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
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


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


<table>
<thead>
<tr>
<th>Goals</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leading search goals</td>
<td>Recreational use</td>
</tr>
<tr>
<td></td>
<td>Professional use</td>
</tr>
<tr>
<td></td>
<td>Educational assignment</td>
</tr>
<tr>
<td></td>
<td>Personal information use</td>
</tr>
<tr>
<td></td>
<td>Other</td>
</tr>
<tr>
<td>Current search goals</td>
<td>Looking for specific item</td>
</tr>
<tr>
<td></td>
<td>Looking for specific information</td>
</tr>
<tr>
<td></td>
<td>Looking for items with common characteristics</td>
</tr>
<tr>
<td></td>
<td>Keeping up to date</td>
</tr>
<tr>
<td></td>
<td>Other</td>
</tr>
<tr>
<td>Interactive intentions</td>
<td>Identifying</td>
</tr>
<tr>
<td></td>
<td>Learning</td>
</tr>
<tr>
<td></td>
<td>Finding an item(s)/information</td>
</tr>
<tr>
<td></td>
<td>Locating</td>
</tr>
<tr>
<td></td>
<td>Accessing</td>
</tr>
<tr>
<td></td>
<td>Evaluating</td>
</tr>
<tr>
<td></td>
<td>Keeping record</td>
</tr>
<tr>
<td></td>
<td>Obtaining an item(s)</td>
</tr>
<tr>
<td></td>
<td>Combination of interactive intentions</td>
</tr>
</tbody>
</table>
Frameworks and models for task


Frameworks and models for task


**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


Tasks in search

Task definition / representation

Search actions
User generated tasks

- Wunderlist (now Microsoft To Do), Google Keep, Apple Reminders
- Work done at MSR AI
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/](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


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.

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
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 explanation.

Current Query

Press button on the right to play this query segment.
### Results

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

Tasks in IIR studies: case study-3

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.
- 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]
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

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

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

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

Satisfaction is an 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 - Intelligemt 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 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

Metrics include
- Engagement: Whether suggested skill used
- Precision-recall: Suggesting used skills (offline engagement proxy)

Learn to rank current skill + previous skills

Engagement: Whether suggested skill used

Precision-recall: Suggesting used skills (offline engagement proxy)

Baseline (always suggest)
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!
Task Management Personas

Based on analysis of Wunderlist customer feedback and interviews:

**Collector**

"I want to save a list of things for later."

Example: You walk around the food hall and discover a new restaurant that you would like to try the next time you are in town.

**Event planner**

"I need to organize and delegate my tasks for decision making."

Example: You are planning a wedding which will happen 6 months from now. The planning, catering, wedding menu, guest list, flowers, etc.

**Tomorrow planner**

"I want to plan ahead so that I can be efficient when I start working on it."

Example: It is Friday afternoon and the weekend is close. Before you go home you would like to prepare a list of things you can work on Monday and new tasks and ideas you didn’t think of this week.

**Daily doer**

"I need to get stuff done. Today and just."

Example: There is a presentation tomorrow and you have a lot of tasks that you needed to work on. You want to have an overview of the next urgent items so that you can plan your day accordingly.

**Analyzer**

"I want to reflect on my behavior and optimize it."

Example: You are reviewing your tasks and events from last week and want to identify areas for improvement and optimize your behavior.
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