Graph-based Change Detection for Natural Language

Akshita Ramya Kamsali
Elmore Family School of Electrical and Computer Engineering, Purdue University

Robot Vision Lab

Introduction & Motivation

From Raw Text to Insights

Need:

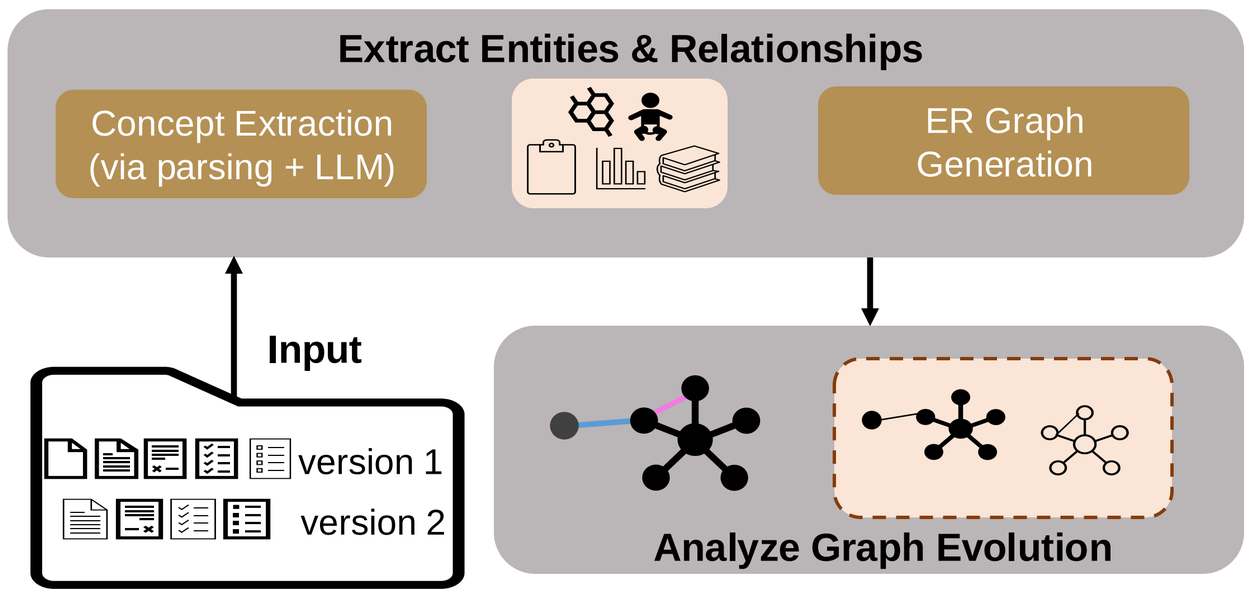
- Automated methods to track evolving, complex text in specialized domains

Solution:

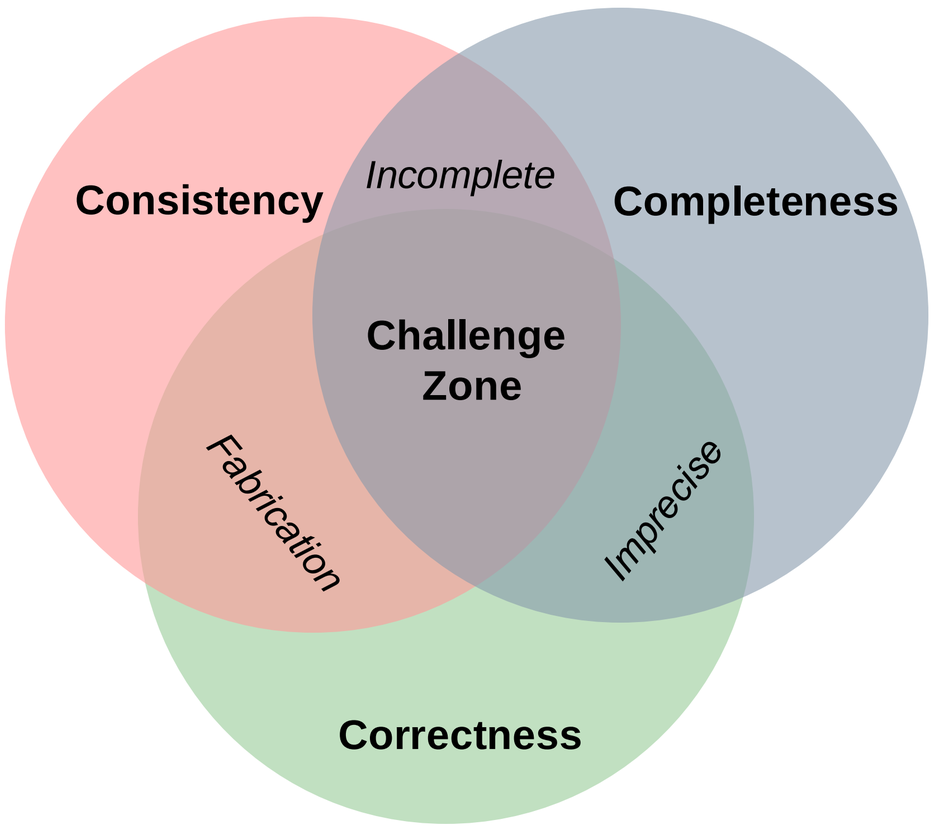
- Entity Relation (ER) graphs structure text, clarifying ambiguities and dependencies

Goal:

