



Twitter API

Laia Subirats i Aina Palacios

25 de novembre de 2022



WIFI: Tallers

Contrassenya: Jump2workshop_22

IT ACADEMY

< Aprèn a programar en 18
setmanes i **reprograma** el
teu futur ./>

Presentació Aina Palacios

- Enginyera de Telecomunicacions especialitzada en Audiovisuals
- Màster en Tecnologies Avançades especialitzada en deep learning en Multimèdia!
- Experiència en programació web i machine learning.
- Mentora a IT Academy de **Data Science** i investigadora a Eurecat



<https://www.linkedin.com/in/ainapc/>



<https://github.com/aina1997>

Presentació Laia Subirats

Experiència laboral:

- **Investigadora @ Eurecat – Centre Tecnològic de Catalunya (2009-Actualitat)**
- **Consultora @ Universitat Oberta de Catalunya (2016-Actualitat)**
- Consultora @ Universitat Pompeu Fabra (Estiu 2011, Estiu 2012)
- Estudiant @ Google Summer of Code (Estiu 2008)
- Estudiant d'estiu @ European Organization for Nuclear Research (CERN) (Estiu 2007)
- Becària @ Telefónica I+D (2006-2008)

Estudis:

- Doctorat en Informàtica @ Universitat Autònoma de Barcelona (2009-2015)
- Màster en Telemàtica @ Universitat Politècnica de Catalunya (2008-2009)
- Enginyeria de Telecomunicacions @ Universitat Pompeu Fabra (2003-2008)

01

Què és una API?

Conceptes bàsics

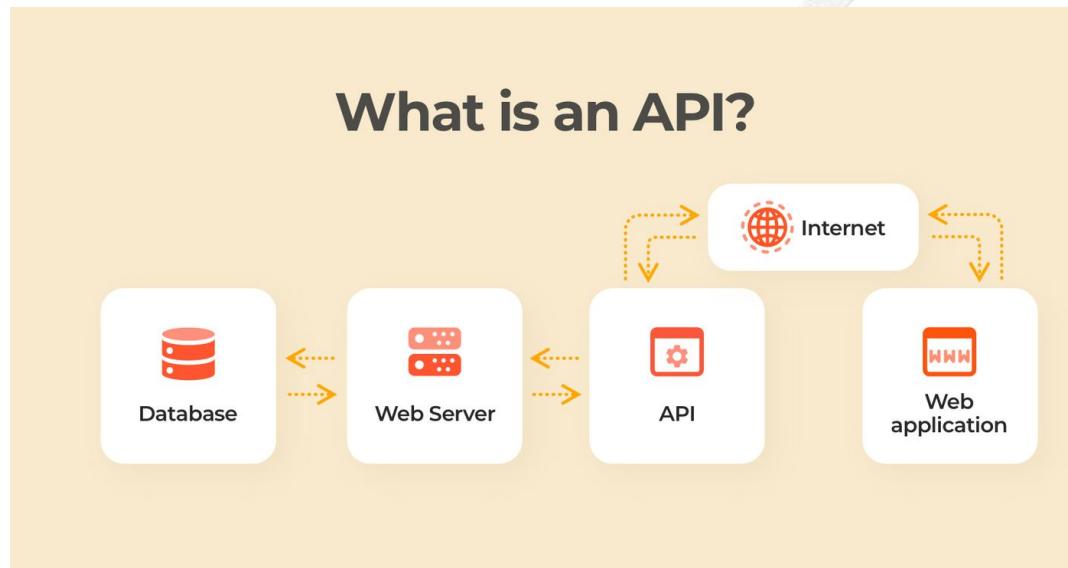




API

Application Programming Interface -> interfície que permet la intercomunicació entre dos sistemes. Els mètodes d'aquestes ja estan predefinides!

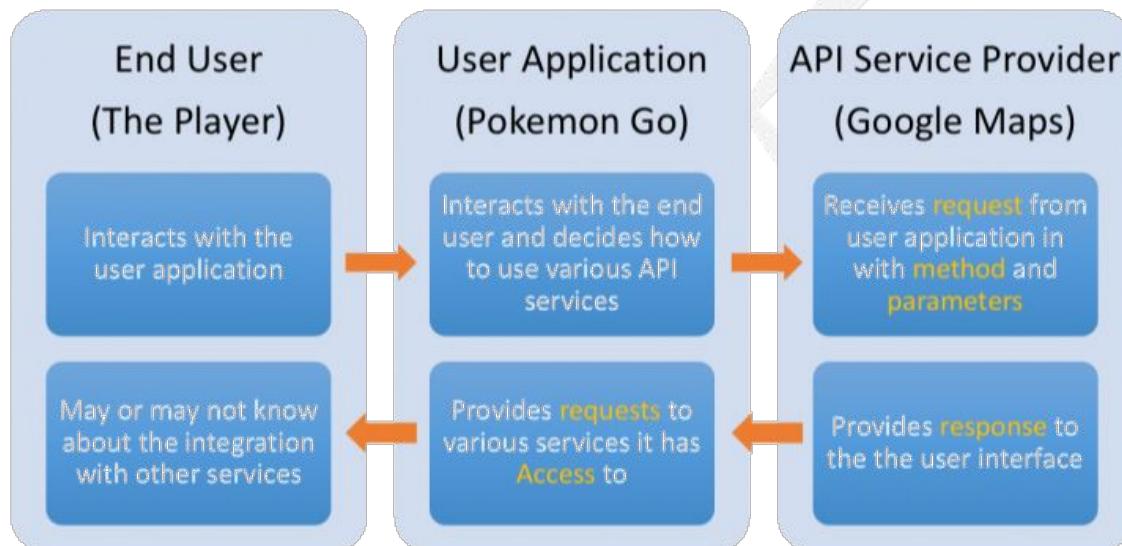
- És independent del llenguatge utilitzat
- Permeten l'entrada a un software que no ha de ser necessàriament de la teva propietat, i fer ús d'aquest sense haver-lo de tenir.
- No necessites saber com funciona l'altre software!





