The "Active Search" Hypothesis: How Search Strategies Relate to Creative Learning

Srishti Palani, Zijian Ding, Stephen MacNeil, Steven P. Dow
Universities of California, San Diego, USA
{srishiti,zding,smacneil,spdow}@ucsd.edu

ABSTRACT
While research shows that web search plays a role throughout the creative process, less is known about how people use web search to learn and frame their thinking about an open problem. People need web search to gather information about a problem area, but this can also influence the rest of the creative process. To understand how web search affects early-stage design, we collected and analyzed search log and self-report data from 34 students in a project-based design class. Participants reported struggling with scoping broad, ill-defined information goals into queries, learning domain-specific language, and assessing the usefulness of information. Analysis found that more active and diverse search behavior (i.e. issuing more frequent and diverse queries, and opening more webpages) related to more progress in early-stage design (i.e. gathering more facts, articulating more insights, and developing better problem frames). Based on these findings, we discuss implications for designing search tools to support peoples’ creative processes.

CCS CONCEPTS
• Information systems → Web search engines; • Applied computing → Education; • Human-centered computing → Empirical studies in HCI.

KEYWORDS
Search Engines; Creativity; Learning; Design; Problem Framing

ACM Reference Format:

1 INTRODUCTION
Large, complex challenges – like keeping the public safe during a pandemic, dealing with climate change, and enabling equitable access to public transportation – often require problem solvers to form a broad and deep understanding of facts, constraints, and existing solutions [12, 18, 29]. This process of gathering information and discovering insights can have a significant impact on how designers approach or 'frame' a problem [11, 23, 31]. Prior research has revealed the importance of web search throughout the creative design process, including to find existing solutions, search for inspiration, learn how to use prototyping tools [14, 17, 40, 41]. Based on a recent diary and survey study [41], researchers have found that people search to support a range of creative tasks across different domains, such as academic writing, cooking, design (e.g. visual, architectural, etc.) [20, 26, 39, 40]. Searchers use specific information resources (e.g. images, videos) strategically to support different stages of the creative process [41]. However, less is known about how specific web search behaviors influence early-stage design and problem framing. This paper builds on prior self-report studies to understand how search behavior relates to learning and problem framing, by gathering search log data to observe in-situ search behavior and survey data to gain qualitative insight into the meaning of quantitative results.

Prior work, by the Search-as-Learning community, has developed tasks [21, 38] and measures [3, 37] using the cognitive learning dimension of Anderson and Krathwohl’s Taxonomy of Learning [24] a well-known education resource. In this taxonomy, six types of cognitive processes are identified: remember, understand, apply, analyze, evaluate and create. Early stages of a design process require the designer to learn about a new domain, and involve all these types of learning: from recalling facts (remember) to synthesizing information to discover insights (understand, analyze, evaluate), and asking questions and posing problems in a fruitful and radical ways, generating ideas (create), etc. [9, 10, 27, 31]. Rich et al. [28] conceptualizes this comprehensive task as "creative learning". While there has been some prior work [16, 22, 38] to understand the relationship between search behavior and learning outcomes, these studies focus on learning rather than creative tasks. This paper builds on this prior work to investigate the following research questions:

**RQ1**: What does search log and self-report data reveal about the information goals, challenges experienced and strategies employed by searchers during early-stage design?

**RQ2**: How do web search behaviors relate to creative learning outcomes (such as gathering facts, discovering of insights, and framing the problem)?

To address these research questions, we observed 34 students in a project-based design classroom as they searched the web for 30 minutes during their early-stage design process. Our analysis of search log data, together with participants’ self-reports about their experiences found that: participants have cognitively-complex information goals – for example, to understand domain-specific language and context, to find patterns and design constraints, and seek inspiration to generate ideas. Participants reported challenges with scoping broad, ill-defined information needs into queries, and assessing the usefulness of information. Analysing the search logs...
found that active searching (i.e. issuing more, longer, and more diverse queries and opening more webpages) relates to higher creative learning gains (such as gathering more facts, discovering more insights and developing more well-defined problem frames). Based on these findings, we discuss implications for designing search tools to better support peoples’ creative learning process.

