LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

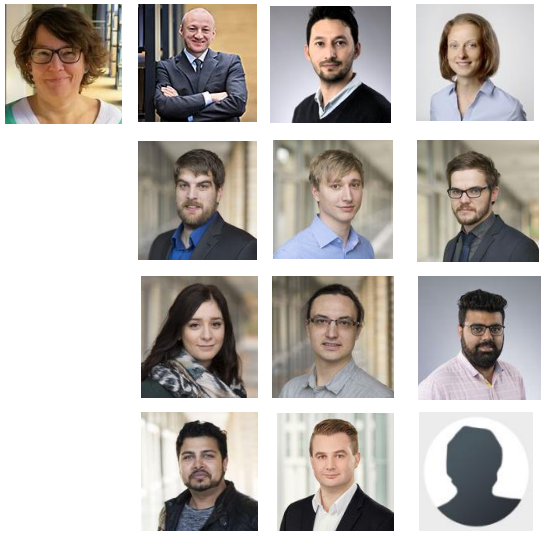
TIB

Wie kann Künstliche Intelligenz das Lernen mit Videos im Web unterstützen?

Prof. Dr. Ralph Ewerth
TIB – Forschungsgruppe Visual Analytics

Fachtagung „Künstliche Intelligenz“, Hannover
Agentur für Erwachsenen- und Weiterbildung
06. März 2020

Forschungsgruppe Visual Analytics – Wer wird sind…



Seit 2015 an der TIB, inzwischen

- 2 Postdoktorand*innen
- 10 Doktorand*innen
- 4 studentische Hilfskräfte
- 1 Administrative Fachangestellte

Forschungsschwerpunkte

- Automatische Analyse von Mediendaten
- Multimodale Suche und Bild-Text-Bezüge
- Digitale Bibliothek als virtueller Lernort

Ziele/Aktivitäten/Ergebnisse:

- Transfer von Ergebnissen in TIB-Dienste
- Drittmittelprojekte (EU, DFG, etc.) & Publikationen
- Förderung von wissenschaftlichem Nachwuchs

Relevante Forschungsthemen

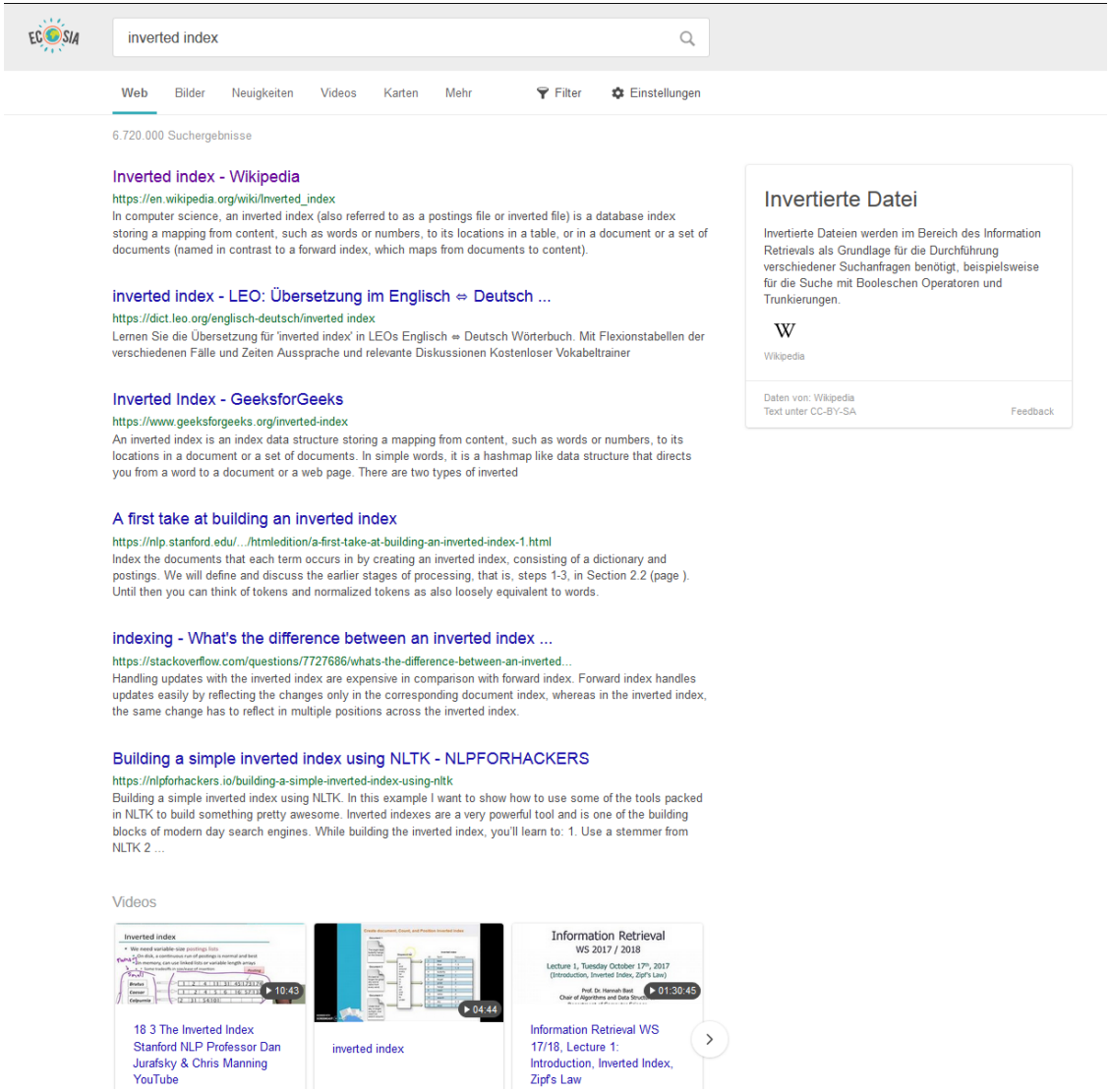
- **Automatische Annotation & semantische Suche (Bilder, Video, Text, etc.)**
- **Digitale Bibliothek als virtueller Lernort**
- **Multimodale & Cross-modale Suche**
- Intelligente Visualisierung
- Informationsextraktion ORKG
- Usability-Aspekte
- Interaktive Exploration von Mediendaten
- Maschinelles Lernen & Deep Learning

Auswahl Forschungsprojekte

- **CLEOPATRA:** Sprach-übergreifende Analyse von Nachrichten
- **iOCW:** SlideWiki – Barrierefreiheit für Blinde und Sehbehinderte
- **iART:** Interaktive Suchmaschine für Bilder der Kunstgeschichte
- **VIVA:** Erschließung des Fernseharchivs der ehemaligen DDR
- **SALIENT:** Informelles Lernen im Web mit Multimediadaten

Motivation –
Informelles Lernen im Web

Motivation – Eine Suchanfrage im Web…



…und evtl. ein Ergebnis davon: Wikipedia & Bildersuche

Von Wikipedia:

“In computer science, an **inverted index** (also referred to as a **postings file** or **inverted file**) is a database index storing a mapping from content, such as words or numbers, to its locations in a table, […]. A **record-level inverted index** (or **inverted file index** or just **inverted file**) contains a list of references to documents for each word.”

[https://en.wikipedia.org/wiki/Inverted_index]

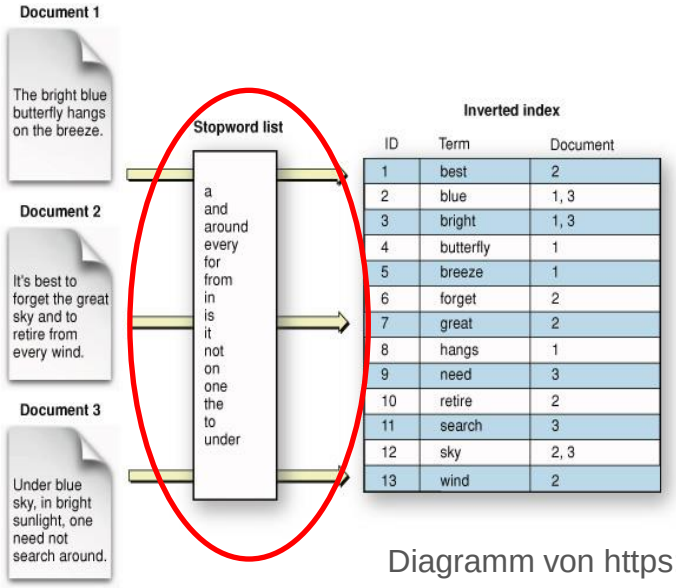
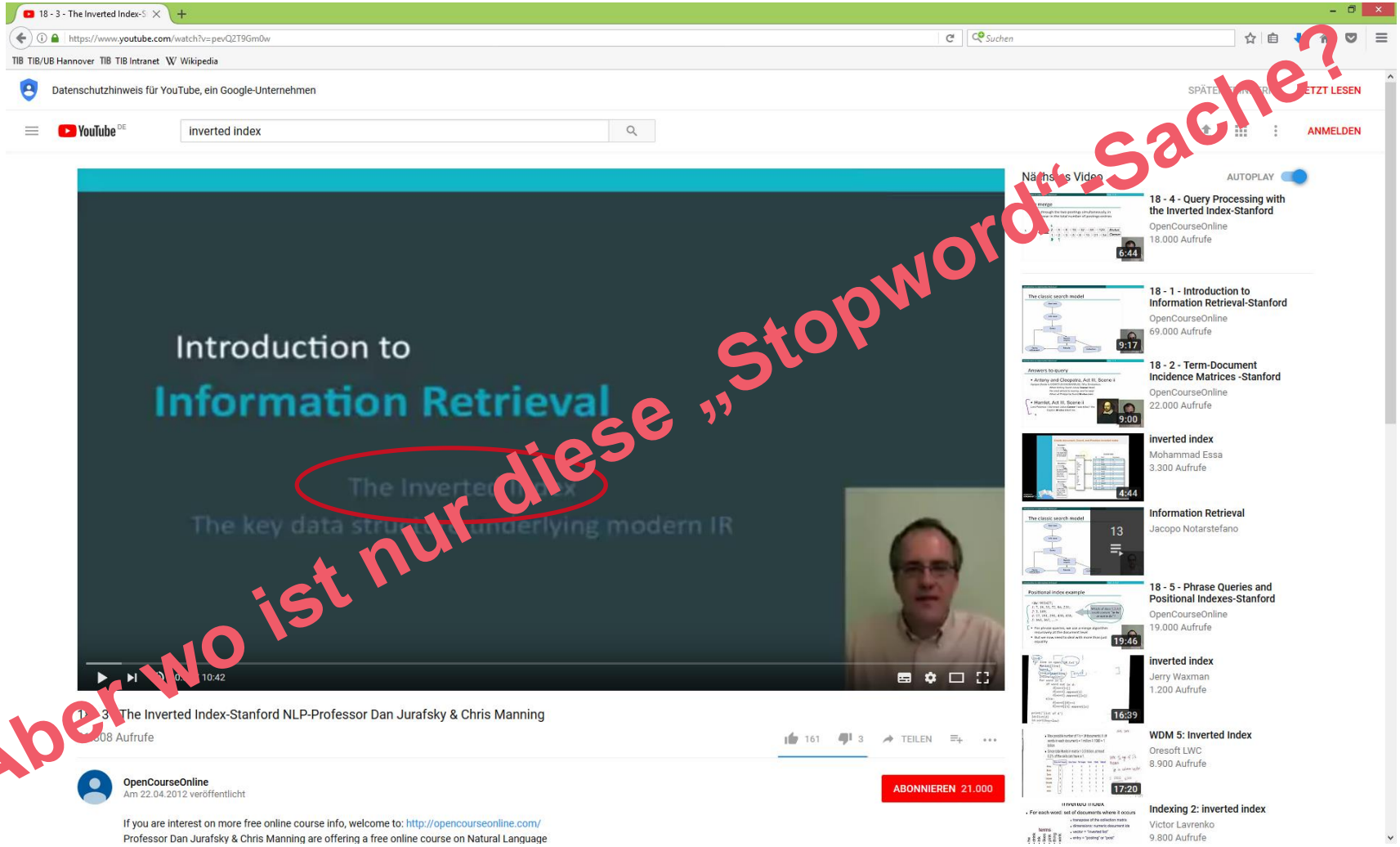


