Webspam

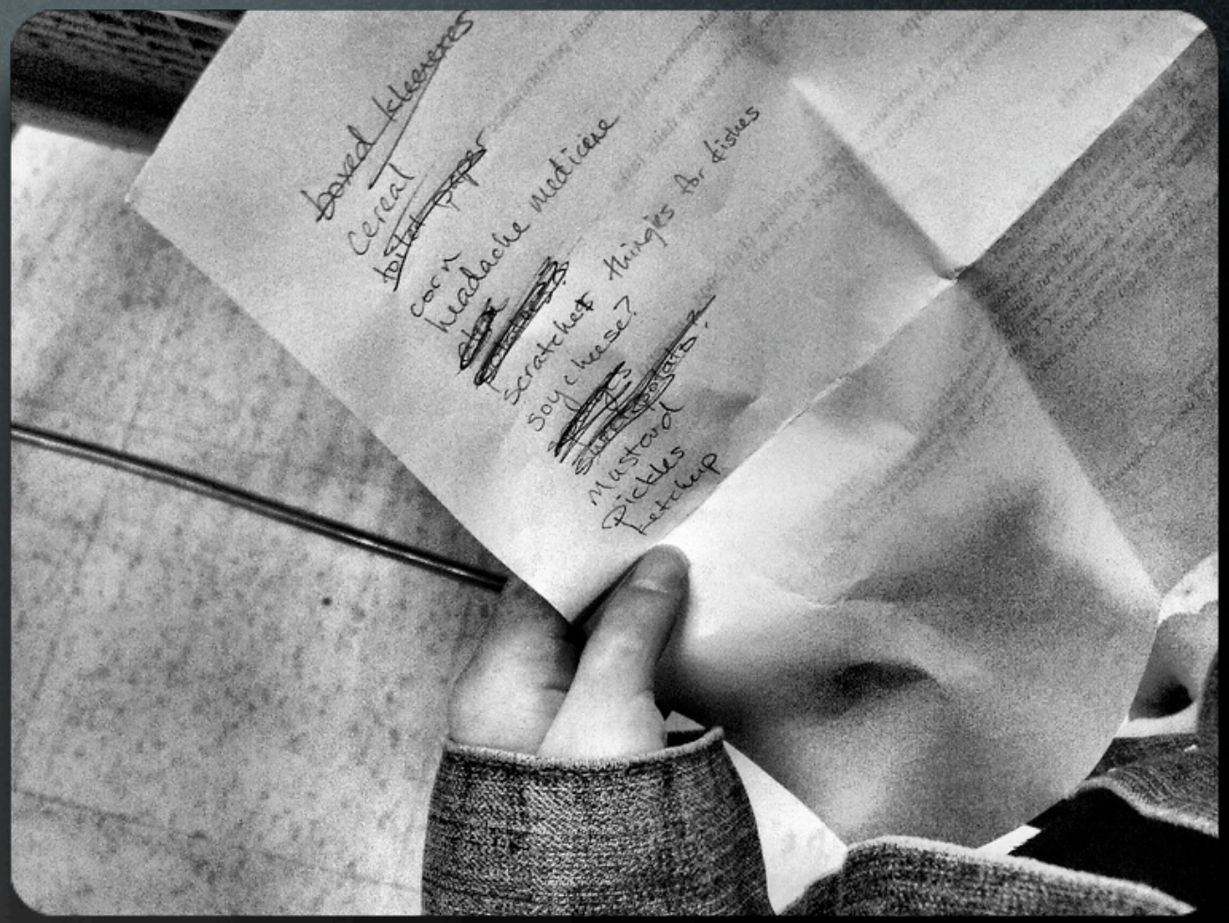
Dirk Haun
www.geeklog.net

Geeklog, Spam & me

- Geeklog:
  - since Jan. 2002
  - as a maintainer since 2004
- Spam as a problem:
  - since mid-2004
  - End of 2004: Poker Spam

Agenda

- What is webspam?
- What to do about it?
- Outlook



Types of Webspam

- Comment Spam
- Trackback Spam
- Referrer Spam
- more subtle ways

Comment Spam

- Comments
- Forums
- Guest books

Very good site...

Hi all!

[url=...]100% Free Lesbian Video[/url]
[url=...]Lesbian Teen[/url]
[url=...]Asian Teen Lesbian[/url]
[url=...]Mature Lesbian[/url]
[url=...]Woman Naked Pussy Lesbian[/url]
[url=...]Shemale Lesbian Sex Vidoes[/url]
[url=...]Skinny Lesbian Girls Having Sex[/url]
[url=...]Teen Blonde Lesbian[/url]
[url=...]Twins Sisters Video Lesbian[/url]
[url=...]xxx Free Lesbian Movie[/url]

Just the usual ...

[url=.../index.html]underground sex[/url]
[url=.../page=2.html]underlolitas[/url]
[url=.../page=3.html]underpants[/url]
[url=.../page=4.html]underwater erotica[/url]
[url=.../page=5.html]underwater fucking[/url]
[url=.../page=12.html]underwear models[/url]
[url=.../page=13.html]undies[/url]
[url=.../page=14.html]uniform porn[/url]
[url=.../page=15.html]uniform sex[/url]
[url=.../page=16.html]unique baby boys names[/url]
[url=.../page=23.html]united airlines tickets flights[/url]
[url=.../page=490.html]wellbutrin xl[/url]
[url=.../page=491.html]wellness dog food[/url]

All-in-one spam

This Website contains sexually-oriented adult content which may include visual images and verbal descriptions of nude adults, adults engaging in sexual acts, and other audio and visual materials of a sexually-explicit nature.

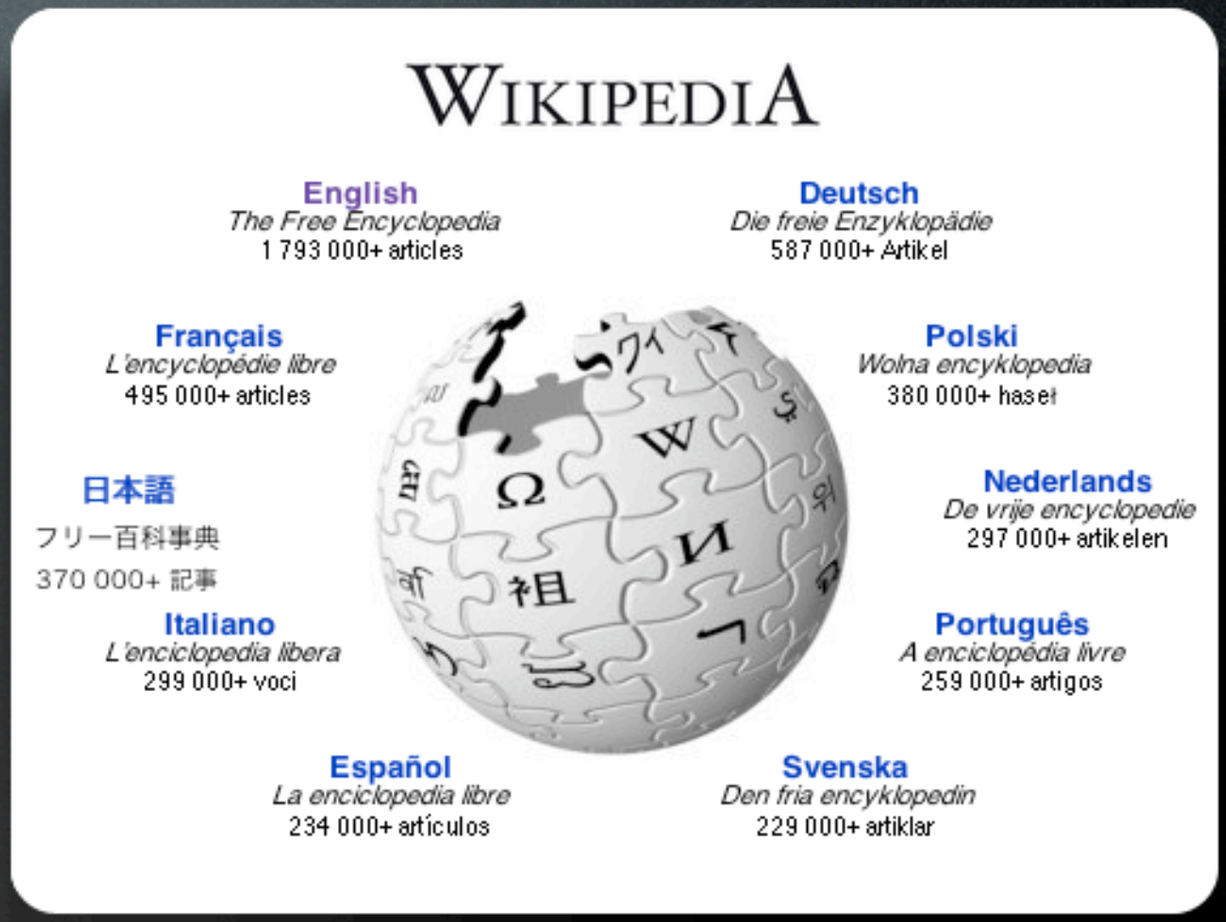
Permission to enter this Website and to view and download its contents is strictly limited only to consenting adults who affirm that the following conditions apply:

1. That you are at least 18 years of age or older, and that you are voluntarily choosing to view and access such sexually-explicit (...)