2 METHOD

2.1 Participants

34 undergraduate students from a project-based design course were recruited, and received 1% extra course credit for their participation. The study was conducted over a period of the first two days in the first week of classes. Participants had a diverse range of prior design experience ($\mu = 2.19, \sigma = 1.07$). Participants had only a little prior knowledge about the topic ($\mu = 1.38, \sigma = 0.72$, on a scale of 1=no knowledge at all, 5=know a lot) All participants reported extensive prior experience with web search, in general everyone reported searching multiple times a day, and using web search for more than five years. 32 of them reported using Google, and 2 using Bing and Baidu as their primary search engine. All of them used Google Chrome as their primary search browser.

2.2 Procedure

At the start of the study, the researcher explained the study procedure and guided participants through how to install and use a Chrome browser plugin (https://tinyurl.com/HistoryMaster) to collect search logs. All search log data was automatically anonymized upon collection. A pre-task survey captured participants’ web search experience, prior design experience, prior domain knowledge, and information seeking goals. As the main task, participants had 30 minutes to search and take notes on one of four topics being studied in a project-based design course (Refer to Supplementary Materials at https://tinyurl.com/SearchTasks for Search Tasks). The breadth and depth of these four multi-faceted topics provides a good opportunity to study web search in the context of early-stage design.

Before and after the task, participants were required to summarize what they knew about the topic in 3-5 sentences or 200-words, and write a problem statement that could be the focus of a quarter-long project. Additionally, in the post-task survey students were asked to report any challenges faced and strategies used when using web search during this early-stage exploratory creative design task. The study lasted 60 minutes: 30 for web search and 30 minutes for filling out the pre- and post-task surveys.

2.3 Measures

2.3.1 Qualitative Insights from Surveys. Survey questions about information seeking goals, challenges faced and strategies used were analyzed using a grounded-theory approach to thematic analysis.

2.3.2 Web Search Logs. To understand search behavior, from the search logs we calculated the following measures for each participant: (i) Number of queries issued; (ii) Length of query (i.e. average number of terms per query); (iii) Diversity of query (i.e. number of unique query terms, stemmed to reduce the query terms to their respective base forms without affixes); (iv) Number of unique web pages opened.

2.3.3 Creative Learning Metrics. To measure creative learning outcomes, we calculated the following measures for each participant: (i) Change in Number of Declared Facts (Remember) is measured by the change in number of distinct facts per statement between pre- and post-task summaries. To reliably measure number of facts, we randomly selected 20% of all 34 pre- and post-summaries for four raters to independently count facts. To account for agreement between four raters we calculated the Fleiss’ $\kappa$. The raters had an inter-rater reliability of 0.74 Fleiss’ $\kappa$. The rest was coded by one of the raters. (ii) Depth of Learning (Understand, Analyze, Evaluate) is measured by three metrics proposed by Wilson and Wilson [37] (refer to Table 1). (iii) Degree of Problem Definition (Create) To understand the process of moving from an ill-defined to a well-defined problem scope (i.e. from level 1: Problem Discovery to level 5: Problem Definition) we adopt Abdulla and Crammoud’s Problem Finding Hierarchy [1]. Level 1: Problem Discovery: the problem statement is very ill-defined or defined very similar to the given problem. There is no relevant information and no insight to build on. Level 2: Problem Formulation: the problem statement is yet to specify the problem, however, there is enough information that they could discover insights from. Level 3: Problem Construction: the problem statement includes some background information, but the problem finder needs to further evaluate the information to specify a well-informed and well-reasoned problem. Level 4: Problem Identification: the problem statement includes some information and has some preliminary insights about the problem; however, there is no specific problem identified or the problem identified is still rather vague. Level 5: Problem Definition: (most well-defined stage of problem finding) the problem statement identifies as specific, well-informed and well-reasoned problem. When classifying the pre- and post-task problem statements, the raters had an agreement of 0.69 Fleiss’s $\kappa$.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Definition</th>
<th>Fleiss’ Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of Facts (Understand)</td>
<td>Usefulness of recalled facts (0-3, where 0: irrelevant or useless facts, 3: facts demonstrate technical understanding)</td>
<td>0.64</td>
</tr>
<tr>
<td>Interpretation of Facts (Analyze)</td>
<td>Synthesis of facts to draw conclusions (0-2, where 0: simply listing facts with no further interpretation, 2: finding patterns across multiple facts)</td>
<td>0.58</td>
</tr>
<tr>
<td>Critique of Facts (Evaluate)</td>
<td>Evaluation of facts to raise questions, identify outliers and inconsistencies (0-1, where 0: true, 1: false)</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Table 1: Depth of Learning Measures corresponding to the Understand, Analyze and Evaluate Cognitive Learning Levels of Anderson and Krathwohl’s Taxonomy of Learning. [24, 37] Fleiss’ $\kappa$ is significant at $p < 0.05$.

