

Mining Web Structure for Advanced Search

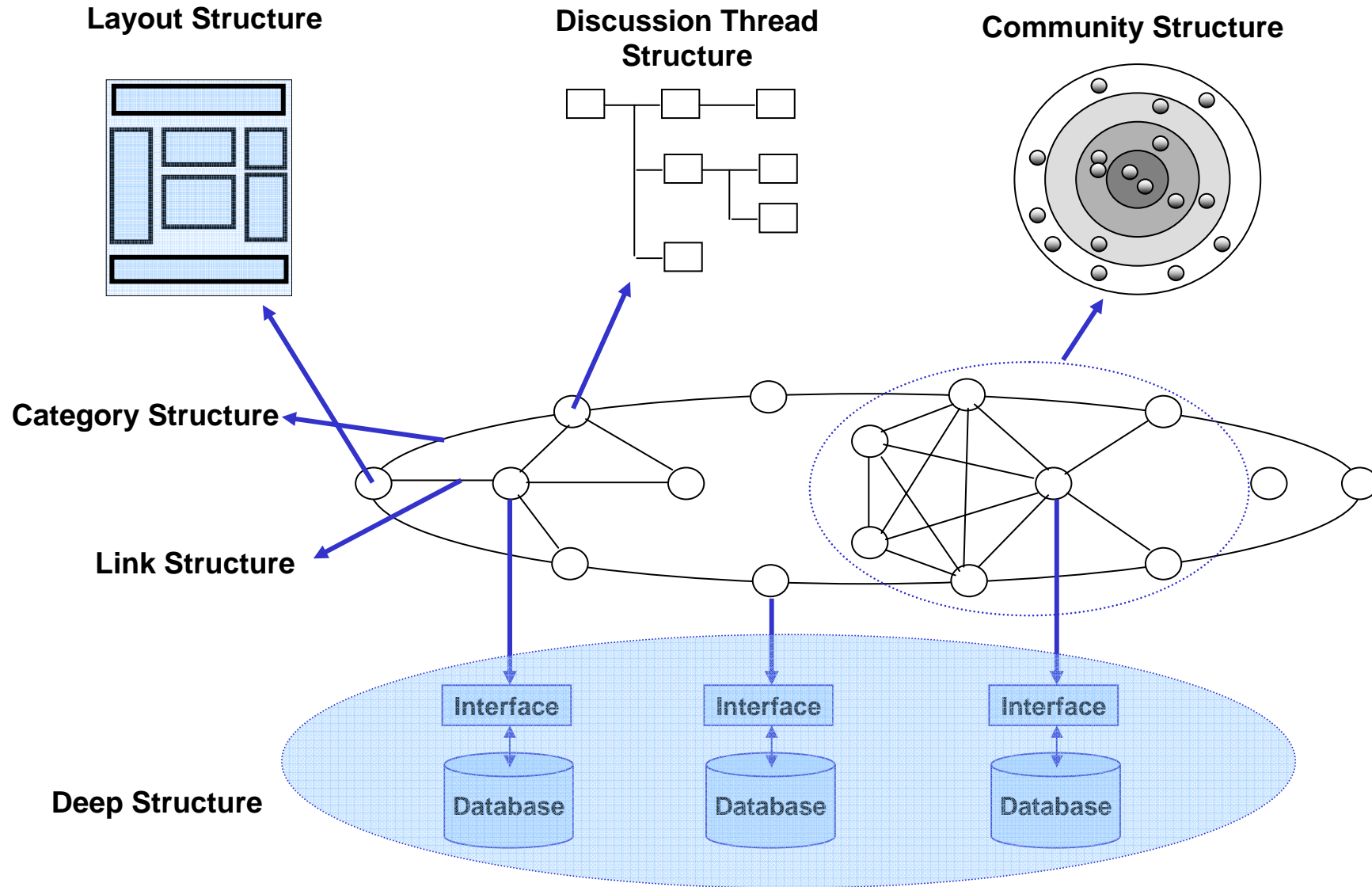
Wei-Ying Ma (马维英 博士)

Research Manager

Web Search & Mining Group

Microsoft Research Asia

Structure of the Web – the Big Picture



Layout Structure

- Compared to plain text, a web page is like a 2D image
 - Visual effects created by rich term types, formats, separators, blank areas, and pictures
 - Different parts of a web page are not equally important

Title: CNN.com International

H1: IAEA: Iran had secret nuke agenda

H3: EXPLOSIONS ROCK BAGHDAD

...

TEXT BODY (with position and font type): The International Atomic Energy Agency has concluded that Iran has secretly produced small amounts of nuclear materials including low enriched uranium and plutonium that could be used to develop nuclear weapons according to a confidential report obtained by CNN...

Hyperlink:

- URL: <http://www.cnn.com/...>
- Anchor Text: Al oaeda...

Image:

- URL: <http://www.cnn.com/image/...>
- Alt & Caption: Iran nuclear ...

Anchor Text: CNN Homepage News ...

Web Page Block – Better Information Unit

Page Segmentation

- Vision based approach

Block Importance Modeling

- Statistical learning



Web Page Blocks

Importance = Low

Importance = Med

Importance = High

Improving PageRank using Layout Structure

- **Z: block-to-page matrix (link structure)**

$$Z_{bp} = \begin{cases} 1/s_b & \text{if there is a link from the } b^{\text{th}} \text{ block to the } p^{\text{th}} \text{ page} \\ 0 & \text{otherwise} \end{cases}$$

- **X: page-to-block matrix (layout structure)**

$$X_{pb} = \begin{cases} f_p(b) & \text{if the } b^{\text{th}} \text{ block is in the } p^{\text{th}} \text{ page} \\ 0 & \text{otherwise} \end{cases}$$

