

PYTHON PARA LINGÜISTAS

CONCEPTOS BÁSICOS DE PYTHON: TIPOS DE DATOS, VARIABLES



ALEJANDRO ARIZA

CENTRE DE LLENGUATGE I COMPUTACIÓ

UNIVERSITAT DE BARCELONA

HISTORIA DE PYTHON



- Creado en 1991
- Disponible para Windows, Mac, Linux...
- La versión actual de Python es Python 3
- Todavía se pueden encontrar programas y librerías creadas con Python 2, pero Python 2 ya no recibe actualizaciones

INSTALACIÓN DE PYTHON

- Python debería estar ya instalado en todos los ordenadores del aula.
- Si queréis instalaros Python en vuestro ordenador personal, recomiendo el uso de Anaconda. Permite controlar diferentes versiones de librerías, crear entornos virtuales, etc con una interfaz de usuario amigable.
- Para realizar los ejercicios de este curso, serán necesarias las siguientes librerías/paquetes:
 - matplotlib, collections, csv, nltk, pandas
 - numpy, scipy, scikit-learn, gensim, spacy
 - nltk.data

EJECUTAR PYTHON

Hay varias formas de ejecutar Python:

- Un programa en Python puede ejecutarse desde la máquina de comandos: `python3 myprogram.py`
- Alternativamente, si solo queremos ejecutar el intérprete de Python para hacer pruebas. En este caso, lo lanzaremos con el siguiente comando: `python3`
- **Jupyter Notebook:** una aplicación que te permite escribir código Python dentro de unos ficheros llamados “notebooks” que disponen de celdas que pueden verse como contenedores de código. De esta forma, te permite ejecutar bloques separados de código, guardando en memoria los resultados de celdas evaluadas con anterioridad. Además, este programa está pensado para el usuario y para análisis/visualización de datos y resultados.

ALGORITMOS Y DATOS

PARTE 1: DATOS

COMPUTACIÓN = DATOS + ALGORITMOS

Recipe for baking a cake....

- 2 sticks butter
 - 2 cups flour
 - 1 cup sugar
 - 4 eggs
 - 1 cup milk
 - 1 tsp baking powder
 - Cocoa powder (1/2 pound)
- DATA (input)**



Mix the sugar, baking powder and flour, mix in beaten eggs, melted butter and bake at 325F for 40 mins.

ALGORITHM

DATOS Y TIPOS DE DATOS

- Los datos son los ingredientes (entrada) de nuestra receta (programa).
- Para que el ordenador entienda estas entradas, existen pre-definidos determinados tipos de datos. Recordemos que el ordenador únicamente entiende 0s y 1s por lo que diferentes tipos de datos ocuparan mayor o menor espacio.
- Cada tipo de datos tiene sus características y operaciones particulares.
- A menudo, los tipos de datos en un lenguaje de programación corresponden a “tipos” reales de datos lo que facilita su comprensión.

TIPOS DE DATOS NUMÉRICOS

- Los tipos de datos numéricos son esenciales para muchos programas informáticos.
- En Python, existen dos tipos de datos numéricos:
 - Integer (Entero): 1, 2, 3, 4, ... 100, ... 1337, ...
 - Float (Flotante): 3.1415, 66.667, 91.5, ...
- Ambos tipos de datos son muy parecidos en sus características.
- Es importante hacerle saber al ordenador el tipo de datos con los que está trabajando por su forma de codificarlos y procesarlos posteriormente.

