CWI
Centrum Wiskunde & Informatica

UCL

User behaviour and task characteristics:

a field study of daily information behaviour

Jiyin He and Emine Yilmaz

March 8, CHIIR 2017, Oslo

Yet another study on tasks and user behaviours

Three types of empirical methods

Observables	Lab studies	Field studies	Log analysis
Tasks	Pre-defined	Often a defined task	Rely on annotation
Task characteristics	By design	Depends on the task	Rely on annotation
Interaction between task characteristics	Difficult to controll	Depends on the task	Rely on annotation
Natural behaviour	No	Yes	Yes

More natural behaviour

Less control, more interpretation

This study

- More natural behaviour
  - ➡ A **field study** of people's **daily** Web searching and browsing activities.
  - ➡ This allows observation of multiple task characteristics and their interactions happening in a natural setting.

- Less interpretation
  - ➡ **Self reported** task and task characteristics annotation.
  - ➡ …because interpreting someone else's search task or intent is difficult (e.g. Russell et al., 2009).

Study procedure

- Pre-study questionnaire
  - Demographics and general habits of information seeking
- A 5-day dairy study
  - Tracking participants' search and browsing activities with a chrome extension
  - Participants review and annotate their own log with task information
- Post-study questionnaire
  - Participants annotate their tasks with task characteristics

Logged information

Event type	Related information	Sub-types
Search events	- query - type of vertical - search engine (G, B, Y)	- search by query - search by vertical switch
Link click events	- anchor text - target URL	- click on SERP - click on a regular page (external or internal link)
Tab events	info about tab operation allowing determining when a user is actually “on” a page	- open-a-new-tab - close-a-tab - switch-to-a-tab - open-link-in-new-tab - tabl-loaded-status
Navigation events	info about how the user arrives on a page	- by link - by direct URL input - by form submission - by forward/backward

