



PRÁCTICA 3

Autómatas y Lenguajes Formales

Última actualización: 19 de noviembre de 2019

Eduardo Martínez Graciá, Mercedes Valdés Vela,
Santiago Paredes Moreno, José Manuel Juárez Herrero

Dpto. de Ingeniería de la Información y las Comunicaciones
Facultad de Informática de la **Universidad de Murcia**

INDICE

1. Introducción
2. Normas de división silábica
3. Normas para el uso de guiones
4. Trabajo y evaluación
5. Ampliación

INTRODUCCIÓN

INTRODUCCIÓN

Algunos sistemas informáticos que procesan lenguajes naturales como el español requieren conocer la descomposición de las palabras en sus sílabas.

- ▶ Síntesis o reconocimiento del habla.
- ▶ Sistemas de edición de texto.

En esta práctica se implementará un sistema de formateo de textos en español que inserta guiones en las posiciones adecuadas.

SÍLABAS EN ESPAÑOL

Una sílaba en español:

ATAQUE		NÚCLEO			CODA	
C	C	V	V	V	C	C

donde C es una consonante y V es una vocal.

- ▶ Consonantes: **b,c,d,f,g,h,j,k,l,m,n,ñ,p,q,r,s,t,v,w,x,y,z** y los grupos **ch, ll** y **rr**.
- ▶ Vocales: **a,e,i,o,u**, sus versiones acentuadas **á,é,í,ó,ú,ü**, junto con **y** a final de palabra (buey).
- ▶ *Vocal silábica*: es obligatoria, debe ser **a,e,i,o,u**, así como sus versiones acentuadas.
- ▶ *Vocales no silábicas*: **i,u** e **y** al final de palabra.

NORMAS DE DIVISIÓN SILÁBICA

NORMAS DE DIVISIÓN SILÁBICA

Estrategia: buscar patrones que identifiquen la porción de letras desde el final del núcleo de una sílaba hasta el comienzo del núcleo de la siguiente.

REGLA	EJEMPLO	SÍLABA 1							SÍLABA 2						
		ATAQUE		NÚCLEO			CODA		ATAQUE		NÚCLEO			CODA	
		C	C	V	V	V	C	C	C	C	V	V	V	C	C
1	u-no					V ₁	-	-	C	-	V ₂				
2.a	o-tro					V ₁	-	-	C ₁	C ₂	V ₂				
2.b	is-la					V ₁	C ₁	-	C ₂	-	V ₂				
3.a	en-tra					V ₁	C ₁	-	C ₂	C ₃	V ₂				
3.b	ist-mo					V ₁	C ₁	C ₂	C ₃	-	V ₂				
4	ins-crito					V ₁	C ₁	C ₂	C ₃	C ₄	V ₂				

La vocal V₁ es la última de la primera sílaba y V₂ es la primera de la segunda.

Las reglas de inserción de guiones no permiten la colocación de estos entre sílabas separadas por hiatos (ra-íz, o-cé-a-no).

DESCRIPCIÓN DE NORMAS (1)

1. $V_1 C V_2$. Una consonante entre dos vocales siempre forma sílaba con la vocal que le sigue (ejemplo: u-no).

Regla 1: $V_1 C V_2 \rightarrow V_1 - C V_2$

2. $V_1 C_1 C_2 V_2$. Dos consonantes contiguas rodeadas de vocales:
 - a) Grupos consonánticos de ataque que permite el español:
 - i. $(C_1 = p \mid c \mid b \mid g \mid f) + (C_2 = r \mid l)$: a-bra-zo, i-glú, o-pri-me, a-cla-ra, o-fre-ce.
 - ii. $(C_1 = d \mid t) + (C_2 = r)$: cua-tro, ce-dro.

En estos casos, $C_1 C_2$ son ataque de nueva sílaba.

Regla 2.a: $V_1 C_1 C_2 V_2 \rightarrow V_1 - C_1 C_2 V_2$

- b) En otro caso, consideramos que C_1 es coda de la sílaba anterior y C_2 es ataque de la siguiente sílaba, de modo que la separación silábica se sitúa entre C_1 y C_2 (is-la).

Regla 2.b: $V_1 C_1 C_2 V_2 \rightarrow V_1 C_1 - C_2 V_2$

DESCRIPCIÓN DE NORMAS (2)

3. $V_1 C_1 C_2 C_3 V_2$. Tres consonantes contiguas entre vocales:
- a) Si $C_2 C_3$ son un grupo como en 2.a), $C_2 C_3$ es ataque de la nueva sílaba y C_1 la coda de la anterior (des-gra-cia).

