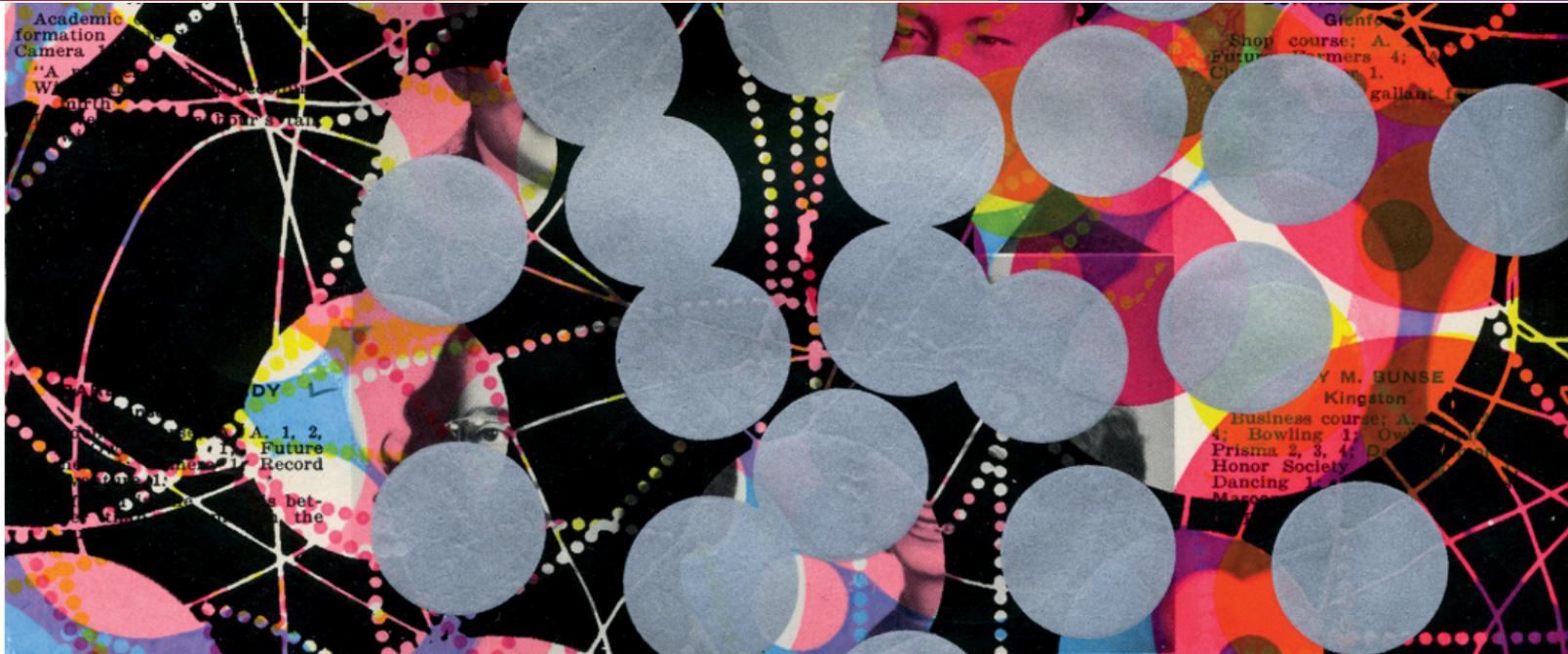




Forschung als Brückenschlag zwischen Datenwissenschaft und Informationswissenschaft

Philipp Schaer, Technische Hochschule Köln, Cologne, Germany

Version: 2020-02-03



ARTWORK: TAMAR COHEN, ANDREW J BUBOLTZ, 2011, SILK SCREEN ON A PAGE FROM A HIGH SCHOOL YEARBOOK, 8.5" X 12"

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

Summary Save Share Comment 13 Text Size Print \$8.95 Buy Copies

WHAT TO READ NEXT



Big Data: The Management Revolution

The Story so far...

2016: Fachhochschule → **Technische Hochschule**

2016-2020: **8 Neuberufungen**

- 3 Informatiker, 2 Medienw., 1 German., 1 Psych., 1 Bibliothekarin
- 1 gemeinsam mit der Fakultät für Informatik / 2 mit ZB MED

2017: Neugründung **Forschungsinstituts IIM + iLab**

2018: Neuer **Studiengang Data and Information Science**

2019: **Forschungsschwerpunktes Knowledge Discovery**

Institut für Informationswissenschaft

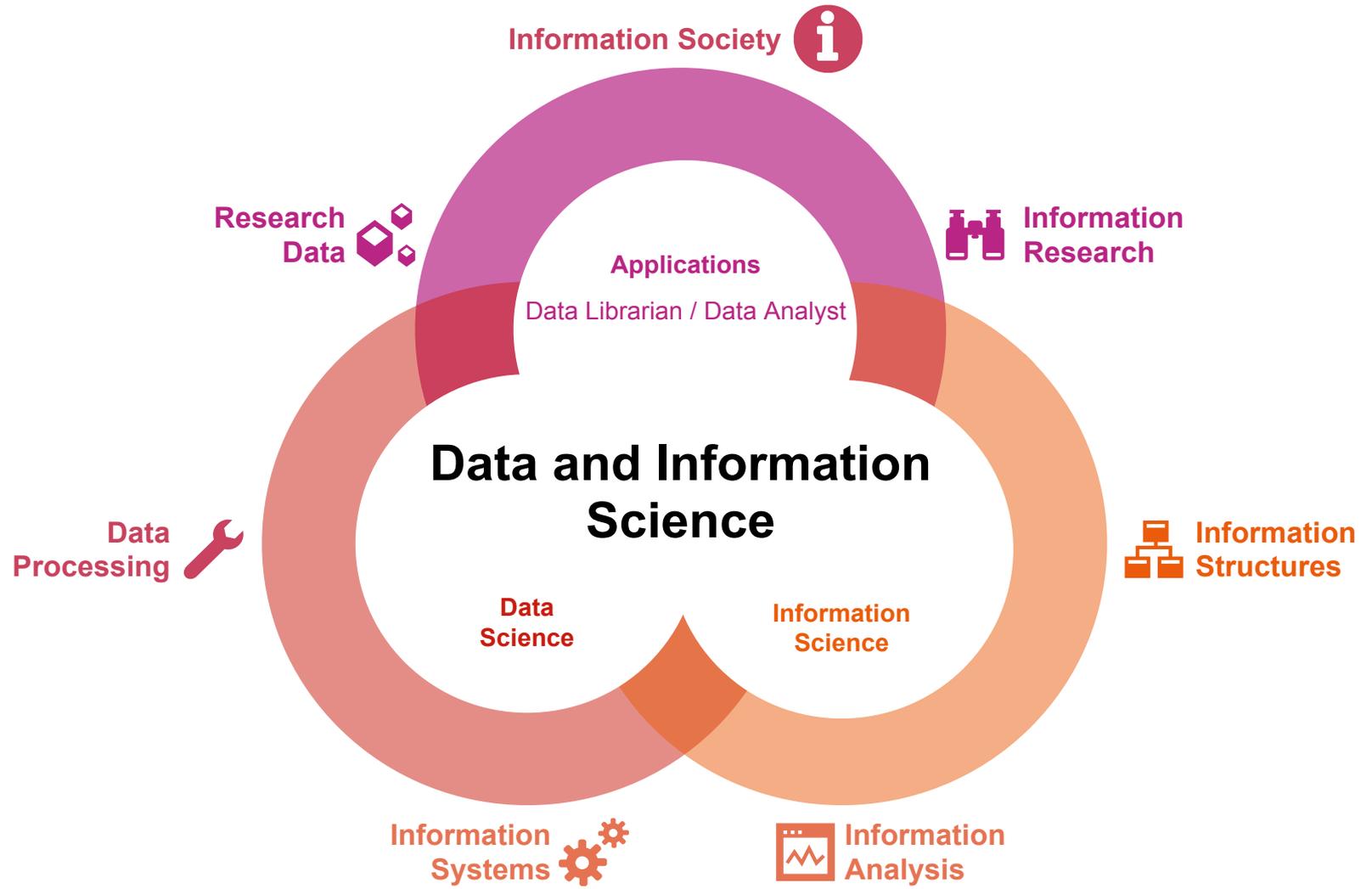
Bibliothek und digitale Kommunikation (BA)

Data and Information Science (B.Sc.)

