Modeling User Behavior and Interactions

Lecture 4: **Search Personalization**

Eugene Agichtein
Emory University

Lecture 4 Outline

1. **Approaches to Search Personalization**

2. **Dimensions of Personalization**
    1. **Which queries to personalize?**
    2. **What input to use for personalization?**
    3. **Granularity: personalization vs. groupization**
    4. **Context: Geograpical, search session**

Approaches to Personalization

1. Pitkow et al., 2002
2. Qiu et al., 2006
3. Jeh et al., 2003
4. Teevan et al., 2005
5. Das et al., 2007

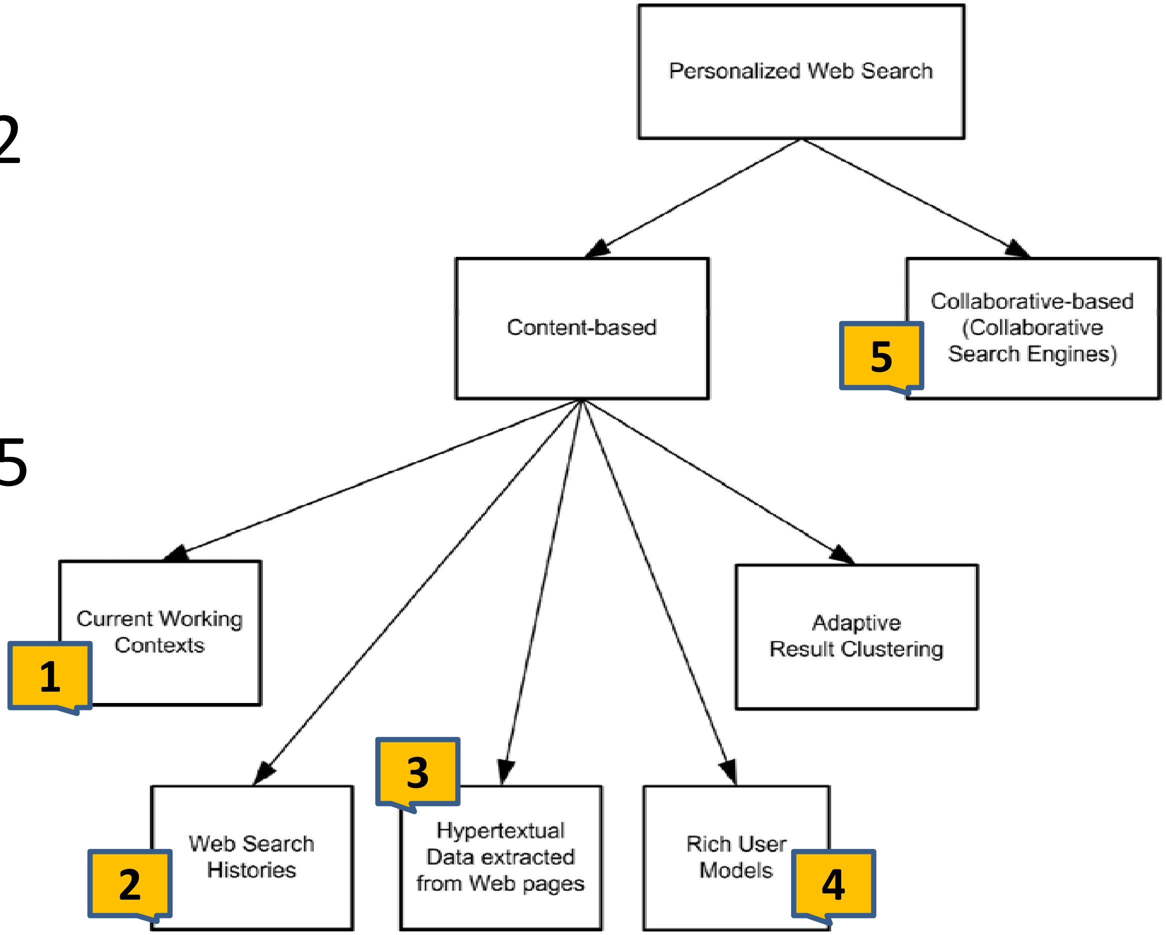


Figure adapted from: *Personalized search on the world wide web*, by Micarelli, A. and Gasparetti, F. and Sciarrone, F. and Gauch, S., LNCS 2007

When to Personalize

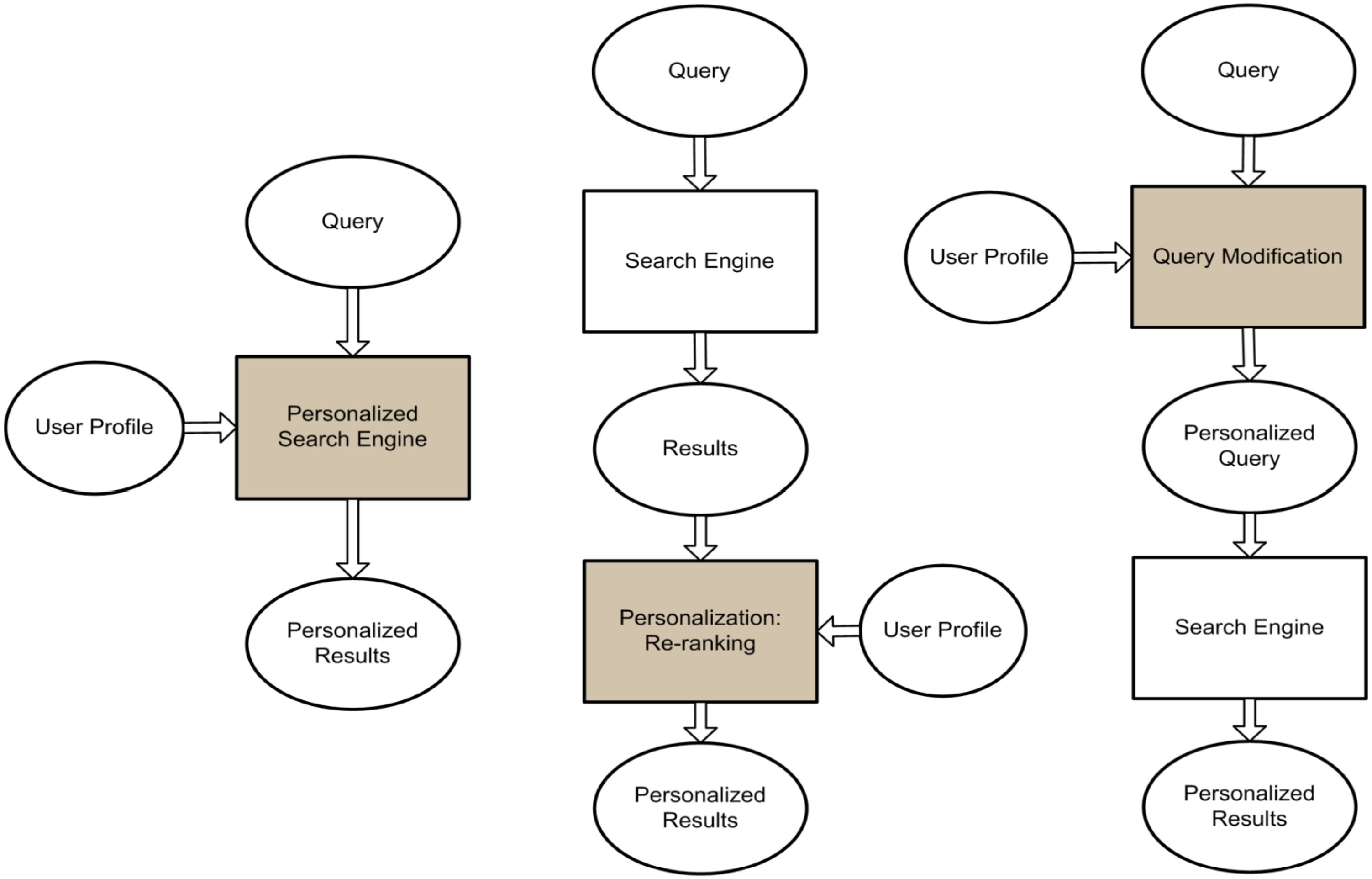
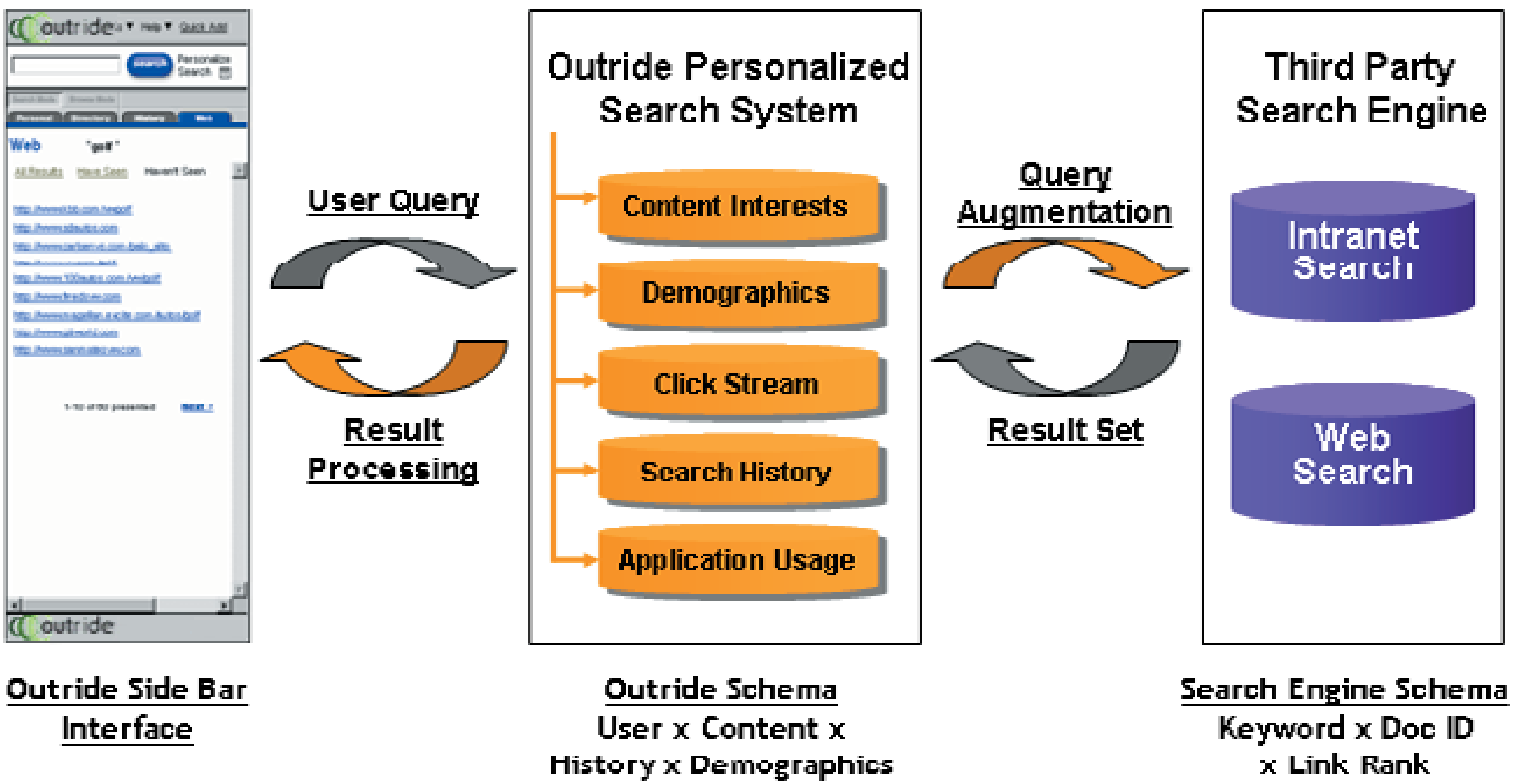