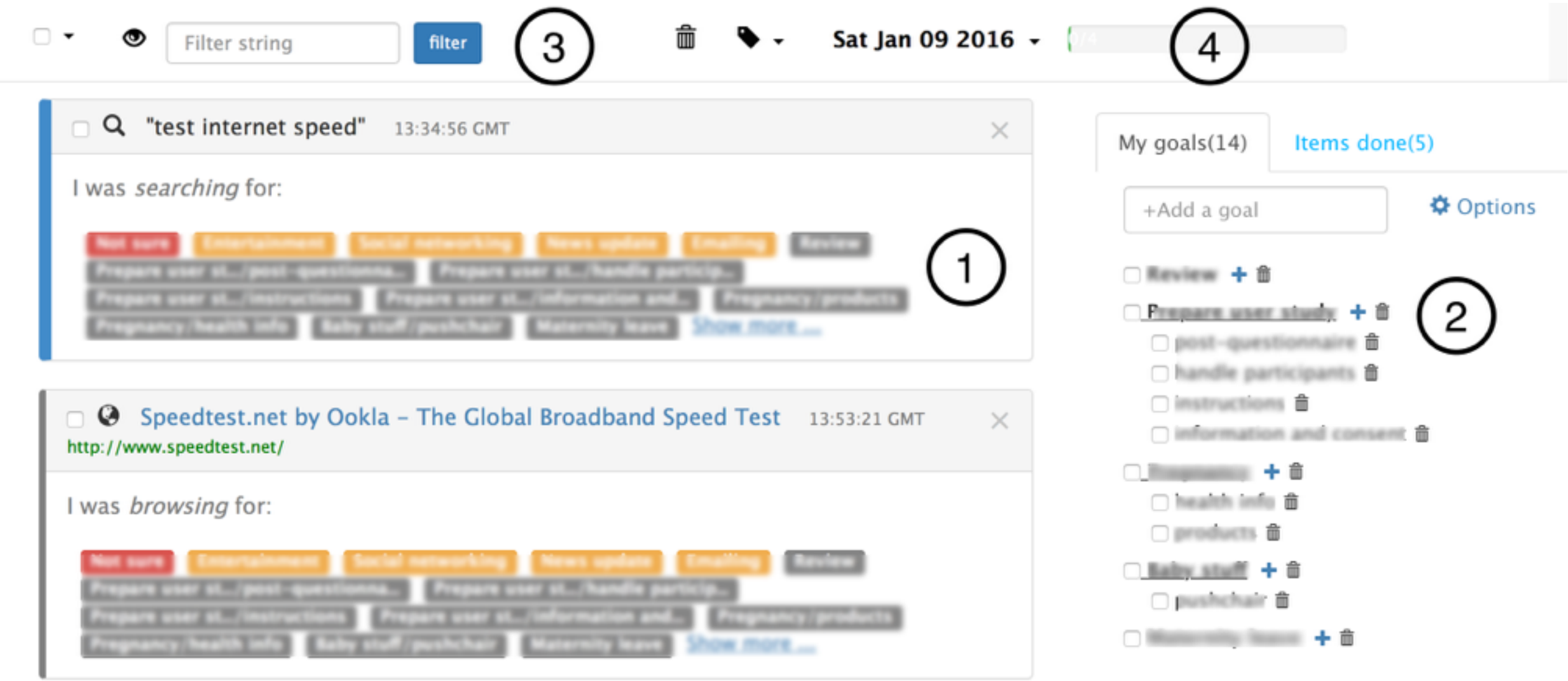
Annotation: tasks

- Daily review of queries issued and pages viewed
  - Remove entries they do not want to share
  - Associate queries/page views with task labels

- Users were encouraged to think of the notion of “tasks” at a level that are typically considered in the literature
  - e.g. “write a report”, “plan a vacation”

- Some general labels were provided
  - Emailing, Social networking, Entertainment, News update, and "Not sure"

Annotation: tasks



Annotation: task characteristics (1)

Task characteristics	**Description**	**Values**
Frequency (FQ)	How frequent would you say the following task have occurred?	(1) One-time task—Routine tasks (5)
Length (TL)	How quickly do you think the following task can be finished?	(1) Very quick (< 1 day)—long term (≥ 1 month) (5)
Stage (STG)	To what extend did you manage to complete the task so far?	(1) Just started—(Almost) finished (5)
Cognitive level (CL)	Different tasks involve cognitive activities of different levels of complexity. At which level would you rate the activities involved to complete the following task?	(1) Remember; (2) Understand; (3) Apply; (4) Analyse; (5) Evaluate; (6) Create.
Collaboration (COL)	To what extend would you say you were responsible for the task?	(1) Solely responsible—Collaborates with many people (5)
Importance (IMP)	How would you rate the importance of the task?	(1) Unimportant—Extremely important (5)

Task characteristics derived and modified from (Li and Belkin 2008)

Annotation: task characteristics (2)

Task characteristics	**Description**	**Values**
Urgency (UR)	How would you rate the urgency of the task?	(1) Not urgent—Extremely urgent (5)
Difficulty (DIF)	How do you feel about the difficulty of the task? (e.g. difficult to find relevant information, or requires great effort in thinking/understanding).	(1) Easy—Extremely difficult (5)
Complexity (COM)	How do you feel about the complexity of the task? (e.g. it may involve many steps or subtasks in order to complete the task).	(1) Simple—Extremely complex (5)
Knowledge of topic (KT)	How would you rate your knowledge on the topic of the task?	(1) No knowledge—Highly knowledgeable (5)
Knowledge of procedure (KP)	How would you rate your knowledge on the procedure to complete the task?	(1) No knowledge—Highly knowledgeable (5)
Satisfaction (SAT)	Were you satisfied with the process of information seeking activities for completing the task?	(1) Unsatisfied —Very satisfied (5)

Task characteristics derived and modified from (Li and Belkin 2008)

Data obtained

- 23 participants
  - 13 males, 10 females (18 - 34 yrs)
  - experience with search engines (md = 5, IQR=1.0)
- 289 user defined tasks
  - 17 with subtasks
  - 135 annotated with task characteristics
- Annotations
  - 2566 queries and 32, 902 page visits annotated
  - 1768 queries and 17, 313 page visits annotated with user defined tasks

User task activities in logs

- This study compared to previous studies in log analysis:
  - Rich interaction types vs. query-only logs
  - Self-annotated vs. externally annotated

RQ1: Whether, and if so how, tasks annotated by users themselves leads to new observations in the scope of tasks and observed statistics in log analysis?

