

Conclusion and review

Domain-specific search (DSS)

The Massive YouTube Ecosystem

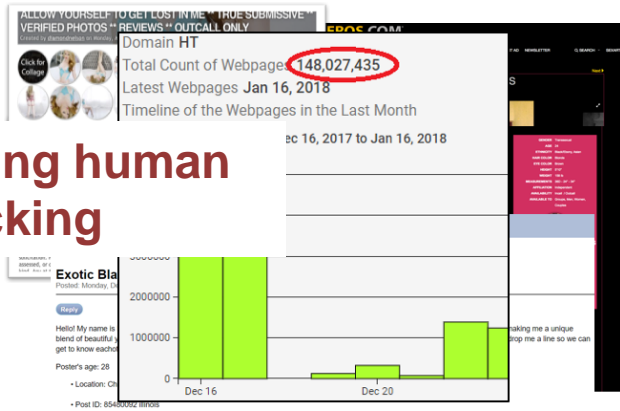
- 56 Countries
- 2nd Largest
- 60 Years
- One Billion
- 100 Hours
- 17%
- One Billion
- Six Billion
- Hundreds of Millions
- 25% Mobile



source: <https://photos.pnnewswire.com/prnfull/20151006/274273-INFO>

Emerging opportunities for DSS

Fighting human trafficking



Stopping Penny Stock Fraud

Penny Stock Fraud Nets Millions Scheme Mastermind Among Those Sentenced to Prison

Internet opens new avenue for penny stock fraud
 NEW YORK □ Most investors take e-mails advertising a 300 percent return on penny stocks sh bin. But those Internet promotions are still irresistible for some use of making a killing.
 ny, July 11, 2004

ommission is increasingly taking legal action against individuals and companies that nline. In one of its recent cases, involving Ives Health Co., the SEC reported a final ner president, M. Keith Ives, for disseminating misleading information on the internet.

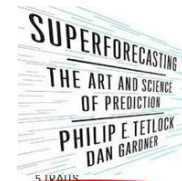
insate investors a total of \$1.25 million for, among other things, falsely claiming the igation the company developed.

Defined by the SEC as stocks that sell below \$5 a share, penny stocks have always been considered speculative and easily manipulated. But stock market experts, seeing an increase in penny stock promotion online, say investors should be wary of

Predicting cyberattacks



Accurate geopolitical forecasting



- performers begin by gathering as much information possible.
- performers nurture and develop the habit of thinking in terms of probabilities when exploring the likelihood of specific events.
- forecasting improves when individuals work in teams.
- performers ensure that they are regularly keeping abreast of their projections.
- The most successful forecasters are willing to admit error and quickly change course on their projections.



General Search

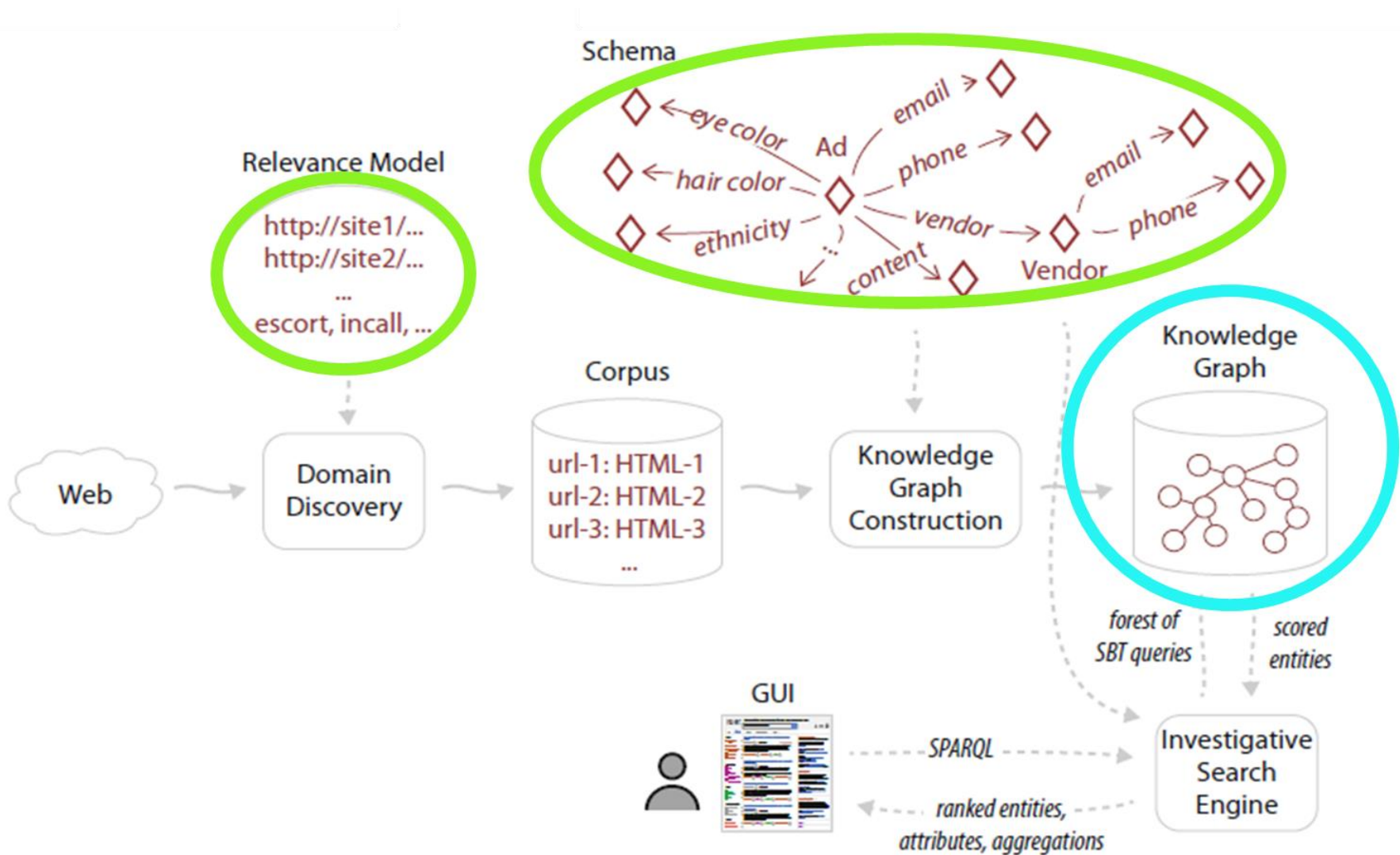
Google Knowledge Graph

DSS

Domain-Specific Knowledge Graphs

How do we construct domain specific knowledge graphs over web data for powerful DSS applications

Knowledge Graphs for DSS



Challenges

Many Document Features

Text paragraphs without formatting

Astro Teller is the CEO and co-founder of BodyMedia. Astro holds a Ph.D. in Artificial Intelligence from Carnegie Mellon University, where he was inducted as a national Hertz fellow. His M.S. in symbolic and heuristic computation and B.S. in computer science are from Stanford University. His work in science, literature and business has appeared in international media from the New York Times to CNN to NPR.

