

REQUISITOS PARA LA EDICIÓN CRÍTICA INFORMATIZADA: UNITE. CONJUNTO DE PROGRAMAS PARA LA UNIFICACIÓN AUTOMÁTICA DE TEXTOS. VERSIÓN PARA MICROORDENADORES SUN

FRANCISCO MARCOS MARÍN y JUAN DE DIOS GODOY GARCÍA
Universidad Autónoma de Madrid

1. INTRODUCCIÓN

UNITE es un conjunto de programas que comparan diferentes versiones de un mismo texto en verso (es decir, entendido como *línea + línea*) con el objetivo de obtener una versión unificada a partir de las comparadas. UNITE no está planteado como la solución definitiva a los problemas de la crítica textual (extensión de los textos, diversidad de grafías, existencia de diferentes versiones, etc.), sino como una herramienta que libere al editor humano de labores rutinarias y centre su trabajo en aquellas fases del proceso en las que sea indispensable la actuación del experto.

La primera de las características de UNITE es que los textos no llevan ningún tipo de marca, etiqueta o membrete incorporado por el editor al texto que copia: no se requiere ninguna labor de pre-edición.

La segunda es que presenta un amplio abanico de posibilidades de automatización, con lo que el usuario puede delimitar perfectamente el campo de actuación del programa.

El paquete estándar está diseñado para comparar un máximo de seis versiones de un mismo texto, diseño que puede ampliarse si se van a comparar más de seis textos simultáneamente. La unidad de comparación es la estrofa, entendida simplemente como un conjunto de versos separados por blanco. Este tipo de unidad permite detectar y solucionar el problema de los versos descolocados dentro de una estrofa. No es necesario que las versiones tengan el mismo número de estrofas ni que éstas estén ordenadas por su numeración, ya que se incluyen utilidades para dar formato y ordenar los textos. Tampoco es necesario que

las estrofas tengan el mismo número de versos, siempre que este número esté comprendido entre 1 y 5 ($> 1 < 5$).

Los únicos requerimientos que se deben cumplir en el formato de los textos son:

- Un número de líneas al principio del fichero para identificar el texto, que debe ser igual en todas las versiones que se van a comparar.
- El texto tiene que estar dividido en estrofas, separadas entre ellas por una línea en blanco. Cada estrofa lleva un número, que debe aparecer delante del primer verso de la estrofa y en la misma línea que éste.
- Aunque no es necesario que el texto esté ordenado, ya que se puede hacer de forma automática, si es obligatorio que todas las versiones estén ordenadas con el mismo formato (opción 2 del menú principal). Tenemos que advertir que las líneas al principio de los ficheros, que identifican los textos y que no serán tenidas en cuenta por los programas, no serán modificadas por las utilidades de formateo. Por ello es responsabilidad del usuario el asegurar que el número de dichas líneas sea igual en todas las versiones.

El conjunto de programas está implementado en lenguaje PASCAL y ha sido desarrollado en los Centros Científicos de IBM en Madrid (España) y Heidelberg (Alemania), para ordenadores IBM 43xx y superiores. Posteriormente ha sido modificado para adaptarlo al entorno UNIX del ordenador SUN y se ha utilizado un modelo 3/60.

Para terminar con esta introducción conviene informar de que en la actualidad se está desarrollando una nueva versión de UNITE, en la que se intenta eliminar cualquier tipo de limitación, tanto en el formato de los textos como en el máximo de versiones y de versos por estrofa. Está previsto, además, que esta última versión de UNITE sea ejecutable en ordenadores personales.

2. ENTORNO DE MENÚS

UNITE se presenta al usuario mediante un sistema de menús en el que aparecen todas las utilidades que componen el paquete de programas. El sistema de menús está compuesto por un menú principal de que cuelgan una serie de menús secundarios. A continuación se describen uno y otros con las utilidades a que dan acceso.¹

1. Cuando reproducimos un menú en el texto lo hacemos respetando la apariencia que tiene en la pantalla: para evitar problemas de compatibilidades de teclado y códigos no hemos usado en ellos ni acentos ni signos especiales.

Unificación textual

(Menú principal)

1. Menú de operaciones con ficheros.
 2. Preparar un texto original.
 3. Identificar los ficheros que contienen los textos.
 4. Ejecutar el proceso de UNIFICACIÓN en modo interactivo.
 5. Ejecutar el proceso de UNIFICACIÓN en modo batch.
 6. Salida temporal al UNIX.

