

Abschlussbericht zum Vorhaben SALIENT

Titel des Vorhabens: SALIENT - Search as Learning: Investigating, Enhancing, and Predicting Learning during Multimodal (Web) Search

Projektnummer/Aktenzeichen: K68/2017

Executive Summary

Das Projekt SALIENT startete im Mai 2018 und lief, nach kostenneutraler Projektverlängerung um sechs Monate, bis Ende Oktober 2021. In dem Vorhaben wurden Lernvorgänge während der Websuche erforscht, wobei die Nutzung multimedialer Onlineressourcen besonders im Fokus war. Damit hebt sich das Projekt von bisherigen Arbeiten im Bereich *Search as Learning* (SaL) ab, die sich bislang hauptsächlich auf die Nutzung textueller Online-Inhalte konzentriert haben.

Das interdisziplinäre Konsortium mit Expert:innen aus Psychologie und Informatik arbeitete im Projektzeitraum erfolgreich zusammen, was unter anderem durch 35 Projektpublikationen dokumentiert ist. Das vorgesehene Arbeitsprogramm wurde umgesetzt und die Ziele des Vorhabens konnten erreicht werden. Eine groß angelegte Studie zur Untersuchung von wissensorientierten Such- und Lernprozessen im Web wurde bereits im Dezember 2018 durchgeführt. Auf Basis dieser Daten konnte u.a. gezeigt werden, dass die Einbeziehung multimedialer Ressource zu einer besseren automatischen Vorhersage des Wissenszuwachses bei den Proband:innen führen kann (Otto et al., 2021). Ein weitere Studie befasste sich mit der Rolle von kognitiven Fähigkeiten und der Auswirkung der Nutzung von Text- und Videoressourcen auf das informelle Lernen im Web (Pardi et al., 2020). Weitere Ergebnisse der ersten Studie wurden in einer Folgestudie im Herbst 2019 validiert. Hierbei wurde ein bisher nicht untersuchter Effekt, der "False-certainty-Effekt" genauer untersucht (von Hoyer et al., 2019; 2022): Nutzende zeigten sich nach der Lernphase im Experiment zuversichtlicher hinsichtlich der Korrektheit ihrer Antwort auf Wissensfragen – allerdings nicht nur bei korrekten, sondern auch bei falschen Antworten. Die Auswertung der erhobenen Daten führte insgesamt zu einer Reihe von Beiträgen in renommierten wissenschaftlichen Konferenzen und Journalen der involvierten Fachrichtungen (z.B. CHIIR-Konferenz). Die während der großen Initialstudie generierten Daten wurden zudem aufbereitet und einschlägig publiziert (Otto et al., 2022), um die Nachnutzung durch andere Forschende zu ermöglichen. Für alle Veröffentlichungen wurden namhafte Konferenzen und Journale der beteiligten Disziplinen mit Peer Review ausgewählt.

Der wissenschaftliche Nachwuchs war aktiv in das Forschungsvorhaben eingebunden. Im Projekt waren fünf Promovierende und eine Postdoktorandin beschäftigt. Eine Dissertation wurde bereits im Jahr 2020 verteidigt und vier weitere Dissertationen wurden in SALIENT begonnen. Deren Fertigstellung hat sich jedoch, größtenteils bedingt durch die Covid19-Pandemie, verzögert und wird im Laufe des Sommers 2022 erwartet. Auch Studierende wurden im Rahmen von 15 Abschlussarbeiten und studentischen Hilfskraftstellen in die Forschungsaktivitäten eingebunden.

Neben den wissenschaftlichen Veröffentlichungen gab es auch Meldungen der Institute und eine Pressemeldung zu SALIENT. Darüber hinaus wurden Neuigkeiten auf der Projektwebseite veröffentlicht. Im Herbst 2019 wurde ein Projekttreffen unter Beteiligung des SALIENT Advisory Boards (Prof. Saskia Brand-Gruwel und Prof. Kevyn Collins-Thompson) durchgeführt. Projektergebnisse wurden auch in Vorträgen auf nicht-wissenschaftlichen Veranstaltungen präsentiert, etwa auf der LEARNTEC oder in den eLearning Summits 2019 und 2020.

Mitglieder des Projektkonsortiums waren wesentlich in die Organisation wissenschaftlicher Veranstaltungen involviert. Zunächst wurde der Workshop "International Workshop on Search as Learning with Multimedia Information (SALMM)" organisiert (Ewerth et al., 2019), der Teil der renommierten ACM Int'l Conference on Multimedia war. Zudem wurde ein „International Workshop on Investigating Learning During Web Search“ (IWILDS) initiiert, der in 2020 und 2021 im Rahmen der ACM Int'l Conference on Information and Knowledge Management (CIKM) ausgerichtet wurde (Hoppe et al., 2020; 2021) und 2022 auf der Konferenz der ACM Special Interest Group on Information Retrieval (SIGIR) durchgeführt wird. Teil des Organisationsteam waren u. a. Anett Hoppe, Yvonne Kammerer und Ran Yu. Über die Workshops wurden zudem neue Kooperationen initiiert (etwa mit Prof. Ladislao Salmerón (University of Valencia, Spanien), Prof. Jiqun Liu (University of Oklahoma, US) und Prof. Claudia Hauff (TU Delft, Niederlande).