Per a què serveix una API?

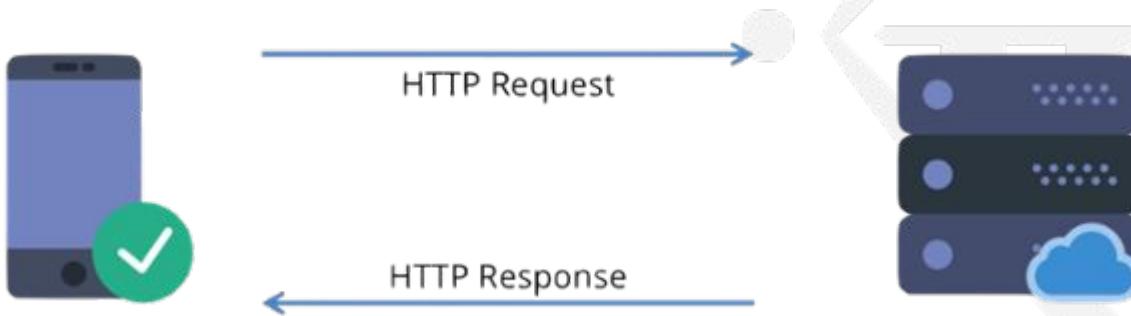
- Intercanvi de dades o funcionalitats entre dos sistemes diferents! (Ex. El temps)
- Pots integrar la funcionalitat al teu programa.
- Pots accedir a una base de dades
- Pots modificar aquesta base de dades!
- Pots esborrar dades!





Elements bàsics d'una API

- **Accés:** Autorització a comunicar-se amb l'API -> KEY
- **Request:** Dada o servei que es demana a l'API. Conté dues parts:
 - **Mètode/Method:** Pregunta que li fas a l'API
 - **Paràmetres:** Informació extra
- **Response:** Resposta o servei que et proporciona l'API dependent de la teva Request





Operacions més comunes: CRUD

C → Create → **POST**

R → Read → **GET**

U → Update → **PUT**

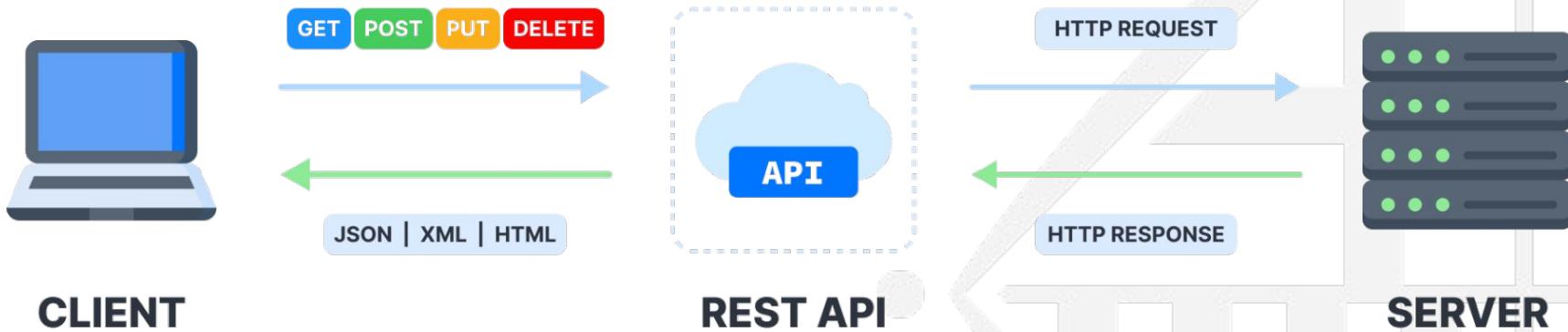
D → Delete → **DELETE**



HTTP
Methods



REST API Model



Podem accedir a l'API a partir d'un URL públic.

També podem accedir a diferent informació amb l'ús de la **query**.



Status Code

Quan l'API ens retorna el response sol està陪伴at d'un codi que ens aporta informació extra:

HTTP Status Codes





API Endpoints

Quan nosaltres estem cridant l'API, aquesta pot tenir diferents rutes que ens permeten accedir a diferent informació. Aquestes rutes s'anomenen endpoints. Mirem alguns exemples:

- Per accedir per exemple a la informació del client, podem utilitzar diferents endpoints:
 - Mètode HTTP serà **GET**
 - Per accedir a tots els clients, l'endpoint serà **/customers**
 - Per accedir a la informació d'un client en específic, podem tenir un altre endpoint com ara **/customers/<customer_id>**
- Podem tenir un mateix endpoint per diferents mètodes:
 - Amb endpoint **/customers** i el mètode **GET** ens retornen la llista de clients.
 - Amb endpoint **/customers** i el mètode **PUT** podem crear un nou client



