Evaluating Stability of Information Needs

Christin Katharina Kreutz[1,*], Philipp Schaer[1] and Ralf Schenkel[2]

[1]*TH Köln – University of Applied Sciences, Cologne, Germany*
[2]*Trier University, Trier, Germany*

Abstract
Scientific digital libraries provide their users access to large amounts of data to satisfy diverse information needs. Factors that can influence users' decisions on the relevancy of a publication or a person are individual and usually only visible through posed queries or clicked on information which is influenced by what interfaces offer. The actual formulation or consideration of information requirements begins earlier in users' exploration processes. Hence, this work investigates how to capture the (in)stability of factors supporting such relevancy decisions through users' different levels of manifestation of two exploratory search tasks: expert and paper search. Independent of the information need we found general stability of manifestations and users' tendency to disregard groups of specific factors, such as non-factual or intransparent ones, completely. Relevancy decisions on information objects are multi-faceted, individual considerations. They can be best supported if users of digital libraries would be provided a multiplicity of explainable indicators from which they could consider their preferred ones.

Keywords
Information Seeking Behaviour, User Study, Information Needs, Digital Libraries.

1. Introduction

Bibliographic digital libraries (DLs) such as the ACM DL, Bibsonomy [1], or dblp [2] provide a wide variety of information to their users. Within these systems, users can decide on the relevancy of information objects and satisfy their information needs, e.g., if a scientific publication is relevant to a topic or if a person is an expert. Users consider a multiplicity of relevancy indicators, they determine the relevancy of a document for a task on more than mere topical fit [3]. In general, which indicator is considered relevant is partially dependent on the application domain [4].

Many studies [5, 6, 7, 8, 9] explain or model users' overall information seeking strategies. Factors, which users consider in their relevancy judgements on information objects, change as their cognitive state changes in their information gathering process [10]. However, research on the satisfaction of information needs focuses on general strategies of users [6, 11] or predicting information seeking intentions [12] and less on changes of factors between different representations of information needs.

In this work, we close this gap by presenting a method to investigate the persistence or change of users' considered factors throughout different manifestations in their information seeking

Joint Proceedings of BIR 2024: 14th International Workshop on Bibliometric-enhanced Information Retrieval and IR4U2 2024: 1st Workshop on Information Retrieval for Understudied Users
*Corresponding author.
ckreutz@th-koeln.de (C. K. Kreutz); pschaer@th-koeln.de (P. Schaer); schenkel@uni-trier.de (R. Schenkel)
0000-0002-5075-7699 (C. K. Kreutz); 0000-0002-8817-4632 (P. Schaer); 000-0001-5379-5191 (R. Schenkel)

strategies. The construction of information exploration and retrieval systems can be improved by analysing how users describe their information needs to humans, not only their formalised queries [13]. Additionally, Ingwersen [14] assumes that the change between a verbalisation and conducting a task leads to users compromising their information needs. In general, the longitudinal stability of users' defining factors of information needs is under-researched.

Therefore, we propose to extract and observe key factors from users' (1) general definition of an information need, (2) in an idealised retrieval process, (3) in the actual task solution with an information system and (4) over time. Our research question is: *How can we observe stability of motives of users' information needs in different expressions?* To analyse this, we conduct a user study with thirteen participants with different expertise levels in using DLs for research purposes and two specific information needs of users of DLs: expert search and identification of relevant publications for a topic.

We make the following contributions: 1. We present a qualitative but structured method to evaluate stability of users' information needs in digital libraries. 2. We present a detailed qualitative analysis with thirteen users of digital libraries to identify key factors of typical information needs and their persistence or stability in different levels of expression and over time. 3. We present Re-FIND [15], an extension to the Formalised Information Needs [16] dataset with study participants' re-definitions of expertise and relevancy.

One of the takeaways from our work is a compilation of factors which are typically part of the information exploration processes of users of DLs. This information can help in the design of user interfaces. The factors contributing to the satisfaction of an information need should all be contained in a single system. Users should not be required to change systems when fulfilling their needs, to avoid a possibly negative effect on their speed and cognitive load [17].

2. Related Work

Areas adjacent to our work are general investigations from the domain of *information needs*, *relevancy indicators* and *typical information needs in DLs*.

Information Needs. Research on users' information seeking behaviour and expression of information needs has a rich history: Taylor [9] proposed a four-level continuum to describe the expression of information needs in the context of a person coming up with the formulation and satisfaction of their information need: Q_1 describes the actual visceral and linguistically inexpressible need for some type of information, Q_2 describes the conscious mental description for some type of information, Q_3 describes the verbalised need for some type of information and Q_4 describes the compromised interaction with an information system to satisfy the need for some interaction.

Belkin et al. [18] defined information needs as persons' anomalous states of knowledge, of which the representation is an important aspect of information retrieval research. They constructed representations of information needs which stem from a description of real users' needs in the context of literature search. These representations were then assessed by study participants.

Kuhlthau [19] describes six stages of information seeking processes: initiation, selection, exploration, formulation, collection and presentation. Similarly, Ellis et al. [20] also defined several categories of information behaviour: starting, chaining, browsing, differentiating, monitoring, extracting, verifying and ending. Bates [5] describes six information search stratagems: footnote chasing, citation searching, identifying central venues for areas (journal run), area scanning, subject searches and author searching. Carevic et al. [21, 22] analyse these strategies in single sessions of users of DLs. Wilson [23] describes information seeking behaviour via four stages: passive attention, passive search, active search and ongoing search. Weigl et al. [24] describe the correspondence and gaps between the three models of Ellis [20], Bates [5] and Wilson [23].

