

# Domain-Specific Knowledge Graph Construction for Semantic Analysis

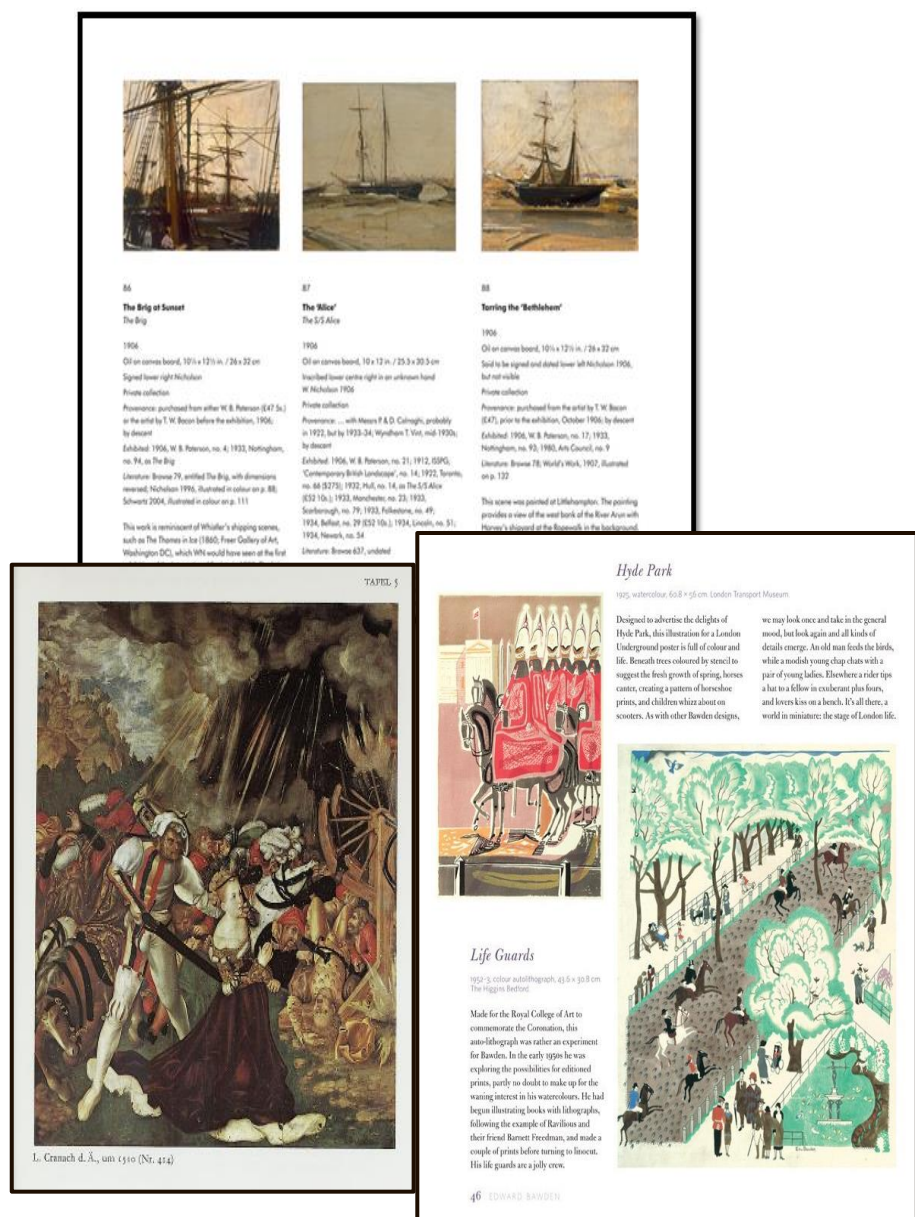
Nitisha Jain

Hasso Plattner Institute, University of Potsdam

## Problem Statement

To enable automated construction of a domain-specific and semantically-rich knowledge graph from cultural heritage datasets.

### Cultural Heritage Datasets



- Heterogeneous data
  - Art Books, Exhibition, Auction Manuals
- Challenges
  - Old archival data
  - Noisy data due to OCR errors
  - Multilingual texts
  - Semantic Ambiguity

### Research Questions

1. How can a domain-specific ontology be learnt automatically from data?
2. How can we extract artwork titles from cultural heritage data through named entity recognition?
3. How can cultural heritage entities be connected by meaningful relations?
4. How can enrichment of an art knowledge graph enable efficient semantic exploration?

### Named Entity Recognition for Artwork Titles



'girl before a mirror'

Artwork titles can be generic and abstract

'head of a woman'  
'Paris' 'a man'  
'untitled'