- Detect document changes using ER graphs to improve text analysis



Challenges



Balancing the 3Cs for Optimal Knowledge Graph Construction

Consistency:

- Outputs of LLMs lack determinism, as they are sampled from a learned probability space
- Leads to inconsistencies in the generated knowledge graphs

Completeness:

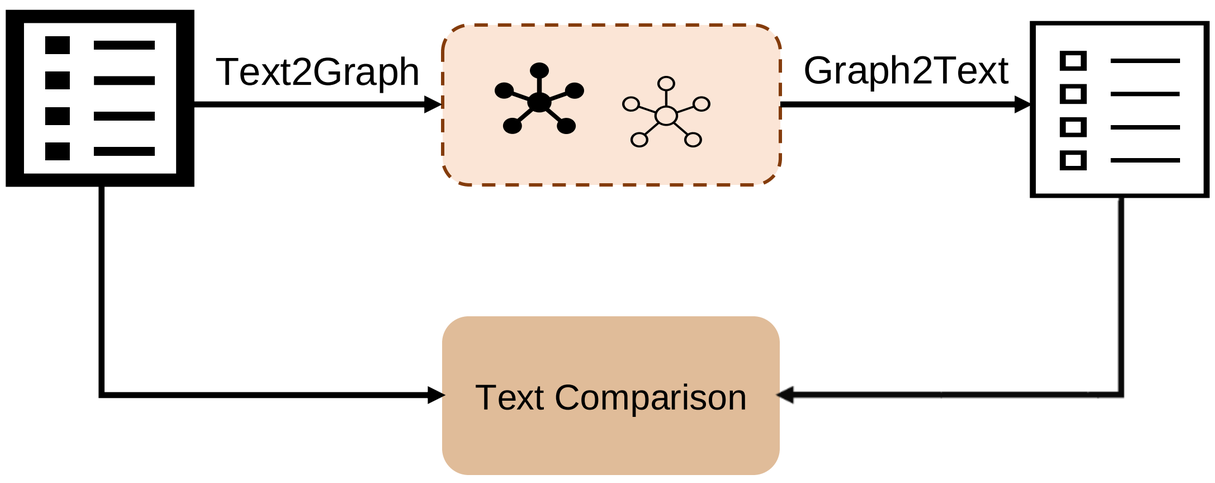
- Parsing documents with LLMs often results in graphs that fail to capture all relevant or necessary information
- Leads to gaps in the representation

Correctness:

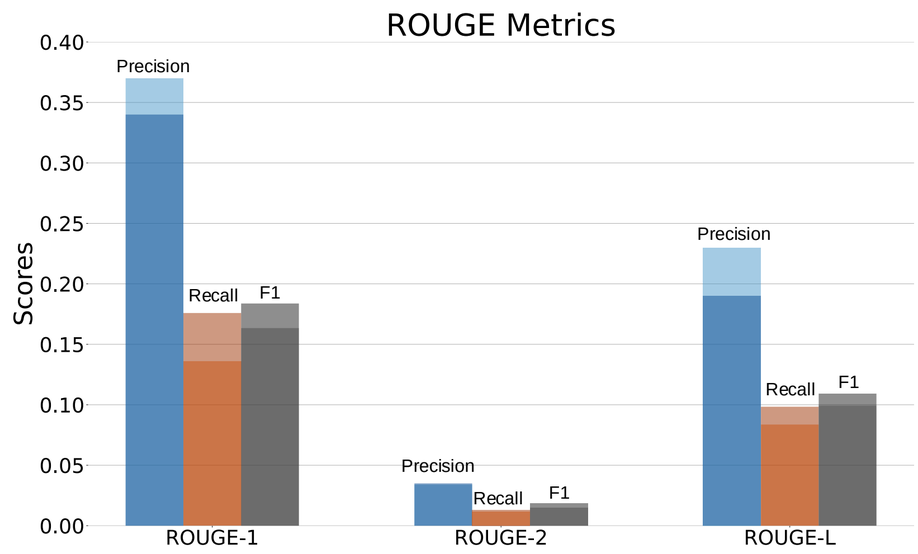
- LLMs are prone to hallucination, sometimes inventing entities or relationships that do not exist in the source document
- Leads to inaccuracies

Methodology

- We employ various entity extractors, including classical CRF-based models and neural-based models, to construct complete and correct ER graphs



Results



- **Precision (Consistency):** Weak ROUGE-2 precision indicates missing relevant entities across repeats
- **Recall & F1-Score (Completeness & Correctness):** Low ROUGE-2 recall and ROUGE-2/ROUGE-L F1-scores highlight missed multi-entity relationships and the need to reduce hallucinations

Future Work

- Fine-tuning models
- Expanding training data
- Improving multi-entity relationship recall (ROUGE-2)
- Exploring hybrid CRF-neural methods
- Enhancing graph construction for complex relationships
- Boosting structural evaluation (ROUGE-L)

References

- Brown, Tom, et al. "Language Models are Few-Shot Learners." *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 1877-1901.
- Ji, Shaoxiong & Pan, Shirui & Cambria, Erik & Marttinen, Pekka & Yu, Philip. (2021). A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. IEEE transactions on neural networks and learning systems. PP. 10.1109/TNNLS.2021.3070843.
- Lample, Guillaume, et al. "Neural Architectures for Named Entity Recognition." *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2016, pp. 260-270.
- Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." *Text Summarization Branches Out*, vol. 8, 2004, pp. 74-81.
- Valle, Kjetil. “Graph-based Representations for Text Classification.” (2011).