Information Search as Knowledge Gain: Towards New Measures

Nilavra Bhattacharya
The University of Texas at Austin
Unites States
nilavra@ieee.org

Jacek Gwizdka
The University of Texas at Austin
Unites States
chiir2021@gwizdka.com

ABSTRACT

Search engines and information retrieval systems are becoming increasingly important as educational platforms to foster learning. Modern search systems still have room to improve in this regard. We posit that learning-during-search is a good candidate for a human-centred metric for information seeking. We discuss ways to measure learning, and propose a conceptual framework for describing searchers' knowledge-change during search. We stress the need for developing better measures for the search process, and discuss why we need to rethink the existing models of information seeking. We conclude by sharing our own experiences that guided our recommendations in this position paper.

CCS CONCEPTS

• **Information systems → Users and interactive retrieval**.

KEYWORDS

search as learning; measuring search process; measuring learning

ACM Reference Format:
Nilavra Bhattacharya and Jacek Gwizdka. 2021. Information Search as Knowledge Gain: Towards New Measures. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

As early as in 1980, Bertam Brookes, in his 'fundamental equation' of information and knowledge: $K[S] + \Delta I = K[S + \Delta S]$ had stated that a searcher's current state of knowledge, $K[S]$, is changed to the new knowledge structure, $K[S + \Delta S]$, by exposure to information ΔI, with the ΔS indicating the effect of the change [8, p. 131]. This indicates that searchers acquire new knowledge in the search process, and the same information ΔI may have different effects on different searchers' knowledge states. Fifteen years later, Marchionini described information seeking as "a process, in which humans purposefully engage in order to change their state of knowledge" [25]. Thus, we have known for quite a while that search is driven by higher-level human needs, and Information Retrieval (IR) is a means to an end, and not the end in itself.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

When we consider information seeking as a process that changes the searcher's knowledge-state, the question arises whether the assessment of knowledge-acquisition-during-search, or *learning*, should subsume the standard IR evaluation metrics and the search interface usability metrics. It seems that to diagnose a problem or to understand a success of a search system, we would still need to control the standard aspects of a search system (e.g., results ranking, search user interface design features). However, a direct assessment of these "lower-level" aspects would lose on importance. On the other hand, support for more rapid learning across a number of searchers, and over a range of different search tasks can be indicative of an IR system that is more effective at supporting intelligence amplification and knowledge building [38]. In the last decade, this recognition that IR systems of tomorrow can become "rich learning spaces" and foster knowledge gain, has led to the emergence of the Search as Learning (SAL) research community [29], and the need to consider learning-during-search as a metric for evaluation of Interactive IR (IIR) systems.

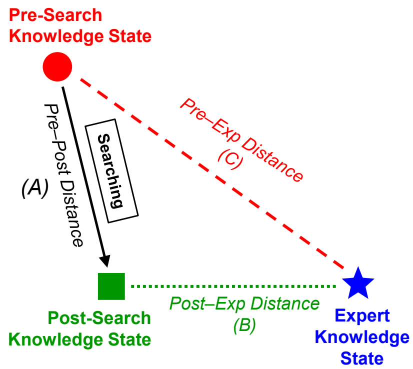
2 METRICS FOR LEARNING & KNOWLEDGE

2.1 Experts vs. Novices

If we consider learning-during-search to be a good candidate for IR evaluation criterion, the next challenge is **how to measure learning**, or knowledge acquisition, possibly in an automated fashion. We can turn to educational psychology literature. A research report by the US National Research Council [10] identified the following key principles about experts' knowledge, illustrating the results of successful knowledge acquisition:

(1) "Experts notice features and meaningful patterns of information that are not noticed by novices."
(2) "Experts have acquired a great deal of content knowledge that is organized in ways that reflect a deep understanding of their subject matter."
(3) "Experts' knowledge cannot be reduced to sets of isolated facts or propositions but, instead, reflects contexts of applicability: that is, the knowledge is 'conditionalized' on a set of circumstances."
(4) "Experts are able to flexibly retrieve important aspects of their knowledge with little attentional effort."

