Search Tasks, Proactive Search & Digital Assistants

Rishabh Mehrotra

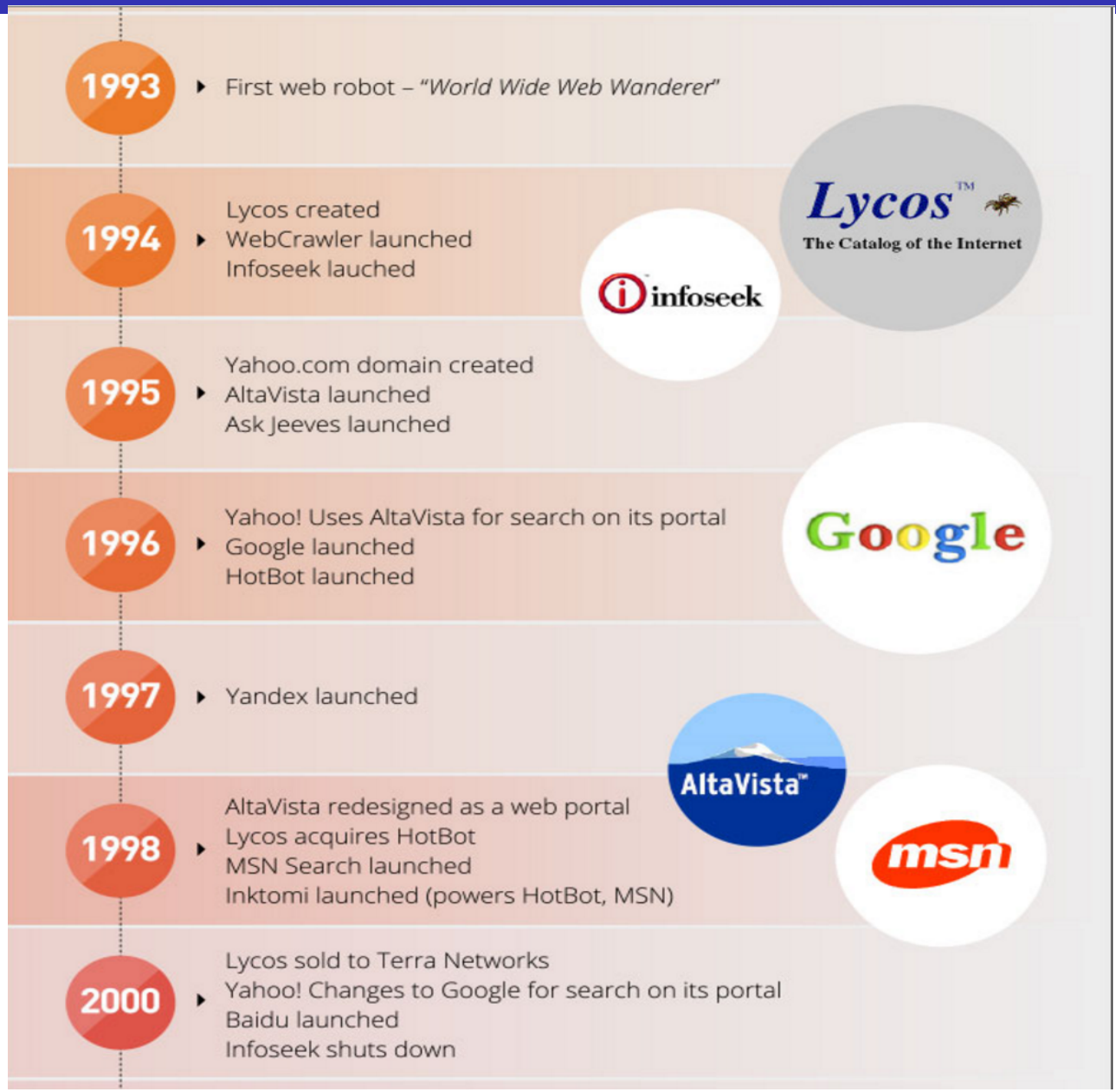
University College London
r.mehrotra@cs.ucl.ac.uk