Vakkari [25] describes the formulation of information needs as an iterative process. Users acquire new information which then influences their perception of the information space. Taylor [10] examines which different factors users of information search systems consider when making relevancy decisions in an information gathering process and when which factors are relevant. Study participants conducted individual web searches where they chose their search stage (see [7, 20]), the relevancy of results and the relevancy criterion that most influenced their decision out of 19 predefined ones following previous literature [3, 4].

More recent research considers information seeking performed by users as a dynamic process where users' goals and intentions vary depending on their current search or exploration steps [8, 26, 27]. In contrast, we compare different expression levels of complex information needs with multiple relevancy indicators. We investigate and capture changes between these levels of the same two information needs instead of many random information search tasks.

Relevancy Indicators. There have been many studies concerned with observing relevancy and its dimensions in information searching behaviour [3, 4, 28, 29, 30]. They all found numerous aspects as part of users' relevancy decisions such as topic, content, format, representation, under-specified factors, a person's own relationship with a document [3], credibility, language, reputation, scope [30] and users' cognitive characteristics [29].

Barry and Schamber [4] conducted a study in different domains and besides domain-specific factors also found overlap in independent relevancy indicators such as currency, reliability and accessibility. Kato et al. [29] found about half of their study participants not using keywords representing cognitive search intents in their searches even though they were aware of or relevant to them.

Aligning with these studies we also observe a multitude of relevancy indicators which surpass mere topical fit, such as bibliometric relevance signals [31].

Typical Information Needs in DLs. When users are working with DLs, there is a plethora of information needs they could fulfil. One can differentiate between information needs for which clear results without question of relevancy can be found and ones where the distinction of relevant and non-relevant results might not be as straightforward: Clearly formulated narrow information search with verifiable results (also called lookup [32]) could be represented by these queries: what are the co-authors of a person [33, 34, 35] or what is the most cited paper in a venue [34]. Broader information needs with a fuzzy definition of relevancy of results (also called exploratory search [32]) could be the following: what papers are about or fit a specific topic [34, 35, 36, 37]. Exploratory information access oftentimes strives to overcome a searchers anomalous state of knowing [25] and support them in information spaces they have not been previously familiar with [27]. Soufan et al. [32] review the exploratory search task in detail.

3. Concept

We observe different stages of the expression of users' typical information needs in DLs, such as *"what papers are about or fit a specific topic"* [34, 35, 36, 37]. Contrasting recent work [8, 26, 27] we do not only focus on queries posed to DL interfaces. We explicitly assume that the actual visible interaction with a system is only part of the complete information seeking process [14]. An initial query was preceded by internal considerations or general tendencies of users.

Using Taylor's [9] four-level continuum for simplicity, we try to map out the levels to capture: *First*, we observe the personal **definition** of an information need without the satisfaction of the information need in mind, e.g., *"define relevancy of a paper for a topic"*. This is a conscious verbalisation of important factors that a user considers relevant when generally thinking about a specific information need. We expect this manifestation to lie between levels Q_2 and Q_3 as this is a notion of the general requirements, which data is required to satisfy an information need. However, there still is no verbalisation of the task solution strategy itself. *Second*, we observe the **ideal** or general satisfaction of an information need without the restriction of the scope of one specific information system, e.g., *"describe your general process of finding relevant papers from a topic of your choice"*. We regard this as a manifestation of a point between levels Q_3 and Q_4 as there is a conscious verbalisation, but no restrictions are imposed which stem from the specialisation of using one information system to satisfy the information need. A person describing their ideal task solution might not necessarily verbalise the considered factors for their relevance decision. *Third*, we observe the **actual** satisfaction of an information need with the restrictions of one specific information system, e.g., *"use this system to find relevant papers from a topic of your choice"*. We estimate this manifestation to correspond to level Q_4. To solve a task, we assume persons unconsciously use or mention factors that they consider to determine the relevancy of information objects. As a *fourth* manifestation we consider a time-delayed **re-definition** of the general information need, i.e., a second iteration of the first manifestation. This enables analysis of a temporal dimension.

To summarise we suggest observing four different manifestations: the general *definition* of an information need, users' *ideal* task solution, them *actual*ly solving a task and a *re-definition* of users' perception on the information need.

Opposing the scenario described by Taylor [9] in our case the descriptions of the underlying information needs come from an outside force (the study setup) instead of forming organically from a user's anomalous state of knowing [18].

4. User Studies and Data Preparation

This work builds upon and extends the Formalised Information Needs [16] dataset (Sect. 4.1) by another study (Sect. 4.2) with the same participants and tasks.

4.1. Formalised Information Needs Dataset (FIND)

We build upon the Formalised Information Needs [16] dataset (FIND) from Kreutz et al.'s [38] user study. They studied thirteen computer and information scientists with different experience levels of using DLs for research tasks: two masters students, six PhD students (first year to last year students), an industry researcher, a dblp staff member, a postdoc and two professors.

The data was collected in a two-session user study focusing on two tasks derived from typical information needs of users of DLs: *1. Find two experts on a topic of your liking.* and *2. Find relevant papers from a topic of your liking which appeared after 2017.* The general form of the first exploratory search task, the definition of an expert on a topic is extended by the component of finding a second one. The general form of the second exploratory search task, the definition of relevancy of papers on a topic is restricted by a recency component.

In the first session, participants described their definitions of an expert and the relevancy of a paper. The dataset contains this information in the form of transcribed interviews. Users also orally described how they usually conduct the two tasks, the dataset offers this information in formalised business process model notation (BPMN) [39] on the ideal task solution following the description by Law et al. [40]. In a second session, participants verified their ideal strategies and used a specific DL (SchenQL [34, 41]) to then solve the tasks. This process with entered search queries, oral explanations by participants and observations from screen capture is included in the dataset as BPMNs also.

