



# DIE INFORMATIONSWISSENSCHAFT IN HILDESHEIM

Christa Womser-Hacker

Workshop des HI

03. Februar 2020 @ Humboldt Uni zu Berlin

[www.uni-hildesheim.de/~womser](http://www.uni-hildesheim.de/~womser)

womser@uni-hildesheim.de

# WER IST DAS?

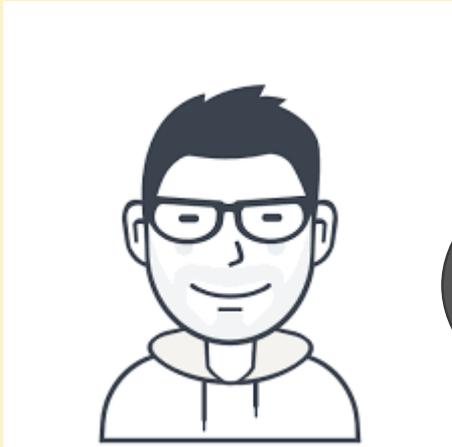
- Vier Professuren (plus 1 Juniorprofessur)
  - Griesbaum, Heid, Mandl, N.N., Womser-Hacker
- Ca. 16 MitarbeiterInnen und interne DoktorandInnen
- 430 Studierende im Bachelor, 80 im Master, seit WS 13/14 etwa 110 Masterabsolventen

# INSTITUT FÜR INFORMATIONSWISSENSCHAFT & SPRACHTECHNOLOGIE



# VERSTÄNDNIS DER INFORMATIONSWISSENSCHAFT

Fokus Mensch



technikgetrieben



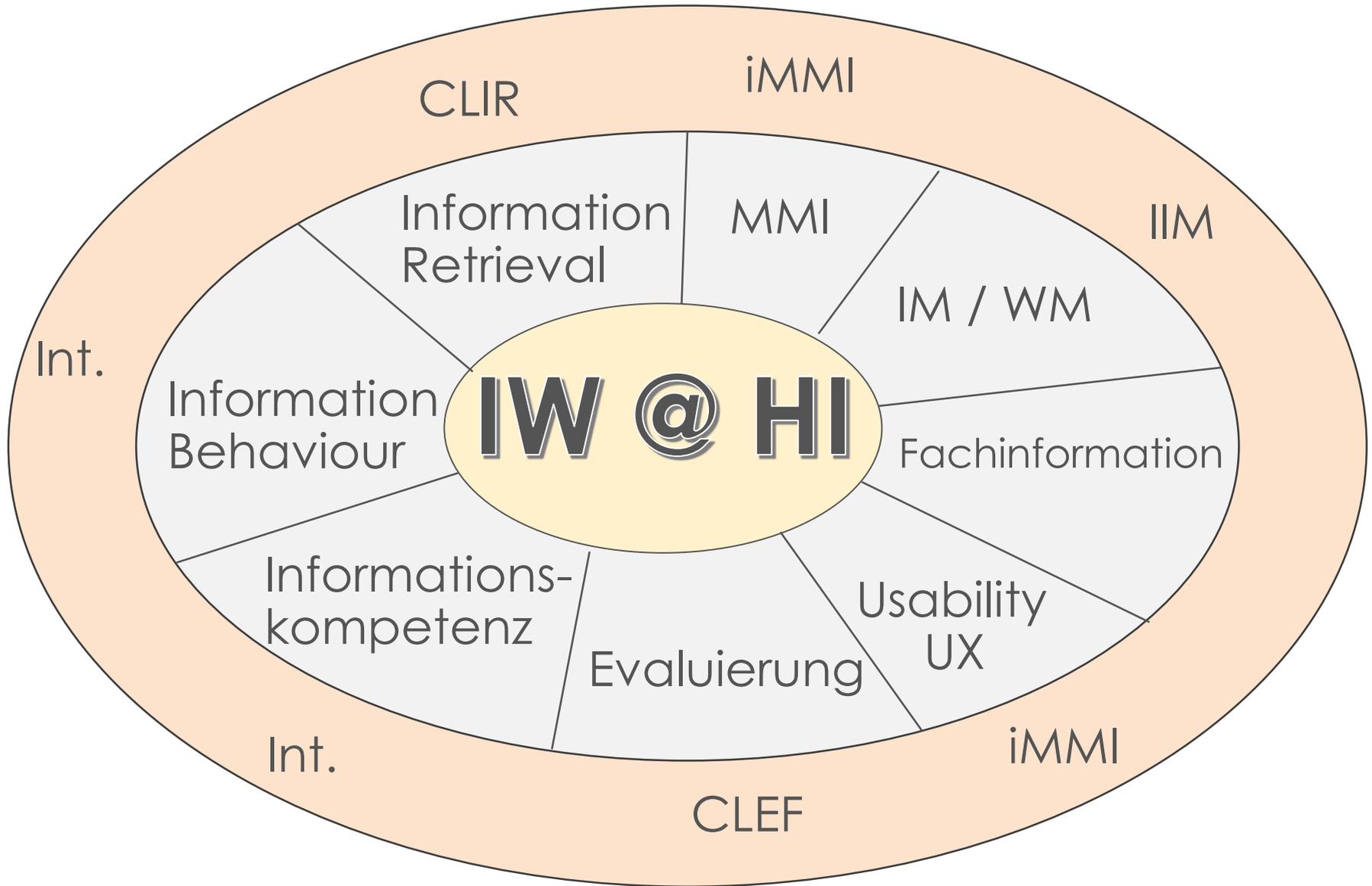
Informations-  
prozesse/  
systeme

# INTERDISZIPLINARITÄT (SYNERGIEN) ...

Informatik <ul style="list-style-type: none"><li>• Systementwicklung</li><li>• Programmierung</li></ul>	Computer-/Linguistik <ul style="list-style-type: none"><li>• Speech / Dialogsyst.</li><li>• MÜ</li><li>• Textanalyse</li><li>• Sentimentanalyse</li></ul> Korpus / Lexikon
Humanities <ul style="list-style-type: none"><li>• Nicht-digitaler Teil</li></ul>	
Interkulturelle Kommunikation <ul style="list-style-type: none"><li>• Kulturanalyse</li><li>• Diskursforschung</li></ul>	Medien... <ul style="list-style-type: none"><li>• Soziale Medien</li><li>• Medientypen</li></ul>
	Psychologie <ul style="list-style-type: none"><li>• Menschl. Kognition</li></ul>

**IW**

# Mehrsprachigkeit



# INFORMATIONSWISS. METHODEN UND INSTRUMENTE

- **Empiriefundierung**
- Information Behaviour Analyse  
→ Anforderungsanalyse
- User-centered Design
- Partizipative, iterative Systementwicklung
- Benutzerorientierte Evaluierung
- Usability / UX Labor (Eyetracking, Crowdsourcing etc.)
- ...

