

# **An Investigative Search Engine for the Human Trafficking Domain**

**Mayank Kejriwal and Pedro Szekely**

**USC Information Sciences Institute**

**October 2017**

# Investigative Queries

What is the **ethnicity** listed in the **ad** that contains the **phone number 6135019502**, **in Toronto Ontario**, with the **title 'the millionaires mistress'**?

```
SELECT ?ad ?ethnicity WHERE {  
  ?ad a :Ad ;  
    :phone '6135019502' ;  
    :location 'Toronto, Ontario' ;  
    :title 'the millionaires  
mistress' ;  
    :ethnicity ?ethnicity . }
```

What is the ethnicity listed in the ad that contains the phone number 6135019502, in Toronto Ontario, with the title 'the millionaires mistress'?

Sri Lankan  
Black  
Caramel

The screenshot shows a profile page on EROS.COM. The profile title is "TS THE Millionaires MISTRESS" and the location is "Trans in Toronto". The profile includes a bio, a list of interests, and a sidebar with personal details. Red arrows point to the profile title, the location, and the "View My Number" button. Green arrows point to the "ETHNICITY" field in the sidebar, the bio text, and the "View My Email" button.

EROS.COM  
TRANSEXUALS

ESCORTS MAP LIVE CAMS LOCAL SINGLES MEET TRANSEXUALS CAMSODA POST AD NEWSLETTER Q SEARCH SEXART

Home / Toronto / What's New / TS THE Millionaires MISTRESS

Previous Next

## TS THE Millionaires MISTRESS

Trans in Toronto

Hi I'am tonya and am glad to announce my residence to this beautiful city.

- I live in the yorkville area of downtown

Thanks for taking some time out to read a lil bit about me! As far as statistics go...

I was born on the beautiful island of JAMAICA of Black & Sri Lankan decent. I've been blessed. At 5'10, I'm neither short nor too tall. At 158lbs, I'm slender, but not skinny. With the all natural-measurements of 22 waist and 34 hips, I'm curvy, but far from voluptuous.

- I am the perfect girl with a special secret!
- Well educated and well travel.

I'm in my early 20's and happily independent, as well as stretch mark-free and cellulite-free.

My caramel-hued skin is just as flawless as you see in my photos... But feel free to peel off my clothes and inspect for yourself! Outgoing, friendly (at first) and open-minded, I'm a true extrovert in every sense of the term.

The words timid, shy and meek have NEVER been used to describe me. I'm always the life of the Party, regarding everyone I encounter as a friend, or a friend I've yet to meet.

So it comes as little suprise that I now find myself as a personal paramour. The opportunity to meet new, fascinating people and connect with them on multiple levels gives me an incomparable thrill. Under the right circumstances, I go from delightfully dirty, to downright devilish in a matter of seconds.

I invite you to experience the transition...you'll be glad you did...as will I... Screening is required, but references are not. A detailed and thorough email/PM goes a long way, so don't be afraid to introduce yourself.

I couldve been anything I wanted to be -she is well educated- instead I chose a more exhilarating path for myself and why not utilize my good looks before they fade?

I'm open to just about anything. Give me a clear idea of who you are and what you want, and I know we'll get along famously. Ciao! Formalities

GENDER	Transsexual
AGE	24
ETHNICITY	Black/Ebony, Asian
HAIR COLOR	Blonde
EYE COLOR	Brown
HEIGHT	5'10"
WEIGHT	158 lb
MEASUREMENTS	36D - 24" - 34"
AFFILIATION	Independent
AVAILABILITY	Incall / Outcall
AVAILABLE TO	Groups, Men, Women, Couples

TS THE Millionaires MISTRESS

View My Number

View My Email

Connect

# Query Challenges

```
SELECT ?ad WHERE
{
  ?ad a :Ad ;
  :hair_color 'Auburn' ;
  :review_site_id 'cg9469f'
;
  :price_per_hour '500' ;
  :name 'Claire Gold' ;
  :ethnicity 'Asian'.
}
```

synonyms "red"  
typos "brunette"

not present

numbers hard to match

Claire is a common name  
Gold is a domain word

slang, e.g., "FOB" for Asian  
inference, e.g., "Japanese"

# Cluster and Aggregate Queries

Find the **ads** that provide **outcall service** in the **network** for **phone number 75924833**

Find the **minimum price** in the **network** of **ads** for **phone number 75924833**

# Outline

**Challenges**

**Approach**

**Results**

# Challenges

## Information extraction

Many websites, different page templates  
Purposeful obfuscation  
Unusual language model

## Easy to use, robust & expressive queries

Users are investigators in law enforcement offices  
Investigators need current and historical data