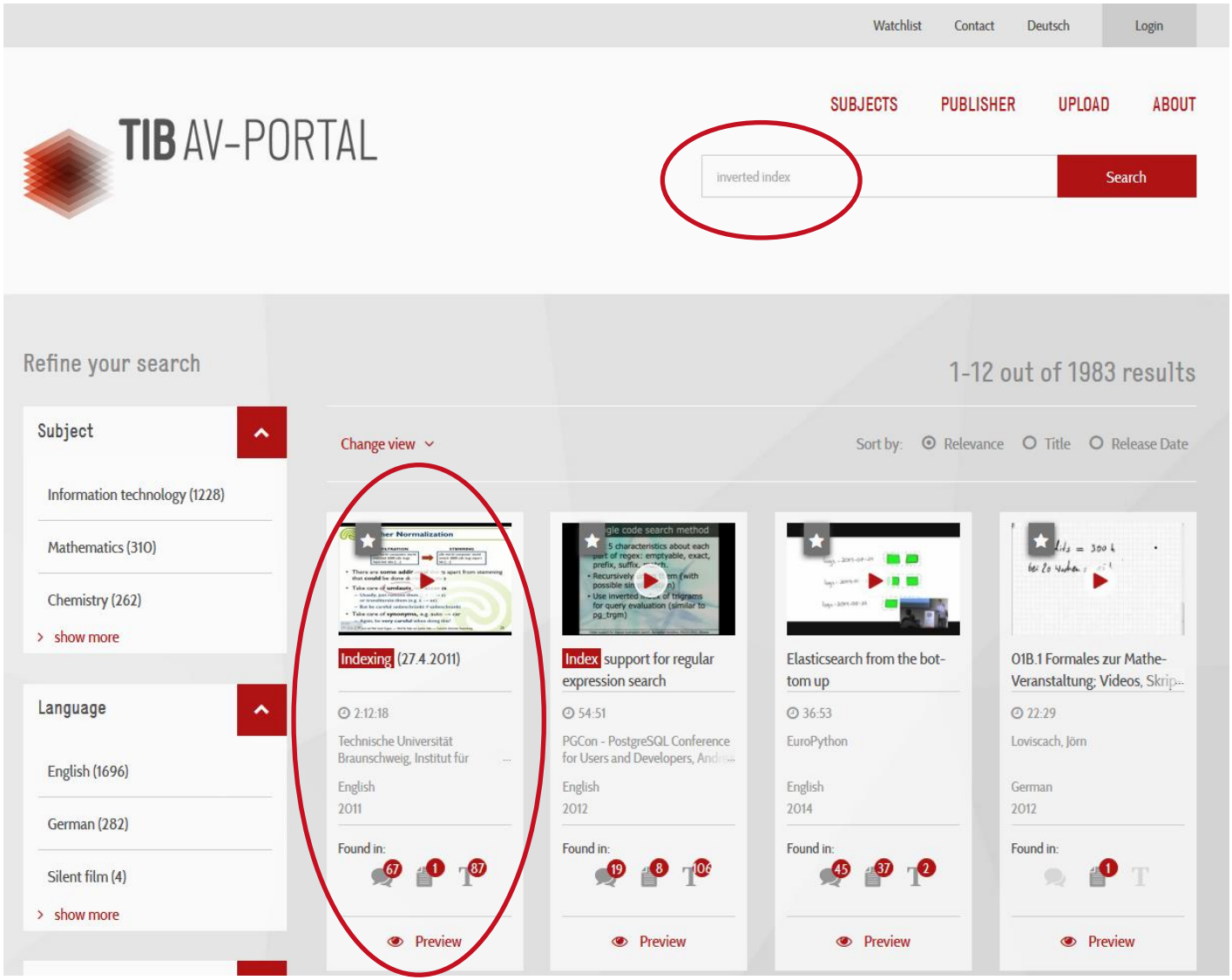
Diagramm von https://stackoverflow.com

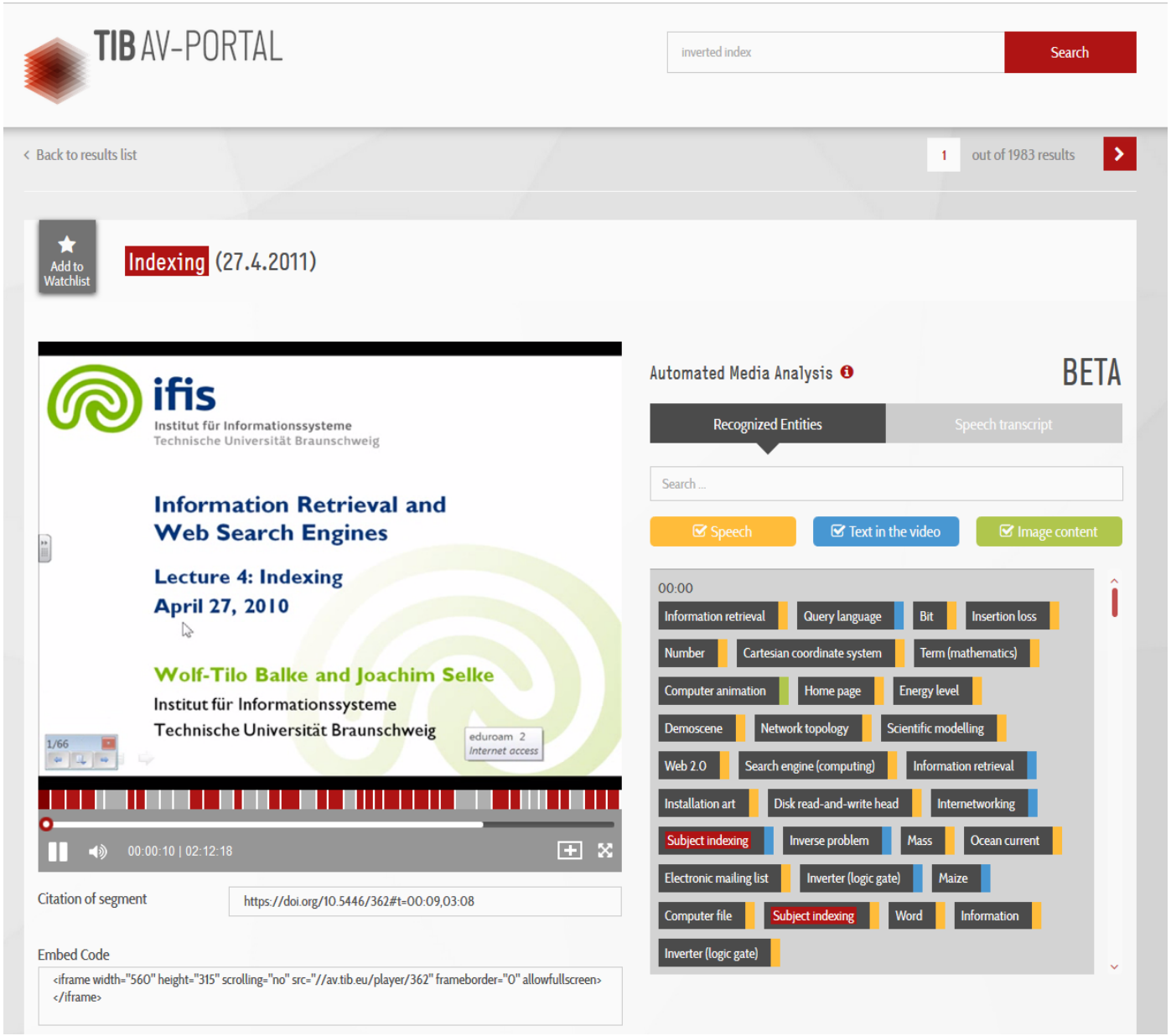
Weiterlernen im Web – mit YouTube?

