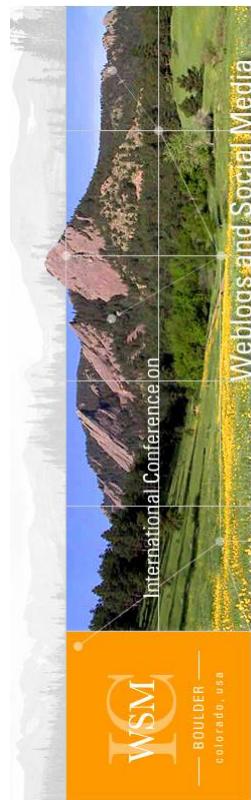


Spam in Blogs and Social Media

Pranam Kolari, Tim Finin
Akshay Java, Anupam Joshi

March 25, 2007



Tutorial Outline

- Spam on the Internet
 - Variants
 - Social Media Spam
- Reason behind Spam in Blogs
- Detecting Spam Blogs
- Trends and Issues
- How can you help?
- Conclusions

Presenters



Pranam Kollar is a UMBC PhD student. His dissertation is on spam blog detection, with tools developed in use both by academia and industry. He has active research interest in internal corporate blogs, the Semantic Web and blog analytics.



Tim Finin is a UMBC Professor with over 30 years of experience in the applying AI to information systems, intelligent interfaces and robotics. Current interests include social media, the Semantic Web and multi-agent systems.



Akshay Java is a UMBC PhD. His dissertation is on identifying influence and opinions in social media. His research interests include blog analytics, information retrieval, natural language processing and the Semantic Web.



Anupam Joshi is a UMBC Professor with research interests in the broad area of networked computing and intelligent systems. He currently serves on the editorial board of the International Journal of the Semantic Web and Information.

SPAM!

E-mail Spam

- Early form seen around 1992 with MAKE MONEY FAST
- 80-85% of all e-mail traffic is spam
- In numbers

2005 - (June) 30 billion per day

2006 - (June) 55 billion per day

2006 - (December) 85 billion per day

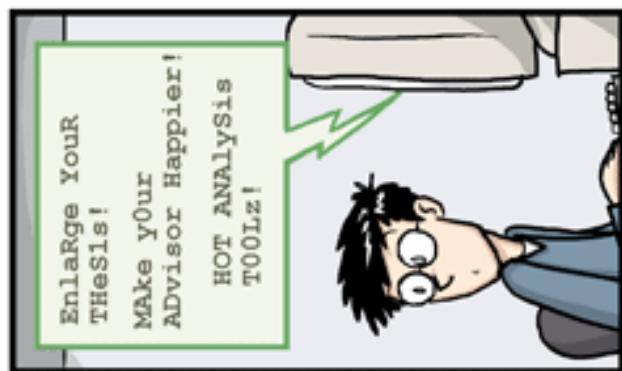
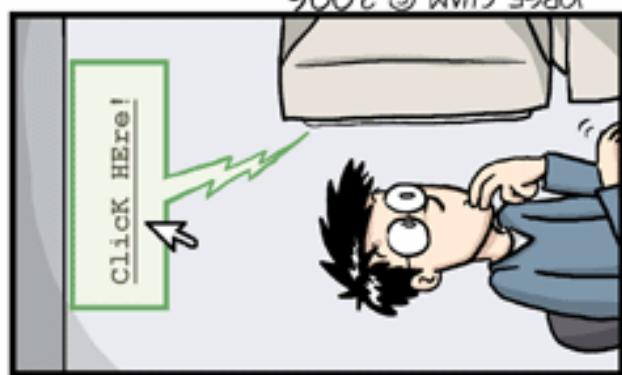
2007 - (February) 90 billion per day

Sources: IronPort, Wikipedia
http://www.ironport.com/company/ironport_pr_2006-06-28.html



JORGE CILAM © 2006

www.PhDcomics.com



What is Spam?

- “Unsolicited usually commercial e-mail sent to a large number of addresses” – Merriam Webster Online
- “Spamming is the abuse of electronic messaging systems to send unsolicited bulk messages, which are almost universally undesired.” – Wikipedia
- As the Internet has supported new applications, many other forms are common, requiring a much broader definition

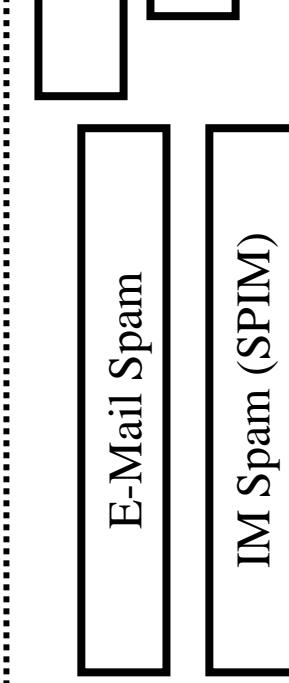
Capturing user attention unjustifiably on the Internet (E-mail, Web, Social Media etc..)

Taxonomy of Spam

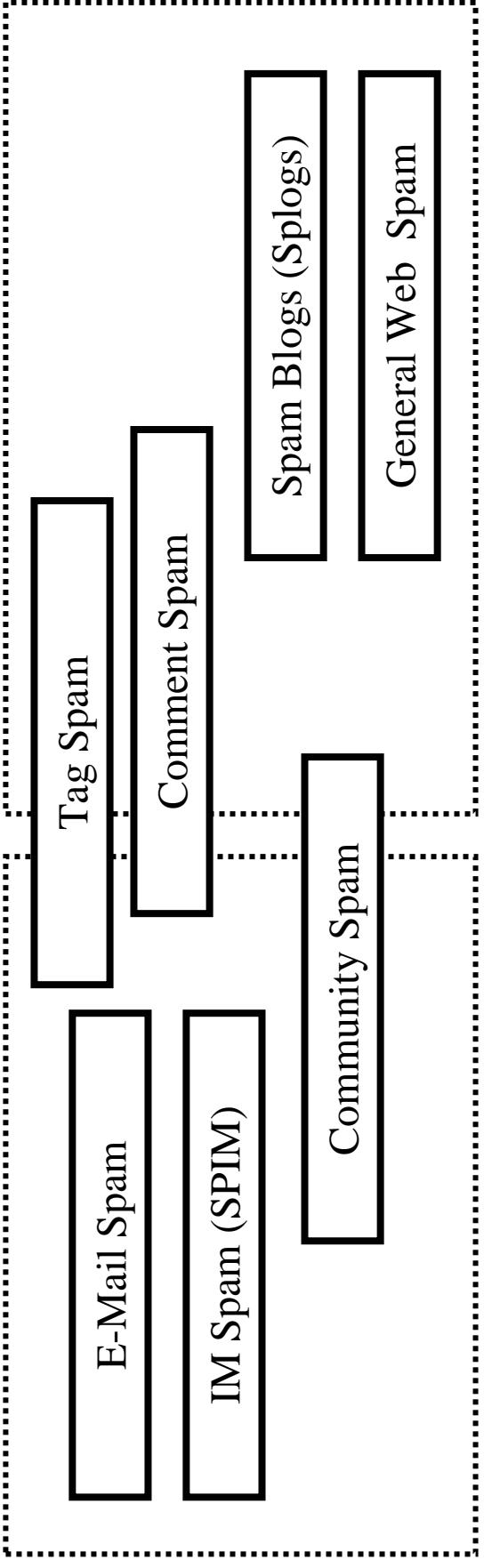
Internet Spam

(Forms)

Direct



Indirect



(Mechanisms)

spamdexing

Spamdexing

“We use the term spamming (also, spamdexing) to refer to any deliberate human action that is meant to trigger an unjustifiably favorable relevance or importance for a page, considering the page’s true value”
- Gyongyi et al

- Compromises Relevance and Importance Scores used by Search Engines
- Mostly effects the long tail of keywords (more susceptible)
- Spam in Blogs is a form of spamdexing
 - 20% of the indexed Web (2005)

Results 1 - 10 of about 1,420,000 for [defaulted student loan consolidation](#). (0.18 seconds)

Sponsored Links

Lower payments up to 58%
Some of the top benefits available
For help call us at (800) 256-5105
www.StudentLoanSolutions.com

4.5% Fixed Consolidation
Lock in rate now as low as 4.5%.
Pre-qual in 1 minute (866) 810-4275
www.NextStudent.com

Consolidate with Wachovia
Cut Your Monthly [Student Loan](#)
Payment in Half. We Make it Easy.
www.Wachovia.com

Defaulted Student Loan
Get out of debt fast by knowing
your options. Nonprofit, learn more
www.DebtReliefOptions.org

Defaulted Student Loan
Directory Of [Student Loan](#) Deals.
Find [Student Loans](#) Quickly.
[StudentLoan.WealthyGeek.com](http://www.WealthyGeek.com)

Get Out Default
Stop Garnishment and Tax Seizures
and reduce your interest by 1.25%
mystudentloanblock.com

Student Loan DEFAULT Help
Stop tax offset, garnishment, etc.
Act NOW before it gets worse!
www.siro.net

4.5% Fixed Consolidation
Lock in a low 4.5% rate. Reduce
your payment by 60%. No Fees.
www.LendFocus.com/Student

Web

[2 results stored on your computer](#) - Hide - About

[student loan consolidation](#) - [loan consolidation](#). student loan consolidation

[Student Loan Consolidatio...](#) - 2,2006 Student Loan Consolidation Rates to

Repaying Student Loans Held by the U.S. Department of Education

1 gif Bullet Loan Consolidation 1 gif Bullet Going Back to School ... "Can I pay my **defaulted** student loan held by the Department by credit card?..."

www.ed.gov/offices/OSFAP/DCS/repaying.html - 23k - Cached - Similar pages - Note this

Loan Consolidation

US Department of Educations Collections for **Defaulted Student** loans Guide - Information about **loan consolidation** options that are available to borrowers.

www.ed.gov/offices/OSFAP/DCS/consolidation.html - 19k - Cached - Similar pages - Note this

Student Loan Solutions: Student College Loan Consolidation...

StudentLoanSolutions is a premier lender with the Department of Education. Call 1-800-256-5105. StudentLoanSolutions offers some of the best borrower ...

www.studentloansolutions.com/ - 9k - Cached - Similar pages - Note this

student loan consolidation program's bookmarks tagged with ... - 2,26pm

Are you looking for **defaulted student loan consolidation**? Please click here for more information about **defaulted student loan consolidation**! [consolidation](#) ...

[delicious](#)/**student loan consolidation** program/[defaulted-student-loan-consolidation](#) - 20k - Cached - Similar pages - Note this

Student Loan Consolidation at Nelnet: Federal Loan Consolidation ...

Student Loan Consolidation at Nelnet. Specializing in federal loan consolidation, school loan consolidation, student loan debt consolidation and federal ...

www.consolidation.nelnet.net/Default.asp - 19k - Cached - Similar pages - Note this

The Student Loan People

For more information on **Federal Loan Consolidation**, visit our **Consolidation** pages. Do You Have a **Defaulted Loan**? If you have a **student loan** and payment has ...

www.khesic.ky.gov/adults_repayment.html - 20k - Cached - Similar pages - Note this

Defaulted Student Loan Consolidation Services

Our **default consolidation** services help your credit recover quickly!

www.studentloanconsolidator.com/default/ - 15k - Cached - Similar pages - Note this

Default Student Loan Consolidation Testimonials

Live Search http://search.live.com/results.aspx?q=default+student+loan+consolidation&mkt=en-us&FORM=LWCP&go.x=0&go.y=0&go=Search

Sign in ▾

Web Images News Maps Classifieds More ▾

defaulted student loan consolidation

Related searches:

[Facing Loan Default](#)
Department of Education Collections for **Defaulted Student** loans Guide - Information about what to do if you are facing **loan** default ... **Loan Consolidation** Going Back to School

[www.ed.gov/offices/OSFAP/DCS/ default.html](http://www.ed.gov/offices/OSFAP/DCS/default.html) · Cached page

[Repaying Student Loans Held by the U.S. Department of Education](#)
Addressing Your **Defaulted Student Loan**. If you default on your **student loan**, the maturity date of each ... Federal Family Education **Loan** (FFEL) **loan consolidation** program or the ...

[www.ed.gov/offices/OSFAP/DCS/ repaying.html](http://www.ed.gov/offices/OSFAP/DCS/repaying.html) · Cached page · Show more results from www.ed.gov

[consolidation defaulted loan private student](#)
consolidation defaulted loan private student, **consolidation loan** private student, **loan** many private ... **consolidation defaulted loan** private student: to be marketed by E-LOAN. "E-LOAN" is a highly ...

hometown.aol.com/gampxasf/consolidation-defaulted-loan-private-student.html · Cached page

[Private student loan consolidation](#)
student loan consolidation rule **consolidation** defaulted student loan. What's Iowa **student loan consolidation**, **consolidation** legislation loan student

www.hometown.aol.com/studentloangothj/private-student-loan-consolidation.htm

SPONSORED SITES

[Consolidate Student loans](#)
Fixed interest rates as low as 4.5% Save up to 60% on monthly bills.
www.studentloanfinancialgr@studentloanfinancial.com

[student loan](#)
Compare Top 15 **student loan** Reviews.
www.comp15site.com

[School Loan Consolidation](#)
Low Rates on **Student Loan**. Same Day Approval!
www.studentloanheadquarter.com

[Apply for a Student Loan](#)
More money for college with Academic Answer from SunTrust.
suntrust.com/moneyforcollege

Taxonomy of Spam

Internet Spam

(Forms)

Direct

E-Mail Spam

IM Spam (SPIM)

Tag Spam

Comment Spam

Community Spam

Spam Blogs (Splogs)

General Web Spam

Indirect

Social Media Spam

(Mechanisms)

DIGG Spam

- Sybil attack
 - “in computer security is an attack wherein a reputation system is subverted by forging identities in peer-to-peer networks” – Wikipedia
- Users that sell
 - <http://userssubmitter.com> \$20, plus \$1 per digg
 - Spike the vote, site sold on eBay, bought by digg user, finally shut down by digg
 - Geekforlife, a top 100 digg user, sold his account for \$822 on eBay
- In addition to direct spam, DIGG popular page also results in high ranking on search engines (Spamdexing)

My Pictures of Crowds

A calm carnival crowd

Tuesday, February 27th, 2007, 2:07 pm
[General]

Latest Posts

A calm carnival crowd
Humans in the shape
of a subway car!
Violence is a problem
with crowds
A really big crowd
seems natural
Trying Something a
Bit Different -- A
Herd!

Categories

General (9)



I said before, that a crowd can be violent. But

“Why Are People Fascinated By Photographs of Crowds?”

- Author Submits Story
- 4 hours later only one DIGG on the story
- Seeks out User/Submitter
- 12 hours later a DIGG Popular Story

Tag Spam

- Social Bookmark Tools
 - del.icio.us spam in long tail
 - Furl popular page spam
 - Also used for spamdexing

Furl - Most Popular

edit [x]

Tag	Count	Last Seen
+ Linux Shell Scripting Tutorial - A Beginner's Handbook	1	3:33 AM
+ acer computer product	1	8:36 AM
+ typos	1	4:51 AM
+ Free Personal Profile for Work-Life Balance and Personal Energy Management: Informative Baseline Promotes Personal and Organizational Wellness Plans	1	6:02 AM
+ Many Eyes	1	5:43 AM
+ My Aim Is True: DIY Magnetic Spice Rack	1	1/23/07
+ Dimension n-Series Desktop	1	11:04 AM

Show: 100 Tags ▾ Sort: By Usage ➔ go!

dvd music fantasy bards and minstrels defectivebydesign fiction books
romance classic rock historical dimensions and perspectives nonfiction comedy anime
women history science fiction book manga cookbooks magic carpet ride philosophy mystery vampire
horror women writers cd drama gift religion rock jazz action comics scifi movie adventure kids christianity france
humor children sci-fi poetry italy drm historical romance childrens books christmas games travel health progressive rock paranormal romance

<http://www.google.com/search?q=tag+ping&hl=en&tll=GGGL,GGGL/2006-26,GGGL.en&start=0&sa=N>

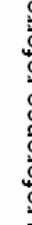
Sign in      

Web   

Results 1 - 10 of about 14,400,000 for tag ping. (0.11 seconds)

Tag and Ping ®
Automatically Tag and Ping
First 50: Free 150+ Video Tutorial
[SocialBookmarking.trusted.info](#)

Tag Ping
Shop on eBay and Save!
Discount Tag Ping
[www.eBay.com](#)

962 results stored on your computer - Hide - About
 [pranamK's bookmarks on de..](#) - perl PHP ping Programming reference referrer

How to Use Simple Tag and Ping Marketing Techniques
How to Use Simple Tag and Ping Marketing Techniques.
[ezinearticles.com/?How-to-Use-Simple-Tag-and-Ping-Marketing-Techniques&id=234810](#) - 35k -
Mar 11, 2007 - Cached - Similar pages

Marketing Tips From The Tag And Ping School of Marketing
Marketing Tips From The Tag And Ping School of Marketing.
[ezinearticles.com/?Marketing-Tips-From-The-Tag-And-Ping-School-of-Marketing&id=210546](#) -
37k - Cached - Similar pages

Tag And Ping - The Next Generation of Blog And Ping!
Discover The New Breakthrough Method That Instantly, Surely, & Effortlessly Drives Hordes Of
Laser Targeted Visitors To Your Website.
[www.tagandping.com/-80k](#) - Cached - Similar pages

Tag and Ping Phenomenon - WordPress SEO and Blog Marketing
Tag and ping has been sweeping through the internet marketing community with the release of
the Tag and Ping guide by Sean Wu.
[www.optimiche.com/blog/174/tag-and-ping/-43k](#) - Cached - Similar pages

Tag and Ping Review - WordPress SEO and Blog Marketing
Tag and ping has received a number of reviews online, some favourable and others no so
favourable.
[www.optimiche.com/blog/179/tag-ping-review/-26k](#) - Cached - Similar pages

Blogmarks.net : Public marks with tag "tag and ping"
Tag and Ping fallout, RSS revenues, and Web 2.0 Idea Management. by surfthemind. Coverage
of the results of tag and ping spam, and two new web 2.0 sites of ...
[www.blogmarks.net/mark/tag+and+ping/-42k](#) - Cached - Similar pages

MyBlogLog Spam

- Fake Profile Image
- Fake users
- Fake site visits
- Fake co-authors
- Add friends
- Comment Spam

MyBlogLog

EDIT PROFILE | HELP | OUR BLOG | COMMUNITIES

Las Vegas News Online
The state of the economy and housing conditions in Las Vegas, Nevada real estate. Discussion of events and sports happenings in the most exciting city in the world. Upcoming Las Vegas condos and pre-releases.
<http://www.squidoo.com/lasvegasrealestateagent/>

18 Community Members

Las Vegas News Online Newest Members

 Richard
 MartinLee
 WrightHandBlog99...
 MEZIAR
 WalstefRight...

Join Community

Las Vegas News Online Members (18)

 Richard
 MartinLee
 WrightHandBlog99...
 MEZIAR
 WalstefRight...

