Task-Based Search and Assistance

SIGIR 2020 Tutorial

Chirag Shah
University of Washington

Ryen White
Microsoft Research AI

Introductions

- Who are we?
- Why are we here?
- How will we run this tutorial?

Chirag Shah
University of Washington

Ryen White
Microsoft Research AI

What is this tutorial about?

- Going **beyond queries** and even **sessions**
- Thinking through the **task** and the **context** in which users seek information
- Extracting **task information**
- Applying task knowledge to **search** and **recommendation** applications
- **Prerequisite**: basic understanding of IR systems and evaluation
- **Material**: slides and bibliography available through Github

Why does this matter?

- People engage in information seeking because they find themselves in **problematic situations** [Wersig ‘79, Belkin ‘83].
- Rather than information need, we should understand what people wish to accomplish (**task, goal**) [Wilson ‘81].
- People have a **task** behind their querying/questioning. Knowing that task could help systems serve the users better.
- Often **people don’t know what they don’t know**. If we are relying on them expressing their need, even in vague terms, we may be at a loss.
- Conversational systems and in general, **intelligent agents**, are supposed to work at the task level, understanding the user and the context.

Outline of the tutorial

1. Introduction
2. Explicating task
3. Case studies
4. Evaluation
5. Challenges and opportunities

Part I: Introduction

Motivating example

What can we say from looking at these queries in a session?

1. Nigerian scam email
2. Nigerian scam unemployment
3. Washington unemployment scam
4. Email for reporting unemployment scam
5. contact for reporting unemployment scam

What can we do differently here?

1. Nigerian scam email
   - No clicks
2. Nigerian scam unemployment
   - Click on a WIRED story
3. Washington unemployment scam
   - Click on a Seattle Times story
4. Email for reporting unemployment scam
   - Clicks on Department of Labor and FTC sites
5. contact for reporting unemployment scam
   - No clicks

Explicating the Task

- Find information about Nigerian scam email
- Find stories related to Nigerian scam email and/or unemployment
- Find stories related to Nigerian scam email and/or unemployment concerning Washington state
- Find email for knowing/reporting Nigerian scam and/or unemployment concerning Washington state

A scenario with an intelligent agent

User: I think I would like to go do some outside activity today. Do I need to wear a mask if I go running?

Agent: It depends where you are running, but if you are concerned about safety and still want an outdoor activity, may I suggest biking?

User: Oh.. ya, sure, that could work. Do I need to know anything?

Agent: While you don’t need to wear a mask while biking, you should still bring one with you. There is also a chance of some rain showers, so plan for that. And yes, definitely carry some water.

What is this scenario addressing?

- Understanding the intention behind a user seeking information. [need to do outdoor activity while being safe]
- People don’t know what they don’t know. [what do I need to know if I go biking?]
- Zero-query recommendations. [giving warning about the weather and water]
- Proactive recommendations. [suggesting biking as an alternative]

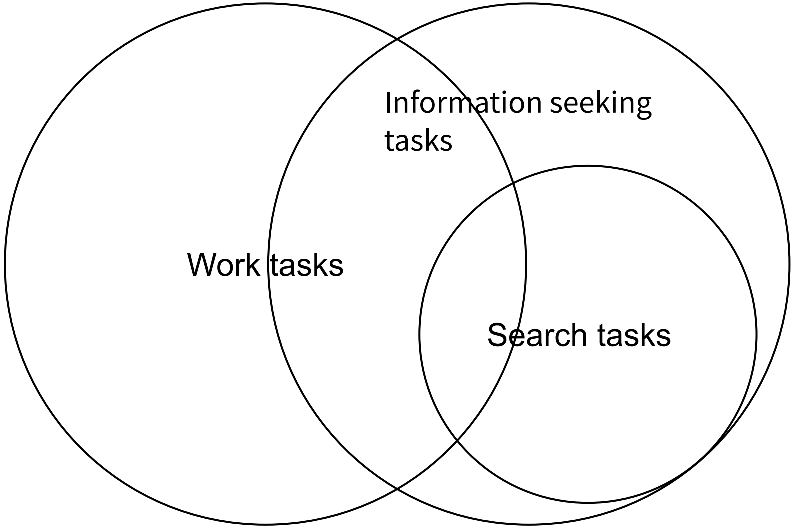
What are the challenges here?

- Abstracting out from a query or a question or even an observation to the task and/or context.
- Generating recommendations based on that task/context and weighing if that would be better than query/question-based recommendation.
- Learning how to do a task.

Part II: Explicating task

What is a task?

- Work task and simulated work task
- Search task
- Information seeking task
- Explicitly vs. implicitly expressed



Frameworks and models for task

McGrath, J. E. (1984). *Groups: Interaction and performance*(Vol. 14). Englewood Cliffs, NJ: Prentice-Hall.

