Detecting, Understanding and Supporting Everyday Learning in Web Search

Ran Yu  
L3S Research Center  
Hannover, Germany  
yu@l3s.de

Ujwal Gadiraju  
L3S Research Center  
Hannover, Germany  
gadiraju@l3s.de

Stefan Dietze  
L3S Research Center  
Hannover, Germany  
dietze@l3s.de

ABSTRACT
Web search is among the most ubiquitous online activities, commonly used to acquire new knowledge and to satisfy learning-related objectives through informational search sessions. The importance of learning as an outcome of web search has been recognized widely, leading to a variety of research at the intersection of information retrieval, human computer interaction and learning-oriented sciences. Given the lack of explicit information, understanding of users and their learning needs has to be derived from their search behavior and resource interactions. In this paper, we introduce the involved research challenges and survey related work on the detection of learning needs, understanding of users, e.g. with respect to their knowledge state, learning tasks and learning progress throughout a search session as well as the actual consideration of learning needs throughout the retrieval and ranking process. In addition, we summarise our own research contributing to the aforementioned tasks and describe our research agenda in this context.

KEYWORDS
Search As Learning, User Modeling, Search Intent Detection, Learning in Web Search

ACM Reference Format:
Ran Yu, Ujwal Gadiraju, and Stefan Dietze. 2018. Detecting, Understanding and Supporting “Everyday Learning in Web Search.” In Proceedings of . ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION
Web search is among the most frequent online activities and has become a ubiquitous task. As is common search practice, a coherent search session, involving a particular search intent, usually involves several queries as well as one or more breaks in between (cf. [17]).

In particular, informational search sessions [7], i.e. sessions pertaining to the search for a particular piece of information expected to be available on the Web, are common and involve a particular learning intent, that is, the intent to acquire knowledge with respect to a certain topic.

Whereas platforms dedicated to online learning, such as MOOC environments, are tailored towards improving the learning performance and experience of online users, contemporary search engines have to satisfy a range of use cases, which may or may not involve learning. Transactional search sessions [7] are a common example of non-learning related online search. In contrast to actual learning-oriented environments in the online or offline sphere, where certain knowledge about the learning intent, the user as well as the learning task usually is available, such information is lacking in general online search settings. Consequently, heterogeneous features observable throughout a Web search session have to be utilised to derive insights about the learning intent, the user and the actual learning task.

Recently, a range of research works have approached this problem, often summarised under the ‘search as learning (SAL)’ umbrella and involving distinct disciplines such as information retrieval, human computer interaction or machine learning.

This paper attempts to provide an overview of a SAL research agenda by (i) summarising research challenges involved in this context, (ii) discussing related works in the area, (iii) presenting insights into early results of the authors’ own work as well as (iv) introducing gaps and future work in this area.

Figure 1 summarises the key emerging research challenges which at the same time define the structure for the remaining sections. Detecting Learning in Web Search (Section 2), refers to the process of distinguishing learning-related activities from other, non-learning, activities in general Web search scenarios. Understanding Users, Learning Tasks, Resources (Section 3) refers to the challenges involved in inferring information about a user, such as his/her knowledge state, the learning task, such as its complexity, or the involved resources from unstructured behavioral data observable throughout an online search session. Finally, Supporting Learning through Retrieval and Ranking (Section 4) refers to the actual consideration of inferred learning needs as part of the retrieval and ranking process or through adapting search interfaces to the user’s learning intent.

2 DETECTING LEARNING IN WEB SEARCH
Whereas only a certain amount of Web search sessions include a particular learning need, identifying such sessions becomes a prerequisite to facilitate further applications for understanding and supporting learning.

An established taxonomy from Broder [7] that has been widely used in the Web search context distinguishes between transactional, navigational and informational search sessions, where in particular the latter involve a learning goal, i.e. the intent to acquire knowledge about a particular topic. Specifically, transactional
search sessions usually aim at conducting a specific online trans-
action, such as, purchasing a ticket, navigational queries merely 
are aimed at leading the user to a dedicated website. In contrast, informational sessions imply the intent of a user to acquire some in-
formation assumed to be present on one or more web pages. In this context, the same query, for instance, Elbphilharmonie may be used 
to either buy tickets in a particular concert venue (transactional), to 
reach the Website https://www.elbphilharmonie.de (navigational) 
or to acquire knowledge about the Elbphilharmonie (informational).

We inspected a real-world query log, which consists of 913 search 
sessions, we found that 49.7% of them were informational search 
sessions with specific learning intent.

