

On December 8, 2022, the UIUC course, CS 410: Text Information Systems is calling for project submittals that explore topics covered in the course. The project team, MLTK_410 is proposing a project submittal covering 'Theme 5: Free Topics' as outlined in the 'CS 410 Project Topics' document. Team members will consist of: captain Luis Chavez (lwchave2@illinois.edu), Adam Michalsky (adamwm3@illinois.edu), Yijing Yang (yijingy2@illinois.edu), and Sri Ram Eswaran (srirame2@illinois.edu).

MLTK_410 is proposing developing an application, 'MLTK', that will integrate MeTa with NLTK to provide a sentiment analysis of any media shared on social media. The application is aimed at social media owners, however for the sake of showcasing the application will be developed such that social media users can also utilize it to the extent provided. MLTK, will allow social media owners to identify and rank the sentiment behind any media provided a collection of documents pertaining to the media type in question. Social media owners should value having access to a sentiment analysis tool as this information can be used to direct platform growth, leveraged against competitors, sell access to marketing agencies, etc.

MLTK's development environment will consist of Git version control through GitHub so as to allow for parallel development and collaboration. The four main tools used in this application are Python, Twitter Scraper, MeTa, and NLTK. The dataset required for this application to work is a collection of documents, so Twitter along Twitter Scraper will be used as the social media platform and document collecting tool to provide the necessary dataset to showcase MLTK. MeTa will then gather relevant documents from the collection of documents, then NLTK will perform a sentiment analysis over the relevant documents. The stack of programs will then be bundled together and labeled MLTK and provided as an application. The application will prompt the user to provide a media type (e.g., books, music, movies, games, etc) and a specific media item (e.g., "Pride and Prejudice", "Bad Habit", "Halloween Ends", "Elden Ring", etc), once provided, the program will output the favorability of the media in question. MLTK's sentiment will be compared against media review outlets, e.g., Rotten Tomatoes for movies, IGN for games, etc to assess functionality.

MLTK's development is split up into four tasks, each amounting to at least 20 hours. Luis Chavez is assigned to Development, Production, and Testing, this includes setting up the development environment required to begin software development, developing the application programming interface (API), setting up the production environment for the software usage tutorial presentation, and coordinating across each tool of the project. Adam Michalsky is assigned to Document Collection, this includes using an open source twitter scraper to retrieve a collection of documents per the user's media type input, formatting retrieved tweets into usable docs for MeTa, formatting metadata into usable docs for MeTa extension (see MeTa/NLTK Development), and providing a sample collection of documents to the MeTa and NLTK developers to facilitate parallel software development. Sri Ram Eswaran is assigned to MeTa Development, this includes using native MeTa features to filter out a list of top X relevant documents per the user's specific media item query, and extending MeTa with a heuristic ranking function to leverage metadata and document data. Yijing Yang is assigned to NLTK development, this includes using native NLTK features to perform a sentiment analysis on the relevant documents provided by MeTa, and extending NLTK with a heuristic ranking function to leverage metadata and document data.