

---

# CS411: Summary and Beyond

---

Kevin C. Chang



**DAIS**

The Database and Information Systems Laboratory

at The University of Illinois at Urbana-Champaign

*Large Scale Information Management*

# About the Final Exam

- Friday 05/06/11 Final Exam 7-10pm 1310 DCL.
- Closed book/note, only scratch paper allowed
- Coverage:
  - Accumulative, with emphasis on 2<sup>nd</sup> half
  - All “special topics” included
- Bring your UIUC ID
- Do not discuss exam on the newsgroup
- Final Tutorial session:  
Wednesday May 4th. 4:30 to 5:30pm in SC1304.

# Final Exam Format

- One set of true/false/short-answer questions
- Followed by several more questions
- Reference: Spring10 Final Exam
  
- 100 points
- I believe you should have enough time (aiming for 120 minutes)



# Suggested Method for Study

- Go over the lecture slides. Study the textbook.
  - Work on problems in hw/lectures.
  - Work on sample exams (see course Web).
  - Questions? Tutorial sessions if solved there.
  - Questions? Office hours.
- 
- Any practice— work it out before looking at solutions
  - Discuss with people in your group

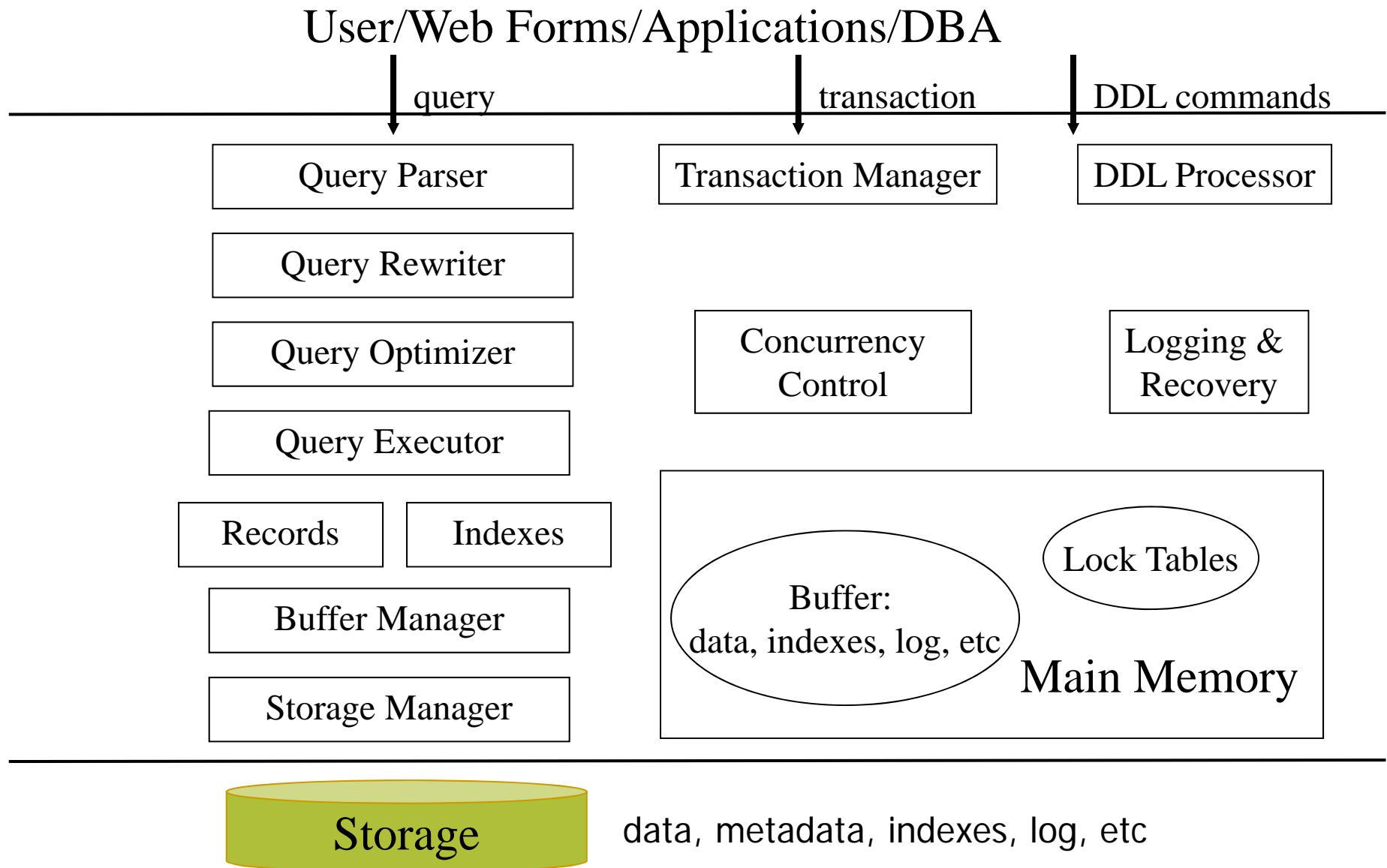
# Motivation:

