

Struct2Unstruct: Creating Tender NER Datasets from Structured Procurement Records using Large Language Models

Asim Abbas

School of Computer Science

University of Birmingham



UNIVERSITY OF
BIRMINGHAM



What is a tender?



Buyer / Tenderer: Issues the Request



An individual or organization, like a government department or private entity, that issues an invitation for a specific project.

REQUEST / NOTICE FLOW



Publishing Tender Notices

Notices are published online, in newspapers, or on noticeboards depending on the nature of work and culture.



Seller / Tenderer: Submits the Bid



An individual or organization that submits a proposal in response to the tenderer's request.

PROPOSAL / BID FLOW



Proposal Contents: Prices, Availability, Deadlines

These details are provided by the seller in response to the invitation issued by the tenderer.

Problem Statement

- ❑ Tender **data scarcity** due to domain sensitivity
- ❑ Data **hallucination** using LLMs
- ❑ Data **annotation** is tedious for NER task

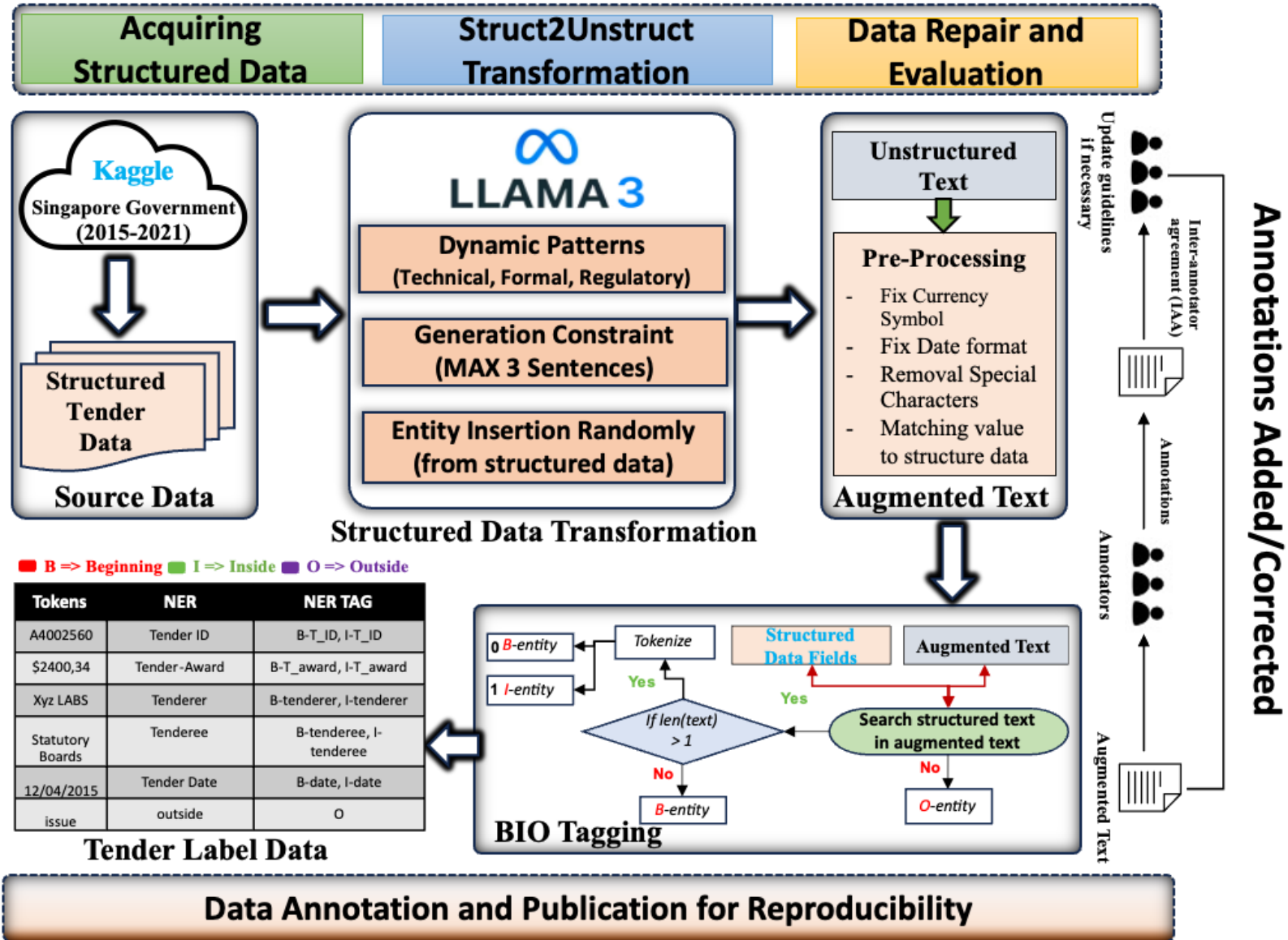
Research Question

To what extent can **LLM-generated synthetic** tender data mitigate data **scarcity** and domain **sensitivity** challenges in NER tasks?

Research Gap

Despite advances in LLM-based synthetic data generation, there is a **lack of rigorous** evaluation of data to ensure the reliability of such data for **downstream tasks** such as NER in sensitive domains like tender and procurement.

Proposed Approach



Thank you.



UNIVERSITY OF
BIRMINGHAM

