

Predicting Ads' CTR with Decision Rules

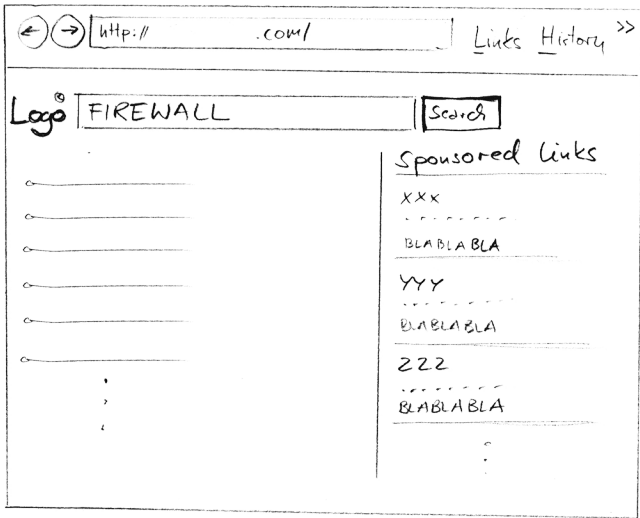
Krzysztof Dembczyński Wojciech Kotłowski Dawid Weiss

Institute of Computing Science
Poznan University of Technology
Poland

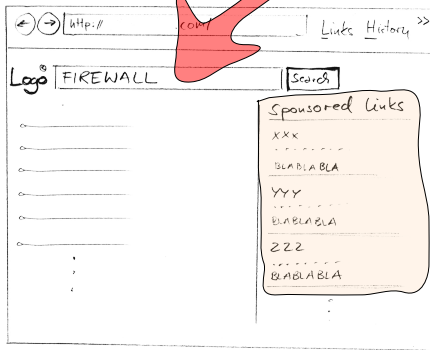


- 1 Introduction
- 2 'Beyond Search' Data Set
- 3 CTR Prediction Model
- 4 Query Intent Analysis
- 5 Summary and Conclusions

