Integrating and Ranking Aggregated Content on the Web

Jaime Arguello[1] Fernando Diaz[2] Milad Shokouhi[3]

[1]UNC Chapel Hill [2]Yahoo! Labs [3]Microsoft Research

May 3, 2012

Outline

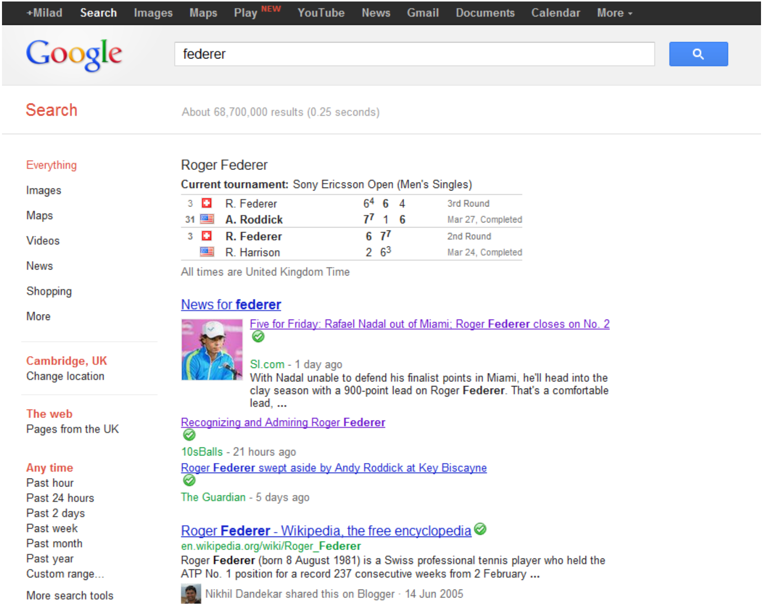
Introduction

Problem Definition

"Aggregating *content* from different sources."

In *aggregated search*, content is retrieved from *verticals* in response to queries.

Examples: Aggregated Search



Examples: Personalized Search

Web Images Groups News Froogle **Desktop** more »

new york | Search Desktop

Desktop Preferences
Remove items

Desktop: All - 39 emails - 0 files - 1 chat - 130 web history

1-10 of about 170 (0.01s)

Sort by relevance **Sorted by date**

Architecture of **New York** City - Great Buildings Online
Architecture of **New York** City -Great Buildings Online Architecture of **New York** City Visit the Home Design Store for great values! DesignWorkshop Classic -HomePAK
greatbuildings.com/places/**new_york**_city - 1 cached - 1:15pm

eBay - **New York** Yankees, Fan Apparel Souvenirs, Cards, a..
New York Yankees, Fan Apparel Souvenirs, Cards, and Autographs-Original items at low prices home |pay |register |register |sign in |sign in/out |services |site map
buy.ebay.com/**new-york**-yankees - 1 cached - 1:14pm

Guggenheim Museum - **New York**
Guggenheim Museum -**New York**
www.guggenheim.org/**new_york**_index.shtml - 1 cached - 1:14pm

BASEBALL-LINKS.COM - John Skilton's Baseball Links
10-13 BASEBALL ROUNDUP Larkin's Career With Reds Is Over After 19 .**New York** Times 2004-10-12 BASEBALL PLAYOFFS San Francisco Chronicle 2004-10-12
www.baseball-links.com/ - 1 cached - 1:13pm

The **New York** Times > Breaking News, World News & Multi..
The **New York** Times >Breaking News, World News Multimedia UPDATED THURSDAY, OCTOBER 14, 2004 4:12 PM ET |Personalize
www.nytimes.com/ - 1 cached - 1:13pm

Examples: Metasearch

dögpile

Web | Images | Video | News | Local | White Pages

Integrating and Ranking Aggregated Content **Go Fetch!**

Google YAHOO! bing Advanced Search Preferences

Web Search Results for "Integrating and Ranking Aggregated ..." (About Results)

Robust rank aggregation for gene list integration **and** meta-analysis
bioinformatics.oxfordjournals.org/...nt/28/4/573.full • Found exclusively on: Google
Jan 12, 2012 ... Thus, the rank aggregation methods can become a useful **and** general solution for the integration task. Results: Standard rank aggregation ...

Integrating Content, Pedagogy, **and** Reflective Practice: Innovative ...
www.westga.edu/...nce/ojdla/fall113/hovermill113.html • Found exclusively on: Yahoo! Search
Integrating Content, Pedagogy, **and** Reflective Practice ... Respondents' comments were **aggregated** by theme ... Course evaluation ratings have also shown a high ...

How to Integrate Social Media **and** Branded **Content** | Digital Tonto
www.digitaltonto.com/...nded-contentand-social-media/ • Found on: Yahoo! Search, Bing
Ranking user **content** as part of a branded **content** ... more powerful when the site is **aggregated** with other **content** ... going to think more seriously about **integrating** ...

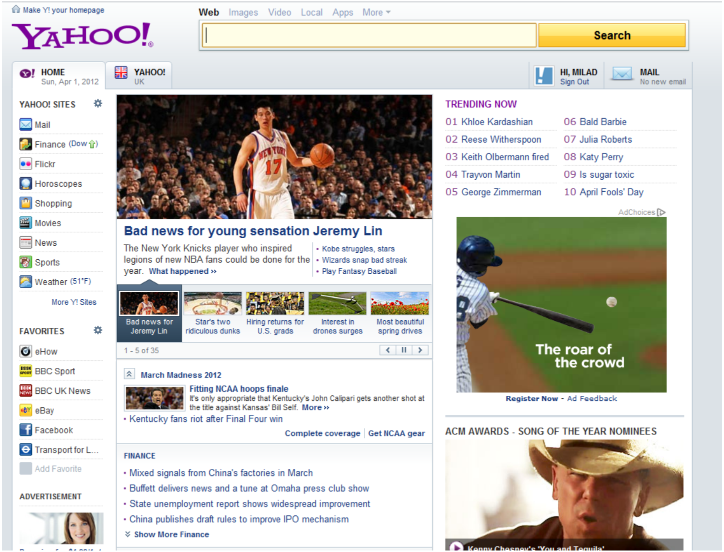
Robust Rank Aggregation for gene list integration **and** meta-analysis
bioinformatics.oxfordjournals.org/...r709.short?rss=1 • Found exclusively on: Google
Jan 12, 2012 ... Abstract. Motivation: The continued progress in developing technological platforms, availability of many published experimental data sets, ...

Tutorials | www2012
www2012.wwwconference.org/program/tutorials • Found exclusively on: Bing
Integrating and Ranking Aggregated Content on the Web (Fernando Diaz, Jaime Arguello **and** Milad Shokouhi) Tuesday April 17th – afternoon. The Web of Things (Carolina ...

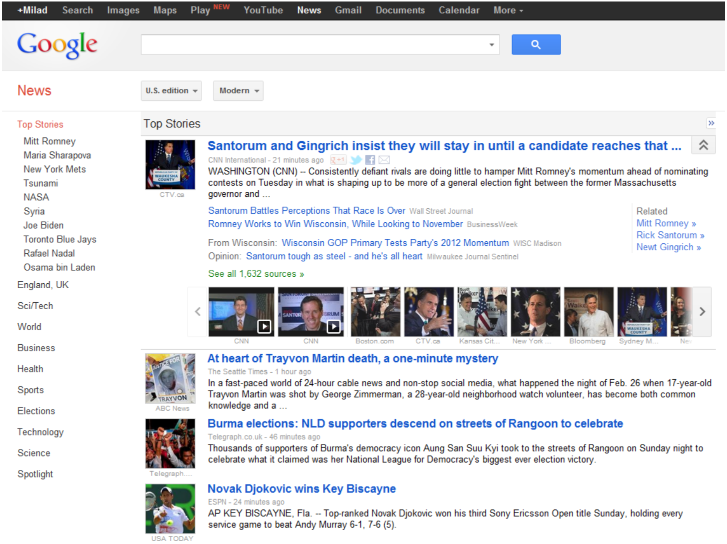
Tutorial abstracts | www2012
www2012.wwwconference.org/...ials/tutorial-abstracts/ • Found exclusively on: Google
Integrating and Ranking Aggregated Content on the Web. In this tutorial, we will present the core problems associated with **content** aggregation, which include: ...

Examples: Content Aggregation

Examples: Content Aggregation



Examples: Content Aggregation



History of Content Aggregation

Pre-Computer Content Aggregation

DAILY EXPRESS, Monday, September 23, 1940.

More Of A Pal Than Ever

SHERLEY'S TONIC AND CONDITION POWDERS

BLACK-OUT ZERO HOUR TO-NIGHT UNTIL 6.20 A.M.

MOON RISES MOON SETS

Daily Express

No. 12,585 Monday, September 23, 1940 One Penny

THOUGHT FOR FOOD H·P SAUCE

Sent to escape the bombers, 89 English children are murdered by a U-boat

CHILDREN'S LINER SUNK WITHOUT WARNING IN GALE

Lord Beaverbrook calls to aircraft workers 'WORK AFTER SIREN HAS SOUNDED'

Lord Beaverbrook, Minister of Aircraft Production, last night issued this message:—

I HAVE seen the statements in the Press about some workers in several aircraft factories taking shelter throughout the period of air-raid warnings.

Outrage in Atlantic

BOATS SWAMPED BY TERRIFIC SEAS 600 MILES FROM LAND

WITHOUT WARNING A U-BOAT FIRED A TORPEDO AT A LINER STRUGGLING THROUGH A STORM IN THE ATLANTIC LAST TUESDAY NIGHT—AND KILLED EIGHTY-NINE ENGLISH CHILDREN.

A number of the children were killed by the explosion when the torpedo hit the ship. Many others were drowned. The terrific seas swamped rafts and overturned lifeboats.

Seven out of the nine adult escorts with the children lost their lives in heroic attempts to save them. All who survived tell of the amazing courage of the children. Some of them were only five years old, yet they stood quietly without whimpering until they were

GALE SEASON

TODAY is Autumn Equinox, when day exactly equals darkness. Strong equinoctial gales are due to start whipping up the Channel, making Hitler's chances of bringing his invasion barges across more and more remote.

Japanese attack Indo-China

JAPANESE troops crossed the border from China into French Indo-China last night. They attacked a French block-

FLATS, CINEMA, CHURCH BOMBED

Daily Express Raid Reporters

EARLY today, during London's sixteenth Blitznight, it was reported that a block of flats had been hit by a bomb.

High explosive and incendiary bombs dropped on the eastern outskirts badly damaged an old parish church, a cinema, a dance-hall, and shops.

After a Sunday confined to sneak-bombing by single planes, London's third alert of the day was followed by the most intense night attack the German raiders have made since last Wednesday.

For more than an hour the guns kept the raiders out. Then one, fly-

STOP PRESS

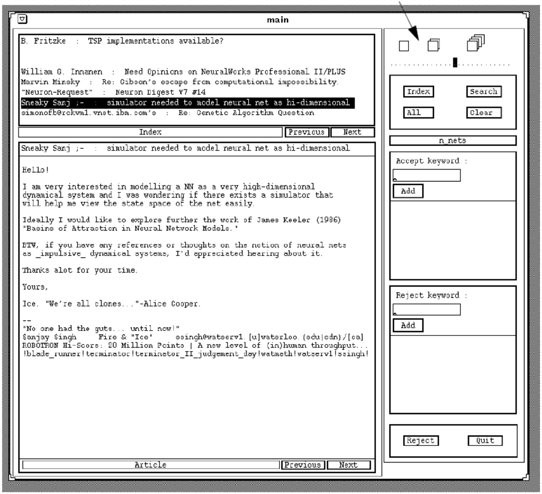
HOUR'S BREAK IN LONDON RAID

There was a break in the London area raid early this morning, but the alert again sounded after about an hour.

BERLIN RAIDED AGAIN

Were newspapers the first content aggregation media?

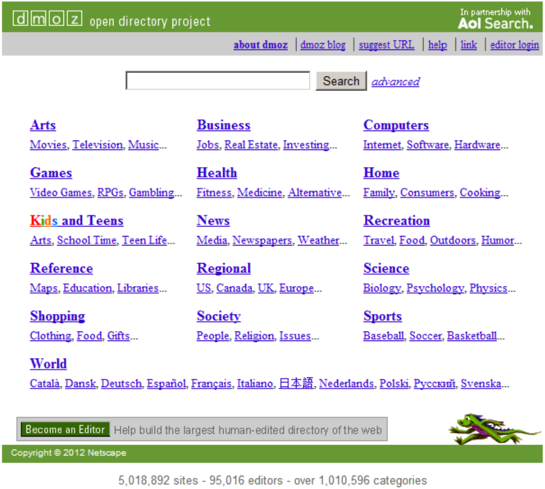
Pre-Web Content Aggregation



Pre-Web systems were mostly used for content *filtering* rather than *aggregation*.