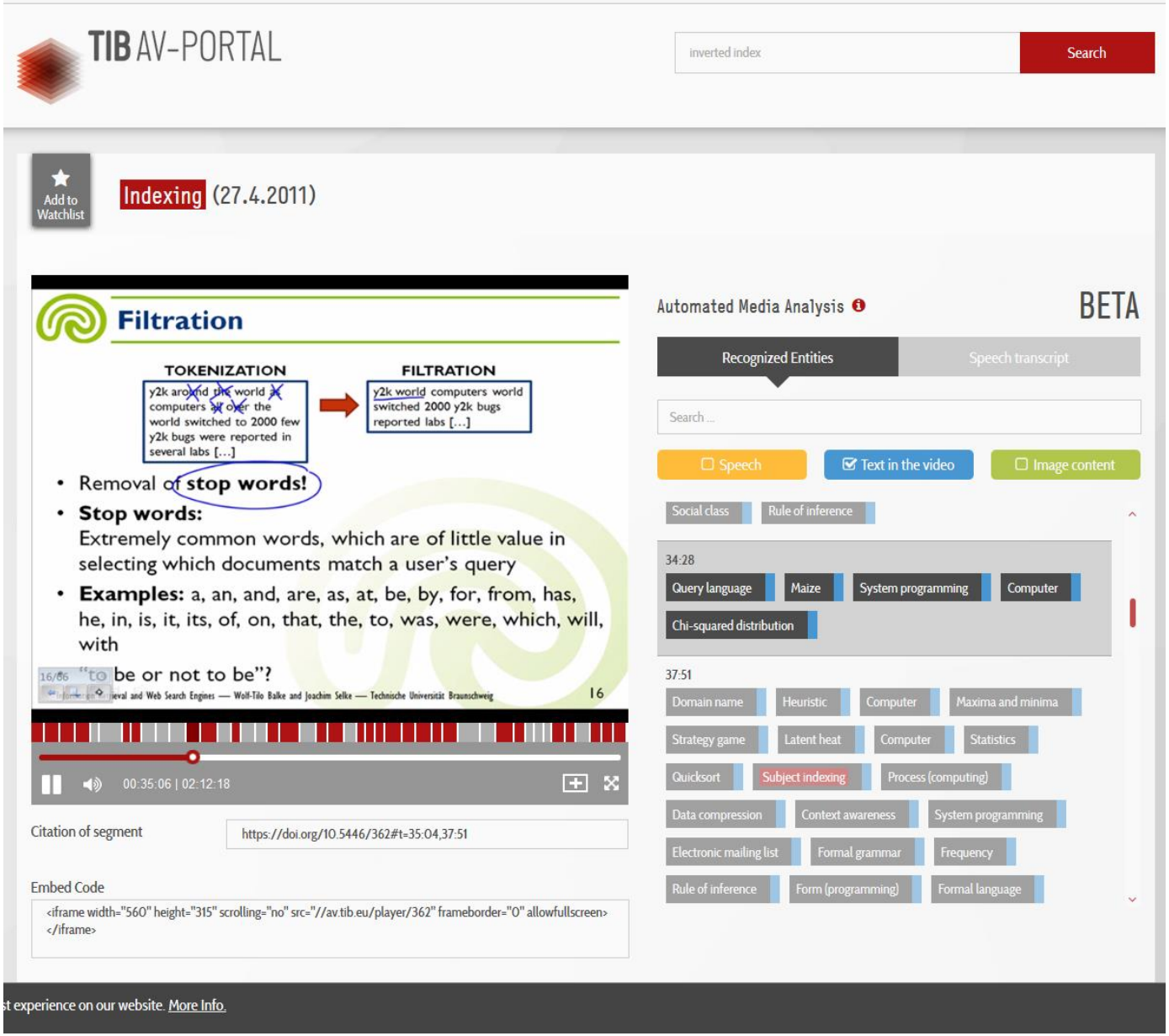


"18 3 The Inverted Index Stanford NLP Professor Dan Jurafsky & Chris Manning YouTube"
Quelle: https://www.youtube.com/watch?v=bnP6TsqyF30

Lernen im Web mit dem TIB AV-Portal?



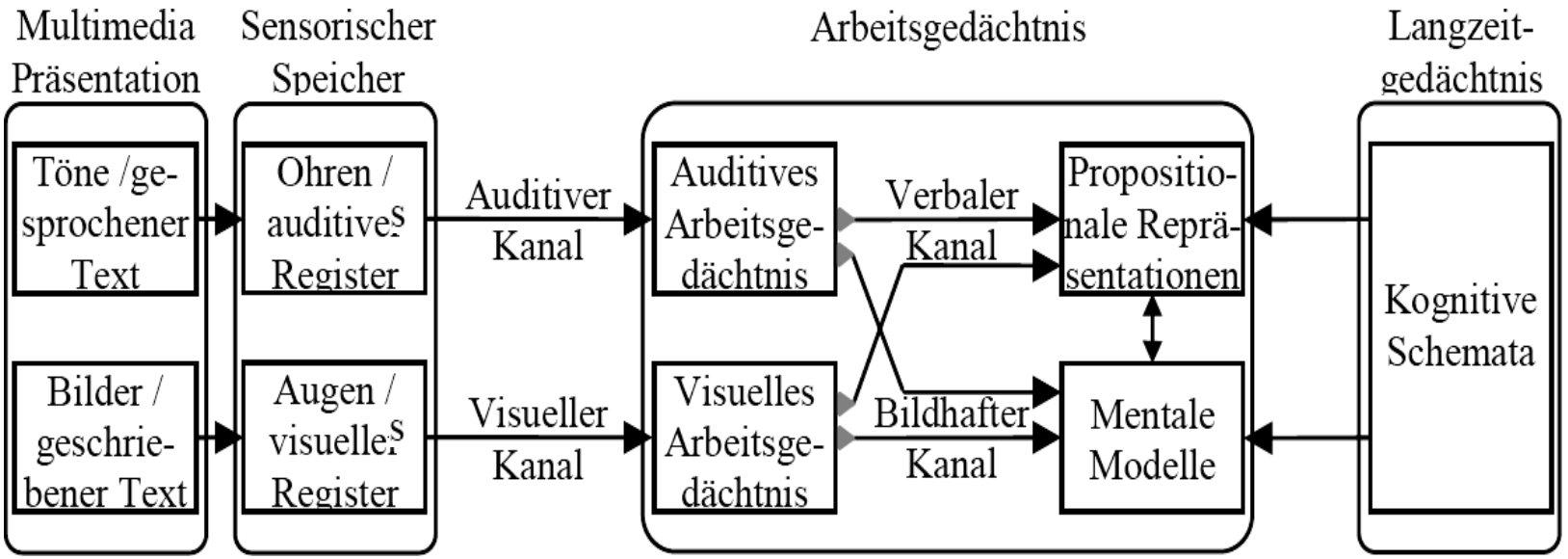

TIB AV-PORTAL
inverted index
Search
Back to results list
1
out of 1983 results
Add to Watchlist
Indexing (27.4.2011)
ifis
Institut für Informationssysteme
Technische Universität Braunschweig
Information Retrieval and Web Search Engines
Lecture 4: Indexing
April 27, 2010
Wolf-Tilo Balke and Joachim Selke
Institut für Informationssysteme
Technische Universität Braunschweig
eduroam 2
Internet access
1/66
00:00:10 | 02:12:18
Citation of segment
https://doi.org/10.5446/362#t=00:09,03:08
Embed Code
<iframe width="560" height="315" scrolling="no" src="//av.tib.eu/player/362" frameborder="0" allowfullscreen></iframe>
Automated Media Analysis
BETA
Recognized Entities
Speech transcript
Search ...
Speech
Text in the video
Image content
00:00
Information retrieval
Query language
Bit
Insertion loss
Number
Cartesian coordinate system
Term (mathematics)
Computer animation
Home page
Energy level
Demoscene
Network topology
Scientific modelling
Web 2.0
Search engine (computing)
Information retrieval
Installation art
Disk read-and-write head
Internetworking
Subject indexing
Inverse problem
Mass
Ocean current
Electronic mailing list
Inverter (logic gate)
Maize
Computer file
Subject indexing
Word
Information
Inverter (logic gate)


TIB AV-PORTAL
inverted index
Search
Add to Watchlist
Indexing (27.4.2011)
Filtration
TOKENIZATION
FILTRATION
y2k world computers world switched 2000 y2k bugs reported labs [...]
Removal of stop words!
Stop words:
Extremely common words, which are of little value in selecting which documents match a user's query
Examples: a, an, and, are, as, at, be, by, for, from, has, he, in, is, it, its, of, on, that, the, to, was, were, which, will, with
be or not to be"?
16
00:35:06 | 02:12:18
Citation of segment
https://doi.org/10.5446/362#t=35:04,37:51
Embed Code
<iframe width="560" height="315" scrolling="no" src="//av.tib.eu/player/362" frameborder="0" allowfullscreen></iframe>
Automated Media Analysis
BETA
Recognized Entities
Speech transcript
Search ...
Speech
Text in the video
Image content
Social class
Rule of inference
34:28
Query language
Maize
System programming
Computer
Chi-squared distribution
37:51
Domain name
Heuristic
Computer
Maxima and minima
Strategy game
Latent heat
Computer
Statistics
Quicksort
Subject indexing
Process (computing)
Data compression
Context awareness
System programming
Electronic mailing list
Formal grammar
Frequency
Rule of inference
Form (programming)
Formal language
More Info.

Erkenntnisse zum multimodalen Lernen

Kognitive Theorie(n) zum Multimedialernen

„Television is easy, print is tough“ (Salomon 1984)



Schnotz 2011

Propositionale Repräsentationen: „…grundlegende Informationseinheiten, die aus einem Prädikat und einem oder mehreren Argumenten bestehen, wobei diese durch das Prädikat in Beziehung gesetzt werden (z. B. ESSEN[handelnde Person: STUDENT, Zeit: MITTAG, Ort: MENSA]).“

https://de.m.wikipedia.org/wiki/Datei:Cognitive_Theory_of_Multimedia_Learning_(Mayer,_2005).png
https://creativecommons.org/licenses/by-sa/3.0/deed.de (CC-by 3.0)
Bildquelle 2: http://www.elearning-psychologie.de/integratives_modell_schnotz_ii.html

„Split-Attention-Effekt“

Theorie der kognitiven Belastung (cognitive load theory, Sweller et al. 1991)

„Wenn man spezielle Aufmerksamkeit aufwenden muss, um eine Abbildung dem zugehörigen Text zuzuordnen, dann fehlt diese Aufmerksamkeit bei der Auseinandersetzung mit den Lerninhalten.“ (Lehner 2018, S. 107)

Es gilt, zwei Alternativen zu vergleichen. Zu diesem Zwecke werden diese Alternativen anhand von zuvor bestimmten Kriterien eingeschätzt. Jedes Kriterium hat eine Gewichtung G, die zwischen 1 (= weniger ausgeprägt wichtig) und 4 (= sehr wichtig) liegt. Jede Alternative wird mit einer Bewertung B, die zwischen 1 (= gering ausgeprägt) und 4 (= stark ausgeprägt) liegt, eingeschätzt. Dann ermittelt man das Produkt P aus Gewicht und Bewertung: P = G x B. Die Werte für die einzelnen Alternativen werden dann aufsummiert.