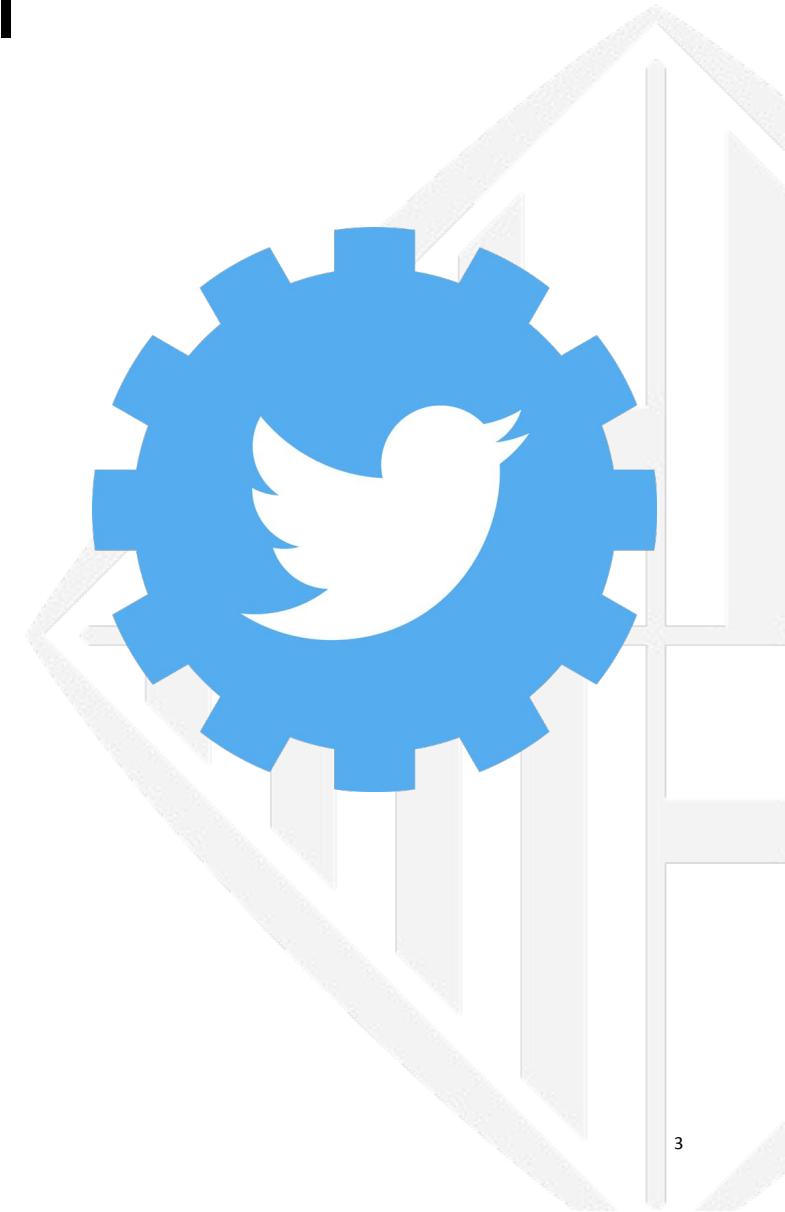
Twitter API

Twitter ens permet accedir a la informació de l'aplicació sense haver de fer web scraping (extreure el contingut directament de la pàgina web) utilitzant la seva [API](#).

Per fer-ho, primerament ens haurem de crear una conta de [Developer Twitter](#). Aquesta pot tardar uns dies a ser acceptada. Una vegada feta, podrem demanar les acreditacions per accedir a l'API!

Pots accedir a molt contingut diferent dins la pàgina web, fent servir els diferents endpoints, querys i request.

L'accés al contingut és limitat: no pots fer més de X request per minuts ni accedir a X quantitat de contingut al mes.





Twitter API + Python

- La gràcia de l'API és que qualsevol llenguatge de programació pot accedir a ella.
- A Python podem fer crides directes a l'API
- Per facilitar l'accés al contingut, Python té **llibreries** que ens ajuden
- Nosaltres podem fer servir la llibreria amb les nostres acreditacions de Twitter
- La llibreria que utilitzarem és **Tweepy**



02

Social Networks

Social Network Analysis





Twitter: A Social Network

Una **xarxa social** és una estructura social composta per individus (o organitzacions) anomenats «nodes» que estan connectats per un o més tipus d'interdependència com ara amistat, parentesc, interessos comuns, intercanvis financers, relacions sexuals, creences, coneixements, prestigi, etc. [\[Wikipedia\]](#)

Dins del món basat en internet, podem entendre com aquella aplicació que permet als usuaris comunicar-se amb altres compartint informació, comentaris, missatges, etc.