# STUDIENGÄNGE (AKTUELL)

- Zentrale Studiengänge:
  - Internationales Informationsmanagement (Bachelor)
  - IIM: Informationswissenschaft (Master)
- Beteiligung an: Informationsmanagement und Informationstechnologie
- Beteiligung an: Internationale Kommunikation und Übersetzen
- JDP: Global Studies in Management and Information Science (GLOMIS)

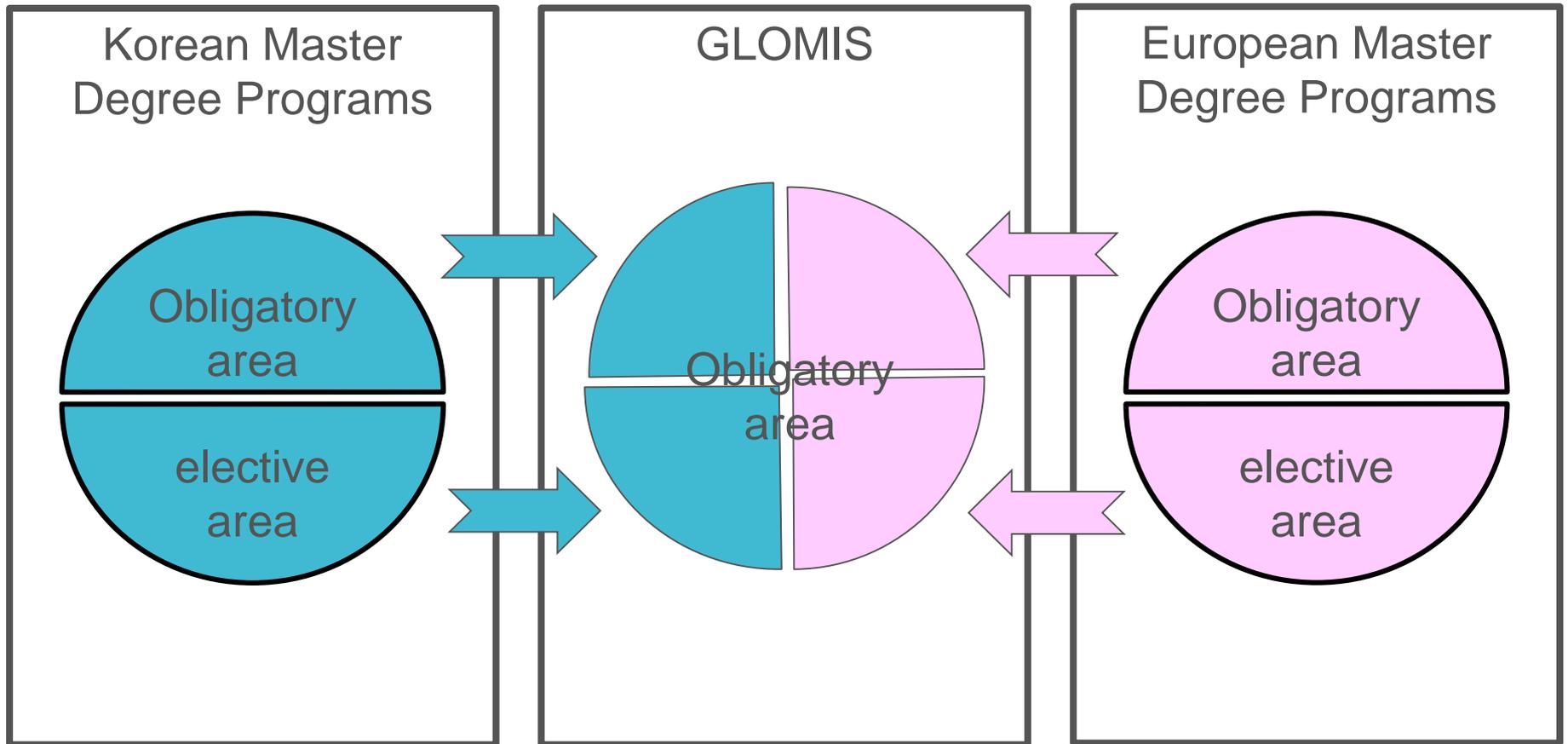


# GLOMIS - JOINT DEGREE PROGRAM

- 1 Jahr in Hildesheim oder Graz
- 1 Jahr in Korea (Paichai oder Chungbuk)
- Gemeinsame Masterurkunde beider beteiligter Universitäten
- Betreuer aus zwei Ländern
- Seit 11 Jahren durchgehend Absolventen



# GLOMIS GENERAL STRUCTURE



# NEUE STUDIENGANGSVARIANTE AB 2020

## BA Digitale Sozialwissenschaft

- IW mit Soziologie und Politikwissenschaft
- Analyse sozialer Prozesse mit digitalen Methoden
- Schwerpunkte in der IW
  - Online Marketing (für Organisationen)
  - Information Retrieval
  - Computerlinguistik

Verantwortlich: Joachim Griesbaum, Thomas Mandl

# NEUE STUDIENGANGSVARIANTE AB 2021

## **BA Global Information Management GIM**

- Gewinnung ausländischer Studienanfänger im BA IIM
- 2 Jahre Teilzeitstudium und parallel Deutschkurse
- Danach Studium auf Deutsch (und Englisch)
  
- [www.uni-hildesheim.de/gim](http://www.uni-hildesheim.de/gim)

Verantwortlich: Thomas Mandl, Ralph Kölle

# BA GIM

YEAR 1	YEAR 2	YEAR 3	YEAR 4
Lectures in English	Lectures in English	Lectures in German & English	Lectures in German
German language class	German language class & certification	Internship in a <b>German organization</b>	Bachelor Thesis

# SCHWERPUNKTE IN DER FORSCHUNG

- Informationskompetenz
  - Informationsverhalten
  - Mensch-Maschine-Interaktion / Accessibility
  - Information Retrieval
- 
- Fokus „Bild“ → Bildanalysen und Informationssuchverhalten mit Bildern
  - Digital Humanities / Erkennung von Hate Speech / Andeutung von Self Harm / Anorexia / Depression in Sozialen Medien
  - Cross-Language Information Retrieval und dessen Evaluierung
  - Immer noch oder wieder Patente ;-)

