User Interactions with Search Systems

Ying-Hsang Liu 1 2  Chang Liu 3  Ralf Bierig 4

1 Charles Sturt University, Australia
2 The Australian National University, Australia
3 Peking University, China
4 Maynooth University, Ireland

The 4th ACM SIGIR International Conference on the Theory of Information Retrieval, ICTIR 2018, September 14, 2018
1. Overview of IIR Studies

2. Experimental User Studies and its Challenges
1. To understand the importance and the added benefits of conducting user studies;

2. To be able to design user experiments for exploring and modelling user search behaviour;

3. To be able to implement user experiments with confidence in collaborative research settings.
2. Overview of IIR Studies

Definitions of IIR

IIR

- "characterizing the study of the interactions of people, information and information retrieval systems" (Cool & Belkin, 2011, p. 4)
- "IIR itself is shaped by (a) research on information seeking and search behavior and (b) ...new methods of interacting with electronic resources" (Ruthven, 2008, p. 44)
A nested model of the information seeking and information searching (Wilson, 1999, p. 263)
2. Overview of IIR Studies

Overview

Users lost

- “... [U]ser issues, the human factors, are addressed right from the beginning to the very end, right from theory, conceptualization, and design process on to development, evaluation, and to provision of services.” (Saracevic, 1997b, p. 26)

HCIR to CHIIR

- Human-Computer Interaction and Information Retrieval ”... to empower people to explore large-scale information bases... expending cognitive and physical energy.” (Marchionini, 2006, p. 20)
Factors affecting information behaviour (Ford, 2015, p. 263)

Figure 6.1 Factors that can influence information behaviour
2. Overview of IIR Studies

Models of IR Interactions

Stratified model (Saracevic, 1997a)
2. Overview of IIR Studies

Models of IR Interactions

**Episode model (N. J. Belkin, 1996)**

Figure 1: Information retrieval as support for information interaction
2. Overview of IIR Studies

Models of IR Interactions

Cognitive model (Ingwersen, 1996, p. 8)

FIGURE 2. Cognitive model of IR interaction. Extension of Ingwersen [3, p. 16]
2. Overview of IIR Studies

Models of IR Interactions

Cognitive framework (Ingwersen & Jarvelin, 2005, p. 274)
## Information foraging (Pirolli, 2007, p. 18)

<table>
<thead>
<tr>
<th>Level</th>
<th>Question</th>
<th>Stance</th>
<th>Analysis Elements</th>
<th>Examples</th>
</tr>
</thead>
</table>
| Rational     | What environmental problem is solved? Why is this solution a good one?    | Design  | • States, resources, state dynamics  
• Constraints, affordances  
• Feasible strategies  
• Optimization criteria | • Optimal foraging theory  
• Information Foraging Theory |
| Knowledge    | What does the system know?                                               | Intentional | • Environment  
• Goals, preferences  
• Knowledge  
• Perception, action | • Knowledge-level analysis |
| Cognitive    | How does the system do it?                                               | Information processing | • Cognitive states  
• Cognitive processes | • ACT-R  
• Soar |
| Biological   | How does the system physically do it?                                    | Biophysical | • Neural process | • Neural models |
2. Overview of IIR Studies

Models of IR interactions

Common Interaction Model (CIM) (Pääkkönen et al., 2017, p. 341)
2. Overview of IIR Studies

Models of IR interactions

Interface Card Model (Zhang & Zhai, 2016)

- Modeling and optimizing interactive retrieval interfaces
- Formal modeling of user states and stopping actions
- Use of reinforcement learning
- Interface layout adapted to inferred user stopping tendencies
2. Overview of IIR Studies

Models of IR interactions

*Probability ranking principle for IIR; Assumptions (Fuhr, 2008)*

- Focus on the functional level of interaction
- Decisions are the major interaction activity
- Users evaluate choices in linear order
- Only positive, correct decisions are of benefit for a user (Fuhr, 2008, pp. 253–254)
Learning Activity 1

What are the similarities and differences among the user models of IR interactions? Apply one of the user-centred information search models to the information seeking/retrieval process. Discuss the applicability of the model to real life information seeking/search.
2. Overview of IIR Studies

Examples of IIR studies

*Experimental methodology, participants and variants (White, 2016a, p. 338)*

<table>
<thead>
<tr>
<th>Experimental Methodology</th>
<th>Typical Number of Participants</th>
<th>Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laboratory studies (rich data)</td>
<td>10–100s</td>
<td>In-laboratory behavior</td>
</tr>
<tr>
<td>Instrumented panels</td>
<td>100–1,000s</td>
<td>Ethnography</td>
</tr>
<tr>
<td>Retrospective log analysis</td>
<td>1,000s–millions</td>
<td>Logs from one searcher</td>
</tr>
</tbody>
</table>
2. Overview of IIR Studies