Machine Learning Meetup, Gurgaon

Saturday 25th June, 2016

Modelling Search Tasks & Behaviors

Evolution of Web Search



Search is Everywhere

Web Search Volume

3 BILLION searches per day

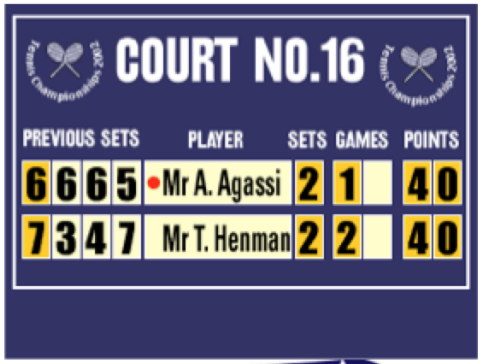
30 TRILLION unique URLs

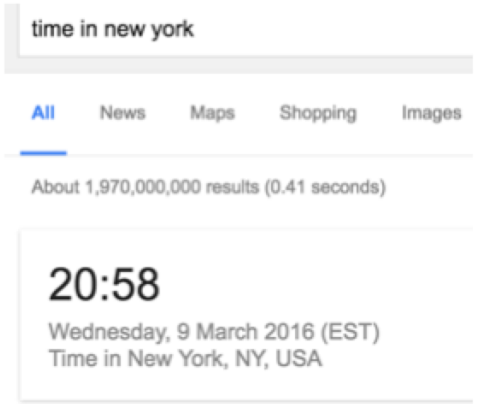
20 BILLION sites crawled per day

60 TRILLION web addresses

Towards Task-based Search Systems

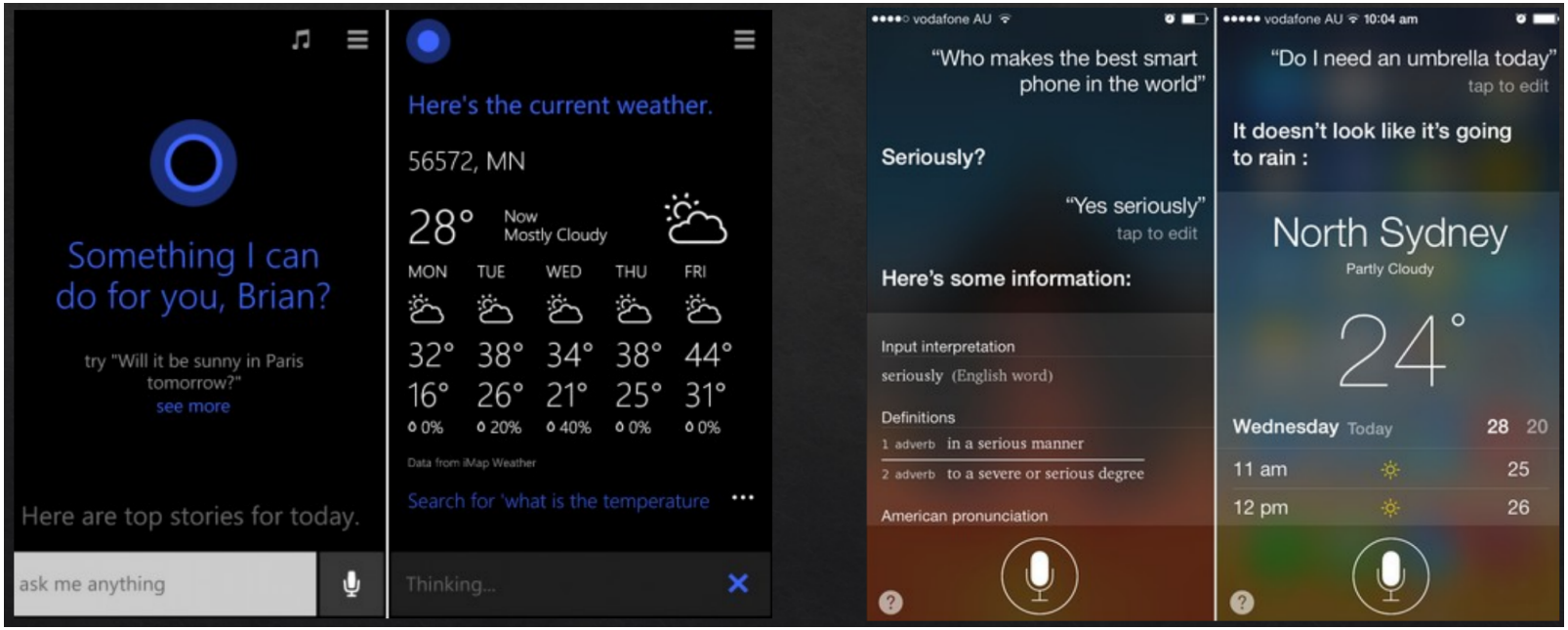
SIMPLE



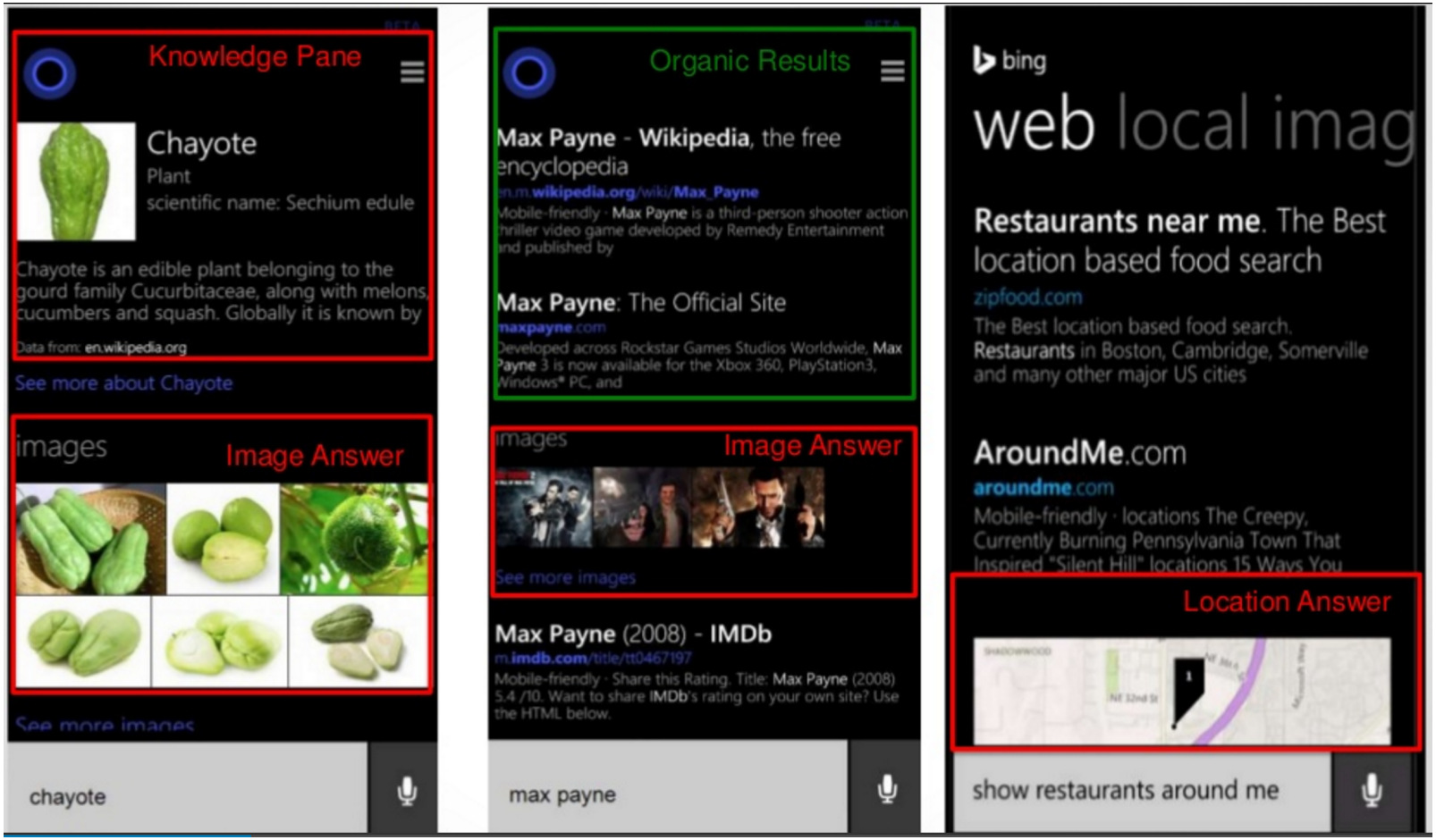


COMPLEX

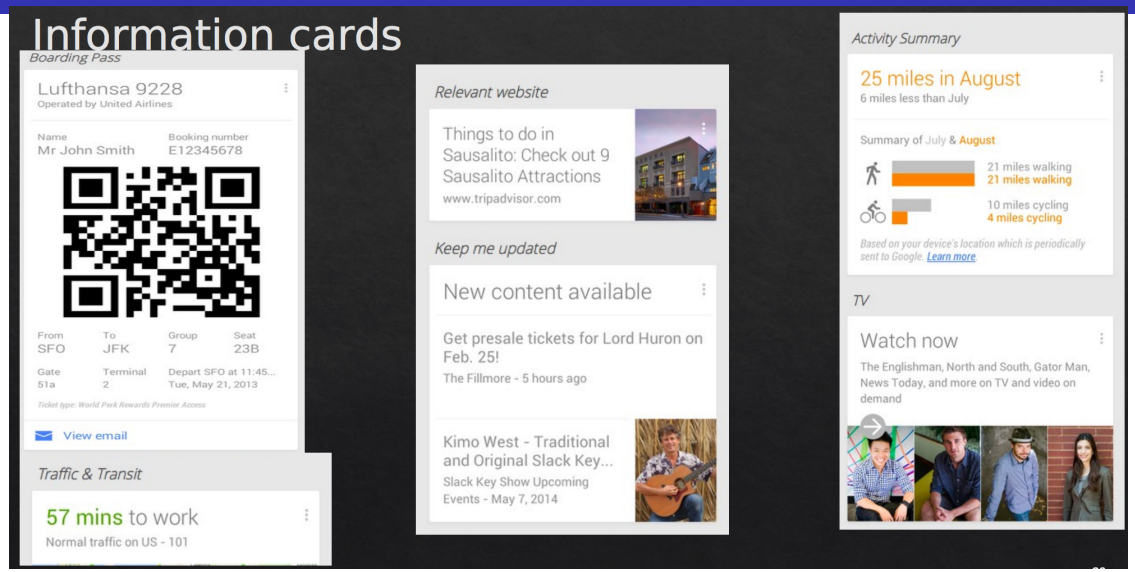
Slight Detour - Proactive IR & Digital Assistants



Slight Detour - Proactive IR & Digital Assistants



Slight Detour - Proactive IR & Digital Assistants



Information Cards:

- From Queries to Cards (Shokouhi et al. SIGIR'15)
- Modelling User Interests for Zero-query Ranking (Yang et al. ECIR'16)

Intelligent Assistants:

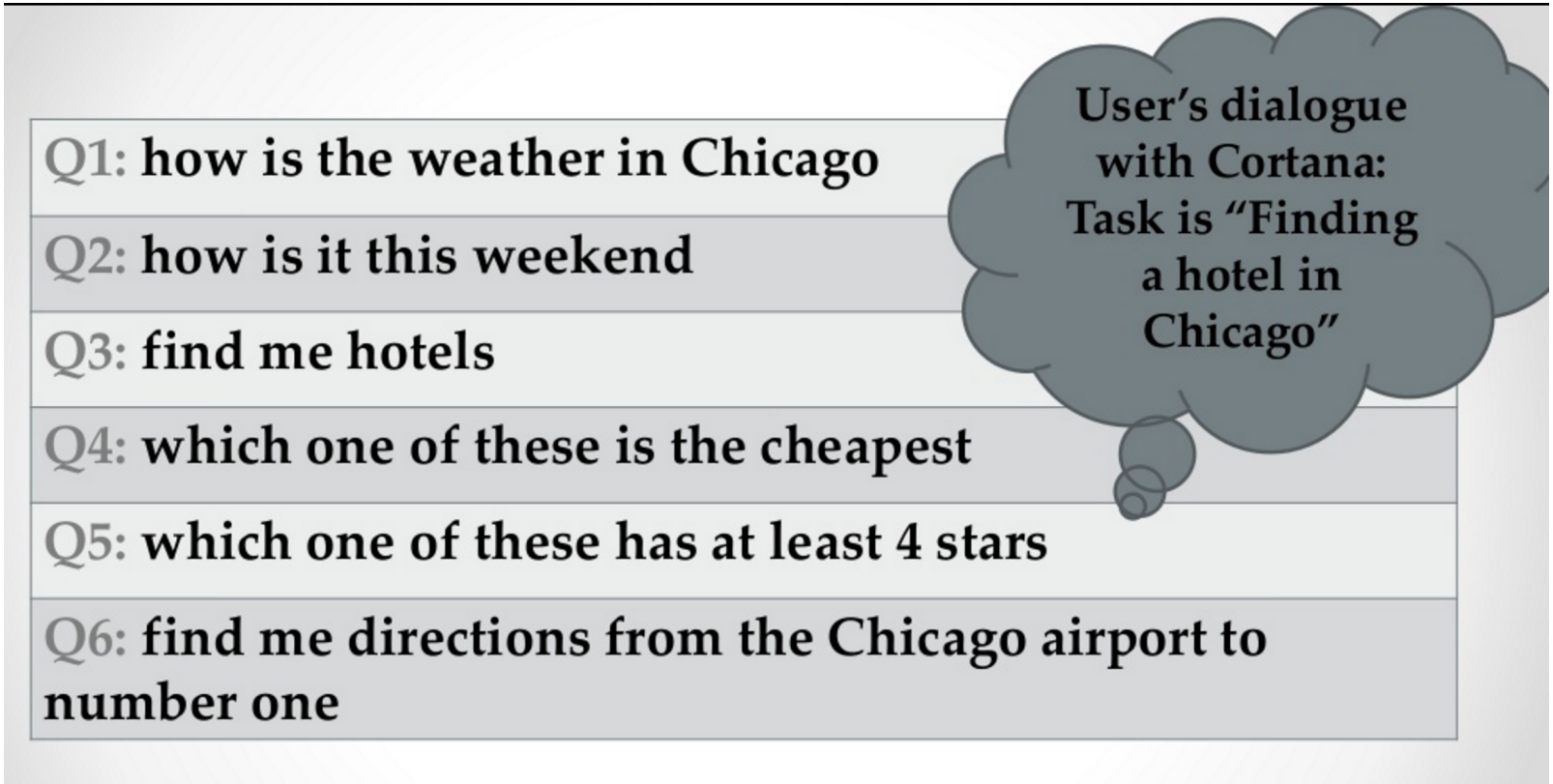
- Understanding User Satisfaction with Intelligent Assistants (Kiseleva et al. CHIIR'16)
- Automatic Online Evaluation of Intelligent Assistants (Jiang et al. WWW'15)

