# User's Guide: Using the DharmaWheel.net Post Scraper Program

## Introduction

This console application is designed to scrape posts from the DharmaWheel.net forum, specifically from the user specified as "Malcolm". It will gather all the posts from this user and save them into text files in your computer. The program supports incremental updates, meaning it can scrape only the latest posts since the last scraping operation.

## Pre-requisites

Before you can run the program, make sure that you have the following software installed on your machine:

* .NET Core 3.1 or later

## How to Use

### Step 1: Launch the Program

To start the application, navigate to the directory where the program is located using your command line or terminal and run the following command:

arduinoCopy code

dotnet run

### Step 2: Choose Scraping Mode

Once the program starts, it will prompt you to choose between scraping all posts or just the latest ones. If you want to scrape only the posts that have been made since the last time you ran the program, enter **yes**. If you want to scrape all the posts, enter **no**.

### Step 3: Wait for the Program to Complete

The program will now log into DharmaWheel.net, navigate to Malcolm's posts, and start scraping them. It will fetch 20 posts at a time and save them in a temporary storage.

Once it has collected 500 posts, it will write them to a file and clear the temporary storage. If you chose to scrape only the latest posts, it will stop once it encounters a post that has already been saved in a previous run.

The program will keep you informed about its progress by printing messages to the console.

### Step 4: Check the Output

The program saves the scraped posts in text files in the same directory where the program is located. The files are named in the format "Malcolm\_posts\_#.txt", where # is the number of the file. If you chose to scrape only the latest posts, the file will be named "Malcolm\_posts\_0\_date.txt", where date is the current date.

Each file contains up to 500 posts, each post consisting of the author's name, the date and time of the post, the title, and the content of the post.

## Customization

If you want to modify the program to scrape posts from a different user, you will need to change the following constants at the top of the Program class:

* **AuthorName**: Change this to the name of the user whose posts you want to scrape.
* **AuthorId**: Change this to the user ID of the user. You can find the user ID by navigating to the user's profile on DharmaWheel.net and looking at the URL.

The login credentials used for DharmaWheel.net are hardcoded. If you want to use a different account, change the **username** and **password** fields to your own username and password.

**Please be aware that with new accounts on the Dharmawheel platform such as “nyingje”, there may be limitations on scraping the most recent posts from the same day. This limitation seems to be tied to the age and activity level of the account used for scraping. It appears that the platform places restrictions on viewing same-day posts for newer accounts, possibly to limit automated activity or to encourage engagement with the community. As such, you may only be able to scrape posts from previous days with a new account. If you need to scrape posts from the same day, you may need to use an older account that has some level of activity. Please make sure to respect the platform's guidelines and terms of use when using this scraping tool.**

## Troubleshooting

If you encounter any issues while running the program, check the console output for error messages. If you see a message saying "Error fetching post page", it means that the program had trouble loading a page. This could be due to a temporary issue with the site or with your internet connection. Try running the program again later.

If you continue to experience issues, make sure that your .NET Core installation is up to date and that you have a stable internet connection.