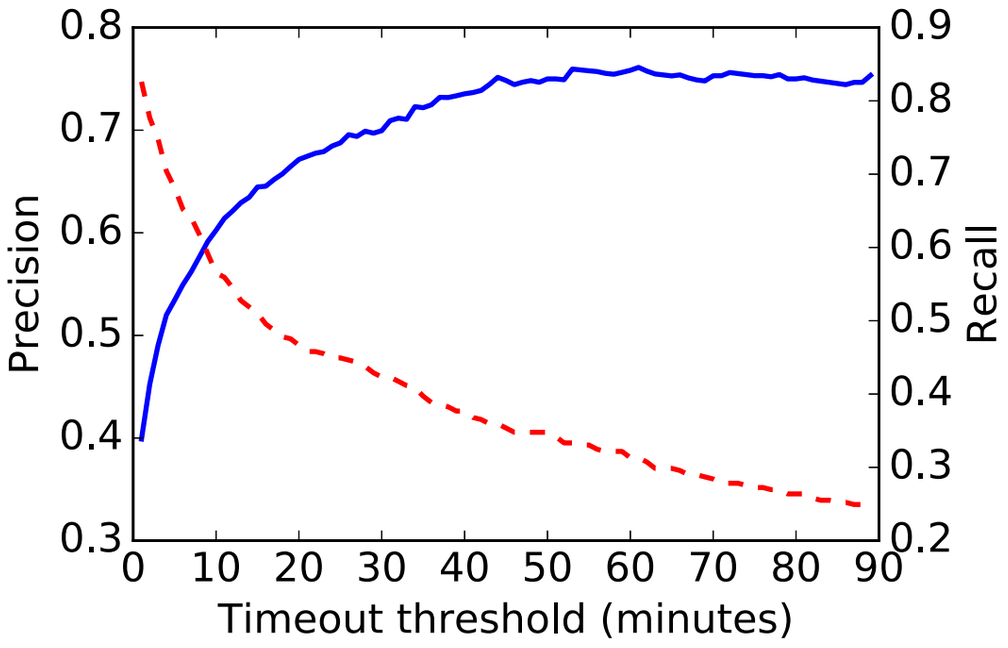
Tasks based log analysis: concepts and terminology

Concept	Physical session	Logical session	(Complex) task
Definition	All user queries or activities within a time window.	Consecutive queries or activities belonging to the same task.	A set of related information needs span over one or more logical sessions.
Terminology			
Jones et al. 2008	Session	Goal	Mission
Lucchese et al. 2011	Time-gap session	Task session	—
Hagen et al 2013	Physical session	Logical session	Mission
This study	Physical session	Logical session	Task

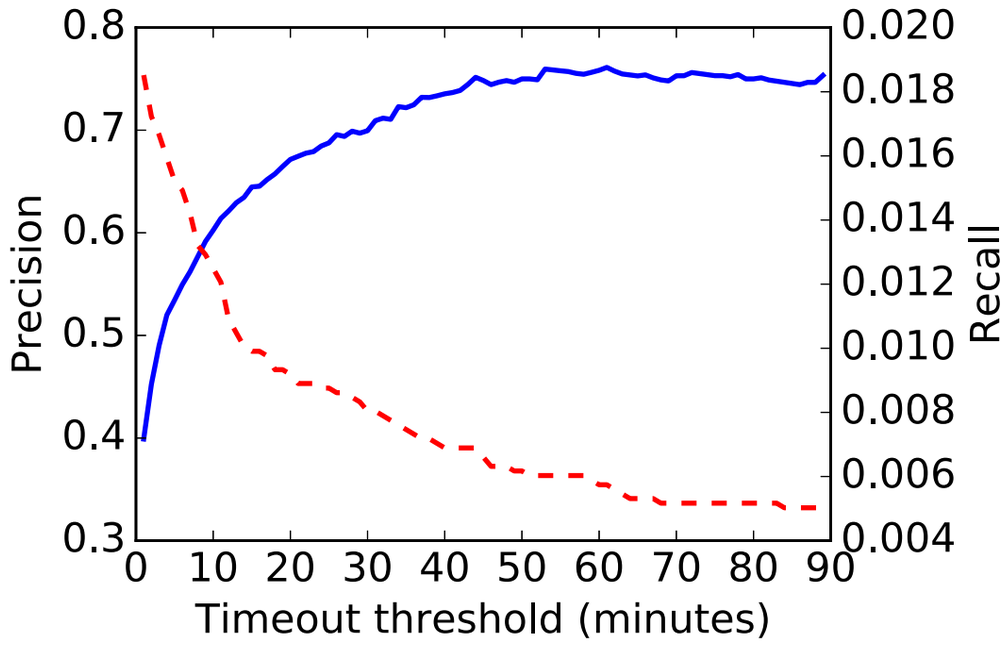
User task activities in logs: key observations

- Zero-query task and sessions:
  - 86% logical sessions; 41% user defined tasks
- Tasks are highly interleaved:
  - On average, 23.9 logical sessions per task; 86% tasks were interrupted and revisited
  - Much higher than reported previously (query-only log)
    - 2.9 logical sessions per task (Hagen et al. 2013)
    - 17% tasks were interrupted (Jones et al., 2008)
  - If only queries are considered, 6.9 logical sessions per task; 68% tasks were interrupted and revisited