Grammatical sentences plus some formatting & links

Dr. Steven Minton - Founder/CTO
Dr. Minton is a fellow of the American Association of Artificial Intelligence and was the founder of the Journal of Artificial Intelligence Research. Prior to founding Fetch, Minton was a faculty member at USC and a project leader at USC's Information Sciences Institute. A graduate of Yale University and Carnegie Mellon University, Minton has been a Principal Investigator at NASA Ames and taught at Stanford, UC Berkeley and USC.

- Press
- **Contact**
- General information
- Directions maps

Frank Huybrechts - COO
Mr. Huybrechts has over 20 years of

Non-grammatical snippets, rich formatting & links

Barto, Andrew G. Professor. Computational neuroscience, reinforcement learning, adaptive motor control, artificial neural networks, adaptive and learning control, motor development.	(413) 545-2109	barto@cs.umass.edu	CS276
Berger, Emery D. Assistant Professor.	(413) 577-4211	emery@cs.umass.edu	CS344
Brock, Oliver Assistant Professor.	(413) 577-0334	oli@cs.umass.edu	CS246
Clarke, Lori A. Professor. Software verification, testing, and analysis; software architecture and design.	(413) 545-1328	clarke@cs.umass.edu	CS304
Cohen, Paul R. Professor. Planning, simulation, natural language, agent-based systems, intelligent data analysis, intelligent user interfaces.	(413) 545-3638	cohen@cs.umass.edu	CS278

Tables

Board of Directors		
Chairman of the Board	Chairman	Chief Executive Officer
President	Chairman	Chief Operating Officer
Chief Executive Officer	Chairman	Chief Representative for Future Operations
Chief Representative for Future Operations	Chairman	President, Medical Company
President, Medical Company	Chairman	President, Agricultural & Food Company
President, Agricultural & Food Company	Chairman	Chief Representative for Research Operations
Chief Representative for Research Operations	Chairman	Chief Representative for Chemical Operations
Chief Representative for Chemical Operations	Chairman	Chief Representative for Energy Operations
Chief Representative for Energy Operations	Chairman	Chief Representative for Infrastructure Operations
Chief Representative for Infrastructure Operations	Chairman	Chief Representative for Real Estate Operations
Chief Representative for Real Estate Operations	Chairman	Chief Representative for Other Operations
Chief Representative for Other Operations	Chairman	Chief Representative for Other Operations

Charts



Short tail

InvestorPlace Stocks Funds Retirement Trading Market Insight Financial Advisors Premium Services

DOW 22,049 +0.1% NASDAQ 6,352 +0.1% S&P 500 2,474 +0.1%

CHARTER COMM RG-A (CHTR)

400.90 -11.15 (2.86%) 20:10 EDT

CHTR STOCK QUOTE DELAYED 20 MINUTES

NAVILLER RATINGS

ALWAYS BE FINANCING TRADE OPPORTUNITIES

GET 100 FREE TRADES

CHTR Stock Chart Historical CHTR Prices

CHTR STOCK GRADE: B

Fundamental Grade: D

Quantitative Grade: A

CHTR Earnings: F

Earnings Growth: F

Earnings Momentum: F

Earnings Realization: F

Analyst Earnings Revisions: D

CHTR Financial Information

Sales Growth: A

InvestorPlace Stocks Funds Retirement Trading Market Insight Financial Advisors Premium Services

DOW 22,049 +0.1% NASDAQ 6,352 +0.1% S&P 500 2,474 +0.1%

LEGACY VENTURES INTL, Inc. (LGVV)

6.01 0.00 (0.00%) 07:41:17

LGVV STOCK QUOTE DELAYED 20 MINUTES

NAVILLER RATINGS

THE WAY TRADING SHOULD ALWAYS BE

LGVV Stock Chart Historical LGVV Prices

LEGACY VENTURES INTL, Inc. (LGVV)

Dividend & Yield: N/A (N/A)

P/E: -

Market Cap: 193,208

EPS: -0.05

Volume: 406

Days Range: 346.38 - 458.83

52wk Range: 236.08 - 458.83

LGVV STOCK PREDICTIONS, ARTICLES, AND LEGACY VENTURES INTL, INC. NEWS

Genre specific (e.g., forums)

The Lion News Money Price News Daily Economic Outlook The "Happier" Price in Euro

Home Forum Portfolios Blog User Mail Help

Welcome Stranger! Please register or log in to enable additional features.

Forum - The Lion.com Central

Return Top | Return List | Reply Thread | Search

From: [redacted] Date: 2017-01-11 09:02:12

HTZ - Hertz Global Holdings

Discussion in 'Stock Message Boards NYSE, NASDAQ, AMEX' started by [redacted] at May 7, 2016.

Stockholics

Do you want to help develop this community? You are looking for contributions from investors and traders like you! What stocks do you follow? What is hot right now? Sign up and get in on the ground floor of the newest, fastest growing financial forum!

HTZ - Hertz Global Holdings

Discussion in 'Stock Message Boards NYSE, NASDAQ, AMEX' started by [redacted] at May 7, 2016.

Long tail

AIR WATER INC. Business & Products R&D Environmental & Social Corporate Profile Investor Relations

Corporate Profile Corporate Outline

profitspi

Search to Interactive charts

Stock Charts

4-Traders

MARKETS NEWS ANALYSIS STOCK PICKS PORTFOLIOS SCREENERS WATCHLISTS TOP FLIP TOOLS

Home Fundamentals

Symbol	Price	%	YTD	2017	2016	2015	2014	2013
Home Fundamentals	16.31	0.0%	16.31	16.31	16.31	16.31	16.31	16.31
HD-1	7.92	0.0%	7.92	7.92	7.92	7.92	7.92	7.92
Shop	6.88	-0.8%	6.88	6.88	6.88	6.88	6.88	6.88
Sea Development	88.38	2.7%	88.38	88.38	88.38	88.38	88.38	88.38
Other	28.23	0.1%	28.23	28.23	28.23	28.23	28.23	28.23

Managers

Name	Age	Stock	Title
Michael Hargrave	68	1975	President, Transportation Director & CEO
Thomas Hargrave	62	1985	Senior Managing Director & CEO
Nathanial Hill	52	1985	CEO
Michael Hargrave	68	1975	President, Transportation Director & CEO
Thomas Hargrave	62	1985	Senior Managing Director & CEO
Nathanial Hill	52	1985	CEO
Michael Hargrave	68	1975	President, Transportation Director & CEO
Thomas Hargrave	62	1985	Senior Managing Director & CEO
Nathanial Hill	52	1985	CEO

By James Rickards

August 8, 2017

https://dailytrading.com/billion-money/

Is Billion Money?

New GE Chief Delays Part of

Pattern Complexity

Closed set

U.S. states

He was born in Alabama...

The big Wyoming sky...