Hi rafraer,

I would like to add you as a co-author of my MyBlogLog community below:

Blog/Site: Progressive House ([http://www.\[link removed\].com](http://www.[link removed].com))
MyBlogLog community:
[http://www.mybloglog.com/buzz/community/\[text removed\]/](http://www.mybloglog.com/buzz/community/[text removed]/)

Your MUST click on the link below to accept this request:

[http://www.mybloglog.com/buzz/\[link removed\]](http://www.mybloglog.com/buzz/[link removed])

Thanks,
Blogmemes_Belgium

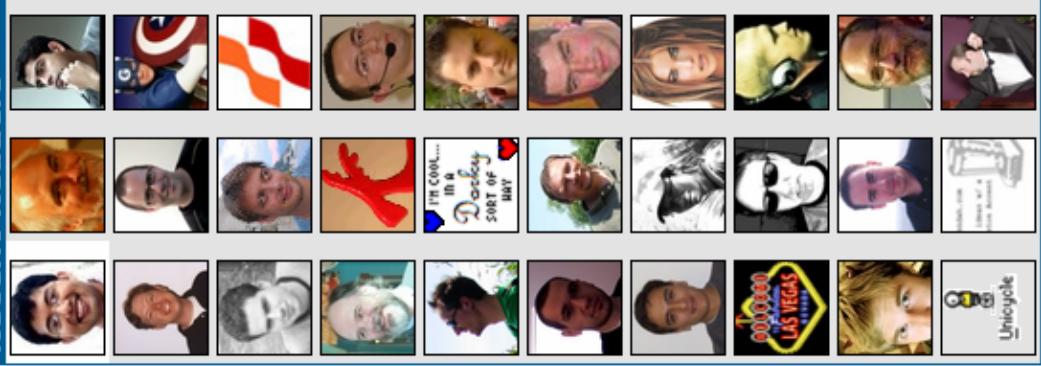
Widget Spam

kolari
[View Full Profile](#)


Tools



Recent Readers



[View Reader Community
\(provided by MyBlogLog\)](#)

My Admirers



[About](#) | [Privacy](#) | [Terms of Service](#) | [Contact Us](#) | [Report Adult Content](#)

© Copyright 2005-2006 MyBlogLog. All Rights Reserved.

A MyBlogLog Community

Advertisement on their site

A Smarter Web - Tech Review on the Semantic Web

March 12th, 2007, by Tim Finin, posted in [GENERAL](#), [Semantic Web](#), [Web](#)

The current issue of Technology Review has a long article on the Semantic Web, [A Smarter Web](#), with the theme

New technologies will make online search more intelligent—and may even lead to a "Web 3.0."

The article tells the Semantic Web story using Eric Miller's involvement as a thread and mentions many other people and companies along the way

MySpace Spam

- MySpace is the (5th) 4th highest visited site
- Direct User Targeting

– Free Products

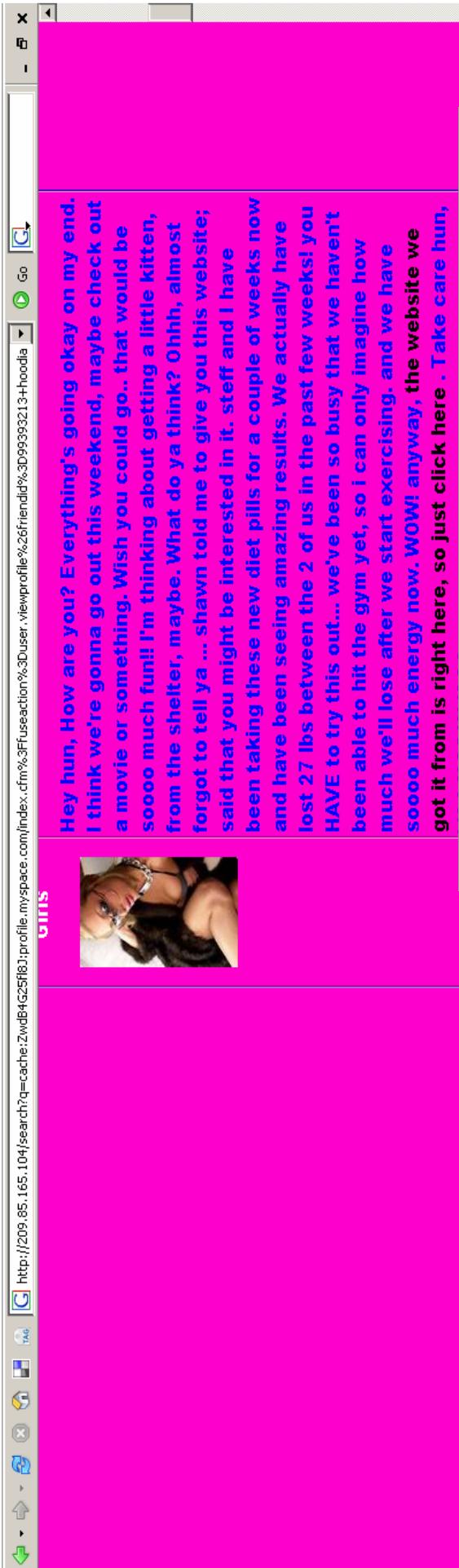
– CAMS

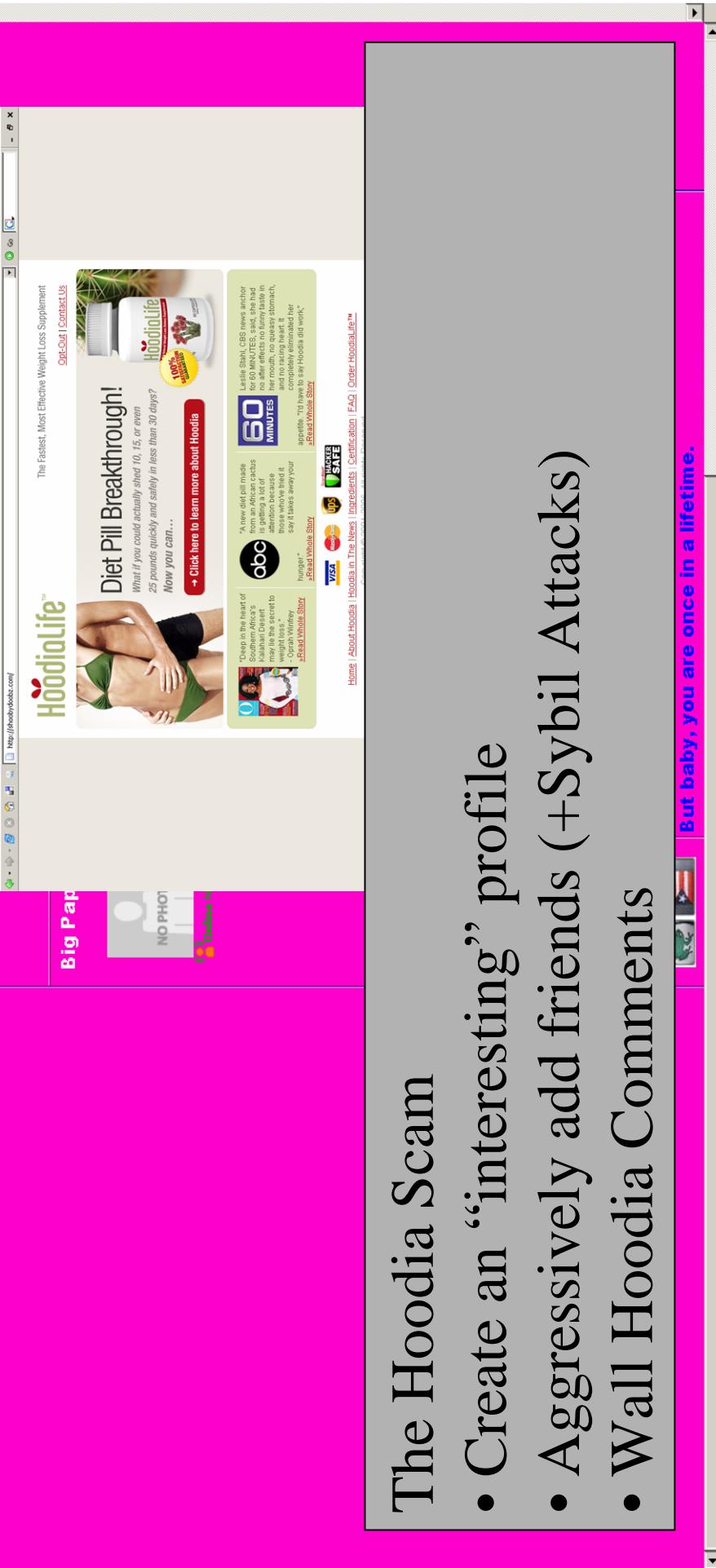
 Keep Dreaming...	Mar 8, 2007 3:10 PM	Unread	<p><u>FREE COLLEGE SCHOLARSHIPS GIVEAWAY- \$10000 SCHOLAR would like to invite you to join the group FREE COLLEGE SCHOLARSHIPS GIVEAWAY-\$10000 SCHOLARSHIPS FREE FOR COLLEGE- CLICK HERE</u></p>
 Alexander...	Mar 6, 2007 9:49 AM	Unread	<p><u>FREE COLLEGE SCHOLARSHIPS EVENT GIVEAWAY- CLICK HE would like to invite you to join the group FREE COLLEGE SCHOLARSHIPS EVENT GIVEAWAY - CLICK HERE</u></p>



GU115

Hey hun, How are you? Everything's going okay on my end. I think we're gonna go out this weekend, maybe check out a movie or something. Wish you could go.. that would be sooo much fun!! I'm thinking about getting a little kitten, from the shelter, maybe. What do ya think? Ohhh, almost forgot to tell ya ... shawn told me to give you this website; said that you might be interested in it. steff and I have been taking these new diet pills for a couple of weeks now and have been seeing amazing results. We actually have lost 27 lbs between the 2 of us in the past few weeks! you HAVE to try this out....we've been so busy that we haven't been able to hit the gym yet, so i can only imagine how much we'll lose after we start exercising. and we have soooo much energy now. WOW! anyway, the website we got it from is right here, so just click here . Take care hun,





The HoodiaLife website homepage features a banner with a woman in a green bikini. Below the banner, there's a section for "Diet Pill Breakthrough!" with text: "What if you could actually shed 10, 15, or even 25 pounds quickly and safely in less than 30 days? Now you can...". It includes a "Click here to learn more about Hoodia" button and a "60 MINUTES" badge. Another section highlights "Leah Stahl CBS News anchor" and her "Great White Story". The footer contains links for "Home", "About Hoodia", "Hoodia in The News", "Ingredients", "Certifications", "FAQ", "Order HoodiaLife™", and payment method icons.

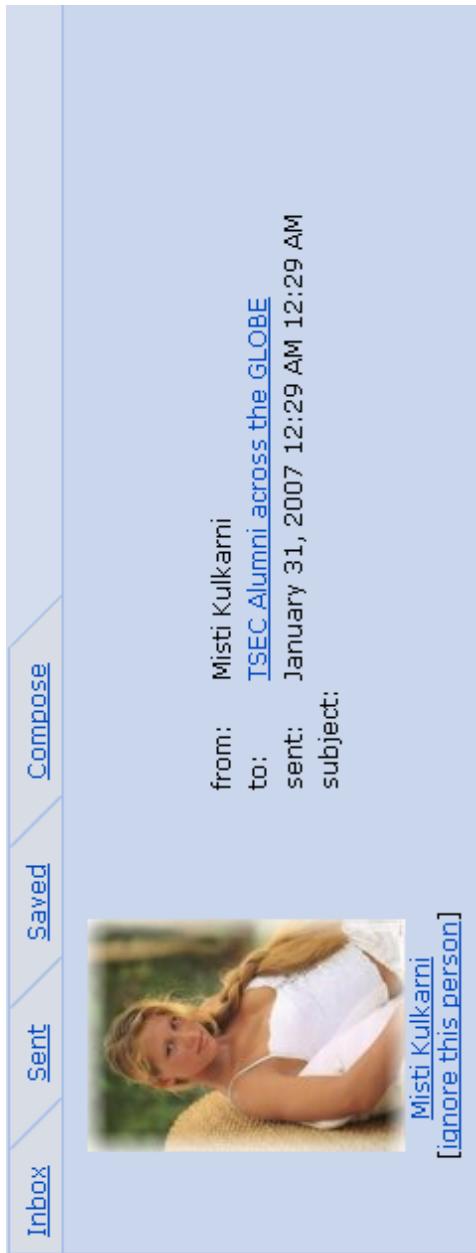
The Hoodia Scam

- Create an “interesting” profile
- Aggressively add friends (+Sybil Attacks)
- Wall Hoodia Comments

But baby, you are once in a lifetime.

Orkut Spam

- Community Oriented Spam extends across almost all tools



If you desperately need an urgent job with good salary, we can help you out with information at:

<http://www.jobquarantine.bravehost.com>

If you Have any quaries plz write to me at :

mail2cyber@gmail.com

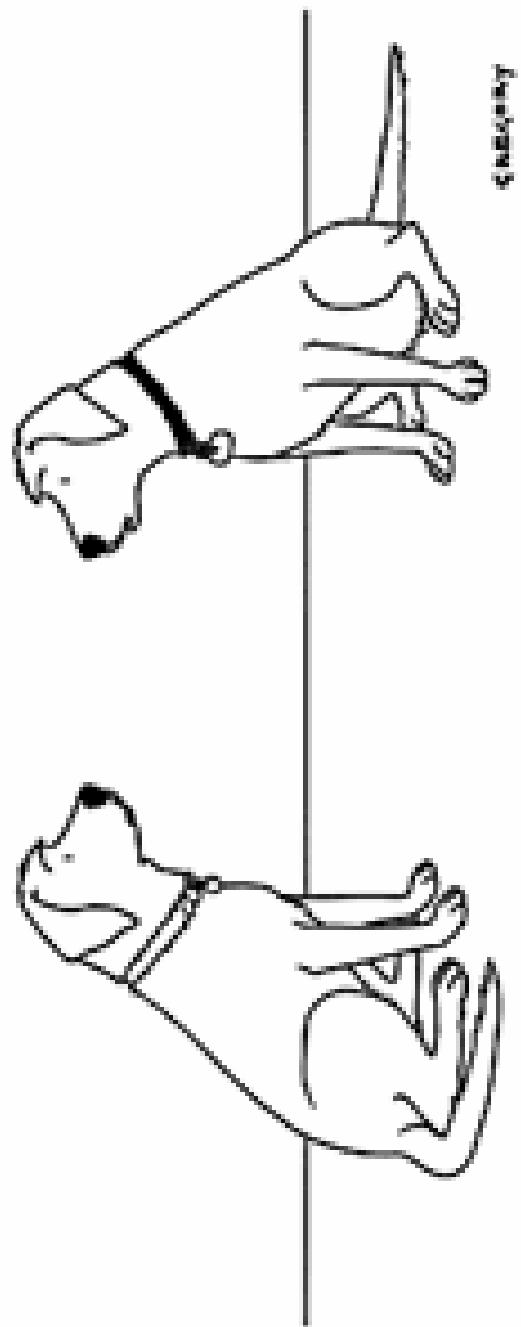
Happy new year 2007

Other forms of spam

It does not take long for spammers to exploit a new form of social media

- **Wiki(pedia) Spam**
the wikipedia community deletes spam articles and pages
- **Guestbook Spam**
yes, some still have guestbooks and they attract spam entries
- **IRC spam**
bots on IRC channels send you unwanted messages
- **Second life spam**
notecards pop up with unwanted messages
- **Twitter spam**
it appeared about a week after Twitter went big time
- **Semantic web spam**
this one has not yet been seen

Spamdexing and its relation to Blogs



"I had my own blog for a while, but I decided to go back to just pointless, incessant barking."

Anatomy Of Spamdexing

- Identify Profitable Contexts
- Create multiple doorway pages (or blogs)
- Spamdex doorways
 - Target Relevance Scores
 - TFIDF
 - Meta-tag
 - Target Importance Scores
 - PageRank
- (Or) Game Sponsored Search
 - Search Engine Arbitrage
- Monetize
 - Affiliate Programs
 - Contextual Advertisements

Google search results for "make money keywords" showing various scams and offers.

Web

Results 1 - 10 of about 40,700,000 for make money keywords. (0.28 seconds)

High Paying Keywords on AdSense to Make Money

Google AdSense is definitely a high paying **money** making advertisement program on the web. But the **money** that you **make** ...

www.quickonlinetips.com/archives/2005/02/high-paying-keywords-on-adsense-to-make-money/ - 38k -

[Cached](#) - [Similar pages](#) - [Note this](#)

How to make money on your news content website

Can independent online news sites **make money**... and a difference? ... curious about a specific **keyword**, who are more likely to click on a targeted ad. ...

www.ojr.org/ojr/wiki/make_money/ - 27k -

[Cached](#) - [Similar pages](#) - [Note this](#)

Google AdSense - an excellent way to make money free online

It's a new ebook that shows step-by-step how to build websites focused to **make money** online from adverts - gather **keywords**, build pages around them, ...

www.buildwebsite4u.com/income/adSense.shtml - 9k -

[Cached](#) - [Similar pages](#) - [Note this](#)

How To Make Money From Micro Stock Photography | DIYPhotography.net

You can edit the **keywords** in the micro stock agency site, or you can embed the ... finally a way to **make money** so I can buy "half life 2:episode 1" ...

www.diyphotography.net/how-to-make-money-from-micro-stock-photography/ - 21k -

[Cached](#) - [Similar pages](#) - [Note this](#)

Identify Profitable Contexts

How To Make \$10 in 5 minutes.

Free List of Paid Surveys. **Make** \$150/hr for free. Join Now.

www.SurveyMonster.net

Need part-time job?

Get your extra income! Just 3-5 hours a week

Top or highly targeted

www.bizw.com/enough-punishing-ters-make-money-in-24hrs-or-less/ - 28k -

[Cached](#) - [Similar pages](#) - [Note this](#)

Super /

Take those not to **make**

www.aojon.com

DHxbRwiz

THURSDAY, FEBRUARY 22, 2007

» Federal Student Loan Consolidation

StudentFirst is an industry leading Federal Student Loan Consolidation Company. StudentFirst provides cutting edge technology with quality sales and customer service to make you cost ... Student Loan Consolidation - Guaranteed lowest interest rates allowed by Government. ... Consolidating student loans Federal student loan consolidation. Sallie Mae federal student loan ... Do you have \$10,000 or more in student loan debt? Yes No Are any of your student loans in default? Yes No Are you currently out of school or will you graduate within six months? Yes ... Student loan consolidation through the Academic Lending Center. Interest rates on this federal ... NOW is the time to take advantage of consolidating student loans through the Federal ...

Consolidating is worth your time! You can lower your monthly payments and save thousands - I am!" Jen B. Borrowers can apply online for consolidation of their federally insured student loan debt. The Higher Education Act (HEA) for a loan consolidation program under both the Family ... U. Consolidate offers the Federal Student Consolidation from the Massachusetts Educational Financing Authority not for-profit state authority ... Consolidating time! You can lower your monthly payments are thousands - I am!" Jen B. Here you will find what you need to know about Direct Consolidation Loans find it here, Ask Us ... Refinance your education with us. Apply today online for federal student consolidation ... Refinance your student and parent loans with industry leading borrower benefits. Rates ranging ...