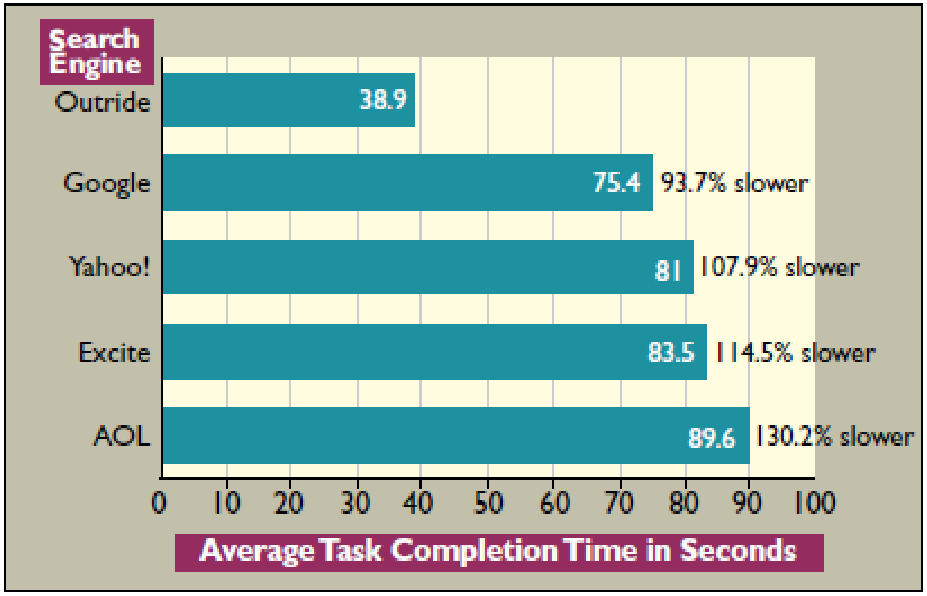
Figure adapted from: *Personalized search on the world wide web*, by Micarelli, A. and Gasparetti, F. and Sciarrone, F. and Gauch, S., LNCS 2007

Example: *Outride*

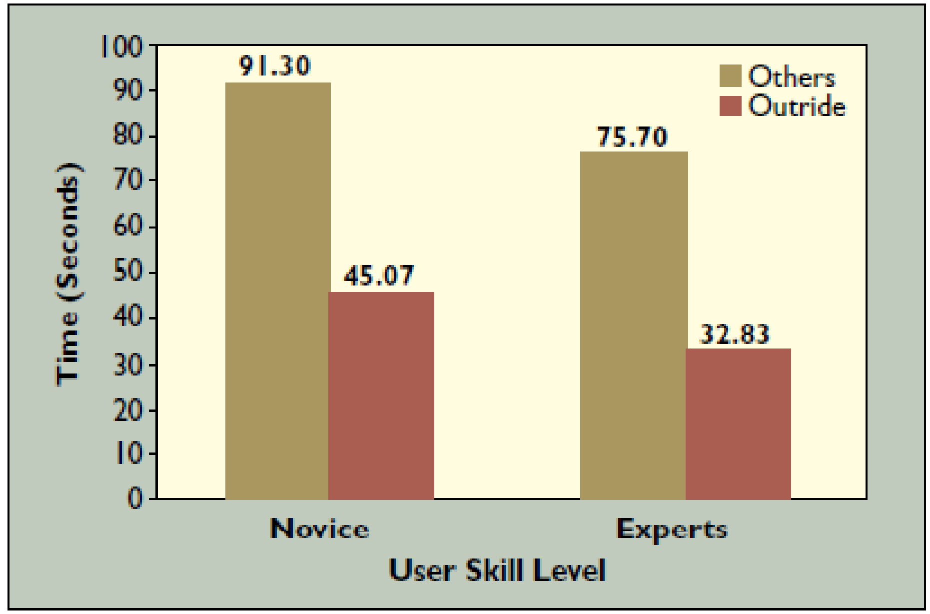


From Pitkow et al., 2002

Outride (Results)



(% slower from Outride enabled search)
Source: ZDLabs/eTesting, Inc. Oct. 2000



(Average time to complete task)
Source: ZDLabs/eTesting, Inc. Oct. 2000

Search Engine	User Actions	Difference (%)
Outride	11.2	
Google	21.2	89.6
Yahoo!	22.4	100.5
AOL	23.1	107.0
Excite	23.3	108.5
Average	22.5	101.4

Source: ZDLabs/eTesting, Inc. Oct. 2000

Engine	Expert Time	Rank	Novice Time	Rank	Average	Rank	% Difference
Outride	32.8	(1)	45.1	(1)	38.9	(1)	0%
AOL	92.3	(5)	87.0	(4)	89.6	(5)	130.2%
Excite	75.7	(3)	91.3	(5)	83.5	(4)	114.5%
Google	72.5	(2)	78.4	(3)	75.4	(2)	93.7%
Yahoo!	85.1	(4)	76.9	(2)	81.0	(3)	107.9%

(in seconds, with placement in parenthesis)
Source: ZDLabs/eTesting, Inc. Oct. 2000

From Pitkow et al., 2002

Input to Personalization

- Behavior (clicks): Qiu and Cho, 2006
  - Use clicks to tune a personalized (topic sensitive) **PageRank** model for each user
  - Use **personalized PageRank** to re-rank web search results
- Profile (user model): SeeSaw (Teevan et al., 2005)

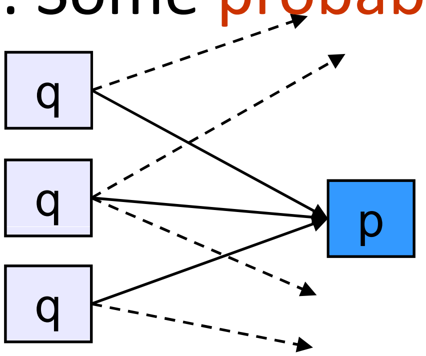
PageRank Computation

I: Set of Incoming links

O: Set of Outgoing links

c: Dampening factor (~0.15) or “teleportation probability”

E: Some probability vector over the Webpages



$$PR(p) = (1-c) \cdot \sum_{q \in I(p)} \frac{PR(q)}{\|O(q)\|} + c\,E(p)$$

E vector can be:

- Uniformly distributed probabilities over all Web Page (democratic)
- Biased distributed probabilities to a number of important pages
  - Top-levels of Web Servers
  - Hub/ Authority pages
- Used for Customization (Personalization)

Topic-Sensitive PageRank

• Uninfluenced PageRank *"Page is important if many important pages point to it"*	• Influenced PageRank *"Page is important if many important pages point to it, and btw, the following are by definition important pages."*

Main Idea

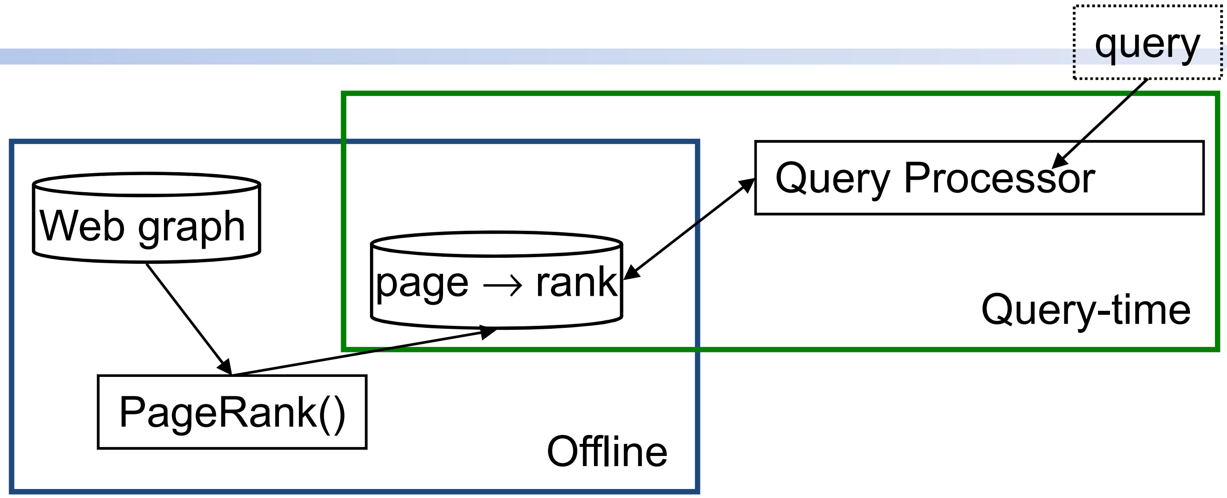
- Assign *multiple* a-priori "importance" estimates to pages with respect to a set of *topics*
- One PageRank score per *basis topic*
  - Query specific rank score (+)
  - Make use of context (+)
  - Inexpensive at runtime (+)

PageRank vs Topic-Sensitive PageRank

PageRank

Input:
Web graph *G*

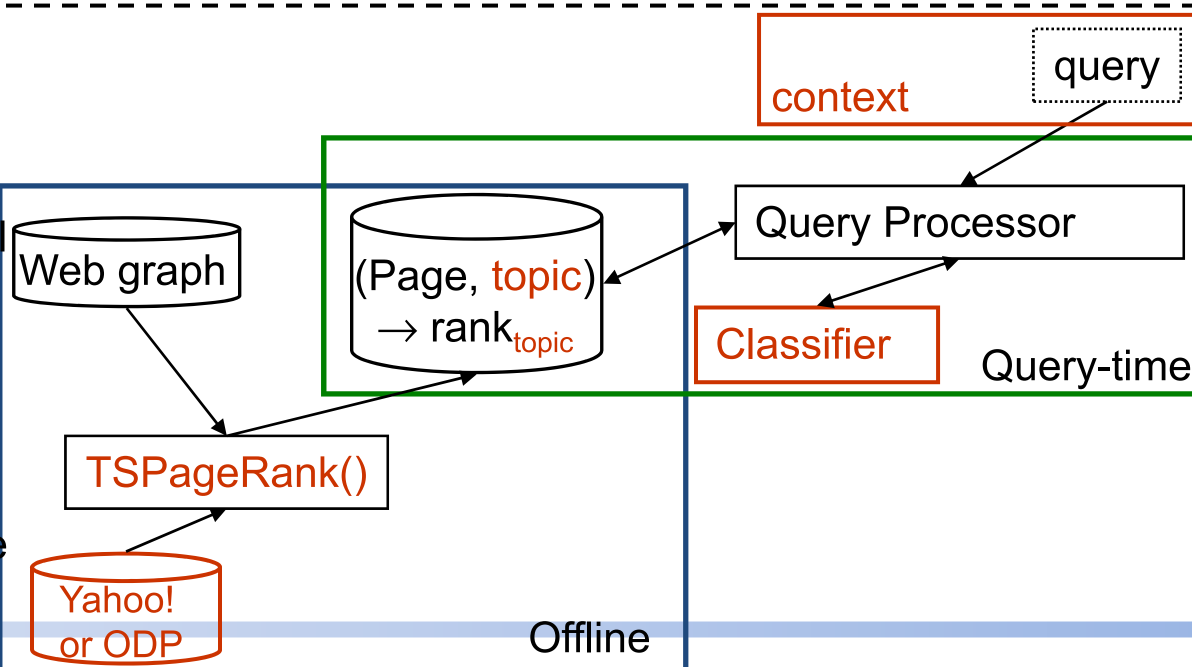
Output:
Rank vector
r : (page → page importance)



Topic-Sensitive PageRank

Input:
Web *W*, Basis topics [c1, ... ,c16]
e.g. 16 categories (first level of ODP)

Output:
List of rank vectors [r1, ... ,r16]
r_j : page → page importance in topic c_j