# INFORMATION(SUCH)VERHALTEN

- Kollaboratives Informationsverhalten
  - Diss Elbeshausen
- Informationsverhalten zwischen Generationen / Kulturen
  - KNOWISH-Projekt mit Gunilla Widèn und Jannica Heinström
  - K-KNOWISH mit DongBack Seo im GLOMIS Kontext
- Informationsverhalten in bestimmten Domänen (Historiker, Politikwissenschaft, Patentingenieure, Berater etc.)
  - „Welt der Kinder“ mit GEI
  - Diss Görtz
  - Vortrag Wiebke Thode im Doktorandenworkshop ECIR 2019

# ...INTERDISZIPLINÄRE PROJEKTE

- Gamification in Informationssystemen
  - EnerVation - Energie-Bewusstsein durch Transparenz, motivationale Systeme und Gamification
- Informationskompetenz
  - Informationskompetenz und Demokratie (IDE)
  - Informationskompetenz (ILO) mit Graz, London, Barcelona, Zadar, Ljubljana, DIPF (eLearning-Module)
  - Information Professionals for Civic Development in Bosnia and Herzegovina (ILI-Doc)
- Bildanalyse in Digital Humanities – Distant Viewing
  - Welt der Kinder: Schulbücher im deutschen Kaiserreich
  - Bildikonographie in Kinder- und Jugendbüchern
  - PorträtApp - Bilderkennung für frühneuzeitliche Porträtgrafik
- Hate Speech Erkennung in Sozialen Medien
  - HASOC - Shared Task
  - HASEKI - Interdisziplinärer Zukunftsdiskurs

# GAMIFICATION / MMI FÜR KINDER

**Projekt: Energie-Bewusstsein durch Transparenz, motivationale Systeme und Gamification - EnerVation**

- interaktives Vermittlungssystem mit Gamification-Elementen zum Energieverbrauch mit der Zielgruppe Kinder
- Entwicklung und Analyse reliabler Instrumente zur Messung der User Experience (UX) für Kinder

# ENERVATION (2018-2020)

- Quizze und Spiele online
- Tests in Schulen
- Partizipative Entwicklung mit Kindern
- Analyse und Optimierung von Energiespartipps
- Usability und UX für/mit Kindern

Wo nutzt du zuhause täglich Wasser? Ordne die passenden Gegenstände zu.

	Duschen oder baden	Hände waschen
	Pflanzen gießen	Trinken
	Duschen oder baden	Hände waschen
	Pflanzen gießen	Trinken
	Duschen oder baden	Hände waschen
	Pflanzen gießen	Trinken
	Duschen oder baden	Hände waschen
	Pflanzen gießen	Trinken

# INFORMATIONSWISSENSCHAFTL. FORSCHUNGSBEREICHE

- Zielgruppengerechte Informationspräsentation und Angebote
  - Webdesign und Software für Kinder
  - Kinderpartizipation und aktive Einbindung in den Entwicklungsprozess
- Adaption von Usability und User Experience Methoden Gamification und Motivation



- Informationskompetenz und Demokratie (**IDE**):  
Bürger, Suchverfahren und Analyse-Algorithmen
  - Fachtagungen
  - Diskussion mit der Zivilgesellschaft

Verantwortlich: Joachim Griesbaum

# ILI-DOC

**DAAD**

Deutscher Akademischer Austauschdienst  
German Academic Exchange Service



Univerzitet u Sarajevu



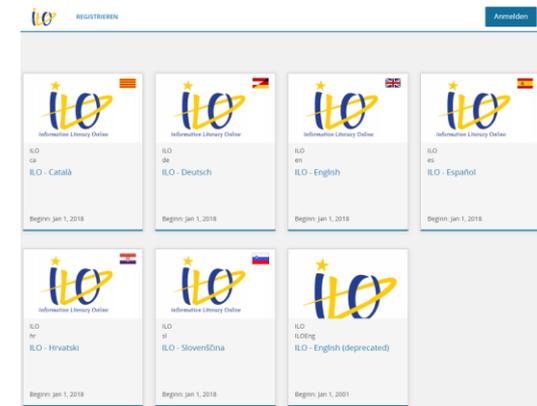
## Information Literacy and Democracy: The Role of Information Professionals for Civic Development in Bosnia and Herzegovina

- School libraries as information hub
- Best practice analysis of IL teaching
- Multiplier event for promoting IL in all 3 parts of Bosnia and Herzegovina

Verantwortlich: Thomas Mandl

# INFORMATION LITERACY ONLINE (ILO)

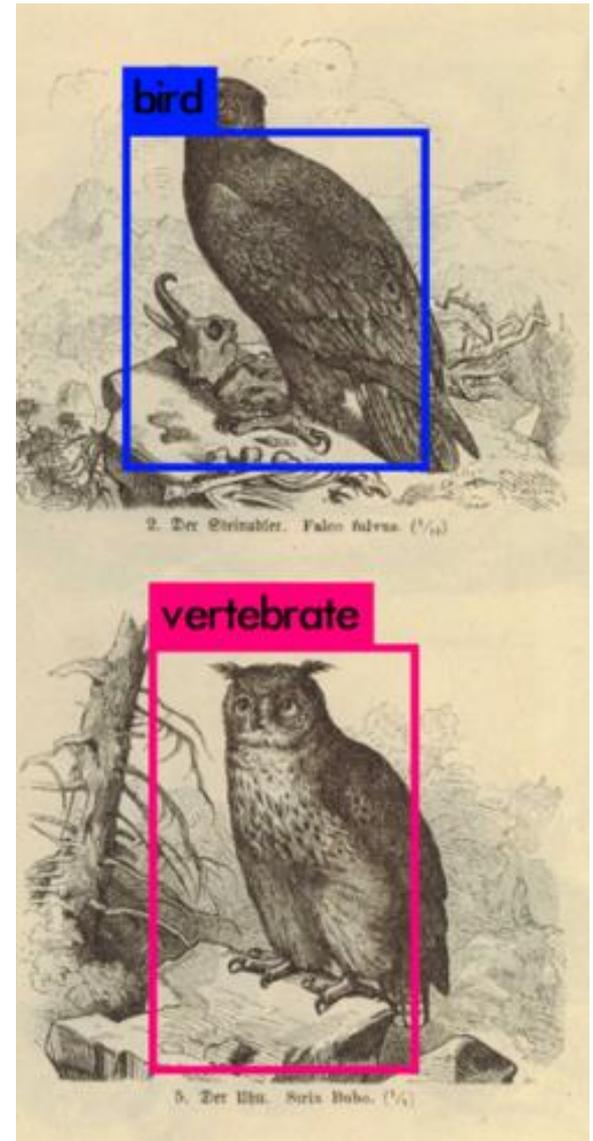
- Erasmus+ Key Action 2 Projekt gemeinsam mit DIPF, Univ. Graz, City University of London, Univ. Barcelona, Univ. Ljubljana und Univ. Zadar (2016 – 2019)
  - MOOC zur Förderung von IK von Studierenden
  - Entwicklung in 6 Sprachen
  - Self-paced jederzeit absolvierbar
  - Sämtliche Inhalte mit CC-BY Lizenz, zusätzlich zu MOOC in offenem Repository
- Laufende Folgeforschung