Examples of IIR studies

**CrowdLogger** *(Feild & Allan, 2013)*

- "... a rich API for tracking user behavior, interacting with web pages and users, communicating with servers, and uploading data privately." *(Feild & Allan, 2013, p. 17)*
- Similar to instrumented panels *(White, 2016a, p. 338)*

[Link]
Examples of IIR studies

Measuring information seeking contexts (Kelly, 2006)

- Naturalistic, longitudinal research design
- Information seeking contexts: User self-identified search tasks and topics
- 7 participants, laptop with client-side logger, 14 weeks
- Similar to observational (in)-laboratory behavior (White, 2016a, p. 338)
Examples of IIR studies

**Search roles (Foss & Druin, 2014)**

- Parent interviews; Child interviews when he/she uses the home computer
- Thematic analysis using grounded theory
- Similar to observational (in)-laboratory behavior (White, 2016a, p. 338)
Examples of IIR studies

Information need formation and search tool use (Saastamoinen & Järvelin, 2016)

- How work task factors affect search
- Mixed-method approach
- Similar to ethnography (White, 2016a, p. 338)

Data collection techniques

1. Start
2. Initial interview
3. Actual data collection:
   - Daily: Log, video, + 1 day observation
   - 2 questionnaires
4. Data collection duration:
   - Companies, universities: ~ a week
   - City: 1 day
5. Exit interview
6. End
Learning Activity 2

Based on your understanding of experimental methodologies, identify the methods used and the proposed research questions in the following CHIIR ’17 studies:

- User Interests in German Social Science Literature Search: A Large Scale Log Analysis (Hienert, 2017)

To what extent the proposed research questions can be answered by the research methods? Can we use alternative methods to address the proposed research questions?
3. Experimental User Studies and its Challenges

Experimental User Studies and its Challenges

**History of evaluation experiments in IR (S. Robertson, 2008)**

- Cranfield experiments (1958–1966)
- Some experiments: SMART, MEDLARS, Sp’arck Jones, Keen, Belkin and Oddy, Okapi, Croft
- TREC (Text REtrieval Conference): Ideal test collection
- TREC-like methods: CLEF (Cross-Language Evaluation Forum), NTCIR (NII Test Collection for IR Systems), INEX (Initiative for Evaluation of XML)

**Library card catalogue**
Types of IR experiment

- **Black-box**: System as a whole, input and output
- **Diagnostic**: Internal structures, components and activities
- **Operational**: Operational environments
- "Given that in the past most experiments have been essentially black-box and laboratory based, our conclusion ... is that more emphasis should be placed on diagnostic and/or operational experiments.”
  (S. E. Robertson & Hancock-Beaulieu, 1992, p. 464)
**Which type?**

- **Cranfield**: Compare the efficiency of four indexing systems (Cleverdon, 1962)
- **TREC-1** "ad hoc" retrieval (retrospective searching) (Harman, 1993)
- **Simulation**: Objectives: "how well an IR system performs, and how performance changes under different conditions and behaviours." (Azzopardi, 2016, p. 1227)
- ”IR is best construed as a sequence of interactions of the person with the IR system” (N. J. Belkin, Hienert, Mayr-Schlegel, & Shah, 2017)
### Examples of types of IR experiment

<table>
<thead>
<tr>
<th>Studies</th>
<th>Objectives</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Edwards &amp; Kelly, 2017, p. 125)</td>
<td>&quot;Examines the differences in the search behaviors and physiologies of people who are engaged or frustrated during search.&quot;</td>
<td>Diagnostic</td>
</tr>
<tr>
<td>(Baskaya, Keskustalo, &amp; Järvelin, 2013, p. 2297)</td>
<td>&quot;How effective is human behavior employing search strategies compared to various baselines under various search goals and time constraints.&quot;</td>
<td>Black-box/Diagnostic</td>
</tr>
<tr>
<td>(Salton, 1972, p. 75)</td>
<td>&quot;To compare conventional retrieval (MEDLARS) and automatic text analysis methods (SMART).”</td>
<td>Black-box</td>
</tr>
<tr>
<td>(Blair &amp; Maron, 1985, p. 289)</td>
<td>&quot;Aimed at evaluating the effectiveness of full-text retrieval... examined IBM’s full-text retrieval system, STAIRS.&quot;</td>
<td>Operational</td>
</tr>
</tbody>
</table>
Learning Activity 3