Regular set

U.S. phone numbers

Phone: (413) 545-1323

The CALD main office can be reached at 412-268-1299

Complex

U.S. postal addresses

University of Arkansas
P.O. Box 140
Hope, AR 71802

Headquarters:
1128 Main Street, 4th Floor
Cincinnati, Ohio 45210

Ambiguous, needing context

Person names

...was among the six houses sold by Hope Feldman that

Pawel Opalinski, Software Engineer at WhizBang Labs.

Unusual language

models

"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 ♥
7o7~7two7~7four77 ♥ HH 80 roses ♥
Hour 120 roses ♥ 15 mins 60 roses"

Courtesy of Andrew McCallum

647-241-1986 New Haven Escort Listing

View Escorts in other cities

647-241-1986 Upscale + Sophisticated | Busty | Curvy Asian -- Delia London - 25

Escort's Phone: **647-241-1986**
Escort's Location: New Haven, Connecticut
Escort's Age: 25
Date of Escort Post: Jun 17th 4:49pm

REVIEWS: [READ AND CREATE REVIEWS FOR THIS ESCORT](#)

There are **42** girls looking in. [VIEW GIRLS](#)

If you are looking for the right combination of Erotic & Sensual then you have come to the right place. Always a great personality, and environment.
NO RUSH SERVICE Discreet & Upscale PLAYFUL 100% REAL PHOTOS.
100% Independent | Dedicated | Verified Provider | dateck dl6472fp 411 p98690
phone: 773 431 8174 ___ REFERENCES REQUIRED BDSM, Domme, & Fetishes Available | www.delialondon.com | Call [647-241-1986](tel:647-241-1986). See my menu of services on my profile
[Ezsex](#) Find me... [BackDoorOpen](#)

Call me on my cell at [647-241-1986](tel:647-241-1986).
Date of ad: 2016-06-17 16:49:00

More posts from 647-241-1986

- [647-241-1986 Oct 28, 2016 Verified Upscale + Sophisticated | Busty | Curvy Asian -- Delia London - 25](#)
- [647-241-1986 Oct 25, 2016 Verified Upscale + Sophisticated | Busty | Curvy Asian -- Delia London NOW IN TOWN...](#)
- [647-241-1986 Oct 09, 2016 Verified Upscale + Sophisticated | Busty | Curvy Asian -- Delia London - 22](#)
- [647-241-1986 Oct 09, 2016 Verified Upscale + Sophisticated | Busty | Curvy Asian -- Delia London In town TODA...](#)
- [647-241-1986 Oct 07, 2016 Visiting... Today Only :: Verified + Reviewed -- // Delia London ... In town for...](#)
- [647-241-1986 Oct 05, 2016 Verified Upscale + Sophisticated | Busty | Curvy Asian -- Delia London NOW IN TOWN...](#)
- [647-241-1986 Aug 16, 2016 NEW FICS Upscale + Sophisticated | Busty | Curvy Asian -- Delia London - 25](#)
- [647-241-1986 Aug 07, 2016 Upscale + Sophisticated | Busty | Curvy Asian -- Delia London - 25](#)
- [647-241-1986 Aug 07, 2016 Upscale + Sophisticated | Busty | Curvy Asian -- Delia London - 25](#)
- [647-241-1986 Jun 19, 2016 NOW IN W.I.U. Upscale + Sophisticated | Busty | Curvy Asian -- Delia London - 25](#)
- [647-241-1986 Jun 15, 2016 In & oncalls Upscale + Sophisticated | Busty | Curvy Asian -- Delia London - 25](#)
- [647-241-1986 May 16, 2016 Upscale + Sophisticated | Busty | Curvy Asian -- Delia London - 24](#)
- [647-241-1986 May 03, 2016 Upscale + Sophisticated | Busty | Curvy Asian -- Delia London - 25](#)
- [647-241-1986 Apr 30, 2016 Upscale + Sophisticated | Busty | Curvy Asian -- Delia London - 24](#)
- [647-241-1986 Mar 07, 2016 Upscale + Sophisticated | Busty | Curvy Asian -- Delia London NOW IN TOWN - 24](#)
- [647-241-1986 Feb 26, 2016 Upscale + Sophisticated | Busty | Curvy Asian -- Delia London - 24](#)
- [647-241-1986 Jan 13, 2016 Erotic x Busty Asian Companion Verified + Reviewed + Safe In town now - 24](#)
- [647-241-1986 Dec 21, 2015 Asian American -- Busty Companion + Kinkstress :: New Pics + Verified Provider - ...](#)
- [647-241-1986 Dec 14, 2015 Upscale + Sophisticated | Busty | Curvy Asian -- Delia London - 26](#)

Recent Escort Classifieds

- [North Jersey, New Jersey \(928-1100\)](#)
- [Gorgeous Beautiful MILF \(928-1100\)](#)
- [Chicago, Illinois \(773-220-2000\)](#)
- [\(LA\) PORN PHOTOS & HD TIMELAPSE SURF CAM NG AGENT CHASE V](#)
- [Chicago, Illinois \(414\) 4-37-4437](#)
- [Petite and sweet Super new and Ready... in out call - 21](#)
- [Chicago, Illinois \(312-600-8628\)](#)
- [CM Get PAMPERED - 22](#)
- [Ft Lauderdale, Florida \(786-504-1860\)](#)
- [you will ME, GORGEOUSLY sexy LATIN - 21](#)
- [Atlanta, Georgia \(401-324-9388\)](#)
- [WoW, Must Take a Look At This - 21](#)
- [Atlanta, Georgia \(401-324-9388\)](#)
- [Seattle, Washington \(206\) 440-1982](#)
- [Seattle, Washington \(206\) 440-1982](#)
- [Atlanta, Georgia \(404-849-8140\)](#)
- [I'm to please all of desires CALL ME 100% Real Curvy White Girl - 21](#)
- [Houston, Texas \(832\) 496-6677](#)
- [Beautiful Salvadoran The One And Only! - 21](#)
- [Phoenix, Arizona \(623-500-7076\)](#)
- [NEW GIRL PERSIAN Gem EXotic Blend - 21](#)
- [Toronto, Ontario \(416-554-3337\)](#)
- [\(United States\) Special 80 for 20 min 22 - Bar of 4 \\$\\$\\$\\$very LITING PORN MBSherLe - 21](#)
- [Toronto, Ontario \(947-702-8825\)](#)

Top Escort Cities

- New York, New York
- Toronto, Ontario
- Dallas, Texas
- Chicago, Illinois
- Atlanta, Georgia
- North Jersey, New Jersey
- Detroit, Michigan
- Los Angeles, California
- San Diego, California
- Orange County, California
- Houston, Texas
- Phoenix, Arizona
- Philadelphia, Pennsylvania
- Boston, Massachusetts
- Washington DC, DC
- Orlando, Florida
- Las Vegas, Nevada
- Miami, Florida