 9. Salida.
- Elegir opción:

1. *Operaciones con ficheros.*

Esta opción permite acceder al menú de operaciones con ficheros, que está compuesto de los siguientes menús y utilidades:

1. OPERACIONES CON FICHEROS

1. Edición de ficheros.
 2. Sacar copias de ficheros.
 3. Formatear la Versión Unificada.
 4. Extraer Fichero de Variantes.
 5. Trabajos con los textos: localización de palabras, concordancias, etc.

 9. Menú principal.
- Elegir opción:

1.1. Edición de ficheros.

Esta opción permite acceder al menú de edición de ficheros, que está compuesto de las siguientes utilidades:

1.1 EDICIÓN DE FICHEROS

1. Fichero de palabras VACIAS y SEMIVACIAS.
 2. Fichero de SIGNOS de la grafía.
 3. Fichero de CARACTERES QUE DERIVAN HACIA OTROS.
 4. Fichero de PARÁMETROS de ejecución por defecto.
 5. Fichero que almacena la VERSIÓN UNIFICADA.
 6. Fichero resumen de la ejecución.
 7. Otros.

 9. Menú anterior.
- Elegir opción:

1.1 Edición del fichero que contiene las palabras «vacías» y «semivacías».

Se conocen con estos nombres aquellas palabras sin contenido semántico o con un contenido semántico muy pobre. Dichas palabras se utilizan para determinar si dos versos están descolocados dentro de una estrofa, aunque coincidan en un número limitado de ellas.

1.2 Edición del fichero que contiene los signos de la grafía que se eliminarán de los textos cuando se unifique dicha grafía.

1.3 Edición del fichero que contiene los caracteres y cadenas de caracteres que derivan hacia otros en las fases de unificación de grafía.

1.4 Edición del fichero que contiene los parámetros para la ejecución del programa.

Mediante la modificación de estos valores el usuario puede controlar el grado de automatización del proceso de edición unificada. Los valores que contiene este fichero son los que toma por defecto cuando no se utiliza la opción de modificarlos interactivamente desde la pantalla.

1.5 Edición del fichero que contiene la versión unificada resultante de la ejecución del proceso de unificación.

1.6 Edición del fichero resumen de la ejecución de los procesos de unificación. Este fichero describe cómo se generaron palabras de la versión unificada como resultado del tratamiento informático e indica para cada una de ellas el proceso de unificación que la generó, la estrofa y verso en el que aparece, y las palabras originales a partir de las cuales se formó. Es el fichero auxiliar para la elaboración posterior del aparato crítico.

1.7 Edición de cualquier otro fichero no considerado en las opciones anteriores.

Es posible ver el contenido de los ficheros mediante una llamada al editor VI de UNIX. Todos ellos pueden ser modificados a conveniencia del usuario siempre que se respeten las indicaciones que aparecen al principio de cada fichero.

1.2 Copiar ficheros

Esta opción permite hacer copias de seguridad y para ello accede al menú que simplifica la copia de ficheros, el cual está compuesto de las siguientes utilidades:

1.2 COPIAR FICHEROS

1. Fichero de palabras VACIAS y SEMIVACIAS (fich_vacias, scr).
 2. Fichero de SIGNOS de la grafía (fich_signos, scr).
 3. Fichero de CARACTERES QUE DERIVAN HACIA OTROS (fich_carugr.scr).
 4. Fichero de PARÁMETROS de ejecución por defecto (valdefec.dat).
 5. Fichero que almacena la VERSIÓN UNIFICADA (berunif.txt).
 6. Fichero resumen de la ejecución (faclara.res).
 7. Otros.

 9. Menú anterior.
- Elegir opción:
Haga copia de seguridad del fichero antes de ejecutar los programas.
Nombre del fichero (<ENTER> para salir):

1.3 Formatear la versión unificada

Esta opción opera una vez realizada la unificación y permite preparar la versión unificada, bien con el texto unificado sólo, bien con las variantes no unificadas detrás de cada línea unificada, para conseguir un fichero manejable, en combinación con la opción 1.4.

1.4 Extraer fichero de variantes

La extracción del fichero de variantes permite disponer de todas las variantes que se han producido al realizar el proceso de unificación, con objeto de tener un conocimiento, por separado, de lo que no ha sido unificado y poder disponer de estos datos a la hora de confeccionar el aparato crítico. Téngase en

cuenta que tanto en este caso como en el anterior se dispone de estos ficheros, opcionalmente, más del fichero resultante de la versión unificada.