## Emergence of the Notion of DBMS

- William C. McGee: Generalization: Key to Successful Electronic Data Processing. J. ACM 6(1): 1-23 (1959)
- When data processing was mostly ad-hoc programs— Need generalization, e.g.,
  - sorting
  - file maintenance
  - report generation



# What has resulted-- DBMS Architecture



# Summary 1- First 1/2 Topics: User Perspective

- Entity-Relationship Model
- Relational Model
- Relational Algebra
- Relational Database Design
- SQL and DBMS Functionalities:
  - Queries and Updates
  - Indexes and Views
  - Constraints and Triggers
  - SQL Programming



# Summary 2- Second ½ Topics: System Perspective

- Indexing
- Query Execution
- Query Optimization
- Transaction Management



ILLINOIS

# Results are Impressive.

**Worldwide Vendor Revenue Estimates from RDBMS Software, Based on Total Software Revenue, 2006 (Millions of Dollars)**

Company	2006 Market Share (%)		2005	2005 Market Share (%)	2005-2006 Growth (%)
	2006	2005			
Oracle	7,168.0	47.1	6,238.2	46.8	14.9
IBM	3,204.1	21.1	2,945.7	22.1	8.8
Microsoft	2,654.4	17.4	2,073.2	15.6	28.0
Teradata	494.2	3.2	467.6	3.5	5.7
Sybase	486.7	3.2	449.9	3.4	8.2
Other Vendors	1,206.3	7.9	1,149.0	8.6	5.0
<b>Total</b>	<b>15,213.7</b>		<b>100.0</b>	<b>13,323.5</b>	<b>14.2</b>
				<b>100.0</b>	

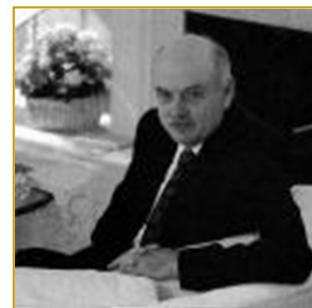
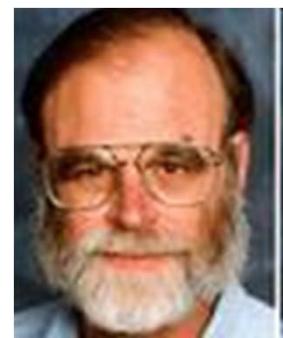
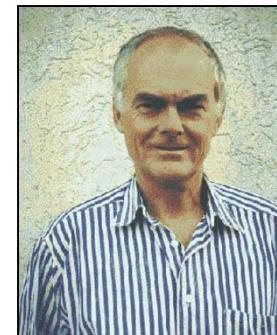
Source: Gartner Dataquest (June 2007)

~ 40<sup>th</sup> GDP as a country.

