Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Machine Learning und die Bibliothek – Chancen und Grenzen

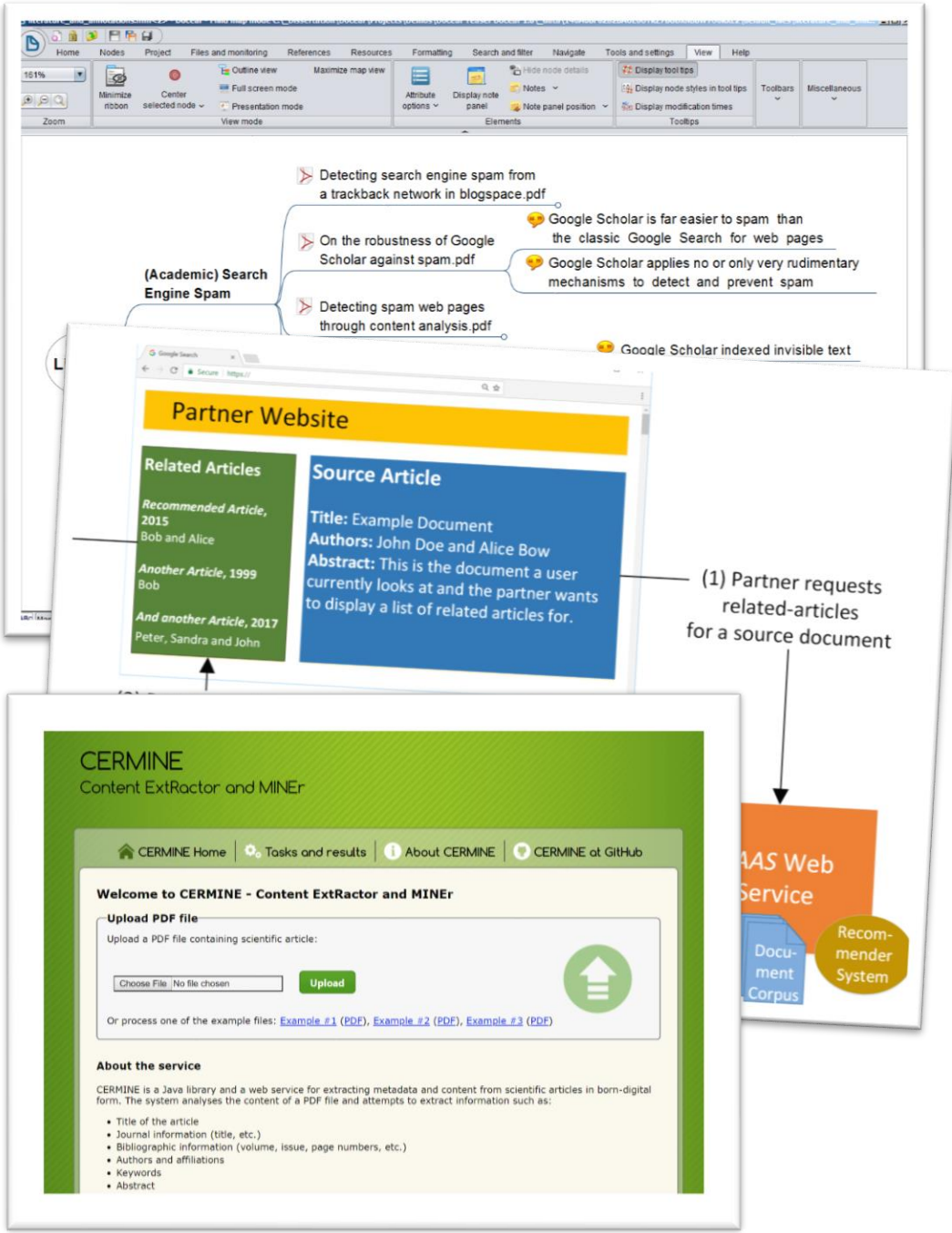
18. BVB-Verbundkonferenz, Weiden
20. September 2018

Dr. Joeran Beel
Professor in Intelligent Systems, Trinity College Dublin

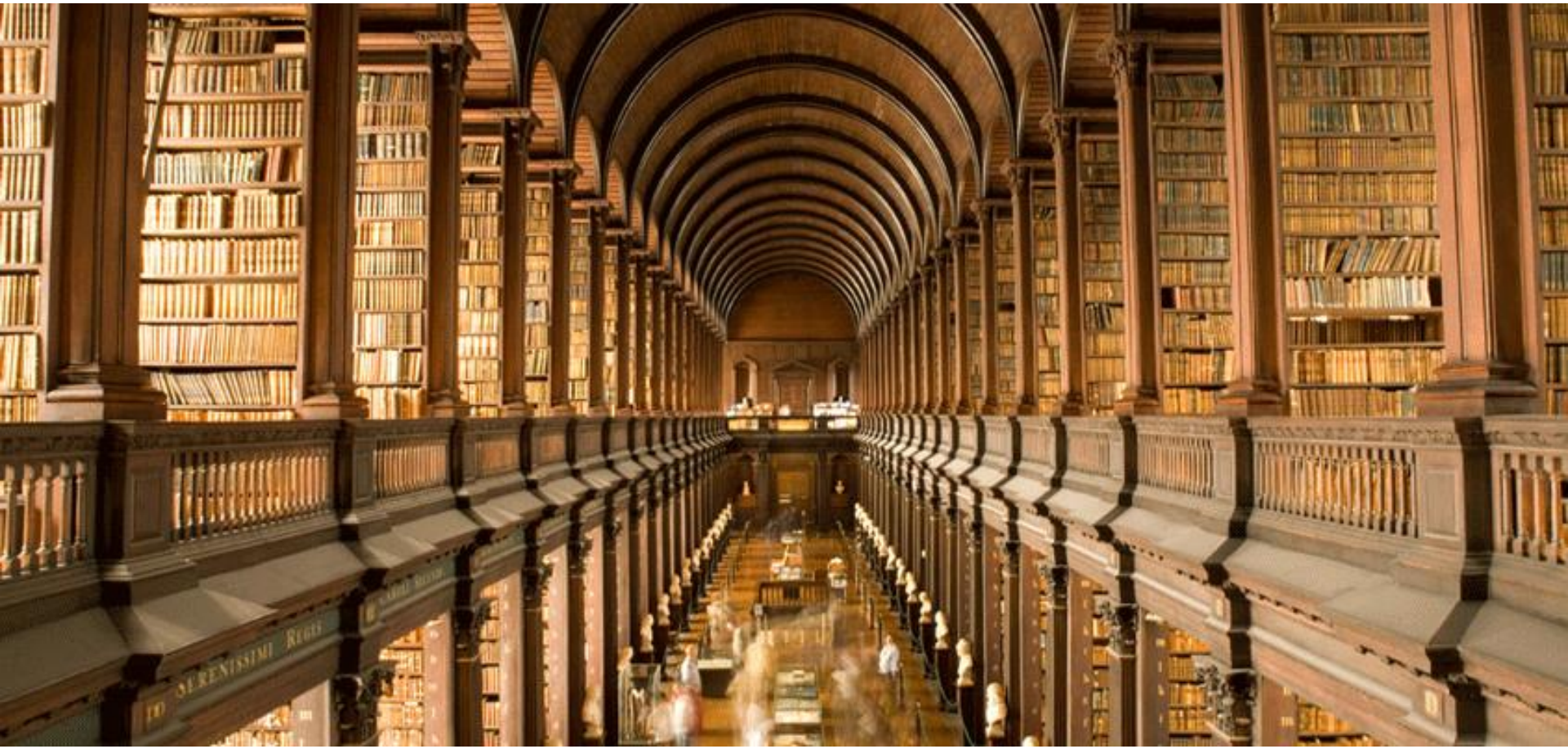
beelj@tcd.ie

Joeran Beel

- **Docear http://docear.org**
- **Mr. DLib http://mr-dlib.org**
- **(Cermine) http://cermine.ceon.pl/index.html**
- **Titel & Referenzextraktion**
- **Akademische Suchmaschinen Optimierung & Spam**

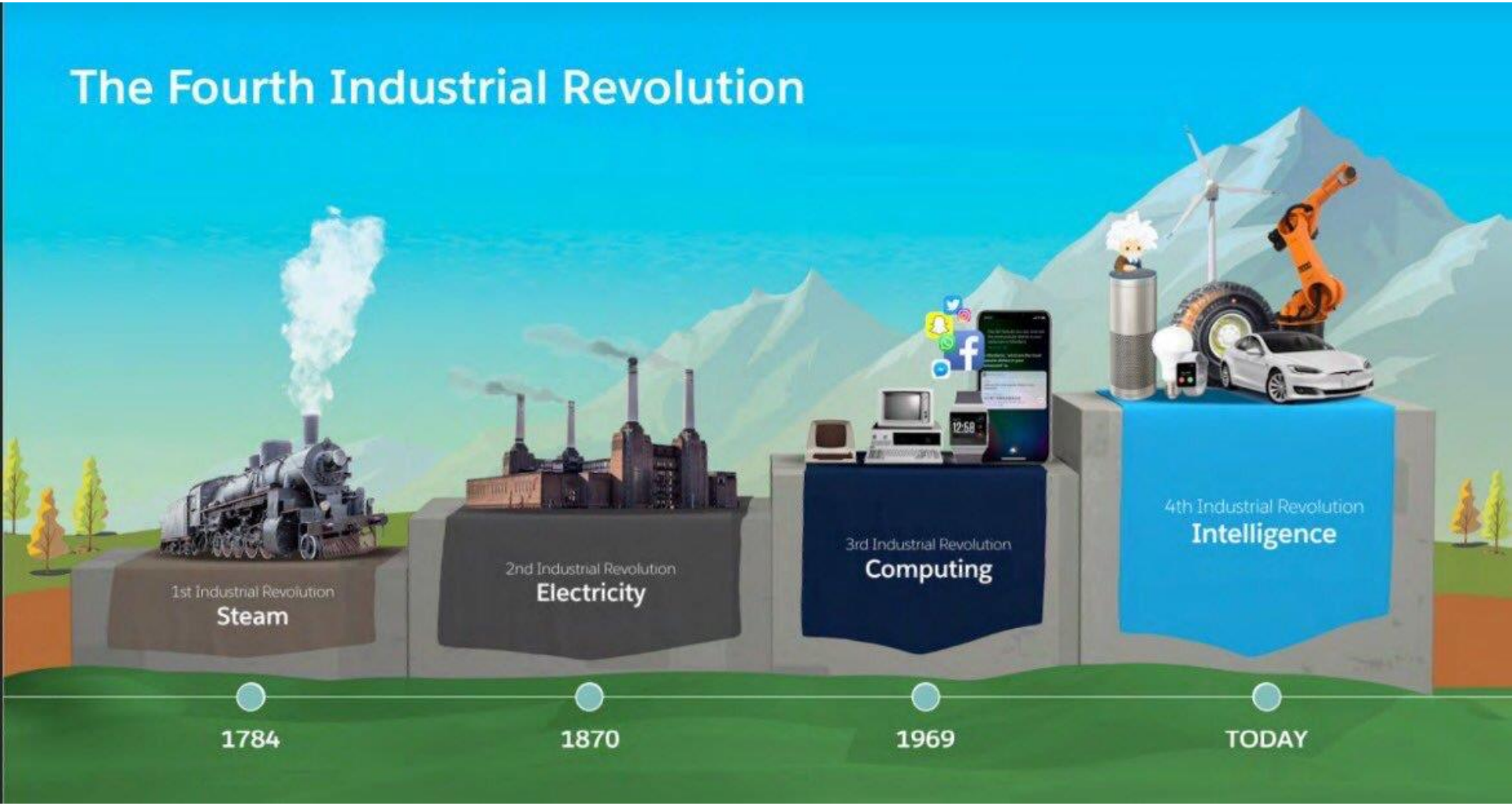


Trinity College Dublin, Bibliothek



https://www.tcd.ie/Philosophy/assets/images/homepage/philosophy-slider-library.gif

Meilensteine der Industrie



https://pbs.twimg.com/media/DORVjvLWsAAYJIx.jpg

Restaurant Reservierung

Wer ist die Maschine?

