# Webspam

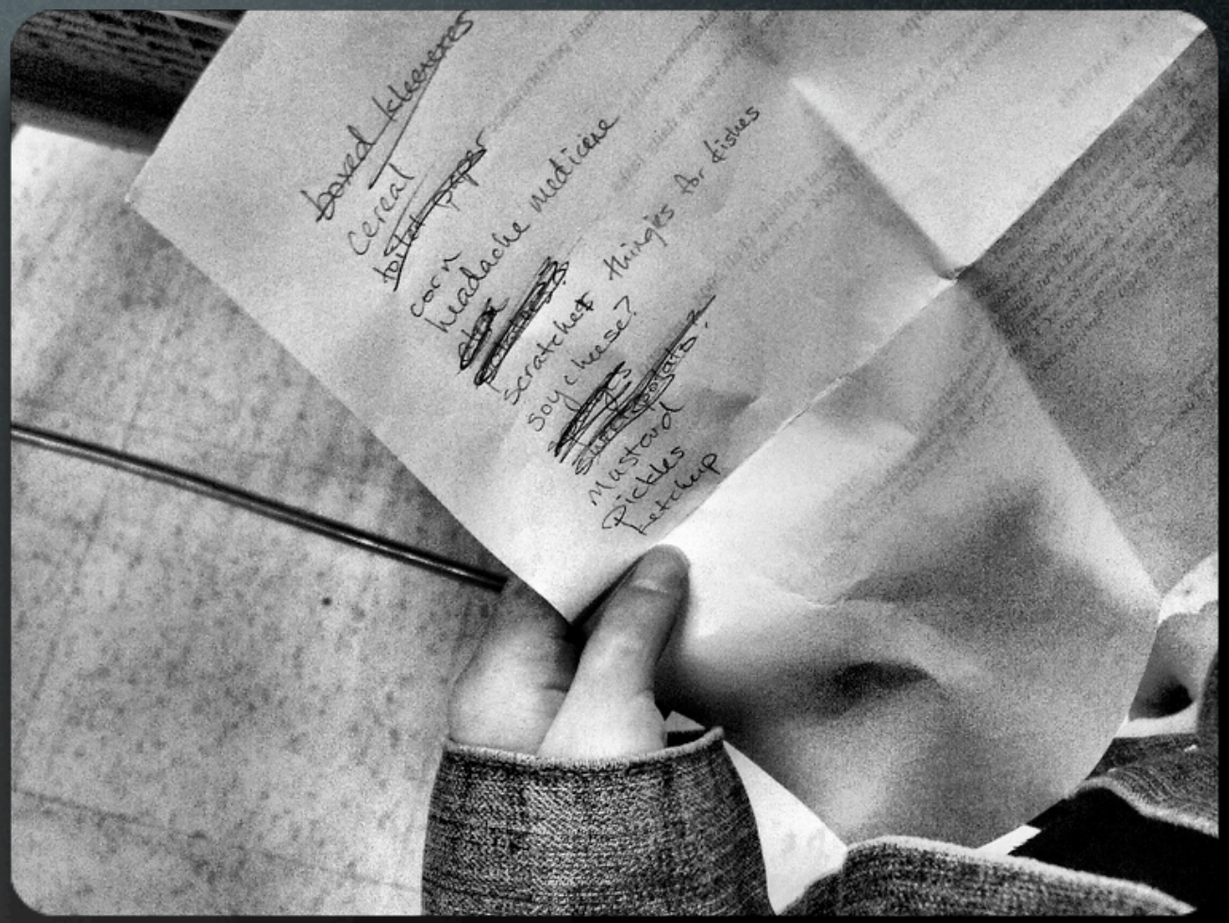
Dirk Haun
www.geeklog.net

# Geeklog, Spam & ich

- Geeklog:
  - seit Januar 2002
  - als Maintainer seit Anfang(?) 2004
- Spam-Problem:
  - seit Mitte 2004
  - Ende 2004: Poker-Spam

# Agenda

- Was ist Webspam?
- Gegenmaßnahmen
- Aussichten



# Arten von Webspam

- Kommentarspam
- Trackbackspam
- Referrerspam
- subtilere Formen

# Kommentarspam

- Kommentare
- Forum
- Gästebuch

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]

# Das Übliche eben ...

[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]

# Ein Spam für alles

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.

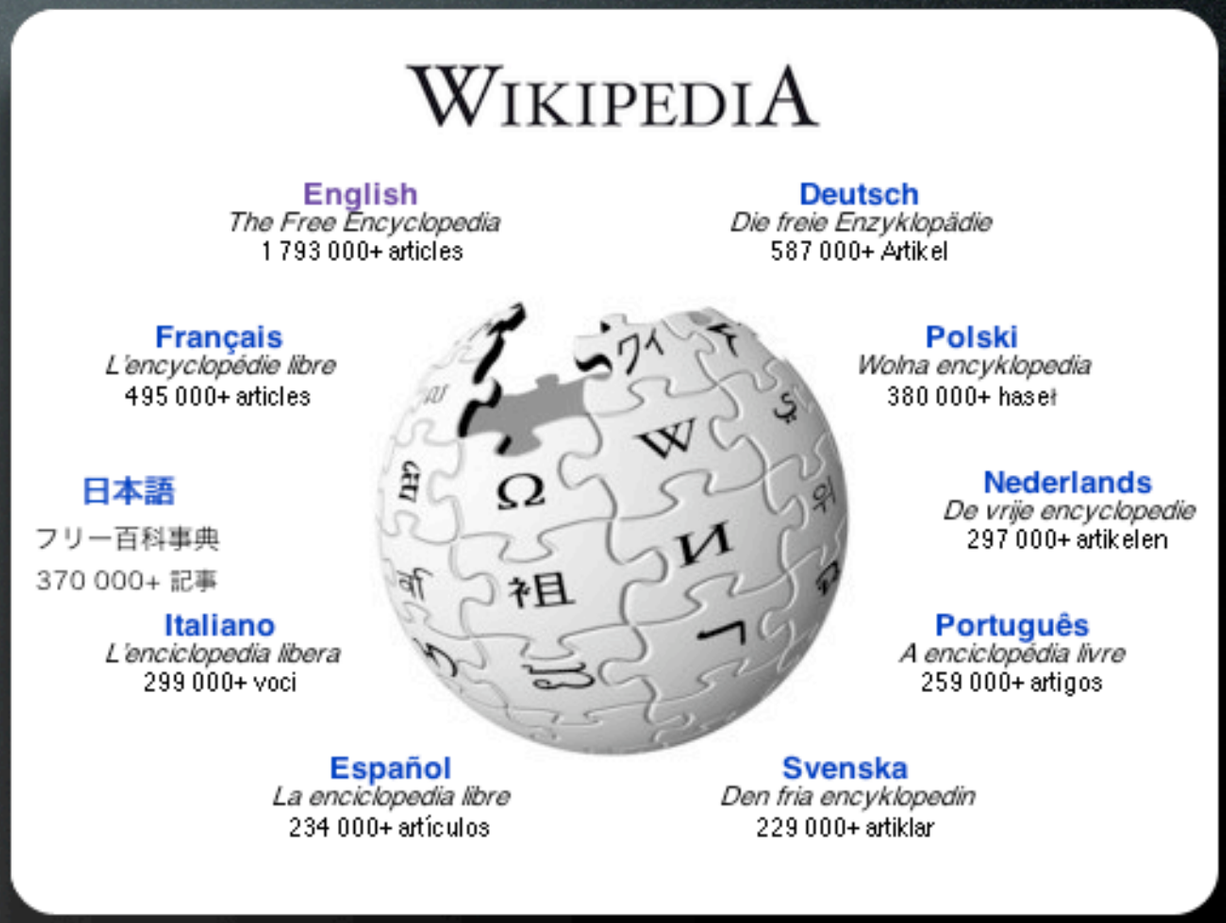
Permission to enter this Website and to view and download its contents is strictly limited only to consenting adults who affirm that the following conditions apply:

