

LERNEN MIT VIDEOS – UNTERSTÜTZT DURCH KI (?)

Prof. Dr. Ralph Ewerth ist Professor an der Fakultät für Elektrotechnik und Informatik der Leibniz Universität Hannover und leitet am Leibniz-Informationszentrum für Technik und Naturwissenschaften – Technische Informationsbibliothek (TIB) – die Forschungsgruppe Visual Analytics. Zudem ist er Mitglied des Forschungszentrums L3S in Hannover. Seine Forschungsschwerpunkte sind Visual Analytics, Multimedia Information Retrieval und informelles Lernen im Web. Prof. Dr. Ewerth hat eine Vielzahl wissenschaftlicher Beiträge veröffentlicht, von denen einige mit Preisen ausgezeichnet wurden (z. B. Best Paper Award bei der ACM International Conference on Multimedia Retrieval 2019, Promotionspreis der Universität Marburg 2009).

In der gegenwärtigen Arbeitswelt, die unter anderem durch rasche Veränderungen und notwendige Anpassungsprozesse aufseiten der Beteiligten gekennzeichnet ist, spielt das informelle Lernen mit unterschiedlichen Medien eine große Rolle. Insbesondere Videos werden immer populärer, um sich neue Kenntnisse und Fertigkeiten anzueignen. So zeigte eine Umfrage des *Pew Research Center* (SMITH/TOOR/VAN KESSEL 2018), dass mehr als die Hälfte der erwachsenen YouTube-Nutzenden die Plattform zum Aneignen neuer Fertigkeiten oder Kompetenzen besucht. Allerdings ist es weder aus psychologisch-didaktischer noch aus

technischer Hinsicht trivial, Videos lernförderlich zu gestalten. In diesem Beitrag werden exemplarisch psychologisch-didaktische Aspekte von Lernvideos beleuchtet, bevor auf Möglichkeiten zur verbesserten Nutzung von Lernvideos, basierend auf künstlicher Intelligenz (KI), eingegangen wird.

LERNEN MIT VIDEOS – DIDAKTISCHE UND PSYCHOLOGISCHE ASPEKTE

Lernvideos dienen dem Ziel, Wissen und Kompetenzen zu vermitteln. Hierbei kann es sich um unterschiedliche Arten von Wissen handeln, zum

Beispiel um Faktenwissen (z. B. „Was ist die Hauptstadt von Brasilien?“) oder prozedurales Wissen über Handlungsabfolgen (z. B. beim Autofahren). Weiterhin kann unterschieden werden, ob ein Video zu Lernzwecken erstellt wurde oder nicht. Dabei haben Videos zwei charakteristische Besonderheiten:

- 1. Multimodalität:** Videos können grundsätzlich Bewegtbilder mit Audioinformationen (sprachliche Erläuterungen, akustische Signale etc.) kombinieren und sind in der Regel multimodal. Dies hat Implikationen für deren Nutzung zu Lernzwecken.
- 2. Dynamik und Flüchtigkeit der Information:** Die dargebotenen Informationen sind in der Regel „flüchtig“ und es folgen im Verlauf eines Videos fortwährend neue visuelle oder auditive Informationen. Dies erfordert von den Lernenden kontinuierliche Aufmerksamkeit. Werden dargebotene Informationen nicht direkt verstanden, so kann dies das Verständnis der nachfolgenden Informationen erschweren oder gar unmöglich machen. Dieser Aspekt verleiht Interaktionsmöglichkeiten eine besondere

Bedeutung – etwa die Möglichkeit, das Video anhalten oder zu einer vorherigen Sequenz zurückspringen zu können.