https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html

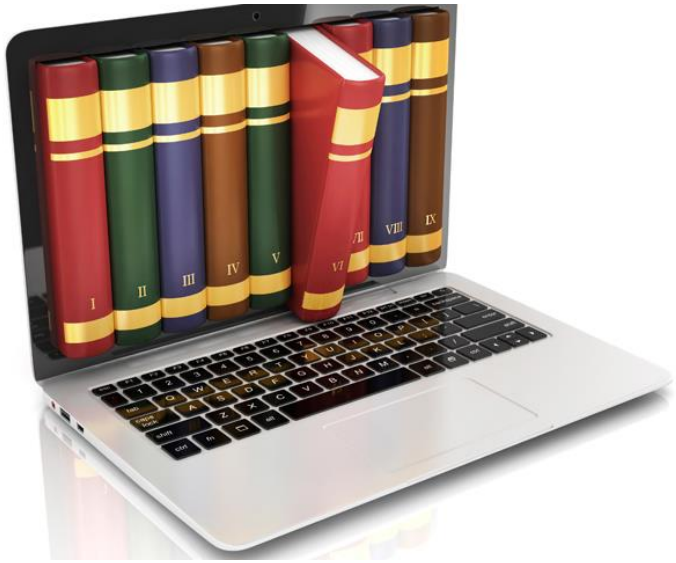
Obama

- **https://www.youtube.com/watch?v=cQ54GDm1eL0**

Meilensteine der Bibliotheken

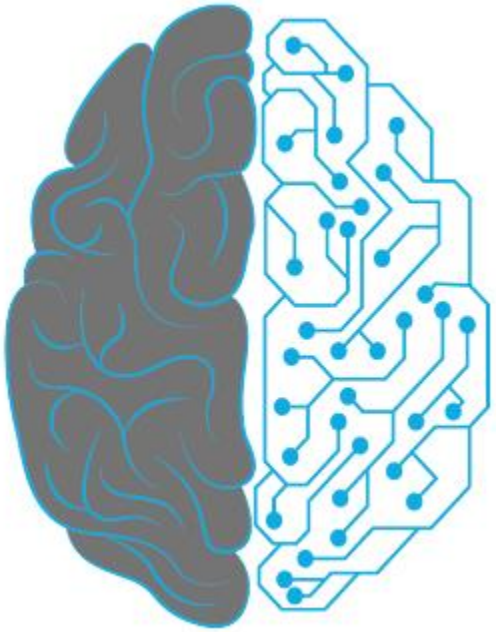
Buchdruck

https://de.wikipedia.org/wiki/Buchdruck#/media/File:Buchdruck-15-jahrhundert_1.jpg



Digitalisierung

https://www.uni-trier.de/fileadmin/organisation/Presse/Unijournal/Forschung_aktuell/Digitalisierung_koya_Fotolia.jpg



Machine Learning / Artificial Intelligence

https://www.wogra.com/wp-content/uploads/2018/04/praxisbeispiele_fuer_kuenstliche_intelligenz_500x333.png

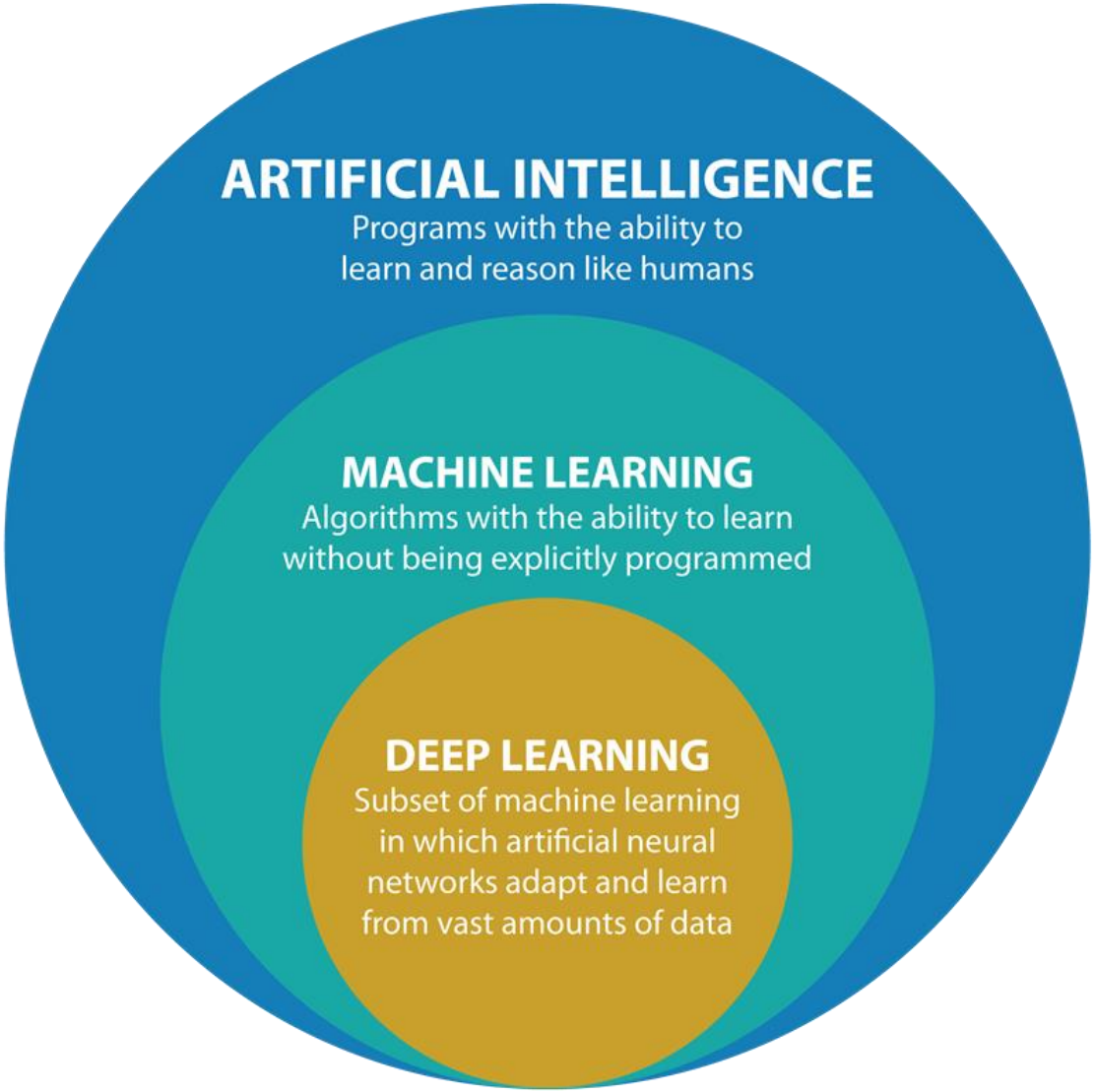
Gliederung

1. **Einführung in Machine Learning**
2. **Anwendungsbeispiele in der (Digitalen) Bibliothek**
3. **Machine Learning als Service**
4. **Aktuelle Qualität von Machine Learning**
5. **Herausforderungen**
6. **Diskussion**

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

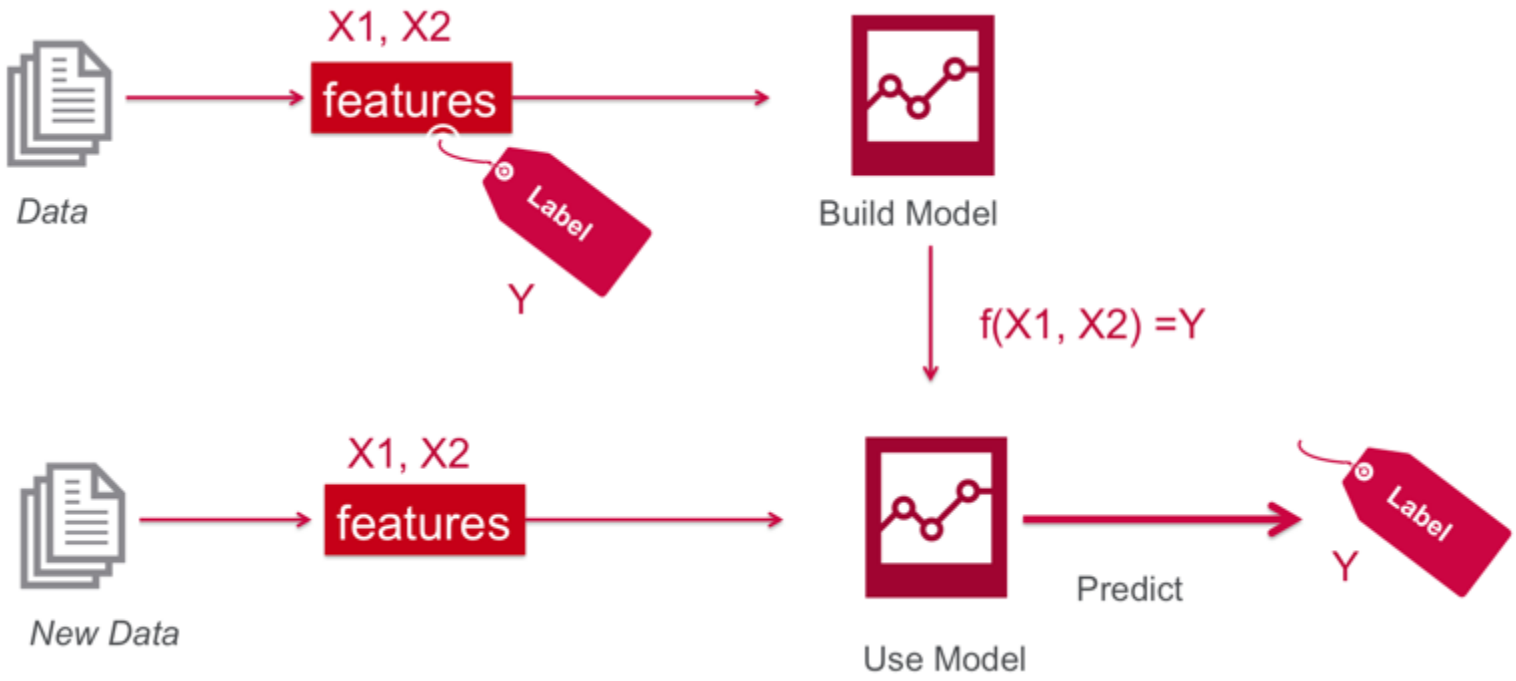
1. Einführung in Machine Learning

ML vs. AI vs. DL



https://www.argility.com/wp-content/uploads/2018/04/image10.png

Typisches Vorgehen



https://mapr.com/blog/demystifying-ai-ml-dl/assets/sl.png

Beispiel

- Finde eine Funktion y=f(x)= a + bX
- Berechne y, gegeben x
- z.B. f(x)= 10 + 0.7x
- f(90)= 5 + 1.5*90 = 140

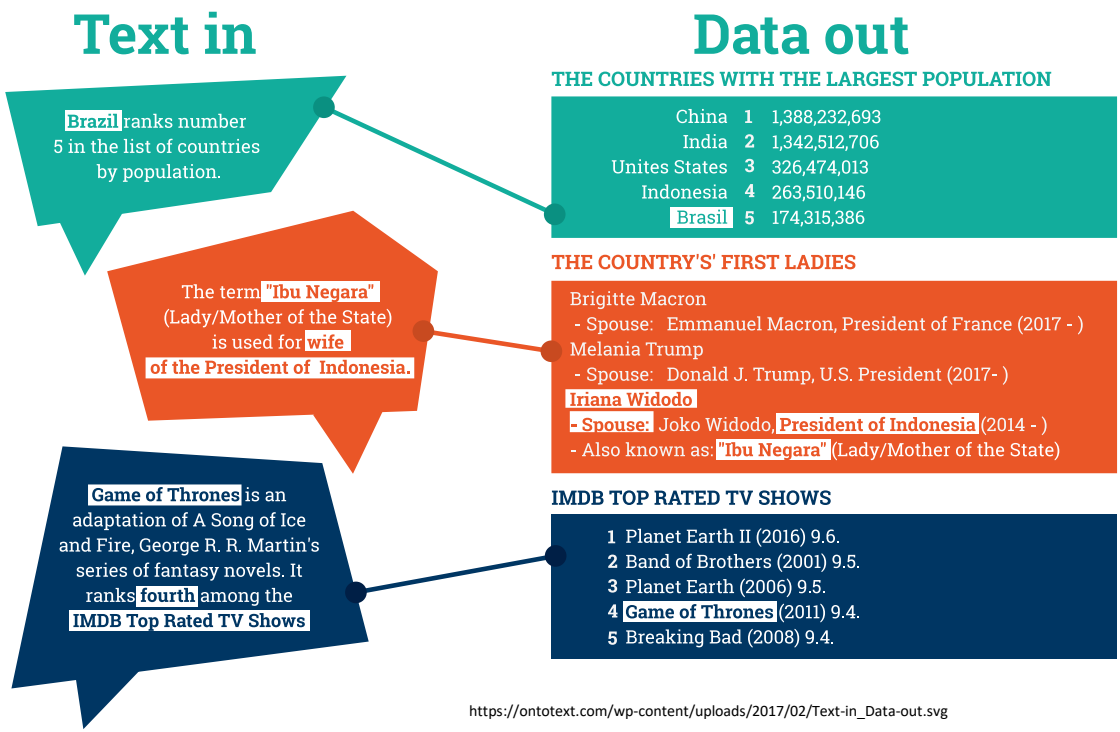
Loss Function / Optimization

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

2. Anwendungsbeispiele in der (Digitalen) Bibliothek

Informationsextraktion

- **OCR**
- **Metadaten (Titel, Autor, Affiliation …)**
- **Referenzen (Titel, Autor, Affiliation …)**
- **Entity Recognotion**
- **« Dark Data »**
  - Abbildungen
  - Tabellen
  - Formeln