4.2. Re-FIND: extending FIND with Re-Definitions

We conducted interviews with the participants of the study by Kreutz et al. [38] to capture the longevity of information needs and factors of expertise and relevancy of papers. Study participants took part voluntarily and did not receive any incentives. Our 15 to 30 minute online study took place three months after the last study of FIND in a single session. For each of the thirteen participants, an investigator conducted an one-on-one interview.

Step *i)*: Interview. In audio-recorded interviews the participants filled out an online questionnaire with an interviewer. First, the study participants were reminded of the first task and their chosen topic in the study by Kreutz et al. [38]. Then, users defined expertise. In the second part of the interview, participants were reminded of the second task and their chosen topic from FIND. Following this step, the study subjects gave their definition of relevancy of papers.

Step *ii)*: Transcription and Factor Extraction. A person uninvolved in the interviews transcribed the interviews using Otter.ai, re-checked and corrected them. An expert in DLs identified and extracted mentioned factors from the original transcribed **definitions** in FIND and the **re-definitions** from step i) regarding *expertise* and *relevancy of a paper*. For the BPMNs from FIND which describe the **ideal** and **actual** task solution, the same expert extracts considered factors in a semi-normalised language, e.g., in the expert search scenario the un-normalised factors *'resulting paper newer than put in paper/keywords'* and *'check recency of paper'* (of papers fitting keywords on topic) would both be mapped to the semi-normalised factor *'published recently on topic'*.

Step *vi*: Categorisation. For all factors regarding expertise and all factors regarding relevancy of papers, the expert in DLs divided the factors in disjunctive categories per task. A previously uninvolved second expert from the area of information retrieval and DLs then evaluated the categorisation. In cases, where both experts labelled a factor as belonging to a different category, the two discussed until they reached a consensus. This resulted in four main categories per information need: `self-determined (S)`, `other-directed (O)`, `under-specified (U)` and `personal (P)`. The `S` and `O` meta-categories contain eight and two sub-categories for expertise (twelve in total) and three each for relevancy of a paper (eight in total). Table 1 holds information on study participants' mentioned factors in our four (definition, ideal + actual process, re-definition) different manifestations of the two information needs, the meta- and sub-categories.

5. Analysis

To assess *How can we observe stability of motives of users' information needs in different expressions?*, we observe the following more fine-grained questions: 1. What factors do users of DLs define expertise and relevancy of papers with? 2. How do individual users (intend to) apply their general definitions? 3. How stable are individual users' general definitions over time? 4. How stable are individual users' manifestations of information needs?

In our study we refer to participants by IDs (see Fig. 1, e.g., *P4* for *green_deer*).

5.1. Factor Analysis of Definitions

For *What factors do users of DLs define expertise and relevancy of papers with?*, we observe factors of definitions and their categories for both information needs.

Expertise. We identified eight of the twelve categories of factors through which participants in FIND's first study characterise expertise. *Author-dependent*, *collaboration*, *quality* and *searchers' context* factors were not mentioned in any definition. On average each participant mentioned 3.08 factors. Concerning the meta-categories, no `personal` factors were mentioned. The *paper-dependent* category and *external* factors were only mentioned once, *venue*s were considered by two persons, the remaining categories were mentioned by four (*academic*, *citation* and *under-specified* factors), six (*productivity*) or seven (*knowledge*) persons. The single factors which have been mentioned by the highest number of participants (3) were *'knowledge of topics related to topic'* and *'some papers on the topic'*. We encounter only one participant mentioning more than two (3) factors from the same category (*knowledge* factors as described by *P8*). One participant (*P4*) only considers one (*under-specified*) factor in total. Applying frequent itemset mining (and the apriori algorithm [42]) leads to the combination (*knowledge*, *productivity*) with the highest two-element-item support of three occurrences, i.e., this combination was mentioned three times by participants.

Relevancy of Papers. We found all eight categories of factors in participants' definitions of relevancy of papers in FIND. On average each participant mentioned 2.69 factors in their definitions. Factors from all meta-categories were mentioned. An *expert-dependent* factor was mentioned by one study participant. Some categories (*author-dependent*, *external*, *searchers' context* and *under-specified* factors) were considered by two participants each, *venue*-dependent factors were mentioned by five, *citation* factors by six and *paper-dependent* ones by seven participants. The single factor which has been mentioned the most (4) was *'published in good*

Table 1
Number of factors from all categories which participants mentioned in their definition in the initial interview from FIND (d), in their ideal (i) and actual (a) task solution and in the re-definition (r).

