**Algoritmo de Huffman:**

El algoritmo de codificación/compresión Huffman fue propuesto en 1952 por el profesor  David A. Huffman como una forma sencilla y óptima de representar cada símbolo del alfabeto con un código de longitud óptima. Para conseguir una asignación óptima, los símbolos son representados mediante códigos cuya longitud es inversamente proporcional a la probabilidad (frecuencia) del símbolo.

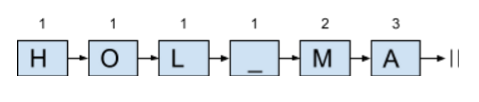
Suponiendo que queremos aplicar el algoritmo de Huffman a una cadena de texto, tomaremos como ejemplo la cadena de caracteres “HOLA MAMA”. En este ejemplo vamos a representar el espacio entre las palabras como un “\_”.

A continuación se presentan una serie de pasos para realizarlo:

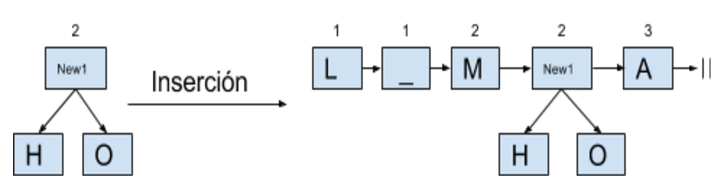
1. Formamos una tabla con la cantidad de apariciones de cada carácter en la cadena:

|  |  |
| --- | --- |
| **Carácter** | **Frecuencia** |
| H | 1 |
| O | 1 |
| L | 1 |
| A | 3 |
| \_ | 1 |
| M | 2 |

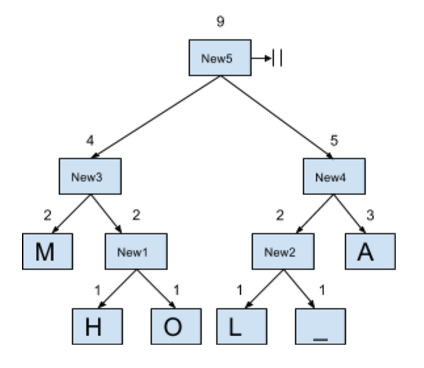
1. Con esta tabla formamos una lista ordenada de árboles binarios, donde cada carácter formará la raíz de un nuevo árbol:



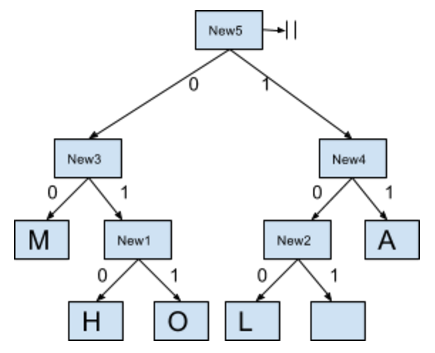
1. Una vez formada la lista agarramos los dos primeros árboles para formar un nuevo árbol, donde la frecuencia del nuevo árbol será la suma de las frecuencias de sus hijos y lo insertamos ordenadamente en la lista de árboles:



1. Aplicando el paso anterior unas cuantas veces más hasta que en la lista solo quede un árbol, se debería llegar a:



1. Cuando esté listo el árbol, se etiquetan las ramas del árbol con 0 las ramas izquierdas y 1 las ramas derechas, esto con el fin de formar un camino a cada hoja del árbol (caracteres de la cadena).



1. Formamos una tabla con todas las rutas del árbol hacia cada carácter.

|  |  |  |
| --- | --- | --- |
| **Carácter** | **Frecuencia** | **Representación** |
| H | 1 | 010 |
| O | 1 | 011 |
| L | 1 | 100 |
| A | 3 | 11 |
| \_ | 1 | 101 |
| M | 2 | 00 |

1. Finalmente nuestra cadena queda codificada como:

010 011 100 11 00 11 00 11

Nota: Para luego decodificar dicha cadena, es necesario disponer del árbol a partir del cual se codificaron los datos, esto se hace comunicando al decodificador la tabla de representaciones de cada símbolo, con ellas se es capaz de reconstruir el árbol original y una vez se tiene el árbol, para el decodificador es sumamente fácil reconstruir la cadena original, esto recorriendo el árbol según la cadena de bits codificada.