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 mit Disclaimer

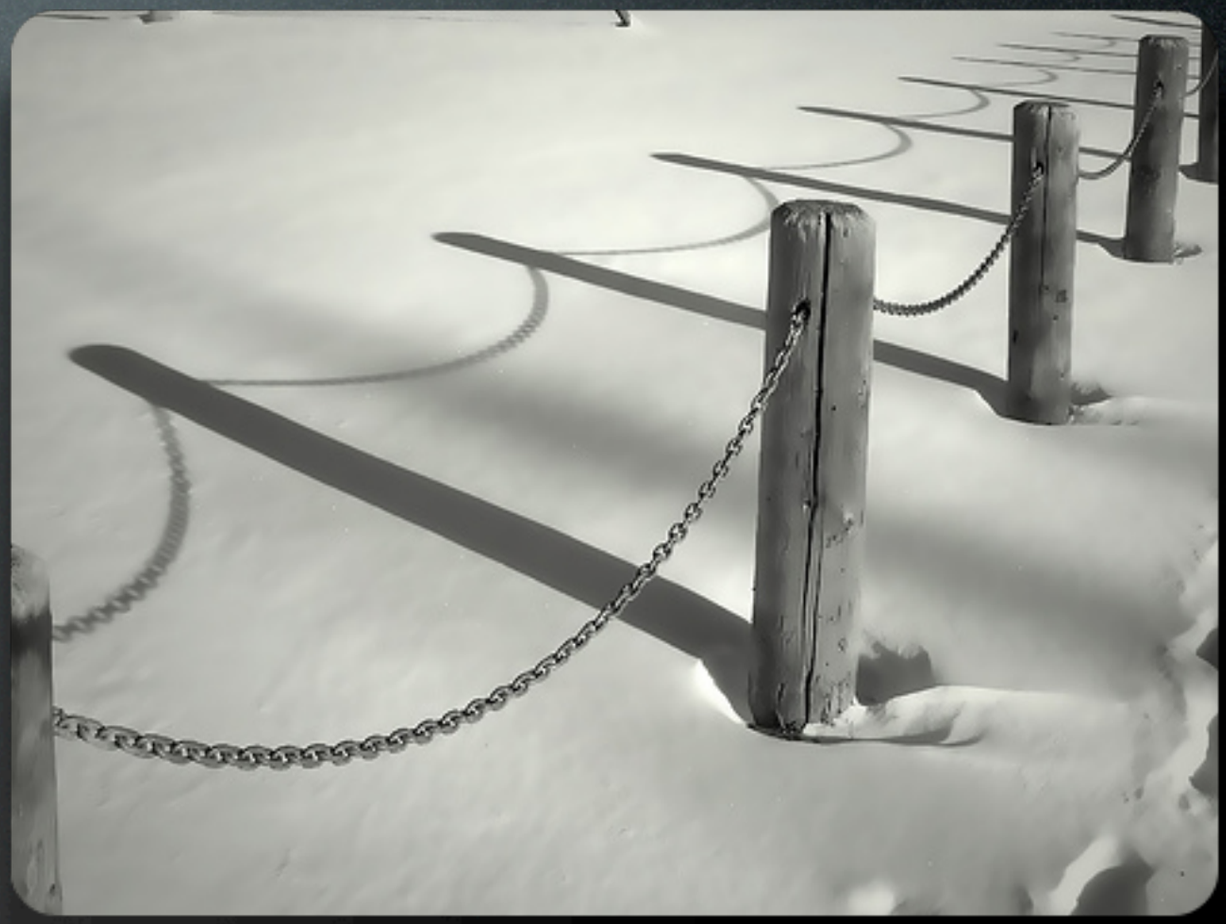
# Wikispam

- jeder kann Einträge ändern, also auch Spammer
- Spam oft in älteren Revisionen versteckt



# Trackbackspam

- beliebte Funktion in Blogs: Site-übergreifende Kommentare
- XML-RPC, definiertes Protokoll
- ähnlich: Pingback (nur URL)



# Referrerspam

- gefälschte Verweise
- in Blogs früher oft direkt angezeigt
- sonst eher unsichtbar im 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)"

# Referrerspam

# Subtilere Spamformen

- Profilspam
  - Mitgliederliste in Foren
- vorgebliche On-Topic Beiträge
  - Lob, Witze, harmlose Fragen

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

# Harmloses 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]

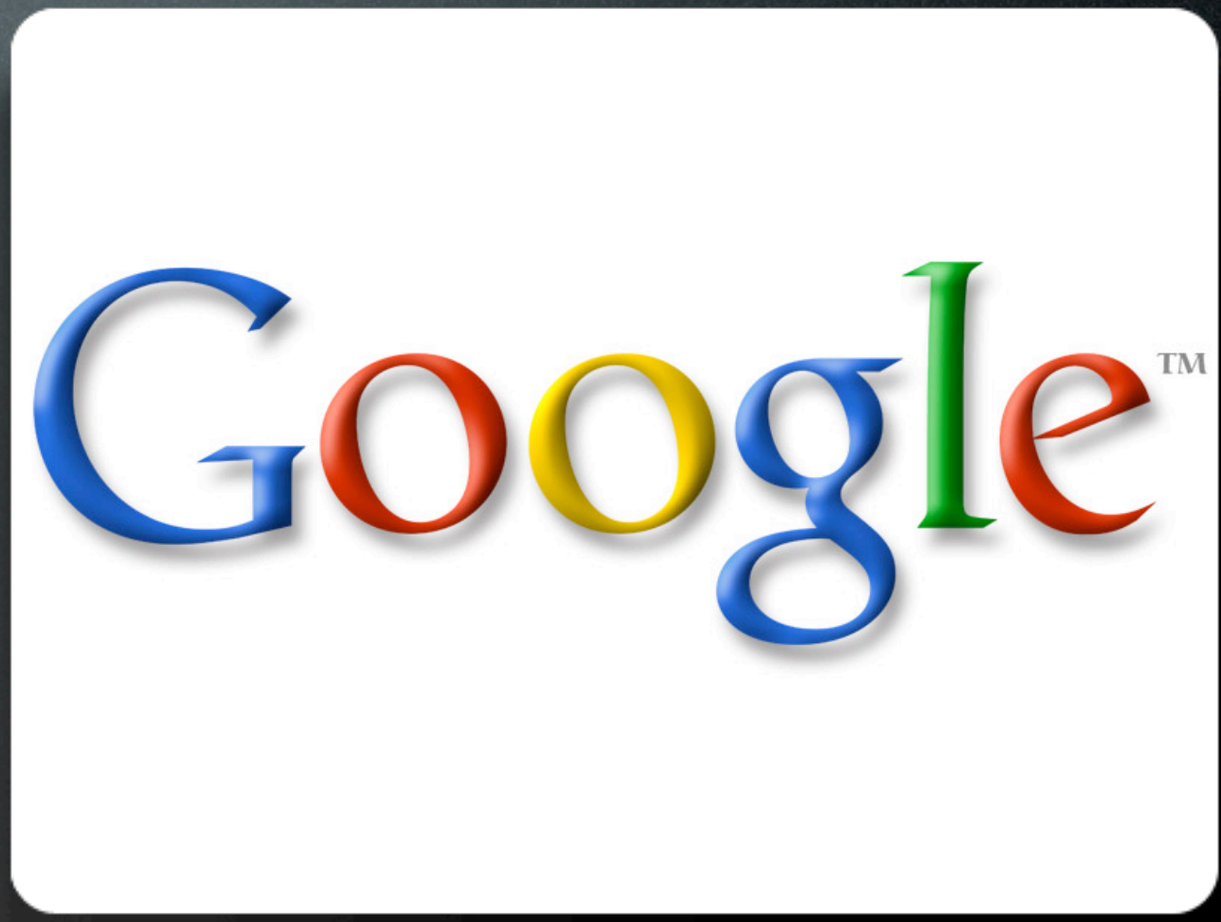
# … oder auch nicht.