Some of the above findings have been used by our community in the past. E.g, user learning has been measured by user's familiarity with concepts and relationships between concepts [28], gains in user's understanding of the topic structure [41], and user's ability to formulate more effective queries [9, 28]. From the above findings, we can think about ways to consider *Expert's Knowledge on the search topic* as 'gold-standard' or 'ground-truth' (by algorithmic parlance), for developing learning based IIR evaluation metrics.



Row No.	Pre-Post Dist (A)	Post-Exp Dist (B)	Pre-Exp Dist (C)	[Level of Knowledge *Change*] Intuition
1	Low	Low	Low	[Low] Expert
2	Low	Low	HIGH	X
3	Low	HIGH	Low	X
4	Low	HIGH	HIGH	[Low] Slow learner
5	HIGH	Low	Low	X
6	**HIGH**	**Low**	**HIGH**	[High] User gained new knowledge; **desirable**
7	HIGH	HIGH	Low	[High] **Knowledge loss:** started near expert-knowledge state before search; went further away from expert knowledge after search
8	HIGH	HIGH	HIGH	[High] **Mis-directed search/learning:** started away from expert-knowledge; went further away

Figure 1: A conceptual framework for knowledge-change during search. We assume Pre-Search Knowledge, Post-Search Knowledge, and the reference Expert Knowledge as three vertices of a triangle (left figure). Ifwe can compute the distance between the triangle's vertices, and then further dichotomize these distances as HIGH vs. Low, then we can have eight possible outcomes (right table). 'X' denotes outcomes violating the triangle inequality.

2.2 Measuring Knowledge-Change

Recent literature on Search-as-Learning adopts three broad approaches to measure learning, or knowledge-change, with their own strengths and limitations. The **first approach** asks searchers to rate their self-perceived pre-search and post-search knowledge levels [14, 26]. This approach is the easiest to construct, and can be generalised over any search topic. However, self-perceptions may not objectively represent true learning. The **second approach** tests searchers' knowledge using factual multiple choice questions (MCQs). The answer options can be a mixture of fact-based responses (*TRUE*, *FALSE*, or *I DON'T KNOW*), [13, 40] or recall-based responses (*I remember / don't remember seeing this information*) [23, 30]. Constructing topic-dependant MCQs may take time and effort, which may be aided by automated question generation techniques[35]. For evaluation, this approach is the easiest, and often automated. However, MCQs allow respondents to answer correctly by guesswork. The **third approach** lets searchers write natural language summaries or short answers, before and after the search [2, 26]. Depending on experimental design, prompts for writing such responses can be generic (least effort) [3] or topic-specific (some effort) [35]. While this approach can provide the richest information about the searcher's knowledge state, evaluating such responses is the most challenging.

2.3 Proposed Conceptual Framework

We can think of a conceptual framework for a searcher's knowledge change during search (**Fig. 1**)[1]. Searchers initiate a search session with a Pre-Search Knowledge state. During search, they undergo a change in knowledge. On conclusion of search, searchers attain the Post-Search Knowledge state. We can attempt to measure this dynamic knowledge-change from a stationary reference point: Expert Knowledge on the search topic (ground-truth). If we imagine these three knowledge states to be the three vertices of a triangle (Fig. 1, left), and if, by some hypothetical metric, we can compute the *distance* between any two of these knowledge-state points, then we have found a way to quantify learning-during-search.

Moving further, if we dichotomize the learning-during-search as 'HIGH' vs 'Low', by establishing a threshold value for the distances, then we can obtain eight possible knowledge-change situations (Fig. 1, right table). Three of these eight situations violate the triangle inequality[2] (denoted by 'X' in the table), and are therefore discarded. The remaining five valid situations are discussed below.

When Pre-Search Knowledge State and Post-Search Knowledge State are both very 'close' to Expert Knowledge (row 1 in table), we can assume the searcher is an **expert**. On the other hand, if Pre-Search Knowledge State and Post-Search Knowledge State are close to each other, but are far away from Expert Knowledge (row 4), the searcher is probably a **novice**, and also a **slow learner**, because on conclusion of search, their knowledge still remained far away from Expert Level. When the Post-Search Knowledge is closer to Expert than Pre-Search Knowledge (row 6), it implies that the searcher gained 'good amount' of new knowledge, and is thus, the **most desirable situation for IIR**.

