Operationalizing Knowledge Gain: Implementing and Testing the DKG Metric in Search Environments

Rafael Tavares da Silva
rafaeltavares.silva@edu.unirio.br
Federal University of the State of Rio de Janeiro

Sean Wolfgand Matsui Siqueira (Orientador)
sean@uniriotec.br
Federal University of the State of Rio de Janeiro

Marcelo Tibau (Coorientador)
marcelo.tibau@uniriotec.br
Federal University of the State of Rio de Janeiro

ABSTRACT

Searching the Web is increasingly recognized as a process of knowledge construction rather than simple information retrieval, a perspective framed by the paradigm of Searching as Learning (SaL). A central challenge in this domain lies in evaluating the extent to which users actually acquire knowledge during search. Traditional approaches either rely on behavioral proxies, scalable but limited in capturing conceptual change, or structured assessments, which provide direct evidence but are often intrusive. The Degree of Knowledge Gain (DKG) metric addresses this gap by modeling reductions in uncertainty through Shannon's entropy and integrating semantic similarity between queries and clicked documents. This paper reports on the operationalization of DKG within the CNPq project *3C-BPA: Comportamento de busca, Complexidade da informação e pensamento Crítico na Busca como um Processo de Aprendizagem*. Two artifacts were developed: an initial search engine prototype embedding DKG computation, and a Chrome extension that estimates DKG in real time while users employ their preferred search engines. The latter artifact overcame earlier limitations by improving ecological validity, reducing costs, and enabling more natural experimentation. An experiment combined pre- and post-tests, the Concurrent Think-Aloud (CTA) protocol, and the plug-in's automated logging. Preliminary findings show that DKG values are sensitive to differences in search strategies, with systematic reformulation and evaluation aligning with greater knowledge gains, while disorientation behaviors corresponded to more modest outcomes. A distinctive feature of this study was the active role of an undergraduate researcher, who contributed to artifact development, experiment setup, participant support, transcription, and ongoing content analysis.

KEYWORDS

Degree of Knowledge Gain, Searching as Learning, Search Artifacts, Search Environment

1 INTRODUCTION

Searching the Web increasingly involves knowledge construction, a perspective framed by Searching as Learning (SaL) [12, 20, 24]. Assessing what users actually learn is challenging: behavioral proxies such as clicks and reformulations scale well but overlook conceptual change [6, 22, 25], whereas structured tests provide precision but at the cost of ecological validity [18]. The Degree of Knowledge Gain (DKG) metric addresses this tension by formalizing learning as a reduction of uncertainty through Shannon entropy and by incorporating semantic similarity between queries and clicked documents [19]. While initial studies demonstrated its interpretability and alignment with learning indicators, the next essential step is its operationalization in real search environments [19–21].

This paper reports on such an effort, conducted within the CNPq project *3C-BPA: Comportamento de busca, Complexidade da informação e pensamento Crítico na Busca como um Processo de Aprendizagem.* A central contribution of the project has been the development of technological artifacts that embed DKG into actual search workflows. This work was carried out in large part through a student research project, where the student implemented and tested two artifacts: a prototype search engine with built-in DKG computation and a browser plug-in capable of estimating the metric in real time while users rely on their preferred search engines. These developments advanced the practical viability of the metric and provided a valuable training ground for junior researchers.

To assess the plug-in's applicability, the authors designed an experiment combining pre- and post-tests with the Concurrent Think-Aloud (CTA) protocol [8], across tasks of varying complexity ranging from introductory science concepts to applied topics such as prosthetics. This experimental setup allowed us to evaluate both the reliability of the DKG as an automated measure and the kinds of strategies participants employed while searching.

The remainder of this paper is structured as follows. Section 2 reviews related work on SaL and automatic measures of knowledge gain. Section 3 introduces the principles of the DKG metric and describes the artifacts developed to operationalize it. Section 4 outlines the experimental design, emphasizing the contributions of the undergraduate researcher to the setup, execution, and analysis. Section 5 presents preliminary results, integrating automated and qualitative evidence of knowledge gain. Finally, Section 6 reflects on the metric's potential, limitations, and future directions, highlighting both methodological advances and the formative role of undergraduate participation.