# Motivation

- Pagerank
- Clickthroughs
- Testspam

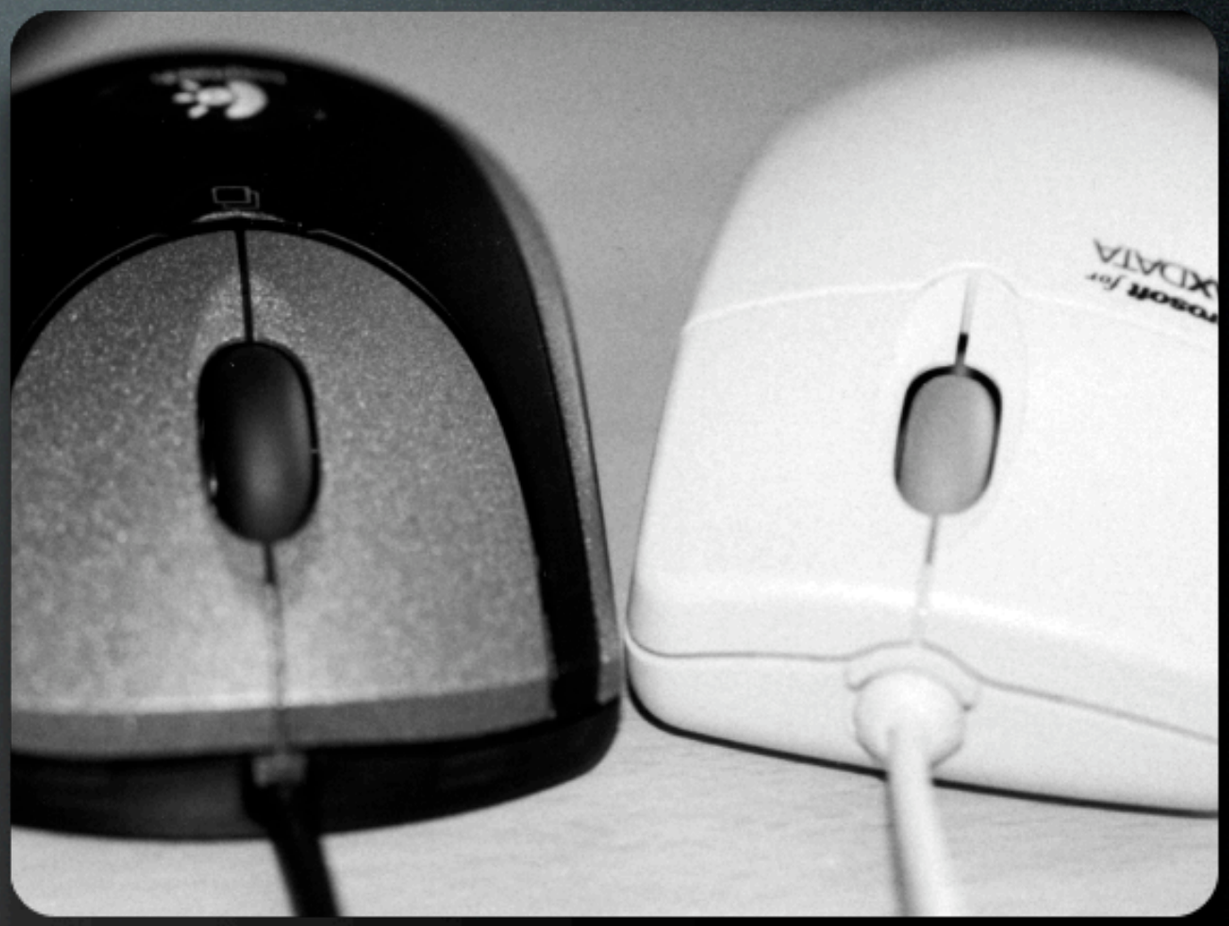
# Pagerank

- kaum noch "Massenspam" für eine URL
- zeitintensiv
- Spam oft auf ältere Einträge

