
EntityRank: Searching Entities Directly and Holistically

Tao Cheng

Joint work with: Xifeng Yan, Kevin Chang

VLDB 2007, Vienna, Austria

Motivating Scenario



Customer service **phone** number of Amazon?

Help Topics**Contact Us**

- > [Placing or Changing Orders](#)
- > [Using Features & Services](#)
- > [Phishing or Spoofed E-mails](#)
- > [Gift Orders & Gift Certificates](#)
- > [Selling Items](#)
- > [Prices, Promotions, & Rebates](#)
- > [Returns & Refunds](#)
- > [Account Assistance](#)
- > [Typographical Errors](#)

Contact Us

Our Help pages or [Your Account](#) often contain the answers to your questions. Therefore, to best serve you in answering your question quickly, we have compiled frequently asked questions by topic. (If the self-service options don't answer your question, you'll have an opportunity to e-mail us.)

First choose the topical link on this page that best describes your question. From this page, read through the frequently asked questions listed and click any of the links to read the corresponding answer. If your question isn't addressed, click the sentence "I have a question that's not listed above" (found on many of these pages), the "e-mail us" button, or the "contact us" link found at the bottom of the page to e-mail our customer service department. We'll be happy to assist you.

Search on Amazon?

- > [Delivery Estimates](#)
- > [Tracking Packages](#)
- > [Where's My Stuff?](#)
- > [Amazon and the Environment](#)
- >> [More...](#)

Privacy & Security

- > [A-to-Z Guarantee Protection](#)
- > [Identifying Phishing E-Mails](#)
- > [Safety & Security Tips](#)
- > [Credit Card Security](#)
- > [Privacy Notice](#)
- >> [More...](#)

[Special Occasion Reminder](#). If you don't receive it on the date you specify, you will know that messages from Amazon.com are being blocked or deleted automatically.

Related Topics

- [Placing or Changing Orders](#)
- [Using Features & Services](#)
- [Phishing or Spoofed E-mails](#)
- [Gift Orders & Gift Certificates](#)
- [Selling Items](#)
- [Prices, Promotions, & Rebates](#)
- [Returns & Refunds](#)



amazon customer service

Search

[Advanced Search](#)
[Preferences](#)

Web

[The Amazon.Com Customer Service Page](#)

Looking for the **Amazon.com customer service** phone number? Wait! Before you read any more, I wanted to let you know that I've been doing a little more ...

[clicheideas.com/amazon.htm](#) - 21k - [Cached](#) - [Similar pages](#) - [Note this](#)

[Amazon.com: Help > Contact Us](#)

Returns to **Amazon Merchants ...** Get Express **customer service** or contact us by e-mail or phone. contact us. Search Help Topics. Did this info help? ...

[www.amazon.com/gp/help/customer/display.html?nodeId=518316](#) - 43k -

[Cached](#) - [Similar pages](#) - [Note this](#)

Search on Google?

[A](#)
[T](#)
[C](#)

[www.slate.com/id/2091623/](#) - 31k - [Cached](#) - [Similar pages](#) - [Note this](#)

[Amazon's customer service number. - By Timothy Noah - Slate Magazine](#)

A journalist, if he's lucky, gets at most one chance in life to leave a lasting legacy. Jacob Riis exposed the horrors of tenement life.

[slate.com/id/2111081/](#) - 29k - [Cached](#) - [Similar pages](#) - [Note this](#)

[What Is the Customer Service Phone Number at Amazon.com?](#)

Questions and answers from educators, legal professionals, research professionals and subscribers to TVC Alert, a research news bulletin.