Verantwortlich: Stefan Dreisiebner

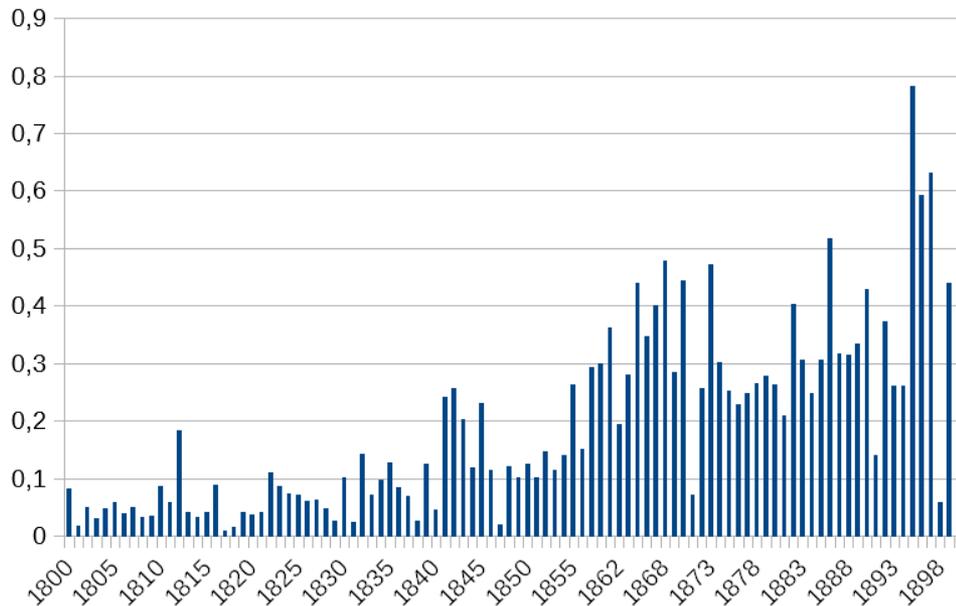
# BILDANALYSE IN DIGITAL HUMANITIES – DISTANT VIEWING

- Bildikonographie in Wissen vermittelnder Kinder- und Jugendliteratur des 19. Jahrhunderts [mit Univ. Leipzig, Stabi. Berlin, TU Braunschweig]
- PorträtApp - Bilderkennung für frühneuzeitliche Porträtgrafik [mit Herzog August Bibliothek Wolfenbüttel]

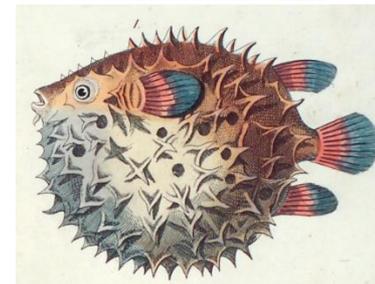
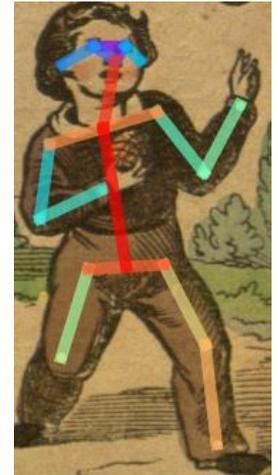


# DISTANT VIEWING

- Moderne Computer Vision Quantitative Analysen in den Kulturwissenschaften
- Algorithmen angewandt auf historische Bestände

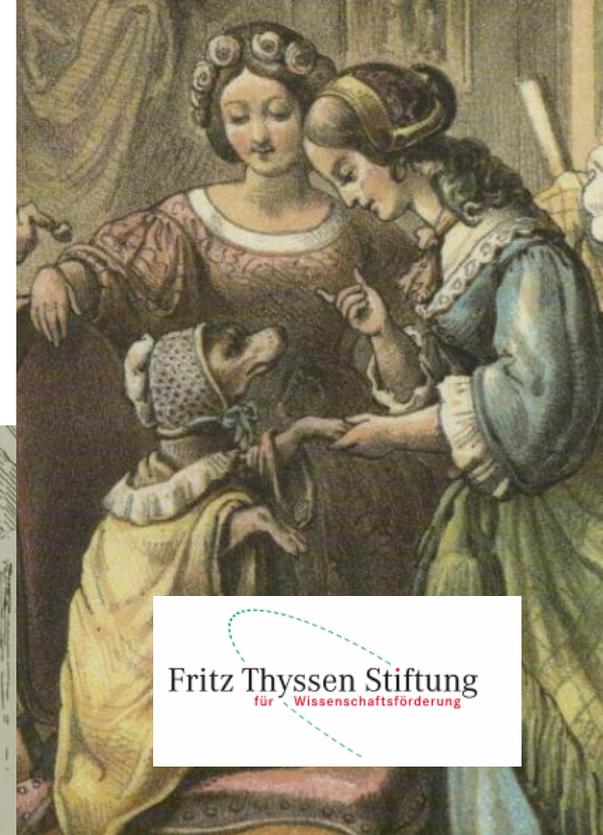
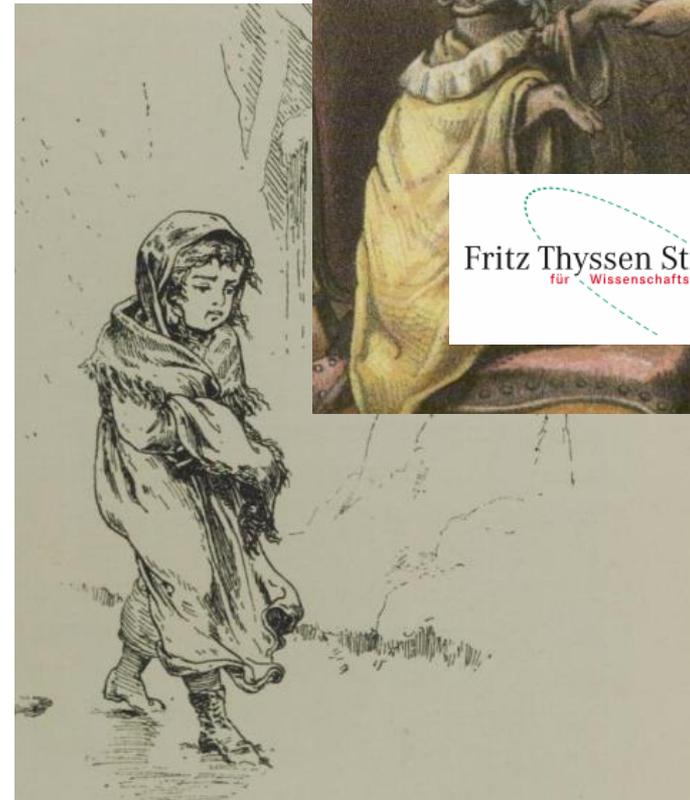