User task activities in logs: physical sessions and task boundaries



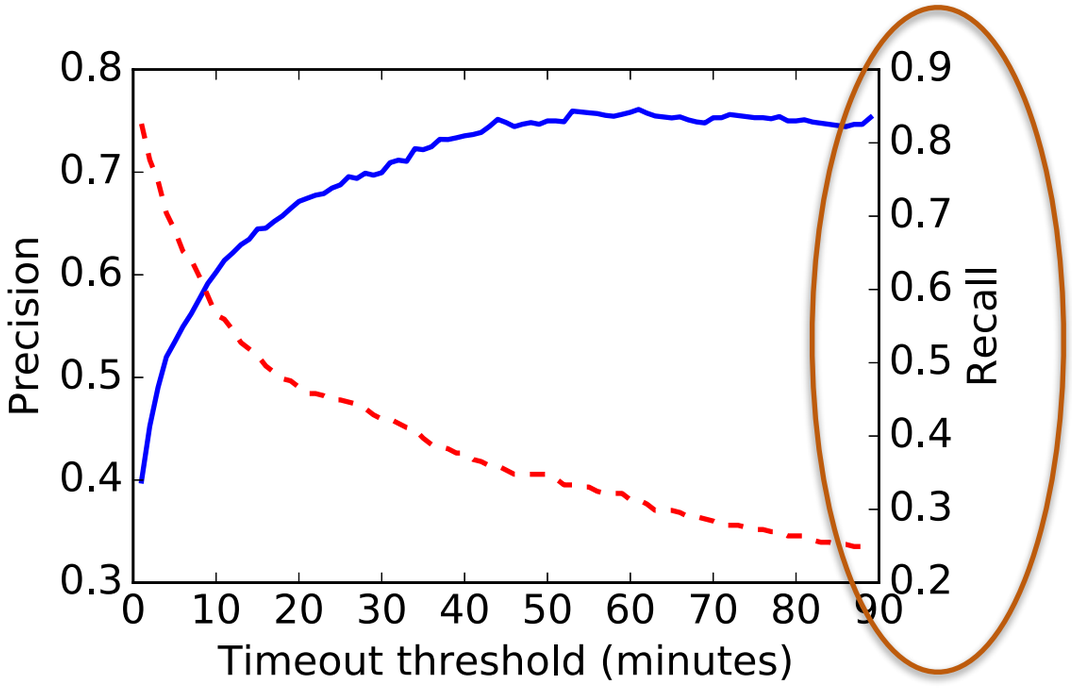
Evaluated on queries only



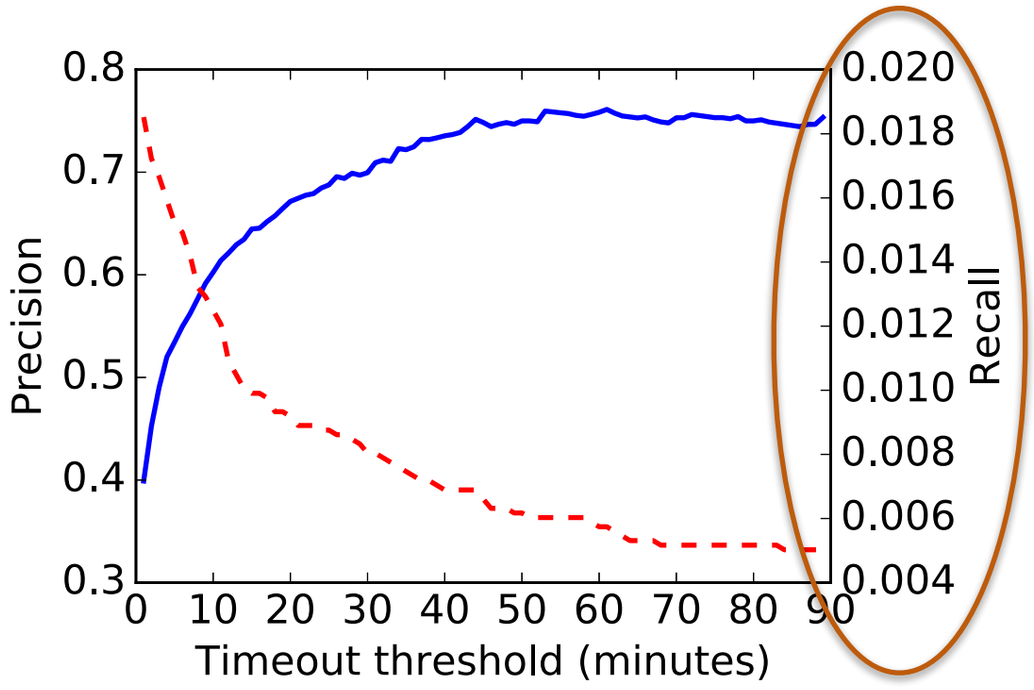
Evaluated on all activities

Using time threshold between *queries* for task detection.

