

# CC3001 Tarea 2 Primavera 2018

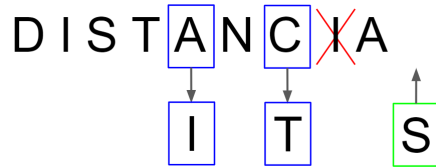
Profs. J  r  my Barbay / Patricio Poblete

Fecha de entrega: viernes 19 de octubre de 2018 a las 23:30

## Conceptos Generales

Una forma de medir la similaridad entre dos strings  $A$  y  $B$  es la llamada “distancia de edici  n (o de Levenshtein)”, definida como el n  mero m  nimo de operaciones de inserci  n, eliminaci  n o sustituci  n de caracteres para transformar un  $A$  en  $B$ .

*Ejemplo:* La distancia de edici  n entre el string “DISTANCIA” y el string “DISTINTAS” es 4, porque el primero se puede transformar en el segundo de la siguiente manera (y no hay forma de hacerlo con menos operaciones):



Supongamos que los strings son  $A = a_1a_2 \cdots a_m$  y  $B = b_1b_2 \cdots b_n$ , y definamos  $d(i, j)$  como el n  mero m  nimo de operaciones para transformar  $a_1a_2 \cdots a_i$  en  $b_1b_2 \cdots b_j$ . En t  rminos de esta funci  n, la distancia de edici  n entre  $A$  y  $B$  es igual a  $d(m, n)$ .

La funci  n  $d(i, j)$  se puede calcular recursivamente de la siguiente manera:

$$\begin{aligned} d(i, 0) &= i \\ d(0, j) &= j \\ d(i, j) &= \min \begin{cases} d(i-1, j) + 1 \\ d(i, j-1) + 1 \\ d(i-1, j-1) + [a_i \neq b_j] \end{cases} \end{aligned}$$

donde la notaci  n  $[P]$  toma el valor 1 si  $P$  es verdadero, y toma el valor 0 si  $P$  es falso.

Usando programaci  n din  mica (tabulaci  n), el c  lculo de  $d(m, n)$  se puede hacer en tiempo del orden de  $mn$ . Junto con calcular la distancia, se puede encontrar cu  l es la secuencia de cambios necesarios para transformar un string en el otro. *Nota:* Es posible que haya m  s de una soluci  n   ptima.

Una aplicación importante de este algoritmo es comparar dos archivos de texto y calcular cuántas líneas deben insertarse, eliminarse o sustituirse para transformar un archivo en el otro. En este caso, para el algoritmo las líneas del archivo juegan el rol de “letras”, y el archivo completo juega el rol de “string”. En el sistema Unix, existe un comando llamado *diff* que implementa esta funcionalidad.

## La tarea

Esta tarea consiste en escribir un programa en Java que compare dos archivos de texto dados. Para ilustrar lo que debe hacer, considere los dos archivos de texto siguientes:

A.txt	B.txt
Qué linda en la rama	Qué linda en la rama
la fruta se ve	las frutas se ven
si lanzo una piedra	ojalá que no me atrapen
tendrá que caer	si lanzo una piedra

El archivo A.txt se puede transformar en el archivo B.txt de la siguiente manera:

	A.txt	B.txt
1	Qué linda en la rama	Qué linda en la rama
2 c	la fruta se ve	las frutas se ven
3 i		ojalá que no me atrapen
3	si lanzo una piedra	si lanzo una piedra
4 d	tendrá que caer	

donde los números de línea se refieren al archivo original, y las letras tienen los siguientes significados:

- c (change): cambiar una línea por otra
- i (insert): insertar una línea antes de la línea indicada
- d (delete): eliminar una línea

Su programa debe leer desde la entrada estándar los nombres de los dos archivos (de a uno por línea) y debe escribir en la salida estándar la lista de operaciones que hay que hacer para transformar el primer archivo en el segundo. Esta lista debe tener el formato

NÚMERO LÍNEA,OPERACIÓN,NUEVA LÍNEA

donde OPERACIÓN puede ser *c*, *i* o *d*, y la NUEVA LÍNEA se incluye solo para los cambios y las inserciones.

En el ejemplo indicado, la salida emitida por el programa debería ser:

```
2,c,las frutas se ven
3,i,ojalá que no me atrapen
4,d
```

## Condiciones

- El programa debe ser implementado en Java.
- No se acepta ningún otro lenguaje de programación.
- Realice un informe de acuerdo a la pauta indicada en las Normas del curso, localizada en sección Material Docente.
- NO SE ACEPTARÁN TAREAS ATRASADAS.
- La tarea es individual y se revisará que no hayan copias.
- Dudas sobre el enunciado pueden discutirlas en el foro del curso.
- La tarea debe entregarse a través de la página de U-Cursos, sección Tareas. NO SE ACEPTARÁ NINGUNA TAREA ENVIADA POR E-MAIL NI SUBIDA A LA SECCION MATERIAL ALUMNOS. Preocúpese de enviar su tarea a tiempo. POR FAVOR REVISEN DESPUÉS DE SUBIR SU TAREA QUE SUBIERON LOS ARCHIVOS CORRECTOS, QUE LOS SUBIERON A LA TAREA CORRECTA Y QUE SE SUBIERON CORRECTAMENTE AL SERVIDOR. Es responsabilidad de usted asegurarse que su tarea fue entregada en forma correcta.