OPERACIONES CON DATOS NUMÉRICOS

Las operaciones esenciales que se pueden hacer con estos tipos de datos son:

- Suma $(+): 2 + 3 = 5$
- Resta $(-): 3 - 2 = 1$
- Multiplicación $(*): 2 * 3 = 3 * 2 = 6$
- División $(/): 5 / 2 = 2.5$
- (Parte entera) División $(//): 5 // 2 = 2$
- (Resto) División $(%): 5 \% 2 = 1$
- Potencia $(**): 5 ** 2 = 25$

TIPO DE DATO LÓGICO/BOOLEANO

- Tipo de dato booleano. Puede tomar dos valores: **TRUE** y **FALSE**. También pueden representarse (en formato numérico) como 1 y 0 respectivamente.
- Otros tipos de datos pueden tener “infinitos” valores.
- ¿Razones por las que necesitamos el tipo de datos booleano?
 - Los ordenadores operan de forma binaria.
 - Permite evaluar condiciones (e.g.: “¿Has comprado levadura?”).
 - Eficiencia en la computación y escritura de código.
 - Hay lenguajes de programación que funcionan sin un tipo de datos booleano explícito.

OPERACIONES CON DATOS BOOLEANOS

- Las operaciones lógicas más comunes son:
 - Logical **AND**
 - Logical **OR**
 - Logical **NOT**
- Hay operaciones lógicas más complejas que veremos a lo largo del curso.

TIPOS DE DATOS PARA TEXTO (STRING)

- Uno de los tipos de datos más importantes es el “String” (cadena de caracteres).
- En lingüística, podemos diferenciar entre diferentes unidades de texto más precisas como caracteres, palabras, frases, párrafos, etc.
- En Python, por defecto, únicamente tenemos “strings”
- Un String consiste en una secuencia de caracteres delimitados por comillas:
 - “s” – un string de un carácter; ‘string’ – un string de una palabra; “un string” – una frase;
 - Un string en Python puede delimitarse por comillas simples (‘hola’) o comillas dobles (“hola”) indistintamente

OPERACIONES CON STRINGS

- Las operaciones básicas con strings incluyen la concatenación y segmentación.
- La concatenación permite unir dos strings en uno:
 - “un ” + “string” se convierte en “un string”
 - El operador + no añade ningún espacio al concatenar.: “Un” + “string” resultaría en “Unstring”
- Segmentar es una operación que te devuelve una parte de un string:
 - Partiendo del string “un string”, podríamos segmentarlo en los siguientes strings: “un”, “ “, “str”, or “string”, etc.

TIPO DE DATO LISTA

- Otro tipo de datos bastante importante es el tipo “list”
- Una lista es una colección de elementos de cualquier tipo (incluido entero, lista, booleano, etc) delimitado por corchetes y separados por comas:
 - Una lista de enteros: [1, 2, 3, 4, 5]
 - Una lista de strings: [“algunas”, “palabras”, “en”, “una”, “lista”]
 - Una lista de enteros Y strings: [6, “palabras”, “y”, 2, “números”, “en”, “una”, “lista”]
- Las operaciones típicas de las listas son concatenación y segmentación (al igual que los strings)

ALGORITMOS Y DATOS

PARTE 2: ALGORITMOS

COMPUTACIÓN = DATOS + ALGORITMOS

Recipe for baking a cake....

- 2 sticks butter
 - 2 cups flour
 - 1 cup sugar
 - 4 eggs
 - 1 cup milk
 - 1 tsp baking powder
 - Cocoa powder (1/2 pound)
- DATA (input)**

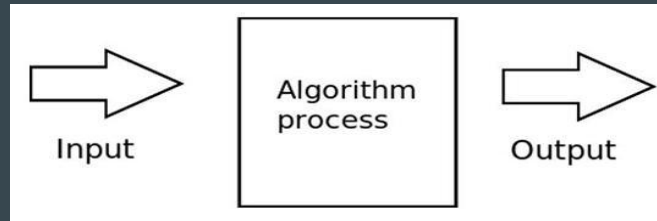


Mix the sugar, baking powder and flour, mix in beaten eggs, melted butter and bake at 325F for 40 mins.

ALGORITHM

ALGORITMOS: TRANSFORMAR LOS DATOS

- De forma simplificada, un algoritmo recibe unos datos como entrada, los procesa de una forma determinada y devuelve el resultado.
- En el ejemplo de la receta, el algoritmo consistiría en recibir los alimentos básicos como entrada y, siguiendo los pasos determinados por la receta, devolver un pastel ya cocinado.



