# User Guide: KeywordsCategorisedWordPDF Program

## Overview

The KeywordsCategorisedWordPDF is a console-based program designed to organize and categorize text data from forum posts. It automates the process of categorizing, sorting, and saving these posts into Microsoft Word and PDF files based on user-specified keywords and their synonyms.

## Purpose

The program is aimed at facilitating data organization and analysis, especially when dealing with a large number of text files. By categorizing posts based on certain keywords, it simplifies the process of extracting relevant information.

## Preparation

1. **Keywords File**: The program requires a **Keywords.txt** file located in the same directory as the program's executable. Each line of this file should contain a keyword or a group of synonyms separated by slashes (/). For example:

dzogchen/great perfection/ati/atiyoga

1. **Text Files**: The text files containing forum posts should be structured in a specific format. Each post should start with an **Author:** line, followed by a **Date:** line, a **Title:** line, and a **Content:** line. Additional lines are considered part of the content until the next **Author:** line appears.

## Operation

1. Run the program. You can do this either from an Integrated Development Environment (IDE) or from the command line.
2. When prompted, enter the folder path containing the text files to be processed. Make sure to provide the full path to the directory.
3. The program will then read all text files in the specified directory, categorize the posts based on your keywords, sort them by date (in ascending and descending order), and save them into Word and PDF files.

## Output

The program creates a new subdirectory named **CategorisedWordsAndPDFs** in the specified folder. In this subdirectory, it generates Word and PDF files for each keyword. These files are named in the following format:

* **Malcolm\_<keyword>\_Ascending.docx/pdf**
* **Malcolm\_<keyword>\_Descending.docx/pdf**

Each file contains posts sorted by date, with each post including the author's name, the date of the post, the title, and the content of the post.

## Notes

* The program matches keywords and their synonyms case-insensitively.
* It uses complete word match, which means it won't categorize a post under a certain keyword if the keyword is part of another word in the post. It has to be a standalone word or phrase.
* The program requires Microsoft Office to be installed for creating Word and PDF files.
* The application uses .NET Framework features, which may require installation or updates depending on your system configuration.