|  |
| --- |
|  |
| Fighting Mongooses  1234 Beacon Way SE  **(206) 555-4095**  Github.com/MANollette/RedditCrawler/ |

Potential clients,

This capstone project is being initially designed as a web crawler geared towards Reddit. Its purpose is to take text-based user input, and monitor subs of your choice for posts that match or contain that input. It will then notify you via email. The core program I aim to complete is a console application that accepts simple text input to this end. It stores the information on what to scan for and what has already been found in a .txt file, and sends an email on that basis.

Our primary audience is any Reddit user who wishes to monitor content on the website for updates, and be quickly notified of them. The primary intent is to assist goods traders and moderators with expedient email notifications of posts, but the product could easily be used by anyone who wishes to monitor these posts.

Currently, Reddit lacks the functionality to notify users when posts about certain items have been made. For abundant items, this poses no problem. However, rarer items can be frustratingly difficult to find because of this. Often, they are few in number and the offer lasts only for a few hours. The wait time between offers is also highly variable. The next offer may show up within a few days, or may take months and even years to come up again. Without a notification system, users are left to repeatedly and manually check for these items. This is incredibly tedious, frustrating, and wears users down.

Here is the current process Reddit users must go through to find items for sale:

1. Navigate to the relevant subreddit
2. Look through posts to find items
3. Contact seller

Out of these steps, step two is the most time sensitive. Often high-demand items will be sold within minutes of being posted. Users must read through posts to determine if offers are still available. Without constant checking, all but the most dedicated of users will be unable to find what they are looking for in a timely manner. Without notification, they are left to wait for the next offer which can take months to reappear.

The three main objectives of our Reddit Web Crawler are to:

* Read/write user search terms
* Read/write subreddit post URLs containing said search terms
* Send email notifications once read/write is completed

The deliverable/result will be a Console Application that will: read/write user search terms to a .txt file, read/write subreddit post URLs that contain these search terms to a separate .txt file, and send emails that will notify users that new posts have been made.

In the event extra time is available for further deliverables we would also like to expand the functionality of our crawler. The first priority would be to enable the console application to be run two ways: with the default console window allowing user input for search criteria or credentials, and in the background as a windows process. This will allow users to set the application to run in the background whenever their computer is on, rendering it invisible but functional for user convenience.

The second priority would be to expand the user interactivity, allowing more direct manipulation of the different criteria entered. For the core functionality, the application will allow the amending of data, or the rewriting of data to the text files. This step would further allow manipulation or removal of individual lines/search items.

The final priority, while more useful, is also more involved. We would like to incorporate an Azure database in the system for the search methods to write to and read from, rather than a locally stored text file. This expands our experience with cloud technology, and provides a much more practical exercise in application development.

*Warm regards,*  
  
Michael Nollette  
Developer  
Tim Thor  
Developer