Related reading: A. Jennings and H. Higuchi [15]

Web 1.0 Content Aggregation



Manual content aggregation since early days of the world-wide-web.

Web 1.0 Content Aggregation

The begining of automatic content aggregation and news recommendation on the web.

Related reading: Kamba et. al [16]

Content Aggregation Today (Exploit/Explore)

- Explore/Exploit. *"What to exploit is easy, but what to explore is where the secret sauce comes in."* says Ramakrishnan.
- Clickthrough rates on Yahoo! articles improved by 160% after personalized aggregation.
- Many content aggregators such as *The New York Times* use a hybrid approach.

Related reading: Krakovsky [17]

Content Aggregation Today (Real-time and Geospatial)

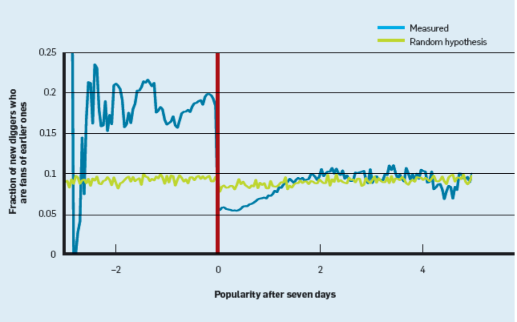
(a) user and sensory context (b) rank of entity types (c) rank of restaurant entities

- real-time updates
- geo-spatial signals

Related reading: Zhuang et. al [34]

Content Aggregation Today (Temporal and Social)

Figure 7. Probability that a digger of a story is a fan of a digger who dugg the same story (blue line) as a function of the time of the digg. Time is relative to the promotion time of the story, with the average calculated over all diggs on all stories. The vertical red line marks time 0 (promotion time), and negative times refer to the "upcoming" phase. The green line is the same measurement but with diggs randomly shuffled.



- "Semantic analysis of content is more useful when no early click-through information is available."
- Social signals are more influential for trending topics.

Related reading: G. Szabo and B. Huberman [33]

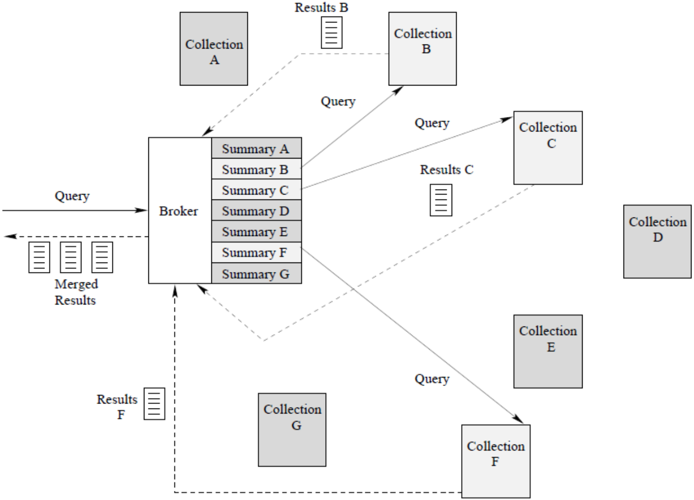
History of Aggregated Search

Stone Age: Federated Search

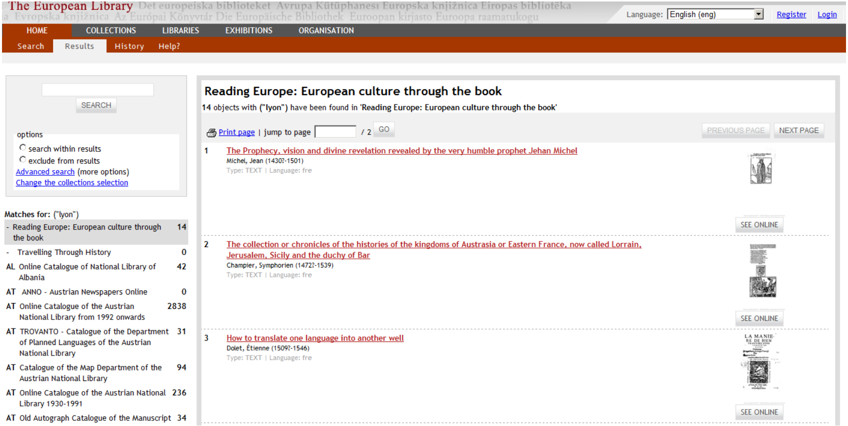
Also known as distributed information retrieval (DIR), allows users to search multiple sources (collections, databases) at once and receive merged results. Typically, in federated search:

- Collections contain text documents only.
- There is no overlap among collections.
- Still common in enterprise and digital libraries.

Federated Search Architecture



Example: The European Library



Federated Search environments

- **Cooperative**
  - The broker has comprehensive information about the contents of each collection.
  - Collection sizes and lexicon statistics are usually known to the broker.
- **Uncooperative**
  - Collections are not willing to publish their information.
  - The broker uses query-based sampling [7] to approximate the lexicon statistics for each collection.

Federated Search Challenges

- Query translation
- Source representation
- Source selection
- Result merging

Related reading: Shokouhi and Si [26]

Query Translation

Example: STARTS Protocol Query

```
@SQuery{
Version{10}: STARTS 1.0
FilterExpression{50}: ((author "Garcia Molina") and (title
"databases"))
RankingExpression{61}: list((body-of-text "distributed") (body-of-text
"databases"))
DropStopWords{1}: T
DefaultAttributeSet{7}: basic-1
DefaultLanguage{5}: en-US
AnswerFields{12}: title author
MinDocumentScore{3}: 0.5
MaxNumberDocuments{2}: 10
}
```

Related reading: Gravano et. al [14]

Source Representation

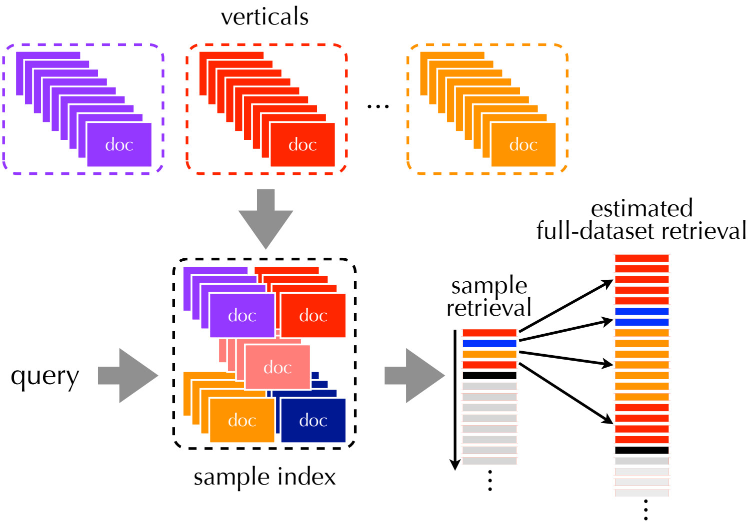
Document Sampling

- Query-based Document Sampling
- Content-driven Sampling
  - Issue random term to the vertical
  - Sample top results
  - Update vertical-specific vocabulary representation
  - Sample new term from emerging representation
  - Repeat
- Demand-driven Sampling
  - Sample query from vertical-specific query-log
  - Sample top results
  - Repeat

Related Reading: Callan and Connell [7] and Shokouhi *et al.* [27]

Source Selection

Example: Relevant Document Distribution Estimation



Source Selection

Example: Relevant Document Distribution Estimation

- **Assumption:** each (predicted) relevant sample represents $\frac{|v|}{|S_v|}$ relevant documents in the original vertical collection

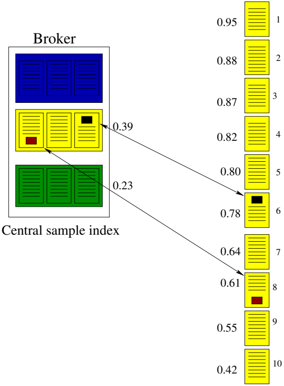
$$\text{ReDDE}(v,q) = \frac{1}{\mathcal{Z}} \sum_{d \in \mathcal{R}_N} \frac{|v|}{|S_v|} \times \mathcal{I}(d \in v)$$

- $\mathcal{Z}$ normalizes across verticals
- $\mathcal{Z} = \sum_{v \in \mathcal{V}} \text{ReDDE}(v,q)$

Related Reading: Si and Callan [28], Fuhr [13]

Result Merging

Example: Semi-supervised Learning for Merging



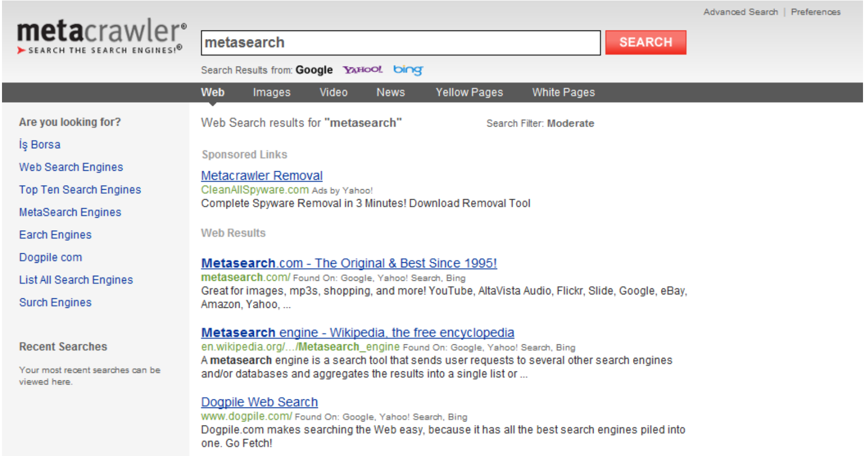
- SSL [29] uses sampled data and regression for merging.
- SSL relies on overlap between the returned results and sampled documents.
- A regression function is trained based on the overlap

Bronze Age: Metasearch engines

Metasearch engines submit the user query to multiple search engines and merge the results. They date back to MetaCrawler (1994), that used to merge the results from WebCrawler, Lycos and InfoSeek.

- In metasearch the query is often sent to all sources.
- Result merging is usually based on position and overlap.

Example: MetaCrawler



Iron Age: Aggregated Search

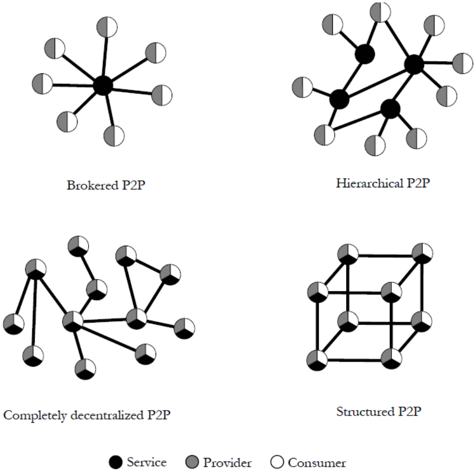
In 2000, the Korean search engine (Naver) introduced *comprehensive search* and started blendeding multimedia answers in their default search results. Google introduced *universal search* in 2007.

Motivation:

- Web data is highly heterogeneous.
- Information needs and search tasks are similarly diverse.
- Keeping a fresh index of real-time data is difficult.