	Neukunden gewinnen			Angebot ausweiten		
	G	B	P	G	B	P
Kompetenzen vorhanden	3	4	**12**	3	4	**12**
kostengünstig	4	3	**12**	4	3	**12**
in kollegiale Netze einbindbar	1	3	**3**	1	4	**4**
geringer zeitlicher Aufwand	3	2	**6**	3	3	**9**
familienverträglich	2	1	**2**	2	3	**6**
SUMME			**35**			**43**

Informationssuche mit Videos

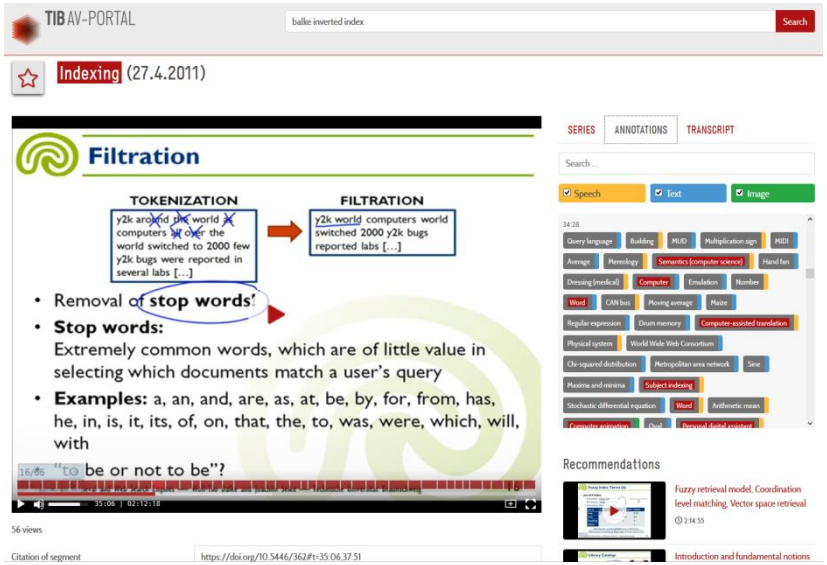
Interaktive Bedienelemente

- Play / Stop
- Kapitelstruktur / Inhaltsverzeichnisse
- Register

Einfache Suche: Informationssuche unterstützt durch interaktive Verzeichnisse (Merkt & Schwan 2014a, 2014b)

Komplexere Aufgaben:

- Engerer Suchhorizont durch Verzeichnisse und Register (Merkt et al. 2011)
- Training führt zu umfassenderer Informationserschließung (Merkt & Schwan 2014a)



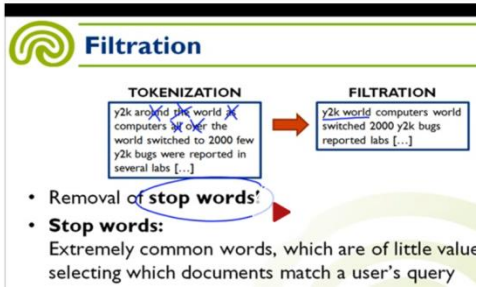
Optimierung der kognitiven Verarbeitung

Nach Theorie des Multimedialernens (beide Kanäle kapazitive Grenzen)

- **Dual-Coding-Prinzip:**
  Text und Bild **besser als** Text
- **Modalitätsprinzip:**
  Gesprochener Text & Bild **besser als** geschriebener Text und Bild

Flüchtigkeit der Information

- Videos erfordern kontinuierliche Aufmerksamkeit
- Visuelle Hinweisreize helfen



Sprachliche Referenzierung (Glaser & Schwan 2015)

- Bilder mit verbaler Referenzierung wurden besser erinnert
- Längere Fixationszeiten für benannte Bildelemente

Glaser, M., & Schwan, S. (2015). Explaining pictures: How verbal cues influence processing of pictorial learning material. *Journal of Educational Psychology, 107*(4), 1006–1018. https://doi.org/10.1037/edu0000044

Weitere Gestaltempfehlungen – MOOC Settings

Finding	Recommendation
Shorter videos are much more engaging.	Invest heavily in pre-production lesson planning to segment videos into chunks shorter than 6 minutes.

Table 1. Summary of the main findings and video production recommendations that we present in this paper.

Aus: Guo, Philip J., Juho Kim, and Rob Rubin. "How video production affects student engagement: An empirical study of MOOC videos." *Proceedings of the first ACM conference on Learning@ scale conference*. ACM, 2014.

Gestaltungsprinzipien von Lernvideos (nach Merkt, 2015 sowie Guo et al. 2014)

Basierend auf wissenschaftlichen psychologischen Erkenntnissen

- Einfacher Zugriff auf Inhalte
- Optimierung der kognitiven Verarbeitung
- Personalisierung und Sichtbarkeit des Lehrenden
- **Kameraperspektive** in Lehr-Lern-Videos

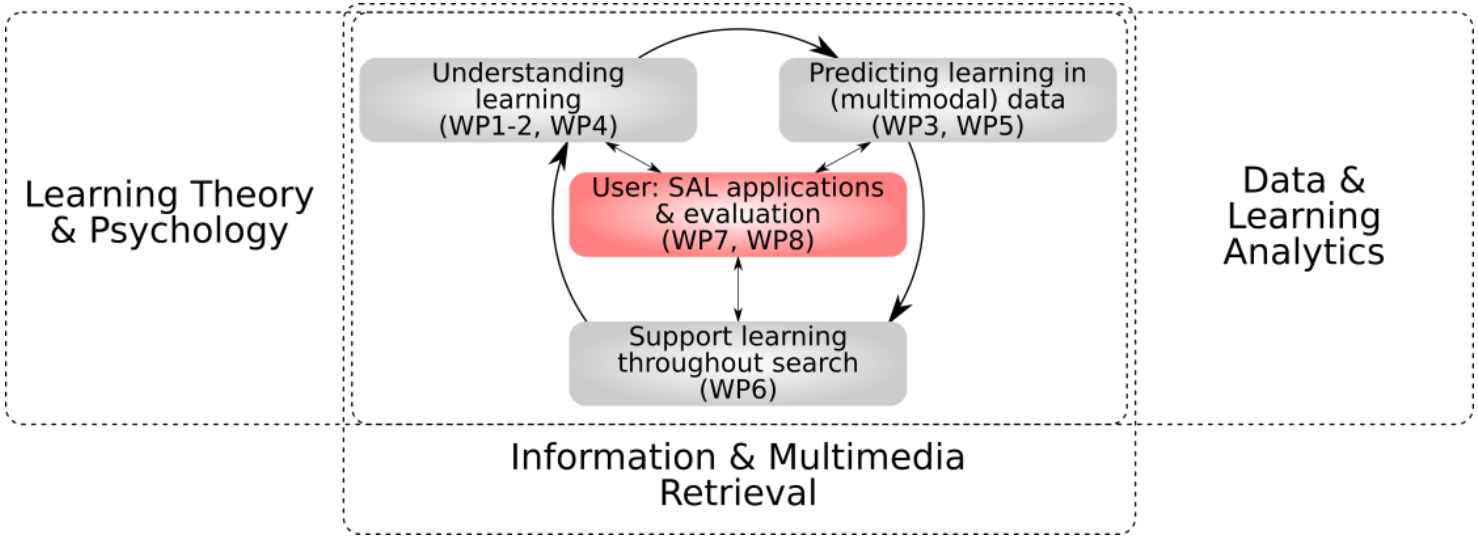
-> **Ich-Perspektive besser** für das Zeigen **prozeduraler Handlungen**

SALIENT –
Informelles Lernen im Web

SALIENT – Search as learning

Search as Learning – Investigating, Enhancing, and Predicting Learning during Multimodal Web Search

- Leibniz-Wettbewerb „Kooperative Exzellenz“ (2018 – 2021), Partner:
  - IWM Leibniz-Institut für Wissensmedien, Tübingen
  - GESIS Leibniz-Institut für Sozialwissenschaften, Köln
  - Forschungszentrum L3S, Hannover

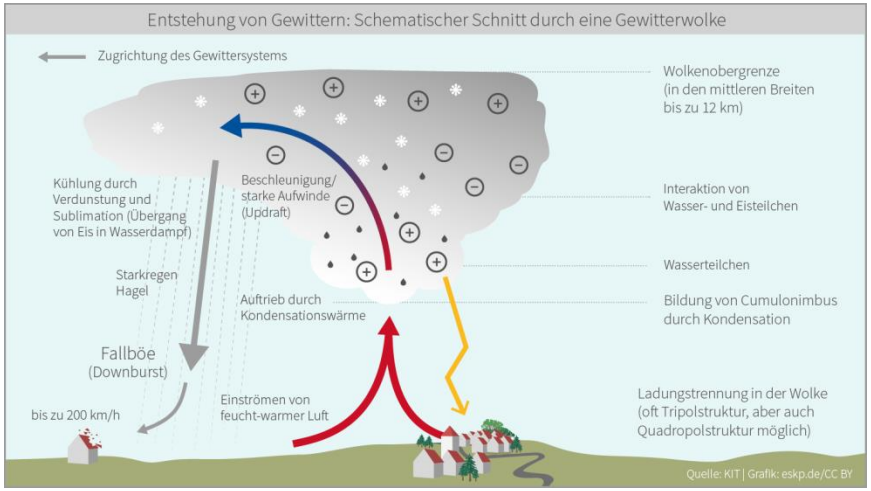


SALIENT: Erste Studien

Erste Studie (durchgeführt am IWM) mit 114 Teilnehmer/innen

Aufgabe: „Wie entstehen Blitz und Donner?“

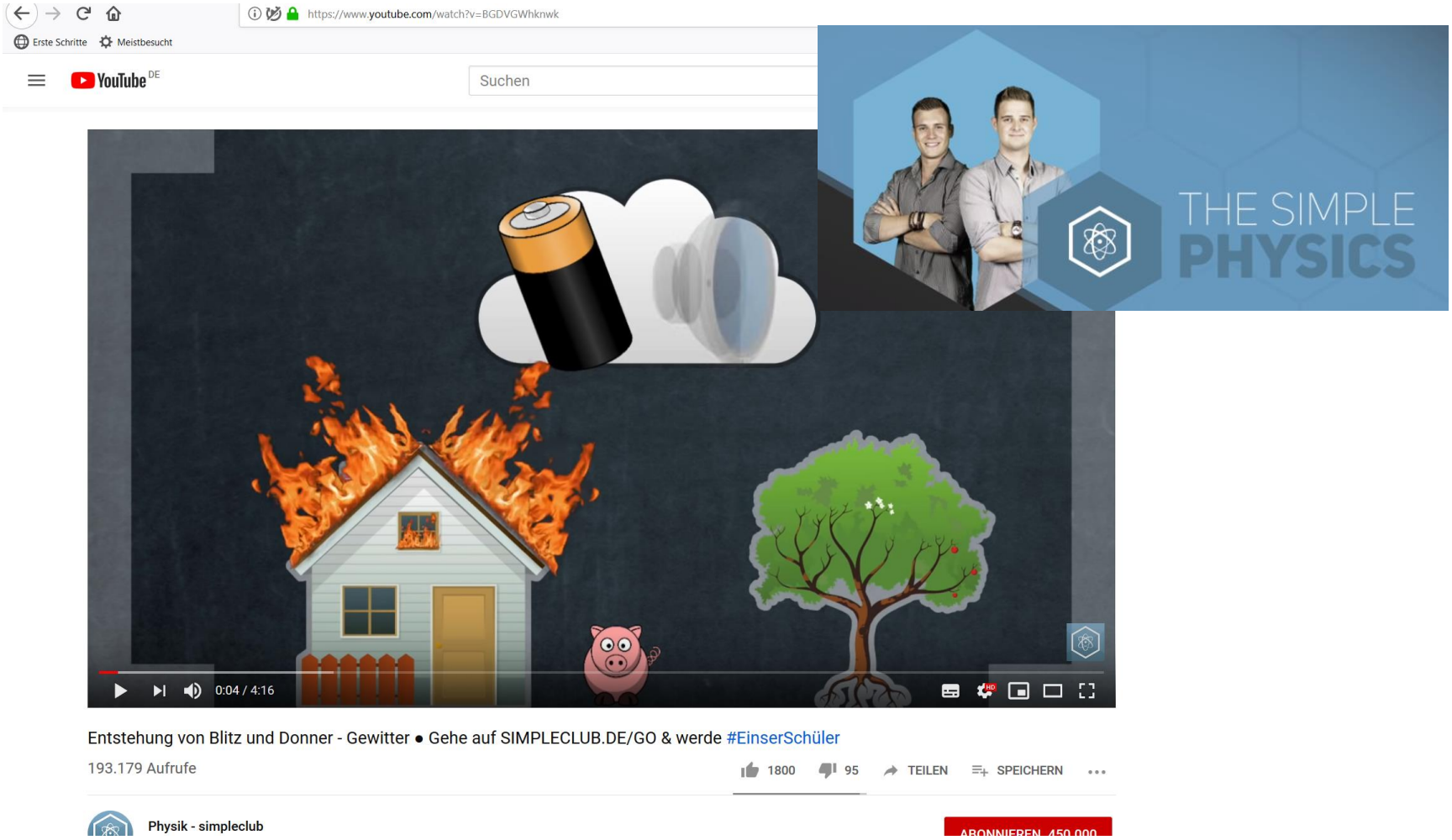
Relevante Konzepte:
Formation Gewitterwolke, Elektrische Aufladung, Blitz, etc.