The last two rows of the table in Fig. 1 present two interesting, albeit undesirable, possibilities. If the Pre-Search Knowledge is closer to Expert, but the Post-Search Knowledge is further away (row 7), it can signify **knowledge loss** (which is also a form of knowledge *change*). On the other hand, if both the Pre-Search and the Post-Search knowledge are far away from Expert, and they are also far away from each other (row 8), then it is a case of **misdirected search**, and therefore, **misdirected learning**. A classic illustration of these two situations is health information seeking. Suppose a user is searching for cause and treatment of a small brownish spot on the wrist. If a physician examined the spot, they would immediately identify the spot to be caused by oil-splatter burn during cooking (Expert Knowledge State). The searcher may however, based on search results, come to the incorrect conclusion that they have skin cancer [1, 36]. Before the search, if the searcher

[1]This is a conceptual expansion of our work in [4].

[2]sum of lengths of any two sides of a triangle is greater than the third side

correctly guessed that the spot was due to oil splatter burn, then the situation would be described by row 7 (knowledge loss, or increase in confusion), whereas if the searcher had no intuition about the cause of the spot before the search, the situation would be described by row 8. Both situations should be avoided by modern IIR systems.

3 MEASURING THE SEARCH PROCESS

Learning-during-search involves two intertwined activities: learning, and searching. In Sec. 2, we discussed approaches to measure learning. The other part of the picture involves measuring the search process itself. Past research efforts has largely been devoted to measuring **search outcomes**: e.g., if a target document was reached, or if relevant results were shown. We argue that a more human-centred approach for measuring search is to try and quantify the **search process**.

3.1 Need for Longitudinal Studies

A major limitation of most IIR research efforts is that the user is examined in the short-term, typically over the course of a single lab session. The trend is similar in other HCI research venues. Kelly et al. [21] stressed the need for longitudinal designs over a decade ago, yet a meta-analysis of 1014 user studies reported in the ACM CHI 2020 conference revealed that more than 85% of the studies observed participants for a day or less. To this day, "longitudinal studies are the exception rather than the norm" [22]. On the other hand, it is quite evident that knowledge acquisition is a longitudinal process, occurring gradually over time. Therefore, most educational curricula in schools and universities are spread across several months and years. "An over-reliance on short studies risks inaccurate findings, potentially resulting in prematurely embracing or disregarding new concepts" [22].

3.2 Need for Updated Theoretical Models

The Information Seeking literature is dominated by a large number of "multiple arrow-and-box" theoretical models. These models divide the information seeking process for complex search-tasks into different stages. Some argue that these models are not not "real models" but more of "short-hand common-sense task flows" [11, 12]. The mantra of these models have always been the same: they have "implications for systems design and practice". Unfortunately, these models, along with a significant body of IIR research, has not been able to go beyond suggestions, to providing concrete design solutions [31]. Moreover, there is great overlap in basic search strategies across many of these models [18], calling into question whether so many models are still relevant. Consequently, current search systems still predominantly use a "one-size-fits-all" approach: one interface is used for all stages of a search, even for complex search endeavours [20].

Again reiterating Kelly et al. [21], we posit that these models, theorised decades ago for bulky desktop computers, are in need of improvement. Information seeking models have to incorporate the continuous or lifelong nature of online information searching, enabled by the proliferation of internet access in various handheld and portable digital devices. For instance, Marchionini [25]'s well known information seeking process (ISP) models the information seeking behaviour into eight stages, with connecting feed-forward and feed-back loops between the stages. However, some argue that users never really go "back" to an earlier state; e.g., "when reformulating the query, users do not really go back to the initial situation, they submit an improved query" [37]. With progress of time, there is continuous update of users' *information need* [19] and *search context* [32]. Thus, the intricate relationships between users' knowledge state, cognitive state, and other factors influencing search (search *context*), are ever-changing. Perhaps then Spink [34]'s model of the IR interaction process, which models interactive search as an infinite continuous process of sequential steps, or cycles[3], is better suited to explain information searching behaviour. Like time, there may not be an absolute beginning or end of a user's information searching process, but only search sessions. The user's cognitive state is always ever-changing and advancing, both during and between these search sessions. So a more realistic model will probably mean a fusion of Marchionini's and Spink's models, where Marchionini's entire ISP process becomes a *cycle* inside the Spink's model, with forward-directed arrows only. These types of realistic models, improved and validated by empirical data, will help to explain phenomena behind next-generation search interactions, such as searching and multi-tasking, multi-tabbed browsing, [38, p. 36] multi-device searching, and multi-session searching [38, p. 61].

