

Data mining project - second checkpoint

In this exercise you design and implement a database that your data will be stored in, and write a command line interface to your scraper that will provide different options and functionality.

Command line interface

1. Wrap up your web scraper to be able to call it with different arguments from the terminal.
2. Examples to arguments you can use: different data types to scrape (e.g. hashtags on instagram, product categories on amazon), timespan/dates to scrape (scrape only data that was created in a certain timespan, etc), different technical parameters (DB params, number of iterations to scrape, etc).
3. Use [click](#) or [argparse](#) packages.
4. Add documentation to the different CLI arguments to the README.md file, including the default values.

Database implementation

1. Design an ERD for your data. Think about which fields should be primary and foreign keys, and how you distinct new entries from already existing ones.
2. Take notice of primary and foreign keys.
3. Write a script that creates your database structure (python or sql), it should be separate from the main scraper code (but should be part of the project and submitted as well).
4. Add to your scraper the ability to store the data it scrapes to the database you designed. It should store only new data and avoid duplicates.
5. Work with Mysql database.
6. If you'd like you can use ORM tools such as [sqlalchemy](#).
7. Add a DB documentation section to the README.md file, including an ERD diagram and explanations about each table and its columns.

Good luck!