

laSalle

UNIVERSITAT RAMON LLULL

PROGRAMACIÓ AVANÇADA I ESTRUCTURES DE DADES



InstaSalle

An algorithm must be seen to be believed.

Donald Knuth

1 Introducció

InstaSalle és una xarxa social desenvolupada per compartir imatges en format de post. L'aplicació permet als usuaris del sistema fer fotografies per mostrar-les a les seves amistats o seguidors. A part de permetre compartir publicacions, també ofereix la possibilitat d'interactuar amb aquestes, on es podran comentar publicacions d'altres usuaris i donar-li m'agrades.

En la passada actualització es va decidir innovar amb l'algorisme d'ordenació del *feed* de l'usuari. Aquesta és la pantalla principal on el mateix veurà les publicacions dels usuaris als que segueix.

L'algorisme tenia l'objectiu d'endregar els posts que veurà l'usuari segons els seus interessos. Així doncs, la intenció d'aquest, és que a l'*instasallegramer* se li mostrin les publicacions que més interès potencial tingui per ell. D'aquesta manera, tindrà més prioritat el contingut dels usuaris amb els que més interacció es tingui.

Encara que els enginyers estiguessin molt orgullosos del seu algorisme, han rebut moltes queixes per part dels usuaris del sistema. És per això que, per por a la fallida de l'empresa, han demanat ajuda als estudiants de La Salle.

Amb la intenció d'ajudar-los, s'ha encomanat als alumnes de l'assignatura de Programació avançada i estructures de dades realitzar diferents propostes d'algorismes d'ordenació, tenint en compte altres criteris com ara la temporalitat o la localització de les publicacions.

2 Format de les dades

Els *datasets* proporcionats estaran en format JSON i contindran la informació dels usuaris del sistema.

Cada usuari tindrà:

- followers: número de seguidors
- follows: número de persones a qui segueix
- connections: relacions d'un mateix contra els altres usuaris
 - username: persona amb qui es relaciona
 - since: des de quan el segueix. Aquesta dada estarà en format **timestamp**
 - visits: quantes vegades ha consultat el seu perfil
 - likes: quants m'agrada li ha proporcionat a les seves publicacions
 - comments: quants comentaris li ha proporcionat a les seves publicacions
- Posts: informació de les publicacions d'un mateix
 - id: identificador del post
 - liked_by: usuaris que han donat m'agrada al post
 - commented_by: usuaris que han comentat el post
 - published: quan es va publicar. Aquesta dada estarà en format **timestamp**
 - location: coordenades des d'on s'ha publicat. Aquestes dades vindran donades per [latitud, longitud]
 - category: categoria identificativa de la publicació. Les categories que es tindran en compte seran les següents:

* Landscape	* Animals	* Music
* Food	* TV Shows	* Travel
* Sports	* Fitness	
* Style	* Science & Tech	* Architecture

3 Funcionalitats

Tal com s'ha esmentat en la introducció, caldrà realitzar diversos algoritmes d'ordenació tenint en compte diferents criteris. Els algoritmes que s'implementaran seran: **Merge Sort**, **Quicksort**, **Selection Sort** i **Radix Sort**. S'especificarà l'algoritme que s'emprarà com a argument del programa.

Segons temporalitat

Els posts es mostraran segons la data de publicació d'aquests on els més recents es mostraran abans.

Caldrà especificar com a argument del programa que es tracta d'una ordenació segons temporalitat. En aquest cas, s'ordenaran tots els posts en general. En conseqüència, no caldrà especificar l'usuari del qual s'haurien d'ordenar els posts.

Segons ubicació

En aquest cas, l'usuari haurà d'introduir una ubicació de referència i s'hauran de mostrar segons proximitat.

Caldrà especificar com a argument del programa que es tracta d'una ordenació segons ubicació. A més a més, caldrà especificar també la latitud i longitud que s'emprarà com a referència. En aquest cas, s'ordenaran tots els posts en general. En conseqüència, no caldrà especificar l'usuari del qual s'haurien d'ordenar els posts. Per tal de calcular la distància entre punts, es recomana fer ús de la fórmula del *Haversine*.

Segons una combinació de prioritats

En aquest cas es tindrà en compte la interacció d'un usuari amb els altres usuaris. La data de publicació dels posts serà un factor important, en tant que interessarà veure els posts més recents de les persones amb qui l'usuari interacciona. Així doncs, s'indicarà al sistema de quin usuari es tracta per poder elaborar el llistat com a argument del programa.

Per exemple, si un usuari té molt contacte amb un usuari però aquest publica només cada dos mesos, interessarà veure publicacions d'altres persones abans. En conseqüència, la temporalitat jugarà un paper clau.

D'aquesta manera, caldrà ordenar els posts que hauria de veure un usuari. Aquests posts seran les publicacions realitzades pels usuaris als quals segueix. Així doncs, serà necessari ponderar cada post, per posteriorment ordenar-los segons l'esmentada ponderació.

Es calcularà aquest pes segons:

- La *likelihood* que té l'usuari en estar interessat en el post segons la categoria

- La interacció que es té amb el publicador del post
- La temporalitat del post

Per saber l'interès potencial serà necessari revisar a quines publicacions ha donat m'agrada l'usuari. Així doncs, es podrà predir quin serà el contingut que li pugui interessar més. D'altra banda, serà necessari revisar la interacció de l'usuari amb les seves connexions. Finalment es tindrà en compte la data de publicació del post.

