Actividad 06

MAESTRO:

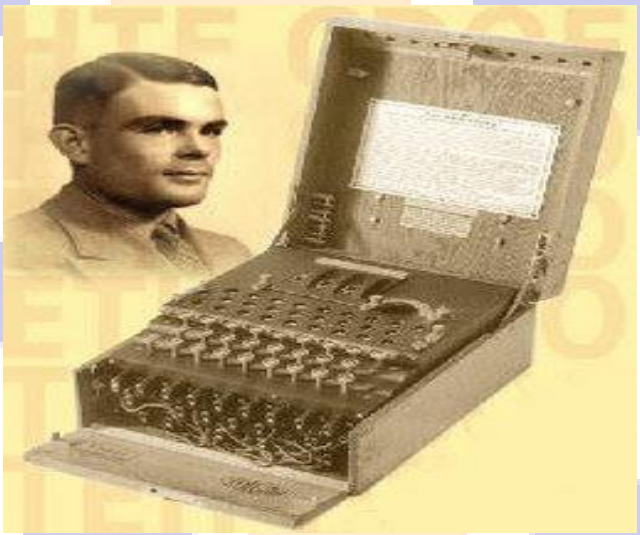
Abelardo Gómez Andrade

ALUMNO:

Rodríguez Tabares Juan

CODIGO:

215615699



CARRERA:

Ingeniería en Computación

MATERIA:

Teoría de la computación

HORARIO:

Martes y jueves

11:00 – 13:00

SECCION:

D07

Lenguajes de Programación

Un Lenguaje Formal, surge a partir de su gramática, y por lo tanto no presenta excepciones en su definición. Esto es así, porque los Lenguajes Formales son los que se utilizan para que se comuniquen los hombres con las máquinas. De esta manera a partir de las gramáticas formales es como surgen los Lenguajes de Programación.

Lex, o Flex en una implementación más reciente, que nos permite especificar un analizador léxico mediante la especificación de expresiones regulares para describir patrones de los tokens. La notación de entrada para la herramienta Lex se conoce como el lenguaje Lex, y la herramienta en sí es el compilador Lex. El compilador Lex transforma los patrones de entrada en un diagrama de transición y genera código, en un archivo llamado lex.yy.c, que simula este diagrama de transición. La mecánica de cómo ocurre esta traducción de expresiones regulares a diagramas de transición es el tema de las siguientes secciones; aquí sólo aprenderemos acerca del lenguaje Lex.

Gramáticas para describir Lenguajes Naturales

Para poder describir a los Lenguajes Naturales, podemos escribir reglas como las que enumeramos en el siguiente conjunto:

P_n = {<oración> → <sujeto> <predicado>,
<sujeto> → <sustantivo> | <artículo> <sustantivo> | <artículo> <sustantivo> <adjetivo>,
<predicado> → <verbo> | <verbo> <modificador>,
<artículo> → el | la | los | las | un,
<sustantivo> → niño | madre | árbol | vehículos,
<adjetivo> → pequeño | alto | joven,
<verbo> → aprende | corre | juega | crece,
<modificador> → velozmente | risueño | temprano}

Procesamiento de Lenguajes Naturales

El Procesamiento del Lenguaje Natural es el campo de conocimiento de la Inteligencia Artificial que se ocupa de la investigar la manera de comunicar las máquinas con las personas mediante el uso de lenguas naturales, como el español, el inglés o el chino.

Virtualmente, cualquier lengua humana puede ser tratada por los ordenadores.

Lógicamente, limitaciones de interés económico o práctico hace que solo las lenguas más habladas o utilizadas en el mundo digital tengan aplicaciones en uso.

Pensemos en cuántas lenguas hablan Siri o Google Assistant. El inglés, español, alemán, francés, portugués, chino, árabe y japonés (no necesariamente en este orden) son las que cuentan con más aplicaciones que las entienden. Google Translate es la que más lenguas trata, superando el centenar… pero hay entre 5000 y 7000 lenguas en el mundo.

Modelos para procesamiento del lenguaje natural

Tratar computacionalmente una lengua implica un proceso de modelización matemática. Los ordenadores solo entienden de bytes y dígitos y los informáticos codifican los programas empleando lenguajes de programación como C, Python o Java.

Los lingüistas computacionales se encargan de la tarea de “preparar” el modelo lingüístico para que los ingenieros informáticos lo implementen en un código eficiente y funcional. Básicamente, existen dos aproximaciones generales al problema de la modelización lingüística:

Modelos Lógicos: gramáticas

Los lingüistas escriben reglas de reconocimiento de patrones estructurales, empleando un formalismo gramatical concreto. Estas reglas, en combinación con la información almacenada en diccionarios computacionales, definen los patrones que hay que reconocer para resolver la tarea (buscar información, traducir, etc.).

Modelos probabilísticos del lenguaje natural: basados en datos

La aproximación es a la inversa: los lingüistas recogen colecciones de ejemplos y datos

(corpus) y a partir de ellos se calculan las frecuencias de diferentes unidades lingüísticas (letras, palabras, oraciones) y su probabilidad de aparecer en un contexto determinado.

Calculando esta probabilidad, se puede predecir cuál será la siguiente unidad en un contexto dado, sin necesidad de recurrir a reglas gramaticales explícitas.