User task activities in logs: physical sessions and task boundaries



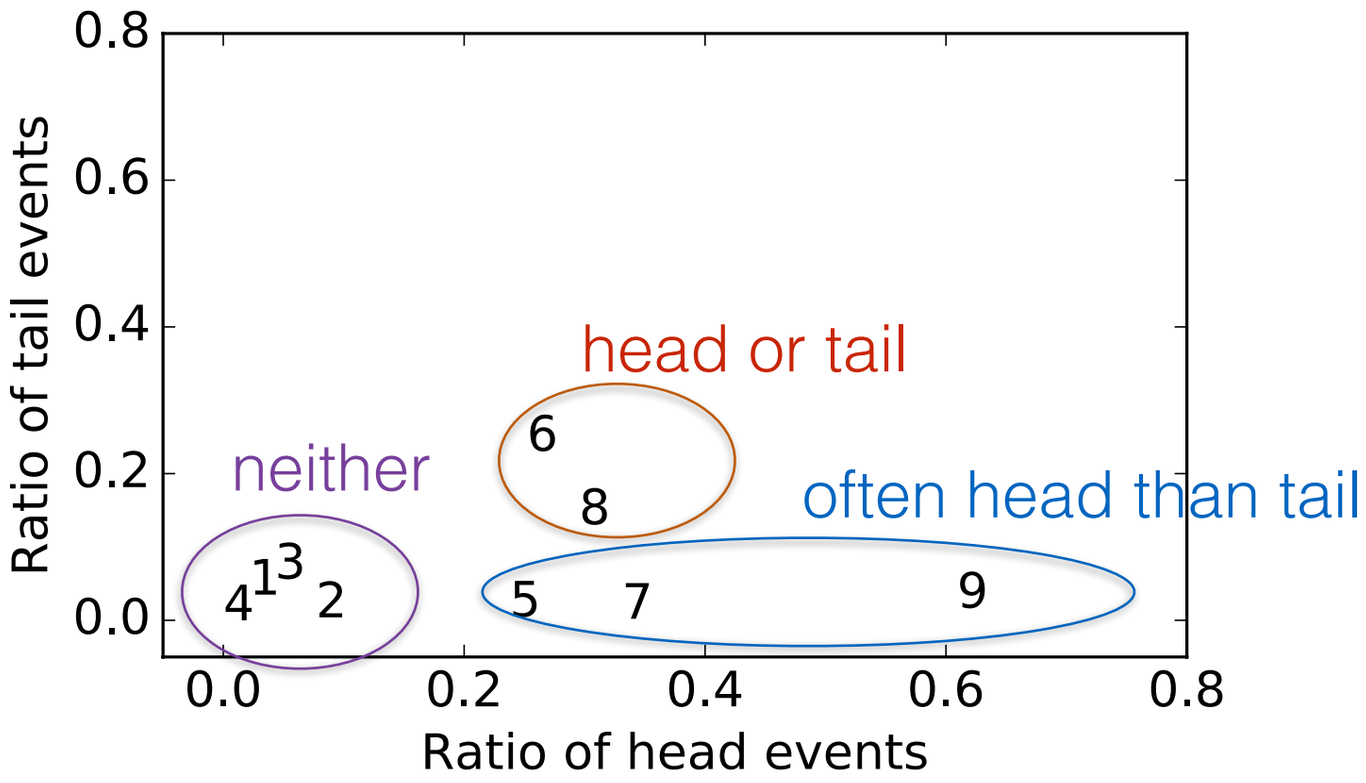
Evaluated on queries only



Evaluated on all activities

There is a majority of task switches happening in between queries that are missed out if we only look at queries to identify task switches.

What types of actions signifies task switch?



ID	Action
1	form submit
2	for/backward
3	link click
4	pagination
5	query
6	tab close
7	tab new
8	tab switch
9	go to URL

User task activities in logs: implications

- This study compared to previous studies in log analysis:
  - Rich interaction types vs. query-only logs
  - Self-annotated vs. externally annotated

RQ1: Whether, and if so how, tasks annotated by users themselves leads to new observations in the scope of tasks and observed statistics in log analysis?

➡ A fair amount of tasks or task sessions do not involve search;

➡ Query-only logs miss those browse/navigation-only task activities, as well as task switches.

Task characteristics and user activities

- The self-reported annotation of tasks and task characteristics allow us to observe
  - interactions between task characteristics
  - examples of tasks in which these characteristics naturally occur

RQ2: how do task characteristics relate to each other and how do these characteristics co-occur within actual Web user tasks?