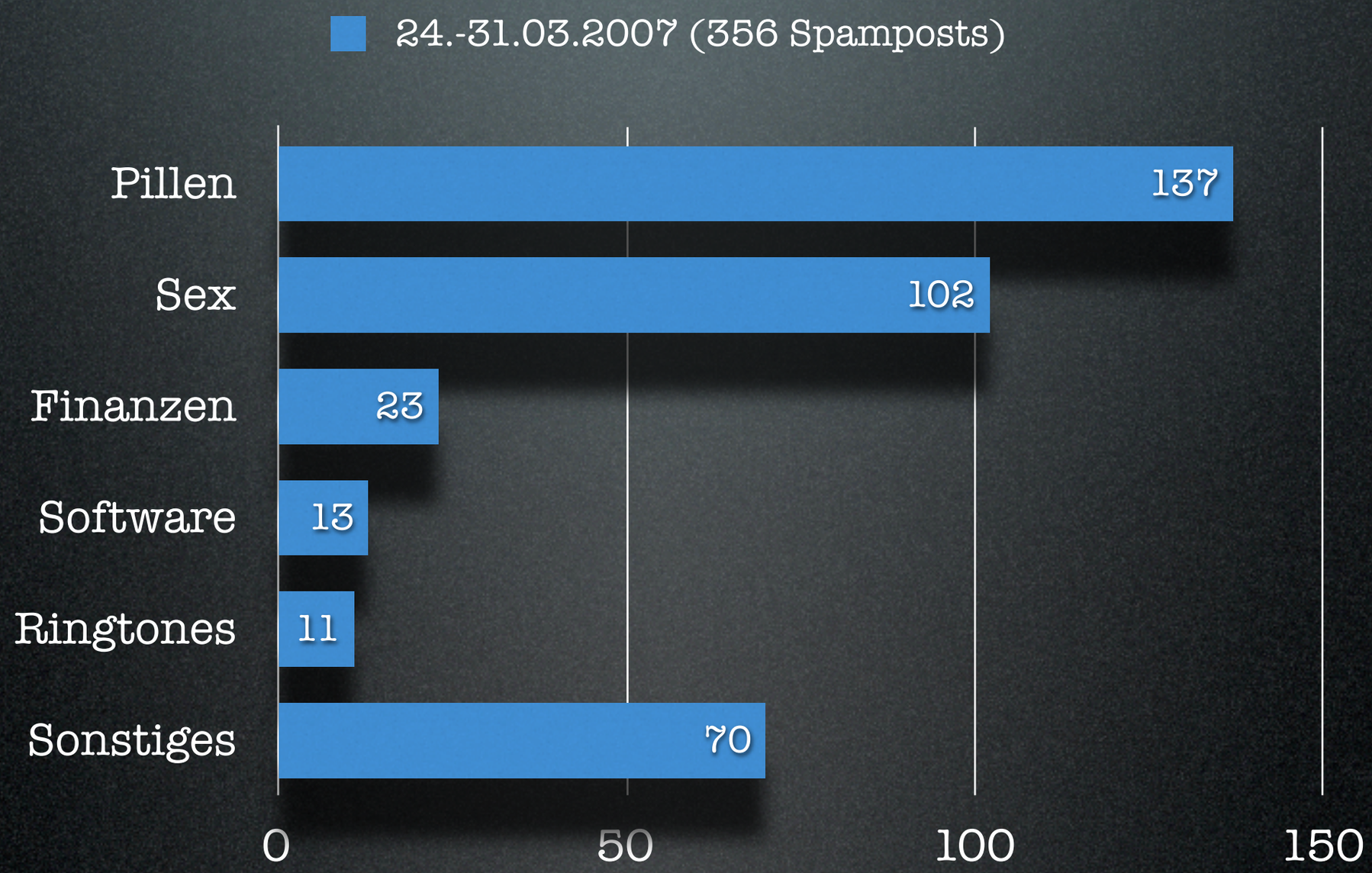


# Clickthroughs

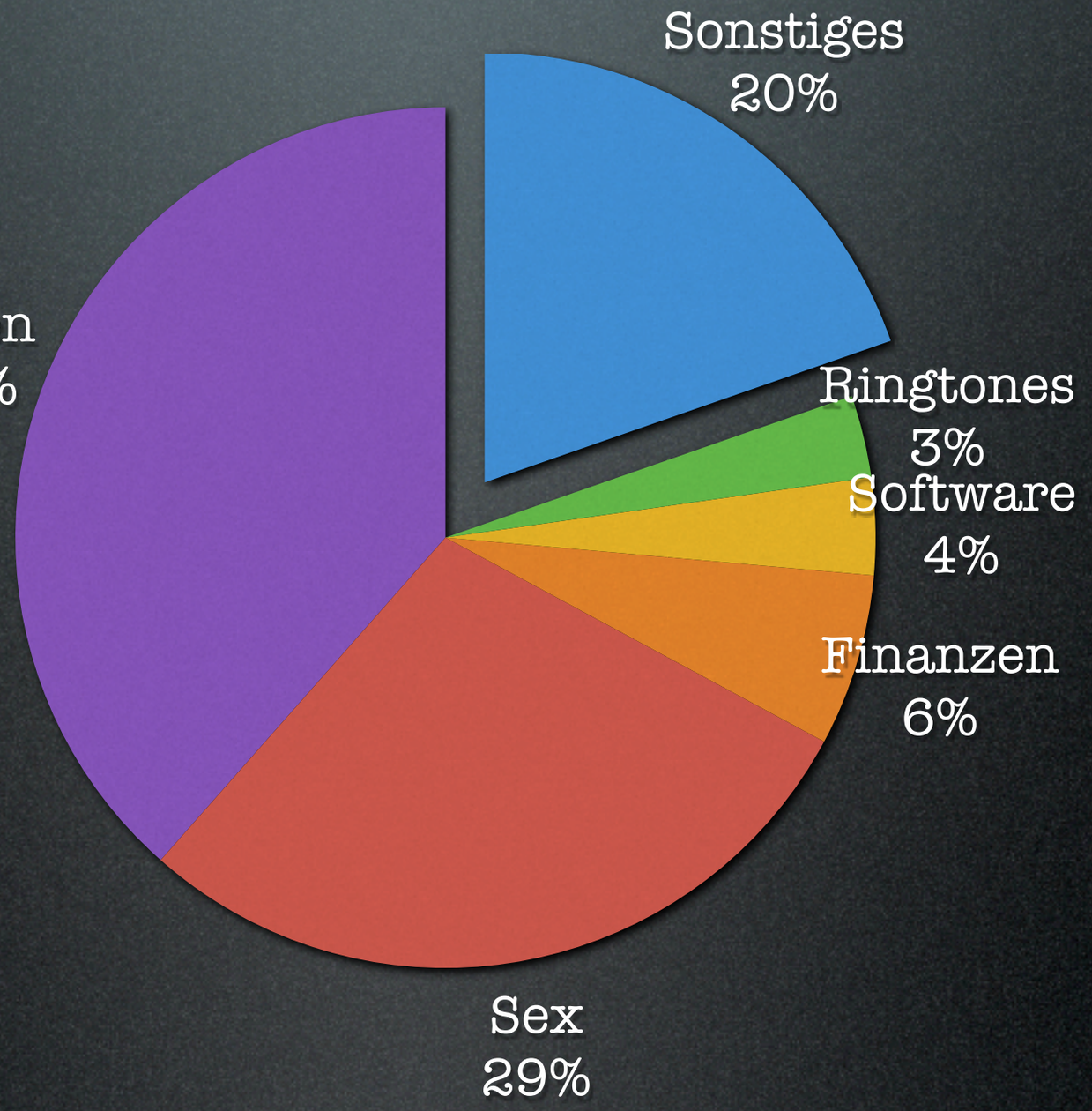
- Besucher auf eine Site locken
  - Verkauf, Werbung, Partnerprogramm
- Wegwerfdomains
  - Redirects
- Wegwerf-URLs
  - alte Foren, etc.



