Learner, Assignment, and Domain: Contextualizing Search for Comprehension

Catherine L. Smith
Kent State University
Kent, Ohio, USA
csmit141@kent.edu

Kelsey Urgo
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina, USA
kurgo@unc.edu

Jaime Arguello
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina, USA
jarguello@unc.edu

Robert Capra
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina, USA
rcapra@unc.edu

ABSTRACT

Modern search systems are largely designed and optimized for simple navigational or fact-finding tasks, with little support for complex tasks involving comprehension and learning. In response, the search-as-learning research community has undertaken a wide range of research questions focused on understanding how various types of learning outcomes are affected by searcher characteristics, the search task, and the search system. Typically, these views embed learning within a search system. In this paper we take a different view, embedding search within a framework for an end-to-end learning system designed to support learning in a formal educational context. Our central goal is to motivate research questions aligned to advance progress on techniques for active support of comprehension and formal learning. Thus we intentionally set aside goals for informal and surface learning. We argue that to be effective, such a search-centric learning system must model four key components: individual students (searcher factors), the educational domain (topic factors), academic assignments (task factors), and progress toward learning goals (the objective function of the end-to-end system). In modeling these components, our hypothetical system makes inferences about students' learning histories, knowledge states, comprehension, and the utilities of different types of information resources. We present examples of possible techniques and data sources for each model. We also introduce the novel concept of leveraging school assignments as rich task context. Our intention is not to propose a functional system, but to frame search-as-learning in the context of comprehension and to inspire research questions arising from an end-to-end view of this important research domain.

CCS CONCEPTS

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

KEYWORDS

search-as-learning, retrieval models, user models, self-regulated learning, interfaces for learning

ACM Reference Format:
Catherine L. Smith, Kelsey Urgo, Jaime Arguello, and Robert Capra. 2022. Learner, Assignment, and Domain: Contextualizing Search for Comprehension. In *Proceedings of the 2022 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22), March 14–18, 2022, Regensburg, Germany*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3498366.3505819

1 INTRODUCTION

Search-as-learning (SAL) is a multi-faceted research area that addresses the complexities of human cognition, algorithm design, and the sometimes-divergent agendas of interdisciplinary work [13, 40]. A central goal of SAL research is to improve learning for search system users. Of course, learning has many dimensions, which researchers approach from different perspectives across important research questions. Some view learning broadly as knowledge acquisition—a natural byproduct of searching for information [27]. Others focus on learning as a type of search-task goal [14]. Still others see learning as the central goal, and the search system as a support for learning [9]. Such differences affect the measures of learning used in research. Learning is sometimes measured against an expert level of knowledge [7]. Often, short or long-term recall from memory serves to gauge knowledge gains [49]. In contrast, constructive measures of learning assess the ability to use knowledge, where measures of comprehension are a benchmark for learning [34]. The design of a search system also implies an underlying perspective on learning. In some views, the goal of the system is to infer learning goals and gains from behavior, where no explicit context is available [56]. In other views the goal is facilitate an explicit learning goal known to the system [50].

Building on prior work from across the SAL community, this paper looks across these many views. In our view, a focus on the explicit goal of comprehension is likely to yield the most challenging and rewarding research questions. We argue for a unified research agenda that aligns perspectives on this measure. Our argument is couched in a sketch for a blue-sky framework designed to facilitate comprehension in the specific context of undergraduate education, where this goal is highly likely for users. Here, comprehension is defined as the construction of meaning from source materials,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHIIR '22, March 14–18, 2022, Regensburg, Germany
© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9186-3/22/03…$15.00
https://doi.org/10.1145/3498366.3505819

where learning is measured as demonstrated ability to use comprehended meaning in novel situations [16]. Comprehension is a well-established construct in the learning sciences, with attendant theoretical underpinnings and validated instrumentation [15]. We include here the important construct of multiple-text comprehension [8]. The undergraduate learning environment is but one of many possible use cases for a search-centric learning system [47]. Our focus on it here does not preclude the applicability of our ideas to other settings such as secondary and graduate education. We selected this setting because we believe it provides a clear view of the interdependencies, challenges, and opportunities in designing for comprehension, and because the ubiquity of undergraduate digital learning environments affords unique opportunities with respect to empirical questions.

Learning for comprehension is a primary goal across undergraduate academic domains, but the manner of learning depends heavily on the domain of learning. Pedagogy, here taken as practice [2], defines a means for achieving and demonstrating comprehension within a learning domain. Digital environments for learning are replete with evidence of domain pedagogy and related knowledge structure. For example, university websites express the varied, complex associations between knowledge domains (e.g., academic areas), learning goals (e.g., theory, practice), and curriculum (e.g, programs and courses). Learning management systems contain pedagogical artifacts such as syllabi and course assignments. Of course, within this structure, individual characteristics of learners drive the manner of learning [10]. Digital learning environments may also provide information about learners' skills, strengths, goals, and needs.

We argue that progress in the SAL research agenda can be accelerated by considering research questions within a unified experimental framework that supports undergraduate education. Our framework includes abundant learning context, accessible *in situ* learning tasks, and learning measures aligned with the goals of comprehension. The paper describes components of the framework, including models of the learner, assignment, domain, corpus, and learner behavior, and how these models might capture and leverage information to support learning and reflection on learning [11]. Our goal here is not to present a functional schematic but a thought-experiment concerning how to better align research questions with developing techniques in support of SAL.

The paper is organized as follows. In the next section, we briefly review selected prior work taken from the perspectives above. A simplified functional overview of the framework follows. We then describe each of eight components in more detail, following a simple example throughout our discussion. We conclude with a brief review of possible interface functionality that leverages the available models to facilitate reflective learning.

2 RELATED WORK

Studies in SAL have investigated how various factors influence learning during search. We present a brief overview of selected papers focused on four sets of factors: characteristics of the searcher [31, 41, 41, 53], the search task [18, 23, 27, 28], search behavior [1, 7, 12, 17, 26, 29, 32, 56] and the search system [9, 16, 24, 42, 54].

Searcher characteristics. Research here focuses on how factors such as prior knowledge [31, 41, 53], working memory capacity, and reading comprehension [34] influence learning. For example, [31] found that searchers with greater prior knowledge had lower knowledge gains, possibly because they were less likely to encounter new information while searching. By contrast, [53] found that searchers with greater prior knowledge had greater knowledge gains, possibly because they were able to search more effectively. [41] found that knowledge gains were greater for domain novices earlier in the search session and that experts gained more knowledge towards the end. [34] found a positive relationship between working memory capacity, reading comprehension, and learning.

Task characteristics. Studies of how task complexity affects learning [18, 23, 27] find an association between complex tasks and types of mental activities [18], lower knowledge gains [23], and learning that was dependent on task progress, with greater knowledge gains later in the search session [27]. Investigating learning during multi-session search, [28] found that knowledge gains plateaued faster when subtasks had no dependencies.

Search behavior. Searchers with better learning outcomes tend to spend more time reading documents [12, 17, 29, 49, 56]; issue queries with more advanced or uncommon vocabulary [7, 12, 17]; issue more diverse queries within a search session [32]; review more results that are relevant and novel [1, 12]; and use resources that are more suitable to the task [26].

