



**UNSA**

Universidad Nacional de  
San Agustín de Arequipa

**UNIVERSIDAD NACIONAL DE SAN AGUSTÍN**

**FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS**

**ESCUELA PROFESIONAL INGENIERÍA DE SISTEMAS**



## **“Detector de plagio”**

CURSO: Estructura de datos y algoritmos

Docente: Ing. Jorge Cristhian Chamby Diaz

---

Integrantes:

- Chura Puma, Mario Franco
- Contreras Mamani Claudia
- Llaique Chullunquia, Angie
- Aco Tito, Anthony Edwin





# Clases del Proyecto

AVLTree.java

Document.java

GUI.java

Main.java

Node.java

Phrase.java

PlagiarismChecker.java

ResultChecker.java





# AVLTree.java

## 1. Definición de un AVL

Un árbol AVL es un árbol binario de búsqueda que satisface la condición de estar balanceado.

- **Propiedad de orden de nodos:** Todas las claves en su subárbol izquierdo son menores que la clave del nodo y todas las claves en el subárbol derecho son mayores.
- **Propiedad de balanceo:** Para cada nodo, la diferencia de altura entre el subárbol izquierdo y el subárbol derecho es a lo mucho 1

## 2. Atributos

El AVL cuenta con solo un atributo, el cual es el nodo raíz:

```
private Node<TYPE> root;
```

## 3. Métodos

El AVL cuenta con los siguientes métodos:

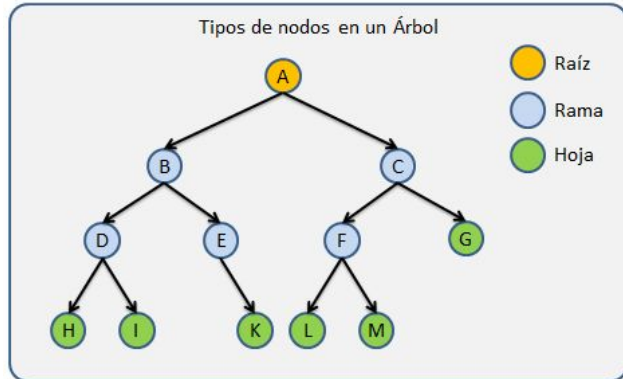
- `public int height(Node<T> node)`
- `public int max(int n1, int n2)`
- `public Node<T> rightRotate(Node<T> current)`
- `public Node<T> leftRotate(Node<T> current)`
- `public Node<T> insert(Node<T> current, T value)`
- `public int getBalance(Node<T> node)`
- `public Node<T> find(AVLTree<T> tree, T v)`



## 1. Definición

Se le llama **Nodo** a cada elemento que contiene un **Árbol**.

**Nodo Raíz:** Se refiere al primer nodo de un **Árbol**, Solo un nodo del **Árbol** puede ser la **Raíz**.



## 2. Atributos

Cada nodo tiene una altura, datos y un hijo izquierdo y derecho

```
private int height;  
private Node<TYPE> left, right;  
private TYPE value;
```



# Document.java

## 1. Definición

Se recibirá el nombre del archivo y se le convertirá en una clase Document.

Esta clase contiene tiene 3 atributos y 2 métodos que nos ayudan tanto a leer y crear un árbol AVL, como también a verificar si existe similitud.

## 2. Atributos

Cuenta tiene 3 atributos:

```
private String fileName;  
private Phrase phrase;  
  
final static int max_number_word=10;
```

## 3. Métodos

Cuenta tiene 2 métodos:

**Método para crear un árbol AVL**

```
public AVLTree<String> createAVL ()
```

**Método para verificar la similitud(plagio)**

```
public boolean match (AVLTree<String>  
tree)
```



Phrase.java

## 1. Definición

Esta clase es una estructura de un tipo de Frase. Cada frase contiene una cantidad máxima de palabras (10).

En los archivos a comparar, se separan en grupos de frases que serán los nodos de los árboles.

Esta clase contiene 3 atributos y 1 método.