f is the block importance function

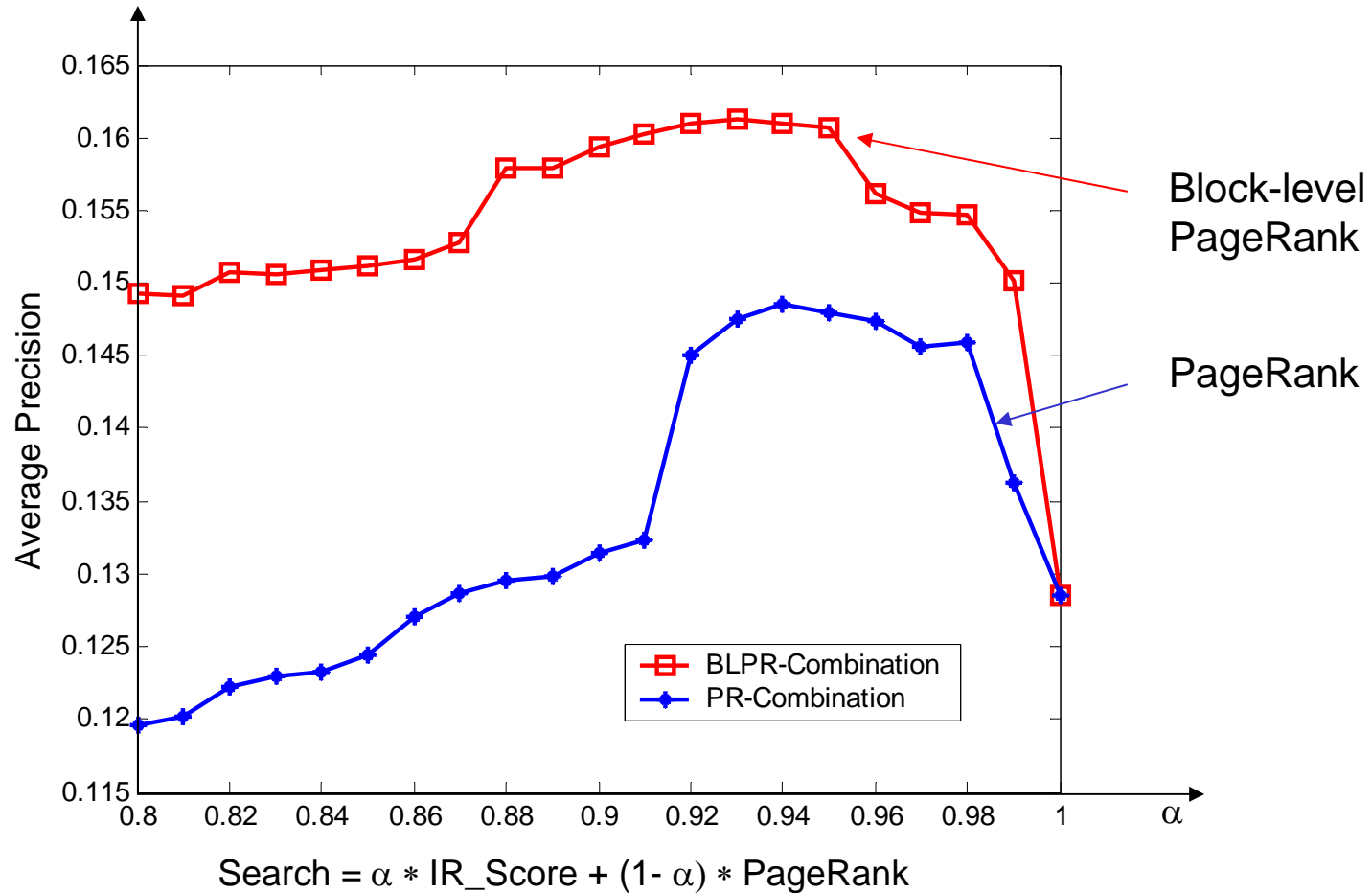
- **Block-level PageRank:**

– Compute PageRank on the page-to-page graph $W_P = XZ$

- **BlockRank:**

– Compute PageRank on the block-to-block graph $W_B = ZX$

Using Block-level PageRank to Improve Search



Block-level PageRank achieves 15-25% improvement over PageRank (SIGIR'04)

Page Layout and Link Analysis for Web Images



Image Graph Model & Spectral Analysis

- **Block-to-block graph:** $W_B = ZX$
- **Block-to-image matrix (container relation):** Y

$$Y_{ij} = \begin{cases} 1/s_i & \text{if } I_j \in b_i \\ 0 & \text{otherwise} \end{cases}$$

- **Image-to-image graph:** $W_I = Y^T W_B Y$
- **ImageRank**
 - Compute PageRank on the image graph
- **Image clustering**
 - Graphical partitioning on the image graph

ImageRank

- Relevance Ranking
- Importance Ranking
- Combined Ranking



Image Clustering (ICME'04)

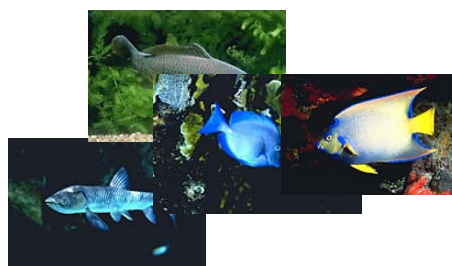
1710 JPG images in 1287 pages are crawled within the website

<http://www.yahooligans.com/content/animals/>

Six Categories



Mammal



Fish



Reptile



Bird



Amphibian



Insect

2-D Embedding of Web Images (ICME'04)