1.5 Trabajos con los textos

Esta opción permite utilizar algunas instrucciones de UNIX para realizar búsquedas de cadenas y obtener información sobre formas léxicas o sintagmáticas que interesen, dentro de su verso. En cada caso se busca dentro de un fichero y se almacenan los resultados en un fichero común al que se van agregando los datos de las búsquedas subsiguientes, si así se desea. Se obtienen concordancias con indicación del texto donde se encuentra la cadena buscada, número de línea, y supresión optativa de las diferencias de mayúsculas y minúsculas y de los espacios en blanco entre palabras que pueden ir unidas o separadas, caso de los clíticos y de otras formas. El número de opciones y combinaciones es elevado, por lo que vale la pena destacar simplemente la opción entre recoger la concordancia con el texto o sólo la indicación numérica de la línea en la que se encuentra, procedimiento éste más rápido y de interés para recuentos estadísticos.

2. Preparar un texto original

Esta opción permite dar formato y ordenar las estrofas de forma automática. el menú para esta utilidad está compuesto de las siguientes opciones:

2. MENÚ DE PROCESOS DE PREPARACIÓN DE LOS TEXTOS

1. Introducir los datos del formato.

Formato por defecto:

espacios antes de la numeración = 4

longitud de la numeración = 3

espacios después de la numeración = 4

líneas que se ignoran al principio del fichero = 4

2. Formatear texto.

3. Ordenar las estrofas del texto.

9. Menú principal.

NOTA: No olvidar que el formateo y la ordenación se harán de acuerdo con los datos introducidos en la opción 1.

Teclear opción:

2.1 Introducir los datos del formato. Es imprescindible ejecutar esta opción antes que cualquiera de las siguientes. En este punto conviene recordar que el formato debe ser el mismo para todas y cada una de las versiones que se van a comparar, y que este formato se le proporcionará al programa de unificación desde la pantalla o modificando el fichero de parámetros. También es importante decir que, aunque se pide el número de líneas inútiles al principio del fichero, éstas no se modifican con el formateo, y únicamente indican dónde empieza el texto al que se va a dar formato.

2.2 Dar formato al texto de acuerdo con los valores introducidos anteriormente. Además establece la separación de un solo blanco entre palabra y palabra y una sola línea en blanco entre estrofa y estrofa.

2.3 Ordenar las estrofas del texto según su numeración. Si el texto no ha recibido formato previamente puede dar lugar a errores de ejecución.

Es conveniente hacer una copia de seguridad antes de dar formato u ordenar un texto; de esta forma se tendrá siempre una versión del fichero original.

3. *Identificar los ficheros que contienen los textos*

Esta opción se utiliza para proporcionar al programa de unificación los nombres de los ficheros que contienen los textos. Este proceso es obligatorio antes de pasar a cualquiera de las dos modalidades de unificación automática (opciones 4 y 5 del menú principal). Como ya señalamos anteriormente, el máximo de versiones para el estándar es de seis. El orden en que se introducen los nombres de los ficheros está en estrecha relación con la prioridad que se quiera dar a cada versión; las versiones con prioridad más alta tendrán más peso a la hora de tomar decisiones en determinadas circunstancias; aquellas en las que por otros criterios no se podría establecer el orden que deben seguir las palabras en el verso unificado.

Otro punto que se debe tener en cuenta es que el programa buscará los ficheros que almacenan los textos en cualquier directorio, siempre que el nombre del fichero incluya el «path» completo para llegar a dicho fichero. Esta particularidad es extensible a todas las utilidades de UNITE a las que se le suministran nombres de ficheros.

3. Introducir los nombres de los ficheros que almacenan los textos (0 cuando no existan más versiones)

PRIORIDAD NOMBRE

1	-----	texto 1
2	-----	texto 2
3	-----	texto 3
4	-----	texto 4
5	-----	texto 5
6	-----	texto 6

Conforme con los datos? [s(sí), n(no), q (para salir del programa)]:

4. *Ejecutar el proceso de unificación en modo INTERACTIVO*

Realiza la unificación automática mostrando en la pantalla los resultados parciales de los diferentes procesos por los que va pasando. Su utilización puede ser conveniente para comprobar la unificación de una o varias estrofas; pero se recomienda cuando los textos que se van a unificar no son cortos, ya que la utilización de la pantalla por el programa bloquea otra posible actividad del usuario. El menú al que da acceso consta de las siguientes opciones:

4. EJECUCIÓN EN MODO - INTERACTIVO -

MENÚ DEL PROCESO DE UNIFICACIÓN

1. Modificar Parámetros de ejecución desde la pantalla.
2. Asignar los valores por defecto a los Parámetros.
3. Unificación.
9. Menú principal.