By adopting Broder’s taxonomy, the task of detecting learning 
in Web search can be formalised as identifying informational Web 
search sessions. Here a Web search session refers to the search activ-
ities within a certain time period that share the same information 
need. Previous work [17] on segmenting such search sessions have 
achieved promising performance. Here we do not go into details 
about the session segmentation task but focus on the automated 
detection of the intent of Web search.

2.1 Related Works

The classification of Web search queries has been explored in sev-
eral different scenarios, for instance, to classify a query into one 
of the categories [23, 24, 27, 29] or subcategories [21] of Broder’s 
taxonomy, or into other classes that are tailored towards specific 
applications [5, 19, 26].

Early studies on intent classification relied on manual approaches, 
for instance, by asking users through surveys [7] or by manual 
annotation of intents through judges [32]. However, while this 
process does not scale well to large datasets, automated classi-
fication approaches have been explored. Both supervised [5, 19, 
21, 23, 26, 27, 29] and unsupervised [5, 24] approaches have been 
applied on the classification of Web search queries. The features 
utilised in the aforementioned approaches are extracted from query 
terms [19, 21, 24, 26], user-click behaviors [26, 27, 29], anchor-
links [27, 29], Web documents’ content [21] and page views [24, 26].

The aforementioned works focus on the classification of single 
query sessions, often limited to data collected through lab studies. 
However, recent studies have shown that users information seeking 
tasks have grown more sophisticated [22] and often require one or 
more queries across multiple search sessions [1, 25, 28].

2.2 Detecting the Intent of Web Search Sessions 
from User Interaction Features

In contrast to such previous works, focused on query-based intent 
detection, our ongoing work to address this problem focuses on 
automatically detecting the intent of search activities at the level 
of search sessions.

Approach. We approach the problem of detecting informational 
Web search sessions with supervised models for classification. We 
extract 22 features according to multiple dimensions of a search 
session, structured into three categories, namely features related 
to Query (i.e. features related to number of query terms and the 
between query similarity), Session (i.e. total number of queries is-
sued, session duration related and session breaks related features) 
and Browsing behaviour (i.e. features related to number of clicks, 
revisited pages and similarity between query and the clicked URL).

For the classification model, we have experimented with several 
different approaches. Considering the scale of the data as well as the 
number and characteristics of the features, we have opted for Deci-
sion Tree (DT), Logistic Regression (LR), Support Vector Machine 
(SVM) [31] and Random Forest (RF) [6] as classification models. We 
tune the hyper parameters of each classifier through grid search.

Preliminary Result.

We apply our model to a dataset of real-world query logs, which 
contains 6860 queries from 124 users corresponding to 913 ses-
sions. Each session has been manually annotated by at least two 
annotators and assigned to one of the three classes. The annotated 
dataset is available online1. The results of using standard precision, 
recall and F1 score for each individual class, as well as their average

1 http://l3s.de/~yu/mission_classification
These can be classified into three main classes, namely, navigational, informational, and transactional. The overall classification performance indicates reasonable results. For instance, the task difficulty is subjective to user’s knowledge state, whereas the user’s knowledge state is a decisive factor on the resource selection.

We also analyzed the information gain of the selected features, and found that features in the browsing category appear more important than features in other categories, with 2 browsing features ranking at the top 2 positions. Query features are also shown to be effective with 3 features among the top 6. Session-based features have the least contribution among all 3 categories.

For simplicity, we use the term “session” in the remaining of this paper to refer to informational Web search sessions in particular, i.e. sessions which involve a particular learning intent.

2.3 Future Work
The overall classification performance indicates reasonable results on average, in particular transactional sessions appear ambiguous for both human annotators as well as supervised models. For this reason, results indicate that more specific classification tasks are likely to yield superior performance. For instance, an application-specific classifier aimed at targeted advertising may focus on only transactional or informational sessions (depending on the advertised offering), so that binary classification can be applied through a more tailored model. Further more, we found limitations arise in particular from the nature of the experimental dataset and the lack of publicly available, up-to-date query logs, future work will be concerned in particular with the application of similar approaches on a more recent and larger scale dataset. This would enable supervised models which are better reflecting contemporary search behavior and at the same time, utilise a wider variety of features.

3 UNDERSTANDING THE LEARNING PROCESS AND OUTCOMES
Since a sound understanding of learning throughout the search process is required in order to support learning, in this section we summarize the existing efforts in relevant topics and introduce our ongoing works on this task.