System characteristics. Research on interfaces and interactive tools designed to support learning, as well as the retrieval system itself, generally suggests that users benefit from interfaces that convey information about the items in the collection [24, 54] and from tools allowing document annotation [16, 42]. However, tools can have unintended effects such as cognitive overload [42] or effects on characteristics of the learner's goals [9]. Studies have also investigated retrieval models optimized to improve learning. In an evaluation of an end-to-end system to support vocabulary learning by presenting unfamiliar vocabulary words in context, the model was optimized to *maximize* learning potential by promoting items with greater term density and to *minimize* effort by favoring shorter texts, and resulted in improved vocabulary learning [49, 50]. Some of the ideas in this position paper resonate directly with this work. For example, we advocate for a retrieval model that favors documents based on students' current learning objectives and state of knowledge. However, we go beyond vocabulary learning to model different factors, including characteristics of the individual learner, the educational domain, the learner's history of assignments, and the current assignment.

3 FUNCTIONAL OVERVIEW

Our framework is designed to support undergraduate students as they complete academic assignments throughout their studies. The goal is to facilitate comprehension within a domain of study while reducing the effort needed to find and use information resources (lectures, books, discussions, library sources, websites, etc.). We seek to support individual students in completing a class assignment by providing information that best supports each individual. We place the framework in a learning system that integrates search, tools (e.g.,

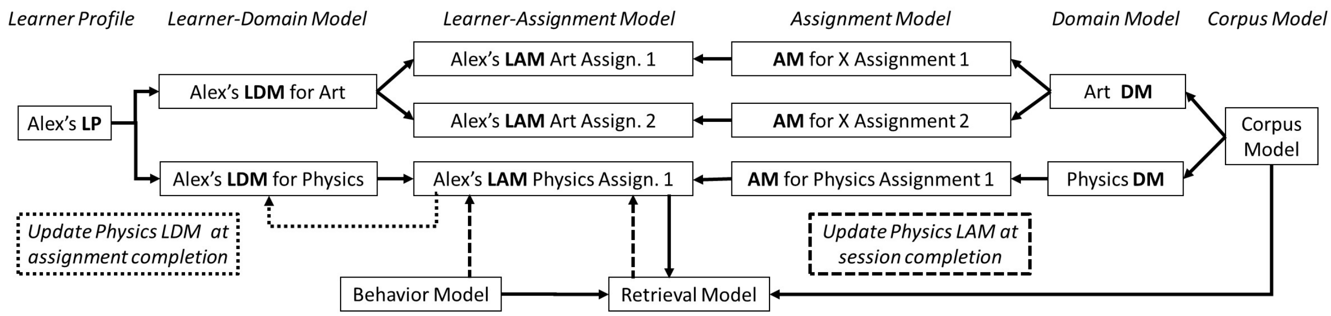

Figure 1: Overview of model framework across multiple assignments and domains.

word processor), and academic context (e.g. assignments). At a minimum, the system needs representations of information resources, the learner, and the assignment, with a mechanism for improving the representation of the student as understanding evolves over multiple assignments within a domain. The framework comprises eight models that together facilitate these goals. For purposes of simplification, the model describing information resources (*Corpus Model* - CM [see 4.3]) is not detailed in this overview. Two models describe the learner. Two models describe the assignment within its broader academic context. The three remaining models work together dynamically to optimize the retrieval system for learning and effort across multiple sessions during assignment completion. Figure 1 depicts the components and their connections, using the example of Alex, a hypothetical student presented throughout the paper. Alex is studying in two domains, art and physics.

Descriptions of the learner: The *learner profile* (LP) describes students' skills and academic history - factors likely to affect the utility of different types of information resources and to indicate prior knowledge. Derived from existing records, these relatively stable traits cross academic domains. For example, Alex's history includes success in a secondary school physics course but they have low reading skill. A detailed description of the learner, the *learner-domain model* (LDM), describes the learner separately for each domain of study. It predicts the utility of different resource types for the domain (*information utilities*), and estimates the learner's *prior knowledge* in the domain. For example, Alex is likely to benefit from materials with an easy reading level but that require prior knowledge in physics. The LDM updates at the completion of each assignment, as described below.

Descriptions of academic context. Students work in a rich academic context, the structure of which is expressed in many information sources. We argue that the conceptual and pragmatic aspects of academic domains can be discovered and described using existing sources. Academic domains are represented by the *domain model* (DM), which has two goals. First, it organizes the domain's academic subject matter as a knowledge graph. Second, it describes characteristics of the subject matter such as pedagogical practices, the types of information resources used in teaching, and the relative utility of those resources. In our example, Alex is studying physics, where fluid dynamics is a central subject area, with assumed prerequisite knowledge and useful types of information resources modeled for the domain. The DM informs models of assignments, which we discuss next.

Supporting students in assignment completion requires a description of the assignment task and its learning objectives. This is the role of the *assignment model* (AM), which predicts subject matter required for completing the assignment, including prerequisite knowledge, concepts to be learned (*target knowledge*), possible paths an average student might take, and *work products* needed for completion (e.g., notes, essay). Finally, it predicts the utilities of resource types likely to be useful. These predictions are explicit in or inferred from the assignment text, or inferred from the domain model (DM). Our example centers on an assignment about "lift." The assignment model predicts that completing the assignment requires understanding of "fluid" and "force," the production of "notes" and an "essay," and that visual simulations have high utility for learning.

Dynamic models. The framework's three dynamic models combine information from the other models to support completion of an assignment. The *learner-assignment model* (LAM) customizes the assignment model (AM) to optimize for the student's learning needs, as predicted by the learner-domain model (LDM). Thus the LAM predicts the knowledge students will learn from the assignment and their possible paths to that knowledge. It also adjusts the predicted utility of information resource types to optimize for students' individual needs. Resulting values are passed to the retrieval model (RM) for use in retrieving the first search engine results page (SERP) of the first work session on the assignment. For Alex, the LAM for the "lift" assignment predicts both limited prior knowledge of "fluid" and "force" and a learning path that includes these concepts. It also predicts high utility for easy reading materials that support understanding of visual simulations.

The retrieval model (RM) uses the feature values from the learner-assignment model (LAM) for the first SERP of the session. Below we propose a learning-to-rank (LTR) model that optimizes for session relevance and learning gain. The RM is dynamic in that within a session, it adjusts feature values between queries using feedback received from the behavior model (BM), as explained next. (The RM is supported directly by a CM (see 4.3), which we have excluded here.)

The behavior model (BM) observes learner interaction with resources opened from the SERP or with saved documents. Observations include eye-tracking for visual attention and keyboard/mouse engagement during writing. The model associates observations with engaged resources to produce three types of estimates. First, it estimates the utility of engaged items from the active SERP. For example, observations of visual attention enable a model of reading behavior that predicts potential for comprehension, as well as effort invested, to estimate the utility of the engaged item. Second, the model uses observations of behavior to estimate the amount and quality of attention given to relevant content areas. For example, when Alex reads and writes bullet points about "flow deflection" and "Newton's laws," the model estimates their knowledge level for these concepts. Finally, the model classifies writing products, for example, differentiating "notes" from "summaries." For example, Alex's bullet points are classified as "notes." These three types of estimation are updated continuously as Alex continues to work. When a new query is submitted, an update is triggered to the retrieval model (RM) before retrieval for that query. The updates adjust feature values on Alex's knowledge of the concepts and information utilities, as well as the status of work products. The BM is reinitialized with the generation of the next SERP.

Other dynamic updates. Two other dynamic updates are important to the framework, one at the detection of a *session break*, and the other at the detection of *assignment completion*. When a session ends, the retrieval model (RM) pulls final values from the behavior model (BM) and updates final feature values on utilities and knowledge state. At the same time, the behavior model (BM) estimates work product status and effort. The learner-assignment model (LAM) then uses values from the RM and BM to update utilities, knowledge state, and the status of work products in preparation for the next task session. When assignment completion is detected, the applicable learner-domain model (LDM) updates using the final values from the LAM, reflecting changes in the learner's domain knowledge and information utilities within the domain.