Recent Blog Posts

- [Why do girls who work in escorts remove their photos with you...](#)
- [The Psychology of Escorts: Impacts that escorts have on their clients](#)
- [The Psychology of Escorts: Impacts that escorts have on their clients](#)
- [Surfside Beach, SC Prostitution BUST! Video](#)

Search Box

Search For Profiles

- [Register Here](#)
- [Login to your account](#)
- [Non Mobile Version](#)
- [Escort Blog](#)
- [Key for Escort Acronyms](#)
- [Top 10 Escort Practices](#)
- [Escort Reviews](#)
- [See Escorts on Webcam](#)
- [Prostitution Laws](#)

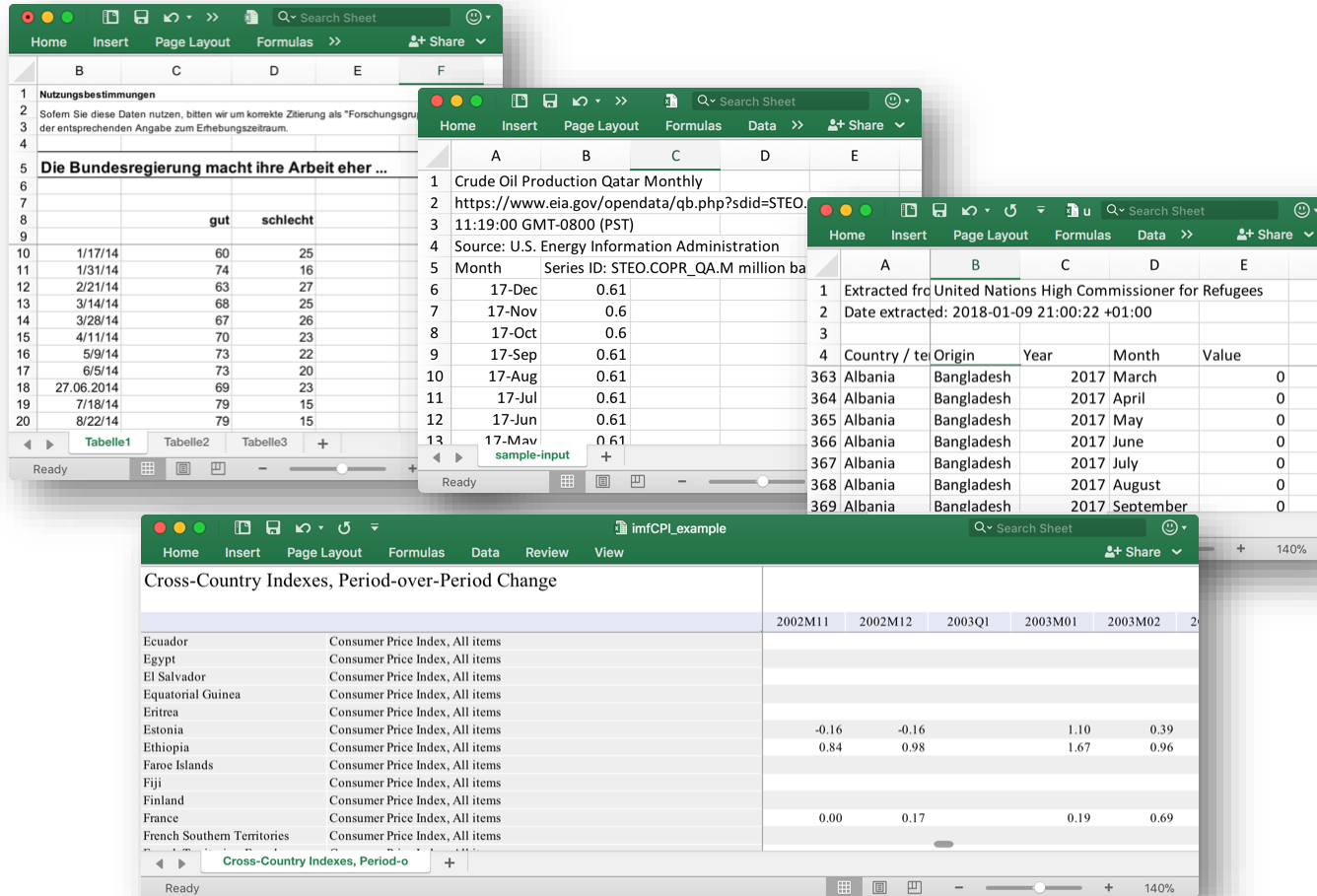
Most Recently Viewed
Today at 5:30pm Pacific



[419-283-6378](tel:419-283-6378)
[Detroit](#)

small amount of relevant content irrelevant content very similar to relevant content

Spreadsheets Created For Human Consumption



Databases with PDF Code Books

	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	event_date	year	time_precision	event_type	actor1	assoc_actor	inter1	actor2	assoc_actor	inter2	interaction	region	country	admin1
2	1/13/18	2018		1 Battle-No ch	Military Forces of Democ			1 ADF: Allied Democratic Fc			2	12 Central Afric Democratic	Nord-K	
3	1/13/18	2018		1 Battle-No ch	Military Forces of Democ			1 ADF: Allied Democratic Fc			2	12 Central Afric Democratic	Nord-K	
4	1/13/18	2018		1 Battle-No ch	Military Forces of Democ			1 ADF: Allied Democratic Fc			2	12 Central Afric Democratic	Nord-K	
5	1/13/18	2018		1 Battle-No ch	Al Shabaab			2 Police Forces of Kenya (20			1	12 Eastern Afric Kenya	Lamu	
6	1/13/18	2018		1 Riots/Protest	Rioters (Kenya)			5 Police Forces of Kenya (20			1	15 Eastern Afric Kenya	Marsab	
7	1/13/18	2018		1 Riots/Protest	Protesters (L Tawergha Co			6			0	60 Northern Afr Libya	Sabha	
8	1/13/18	2018		1 Violence aga	Unidentified Armed Group			3 Civilians (Int: IOM: Interna			7	37 Northern Afr Libya	Sabha	
9	1/13/18	2018		1 Riots/Protest	Protesters (Morocco)			6			0	60 Northern Afr Morocco	Orienta	
10	1/13/18	2018		1 Riots/Protest	Protesters (Nigeria)			6			0	60 Western Afri Nigeria	Benue	
11	1/13/18	2018		1 Violence	Civilians = 7							47 Western Afri Nigeria	Kaduna	
12	1/13/18	2018		2 Riots/Protest	Protesters (Nigeria)							50 Western Afri Nigeria	Lagos	
13	1/13/18	2018		1 Battle-No ch	Unidentified Armed Group							13 Eastern Afric Somalia	Banaac	
14	1/13/18	2018		1 Battle-No ch	Unidentified Armed Group							12 Eastern Afric Somalia	Shabee	
15	1/13/18	2018		1 Battle-No ch	Unidentified Armed Group							16 Southern Afr South Africa	Capetown	

These single numbers represent the actors noted in “Actor 1” and “Actor 2” columns, and are placed in “Inter 1” and “Inter 2” respectively. “Inter 1” and “Inter 2” are the basis of the “Interactions” column. Interaction numbers are always the smallest possible number (for example, 37 instead of 73), regardless of the order of “Actor 1” and “Actor 2”. For single actor events, the empty second actor category is coded as “0”.