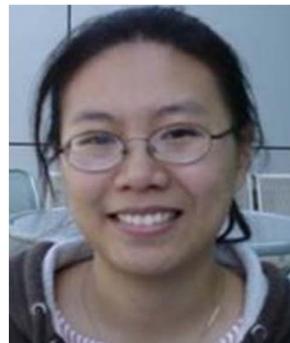


ILLINOIS

Why such great achievements?  
Many things happened behind the scene.



# And Many More Behind the Scene...



ILLINOIS

11

# Special Thanks to the Many Beers... But.



# Beyond: *Data management in the information age*

- Information abounds in our civilization and inundates our daily life.
- Beyond ~~database~~ management:  
*Data management issues everywhere*
- My research projects



# Advanced Database System?

*What is the ultimately advanced DB?*



ILLINOIS

14

How do you greet people in your culture?

What have you been  
*searching* lately?

# What have you been searching lately?

- The **university** and **areas** of Kevin Chang?
- The **email** of Marc Snir?
- Customer service **phone** number of Amazon?
- What **profs** are doing databases at UIUC?
- The **papers** and **presentations** of SIGMOD 2007?
- Due **date** of SIGMOD 2008?
- Sale **price** of “Canon PowerShot A400”?
- “Hamlet” **books** available at bookstores?

# Data of all sorts--- Prevalent on the Web!

**Research**

## Kevin Chen-Chuan Chang

**Departing**

Select	Carrier	Flight #	Departing		Arriving		Aircraft Type	Cabin	AA Flight Miles	Meals	Travel Time
			City	Date & Time	City	Date & Time					
●	AA AMERICAN AIRLINES OPERATED BY AMERICAN EAGLE	4401	CMI Champaign	07/26/2006 08:24 AM	ORD Chicago	07/26/2006 09:14 AM	ER4	Economy <a href="#">View Seats</a>	135	\$11	7 hr 4 min
	AA AMERICAN AIRLINES	781	ORD Chicago	07/26/2006 11:03 AM	SFO San Francisco	07/26/2006 01:28 PM	M83	Economy <a href="#">View Seats</a>	1846	\$11	
●	AA AMERICAN AIRLINES OPERATED BY AMERICAN EAGLE	4401	CMI Champaign	07/26/2006 08:24 AM	ORD Chicago	07/26/2006 09:14 AM	ER4	Economy <a href="#">View Seats</a>	135	\$11	8 hr 44 min
	AA AMERICAN AIRLINES	1655	ORD Chicago	07/26/2006 12:44 PM	SFO San Francisco	07/26/2006 03:08 PM	M83	Economy <a href="#">View Seats</a>	1846	\$11	
●	AA AMERICAN AIRLINES OPERATED BY AMERICAN EAGLE	4401	CMI Champaign	07/26/2006 08:24 AM	ORD Chicago	07/26/2006 09:14 AM	ER4	Economy <a href="#">View Seats</a>	135	\$11	10 hr 44 min

**735 ALVARADO CT  
STANFORD, CA 94305**





**Bedrooms / Bathr**  
Living Area square  
Approximate Lot  
Property Age: 46  
Property Class: R  
MLS #: 638342  
Status: Active

- 10 of 7,148 results for **databases** :

Sort by: **Relevance** [GO!](#)

**1. Fundamentals of Database Systems (5th Edition)**  
by Ramez Elmasri, Shamkant B. Navathe (Hardcover - March 7, 2006)  
Avg. Customer Rating:   
[Rate this item](#)  
Other Editions: [Hardcover](#) | [Textbook Binding](#) | [See all \(8\)](#)

Usually ships in 24 hours  
1st Price: \$100.00 [Used & new from \\$70.43](#)  
Buy new: \$100.00

