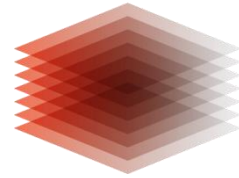

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY



TIB

Open Research Knowledge Graph für technische Fachinformation

Prof. Dr. Ralph Ewerth
Forschungsgruppe Visual Analytics
Frankfurt, 23. Mai 2018

TIB Historie: Klassische Zentrale Fachbibliothek für Technik und Naturwissenschaften

Einziger Bestand im Bereich Technik und Naturwissenschaften

- überwiegend Alleinbesitz

Schwerpunkte:

- **graue Literatur** (z.B. Konferenz- und Forschungsberichte)
- Literatur aus **Osteuropa** und **Ostasien**

Umfassender Bestand dient als Grundstock für Literaturversorgung

- Veränderungen durch Digitalisierung
- nach wie vor große Bedeutung für Wissenschaft und Unternehmen

1959 Gründung der TIB als Deutsche Zentrale Fachbibliothek für Technik und deren Grundlagenwissenschaften // zugleich Universitätsbibliothek

1977 Finanzierung Rahmenvereinbarung Förderung durch Bund und Länder

2005 Erste DOI-Registrierungsagentur für Forschungsdaten

2012 Übernahme des IWF Bestands - Kompetenzzentrum für nicht-textuelle Materialien // Digitales Langzeitarchivierungssystem

2015 Programmbereich C „Forschung und Entwicklung“ // Forschungsgruppe und Professur „Visual Analytics“

2016 Stiftung öffentlichen Rechts des Landes Niedersachsen – TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek

2016 Verabschiedung e-Preferred Content-Strategie

2017 Gemeinsame Berufung des Direktors mit Leibniz Universität Hannover // Forschungsgruppen „Data Science & Digital Libraries“, „Scientific Data Management“ // Joint Lab mit Forschungszentrum L3S

TIB: Moderne Bibliothek und Informationszentrum mit digitalen Diensten

Digitale Dienste ergänzen klassische Bibliotheksangebote und übersetzen Bibliotheksdienstleistungen in die digitale Welt:

- *Integriertes TIB-Portal* zu Recherche, Nachweis und Zugriff auf **>85 Millionen bibliographische Daten und Objekte, 18 Millionen ausgelieferte Volltexte**
- *Verhandlung von Lizenzen* zur digitalen Literaturversorgung für **>40 Konsortien**
- *AV-Portal für wissenschaftliche Videos* mit semantischer Erschließung **>13.000 Videos für täglich >1.000 Besucher**
- *Dienste zur Retrodigitalisierung* und digitaler Langzeitarchivierung (LZA-as-a-Service) **>1,6 Millionen langzeitarchivierte Objekte in 40 Formaten**
- *Open-Access-Transformation* durch deutsches arXiv-Netzwerk, SCOAP3-DH, OA-Repositoryn (z.B. Leibniz Open) und OA-Publikationsfonds
- *Forschungsdatenmanagement* entsprechend den FAIR-Prinzipien (mit **CKAN Data Manager**), wie Referenzierung mit **DOIs** durch **DataCite** oder Langzeitarchivierung mit **RADAR**



TIB: Informationszentrum für die Digitalisierung und wissenszentrierte Informationsflüsse

Vernetzte, digitale Dienste und systematische Realisierung von Synergien zwischen verschiedenen Diensten

*Kurz-/mittelfristige Ziele: **Digitalen Wandel in der Informationsversorgung aktiv gestalten***

- Intensivierung der Open-Access-Transformation
- Weiterentwicklung von Diensten für Forschungsdatenmanagement

*Mittel-/langfristige Ziele : **Führende Rolle beim Wandel von dokumentenbasierten zu wissensbasierten Informationsflüssen einnehmen***