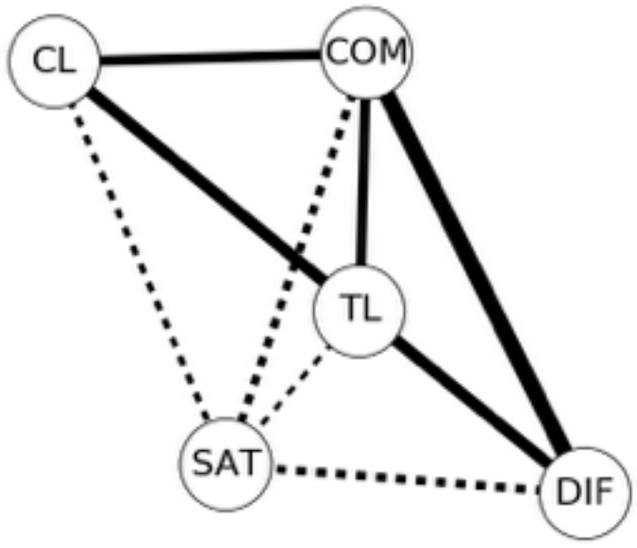
How do task characteristics relate to each other: method

- A correlation analysis on task characteristics
  - 12 task characteristics
  - 135 user annotated tasks from post-study questionnaire
  - Measure: Kendall's τ
- Clustering of characteristics using correlation as similarity measure
  - To discover groups of mutually correlated characteristics
  - Clustering method: Affinity Propagation (Frey and Dueck 2007)

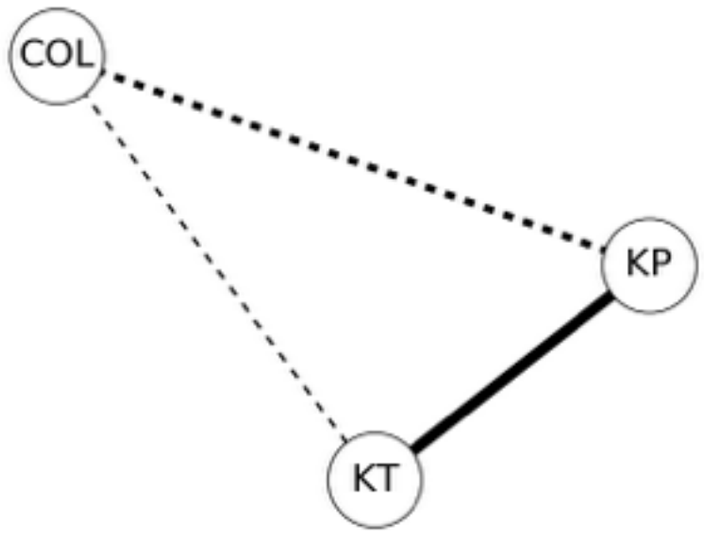
Groups discovered

Group	Members
1	cognitive complexity level (CL) task complexity (COM) task difficulty (DIF) task length (TL) task satisfaction (SAT)
2	collaboration (COL) knowledge of topic (KT) knowledge of procedure (KP)
3	importance (IMP) task stage (STG) task urgency (UG)
4	task frequency (FQ)

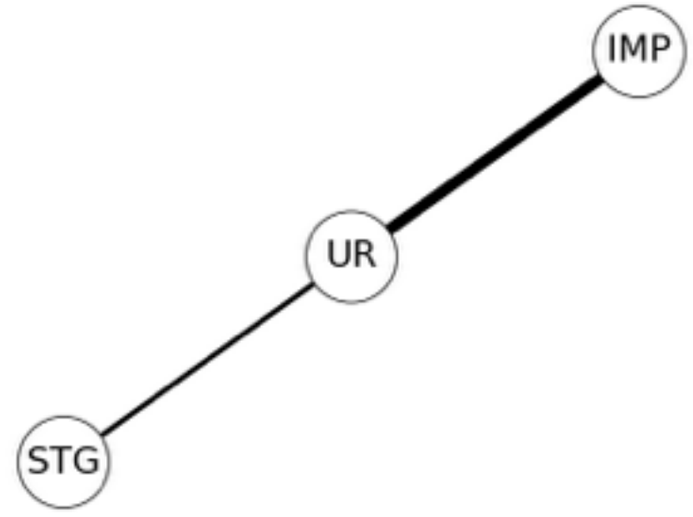
Interaction between task characteristics: within groups