Die psychologische Forschung betrachtet seit rund 40 Jahren das Lernen mit Videos. Eine Publikation zu einer der ersten Studien hatte den Titel: „Television is ‘easy’ and print is ‘tough’“ (SALOMON 1984). In der Studie wurden die teilnehmenden Schüler und Schülerinnen aufgeteilt: Eine Gruppe sah ein Video (ohne Ton), die andere Gruppe erhielt einen gedruckten Text zur Erarbeitung eines Themas. Es zeigte sich, dass die „Videogruppe“ weniger kognitive Anstrengungen in das Verständnis des Sachverhalts investierte als die „Textgruppe“. Auch schnitt die Textgruppe in einem Test besser ab. Die Ergebnisse konnten allerdings in Folgestudien oftmals nicht repliziert werden.

Die Theorie des Multimedialernens (*Multimedia Learning Theory*, vgl. MAYER 2005) gibt Orientierungspunkte zum Lernen mit Videos. Sie berücksichtigt, dass die Informationen von den Lernenden sowohl über den visuellen als auch über den auditiven Sinneskanal aufgenommen werden. Nach der „Cognitive Load

Theory“ ist die kognitive Kapazität beider Sinneskanäle jeweils begrenzt (SWELLER/CHANDLER 1991). Eine optimale Nutzung liegt vor, wenn beide Sinneskanäle so weit wie möglich angesprochen werden. Ein Beispiel für eine optimale Nutzung wäre, wenn eine Dozentin die Animation eines Sachverhalts in gesprochener Form erläutert; nicht optimal wäre hingegen, wenn der erklärende Text in einer Animation im Bild ohne direkte Bezüge eingeblendet würde, da dann sowohl die Animation des Sachverhalts als auch der geschriebene Text vom visuellen Sinneskanal verarbeitet werden müssten und diesen möglicherweise „überlasten“. Diese Beispiele beziehen sich auf das **Modalitätsprinzip** aus der Theorie des Multimedialernens: Gesprochener Text und Bild sind in der Regel besser als geschriebener Text und Bild. Das Prinzip der dualen Codierung (**Dual-Coding-Prinzip**) ist ebenfalls ein Element der Theorie des multimedialen Lernens und ein Anhaltspunkt, dass Videos lernförderlich sein können. Es besagt, dass die gemeinsame Verwendung von Text und Bild lernförderlicher sein kann, als wenn nur Text eingesetzt wird. Dies gilt allerdings nicht pauschal, sondern hängt von vielen Faktoren, wie etwa der Expertise der

Lernenden oder der Komplexität des Themas, ab.

An dieser Stelle sei nur kurz exemplarisch auf zwei weitere Arbeiten aus der Psychologie eingegangen, die sich mit der Gestaltung sowie möglichen Effekten von Lernvideos beschäftigen. So wurde etwa die Lernförderlichkeit von Animationen (von schematischen Inhalten bis hin zu fotorealistischen Inhalten) in einer Metastudie untersucht und belegt, insbesondere für prozedural-motorische Inhalte (HÖFFLER/LEUTNER 2007). Gezeigt werden konnte auch, dass einfache Aufgaben zur Informationssuche mittels Videos durch interaktive Verzeichnisse unterstützt werden können. Allerdings können Verzeichnisse und Register bei komplexeren Aufgaben zu einem engeren Suchraum mit nachteiligen Ergebnissen führen. Durch ein Training der Lernenden können die Medienkompetenz verbessert und eine umfassende Informationserschließung erreicht werden (MERKT/SCHWAN 2014a).

GESTALTUNG VON LEHRVIDEOS

Guo et al. (GUO/KIM/RUBIN 2014) haben Lehrvideos einer MOOC-Plattform (MOOC: *Massive Open Online*

Course) untersucht. Dazu haben sie für Videos der Plattform edX (<https://www.edx.org/>) das Engagement der Studierenden auf Basis von 6,9 Millionen Sessions für vier edX-Kurse analysiert. Zur Messung des Engagements wurde erhoben, wie oft und wie lange Videos angesehen und ob anschließende Aufgaben von den

Lernenden bearbeitet wurden. Weiterhin wurden Interviews mit den technischen Mitarbeitenden von edX geführt, die für die Produktion und die Bereitstellung der Videos mitverantwortlich waren. Die daraus abgeleiteten Erkenntnisse und Empfehlungen für die Gestaltung von Lernvideos zeigt Tabelle 1.