There are several factors that potentially affect the learning performance and the required support throughout the search session. These can be classified into three main classes, namely, user, e.g. the initial knowledge state and behavioral pattern, learning task, such as the task difficulty and novelty, and resource, e.g. the complexity or relevance of a resource. These factors are strongly inter-dependent. For instance, the task difficulty is subjective to user’s knowledge state, whereas the user’s knowledge state is a decisive factor on the resource selection.

Given the sparsity and heterogeneity of data throughout a search session, the works discussed in this section aim at inferring the aforementioned notions by considering a range of features observable throughout a search session.

3.1 Related Works
Previous works assessed the relation between learning and user search behavior from several different perspectives. Eickhoff et al. [11] investigated the correlation between features extracted from search session as well as search engine result page (SERP) documents with learning needs related to either procedural or declarative knowledge. The influence of distinct query types on knowledge gain was studied by Collins-Thompson et al. [9], finding that intrinsically diverse queries lead to increased knowledge gain. Hagen et al. [18] investigated the relation between the writing behavior and the exploratory search pattern of writers and revealed that query terms can be learned while searching and reading. Vakkari [34] provided a structured survey of features indicating learning needs as well as user knowledge and knowledge gain throughout the search process. Zhuang et al. [39] investigated the possibility of using 37 user search behavioral features to predict the user engagement, which correlates with learning, with supervised classifiers. By matching the learning tasks into different learning stages of Anderson and Krathwohl’s taxonomy [2], Jansen et al. studied the correlation between search behaviors of 72 participants and their learning stage [20]. Gwizdka et al. [15] proposed to assess learning outcomes in search environments by correlating individual search behaviors with corresponding eye-tracking measures. White et al. [35] investigated the difference between the behavior of domain experts and non-experts in seeking information on the same topic. By analyzing the activity log of experts and non-experts across different domains, the authors found that the distribution of features such as number of queries and query length differed across the levels of expertise. Zhang et al. [37, 38] explored using search behavior as an indicator for the domain knowledge of a user based on data acquired through a lab study (n = 35). Further, Cole et al. [8], observed that behavioral patterns provide reliable indicators about the domain knowledge of a user, even if the actual content or topics of queries and documents are disregarded entirely. Gwizdka and Spence [16] have shown that a searcher’s perception of task difficulty is a subjective factor that depends on the domain knowledge and some other individual traits. Arguello [3] proposed to use logistic regression to predict task difficulty in a search environment using behavioral features.

| Method | Navigational | Informational | Transactional | Weighted average | All Accu |
|--------|--------------|---------------|---------------|------------------|----------|
|        | P  R  F1    | P  R  F1      | P  R  F1      | P  R  F1         | P  R  F1 |
| DT     | 0.764 0.731 0.747 | 0.644 0.839 0.728 | 0.241 0.076 0.116 | 0.599 0.653 0.611 | 0.653    |
| SVM    | 0.786 0.760 0.773 | 0.656 0.927 0.768 | 0.800 0.022 0.042 | 0.724 0.694 0.623 | 0.694    |
| LR     | **0.809** 0.709 0.756 | 0.651 **0.938** 0.769 | 0.556 0.054 0.099 | 0.680 0.691 0.630 | 0.691    |
| RF     | 0.782 0.731 0.756 | 0.648 0.923 0.761 | 0.556 0.027 0.052 | 0.670 0.685 0.617 | 0.685    |

Table 1: Performance of different classifiers.
The aforementioned prior works have either studied a limited set of features or have addressed only specific learning scenarios and learning types. In particular, the generalizability of knowledge gain measures in previous works has not been investigated.

### 3.2 Analyzing Knowledge Gain of Users in Informational Search Sessions on the Web

We extend the current understanding of user knowledge gain in informational search sessions. Using real-world information needs and search sessions on the Web, we investigate the possibility of using search activity related features to predict knowledge gain (Section 3.3).

In particular, our recent work [14] investigated the impact of information needs on the search behavior and knowledge gain of users. To further the current understanding of the impact of informational search on a user’s knowledge, we recruited 500 distinct users from a crowdsourcing platform and orchestrated search sessions spanning 10 different information needs. We followed the recommended guidelines for effective crowdsourcing [12, 13]. By employing scientifically formulated knowledge tests to calibrate a user’s knowledge before a search session, and assess it after the session, we were able to quantify knowledge gain. The collected data has been released for the purpose of supporting research in the field.

Our investigation revealed a significant effect of information need on user queries and navigational patterns, but no direct effect on the knowledge gain. Users on average exhibited a higher knowledge gain through search sessions pertaining to topics they were less familiar with. For more details and findings please refer to the original paper [14].