## 2. Atributos

Cuenta tiene 3 atributos:

```
private StringBuilder data;  
private int numbersWord;  
final static int max_number_word=10;
```

## 3. Métodos

Método que ayuda a agregar palabras a una frase dada y cada vez que la frase llegue a 10 palabras quitaremos la primera palabra y agregue la nueva palabra al final de la oración:

```
public void addword(String word)
```

### Características de StringBuilder :

- Son mutables (Su tamaño y contenido pueden modificarse)
- Más eficiente que StringBuffer siempre que no se requiera trabajar con múltiples hilos (threads) .



## 1. Definición

Esta clase ayuda a cargar y leer los archivos a verificar(BD). También carga el archivo a comparar y examina con los demás.

Esta clase contiene 4 atributos y 2 métodos.

## 2. Atributos

Cuenta tiene 3 atributos:

```
static List<String> listDocuments = new ArrayList<String>();  
static List<AVLTree<String>> trees= new ArrayList<AVLTree<String>>();  
static String respuesta="";
```

## 3. Métodos

Método que indica que no hubo errores con la lectura de rutas de los archivos de la BD.

```
public boolean LoadFiles (String[] paths)
```

Método que otorga los resultados del sistema de detección de plagio con respecto al archivo con/sin plagio

```
public ResultChecker verifyPlagiarism (String path)
```



# ResultChecker.java

## 1. Definición

Esta clase contiene un arreglo de booleanos donde están las comprobaciones de los archivos. Se explica si es o no plagio.

También contiene los títulos de los archivos de la BD.

## 2. Atributos

```
private boolean [] result;  
private String [] titulos;  
private int i;
```

## 3. Métodos

Método que retorna los resultados de las comprobaciones.

```
public String imprimir()
```





## 1. Definición

La clase Gui es la que dará forma a la interfaz gráfica, por medio de los botones se cargarán los archivos y se verá el resultado de cada comparación.

## 3. Métodos

Contiene 2 métodos:

Método en los que se edita y se especifica las características de los atributos y las que nos mostrará la interfaz

```
public void Display()
```

Método para entrar al directorio y escoger los files

```
private static String getDirectory()
```

## 2. Atributos

```
public final JTextField dirPath;
public static JTextArea processArea;
public final JButton checkButton;
public final JButton selectButton;
public final JPanel contentPanel;
public JPanel subPanel;
public final JLabel dirLabel;
public final JLabel processLabel;
public JScrollPane pane;
public static JProgressBar pbar;
public final JButton selectButtonArchivo;
```

## 3. ActionListener

Contiene 3 ActionListener:

class SubidaBd implements ActionListener -->Nos permitirá escoger los archivos de base de datos

class SubidaArchivo implements ActionListener -->Nos permitirá escoger el archivo a analizar

class Ejecutar implements ActionListener -->Este ejecutara el análisis y nos mostrará en el textArea el resultado..



Main.java

## 1. Definición

La clase main es la que ejecuta la interfaz. Así con esta clase se inicia nuestro proyecto.

```
public static void main(String[] args) throws IOException, InterruptedException {  
    SwingUtilities.invokeLater(new Runnable() {  
        @Override  
        //Se llama a la interface  
        public void run() {  
            GUI gui = new GUI();  
            gui.Display();  
        }  
    });  
}
```