3.3 Neuro-physiological methods

Neuro-physiological methods (NP methods) [17] provide an interesting avenue to observe users while they interact with information systems. Two popular NP methods are eye-tracking [7, 15] and EEG [27]. Eye-tracking can captures eye-movements of users while they examine information on a screen. EEG captures (changes in) activation in different brain regions as users consume information. NP methods provide opportunities to investigate and understand how users gain knowledge during search. E.g, searchers use words or phrases they read in previous search results, in their future query reformulations. Eye-tracking can detect and model this phenomenon. As a result, a number of recent efforts have tried to investigate learning (during search) using one or more NP methods [2, 3, 24, 35]. However, a major limitation is that the NP methods (still) require lab-based settings for data collection. Taking lessons from the COVID-19 pandemic, as well as for scalability reasons, the IIR community needs search process metrics that can be measured remotely. Consumer wearable devices (smartwatches) are a promising direction, since they can record physiological data such as heart rate, skin temperature, and galvanic skin response. White and Ma [39] collected such data at a population scale, and correlated them with the population's search activities, to obtain improvements in relevance of result rankings.

4 CONCLUSION

Our propositions in this paper are shaped by our own experience in IIR research. The Information Processing Model from Educational Psychology states that information is most likely to be retained by a learner if it makes *sense*, and has *meaning* [33, p. 55]. When a piece of information *fits into the world-view* of the learner, it is said to make sense; when information is *relevant* to the learner, it has meaning.

[3]where each cycle consists of one or more interactive feedback occurrences of user's query input, IR system output, and user's interpretation and judgement of the output

Our past research have primarily been in the second aspect of information retention: *relevance judgement*. After several user-studies and analysing multimodal sources of data, we generally conclude that relevant information attracts more visual attention, longer eye-dwell time, and more brain activations [3, 15, 16], compared to irrelevant information. Metrics which can capture the entire duration of an experimental trial, or the real-time flow of interactions, usually perform better as predictors, than metrics which aggregate the entire trial into a set of single numbers [6, 7, 16]. Hence we call for new and improved measures of the search process.

In the domain of Search as Learning, we employed word [5] and sentence [3] embeddings to semantically compare searcher's responses to expert knowledge. Word embeddings provided better visualization of results, showing clear separation of Pre-Search Knowledge from Post-Search and Expert Knowledge [5]. We also co-related Knowledge Change measures with interaction and eye-tracking measures. We saw that people who learnt 'less' spent more reading effort on SERPs [3]. Conversely, people who learnt 'more' were doing less reading overall; but most of their reading was on content pages. These high learners used more specialized terms in their queries, and reported higher mental workload (NASA-TLX).

In conclusion, we reiterate that learning-during-search is a good candidate for an for evaluating IR systems. We need more research to understand relationships between the individual's search process and their learning outcomes. Process measures can shed light on the various subtle aspects of human behaviour. If we understand them well, we can teach people to be more successful in their information seeking efforts, and maximize their learning outcomes.

REFERENCES

[1] Daily Mail Australia. 2019. Why Googling symptoms leads to cancer diagnosis and worse for your health. https://www.dailymail.co.uk/news/article-6927695. (Accessed on 02/07/2021).

[2] Nilavra Bhattacharya and Jacek Gwizdka. 2018. Relating Eye-Tracking Measures with Changes in Knowledge on Search Tasks. In *Symposium on Eye Tracking Research & Applications (ETRA'18)*.

[3] Nilavra Bhattacharya and Jacek Gwizdka. 2019. Measuring Learning During Search: Differences in Interactions, Eye-Gaze, and Semantic Similarity to Expert Knowledge. In *CHIIR'19*.