Categories	Task types and explanations
Generate	Planning tasks: Generating plans. Key notion: action-oriented plan
	Creativity tasks: Generating ideas. Key notion: creativity
Choose	Intellective tasks: Solving problems with a correct answer. Key notion: correct answer
	Decision-making tasks: Dealing with tasks for which the preferred or agreed upon answer is the correct one. Key notion: preferred answer
Negotiate	Cognitive conflict tasks: Resolving conflicts of viewpoint (not of interests). Key notion: resolving policy conflicts
	Mixed-Motive tasks: Resolving conflicts of motive-interest. Key notion: resolving pay-off conflicts
Execute	Contests/Battles: Resolving conflicts of power; competing for victory. Key notion: winning
	Performances: Psychomotor tasks performed against objective or absolute standards of excellence, e.g., many physical tasks; some sports events. Key notion: excelling

Frameworks and models for task

Xie, H. (1997). Planned and situated aspects in interactive IR: Patterns of user interactive intentions and information seeking strategies. In Proceedings of the ASIST Annual Meeting (Vol. 34,

Goals	Categories
Leading search goals	Recreational use
	Professional use
	Educational assignment
	Personal information use
	Other
Current search goals	Looking for specific item
	Looking for specific information
	Looking for items with common characteristics
	Keeping up to date
	Other
Interactive intentions	Identifying
	Learning
	Finding an item(s)/information
	Locating
	Accessing
	Evaluating
	Keeping record
	Obtaining an item(s)
	Combination of interactive intentions

Frameworks and models for task

Marchionini, G. (1989). Information seeking strategies of novices using a full-text electronic encyclopedia. Journal of the American Society for Information Science, 40(1), 54–66.

Byström, K. (2002). Information and information sources in tasks of varying complexity. Journal of the American Society for information Science and Technology, 53(7), 581-591.

Byström, K., & Hansen, P. (2002). Work tasks as units for analysis in information seeking and retrieval studies. Emerging frameworks and methods, 239-251.

Byström, K., & Hansen, P. (2005). Conceptual framework for tasks in information studies. Journal of American Society for Information Science and Technology, 56(10), 1050–1061.

Frameworks and models for task

Li, Y., & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. Information Processing & Management, 44(6), 1822–1837.

Generic facets: Source of task, Task doer, Time, Product, Process, Goal

Common attributes: Task characteristics, User’s perception of task

Frameworks and models for task

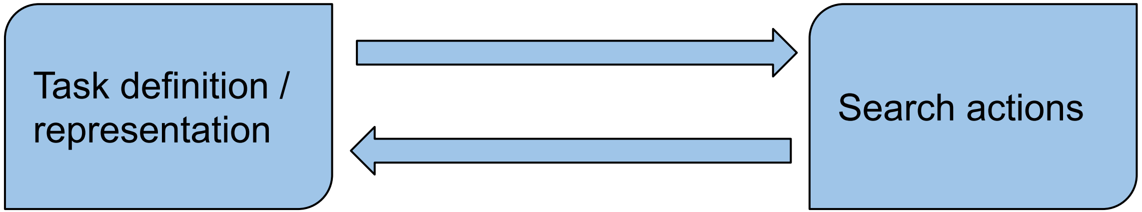
Mehrotra, R., & Yilmaz, E. (2015). Terms, topics & tasks: Enhanced user modelling for better personalization. In Proceedings of the 2015 International Conference on The Theory of Information Retrieval (pp. 131-140).

He, J., & Yilmaz, E. (2017). User behaviour and task characteristics: A field study of daily information behaviour. In Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval (pp. 67-76).

Mitsui, M., & Shah, C. (2019). Bridging Gaps: Predicting User and Task Characteristics from Partial User Information. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 415-424). ACM. Paris, France.

Liu, J., Sarkar, S., & Shah, C. (2020). Identifying and Predicting the States of Complex Search Tasks. Proceedings of ACM Conference on Human Information Interaction and Retrieval (CHIIR). March 14-18. 2020. Vancouver, Canada.

Tasks in search



User generated tasks

- Wunderlist (now Microsoft To Do), Google Keep, Apple Reminders
- Work done at MSR AI
  - **Nouri, E., Fourney, A., Sim, R., & White, R. W. (2019)**. Supporting complex tasks using multiple devices. In Proceedings of WSDM'19 Task Intelligence Workshop (TI@ WSDM19).
  - **Nouri, E., Sim, R., Fourney, A., & White, R. W. (2020)**. Step-wise Recommendation for Complex Task Support. In Proceedings of the 2020 Conference on Human Information Interaction and Retrieval (pp. 203-212).

Researcher assigned tasks in IR: TREC

spider bites [My friend may have been bitten by a spider and I need to know what to do]

  *   Identify which spiders are harmless and which are dangerous (eg black widow)
  *   Identify a spider's bite from a bug's bite (or something else that mimics a spider's bite)
  *   Find out which are the symptoms of spider bites and its side effects
  *   Seek emergency medical care
  *   Symptoms of dangerous spider bites
  *   Take appropriate medicine (ex. tetanus booster shot or an anti venom)
  *   Learn how to provide first aid to someone who is bitten by a spider
  *   Ways of non medicine treatment (ex. Wash the site of the spider bite well with soap and water, apply a cool compress,etc)
  *   Check out whether you are allergic to spiders

Researcher assigned tasks in Interactive IR

- How scholars define and assign tasks for IIR studies
- What kind of data they collect and how they analyze
- Some common themes
  - Using task type as a controlled/independent variable
  - Seeing the effects of task type to behaviors, intentions, and other dependent variables
- Systematic review of assigned search tasks: https://ils.unc.edu/searchtasks/

Tasks in conversational assistants

- **Limitation**: multi-turn is hard with error increasing exponentially with each turn.
- **Challenge**: modalities are different than ‘classical’ IR interactions.
- **Opportunity**: naturalistic, wider applicability, possibility for being proactive and truly *intelligent*.

Tasks in conversational assistants

Trippas, J. R., Spina, D., Scholer, F., Awadallah, A. H., Bailey, P., Bennett, P. N., ... & Sanderson, M. (2019). Learning about work tasks to inform intelligent assistant design. In Proceedings of the 2019 Conference on Human Information Interaction and Retrieval (pp. 5-14).

