2022

Expeditionary Domain Awareness - Intelligence Support to NECC & NECC Support to Intelligence Analysis (NECC focus)

Das, Arijit

Monterey, California: Naval Postgraduate School

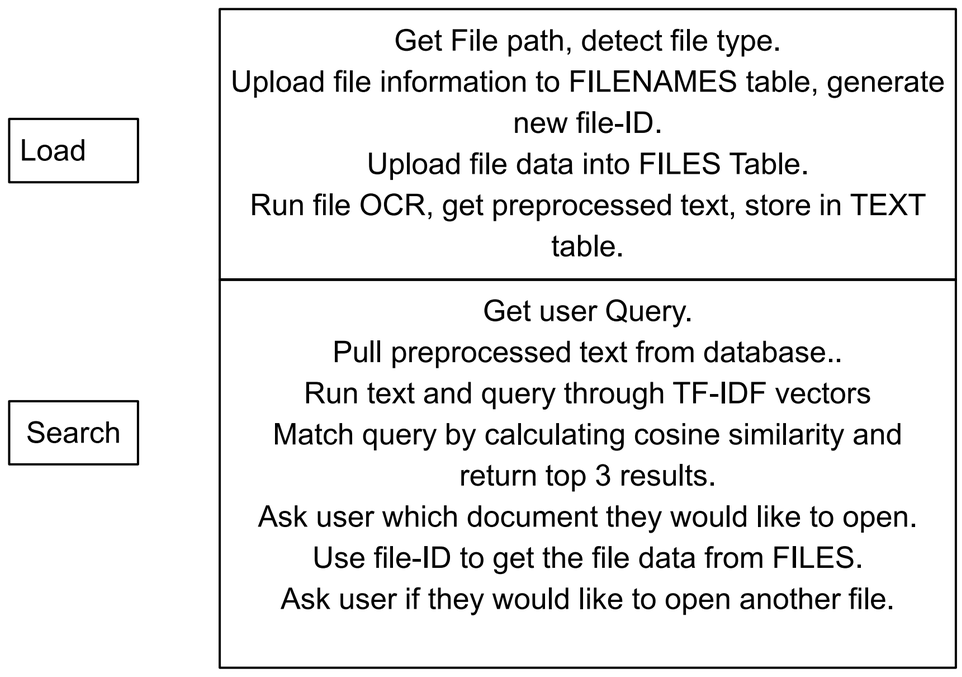
Expeditionary Domain Awareness – Intelligence Support to NECC & NECC Support to Intelligence Analysis

Naval Postgraduate School

Background

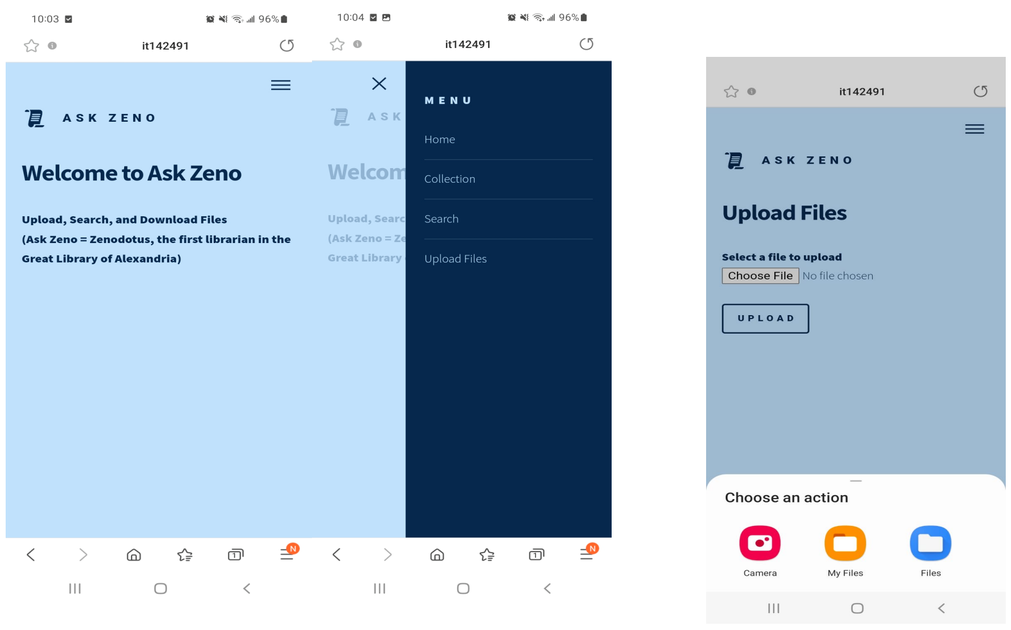
- FY21 effort identified that datasets were summarized into reports or documents.
- Documents accumulate over time and have to be stored and searched over time.
- There are various formats namely PDF, Word, PowerPoint, images and plaintext.
- A Document store architecture has to consider front-end, middleware and back-end.
- NPS team studied the technology challenges taking into consideration scaling, costing and DoD licensing.

