

Unexplored Frontiers: A Review of Empirical Studies of Exploratory Search

Alan Medlar

University of Helsinki
Finland

alan.j.medlar@helsinki.fi

Denis Kotkov

University of Helsinki
Finland

denis.kotkov@helsinki.fi

Dorota Głowacka

University of Helsinki
Finland

dorota.glowacka@helsinki.fi

Abstract

This article reviews how empirical research of exploratory search is conducted. We investigated aspects of interdisciplinarity, study settings and evaluation methodologies from a systematically selected sample of 231 publications, including a total of 172 articles with empirical studies. Our results show that exploratory search is highly interdisciplinary, with the most frequently occurring publication venues including high impact venues in information science, information systems and human-computer interaction. However, taken in aggregate, the breadth of study settings investigated was limited. We found that a majority of studies (77%) focused on evaluating novel retrieval systems as opposed to investigating users' search processes. Furthermore, a disproportionate number of studies were based on scientific literature search (20.7%), a majority of which only considered searching for Computer Science articles. Study participants were generally from convenience samples, with 75% of studies composed exclusively of students and other academics. The methodologies used for evaluation were mostly quantitative, but lacked consistency between studies and validated questionnaires were rarely used. In discussion, we offer a critical analysis of our findings and suggest potential improvements for future exploratory search studies.

1 Introduction

Exploratory search was conceptualized as the interdisciplinary study of how search could be used to facilitate complex tasks related to learning and investigation [Marchionini, 2006]. These search tasks are open-ended, give rise to evolving information needs and are highly subjective in terms of the relevance of search results to those needs [White and Roth, 2009; Athukorala et al., 2016a]. Research in exploratory search, therefore, necessitates empirical studies that blend aspects of information science, information retrieval and human-computer interaction to build retrieval systems that support exploratory search activities and to understand related search behaviors. While the complexity of exploratory search has led to definitions that are broad in scope [White and Roth, 2009], empirical studies tend to define exploratory search indirectly according to the attributes of search tasks believed to elicit exploratory behavior [Wildemuth and Freund, 2012; Li and Belkin, 2010].

In this article, we present a systematic review of empirical studies of exploratory search. The motivations for this review are twofold. First, since White and Roth there have been no reviews of exploratory search that surveyed the field as a whole [White and Roth, 2009]. Furthermore, while there have been reviews that focused on specific aspects of exploratory search, such as search tasks [Wildemuth and Freund, 2012] and information seeking models [Palagi et al., 2017], there have been no systematic reviews of the research literature. The period considered in our review was chosen to follow on from White and Roth. Second, we believe that analyzing empirical studies will help to clarify what the research community believes exploratory search is in practice or, at the very least, what aspects of exploratory search receive more attention than others. Furthermore, the related work sections of articles only give a limited snapshot of current research, whereas systematic reviews can provide an unbiased view of a given field of study.

This article aims to provide as complete a picture of research in exploratory search as possible. We look at the interdisciplinary nature of exploratory search, identifying the most common publication venues and how authors from different disciplines position their research contributions. We enumerate what is studied in empirical studies, covering retrieval systems, search domains and simulated work tasks. We additionally investigate the experimental design characteristics of exploratory search studies, including the types of participants studied and what data is collected during evaluation. Lastly, we provide a discussion where we put our findings into context. We detail what we see as problems associated with current research practices and suggest potential improvements for future exploratory search studies

2 Background

Marchionini’s original conception of exploratory search was of a search process that blends the use of queries with browsing strategies to satisfy complex information needs [Marchionini, 2006]. To distinguish exploratory search from other kinds of information search, Marchionini describes three types of search activity: lookup, learn and investigate. Lookup search (or known item search) requires the formulation of well-defined, unambiguous search queries to retrieve discrete, well-structured facts or specific documents. Learn and investigate, however, are seen as core activities in exploratory search, necessitating the use of browsing strategies, multiple search iterations and the assessment of many potentially relevant search results [Athukorala et al., 2016a,b; Mitsui et al., 2018; Kotzyba et al., 2017; Palmquist and Kim, 2000]. For example, searchers engaged in learning tasks, such as self-directed learning, need to evaluate information from various sources (websites, scientific literature, videos, etc.) for the purpose of knowledge acquisition. Whereas investigation tasks involve information synthesis, discovery and planning, e.g., finding relevant articles for a scientific literature review [Medlar et al., 2016], which is more reliant on recall (i.e., maximising the number of relevant documents identified).

In contrast, White and Roth define exploratory search as both an information seeking problem context and a search process [White and Roth, 2009]. As a problem context, users engaged in exploratory search are motivated by either a need to address a knowledge gap [Belkin et al., 1982] or satisfy their curiosity [Pace, 2004]. As a search process, exploratory search involves significant cognitive processing and interpretation, and is, therefore, distinct from traditional information retrieval scenarios [Marchionini, 2006]. White and Roth emphasize the role of uncertainty in exploratory search. Users engaged in exploratory search may be unfamiliar with their search

domain, be unsure how to achieve their goals or be unsure about what their search goals are. White and Roth carry the theme of uncertainty through to a model of exploratory search behavior composed of two phases: exploratory browsing and focused searching [White and Roth, 2009]. Exploratory browsing is necessary due to the uncertainty of the initial problem context, whereas focused searching results from decreasing uncertainty as users learn about the search domain. However, others have suggested that these distinctions are unnecessary and exploratory search can be sufficiently described in terms of other information seeking models [Savolainen, 2018].