What are the objectives of your study? What type of IR experiment will you do to address the objectives? Explain the relationship between the objectives of the study and the type of IR experiment.
IR Experiment Decisions

**Pragmatics of IR experimentation (Tague-Sutcliffe, 1992)**

- What kind of test? (from experiment to operational test)
- How to operationalise the variables?
- What database to use? Where to get queries?
- How to process queries?
- How will treatments be assigned to experimental units? (experimental design)
- How to collect and analyse the data?
3. Experimental User Studies and its Challenges

Operationalisation of searcher characteristics

**MeSH (Medical Subject Headings) search effectiveness**

- Usefulness of MeSH terms for different kinds of users
- Searcher characteristics of domain knowledge and search training
- Four types of searchers


<table>
<thead>
<tr>
<th>Searcher type</th>
<th>Domain knowledge</th>
<th>Search training</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Novice (SN)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Domain Expert (DE)</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Search Expert (SE)</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Medical Librarian (ML)</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

*Note. Plus (+) and minus (−) indicate the high-level and low-level of the specified searcher characteristics respectively.*
Search expertise (Bailey & Kelly, 2016, p. 237)

Domain expertise (White, Dumais, & Teevan, 2009, p. 138)
Operationalisation of searcher characteristics

**Instruments used in individual differences studies (O’Brien, Dickinson, & Askin, 2017, p. 247)**

<table>
<thead>
<tr>
<th>Individual difference</th>
<th>Example of instruments</th>
<th>No. instruments per category</th>
<th>No. articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning, thinking and cognitive style</td>
<td>Cognitive Styles Analysis [17]</td>
<td>23</td>
<td>56</td>
</tr>
<tr>
<td>Personality</td>
<td>NEO Five-factor Inventory (NEO-FFI) [9]</td>
<td>7</td>
<td>20</td>
</tr>
<tr>
<td>Attitudes</td>
<td>Internet Perceptions Questionnaire [1]</td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>

Note: Counts in square brackets indicate the number of articles that employed the given instrument.
Experimental components for evaluation of IIR

**Simulated work task (Borlund, 2000, p. 76)**

- The involvement of potential users as test persons;
- The application of dynamic and individual information needs (real and simulated information needs);
- the use of multidimensional and dynamic relevance judgements.
Simulated work task (Borlund & Schneider, 2010, p. 156)

**Simulated situation:**

**Simulated work task situation:** After your graduation you will be looking for a job in industry. You want information to help you focus your future job seeking. You know it pays to know the market. You would like to find some information about employment patterns in industry and what kind of qualifications employers will be looking for from future employees.

**Indicative request:** Find, for instance, something about future employment trends in industry, i.e., areas of growth and decline.
### Experimental vs. operational database

- Test collections: Cranfield, Medlars, Commun ACM, ERIC, Reuters, TREC and TREC-like (Sanderson, 2010; Tague-Sutcliffe, 1992)
- TREC Interactive Track (Dumais & Belkin, 2005)
- Use of test collections in IIR (Y.-H. Liu, Thomas, Bacic, Gedeon, & Li, 2017; Y.-H. Liu & Wacholder, 2017)
Determination of experiment structure (Lazar, Feng, & Hochheiser, 2017, p. 48)

- Design study
  - Number of independent variables > 1?
    - No
      - Basic design
        - Determine number of conditions
          - Between group
          - Within group
    - Yes
      - Factorial design
        - Determine number of conditions
          - Between group
          - Within group
          - Split-plot
### Assignment of experimental conditions

- “A plan for assigning subjects to experimental conditions and the statistical analysis associated with the plan.” (Kirk, 2013, p. 2)
- Within-subject design vs. between-subject design (Hofmann, Li, & Radlinski, 2016; Hornbæk, 2013; Kelly, 2009)
- Sample size and power analysis (Champely, 2015; Faul, Erdfelder, Lang, & Buchner, 2007; Kelly, 2015)
- Statistical testing for IR experiments (Hull, 1993)
- Is it statistically significant? (Carterette, 2017)
Experimental Design

Example of factorial design

- 4 × 2 × 2 Factorial design; 4 interfaces, controlled search topic pairs and cognitive styles (Y.-H. Liu et al., 2017)

- 4 × 4 Graeco-Latin Square to arrange experimental conditions

- Power Analysis for ANOVA Design; medium effect size of .25, $\alpha < .05$ and $N = 256$, statistical power of .93 (Cohen, 1988; Faul et al., 2007)

4 × 4 Graeco-Latin Square

(de Mendiburu, 2017; R Core Team, 2017)