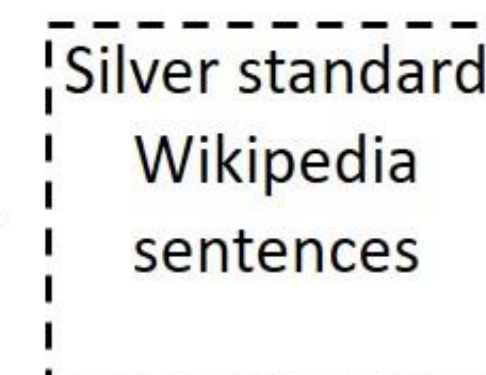
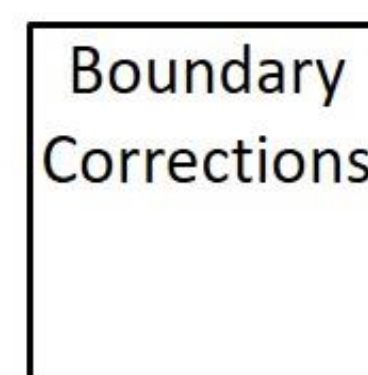
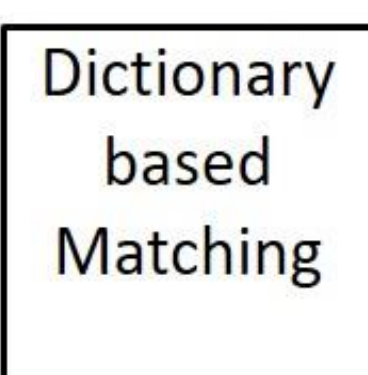
- Existing NER tools do not scale
- Incorrect named entity type tagging
- Incorrect boundary detection



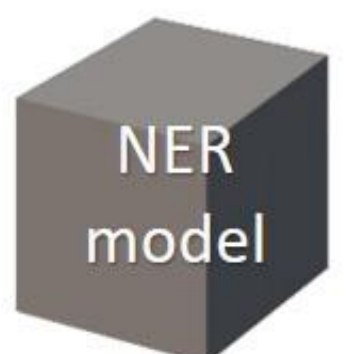
Person or Artwork?



Art Dataset

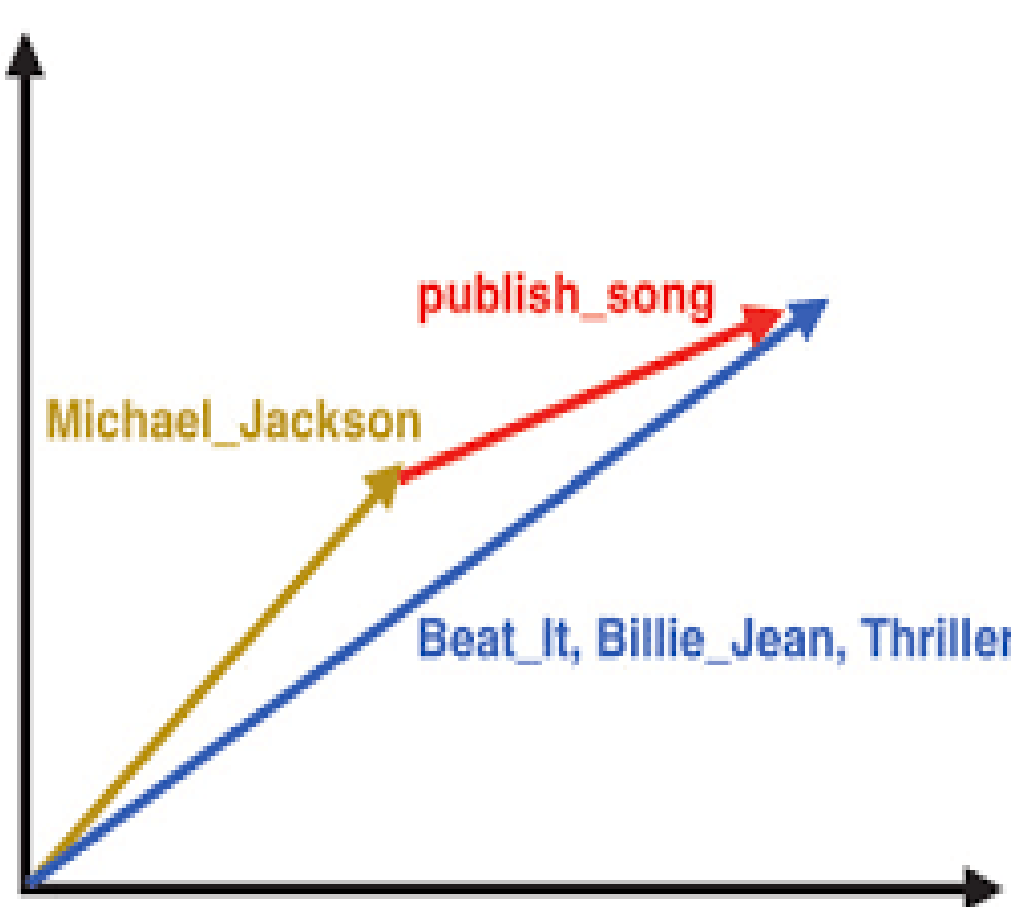


Annotated Dataset



NER Framework

### Automated Ontology Learning via Knowledge Graph Embeddings



- Embed components of KG (entities, relations) into continuous vector spaces
- Easy manipulation of data, preserve KG data structure

- Leverage semantic information present in KGs e.g. <Van Gogh isA Painter>

Goal - Link prediction with KG embeddings for ontological triples

### Evaluation

Quality of a knowledge graph and ontology will be determined by how useful it is for the domain. Feedback from domain experts crucial.

- **Ontology Learning** - Manually create small gold standard test data for evaluation
- **Knowledge Graph construction**
  - Intrinsic Evaluation - Measure completeness and correctness of extracted facts
  - Extrinsic evaluation - Downstream tasks such as search and retrieval of artwork entities