**2. Databases Demystified (Demystified)**  
by Andrew Oppel (Paperback - March 1, 2004)  
Avg. Customer Rating:   
[Rate this item](#)  
Other Editions: [Paperback](#) | [Digital \(Adobe Reader\)](#) | [See all \(3\)](#)

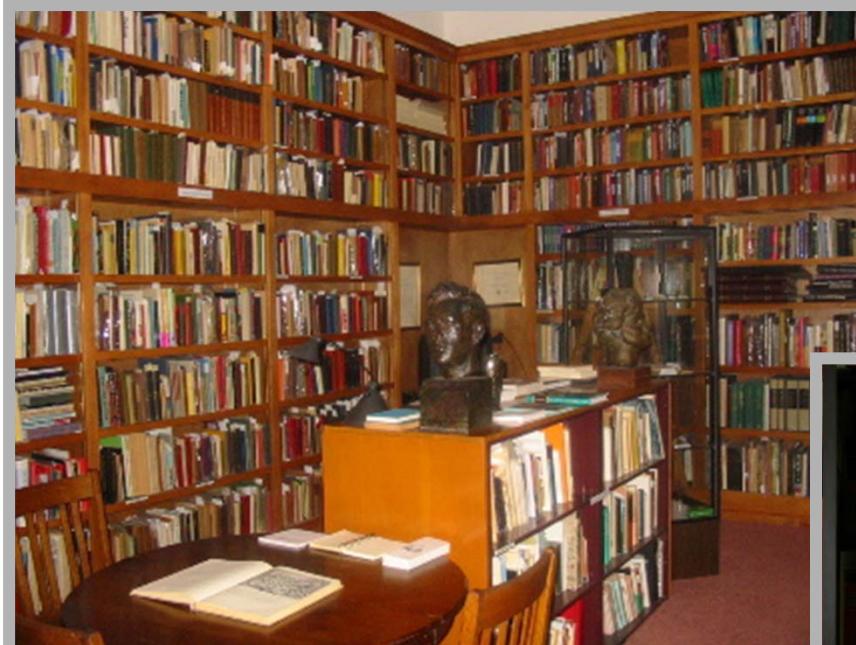
Usually ships in 24 hours  
1st Price: \$49.95 [Used & new from \\$8.92](#)  
Buy new: \$12.97

**Excerpt from Page 1:** "... This chapter introduces fundamental concepts and definitions regarding **databases**, including properties common to **databases**, prevalent **database** models, a brief ..."  
[See more references to databases in this book.](#)

**Surprise me! See a random page in this book.**

**3. Refactoring Databases : Evolutionary Database Design (Addison Wesley Signature Series)**  
by Scott W. Ambler, Pramodkumar J. Sadalage (Hardcover - March 3, 2006)

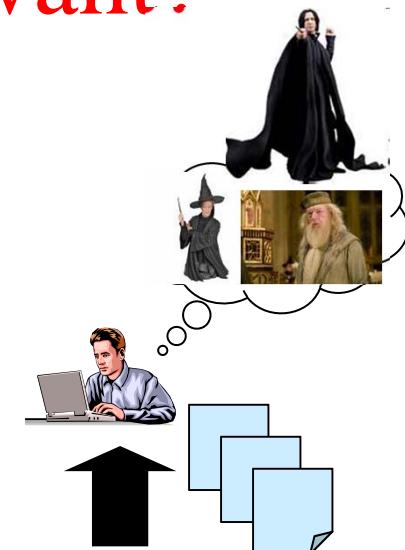
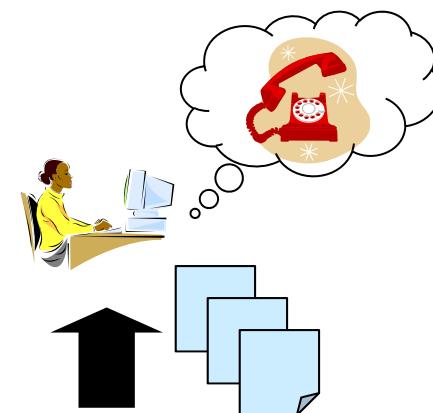
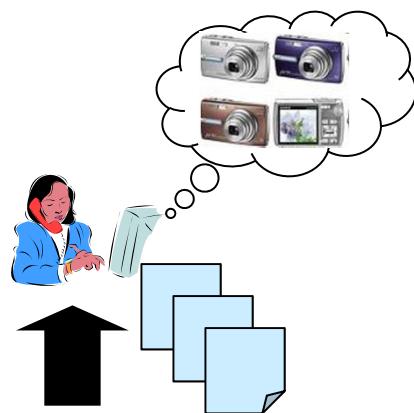
# The Web is a Big Library.