Related Problems

Peer-to-Peer Search



Related reading: Lu [22]

Related Problems

Data fusion

- Several rankers on the same data collection
- Each ranker is considered as a *voter*

#	51 voters	5 voters	23 voters	21 voters
1st	Andrew	Catherine	Brian	David
2nd	Catherine	Brian	Catherine	Catherine
3rd	Brian	David	David	Brian
4th	David	Andrew	Andrew	Andrew

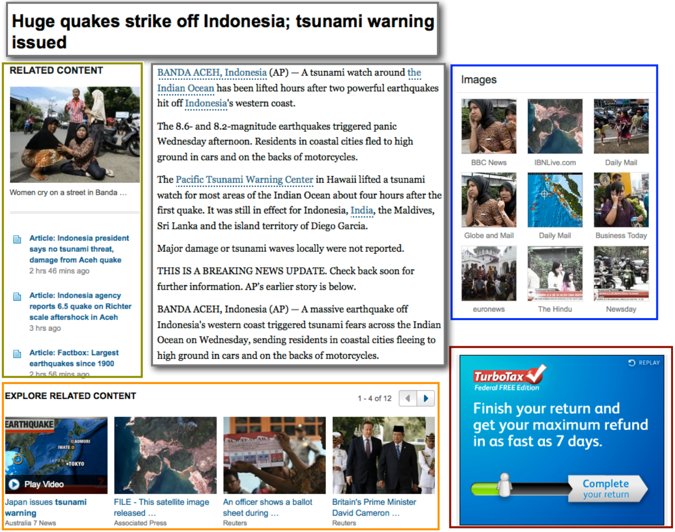
Borda scores: Andrew 153, Catherine 205, Brian 151, David 91

Related reading: `http://bit.ly/HPBFoG`

Problem Definitions

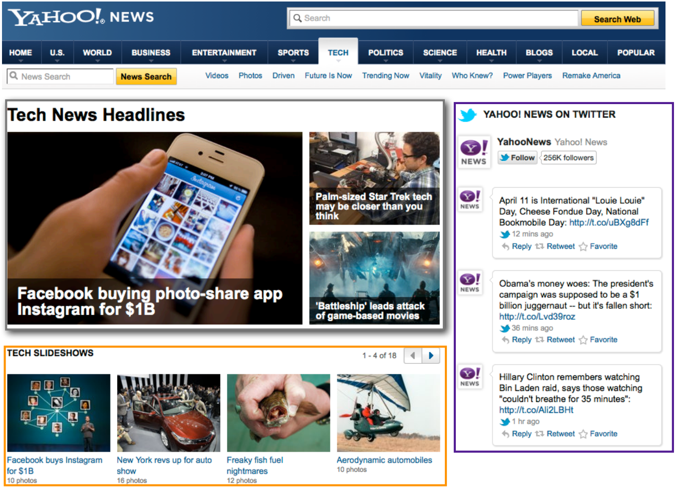
Content Aggregation

Examples



Content Aggregation

Examples

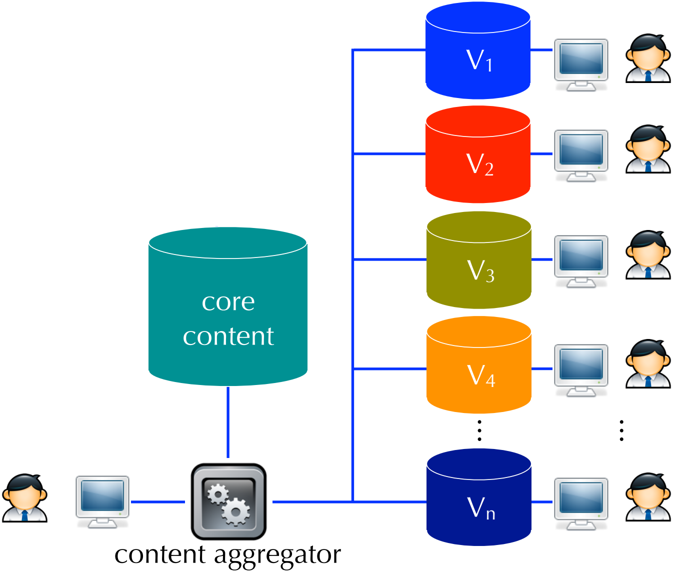


Content Aggregation

Definitions

- **Core content:** the content that is always presented and is the main focus on the page
- **Vertical content:** related content that is optional and supplements the core content

Content Aggregation



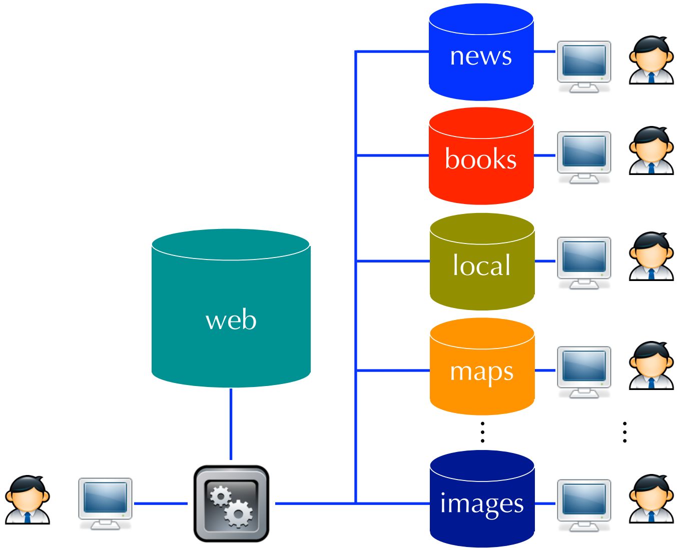
Problem Definition

- Given a particular **context**, predict *which* verticals to present and *where* to present them
- A **context** is defined by the core content, the information request (i.e., the query), and/or the user profile

Content Aggregation in Web Search

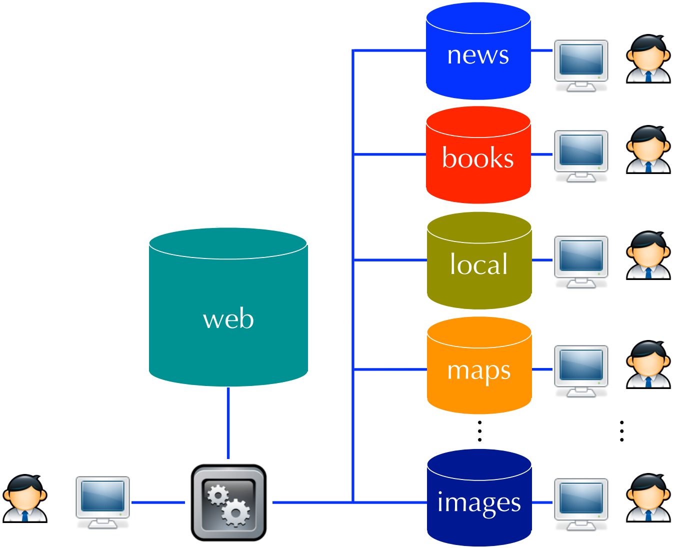
Aggregated Web Search

- Integrating vertical results into the core Web results



What is a Vertical?

- A specialized search service that focuses on a particular type of information need



Example Verticals

lyon news | **Search**

Lyon - News Results

Brandon **Lyon** nearing end of long road back Houston Chronicle - 11 hours ago

Lyon County & NNDA seeks input on regional plan at workshop in Silver Springs Fernley Leader - Mar 28 01:53am

Deputies in Douglas, **Lyon** counties increasing traffic patrols

Lake Tahoe News - Mar 31 09:58am

lyon restaurants | **Search**

Restaurants in Lyon, France

travel.yahoo.com

1 Adrets (Les) ★★★★★ (4 Reviews)
+33 4 7838 2430 - 30, rue du Boeuf, Lyon
Cuisine: French

2 Pierre Orsi ★★★★★ (1 Review)
+33 4 7889 5768 - 3, place Kléber, Lyon
Cuisine: Contemporary

3 Grand Cafe des Negoc... ★★★★★ (2 Reviews)
+33 4 78 42 5005 - 2, place Francisque Régaud, Lyon
Cuisine: Bistros & Brasseries

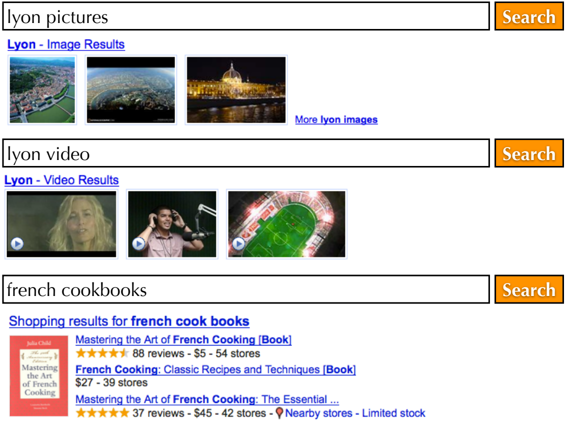
More Restaurants in Lyon »

lyon map | **Search**

Lyon
France
maps.google.com

Hotels - Restaurants - Fourvière - Cathedrale St Jean - Place Bellecour - Lugdunum - Parc De La Tete D Or - Cour des Loges

Example Verticals



Example Verticals

lyon weather | Search

Weather for **Lyon, France**

61°F | °C
Clear
Wind: N at 12 mph
Humidity: 39%

Sun	Mon	Tue	Wed
64° 39°	72° 50°	68° 50°	61° 41°

Detailed forecast: The Weather Channel - Weather Underground - AccuWeather

lyon flight | Search

Flights to **Lyon, France** (LYS)

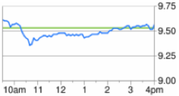
No non-stop flights from Raleigh, NC

From	Duration	Airlines
Paris, France	1h 11m	Air France
Brussels, Belgium	1h 26m	Brussels, Air France, easyJet
Bordeaux, France	1h 12m	Air France, easyJet

+ Show all non-stop routes to Lyon

bank of america | Search

BAC - Bank of America Corp (NYSE)



9.57 +0.04 (0.42%)
Mar 30 4:01pm ET - Disclaimer

Open:	9.61	**Volume:**	250,234,655
High:	9.64	**Avg Vol:**	293,659,000
Low:	9.35	**Mkt Cap:**	103.08B

bonjour in english | Search

Translate **"bonjour"** from French
translate.google.com
bonjour - hello

What is a Vertical?

- A specialized search service
- Different verticals retrieve different types of media (e.g., images, video, news)
- Different verticals satisfy different types of information needs (e.g., purchasing a product a product, finding a local business, finding driving directions)

Aggregated Web Search

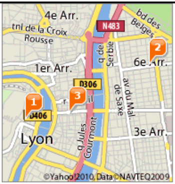
lyon | **Search**

Restaurants in Lyon, France

travel.yahoo.com

1. Adrets (Les) ★★★★★ (4 Reviews)
   +33 4 7838 2430 - 30, rue du Boeuf, Lyon
   Cuisine: French
2. Pierre Orsi ★★★★★ (1 Review)
   +33 4 7889 5768 - 3, place Kléber, Lyon
   Cuisine: Contemporary
3. Grand Cafe des Negoc... ★★★★★ (2 Reviews)
   +33 4 78 42 5005 - 2, place Francisque Régaud, Lyon
   Cuisine: Bistros & Brasseries

More Restaurants in Lyon »



Industrial Workspace Products - **Lyon** Workspace Products ...

www.**lyon**workspace.com/

Lyon Workspace Products of Montgomery, IL offers workspace products such as steel lockers, heavy duty steel shelving, steel storage racks, and industrial ...

Lyon - Image Results

More **lyon** images

Lyon travel guide - Wikitravel

wikitravel.org/en/**Lyon**

Open source travel guide to **Lyon**, featuring up-to-date information on attractions, hotels, restaurants, nightlife, travel tips and more. Free and reliable advice ...

Lyon - Video Results

Aggregated Web Search

- **Task:** combining results from multiple specialized search services into a single presentation
- **Goals**
  - To provide access to various systems from a single search interface
  - To satisfy the user with the aggregated results
  - To convey how the user's goal might be better satisfied by searching a particular vertical directly (if possible)

Aggregated Web Search

Motivations

- Users may not know that a particular vertical is relevant
- Users may want results from multiple verticals at once
- Users may prefer a single-point of access

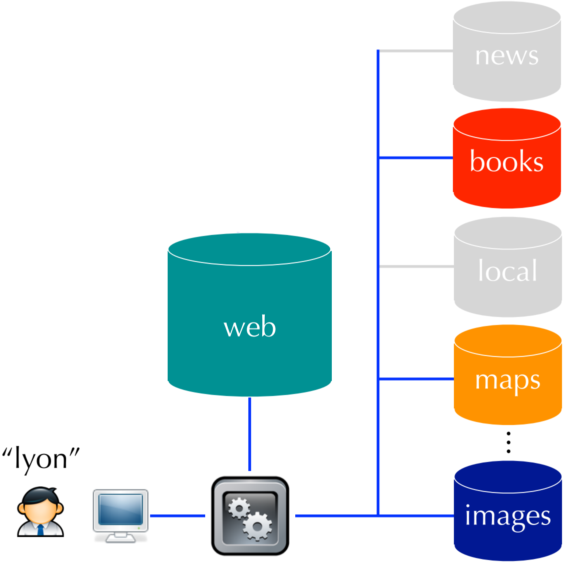
Aggregated Web Search

Task Decomposition

- **Vertical Selection**
- **Vertical Results Presentation**

Vertical Selection

- Predicting *which* verticals to present (if any)



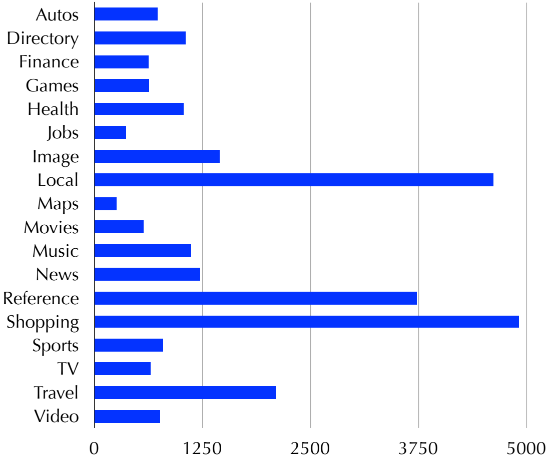
Vertical Selection

- Given a query and a set of verticals, predict which verticals are relevant
- In some situations, this decision must be made without issuing the query to the vertical
- Later on we'll discuss sources of *pre*-retrieval evidence