To summarize, the framework maintains a model of what the student knows and is learning within each academic domain; the types of information resources most likely to benefit the learner in that domain; and work the student is completing to advance learning.

4 COMPONENTS

This section of the paper provides more detail on each component, in an order slightly different from the functional view above. First, we address the learner profile (LP) and domain model (DM), as both provide information necessary to other models, and are independent from those models. Next, we detail the assignment model (AM) and learner-domain model (LDM), both of which depend on one other model and which are interdependent. Then, the learner-assignment model (LAM) is covered, as it depends on both the AM and the LDM. The section ends with the retrieval model (RM) and the behavior model (BM).

4.1 Learner Profile (LP)

The LP captures three types of information about the learner: neurodiversity (e.g., dyslexia), skills the learner has acquired to help them learn (e.g., self-regulated learning skills), and academic focus and history (e.g., major, transcripts). Profile information helps predict aspects of the corpus likely to be useful to an individual learner, as well as the learner's prior knowledge. The LP uses established instruments from the learning sciences.

Assessment of neurodiversity depends on a wide range of instruments. For example, validated and reliable professional screening instruments, that include self-report questionnaires, can detect conditions such as dyslexia and attention deficit/hyperactivity disorder [20, 48, 55].

The LP also captures evidence of learning skills, including reading level, self-regulated learning (SRL, the ability to monitor and control one's own learning), and self-efficacy (i.e., one's perceived ability to complete a task). All of these skills may be assessed with well-established instruments. For example, the Independent Reading Level Assessment framework, captures a learner's reading level [37]. Effective SRL improves learning outcomes [46] and may be assessed using the Motivated Strategies for Learning Questionnaire [35]. Self-efficacy affects an individual's ability to learn [45], which may be measured using self-reported scales [44]. SRL skills and self-efficacy levels are useful for understanding the learning support a student may require. In Section 5, we discuss how the system can prompt and encourage learners to engage SRL skills and reflect on their learning.

Finally, the LP captures the learner's academic focus and history. This includes the learner's major, minor, and course history, such as preparation in upper secondary school.

In our running example, Alex's profile captures, among other information, a low reading level, major in physics, and successful completion of a physics course during secondary education. Alex's profile also reveals mid-range SRL skills and self-efficacy. This information will provide context for subsequent models (e.g., LDM) to predict Alex's knowledge from preparatory studies and the types of resources most likely to be useful for Alex.

4.2 Domain Model (DM)

The DM is responsible for capturing, inferring, and storing important information about an educational domain (e.g., physics). To be useful, this information must be structured for use by other models. Here we chose to use a knowledge graph comprising *nodes*, *node attributes*, and *edges*. Nodes correspond to units of information within the domain; attributes represent metadata about the nodes; and edges represent relations between nodes. What these elements encode is an open research question and is likely to depend on the domain. In this section, we explore information that might be encoded and discuss some possible resources (e.g., course syllabi) for capturing or inferring this information.

Nodes represent important units of information in the domain. For example, for the humanities, the nodes might represent people, places, organizations, and events. For the social sciences the nodes might represent concepts, procedures, and organizations. As an example, textbook indices can be useful for determining "index-worthy" nodes to include in a specific DM.

Node attributes encode information about nodes to help the system predict relevance, information utilities, and a learner's state of knowledge. To illustrate, throughout the remainder of the paper we discuss possible attributes associated with *conceptual* nodes. Education research defines conceptual knowledge as knowledge about concepts, categories, principles, theories, and models [4].

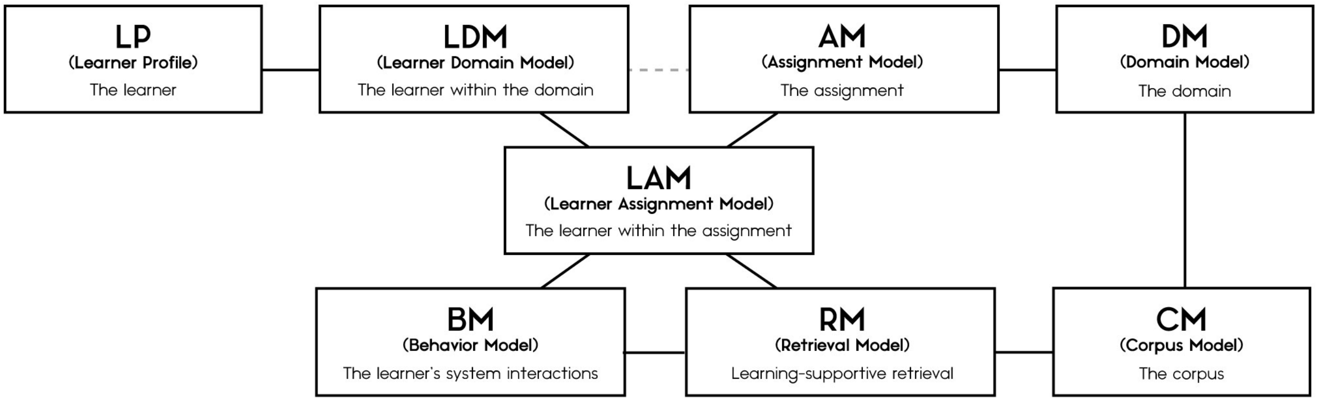

Figure 2: Overview of framework and information captured by each component (e.g., LP captures information about the learner).

Concepts are useful for organizing a body of knowledge in a systematic and interconnected manner [4, p.79]. Therefore conceptual nodes are likely to be prevalent in most DMs. As discussed next, we see four different types of attributes associated with conceptual nodes: *alternative identifiers*, *topic language*, *learning complexity*, and *information utilities*. Other attributes are likely to be pertinent and we see this as an open and rich research question.

First, *alternative identifiers* provide the system with alternative ways to refer to a node. For example, the DM should indicate that "Newton's 1st Law" and "inertia" refer to the same construct in classical physics. This attribute may be sourced from resources such as textbook indices (e.g., "see also") and anchor text in educational websites (e.g., Wikipedia). Second, the system needs a *topic language* that represents the node. This could take the form of a language model or low-dimensional representation (e.g., using a neural embedding technique) derived from external corpora. The topic language helps other models, for example associating nodes with documents in the corpus (e.g., the CM, see section 4.3) or with a learner's work products. For example, concepts in a student's written text may suggest an increase in knowledge of the corresponding node. Third, *learning complexity* is the complexity of domain knowledge required to understand the concept represented by a node. For example, economics concepts like "supply" and "demand" are fundamental, while "inflation" and "deflation" are more complex. Learning complexity helps the AM (see section 4.4) estimate knowledge dependencies in the domain. Many resources capture this type of evidence, including educational websites, textbooks, and course syllabi tailored for specific learners (i.e., novices vs. experts). Additionally, position information may also provide useful evidence. For example, foundational topics are likely to precede advanced topics in textbooks and course syllabi. Finally, attributes associated with *information utility* characterize educational materials used to teach a concept. Such information characteristics might include the type of resource (e.g., textbook, news article), the type of media (e.g., image, interactive visualization), and the genre (e.g., definition, summary, example, comparison). To illustrate, the concept of Bernoulli's principle may be taught most frequently with definitions and examples, such as lift acting on an airplane wing. Machine learning may be used to capture and store these utilities. For example, given a specific feature representation (i.e., information characteristics), one could train a simple linear model to predict the presence/absence of topic in a document. In this respect, the learned weights in this model will encode the characteristics of educational resources that are desirable and undesirable for the given topic. These weights can then be used to promote content that is appropriate from an educational standpoint, even if the material is not explicitly educational.