Slight Detour - Proactive IR & Digital Assistants

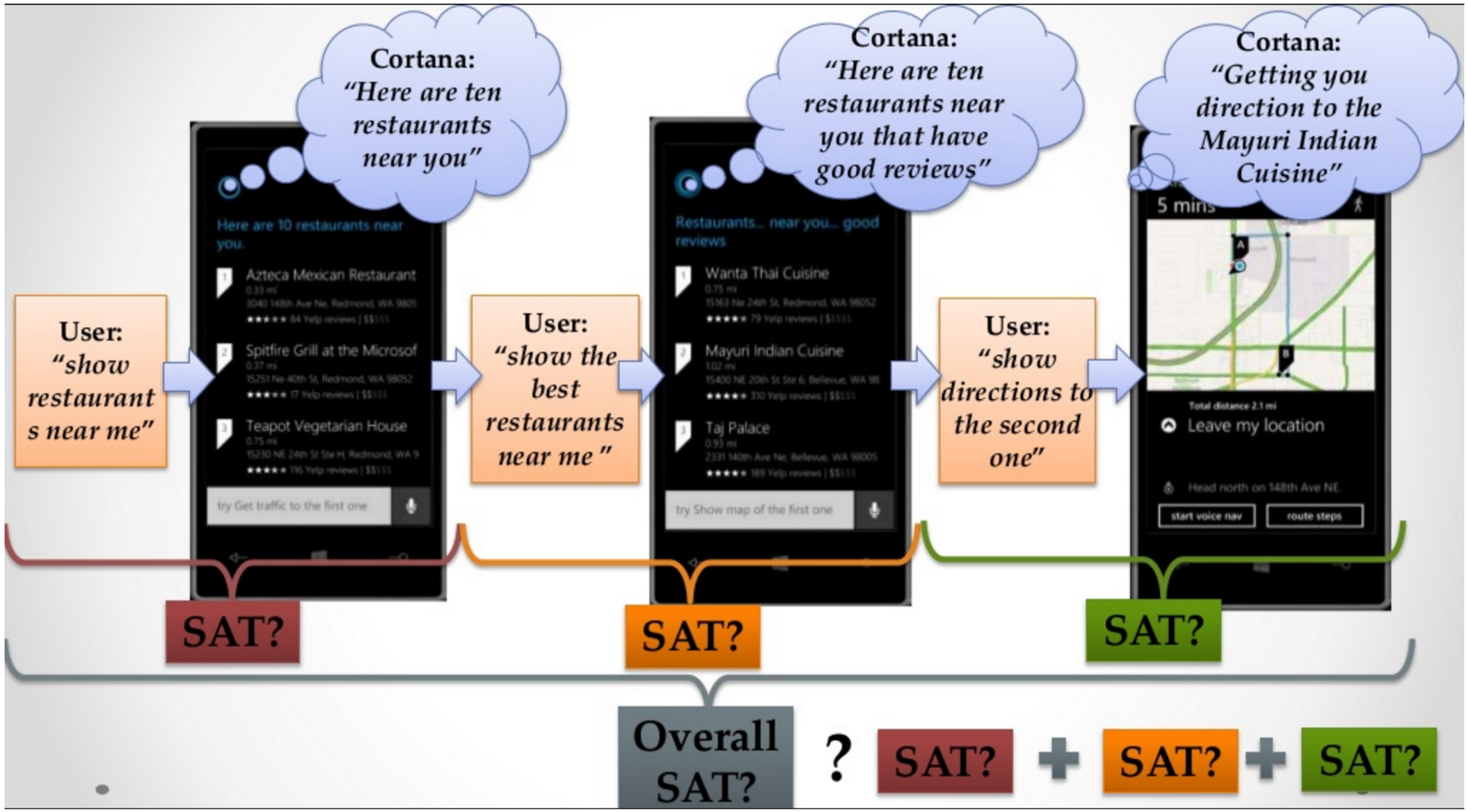
Modelling user interaction on Information Cards:

- Click-based (pseudo) relevance labels may not be appropriate for evaluating all types of cards.
- Features
  - Reactive History
  - Proactive History
  - Lexical/Topical Features
  - Local/Temporal Features
  - Constant Features

Slight Detour - Proactive IR & Digital Assistants



Slight Detour - Proactive IR & Digital Assistants[1]



[1] Understanding User Satisfaction with Intelligent Assistants, Kiseleva et al. ACM CHIIR 2016

Putting it together

- Digital Assistants - entry points for all web activities
- Goal is to move towards **Task Completion**

Understanding Search Tasks

Search Tasks

An (atomic?) information need that may result in issuing of one or more queries.

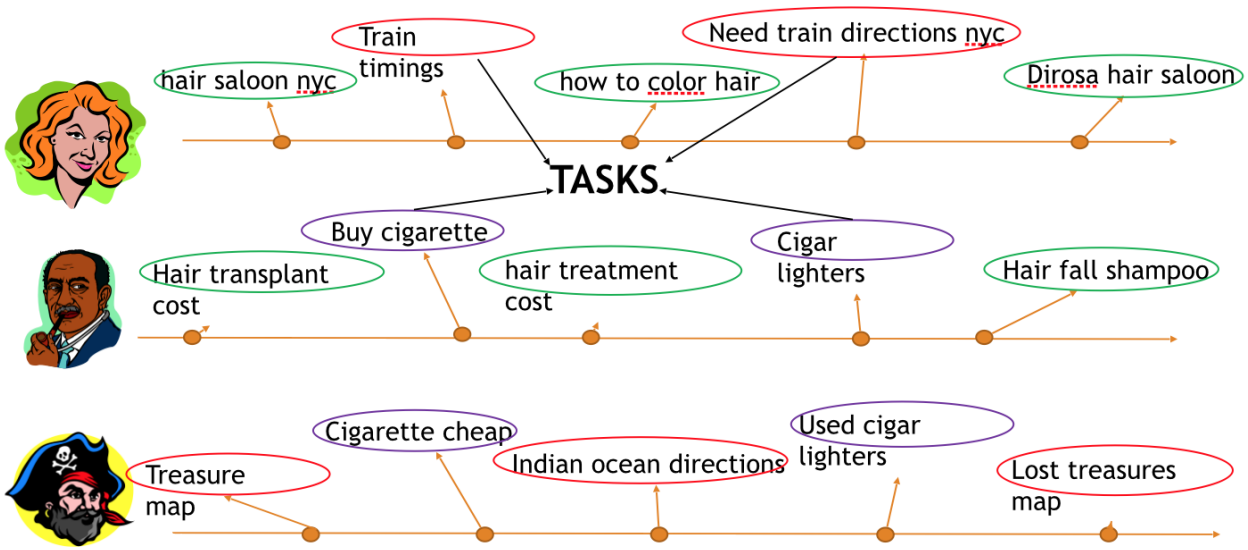
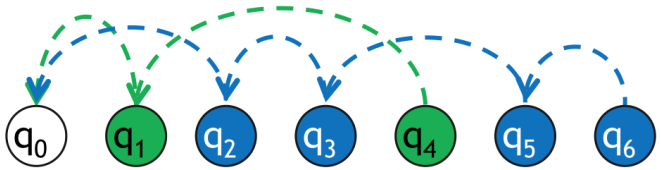