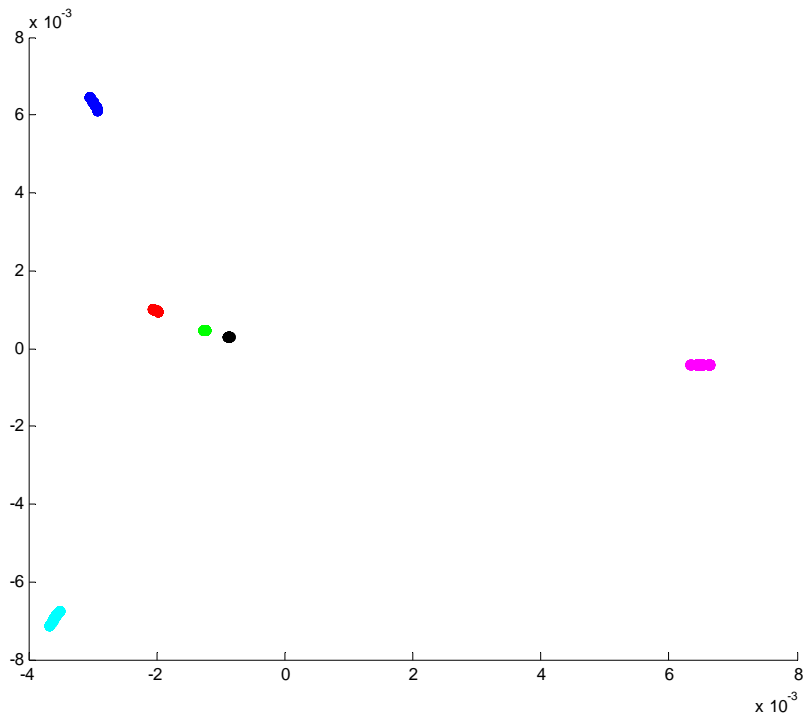


Image graph constructed from
block-level link analysis

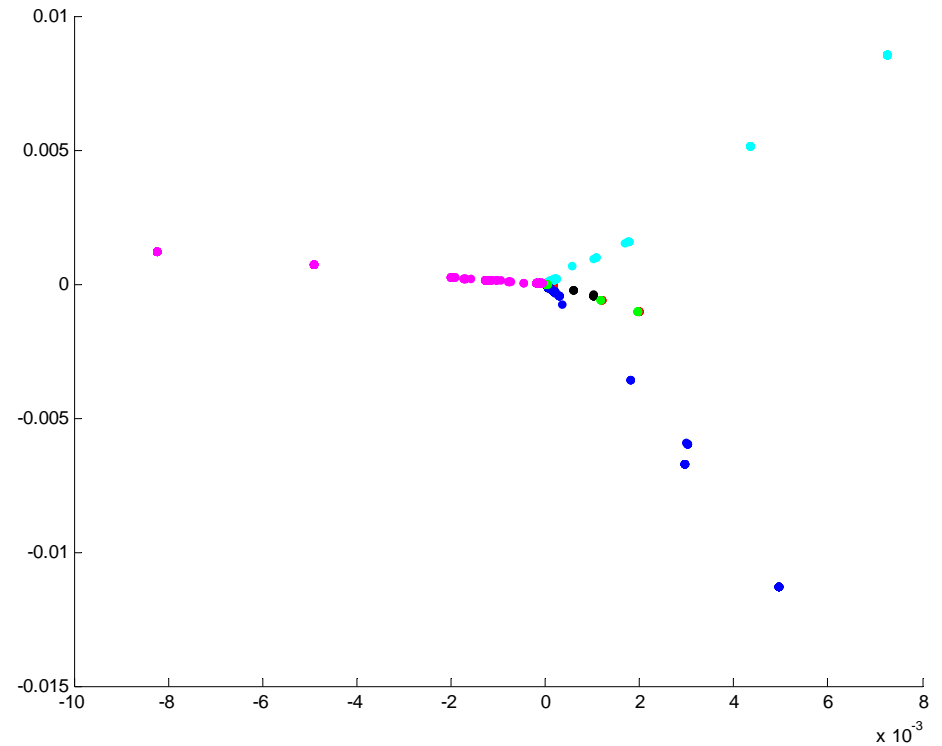
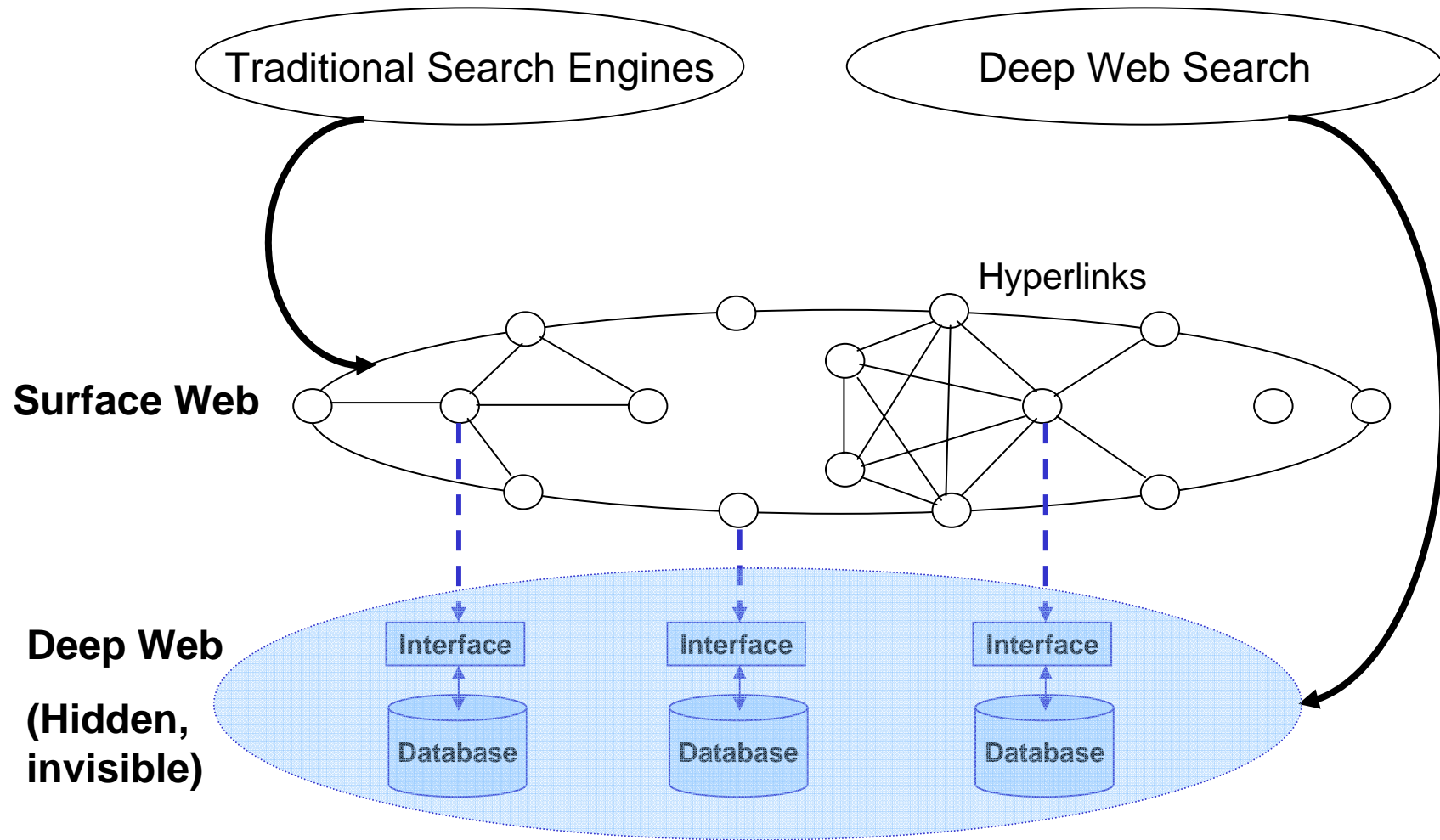
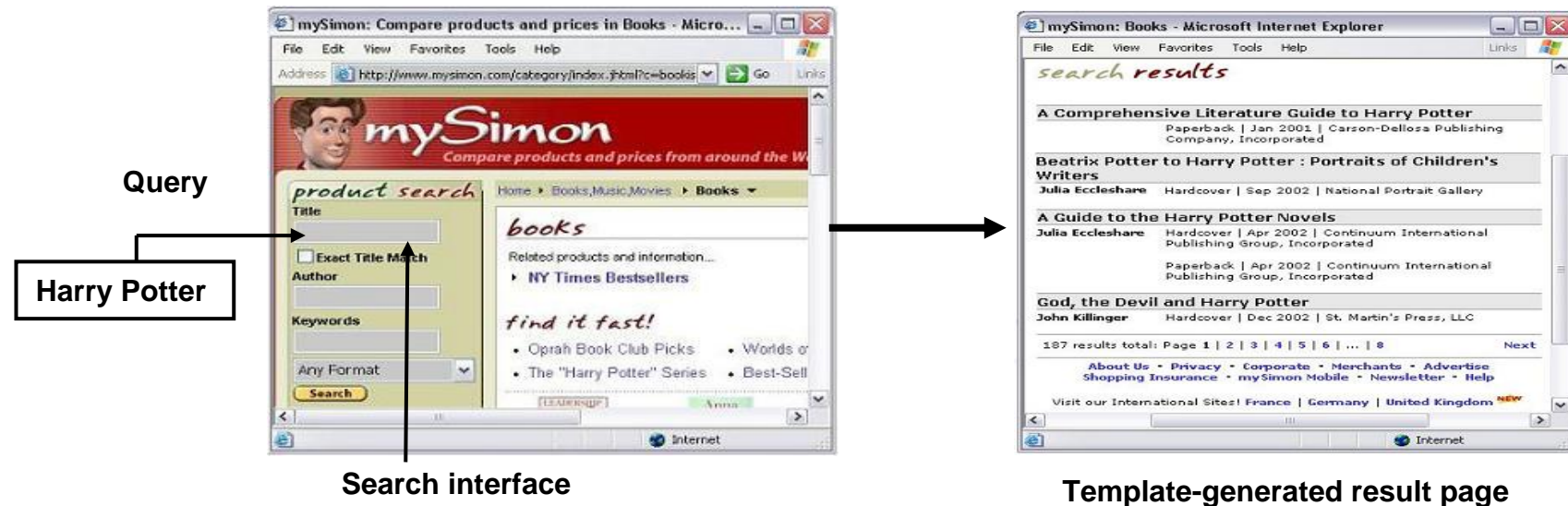


