SaL-Lightning Dataset: Search and Eye Gaze Behavior, Resource Interactions and Knowledge Gain during Web Search

Christian Otto
TIB – Leibniz Information Centre for Science and Technology
Hannover, Germany
christian.otto@tib.eu

Markus Rokicki
L3S Research Center, Leibniz University Hannover
Hannover, Germany
rokicki@l3s.de

Georg Pardi
IWM – Leibniz-Institut für Wissensmedien
Tübingen, Germany
g.pardi@iwm-tuebingen.de

Wolfgang Gritz
TIB – Leibniz Information Centre for Science and Technology
Hannover, Germany
wolfgang.gritz@tib.eu

Daniel Hienert
GESIS – Leibniz Institute for the Social Sciences
Cologne, Germany
daniel.hienert@gesis.org

Ran Yu
Data Science & Intelligent Systems (DSIS), University of Bonn
Bonn, Germany
ran.yu@uni-bonn.de

Johannes von Hoyer
IWM – Leibniz-Institut für Wissensmedien
Tübingen, Germany
j.hoyer@iwm-tuebingen.de

Anett Hoppe[*]
TIB – Leibniz Information Centre for Science and Technology
Hannover, Germany
anett.hoppe@tib.eu

Stefan Dietze[†]
GESIS – Leibniz Institute for the Social Sciences
Cologne, Germany
stefan.dietze@gesis.org

Peter Holtz
IWM – Leibniz-Institut für Wissensmedien
Tübingen, Germany
p.holtz@iwm-tuebingen.de

Yvonne Kammerer[‡]
IWM – Leibniz-Institut für Wissensmedien
Tübingen, Germany
y.kammerer@iwm-tuebingen.de

Ralph Ewerth[§]
TIB – Leibniz Information Centre for Science and Technology
Hannover, Germany
ralph.ewerth@tib.eu

ABSTRACT

The emerging research field Search as Learning (SAL) investigates how the Web facilitates learning through modern information retrieval systems. SAL research requires significant amounts of data that capture both search behavior of users and their acquired knowledge in order to obtain conclusive insights or train supervised machine learning models. However, the creation of such datasets is costly and requires interdisciplinary efforts in order to design studies and capture a wide range of features. In this paper, we address this issue and introduce an extensive dataset based on a user study, in which 114 participants were asked to learn about the formation of lightning and thunder. Participants' knowledge states were measured before and after Web search through multiple-choice questionnaires and essay-based free recall tasks. To enable future research in SAL-related tasks we recorded a plethora of features and person-related attributes. Besides the screen recordings, visited Web pages, and detailed browsing histories, a large number of behavioral features and resource features were monitored. We underline the usefulness of the dataset by describing three, already published, use cases.

KEYWORDS

Web Learning, Knowledge Gain, User Study

1 INTRODUCTION

Search as learning (SAL) is an interdisciplinary field that combines the insights of (multimedia) information retrieval, human-computer interaction, and cognitive psychology. Its objective is to understand and support individual learning processes during Web search. Research questions concern

(1) detection and prediction of learning processes taking place,
(2) assessment of resource suitability for learning in general and for specific users,
(3) adapted retrieval and ranking for optimized learning efficiency, and
(4) classification of users and user knowledge from in-session behavior.

Research on this kind of questions relies on study-based data that has to capture (a) search behavior of various nature and (b) knowledge metrics of users (through pre- and posttests). As they have to be conducted in controlled environments, their design and execution is costly. They have to reflect realistic scenarios in order to be indicative for real-life applications: A restriction of web pages or intrusive recording equipment Second, assembling meaningful questionnaires for pre- and post-tests, for instance, is a challenging task, as topic domain and item difficulty have to be well calibrated. Lastly, the logging process itself is non-trivial since

available software, to our knowledge, usually only covers part of the features of interest. Therefore handcrafted, non-intrusive logging mechanisms need to be implemented manually.