Wildemuth and Freund highlight how experimental studies tend to define exploratory search indirectly in terms of cognitive and behavioral search task attributes [Wildemuth and Freund, 2012]. Cognitive attributes of exploratory search tasks include: (i) learning and investigation related search goals [Marchionini, 2006], (ii) conceptually broad or underspecified search task descriptions [Kules and Capra, 2008], (iii) uncertainty with respect to the clarity of the information need, the topic, the available documents or search outcomes [White and Roth, 2009], (iv) ill-structured underlying questions, problems or information needs [MacMullin and Taylor, 1984], (v) dynamic, in terms of having evolving information needs [Bates, 1989], (vi) multi-faceted, with search outcomes covering multiple concepts (e.g., Toms et al. [2007]), (vii) “not too easy”, in terms of the inherent task difficulty [Kules and Capra, 2008] or procedural complexity, e.g., involving many subtasks [Diriye et al., 2010], and (viii) accompanied by cognition, such as decision-making or sensemaking [Dervin, 1998]. Whereas, from a behavioral perspective, exploratory search tasks can feature: (i) open-ended problems for which there are no specific answers [Marchionini, 1993], (ii) multiple target items resulting from open-ended learning tasks [Marchionini, 2006], and (iii) occur over time and, therefore, involve multiple queries and search sessions [White and Roth, 2009]. Our review complements Wildemuth and Freund’s analysis by providing an overview of the field to understand what is studied in exploratory search and how it is evaluated experimentally.

In summary, exploratory search is broadly defined with experimental studies describing it in terms of numerous cognitive and behavioral attributes. In our meta-analysis, we investigate bibliographic data (e.g., publication venue, keywords), what is studied (retrieval systems, search domains, work tasks), and how evaluation is conducted (experimental design, choice of participants, data collection). Our goal is to understand how investigators present exploratory search in practice and expose underexplored areas of the problem space.

3 Review Procedure

We conducted our review using an analytical approach based on a representative sample of published work spanning a diverse range of publication venues. The selection of publications was based on the QUOROM statement [Moher et al., 1999], that defines a procedure for conducting meta-analyses. Figure 1 provides an overview of how papers were identified for inclusion in the study.

3.1 Literature Search

Articles related to exploratory search can be found in multiple scientific conferences and journals covering various research disciplines. We selected three scientific repositories to conduct literature

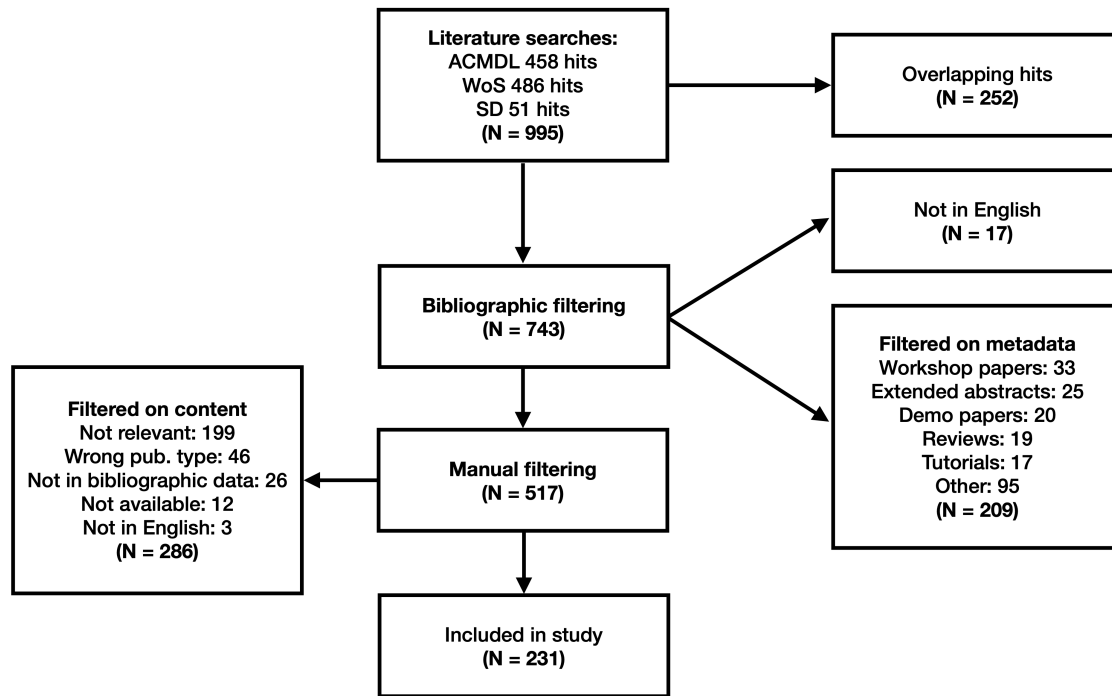


Figure 1. Article selection procedure overview.

searches: ACM Digital Library (ACMDL), Web of Science (WoS, formerly Web of Knowledge) and Science Direct (SD). ACMDL provides access to ~ 3 million articles related to computing through the ACM Guide to Computing Literature. WoS’s core collection includes ~ 80 million articles across all fields of science. SD contains ~ 18 million articles from Elsevier publication venues across all fields of science.

For each scientific repository, we retrieved all publications where the exact phrase “exploratory search” appears in the bibliographic data (i.e., title, abstract, keywords). WoS did not allow us to limit search terms to bibliographic data, so we searched using “All fields”. In general, we avoided full-text search as it returned too many articles that were not primarily about exploratory search. The literature search was conducted in mid-2022 and, therefore, we limited our search to articles published between 2010–2021, inclusive. We did not limit search results on the basis of any other criteria, such as publication venue.