1. Zielerreichung und Umsetzung der Meilensteine

Der Meilensteinplan wurde dem tatsächlichen Projektbeginn (Mai 2018 statt Januar 2018) angepasst. Zu **Meilenstein 1** wurde im Arbeitspaket (AP) 1 ein interdisziplinäres theoretisches Framework zur Modellierung von Search-as-Learning-Prozessen entwickelt. Hierfür wurden für Search as Learning (SaL) relevante Konzepte und Theorien aus den beteiligten Fachdisziplinen identifiziert und in einem integrativen Modell zusammengeführt. Entsprechend der Projektplanung wurde eine vorläufige Version des Frameworks Ende 2018 festgehalten und im Nachgang kollaborativ verfeinert. Im März 2020 fand sich ein größerer Kreis des Salient-Konsortiums von neun Mitgliedern in Tübingen zusammen, um das Modell zu finalisieren. Der resultierende Artikel ist im Journal *Frontiers in Psychology* (von Hoyer et al. 2022) erschienen.

Der Projektantrag sah eine Durchführung der ersten Labor- und Feldstudien bis Mai 2019 vor (**Meilenstein 2**). Koordiniert von den beiden IWM-Arbeitsgruppen (Holtz/Cress und Kammerer/Gerjets) wurde bereits Dezember 2018 eine groß angelegte Eyetracking- und Logfile-Studie zum Lern- und Suchverhalten mit 114 Teilnehmenden durchgeführt. Hierzu wurde von L3S und TIB softwaretechnisch ein Versuchsetting entwickelt, das der Untersuchung des Einflusses multimodaler Webinhalte auf den Lernerfolg dient. Dies umfasste ein umfangreiches Logging-System (Screenrecording, Tastatur- und Mausinteraktionen, Anfragen, HTML/CSS-Daten, etc.), das die detailgetreue Reproduktion von Lernsessions in Nachuntersuchungen gestattet. Die Studiendaten sind die Grundlage zahlreicher Projektveröffentlichungen (von Hoyer et al. 2019, SALMM; Pardi et al., 2020, CHIIR; Otto et al., 2021, AIED). Erstmals beobachtet werden konnte in der Studie der „false certainty effect“ (von Hoyer et al., 2019, SALMM), zu dem Folgestudien zur Reproduzierbarkeit durchgeführt und Maßnahmen zu dessen Vermeidung untersucht wurden (von Hoyer, Kimmerle & Holtz, 2022). Weiterhin wurden die Daten genutzt, um Erkenntnisse bezüglich der Merkmalsextraktion aus multimodalen Webressourcen abzuleiten (siehe **Meilenstein 3**). Zusätzliche Einblicke ermöglichte das am GESIS entwickelte Tool „Reading Protocol“, welches die Rückverfolgung der von Nutzenden gesehenen Inhalte für alle Webseiten auf der Wortebene gestattet. Es wird somit potenziell möglich, genau zu bestimmen, in welcher Webressource bzw. aus welcher Modalität ein bestimmtes neues Inhaltskonzept gelernt wurde. Die gesammelten Forschungsdaten wurden aufbereitet und (sofern rechtlich zulässig) zur Nachnutzung bereitgestellt („SaL Dataset“: <https://doi.org/10.25835/0062363>). Eine Beschreibung der Datenbasis wurde in einem „Resource Paper“ in einer Konferenzpublikation vorgestellt (Otto et al., 2022, CHIIR).

Meilenstein 3 zielte auf die Optimierung von Algorithmen zur Merkmalsextraktion für Search-as-Learning-Szenarien ab. Dies begann mit der Entwicklung der umfangreichen Infrastruktur zur Begleitung der durchgeführten Laborstudie (siehe Meilenstein 2). Die Nachlese der Studie bestand in der Realisierung eines Frameworks zur multimodalen Merkmalsextraktion, das die aufgenommenen Webdaten für die weitere Analyse aufbereitet und bereinigt, sowie aufwändige Methoden zur Beschreibung der gesehenen Inhalte implementiert. Dies nahm etwas mehr Zeit in Anspruch als erwartet, geschuldet dem Detailgrad der Analysen. So werden für die Beschreibung der visuellen Informationen etwa ausschließlich die von den Nutzenden inspizierten Anteile der Webressourcen analysiert und nicht die komplette Webseite. Hierzu wurden auf Basis der Videomitschnitte und Browsingprotokolle such- und lernrelevante Webseiten voneinander unterschieden und anschließend mit Deep-Learning-Verfahren automatisch analysiert.

