



Calhoun: The NPS Institutional Archive
DSpace Repository

Faculty and Researchers

Faculty and Researchers' Publications

2022

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

Das, Arijit

Monterey, California: Naval Postgraduate School

<https://hdl.handle.net/10945/71923>

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

Downloaded from NPS Archive: Calhoun



Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

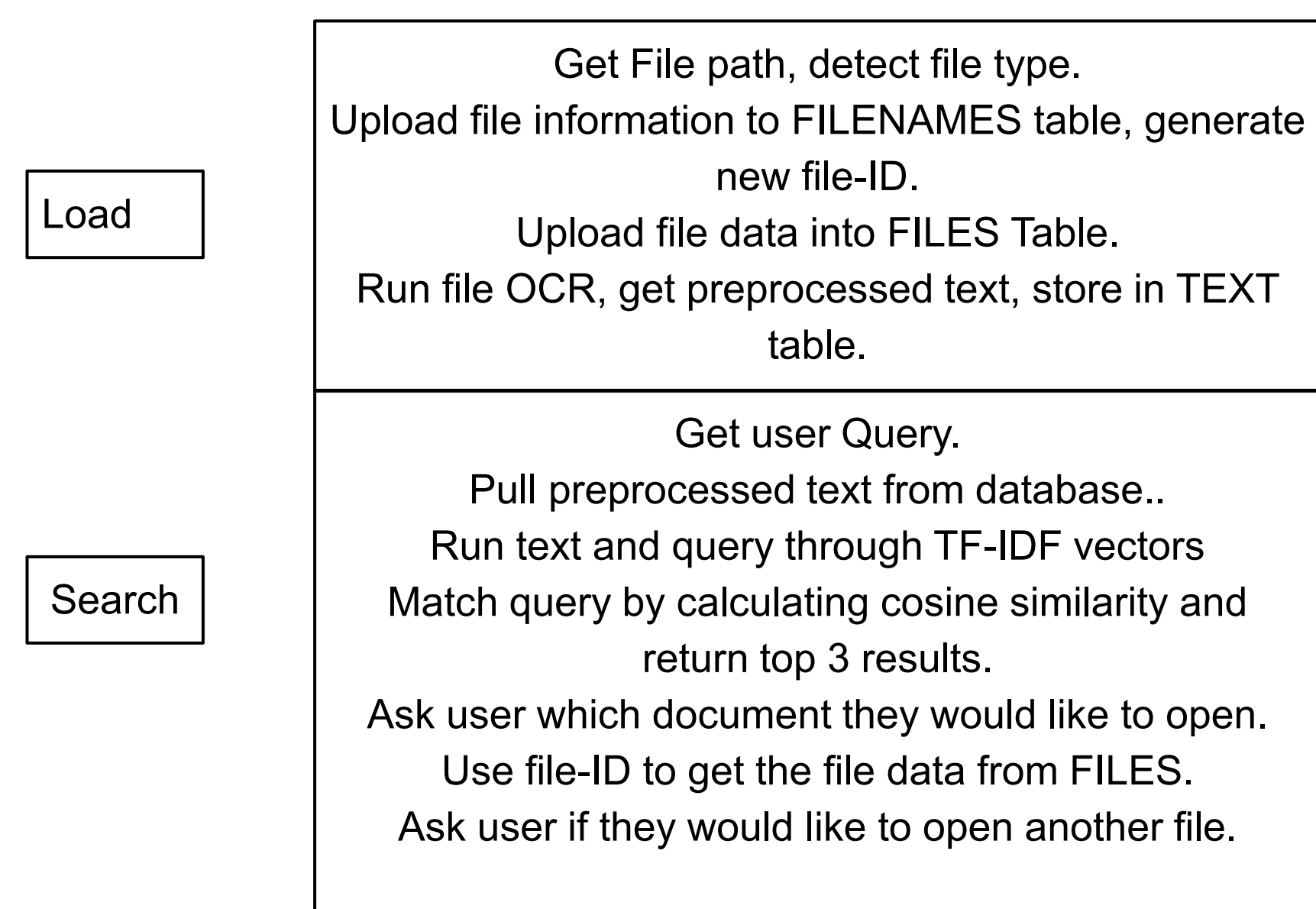
Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

