

## 75.06 Organización de Datos

### **Trabajo Práctico Primer Cuatrimestre 2004 – Curso Servetto**

En esta entrega se implementará un programa compresor/descompresor optimizado para archivos de texto y con la funcionalidad de multivolumen, es decir, en vez de generar un único archivo comprimido se generarán N archivos que no superen un tamaño especificado (salvo el último si corresponde).

La sintaxis del ejecutable a realizar, donde XX debe reemplazarse por el número de grupo es:

- **grupoXX -c -tNNN archivoentrada archivosalida**

donde:

**-c** : indica que es una operación de compresión

**-tNNN** : es un switch adicional que especifica que el archivo de salida debe ser multivolumen, donde todos los archivos generados (salvo el último) deberán medir EXACTAMENTE NNN kbytes (1 kbyte = 1024 bytes). NNN no tiene una longitud fija de caracteres que puede ocupar (no solo tres).

**archivoentrada** : nombre del archivo a comprimir. Puede no ser de texto, pero el compresor siempre estará optimizado para comprimir mejor cuando lo sea.

**archivosalida** : nombre del archivo comprimido a generar. En caso de haberse utilizado multivolumen (switch t), se especificará únicamente el nombre del primero de los archivos.

- **grupoXX -d archivocomprimido archivosalida**

donde:

**-d** : indica que es una operación de descompresión.

**archivocomprimido** : nombre del primer archivo comprimido. El descompresor deberá detectar si se utilizó multivolumen y en caso afirmativo continuar la descompresión del resto de los volúmenes en forma automática, es decir, sin pedir el nombre del resto de los volúmenes si los mismos se encuentran en el mismo directorio del archivo principal. De no ser así, se le deberá pedir al usuario que especifique el path en donde se encuentra el volumen no encontrado a medida que se van requiriendo.

**archivosalida** : nombre del archivo de salida descomprimido.

Para la implementación se deberá optar por algunos de los siguientes métodos:

- Move to front + Block Sorting + otro compresor (estadístico o no)
- Aritmético con contextos (mínimo orden 2)
- LZHUFF
- PPMC (mínimo orden 2)
- LZP (mínimo contexto 4)

modificándolos para que compriman mejor archivos de texto en castellano (y por ende, compriman un poco peor archivos binarios).

En el informe se deberá incluir:

- Justificación del compresor elegido.
- Descripción de ajustes hechos en el compresor para la optimización de textos.
- Comparación con otros ajustes posibles (mínimo 2) y justificación de la elección en base a pruebas de tiempo y tamaño. Para lo cual se deberá indicar el ambiente de trabajo utilizado.