White, R. W., Fourney, A., Herring, A., Bennett, P. N., Chandrasekaran, N., Sim, R., ... & Encarnación, M. J. (2019). Multi-device digital assistance. Communications of the ACM, 62(10), 28-31.

Part III: Case studies

Tasks in IIR studies: case study-1

Task-1: Copy Editing (Goal: Specific, Product: Factual)

You are a copy editor at a newspaper and you have only 20 minutes to check the accuracy of the six italicized statements in the excerpt of a piece of news story below. Please find and save an authoritative page that either confirms or disconfirms each statement.

Task-2: Relationships (Goal: Amorphous, Product: Intellectual)

You are writing an article about coelacanths and conservation efforts. You have found an interesting article about coelacanths but in order to develop your article you need to be able to explain the relationship between key facts you have learned. In the following there are five italicized passages, find an authoritative web page that explains the relationship between two of the italicized facts.

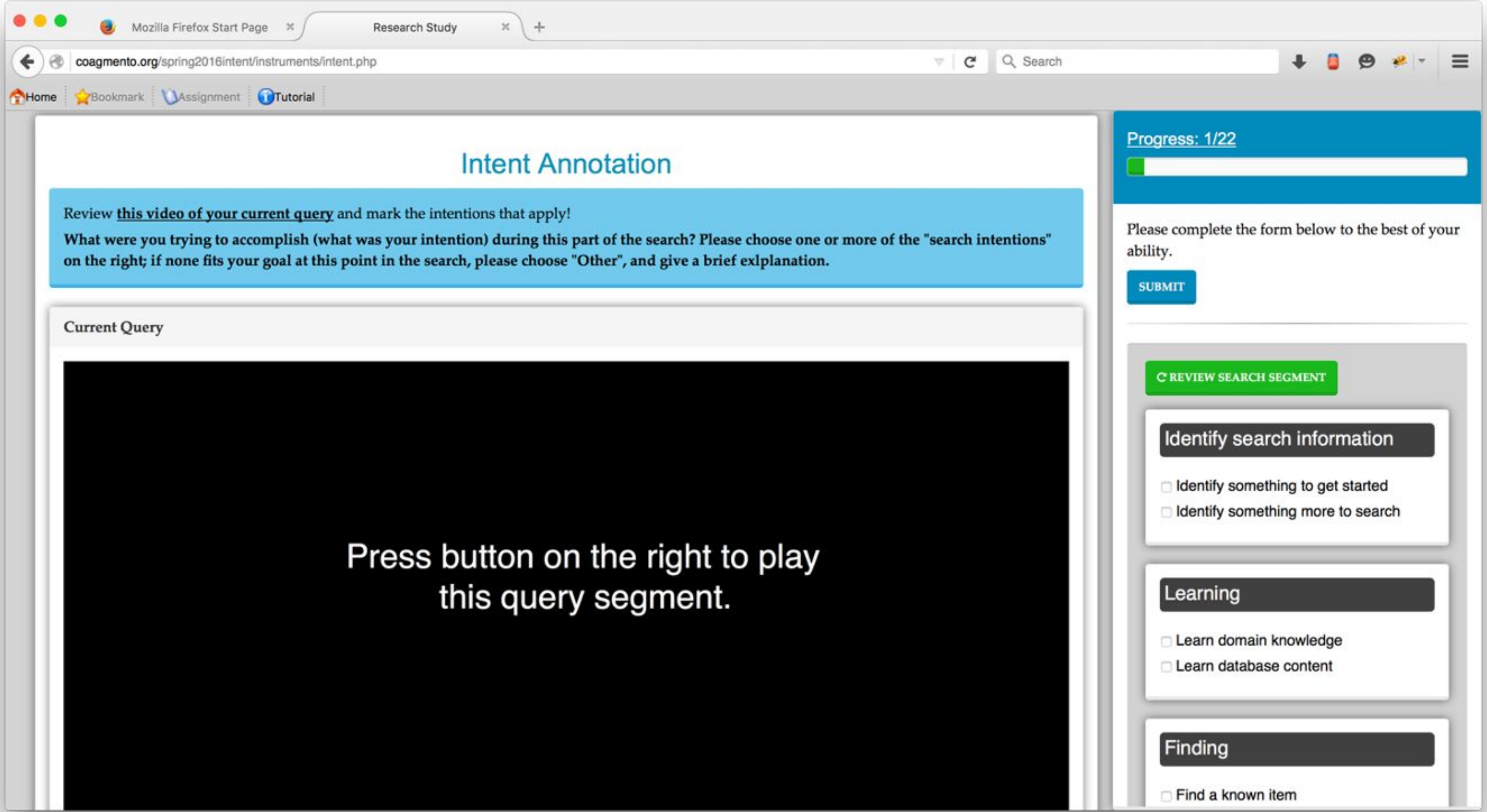
Explicating task from search actions/logs

- Datasets:
  - Lab study data: 40 users with searches on 2 topics
  - TREC Session Track data: 260 users with searches on 60 topics
- Features: query length, dwell time on SERP, dwell time on content pages, no. of pages visited
- Goal: predict the ‘goal’ (specific, amorphous) and ‘product’(factual, intellectual) aspects of the task.
- Results: Multilayer perceptron (MLP) gives best results (66% to 74% accuracy) in most cases, often **first query prediction tying with whole session**.

Mitsui, M., Liu, J., & Shah, C. (2018). How Much is Too Much? Whole Session vs. First Query Behaviors in Task Prediction. In Proceedings of ACM SIGIR 2018 Conference. 4 pp. July 8-12, 2018. Ann Arbor, MI.