<http://www.nps.edu/library>

Background

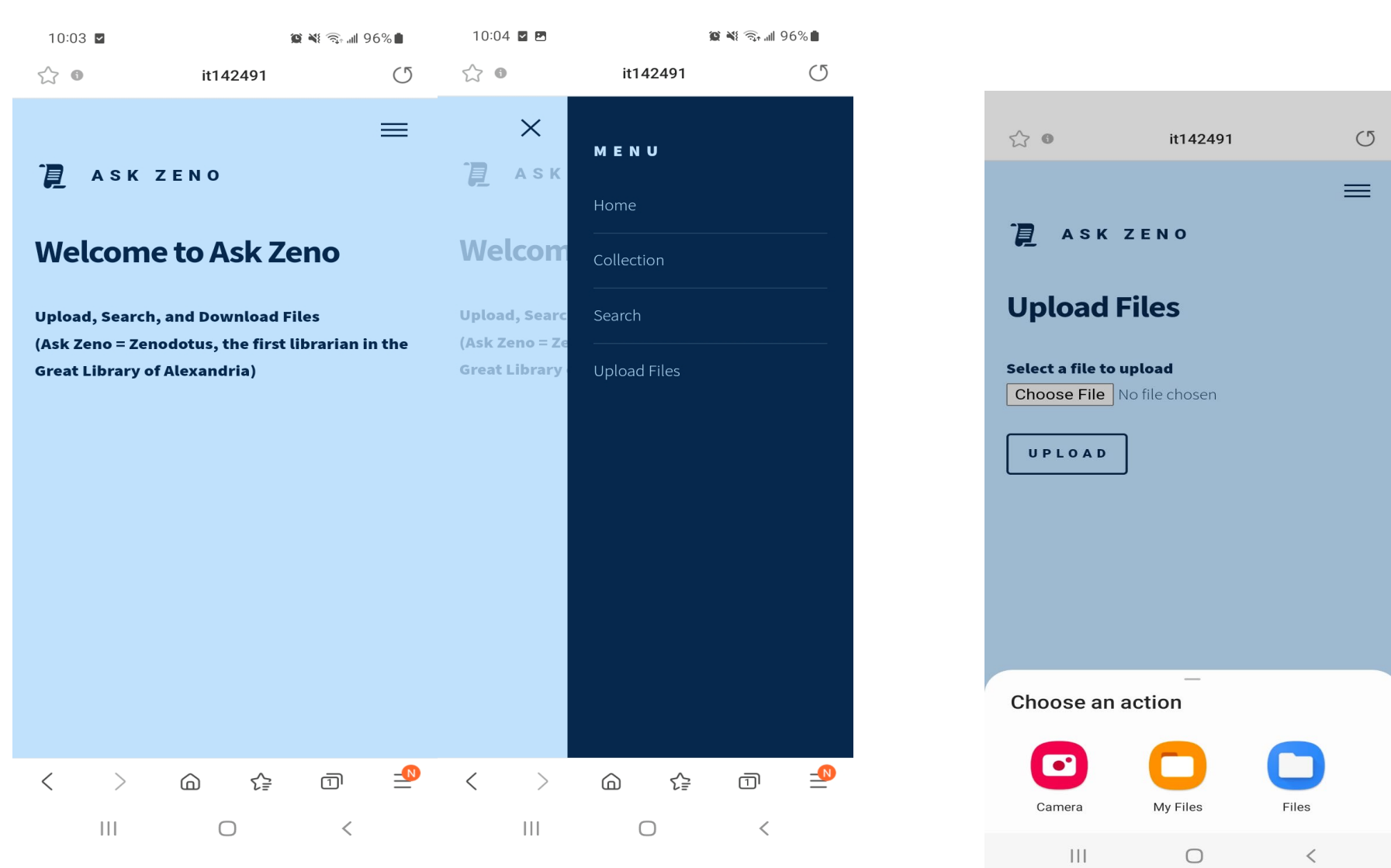
- FY21 effort identified that datasets were summarized into reports or documents.
- Documents accumulate over time and have to be stored and searched over time.
- There are various formats namely PDF, Word, PowerPoint, images and plaintext.
- A Document store architecture has to consider front-end, middleware 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.

