

More than words. More meaningful and relatable search results are more relevant and accessible to humans.

Improving Search Accessibility at the South African SDG Hub using Semantic Matching and Topic Modelling.

INTRO

- The South African SDG^{1,2} Hub's search engine returns irrelevant search results.
- Goal: improve search engine by returning relevant results that are more accessible:
 - semantic matched, and
 - visualised.

METHODS

Search Engine:

- Semantic vector representations were produced using transformers. Two were carefully selected based on size and semantic matching performance.
- Cosine similarity between the query vector and the SDG metatext vector was used as a distance to rank and match SDG papers.
- A survey was conducted to see which search engine was better.

Topic Modeling:

- Latent Dirichlet Allocation (LDA) applied to the text from the returned search results.
- Coherence score was used to find the optimal number of topics
- Returned topic word clouds inform of themes (or topics) contained in the search results.
- The topic number and topic colour overlaid on the each returned result.

Shaun Johnson, Jesse Parvess

RESULTS

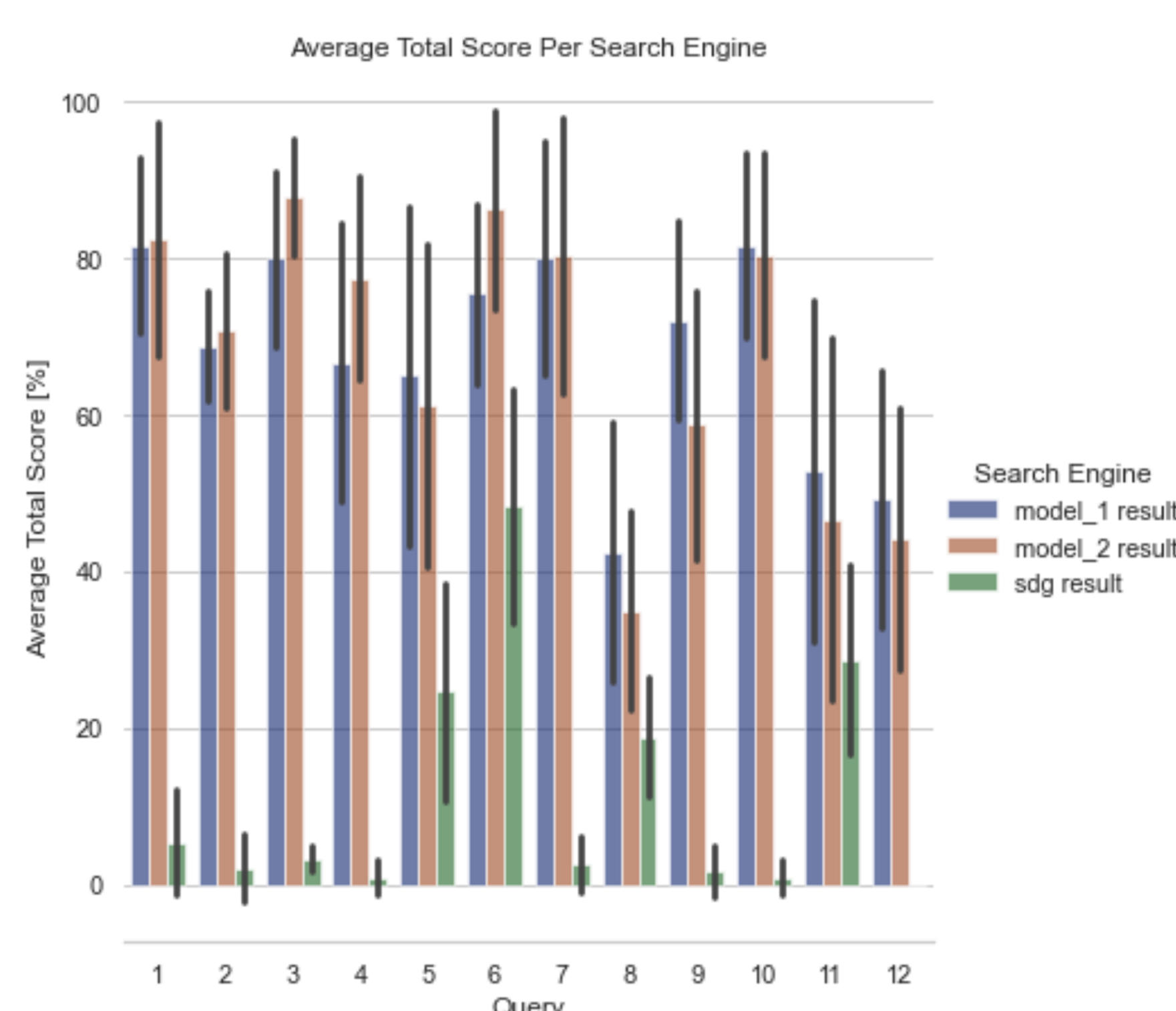


Figure 1. Survey results comparing search engines

Searches	Search Engine
0	--- Query-effectiveness of judicial system ---
1	[Topic5] legal efficiency and consistency
2	[Topic4] the supreme court as a constitutional watchdog
3	[Topic6] reviewing reasonableness an appropriate standard for evaluating state action and inaction
4	[Topic2] research recap can information improve the functioning of courts
5	[Topic1] judicial administration

Figure 2. Survey results with LDA topic colour

DISCUSSION

- Figure 1 demonstrates that users find our implemented search engines return more relevant results.
- Figure 2 demonstrates how LDA can be used to make these results more accessible by assigning each result with a topic and colour.

