

PYTHON PER A LINGÜISTES

INTRODUCCIÓ

Alex Peiró Lilja

Servei de Tecnologia Lingüística (STeL)

Universitat de Barcelona

INFORMACIÓ GENERAL

- Aula 4.4
- Duració: 30 hores – 10 sessions
- Horari: Dimecres de 16:30h a 19:30h, del 19/04/2023 fins al 21/06/2023
- Instructor: Alex Peiró Lilja (alex.peiro.lilja@gmail.com)

A QUI ESTÀ DIRIGIT?

- Curs d'introducció a la programació i processament del llenguatge natural (text) per a lingüistes amb Python. El curs també està pensat per a altres perfils amb interès en programació (nivell d'iniciació) per a la seva aplicació a l'anàlisi de text.
- El curs està dissenyat assumint que els participants parteixen d'una posició amb zero experiència en programació
- S'explicaran conceptes de programació com els tipus de dades o l'estructuració d'idees a mesura que vagin presentant-se en els exemples

OBJECTIUS DEL CURS

- Objectius principals:
 - Comprensió dels conceptes bàsics de la programació
 - Adquirir coneixements d'estructuració de programes
 - Traduir un problema lingüístic a un problema de programació
- Sessions pràctiques:
 - Estructures de control, bucles, funcions, manipulació de fitxers de text
 - Ús de llibreries externes Python pensades per al processament de text
 - Diferents corpus

