Domain-Specific Insight Graphs (DIG)

Pedro Szekely

May 2017

University of Southern California
Use the web to answer investigative questions
Use Case: Human Trafficking

100 million pages
>5,000 Web sites

help victims & prosecute traffickers
Investigating a Reported Victim

San Diego, where else?

Jessica Sweden Independent San Diego Fiery Red Escort
san-diego.date-check.com/Ashow-escort-pub.asp?ProID=js2101f
Nice to meet you. My name is Jessica. I am from Sweden currently living in San Diego. I am 33 years old, fiery red hair and ocean blue eyes. Petite 130 pounds ...

Highly reviewed Swedish redhead. Upscale incall. - San ...
sandiego.backpage.com › ... › San Diego escorts › Backpage ›
Aug 6, 2013 - My name is Jessica. I am from Sweden currently living in San Diego. I am 34 years old, fiery red hair and ocean blue eyes. Petite 130 pounds ...

Jessica Swedish GFE | San Diego Escort Massage Directory
sandiego.sugarnights.com/escorts/jessica-swedish-gfe
I am from Sweden, currently living in San Diego. 33 years old, fiery red hair and ocean blue eyes. Petite 130 pounds and 5’4” tall with nice postioned D breasts.

San Diego, where else?
Locations Where A Potential Victim Was Advertised

Phone: 619-319-7315

13 of 15 Total Ads

No Other Telephone Numbers

1 Email Address

Summary of Provider Attributes

<table>
<thead>
<tr>
<th>NAME</th>
<th>Ads</th>
<th>Ads Height (cm)</th>
<th>Ads Weight (kg)</th>
<th>Ads Hair Color</th>
<th>Ads Ethnicity</th>
<th>Ads Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jessica</td>
<td>265</td>
<td>159</td>
<td>36</td>
<td>2</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>Eve</td>
<td>238</td>
<td>159</td>
<td>36</td>
<td>2</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>Swedehfgf</td>
<td>190</td>
<td>159</td>
<td>36</td>
<td>2</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>Emily</td>
<td>190</td>
<td>159</td>
<td>36</td>
<td>2</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>Sarah</td>
<td>190</td>
<td>159</td>
<td>36</td>
<td>2</td>
<td>26</td>
<td>3</td>
</tr>
</tbody>
</table>

LOCATION

- San Diego, California (779)
- San Marcos, California (155)
- Orange County, California (132)
- Los Angeles, California (31)
- Chicago, Illinois (18)
- Washington, District of Columbia (8)
- Manhattan, New York (1)

Location Drops Timeline

<table>
<thead>
<tr>
<th>Date</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>31 October</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28 May</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

San Diego: Something HOT when you are cold? - San Diego woman seeking men personal. - backpage.com
San Diego, California, United States
May 28, 2017 - backpage.com
DIG Technology

raw • messy • disconnected
hard to query, analyze & visualize

clean • organized • linked
easy to query, analyze & visualize
Steps To Build a DIG

Data Acquisition → Information Extraction → Schema Alignment → Entity Resolution → Graph Construction → User Interface

Elastic Search → Graph DB

schema.org, geonames
Data Acquisition

batch ◆ real-time

Web pages ◆ Web service
database ◆ CSV ◆ Excel
XML ◆ JSON
Information Extraction

Text
Web pages
Web tables
Images
PDF
“YOU don't wanna miss out on ME :) Perfect lil booty Green eyes Long curly black hair Im a Irish, Armenian and Filipino mixed princess :) ❤️ Kim ❤️ 707-7two7-7four77 ❤️ HH 80 roses ❤️ Hour 120 roses ❤️ 15 mins 60 roses”

name: Kim
eye-color: green
hair-color: black
phone: 707-727-7477
rate: $60/15min $80/30min $120/60min
# FOR SALE: STOEGER M3500

<table>
<thead>
<tr>
<th>Price:</th>
<th>$500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seller:</td>
<td>Private Party</td>
</tr>
<tr>
<td>Account:</td>
<td>Registered on 5/9/2013</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Listed On:</th>
<th>Thursday, September 17, 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listed In:</td>
<td>Shotguns</td>
</tr>
<tr>
<td>Location:</td>
<td>Keenesburg, Denver, Colorado</td>
</tr>
<tr>
<td>Shipping:</td>
<td>No</td>
</tr>
</tbody>
</table>