Von Helmholtz-Wissensplattform "Erde und Umwelt", ESKP – https://www.eskp.de/charakteristik-und-entstehung-von-gewitterstuermen/, CC-BY 4.0, https://commons.wikimedia.org/w/index.php?curid=50096272

Eines der ausgewählten Videos

https://www.youtube.com/watch?v=BGDVGWhknwk



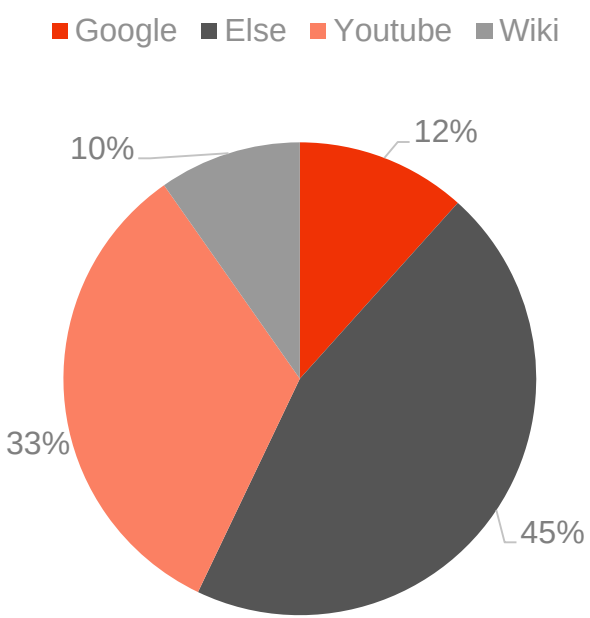
SALIENT: Erste allgemeine Ergebnisse

Erste Studie (durchgeführt am IWM)
Lernen mit dem Web: Wie entstehen Blitz und Donner?

Wo und wie lange suchten Teilnehmer*innen:

1. Übrige Seiten 12,3 min
2. YouTube: 8,5 min
3. Google: 2,5 min
4. Wikipedia: 2,3 min

Insgesamt durchschnittlich 25,6 min



SALIENT: „False-Certainty-Effekt“

Dunning Kruger Effekt (Kruger & Dunning, 1999)

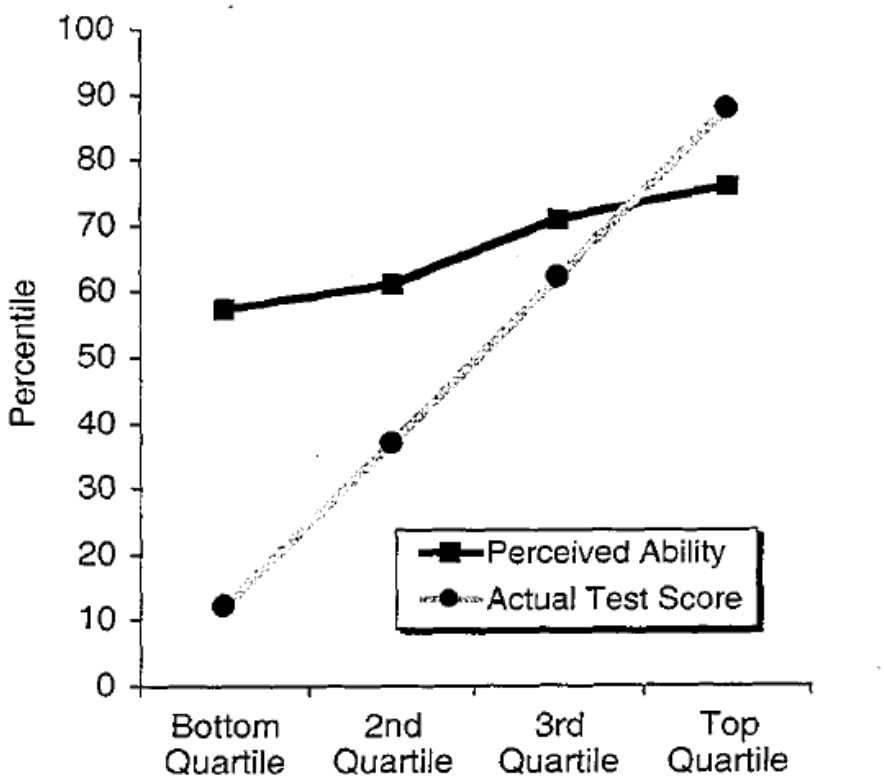
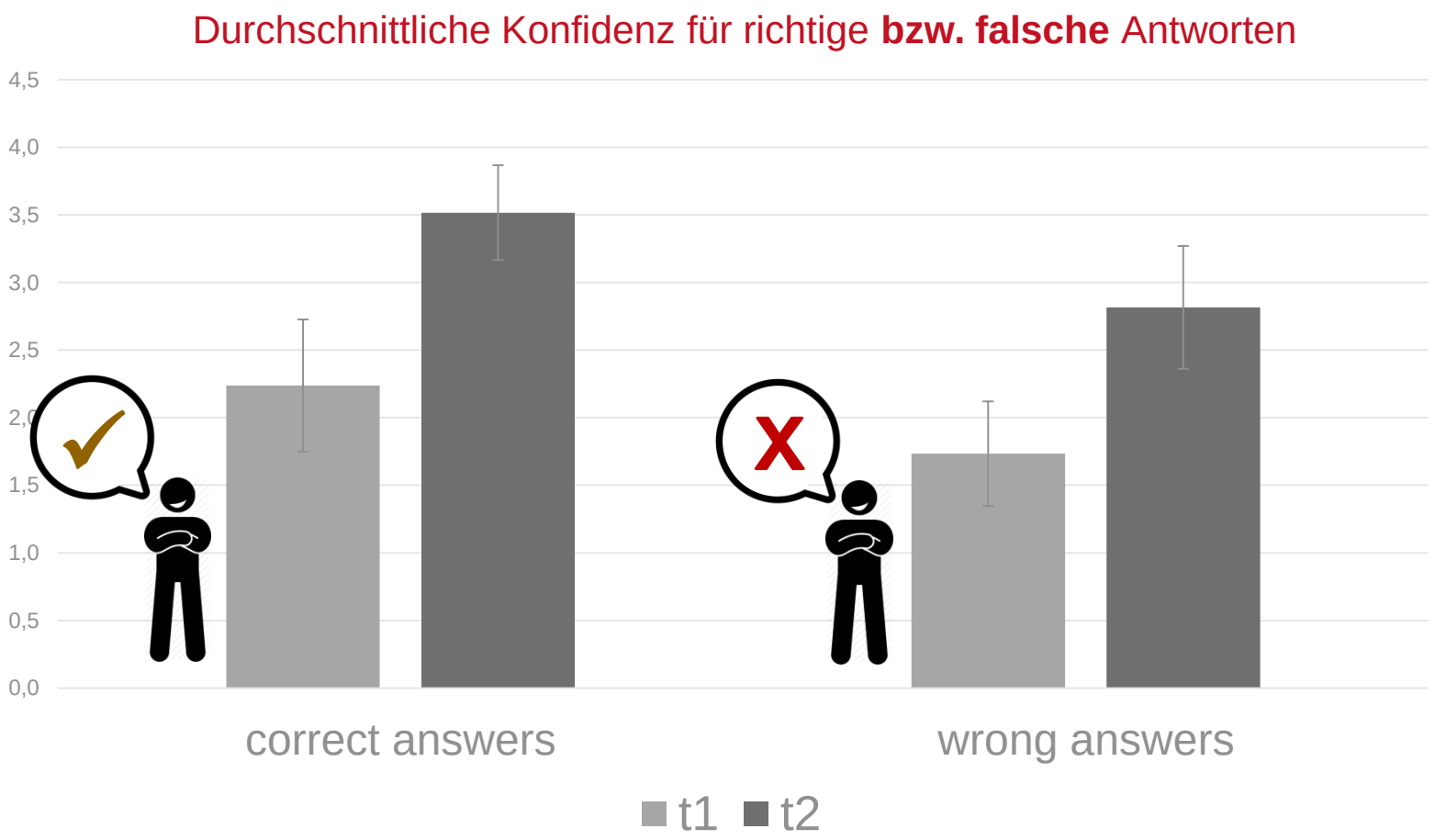


Figure 1. Perceived ability to recognize humor as a function of actual test performance (Study 1).

False-Certainty-Effekt?



SALIENT –
Cross-modale Bezüge

Semantische Bild-Text-Bezüge

Unkorreliert (cmi=0, sc=0,stat=0):

An elephant is walking through a swamp.

Ergänzend (cmi=1, sc=1,stat=0)

We couldnt wait to enjoy some delicious grilled food. The bands had begun and they were sounding great! There is no way i was going to be going on this ride. People were cooling off by jumping right in the lake!

Interdependent (cmi=0, sc=1,stat=0):

Gravity will never be the same.

Illustration (cmi=1, sc=1,stat=T):

A turnbuckle, stretching screw or bottlescrew is a device for adjusting the tension or length of ropes, cables, tie rods, and other tensioning systems. It normally consists of two threaded eye bolts, one screwed into each end of a small metal frame, one with a left-hand thread and the other with a right-hand thread.