Die Anforderungen an die Merkmalsextraktion wurden im interdisziplinären Dialog iterativ eruiert. So wurden etwa Merkmale von Bildern gesammelt, die aus lernpsychologischer Sicht sinnvolle Informationen über visuelle Lernmedien darstellen. Diese reichten von einer prozentualen Aufteilung des Lerninhalts in verschiedene Klassen (Bild, Text, Titel, Liste, Menüs, Hintergrund) bis hin zur Einordnung des Bildinhalts anhand von Bildtypen (Diagramme, Informationsvisualisierungen, Fotos, etc.). Darauf aufbauend wurde eine erste Korrelationsanalyse zwischen dem beobachtbaren Nutzendenverhalten in Bezug auf multimediale sowie textuelle Inhalte und dem erfolgten Wissenszuwachs durchgeführt. Die ermittelten Korrelationen sind eher schwach, geben aber Hinweise für das Design zukünftiger Studien, um die beobachteten Effekte zu bestätigen oder zu widerlegen (etwa die Tendenz von

Nutzenden mit geringem thematischen Vorwissen, auf Videos zurückzugreifen). Es konnte aber gezeigt werden, dass die Multimedia-Merkmale helfen, die Vorhersage des Wissenszuwachses bei den Lernenden zu verbessern (Otto et al., 2021). Eine weitere Studie analysierte, ob mittels der Textkomplexität der konsultierten Webressourcen ein Wissenszuwachs vorhergesagt werden kann (Gritz et al. 2021). Weiterhin wurde im Rahmen der Dissertation von Christian Otto eine weitere Studie durchgeführt, welche die Extraktion textueller, visueller und intermodaler Merkmale in MOOC-Videos vorstellt (Shi et al., 2019) und ihren Nutzen zur Vorhersage des Wissenszuwachses untersucht (Otto et al., 2022, AIED). Dazu werden neben bekannten Merkmalen wie der Lesbarkeit oder Textkomplexität auch neue intermodale Merkmale vorgestellt. So wurde etwa das mündliche Hervorheben von Folieninhalten herangezogen. Es zeigte sich, dass die Berücksichtigung von multimodalen bzw. textuellen Merkmalen hilfreich sein kann für die Vorhersage des Wissenszuwachses.

Meilenstein 4 beinhaltet die Planung und Durchführung weitergehender Laborexperimente, die die Beobachtungen aus den Initialstudien weiter verfolgen. Wie bereits erwähnt wurde in der ersten Studie ein neuer Effekt beobachtet („false certainty effect“) (von Hoyer et al., 2019). Um diesen genauer zu untersuchen, wurde im Juli 2019 eine Folgestudie durchgeführt. Im Rahmen der Dissertation von Johannes von Hoyer wurden weitere experimentelle Studien zum „false certainty effect“ durchgeführt. Der Effekt konnte mit verschiedenen Wissensthemen repliziert werden und es wurde gezeigt, dass der Effekt kausal durch eine kurze Online-Lernaktivität bedingt wird. Außerdem scheint das Ausmaß des Wissenserwerbs zwar eine Rolle für die Größe des Effekts zu spielen, aber nicht ursächlich zu sein, d.h. die Aktivität des Lernens an sich reicht aus (von Hoyer, Kimmerle & Holtz, 2022). Weitere Ergebnisse zeigten, dass der Effekt nicht nur bei Online-Lernaktivitäten beobachtet werden kann. In einem Folgeprojekt soll der Frage nachgegangen werden, ob sich der Effekt durch ein Training der metakognitiven Fähigkeit abmildern lässt. Des Weiteren wurden im Rahmen der Dissertation von Georg Pardi zwei Laborexperimente und drei Online-Studien durchgeführt, in denen der Einfluss der Wissensart auf das Selektionsverhalten hinsichtlich der Modalität von Suchergebnissen während der Websuche untersucht wurde. Hierbei konnte etwa beobachtet werden, dass die Präferenz für verschiedene Modalitäten (z.B. Videos oder Webseiten) durch die Wissensart der Aufgabe und den Grad an visuospatialen Darstellungen beeinflusst wird (Pardi et al., eingereicht bei „Computers & Education Open“). Weitere Ergebnisse werden in einem weiteren Manuskript zur Begutachtung eingereicht.

Im Projektantrag war die Organisation von zwei wissenschaftlichen Workshops zu den in SALIENT behandelten Themen vorgesehen (**Meilenstein 5**). Dies wurde wie geplant realisiert: Im Rahmen der renommierten ACM Int'l Conference on Multimedia wurde der Workshop „Search as Learning with Multimedia Information“ (SALMM 2019) ausgerichtet. Organisiert wurde der Workshop von Ralph Ewerth, Stefan Dietze, Anett Hoppe und Ran Yu. Für das Programm konnten zwei namhafte Wissenschaftler als Keynote-Sprecher gewonnen werden (Ladislao Salmerón, Universität von Valencia und Raphael Troncy, Eurecom Data Science). Aus dem Workshop entstanden weitergehende Kooperationen. So wirkte Prof. Salmerón als Organisator des zweiten SALIENT-Workshops „Int'l Workshop on Investigating Learning During Web Search (IWILDS)“ mit. Dieser wurde erstmalig im Rahmen der ACM Int'l Conference on Information and Knowledge Management (CIKM) 2020 ausgerichtet (Organisationskommittee: Anett Hoppe, Yvonne Kammerer, Ladislao Salmerón, Ran Yu). Für Keynotes wurden Carita Kiili (Tampere Universität, Finnland) und Claudia Hauff (Technische Universität Delft, Niederlande) gewonnen. Der IWILDS-Workshop wird als Serie fortgeführt: IWILDS '21 fand im Rahmen der CIKM 2021 statt und wurde von Anett Hoppe, Jiqun Liu (University of Oklahoma, USA) und Irina Brich (IWM, Tübingen) organisiert, im Jahr 2022 wurde der Workshop bei der ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22) angenommen.