Tasks in IIR studies: case study-2

- Task = Topic + Intention
- 40 participants – journalism majors
- 80 Sessions (40 users x 2 sessions)
- 20 minutes per search session
- 693 query segments
    - Each labeled with 20 possible intentions (by user)
- Task characteristics: Goal (Specific, Amorphous), Product (Factual, Intellectual)
- Features related to queries, SERPs, content pages
- Classifiers with 66% data for training



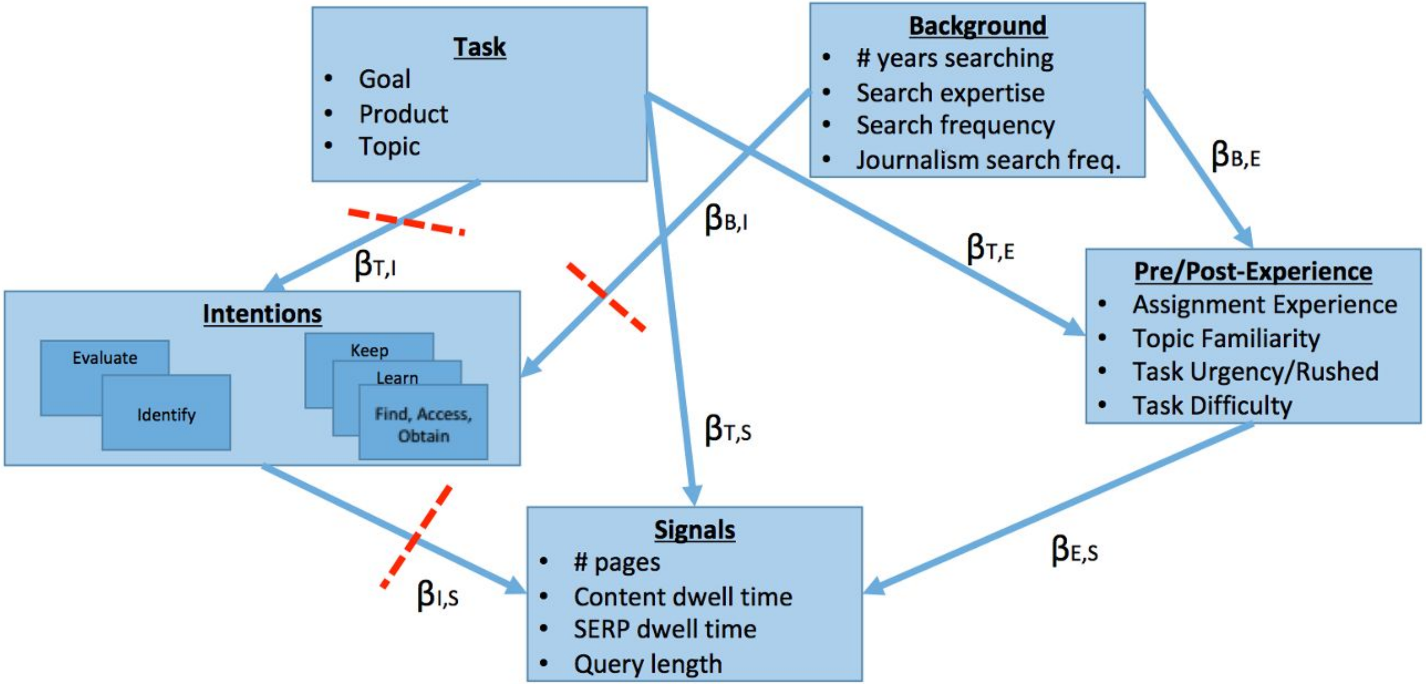
Mozilla Firefox Start Page
Research Study
coagmento.org/spring2016intent/instruments/intent.php
Search
Home
Bookmark
Assignment
Tutorial
Intent Annotation
Review this video of your current query and mark the intentions that apply!
What were you trying to accomplish (what was your intention) during this part of the search? Please choose one or more of the "search intentions" on the right; if none fits your goal at this point in the search, please choose "Other", and give a brief exlplanation.
Current Query
Press button on the right to play this query segment.
Progress: 1/22
Please complete the form below to the best of your ability.
SUBMIT
REVIEW SEARCH SEGMENT
Identify search information
Identify something to get started
Identify something more to search
Learning
Learn domain knowledge
Learn database content
Finding
Find a known item

Results

Intention	% Positive	ACC(feat)	ACC(STR)	ACC(MFQ)
Access Common (AC)	24.30	**0.761(CP)**	0.628	0.755
Access Page (AP)	10.80	**0.900(CP+BK)**	0.812	0.894
Access Specific (AS)	27.60	**0.731(ALL)**	0.598	0.721
Evaluate Best (EB)	19.80	**0.815(ALL)**	0.669	0.792
Evaluate Correctness (EC)	26.30	**0.754(ALL)**	0.610	0.735
Evaluate Duplication (ED)	7.50	**0.929(ALL)**	0.856	0.922
Evaluate Specific (ES)	23.80	**0.776(ALL)**	0.646	0.766
Evaluate Usefulness (EU)	25.30	**0.775(ALL)**	0.638	0.763
Find Characteristic (FC)	20.10	**0.806(ALL)**	0.675	0.797
Find Known (FK)	17.00	**0.832(ALL)**	0.705	0.825
Find Without Predefined (FP)	8.00	**0.926(ALL)**	0.858	0.922
Find Specific (FS)	57.10	**0.608(ALL)**	0.511	0.579
Identify More (IM)	37.50	**0.668(ALL)**	0.540	0.641
Identify Specific (IS)	29.00	**0.817(ALL)**	0.568	0.688
Keep Record (KR)	33.40	**0.714(ALL)**	0.551	0.659
Learn Database (LD)	16.20	**0.839(BK)**	0.729	0.837
Learn Domain Knowledge (LK)	33.20	**0.712(ALL)**	0.548	0.657
Obtain Part (OP)	18.90	**0.802(ALL)**	0.679	**0.802**
Obtain Specific (OS)	43.20	**0.645(ALL)**	0.513	0.581
Obtain Whole (OW)	8.30	**0.918(CP)**	0.850	0.917

Mitsui, M., Liu, J., Belkin, N., & Shah, C. (2017). Predicting Information Seeking Intentions from Search Behaviors. In Proceedings of ACM SIGIR 2017 Conference. 4 pp. August 7-11, 2017. Tokyo, Japan.