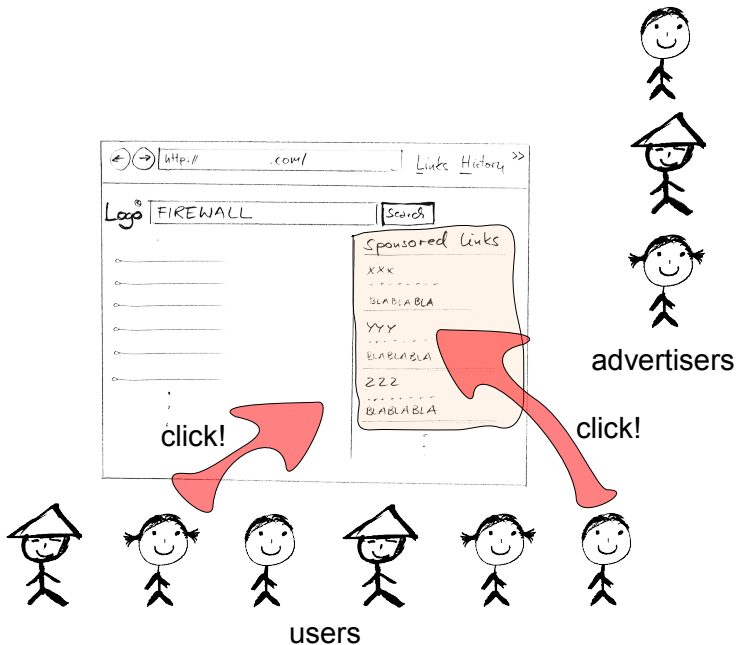
CTR, CPC, REVENUE MODEL



keyword
bid

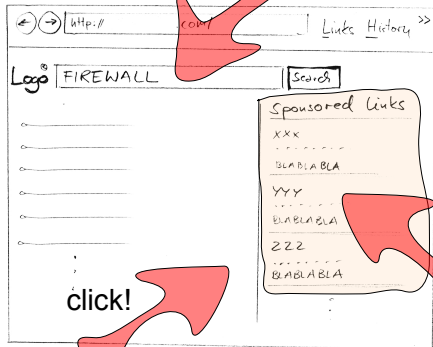


advertisers

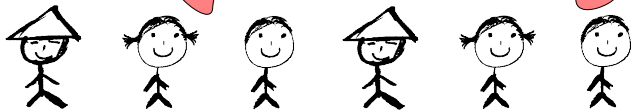




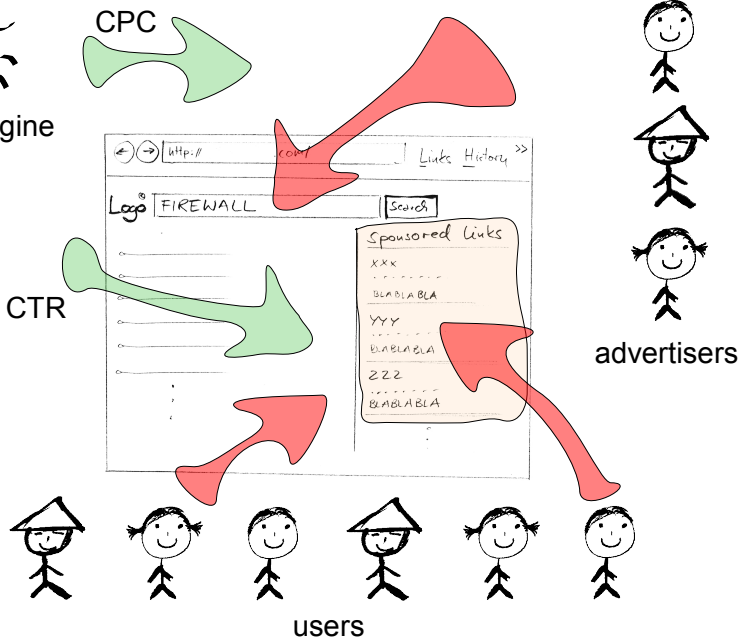
keyword
bid



advertisers



users



$$\text{REVENUE} \propto \text{CPC} \times \text{CTR}$$



search engine



... firewall ...



advertiser



search engine



computer science



CTR = 0.8

... firewall ...



civil engineering

CTR = 0.2



advertiser



search engine

o o o

- what CTR assign to this keyword?
- how to help the advertiser choose the right keywords?



computer science

←
CTR = 0.8

... firewall ...

→ *civil engineering*

CTR = 0.2



advertiser

o o o

- Is this a good keyword?
- will it attract the right customers?



search engine



... firewall ...



advertiser



decision rule
based model



historical data

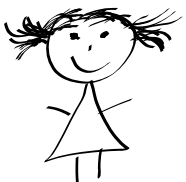


search engine

IF firewall AND doors: CTR--
IF firewall AND network: CTR++



... firewall ...



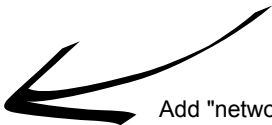
advertiser



decision rule based model



historical data



Add "network" to your ad's description

CTR PREDICTION MODEL

- handling interactions between features
- (possibly) interpretable if-then decision rules (white box)

AD IMPROVEMENT HINTS

- feedback for advertisers

- 1 Introduction
- 2 'Beyond Search' Data Set**
- 3 CTR Prediction Model
- 4 Query Intent Analysis
- 5 Summary and Conclusions

“Unfortunately, the data set [...] does not contain all the elements that have been previously used to derive features of ads [...].”



“We think some of these features could be reconstructed by issuing queries to the Live! search engine (using its API).”

“Additional features could be based on the context of original queries for which ads were displayed and clicked.”

BEYOND SEARCH DATA SET

Ads and domains

Domain	ads count
www.shopzilla.com	20032
www.bizrate.com	17592
www.pronto.com	12831
clickserve.dartsearch.net	11090
www.amazon.com	10828
clk.atdmt.com	7097
clickserve.cc-dt.com	5704
track.did-it.com	5264
...	