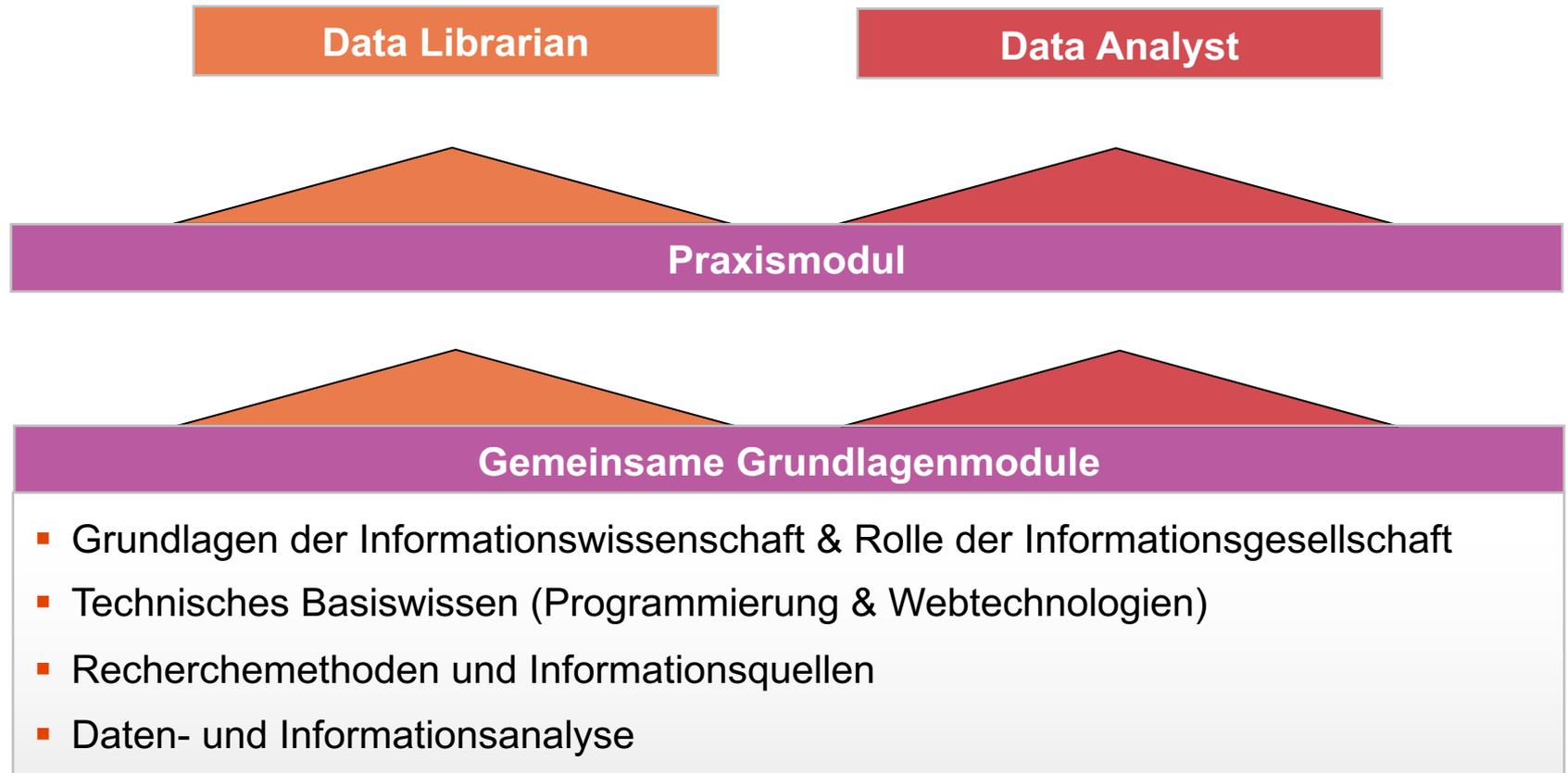
Online-Redaktion (BA)

Master of Library and Information Science (MA) – berufsbegl.

Markt- und Medienforschung (MA)