INTERMEDIATE STEPS IN ALGORITHMS

- La “transformación” desde la entrada hasta la salida puede contener pasos intermedios:
 - 1) “Mezclar el azúcar con la harina y la levadura”
 - 2) “Mezclar con huevos batidos”
 - 3) “Derretir la mantequilla”
 - 4) “Mezclar con la mantequilla derretida”
 - 5) “Hornear a 220°C durante 40 mins”
- Estos pasos intermedios también reciben datos como entrada y devuelven un resultado que servirá de entrada para el siguiente paso. Para estar al corriente de todos estos datos, utilizamos lo que se denomina como “variables”.

DEFINICIÓN Y USO DE VARIABLES

¿QUÉ SON LAS VARIABLES?

- “Las variables se utilizan para guardar información que va a ser referenciada y manipulada en un programa informático”
- “Las variables proveen de una forma de etiquetar los datos con un nombre descriptivo para que nuestros programas se entiendan más fácilmente por otros lectores o nosotros mismos”
- “Puede ser útil pensar en las variables como contenedores de información. Su función principal es etiquetar y guardar datos en memoria. De esta forma, estos datos pueden ser luego utilizados a lo largo de tu programa.”

¿QUÉ SON LAS VARIABLES? EJEMPLO

Descripción: La Tierra es un planeta del sistema solar que gira alrededor de su estrella —el Sol— en la tercera órbita más interna. Es el más denso y el quinto mayor de los ocho planetas del sistema solar. También es el mayor de los cuatro terrestres o rocosos.

Distancia desde el Sol: 149,6 millones km

Edad: 4,543 miles de millones años

Radio: 6.371 km

Masa: $5,972 \times 10^{24}$ kg

Superficie: 510,1 millones km²

ANALOGÍA EN EL EJEMPLO DEL PASTEL

- Una variable sería el utensilio de cocina que utilizas para guardar los ingredientes
- Al principio del programa los utensilios están vacíos
- Después del paso 1, contiene la mezcla de “azúcar, harina y levadura”
- Después del paso 2: “azúcar, harina, levadura y huevos batidos”
- Después del paso 4: “azúcar, harina, levadura, huevos batidos y mantequilla derretida”
- Al final del programa contiene un pastel.

DEFINIR UNA VARIABLE EN PYTHON

- La sintaxis para la definición de una variable en Python es: `variable = value`
- Definición de un string: `frase = "Esto es un string."`
- Definición de un booleano: `es_verde = True`
- Definición de un entero: `num_palabras = 17`
- Los valores de una variable pueden ser la evaluación de una expresión. En el caso de una expresión matemática sería: `numero_pi = 3.0 + 0.1 + 0.04`

ASIGNAR NOMBRES A VARIABLES

- Los nombres de las variables deben seguir unas convenciones específicas o de lo contrario, aparecerán errores en la ejecución del código:
 - Se permite tanto letras mayúsculas como minúsculas (A-Z, a-z).
 - Se permite añadir números (0-9, **salvo en el primer carácter** del nombre de la variable).
 - Los caracteres especiales **NO** se permiten, a excepción de la barra baja.
 - Algunas palabras están reservadas para comandos de Python específicos.
- Existen ciertas buenas prácticas a la hora de dar nombres a las variables (deben tener sentido). Por ejemplo, si guardamos en una variable el diámetro de la Tierra, un buen nombre sería “diametro_tierra”.

PRÁCTICA: VARIABLES Y TIPOS DE DATOS

- En la primera práctica del curso, veremos unos ejemplos de asignación de variables usando diferentes tipos de datos y sus operadores correspondientes.
- Utilizaremos algunas funciones ya pre-definidas en Python:
 - `print()` – imprime el valor de la variable que recibe como entrada
 - `len()` – cuenta en número de caracteres en un string o de elementos en una lista
- En la próxima sesión, aprenderemos nuevos conceptos acerca de funciones.

¡GRACIAS!
¿PREGUNTAS?