



Autors: **Labi** Raksta ID= 37

## Roboti (boti) Internetā!

Internetā par robotiem vai vienkārši botiem sauc aplikācijas(programmas), kuras nodarbojas ar informācijas meklēšanu globālajā tīmeklī, tie ieseknē informāciju no mājaslapām un tajās esošajiem failiem, lai mēs vēlāk šos datus varētu atrast meklētājā (piemēram, Google). Šādu botu ir ļoti daudz (botu sarakstu var atrast (WEB) <http://www.botsvsbrowsers.com>). Tomēr ir arī "sliktie" boti, kas nodarbojas ar spiegošanu internetā, piemēram, kādam uz servera ir ekskluzīvs fails, kuru viņš nevienam negrib rādīt, "sliktais" bots to atrod un parāda pārējiem. Kā arī "sliktie" boti mēdz "uzkarināt" serverus, kamēr skenē visus uz servera esošos failus. Boti mēdz arī "ielogoties" dažādās mājaslapās un nodarboties ar spamu vai flūdu. Kā izvairīties no botiem?

Ir divas iespējas, kā tikt vaļā no botiem - 1) aizliedzot tos <meta> tagos 2) izveidojot failu robots.txt Tikai jāatceras, ka daudzi "sliktie" boti māk apiet šos aizliegumus, bet "labie" boti gan "klausā aizliegumiem", tāpēc lietojot šos paņēmienus var paslēpties no meklētājiem nevis pilnībā tikt vaļā no "sliktajiem" botiem.

### 1) <meta> tagi

Izmantošanas piemērs:

```
<html>
<head>
<title>Virsraksts</title>
<META NAME="ROBOTS" CONTENT="NOINDEX, NOFOLLOW">
</head>
```

Meta tagam ir divi atribūti NAME, kurš paliek nemainīgs, un CONTENT. CONTENT iespējamās vērtības varbūt NOINDEX (neļauj skenēt šo lapu) un NOFOLLOW (neļauj pāriet uz linkiem, kas atrodas šajā lapā), tos var lietot pa vienam vai kopā, kā šajā piemērā.

### 1) Robots.txt

Failu robots.txt var izveidot jebkurā teksta redaktorā, piemēram, notepad++ , un tad jāievieto servera galvenajā mapē (parasti tās nosaukums ir public\_html vai www vai htdocs).Izmantošanas piemēri (šis teksts jāraksta robots.txt failā):

*Lai aizliegtu pieeju visiem botiem (kas nemāk apiet šo aizliegumu):*

User-agent: \*

Disallow: /

*Lai aizliegtu pieeju konkrētam botam:*



User-agent: BadBot

Disallow: /

*Lai aizliegtu pieeju visiem botiem izņemot vienu konkrētu botu:*

User-agent: Google

Disallow:

User-agent: \*

Disallow: /

*Lai aizliegtu pieeju konkrētiem failiem:*

User-agent: \*

Disallow: /~joe/junk.html

Disallow: /~joe/foo.html

Disallow: /~joe/bar.html

Tā kā fails robots.txt atrodas publiski pieejamā mapē, tad tajā nevajadzētu rakstīt failus, kuri ir slepeni, jo jebkurš var ielādēt šo failu un apskatīties, kas tajā ir rakstīts. => Šajos piemēros "User-agent" ir bota nosaukums, bet "Disallow" ir tā mape vai fails, kuru aizliegts skenēt("skatīties").

Informācija ņemta no (WEB) <http://robotstxt.org>

Uzrakstīts:15:17 10-08-10