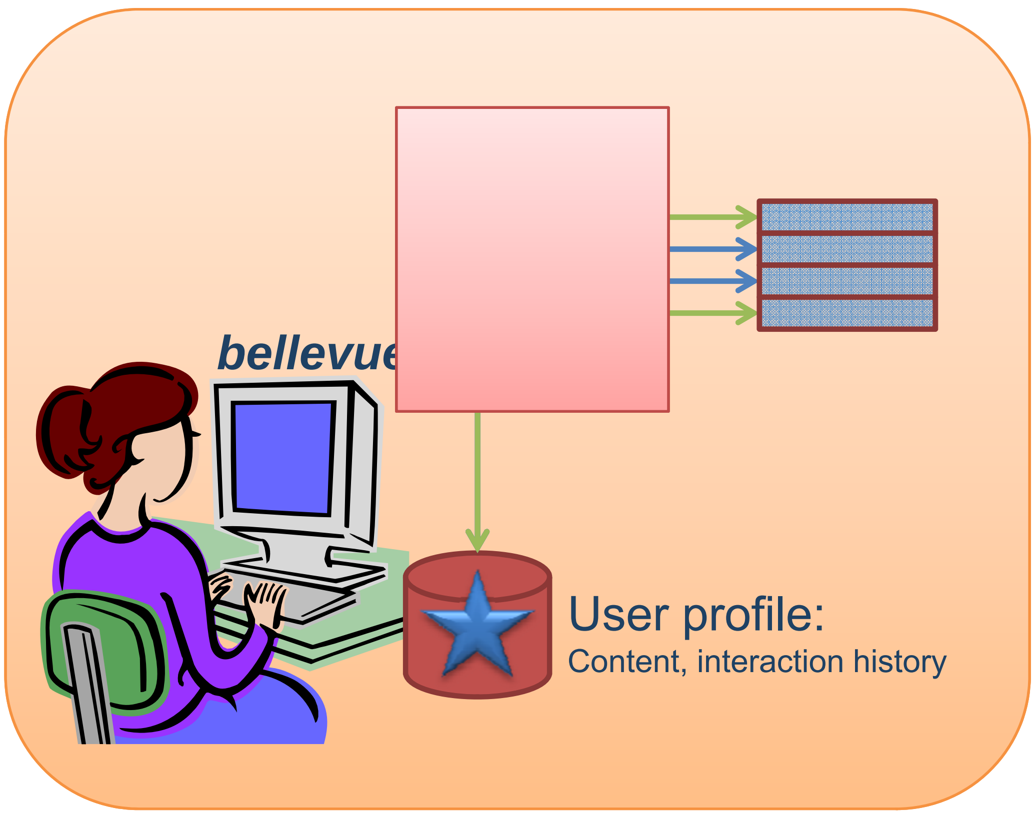
Input to Personalization

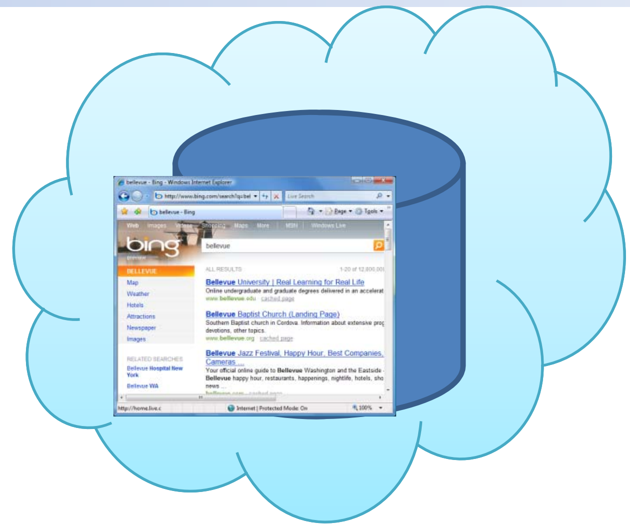
- Behavior (clicks): Qiu and Cho, 2006
  – Use clicks to tune a personalized (topic sensitive) PageRank model for each user
    - Map clicked results to ODP
  – Use personalized PageRank to re-rank web search results

- **Profile (user model): SeeSaw (Teevan et al., 2005)**

PS Search Engine (Profile-based)

[Teevan et al., 2005]

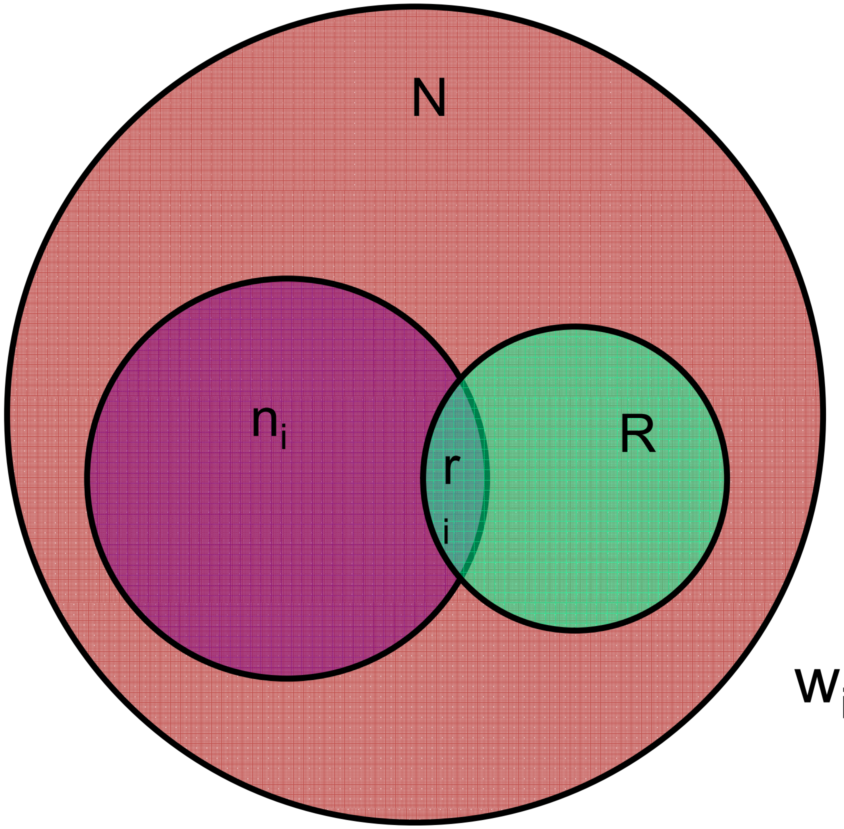




Result Re-Ranking

- Ensures privacy
- Good evaluation framework
- Can look at rich user profile
- Look at light weight user models
  - Collected on server side
  - Sent as query expansion

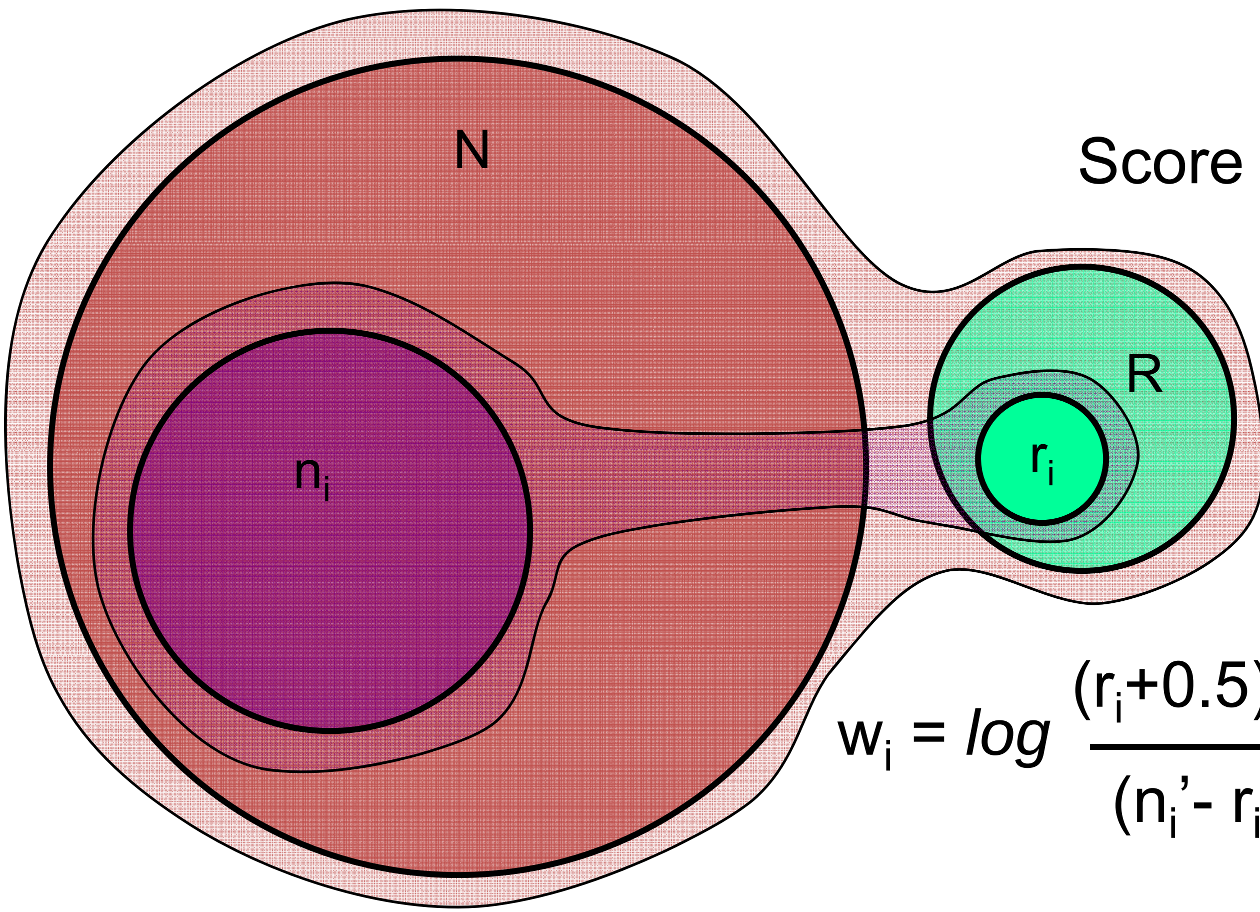
BM25 with Relevance Feedback



$$\text{Score} = \sum tf_i * w_i$$

$$w_i = \log \frac{(r_i+0.5)(N-n_i-R+r_i+0.5)}{(n_i-r_i+0.5)(R-r_i+0.5)}$$

User Model as Relevance Feedback



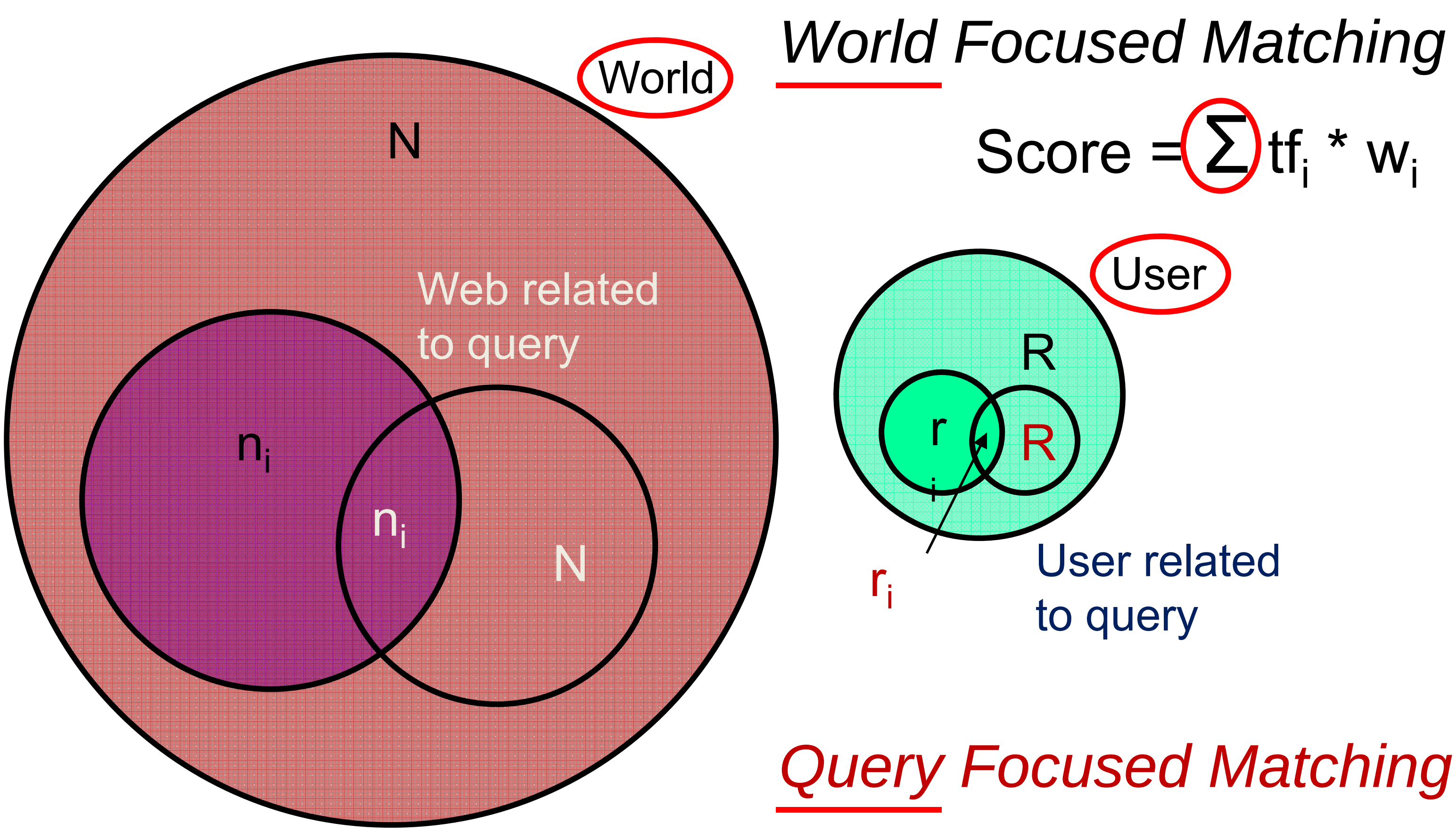
$$Score = \Sigma \, tf_i * w_i$$

$$N' = N+R$$

$$n_i' = n_i+ri$$

$$w_i = log \frac{(r_i+0.5)(N'-n_i'-R+r_i+0.5)}{(n_i'- r_i+0.5)(R-r_i+0.5)}$$

User Model as Relevance Feedback



User Representation

- Stuff I've Seen (SIS) index
  - MSR research project [Dumais, et al.]
  - Index of everything a user's seen
- Recently indexed documents
- Web documents in SIS index
- Query history
- None