Edges in the DM represent important relations between nodes, with associations at different levels of specificity. The edges help the system support and infer learning. To support learning, the edges enable identification of concept dependencies. For example, if a learner wants to understand "inertia," understanding "conservation of momentum" is necessary. To infer learning, the edges enable the system to propagate inferences about what the learner knows. For example, a learner who understands the principle of inertia probably also understands the foundational concepts of "velocity" and "force." Next, we describe three types of edges in our hypothetical DM and present some example resources and techniques for inferring such relations. However, as with node attributes, we see this as an open research question.

First, simple edges indicate the extent to which nodes are related, without specifying the type of relation. Weighted edges representing the strength of concept relations can be derived using a large corpus and a co-occurrence statistic (e.g., pointwise mutual information). Second, directed edges capture prerequisite relations between concepts. For example, edges can represent that understanding "concept B" usually requires prior understanding of "concept A." Inferences to foundational concepts for understanding "lift" might derive from frequent use of "fluid" and "force" within text passages that introduce "lift" (e.g., definitions, summaries, introductory materials). Knowing prerequisite relations helps the system do three things: make inferences about the learner's preparation for learning a concept; recommend resources based on prerequisite concepts; and propagate evidence about what the learner knows (e.g., a learner who knows B likely knows prerequisite A). Prior research has proposed techniques for automatically predicting prerequisite relations between concepts from educational materials [25, 33, 43]. These techniques may provide a starting point for implementing this component of the DM. Finally, edges may encode information about nodes that are frequently learned in conjunction. This is particularly important for concept learning. Concepts exist within a greater context and are often defined by their relation to other concepts, such as similarities and differences. For example, the concept of "velocity" is often differentiated from "speed." Such relations could be mined from textual corpora by considering concepts that are frequently compared and contrasted.

Generally, the DM supports the system as a whole by providing the structure and content of essential information needed by various models. Prior work in SAL has introduced the idea of manually generated topical outlines [9]. Our framework extends this idea by proposing automatically generated relationships among concepts contextualized by the assignment, as we discuss further below.

4.3 Corpus Model (CM)

The CM has three main tasks. First, it segments the corpus into units of analysis smaller than the document level, such as sections, paragraphs, tables, figures, examples, images, and videos. Second, the CM maps units in the corpus to nodes in the DM by identifying all the locations in the corpus that mention a node. This mapping process is analogous to entity-linking. A wide range of possible approaches to this task have been developed (see Reinanda et al. [39] for a review). Finally, the CM characterizes units in the corpus by automatically generating document features that convey important characteristics that support learning. Examples include the document reading level, source information (bibliographics), degree of explicit structure (e.g., text density), the presence of visual information, and content characteristics such as "objective" versus "subjective." Identifying these characteristics and their value in support of learning are open research questions (see for example [3]), and evidence from the learning sciences provides additional starting points supported by learning theory.

4.4 Assignment Model (AM)

Undergraduate assignments are often difficult to parse [5, 19], requiring significant inference from the content of lectures, labs, and other learning experiences. The AM seeks to leverage assignment texts in two ways. First, it detects explicit task attributes found in the assignment and infers missing attributes. Second, it uses information from the DM to contextualize the assignment with respect to target and prior knowledge, possible paths for completing the assignment, and the characteristics of resources likely to be helpful in completing the assignment.

Initially, the AM assesses an assignment to identify target knowledge (concepts to be learned) and a set of explicit assignment attributes such as work products (e.g., essay, notes), learning objectives (e.g., evaluate, summarize), and resources (e.g., references, links). Target knowledge and assignment attributes encode information that helps the AM predict the structure of possible subtasks for the assignment. Where explicit attributes are missing or ambiguous, they are inferred using a language model for assignments within the domain and existing learning frameworks such as [4]. If the assignment lacks explicit learning objectives (as many do [5, 19]), the AM infers them. Where an assignment involves student selection of their own topic, such as in expository writing courses, the system infers learning objectives with respect to the domain of expository writing, which includes constructs such as topic development. Also, while assignments often contain an explicit description of the final work product to be submitted for a grade, interim products may also be inferred from the task structure.

Next, starting with nodes representing target knowledge in the DM, the AM generates a graph for the assignment domain, the *assignment graph*, which contextualizes the assignment within the applicable DM (see section 4.2, above). Two weights are generated for each node. *Prior-knowledge weights* use learning complexity and other data to estimate the probability that an average student will understand the concept before starting the assignment. *Learning weights* estimate the probability that the concept will be understood if the assignment is completed successfully. Nodes dependent on target knowledge also contextualize the assignment with respect to upcoming learning of concepts that depend on the target. The AM then associates node concepts with applicable attributes from the DM such as information utilities (see section 4.2, above). Finally, the AM associates the assignment attributes with nodes in the assignment graph. The resulting final graph represents task structure for the assignment, generally as a set of subtasks with attendant learning goals, work products, and information utilities.

In our running example, the first assignment in the physics class is vague and open-ended. It states: *The purpose of this assignment is to explain the phenomenon of lift in an essay focused on concepts covered in the course so far (see the list below). Discuss how the concepts explain lift, and be sure to use evidence and logic to show how the concepts are related. See chapter 3 of your text.*

Here, the target knowledge "lift" is explicit, as are the other concepts in the list (not shown). While no explicit learning objectives are found, the assignment includes general goals using the verbs "explain" and "show." From the target concepts and the extracted verbs, the AM generates the learning objective, "Show how principles explaining lift are related." Finally, the AM identifies "essay" as an explicit work product, the referenced chapter 3 as an explicit resource, and infers "notes" as an interim product.

Next, the AM uses the physics DM to generate the weighted assignment graph (see Figure 3). The graph contains subtopic structure, with two subtopics, one for "Bernoulli's principle" and the other for "Newton's 3rd law." The edges also suggest possible paths through topics. For example, understanding of "pressure," "velocity," and "fluid" is required to understand Bernoulli's principle. The AM then generates prior-knowledge weights predicting that average students have only intuitive knowledge of some foundational concepts (not depicted), and no knowledge of Newton's laws or

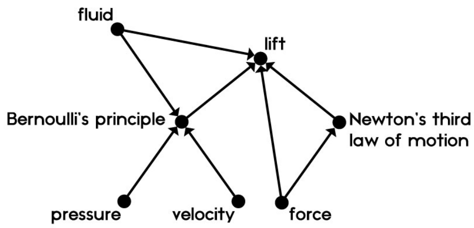

Figure 3: Example assignment subgraph for "lift".

Bernoulli's principle. Learning weights predict that all concepts in the graph will be understood with successful completion of the assignment. The AM then associates the assignment graph nodes with information utilities from the DM, including a utility for the referenced textbook chapter. For example, summary textual materials have the highest utility for foundational concepts, while animated visual simulations have the highest utility for learning Bernoulli's principle. Finally, the AM associates assignment attributes with nodes on the graph, including inferred work product "notes" and the learning goal "explain" for both subtasks.

4.5 Learner-domain model (LDM)

The LDM combines information from the LP with information about a single domain of study. The LDM instantiates itself with an initial estimation of a learner's knowledge in the domain, then updates that representation as the student continues to learn in the domain. The LDM also maintains a record of the student's personal information utilities for learning resources in the domain. These functions occur at different moments and frequencies, with instantiation occurring once for a domain, and updates occurring each time the learner completes an assignment within the domain.