Ad ID	domains count
85593319	4764
1269587	713
87598825	562
5941421	491
2726279	470
442332	429
1123716	385
3789467	262
...	

Ads and domains

89316547	www.bizrate.com www.shopzilla.com	www.smarter.com, travel.nextag.com
89288811	pixel1060.everesttech.net www.firestonecompleteautocare.com	www.bizrate.com clickserve.dartsearch.net
89285594	www.calibex.com www.shopzilla.com	www.anntaylor.com search.hsn.com
89274298	www.bestonlineshopping.com www.dpbolvw.net	www.kqzyfj.com clk.atdmt.com

Ads and Queries

174544 kids pop cds
174544 Photo Fixer
174544 classes of species
174544 kent cartridge company
174544 develops following Greek models from
174544 du-all.com
174544 psychological research journal depression in women
174544 jake leevey
174544 walker north country marathon
174544 MSN Mapsdulles international airport map
174544 phenyl magnesium bromide in gridnarf synthesis
174544 coke
174544 2000 degrees heat
174544 COSHOCTON AIR SHOW
174544 adult freindfinder.com

Ads and Queries

634888	1	0	10 dollar scrubs
634888	1	0	sugarpea scrubs
634888	1	0	landue scrubs
634888	1	0	scrubs and cubs
634888	1	1	costumes scrubs
634888	1	0	Turbinado Scrubs
634888	1	0	scrub top with 5 pockets
634888	1	0	scrubs RI
634888	1	0	tulsa A.A. scrubs
634888	2	0	scrubs rx
634888	1	0	CASSANDRAO SCRUBS

Final Pruning

- Only ads associated with a single domain.
- Only ads with a minimum impression count (10, 30, 60, 200).

Final Pruning

threshold	unique ad IDs	impressions
all	463,202	110,955,561
10	373,437	110,553,933
30	279,786	108,851,279
60	211,419	105,916,306
200	101,529	93,538,958

- 1 Introduction
- 2 'Beyond Search' Data Set
- 3 CTR Prediction Model**
- 4 Query Intent Analysis
- 5 Summary and Conclusions

MODELING THE CTR

Assumption

The result of comparison of probabilities for two ads does not depend on position and results page **as long as** both position and page are **the same** for both ads:

$$\frac{P(\textit{click}|\textit{page}_1, \textit{pos}_1, \textit{ad}_1)}{P(\textit{click}|\textit{page}_1, \textit{pos}_1, \textit{ad}_2)} = \frac{P(\textit{click}|\textit{page}_2, \textit{pos}_2, \textit{ad}_1)}{P(\textit{click}|\textit{page}_2, \textit{pos}_2, \textit{ad}_2)}$$

Conclusion

Probability factorizes into the following form:

$$P(\textit{click}|\textit{page}, \textit{pos}, \textit{ad}) = f_{12}(\textit{pos}, \textit{page}) \times f_3(\textit{ad})$$

Additionally, we assume: $f_{12}(\textit{pos}, \textit{page}) = f_1(\textit{pos}) \times f_2(\textit{page})$.

Parameter estimation

Naïve method

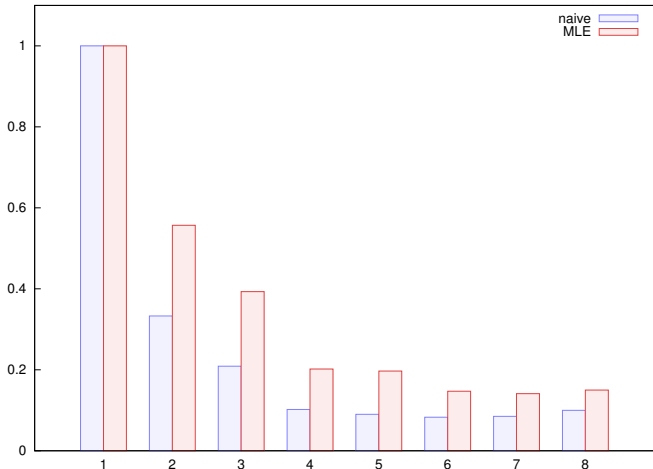
$$f_1(pos) = \frac{|\{clicked, pos\}|}{|\{pos\}|} \quad f_2(page) = \frac{|\{clicked, page\}|}{|\{page\}|}$$

Overestimates top positions and pages, underestimates the bottom ones.

