

DICCIONARIO EN CASTELLANO PARA T_EX

Juan Luis Varona

Resumen

Se presenta un diccionario electrónico en castellano destinado a ser usado, fundamentalmente, por verificadores ortográficos de T_EX.

El diccionario es una actualización del que estaba disponible hasta ahora y que ya se emplea, por ejemplo, en Excalibur para Macintosh, en WinEdt para Windows, y en amSpell para MS-DOS. Consiste en una recopilación de casi 300 mil palabras.

1. Verificadores ortográficos

Cuando escribimos un texto, sea del tipo que sea, es bastante común que, por mucho cuidado que pongamos, cometamos alguna errata. No sólo es posible que nos equivoquemos al recordar la ortografía de una palabra, sino que nuestra impericia escribiendo nos lleva a menudo a cambiar unas letras por otras al teclear. Para intentar minimizar esto en lo posible, unos programas llamados verificadores ortográficos vienen en nuestra ayuda.

Un verificador ortográfico comprueba que cada palabra de un texto es una de las palabras posibles en el idioma. Obviamente, un verificador nunca descubrirá un error cuando, al escribir mal una palabra, hemos escrito otra también existente, incluso aunque no tenga ningún sentido gramatical en la frase en que aparece.

En lo que a nosotros —usuarios de T_EX— respecta, es importante tener en cuenta que un verificador ortográfico que quiera ser útil para comprobar documentos escritos en T_EX debe comprender, al menos mínimamente, la sintaxis del lenguaje: de lo contrario, marcará como errores los comandos de T_EX, las fórmulas, etc. Así pues, deberemos utilizar un verificador ortográfico especializado en T_EX.

Por otra parte, el diccionario de palabras admisibles puede estar construido de dos maneras, lo que da lugar a dos tipos de verificadores ortográficos. La primera consiste simplemente en una enumeración de todas las palabras

posibles. Así, cada palabra debe aparecer con todas sus variaciones permitidas al añadirle distintos prefijos y sufijos. Por ejemplo, los sustantivos y adjetivos deben aparecer en masculino, femenino, singular y plural; y los verbos conjugados.

La segunda no es una simple lista de palabras con todas sus derivaciones posibles. Aquí, podíamos pensar que en el diccionario tenemos sólo la raíz de la palabra y, además, unas reglas de derivación. Por ejemplo, para cada sustantivo deben aparecer las reglas de cómo formar masculinos, femeninos y plurales; y cada verbo debe ir asociado con sus reglas de conjugación.

Programas del primer tipo son Excalibur para Macintosh, amSpell para PC con MS-DOS y WinEdt para Windows. Del segundo tipo, podemos destacar ispell, que está disponible para equipos unix.

En general, los del primer tipo son más sencillos, pero requieren mayores ficheros de palabras. Los del segundo tipo requieren un trabajo de diseño y programación más importante, pero pueden llegar a ser más efectivos. Por ejemplo, para incluir aumentativos y diminutivos sólo hace falta añadir unas pocas reglas. En uno de los del primer tipo requiere incluir en la lista cada palabra en aumentativo y diminutivo.

Por otra parte, unos idiomas se adaptan mejor que otros a utilizar verificadores de uno u otro tipo. Por ejemplo, en inglés no hay femeninos ni plurales de adjetivos, y los verbos se conjugan muy poco. Así, no merece mucho la pena el esfuerzo de programación que se necesita para utilizar verificadores del segundo tipo. El español, en cambio, tiene gran cantidad de formas verbales que, en un verificador del primer tipo, es necesario incluir individualmente. Otros idiomas, como el alemán y el vasco, aun tienen más mecanismos de formación de palabras, como la declinación y la yuxtaposición (que puede llegar a ser bastante común), lo que hace que el número de palabras permitidas crezca desmesuradamente. Para estos idiomas, una estrategia del segundo tipo parece ser bastante más recomendable. El castellano podría ser un caso intermedio.

2. La lista de palabras en castellano

En lo que a nosotros respecta, nos vamos a ocupar aquí únicamente de verificadores ortográficos del primer tipo, de los basados en una lista de palabras. Desde hace tiempo, existen diversos programas que pueden ser usados para comprobar la corrección ortográfica de textos escritos en \TeX , unos gratis y otros no excesivamente caros. Pero no había ninguna lista de palabras en español suficientemente completa y de distribución gratuita. Obviamente, sí que existían implementaciones en programas comerciales pero que, en la

práctica, no servían para nada si lo que queríamos era comprobar un documento \TeX . Así pues, hace ya años, el que escribe se planteó elaborar una lista lo bastante amplia como para que pudiera ser usada con comodidad.