Regla 3.a: $V_1 C_1 C_2 C_3 V_2 \rightarrow V_1 C_1 - C_2 C_3 V_2$

- b) Si no, comprobar si $C_1 C_2$ forman una coda silábica que no está al final de la palabra:
- i. ($C_1 = b \mid d \mid n \mid m \mid l \mid r$) + ($C_2 = s$): abs-te-mio, ins-tar, háms-ter, sols-ti-cio, pers-pi-caz.
- ii. ($C_1 = s$) + ($C_2 = t$): post-da-ta, ist-mo.

Regla 3.b: $V_1 C_1 C_2 C_3 V_2 \rightarrow V_1 C_1 C_2 - C_3 V_2$

4. $V_1 C_1 C_2 C_3 C_4 V_2$. Hay división silábica entre la segunda y la tercera consonante cuando $C_1 C_2$ son un grupo como en 3.b) y $C_3 C_4$ como en 3.a) (post-pro-duc-ción, trans-plan-te, ads-cri-bir).

Regla 4: $V_1 C_1 C_2 C_3 C_4 V_2 \rightarrow V_1 C_1 C_2 - C_3 C_4 V_2$

NORMAS PARA EL USO DE GUIONES

NORMAS PARA EL USO DE GUIONES

Se seguirán las siguientes normas de la RAE para la inserción de guiones en un texto dado:

1. El guion no debe separar letras de una misma sílaba (incorrecto: tel-éfono).
2. Dos o más vocales seguidas nunca se separan al final de renglón, formen diptongo, triptongo o hiato (incorrecto: ca-usa).
3. La h intercalada se tratará como si no existiese (incorrecto: ad-herente).
4. Los grupos **ch**, **ll** y **rr** no se pueden dividir con guion de final de línea (incorrecto: per-ro).

TRABAJO Y EVALUACIÓN

TRABAJO (1)

Implementa un módulo `practica3.py`:

1. Entrada por consola:
 - ▶ Fichero `X.txt` que contiene un texto que hay que dividir mediante guiones.
 - ▶ Longitud de línea `L`, entero igual o superior a 10.
2. Salida: fichero `X_g.txt`.
3. Se insertarán guiones en las posiciones adecuadas para que la longitud de las líneas se acerque todo lo posible a `L` sin superarlo.
4. Otras transformaciones en el texto:
 - ▶ Múltiples espacios en blanco sustituidos por un único espacio.
 - ▶ Múltiples saltos de línea sustituidos por un único salto.
 - ▶ Los tabuladores serán sustituidos por un espacio simple.

TRABAJO (2)

5. Detección de errores: si no se detecta una situación válida al buscar el corte silábico, el programa termina imprimiendo un mensaje por salida de error que indique la línea y el término que no se ha podido dividir. No genera ningún fichero de salida.
6. Restricciones: no se puede emplear ningún método de cadenas de caracteres de Python, como `split()` o la comparación de cadenas. Sí se puede emplear el operador de concatenación o el troceado para generar la salida.
7. El código de `practica3.py` debe estar estructurado en métodos. Se penalizará el uso de métodos muy largos, mal estructurados, mal documentados o con nombres de variables poco significativas.

EVALUACIÓN

En caso de que se cumplan todos los requisitos indicados, esta práctica se evaluará con **4 puntos**.

AMPLIACIÓN

SPEECH RECOGNITION DE GOOGLE (1)

Puedes obtener **5 puntos** si integras la librería **Speech Recognition** de Google en tu sistema de guiones.

Tendrás que comenzar instalando PyAudio:

<http://people.csail.mit.edu/hubert/pyaudio/>

Después puedes instalar la librería **Speech Recognition**:

<https://pypi.org/project/SpeechRecognition/>

Indicaciones:

- ▶ Necesitas tener activa la conexión a Internet porque la traducción a texto se realiza en los servidores de Google.
- ▶ Imprime por consola el texto con los guiones conforme se vaya reconociendo.
- ▶ Sería conveniente que uses alguna palabra como comando para finalizar el reconocimiento de voz.

SPEECH RECOGNITION DE GOOGLE (2)

```
1  import speech_recognition as sr
2  r = sr.Recognizer()
3  with sr.Microphone() as source:
4      try:
5          audio = r.listen(source, timeout=2)
6          try:
7              texto = r.recognize_google(audio,
8                                     language="es-ES"))
9              print(texto)
10         except sr.UnknownValueError:
11             print("Google no entiende")
12         except sr.RequestError as e:
13             print("Error: {0}".format(e))
14     except sr.WaitTimeoutError:
15         print("Timeout de 2 segundos")
```