2 RELATED WORKS

Research in Searching as Learning (SaL) has highlighted that search is not only a means of locating information but also an active process of knowledge construction [23]. Within this paradigm, one of the main challenges lies in assessing the extent to which learning occurs during a search session. Approaches to this problem can be

In: V Concurso de Trabalhos de Iniciação Científica (CTIC 2025). Anais Estendidos do XXXI Simpósio Brasileiro de Sistemas Multimídia e Web (CTIC'2025). Rio de Janeiro/RJ, Brasil. Porto Alegre: Brazilian Computer Society, 2025.

ISSN 2596-1683

divided into those that directly measure user knowledge, such as pre/post-tests, open-ended responses, and self-reports, and those that rely on indirect signals derived from system interactions, such as behavioral logs or query traces [19].

A substantial body of work has explored these indirect approaches, often referred to as measures of assumed evidence of learning. Early contributions [2] and subsequent studies [4, 6, 14, 15, 25, 26] analyzed behavioral proxies such as query reformulations, document sequences, and interaction patterns as indicators of evolving knowledge states. More recent efforts extend this line of inquiry with knowledge-graph models that approximate conceptual growth during exploratory tasks [5, 9], or with multimodal data such as eye-tracking to capture reading strategies and their relation to knowledge acquisition [7]. Complementary work has emphasized task-oriented and contextual factors, from query refinements to motivational attributes, which shape how knowledge is acquired during search [1, 10].

These advances underscore the potential of search systems to support retrieval and stimulate and measure learning. Nevertheless, current behavioral proxies face two persistent limitations: First, they often model learning as an emergent byproduct of interaction, without fully addressing its cognitive structure; second, while scalable, they struggle to reflect qualitative differences in knowledge acquisition across tasks of varying complexity.

The Degree of Knowledge Gain (DKG) metric addresses these gaps by combining Shannon entropy with semantic similarity between queries and clicked documents to formalize the refinement of knowledge during search sessions [19]. Unlike traditional proxies, DKG models learning as a progressive reduction of uncertainty, anchored in observable behaviors, which positions it as a promising approach for operationalization in search environments that aim to evaluate or stimulate learning. The present work contributes to this agenda by advancing the integration of DKG into real-world search settings, linking metric formalization with practical tools and experiments conducted in the context of educational projects.

3 THE DKG METRIC AND ARTIFACT DEVELOPMENT

The Degree of Knowledge Gain (DKG) metric provides a formal model to quantify learning during search. It is grounded in Shannon's Entropy (H), which expresses uncertainty as a probability distribution of possible outcomes [16]. In the search context, entropy represents the uncertainty of a user's knowledge state before encountering new information; as learning occurs, entropy decreases. Extending this notion, DKG integrates semantic similarity between queries and clicked documents, ensuring that knowledge gain reflects both novelty and conceptual refinement.

The formulation of additive information gain [13] underpins the qualitative dimension of DKG; each new information element β adds incrementally to what is already known from α:

$$i(\alpha \wedge \beta, \Gamma) = i(\alpha, \Gamma) + i(\beta, \Gamma \cup \alpha) \tag{1}$$

where Γ denotes the evolving knowledge base. This iterative view of knowledge guided the design of artifacts to compute DKG in real time.

The formal expression of the Degree of Knowledge Gain (DKG) is given in [19] as:

$$\deg_{KG} = \left(1 - \sum_{i=1}^{n} p_i \log \frac{m_i}{p_i}\right) \times 0.01 \tag{2}$$

where:

- p_i represents the probability of detecting a clicked document at position i, estimated from empirical click-through data [3];
- m_i denotes the observation factor, which integrates the click indicator (c_i) with the Jaccard similarity between successive queries.

In this way, m_i accounts for the combined influence of user interactions (such as query reformulations and clicks) and system behavior (ranking output), whereas p_i models the likelihood of those interactions. The scaling term ensures the metric remains interpretable.