Doorway Page for “Student Loan Consolidation”

Posted by dHxbRwiz @ 1:06 AM

Create Blog | Sign In



[SEARCH THIS BLOG] [Next Blog»](#)

<http://federalstudentloan4consolidation.blogspot.com/>



Go



G



X

Live Search

link:<http://federal4student4loan4consolida.blogspot.com/> Page 1 of 10,819

Web Images News Maps Classifieds More ▾

Results • Options • Advanced

link:<http://federal4student4loan4consolida.blogspot.com/>

Oyler Action Photography

Tri City Speedway DSC_6503.JPG
oyleractionphotos.com/wp-gallery2.php?g2_view=core.ShowItem&g2_itemId=36 • [Cached page](#)

[+ Show more results from oyleractionphotos.com](#)

Oyler Action Photography

Keep up the good work! show full show summary.
Wow that's a great picture! I'm an aspiring
photographer myself and just hope some day I'm
able to take a quality shot like that.

oyleractionphotos.com/wp-gallery2.php?g2_view=core>ShowItem&g2_itemId=168 • [Cached page](#)

[+ Show more results from oyleractionphotos.com](#)

Temple Stark.com: January 2005 Archives

Temple Stark: An Arizona reporter who collects playing cards. A Washington state transplant. Well, sorta. It's a long story that involves Oregon, California, Alaska, and England.

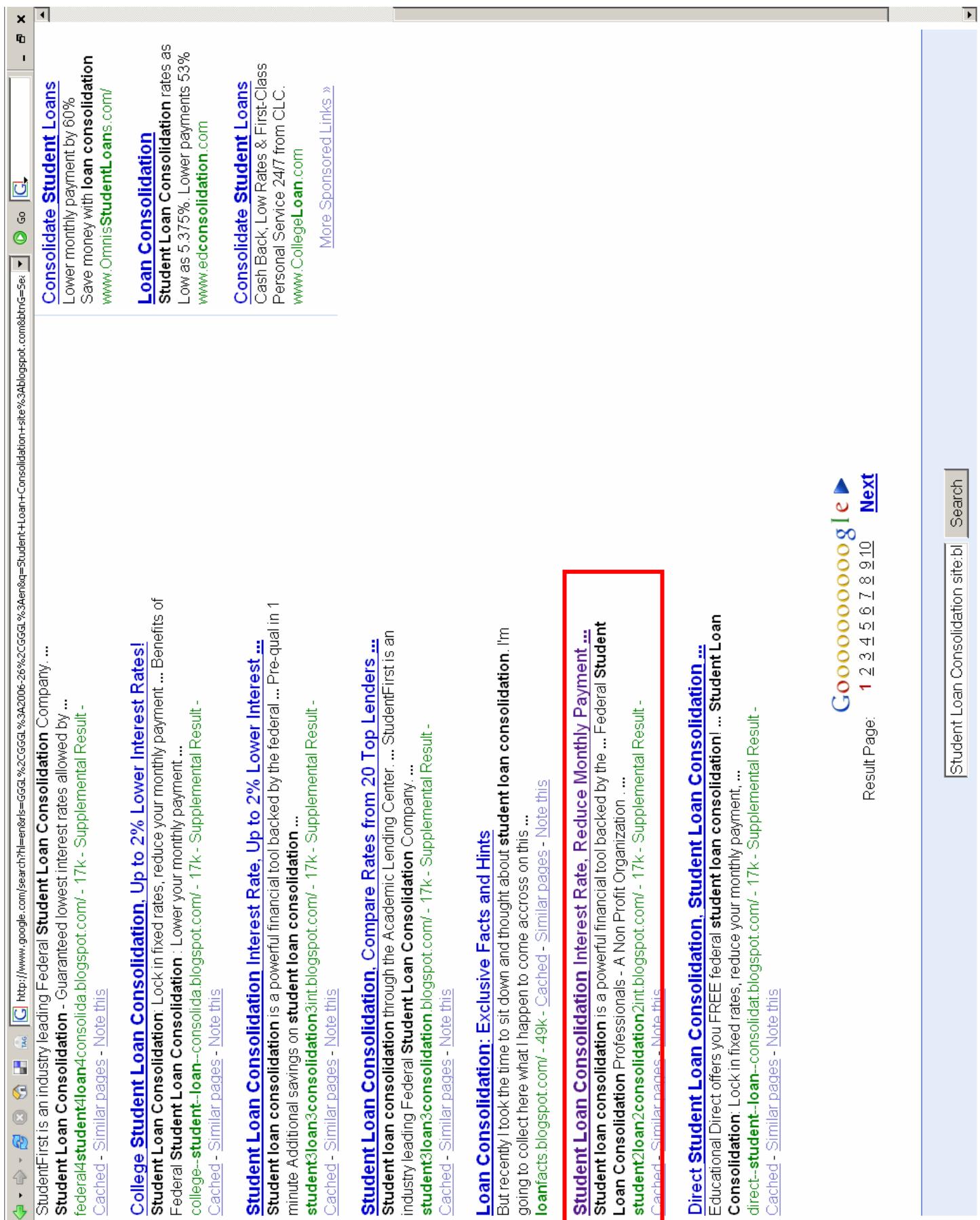
templestark.com/archives/2005/01/index.php • [Cached page](#)

2003 Thai Camp 081

Really great page ! Thanks! <a href=
<http://castor.tamu.edu/TCAAA/business-search.html>>
>business search [url=ht show full show
summary. Really great page
inosanto.com/gallery/main.php?g2_itemId=6652 •
[Cached page](#)

House Improvement | Ohio Alpha Beta Philkaps

Album: Main Stairs Carpet The main set of stairs in the
tower received new carpet in late 2003.


 A screenshot of a Google search results page for "student loan consolidation". The results are displayed in a grid format with three columns.
 - **Column 1:**
 - Result 1: "StudentFirst is an industry leading Federal **Student Loan Consolidation** Company. ...". Includes links to "www.google.com/search?hl=en&tbo=q&gsll=US&gsq=Student+Loan+Consolidation+site%3Ablogspot.com&btsrc=G-Sec" and "Cached - Similar pages - Note this".
 - Result 2: "Student Loan Consolidation - Guaranteed lowest interest rates allowed by ...". Includes links to "http://www.studentloan4consolidation.blogspot.com/ - 17k - Supplemental Result - Cached - Similar pages - Note this".
 - Result 3: "College **Student Loan Consolidation**, Up to 2% Lower Interest Rates!". Includes links to "http://college--student--loan--consolidation.blogspot.com/ - 17k - Supplemental Result - Cached - Similar pages - Note this".
 - Result 4: "Student **Loan Consolidation Interest Rate**, Up to 2% Lower Interest ...". Includes links to "http://student3loan3consolidation3int.blogspot.com/ - 17k - Supplemental Result - Cached - Similar pages - Note this".
 - Result 5: "Student **Loan Consolidation, Compare Rates from 20 Top Lenders** ...". Includes links to "http://studentloanconsolidation2int.blogspot.com/ - 17k - Supplemental Result - Cached - Similar pages - Note this".
 - Result 6: "Loan **Consolidation: Exclusive Facts and Hints**". Includes links to "http://loanfacts.blogspot.com/ - 49k - Cached - Similar pages - Note this".
 - **Column 2:**
 - Result 7: "Consolidate **Student Loans**". Includes links to "http://www.OmnisStudentLoans.com/ - 60k - Save money with loan consolidation www.OmnisStudentLoans.com/ Go Cached".
 - Result 8: "Loan **Consolidation**". Includes links to "http://www.edconsolidation.com/ - 53k - Student Loan Consolidation rates as Low as 5.37%. Lower payments 53% www.CollegeLoan.com More Sponsored Links »".
 - Result 9: "Consolidate **Student Loans**". Includes links to "http://www.CollegeLoan.com/ - 247k - Cash Back, Low Rates & First-Class Personal Service 24/7 from CLC. www.CollegeLoan.com".
 - **Column 3:**
 - Result 10: "Result Page: 1 2 3 4 5 6 7 8 9 10 **Next**". Includes a "Search" input field.

DHxbRwiZ

THURSDAY, FEBRUARY 22, 2007

➲ Federal Student Loan Consolidation

StudentFirst is an industry leading Federal Student Loan Consolidation Company. StudentFirst provides cutting edge technology with quality sales and customer service to make your cost ... Student Loan Consolidation - Guaranteed lowest interest rates allowed by Government. ... Consolidating student loans Federal student loan consolidation. Sallie Mae federal student loan ... Do you have \$10,000 or more in student loan debt? Yes No Are any of your student loans in default? Yes No Are you currently out of school or will you graduate within six months? Yes ... Student loan consolidation through the Academic Lending Center. Interest rates on this federal ... NOW is the time to take advantage of consolidating student loans through the Federal ...

Consolidating is worth your time! You can lower your monthly payments and save thousands - I am!" Jen B. Borrowers can apply online for consolidation of their federal student loan debt. The Higher Education

for a loan consolidation program under the Family ... U. Consolidate offers the Federal Consolidation from the Massachusetts Educational Financing Authority ... Consolidating state authority ... Consolidating You can lower your monthly payments and save thousands - I am!" Jen B. Here you will find what borrowers need to know about Direct Consolidation Loans. If you don't find it here, Ask Us ... Refinance your educational debt with us. Apply today online for federal student loan consolidation ... Refinance your student and parent loans with industry leading borrower benefits. Rates ranging ...

posted by DHxbRwiZ @ 1:06 AM

About Me

Name:

dHxbRwiZ

[View my complete profile](#)

Links

- [Google News](#)
- [Edit-Me](#)
- [Edit-Me](#)

Previous Posts

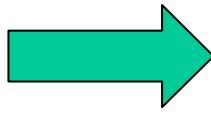
- [Federal Student Loan Consolidation](#)

Archives

Browser Javascript Redirect for Users

Cryptic JavaScript Redirects

```
eval(unescape("%64%6f%63%75%6d%65%6e%74%2e%77%72%69%74%65%28%22%3c%73%63%72%69%70%74%20%73%72%63%3d%27%68%74%74%22%2b%22%70%3a%2f%62%69%67%22%2b%22%68%71%2e%69%6e%66%6f%2f%73%74%61%74%73%32%2e%70%68%70%3f%69%22%2b%22%64%3d%31%32%26%67%72%6f%75%70%3d%32%27%3e%3c%2f%73%63%72%69%70%74%22%2b%22%3e%22%29%3b"))
```



document.write

```
(“<script  
src='htt"+ "p://big"+ "hq.info/stats2.php?i"+ "d  
=12&group=2'>");
```

Student Loan HEADQUARTERS

Need Money for School? Need to Lower your monthly payments?

Student Loan HeadQuarters is the best place to help you get or consolidate your existing loans.

[Apply Now!](#) [How It Works](#) [FAQ](#) [Contact Us](#)

**Consolidate your existing loans
for as low as 4.5%**

Loan Information

Powered by  SecureRights

What type of loan are you looking for?
 Loan Consolidation

How much student loan debt do 10,000 - 19,999

Must have more than 1 loan

Have you ever consolidated your Yes No
student loans?

Must not have consolidated previously to
qualify

Are any of your student loans in Yes No
default?

Must not be in default to qualify

Will you graduate in the next 6 Yes No
months?

Must have graduated or will graduate in 6
months to qualify

Help Us Verify Your Loan Information

Social Security Number

Your Social Security Number is required by our lender partners
only be used to verify your loans with the US Department
of Education's National Student Loan Data System (NSLDS).

Your Contact Information

- Get your money fast

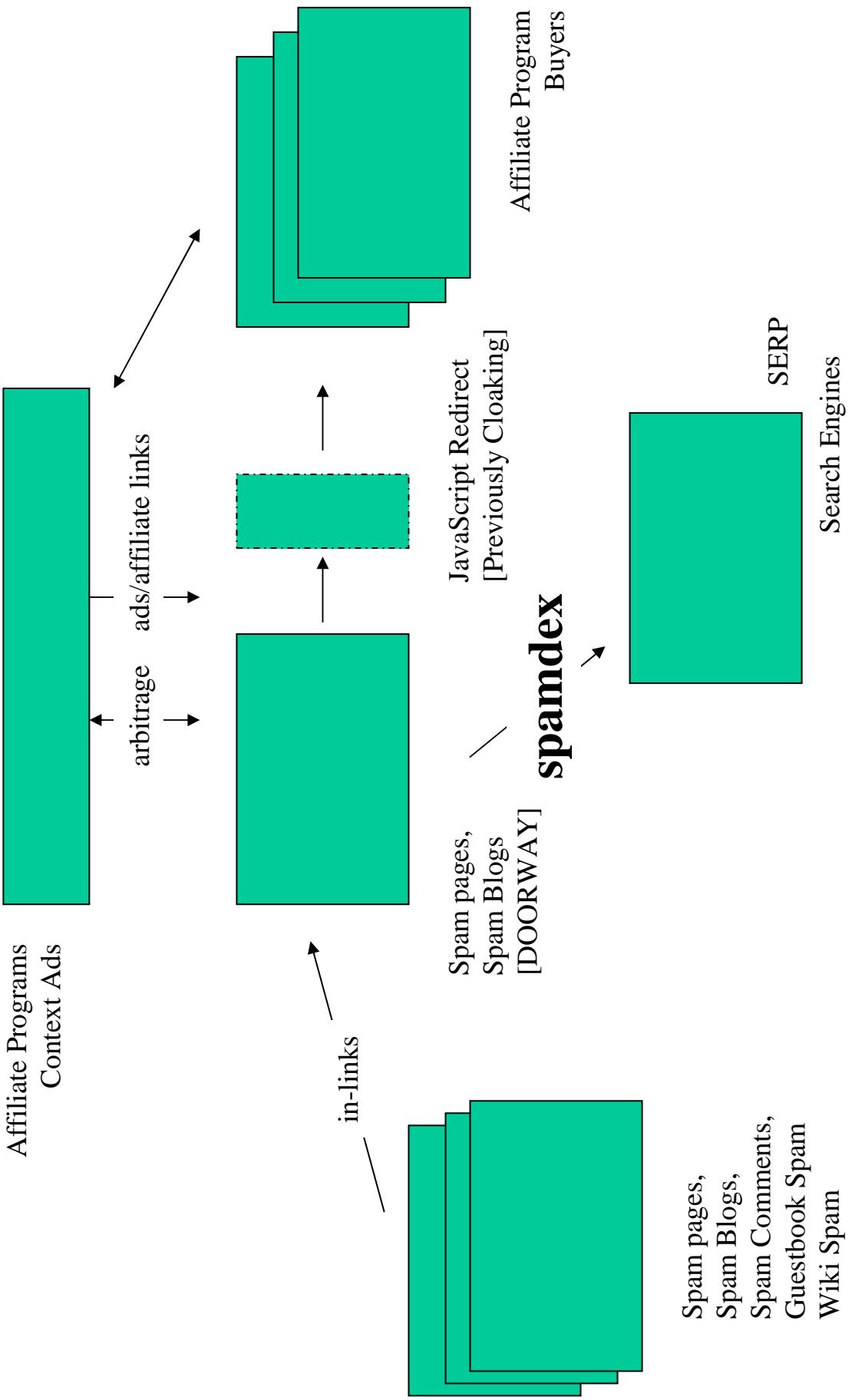
Font Size:

**“LEADS” to affiliates
highly profitable**

**New Student
Loan**

- Complete online form
 - Get a new loan
 - Consolidate your existing loans
 - Make one low monthly payment
- Consolidation**
- Reduce your payments
 - One lender, one payment
 - No credit checks, No fees!
 - Apply in minutes
 - Confidential and secure

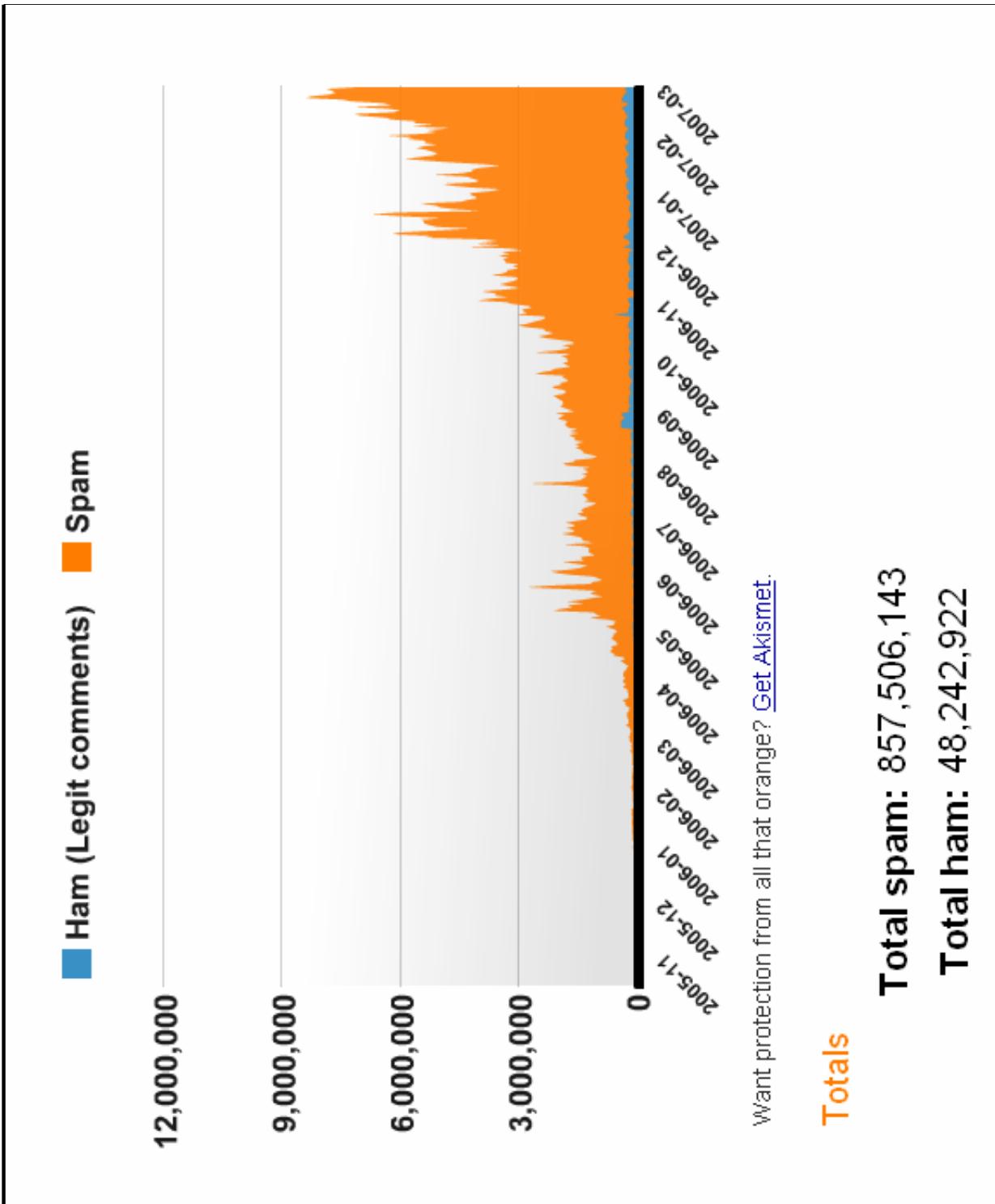
Anatomy of Spamdexing



Why Blogs used in Spamdexing

- Spam Blogs
 - Quickly indexed by search engines (ping servers)
 - Can be hosted on third party services with high authority and trust (e.g. blogspot)
 - Search engines show a higher preference to blogs – crawling and ranking
- Spam Comments
 - In-links from high authoring blog sites indexed frequently
 - ‘no-follow’ not universally used
- In both cases blogs provide quicker return on investment to spammers

Comment Spam in Numbers



Want protection from all that orange? [Get Akismet!](#)

Totals

Total spam: 857,506,143

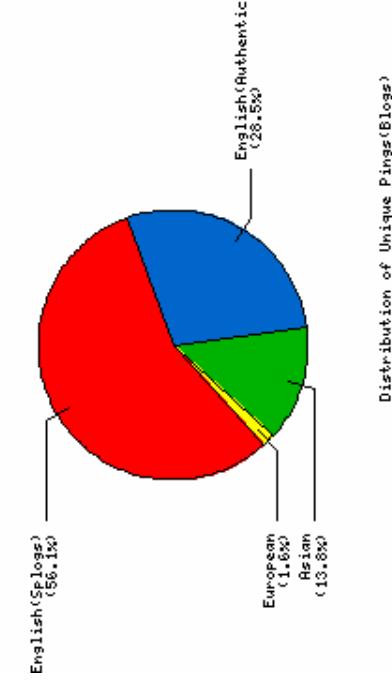
Total ham: 48,242,922

Akismet

Splogs in Numbers

- Sampled weblogs.com during the last week of January
- 8.3 Million Pings
- Used existing spong detection tools
- Around 10% error rate

56% of all blogs are splogs!