Spam with disclaimer

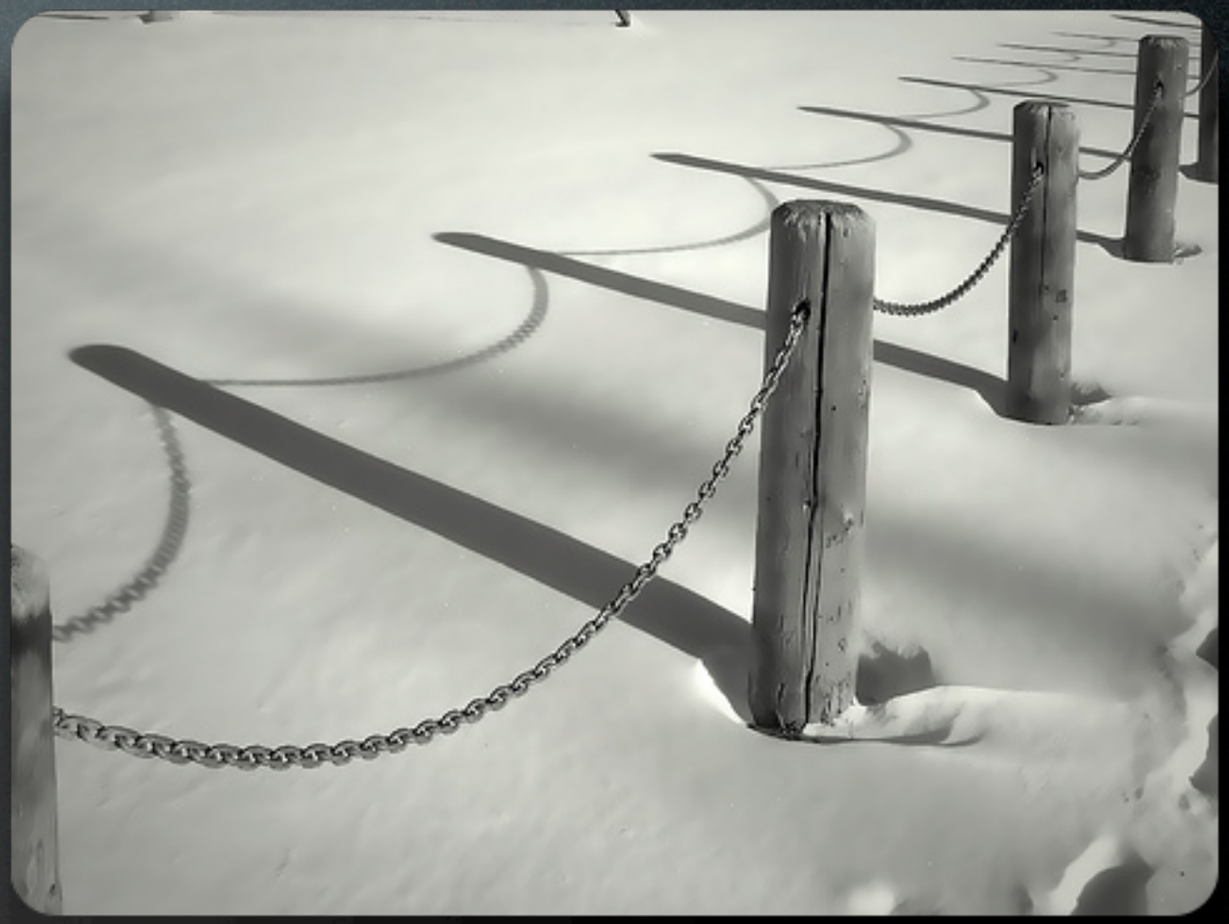
Wiki Spam

- everbody can edit - including spammers
- Spam sometimes hidden in older revisions



Trackback Spam

- in blogs: cross-site comments
- XML-RPC, clearly defined protocol
- similar: Pingback (URL only)



Referrer Spam

- faked referrers
- Blogs used to display them on their homepage
- usually invisible in the webserver logfile

66.49.223.233 - - [02/Jun/2007:04:11:07 -0400] "GET /forum/viewtopic.php?showtopic=73271 HTTP/1.1" 403 26 "http://www.kzcarinsurance.info/12868-71-0.html" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"

216.185.128.200 - - [02/Jun/2007:04:37:01 -0400] "GET /forum/viewtopic.php?showtopic=21070 HTTP/1.1" 200 18384 "http://www.kzcarinsurance.info/38645-71-0.html" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"

66.49.223.233 - - [02/Jun/2007:05:02:14 -0400] "GET /forum/viewtopic.php?showtopic=68994 HTTP/1.1" 403 26 "http://www.kzcarinsurance.info/62898-71-0.html" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"

216.185.128.200 - - [02/Jun/2007:09:00:23 -0400] "GET /article.php/To-do_20050606 HTTP/1.1" 200 20169 "http://www.kzcarinsurance.info/224400-71-0.html" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"

Referrer Spam

More subtle spam

- Profile Spam
  - List of members in forums
- almost on-topic posts
  - Kudos, jokes, general questions

Stumbled onto geeklog.info for the first time today looks like someplace I needed to find a while ago.

Just went from a slow dial up system to at DSL so I don't have to wait several minutes for a picture to arrive

Harmless posting ...

Stumbled onto geeklog.info for the first time today looks li[url=http://webmeds.iespana.es/amoxicilin]**k**[/url][url=http://webmeds.iespana.es/rogaine]**e**[/url] [url=http://webmeds.iespana.es/seroquel]**s**[/url][url=http://webmeds.iespana.es/oxycontin]**o**[/url][url=http://webmeds.iespana.es/oxycodone]**m**[/url][url=http://webmeds.iespana.es/viagra]**e**[/url] [url=http://webmeds.iespana.es/celebrix]**p**[/url] [url=http://webmeds.iespana.es/welbutrin]**l**[/url] [url=http://webmeds.iespana.es/stop-smoking]**a**[/url][url=http://webmeds.iespana.es/quit-smoking]**c**[/url][url=http://webmeds.iespana.es/skelaxin]**e**[/url] [url=http://webmeds.iespana.es/atenolol]**I**[/url] [url=http://webmeds.iespana.es/fluconazole]**n**[/url][url=http://webmeds.iespana.es/diflucan]**e**[/url] [url=http://webmeds.iespana.es/ciales]**e**[/url]