Compared with click-count or dwell-time proxies, DKG integrates rank-weighted detectability (p_i) and inter-query similarity (via m_i), offering a single interpretable trajectory of knowledge refinement during a session.

3.1 First Artifact: Search Engine Prototype

The initial implementation effort focused on building a software prototype that simulated a conventional search engine while embedding DKG computation in its logic. The objective was to automate the calculation of DKG during live search sessions, enabling immediate tracking of user knowledge gain. This artifact required approximately two months of development and was used in preliminary experiments. However, significant limitations emerged. The quality of retrieved results did not always match participants' search intentions, which undermined user experience and the validity of collected data. Furthermore, the external API used to supply results imposed high token costs per session, making large-scale experimentation unfeasible. While this first artifact served as a proof of concept, its constraints highlighted the need for a more sustainable solution.

3.2 Second Artifact: Chrome Extension

To overcome these challenges, a second artifact was developed: a plug-in for the Google Chrome browser. Unlike the standalone prototype, this extension integrated seamlessly into users' natural search environment. By computing DKG unobtrusively in the background, the plug-in preserved the user's preferred search engine, thereby improving ecological validity and user satisfaction. From a technical standpoint, the extension stored all session data in Supabase, which provided structured persistence with an adequate free tier. This eliminated the prohibitive costs of API-based retrieval and allowed consistent modeling of query, click, and semantic similarity data. Methodologically, it also ensured that experiments could be conducted under more naturalistic conditions, reducing intrusiveness and enhancing reliability.

Implementation at a glance:

- Stack: Chrome MV3 extension; background logger; Supabase storage.

- Signals logged: queries, clicks, ranks, inter-query Jaccard, timestamps (for m_i, p_i).
- Privacy: local pseudonyms; no page content stored, only interaction metadata.

Experiments conducted with five participants confirmed the viability of the plug-in. The decision to work with five users follows a well-established tradition in usability studies, which consistently demonstrates that such samples are sufficient to capture core interaction issues and achieve observational saturation [11]. In contrast to the first artifact, the extension delivered relevant results, required no maintenance costs, and supported a more scalable approach to experimentation.

3.3 Comparison and Contribution

The progression from the standalone search engine to the Chrome extension illustrates an important conceptual shift in operationalizing the DKG metric. The first artifact, while limited, validated the feasibility of computing DKG in real time. The second artifact advanced this goal substantially, combining technical sustainability with methodological rigor. This development effort, carried out within the context of an undergraduate research project, demonstrates how theoretical models such as DKG can be embedded into real-world search settings. By enabling scalable, ecologically valid experiments, the artifacts[1] represent a technical contribution and a methodological bridge between formal metrics and practical applications of Searching as Learning.

4 METHODOLOGY

The study was conducted to evaluate the Degree of Knowledge Gain (DKG) metric in real search scenarios through a combination of automatic logging and qualitative inquiry. Five university volunteers (convenience sampling; prior search familiarity self-reported) completed three tasks of increasing complexity under controlled conditions. Ethical approval was obtained from the institutional review board (CAAE 82881624.0.0000.5285), and all participants gave informed consent. Sessions were anonymized, recorded, and later transcribed. A key contribution to this work was the involvement of the undergraduate student supported by the project *Comportamento de busca, Complexidade da informação e pensamento Crítico na Busca como um Processo de Aprendizagem (3C-BPA)*. His participation was central in several stages of the experiment: First, he was responsible for preparing the experimental environment, which included configuring the logging plug-in, setting up the online platform for pre- and post-tests, and ensuring smooth integration between these components. During the sessions themselves, the student assisted participants by clarifying task instructions and ensuring that the Concurrent Think-Aloud (CTA) protocol [8] was followed consistently.

Beyond data collection, the student also contributed to the analytic process, preparing transcripts that ground the preliminary qualitative findings and is currently advancing the content analysis. This involves coding participant interactions with the Online Information Searching Strategies (OISS) framework [17] and the ESKiP Taxonomy of Query State [22], both of which enable systematic mapping of strategies and query transitions.