When a student begins the first assignment in a new domain, the LDM estimates what the learner already knows within the domain. To do this, the LDM uses the relatively static data in the student's LP and the average prior-knowledge weights from the AM. With these data, the LDM creates a weighted *learner graph* that estimates the learner's knowledge in the domain by adjusting the initial AM weights to represent the probability that *this* learner understands each node concept. In our running example, Alex's profile indicates successful completion of a preparatory physics course. For the "lift" assignment, the AM predicts that average students have no prior knowledge of Newton's 3rd law. The LDM instantiates itself by estimating Alex's knowledge of each concept in the AM's assignment graph. The resulting learner graph weights reflect the likelihood that Alex has some prior knowledge of Newton's 3rd law. During instantiation, the LDM also adjusts information utilities for nodes in the graph, for example, predicting that because Alex has low reading skills, utility will be highest for introductory texts with an easy reading level.

After instantiation, the LDM is updated every time an assignment is completed in the applicable domain. After each assignment is completed, the LDM uses learning estimates from the just-completed LAM (see Section 4.6 below) to revise the prior weights in the learner graph. The LDM also updates the learner's prior information utilities using utility estimates from the LAM. In our example, when Alex completes the "lift" assignment the LAM estimates that they understand Bernoulli's principle, Newton's 3rd law, lift, and related concepts. The LDM thus increases the weights on these concepts in the learner graph and propagates knowledge increases to related nodes. Additionally, the LDM updates Alex's personal information utilities for physics resources. The utility estimates from the LAM reflect Alex's engagement with resources while working on the assignment, as discussed in Section 4.6 below. In this manner, for each subsequent assignment in physics, Alex's LDM is updated to reflect knowledge gains and changes in personal information utilities.

4.6 Learner-assignment model (LAM)

The LAM contextualizes the AM, representing one learner working on the assignment. After instantiation, the model is updated every time the learner completes a work session on the assignment. Thus, between task sessions the model estimates the learner's knowledge state and information utilities, while also tracking progress on assignment subtasks and products. Below, each function is discussed in turn, including derivation of initial values, updates to values, and how these relate to other models. Here we assume the assignment is not the first one for the domain. Of course, there are likely to be complex interdependencies among information utilities, changes in knowledge, and task progress, but we do not speculate on those dependencies. Rather, we provide very general examples of possible associations between these factors and outcomes of interest. We argue that these interdependencies comprise open and important research questions.

Knowledge state. The LAM maintains an updated representation of the learner's prior knowledge and the target knowledge to be gained while completing the assignment - the *learner-assignment graph.* The goal is to help the RM prioritize resources needed to close the gap between prior and target knowledge as work continues on the assignment. Like the AM (see section 4.4 above), the graph has weights on each node, here representing the learner's prior knowledge (prior knowledge weights) and the assignment target knowledge (learning weights). Prior-knowledge weights are instantiated using the prior-knowledge weights from the learner's LDM, while learning weights from the AM are retained without adjustment. Both sets of weights are then used to generate features values passed to the RM for use with the first query of the first task session. After instantiation, updates to the knowledge representation derive from estimates provided by the RM at the end of a work session (see section 4.7 below), when the LAM updates its prior-knowledge weights to reflect estimated learning in the just-ended session. With respect to our running example, when Alex completes a work session on the "lift" assignment, the RM returns values indicating a high likelihood that Alex now understands "fluid" and "force." The LAM updates its prior-knowledge weights on these concepts in preparation for their next work session on the assignment.

Utility estimation. Different types of information resources have varying levels of usefulness for learners, depending on a learner's profile and prior knowledge in a domain, as modeled by the LDM. The LAM estimates these utilities specifically for this learner working on this assignment. The estimates comprise a set of feature values passed to RM at the beginning of each assignment task session, where one or more features describe each resource type (see section 4.7, below). Before the first work session on the assignment, the LAM instantiates itself with estimates based on average utilities found in the AM, which the LAM updates using the learner's current personal utility estimates for the domain, found in the applicable LDM. In our physics example, the AM predicted that for average students, summary textual materials would have high utility for review of foundational concepts. For Alex, who has low reading skill, the LAM reduces utility estimates for summary text and increases utility estimates for animated visual simulations. After instantiation, updates to LAM utilities derive only from data provided by the RM, as detailed in section 4.7 below. Thus, in preparation for the start of a subsequent work session on the assignment, the LAM updates its prior utility estimates to reflect the just-ended session. In our example, Alex experiences very low utility from written text during the most recent session, so the LAM further reduces its estimated utility for text and increases the estimate for visual resources, which had more utility for Alex. When the next work session begins with a new query, the RM uses Alex's now-updated prior utilities from the LAM to favor resources likely to be useful to Alex for this assignment.

Task structure. The LAM also serves to track a learner's path along predicted possible subtask paths in the learner-assignment graph. This task structure is instantiated using the assignment graph and task attributes found in the AM. The resulting representation of the task generates initial feature values, which inform the RM of the characteristics of the learning task likely to be underway at the beginning of the first session. As with utilities and knowledge, the LAM's initial model is updated at each session break. Here the LAM obtains descriptions of work products directly from the BM (see section 4.8 below). For example, at a session break, observations made by the BM result in a prediction that Alex wrote "notes" on several concepts. Using these estimates, the LAM updates its representation of the task to reflect Alex's work products for subtasks related to these concepts. It then records the status of predicted work products, estimates progress on predicted subtasks, and adds unpredicted products. These data, along with the adjusted learning weights, are then used to model possible new subtasks and concept areas for the next task session. Finally, the LAM generates values on task-related features for use by the RM in the first query of the next session.

4.7 Retrieval Model (RM)

The goals of the RM are to retrieve information that is both relevant to the target knowledge of an assignment, and that allows the learner to leverage their prior knowledge while favoring resources consistent with their *information utilities* in the domain. We have chosen to frame the RM as a learning-to-rank (LTR) model. Generally, LTR algorithms use features that characterize the query (e.g., length), the document (e.g., clickthrough rate), and the query-document pair (e.g., BM25 score). For the learning tasks we are modeling, input features also capture characteristics of the individual learner (e.g., prior knowledge) and their learning objective (e.g., understand a concept). LTR algorithms can exploit useful feature interactions, accounting for the effects of one feature on others, which is essential in modeling the complex context of a unique learner working on a specific assignment. LTR approaches can be trained and updated using implicit feedback [22] such as scaled behavioral measures correlated with performance goals. The RM has four main functions, which support the objectives of the framework when generating a ranking.

Knowledge features. Because the RM favors information that enables the learner to acquire the target knowledge and leverage their prior knowledge to facilitate comprehension, the RM obtains information from the LAM (see section 4.6 above). This information is passed to the RM in the form of feature vectors that characterize target knowledge (ϕ_{TK}) and another that captures prior knowledge (ϕ_{PK}). These feature vectors take the form of probability values over the node concepts. For ϕ_{TK}, high values indicate that the learner needs to acquire knowledge about those node concepts. Similarly, for ϕ_{PK}, high values indicate that the learner has high levels of prior knowledge about those node concepts. To use this information for ranking, the RM leverages information from the CM, which associates each document or information unit with node concepts. Let ϕ_{D} denote a vector of probabilities that map document D to nodes in the DM. In this respect, the RM generates features that capture the extent to which document D contains information relevant to the learner's target knowledge (i.e., $\mathbf{sim}(\phi_{\text{D}}, \phi_{\text{TK}})$) and their prior knowledge (i.e., $\mathbf{sim}(\phi_{\text{D}}, \phi_{\text{PK}})$). Using these two types of similarity features, the RM will favor documents that are both novel and understandable. We have proposed an approach that explicitly models the similarities between ϕ_{D} and ϕ_{TK} and between ϕ_{D} and ϕ_{PK}, however, alternative solutions pertain.