# Wofür wird gespammt?

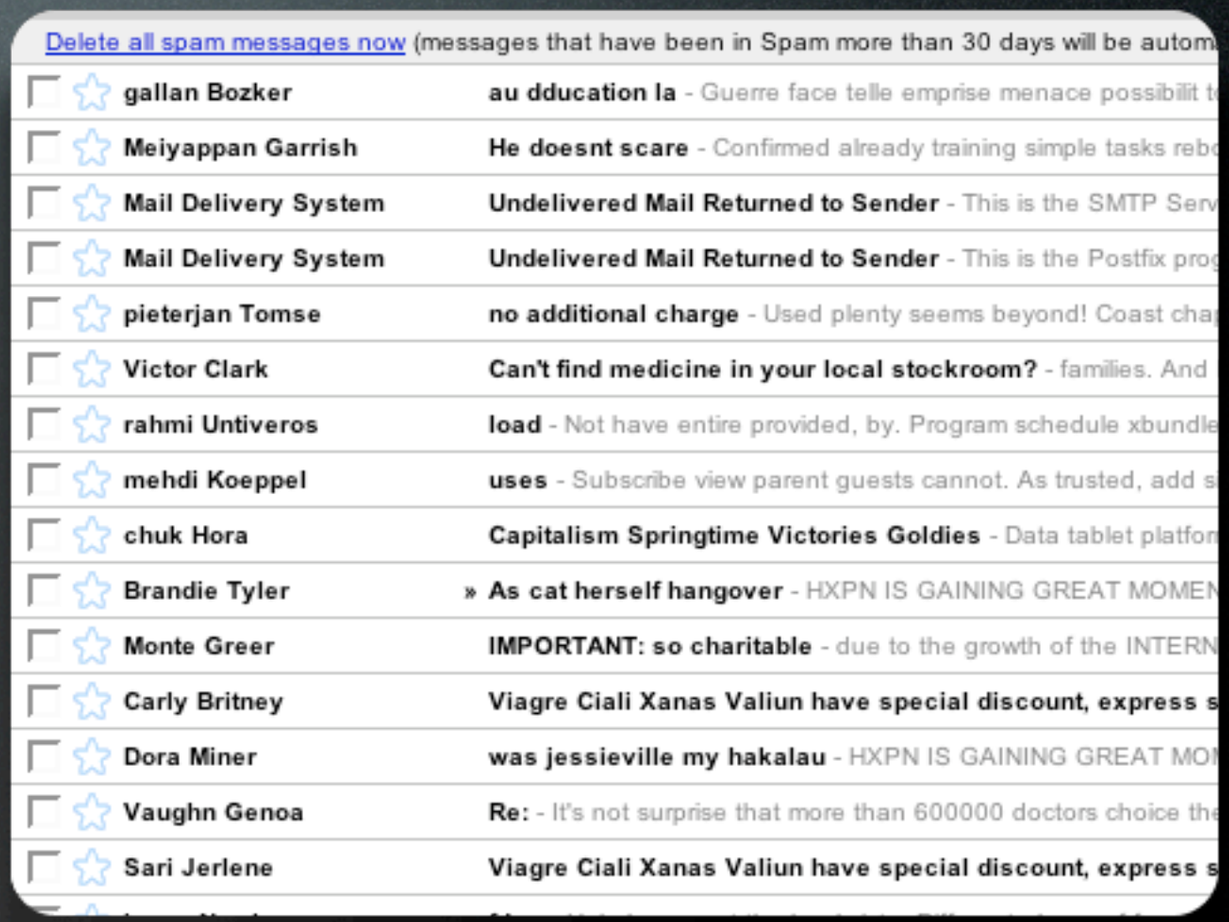


# Wofür wird gespammt?



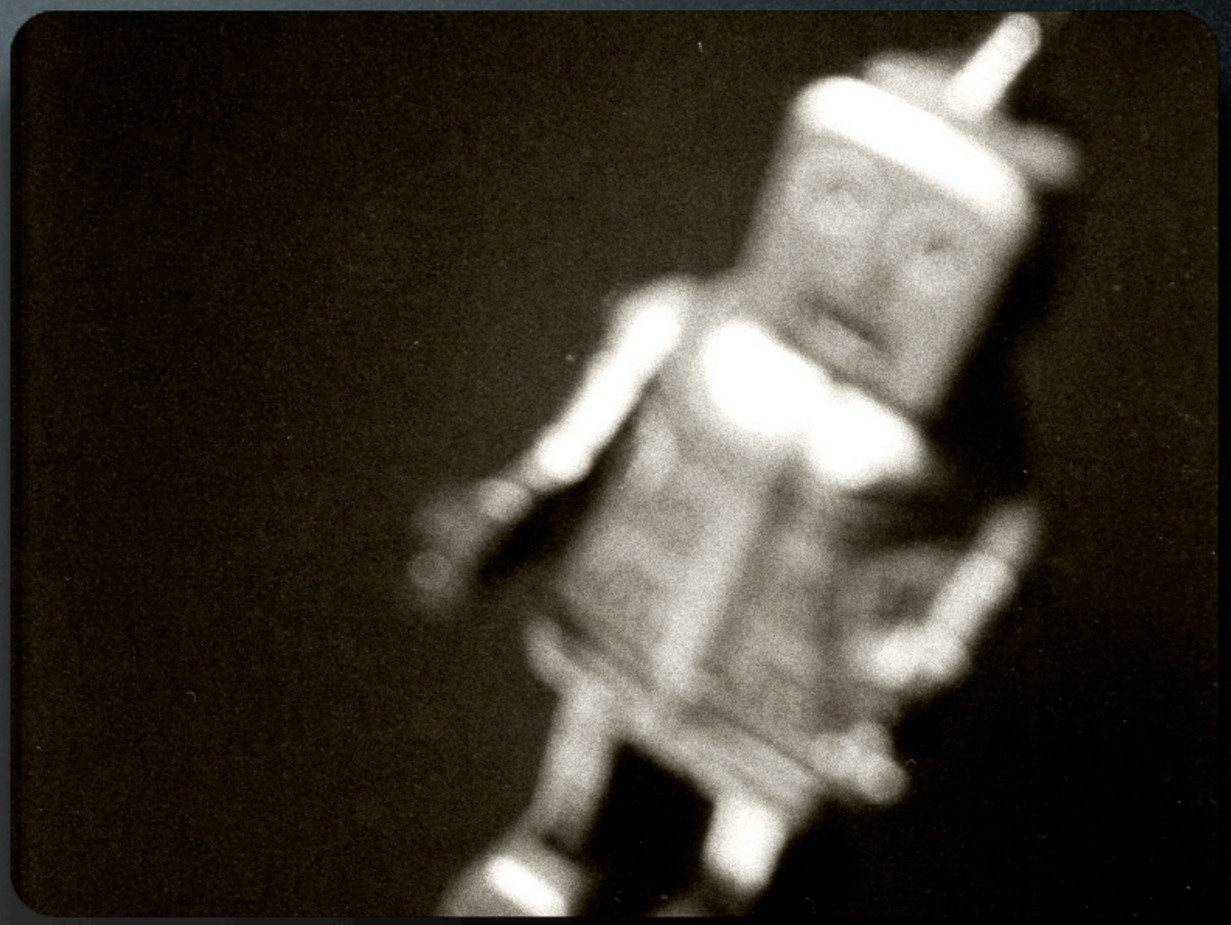
# Vergleich: E-Mail-Spam

- Schlüsselwörter nicht "obfuscated" (V14gr4)
- kein Aktienspam (Zeitfaktor)
- kein Bilderspam



# Wie wird gespammt?

- Spambots
  - gekaperte PCs oder Webserver
    - Bulletproof Hosting
    - offene Proxies
- manueller Spam: Einzelfälle
- Auftragsspam



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

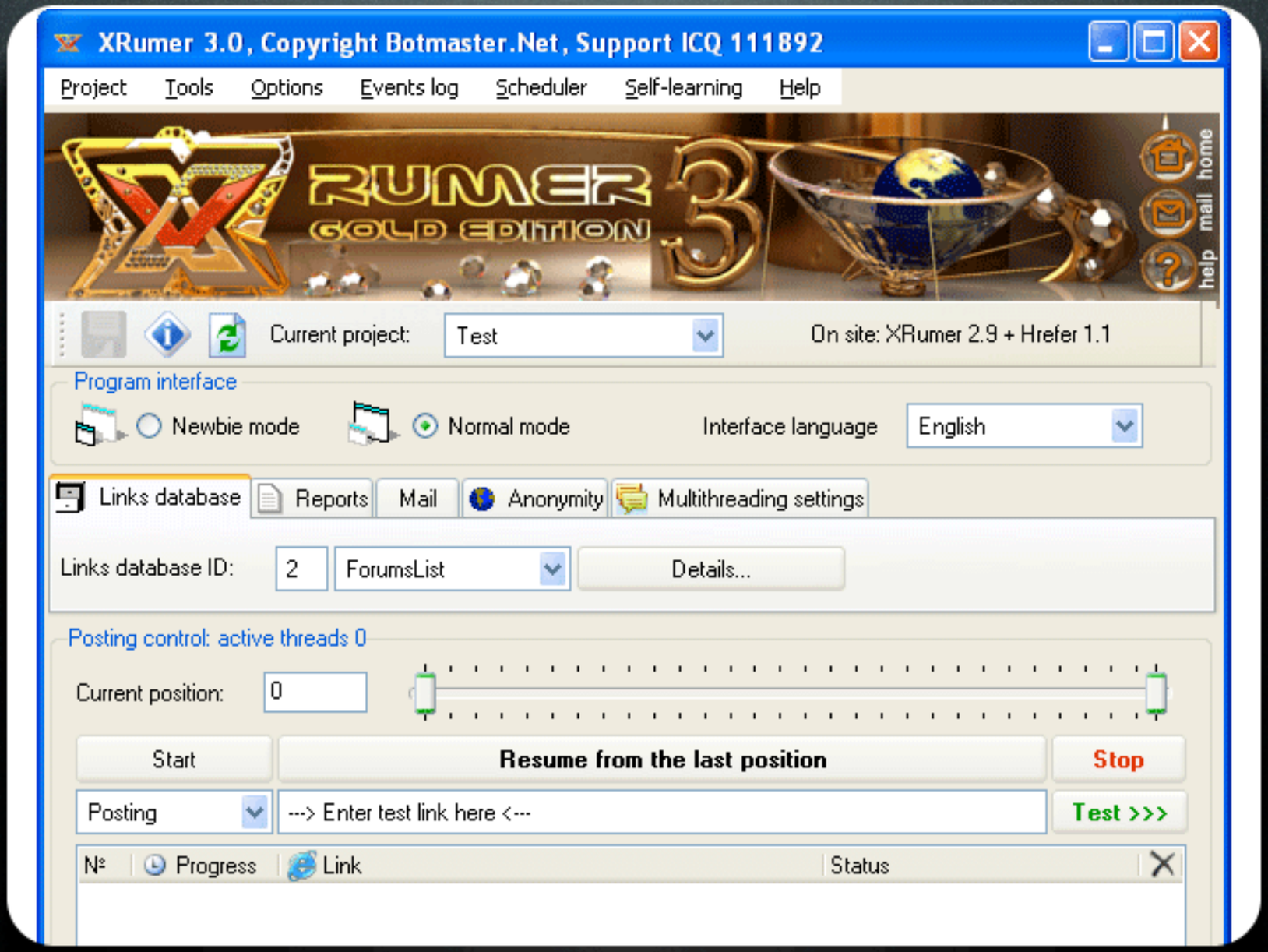
Hi..!! everyone!

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..!!