3 RESULTS

10 participants chose to work on the topic of the Last Mile problem; 10 on Safe Roadways; 8 on Equitable Access and 6 on Autonomous Vehicles. Since there were no significant differences in search behavior across the topics, for the remainder of this paper we do not differentiate between search task topics, and treat them as independent trials of the study.
3.1 What information goals do searchers have during early-stage design?

Participants reported using web search to fulfill the following information goals: (i) To get an overview of the information space: 22 participants mentioned wanting to know key concepts and terminology in their chosen topic and related topic areas. As P39 wrote, they searched “to learn more about related topics and potential avenues to go down; for basic understanding of concepts”. 14 participants mentioned wanting to know more about the “history and current practices to get the context and background” (P16). 4 participants mentioned wanting to search for perspectives of users and experts. As P99 stated, “search forums to solutions to see if other users or experts have encountered similar issues... to collect related images and concepts.” (ii) To discover design patterns and criteria: 27 participants mentioned that they used web search to analyze and evaluate found information by trying to determine patterns and critique how different pieces relate to one another through differentiating, organizing, and attributing. For example, P12 stated that they use web search to, “compare and check ideas to come up with a criteria and give me some direction”, and as P03 stated, “to look for counterpoints or alternatives”. (iii) To seek inspiration and generate ideas: 21 participants mentioned using web search to get inspiration to plan to or to generate ideas. For example, P48 says that they search, “to find design inspiration when I am starting a design... to check and compare existing solutions.” Similarly, P115 states that they “seek design inspiration from what others have done as well as find resources to make the design possible”.

3.2 What search strategies emerged to meet these information goals?

Results from the post-task survey enrich our understanding of challenges faced and strategies employed by searchers to fulfill the above-mentioned information goals. 29 participants reported struggling to formulate their informational needs as a query. For example, P9 stated, “I didn’t know what to search for... I don’t know what I don’t know and what I’m missing out on. I have this FOMO [Fear of Missing out] like feeling”. To try to better articulate their information need in an effective query, participants talked about issuing multiple queries in quick succession in an iterative manner. For example, P8 issued their first query "congestion" followed by the queries: "Car congestion san diego", "car congestion san diego map", "car congestion san diego hotspots", "car congestion san diego hotspots map", "road congestion solutions san diego", "environmental-friendly road congestion solutions san diego". Commenting on this P8 said that their strategy was to “start searches broad and then add terms to narrow down by adding terms”. This strategy was used by other participants to specify contexts (e.g. P8’s "san diego"), information sources (e.g. P8’s "map"), or other constraints (e.g. P8’s "environmental-friendly"). Another strategy to help articulate information needs as queries was to ask natural language questions – like P6 stated, “I didn’t know how to phrase it as a search. I just searched the way I would tell it to my friend and hoped something interesting came up” (for the query: “why unsafe driving behavior on the rise”). These natural language questions, and adding terms to specify the query tend to make these queries longer than keyword queries [2].

12 participants reported challenges learning domain-specific terminology. P16 illustrates this in their use of the term “hub” “I didn’t know what the term for a bus or train station generally was in urban design. Now that I found it in an article, it makes it so much easier to search” P16 used the term ‘hub’ across 7 consequent queries. Additionally, 10 participants reported challenges assessing the usefulness of information. P19 discusses their strategy to assess usefulness of search results “I now know this is a reliable source since a lot of articles refer to it”. Similarly, P18 also said, “it occurs as the top result across multiple queries so it must be relevant and trustworthy”. Participants also reported “opening webpages in new tabs for reading later” (almost like a “bookmarking” strategy).

3.3 How do web search behaviors affect creative learning outcomes?

From the search log data we defined key searching behaviors: the number, depth, diversity of queries and number of webpages opened. Doing more of these search behaviors were indicative of more “active searching behavior”.

3.3.1 Active, Diverse Searching Behavior Correlate with Learning More Facts. To analyze how search behavior relates to the change in number of declared facts post- compared to pre-search, we performed correlation analyses (see Table 2). We found a significant correlation where searchers who saw the greatest increase in number of declared facts also tended to have issued more, longer, more diverse queries, and opened more web pages.