Entwicklung von  
Abbildungen  
in Kinder- und  
Jugendliteratur



# DISTANT VIEWING

- Klassifikation
  - Drucktechnologie  
(Holzschnitt, Holzstich, Kupferstich, Lithographie)
- Analyse von Wiederverwendung
  - Ähnliche Bilder
- Erkennung von Objekten
  - Klassifikation
  - Welches Wissen wurde verbreitet?
  - Nutzung von automatisch erkannten Objekten für Expertensystem

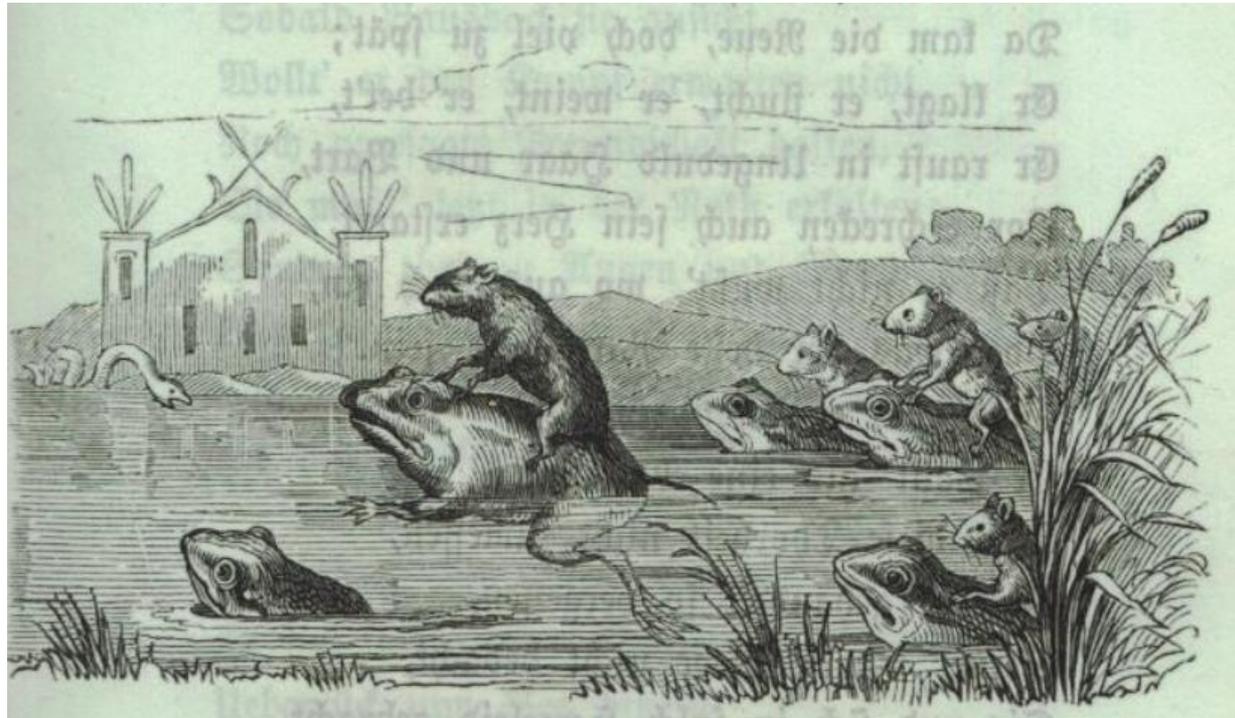


Das Jahr und was es den Kindern bringt! Düsseldorf 1861. Elkan, Bäumer & Co.  
[urn:nbn:de:gbv:084-09110211315](https://nbn-resolving.org/urn:nbn:de:gbv:084-09110211315)

Aus der Kinderwelt : ein Buch für jüngere Kinder Stuttgart,  
1908. Union Dt. Verlagsgesellschaft.  
[urn:nbn:de:gbv:084-12012713207](https://nbn-resolving.org/urn:nbn:de:gbv:084-12012713207)

# SCHWIERIGKEITEN

- Papier-Qualität
- Durchscheinen
- Verschiedene Auflösungen und Stile
- Fiktive Objekte
- Modelle auf Fotos optimiert
- ...