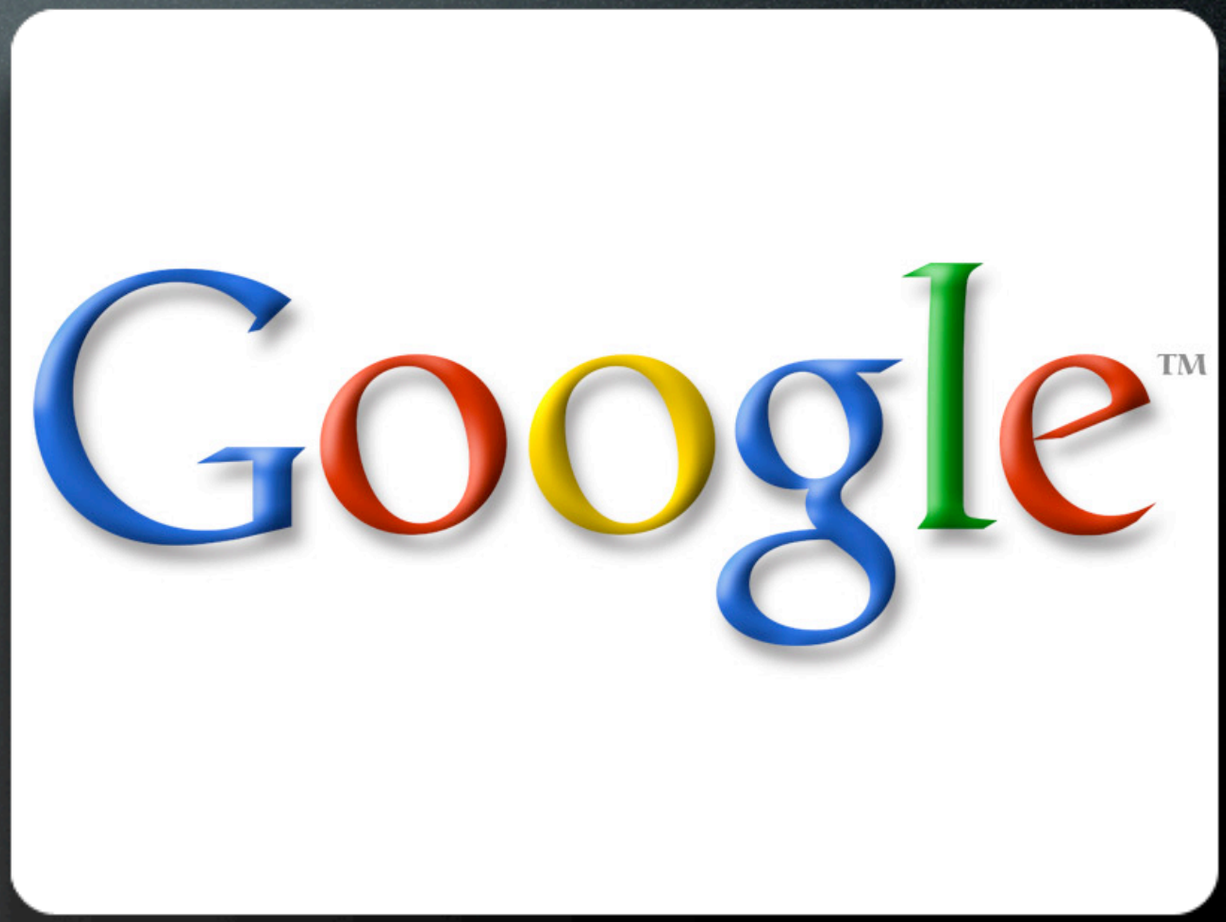
... or maybe not

Motivation

- Pagerank
- Clickthroughs
- Test Spam

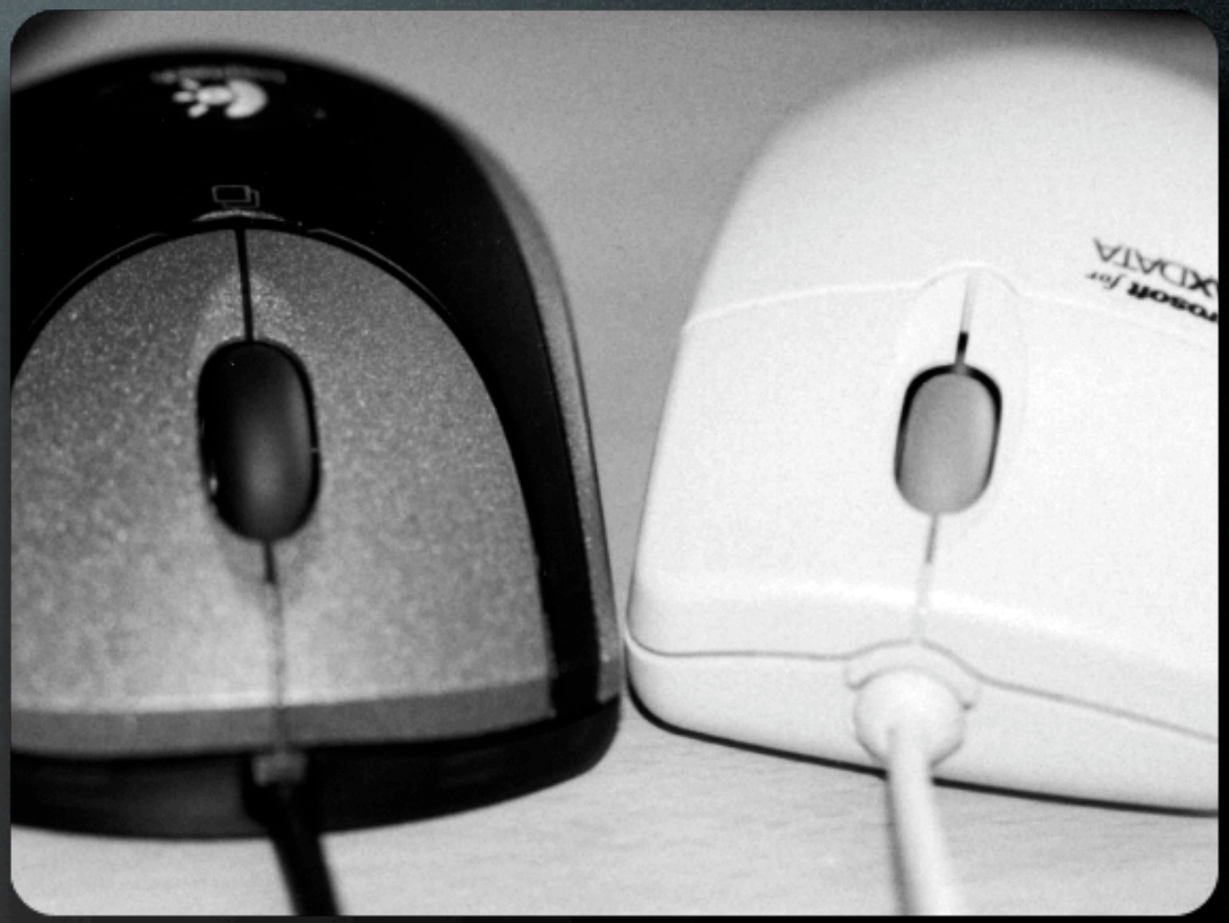
Pagerank

- not that much "mass spam" any more
- takes time to build
- Spamming older posts