Image graph constructed from
traditional page level link analysis

Searching the Deep Web

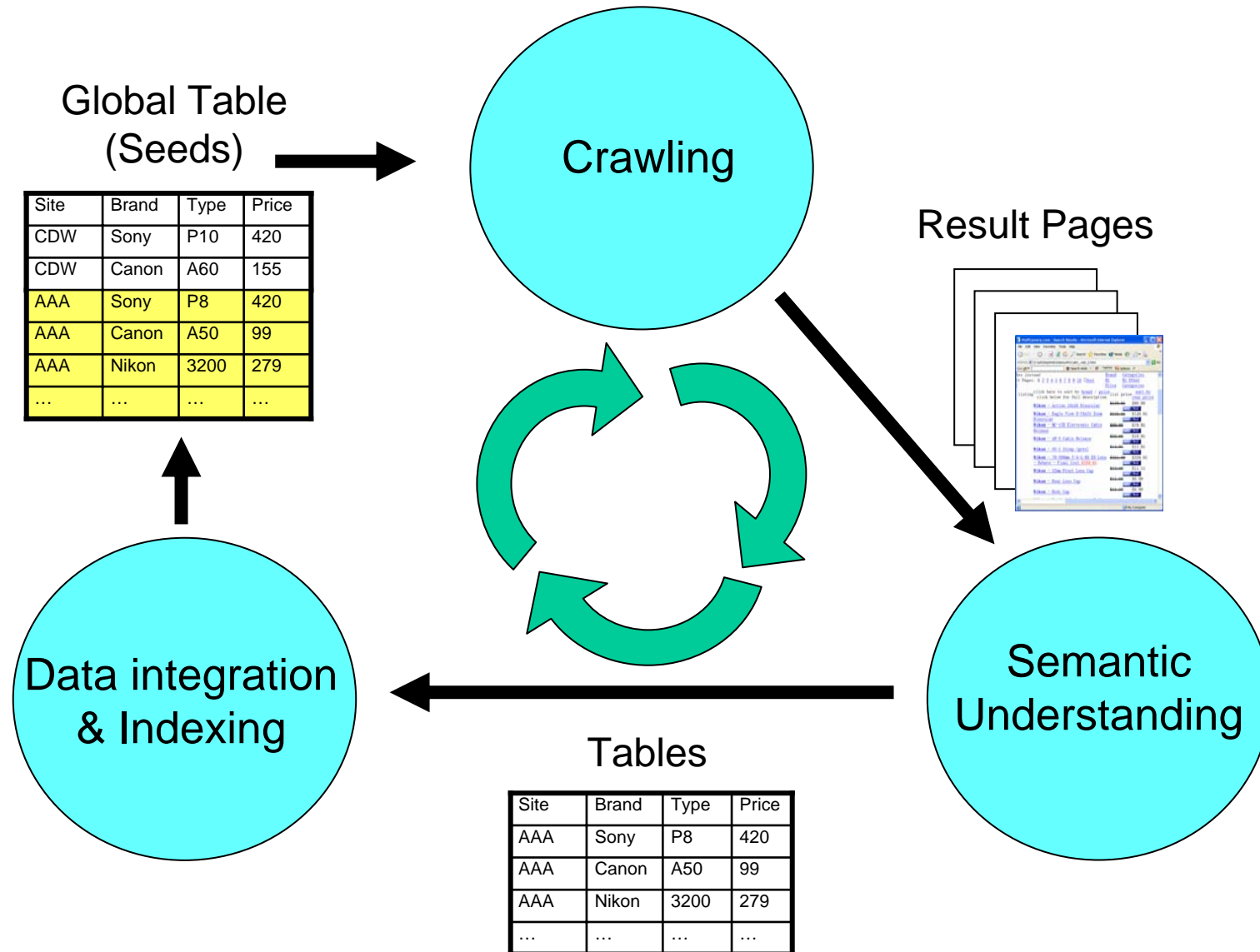


Deep Web Search – Technical Challenges



- How to crawl the data?
 - Data are only accessible by querying the search interfaces
- How to structure the data from a deep web site?
 - Structure information is lost in generated HTML page
- How to consolidate the data from multiple deep web sites?
 - Data integration & indexing

Crawling & Mining the Deep Web (VLDB'04)



Semantic Understanding of a Deep Web Site

- 3-Layer Schemas
 - *Global Schema (GS)*
 - Manually defined OR automatically generated
 - *Interface Schema (IS)*
 - Extracted through form analysis
 - *Result Schema (RS)*
 - Extracted by wrappers
- Schema Matching

Global Schema

Book < Title, Author, Publisher, ISBN >

Search Interface

Writer :

Title :

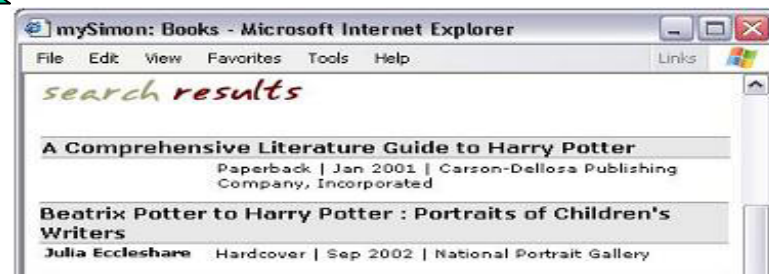
Keyword :

Any Format













GO !

Result Schema

Rowling	Harry potter	Avon Ltd	...
Julia ...	A Guide to ...	Scholastic	...
...



