# Overview

This program is used for extracting words from text file, transfer words to original format and then save in database

# Functionality

* 1. Extracting from text--extxt
     1. Variables

File Path: Input text file full path

* + 1. Return value

Return all words except stop words. Stop word defined in external library. Return as a list.

* 1. Transfer word to original(lemmatization)--transword
  2. Compare with database and insert new words—insnew
     1. Notification

When insert finished

* 1. Output new words to text file—output

Select range by source or after a specific date, output formatted text file

# Diagram

# DateBase

In database, currently there are following columns

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Description | Type/Sample | Comments |
| Index | Given index for words | Integer/1 |  |
| TEXT | Word text data | String/accept |  |
| Date | Date includes the word | YYYY/MM/DD/2020/12/31 |  |
| Source | Source of the word | String/Verbs1 |  |