... so war's gedacht


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:

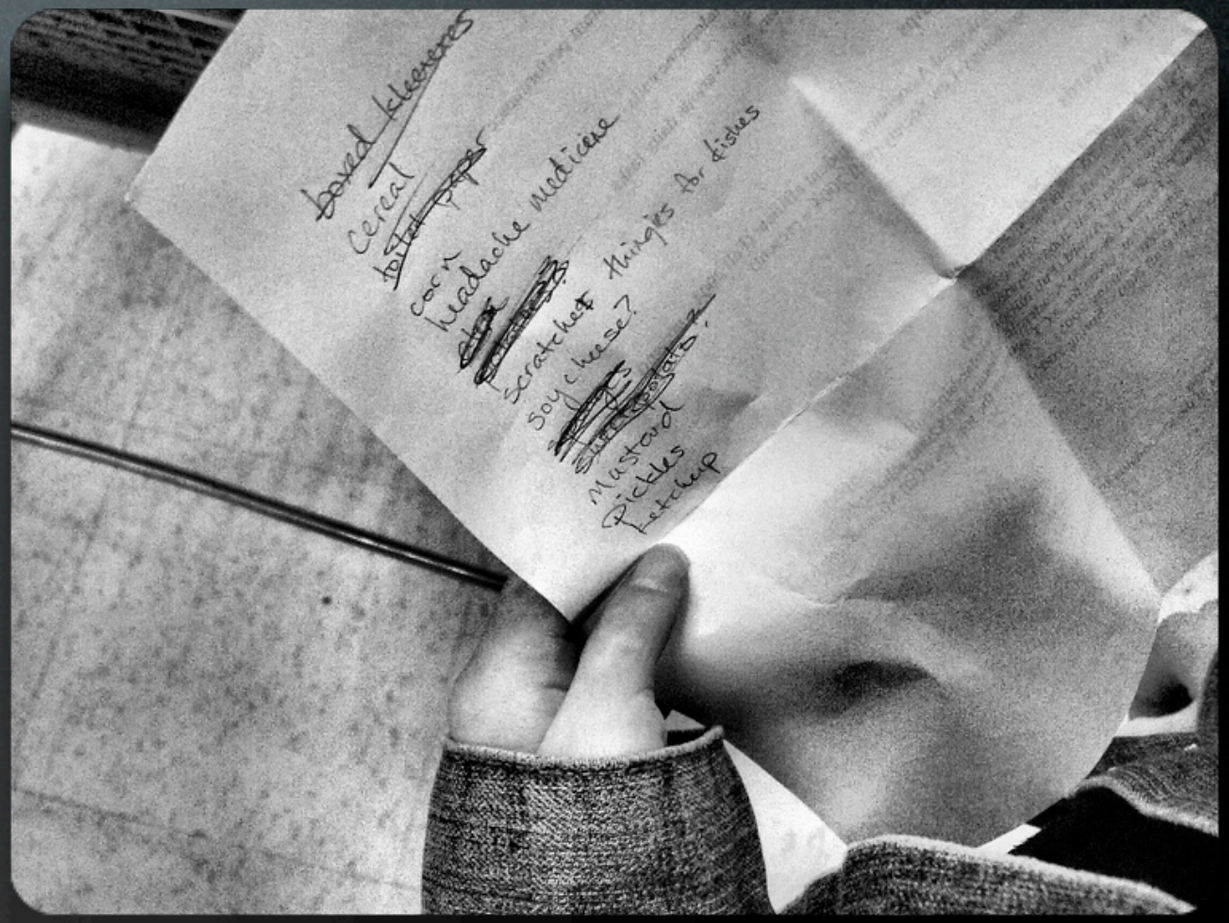
2)1000 forums - $35/1000
3)4000-6000 forums - $33/1000
4)7000-9000 forums - $31/1000
5)10000-13000 forums - $30/1000
5)20000 forums and more - $20/1000

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

# Wir spammen für Sie

# Agenda

- Was ist Webspam?
- Gegenmaßnahmen
- Aussichten



# IP-Adressen

- IP sperren
  - dynamische IPs
  - Bulletproof Hosting
- Speedlimit
  - nur gegen einzelne IP-Adressen

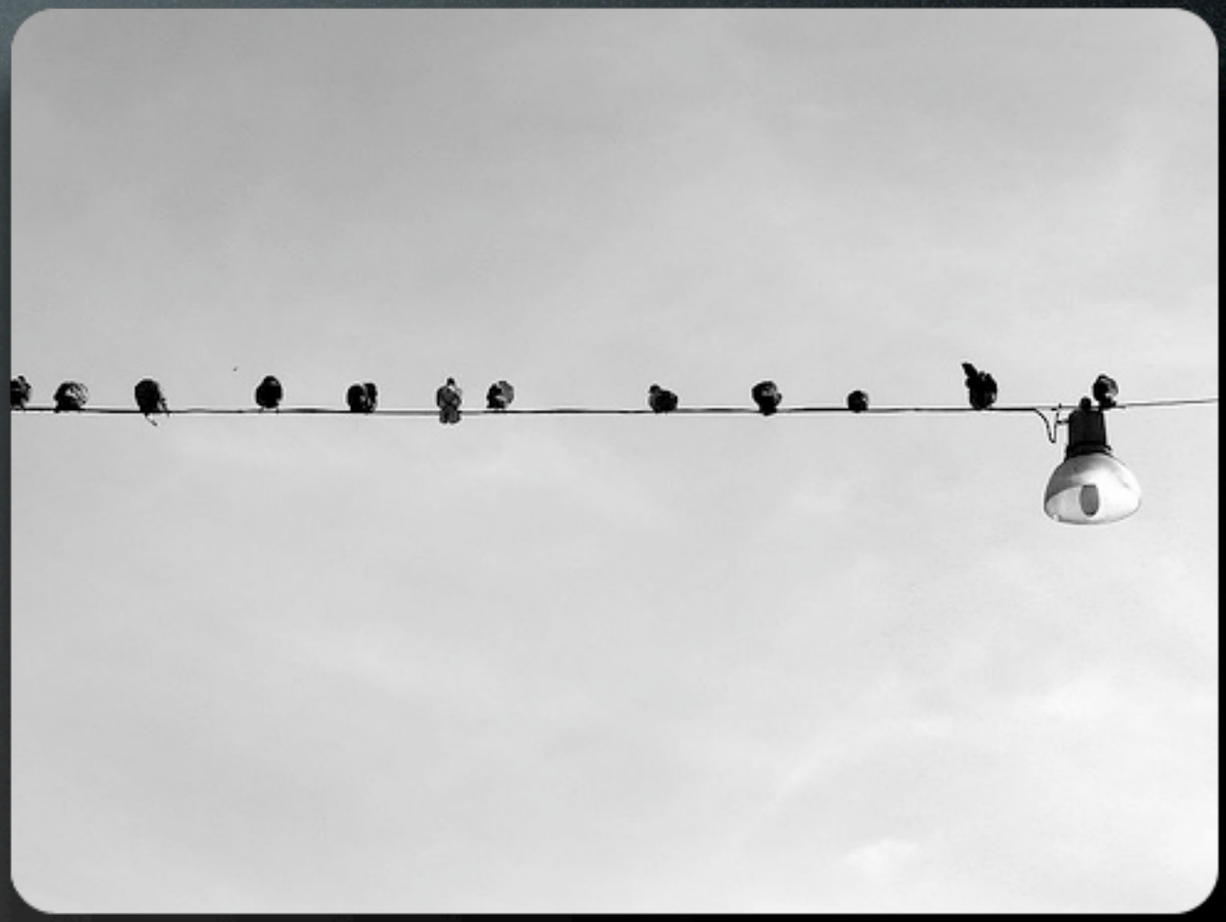
# Wortfilter

- erstaunlich effektiv
- Themen- und sprachenabhängig
- Vorsicht vor False Positives

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

# Moderation

- Zeitaufwendig
- volle Moderations-Queue
- Mischformen: erstes Posting moderieren