Tabelle 1:

Ergebnisse der Studie zu MOOC-Videos und daraus abgeleitete Empfehlungen für Lehrvideos (nach GUO/KIM/RUBIN 2014)

STICHWORT	BEOBACHTUNG	EMPFEHLUNG
DAUER	Kürzere Videos sind attraktiver und einnehmender.	Stark in die Planung einzelner, kurzer Lektionen investieren, um Videos mit einer Länge von maximal sechs Minuten zu erreichen.
VORTRAGENDE	Videos, die das Gesicht der sprechenden Person zeigen oder zeitweise einblenden, sind attraktiver (als nur Folien zu zeigen).	In das Editieren (Schnitt) des Videos nach der Aufnahme investieren, um die vortragende Person im Video zu geeigneten Zeitpunkten zu zeigen.
PRODUKTION	Videos, die mit einer persönlichen Note produziert werden, können einnehmender sein als professionelle Studioaufnahmen.	Videos sollten in einem informellen Rahmen aufgenommen werden; es ist nicht unbedingt notwendig, in eine professionelle Produktion zu investieren.
PROFESSIONALITÄT	Selbst qualitativ hochwertige, vorab aufgezeichnete Vorlesungen im Hörsaal sind nicht so ansprechend wie kurze Teil-Lektionen.	Wenn Dozierende Vorlesungen im Hörsaal aufzeichnen, sollten sie dennoch für das MOOC-Format (kurze Videos etc.) planen und anpassen.
FOLIENSTIL	Videos im „Stil der Khan Academy“, bei denen die Folieninhalte live erstellt werden, sind ansprechender als fertige Foliensätze.	Es sollte versucht werden, Folieninhalte im Laufe des Videos herzuleiten und die Erklärungen frei vorzutragen.
SPRECHSTIL	Videos, in denen Vortragende relativ schnell und mit Begeisterung sprechen, sind attraktiver.	Lehrende sollen ihren Enthusiasmus zum Ausdruck bringen; es ist z. B. nicht notwendig, zur Betonung absichtlich langsamer zu sprechen.
ZIELSZENARIO	Studierende nutzen Vorlesungs- und Tutorialvideos unterschiedlich.	Vorlesungen auf das erste bzw. einmalige Konsumieren der Videos ausrichten; bei Tutorials ist eine Unterstützung für das Suchen und wiederholte Ansehen von Segmenten notwendig.

Eine ebenfalls sehr gute Übersicht zu Gestaltungsempfehlungen für Videos in der Hochschullehre gibt Merkt (MERKT 2015). Er stellt heraus, dass Lernende sich aktiv mit den Videoinhalten auseinandersetzen und nicht nur passiv rezipieren sollten. Ein grundlegendes Element sei, die Geschwindigkeit der Informationswiedergabe an die Verarbeitungskapazität der Lernenden anpassen zu können und Möglichkeiten zur aktiven Steuerung der Videowiedergabe anzubieten. Zudem wird die Bedeutung einfacher Zugriffsmöglichkeiten auf die Inhalte eines Videos betont. Zur aktiven Auseinandersetzung mit den Videoinhalten sind handschriftliche Notizen eher hilfreich als digital angefertigte; Lernende sollten entsprechend ermuntert werden. Hervorgehoben wird auch der Usability-Aspekt von Lern- oder Videoplattformen. Es ist wichtig, dass ein solches System einfach zu bedienen ist, damit die kognitiven Kapazitäten der Lernenden für die eigentlichen Lerninhalte genutzt werden können – und nicht für die Bedienung einer (zu) komplexen Nutzerschnittstelle „verschwendet“ werden.