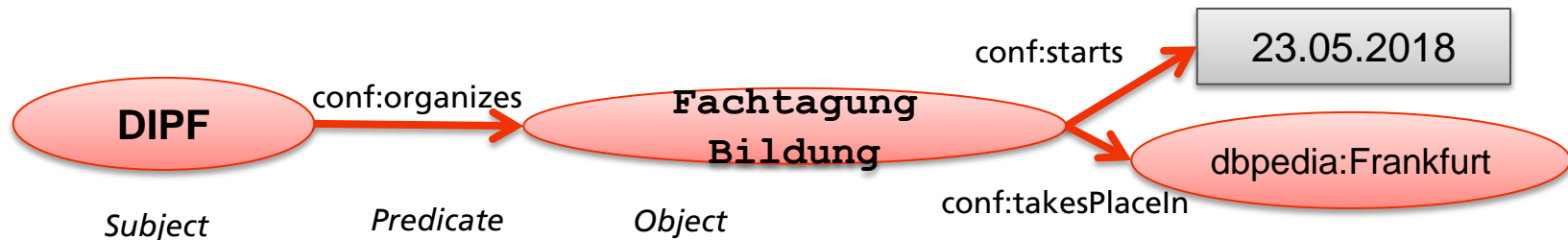
- *Open Research Knowledge Graph*: Repräsentiert Forschungsergebnisse semantisch als Wissensgraph, ermöglicht gänzlich neue Form der Zusammenarbeit in der Forschung

Paradigmenwechsel im Informationsaustausch in der Forschung

Wissensbasierte Informationsflüsse in
Wissenschaft und Technik

RDF & Linked Data: Crashkurs

1. Graph-basiertes RDF-Datenmodell mit S-P-O Aussagen (Fakten)



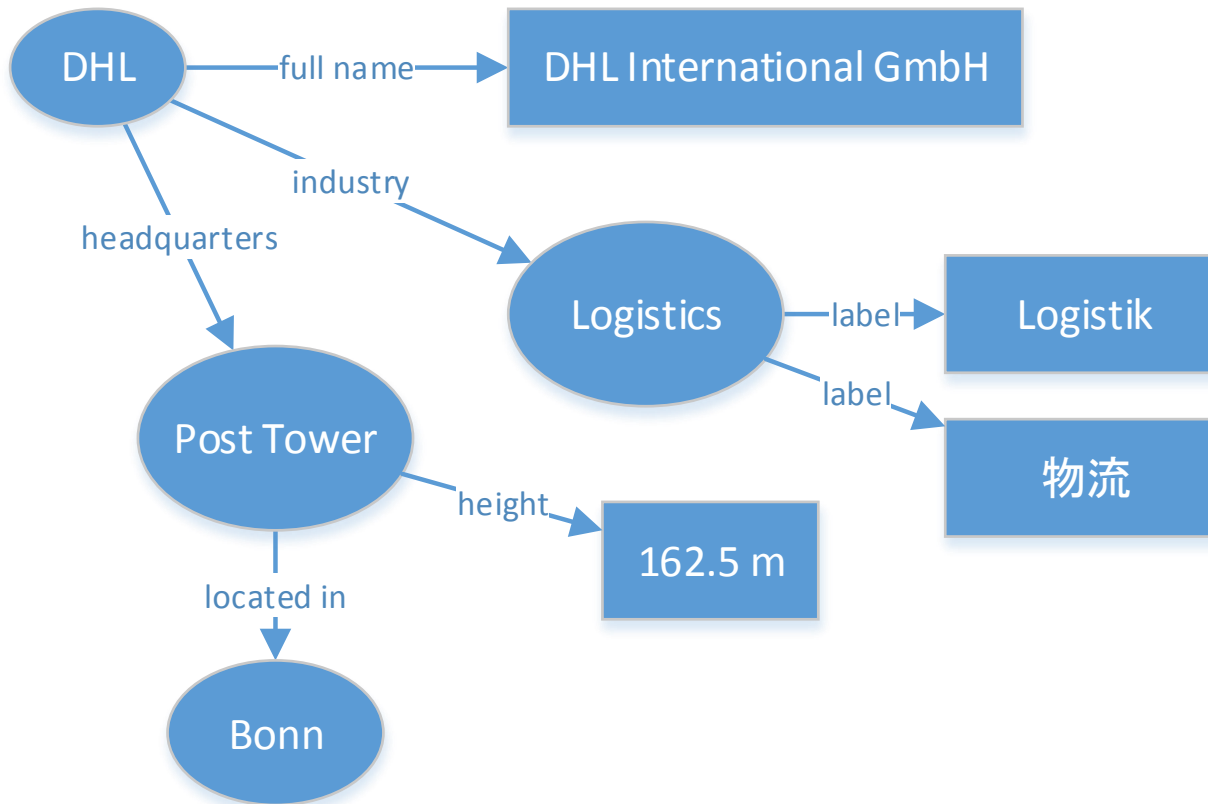
2. Serialisiert als RDF Triple:

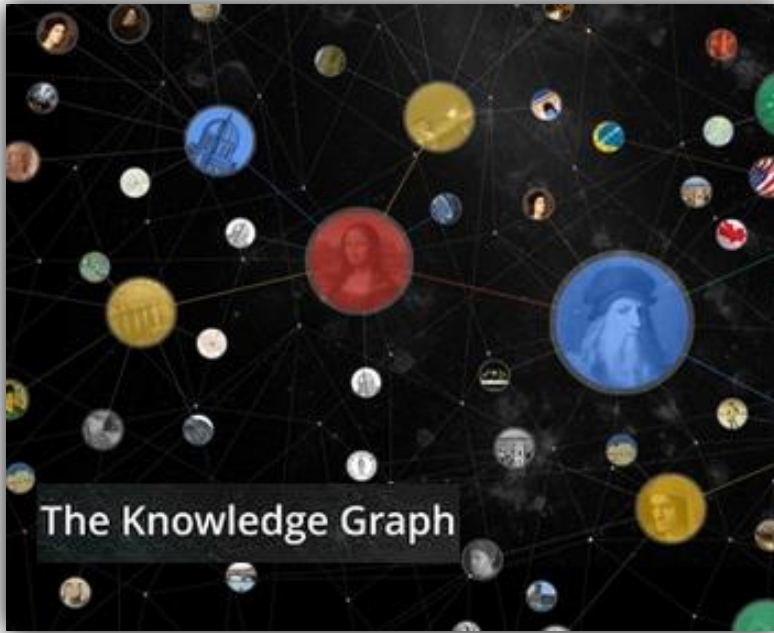
DIPF	<code>conf:organizes</code>	Fachtagung B.
Fachtagung B.	<code>conf:starts</code>	<code>"2018-05-23"^^xsd:date</code>
Fachtagung B.	<code>conf:takesPlaceIn</code>	<code>dbpedia:Frankfuer</code>

3. Publikation mit URL im Web, Intranet, Extranet

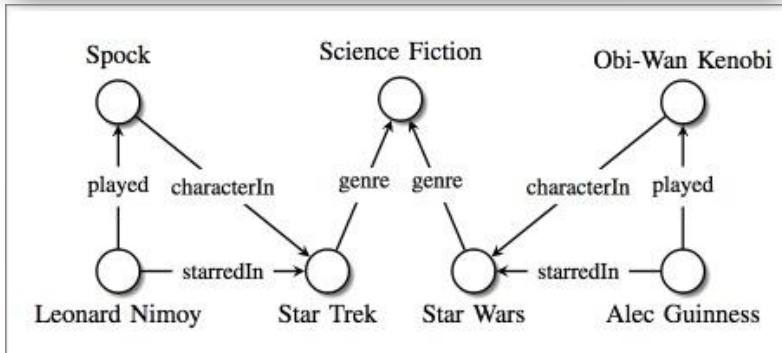
Linked Data

Creating Knowledge Graphs with RDF





GND	
Link zu diesem Datensatz	http://d-nb.info/gnd/1021356255
Typ	Person (piz)
Person	Appelbaum, Anne
Geschlecht	weiblich
Zeit	Lebensdaten: 1908-1998
Land	Deutschland (XA-DE); USA (XD-US)
Geografischer Bezug	Wirkungsort: New York, NY
Beruf(e)	Psychologin
Beziehungen zu Personen	Cassirer, Ernst (Vater)
Beteiligt an	1 Publikation 1. <i>Nachgelassene Manuskripte und Texte / Bd. 1. Zur Metaphysik der symbolischen Formen</i> 1995