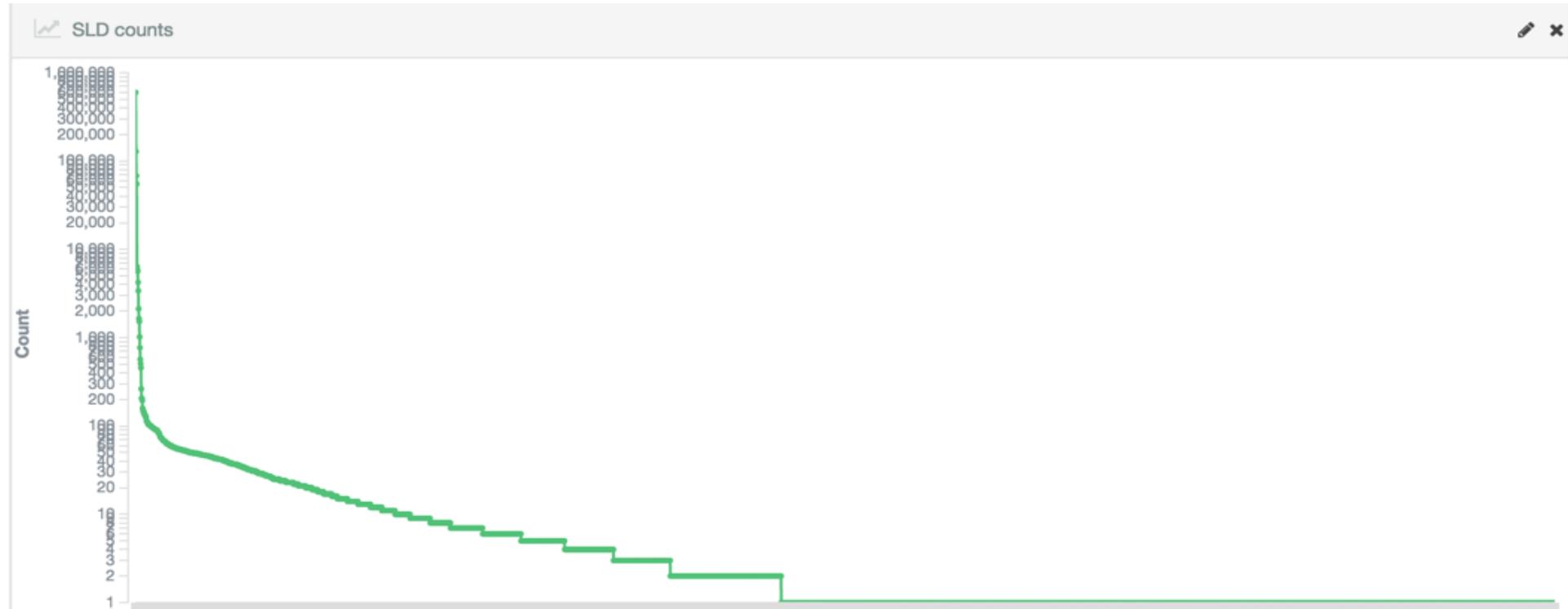
## Provenance

Information extraction is noisy  
Investigators need to verify information against original sources

## Scale

120 million pages covering 3 years

# Heterogeneous Web Pages



~1 Million Pages and 4,354 Second Level Domains

# Unusual Language Model

Italian 19 hello guys....My name is charlotte , New to town from kansas

[ GORGIOUS BLONDE beauty] ? FROM Florida ? (Petite) ? [ CURVy ]?

