

Identification of the origin of place names via Geographic Information Retrieval

Horde Vo, A.⁽¹⁾; Duckham, M.⁽¹⁾; He, E.⁽¹⁾; Ranamuka N.G.⁽¹⁾; Benli, R.⁽²⁾
(1) RMIT University (2) Geographic Names Victoria, Victoria State Government

Introduction

Who is commemorated behind a place name in Australia?

Approach

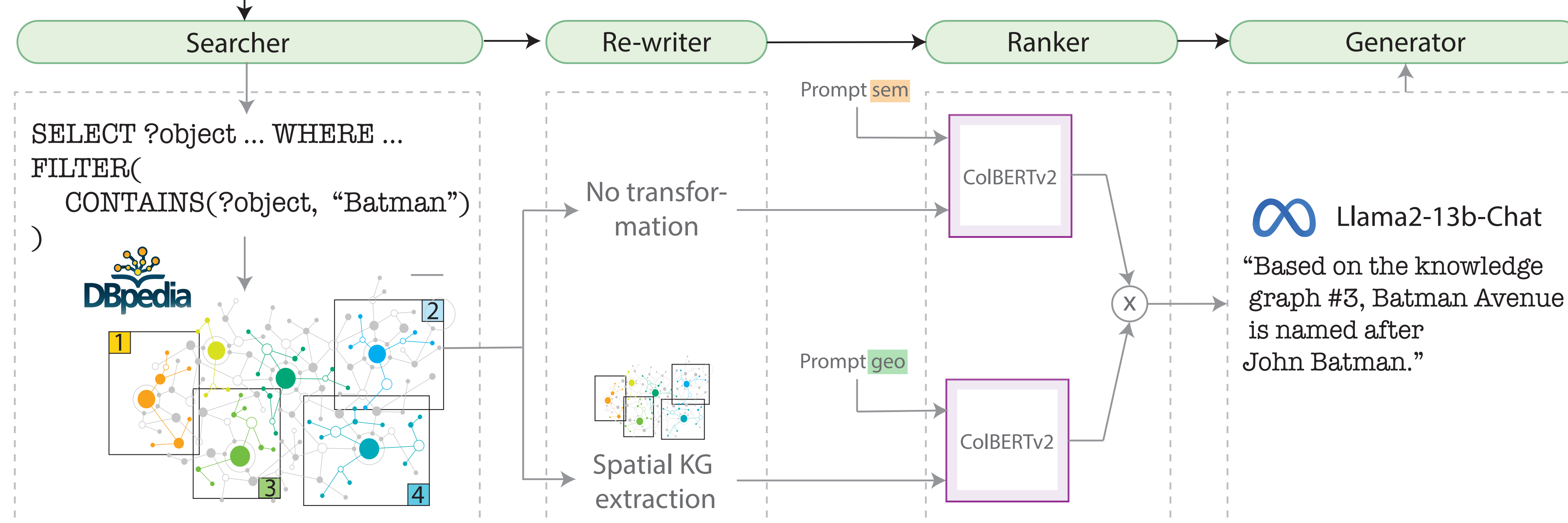
We search for candidate-answers that fulfill two cumulative conditions: a semantic and a spatial matching.

Gazetteer:

Place: Batman Avenue
City: Melbourne
State: Victoria
Country: Australia
Named after: ?

Who is **Batman Avenue** most likely named after, in **Melbourne, Victoria, Australia**? → Prompt_{sem}

