

ECE 498KL: eCrime and Internet Service Abuse

# Web Cloaking

Kirill Levchenko

November 27, 2018

**I** ILLINOIS

Electrical & Computer Engineering

COLLEGE OF ENGINEERING

# Radare2: First impressions

Oct 29, 2016 • Stephen Checkoway

*This post has been translated into [Swedish](#) by Weronika Pawlak.*

**Update (2016-11-01):** The Radare2 creator has informed me that all of the issues I mention below have been fixed.



**Stephen Checkoway** @stevecheckoway · Oct 29, 2016



Yesterday, I played with radare2 for the first time. It didn't really go well. [cs.uic.edu/~s/musings/rad...](https://cs.uic.edu/~s/musings/rad...)



**pancake**   
@trufae

.@stevecheckoway i fixed all the issues you say in the post. but anyway, r2 is far from perfect, but everything works if you use it properly

♥ 3 12:03 PM - Nov 1, 2016



# piecesauto-pro.fr | professional translator

HOME

## Radare2: Första intrycket

[Leave a reply](#)

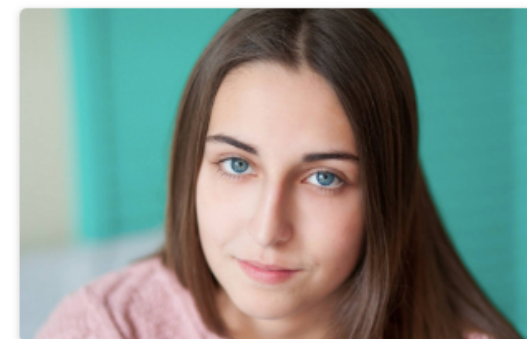
Link: <https://www2.cs.uic.edu/~s/musings/radare2-first-impressions/>

**Update (2016-11-01): Radare2 skapare har informerat mig om att alla de frågor jag nämner nedan har rättats till.**

[Stephen Checkoway @stevecheckoway](#)

Igår spelade jag med radare2 för första gången. Det gjorde inte riktigt gå bra. [https://www.cs.uic.edu/~s/funderingar/radare2-första-intryck/...](https://www.cs.uic.edu/~s/funderingar/radare2-första-intryck/)

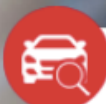
[Piecesauto-Pro.FR](#)