Information utilities. The RM is also responsible for favoring content that is consistent with the information utilities inferred by the LAM. As previously mentioned, information utilities are influenced by characteristics of the individual learner (e.g., reading level) and their current objective (e.g., to understand a concept). Information utilities are passed from the LAM to the RM in the form of a feature vector (ρ_{IU}). While ρ_{IU} can take different forms, one alternative is for ρ_{IU} to contain probabilities over a predefined set of resource attributes (e.g., reading level, visual content, examples, etc.). Here, a high probability indicates that the learner is likely to benefit from information with those characteristics. To use this information for ranking, the RM leverages information from the CM, which characterizes each document and information unit in the corpus along a set of dimensions. Let ρ_{D} denote a vector of probabilities that characterize document D along the same set of predefined characteristics. Same as above, the RM can generate features that capture the extent to which document D matches the information utilities from the LAM (i.e., $\mathbf{sim}(\rho_{\text{D}}, \rho_{\text{IU}})$).

Task context. In addition to features on target knowledge, prior knowledge, and information utilities, the RM also uses features about the learner's current task in the context of the assignment, derived from the LAM. This information may include characteristics of the current subtask (e.g., goal and knowledge type), as well as the stage of completion. Both factors are likely to influence the types of information that will benefit the learner. To illustrate, if the current subtask involves the learning objective "understand," the learner may benefit from definitions and examples. On the other hand, if it involves the learning goal "compare," the learner may benefit from multiple perspectives. In terms of task stage, during the beginning stages of the task, the learner may benefit from isolated units of knowledge such as definitions, while in later stages information situating concepts in a larger context (e.g., concept comparisons) may be more useful. The RM is responsible for tailoring the search results based on these contextual factors. To this end, these task attributes from the LAM can be added directly as input features in the LTR model. By exploiting feature interactions, the LTR model can learn to use these contextual cues to favor documents or information units with certain characteristics, captured in ρ_{D} for each document D.

Updating knowledge states, utilities, and task context. Finally, at the end of the search session, the RM (with the help of the BM) is responsible for updating the LAM with respect to the learner's knowledge state, information utilities, and the current subtask. Specifically, the RM and BM work together to infer the knowledge acquired during the search session, the characteristics of resources the learner successfully engaged with for the purpose of learning, and the extent of progress made during the current subtask. As described next, the BM makes these inferences based on the learner's interactions on the SERP, reading activities on pages and documents viewed, and writing activities.

4.8 Behavior Model (BM)

The goal of the BM is to observe the learner's interaction with learning and information resources to estimate the level and quality of engagement for the prediction of information utilities and comprehension. The model has four elements that form a fast and dynamic unit that provides information for the RM at each query iteration. The units include a reading component, a writing component, a information utility model, and a vocabulary model. These components are reinitialized for each SERP generated during a session. The BM also has a less dynamic element that stores persistent information about work products engaged during an active session. The technologies and modeling that underlie these hypothetical components are active research areas within SAL [7, 51, 52, 56] and other communities [21, 30, 36, 38].

Using psychophysical instrumentation of engagement (eye/tracking, mouse movement, keystrokes), the reading component observes visual attention and the writing component classifies work products. The reading component classifies visual attention as to type (e.g., skim, scan, read) and difficulty level (e.g., struggling) to estimate a probability of comprehension for engaged resources. The writing component classifies digital work products produced or engaged during interaction (e.g. note, code, calculation) and the type of effort used during engagement (e.g., copy/paste, draft, edit). Both components also observe engagement with vocabulary to track terms engaged during reading and writing. These data are passed to the vocabulary model. The two components continue observation throughout interaction with each individual SERP and open document.

When a new query is submitted, the BM uses accumulated data from the reading and writing components to estimate information utilities and the learner's understanding of concepts. The utility model uses data from both components to derive a utility estimate for each resource type engaged during the period. The vocabulary model uses behavior and vocabulary data from both components to estimate a probability that the learner understands the concepts engaged. Feature values expressing utility estimates and concept knowledge are then generated for use by the RM for the new query.

In our running example, the BM observes Alex's behavior with opened documents: scanning an academic article and closing it; scanning the top figure in a web page; rapidly scrolling and scanning several blocks of text in an advanced physics textbook; and finally reading for comprehension in an introductory physics textbook, where Alex begins by skimming, then returns to the top to read through sections discussing "flow deflection" before giving extended attention to a page of figures. Here Alex uses visual attention indicative of the intention to read for comprehension (i.e., not skimming) where there is little evidence of difficulty (e.g., no excessive fixations or regressions). Alex then switches to writing notes on "Newton's 3rd law" before submitting a new query. The BM then predicts that academic articles and advanced texts have low utility for Alex, and introductory texts and figures have high utility. It also estimates that Alex understands the concept of "flow" as it relates to "Newton's 3rd law." As mentioned above, before retrieval for the new query, the RM uses values from the BM to update its prior feature values on Alex's utilities and concept knowledge. Retrieval for the new query will demote resources with lower utility and promote the more useful resources. The next retrieval will also reflect the prediction that Alex understands the concept of flow as it pertains to "lift" and "Newton's third law."

In addition, the writing component of the BM classifies attention and effort given to work products during the session. As new or existing products are engaged during a session they are identified and classified by type. As the session continues, writing effort is periodically quantified and classified or reclassified for each product. At session termination, final values for each product are available to the LAM for updating task structure and progress, as described above in section 4.6 above.

5 SUPPORTING REFLECTIVE LEARNING

Above, we have outlined how our proposed models can leverage knowledge about the student, the domain, and the assignment with the goal of facilitating the student's comprehension and productivity. Because the LAM maintains information about target knowledge, knowledge state, information utilities, and task features, we also propose that this information be used to generate interface features, with the goal of assisting the learner in their broader learning task. Below, we propose six ways the framework might support learning.

Explanations. Information in the LAM may be used to provide explanations about why particular resources are recommended. For example, a student might start their assignment by querying "how does Bernoulli's principle explain lift?" The RM might present results about these concepts, but also include results about the foundational concepts of "pressure" and "velocity" if the LAM predicts that the learner has little prior knowledge. To make it clear why these these concepts are included, the system could provide short explanations such as "Why recommended: understanding pressure is a key part of understanding lift." The explanations could also help students understand how the document might be useful to their specific context. For example, each result on a SERP might display a short annotation such as "has clear diagrams," "describes starting points," and "recommended because you know calculus."

Learning maps. The DM provides a rich source of information that could be used to show the learner a concept map for a domain. This process could visualize the important concepts in the domain, how concepts are related, which concepts are prerequisites for others, and clusters of concepts that are frequently learned together. The map could help learners understand their learning trajectory and visualize their current status relative to learning goals. For example, the map could color-code relevant nodes in the LDM to indicate the system's current representation of the learner's knowledge in the domain.

Feedback and Guidance. Both the BM and DM contain information possibly useful as feedback to learners, including queries entered, results examined, and progress made on the assignment. For example, the BM might indicate that the learner is using documents with a high domain-specific reading level and provide feedback suggesting engagement with introductory texts. The system could provide suggestions on next steps (e.g., concepts, work products). These types of feedback and suggestions could be integrated into the dashboard and also displayed as direct feedback on SERPs.

Reflection dashboard - SRL. Dashboards can help users reflect on their plans for working on an assignment [11] and their search interactions [6]. Information from the LAM could be used to provide displays to help learners consider the resources they will use, track progress on assignments, and reflect on their learning. Specifically, it could provide evaluative information about engagement with resource types (e.g., reading articles, skimming websites), use of resources (e.g., number and types of resources engaged), and the level and types of engagement while writing (e.g., amount pasted in, concepts noted). Interactive elements could express the model's current representation of the learner's progress on an assignment, with the option of the learner adding explicit plans and progress reflections.