Tasks in IIR studies: case study-3



Mitsui, M., & Shah, C. (2019). Bridging Gaps: Predicting User and Task Characteristics from Partial User Information. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 415-424). ACM. Paris, France.

Results

- The relationship between task and behavior is better explained when considering intervening user characteristics
  - e.g., general search background, topic familiarity, task difficulty
- Traditional task-behavior relationships still hold
  - e.g., Task product -> browsing behaviors
- The specifics of these relationships may change in the presence of indirect effects
  - e.g., Task product-> Difficulty -> # pages

Summary

- Researchers have argued for decades that IR systems should help people get tasks done.
- Many IR experiments and IIR studies use predefined tasks to study user behaviors.
- Recently, more works have started appearing that do the reverse - using user behaviors to determine the task.
- Why? So they can make task-level recommendations. Example: e-commerce.
- Task = Topic (what) + Intention (why)
- Recommendation = Strategy (how)

Part IV: Evaluation

Overview

- Evaluation of intelligent systems important to understand their effectiveness

- Need to evaluate these (complex) systems holistically, per component [Balog, 2015]

- Many standard evaluation methods (user studies, etc.) apply; see Kelly [2009]

- Focus on **metrics for intelligent systems** and answer questions such as:
    - How do we measure user satisfaction?
    - How do we determine task progress and task completion?
    - How do we assess the performance of systems supporting proactive experiences?

Metrics

Process Metrics

Covered in this tutorial

- Time and effort
- Engagement
- Progress

Others include

- Cognitive load [Card et al., 1983]
- Learning [Agosti et al., 2014]
- Affect [Feild et al., 2010]
- Usability [Albert & Tullis, 2013]

Outcome Metrics

Covered in this tutorial

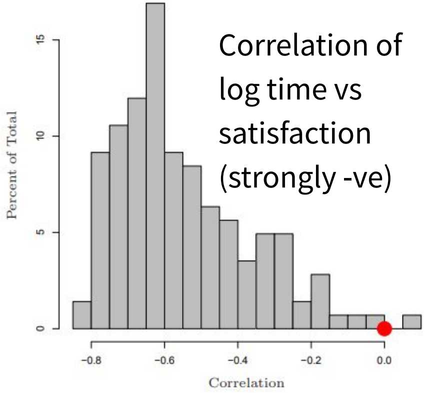
- Relevance
- Success
- Satisfaction

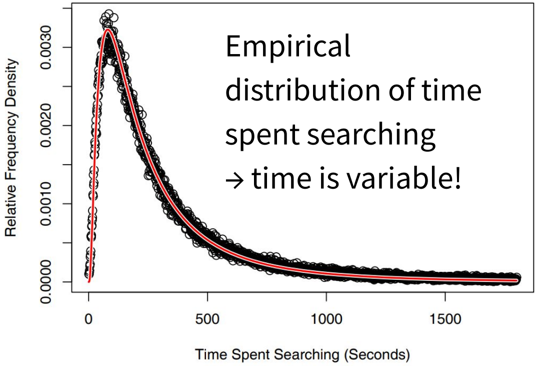
Others include

- Novelty and diversity [Clarke et al., 2008]
- Creativity [Shneiderman, 2000]
- Adoption and retention [White et al., 2010]

Time

Time on task is a key productivity and usability metric [Czerwinski et al., 2004]





In search

- Task completion time as search evaluation metric [Xu & Mease, 2009]
- Time-biased gain (consider factors affecting time) [Smucker & Clarke, 2012]

Subjective perceptions of elapsed time are problematic

- Attentional demand [Zakay & Block, 1996], experience [Thomas et al., 2004], etc.
- For easy tasks, people tend to overestimate time [Boltz et al., 1998]

Time forecasting is also problematic

- Biases such as "planning fallacy" (overconfidence) [Kahneman & Tversky, 1979]
- Estimating task duration using deep learning [White et al., 2019]

Effort

Quantifiable measures of work to complete task, e.g.,

- Search: Number of queries/clicks
- Assistance: Number of actions/steps
- Conversation: Number of dialog turns

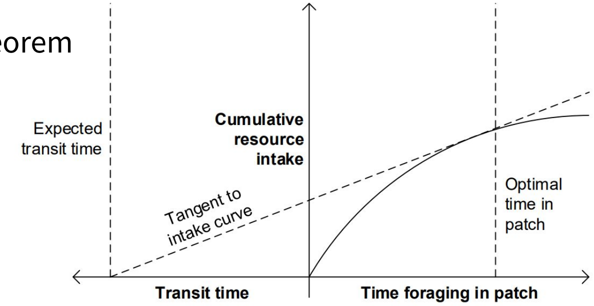
Research on information foraging theory (IFT) trades off costs and benefits of effort and reward [Pirolli et al., 1996]

- C/W/L applies IFT principles for search [Azzopardi et al., 2018]

