

TRATAMIENTO DE TEXTOS DE QUIMICA EN CASTELLANO

Osvaldo Cairó Battistutti *

Sección de Computación
Departamento de Ingeniería Eléctrica
CINVESTAV - IPN

Apartado Postal 14-740

07000 - México DF

RESUMEN

En este artículo se presenta un sistema que lleva a cabo la comprensión de textos en lenguaje natural aplicado al campo de la química.

El sistema funcionalmente opera en dos fases. La primera fase corresponde al procesamiento del lenguaje natural y la segunda fase corresponde al mecanismo de inferencia del sistema.

El objeto de este artículo es presentar las técnicas utilizadas en ambas fases.

La información básica del sistema consiste esencialmente de las características físicas y químicas de los elementos de la tabla periódica. En base a esta información el sistema puede deducir propiedades, como así también combinaciones entre elementos (fórmulas y nombres de fórmulas).

El sistema fue implementado en LISP.

* Estudiante de Doctorado de la Sección de Computación.

Departamento de Ingeniería Eléctrica.

1 - INTRODUCCION

El sistema desarrollado lleva a cabo la comprensión de textos en lenguaje natural (dado el carácter de la lingüística computacional, complicado en todos sus aspectos) aplicado al campo de la química, y concerniente a las propiedades físicas y químicas de los elementos de la tabla periódica.

La entrada del sistema es un enunciado y la salida del mismo es la respuesta a la interpretación semántica del texto ingresado.

Funcionalmente, el sistema opera en dos fases. En la primera fase se hace un análisis sintáctico y semántico del texto ingresado y en la segunda, el sistema hace inferencias para poder responder a las exigencias del usuario.

Para la primera fase, el sistema hace uso de una Red de Transición Aumentada (ATN) para ⁹²efectuar los análisis sintáctico

y semántico.

Sintaxis y semántica se presentan, en este trabajo, completamente fusionadas en el tratamiento del lenguaje natural y por lo tanto los análisis se realizan en forma concurrente a lo largo del procesamiento del texto ingresado.

El ATN lo podemos definir como un analizador sintáctico (intérprete) que contiene una gramática que se aplica de manera recursiva y de arriba hacia abajo y un diccionario de datos, los cuales son independientes entre sí. Es decir, que el intérprete, la gramática y el diccionario se encuentran separados, y por lo tanto cualquiera de ellos puede sufrir modificaciones sin alterar el resto del sistema.

Prueba de ello es que el sistema funcionó incluso con una gramática general para el castellano y los resultados fueron excelentes. Solo fue necesario definir otra gramática y otros items del diccionario de datos pero el intérprete utilizado fue el mismo para ambas.

Para la segunda fase, citada anteriormente, el sistema utiliza una máquina de inferencias, un conjunto de reglas, y una base de hechos, las cuales son independientes entre sí. Es decir, cualquiera de ellas puede sufrir modificaciones sin alterar el resto del sistema.

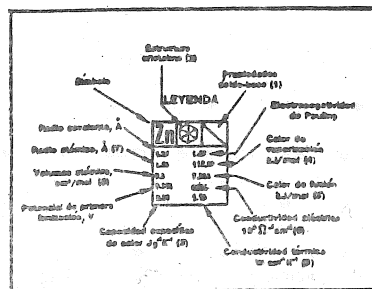
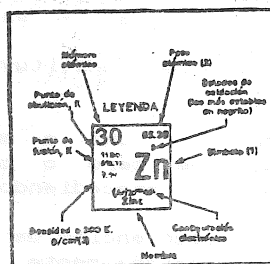
Las reglas contienen la información que el experto utiliza para la resolución de sus problemas y representa el mecanismo válido de razonamiento o inferencia, no existe ambigüedad. El intérprete es el que accesa las reglas.

La información básica con la que se dota al sistema se presenta en la figura 1 y consiste esencialmente de las características físicas y químicas de los elementos de la tabla periódica.

En base a esta información, el sistema deducirá propiedades de los elementos como por ejemplo el ser o no alcalinos, alcalinotérreos, gases, líquidos, sólidos, halógenos, metales, no metales, metaloides, etc, o bien determinar a que grupos o periodos pertenecen. También podrá realizar combinaciones entre elementos para así obtener fórmulas y los nombres de fórmulas correspondientes.

Finalmente, puedo mencionar que el sistema fue implementado en LISP, corre a una gran velocidad y el espacio total de memoria ocupado por el sistema es de 63.146 bytes.

GRAFICA SOBRE LAS PROPIEDADES DE CADA ELEMENTO DE LA TABLA PERIODICA.



2 - EL PROCESAMIENTO DEL LENGUAJE NATURAL

Como se mencionó anteriormente, el sistema hace uso de un ATN para el procesamiento del lenguaje natural. La definición sintáctica del lenguaje de escritura se presenta a continuación y se da en la Forma Normal de Backus.

