INFO-116

*Semantic Technologies* 

Made by

NUMBERS HERE

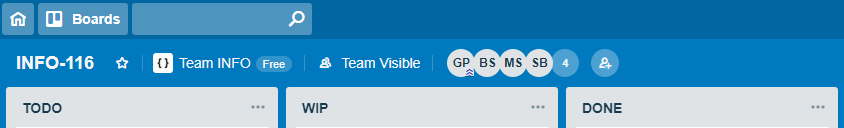
INFO 116

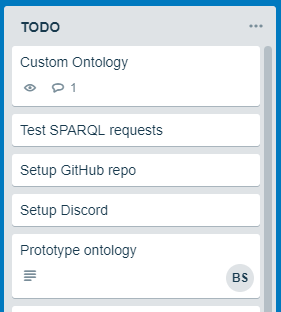
*Group assignment*

Floating Titanic sinking NY show

We chose the article named “Floating Titanic sinking NY show” for our assignment.

We first started with selecting a group leader and he setup the infrastructure for our group. We created a online task board using a website named Trello.com. It allows us to create task that each member can assign to themself. This helps us know what each person is working on at any time. We split our board into three sections, “TODO”, “WIP” and “DONE”. This is the three states that each task can exist in.

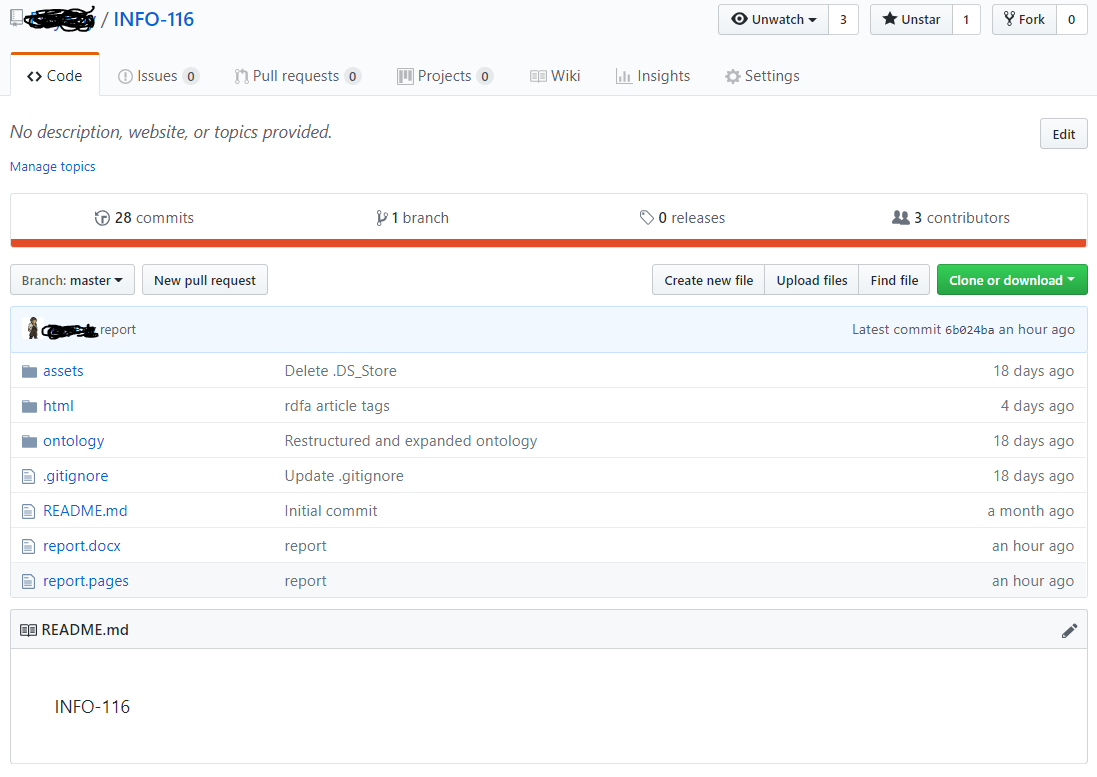


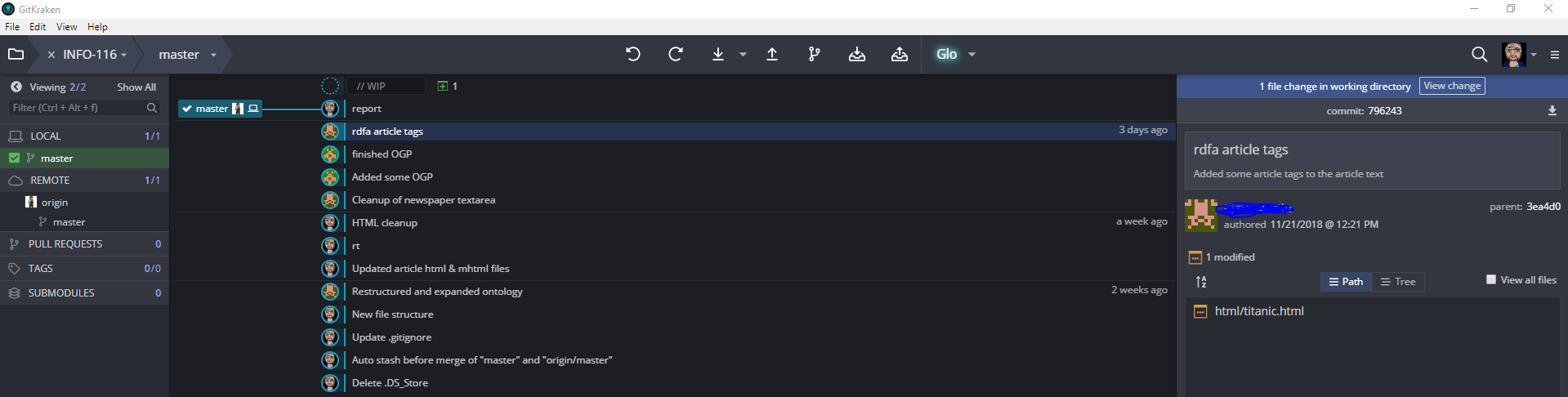


We split the assignment into as many small tasks as possible. To the left you can see a tiny portion of our tasks from the TODO (to do) section. Once a person starts working on a task, he drags it over into the “WIP” (work in progress) section and signs his name on that task. When the task is done, we just move em over to DONE and signs it out as finished.

After setting up the Trello board the group leader created a Discord server for every group member to join. Discord is a application that allows people to host their own communication portal where each member has access to text and voice chat and can easily share files and pictures with each other. This made the group communicate a lot better and where able to share info about where and when to meet up and work on the assignment.

The last thing we did for infrastructure within the group was to setup a GitHub repository. This allows every member to work on the project simultaneously and gives us a centralized online hub to host our project while we work on it as well as a version control tool, just in case something goes wrong.

We also got a program named GitKraken to help us with managing the git using a visual interface.



We started out by extracting the HTML code from “The British Newspaper Archive” because the file provided to us from mitt.uib was a MHTML file that contained a lot of extra bloat such as portable css and javascript and other stuff preventing us from running the file through any websites validating structure data. After grabbing the page HTML, we started cleaning up the file in a text editor to make it more readable.

After cleaning up the HTML file..