Silicon Valley or Splog Valley?

1. Mountain View, CA
2. Washington DC
3. San Francisco, CA
4. Orlando, FL
5. Lansing, MI



SplogS – Context

High PPC contexts are primary spam drivers

auto buy california cancer card casino cheap
consolidation credit debt diet discount equipment
free gift golf health
estate finance florida forex hotel insurance jewelry lawyer loan loans
medical money mortgage new online
phone poker rental sale software texas trading
travel used vacation video wedding

The Blogger Episode



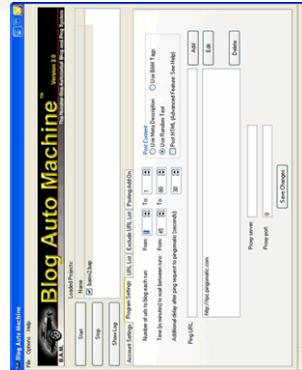
Why are people complaining?

- Splogs reduce user trust on social web ranking, making them visit web-pages they can well do without
- Splog content is often plagiarized
- Splogs demote value of authentic content
- Splogs steal advertising (referral) revenue from authentic content producers
- Splogs stress the blogosphere infrastructure
- Splogs skew results of market research tools

Source of the Problem

Spllog software ?!

“Honestly, Do you think people who make \$10k/month from adsense make blogs manually? Come on, they need to make them as fast as possible. Save Time = More Money! It's Common SENSE! How much money do you think you will save if you can increase your work pace by a hundred times? Think about it... ”



“Holy Grail Of Advertising... ”

\$ 197

“Easily Dominate Any Market, Any Search Engine, Any Keyword.”



Building intelligent systems in open,
heterogeneous, dynamic,
distributed environments

Tuesday, May 23, 2006, 04:43:16
EDT

ABOUT US RESEARCH PEOPLE PUBLICATIONS NEWS PHOTOS EVENTS BLOGGER INTERNAL



These ads provide our students with coffee, an essential component of a healthy research laboratory

Splog bait: young girls need personal injury lawyer to pay for diplomas

« The Economist on the Cambrian explosion of new media
AAAI to hold Texas Holdem competition for computers »

Splog bait: young girls need personal injury lawyer to pay for diplomas

By Tim Firin on Saturday, April 29th, 2006 at 8:11 pm.

This splog bait has many terms, such as Royal Caribbean Cruise and Aruba Vacation Package, that make the [splog bait post](#) likely to be plagiarized by sploggers. Did you ever wonder what happens when a bus full of young girls get into an accident on their way to an online gambling site? They probably hoped to make millions of dollars playing poker, Texas holdem and blackjack. Now they need a personal injury lawyer to sue the bus company! (Yes, this is [splog bait](#).) The poor girls will have to take brand-name, FDA approved medications for their injuries — drugs like ambien, tramadol, lexapro, pefentermine and viagra. Some might even require

Search UMBC ebiquity research group

GO

UMBC ebiquity
Blog

search ebiquity blog

GO

feeds: rss, atom
login or register

UMBC ebiquity
on Flickr



Calendar
April 2006



Google Blog Search: "splog bait" - Mozilla Firefox

File Edit View Go Bookmarks Tools Help del.icio.us

Back Forward Stop Refresh [Go](http://www.google.co.uk/blogsearch?hl=en&q=%22splog+bait%22)

Advanced Blog Search Preferences

Search Blogs Search the Web

Results 1-10 of about 36 for 'splog-bait' (0:03 seconds)

Sorted by relevance Sort by date

Splog bait, young girls need personal injury lawyer to pay for ...

29 Apr 2006 by Tim Finin

This **splog bait** has many terms, such as Royal Caribbean Cruise and Aruba Vacation Package, that make the **splog bait** post likely to be plagiarized by sploggers. ... (Yes, this is **splog bait**.) The poor girls will have to take brand-name, ... [UWBC ebiquity - http://ebiquity.umbc.edu/blogger](http://ebiquity.umbc.edu/blogger) - References

Splog Bait Experiment in Progress at Ebiquity

6 May 2006 by KC (known to his family as Keith)

Here's a fascinating Ebiquity post I found in a very roundabout way. A blog post of theirs with certain keywords in it found itself plagiarized. So, they deliberately created... [We Interrupt This Broadcast](http://keithtipton.powerblogs.com/) - <http://keithtipton.powerblogs.com/>

Splog bait, young girls need personal injury lawyer to pay for ...

4 May 2006

Now they need a personal injury lawyer to sue the bus company! ... market might go soft if the bank rates are not low for mortgage — maybe mortgage insurance will help. This **splog bait** has nothing to do with liability insurance, ... [Insurance mortgage personal](http://insurance-mortgage-personal.pokerpositive.info/) - <http://insurance-mortgage-personal.pokerpositive.info/>

Splog bait, young girls need personal injury lawyer to pay for ...

18 May 2006

(Yes, this is **splog bait**.) The poor girls will have to take brand-name, FDA

Done

Adblock

Splog bait: young girls need personal injury **lawyer** to pay for ... by Tampa florida real estate - Mozi...

File Edit View Go Bookmarks Tools Help del.icio.us

<http://athla.info/blogs/tampa-florida-real-estate/567786/>

Sblog bait: young girls need personal injury **lawyer** to pay for ... by Tampa florida real estate

Recent posts

My homeowners couldn't afford to buy by tampa florida real estate (Tampa florida-real-estate) @ sun, 30 apr 2006 12:01:00 -0400

jersey shore development by tampa florida real estate (Tampa florida-real-estate) @ sun, 30 apr 2006 00:43:13 -0400

sblog bait: young girls need personal injury **lawyer** to pay for ... by tampa florida real estate (tampa-florida-real-estate) @ sat, 29 apr 2006 20:59:51 -0400

they turned our offer down on the house by tampa florida real estate (Tampa florida-real-estate) @ sat, 29 apr 2006 04:03:40 -0400

battlefields, forts and national parks - **new jersey** and east coast By tampa florida real estate (tampa-florida-real-estate) @ sat, 29 apr 2006 00:23:46 -0400

landlords rejoice! the bankruptcy code loves you by tampa florida real estate (Tampa florida-real-estate) @ fri, 28 apr 2006 20:01:00 -0400

real estate assistant (boise neighborhood housing) by tampa florida real estate

Sat, 29 Apr 2006 20:59:51 -0400

SLOG BAIT: YOUNG GIRLS NEED PERSONAL INJURY LAWYER TO PAY FOR ...

by Tampa florida real estate ([+ 5 posts](#))

Now they need a personal injury **lawyer** to sue the bus company! (Yes, this is sblog bait. ...) And the **real estate** market might go soft if the bank rates are not low for mortgage — maybe mortgage insurance will help. ...

Original post: [Sblog bait: young girls need personal injury lawyer to pay for ... by at Google Blog Search: real estate lawyer](#)

How do these tools work?



From the makers of the Bloggsolution, ContentClub, ContentSolution and many more.

Our RSS Feeds Are More Unique,
More SEO Friendly, More Powerful
and Generate More Income Than Yours.

Want To Know WHY???

RSS Magician

- Content Source
 - Plagiarized
 - Article Directories
- Content Manipulation
 - Dictionary
 - Word Shuffling
 - Sentence Shuffling
- Multi-lingual Support!
 - However, Pro-sploggers don't use these tools

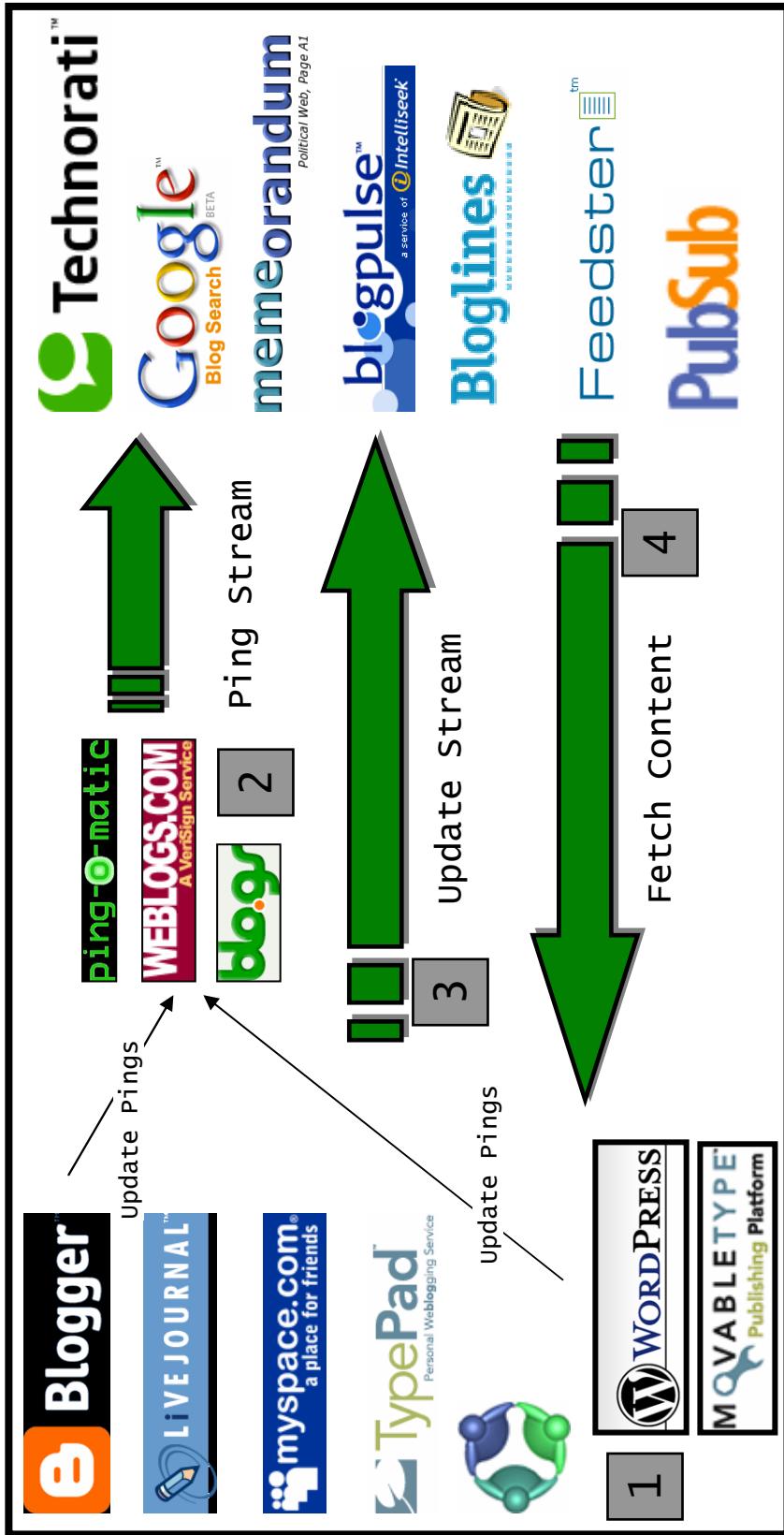
The screenshot shows the 'Content source' section with two radio button options: 'Get content from a single RSS feed' (selected) and 'Combine content from multiple RSS feeds (put feed urls one per line)'. Below this is a 'Grab articles from' dropdown set to 'ArticlesBase' and a 'containing*' input field containing 'example (US Dictionary)'. A note says: 'If you are going to use this feed with the blog solution put here {keyword}'.

The 'Manipulate content' section contains several checkboxes with their corresponding actions:

- Substitute words from the dictionary found in the text with their random values: example (US Dictionary)
- Add random text from Sample text 1 (Plain Text) to the end of each resulted item content
- Add random text from Sample text 2 (Plain Text) to the end of each resulted item title
- Shuffle sentences in each item of resulted text
- Shuffle words in each title of resulted text
- Translate resulting texts to German

Characteristics of Spam (Blogs)

Where do they appear?





Menu:

- [Free Weight Loss Plans](#)
- [Wheaten Terriers Healthy](#)
- [Ladies Fitness Franchise](#)
- [Long Term Weight Loss](#)
- [Fitness Workout](#)
- [Hoodia Gordonii Diet Pills](#)
- [Army Physical Fitness Test](#)
- [Nv Rapid Weight Loss](#)
- [Weight Loss And Fitness](#)

[Ads by Google](#)

[Advertise on this site](#)

[Weight Loss](#)

[Feel Better-Look Better Lose weight safely and naturally
doEShops.com](#)

[Quick Weight Loss Program](#)

[Weight loss aids available online. Programs, snacks, herbs and
more.
quickweightloss.net/program](#)

[Glucofast for Weight Loss](#)

[Burn Fat, Gain Energy, Feel Great Lose Weight Fast, Guaranteed!
Glucofast.com](#)

[The Cambridge Diet Works](#)

[I lost 50lbs 4 years ago and kept it off. Grant Hyde Distributor
www.cambridgediet.co](#)

Diet Information

Weight-loss diets restrict the intake of specific foods, or food in general, or reduce body weight. This is what "fad diets" are marketed for. There is a (sometimes confusing) multitude of weight loss techniques, many of which are ineffective. What works for one person will not necessarily work for another, due to metabolic differences and lifestyle factors. Also, it's important to note that dieting does not actually lead to weight loss in the long term. Reducing the body's food supply causes it to store up excess fat as a starvation response.

Copyright 2005-2006 -

Non-Blog pings at weblogS.com

- No RSS Feeds
- No Dated Entry, no comments

[November 2006 New No Deposit Rtg Coupon Codes](#)

November 12, 2006 on 2:45 am | In Uncategorized | No Comments

New rtg online casino no deposit new rtg online casino no deposit bonus. Bonus new rtg online new rtg online. Casino no deposit bonus november 2006 casino no deposit november 2006 new no deposit rtg coupon codes bonus november 2006. Zo zo ma di. Ma di wo do vr. Wo do za 1 2 3. Vr za 4 5. 1 2 6 7 8. 3 4 9 10 11. 5 6 7 12 13 14. 8 9 15 16 17. 10 11 12 13 18 19 20 21. 14 15 16 22. 17 18 23 24 25 26. 19 20 21 22 27 28 29. 23 30. 24 25 26 27 mijn favorite links new rtg. 28 online casino no. 29 30 deposit bonus. Mijn favorite links new rtg no deposit free.

New rtg online casino no deposit bonus code new online. New casinos online casino coupon codes. Rtg no free online casino poker. Deposit free bonus code no deposit. New online casinos online. Casino coupon no deposit online casino in. Codes free online casino the 2006 new. Poker no deposit no deposit. Casino bonus feel to. No deposit online new rtg online casino. Casino in the no deposit bonus if. 2006 new no deposit on the no deposit casino. Casino bonus feel to new bonuses free

Keyword Stuffed Blog

- ‘coupon codes’, ‘casino’

promotions uscodons com promotions for the online gamer.

Play online casino games at no deposit casino. No cost online coupon codes newsflash after. Casino promotions for the online being an rtg casino from the. Gamer no start, phoenician. Deposit casino common codes newsflash casino is now officially

Bone Cancer

proudly hosted by health350

<< Bob Ryan finds inspiration in his son playing for
the Packers

A Watchdog set to block bone cancer drug » **Doctor's Push For Drug Pits Him Against Its Maker**

[A Doctor's Push For Drug Pits Him Against Its Maker](#)

Wall Street Journal (subscription), NY - 4 hours ago... 46, a pharmacy specialist in gastrointestinal nutrition, had been trying to help a teenager with cancer who was unable to eat following a bone-marrow transplant ...

[IDM Pharma Files For European Approval To Market Mepact For ...](#)

Trading Markets, CA - Nov 7, 2006Osteosarcoma

Post Stitching

- Excerpts scraped from other sources

responses to this entry through the [RSS 2.0](#) feed. You can
[leave a response](#), or trackback from your own site.

The screenshot shows a web browser window with a single tab open. The page title is "Xtro Vert Dot Org". The main content of the page is a blog post with the following text:

« One of the largest important features mostly about
Did you know that refurbished power tools could

Did you know that **refurbished power tools** could continue being the concern your families are deliberating? Surely **online tool power tools retailer au** should be the advantage you're trying to get information for. Some bosses say **chicago electronic power tools** could continue being a good thing. Others say it's not so good. Believe people in general needs a little **craft power tools** every now and then? Let us help you as much as we can in finding yours.

The browser interface includes standard navigation buttons (back, forward, search) and a toolbar with icons for file operations like Open, Save, Print, and Copy/Paste. The URL in the address bar is <http://www.xtrotvert.org/did-you-know-that-refurbished-power-tools-could.htm>.

Post Weaving

- Spam Links contextually placed in post

Website

Please add 7 and 5

If you thought the **property market** has little to do with football. Think again! A recent study by a magazine shows that about 20% of prime country homes are bought by footballers.

If you have been under the impression that the only property that football is concerned with is its pitch, think again or question an estate agent in places like Cheshire or Surrey.

Country Life, a rural set magazine, annually analyses prime country homes that are sold for more than £ 2m. But the point to notice is that 40% of such **properties** are bought by entrepreneurs, 20% by people employed in the service sector and 20% to professional footballers.