To the best of our knowledge, there are currently only two SAL-focused datasets available: Proaño-Ríos and González-Ibáñez[20] provide a set of 83 expert-generated learning paths on a diversity of topics. Each expert assembles a set of three web resources useful towards a certain learning goal, including a justification of their choice. However, data on real-world user behavior is not included. Gadijaru et al. [8] present a dataset comprised of 420 crowdsourced learning sessions, investigating the information needs on the search behavior and knowledge gain of users. Our dataset improves their contribution by presenting data gathered in a controlled lab study and it captures behavioral, resource, and gaze data. Other resources focus on either search or learning: (1) *Search Focus* – Datasets for the conception and optimization of search systems provide the basis for improved automatic analysis of queries [10, 11], identification of user tasks [12, 26], the influence of found resources on the user's viewpoints [1], and novel interaction methods such as conversational search [19]; (2) *Learning focus* – Datasets from the educational domain often explore recommendation tasks [25], provide data on user behavior in restricted learning environments [9] or specific instructional practices [28]. Finally, there is an active area of research on predicting the memorability of visual resources [4, 14] and the impact of resource modality on learning success [13]. The scope of these datasets is usually limited to a single feature type; none of them collects user behavior information in a realistic and open, learning-related Web search scenario.

In this paper, we contribute to SAL and related areas by providing a unique dataset to overcome some of the aforementioned limitations. Our SaL-Lightning dataset contains detailed recordings, pre- and post-knowledge assessments of 114 participants, interaction data on real-world search behavior, as well as resource features of a user study. This data diversity has the potential to help researchers answer diverse questions tied to the entire online learning framework, from individual psychological aspects, over usability tests and data visualization over retrieval and ranking issues existing in the technology that enables this process. We prove this claim by presenting three already published works.

The remainder of the paper is structured as follows. First, we go into great detail regarding the data acquisition process before laying out how we curated the gathered information (Section 2). Afterward, we showcase already existing use cases (Section 3). Section 4 concludes with a summary and possible future work.

2 DATASET DESCRIPTION

This section gives an in-depth explanation of the dataset acquisition process:

(1) the initial user study in Section 2.1,
(2) the technical environment in Section 2.2 and
(3) the detailed description of the different data subsets in Section 2.3.

2.1 User Study

2.1.1 Participants and Task. The participants (N=114), German speaking university students (95 female, μ_{age}=22.88, σ_{age}=2.93)

from different majors were asked to solve a realistic learning task to understand the principles of thunderstorms and lightning. This topic has been used before to study multimedia learning (e.g., [16, 22]) and has been chosen since it requires learners to gain knowledge about different physical and meteorological concepts and their interplay, i.e., they need to learn about causal chains of events and acquire declarative as well as procedural knowledge [2].The acquisition of information about such task a can be accomplished through studying different representation formats, such as text, pictures, videos, or combinations of those. This circumstance is beneficial for our goal to get a general idea about optimal multimedia learning resource design, especially in SAL scenarios.

2.1.2 Procedure and Measurements. The experiment consisted of an online and a laboratory part. In the online part, which had to be completed around one week before the lab appointment, participants had to respond for the first time to the 10-item multiple-choice and 4-item transfer knowledge test based on previous work [22]. Further, participants worked on questionnaires regarding their achievement motivation [6] and their Web-specific epistemic justification beliefs [3]. At the lab appointment, participants first completed tests assessing their reading comprehension [23] and working memory capacity [5]. The participants were asked to write a first essay (t1) about the topic of the formation of thunderstorms and lightning. Afterward, they were instructed to learn about this specific topic by searching the Web in a self-regulated manner. Participants were informed about the time limit of max. 30 minutes for their web search, and that they could also end the task early. They were encouraged to use every kind of Web resource they would like. After the learning phase, they were asked to write again everything what they now knew about the topic in a free essay (t2) format. Lastly, the participants were asked to answer the multiple-choice questionnaires (t2) again followed by a questionnaire assessing task engagement [15] and cognitive reflection tasks [7].

