
Capítulo 5

ANALIZADOR DE FICHEROS

La información relativa a nombres de dominio que suministra el sistema ANS está recogida dentro de unos ficheros denominados “*master files*”, siendo cada fichero único para cada una de las zonas en las que se divide el espacio de nombres de dominio. Estos ficheros tienen una sintaxis particular, siendo la funcionalidad del bloque que se describe en este capítulo la de generar un fichero *equivalente* al “*master file*”, de forma que el sistema ANS pueda obtener directamente los datos de este fichero.

DESCRIPCIÓN SINTAXIS DE FICHEROS DE ZONA

Como es sabido, cualquier dato dentro del sistema ANS concerniente a un determinado nombre de dominio se guarda dentro de una estructura de datos denominada registro fuente. La estructura de un registro fuente es la siguiente:

NOMBRE DE DOMINIO
CLASE
TIPO
TIEMPO DE VIDA
LONGITUD DE LOS DATOS
DATOS

Los “*master files*” son ficheros de texto que contienen los registros fuentes para cada de las distintas zonas del espacio de nombres. Como norma general, cada registro fuente ocupará una única línea dentro de cada fichero, por lo que cada una de las líneas tendrá el siguiente formato:

NOMBRE DE DOMINIO + CLASE + TTL + TIPO + DATO

- *NOMBRE DE DOMINIO*: El del registro fuente al cual se refiera.
- *CLASE*: La del registro fuente(en el caso que nos ocupa siempre será de tipo INTERNET). En su defecto se tomará como clase la del registro fuente de tipo autoritario(SOA) para la zona a la que pertenece el registro fuente.
- *TTL*: Tiempo de vida del registro fuente. En su defecto se tomará como TTL del registro fuente el valor del parámetro “*mínimo tiempo de vida*”(parámetro del registro fuente tipo SOA) asignado a la zona.
- *TIPO*: Tipo de dato del registro fuente(NS, CNAME, SOA, ATMA, ...)
- *DATO*: Son los datos que contiene el registro fuente. Su formato variará según el tipo del registro fuente.

Además existen una serie de reglas que ha de cumplir la sintaxis de los registros fuentes dentro de los “*master files*” de las distintas zonas. Un ejemplo regla sintáctica es que si al principio de una línea, en vez de aparecer un nombre de dominio, aparece un espacio en blanco, se tomará como nombre de dominio el aparecido en la línea previa.

El tipo SOA es un tipo de registro fuente especial. De este tipo de registro sólo puede aparecer uno por cada “*master file*” o zona, estando ubicado al principio de fichero. Puede y suele ocupar más de una línea, y en él se recogen los parámetros de mantenimiento de la zona en cuestión.

Por otra parte se pueden encontrar una serie de líneas que no responden a la sintaxis genérica de un registro fuente dentro de un fichero. Estas líneas poseen el siguiente formato:

\$origen <nombre de dominio>

- *\$origen <nombre de dominio>*: Permite definir un nombre de dominio “*origen*”, de forma que a partir de esta línea, todos los nombres de dominio que aparezcan en registros fuentes posteriores sean relativos a este “origen” en cuestión. El carácter ‘@’ será un carácter comodín que equivale al nombre de dominio definido como origen, y podrá ser utilizado dentro del “*master file*” para referirse a dicho nombre de dominio “origen”.

\$incluye <nombre de fichero>

- *\$incluye <nombre de fichero>*: Sirve para poder anexar nuevos ficheros de datos al “*master file*” de la zona que se esté analizando. De esta forma los datos relativos a una determinada zona podrán estar distribuidos en más de un fichero.

Los comentarios dentro del master file se sitúan en cualquier línea, y van desde la primera ocurrencia del carácter ‘;’ hasta el final de línea.

A continuación se muestra un fichero de datos (“*master file*”) de ejemplo:

```
.....
..... Master File example
.....
data.example.com.      IN      SOA      name1.data.example.com. name4.data.example.com. (
                        1994041800   ;serie(fecha)
                        1800      ;refresco
                        300       ;reintento
                        604800    ;expiración
                        3600     )      ;minimo
                        IN      NS      name1.data.example.com.
                        IN      NS      ns.example.com.
;
$ORIGEN data.example.com.
salmon IN      ATMA 39.246f.00e7c9c031200010001.000012345678.00
char   IN      ATMA 39.246f.00e7c9c031200010001.000012345678.00
```

FILTRADO DE FICHEROS DE ZONA.

