

# Evaluating Automated and Hybrid Neural Disambiguation for African Historical Named Entities



## Introduction

Names can be ambiguous as many people may share the same name or the same person may be referred to by different names. Ambiguous names are a particular problem within a historical South African context as different narrators of historical events often use the different names for the same person.

The 500 Year Archive (FHYA) is a collection of artifacts from the last 500 years of South African history. This project seeks to use language models (LMs) to construct a Named Entity Disambiguation (NED) system to disambiguate names appearing within historical documents within the FHYA as well as the metadata associated with each of the artifacts.

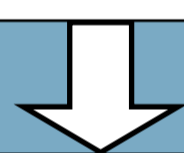
## Research Questions

- How accurate are language models (LMs) for NED applied to the text from the 500 Hundred Year Archive?
- What is the effect of additional annotated data and hand-crafted features on the accuracy of the NED model?

### NED System Outline

#### Input Text

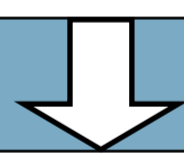
Shaka was a Zulu chief.



#### Mention Detection

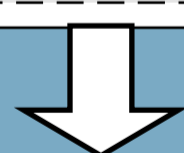
Mention: Shaka  
Context: Shaka was a Zulu chief.  
Position: (0,1)

**Shaka** was a Zulu chief.



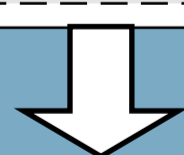
#### Candidate Generation

Candidates: [(#1, Shaka King), (#2, Shaka), (#3, Shaka Smart)]



#### Entity Selection

Candidates: [(#2, **Shaka**), (#3, Shaka Smart), (#1, Shaka King)]