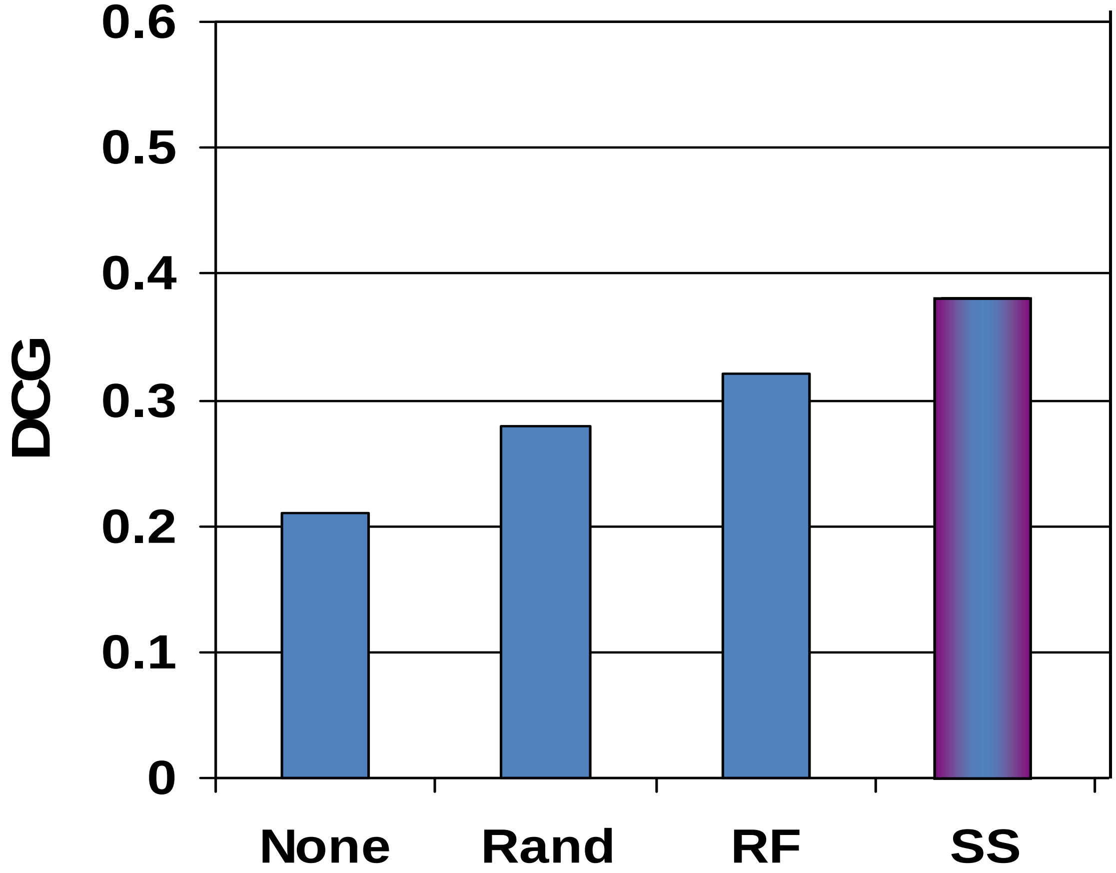
World Representation

- Document Representation
  - Full text
  - Title and snippet
- Corpus Representation
  - Web
  - Result set – title and snippet
  - Result set – full text

Parameters

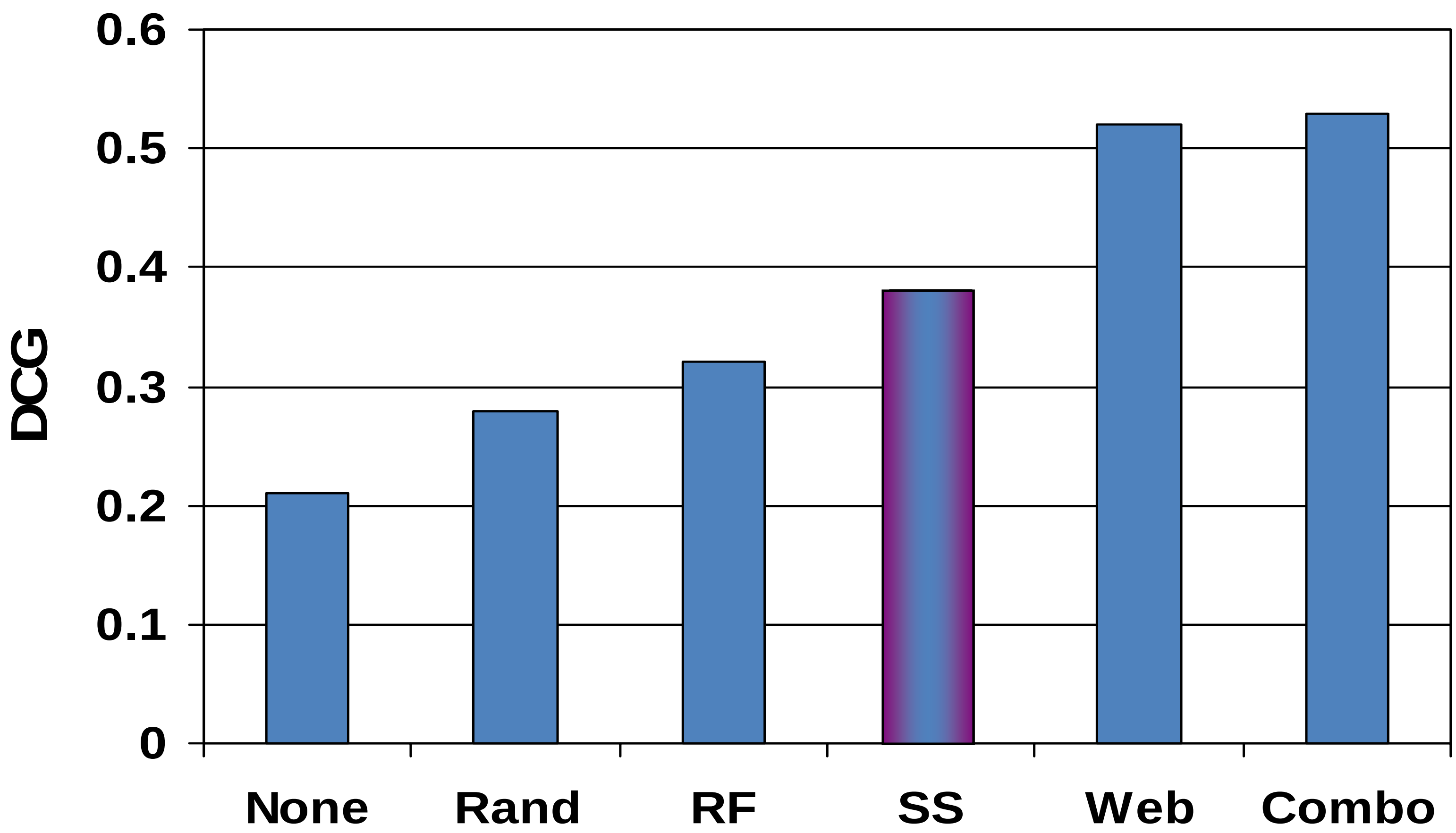
- Matching
  - Query focused
  - World focused
- User representation
  - All SIS
  - Recent SIS
  - Web SIS
  - Query history
  - None
- World representation
  - Full text
  - Title and snippet
- Query expansion
  - Web
  - Result set – full text
  - Result set – title and snippet

Results: *Seesaw* Improves Retrieval



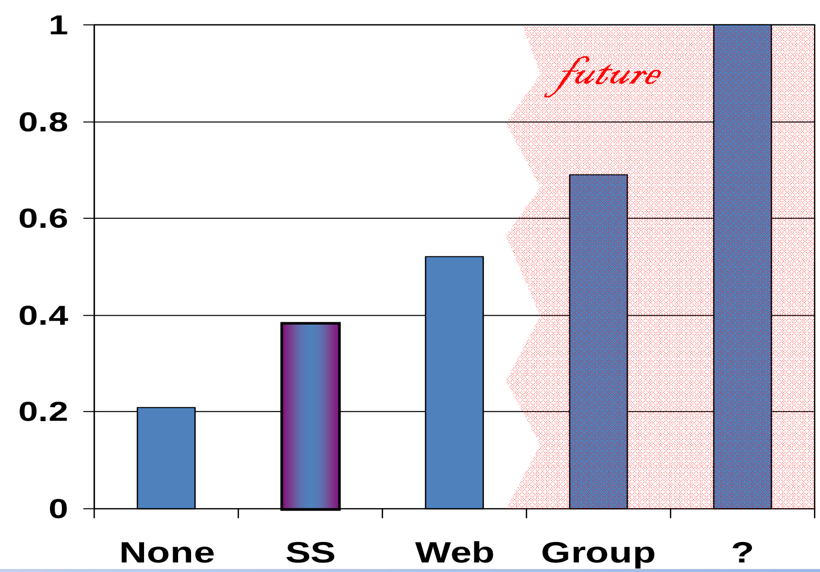
- No user model
- Random
- Relevance Feedback
- *Seesaw*