2.2 Technical environment

All search and learning activities of participants were conducted within a tracking framework consisting of two layers. The first layer was the SMI (SensoMotoric Instruments) ExperimentCenter (3.7) software environment that enabled us to track participants' eye movements as well as their activities during Web search in the formats of screen recordings and navigation log files. For the second layer we utilized browser plugins to gather resources of all visited HTML files and adapted Talibi et al's method [24] to track navigation and interaction data (e.g., mouse movements). For more details we refer to [17].

2.3 Dataset Structure

The following section describes the information per user provided by the individual data subsets. Apart from the screen recordings and HTML data, which we cannot make publicly available due to licensing restriction, all dataset parts are available under https://doi.org/10.25835/0062363.

2.3.1 Resource Data - Screen Recordings. The screen recordings show the entire search process of the participants over the duration of the study and are aligned with the provided logs (Section 2.3.3)

and HTML data (Section 2.3.2). The screen recording's video format is MP4, and they have been recorded with a resolution of $1920x1080$ at 30 frames per second. The audio track is not included. They are not longer than 30 minutes and start at the point in time the learning session starts. We manually cut the start of the video that showed the participants entering their session IDs. Since we are not able to share the screenrecordings, we provide the multimedia features generated in [17], specifically statistics about the document layout as well as the results of the image type classification for all seen websites.

2.3.2 Resource Data - HTML. Since online content is not persistent, to achieve our goal to enable research on the actual data seen by the participants, we decided to record the content of each visited website, including but not limited to *html, *css, *js, and image files. Due to technical difficulties this process was not entirely successful, forcing us to fill in the gaps at later points in time. In detail, we managed to capture 87.9% of the data at the time of the study, another 4% in March 2020 and finally, another 2.5% in September 2021. For the remaining 5.6% (181 URLs) we were not able to record any data. With very few exceptions (a few websites that are not available anymore) these were search engine result pages from Google and YouTube that do not contain any learning relevant information, and when crawled at a later point in time, can differ strongly from the original. For these two reasons we decided to exclude them from the dataset. For full transparency we disclosed the date of acquisition in the provided timeline.

2.3.3 Behavioral Data - Browsing Timeline. Each participant's browsing log is represented by one TSV (tab separated value) file as outlined by Table 1, chronologically displaying the visited websites with a timestamp in seconds passed since the start of the session. Additionally, we disclose the path to the respective HTML files and their acquisition date as mentioned in Section 2.3.2.

p_id	timestamp	url	html_files	date_of_acquisition

Table 1: The fields (columns) in the timeline file; each web resource is associated with a directory of HTML files and acquisition date.

2.3.4 Behavioral Data - Gaze. As mentioned in Section 2.2, we used an eye-tracker to record the learner's eye movements over the course of the Web search session. We exported the gaze information from the eye-tracking software as raw data and separated the fixations and saccades with an I-DT algorithm [21] by marking the entries via the $fixation \in \{0, 1\}$ flag. Please note that the y-coordinates are relative to the entire website and not the viewport, i.e., values larger than 1080 are possible. Further, the data can contain incorrect entries originating from tracking errors (e.g., negative values). Similar to the data subsets introduced in the sections before, this data subset contains one TSV file for each participant, chronologically listing the coordinates of the left and right pupil with millisecond precision. Additionally, the URL visible at that point in time is displayed (Table 2). This data subset allows for further experiments on website examination behavior and their influence on the learning process.

p_id	timestamp	left_x	left_y	right_x	right_y	url	fixation

Table 2: The fields (columns) in the gaze data files, chronologically displaying the gaze coordinates for each eye.