[www.virtualchase.com/ask_answer/amazon_phone.html](#) - 6k -

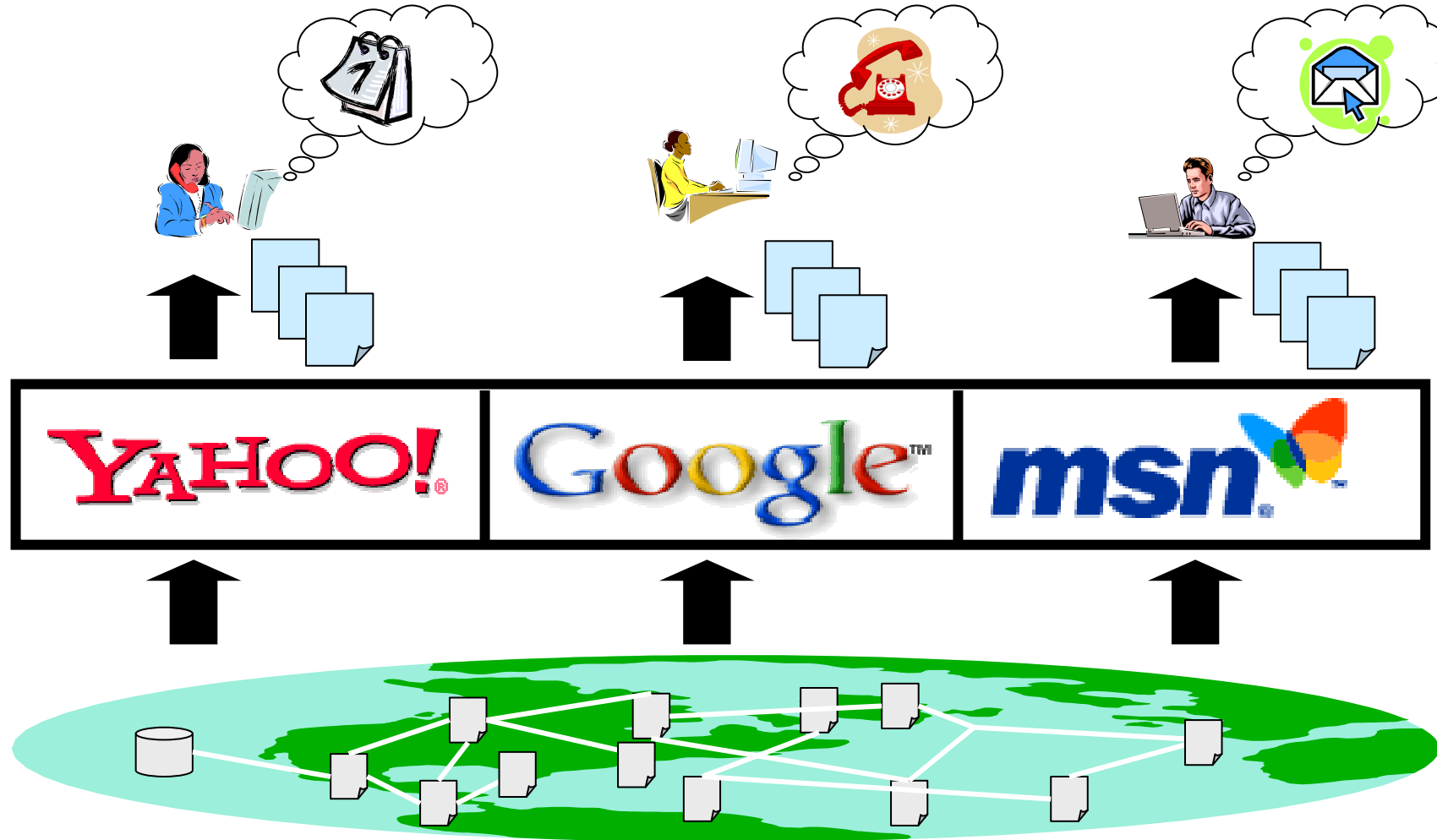
[Cached](#) - [Similar pages](#) - [Note this](#)

Many many similar cases:

- The **email** of Luis Gravano?
- What **profs** are doing databases at UIUC?
- The **papers** and **presentations** of ICDE 2007?
- Due **date** of SIGMOD 2008?
- Sale **price** of “Canon PowerShot A400”?
- “Hamlet” **books** available at bookstores?

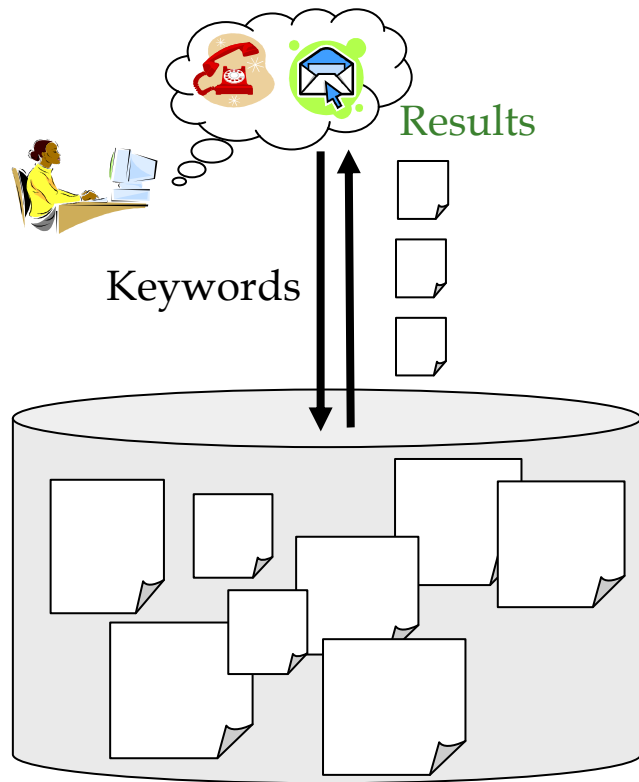
Often times, we are looking for data **entities**, e.g. emails, dates, prices, etc, not pages.

What you search is not what you want.