Aktuelle Herausforderungen in der Wissenschaftskommunikation

Herausforderungen

- Digitalisierung der Wissenschaften
- Open Science
- Veränderungen in der Wissenschaftskommunikation
- Monopolisierung durch kommerzielle Akteure
- Große Masse an Publikationen
- „Reproducibility Crisis“

Strategisches Ziel: Paradigmenwechsel im Informationsaustausch

Wissensbasierte Informationsflüsse in Wissenschaft und Technik



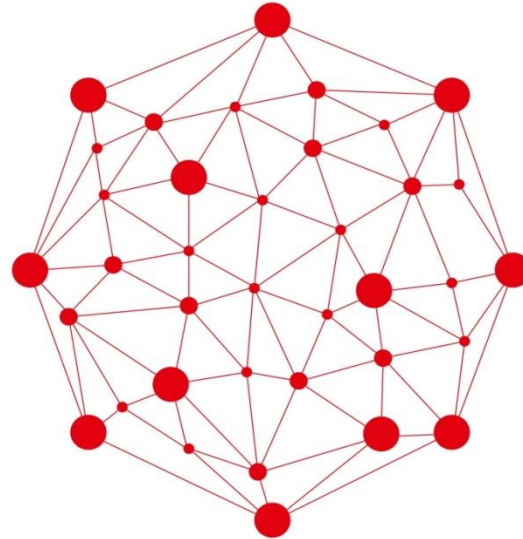
Open Research Knowledge Graph für die TIB-Fachdomänen – übergreifende Konzepte

Übergreifende Konzepte

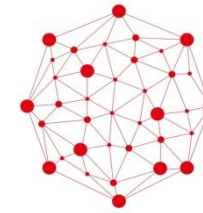
- Forschungsprobleme
- Definitionen
- Forschungsansätze
- Methoden

Artefakte

- Publikationen
- Daten
- Software
- Bild/Audio/Video
- Wissensgraphen / Ontologien



Open Research Knowledge Graph macht übergreifende und fachspezifische Konzepte **eindeutig identifizierbar** und **vernetzt diese semantisch** (mit klar beschriebenen Relationen) untereinander und mit relevanten weiteren Artefakten



Open Research Knowledge Graph für die TIB- Fachdomänen – fachspezifische Konzepte

Mathematik	Physik	Chemie	Informatik	Technik	Architektur
<ul style="list-style-type: none">• Definitionen• Theoreme• Beweise• Methoden• ...	<ul style="list-style-type: none">• Experimente• Daten• Modelle• ...	<ul style="list-style-type: none">• Stoffe• Strukturen• Reaktionen• ...	<ul style="list-style-type: none">• Konzepte• Implementierungen• Evaluationen• ...	<ul style="list-style-type: none">• Normen / Standards• Prozesse• Bauelemente• Einheiten, Sensordaten	<ul style="list-style-type: none">• Normen / Regularien• Bauelemente• Modelle• ...



CRISPR



Nur im Bibliothekskatalog der TIB suchen

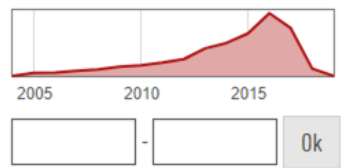
**Suche nach CRISPR:
>4.000 Ergebnisse**

Sortieren nach: Relevanz | [Aktualität](#) | [Titel](#)

Seite 1 von 4.373 Ergebnissen

Treffer erschließen

Erscheinungsjahr



Medientyp

- Aufsatz (Zeitschrift) (3.961)
- Patent (205)
- Hochschulschrift (93)
- Aufsatz (Konferenz) (34)
- Sonstige (30)

[+ Weitere](#)

Datenquelle

- British Library Online Contents (1.369)
- CiteSeerX (558)



CRISPR Critters and CRISPR Cracks

Charo, R. Alta / Greely, Henry T. | Taylor & Francis Verlag | 2015
 This essay focuses on possible nonhuman applications of **CRISPR**/Cas9 that are likely to be widely overlooked because they are unexpected