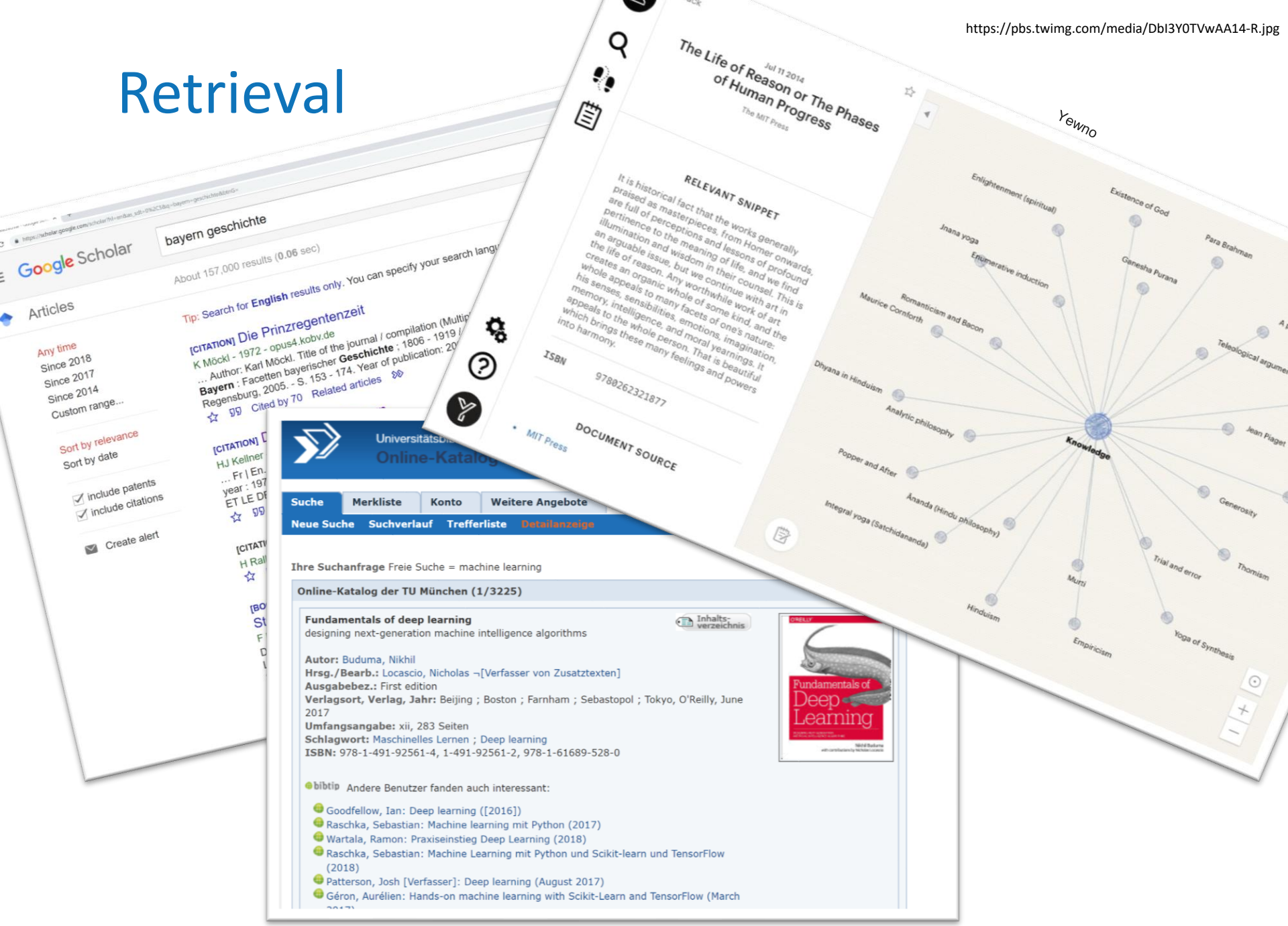
https://ontotext.com/wp-content/uploads/2017/02/Text-in_Data-out.svg

Natural Language Processing

- **Verschlagwortung**
- **Kategorisierung**
- **Zusammenfassungen**
- **Übersetzungen**
- **Schreibstil**
- **Disambiguation/ De-Duplication (Autoren, Dokumente, …)**

https://pbs.twimg.com/media/DbI3Y0TVwAA14-R.jpg

Retrieval



User Experience

https://s3.amazonaws.com/ArchiveImages/LJ/2016/08/Singapore_Robot_Checks_Shelves-550x366.jpg

128 FACHBEITRÄGE

Dierolf | Skarupianski

„Alexa, frage KIT-Bibliothek wo ich lernen kann“

Einsatz von digitalen Sprachassistenten im Umfeld von Bibliotheken und Erweiterung um eigene Dienste

we Dierolf und Michael Skarupianski

ma sprachgesteuerte persönliche Assistenten ist im wahrsten Sinne des aller Munde. Die KIT-Bibliothek hat daher Ende 2017 entschieden, sich Thema unter technischen Gesichtspunkten exemplarisch am Beispiel ka zu beschäftigen. Dazu wurde Anfang 2018 begonnen, den Alexa liothek“ zu entwickeln, um Erfahrungen mit den von Amazon bereit- wicklerwerkzeugen und dem Review-Prozess zu sammeln, der ähn- App-Entwicklung durchlaufen werden muss. Dieser Beitrag gibt wonnenen Erfahrungen und Probleme und verweist auf ähnliche anderen Bibliotheken.

be found in many households and more and more people are the end of 2017 the KIT Library decided to engage in this top- a Skill called „KIT-Bibliothek“. The reasons for doing this step aining experience and was a technique-driven approach. a skill is similar to the development of a smartphone app ss to publish the skill. This reading delivers an insight in- but shows also the challenges and gives a short over- other libraries.

Sprachassistenz – mehr als nur ein Hype?

Ob Apples Siri, der Google Assistant oder Amazons Alexa – fast wie eine Flut kommen jeden Monat neue Geräte mit eingebauten Sprachassistenten auf dem Markt und buhlen um die Gunst der Kunden. Unbestrittener Vorreiter war dabei Apples Siri auf dem iPhone, jedoch unterscheidet man mittlerweile den reinen Software-Assistenten (z.B. Siri) und das Hardware-Gerät, auf dem dieser ausgeführt wird (z.B. iPhone).

Die Software-Herstell...

...elbige Hardware einzu- ...Die Tatsache, dass Amazon der erste Herstel- ...ler mit einem reinen smarten Lautsprecher war, die aggressive Vermarktung und ein niedriger Einstiegspreis sind der Grund für einen großen Ansturm auf die smarten Lautsprecher von Amazon. Genaue Zahlen gibt es von den Herstellern nicht, allerdings hat Amazon alleine im dritten Quartal 2017 mindestens 20 Millionen Echo-Geräte verkauft. Laut einer repräsentativen Umfrage haben 16% der US-Amerikaner einen smarten Lautsprecher zu Hause stehen. Dabei hat Amazon einen Marktanteil von fast 70% [1]. Aus diesen Gründen fiel die Wahl bei der Auswahl eines möglichen Assistenten für die KIT-Bibliothek auf Amazons Alexa.

Wie bereits erwähnt ist Alexa der reine Software-Assistent. Um diesen zu benutzen wird eine Hardwarekomponente benötigt. Die erste verfügbare Hardware war der von Amazon selbst vertriebene Lautsprecher „Amazon Echo“. Die günstigste Variante nennt sich „Echo Dot“, ist kaum größer als ein Eishockey Puck und besteht aus einem Lautsprecher, vier Tasten sowie 7 Fernfeld-Mikrofonen (diese sorgen dafür, dass man auch trotz Nebengeräuschen oder aus dem nächsten Zimmer verstanden wird). Das obere Ende der Produktpalette der Smarten Lautsprecher bildet der „Echo Show“ mit eingebautem Bildschirm. Dazwischen gibt es noch den klassischen „Echo“ (etwa dreimal so groß wie der „Echo Dot“ und mit besserem Lautsprecher, den „Echo Plus“ mit eingebautem Smart-Home Hub und den „Echo Spot“, einen Radiowecker großen Lautsprecher mit kleinem Bildschirm. Der Normalpreis für den Echo Dot liegt bei 60 ...llerdings lockt Amazon alle paar Wochen mit Son- ...ngeboten und bietet diesen schon ab 35 € an, Oberklasse-Modell „Echo Show“ mit Bildschirm der Normalpreis bei 220 €.

...den Geräten bietet Amazon seinen Assistenten ...den Apps für Android (Alexa-App und Amazon sowie in der Fernbedienung zur hauseigenen -Box FireTV an.

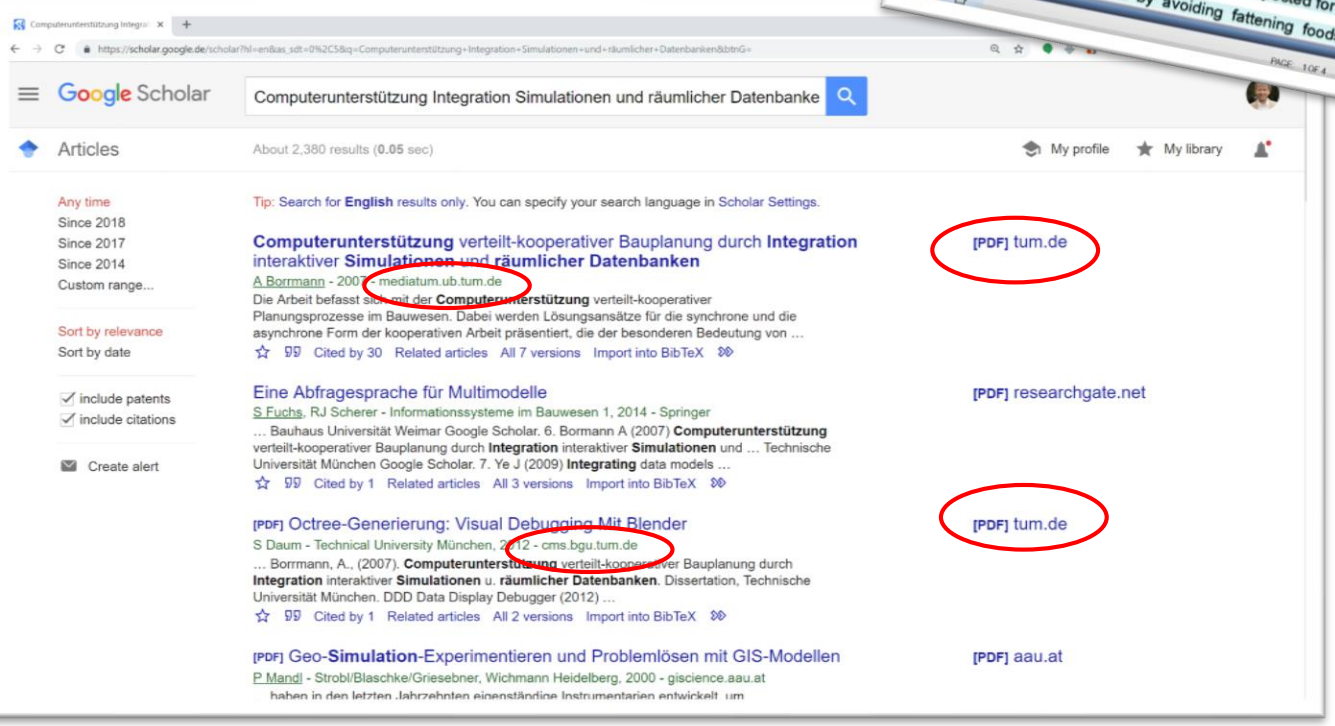
...tzlich haben alle Geräte eines gemeinsam: Sie bestehen aus einem oder mehreren Mikrofonen um die Frage des Benutzers aufzunehmen und einem Lautsprecher um die Antwort auszugeben. Zum Betrieb benötigen Sie zudem eine Internetverbindung. Auf dieser Basis erlaubt es Amazon Fremdfirmen,