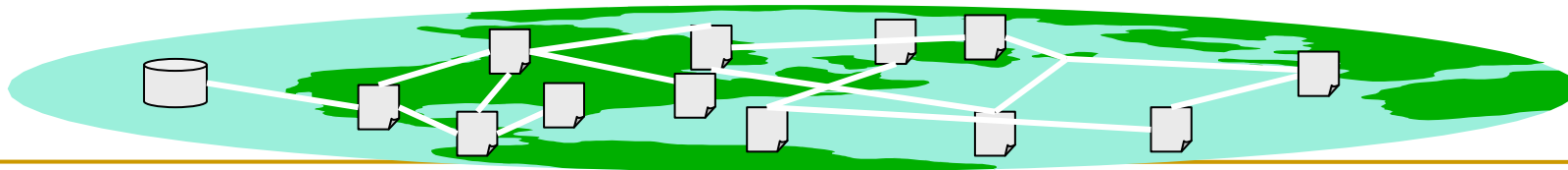
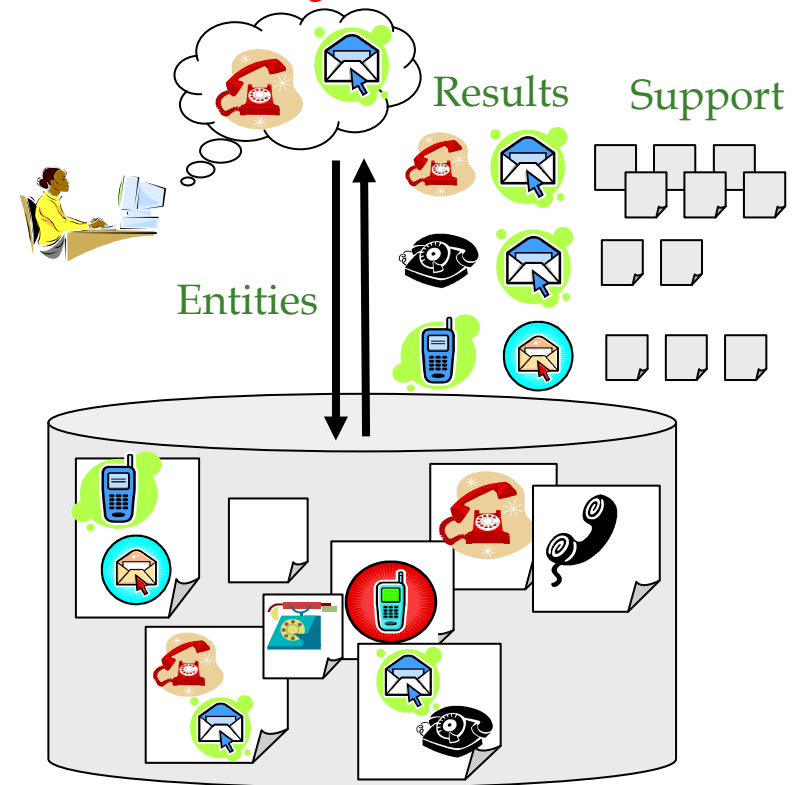


From pages to entities

Traditional Search



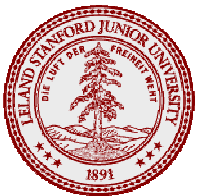
Entity Search



Concretely, what do we mean by Entity Search?

Online Demo.

- Yellowpage: *Comprehensive corpus.*

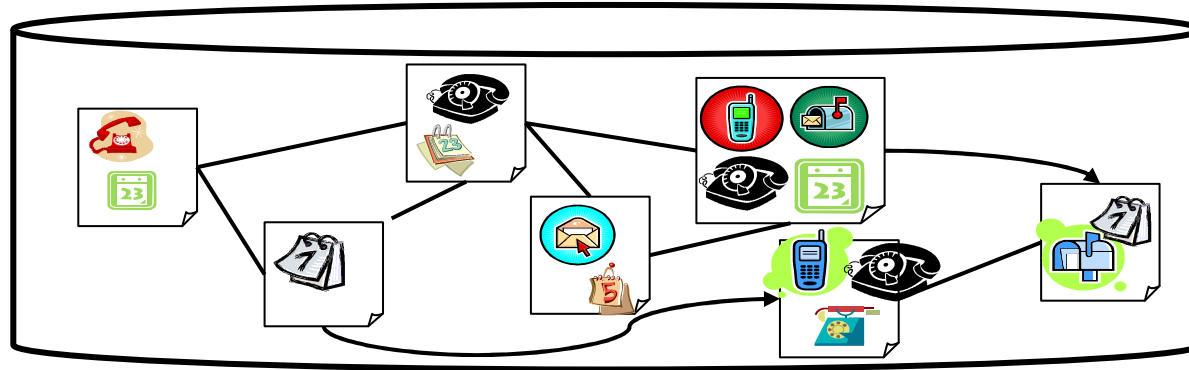


Special Thanks:

~100M Pages from Stanford WebBase

Entity Search Problem:




Given:



Input: Keywords & **Entities** (optionally with a pattern)

E.g. Amazon Customer Service **#phone**

Output: Ranked Entity Tuples

	0.90
	0.80
	0.60
...	...