My name is Weronika Pawlak. I am a student of the Philology fac-



## Votre expert en pièces détachées trouvera la pièce adaptée à votre voiture



### DÉFINISSEZ VOTRE VÉHICULE

Choisissez une marque



Choisissez un modèle



Choisissez une motorisation



RECHERCHER

### SAISIE DU NUMÉRO D'IMMATRICULATION



AB-123-CD

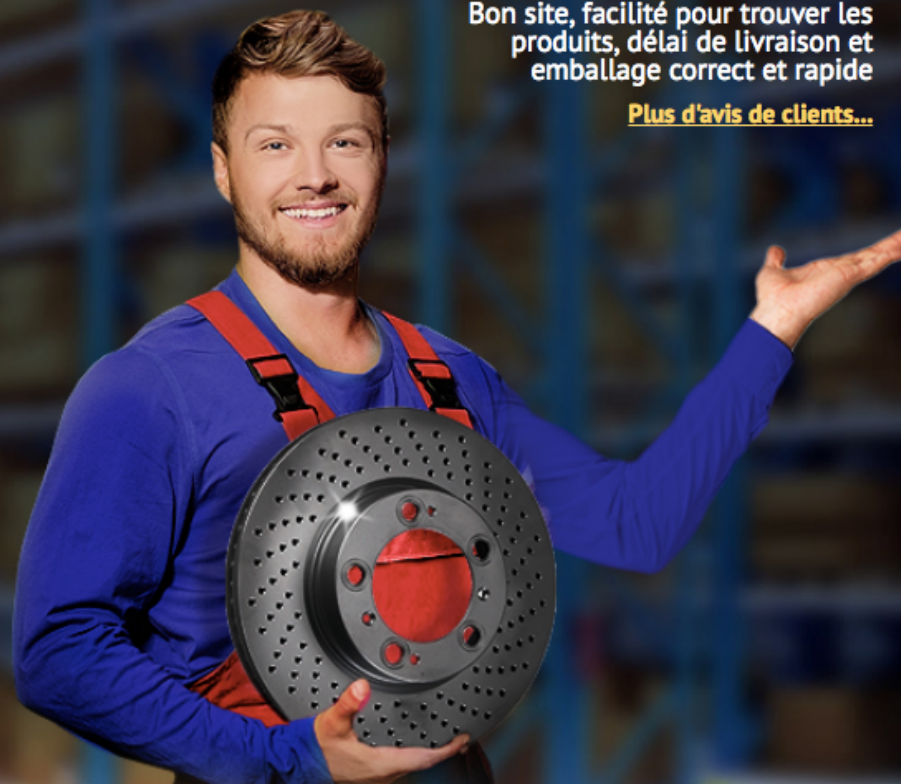


## D'avis de clients

★★★★★ 4.93/ 5.00

Bon site, facilité pour trouver les produits, délai de livraison et emballage correct et rapide

[Plus d'avis de clients...](#)



Pièces détachées d'origine



Choisissez une voiture

Why is a *French auto parts store* hosting a Swedish translation of an English blog post that is a review of a reverse-engineering tool?

# Result Ranking

- ❖ Result rank is combination of content and page rank
- ❖ **Content rank** based on page content (relevance)
  - Optimize using keywords, title, site layout, etc.
- ❖ **Page rank** based on incoming links (authority)
  - Optimize by manipulating Web link structure
- ❖ Specifics of algorithm are a closely guarded secret
  - Creates an SEO mythology based on inferred behavior



# THE FLOW OF **LINK JUICE** A QUICK GUIDE



WooRank / [SEO Guides](#) / [What is Link Juice?](#)

Link juice is the term used in the SEO world to refer to the value or equity passed from one page or site to another. This value is passed through hyperlinks. Search engines see links as votes by other websites that your page is valuable and worth promoting.

There are many ways to earn links from the web through direct and indirect efforts. Direct effort refers to [link building strategies](#), such as document sharing, guest posting, social media marketing, press release publishing and more. The indirect effort is gained from presenting excellent content on your site that causes readers to share it around the web, linking the pages naturally. The link equity that passes from these sites to your site is the link juice, and this link juice differs in its authority depending on the sites linking to you.