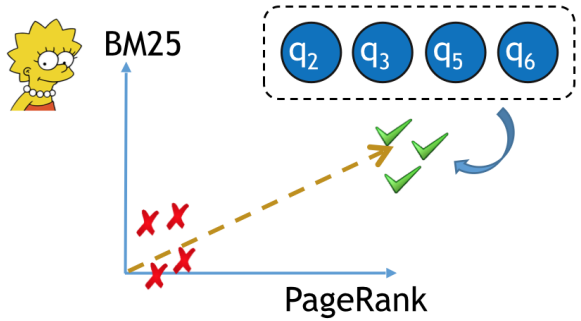
Image from Verma *et al.* 2014

Why Search Tasks

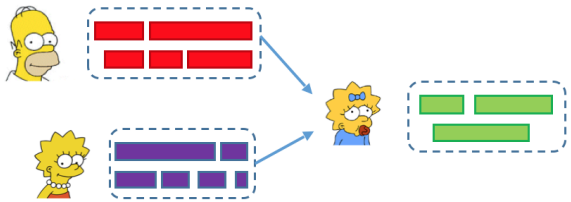
- Long-term task extraction



- In-task Personalization



- Cross-user collaborative ranking



- Search-task satisfaction prediction

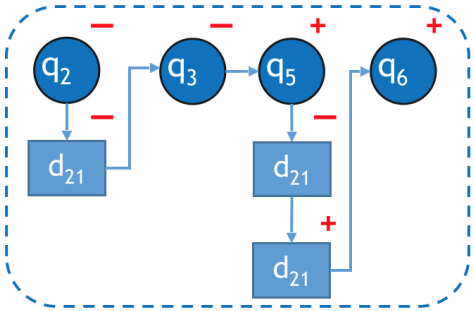


Image from Wang *et al.* 2014

Outline

Understanding Search Behavior

IR research = sessions, topics, queries

We focus on:

1. Multitasking
2. User groups based on multitasking
3. Topical heterogeneity

Research Questions

1. **RQ1**: *Extent of multi-tasking in search sessions?*
2. **RQ2**: *Evidence of user-level heterogeneity based on task behavior?*
    1. **RQ2.1**: *Does search task effort vary across user groups?*
    2. **RQ2.2**: *Association between users' interests and task multiplicity?*
3. **RQ3**: *Characterizing various heterogeneities in search behavior.*

Data Context

No of Queries	620M
No of Sessions	190M
No of Users	2M
Avg No of queries per session	3.18
Avg no of sessions per user	76.01
Avg no of tasks per session	2.08

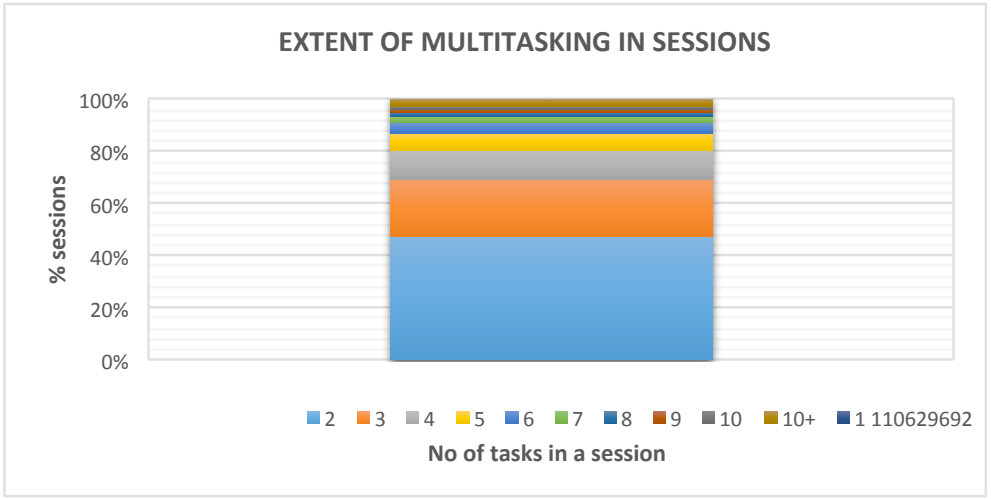
Table: Data summary

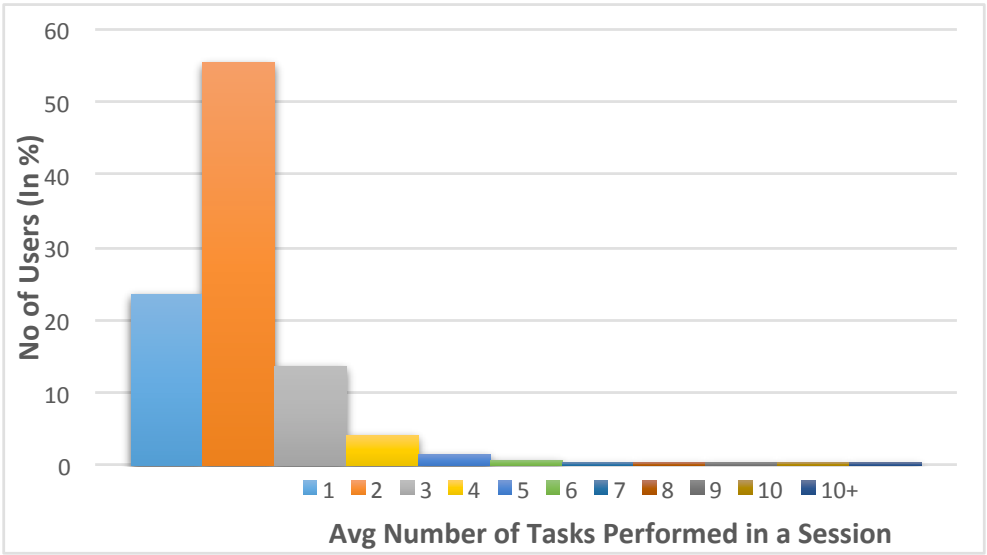
Time	Query	SessionID	TaskID	Topic
05/29/2012 14:06:04	adele songs	1	1	Arts
05/29/2012 14:11:49	wedding venue	1	2	Society
05/29/2012 14:12:01	video download	1	3	Arts
05/29/2012 14:06:04	Obama care	2	4	News
05/29/2012 14:11:49	running shoes	2	5	Shopping
05/29/2012 14:12:01	sports shoes	2	5	Shopping
05/29/2012 14:22:12	wedding cards	2	2	Society

Table: Sample search sessions

Extent of Multitasking

Quantifying the extent of multi-tasking





User Groups

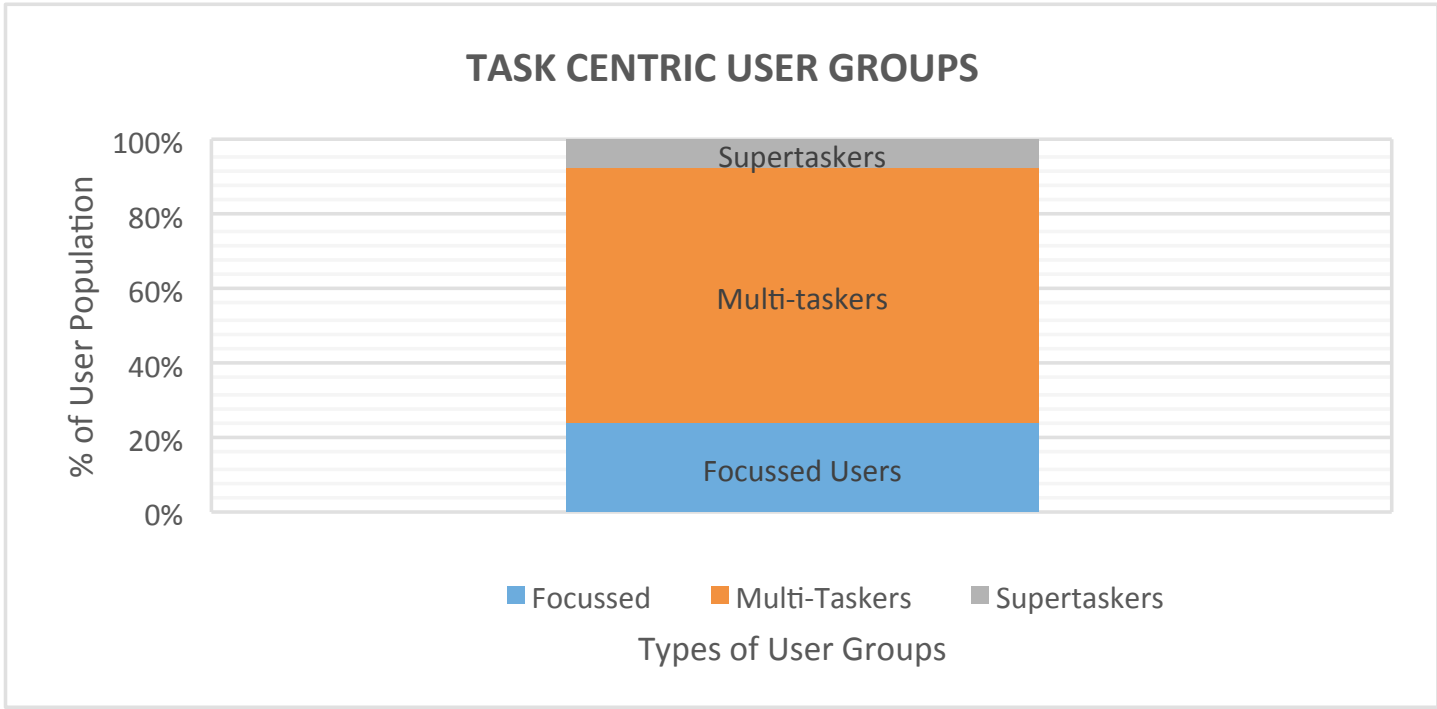


Figure: User groups based on multi-tasking behaviors

User Groups: Example Sessions

User Type	Example Queries from a Typical Session
Focussed User	"test guide.com", "CNA Practice Test", "CNA State Board Exam", "CNA Testing Schedule and Locations", "CNA State Board Practice Test", "CNA Practice Test 2014", CNA 50 Questions Test", "Free GED Practice Test 2014"
Multi-Tasking User	"Gravity FSX 2.0", "Full Suspension Mountain Bikes", "Walmart Cards", "Walmart Instant Card Application", "Gravity FSX 2.0 price", "full suspension bike sale"
Supertasking User	"hairstyles for women over 50", "thin wavy hairstyles for women", "facebook", "fb sign in", pulled pork crock pot recipe easy", "Slow-Roasted Pulled Pork", "barefoot contessa", "miley cyrus hair styles", "hairstyler.com"

Table: Example query sessions from the different user groups.