- 4.1 Modificar los parámetros de ejecución desde la pantalla.
- 4.2 asignar los valores por defecto a los parámetros de la ejecución. Como

ya mencionamos, estos parámetros están almacenados en un fichero que es accesible con la opción 4 del menú de edición de ficheros.

4.3 Inicia el proceso de unificación automática. Antes de esto es necesario haber asignado valores a los parámetros con algunas de las dos opciones anteriores.

5. *Ejecutar el proceso de unificación en modo BATCH*

Una vez que se lanza el proceso batch se devuelve el control al sistema de menús, con lo que se permite realizar cualquiera otra tarea, incluso fuera del menú de UNITE. Cuando finalice la ejecución el ordenador advertirá con un mensaje al usuario.

6. *Salida temporal al UNIX*

Esta opción permite pasar del entorno de menús de UNITE al *shell* de UNIX. En el *shell* se puede ejecutar cualquiera de los comandos UNIX. Con el comando «exit» se vuelve al menú principal de UNITE.

9. *Salida del menú principal*

La tecla 9 permite abandonar el menú y retornar definitivamente a UNIX.

3. PROCESO DE UNIFICACIÓN

En este apartado se describen todas las fases por las que atraviesa el proceso de unificación. Conviene recordar que el funcionamiento de todas y cada una de estas fases está controlado por los valores almacenados en el fichero de parámetros o asignados desde la pantalla (opciones 1.4 y 4.1) del sistema de menús). Con la modificación de estos valores, el usuario delimita la capacidad del programa para tomar decisiones en el proceso de unificación.

3.1 *Lectura de los textos*

La lectura de los textos se realiza de acuerdo con los valores del formato especificados, de ahí la importancia de que estos valores se correspondan con aquellos que sirvieron para dar formato a los textos.

La unidad básica de lectura es la estrofa, formada por un número variable de versos nunca mayor que 5. Otro punto destacable es la numeración de dicha estrofa, utilizada para establecer la correspondencia entre las leídas de las diferentes versiones. La elección de la estrofa como unidad básica implica unas mayores necesidades de recursos del sistema en el proceso de unificación; pero tiene la ventaja de que se puede detectar y corregir la descolocación de los versos dentro de la estrofa en unas versiones con respecto a otras.

Como señalamos en la introducción, está previsto desarrollar una nueva versión que elimine la limitación del máximo de versos por estrofa. Esto implicaría que se podría utilizar el programa para cualquier tipo de texto escrito en verso.

3.2 Unificación de grafía

En esta parte del proceso se modifican las grafías originales de los textos hacia formas gráficas más generales. Con ello se pretende eliminar en lo posible las diferencias producidas por dichas grafías y obtener un mayor número de palabras iguales en el proceso de comparación.

La unificación de grafía está dividida en cinco fases, en cada una de las cuales se realiza un conjunto de modificaciones sobre la grafía original. Dichas fases no se ejecutan de forma consecutiva, sino que entre cada dos de ellas se intercala un proceso de comparación. De esta forma se conserva en lo posible la grafía original de los textos, ya que cuando una palabra llega a ser igual en alguna de las modificaciones, no es alterada por las siguientes. Además de esto existe una fase llamada inactiva en la que no se modifica en nada la grafía original; dicha fase ejecutada en primer lugar permite comparar los textos con su grafía original y salvaguardar de posteriores modificaciones aquellas palabras que sean iguales.

A continuación se describen cada una de las subfases:

- Fase inactiva. No modifica la grafía de los textos.
- Fase en la que se eliminan los signos de la puntuación. Elimina todos los caracteres almacenados en el fichero de signos de puntuación (opción 1.2 del sistema de menús).
- Fase en la que se sustituyen las letras mayúsculas por las correspondientes minúsculas y las vocales acentuadas por las no acentuadas.
- Fase en la que se derivan caracteres y grupos de caracteres hacia formas más generales. Dichas modificaciones están almacenadas en el fichero accedido con la opción 1.3 del sistema de menús.
- Simplificación de las letras consecutivas iguales dentro de una misma palabra.