## How Does Link Juice Work?

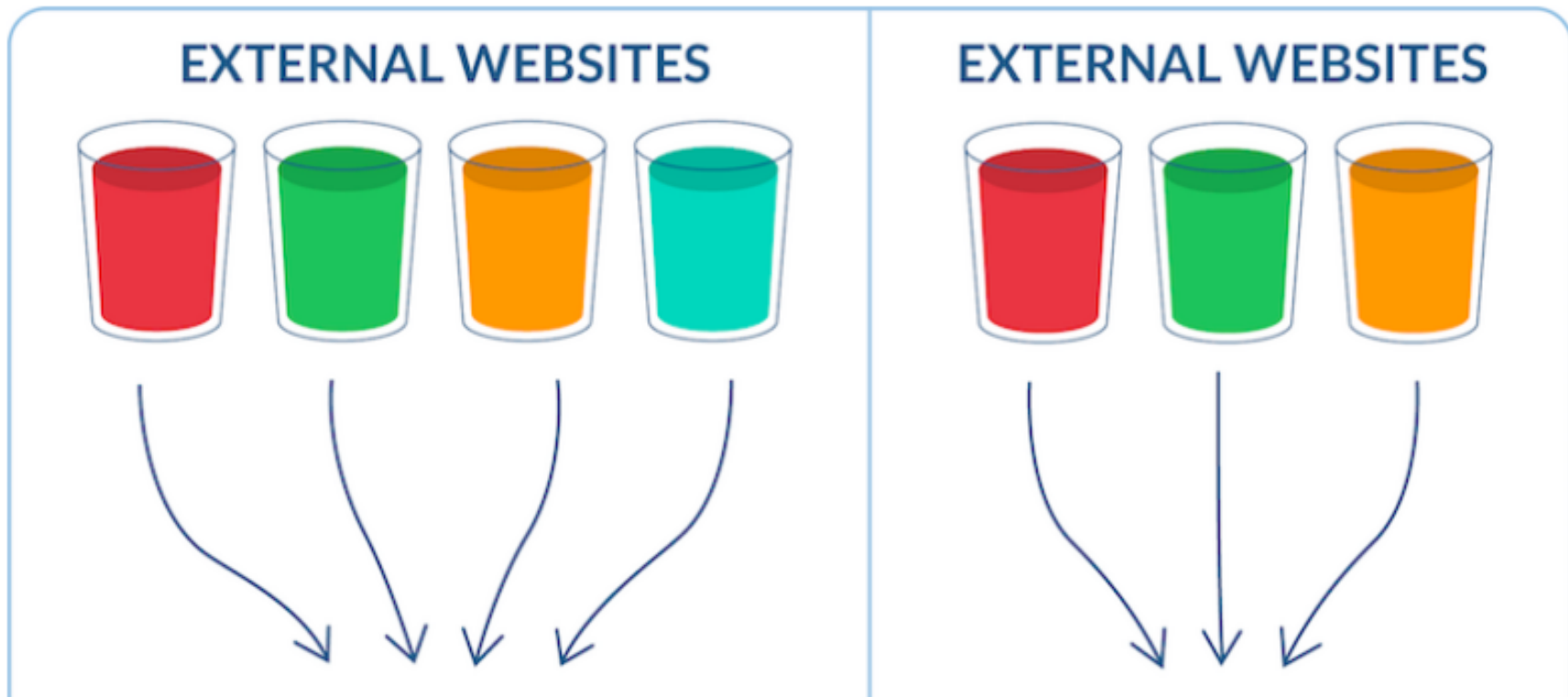
Suppose you have sites A and B. If all other ranking factors are constant and site A has one



pages naturally. The link equity that passes from these sites to your site is the link juice, and this link juice differs in its authority depending on the sites linking to you.