Structured Data Extraction

<p>Staff</p> <ul style="list-style-type: none"> • 2-8514, Heavy005, Glen Bancke (CTSR, Engineer) • 2-6097 Room 216, Loides Collazo (Garcia Center, Project Staff Associate) • 2-8480, Room 105, John Guttler (CTSR, Engineer) • 2-4174, Room 314, Gertha Benoit-Hollis (Dept. Asst to Chair) • 2-8501, Room 311, Kim-Kong Huang (Xray, Scientist) • 2-8480, Room 105, Sean Kubik (CTSR, Engineer) • 2-8480, Room 105, Josh Margolies (CTSR, Engineer) • 2-8484, Room 314, Debbie Michienzi (Dept. Staff Assistant) • 2-6663, Room 217, Jim Quinn (Dept. Dir of Lab) • 2-4567, Room 228, Lynn Russo (CTSR, Engineer) • 2-8480, Room 105, Eileen Zappia (CTSR, Administrative Assistant) <p>Post Docs and Research Scientists</p> <ul style="list-style-type: none"> • 2-8515, HwyE 105, Martin Friis (CTSR) • 2-8501, Room 311, Dhanaraj G (Xray) • 2-4553, Room 216, Shouren Ge (Garcia) • 2-1809, xxxxxxxx, Balvinder Gogia (RD, S-CAT) • 2-8501, Room 311, X Huang (Xray) • 2-xxxx, Room 218, Boris (Dave) Kharas (RD/CTSR) • 2-3209, Room 216, Tadrozzi Kova (Garcia) 	<p>LYCOS PEOPLE SEARCH</p> <p>Find great details on anyone. White pages, web results, professional profile, alumni info, and more...</p> <ul style="list-style-type: none"> White Pages Plus Web Background Check Professional Profile Alumni Email <p>First Name or Initial: _____ Last Name: _____</p> <p>Street # and Name: _____</p> <p>Unit #: (e.g. Apt) _____ City: _____</p>	<p>Browse categorically:</p> <p>Whats: All Categories: PCs: Multimedia & Graphics</p> <p>Word list for Multimedia & Graphics</p> <p>101 - 200 of 385 terms.</p> <ul style="list-style-type: none"> > ditect (searchSmall@t) > display (whats) > display_mode (searchSmall@t) > display_modes (searchSmall@t) > dither (searchSmall@t) > dithered (searchSmall@t) > dithering (searchSmall@t) > document_reader_ (whats) > dogcog (whats) > Dolby Digital (searchSmall@t) > Double Density CD (searchStorage) > Double Density Compact Disk (searchStorage) > drop_shadow (whats) > DTV (searchSmall@t) > HDC (searchSmall@t) > HDCP (searchSecurity) > HDMI (searchSmall@t) > Hewlett-Packard Graphics Language (searchCD) > High Definition Compatible Digital (searchSmall@t) > High Definition Multimedia Interface (searchSmall@t) > High bandwidth Digital Content Protection (searchSecurity) > hologram (whats) > homecam (searchMobileComputing) > HP-GL (searchCD) > HPSL (searchCD) > USB (whats) > hue_saturation_and_brightness (whats) 			
<p>Phone Books</p>	<p>People Finders</p>	<p>Dictionary Definitions</p>			
<p>Digital Camera Deals</p> <table border="0"> <tr> <td>  <p>\$63.99</p> <p>Mustek 2.1MP W/ Li-ion Battery Digital Camcorder, Model GSmart MINI 3</p> <p>FedEx Saver Shipping \$5.99</p> </td> <td>  <p>\$69.99</p> <p>Weekend Special! Mercury 3.1 Deluxe Classic Cam, 4X DG, 1.5" LCD, SD Card Digital Camera</p> <p>FedEx Saver Shipping \$5</p> </td> <td>  <p>\$76.00</p> <p>Weekend Special! DXG USA DXG-308 3.1MP, 4xDS, 1.5" LCD, SD/MMC Digital Camera</p> <p>Free FedEx Saver Shipping</p> </td> </tr> </table>	 <p>\$63.99</p> <p>Mustek 2.1MP W/ Li-ion Battery Digital Camcorder, Model GSmart MINI 3</p> <p>FedEx Saver Shipping \$5.99</p>	 <p>\$69.99</p> <p>Weekend Special! Mercury 3.1 Deluxe Classic Cam, 4X DG, 1.5" LCD, SD Card Digital Camera</p> <p>FedEx Saver Shipping \$5</p>	 <p>\$76.00</p> <p>Weekend Special! DXG USA DXG-308 3.1MP, 4xDS, 1.5" LCD, SD/MMC Digital Camera</p> <p>Free FedEx Saver Shipping</p>	 <p><i>The Finding of Moses.</i> 1904. Oil on canvas. Private Collection, UK. More... TO BUY THIS PRINT CLICK HERE</p>  <p><i>Bacchante.</i> 1907. Oil on panel. Private Collection, UK.</p>  <p><i>Caracalla and Geta, Bear Fight in the Coliseum: AD 203.</i> 1907. Oil on panel. 123 x 154 cm. Private collection.</p>	<p>2. <i>Here for the Party</i> by Gretchen Wilson</p> <p>Usually ships in 24 hours (May 11, 2004) Audio CD</p> <p>Price: \$13.49 You Save: \$5.49 (29%)</p> <p>LOW PRICE!</p> <p>Click here for more info</p> <p>3. <i>Julie Roberts</i> by Julie Roberts</p> <p>Usually ships in 24 hours (May 25, 2004) Audio CD</p> <p>Price: \$11.99 You Save: \$1.99 (14%)</p> <p>Click here for more info</p>
 <p>\$63.99</p> <p>Mustek 2.1MP W/ Li-ion Battery Digital Camcorder, Model GSmart MINI 3</p> <p>FedEx Saver Shipping \$5.99</p>	 <p>\$69.99</p> <p>Weekend Special! Mercury 3.1 Deluxe Classic Cam, 4X DG, 1.5" LCD, SD Card Digital Camera</p> <p>FedEx Saver Shipping \$5</p>	 <p>\$76.00</p> <p>Weekend Special! DXG USA DXG-308 3.1MP, 4xDS, 1.5" LCD, SD/MMC Digital Camera</p> <p>Free FedEx Saver Shipping</p>			
<p>Items for Sale</p>	<p>Digital Exhibits</p>	<p>Multimedia & Graphical Files</p>			


Existing Approaches


- Basic Idea
 - Convert HTML into a sequence of tokens or a tag tree
 - Discover pattern
- Representative Methods
 - Wrapper generation
 - Manually write a wrapper
 - Induct a wrapper [Liu 2000], [Kushmerick 1997]
 - Extract structured data from Web pages that shared a common template
 - Equivalence classes [Arasu 2003]
 - RoadRunner [Crescenzi 2001]
 - Extract data record within a Web page
 - OMINI: record-boundary discovery
 - IEPAD: pattern discovery on PAT tree
 - MDR: repeated nodes discovery
 - Extract data from tables in a Web page
 - Classify tables into genuine table or non-genuine table [Wang 2002]
 - Extract data from data tables [Chen 2002], [Lerman 2001]

Problems with Existing Approaches

- Information from HTML token sequence or tag tree is not reliable
 - Error-prone
 - Same tag trees could present totally different objects
- Visual cues are much more reliable but not used
 - Records are put into several grids and aligned
 - Attributes are displayed in regular styles and aligned

Object Block & Object Element

 [DXG-308 DXG Technology USA 3.1 Megapixel Digital Camera with 1.5...](#)
\$79.12 - [Compare](#)
 IMAGE RESOLUTION UP TO 2976 X 2232 INTERPOLATED 4X DIGITAL ZOOM USES SD OR MMC MEMORY CARD UP TO 512MB RECORDS MOVIE FILES FOR 80 SECONDS IN 512 X 384 USES 2 ...
 The Twister Group: [4.4 / 5](#)



 [DXG USA - DXG-308 3.1 MP Ultra Compact Slim Digital Camera](#)
\$89.89
 ... Cable User's Manual **DXG USA - DXG-308** 3.1 MP Ultra Compact Slim Digital Camera With 1.5 LCD # DXG308, **DXG-308**, D-XG308, DX-G308, **308 DXG USA - DXG-308** 3.1 MP ...
 A1PlusElectronics.com

 [DXG DXG-308 / 3.1 Megapixel / 4x Digital Zoom / Digital Camera ...](#)
\$98.45
 ... to your computer or television fast and easy with the **DXG-308's** USB and ... comes with all manufacturer supplied accessories, and full manufacturer's **USA** warranty.
 Sale Stores Corp.: [3.2 / 5](#)



Element

Object Block

RcdBlock1	RcdBlock2	RcdBlock3
		
<p>DXG-308 DXG Technology USA 3.1 Megapixel Digital Camera with 1.5... IMAGE RESOLUTION UP TO 2976 X 2232 INTERPOLATED 4X DIGITAL ZOOM USES SD OR MMC MEMORY CARD UP TO 512MB RECORDS MOVIE FILES FOR 80 SECONDS IN 512 X 384 USES 2 ... The Twister Group: 4.4 / 5</p>	<p>DXG USA - DXG-308 3.1 MP Ultra Compact Slim Digital Camera ... Cable User's Manual DXG USA - DXG-308 3.1 MP Ultra Compact Slim Digital Camera With 1.5 LCD # DXG308, DXG-308, D-XG308, DX-G308, 308 DXG USA - DXG-308 3.1 MP ... A1PlusElectronics.com</p>	<p>DXG DXG-308 / 3.1 Megapixel / 4x Digital Zoom / Digital Camera to your computer or television fast and easy with the DXG-308's USB and ... comes with all manufacturer supplied accessories, and full manufacturer's USA warranty. Sale Stores Corp.: 3.2 / 5</p>

