



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

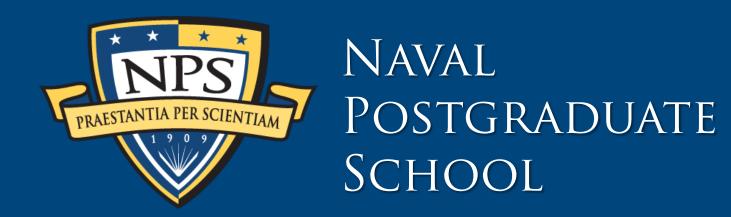


Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

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

http://www.nps.edu/library

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



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 frontend, middleware and back-end.
- NPS team studied the technology challenges taking into consideration scaling, costing and DoD licensing.

Data Flow

	Get File path, detect file type.						
	Upload file information to FILENAMES table, generate						
	new file-ID.						
Load	Upload file data into FILES Table.						
	Run file OCR, get preprocessed text, store in TEXT						
	table.						
	Get user Query.						
	Pull preprocessed text from database						
	Run text and query through TF-IDF vectors						
Search	Match query by calculating cosine similarity and						
	return top 3 results.						
	Ask user which document they would like to open.						
	Use file-ID to get the file data from FILES.						
	Ask user if they would like to open another file.						

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

Technology

User Interface

10:03 🗹	(2)	\$\$ 🖘 III 96% 🖿	10:04 🖬 🖪	9	s 💐 🗊 💷 96% 🛢			
☆ ❻	it142491	C	☆ 6	it142491	C			
ASK	ZENO	≡	Х Дазк	MENU		☆ 0	it142491	
Nelcom	e to Ask Zer	າວ	Welcom	Home Collection		D ASK	ΖΕΝΟ	
Upload, Search, and Download Files (Ask Zeno = Zenodotus, the first librarian in the Great Library of Alexandria)		Upload, Searc	Search		Upload	Files		
		(Ask Zeno = Ze Great Library	Upload Files		Select a file to Choose File			
						UPLOAD]	
						Choose an	action	
< >	<u>م</u> د	⊡ _0	$\langle \rangle$		0	\bigcirc		
		<		uu <i>≀≡</i>	<	Camera	My Files	Files
111	U		111	U		III	0	<

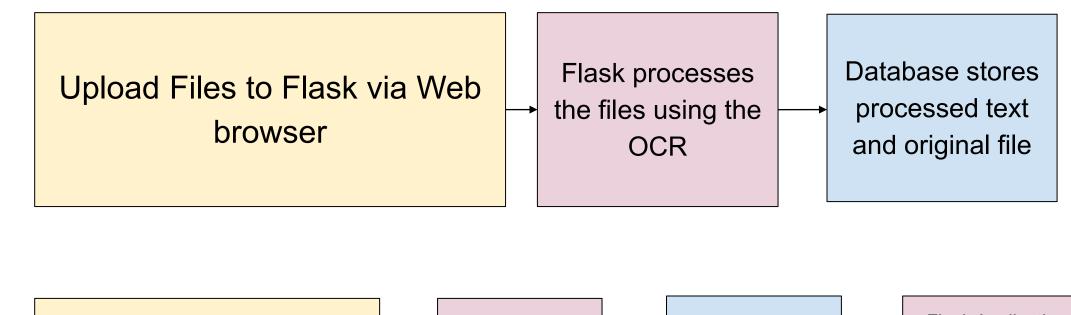
The picture on the left is on the laptop while that on the right is the phone, for the same application.

- 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.

Architecture



- Algorithms recalculate on each document load.
- Topic Sponsor made aware of the system needed.

Future Work

- Data scaling using HDFS.
- Building a Terabyte size repository.
- Involving a community for user interface testing.
- User searches for Flask Application Flask Database send returns relevant Application processed text document via Web documents after requests texts to Flask creating TF-IDF Browser from Database Application matrix 3-Tier Architecture.
 - Better understanding of DoD licensing for software.
 - Evaluating in-house expertise versus contracting.
 - Research prior work done in DoD/DoN IT.

NRP Project ID:

NPS-22-N207-A



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

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