Week 9: Unstructured data; storage

Objective: In this lab session, you will learn:

1. Image data storage
2. Text data storage

MongoDB is a NoSQL database management system known for its flexibility, scalability, and ease of use. Unlike traditional relational databases, MongoDB stores data in flexible, JSON-like documents, making it suitable for various applications.

This section will use MongoDB Atlas and Google Colab.

MongoDB Atlas is a fully managed cloud database service provided by MongoDB, offering a hassle-free way to deploy, manage, and scale MongoDB databases in the cloud.

**Step 1. Register on MongoDB Atlas**

* Visit the MongoDB Atlas website [here](https://www.mongodb.com/cloud/atlas).
* Click the “Try Free” button to create an account and follow the registration process.

**Step 2: Setting up Users**

* After logging in to MongoDB Atlas
* Create a new project, and select the project
* Navigate to the “Database Access” tab from the left sidebar.
* Click the “Add New Database User” button to create a new user
* Enter the username and password for the new user and assign appropriate permissions

**Step 3: Setting up IPs**

* In the MongoDB Atlas dashboard, go to the “Network Access” tab from the left sidebar.
* Click on the “Add IP Address” button.
* Add your current IP address to the whitelist to allow connections from your location.
* Add IP address: 0.0.0.0/0 to access the database from Google Collab

**Step 4: Creating Clusters**

* Go to the “Clusters” tab from the left sidebar and click on the “Create New Cluster”
* Choose the free cluster – M0, choose AWS as a cloud provider, Ireland as the region
* Click on the “Create Cluster” button to provision the cluster.

**Step 5: Create a database and collection name**

* Go to the cluster then click on add my own data
* Enter database and collection name. Enter one collection name while defining database.
* Come back to the created database, and click on the + to create another collection
* For the Image data, create CBIR db. and two collections: processed images and image metadata
* For the text data, create an NLP db. And two collections: processed doc and doc metadata

A screenshot of a computer

Description automatically generated

**Step 6: Getting Connection String**

* Once the cluster is created, click on the “Connect” button.
* Select “Connect Your Application” and copy the connection string provided. This string will be used to connect to MongoDB Atlas from Google Colab.
* Make sure you refer to the correct database and collection every time you make the connections

**Step 7: Basic Installation**

* Open Google Colab in your browser.
* Create a new notebook and name it week9\_tutorial.

**Step 8: Basic Installation**

* In the first code cell, install the pymongo library by running ‘pip install pymongo’.
* Import the necessary libraries in the next code cell:



**Step 9: Connect to MongoDB Atlas:**

* Paste the connection string obtained from MongoDB Atlas into your code.
* Replace <password> with the password for the MongoDB user you created and <dbname> with the database name you want to connect to.
* Use the following code snippet:



**Step 10: Accessing the Database:**

* Once connected, you can access your MongoDB database and perform operations like querying, inserting, updating, and deleting data.

**Step 11: Test the connection by listing the collections**

**A close-up of a computer code

Description automatically generated**

**Step 12: Storing the Image metadata files into the database**

* Upload the metadata file BB
* Select the collection where you want to query/load into the CBIR database

**A screenshot of a computer

Description automatically generated**

**Step 13: Load the processed image data from a folder to the database**

* Upload the image data folder from BB
* Select the collection where you want to query/load into the CBIR database

**A screenshot of a computer code

Description automatically generated**

**Step 14: Query the database to see if the processed image was loaded**

* Let’s check if one of the images, for example, image12.jpg, is in the DB

A screenshot of a computer code

Description automatically generated

**Text data storage in MongoDB Atlas**

Building on the earlier connection created, metadata was loaded, and textual data from week 8 was processed in MongoDB.

Step 1: Create a new connection; refer to NLP as a database.

A screen shot of a computer

Description automatically generated

Step 2: load the metadata of the processed document file.

A screenshot of a computer program

Description automatically generated

Step 3: load the processed document files

A screenshot of a computer

Description automatically generated

**Extra activities.**

This is part of the image texture extraction from week 8.

Here, Features like Contrast, and dissimilarity among others from GLCM are computed.

A screenshot of a computer program

Description automatically generated