

¿Qué sabes de programación?

Escritura T9 – Entrega uno de tres

¿Te has percatado de que en tu teléfono celular existe una función que te permite escribir palabras de una manera realmente rápida, sin la necesidad de pulsar varias veces cada uno de los números para que obtengas la letra correspondiente?

Déjame explicártelo más detalladamente.

Tu teléfono tiene asignadas varias letras a cada una de las teclas numéricas, como se muestra en la tabla 1:

| Número | Letras | Número | Letras |
|--------|----------|--------|-------------|
| 1 | Ninguna | 6 | M, N y O |
| 2 | A, B y C | 7 | P, Q, R y S |
| 3 | D, E y F | 8 | T, U y V |
| 4 | G, H y I | 9 | W, X, Y y Z |
| 5 | J, K y L | 0 | Ñ |

Tabla 1: Letras asignadas a los números de un teléfono celular

Si deseas escribir la palabra **HOLA**, lo que normalmente tendrías que pulsar sería **4 4 6 6 6 5 5 5 2**, con lo cual irías obteniendo **G, H, HM, HN, HO, HOJ, HOK, HOL, HOLA**. Pero, por otro lado, los teléfonos tienen una forma de escritura que es conocida como T9, en la cual solamente tienes que pulsar las primeras teclas de la palabra que deseas, y al final vas obteniendo una serie de palabras válidas en el idioma configurado (estoy suponiendo que es el español), así, si escribes **4652** obtienes **HOLA** (además de alguna otra opción tal como **HOJA**).

¿Te has preguntado cómo es que esto funciona?

Conforme vas escribiendo números, el sistema de búsqueda va determinando las posibles combinaciones válidas de las letras que están representadas en dichos pulsos, como es posible que aún no hayas escrito todas las letras de la palabra que necesitas, el sistema debe de conservar todas las posibles combinaciones válidas de letras, además de determinar las palabras completas que se han identificado hasta el momento.

Siguiendo el mismo ejemplo de hace un momento, si has pulsado **465**, puedes obtener la palabra **GOL**, pero el sistema debe ser capaz de conservar **GOL, HOL, HOJ** como posibilidades de palabras válidas.

Una de las formas en las que puedes lograr esto es teniendo un diccionario de las palabras válidas del lenguaje en la memoria. Tomando como base este esquema, lo que puedes hacer para determinar si existe una palabra válida es tomar todas las combinaciones posibles de las letras de los números, y buscar si existe una palabra así en dicho diccionario. Se puede ver un esquema en la tabla 2 para la combinación **465**. Como puede apreciarse, la única palabra válida hasta ahora es **GOL**, lo cual no te indica nada acerca de la palabra que estás intentando escribir (**HOLA**).

Una vez que escribes el siguiente dígito este esquema cambia completamente, tal y como se muestra en la vista simplificada de la tabla 3.

Como puedes apreciar, ahora **GOLA, GOLB** y **GOLC** no son palabras válidas dentro del diccionario, pero se han añadido dos nuevas palabras que sí lo son: **HOJA** y **HOLA**, las cuales representan las posibles combinaciones de las letras que has pulsado.

| Combinación | Válida | Combinación | Válida | Combinación | Válida |
|--------------------|---------------|--------------------|---------------|--------------------|---------------|
| GMJ | No | GMK | No | GML | No |
| GNJ | No | GNK | No | GNL | No |
| GOJ | No | GOK | No | GOL | Sí |
| HMJ | No | HMK | No | HML | No |
| HNJ | No | HNK | No | HNL | No |
| HOJ | No | HOK | No | HOL | No |
| IMJ | No | IMK | No | IML | No |
| INJ | No | INK | No | INL | No |
| IOJ | No | IOK | No | IOL | No |

Tabla 2: Combinaciones válidas para la pulsación de teclas **465**

| Combinación | Válida | Combinación | Válida | Combinación | Válida |
|--------------------|---------------|--------------------|---------------|--------------------|---------------|
| GMJA | No | GMJB | No | GMJC | No |
| ... | ... | ... | ... | ... | ... |
| GOLA | No | GOLB | No | GOLC | No |
| ... | ... | ... | ... | ... | ... |
| HOJA | Sí | HOJB | No | HOJC | No |
| HOKA | No | HOKB | No | HOKC | No |
| HOLA | Sí | HOLB | No | HOLC | No |
| ... | ... | ... | ... | ... | ... |
| IOJC | No | IOKC | No | IOLC | No |

Tabla 3: Combinaciones válidas para la pulsación de teclas **4652**

Para que este método pueda funcionar es necesario que tengas todo el diccionario de palabras en español e ir, con cada nueva pulsación de tecla, buscando en éste las palabras que coincidan con dichas combinaciones (la cantidad de combinaciones aumenta exponencialmente con cada nueva pulsación de tecla, es decir, para las teclas **465** se tienen $3 * 3 * 3 = 27$ combinaciones, pero para la combinación **4652** se tienen $3 * 3 * 3 * 3 = 81$ combinaciones) lo cual puede ser demasiado costoso en cuanto a las búsquedas sobre el diccionario (sobre todo si éste tiene muchas palabras y aunque se tenga almacenado en forma de árbol).

Un mejor método de identificar las palabras que son válidas para una combinación de teclas dada es la de almacenar el diccionario en una forma que te facilite la búsqueda, y que a la vez te permita ir estableciendo las combinaciones que van siendo válidas hacia nuevas palabras del mismo diccionario. La forma en la que lo puedes almacenar y que resulta bastante conveniente para el problema en cuestión es en un árbol en el que cada nivel contiene sólo las letras siguientes de las posibles palabras del diccionario, es decir, si la palabra comienza con la letra **a**, le pueden seguir de la letra **a** a la letra **z** sin ningún problema, pero si la palabra comienza con la letra **q**, entonces no le puede seguir la letra **k**, ya que no hay palabra alguna en español que comience con la combinación de letras **qk**, lo cual limita bastante el espacio de búsqueda sobre el diccionario.

Una vez que has tomado en cuenta este hecho, el pulsar las teclas **465** te conduce solamente a tres posibilidades dentro del árbol que representa el diccionario, y éstas son **GOL**, **HOJ** y **HOL**, de las cuales sólo la primera representa una palabra completa dentro del mismo. Dentro de la estructura de datos

que te representará el diccionario, se tienen que tomar medidas en cuanto a si ya se completó una palabra o solamente es la parte inicial de alguna de ellas.

Por otro lado, hay que considerar el espacio de almacenamiento que se tiene en un teléfono celular, ya que éste no es tan amplio como el de una PC. Si el árbol que representa el diccionario se tuviera que crear cada vez que se realiza una búsqueda en el mismo, se estaría desperdiciando un espacio valioso dentro de la memoria del dispositivo. Por ello es conveniente que crees el árbol a priori y que sea éste el que se almacene en la memoria, sin necesidad de procesar el diccionario, esto lleva a un ahorro de memoria y a que siempre esté disponible para futuras consultas.

En la segunda entrega de este artículo podrás encontrar la descripción de las estructuras de datos con las cuales se construye el árbol, la estructura de datos con la cual se almacena y los algoritmos que nos permiten su creación y su consulta. Y en la tercera, y última entrega, encontrarás el código de la solución completa correctamente documentado y explicado.

Por el momento no me queda más que desear que esta primera entrega haya sido de tu agrado y recordarte que se ha abierto el curso de preparación para el concurso ICPC del capítulo estudiantil del ACM, edición 2005.