

# Exploring the Representation of a Concept in News Stories - Code Execution Guidelines

Tomás Mc Phillips - 15382091

**Note:** This code is connected to the gmail account chantritualresearch@gmail.com and the script will search for unread emails scrape their data and mark them as read. If no unread emails are available this will be shown on the terminal and the results outputted to an additions.txt file in the same directory, otherwise the data will be added to the local instance of Mongo DB running on your machine which may have an impact on the research of Dr. Cummins. This tool is hosted on csi420-01-vm4.ucd.ie and is currently being used for research while migration to Dr. Cummins' personal server is underway. Therefore no submission of tags for stories should be performed when visiting. Instructions on how to collect the news stories and begin the web application locally are provided below.

**Note:** It is necessary to pull files from git as the submission was too large for Moodle  
All files for this project can be pulled from git:

Git Repo: <https://github.com/mosbyte/Exploring-The-Representation-of-a-Concept-in-News-Stories.git>

## To Perform the Collection of News Stories

1. Python version 3 and pip version 3 are requirement. The following dependencies are needed.
2. Download and install MongoDB based on your operating system following the guidelines of this online tutorial from MongoDB, <https://docs.mongodb.com/manual/installation/>.
  - a. `pip install BeautifulSoup`
  - b. `pip install pymongo`
  - c. `pip install --upgrade google-api-python-client google-auth-http2 google-auth-oauthlib`
3. To use an example data set and add to it, open the command line and change directory to fyp\_mern/. Run the first command for bash terminals and the second for windows. The paths should be changed to suit your system.
  - a. `sudo mongorestore --db chant_ritual --drop /pathto/chant_ritual/`
  - b. `mongorestore -h localhost -d chant -o C:\pathto\chant_ritual`

The Python files in the python directory perform the following operations:

**collect\_new\_stories.py** - Run this script to collect news stories for chant and ritual and add them to their appropriate database collection. The collection process handles chant first and then ritual.

**gmail\_collect.py** - This file will connect to the gmail account chantritualresearch@gmail.com using the credentials.json and token.pickle file. It will open any unread emails for chant and ritual, scrape their RSS link and mark the email as closed.

**xml\_parse.py** - This file contains the Scrape class which is responsible for web scraping a RSS link received, cleaning the data set and then returning the data so it can be stored in the database.

**mongo\_news\_db.py** - Stories retrieved are added to their correct database collection based on the database term received, either chant or ritual.

**cloneDB.py** - Cloning a database and its collections can be performed with this script. The user must specify in the file the exact database name they wish to clone and the name which they wish it to be cloned under.

To collect news stories over the command line, change directory to where the `collect_new_stories.py` file is contained and run:

```
python collect_new_stories.py
```

## To Host the Web Application Locally

The file structure of this stack has the routes for chant and ritual, the Node.JS web server and mongoose schema defined in the `fyp_mern` directory. All front end components exist in the `fyp_mern/client` directory with reused components being in a `components` directory.

1. Install Node.JS on your system. Online tutorial: <https://nodejs.org/> for all platforms.
2. In the directory `fyp_mern/client` run the following command
  - a. `npm -i axios bootstrap recharts react-charts react-router-dom`
3. In the directory `fyp_mern/` run the following command
  - a. `npm -i express body-parser cors mongoose`
4. In the `fyp_mern` directory run the following command in the terminal suited to your system. This will startup the node web server locally on port 5000 and will access the database of example data restored in part 4 above.
  - a. BASH: `nodejs server.js`
  - b. Windows: `node server.js`
5. In a new terminal open change to the directory `fyp_mern/client` and run the following command to host the ReactJS front end locally on port 3000:
  - a. `npm start`