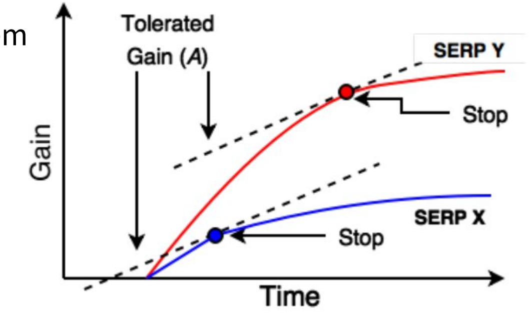
ESL [Cooper, 1968] - expected number of docs read to find relevant docs

User satisfaction can depend on amount of effort to complete complex tasks [Kelly, 2015]

Marginal value theorem (general)



Marginal value theorem (applied to search)



Engagement

“***a user’s response to an interaction that gains, maintains, and encourages their attention, particularly when they are intrinsically motivated***” [Jacques, 1996]

Engagement refers to the emotional, cognitive, and behavioral connection that exists at any point in time and over time, between users and the system [White, 2016]

Includes work on user activity tracking (not just result clicks - hover, swipe, gaze, etc.)

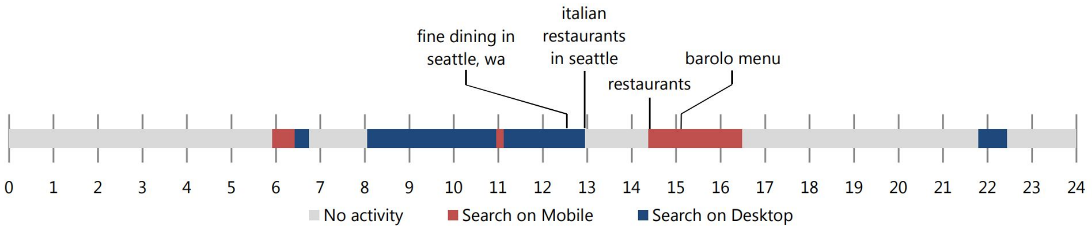
- Also, self reports and cognitive measures (physiological, perceptual)

Signals can be combined - watch for interaction effects (e.g., think aloud + activity)

Engagement affected by many factors

- User, task, UX, biases, etc. [O’Brien & Toms, 2008; Lalmas et al., 2014]

Progress



How far through a task a user has gotten / how close to completion they are

Tasks can span multiple sessions [Kotov et al., 2011] and devices [Wang et al., 2013]

Progress of individuals or teams tracked using dedicated tools [Bellotti et al., 2004]

Also studied in task-oriented dialog systems

- e.g., number of slots filled (*x* of *y*) [Budzianowski et al., 2018]

Possible to clearly measure for stepwise tasks (cooking, reservations)

We may only observe some user actions, making it difficult to reliably track progress

Relevance

Relevance metrics help estimate support for task completion (proxy for task success)

- Usually per query, but session-based metrics also proposed [Jarvelin et al., 2008]

Many metrics proposed, e.g., MAP, DCG, P@k, RBP, INST, etc.

- All encode different user models [Moffat et al., 2017]

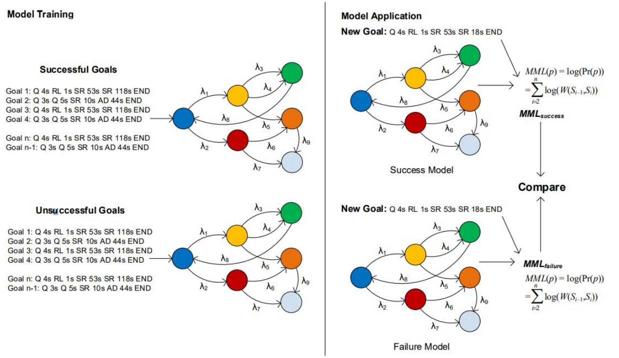
Applied offline with third party judgments (user and task effects are important - more later)

Relevance personal, situational [Mizzaro, 1997] - changes w/ task stage [Taylor et al., 2007]

Used in TREC Tasks Track (2015-2017) - alongside **utility** and **task understanding**

Does not always reflect task completion - search is just beginning, esp for complex tasks

Success



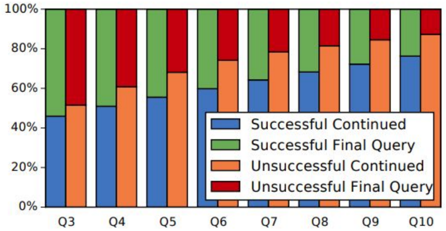
Success modeling

Measures goal completion - can be successful with DSAT

Behavior > traditional metrics in modeling task success [Hassan et al., 2010]

Success can be **objective** (e.g., factually correct) or **subjective** (e.g., perceived correct)

- Objective success is typically focus - subjective can be biased

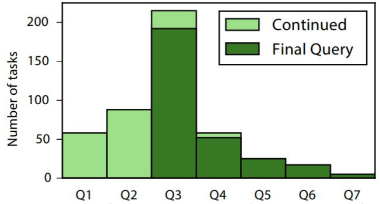


Query success over session

Struggling is common in search task completion [Odjik et al., 2015]

Task success in task-oriented dialog systems often tied to task completion (place order, make reservation, etc.)