Data Flow



Top flow shows the Document load process and the bottom shows the Document search.

User Interface



The picture on the left is on the laptop while that on the right is the phone, for the same application.

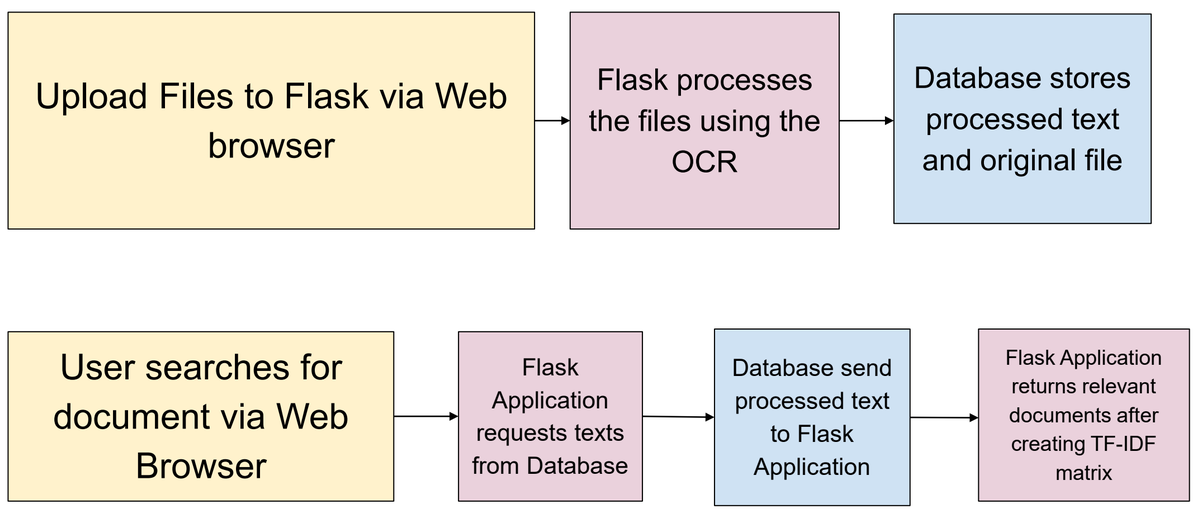
Technology

- Backend Store: Oracle Database.
- Middleware: FLASK Python Framework.
- OCR: OPENCV Libraries.
- OCR Training Model: Tesseract.
- Ranking Documents: TD-IDF vector Algorithms.
  - Cosine Similarity.
- Jython: Running Python code via Java on phone.
- Browser: JavaScript/HTML5.

Findings / Challenges

- Building a sandbox to evaluate architecture.
- Choice of language critical (Python vs. Java).
- Porting codebase from laptop to phone.
- Extracting keywords from binary files.
- Algorithms recalculate on each document load.
- Topic Sponsor made aware of the system needed.

Architecture



3-Tier Architecture.

Future Work

- Data scaling using HDFS.
- Building a Terabyte size repository.
- Involving a community for user interface testing.
- Better understanding of DoD licensing for software.
- Evaluating in-house expertise versus contracting.
- Research prior work done in DoD/DoN IT.

Researchers: Mr. Arijit Das, Computer Science; Mr. Walter Kendall, Information Science, Mr. Peter Ateshian, Computer Science; Dr. Neil Rowe, Computer Science; Ms. Aroshi Ghosh, Summer Intern, MIT undergraduate student
Topic Sponsor: N2/N6 – Information Warfare

NRP Project ID: NPS-22-N207-A

This research is supported by funding from the Naval Postgraduate School, Naval Research Program (PE 0605853N/2098). Approved for public release; distribution is unlimited.