

Kantor Raúl, Machi Guido.

Desarrollo de una herramienta informática de abreviación

Departamento de Ciencias de la Computación
Facultad Ciencias Exactas, Ingeniería y Agrimensura
Universidad Nacional de Rosario

kantor@fceia.unr.edu.ar guido@fceia.unr.edu.ar

Resumen

La idea apunta a mejorar la velocidad de captura de datos en el eslabón más fundamental de la cadena informativa: el ser humano (razón última de todo circuito de información).

Una relevante cantidad de la información que se recibe de las computadoras es en forma de texto. Se trata de, automatizando el conocido concepto de abreviación, producir textos que se lean más rápidamente, sin perder capacidad de comprensión.

La propuesta tiene un aspecto que se encuadra en el marco del llamado "desarrollo experimental" que consiste en la realización de trabajos sistemáticos para la producción de un nuevo producto basado en conocimientos existentes y disponibles.

Así se ha diseñado un utilitario (o en realidad una familia de utilitarios con la misma filosofía operacional) que se inserta en disponibilidad en distintas herramientas informáticas que producen texto como salida para el usuario y, funcionando como un filtro en tiempo real, reemplaza el texto producido, por una versión en la que son abreviadas sus palabras. Antes de entrar en la implementación concreta de esta herramienta se necesita verificar si el producto obtenido es verdaderamente útil. Esto es si el usuario aumenta su velocidad de lectura sin perder la capacidad de comprensión de los textos. Se trata de la segunda etapa del proyecto, que estamos transitando actualmente, en la que se deben aplicar criterios comparativos. Para ello estamos diseñando mecanismos que presenten textos de similar dificultad con diversos niveles y procedimientos de abreviación. Determinar la velocidad de lectura es sencillo. La dificultad principal reside en encontrar parámetros que permitan comparar niveles de comprensión

Palabras clave: comprensión de lectura, velocidad de lectura, software libre.

Contexto

El proyecto está radicado en la Facultad Ciencias Exactas, Ingeniería y Agrimensura de la Universidad Nacional de Rosario. La unidad ejecutora tiene sede en el Departamento de Ciencias de la Computación de la citada Facultad. Está acreditado en la U.N.R. (Resolución C.S. N° 1102/2010) Diversas razones atrasaron el cronograma original por lo que está en trámite un pedido de continuidad solicitando una prórroga al 31/12/2013.

Introducción:

En los últimos años hemos presenciado un notable progreso en las prestaciones del procesamiento electrónico de la información. Han bajado dramáticamente los costos de los elementos tecnológicos y paralelamente se han incrementado enormemente las capacidades de los mismos.

La idea es desarrollar un producto que aumente la capacidad de recepción de información del eslabón más importante de la cadena informativa: el ser humano.

Observemos que respecto a facilitar el ingreso de datos se han hecho substanciales avances. Basta comparar los antiguos modos con que "el operador" (un especialista), suministraba datos a un sistema informativo (Recordemos, como figura casi "prehistórica", al perfograboverificador) con los que se ofrecen actualmente. (Íconos, ratón, pantalla táctil, etc.).

El proyecto apunta a la actividad de recibir información desde el procesador considerando que existe la posibilidad de conseguir una interesante mejora.

Nos referimos a la información que es suministrada en **forma de texto**. La idea, es aplicar un recurso conocido, de una manera automática.

Desde el comienzo del empleo de la escritura (con referencia tanto al tiempo histórico de la humanidad, como al tiempo de aprendizaje de cada individuo), surge la posibilidad de simplificarla **abreviando**.

Nuestro proyecto se propone desarrollar un utilitario (o en realidad una familia de utilitarios con la misma filosofía operacional) que se inserte en disponibilidad en distintas herramientas informáticas que producen texto como salida para el usuario. (Procesadores de texto, administradores de correo, etc.) partiendo de la construcción de un prototipo para verificar si, efectivamente, se produce una mejora en la velocidad de recepción.

Características del producto :

Se trata de un utilitario, que puede ser activado o desactivado a voluntad y con distintos criterios y niveles de aplicación, funcionando como un filtro en tiempo real sobre el texto que se “debería presentar” al usuario, reemplazándolo por una versión en la que son abreviadas algunas de sus palabras.

El mecanismo reemplaza unívocamente palabras o conjuntos de palabras por “**equivalentes**” de más corta y/o de mejor recepción expresión visual...

El concepto utilizado es “equivalentes” pues, no necesariamente nos debemos restringir a la clásica definición de abreviatura¹.

Llamaremos de todos modos abreviación al resultado de este proceso y TXTABR el texto abreviado.

Los mecanismos técnicos para conseguir dicho resultado son actualmente totalmente posibles, funciona en forma similar a las herramientas de autocorrección, y los recursos requeridos están disponibles en cualquier procesador.