CRISPR BIOLOGY CRISPR-Cas: Adapting to change

Jackson, S. A. | British Library Online Contents | 2017



CRISPR decoys: Competitive inhibitors of CRISPR immunity

Maniv, I. / Hatoum-Aslan, A. / Marraffini, L.A. | British Library Online Contents | 2013



CRISPR-Cas

Das Immunsystem der Prokaryoten
 Marchfelder, Anita / Maier, Lisa-Katharina / Heidrich, Nadia et al. | Wiley | 2013

Beispiel aus der Chemie: CRISPR/Cas Genome Editing

arXiv.org

New Results

A practical guide to CRISPR/Cas9 genome editing in Lepidoptera

Linlin Zhang,  Robert Reed

doi: <https://doi.org/10.1101/130344>

Now published in *Diversity and Evolution of Butterfly Wing Patterns* doi: [10.1007/978-981-10-4956-9_8](https://doi.org/10.1007/978-981-10-4956-9_8)

Abstract

Info/History

Metrics

 Preview PDF

Abstract

CRISPR/Cas9 genome editing has revolutionized functional genetic work in many organisms and is having an especially strong impact in emerging model systems. Here we summarize recent advances in applying CRISPR/Cas9 methods in Lepidoptera, with a focus on providing practical advice on the entire process of genome editing from experimental design through to genotyping. We also describe successful targeted GFP knock-ins that we have achieved in butterflies. Finally, we provide a complete, detailed protocol for producing targeted long deletions in butterflies.

Copyright The copyright holder for this preprint is the author/funder. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

 Previous

Next 

Posted June 22, 2017.

 Download PDF

 Share

 Email

 Citation Tools

 Tweet

 Like 0

 G+

Subject Area

Genetics

Subject Areas

All Articles

Animal Behavior and Cognition

Biochemistry

Bioengineering

Bioinformatics

Biophysics

Cancer Biology

Semantische Sacherschließung mit einem Wissensgraph

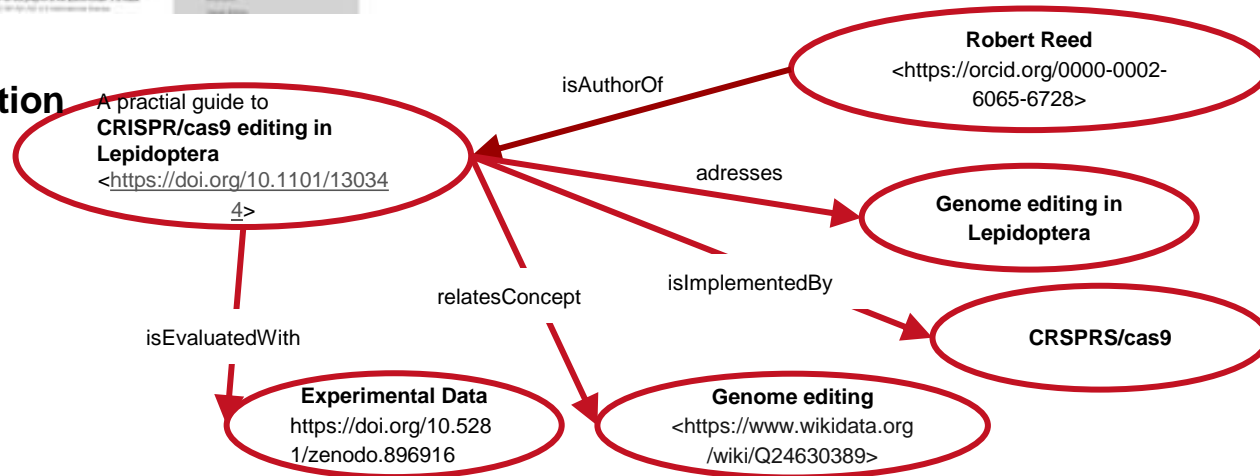
1. Original Publication



2. Graph Curation Form

Author	Robert Reed
Research Problem	Genome editing in Lepidoptera
Methods	CRISPR/cas9
related Concepts	Lepidoptera; Genome editing; CRSIPR
Experimental Data	https://doi.org/10.5281/zenodo.896916