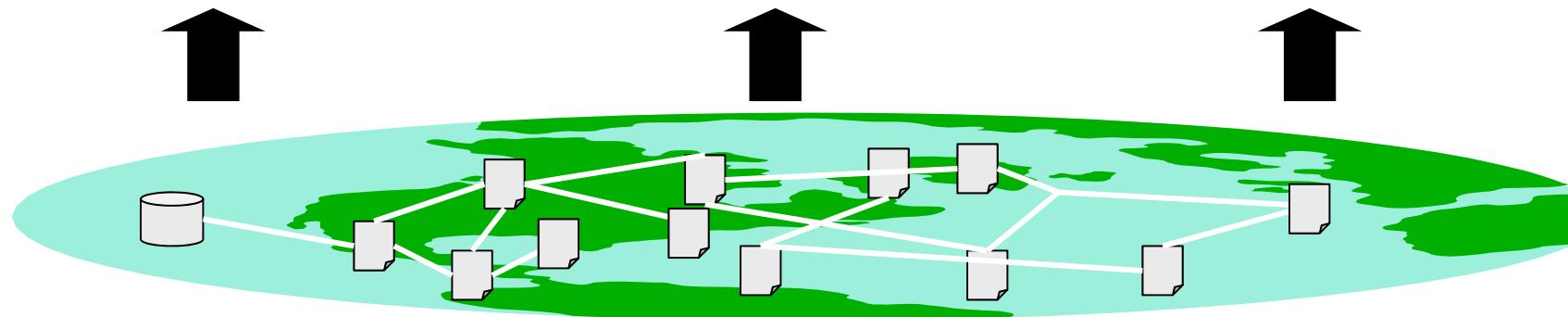
# Huge Supermarket!



# What you search is what you want?



**Search Engine**



## Our View

*Web is “interlinked pages” of “data objects”.*



# What Are We “Still” Working On?

- Although relational *database systems* are almost commodity software now, researchers are even more fascinated by general *data management* issues.
- New challenges naturally arise–
  - databases vs. information retrieval
  - structured vs. unstructured data
  - querying vs. analysis vs. searching
  - closed “base” vs. the open Web

# Today's Search Engine--

- **Only keyword matching--** *Guess what your target page will “say”*
  - e.g., to find Kevin Chang's email
- **Only individual pages --** *Search does not go beyond one page*
  - e.g., to find Kevin Chang's “most likely” email at UIUC
  - e.g., to find all CS profs email
- **Only follow links--** *Databases remain untapped territory*
  - e.g., to find all flights to San Francisco
  - e.g., to find all jobs in Urbana-Champaign

# Getting Structure Data from the Web: *Integration and Mining*

Getting structured data from --

- The “deep” Web
  - semantic-rich, structured data hidden “deeply” inside databases on the Web
  - need “integration” to access these databases
- The “surface” Web
  - semantic-rich, structured data hidden “implicitly” on the surface Web
  - need “mining” to find these relations