Reflection dashboard - learning. We also see opportunities for the system to leverage its models of the domain, assignment, and learner's behaviors to help the learner reflect on the concepts they've engaged, how well they understand those concepts, and how the concepts are related. The emphasis here is on information that facilitates reflection rather than questions focused on validating memory of knowledge. For example, "How confident are you that you understand lift?" encourages reflection and self-judgement of learning. Questions like "What are the key concepts in lift?" check memory for facts. Both types of questions have important roles but we argue the proposed framework is well-situated for reflection during the learning process. Prior research has explored methods for reflection and shown that presenting users with system-generated questions that facilitate reflection can have positive impacts for low-knowledge learners [51].

6 DISCUSSION AND CONCLUSION

Having contemplated our design at length, we are left with the question of how an operational system could be developed collaboratively. There is precedence for large "moonshot" projects, but success will more likely come from an incremental approach. This raises questions on development priorities, scale, and scope. Domain-related prototyping efforts such as for the CM, DM, and AM, can be constrained by starting with relatively limited and simple domains. Other efforts are dependent on collaboration in the learning sciences. For example, advances in behavior models for reading comprehension and writing output are likely to progress more rapidly when aligned closely with theory development in these areas. We envision a tight coupling between domain-related models and behavior models for the inference of knowledge gain. The approximation and representation of knowledge gain is thus dependent on progress in these two general areas. In our design, a sufficiently accurate model of knowledge gain is necessary for modeling learning progress, and subsequent refinement of all the models to optimize for learning outcomes. The payoff, of course, is the opportunity for longitudinal study of learning outcomes at scale, where the effects of multiple factors may be observed. We argue that SAL research can be advanced by alignment of research questions with this set of goals.

In this paper, we have argued for a view of SAL that embeds searching within a larger view of the learning process. Our goal is to motivate research questions that center on facilitating comprehension for undergraduate students undertaking assignment work-tasks known to the system. The proposed framework provides a rough approximation for the types of functionality that might enable such an approach. We have presented examples of possible techniques and data sources for each model, and identified areas where key research is needed to produce effective models. We have introduced the novel concept of leveraging school assignments as rich task context, as well as a novel framework for constructing an end-to-end system for SAL.

REFERENCES

[1] Mustafa Abualsaud. 2017. *Learning Factors and Determining Document-level Satisfaction In Search-as-Learning*. Master's thesis. University of Waterloo, Waterloo, Ontario, Canada.

[2] Robin Alexander. 2008. Defining Pedagogy. In *Pedagogy and practice: culture and identities*, Patricia Murphy, Kathy Hall, and Janet Soler (Eds.). Sage, 28–39.

[3] Garrett Allen, Brody Downs, Aprajita Shukla, Casey Kennington, Jerry Alan Fails, Katherine Landau Wright, and Maria Soledad Pera. 2021. BiGBERT: Classifying Educational Web Resources for K-12 Grades. In *European Conference on Information Retrieval*. Springer, 176–184.

[4] Lorin W. Anderson, David R. Krathwohl, Peter W. Airasian, Kathleen A. Cruikshank, Richard E. Mayer, Paul R. Pintrich, James Raths, and Merlin C. Wittrock. 2000. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition* (1 edition ed.). Pearson, New York.

[5] Ella August, Karen Burke, Cathy Fleischer, and James A. Trostle. 2019. Writing Assignments in Epidemiology Courses: How Many and How Good? *Public Health Reports* 134, 4 (July 2019), 441–446. Publisher: SAGE Publications.

[6] Scott Bateman, Jaime Teevan, and Ryen W. White. 2012. The search dashboard: how reflection and comparison impact search behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Austin Texas USA, 1785–1794. https://doi.org/10.1145/2207676.2208311

[7] Nilavra Bhattacharya and Jacek Gwizdka. 2019. Measuring Learning During Search: Differences in Interactions, Eye-Gaze, and Semantic Similarity to Expert Knowledge. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. ACM, Glasgow Scotland UK, 63–71.

[8] M. Anne Britt and Jean-François Rouet. 2012. Learning with Multiple Documents. In *Enhancing the Quality of Learning*, John R. Kirby and Michael J. Lawson (Eds.). Cambridge University Press, Cambridge, 276–314. https://doi.org/10.1017/CBO9781139048224.017

[9] Arthur Câmara, Nirmal Roy, David Maxwell, and Claudia Hauff. 2021. Searching to Learn with Instructional Scaffolding. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval (CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 209–218.

[10] Simon Cassidy. 2012. Exploring individual differences as determining factors in student academic achievement in higher education. *Studies in Higher Education*

37, 7 (2012), 793–810. Publisher: Taylor & Francis.

[11] Patricia Chen, Omar Chavez, Desmond C Ong, and Brenda Gunderson. 2017. Strategic resource use for learning: A self-administered intervention that guides self-reflection on effective resource use enhances academic performance. *Psychological Science* 28, 6 (2017), 774–785. Publisher: Sage Publications Sage CA: Los Angeles, CA.

[12] Kevyn Collins-Thompson, Soo Young Rieh, Carl C. Haynes, and Rohail Syed. 2016. Assessing Learning Outcomes in Web Search: A Comparison of Tasks and Query Strategies. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (CHIIR '16)*. ACM, New York, NY, USA, 163–172.

[13] Carsten Eickhoff, Jacek Gwizdka, Claudia Hauff, and Jiyin He. 2017. Introduction to the special issue on search as learning. *Information Retrieval Journal* 20, 5 (2017), 399–402. Publisher: Springer.

[14] Carsten Eickhoff, Jaime Teevan, Ryen White, and Susan Dumais. 2014. Lessons from the journey: a query log analysis of within-session learning. In *Proceedings of the 7th ACM international conference on Web search and data mining*. ACM, 223–232.

[15] Jack M Fletcher. 2006. Measuring reading comprehension. *Scientific studies of reading* 10, 3 (2006), 323–330.

[16] Luanne Freund, Rick Kopak, and Heather O'Brien. 2016. The effects of textual environment on reading comprehension: Implications for searching as learning. *Journal of Information Science* 42, 1 (2016), 79–93. Publisher: SAGE Publications.

[17] Ujwal Gadiraju, Ran Yu, Stefan Dietze, and Peter Holtz. 2018. Analyzing Knowledge Gain of Users in Informational Search Sessions on the Web. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18)*. ACM, New York, NY, USA, 2–11.

[18] Souvick Ghosh, Manasa Rath, and Chirag Shah. 2018. Searching As Learning: Exploring Search Behavior and Learning Outcomes in Learning-related Tasks. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18)*. ACM, New York, NY, USA, 22–31.

[19] Roger Graves, Theresa Hyland, and Boba M. Samuels. 2010. Undergraduate Writing Assignments: An Analysis of Syllabi at One Canadian College. *Written Communication* 27, 3 (2010), 293–317. Publisher: SAGE Publications.

[20] K. Green, F.e. Tønnessen, K. Tambs, M. Thoresen, and E. Bjertness. 2009. Dyslexia: Group Screening among 15-16-Year-Olds in Oslo, Norway. *Scandinavian Journal of Educational Research* 53, 3 (June 2009), 217–227. Publisher: Routledge.

[21] Matthias Hagen, Martin Potthast, Michael Völske, Jakob Gomoll, and Benno Stein. 2016. How writers search: Analyzing the search and writing logs of non-fictional essays. In *Proceedings of the 2016 ACM on conference on human information interaction and retrieval*. 193–202.

