

PYTHON PER A LINGÜISTES

CONCEPTES BÀSICS DEL LENGUATGE DE PROGRAMACIÓ PYTHON

Alex Peiró Lilja

Servei de Tecnologia Lingüística (STeL)

Universitat de Barcelona

HISTÒRIA DE PYTHON

- Llenguatge presentat l'any 1991 per Guido van Rossum
- Disponible per a Windows, Mac i Linux
- Actualment es treballa amb la versió Python 3
- A l'hora de buscar noves llibreries o paquets, assegurar-se d'instal·lar-los en la versió 3



INSTAL·LACIÓ DE PYTHON

- Python hauria d'estar instal·lat en tots els ordinadors de l'aula
- Per Windows:
 - Instal·lar Anaconda. Ofereix una interfície fàcil de gestionar. Permet crear entorns virtuals i treballar amb diferents versions de llibreries
- Per realitzar els exercicis del curs comptarem amb una sèrie de llibreries/paquets:
 - Matplotlib, collections, csv, nltk, pandas
 - Numpy, scipy, scikit-learn, genism, spacy
 - nltk.data

EXECUCIÓ DE PYTHON

- Un programa en Python es pot executar des de la màquina d'ordres (terminal):
 - `$ python3 myprogram.py`
- D'altra banda, si volem executar ordres específiques per fer proves, podem habilitar l'interpret de Python en el propi terminal:
 - `$ python3`
- Jupyter Notebook: aplicació web on podem editar codi Python dins uns fitxers anomenats "notebooks", els quals disposen de cel·les que poden contenir codi, o bé text, imatges, equacions, etc. D'aquesta manera es poden executar blocs de codi per separat, guardant en memòria les executades amb anterioritat. Molt útil per visualització i anàlisi de resultats i dades.
- Google Colab: = Jupyter Notebook. Els arxius estan connectats a un servidor remot amb l'entorn de Python i les seves llibreries i paquets al dia. També ofereixen la possibilitat d'utilitzar GPUs per executar programes d'alta complexitat.

ALGORISMES Y DADES

PART I: DADES

Computació = Dades + Algorismes

Recepta per coure un pastís:

- 2 barres de mantega
 - 2 tasses de farina
 - 1 tassa de sucre
 - 4 ous
 - 1 tassa de llet
 - 1 culleradeta de llevat en pols
 - 1/2 lliura de cacau en pols
- DADES (Input)**



Barregeu el sucre, el llevat i la farina. Barregeu-hi els ous batuts, la mantega fosa i coure a 220°C durant 40 minuts

ALGORISME

DADES I TIPUS DE DADES

- Les dades són els ingredients (entrada) de la nostra recepta (programa)
- Per tal de que l'ordinador entengui aquestes entrades, existeixen uns tipus de dades pre-definits. Recordem que l'ordinador únicament entén 0s i 1s. Així que alguns tipus de dades ocuparan més o menys espai de memòria.
- Cada tipus de dada té les seves característiques i operacions particulars
- Sovint, els tipus de dades en un llenguatge de programació corresponen a “tipus” reals de dades, facilitant la seva comprensió.

TIPUS DE DADES NUMÈRIQUES

- Els tipus de dades numèriques són essencials per a molts programes informàtics
- En Python, existeixen bàsicament dos tipus de dades numèriques:
 - Integer (Enter): 1, 2, 3, 4, ..., 100, ..., 1578, ...
 - Float (Flotant): 3.1415, 34.334, 9.5, 0.001, ...
- Ambdós tipus de dades són molt semblants en les seves característiques
- Important fer-li saber a l'ordinador el tipus de dades amb el que està treballant per la seva forma de codificar-lo i processar-lo

OPERACIONS AMB DADES NUMÈRIQUES

- Les operacions essencials que es poden realitzar amb aquest tipus de dades:

○ Suma	$(+) : 2 + 3 = 5$
○ Resta	$(-) : 3 - 2 = 1$
○ Multiplicació	$(*) : 2 * 3 = 3 * 2 = 6$
○ Divisió	$(/) : 5 / 2 = 2.5$
○ (Part sencera) Divisió	$(//) : 5 // 2 = 2$
○ (Residu) Divisió	$(\%) : 5 \% 2 = 1$
○ Potència	$(**) : 5 ** 2 = 25$

TIPUS DE DADA LÒGICA/BOOLEAN

- Tan sols poden ser dos tipus de valors: **TRUE** o **FALSE**. També poden representar-se en format numèric com 1 o 0, respectivament
- Diferent als tipus de dades numèriques que poden tenir “infinit” valors.
- Raons per les quals necessitem les dades de tipus boolean:
 - Els ordinadors operen de forma binària
 - Permeten avaluar condicions (e.g: “Has comprat el llevat?”)
 - Eficiència en la computació i l’escriptura de codi (més descriptius)
 - Hi ha altres llenguatges de programació que funcionen sense un tipus de boolean explícit

OPERACIONS AMB BOOLEANS

- Les operacions lògiques més comunes són:
 - Logical **AND** (comprovar que es compleixen dues o més condicions)
 - Logical **OR** (comprovar que almenys una de les condicions es compleix)
 - Logical **NOT** (comprovar que no es compleix una condició)
- Al llarg del curs es mostraran aquestes operacions lògiques i d'altres més complexes

TIPUS DE DADES PER A TEXT: STRING

- El tipus de dada que representa el text que tots entenem és l'“String” (cadena de caràcters)
- En lingüística, podem diferenciar diferents unitats de text més precises com les paraules, frases, paràgrafs, etc.
- En canvi Python únicament representa “strings”
- Un *String* consisteix en una seqüència de caràcters delimitats per unes cometes:
 - “s” : un *string* d'un sol caràcter
 - “string” : un *string* d'una paraula
 - “un string” : un *string* d'una frase
 - És el mateix representar la cadena amb cometes simples ('hola') com amb cometes dobles (“hola”)

OPERACIONS AMB STRINGS

- Les operacions bàsiques amb *strings* inclouen la concatenació i la segmentació
- La concatenació permet seguir dos strings en un de sol:
 - “un ” + “string” es converteix en “un string”
 - L'operador + no afegeix un espai natural entre els dos *strings* concatenats: “Un” + “string” resultaria “Unstring”
- Segmentar és una operació que et retorna una part de l'*string*:
 - Partint de l'*string* “un string”, podem segmentar-lo com: “un”, “ ”, “str”, “string”, etc.

