

Práctica: Clasificación de Textos en Lenguaje Natural

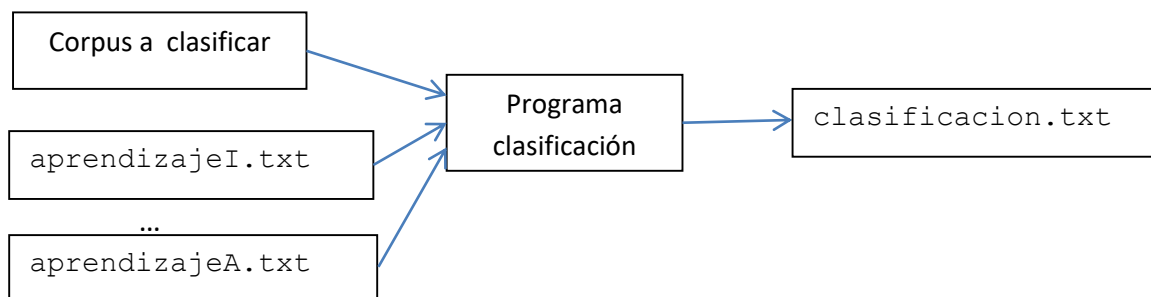
Objetivo: Construir un Sistema para la clasificación de los mensajes que emiten las empresas en las redes sociales. Los mensajes se dividen en informativos (afirmaciones objetivas sobre la empresa o sus actividades), diálogo (respuestas a los usuarios, etc.) o acciones (mensajes que piden votos a los usuarios, que piden clicks en enlaces, etc.)

Contenidos:

Parte 2 Clasificación

En esta parte se clasificarán los documentos presentes en un corpus.

Escribe un programa que tome como entrada las estimaciones de probabilidad de cada palabra y un corpus con documentos a clasificar (con el formato de corpus de 1.1) y devuelva los documentos clasificados en un fichero `clasificación.txt` donde cada línea del fichero de salida tiene la clase (I,D,A) del documento de entrada:



Deberás clasificar `corpustodo.txt` generando el fichero `clasificación.txt`
Deberás también clasificar el corpus suministrado por el profesor `corpustest.txt` generando el fichero `clasificacióntest.txt`

Entregable

En el Campus Virtual

- Programas:
 - o Clasificación

- **Fichero:**
 - o `clasificación.txt`, `clasificacióntest.txt`
- **Informe:**
 - o Descripción de la implementación: Lenguaje, Estructura de los programas, utilización (o no) de librerías, Participación de cada miembro del grupo en las tareas.
 - o Porcentaje de acierto del programa sobre `corpus_todo.txt`.

En el Campus Virtual

Nota

- **Calificación de la práctica:**
- Entregables (6/10)
- Rendimiento del programa sobre el corpus del profesor (4/10):
 - o 99-100% del porcentaje de acierto del mejor grupo 4 puntos
 - o 97-99% del porcentaje de acierto del mejor grupo 3 puntos
 - o 95-97% del porcentaje de acierto del mejor grupo 2 puntos
 - o 85-95% del porcentaje de acierto del mejor grupo 1 puntos
 - o menos del 85% del porcentaje de acierto del mejor grupo 0 puntos