2.3.5 Behavioral Data - Browsing Events. The investigation of the Web search behavior requires detailed logs about the learners' interactions with a website, going beyond logging what type of resources they visited. We recorded over 1 million user interaction events of 11 different types. The *focus*, *blur*, and *beforeunload* describe whether a website has come into focus, lost focus, or is about to be closed. The *resize* event tracks if a participant chose to resize the current browser window and captures the resulting window size in the value column encoded as pixel sizes $x|y$. Similarly, if a learner scrolled on a website, a *scroll* event is triggered and we log the scroll distance in vertical and horizontal direction in the value column as $vertical|horizontal$. The *mousemove* event tracks the learners' mouse movements by logging x and y coordinates in the respective columns. Mouse clicks were captured in the *click* event, tracking their location (x and y columns) and the clicked HTML element in the target column as XPath. We recorded keypresses in the *keypressed* event, but omitted recording the key values due to privacy reasons in case the learner chose to login somewhere during the session. However, the chosen queries are available in the URLs of the respective search engines. Lastly, we captured *copy* and *paste* events and, if available, the intended target elements. The structure of the data record per event is displayed in Table 3.

p_id	timestamp	track_id	type	value	x	y	target	url

Table 3: The columns in the event data files for each participant chronologically displaying the browsing interaction events.

2.3.6 Behavioral Data - Browsing Tracks. The browser tracking tool associates events to websites by means of *tracks*. Upon navigating to a website, a track is created and exists until the user navigates somewhere else within the same browser tab or closes it. This setup is geared towards realistic search sessions with multiple concurrent tabs. For each track, our dataset contains the time of creating the track, URL, and title of the website, as well as the viewport dimensions. Additionally, the data contains the lifetime of the track, as well as the time the track was active, i.e., it was displayed to the user in the active browser tab.

p_id	timestamp	track_id	url	title
viewport width	viewport height	time stay	time active	

Table 4: The columns in the track data files, capturing information such as URL and active time for a visited website.

2.3.7 Knowledge Data and Questionnaires. As mentioned in 2.1.2, we measured the knowledge state of learners at multiple points in time. Additionally, through questionnaires and tests, we captured cognitive abilities and assessments of participants across the study. Thereby, several sub datasets were generated for which we provide the documentations with explanations of measured variables and, if possible, the original German items. This section will give brief

explanations of these files, while detailed documentation can be found in the dataset.

demo_knowledge_sum.csv: This file contains demographic information of participants and the summary of the knowledge-related scores (multiple-choice, essay) and cognitive abilities (working memory capacity, reading comprehension, cognitive reflection) (Table 5). Reading comprehension was measured through a standardized German screening instrument for adolescents and young adults. Working memory capacity was measured through a reading span task. Detailed information is provided by Pardi et al. [18].

Feature	Description
p_id	Participant ID
d_sex	1 = female; 2 = male
d_age	Age of participant
d_field	Field of study
d_no_sem	Number of semesters
d_lang	Mother tongue
k_mc_sum_t1	# of correct mc questions before search
k_mc_sum_t2	# of correct mc questions after search
kg_mc	Knowledge gain multiple choice questions
essay_C1	# of correct concepts before search
essay_C2	# of correct concepts after search
kg_essay	Knowledge gain essays
LGVT_speed	# of words read
LGVT_core	Points for correctly solved sentences
WMC_Recalls	# of correctly recalled sets
WMC_Sentence	# of correctly solved sentences
CRT_sum	# of correctly solved cognitive reflection tasks

Table 5: The columns in the demo_knowledge_sum data files. One participant per row.

mc_data.csv: This file contains the scores for all multiple-choice questions answered by participants before the lab session (t1) and after the search in the lab (t2). Includes also the confidence rating of participants for each question and the information if the answer was guessed.

essay_data.csv: This file contains the raw essays written by the participants before (t1) and after the search (t2).

internet_specific_epistemic_justification.csv: This file contains the measured Web-specific epistemic justification based on a translated version of [3].

selfassessment_data.csv: To measure participant's self-assessed performance on the knowledge tests, we used both global self-assessment (estimated numbers of items answered correctly, estimated placement as compared to others and perceived ability to explain the concepts of the learning topic) as well as local on-item

confidence rating, indicating how confident participants were that their given answer was correct.

