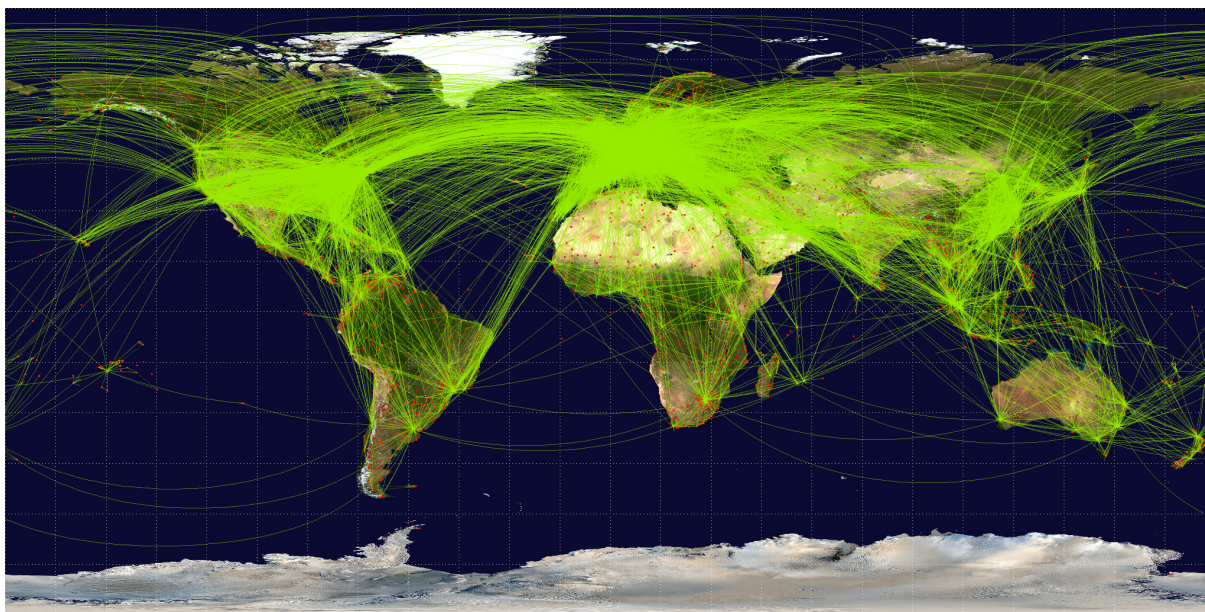


Pràctica d'Estructures de Dades

Estructures de Dades i Algorítmica (3105G07010)

Grau en Enginyeria Informàtica - Grau en Disseny i Desenvolupament de Videojocs

Curs 2019-20



Estructura de Dades d'Aviació

Avui en dia existeixen moltes bases de dades amb informació de vols, algunes d'elles de franc i d'altres de pagament. A <https://openflights.org/data.html>, per exemple, podem trobar una bona quantitat de dades gratuïtes d'aeroports, companyies aèries i rutes. D'altra banda, <https://www.transtats.bts.gov> és un web del departament de transport dels EUA que permet filtrar dades de vols dels EUA i extreure'n estadístiques de manera completament lliure i gratuïta.

L'objectiu d'aquesta pràctica és treballar el disseny d'estructures de dades eficients per processar consultes sobre grans volums de dades d'aviació en memòria principal.

Funcionament del programa

El nostre programa haurà de carregar una sèrie de fitxers de text i processar a continuació una sèrie de consultes sobre les dades llegides.

Llegirem el nom dels fitxers (pot incloure la ruta sencera) pel canal d'entrada estàndard (una línia per cadascun) en el següent ordre:

```
aeroports
companyies aèries
avions
rutes
vols
consultes
```

Un exemple podria ser:

```
../openflights/airports.dat
../openflights/airlines.dat
../openflights/planes.dat
../openflights/routes.dat
../472598099_T_ONTIME_REPORTING/T_ONTIME_REPORTING_2019_6.csv
./queries.txt
```

Les consultes estaran contingudes en el darrer fitxer (`./queries.txt` a l'exemple). D'aquesta manera serà més fàcil automatitzar les proves. Noteu que, fins i tot, si les rutes dels fitxers estan contingudes al seu torn en un fitxer de text, de nom per exemple `files.txt`, podrem executar el programa des de línia de comanda fent simplement:

```
bash-4.2$ ./main < files.txt
```

Assumirem que ens donen les rutes dels fitxers en l'ordre correcte, i que el seu format és correcte (no cal fer control d'errors).