		Name	blue_dog				flowergirl				gray_fox				green_deer				Holla_Waldfee				Jigglypuff				Knuspermüsli			
		ID	P1				P2				P3				P4				P5				P6				P7			
	Meta	Category	d	i	a	r	d	i	a	r	d	i	a	r	d	i	a	r	d	i	a	r	d	i	a	r	d	i	a	r
Expertise	S	academic					1					1				2							1	1		2		1		
Expertise	S	author-dependent											1																	
Expertise	S	collaboration						1												1						1				1
Expertise	S	knowledge	2				2	1		3	1											1								1
Expertise	S	paper-dependent			1				3			2	1				1				2			6	1		1	2	1	1
Expertise	S	productivity	1			1	1		1					1		1	1	1		2	1	1	1	1		2	1	1	1	1
Expertise	S	quality				1							1																	
Expertise	S	venue	2			2					1	2	1	2				1										2		
Expertise	O	citation	2	1	3							1	2	1		2	3	2	1	1	2				2	1		1	1	
Expertise	O	external				1		2				1				1						1		1			1			
Expertise	P	searcher's context							1			1																1		
Expertise	U	under-specified		4			1	1				1		1	1	1			1	1		1		1		1		2		1
# Factors considered in scenario			7	5	4	5	5	5	5	3	2	9	6	5	1	7	5	4	2	5	5	4	2	10	3	7	3	10	3	5
Relevancy of Papers	S	author-dependent							1	1		1		1		1						1			1					
Relevancy of Papers	S	paper-dependent		1	1		1	1		4	1	1	1	1	1	3	1			4	1	3	1	4	4	3		1	1	2
Relevancy of Papers	S	venue	1	1		2				1			1	1								1	1			1				1
Relevancy of Papers	O	citation	1	1	2	2							1			2	1	1			1	1		1			1	2	2	1
Relevancy of Papers	O	expert-dependent						1				1							1											1
Relevancy of Papers	O	external				1			1	1								3									2			
Relevancy of Papers	P	searcher's context							1		2	1								1				1		1				
Relevancy of Papers	U	under-specified			1											1														
# Factors considered in scenario			2	3	4	5	1	2	3	7	3	4	3	3	1	7	2	4	1	5	2	6	2	6	5	5	3	3	3	5

		Name	leesedol				MaxMustermann				Pink Rabbit				SchlauerFuchs				TigerBalm				wild bear			
		ID	P8				P9				P10				P11				P12				P13			
	Meta	Category	d	i	a	r	d	i	a	r	d	i	a	r	d	i	a	r	d	i	a	r	d	i	a	r
Expertise	S	academic				1	1			1		2		1		1				1	1		2	2		
Expertise	S	author-dependent								1						1	1			1						
Expertise	S	collaboration						1													1					1
Expertise	S	knowledge	3				1			1	2			1	2					1	1			1		1
Expertise	S	paper-dependent		1	1			2	3			1	2				4			1	2			1	3	
Expertise	S	productivity		1	1	1	2	2		1		1		1		3	4	2	2	2	2	2		2	1	2
Expertise	S	quality		1																1						
Expertise	S	venue						1		1		1				1		1				1		1		
Expertise	O	citation	1	2	1				1			1	1			1	2	2		2	2		1	1		
Expertise	O	external		1		1														1						
Expertise	P	searcher's context		1	3				2			2	1			1	1				1					
Expertise	U	under-specified		1		1	1	1	1			1												4		
# Factors considered in scenario			4	8	6	4	5	7	7	5	2	9	4	3	2	8	12	5	2	10	10	3	3	12	4	4
Relevancy of Papers	S	author-dependent						5	2		1	1		2				1		2		1	1			
Relevancy of Papers	S	paper-dependent	4	5	3	3		1	2	4		1	2	1	1	3	3	2	1	2	4	3	2	4	4	
Relevancy of Papers	S	venue					1				1	1	1	1		1		1		1		1	1	2	1	1
Relevancy of Papers	O	citation		1	2		1		1		1	1					2	1	1			1	1	1		
Relevancy of Papers	O	expert-dependent						1												1						
Relevancy of Papers	O	external					1															1				
Relevancy of Papers	P	searcher's context		1		2			1	1									1							1
Relevancy of Papers	U	under-specified		2		1	2		2	1							1						1	2		
# Factors considered in scenario			4	9	5	6	5	7	8	6	3	4	3	4	1	4	6	5	3	6	4	7	6	9	5	2

conference'. Three persons each only gave one *paper-dependent* factor in their definitions, another participant (*P8*) also only considered this category but formulated four distinct factors. One participant (*P5*) gave an *expert-dependent* factor as their single factor. Applying frequent itemset mining (and the apriori algorithm [42]) leads to the combination (*citation*, *venue*) with the highest two-element-item support of four occurrences, i.e., this combination of categories is mentioned four times by participants.

Discussion. Participants tend to only consider few factors in their abstract definitions for expertise and relevancy of papers. For papers, the amount of considered factors is lower than for expertise. One participant (*P4*) seems to only give brief explanations while another seems to elaborately focus on one category in particular (*P8*). The other participants seem to brush multiple categories with one or two relevant factors for each. For most participants no clear tendencies besides a liking for *knowledge* in the expert search and for *paper-dependent* factors in the relevancy of papers task could be found.

We conclude that the factors users' define expertise and relevancy of papers with are very diverse and highly individual. For both information needs the study participants mostly describe factors which are from the `self-determined` meta-category. That means information which is solely dependent on the expert or relevant paper candidate, not outside factors which a person or paper cannot directly affect such as citation counts.

5.2. (Intended) Application of Definitions

To answer *How do individual users (intend to) apply their general definitions?*, we align users' definitions with their ideal and actual task solution strategies.

Definitions - Ideal Task Solution. **Expertise.** Factors from the categories *author-dependent*, *collaboration*, *quality* and *searchers' context* were not described in participants' definitions but were mentioned in either five (*searchers' context*), three (*collaboration*) or two (*author-dependent*, *quality*) ideal task solutions. On average participants mentioned 8.08 factors in ideal solutions. In 17 cases participants described factors from a category in the definitions, that has also been present in their ideal task description. Aside from two participants, all others have at least one category present in both manifestations. Six participants described the same two categories in both their definitions and ideal task solution.

Relevancy of Papers. On average participants stated 5.31 factors in ideal task solutions. In 13 cases we found the categories which participants mentioned in their initial definitions again in their ideal task descriptions. Only two participants had no overlap between their definitions and ideal task description.