Von den Mitgliedern des Konsortiums wurden weitere wissenschaftliche Veranstaltungen organisiert, die sich mit Themen um das Lernen im Web und während der Suche drehen. Hierzu zählten die International Workshops on Linked Data for Learning and Education (LILE) in den Jahren 2018 und 2019 (ausgerichtet in Zusammenhang mit der ACM Web Science Conference, organisiert von Ran Yu, Stefan Dietze, Ralph Ewerth). Zudem trug Yvonne Kammerer zur Zusammenstellung eines Special Issues in einem namhaften Journal bei (Special Issue „How adolescents read and learn on the Web: internal and external factors“ im Journal for the Study of

Education and Development). Teilergebnisse wurden zudem einem nicht-wissenschaftlichen Publikum vorgestellt, etwa in Vorträgen auf der LEARNTEC 2019 (<https://www.learntec.de/>), vor privatwirtschaftlichem Publikum mit Vorträgen auf zwei Tagungen der eLearning Summits 2019 sowie 2020 veranstaltet vom e-Learning Journal, bzw. im Rahmen eines Interview-Workshops der deutschen Journalistenschule („Deutsche Journalistenschule trifft Leibniz“, 2021).

2. Aktivitäten und Hindernisse

Die Projektaktivitäten wurden gemäß der Projektplanung und des Meilensteinplans wie oben beschrieben vorangetrieben. Im Projektverlauf ermittelte zusätzliche Forschungsbedarfe (etwa zum False-Certainty-Effekt) wurden weiter verfolgt und in die Planung integriert.

Aufgaben mit verstärktem Kommunikations- und Abgleichungsbedarf nahmen im Projektverlauf mehr Zeit in Anspruch als ursprünglich geplant. Hier ist etwa speziell die Entwicklung des theoretischen Modells zu nennen (Meilenstein 1). Im ersten Projektabschnitt wurde wie geplant ein erster Entwurf entwickelt, dessen Verfeinerung und Finalisierung jedoch länger als geplant dauerte. Es kam über den Projektverlauf zu Aktualisierungen, bedingt durch eine Weiterentwicklung der interdisziplinären Terminologie im Projektteam, aber auch teilweise durch die in den Untersuchungen erlangten Ergebnisse. Trotzdem kann gerade dieser Meilenstein als sehr sinnvoll hervorgehoben werden. Der hier geforderte interdisziplinäre Dialog war Grundlage für die folgende Projektarbeit. Die hergestellten Zusammenhänge zwischen psychologischer und informatikseitiger Terminologie haben die Zusammenarbeit weitgehend erst möglich gemacht.

Ein Hindernis für die Projektentwicklung stellte die Covid-19-Pandemie, die im März 2020 begann, dar. Bereits vor der Pandemie wurde ein großer Teil des regelmäßigen Austauschs (zwei- bis vierwöchentliche Meetings und thematische Unterarbeitsgruppen) in Videokonferenzen durchgeführt. Die vierteljährlichen persönlichen Treffen an einem der Projektstandorte wurden jedoch als große Bereicherung wahrgenommen und ermöglichten das konzentrierte Zusammenarbeiten an spezifischen Querschnittsthemen (etwa das noch Anfang März 2020 durchgeführte Meeting zur Finalisierung des theoretischen Frameworks in Tübingen). Ähnliches gilt für die danach ausschließlich online durchgeführte Fachkonferenzen, die zwar die Vorstellung eigener Beiträge weiterhin ermöglichten, allerdings mit nur eingeschränkten Möglichkeiten zur Vernetzung mit anderen Wissenschaftler:innen. Über die soziale Komponente hinaus waren weiteren Elemente der Projektdurchführung durch die pandemische Situation eingeschränkt. So konnten geplante Laborstudien für einen längeren Zeitraum nicht durchgeführt werden. Dies hatte vor allem Einfluss auf die Promotionsvorhaben in der Psychologie, deren Abschluss im Laufe des Sommers 2022 erwartet wird.