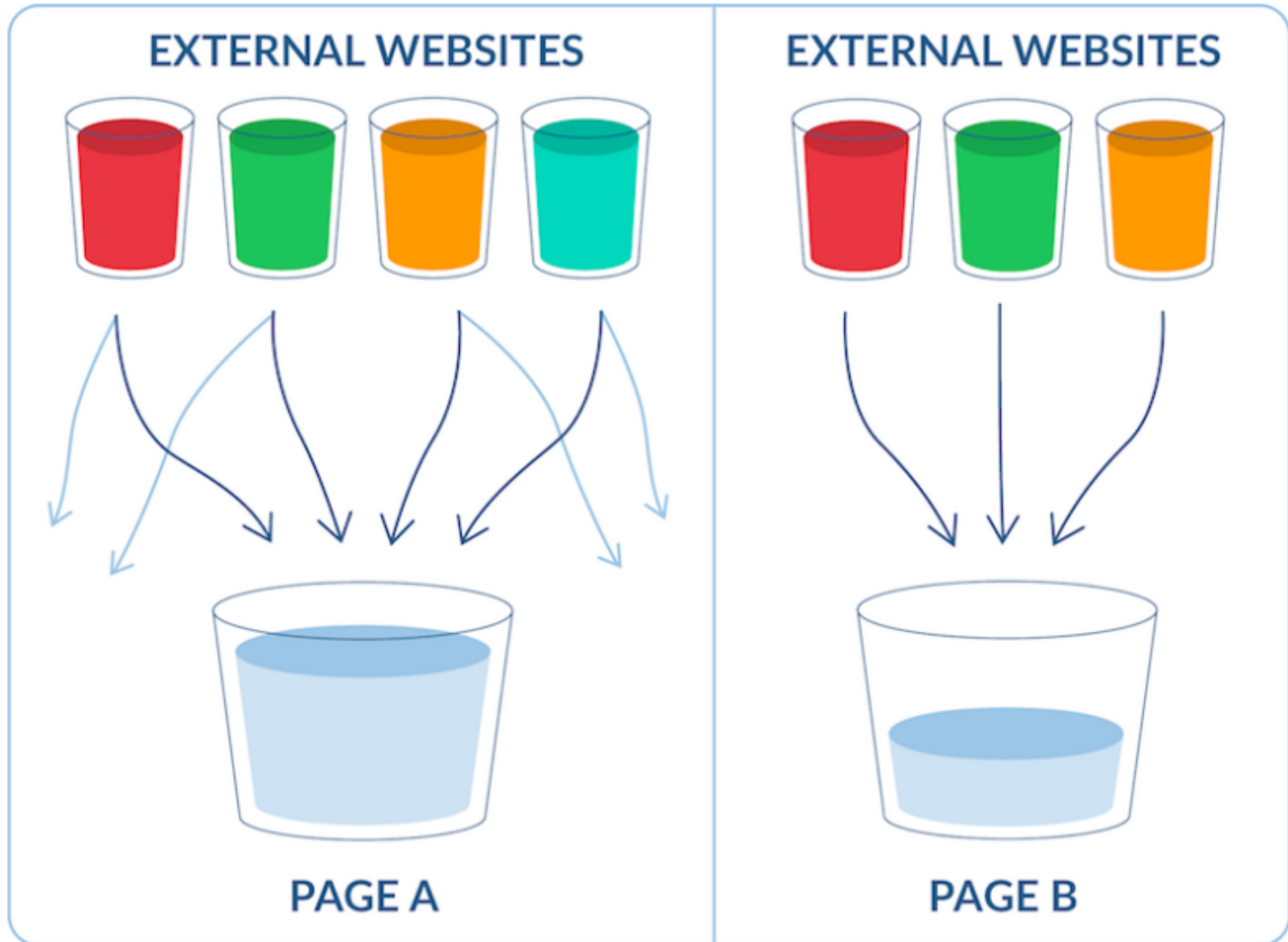
## How Does Link Juice Work?

Suppose you have sites A and B. If all other ranking factors are constant and site A has one link while site B has no links, site A will rank higher in search results due to the link juice it receives from the external site linking to it. What happens if site B also gains one link? This depends on the amount of juice each link passes. Look at the diagram below. Site A receives links from four sites while B receives links from two sites. All the linking sites receive link juice from other sites too. Since A receives links from more sites, there is more link juice being transferred to A and consequently, A is likely to rank higher than B in search results. Note: These results assume the sites linking to A and B have a similar authority.