Vertical Distribution

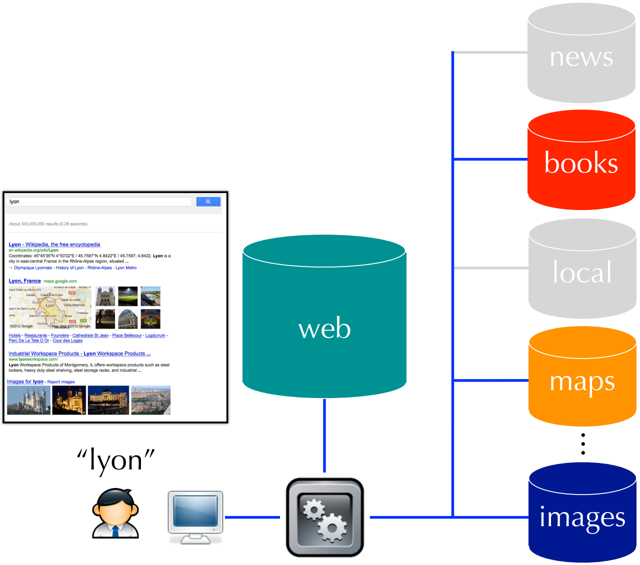
25K queries sampled randomly from Web traffic

- Number of queries for which the vertical was considered relevant



Vertical Results Presentation

- Predicting *where* to present them (if at all)



Vertical Results Presentation

- Given a query, a set of (selected) verticals, and a set of layout constraints, predict where to present the vertical results
- In some situations, it may be possible to suppress a predicted relevant vertical based on its results
- Later on we'll discuss sources of *post*-retrieval evidence

Sources of Evidence

Content Aggregation

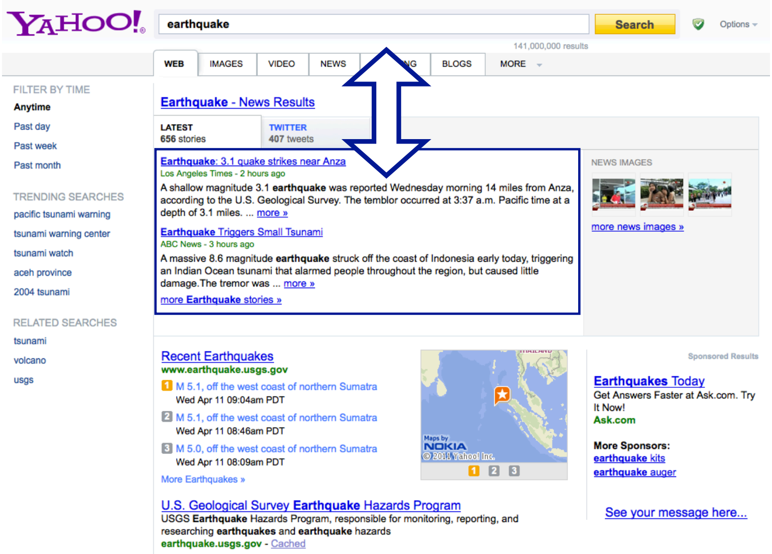
- Given a particular **context**, predict *which* verticals to present and *where* to present them
- A **context** is defined by the core content, the information request (i.e., the query), and/or the user profile
- **Vertical selection:** predicting which verticals to present (if any)
- **Vertical results presentation:** predicting where to present each vertical selected (if at all)
- Different content aggregation environments may be associated with different sources of evidence

Sources of Evidence

- Relationship between core content and vertical content

Sources of Evidence

- Relationship between explicit request and vertical content

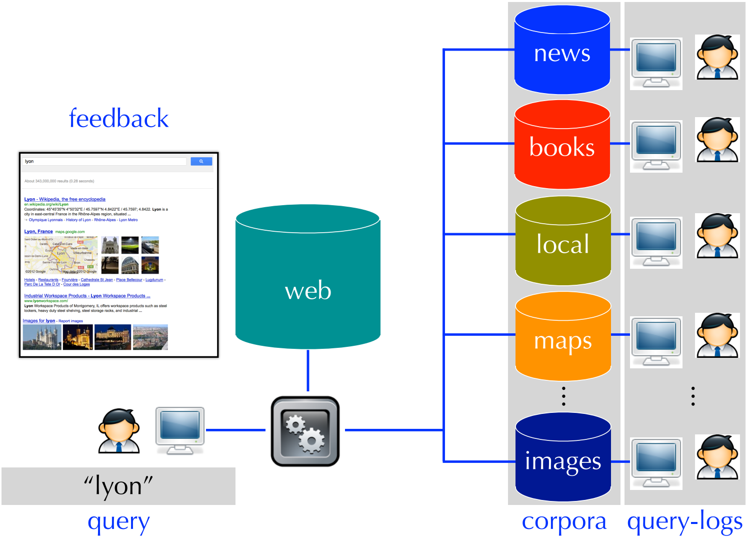


Sources of Evidence

- Relationship between user profile and vertical content

Sources of Evidence for Aggregated Web Search

Sources of Evidence



Types of Features

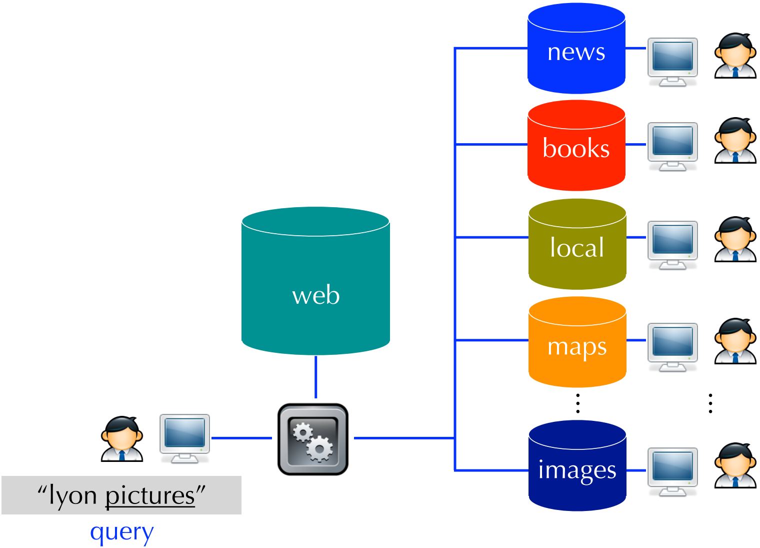
- **Pre-retrieval features:** generated before the query is issued to the vertical
- **Post-retrieval features:** generated after the query is issued to the vertical and before it is presented
- **Post-presentation features:** generated after the vertical is presented
    - possibly available from previous impressions of the query

Pre-retrieval Features

- Query features
  - the query's topical category (e.g., travel)
  - key-words and regular expressions (e.g., lyon pictures)
  - named entity types (e.g., city name)
- Vertical corpus features
  - similarity between the query and sampled vertical results
- Vertical query-log features
  - similarity between the query and vertical query-traffic

Query Features

- Derived from the query, independent of the vertical



Query Features

- Terms appearing in the query
- Dictionary look-ups: "yhoo" $\rightarrow$ finance vertical
- Regular expressions:
  - "obama news" $\rightarrow$ news vertical
  - "ebay.com" $\rightarrow$ no vertical
- Named-entity types: "main st., pittsburgh" $\rightarrow$ maps vertical
- Query category: "lyon attractions" $\rightarrow$ travel vertical

Related Reading: Arguello *et al.* [4], Ponnuswami *et al.* [23], and Li *et al.* [20]

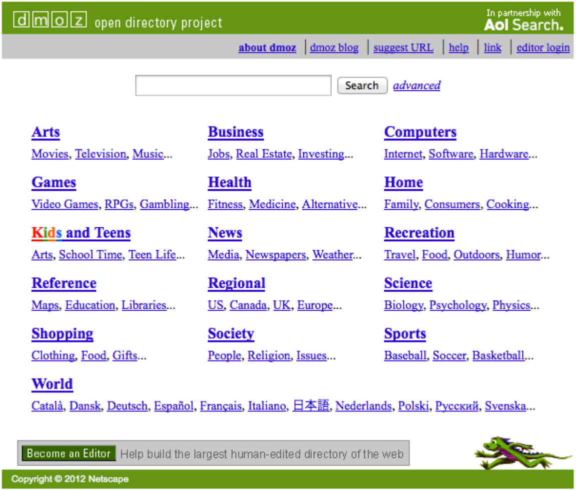
Query Features

- Query-log-based co-occurrence between the query and vertical specific key-words
  - Co-occurrence can be measured using the χ^2 static
  - Example image vertical keywords: photo(s), pic(s), picture(s), image(s)

Related Reading: Arguello *et al.* [2]

Query Category Features

- Use a corpus with known document-to-category assignments (binary or soft)



Query Category Features

- Query-category assignment based on document-category assignments of top-N results

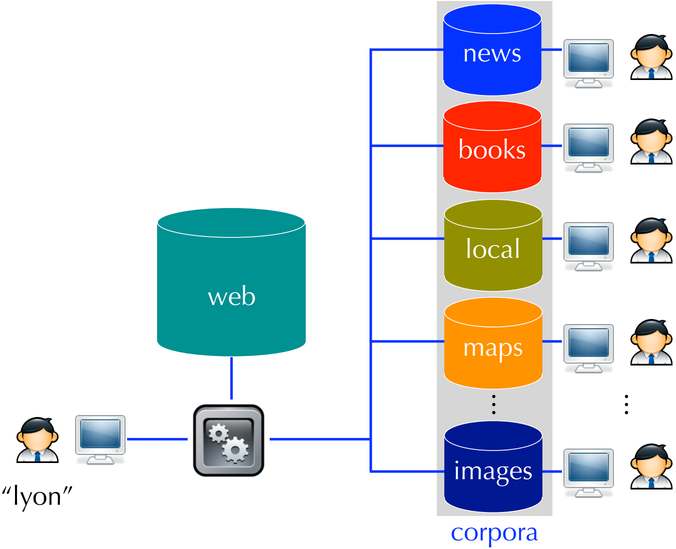
$$P(c|q) = \frac{1}{\mathcal{Z}} \sum_{d \in \mathcal{R}_N} P(c|d) \times \mathsf{score}(d, q)$$

- Z normalizes across categories
- $Z = \sum_{c \in \mathcal{C}} P(c|q)$

Related Reading: Shen *et al.* [25]

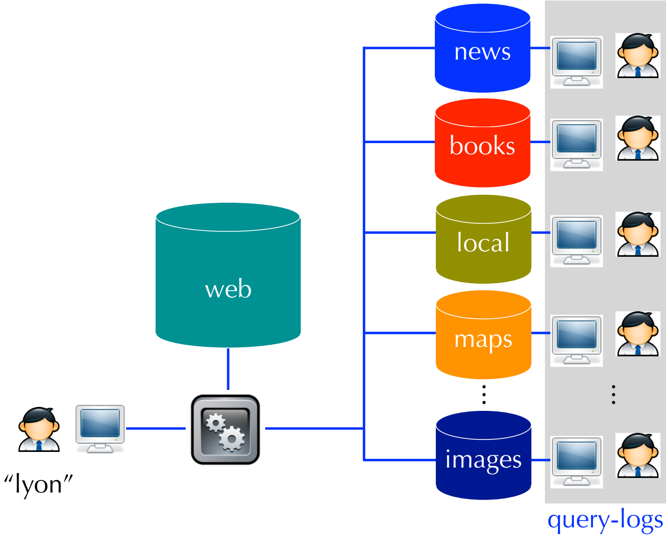
Vertical Corpus Features

- Derived from (sampled) vertical documents (e.g. ReDDE)



Vertical Query-log Features

- Derived from queries that were issued directly to the vertical by users



Similarity to Vertical Query-Traffic

- Similarity between the query and queries issued directly to the vertical by users

$$\mathsf{QLOG}(v,q) = \frac{1}{\mathcal{Z}} \prod_{w \in q} P(w|\theta_v^{\mathsf{qlog}})$$

- $\mathcal{Z}$ normalizes across verticals
- $\mathcal{Z} = \sum_{v \in \mathcal{V}} \mathsf{QLOG}(v,q)$

Related Reading: Arguello *et al.* [4]

Types of Features

- **Pre-retrieval features:** generated before the query is issued to the vertical
- **Post-retrieval features:** generated after the query is issued to the vertical and before it is presented
- **Post-presentation features:** generated after the vertical is presented
  - possibly available from previous impressions of the query

Post-Retrieval Features

- Derived from the vertical's response to the query
- Derived from the vertical's *full* retrieval, or only the few results that will potentially be presented
- Results from different verticals are associated with different meta-data
    - publication date: news, blog, micro-blog, books
    - geographical proximity: news, micro-blog, local, maps
    - reviews: community Q&A, local, shopping
    - price: shopping, books, flights
- **Challenge:** Post-retrieval features tend to be different for different verticals

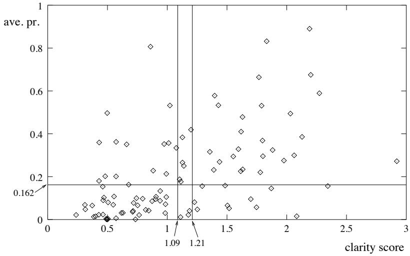
Post-Retrieval Features

- Number of results
- Retrieval score distribution
- Text similarity features: cross-product of
  - **Extent:** title, url, summary
  - **Similarity Measure:** clarity, cosine, jaccard, query-likelihood
  - **Aggregator:** min, max, mean, std. dev.
- Recency features: cross-product of
  - **Extent:** creation date, last modified date
  - **Recency Measure:** time difference, exp. decay
  - **Aggregator:** min, max, mean, std. dev.