Für die Untersuchung der entworfenen Algorithmen für Retrieval und Ranking wurde eine wissenschaftliche Suchplattform entwickelt. Hier wurde die Planung angepasst, von einer ursprünglich angedachten Laborstudie hin zu einem Setting, das vollständig online funktioniert, auf Basis der hierzu neu entwickelten multimodalen Suchplattform SALLY für wissenschaftliche Publikationen und Videos. Dies führte zu einem zusätzlichen Entwicklungsaufwand und entsprechenden Verzögerungen. Eine initiale Studie untersucht, wie verschiedene Arten, multimodale Suchergebnisse zu präsentieren sich auf die Qualität der Ergebnisse und die Nutzerzufriedenheit auswirken. Die Datenerhebung wurde abgeschlossen und in Eigenleistung mit der Auswertung begonnen, verbunden mit dem Ziel einer weiteren Publikation.

3. Ergebnisse und Erfolge

Die Projektarbeiten führten bisher zu 35 Publikationen, davon neun in internationalen Fachzeitschriften, 24 in Form von Konferenzartikeln, ein Buchkapitel und ein Abstract mit Einladung zum Vortrag auf einer Fachkonferenz. Dabei wurden 31 dieser Veröffentlichungen einem Peer-Review-Verfahren unterzogen, bei den restlichen vier Publikationen handelt es sich um Begleitartikel zu organisierten Veranstaltungen bzw. Special Issues, die ausschließlich von Verlagsseite geprüft wurden. Drei der Veröffentlichungen im Rahmen der Konferenzen besondere Aufmerksamkeit. So erhielt eine Publikation den **Best Paper Award** der ACM Int'l Conference on Multimedia Retrieval (ICMR) (Otto et al. 2019); sie beschäftigte sich neben Grundlagenforschung zur Charakterisierung cross-modaler Bezüge auch mit Ansätzen zur automatischen Erkennung von Metriken zur Beschreibung von Bild-Text-Paaren. Eine weitere Publikation war als **Best Paper Candidate** der ICALT 2020 (IEEE Int'l Conference on Learning

Technologies and Technology-Enhanced Learning) nominiert; hier wurde ein Prototyp zur verbesserten Empfehlung von Lernvideos in der beruflichen Bildung vorgestellt (Tavakoli et al., 2020). Bei der ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR) erhielt eine weitere SALIENT-Publikation einen **Best Short Paper Award** (Pardi et al., 2020); diese Arbeit untersuchte, inwieweit kognitive Fähigkeiten wie die *Working Memory Capacity* bzw. die Lesekompetenz den Lernerfolg im Experiment vorhersagen lassen.

Hervorzuheben ist auch die erfolgreiche Verteidigung der Dissertation „*Improving Knowledge Accessibility on the Web - from Knowledge Base Augmentation to Search as Learning*“ der Projektmitarbeiterin Ran Yu im Februar 2020, welche unter anderem Ergebnisse des SALIENT-Projekts beinhaltet. Vier weitere Promotionen wurden im Projektkontext begonnen und befinden sich in verschiedenen Stadien der Fertigstellung. Hierbei kam es zu Verzögerungen durch die Pandemie, speziell bei den Arbeiten in der Psychologie, die von Laborstudien abhängen. Alle Dissertationen sollen im Laufe des Sommers 2022 eingereicht werden.

Zum Projektstart veröffentlichte die Hannoversche Allgemeine Zeitung ein Interview mit Antragsteller Stefan Dietze (<https://uebermorgen.haz.de/2018/08/kuenstliche-intelligenz/>).

4. Chancengleichheit, Karriereförderung und Internationalisierung

Mit den Projektmitteln wurden fünf Promovierende (männlich: 4, weiblich: 1) und eine Postdoktorandin (w) finanziert. Unter diesen Mitarbeitenden waren fünf Deutsche und eine Chinesin (Doktorandin). Zur Auswertung der durchgeführten Nutzerstudien wurden zudem fünf wissenschaftliche Hilfskräfte angestellt (männlich: 2, weiblich: 3, deutsch). Vierzehn Studierende (männlich: 9, weiblich: 5) fertigten im Projektkontext eine Abschlussarbeit an, darunter 12 deutsche und zwei chinesische Studierende.

Von den sechs beantragten Stellen wurde eine zu Projektbeginn offen ausgeschrieben, die restlichen Stellen wurden mit bekannten bestens qualifizierten Kandidat:innen besetzt (davon vier, die bereits an den Instituten tätig waren). Es ist zu beachten, dass speziell die Durchführung und Auswertung der Nutzerstudien gute Deutschkenntnisse erforderte. Damit ist die vornehmlich deutsche Nationalität der Projektmitarbeiter:innen zu erklären. Das Projektteam wurde ergänzt durch weitere Mitarbeiter:innen aus den teilnehmenden Instituten (siehe Punkt 6).

5. Strukturen und Kooperationen

Das Projektkonsortium bestand im Wesentlichen aus den im Antrag genannten Partnerinstituten. Eine Änderung ergab sich aus dem Wechsel des Antragstellers Stefan Dietze, der im Juli 2018 einem Ruf an die Heinrich-Heine-Universität Düsseldorf bzw. an das GESIS – Leibniz-Institut für Sozialwissenschaften folgt. Dies geschah zu Projektbeginn, so dass GESIS als inhaltlich passender Partner in das Konsortium aufgenommen werden konnte. Es ergaben sich hierdurch zusätzliche Kooperationen mit dortigen Mitarbeitern, die das Projekt unterstützten (siehe Punkt 7 „Zusätzliche eigene Ressourcen“).

