

ANÁLISIS Y SIMPLIFICACIÓN DE ORACIONES

Práctica 2: Conocimiento y Razonamiento Automatizado

2022-2023

1. Objetivos

El objetivo de la práctica es:

- Crear una Gramática de Cláusulas Definidas que permita analizar sintácticamente una oración en Inglés y crear el árbol de constituyentes.
- Simplificar oraciones complejas convirtiéndolas en oraciones simples.

Debido a la complejidad de la gramática de la lengua inglesa se trabajará con una versión reducida de la misma que permita analizar un conjunto limitado de frases.

2. Práctica a entregar

Abordar la creación de una gramática que valide y represente el árbol de constituyentes y simplifique oraciones complejas de los siguientes tipos (entre paréntesis):

- Oración Simple (o).
- Oración Coordinada (oc).
- Oración Subordinada de Relativo (or).
- Oraciones Compuestas, es decir combinaciones de oraciones simples, coordinadas y de relativo (ocm).

En cuanto a los grupos sintácticos, se deben incluir reglas que soporten los siguientes tipos:

- Grupo Nominal (gn).
- Grupo Adjetival (gadj).
- Grupo Adverbial (gadv).
- Grupo Preposicional (gp).
- Grupo Verbal (gv).

Lo que implica la definición de los siguientes tipos de términos:

- Determinantes (det).
- Nombres (n).
- Nombres propios (np).
- Verbos (v).
- Adjetivos (adj).
- Adverbios (adv).
- Conjunciones (conj).
- Preposiciones (prep).

La simplificación de oraciones coordinadas, subordinadas de relativo o compuestas consistirá en la conversión de cualquiera de estas oraciones en tantas oraciones simples como se componga la original. A continuación, se muestran algunos ejemplos:

- "JOSÉ estudia Filosofía, pero MARÍA estudia Derecho" se debe descomponer en: "JOSÉ estudia Filosofía y MARÍA estudia Derecho".
- "JOSÉ come y bebe mientras MARÍA lee" se debe descomponer en "JOSÉ come", "JOSÉ bebe" y "MARÍA lee".
- "JOSÉ, que es muy alto, tiene el pelo rubio" se debe descomponer en "JOSÉ es muy alto" y "JOSÉ tiene el pelo rubio".
- "JOSÉ, que es muy alto, tiene el pelo rubio, pero MARÍA es morena" se debe descomponer en "JOSÉ es muy alto", "JOSÉ tiene el pelo rubio" y "MARÍA es morena".

El código de la práctica se entregará en dos ficheros: el programa propiamente dicho y la utilidad **draw.pl** que deberá ser consultada por el primero para representar gráficamente los árboles de constituyentes de las oraciones analizadas y simplificadas.

3. Conjunto de oraciones

Además, la gramática debe analizar y simplificar correctamente las siguientes oraciones tipo:

1. JOSÉ es moreno y MARÍA es alta.
2. JOSÉ estudia Filosofía, pero MARÍA estudia Derecho.
3. MARÍA toma un café mientras JOSÉ recoge la mesa.
4. JOSÉ toma café y lee el periódico.
5. JOSÉ y HECTOR comen patatas fritas y beben cerveza.
6. JOSÉ come patatas fritas, pero MARÍA prefiere paella, aunque HECTOR toma café e IRENE lee una novela.
7. IRENE canta y salta mientras JOSÉ estudia.
8. HECTOR come patatas fritas y bebe zumo mientras JOSÉ canta y salta, aunque MARÍA lee una novela.
9. JOSÉ, que es ágil, escala en el rocódromo por las tardes.
10. JOSÉ, que es muy delicado, come solamente manzanas rojas.
11. El procesador de textos, que es una herramienta bastante potente, sirve para escribir documentos.
12. El procesador de textos es una herramienta muy potente que sirve para escribir documentos, pero es bastante lento.
13. El ratón que cazó el gato era gris.
14. El hombre que vimos ayer era mi vecino.

Se debe construir un conjunto de oraciones en Inglés que constituya un conjunto de prueba inicial para el sistema, que incluya al menos las correspondientes a las anteriormente indicadas en Español.

Para refrescar los conceptos de análisis sintáctico (tipos de sintagmas, tipos de oración, funciones de los sintagmas, ejemplos resueltos,...) se recomienda como referencia inicial la página: <https://syntaxis.org>. El analizador que incluye no siempre es capaz de analizar las oraciones propuestas en este enunciado, pero puede servir para resolver los ejemplos de la propia página web.

4. Mejoras del programa

Se valorará positivamente cualquier mejora de la gramática sobre lo expuesto anteriormente. Algunas ideas son las siguientes:

- Introducir algún preprocesamiento de la oración que agilice el proceso de análisis de una estructura sintáctica compleja.
- Implementar algún mecanismo para flexionar las palabras del diccionario, añadiendo el plural a las palabras en singular o el femenino para las palabras de género masculino (o viceversa).
- Asignar funciones a cada uno de los grupos gramaticales indicando si funcionan como Complemento Directo, C. Indirecto, Atributo, C. Circunstancial, etc.
- Construir un traductor básico español-inglés / inglés-español inspirado en el programa **traductor.pl**

- Cualquier otra que cuente con el visto bueno del profesor.

5. Detalles de la entrega

La práctica se DEBERÁ realizar en equipos de 3 personas. Además del código fuente del programa (.pl), convenientemente estructurado y comentado, se deberá incluir una breve memoria (PDF – extensión orientativa 15-20 pgs.) indicando los miembros del equipo, reparto de tareas entre los mismos, conjunto de frases analizadas/simplificadas con éxito, mejoras realizadas, errores y/o aspectos no implementados, así como una bibliografía con las fuentes consultadas para la resolución de la práctica. La entrega se realizará a través de la plataforma en un único **fichero .zip**