- El format dels quatre primers fitxers (aeroports, companyies aèries, avions, i rutes) és l'especificat al web <https://openflights.org/data.html>
- El format del cinquè fitxer (vols) el trobareu descrit al Moodle en un fitxer anomenat `T_ONTIME_REPORTING_ReadMe.csv` (els codis que hi apareixen són codis IATA).
- El format del sisè fitxer (consultes) està descrit a l'apartat "Consultes" d'aquest document.

A la carpeta `/u/prof/u2001239/eda/p1/` de `bas.udg.edu` us hem deixat un executable i uns fitxers d'exemple per tal que pugueu veure quin és el resultat esperat. El format de sortida del vostre programa és lliure però, per tal de facilitar la correcció, us agraïrem que s'assembli el màxim possible al de l'executable que us proporcionem.¹

Per executar el programa dels professors des del vostre compte de `bas.udg.edu`, podeu fer:

```
bash-4.2$ /u/prof/u2001239/eda/p1/main < /u/prof/u2001239/eda/p1/files.txt
```

Si feu:

```
bash-4.2$ cat /u/prof/u2001239/eda/p1/files.txt
```

¹Si detecteu algun error us agraïrem que ho comenteu al fòrum.

podreu veure les rutes dels fitxers utilitzats i copiar-los si us interessa amb la comanda `cp`. Tingueu present però que el fitxer de vols és molt gran (>80MB) i la vostra quota de disc a bas és limitada.

De fet no cal ni tan sols que copieu els fitxers. Podeu executar el vostre programa sobre els fitxers que hi ha al compte dels professors fent:

```
bash-4.2$ ./main < /u/prof/u2001239/eda/p1/files.txt
```

suposant que `./main` sigui el vostre executable.

Si no voleu treballar a `bas.udg.edu`, podeu baixar-vos els quatre primers fitxers (aeroports, companyies aèries, avions, i rutes) del web <https://openflights.org/data.html>. El fitxer de vols el trobareu al Moodle.

Consultes

Les consultes que hem de ser capaços de processar són les següents (el format s'indica a sota):

1. Comptar quantes rutes nacionals i internacionals té cada companyia aèria, i llistar el nom de les companyies (segons l'ordenació natural dels *strings*) junt amb el seu nombre de rutes nacionals i internacionals. Ometre les companyies que no tinguin cap ruta. Una ruta es considera nacional si tant l'aeroport de sortida com l'aeroport d'arribada estan en el país de la companyia aèria que la cobreix.

"COUNT ROUTES"

2. Quin és el model d'avió més utilitzat per determinada companyia aèria (el que apareix en més rutes; tots els més utilitzats en cas d'empat).

"MOST USED PLANE", IATA

3. Llistat ordenat dels *n* (0 per indicar tots) aeroports amb més retards de sortida (suma de retards sense comptar sortides abans d'hora) en una data determinada (ordre decreixent, i indiferent en cas d'empat).

"DEPARTURE DELAY AIRPORTS", *n*, *year*, *month*, *day_of_month*

4. Llistat ordenat dels *n* (0 per indicar tots) aeroports amb més retards d'arribada (suma de retards sense comptar arribades abans d'hora) en una data determinada (ordre decreixent, i indiferent en cas d'empat).

"ARRIVAL DELAY AIRPORTS", *n*, *year*, *month*, *day_of_month*

5. Llistat ordenat de les *n* (0 per indicar totes) companyies aèries amb més retards de sortida (suma de retards sense comptar sortides abans d'hora) independentment de la data (ordre decreixent, i indiferent en cas d'empat).