### 3.3 Towards Predicting User Knowledge Gain in Informational Search Sessions

Based on the advanced understanding of the relation between user knowledge and their search behavior, we investigate the possibility of using search activity related features to predict knowledge gain and the knowledge state of a user – avoiding the need for explicit post-search knowledge assessments [36].

In this work [36], we aim at classifying the knowledge state (gain) of a user at the end of a given search session. For the sake of this work, a user’s knowledge state with respect to a particular information need is defined as the predicted user capability (accuracy) to correctly respond to a set of test questions about the respective information need. We classify the user knowledge state into 3 classes according to the user capability: low knowledge state, moderate knowledge state and high knowledge state. We hence define the user’s knowledge gain as the amount of knowledge state change, and consequently classify the knowledge gain into 3 classes: low knowledge gain, moderate knowledge gain and high knowledge gain.

**Approach.** We approach the problem with supervised models for classification. To this end, each session is represented by a feature vector, consisting out of 79 features related to: i) query (e.g. number of query terms, query complexity), ii) SERP (e.g. number of clicks, click-through ratio), iii) browsing behavior (e.g. number of pages viewed, average time stay per page), and iv) mouse movement (e.g. total scroll distance, number of mouseovers). We applied several feature selection techniques on the considered set of feature, and a range of standard models for the classification tasks, namely, Naive Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), Random Forrest (RF) and Multilayer Perceptron (MP).

**Preliminary Result.** Using the search activity log and the knowledge test data we collected through crowdsourcing (see Section 3.2), we trained and evaluated our classification models. The experimental results underline that a user’s knowledge gain and knowledge state can be modeled based on a user’s online interactions observable throughout the search process. Through feature analysis, we provide evidence for an improved understanding between individual user behavior and the corresponding knowledge state and change.

### 3.4 Future Work

As part of future work, we aim to reproduce and refine the findings in more varied search sessions, where durations and learning intents are more diverse, also involving considerably longer/shorter search sessions and, for instance, procedural knowledge rather than intents focused on declarative knowledge only. This would provide the opportunity to observe evolution-oriented features, for instance, considering the evolution of queries, their length and complexity. In addition, in crowd-based quasi experiments understanding of the actual users is very limited and data collected as such is expected to exhibit a certain amount of noise. For these reasons, we aim at conducting equivalent experiments in more controlled lab environments, where reliable information about both user interactions as well as the actual users can be obtained. Furthermore, whereas our previous work has focused on user interaction features, ongoing research investigates resource-centric features which take into account the characteristics of resources involved within the user interactions.

Potential applications for this work include the consideration of user knowledge and its expected learning progress as part of Web search engines and information retrieval approaches, or within informal learning-oriented search settings, such as libraries or knowledge- and resource-centric online platforms.

### 4 SUPPORTING LEARNING THROUGHOUT WEB SEARCH

The application-oriented objective are concerned with eventually supporting users in their learning tasks through i) optimizing user interaction and interfaces, and ii) enhancing the retrieval and ranking process. In this section, we review the status of existing techniques, and discuss the potential future directions.

#### 4.1 Related Works

**User interface and interaction.** Learning oriented online platforms (e.g. coursera\(^3\), mooc\(^4\), Didactalia\(^5\)) have been constantly

---

\(^3\)https://www.coursera.org/
\(^4\)http://mooc.org/
\(^5\)https://didactalia.net
optimized to improve the learning performance of users. Examples are, for instance, the use of learning dashboards to inform users about his/her learning progress or provide discussion forums to enable collaboration among users. However, within general-purpose search engines, there is a lack of attention for the support of learning, also due to the general-purpose nature of such environments and the variety of tasks conducted there. A central question for research in that area is whether and how interfaces can be adopted to improve learning performance even in such general-purpose environments. An attempt has been made by Arora et al. [4], by aiming at improving user engagement in learning oriented search tasks through providing richer representation of retrieved Web documents. Specifically, they explored methods of finding useful semantic concepts within retrieved documents, with the objective of creating improved document surrogates for presentation in the SERP.

Retrieval and ranking. As current search engines are optimized by considering an information need disregarding the learning intent behind a query, relatively little research has been carried out on optimising retrieval and ranking algorithms towards particular learning needs. For instance, Dave et al. [10] discussed the potential of two ranking models with varied objectives (i.e. paragraph retrieval model, dependency based re-ranking) on enhancing the performance of learning-centric search engines. Recently, Syed and Collins-Thompson [33] proposed to optimize the learning outcome of the vocabulary learning task by selecting a set of documents while considering keyword density and domain knowledge of the learner. Their theoretical framework provides a sound basis for furthering the study on learning-oriented retrieval techniques.