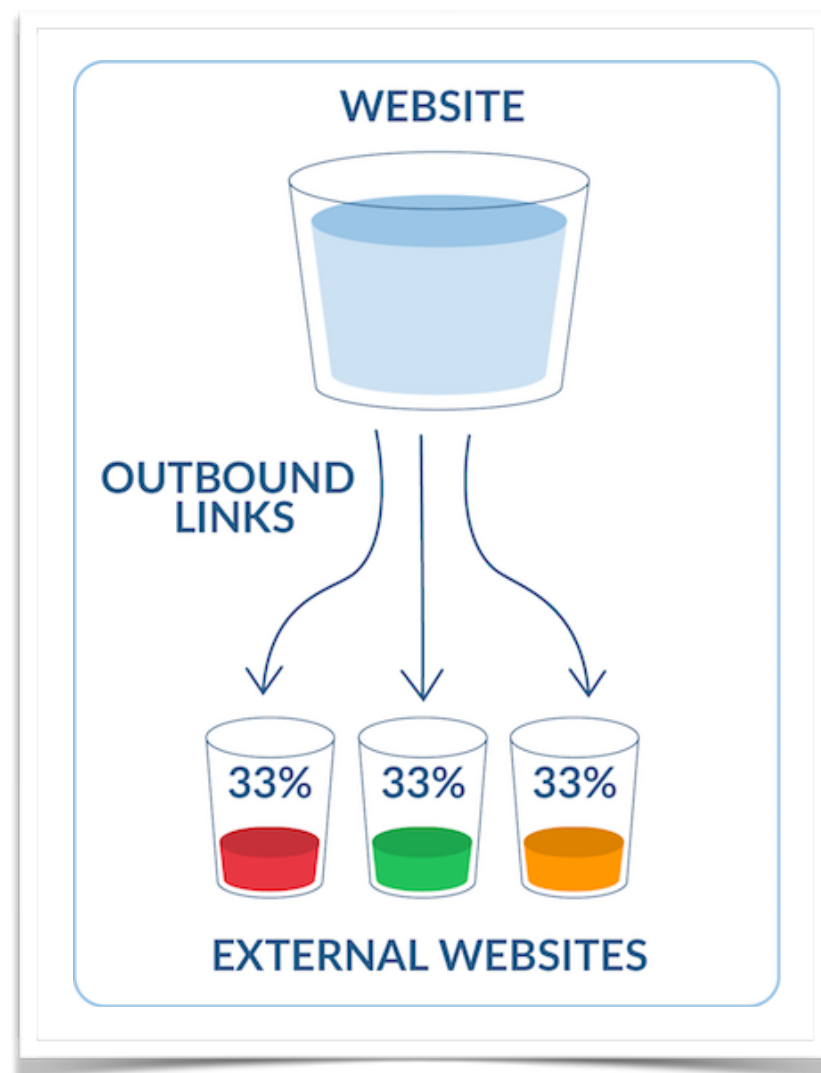


links from four sites while B receives links from two sites. All the linking sites receive link juice from other sites too. Since A receives links from more sites, there is more link juice being transferred to A and consequently, A is likely to rank higher than B in search results. Note: These results assume the sites linking to A and B have a similar authority.



# Link Juice Outflow

- ❖ Link juice also called *link equity*
- ❖ Linking page divides its link juice equally between its outgoing links
  - Real search engines may implement slightly different algorithm
- ❖ Volume of *incoming* link juice determines page rank



# Stealing Link Juice

- ❖ ***Can you steal link juice?***
- ❖ Page does not *lose* link juice through outgoing links
  - Page rank (authority) determined by *incoming* links only
  - Outgoing links do not adversely affect linking page
- ❖ Sites have no legal ownership claim to link juice
  - Page rank is a a third party assessment of a site's authority

# Spamming for Link Juice

- ❖ Spam sites with user-generated content (e.g. blog comments) with links to target site
- ❖ Spam comments are a form of free advertising
- ❖ Spammed site sends also link juice to target site
- ❖ Spamming hurts *user experience* on spammed site
- ❖ Users may *hold spammed site responsible* for content of target pages

# Spamming for Link Juice

- ❖ Sites with user-generated content attractive targets
- ❖ Countermeasure: `nofollow` attribute

```
<a rel="nofollow" href="http://seo-target.com">
```

- ❖ Tells search engines not to send link juice along link
- ❖ Eliminates page rank incentive to spam

# Link Juice from Compromise

- ❖ What can you do if you compromise a site with high page rank?
  - Complete control over compromised site's content
- ❖ **Simple:** Put links to target site (spamming)
- ❖ **Problem:** Site owner will notice change to site
- ❖ **Solution:** show *original* version of site to normal visitors and a *different* version of site to search engine crawlers

# Cloaking

- ❖ **Cloaking:** delivering semantically different content to different site visitor groups
  - Entails deception of at least one visitor group
  - Minor differences (e.g. mobile vs desktop version) not cloaking
- ❖ Search engines see one version of site
  - Site optimized for search engines
- ❖ Real users see another version of site
  - Site optimized for cloaker's end goals



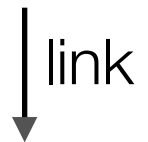
# Cloaking for SEO

- ❖ Hide compromise by showing *original site* to real users
- ❖ Show *SEO content* to search engines (SEs)
  - Links to a target site that attacker wants to promote

Users see:

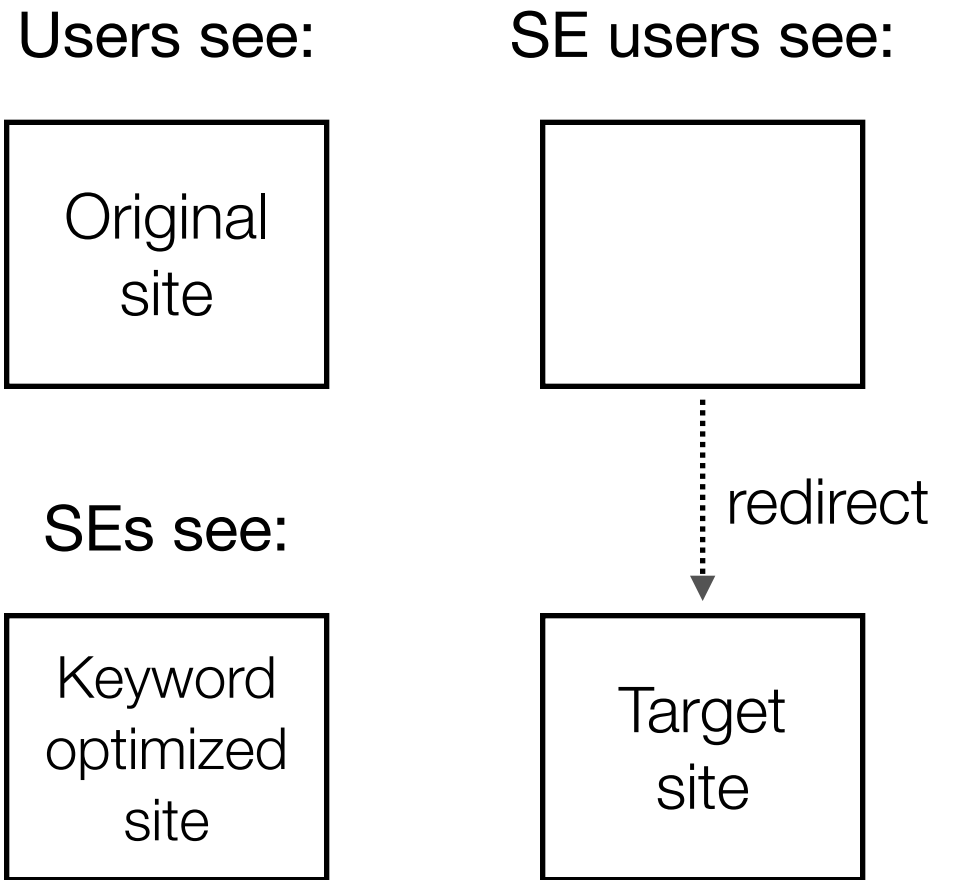


SEs see:



# Cloaking for Direct Monetization

- ❖ **Goal:** bring users searching for particular terms to your site (usually site where you can directly monetize visit)
  - **Monetize:** extract money from user



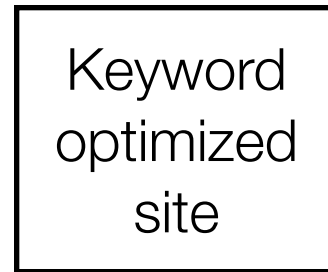
# Cloaking for Direct Monetization

- ❖ Direct visitors see original site
  - Hide from site owner
- ❖ Search engines see keyword-optimized site
  - Site gets indexed for keywords
- ❖ Visitors from search results are redirected to monetization site

Direct visitors:



SEs see:



SE visitors see:



Live Demo

# Detecting Search Engines

- ❖ **User-Agent:** SE crawlers use specific User-Agent

# Google crawlers (user agents)

See which robots Google uses to crawl the web

"Crawler" is a generic term for any program (such as a robot or spider) used to automatically discover and scan websites by following links from one webpage to another. Google's main crawler is called [Googlebot](#). This table lists information about the common Google crawlers you may see in your referrer logs, and how they should be specified in [robots.txt](#), the [robots](#) meta tags, and the X-Robots-Tag HTTP directives.