Definición sintáctica del lenguaje del ATN

```
<red tran> ::= (<nombre subred> (<subred>))+
<subred> ::= (<arco>)+
<arco> ::= <nombre categoría>+ <arc>* <estado>
           <nombre subred> /
           <nombre categoría>+ <arc>+ [<arco>]
<arc> ::= <estado> <acción> /
           <estado> <paquete regla>
<nombre subred> ::= <identificador>
<nombre categoría> ::= <identificador>
<acción> ::= ARREGLA / COMPARA / CONTROLA / CARGA /
             CARGAO
<paquete regla> ::= FAMILIA1 / FAMILIA2 / FAMILIA3 / FAMILIA4
<estado> ::= 0 / 1
<identificador> ::= <caract. alfabéticos>+ <caract. numéricos>*
```

Se presenta a continuación la definición semántica del lenguaje del ATN.

Definición semántica del lenguaje del ATN

En la primera línea de la definición sintáctica se establece que una <red tran> está constituida por una o más ocurrencias de un <nombre subred>, seguido de una <subred>.

Un <nombre subred> indica el nombre de una <subred> o subgráfica. La <subred> constituye el cuerpo de una subgráfica, y está formada por una o más ocurrencias de un <arco>.

Por medio de los <arco> se especifica qué elementos constituyen el lenguaje de escritura del ATN.

Se establece que un <arco> está constituido por : una o más ocurrencias de un <nombre categoría>, seguido de cero o más ocurrencias de un <arc> y finalmente de un <estado> precediendo a un <nombre subred>.

Otro modo de definir un <arco> es : una o más ocurrencias de un <nombre categoría> seguido de uno o más <arc> y opcionalmente de un <arco>.

Un <arc> se define como un <estado> seguido de una <acción> o bien, de un <estado> seguido de un <paquete regla>. Los <arc> son usados para especificar acciones que deben ejecutarse durante el análisis de un texto, y para especificar el llamado a <paquete regla>.

Un <nombre categoría> indica el nombre de una categoría o clase gramatical. Por medio de los <nombre categoría> se establece cuales clases gramaticales son permitidas para definir el ATN. Las clases gramaticales tratadas en este trabajo son presentadas detalladamente cuando se describe el diccionario del ATN.

Una <acción> indica el nombre de una función. Las acciones se ejecutan durante el análisis de enunciados. Si al ejecutar una <acción> se detecta un error, el proceso de análisis se detiene y se emite el correspondiente mensaje de error. Las acciones a ejecutarse son : ARREGLA, COMPARA, CONTROLA, CARGA, CARGAO. Las mismas se describen detalladamente cuando se presenta la red QUIMICA.

Un <paquete regla> indica el llamado a un paquete de reglas. Los paquetes de reglas son usados como medio de selección de reglas. Cada paquete está formado por varias reglas que comparten ciertas características.

Un <estado> puede definirse como un 0 o un 1. El estado 0 se utiliza para preceder a una <acción>, e indica que la misma debe ejecutarse para poder continuar con el análisis sintáctico y semántico del texto ingresado. También el <estado> 0 se utiliza para preceder a un llamado a un <paquete regla>. El <estado> 1 se utiliza para preceder a un <nombre subred> e indica que el control debe pasarse a la subgráfica especificada por <nombre subred>.

Se presenta a continuación la gramática general del sistema que fue utilizada para tratar oraciones en castellano sobre química. El formato de la misma responde a la definición del ATN presentado anteriormente.

La gramática general

```
QUIMICA  C (PRON_INT1 1 FRA1)
          (PRON_INT2 1 FRA6)
          (VERB_3 0 ARREGLA 1 FRA2)
          (PREP_2 PRON_INT2 SUSCOM_1 PRON_ACUS 0 CONTROLA
           VERB_5 0 COMPARA VERB_6 ARTDET 0 COMPARA
           SUSCOM_1 0 COMPARA X 0 CARGA 0 FAMILIA2)
          (VERB_7 0 ARREGLA SUSCOM_5 PREP_1 X4 0 CARGAO
           PREP_3 ARTDET 0 COMPARA ADJIND 0 COMPARA
           SUSCOM_1 0 COMPARA CONE 1 FRA7)
          (VERB_8 0 ARREGLA ARTDET ADJIND 0 COMPARA
           SUSCOM_1 0 COMPARA PREP_4 X2 0 CARGAO CONEC
```

X2 0 CARGAO CONE 1 FRA7)
 C(VERB_9 0 ARREGLA 1 FRA8)))

 CFRA1 C (VERB_1 0 COMPARA 1 FRA2)))

 CFRA2 C (ARTDET 0 COMPARA 1 FRA3)))

 CFRA3 C (SUSCOM_1 0 COMPARA 1 FRA4)
 CADJCAL_1 0 COMPARA SUSCOM_1 0 COMPARA
 SUSCOM_2 0 COMPARA 0 CONTROLA PREP_1 ARTDET
 0 COMPARA X1 0 CARGAO 0 FAMILIA2)
 CSUSCOM_3 0 COMPARA PRON_REL1 1 FRA5)
 CSUSCOM_4 0 COMPARA PREP_1 SUSCOM_3 0 COMPARA
 PRON_REL1 1 FRA5)))