Related Reading: Arguello *et al.* [2], Diaz [10]

Post-Retrieval Features

Example: Clarity Score



- Measures query ambiguity with respect to collection.
- If the returned results are not topically similar, the peformance might be poor.

$$\textit{clarity score} = \sum_{w \in V} P(w|Q) \log_2 \frac{P(w|Q)}{P_{\textit{coll}}(w)} \tag{1}$$

Related reading: Cronen-Townsend et al. [9]

Types of Features

- **Pre-retrieval features:** generated before the query is issued to the vertical
- **Post-retrieval features:** generated after the query is issued to the vertical and before it is presented
- **Post-presentation features:** generated after the vertical is presented
  - possibly available from previous impressions of the query

Post-Presentation Features

- Derived from implicit feedback
- Click-through rates associated with query-vertical-position triplets
- Average dwell-times on vertical results
- Other feedback signals have not been explored in published work
  - Mouse movement
  - Scrolls
  - ....

Related Reading: Diaz [10], Ponnuswami *et al.* [23], Song *et al.* [30]

Post-Presentation Features

Example: Clickthrough

Table 2: Mean newsworthiness grade in each bin. Twenty samples from each bin of Winter 2008 queries were judged to be newsworthy or not. Treating newsworthy queries as having value 1 and 0 otherwise, we computed the mean and standard deviation of grades in each bin. We also present a third column which averages the mean bin values for both users

bin	user 1	user 2	mean
1	0.889 ± 0.314	1.000 ± 0.000	0.944
2	0.722 ± 0.448	0.944 ± 0.229	0.833
3	0.765 ± 0.424	0.944 ± 0.229	0.855
4	0.556 ± 0.497	1.000 ± 0.000	0.778
5	0.600 ± 0.490	0.941 ± 0.235	0.771
6	0.400 ± 0.490	0.706 ± 0.456	0.553
7	0.368 ± 0.482	0.647 ± 0.478	0.508
8	0.111 ± 0.314	0.412 ± 0.492	0.261
9	0.100 ± 0.300	0.222 ± 0.416	0.161
10	0.000 ± 0.000	0.200 ± 0.400	0.100

Related reading: Diaz [11]

Post-Presentation Features

- Can be derived from previous impressions of the query
- However, this assumes that the query is a *head* query
- Post presentation features can also be derived from *similar* queries
- **Assumption:** semantically related queries are associated with similar implicit feedback

$$\mathsf{click}(v,q) = \frac{1}{\mathcal{Z}} \sum_{q' \in \mathcal{Q} \ | \ \mathbf{sim}(q,q') > \tau} \mathbf{sim}(q,q') \times \mathsf{click}(v,q')$$

Related Reading: Diaz *et al.* [12]

Post-Presentation Features

nuances

- Some verticals do not require being clicked
  - weather, finance, translation, calculator
- Visually appealing verticals may exhibit a presentation bias
- A previous study found a click-through bias in favor of *video* results
- That is, users clicked on video results more often irrespective of position and relevance
- Suggests the need to model feedback differently for different verticals

Related Reading: Sushmita *et al.* [31]

Feature Importance for Vertical Selection

Which features help the most?

- Individually removing different types of features resulted in worse performance
- The most useful features corresponded to the topical categories of the query

all	0.583	
no.geographical	0.577▼	-1.01%
no.redde	0.568▼	-2.60%
no.soft.redde	0.567▼	-2.67%
no.category	0.552▼	-5.33%

■ corpus ■ query-log ■ query

Related Reading: Arguello *et al.* [4]

Feature Importance for Vertical Selection

Which features help the most?

- Explains why performance was superior for topically-focused verticals

travel	0.842
health	0.788
music	0.772
games	0.771
autos	0.730
sports	0.726
tv	0.716
movies	0.688
finance	0.655
local	0.619
jobs	0.570
shopping	0.563
images	0.483
video	0.459
news	0.456
reference	0.348
maps	0.000
directory	0.000

Related Reading: Arguello *et al.* [4]

Outline

Modeling

Content Aggregation

Tasks

- **Vertical Selection:** predicting *which* verticals to present
- **Vertical Presentation:** predicting *where* to present each vertical selected
  - May include deciding whether to suppress a selected vertical based on post-retrieval evidence

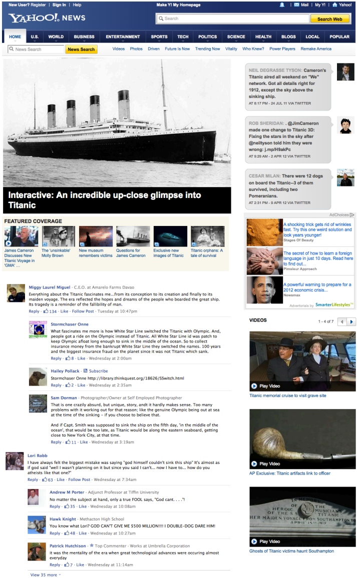
Content Aggregation

Layout Assumptions

- Content aggregation requires assuming a set of layout constraints
- Example layout constraints:
  - The core content is always presented and is presented in the same position
  - If a vertical is presented, then its results must be presented together (horizontally or vertically)
  - If a vertical is presented, then a minimum and maximum number of its results must be presented
  - If a vertical is presented, it can only be presented in certain positions
  - ....

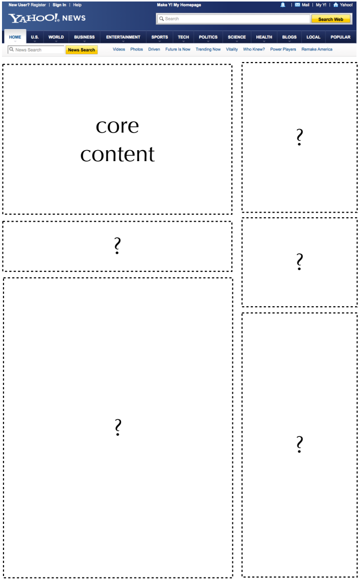
Content Aggregation

Layout Assumptions



Content Aggregation

Layout Assumptions



Aggregated Web Search

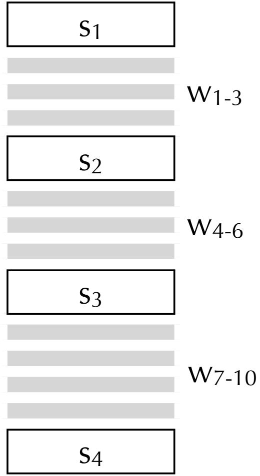
Layout Assumptions

- The core Web content (e.g., w$_{1-10}$) is always presented
- If a vertical is presented, then its results must be presented together (horizontally or vertically)
- If a vertical is presented, it can only be presented in certain positions relative to the top Web results, for example:
    - above w$_{1}$
    - between w$_{3-4}$
    - between w$_{6-7}$
    - below w$_{10}$

Aggregated Web Search

Layout Assumptions

- Because of these layout constraints, aggregated Web search is sometimes referred to as *slotting* or *blending*



Aggregated Web Search

Tasks

- **Vertical Selection:** predicting *which* vertical(s) to present
- **Vertical Presentation:** predicting *where* in the Web results to present them

Aggregated Web Search

Modeling

- Use machine learning to combine different types of features
- Vertical selection and presentation may be associated with different features
  - Post-retrieval features may not be available for selection
- **Gold-standard Training/Test Data**
  - Editorial vertical-relevance judgements: a human assessor determines that a particular vertical should be presented for a given query
  - User-generated clicks and skips collected by presenting the vertical (at a specific location) for all queries, or a random subset

Aggregated Web Search

Challenges

- Different verticals are associated with different types of features
  - Some verticals retrieve non-textual results (no vertical-corpus features)
  - Some verticals do not have direct search capabilities (no vertical query-log data)
  - Results from different verticals are associated with different meta-data (news articles have a publication date, local results have a geographical proximity to the user)
- A feature that is common to multiple verticals may be correlated differently with relevance (recency of results may be more predictive for the news vertical than the images vertical)

Aggregated Web Search

Challenges

- Requires methods that can handle different features for different verticals
- Requires methods that can exploit a vertical-specific relation between features and relevance

Vertical Selection

Classification Approach

- Learn independent vertical-specific binary classifiers
- Use binary ground truth labels for training
- Make independent binary predictions for each vertical

Classification Approach

Logistic Regression

$$P(v|q) = \frac{1}{1 + \exp\left(w_o + \sum_i w_i \times \phi_v(q)_i\right)}$$

- v = the vertical
- ϕ_v = feature generator specific to v (may include vertical-specific features)
- LibLinear: `http://www.csie.ntu.edu.tw/~cjlin/liblinear/`

Classification Approach

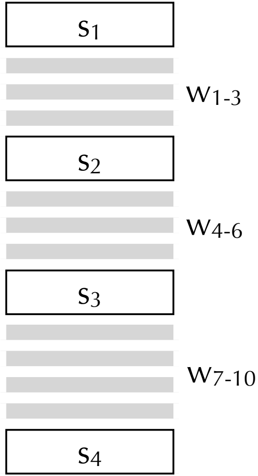
Logistic Regression

- **Advantages**
  - Easy and fast to train
  - Regularization parameter balances the importance of different features (helps improve generalization performance)
  - Outputs a confidence value $P(v|q)$, which can be treated as a hyper-parameter
- **Disadvantages**
  - Cannot exploit complex interactions between features

Vertical Presentation

Modeling

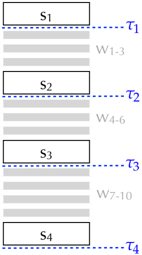
- **Slotting:** assume that vertical results can only be presented into specific locations or *slots*.



Vertical Presentation

Classification Approach

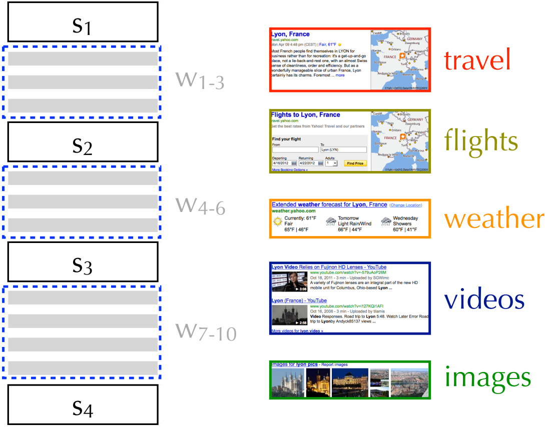
- Learn independent vertical-selectors using binary labels
- Present vertical v in slot s_i if $P(v|q) > \tau_j \; \forall \; j \geq i$
- Tune parameters τ_{1-4} using validation data



Related Reading: Arguello *et al.* [2], Ponnuswami *et al.* [23]

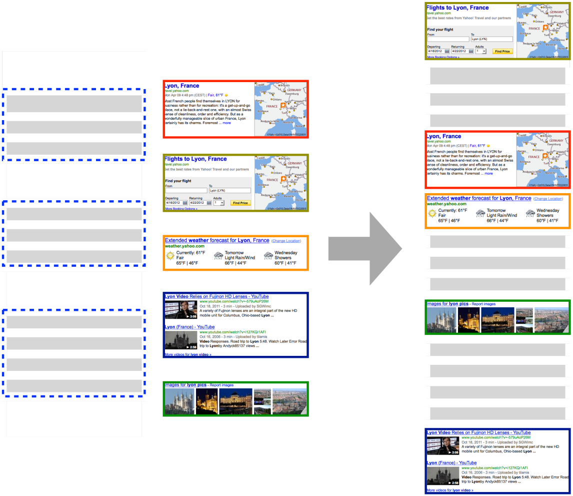
Vertical Presentation

Ranking Approach



Vertical Presentation

Block Ranking



Vertical Presentation

Learning To Rank

- Maching learning algorithms that learn to order elements based on training data
- Features derived from the element or from the query-element pair
- **Point-wise methods:** learn to predict an element's relevance grade independent of other elements
- **Pair-wise methods:** learn to predict whether one element is more relevant than another
- **List-wise methods:** learn to maximize a metric that evaluates the ranking as a whole (e.g., NDCG)

Block Ranking

Challenges

- LTR models require using a common feature representation across all elements
- LTR models learn an element-type agnostic relation between features and relevance
- Vertical block ranking requires methods that can handle different features for different verticals
- Vertical block ranking requires methods that can exploit a vertical-specific relation between features and relevance

SVM Rank

Learning To Rank

- These requirements can be satisfied by modifying the feature representation
  1. Use the union of all features
  2. If a feature is common to multiple verticals, make a vertical specific copy
  3. Zero all vertical-specific copies that do not correspond to the vertical in question

Query Similarity

- Similar queries should have similar predictions
- Gold-standard labels for training can also be shared between similar (labeled and unlabeled) queries.
- Related to semi-supervised learning.