Results: Feature Contribution



Summary

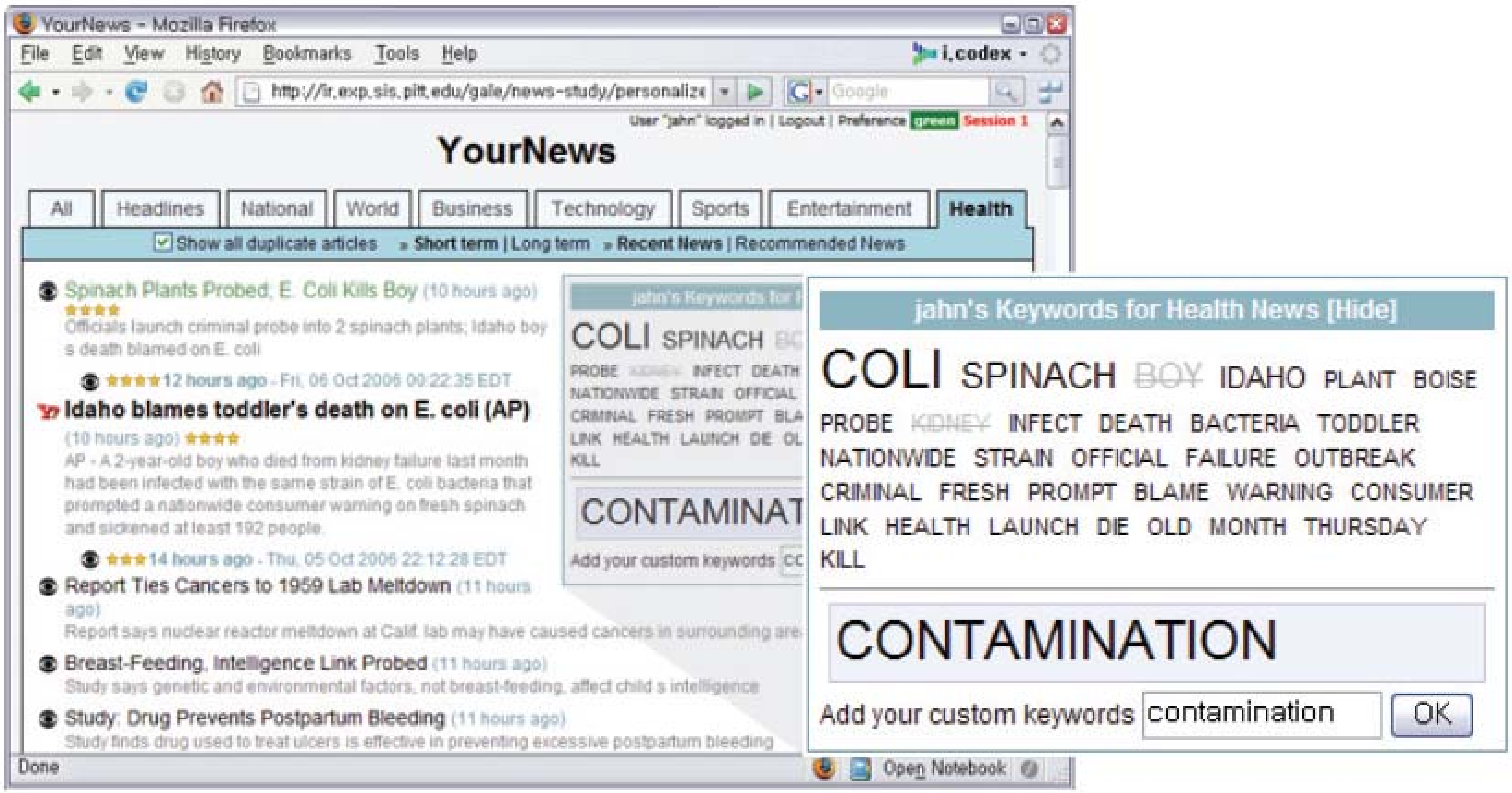
- Rich user model important for search personalization
- *Seesaw* improves text based retrieval
- Need other features to improve Web
- Lots of room for improvement



Evaluating Personalized Search

- Explicit judgments (offline and in situ)
  - Evaluate components before system
  - NOTE: What's relevant for you
- Deploy system
  - Verbatim feedback, Questionnaires, etc.
  - Measure behavioral interactions (e.g., click, reformulation, abandonment, etc.)
  - Click biases –order, presentation, etc.
  - Interleaving for unbiased clicks
- Link implicit and explicit (Curious Browser plugin)
- Beyond a single query -> sessions and beyond

User Control in Personalization (RF)



J-S. Ahn, P. Brusilovsky, D. He, and S.Y. Syn. Open user profiles for adaptive news systems: Help or harm? WWW 2007

Study: Comparing Personalization Strategies

[Dou et al., 2007]

- 10,000 users, 56,000 queries, and 94,000 clicks over 12 days.
- Used the first 11 days' worth of data to form user profiles and clicks.
- Simulated the application of five different personalization algorithms on the remaining 4,600 queries from the last day of the log.
- Retrieved top 50 results for each query from the comparison search engine and assumed that clicking a link indicated a relevance judgment for the query

Results: Which Strategy is Most Effective?

[Dou et al., 2007]

- Compared two click-based (behavior) personalization strategies to three profile-based strategies
- Click-based strategies appear more effective than profile-based (but carefully combining historical profile data helps slightly)
- Search context crucial
- Personalization effectiveness varies by query

Evaluated using naïve click models

Lecture 4 Outline

- ✓ **Approaches to Search Personalization**

1. **Dimensions of Personalization**
   - ✓ **What input to use for personalization?**
   - ➢ **Which queries to personalize?**
   1. **Granularity: personalization vs. groupization**
   2. **Context: Geograpical**

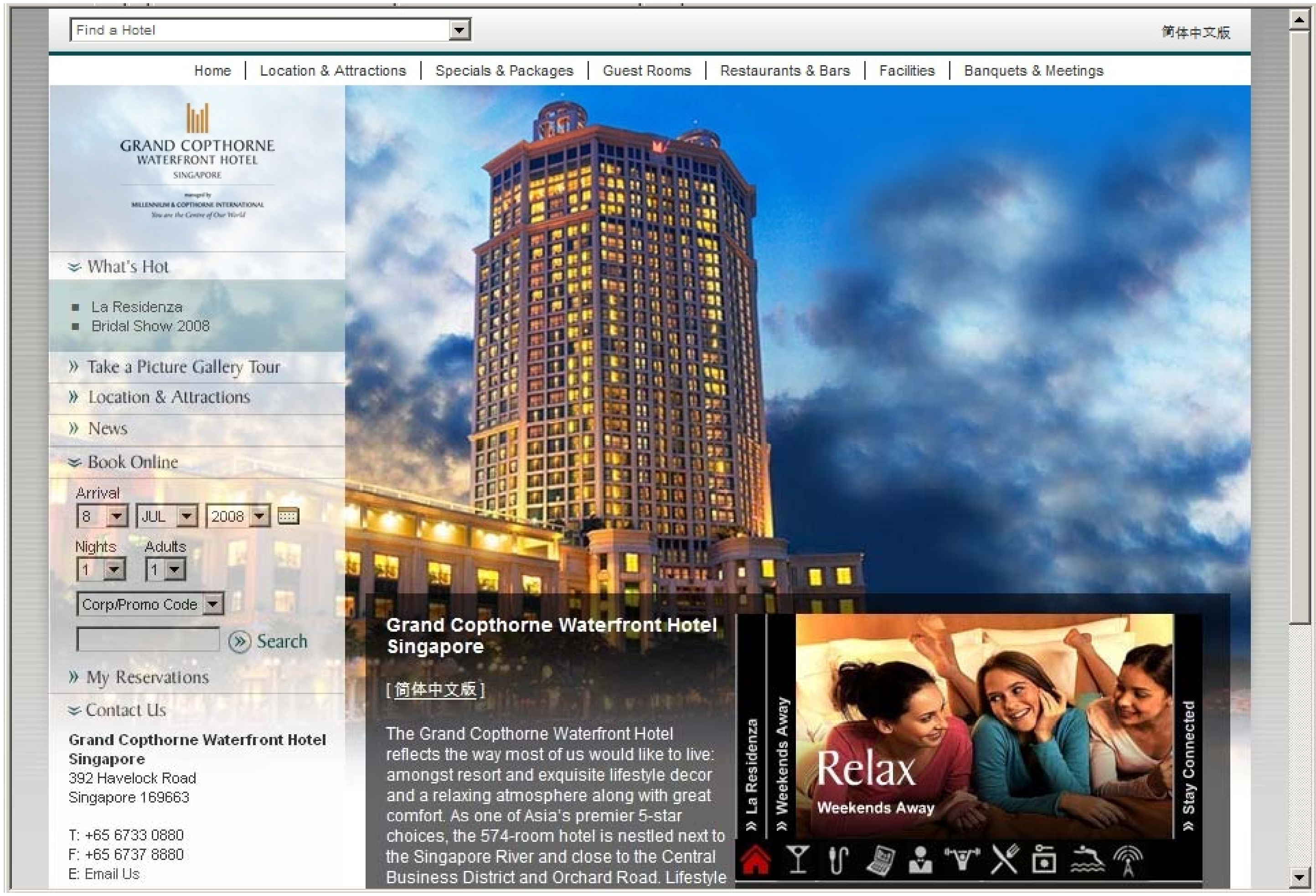
Understanding Query Ambiguity

SIGIR 2008

Jaime Teevan, Susan Dumais, Dan Liebling

Microsoft Research

"grand copthorne waterfront"

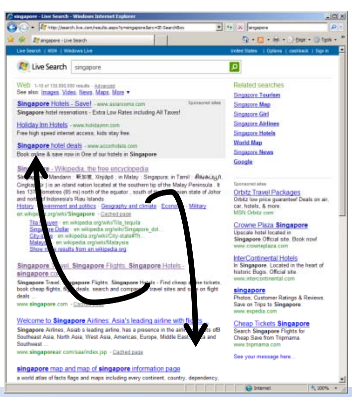


“singapore”

Travel Warnings | Travel Alerts | Country Specific Information
U.S. DEPARTMENT of STATE
TRAVEL.STATE.GOV
BUREAU OF CONSULAR AFFAIRS
Search
GO
Most Requested
Feedback | Contact Us
Tuesday July 8, 2008
INTERNATIONAL TRAVEL
PASSPORTS
VISAS
CHILDREN & FAMILY
LAW & POLICY
Home > International Travel Home > International Travel Information
International Travel Home
A-Z Index of Topics
New Requirements for U.S. Citizens
International Travel Information
Tips for Traveling Abroad
Document Requirements
Living Abroad Tips
Additional Resources
About Overseas Citizens' Services
Printer friendly version
Email
Singapore
Country Specific Information
Recent Embassy Notices for American Citizens
--------- Select a Country ---------
On this page »
Country Description
Entry/Exit Requirements
Safety and Security
Crime
Information for Victims of Crime
Medical Facilities and Health Information
Medical Insurance
Traffic Safety and Road Conditions
Aviation Safety Oversight
Special Circumstances
Criminal Penalties
Children's Issues
Registration/Embassy Location
REGISTER YOUR TRIP ONLINE
Travel Information
Current Travel Warnings
Current Travel Alerts
Country Specific Information
May 12, 2008
COUNTRY DESCRIPTION: Singapore is a small, stable, highly developed country with an elected parliamentary system of government. Tourist facilities are modern and widely available. Singapore's resident population of over 4.6 million inhabitants (including permanent residents and foreign workers) comprises 75% Chinese, 14% Malay, 9% Indian
MALAYSIA
Singapore
Relax

How Do the Two Queries Differ?

- grand copthorne waterfront v. singapore
- Knowing query ambiguity allow us to:
  - Personalize or diversify when appropriate
  - Suggest more specific queries
  - Help people understand diverse result sets



Related searches

Singapore Tourism

Singapore Map

Singapore Girl

Singapore Airlines

Singapore Hotels

World Map

Singapore News

Understanding Ambiguity

- Look at measures of query ambiguity
  - Explicit
  - Implicit
- Explore challenges with the measures
  - Do implicit predict explicit?
  - Other factors that impact observed variation?
- Build a model to predict ambiguity
  - Using just the query string, or also the result set
  - Using query history, or not

Which Queries to Personalize?

[Teevan et al., 2008]

- Personalization benefits **ambiguous** queries
- Inter-rater reliability (Fleiss' kappa)
  - Observed agreement (P_a) exceeds expected (P_e)
  - $\kappa = (P_a - P_e) / (1 - P_e)$
- Relevance entropy
  - Variability in probability result is relevant (P_r)
  - $S = -\Sigma\, P_r \log P_r$
- Potential for personalization
  - Ideal group ranking differs from ideal personal
  - $P4P = 1 - nDCG_{group}$

Teevan, J, S. T. Dumais, and D. J. Liebling. *To personalize or not to personalize: modeling queries with variation in user intent.*, SIGIR 2008

Predicting Ambiguity

[Teevan et al., 2008]

		History	
		No	**Yes**
Information	**Query**	Query length Contains URL Contains advanced operator Time of day issued Number of results (df) Number of query suggests	Reformulation probability # of times query issued # of users who issued query Avg. time of day issued Avg. number of results Avg. number of query suggests
	Results	Query clarity ODP category entropy Number of ODP categories Portion of non-HTML results Portion of results from .com/.edu Number of distinct domains	Result entropy Avg. click position Avg. seconds to click Avg. clicks per user $\boxed{\text{Click entropy}}$ Potential for personalization

