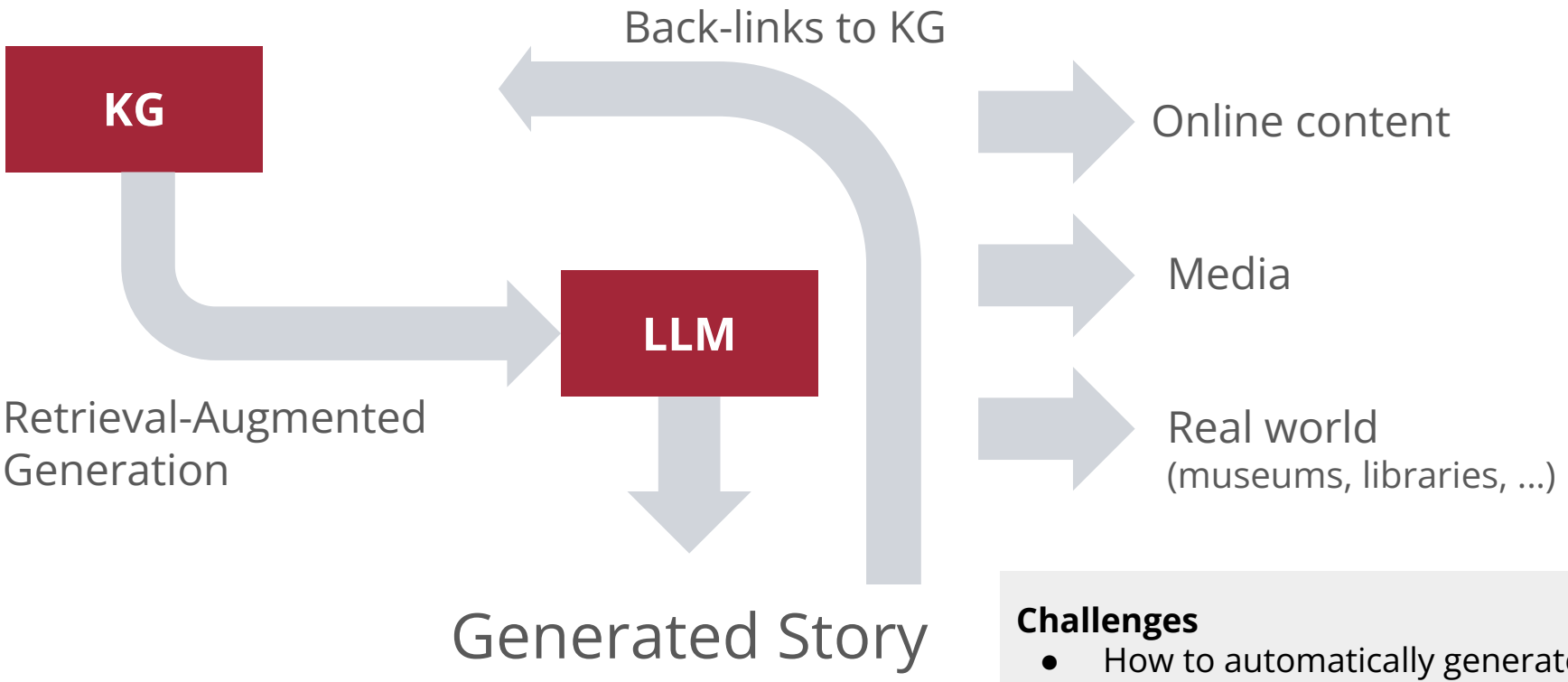
HUMAN-COMPUTER INTERACTION

Goal of the narrative
- Education
- Memorisation

Personalisation on the storyteller
- Style
- Metaphors

Personalisation on the Listener
- Cultural gap
- Generational gap
- Preferences

Enriched storytelling



Challenges

- How to automatically generate?
- How to guide and control this generation?

Takeaways

- **AI and Knowledge Graphs** are key elements for understanding, analyzing, and generating narratives.
- Including **semantically precise event relations** can improve the story understanding and increase the performance of downstream applications
- **Data augmentation strategies** using generative AI and common sense are effective in this domain
- **LLM** will become even more central in **future research in storytelling,** but still challenges open

pasquale.lisena@eurecom.fr

Thank you!

Q&A

This presentation: bit.ly/kflow-momi2025