Challenge:

How to rank Entities?

Why a novel Problem?

Characteristics I: Contextual

-Utilize Entities' Surrounding Context

The Amazon.com Customer Service Phone Number

When you call, please be nice to your customer service rep! -- your situation is not their fault. They are on your side and trying to help you.

Happy shopping!

The numbers!

Amazon.com Customer Service Phone Number
US Customer Service
Phone toll-free in the US and Canada:
(800) 201-7575

Phone from outside the US and Canada:
(206) 346-2992 or (206)-266-2992
Fax: (206) 266-2950
E-mail: orders@amazon.com (I think this will still work, but no guarantees)

According to good sources, Amazon is no longer outsourcing much of its customer service work to iSky.

You asked for it! e-Bay's Phone Numbers!

e-Bay, Inc.
408-376-7400
Toll Free: 1-800-322-9266

Content

Context

Characteristics II: Uncertain

-Extractions are non-“prefect”

Steve Lawrence, Luis Gravano: Learning to find answers to questions on the Web. *ACM Trans. Internet Techn.* 4(2): 129-162 (2004)

Luis Gravano, Amélie Marian: Optimizing Top-k Selection Queries over Multimedia Repositories. *IEEE Trans. Knowl. Data Eng.* 16(8): 992-1009 (2004)



Xantrex Technologies XPower Plus 400-Watt Inverter 851-0400

Xantrex (May 14, 2003)

Average Customer Review: ★★★★★ (58)

In Stock

List Price: ~~\$59.99~~

Characteristics III: Holistic

-Many evidences from multiple sources

Amazon.com: The Death of Customer Service

By Antoine du Rocher

SAN FRANCISCO, 28 December 2003—Customer service, small surprise, has been one of the casualties of America's drive towards cost-cutting in the age of e-business. The movement of customer service call centers off-shore is one-upped by companies like Amazon.com, which increasingly are hiding their customer service telephone numbers and other contact information, in order to prevent dissatisfied customers from calling in for service at all.

Amazon US Customer Service

1.800.201.7575 (Toll free, US and Canada)
1.206.346.2992 or 1.206.266.2992 (Outside US and Canada)
1.877.586.3230 (Canada only)



Gregory (Grisha) Chockler
Research Staff Member,
[IBM Haifa Research Laboratory](#)

During 2003-2005, I was a postdoctoral associate with the [Theory of Distributed Systems](#) group, [MIT/CSAIL](#).

Ph.D., [CS and Eng. School](#), The Hebrew University of Jerusalem, Israel, 2003.

AMAZON.COM customer service phone number (US): (800) 201-7575

Digging up buried info, like how to quit AOL

August 17, 2006

By [Jim Rossman](#) / The Dallas Morning News



Jim Rossman is your Tech Adviser offering advice and tips for computer hardware and programs. Helpful links are included. [Jim Rossman](#) is technical manager for Macintosh support for Belo Corp.

Reaching eBay, Amazon

While I'm at it, another hard-to-find phone number is the customer service line for eBay.

The main number for eBay is 1-888-749-3229. Once connected, press 2 for customer service.

Another handy phone number for eBay users is the customer care number for PayPal, 1-888-221-1161.

I'll throw in one more — Amazon's number is 1-800-201-7575.

Characteristics IV: Discriminative

- Web Pages are of Varying Quality

CNN.com Web | [CNN News](#) | [CNN Videos](#)

HOME WORLD U.S. POLITICS ENTERTAINMENT HEALTH TECH TRAVEL LIVING BUSINESS SPORTS

Hot Topics » [Blackwater USA](#) · [Subprime Lending](#) · [Madeleine McCann](#) · [CNN Heroes](#) · [More Topics](#)

Breaking News The Fed cuts by one-half percent the heavily impacts how much interest c

updated 2:41 p.m. EDT, Tue September 18, 2007 [Make CNN Your Home Pa](#)

Q: What's this web site?

A: I'm Molly E. Holzschlag, and this web site shares my **web development work and personal though** Given that, one hopes I have an interesting enough personality to keep you entertained for at least a littl

THURSDAY 8 DECEMBER 2005

BEST. SPAM. EVER.

TELL ME YOUR BEST COMMENT SPAM EVER!

I just got a really good one, if not the best:

“Thanks for useless info!”

I would understand this better had it been a real comment instead of just spam.

What's your favorite comment spam ever?