Related Reading: Li *et al.* [20], Chapelle *et al.* [8]

Summary

- Vertical Selection
  - classification problem
  - can have a disjoint feature set amongst verticals
- Vertical Presentation
  - ranking problem
  - a common for verticals feature set is desirable

Evaluation

Vertical Selection

No Vertical

inaguration | search

Inauguration Day - Wikipedia
The swearing-in of the President of the United States occurs upon the commencement of a new term of a President of the United States. The United States Constitution mandates that the President make the following oath or...
http://en.wikipedia.org/wiki/United_States_presidential_inauguration

Joint Congressional Committee on Inaugural Ceremonies
Charged with planning and conducting the inaugural activities at the Capitol: the swearing-in ceremony and the luncheon honoring the President and Vice President.
http://inaugural.senate.gov

Inauguration Day 2009
Official site for the 2009 Inauguration of Barack Obama. Provides information about events, tickets, and inaugural balls and parades.
http::inaugural.senate.gov/2009

Inaugural Addresses of the Presidents of the United States
From George Washington's first address in 1789 to the present. Includes a note on the presidents who took the oath of office without a formal inauguration.
http://www.bartleby.com/124

News Vertical

inauguration | search

News Results for Inauguration

- **Online inauguration videos set records** CNN - 3 hours ago
- **Castro watched inauguration, Argentine leader says** CNN - 3 hours ago
- **Photographer: Inauguration like no moment I've ever witnessed** CNN - 4 hours ago

Inauguration Day - Wikipedia
The swearing-in of the President of the United States occurs upon the commencement of a new term of a President of the United States. The United States Constitution mandates that the President make the following oath or...
http://en.wikipedia.org/wiki/United_States_presidential_inauguration

Joint Congressional Committee on Inaugural Ceremonies
Charged with planning and conducting the inaugural activities at the Capitol: the swearing-in ceremony and the luncheon honoring the President and Vice President.
http://inaugural.senate.gov

Inauguration Day 2009
Official site for the 2009 Inauguration of Barack Obama. Provides information about events, tickets, and inaugural balls and parades.
http::inaugural.senate.gov/2009

Inaugural Addresses of the Presidents of the United States
From George Washington's first address in 1789 to the present. Includes a note on the presidents who took the oath of office without a formal inauguration.
http://www.bartleby.com/124

Evaluation Notation

Vertical Selection

$\mathcal{Q}$	set of evaluation contexts (queries)
$\mathcal{V}$	set of candidate verticals
	e.g. $\{\text{Web}, \text{news}, \ldots\}$
$\mathcal{V}_q$	set of verticals relevant to context q
$\tilde{v}_q$	predicted vertical for context q

Vertical Selection

Accuracy

- **relevance:** a vertical is *relevant* if satisfies some possible intent.
- **objective:** predict appropriate vertical when relevant; otherwise, predict no relevant vertical.
- **metric:** accuracy

Related reading: Arguello *et al.* [5]

Vertical Selection

Accuracy

$$\mathcal{A}_q = \begin{cases} \mathcal{I}(\tilde{v}_q \in \mathcal{V}_q) & \mathcal{V}_q \neq \emptyset \\ \mathcal{I}(\tilde{v}_q = \emptyset) & \mathcal{V}_q = \emptyset \end{cases}$$

Vertical Selection

Utility

- **relevance:** a vertical is *relevant* if satisfies the intent of a particular user at a particular time.
- **objective:** predict appropriate vertical when relevant; otherwise, predict no relevant vertical.
- **metric:** utility of whole page layout

Related reading: Diaz and Arguello [12]

Vertical Selection

Utility

$$u(v_q^*, \tilde{v}_q) = \begin{cases} 1 & v_q^* = \tilde{v}_q \\ \alpha & (v_q^* = \textsf{Web}) \wedge (\tilde{v}_q \neq \textsf{Web}) \\ 0 & \textsf{otherwise} \end{cases}$$

where $0 \leq \alpha \leq 1$ represents the user's discounted utility by being presented a display above the desired web results.

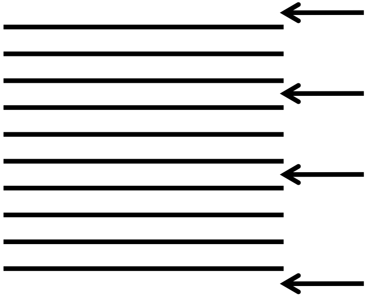
Vertical Presentation

Preference

- **relevance:** a presentation is good if the user can easily find more relevant content before less relevant content.
- **objective:** predict appropriate vertical preferences.
- **metric:** similarity to 'optimal' ranking.

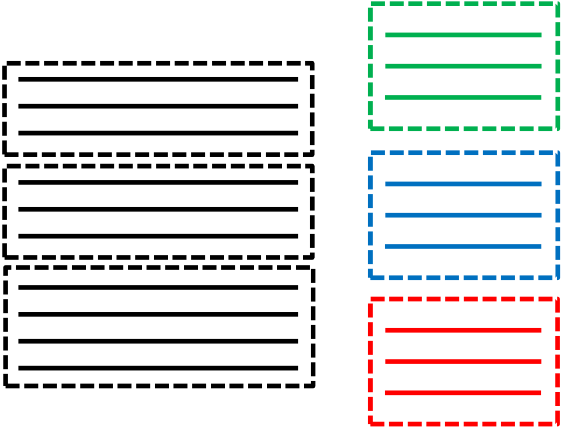
Related Reading: Diaz [11]

Candidate Slots



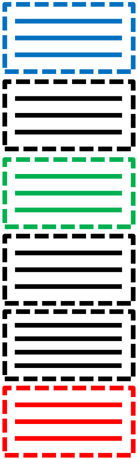
Candidate Modules

$\mathcal{C}_q$



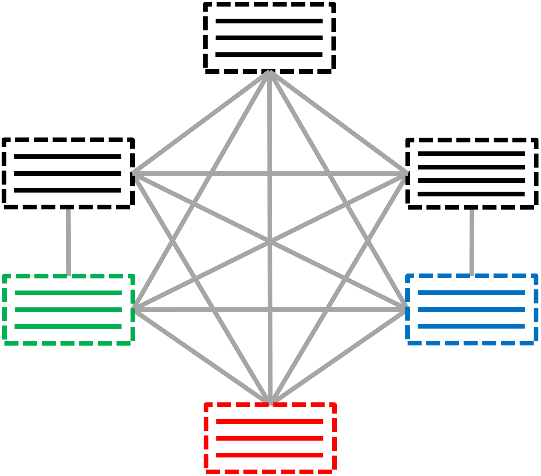
Aggregation

$\tilde{\sigma}_q$



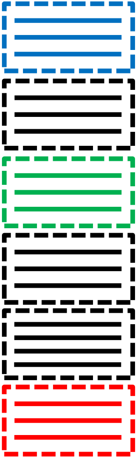
Metric

Module Preferences



Aggregation

σ_q^*



Metric

Evaluation

σ_q^*	optimal ranking from preference judgments
$\tilde{\sigma}_q$	predicted ranking by the system
$K(\sigma_q^*, \tilde{\sigma}_q)$	similarity between predicted and optimal rankings
	e.g. Kendall τ, Spearman ρ

Related reading: Arguello *et al.* [3]

Evaluation

- User study: small scale laboratory experiment resulting in a deep, focused understanding of the user-level effects.

- Batch study: medium scale laboratory experiment producing data and metrics for comparing systems.

- Production data: large scale production experiment gathering realistic user reactions to different systems.

Evaluation

- User study: small scale laboratory experiment resulting in a deep, focused understanding of the user-level effects.
  - advantages: fine-grained analysis of system behavior often situated in a real world task.
  - disadvantages: expensive; difficult to reuse results on drastically different systems; synthetic environment.
- Batch study: medium scale laboratory experiment producing data and metrics for comparing systems.
- Production data: large scale production experiment gathering realistic user reactions to different systems.

Evaluation

- User study: small scale laboratory experiment resulting in a deep, focused understanding of the user-level effects.
  - advantages: fine-grained analysis of system behavior often situated in a real world task.
  - disadvantages: expensive; difficult to reuse results on drastically different systems; synthetic environment.
- Batch study: medium scale laboratory experiment producing data and metrics for comparing systems.
  - advantages: repeatability; many metrics
  - disadvantages: expensive; synthetic environment.
- Production data: large scale production experiment gathering realistic user reactions to different systems.

Evaluation

- User study: small scale laboratory experiment resulting in a deep, focused understanding of the user-level effects.
  - advantages: fine-grained analysis of system behavior often situated in a real world task.
  - disadvantages: expensive; difficult to reuse results on drastically different systems; synthetic environment.
- Batch study: medium scale laboratory experiment producing data and metrics for comparing systems.
  - advantages: repeatability; many metrics
  - disadvantages: expensive; synthetic environment.
- Production data: large scale production experiment gathering realistic user reactions to different systems.
  - advantages: naturalistic experiment; large scale.
  - disadvantages: repeatability difficult.

Batch Study

- **editorial pool**: sampling editors to assess relevance (e.g. in-house editorial pool, mechanical turk).
- **query pool**: sampling queries to assess performance.
- **editorial guidelines**: defining precisely what is meant by relevance.

Batch Study

- **vertical selection**: queries labeled with all possible relevant verticals [4].
- **vertical presentation**: queries labeled with preferences between verticals [3].

Production Data

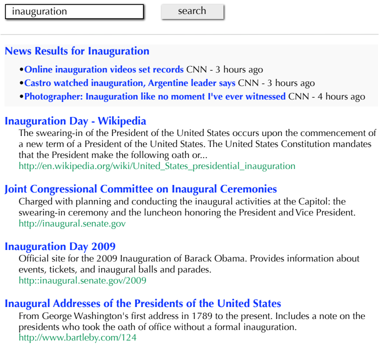
- **user pool**: sampling users to assess relevance (e.g. random, stratified).
- **implicit feedback**: defining user interactions correlated with relevance (e.g. clicks, hovers).

Production Data

- **vertical selection**: infer relevance from clicks on vertical displays.
- **vertical presentation**: infer preferences from clicks on vertical displays.

Production Data

Vertical Selection



- **click** on vertical content suggests relevance.
- **skip** over vertical content suggests non-relevance.
- **click through rate** summarizes the inferred relevance of a vertical.

Production Data

Vertical Ranking

inaguration search

News Results for Inauguration

- **Online inauguration videos set records** CNN - 3 hours ago
- **Castro watched inauguration, Argentine leader says** CNN - 3 hours ago
- **Photographer: Inauguration like no moment I've ever witnessed** CNN - 4 hours ago

Inauguration Day - Wikipedia
The swearing-in of the President of the United States occurs upon the commencement of a new term of a President of the United States. The United States Constitution mandates that the President make the following oath or...
http://en.wikipedia.org/wiki/United_States_presidential_inauguration

Joint Congressional Committee on Inaugural Ceremonies
Charged with planning and conducting the inaugural activities at the Capitol: the swearing-in ceremony and the luncheon honoring the President and Vice President.
http://inaugural.senate.gov

Inauguration Day 2009
Official site for the 2009 Inauguration of Barack Obama. Provides information about events, tickets, and inaugural balls and parades.
http::inaugural.senate.gov/2009

Inaugural Addresses of the Presidents of the United States
From George Washington's first address in 1789 to the present. Includes a note on the presidents who took the oath of office without a formal inauguration.
http://www.bartleby.com/124

- **skip** over a to **click** on result b implies b is preferred to a

Aggregating Performance

- Traffic-Weighted Average
  $\frac{1}{|\mathcal{D}|}\sum_{\langle u,t,q\rangle\in\mathcal{D}} \mathsf{perf}(u,t,q)$
  - focuses evaluation on the frequent queries
- Query-Stratified Average
  $\frac{1}{|\mathcal{Q}|}\sum_{q\in\mathcal{Q}} \frac{1}{\mathcal{D}_q}\sum_{\langle u,t,q\rangle\in\mathcal{D}_u} \mathsf{perf}(u,t,q)$
  - focuses evaluation on robustness across *queries*
- User-Stratified Average
  $\frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} \frac{1}{\mathcal{D}_u}\sum_{\langle u,t,q\rangle\in\mathcal{D}_u} \mathsf{perf}(u,t,q)$
  - focuses evaluation on robustness across *users*

Summary

- Many ways to evaluate aggregation performance.
  - Most important metric should be correlated with user satisfaction (e.g. whole-page relevance).
  - All metrics tell you *something* about the aggregation system.