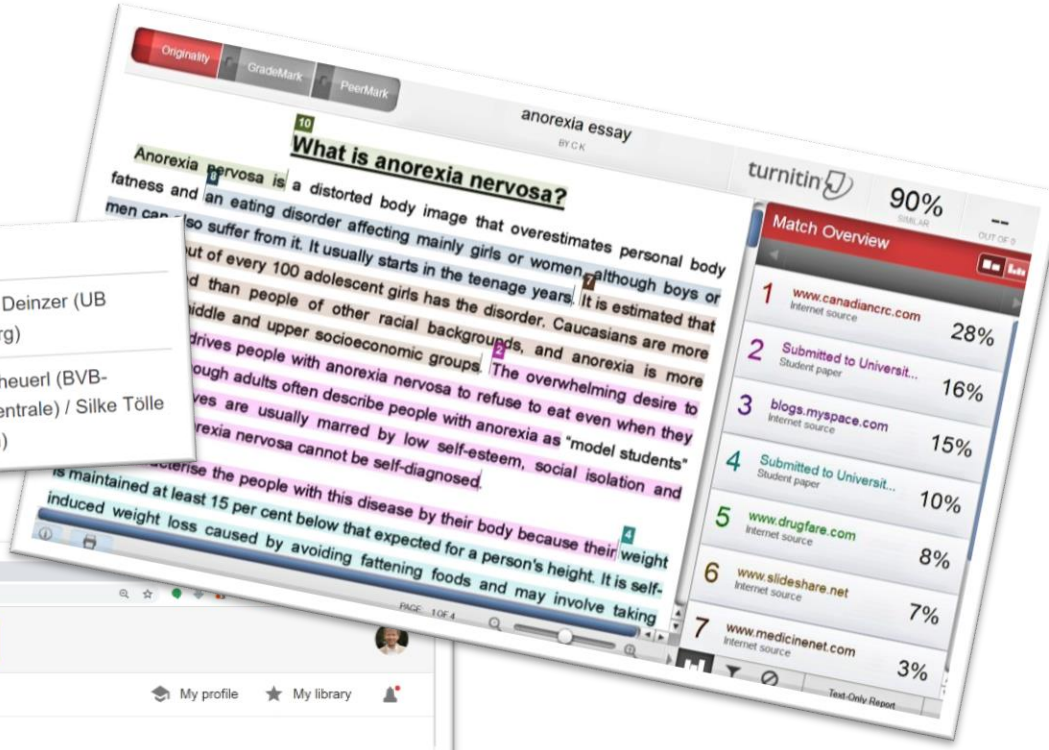
b.i.t.online 21 (2018) Nr. 2

12:45 - 14:05	Moderation: Dr. Mathias Kratzer	
	Mit Alexa in die Komfortzone unserer Nutzerinnen und Nutzer. Ein Sprachassistent als Bibliotheksservice (KSI)	Steffen Illig / Johannes Gütling (UB Bamberg)

… und mehr

14:35 - 16:00	Moderation: Dr. Hildegard Schäffler	
	Predatory Publishing – Sind Bibliotheken hier in der Pflicht? (KER)	Dr. Gernot Deinzer (UB Regensburg)
	"Endlich automatisiert" – Literatursichtung, Auswahl und Bestellung leicht gemacht (AG Lokalsysteme + hbz)	Robert Scheuerl (BVB-Verbundzentrale) / Silke Tölle (hbz, Köln)



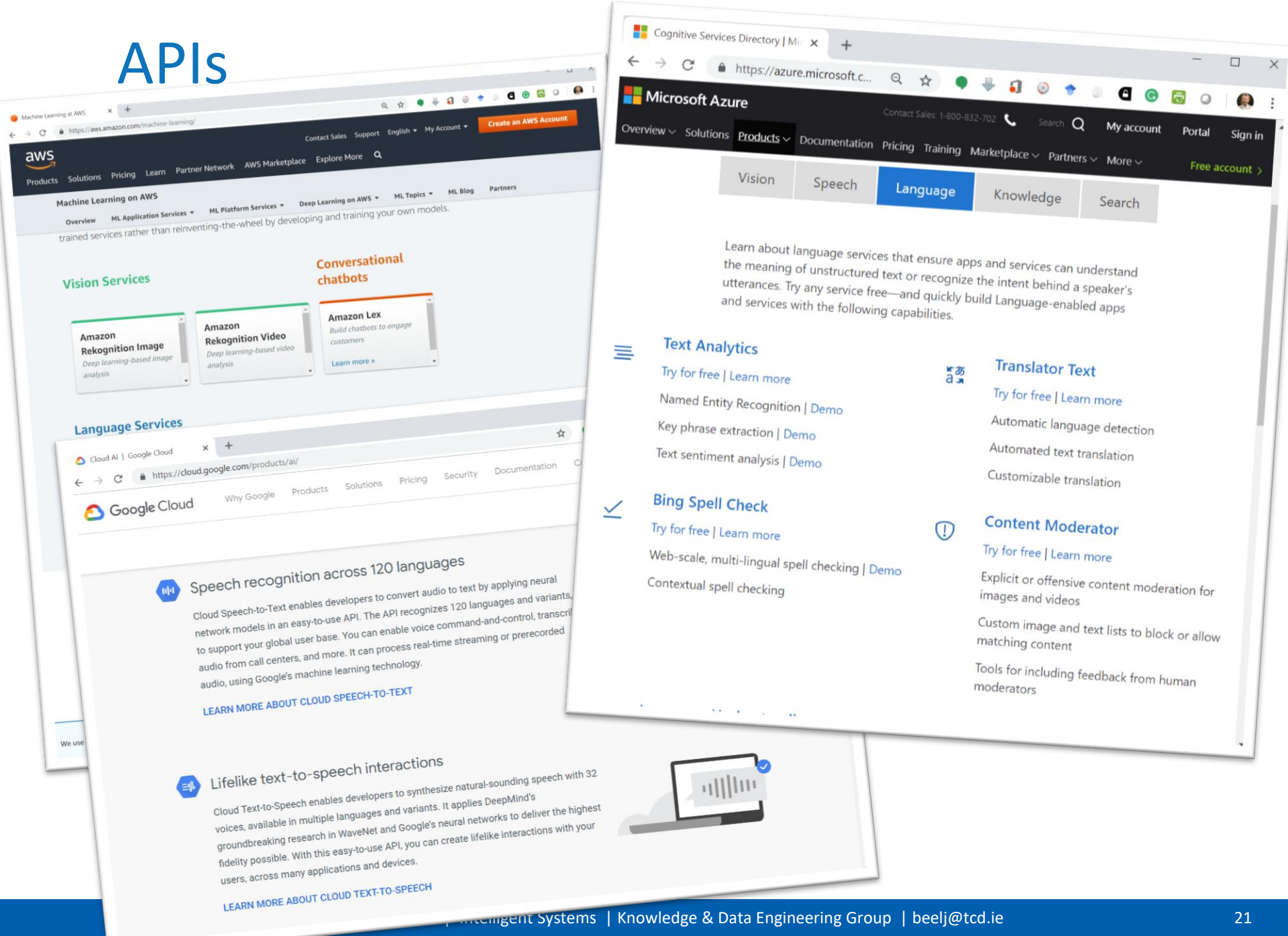


https://scholar.google.de/scholar?q=Computerunterstützung+Integration+Simulationen+und+räumlicher+Datenbanken

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

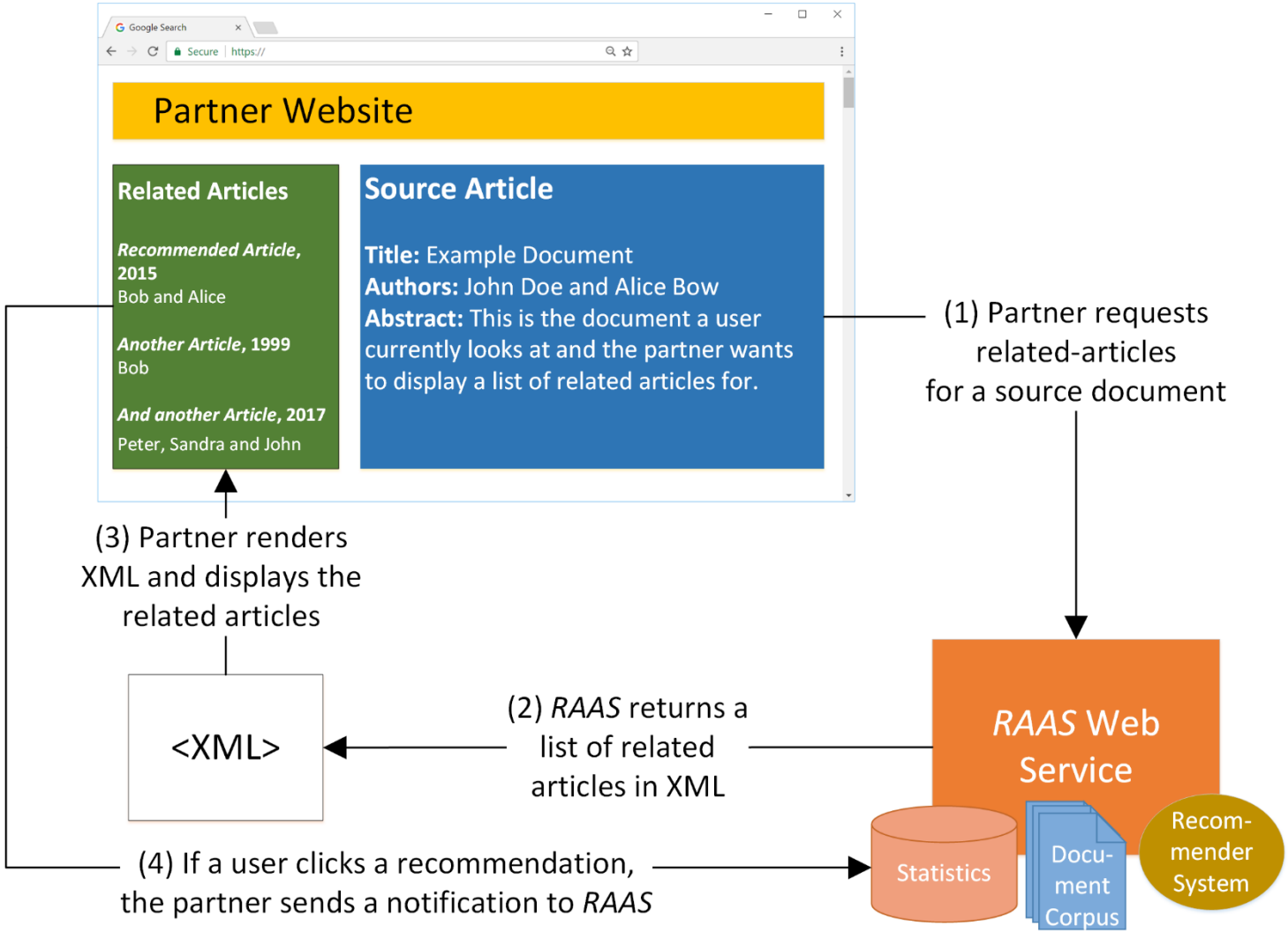
3. Machine-Learning als Service

APIs


Machine Learning at AWS
https://aws.amazon.com/machine-learning/
Contact Sales
Support
English
My Account
Create an AWS Account
aws
Products
Solutions
Pricing
Learn
Partner Network
AWS Marketplace
Explore More
Machine Learning on AWS
Overview
ML Application Services
ML Platform Services
Deep Learning on AWS
ML Topics
ML Blog
Partners
trained services rather than reinventing-the-wheel by developing and training your own models.
Vision Services
Conversational chatbots
Amazon Rekognition Image
Deep learning-based image analysis
Amazon Rekognition Video
Deep learning-based video analysis
Amazon Lex
Build chatbots to engage customers
Learn more »
Language Services
Cloud AI | Google Cloud
https://cloud.google.com/products/ai/
Google Cloud
Why Google
Products
Solutions
Pricing
Security
Documentation
Speech recognition across 120 languages
Cloud Speech-to-Text enables developers to convert audio to text by applying neural network models in an easy-to-use API. The API recognizes 120 languages and variants, to support your global user base. You can enable voice command-and-control, transcri audio from call centers, and more. It can process real-time streaming or prerecorded audio, using Google's machine learning technology.
LEARN MORE ABOUT CLOUD SPEECH-TO-TEXT
Lifelike text-to-speech interactions
Cloud Text-to-Speech enables developers to synthesize natural-sounding speech with 32 voices, available in multiple languages and variants. It applies DeepMind's groundbreaking research in WaveNet and Google's neural networks to deliver the highest fidelity possible. With this easy-to-use API, you can create lifelike interactions with your users, across many applications and devices.
LEARN MORE ABOUT CLOUD TEXT-TO-SPEECH
Cognitive Services Directory | Mi
https://azure.microsoft.c...
Microsoft Azure
Contact Sales: 1-800-832-702
Search
My account
Portal
Sign in
Overview
Solutions
Products
Documentation
Pricing
Training
Marketplace
Partners
More
Free account
Vision
Speech
Language
Knowledge
Search
Learn about language services that ensure apps and services can understand the meaning of unstructured text or recognize the intent behind a speaker's utterances. Try any service free—and quickly build Language-enabled apps and services with the following capabilities.
Text Analytics
Try for free | Learn more
Named Entity Recognition | Demo
Key phrase extraction | Demo
Text sentiment analysis | Demo
Translator Text
Try for free | Learn more
Automatic language detection
Automated text translation
Customizable translation
Bing Spell Check
Try for free | Learn more
Web-scale, multi-lingual spell checking | Demo
Contextual spell checking
Content Moderator
Try for free | Learn more
Explicit or offensive content moderation for images and videos
Custom image and text lists to block or allow matching content
Tools for including feedback from human moderators

Mr. DLib

http://mr-dlib.org/



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

4. Derzeitige Qualität von Machine Learning

Qualität z.B. von Empfehlungen

Online Catalogue TU München

https://opac.ub.tum.de/InfoGuideClient.tumsis/start.do?Query=-1%3D"machine+learning"&Language=...