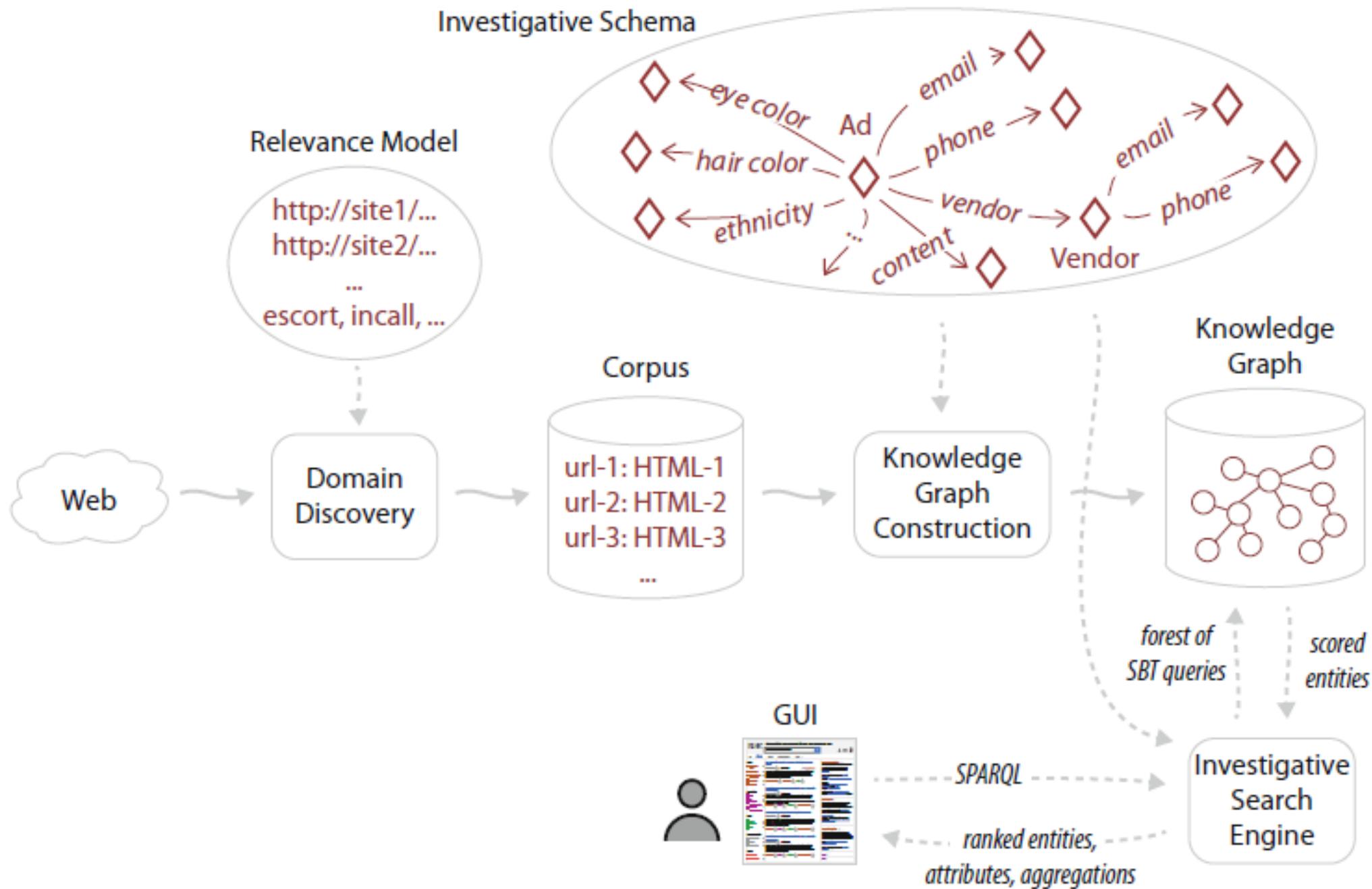
NO DISAPPOINTMENTS. 34C..Brazilian,ITALIAN beauty....

Hey gentleman im Newyork and i'm looking for generous

Hi guy's this is sexy newyork . & ready to party.

AVAILABLE NOW! ?? - (1 two 1) six 5 six - 0 9 one 2 - 21

# Approach



# **Step 1: Knowledge Graph Construction**

# Easy To Use Extractors

## Glossary-based

User-provided lists of words/phrases

Uses: names, cities, states, countries, ethnicities, eye/hair colors, ...

## Semi-supervised HTML template-based

Identifies page templates and extracts content

Uses: 30 most important web sites

## Rule-based

Simple patterns of tokens

Uses: phones, review ids, social media names, person names, age, ...

## Custom extractors

Uses: de-obfuscating phone and email extractor

# Unsupervised Extractor Generator

**Text** → ALLOW YOURSELF TO GET LOST IN ME \*\* TRUE SUBMISSIVE \*\*  
VERIFIED PHOTOS \*\* REVIEWS \*\* OUTCALL ONLY

**Name** → Created by [diamondnelson](#) on Monday, August 1st, 2016 in San Francisco Escorts • PostID 6863205

**PostingDate** →

**Location** →

**PostID** →

**HyperLink** → [Review: theeroticreview.com/reviews/show.asp?id=302542](#)

**Text** → The working men, I'll go by and they'll whistle. At first they whistle because they think, 'Oh, it's a girl. She's got blonde hair and pretty blue eyes,' and then they say, 'Oh my Gosh, it's Diamond Nelson!

**Phone** → **Phone** 415-813-8165

**Email** → **Email** [diamondnelsonxoxo@gmail.com](mailto:diamondnelsonxoxo@gmail.com)

**HyperLink** → **Website** [www.goddessdiamondnelson.com](http://www.goddessdiamondnelson.com)

Visit my website to for book request, donations, and to see my private photo gallery: [www.goddessdiamondnelson.com](http://www.goddessdiamondnelson.com).