 CFRA4 C (X2 0 CARGAO 0 FAMILIA1)
 CART_CON1 X3 0 CARGAO 0 FAMILIA1)))

 CFRA5 C (VERB_4 ART_CON2 VERB_6 ARTDET 0 COMPARA
 SUSCOM_1 0 COMPARA X 0 CARGA CONEC X 0
 CARGA 0 FAMILIA3)))

 CFRA6 C (SUSCOM_3 1 FRA5)
 CSUSCOM_4 1 FRA5)))

 CFRA7 C (X 0 CARGA 1 FRA7)
 CNIL 0 FAMILIA2)))

 CFRA8 C (DETIND 0 COMPARA ARTDET 0 COMPARA SUSCOM_6
 0 COMPARA 0 CONTROLA ART_CON1 SUSCOM_1 0
 COMPARA X 0 CARGA 0 FAMILIA4)
 CARTDET 0 COMPARA SUSCOM_6 0 COMPARA X4 0
 CARGAO ART_CON1 SUSCOM_1 0 COMPARA X 0 CARGA
 0 FAMILIA4)))

Se presenta a continuación una descripción de las acciones utilizadas en esta red.

Las acciones utilizadas

ARREGLA : Esta es una acción que obligatoriamente debe ir después de verbos imperativos. La flexibilidad de la gramática permite que estos verbos puedan ingresarse en 2da y 3era persona, sin embargo esto nos puede llevar a rechazar un texto por no concordar con la sintaxis especificada. Por ejemplo en el siguiente enunciado :

" Muestra los elementos alcalinos "

En este enunciado " Muestra " es un verbo en segunda persona y " los " es un artículo en tercera persona del plural. Si luego de estas palabras se aplica una comparación de números, el resultado será negativo ya que los mismos no coinciden. **ARREGLA** tiene como objetivo solucionar estos inconvenientes.

COMPARA : Esta es una acción destinada a comparar género, número y persona de los estados analizados. Si **COMPARA** detecta un error se detiene el análisis del enunciado y se da un aviso de error. Por ejemplo en el siguiente enunciado :

" Cuales son el elementos líquidos ? "

En este enunciado existe un error de concordancia entre el número del verbo (3era. persona del plural) y el número del artículo (3era. persona del singular). **COMPARA** detecta este error y emite el siguiente mensaje :

" Por Favor Revise Género, Número y Persona, Hay un Error en su Oración "

CONTROLA : Esta acción tiene por objetivo evitar conflictos de concordancia de números entre ciertas palabras.

Dada la flexibilidad de la gramática, se puede llegar a tener que comparar, en un determinado momento, el número de una palabra en tercera persona del plural con el número de una palabra en tercera persona del singular..pn 7

Sin embargo, en ciertos casos esto no significa que exista error en la construcción de la oración.

Por ejemplo en el siguiente enunciado :

" Muestre los principales elementos conductores de la electricidad "

En este enunciado, luego de la palabra " conductores " (sustantivo, 3era. persona del plural) sigue la palabra " de " (preposición, neutra). Al existir una palabra con número neutro se ignora, y la comparación se efectúa con la siguiente palabra, en este caso " la " (artículo, 3era. persona del singular). **CONTROLA** evita que el proceso de análisis se interrumpa, ya que aunque no existe concordancia de números entre estas dos palabras, la oración es sintácticamente correcta.

CARGA : Es una función que tiene por objetivo cargar en una pila los elementos de la tabla periódica que se mencionan en el texto ingresado. Esta pila luego será utilizada en el análisis semántico del mencionado texto.

Por ejemplo en el siguiente enunciado :

" Clasifique los siguientes elementos en líquidos y sólidos : mercurio cesio francio oro plata cobre. "

Durante el análisis la acción CARGA irá almacenado en una pila (pila1), los elementos de la tabla periódica que se encontraron en el texto. Al finalizar el análisis pila1 tendrá la siguiente información :

pila1 : (cobre plata oro francio cesio mercurio)

CARGAO : Es una función que tiene por objetivo cargar en una pila las propiedades de los elementos de la tabla periódica que se mencionan en el texto ingresado. Esta pila luego será útil en el análisis semántico del texto ingresado.
Por ejemplo en el siguiente enunciado :

" Clasifique los siguientes elementos en líquidos y sólidos :
mercurio cesio francio oro plata cobre. "

Durante el análisis la acción CARGAO irá almacenado en una pila (pila2), las propiedades de los elementos de la tabla periódica que va encontrando en el texto. al finalizar el análisis pila2 tendrá la siguiente información :

pila2 : (sólidos líquidos).

El Diccionadrio de datos

Si bien el lenguaje utilizado es restringido, la cantidad de verbos, sustantivos, preposiciones, adjetivos, pronombres y articulos utilizado hace necesario dividir estas categorías gramaticales en subcategorías para poder diferenciar el significado q
Se muestran a continuación las categorías y subcategorías gramaticales a las que pertenecen las palabras utilizadas en este trabajo.