The literature search resulted in 995 search results (ACMDL: 458, WoS: 486, SD: 51). We identified duplicate articles between repositories based on DOIs (Digital Object Identifier) and, when the DOI was absent, article titles (duplicates identified via titles were manually checked). After duplicate search results were removed, 743 publications remained.

3.2 Article Selection

We imported the metadata into Excel and manually screened articles to narrow down the set of papers to: (i) peer-reviewed conference or journal articles, that were (ii) written in English.

As metadata related to the type of publication was not always available, we inferred it using a

combination of page count, article title (e.g., “A review of ...”), abstract (“In this survey ...”) and publication venue (“Proceedings of the First International Workshop on ...”). This procedure excluded: workshop papers (33), extended abstracts (25), demo papers (20), review articles (19), tutorials (17), and other miscellaneous articles (95), including doctoral consortium papers, dissertations, and keynote speeches. We excluded a further 17 papers that were not written in English. To ensure this process was reliable, two of the authors independently marked papers for exclusion. Inter-rater reliability was assessed using Cohen’s kappa ($\kappa = 0.77$, $p < .00001$), indicating substantial agreement [Landis and Koch, 1977]. Disagreements were discussed between all authors and resolved by consensus. This procedure left 517 scientific articles.

3.3 Validity Assessment

In the validity assessment step, we screened out articles that were not substantively about exploratory search. We excluded papers that were not related to information retrieval or information seeking (e.g., exploratory search is an unrelated research topic in the optimization literature), or where, despite appearing in bibliographic data, exploratory search is only peripherally mentioned in related work or discussion.

We downloaded the corresponding PDF files for all search results. Of the 517 papers from the article selection step, we excluded a further 286 papers for the following reasons: (i) articles not about exploratory search (199), (ii) incorrect publication type, only discernible from the full text due to missing metadata (46), (iii) “exploratory search” not found in title, abstract or keywords from WoS (26), (iv) unavailable online or inaccessible from our research institution (12), and (v) body of the article was not written in English (3). Assessment was performed independently by all three authors. Inter-rater reliability was assessed using Fleiss’ kappa, ($\kappa = 0.71$, $p < .00001$), indicating strong agreement. Any cases where the decision to exclude an article was unclear were discussed between the authors to reach a consensus. After validity assessment, we were left with 231 papers, including journal articles, full-length and short conference papers, for use in our study.

Following Webster and Watson, we took a concept-centric approach to our review where pre-defined units of analysis were checked and coded for in each paper as it was read [Webster and Watson, 2002]. These coded concepts were compiled into frequency tables, which we present next.

4 Results

In our results, we analyzed the interdisciplinarity of research in exploratory search (i.e., *who conducts exploratory search research?*), we enumerate retrieval systems, search domains and simulated work tasks (*what is the focus of exploratory search research?*), and, lastly, detail the study characteristics of user studies based on simulated work tasks (*how do we conduct evaluation in exploratory search studies?*). Several of the tables presented below are stratified by discipline (information science, information systems and human-computer interaction). However, for concision, this is only shown where per-discipline results differed substantially from the overall results.

Publication Venue	Freq.
Conference on Human Information Interaction and Retrieval* (CHIIR)	15
Conference on Intelligent User Interfaces (IUI)	13
Conference on Information and Knowledge Management (CIKM)	10
Information Processing & Management (IPM)	9
Journal of the Association for Information Science and Technology (JASIST)	9
Conference on Research and Development in Information Retrieval (SIGIR)	9
Association for Information Science and Technology (ASIS&T)	8
Conference on Human Factors in Computing Systems (CHI)	6
Joint Conference on Digital Libraries (JCDL)	5
The Web Conference (WWW)	4

Table 1. Most frequently occurring publication venues (* = publication count for CHIIR includes both CHIIR (6) and IiX (9)).

4.1 Interdisciplinarity

To characterize the interdisciplinary nature of research in exploratory search, we analyzed which journals and conferences each article was published in and how authors used article keywords to describe their work.

4.1.1 Publication Venues

Table 1 shows the 10 most frequently occurring publication venues in our data set, accounting for 38.1% of articles. These venues highlight the interdisciplinary nature and high standard of research in exploratory search, encompassing information science (JASIST, ASIS&T, JCDL), information systems (IPM, CIKM, SIGIR, WWW) and human-computer interaction (CHIIR, IUI, CHI).

Based on this finding, we categorised all 231 articles based on whether they were published in venues specialising in information science (18.6%), information systems (35.5%), human-computer interaction (26.8%), and other (19.1%). The other category represents a wide variety of conferences and journals, specializing in fields from artificial intelligence to educational technologies, and multidisciplinary venues, such as PLOS One.

4.1.2 Author-defined Keywords

Table 2 shows the top-10 author-defined keywords for articles from information science, information systems and human-computer interaction. We lightly edited the keywords by making plurals singular and unified any differences between British and American spelling. While a majority of articles had author-defined keywords (91%), the proportion varied between disciplines with information science being the lowest (81%).

The keyword distributions highlight the differences between disciplines that conduct research into exploratory search: there is a strong theme of information seeking in the information science articles (e.g., “*information seeking behavior*”, “*search strategies*”, “*information foraging*”).

Information Science			Information Systems			Human-Computer Interaction		
Keywords	Freq.	%	Keywords	Freq.	%	Keywords	Freq.	%
exploratory search	16	37.2	exploratory search	54	65.9	exploratory search	44	71.0
information seeking	6	14.0	information retrieval	7	8.5	information seeking	6	9.7
evaluation	4	9.3	faceted search	5	6.1	user study	6	9.7
user study	4	9.3	search user interface	4	4.9	user interface	4	6.5
interactive ir	4	9.3	user study	4	4.9	sensemaking	4	6.5
faceted search	4	9.3	linked data	4	4.9	information viz.	4	6.5
information seeking behavior	3	7.0	semantic web	4	4.9	collaborative search	3	5.6
search strategies	2	4.7	information viz.	4	4.9	search behavior	3	5.6
information foraging	2	4.7	interactive ir	4	4.9	visual analytics	3	5.6
serendipity	2	4.7	recommender system	3	3.7	visualization	3	5.6

Table 2. Author-defined keywords from exploratory search articles in different disciplines.