Universitätsbibliothek der TUM
Online-Katalog

Search | My List | Account | Additional Services

New Search | Previous Searches | Results

Your search request Keyword Anywhere = machine learning

Narrow results

- **Branch Library**
  - TB Math./Inf. (1386)
  - TB Stammgelände (526)
  - TB Chemie (71)
  - TB Maschinenwesen (63)
  - TB Weihenstephan (59)
  - Display more ...
- **Subject**
  - Informatik (1959)
  - Mathematik (162)
  - Biologie (140)
  - Medizin (108)
  - Fertigungstechnik (70)
  - Display more ...
- **Subject Heading**
  - Maschinelles Le... (1107)
  - Künstliche Inte... (353)
  - Data Mining (326)
  - Soft Computing (207)
  - Mustererkennung (181)
  - Display more ...
- **Year**
  - >= 2014 (961)
  - 2010 - 2013 (791)
  - 2001 - 2009 (740)
  - <= 2000 (733)
- **Media Type**
  - Buch (2953)
  - Konferenzbericht (1561)
  - eMedien (1227)
  - Online-Zugriff (1222)
  - eBook (1169)

Result(s) Online Catalogue of the TU München (3225) |« < 1 2

1. Fundamentals of deep learning
   Buduma, Nikhil
   2017
   0303/DAT 708f 2017 L 749+11
   on loan (total of reservations: 0)
   Get document | Add to My list
2. Deep learning
   Goodfellow, Ian
   2016
   0303/DAT 708f 2017 L 142
   available for request
   Get document | Add to My list
3. Machine Learning mit Python und Scikit-learn und TensorFlow
   Raschka, Sebastian
   2018
   0003/DAT 708f 2015 L 246(2)+10
   on loan (total of reservations: 5)
   Get document | Add to My list
4. Advances in machine learning
   2009
   0102/DAT 001z 2001 A 999-5828
   available for request
   Get document | Add to My list
5. Deterministic and statistical methods in machine learning
   2005
   0102/DAT 001z 2001 A 999-3635
   available for request
   Get document | Add to My list
6. Hands-on machine learning with Scikit-Learn and TensorFlow
   Géron, Aurélien

https://opac.ub.tum.de/InfoGuideClient.tumsis/start.do?Query=-1%3D%22machine+learning%22&Language=en&x=0&y=0&BVBLocation=wotum

Online-Katalog der TU München (1/3225)

Fundamentals of deep learning
designing next-generation machine intelligence algorithms

Autor: Buduma, Nikhil
Hrsg./Bearb.: Locascio, Nicholas ¬[Verfasser von Zusatztexten]
Ausgabebez.: First edition
Verlagsort, Verlag, Jahr: Beijing ; Boston ; Farnham ; Sebastopol ; 2017
Umfangsangabe: xii, 283 Seiten
Schlagwort: Maschinelles Lernen ; Deep learning
ISBN: 978-1-491-92561-4, 1-491-92561-2, 978-1-61689-528-0

bibtip Andere Benutzer fanden auch interessant:

- Goodfellow, Ian: Deep learning ([2016])
- Raschka, Sebastian: Machine learning mit Python (2017)
- Wartala, Ramon: Praxiseinstieg Deep Learning (2018)
- Raschka, Sebastian: Machine Learning mit Python und Scikit-learn (2018)
- Patterson, Josh [Verfasser]: Deep learning (August 2017)
- Géron, Aurélien: Hands-on machine learning with Scikit-Learn and 2017)

https://opac.ub.tum.de/InfoGuideClient.tumsis/singleHit.do?methodToCall=showHit&curPos=1&identifier=-1_FT_1693924031

Qualität der Automatisierten Erschließung in der DNB

Deutsche Nationalbibliothek: In Frankfurt lesen jetzt zuerst Maschinen

Überdruss an Kernaufgabe

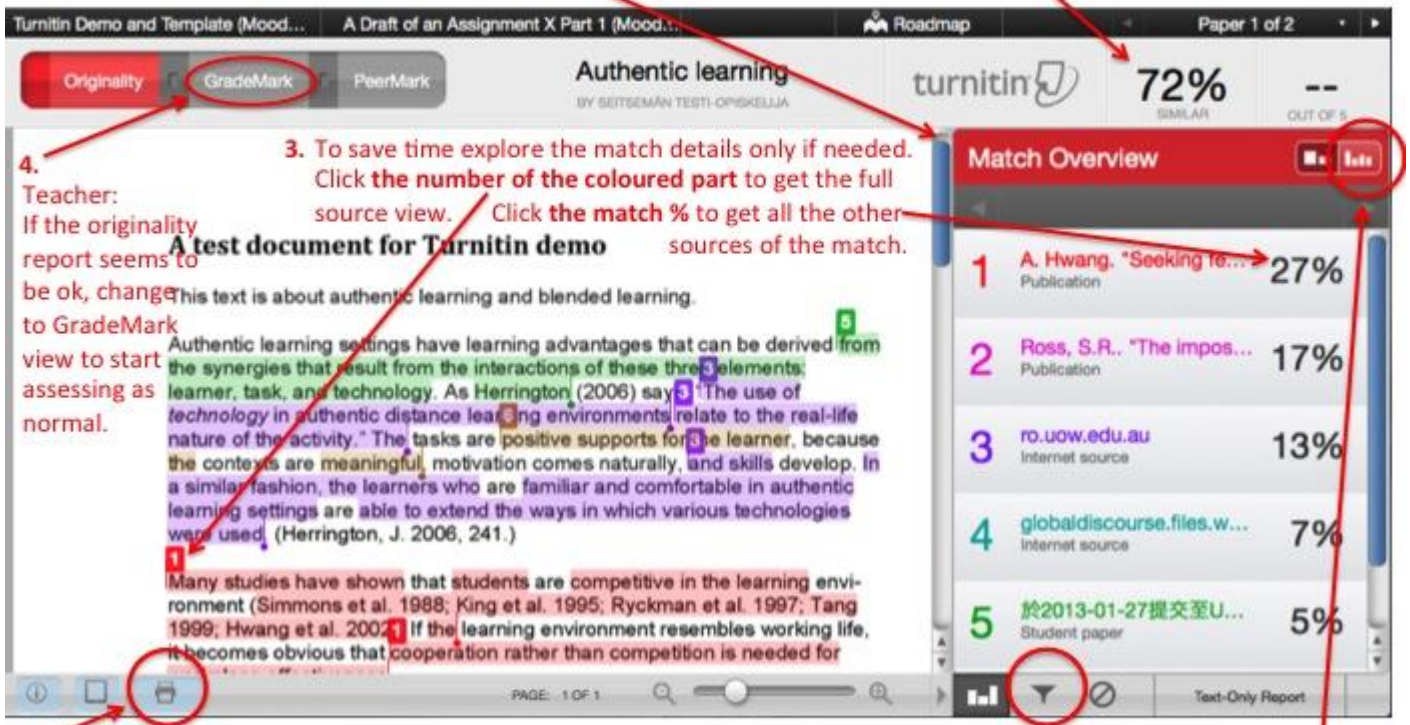
Die Resultate einer automatischen Inhaltserschließung hingegen sind, wie Untersuchungen der DNB selbst zeigen, vergleichsweise ernüchternd. Die Qualität der algorithmisch generierten Schlagwörter schwankt deutlich. Bei Wissenschaftsfächern mit eindeutigem Fachvokabular ist sie besser, bei anderen schlechter. Über alle Fächer verteilt erweisen sich nur 50 Prozent als „nützlich“ oder „sehr nützlich“, 50 Prozent sind „wenig nützlich“ oder schlicht falsch. Zwei Beispiele: Der Titel „Die Macht der Bilder: A. Lorenzetti – P. Picasso – E. Kienholz, 1338-1968“ erhält die maschinell ermittelten Schlagwörter „Kiefernholz; Bildmaterial; Macht; Lorenzetti, Ambrogio“, das Buch „Jüdisches Leben in Biebesheim“ die Schlagwortfolge „Biebesheim am Rhein; Leben“.

http://www.faz.net/aktuell/feuilleton/buecher/maschinen-lesen-buecher-deutsche-nationalbibliothek-setzt-auf-technik-15128954-p2.html

Qualität von Plagiatserkennung

1. Browse the originality report and take a quick look at the coloured parts (=simililarities with different sources) of the text by scrolling the bar.

2. Do not pay too much attention to the similarity percentages. All matches (coloured parts) must be interpreted in the context.



3. To save time explore the match details only if needed. Click **the number of the coloured part** to get the full source view. Click **the match %** to get all the other sources of the match.

4. Teacher: If the originality report seems to be ok, change to GradeMark view to start assessing as normal.

6. Teacher: if plagiarism is suspected, download a pdf file of the current view and attach an interpretation of it for the investigator.

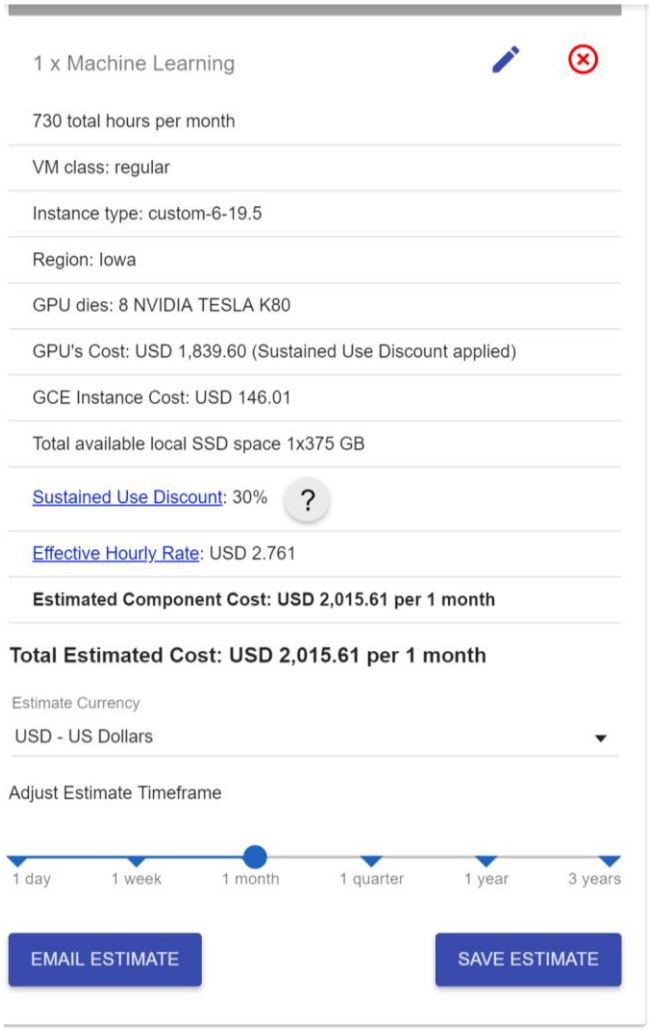
5. Filter matches if needed:
   - Similarity settings – bibliography, direct quotations or small matches – can be edited in Filters & Settings view
   - To exclude some of the matching sources, open All Sources view.

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

5. Herausforderungen

Übersicht

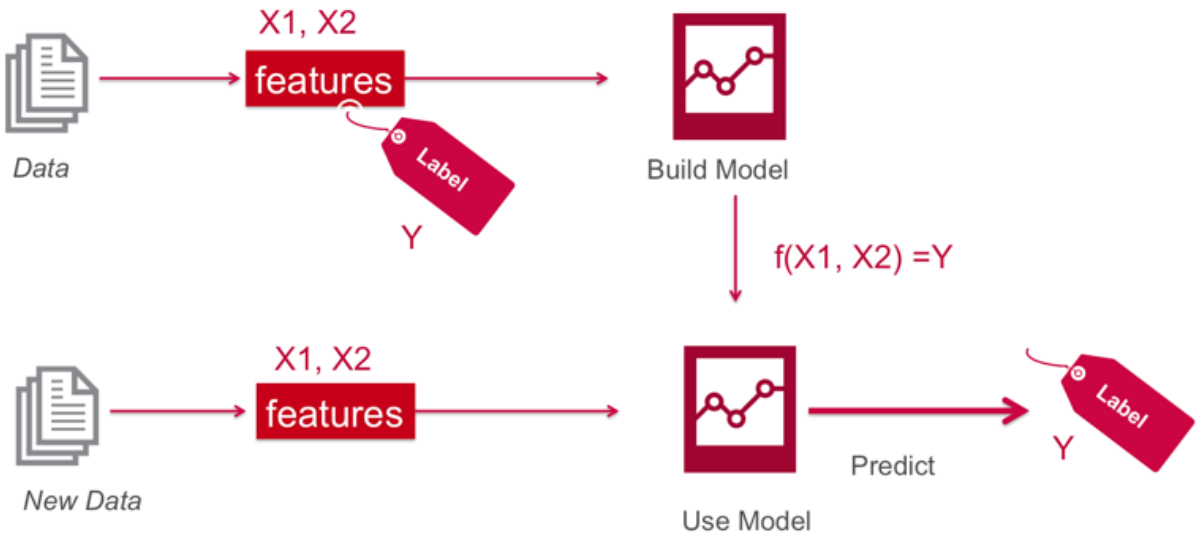
- **Algorithmen**
- **Infrastruktur**
  - Speicher
  - Rechenkapazität
  - Mitarbeiter
- **Verfügbarkeit von Daten**
  - Lizenzen
  - Nischengebiete
  - Sprachen
  - GDPR
- **Qualität on Daten**
  - Rauschen
  - Bias
  - Ground Truth



https://cloud.google.com/products/calculator/#id=82b0926f-6269-41c3-9dba-51b2bb7fcbea

Ground Truth

- **„Labels“ stellen die „Wahrheit“ dar die erreicht werden soll, und auf derer basierend Machine Learning evaluiert wird.**
- **Machine Learning kann nur so gut sein wie die Ground Truth**



https://mapr.com/blog/demystifying-ai-ml-dl/assets/sl.png

Bias in Ground Truth

big think Smarter Faster™ | *Your Daily Microdose of Genius* | Discover ▾ Experts ▾ Playlists ▾ Podcast

Predictive policing: Data can be used to prevent crime, but is that data racially tinged?