CRT_data.csv: To measure an individual's tendency for cognitive reflection, participants worked on five items of the cognitive reflection task (CRT [7]) translated into German. Within the dual-process model of reasoning, there is a distinction between faster responses with little deliberation and slower and more reflective responses. Solving more of the CRT items shows a higher disposition for the latter one, i.e., reflective cognition.

achievement_data.csv: We used the German version of the achievement motives scale [6] to measure hope of success (HS) and fear of failure (FF). This scale contains 10 items that are rated on a scale from 1 to 4, assessing those two dimensions. The achievement motive of an individual describes the general tendency to approach or avoid success in an evaluative situation. The HS score is associated with a range of variables beneficial for learning success, such as, performance in reasoning, persistence, or task enjoyment.

dssq_data.csv: The Dundee Stress State Questionnaire [15] measures subjective states in a performance context. Participants had to indicate their agreement to 7 items with labeled endpoints immediately after the learning phase. The mean score of those items indicates an individual's self-reported task engagement during the learning phase. Differences in task engagement can act as a moderator for task performance.

3 USE CASES

The following paragraphs describe three studies that were conducted based on the SaL-Lightning dataset.

Knowledge Prediction. One of the most sought after tasks in SAL is the prediction of the learning outcome given a set of learning resources and, optimally, considering the learner's individual needs. The prediction of the user's knowledge would allow retrieval algorithms to present more accurate search results given a set of resources for a respective topic. Otto et al. [17] investigated how the layout and content of the multimedia data on the websites influence, besides behavioral and resource features, the knowledge gain of the participants.

Further, they conducted correlation analyses of (multimodal) resource features and pre-/post-knowledge scores. These analyses provided (weak) indications whether certain resources or resource types (e.g., videos) are more frequently used by novices, intermediates, or experts.

Since our pre- and post-study questionnaires have measured the knowledge state both using multiple-choice questions and essays, the dataset can be used to advance the state of the art in this research area. For example, the relationship between gaze behavior and learning outcome is still rather unexplored. Another interesting aspect that has not yet been investigated is the evolution of the users' knowledge (gain) by analyzing their search queries over time.

Cognitive Abilities and Search Tasks. Pardi et al. [18] investigated the connection between behavioral data, namely the time spent on different information resources, and the learners' cognitive abilities and learning outcome. Therefore, they classified based on the URL and screen recordings the time participants spent on text-dominated websites (potentially accompanied by images) and the time spent

on online videos. Furthermore, they analyzed the assessed working memory capacity and reading comprehension of participants in connection to their learning outcome. The learning outcome was derived from the pre- and post essays users wrote before and after their web search. Based on these data, the correlations between learning outcome (pre, post), cognitive abilities (working memory, reading comprehension), and time spent on resources (search activities, on video, on text-dominated websites) were analyzed.

Metacognitive Judgments in SAL. Hoyer et al. [27] used the global and local confidence ratings to investigate how the evaluation of one´s own knowledge changes after a learning phase. The authors found that after the 30-minutes self-regulated online learning phase, learners are able to gain knowledge and are in general capable of accurately estimating their knowledge on measures of global confidence ratings. However, there was also an increase in overclaiming, which indicates that more knowledge is claimed after learning. Additionally, the authors report an unexpected false certainty effect indicated by increased local confidence ratings given to incorrectly answered questions after the learning phase. Since accurate local metacognitive judgments are essential for controlling learning processes, this result points to a possible detrimental effect of short online learning and should be explored further in future experiments.

4 CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a comprehensive, multilateral dataset for research in the field of Search as Learning. It enables interdisciplinary research in various disciplines related, but not limited to, multimodal information retrieval and learning psychology, and usability. The potential is underlined by three already published publications that made use of the SaL-Lightning dataset. Our plan for the future is to explore further interdisciplinary topics related to Search as Learning, in addition to the ones already mentioned. Besides the exploration of psychological phenomena based on the provided knowledge metrics, a variety of computer science applications are also unexplored, for instance, multimedia resource recommendation, web document layout analysis, analysis of user behavior during search, and query refinement.

ACKNOWLEDGMENTS

REFERENCES

[1] Khalid Al-Khatib, Michael Völske, Shahbaz Syed, Nikolay Kolyada, and Benno Stein. 2020. *Webis ChangeMyView Corpus 2020 (Webis-CMV-20).* https://doi.org/10.5281/zenodo.3778298

[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. *Educational Horizons* 9, 2 (2001), 137–175.