- Might be the case in search too, but often unobservable

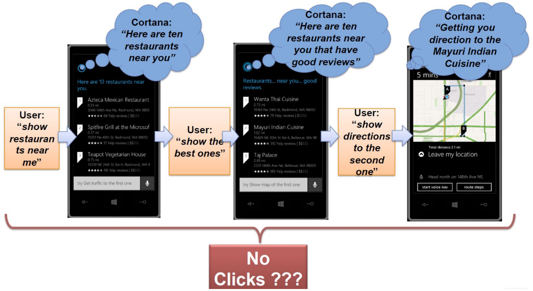


Pivotal query within a task

Success in conversational systems: human eval [Liu et al., 2017]

Satisfaction


Cursor movement positions

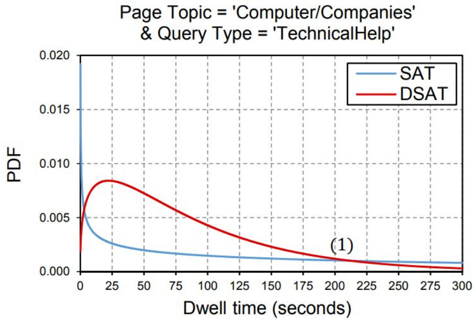


Satisfaction is emotional response - more general than success

- Psychology [Lopez & Snyder, 2011], commerce [Oliver, 2014]

Satisfaction modeling mostly at session level [Hassan et al., 2011]

- Led to 30s dwell time as SAT click [Fox et al., 2005]



Dwell time used as a measure of satisfaction

- Task effects [Kelly & Belkin, 2004]; Topic and complexity effects [Kim et al., 2014]
- Absence of clicks can also be a good sign - good abandonment [Li et al, 2009]
- Cursor-based modeling [Huang et al., 2012; Guo & Agichtein, 2012]

Research has considered user satisfaction on intelligent assistants

- Using click, touch, and voice interactions yields ~80% accuracy [Kiseleva et al., 2016]

Factors Affecting Task Performance

Task attributes, including

- Type e.g., [Mitsui and Shah, 2018]
- Topic e.g., [Mehrotra et al., 2017]
- Difficulty e.g., [Wildemuth et al., 2014]
- Complexity e.g., [Byström & Järvelin, 1995]
- Urgency e.g., [Mishra et al., 2014]

User attributes, including

- Subject matter expertise e.g., [White et al., 2009]
- Familiarity with task/topic e.g., [Kelly & Cool, 2002]

And, of course, **system support** - we'll now look at some examples of this ...

Case Studies

Four examples of scenarios requiring different task support and different metrics

1. Intelligent notifications
    - Offering non-redundant task reminders

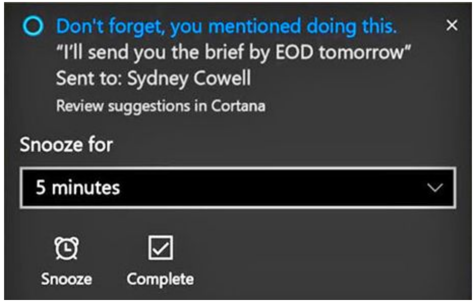
2. Skill discovery
    - Suggesting relevant skills based on context

3. Contextual search
    - Re-ranking search results based on previous searches

4. Conversational systems
    - Multi-modal support for complex tasks

1 - Intelligent Notifications

Cortana provide notifications for pending tasks (commitments)



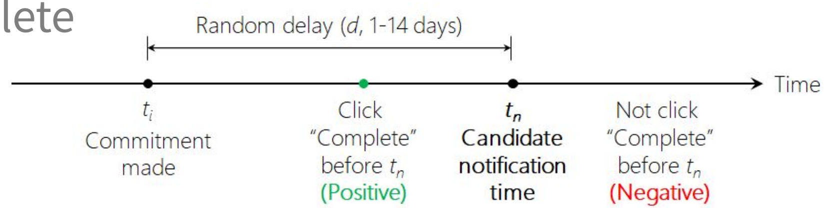
Do not want to suggest commitments that are already completed

"Mark Complete" is a key affordance

- Users may not explicitly indicate when tasks are complete

Use data from cases where users mark tasks complete

- Auto detect completion [White et al., 2019]



Metrics include

- Satisfaction: Fewer redundant interrupts for users [holistic]
- Accuracy: % correctly suppressed notifications (offline SAT proxy) [component]

2 - Skill Discovery

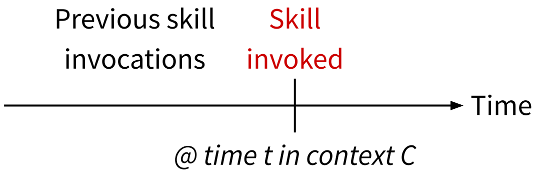
Skill discovery in headless devices

- Users unaware of what can do intelligent assistants can do

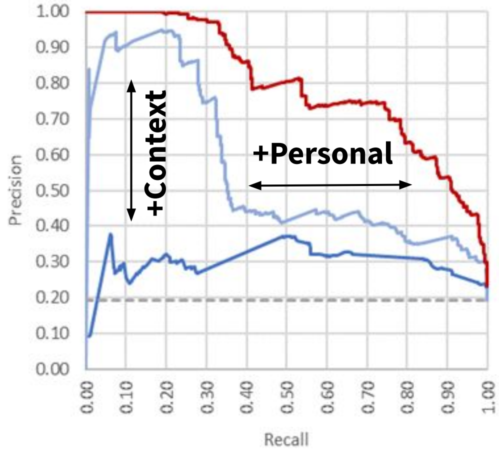
Suggesting right skill(s) at the right time - based on user context

Learn recommendation model based on skill usage [White, 2018]

Ranking problem



Learn to rank current skill + previous skills



Baseline (always suggest)
Popularity
Popularity + Context
Popularity + Context + Personal