# Funcionalidad



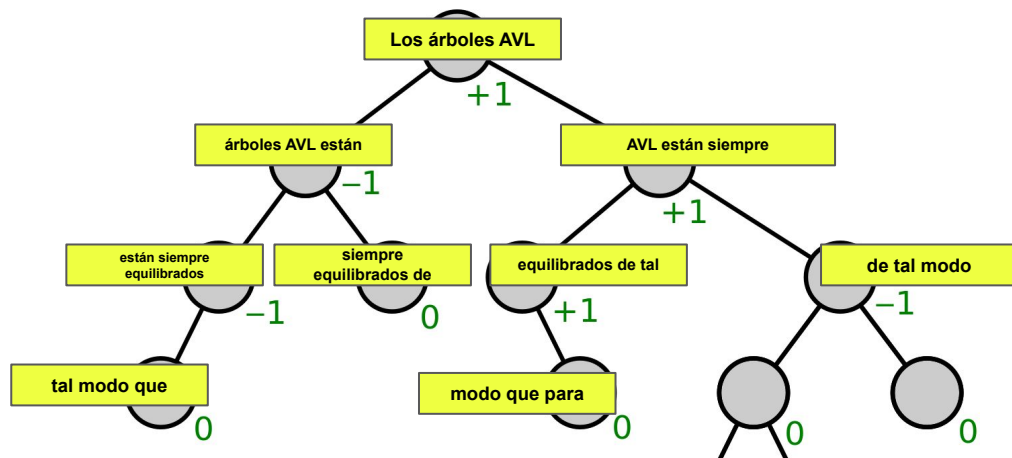
# Para la base

## Clase Phrase – addword(String word)

reut2-004.sgm

Los árboles AVL están siempre equilibrados de tal modo que para todos los nodos, la altura de la rama izquierda no difiere en más de una unidad de la altura de la rama derecha o viceversa. Gracias a esta forma de equilibrio (o balanceo), la complejidad de una búsqueda en uno de estos árboles se mantiene siempre en orden de complejidad  $O(\log n)$ . El factor de equilibrio puede ser almacenado directamente en cada nodo o ser computado a partir de las alturas de los subárboles.

Si el número de palabras por frases es 3 nuestro árbol se llenaría de la siguiente manera

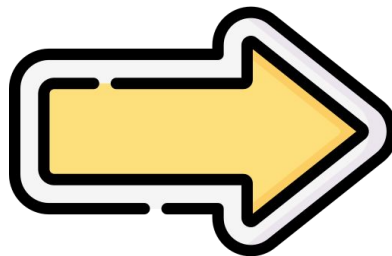
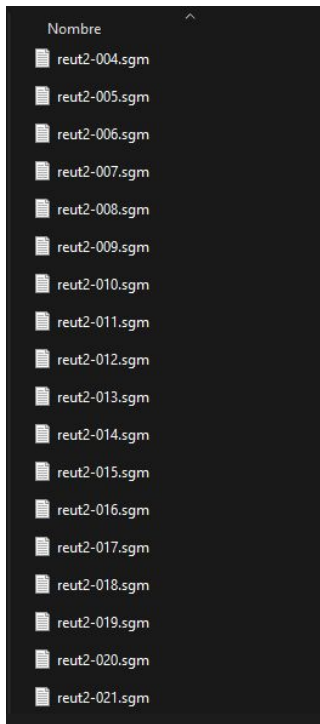




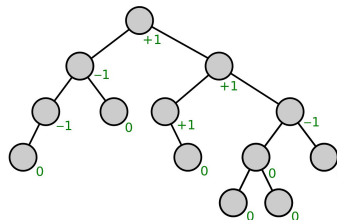
# Para la base

## Clase Document – createAVL()

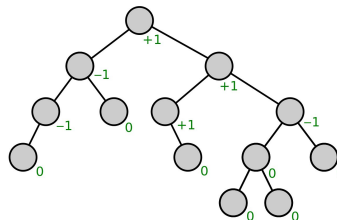
### Archivos .sgm



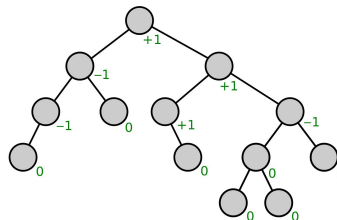
reut2-004.sgm



reut2-005.sgm



reut2-006.sgm





# Para los archivos a verificar si son plagio

## Clase Phrase – addword(String word)

### Mi archivo

Por ser un Árbol Binario de Búsqueda respeta la propiedad de orden de tal modo en todos sus nodos, es decir, todas las claves en su subárbol izquierdo son menores que la clave del nodo y todas las claves en el subárbol derecho son mayores.