Otto / Springstein / Anand / Ewerth @ ACM Int'l Conference on Multimedia Retrieval 2019 (ICMR)

- ***Best Paper Award*** **-**

(Quellen Bild-Text-Paare: siehe Referenzen [17, 18, 26, 39] im Paper, MSCOCO, VIST, ImageNet, Wikipedia, Hussain et al. 2017)

Semantische Bild-Text-Bezüge

Anker (cmi=1, sc=1,stat=Bild)

A man in dress uniform talks with a woman wearing camouflage.

Kontrast:(cmi=1, sc=-1,stat=0)

Here were so **few** people walking and talking on their phones. It was fun to see the **brown** costumes that patrons were wearing. My sister and his **enemy** marched in the parade. Me and my brother wore funky **purple** hats in honor of the irish holiday.

Illustration (widersprüchlich): (cmi=1, sc=-1,stat=T)

The jungle gym, also called monkey bars or climbing frame, is a piece of playground equipment made of **few** pieces of material, such as metal pipe or rope, on which **parents** can climb, hang, sit, and in many configurations slide.

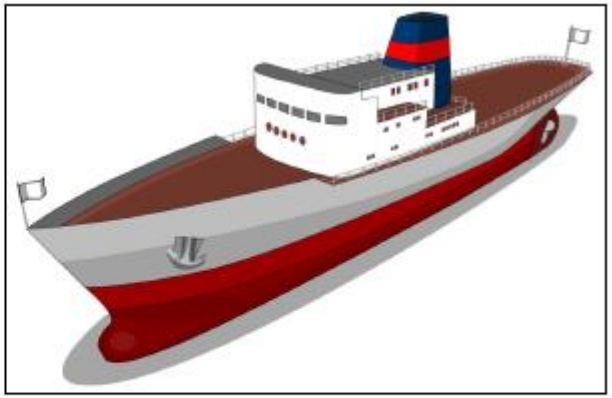
Anker (widersprüchlich): (cmi=1, sc=-1,stat=B)

A **woman** is posing with two people dressed **down** as stormtroopers.

Otto / Springstein / Anand / Ewerth @ ACM Int'l Conference on Multimedia Retrieval 2019 (ICMR)
- ***Best Paper Award*** **-**
(Quellen Bild-Text-Paare: siehe Referenzen [17, 18, 26, 39] im Paper, MSCOCO, VIST, ImageNet, Wikipedia, Hussain et al. 2017)

Differenzierter Blick auf Illustrationen

Weitere neue Metrik: **Relativer Abstraktionsgrad**



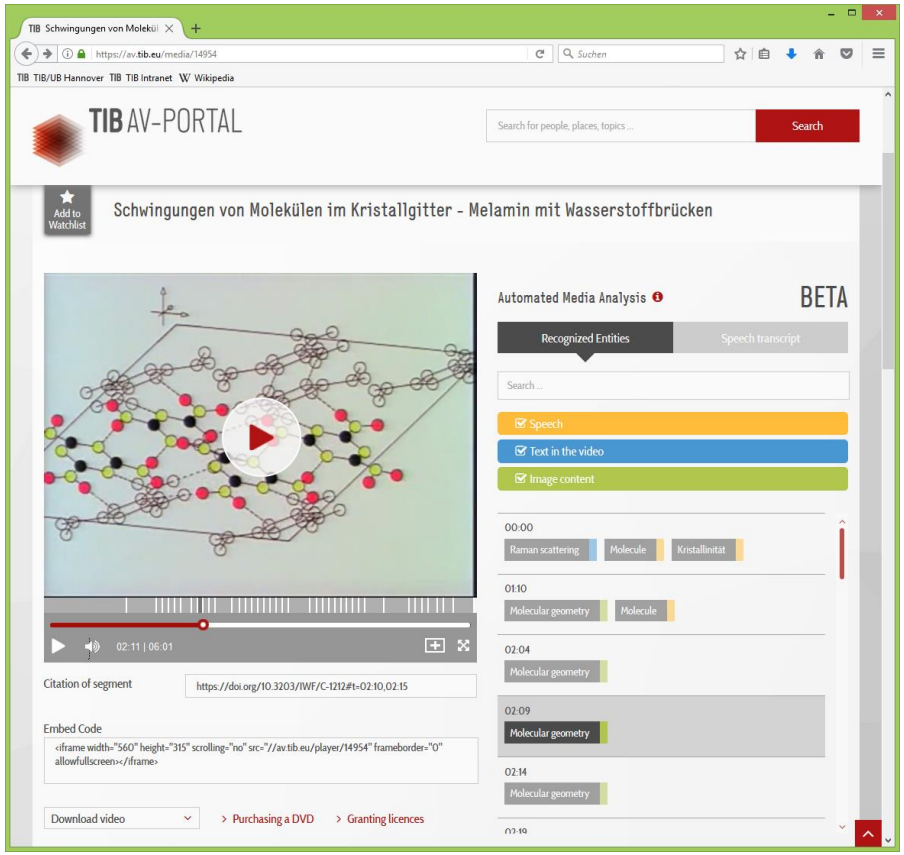
A ship is a large watercraft that travels the world's oceans and other sufficiently deep waterways, carrying passengers or goods, or in support of specialized missions, such as defense, research and fishing.

TIB AV Portal https://av.tib.eu

Semantische Suche im TIB AV-Portal

Automatische Videoanalyse
- Spracherkennung
- Video OCR
- Visual concept detection (VCD)

Examples for VCD:
- 12533 (Agricultural machinery)
- 14956 (Chemical experiment)
- 14954 (Molecular geometry)
- 12793 (Rocket)
- 12858 (Satellite)
- 11539 (Water transport)
- 12775 (Motor vehicle)
- 16154 (Indoor)



Liste von Konzepten

Universe, Astronomical object, Star, Galaxy, Gas giant, Nebula, Terrestrial planet, Asteroid, Black hole, Aerospace engineering, Spacecraft, Rocket, Artificial satellite, Interior space, Facade, Building, City, Bridge, Power station, Hydraulic engineering, Conceptual model, Software, Source code, JSON, XML, UML, Plant, Bacteria, Virus, Amphibian, Arachnida, Bird, Crustacea, Fish, Insect, Mammal, Mollusca, Reptile, Chemical compound, Chemical experiment, Atom, Atomic_model, Molecular geometry, Elementary particle, Laser, Magnet, Physics experiment, Particle, Pendulum, Visualization, Computer animation, Engineering drawing, Diagram, Drawing, Table, Program flowchart, Map, Lecture/Conference, Meeting/Interview, Panel painting, Aerial photography, Satellite imagery, Machine, Motor vehicle, Aircraft engineering, Helicopter, Railroad, Agricultural machinery, Water transport, Electronic component, Transformer, Computer, Turbine

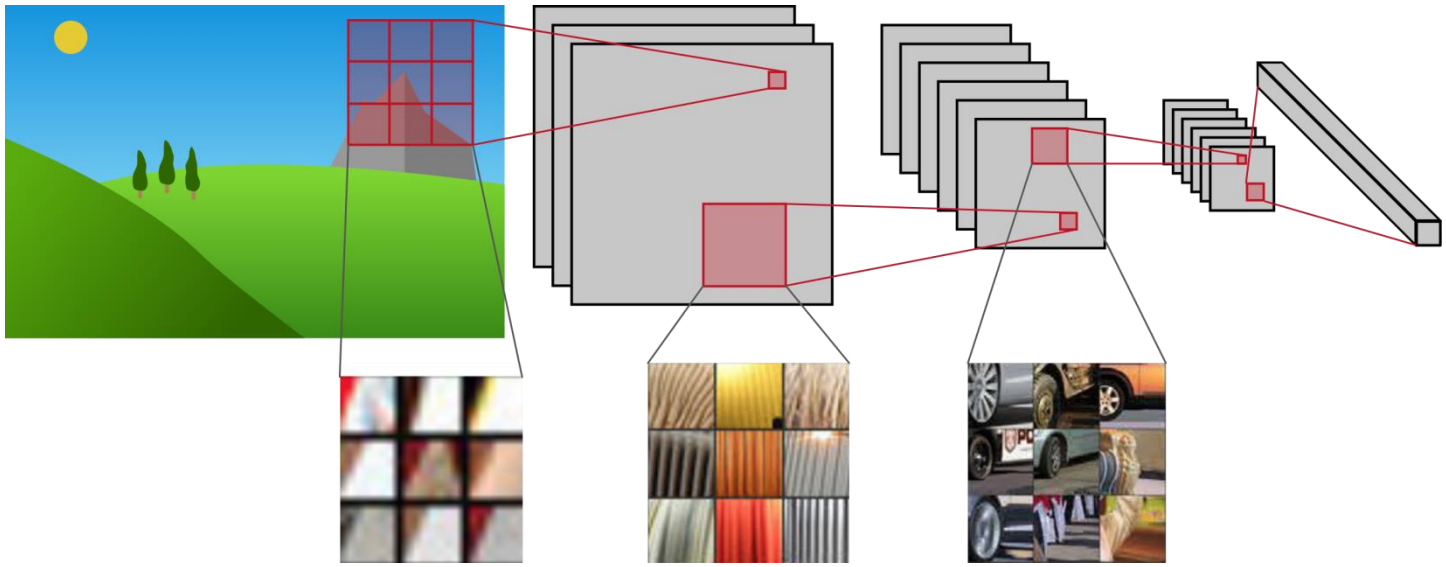
Liste von Konzepten

Universe, Astronomical object, Star, Galaxy, Gas giant, Nebula, Terrestrial planet, **Asteroid**, Black hole, Aerospace engineering, **Spacecraft**, Rocket, Artificial satellite, Interior space, Facade, **Building**, City, **Bridge**, Power station, Hydraulic engineering, Conceptual model, Software, **Source code**, JSON, **XML**, **UML**, Plant, Bacteria, Virus, Amphibian, Arachnida, Bird, Crustacea, Fish, Insect, Mammal, Mollusca, Reptile, Chemical compound, **Chemical experiment**, Atom, Atomic_model, Molecular geometry, Elementary particle, Laser, Magnet, Physics experiment, Particle, Pendulum, Visualization, Computer animation, Engineering drawing, **Diagram**, **Drawing**, Table, Program flowchart, **Map**, **Lecture/Conference**, Meeting/Interview, Panel painting, Aerial photography, Satellite imagery, Machine, Motor vehicle, Aircraft engineering, Helicopter, Railroad, Agricultural machinery, **Water transport**, Electronic component, Transformer, **Computer**, Turbine