What event **occured** in **Melbourne, Victoria, Australia**? → Prompt_{geo}



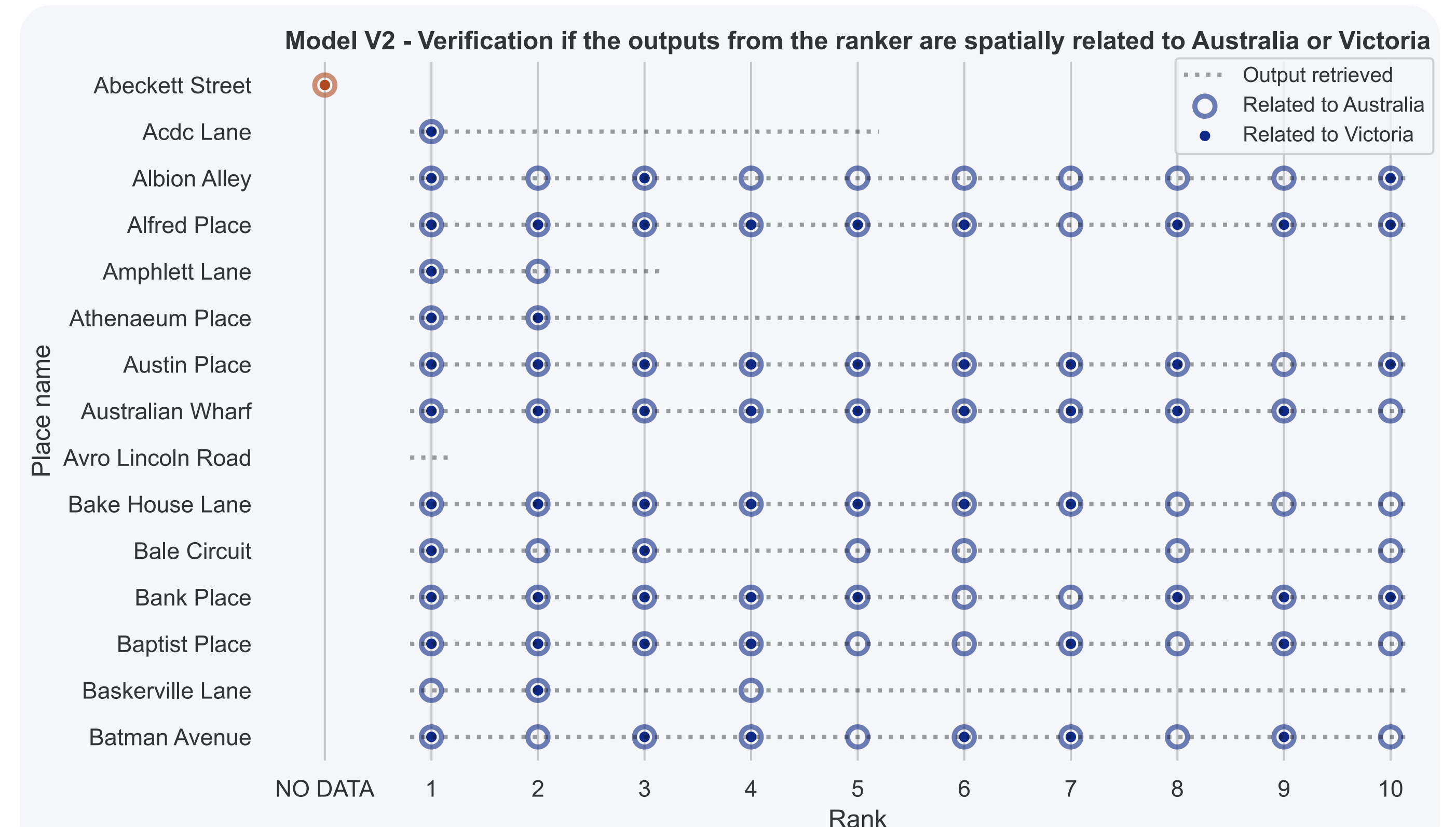
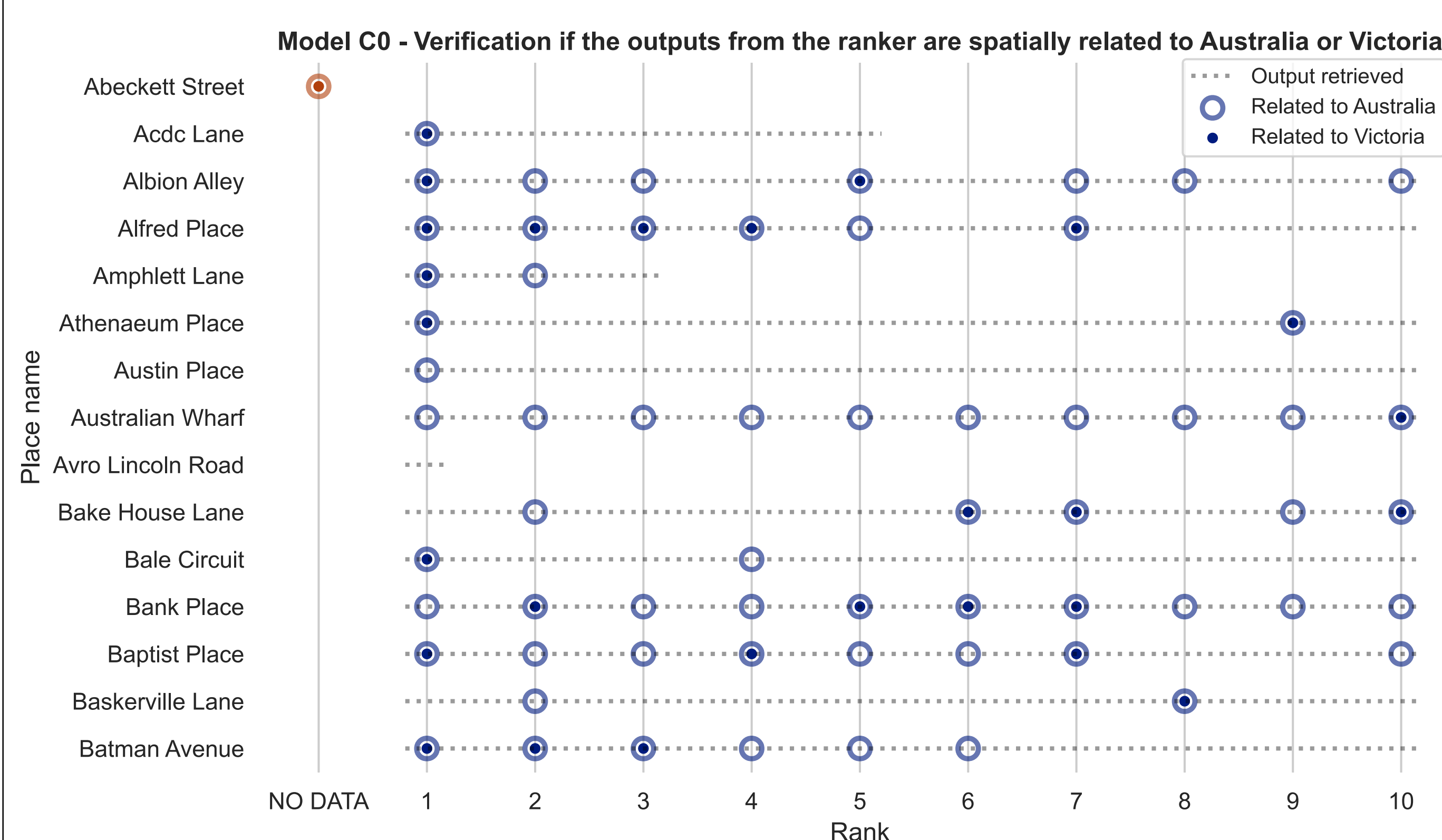
- Even if place names are recorded in gazetteers, their origins are still **not exhaustively tracked**.
- By injecting large knowledge graphs (KGs), we propose a geographic information retrieval (GIR) model, based on a **retrieval-augmented generation (RAG)** architecture.
- With the combination of **semantic and spatial correlations**, we aim to **find exact answers** from existing documents. If missing, we **recommend** the most plausible candidates.

Results

How spatially efficient is the ranker?

Naive RAG (baseline - CoBERTv2)

Spatially optimized RAG (ours)



Conclusions

- Retrieving an origin for place names is **enhanced by the reading capabilities of large language models** (+88% for the precision after the generator in comparison with GPT-4o-mini). However, standalone large language models **insufficiently treat space as a discriminative factor**.
- The normalized discounted cumulative gain (nDCG) measures if relevant candidates are often recommended with higher ranks. Through a **multi-view representation in the ranker**, the **nDCG is increased by +16% in recommending candidates linked to Australia**, and by +30% for candidates related to the state-level of Victoria, in comparison with the state-of-the-art baseline.

Acknowledgements

- This project was undertaken with the assistance of computing resources from RACE (RMIT AWS Cloud Supercomputing).
- We thank Nenad Radosevic for his insightful help at the Geographic Knowledge Lab (RMIT University).