# Project 1: MetaQuerier

*Knocking the Door to the Deep Web*

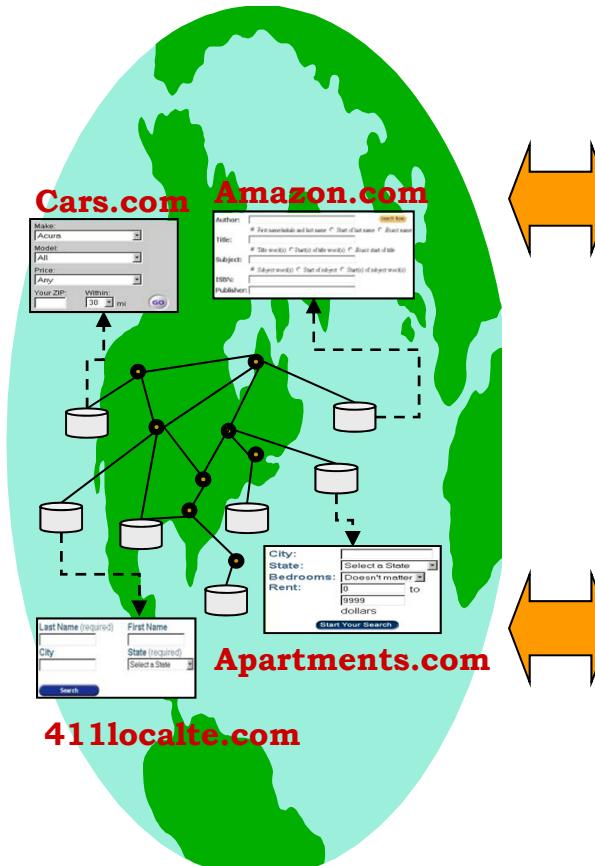


ILLINOIS

25

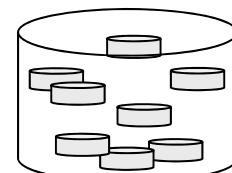
# MetaQuerier:

## *Exploring and integrating deep Web*



### *Explorer*

- source discovery
- source modeling
- source indexing



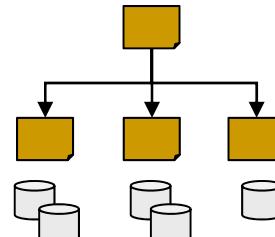
**FIND** sources

db of dbs



### *Integrator*

- source selection
- schema integration
- query mediation



**QUERY** sources

unified query interface



ILLINOIS

26

# Case 1: Find all apartments on the same map.

http://www.apartments.com/Results.aspx?page=results&stype=city&city=champaign&state=il&rad=20

**Cazoodle Apartment BETA** One Search, All Apartments, Entire Web!

City & State or Zipcode: champaign, il      Beds: Studio+      Baths: 1+      Rent: Min to Max      Posted within Anytime

**Search**      **More options**

**Listings 1-10 of 785**      **Sort by** Relevance

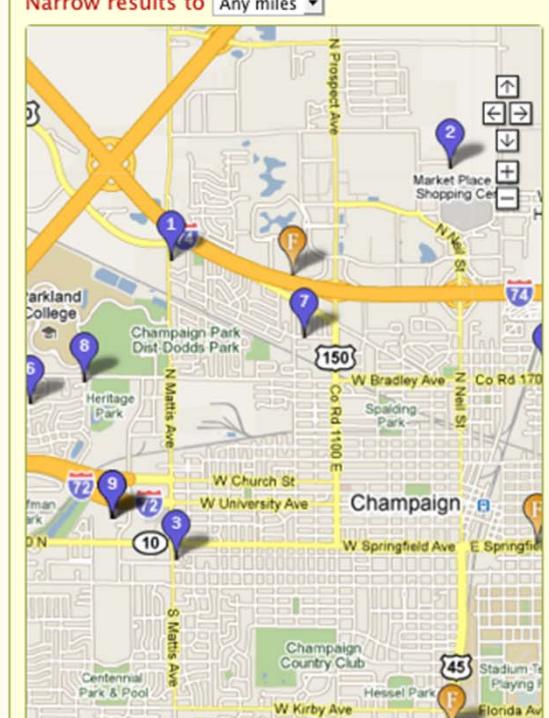
**Apartment Finder** \$839 2 beds  
310 E Springfield Ave. We C.A.T.E.R. services|Convenient Grocery Store|Theatre|Tanning|Wireless 2 Story Atrium Lounge|Bicyc ...  
Posted on 2010-01-29. [Share](#)

