

# Expresiones regulares

Materia: Procesamiento de lenguaje natural  
Blanca Vázquez

# Expresiones regulares

---

- Conocidas como *regex* o *regexp* por su contracción de las palabras inglesas *regular expression*.
- Es una secuencia de caracteres que conforma un patrón de búsqueda.
- Se utilizan principalmente para la búsqueda de patrones de cadenas de caracteres u operaciones de sustituciones.

# ¿Dónde usar *regex*?

---

- Reemplazar textos
- Extraer partes específicas
- Dividir textos grandes en partes más pequeñas
- Verificar si una palabra existe dentro de un documento
- Verificar si un valor ingresado es válido

# Propósito

---

- Las *regex* pueden utilizarse en cualquier lenguaje: Python, Julia, Perl, JavaScript, HTML, ...
- Toda *regex* tiene algún autómata finito asociado.
- La curva de aprendizaje, al inicio, puede ser complicada.

“Some people, when confronted with a problem, think ‘I know, I’ll use regular expressions.’

Now they have two problems.”

**Jamie Zawinski**

# Sintaxis de una expresión regular

---

- Una *regex* es un patrón de texto que consiste en caracteres comunes (por ejemplo, de la **a** la **z** o números del **0 al 9**) y caracteres especiales (metacaracteres).
- El objetivo es describir las cadenas que coincidirá cuando se aplica a un texto.

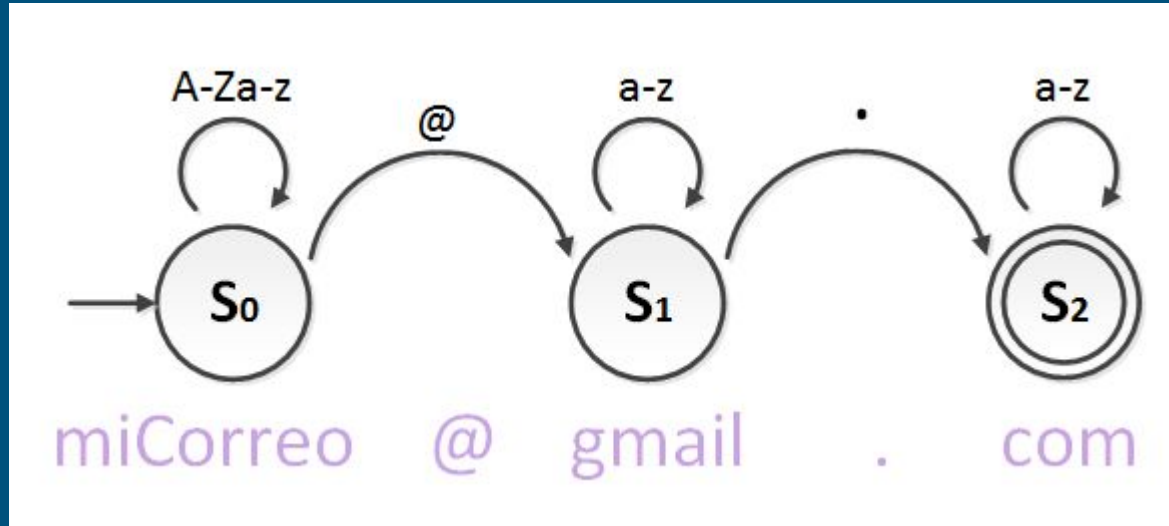
Ejemplo:  
Genera una regex que  
extraiga números

**\d+**

metacaracter



# Diagrama de estados



Autómata finito

# Time to code

---



# Variaciones en fechas

---

La fecha 25 de octubre de 2024,  
puede ser escrita como:

25-10-2024

25/10/2024

25/10/02

10/25/2024

25 Oct 2024

25 de octubre de 2022

Oct 25, 2024

Octubre 25, 2024