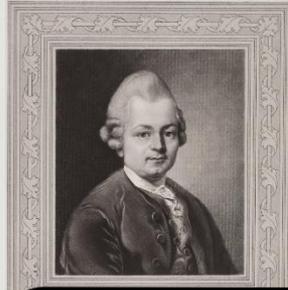
Froschmäuseler. 1861. [urn:nbn:de:gbv:084-11050917510](https://nbn-resolving.org/urn:nbn:de:gbv:084-11050917510)

# PortApp

Automatische Bilderkennung  
frühneuzeitlicher Porträtgrafik  
als App



HERZOG  
AUGUST  
BIBLIOTHEK



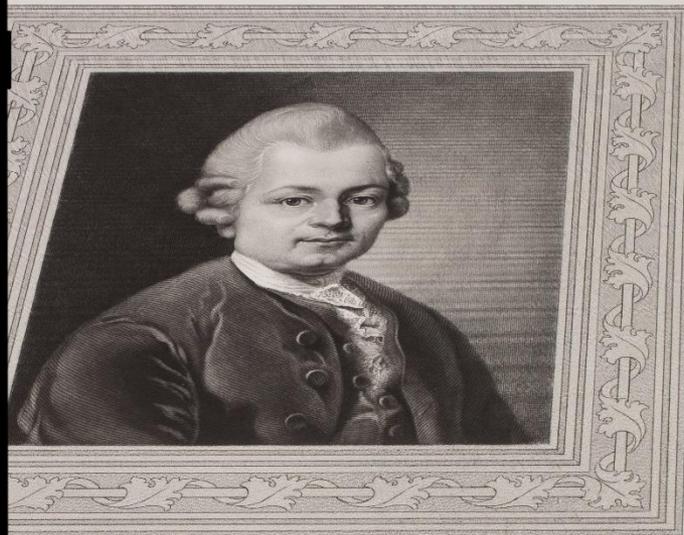
**Dargestellte Person**  
Lessing, Gotthold Ephraim

**Künstler**  
Graff, Anton (Maler Vorlage)

**Technik**  
Stahlstich

**Datierung**  
nach 1771

**Bestand**  
Herzog August Bibliothek,



*G. E. Lessing.*

HERZOG AUGUST BIBLIOTHEK  
WOLFENBÜTTEL



VolkswagenStiftung

PorträtApp:  
Automatische  
Bilderkennung  
frühneuzeitlicher  
Porträtgrafik als  
App

Verantwortlich:  
Thomas Mandl

# PORTRÄTAPP 2020-2022

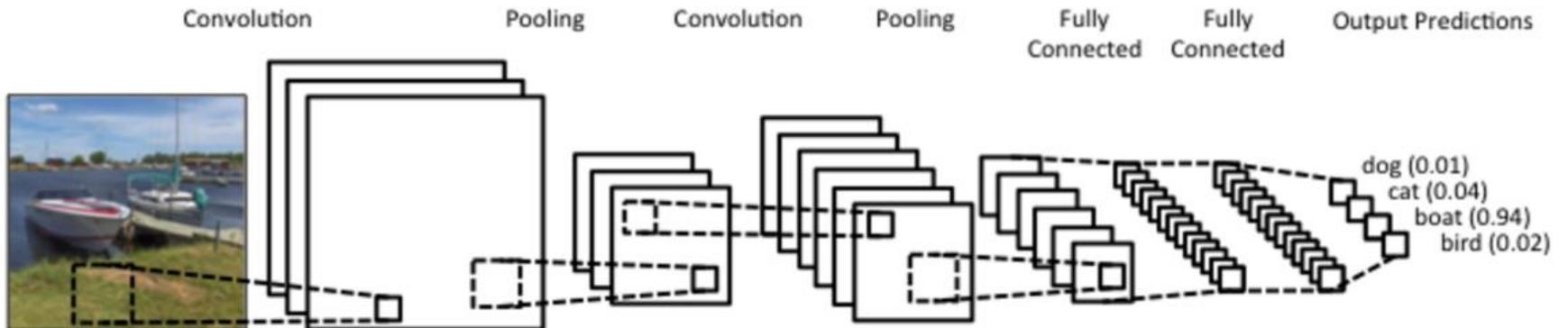
- Einsatz von Deep Learning Algorithmen
- Erkennen von Traditionslinien in Autoren-Porträts der frühen Neuzeit
- DH Tools für Porträt-Forschung
- App für die Suche nach sehr ähnlichen Porträts



# CONVOLUTIONAL NEURAL NETWORK (CNN)



Scholar 0.01  
King 0.8  
Priest 0.1  
...



# TEXT KLASSIFIKATION

- Hate Speech Erkennung
  - Organisation Shared Task HASOC: Hate Speech and Offensive Content Identification in Indo-European Languages
  - Interdisziplinäres Projekt: Hate Speech und seine Erkennung durch KI – interdisziplinär – international – erklärbar? (HASEKI)





# HASOC TRACK AT FIRE 2019

## HATE SPEECH AND OFFENSIVE CONTENT IDENTIFICATION IN INDO-EUROPEAN LANGUAGES



Thomas Mandl<sup>1, 2</sup>, Sandip Modha<sup>2</sup>, Prasenjit Majumder<sup>2</sup>,  
Daksh Patel<sup>3</sup>, Mohana Dave<sup>4</sup>, Chintak Mandlia<sup>5</sup>, Aditya  
Patel<sup>3</sup>

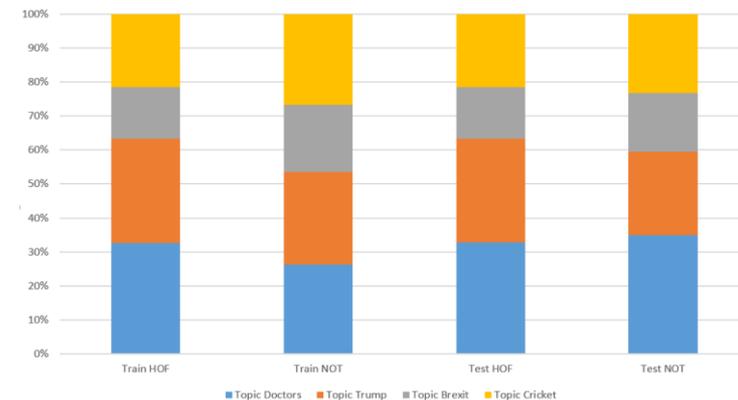
<sup>1</sup> University of Hildesheim, Germany, <sup>2</sup> DA-IICT, Gandhinagar, India,  
<sup>3</sup> Dalhousie University Halifax, Canada, <sup>4</sup> LDRP-ITR, Gandhinagar, India,  
<sup>5</sup> infoAnalytica Consulting Pvt. Ltd

# HASOC SHARED TASK

# HASOC

Content Identification

- Offensive content detection in social media
- multilingual, multi-level, multi-class classification problem for in three languages
  - English
  - German
  - Hindi
- Organized at FIRE – India
  - Great interest worldwide
- Analysis of best algorithms
  - Further research on bias and reliability



**FIRE 2019**

Forum for Information Retrieval Evaluation  
Indian Statistical Institute, Kolkata  
12th - 15th December