**Apartment Finder** \$199 Studio  
1212 W Anthony Dr. Expandable Cable TV & one movie channel|Food and Beverage Vending Machines|Full Size Refrigerator|Fr ...  
Posted on 2010-01-29. [Share](#)

**Apartment Finder** \$550 1 bed  
1710 S Neil St. Ceramic Tile|Garage|Hardwood Floors|Maintenance|Professional Management|Balconies|Carpet|Covered Par ...  
Posted on 2010-01-29. [Share](#)

**Apartment Finder** \$545 2 beds/1 baths  
2000 N Mattis Ave. Kitchen ...  
Posted on 2010-01-31. [More info](#) . [Details](#) . [Share](#)

**Narrow results to** Any miles



1 2 3 4 5 6 7 8 9 10



ILLINOIS

## Case 2: Better prices, more merchants.

**NexTag®**  
Comparison Shopping

Products [Degrees](#)

Most Popular | Top Br

All Categories : Computers : Notebooks : [LENOVO](#)

**Lenovo Thinkpad T400 7**  
2GB DDR3,...

2.0 GHz - Dual Core, Core 2 Duo - 2  
XP Pro

The Lenovo ThinkPad T400 notebook, such as its 9-cell battery, switchable

Part #: 741723U [more ...](#)

**Sellers Found: 7**  
Available Since: 01-23-2009  
Lowest Price: \$907.00

**Compare Prices** [Write Product Review](#)

Sellers	Seller Ratings
California Computer	★★★★★ 120 Seller Reviews
2020_PC	Trusted Seller ★★★★★ 171 Seller Reviews
ANTOnline	Trusted Seller ★★★★★ 3720 Seller Reviews
	Trusted Seller ★★★★★ 544 Seller Reviews

**62 Offers (\$819.48 ~ \$1,299.99)**

Merchant	Merchant Rating	Stock	Condition	Price	Action
Jlpshop.com	★★★★★	In Stock	New	\$819.48	<a href="#">Go to Store</a>
Amazon.com	★★★★★	In Stock	New	\$859.00	<a href="#">Go to Store</a>
J&R	★★★★★	In Stock	New	\$859.00	<a href="#">Go to Store</a>
Jlpshop.com	★★★★★	In Stock	New	\$875.75	<a href="#">Go to Store</a>
PCSuperStore	★★★★★	In Stock	New	\$919.27	<a href="#">Go to Store</a>
CostCentral	★★★★★	In Stock	New	\$923.02	<a href="#">Go to Store</a>
2020PC.COM	★★★★★	In Stock	New	\$933.00	<a href="#">Go to Store</a>
Frontier PC USA	★★★★★	In Stock	New	\$938.99	<a href="#">Go to Store</a>
PCSuperDeals	★★★★★	In Stock	New	\$942.95	<a href="#">Go to Store</a>
Mwave.com	★★★★★	Out of Stock	New	\$944.99	<a href="#">Go to Store</a>
Shopdigitaliving	★★★★★	Out of Stock	New	\$951.49	<a href="#">Go to Store</a>
BullDog SuperStore	★★★★★	In Stock	New	\$952.37	<a href="#">Go to Store</a>