Studieninhalte und Studienschwerpunkte

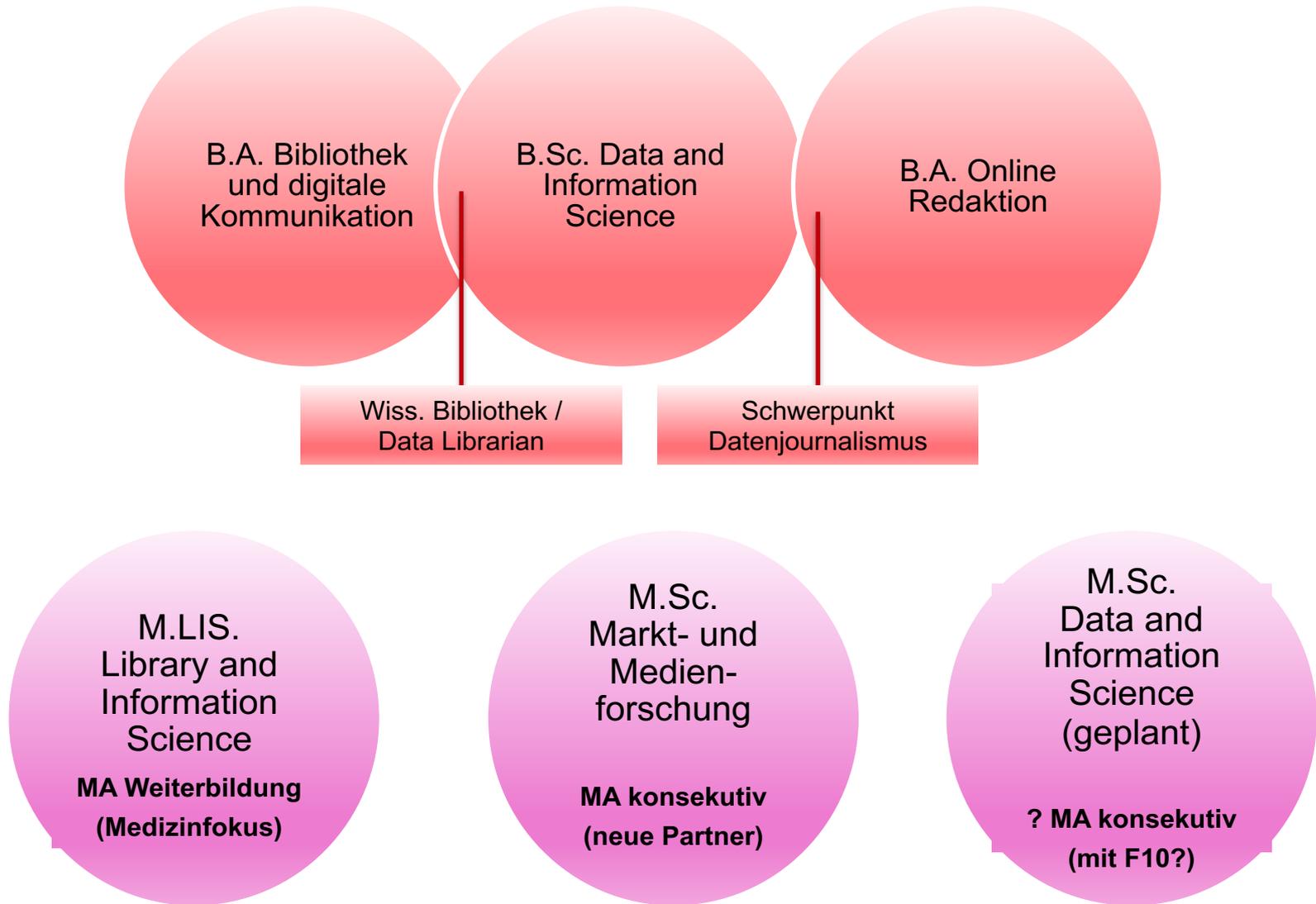


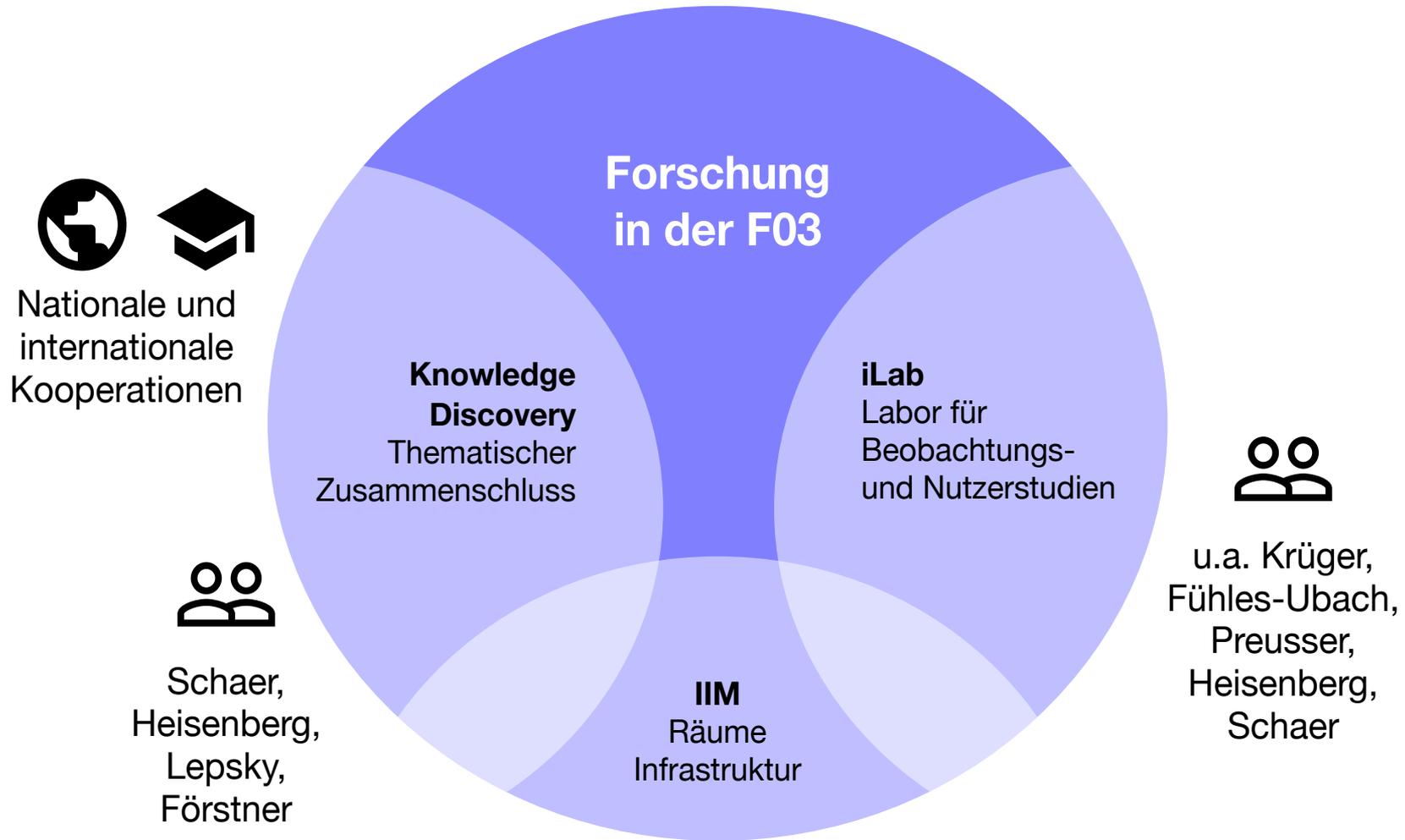
Modulübersicht

Data and Information Science – Studienverlaufsplan – Data Librarian

1. Semester	Programmierung • Webentwicklung	Informationserschließung • Wissensrepräsentation	Digitale Informationsgesellschaft Informationsethik Berufsfelderkundung (Profil2)	Informationsvisualisierung	Information in Unternehmen
2. Semester	Programmierung • Softwareentwicklung	Information-erschließung Strukturierte Dokument- beschreibung	Datenmodellierung	Statistische Datenanalyse	Informationsquellen • Informationsrecherche
3. Semester	Informationssysteme • Content- & Dokumenten- managementsysteme	Information Retrieval	Datenbanksysteme	Data Mining	Informationsanalyse
4. Semester	Praxismodul	Praxismodul	Praxisphase	Praxisphase	Praxisphase
5. Semester	Suchmaschinentechnologie • Webtechnologien	Projektarbeit I <i>Data Librarian</i> Interdisziplinäres Projekt	Informationsrecht & Datenschutzrecht • Wissenschaftliches Arbeiten	ASC • Studienportfolio	Information Consultancy, Wissenschaftskommunikatio n & Wissenschaftspolitik <i>Data Librarian</i>
6. Semester	Projektarbeit II <i>Data Librarian</i>		Forschungsdaten I <i>Data Librarian</i> Lizenzmanagement, Digitales Publizieren & Open Access	Informetrie, Bibliometrie, Szientometrie Empirische Forschungsmethoden	Automatische Erschließung <i>Data Librarian</i>
7. Semester		Forschungsdaten II <i>Data Librarian</i> • Digitalisierung & Langzeitarchivierung	Bachelorarbeit — Seminar zur Bachelorarbeit	Bachelorarbeit	Bachelorarbeit