Special Topics in Aggregation

Special Topics

- Dealing with non-stationary intent
- Dealing with new verticals
- Explore-exploit methods

Dealing with non-stationary intent

Dealing with non-stationary intent

- Vertical relevance depends on many factors.

- Addressing most factors can be addressed with enough data or careful sampling.

Dealing with non-stationary intent

- Vertical relevance depends on many factors.
  - core context: relevance may depend on the user's immediate intent

- Addressing most factors can be addressed with enough data or careful sampling.

Dealing with non-stationary intent

- Vertical relevance depends on many factors.
  - core context: relevance may depend on the user's immediate intent
  - geography: relevance may depend on the user's location (e.g. country, city, neighborhood)
- Addressing most factors can be addressed with enough data or careful sampling.

Dealing with non-stationary intent

- Vertical relevance depends on many factors.
  - core context: relevance may depend on the user's immediate intent
  - geography: relevance may depend on the user's location (e.g. country, city, neighborhood)
  - time: relevance may depend on recent events (e.g. holidays, news events)
- Addressing most factors can be addressed with enough data or careful sampling.

Dealing with non-stationary intent

- Vertical relevance depends on many factors.
  - core context: relevance may depend on the user's immediate intent
  - geography: relevance may depend on the user's location (e.g. country, city, neighborhood)
  - time: relevance may depend on recent events (e.g. holidays, news events)
- Addressing most factors can be addressed with enough data or careful sampling.
  - core context:
  - geography:
  - time:

Dealing with non-stationary intent

- Vertical relevance depends on many factors.
  - core context: relevance may depend on the user's immediate intent
  - geography: relevance may depend on the user's location (e.g. country, city, neighborhood)
  - time: relevance may depend on recent events (e.g. holidays, news events)
- Addressing most factors can be addressed with enough data or careful sampling.
  - core context: sample to include most expected contexts
  - geography:
  - time:

Dealing with non-stationary intent

- Vertical relevance depends on many factors.
  - core context: relevance may depend on the user's immediate intent
  - geography: relevance may depend on the user's location (e.g. country, city, neighborhood)
  - time: relevance may depend on recent events (e.g. holidays, news events)
- Addressing most factors can be addressed with enough data or careful sampling.
  - core context: sample to include most expected contexts
  - geography: sample to include many locations
  - time:

Dealing with non-stationary intent

- Vertical relevance depends on many factors.
  - core context: relevance may depend on the user's immediate intent
  - geography: relevance may depend on the user's location (e.g. country, city, neighborhood)
  - time: relevance may depend on recent events (e.g. holidays, news events)
- Addressing most factors can be addressed with enough data or careful sampling.
  - core context: sample to include most expected contexts
  - geography: sample to include many locations
  - time: sample to include many events?

Dealing with non-stationary intent

- Easy to predict appropriate news intent *retrospectively*
- Difficult to predict appropriate news intent *online*
- Possible solutions
  - Online learning of event models (e.g. statistical model of current events) [21]
  - Model with temporally-sensitive but context-independent features (e.g. 'rate of increase in document volume') [11]

Dealing with non-stationary intent

Online Language Models

- Language model: a statistical model of text production (e.g. web documents, news articles, query logs, tweets).
  - can compute the likelihood of a model having produced the text of a particular context (e.g. query, news article)
  - conjecture: relevance is correlated with likelihood
- Online language model:
  - Topic detection and tracking (TDT): model emerging topics using clusters of documents in a news article stream [1].
  - Social media modeling: model dynamic topics in social media (e.g. Twitter) [21, 24].

Dealing with non-stationary intent

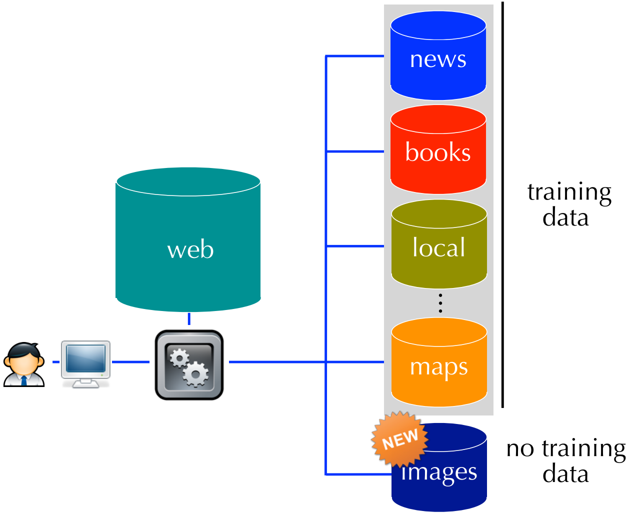
Context-Independent Features

- Approach: Instead of explicitly modeling the words associated with topics, model the topic-independent second order effects.
    - 'how quickly is this query spiking in volume?'
    - 'how quickly is the number documents retrieved spiking in volume?'
- Topic-independent features generalize across events and *into the future* (as long as the new events behave similar to historic events)

Domain Adaptation for Vertical Selection

Problem

- Supervised vertical selection requires training data (e.g., vertical-relevance judgements).



Problem

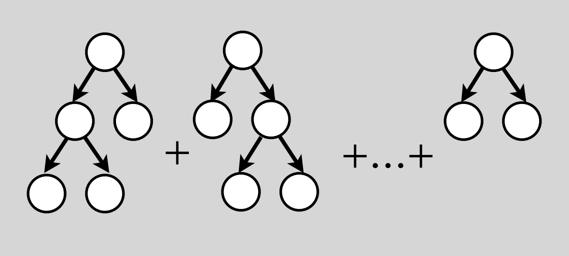
- A model trained to predict one vertical may not generalize to another
- Why?
  - A feature that is correlated with relevance for one vertical may be uncorrelated or negatively correlated for another (e.g., whether the query contains "news")

Task Definition

- **Task:** Given a set of source verticals $\mathcal{S}$ with training data, learn a predictive model of a target vertical t associated with no training data.
- **Objective:** Maximize effectiveness on the target vertical

Learning Algorithm

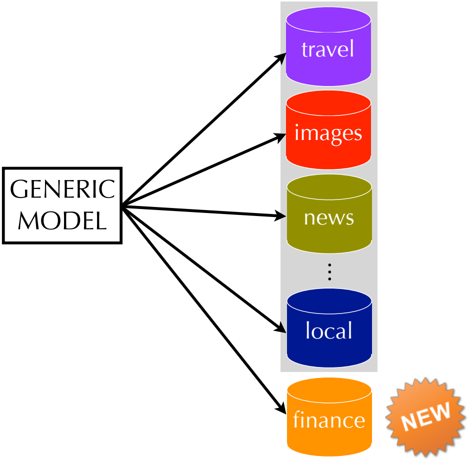
- Gradient Boosted Decision Trees (GBDT)



- Iteratively trains decision tree predictors fitting the residuals of preceding trees
- Minimizes logistic loss

Generic Model

- Train a model to maximize (average) performance for all verticals in $\mathcal{S}$ and apply the model to the target t.



Generic Model

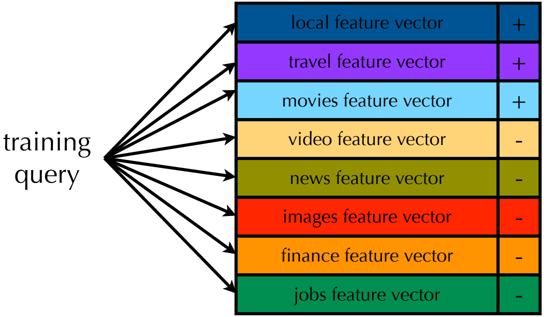
- **Assumption:** if the model performs well across *all* source verticals $\mathcal{S}$, it will generalize well to the target t.

Training a Generic Model

- Share a common feature representation across all verticals
- Pool together each source vertical's training data into a single training set
- Perform standard GBDT training on this training set

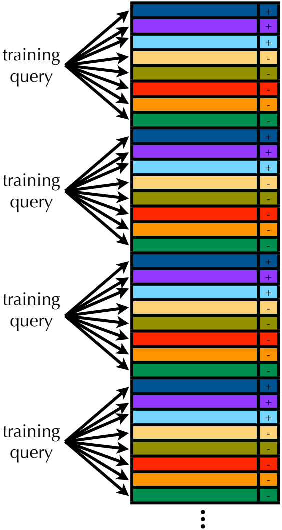
Training a Generic Model

- Each training set query is represented by $|\mathcal{S}|$ instances (one per source vertical)



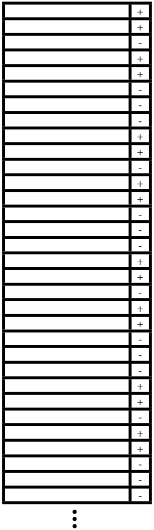
Training a Generic Model

- **Training set:** union of *all* query/source-vertical pairs



Training a Generic Model

- Perform standard GBTD training on this training set
- The model should automatically ignore features that are *inconsistently* correlated with the positive class

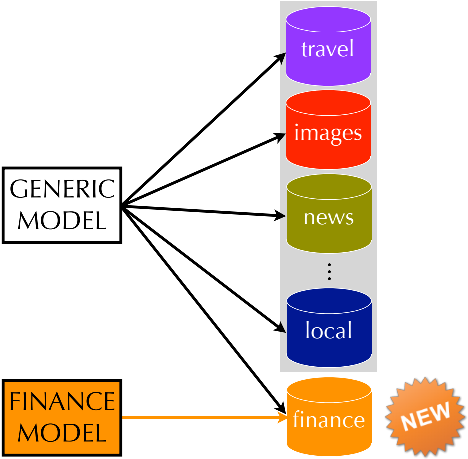


Portable Feature Selection

- **Goal:** Automatically identify features that may be uncorrelated (or negatively correlated) with relevance across verticals in $\mathcal{S}$
- **Method**
  - Treat each feature as a single-evidence predictor
  - Measure its performance on each source vertical in $\mathcal{S}$
  - Keep only the ones with the greatest (harmonic) average performance
  - **Assumption:** if the feature is correlated with relevance (in the same direction) for all verticals in $\mathcal{S}$, it will generalize well to the target t.