Filed under: [humor](#), [blogging](#), [pop culture](#), [software](#), [society](#)

Posted by: Molly | 5:22 am |

33 RESPONSES TO “BEST. SPAM. EVER.”

Pingu Says:
December 8th, 2005 at 5:29 am

I agree.

investorconnect.com

Home Alternative Investments Commodity And Markets Financial Planning International Investing

Most Relevant Link: [Investors](#)

Other Relevant Links:

- [Alternative Investments](#)
- [Commodity And Markets](#)
- [Financial Planning](#)
- [International Investing](#)
- [Investing](#)
- [Investing Online](#)
- [Investment Advisors](#)
- [Investment Fraud](#)
- [Investment Managers](#)
- [Investment Newsletters](#)
- [Investment Property](#)
- [Investment Research](#)
- [Investment Strategy](#)
- [Mutual Funds](#)
- [Online Brokers](#)
- [Real Estate Investments](#)
- [Smart Investing](#)
- [Stock Certificates](#)

Popular Links:

- Financial Aid and Grants**
 - [Fasfa.gov](#)
 - [Free Grant Money](#)
 - [Fasfa.com](#)
 - [Grant Money](#)
 - [Grant Application](#)
- Collection Services**
 - [Debt Collection](#)
 - [Collection Lawyer](#)
 - [Judgment Collection](#)
 - [Auto Repo](#)
 - [Sample Collection Letter](#)
- Banking**
 - [Chase Bank](#)
 - [Banks](#)
 - [Orchard Bank](#)
 - [Wachovia Bank](#)
 - [National City](#)

Related Links

- [Investor](#)
- [Investment](#)
- [Investing](#)
- [Investing Information](#)
- [Online Investing](#)
- [Invest](#)
- [Money Investing](#)
- [Investing In Mutual Funds](#)
- [Retirement Investing](#)
- [Venture Capital](#)
- [Investing Online](#)
- [Investment Opportunity](#)
- [Investment Online](#)
- [Investment Company](#)
- [Wall Street](#)

Financeinvesting Categories

- Investing**
 - [Online Investing](#)
 - [Offshore Investing](#)
 - [Bonds](#)
 - [Mutual Funds](#)
- Stocks**
 - [Stock Market](#)
 - [Stock Trading](#)
 - [Stock Options](#)
 - [Stock Investments](#)
- Banking**
 - [Bill Payments](#)
 - [Online Banking](#)
 - [Checks](#)
 - [Banking](#)
- Finance**
 - [Financial Aid](#)
 - [Bankruptcy](#)
 - [Personal Finance](#)
 - [Debt Relief](#)
- Offshore**
 - [Offshore Trusts](#)
 - [Offshore Merchant Account](#)
 - [Offshore Banking](#)
 - [Offshore Investing](#)
- Retirement**
 - [Social Security](#)
 - [Retirement Planning](#)
 - [401k Plans](#)
 - [Annuities](#)