[4] Nilavra Bhattacharya and Jacek Gwizdka. 2020. Visualizing and Quantifying Vocabulary Learning During Search. In *CIKM 2020 Workshop: 1st International Workshop on Investigating Learning During Web Search (IWILDS, CIKM'20)*. http://ceur-ws.org/Vol-2699/paper22.pdf

[5] Nilavra Bhattacharya and Jacek Gwizdka. 2020. Visualizing and Quantifying Vocabulary Learning During Search. In *Proceedings of the CIKM 2020 Workshops, October 19-20, 2020, Galway, Ireland*.

[6] Nilavra Bhattacharya, Somnath Rakshit, and Jacek Gwizdka. 2020. Towards Real-time Webpage Relevance Prediction Using Convex Hull Based Eye-tracking Features. In *Symposium on Eye Tracking Research & Applications (ETRA '20)*.

[7] Nilavra Bhattacharya, Somnath Rakshit, Jacek Gwizdka, and Paul Kogut. 2020. Relevance Prediction from Eye-Movements Using Semi-Interpretable Convolutional Neural Networks. In *Conference on Human Information Interaction and Retrieval (CHIIR'20)*.

[8] Bertram C Brookes. 1980. The foundations of information science. Part I. Philosophical aspects. *Journal of information science* 2, 3-4 (1980), 125–133.

[9] Yijin Chen, Yiming Zhao, and Ziyun Wang. 2020. Understanding online health information consumers' search as a learning process. *Library Hi Tech* (2020).

[10] National Research Council. 2000. *How People Learn: Brain, Mind, Experience, and School: Expanded Edition.* The National Academies Press, Washington, DC. https://doi.org/10.17226/9853

[11] Andrew Dillon. 2006. No more information seeking models please. https://adillon.ischool.utexas.edu/2006/10/03/no-more-information-seeking-models-please/. [Online; accessed 2020-05-28].

[12] Andrew Dillon. 2020. Personal communication. [dated 2020-04-29].

[13] Ujwal Gadiraju, Ran Yu, Stefan Dietze, and Peter Holtz. 2018. Analyzing knowledge gain of users in informational search sessions on the web. In *Conference on Human Information Interaction & Retrieval (CHIIR)*.

[14] Souvick Ghosh, Manasa Rath, and Chirag Shah. 2018. Searching as learning: Exploring search behavior and learning outcomes in learning-related tasks. In *Conference on Human Information Interaction & Retrieval (CHIIR)*.

[15] Jacek Gwizdka. 2014. Characterizing Relevance with Eye-Tracking Measures. In *Proceedings of the 5th Information Interaction in Context Symposium (IIiX '14)*. ACM, New York, NY, USA, 58–67. https://doi.org/10.1145/2637002.2637011 00028.

[16] Jacek Gwizdka, Rahilsadat Hosseini, Michael Cole, and Shouyi Wang. 2017. Temporal Dynamics of Eye-Tracking and EEG during Reading and Relevance Decisions. *Journal of the Association for Information Science and Technology* (2017).

[17] Jacek Gwizdka, Yashar Moshfeghi, and Max L. Wilson. 2019. Introduction to the Special Issue on Neuro-Information Science. *Journal of the Association for Information Science and Technology* 70, 9 (2019), 911–916.

[18] Orland Hoeber, Dolinkumar Patel, and Dale Storie. 2019. A Study of Academic Search Scenarios and Information Seeking Behaviour. In *Conference on Human Information Interaction and Retrieval (CHIIR)*. 231–235.

[19] Xiaoli Huang and Dagobert Soergel. 2013. Relevance: An Improved Framework for Explicating the Notion. *Journal of the American Society for Information Science and Technology* 64, 1 (2013), 18–35. https://doi.org/10.1002/asi.22811

[20] Hugo C Huurdeman and Jaap Kamps. 2014. From multistage information-seeking models to multistage search systems. In *Proceedings of the 5th Information Interaction in Context Symposium*. 145–154.

[21] D. Kelly, S. Dumais, and J. O. Pedersen. 2009. Evaluation challenges and directions for information-seeking support systems. *IEEE Computer* 42, 3 (2009).