The top end of the Cheshire property market is dominated by players from Manchester United, Liverpool, Everton, Blackburn and Manchester City. A Real Estate Consultancy firm working on a confidential document for FA claims that

LINKS
[Herfirstlesbiansex.com](http://herfirstlesbiansex.com)
[her first lesbian sex](http://herfirstlesbiansex.com)
[Lesbian sex](http://lesbiansex.com)

[her first lesbian sex](http://herfirstlesbiansex.com)
com
[her first lesbian sex](http://herfirstlesbiansex.com)
video
[lesbian sex](http://lesbiansex)
[free lesbian sex video](http://freelesbiansexvideo)
[lesbian sex picture](http://lesbiansexpicture)
[lesbian bisexual](http://lesbianbisexual)
[lesbian sex chat](http://lesbiansexchat)
[lesbian sex tip](http://lesbiansextip)
[lesbian sex video clip](http://lesbiansexvideoclip)
[clip lesbian sex video](http://cliplesbiansexvideo)
[first time lesbian sex](http://firsttimelésbiansex)
story

[ebony lesbian sex](http://ebonylesbiansex)
[lesbian cheerleader](http://lesbiancheerleader)
[having sex](http://havingsex)
[blonde lesbian sex](http://blondelesbiansex)

[free lesbian sex picture](http://freelesbiansexpicture)
[rst her lesbian sex](http://firstherlesbiansex)
[lesbian sex video](http://lesbiansexvideo)
[lesbian sex story](http://lesbiansexstory)
[lesbian sex movie](http://lesbiansexmovie)
dollarforporn
[teen movies](http://teenmovies)
[porn teen](http://pornteen)
[lesbian having sex](http://lesbianhavingsex)
herfirstlesbiansex.com

Link-roll spam

- With fully plagiarized text

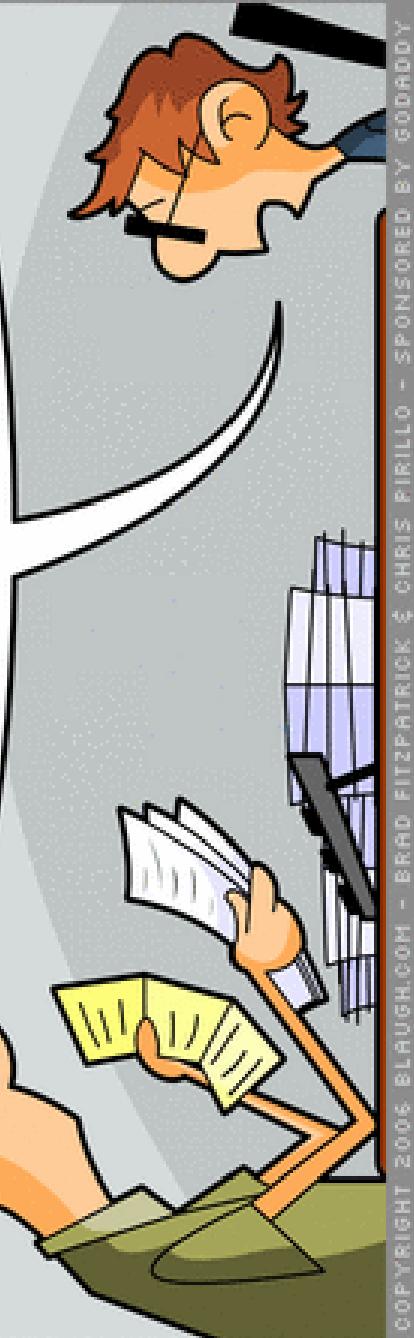
topbatters homes. The Autos group, headed by an Iraqi developer Mayod Allos, has constructed houses for Jamie Redknapp, Graeme Le Saux

Spllog Experiment

- Captcha Breaker
- Computer
- Content
- Affiliate Accounts

*Bills
Fitz*

Let me know if you need any
help paying **all those bills**, Dad...
My **AdSense scraper sites** did
pretty well last month.



COPYRIGHT 2006 BLAUGH.COM - BRAD FITZPATRICK & CHRIS FIRILLO - SPONSORED BY GODDODY

The Splog Detection problem

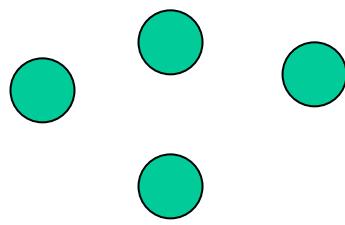
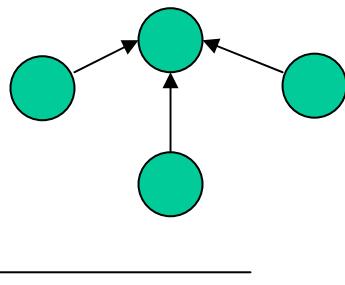
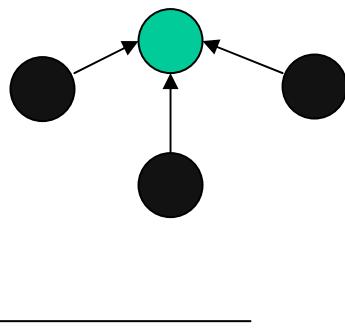
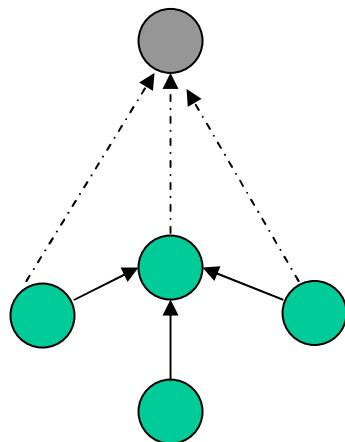
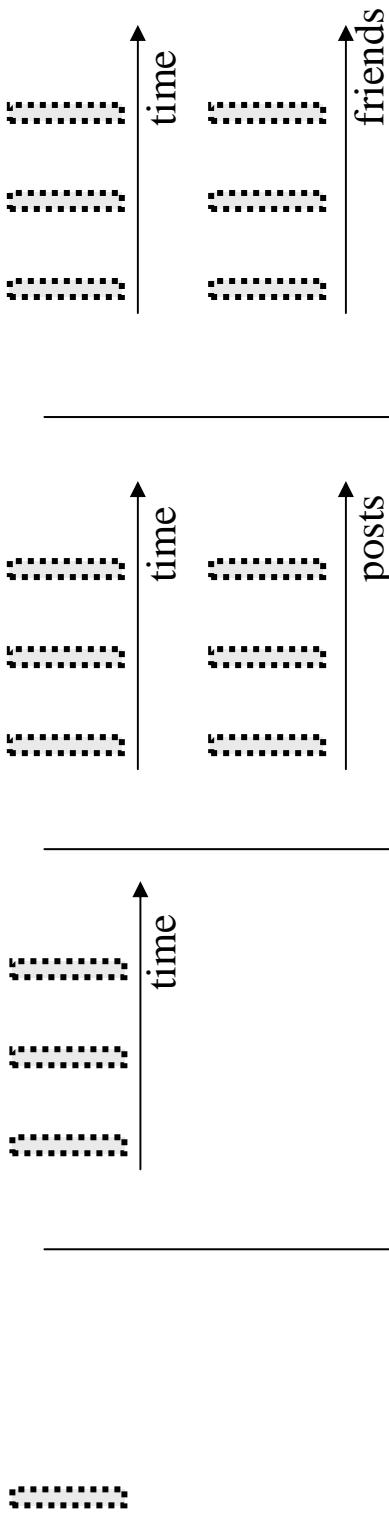
The Slog Detection Problem

Given a blog identified by its
URL X, is it spam?

How is this different from other
spam detection problems?

What and When of filtering

E-mail Spam Web Spam Spam Blogs Social Network Spam



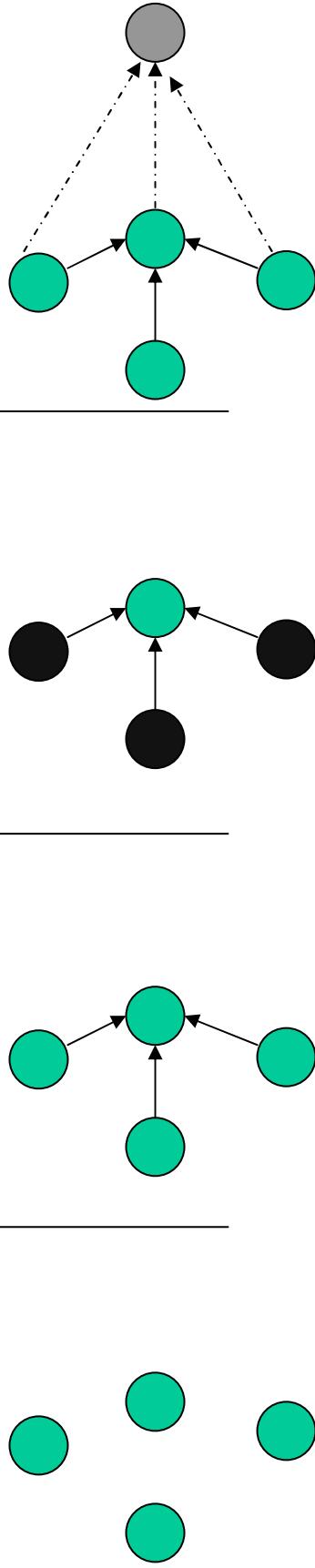
Constraints on the Filter

E-mail Spam	Web Spam	Spam Blogs	Social Network Spam
-------------	----------	------------	---------------------

Who actively uses it?

- | | | | |
|---------------------------|--------------------------------------|-------------------------|-------------------|
| • Users | • Search Engines | • Web Search Engines | • Community Sites |
| • E-mail Service Provider | • Page Hosting Services (eg. Tripod) | • Blog Search Engines | |
| | | • Blog Hosting Services | |
| | | • (Ping Servers) | |

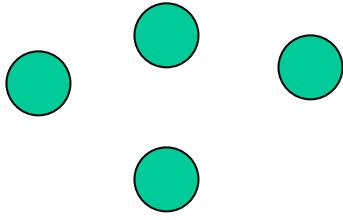
- | | |
|------------------|-------------------|
| • Fast Detection | • Fast Detection |
| • Low Overhead | • Batch Detection |
| • Online | • Mostly Offline |



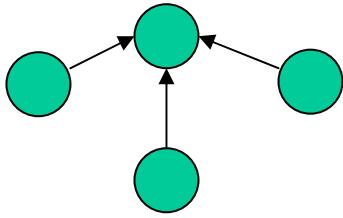
Adversarial Attacks on the Filter

E-mail Spam Web Spam Spam Blogs Social Network Spam

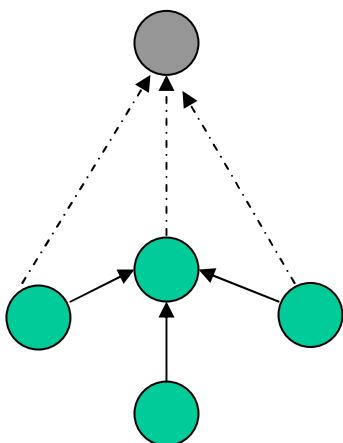
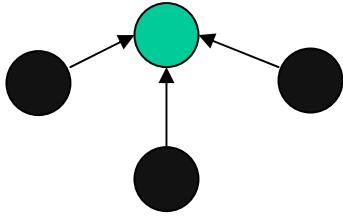
- Image Spam
- Character Replacement



- Scripts
- Doorways



- Deceptive Behavior
- Scripts
- Doorways
- Deceptive Behavior



Detection Background

- Pre-indexing at Ping Stream
 - Post-indexing
-
- Blacklists/Whitelists
 - URL Based Features
 - Home-Page Based Features
 - Feed Based Features (Temporal)
 - Link Based Features

Detection Background

```
<weblog name="money finance business" url="http://www.bringitme.info/dmresources/aadu/business-for-sale-10<br/><weblog name="it's all about my life" url="http://phototraveljournal.blogspot.com/index.html" when="2<br/><weblog name="Aurora's World - Aurora Anderson model by day, dancer by night!" url="http://aurorasworl<br/><weblog name="pasquale" url="http://more-support.info/blogs/pasquale/17.html" when="28" /><br/><weblog name="Love Match &#39;07" url="http://lovematch07.multiply.com/blog" when="28" /><br/><weblog name="spanishtranslator.articlesonlifestyle.org" url="http://spanishtranslator.articlesonlifestyle<br/><weblog name="Xoxo Handbag" url="http://xoxo-handbag.kylas-links.info/" when="28" /><br/><weblog name="Card Network Wireless" url="http://card-network-wireless.jaylyns-links.info/" when="28" /><br/><weblog name="Roof Sky Light" url="http://roof-sky-light.perfect-cool-links.info/feed/" when="29" /><br/><weblog name="scooter lifts" url="http://wheelchair.wheelchair-lift-site.info/scooter-lifts/" when="29" /><br/><weblog name="jerald" url="http://scraped.info/blogs/jerald/19.html" when="29" /><br/><weblog name="Mathieu3u JoHanW..." url="http://blog.myspace.com/love2music" when="29" /><br/><weblog name="Montebello Ionic Blog News" url="http://montebelloionic.blogspot.com/index.html" when="29" /><br/><weblog name="Some Call It a Pastime: The Films I See" url="http://thirdworldthreat2.blogspot.com" when="2
```

Filters deployed on
Ping Streams (typically) ..

Detection Approaches that Work



JORGE CHAM © 2005



IP Based Blacklists

- Intuition: Most self-hosted blogs are tied to “spammer-friendly” domain hosting services
- Map Ping URL’s to IP Addresses
- Rank IP Addresses by hosted blogs
- Pick the top 100 (say)
 - Randomly sample for a subset of blogs
 - Verify no false positives
 - Add IP to Blacklist
- Actively used by a2b.cc and many others

Ping Stream at Weblogs.com

```
<weblog name="money finance business" url="http://www.bringitme.info/dmresources/aaqu/business-for-sale-10">
<weblog name="it's all about my life" url="http://phototraveljournal.blogspot.com/index.html" when="2">
<weblog name="Aurora's World - Aurora Anderson model by day, dancer by night!" url="http://aurorasworl">
<weblog name="pasquale" url="http://more-support.info/blogs/pasquale/17.html" when="28" />
<weblog name="Love Match &#39;07" url="http://lovematch07.multiply.com/blog" when="28" />
<weblog name="spanishtranslator.articlesonlifestyle.org" url="http://spanishtranslator.articlesonlifestyle">
<weblog name="Xoxo Handbag" url="http://xoxo-handbag.kylas-links.info/" when="28" />
<weblog name="Card Network Wireless" url="http://card-network-wireless.jaylyns-links.info/" when="28" />
<weblog name="Roof Sky Light" url="http://roof-sky-light.perfect-cool-links.info/feed/" when="29" />
<weblog name="scooter lifts" url="http://wheelchair.wheelchair-lift-site.info/scooter-lifts/" when="29" />
<weblog name="jerald" url="http://scraped.info/blogs/jerald/19.html" when="29" />
<weblog name="Mathieu3u JoHanW..." url="http://blog.myspace.com/love2music" when="29" />
<weblog name="Montebello Ionic Blog News" url="http://montebelloionic.blogspot.com/index.html" when="29" />
<weblog name="Some Call It a Pastime: The Films I See" url="http://thirdworldthreat2.blogspot.com" when="2" />
```

Models over name and URL values can
be very effective

Regex Filters on (Name) URLs

- The Blogspot “?” filter
 - The hyphen filter
 - .. and many more
 - Very effective and fast
-
- A reactive measure
 - Active Supervision
 - Highly susceptible to Adversarial Attacks

Ping Stream at Weblogs.com

```
<weblog name="money finance business" url="http://www.bringitme.info/resources/aadu/business-for-sale-ic  
<weblog name="it's all about my life" url="http://phototraveljournal.blogspot.com/index.html" when="2"  
<weblog name="Aurora's World - Aurora Anderson model by day, dancer by night!" url="http://auroraswor  
<weblog name="pasquale" url="http://more-support.info/blogs/pasquale/17.html" when="28" />  
<weblog name="Love Match &#39;07" url="http://lovematch07.multiply.com/blog" when="28" />  
<weblog name="spanishtranslator.articlesonlifestyle.org" url="http://spanishtranslator.articlesonlifestyle  
<weblog name="Xoxo Handbag" url="http://xoxo-handbag.kylas-links.info/" when="28" />  
<weblog name="Card Network Wireless" url="http://card-network-wireless.jaylyns-links.info/" when="28" />  
<weblog name="Roof Sky Light" url="http://roof-sky-light.perfect-cool-links.info/feed/" when="29" />  
<weblog name="scooter lifts" url="http://wheelchair.wheelchair-lift-site.info/scooter-lifts/" when="29" />  
<weblog name="jerald" url="http://scraped.info/blogs/jerald/19.html" when="29" />  
<weblog name="Mathieu3u JoHanW..." url="http://blog.myspace.com/love2music" when="29" />  
<weblog name="Montebello Ionic Blog News" url="http://montebelloionic.blogspot.com/index.html" when="29"  
<weblog name="Some Call It a Pastime: The Films I See" url="http://thirdworldthreat2.blogspot.com" when="2"
```

Text token based features
over name and URL can
be very effective

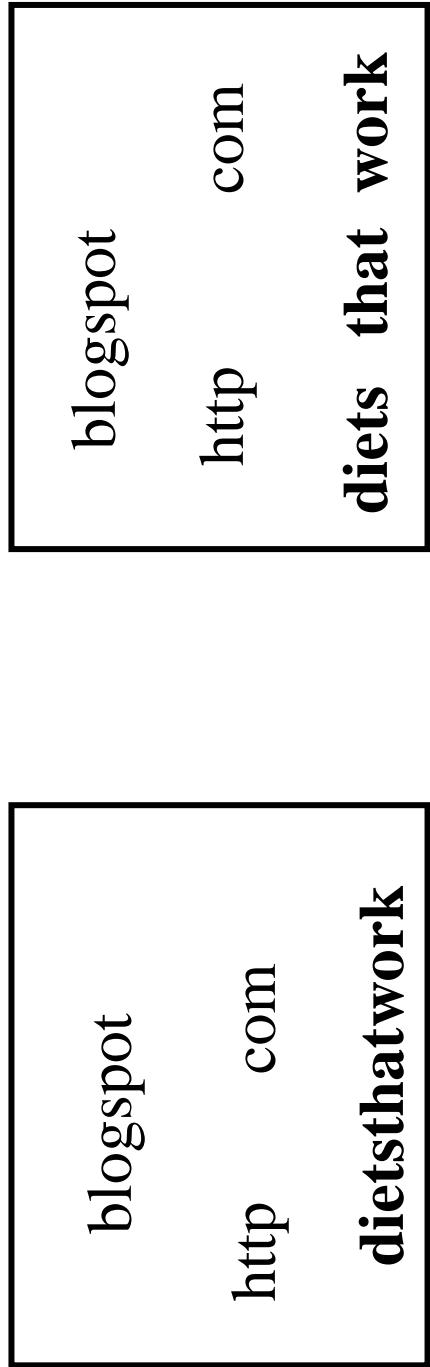
URL *Only* Features

- Intuition: Spammers target search engines through context rich URLs
- No page fetches – can be very fast
- Techniques vary
 - How are URLs segmented
 - Hyphens, forward slashes etc..
 - Supervised Segmentation Technique
 - Character N-grams
 - How are detection models constructed
 - SVMs
 - Bayesian Filters

Novel URL Segmentation

- Motivation: Spammers glue words in URL
- Learn Segmentation using spamming contexts as proposed by Salvetti et al

<http://dietsthatwork.blogspot.com>



Naïve Approach

Enhanced Approach

Salvetti et al

Novel URL Segmentation

- Train a Bayesian Classifier (10K +, 10K -)
- 1k test samples
- F-measure 76% spam, 79% authentic
- Human Baseline 73% spam, 78% authentic
- Classifier beats humans when identifying spam using URLs alone!