Componentes del procesamiento del lenguaje natural

Análisis morfológico o léxico. Consiste en el análisis interno de las palabras que forman oraciones para extraer lemas, rasgos flexivos, unidades léxica compuestas. Es esencial para la información básica: categoría sintáctica y significado léxico.

Análisis sintáctico. Consiste en el análisis de la estructura de las oraciones de acuerdo con el modelo gramatical empleado (lógico o estadístico).

Análisis semántico. Proporciona la interpretación de las oraciones, una vez eliminadas las ambigüedades morfosintácticas.

Análisis pragmático. Incorpora el análisis del contexto de uso a la interpretación final. Aquí se incluye el tratamiento del lenguaje figurado (metáfora e ironía) como el conocimiento del mundo específico necesario para entender un texto especializado.

Expresiones Regulares

Una expresión regular es un modelo o una forma de comparar con una cadena de caracteres. Esta comparación es conocida con el nombre de pattern matching o reconocimiento de patrones, permite identificar las ocurrencias del modelo en los datos tratados. La utilización principal de las expresiones regulares en Perl consiste en la identificación de cadenas de caracteres para la búsqueda modificación y extracción de palabras clave. De esta manera se pueden dividir las expresiones regulares en varios tipos que son: expresiones regulares de sustitución, expresiones regulares de comparación y expresiones regulares de traducción.

Expresiones regulares de comparación.

Nos permiten evaluar si un patrón de búsqueda se encuentra en una cadena de caracteres, de modo que mediante este tipo de expresiones regulares obtendremos un valor lógico verdadero o falso según se encuentre el patrón deseado. La sintaxis de este tipo de expresiones regulares es la siguiente:

valor por comparar =~ patrón de búsqueda

Ejemplo:

```
# imprimimos todas las líneas que contengan "html". Ej.:

"htmlexe" if ($linea =~ /html/) {

     print $linea;

}
```

Los patrones de búsqueda pueden integrar información relativa al contexto, tal como la búsqueda de líneas que empiecen por la cadena de caracteres, la extracción de palabras que tengan prefijos o sufijos particulares

- Patrones definidos por una clase de caracteres.

  A menudo resulta práctico extraer las palabras que contienen una cifra, una vocal, o caracteres de control particulares. El modelo así definido no se indica por un carácter particular sino por un clase de caracteres mediante el operador [ ]. He aquí algunas posibles construcciones:

```
[aeiou]          # Cualquier vocal
[0-9]            # Cualquier número del 0 al 9.
[0123456789]     # Igual [0-9]
[0-9a-z]          # Cualquier letra o cualquier
numéro  [\~\@;:\^_]         #  Cualquiera  de  los
caracteres(~,@,;,:^,_)
```

  Referencia de patrones.

  Se utilizan para referenciar patrones en las expresiones regulares. Perl trabaja con dos tipos de operadores de comparación:

  - $1,$2,...,$9. Sirven para referenciar uno de los patrones de búsqueda de la expresión regular.
  - \1,\2,...,\9. Este operador tiene la misma utilidad que el anterior se utiliza para referenciar patrones.

- Utilización de caracteres reservados.

  En la especificación del modelo, cada carácter se interpreta para determinar las ocurrencias en los datos. Sin embargo, los caracteres siguientes:

  + ? . * ^ $ ( ) [ ] { } | & \

- Combinación de expresiones regulares.

  Se realiza con los operadores | y & que equivalen al or y al and lógico respectivamente. Por ejemplo, con el operador | podemos representar una lista de alternativas, es decir:

```
if ($car =~ /ford | audi/)
{    print $car;
```

Aplicación adicional

- Extracción de información: se trata de un tipo de recuperación de información, cuyo objetivo es extraer automáticamente información estructurada (o semiestructurada) desde documentos legibles por un ordenador.
- Respuesta a preguntas: es uno de los sistemas más complejos de recuperación de la información y por el cual una máquina debería ser capaz de recuperar respuestas planteadas en lenguaje natural. Estos buscadores de respuestas intentan reconocer un amplio rango de tipos de preguntas,

incluyendo “cómo”, “cuándo”, “dónde”, “por qué”, hechos, listas, definiciones, etc.

- Traducción automática: (en inglés, Machine Translation) es un área de la lingüística computacional que investiga el uso de un sistema capaz de traducir texto o habla de un lenguaje natural a otro.
- Síntesis de voz: o también Síntesis del Habla, es la creación de habla artificial. Los sintetizadores de voz, que son los sistemas utilizados para este propósito, pueden ser utilizados tanto en productos softwares como hardware.
- Comprensión del lenguaje: es el campo de la ciencia que estudia la manera en la que los seres humanos utilizamos los símbolos para expresar ideas y sentimientos y cómo nuestro cerebro procesa y entiende esa comunicación.
- Reconocimiento del habla: es una disciplina de la Inteligencia Artificial que estudia cómo permitir la comunicación entre personas y máquinas.
- Generación de lenguajes naturales: como el proceso de creación de un texto en lenguaje natural para la comunicación con fines específicos.

Bibliografia

http://www2.iib.uam.es/bioinfo/curso/perl/tutoriales/cicei/cap6.html

https://www.iic.uam.es/inteligencia/que-es-procesamiento-del-lenguaje-natural/

www2.iib.uam.es/bioinfo/curso/perl/tutoriales/cicei/cap6.html