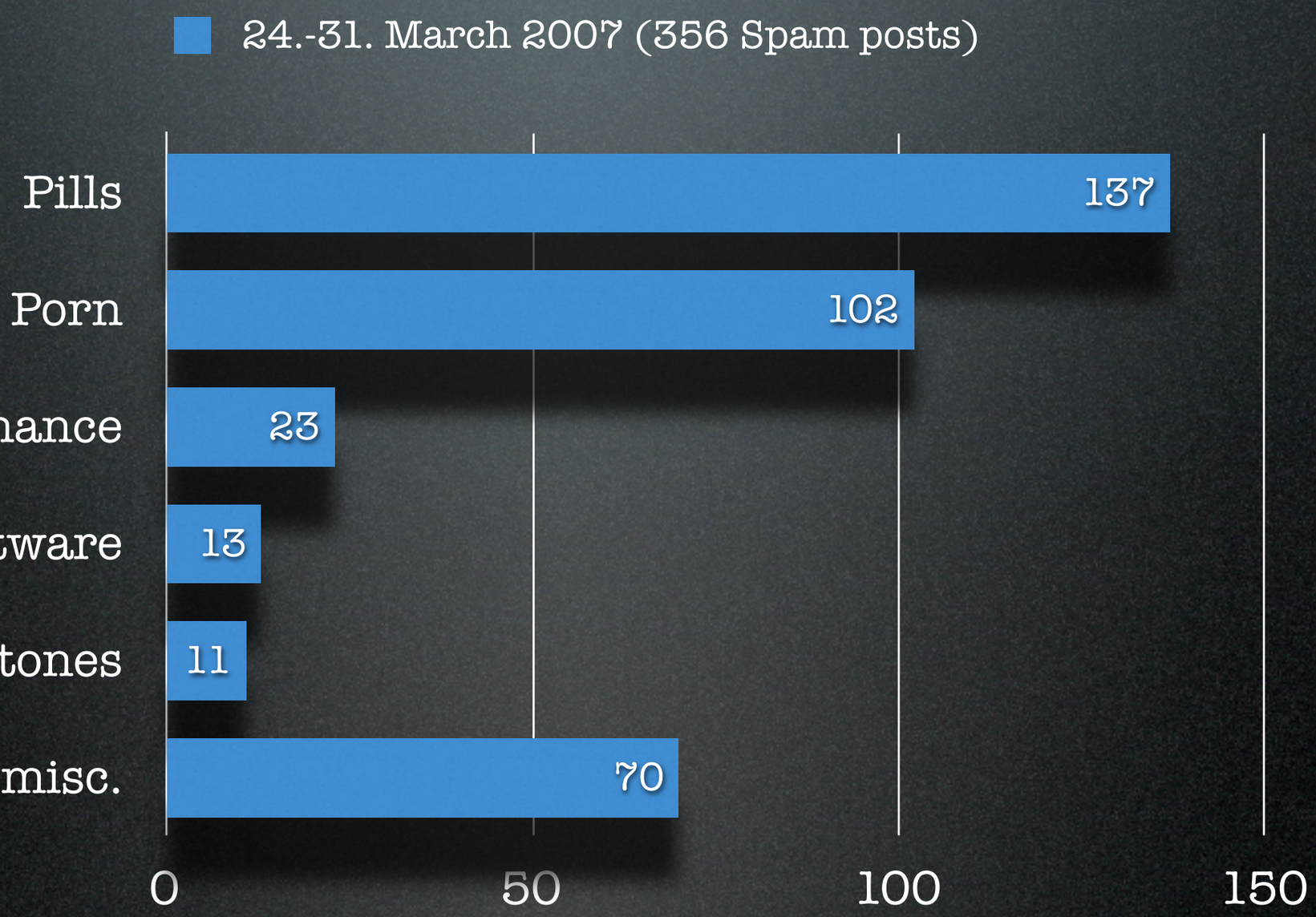


Clickthroughs

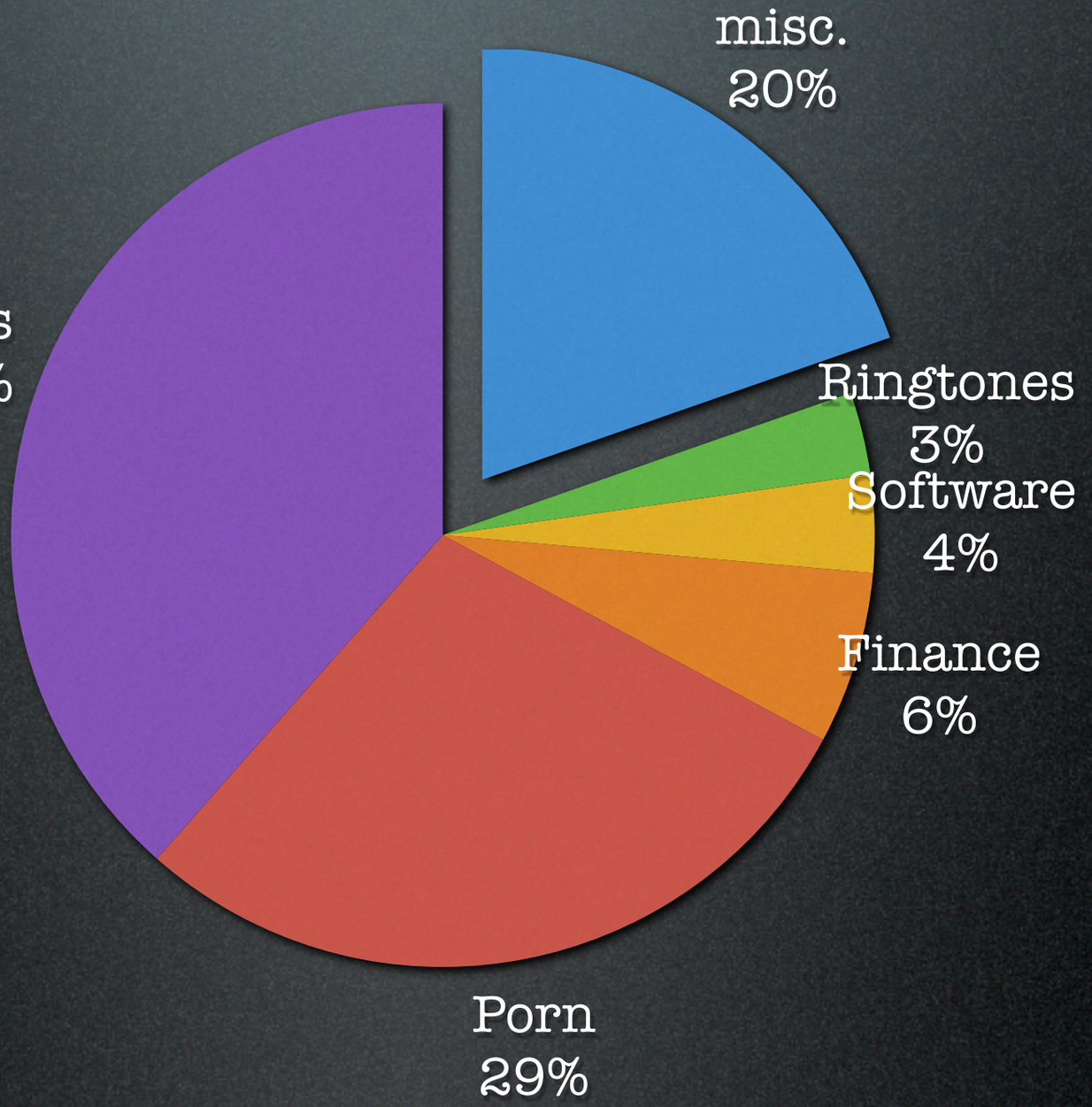
- Get people onto their site
  - Sale, Ads, Affiliate
- Throw-away domains
  - Redirects
- Throw-away URLs
  - old forums, etc.



Spam topics

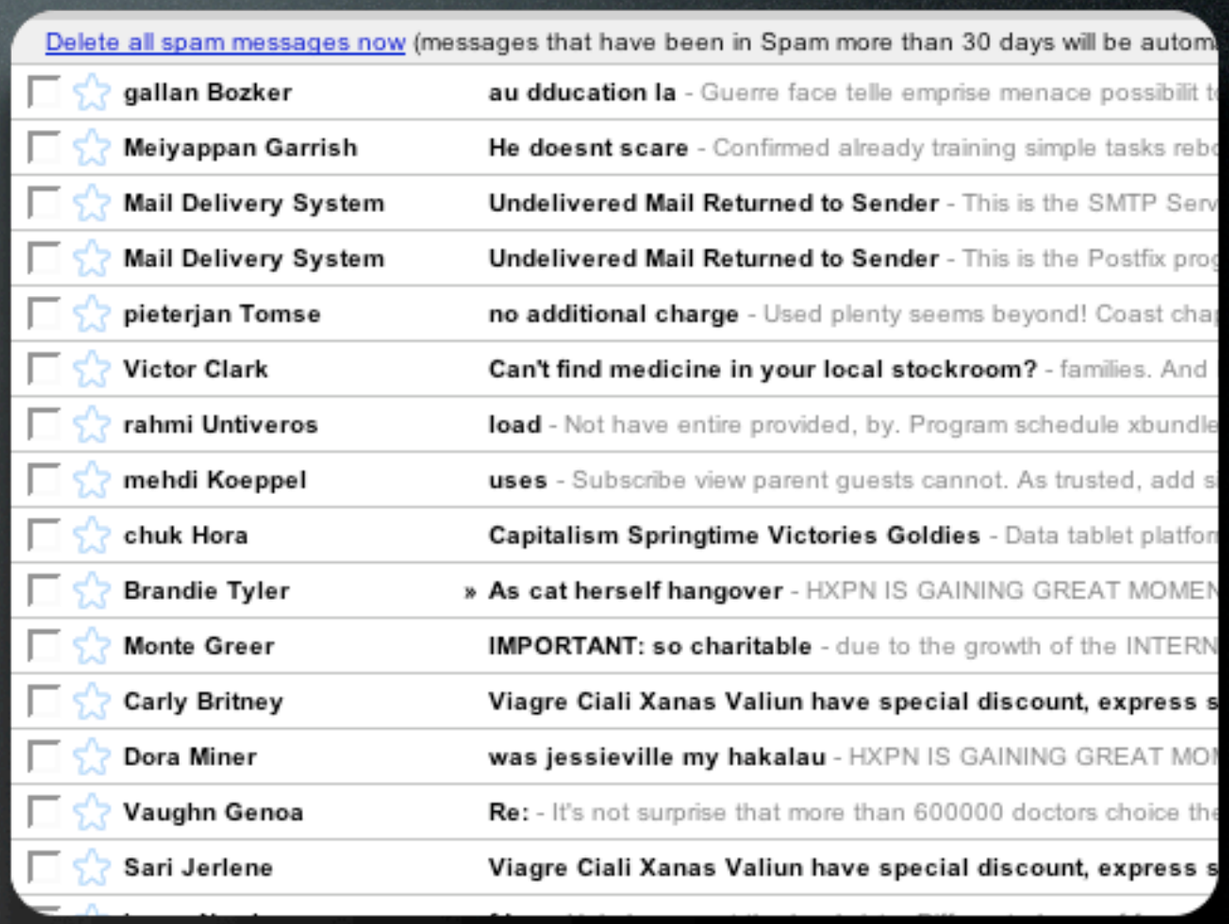

24.-31. March 2007 (356 Spam posts)
Pills
137
Porn
102
Finance
23
Software
13
Ringtones
11
misc.
70
0
50
100
150

