

USL: UN SISTEMA PARA INTERROGAR EN CASTELLANO

A BASES DE DATOS RELACIONALES

Luis de Sopena.

Es un sistema experimental desarrollado en el Centro Científico de IBM en Heidelberg (Alemania) inicialmente por H. Lehmann, N. Ott y M. Zoeppritz. Información básica sobre él puede encontrarse en (3, 4). Se han implementado tres versiones de USL: alemana, inglesa y española, que es la que aquí vamos a considerar (5).

1. OBJETIVOS

USL constituye una interfase interactiva con un Sistema de Gestión de Bases de Datos Relacionales. Traduce las frases de entrada escritas en Lenguaje Natural a sentencias del lenguaje formal de interrogación de la base de datos. Con dicha expresión accede a la Base, y una vez obtenida la respuesta, se la presenta al usuario, bien directamente, bien después de llevar a cabo alguna operación adicional sobre ella.

El principio fundamental de diseño de USL es la independencia respecto del dominio de la aplicación, es decir que el sistema es utilizable para diferentes dominios. Para lograr esto han debido desarrollarse las reglas de la gramática española y las rutinas de interpretación adecuadas para el subconjunto del LN que se necesita para manipular las informaciones de una base de Datos.

USL se dirige a usuarios no especialistas en Informática, pero con un cierto conocimiento del dominio de la aplica-

ción: .

Por otra parte, no intenta simular diálogos ni construir modelos de adquisición de conocimientos.

2. DESCRIPCION DEL SISTEMA

El método utilizado para el proceso de traducción mencionado se basa en el análisis sintáctico y semántico del subconjunto del castellano escogido.

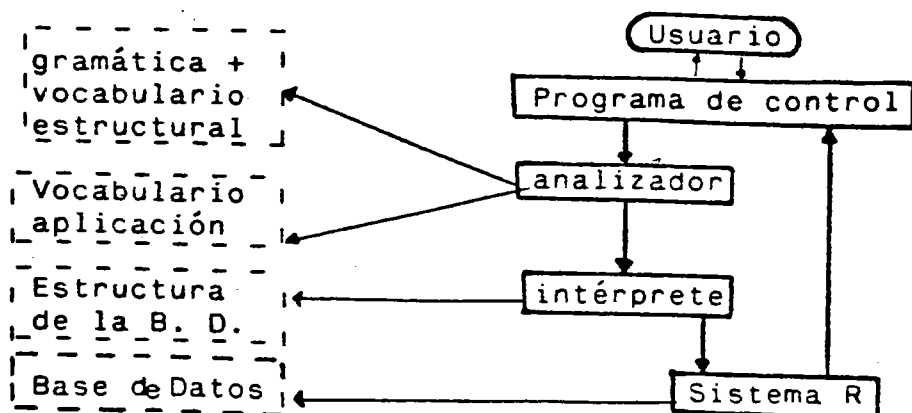
El sistema se compone de:

- Un analizador llamado USAGE (2), que acepta gramáticas escritas en un formato BNF modificado, y permite la utilización de rutinas semánticas.
- Una gramática de unas 450 reglas, donde se codifican las estructuras sintácticas permitidas.
- Un vocabulario de unas 400 palabras estructurales, incluyendo cuantificadores, preposiciones, pronombres, nombres de meses, los verbos tener y ser, etc.
- Un conjunto de rutinas de interpretación, especificadas en las reglas de la gramática, que llevarán a cabo de traducción al lenguaje de la base de datos.
- Un sistema de Base de Datos Relacionales: el sistema R (1), que dispone de su propio lenguaje formal de interrogación, SQL.

Además, el usuario deberá definir para cada aplicación particular:

- Las relaciones base que constituyen la base de datos, y las relaciones virtuales (visualizaciones) definidas sobre ellas.
- El vocabulario asociado con la aplicación.

El diagrama del sistema es:



Las palabras estructurales y el significado de las estructuras sintácticas se suponen independientes del dominio y son en consecuencia interpretadas por las rutinas semánticas de un modo general. Por el contrario, los nombres, verbos y adjetivos se consideran dependientes de la aplicación, y deben de ser definidos por el usuario de acuerdo con la estructura de las relaciones de la Base de Datos. Los nombres de contenidos y los valores numéricos no necesitan ser definidos.

Los criterios seguidos para la representación del conocimiento en USL son los siguientes: el conocimiento conceptual se encuentra contenido en las relaciones virtuales, que están definidas sobre las relaciones físicamente almacenadas en la base, y ligadas a las palabras del vocabulario; el conocimiento factual está almacenado directamente en la base de datos, y no ha sido duplicado en forma de diccionario o de red semántica.

En USL hay una separación total entre el analizador y la base de datos: durante el análisis no se accede a la base, durante la interpretación solamente se utiliza información sobre su estructura.

3. COBERTURA DEL LENGUAJE

Para describir la cobertura del castellano que el sistema proporciona usaremos ejemplos clasificados según la estructura o construcción sintáctica:

- Interrogativas generales: Pertenece Holanda a la OTAN?
- Interrogativas parciales: Qué país exporta petróleo?
- Ordenes : Lista los países árabes
- Enunciativas: Juan vive en Madrid
- Complementos del nombre : Qué empleados de Juan viven en Madrid?
- Negación : Qué países no importan petróleo?
- Adjetivos: Cuales son los empleados solteros.
- Aposiciones: Qué empleados de García, el jefe, están casados?
- Oraciones de relativo: Cuáles son los países que exportan vino?
- Coordinación: Qué edad y qué salario tienen los empleados; quién vive en Madrid o en Barcelona?
- Cuantificadores: Tienen todos los jefes al menos 5 empleados?
- Comparación : Quien gana más que Juan?
- Posesivos: Qué empleado gana más que su jefe.
- Complementos de tiempo: Quién nació el 5 de marzo de 1.959?
- Funciones de suma, media, máximo, mínimo:Cuál es el salario medio?
- Subtotales: Cuál es la suma de los salarios de los empleados de cada jefe?

4. APLICACIONES Y RESULTADOS

En su estado actual, USL puede ser directamente utilizado para interrogar y actualizar una Base de Datos Relacional, y los usuarios pueden formular un número virtualmente ilimitado de preguntas diferentes utilizando el castellano como lenguaje de comunicación.

Aunque la versión española de USL aún no ha sido sometida a tests utilizando usuarios y aplicaciones reales, se han construido dos bases de datos experimentales para la puesta a punto y demostración del sistema, una con datos geográficos y otra con datos sobre los empleados de una empresa. Los ejemplos antes mencionados se han tomado de dichas bases.

Son interesantes, la independencia del sistema respecto del dominio, el hecho de que sea una interfase con un sistema general de Bases de Datos de estructura relacional, así como las facilidades de uso y definición de palabras.

REFERENCIAS

- /1/ ASTRAHAN M. M. et. al.
"System R: Relational approach to database management"
ACM Transactions on Database Systems, vol. 1, nº 2,
junio 1976.
- /2/ BERTRAND O., J. DAUDENARDE
"Usage-processing and natural language"
IBM Paris Scientific Center, 1977.
- /3/ LEHMANN H.
"Interpretation of natural language in an information
system", IBM Journal of Research and Development,
vol. 22, 1978.
- /4/ OTT N., ZOEPPRITZ M.
"USL - An experimental information system based on
natural language", in Bolc (Ed.): Natural Communication
with computers, vol. 2, Carl Hanser Verlag, Muenchen-
Wien, 1979.
- /5/ SOPEÑA L.
"Grammar of Spanish for the user specialty languages
system", IBM Heidelberg Scientific Center, 1982.