Spam
&
The Power of Social Networks

Ricardo Baeza-Yates
Yahoo! Research
Barcelona, Spain & Santiago, Chile

Thanks to Tim Converse (Yahoo! Search) & R. Ravi (CMU)

Models of Trust Workshop: May 22, 2006

Agenda

• Spam
• What we have to fight it
• Current content spam challenges
• Social networks
• Current advertising spam challenges
• What we do to fight content spam
• Final Remarks
Types of Spam

- Content Spam (e.g. Cloaking, Quilt-like pages)
- Link Spam (e.g. Link Farms)
- Metadata spam (e.g. Misleading info)
- Ad spam (e.g. Click spam, Bid spam)

Adversarial Web Retrieval

Spam is an Economic Activity

- Depending on the goal and the data spam is easier to generate
- Depending on the type & target data spam is easier to fight

- Disincentives for spammers?
  - Social
  - Economical
- Exploit the power of social networks and their work
Crawled Data

- **WWW**
  - Web Pages & Links
  - Blogs
  - Dynamic Sites

- **Sales Providers (Push)**
  - Advertising
  - Items for sale: Shopping, Travel, etc.

- **News Index**
  - RSS Feeds
  - Contracted information

Produced data

- **Yahoo’s Web**
  - Ygroups
  - YCars, YHealth, Ytravel

- **Produced Content**
  - Edited (news)
  - Purchased (news)

- **Direct Interaction:**
  - Tagged Content
    - Object tagging (fotos, pages, ?)
    - Social links
  - Question Answering

```
Models of Trust Workshop, WWW2006, Edinburgh
```
Observed Data

- **Query Logs**
  - spelling, synonyms, phrases (named entities), substitutions

- **Click-Thru**
  - relevance, intent, wording

- **Advertising**
  - relevance, value, terminology

- **Social**
  - links, communities, dialogues...

---

*Current challenges (1)*

- **Scraper spam**
  - Copies good content from other sites, adds monetization (most often Google AdSense)
  - Hard to identify at the page level (indistinguishable from original source), monetization not reliable clue (there is actually good content on the web that uses AdSense/YPN!)

- **Synthetic text**
  - Boilerplate text, randomized, built around key phrases
  - Avoids duplicate detection

- **Query-targeted spam**
  - Each page targets a single tail query (anchortext, title, body, URL). Often in large auto-constructed hosts, host-level analysis most helpful

- **DNS spam**
Current challenges (2)

• **Blog spam**
  – Continued trend toward blog “ownership” rather than comment spam
  – Orthogonal to other categories (scrapers, synthesizers). Just a hosting technique, plus exploiting blog interest

• **Example:**
  – 68,000 blogspot.com hosts all generated by the same spammer
    • 1) nursingschoolresources.blogspot.com
    • 2) transplantresources.blogspot.com
    • ... 67,798) beachesresourcesforyou.blogspot.com
    • 67,799) startrekresourcesforyou.blogspot.com

The wisdom of spammers

• Many world-class athletes, from all sports, have the ability to get in the right state of mind and when looking for **russian women looking for love** the state of mind is most important. [...] You should have the same attitude in looking for **russian women looking for love** and we make it easy for you.

• Many world-class athletes, from all sports, have the ability to get in the right state of mind and when looking for **texas boxer dog breeders** the state of mind is most important. [...] You should be thinking the same when you are looking for **texas boxer dog breeders** and we make it easy for you.
The power of social networks & media

- Example: Flickr – community phenomenon
- Millions of users share and tag each others’ photographs (why???)
- The wisdom of crowds can be used to search
- The principle is not new – anchor text used in “standard” search

The Wisdom of Crowds
Challenges in social media

• What’s the ratings and reputation system?
• How do you cope with spam?
  – The wisdom of the crowd can be used against spammers

• The bigger challenge: where else can you exploit the power of the people?
• What are the incentive mechanisms?
  – Example: ESP game
• *Spammers many times are (or look like) social networks*
  – But the Web has larger social networks

• **Examples**
  – Any statistical deviation is suspicious
  – Any bounded amount of work is suspicious
  • Truncated PageRank (joint work with people at Univ. of Rome)
    – *Spammers link support have shorter incoming paths*