[1]Available at https://github.com/rafaeltavaares/iniciacao-cientifica-dkg

The quantitative side of the study combined pre/post-test scores with automatically computed DKG values derived from query and click logs. Mixed-methods triangulation was then applied, linking external test outcomes with DKG measurements and coded strategies[2], illustrating both the methodological robustness of the study and the formative role of undergraduate research in advancing the development and application of the DKG metric.

5 FINDINGS

The preliminary review of transcripts points to a variety of search behaviors, shaped both by participants' prior familiarity with search engines and by their ability to regulate the process in real time. Procedural strategies linked to Trial-and-Error were especially common. Participants frequently modified queries, experimented with alternative keywords, and explored different websites when early attempts did not succeed. User 4 illustrates this pattern, persistently refining queries after receiving unhelpful results, showing determination and iterative problem-solving.

Metacognitive strategies appeared more prominently among those who achieved higher post-test gains. Purposeful Thinking, such as narrowing a query's scope or conducting targeted searches within specific sites, was particularly visible. For example, User 3 consistently evaluated and cross-checked multiple sources, weighing credibility and synthesizing information across sites. Indicators of disorientation were also observed. User 2, for instance, showed frustration when initial searches failed, at times giving up quickly or reacting negatively to failure. Such reactions suggest that limited cognitive regulation may hinder effective learning through search.

When analyzed through the ESKiP taxonomy, participants were found to rely most heavily on Specialization (SC) and Word Substitution (WS). These moves allowed them to narrow broad queries or adapt vocabulary to better match expected document language. User 5, for instance, began with general queries before progressively adding detail to achieve more precise results. In contrast, Generalization (GE) moves were rare, suggesting that participants tended to reduce rather than expand the scope of queries when encountering obstacles. Across sessions, higher DKG values co-occurred with more frequent purposeful reformulations and cross-source evaluation, whereas segments marked by disorientation aligned with flatter DKG trajectories (preliminary pattern consistent in all sessions).

Taken together, these preliminary findings suggest that the DKG metric is responsive to variations in strategy use. While automated measures confirm that learning occurs, transcript analysis reveals why the outcomes differ. Persistent reformulation and structured evaluation, as seen with Users 4 and 3, aligned with stronger gains, whereas frustration and less systematic behavior, as with User 2, corresponded to more modest improvements.

6 CONCLUSION

This study contributes to the ongoing effort of operationalizing the Degree of Knowledge Gain (DKG) metric within real search environments by developing and testing two artifacts: a prototype search engine and a browser extension, which demonstrated how

[2]The coding scheme and the computed values are available in the same repository as the artifacts.

theoretical formulations of knowledge gain can be embedded into practical tools. The transition from the standalone prototype to the Chrome extension marked a crucial methodological advance, as it ensured greater ecological validity, reduced costs, and enabled experimentation in more naturalistic contexts. The mixed-methods experiment provided preliminary evidence that the DKG metric is sensitive to differences in search strategy use. Participants who engaged in systematic query reformulation and source evaluation exhibited higher levels of knowledge gain, as reflected in both pre/post-test comparisons and DKG scores. Conversely, disorientation and limited cognitive regulation corresponded to more modest improvements. These results underscore the potential of DKG as both a diagnostic and evaluative tool for Searching as Learning.

A key aspect of this work was the formative role of undergraduate research. The student supported by the 3C-BPA project was central to the design, implementation, and execution of the study: from building and refining the artifacts, through preparing the experimental setup, to assisting in data collection and conducting ongoing content analysis. His involvement expanded the practical viability of the DKG metric and exemplified how student research can drive innovation while providing meaningful training in empirical methods. While the findings reported here are preliminary, they highlight the promise of the DKG metric as a scalable and interpretable measure of knowledge acquisition during search.

In forthcoming work, the authors intend to report a full, artifact-centered study (expanded participant pool, task variety, reliability analysis, and descriptions of DKG components), detailing application outcomes and derived design recommendations. Also, the authors intend to triangulate CTA with complementary protocols (e.g., concurrent/retrospective hybrids and design-thinking probes [8]) to strengthen process validity.