Whereas information systems focuses on technological aspects of search systems (“*information retrieval*”, “*linked data*”, “*semantic web*”) and human-computer interaction has numerous keywords related to information presentation (“*user interface*”, “*information visualization*”, “*visual analytics*”). There are also commonalities between disciplines, with keywords found in multiple fields highlighting the dialogue-like back and forth between user and search system (“*interactive information retrieval*”) and the central role of experimentation (“*user study*”).

4.2 Study Settings

Of the 231 papers included in our study, 172 included one or more user studies. We used this subset of papers to understand what empirical studies of exploratory search investigate in terms of retrieval systems, search domains and simulated work tasks.

4.2.1 Retrieval Systems

Table 3 shows the named retrieval systems from the 172 articles that included user studies. We further categorized each system as *experimental* and *production* systems. Experimental systems tend to be prototypes with articles often evaluating new features, whereas production systems are available on the open Internet, such as commercial search engines (e.g., Bing) and digital libraries (Sowiport). We did not count baseline systems unless they were a named retrieval system (i.e., “list-based interfaces” or baselines comprised of the experimental system without specific features were not included). Systems that appeared only once in the data set were collapsed into a “miscellaneous” subcategory.

A majority of exploratory search user studies are based on the evaluation of experimental search systems (77%) with the most frequently occurring systems being Coagmento [Shah, 2010] and SciNet [Glowacka et al., 2013]. These studies tended to evaluate novel interface components that support users performing exploratory search (e.g., Qvarfordt et al. [2013]). Whereas, studies based on production systems focused primarily on investigating exploratory search behavior. For example, Athukorala et al. [2016a] performed a comparative study of lookup and exploratory search behavior using Google Scholar.

	System Name	Freq.	%
Experimental	Coagmento	8	4.6
	SciNet	6	3.4
	Discovery Hub	4	2.3
	Cruise	3	1.7
	CollabSearch	2	1.1
	ConceptCloud	2	1.1
	Paths	2	1.1
	Querium	2	1.1
	uRank	2	1.1
	Misc.	103	59.2
Production	User-selected Search Engine	15	8.6
	Bing	5	2.9
	Google	5	2.9
	Sowiport	3	1.7
	Google Scholar	2	1.1
	Misc.	10	5.7
Total		174*	

Table 3. Named retrieval systems in our data set (* = total is greater than the number of papers with user studies (N=172) as two papers compared two of the listed systems).

4.2.2 Search Domains

Table 4 shows the domains used in exploratory search user studies stratified by discipline. We grouped our coding of domains into 5 categories: *data*, *documents*, *multimedia*, *scientific literature* and *web*. Data included structured data, such as knowledge graphs (most commonly DBpedia [Lehmann et al., 2015]) and log data. Documents included unstructured corpus data, such as news articles. Multimedia included audio, video and image data, with “various” indicating multiple media types. Scientific literature was coded by discipline, with an additional “multiple disciplines” subdomain for search tasks where the participant’s background was relevant. Lastly, web included different domains of web search, such as travel and politics, following articles’ descriptions. In all domains, we included a “miscellaneous” category to group infrequently occurring subdomains. Each article could include multiple domains or multiple subdomains within a given domain.

All domain categories were present in all three disciplines. The most common domains overall were web (42.9%) and scientific literature (20.7%). The least common domain was multimedia (8.5%). We note that user studies could include multiple subdomains. For example, web search covers 42.9% of *subdomains* listed in Table 4, but only 30.8% of *articles*. Studies in the data domain were primarily from information systems where exploratory search is studied in connection to semantic web (e.g., Thakker et al. [2013]). Interestingly, a majority of the studies in scientific literature domain were related to Computer Science (accounting for 17.0% of subdomains and 20.9% of articles). This was driven by studies in information systems and human-computer interaction, which likely had access to Computer Science students who needed to perform learning tasks in the completion of their dissertations. For example, di Sciascio et al. investigated ideas from social search to enhance the exploratory search of scientific literature for Master’s and doctoral students [di Sciascio et al., 2018].

4.2.3 Tasks

Table 5 shows three categories of search task performed during exploratory search user studies: exploratory search, lookup search and no simulated work task. We note that several articles contained multiple user studies. A majority of studies (63.6%) included simulated work tasks [Borlund, 2003] for exploratory search. Following Athukorala et al. [2016a], we further subdivided exploratory search tasks into knowledge acquisition (learning tasks with open-ended search goals [Wu et al., 2012]), planning (investigation tasks that are focused on producing a general overview of a topic [Wu et al., 2012]) and comparison (gathering information about two or more topics to identify similarities and differences [Liu and Belkin, 2010]). Very few studies contrasted exploratory search with lookup search tasks (5.9%). We categorized studies with no simulated work task into usability assessment, search results assessment and log data analysis (30.5%).

Exploratory search studies with simulated work tasks feature a reasonably balanced combination of learning and investigation tasks. Several studies contrast exploratory and lookup search to identify differences in user search behavior. Lastly, studies with no simulated work task tend to focus on usability assessment, for example, in interface evaluation and co-design activities.