Encara que el criteri a seguir no estigui definit explícitament, es recomana fer ús de pesos parcials per aconseguir el pes total. A més a més, es recomana abordar el càlcul de forma fragmentada fent ús d'estructures de dades com ara, llistes o, per aquells que vulguin anar un pas més enllà, taules de hash.

El criteri seguit per elaborar els pesos haurà d'estar correctament detallat en la memòria. En l'apartat de bibliografia es detallaran enllaços que us poden ser útils.

4 Consideracions

4.1 Grups

La pràctica es pot realitzar de manera individual o en parelles.

4.2 Llenguatge

La pràctica es pot implementar en el llenguatge de programació que vingui de gust.

4.3 Detecció de CÒPIA

Una còpia és un insult i falta de respecte no només al professorat sinó a la resta de companys que han dedicat esforç i hores per realitzar la feina honestament.

Una pràctica de l'assignatura es considera una activitat altament significativa d'aquesta. **La detecció d'una còpia, sigui aquesta parcial o total; d'un company o Internet, es considerarà una acció premeditada.** Per aquests motius, aplicant la Normativa de Còpies de la universitat, la còpia es considerarà una **falta molt greu** implicant les següents accions:

1. Qualificació de 0 en la convocatòria oficial.
2. Pèrdua de la següent convocatòria oficial (amb un NP).
3. Obertura de l'expedient acadèmic.

5 Entrega

5.1 Format

S'haurà d'entregar un arxiu **ZIP** al pou corresponent a l'eStudy amb el següent contingut:

- Carpeta amb el codi i/o estructura del projecte generat per l'IDE.
- Memòria: arxiu **PDF**.
- README: arxiu **TEXT** o **MD** on s'expliqui tot el que fos necessari pel testeig del codi entregat (llenguatge, versió d'aquest, IDE de desenvolupament, com executar-lo, etcètera)

5.2 Memòria

La memòria ha de contenir els apartats següents:

- Portada (amb el nom complet i *login* dels membres del grup)
- Índex numerat (apartats i número de pàgina que ocupa cadascun dins el document)
- Explicació **detallada** de com s'han codificat els criteris d'ordenació i les estructures de dades emprades
- Comparativa d'algoritmes d'ordenació acompanyat de gràfiques
- Mètode de proves utilitzat (explicació del mètode de proves que s'han seguit)
- Problemes observats
- Conclusions (a nivell personal i a nivell tècnic)
- Bibliografia (segons la norma ISO 690:2010)

La memòria ha d'estar escrita amb un llenguatge formal i sense faltes d'ortografia. Aquesta és tant o més important que la codificació en si, atès que reflecteix el coneixement assolit. L'objectiu d'aquesta pràctica no és només aprendre a codificar algoritmes d'ordenació sinó que s'apreguin els avantatges i inconvenients de cada algoritme. Per tant, l'apartat de comparativa d'algoritmes tindrà un pes important.

Es recomana fer testos amb els vostres codis i que els resultats d'aquests es mostrin en taules que es puguin comentar. D'altra banda, es tindrà en compte que el contingut de la memòria sigui de qualitat i no de quantitat. Es recomana fer ús de $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$.

Per la comparativa dels algorismes es proporcionarà un conjunt de *datasets*. Podeu, a més, usar els vostres propis *datasets* per la comparativa.

5.3 Data

La data màxima d'entrega de la pràctica sobre 10 serà el dia 11 de Desembre a les 23:55h. No s'acceptarà cap entrega posterior a aquest dia i hora.

5.3.1 Entrega no realitzada

En cas de no entregar la pràctica a temps s'obtindrà una nota de NP i la pràctica haurà de ser entregada com a tard el 17 de febrer de 2019 a les 23:55h amb una nota màxima de 5. Després d'aquesta data no existiran més oportunitats d'entregar, fet que implicaria un suspens directe de l'assignatura.

5.3.2 Entrega suspesa

Si una entrega sobre 10 es troba suspesa sigui per memòria o perquè alguna funcionalitat no funciona correctament, en funció del cas es proposarà una entrega extra amb una nota màxima d'entre 6 i 9.

Per tant, si tot i no haver aconseguit acabar la pràctica o fer-la funcionar en la seva plenitud hi ha prou esforç darrera i bastant de feina realitzada ¹, s'aconsella entregar sobre 10 posant en el README què ha fallat per tal d'estudiar el cas i proposar l'entrega extra.

Ara bé, si es detecta un intent d'aconseguir una entrega extra en una pràctica que no funciona res o es detecta que no hi ha hagut una implicació al darrere per tal d'arribar a l'entrega sobre 10, no s'atorgarà l'entrega extra anant directe sobre 5. A més, per la persona o persones implicades en aquesta entrega no s'estudiarà cap possible reentrega de pròximes pràctiques/projecte.

¹Només s'estudiarà el cas si s'entrega una memòria formal

References

- [1] Alfred Lua. *How the Instagram Algorithm Works in 2018: Everything You Need to Know* <https://blog.bufferapp.com/instagram-feed-algorithm>, 2018.
- [2] -. *Distance on a sphere: The Haversine Formula*. <https://community.esri.com/groups/coordinate-reference-systems/blog/2017/10/05/haversine-formula>, 2017.
- [3] Prateek Garg. *DBasics of Hash Tables*. <https://www.hackerearth.com/practice/data-structures/hash-tables/basics-of-hash-tables/tutorial/>, -.