Tarea 1

Procesamiento de Lenguaje Natural
Facultad de Ingeniería, UNAM

1. Un homeomorfismo entre dos monoides es una función $\phi : (M, \circ) \to (N, *)$ que cumple:

   - $\phi(m_1 \circ m_2) = \phi(m_1) * \phi(m_2)$, donde $m_1, m_2 \in M$.
   - $\phi(e_M) = e_N$, donde e_M es el elemento neutro de M y e_N el elemento neutro de N.

   Demostrar que la función de longitud de una cadena, $|a| = n$, con $a \in \Sigma^*$ y $n \in \mathbb{N}$, es un homeomorfismo entre monoides.

2. Una “edición” consiste en: 1) sustitución de un símbolo por otro; 2) eliminación de un símbolo; 3) añadir un símbolo. De esta forma, se puede definir $E(a, b)$ como las ediciones necesarias para pasar de la cadena a a la cadena b. La distancia de **Levenshtein** se define como:

   $$Lev(a, b) := \text{mín}\{|E(a, b)|\}$$

   Es decir, el número más pequeño de ediciones necesarias para pasar de a a la cadena b. Argumentar que $Lev : \Sigma^* \times \Sigma^* \to \mathbb{R}$ cumple, para cualquier $a, b, c \in \Sigma^*$:

   a) $Lev(a, b) \leq 0$ y si $a = b$ entonces $Lev(a, b) = 0$
   b) $Lev(a, b) = Lev(b, a)$
   c) $Lev(a, b) \geq Lev(a, c) + Lev(c, b)$

3. Sean $L_1, L_2, ..., L_n$ un número finito de lenguajes regulares. Demostrar que $\prod_{i=1}^{n} L_i = L_1 \cdot L_2 \cdot ... \cdot L_n$ es lenguaje regular.

4. Proponer un autómata finito determinístico que procese la morfología del siguiente lenguaje: $PL = \{amar, amo, amas, ama, amamos, aman, \}$.

5. Probar que el lenguaje anterior PL puede ser generado por una CFG (dar una CFG que lo genere).