Effort Metrics across User Groups

TTFC: Time to First Click
TTLC: Time to Last Click
PCC: Page Click Count
PgCC: Pagination Click Count

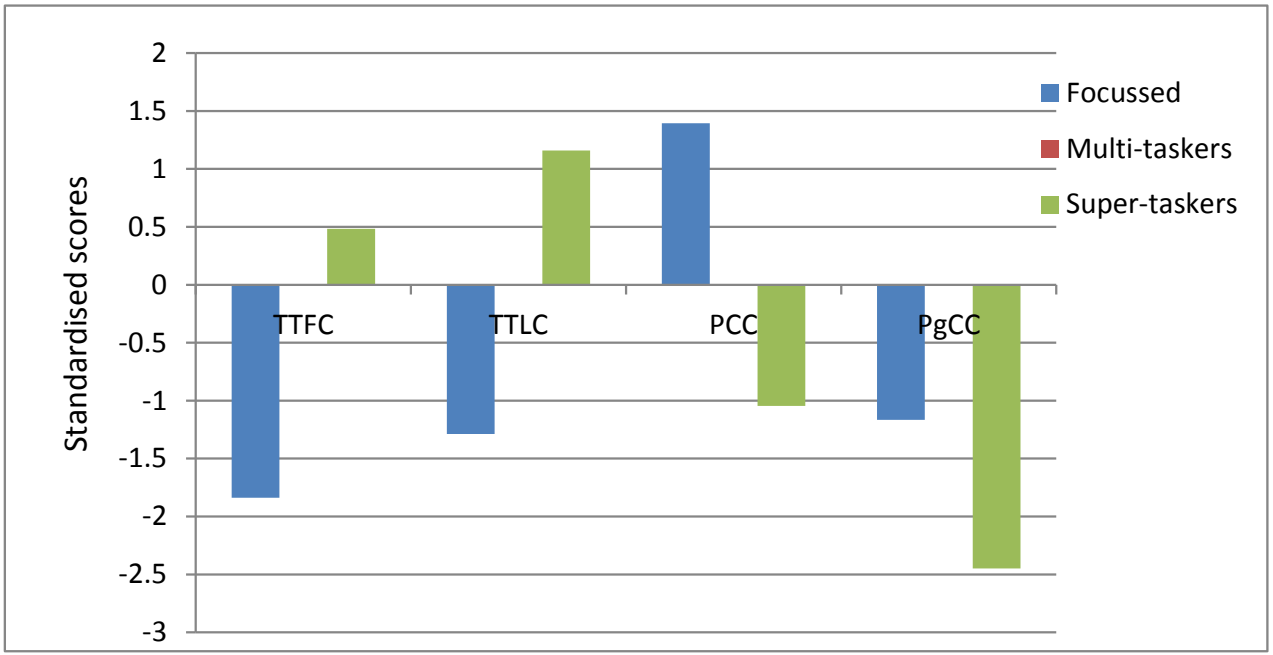
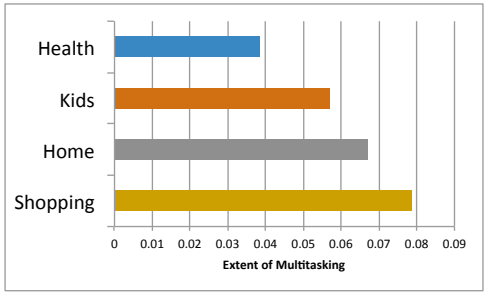
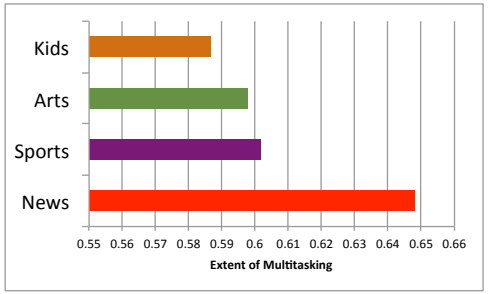


Figure: User groups based on multi-tasking behaviors

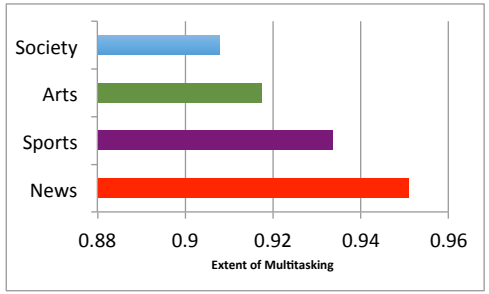
Topical Heterogeneity



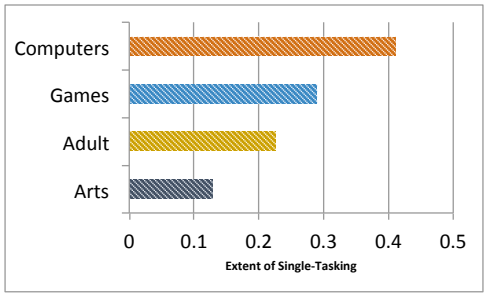
(a) Focused



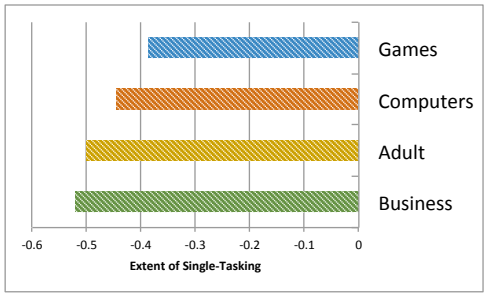
(b) Multitasker



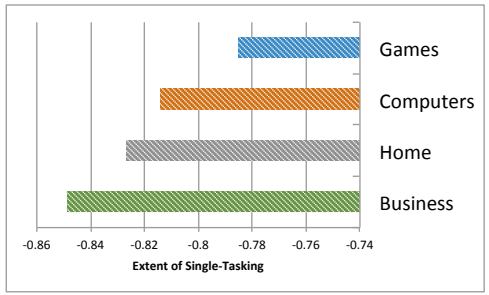
(c) Supertasker



(d) Focused



(e) Multitasker



(f) Supertasker

Figure: Top topics prone to multi-tasking (Top) and single-tasking (Bottom) across different user groups.

Takeaways[2]

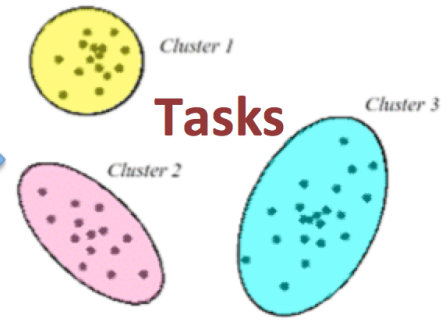
- Sessions should NOT be the focal units in IR
- More than 70% of sessions are multi-task
- Users differ based on their multi-tasking habits
- Search effort varies across user groups
- Some topics prone to multi-tasking than others

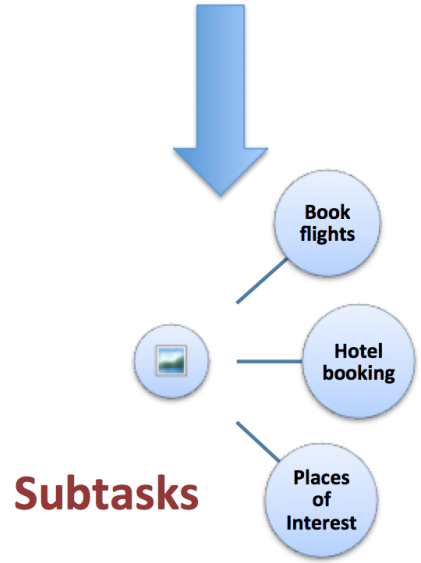
[2] R. Mehrotra, P. Bhattacharya, E. Yilmaz; **Characterizing Users Multi-Tasking Behavior in Web Search**, at ACM SIGIR Conference CHIIR 2016