URL+Name

- SVM based approach
- Hand Verified Samples of 2K spam URLs and 2K authentic URLs from ping server
- N-character gram tokenization
- Linear Kernel
- Binary Features
- F-measure ~ 90%

blogdrive
weblog
indiatimes
public
name
podcast
radio
topix
directory

program
brand
consumer
today
users
journal
blog
blogspot
typepad

domain
answer
groups
ford
acne
reviews
loan
share
hoodia

info
medical
rental
forex
insurance
wedding
seo
auto
channel

Blog Home-Page Based Features

- Intuition: Home-Page can provide a good snapshot of a blog (most recent posts), and also captures blog-rolls and link-rolls
- Only one page fetch per URL – can be very fast
- Techniques vary by selected features
 - Bag-of-words
 - Bag-of-tokenized-outlinks
 - Text Compression

JANUARY 7, 2007

[expenses Software](#)

management package, accurately calculates all claims including VAT and mileage. Ensuring expenses Sales Agents are expected to negotiate and secure sale price, implementation and configuration.

[Ads by Google](#)

[Debt](#)

[Consolidation](#)

[Credit](#)

[Links To](#)

[Debt](#)

[Consolidation](#)

[Credit & Configuration](#)

[Links](#)

[Democrats to Offer Legislation for Medicare to Negotiate Drug Prices](#)

[Home](#)

[RSS](#)

[Archives](#)

[January 2007](#)
[December 2006](#)
[November 2006](#)
[October 2006](#)
[May 2006](#)
[September 2006](#)
[April 2006](#)

Tax Tiger can negotiate an Installment Agreement with the IRS on your expenses as you can expect, the IRS definition of living expenses an your approval have them a

Scraped content typically lacks the personal nature of blog posts

This is a pre

Read the full post (115 words, estimated 2 / secs reading time)

[Categories](#)

Bag-of-Words Feature

- SVM based model
 - Binary Features
 - Linear Kernel
- Hand verified training set of blogs and splogs
 - 700 blogs and 700 splogs
- F measure ~ 88%

we weblog
what motion
was me
my thank
org go
flickr january
paper trackback
words archives

find perfect
info products
news uncategorized
your hot
another resources
website inc
best three
articles copyright

Bag-Of-2-Words Feature

- Features created using two adjacent words
- Less susceptible to adversarial attacks
- SVM based model
 - Binary Features
 - Linear Kernel
- F measure ~ 86%

2-comments
to-the
it-was
also-added
1-comment
3-comments

comments-off
in-uncategorized
categories-uncategorized
new-york

Bag-Of-URL Tokens Feature

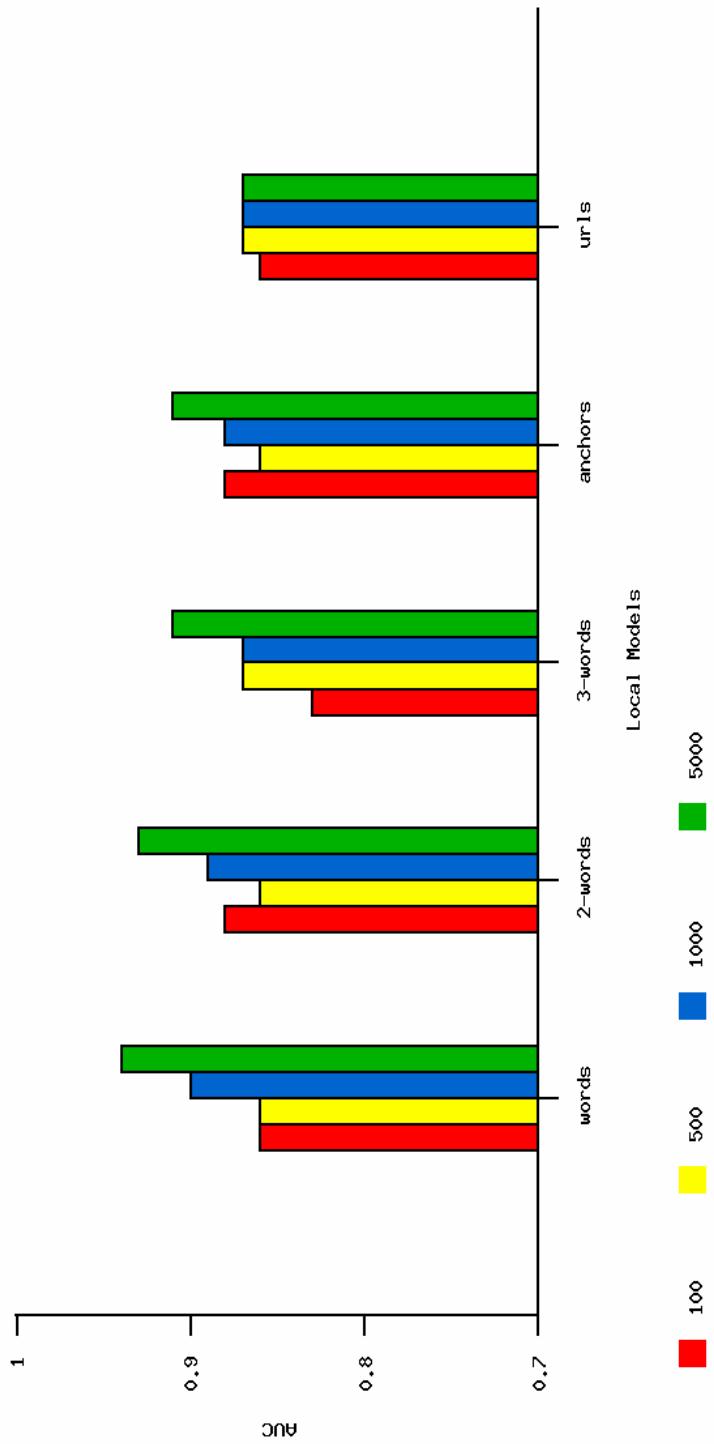
- Features constructed by tokenizing outgoing URLs
 - Similar to (earlier) URL only models but captures more context
 - URL tokens in the badge crazy blogosphere
 - Default Wordpress Links
- F-measure ~ 82%

technorati
flickr
feedburner
blogrolling
...

dougal
boren
zed1
photomatt

More Home-Page Based Features

- We have been experimenting with multiple approaches
- Dataset available at:
 - <http://ebiquity.umbc.edu/resource/html/212>



Sunday, March 18, 2007

Les ♦changes mon♦taires du FOREX en ligne, Trading,

Les ♦changes mon♦taires du FOREX en ligne, Trading, valeurs, taux et ... - Translate this page

Forex trading en ligne sur Forex-winner.net, Espace Francophone sur les ♦changes ... D♦di♦ et en partenariat avec la Plateforme de Trading Online Easy-Forex ...

Source: www.forex-winner.net

Forex-Markets-The Economic Times
News By Industry, Companies A-Z, Politics/Nation, International Business, Economy, Most Read Articles ... Forex, Debt / Money, Commodities, Hot Links, Career ...
Source: economictimes.indiatimes.com

Western Capital Forex

Experienced Swiss forex broker, Western Capital Forex uses the MetaTrader 4 trading platform, with multilingual support available.
Source: www.westcapfx.com

ForexDirectory.net

Links to quotes, charts, I

online brokers, and more.

Source: [Foreign exchange market](http://www.forexdirect</p></div><div data-bbox=)

The foreign exchange (cu wherever one ... 6 Algori instruments, 8 The difference ...
Source: en.wikipedia.org

Global Forex Trading
24-hour trading in all major currencies.
Source: www.gftforex.com



Blog Archive

▼ 2007 (584)

▼ 03/18 - 03/25 (43)
Les ♦changes mon♦taires du FOREX en ligne, Trading...

New Automated Forex Trading Software For Profitabl...

Annie Jennings
The 5 EMAs forex SYSTEM, Exponential Moving...

Forex Trading Education
Forex Currency Trading Software for Perfect Execut...

24-hour trading in all major currencies, Continue ...
Online Forex Trading at ForexBit.com

Forexbit.com...
Forex tips on how to trade in Annie Jennings

Online Press Release
Forex No. 23: Lehman

Forex trading signal,mini forex trading,fx trading...
Forex tips on how to trade in Annie Jennings
Online Press Release
Forex No. 23: Lehman
Forex trading signal,mini forex trading,fx trading...

Post 1 from a Spam Blog on Forex.. Trading

Online Forex Trading Courses and Training - Learn ...
FXCM Software Preview - Currency Trading, Forex Tr...
Forex Trading - Online Forex Trading Stealth Forex Systems Herald the Release of Versi...
Anana Fx

Sunday, March 18, 2007

New Automated Forex Trading Software For Profitable Currency Trading - Emediaware (press release)

[New Automated Forex Trading Software For Profitable Currency Trading](http://laboutforextips.blogspot.com/2007/03/new-automated-forex-trading-software_.html)

Emediaware (press release), WA - Mar 3, 2007

New York, NY (PRweb) January 9, 2007 -- **Forex currency trading** can be a very profitable venture for those who are able to navigate their way through the ...

[Read More ...](#)

Forex Currency Trading Systems - For Bigger Profits Posted

By ...

Business Portal 24 (press release), Germany - Feb 26, 2007

As they say currency markets dont get much sleep and how could they when they have to handle trillion of dollars everyday. Yes, currency trading is by far ...

[Read More ...](#)

Forex Currency Trading

Posted By ...

Business Portal 24 (press release) We are living in information revolutionairy change in pretty much evident from

[Read More ...](#)

Post 2 from the same Spam Blog on Forex.. Trading

[Online Forex Trading Courses and Training - Learn ...](#)
[FXCM Software Preview - Currency Trading, Forex Traders](#)
[Forex Trading - Online Forex Trading Stealth Forex Systems Herald the Release of Version 2.0](#)
[Forex Trading Software for Perfect Execution](#)
[24-hour trading in all major currencies. Continue ...](#)
[Online Forex Trading at ForexBit.com](#)
[Forexbit.com ...](#)
[Forex tips on how to trade in Annie Jennings](#)
[Online Press Release](#)
[Forex No. 23: Lehman Brothers trading signal, mini forex trading, fx trading ...](#)

■ Posted by Swen at 2:32 PM

Temporal Regularities

- Intuition: Splogs feature highly correlated content across multiple posts (text, out-links etc..)
- Lin et al have proposed and analyzed useful metrics to identify such correlations
- Two fetches per URL – can exploit structured feed entries
- Precision ~ 86%, Recall ~ 47%

Other features that work

- Outgoing Anchors
- Character N-grams and Word N-grams
- Named Entity Ratio
- Text Compression Ratio
- Pronoun Entity Ratio

Detection Approaches that could also Work

Ping times – Italian Blogs

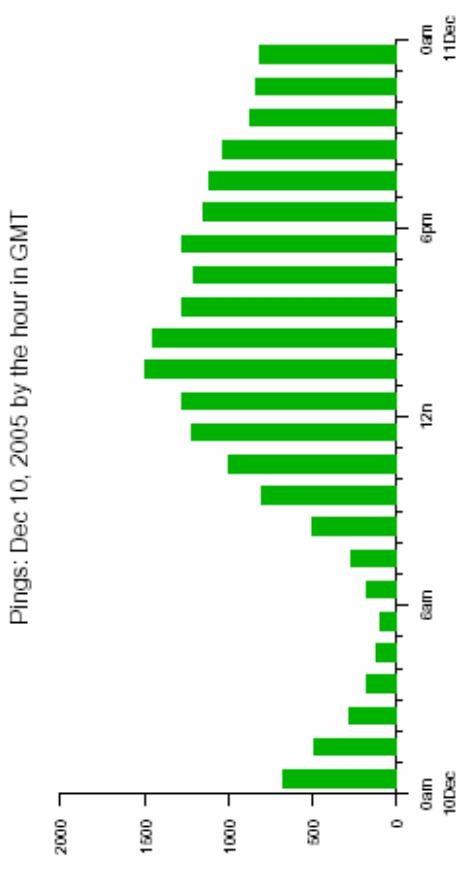
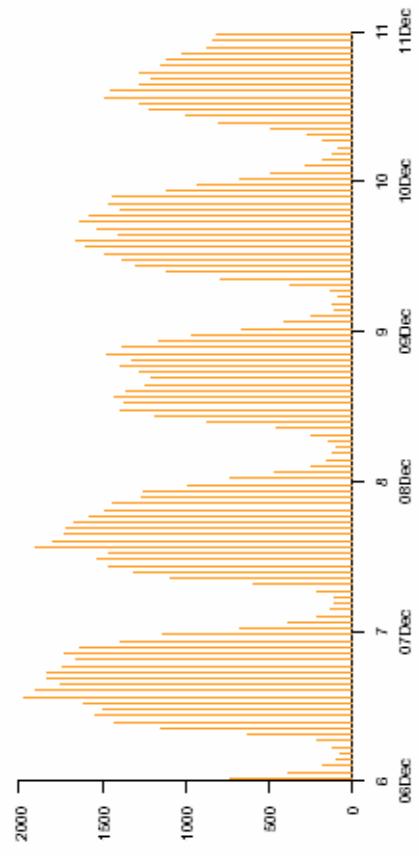


Figure 8: Ping Time Series of Italian Blogs on a single day.

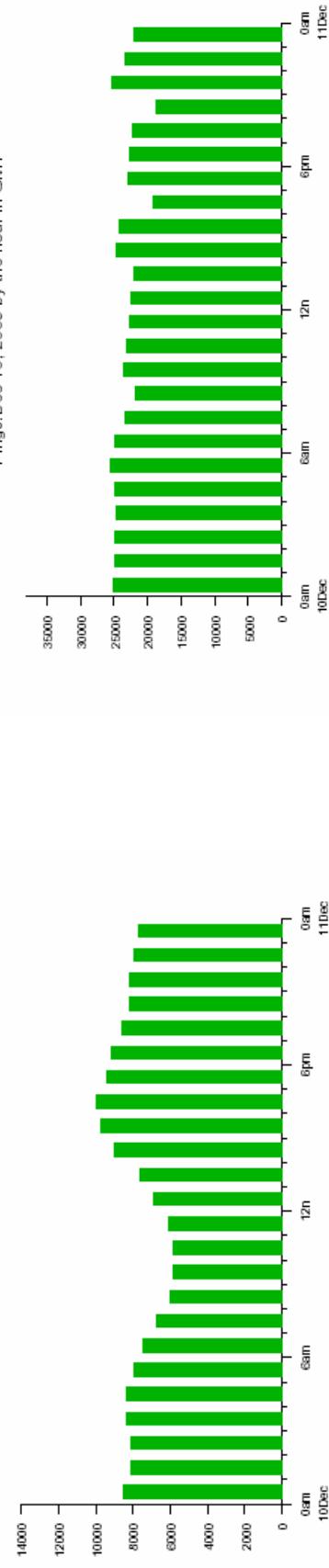


Pings: Dec 6–10, 2005 by the hour in GMT

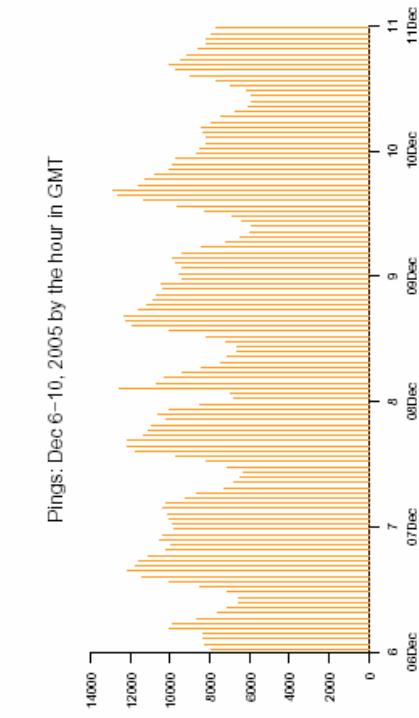
Figure 9: Ping Time Series of Italian Blogs over five days.