Discussion. Especially for the paper relevancy task participants mentioned considerably more factors compared to their initial definitions. This might stem from the data collection process for this manifestation. Participants were able to verify, extend and modify their previously only verbalised definitions. These models were visualisations which might have lead to them being very aware of all factors and potentially over-modelling their ideal task solution. This observation is supported by the statement of a participant in the post-task questionnaire of FIND: *"I strongly idealized my search behaviour. (...) I had the impression that my* real *search behaviour is much simpler."*. **We found many overlaps between users' definitions and ideal task solutions, but the ideal processes were more detailed and contained considerably more aspects.**

Definition - Actual Task Solution. **Expertise.** Nine participants used factors from a category when conducting the task which is also part of their definitions. Five persons did not have any overlap in categories.Four participants actually used two categories from the definitions, one used three of their previously described categories. On average participants used 5.69 factors in expert search. In general the utilised or described categories mostly consist of factors from the `self-determined` and in a smaller amount from the `other-directed` meta-category. Factors from the `under-specified` meta-category seem to be considerably less important in the actual task solution (see Fig. 1, upper part).

Relevancy of Papers. Eleven participants used factors from a category when doing the task which they have also mentioned in their previous definitions. Participants used 4.08 factors on average. Here factors from the `under-specified` meta-category also seemed to be irrelevant.

Discussion. We found many overlaps in definitions and task solutions, especially for relevancy of papers. However, we found study participants' disregard for `under-specified` factors in actually solving the task which were mentioned in definitions. This might be attributed to the fact that participants had to really solve the tasks and had to formulate or check information which was clearly visible in the digital library interface. A contributing factor for some differences between definitions and actual task solution might be caused by users intentionally varying their search tactic and information goals depending on the used collection [43].

5.3. Stability of Definitions over Time

For answering *How stable are individual users' general definitions over time?*, we observe the overlap of between factors of definitions and re-definitions as well as their categories for both information needs individually.

Expertise. On average participants describe 4.38 factors in re-definitions. Five participants have no overlap in categories between their two definitions. Four persons have overlaps in at least two categories. Five persons mentioned the exact same single factor in both their definitions and their re-definitions.

Relevancy of Papers. Participants mentioned 5 factors on average in their re-definitions. Two participants had no overlapping categories between their two definitions. Four persons' factors overlapped in two categories. Four persons mentioned one exact same factor in both their definitions and their re-definitions.

Discussion. Definitions of participants were typically much shorter than their re-definitions, we did find considerably more factors in re-definitions for both tasks. Likewise, the area spanned by the meta-categories' percentages in Fig. 1 is larger for re-definitions. **We found considerable similarities, not only on the described categories but also on the individual factor level.**

5.4. Stability over Manifestations

For answering *How stable are individual users' manifestations of information needs?*, we observe the similarities and differences between users' definitions, ideal strategies and actual solutions for the two information needs.

Expertise. Ten participants used factors belonging to the meta-category `self-determined` in all four observed settings. Factors belonging to the meta-category `other-directed` were mentioned by three participants in all settings. The remaining meta-categories were not present throughout all settings for any of the study participants. On the level of single factors we

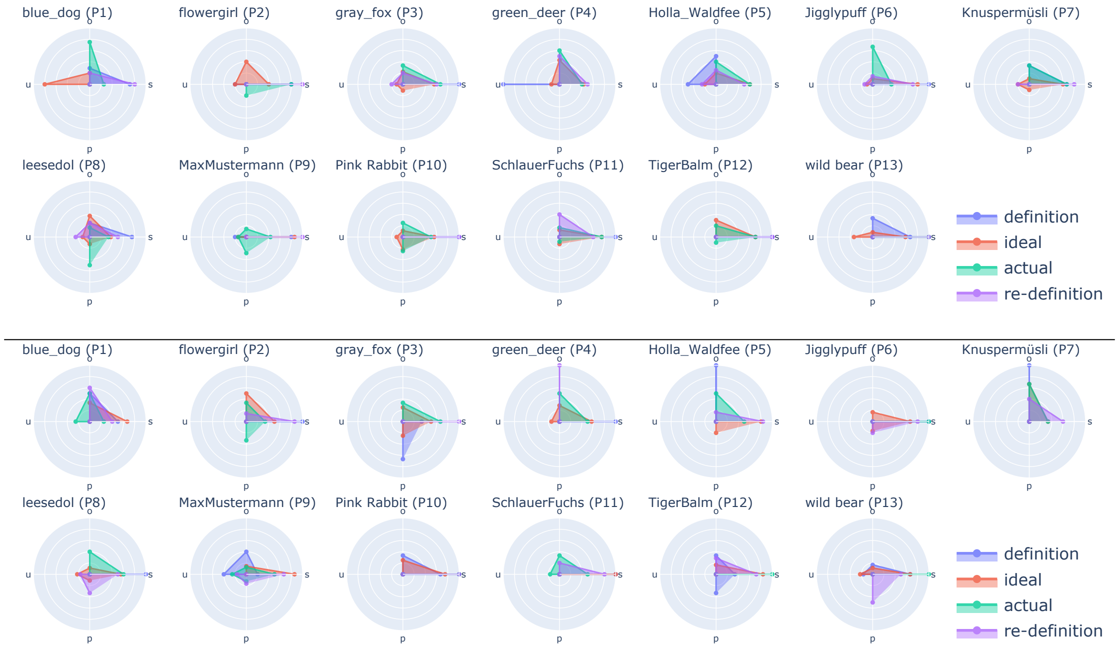

Figure 1: Meta-categories' presence in the four manifestations of expertise (upper 13) and relevancy (lower 13) for all participants and their IDs which they are referred to.

found some present throughout all four settings for few participants: *productivity* for two and *paper-dependent* as well as *venue* for one person. Fig. 1 (upper part) visualises the percentage of study participants' manifestations with regards to our four defined meta-categories. We observe some participants having very comparable percentages for meta-categories `self-defined` and `other-directed` (*P3*, *P4* and *P11*) and some participants roughly having the same core representations in some representations (*P5*, *P7*, *P8*, *P9*, *P10* and *P12*). The reminder of the participants have unclear or very diffuse strategies over the four representations.