Characteristics V: Associative

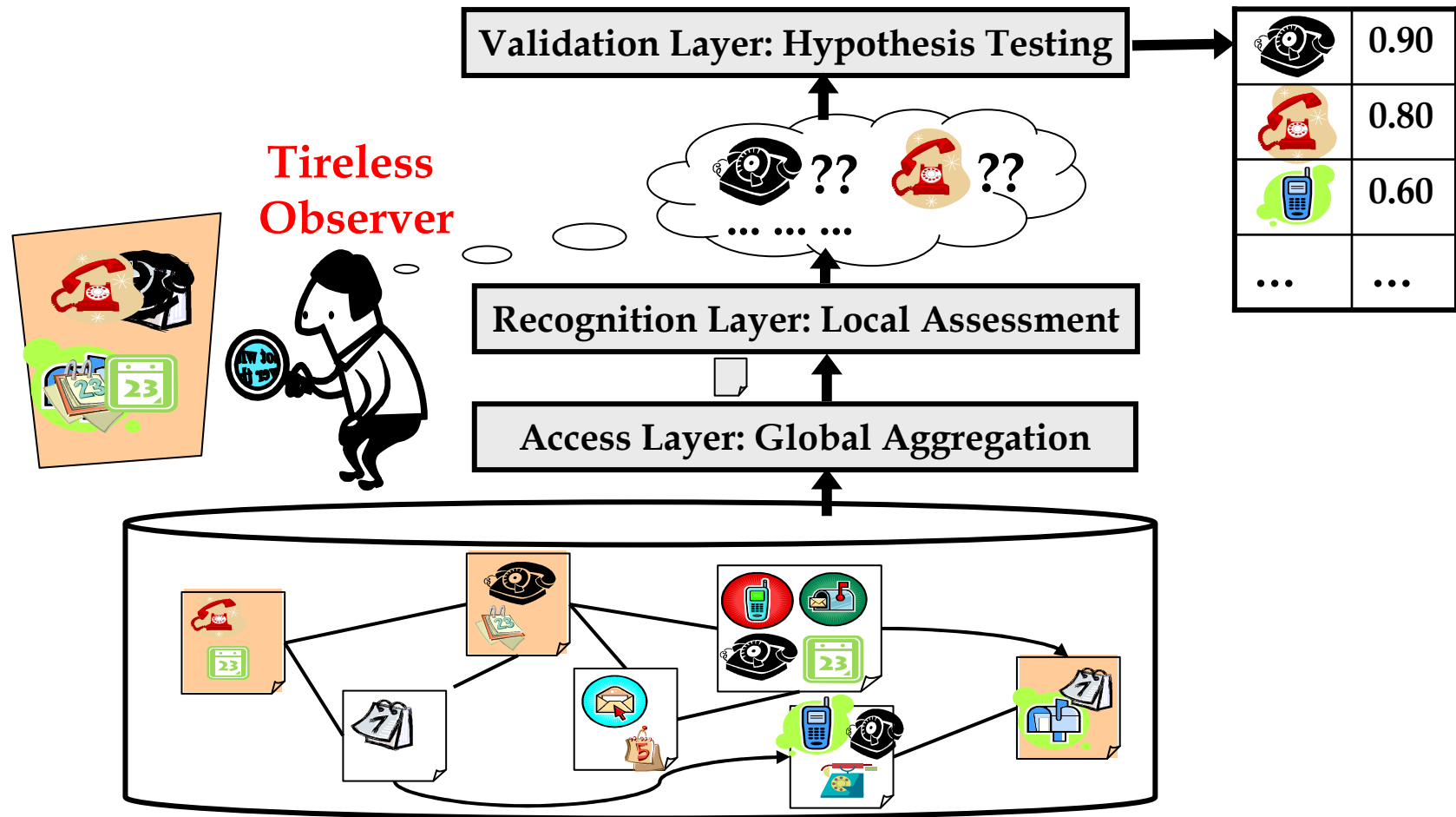
-Tell True Associations from Accidental

- Example: Finding Prof. Luis Gravano's Email

Observation: info@acm.org appears very frequently with keywords "Luis", "Gravano"

However, such association is only accidental as info@acm.org appears on many pages.

EntityRank: The Impression Model



Recognition Layer: Local Assessment

Input:

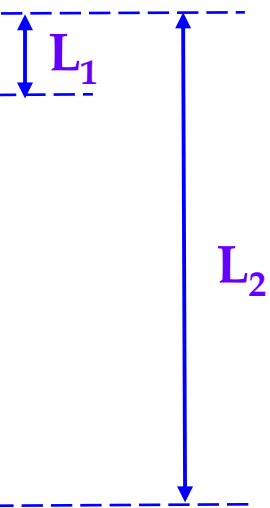
d :

US Amazon Customer Service
 Phone toll-free in the US and Canada:
 (800) 201-7575
 Extraction Conf = 1.0
 E-mail: orders@amazon.com (I think this will still work, but no guarantees)

According to good sources, Amazon is no longer outsourcing much of its customer service work to iSky.

You asked for it! e-Bay's Phone Numbers!

e-Bay, Inc.
 408-376-7400
 Date: 9.02.2006
 Extraction Conf = 0.3



Output:

(800) 201-7575 $p(q((800)201-7575) | d)$

408-376-7400 $p(q(408-376-7400) | d)$

9.02.2006 $p(q(9.02.2006) | d)$

✓	Contextual
✓	Uncertain
	Holistic
	Discriminative
	Associative

Access Layer: Global Aggregation

Input:

d_1

SAN FRANCISCO, 28 December 2003—Casualties of America's drive towards cost-cutting customer service call centers off-shore is one increasingly are hiding their customer service order to prevent dissatisfied customers from

Amazon US Customer Service
1-800-201-7575 (Toll free, US and Canada)

d_2



Gregory (Grisha) Chockler
Research Staff Member,
[IBM Haifa Research Laboratory](#)

During 2003-2005, I was a postdoc in the [Systems](#) group, [MIT/CSAIL](#)

Ph.D., [CS and Eng. School](#), The Hebrew University of Jerusalem

AMAZON.COM customer service phone number (US): (800) 201-7575

d_3

Amazon.com Customer Service Phone
US Amazon Customer Service
Phone toll-free in the US and Canada:
(800) 201-7575

E-mail: orders@amazon.com (I think toll-free guarantees)

$$p(q(800-201-7575) | d_1)$$

$$p(q(800-201-7575) | d_2)$$

$$p(q(800-201-7575) | d_3)$$

Output:

$$p_o = \sum_d p(q(800-201-7575) | d) p(d)$$

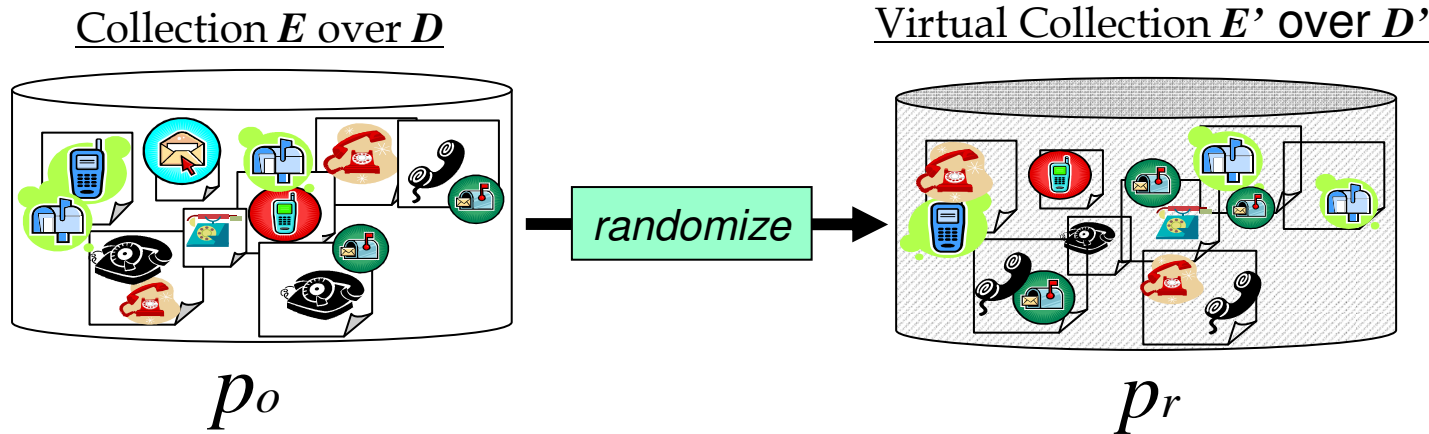
Holistic

Discriminative

✓	Contextual
✓	Uncertain
✓	Holistic
✓	Discriminative
	Associative

Validation Layer: Hypothesis Testing

Input:



Output:

$$\text{Score}(q(t)) = 2\left(p_o \log \frac{p_o}{p_r} + (1 - p_o) \log \frac{1 - p_o}{1 - p_r}\right)$$

✓	Contextual
✓	Uncertain
✓	Holistic
✓	Discriminative
✓	Associative

EntityRank: The Scoring Function

- **Query:** $q(\langle E_1, \dots, E_m \rangle) = \alpha(E_1, \dots, E_m, k_1, \dots, k_l)$ over \mathcal{D}
- **Result:** $\forall t \in E_1 \times \dots \times E_m$: Rank all t by computing $Score(q(t))$ as follows.

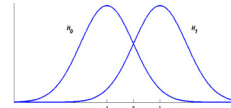
(1) $Score(q(t)) = p_o \cdot \log \frac{p_o}{p_r}$, where

(2) $p_o \equiv p(q(t)|D) = \sum_{d \in D} PR[d] \times \max_{\gamma} \left(\prod_{e_i \in \gamma} e_i.conf \times \alpha_B(\gamma) \times p(s|\gamma) \right)$

(3) $p_r \equiv p(q(t)|D') = \prod_{j=1}^m \left(\sum_{e_j \in d, d \in D} p(d) \right) \times \prod_{i=1}^l \left(\sum_{k_i \in d, d \in D} p(d) \right) \times \prod_{j=1}^m e_j.conf \times \frac{\sum_s p(q(t)|s)}{|s|}$

Validation Global Aggregation Local Recognition

Query Processing



Hypothesis Test → **Result**

800-202-7575: 0.4

800-322-9266: 0.1

Aggregation

800-202-7575: 0.6

800-202-7575: 0.5

800-322-9266: 0.2

Sort-merge Join

Σ



Doc Posting

d ₁	8, 25
d ₃	5
d ₆	10
d ₇	3
d ₉	7, 33

Amazon

Doc Posting

d ₃	11
d ₅	66
d ₇	8, 24

Customer

Doc Posting

d ₃	12
d ₇	9
d ₈	44

Service

Doc Posting

d ₂	(42,851-0400,0.8)
d ₃	(18,800-202-7575,1.0)
d ₇	(13,800-202-7575,1.0)
	(78,800-322-9266,1.0)

#phone

Experiment Setup

- **Corpus:** General crawl of the Web(Aug, 2006), around 2TB with 93M pages.
- **Entities:** Phone (8.8M distinctive instances)
Email (4.6M distinctive instances)
- **System:** A cluster of 34 machines

Comparing EntityRank to the Following Different Approaches

	C ontextual	U ncertain	H olistic	D iscriminative	A ssociative
N aïve			✓		
L ocal	✓	✓			
G lobal			✓	✓	
C ombine	✓	✓	✓		
W ithout	✓	✓	✓	✓	
E ntity R ank	✓	✓	✓	✓	✓

