Project Mentor

CAPT Mark Stoops is a career Naval Officer and pilot with over 4,000 flight hours.

Before joining the Navy, he earned a BSEE from Virginia Tech and worked as an engineer and computer programmer.

After joining the Navy, he earned an EMBA from Naval Postgraduate School. He now works for the Navy’s Operational Test & Evaluation Command in Norfolk.

Problem Statement

“We get stacks of documents as thick as phone books... most don’t read past the executive summary.”

- Capt. Mark Stoops, USN

Problem Statement

“Documents deliver information. The links & relationships between them contain meaning. Curation of these relationships can offer clarity.”
Solution Product

Documeta is a visualized document insight & comprehension tool

Solution Statement

Extracting keyword metadata from documents reveals the relevant information on a topic. Visualizing these discoveries interactive, and meaningful way acts as a vehicle of tangential learning.

Solution Objectives & Goals

- Agnostic Document Analysis
- Metadata Extraction
- Visualization & Navigation
- Out-of-Box Web Application
Core Components

- **documeta** will be built in three primary branches of development
  - User Interface
  - Infrastructure
  - Backend

Major Functional Components

- **documeta** will be deployed as a containerized web application, focusing on:
  - Accessibility
  - Portability
  - Security
D3 Javascript API

- D3.js API contains several hundred functions
- Operations can be grouped into the following logical units:
  - Selections
  - Transitions
  - Arrays
  - Math
  - Color
  - Scales
  - SVG
  - Time
  - Layouts
  - Geography
  - Geometry
  - Behaviors
Infrastructure Technologies

Docker
- Used to run software packages called *containers*.
- Containers are isolated from one another and bundle their own application, tools, libraries, and configuration files.
- Containers can communicate with each other through well-defined channels.

Kubernetes
- An open-source container-orchestration system for:
  - Automating deployment
  - Scaling and managing of containerized applications

Memcached
- A general purpose distributed memory caching system
- Used to speed up dynamic database driven websites by caching data and objects in RAM
- Reduces the number of read iops from an external data source

Traefik
- A reverse proxy and load balancer that provides network metrics and integrates with major cluster technology.
Backend Technologies

Python
- Interpreted, object-oriented programming language
- Support for dynamic typing and dynamic binding
- Essential data structures built-in
- Ideal for Rapid Application Development (Agile)

MySQL
- Relational database management system
- Strong index performance
- Support for JSON objects

RESTful API
- Based on the representational state transfer architectural style for web-based communications
- Application programming interface utilizing HTTP requests
  - GET, POST, DELETE, PUT
  - Uniform and predefined set of stateless (server does not store information about client state) web operations

Flask web framework
- "Micro" web framework where you plug in extensions as needed
  - Lightweight, no unused features
  - Flask-RESTful extension supports rapid development of RESTful APIs
Glossary

ACL: Access Control List, a list specifies to the operating system the access and respective operations a user has towards a listed object.

Agile: A software development methodology focused on incremental, iterative development cycles; where requirements and solutions evolve through collaborations with customers via continuous prototyping.

API: Application Programming Interface, an abstract set of functions, communication protocols, and other various tools that specify how software components can interact.

Async: Asynchronous processing, units of work running separately from the main thread (process is not blocked on async operation).

Containers: A standard unit of software that packages up all code and dependencies into a sandboxed, virtual environment.

DNS: Domain Name System, a decentralized naming system used to associate various information with assigned domain names.

DoD: Department of Defense.

DOM: Document Object Model.

JSON: JavaScript Object Notation.

KRB5: Kerberos 5 (current release), computer network authentication protocol that works on the notion of tickets (allow nodes communicating over a non-secure network to prove their identity to one another in a secure manner).

Method Chain: Technique used in Object Oriented Programming where multiple methods are invoked in sequence, with each method operating on the object returned by the previous call.

RDBMS: Relational Database Management System.

REST: Representational State Transfer.

Tangential Learning: A process by which people educate themselves on a topic when it is presented in a context that they enjoy (e.g. video games that teach educational topics).

UI: User Interface.

UX: User Experience.