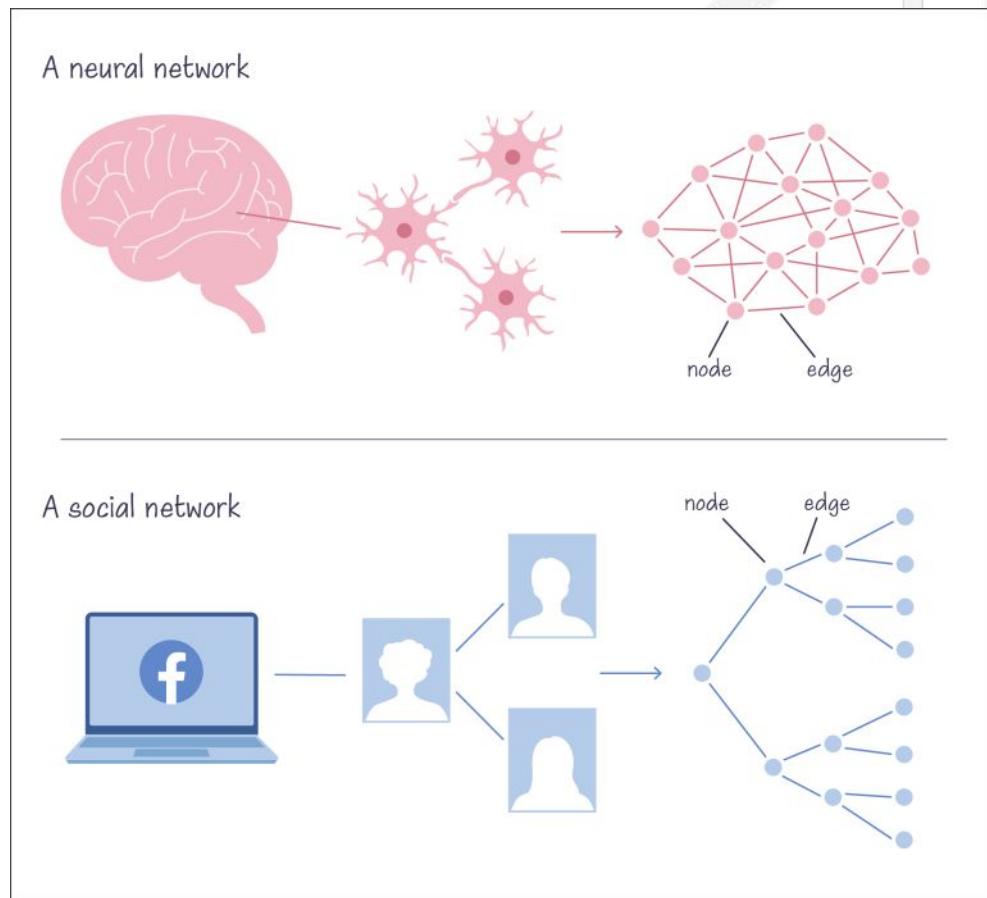
Twitter permet als seus usuaris interactuar amb petits posts anomenats “tweets”



Social Network Anàlisis

És possible estudiar patrons socials unint els coneixements de les ciències socials i el comportament, les matemàtiques, l'estadística i les ciències de la computació!

Les estructures socials es poden estudiar amb l'ús de les xarxes i la **teoria del graf!**



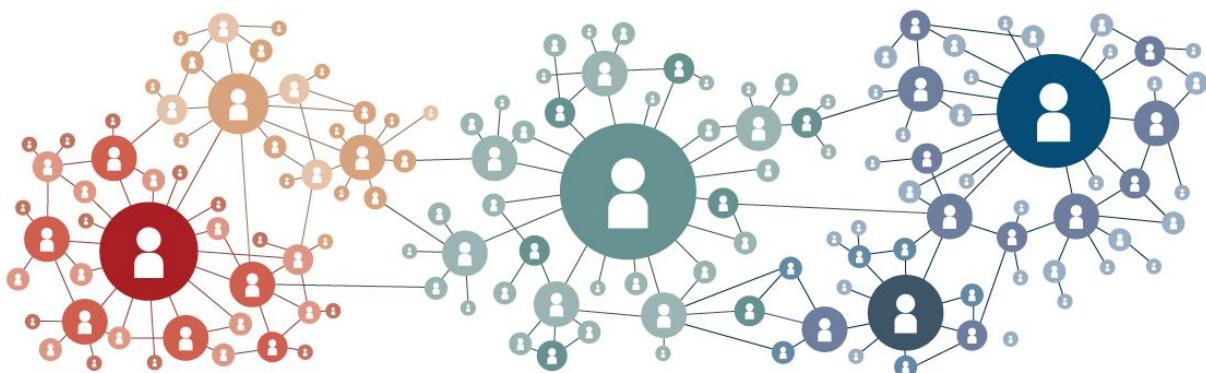


Teoria del Graf

La teoria de grafs és una branca de les matemàtiques i la informàtica que es dedica a l'estudi dels grafs, estructures matemàtiques amb l'objectiu de modelitzar relacions entre parells d'objectes.

El **graf** estarà compost per nodes o vèrtex (objecte) i les arestes (relació entre nodes).

Les arestes poden representar la relació entre nodes per distància, i els nodes poden representar certa jerarquia pel tamany.