Studienreformen Bachelor 2017/2018





RheinEnergie
Stiftung | **Jugend | Beruf**
Wissenschaft



Bundesministerium für Bildung und Forschung



Bundesministerium für Gesundheit



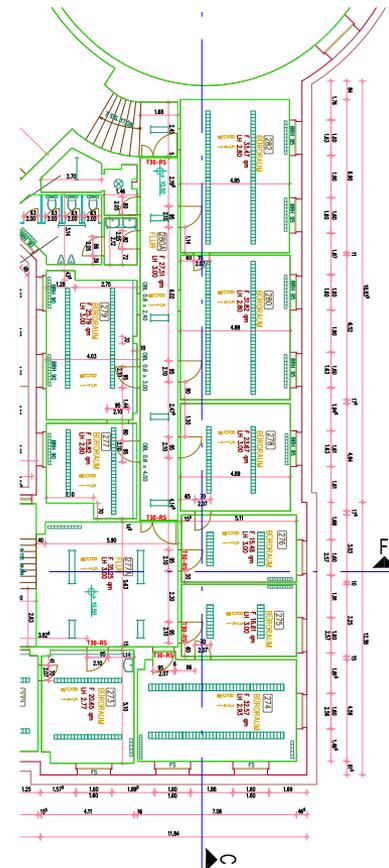
Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen



Institute of Information Management (IIM)

Forschungsinstitut der Fakultät für Informations- und Kommunikationswissenschaft

- Beherbergt alle **Drittmittelprojekte** > 20.000 EUR
- Keine Dienstleistungs- oder Studiengangspunkte, wie ZBIW o. MALIS (bleiben am IWS)
- Eigene **Räume, Co-Working-Spaces** und **Budget**
- “Inkubator” für **drittmittelgeförderte** Forschung
- **Promotionsstudierende** unter einen Dach
- **Gemeinsame Infrastruktur und Services**
 - Groupware-Server, High-Performance-Server-Kapazitäten, Daten- und Lizenzen,
 - Anträge, Förderempfehlungen, etc. ...



Drittmittelprojekte am IIM seit 2016

Drittmittelprojekte am IIM seit 2016

- **Search and Indexing for DLR.de** (Lepsky)
- **Smart Harvesting - WebScrapers for Digital Libraries** (Schaer)

DFG



Drittmittelprojekte am IIM seit 2016

- **PRIOR** - Prepublication Radar (Schaer)
- **PREFECT** - prediction models for famine catastrophes (Heisenberg)
- **ESUPOL** - Political Influence of Search Engines (Schaer)
- **Search and Indexing for DLR.de** (Lepsky)
- **Smart Harvesting** - WebScrapers for Digital Libraries (Schaer)

Drittmittelprojekte am IIM seit 2016

- **Data Literacy** - Lehrkonzept zur Datenkompetenz (Fühles-Ubach)
- **UseAltMe** - Altmetrics on the article level (Tunger)
- **GeKo:mental** - Mental health literacy (Werner)
- **InTraHealth** - Improving access to health care for intersex and transgender people (Werner)
- **STELLA** - Infrastructures For Living Labs - Reproducibility (Schaer)
- **PRIOR** - Prepublication Radar (Schaer)
- **PREFECT** - prediction models for famine catastrophes (Heisenberg)
- **ESUPOL** - Political Influence of Search Engines (Schaer)
- **Search and Indexing for DLR.de** (Lepsky)
- **Smart Harvesting** - WebScrapers for Digital Libraries (Schaer)

iLab @ TH-Köln

Das iLab steht für verschiedene Ideen und Konzepte

- **Information**, das zentrale Forschungssujet
- **Interaktion**, also das Zusammenspiel von Nutzer und Erzeuger
- **Innovation**, als Motor

Labor für Nutzer-, Beobachtungs- und Innovationsstudien

- **Komplett flexibles Raumkonzept**

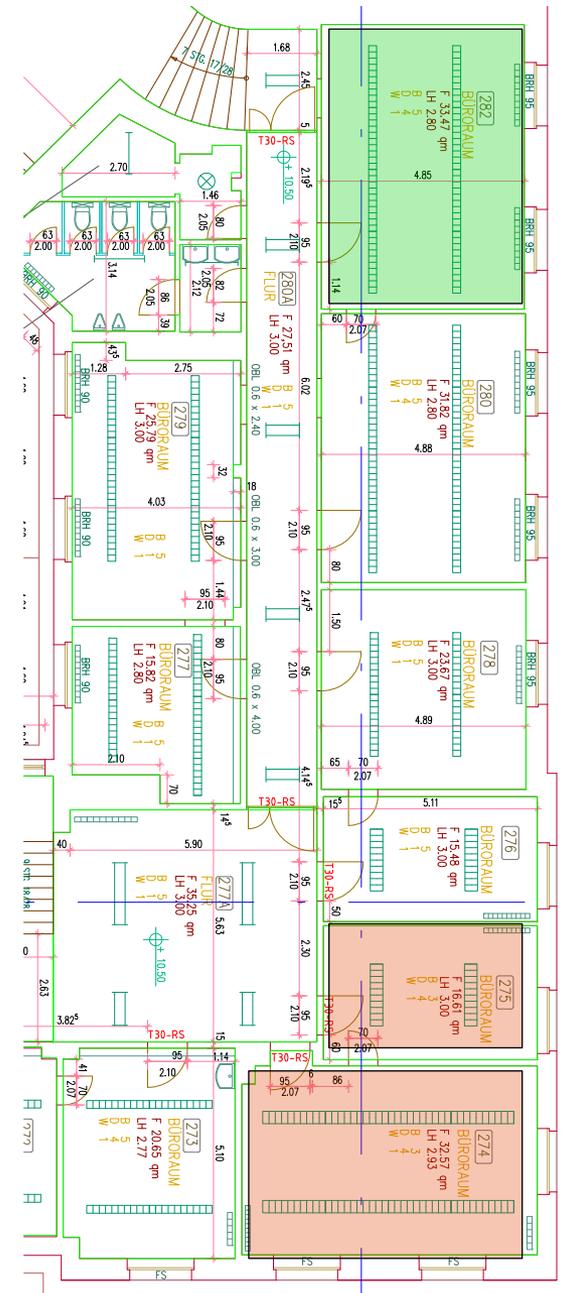
iLab @ TH-Köln

iLab-Innovation

- Ein-Raum-Setup
- Design-Thinking-Workshops

iLab-Observation / iLab-Control

- Zwei-Raum-Setup
- Eye-Tracking/Biosensoren (EEG, Puls, ...)
- Stationäre und mobile Rechner
- Usability, Nutzerstudien und Prototyp-Evaluierungen