PRONOMBRES

<PRON_INT1> ::= Interrogativo.
Cual / Cuales

<PRON_INT2> ::= Interrogativo.
Qué

<PRON_REL1> ::= Relativo.
que

<PRON_ACUS> ::= Personal.
se

VERBOS

<VERB_3> ::= Imperativo
Dé / Diga / Dá / Dí / Muestre / Muestra / Obtén /
Obtenga / Mencione / Menciona / Escribe / Escriba

<VERB_7> ::= Imperativo.
Asigne / Asigna

<VERB_8> ::= Imperativo.
Clasifica / Clasifique / Agrupe / Agrupa.

<VERB_9> ::= Imperativo.
Borre / Borra / Elimine / Elimina

<VERB_1> ::= Indicativo.
son

<VERB_2> ::= Indicativo.
es

<VERB_4> ::= Indicativo.
originan / constituyen / forman

<VERB_5> ::= Indicativo.
puede

<VERB_6> ::= Infinitivo.
combinar / juntar / unir / ligar

SUSTANTIVOS

<SUSCOM_1> ::= Común.
elemento / elementos / átomo / átomos

<SUSCOM_2> ::= Común.
conductor / conductores / transportador
transportadores / transmisor / transmisores

<SUSCOM_3> ::= Común.
fórmula / fórmulas

<SUSCOM_4> ::= Común.
nombre / nombres

<SUSCOM_5> ::= Común.
intervalo / intervalos

<SUSCOM_6> ::= Común.
propiedad / propiedades

PREPOSICIONES

<PREP_1> ::= Separable.
de

<PREP_2> ::= Separable.
Con

<PREP_3> ::= Separable.
a

<PREP_4> ::= Separable.
en

DETERMINANTES

<DETIND> ::= Indefinido.
todo / toda / todos / todas

<ARTDET> ::= Artículo Determinado.
el / la / los / las

<ART_CON1> ::= Artículo Contracto.
del

<ART_CON2> ::= Artículo Contracto.
al

ADJETIVOS

<ADJCAL_1> ::= Calificativo.
principal / principales / mejor / mejores

<ADJIND> ::= Indefinido.
siguiente / siguientes

CONECTIVOS

<CONEC> ::= Conjunción de Coordinación.
y

<CONE> ::= Conjunción de Coordinación.
:

Fue necesario definir además, categorías gramaticales que
tuvieran como instancias a los elementos de la tabla periódica.
El nombre dado a cada una de estas categorías fue elegido
arbitrariamente.
La definición sintáctica y semántica de las mismas se dan a
continuación :

<X> ::= sodio / plata / oro / magnesio / cobre / berilio.....

Un <x> permite tomar como instancias a los nombres de elementos
de la tabla periódica.

<X1> ::= electricidad / calor

Un <X1> permite considerar como instancias a las propiedades
transmisibles de elementos de la tabla periódica.

<X2> ::= líquidos / sólidos / alcalinos / alcalinotérreos.....

Un <X2> toma como instancias atributos de los elementos de la
tabla periódica.

<X3> ::= primer_periodo / grupo_1 / grupo_2b /

Un <X3> se utiliza para denotar propiedades de pertenencia de los
elementos de la tabla periódica, los que necesitan de el artículo
contracto " del " para poder referirlos.

<X4> ::= NA / PE / PF / DEN / PA / EO / SIM / EP / CE / CT

Un <X4> se utiliza para denotar atributos generales de los elementos de la tabla periódica. Hemos considerado a los siguientes :

NA	—————>	Número Atómico
PE	—————>	Punto de Ebullición
PF	—————>	Punto de Fusión
DEN	—————>	Densidad
PA	—————>	Peso Atómico
EO	—————>	Estado de Oxidación
SIM	—————>	Símbolo
EP	—————>	Electronegatividad de Pawling
CE	—————>	Conductividad Eléctrica
CT	—————>	Conductividad Térmica

Se presenta a continuación un ejemplo, para ilustrar al lector como efectúa el ATN los análisis sintáctico y semántico. Consideremos la oración :

Muestre los elementos líquidos.

El análisis se efectúa palabra por palabra. El intérprete del ATN busca en el diccionario de datos las características gramaticales de " Muestre ", que son :

Clase Gramatical :	VERB_3	/	Modo :	Imperativo
Tiempo	: Presente	/	Persona :	3era. singular

Con la clase gramatical <VERB_3> se comienza a recorrer la red de transición. Cuando <VERB_3> se instancia con el primer elemento de un arco, el análisis continúa a través de dicho arco. En este caso, se trata del arco :

(VERB_3 0 ARREGLA 1 FRA2)

El análisis continúa con el <estado> 0, que indica que a su derecha existe una <acción> (ARREGLA) que debe ejecutarse. Esta <acción> permite que la palabra siguiente a " Muestre " pueda estar tanto en 3era. persona del singular, como en 3era. persona del plural. Posteriormente se indica por medio de un <estado> 1, que el control debe pasarse a la <subred> FRA2.

