|  |
| --- |
|  |
| Vorbit: Reddit Data Mining and Machine Learning Software |
| Neumont University Capstone |
|  |
| **Felix Milea-Ciobanu** |
| **12/01/2013** |

|  |
| --- |
|  |

Project Success

Vorbit is a data mining and machine learning tool for the social news website Reddit that provides its users with two major features: first, the ability to gather analytical data about user posting behavior on specific segments of Reddit; secondly, to make automated posts to Reddit using machine learning bots that learn from the gathered data. These features are especially useful to Reddit moderators, who already use an array of tools to perform similar functions on the subreddits they manage.

The data mining component allows the user to store the mined data into modular data sets that can be easily exported outside of the application. Vorbit’s wide range of configuration options allows users to gather very specific, comprehensive data. Similarly, the machine learning component can be configured to behave in very specific ways and to post content that is indistinguishable from that authored by a human.

With Vorbit, Reddit moderators can learn about their communities and manage automated posting tasks all from one application. Its wide range of configuration options, ease of use, and advanced features makes it the tool of choice for Reddit moderators and hobbyists alike.

Target Users

Vorbit targets technically oriented Reddit users, specifically subreddit (user-managed sub-community on the site) moderators, who already use an array of tools to gather analytical data about the communities they manage and bots to automate many moderation tasks. A typical moderator will be interested in monitoring user posting behavior in their community and compare it with that of other segments of the site. Vorbit’s machine learning component provides moderators with the advantage of posting automated messages in reply to posts in their community that are indistinguishable from those composed by a human moderator.

Alternatively, technically oriented Reddit users can use Vorbit’s machine learning component to automatically generate comments to posts on the site that emulate those made by real users for the purpose of gaining karma (a point system used to measure the quality of a user’s content) on the site.

Project Backlog

## Data Mining Component User Stories

##### The user can increase the number of processes used by a data mining task.

###### User will be notified that the action was successful or not.

##### The user can sort the data currently collected by the mining task.

###### User is presented with a table representation of sorted data.

##### The user can filter the data currently collected by the mining task.

###### User is presented with a table representation of filtered data.

##### The user can export mined data outside of the application.

###### User will be notified that the action was successful or not.

## Machine Learning Component User Stories

##### The user can create and remove a new machine learning bot.

###### User will be notified that the action was successful or not.

##### The user can configure the posting rate of an existing machine learning bot.

###### User will be notified that the configuration was successfully saved.

##### The user can configure the subreddits an existing machine learning bot can post to.

###### User will be notified that the configuration was successfully saved.

##### The user can configure the type of posts an existing machine learning bot can reply to.

###### User will be notified that the configuration was successfully saved.

##### The user can instruct a machine learning bot to start or stop automatically generating and posting content to Reddit.

###### User will be notified that the action was successful or not.

##### The user can add and remove a Reddit account from an existing machine learning bot.

###### User will be notified that the action was successful or not.

##### The user can view the posts made by a machine learning bot.

###### User is presented with a table representation of submitted posts.

##### The user can view the current status of a machine learning bot.

###### User is presented with a description of the current status of bot.

2-Week Plan

## Week #1 User Stories

##### The user can create and remove a new data mining task.

###### User will be notified that the action was successful or not.

##### The user can configure the behavior of an existing data mining task.

###### User will be notified that the configuration was successfully saved.

##### The user can start and stop a data mining task.

###### User will be notified that the action was successful or not.

## Week #2 User Stories

##### The user can view the data currently collected by the mining task.

###### User is presented with a table representation of collected data.

##### The user can view the current status of a data mining task.

###### User is presented with a description of the current status of the task.