LongSAL: A Longitudinal Study to Understand University Students' Learning During Search

Nilavra Bhattacharya

School of Information

University of Texas at Austin

A thesis submitted for the degree of

Doctor of Philosophy

May 2023

For Yihui Xie

Acknowledgements

This is where you will normally thank your advisor, colleagues, family and friends, as well as funding and institutional support. In our case, we will give our praises to the people who developed the ideas and tools that allow us to push open science a little step forward by writing plain-text, transparent, and reproducible theses in R Markdown.

We must be grateful to John Gruber for inventing the original version of Markdown, to John MacFarlane for creating Pandoc (`http://pandoc.org`) which converts Markdown to a large number of output formats, and to Yihui Xie for creating `knitr` which introduced R Markdown as a way of embedding code in Markdown documents, and `bookdown` which added tools for technical and longer-form writing.

Special thanks to Chester Ismay, who created the `thesisdown` package that helped many a PhD student write their theses in R Markdown. And a very special thanks to John McManigle, whose adaption of Sam Evans' adaptation of Keith Gillow's original maths template for writing an Oxford University DPhil thesis in LaTeX provided the template that I in turn adapted for R Markdown.

Finally, profuse thanks to JJ Allaire, the founder and CEO of RStudio, and Hadley Wickham, the mastermind of the tidyverse without whom we'd all just given up and done data science in Python instead. Thanks for making data science easier, more accessible, and more fun for us all.

Ulrik Lyngs
Linacre College, Oxford
2 December 2018

Abstract

This *R Markdown* template is for writing an Oxford University thesis. The template is built using Yihui Xie's `bookdown` package, with heavy inspiration from Chester Ismay's `thesisdown` and the `OxThesis` LaTeX template (most recently adapted by John McManigle).

This template's sample content include illustrations of how to write a thesis in R Markdown, and largely follows the structure from this R Markdown workshop.

Congratulations for taking a step further into the lands of open, reproducible science by writing your thesis using a tool that allows you to transparently include tables and dynamically generated plots directly from the underlying data. Hip hooray!

Contents

List of Figures

List of Tables

List of Abbreviations

1-D, 2-D . . . One- or two-dimensional, referring **in this thesis** to spatial dimensions in an image.

Otter . . . . . One of the finest of water mammals.

Hedgehog . . . Quite a nice prickly friend.

1 Introduction

1.1 Search as Learning: Overview

Searching for information is a fundamental human activity. In the modern world, it is frequently conducted by users interacting with online search systems (e.g., web search engines), or more formally, **Information Retrieval** (IR) systems. As early as in 1980, Bertam Brookes, in his 'fundamental equation' of information and knowledge, had stated that an information searcher's current state of knowledge is changed to a new knowledge structure by exposure to information (Brookes, 1980, p. 131). This indicates that searchers acquire new knowledge in the search process, and the same information will have different effects on different searchers' knowledge states. Fifteen years later, (Marchionini, 1995) described information seeking as "a process, in which humans purposefully engage in order to change their state of knowledge". Thus, we have known for quite a while that search is driven by higher-level human needs, and IR systems are a means to an end, and not the end in itself. **Interactive information retrieval** (IIR), a.k.a. human-computer information retrieval (HCIR) (Marchionini, 2006) refers to the study and evaluation of users' interaction with IR systems and users' satisfaction with the retrieved information (Borlund, 2013).

1. Introduction

Despite their technological marvels, modern IR systems falls short in several aspects of fully satisfying the higher level human need for information. In essence, IR systems are software that take, as input, some query, and return as output some ranked list of resources.

"Within the context of information seeking, (search engines and IR systems) **feel** like they play a prominent role in our lives, when in actuality, they only play a small role: the **retrieval** part of

$$information$$

...