El análisis continúa con la siguiente palabra del enunciado, en este caso " los ", cuyas características gramaticales son :

Clase gramatical :	ARTDET	/	Género :	M (masculino)
Número :	P (plural)	/	Persona :	3era. plural

Con <ARTDET> se comienza a recorrer la <subred> FRA2 hasta detectar el <arco> que contiene esta <categoría gramatical>. En este caso se trata del arco :

(ARTDET 0 COMPARA 1 FRA3)

El análisis continúa con el <estado> 0, que indica que a su derecha existe una acción que debe ejecutarse : COMPARA. La parte final del arco indica, por medio del <estado> 1, que el control debe pasarse a la subred FRA3.

Se analiza a continuación la siguiente palabra del enunciado, en este caso : elementos. Sus características gramaticales son :

Clase gramatical : SUSCOM_1	/	Género : M (masculino)
Número : P (plural)	/	Persona : 3era. plural

Con <SUSCOM_1> se comienza a recorrer la <subred> FRA3. El <arco> que comienza con esta <categoria gramatical> es el siguiente :

(SUSCOM_1 O COMPARA 1 FRA4)

El análisis continúa con el <estado> 0 y la <acción> COMPARA. Posteriormente se detecta el <estado> 1 y se transfiere el control a la <subred> FRA4.

Se analiza la siguiente palabra del enunciado : líquidos. Sus características gramaticales son :

Clase gramatical : X	/	Género : Masculino
Número : P (plural)	/	Persona : 3era. plural

Con <X2> se recorre la <subred> FRA4. El <arco> que comienza con la <categoria gramatical> es el siguiente.

(X2 O CARGAO O FAMILIA1)

El análisis continúa con el <estado> 0 y la <acción> CARGAO. Esta <acción> tiene el objetivo de cargar la palabra " líquidos " en una variable de trabajo. Posteriormente el análisis continúa con el <estado> 0 y se detecta un llamado a un <paquete regla>. En este paquete se hace uso de la variable de trabajo y según su contenido se decidirá que regla debe de aplicarse. Para este ejemplo en particular se aplica la regla REGLA4.

REGLA4 nos permite determinar qué elementos de la tabla periódica son líquidos.

Una vez que se ejecuta esta regla, el sistema presenta los resultados obtenidos.

Para que finalmente el sistema pueda proporcionar resultados, el enunciado debe estar bien formado. Se considera bien formado, si durante los análisis sintáctico y semántico no se detectan errores y se llega finalmente al llamado de un <paquete regla>.

3 - EL MECANISMO DE INFERENCIA DEL SISTEMA

La parte medular en el sistema es el mecanismo de inferencia del mismo. Este opera con : una memoria de trabajo, una base de conocimientos, una base de hechos y una máquina de inferencia que aplica las reglas pertinentes a la resolución del problema en curso.

La memoria de trabajo contiene la información (datos o conclusiones) inferida hasta el momento.

La base de conocimientos está constituida por el conocimiento específico y procedimental acerca de la clase de problemas en los que el sistema es experto.

La máquina de inferencia es el conjunto de programas y estructuras de control que hacen deducciones a partir de la selección, interpretación y aplicación de las reglas a la información presente en la memoria de trabajo.

Para que una regla pueda aplicarse es necesario que parte del contenido de la memoria de trabajo coincida (sea igual o pueda hacerse igual mediante sustitución) con parte de la regla. En el sistema existen dos formas básicas de interpretación de las reglas : el encadenamiento hacia adelante y el encadenamiento hacia atrás.

El Lenguaje de Escritura de las Reglas

A continuación se presenta la sintaxis del lenguaje de escritura de reglas especificado en la Forma Normal de Backus.

```

<paquete regla> ::= <regla>+

<regla> ::= (DEFUN <nombre regla> ( <arg1> [<arg2>] )
      <cuerpo regla>

<cuerpo regla> ::= (( <oper> <premisa>+ ) <alternativa1>
      <alternativa2> /
      ( <predicado> <alternativa1> <alternativa2>

<premisa> ::= ( <oper> <predicado> <predicado>+ ) /
      <predicado> <predicado>+

<alternativa1> ::= <cuerpo regla> / <acción> )

<alternativa2> ::= <cuerpo regla> / <acción> )

<oper> ::= AND / OR

<predicado> ::= ( MIEMBRO <op1> <op2> ) /
      ( IGUAL <op1> <op2> ) /
      ( MAYOR <op1> <op2> ) /
      ( MENOR <op1> <op2> ) /
      ( NEGATIVO <op1> <op2> ) /
      ( POSITIVO <op1> <op2> )

<acción> ::= ( REGRESA <nombre regla> ) /
      ( AGREGA <nombre regla> ) /
      ( ALMA_MIN <nombre regla> ) /
      ( ALMA_MAX <nombre regla> )

<nombre regla> ::= REGLA1 / REGLA4 / REGLA5 /...../REGLAn

<op1> ::= NA / PE / PF / DEN / PA / EO / SIM / EP /
      CE / CT / <nombre variable>

<op2> ::= <números reales> / <nombre variable>

<nombre variable> ::= <identificador>

<identificador> ::= <caracteres alfabéticos>+

```

<arg1> ::= LST
<arg2> ::= SEN / ATOMO

Definición Semántica del Lenguaje

En la primera línea de la definición sintáctica, se establece que un <paquete regla> está formado por una o más reglas. Una <regla> está constituida por una palabra terminal DEFUN, un <nombre regla>, un <arg1> seguido opcionalmente de un <arg2> y finalmente de un <cuerpo regla>. De esta manera se establece como debe formarse una regla, de tal manera que la misma pueda ser reconocida por el intérprete.