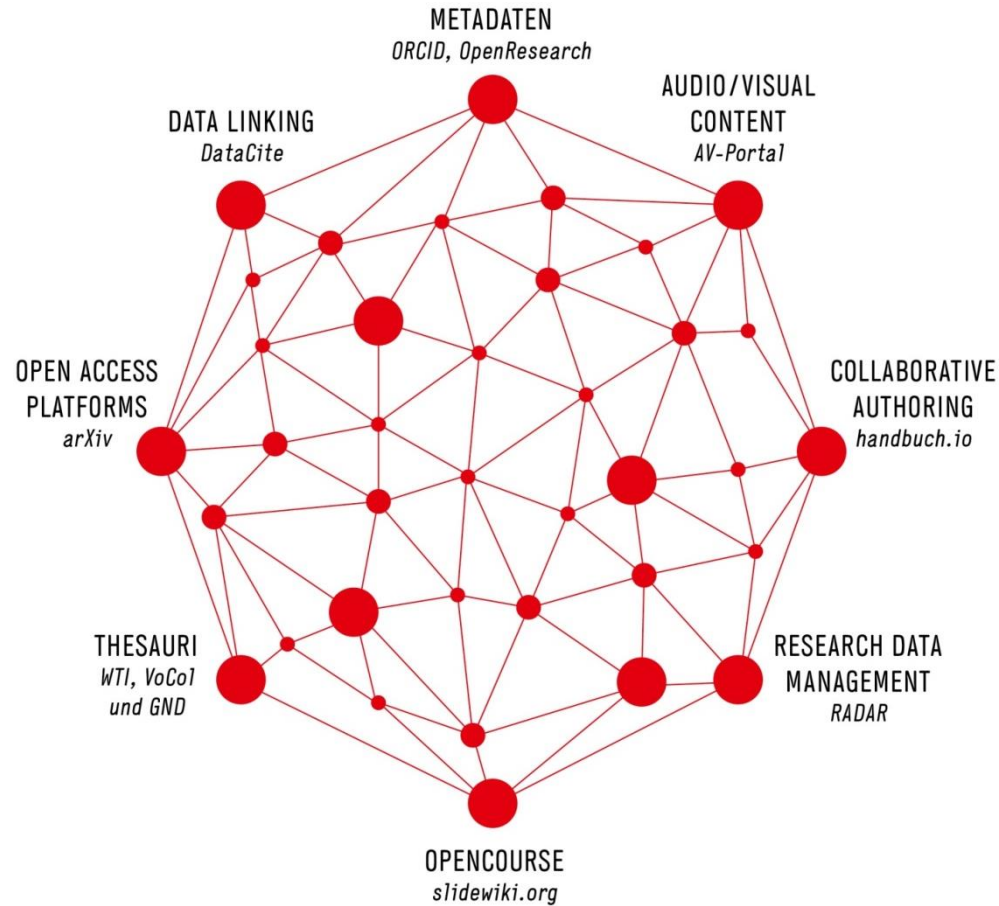
3. Graph representation



Automatische Generierung von Vergleichen/Surveys

Engineered Nucleases	Site-specificity	Safety	Ease-of-use / costs / speed
zinc finger nucleases (ZFN)	++ 9-18nt	+	-- \$\$\$: screening, testing to define efficiency
transcription activator-like effector nucleases (TALENs)	+++ 9-16nt	++	++ Easy to engineer 1 week / few hundred dollar
engineered meganucleases	+++ 12-40 nt	0	-- \$\$\$ Protein engineering, high-throughput screening
CRISPR system/cas9	++ 5-12 nt	-	+++ Easy to engineer few days / less 200 dollar