Maximum Likelihood Estimation

- Expectation-maximization algorithm.
- Tests on artificial data showed high accuracy.

Position and the CTR



The likelihood of seeing the ad as a function of its position, $f_1(pos)$ for both naïve and MLE methods.

Predicting CTR

To predict the CTR we learn a function $f(x)$ using a training set $\{y_a, x_a\}_1^N$, in which each example corresponds to a single ad a :

- y_a – CTR estimated using MLE ($y_a = p_a(a)$)
- x_a – feature vector related to the quality of a (title, body, URL).

ENDER

- We treat $f(x)$ as an ensemble of decision rules
- We use the ENDER algorithm to generate the rules.
- ENDER is based on **boosting**: in each iteration, a single rule is generated by concentrating on examples which were hardest to classify correctly by previous rules.
- For CTR prediction, ENDER minimizes the squared-error loss.

Decision Rules

The main advantage of decision rules is their simple and human-interpretable form that can model interactions:

if **term = “costume”** and **term = “halloween”**
and **URL = “com”**
and **#segments in URL ≥ 3**
and **URL length ≤ 21**
then **CTR is increased by 0.017**

Number of covered ads = 87



halloween costume



ALL RESULTS

Images

Videos

Results are included for [halloween costumes](#). Show just the results for [halloween costume](#).

ALL RESULTS

1-10 of 22,600,000 results · [advanced](#)

[Unique Halloween Costumes](#) - HalloweenMart.com

Sponsored sites

Costumes For Kids & Adults - Safe Site - Fast Ship

[Halloween Costumes for Adults, Teens & Kids - Halloween Supplies : Buy](#)

...

BuyCostumes.com is the web's most popular store for **Halloween costumes**, accessories and props as well as party supplies and decorations. Visit our **Halloween costume** and party ...

[www.buycostumes.com](#) · [cached page](#)

[Halloween Costumes For Adults & Kids - Costume SuperCenter.com](#)

Question? Call us at 1-888-575-5575 Mon-Thur: 8:30am - 9:00pm | Fri: 8:30am - 6pm | Sat - Sun: 10am - 5pm

[www.costumesupercenter.com](#) · [cached page](#)

[Halloween Costumes and Costume Accessories for Adults, Teens and Kids](#)

The absolute largest selection of **Halloween costumes**, **costume** accessories, props and **Halloween** decorations available anywhere. Quick ship. Low prices. We are The **Halloween**

...

[www.halloweenexpress.com](#) · [cached page](#)

Sponsored sites

[Halloween Costumes \(Save\)](#)

Low Prices, Fast + Cheap Shipping! Shop the Latest **Halloween Costumes**.

[www.CostumeHub.com/Halloween](#)

[Halloween Costume](#)

Halloween Costumes & Accessories! Shop Today For The Best Selection!

[CostumeExpress.com](#)

[Halloween Costumes](#)

Shop Over 15,000 **Costumes** - All Sizes, Styles & Low Prices!
[BuyCostumes.com/Halloween](#)

[halloween costume](#)

Shop, Compare, Decide. Save on **halloween costume** at Bing™ cashback!

