

PROVINCIA DE SANTA FE
INSTITUTO PROVINCIAL DE ESTADÍSTICA Y CENSOS

APLICACIÓN
PARA LA
NORMALIZACIÓN DE NOMBRES DE CALLES
DE ARCHIVOS DE USUARIOS
DE LA
EMPRESA PROVINCIAL DE LA ENERGÍA

Febrero de 2002

**APLICACIÓN
PARA LA
NORMALIZACIÓN DE NOMBRES DE CALLES
DE ARCHIVOS DE USUARIOS DE LA EPE**

1- OBJETIVOS

2- METODOLOGÍA

3- INSUMOS

4- PRODUCTO

5- DESCRIPCIÓN DEL PROCESO

6- EVALUACIÓN

Análisis: Stella Maris Agueria

Programación: Néstor Mussio

1- OBJETIVOS

El objetivo principal es la normalización del atributo “nombre de calle” que forma parte del campo domicilio en el archivo de usuarios de la Empresa Provincial de la Energía.

Secundariamente se valida el domicilio así obtenido contra el mapa digital de manera de verificar la inclusión en el mismo ó, en caso contrario, buscar por otros medios la causa de la inconsistencia.

2- METODOLOGÍA

La aplicación utiliza el concepto de “fuzzy logic”, en castellano “lógica difusa”. De manera que, cuando no puede resolver un nombre de forma 0 directa, trata de acotar las posibilidades a un conjunto finito de probables soluciones, en base a un análisis de los elementos del nombre a partir de un cierto conjunto de criterios generales.

El conjunto solución puede resultar vacío si el análisis no conduce a nada; puede contener un cierto número pequeño de posibles soluciones cuya elección queda en manos del operador; puede contener un gran número de posibles soluciones que indiquen que el nombre analizado es muy ambiguo y entonces debe dársele tratamiento especial para resolverlo; por último, puede contener un único elemento que deviene en solución automática del nombre analizado.

3- INSUMOS

La aplicación utiliza archivos en formato DBF estándar. Para poder ejecutar la tarea necesita tres archivos, a saber:

- El archivo de usuarios de la EPE conteniendo los domicilios, discriminados en calle y número de puerta, a normalizar y codificar.

<u>Atributo</u>	<u>Formato</u>	<u>Ancho</u>	<u>Dec.</u>	<u>Comentario</u>
<i>PLAN</i>	Carácter	2		Secuencia geográfica
<i>RUTA</i>	Carácter	3		Secuencia geográfica
<i>FOLIO</i>	Carácter	5		Secuencia geográfica
<i>TARIFA</i>	Carácter	4		Discrimina tipo de usuario
<i>LOCALIDAD</i>	Carácter	30		Nombre de la localidad
<i>DOMICILIO</i>	Carácter	60		Domicilio completo
<i>CALLE</i>	Carácter	60		Nombre calle
<i>NRO</i>	Carácter	15		Nro. de puerta
<i>RESTO</i>	Carácter	20		Otros datos del domicilio
<i>CODIGO_NEW</i>	Numérico	6		Código de Calle

- El archivo del codificador de calles conteniendo los datos normalizados de las calles.

<u>Atributo</u>	<u>Formato</u>	<u>Ancho</u>	<u>Dec.</u>	<u>Comentario</u>
<i>PROV</i>	Numérico	2		Código de provincia
<i>DPTO</i>	Numérico	3		Código de departamento
<i>LOCA</i>	Numérico	3		Código de localidad
<i>CODIGO</i>	Numérico	6		Código de calle
<i>ANCHO</i>	Numérico	2		Ancho
<i>NOM_USUAL</i>	Carácter	30		Nombre usual
<i>SENT</i>	Carácter	1		Sentido de crecimiento
<i>ALT_COR</i>	Numérico		6	Altura de corte
<i>TIPO</i>	Numérico	1		Tipo de calle
<i>NOM_NOR</i>	Carácter	30		Nombre normalizado

- La **AAT** ó **Tabla de Atributos de Arcos** de la cobertura de ejes del mapa digital que tenga los atributos de las calles con rango de alturas para cada **arco**.

<u>Atributo</u>	<u>Formato</u>	<u>Ancho</u>	<u>Dec.</u>	<u>Comentario</u>
<i>PROV</i>	Numérico	2		Código de provincia
<i>FNODE_</i>	Numérico	11		Atributo interno
<i>TNODE_</i>	Numérico	11		Atributo interno
<i>LPOLY_</i>	Numérico	11		Atributo interno
<i>RPOLY_</i>	Numérico	11		Atributo interno
<i>LENGTH</i>	Numérico	13	6	Longitud del arco
<i>E0041_</i>	Numérico	11		Atributo interno
<i>E0041_ID</i>	Numérico		11	Identificador del arco
<i>CODIGO</i>	Numérico	6		Código de calle
<i>DESDEI</i>	Numérico	6		Nro. de puerta desde izq.