ADDENDUM

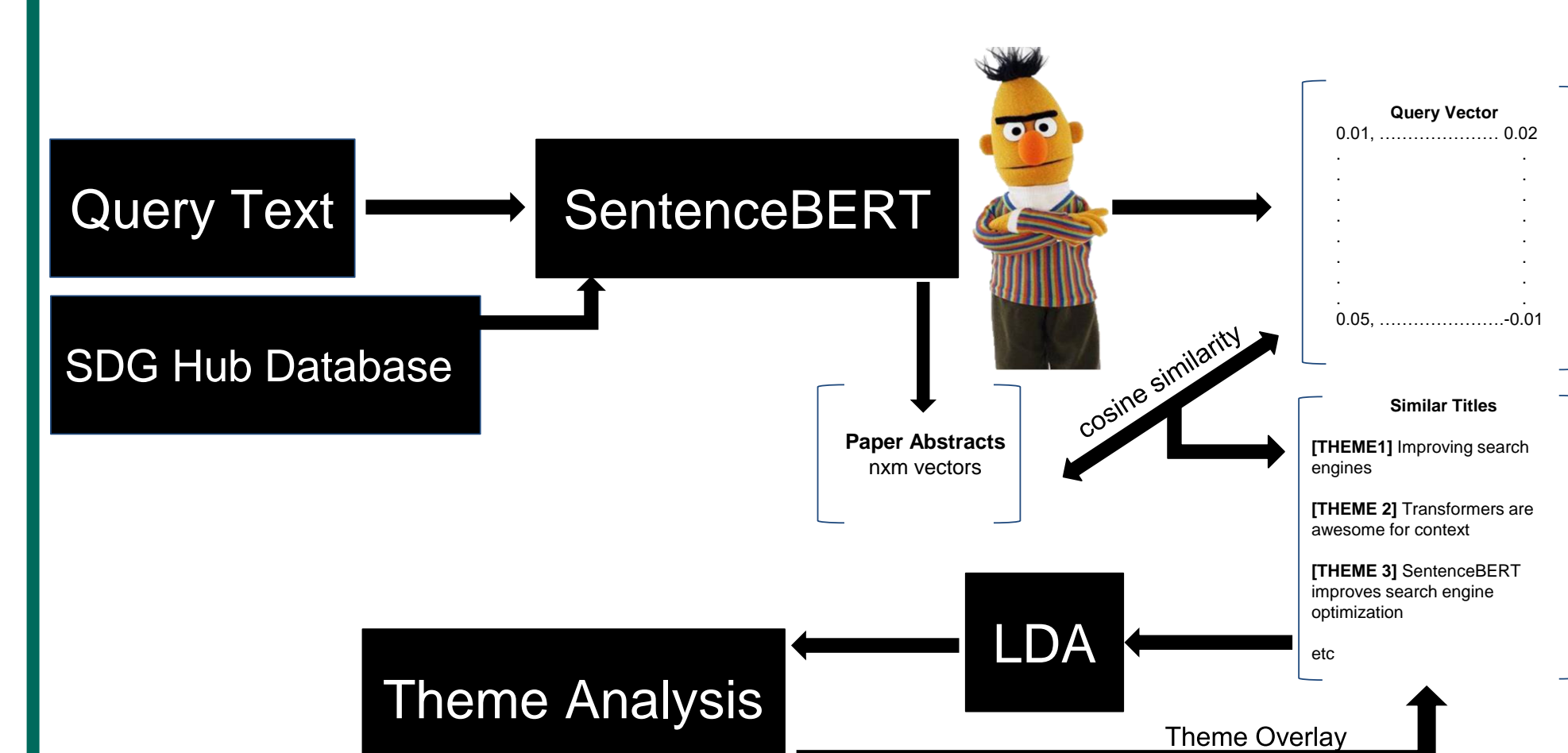


Fig A. Solution Overview

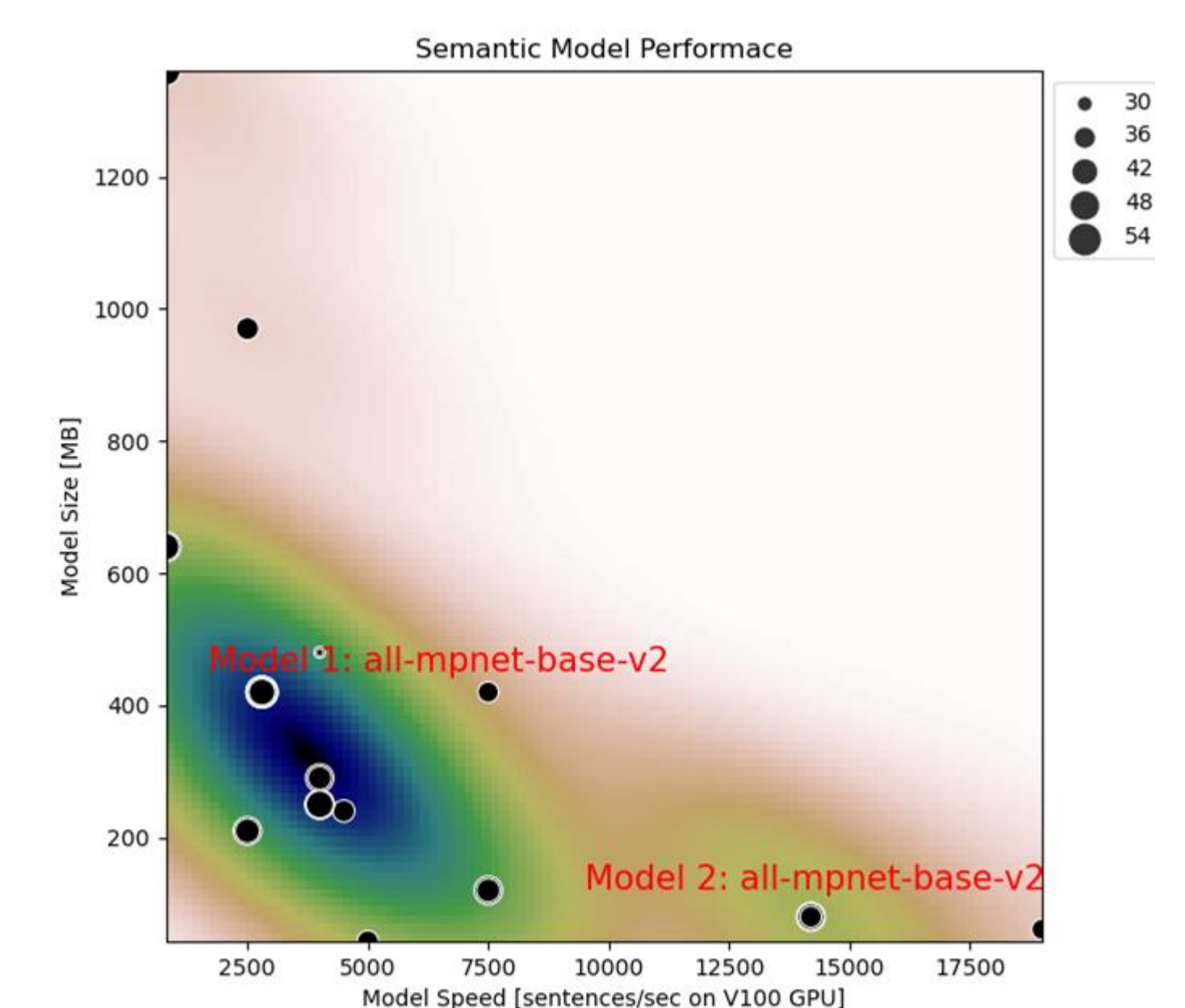


Fig B. Sentence Transformer Analysis

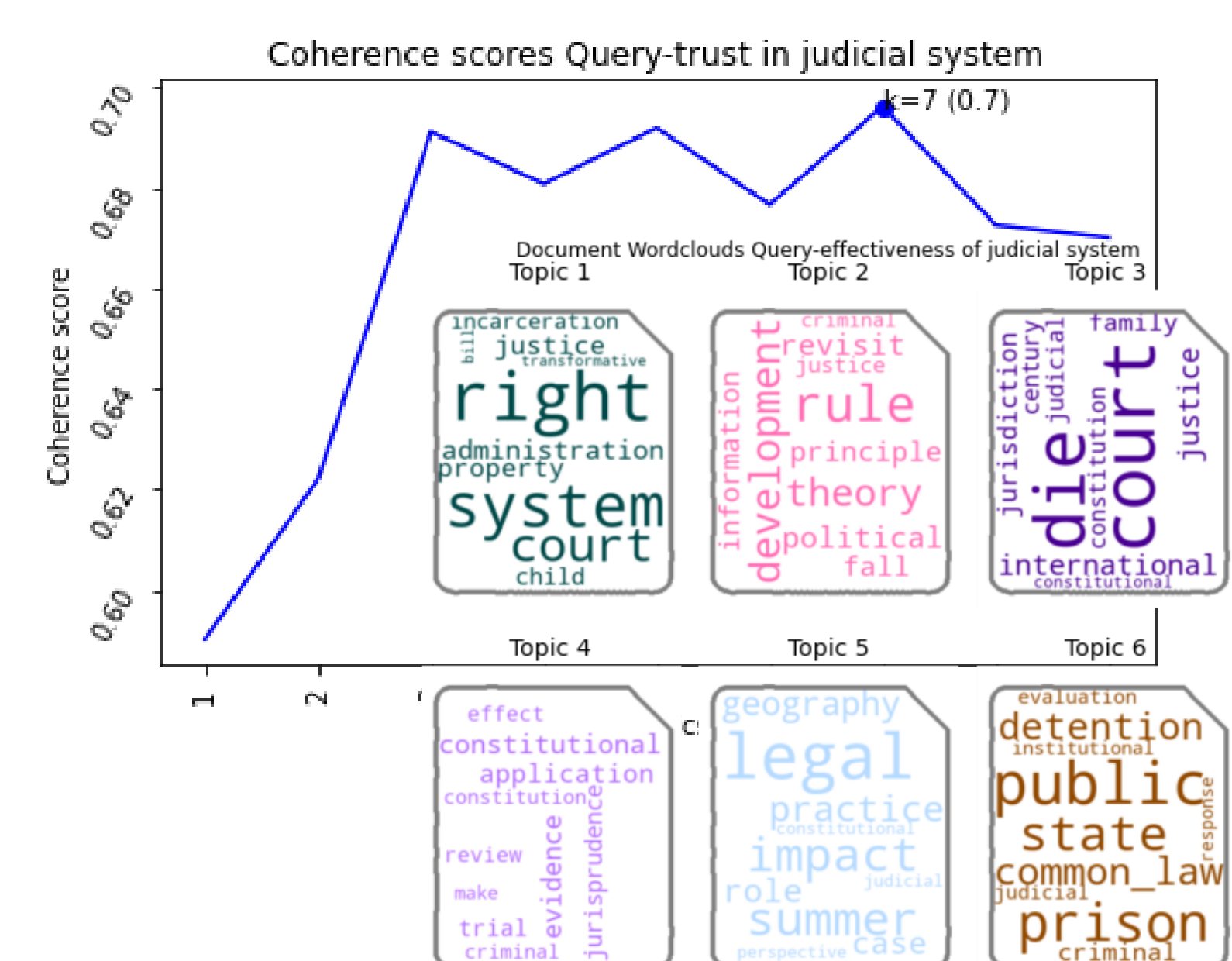


Fig C. Coherence Scores and Word Cloud Results

- Sustainable Development Goals (SDGs).
- <https://sasdhub.up.ac.za>

Department of Computer Science

Faculty of Engineering,
Built Environment and
Information Technology

Fakulteit Ingenieurswese, Bou-omgewing en
Inligtingtegnologie / Lefapha la Boetšenere,
Tikologo ya Kago le Theknolotši ya Tshedimošo

Capstone Project - MIT 808

Course Coordinators:
Dr. Vukosi Marivate (vukosi.marivate@cs.up.ac.za)
Abiodun Modupe (abiodun.modupe@cs.up.ac.za)

Scan me

