Para el programa yo creé 3 clases:

1. La clase pair
2. La clase pair list
3. La clase file reader

La clase pair lo que hace es definir la posición “x” y “y” de la palabra.

Debido a que una palabra puede tener múltiples posiciones, la clase pair list, que es un ArrayList, que almacena todas las posicione de una palabra.

La clase file reader es la más importante, ya que se encarga de toda la parte lógica.

Aquí, se crea un ArrayList de las palabras que están en el archivo de excepciones. Las palabras están separadas por salto de línea. Luego se elimina estas palabras del AVL.

Para almacenar las palabras utilizo un AVLTreeMap, donde el key de cada entry es cada palabra, aprovechando el ordenamiento del AVL, y el value asociado a cada palabra es el ArrayList de pair-posiciones.

Para ingresar los entries en el AVL lo que hago es pasar a la palabra por varios filtros:

Por ejemplo: si tiene un tamaño menor que 2 la convierte en un string vacío, si determina si es un digito, elimina los signos de puntuación al inicio y al final de la palabra. El último filtro es una conversión de la palabra con la primera letra en mayúscula y el resto en minúsculas, para facilitar y tener un formato.

Luego, se elimina las palabras de excepción del AVL, para que no sean impresas y los strings vacíos.

Por lo tanto al pulsar el botón “Generar índice invertid” imprime en consola y escribe en archivo de texto solo las palabras restante o válidas que están en el AVL. Y abre el archivo de texto para visualizar el índice.

Para el sistema de búsqueda hago que la palabra ingresada se convierta en al formato de la primera letra en mayúscula y luego minúsculas. Si encuentra la palabra la imprime, sino avisa que no la encontró.

Lo mismo sucede para la búsqueda de un grupo de palabras, pero estas palabras deben estar separadas por comas, para la búsqueda, ya que el delimitador.

Adicionalmente, se puede ingresar nuevas palabras como excepciones al archivo de excepciones.

Lo importante es que se debe brindar siempre las rutas de la ubicación de los archivos de texto a leer y el de excepciones.