
Scraper Bots

X34719_ca

L'empresa Digital Pond disposa del registre de visites a la seva web (també anomenat el *log*) des de tots els ordinadors del món. El registre és cronològic i té el següent format:

```
72.91.13.24 /favicon.ico 11797
62.55.11.182 /index.html 15244
13.13.4.81 /js/main.js 17912
171.82.6.28 /login 21892
72.91.13.24 /index.html 29265
171.82.6.28 /detail?id=13 54401
...
```

A cada línia del registre hi ha un accés a una pàgina del servidor web, amb la IP (adreça d'internet numèrica de l'ordinador que inicia la connexió), la ruta o *path* dins de la web (p.e. */index.html*), i finalment l'instant de temps de la connexió (que es mesura en mil·lisegons des del moment d'iniciar el servidor web).

Digital Pond ens demana que fem un programa per detectar **scraper bots**, que són programes que rastregen la web a molta velocitat, descarregant totes les pàgines que van trobant. Aquests programes poden ser ben intencionats (com el *Google Bot*), però també poden ser utilitzats per a l'extracció de dades de la web sense permís.

Sigui a_1, a_2, \dots, a_n la seqüència d'accessos des d'una IP donada, per exemple, 89.68.225.93. Considerarem que aquesta IP és un *bot* si la seqüència d'accessos conté una subseqüència de 20 accessos consecutius a_i, \dots, a_{i+19} amb $1 \leq i \leq n - 19$ tal que $time(a_{k+1}) - time(a_k) < 100$, amb $i \leq k \leq i + 18$, i a on $time(a)$ és el temps de l'accés a . Un reguitzell d'accessos amb diferències de temps tan petites és un ritme de visites impossible per a un humà i per tant demostra que l'accés prové d'una màquina.

Entrada

El llistat d'accessos a la web, un per línia: la IP, el *path* i un temps en mil·lisegons, que representa el temps transcorregut des de l'inici de l'execució del servidor.

Sortida

Si no hi ha bots, una línia amb "No bots". En cas contrari, tantes línies com *bots* s'hagin trobat, i en cadascuna la IP seguida de tots els *paths* consultats pel bot (incloent els que estan fora de la subseqüència que l'ha delatat), separats per un espai, i per ordre lexicogràfic. Les diferents IPs també s'han de sortir per ordre lexicogràfic, malgrat les IPs siguin numèriques. (En ordre lexicogràfic, la IP "111.2.2.2" és més petita que "2.2.2.2", perquè, si es miren els dígit com lletres, el dígit 1 és més petit que el dígit 2.)

Exemple d'entrada 1

```
63.244.195.58 /help/faq/shipping 76
235.184.182.163 /products/142 77
235.184.182.163 /products/142 103
```

```
63.244.195.58 /products/97 351
235.184.182.163 /index.html 612
63.244.195.58 /help/faq/shipping 949
235.184.182.163 /login 1246
235.184.182.163 /index.html 1323
```

```
63.244.195.58 /login 1876
63.244.195.58 /help/faq/shipping 1886
235.184.182.163 /products/142 2256
235.184.182.163 /products/97 2496
63.244.195.58 /help/faq/shipping 2553
63.244.195.58 /checkout 2738
63.244.195.58 /login 3332
235.184.182.163 /products/142 3372
```

Exemple de sortida 1

No bots

Exemple d'entrada 2

```
191.55.50.46 /h 121
191.55.50.46 /2 173
191.55.50.46 /g 583
191.55.50.46 /a 1180
191.55.50.46 /3 1430
212.207.160.47 /3 2016
212.207.160.47 /b 2042
212.207.160.47 /3 2048
212.207.160.47 /d 2128
191.55.50.46 /h 2175
212.207.160.47 /c 2209
212.207.160.47 /3 2249
212.207.160.47 /c 2266
212.207.160.47 /f 2365
212.207.160.47 /2 2422
212.207.160.47 /3 2483
212.207.160.47 /3 2537
212.207.160.47 /2 2542
212.207.160.47 /2 2558
212.207.160.47 /e 2639
212.207.160.47 /2 2670
212.207.160.47 /c 2700
212.207.160.47 /f 2798
212.207.160.47 /a 2823
212.207.160.47 /g 2922
212.207.160.47 /3 3009
191.55.50.46 /b 3114
191.55.50.46 /g 4052
191.55.50.46 /2 4546
191.55.50.46 /3 5526
```

Exemple de sortida 2

212.207.160.47 /2 /3 /a /b /c /d /e /f /g

Informació del problema

Autor : Pau Fernández

Generació : 2025-03-17 15:47:57

© Jutge.org, 2006–2025.

<https://jutge.org>