Im November 2019, etwa zur Hälfte der Projektlaufzeit, wurde in Köln ein Projekttreffen unter Einbeziehung des Advisory Boards durchgeführt. Prof. Dr. Saskia Brand-Gruwel (Open University, Niederlande) und Prof. Dr. Kevyn Collins-Thompson (University of Michigan, US) waren zu Gast und ließen sich über den Projektfortschritt informieren. Ihre wertvollen Impulse und konstruktiven Vorschläge gingen in die nachfolgenden Projektarbeiten ein.

Im Rahmen der ausgerichteten wissenschaftlichen Workshops wurden zudem neue Kooperationspartner gefunden. So konnte Ladislao Salmerón, einer der Keynote-Speaker aus dem ersten Workshop (SALMM'19), als Co-Organisator für den folgenden IWILDS-Workshop 2020 gewonnen werden. Vorträge der Projektmitarbeiter:innen wurden zum Knüpfen neuer Kontakte genutzt und führten zu weiteren Veröffentlichungen im Kontext von SALIENT-Themen (etwa Merkt et al., 2022, Computers & Education).

6. Qualitätssicherung

Publikationen aus dem Projektkontext erfolgten ausschließlich in namhaften, inhaltlich passenden Foren mit Peer-Review-Prozess (Ausnahmen bilden hier die Begleitpublikationen zu organisierten Veranstaltungen/Special Issues), möglichst mit Open-Access-Option. Die Artikel, die zur Begutachtung eingereicht wurden, sind nach Möglichkeit auf gängigen Preprint-Servern abgelegt worden (z. B. arxiv.org), um die Inhalte allgemein zugänglich zu machen.

Der plangemäÙe Ablauf des Projekts wurde durch regelmäßige telefonische und persönliche Treffen des Projektteams sichergestellt. Video-Telefonate zur Koordination fanden mindestens monatlich statt. In abstimmungsintensiven Phasen, wie zum Beispiel rund um die Erstellung einer neuen Publikation, wurde die Frequenz nach Bedarf erhöht. Zudem fanden sich bedarfsgemäß Teilgruppen des Konsortiums zu thematischen Schwerpunkten zusammen, um spezifische Fragestellungen zu erarbeiten. So gab es etwa Fokusgruppen zur technischen Umsetzung der Studien, zu Teilkomponenten des theoretischen Modells oder zur Diskussion der Merkmalsmenge zur Multimediaanalyse.

7. Zusätzliche Ressourcen

Über die finanzierten Stellen hinaus haben alle Institute Eigenleistungen eingebracht.

Personalmittel: Alle Antragsteller:innen haben sich intensiv am Projektfortschritt beteiligt. Die leitenden Wissenschaftler:innen aus GESIS, IWM, L3S und TIB werden nicht aus Projektmitteln bezahlt, nahmen jedoch an den Treffen (telefonisch und in Person) teil, überwachten den Projektfortschritt, wirkten an Publikationen mit und betreuten die Doktorand:innen und Abschlussarbeiten. Die TIB übernahm zusätzlich die Projektleitung und die damit verbundenen Management- und Koordinationsaufgaben. An drei Instituten waren Postdoktorand:innen am Projekt beteiligt, die teilweise bereits an der Antragstellung beteiligt waren. Diese Mitarbeitenden wurden ganz oder teilweise aus Haushaltsmitteln der jeweiligen Einrichtung bezahlt. Hier sind zu nennen Dr. Daniel Hienert (GESIS) sowie die Mit Antragstellenden Dr. Peter Holtz (IWM) und Dr. Yvonne Kammerer (IWM). Weiterhin unterstützte Dr. Peter Mutschke von Juli 2018 bis Januar 2020 als leitender Angestellter des GESIS das Projekt. Für kleinere Programmier- und Annotationsaufgaben wurden an der TIB Hilfskräfte stundenweise beauftragt, die nicht durch Projektmittel finanziert wurden. Seit Juni 2020 unterstützte Wolfgang Gritz (M.Sc., TIB) mit 25% seiner Stelle (E13) maßgeblich die Datenanalyse und die Entwicklung der Suchplattform SALly.

Sachmittel: Der als Projektbeitrag organisierte Workshop SALMM (in Zusammenhang mit der ACM Multimedia '20) wurde aus Mitteln der TIB finanziert, etwa die Reisekosten für die Mitglieder des Organisationskommittees und die Keynote-Speaker. Alle Mitarbeiter:innen des Projekts wurden mit der notwendigen technischen Ausrüstung ausgestattet, etwa in Form von Serverleistungen (GESIS, IWM, L3S, TIB) bzw. Laborkapazitäten (IWM).