Domain	Subdomain	Info. Science		Info. Systems		HCI		All (incl. other)	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%
Data	DBpedia			6	8.3			8	3.8
	Experimental Results			1	1.4	2	3.2	3	1.4
	Search Logs			2	2.8			2	0.9
	Academics			2	2.8			2	0.9
	Misc.	1	2.2	3	4.2	4	6.3	7	3.3
Documents	News			1	1.4	1	1.6	6	2.8
	Digital Heritage	3	6.5			1	1.6	4	1.9
	Health			2	2.8	1	1.6	4	1.9
	Misc.	4	8.7	7	9.7	7	11.1	23	10.8
Multimedia	Images	1	2.2	4	5.6	2	3.2	9	4.2
	Various*	1	2.2	1	1.4	2	3.2	5	2.4
	Video	2	4.3	1	1.4			3	1.4
	Audio	1	2.2					1	0.5
Scientific	Computer Science	5	10.9	12	16.7	17	27.0	36	17.0
Literature	Multiple Disciplines**	1	2.2			1	1.6	3	1.4
	Social Sciences	3	6.5					3	1.4
	Life Sciences					1	1.6	2	0.9
Web	Technology	6	13.0	4	5.6	4	6.3	16	7.5
	Environment	3	6.5	4	5.6	4	6.3	12	5.7
	Health	1	2.2	3	4.2	3	4.8	10	4.7
	Travel	2	4.3	3	4.2	2	3.2	8	3.8
	Education	1	2.2	1	1.4	1	1.6	7	3.3
	History	2	4.3	4	5.6			7	3.3
	News	4	8.7			1	1.6	5	2.4
	Misc.	5	10.9	11	15.2	9	14.3	26	12.3
Total		46		72		63		212***	

Table 4. Overview of domains used in exploratory search articles with user studies stratified by discipline (N=172). The “All” column also includes studies outside of information science, information systems and human-computer interaction. (* = containing multiple multimedia data types, ** = domain decided based on participant’s background, *** = many user studies covered multiple subdomains).

	Task Type	Freq.	%
Exploratory Search	Knowledge Acquisition	62	33.2
	Planning	50	26.7
	Comparison	7	3.7
Lookup Search	-	11	5.9
No Simulated Work Task	Usability Assessment	30	16.0
	Search Results Assessment	16	8.6
	Log Data Analysis	11	5.9
Total		187*	

Table 5. Types of task performed in exploratory search articles with user studies (N=172). (* = several articles contained user studies that involved multiple task types).

Experimental Design	Freq.	%
Within-subject	62	50.0
Observational	31	25.0
Between-subject	23	18.5
Mixed	8	6.5

Table 6. Experiment design in studies with simulated work tasks (N=124).

4.3 Evaluation Methodologies

To understand the experimental characteristics of exploratory search user studies, we focused on the 115 articles that included 124 studies with simulated work tasks. We investigated the experimental design, choice of study participants and methods of data collection.

4.3.1 Experimental Design

In user studies where participants performed simulated work tasks, sample sizes ranged from 2 to 381. The median sample size was 24 and the mean sample size was 35.7 (95% CI [28.0, 43.5]). A majority of exploratory search user studies were experiments (75%), i.e., non-observational studies (see Table 6). The most common experimental design was within-subject, which had lower average sample sizes (median = 20, mean = 32.1 (95% CI [19.6, 44.6])) than between-subject studies (median = 41.5, mean = 50.0 (95% CI [36.4, 63.5])). In general, there is a lack of large-scale studies in exploratory search: we found 5 user studies with sample sizes greater than 100 participants. The largest of these studies, however, pooled observations from multiple past experiments [Hendahewa and Shah, 2017] or were focused on evaluating a questionnaire in the context of exploratory search [O’Brien and Toms, 2013].

Category	Type of Participant	Freq.	%
Academics	Postgraduates	44	35.5
	Undergraduates	39	31.5
	Students	24	19.4
	Doctoral Students/Researchers	22	17.7
	Academic Staff	18	14.5
	Postdoctoral Researchers	11	8.9
	From a University*	11	8.9
Non-academics	Non-descript	13	10.5
	General Public	11	8.9
	Domain Experts/Employees/Teachers/Authors	6	4.8
	High School Students	1	0.8
	Mechanical Turk	1	0.8

Table 7. Participants in studies with simulated work tasks (N=124).

4.3.2 Choice of Participants

We extracted details about study participants from the 124 studies that used simulated work tasks (see Table 7). As participant descriptions are informal and often ambiguous (e.g., “postgraduate” and “doctoral student” may be synonyms in some countries, but not others), we categorized participants into *academics* and *non-academics*, but otherwise did not want to over-interpret their descriptions.

In a majority of studies, participants either included academics (83.9%) or were exclusively composed of academics (75.0%). We additionally extracted disciplines from the descriptions of academic participants. The two most common disciplines were *Unstated* (38.9%) and *Computer Science* (30.6%) (data not shown). The most common types of non-academic participant were without any description (10.5%) or described as the general public (8.9%). It is possible that non-academic participants had academic backgrounds.

4.3.3 Data Collection

Table 8 shows the data collected in the 124 studies that included simulated work tasks. We categorized each measure as *interaction*, *performance* or *usability* [Kelly, 2009]. Interactions are observations used to understand search behavior (e.g., search queries, think-aloud, etc.). We collectively refer to interface-related interactions other than search queries, clicks and bookmarks as *interface interactions* (i.e., interactions specific to a given retrieval system). Performance measures are used to judge the success of a simulated work task (e.g., task time). Usability included questionnaires and interviews to collect information about users’ experiences and preferences.