[3] Ivar Bråten, Christian Brandmo, and Yvonne Kammerer. 2019. A Validation Study of the Internet-Specific Epistemic Justification Inventory With Norwegian Preservice Teachers. *Journal of Educational Computing Research* 57, 4 (2019), 877–900. https://doi.org/10.1177/0735633118769438 arXiv:https://doi.org/10.1177/0735633118769438

[4] Romain Cohendet, Claire-Hélène Demarty, Ngoc Q. K. Duong, and Martin Engilberge. 2019. VideoMem: Constructing, Analyzing, Predicting Short-Term and Long-Term Video Memorability. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019.* IEEE, 2531–2540. https://doi.org/10.1109/ICCV.2019.00262

[5] Andrew R. A. Conway, Michael J. Kane, Michael F. Bunting, D. Zach Hambrick, Oliver Wilhelm, and Randall W. Engle. 2005. Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review* 12, 5 (2005), 769–786. https://doi.org/10.3758/BF03196772

[6] Stefan Engeser. 2005. Messung des expliziten Leistungsmotivs: Kurzform der Achievement Motives Scale. (2005). https://www.uni-trier.de/fileadmin/fb1/prof/PSY/PGA/bilder/Engeser_2005_Kurzform_der_AMS.pdf

[7] Shane Frederick. 2005. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives* 19, 4 (December 2005), 25–42. https://doi.org/10.1257/089533005775196732

[8] 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 and Retrieval, CHIIR 2018, New Brunswick, NJ, USA, March 11-15, 2018,* Chirag Shah, Nicholas J. Belkin, Katriina Byström, Jeff Huang, and Falk Scholer (Eds.). ACM, 2–11.

[9] Raphaël Grolimund. 2017. *Individual work of Bachelor students monitored using a dynamic assessment dashboard.* https://doi.org/10.5281/zenodo.290129

[10] Matthias Hagen, Martin Potthast, Benno Stein, Christof Bräutigam, and Anna Beyer. 2010. *Webis Query Segmentation Corpus 2010 (Webis-QSeC-10).* https://doi.org/10.5281/zenodo.3256198

[11] Matthias Hagen, Martin Potthast, Benno Stein, Marcel Gohsen, and Anja Rathgeber. 2017. *Webis Query Spelling Corpus 2017 (Webis-QSpell-17).* https://doi.org/10.5281/zenodo.3570912

[12] Matthias Hagen, Benno Stein, Jakob Gomoll, and Anna Beyer. 2013. *Webis Search Mission Corpus 2012 (Webis-SMC-12).* https://doi.org/10.5281/zenodo.3265962

[13] Avery Hlousek and Bentley Krause. 2020. *The impact of learning modality on team-based learning (TBL) outcomes in anatomical sciences education.* https://doi.org/10.5061/dryad.cfxpnvx35

[14] Phillip Isola, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. 2011. What makes an image memorable?. In *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011.* IEEE Computer Society, 145–152. https://doi.org/10.1109/CVPR.2011.5995721

[15] Gerald Matthews, Sian E Campbell, Shona Falconer, Lucy A Joyner, Jane Huggins, Kirby Gilliland, Rebecca Grier, and Joel S Warm. 2002. Fundamental dimensions of subjective state in performance settings: task engagement, distress, and worry. *Emotion (Washington, D.C.)* 2, 4 (December 2002), 315—340. https://doi.org/10.1037/1528-3542.2.4.315

[16] Richard E Mayer and Roxana Moreno. 1998. A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of educational psychology* 90, 2 (1998), 312.