ESTRUCTURA DE DADES: LLISTA

- L'estructura més bàsica i essencial en Python és la llista
- Una llista és una col·lecció d'elements de qualsevol tipus (enters, flotants, *strings*, booleans...) de manera ordenada.
- Dades delimitades per claudàtors i separats per comes:
 - Llista d'enters: [1, 2, 3, 4, 5]
 - Llista d'*strings*: ["algunes", "paraules", "en", "una", "llista"]
 - Llista d'enters i *strings*: [6, "paraules", "i", 2, "números", 26, "llista", "una"]
- Les operacions típiques també són concatenacions i segmentacions

Computació = Dades + Algorismes

Recepta per coure un pastís:

- 2 barres de mantega
 - 2 tasses de farina
 - 1 tassa de sucre
 - 4 ous
 - 1 tassa de llet
 - 1 culleradeta de llevat en pols
 - 1/2 lliura de cacau en pols
- DADES (Input)**



Barregeu el sucre, el llevat i la farina. Barregeu-hi els ous batuts, la mantega fosa i coure a 220°C durant 40 minuts

ALGORISME

ALGORISMES Y DADES

PART II: ALGORISMES

ALGORISMES: TRANSFORMAR LES DADES

- De manera simplificada, un algorisme rep unes dades d'entrada, les processa d'una manera determinada, retorna un resultat.
- Seguint l'exemple de la recepta, l'algorisme agafaria els ingredients com a dades d'entrada i, seguint els passos determinats per la recepta, retornar el pastís preparat.



PASSOS INTERMEDIS EN ELS ALGORISMES

- La “transformació” des de l’entrada fins a la sortida pot contenir passos intermedis:
 1. “Barrejar el sucre amb la farina i el llevat”
 2. “Barrejar amb els ous batuts”
 3. “Desfer la mantega”
 4. “Barrejar amb la mantega fosa”
 5. “Posar al forn a 220°C durant 40 mins”
- Aquests passos intermedis també reben dades d’entrada i retornen dades de sortida, que alhora seran dades d’entrada pel següent pas. Per tal de mantenir i supervisar totes aquestes dades, utilitzem les **variables**.

DEFINICIÓ I ÚS DE VARIABLES

QUÈ SÓN LES VARIABLES?

- “Les variables es fan servir per enregistrar informació que serà referenciada i manipulada en un programa informàtic”
- “Les variables proveeixen d’una forma d’etiquetar les dades amb un nom descriptiu per a que els nostres programes s’entenguin més fàcilment per altres lectors o per nosaltres mateixos”
- “Pot ser útil pensar en les variables com contenidors d’informació. La seva funció principal és etiquetar i enregistrar dades en memòria. D’aquesta manera, aquestes dades poden ser després usades al llarg del nostre programa.”

QUÈ SÓN LES VARIABLES?

- Exemple del pastís:
 - Una variable seria l'estri de cuina que utilitzem per emmagatzemar els ingredients
 - Al principi del programa l'estri està buit
 - Després del pas 1, conté la barreja de “sucre, farina i llevat”
 - Després del pas 2: “sucre, farina, llevat i ous batuts”
 - Després del pas 4: “sucre, farina, llevat, ous batuts i mantega fosa”
 - Al final del programa conté un pastís

DEFINIR UNA VARIABLE EN PYTHON

- La sintaxi per a la definició d'una variable en Python es: **variable** = valor
- Definició d'un *string*: **frase** = "Això és un string"
- Definició d'un boolean: **es_verd** = True
- Definició d'un enter: **num_paraules** = 17
- El valor d'una variable pot ser també l'avaluació d'una expressió. Per exemple, una expressió matemàtica: **numero_pi** = 3.0 + 0.1 + 0.04

ASSIGNACIÓ DE NOMS A VARIABLES

- Els noms de les variables han de seguir unes convencions específiques, o en cas contrari apareixeran errors d'execució en el codi:
 - Es permeten tant lletres majúscules com minúscules (A-Z, a-z)
 - Es permeten afegir números (0-9, tret en el primer caràcter del nom de la variable)
 - Tot i que Python accepti accents als noms de les variables, no és recomanable per temes de rendiment
 - Els caràcters especials **NO** estan permesos, excepte el guió baix '_'
 - Algunes paraules estan reservades per a ordres de Python específiques
- Existeixen bones pràctiques a l'hora de nombrar les variables (donar sentit al contingut). Per exemple, si guardem en una variable el diàmetre de la Terra, un bon nom seria "diametre_terra"

ASSIGNACIÓ DE NOMS A VARIABLES

- En la primera pràctica d'aquest curs, veurem alguns exemples d'assignació de variables utilitzant diferents tipus de dades i els seus operadors corresponents
- Utilitzarem algunes funcions ja pre-definides en Python:
 - `print()` → imprimeix el valor de la variable que rep d'entrada
 - `len()` → compta el número de caràcters o elements d'una llista d'entrada
- En la propera sessió, aprendrem més sobre les funcions

GRÀCIES!
PREGUNTES?