

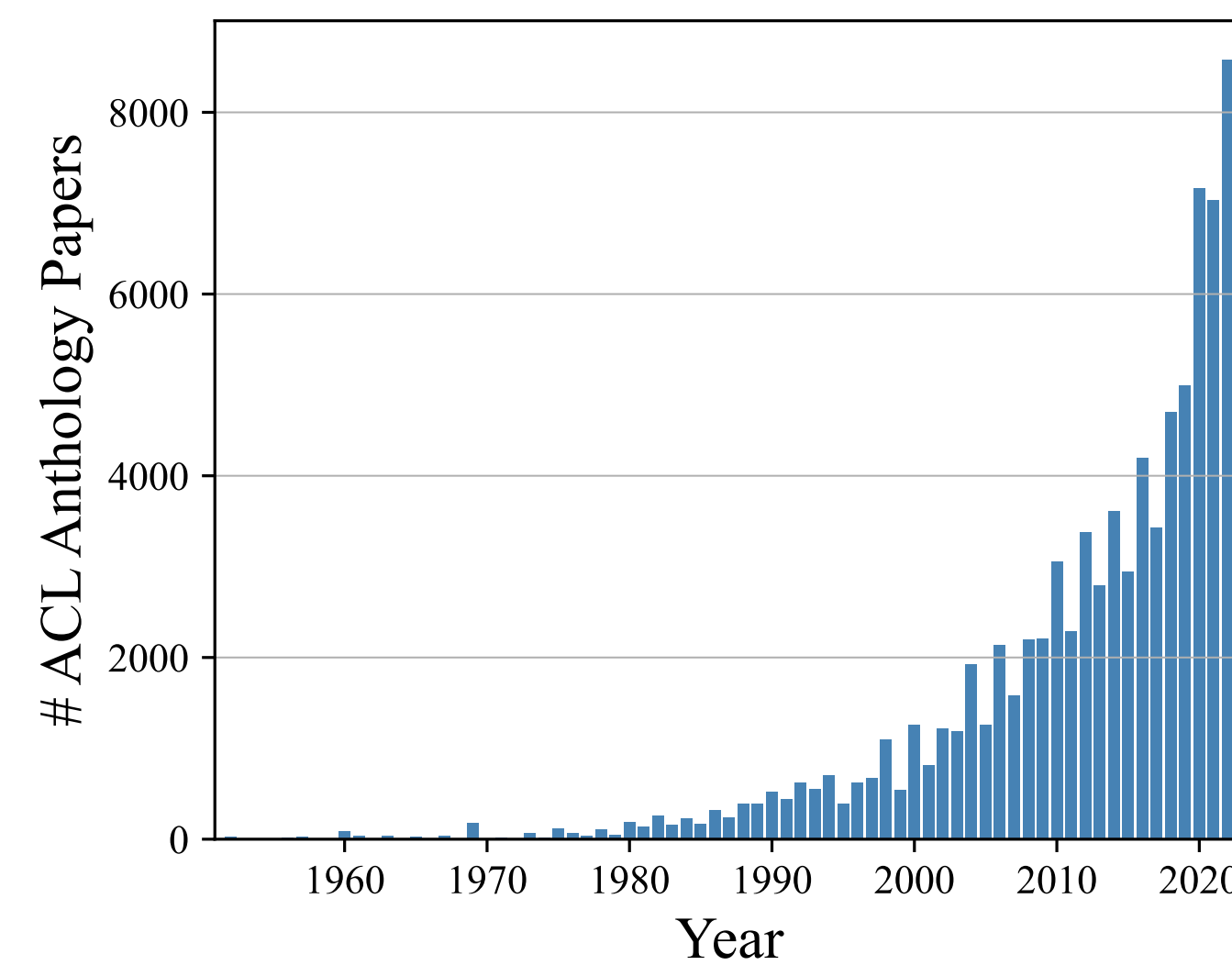
# A Knowledge Graph for NLP Research

Tim Schopf and Florian Matthes

{tim.schopf,matthes}@tum.de

## Motivation

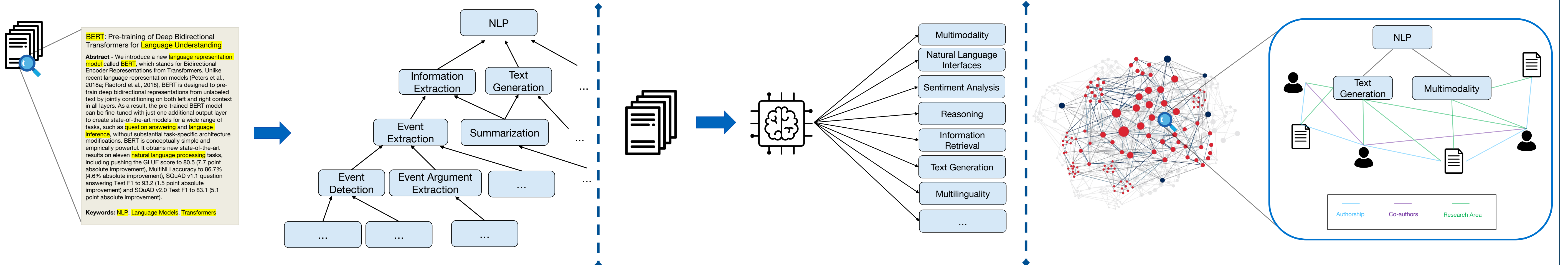
- Scientific knowledge is usually available in large quantities as **unstructured texts**.
- This makes it **difficult to get an overview** of new or unknown scientific fields.
- Furthermore, it is **challenging to stay up-to-date** with newly published research.
- This is a growing challenge in NLP, where **new research** is being **published** at an **exponential growth** rate.
- Structuring the NLP knowledge** by linking semantically related NLP fields offers the potential for **enhanced exploration** of the domain.



## Exploratory Search

- Process of **obtaining insights within a new domain** and is often directed towards a complex **open-ended goal**.
- Exploratory Search** (open-ended goal)  $\neq$  **Information Retrieval** (specific goal)
- Exploratory Search is especially important for researchers:
  - What **approaches** are currently **state-of-the-art**?
  - Who are the most **important researchers**?
  - What are **current research trends**?
  - ...

## Natural Language Processing Knowledge Graph



### 1. Construction of an NLP ontology:

Semi-automatic approach that uses LLMs, specialized fine-tuned NER and RE models, and domain experts to construct an ontology of NLP concepts from scientific publications.

### 2. Classification of research publications:

Use few-shot and zero-shot approaches to automatically assign NLP research publications to their respective concepts in the NLP ontology based on titles and abstracts.

### 3. NLP Knowledge Graph:

The NLP Knowledge Graph contains „Field of Study“, „Publication“, and „Researcher“ entities. Researchers are connected to NLP concepts based on their research publications.

### Research Hypothesis:

Structuring natural language processing publications into a knowledge graph comprising semantic relationships between scholarly entities can help researchers efficiently search, discover, and explore new and relevant knowledge.

### Objective:

The NLP knowledge graph is an approach to tackle the information overload challenge of researchers. It aims to help researchers in obtaining an overview of NLP-related topics and find relevant papers more efficiently. A user-friendly web application and a natural language conversational interface will be developed to make the NLP knowledge graph easy to use.

