



THE UNIVERSITY OF BRITISH COLUMBIA

School of Information  
Faculty of Arts

# Index-Aware Knowledge Grounding of Retrieval-Augmented Generation in Conversational Search for Archival Diplomatics

Qihong Zhou, Binming Li, Victoria Lemieux

DECEMBER 9, 2025





THE UNIVERSITY OF BRITISH COLUMBIA

School of Information  
Faculty of Arts

**Acknowledgements:** Natural Science and Engineering Council (NSERC) and Social Sciences and Research Council of Canada (SSHR) InterPARES Trust AI Partnership





# PROBLEM STATEMENT



# Problem Statement

- Standard RAG processing often does not yield accurate results even with high computational demands, requires large amounts of data, has considerable energy consumption
- Results in inaccuracy of results and high operational costs and environmental concerns





# Research Question

---

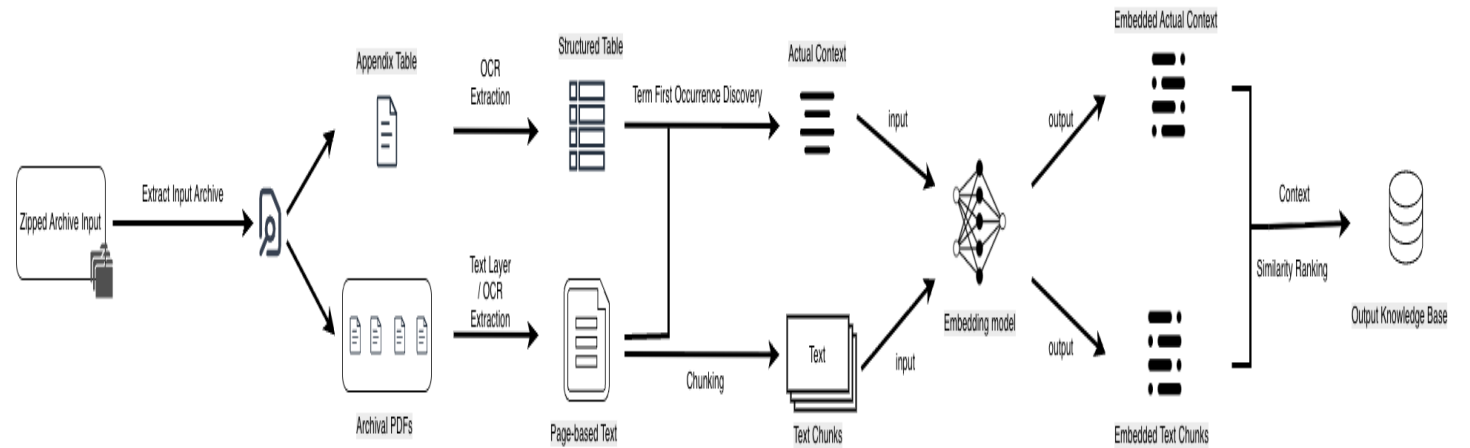
- Can we adjust the conversational search pipeline to improve accuracy and reduce high data and computational costs?
- We decided to focus on the chunking strategy used in RAG

# METHODOLOGY



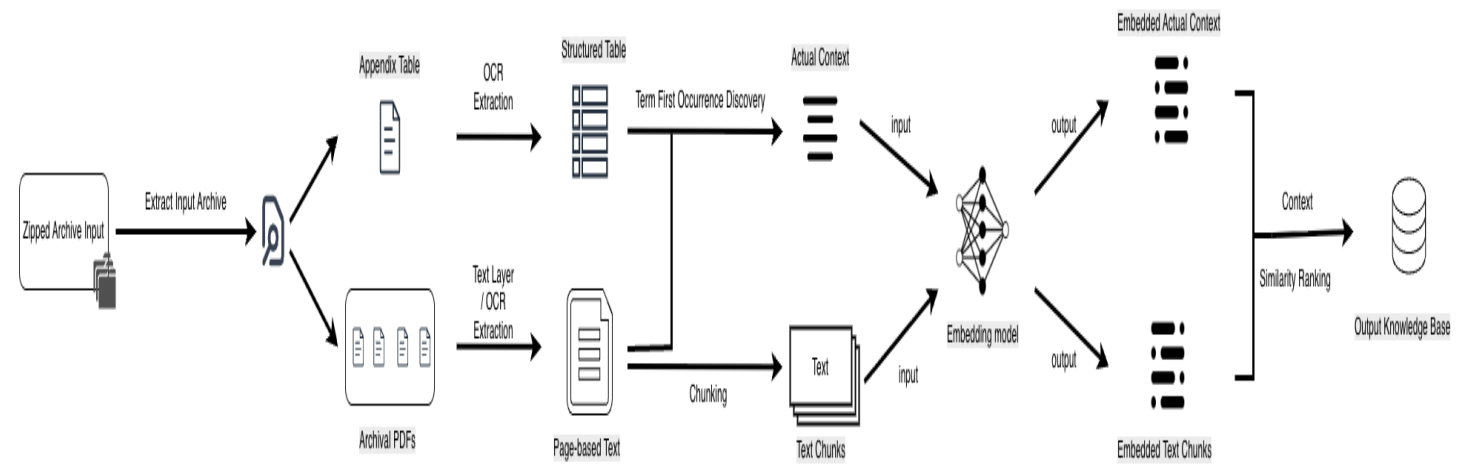


# Methodology (Chunking Strategy)



- Appendix Table OCR
- Term's first occurrence discovery
- Re-chunking
- Top related text chunks retrieval

# Methodology (Approach to Evaluation)



- Evaluation

- 2 platforms

- Clio-X (Clix.org)



- Hugging Face



- 4 Test Prompts (2 reported in paper)

## Limitations

- Evaluation/testing not isomorphic
- Did not control for all steps in the conversational search pipeline
- Ran test questions only once



# RESULTS



# Results

- 3 Personas:
  - Clio-X Standard Chunking: 'young turk'
  - Hugging Face: 'windbag'
  - Clio-X Index-Aware Chunking: 'CAS student'
- Results not necessarily improved by throwing more GPUs at a conversational search task
- Refinements to the conversational search pipeline, **such as our adjustments to the chunking strategy**, can achieve incremental improvements in the accuracy and completeness of results without requiring time-consuming and labour-intensive model training.
- Future work: adjust other components of the conversational search pipeline





THE UNIVERSITY OF BRITISH COLUMBIA

School of Information

Faculty of Arts

THE UNIVERSITY OF BRITISH