Interaction codes include:

- 10- SOLE MILITARY ACTION
- 11- MILITARY VERSUS MILITARY
- 12- MILITARY VERSUS REBELS
- 13- MILITARY VERSUS POLITICAL MILITIA

PDF

Data In Web Tables



UNITED NATIONS SECURITY COUNCIL

ABOUT PRESIDENCY MEMBERS PROGRAMME OF WORK DOCUMENTS MEETINGS SUBSIDIARY ORGANS 2231 (2015) REPERTOIRE

■ About SC Documents
 ■ Resolutions
 ■ Presidential Statements
 ■ Press Statements
 ■ Notes by the President
 ■ Exchange of Letters
 ■ Reports submitted by / transmitted by the Secretary-General
 ■ Reports of the Security Council Missions
 ■ Annual Reports
 ■ Volumes of Resolutions
 ■ Round-ups
 ■ Highlights of Security Council Practice
 ■ Annual Highlights

Security Council Resolutions

Resolutions adopted by the Security Council in 2017

Resolution ID	Date	Title
S/RES/2397 (2017)	22 December 2017	Non-proliferation/Democratic People's Republic of Korea
S/RES/2396 (2017)	21 December 2017	Threats to international peace and security caused by terrorist acts
S/RES/2395 (2017)	21 December 2017	Threats to international peace and security caused by terrorist acts
S/RES/2394 (2017)	21 December 2017	The situation in the Middle East
S/RES/2393 (2017)	19 December 2017	The situation in the Middle East

07年 1月 2月 3月 4月 5月 6月 7月 8月 9月 10月 11月 12月 (参)

(参) = 第21回参院選第1回全国電話調査 (%)

	1月	2月	3月	4月	5月	6月	7月 (参)	8月	9月	10月	11月	12月
支持する	51	41	44	44	50	37	38	29	34	58	54	51
支持しない	34	43	39	37	34	47	49	58	55	27	35	33
内閣	安倍内閣						福田内閣					

政党支持率 (%)

	1月	2月	3月	4月	5月	6月	7月 (参)	8月	9月	10月	11月	12月
自民党	36.4	31.6	31.4	33.4	34.3	30.2	31.8	27.7	27.4	32.8	34.0	31.5
民主党	14.8	13.3	13.9	12.0	13.0	15.1	20.5	26.0	24.5	18.9	21.6	19.8
公明党	4.1	3.3	3.8	4.0	3.8	3.5	4.3	3.5	2.6	4.1	3.1	3.3
共産党	2.2	3.0	1.7	2.1	2.6	3.0	3.8	3.7	2.6	2.9	1.5	2.2
社民党	1.3	1.9	1.4	0.8	1.0	1.5	1.3	2.3	1.3	1.3	1.5	1.0
国民新党	0.1	0.1	0.1	0.1	0.1	0.3	0.2	0.4	0.3	0.3	0.2	0.2
新党日本	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.4	0.2	0.1	0.0	0.1
その他の政治団体	0.0	0.1	0.0	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.1
支持なし	34.8	39.3	37.8	41.3	38.9	38.5	22.1	30.8	34.8	32.1	32.1	32.1
わからない、無回答	6.4	7.4	10.0	6.3	6.2	7.6	15.9	5.3	6.3	7.7	6.1	9.7

調査概要

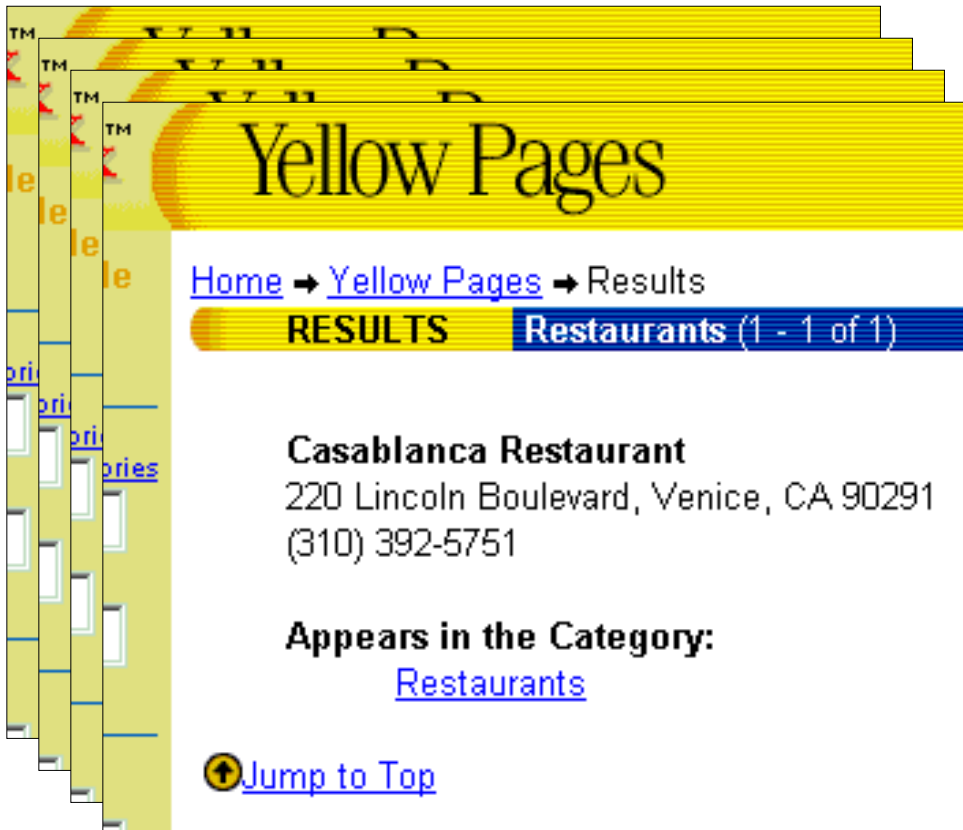
調査対象：全国の20歳以上の男女
 調査方法：電話法 (RDD追跡法)

調査時期	調査相手(人)	回答数(人)	回答率(%)
1月 1月6日(土)～8日(月)	1,725	1,030	59.7

Knowledge Graph Construction: Long-tail vs. Short-tail

Extracting Data from Semi-structured Sources

Wrappers



Yellow Pages

Home → [Yellow Pages](#) → Results

RESULTS Restaurants (1 - 1 of 1)