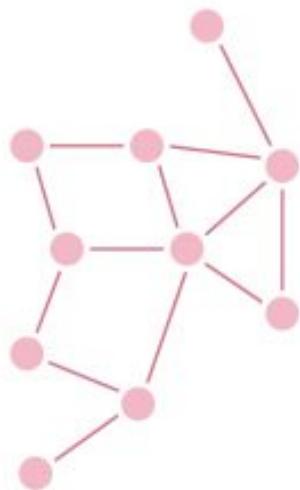




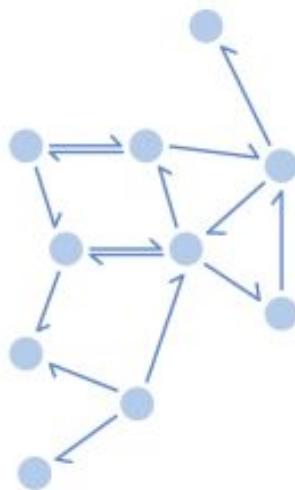
Característiques del Graf

Types of graphs

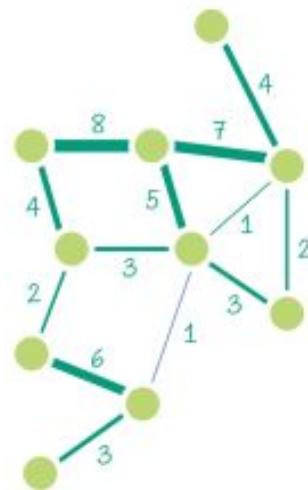
undirected



directed

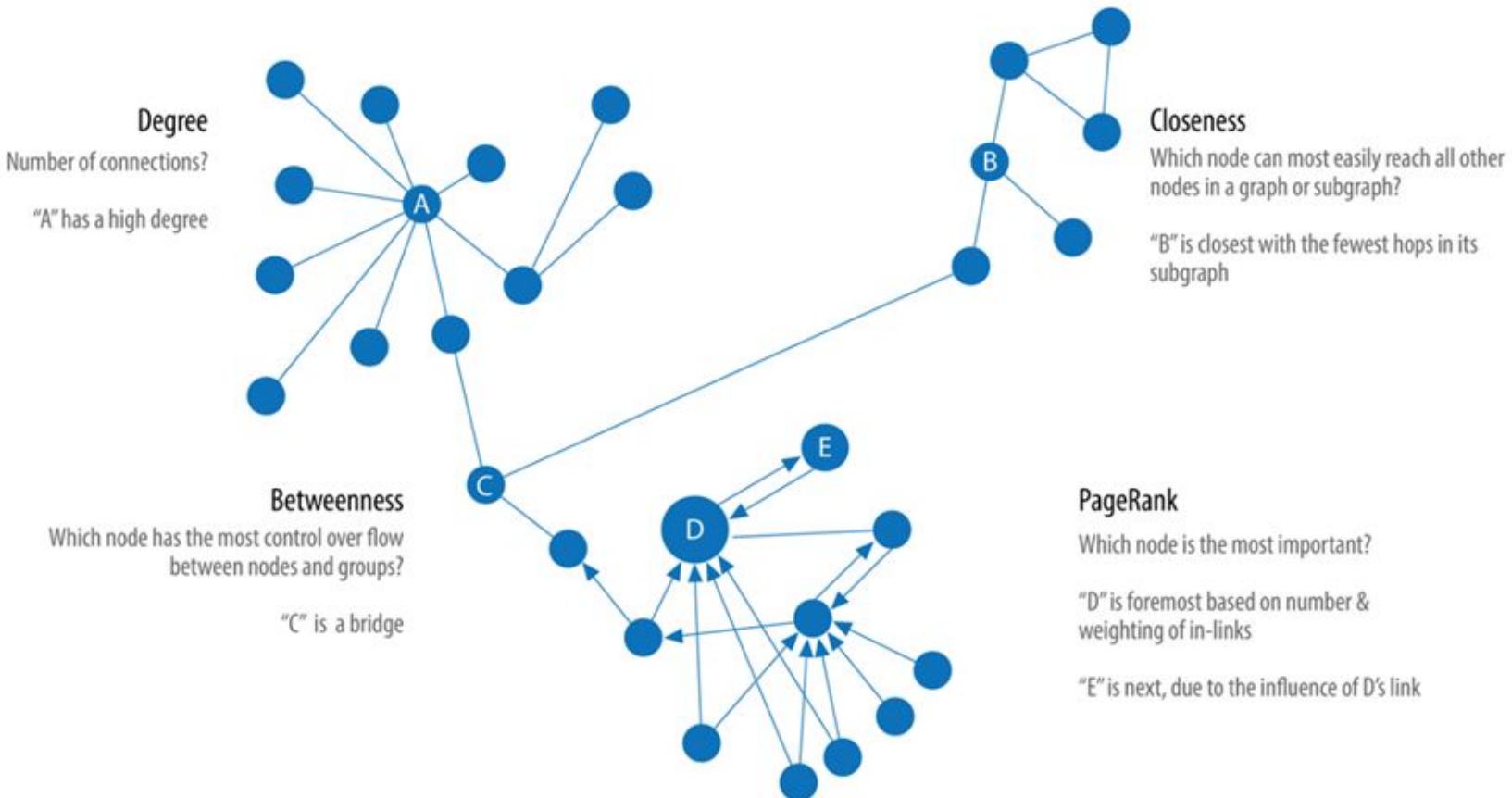


weighted





Característiques del Graf





Característiques del Graf

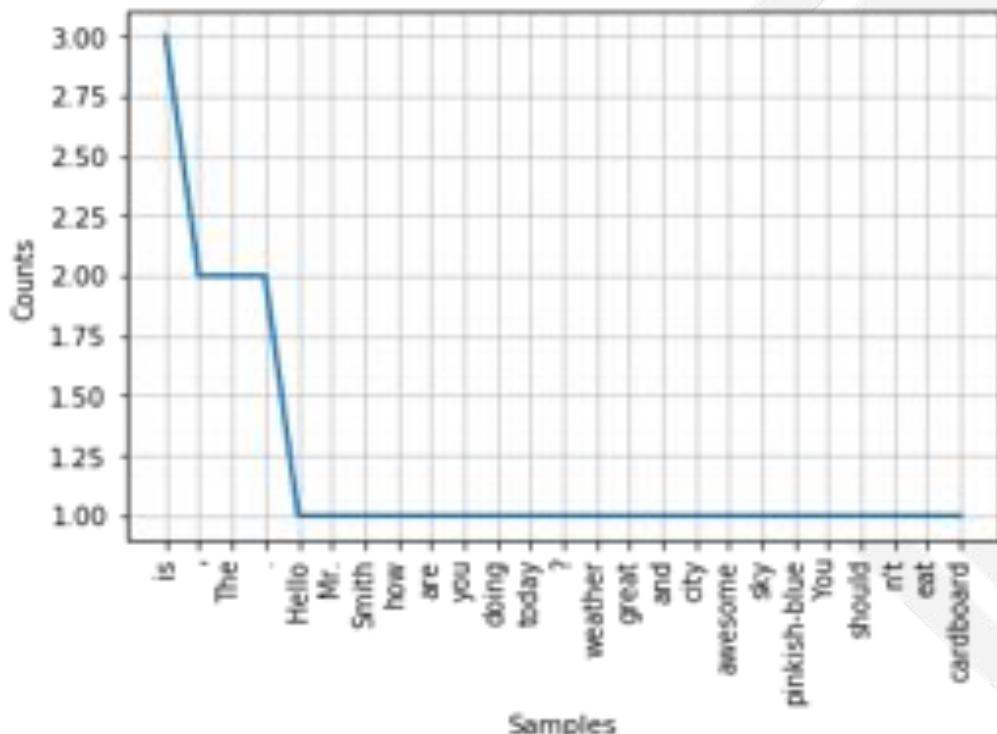
03

Analítica de text i processament del llenguatge natural

Llibreria NLTK

Tokenització

- De frase
- De paraula
- Distribució de freqüències



Stopwords o paraules crossa

Treure paraules crossa en l'idioma del text

Normalització

- Stemming: Stemming is a process of linguistic normalization, which reduces words to their word root word or chops off the derivational affixes. For example, connection, connected, connecting word reduce to a common word "connect".
- Lemmatization: Lemmatization reduces words to their base word, which is linguistically correct lemmas. It transforms root word with the use of vocabulary and morphological analysis. Lemmatization is usually more sophisticated than stemming. Stemmer works on an individual word without knowledge of the context. For example, The word "better" has "good" as its lemma. This thing will miss by stemming because it requires a dictionary look-up.

Part-of-Speech (POS) tagging

Part-of-Speech(POS) tagging is to identify the grammatical group of a given word. Whether it is a NOUN, PRONOUN, ADJECTIVE, VERB, ADVERBS, etc. based on the context. POS Tagging looks for relationships within the sentence and assigns a corresponding tag to the word.

```
[('Albert', 'NNP'), ('Einstein', 'NNP'), ('was', 'VBD'), ('born', 'VBN'), ('in',  
'IN'), ('Ulm', 'NNP'), ('.', '.'), ('Germany', 'NNP'), ('in', 'IN'), ('1879', 'CD'),  
('.', '.')]
```

Textblob

- Polarity [-1,1]
- Subjectivity [0,1]
- >>> testimonial = TextBlob("Textblob is amazingly simple to use. What great fun!")
- >>> testimonial.sentiment
- Sentiment(polarity=0.3916666666666666,
subjectivity=0.4357142857142857)
- >>> testimonial.sentiment.polarity
- 0.3916666666666666

```
'created_at',           # date and hour of tweet creation
'id',                  # tweet id (numeric)
'id_str',              # tweet id (string)
'text',                # tweet text (can be truncated)