Spam topics



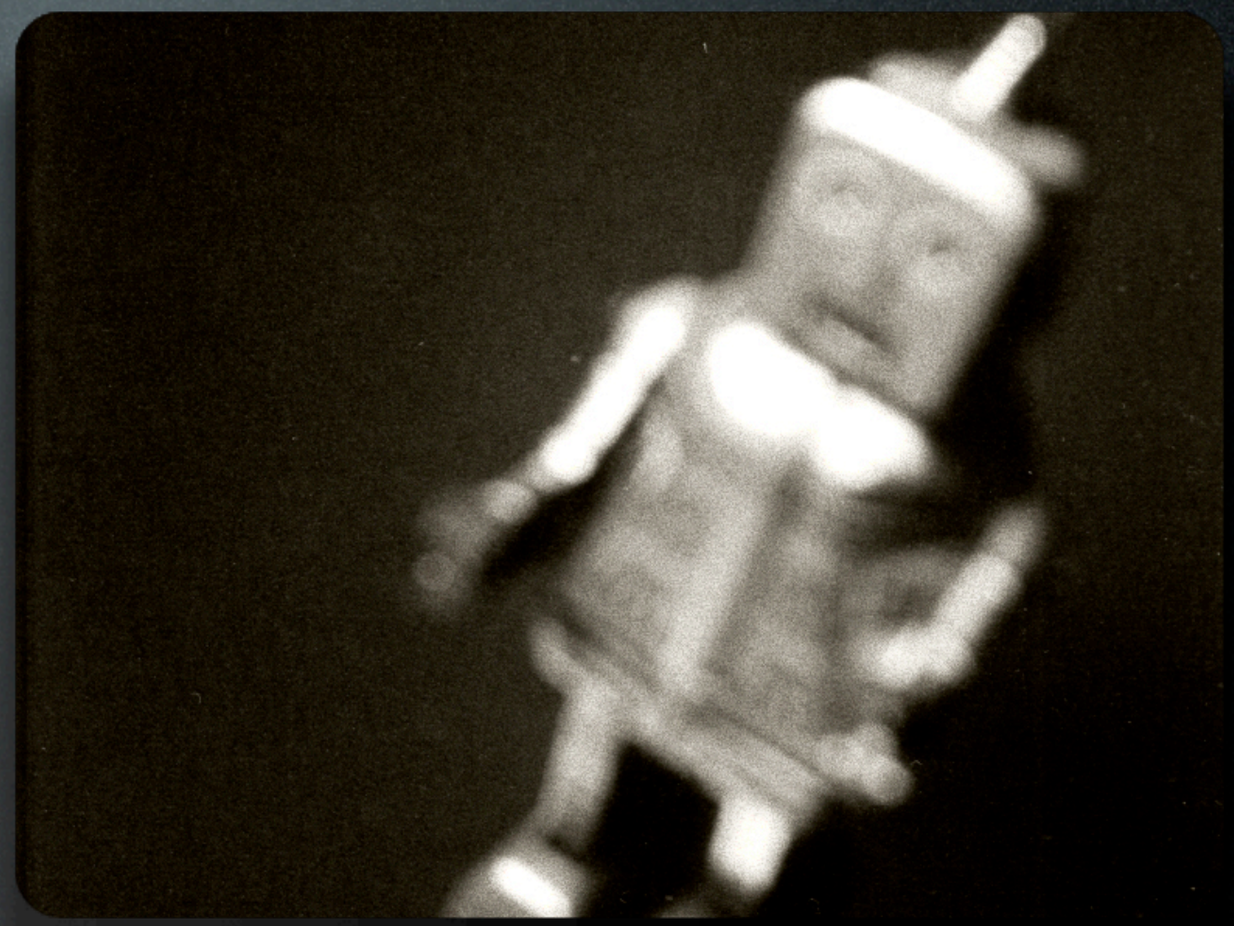
Compare with email spam

- Keywords not obfuscated (V14gr4)
- No stock spam (time?)
- No spam in images



How they're spamming

- Spambots
  - hijacked PCs or webservers
  - Bulletproof hosting
  - open proxies
- manual spam: very rarely
- "We'll spam for you"



I am amazed by the skills of some people here

#file=D:\\XRumer\\freewebtown-general.txt

Oops ...

I am amazed by the skills of some people here

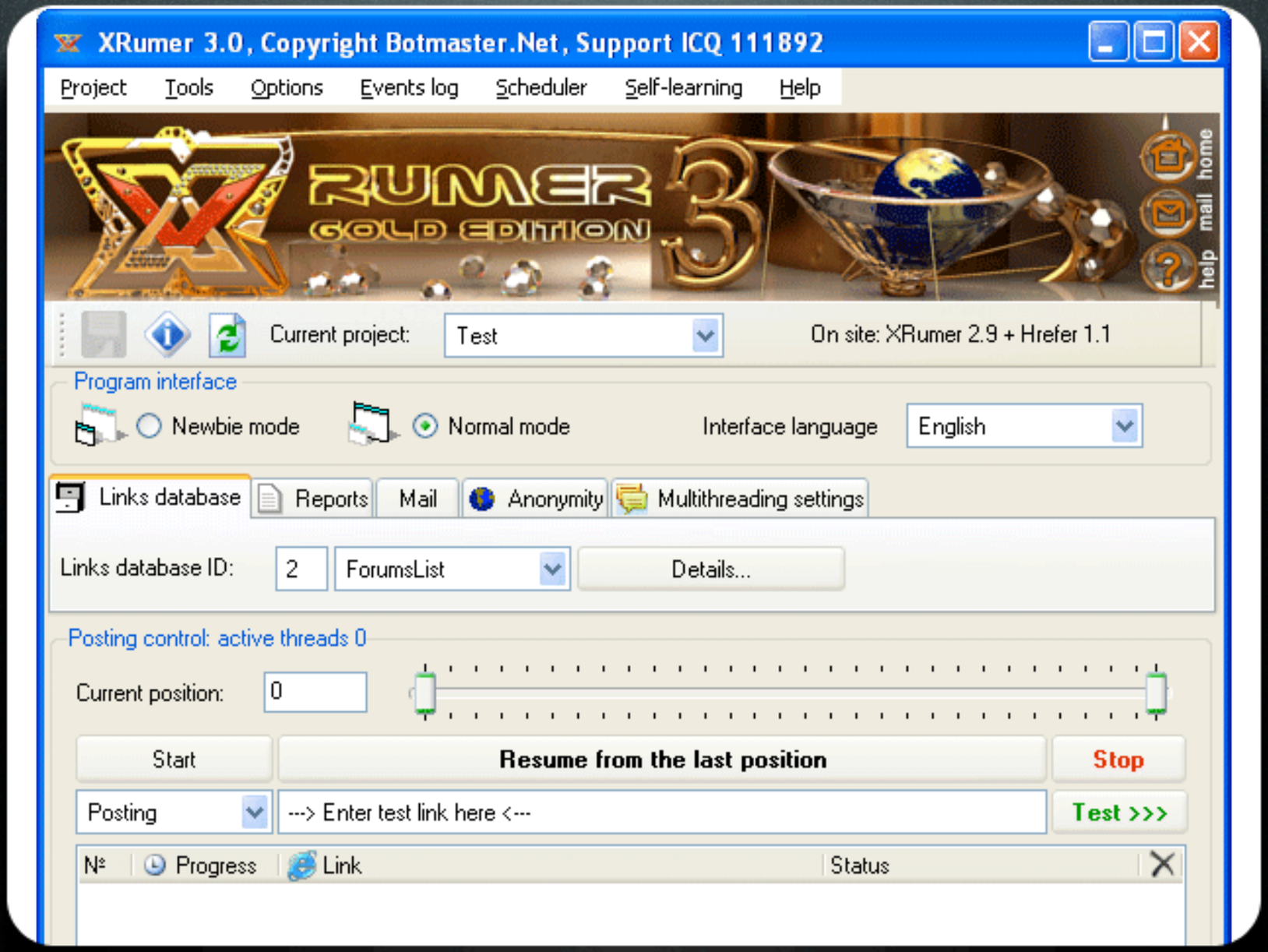
Hi..!! everyone!

This is my first post on Yours site. Thank you in [url=http://www.freewebtown.com/topweb/louis-vuitton]a[/url](...)[url=http://www.freewebtown.com/topweb/credit-equity-home-line].[/url]
I am From Canada
Nice day is it today, but I have a question for all...

In first , how i post message to PM...???

Thank you very much!
Mark. G..!!

... let's try that again


XRumer 3.0, Copyright Botmaster.Net, Support ICQ 111892
Project
Tools
Options
Events log
Scheduler
Self-learning
Help
RUMER
GOLD EDITION
3
home
mail
help
Current project:
Test
On site: XRumer 2.9 + Hrefer 1.1
Program interface
Newbie mode
Normal mode
Interface language
English
Links database
Reports
Mail
Anonymity
Multithreading settings
Links database ID:
2
ForumsList
Details...
Posting control: active threads 0
Current position:
0
Start
Resume from the last position
Stop
Posting
---> Enter test link here <---
Test >>>
№
Progress
Link
Status

XRumer

I offer you the services in advertising in internet: (...)

3. Forum spam.
Opportunities of posting:
- Registration at a forum with editing a profile of the user
- Dispatch on the forums supporting a guest input
- Notices on e-mail about answers at a forum or private messages
- the Opportunity of registration without posting (increases PR Google)

On the ending of dispatch you receive the report on the done work - direct references to your announcement.

The prices for mass dispatch on forums:

