

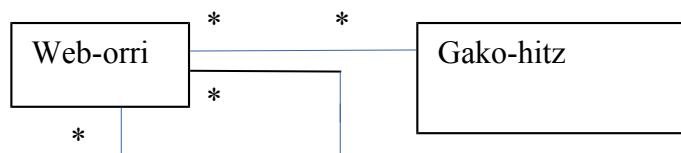
Datu-Egiturak eta Algoritmoak: proiektua

Web-a kudeatzeko aplikazioa

Helburuak

Web-orri kopuru handia (milaka, hamar milaka edo ehun milaka) kudeatuko dituen aplikazioa sortu nahi dugu. Datuak Web Data Commons - Hyperlink Graph 2012tik hartu dira - Download Instructions: <http://webdatacommons.org/hyperlinkgraph/2012-08/download.html>

Web-orri bat beste web-orri batzurekin lotuta dago esteken bitartez. Hurrengo domeinu-eredua erabili behar da.



Hau da, web-orri batek beste hainbat web orritara nabigatzeko estekak ditu. Halaber, web-orri bat zenbait gako-hitzekin lotuta dago.

Web-orrien fitxategietako lerroak honelakoak dira:

"index" fitxategia	"pld-arc-1-N" fitxategia
0-00.pl 0	0 --> 777835 781154 795437 862696 1102570
0-100.com.cn 1	1 --> 598847 1297833 1438234 2136133
0-100editions.net 2	2 --> 413745 1141599
0-18.gr 3	...
0-200.com 4	
0-24-sex.de 5	
0-24.ro 6	
0-3-6.com 7	
0-311.com 8	
0-360.com 9	
0-3ani.ro 10	
0-5-30.com 11	
0-5.co.il 12	
0-5ans.com 13	
0-60cartimes.com 14	
0-700.pl 15	
0-75.pl 16	
0-adult.net 17	
0-apr-creditcards.com 18	
0-art.co.uk 19	
0-brune-nue.com 21	
...	

“index” izeneko fitxategian osoko bat (zerotik hasita) esleitzen zaio web-orri bakoitzari. “pld-arc-1-N” izeneko fitxategiak “ $x \rightarrow u v w$ ” motako lerroak ditu. Honek esan nahi du x web-orrian u , v eta w web orrietara nabigaltzeko estekak daudela.

“words.txt” izeneko fitxategiak hitz-zerrenda posible guztiak dauzka:

```
credere
credibilities
credibility
credible
credibleness
credibly
credit
creditabilities
creditability
creditable
creditableness
creditably
credited
crediting
creditive
...
```

Funtzionalitateen artean honako eragiketak izan ditzakegu:

- web-orri baten bilaketa
- elementu berri bat gehitu (web-orri edo gako-hitz)
- web-orrien zerrenda ordenatua lortu

Zehazki, hauek izan daitezke funtzionalitate batzuk:

- String **id2String**(int x)
// Postbaldintza: osoko bat emanda, dagokion web-orria itzuliko du.
Adibidez: id2String(18) \rightarrow 0-apr-creditcards.com
- int **string2Id**(String s)
Adibidez: string2Id(“0-apr-creditcards.com”) \rightarrow 18
- ArrayList<String> **irteerakoEstekak**(String s)
// post: web-orri baten izena emanda, estekatzen dituen web-orriak itzultzen ditu.
- ArrayList<String> **webOrdenatua**()
// post: web-orrien zerrenda itzultzen du, alfabetikoki ordenatuta.
// ez du web-orrien jatorrizko zerrenda aldatzen.
- ArrayList<String> **word2Webs**(String s)
// pre: “s” gako-hitz bat da
// post: “s” gako-hitza daukaten web-orriak itzultzen ditu
Adibidez: word2Webs(“money”) \rightarrow
 <007waystomakemoney.com, 1000moneymakingideas.com, ...>
- ArrayList<String> **web2Words**(String w)
// post: “w” web-orrian agertzen diren gako-hitzak itzultzen ditu.
Adibidez: web2Words(“1000moneymakingideas.com”) \rightarrow
 <money, making, ideas>
- ...

1. eginkizuna

Elementuen karga, bilaketak eta zerrendatzea egingo dituen sistemaren diseinu eta implementazioa.

Azken helburua: sistemak honako eragiketak egingo ditu modu eraginkorrean:

- Datuak kargatu fitxategietatik
- Web-orri baten bilaketa
- Web-orri berri baten txertaketa
- Web-orri bat ezabatu
- Web-orri bat emanda, honek estekatzen dituen web-orrien zerrenda bueltatu
- Gako-hitz bat emanda, gako-hitz hau duten web-orrien zerrenda bueltatu
- Web-orrien zerrenda fitxategitan gorde
- Web-orrien zerrenda ordenatua lortu (alfabetikoki)

Hau entregatu beharko da:

- Eskatutakoa inplementatzen duten programak (zuzen exekutatu dira)
- Dokumentazioa, hasierako problema, aztertutako aukerak, inplementazioak eta eraginkortasuna deskribatuz (ikus dokumentazio-eredua)

Data garrantzitsuak:

Ostirala, 2020-IX-18 (ez da entrega):

- Klaseen espezifikazioa. Klase nagusien diagrama.
- Proben diseinua.
- Datu-egitura nagusien diseinua

Ostirala, 2020-X-9, azken eguna 1.eginkizuna entregatzeko. Honako hauek dauzkan dokumentua entregatuko da:

- Problemaren deskribapen nagusia
- Aztertutako aukeren deskribapena eta hartutako soluzioen deskribapen orokorra, irizpide ezberdinen arabera arrazoituta (adibidez, eraginkortasuna)
- Algoritmoen diseinua eta inplementazioa
- Proben emaitza enpirikoak (denborak)

Gainera Checklist-a bete eta entregatu beharko duzue, eskatutako guztia egin duzuela egiaztatzeko.