KÜNSTLICHE INTELLIGENZ – WAS IST DAS EIGENTLICH?

Die künstliche Intelligenz (KI) ist ein Teilgebiet der Informatik und integriert je nach Ausrichtung interdisziplinäre Elemente aus den Neurowissenschaften und der Psychologie. Als „künstlich intelligent“ werden Maschinen bzw. Programme bezeichnet, die bei bestimmten Aufgaben menschenähnliche Leistungen erbringen – oder manchmal sogar besser abschneiden (EWERTH et al. 2017). Die „Maschine“ kann dabei der ganz normale Computer unter dem Büroschreibtisch sein, aber auch ein Roboter oder ein intelligentes Fahrzeug. Aufgaben sind – historisch betrachtet – repetitive, langweilige Tätigkeiten, die dem Menschen abgenommen werden, wie das Herstellen von Bauteilen in einer Produktionsstrecke oder das automatische Erkennen und Aussortieren fehlerhafter Teile in der Produktion. Mehr und mehr können die Systeme jedoch auch komplexere Fähigkeiten lernen; so entstanden in den letzten Jahren zuverlässige automatische Übersetzer für gängige Sprachen (z. B. DeepL: <https://www.deepl.com/translator>) und KI-basierte Ansätze, die automatisch Bildbeschreibungen generieren (ANDERSEN et al. 2018)

oder allgemein Objekte und Szenen erkennen (vgl. KRIZHESVSKY/SUTSKEVER/HINTON 2012). Mit den letztgenannten Ansätzen lassen sich z. B. große Bild- und Videobestände automatisch erschließen und Bildinhalte besser auffindbar machen (vgl. MÜHLING et al. 2019). Um solche Ansätze geht es in diesem Beitrag.

LERNEN MIT VIDEOS – UNTERSTÜTZT DURCH KI

KI-Methoden können in verschiedener Weise eingesetzt werden, um das Lernen mit Videos zu unterstützen. Im Folgenden werden einige Methoden, die einen Bezug zu den Gestaltungsempfehlungen zu Lehr-/Lernvideos haben, kurz vorgestellt.

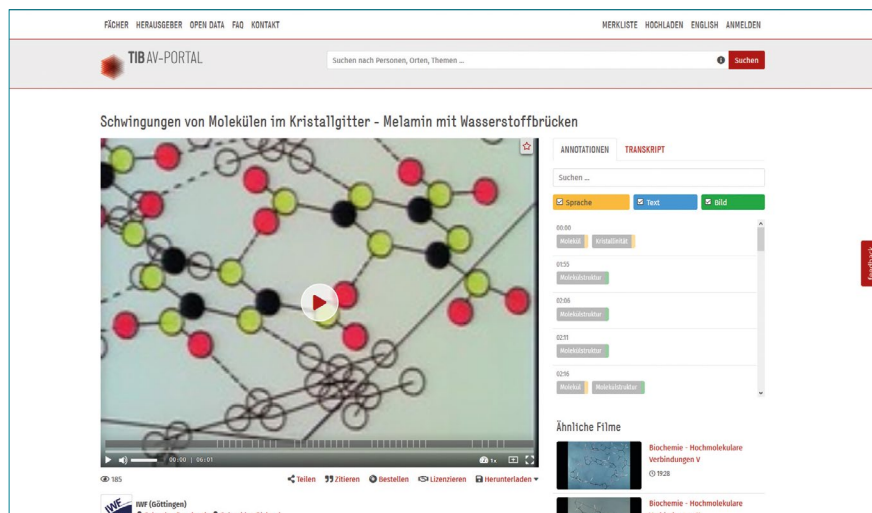
Zeitliche Strukturierung von Videos. Dieser Vorverarbeitungsschritt ist für die Strukturierung der Lehr-/Lernvideos zur Unterstützung der Suche essenziell: Eine „Schnitterkennung“ zerteilt das Video in seine grundlegenden Einheiten, die einzelnen (Kamera-)Einstellungen. In Lehrvideos können dies die Wechsel von Folieninhalten sein. Zur Erkennung der Einstellungswechsel werden Methoden des maschinellen Lernens, eines Teilgebiets der KI, eingesetzt. Klassische Verfahren hierfür