In the following table, the **user agent token** is used in the `User-agent:` line in robots.txt to match that specific crawler. Some crawlers respond to more than one token, as shown in the table; you need to use only one matching token for a crawler. This list is not complete, but covers most of the crawlers you can see on your website.



These values can be spoofed. If you need to verify that the visitor is Googlebot, you should [use reverse DNS lookup](#).

Crawler	User agent tokens (used in robots.txt)	Full user agent string (as seen in website log files)
<a href="#">APIs-Google</a>	<ul style="list-style-type: none"><li><code>APIs-Google</code></li></ul>	<code>APIs-Google (+https://developers.google.com/webmasters/APIs-Google.html)</code>
<a href="#">AdSense</a>	<ul style="list-style-type: none"><li><code>Mediapartners-Google</code></li></ul>	<code>Mediapartners-Google</code>

Googlebot  
Video

- Googlebot-Video
- Googlebot

Googlebot-Video/1.0

Googlebot [🔗](#)

(Desktop)

- Googlebot

- Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)
- Mozilla/5.0 AppleWebKit/537.36 (KHTML, like Gecko; compatible; Googlebot/2.1; +http://www.google.com/bot.html) Safari/537

or (rarely used):

- Googlebot/2.1 (+http://www.google.com/bot.html)

Googlebot [🔗](#)

(Smartphone)

- Googlebot

Mozilla/5.0 (Linux; Android 6.0.1; Nexus 5X Build/MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2272.96 Mobile Safari/537.36 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)

Mobile

- Mediapartners-

(Various mobile device types) (compatible;



# Detecting Search Engines

- ❖ **User-Agent:** SE crawlers use specific User-Agent
- ❖ **Reverse DNS:** Crawler host name will belong to SE

# Verifying Googlebot

You can verify if a web crawler accessing your server really is Googlebot (or another [Google user-agent](#)). This is useful if you're concerned that spammers or other troublemakers are accessing your site while claiming to be Googlebot. Google doesn't post a public list of IP addresses for webmasters to whitelist. This is because these IP address ranges can change, causing problems for any webmasters who have hard-coded them, so you must run a DNS lookup as described next.

## To verify Googlebot as the caller:

1. Run a reverse DNS lookup on the accessing IP address from your logs, using the `host` command.
2. Verify that the domain name is in either `googlebot.com` or `google.com`
3. Run a forward DNS lookup on the domain name retrieved in step 1 using the `host` command on the retrieved domain name. Verify that it is the same as the original accessing IP address from your logs.

## Example 1:

```
> host 66.249.66.1
1.66.249.66.in-addr.arpa domain name pointer crawl-66-249-66-1.googlebot.com.

> host crawl-66-249-66-1.googlebot.com
crawl-66-249-66-1.googlebot.com has address 66.249.66.1
```

## Example 2:

```
> host 66.249.90.77
77.90.249.66.in-addr.arpa domain name pointer rate-limited-proxy-
```

# Detecting Search Engines

- ❖ **User-Agent:** SE crawlers use specific User-Agent
- ❖ **Reverse DNS:** Crawler host name will belong to SE
- ❖ **IP address:** SE crawlers usually use known addresses
- ❖ **robots.txt:** SE crawlers will access robots.txt
- ❖ **Other ways?**

# Detecting Search Traffic

- ❖ How to determine user clicked on search result?
- ❖ **Referrer** header shows previous page
  - Empty if user typed in URL directly
- ❖ Would normally include search terms in GET request
- ❖ Google search only shows that user came from Google
  - Does not reveal search terms

Referrer demo

# Handling Search Traffic

- ❖ Redirect search traffic directly to monetization page
  - Usually affiliate marketing URL
  - More on this next lecture

# Handling Non-Search Traffic

- ❖ Assume visitor is there for original site
- ❖ (Optional) Set cookie to remember this



# Leonard Cassuto



**HOME**

**NEWS**

**BOOKS**

**OTHER WRITING**

**REVIEWS**

**APPEARANCES**

**TEACHING**

**BIO/BIBLIO**

**CONTACT**

## News

[more news >>](#)