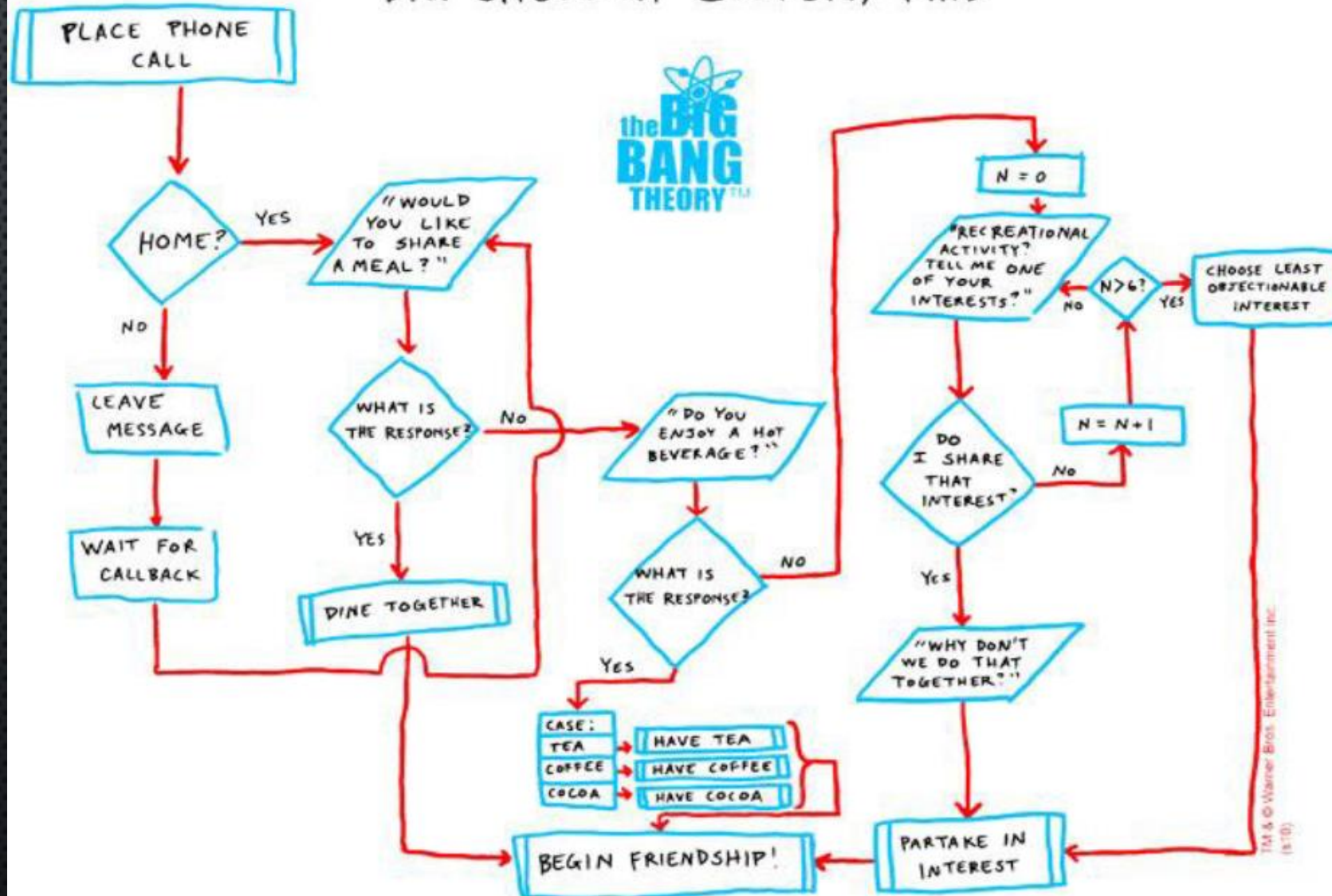
AGRAÏMENTS I BIBLIOGRAFIA

- Diapositives del curs son adaptacions del material impartit per Alejandro Ariza en el curs 2022
- “Introducción a la programación para humanistas” (Leticia Martín-Fuertes Moreno)
- “The NLTK book” i “Python Programming: An Introduction to Computer Science” són recursos addicionals adequats per familiaritzar-se amb Python i NLTK

PENSAMENT ALGORÍSMIC

THE FRIENDSHIP ALGORITHM

DR. SHELTON COOPER, Ph.D



QUÈ ES LA COMPUTACIÓ?

COMPUTACIÓ

La computació es pot considerar com la ciència o el tractament automàtic de la informació

Computació = Dades + Algorismes

Algorisme [Viquipèdia] : “un algorisme és una seqüència finita d'instruccions realitzables, no ambigües, l'execució de les quals condueix a una resolució d'un problema...l'algorisme pot ser dissenyat per rebre i aprofitar una determinada entrada, donant com a resultat una sortida... “

Dades [Viquipèdia] : “ les dades són qualsevol seqüència d'un o més símbols als quals se'ls dona significat mitjançant actes específics d'interpretació.“

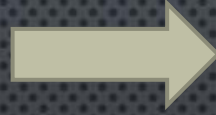
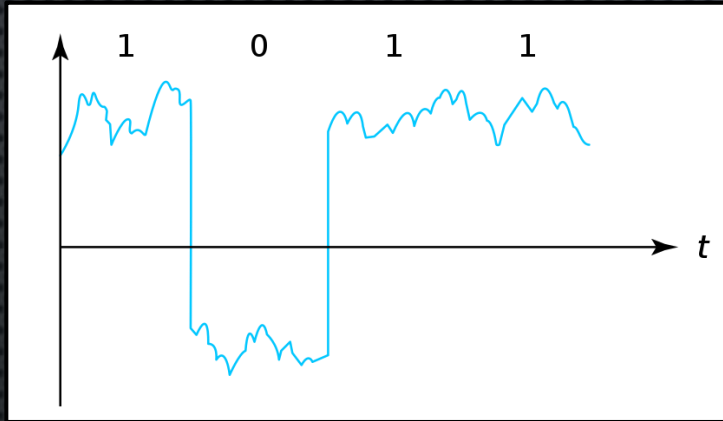
PER QUÈ ÉS ÚTIL LA COMPUTACIÓ?

1. Durant els últims anys, la quantitat de dades disponibles ha crescut, i segueix creixent, exponencialment. Els humans no podem tractar tantes dades manualment.
 1. Text
 2. Veu
 3. Imatge
 4. Vídeo...
2. Per tant, amb els algorismes apropiats, les computadores poden processar tota aquesta quantitat de dades de manera eficient i molt més ràpid
3. Permet als científics i investigadors avançar amb major velocitat, tant en la preparació i execució dels experiments com en l'anàlisi dels resultats