Relevancy of Papers. Ten participants used factors belonging to the meta-category `self-determined` in all four observed settings. Factors belonging to the meta-category `other-directed` (and even the same category, namely *citation*) were mentioned by two participants in all settings. We found some other single categories present in all four manifestations: the *paper-dependent* category is considered by five participants, *venue* is always considered by one person. Fig. 1 (lower part) visualises the percentage of study participants' manifestations with regards to the meta-categories. Few participants have very comparable percentages for meta-categories `self-defined` and `other-directed` (*P1* and *P12*), some participants roughly have the same core meta-categories in some representations (*P3*, *P4*, *P6*, *P9*, *P10*, *P11* and *P13*). The reminder of the participants have rather unclear or very diffuse strategies over the four representations.

Discussion. We see clear tendencies of study participants over all four data points. Participants' tend to use or disregard complete categories over both tasks. Four study participants (*P1*, *P4*, *P5*, *P13*) almost completely ignore searchers' context or external factors. These persons seem to prefer quantifiable, clearly defined and explainable factors which do not depend on some vaguely defined outside entity. This could, e.g., show by taking a look at meta-category `other-directed`: search engines' ranking of a paper (category *external*) would be disregarded while the number of citations a paper received (category *citation*) could be a considered. Three study participants (*P10*, *P11*, *P12*) tend to clearly describe factors they consider relevant, they at most mention one factor from the *under-specified* category over all manifestations and both tasks. From these observations we conclude the individual's considered factors being somewhat stable independent from the actual information needs.

For expertise we found factors from these categories mentioned the most: *productivity* (52), *citation* (44) and *paper-dependent* (43). For relevancy of papers we found factors from these categories mainly mentioned: *paper-dependent* (96) and *citation* (36). **In general, we found different usage patterns of participants and considerable stability of (meta-) categories.**

5.5. Discussion

Concerning our overall question of how to capture the stability of motives in different levels of manifestations of users' information needs, we observed generally stable meta-categories. **We observe that the decision if someone is considered an expert or a paper is relevant is a multi-faceted decision, independent of the current manifestation of the information need.** DL interfaces should support users better by providing a multiplicity of factors for them to pick the ones they want to base their decision regarding an information object on. Clearly quantifiable or explainable factors such as citation counts or authors' affiliations seem to play a bigger role in decision making compared to highly personal factors, e.g., if a paper at hand tackles a user's research interest. Participants mentioning under-specified factors might be attributed to the general feeling of knowing [44] and the narrower tip of the tongue [45] effect. Sometimes they even mentioned their feeling of missing an additional aspect in their definitions, e.g., *P13* wanted to refer to a specific indicator but stated they *'...forgot all the names'*. Relevancy indicators in actual task solution could be impacted by users' greater ease to formulate requests directed at humans as in the other three scenarios compared to a DL system [13].

We identified part of Kato et al.'s [29] characterisation of cognitive search intents in factors mentioned in our study participants' factors of definitions (twice), ideal solution (once) and re-definitions (once) for the relevancy of papers task.

Our observed categories can be aligned with the search stratagems defined by Bates [5]: footnote chasing and citation searching are represented by the category `citation`, journal run is described in the category `venue`, subject searches and area scanning are both merged in `paper-dependent` and author searching is described by the category `author-dependent`.

Further, we found that representing users' ideal and actual task solution via BPMNs following Law et al.'s [40] method is suitable for finding factors and categories of information needs.

Finally, our evaluation is limited by the small amount of study participants and their diversity in expertise; academic positions [46] and experience [47] both influence users' search behaviour.

6. Limitations

Our study's meaningfulness could be limited by a number of factors which are discussed in the following: We only observe a *small amount of study participants*. They are investigated in detail over multiple setting and in context of two different information needs. The universality of our findings should be investigated in bigger sized user groups and with a more diverse set of information needs.

Our study participants have *different academic positions*, they range from students to professors. Academic position [46] and experience [47] influences users' search behaviour. Less research-experienced persons could be unfamiliar with expert search.

Participants could have differences in their definitions of expertise and relevancy of papers due to them *learning* more about these topics during the three month period between Study I and Study II. This could be especially true for the two masters students but then in terms should be a less pronounced effect when observing the definitions of both professors. Observing the respective participants' data with this knowledge does not support this assumption.

The BPMNs describing users' ideal and performed strategy to satisfy information needs contain *non-standardised data*. The expressions used in the models are directly derived from study participants' verbalisations [40]. This could lead to ambiguities when extracting factors from the models.

The utilised dimensions in the performed strategy to satisfy the information needs with an example system at hand could be influenced by the *options the system offers*. A counter-argument for this assumption are users describing their intentions independent of what the system offers, e.g. them stating they are searching on how to order the displayed persons by their h index.

7. Conclusion

Our guiding question *How can we observe stability of motives of users' information needs in different expressions?* was approximated with studying four manifestations of two typical task of digital library users. We found similarities and differences between different levels of manifestations. `Self-determined` and `other-defined` factors generally seemed to be favoured over ones belonging to the `under-specified` or `personal` meta-categories. Study participants displayed tendencies to consider or completely disregard categories over both tasks which hints at this preference being task-independent. For the expertise task we found twelve categories of factors which users considered throughout all four representations, the most prominent ones being persons' *productivity*, *citation*s and *paper-dependent* factors. For relevancy of papers we found eight categories with the most popular one by far being the *paper-dependent* one followed by *citation*s. Satisfying our observed typical information needs required a multiplicity of factors which should all be provided by DLs for users' convenience.