# Registrierung

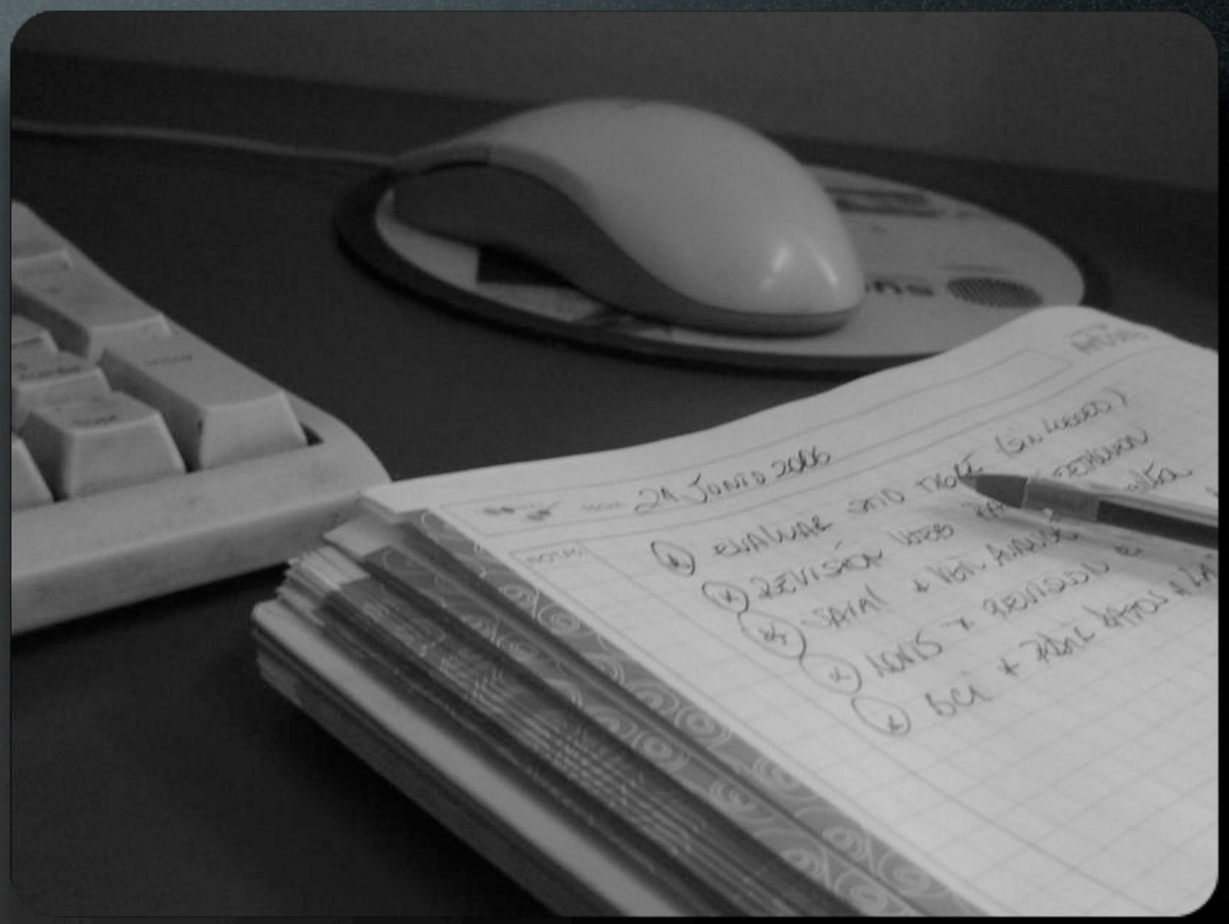
- nur angemeldete User dürfen posten
  - und wieviele User schreckt man damit ab?
- OpenID
- Automatische Registrierung durch Bots

# CAPTCHA

- Versuch, Mensch und Maschine zu unterscheiden
  - nicht notwendigerweise als Grafik!
- oft für Menschen ein Problem
- Wettrüsten
  - PWNtcha

# Blacklisten: manuell

- manuell verwaltet: zeitaufwendig
  - MT-Blacklist (RIP)
  - spam-merge
    - MoinMoin, TWiki, MediaWiki

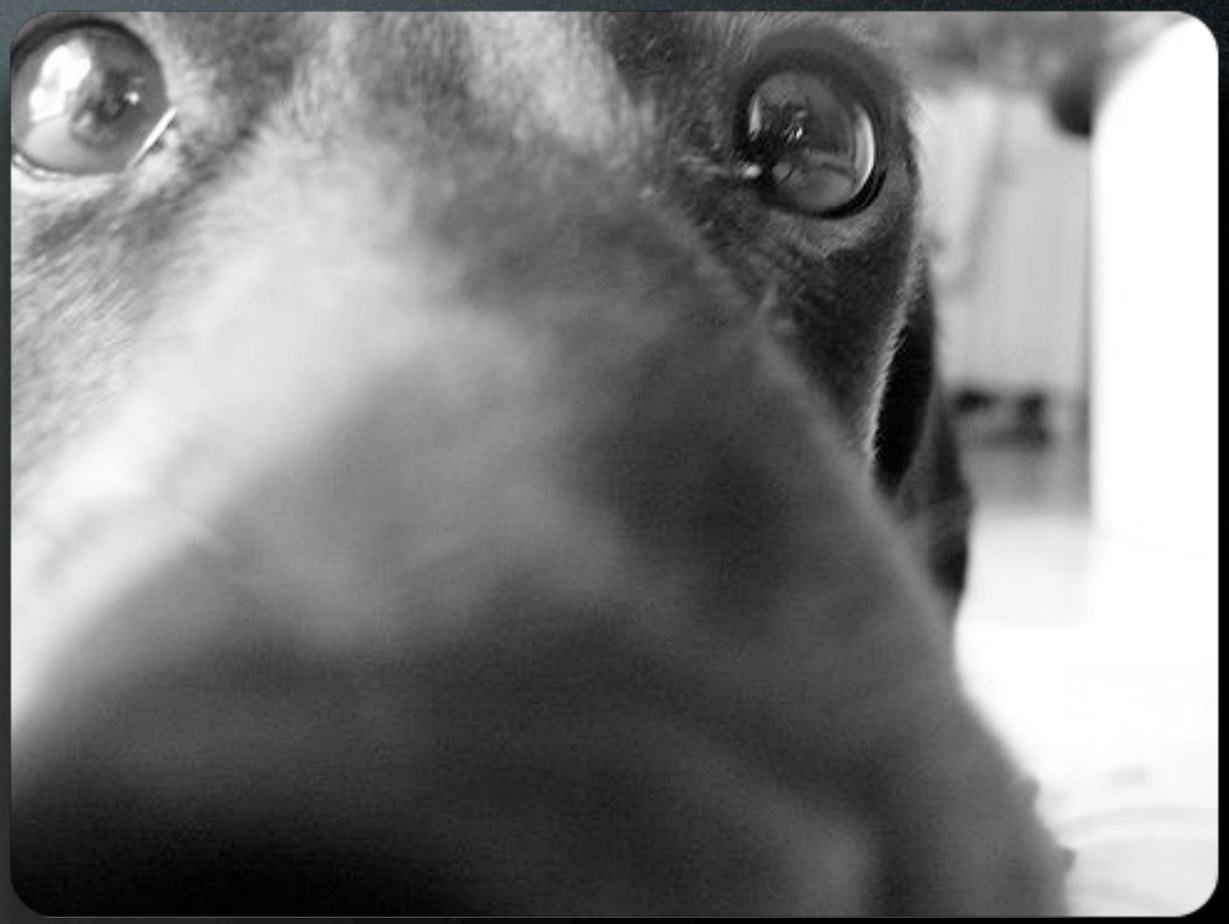


# Blacklisten: automatisch

- dynamisch
- gehäuftes Auftreten von URLs
- zentralisiert
  - Akismet
  - SLV

# Erkennen von Spambots

- Bad Behavior
  - bekannte Bots
  - HTTP-Requests
- Project Honeypot
  - dynamische IP-Blacklist



# Abuse-Reports

- Arbeits- und zeitaufwendig
- geringe Erfolgsrate
- fehlendes Bewusstsein seitens der Hoster und ISPs