LLENGUATGES DE PROGRAMACIÓ

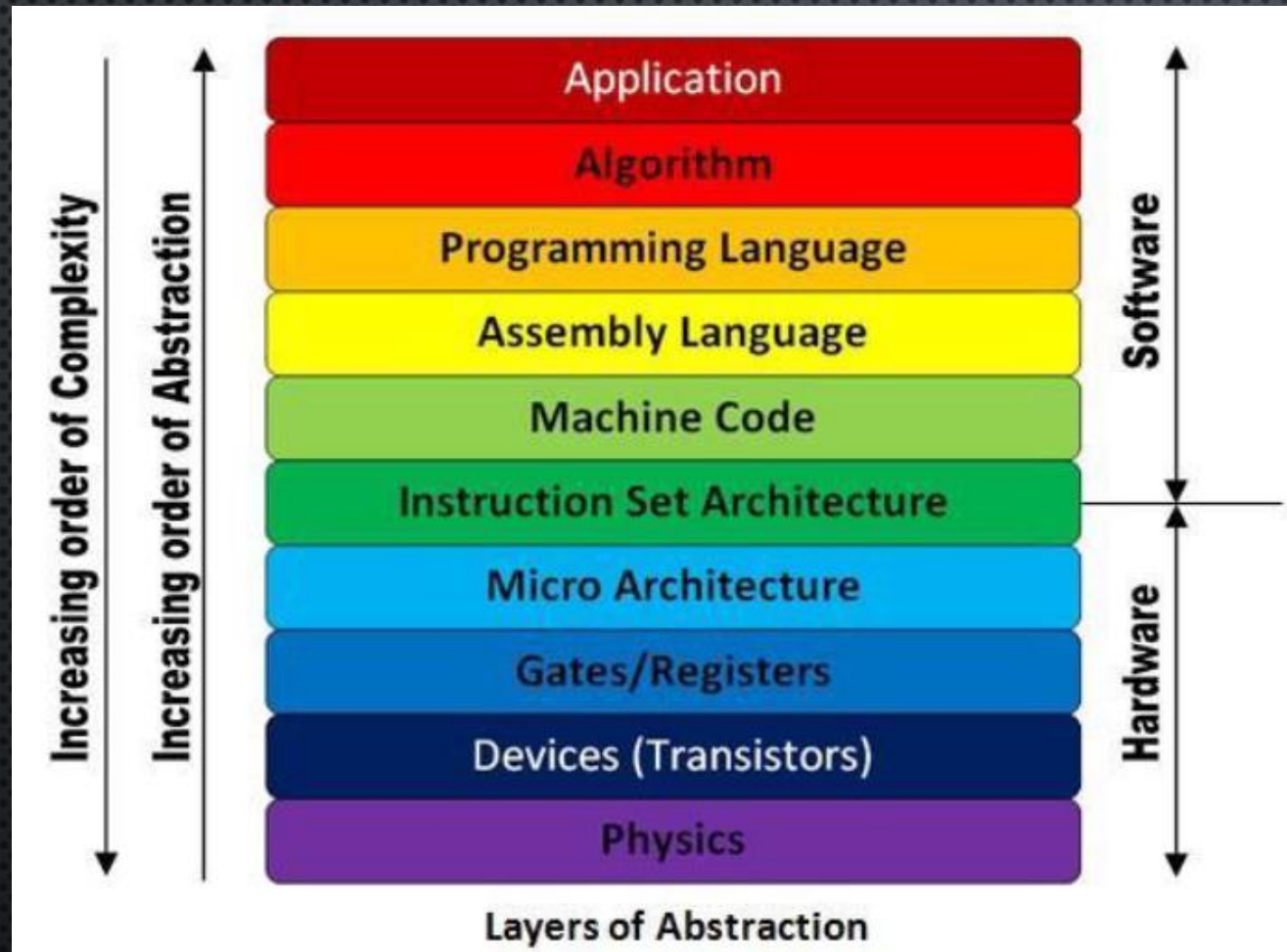
LA NOSTRA COMUNICACIÓ AMB ELS ORDINADORS

DEL 0 I 1 AL LLENGUATGE HUMÀ



- Similar al cervell humà, que funciona també per impulsos elèctrics dels neurotransmissors
- A través de bilions de neurones interconnectades, som capaços d'aprendre i entendre llenguatge natural, artificial, matemàtiques, abstracció...
- No sabem comunicar-nos a través d'impulsos elèctrics tal com fa la computadora
- “Ensenyar” als ordinadors un llenguatge més natural per entendre'ns

CAPE D'ABSTRACCIÓ



- A mesura que afegim més capes d'abstracció, més humana sembla la comunicació amb la màquina
- Els nivells més baixos són els més propers al llenguatge màquina
- En aquest curs aprendrem el nivell 'Programming Language'
- A més abstracció millor interpretació, però menor velocitat de càlcul

LLENGUATGE DE PROGRAMACIÓ

- Els llenguatges de programació són un pont entre el llenguatge màquina base (0s i 1s) i el llenguatge natural humà
- Com a “llenguatge”:
 - Estructura i regles predefinides (Sintaxi)
 - S'aprèn a través de la pràctica
- L'ordinador té una sèrie d'instruccions per poder traduir el llenguatge de programació a codi binari (0s i 1s)

LLENGUATGE DE PROGRAMACIÓ

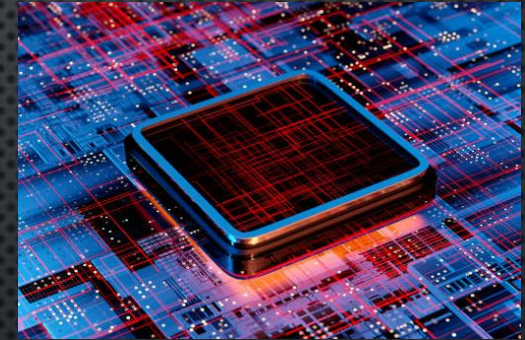


Humà



```
0 # loading from the website
1 requests.get(url)
2 # checking response.status_code (if you get 502, try removing the code)
3 if response.status_code != 200:
4     print(f"Status: {response.status_code} - Try removing the code")
5 else:
6     print(f"Status: {response.status_code}\n")
7 # using BeautifulSoup to parse the response object
8 soup = BeautifulSoup(response.content, "html.parser")
9 # finding Post images in the soup
10 images = soup.find_all("img", attrs={"alt": "Post image"})
11 # downloading images
12 images:
```