Metrics include

- Engagement: Whether suggested skill used [holistic]
- Precision-recall: Suggesting used skills [component]
  (offline engagement proxy)

3 - Contextual Search

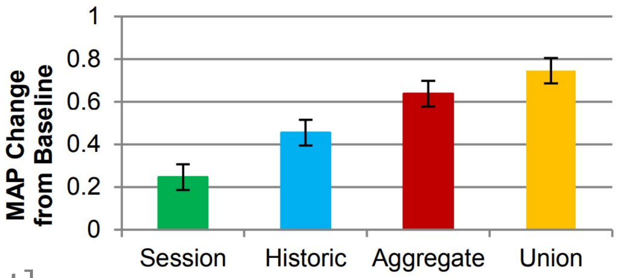
Work in Bing on contextual search [White et al., 2010; Bennett et al., 2012]

Shipped many different types of contextualization (location, reading level, etc.)

Previous queries (short- and long-term) models more of the search task

Need ground truth to evaluate performance and train contextual rankers

Need context-sensitive labels and metrics (clicks+context)



Metrics include

- Success: % of tasks completed [holistic]
- Relevance: MAP (offline success proxy) [component]

4 - Conversational Systems

Multi-device experiences (MDX) [White et al., 2019]

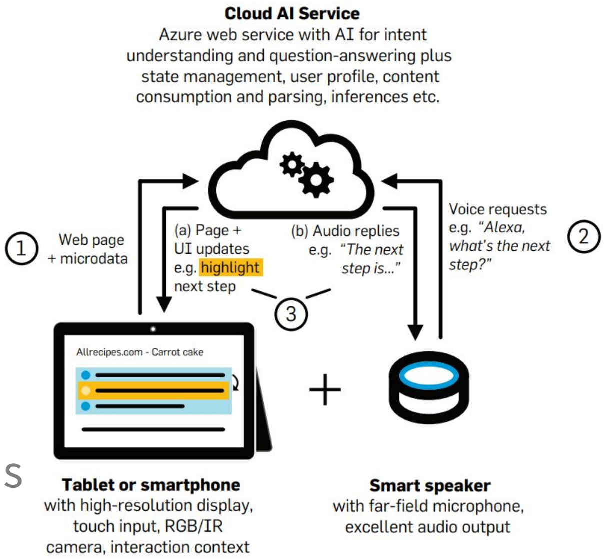
- Combination of smart speaker + smartphone/tablet

Focus on complex tasks, in this case, recipe preparation

Models for ASR, intent understanding, Q&A, recommendations

Metrics include

- Time: Time on task [holistic]
- Effort: Number of dialog turns [holistic]
- Answer correctness (Q&A) [component]
- Number of recognition errors (ASR) [component]
- Accuracy (intent understanding) [component]



Challenges in Evaluation (1 of 2)

Task-based systems can be highly complex

- Difficult to attribute outcome to one component / model component interactions
- E.g., in MDX - ASR, intent understanding, Q&A, recommendations (all interact)

Task activity is often unobservable, both process and completion

- Many task-related events are invisible to systems and not archived
- Users reluctant to indicate progress or completion

End-to-end task completion for many tasks spans multiple applications and devices

- Focusing on a single application (search engine) or device (PC) is too limited
- Need to evaluate task performance across applications and devices

Challenges in Evaluation (2 of 2)

Many metrics, for all systems - especially for complex systems / task scenarios

- Metrics may lead to different system orderings
- Important to prioritize metrics a priori

Many factors influence task performance but are not codified

- Consider task difficulty, etc. in system evaluations
- Report performance for different task management personas (see next slide)

A lot of ML and data analytics work in this area relies on third-party labels

- Classifying tasks or activities with a taxonomy is difficult [Russell et al., 2009]
- First-party labeling reliable - retrospective [Kelly, 2004], in-situ [Liono et al., 2020]

Many of these challenges are also research **opportunities!**

Credit: Carina Stefes

Task Management Personas

Based on analysis of Wunderlist customer feedback and interviews:

Capture-Organize ←→ Achieve → Analyze

Evaluation - Summary

Evaluating task-based systems is important but also challenging

Many missing signals that paint an incomplete picture

Triangulating signals from multiple applications is preferable (with user consent)

Each scenario has own task setup (see Case Studies)

- Careful thought required in devising metrics per application

Note: Focus is on metrics - but we also need to consider **evaluation methodologies**

- Includes the realism of any assigned tasks (simulated work tasks [Borlund, 2000])
- Setting, participants, baselines, etc.

Part V: Conclusions and Future Directions

Overall Conclusions

Intelligent systems should help people get tasks done

- Task-based search and assistance critical area that needs more study

Task and user characteristics play a large part in task behavior and task performance

Task information can be inferred from user behaviors (+ content analysis)

Systems can support tasks in many ways, e.g., recommendations for e-commerce

Evaluation is challenging, especially given limited information about users and tasks

- Need to consider a range of metrics, per application, holistic and per component
- Other metrics also important: learning, creativity (short- and long-term) … **+ more**

Examples of Future Directions

Systems need to provide end-to-end task support

- Seamless integration with existing tools
- Cross application - no more silos
- AI can help in all three task phases: capture, focus, do [Allen, 2015]

Task understanding

- Need to better model task intents (task2vec)
- Need more signals on task progress and task completion (behavior, content)

Task completion

- Last mile in search interaction - search engines → task completion engines
- Support completion when we know task (e.g., to-do tasks) (task2search)

Thank you!

http://chiragshah.org
http://ryenwhite.com