

Trabajo Práctico No. 1

La resolución de este trabajo práctico debe ser enviada a través del moodle de la materia <http://dc.exa.unrc.edu.ar/moodle>, antes de las **23:55 del martes 22 de Mayo de 2018**.

El código provisto como parte de la solución a los ejercicios deberá estar documentado apropiadamente (por ejemplo, con comentarios en el código). Aquellas soluciones que no requieran programación, como así también la documentación adicional de código que se desee proveer, debe entregarse en **archivos de texto convencionales o archivos en formato PDF** unicamente, con nombres que permitan identificar facilmente su contenido.

Tanto la calidad de la solución, como el código y su documentación serán considerados en la calificación. Recuerde además que los trabajos prácticos **no tienen recuperación**, y que el trabajo en la resolución de los ejercicios debe realizarse en grupos de 3 personas.

Ej. 1. En Biología resulta natural representar información a través de *secuencias*, como en el caso de las secuencias de ADN o ARN. En este contexto, en general resulta sumamente útil encontrar similitudes entre cadenas, y suelen aparecer problemas de cómputo de distancia particulares entre cadenas, intentar detectar patrones en las mismas, o determinar si una secuencia es subsecuencia de otra. Un problema particularmente relevante en Biología es el *alineamiento de secuencias*, que posee numerosas aplicaciones, incluyendo las que permiten determinar relaciones de parentesco entre secuencias. El problema de *alineamiento óptimo de secuencias* consiste en lo siguiente. Dado un alfabeto Σ , un símbolo destacado $_ \notin \Sigma$, y dos secuencias $s, t \in \Sigma^*$, un *alineamiento* de s y t es un par de cadenas $s', t' \in (\Sigma \cup \{ _ \})^*$ tales que:

- al eliminar los caracteres $_$ de s' , se obtiene s ,
- al eliminar los caracteres $_$ de t' , se obtiene t ,
- s' y t' tienen la misma longitud l , y $l \leq m + n$, con m (resp. n) la longitud de s (resp. t),
- no existe ninguna posición i tal que $s'[i] = t'[i] = _$.

Un alineamiento $\alpha(s, t) = \{s', t'\}$ se denomina *óptimo* respecto a una función:

$$D : (\Sigma \cup \{ _ \})^* \times (\Sigma \cup \{ _ \})^* \rightarrow \mathcal{R}$$

si maximiza (o minimiza, de acuerdo al caso) dicha función, es decir, si entre todos los alineamientos posibles es aquel que optimiza D .

El objetivo de este ejercicio es desarrollar una aplicación que, dado un alfabeto Σ , un símbolo destacado $_ \notin \Sigma$, una función de costos D y dos cadenas s y t , encuentre un alineamiento óptimo para s y t , respecto a D . La solución debe ser *genérica*, en el sentido que D es un parámetro de la solución, es decir debe poder reusarse con mínimo esfuerzo si se desea utilizar cualquier función de costos. A modo de simplificación, podemos considerar sólo funciones de costo cuyo objetivo de optimización sea la minimización (es decir, a menor valor, mejor es el alineamiento). Como caso de prueba puede utilizarse como función de costos D a la distancia de edición entre los dos alineamientos.

Si bien este ejercicio no incluye ninguna restricción en relación al tiempo de ejecución de la solución, se espera que la misma sea eficiente, y la eficiencia en la solución será considerado un atributo de calidad relevante del ejercicio.