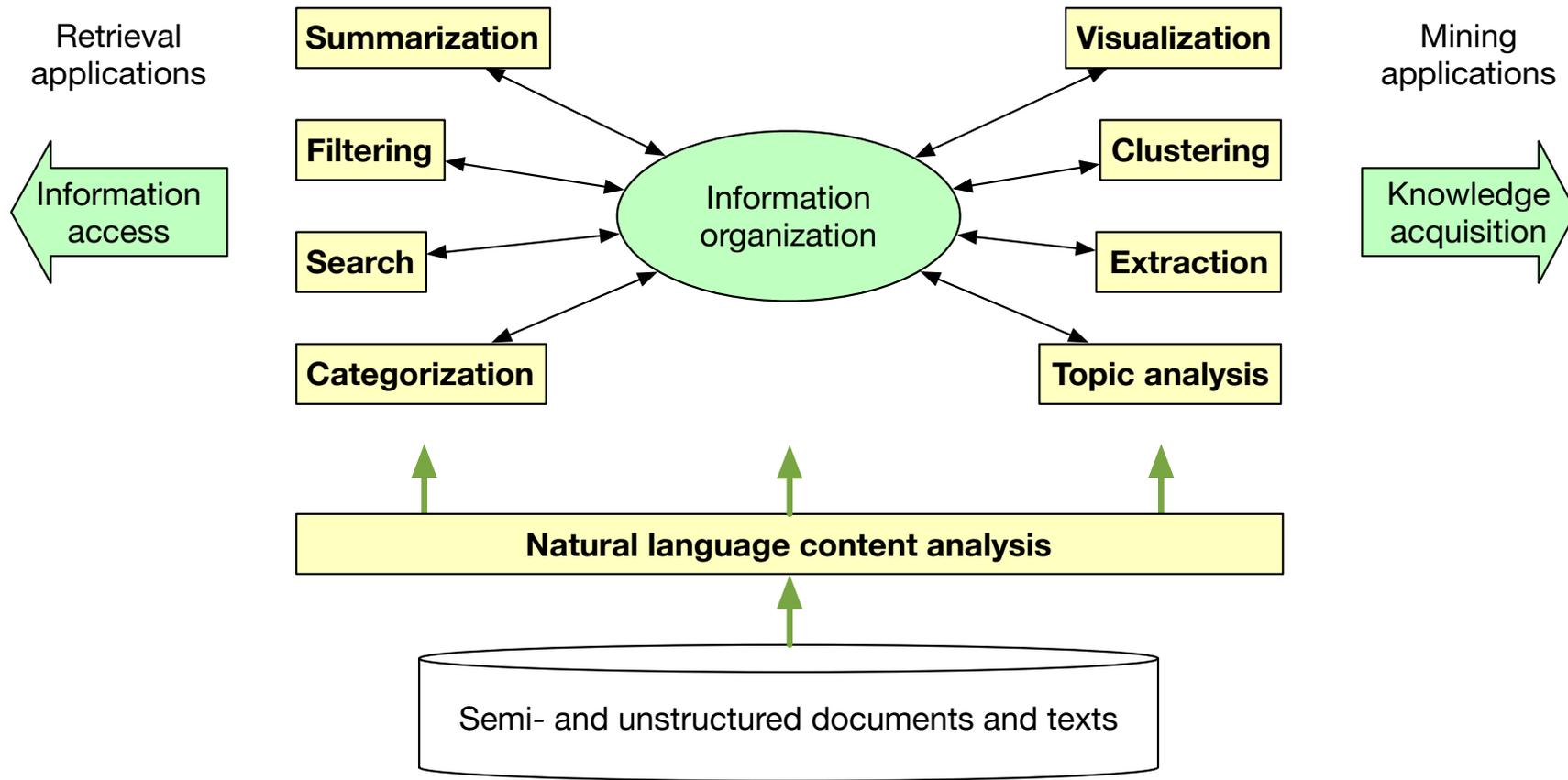
Neuer Forschungsschwerpunkt KD

2019: TH Köln fördert die Gründung eines neuen **Forschungsschwerpunktes** “Knowledge Discovery”

Aktuell 4 beteiligte Professuren

- Information Retrieval
- Information Research
- Informationserschließung
- Information and Data Literacy

Knowledge Discovery als Überbau



Suchvorschläge/Auto-Vervollständigung



ang

- angies list
- angel
- angelina jolie
- angina
- angela kinsey
- angela bassett
- angiogram
- angry birds
- angioedema
- angler fish

Google Search I'm Feeling Lucky




angela

- angela kinsey
- angela bassett
- angela merkel
- angela davis
- angela's christmas
- angela rye
- angela simmons
- angela bassett age
- angela rummans
- angela yee

Google Search I'm Feeling Lucky

[Report inappropriate predictions](#)



angela merk

- angela merkel
- angela merkel news
- angela merkel bush funeral
- angela merkel husband
- angela merkel net worth
- angela merkel biography
- angela merkel twitter
- angela merkel approval rating
- angela merkel age
- angela merkel g20

Google Search I'm Feeling Lucky

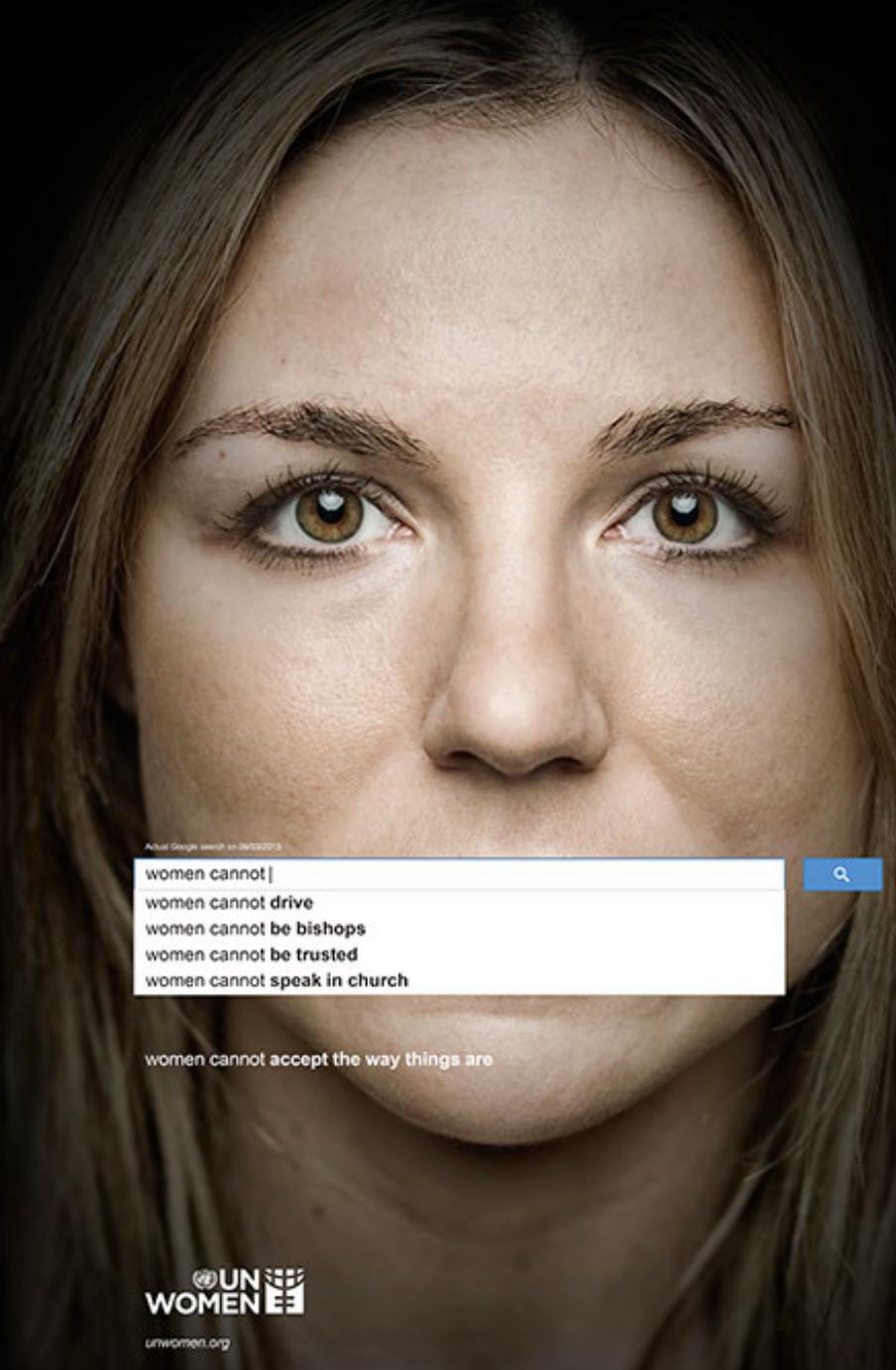
[Report inappropriate predictions](#)

angela merkel

- angela merkel
- angela merkel news
- angela merkel bush funeral
- angela merkel husband
- angela merkel net worth
- angela merkel biography
- angela merkel twitter
- angela merkel approval rating
- angela merkel age
- angela merkel g20