Por ser un

ser un Árbol

un Árbol Binario

.....

de tal modo

derecho son mayores

Nuestro archivo en el que vamos a verificar si hay plagio usa el método “**addword**” para formar frases de 3 palabras las cuales se irán buscando una a una si hay frases coincidentes





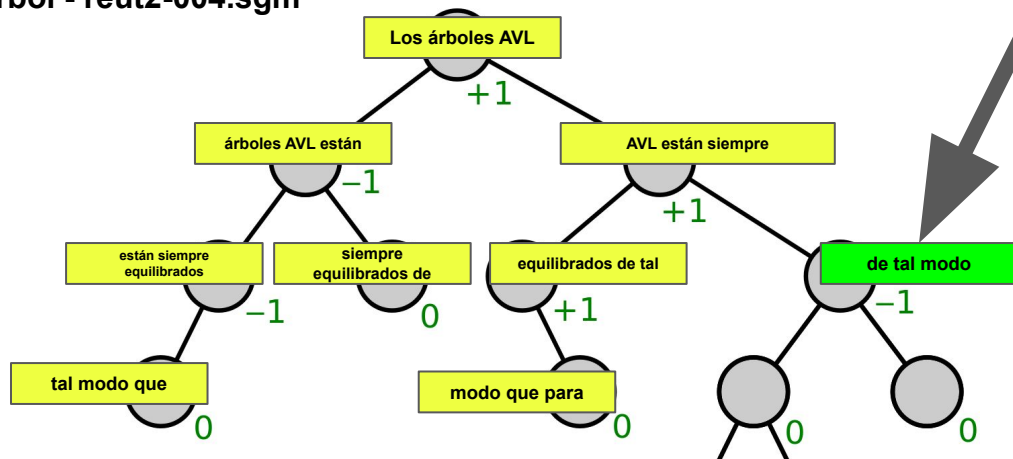
# Para comparar "Mi archivo.txt" con la base

## Clase Document – matching\_count(AVLTree<String> tree)

Mi archivo



Árbol - reut2-004.sgm



# Resultados De Experimentos

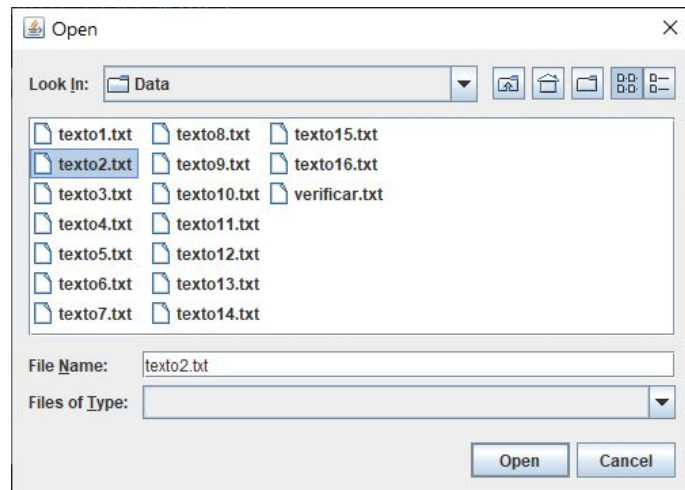




## Base de datos

Nombre	Fecha de modificaci...
texto1.txt	14/08/2022 10:31 ...
texto2.txt	15/08/2022 12:26 ...
texto3.txt	15/08/2022 12:51 ...
texto4.txt	14/08/2022 10:31 ...
texto5.txt	14/08/2022 10:31 ...
texto6.txt	15/08/2022 01:42 ...
texto7.txt	14/08/2022 10:31 ...
texto8.txt	14/08/2022 10:31 ...
texto9.txt	14/08/2022 10:31 ...
texto10.txt	14/08/2022 10:31 ...
texto11.txt	14/08/2022 10:31 ...
texto12.txt	15/08/2022 01:49 ...
texto13.txt	14/08/2022 10:31 ...
texto14.txt	14/08/2022 10:31 ...
texto15.txt	14/08/2022 10:31 ...
texto16.txt	14/08/2022 10:31 ...
verificar.txt	15/08/2022 01:56 ...

