



Article metadata is initially scraped from RSS feeds by feed scraper classes (each one specialized to a particular source) the obtained metadata is passed to the feed manager, and stored as CSV in both a temporary backup file, as well as added to a folder that represents a scrape queue. The metadata controller gradually works through these queued scrapes and compares it to it's existing store of meta data (a permanent metadata CSV store file.) Articles that don't yet exist get a UUID generated for them, and added to the CSV store as well as sent to the database manager to get sent to the database.

Periodically, the article manager queries the database to find articles that don't yet have content associated with them. The links are sent to the content manager, and based on the source, the content manager will run the scraper for that source and article link. The raw scrape HTML is stored, and the cleaned version is returned to the content manager, which stores this as well, with the associated UUID as the filename. The content manager then also sends the content text to the database manager to get stored in the associated metadata row