En el <cuerpo regla> se dan las especificaciones que establecen qué hace la regla. Es aquí donde el usuario, por medio de predicados y operadores, establece el objetivo de la regla.

Un <cuerpo regla> puede estar constituido de dos maneras. La primera establece que un <cuerpo regla> está formado por dos paréntesis abierto, un <oper>, una o más ocurrencias de una <premisa>, un paréntesis cerrado, una <alternativa1> y una <alternativa2>.

La segunda alternativa establece que un <cuerpo regla> está formado por un paréntesis abierto, un <predicado>, una <alternativa1> y una <alternativa2>.

Una <premisa> está formada por un <oper>, un <predicado>, seguido de una o más ocurrencias de otro <predicado>. También una <premisa> puede estar formada por un <predicado> seguida de una o más ocurrencias de otro <predicado>.

Una <alternativa1> está formada por un <cuerpo regla>, o bien por una <acción> seguida de un paréntesis cerrado.

Una <alternativa2> está formada por un <cuerpo regla>, o bien por una <acción> seguida de un paréntesis cerrado.

La <alternativa1> se ejecuta cuando en el cuerpo de la regla resulta verdadero el cumplimiento de las premisas o predicados, en caso contrario se ejecuta la <alternativa2>.

Un <oper> está formado por un símbolo terminal que indica un operador del tipo lógico, como AND (conjunción) y OR (disyunción).

Un <predicado> está formado por un nombre nemotécnico que indica la relación existente entre los argumentos <op1> y <op2>. Los predicados definidos por las reglas tratadas en este sistema son: MIEMBRO, IGUAL, MAYOR, MENOR, NEGATIVO y POSITIVO.

MIEMBRO : Se utiliza para verificar si <op1> pertenece a <op2>. Cuando una ocurrencia de <op1> aparece en <op2>, el valor de verdad de este predicado es TRUE. En caso contrario es FALSE.

IGUAL : Se utiliza para comparar <op1> con <op2>. Si los operandos son iguales, entonces el valor de verdad del predicado es TRUE. En caso contrario es FALSE.

MAYOR : Se utiliza para comparar <op1> con <op2>. Si <op1> es mayor que <op2>, entonces el valor del predicado es TRUE. En caso contrario es FALSE.

MENOR : Se utiliza para comparar <op1> con <op2>. Si <op1> es menor que <op2>, entonces el valor de verdad del predicado es TRUE. En caso contrario es FALSE.

NEGATIVO : Se utiliza para verificar si <op1> tiene en <op2> algún valor negativo. En caso afirmativo, el valor de verdad del predicado es TRUE. En caso contrario es FALSE.

POSITIVO : Se utiliza para verificar si <op1> tiene en <op2> algún valor positivo. En caso afirmativo, el valor de verdad del predicado es TRUE. En caso contrario es FALSE.

Una <acción> define una función. Las funciones definidas para las reglas tratadas por este sistema son : REGRESA, AGREGA, INICIA, ALMA_MIN, ALMA_MAX.

REGRESA : Se usa para provocar la bifurcación del control del sistema hacia la regla especificada por <nombre regla>. Cuando <nombre regla> especifica la regla en la cual aparece REGRESA, se produce el llamado recursivo a la regla.

AGREGA : Se usa para agregar un elemento a una lista de resultados. Luego pasa el control del sistema a la regla especificada por <nombre regla>. Implícitamente ejecuta un REGRESA.

ALMA_MIN : Se usa para asignar un valor mínimo a una variable. Luego se transfiere el control del sistema a la regla especificada por <nombre regla>.

ALMA_MAX : Se usa para asignar un valor máximo a una variable. Luego se transfiere el control del sistema a la regla especificada por <nombre regla>.

Un <op1> está formado por símbolos terminales, como NA (número atómico), PE (punto de ebullición), DEN (densidad) o por un <nombre variable>.

Un <op2> está formado por cualquier número real o por un <nombre variable>.

Un <arg1> se define como una LST. Una LST es el nombre de una variable, que recibirá como valor una lista de elementos sobre los cuales se quiere aplicar una regla.

Un <arg2> se define como SEN o ATOMO. La variable SEN se utiliza como bandera. Un ATOMO es el nombre de una variable, que recibirá como valor un elemento de la tabla periódica sobre el cual se quiere aplicar una regla.

Los siguientes son ejemplos de reglas de inferencia del sistema, definidas siguiendo la sintaxis del lenguaje de escritura de reglas presentado anteriormente. Con el objeto de que el lector entienda el significado de las mismas precede a cada una de ellas un texto en lenguaje natural.