May 2, 2018 *by* ERIC SIEGEL

High angle view of cityscape model on gray background with data point overlay. (Getty/Big Think)

As predictive analytics advances decision making across the public and private sectors, nowhere could this prove more important – nor more risky – than in law enforcement. If the rule of law is the cornerstone of society, getting it right is literally foundational. But the art of policing by data, without perpetuating or even magnifying the human biases captured within the data, turns out to be a very tricky art indeed.

Zitationen als Ground-Truth

- Negativbeispiel

Mr. DLib: Recommendations-as-a-Service (RaaS) for Academia

Joeran Beel
Trinity College Dublin
ADAPT Centre, Ireland
j@beel.org

Akiko Aizawa
National Institute of Informatics
Tokyo, Japan
aizawa@nii.ac.jp

Corinna Breitinger
University of Konstanz
Germany
corinna.breitinger@uni.kn

Bela Gipp
University of Konstanz
Germany
bela.gipp@uni.kn

ABSTRACT

Only few digital libraries and reference managers offer recommender systems, although such systems could assist users facing information overload. In this paper, we introduce Mr. DLib's recommendations-as-a-service, which allows third parties to easily integrate a recommender system into their products. We explain the recommender approaches implemented in Mr. DLib (content-based filtering among others), and present details on 57 million recommendations, which Mr. DLib delivered to its partner GESIS Sowiport. Finally, we outline our plans for future development, including integration into JabRef, establishing a living lab, and providing personalized recommendations.

KEYWORDS

recommender system, digital library, reference management software, recommendation-as-a-service, RaaS, API, web service

1 INTRODUCTION

Recommender systems in academia help researchers and scientists overcome information overload. However, only a few operators of academic services, such as digital libraries and reference managers, offer recommender systems to their users. The lack of recommender systems is possibly due to the high costs of implementing and maintaining such systems. Furthermore, operators may lack the expertise to design recommender systems. In this paper, we introduce Mr. DLib's recommendations-as-a-service (RaaS), which allows operators of academic services to easily integrate recommendations in their system. For the operator, the development and maintenance effort is minimal and no expertise in designing recommender systems is required.

Only a few other companies and organizations offer recommender-systems as-a-service for academia. BibTip [15] and bX [10] are commercial RaaS provider that offer co-occurrence-based recommendations. This approach is a generic recommendation approach applicable to a variety of items (documents, movies etc.), and suitable for rather large systems with many users [3]. Additionally, coverage is rather low, i.e. the approach recommends only a fraction of the documents in a library's catalogue. CORE [13, 14] and Babel [17] offer RaaS through an API, JavaScript client, and browser plug-in. Both services recommend open-access documents (CORE indexed approx. 68 million documents, Babel approx. 40 million), and Babel states that they are welcoming researchers to evaluate novel algorithms in Babel.

While these services have their strengths and weaknesses, only Mr. DLib offers all the following features: open source; not-for-profit; supportive of research, i.e. information about the RaaS, its architecture, and effectiveness is published [5, 7, 8, 11, 16]; integration into reference managers *and* digital libraries (of any size); recommendations for open-access articles *and* the option for partners to manage private document collections; capable of recommending all documents in a corpus (100% coverage).

2 MR. DLIB'S RECOMMENDER SYSTEM

Mr. DLib was originally developed as a Machine-readable Digital Library at the University of California, Berkeley and introduced at JCDL 2011 [2]. Since September 2016, Mr. DLib provides recommendations-as-a-service. The concept is illustrated in Figure 1. A user browses a web site of Mr. DLib's partner, e.g. a digital library. (1) When the user looks at a specific article's detail page, the web site requests a list of related articles from Mr. DLib's RESTful Web Service as a HTTP GET request:

GET /v1/documents/{document_id}/related_documents/ [1]

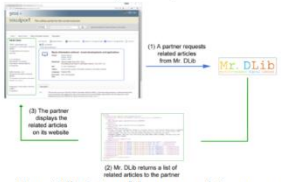


Figure 1: Illustration of the recommendation process

(2) Upon receiving the request, Mr. DLib computes a list of related articles, and returns the list as XML. (3) The partner website converts the XML to HTML and displays the recommendations on its web page (or mobile app, or desktop application).[2] Before this process begins, Mr. DLib indexes the metadata of the partner's documents (title, authors, abstract, venue, keywords). Mr. DLib uses Apache Lucene/Solr's *More-Like-This* function to calculate document relatedness. We are also experimenting with alternative recommendation approaches, such as stereotype and most-popular recommendations [5]. Additionally, we are

[1] Example: https://api.mr-dlib.org/v1/documents/gesis-bib-136994/related_documents/

[2] Example: http://sowiport.gesis.org/search/id/gesis-bib-136994

experimenting with key-phrase extraction and a bibliometric re-ranking based on readership data from Mendeley [7, 8, 16].

The production system api.mr-dlib.org and development system api-dev.mr-dlib.org each run on a dedicated server (i7-6700k, 32GB RAM, SSD hard drives). The development system is also used for resource intensive tasks, including document indexing, key-phrase extraction, and the calculation of bibliometrics. The uptime of the servers is constantly monitored[3], and the average response time to deliver recommendations is 682ms. We further run a beta system api-beta.mr-dlib.org on a virtual machine (4 cores, 14 GB RAM).

3 MR. DLIB'S DIGITAL LIBRARY PARTNER

Mr. DLib's first partner is the digital library Sowiport[4], which is Germany's largest social science repository, operated by the GESIS institute [12]. Mr. DLib has indexed around 10 million documents from Sowiport. While GESIS agreed to let their documents be recommended to all partners of Mr. DLib, GESIS chose to recommend only its own content on Sowiport.

Between September 18th 2016 and February 11th 2017, Mr. DLib delivered 57,435,086 recommendations to Sowiport. Users clicked on 77,468 recommendations. This equates to an overall click-through rate (CTR) of 0.13%. This CTR is rather low, and, as shown in Figure 2, there was a notable variance (e.g. 0.19% in September and 0.10% in December). The variance may be caused by different algorithms used. In addition, recommendations are delivered when web spiders, such as the Google Bot, crawl the Sowiport website. In contrast, clicks are logged with JavaScript, which is usually not executed by web spiders.

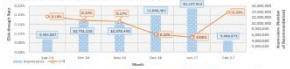


Figure 2: Recommendations and CTR by Month

4 LICENSE AND POLICY

Mr. DLib advocates an 'open culture' and publishes its code as open source on GitHub[5]. Project details are described in a public WIKI (Confluence)[6], and issues managed in a public ticket tracker (JIRA) [7]. Data from our research is published on Harvard's Dataverse[8] (if we can ensure privacy and copyrights of our partners). We invite other researchers to evaluate their recommendation algorithms with Mr. DLib and our partners[9].

5 OUTLOOK

In the future, we plan to add more partners (e.g. JabRef [11] and Docear [1, 4, 6]), import more documents (e.g. CORE [14]), improve the recommendation quality (e.g. cross-language recommendations [18]), and make them more personalized. We will also implement a JavaScript client that allows an easier integration of Mr. DLib's recommender system into partner websites, and monitors the number of delivered and clicked recommendations more reliably. In the long term, we will extend the scope of Mr. DLib to recommend not only related articles, but also other items relevant in academia, such as calls for papers and research grants, as well as recommending researchers as potential collaborators. In addition, we plan to establish an interactive evaluation task [9] that allows other researchers to evaluate new recommendation approaches in Mr. DLib's recommender system.

5 ACKNOWLEDGEMENTS

This work was supported by a fellowship within the FITweltweit programme of the German Academic Exchange Service (DAAD) and a scholarship of the Carl Zeiss Foundation. This publication has also emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number 13/RC/2106. We are further grateful for the support provided by Zeljko Carevic, Philipp Mayr, Siddharth Dinesh, Sophie Siebert, Stefan Feyer, Sara Mahmoud, Gabor Neusch, and Felix Beierle.

REFERENCES

[1] Beel, J. et al. 2011. Docear: An Academic Literature Suite for Searching, Organizing and Creating Academic Literature. *11th Annual International ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (2011), 465–466.

[2] Beel, J. et al. 2011. Introducing Mr. DLib, a Machine-readable Digital Library. *Proceedings of the 11th ACM/IEEE Joint Conference on Digital Libraries (JCDL'11)* (2011), 463–464.

[3] Beel, J. et al. 2016. Research Paper Recommender Systems: A Literature Survey. *International Journal on Digital Libraries*, 4 (2016), 305–338.

[4] Beel, J. et al. 2009. SciPlore MindMapping' - A Tool for Creating Mind Maps Combined with PDF and Reference Management. *D-Lib Magazine*. 15, 11.

[5] Beel, J. et al. 2017. Stereotype and Most-Popular Recommendations in the Digital Library Sowiport. *Proceedings of the 15th International Symposium of Information Science (ISI)* (2017).

[6] Beel, J. et al. 2014. The Architecture and Datasets of Docear's Research Paper Recommender System. *D-Lib Magazine*. 20, 11/12 (2014).

[7] Beel, J. and Dinesh, S. 2017. Real-World Recommender Systems for Academia: The Gain and Pain in Developing, Operating, and Researching them. *Fifth Workshop on Bibliometric-enhanced Information Retrieval (BIR)*, 6–17.

[8] Beierle, F. et al. 2017. Exploring Choice Overload in Related-Article Recommendations in Digital Libraries. *5th International Workshop on Bibliometric-enhanced Information Retrieval (BIR)*.

[9] Bogers, T. et al. 2017. Lessons Learned from the CHiC and SBS Interactive Tracks: A Wishlist for Interactive IR Evaluation. *CHIIR 2017 Second Workshop on Supporting Complex Search Tasks (SCST 2017)* (2017), 11–14.

[10] ExLibris 2013. bX Usage-Based Services transform your discovery experience! *Web page*, *http://www.exlibrisgroup.com/category/bXUsageBasedServices*.

[11] Feyer, S. et al. 2017. Integration of the Scientific Recommender System Mr. DLib into the Reference Manager JabRef. *Proceedings of the 39th European Conference on Information Retrieval (ECIR)* (2017).

[12] Hienert, D. et al. 2015. Digital Library Research in Action – Supporting Information Retrieval in Sowiport. *D-Lib Magazine*. 21, 3/4 (2015).

[13] Knoth, P. et al. 2017. Towards effective research recommender systems for repositories. *Proceedings of the Open Repositories Conference* (2017).

[14] Knoth, P. and Zdrahal, Z. 2012. CORE: three access levels to underpin open access. *D-Lib Magazine*. 18, 11/12 (2012).

[15] Monnich, M. and Spiering, M. 2008. Einsatz von BibTip als Recommendersystem im Bibliothekskatalog. *Bibliotheksdienst*. 42, 1 (2008).

[16] Siebert, S. et al. 2017. Extending a Research Paper Recommendation System with Bibliometric Measures. *5th International Workshop on Bibliometric-enhanced Information Retrieval (BIR)*.

[17] Wesley-Smith, I. and West, J.D. 2016. Babel: A Platform for Facilitating Research in Scholarly Article Discovery. *Proceedings of the 25th International Conference Companion on World Wide Web* (2016), 389–394.

[18] Zhou, D. et al. 2016. A study of user profile representation for personalized cross-language information retrieval. *Asilb Journal of Information Management*. 68, 4 (2016), 448–477.