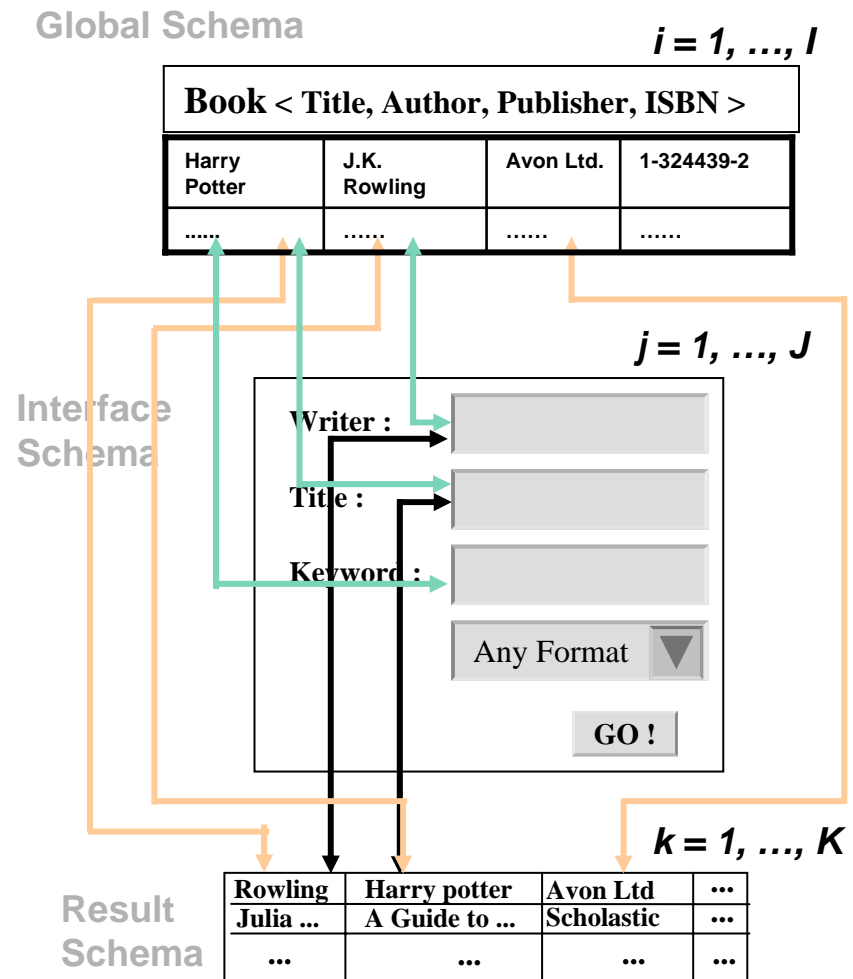
  \$63.99 Mustek 2.1MP W/ Li-ion Battery Digital Camcorder, Model GSmart MINI 3 FedEx Saver Shipping \$5.99 ADD TO CART	  \$69.99 Weekend Special! Mercury 3.1 Deluxe Classic Cam,4X DG, 1.5"LCD,SD Card Digital Camera FedEx Saver Shipping \$5 ★★★★★ ADD TO CART	  \$76.00 Weekend Special! DXG USA DXG-308 3.1MP, 4xDG, 1.5" LCD, SD/MMC Digital Camera Free Fedex Saver Shipping ★★★★★ ADD TO CART
---	--	---



RcdBlock1	RcdBlock2	RcdBlock3
		
		
\$63.99	\$69.99	\$76.00
<p>Mustek 2.1MP W/ Li-ion Battery Digital Camcorder, Model GSmart MINI 3</p>	<p>Weekend Special! Mercury 3.1 Deluxe Classic Cam,4X DG, 1.5"LCD,SD Card Digital Camera</p>	<p>Weekend Special! DXG USA DXG-308 3.1MP, 4xDG, 1.5" LCD, SD/MMC Digital Camera</p>
FedEx Saver Shipping \$5.99	FedEx Saver Shipping \$5	Free Fedex Saver Shipping
ADD TO CART	★★★★★ ADD TO CART	★★★★★ ADD TO CART

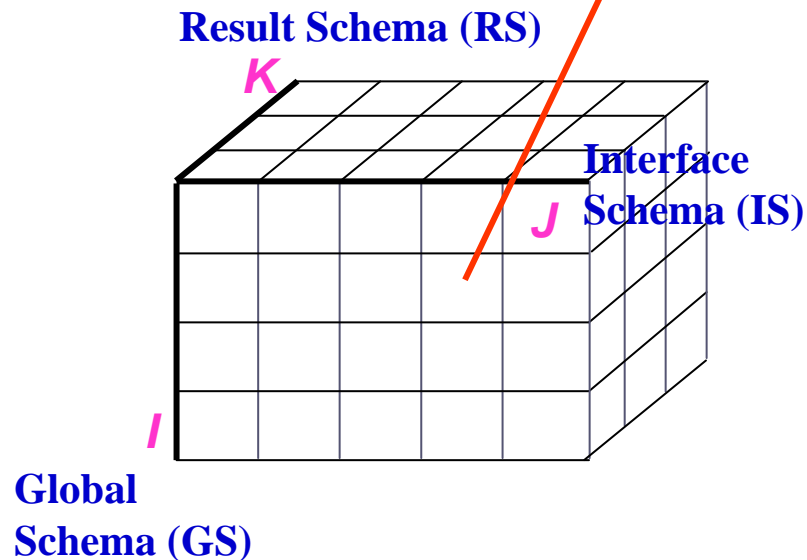
Schema Matching

- Instance-based probing
 - Content overlap is bigger between the same or similar attributes
- Method
 - **Input:**
 - A global schema for the domain
 - Some seed sample instances
 - **Probing**
 - Exhaustively submit all attribute values of the sample records to each input element of the search interface
 - Count the reappearance of each query value in the result pages
 - **Matching**



Schema Matching (Cont)

- Schema Cube



The total number of re-appeared query values in a result column after the exhaustive probing

- Project the Schema Cube to 3 matrices to identify the mapping between a pair of schemas

Mappings in the Matrices

- Mutual Information: $I(A, B) = \Pr(A \wedge B) \times \log \frac{\Pr(A \wedge B)}{\Pr(A) \times \Pr(B)}$

