

Creación del árbol del documento

Información del documento

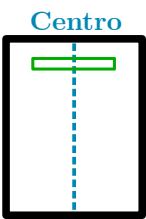
- * Texto
- * Posiciones y coordenadas de cada línea y columna

Asignar tipo de título

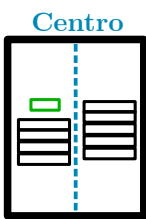
Para cada bloque separado verticalmente:

Si es una sola columna y el texto está centrado en la página, es título tipo 1.

Si son más de una columna y el texto está centrado en la columna, es título tipo 2.



Tipo 1



Tipo 2

Asignar nivel de título

Para cada título de tipo 1:

Comparar el bloque contra cada expresión regular registrada, si hay coincidencia se le asigna el nivel correspondiente.

Si no hay coincidencia se asigna nivel 1 que indica que es un título principal.



Nivel 1



Nivel 2



Nivel 3

Recorrer el archivo (armado del árbol)

Se parte de un nodo raíz o nivel 0.

Cuando se encuentra un título nivel 1, se crea un nuevo nodo. Si se definieron más niveles, se repite el proceso.

Al terminar los niveles de los títulos se analiza si la línea cumple con la expresión regular, de ser así se crea un nuevo nodo.

Todo el texto que no sea título o parte del identificador del nodo se guarda como contenido del nodo en el que se encuentre.