(EWERTH/FREISLEBEN 2009; BARALDI/GRANA/CUCCHIARA 2015) sind in der Regel nicht auf die Eigenschaften von Lernvideos zugeschnitten. Abrupte Wechsel im Bildinhalt und „Schnitte“ in Lernvideos sind nicht zwingend mit inhaltlichen Themen Grenzen, Themenübergänge nicht zwangsläufig mit einer abrupten Bildänderung verbunden. Verfahren zur Erkennung von Themenwechseln nutzen daher Sprachtranskripte und Texteinblendungen (TUNA et al. 2015) oder detektierte Folienwechsel (JEONG et al. 2015).

Automatische Erzeugung von Metadaten. Die Erzeugung von Metadaten dient in der Regel der verbesserten Auffindbarkeit von bestimmten Segmenten, Einstellungen oder Szenen in Videos (etwa wenn ein bestimmter Begriff genannt wird). Hierzu werden Methoden des maschinellen Lernens zur Mustererkennung eingesetzt. Ein Beispiel einer Videoplattform, die solche Methoden zur Erschließung einsetzt, ist das TIB AV-Portal (<https://av.tib.eu>) der Technischen Informationsbibliothek: Es setzt Methoden zur Spracherkennung, zur Erkennung von eingeblendetem Text und zur Bildklassifikation ein. Jede Kameraeinstellung wird automatisch mit

Abbildung 1:

Screenshot des TIB AV-Portals. Unten im Video ist das Ergebnis der Schnitterkennung eingeblendet (grauer Streifen mit vertikalen Strichen, welche die Schnitte repräsentieren). Rechts sieht man die Metadaten, die automatisch mittels Spracherkennung (gelb), Texterkennung im Bild (blau) und Bild-/Konzepterkennung (grün) generiert wurden.



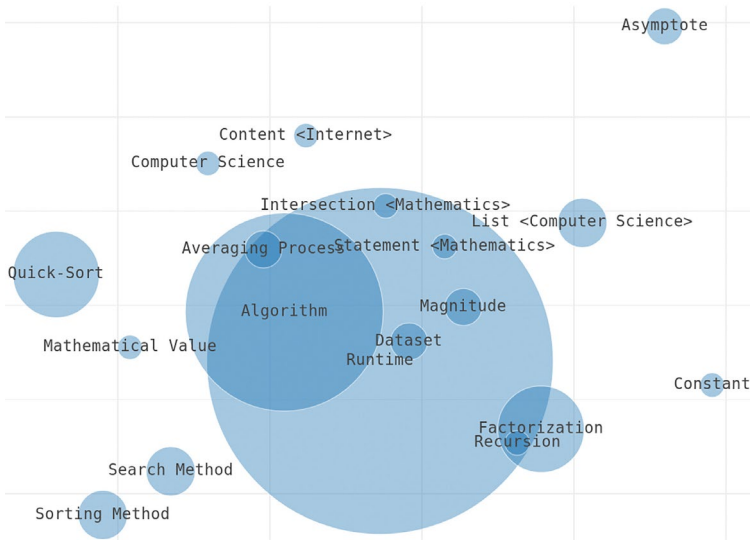
dem zugehörigen Sprachtranskript, den erkannten Schlüsselwörtern im Sprachtranskript sowie mit den erkannten eingeblendeten Begriffen versehen (siehe Abbildung 1). Auf diese Weise wird eine einstellungsgenaue, semantische Suche in den Videos ermöglicht, die bei Aufgaben der Informationssuche lernunterstützend sein kann (siehe oben).