Group 1



Group 2

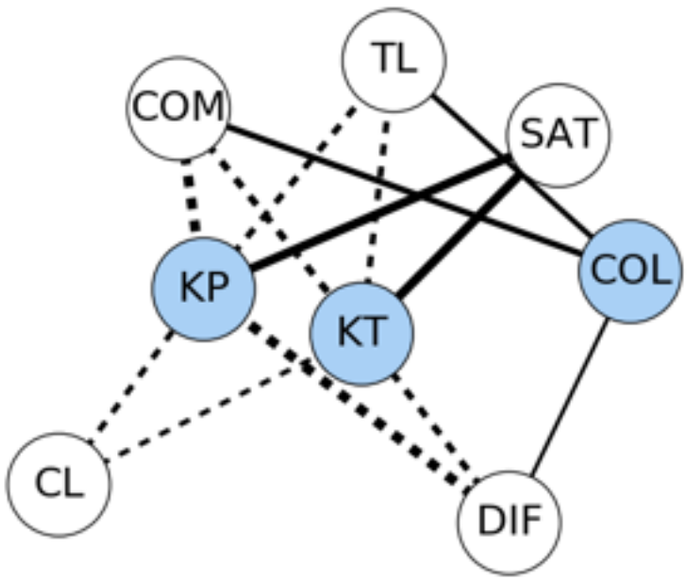


Group 3

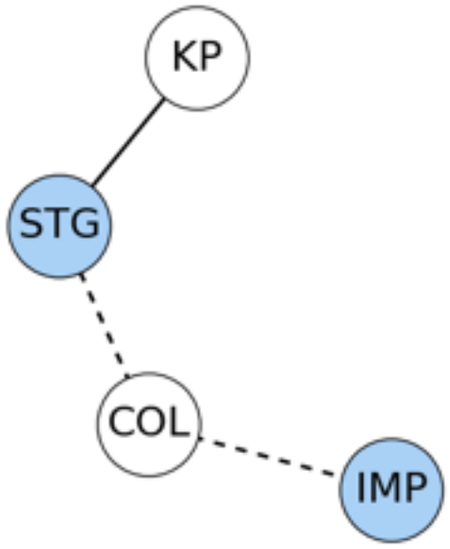
- - Negative correlation

— Positive correlation

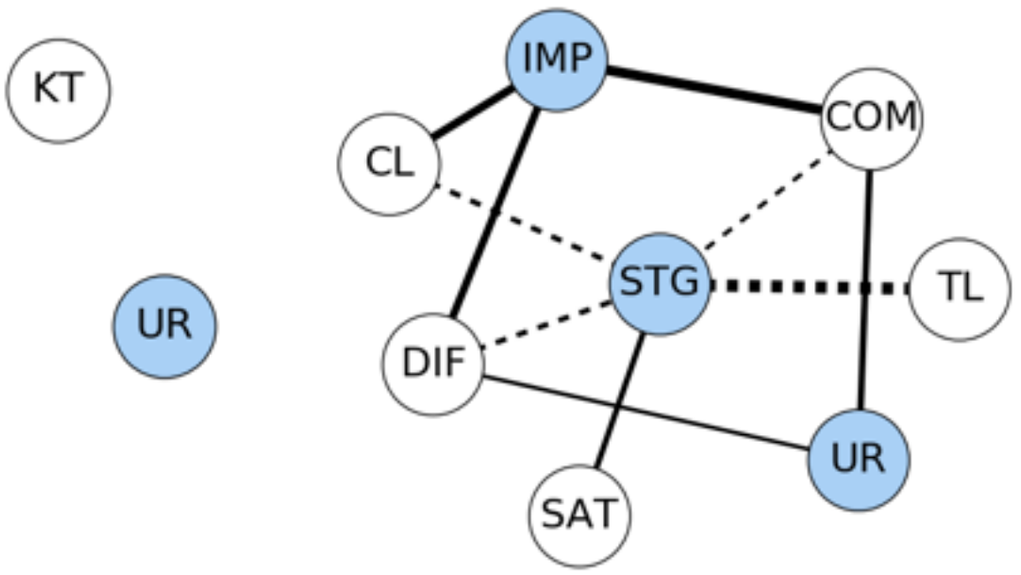
Interaction between task characteristics: between groups



Group 1 - 2



Group 2 - 3



Group 1 - 3

- - Negative correlation

— Positive correlation

How do these characteristics co-occur within actual Web user tasks? Method

- Task abstraction
  - Tasks are aggregated into topics (only for obvious cases to avoid over-interpretation), e.g. look for jobs
  - Task examples are anonymised by masking the identifiable information in the task description with “X”.
- A qualitative case study with cognitive complexity
  - How popular topics are distributed over different cognitive complexity levels?
  - What are typical topics at each level?

Task characteristics in naturalistic user tasks: a case with cognitive complexity