Google Search I'm Feeling Lucky

[Report inappropriate predictions](#)



Actual Google search on 06/03/2013

women cannot |

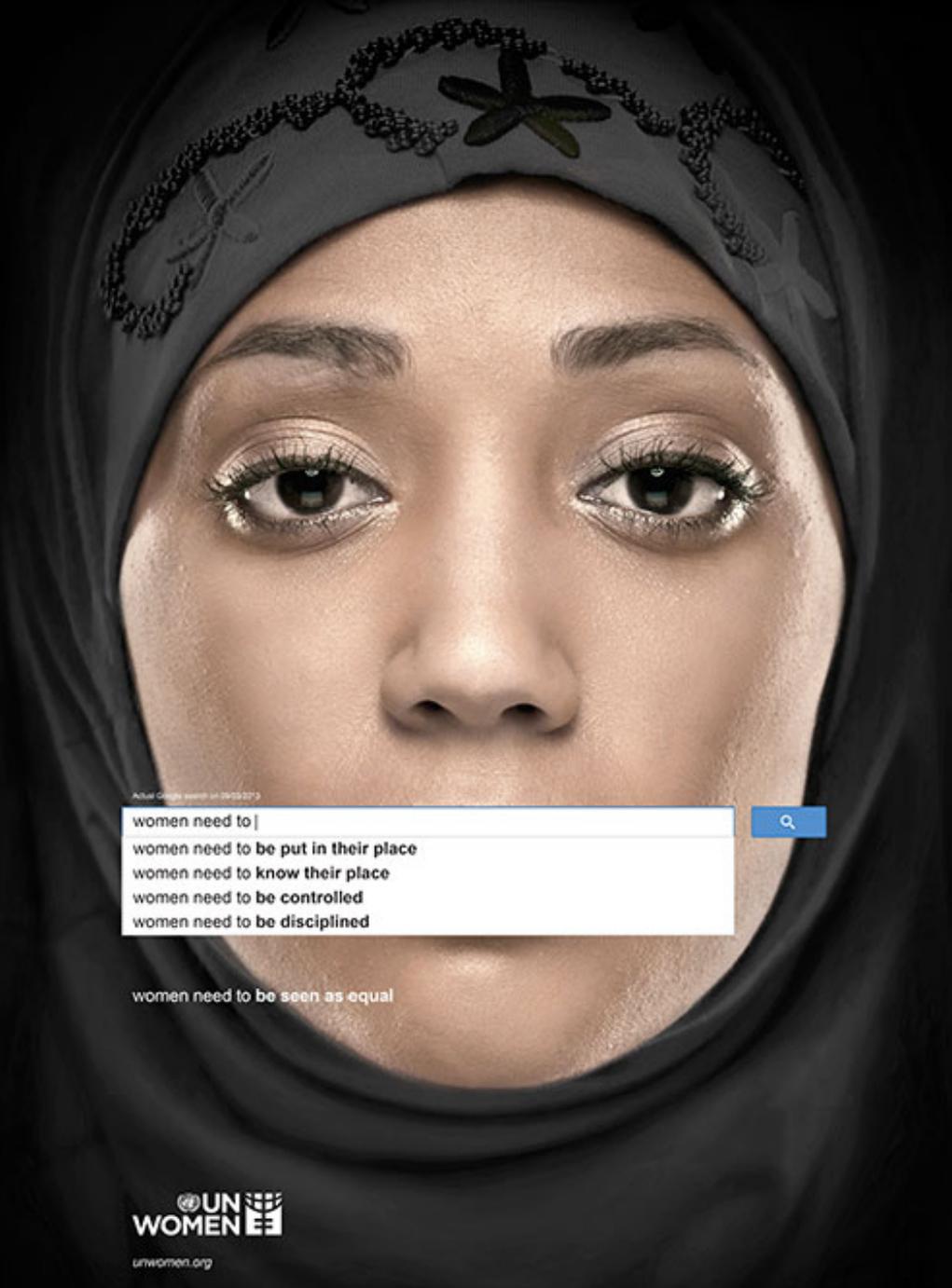
- women cannot drive
- women cannot be bishops
- women cannot be trusted
- women cannot speak in church



women cannot accept the way things are



unwomen.org



Actual Google search on 06/03/2013

women need to |

- women need to be put in their place
- women need to know their place
- women need to be controlled
- women need to be disciplined



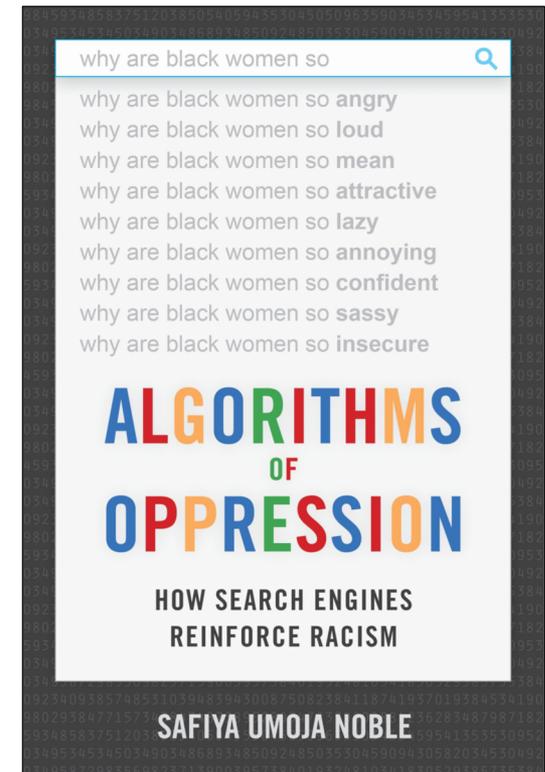
women need to be seen as equal



unwomen.org

Search Engine Bias

- Es geht oft um **Einzelphänomene**



Datensammlung in ESUPOL

Seit Februar 2017 **tägliche Abfrage** Google, Bing, DuckDuckGo

- **43.077.293 Suchanfragen** mit
- **296.321.947 Einzelvorschlägen**
- ~ 6,8 Vorschlageterme pro Anfrage
- ~ **3.100 Politikernamen** (Bundestag, Bundesrat, etc.)