Outline

Extracting Tasks & Subtasks

1. **Task extraction**
2. Subtask extraction

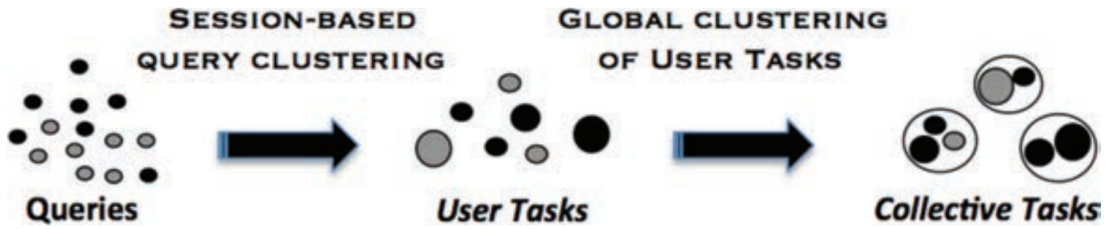




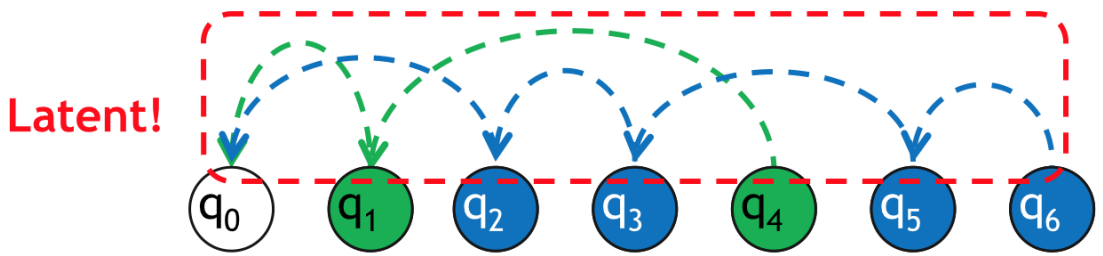
Extracting Search Tasks

Various proposed strategies:

- Clustering session based queries [Lucchese, WSDM'11]



- Structured Learning Approach [Wang, WWW'13]

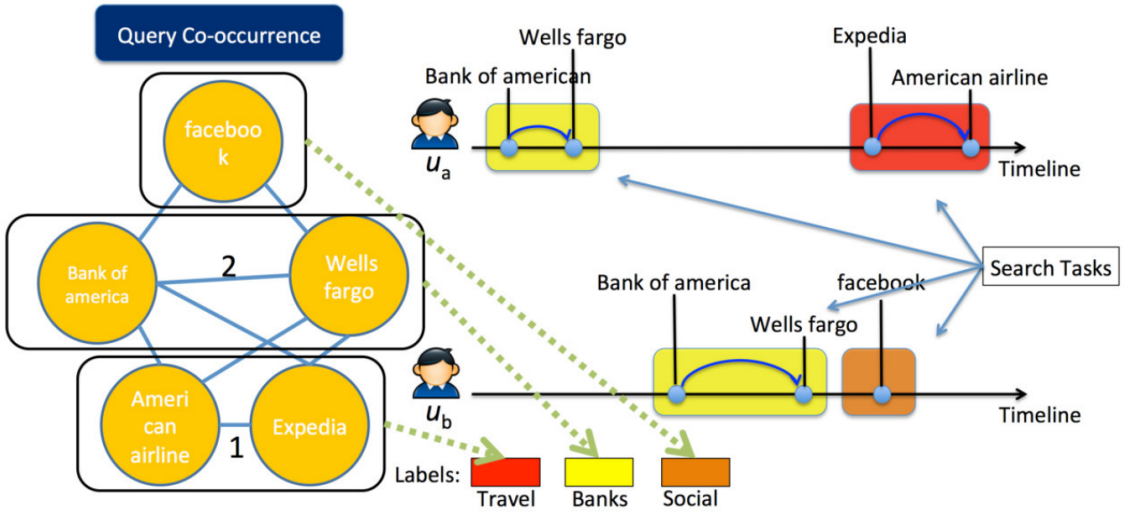


$$\mathcal{T}_1 = \{q_1, q_4\}\ \mathcal{T}_2 = \{q_2, q_3, q_5, q_6\}$$

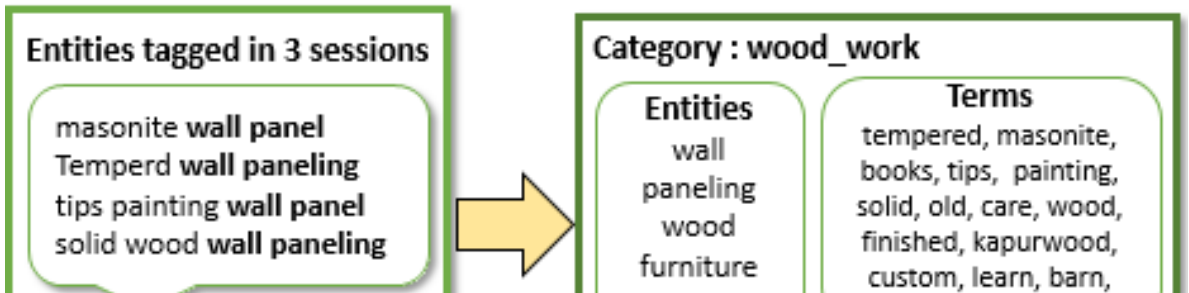
Extracting Search Tasks

Various proposed strategies:

- Hawkes Process based Task Extraction [Li, KDD'14]



- Entity-based Task Extraction [Verma, CIKM'14][White, CIKM'14]



Extracting Search Tasks

Problems abound:

- Link query to on-going task = long chains
- *Impure* tasks
- Pairwise predictions need extra post-processing
- Rely on large corpus of pre-tagged queries
- Do not aggregate across users

Random Fields based Task Extraction[3]

Propose a latent variable model for extracting tasks:

- Assumes k-underlying search tasks
- Places a Markov Random Field over the latent task layer
- Assumes access to a *task-oracle*
- Each latent task is a multinomial distribution over query terms

[3] R. Mehrotra, E. Yilmaz; **Extracting Search Tasks via Task Random Field Model**, SIGIR 2016 (**under review**)

Random Fields based Task Extraction

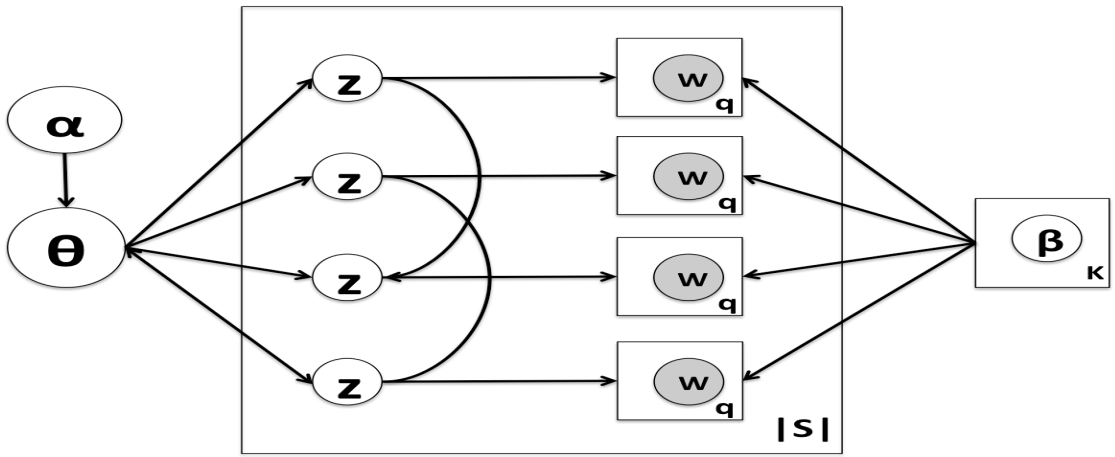


Figure: The proposed *task-RF* model for extracting search tasks.

