

# Data Mining

## PageRank

<http://www.salle.url.edu>

**Enginyeria La Salle – Universitat Ramon Llull**

### **Pràctica 1 (Week 3): Trobar el *PageRank* de pàgines de California**

## Objectiu

Els principals objectius d'aquesta pràctica son:

- Agafar familiaritat amb l'entorn Python
- Prendre consciència de la complexitat del coneixement en Xarxa
- Implementar un algorisme a partir d'equacions formals
- Conèixer l'algorisme PageRank i manegar els seus paràmetres

## Requirements

La pràctica es pot fer en qualsevol sistema operatiu: Windows / Mac OS X / Linux però es necessita d'un intèrpret python. Sinó esteu acostumats a l'entorn Python adreceu-vos a la guia penjada a l'e-study:

<https://estudy.salle.url.edu/mod/resource/view.php?id=463912>

Els recursos per fer la pràctica es troben a:

<https://estudy.salle.url.edu/mod/resource/view.php?id=463919>

<https://estudy.salle.url.edu/mod/resource/view.php?id=463920>

El pou es troba a:

<https://estudy.salle.url.edu/mod/assign/view.php?id=463921>

Si teniu algun problema, comenceu una discussió a :

<https://estudy.salle.url.edu/mod/forum/view.php?id=463902>

També recomano l'ús d'un editor avançat de programació tals com:

**Sublime Text Editor** : <http://www.sublimetext.com/>

**Notepad++** : <http://notepad-plus-plus.org/>

**Temps estimat**: 1,5 hores. Maxim 3 hores

**Deadline**: 18 d'Octubre a les 23:59:59 CET

## Descripció

La pràctica proporciona els següents recursos:

- **Assignment1.pdf:** Enunciat de la pràctica
- **processPageRank.py:** Esquelet del codi per computar PageRank
- **gr0.California:** Graph de totes les web implicant la cerca California, amb dues parts diferenciades:
  - n            *ID*            *URL*
    - on ID defineix un id de node i URL la seva web
  - e            *IDFROM*            *IDTO*
    - on indica que la pàgina IDFROM enllaça a la plana IDTWO

**Funcionament:**

1. Intenta entendre el codi de l'esquelet que s'ha proporcionat
2. Identifica les parts on has d'implementar el PageRank
3. Implementa l'algorisme de PageRank de classe (a la Pantalla) l'algorisme en el codi
4. Juga amb el paràmetre  $\beta$  i observa com varien els resultats. Fes una reflexió personal en el fitxer README.txt
5. Comprimeix el codi python i README.txt i puja el codi dins el pou amb un zip amb el teu nom i login.
  - Nota: La pràctica se t'avaluarà amb una Beta determinada tinguent en compte el temps de computació els resultat obtingut

Felicitats! No només has après *Python* sinó que també has après a extreure dades en xarxa via PageRank!