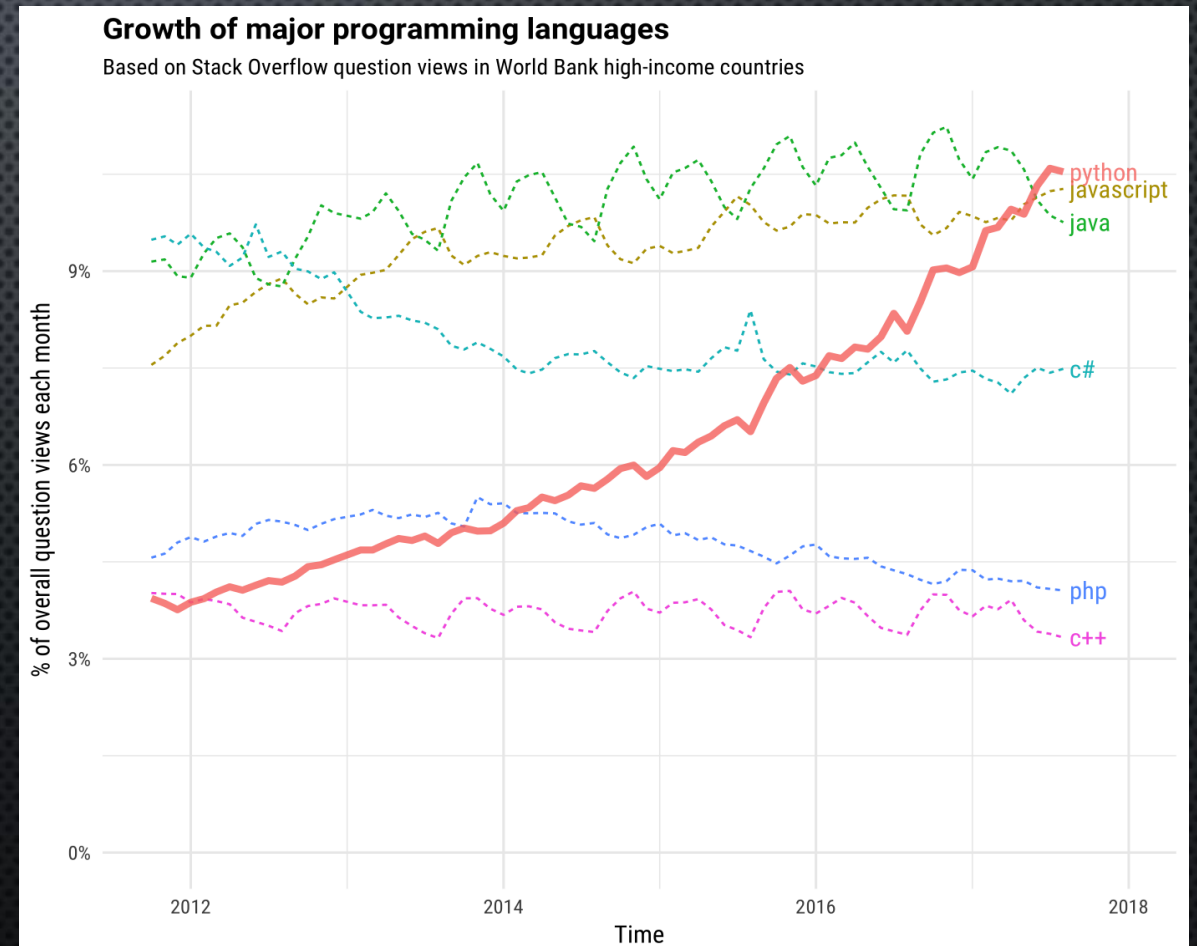
Llenguatge
de
Programació



010001010100
010101001010
111100101001

PYTHON COM A LENGUATGE DE PROGRAMACIÓ

- Llenguatge de programació d'alt nivell que intenta minimitzar el nombre de línies de codi
- Intenta apropar-se al llenguatge natural
- Llenguatge amb gran aplicabilitat i altament acceptat en el camp científic
- Cada cop més popular
- Gran disposició de llibreries per realitzar tasques de tota mena



GRÀCIES!
PREGUNTES?