---

**Content match = meeting of Publishers, Advertisers, Users**

and Spammers! Grrr...
Contextual ads

1. Advertisement for Amazon.com featuring J.S. Bach:
   - Music by J.S. Bach at Amazon.com
   - Amazon.com has a huge selection of merchandise, including videos and DVDs at great savings. Free Super Saver Shipping available.
   - Find "J.S. Bach" from $55.00 at Buy.com
   - Buy now at Buy.com, with over 1 million products. Customers can buy with confidence at Buy.com.
   - If It Makes Music, It's on eBay
   - You can find J.S. Bach music and collectibles right here. Today, you'll find the artists you're looking for on eBay.

2. Advertisement for STLtoday.com:
   - Bush policies endanger rights, protesters say
   - By Bumgarner Makhso, Post-Dispatch Washington Bureau
   - Washington, D.C. - Armed with bold placards and bright-colored banners, abortion-rights supporters from Missouri and Southern Illinois marched in step with throngs of other protesters Sunday to rally against what they see as an "attack."
Click spam

- **Rival click fraud**: Rival of advertising company employs clickers for clicking through ads to exhaust budget

- **Publisher click fraud**: Publisher employs clickers to reap per-click revenue from ads shown by search firm

- **Bidder click fraud**: Keyword bidders employ clickers to raise rate used in (click-thru-rate * bid) ranking used to allocate ad space in Google (or to pay less!)
Other bidders for ad space are static (in their bids and CTRs)
Only decisions to make are computation of CTR
Payment of advertiser inversely proportional to CTR
Single ad space allocated via (second-price) auction to highest submission of (bid*CTR)
The set of click-based algorithms are shown to be fraud-resistant for an appropriate definition
Loosely shown to be equivalent to pay-per-impression pricing

Other Possible Ad Spam

Rival buys misleading or fraudulent ads
  Queries
  Bids
  Ads
Rival submits queries that brings up competitor ad but without clicking on it
  Reduces rival’s CTR and hence its ranking for ad space
Current goals for spam effort

- Prevent spam from distorting ranking, but preserve:
  - Relevance
    - “Perfect spam” is a sensible category
  - Freshness
    - Can’t slow down discovery just because spammers exploit it
  - Comprehensiveness
    - Navigational queries for spam should succeed

- Focus on two kinds of spam only:
  - 1) Spam that is succeeding in ranking inappropriately highly
  - 2) Spam that consumes huge amounts of system resources
    (Everything else is “dark matter”)

How do we fight content spam at Yahoo!?

- Mix of algorithmic and (we admit it) editorial techniques
  - Editors do catch egregious spammers, label them directly
  - This would be fine by us if it scaled to 20B docs.

- Editors interact with algorithms by:
  - Providing test cases
  - Judging output of algorithms
  - Acting as domain expert and “canary in coalmine”

- Algorithmic techniques
  - Spam-resistant general ranking (first line of defense)
  - “Corrective” algorithms
  - Identification algorithms
What are our goals as indexers?

- Prevent spam from distorting ranking
  - Main obligation is to preserve result quality for users
  - Otherwise would like to avoid imposing judgement
    - Contrasts with aggressive deletion in Inktomi era
- Spam fighting trades off with other quality goals
  - Spam-fighting trades off with comprehensiveness
  - Spam-fighting trades off with relevance
- Goal is to identify and characterize spam algorithmically
  - Relevance and product goals drive exact outcome once identified

How to collaborate

- True spam-data sharing between indexers is problematic
  - Competitive arena
- Can indexers share spammer lists with publishers?
  - If not interested in simple spammers, how about spammers who violate TOS (scrapers of copyrighted material)?
- If publishers had a choice, what would they want indexers to do with spam on the publisher’s own service?
  - Delete, demote, report?
Many Open Problems in Ad Spam

- Trust Models
- Disincentive mechanisms
- Detection Algorithms (preprocessing, on-line)
- .......

Spam & The Power of Social Networks

Ricardo Baeza-Yates
Yahoo! Research
Barcelona, Spain & Santiago, Chile

Questions?

Models of Trust Workshop: May 22, 2006