'truncated',            # boolean if tweet is truncated or not
'source',               # source of publication (phone, computer)
'lang',                 # language of tweet (machine detected)
'in_reply_to_status_id', # if it's a reply, the id of the replied tweet, if not is null
'user_id',              # user id (numeric)
'user_id_str',          # user id (string)
'user_name',             # user full name
'user_screen_name',     # user screen name, used to identify the user
'user_location',         # user location
'user_url',              # url provided in user's profile
'user_description',      # user description in profile
'user_verified',          # boolean if user is verified or not
'user_followers_count',   # user number of followers
'user_friends_count',    # user number of people this account is following
'user_favourites_count', # user number of favourite tweets
'user_statuses_count',   # number of tweets (incl. retweets) by the user
'user_created_at',        # date and hour of user creation
'user_geo_enabled',       # boolean if geolocation is enabled
'user_lang',              # user language
'place_id',              # id representing the place
'place_url',              # url representing the location or additional place metadata
'place_place_type',       # type of location (example: city)
'place_name',              # short representation of place name
'place_full_name',         # long representation of place name
'place_country_code',      # shortened country code that contains the place
'place_country',            # name of the country containing the place
'place_bounding_box_coordinates', # coordinates enclosing this place
'extended_tweet_full_text', # if tweet is truncated, full text of the tweet
'extended_tweet_entities_hashtags', # array of hashtags parsed from the tweet
'extended_tweet_entities_urls',    # array of urls in tweet
'extended_tweet_entities_media' ,  # array of media contained in tweet (photo or video)
```

Wordclouds

indicaciones tarde
sabía ayudarnos
quizá limitaciones
tercios falsa arterias sobre
pueda **afectación** poeta
constante tratamiento
carísimos iba aguda
 parches iba **aortica** aguda
segura Italia mal
terapia chia arterial diario rotura
impresión holanda intimamente
ganas llevamos pecho
responsable únicos
sensibilidad.