Algunas características:

a) El proceso es no destructivo respecto a la información original. El texto base se mantiene disponible (inmediatamente disponible), para ser recuperado a solicitud. Puede, a pedido, automáticamente “desabreviar” tanto palabras sueltas como párrafos o el texto completo.

b) El proceso se realiza localmente. Quien produce el texto original no es el responsable de su abreviación. No se

¹ Abreviación : Representación de las palabras con sólo una o varias de sus letras , empleando a veces sólo mayúsculas y poniendo un punto al final. (Según el diccionario)

requiere por tanto acuerdo previo, y puede haber tantas versiones diferentes abreviadas como receptores. (eventualmente pueden existir diversos niveles de abreviación para un mismo texto y para un mismo receptor)

c) Si bien existen criterios para abreviar palabras que podemos considerar comunes (por ej. : **ej.**), uno de los logros del proyecto es el respeto de la enorme carga individual que tiene un proceso de abreviación.

d) El concepto tiene un aspecto global, independiente de los diversos idiomas, y múltiples niveles de aplicación y refinamiento que van de un idioma específico, pasando por comunidades de interés hasta llegar al diccionario de un usuario en particular.

Definida la idea central, se abren inmediatamente diversas ramificaciones para ser desarrolladas.

El mecanismo del motor en sí y la optimización de su rendimiento, el modo de inserción en cada uno de los utilitarios a los que sirva, son aspectos que ofrecen, a partir de la gran cantidad de material afín existente, un territorio de mejora continua, con seguros resultados aplicables.

Por otra parte la localía de los diccionarios ofrece la posibilidad de su construcción en el ámbito de comunidades específicas.

Finalmente está el amplio terreno que existe para la construcción de los símbolos “equivalentes”.

La concepción que una abreviatura se debe construir con letras seguidas de un punto es un concepto proveniente de restricciones tipográficas, que no tienen nuestros procesadores

Líneas posibles a recorrer (algunas), son la inserción de símbolos, el uso de colores, el empleo de sub y superíndices, diversos tipos de letras y diversos alfabetos.²

El carácter de la idea es universal, su implementación local.

Se intenta efectivizar la construcción bajo la modalidad del llamado software libre.

Es entonces que la propuesta se desarrolla en código abierto, sin reservar más derecho que el de que no se pueda, de ninguna manera, establecer limitaciones a su disposición pública.

² Existen lenguas donde sólo figuran las consonantes y las vocales se indican como marcas bajo o sobre las letras.

Es interesante comparar la cantidad de espacio que utilizan diversas traducciones de un mismo texto...

Se ofrecen públicamente los progresos del proyecto, poniendo a disposición de los posibles usuarios del producto cada uno de los logros conseguidos, junto con el código abierto de los programas para quienes les interese conocer y eventualmente participar en el aspecto de la elaboración técnica subyacente.

Se contempla la posibilidad de que incorporen sugerencias externas tanto desde el punto de vista técnico del producto, como en cuanto a propuestas mejoras, o agregados a sus prestaciones, manteniendo, para ello, desde el proyecto una conducción centralizada de versiones y documentación.

Se ha seleccionado como lenguaje a utilizar el denominado Python en el que hemos desarrollado un prototipo que consta entonces de un motor de abreviación y de un "diccionario personalizado" que es ampliado y perfeccionado por el usuario a partir de un núcleo base.

Utilidad del Producto:

Estimamos que conviene, previo a continuar con el complejo proceso de concreta implementación de la herramienta en diversos productos del mercado, comprobar su utilidad. Esto es verificar si efectivamente con su uso se gana en velocidad de lectura sin pérdida de comprensión.

Es en esa etapa en la que nos encontramos actualmente.

La idea es desarrollar un mecanismo de testeo que compare las velocidades de lectura y la capacidad de comprensión de textos.

El lector sujeto de la prueba de haberse "entrenado" en la lectura de diversos TXTABRs,

En la prueba concreta se encontrará con una pantalla en blanco y al presionar una tecla aparecerá un TXTABR de aproximadamente unas veinticinco líneas.

(el tamaño del Resumen que encabeza este documento).

Debe presionar una tecla cuando termina la lectura, para medir el tiempo que le llevó la misma..

Inmediatamente debe contestar una serie de preguntas que apuntan a medir la comprensión de lo leído.

Luego debe realizar lo mismo con un texto de similares características pero en su forma usual.

Los principales problemas que presenta este mecanismo de prueba son a) Establecer cuando dos textos son "similares" y b) el criterio sobre determinar las preguntas para medir la comprensión. En principio pensamos en múltiple choice para facilitar la automatización del proceso. La intención es conseguir cantidad de voluntarios que se presten a las pruebas, probablemente creando una página en la WEB, de donde además se pueda descargar el producto, mantener la discusión sobre sus posibilidades y recibir críticas y aportes.

Objetivo:

Si se logra reducir el tiempo que se dedica a la lectura de correos diariamente, en un modesto porcentaje, ya está justificada la existencia del producto.

Formación de Recursos Humanos:

En esta etapa del proyecto se incorporaran al proyecto alumnos de la carrera de Licenciatura en Ciencias de la Computación para la realización de diversas tareas de implementación. A esos efectos se realizará un llamado a concurso. Se estima que para los alumnos participantes será una buena oportunidad de formación tanto desde el punto de vista de realización de una práctica supervisada, como de una primera introducción a un proceso de Investigación y Desarrollo.