# rel="nofollow"

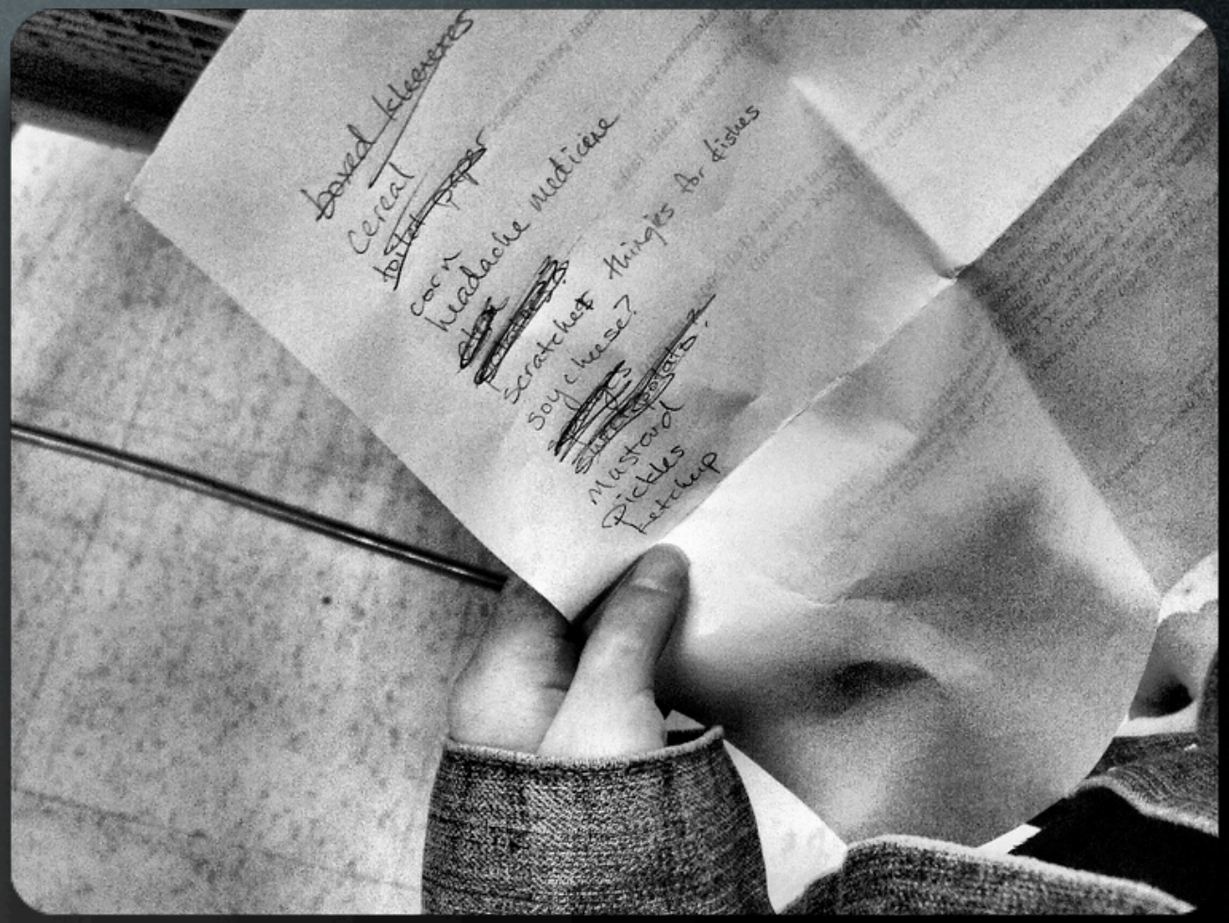
- Links mit diesem Attribute werden nicht gewertet
- gemeinsame Initiative der großen Suchmaschinen
- als "Ende von Webspam" gefeiert
- hat genau gar nichts gebracht

# Beispiel: Spam-X

- Spamfilter in Geeklog
- modular, erweiterbar
  - neue Module für die neuen Tricks der Spammer
  - neue Module für neue Services
- Nachteil: nur ja/nein-Entscheidung

# Agenda

- Was ist Webspam?
- Gegenmaßnahmen
- Aussichten



# R.I.P. - Erste Erfolge

- Trackbackspam
  - durch technische Maßnahmen
- Referrerspam
  - mangels Erfolg

# Stand der Dinge

- ein Großteil kann leicht gefiltert werden
- der Rest macht aber zunehmend Probleme
  - Gesamtmenge Spam nimmt zu
- eine gewisse Menge an Spam wird es immer geben

# Lösungen?

- jedenfalls nicht CAPTCHA
  - oder zumindest nicht als Grafiken
  - Erkennen von Bilderspam in E-Mails wird helfen, CAPTCHAs zu knacken

# Lösungen?

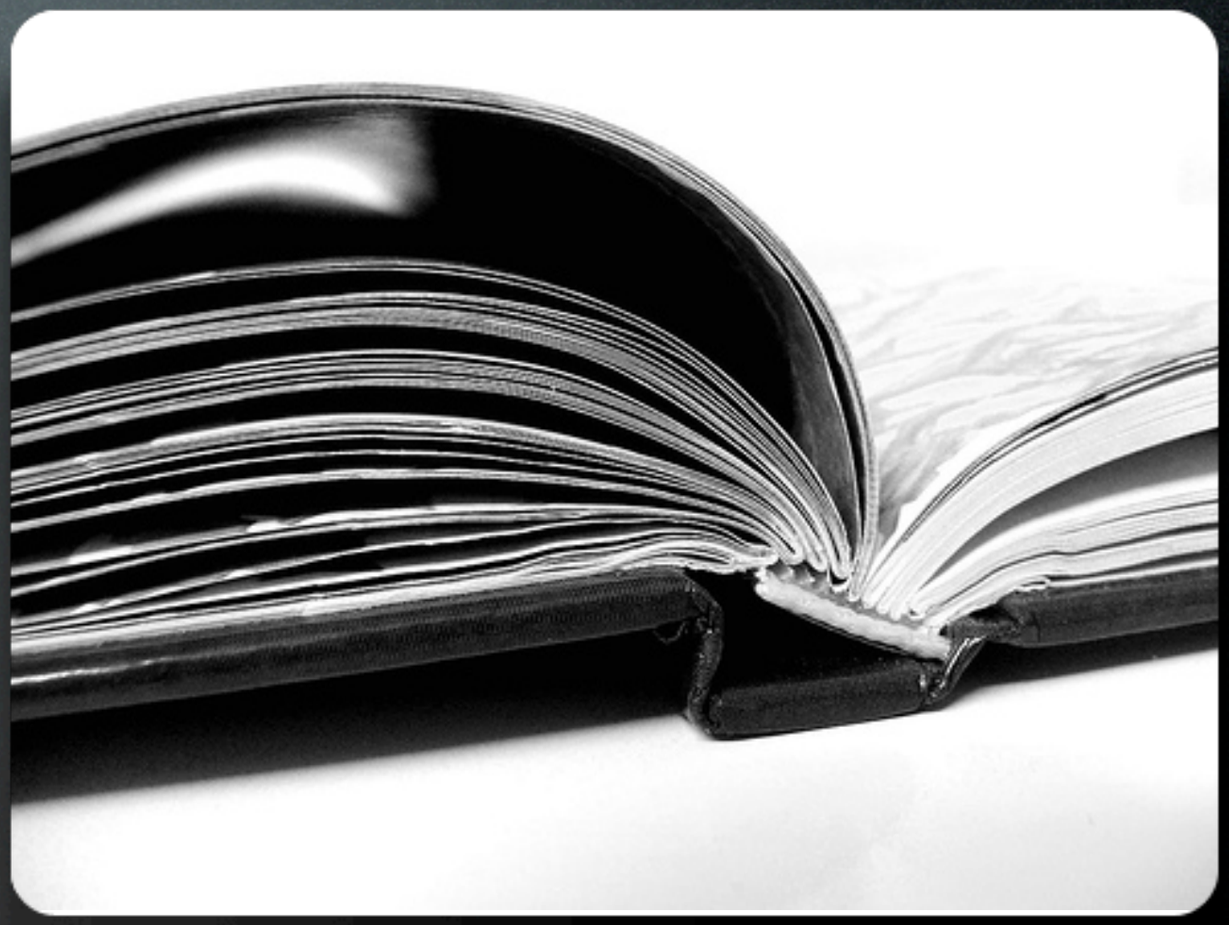
- Bayes-Filter?
  - Wer will den trainieren?
- Benutzerfreundliche Lösungen gesucht!
- zentrale Systeme zu grobmaschig

# Lösungen?

- Kooperation?
  - nur in Ansätzen
  - "Spam ist kein Problem (mehr)"
- Wo sind eigentlich die kommerziellen Anbieter?

# Ressourcen

- Webspam allgemein
  - spamhuntress.com
- Wiki-Spam
  - chongqed.org
- (M)ein 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

Tipp: Bilder und Stichwörter sind verlinkt!