8. Ausblick

Im Rahmen der durchgeführten Studien zeigte sich, dass besonders Videos von den Proband:innen als Lernressourcen genutzt wurden. Dies untermauert zunächst die Grundannahme des Projektantrags, dass die Unterstützung von Online-Lernaktivitäten nicht bei der Auswertung von textuellen Daten aufhören kann, sondern diese multimodal betrachtet werden müssen. Es konnte auch beobachtet werden, dass einige Teilnehmer:innen direkt auf spezialisierte Videoplattformen zugriffen, ohne den Umweg über eine generalisierte Suchmaschine. Dies zeigt, dass eine spezifischere Erforschung des Mediums Video, als inhärent multimodales Medium, erstrebenswert ist. Es stellen sich hierbei Fragen nach der Formulierung von passenden Suchanfragen, zu deren korrekten Auswertung und Zuordnung zu geeigneten Lernressourcen (etwa den Zugriff auf spezifische Fachinhalte in längeren Videos), wie Konzepte und Zusammenhänge von den Lernenden aufgenommen und integriert (etwa die Frage nach dem Erwerb von Fachsprache) werden, und inwieweit Lernende bei der Zusammenstellung geeigneter Videosequenzen für den Lernbedarf unterstützt werden können.

Momentan sind drei Drittmittelanträge für verwandte Themenbereiche zur Begutachtung eingereicht (bei VolkswagenStiftung und Leibniz-Gemeinschaft). Weiterhin ist ein Antrag für eine DFG-Forschungsgruppe geplant, an dem mehrere Forschende aus SALIENT beteiligt sind.

Referenzen

- Ewerth, R., Dietze, S., Hoppe, A., & Yu, R. (2019). SALMM'19: First International Workshop on Search as Learning with Multimedia Information. In Proc. of the 27th ACM International Conference on Multimedia (MM '19). ACM, New York, NY, USA, 2724–2725.
- Gritz, W., Hoppe, A., & Ewerth, R. (2021). On the Impact of Features and Classifiers for Measuring Knowledge Gain during Web Search - A Case Study. In G. Cong & M. Ramanath (Eds.), Proceedings of the CIKM 2021 Workshops co-located with 30th ACM International Conference on Information and Knowledge Management (CIKM 2021), Gold Coast, Queensland, Australia, November 1-5, 2021 (Vol. 3052). CEUR-WS.org.
- Hoppe, A., Holtz, P., Kammerer, Y., Yu, R., Dietze, S., & Ewerth, R. (2018). Current Challenges for Studying Search as Learning Processes. Linked Learning Workshop – Learning and Education with Web Data (LILE), in conjunction with ACM Conference on Web Science.
- Hoppe, A., Yu, R., Kammerer, Y., & Salmerón, L. (2020). IWILDS'20: The 1st International Workshop on Investigating Learning during Web Search. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM '20). ACM, New York, NY, USA, 3535–3536.
- Hoppe, A., Yu, R., Brich, I., & Liu, J. (2021). IWILDS'21: Second International Workshop on Learning During Web Search. Proceedings of the 30th ACM International Conference on Information & Knowledge Management. ACM, New York, NY, USA, 4880–4881.
- Merkt, M., Hoppe, A., Bruns, G., Ewerth, R., & Huff, M. (2022). Pushing the button: Why do learners pause online videos? *Computers & Education*, 176, 104355.
- Otto, C., Holzki, S. & Ewerth, R. (2019). “Is This an Example Image?” - Predicting the Relative Abstractness Level of Image and Text. In L. Azzopardi, B. Stein, N. Fuhr, P. Mayr, C. Hauff & D. Hiemstra (Eds.), *Advances in Information Retrieval - 41st European Conference on IR Research, ECIR 2019, Cologne, Germany, April 14-18, 2019, Proceedings, Part I* (Vol. 11437, pp. 711–725). Springer.
- Otto, C., Rokicki, M., Pardi, G., Gritz, W., Hienert, D., Yu, R., von Hoyer, J., Hoppe, A., Dietze, S., Holtz, P., Kammerer, Y. & Ewerth, R. (2022). SaL-Lightning Dataset: Search and Eye Gaze Behavior, Resource Interactions and Knowledge Gain during Web Search. In D. Elsweiler (Ed.), *CHIIR '22: ACM SIGIR Conference on Human Information Interaction and Retrieval, Regensburg, Germany, March 14 - 18, 2022* (pp. 347–352). ACM.
- Otto, C., Springstein, M., Anand, A. & Ewerth, R. (2020). Characterization and classification of semantic image-text relations. *Int. J. Multim. Inf. Retr.*, 9, 31–45.
- Otto, C., Yu, R., Pardi, G., von Hoyer, J., Rokicki, M., Hoppe, A., Holtz, P., Kammerer, Y., Dietze, S. & Ewerth, R. (2021). Predicting Knowledge Gain During Web Search Based on Multimedia Resource Consumption. In I. Roll, D. S. McNamara, S. A. Sosnovsky, R. Luckin & V. Dimitrova (Eds.), *Artificial Intelligence in Education - 22nd International Conference, AIED 2021, Utrecht, The Netherlands, June 14-18, 2021, Proceedings, Part I* (Vol. 12748, pp. 318–330). Springer.
- Otto, C., Rokicki, M., Pardi, G., Gritz, W., Hienert, D., Yu, R., von Hoyer, J., Hoppe, A., Dietze, S., Holtz, P., Kammerer, Y. & Ewerth, R. (2022). SaL-Lightning Dataset: Search and Eye Gaze Behavior, Resource Interactions and Knowledge Gain during Web Search. In D. Elsweiler (Ed.), *CHIIR '22: ACM SIGIR Conference on Human Information Interaction and Retrieval, Regensburg, Germany, March 14 - 18, 2022* (pp. 347–352). ACM.
- Otto, C., Springstein, M., Anand, A. & Ewerth, R. (2020). Characterization and classification of semantic image-text relations. *Int. J. Multim. Inf. Retr.*, 9, 31–45.
- Papenmeier, A., Kern, D., Hienert, D., Kammerer, Y., & Seifert, C. (2022). How Accurate Does It Feel? - Human Perception of Different Types of Classification Mistakes. In S. D. J. Barbosa, C. Lampe, C. Appert, D. A. Shamma, S. M. Drucker, J. R. Williamson, & K. Yatani (Eds.), *CHI '22: CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April 2022 - 5 May 2022* (pp. 180:1–180:13). ACM.
- Pardi, G., Kammerer, Y. & Gerjets, P. (2019). Search and Justification Behavior During Multimedia Web Search for Procedural Knowledge. In P. Boldi, B. F. Welles, K. Kinder-Kurlanda,