<i>DESDED</i>	Numérico	6	Nro. de puerta desde
der.			
<i>HASTAI</i>	Numérico	6	Nro. de puerta hasta izq.
<i>HASTAD</i>	Numérico	6	Nro. de puerta hasta
der.			

4- PRODUCTO

El producto resultante del proceso es el archivo de usuarios modificado, extendido, con un nuevo atributo, de formato numérico, llamado CODIGO_NEW, insertado para cada domicilio en el archivo de usuarios de la EPE.

Secundariamente se obtiene una página con estadísticas del proceso y también pueden generarse listados de calles asignadas y no asignadas.

CODIGO_NEW contiene el código de la calle, según el codificador de calles normalizadas, a la que pertenece el domicilio. Dicho código será un valor numérico entre 1 y 99999 inclusive. Se utilizaron códigos cuyos valores más comunes son múltiplos de cinco. El valor inicial de CODIGO_NEW, antes de procesar los domicilios, es cero.

Después del proceso, de no haberse identificado la calle a la que pertenece el domicilio, el atributo contiene el código que identifica la causa del fracaso en la identificación:

- **Domicilio ambiguo (111111):** el nombre del domicilio podría pertenecer a más de diez calles diferentes. Esto sucede cuando la información del archivo de usuario es insuficiente ó muy confusa.
- **Domicilio inexistente (999999):** el nombre del domicilio no está contenido, ni siquiera confusamente, en ningún nombre de calle.

- **Domicilio rechazado por el operador (777777):** aunque la aplicación sugirió unas pocas calles a las que el domicilio podría pertenecer, el operador por razones subjetivas no aceptó ninguna de las sugerencias y el domicilio quedó sin codificar.
- **Domicilio con número de puerta inexistente (444444):** Se detectó la calle a la que presuntamente pertenece el domicilio, pero al confrontar la altura en el mapa digital, resultó inexistente.

5- DESCRIPCIÓN DEL PROCESO.

La primera vez que se ejecuta la aplicación para el archivo de usuarios de una localidad dada, la aplicación crea un **glosario** que contiene el universo de palabras que componen los nombres de las calles contenidas en el codificador de calles correspondiente a la localidad.

La creación de este glosario tiene una instancia de consulta al operador para resolver las abreviaturas no estándar contenidas en el glosario, a fin de vincularlas con sus correspondientes significados.

Para codificar los nombres de calles del archivo de usuarios la aplicación tiene dos instancias. Una primera instancia **automática**, que se ejecuta batch e intenta codificar todos los nombres posibles. Una segunda instancia **asistida**, en la cual el operador es asistido por la aplicación para resolver los nombres que no hayan sido codificados en la instancia anterior.

En todos los casos la aplicación recorre los mismos pasos: reconocer el nombre como idéntico al del codificador; reconocer las palabras que componen el nombre como pertenecientes al glosario por identidad, por abreviación ó por error ortográfico; reconocer la ó las calles a las que pertenece el conjunto intersección de todas las palabras encontradas en el glosario; verificar y acotar el conjunto solución según el

rango de números de puerta existente en el mapa digital; sugerir las posibles soluciones ó determinar la imposibilidad de resolver el caso a partir de la información disponible.

En la instancia asistida el operador tiene la posibilidad en todo momento de examinar cualquiera de los casos, aún los definidos como sin solución, y en última instancia forzar una solución si a su criterio se justifica. Un elemento adicional con el que cuenta el operador para los casos difíciles son los atributos de ruta y folio del archivo de usuarios que le permiten aproximar una secuencia geográfica de los registros.

6- EVALUACIÓN.

La experiencia nos a demostrado que si bien esta metodología no resuelve la totalidad de los casos, si resuelve automáticamente más de la mitad de los mismos y sugiere interactivamente la solución correcta para otro veinte a treinta por ciento de los casos.

Es destacable que los casos resueltos tienen un porcentaje de certeza superior al 98%. Los casos sugeridos permiten llegar a una solución inmediata en más del 70 % de ellos. Este rendimiento ha sido obtenido a partir de los archivos de la EPE para ciudades como Rafaela, Santa Fe y Rosario.

Es posible afinar más la aplicación ya que hay unos pocos casos cuya solución aparece como sencilla pero sin embargo no llega a encontrarla, a veces ni a sugerirla.

Hemos verificado que la calidad de los casos resueltos depende casi exclusivamente de la veracidad, integridad e historicidad del codificador de calles y de la calidad del mapa digital que se utiliza para verificar los números de puerta.

Si bien la aplicación nació para resolver la codificación de los archivos de usuarios de la EPE, es adaptable a otros diseños que contengan al menos los atributos de nombre de calle, nro. de puerta y, si es posible, atributos de secuencia geográfica de relevamiento de la información tales como ruta y folio ó manzana y nro. de secuencia de la vivienda ó análogos.