### Swedish translation of interview with me

Take note: Weronika Pawlak, a translator at Jagiellonian University in Poland, has just published [this translation into Swedish](#) of an interview with me that was originally published in The Chronicle of Higher Education.

previous: [Danish translation of a page from this site!](#)

next: [All Publicity is Good Publicity Dept.](#)

[more news >>](#)

# Leonard Cassuto



## News

[more news >>](#)

### Danish translation of a page from this site!

In a treat for my many Danish readers, the translator Mille Eriksen has translated a page from this site. Read all about *Hard-Boiled Sentimentality* in Danish [here](#).

previous: [Interview on Jack London](#)

next: [Swedish translation of interview with me](#)

[more news >>](#)

HOME

NEWS

BOOKS

OTHER WRITING

REVIEWS

APPEARANCES

TEACHING

BIO/BIBLIO

CONTACT


# piecesauto-pro.fr | professional translator

HOME

## En Konversation Med Leonard Cassuto på "Graduate School Röra"

[Leave a reply](#)

Link: <http://www.chronicle.com/article/A-Conversation-With-Leonard/234101/>

"Vi vidmakthålla en kultur som fortlä  pande doktorander"

Av Rebecca Schuman NOVEMBER 08, 2015

Tim Foley Krönikan

Leonard Cassuto är arg som fan om tillståndet för forskarutbildning i Usa, och han kommer inte att ta det längre. Eller, okej, han är passionerat i fråga, och han honnas att hans nya bok: *forskarskolan*



[Piecesauto-Pro.FR](#)



My name is Weronika Pawlak. I am a student of the Philology faculty at the Jagiellonian University, Poland. Ever since the start of the university studies I



# PROMO D'ENFER ! ZOOBIO FAIT FONDRE TOUS LES PRIX AVEC SES 14 % DE RÉDUCTION !

06:34:31  
HEURES MINUTES SECONDES

Livraison gratuite à partir de 59.00 EUR



+33 970 733 833 Lu au Ve de 09h00 à 18h00 | Bulletin d'Information | Contact

Que cherchez-vous?



CONNEXION ▾

PANIER / 0,00 €



CHIEN

CHAT

OISEAU

PETIT ANIMAL

AQUARIOPHILIE

CHEVAL

TERRARIOPHILIE

MARQUES

**% MEILLEURES OFFRES**

## ANIMALERIE

### Boutique pour chiens

- Nourriture
- Laisses et colliers
- Jouets
- Transport & voyage
- Couchages
- Toiletage et hygiène

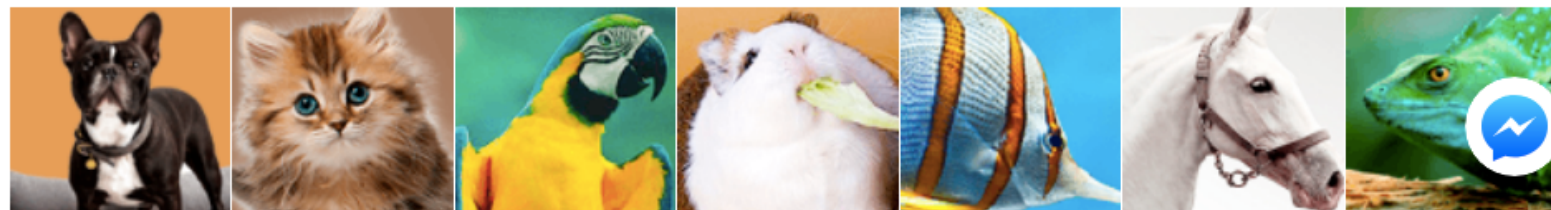
Tous >

### Boutique pour chats

- Nourriture
- Soins et hygiène
- Arbres à chats
- Jouets pour chats
- Couchages
- Gamelles et distributeurs de nourriture



## ANIMALERIE EN LIGNE POUR VOS ANIMAUX PRÉFÉRÉS



## Les meilleures Offres



Nous sommes les seuls à vous permettre de faire des économies sans compromettre la qualité du service!