[22] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2005. Accurately Interpreting Clickthrough Data as Implicit Feedback. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05)*. Association for Computing Machinery, New York, NY, USA, 154–161.

[23] Rishita Kalyani and Ujwal Gadiraju. 2019. Understanding User Search Behavior Across Varying Cognitive Levels. In *Proceedings of the 30th ACM Conference on Hypertext and Social Media (HT '19)*. ACM, New York, NY, USA, 123–132.

[24] Yvonne Kammerer, Rowan Nairn, Peter Pirolli, and Ed H. Chi. 2009. Signpost from the masses: learning effects in an exploratory social tag search browser. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, 625–634.

[25] Chen Liang, Zhaohui Wu, Wenyi Huang, and C. Lee Giles. 2015. Measuring Prerequisite Relations Among Concepts. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 1668–1674.

[26] Chang Liu and Xiaoxuan Song. 2018. How do Information Source Selection Strategies Influence Users' Learning Outcomes'. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18)*. ACM, New York, NY, USA, 257–260.

[27] Hanrui Liu, Chang Liu, and Nicholas J. Belkin. 2019. Investigation of users' knowledge change process in learning-related search tasks. *Proceedings of the Association for Information Science and Technology* 56 (2019), 166–175.

[28] Jingjing Liu, Nicholas J. Belkin, Xiangmin Zhang, and Xiaojun Yuan. 2013. Examining users' knowledge change in the task completion process. *Information Processing & Management* 49, 5 (2013), 1058–1074.

[29] Yihan Lu and I-Han Hsiao. 2017. Personalized Information Seeking Assistant (PiSA): from programming information seeking to learning. *Information Retrieval Journal* 20 (2017), 433–455.

[30] Diane Meziere, Lili Yu, Erik Reichle, Titus von der Malsburg, and Genevieve McArthur. 2021. Using Eye-Tracking Measures to Predict Reading Comprehension. (2021). Publisher: PsyArXiv.

[31] Heather L. O'Brien, Andrea Kampen, Amelia W. Cole, and Kathleen Brennan. 2020. The Role of Domain Knowledge in Search as Learning. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval (CHIIR '20)*. ACM, 313–317.

[32] Srishti Palani, Zijian Ding, Stephen MacNeil, and Steven P. Dow. 2021. The "Active Search" Hypothesis: How Search Strategies Relate to Creative Learning. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. ACM, New York, NY, USA, 325–329.

[33] Liangming Pan, Chengjiang Li, Juanzi Li, and Jie Tang. 2017. Prerequisite Relation Learning for Concepts in MOOCs. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, 1447–1456.

[34] Georg Pardi, Johannes von Hoyer, Peter Holtz, and Yvonne Kammerer. 2020. The Role of Cognitive Abilities and Time Spent on Texts and Videos in a Multimodal Searching as Learning Task. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. ACM, 378–382.

[35] Paul R. Pintrich, David A. F. Smith, Teresa Garcia, and Wilbert J. Mckeachie. 1993. Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (Mslq). *Educational and Psychological Measurement* 53, 3 (Sept. 1993), 801–813. Publisher: SAGE Publications.

[36] Martin Potthast, Matthias Hagen, Michael Völske, and Benno Stein. 2013. Crowdsourcing interaction logs to understand text reuse from the web. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1212–1221.

[37] Nicole C Ralston, Jacqueline M Waggoner, Beth Tarasawa, and Amy Jackson. 2016. Concurrent Validity of the Independent Reading Level Assessment Framework and a State Assessment. *Journal of At-Risk Issues* 19, 2 (2016), 8.

[38] Erik D Reichle, Simon P Liversedge, Denis Drieghe, Hazel I Blythe, Holly SSL Joseph, Sarah J White, and Keith Rayner. 2013. Using EZ Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review* 33, 2 (2013), 110–149. Publisher: Elsevier.

[39] Ridho Reinanda, Edgar Meij, and Maarten de Rijke. 2020. Knowledge Graphs: An Information Retrieval Perspective. *Foundations and Trends in Information Retrieval* 14 (2020), 289–444.

[40] Soo Young Rieh, Kevyn Collins-Thompson, Preben Hansen, and Hye-Jung Lee. 2016. Towards searching as a learning process: A review of current perspectives and future directions. *Journal of Information Science* 42, 1 (2016), 19–34.

[41] Nirmal Roy, Felipe Moraes, and Claudia Hauff. 2020. Exploring Users' Learning Gains within Search Sessions. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval (CHIIR '20)*. ACM, 432–436.

[42] Nirmal Roy, Manuel Valle Torre, Ujwal Gadiraju, David Maxwell, and Claudia Hauff. 2021. Note the Highlight: Incorporating Active Reading Tools in a Search as Learning Environment. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. Association for Computing Machinery, New York, NY, USA, 229–238.

[43] Sudeshna Roy, Meghana Madhyastha, Sheril Lawrence, and Vaibhav Rajan. 2019. Inferring Concept Prerequisite Relations from Online Educational Resources. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (2019), 9589–9594.

[44] Dale H. Schunk. 1981. Modeling and attributional effects on children's achievement: A self-efficacy analysis. *Journal of Educational Psychology* 73, 1 (1981), 93–105. Place: US Publisher: American Psychological Association.

[45] Dale H. Schunk. 1996. Self-Efficacy for Learning and Performance. *American Educational Research Association*.

[46] Traci Sitzmann and Katherine Ely. 2011. A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin* 137, 3 (2011), 421–442. Place: US Publisher: American Psychological Association.

[47] Catherine L Smith and Soo Young Rieh. 2020. The Other Side of the Same Coin: From Learning-centric Search Systems to Search-centric Learning Systems.. In *CIKM (Workshops)*.

[48] Margaret Snowling, Piers Dawes, Hannah Nash, and Charles Hulme. 2012. Validity of a Protocol for Adult Self-Report of Dyslexia and Related Difficulties. *Dyslexia* 18, 1 (2012), 1–15.

[49] Rohail Syed and Kevyn Collins-Thompson. 2017. Optimizing search results for human learning goals. *Information Retrieval Journal* 20, 5 (2017), 506–523.

[50] Rohail Syed and Kevyn Collins-Thompson. 2017. Retrieval Algorithms Optimized for Human Learning. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. ACM, 555–564.

[51] 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 *Proceedings of The Web Conference 2020 (WWW '20)*. Association for Computing Machinery, New York, NY, USA, 1693–1703.

[52] Pertti Vakkari, Michael Völske, Martin Potthast, Matthias Hagen, and Benno Stein. 2021. Predicting essay quality from search and writing behavior. *Journal of the Association for Information Science and Technology* 72, 7 (2021), 839–852. Publisher: Wiley Online Library.

[53] Teena Willoughby, S. Alexandria Anderson, Eileen Wood, Julie Mueller, and Craig Ross. 2009. Fast searching for information on the Internet to use in a learning context: The impact of domain knowledge. *Computers & Education* 52 (2009), 640–648.

[54] Max L. Wilson, Paul André, and mc schraefel. 2008. Backward highlighting: enhancing faceted search. In *Proceedings of the 21st annual ACM symposium on User interface software and technology (UIST '08)*. ACM, 235–238.

[55] Ulrika Wolff and Ingvar Lundberg. 2003. A technique for group screening of dyslexia among adults. *Annals of Dyslexia* 53, 1 (Jan. 2003), 324–339.

[56] Ran Yu, Ujwal Gadiraju, Peter Holtz, Markus Rokicki, Philipp Kemkes, and Stefan Dietze. 2018. Predicting User Knowledge Gain in Informational Search Sessions. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. ACM, 75–84.