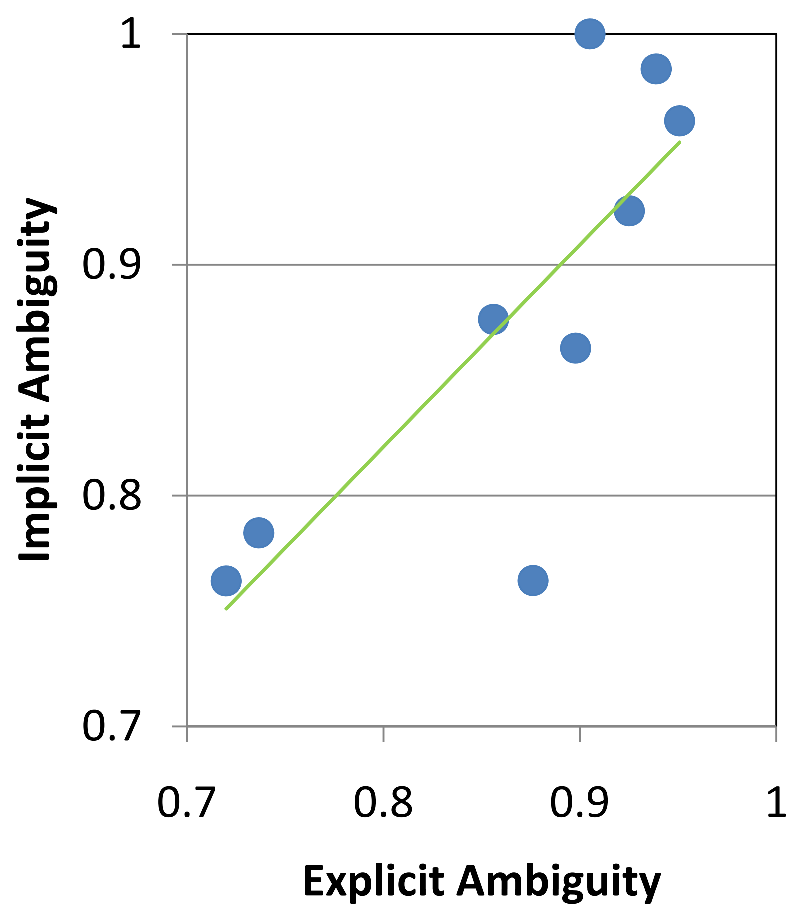
Teevan, J, S. T. Dumais, and D. J. Liebling. *To personalize or not to personalize: modeling queries with variation in user intent.*, SIGIR 2008

Collecting Implicit Relevance Data

- Variation in clicks
  - Proxy (click = *relevant*, not clicked = *irrelevant*)
  - Other implicit measures possible
  - Disadvantage: Can mean lots of things, biased
  - Advantage: Real tasks, real situations, lots of data
- 44k unique queries issued by 1.5M users
  - Minimum 10 users/query
- 2.5 million result sets "evaluated"

How Good are Implicit Measures?

- Explicit data is expensive
- Implicit good substitute?
- Compared queries with
  - Explicit judgments and
  - Implicit judgments
- Significantly correlated:
  - Correlation coefficient = 0.77 (p<.01)



Which Has Lower Click Entropy?

- www.usajobs.gov v. federal government jobs
- find phone number v. msn live search
- singapore pools v. singaporepools.com

Click entropy = 1.5

Result entropy = 5.7

Click entropy = 2.0

Result entropy = 10.7

Results change

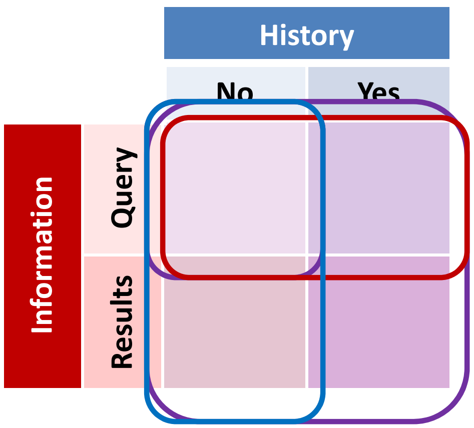
Challenges with Using Click Data

- Results change at different rates
- Result quality varies
- Task affects the number of clicks

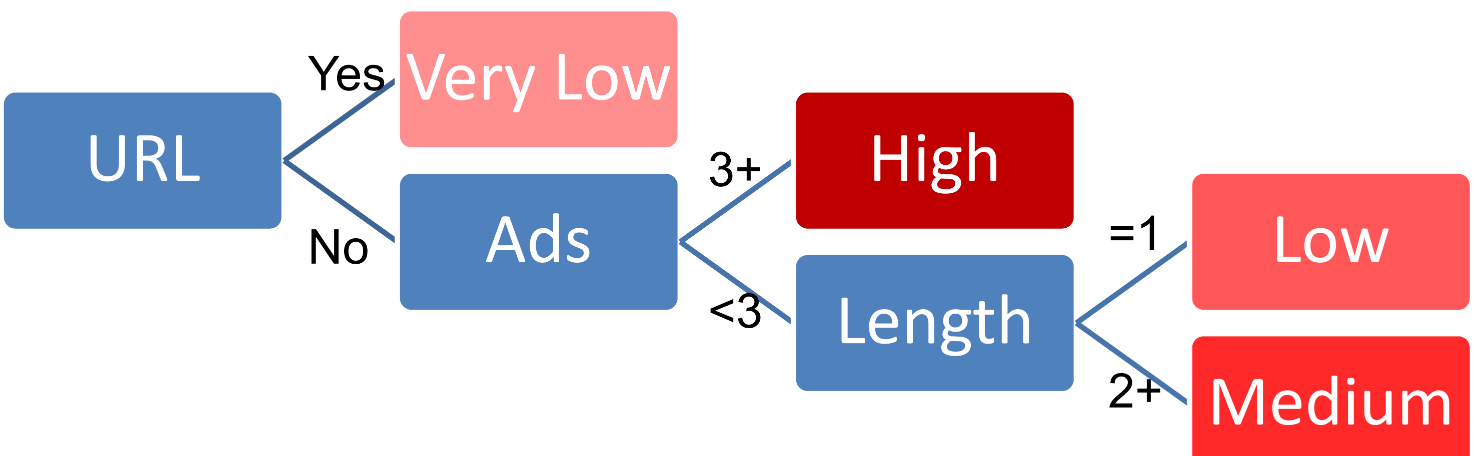
- We don't know click data for unseen queries

➢ Can we predict query ambiguity?

Result Summary [Teevan et al., 2008]



- All features = good prediction
  - 81% accuracy (↑ 220%)
- Just query features promising
  - 40% accuracy (↑ 57%)
- No boost adding results or history



Teevan, J, S. T. Dumais, and D. J. Liebling. *To personalize or not to personalize: modeling queries with variation in user intent.*, SIGIR 2008

Lecture 4 Outline

- ✓ **Approaches to Search Personalization**

1. **Dimensions of Personalization**
   - ✓ What input to use for personalization?
   - ✓ Which queries to personalize?
   - ➢ **Granularity: personalization vs. groupization**
   1. **Context: Geograpical, search session**

Connection: Collaborative Filtering and Recommender Systems

- Identify related groups
  - Browsed pages [Almeida & Almeida 2004; Sugiyama et al. 2005]
  - Queries [Freyne & Smyth 2006; Lee 2005]
  - Location [Mei & Church 2008], company [Smyth 2007], etc.
- Use group data to fill in missing personal data
  - Typically data based on user behavior

Discovering and Using Groups to Improve Personalized Search

Jaime Teevan, Merrie Morris, Steve Bush

Microsoft Research

WSDM 2009

[Slides from Teevan et al., WSDM 2009]

Diego Velasquez, *Las Lanzas*

[Slides from Teevan et al., WSDM 2009]

People Express Things Differently

- Differences can be a challenge for Web search
  - *Picture of a man handing over a key.*
  - *Oil painting of the surrender of Breda.*
- Personalization
  - Closes the gap using more about the person
- ***Group***ization
  - Closes the gap using more about the ***group***

[Slides from Teevan et al., WSDM 2009]

How to Take Advantage of Groups?

- Who do we share interests with?
- Do we talk about things similarly?
- What algorithms should we use?

[Slides from Teevan et al., WSDM 2009]

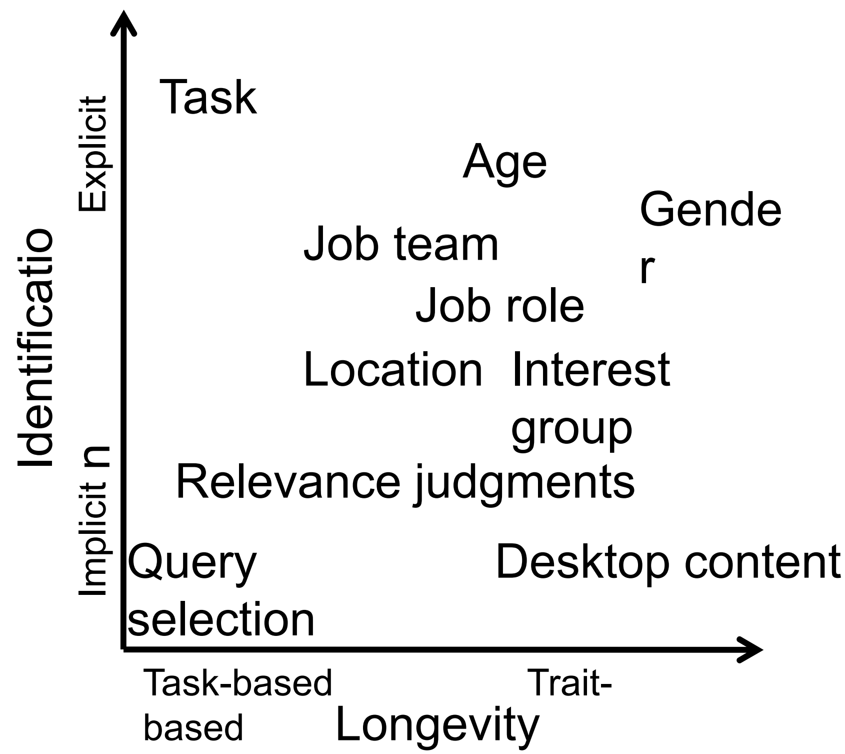
Approach

- Who do we share interests with?
  - Similarity in query selection
  - Similarity in what is considered relevant
- Do we talk about things similarly?
  - Similarity in user profile
- What algorithms should we use?
  - **Groupize** results using groups of user profiles
  - Evaluate using groups' relevance judgments

[Slides from Teevan et al., WSDM 2009]

Interested in Many Group Types

- Group longevity
  - Task-based
  - Trait-based
- Group identification
  - Explicit
  - Implicit



[Slides from Teevan et al., WSDM 2009]

Queries Studied

Trait-based dataset

- Challenge
  - Overlapping queries
  - Natural motivation
- Queries picked from 12
  - Work

    c# delegates, live meeting
  - Interests

    bread recipes, toilet train dog

Task-based dataset

- Common task
  - Telecommuting v. office

    pros and cons of working in an office

    social comparison telecommuting versus office

    telecommuting

    working at home cost benefit

Data Collected

- Queries evaluated
- Explicit relevance judgments
  - 20 - 40 results
  - Personal relevance
    - *Highly relevant*
    - *Relevant*
    - *Not relevant*
- User profile: Desktop index

[Slides from Teevan et al., WSDM 2009]

Answering the Questions

- Who do we share interests with?
- Do we talk about things similarly?
- What algorithms should we use?

[Slides from Teevan et al., WSDM 2009]

Who do we share interests with?

- Variation in query selection
  - Work groups selected similar work queries
  - Social groups selected similar social queries
- Variation in relevance judgments
  - Judgments varied greatly (κ=0.08)
  - Task-based groups most similar
  - Similar for one query ≠ similar for another

[Slides from Teevan et al., WSDM 2009]

Do we talk about things similarly?

- Group profile similarity
  - Members more similar to each other than others
  - Most similar for aspects related to the group

	In task group	Not in group	Difference
All queries	0.42	0.31	34%
Group queries	0.77	0.35	120%

- Clustering profiles recreates groups
- Index similarity ≠ judgment similarity
  - Correlation coefficient of 0.09

[Slides from Teevan et al., WSDM 2009]

What algorithms should we use?