One future direction would be conducting a quantitative study using the found (meta-) categories and factors to investigate user groups of different expertise levels. A goal could be the identification of prevalence for specific categories of user groups and an analysis of importance of these dimensions in the actual decision making processes. A complementary line of research could be the formalisation of underlying factors and categories of the task solution strategies into a user model. This could be useful for simulating the relevancy decision on results in information access. Additionally, the formalisation would deliver testable hypotheses on user behaviour from which we could gain insights on the validity of our assumptions on users [48].

References

[1] A. Hotho, R. Jäschke, D. Benz, M. Grahl, B. Krause, C. Schmitz, G. Stumme, Social bookmarking am beispiel bibsonomy, in: Social Semantic Web: Web 2.0 - Was nun?, 2009, pp. 363–391. doi:10.1007/978-3-540-72216-8_18.

[2] M. Ley, DBLP - some lessons learned, Proc. VLDB Endow. 2 (2009) 1493–1500. doi:10.14778/1687553.1687577.

[3] C. Cool, N. J. Belkin, O. Frieder, P. P. Kantor, Characteristics of texts affecting relevance judgments, 1993.

[4] C. L. Barry, L. Schamber, Users' criteria for relevance evaluation: A cross-situational comparison, Inf. Process. Manag. 34 (1998) 219–236. doi:10.1016/S0306-4573(97)00078-2.

[5] M. J. Bates, The design of browsing and berrypicking techniques for the online search interface, Online Review 13(5) (1989) 407–424.

[6] O. Hoeber, D. Storie, Graduate student search strategies within academic digital libraries, International Journal on Digital Libraries (2023) 1–14.

[7] C. C. Kuhlthau, Seeking Meaning: A Process Approach to Library and Information Services, 2004. URL: https://books.google.de/books?id=feDgAAAAMAAJ.

[8] J. Liu, C. Shah, Leveraging user interaction signals and task state information in adaptively optimizing usefulness-oriented search sessions, in: JCDL '22, 2022, p. 23. doi:10.1145/3529372.3530926.

[9] R. S. Taylor, Question-negotiation and information seeking in libraries, Coll. Res. Libr. 76 (1968) 251–267.

[10] A. R. Taylor, User relevance criteria choices and the information search process, Inf. Process. Manag. 48 (2012) 136–153. doi:10.1016/j.ipm.2011.04.005.

[11] O. Hoeber, D. Patel, D. Storie, A study of academic search scenarios and information seeking behaviour, in: CHIIR '19, 2019, pp. 231–235. doi:10.1145/3295750.3298943.

[12] M. Mitsui, J. Liu, N. J. Belkin, C. Shah, Predicting information seeking intentions from search behaviors, in: SIGIR '17, 2017, pp. 1121–1124. URL: https://doi.org/10.1145/3077136.3080737. doi:10.1145/3077136.3080737.

[13] J. Arguello, A. Ferguson, E. Fine, B. Mitra, H. Zamani, F. Diaz, Tip of the tongue known-item retrieval: A case study in movie identification, in: CHIIR '21, 2021, pp. 5–14. doi:10.1145/3406522.3446021.

[14] P. Ingwersen, Search procedures in the library - analysed from the cognitive point of view, Journal of Documentation 38 (1982) 165–191. doi:10.1108/eb026727.

[15] C. K. Kreutz, S. Myshkina, P. Schaer, R. Schenkel, Re-FIND: Re-Definitions for the Formalised Information Needs dataset, 2024. URL: https://doi.org/10.5281/zenodo.10791641. doi:10.5281/zenodo.10791641.

[16] C. K. Kreutz, S. Myshkina, M. Blum, P. Schaer, R. Schenkel, B. Weyers, Formalised information needs dataset, 2023. doi:10.5281/zenodo.7826530.

[17] S. Jeuris, J. E. Bardram, Dedicated workspaces: Faster resumption times and reduced cognitive load in sequential multitasking, Comput. Hum. Behav. 62 (2016) 404–414. doi:10.1016/j.chb.2016.03.059.

[18] N. J. Belkin, R. N. Oddy, H. M. Brooks, Ask for information retrieval: Part II. results of a