## Subida de archivos





# PRIMERA PRUEBA: PLAGIO DE TEXT02.TEXT

verificar.txt: Bloc de notas

Archivo Edición Formato Ver Ayuda

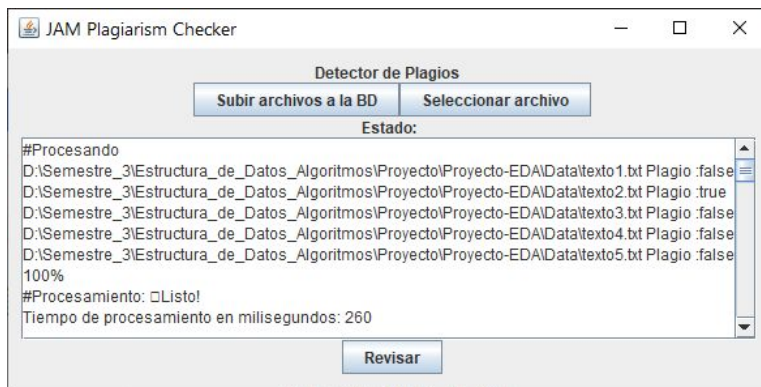
Bean shippers were reluctant to offer nearby shipment and only limited sales were booked for March shipment at 1,750 to 1,780 dlrs per tonne to ports to be named.

texto2.txt: Bloc de notas

Archivo Edición Formato Ver Ayuda

Bean shippers were reluctant to offer nearby shipment and only limited sales were booked for March shipment at 1,750 to 1,780 dlrs per tonne to ports to be named.

In English this semester we have studied three different texts. All three texts were based on original, fictional worlds. The fictional world which stood out above the rest and really amazed me would have to be 'Nineteen Eighty-Four'. 'Nineteen Eighty-Four' was the most realistic out of the three.





# SEGUNDA PRUEBA: PLAGIO DE TEXT03.TEXT Y TEXT04.TXT

verificar.txt: Bloc de notas

Archivo Edición Formato Ver Ayuda

lobal change has become a popular word in scientific debates on long-range structural change in the earth's ecology. Globalization has in the past played a major role in the controversial environmental debates. Many problems resulted in this area of discussion, in regard to the intricate linkages between globalization, government Juror Three is the last juror to change his vote; nothing anyone says can convince him that there is a "reasonable doubt" in the case. This man was the most stubborn of all. He refused to pay attention to things that were being stated in order save the defendant from death. It was as if the word guilty seemed to dwell in his mind and was unable to be altered. He becomes outraged that the others are changing their votes and letting this kid "slip through their fingers." He says that the whole case is based on the testimony of the woman across the el-tracks.

texto3.txt: Bloc de notas

Archivo Edición Formato Ver Ayuda

Global change has become a popular word in scientific debates on long-range structural change in the earth's ecology. Globalization has in the past played a major role in the controversial environmental debates. Many problems resulted in this area of discussion, in regard to the intricate linkages between globalization, government, trade and transport, and environmental decay. The current debate on the environmental effects of globalization is particularly concerned with the question whether a worldwide liberalization of trade may provoke environmental collapse. Three major environmental concerns related to trade are the domestic environmental effects caused by the use of imported products, the foreign environmental effects caused by the production of exported goods, and the environmental effects caused by transport movements needed for international trade.

texto4.txt: Bloc de notas

Archivo Edición Formato Ver Ayuda

Juror Three is the last juror to change his vote; nothing anyone says can convince him that there is a "reasonable doubt" in the case. This man was the most stubborn of all. He refused to pay attention to things that were being stated in order save the defendant from death. It was as if the word guilty seemed to dwell in his mind and was unable to be altered. He becomes outraged that the others are changing their votes and letting this kid "slip through their fingers." He says that the whole case is based on the testimony of the woman across the el-tracks. The jurors play out the murder to themselves, and talk about the lady across the street. They notice several things about her. The lady claimed that she saw the murder through the last two cars on the train that was passing. A juror also noticed that she had indents in her nose which means she wore glasses. She never mentioned the glasses in the trial. It was finally agreed with eleven jurors that there was "reasonable doubt" that the lady could not have successfully seen the murder without her glasses, and through a train. Everybody is angered and the votes gradually change to not guilty, some come





# TERCERA PRUEBA: ARCHIVO VACÍO

verificar.txt: Bloc de notas

Archivo Edición Formato Ver Ayuda



# CONCLUSIONES