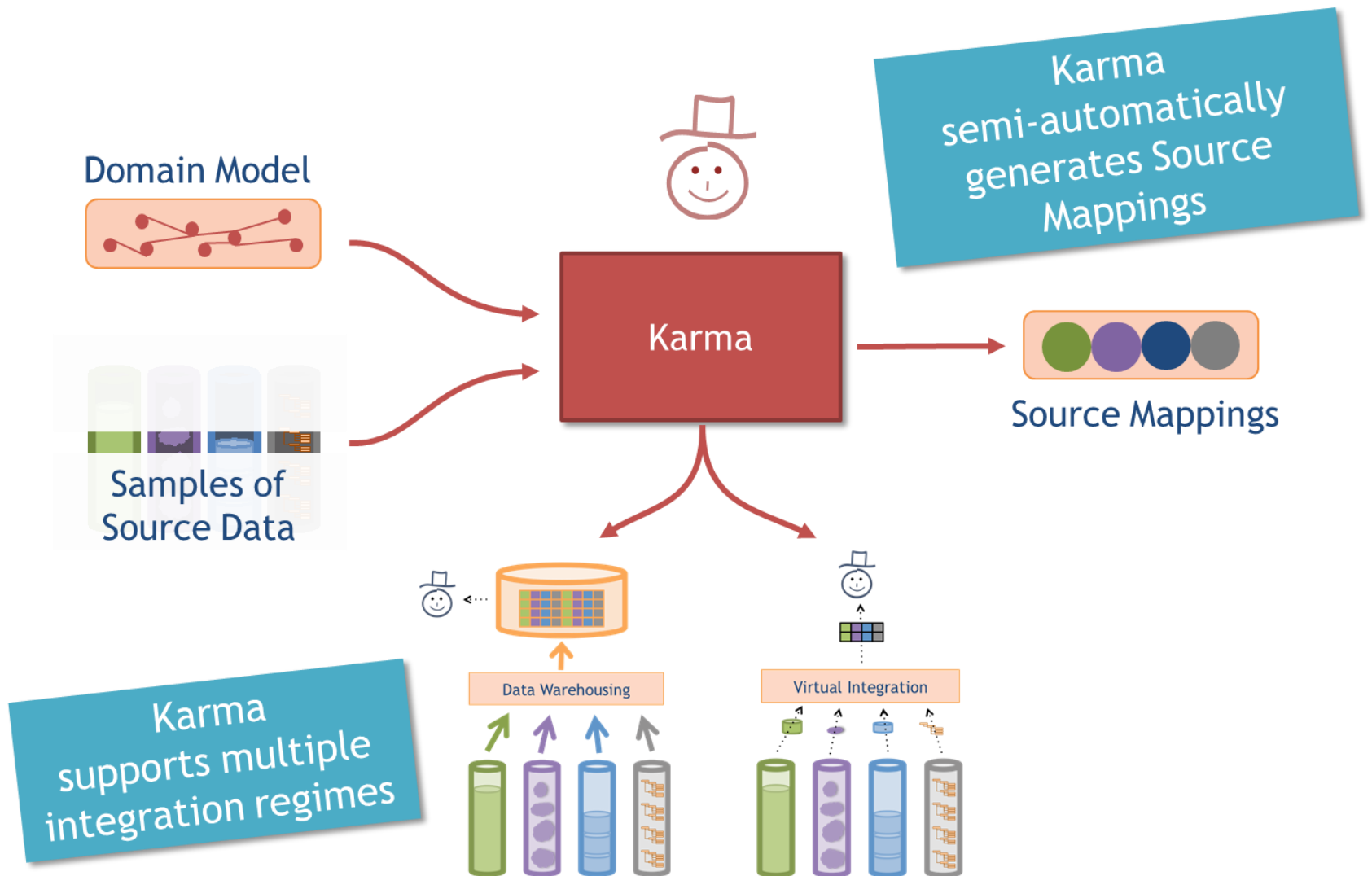
Casablanca Restaurant
220 Lincoln Boulevard, Venice, CA 90291
(310) 392-5751

Appears in the Category:
[Restaurants](#)

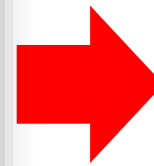
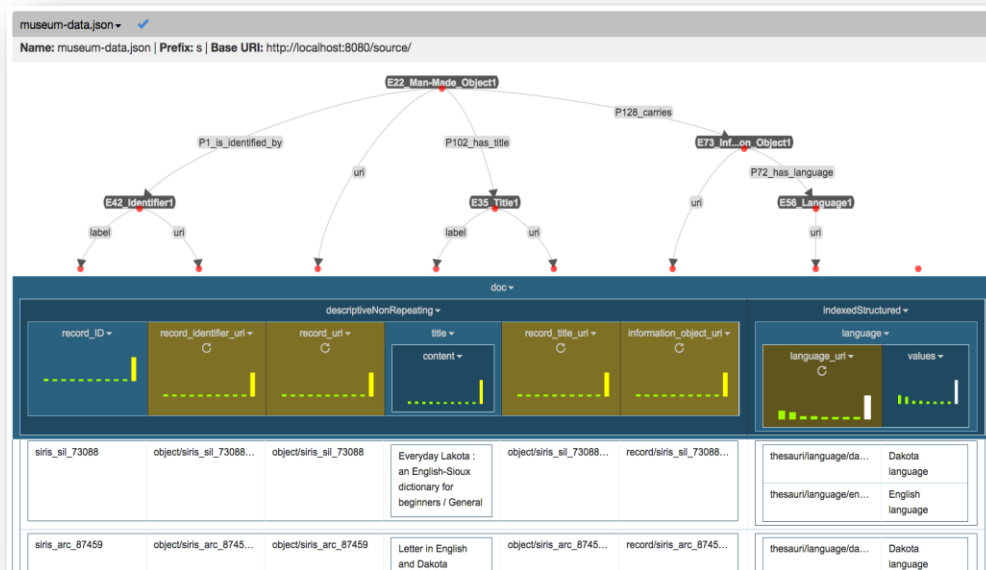
[↑ Jump to Top](#)

NAME	Casablanca Restaurant
STREET	220 Lincoln Boulevard
CITY	Venice
PHONE	(310) 392-5751

Information Integration in Karma



Karma **semi-automatically builds** semantic models
... and provides a **nice GUI** to edit them



Knowledge
Graphs

Karma uses **semantic models** to **create** knowledge graphs

Practical Considerations for Extractions

Long tail vs. short tail

- How good (precision/recall) is necessary?
 - High precision when showing KG nodes to users
 - High recall when used for ranking results
- How long does it take to construct?
 - Minutes, hours, days, months
- What expertise do I need?
 - None (domain expertise), patience (annotation), scripting, machine learning guru
- What tools can I use?
 - Many ...