- Search engines **don't help us identify what we need** – that's up to us; search engines don't question what we ask for, though they do recommend queries that use similar words.
- Search engines **don't help us choose a source** – though they are themselves a source, and a heavily marketed one, so we are certainly compelled to choose search engines over other sources, even when other sources might have better information.
- Search engines **don't help us express our query** accurately or precisely – though they will help with minor spelling corrections.
- Search engines do help retrieve information—this is the primary part that they automate.
- Search engines **don't help us evaluate the answers we retrieve** – it's up to us to decide whether the results are relevant, credible, true; Google doesn't view those as their responsibility.
- Search engines **don't help us sensemake** – we have to use our minds to integrate what we've found into our knowledge."

– (Ko, 2021)

In recent years, the IIR research community has been actively promoting the **Search as Learning** (SAL) research direction. This fast growing community of researchers propose that search environments should be augmented and reconfigured to foster learning, sensemaking, and long-term knowledge-gain. Various workshops and seminars have been organized to develop research agendas at the interaction of IIR and the Learning Sciences (Agosti et al., 2014; Allan et al., 2012; Collins-Thompson et al., 2017; Freund et al., 2013, 2014; Gwizdka et al., 2016) Additionally, special issues on Search as Learning have also been published in the *Journal of Information Science* (Hansen & Rieh, 2016) and in the *Information Retrieval Journal* (Eickhoff et al., 2017). Articles in these special issued presented landmark literature reviews (Rieh et al., 2016; Vakkari, 2016), research agendas, and ideas in this direction. Overall, these works generally advocate that future research in this domain should aim to:

- understand the contexts in which people search to learn
- understand factors that can influence learning outcomes
- understand how search behaviours can predict learning outcomes
- develop search systems to better support learning and sensemaking
- help searchers be more critical consumers of information
- understand the cognitive biases fostered by existing search systems
- develop search engine ranking algorithms and interface tools that foster long term knowledge gain

Parallelly, the Educational Science and the Learning Science research communities have also been organizing workshops and formulating research agendas to conceptualize forms of 'new learning' (Cope & Kalantzis, 2013; Kalantzis & Cope, 2012; New London Group, 1996) that are afforded by innovations in digital

technologies and e-learning ecologies (Cope & Kalantzis, 2017). Higher education researchers have been increasingly studying how students' information search and information use behaviour affect and support their learning (Weber et al., 2019, 2018; Zlatkin-Troitschanskaia et al., 2021). Efforts are underway to conceptualize a theoretical framework around new forms of e-Learning that is aided and afforded by digital technologies (Amina, 2017; Cope & Kalantzis, 2017). In the community's own words: "learning today is more about navigation, discernment, induction, and synthesis" of the wide body of information present ubiquitously at every student's fingertips (Amina, 2017). Therefore "knowing the source, finding the source, and using the information aptly is important to learn and know now more than ever before" (Cope & Kalantzis, 2013). All of these interests in the intersection of searching and learning goes to emphasise that understanding learning during search is critical to improve human-information interaction.

1.2 Problem Statement

A major limitation in the area of Search as Learning, Interactive IR (IIR), and more broadly, in Human-Computer Interaction (HCI) research is that, the user is examined in the short-term, typically over the course of a single experimental session in a lab (Karapanos et al., 2021; Kelly et al., 2009; Koeman, 2020; Zlatkin-Troitschanskaia et al., 2021). Very few studies exist in the search-as-learning domain that have observed *the same participant* over a longer period of time than a single search session (Kelly, 2006a, 2006b; Kuhlthau, 2004; Vakkari, 2001b; White et al., 2009; Wildemuth, 2004). This ephemeral approach has acute implications in any domain where learning is involved because "learning is a *process* that leads to *change* in knowledge ... (which) unfolds over time" (Ambrose et al., 2010), and "...does not happen all at once"(White, 2016b).

To the best of the author's knowledge, almost no new longitudinal studies were reported in major search-as-learning literature in the last

five years, that systematically studied students' information search behaviour and information-use over the long term, in their *in-situ* naturalistic environment and contexts, and linked those behaviours quantitatively to the students' learning outcomes and individual differences. Higher education students are increasingly using the Internet as their main learning environment and source of information when studying. Yet, the short term nature of research in this domain creates significant gaps in our knowledge regarding how students' information search behaviour and information use develop over time, and how it affects their learning (Zlatkin-Troitschanskaia et al., 2021). When research in this area "relies so heavily on (short-term) lab studies, can we realistically say we are comprehensively studying human-tech interactions – when many of those interactions take place over long periods of time in real-world contexts? ... An over-reliance on short studies risks inaccurate findings, potentially resulting in prematurely embracing or disregarding new concepts." (Koeman, 2020).