[17] Christian Otto, Ran Yu, Georg Pardi, Johannes von Hoyer, Markus Rokicki, Anett Hoppe, Peter Holtz, Yvonne Kammerer, Stefan Dietze, and Ralph Ewerth. 2021. Predicting Knowledge Gain During Web Search Based on Multimedia Resource Consumption. In *Artificial Intelligence in Education - 22nd International Conference, AIED 2021, Utrecht, The Netherlands, June 14-18, 2021, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 12748),* Ido Roll, Danielle S. McNamara, Sergey A. Sosnovsky, Rose Luckin, and Vania Dimitrova (Eds.). Springer, 318–330. https://doi.org/10.1007/978-3-030-78292-4_26

[18] 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 *CHIIR '20: Conference on Human Information Interaction and Retrieval, Vancouver, BC, Canada, March 14-18, 2020,* Heather L. O'Brien, Luanne Freund, Ioannis Arapakis, Orland Hoeber, and Irene Lopatovska (Eds.). ACM, 378–382. https://doi.org/10.1145/3343413.3378001

[19] Gustavo Penha, Alexandru Balan, and Claudia Hauff. 2019. Introducing MANtIS: a novel Multi-Domain Information Seeking Dialogues Dataset. *CoRR* abs/1912.04639 (2019). arXiv:1912.04639 http://arxiv.org/abs/1912.04639

[20] Verónica Proaño-Ríos and Roberto I. González-Ibáñez. 2020. Dataset of Search Results Organized as Learning Paths Recommended by Experts to Support Search as Learning. *Data* 5, 4 (2020), 92. https://doi.org/10.3390/data5040092

[21] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Eye Tracking Research & Application Symposium, ETRA 2000, Palm Beach Gardens, Florida, USA, November 6-8, 2000,* Andrew T. Duchowski (Ed.). ACM, 71–78. https://doi.org/10.1145/355017.355028

[22] Florian Schmidt-Weigand and Katharina Scheiter. 2011. The role of spatial descriptions in learning from multimedia. *Comput. Hum. Behav.* 27, 1 (2011), 22–28.

[23] Wolfgang Schneider, Matthias Schlagmüller, and Marco Ennemoser. 2007. *LGVT 6-12: Lesegeschwindigkeits-und-verständnistest für die Klassen 6-12.* Hogrefe Göttingen.

[24] Davide Taibi, Francesca Bianchi, Philipp Kemkes, and Ivana Marenzi. 2018. Learning Analytics for Interpreting. In *Proceedings of the 10th International Conference on Computer Supported Education, CSEDU 2018, Funchal, Madeira, Portugal, March 15-17, 2018, Volume 1,* Bruce M. McLaren, Rob Reilly, Susan Zvacek, and James Onohuome Uhomoibhi (Eds.). SciTePress, 145–154. https://doi.org/10.5220/0006774801450154

[25] Katrien Verbert, Hendrik Drachsler, Nikos Manouselis, Martin Wolpers, Riina Vuorikari, and Erik Duval. 2011. Dataset-driven research for improving recommender systems for learning. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge, LAK 2011, Banff, AB, Canada, February 27 - March 01, 2011*, Phillip Long, George Siemens, Gráinne Conole, and Dragan Gasevic (Eds.). ACM, 44–53. https://doi.org/10.1145/2090116.2090122

[26] Michael Völske, Matthias Hagen, and Benno Stein. 2019. *Webis Query-Task-Mapping Corpus 2019 (Webis- QTM-19)*. https://doi.org/10.5281/zenodo.3257431

[27] Johannes von Hoyer, Georg Pardi, Yvonne Kammerer, and Peter Holtz. 2019. Metacognitive Judgments in Searching as Learning (SAL) Tasks: Insights on (Mis-) Calibration, Multimedia Usage, and Confidence. In *Proceedings of the 1st International Workshop on Search as Learning with Multimedia Information* (Nice, France) *(SALMM '19)*. Association for Computing Machinery, New York, NY, USA, 3–10. https://doi.org/10.1145/3347451.3356730

[28] Mark Warschauer, Lynn Reimer, Kameryn Denaro, Gabe Orona, Katerina Schenke, Tutrang Nguyen, Amanda Niili, Di Xu, Sabrina Solanki, and Tate Tamara. 2021. *Evaluating Promising Practices in Undergraduate STEM Lecture Courses*. https://doi.org/10.7280/D11M5Q