[3] http://monitoring.mr-dlib.org/

[4] http://sowiport.gesis.org/

[5] http://source-ecle.mr-dlib.org/

[6] http://wiki.mr-dlib.org/

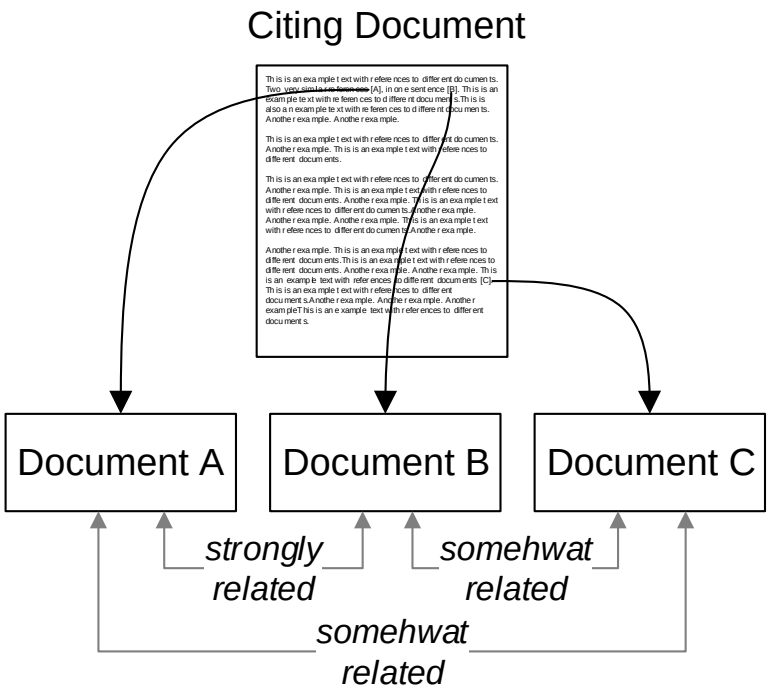
[7] http://tickets.mr-dlib.org/

[8] http://data.mr-dlib.org/

[9] http://mr-dlib.org/research/

Abgeleitete Ground Truth

- **Z.B. Virtual Citation Proximity oder Expertenklassifikation**



6. Diskussion

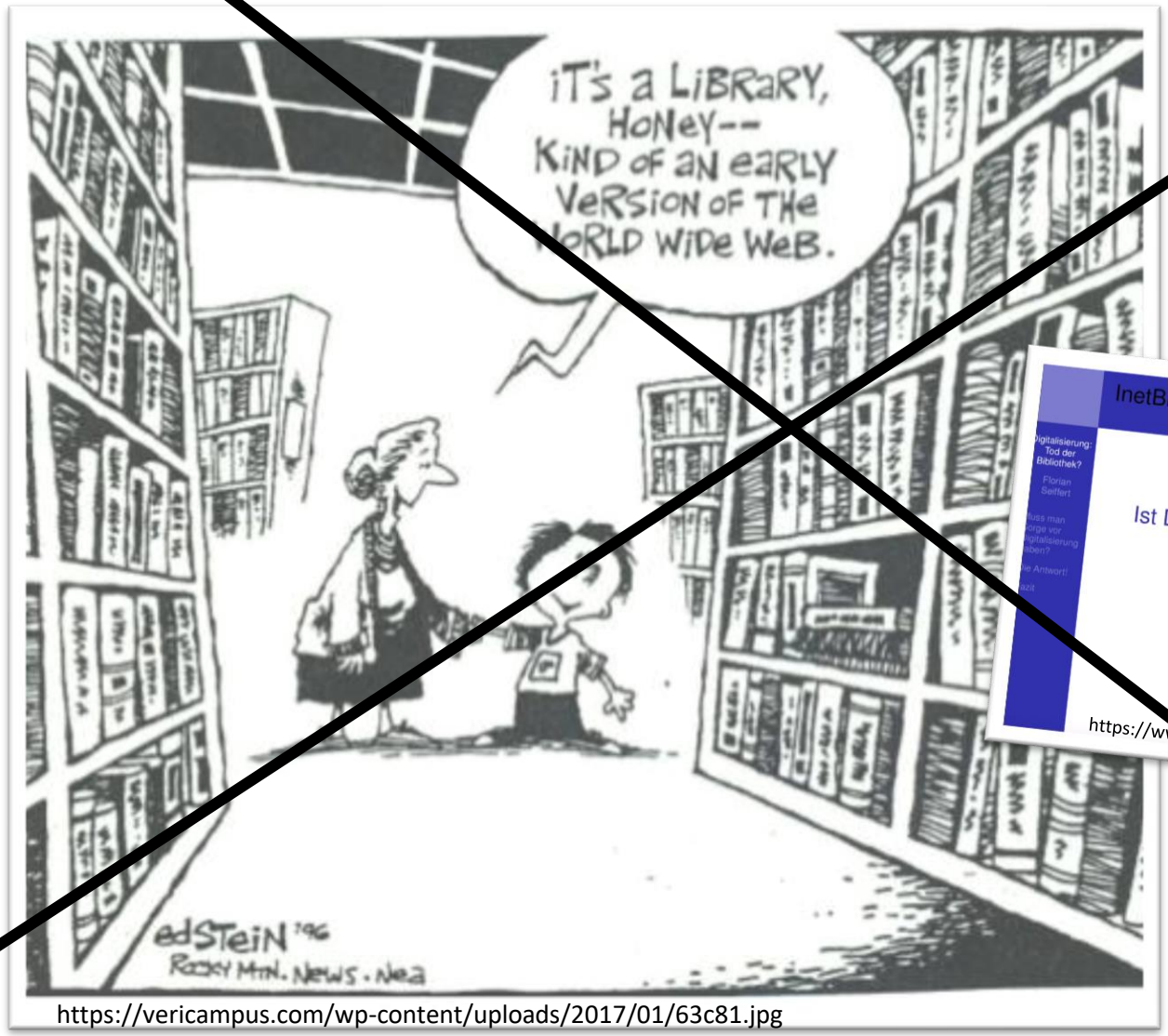
AI/ML und die Bibliothek

“In building the next generation of digital libraries, artificial intelligence will play several important roles”

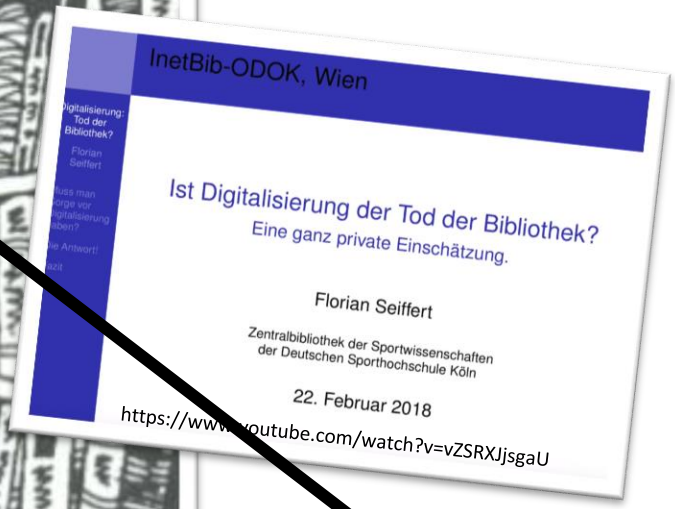
Prof. Daphne Koller, Stanford University

1996

Werden Bibliotheken überflüssig?



https://vericampus.com/wp-content/uploads/2017/01/63c81.jpg



Aufgaben der (Universitäts-) Bibliothek

„Beschafft, erschließt und vermittelt Fachinformationen […] die Forschung, Lehre und Studium unterstützt“

Universitätsbibliothek Paderborn
https://core.ac.uk/download/pdf/50519806.pdf

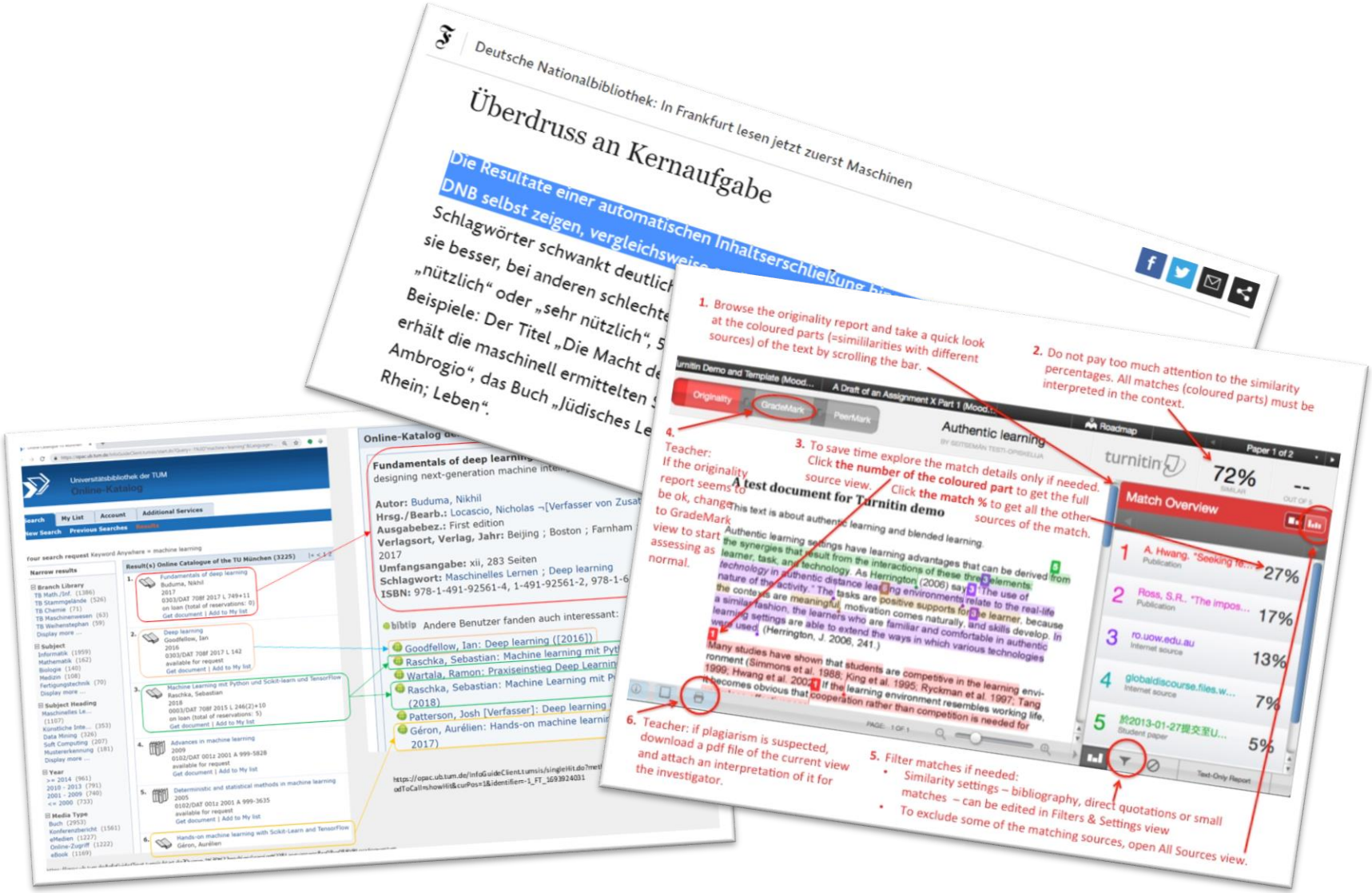
„Fördert durch Beratung und Schulung die Kompetenz zur Informationssuche, -beschaffung und –weiterverarbeitung“

Universitätsbibliothek Düsseldorf https://www.uni-duesseldorf.de/home/fileadmin/redaktion/ZUV/Justitiariat/Amtliche_Bekanntmachungen/2011/2011_10_18_AB_18.pdf