Example Query Results

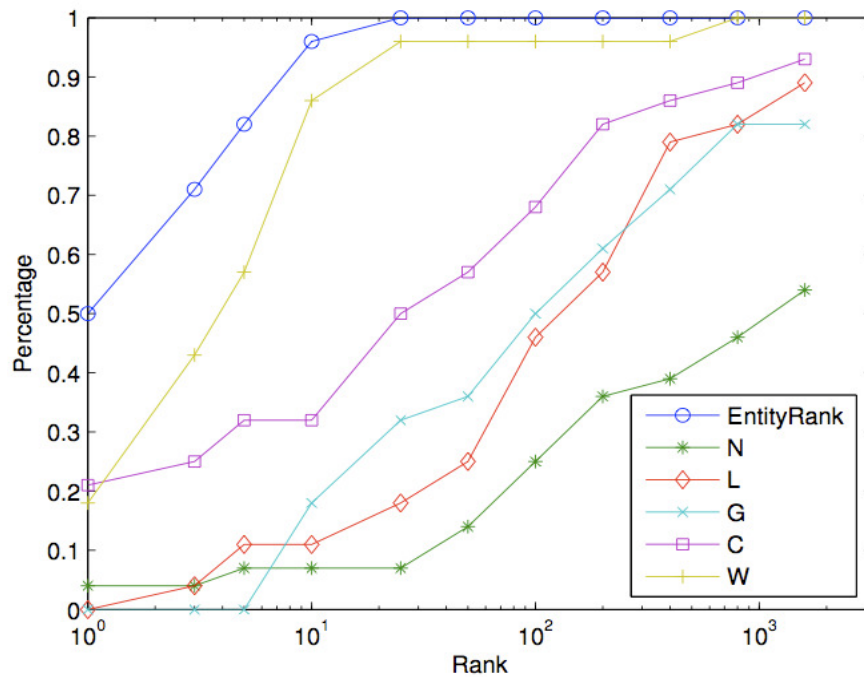
Query	Telephone	ER	L	N	G	C	W
Citibank Customer Service	800-967-2400	1	4	7	43	1	1
New York DMV	800-342-5368	2	2	213	882	5	3
Amazon Customer Service	800-201-7575	1	1	52	83	1	1
Ebay Customer Service	888-749-3229	1	7	859	118	2	13
Thinkpad Customer Service	877-338-4465	5	12	249	127	19	4
Illinois IRS	800-829-3676	1	1	157	697	3	2
Barnes & Noble Customer Service	800-422-7717	1	2	2158	1141	7	1

Query	Email	ER	L	N	G	C	W
Bill Gates	bgates@microsoft.com	4	44	2502	376	21	23
Oprah Winfrey	oprah@aol.com	2	6	745	80	4	3
Elvis Presley	elvis@icomm.com	5	56	1106	267	20	8
Larry Page	larrypage@google.com	8	24	9968	26932	12	11
Arnold Schwarzenegger	governor@governor.ca.gov	4	45	165	169	5	6

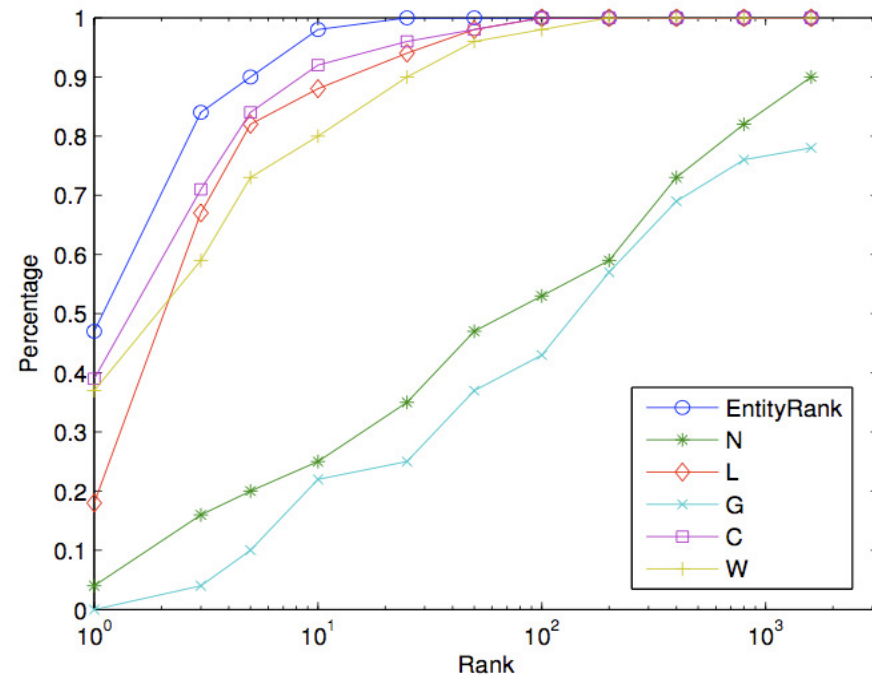
Comparison...

- EntityRank EntityRank
- N Naïve approach
- L Local only
- G Global only
- C Combine L by simple summation
- W L+G without hypothesis testing

%Satisfied Queries at #Rank



Query Type I:
Phone for Top-30 Fortune500 Companies



Query Type II:
Email for 51 of 88 SIGMOD07 PC

Conclusions

- Formulate the entity search *problem*
- Study and define the *characteristics* of entity search
- Conceptual *Impression Model* and concrete *EntityRank* framework for ranking entities
- An online *prototype* with real Web corpus

Thanks!

Questions?