C. Wilson, I. Peters & W. M. Jr. (Eds.), *Companion Publication of the 10th ACM Conference on Web Science, WebSci 2019, Boston, MA, USA, June 30 - July 03, 2019*. (pp. 17–20). ACM.

Pardi, G., Kammerer, Y. & Gerjets, P. (2019). Search and Justification Behavior During Multimedia Web Search for Procedural Knowledge. In P. Boldi, B. F. Welles, K. Kinder-Kurlanda, C. Wilson, I. Peters & W. M. Jr. (Eds.), *Companion Publication of the 10th ACM Conference on Web Science, WebSci 2019, Boston, MA, USA, June 30 - July 03, 2019*. (pp. 17–20). ACM.

Pardi, G., von Hoyer, J., Holtz, P. & Kammerer, Y. (2020). The Role of Cognitive Abilities and Time Spent on Texts and Videos in a Multimodal Searching as Learning Task. In H. L. O'Brien, L. Freund, I. Arapakis, O. Hoerber & I. Lopatovska (Eds.), *CHIIR '20: Conference on Human Information Interaction and Retrieval, Vancouver, BC, Canada, March 14-18, 2020* (pp. 378–382). ACM.

Shi, J., Otto, C., Hoppe, A., Holtz, P., & Ewerth, R. (2019). Investigating Correlations of Automatically Extracted Multimodal Features and Lecture Video Quality. *Proceedings of the 1st International Workshop on Search as Learning with Multimedia Information*, 11–19.

von Hoyer, J., Pardi, G., Kammerer, Y. & Holtz, P. (2019). Metacognitive Judgments in Searching as Learning (SAL) Tasks: Insights on (Mis-) Calibration, Multimedia Usage, and Confidence. *Proceedings of the 1st International Workshop on Search as Learning with Multimedia Information*, 3–10.

von Hoyer, J. F., Kimmerle, J. & Holtz, P. (2022). Acquisition of false certainty: Learners increase their confidence in the correctness of incorrect answers after online information search. *Journal of Computer Assisted Learning*, 38, 833–844.

von Hoyer, J., Hoppe, A., Kammerer, Y., Otto, C., Pardi, G., Rokicki, M., Yu, R., Dietze, S., Ewerth, R. & Holtz, P. (2022). The Search as Learning Spaceship: Toward a Comprehensive Model of Psychological and Technological Facets of Search as Learning. *Frontiers in Psychology*, 13.

Yu, R., Gadiraju, U. & Dietze, S. (2018). Detecting, Understanding and Supporting Everyday Learning in Web Search. *CoRR*, [abs/1806.11046](https://arxiv.org/abs/1806.11046).

Yu, R., Gadiraju, U., Holtz, P., Rokicki, M., Kemkes, P. & Dietze, S. (2018). Predicting User Knowledge Gain in Informational Search Sessions. In K. Collins-Thompson, Q. Mei, B. D. Davison, Y. Liu & E. Yilmaz (Eds.), *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018* (pp. 75–84). ACM.

Yu, R., Gadiraju, U., Fetahu, B., Lehmborg, O., Ritze, D. & Dietze, S. (2019). KnowMore - knowledge base augmentation with structured web markup. *Semantic Web*, 10, 159–180.

Yu, R. (2020). *Improving Knowledge Accessibility on the Web - from Knowledge Base Augmentation to Search as Learning*. Dissertation. University of Düsseldorf, Germany.