Generierung von inhaltlichen Zusammenfassungen. Während die Generierung von Videozusammenfassungen für Spielfilme seit

Langem ein Forschungsgegenstand im Bereich Multimedia ist, ist dies für Lehr-/Lernvideos noch weniger verbreitet. Die Herausforderung ist, die für eine Zusammenfassung besonders wichtigen Inhalte zu erkennen und sinnvoll zu repräsentieren (z. B. als Text, als kurzes Video oder Diagramm). In eigenen Vorarbeiten haben wir ein System entwickelt, das die Häufigkeit der wesentlichen Begriffe und deren gemeinsames Auftreten visualisiert (vgl. ZHOU et al. 2019). Hierzu haben wir sogenannte Wordembeddings

Abbildung 2:

Visuelle Zusammenfassung des Videos „Bubblesort, Quicksort, Laufzeit“ (<https://av.tib.eu/media/9557>). Die Größe eines Kreises gibt die Häufigkeit des jeweiligen Begriffs im Video wieder, die Anordnung repräsentiert die inhaltliche Nähe der Begriffe.



zur Repräsentation von Wörtern, die mittels neuronaler Netze gelernt werden und semantische Ähnlichkeiten gut modellieren, verwendet (JOLIN et al. 2016). Diese Methode transformiert ein Wort in einen hochdimensionalen Vektor (Word2Vec); inhaltlich ähnliche Wörter befinden sich in dem hochdimensionalen Raum näher beieinander als inhaltlich unähnliche Wörter. Dies lässt sich ausnutzen, um Begriffe zu visualisieren. Ein Beispiel einer so erstellten Zusammenfassung ist in Abbildung 2 zu sehen.

Empfehlung von ähnlichen Videos.

Das Finden von zusätzlichen Informationen oder einer anderen Perspektive zu einem Thema kann ein wichtiger Aspekt des Lernens sein. Hier können Computersysteme durch Empfehlungen helfen. Im Kontext von Lern-/Lehrvideos ist bei der Berechnung der Ähnlichkeit zweier Videos die Berücksichtigung der Inhalte besonders wichtig, die Bildinformation spielt oft eine nachgeordnete Rolle. Daher haben wir in einem Ansatz *Wordembeddings* in Verbindung mit der *Dewey Decimal*

Annotation verwendet, um die inhaltliche Ähnlichkeit von Videos zu berechnen (vgl. MEDREK et al. 2018). Die berechneten Ähnlichkeiten werden dann genutzt, um thematisch passende Videos vorzuschlagen. In der dazugehörigen Nutzerstudie wurde die Nützlichkeit der Empfehlungen belegt.

Qualitätsbewertung von Videos.

In einer Studie (vgl. SHI et al. 2019) haben wir die Frage adressiert, inwiefern die Qualität von Lehr-/Lernvideos mit automatisch berechenbaren Merkmalen zusammenhängt. Die Idee ist, bei Qualitätsbewertungen nicht von der Popularität von Videos und subjektiven Merkmalen abhängig zu sein, sondern möglichst objektive, inhaltsbasierte Kriterien zu verwenden. Hierzu haben wir verschiedene Kriterien definiert: Sprachqualität, Geschwindigkeit des Vortrags, Abdeckung der Folieninhalte durch Erläuterungen, Detailgrad der Erläuterungen, Übersichtlichkeit der verschiedenen Folieninhalte (Text, Bilder, Formeln) und andere. In einer Studie wurde dann die Erfüllung der Qualitätskriterien für eine Reihe von Videos durch 13 Teilnehmende bewertet. Es wurde analysiert, wie diese Bewertungen mit den technischen Merkmalen für

Audio- und Bildinhalte korrelieren. Für einige Kriterien konnten schwache bis moderate Korrelationen festgestellt werden, so etwa für Klarheit der Sprache und einige Audiomerkmale wie Lautstärke oder das Signal-Rausch-Verhältnis. Diese Studie dient als Ausgangspunkt, um technische Merkmale zu finden, die künftig von maschinellen Lernverfahren zur Schätzung der Qualität von Videos genutzt werden können.