[bing.com/cashback](#)

👉 [Bing cashback](#)

RELATED SEARCHES

[Kids Halloween Costumes](#)

[Homemade Halloween Costumes](#)

[Party City](#)

[Couples Halloween Costumes](#)

[Halloween Costume Ideas](#)

[Plus Size Halloween Costumes](#)

[Disney Halloween Costumes](#)

[Funny Halloween Costumes](#)

Recommendation Rules

From decision rules we can derive **recommendation rules** to make recommendations concerning ads:

if **term = “costume”** and **term = “halloween”**
and **#segments in URL ≥ 3**
and **URL length ≤ 21**
then **URL \leftarrow “com”**

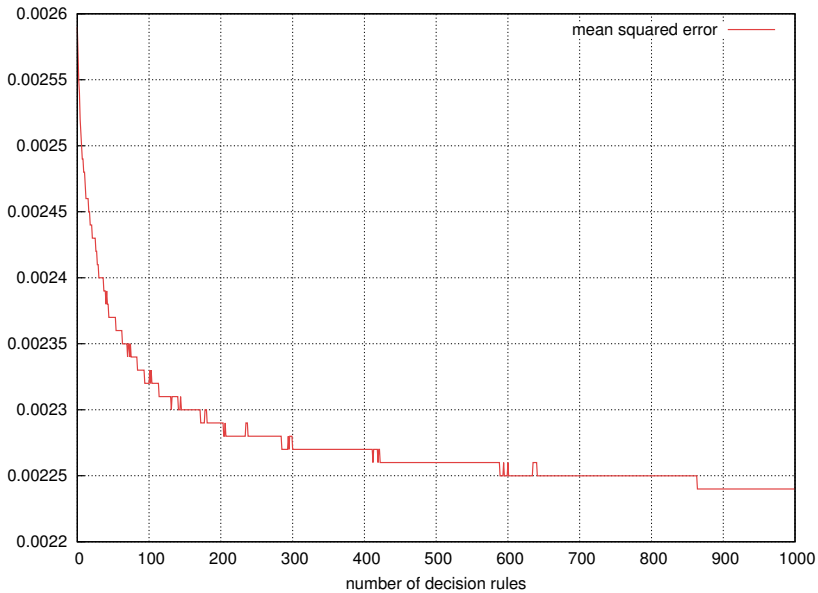
Experiment

- Single-domain ads, shown at least 200 times (101 529 ads).
- Each ad described by features related to query terms and its URL.
- We generated 1000 rules.

Results

classifier	MSE (1e-3)	improvement
Baseline	2.59	—
ENDER	2.24	13.5%

Mean squared error (MSE) on test data.



- 1 Introduction
- 2 'Beyond Search' Data Set
- 3 CTR Prediction Model
- 4 Query Intent Analysis**
- 5 Summary and Conclusions

Can we determine from the query if the user's intention is to click on the ad or browse query results?

The goal is to estimate the probability of clicking the add for a given query content.

QUERY INTENT MODEL

$$P(\textit{click}|\textit{page}, \textit{pos}, \textit{query}) \propto \exp \left\{ f_1(\textit{page}) + f_2(\textit{pos}) + f_3(\textit{query}) \right\}$$

- Fitting the model by minimization of the exponential loss (boosting).
- Functions f_1 and f_2 are non-parametric. Function f_3 is an ensemble of decision rules (ENDER).
- Each impression considered as a separate training example.

Problem

As many observations as query impressions (~ 100 million).

Solution

- 1 First estimate $f_1(\text{page})$ and $f_2(\text{pos})$ by grouping all impressions with the same page and position.
- 2 Then estimate f_3 by grouping all impressions with the same query.

Decision Rules

- The rules show the relation between keywords used in queries and the intent to click an ad:

if **keyword = “cheap”**

then **Intent to click an ad is increased by 0.7106**

Number of covered queries = 69,816



cheap



ALL RESULTS

ALL RESULTS

1-10 of 276,000,000 results · [advanced](#)

Sponsored sites

RELATED SEARCHES

[Cheap Tickets](#)

[Cheap DVD](#)

[Cheap Flights](#)

[Cheap Definition](#)

[Cheap Airline](#)

[Cheap Computer](#)

[Cheap Loans](#)

[Cheap Ford](#)

SEARCH HISTORY

[cheap](#)

[yellow pages](#)

[halloween costume](#)

[youtube](#)

[mtv](#)

[See all](#)

[Super Cheap Airfare Deals](#) - [www.bookingbuddy.com](#)

Compare **cheap** airfare from 40+ major airlines & discount travel sites!