The generative process:

- Draw task proportions vector $\theta \sim Dir(\alpha)$
- Draw task labels z for all queries from the joint distribution defined in Eq.(1)
- For each query (q_i) in the session:
  - For each query term $(w_{q_i,j})$ draw $w_{q_i,j} \sim multi(\beta_{z_i})$

Random Fields based Task Extraction

A query marginal factorizes into its query terms:

$$p(q|z,\beta) = \prod_{j=1}^{M_q} p(w_j|z,\beta) \tag{1}$$

The joint distribution of θ, z, q and w can be written as:

$$p(\theta, z, w|\alpha, \beta, \lambda) = p(\theta|\alpha)p(z|\theta,\lambda)\prod_{i=1}^{N}\prod_{j=1}^{M_{q_i}} p(w_{q_i,j}|z_i,\beta) \tag{2}$$

Under the MRF model, the joint probability of all task assignments $z = z_i{}_{i=1}^{N}$ can be written as:

$$p(z|\theta,\lambda) = \frac{1}{A(\theta,\lambda)}\prod_{i=1}^{N} p(z_i|\theta)\ exp\{\lambda \sum_{(m,n)\in E} \mathbb{1}(z_m = z_n)\} \tag{3}$$

Random Fields based Task Extraction

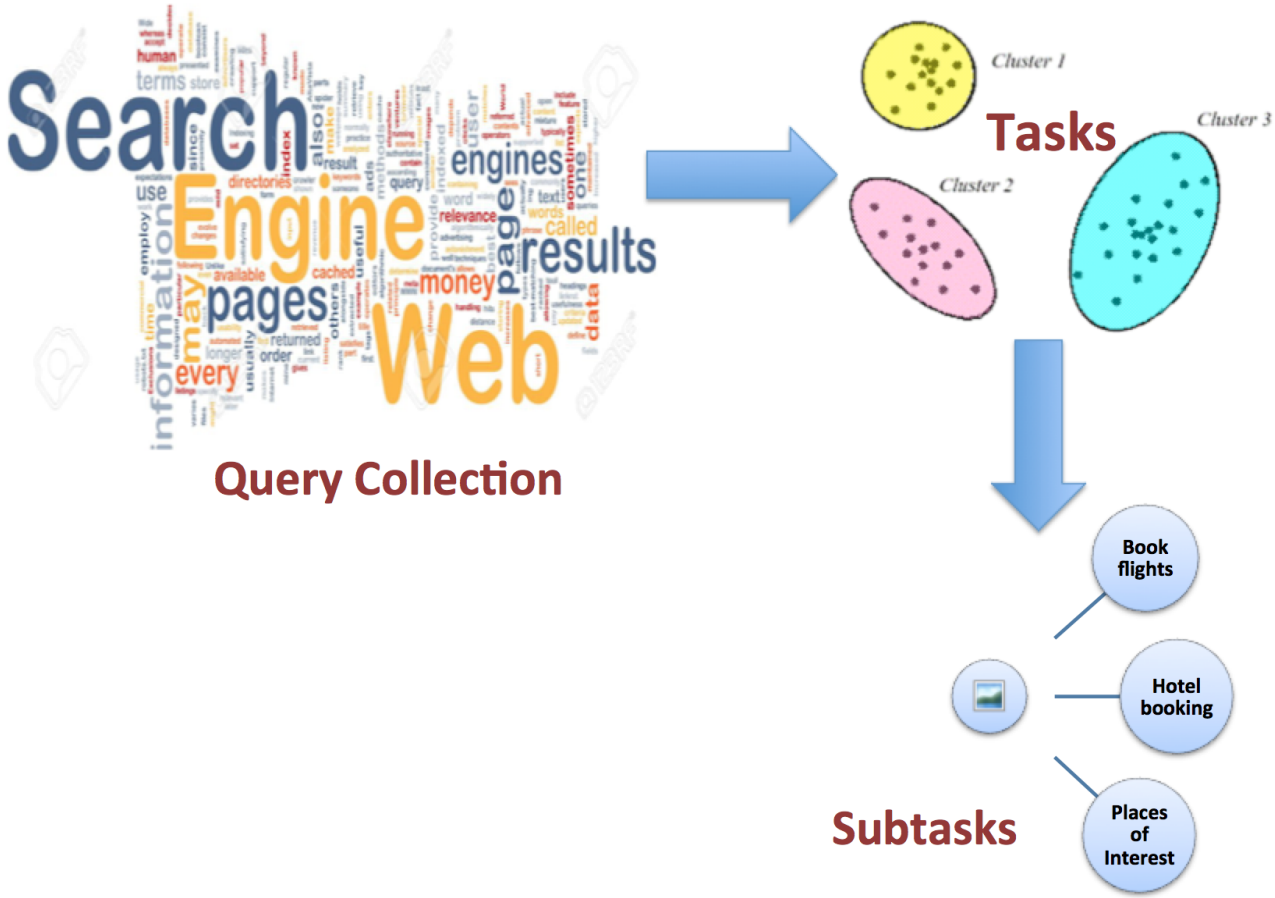
Task Oracle: classifier which predicts whether or not 2 queries belong to the same task.

Query Based Features	
cosine	cosine similarity between the term sets of the queries
edit	norm edit distance between query strings
Jac	Jaccard coeff between the term sets of the queries
Term	proportion of common terms between the queries
url-term-match-sum	$\sum_{u \in URL_i} m(q_j, u) + \sum_{u \in URL_j} m(q_i, u)$
url-term-match-max	$max_{u \in URL_i} m(q_j, u) + max_{u \in URL_j} m(q_i, u)$
Embedding	cosine distance between embedding vectors of the two queries
URL Based Features	
dom-match	the number of domain matches between URLs from both queries
Min-edit-U	Minimum edit distance between all URL pairs from the queries
Avg-edit-U	Average edit distance between all URL pairs from the queries
Jac-U-min	Minimum Jaccard coefficient between all URL pairs from the queries
Jac-U-avg	Average Jaccard coefficient between all URL pairs from the queries

Table: Features used by the classifier.

Extracting Tasks & Subtasks

1. Task extraction
2. **Subtask extraction**

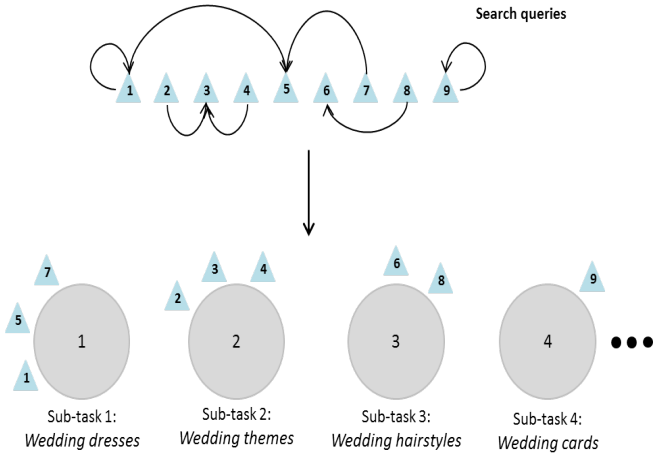


Subtask Extraction[4]

- Complex task decompose to more focussed subtasks
- Number of subtasks is unknown
- Given: collection of *on-task* queries
- Couple Bayesian Nonparametrics & Word Embeddings

[4] Subtask Extraction: R. Mehrotra, P. Bhattacharya, E. Yilmaz; **Exploiting Distributional Representations with Distance Dependent CRPs for Extracting Sub-Tasks**. Accepted at NAACL 2016

Chinese Restaurant Process



$$p(z_i = j \| D, \alpha) \propto \begin{cases} f(d_{ij}) & \text{if } j \neq i \\ \alpha & \text{if } j = i \end{cases}$$

1. For each query $i \in [1, N]$, draw assignment $z_i \sim dist - CRP(\alpha, f, D)$.
2. For each sub-task, $k \in \{1, ...\}$, draw a parameter $\theta_k^* \sim G_0$.
3. For each query $i \in [1, N]$,
   1. If $c_i \notin z_{q_{1:N}}^*$, set the parameter for the i^{th} query to $\theta_i = \theta_{q_i}$. Otherwise draw the parameter from the base distribution, $\theta_i \sim Dirichlet(\lambda)$.
   2. Draw the *ith* query, $w_i \sim Mult(M, \theta_i)$.

Quantifying Task based Distances

Task: plan a wedding
Sample queries: wedding planning, wedding *checklist*, bridal *dresses*, wedding *cards*

- Leverage word embeddings.
- Classify each word as **background** word or **subtask-specific** word.
- Use a weighted combination of their embedding vectors to encode a query's vector:

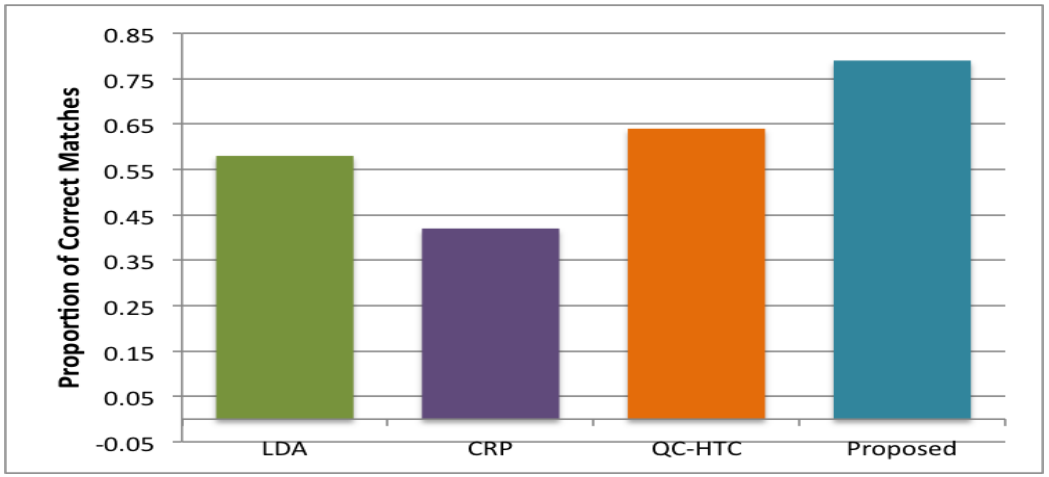
$$V_q = \frac{1}{n_{terms}} \sum_i \frac{n_{q_{t_i}}}{\sum_q n_q} V_{t_i} \tag{4}$$