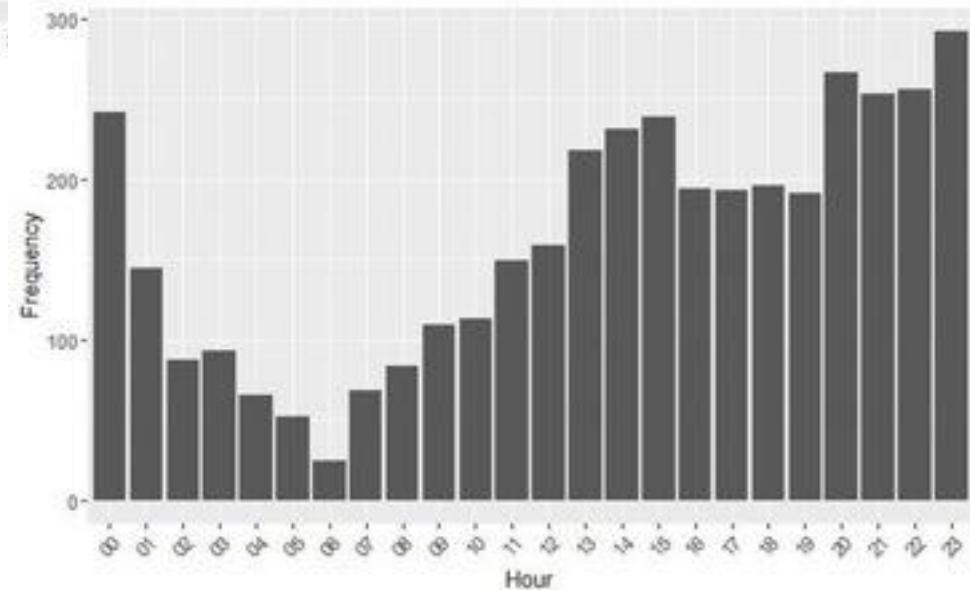
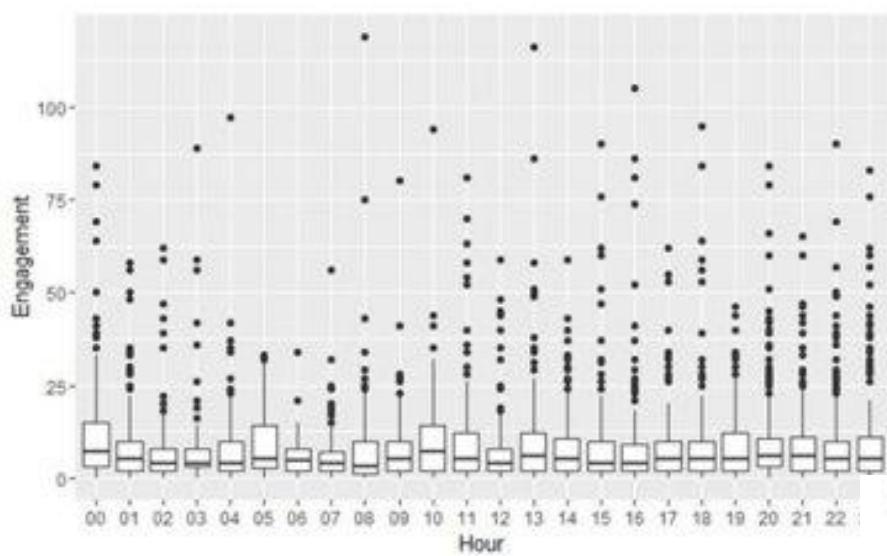
prestamo prestado
recibo otorgo asequibles
se concede enviodeuda
amortización globo
tasa usted dama

constuir plazos
advertencia el envío
sabiduría préstamos
electrónico finanzas matrícula
matrícula

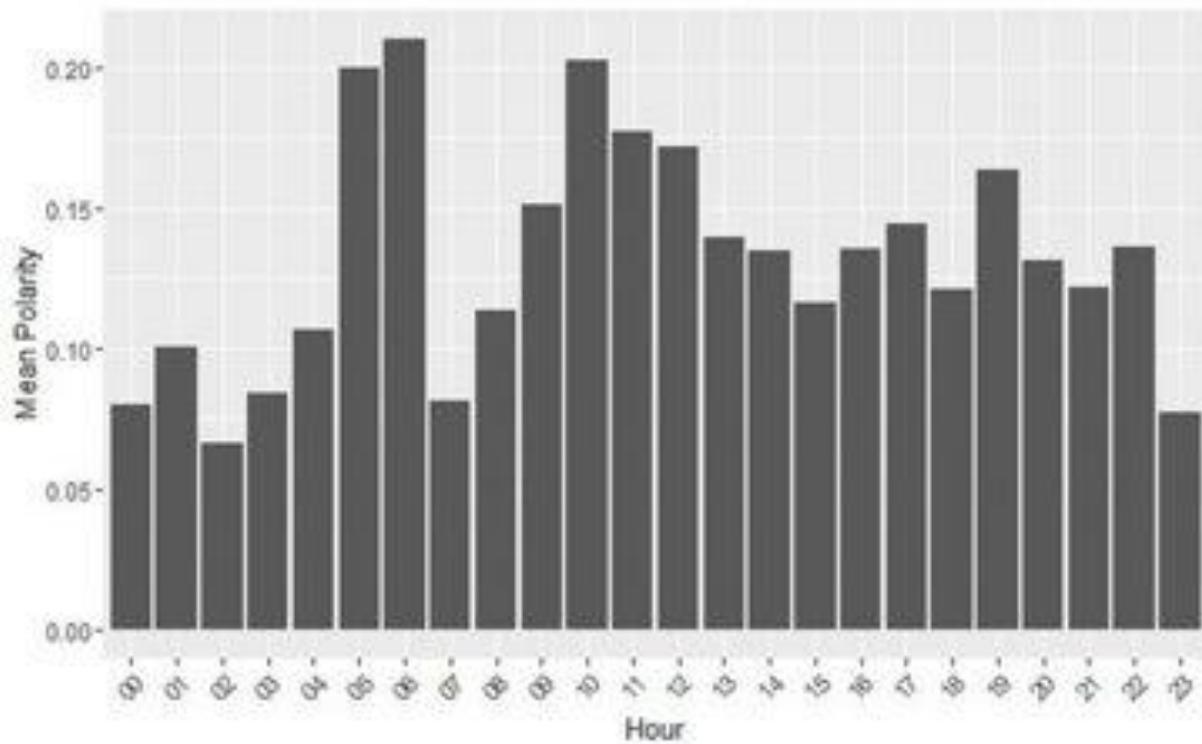
significativa información
programas elásticas discapacidad
aortica efectos pueda urinarias
mejorar vejiga **después** people
articulares **alimentaries** acabo lleva
badas **tratamiento** vías
bebé tvca **muerto** urinario
lesiones orina examen
ghent autoritat dilatación cero
intolerancies régimen formación interesado
syndrome andalucía heparina
hiperaxitud prolapse