Open Research Knowledge Graph vernetzt existierende Dienste



Herausforderungen bei der Entwicklung des ORKG

Datengenerierung für den ORKG

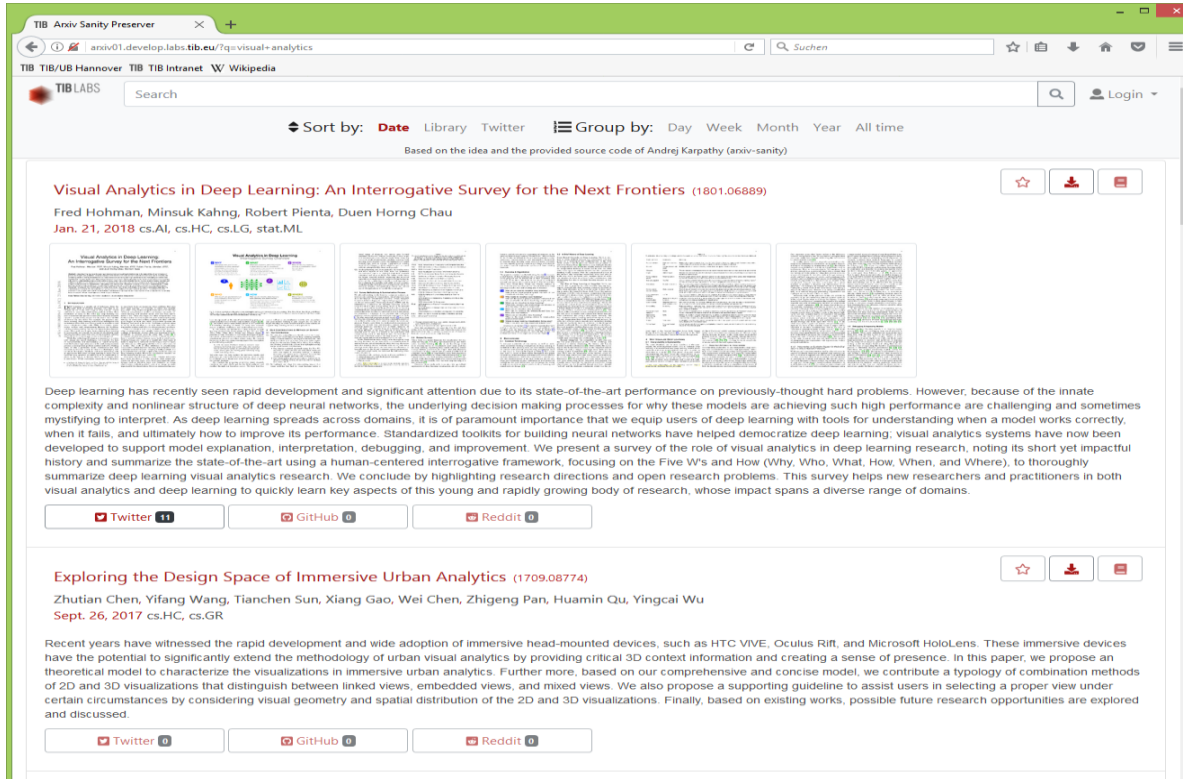
- Manuelle Erstellung von Informationen und Metadaten
- Kuratierung von Information
- Automatische Informationsextraktion

Etablierung des ORKG

- Motivation der Wissenschaftlergemeinschaft
- Belohnungssysteme

Interaktion und intelligente User Interfaces

Informationsextraktion aus Texten: Problemstellung, Lösung, Methoden, Experimente, etc.



The screenshot shows a web browser window titled "TIB Anvis Sanity Preserver" with the URL "anvis01.develop.labs.tib.eu/?q=visual+analytics". The page features a search bar, navigation links for "TIB LABS", "TIB/UB Hannover", "TIB Intranet", and "Wikipedia". Below the navigation is a sorting and grouping menu: "Sort by: Date Library Twitter" and "Group by: Day Week Month Year All time".

The main content area displays two research papers:

- Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers (1801.06889)**
 Fred Hohman, Minsuk Kahng, Robert Pienta, Duen Horng Chau
 Jan. 21, 2018 cs.AI, cs.HC, cs.LG, stat.ML
 This paper discusses the rapid development of deep learning and the challenges of interpreting its complex, nonlinear structure. It presents a survey of visual analytics in deep learning research, focusing on a human-centered interrogative framework (Five W's and How) to summarize research directions and open research problems.
- Exploring the Design Space of Immersive Urban Analytics (1709.08774)**
 Zhutian Chen, Yifang Wang, Tianchen Sun, Xiang Gao, Wei Chen, Zhigeng Pan, Huamin Qu, Yingcai Wu
 Sept. 26, 2017 cs.HC, cs.GR
 This paper explores the potential of immersive head-mounted devices (HTC VIVE, Oculus Rift, Microsoft HoloLens) to extend urban visual analytics methodology. It proposes a theoretical model to characterize visualizations in immersive urban analytics and a supporting guideline for selecting proper views under certain circumstances.