REFERENCES

[1] Arthur Câmara. 2024. *Designing Search-as-Learning Systems*. Ph.D. Thesis. Delft University of Technology. ⟨DOI: 10.4233/uuid:0fe3a6bb-1bc1-40e2-86b0-ec3d3aef9c77⟩, ⟨ISBN: 978-94-6384-569-4⟩.

[2] Yu Chi, Shuguang Han, Daqing He, and Rui Meng. 2016. Exploring Knowledge Learning in Collaborative Information Seeking Process. In *SAL@SIGIR*.

[3] Georges E. Dupret and Benjamin Piwowarski. 2008. A User Browsing Model to Predict Search Engine Click Data from Past Observations.. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Singapore, Singapore) *(SIGIR '08)*. Association for Computing Machinery, New York, NY, USA, 331–338. doi:10.1145/1390334.1390392

[4] Dima El Zein and Célia da Costa Pereira. 2022. User's Knowledge and Information Needs in Information Retrieval Evaluation. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (Barcelona, Spain) *(UMAP '22)*. Association for Computing Machinery, New York, NY, USA, 170–178. doi:10.1145/3503252.3531325

[5] Dima El Zein and Célia Da Costa Pereira. 2023. The Evolution of User Knowledge during Search-as-Learning Sessions: A Benchmark and Baseline. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 454–458. doi:10.1145/3576840.3578273

[6] Wolfgang Gritz, Anett Hoppe, and Ralph Ewerth. 2021. On the Impact of Features and Classifiers for Measuring Knowledge Gain during Web Search-A Case Study. (2021).

[7] Wolfgang Gritz, Anett Hoppe, and Ralph Ewerth. 2024. On the Influence of Reading Sequences on Knowledge Gain During Web Search. In *Advances in Information Retrieval*, Nazli Goharian, Nicola Tonellotto, Yulan He, Aldo Lipani, Graham McDonald, Craig Macdonald, and Iadh Ounis (Eds.). Springer Nature Switzerland, Cham, 364–373.

[8] Todd Kelley, Brenda Capobianco, and Kevin Kaluf. 2015. Concurrent think-aloud protocols to assess elementary design students. *International Journal of Technology & Design Education* 25, 4 (2015), 521 – 540.

[9] Chang Liu, Xiaoxuan Song, and Preben Hansen. 2023. Characterising users' task completion process in learning-related tasks:A search pace model. *Journal of Information Science* 49, 6 (2023), 1462–1480. arXiv:https://doi.org/10.1177/01655515211060527 doi:10.1177/01655515211060527

[10] Yaxi Liu, Chunxiu Qin, Xubu Ma, Jiangping Chen, Hongle He, and Junbo Mao. 2025. Characterising exploratory search tasks: Evidence from different fields. *Journal of Information Science* 0, 0 (2025), 01655515251330611. arXiv:https://doi.org/10.1177/01655515251330611 doi:10.1177/01655515251330611

[11] Jie "Jennifer" Lu. 2025. A systematic literature review of usability definitions and psychometric properties of instruments in the field of learning design and technology. *Journal of Research on Technology in Education* 0, 0 (2025), 1–22. arXiv:https://doi.org/10.1080/15391523.2025.2459165 doi:10.1080/15391523.2025.2459165

[12] Gary Marchionini. 2006. Exploratory Search: From Finding to Understanding. *Commun. ACM* 49, 4 (April 2006), 41–46. doi:10.1145/1121949.1121979

[13] Peter Milne. 2012. Probability as a Measure of Information Added. *Journal of Logic, Language and Information* 21, 2 (April 2012), 163–188. doi:10.1007/s10849-011-9142-0

[14] Christian Otto, Markus Rokicki, Georg Pardi, Wolfgang Gritz, Daniel Hienert, Ran Yu, Johannes von Hoyer, Anett Hoppe, Stefan Dietze, Peter Holtz, Yvonne Kammerer, and Ralph Ewerth. 2022. SaL-Lightning Dataset: Search and Eye Gaze Behavior, Resource Interactions and Knowledge Gain during Web Search. In *ACM SIGIR Conference on Human Information Interaction and Retrieval*. ACM, Regensburg Germany, 347–352. doi:10.1145/3498366.3505835