Son líquidos, los elementos de la tabla periódica, cuyo punto de fusión (PPF) está comprendido entre los valores 234.279 y 302.901

```

(DEFUN REGLA4 (LST)
  ((AND
    (MAYOR PF 234.279)
    (MENOR PF 302.901) ) (AGREGA REGLA4) )
  (REGRESA REGLA4) )

```

Son gaseosos, los elementos de la tabla periódica, cuyo punto de ebullición (PE) está comprendido entre los valores 3.99 y 240.

```

(DEFUN REGLAS (LST)
  ((AND
    (MAYOR PE 3.99)
    (MENOR PE 240.01) ) (AGREGA REGLAS) )
  (REGRESA REGLAS) )

```

4 - RESULTADOS OBTENIDOS POR EL SISTEMA

Los siguientes son algunos ejemplos de resultados obtenidos por el sistema.

Cabe aclarar que los textos que ingresa el usuario estarán en negritas, precedidas por los símbolos U>, en tanto que las respuestas del sistema en tipo estándar, precedidas por los símbolos S>.

Errores que detecta el Analizador Sintáctico

U>
(Cuales son los elemento gaseosos ?)

S>
POR FAVOR VERIFIQUE GENERO, NUMERO Y PERSONA, HAY UN ERROR EN SU ORACION.

NOTA : La palabra " elemento " está en singular y debe estar en plural.

U>
(Cuales son elementos gaseosos ?)

S>
LA ORACION INGRESADA NO CONCUERDA CON LA GRAMATICA, POR FAVOR REVISE SU ORACION.

Nota : La oración ingresada no concuerda con la gramática, ya que la misma exige que después del verbo " son " debe ir un artículo : los.

U>
(Muestre el principal elemento conductor de la electricidad.)

S>
FALTA LA PALABRA SOLICITADA Y UD. HA DECIDIDO NO INGRESARLA

NOTA : Esto ocurre cuando el analizador detecta una palabra desconocida en el diccionario de datos, le dice al usuario que la ingrese y éste decide no hacerlo. Supongamos que " el " es una palabra que no se encuentra en el diccionario de datos, entonces el sistema despliega en la pantalla, el siguiente mensaje :

DESEA INGRESAR LA PALABRA SOLICITADA " el " : (S / N) ?

Si el usuario elige la opción N, el sistema muestra el mensaje presentado anteriormente.

Si el usuario elige la opción S, el sistema le pedirá las características gramaticales de la palabra desconocida. Luego incorporará la nueva palabra y sus características gramaticales al diccionario de datos.

Con esta opción, los análisis sintáctico y semántico sobre el enunciado no se interrumpen y el diccionario aumenta su volumen.

Acciones que modifican la Base de Hechos del Sistema

U>

(Borre el elemento oro.)

S>

O.K.

NOTA : Cuando se borra o elimina un elemento, se borran también todas sus propiedades.

U>

(Elimina todas las propiedades del elemento plata.)

S>

O.K

U>

(Borra la propiedad PF del elemento magnesio.)

S>

O.K

U>

CARGON NA 18 PE 87\$30 PF 83\$81 DEN 1\$784 PA 39\$948 SIM Ar CT
O\$0001772 NIL) <RETURN>

S>

O.K

NOTA : Esta declaración se utiliza para incrementar la base de hechos del sívces lo desee el usuario.

Información que obtiene el usuario por medio de consultas

U>
(Muestre los elementos alcalinos)

S>

1 - Litio	2 - Sodio
3 - Potasio	4 - Rubidio
5 - Cesio	6 - Francio

U>
(Cuales son los elementos alcalinotwreos ?)

S>

1 - Berilio	2 - Magnesio
3 - Calcio	4 - Estroncio
5 - Bario	6 - Radio

U>
(Mencione los elementos lhquidos.)

S>

1 - Cesio	2 - Francio
3 - Mercurio	4 - Galio
5 - Bromo	

U>
(Obtenga los elementos metaloides.)

S>

1 - Boro	2 - Silicio
3 - Germanio	4 - Antimonio
5 - Bismuto	6 - Telurio
7 - Polonio	8 -Mercurio

NOTA: Es de destacar que las reglas fueron escritas en función de ciertos conceptos típicos de química, en los cuales me he basado. En este caso se establece que un elemento es metaloide si su electronegatividad está comprendida entre los valores 1.8 y 2.1 y su conductividad eléctrica es menor que 0.03. Como respuesta a esta orden aparece el elemento mercurio, ya que su electronegatividad es igual a 2 y su conductividad eléctrica es igual 0.0104, sin embargo este elemento no es un metaloide.

U>
(Cual es el principal elemento conductor de la electricidad ?)

S>

El principal elemento es	: plata
Su conductividad eléctrica es	: 0.630 10**6/ohm*cm

NOTA : Es de destacar que este resultado surge de la aplicación de una regla dinámica, es decir que identifica a " principal " como máximo y luego se busca el elemento cuya conductividad eléctrica sea más alta. Si por ejemplo borramos el elemento plata y nuevamente aplicamos la pregunta, la respuesta será :

El principal elemento es : cobre
Su conductividad eléctrica es : $0.596 \cdot 10^{16} / \text{ohm} \cdot \text{cm}$

U>
(Muestre los principales elementos conductores de el calor.)