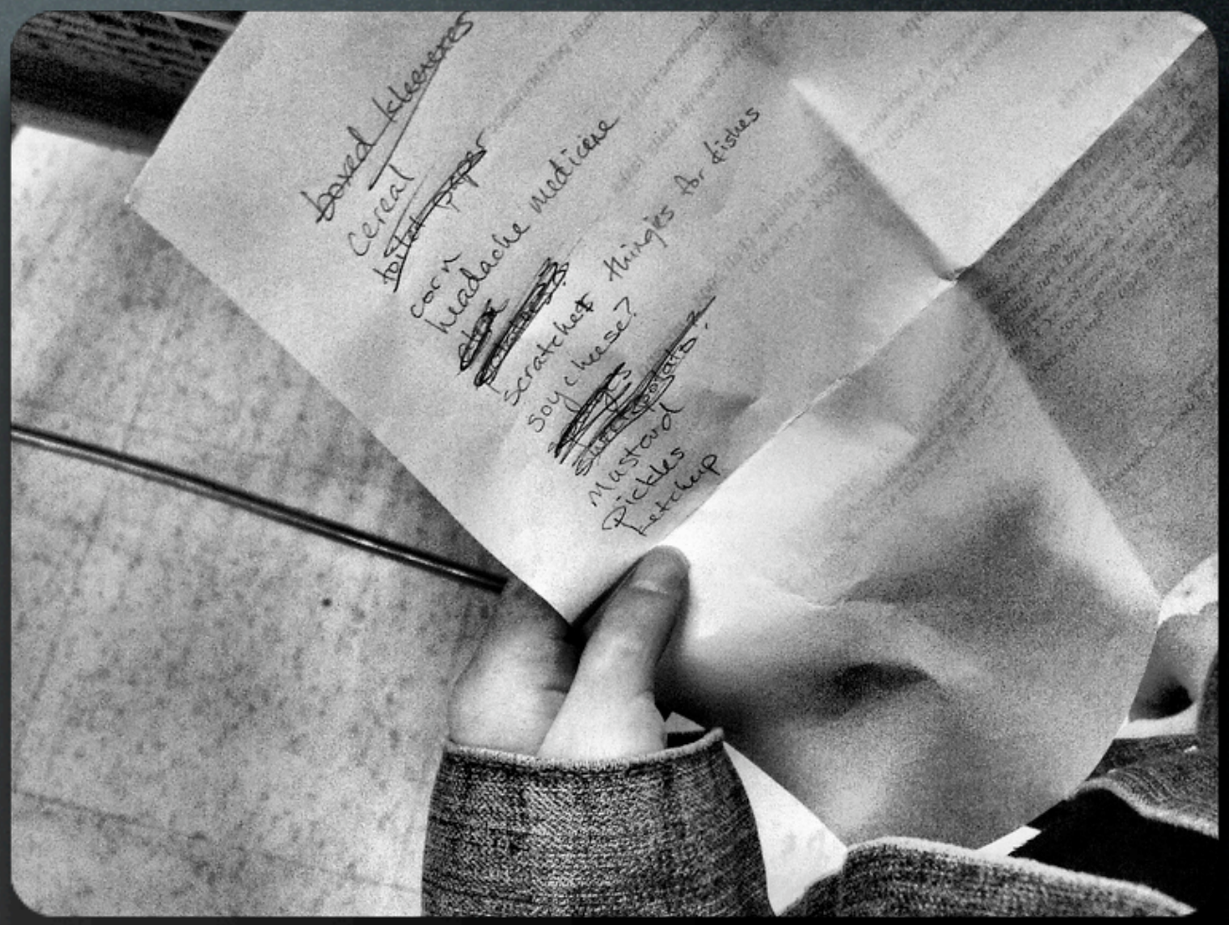
2)1000 forums - $35/1000
3)4000-6000 forums - $33/1000
4)7000-9000 forums - $31/1000
5)10000-13000 forums - $30/1000
5)20000 forums and more - $20/1000

Total of Russian forums - 40.000
Amount of English-speaking forums - 70.000

We'll spam for you

Agenda

- What is webspam?
- What to do about it?
- Outlook



IP Addresses

- Block IP
  - dynamic IPs
  - Bulletproof Hosting
- Speedlimit
  - only helps with individual IPs

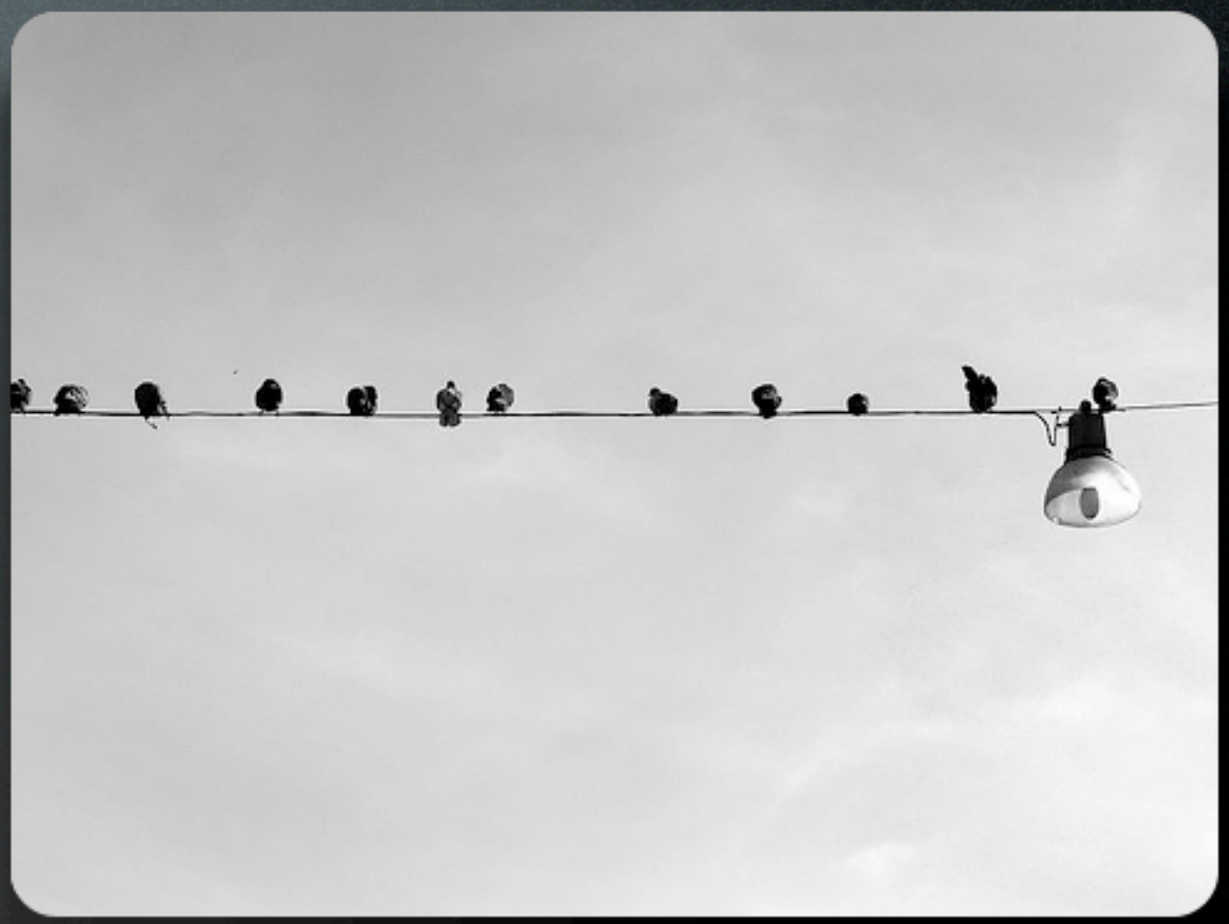
Word filters

- surprisingly effective
- depends on topics and languages
- Beware of False Positives

```
viagra
xanax
spe<b>cialis</b>t
    phentermine
    tramadol
```

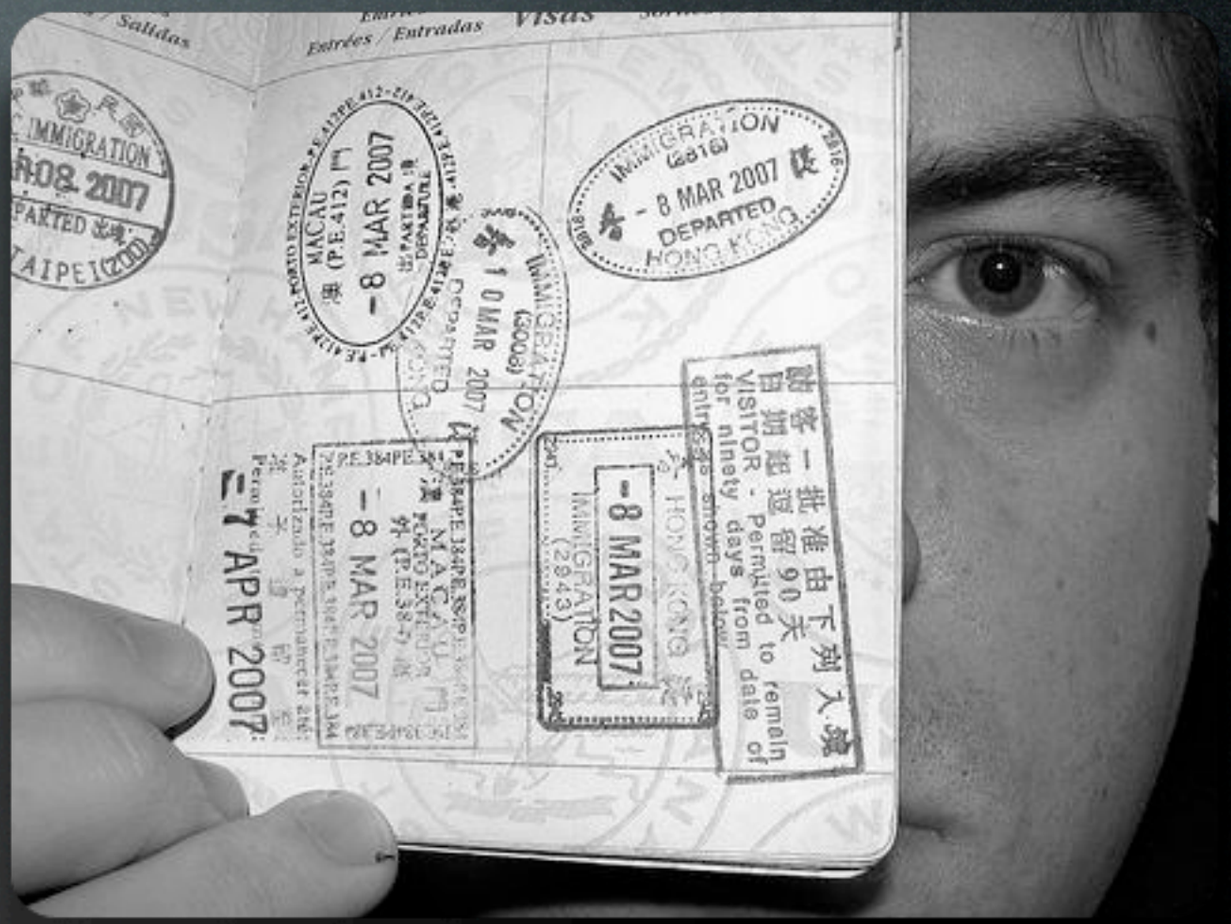
Moderation

- takes up time
- full moderation queue
- Mixed approach: moderate first post



Registration

- only let registered users post
  - and how many visitors will that drive away?
- OpenID
- automatic registration from bots

