

MBA – Športový server

Peter Sabolčák, Martin Brehovský

Analýza

- ◆ Dáta zaznamenávané počas jedného kalendárneho roka
- ◆ Server slúži ako hlavný server katedry telesnej výchovy
- ◆ Prístupy zväčša od študentov a ty majú verejné adresy
- ◆ Veľa prístupov od webových crawlerov, ktoré generovali neúmerný "odpad" v logoch

Analýza II

◆ Začiatok predspracovania dát

- access.log
 - ◆ 332 MB
 - ◆ 1683557 riadkov
 - ◆ Riadky vo formáte

```
IP - - [DATUM] "GET/POST /ADRESA HTTP" NUM NUM "BROWSER INFO"
```

◆ Po odstránení nepoužiteľných záznamov

- access.log.new
 - ◆ 76 MB
 - ◆ 348483 riadkov
 - ◆ Riadky vo formáte

```
IP - - [DATUM] "GET/POST /ADRESA HTTP" NUM NUM "BROWSER INFO"
```

◆ Zmena formátu

- ◆ 9 MB
- ◆ 208339 riadkov
- ◆ Riadky vo formáte

```
IP "/ADRESA"
```

◆ Finálna verzia – vybratie iba požadovaných sekcií a zmazanie duplikátov

- ◆ 133 kB
- ◆ 4881 riadkov
- ◆ Riadky vo formáte

```
IP ID_KATEGORIE
```

Kategórie

- ◆ 36 kategórii
- ◆ Každá zviazaná s nejakým športom alebo človekom, ktorý daný šport reprezentuje na katedre
- ◆ Kategórie sú jednotlivé položky MBA
- ◆ IP adresy nám určujú jednotlivé transakcie MBA

1 /user1
2 /user2
3 /user3
4 /sporty/aerobic
5 /sporty/basketbal
6 /sporty/bowling
7 /sporty/budo
8 /sporty/forbal
9 /sporty/fotbal
10 /sporty/frisbee
11 /sporty/futsal
12 /sporty/hokej
13 /sporty/hrst
14 /sporty/kanoistika
15 /sporty/kobeh
16 /sporty/kopos
17 /sporty/luk
18 /sporty/plavani
19 /sporty/softball
20 /sporty/squash
21 /sporty/stolnitenis
22 /sporty/tenis
23 /sporty/turistik
24 /sporty/volejbal
25 /sporty/ztv
26 /tenis-rezervace
27 /klub/hokejbal
28 /klub/lyzebeh
29 /klub/lyzesjezd
30 /klub/nohejbal
31 /klub/posilovn
32 /klub/tenis
33 /klub/turisti
34 /klub/voda
35 /klub/volejbal
36 /dokument/

Tabuľky na výstupu

◆ Tabuľka četností

- Riadky a stĺpce sú tvorene kategóriami
- Bunka na riadku i a stĺpcu j nám udáva počet transakcii, ktoré obsahujú kategórie i a j

◆ Tabuľka podpory

- V bunkách je percentuálne vyjadrenie podpory medzi kategóriami i a j

◆ Tabuľka spoľahlivosti

- V bunkách je percentuálne vyjadrenie spoľahlivosti medzi kategóriami i a j

◆ Tabuľka zlepšení

- V bunkách je percentuálne vyjadrenie o čo lepšie je pri predikcii použiť náš výsledok ako náhodný predpoklad

Automatizácia procesu

- ◆ Použitý náš vlastný program, ktorý nám zo záznamov v tvare IP ID_KATEGORIA vytvorí tabuľku četností a vytvorí ostatne tabuľky, ktoré obsahujú parametre MBA(podpora, spoľahlivosť, zlepšenie)
- ◆ Výstup do súboru
- ◆ Vytvorí na konci podľa zvolených prahov jednotlivé asociačné pravidla

Výsledky II

	A	B	C	D
1		Zlepseni	Spolahlivost	Podpora
2	futsal => fotbal	4,43	56,31	6,49
3	fotbal => futsal	4,43	51,13	6,49
4	stolni tenis => frisbee	3,36	45,45	5,22
5	frisbee => stolni tenis	3,36	38,58	5,22
6	orientacni beh => horolezecka stena	3,22	56,54	5,49
7	horolezecka stena => orientacni beh	3,22	31,27	5,49
8	stolni tenis => bowling	3,14	54,72	6,29
9	florbal => fotbal	3,14	39,87	5,00
10	fotbal => florbal	3,14	39,35	5,00
11	bowling => stolni tenis	3,14	36,03	6,29
12	kanoistika => horolezecka stena	3,13	54,96	5,45
13	horolezecka stena => kanoistika	3,13	31,04	5,45
14	frisbee => bowling	3,01	52,50	7,11
15	bowling => frisbee	3,01	40,73	7,11
16	klub - posilovani => posilovani	2,98	76,91	8,05
17	posilovani => klub - posilovani	2,98	31,17	8,05
18	softball => bowling	2,97	51,80	5,00
19	bowling => softball	2,97	28,64	5,00
20	orientacni beh => aerobic	2,85	54,22	5,27
21	lukostrelba => horolezecka stena	2,85	50,00	6,35
22	horolezecka stena => lukostrelba	2,85	36,17	6,35
23	aerobic => orientacni beh	2,85	27,69	5,27
24	stolni tenis => horolezecka stena	2,65	46,52	5,35
25	futsal => bowling	2,65	46,18	5,33
26	bowling => futsal	2,65	30,52	5,33
27	horolezecka stena => stolni tenis	2,65	30,46	5,35
28	frisbee => horolezecka stena	2,63	46,14	6,25
29	horolezecka stena => frisbee	2,63	35,59	6,25
30	lukostrelba => bowling	2,59	45,16	5,74

Záver

- ◆ Podľa očakávania sa kategórie, ktoré sú nejakým spôsobom spojené ocitli na vysokých priečkach
- ◆ Menšie prekvapenia v podobe náväznosti medzi frisbee a bowlingom alebo stolným tenisom a bowlingom
- ◆ Pomerne veľa kategórií, ktoré aj keď mali veľkú hodnotu zlepšenia ale ostatné hodnoty mali nízke (niekedy až pod hranicou 0,1%)
- ◆ Spôsobené nevyváženými prístupmi medzi jednotlivými kategóriami