Russian Web Spam Evolution: Yandex Experience
About Yandex
#1 search in Russia

Share of Searches

Yandex

60% 61.9%

Google

21% 26.1%

mail.ru

8% 8.6%

Rambler

6% 1.3%

1 Source: LiveInternet.ru, February 2013. Search traffic reflects Russian users to Russian websites and includes desktop and mobile.
Market Share in Turkey
Spam types

- Spam in domain names
- Link spam
- Emulation of user, fake behavior
- Text spam
Spam in domain names

www.total-insurance.org
www.newyorklife.com
www.insuranceforeveryday.com
www.insurnow.com/insurance-policy
www.weinsureproperty.com
www.insurnow.com
Spam in domain names

Domains which include many query trigrams, % in top-10 for 10K queries (approximated)
Links

Unique situation in Russia: paid links exchange.
Combining anchor text categorization and graph analysis for paid links detection.
User behavior

- Text spam – visible, easy to analyze
- Links – visible, easy to analyze
- Domain names – visible, very easy to analyze
- User behavior spam – extremely difficult to comprehend!
- “Eyes wide shut”
Text spam

- Hidden text
- Keyword stuffing
- Unique content is required
- Random texts, generated texts
- Synonimization
- Overoptimization
Evolution of text spam: overoptimization
Evolution of text spam: overoptimization

80% of terms are on 2nd screen
Example of text spam (doorway)

Quite simple

queries
Tricks

Obfuscation (fragment):

```
OOOO = new Array();

OOOO[0] = `This page requires a browser version 3.10 or newer
...`;

//document.body.innerHTML += new Array();

Tricks()
```
for (i = 0; i < t_arr.length; i++) {
    var lnk = document.createElement('a');
    lnk.id = 'apossum';
    lnk.href = 'http://' + t_arr[i].neg;
    document.body.appendChild(lnk);
    if (document.defaultView) {
        var a_color = document.defaultView.getComputedStyle(lnk, null).getPropertyValue('color');
    } else {
        var a_color = lnk.currentStyle['color'];
    }
    document.body.removeChild(lnk);
    if (a_color == '#142536' || a_color == 'rgb(20, 37, 54)') {
        if (t_arr[i].krl) {
            this.bsl.push(t_arr[i].neg);
        } else {
            this.dlo.push(t_arr[i].neg);
        }
    } else {
    }
}
if (user.history("yandex.ru, google.com, yahoo.com, wikipedia.org") == FALSE)
{
    if user.history("antispam.yandex-team.ru") == TRUE)
    {
        user.type = spam_analyst; // they called him “apossum”
    }
}
...
if (user.type == spam_analyst)
{
    do_not_show_spam_content();  // !!!
}
Important things

• Security is important!
• JS-tricks appear (2008)
JS problem

- Search engines have some difficulties with JS
- JS can produce another JS
- To analyze JS you must run it
- JS can be obfuscated and so on…
• JS
• Bad
• Income for new spam technologies
Aggressive advertising

- JS
- Bad user experience

Average dwell time, compared to normal ads

- Income for new spam technologies
Aggressive advertising

Percent of URLs with aggressive ads in all shown in top-10 URLs for 300 queries

Blackhats tries to increase banners’ effectiveness.
Social engineering + malware.
Social engineering

- **Fakes**
  - smth like fake antivirus

- **Phishing**
  - popular websites, social networks

- **Pharming**
  - Hosts
  - DNS

Malware is widely used
Social engineering

- Fake lottery

Wrong URL (but who cares...)

Enter your mobile phone number and get paid subscription.
Conclusions

• Fast reaction is absolutely necessary
• JS analysis: must have
• Malware detection systems must be improved
• Educational projects to help newbies
• Keep fighting!
Thank you!