<https://hasocfire.github.io>

# RESULTS: GERMAN SUB-TASKS

Standing	Team name	Run_no	Marco F1	Weighted F1	
1	Hate Monitors [48]	1	0.6162	0.7915	← BERT
2	LSV-UdS [10]	1	0.6064	0.7997	
3	3Idiots [27]	1	0.5774	0.7887	← BERT
4	NITK-IT_NLP	1	0.5739	0.6796	
5	DLRG [41]	2	0.5519	0.7566	
6	CS	1	0.5506	0.7131	
7	BRUMS [42]	1	0.5464	0.787	

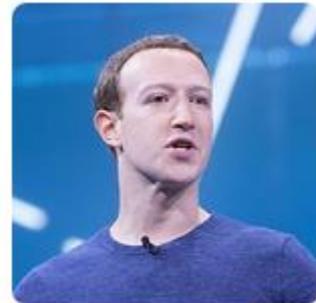
Standing	Team name	Run_no	Marco F1	Weighted F1	
1	LSV-UdS [10]	1	0.3468	0.7749	← BERT+ext.res.
2	HateMonitors [48]	1	0.2769	0.7537	
3	3Idiots [27]	2	0.2758	0.7779	← BERT
4	CS	1	0.274	0.757	
5	FalsePositive [22]	3	0.268	0.7458	

Modha, Sandip; Mandl, Thomas; Majumder, Prasenjit; Patel, Daksh (2019): **Overview of the HASOC track at FIRE 2019: Hate Speech and Offensive Content Identification in Indo-European Languages**. In: Proceedings of the 11<sup>th</sup> annual meeting of the Forum for Information Retrieval Evaluation. pp. 167-190. <http://ceur-ws.org/Vol-2517/>

# NEUES PROJEKT 2020-2021

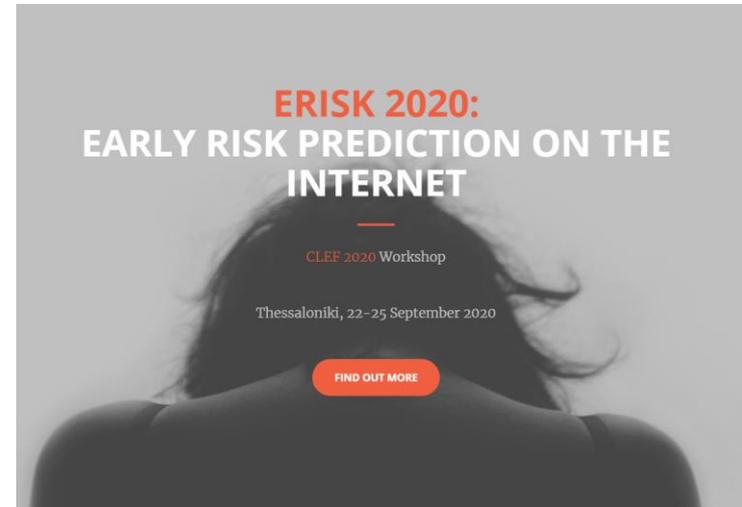


- Das Phänomen Hate Speech und seine Erkennung durch KI - interdisziplinär – international – erklärbar? (HASEKI)
- Kooperation mit Sprachwissenschaft und Politikwissenschaft
- Ziele
  - Analyse von Evaluierungsdaten u.a. auf Bias
  - Schaffung digitaler Souveränität durch Online Prüfsystem
  - Testen von Verfahren der Explainable AI
  - Vorschläge zur politischen Regulierung von Content Moderation Systemen
  - Vermittlung des Wissensstands an die Zivilgesellschaft
  - Öffentlich zugängliche Forschung zu gesellschaftlich relevanter Thematik



# ERISK – EARLY RISK PREDICTION ON THE INTERNET

- CLEF Task
  - eRisk erforscht
    - die Evaluationsmethoden,
    - Effektivitätsmetriken und
    - praktische Anwendungen  
(insbesondere in Bezug auf  
Gesundheit und Sicherheit)
- ... der Risikofrüherkennung im  
Internet.



# ERISK TASKS/CHALLENGES 2020

- Task 1: Früherkennung von Anzeichen von Selbstverletzung
- Task 2: Messung der Schwere der Anzeichen einer Depression
  
- Tasks vergangener Jahre:
  - Früherkennung von Anzeichen einer Depression
  - Früherkennung von Anzeichen von Magersucht

# DATEN UND ZIELE

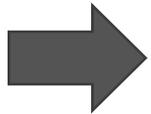
- Datengrundlage:
  - Social Media Posts von erkrankten und nicht erkrankten Personen

```
<WRITING>  
  TITLE /  
  <DATE>2015-10-27 06:52:23</DATE>  
  <TEXT>I don't drink, so I guess  
  that's PJ's by default. I can't  
  wait :) haha</TEXT>  
  <INFO>Reddit post</INFO>  
</WRITING>
```

- Ziel
  - Task 1: Möglichst frühe Erkennung einer Erkrankung: Systembeurteilung nach jedem Post
  - Task 2: Möglichst korrekte Bestimmung des Schweregrads einer Depression: Systembasiertes Ausfüllen eines standardisierten Fragebogen der Psychologie

# EINBEZIEHUNG VON PATENTABBILDUNGEN IN DEN RETRIEVALPROZESS

- Bisher ging es bei der Patentsuche im Wesentlichen um **Text** und um die **IPC**
  - Favorisierung von bestimmten Bestandteilen (Claims, Description etc.)
  - Lösen der Benennungsproblematik
  - Klassifikation
  - Früheres Projekt **TRENDMINING in Patents**



Bilder nehmen eine wichtigere Rolle im Informationsprozess ein

- Analyse des Informationsverhaltens von Patentspezialisten
- Entwicklung von Anforderungen für die Systemfunktionalität
- Prototypentwicklung und -evaluierung

