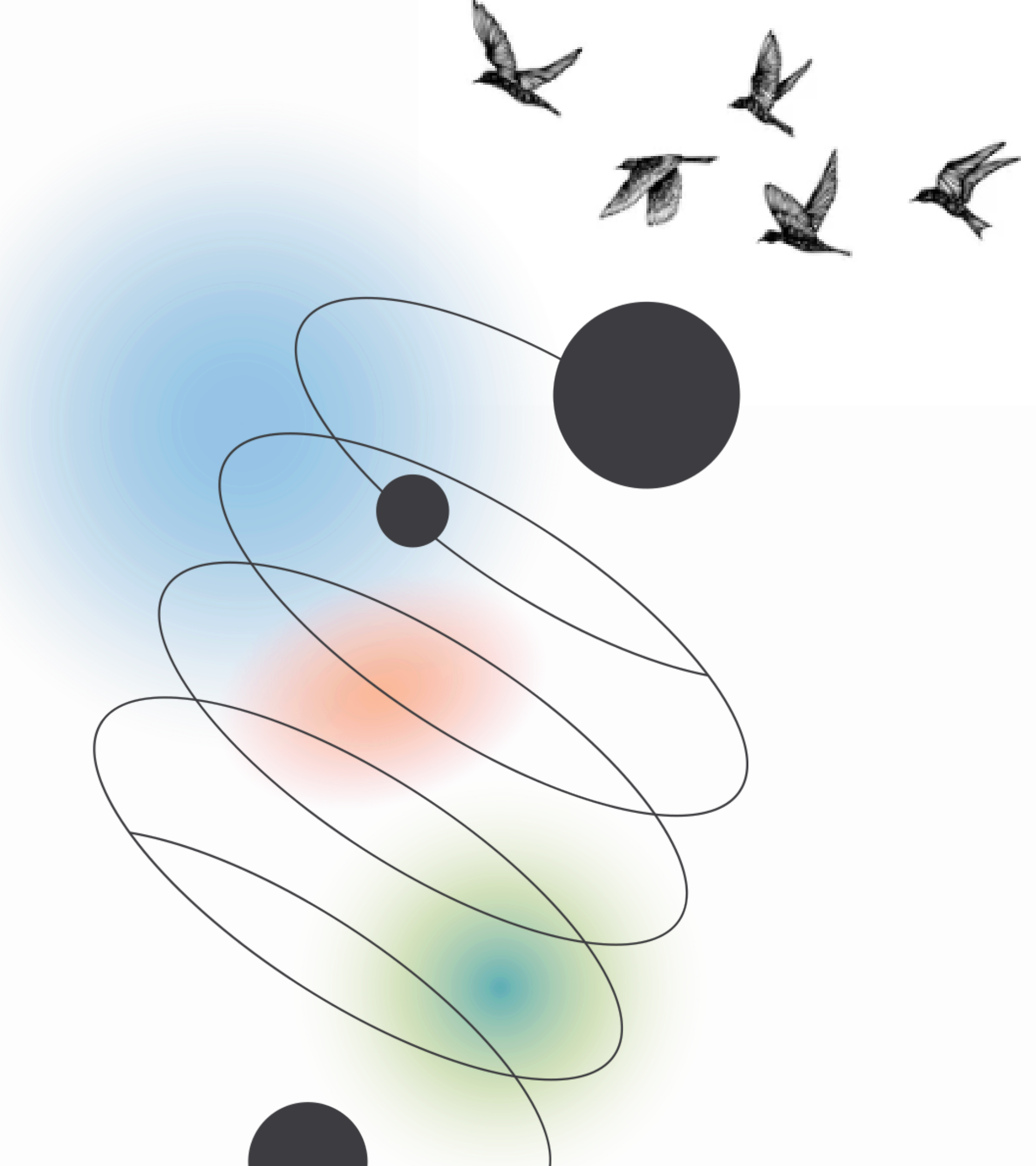# Information Retrieval Project

Steps Followed

1 GCP

- Created Inverted