3.3.2 Active Searching Behavior Relates to Articulating Deeper Insights and More Well-Defined Problem Statements. To understand the relationship between search behavior and the depth of learning measures, we performed ordinal logistic regression analyses. First, we explore the relationship between search behavior and how well searchers ‘understand’ the information space (as measured by Quality of Facts) [37]. Searchers who issued more queries, and opened more web pages had significantly higher increases in quality of facts mentioned post- rather than pre-task. There were no significant differences in the length or diversity of queries issued with respect to the change in quality of facts reported (see Table 3(a)). Second, we explored how search behavior corresponds to how participants “analyze” the information space (as measured by Interpretation of Facts) [37]. We found that searchers who issued more longer and more diverse queries had significantly higher increases in their interpretation scores post- rather than pre-task. There was no significant difference in change of interpretation scores with respect to the number of web pages opened (see Table 3(b)).

Third, we explored how search behavior relates to how well participants “evaluate” the information space (Critique of Facts) [37]. Searchers who issued longer and more diverse queries had significantly higher increases in critique scores, post- rather than pre-task. There were no significant differences in change of critique.

<table>
<thead>
<tr>
<th>Measure</th>
<th>r</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Queries</td>
<td>0.69</td>
<td>&lt;0.01**</td>
</tr>
<tr>
<td>Length of Queries</td>
<td>0.38</td>
<td>0.03*</td>
</tr>
<tr>
<td>Diversity of Query</td>
<td>0.18</td>
<td>0.04*</td>
</tr>
<tr>
<td>Number of Webpages</td>
<td>0.65</td>
<td>0.02*</td>
</tr>
</tbody>
</table>

Table 2: A correlation analysis found significant positive relationship between higher gain in facts stated and issuing more, longer, more diverse queries, and opening the more web. * significant at $p < 0.05$, ** significant at $p < 0.01$
This paper sheds light on how people use the web to search for information during early-stage design. By analyzing search log and survey data, we find that designers have cognitively-complex information goals during early-stage design – goals that go beyond the recall and lookup tasks that current search tools are optimized to fulfill [25, 28, 35, 41]. Designers’ information goals include learning about key concepts and terminology, history and current practices, and the perspectives of users and experts to get an overview of the information space. This exploratory behavior is consistent with research that describes how designers explore a problem space to find a problem [15] and then iteratively re-framing that problem by discovering and integrating new information [10, 27, 31]. This divergent exploration and convergent synthesis is a hallmark of the design process [18] and is exemplified in the Design Thinking model [33]. Additional information goals included wanting to discover design patterns and criteria, and seek inspiration to generate hypotheses and ideas.

By analyzing search log and survey data, we find that designers generally correspond to higher increases in creative learning outcomes (such as gathering more facts, articulating deeper understanding, and more well-defined problems). This strategy of iteratively probing the information space is similar to the design-thinking strategy of using prototyping as a method for actively probing and getting feedback from the design space and user community [10, 18, 27, 31]. By issuing more frequent and diverse queries and opening more webpages, researchers might have had more opportunities to learn terms to better articulate their queries, explore a subset of the information space and develop their relevance judgement criteria. Promoting discovery of domain-specific language and query diversity should be an important design goal for future search tools. They should build on work like [4, 6, 7, 17], to guide novice designers to articulate more diverse queries and support discovery of domain-specific language, hypothesis and idea generation.

We observed significant relationships between active searching behavior and creative learning outcomes, however, we cannot make any causal claims. It could be that these relationships are a better reflection of the designers’ skills or aptitude, rather than a function of specific strategies. Future work needs to conduct a large-scale analysis of naturalistic search behavior during design to further test the hypotheses generated by this short paper. Furthermore, the study has limitations as it tried to balance ecological validity with experimental control. For instance, we controlled all participants’ search sessions to be 30 minutes long to help us compare between participants. However, this might be different from the individuals’ usual searching behavior. Since complex search tasks such as exploratory search are often carried out over multiple sessions and devices [25, 35, 40], we encouraged participants to continue searching beyond this session, and to think of this as their first search session. Future research is required to overcome these limitations, build out and test suggested design implications of these findings.