# BEISPIELE

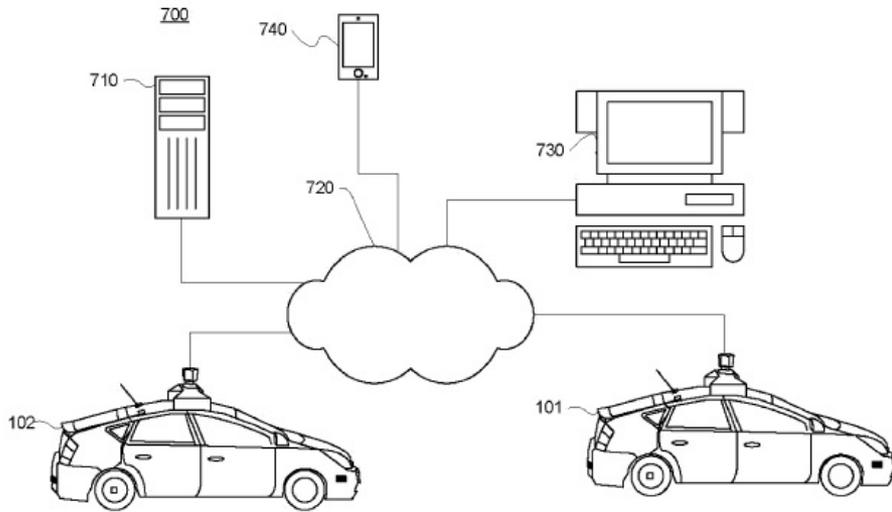
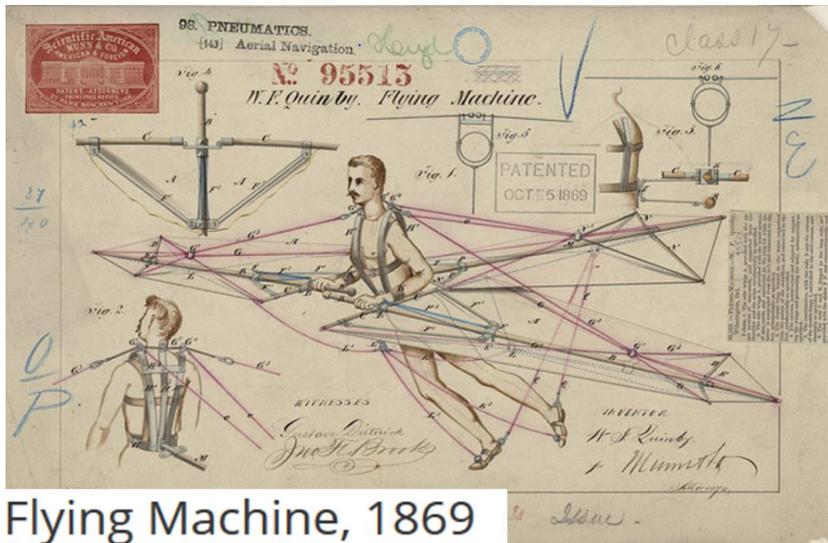
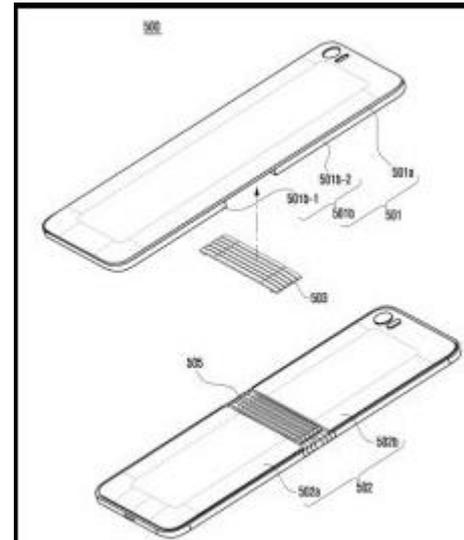
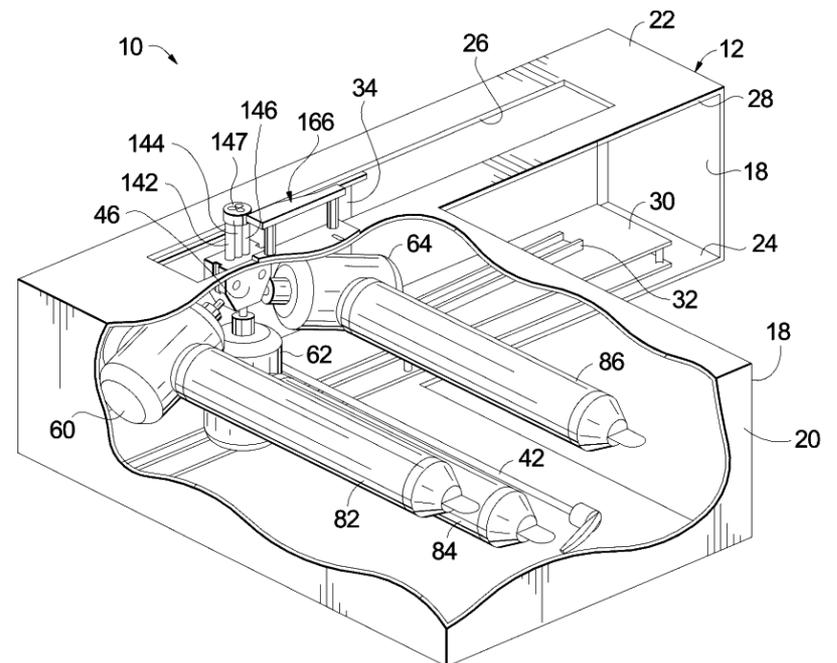


FIGURE 7



Flying Machine, 1869



# MEINE PERSÖNLICHEN HIGHLIGHTS

- Forschung mit Doktoranden und Master-Studierenden
- Interdisziplinarität macht Spaß
- Direkter Bezug zur „Informationsrealität“:  
Kooperation mit den  
Infrastruktureinrichtungen (DIPF, FIZ Karlsruhe,  
TIB etc.) sowie Industrie
- Internationalisierung
  - ENWI → Kooperation mit den  
skandinavischen Standorten
  - Kooperation mit den südosteuropäischen  
Standorten
  - Kooperation mit Korea
  - CLEF

# HERAUSFORDERUNGEN

- Rekrutierung neuer MitarbeiterInnen für Lehre und Projekte wird immer schwieriger
- WZVG (Beschäftigungsverhältnisse)
- Interdisziplinarität ist sehr erwünscht, macht aber z.T. Probleme (z.B. Fächerkultur, Begutachtung)
- Wachsende Internationalisierung (in der Verwaltung noch schwierig)

# DANKE

... an das Publikum!

... für die Bereitstellung von Folien an:

Linda Achilles

Stefan Dreisiebner

Chan Jong Im

Thomas Mandl

Sophie März

Wiebke Thode

Lea Wöbbekind