The most common interaction measures were easy to collect events, such as queries, clicks and bookmarks, with few studies capturing qualitative indicators of search behavior (think-aloud, eye tracking, etc.) that would require additional coding or more complex analysis. Similarly,

Interaction	Freq.	%	Performance	Freq.	%	Usability	Freq.	%
Search Queries	61	49.2	Task Time	36	29.0	Self-defined Questionnaires	53	42.7
Interface Interactions	61	49.2	Free-form Summary	23	18.5	Interviews	41	33.1
Clicked Documents	53	42.7	Relevant Documents	18	14.5	SUS (Usability)	12	9.7
Bookmarks	42	33.9	Question Answers	12	9.7	ResQue (UX)	11	8.9
Think-Aloud	16	12.9	Relevant Snippets	5	4.0	NASA TLX (Workload)	7	5.6
Screen Capture	14	11.3	Concept Maps	3	2.4	UES (Engagement)	3	2.4
Snippets	12	9.7	Research Question	1	0.8	USE (Usability)	1	0.8
Eye Tracking	10	8.1	Essays	1	0.8	PANAS (Affect)	1	0.8
Chats	8	6.4				TAM (Acceptance)	1	0.8
Video Recording	3	2.4						
User Notes	3	2.4						
Audio Recording	1	0.8						

Table 8. Data collected in user studies with simulated work tasks (N=124).

the most frequently occurring performance measures were those that can be quantified, such as task time, free-form summaries (graded by a domain expert) and relevant documents (retrieved or bookmarked documents classified by a domain expert or the participant). While there have been attempts to develop novel performance indicators, such as concept maps [Egusa et al., 2010, 2014], to evaluate knowledge acquisition, such methods have not been widely-adopted by the research community. Lastly, very few studies used standardized questionnaires, but asked similar questions in self-defined (unvalidated) questionnaires. Furthermore, many of the studies that used standardized questionnaires tended to use more than one. For example, Medlar et al. [2021] used both SUS [Brooke, 1996] and ResQue [Pu et al., 2011] to investigate usability and user experience, respectively.

5 Discussion and Limitations

This review has presented a meta-analysis of empirical studies in exploratory search. We were motivated by the lack of reviews of exploratory search and chose the period of 2010-2021 to continue on from White and Roth’s review [White and Roth, 2009]. Furthermore, we wanted to focus on empirical research to enumerate what is studied (systems, domains and tasks) and how it is studied (experimental designs, participants, data collection). The following discussion is a critical analysis of our findings and how they should influence future research in exploratory search.

5.1 Interdisciplinarity

Our review findings confirm the interdisciplinary nature of exploratory search with a majority of articles published in journals and conferences dedicated to information science, information systems and human-computer interaction (80.9%). Our findings also highlight the quality of that research: the ten most frequently occurring publication venues account for a large proportion of our data set (38.1%) and include the top-ranked venues in each discipline. While our analysis of author-defined keywords suggests that the core focus of articles is generally discipline-specific, we

believe this is related to how research is positioned for a given audience as (i) the most prolific authors in our data set published articles in more than one discipline and (ii) outside of search domains, we found very little distinction between disciplines at the level of granularity presented in this review.

We were surprised to find so few articles published in venues dedicated to disciplines that utilize search, such as medicine (e.g., [Pang et al. \[2016\]](#)) and education (e.g., [Cortinovis et al. \[2019\]](#)), in our data set. Future studies should seek to widen the scope of domains in which exploratory search is applied. This could lead to a more comprehensive understanding of how exploratory search manifests in different search contexts and identify novel challenges. Furthermore, such studies should be published in venues that correspond to the specific domain under study, e.g., studies on legal search in law journals, to ensure a high degree of rigor during the peer-review process.

5.2 Study Settings

We identified significantly more empirical studies based on experimental retrieval systems (77%) than production systems, such as commercial search engines (23%). This shows that a majority of studies were focused on the evaluation of new technologies and interfaces to support searchers rather than investigating exploratory search processes. Furthermore, there appears to be very little software reuse between studies, with SciNet [[Glowacka et al., 2013](#)] and Coagmento [[Shah, 2010](#)] being the main exceptions. However, SciNet is not publicly available and Coagmento is intended for collaborative search and not specifically exploratory search. Future work in exploratory search should have a greater emphasis on software reuse and reproducibility. The source code of retrieval systems should be open source and, where possible, log data from anonymized search sessions made available for conducting meta-analyses (e.g., [Medlar and Glowacka \[2018\]](#)).

A majority of empirical studies used simulated work tasks to situate users in exploratory search scenarios (63.6%). While we identified similar numbers of simulated work tasks for learning and investigation, very few studies included lookup search tasks. Therefore, few studies could attempt to create search systems that could adapt to whether users were conducting lookup or exploratory search [[Athukorala et al., 2016a,b](#); [Medlar et al., 2017](#)]. While empirical studies have been conducted in numerous search domains, there is a clear bias toward web search (43.0%) and scientific literature search (20.7%). Furthermore, some domains were either only found in a single discipline (the data domain was almost exclusively found in articles from information systems) or concentrated in a single subcategory (81.8% of studies based on scientific literature search were limited to Computer Science articles). Future work on exploratory search must broaden the scope of research beyond current search domains to enable the study of more diverse user populations, search tasks and information modalities (e.g., [Kropotov et al. \[2021\]](#); [Liu et al. \[2022\]](#)).