design study, J. Documentation 38 (1982) 145–164. doi:10.1108/eb026726.
[19] C. C. Kuhlthau, Inside the search process: Information seeking from the user's perspective, J. Am. Soc. Inf. Sci. 42 (1991) 361–371. doi:10.1002/\%28SICI\%291097-4571\%28199106\%2942\%3A5\%3C361\%3A\%3AAID-ASI6\%3E3.0.CO\%3B2-\%23.
[20] D. Ellis, D. Cox, K. Hall, A comparison of the information seeking patterns of researchers in the physical and social sciences, J. Documentation 49 (1993) 356–369. doi:10.1108/eb026919.
[21] Z. Carevic, M. Lusky, W. van Hoek, P. Mayr, Investigating exploratory search activities based on the stratagem level in digital libraries, Int. J. Digit. Libr. 19 (2018) 231–251. doi:10.1007/s00799-017-0226-6.
[22] Z. Carevic, P. Mayr, Extending search facilities via bibliometric-enhanced stratagems, in: BIR@ECIR '15, volume 1344, 2015, pp. 40–46. URL: https://ceur-ws.org/Vol-1344/paper5.pdf.
[23] T. D. Wilson, Information behaviour: an interdisciplinary perspective, Inf. Process. Manag. 33 (1997) 551–572. doi:10.1016/S0306-4573(97)00028-9.
[24] D. M. Weigl, K. R. Page, P. Organisciak, J. S. Downie, Information-seeking in large-scale digital libraries: Strategies for scholarly workset creation, in: JCDL '17, 2017, pp. 253–256. doi:10.1109/JCDL.2017.7991583.
[25] P. Vakkari, Exploratory searching as conceptual exploration (2010).
[26] L. Azzopardi, Modelling interaction with economic models of search, in: SIGIR '14, 2014, pp. 3–12. doi:10.1145/2600428.2609574.
[27] T. Ruotsalo, J. Peltonen, M. J. A. Eugster, D. Glowacka, P. Floréen, P. Myllymäki, G. Jacucci, S. Kaski, Interactive intent modeling for exploratory search, ACM Trans. Inf. Syst. 36 (2018) 44:1–44:46. doi:10.1145/3231593.
[28] J. Arguello, B. Choi, R. Capra, Factors influencing users' information requests: Medium, target, and extra-topical dimension, ACM Trans. Inf. Syst. 36 (2018) 41:1–41:37. doi:10.1145/3209624.
[29] M. P. Kato, T. Yamamoto, H. Ohshima, K. Tanaka, Investigating users' query formulations for cognitive search intents, in: SIGIR '14, 2014, pp. 577–586. doi:10.1145/2600428.2609566.
[30] I. Xie, E. B. III, Search result list evaluation versus document evaluation: similarities and differences, J. Documentation 69 (2013) 49–80. doi:10.1108/00220411311295324.
[31] T. Breuer, P. Schaer, D. Tunger, Relevance assessments, bibliometrics, and altmetrics: a quantitative study on pubmed and arxiv, Scientometrics 127 (2022) 2455–2478. doi:10.1007/s11192-022-04319-4.
[32] A. Soufan, I. Ruthven, L. Azzopardi, Searching the literature: An analysis of an exploratory search task, in: CHIIR '22, 2022, pp. 146–157. doi:10.1145/3498366.3505818.
[33] S. Gómez-Villamor, G. Soldevila-Miranda, A. Giménez-Vañó, N. Martínez-Bazan, V. Muntés-Mulero, J. L. Larriba-Pey, BIBEX: a bibliographic exploration tool based on the DEX graph query engine, in: EDBT '08, volume 261, 2008, pp. 735–739. doi:10.1145/1353343.1353439.
[34] C. K. Kreutz, M. Blum, R. Schenkel, SchenQL: a query language for bibliographic data with aggregations and domain-specific functions, in: JCDL '22, 2022, p. 37. doi:10.1145/3529372.3533282.

[35] Y. Zhu, Graph-based interactive bibliographic information retrieval systems, 2017.
[36] C. Betts, J. Power, W. Ammar, Grapal: Connecting the dots in scientific literature, in: ACL '19, 2019, pp. 147–152. doi:10.18653/v1/p19-3025.
[37] S. Bloehdorn, P. Cimiano, A. Duke, P. Haase, J. Heizmann, I. Thurlow, J. Völker, Ontology-based question answering for digital libraries, in: ECDL '07, volume 4675, 2007, pp. 14–25. doi:10.1007/978-3-540-74851-9_2.
[38] C. K. Kreutz, M. Blum, P. Schaer, R. Schenkel, B. Weyers, Evaluating digital library search systems by using formal process modelling, in: JCDL '23, 2023, pp. 1–12. URL: https://doi.org/10.1109/JCDL57899.2023.00058. doi:10.1109/JCDL57899.2023.00058.
[39] T. Allweyer, BPMN 2.0 - Business Process Model and Notation: Einführung in den Standard für die Geschäftsprozessmodellierung, 2020.
[40] Y. C. Law, W. Wehrt, S. Sonnentag, B. Weyers, Obtaining semi-formal models from qualitative data: From interviews into bpmn models in user-centered design processes, International Journal of Human–Computer Interaction 39 (2023) 476–493. doi:10.1080/10447318.2022.2041899. arXiv:https://doi.org/10.1080/10447318.2022.2041899.
[41] C. K. Kreutz, M. Wolz, J. Knack, B. Weyers, R. Schenkel, SchenQL: in-depth analysis of a query language for bibliographic metadata, Int. J. Digit. Libr. 23 (2022) 113–132. doi:10.1007/s00799-021-00317-8.
[42] R. Agrawal, T. Imielinski, A. N. Swami, Mining association rules between sets of items in large databases, in: SIGMOD '93, 1993, pp. 207–216. doi:10.1145/170035.170072.
[43] D. McKay, G. Buchanan, One of these things is not like the others: How users search different information resources, in: TPDL '11, volume 6966, 2011, pp. 260–271. doi:10.1007/978-3-642-24469-8_28.
[44] J. T. Hart, Memory and the feeling-of-knowing experience., Journal of educational psychology 56 4 (1965) 208–16.
[45] R. S. Brown, D. McNeill, The "tip of the tongue" phenomenon, Journal of Verbal Learning and Verbal Behavior 5 (1966) 325–337.
[46] X. Niu, B. M. Hemminger, A study of factors that affect the information-seeking behavior of academic scientists, J. Assoc. Inf. Sci. Technol. 63 (2012) 336–353. doi:10.1002/asi.21669.
[47] C. Hsin, Y. Cheng, C. Tsai, Searching and sourcing online academic literature: Comparisons of doctoral students and junior faculty in education, Online Inf. Rev. 40 (2016) 979–997. doi:10.1108/OIR-11-2015-0354.
[48] K. Balog, C. Zhai, User simulation for evaluating information access systems, 2023. arXiv:2306.08550.