Global Schema

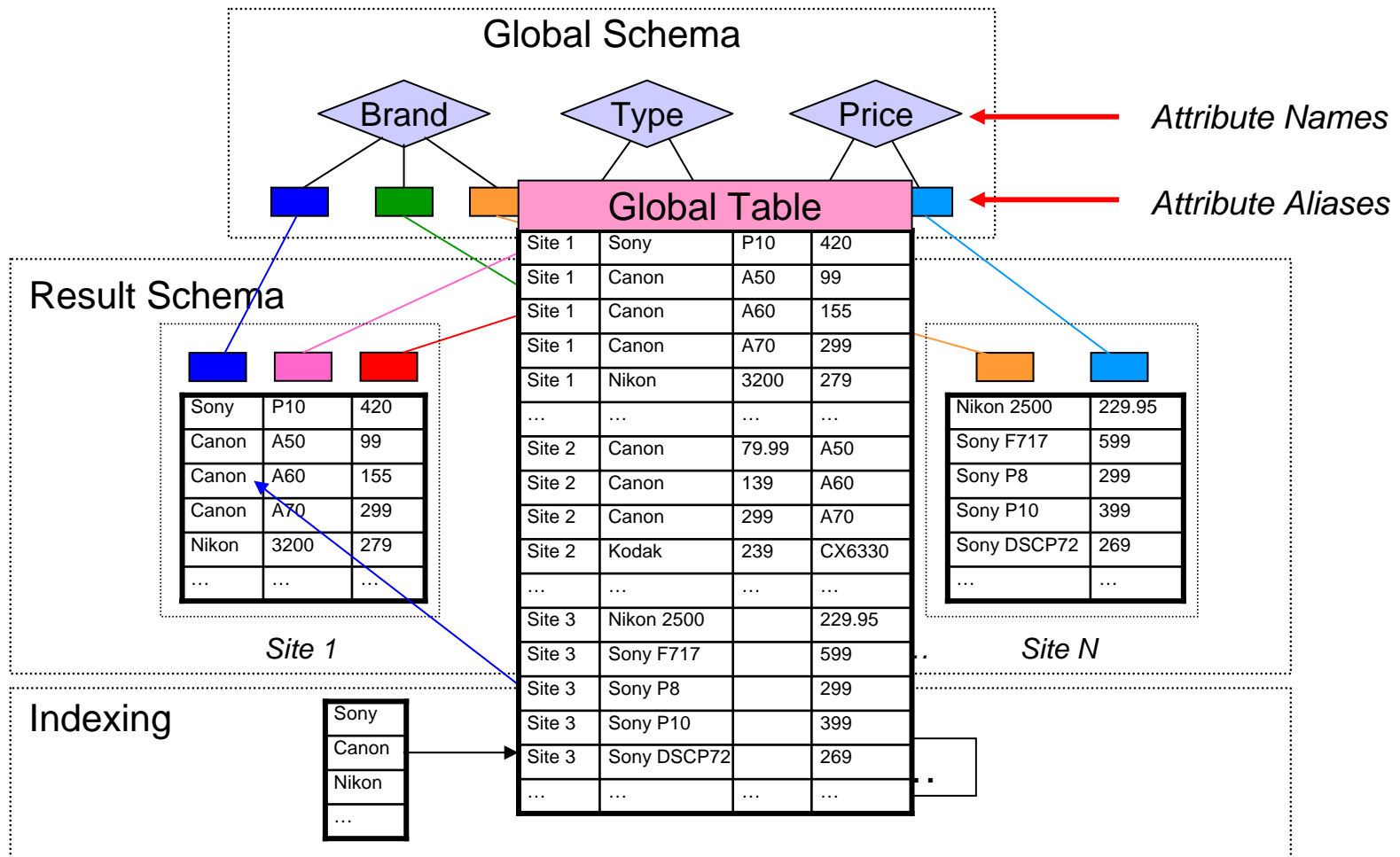
	T	A	P	I																	
Keyword	226	409	381	0	\Rightarrow <table border="0"> <tr> <td>0.012</td> <td>0.020</td> <td>-0.026</td> <td>0</td> </tr> <tr> <td>-0.019</td> <td>0.042</td> <td>0.004</td> <td>0</td> </tr> <tr> <td>-0.010</td> <td>-0.008</td> <td>0.086</td> <td>0</td> </tr> <tr> <td>0.060</td> <td>-0.005</td> <td>-0.031</td> <td>0</td> </tr> </table>	0.012	0.020	-0.026	0	-0.019	0.042	0.004	0	-0.010	-0.008	0.086	0	0.060	-0.005	-0.031	0
0.012	0.020	-0.026	0																		
-0.019	0.042	0.004	0																		
-0.010	-0.008	0.086	0																		
0.060	-0.005	-0.031	0																		
Author	44	427	434	0																	
Publisher	45	8	413	0																	
Title	373	236	218	0																	

Interface Schema

\Downarrow

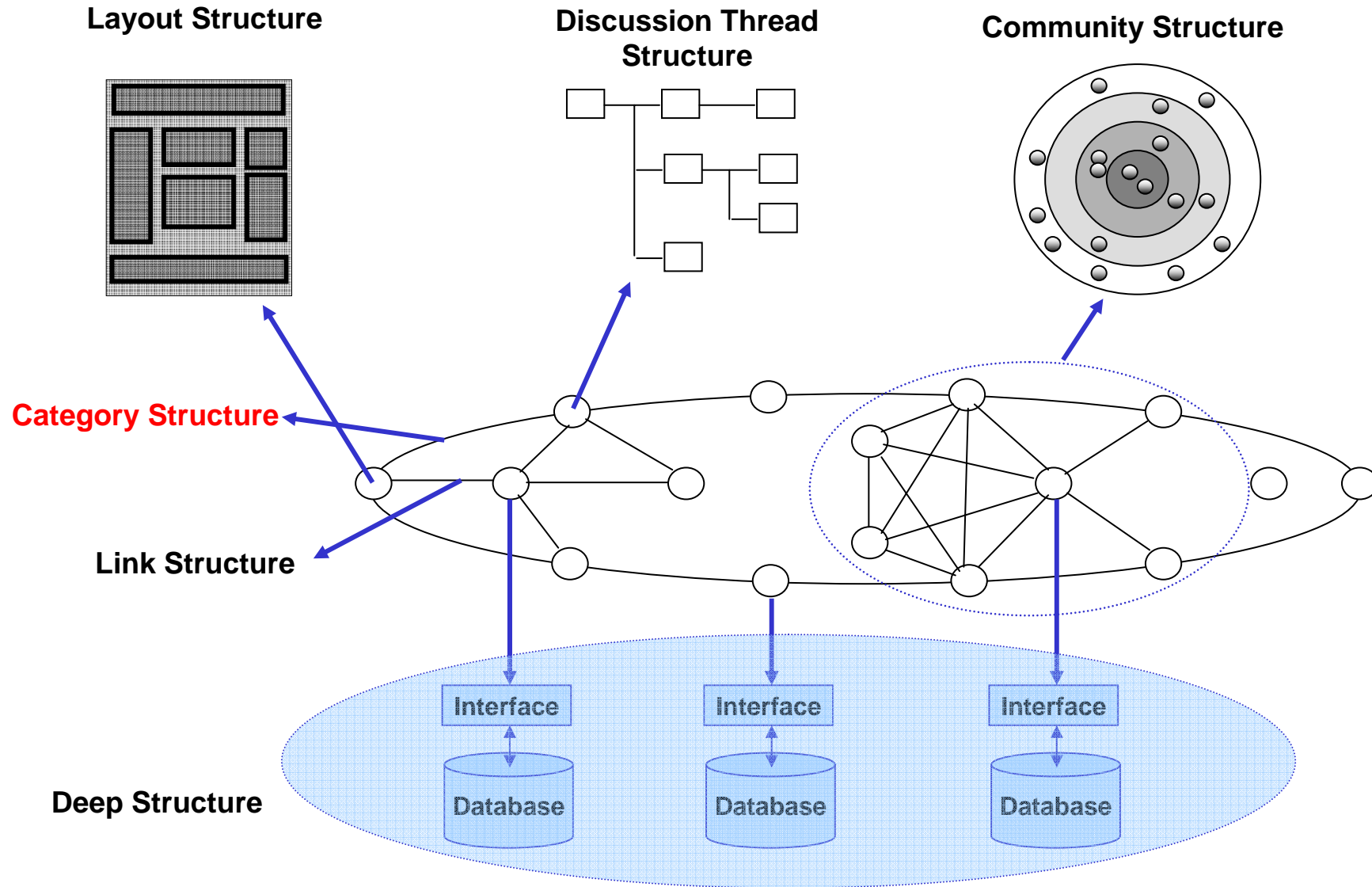
Mapping Result	=	<table border="0"> <tr> <td>0</td> <td>0</td> <td>0</td> <td>0</td> </tr> <tr> <td>0</td> <td>1</td> <td>0</td> <td>0</td> </tr> <tr> <td>0</td> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>1</td> <td>0</td> <td>0</td> <td>0</td> </tr> </table>	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0
0	0	0	0															
0	1	0	0															
0	0	1	0															
1	0	0	0															

Data Integration & Indexing



- Query: Canon A70 digital camera with price less than \$300
- Automatic slot filling
 - Site 1: <http://www.abtelectronics.com/search.php3?Brand=canon&Type=A70&MaxPrinice=300>
 - ...