Model Adaptation

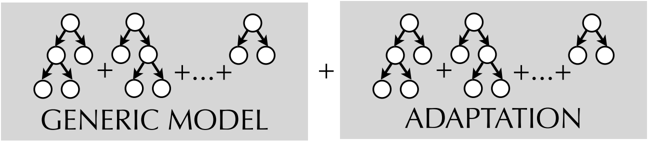
- Use the generic model's most confident predictions on the target t to "bootstrap" a vertical-specific model



Tree-based Domain Adaptation

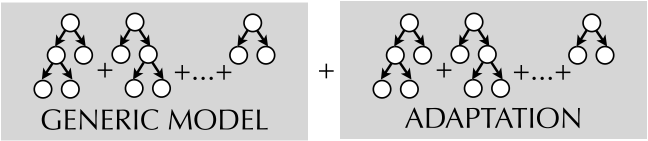
[Chen *et al.* 2008]

- **TRADA:** adapting a source-trained model using a small amount of target domain training data
- A GBDT model can be fit to new data (from whatever domain) by simply appending new trees to the current ensemble

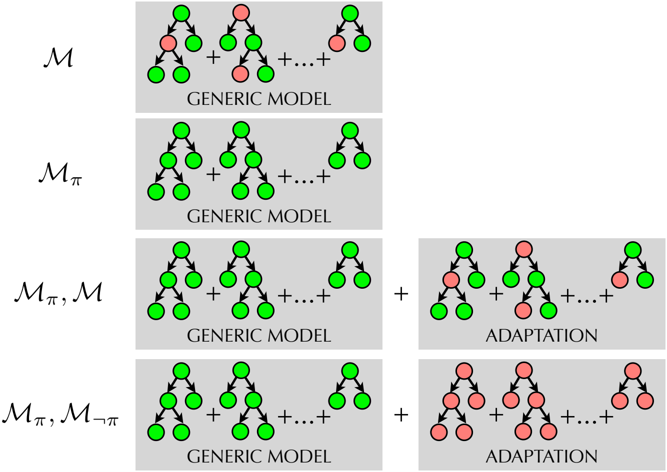


Training an Adapted Model

- Use a generic model to make target-vertical predictions on unlabeled data
- Consider the most confident N% predictions as positive examples and the remaining ones as negative examples
- Adapt the generic model by appending trees while fitting the residuals of the generic model to its own predictions



Evaluation



portable features

non-portable features

Generic Model Results

Average Precision

vertical	generic (all feats.)	generic (only portable feats.)
finance	0.209	0.392▲
games	0.636	0.683
health	0.797	0.839
jobs	0.193	0.321▲
images	0.365	0.390
local	0.543	0.628▲
movies	0.294	0.478▲
music	0.673	0.780▲
news	0.293	0.548▲
travel	0.571	0.639▲
video	0.449	0.691▲

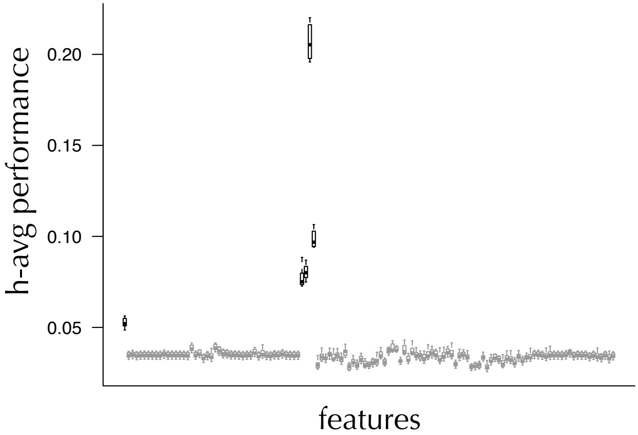
Model Adaptation Results

Average Precision

vertical	generic (only portable feats.)	trada (all feats.)	trada (only non-portable feats.)
finance	0.392	0.328	0.407
games	0.683	0.660	0.817▲
health	0.839	0.813	0.868
jobs	0.321	0.384	0.348
images	0.390	0.370	0.499▲
local	0.628	0.601	0.614
movies	0.478	0.462	0.587▲
music	0.780	0.778	0.866▲
news	0.548	0.556	0.665▲
travel	0.639	0.573▼	0.709▲
video	0.691	0.648	0.722

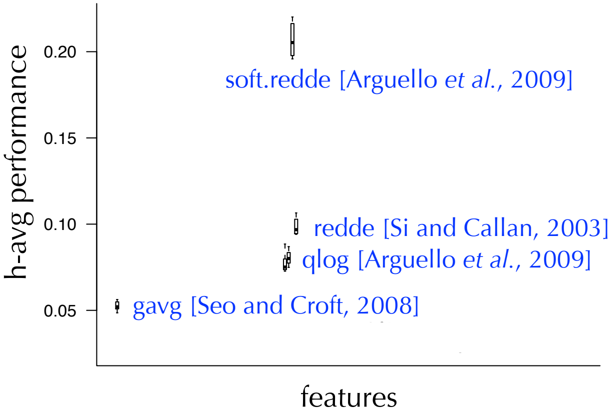
Feature Portability

- Some features were more portable than others (high h-avg performance across source verticals)



Feature Portability

- The most portable features correspond to unsupervised methods designed for homogeneous federated search



Summary

- A generic vertical selector can be trained with some success
- Focusing on only the most portable features (identified automatically) improves performance
- Vertical-specific non-portable evidence is useful but requires training data
- Can be harnessed using pseudo-training examples from a generic model's predictions at *no additional cost*

Explore/exploit methods

Explore/Exploit

- Exploit: Choose articles/sources with highest expected quality for short-term reward.
- Explore: Choose articles/sources with lower expected reward for long-term reward.
- Typical solution: Multi-arm bandits

Related reading: Li et al. [18]

Multi-armed bandit

- Problem: Each news article/source can be considered as an arm.
- Task: Present stories (pull arms) to maximize long term reward (clicks).

Related reading: Li et al. [18, 19]

Multi-armed bandit

- Naïve strategy: with probability ϵ, present a *random* article.
- Better strategy: sample according to confidence that the article is relevant (e.g. Exp4, Boltzmann sampling)
- Extension: incorporating prior information (a.k.a. contextual bandits)

Related reading: Auer *et al.* [6], Sutton and Barto [32]

Future Directions

Short Term

- Diversity in vertical presentation
  - verticals can be redundant (e.g. news, blogs)
  - verticals can be complementary (e.g. news, video)
  - should evaluate *whole page relevance* when aggregating
- Attention modeling
  - some verticals can visually attract a user's attention without being relevant (e.g. graphic advertisements)
  - need a better understanding of attention in response to automatic page layout decisions

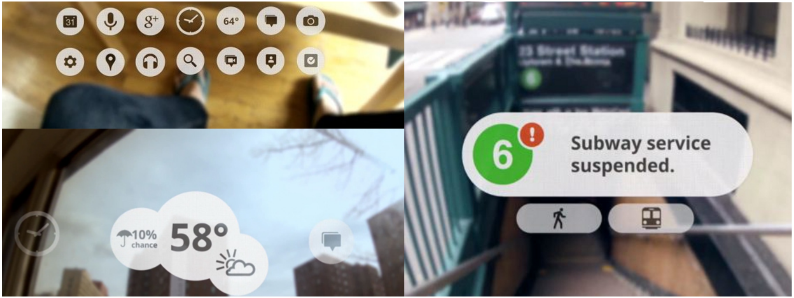
Medium Term

- Relaxed layout constraints
  - the majority of work has focused on aggregation into a conservative ranked list layout.
  - can we consider arbitrary layouts?
- Detecting new verticals
  - currently the set of candidate verticals is manually curated
  - can we automatically detect that we should begin modeling a new vertical?

In Future, You are the Core Content

Related Reading: Kinect
http://www.xbox.com/en-US/kinect

In Future, You are the Core Content



Related Reading: Project Glass `http://bit.ly/HGqZcV`

References

[1] James Allan (ed.), *Topic detection and tracking: Event-based information organization*, The Information Retrieval Series, vol. 12, Springer, New York, NY, USA, 2002.

[2] Jaime Arguello, Fernando Diaz, and Jamie Callan, *Learning to aggregate vertical results into web search results*, CIKM 2011, ACM, 2011, pp. 201–210.

[3] Jaime Arguello, Fernando Diaz, Jamie Callan, and Ben Carterette, *A methodology for evaluating aggregated search results*, ECIR 2011, Springer Berlin / Heidelberg, 2011, pp. 141–152.

[4] Jaime Arguello, Fernando Diaz, Jamie Callan, and Jean-François Crespo, *Sources of evidence for vertical selection*, SIGIR 2009, ACM, 2009, pp. 315–322.

[5] Jaime Arguello, Fernando Diaz, Jamie Callan, and Jean-François Crespo, *Sources of evidence for vertical*

selection, Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, 2009, pp. 315–322.

[6] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire, *The non-stochastic multi-armed bandit problem*, SIAM Journal on Computing **32** (2002), no. 1, 48–77.

[7] Jamie Callan and Margaret Connell, *Query-based sampling of text databases*, TOIS **19** (2001), no. 2, 97–130.

[8] O. Chapelle, B. Schölkopf, and A. Zien (eds.), *Semi-supervised learning*, MIT Press, Cambridge, MA, 2006.

[9] Steve Cronen-Townsend, Yun Zhou, and W. Bruce Croft, *Predicting query performance*, Proceedings of the 25th annual international ACM SIGIR conference on Research

and development in information retrieval (Tampere, Finland), SIGIR '02, ACM, 2002, pp. 299–306.

[10] Fernando Diaz, *Integration of news content into web results*, WSDM 2009, ACM, 2009, pp. 182–191.

[11] Fernando Diaz, *Integration of news content into web results*, Proceedings of the Second ACM International Conference on Web Search and Data Mining, 2009.

[12] Fernando Diaz and Jaime Arguello, *Adaptation of offline vertical selection predictions in the presence of user feedback*, SIGIR 2009, 2009.

[13] Norbert Fuhr, *A decision-theoretic approach to database selection in networked ir*, ACM Trans. Inf. Syst. **17** (1999), no. 3, 229–249.

[14] Luis Gravano, Chen-Chuan K. Chang, Hector Garcia-Molina, and Andreas Paepcke, *Starts: Stanford*

protocol proposal for internet retrieval and search, Technical Report 1997-68, Stanford InfoLab, April 1997, Previous number = SIDL-WP-1996-0043.

[15] A. Jennings and H. Higuchi, *A personal news service based on a user model neural network*, IEICE Transactions on Information and Systems **75** (1992), no. 2, 192–209.

[16] Tomonari Kamba, Krishna Bharat, and Michael C. Albers, *The krakatoa chronicle - an interactive, personalized, newspaper on the web*, In Proceedings of the Fourth International World Wide Web Conference, 1995, pp. 159–170.

[17] Marina Krakovsky, *All the news that's fit for you*, Communications of the ACM **54** (2011), no. 6, 20–21.

[18] Lihong Li, Wei Chu, John Langford, and Robert E. Schapire, *A contextual-bandit approach to personalized news article recommendation*, Proceedings of the 19th

international conference on World wide web (Raleigh, North Carolina, USA), WWW '10, ACM, 2010, pp. 661–670.

[19] Lihong Li, Wei Chu, John Langford, and Xuanhui Wang, *Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms*, Proceedings of the fourth ACM international conference on Web search and data mining (Hong Kong, China), WSDM '11, ACM, 2011, pp. 297–306.

[20] Xiao Li, Ye-Yi Wang, and Alex Acero, *Learning query intent from regularized click graphs*, Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval (New York, NY, USA), SIGIR '08, ACM, 2008, pp. 339–346.

[21] Jimmy Lin, Rion Snow, and William Morgan, *Smoothing techniques for adaptive online language models: topic*

tracking in tweet streams, Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (New York, NY, USA), KDD '11, ACM, 2011, pp. 422–429.

[22] Jie Lu, *Full-text federated search in peer-to-peer networks*, Ph.D. thesis, Language Technologies Institute, Carnegie Mellon University, 2007.

[23] Ashok Kumar Ponnuswami, Kumaresh Pattabiraman, Qiang Wu, Ran Gilad-Bachrach, and Tapas Kanungo, *On composition of a federated web search result page: Using online users to provide pairwise preference for heterogeneous verticals*, WSDM 2011, ACM, 2011, pp. 715–724.

[24] Ankan Saha and Vikas Sindhwani, *Learning evolving and emerging topics in social media: a dynamic nmf approach with temporal regularization*, WSDM, 2012, pp. 693–702.

[25] Dou Shen, Jian-Tao Sun, Qiang Yang, and Zheng Chen, *Building bridges for web query classification*, SIGIR 2006, ACM, 2006, pp. 131–138.

[26] Milad Shokouhi and Luo Si, *Federated search*, Foundations and Trends in Information Retrieval **5** (2011), no. 1, 1–102.

[27] Milad Shokouhi, Justin Zobel, Saied Tahaghoghi, and Falk Scholer, *Using query logs to establish vocabularies in distributed information retrieval*, IPM **43** (2007), no. 1, 169–180.

[28] Luo Si and Jamie Callan, *Relevant document distribution estimation method for resource selection*, SIGIR 2003, ACM, 2003, pp. 298–305.

[29] Luo Si and Jamie Callan, *A semisupervised learning method to merge search engine results*, ACM Trans. Inf. Syst. **21** (2003), no. 4, 457–491.

[30] Yang Song, Nam Nguyen, Li-wei He, Scott Imig, and Robert Rounthwaite, *Searchable web sites recommendation*, WSDM 2011, ACM, 2011, pp. 405–414.

[31] Shanu Sushmita, Hideo Joho, Mounia Lalmas, and Robert Villa, *Factors affecting click-through behavior in aggregated search interfaces*, CIKM 2010, ACM, 2010, pp. 519–528.

[32] Richard Sutton and Andrew Barto, *Reinforcement learning*, MIT Press, 1998.

[33] Gabor Szabo and Bernardo A. Huberman, *Predicting the popularity of online content*, Communications of the ACM **53** (2010), no. 8, 80–88.

[34] Jinfeng Zhuang, Tao Mei, Steven C.H. Hoi, Ying-Qing Xu, and Shipeng Li, *When recommendation meets mobile: contextual and personalized recommendation on the go*, Proceedings of the 13th international conference on Ubiquitous computing (Beijing, China), UbiComp '11, ACM, 2011, pp. 153–162.