[15] Christian Otto, Ran Yu, Georg Pardi, Johannes von Hoyer, Markus Rokicki, Anett Hoppe, Peter Holtz, Yvonne Kammerer, Stefan Dietze, and Ralph Ewerth. 2021. Predicting Knowledge Gain During Web Search Based on Multimedia Resource Consumption. In *Artificial Intelligence in Education*, Ido Roll, Danielle McNamara, Sergey Sosnovsky, Rose Luckin, and Vania Dimitrova (Eds.). Springer International Publishing, Cham, 318–330.

[16] Alfonso Prieto-Guerrero and Gilberto Espinosa-Paredes. 2019. 7 - Nonlinear signal processing methods: DR estimation and nonlinear stability indicators. In *Linear and Non-Linear Stability Analysis in Boiling Water Reactors*, Alfonso Prieto-Guerrero and Gilberto Espinosa-Paredes (Eds.). Woodhead Publishing, 315 – 398. doi:10.1016/B978-0-08-102445-4.00007-2

[17] Ilknur Reisoğlu, Ayça Çebi, and Tuğba Bahçekapılı. 2019. Online information searching behaviours: examining the impact of task complexity, information searching experience, and cognitive style. *Interactive Learning Environments* 0, 0 (2019), 1–18. doi:10.1080/10494820.2019.1662456

[18] Mark A Schmuckler. 2001. What is ecological validity? A dimensional analysis. *Infancy* 2, 4 (2001), 419–436.

[19] Marcelo Tibau. 2024. *Quantifying Knowledge Gain in Online Searches: The DKG Metric*. Tese (Doutorado em Informática). Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Rio de Janeiro. Programa de Pós-Graduação em Informática.

[20] Marcelo Tibau, Sean Wolfgand Matsui Siqueira, and Bernardo Pereira Nunes. 2022. The Impact of Non-Verbalization in Think-Aloud: Understanding Knowledge Gain Indicators Considering Think-Aloud Web Searches. In *Proceedings of the 33rd ACM Conference on Hypertext and Social Media* (Barcelona, Spain) *(HT '22)*. Association for Computing Machinery, New York, NY, USA, 107–120. doi:10.1145/3511095.3531272

[21] Marcelo Tibau, Sean Wolfgand Matsui Siqueira, and Bernardo Pereira Nunes. 2023. Accounting for the knowledge gained during a web search: An empirical study on learning transfer indicators. *Library & Information Science Research* 45, 1 (2023), 101222. doi:10.1016/j.lisr.2022.101222

[22] M. Tibau, S. W. M. Siqueira, B. Pereira Nunes, T. Nurmikko-Fuller, and R. F. Manrique. 2019. Using Query Reformulation to Compare Learning Behaviors in Web Search Engines. In *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, Vol. 2161-377X. 219–223. doi:10.1109/ICALT.2019.00054

[23] Kelsey Urgo and Jaime Arguello. 2022. Learning assessments in search-as-learning: A survey of prior work and opportunities for future research. *Information Processing & Management* 59, 2 (2022), 102821. doi:10.1016/j.ipm.2021.102821

[24] P. Vakkari. 2016. Searching as learning: A systematization based on literature. In *Journal of Information Science, 42(1)*. 7–18.

[25] Luyan Xu, Xuan Zhou, and Ujwal Gadiraju. 2020. How Does Team Composition Affect Knowledge Gain of Users in Collaborative Web Search?. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media* (Virtual Event, USA) *(HT '20)*. Association for Computing Machinery, New York, NY, USA, 91–100. doi:10.1145/3372923.3404784

[26] Ran Yu, Rui Tang, Markus Rokicki, Ujwal Gadiraju, and Stefan Dietze. 2021. Topic-independent modeling of user knowledge in informational search sessions. *Information Retrieval Journal* 24, 3 (June 2021), 240–268. doi:10.1007/s10791-021-09391-7