Indexes (Title, Body, Anchor), Document-Title mappings (GCP Cluster)
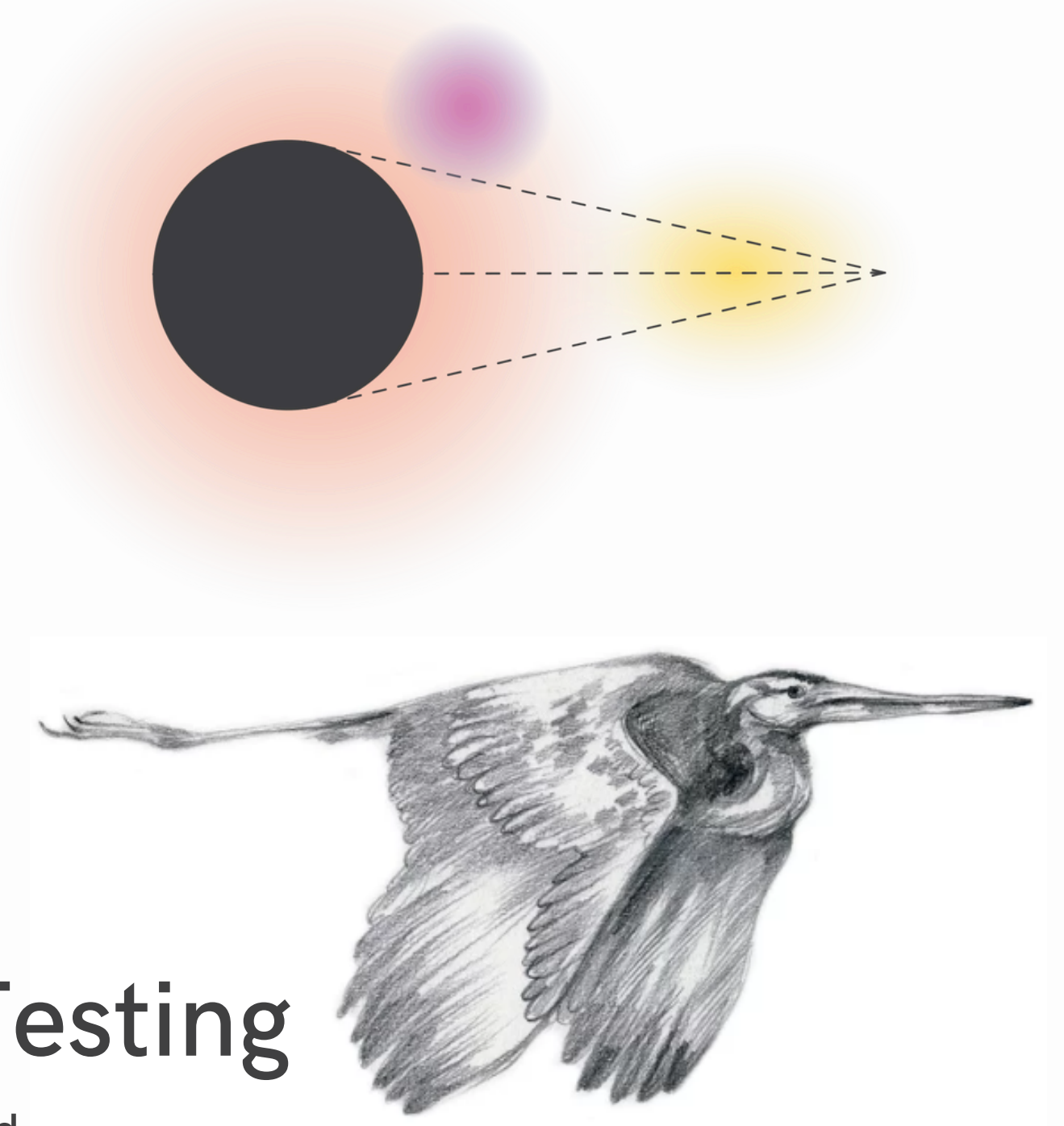
2 Colab