Data Tables

Arsène Wenger



Wenger in July 2015

Personal information	
Full name	Arsène Wenger ^[1]
Date of birth	22 October 1949 (age 67)
Place of birth	Strasbourg, Alsace, France
Height	6 ft 3 in (1.91 m) ^[2]
Playing position	Midfielder
Club information	

Entity Table

Table 4: Average (mean) earnings (£) of UK employees by 2010

	Women F/T £	Women P/T £	Men F/T £	Men P/T £
Managers and senior officials	18.66	15.74	24.67	xxx
Professional occupations	20.43	22.82	22.47	27.55
Associate professional and technical	14.85	14.77	16.84	15.41
Administrative and secretarial	10.80	9.54	12.05	9.73
Skilled trades	8.86	7.89	11.59	10.63

Matrix Table

20 Strongest Performing Metro Areas

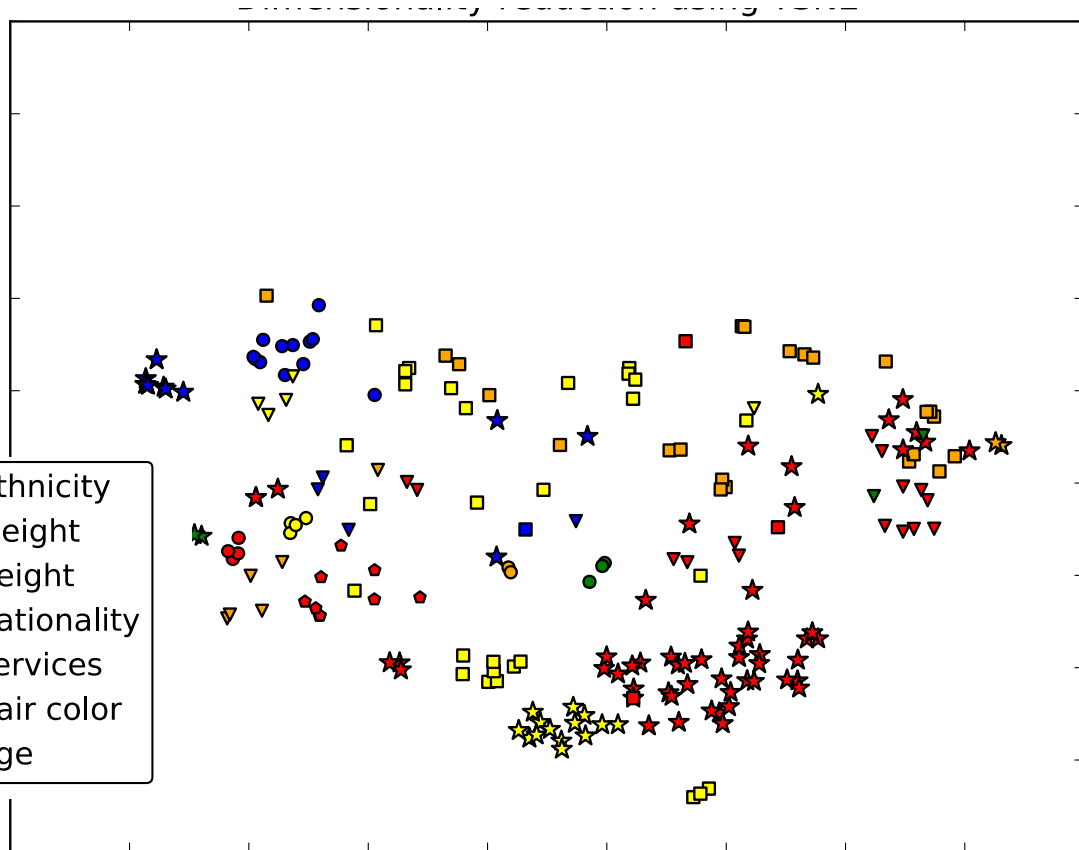
1. San Antonio, TX
2. Oklahoma City, OK
3. Austin, TX
4. Houston, TX
5. Dallas, TX
6. McAllen, TX
7. Little Rock, AR
8. Baton Rouge, LA
9. Tulsa, OK
10. Omaha, NE-IA
11. El Paso, TX
12. Wichita, KS
13. Washington, DC-VA-MD-WV
14. Des Moines, IA
15. Albuquerque, NM
16. Virginia Beach, VA-NC

List Table

Table Type Classification

- Feature-based supervised classification
 - Cafarella'08
 - Crestan'11
 - Eberius'15
- Deep Learning
 - Nishida'2017

Semantic + Structure Embedding



- | | | |
|---------------|-----------------|-----------------|
| ☆☆☆ telephone | □□□ measurement | ▽▽▽ ethnicity |
| ☆☆☆ other | ●●● web | ☆☆☆ weight |
| ☆☆☆ price | ☆☆☆ clients | □□□ height |
| □□□ location | ▽▽▽ date | ▽▽▽ nationality |
| ●●● language | ●●● orientation | □□□ services |
| ▽▽▽ email | ●●● name | ●●● hair color |
| ●●● gender | ☆☆☆ eye color | ▽▽▽ age |

Data Extraction Techniques

- Glossary
- Regular expressions
- Natural language rules
- Named entity recognition
- Sequence labeling (Conditional Random Fields)