referencia reconocimiento establecer
sanitarios situación fuentes implementar inclusión profesionales
información partida sanitaria personas **formación** **CSUR** pacientes
trabajo impulsar **nacional** raras
sociales **atención** nivel
enfermedades discapacidad gastos
estrategia prioriza servicios acceso
protocolos garantizar específica
promover derecho objetivos
financiación fomentar
trabajadores **atención** participación
investigación presupuestaria

Temporal analysis



Polarity



Dies mundials UNESCO

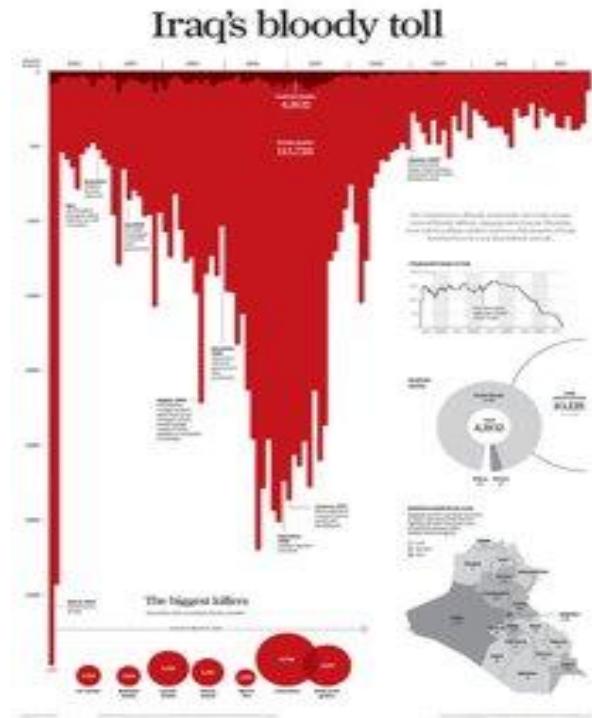
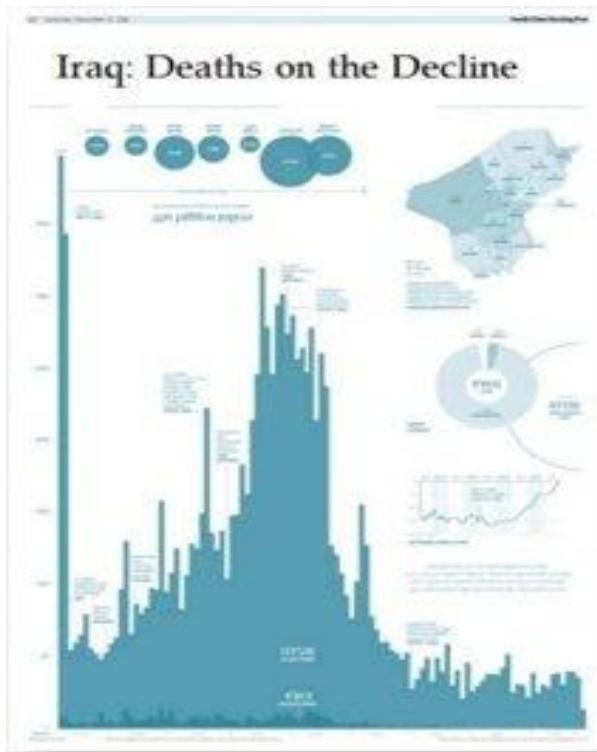
<https://www.unesco.org/en/days>



Data-to-viz

- <https://www.data-to-viz.com>
- <http://graphics.wsj.com/infectious-diseases-and-vaccines/>

En una imatge sempre hi ha una història



Análisis de la depresión y la ansiedad en el aborto en Twitter

<http://hdl.handle.net/10609/138932>

- Anàlisi de tuits seguint terminologia clínica
- Núvols de paraules segons els temes (trobar nombre òptim de temes)
- Important: com “netejar” els tuits i utilitzar keywords

Gràcies per la vostra col·laboració!



Ajuntament
de Barcelona

