

# Práctica 9

## Procesamiento de Lenguaje Natural

Fecha de entrega: 10 de junio de 2018

### Parte 1

Se quiere construir un sistema de enseñanza del español como lengua extranjera, utilizando DCGs en SWI Prolog. Uno de sus módulos necesita poder convertir frases expresadas en estilo directo a estilo indirecto y viceversa. Como ejemplo del procesamiento a realizar nos han proporcionado las siguientes frases:

Estilo directo: María me dijo: “Juan es mi amigo”

Estilo indirecto: María me dijo **que** Juan **era su** amigo

Estilo directo: Miguel me dijo: “Estoy contento de verte”

Estilo indirecto: Miguel me dijo **que estaba** contento de verme

Estilo directo: Lucía me dijo: “Necesito un cambio en mi vida”

Estilo indirecto: Lucía me dijo **que necesitaba** un cambio en **su** vida

Estilo directo: Luis me preguntó: “¿Estás ocupada esta noche?”

Estilo indirecto: Luis me preguntó **que si estaba** ocupada **esa** noche

El módulo leerá una frase introducida como lista de palabras en uno de los dos estilos (que tendrá que averiguar), analizará su corrección (concordancias de género, número y persona) y producirá como salida la frase equivalente en el estilo contrario. Estará continuamente leyendo frases de entrada y produciendo las frases de salida correspondientes hasta que lea una frase vacía y entonces parará.

Se valorará la escalabilidad del módulo desarrollado para permitir la transformación de otras frases diferentes aunque del mismo tipo.

### Parte 2

Elige 6 secciones habituales de los periódicos (Internacional, Deportes, Economía,...). Busca textos breves en inglés, correspondientes a noticias publicadas en periódicos digitales y etiquétalos con la sección en la que tendrían que aparecer. Introdúcelos en Weka y utiliza el filtro StringToWordVector para convertirlos en vectores de palabras, con los valores de los parámetros que consideres más adecuados. Selecciona dos terceras partes de estos textos para entrenar un clasificador y la otra tercera parte como conjunto de validación. Valora los

resultados obtenidos al clasificar los textos del conjunto de validación con diferentes clasificadores y parámetros. Intenta explicar por qué se obtienen esos resultados. A continuación, prueba con herramientas de clustering aplicadas a la totalidad de los textos (conjunto de entrenamiento + conjunto de valoración) y con distintos parámetros. Valora los resultados obtenidos y analiza lo que ha ocurrido. Compara los resultados de los clasificadores con los de las herramientas de clustering y argumenta las razones para elegir unos u otros.

## Entrega

La entrega se realizará a través del campus virtual en un archivo zip que contendrá las memorias elaboradas para cada parte, el archivo .pl de la primera parte y los conjuntos de entrenamiento y validación preparados para la segunda parte. La memoria de la segunda parte incluirá las URL de los textos seleccionados. En la portada de la memoria debe aparecer el número de grupo y los nombres completos de sus integrantes. Además el nombre del archivo será P9GXX, siendo XX el número de grupo.