Von der Forschung zum Betrieb: Konzeptdetektion im AV-Portal

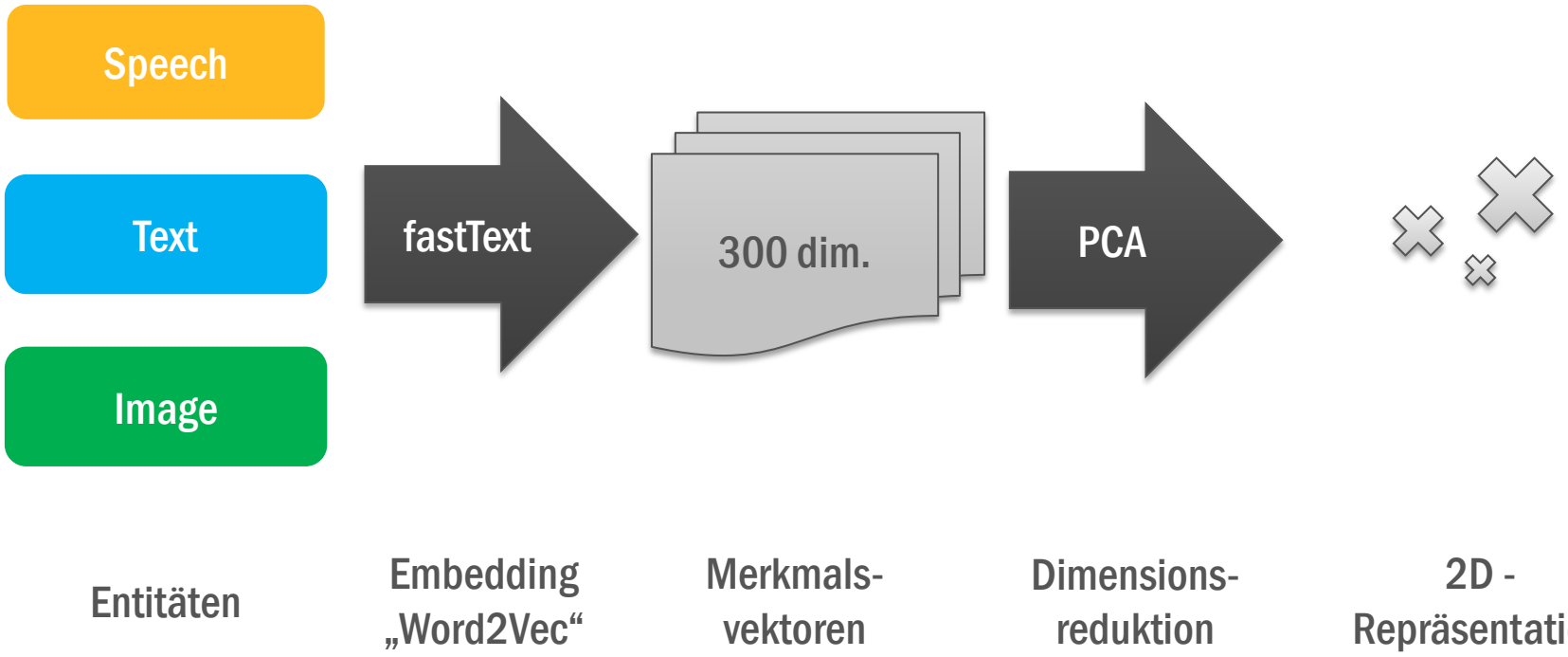
Neue Konzeptdetektion mit Web-überwachtem Lernen realisiert

- 73 visuelle Konzepte für TIB-Fächer
- 50.000 Bilder mit Googles Bildersuche als Trainingsdaten gesammelt
- Inception-Resnet-v2 Netwerkarchitektur (Google)
- Vortrainiert mit 1 Million Bilder

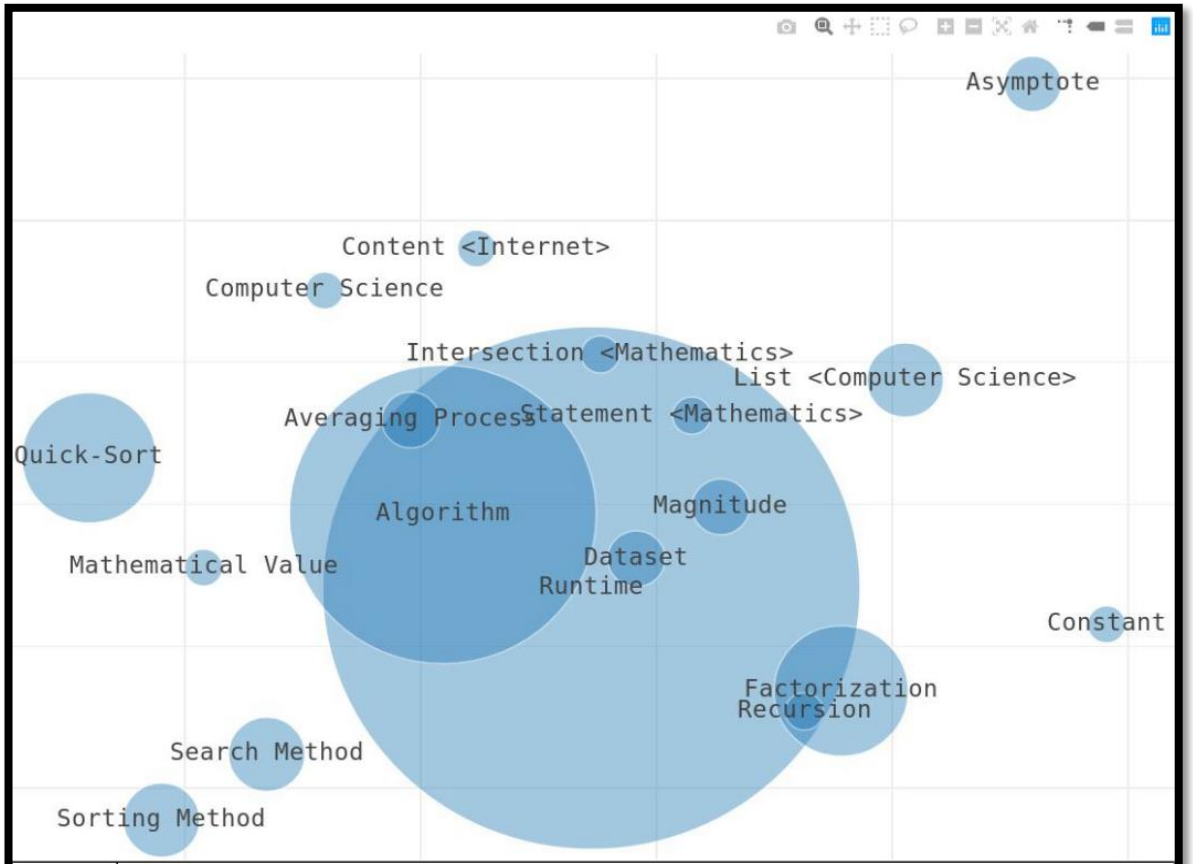


Visuelle Inhaltsverzeichnisse für Videos

Erstellung einer 2D-Repräsentation der wichtigsten Entitäten im Video



Visuelle Inhaltsverzeichnisse für Videos



Time	Content
00:00	corner algorithm, search
00:20	comparing weight next, important rights
08:19	chosen, careful half
09:26	runtime full factor, algorithm

https://av.tib.eu/media/9557

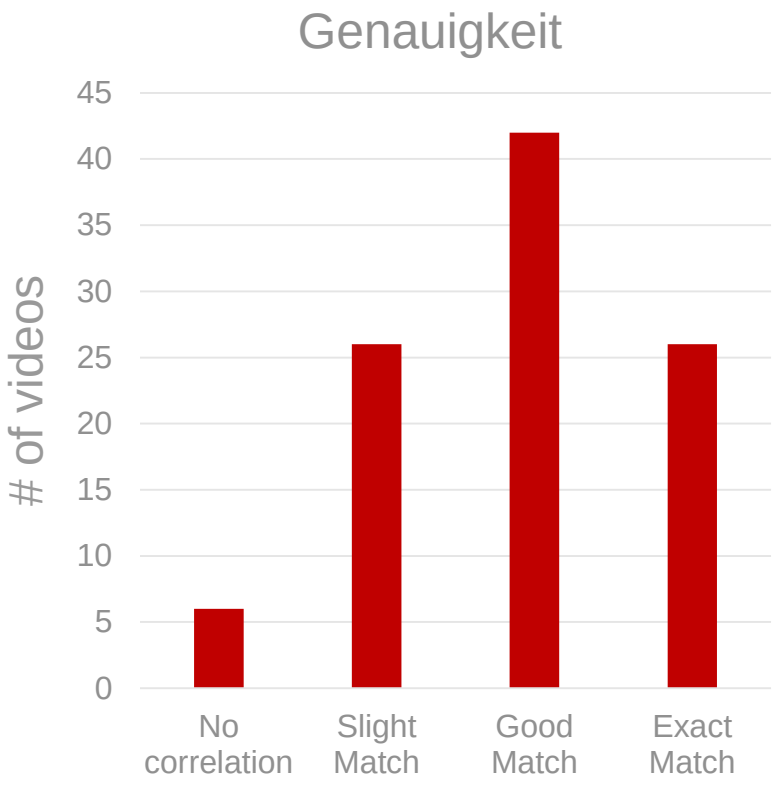
Visualization: *Plot.ly* API

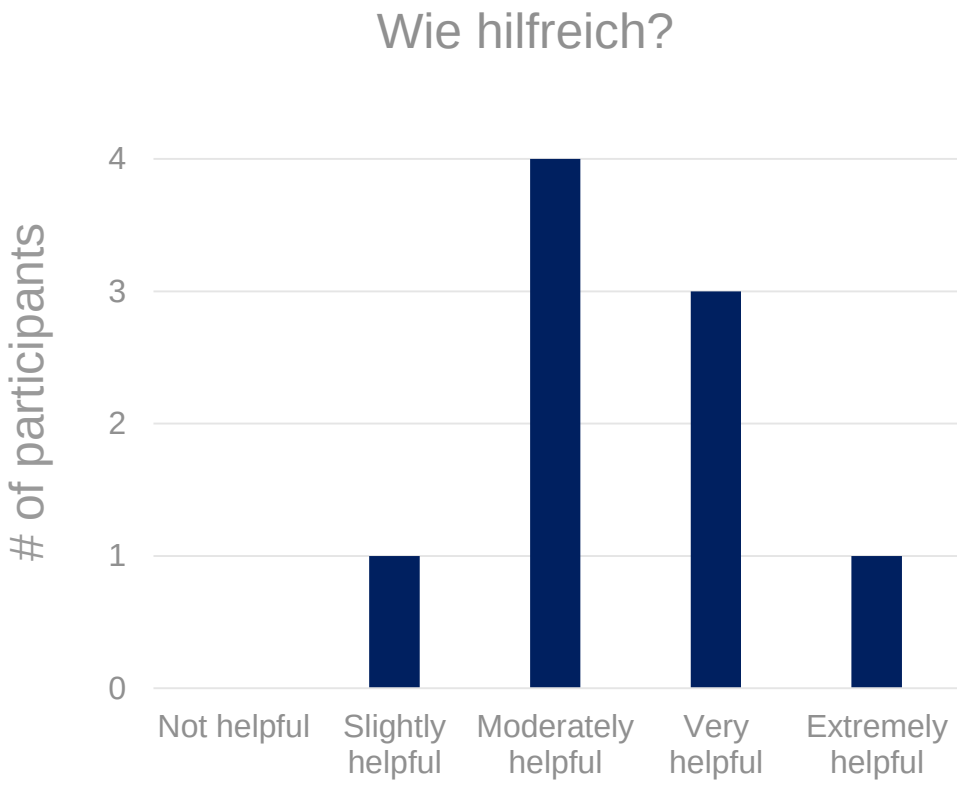
Size of Bubble: entity frequency

Distance: semantic similarity

Table: keyphrases by segment

Experimentelle Ergebnisse





→ 68% der Visualisierungen sind eine gute (oder besser) Inhaltsbeschreibung
→ 6% korrelieren nicht