[22] Lisa Koeman. 2020-06-18. HCI/UX research: what methods do we use? – Lisa Koeman – blog. https://lisakoeman.nl/blog/hci-ux-research-what-methods-do-we-use/. (Accessed on 11/08/2020).

[23] Sanne Kruikemeier, Sophie Lecheler, and Ming M Boyer. 2018. Learning from news on different media platforms: An eye-tracking experiment. *Political Communication* 35, 1 (2018), 75–96.

[24] Jiaxin Mao, Yiqun Liu, Noriko Kando, Min Zhang, and Shaoping Ma. 2018. How Does Domain Expertise Affect Users' Search Interaction and Outcome in Exploratory Search? *ACM Transactions on Information Systems* 36 (July 2018).

[25] Gary Marchionini. 1995. *Information Seeking in Electronic Environments*. Cambridge University Press.

[26] Heather L O'Brien, Andrea Kampen, Amelia W Cole, and Kathleen Brennan. 2020. The Role of Domain Knowledge in Search as Learning. In *Conference on Human Information Interaction and Retrieval (CHIIR)*.

[27] Zuzana Pinkosova, William J McGeown, and Yashar Moshfeghi. 2020. The cortical activity of graded relevance. In *Conference on Research and Development in Information Retrieval (SIGIR)*. 299–308.

[28] Peter Pirolli, Patricia Schank, Marti Hearst, and Christine Diehl. 1996. Scatter/Gather Browsing Communicates the Topic Structure of a Very Large Text Collection. In *Conference on Human Factors in Computing Systems (CHI'96)*.

[29] Soo Young Rieh. 2020. Research Area 1: Searching as Learning. https://rieh.ischool.utexas.edu/research. [Online; accessed 2020-04-19].

[30] Nirmal Roy, Felipe Moraes, and Claudia Hauff. 2020. Exploring Users' Learning Gains within Search Sessions. In *Conference on Human Information Interaction and Retrieval (CHIIR)*.

[31] Tefko Saracevic. 1999. Information Science. *Journal of the American Society for Information Science* 50 (1999), 1051–1063. https://doi.org/10.1002/(SICI)1097-4571(1999)50:12<1051::AID-ASI2>3.0.CO;2-Z

[32] Tefko Saracevic. 2016. The Notion of Relevance in Information Science: Everybody Knows What Relevance Is. But, What Is It Really? *Synthesis Lectures on Information Concepts, Retrieval, and Services* (2016).

[33] David A Sousa. 2017. *How the Brain Learns, Fifth Edition.* Corwin Press.

[34] Amanda Spink. 1997. Study of Interactive Feedback during Mediated Information Retrieval. *Journal of the American Society for Information Science* (1997).

[35] Rohail Syed, Kevyn Collins-Thompson, Paul N Bennett, Mengqiu Teng, Shane Williams, Dr Wendy W Tay, and Shamsi Iqbal. 2020. Improving Learning Outcomes with Gaze Tracking and Automatic Question Generation. In *The Web Conference (WWW)*.

[36] Hindustan Times. 2015. Stop it! Google is not a real doctor and no you don't have cancer. https://www.hindustantimes.com/health/story-EjdPqGmAe2CugNyNdUClSI.html. (Accessed on 02/07/2021).

[37] Vu Tuan Tran and Norbert Fuhr. 2012. Using Eye-Tracking with Dynamic Areas of Interest for Analyzing Interactive Information Retrieval. In *Conference on Research and Development in Information Retrieval (SIGIR'12)*.

[38] Ryen White. 2016. *Interactions with search systems.* Cambridge University Press.

[39] Ryen White and Ryan Ma. 2017. Improving Search Engines via Large-Scale Physiological Sensing. In *Conference on Research and Development in Information Retrieval (SIGIR'17)*.

[40] Luyan Xu, Xuan Zhou, and Ujwal Gadiraju. 2020. How Does Team Composition Affect Knowledge Gain of Users in Collaborative Web Search?. In *Conference on Hypertext and Social Media (HT'20)*.

[41] Pengyi Zhang and Dagobert Soergel. 2016. Process patterns and conceptual changes in knowledge representations during information seeking and sense-making: A qualitative user study. *Journal of Information Science* 42, 1 (2016).