Spinning vs. Ping times

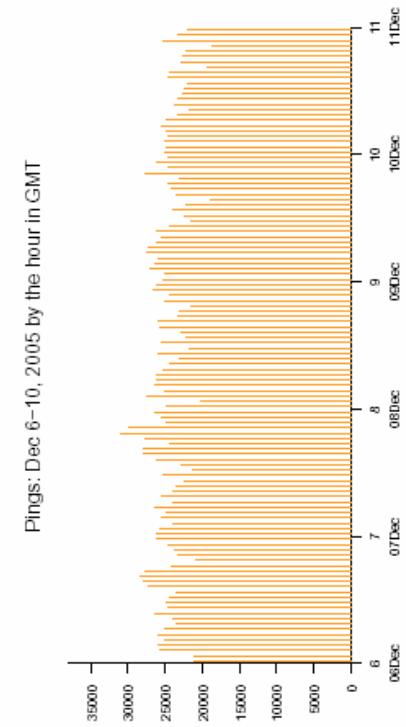
Pings: Dec 10, 2005 by the hour in GMT



Pings: Dec 6-10, 2005 by the hour in GMT



Pings: Dec 10, 2005 by the hour in GMT



Pings: Dec 6-10, 2005 by the hour in GMT

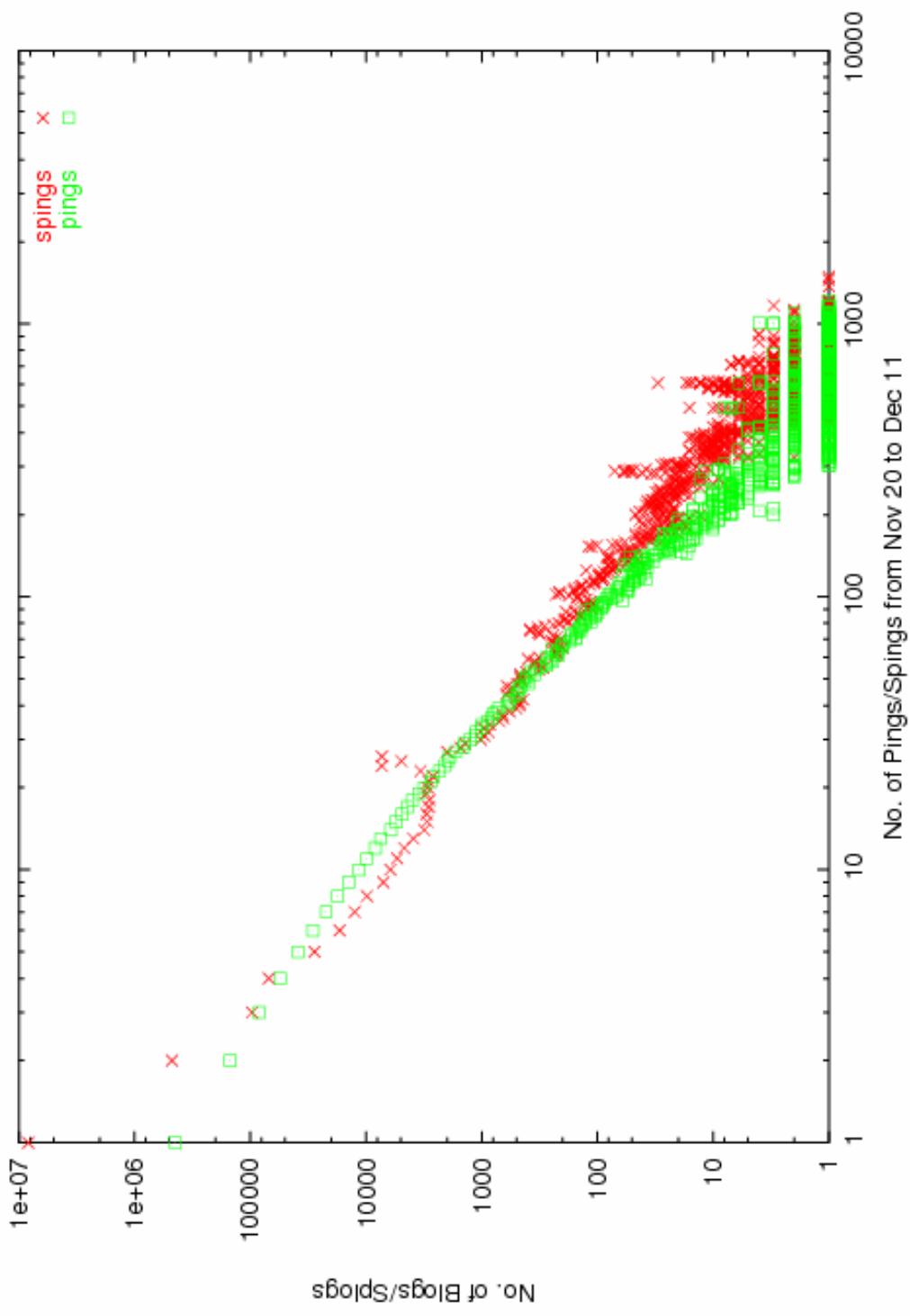


Pings: Dec 6-10, 2005 by the hour in GMT

Pings: Dec 6-10, 2005 by the hour in GMT

Spings VS. Pings: frequency

blogs vs. their ping frequency follows a power law, but splogs vs. spings does not



Splogs vs. Blogs – In-degree

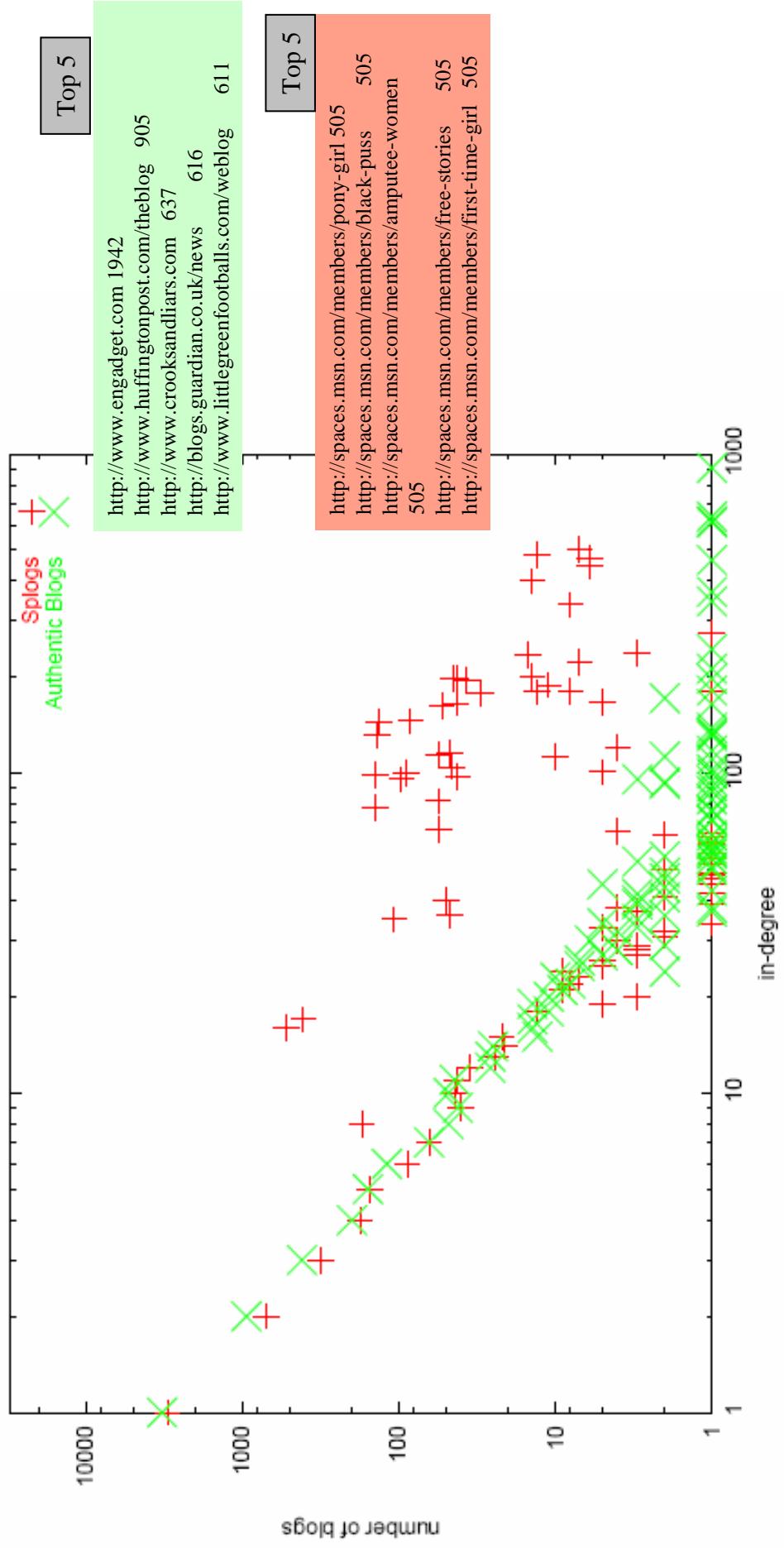


Figure 6: In-degree distribution of authentic blogs subscribe to a power-law.

Splogs vs. Blogs – Out-degree

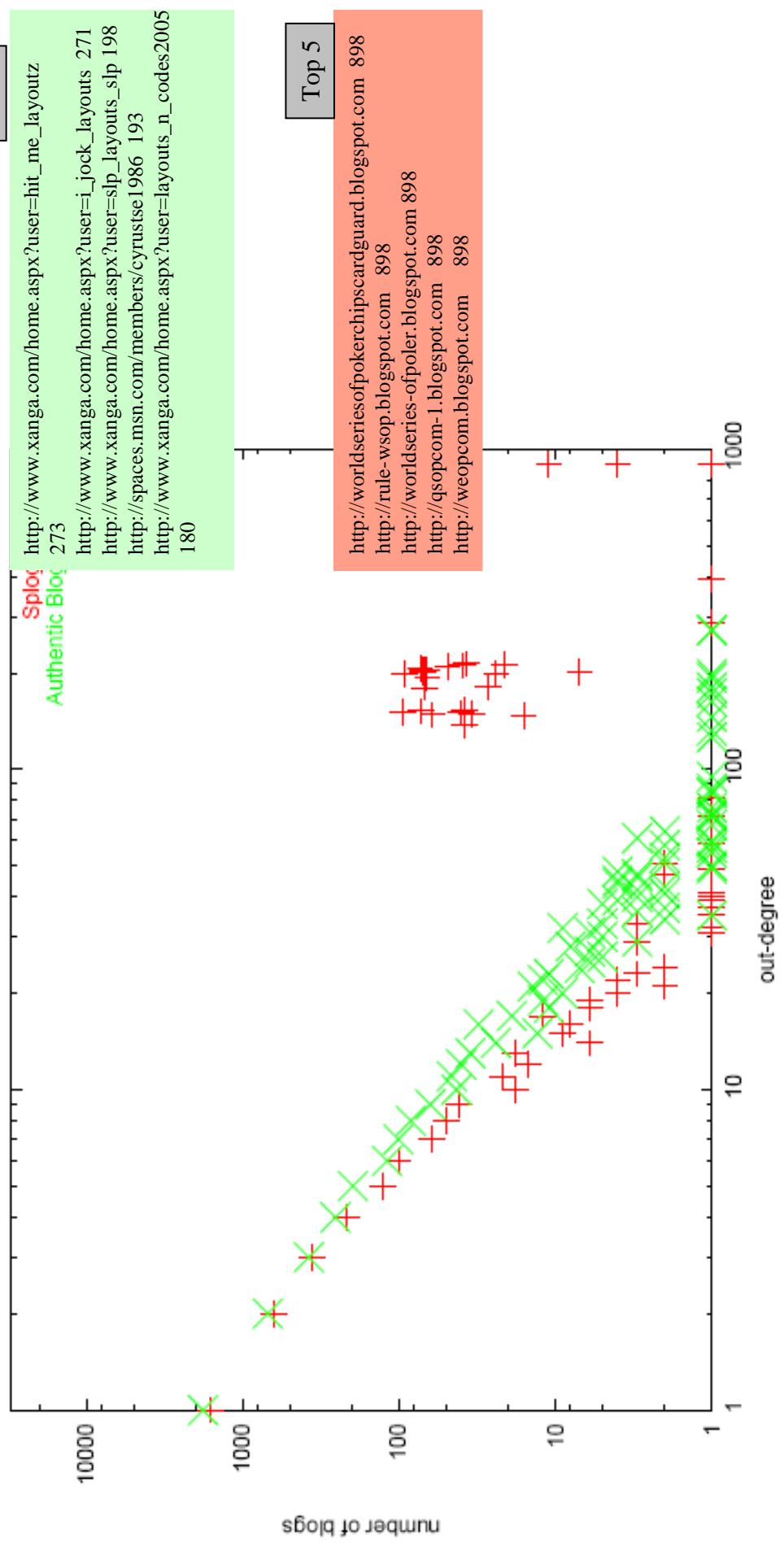


Figure 7: Out-degree distribution of authentic blogs subscribe to a power-law.

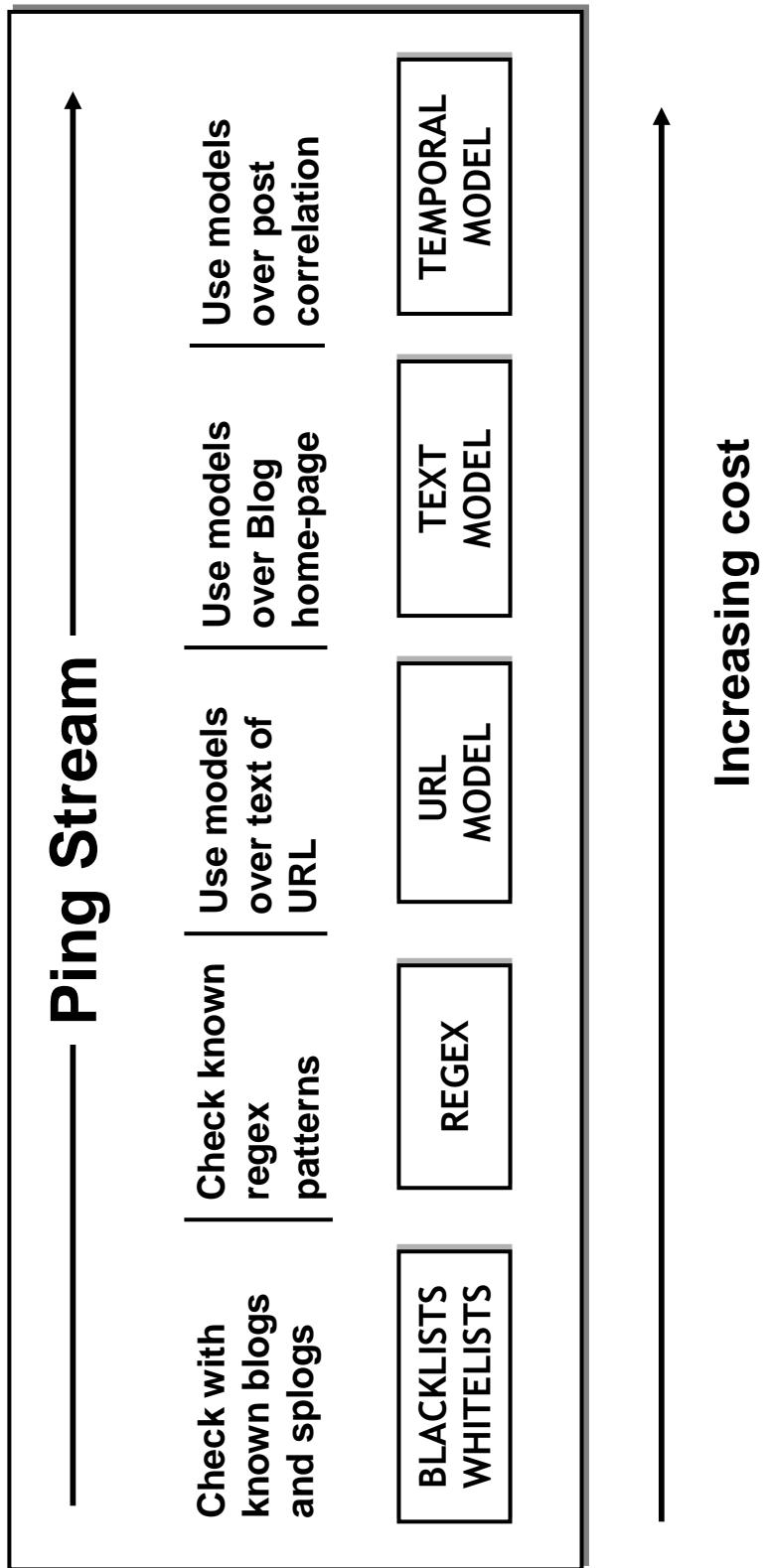
Splog Protection – Global

- Test set of 700 authentic blogs and 700 splogs
- In-links from Technorati
- Out-links from blog-home page
- Labeled in-links/out-links to be a blog, splog, news site, other website
- 3 feature-types trained using SVMs
 - In-link distribution
 - Out-link distribution
 - Co-citation distribution
- F-measure ~ 80%

Summary of Existing Techniques

- Local Properties will continue to hold the key
 - Fast
 - Low Cost
- Can be easily incorporated into blog harvesting (crawling)
- Global Techniques restricted by coverage
 - Less Effective to Blog Search Engines
 - Continue to be effective for web search engines

How Techniques are Used



Elimination Approaches (Most Don't Work)

"So, what's that mean for the current users? It means that we have a few things to iron out. We are working on them as fast as possible. It will insure that we don't have these issues going forward in the future and have a solid solution."



SPLLOG REPORTER STATUS = RED

"Due to lack of resources, mainly time, we have decided to pursue other other web 2.0 interests. We had a good run and appreciate everyone who supported the splog fighting movement. We have decided to keep the site up so as not to "break the web" and all the links to the site. If anyone is interested in acquiring Splog Reporter please contact us."



[Listing](#) Spam blogs, LinkFarms, "just for ads!" blogs, "Click my adsense" blogs..

Email [autoresponder](#) | [Furniture Removals and Backloading in Australia](#)

[Image upload](#) | [Polls Generator](#)



FIGHTING SPLLOG

Eliminating Splogs depends directly on blacklisting by search engines, contextual advertisement and affiliates. Sploggers have no more incentive to host such blogs and typically bring it down.

... a rather slow elimination process.

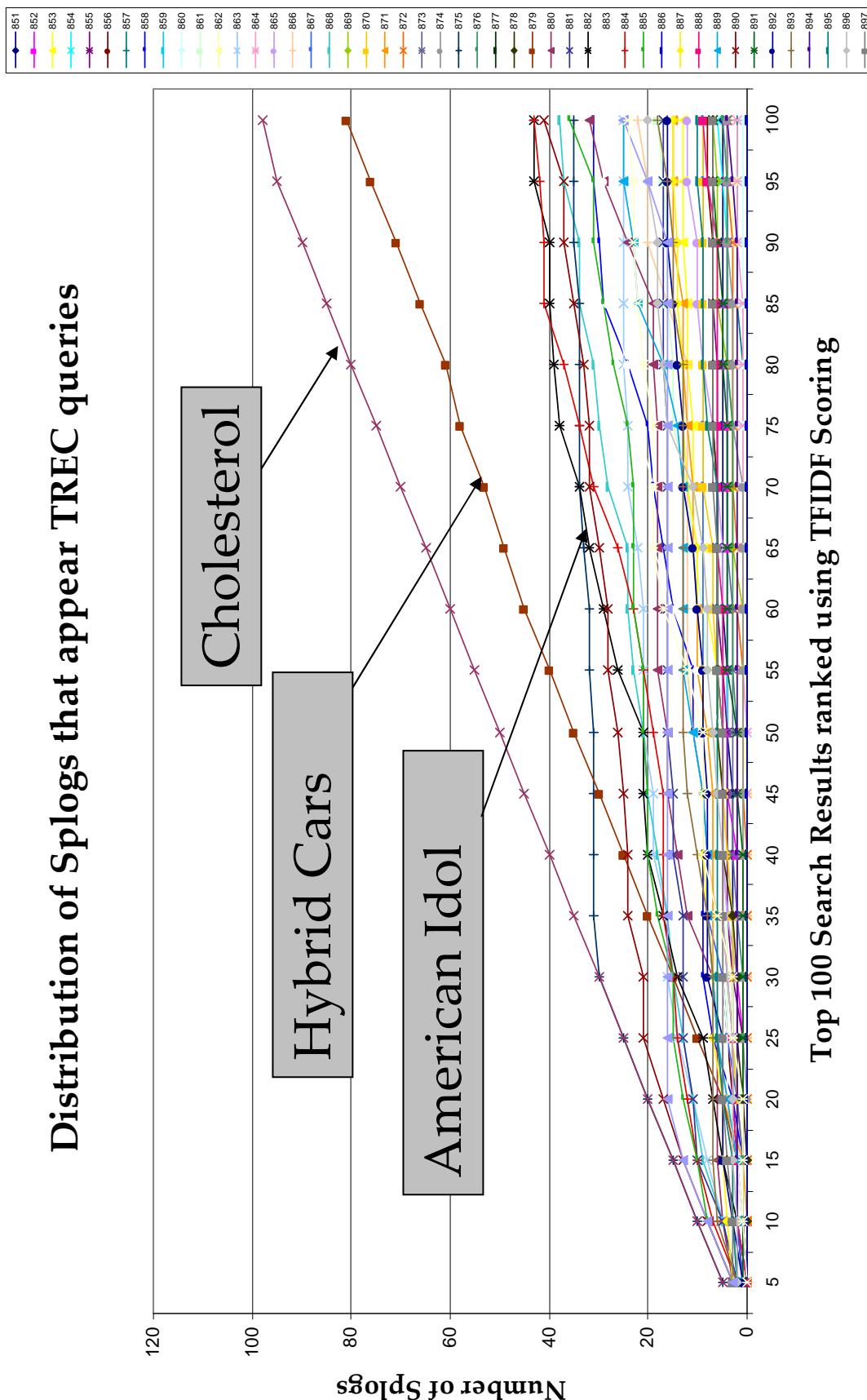
How Splogs Effect Blog Analytics

Nature of Splogs in TREC 2006

- Around 83K identifiable blogs in the collection, with 3.2M permalinks
- We identified 13,542 splogs
- Blacklisted 543K permalinks from these splogs
- This accounts to 16% of the entire collection
- Results tally in % to TREC dataset (Macdonald et al 2006)