Please give me one hour advance notice when booking your appointment! No explicit talk or else the call will be immediately terminated.

Diamond Nelson 415-813-8165

All donations are for my time and companionship only. Legally, I cannot and will not discuss or agree to any type of solicitation. No fees or tips will be quoted, negotiated, assessed, or collected in exchange for sexual conduct of any kind. Anv at tempt to compromise mv position on these

**QR Code**

**AreaCode** 415

**Incall Map**



**Locations** [Belmont](#), [Brisbane](#), [Burlingame](#) ([View All...](#))

**Preferred411** <https://www.theeroticreview.com/revie...>

# Rule-Based Extractor

**Create Word Token**

optional  part of output  match lemma  alphanumeric

**Words:**  
Enter words here.

**Part of speech:**

- noun
- pronoun
- proper noun
- determiner
- symbol
- adjective
- conjunction
- verb
- pre/post-position
- adverb
- particle
- interjection

**Capitalization:**

- exact
- lower
- upper

Length 1:  Length 2:  Length 3:

Prefix:  Suffix:   ignore  
vocabulary

**Create Punctuation Token**

optional  part of output

**Punctuation Symbols:**

- ,
- !
- <
- >
- =
- %
- \
- /
- \*
- \$
- @
- (
- )
- [
- ]
- {
- }
- ~
- ^
- \_
- `
- ~
- ^
- \_
- `

**Create Number Token**

optional  part of output

**Numbers:**

Length 1:  Length 2:

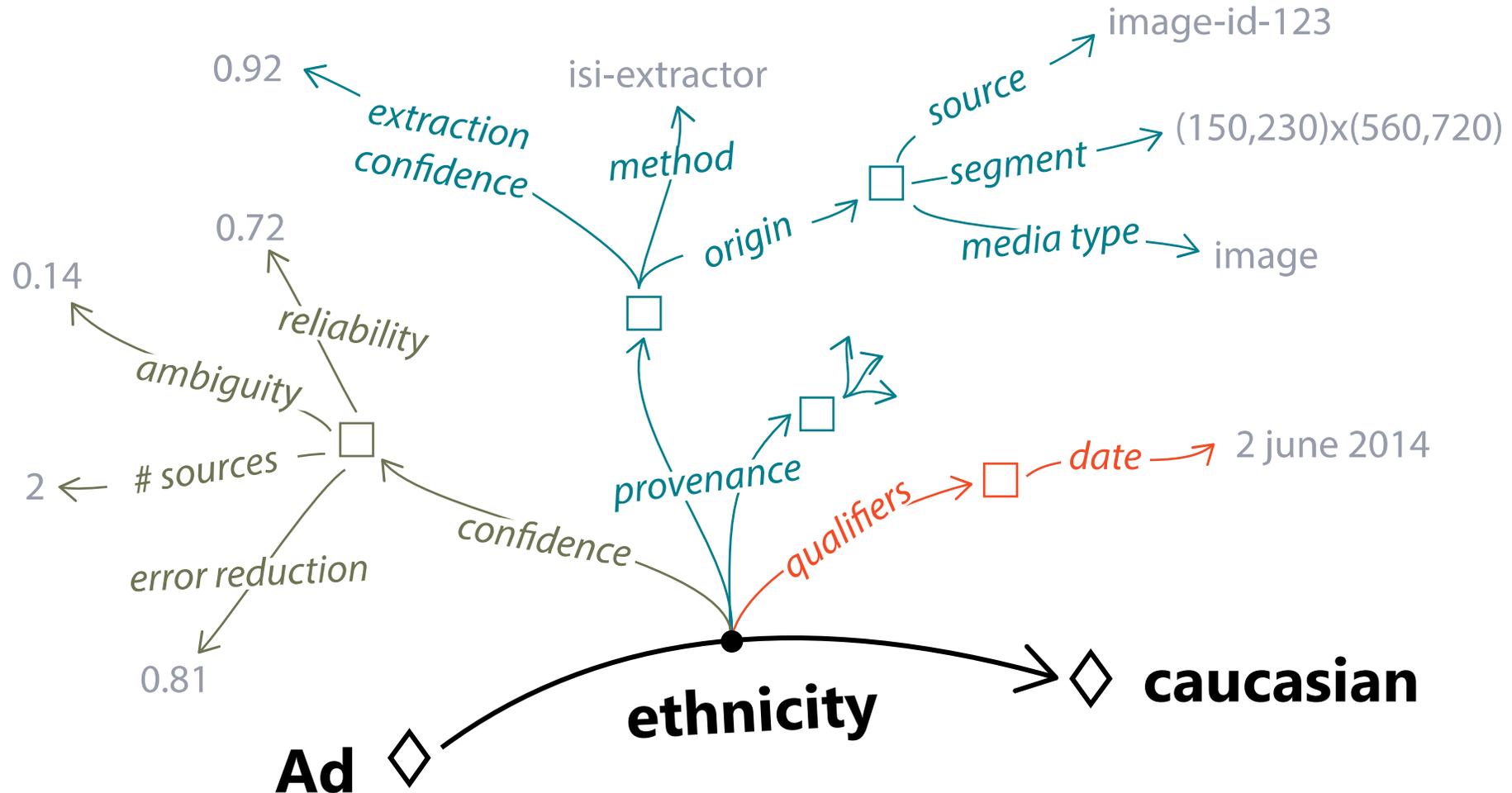
Length 3:

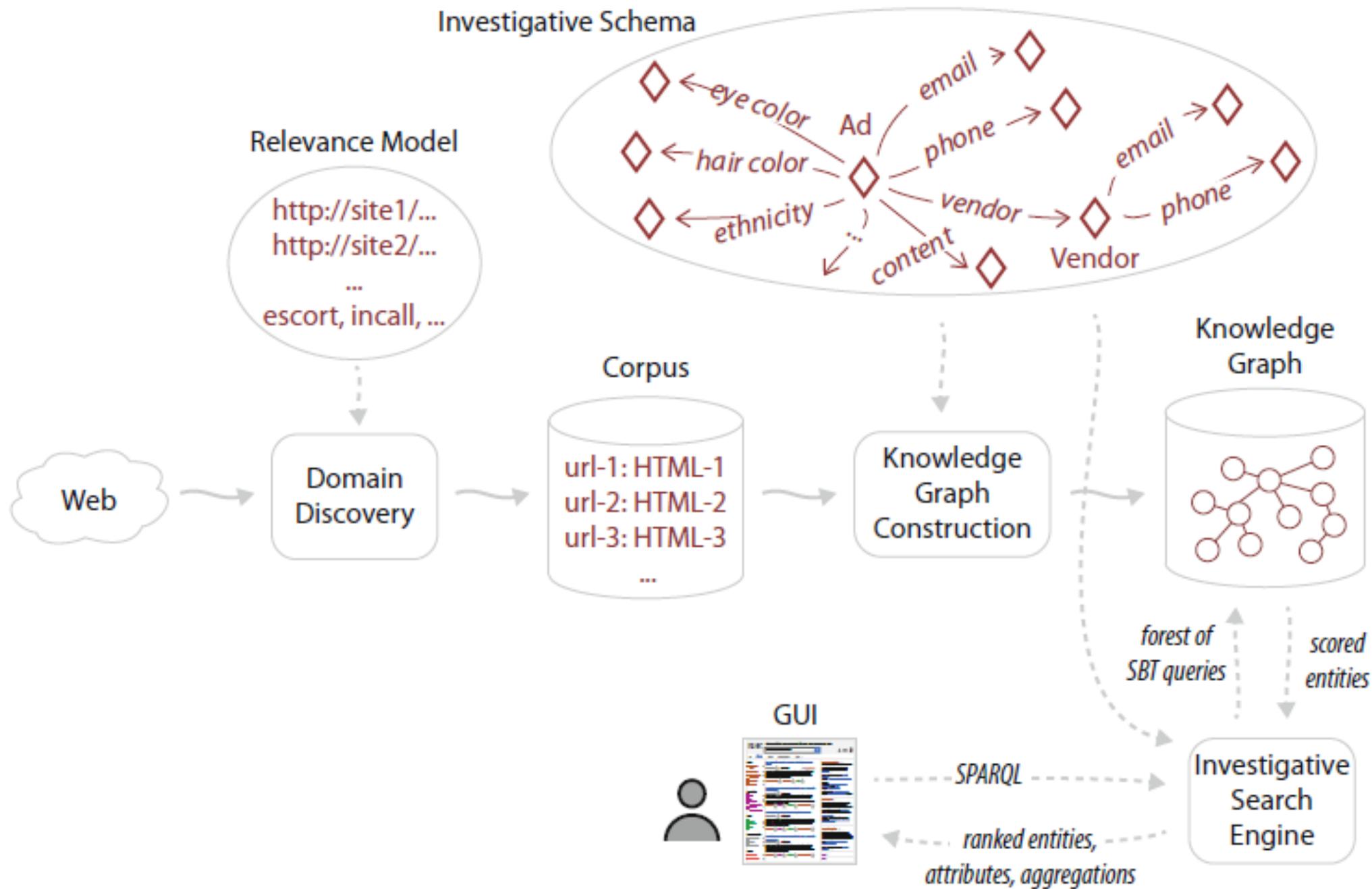
Min:  Max:

cancel Save

# KG Representation

Entities + properties + provenance + confidence + qualifiers

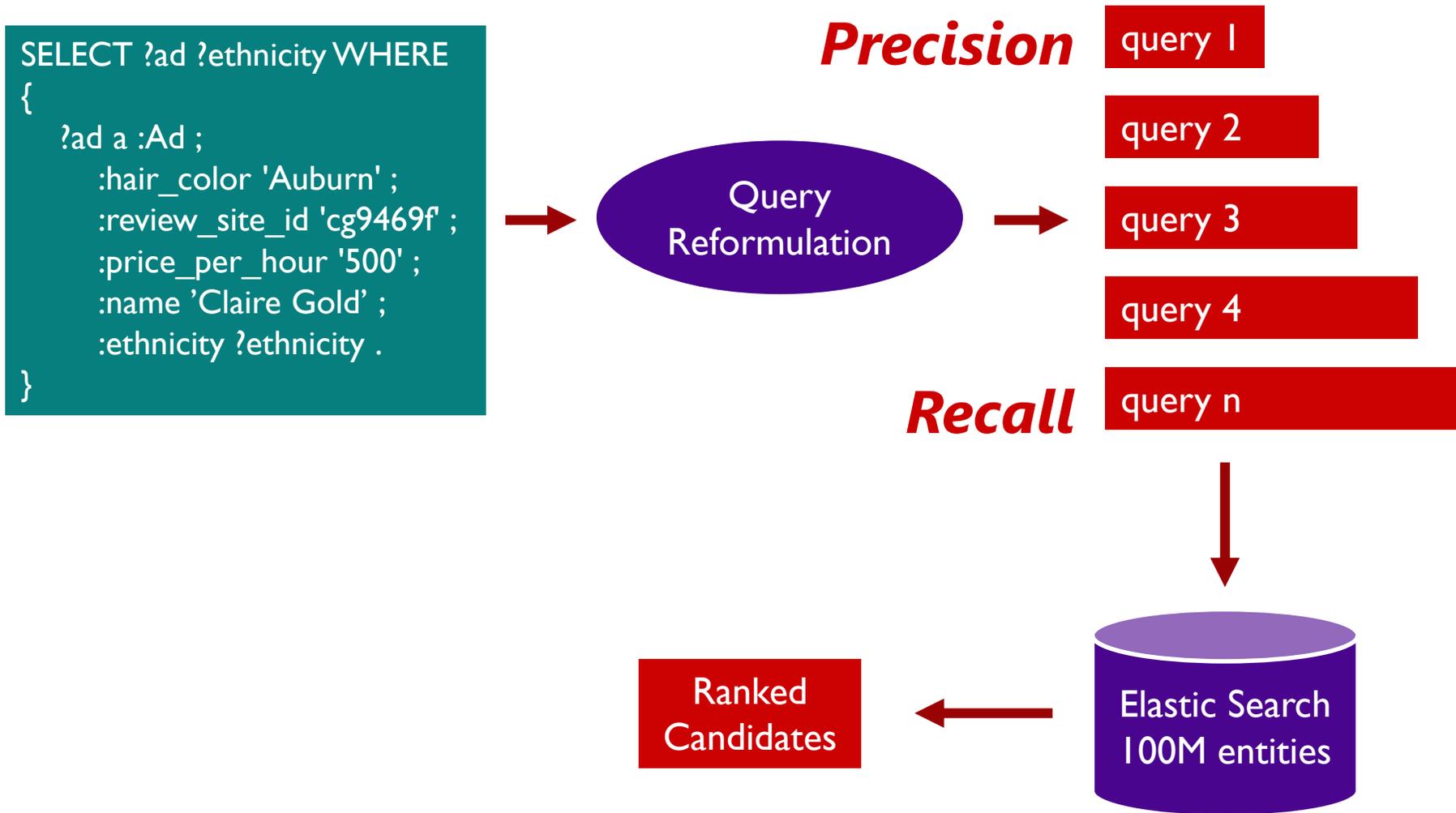




# **Step 2: Information Retrieval on Knowledge Graphs**

# Candidate Generation

Keyword expansion • Context broadening • Constraint relaxation



Search Terms

\* Web Domain (TLD): backpage

Facets

Telephone Number (Top 10)

View More

- 714-980-0067
- 714-650-7980
- 905-990-2533
- 212-206-0200
- 212-756-8800
- 647-767-5240
- 800-538-6740
- 302-543-8385
- 416-523-0044
- 401409103