[Airfare From \\$39](#) - [www.Travelzoo.com](#)

Find the Cheapest Airfare Now. Compare Flights from \$39!

[Cheapest Flights](#) - [www.Fly.com](#)

Save on Flights Now! Compare Fares Before You Book.

[Cheap Flights, Airline Tickets, Airfare, Hotels, Vacations, Rental Car...](#)

Find **cheap** flights and airline tickets with no booking fees at CheapTickets: **cheap** airfare, hotels, travel deals, vacations, car rentals, cruises and more.

[www.cheaptickets.com](#) · [cached page](#)

[Cheap Laptops UK Used Laptops UK Refurbished IBM Notebooks UK](#)

Cheap IT offers top refurbished UK laptops with wi-fi and windows xp installed. Massive range of discounted laptops with up to 40% off RRP, Free delivery, the UK's widest range

[www.cheap-it.com](#) · [cached page](#)

[Cheap Trick.com](#)

From BILLBOARD: "**Cheap Trick** Preps The Latest" May 29, 2009 at 9:16 am · Filed under News. **Cheap Trick** is stepping outside of the label world to release its next studio album ...

[www.cheaptrick.com](#) · 10 hours ago

Sponsored sites

[Discount Airline Tickets](#)

Great Deals on All Flights. Save up to 40% with Priceline.

[www.priceline.com](#)

[Discount Airfare](#)

Get quotes & make reservations. Save time and money!

[www.LowFares.com](#)

[Midwest Airlines](#)

Low Fares and Extra Legroom. No Booking Fees.

[www.midwestairlines.com](#)

[Secret Airfare Loopholes](#)

Fired Travel Agent Exposes Methods For You To Save Big On Airfare

[www.MyAirfareSecrets.com](#)

[Cheap Airline Tickets](#)

Super Saver Airfares International and Domestic Flights

[www.AirGorilla.com](#)

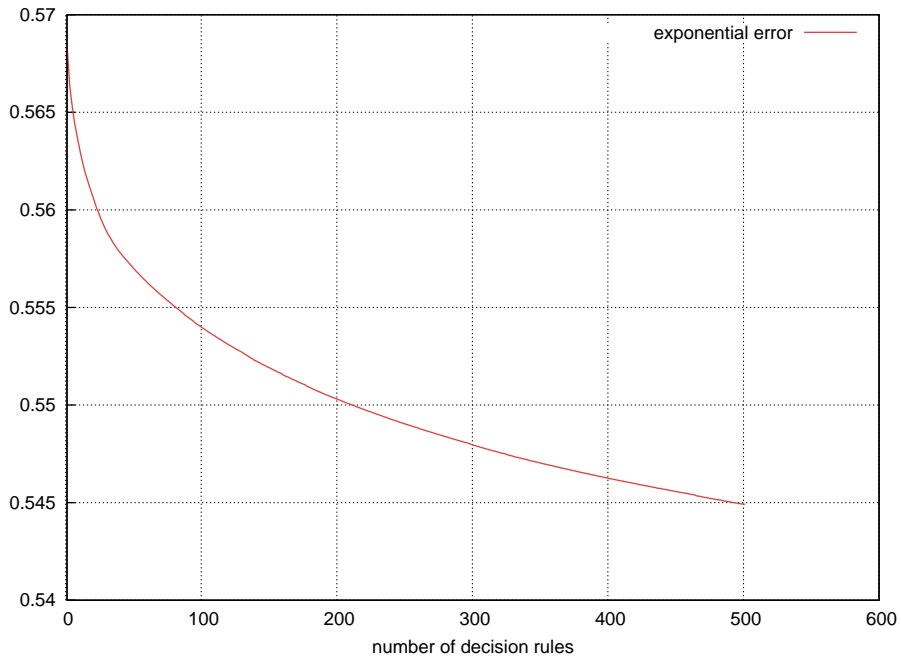
Experiment

- All the impressions (110,955,561).
- 24,879,232 distinct queries.
- Each query is treated as bag of words (we used 10,000 terms).
- We generated 500 rules.