Evaluating Quality of Subtasks

Qualitative Analysis:

Proposed Approach			LDA		
sub-task 1	**sub-task 2**	**sub-task 3**	**sub-task 1**	**sub-task 2**	**sub-task 3**
wedding hairstyles	used wedding dresses	wedding card holders	wedding insurance	christian wedding vows	make your own wedding invitations
wedding hair dos	colorful bridal gowns	indian wedding program cards	destination wedding dresses	brides	wedding cakes pictures
curly wedding hairstyles	preowned wedding dresses	wedding program paper	wedding planning book	cheap wedding dresses	planners
pictures of wedding hair	wedding attire	regency wedding cards	party supply stores	tea length wedding dresses	wedding colors
CRP			**TW**		
sub-task 1	**sub-task 2**	**sub-task 3**	**sub-task 1**	**sub-task 2**	**sub-task 3**
wedding planning kit	wedding theme	wedding insurance	wedding insurance	christian wedding vows	cheap dresses
destination wedding	wedding guide	weddings in vegas	destination wedding	plus size bridesmaid	wedding cakes pictures
wedding table decorations	save the date ideas	wedding cakes pictures	financing wedding rings	wedding colors	pricing weddings
1938 wedding pictures	wedding vacation	planning a wedding	party supply stores	tea length wedding dresses	macys wedding dresses

User Study:



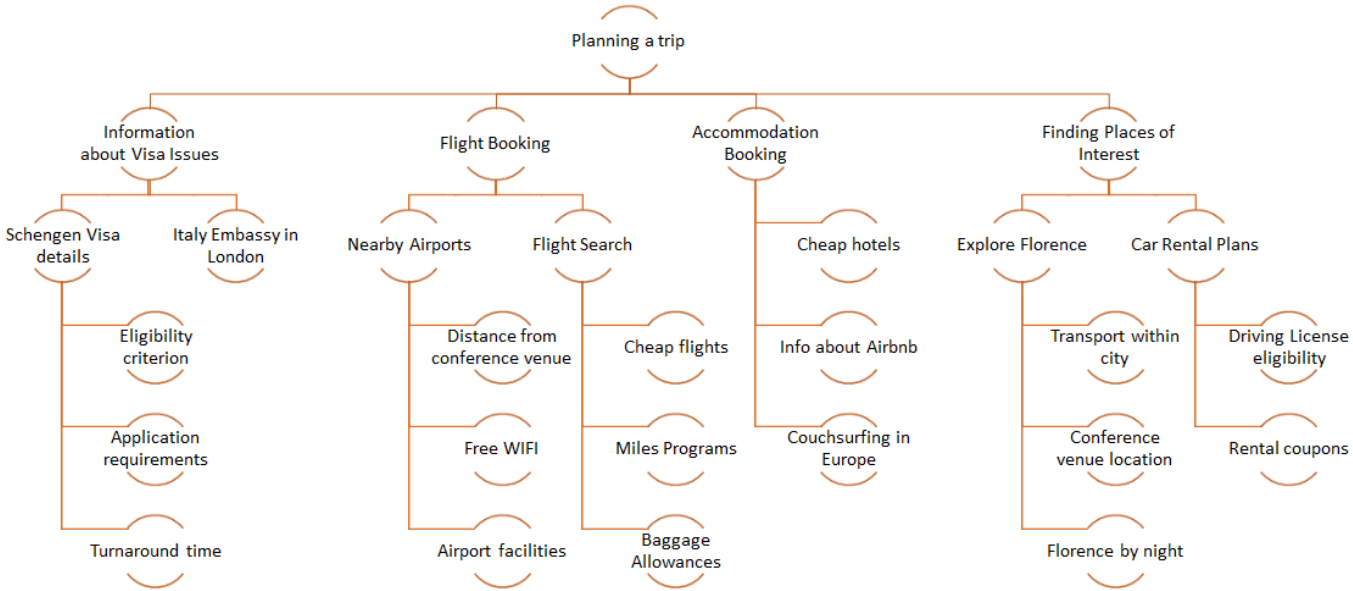
Takeaways

- Latent variable model for task extraction based on Task oracle
- Extracts more coherent tasks
- Nonparametric model for extracting subtasks
- Couple nonparametric approaches with word embeddings
- Human judgements & coherence estimates show improved results

Outline

Hierarchies of Tasks & Subtasks

- Complex search tasks places **intense cognitive burden** on the user
- Users need to explore domain-space
- Flat structure-less tasks **lack insights** about demarcation of subtasks
- Provides tremendous opportunity to **contextually target the user** (recommendations/advertisements)



Constructing Bayesian Hierarchies

Tree as partitions & mixtures:

Tree: partition over the group of queries (Q) where probability of a query group is defined via Dynamic Programming as

$$p(Q|T) = \pi_T f(Q) + (1 - \pi_t) \prod_{T_i \in ch(T)} p(leaves(T_i)|T_i) \quad (5)$$

Gamma-Poisson Model of Query Affinities:

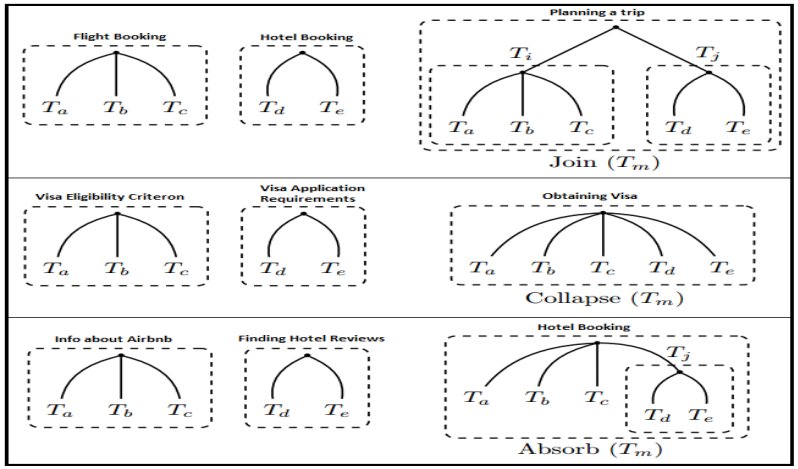
Query-Term Based Affinity (r^1)	
cosine	cosine similarity between the term sets of the queries
edit	norm edit distance between query strings
Jac	Jaccard coeff between the term sets of the queries
Term	proportion of common terms between the queries
URL Based Affinity (r^2)	
Min-edit-U	Minimum edit distance between all URL pairs from the queries
Avg-edit-U	Average edit distance between all URL pairs from the queries
Jac-U-min	Minimum Jaccard coefficient between all URL pairs from the queries
Jac-U-avg	Average Jaccard coefficient between all URL pairs from the queries
Session/User Based Affinity (r^3)	
Same-U	if the two queries belong to the same user
Same-S	if the two queries belong to the same session
Embedding Based Affinity (r^4)	
Embedding	cosine distance between embedding vectors of the two queries

Table: Query-Query Affinities.

Constructing Bayesian Hierarchies

Forming Arbitrary Tree Structure:

Bayesian Rose Trees (NIPS 2012)



Agglomerative Model Selection:

- Bottom-up greedy agglomerative fashion
- At each iteration: merges two of the trees in the forest
- Which pair of tree to merge(& how) - merger that yields the largest **Bayes factor improvement**

Takeaways[5]

- We can extract recursive hierarchies of tasks & subtasks
- Nonparametric tree partition model for hierarchies
- Allows for arbitrary structures
- Improves term prediction & task extraction
- Human labelled judgements show:
  - we extract coherent tasks
  - Subtasks are valid
  - Subtasks are useful for completing the overall task

[5] R. Mehrotra, E. Yilmaz; **Towards Hierarchies of Search Tasks & Subtasks**, In Proceedings of the Poster Track WWW 2015

Thank You!

References

R. Mehrotra, P. Bhattacharya, and E. Yilmaz.

Characterizing users' multi-tasking behavior in web search.
In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, 2016.

R. Mehrotra, P. Bhattacharya, and E. Yilmaz.

Deconstructing complex search tasks: A bayesian nonparametric approach for extracting sub-tasks.
In *Proceedings of NAACL-HLT*, pages 599–605, 2016.

R. Mehrotra, P. Bhattacharya, and E. Yilmaz.

Sessions; tasks & topics - uncovering behavioral heterogeneities in online search behavior.
In *Proceedings of the 2016 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2016.

R. Mehrotra and E. Yilmaz.

Towards hierarchies of search tasks & subtasks.
In *WWW 2015*.

R. Mehrotra and E. Yilmaz.

Terms, topics & tasks: Enhanced user modelling for better personalization.
In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, pages 131–140. ACM, 2015.

R. Mehrotra, E. Yilmaz, and M. Verma.

Task-based user modelling for personalization via probabilistic matrix factorization.
In *RecSys Posters*, 2014.

D. Odijk, R. W. White, A. Hassan Awadallah, and S. T. Dumais.

Struggling and success in web search.
In *CIKM 2015*.

White, Bennett, and Dumais