Ahora se presenta una actualización de la lista, que contiene más de 40.000 palabras nuevas. Concretamente, la lista actual consta de 292.342 palabras, mientras que la anterior tenía 247.049. Es importante hacer notar que incluye multitud de palabras relacionadas con la matemática, física e informática, que son las disciplinas donde más se utiliza \TeX . Muchas de ellas no suelen estar presentes en diccionarios de carácter general.

Es de destacar que el usuario tiene acceso al diccionario fuente y puede modificarlo a su antojo para su uso personal. Aunque sería deseable que la versión que se distribuye en general permanezca unificada. Así, las futuras mejoras y añadidos de palabras se hacen siempre sobre la misma lista. El mecanismo para indicar la versión de la lista es simple: basta indicar el número de palabras que contiene.

3. Implementaciones para diversos sistemas operativos

La lista de palabras que aquí se presenta puede usarse con los siguientes verificadores ortográficos:

3.1. Excalibur

Es un programa para Macintosh, desarrollado por Robert Gottshall y Rick Zaccone. Es gratuito. Su página web es <http://www.eg.bucknell.edu/~excalibr/excalibur.html>, y puede conseguirse también, mediante ftp anónimo, en <ftp://ftp.eg.bucknell.edu/pub/mac/> (en el subdirectorio `Excalibur-dictionaries/` se pueden encontrar diccionarios en multitud de idiomas). En CTAN, se encuentra en el directorio `tex-archive/systems/mac/support/excalibur/`.

Una característica de Excalibur resulta muy útil a los escritores en \TeX . Una palabra que contenga letras no inglesas (acentos o eñes, en el caso del español) puede escribirse con notación de 7 bits o de 8 bits. No es este lugar para comentar las ventajas y desventajas que tiene uno y otro método, puesto que cualquier usuario medianamente avanzado las conoce. Con Excalibur, basta tener la lista de palabras escrita con notación de 8 bits. Así, si una palabra de un texto está escrita con notación de 7 bits, Excalibur se encarga de traducir los 7 bits a 8 bits (en cualquier combinación entendible por \TeX) antes de comparar con las palabras de la lista.

3.2. WinEdt

Es un potente y versátil editor destinado a la creación de documentos en T_EX que funciona en Windows. Entre otras cosas, tiene un verificador ortográfico. Su autor es Aleksander Simonic y es un programa *shareware*. El precio de una licencia personal para uso educativo es de 40 dólares. Su página web es <http://home.istar.ca/~winedt> o <http://www.winedt.com>. En CTAN, puede encontrarse en el directorio `tex-archive/systems/win32/winedt/`. Asimismo, en el subdirectorio `dict/` se pueden encontrar diccionarios en bastantes idiomas.

3.3. amSpell

Es un verificador ortográfico para documentos T_EX que funciona bajo MS-DOS. Su autor es A. Merckens y es gratuito. Puede descargarse, en CTAN, en el directorio `tex-archive/support/amsPELL/`, aunque allí sólo aparece el diccionario inglés. El diccionario en español, que es una adaptación del que aquí se presenta realizada por Agustín Martín Domingo, puede conseguirse por ftp anónimo en <ftp://tex.unirioja.es/pub/tex/dict-pc/>.

Para finalizar, únicamente añadir que todos estos programas, y las últimas versiones del diccionario español adaptado a ellos, pueden obtenerse en la Universidad de La Rioja (de donde procede el que escribe esta nota), mediante ftp anónimo, en <ftp://tex.unirioja.es/pub/tex/>. Una vez ahí, los directorios `dict-mac/`, `dict-win/` y `dict-pc/` contienen, respectivamente, los diccionarios para Macintosh (Excalibur), Windows (WinEdt) y MS-DOS (amSpell) aquí descritos.

Agradecimientos: Gracias a todos los que han colaborado notificándome erratas de la versión anterior de la lista de palabras, o enviándome nuevas palabras para incluir. En particular, a Héctor M. Monacci y a Josu Zabaleta.

Juan Luis Varona,
Departamento de Matemáticas y Computación, Universidad de La Rioja,
C/ Luis de Ulloa s/n, 26004 Logroño, ESPAÑA.
e-mail: jvarona@dmc.unirioja.es
URL: <http://www.unirioja.es/dptos/dmc/jvarona/welcome.html>