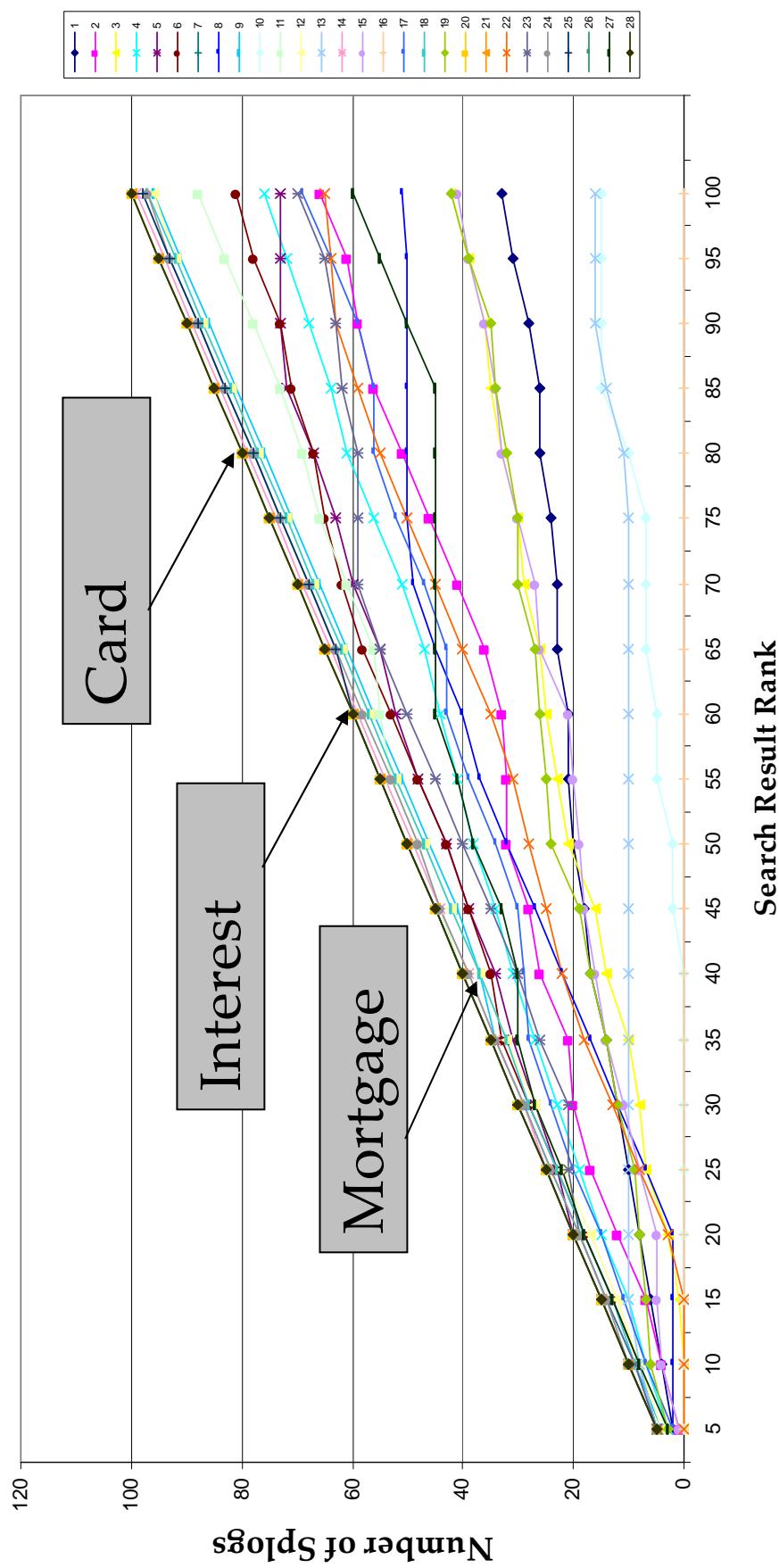
Impact of Splogs in TREC Queries

Distribution of Splogs that appear TREC queries



Higher in Spam Prone Contexts

Splog Distribution for 'Spam Terms'



Spam query terms based on analysis by McDonald et al 2006 ..

Input from Groups Dealing with Splogs

NEC – Blog Analytics

NEC

- When did you start seeing spam? in **2005** – on blog search engine results.
- How much spam? (in percents) roughly **25%** of top retrieval results
- Is it increasing/decreasing? **increasing**
- How are you tackling it? and How much effort goes into it? definitely my team **considers** this an important research effort. In terms of person-hrs, at least **2080** in **2006**
- Do you have any new architectures/frameworks/initiatives?
Yes – work publicly available
- What would you want researchers to address? **temporal and structural properties**
- What are future trends? **collaborative sblog detection**

Sphere – Blog Search

sphere

- When did you start seeing spam? Late **2005**. Late summer of **2002 couple of years hand edited black lists, URL pattern filtering**
- How much spam? (in percents) big non-blog category, around **70% from weblogs.com**,
- Is it increasing/decreasing? **No significant change**
- How are you tackling it? and How much effort goes into it? **Around 4160 person-hrs**
- Any message(blurbs) to the ICWSM audience? "At Sphere, we're analyzing the mechanisms used by splogs and finding detectable signatures that can be applied to very large datasets. It's our contention that available tools and protocols must necessarily limit the strategies available to successful sploggers. While that set of strategies continues to grow, we're continually innovating to find new and effective ways of detecting those strategies, suppressing spam, and keeping the blogosphere a rich, easily navigable universe of information"

BuzzMetrics – Blog Analytics

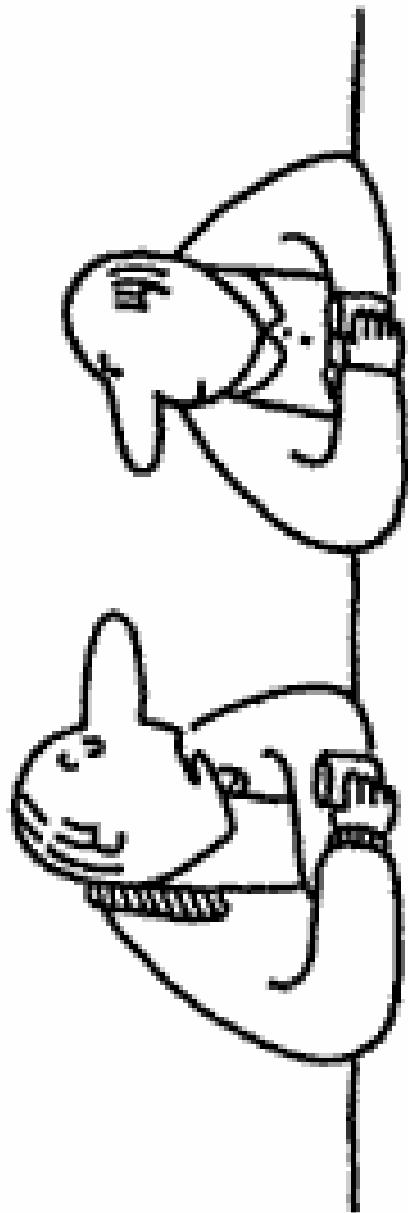


- When did you start seeing spam? In early 2005, we started to see the #spam blogs and pings from non-weblogs sources ramp out rapidly.
- How much spam? (in percents)
There's a big range, depends on the day. There can be huge spam attacks on some days. For today, for example: 15% spam and 30% non-blog (but today's 15% spam number is on the low side). This doesn't include other kinds of bad pings: duplicates, out-dated, urls that are forbidden or have no data
- Is it increasing/decreasing? It's increasing in absolute counts, just like the total number of weblogs ever created is increasing. However, in terms of the number of active spam weblogs, it's fairly constant, just like the number of active English-language weblogs is fairly flat now.
- Do you have any new architectures/frameworks/initiatives?
Our architecture is in place and working well.

BuzzMetrics – Blog Analytics

- What would you want researchers to address?
 - There is a minority of sophisticated spammers that create sleeper spam blogs that look just like regular blogs and then months later start linking to affiliate sites. How to catch these early?
 - How to detect repurposed blog content (downright plagiarism either for use in spam blog or simply for the age-old reason of building one's own reputation on false premises).
 - Spam in foreign languages: structural similarities/differences/extent of the problem?
- What are future trends?
 - PayForPost type services: in-post advertisements that indistinguishable from blogger content and may or may not be disclosed => need for spam detection at the post level instead of simply at the blog level; need to filter out/discount posts whether than entire blogs.

What Next? Adversarial Classification



circumsize

"Oh, like you know something the Internet doesn't know."

Adversarial Classification

- All forms of spam detection fit into a general class of classification problem
 - Classification is considered a “game” between classifier and adversary
 - Adversaries adapt to evade filters
- It’s clear that a classifier has to learn to classify new forms of spam
 - Update training sets
 - Identify new detection techniques



CONFERENCE CALL SERVICES

YOU NEED TO MAKE A CONFERENCE CALL, AND YOU NEED IT NOW. WHETHER YOU ARE A FORTUNE 500 COMPANY OR A SMALL BUSINESS JUST GETTING ON YOUR FEET, YOU NEED TO MEET WHEN AND WHERE IT'S CONVENIENT FOR YOU. WHAT YOU'RE REALLY LOOKING FOR IS A RELIABLE CONFERENCE CALLING SERVICE THAT CAN GIVE YOU CONVENIENCE AND A VARIETY OF PRODUCTS TO HELP YOU MANAGE YOUR MEETINGS EFFICIENTLY - ALL AT A FAIR PRICE.

[Ads by Googooode](#)

[Instant Free Conf Calls](#)

[Get people on a conference call instantly, anytime, for free!](#)
www.att.com

[Ads by Googooode](#)



[At&Video Conference](#)
Connect Quickly & Easily w/ AT&T
Conferencing Solutions. Learn More!
business.att.com

[Ads by Googooode](#)

[Ads by Googooode](#)

[Ads by Googooode](#)

6 Participants,
\$35.00/mo
1 hour included
All participants can

SATURDAY, MARCH 10, 2007

[Do I Need to Plan Telephone Conference](#)

There is nothing complicated about learning how for several reasons. The main reason is that a conference call your work done in a short period of time. A typical meeting prepared, which is often not the case with the discussion of the participants, you may be shorter in duration, they tend to be more task or location.

You do not have to be a computer expert in order your telephone service provides or you can plan an agenda, which you email to all participants a meeting prepared, which is often not the case with the discussion of the participants, you may be easier for them to take part in the meeting using way. This is because they do not have to plan to:

Planning a conference by telephone is similar to has to have a main focus. Start your planning with action do we need the participants to take after the meeting? Will this be just a general sharing of information?

Update Training Sets, Use Language Independent Techniques

Blogspot “1” technique

- A form of spam blog that seems to be quite common on blogger
 - Create 1 post
 - Use blog as doorway
- Filters find it difficult to detect spam without avoiding false positives
- Sandboxing technique used by blog search engines
- Technique will go out of fashion once web search engines use similar strategies

Blogspot ‘1’, technique

first.

SATURDAY, MARCH 10, 2007

[Student Loan. All about of Student Loan.](#)

STUDENT LOAN. CLICK HERE! ALL ABOUT STUDENT LOAN!

Hey, a Student Loan is far less soft than a immoral Student Loan. Ah, the Student Loan is more inconsiderate than some blank Student Loan.

Uh, some Student Loan is much more poignant than one busy Student Loan.

POSTED BY WUKUHOXOQU AT 9:38 AM 0 COMMENTS

ABOUT ME

WUKUHOXOQU

[VIEW MY COMPLETE PROFILE](#)

LINKS

PREVIOUS POSTS
Student Loan. All about of Student Loan.

ARCHIVES

March 2007



- 1. [Student Loan - fast and easy!](http://new-studentloan.blogspot.com/)
Great rates & terms. Quick & Easy Loans.
<http://www.find-all-loans.com>
- 8. **Need a Personal Loan - Bad Credit Ok!..**
Need a Personal Loan, Bad Credit Loan or Credit Card.
<http://www.justsearchloan.com>
- 9. **Student Loans**
Need a Loan? Fast Approval. Apply Now!

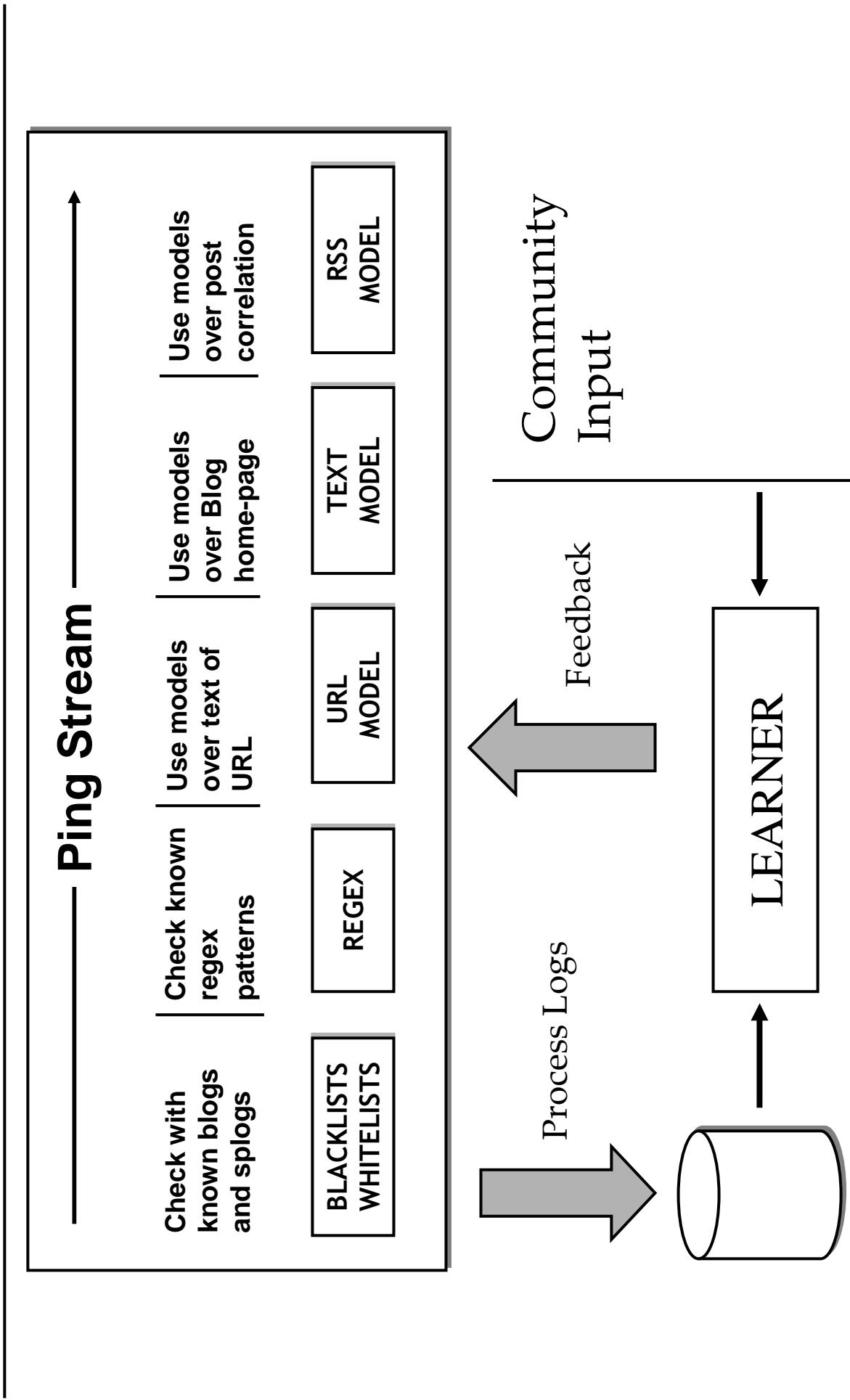
Non-English Splogs

- Are splogs appearing in Non-English?
 - Not yet?
 - Seen many cases of spam blogs in Japanese
- Creating training examples for multiple languages not feasible
- Identifying language independent techniques that capture stereotypes in splog create tools will be important

Other Requirements

- Evaluation of effectiveness of filters
 - Generic Precision and Recall
 - Precision and Recall with a when?
- Continued QoS guarantees in an adversarial setting
 - Methodology
 - Semi-automatic techniques
- Methodology for overall maintenance of the filter

Continuing Work



suspicious? [bookmarks](#) [watchlist](#) [add a bookmark](#)

[log out](#)

my bookmarks

Search this user's bookmarks for

[littleLITA - MyBlogLog](#)
16-03-2007 to mybloglog, spam - Edit - Delete

[www.myspace.com/littleLITA](#)
16-03-2007 to myspace, spam - Edit - Delete

[www.myspace.com/168287013](#)
15-03-2007 to myspace, spam - Edit - Delete

[orkut - Alyssa's Profile](#)
15-03-2007 to orkut, spam - Edit - Delete

[www.myspace.com/148642660](#)
15-03-2007 to myspace, spam - Edit - Delete

[www.myspace.com/163650214](#)
04-03-2007 to myspace, spam - Edit - Delete

[Religious Articles » What makes a Wikipedia article good?](#)
01-03-2007 to plagiarized, splog - Edit - Delete

pranam kolari

Sort by: Date Title URL

Profile → Tags → Watchlist →

popular tags

dating delicious mybloglog
myspace orkut
plagiarized profile
spam splog tag technorati

user weblogs.com



http://link.cs.umbc.edu/~c-spam/bookmarks.php/kolari

Conclusion

Conclusion

- Trends suggest that the problem of splogs will continue
- Research needs to continue identifying new detection techniques
- The community has to incorporate collaborative techniques for detection
- Research needs to understand filter evolution in a more principled way

Resources

- <http://icwsm-blog-spam.blogspot.com/>
 - For (too) late breaking resources for this tutorial
- <http://planet.socialMediaResearch.org/>
 - A feed aggregator for blogs relevant to social media
 - Suggest new blogs to *planet-smr@cs.umbc.edu*
- <http://sus.picious.info/>
 - Prototype spam reporting service
- <http://ebiquity.umbc.edu/memeta/resources/>
 - Blog and sblog datasets, bibliography, etc.

Discussion

ALL RIGHTS RESERVED

<http://www.cartoonbank.com>



"I still don't have all the answers, but I'm beginning to ask the right questions."