- Calculate personalized score for each member
  - Content: User profile as relevance feedback

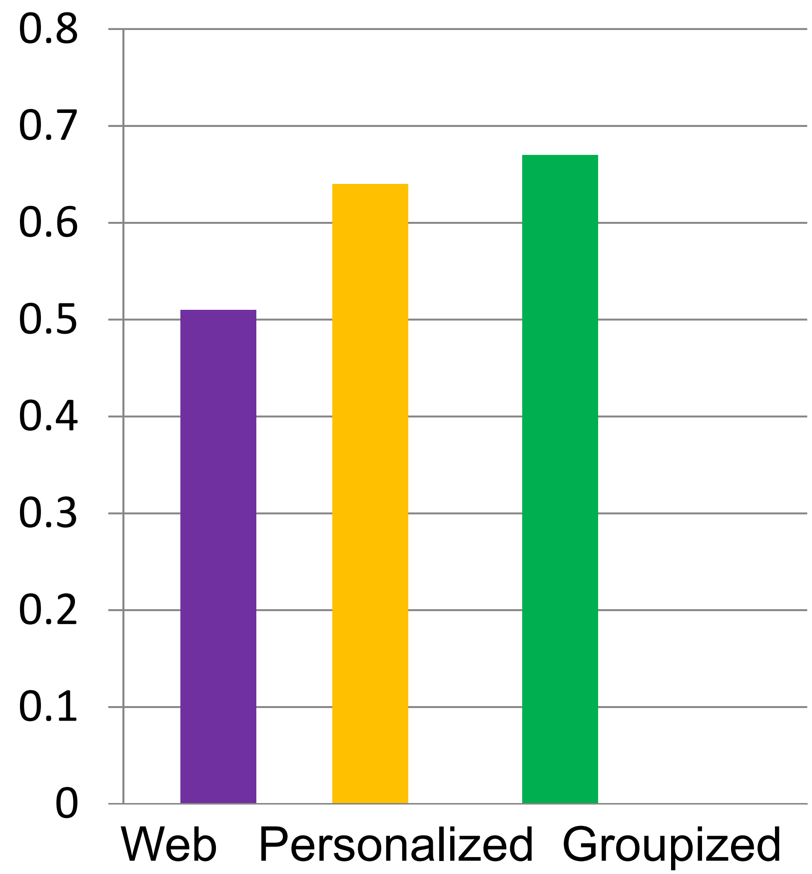
$$\sum_{\text{terms } i} tf_i \log \frac{(r_i+0.5)(N-n_i-R+r_i+0.5)}{(n_i-r_i+0.5)(R-r_i+0.5)}$$

  - Behavior: Previously visited URLs and domains
- Sum personalized scores across group
- Produces same ranking for all members

[Slides from Teevan et al., WSDM 2009]

Performance: Task-Based Groups

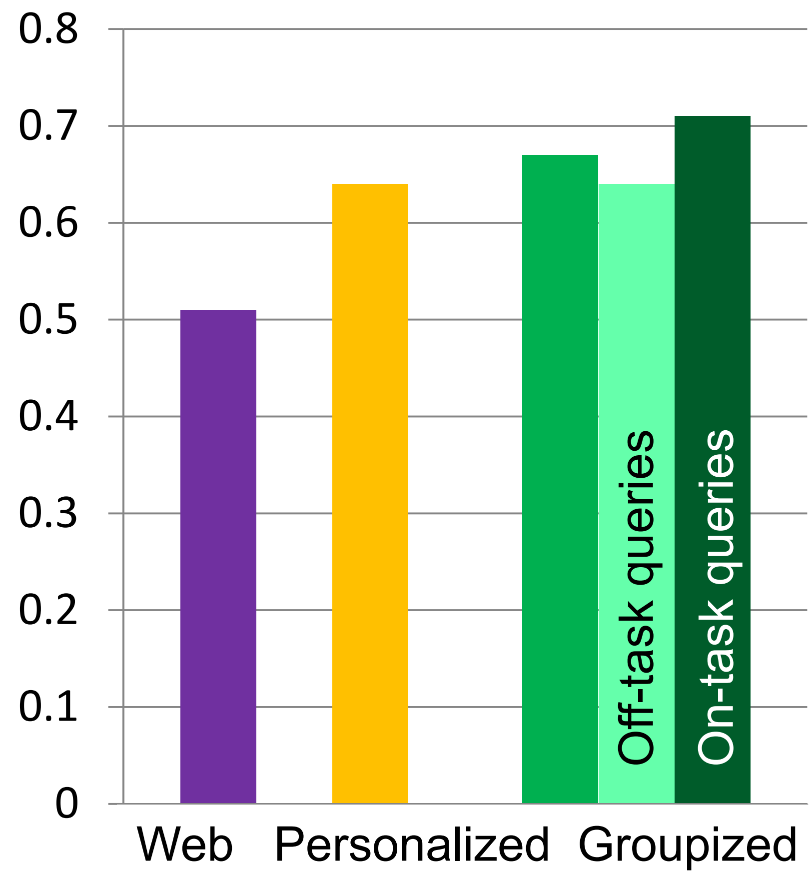
- Personalization improves on Web
- Groupization gains +5%



[Slides from Teevan et al., WSDM 2009]

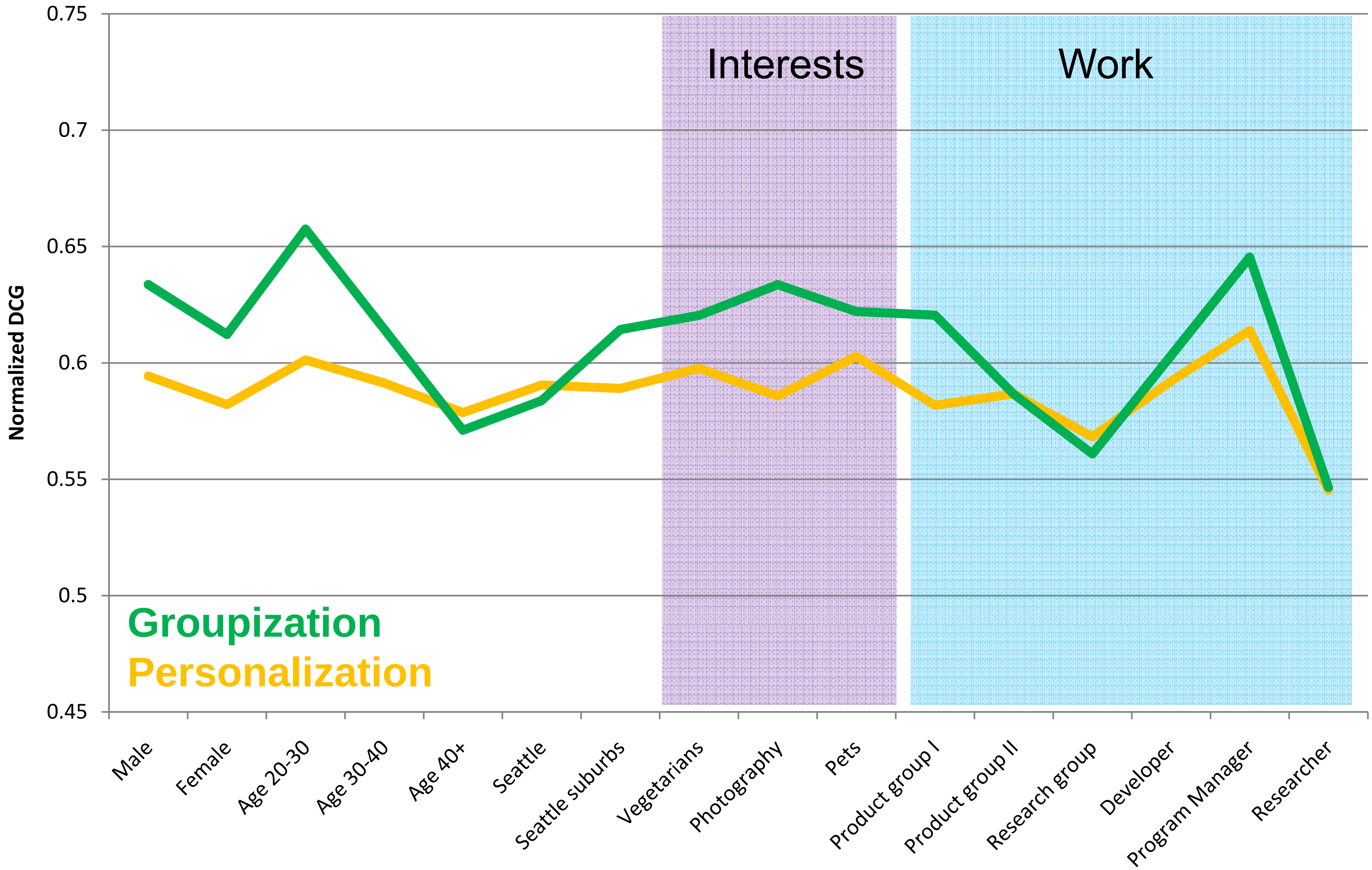
Performance: Task-Based Groups

- Personalization improves on Web
- Groupization gains +5%
- Split by query type
  - On-task v. off-task
  - Groupization the same as personalization for off-task queries
  - 11% improvement for on-task queries



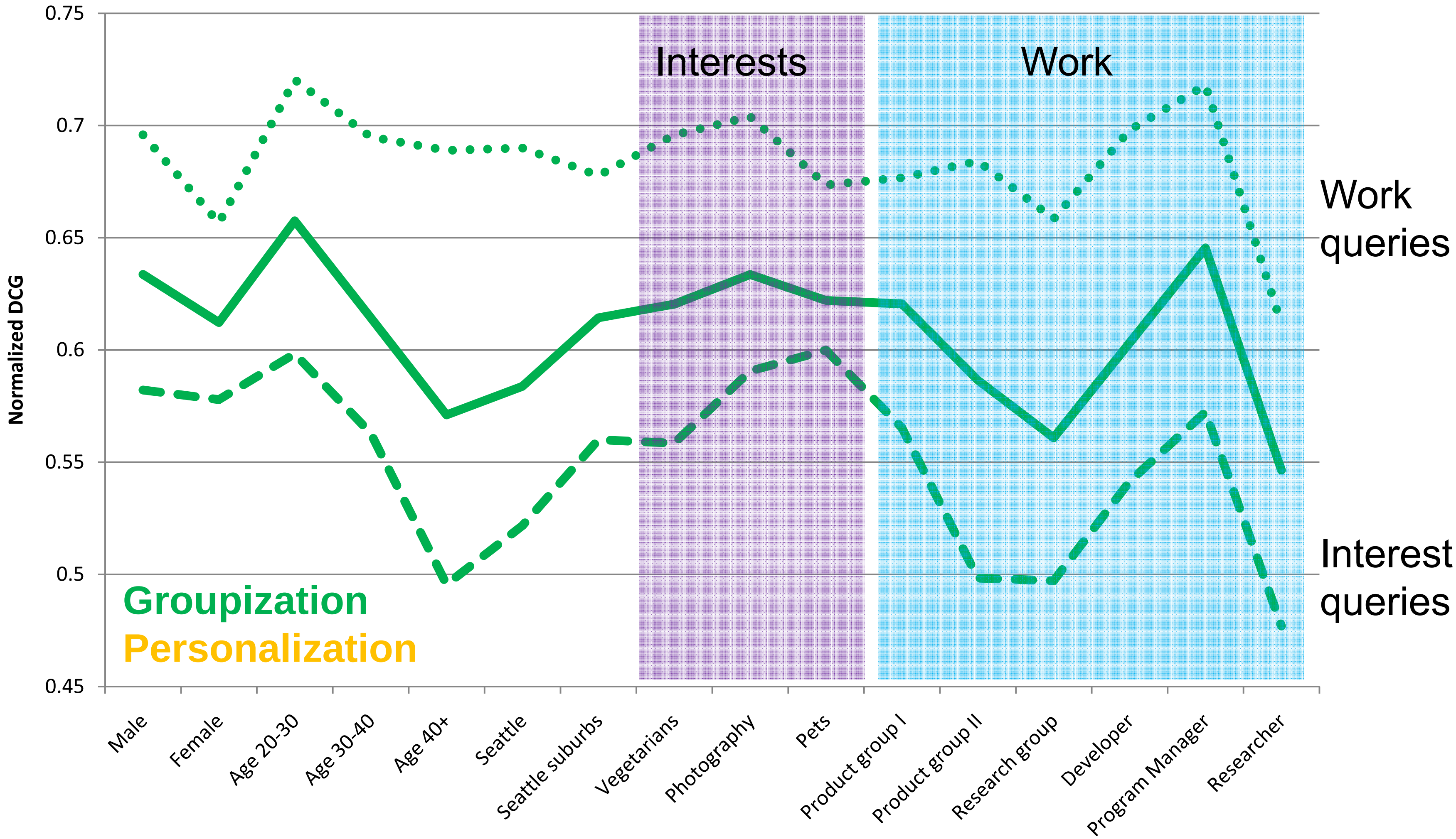
[Slides from Teevan et al., WSDM 2009]