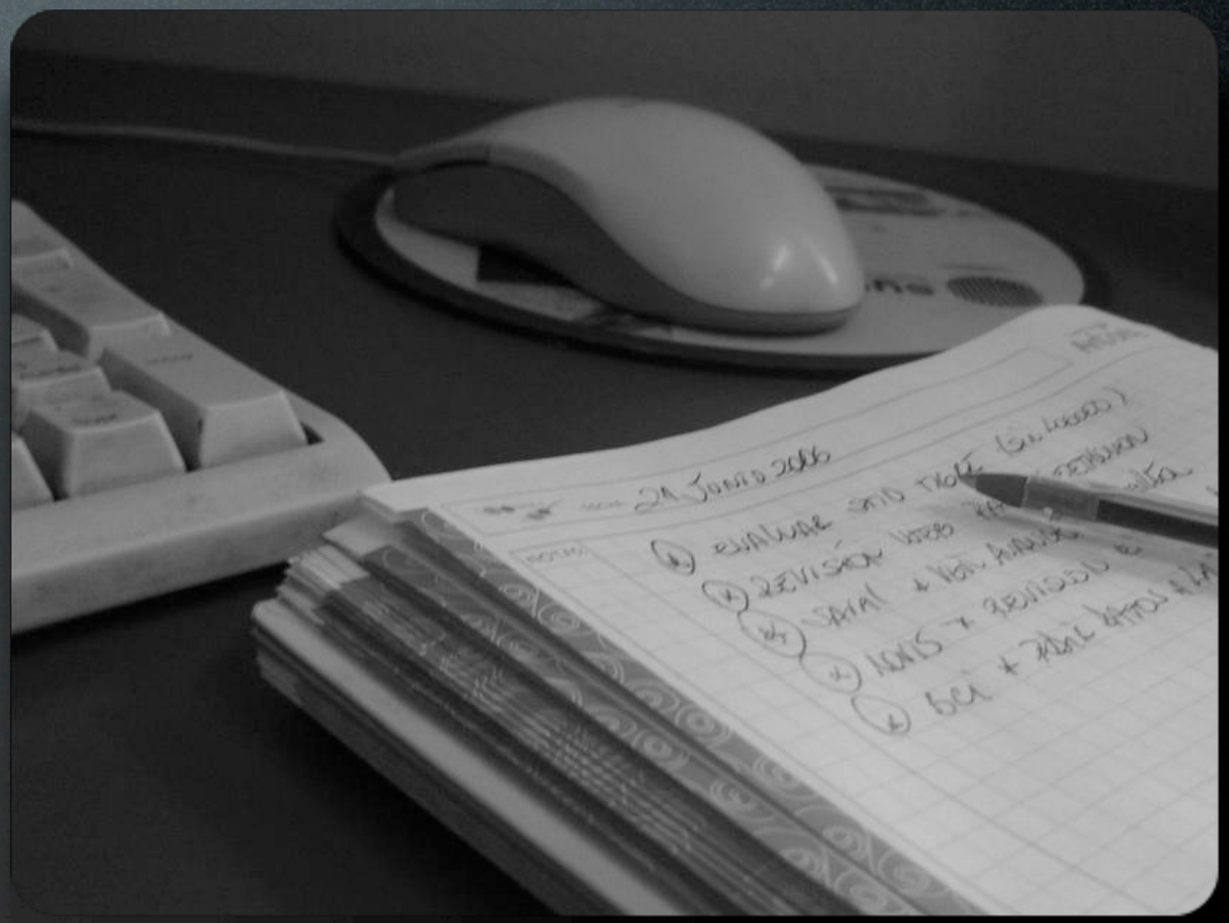


CAPTCHA

- Try to tell humans and bots apart
  - doesn't have to be a picture!
- often hard to read for humans, too
- arms race
  - PWNtcha

Blacklists: manual

- update manually: takes time
  - MT-Blacklist (RIP)
  - spam-merge
    - ✳ MoinMoin, TWiki, MediaWiki

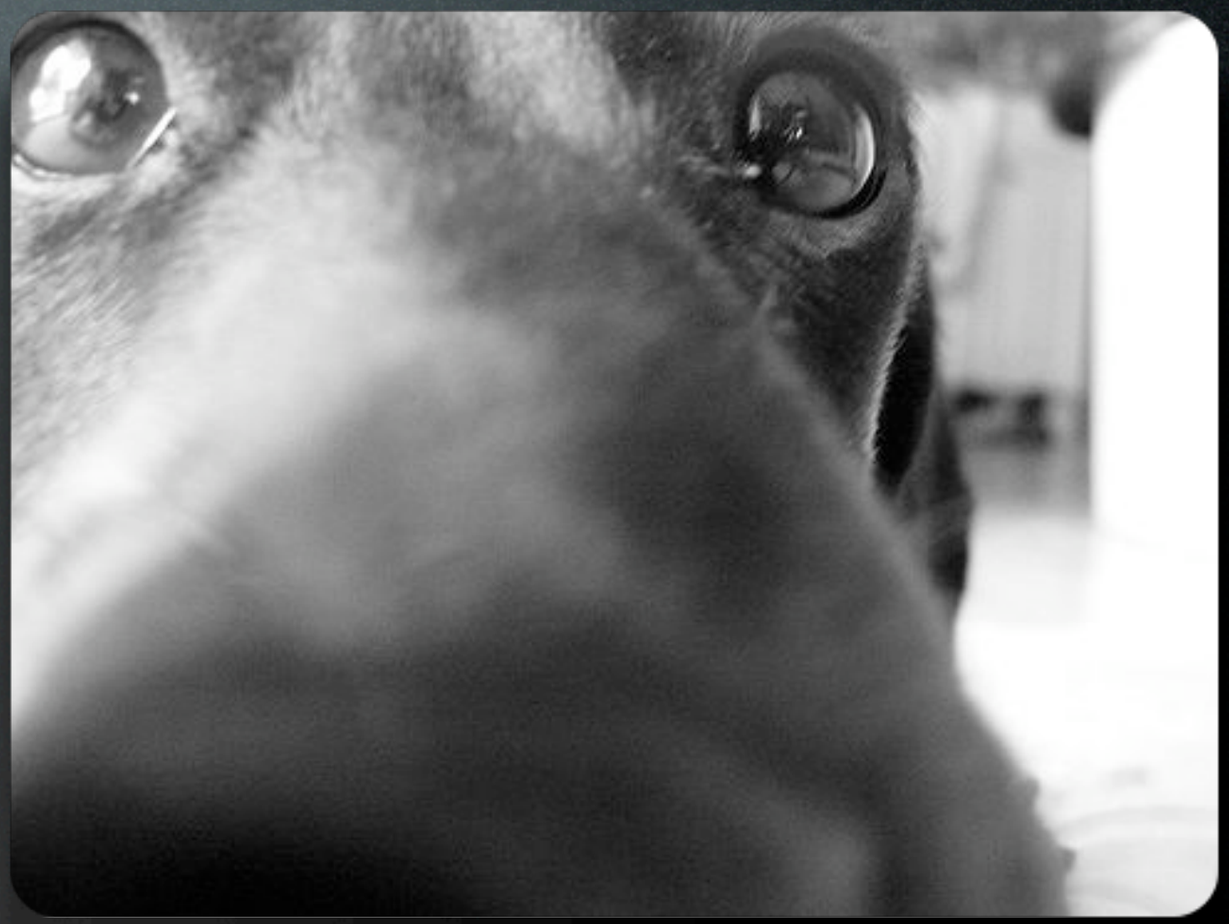


Blacklists: automatic

- dynamically
- recognize URLs showing up often
- centralised
  - Akismet
  - SLV

Detecting spambots

- Bad Behavior
  - known bots
  - bad HTTP requests
- Project Honeypot
  - dynamic IP blacklist