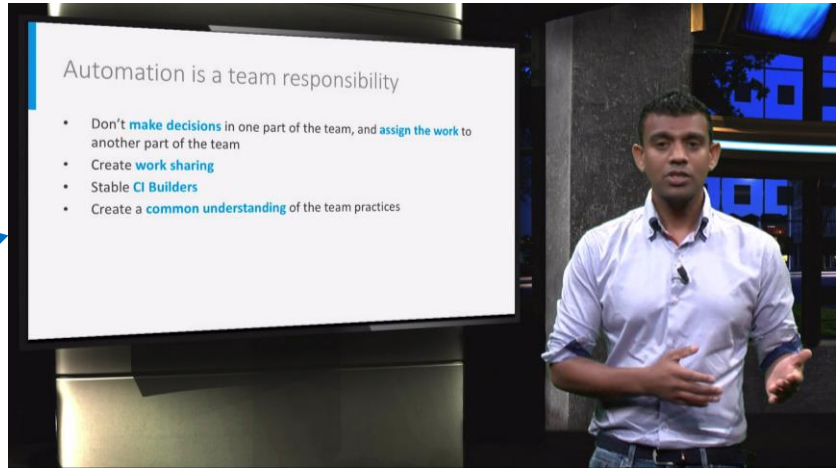
SALIENT –
Qualitätsbewertung von Lehrvideos

Qualitätsbewertung von Lehrvideos

Lehreinheit „Softwareengineering“:
23 Lektionen

Zu jeder Lektion

- Video
- Folien
- Transkripte

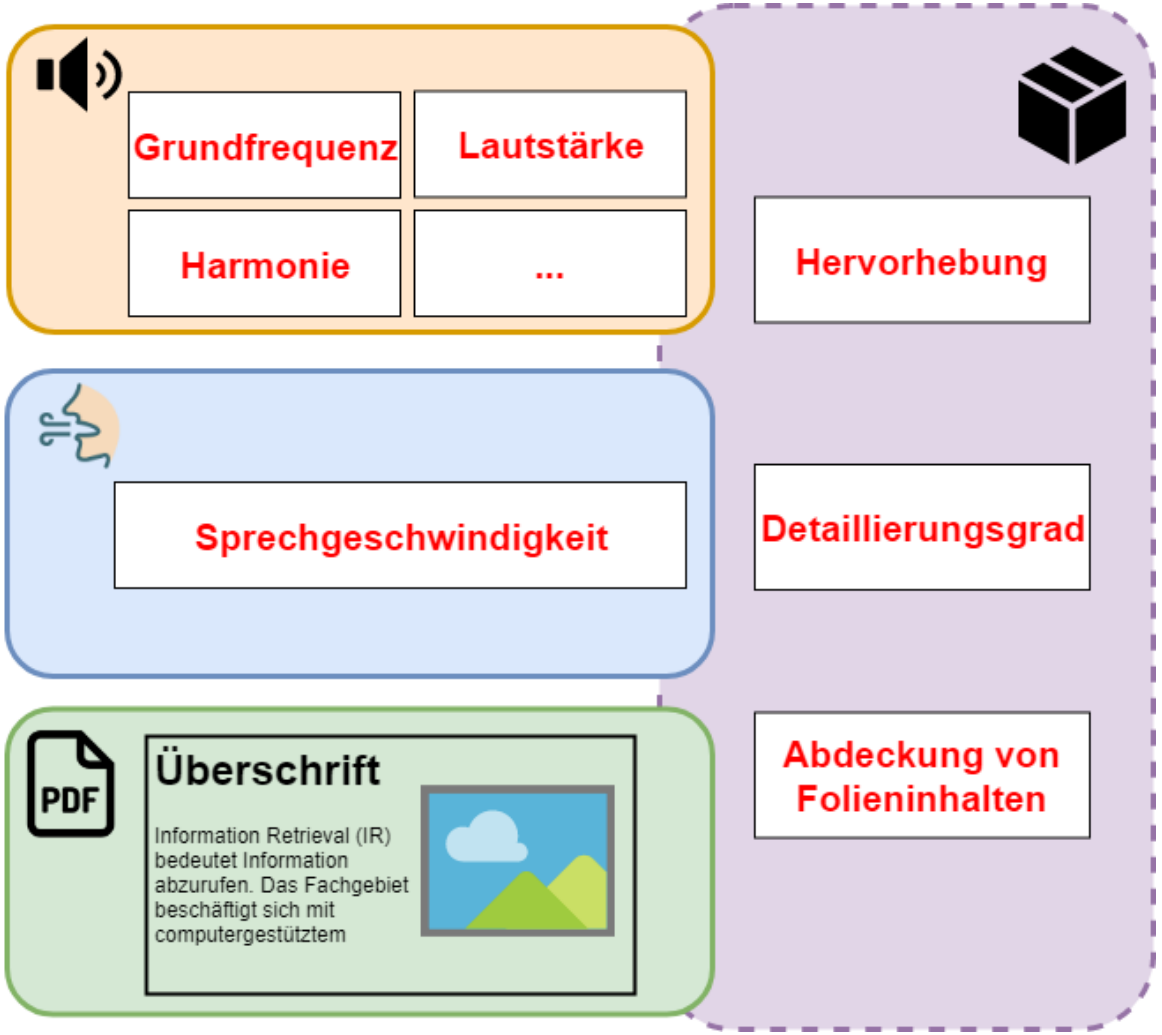


Automation is a team responsibility

- Don't **make decisions** in one part of the team, and **assign the work** to another part of the team
- Create **work sharing**
- Stable **CI Builders**
- Create a **common understanding** of the team practices

Zeitinterval (Δt = einige Sek.)
„Another important thing you should not do, is to make decisions in one part of the team“

Qualitätsbewertung von Videos



Korelationsanalyse

Automatisch extrahierte Merkmale	Manuell annotierte Merkmale	Korrelation	Kommentar
Grundfrequenz	Klare Sprache	0,60	Repräsentation der klaren Sprache in verschiedenen Aspekte
Lautstärke		0,43	
Harmonie		0,43	
Artikulationsrate	Füllwörter	0,46	• Anzahl Füllwörter • Anzahl gesprochener Silben
Varianz von Detailierungsgrad	Detailierung	0,38	Repräsentation der modalitätsübergreifenden Wahrnehmung

Zusammenfassung

Forschungsschwerpunkte der Gruppe Visual Analytics

Informelles Lernen mit Videos im Web

Forschungsprojekt **SALIENT**

Erste Ergebnisse

- Psychologische Perspektive
- Cross-modale Bild-Text-Bezüge
- Visuelle Inhaltsverzeichnisse von Videos
- Qualitätsbewertung von Videos

Berichterstattung über Forschungsgruppe

Geolocation – Schätzung des Aufnahmeorts

- Bericht in Heise Magazine | c't (https://www.heise.de/ct/)
- Exponat auf **MS Wissenschaft** 2019 ☺

MS Wissenschaft – das schwimmende Science Center (https://ms-wissenschaft.de)

Literatur

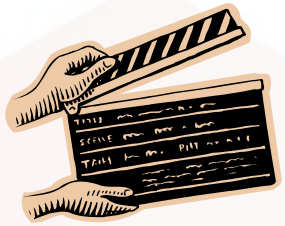
1. J. P. Guo, J. Kim, & R. Rubin (2014). “How video production affects student engagement: An empirical study of MOOC videos.”. In *Proceedings of the first ACM conference on Learning@ scale conference*. ACM, 2014, 41-50.
2. J. Kruger & D. Dunning (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, *77*(6), 1121.
3. M. Lehner (2018). *Erklären und Verstehen*. UTB GmbH.
4. M. Lehner (2019). *Didaktik*. UTB GmbH.
5. R. Mayer & R. E. Mayer (Eds.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge university press.
6. M. Merkt, S. Weigand, A. Heier, & S. Schwan (2011). Learning with videos vs. learning with print: The role of interactive features. *Learning and Instruction*, *21*(6), 687-704.
7. M. Merkt & S. Schwan (2014). “Training the use of interactive videos: effects on mastering different tasks.” In: *Instructional Science* 42.3 (2014a): 421-441.
8. M. Merkt & S. Schwan (2014). “How does interactivity in videos affect task performance?.” In: *Computers in Human Behavior* 31 (2014b): 172-181.
9. M. Merkt: 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, 2015

Literatur

1. C. Otto, S. Holzki, & R. Ewerth: “Is this an example image?” – Predicting the Relative Abstractness Level of Image and Text. In: Proceedings of the 41st European Conference on Information Retrieval (ECIR), Cologne, Germany, 2019, 711-725. https://arxiv.org/abs/1901.07878
2. W. Schnotz (2011). *Pädagogische Psychologie kompakt*. Weinheim: Beltz, PVU, 2. Auflage.
3. J. Shi, C. Otto, A. Hoppe, P. Holtz, & R. Ewerth (2019). Investigating Correlations of Automatically Extracted Multimodal Features and Lecture Video Quality. In: Proceedings of the 1st International Workshop on Search as Learning with Multimedia Information @ACM Multimedia Conference, Nice, France, 2019, 11-19. https://dl.acm.org/citation.cfm?doid=3347451.3356731
4. J. Sweller & P. Chandler (1991). Evidence for cognitive load theory. In: *Cognition and instruction* 8.4 (1991): 351-362.
5. J. von Hoyer, G. Pardi, Y. Kammer, & P. 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*. ACM, 2019.
6. H. Zhou, C. Otto, & R. Ewerth (2019). Visual Summarization of Scholarly Videos using Word Embeddings and Keyphrase Extraction. In: Proc. of *International Conference on Theory and Practice of Digital Libraries (TPDL)*, Oslo, Norway, 2019, 327-335.

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB



Vielen Dank für Ihre Aufmerksamkeit!

Kontakt:

Mail: ralph.ewerth@tib.eu
Telefon: +49 (511) 762-19651

Mitglied der Leibniz-Gemeinschaft