ZUSAMMENFASSUNG

In diesem Beitrag wurden exemplarisch einige psychologische und didaktische Aspekte von Lernvideos betrachtet sowie einige KI-basierte Methoden zu deren Erschließung und Aufbereitung vorgestellt. Zunächst haben wir uns zwei Besonderheiten von Videos vergegenwärtigt, nämlich die Multimodalität und die Flüchtigkeit der darin gezeigten Information. Dies war der Ausgangspunkt, um einerseits den Bezug zur Theorie des Multimedialernens herzustellen, andererseits haben wir Gestaltungsempfehlungen für Lehr-/Lernvideos zusammengefasst. Seitens der Informatik haben wir gezeigt, wie KI-basierte Methoden zur automatischen Erkennung und Erschließung von Videoinhalten, zur Generierung

von visuellen Inhaltsverzeichnissen für Videos, zur Empfehlung von ähnlichen Videos sowie zur Qualitätsbewertung von Lehrvideos eingesetzt werden können. Natürlich können in diesem Kurzbeitrag nicht alle Aspekte und Facetten des Lernens mit Videos behandelt werden. Sowohl aus psychologisch-didaktischer Perspektive als auch seitens der Informatik sind noch viele Forschungsfragen zu klären. Eine der Fragen lautet zum Beispiel, in welcher Form interaktive Elemente, wie Wissenstests, das Lernen mit Videos unterstützen können und wie dies mit KI-Systemen umgesetzt werden kann.

LITERATUR UND QUELLEN

ANDERSEN, PETER/HE, XIAODONG/BUEHLER, CHRIS/TENEY, DAMIEN/JOHNSON, MARK/GOULD, STEPHEN/ZHNAG, LEI (2018): Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, S. 6077–6086.

BARALDI, LORENZO/GRANA, COSTANTINO/CUCCHIARA, RITA (2015): Shot and Scene Detection via Hierarchical Clustering for Re-using Broadcast Video. International Conference

on Computer Analysis of Images and Patterns. Cham, S. 801–811

EWERTH, RALPH/FREISLEBEN, BERND (2009): Unsupervised Detection of Gradual Video Shot Changes with Motion-Based False Alarm Removal. International Conference on Advanced Concepts for Intelligent Vision Systems, S. 253–264.

EWERTH, RALPH/SPRINSTEIN, MATTHIAS/PHAN-VOGTMANN, LO AN/SCHÜTZE, JULIANE (2017): Are Machines Better Than Humans in Image Tagging? – A User Study Adds to the Puzzle. Advances in Information Retrieval – 39th European Conference on IR Research, ECIR 2017, Proceedings, S. 186–198

GUO, PHILIP J./KIM, JUHO/RUBIN, ROB (2014): How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos. Proceedings of the first ACM Conference on Learning@Scale. ACM.

HÖFFLER, Tim N./LEUTNER, DETLEV (2007): Instructional animation versus static pictures: A meta-analysis. Learning and instruction, 17 (6), S. 722–738.

JEONG, HYUN JI/KIM, TTAK-EUN/
KIM, HYEON GYU/KIM, MYOUNG HO
(2015): Automatic detection of slide
transitions in lecture videos. In: Mul-
timedia Tools and Applications, 74
(18), S. 7537–7554.

JOULIN, ARMAND/GRAVE, EDOU-
ARD/BOJANOWSKI, PIOTR/MIKOLOV,
TOMAS (2016): Bag of Tricks for Effi-
cient Text Classification. Online unter:
[https://www.aclweb.org/anthology/
E17-2068/](https://www.aclweb.org/anthology/E17-2068/) (Zugriff am 30.08.2020).