Structure of the Web – the Big Picture



Cluster Search - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://msra-idss-04:8080/clusteriii/ClusterMain2.aspx?query=jaguar&Source=Google>

MSRA WSM | jaguar | Search | Google | Clustering | Options | Highlight | jaguar

Clustered Results

- [-] jaguar(200)
 - [+] jaguar cars(27)
 - [+] panthera onca(16)
 - mac os(19)
 - jaguar club(11)
 - [+] jaguar xjs(6)
 - [+] atari jaguar(9)
 - [+] classic jaguar(5)
 - jaguar models(5)
 - jungle, master(5)
 - [+] jaguar xk(4)
 - jaguar xj6(4)
 - [+] jaguar xtype(4)
 - maya, civilization(4)
 - largest cat(3)
 - [+] game(7)
 - jaguar etype(3)
 - others(86)
 - More



Jaguar

Panthera onca



Key Facts

Body Length(mm) - 1200-1800

Weight (kg) - 70-120

Litter Size - 1-4 average

Life Span - 12-16 years

Status - Near Threatened

In appearance the Jaguar is often confused with the Leopard - both cats, depending to a degree on sub-species have a similar brownish/yellow base fur colour which is distinctively marked with dark rosette markings. However, the jaguar can be distinguished by the presence of small dots or irregular shapes within the larger rosette markings, a more stocky and muscular body and a shorter tail. Melanistic or black jaguars (see below) are common in certain parts of its range and are often confusingly labelled 'Black Panthers', a name which is also applied to black Leopards. In this melanistic form the cats are more difficult to separate, however the jaguars large head and stocky forelimbs are often a good way to differentiate between the two cats.

Done | Local intranet

Cluster Search - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://msra-idss-04:8080/clusteriii/ClusterMain2.aspx?query=Kai%2DFu%20Lee%20Ya%20Qin%20Zhang&Source=Google> Go

MSRA WSM Kai-Fu Lee Ya-Qin Zhang Search Google Clustering Options Highlight Kai-Fu Lee Ya-Qin Zhang

Clustered Results Cluster > kai-fu lee ya-qin zhang(40)

- [-] kai-fu lee ya-qin zhang(40)
 - [+] managing director(16)
 - [+] microsoft research(8)
 - [+] vice president(8)
 - [+] dream team(4)
 - [+] others(8)
 - More

[1.Labs: Asia Fast Facts](#)
 ... In August 2000, **Ya-Qin Zhang**, Microsoft Research Asia's former assistant director and ... Research Asia's original managing director, Dr. **Kai-Fu Lee**, was promoted ...
http://research.microsoft.com/aboutmsr/labs/asia/beijing_facts.aspx
 Source ID [1]

[2.Jianfeng Gao: Resume](#)
 ... 1. Dr. **Kai-Fu Lee** : Vice President of Microsoft Corporation, former Managing Director of Microsoft Research China. Email: kfl@microsoft.com. 2. Dr. **Ya-Qin Zhang**: ...
<http://research.microsoft.com/~jfgao/resume.htm>
 Source ID [2]

[3.msmobiles.com - Chinese research head named VP of MS Mobile and ...](#)
 ... **Kai-Fu Lee, Zhang**'s predecessor as the former managingi director of MSR Asia, is ... **Ya-Qin Zhang**'s assignment is supposed to be officially announced by Redmond ...
<http://www.msmobiles.com/news.php/1936.html>
 Source ID [3]

[4.Technology Review: The World's Hottest Computer Lab](#)
 ... **Kai-Fu LEE**: To be successful in China, you have to be sincere in doing what' s good ... The lab' s two previous directors, **Lee** and **Ya-Qin Zhang**, have been ...
<http://www.google.comhttps://www.techreview.com/articles/huang0604.asp?p=2>
 Source ID [4]

[5.theSpoke](#)
 ... It's about How to be a success Microsoft guy,which tells us a lot of the storys about the people in Microsoft,including Mr. **Kai-fu Lee**,Mr.**Ya-Qin Zhang**,etc. ...
http://www.thespoke.net/MyBlog/zengyiCSTC/MyBlog_Comments.aspx?ID=8284
 Source ID [5]

[6 Microsoft builds R&D Dream Team in Beiiing](#)

Local intranet

Application: Research Community Search

- People, papers, conferences, & interest groups
 - Each treated as a web object
 - Extract & integrate information from multiple sources
- Advanced search functions
 - Object-level link analysis
 - Importance ranking
 - Relationship mining
 - Trend analysis
 - Structured query

Summary

- Mining web structure is the key to develop the next-generation search engine
 - Layout structure
 - Deep structure
 - Category structure
 - Community structure
 - ...

Towards Next Generation Web Search

- From Pages to Blocks
 - Analyze the Web at finer granularity
- From Surface Web to Deep Web
 - Unleash the huge asset of high-value information
- From Unstructure to Structure
 - Provide well organized search results
- From Relevance to Intelligence
 - Combine knowledge discovery with search
- From Desktop Search to Mobile Search
 - Bridge physical world search to digital world search

References

SIGIR 2004

- Dou Shen, Zheng Chen, Hua-Jun Zeng, Benyu Zhang, Qiang Yang, Wei-Ying Ma, Yuchang Lu, Web-page Classification through Summarization, SIGIR'2004, July 2004.
- Hua-Jun Zeng, Qi-Cai He, Zheng Chen, Wei-Ying Ma. Learning To Cluster Search Results, SIGIR'2004, July 2004.
- Ji-Rong Wen, Ni Lao and Wei-Ying Ma, Probabilistic Model for Contextual Retrieval, SIGIR 2004, July 2004 .
- Deng Cai, Shipeng Yu, Ji-Rong Wen and Wei-Ying Ma, Block-based Web Search, SIGIR 2004, July 2004 .
- Deng Cai, Xiaofei He, Ji-Rong Wen and Wei-Ying Ma, Block-Level Link Analysis, SIGIR 2004, July 2004 .
- Xiaofei He, Deng Cai, Haifeng Liu and Wei-Ying Ma. Locality Preserving Indexing for Document Representation, SIGIR'2004, July 2004.

WWW 2004

- Ruihua Song, Haifeng Liu, Ji-Rong Wen and Wei-Ying Ma, Learning Block Importance Models for Web Pages, World Wide Web conference (WWW 2004), 203-211, May, 2004.
- Wensi Xi, Benyu Zhang, Yizhou Lu, Zheng Chen, Shuicheng Yan, Huajun Zeng, Wei-Ying Ma, and Edward A. Fox. Link Fusion: A Unified Link Analysis Framework for Multi-Type Interrelated Data Objects , World Wide Web conference (WWW 2004), 203-211, May, 2004.

VLDB 2004

- Jiying Wang, Ji-Rong Wen, Fred Lochovsky and Wei-Ying Ma, Instance-based Schema Matching for Web Databases by Domain-specific Query Probing, The 30th International Conference on Very Large Data Bases (VLDB 2004), Toronto, Ontario, Canada, August 2004.

ACM Multimedia 2004

- Deng Cai, Xiaofei He, Zhiwei Li, Wei-Ying Ma and Ji-Rong Wen, Hierarchical Clustering of WWW Image Search Results Using Visual, Textual and Link Analysis ,12th ACM International Conference on Multimedia, New York City, USA, Oct. 2004 .
- Xin-Jing Wang, Wei-Ying Ma, Gui-Rong Xue, and Xing Li, Multi-Model Similarity Propagation and its Application for Web Image Retrieval,12th ACM International Conference on Multimedia, New York City, USA, Oct. 2004.
- Xiaofei He, Wei-Ying Ma, Hong-Jiang Zhang, Learning an Image Manifold for Retrieval,12th ACM International Conference on Multimedia, New York City, USA, Oct. 2004.
- Xin Zheng, Deng Cai, Xiaofei He, Wei-Ying Ma and Xueyin Lin, Locality Preserving Clustering for Image Database ,12th ACM International Conference on Multimedia, New York City, USA, Oct. 2004.

Thank You!