<table>
<thead>
<tr>
<th>Interfaces:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Google</td>
</tr>
<tr>
<td>B: Per Query, ProQuest</td>
</tr>
<tr>
<td>C: Per Query, ProQuest+EBSCO</td>
</tr>
<tr>
<td>D: Per Doc, EBSCO</td>
</tr>
</tbody>
</table>

Search topic pairs:
1. (105, 16)
2. (58, 78)
3. (9, 47)
4. (94, 4)

```
[1,]  "D"  "3"  "A"  "2"  "C"  "4"  "B"  "1"
[2,]  "A"  "4"  "D"  "1"  "B"  "3"  "C"  "2"
[3,]  "C"  "1"  "B"  "4"  "D"  "2"  "A"  "3"
[4,]  "B"  "2"  "C"  "3"  "A"  "1"  "D"  "4"
```
Experimental Procedure

**Experimental procedure (Y.-H. Liu et al., 2017; Wittek, Liu, Darányi, Gedeon, & Lim, 2016)**

- **Data collection**
  - User characteristics (background questionnaire and cognitive style test)
  - User perceptions (exit questionnaire)
  - Search behaviours (search logs, mouse clicks and documents saved)
  - Physiological signals (eye gaze and EEG)
Facet classification of tasks (Li & Belkin, 2008, pp. 1834–1835)

<table>
<thead>
<tr>
<th>Li &amp; Belkin (2008) Facet Analysis of Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source of Task</strong></td>
</tr>
<tr>
<td>– Self, Group, Assigned</td>
</tr>
<tr>
<td><strong>Task Doer</strong></td>
</tr>
<tr>
<td>– Individual, Group</td>
</tr>
<tr>
<td><strong>Time</strong></td>
</tr>
<tr>
<td>– Frequency</td>
</tr>
<tr>
<td>– Length</td>
</tr>
<tr>
<td>– Stage</td>
</tr>
<tr>
<td><strong>Product</strong></td>
</tr>
<tr>
<td>– Physical, Intellectual, Decision, Factual</td>
</tr>
<tr>
<td><strong>Process</strong></td>
</tr>
<tr>
<td>– One-time, Multiple</td>
</tr>
<tr>
<td><strong>Items</strong></td>
</tr>
<tr>
<td>– Named or Not</td>
</tr>
<tr>
<td>– Whole or Part</td>
</tr>
<tr>
<td><strong>Goal</strong></td>
</tr>
<tr>
<td>– Quality</td>
</tr>
<tr>
<td>• Specific, Amorphous, Mixed</td>
</tr>
<tr>
<td>– Quantity</td>
</tr>
<tr>
<td>• Single or multiple goals</td>
</tr>
<tr>
<td><strong>Common attributes of task, e.g.</strong></td>
</tr>
<tr>
<td>– Objective/Subjective task complexity, Urgency, Salience, Difficulty, ...</td>
</tr>
</tbody>
</table>
### Classifying types of search tasks (Toms, 2011, p. 57)

<table>
<thead>
<tr>
<th>Authors</th>
<th>Variation in search task classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meadow (1992)</td>
<td>Known item; specific information need; general information need; explore a database</td>
</tr>
<tr>
<td>Marchionini (1989)</td>
<td>Closed; open</td>
</tr>
<tr>
<td>Qiu (1993)</td>
<td>General; specific</td>
</tr>
<tr>
<td>Choo (2001)</td>
<td>Undirected viewing; conditioned viewing; informal search; formal search</td>
</tr>
<tr>
<td>Thatcher (2008)</td>
<td>Directed; general purpose</td>
</tr>
<tr>
<td>J. Kim (2006)</td>
<td>Factual; interpretive; exploratory</td>
</tr>
<tr>
<td>Broder (2002)</td>
<td>Information; transaction; navigation</td>
</tr>
<tr>
<td>Morville (2005)</td>
<td>Sample (few good items); existence (search known item); exhaustive (full recall)</td>
</tr>
<tr>
<td>Kellar, Watters and Shepherd (2007)</td>
<td>Fact finding; information gathering; browsing; transactions; other</td>
</tr>
<tr>
<td>Toms et al. (2008)</td>
<td>Decision making; fact finding; information gathering</td>
</tr>
</tbody>
</table>
Search task complexity (Ingwersen & Jarvelin, 2005, p. 295)
## Experimental User Studies and its Challenges