Topic	Remember	Understand	Apply	Analyse	Evaluate	Create	Tot
Shopping	**10 (56%)** “Amazon-Heater”	—	2 (11%) “sort out X”	3 (17%) “baby products”	3 (17%) “buy contact lenses”	—	18 (13%)
Writing	1 (9%) “compile X paper	—	2 (18%) “Complete X tutorial	—	**4 (36%)** “X Essay”	**4 (36%)** “X paper”	11 (8%)
Travel	**3 (30%)** “weekend travel”	1 (10%) “X trip”	1 (10%) “Book trip to X”	1 (10%) “Flight home”	2 (20%) “book tickets for X ”	2 (20%) “Plan trip X”	10 (7%)
Job	1 (14%) “Look for jobs”	—	1 (14%) “Tutor jobs”	1 (14%) “Internship apply”	**3 (43%)** “job hunt”	1 (14%) “Find job”	7 (5%)
Project	—	1 (17%) “Project management”	**2 (33%)** “X project”	1 (17%) “X proj”	**2 (33%)** “research project-X	—	6 (4%)
Research	—	—	**3 (50%)** “Research”	1 (17%) “...research for X”	1 (17%) “X research”	1 (17%) “X study”	6 (4%)
Program-ming	—	1 (20%) “test X”	—	**3 (60%)** “port X to java”	1 (20%) “...interface for X”	—	5 (3%)
Watch X	2 (40%) “streaming”	—	—	**3 (60%)** “Binge watch X”	—	—	5 (3%)
Other	21 “check location X”	10 “stock”	17 “Find solutions to X”	5 “learn about X”	9 ““buy flat”	5 ““study X”	67 (49%)
Total	38	13	28	18	25	13	135

Task characteristics in naturalistic user tasks: a case with cognitive complexity

Topic	Remember	Understand	Apply	Analyse	Evaluate	Create	Tot
Shopping	**10 (56%)** “Amazon-Heater”	—	2 (11%) “sort out X”	3 (17%) “baby products”	3 (17%) “buy contact lenses”	—	18 (13%)
Writing	1 (9%) “compile X paper	—	2 (18%) “Complete X tutorial	—	**4 (36%)** “X Essay”	**4 (36%)** “X paper”	11 (8%)
Travel	**3 (30%)** “weekend travel”	1 (10%) “X trip”	1 (10%) “Book trip to X”	1 (10%) “Flight home”	2 (20%) “book tickets for X ”	2 (20%) “Plan trip X”	10 (7%)
Job	1 (14%) “Look for jobs”	—	1 (14%) “Tutor jobs”	1 (14%) “Internship apply”	**3 (43%)** “job hunt”	1 (14%) “Find job”	7 (5%)
Project	—	1 (17%) “Project management”	**2 (33%)** “X project”	1 (17%) “X proj”	**2 (33%)** “research project-X	—	6 (4%)
Research	—	—	**3 (50%)** “Research”	1 (17%) “...research for X”	1 (17%) “X research”	1 (17%) “X study”	6 (4%)
Program-ming		1 (20%)		**3 (60%)**	1 (20%)		5
Watch X	“s						5
Other	“che						67 (9%)
Total							135

The same topic can span over multiple cognitive complexity levels

➡ *When people describe their tasks, although sometimes it seems that they are doing the same thing, the actual intention and activities involved can be very different.*

Task characteristics in naturalistic user tasks: a case with cognitive complexity

Topic	Remember	Understand	Apply	Analyse	Evaluate	Create	Tot
Shopping	**10 (56%)** “Amazon-Heater”	—	2 (11%) “sort out X”	3 (17%) “baby products”	3 (17%) “buy contact lenses”	—	18 (13%)
Writing	1 (9%) “compile X paper	—	2 (18%) “Complete X tutorial	—	**4 (36%)** “X Essay”	**4 (36%)** “X paper”	11 (8%)
Travel	**3 (30%)** “weekend travel”	1 (10%) “X trip”	1 (10%) “Book trip to X”	1 (10%) “Flight home”	2 (20%) “book tickets for X ”	2 (20%) “Plan trip X”	10 (7%)
Job	1 (14%) “Look for jobs”	—	1 (14%) “Tutor jobs”	1 (14%) “Internship apply”	**3 (43%)** “job hunt”	1 (14%) “Find job”	7 (5%)
Project	—	1 (17%) “Project management”	**2 (33%)** “X project”	1 (17%) “X proj”	**2 (33%)** “research project-X	—	6 (4%)
Research	—	—	**3 (50%)** “Research”	1 (17%) “...research for X”	1 (17%) “X research”	1 (17%) “X study”	6 (4%)
Program-ming	—	1 (20%) “test X”	—	**3 (60%)** “port X to java”	1 (20%) “...interface for X”	—	5 (3%)
Watch X	“						5 3%)
Other	“che						67 49%)
Total							135

The different cognitive complexity levels are not evenly distributed across task topics

➡ *Some task topics are more likely to involve certain levels of cognitive complexity than others*

Task characteristics and user activities: implications

- The discovery of groups of mutually correlated task characteristics has implications for task designs for lab studies.
  - ➡ e.g., task collaboration is seen related to complex/difficult tasks, implying that studies of complex/difficult tasks may need to consider collaboration as an additional variable.

- Tasks that share similar descriptions can vary greatly in their characteristics (as perceived by the user him/herself).
  - ➡ It would be difficult for external annotators to interpret/classify user tasks and their characteristics
  - ➡ To support users with their tasks, we need to know not only what task the user is engaged with, but also what status the task is in, as different types of supports may be needed