#### Labeled Text

**Shaka**[#2] was a Zulu chief.

### NED Systems

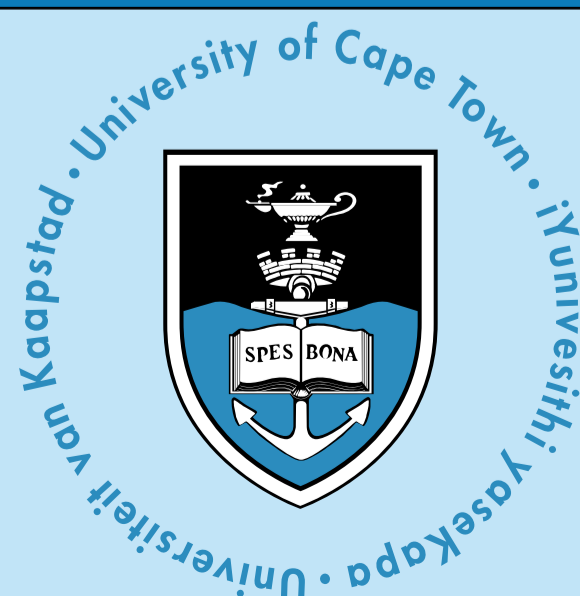
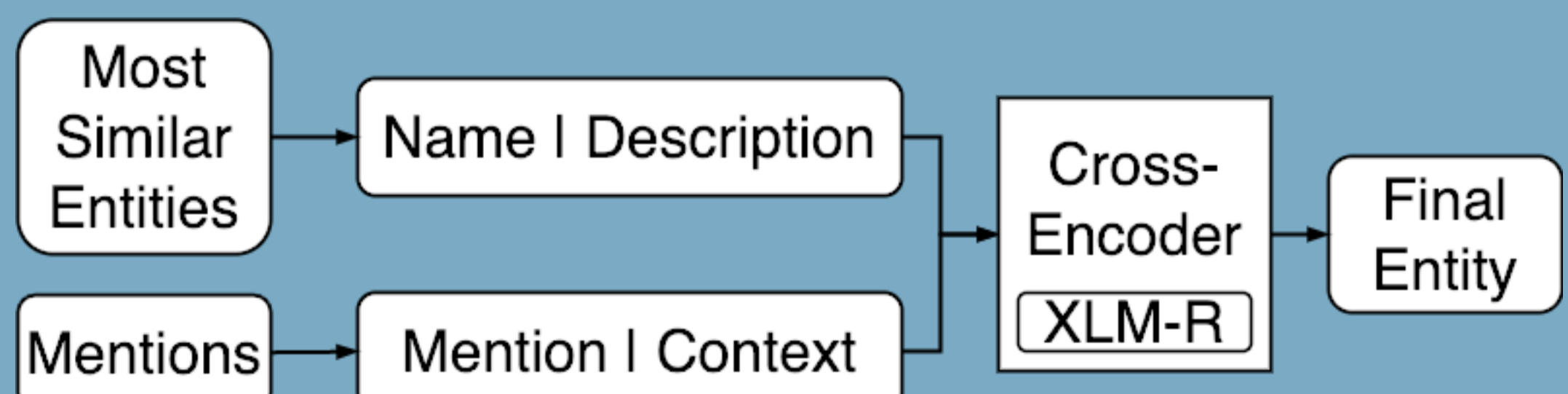
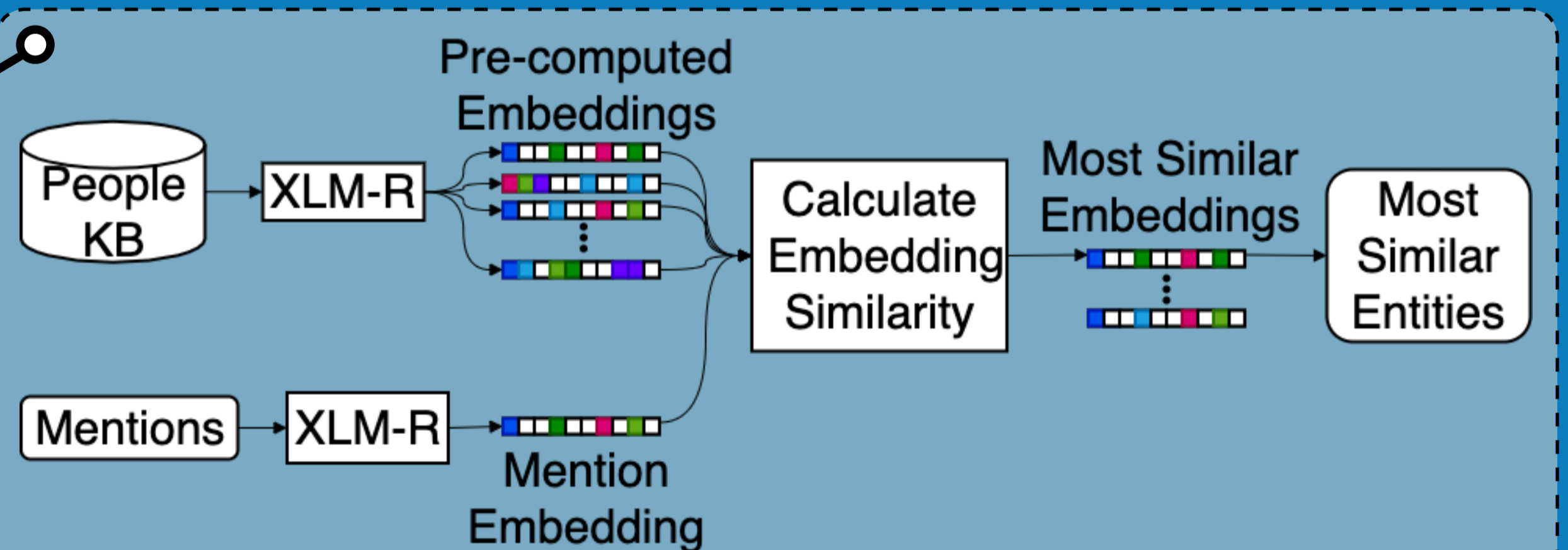
The research questions will be answered based on the performances of two NED systems:

#### Automatic System

Makes use of the language model XLM-R to disambiguate names taking a supervised learning approach.

#### Hybrid System

Extends the Automatic system by incorporating hand-crafted heuristics.



**Jarryd Dunn**  
Masters student  
dnnjar001@myuct.ac.za

**Prof. Hussein Suleman**  
Supervisor

School of Information Technology  
Computer Science Building  
University of Cape Town  
Private Bag X3  
Rondebosch, 7701