### Table 3: More active and diverse searching relates to deeper learning and more well-defined problems. Ordinal Logistic Regression analyses results

<table>
<thead>
<tr>
<th>Measure</th>
<th>Odds Ratio</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Quality of Facts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Queries</td>
<td>1.28</td>
<td>&lt;0.01**</td>
</tr>
<tr>
<td>Length of Queries</td>
<td>0.20</td>
<td>1.20</td>
</tr>
<tr>
<td>Diversity of Query</td>
<td>1.09</td>
<td>0.08</td>
</tr>
<tr>
<td>Number of Webpages</td>
<td>1.04</td>
<td>0.03*</td>
</tr>
<tr>
<td>(b) Interpretation of Facts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Queries</td>
<td>1.71</td>
<td>&lt;0.01**</td>
</tr>
<tr>
<td>Length of Queries</td>
<td>1.23</td>
<td>0.03*</td>
</tr>
<tr>
<td>Diversity of Query</td>
<td>1.11</td>
<td>0.03*</td>
</tr>
<tr>
<td>Number of Webpages</td>
<td>0.75</td>
<td>0.17</td>
</tr>
<tr>
<td>(c) Critique of Facts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Queries</td>
<td>1.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Length of Queries</td>
<td>1.42</td>
<td>0.04*</td>
</tr>
<tr>
<td>Diversity of Query</td>
<td>1.64</td>
<td>0.02*</td>
</tr>
<tr>
<td>Number of Webpages</td>
<td>0.96</td>
<td>0.89</td>
</tr>
<tr>
<td>(d) Degree of Problem Definition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Queries</td>
<td>1.27</td>
<td>&lt;0.01**</td>
</tr>
<tr>
<td>Length of Queries</td>
<td>1.04</td>
<td>0.03*</td>
</tr>
<tr>
<td>Diversity of Query</td>
<td>1.10</td>
<td>0.02*</td>
</tr>
<tr>
<td>Number of Webpages</td>
<td>0.75</td>
<td>1.20</td>
</tr>
</tbody>
</table>

trying to surface patterns and previously unknown connections can be an effective technique to also generate ideas through creative combination (i.e. coming up with something new by combining two concepts/ideas) [5, 13, 32, 36] and analogical reasoning (i.e. the process of making connections through examples) [17, 19, 34].

Searchers reported challenges scoping broad, ill-defined information needs into queries, learning domain-specific language, and assessing the usefulness of information. Theses challenges are related to those faced by design novices [8, 10, 15, 18, 30]. Active, diverse searching strategies (such as issuing more, longer, and diverse queries, and opening more web pages) related to higher creative learning gains (such as more, better quality facts and insights, and more well-defined problems). This strategy of iteratively probing the information space is similar to the design-thinking strategy of using prototyping as a method for actively probing and getting feedback from the design space and user community [10, 18, 27, 31]. By issuing more frequent and diverse queries and opening more webpages, researchers might have had more opportunities to learn terms to better articulate their queries, explore a subset of the information space and develop their relevance judgement criteria. Promoting discovery of domain-specific language and query diversity should be an important design goal for future search tools. They should build on work like [4, 6, 7, 17], to guide novice designers to articulate more diverse queries and support discovery of domain-specific language, hypothesis and idea generation.

We observed significant relationships between active searching behavior and creative learning outcomes, however, we cannot make any causal claims. It could be that these relationships are a better reflection of the designers’ skills or aptitude, rather than a function of specific strategies. Future work needs to conduct a large-scale analysis of naturalistic search behavior during design to further test the hypotheses generated by this short paper. Furthermore, the study has limitations as it tried to balance ecological validity with experimental control. For instance, we controlled all participants’ search sessions to be 30 minutes long to help us compare between participants. However, this might be different from the individuals’ usual searching behavior. Since complex search tasks such as exploratory search are often carried out over multiple sessions and devices [25, 35, 40], we encouraged participants to continue searching beyond this session, and to think of this as their first search session. Future research is required to overcome these limitations, build out and test suggested design implications of these findings.

### 4 DISCUSSION

This paper sheds light on how people use the web to search for information during early-stage design to help them frame a problem. By analyzing search log and survey data, we find that designers have cognitively-complex information goals during early-stage design – goals that go beyond the recall and lookup tasks that current search tools are optimized to fulfill (ref. 3.1) [25, 28, 35, 41]. Designers’ information goals include learning about key concepts and terminology, history and current practices, and the perspectives of users and experts to get an overview of the information space. This exploratory behavior is consistent with research that describes how designers explore a problem space to find a problem [15] and then iteratively re-framing that problem by discovering and integrating new information [10, 27, 31]. This divergent exploration and convergent synthesis is a hallmark of the design process [18] and is exemplified in the Design Thinking model [33]. Additional information goals included wanting to discover design patterns and criteria, and seek inspiration to generate hypotheses and ideas.
REFERENCES