S>

Elemento	Conductividad en $10^6 / \text{ohm} \cdot \text{cm}$.
1 - plata	----> 4.19
2 - cobre	----> 4.01
3 - oro	----> 3.17
4 - aluminio	----> 2.37
5 - berilio	----> 2.00
6 - calcio	----> 2.00
7 - tungsteno	----> 1.74
8 - manganeso	----> 1.56
9 - rodio	----> 1.50
10 - silicio	----> 1.48

NOTA : Es de destacar que en este caso se asocia la palabra " principales " con una lista de elementos que deben cumplir una determinada condición.

U>
(Clasifique los siguientes elementos en alcalinos y alcalinotérreos : litio berilio calcio rubidio cesio oro plata bario.)

S>

Alcalinos	:	litio	-	rubidio	-	cesio
Alcalinotérreos	:	berilio	-	calcio	-	bario

NOTA : En este caso estamos en presencia de una meta_regla. Es decir que luego de analizar sintácticamente la orden hacemos un reordenamiento de la información ingresada.
En este caso la meta_regla actúa como una vía de selección para el conflicto de resolución, detectando y aplicando la regla pertinente.

U>
(Asigne intervalos de PE a los siguientes elementos : escandio titanio vanadio cromo manganeso hierro cobalto níquel cobre cinc.)

S>

El valor inferior es	:	cinc	---->	1180 K.
El valor superior es	:	vanadio	---->	3682 K.

NOTA : Este tipo de órdenes son muy útiles cuando queremos establecer los valores mínimos y máximos de una determinada propiedad. Por medio de estas órdenes podemos establecer cuales son los principales y cuales los menores conductores de calor y electricidad, por ejemplo.

U>
(Con qué elementos se puede combinar el elemento sodio ?)

S>

1 - carbono	2 - nitrógeno
3 - oxígeno	4 - flúor
5 - fósforo	6 - azufre
7 - cloro	8 - arsénico
9 - selenio	10 - bromo
11 - antimonio	12 - telurio
13 - yodo	14 - renio
15 - astato	

NOTA : Para este tipo de cuestionamientos se da como respuesta todos los elementos que satisfacen las hipótesis de la regla aplicada. Cabe mencionar que no importa la probabilidad con que estas combinaciones puedan ocurrir. Así por ejemplo tenemos que el flúor con el sodio tienen una probabilidad del 90 % de ocurrir la ligación, en tanto que el renio con el sodio tienen una probabilidad del 19 % de ocurrir la ligación. Estas probabilidades tendrán una mayor credibilidad cuando se trate de elementos con un determinado carácter iónico.

U>
(Muestre las fórmulas que originan al combinarse los elementos cloro y sodio.)

S>

Fórmula	:	ClNa	ClNa
Probabilidad	:	70 %	

U>
(Muestre los nombres de fórmulas que originan al combinarse los elementos flúor y manganeso.)

S>

Nombres	:	Fluoruro de Manganeso VII
		Fluoruro de Manganeso VI
		Fluoruro de Manganeso II
		Fluoruro de Manganeso IV
		Fluoruro de Manganeso III

Probabilidad : 76 %

NOTA : Cabe mencionar que se da esta notación por ser la moderna y la más comunmente utilizada por los químicos en la actualidad.

5 - CONCLUSIONES

1- En este trabajo los datos se encuentran independientes del programa. Esta característica se da en las dos fases del sistema.

En la primera fase (procesamiento de lenguaje natural), el intérprete está separado de la gramática y del diccionario. En la segunda fase (mecanismos de inferencia), la máquina de inferencia está separada de la base de conocimientos y de la base de hechos.

Si se cambia la gramática, la base de conocimientos y la base de hechos del sistema, el mismo puede operar en otras áreas del saber.

- 2- El procesador de lenguaje natural (ATN) es muy general y la gramática utilizada para tratar textos de química es muy diversa, lo que permite al usuario emplear diferentes modos de expresar sus enunciados.
- 3- La interfase de lenguaje natural facilita la interacción del hombre con la máquina. Esta interfase hace más amigable al sistema. Este guía al usuario por medio de menús y mensajes.
- 4- El sistema puede arrancar sin ningún dato en el diccionario y en la base de hechos, y a medida que opera puede ir incorporando información. Por lo tanto, se puede afirmar que es dinámico.
- 5- El sistema está capacitado para aceptar nuevas reglas de inferencia. El usuario, haciendo uso del lenguaje de escritura de reglas, puede definirlas de acuerdo a sus necesidades.

BIBLIOGRAFIA

- [1] Cairó, Osvaldo. " Un Sistema Basado en Conocimientos ". Tesis de Maestría. Computación. Dpto de Ing. Eléctrica. CINVESTAV - IPN, México, 1987.
- [2] Winograd, Terry. " Understanding Natural Language ". Academic press, New York, 1972.