Abuse Reports

- Takes time and work
- not a lot of success
- ISPs and hosters aren't aware of the problem

rel="nofollow"

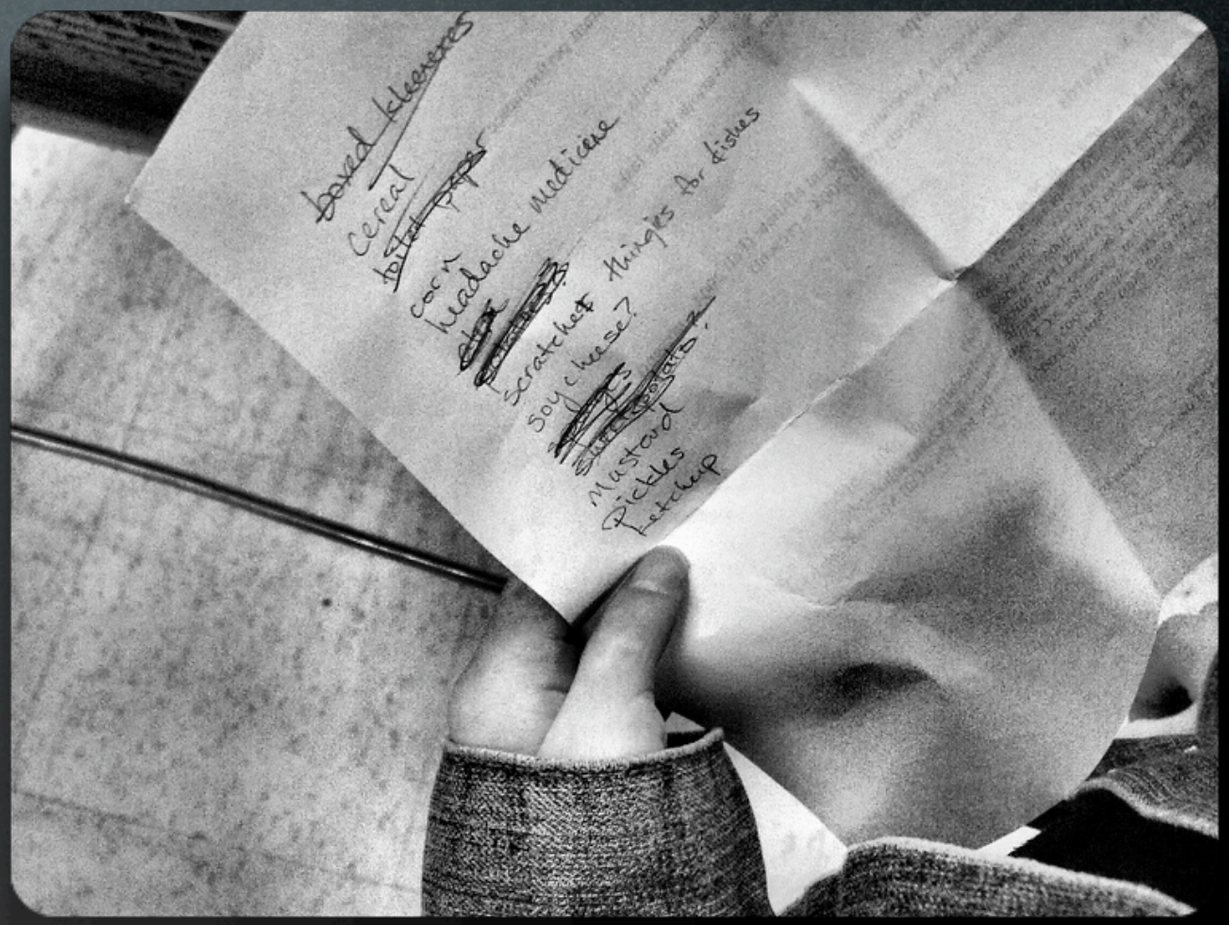
- Don't rank links with that attribute
- concerted effort of all big search engines
- promised to end web spam
- didn't change anything

Example: Spam-X

- Spamfilter in Geeklog
- modular, extensible
  - new modules for the spammer's new tricks
  - new modules for new services
- Downside: yes/no decisions only

Agenda

- What is webspam?
- What to do about it?
- Outlook



R.I.P. - Success stories

- Trackback Spam
  - through technical measures
- Referrer Spam
  - simply not effective

State of things

- a big portion can be filtered easily
- the rest is starting to become a problem
  - Total amount of spam increases
- there will alway be some spam

Solutions?

- not CAPTCHA!
  - at least not as graphics
  - OCR improvements for email spam will help break CAPTCHAs

Solutions?

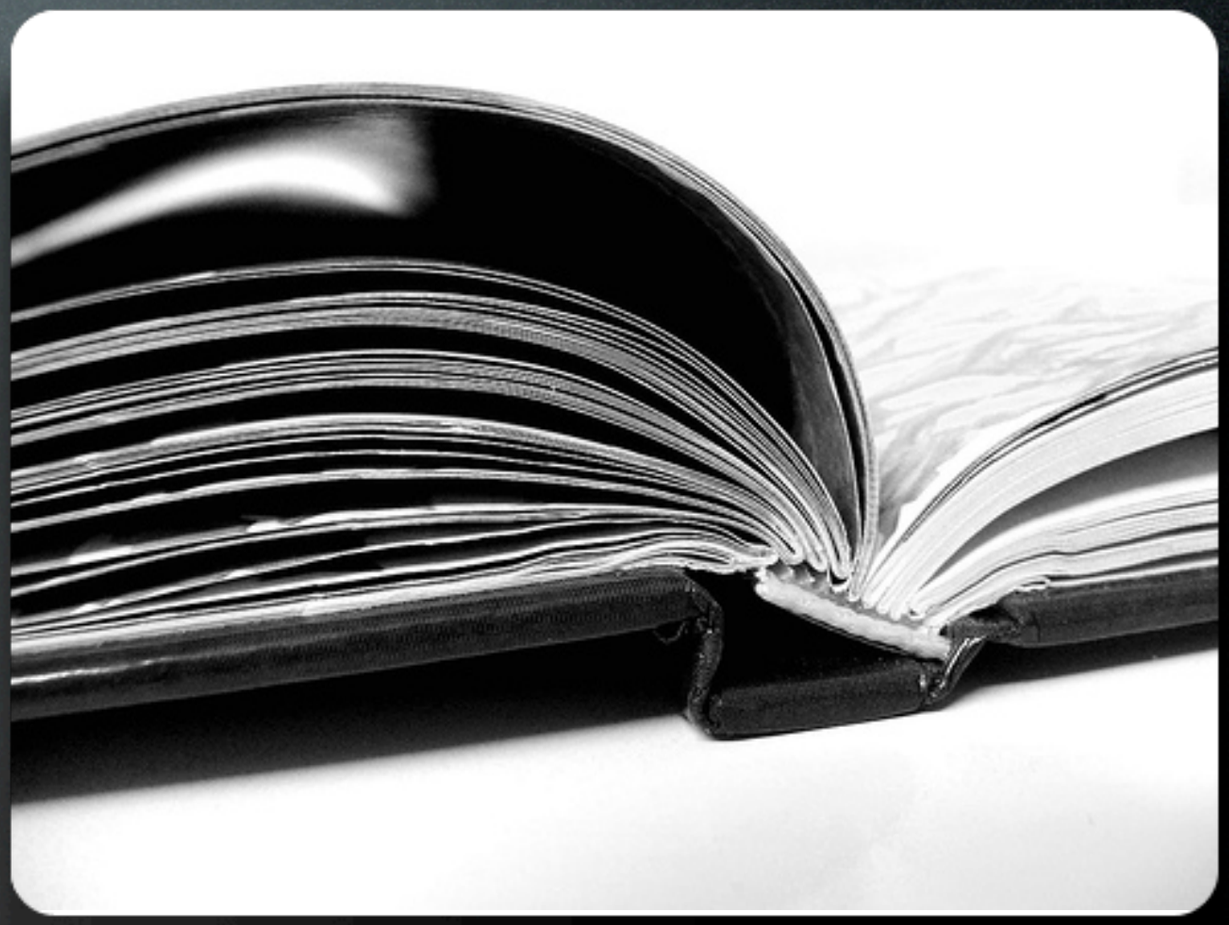
- Bayes-Filter?
  - Who wants to train them?
- We need user-friendly solutions!
- centralized systems may be not accurate enough

Solutions?

- Cooperation?
  - not much
  - "Spam is not a problem any more"
- Where are the commercial solutions?

Resources

- Webspam in general
  - spamhuntress.com
- Wiki-Spam
  - chongqed.org
- My blog
  - spam.tinyweb.net



Credits

- Photos via flickr.com, thanks to: freezelight, Hopkinsii, striatic, chotda, lagiuspo, It'sGreg, lorZ, YnR, kevinthoule, acagamic, R80o (Mark Strozier), Kevin, loungerie, brappy!, ^Sandra^, longwayround, sheeshoo, Orgasmic kmlz, awinn233, teotwawki, Hugo*, rofanator, gyst, Gigglejuice, manuki

Hint: Pictures and keywords are hyperlinked!