Searching Knowledge Graphs

Many problems with 'strict' execution

```
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

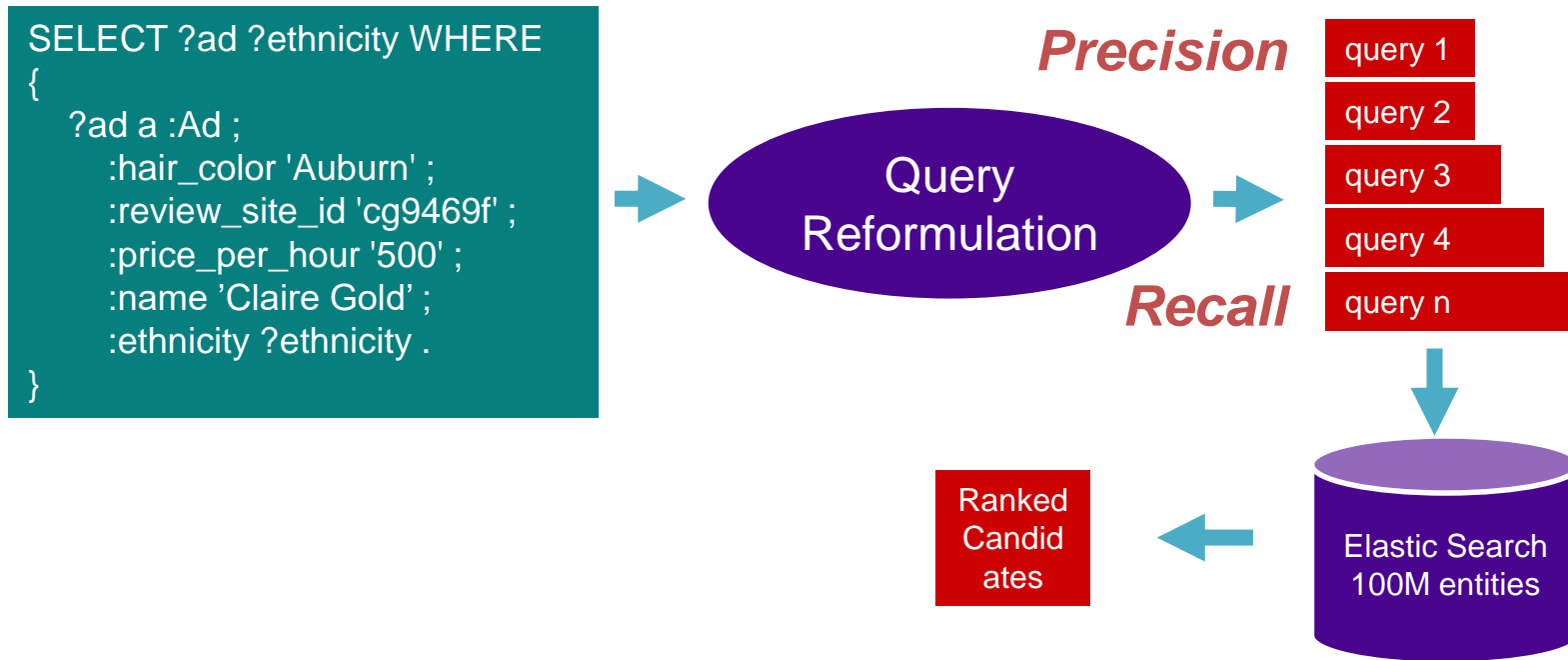
Claire is a common name
Gold is a domain word
slang, e.g., "FOB" for Asian
inference, e.g., "Japanese"



**No
results**

Candidate Generation

Keyword expansion • Context broadening • Constraint relaxation



Knowledge Graph Completion

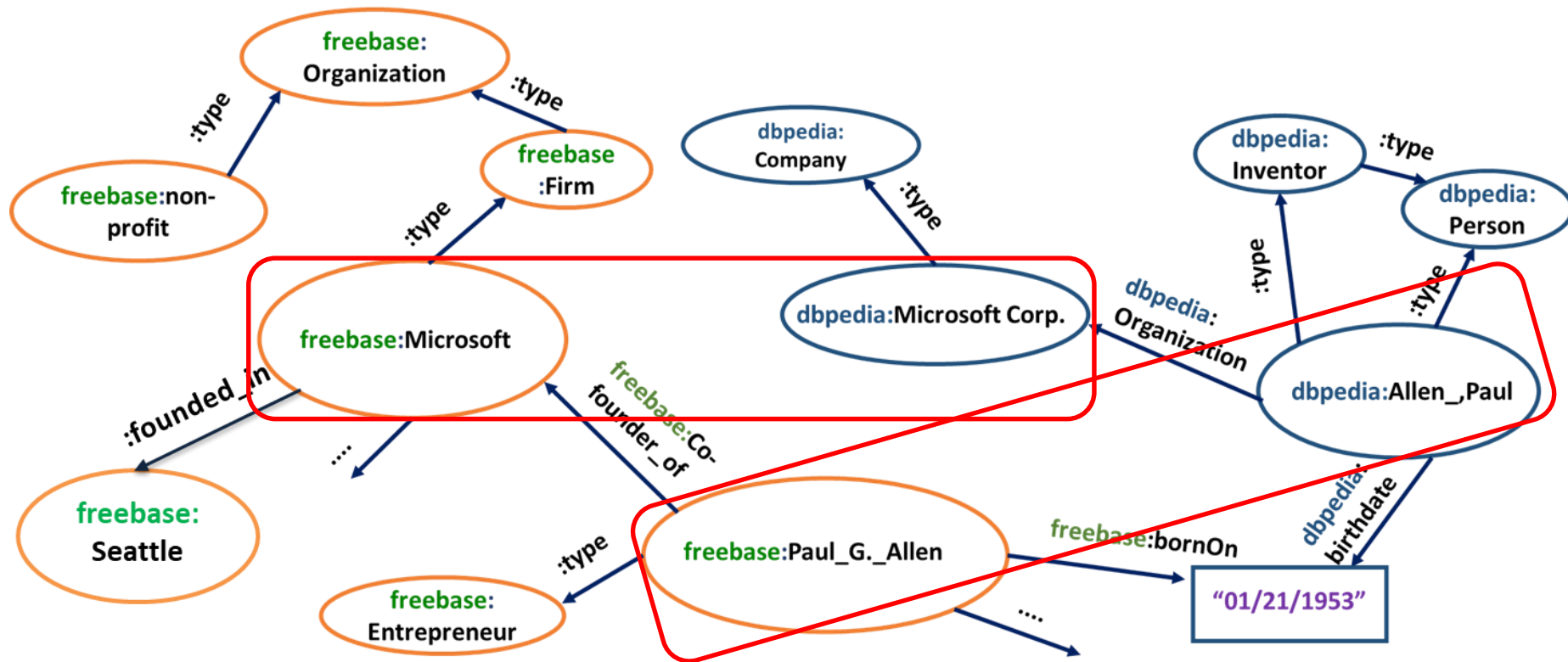
For Web extractions, noise is inevitable

- Thousands of web domains
- Many page formats
- Distracting & irrelevant content
- Purposeful obfuscation
- Poor grammar & spelling
- Tables

To reach its potential, a **constructed KG** must be **completed or identified**

Entity Resolution

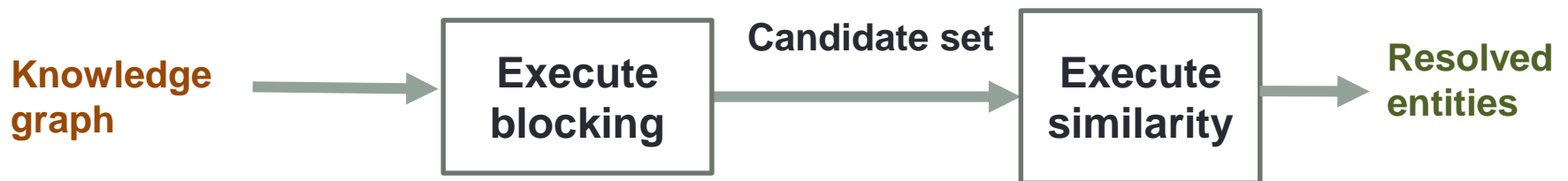
Many nodes refer to the **same underlying entity**



Entity Resolution is fundamentally non-linear

- Theoretically quadratic in the number of nodes, even if 'resolution rule' was known
- In practice, number of 'duplicates' tends to grow linearly, and duplicates overlap in non-trivial ways
- How to devise efficient algorithms?

50 years of research has agreed on a two-step solutions



Other research frontiers for KG Completion

- **Other things explored in the literature:**
 - **Domain knowledge**
 - Collective ER methods have tried to exploit these systematically
 - **Multi-type Entity Resolution**
 - Extremely useful for knowledge graphs, lots more work to be done
 - **Entity Resolution+Ontologies+IE Confidences:**
 - Probabilistic Graphical Models like Probabilistic Soft Logic
 - **Knowledge graph embeddings**
 - Useful for link prediction and triples classification
 - Recall the Microsoft-founded_in-Seattle example earlier



USC University of
Southern California

Information Sciences Institute



THANK YOU! QUESTIONS...