Results

classifier	ExpLoss	improvement
Baseline	0.568	—
ENDER	0.545	4.13%

Exponential loss (ExpLoss) on test data.



Interesting Rules:

if **keyword = “yellow”** and **keyword = “pages”**
then **Intent to click an ad is increased by 0.49**

Number of covered queries = 15,846



yellow pages



ALL RESULTS

ALL RESULTS

1-10 of 185,000,000 results · [advanced](#)

[Yellow Pages Phone Book](#) - Phone.TheYellowPages.com

Sponsored sites

Search by Keyword or Business Name, Get Phone, Address & Coupons

[Yellow Pages Phone Book](#) - Get Phone, Address, Directions, Web

Search Over 16 Million Businesses, Get Phone, Address, Directions, Web

RELATED SEARCHES

[White Pages](#)

[Phone Directory](#)

[MSN Yellow Pages](#)

[Verizon Yellow Pages](#)

[Real Yellow Pages](#)

[Yahoo Yellow Pages](#)

[YellowBook](#)

[Yellow Pages Dex](#)

Best match

[Yellow Pages Local Directory - YELLOWPAGES.COM](#)

[www.yellowpages.com](#) · official site

Find online **Yellow Pages** business listings, phone numbers, addresses, maps, driving directions and more in the YELLOWPAGES.COM online directory.

[People Search](#)
[Business Search](#)
[Maps & Directions](#)
[Mobile](#)

[Local](#)
[About](#)
[Contact](#)
[Careers](#)

Similar to this

[White Pages](#)

[Yellow](#)

[SuperPages](#)

[YellowPages.ca](#)

SEARCH HISTORY

[yellow pages](#)

[halloween costume](#)

[youtube](#)

[mtv](#)

[credit union](#)

[See all](#)

[Yellow Pages Local Directory - YELLOWPAGES.COM](#)

Careers; Browse **Yellow Pages**; About YELLOWPAGES.COM; Contact Us; Site Map; Help; Advertise With Us; AT&T; AT&T Wireless; att.net; TollFreeDA.com; AnyWho; Ingenio **Yellow Pages**

[www.yellowpages.com/?From=Branding_yprnd_yellow+pages](#) · [cached page](#)

Interesting Rules:

if **keyword = “youtube”**
then **Intent to click an ad is decreased by -1.04**

Number of covered queries = 26,436



youtube



YOUTUBE

Videos

Categories

Downloader

Shows

Community

Proxy

xRank

ALL RESULTS

1-20 of 57,600,000 results · [advanced](#)

Best match

[YouTube - Broadcast Yourself](#)

www.youtube.com · official site

YouTube is a video sharing website where users can upload, view and share video clips.

Sign In

Most Viewed

Videos

Favorites

Subscriptions

Sign Up

Upload

Channels

Similar to this

[MySpace](#)

[Facebook](#)

[Photobucket](#)

[Wikipedia](#)

Search within youtube.com

Search

RELATED SEARCHES

[You Tube Video Clips](#)

[YouTube Music](#)

[YouTube MySpace](#)

[YouTube Japan](#)

[Music Videos](#)

[YouTube Seattle](#)

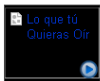
[YouTube Technology](#)

[YouTube Marketing](#)

[Videos of youtube](#)



Gregg Pritchard - Singer - Britains...
YouTube 5:55



Lo que tú Quieras Oír
YouTube 10:14



Akon - I'm So Paid...
YouTube 4:37



Natalie Okri - Britain's
YouTube 2:44

- 1 Introduction
- 2 'Beyond Search' Data Set
- 3 CTR Prediction Model
- 4 Query Intent Analysis
- 5 Summary and Conclusions**

Summary

- Rule-based model for CTR prediction.
- Rule-based model for intent analysis.

„The quest for assets.”

- Insufficient ad description data.
- Unclear ad ID – ad description relationship.