Email Address (Top 10)

View More

- mikepromotions@optusnet.c
- astrid6060@hotmail.com
- jsres44@yahoo.ca
- sexysamanthafoxxx@gmail.c
- contact.rosedegrays@gmail.c
- tamara-...
- ankitankit0370@gmail.com
- franchisescaanne.marie@yaho
- adrianovalentino@hotmail.co
- supermatchescort@gmail.co

Review ID (Top 10)

View More

Sort By A-Z

Export

SEARCH

RESET

CANCEL

HELP

IMAGES

DATE POSTED	Date Posted Begin 10/5/2017	Date Posted End 10/24/2017
IDENTIFIER	Telephone Number	Email Address
	Review ID	Social Media ID
LOCATION	City dallas, houston	State/Region
	Country	
PROVIDER	Provider Age	Provider Ethnicity asian, caucasian
	Provider Eye Color	Provider Gender
	Provider Hair Color red	Provider Height
	Provider Name jessica, ashley	Provider Price
	Provider Weight	Service Provided

Enter or Paste Image Links

UPLOAD FILES

Drop files here...

SHOW 25 MORE WEBPAGE RESULTS

Search Terms

- Page: unrushed X City: chicago X
- Services Provided: fetish friendly X
- Ethnicity of Provider: hispanic X

Price of Provider (Top 10)

View More Sort By: [icon] [icon]

Select All 10 Deselect All

<input type="checkbox"/> 60 \$ per hour	260
<input type="checkbox"/> 200 \$ per hour	124
<input type="checkbox"/> 180 \$ per hour	114
<input type="checkbox"/> 150 \$ per hour	73
<input type="checkbox"/> 150 per hh	61
<input type="checkbox"/> 80 per qv	50
<input type="checkbox"/> 160 \$ per hour	46
<input type="checkbox"/> 120 per hh	44
<input type="checkbox"/> 130 per hh	33
<input type="checkbox"/> 60 per qv	30

Website (8)

View More Sort By: [icon] [icon]

Select All 8 Deselect All

<input type="checkbox"/> eroticmugshots.com	3,927
<input type="checkbox"/> backpage.com	74
<input type="checkbox"/> escortsincollege.com	9
<input type="checkbox"/> escortphonelist.com	8
<input type="checkbox"/> classivox.com	2
<input type="checkbox"/> adultsearch.com	1
<input type="checkbox"/> escortsinthe.us	1
<input type="checkbox"/> massagetroll.com	1

High Risk

None

25 of 46,190,422 Results [How are search results found?](#)

Export

2.06 100hh Special ?H?t C??a? ? P?a???? ? ?? R?s? 100% W??t? T??  
 V?s?t - **Chicago** escorts - backpage.com

Oct 6, 2015

backpage.com

Locations:

**chicago**

Telephone Numbers:

323-██████████

No Email Addresses

No Social Media IDs

No Review IDs

Provider Names:

raina

Url: <http://chicago.backpage.com/FemaleEscorts/100hh-special-ht-ca-pa-rs-100-wt-t-vst/2-██████████>

Description:

Hey Freaky Boys Im Raina (Cuban & Colombian Doll ?100% REAL PICS? ?INDEPENDENT? ? **UNRUSHED** & DISCREET? ?420PARTY/PARTYFRIENDLY with Donation? ? **FETISH FRIENDLY** 5'6 129 36DD ???5 STAR EXPERIENCE?? Ok TEXTING? ?PRIVATE CALLS? ?NO THUGS Or Law Enforcement? Sick Of The BORING LAZY NOT OPEN-Minded Girl Than CALL ME NOW 323-██████████ Raina

Cached Ad Webpage: [Open](#) [What is a cached webpage?](#)

Services Provided:

**fetish friendly**

Prices:

**100 per hh**

Provider Ages:

**20**

Provider Ethnicities:

**hispanic**

No Provider Eye Colors

No Provider Hair Colors

Provider Heights:

**167**

No Provider Weights

12 Images

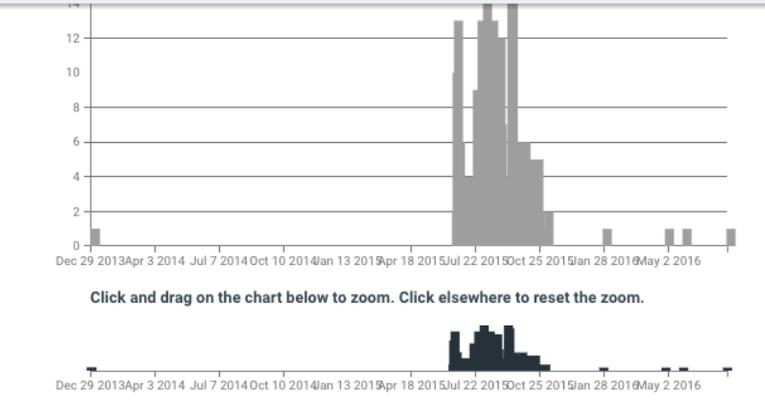


SHOW 25 MORE RESULTS

### 📞 7 Other Telephone Numbers 📄 Copy

PHONE	ADS CO-OCCURRING WITH 323-800-6151	ADS NOT CO-OCCURRING
<input type="checkbox"/> 971-██████████	20	491
<input type="checkbox"/> 720-██████████	17	620
<input type="checkbox"/> 209-██████████	11	920
<input type="checkbox"/> 928-██████████	11	551
<input type="checkbox"/> 916-██████████	9	132
<input type="checkbox"/> 510-██████████	7	139
<input type="checkbox"/> 720-██████████	3	1,397

[? What does this chart represent?](#)



### ✉️ 2 Email Addresses 📄 Copy

EMAIL	ADS CO-OCCURRING WITH 323-800-6151	ADS NOT CO-OCCURRING
<input type="checkbox"/> ██████████@gmail.com	12	290
<input type="checkbox"/> ██████████@gmail.com	1	69

[? What does this chart represent?](#)

### 📍 Location Drops Timeline

### 🌐 8 Websites 📄 Copy

WEBSITE	ADS
<input type="checkbox"/> liveescortreviews.com	220
<input type="checkbox"/> backpage.com	120
<input type="checkbox"/> eroticmugshots.com	66
<input type="checkbox"/> escorttads.xxx	7
<input type="checkbox"/> escortphonest.com	7
<input type="checkbox"/> escortsincollege.com	7
<input type="checkbox"/> escortsintexas.us	7
<input type="checkbox"/> classivox.com	1

### 📄 25 of 435 Ads 📄 Export

**100 special ? S??sat???a? C??a? ? B?? B??s ?P?????t**  
**C????s ? 2Girls Avail 100% W??t? T?? V?s?t - Indianapolis escorts - backpage.com**

📅 Oct 18, 2015      🌐 backpage.com  
 Locations: No Telephone Numbers      🚩 📄  
 📍 indianapolis  
 No Email Addresses      No Social Media IDs  
 No Review IDs      Provider Names:  
 🟢 raina

**100 special ? S??sat???a? C??a? ? B?? B??s ?P?????t**  
**C????s ? 2Girls Avail 100% W??t? T?? V?s?t - Indiana escorts - backpage.com**

📅 Oct 18, 2015      🌐 backpage.com  
 Locations: No Telephone Numbers      🚩 📄  
 📍 indiana  
 No Email Addresses      No Social Media IDs  
 No Review IDs      Provider Names:  
 🟢 raina

### 🚫 No Social Media IDs

SOCIAL MEDIA ID  
None

### ★ No Review IDs

REVIEW ID  
None

**? S??sat???a? C??a? ? B?? B??s ?P?????t C????s ? P?**  
**a???? 100% W??t? T?? V?s?t - Indianapolis escorts - backpage.com**

# **In-Use Impact**

**100 million** escort ads

**3 years** data coverage

**2 billion** triples

**100** law enforcement  
offices

**3** convictions

Water main break snarls traffic near Kennedy Airport



New York's rent freeze is over



Key witness in NYPD corruption probe was an alleged loan...

## Pimp gets 7 to 14 years in prison for peddling teen prostitutes

By Rebecca Rosenberg

September 15, 2015 | 1:38pm



Froilan Rosado (left) and his lawyer in court on Tuesday.

**7-TO-14 YEARS**

## Sex trafficker sentenced after woman escapes from 6-story window

POSTED 5:59 PM, SEPTEMBER 3, 2014, BY JEREMY TANNER, UPDATED AT 06:38PM, SEPTEMBER 3, 2014

FACEBOOK 26 TWITTER TUMBLR REDDIT PINTEREST

This is an archived article and the information in the article may be outdated. Please look at the time stamp on the story to see when it was last updated.



COFFEECAST  
56°  
Mist  
Find Your Location

YOU MAY LIKE



**50-TO-LIFE**

TIMBUK2  
SAN FRANCISCO



\$219



\$118



## Man sentenced to 97 years in human trafficking case

By Vivian Ho Updated 7:07 am, Friday, April 22, 2016

Facebook Twitter Pinterest Reddit Google+

123



**97-TO-LIFE**