Performance: Trait-Based Groups



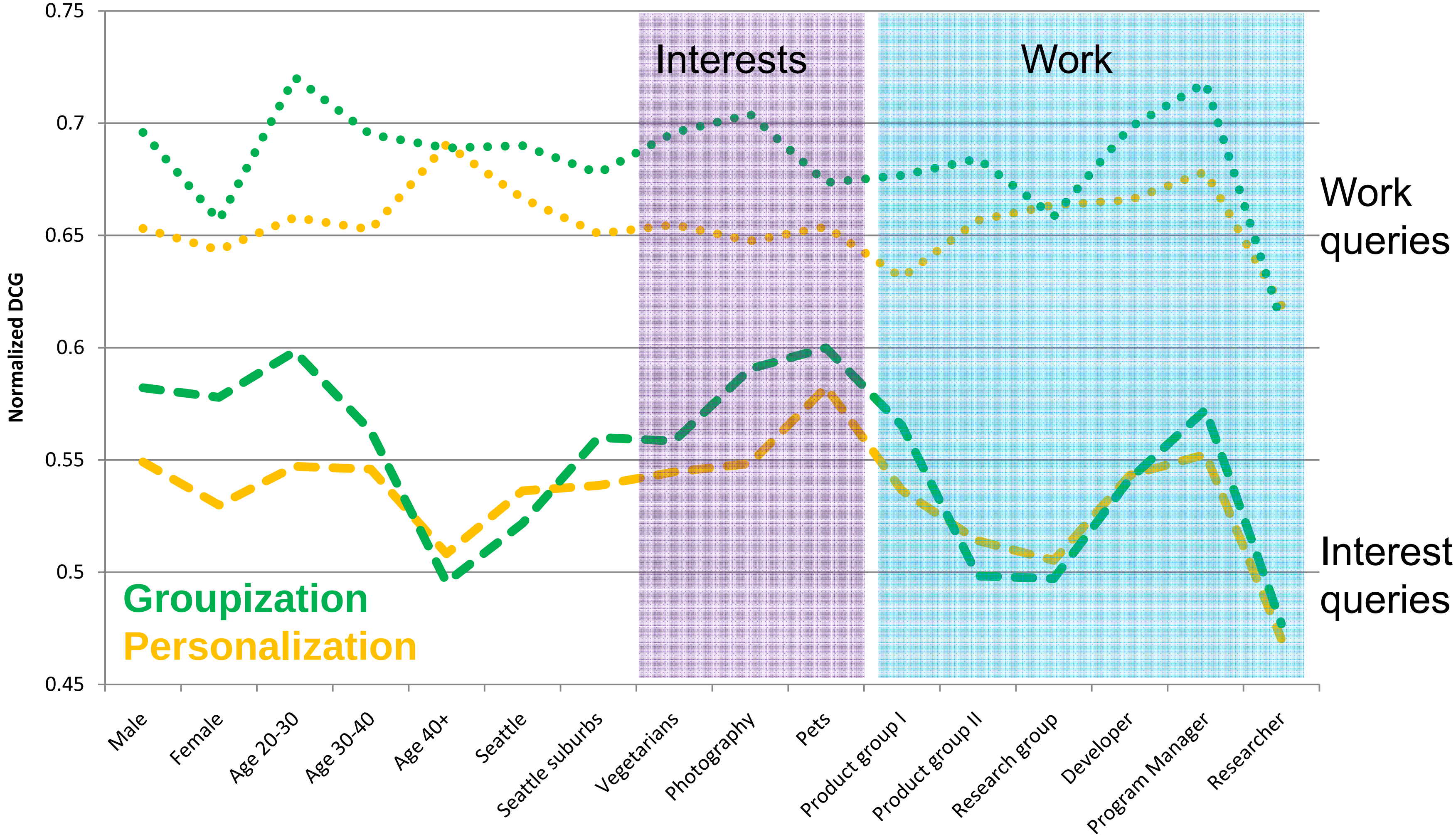
[Slides from Teevan et al., WSDM 2009]

Performance: Trait-Based Groups



[Slides from Teevan et al., WSDM 2009]

Performance: Trait-Based Groups



Lecture 4 Outline

✓ **Approaches to Search Personalization**

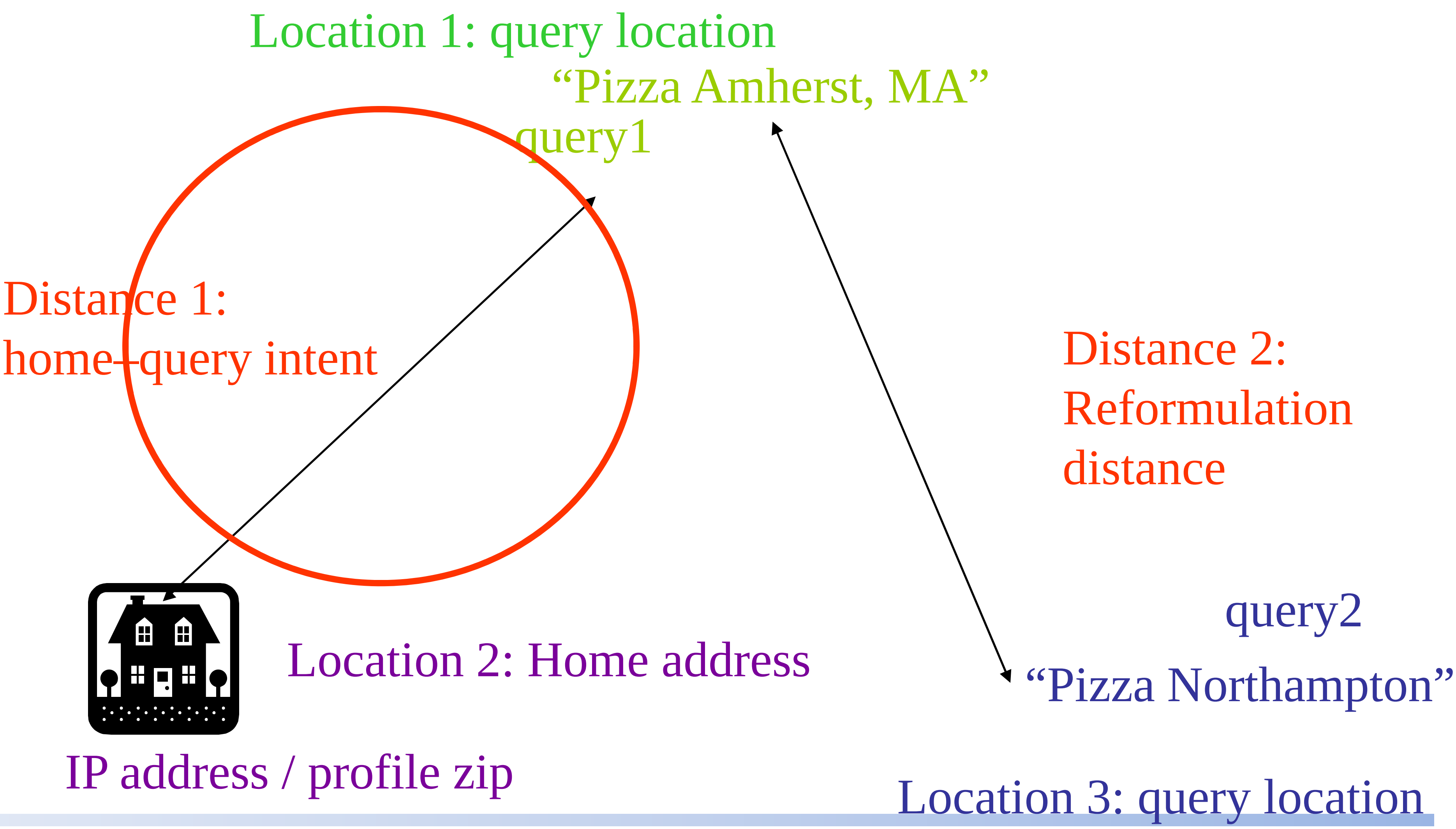
1. **Dimensions of Personalization**
   - ✓ What input to use for personalization?
   - ✓ Which queries to personalize?
   - ✓ Granularity: personalization vs. groupization
   - ➢ **Context: Geographical, search session**

Local Search (Geographical Personalization)

- Location is context
- *Local search* uses geographic information to modify the ranking of search results
  - location derived from the query text
  - location of the device where the query originated
- e.g.,
  - "underworld 3 cape cod"
  - "underworld 3" from mobile device in Hyannis

Geography and Query Intent

[Baeza-Yates and Jones] 2008

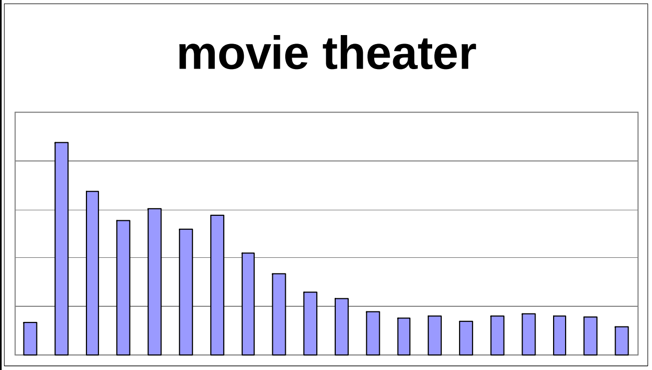


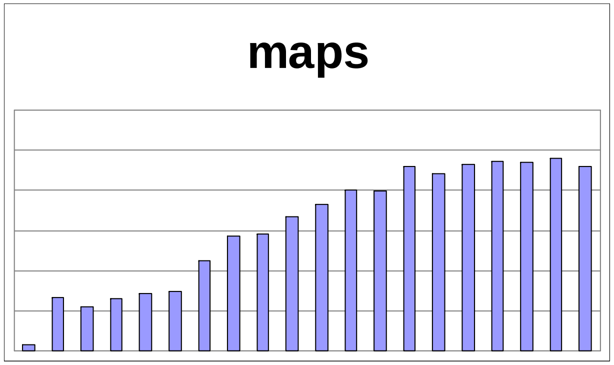
Topic-Distance Profiles

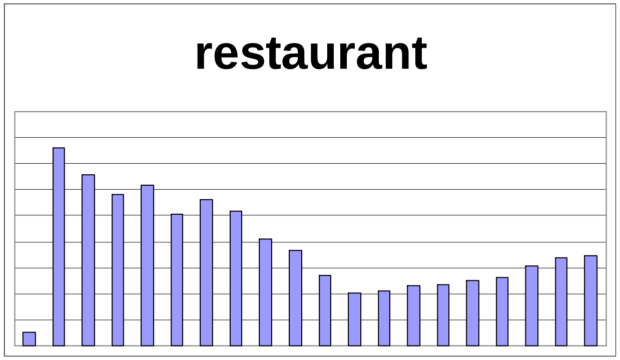
[Baeza-Yates and Jones] 2008

- 20 bins
  - 0 distance
  - Equal fractions of the rest of the data
- Does distribution into distance bins topics vary by topic?

Movie theater Distant places Near-by







Lecture 4 Outline

✓ **Approaches to Search Personalization**

1. **Dimensions of Personalization**
   - ✓ What input to use for personalization?
   - ✓ Which queries to personalize?
   - ✓ Granularity: personalization vs. groupization
   - ✓ Context: Geographical, search session

Key References and Further Reading

➢ Marti Hearst, ***Search User Interfaces***, 2009, Chapter 9: "Personalization in Search", Cambridge University Press, http://searchuserinterfaces.com/

Pitkow, J., Schütze, H., Cass, T., Cooley, R., Turnbull, D., Edmonds, A., Adar, E., and Breuel, T. Personalized search. Communications of ACM, 2002

Teevan, J., Dumais, S. T., and Horvitz, E. 2005. *Personalizing search via automated analysis of interests and activities*. , in Proc. of SIGIR 2005

Dou, Z., Song, R., and Wen, J. *A large-scale evaluation and analysis of personalized search strategies*, in Proc. of WWW 2007

Das, A. S., Datar, M., Garg, A., and Rajaram, S. *Google news personalization: scalable online collaborative filtering*. In Proc. of WWW 2007

Qiu, F and J. Cho. *Automatic Identification Of User Interest For Personalized Search.*, in Proc. of WWW 2006

Teevan, J, S. T. Dumais, and D. J. Liebling. *To personalize or not to personalize: modeling queries with variation in user intent*., in Proc. of SIGIR 2008

Teevan, J, Morris M, and Bush, S. *Discovering and Using Groups to Improve Personalized Search.* WSDM 2009