## NLP Knowledge Graph Construction

### NLP Ontology Construction

- Use **domain experts** to construct an accurate high-level ontology of NLP concepts.
- Prompt **LLMs** to generate further NLP-related triples from their implicitly learned knowledge base.
- Train **specialized fine-tuned NER and RE models** to extract additional entities and relationships from NLP publications.
- Construct the final NLP ontology by **merging the domain expert ontology, the LLM triples, and the extracted entities and relations from NLP publications**.
- Use a **human-in-the-loop** approach to evaluate and adjust the results of the different steps in the NLP ontology construction pipeline.

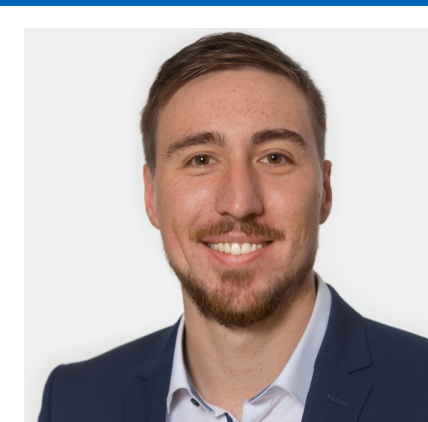
### Assigning Publications to NLP Concepts

- Use a **weakly trained classifier** to categorize publications according to the small number of high-level NLP ontology concepts created by domain experts.
- For the large number of sparse lower-level NLP concepts, use **few-shot and zero-shot classification** approaches. In addition, citation information can be used to assign articles to the same NLP concept.

## Comparison to Existing Approaches

- In contrast to commonly used scholarly search engines like e.g. google scholar or semantic scholar, this approach is highly **domain-specific**.
- NLP concepts** allow for explicit searching and exploration of articles for specific subfields of NLP rather than the usual practices of keyword-based searching and exploration of articles based on citation information.
- The **hierarchical relationships** between NLP concepts allow easy navigation from well-known hypernym concepts to more specific hyponym concepts. This facilitates exploration and gaining an overview of NLP fields.
- Scientific relationships** between NLP concepts show how different subfields of NLP are interconnected and facilitate understanding of the relationships between different fields.
- Provides very **accurate** domain-specific knowledge and **doesn't hallucinate**.

## Contact



Tim Schopf  
 Technische Universität München, sebis  
 tim.schopf@tum.de  
 +49 89 289-17105