4.2 Future Work

User interface and interaction. Within general-purpose search engines, there is a lack of attention for the support of learning, also due to the general-purpose nature of such environments and the variety of tasks conducted there. A central question for research in that area is whether and how interfaces can be adopted to improve learning performance even in such general-purpose environments. Studies reveal that people engage more in many search tasks involving collaboration with others rather than while searching by themselves [30]. To further this investigation and develop tools to support learning by enabling collaboration between users, our ongoing work is concerned with developing a search interface that encourages experienced learners to guide learners who will use the system in the future and assess its impact on the knowledge gain throughout a search session. Suggestions are ranked according to the feedbacks from experienced learners. Throughout large-scale quasi-experiments and facilitated by pre- and post-tests, we aim to quantify the influence of the collaborative search interface on the learning outcome. Future work is concerned with alternative means to improve interfaces and interactions towards increasing the learning outcomes during Web search.

Retrieval and ranking. In Section 3.3, we have discussed means to infer a user’s knowledge state (gain) in online search sessions. On this basis, future work is aimed at optimizing ranking algorithms to recommend resources that fit not only the traditional notion of an information need, but also a user’s knowledge state. Whereas traditional ranking algorithms tend to suggest Web documents disregarding, for instance, a user’s reading level, improved ranking techniques will favor resources which are neither too easy nor too hard for a particular user’s need. This builds on the assumption that, based on the assessment of the relation between a Web resource and user’s knowledge gain, a ranking algorithm can recommend resources not only fitting into the user’s knowledge state, but also maximizing the user’s knowledge gain and learning efficiency.

5 CONCLUSIONS

This paper has provided an overview of challenges and research approaches towards detecting, understanding and supporting learning throughout Web search. One crucial challenge in this context is the lack of explicit information about users, their learning intent, task or progress throughout an online search session, requiring the utilisation of a wide variety of informal features to derive such information. In particular, supervised machine learning techniques and extensive feature analysis have been deployed as part of previous work, yet works are usually focused on specific learning scenarios, isolated feature sets or single-query scenarios, rather than entire search sessions.

Another major obstacle is the lack of large-scale datasets to facilitate SAL research by providing both diverse features of user interactions and behavior as well as high-quality ground truth data about the involved users, their knowledge state and knowledge gain throughout the captured search sessions.

In addition to summarising research challenges and related works, we have introduced some of our own contributions to the respective tasks. These consist of supervised approaches for detecting learning-related (informational) search sessions, for predicting the knowledge state and gain of online users and the preliminary analysis of experimentally obtained search sessions and the correlation of observed variables with user knowledge. Ongoing and future work will expand on these works, consider more varied feature sets, in particular resource-centric features, and will in particular be concerned with obtaining, providing and analysing search session data collected in more controlled lab environments.

REFERENCES

[1] Eugene Agichtein, Ryen W. White, Susan T. Dumais, and Paul N. Bennett. 2012. Search, Interrupted: Understanding and Predicting Search Task Continuation. In Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ’12). ACM, New York, NY, USA, 315–324.

[2] Lorin W Anderson, David R Krathwohl, P Airasian, K Cruikshank, R Mayer, P Pintrich, J Raths, and M Wittrock. 2001. A taxonomy for learning, teaching and assessing: A revision of Bloom’s taxonomy. New York, Longman Publishing.

[3] Jaime Asguello. 2014. Predicting Search Task Difficulty. In ECIR Vol. 14. 88–99.

[4] Piyush Arora. 2015. Promoting User Engagement and Learning in Amorphous Search Tasks. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 1051–1053.

[5] Ricardo Baes-Cuytes, Liliana Calderón-Benívides, and Cristina González-Caro. 2006. The Intention Behind Web Queries. In Proceedings of the 13th International Conference on String Processing and Information Retrieval (SPIRE’06). Springer-Verlag, Berlin, Heidelberg, 98–109.

[6] Leo Breiman. 2001. Random Forests. Mach. Learn. 45, 1 (Oct. 2001), 5–32.

[7] Andrey Broder. 2002. A Taxonomy of Web Search. SIGIR Forum 36, 2 (9 2002), 3–10. Issue Fall 2002.

[8] Michael J Cole, Jacek Owidzka, Chang Liu, Nicholas J Belkin, and Xiangmin Zhang. 2013. Inferring user knowledge level from eye movement patterns. Information