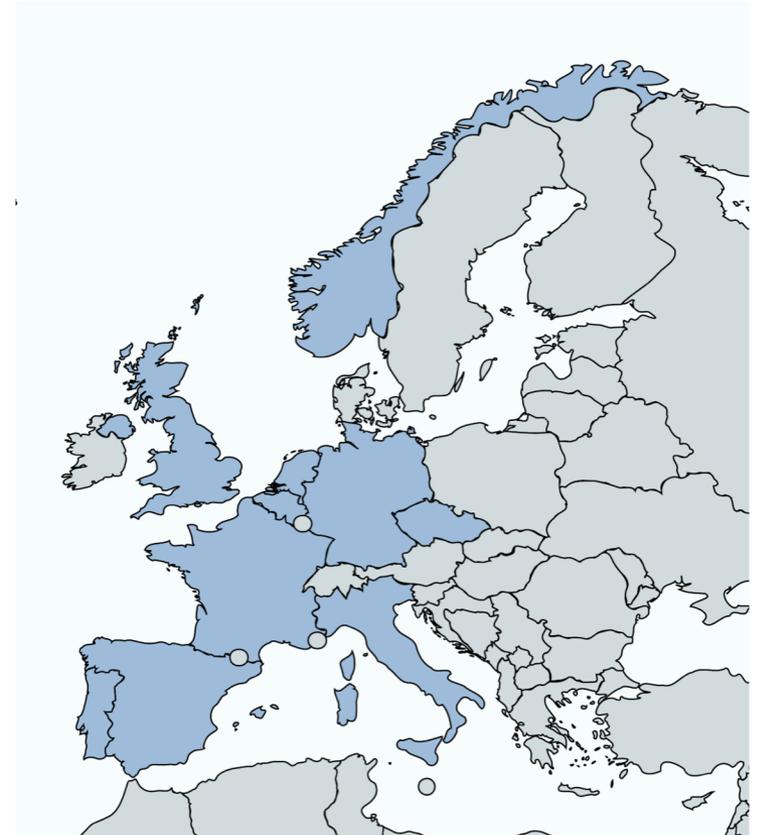
Unterteilung nach

- Geschlecht
- Partei
- Stadt/Region
- Migrationshintergrund
(z.B. Türkische Wurzeln)
- ...

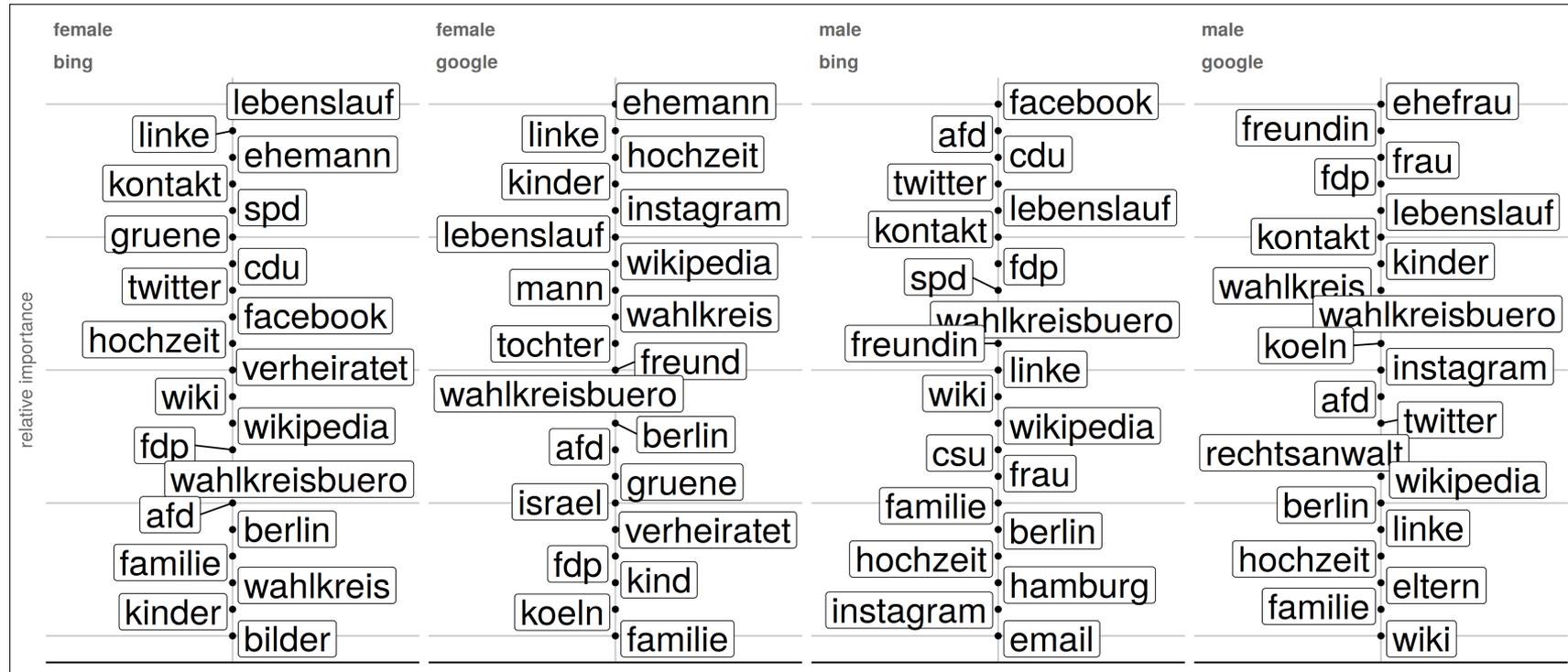


Proxy-Server-Netz und Harvesting-Pipe

- Nutzung von **Proxy-Servern** (Europa und USA)
- Länderübergreifende Vergleiche (z.B. EU-Wahl, Brexit)
- Ermöglicht parallele Anfragen
- Pro Proxy ca. 5 € / Monat