- Created Pagerank, Pageview, Document - Document Length (Colab)
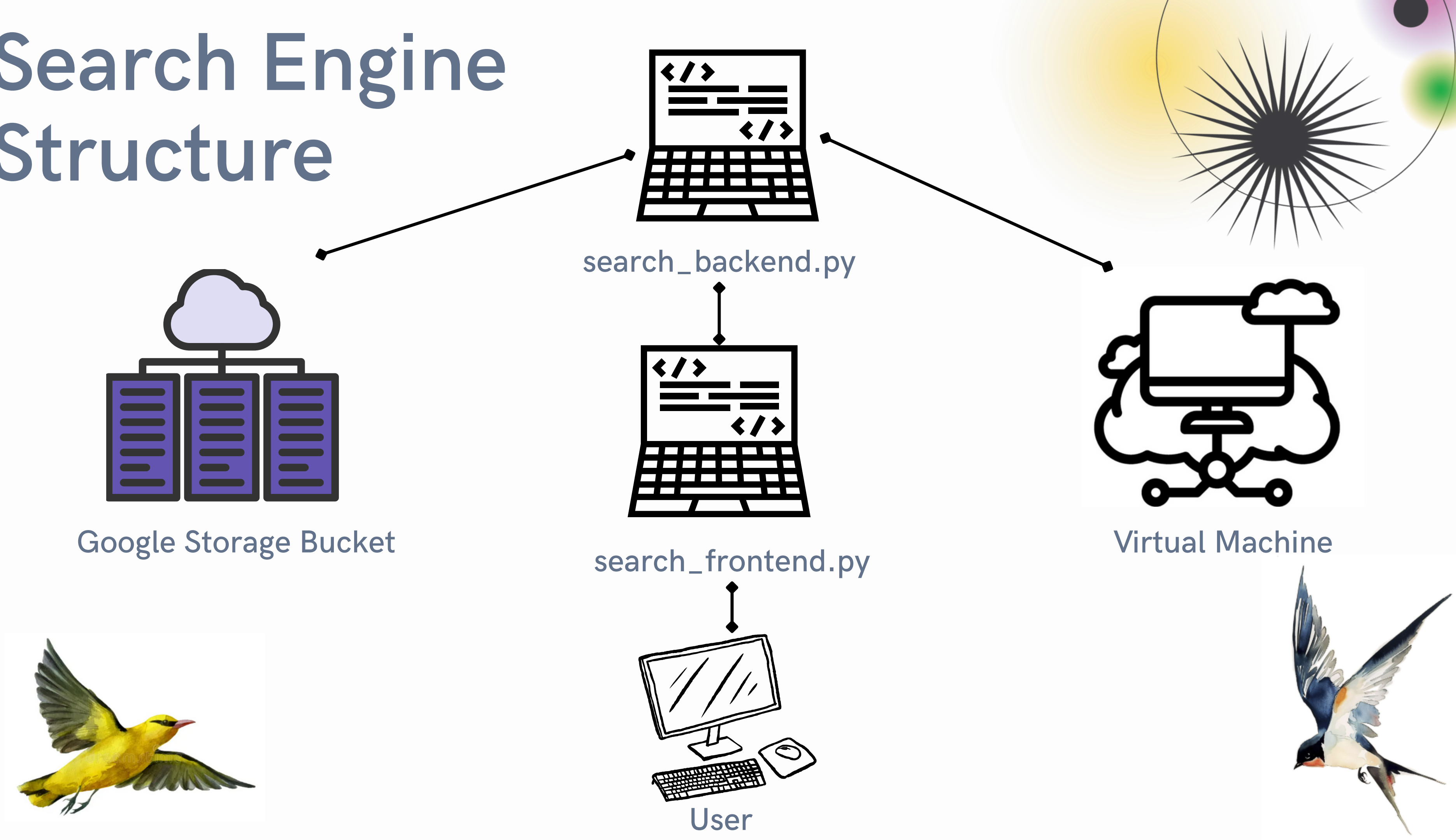- Ran initial tests on engine

3 VM Instance

- Created VM instance (e2-standard-2, 2 vCPUs, 8 GB memory)
- Downloaded retrieval data locally

4 Further Testing

- Ran grid search to find optimal weight components
- Ran advanced tests

Search Engine Structure

search_backend.py

Google Storage Bucket

search_frontend.py

Virtual Machine

User

Final Results

Cos Similarity

```
=================================================
Total Queries:          30
Avg Latency:            766.46 ms
-------------------------------------------------

MAP@5 (Avg P@5):    0.5333
MAP@10 (Avg P@10): 0.3700
Avg F1@30:              0.1546
```

BM-25

```
=================================================
Total Queries:          30
Avg Latency:            745.76 ms
-------------------------------------------------

MAP@5 (Avg P@5):    0.7000
MAP@10 (Avg P@10): 0.6133
Avg F1@30:              0.3371
```