5.3 Evaluation Methodology

From a methodological perspective, our review highlights many of the issues with empirical studies in exploratory search. Sample sizes are generally low, with calculations for statistical power either rarely reported or not conducted. Either way, within-subject and between-subject experimental designs likely had similar statistical power with median sample sizes of 20 and 41.5, respectively. In general, studies used convenience samples of participants from the authors' universities. Indeed,

75% of studies were based exclusively on students and academics. This lack of engagement with participants outside of academia is concerning and shows that our understanding of exploratory search is limited to WEIRD populations [Henrich et al., 2010]. Only a small proportion of studies included members of the general public and we only identified a single study that used a crowdsourcing platform [Cui et al., 2014]. Future research on exploratory search should focus on recruiting study participants from the broader population. In particular, experiments could be designed for crowdsourcing platforms (i.e., including control mechanisms to ensure reliability) which have been shown to be more representative of the general population than laboratory studies [Difallah et al., 2018] and will allow for larger sample sizes.

While empirical studies of exploratory search used numerous methods in evaluation, they tended to be based on quantitative interaction and performance variables that are easily measured, such as search queries, clicks and task time. In general, there was little consensus on what variables were used to measure interaction behavior, whereas performance measures tended to be task-specific, e.g., using free-form summaries to simulate essay planning tasks. Moreover, qualitative research methods, such as think-aloud, were rarely used despite being one of the few techniques available to investigate the cognitive and interpretive aspects of users' search processes. While interviews were used in many studies, they suffer from being retrospective and are unable to capture the fleeting thoughts and iterative nature of exploratory search. Anecdotally, we observed that interviews were rarely coded, but mined for quotations to support quantitative results. Lastly, while usability was measured using questionnaires, these were generally non-validated, with standardized questionnaires rarely used. Future research in exploratory search should make better use of qualitative research methods to understand users' search processes and validated questionnaires should be preferred to ensure that results are robust and comparable between studies.

5.4 Limitations

The aim of this article was to provide a comprehensive review of empirical studies in exploratory search. The review process was systematic and, therefore, reproducible. However, our review also has several limitations. First, our review could only provide an overview of exploratory search as the process of conducting a systematic review necessarily leads to the nuances of individual studies to be lost. While our review did not elaborate on the details of individual studies, we believe that it can be used as a starting point for more focused reviews in the future. Second, despite exploratory search being recognized as an important topic in information retrieval and search, it is a relatively niche area of research. Prolific research groups, such as the authors of SciNet, for example, have the potential to skew the distributions reported in our results section. Third, many of the articles that were not categorized as information science, information systems or human-computers interaction added little value to the review. However, we decided to keep these articles to avoid biasing results to only top venues. Fourth, our study excludes papers that define themselves as searching as learning. While several of the articles in our review used both terms, our study design was such that it required papers to be explicitly related to exploratory search. We similarly omitted papers positioned as interactive information retrieval and cognitive search that could have fit the definition of exploratory search. Fifth, we may have omitted important papers due to the selection of databases. We selected databases based on our understanding of

the type of publication venue that would accept work in exploratory search and, therefore, believe that the number of omissions is likely to be small in number and would not have substantially impacted our overall conclusions.

Acknowledgments

This work has been supported by Helsinki Institute for Information Technology HIIT.

References

- Kumaripaba Athukorala, Dorota Głowacka, Giulio Jacucci, Antti Oulasvirta, and Jilles Vreeken. Is exploratory search different? a comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology*, 67(11): 2635–2651, 2016a.
- Kumaripaba Athukorala, Alan Medlar, Antti Oulasvirta, Giulio Jacucci, and Dorota Glowacka. Beyond relevance: Adapting exploration/exploitation in information retrieval. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pages 359–369, 2016b.
- Marcia J Bates. The design of browsing and berrypicking techniques for the online search interface. *Online review*, 1989.
- Nicholas J Belkin, Robert N Oddy, and Helen M Brooks. ASK for information retrieval: Part i. Background and theory. *Journal of Documentation*, 38(2):61–71, 1982.
- Pia Borlund. The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3):8–3, 2003.
- John Brooke. SUS: A “quick and dirty” usability scale. *Usability Evaluation in Industry*, 189(3): 189–194, 1996.
- Renato Cortinovis, Alexander Mikroyannidis, John Domingue, Paul Mulholland, and Robert Farrow. Supporting the discoverability of open educational resources. *Education and Information Technologies*, 24:3129–3161, 2019.
- Licong Cui, Rebecca Carter, and Guo-Qiang Zhang. Evaluation of a novel conjunctive exploratory navigation interface for consumer health information: a crowdsourced comparative study. *Journal of Medical Internet Research*, 16(2):e45, 2014.
- Brenda Dervin. Sense-making theory and practice: An overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, 2(2):36–46, 1998.
- Cecilia di Sciascio, Peter Brusilovsky, and Eduardo Veas. A study on user-controllable social exploratory search. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces*, pages 353–364, 2018.

-
- Djellel Difallah, Elena Filatova, and Panos Ipeirotis. Demographics and dynamics of mechanical turk workers. In *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, pages 135–143, 2018.
- Abdigani Diriye, Max L Wilson, Ann Blandford, and Anastasios Tombros. Revisiting exploratory search from the HCI perspective. *HCIR 2010*, 99, 2010.
- Yuka Egusa, Hitomi Saito, Masao Takaku, Hitoshi Terai, Makiko Miwa, and Noriko Kando. Using a concept map to evaluate exploratory search. In *Proceedings of the 3rd Symposium on Information Interaction in Context*, pages 175–184, 2010.
- Yuka Egusa, Masao Takaku, and Hitomi Saito. How concept maps change if a user does search or not? In *Proceedings of the 5th Information Interaction in Context Symposium*, pages 68–75, 2014.
- Dorota Glowacka, Tuukka Ruotsalo, Ksenia Konuyshkova, Kumaripaba Athukorala, Samuel Kaski, and Giulio Jacucci. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, pages 117–128, 2013.
- Chathra Hendaheewa and Chirag Shah. Evaluating user search trails in exploratory search tasks. *Information Processing & Management*, 53(4):905–922, 2017.
- Joseph Henrich, Steven J Heine, and Ara Norenzayan. The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3):61–83, 2010.
- Diane Kelly. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval*, 3(1–2):1–224, 2009.
- Michael Kotzyba, Tatiana Gossen, Johannes Schwerdt, and Andreas Nürnberger. Exploration or fact-finding: Inferring user’s search activity just in time. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 87–96, 2017.
- Ivan Kropotov, Alan Medlar, and Dorota Glowacka. Exploratory search of GANs with contextual bandits. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 3157–3161, 2021.
- Bill Kules and Robert Capra. Creating exploratory tasks for a faceted search interface. *HCIR 2008*, pages 18–21, 2008.
- J. Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174, 1977.
- Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6(2):167–195, 2015.