### Search tasks

**Exploratory search** (Wildemuth & Freund, 2012)

<table>
<thead>
<tr>
<th>Categories</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td>Learning and investigation as goals</td>
</tr>
<tr>
<td></td>
<td>General problem, not specific</td>
</tr>
<tr>
<td></td>
<td>Uncertainty is involved</td>
</tr>
<tr>
<td></td>
<td>Ill-structured problem</td>
</tr>
<tr>
<td></td>
<td>Dynamic, evolution during search</td>
</tr>
<tr>
<td></td>
<td>Multi-faceted</td>
</tr>
<tr>
<td></td>
<td>“Not too easy”</td>
</tr>
<tr>
<td></td>
<td>Accompanied by sensemaking, decision making or other cognition</td>
</tr>
<tr>
<td>Behavioral</td>
<td>Open-ended problem</td>
</tr>
<tr>
<td></td>
<td>Target is multiple items</td>
</tr>
<tr>
<td></td>
<td>Occurs over time</td>
</tr>
</tbody>
</table>

Table 1: Summary of exploratory search task attributes
Search tasks

Exploratory search (White, 2016b, p. 130)

Figure 4.8. Exploratory search activities (adapted from Marchionini, 2006a).
### Examples of search tasks (Athukorala, Głowacka, Jacucci, Oulasvirta, & Vreeken, 2016, p. 2643)

<table>
<thead>
<tr>
<th>Task (Abbrev.)</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge acquisition (Know)</td>
<td>You are going to start a new research project on the topic <em>Reinforcement learning</em> (or <em>Active learning</em>). You would like to learn as much information as possible about this topic, e.g. applications, problems, specific algorithms</td>
</tr>
<tr>
<td>Planning (Plan)</td>
<td>You are planning to give a talk on the topic <em>Deep neural networks</em> (or <em>Clustering techniques</em>). Plan the structure of your presentation, including short titles of the headings of your slides and using bullet points describe the content</td>
</tr>
<tr>
<td>Comparison (Comp)</td>
<td>Collect literature to write a short essay describing similarities and differences between <em>Supervised learning and Unsupervised learning</em> (or <em>Transfer learning and Multitask learning</em>).</td>
</tr>
<tr>
<td>Fact-finding (Fact)</td>
<td>Define the term SVM (or UCB) as in the first article that proposed it.</td>
</tr>
<tr>
<td>Navigation (Navi)</td>
<td>Navigate to the article that presents the most commonly used topic model–latent Dirichlet allocation—for the first time (or Navigate to the article that solves the—multi-armed bandit—problem for the first time.)</td>
</tr>
<tr>
<td>Question answering (Question)</td>
<td>What are the most common sampling methods used in machine learning? (List three) (or What are the kernels used in machine learning? (List three))</td>
</tr>
</tbody>
</table>
Examples of Search Tasks


- **ID:** 39
- **Title:** Hypertension
- **Need:** Identify genes as potential genetic risk factors candidates for causing hypertension.
- **Context:** A relevant document is one which discusses genes that could be considered as candidates to test in a randomized controlled trial which studies the genetic risk factors for stroke.

**OHSUMED (Y.-H. Liu et al., 2017)**

> Imagine that you are 63-year-old male with acute renal failure probably 2nd to aminoglycosides/contrast dye. You would like to find information about acute tubular necrosis due to aminoglycosides, contrast dye, outcome and treatment.

*Figure 2. An example OHSUMED search topic, reworded for the participants.*
3. Experimental User Studies and its Challenges

User contexts, individual and task differences

*User contexts, individual and task differences (Buscher, White, Dumais, & Huang, 2012)*

- Use of large scale behavioural log data: Queries, clicks, cursor movements, scrolling, and text highlighting
- User clusters based on behavioural data
- Effect of task type (navigational vs. non-navigational) on user clusters
Learning Activity.4

What kind(s) of search tasks will you investigate in your study? How will you collect and/or design the search tasks?
Physiological signals (Barral et al., 2015)

- Relationship between physiological signals and perceived relevance
- User experiment with 40 participants
- Tasks: Perform searches on topic of interest

Experimental task and user interface
Feature engineering

- Feature window for decision moment
- Feature generation within the window
- Exploratory feature analysis

Workflow of data analysis (Barral et al., 2015, p. 390)

- Experimental study
  - N = 40
  - Data cleaning
    - N = 36
    - Basic signal analysis
    - N = 18
      - Feature engineering
      - N = 18
      - Exploratory feature analysis
      - Feature ranking
      - Predictive models
Personalizing information retrieval (C. Liu, Liu, & Yan, 2018)
Visual and search behaviour (Y.-H. Liu et al., 2017; Wittek et al., 2016)
Summary of Tutorial

- Search models for IIR and their assumptions;
- Benefits and challenges of different approaches to user issues in IIR;
- Design of user experiments for exploring and modelling user search behaviour;
- Implementation of user experiments collaboratively.
References I

References II


References III


References V