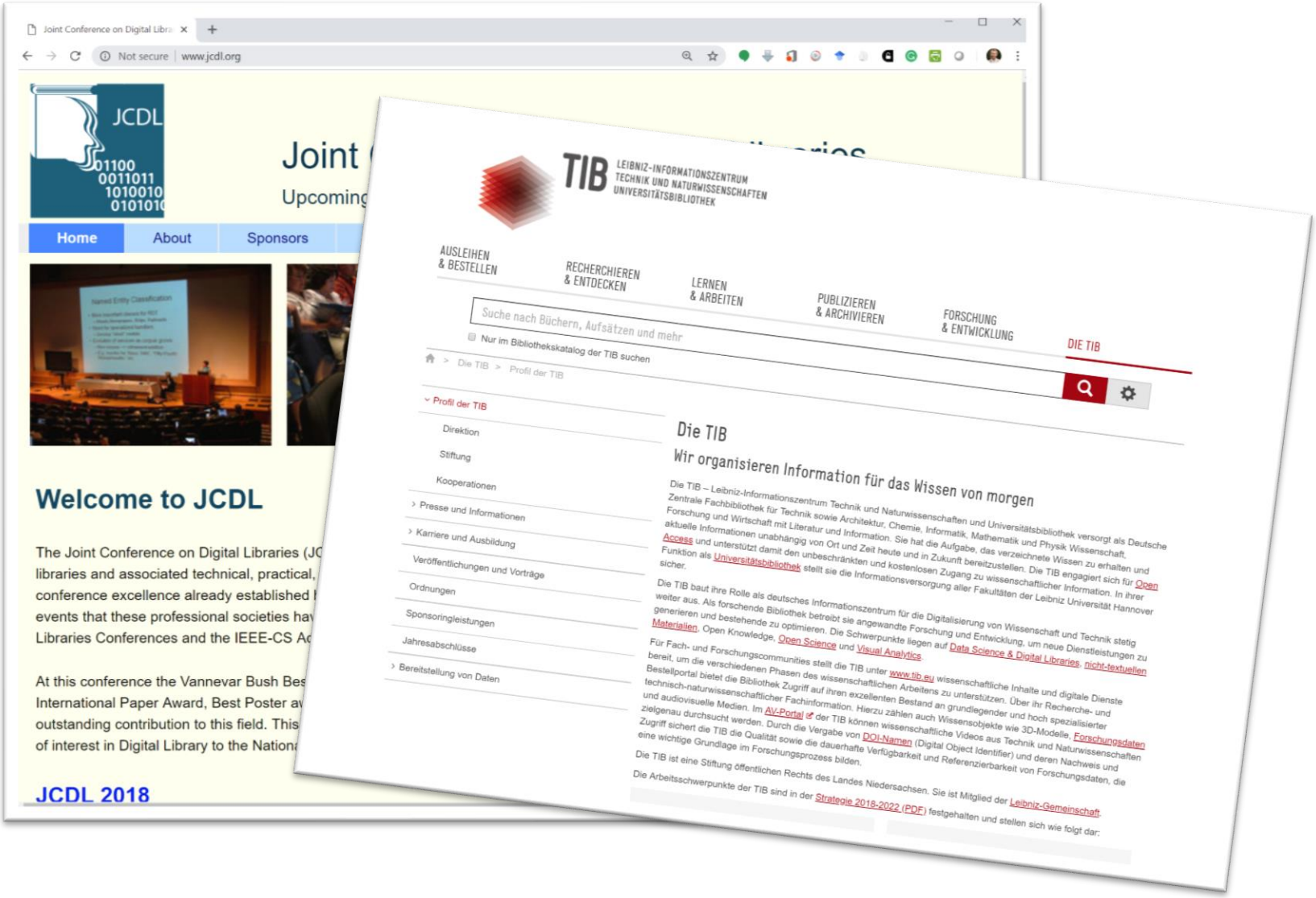
Typische Probleme/Aufgaben der Bibliotheken werden nicht weniger

- **Steigende Anzahl von Publikationen**
- **Bisher nicht erfasster Inhalt (Bilder, Formeln, …)**
- **Interdisziplinarität**
- **Predatory Publishing**
- **Schlechte Qualität von Veröffentlichungen**
- **Unzureichende Ausbildung von Studenten und Forschern**
- **…**

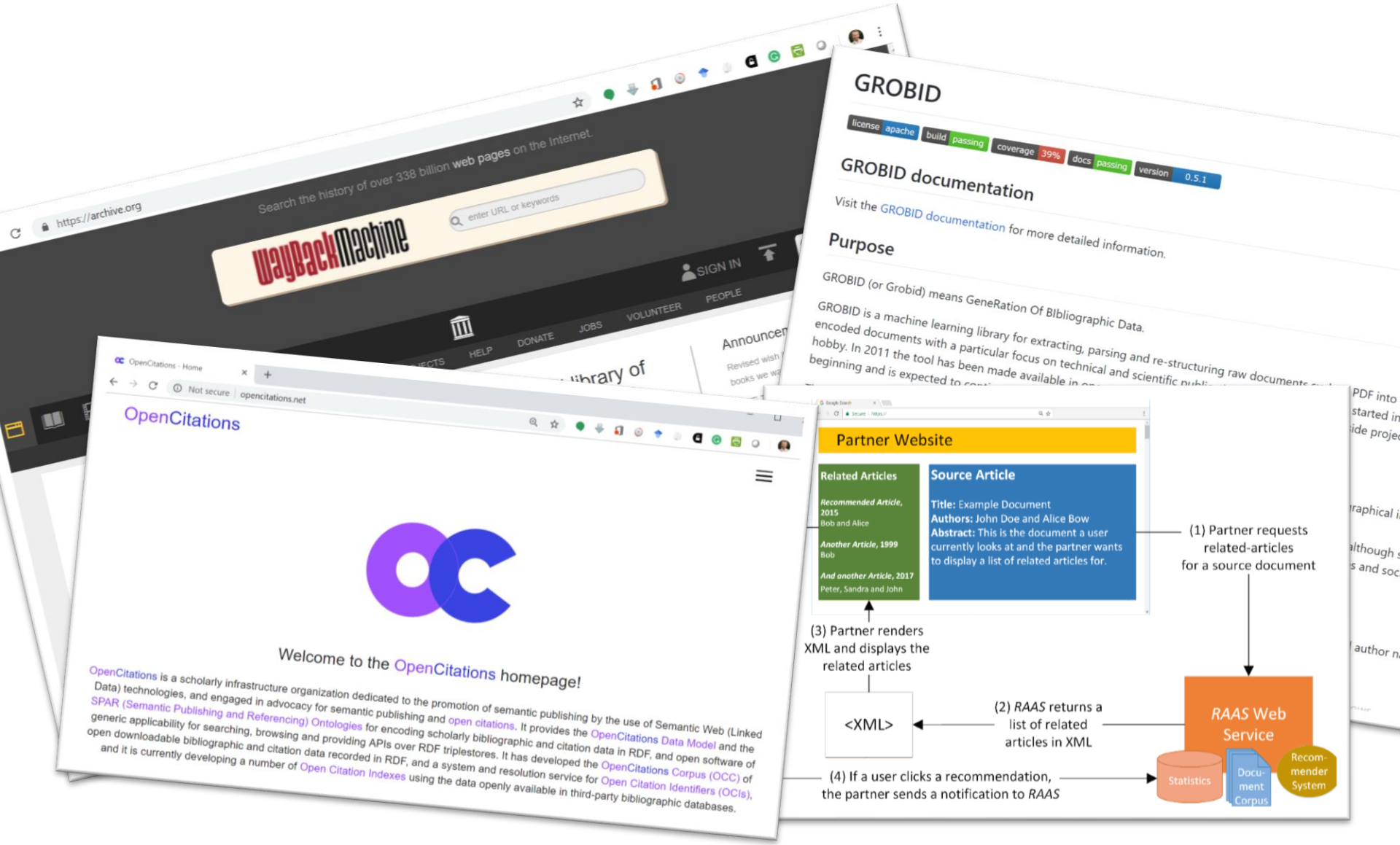
Machine Learning als Tool zum Lösen von Problemen / Unterstützen



Bibliotheken als Forschungszentren

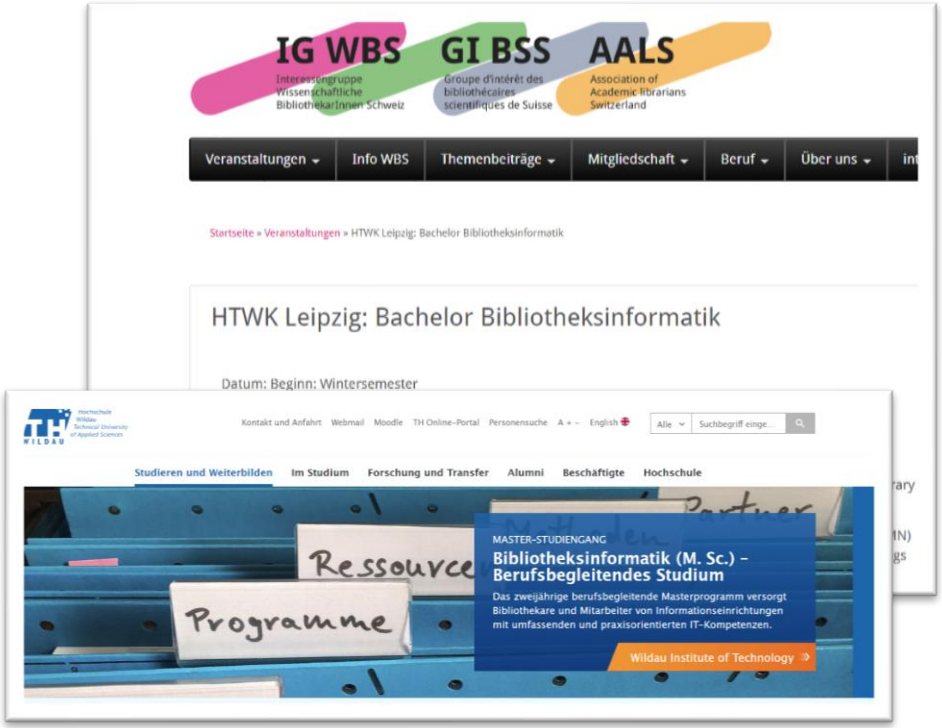


Bibliotheken als Plattformbetreiber



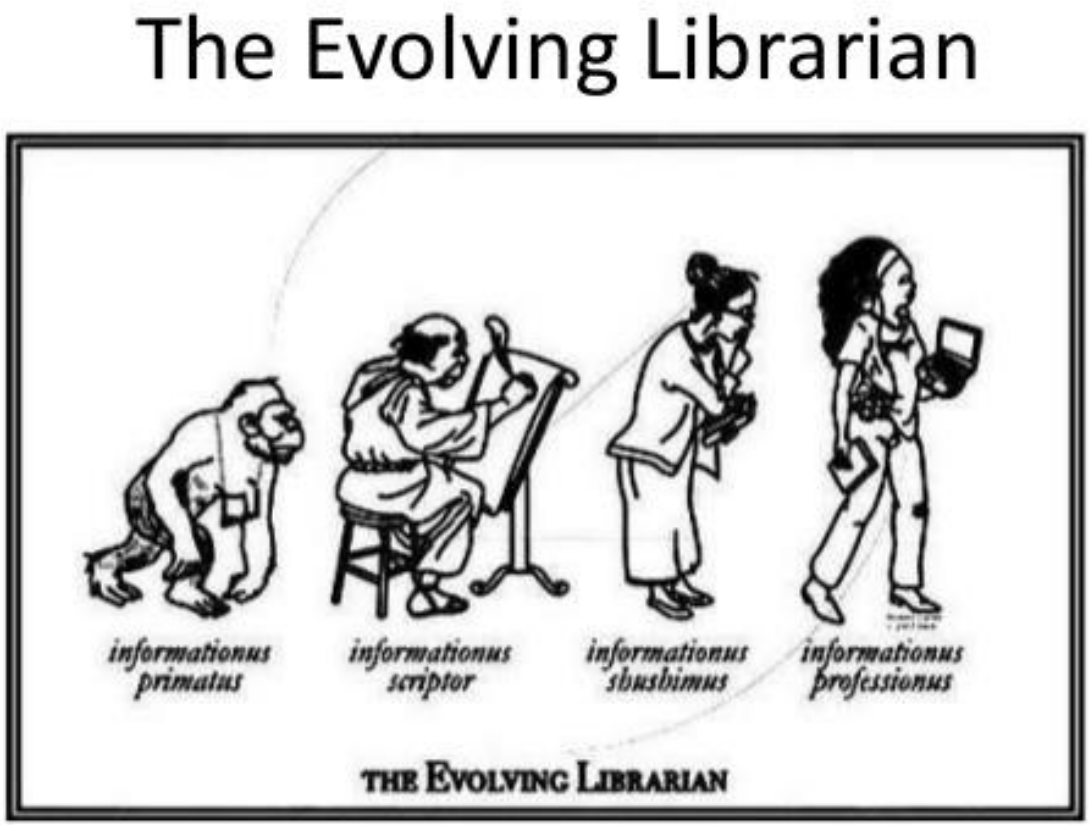
Bibliotheken als Ausbilder

- **Bibliotheksinformatiker**
- **Machine Learning im Akademischen / Bibliothekarischen Kontext**
- **Medienkompetenz: Kritisches Hinterfragen der Suchergebnisse z.B. bei Google Scholar**
- **Akademisches Marketing**
  - Akademische Suchmaschinen Optimierung
  - Pflegen von Google Scholar, ResearchGate … Profilen



„Fördert durch Beratung und Schulung die Kompetenz zur Informationssuche, -beschaffung und –weiterverarbeitung… und Informationsverbreitung“

Entwicklung



Source: A visualização da imagem anterior desperta para a leitura do artigo:
"Evolving Web, evolving librarian" de Amy and Robert Favini
balcaodebiblioteca.blogspot.com

https://image.slidesharecdn.com/suelbuhmlg04022013rev-130313090847-phpapp01/95/the-evolving-librarian-health-and-medical-librarians-in-a-changing-environment-2-638.jpg?cb=1363166232

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Fragen?

Joeran Beel
j@beel.org
www.beel.org