Current search engines and information retrieval systems "do not help us know what we want to know, ... do not help us know if what we've found is relevant or true; and they do not help us make sense of

$$the retrieved information$$

. All they do is quickly retrieve what other people on the internet have shared" (Ko, 2021). Unless we have more long-term understanding of the nature of knowledge gain during search, the limitations of current search systems will continue to persist. Increased knowledge and understanding of students', and more broadly searchers', information searching and learning behaviour over time will help us to overcome the limitations of current IR systems, and transform them into rich learning spaces where "search experiences and learning experiences are intertwined and even synergized" (Rieh, 2020). The internet and digital educational technologies offer great opportunities to transform learning and the education experience. Enabled by our increased comprehension of the longitudinal searching-as-learning process, improved and validated by empirical data, we can create a new wave of fundamentally

transformative educational technologies and "e-learning ecologies, that will be more engaging for learners, more effective (than traditional classroom practices), more resource efficient, and more equitable in the face of learner diversity" (Cope & Kalantzis, 2017).

1.3 Purpose of this Dissertation Proposal

To address the gaps in our knowledge of how information searching influences students' learning process over time, this dissertation proposal proposes to conduct a semester-long longitudinal study (approx. 16 weeks) with university student participants. The overarching research aim is to identify how students' online searching behaviour correlate with their learning outcomes for a particular university course. Building upon principles from the Learning Sciences (Ambrose et al., 2010; National Research Council, 2000; Novak, 2010; Sawyer, 2005), and empirical evidences from the Information Sciences (Rieh et al., 2016; Vakkari, 2016; White, 2016a), this dissertation proposal aims to

- situate students as learners in their naturalistic contexts, and characterised by their individual differences,
- measure students' information search and information use behaviour over time, and
- correlate the information search behaviour with the learning outcomes for the university course.

Learning, or addressing a gap in one's knowledge, has been well established as an important motivator behind information-seeking activities (Section 1.1). Therefore, search systems that support rapid learning across a number of searchers, and a range of tasks, can be considered as more effective search systems (White, 2016a, p. 310). This dissertation proposal takes a step in this direction. "It opens great expectations for many-sided, great contribution to our knowledge on

the relations between search process and learning outcomes" (Bhattacharya, 2021 anonymous reviewer).

1.4 Outline

This dissertation proposal document is structured as follows. First, principles of learning and relevant background from the domain of Educational Sciences are presented in Chapter 2. Next, relevant empirical evidences from the Information Searching Literature are discussed in Chapter 3. Chapter 4 presents the research questions, the overarching hypotheses, and discusses their rationale in the context of the existing research gaps. Chapter 5 describes the research methods, including the longitudinal study design, experimental procedures, data collection and analyses plans, anticipated limitations, and expected schedule to complete the dissertation.

2

Background: Knowledge and Learning

This first chapter on background literature discusses relevant concepts from the disciplines of Education and Learning Sciences. First, we introduce some relevant terminology, and the concepts of deep or meaningful learning. Then we discuss several research backed principles that have been shown to lead to meaningful learning. Next, we discuss how learning, sensemaking, and searching for information are related, and how modern technologies provide affordances for new forms of learning and knowledge work in the 21st century. We also discuss some concepts about individual differences of learners as well as techniques that can promote better learning. In the last section, we state what implications these findings have for shaping the proposed study in this dissertation proposal.

2.1 Terminology

The Webster dictionary[1] defines **knowledge** in two ways. The first definition is "the range of one's information or understanding". (Vakkari, 2016) says it is "the totality what a person knows, that is, a **personal knowledge** or **belief system**. It may include both justified, true beliefs and less justified, not so true beliefs,

[1] https://developer.chrome.com/docs/extensions/reference/history/#transition-types