CHETAH: Fast and Intelligent Search Engine for NGOs and UN

Shivika K Bisen, MS Data Science

BACKGROUND & PROBLEM STATEMENT

- Horizontal search engine are biased towards sponsored links
- NGO/UN sector reports are under-represented
- Reports are long and takes long time to read
- Annual reports does not helps with program designing goals

OBJECTIVES

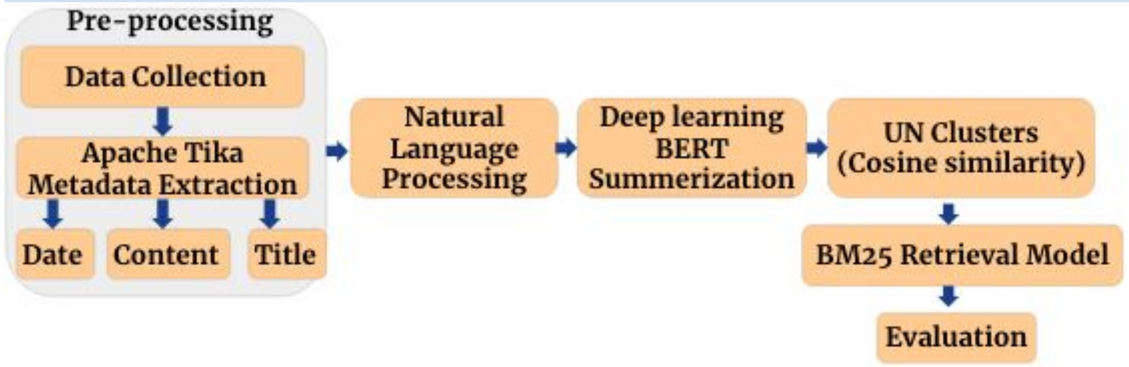
- Develop a fast vertical search engine for Program designing.
- Create a dataset of UN & NGO reports (annual + program reports).
- AI (BERT) based summarization of reports

RESEARCH OVERVIEW

Chetah search engine targets Foundations, NGO program managers and UN agencies policy makers who design programs and grant funds. Based on the query, it returns a list of link of relevant evidence based reports. Creates summary of reports with respect to UN cluster. This helps in helps in engagement with authorities and industry experts.

Data includes 212 reports from 3 organizations: UNICEF, IFRC, IWA. To have a balanced dataset all 12 UN Clusters reports are taken.

METHOD



Vector Space Model
Calculates angle between Query and document vectors

↔

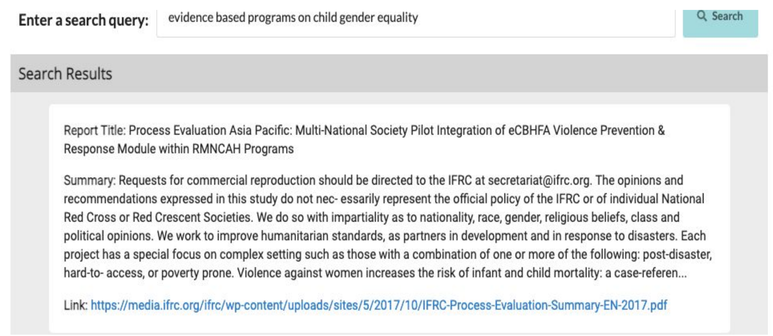
Probabilistic IR Model
Calculates the probablity of relevance

BERT DEEP LEARNING SUMMARIZATION

Bidirectional Encoder Representations from Transformers (**BERT**) is a Transformer-based deep learning technique. It understands the context of the text. This improves the retrieval givesquick insight on the report.

RESULT

CHETAH

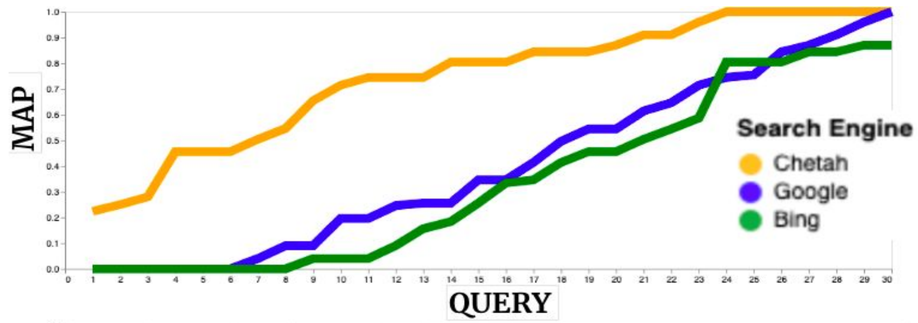


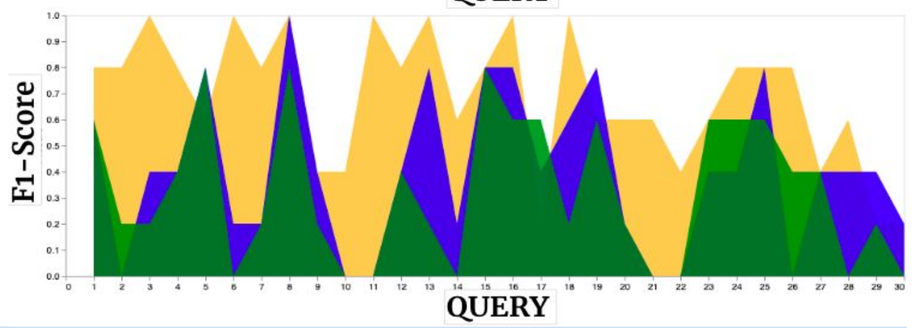
Chetah is implemented on Django architecture

MODEL → VIEW → CONTROLLER

INSIGHTS

- Chetah performs better compared to Google and Bing, for evidence based UN/NGO queries.
- Links reports to UN cluster.
- MAP plot (top):
- F1 score plot (bottom) : Chetah as 0.78 F1 score.
- Time: 0.46 seconds





NEXT STEP

- Include more reports in the dataset and increase search engine audience. It gets betters
- Optimize search algorithms to make search engines based on sections
- Trend analysis of topics on the reports

REFERENCE

Acknowledgement: HoJoon Kim, Mariah Jacob

ChengXiang Zhai et al. 2016. Text Data Management and Analysis.