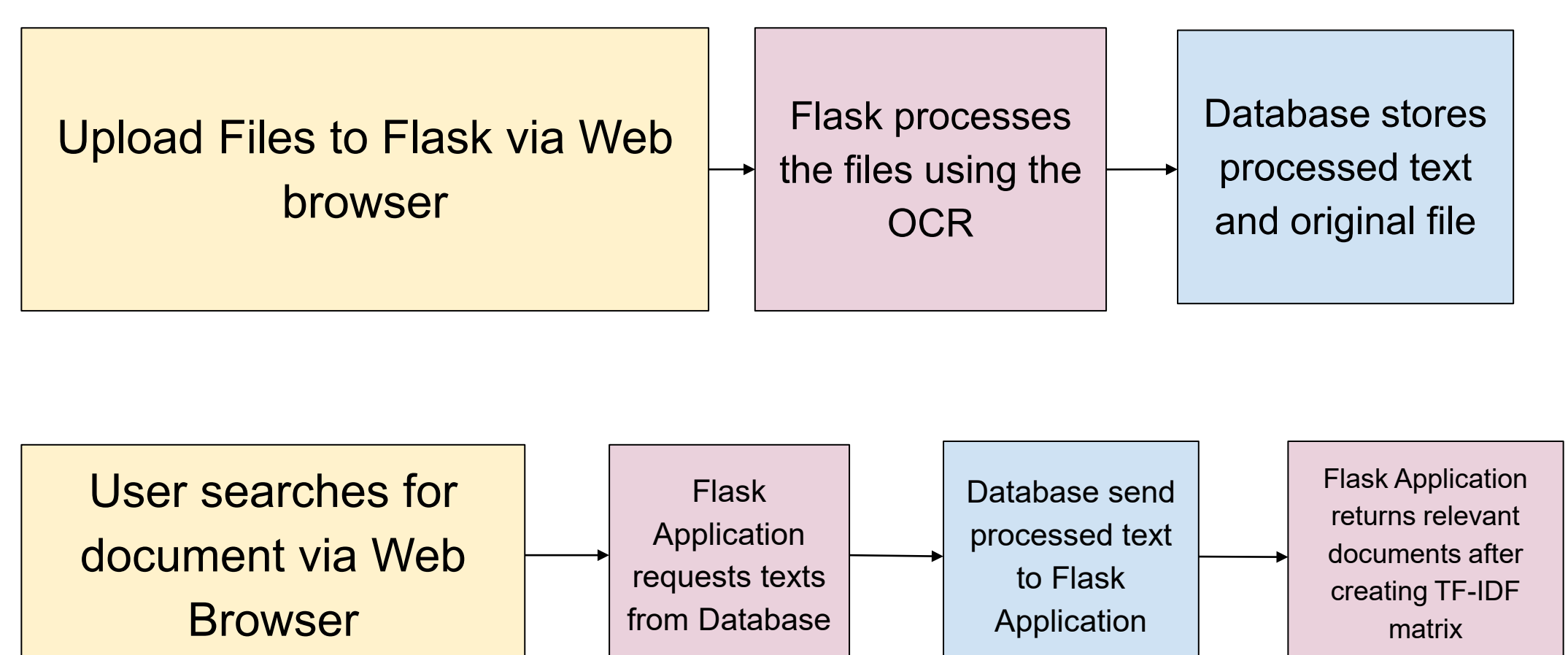
Technology

- 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