- Manufacturer: Stoeger
- Caliber: 12 Gauge
- Action: Semi-automatic
- Firearm Type: Shotgun

I have a Stoeger m3500. It is a year old. It has 200 rounds through it from clay shooting. Its in perfect condition. If you have any questions email or text me 9703427061. I'm asking 500
<table>
<thead>
<tr>
<th>Add Page</th>
<th>page_2.html</th>
<th>page_3.html</th>
<th>page_4.html</th>
<th>page_5.html</th>
</tr>
</thead>
<tbody>
<tr>
<td>0036</td>
<td>Winchester model 140</td>
<td>Beretta hand gun for sale</td>
<td>1956 Mossin nagant</td>
<td>Ruger New Model Single-Six 9 1/2##8243;</td>
</tr>
<tr>
<td>Code1354</td>
<td>37122</td>
<td>37201</td>
<td>37122</td>
<td>37211</td>
</tr>
<tr>
<td>Description1392</td>
<td>Winchester model 140 12 guage semiauto. 28inch ribbed and vented barrel with modified choke. Gun fires and cycles well. Text for pics.</td>
<td>9mm Beretta hand gun for sale at a very good price with a delivery to any interested buyer in the state. contact me at <a href="mailto:silven2016@yandex.com">silven2016@yandex.com</a> or text me at 1956 Mossin nagant. 7.62##215;54 bolt action Russian war rifle. Missing bayonet but rifle fires and cycles well.</td>
<td>Blue, 9½” barrel, original walnut grips, NcStar 4x32E scope/ mount + lighted reticle and lens covers, 2 1/2 lb trigger (professional job), PRO-TECH OUTDOORS cordura shoulder/belt holster, original plastic box/lock &amp; key/.22LR cylinder/factory test envelope w/ fired casing/ instruction manual/original rear sight/ original trigger spring/protective gun sleeve. Everything appears as new except slight .22WRM cylinder spin mark with only 1 to 1½ 50-round boxes fired. .22LR cylinder never used. Scope is set for 50 yds. using Hornady 45 gr FTX .22WMR ammo. Total package f.o.b. Nashville, TN, $539. Must have TN driver’s license and/ or TN CCW permit, or delivery to FFL required with photo of Driver’s License and cleared payment.</td>
<td></td>
</tr>
<tr>
<td>Expires1366</td>
<td>45 days, 23 hours</td>
<td>This ad has expired</td>
<td>46 days, 4 hours</td>
<td>89 days, 7 hours</td>
</tr>
<tr>
<td>Facebook2081</td>
<td>2015</td>
<td>2015</td>
<td>2015</td>
<td>2016</td>
</tr>
<tr>
<td>Firearms1276</td>
<td>raquo; Shotguns »</td>
<td>raquo; Pistols » Semi-Auto » Beretta hand gun for sale</td>
<td>raquo; Rifles » Bolt Action » 1956 Mossin nagant</td>
<td>raquo; Pistols » Revolver » Ruger New Model Single-Six 9 1/2##8243;</td>
</tr>
<tr>
<td>Group2122</td>
<td>Non-Felon / Legal</td>
<td>Non-Felon / Legal</td>
<td>Non-Felon / Legal</td>
<td>Mentally Capable / Non-Felon / Legal</td>
</tr>
<tr>
<td>ID1421</td>
<td>47955db13ce85f8e</td>
<td>930552dc84196efa</td>
<td>8655da9d2455236</td>
<td>6256c6aa4070496</td>
</tr>
</tbody>
</table>
Schema Alignment

karma.isi.edu

Relational Sources

Hierarchical Sources

Services

Schema.org

{ JSON-LD }
Linking Using Image Similarity
DIG Applications

**Human Trafficking**
Identify victims, prosecute traffickers

**Cyber Attacks**
Predict cyber attacks from dark web data

**Firearms Trafficking**
Identify illegal sales

**Patents**
Identify patent trolls

**Securities Fraud**
Identify fraudulent stocks in the Penny Stock market