En el punto anterior se ha mostrado una serie de reglas que deben cumplir los “*master files*” y cómo estos ficheros poseen una sintaxis que hace que los datos de los registros fuentes que contienen dependan del contexto dentro del fichero. Para que el sistema ANS pueda tener unos ficheros de datos que no puedan dar lugar a ningún tipo de confusión en su lectura. Se ha realizado un filtro que toma dichos ficheros de datos y genera unos ficheros de salida equivalentes en los que cada una de sus líneas contiene un registro fuente con todos sus campos rellenos. De esta forma la lectura de los registros fuentes en un fichero de datos de una zona se hará línea a línea de forma secuencial.

El filtro anterior se ha desarrollado con la ayuda de la herramienta *FLEX* existente en sistemas operativos UNIX. *FLEX* es un analizador léxico, y su función principal es la de, una vez definidos una serie de patrones de caracteres, los identifica dentro de un fichero de texto, pasando a ejecutar una serie de acciones sobre dicha cadena de caracteres, para, al final, colocarla dentro de un fichero de salida. Los ficheros escritos en *FLEX*, por convenio, poseen una extensión “.yy”, pasándose como parámetro al ejecutable *flex*, que genera un fichero en lenguaje C del cual se obtiene el filtro deseado.

En un fichero *flex* primeramente hay que definir los distintos patrones de caracteres que se van a poder encontrar dentro del fichero de datos. Estos son los siguientes:

NOMBRE DE DOMINIO: Sucesión de etiquetas, separadas por un punto o un carácter @. Cada etiqueta puede estar formada por una letra del abecedario, un número o los caracteres '- ' ó '_ '.

CLASE: Se corresponde con la cadena "IN".

TIPO: Cualquiera de los distintos tipos de registro fuente que se puede encontrar: "SOA", "ATMNSAP", "ATME164", "NS", "PTRNAME", "CNAME".

TTL: Un número.

DIRECCIÓN ATM: Sucesión de números hexadecimales o decimales.

Una vez definidos los distintos patrones de caracteres que se pueden encontrar dentro de un fichero de datos, se pasa a definir una serie de posibles estados en la lectura de registros fuentes. Por estos estados se irá transcurriendo a la hora de ir extrayendo registros fuentes del fichero de datos que se esté leyendo. Un registro fuente dentro de un fichero de datos tendrá el siguiente formato:

nombre de dominio + clase + ttl + tipo + dato

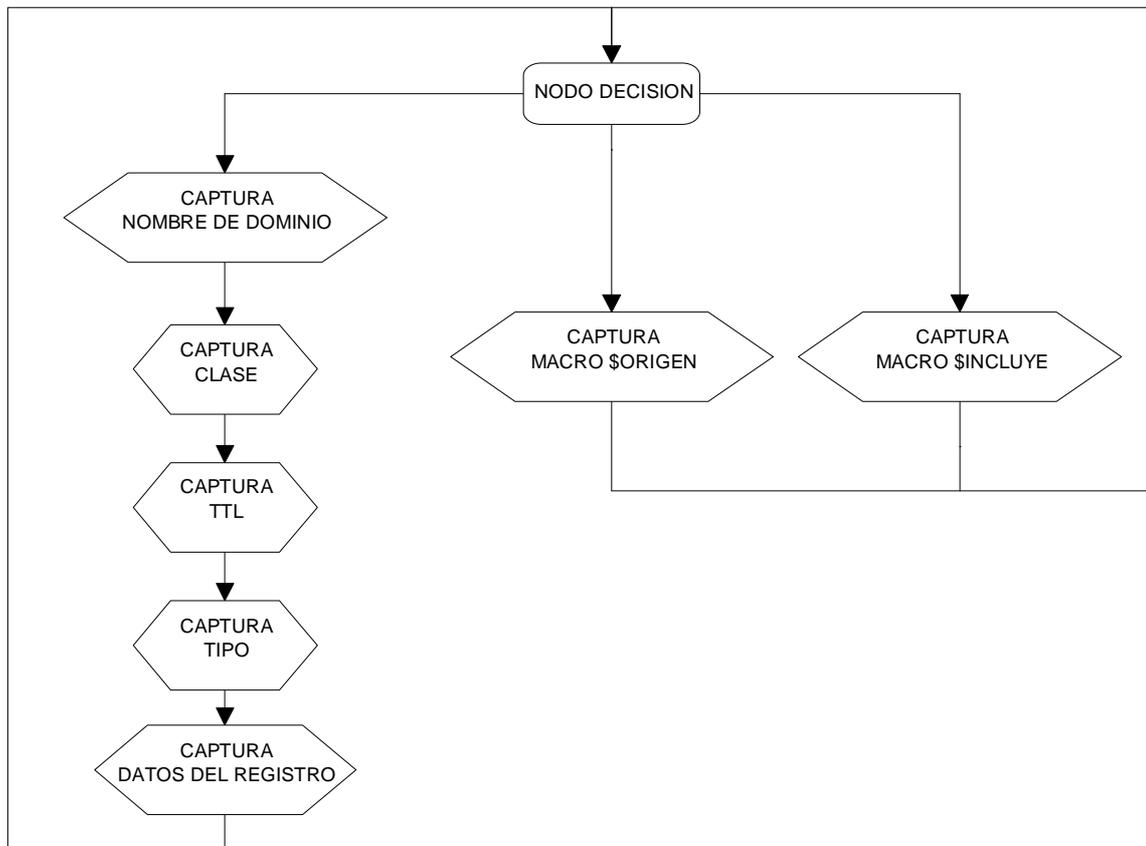
En la fase de lectura de dicho registro fuente se pueden definir estos estados:

- 1) **CAPTURA DE NOMBRE DE DOMINIO:** Este estado se corresponde con la acción de encontrar un principio de línea seguido de un nombre de dominio. Se obtiene el nombre de dominio del registro fuente y se pasa al siguiente estado.
- 2) **CAPTURA CLASE:** Este estado se corresponde con la acción de encontrar una determinada clase. Se pasa al siguiente estado.
- 3) **CAPTURA TTL:** Este estado se corresponde con la acción de encontrar un determinado TTL. Se pasa al siguiente estado.
- 4) **CAPTURA TIPO:** Este estado se corresponde con la acción de encontrar un determinado TIPO válido. Se pasa al siguiente estado que dependerá del tipo encontrado.
- 5) **CAPTURA DATOS:** Depende de cada tipo de registro fuente, ya que el formato de cada campo dato varía según dicho valor.

Los anteriores estados se corresponden con las fases de lectura de un registro fuente dentro de un fichero de datos. Para otras tareas específicas, que no se corresponden con la lectura de registros fuentes, se definen dos nuevos estados:

- 6) **CAPTURA ORIGEN:** Este estado se corresponde con la acción de encontrar una línea que contenga la macro “\$origen”. En esta línea se define la variable origen del fichero, y se le da como valor el de un nombre de dominio.
- 7) **CAPTURA INCLUYE:** Este estado se corresponde con la acción de encontrar una línea que contenga la macro “\$incluye”. En esta línea se anexa un nuevo fichero de datos al ya existente, y contendrá nuevos registros fuente válidos para la zona.

El filtrado de un fichero de datos (“*master files*”) dado se realizará siguiendo una determinada sucesión de posibles estados, y generando otro fichero de salida *equivalente* que no dé lugar a posibles interpretaciones del mismo. La siguiente gráfica muestra los distintos estados en los que puede encontrarse el filtro en su fase de lectura de registros, así como la relación existente entre los distintos estados:



APLICACIÓN AL SISTEMA ANS.

El espacio de nombres de dominio se divide en una serie de zonas, cada una de las cuales definidas por un fichero de datos o “*master file*”. Estos ficheros de datos poseen una determinada estructura de acuerdo con lo que permite la norma del protocolo ANS. Con ayuda del filtro desarrollado se generarán unos ficheros equivalentes a los *master files* que definen cada zona. Estos nuevos ficheros de datos contendrán un registro fuente por cada línea, sin omisión de ningún tipo de dato, estando definidos todos los campos de todos los registros fuentes. Estos ficheros los utilizará el bloque lector de ficheros del sistema ANS, para obtener datos de cada zona, simplificando en gran medida la tarea de obtención de datos de los *master files*.