**Lazoodle Shopping BETA Electronics** A Tiny, Yet Powerful Shopping Engine for Electronics!

Laptop > Lenovo Thinkpad T400 741723U

**Lenovo Thinkpad T400 741723U**

Intel Core 2 duo 2.26GHz, Ram 2GB DDR3, HDD 320GB, Screen 14.1", Black

Price history

62 Offers (\$819.48 ~ \$1,299.99)

Merchant	Merchant Rating	Stock	Condition	Price	Action
Jlpshop.com	★★★★★	In Stock	New	\$819.48	<a href="#">Go to Store</a>
Amazon.com	★★★★★	In Stock	New	\$859.00	<a href="#">Go to Store</a>
J&R	★★★★★	In Stock	New	\$859.00	<a href="#">Go to Store</a>
Jlpshop.com	★★★★★	In Stock	New	\$875.75	<a href="#">Go to Store</a>
PCSuperStore	★★★★★	In Stock	New	\$919.27	<a href="#">Go to Store</a>
CostCentral	★★★★★	In Stock	New	\$923.02	<a href="#">Go to Store</a>
2020PC.COM	★★★★★	In Stock	New	\$933.00	<a href="#">Go to Store</a>
Frontier PC USA	★★★★★	In Stock	New	\$938.99	<a href="#">Go to Store</a>
PCSuperDeals	★★★★★	In Stock	New	\$942.95	<a href="#">Go to Store</a>
Mwave.com	★★★★★	Out of Stock	New	\$944.99	<a href="#">Go to Store</a>
Shopdigitaliving	★★★★★	Out of Stock	New	\$951.49	<a href="#">Go to Store</a>
BullDog SuperStore	★★★★★	In Stock	New	\$952.37	<a href="#">Go to Store</a>

# Project 2: WISDM

*Uncovering Structured Data  
on the Surface Web*



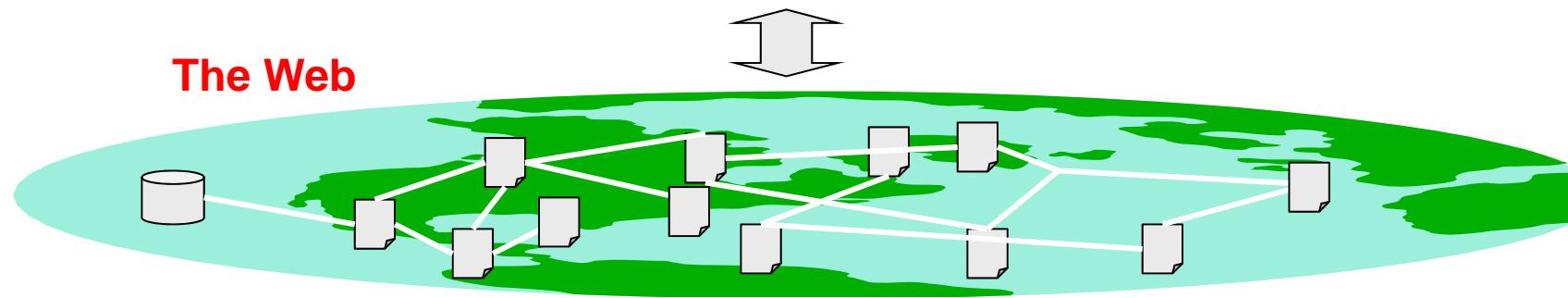
ILLINOIS

29

# The WISDM goal



**WISDM: *Web Indexing and Search for Data Mining***



ILLINOIS

30

So, that's the end of CS411-  
And hopefully the start of  
your DATA career



ILLINOIS

31

Wish you a great summer!

If you are still looking for  
a cool way to spend summer:

jobs @ **Cazoodle** .com



ILLINOIS

32

# *Thank You!*

***For more information:***

<http://www-faculty.cs.uiuc.edu/~kcchang/>  
[kcchang@cs.uiuc.edu](mailto:kcchang@cs.uiuc.edu)



ILLINOIS

33