-
- Yuelin Li and Nicholas J. Belkin. An exploration of the relationships between work task and interactive information search behavior. *Journal of the American Society for Information Science and Technology*, 61(9):1771–1789, 2010.
- Jingjing Liu and Nicholas J. Belkin. Personalizing information retrieval for multi-session tasks: The roles of task stage and task type. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 26–33, 2010.
- Yang Liu, Alan Medlar, and Dorota Glowacka. Rogue: A system for exploratory search of gans. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3278–3282, 2022.
- Susan E. MacMullin and Robert S. Taylor. Problem dimensions and information traits. *The Information Society*, 3(1):91–111, 1984.
- Gary Marchionini. Information seeking in full-text end-user-oriented search systems: The roles of domain and search expertise. *Library & Information Science Research*, 15(1):35–69, 1993.
- Gary Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.
- Alan Medlar and Dorota Glowacka. How consistent is relevance feedback in exploratory search? In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1615–1618, 2018.
- Alan Medlar, Kalle Ilves, Ping Wang, Wray Buntine, and Dorota Glowacka. PULP: A system for exploratory search of scientific literature. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1133–1136, 2016.
- Alan Medlar, Joel Pyykkö, and Dorota Glowacka. Towards fine-grained adaptation of exploration/exploitation in information retrieval. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, pages 623–627, 2017.
- Alan Medlar, Jing Li, and Dorota Glowacka. Query suggestions as summarization in exploratory search. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, pages 119–128, 2021.
- Matthew Mitsui, Jiquan Liu, and Chirag Shah. How much is too much? Whole session vs. first query behaviors in task type prediction. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1141–1144, 2018.
- Dr Moher, Deborah Cook, Susan Eastwood, Ingram Olkin, Drummond Rennie, and Donna Stroup. Improving the quality of reports of meta-analyses of randomised controlled trials: The quorum statement. *Lancet*, 354:1896–900, 12 1999.
- Heather L. O’Brien and Elaine G. Toms. Examining the generalizability of the user engagement scale (UES) in exploratory search. *Information Processing & Management*, 49(5):1092–1107, 2013.

-
- Steven Pace. A grounded theory of the flow experiences of web users. *International Journal of Human-Computer Studies*, 60(3):327–363, 2004.
- Emilie Palagi, Fabien Gandon, Alain Giboin, and Raphaël Troncy. A survey of definitions and models of exploratory search. In *Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics*, pages 3–8, 2017.
- Ruth A. Palmquist and Kyung-Sun Kim. Cognitive style and on-line database search experience as predictors of Web search performance. *Journal of the American Society for Information Science*, 51(6):558–566, 2000.
- Patrick Cheong-Iao Pang, Shanton Chang, Karin Verspoor, and Jon Pearce. Designing health websites based on users’ web-based information-seeking behaviors: A mixed-method observational study. *Journal of Medical Internet Research*, 18(6):e145, 2016.
- Pearl Pu, Li Chen, and Rong Hu. A user-centric evaluation framework for recommender systems. In *Proceedings of the 5th ACM Conference on Recommender Systems*, pages 157–164, 2011.
- Pernilla Qvarfordt, Gene Golovchinsky, Tony Dunnigan, and Elena Agapie. Looking ahead: Query preview in exploratory search. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 243–252, 2013.
- Reijo Savolainen. Berrypicking and information foraging: Comparison of two theoretical frameworks for studying exploratory search. *Journal of Information Science*, 44(5):580–593, 2018.
- Chirag Shah. Coagmento—a collaborative information seeking, synthesis and sense-making framework. *Integrated demo at CSCW*, 2010.
- Dhavalkumar Thakker, Vania Dimitrova, Lydia Lau, Fan Yang-Turner, and Dimoklis Despotakis. Assisting user browsing over linked data: Requirements elicitation with a user study. In *Web Engineering: 13th International Conference, ICWE 2013.*, pages 376–383. Springer, 2013.
- Elaine G. Toms, Heather O’Brien, Tayze Mackenzie, Chris Jordan, Luanne Freund, Sandra Toze, Emilie Dawe, and Alexandra Macnutt. Task effects on interactive search: The query factor. In *International Workshop of the Initiative for the Evaluation of XML Retrieval*, pages 359–372. Springer, 2007.
- Jane Webster and Richard T. Watson. Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, pages xiii–xxiii, 2002.
- Ryen W. White and Resa A. Roth. Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1):1–98, 2009.
- Barbara M. Wildemuth and Luanne Freund. Assigning search tasks designed to elicit exploratory search behaviors. In *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval*, pages 1–10, 2012.
- Wan-Ching Wu, Diane Kelly, Ashlee Edwards, and Jaime Arguello. Grannies, tanning beds, tattoos and nascar: Evaluation of search tasks with varying levels of cognitive complexity. In *Proceedings of the 4th Information Interaction in Context Symposium*, pages 254–257, 2012.