KRIZHESVSKY, ALEX/SUTSKEVER,
ILYA/HINTON, GEOFFREY E. (2012):
ImageNet Classification with Deep
Convolutional Neural Networks.
Advances in Neural Information Pro-
cessing Systems, S. 1097-1105.

MAYER, RICHARD E. (Hrsg.) (2005):
The Cambridge handbook of multi-
media learning. Cambridge.

MEDREK, JUSTYNA/OTTO, CHRIS-
TIAN/EWERTH, RALPH (2018): Recom-
mending Scientific Videos based on
Metadata Enrichment using Linked
Open Data. In Proceedings of Inter-
national Conference on Theory and
Practice of Digital Libraries (TPDL),
Porto, Portugal, S. 286–292.

MERKT, MARTIN/SCHWAN, STEPHAN
(2014a): Training the use of interac-
tive videos: effects on mastering dif-
ferent tasks. Instructional Science, 42
(3), S. 421–441.

MERKT, MARTIN/SCHWAN, STEPHAN
(2014b): How does interactivity in
videos affect task performance? In:
Computers in Human Behavior, 31,
S. 172–181.

MERKT, MARTIN (2015): Didakti-
sche Optimierung von Videos in
der Hochschullehre. Online unter:
[https://www.e-teaching.org/news/
eteaching_blog/didaktische-optimierung-
von-videos-in-der-hochschullehre](https://www.e-teaching.org/news/eteaching_blog/didaktische-optimierung-von-videos-in-der-hochschullehre)
sowie [https://www.e-teaching.org/
etresources/pdf/erfahrungsbericht_2015
_merkt_didaktische_optimierung_
video.pdf](https://www.e-teaching.org/etresources/pdf/erfahrungsbericht_2015_merkt_didaktische_optimierung_video.pdf) (Zugriff am 30.08.2020).

MÜHLING, MARKUS/MEISTER, MANJA/
KORFHAGE, NIKOLAUS/WEHLING,
JÖRG/HÖRTH, ANGELIKA/EWERTH,
RALPH/FREISLEBEN, BERND (2019).
Content-based video retrieval in his-
torical collections of the German
Broadcasting Archive. In: Internatio-
nal Journal on Digital Libraries, 20 (2),
S. 167–183.

SALOMON, GAVRIEL (1984): Television is „easy“ and print is „tough“: The differential investment of mental effort in learning as a function of perceptions and attributions. In: *Journal of Educational Psychology*, 76 (4), S. 647–658. Online unter (hinter einer Bezahlschranke): <https://doi.org/10.1037/0022-0663.76.4.647> [Zugriff am 30.08.2020],

SHI, JIANWEI/OTTO, CHRISTIAN/HOPPE, ANETT/HOLTZ, PETER/EWERTH, RALPH (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 @ ACM Multimedia Conference*. Nizza, Frankreich, S. 11–19.

SMITH, AARON/TOOR, SKYE/VAN KESSEL, PATRICK (2018): Many Turn to YouTube for Children's Content, News, How-To Lessons. Washington. Online unter: <https://www.pewresearch.org/internet/2018/11/07/many-turn-to-youtube-for-childrens-content-news-how-to-lessons/> (Zugriff am 30.08.2020).

SWELLER, JOHN/CHANDLER, PAUL (1991). Evidence for cognitive load theory. In: *Cognition and instruction*, 8 (4), S. 351–362.

TUNA, TAYFUN/JOSHI, MAHIMA/VARGHESE, VARUN/DESHPANDE, RUCHA/SUBHLOK, JASPAL/VERMA, RAKESH (2015): Topic Based Segmentation of Classroom Videos. 2015 IEEE Frontiers in Education Conference, El Paso, TX, S. 1–9.

ZHOU, HANG//OTTO, CHRISTIAN/EWERTH, RALPH (2019). Visual Summarization of Scholarly Videos using Word Embeddings and Keyphrase Extraction. In: *International Conference on Theory and Practice of Digital Libraries (TPDL)*, Oslo, Norwegen, S. 327–335.