"DEPARTURE DELAY AIRLINES", *n*

6. Llistat ordenat de les n (0 per indicar totes) companyies aèries amb més retards d'arribada (suma de retards sense comptar arribades abans d'hora) independentment de la data (ordre decreixent, i indiferent en cas d'empat).

"ARRIVAL DELAY AIRLINES", n

7. Dia amb més retards tant de sortida com d'arribada (sense comptar sortides i arribades abans d'hora) en un subconjunt donat d'aeroports (conjunt buit per indicar tots els aeroports). En cas d'empat llistar tots els dies amb retards màxims.

"DAY WITH MORE DELAYS",IATA_1,...,IATA_N

Veure el fitxer /u/prof/u2001239/eda/p1/queries.txt per un exemple. Noteu que les consultes poden venir en qualsevol ordre.

Observacions

- Tractant-se d'una pràctica d'estructures de dades, es valorarà l'adequació de les estructures de dades utilitzades d'acord amb la informació a emmagatzemar i a les consultes previstes.
- Es prohibeix guardar consultes precalculades (informació redundant). Cal calcular la resposta a cada consulta en el moment que es produeixi la consulta.
- Noteu que la lectura dels fitxers és delicada: tenim camps separats amb ',' però poden haver-hi "cadenaes" amb ',' a dins. Per tal de facilitar la lectura us proporcionem la funció getNextLineAndSplitIntoTokens al Moodle.
- Cal tenir present que poden haver-hi camps buits: ,, o bé "",. Aquests seran transformats en *strings* buits per la funció getNextLineAndSplitIntoTokens.
- Cal decidir com tractar \N (NULL en SQL) en els camps numèrics (0, -1, ...)
- En alguns fitxers cal ignorar la primera línia (capçaleres, ...)
- Quan processem les rutes haurem de buscar companyies aèries i aeroports per ID o bé codi IATA o bé codi ICAO, i l'equipament per codi IATA. En el cas que no existeixi la companyia, algun dels dos aeroports, o algun equipament (model d'avió), ignorarem la ruta.
- De la mateixa manera, ignorarem els vols de companyies aèries o aeroports desconeguts.

Instruccions de lliurament

Caldrà penjar al Moodle un fitxer comprimit **zip** o bé **tgz** amb un nom que segueixi el format

CodiUsuari_CognomsNom_p1

d'acord amb el document "*Instruccions per al lliurament de les activitats de laboratori*". Aquest fitxer comprimit ha de contenir:

1. El codi font.
2. Un Makefile que permeti generar l'executable a partir del codi font.
3. Un joc de proves. Com que l'espai de disc és limitat, podeu lliurar fitxers de dades petits, o bé utilitzar els fitxers disponibles a l'espai dels professors i modificar tan sols les consultes.
4. Un *script* que permeti executar el vostre programa sobre el joc de proves (executant una prova darrera l'altra, sense preguntar res).
5. Un fitxer `readme.txt` amb aclariments que puguin facilitar la correcció. Sobretot, si hi ha alguna cosa que no funciona bé, expliqueu-ho en el `readme.txt` en lloc de posar comentaris al lliurament del Moodle (que a vegades passen desapercebuts). Encara que funcioni tot bé, digueu-ho en un fitxer `readme.txt`.
6. Una carpeta `html` amb la documentació generada amb doxygen² (eina que presentarem properament a classe de laboratori).
7. Un document `informe.pdf` amb una explicació i justificació detallada de les estructures de dades utilitzades.

A més del lliurament al Moodle, al directori `~/eda/p1` del vostre compte de `bas.udg.edu`, hi heu de posar, descomprimit, el mateix que hagueu penjat al Moodle, exceptuant la documentació (punts 6 i 7), que ocupa massa i a bas no ens és útil.

Tots els fitxers han d'estar en format Unix.

Consulteu el document *"Instruccions per al lliurament de les activitats de laboratori"* per més detalls.

La data límit de lliurament està especificada a l'activitat corresponent del Moodle.

²<http://www.doxygen.nl/>