# User Guide: KeywordExtractor Program

## Overview

KeywordExtractor is a console-based program designed to process text files and identify the occurrence of specific keywords within these files. It categorizes data based on user-specified keywords or their synonyms, which can help in extracting specific information from a large number of text files.

## Purpose

The program's main purpose is to assist with data analysis by identifying and counting the occurrences of certain keywords in a set of text files. This can be particularly useful for tasks like text mining, sentiment analysis, or natural language processing, where understanding the frequency of specific words can provide valuable insights.

## Preparation

1. **Keywords File**: The program requires a **OriginalKeywords.txt** file located in a specific directory of your computer. This file should contain the keywords to be identified in the text files. Each line of the file can contain a keyword or a keyword and its synonyms, separated by slashes (/). For example:

dzogchen/great perfection ati/atiyoga advaita/advaita vedanta vedanta

1. **Text Files**: The text files that the program will process should be placed in the directory which you will specify when running the program.

## Operation

1. Run the program. You can do this either from an Integrated Development Environment (IDE) or from the command line.
2. The program will prompt you to enter the directory path containing the text files to be processed. If the directory you entered does not exist, the program will ask you to enter a valid directory.
3. After a valid directory is entered, the program will read all text files in the directory and count the occurrences of each keyword in the text files.

## Output

1. The program generates a **KeywordsOccurrences.txt** file, which contains each keyword (or group of synonyms) along with its occurrence count in the processed text files. The keywords are ordered by their occurrence count in descending order. The format of each line in this file is **keyword: count**.
2. It also generates a **Keywords.txt** file that contains all keywords that have at least one occurrence in the processed text files.

## Notes

* The program is case-insensitive, which means it treats uppercase and lowercase letters as the same when matching keywords.
* It uses complete word match. This means it won't count a keyword if it's part of another word in the text. The keyword must be a standalone word or phrase.
* Special characters in keywords are escaped automatically, so you don't need to worry about escaping them in the **OriginalKeywords.txt** file.