A primera vista puede parecer que algunas de estas fases son demasiado específicas para algún tipo de textos (vocales acentuadas, letras consecutivas iguales); pero conviene hacer hincapié en las fases de derivación de caracteres hacia formas generales y eliminación de los signos de puntuación, puesto que dichas modificaciones están controladas por el usuario. Además de esto, el usuario puede establecer qué bases se ejecutarán y en qué orden y por tanto puede adaptar el procesos de unificación a las características de sus textos.

3.3 *Unificación de posiciones para una misma palabra*

En esta fase del proceso se soluciona el problema producido por aquellas palabras que aún siendo iguales en los textos comparados figuran en diferentes posiciones de dichos textos. Cuando se detecta este caso, el programa intenta asignar una única posición siguiendo los criterios del número de veces que figura en cada una de las posiciones, prioridad de cada una de las versiones y comprobación de que la posición elegida tenga ya otra palabra asignada.

El control del usuario sobre este proceso consiste en la posibilidad de decidir si se ejecuta o no y de establecer el rango de posiciones del verso sobre las que actúa.

3.4 *Unión de palabras*

Consiste en unir dos palabras en una sola si con ello se consigue una nueva palabra que se pueda unificar con alguna de otra de las versiones comparadas.

3.5 *Separación de palabras*

En esta fase se separan determinadas partes de una palabra para intentar unificar dichas partes, que ahora son nuevas palabras, con otras palabras pertenecientes a otras versiones.

3.6 *Formación de palabras a partir de las letras comunes de otras*

Este proceso es el último recurso de la unificación ya que, cuando los anteriores no han conseguido hacer iguales determinadas palabras, se intenta crear una nueva palabra a partir de las letras comunes de las que se están comparando. Se realizan sucesivas comparaciones y se elige como válida aquella que pro-

duzca una relación entre el número de letras comunes y diferentes más positiva; en todo caso, esta relación debe ser igual o superior a 1.

A continuación se comentan algunos aspectos comunes a los puntos 3.4, 3.5 y 3.6.

Antes de dar por válida la ejecución de cualquiera de los anteriores procesos, se hace un estudio para comprobar si dicha ejecución producirá resultados positivos, es decir, ayudará a elegir una nueva palabra para la versión unificada. Con esto se impide que una modificación no deseada interfiera en los posibles resultados de un proceso posterior.

Los tres procesos anteriores trabajan con palabras localizadas en diferentes posiciones; por ello, después de dar por válida la ejecución de un determinado proceso, el programa utiliza los criterios comentados en el punto 3.3, para establecer en qué posición de la versión unificada deben figurar las nuevas palabras.

Otro aspecto destacable es que estos procesos se ejecutan mezclados y en diferentes fases. En la primera fase se exige que las nuevas palabras sean iguales a algunas de las ya existentes. En fases posteriores se van eliminando restricciones, permitiendo que las nuevas palabras no sean totalmente iguales a las ya existentes; pero estableciendo como límite que la relación entre el número de letras iguales y diferentes sea mayor o igual que 1. Este bucle se ejecuta mientras queden posiciones en la versión unificada sin palabra asignada y los procesos anteriores produzcan resultados positivos.

Por último, conviene señalar que el usuario controla la ejecución de los anteriores procesos igual que lo hace para la unificación de posiciones, es decir, decidiendo si se ejecutan o no y el rango de posiciones sobre las que actúan.

3.7 *Salida de resultados*

El proceso de unificación automática genera dos ficheros de resultados. El primero de ellos almacena la versión unificada, mientras que el segundo registra una serie de datos referentes a los tres procesos vistos anteriormente y que resumen su ejecución.

Fichero que almacena la versión unificada

Al igual que los textos comparados, está dividido en estrofas separadas cada una de ellas por una línea en blanco. Además de las palabras que fueron seleccionadas para la versión unificada, figuran también las variantes que no fueron seleccionadas, con un número asignado para identificar la versión original en la que aparecían. Opcionalmente también pueden aparecer las estrofas originarias de cada versión delante de la estrofa unificada.

Fichero que resume la ejecución de los procesos de unificación

La creación de este fichero es opcional y está controlada por un parámetro modificable por el usuario. Su generación es muy conveniente cuando se necesita una aclaración de la ejecución de los citados procesos. Para cada palabra que generan dichos procesos existe una línea del fichero en la que se indican la estrofa y verso en los que figura, el proceso que la generó, la palabra unificada resultado del proceso y las palabras originales (acompañadas del identificativo de versión correspondiente) que generaron dicha palabra unificada; éstas últimas aparecen en formato original y no con el que realmente trabajan los procesos de unificación.