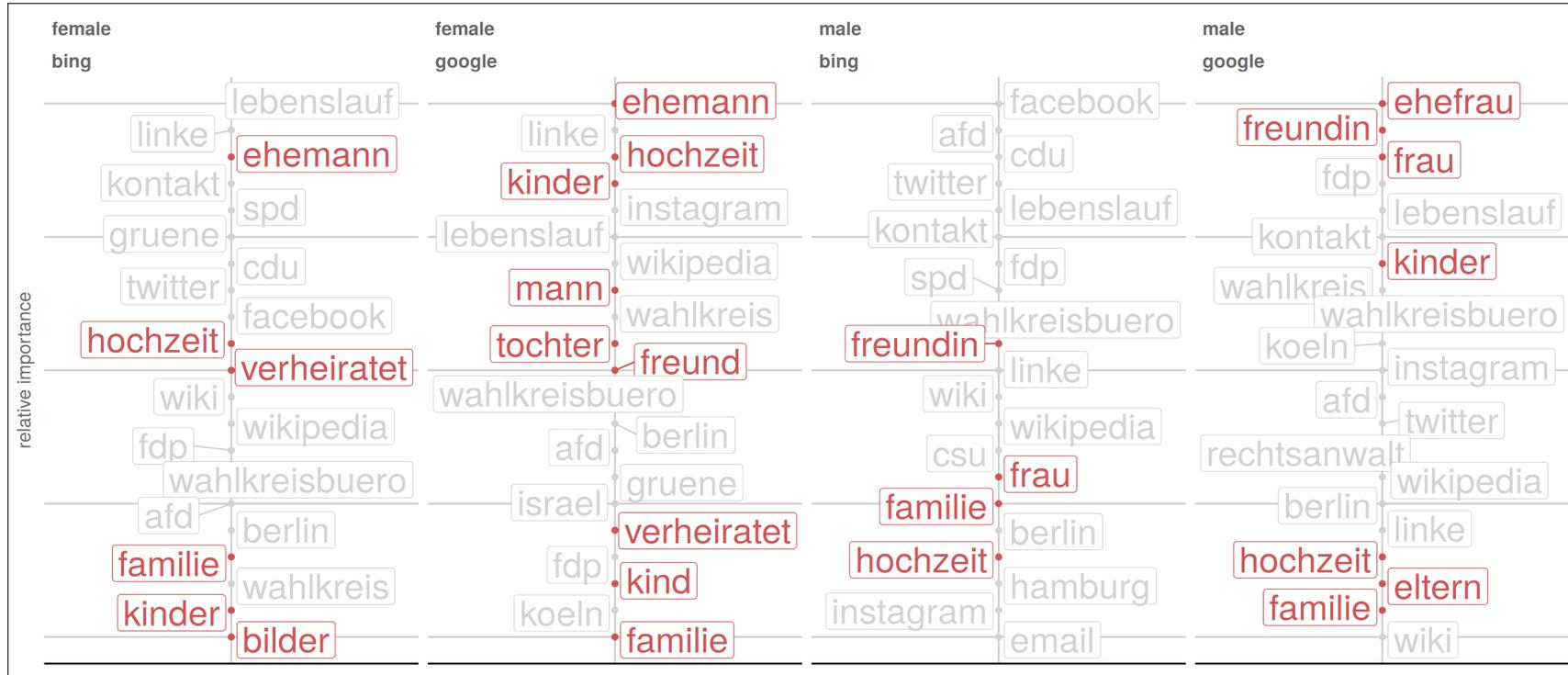
Was haben wir gefunden?



Wir haben verschiedene Gruppen von Begriffen identifiziert:

- Organisatorisches (Partei, Firmen, etc.)
- Orte (Wahlkreise, etc.)
- Privates...

Schauen wir mal genauer hin...



Wir haben das ganze mal etwas weiter untersucht:

- Privates wird statistisch signifikant weniger bei Männern angezeigt
- Privates wird statistisch signifikant weniger bei Älteren angezeigt

Data Literacy

Förderung des Stifterverbandes

- Kooperation der Informationswissenschaft und Informatik
- 300.000 EUR für zwei Jahre

Verankerung einer Vermittlung von **Datenkompetenz**

- In die Lehrinhalte **aller Studiengänge**
- Ab dem **zweiten Fachsemester** der Bachelorstudiengänge
- Bewusstsein für die Relevanz und die Vielfalt von Daten
- Rechtliche Aspekte, Datenverifizierung, uvm.

Alles neu macht der ... Landtag!



TH Köln

Graduiertenzentrum

Das Graduiertenzentrum der TH Köln die zentrale Anlaufstelle für alle, die sich für eine Kooperative Promotion an der TH Köln interessieren oder bereits an der TH Köln promovieren. » Mehr

GI NRW
GRADUIERTENINSTITUT
PROMOTIONEN – KOOPERATIV

Gesetzesänderung: Promotionskolleg mit Promotionsrecht

Am 11.07.19 wurde im Landtag beschlossen, dem zukünftigen Promotionkolleg in NRW das Promotionsrecht zu verleihen.

➤ Änderung des Hochschulgesetzes

Stellungnahme zum geplanten Promotionskolleg NRW

Im Rahmen der bevorstehenden Novellierung des Hochschulgesetzes des Landes NRW ist eine politische Debatte um die Zukunft der Promovierenden an Hochschulen für angewandte Wissenschaften/Fachhochschulen (HAW/FH) entbrannt. Prof. Dr. Martin Sternberg, Vorsitzender des Graduierteninstituts NRW, nimmt Stellung.

➤ Stellungnahme von Prof. Dr. Martin

Sternberg (PDF, 309 KB)

Open Science - Open Source

The screenshot shows a web browser window displaying the GitHub repository page for 'irgroup/gelic'. The browser's address bar shows 'github.com/irgroup/gelic'. The repository name 'irgroup / gelic' is prominently displayed at the top, along with statistics: 5 Unwatch, 1 Unstar, and 0 Fork. Below this, navigation tabs include 'Code', 'Issues 0', 'Pull requests 0', 'Actions', 'Projects 0', 'Wiki', 'Security', 'Insights', and 'Settings'. The repository description is 'German Library Indexing Collection', with an 'Edit' button. A list of topics includes 'information-retrieval', 'libraries', and 'test-collection'. Repository statistics show 45 commits, 1 branch, 0 packages, 0 releases, and 5 contributors. A 'Branch: master' dropdown and a 'New pull request' button are visible. A list of recent commits is shown, with the latest by 'iboeckma' adding a file for 'erschliessung.py' 20 days ago. Below the commit list, the 'README.md' file is selected, showing the title 'GeLIC - German Library Indexing Collection' and a description: 'Gelic is Proto-Germanic for "like, alike, similar" (see Wiktionary)'. It also lists the repository's contents, including a 'components' folder.

irgroup / gelic

Unwatch 5 Unstar 1 Fork 0

Code Issues 0 Pull requests 0 Actions Projects 0 Wiki Security Insights Settings

German Library Indexing Collection Edit

information-retrieval libraries test-collection Manage topics

45 commits 1 branch 0 packages 0 releases 5 contributors

Branch: master New pull request Create new file Upload files Find file Clone or download

iboeckma added file for erschliessung.py Latest commit 1b1110f 20 days ago

File	Description	Time
components	simplification of the file and folder names	2 months ago
scripts	added file for erschliessung.py	20 days ago
src	Update README.md	3 months ago
README.md	added the scripts erschliessung.py and synonyms.py	2 months ago

README.md

GeLIC - German Library Indexing Collection

Gelic is Proto-Germanic for "like, alike, similar" (see Wiktionary).

This repository contains the following items:

- components contains the documents, relevance assessments and topics

Forschung als Brückenschlag

Drittmittelfinanzierte Projekte sind per Design **interdisziplinär**

- Einbindung **mehrer Projektpartner** aus unters. **Domänen**
- Projekte sind meist **datengetrieben** und arbeiten **quantitativ**
- Oft **Informatik- oder Entwicklungsanteil**

Anwendungsfeld ist häufig informationswissenschaftlich

- Bibliothekswissenschaften (STELLA, Smart Harvesting)
- Bibliometrie (UseAltMe)
- Informationsmanagement (DLR)
- Search Engine Bias (ESUPOL)
- (Daten-)Journalismus (PRIOR, GeKo:mental, ...)

Forschung als Brückenschlag

In der **gelebten Praxis** (z.B. in konkreten Forschungsprojekten) gehen Data und Information Science bereits **Hand in Hand**

- Es braucht sowohl Data Science als auch Information Science-Methoden u.a. für Datengewinnung, Filterung und Aufbereitung, Analyse, Kontexteinbindung der Ergebnisse, ...

Nun muss dieser Anspruch auch der **Lehre weiter wirken**

- Data and Information Science (B.Sc.)
- Kooperationen über Fakultätsgrenzen hinweg (M.Sc.)



Auf zu neuen Ufern!

<http://iim.th-koeln.de> philipp.schaer@th-koeln.de