Each paper entry includes a grid of thumbnail images representing the paper's content and social media sharing buttons for Twitter, GitHub, and Reddit.

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.

Semantische Bild-Text-Bezüge

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



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



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.

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.

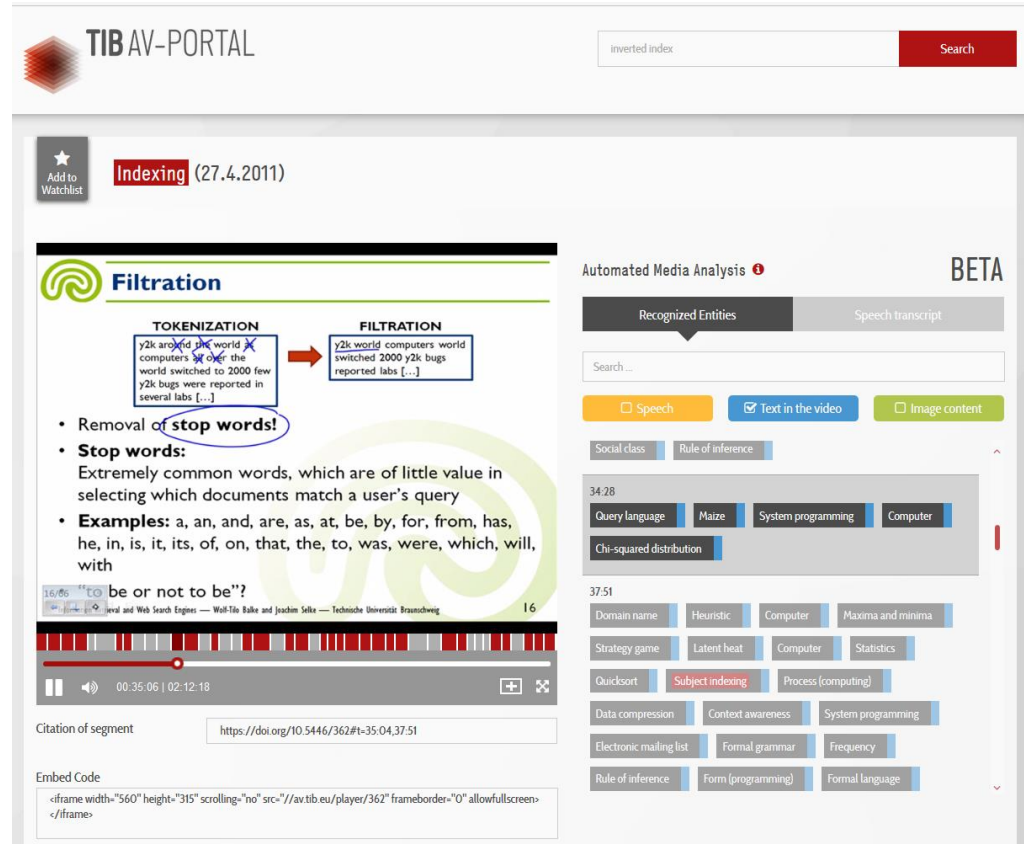


A **woman** is posing with two people dressed **down** as stormtroopers.

TIB AV-Portal

Erkennung von relevanten Informationen in Videos

Verknüpfung mit anderen Informationsquellen und Dokumenten



The screenshot displays the TIB AV-Portal interface. At the top, there is a search bar with the text "inverted index" and a red "Search" button. Below the search bar, the "TIB AV-PORTAL" logo is visible on the left, and a star icon with "Add to Watchlist" is on the right. A red box indicates "Indexing (27.4.2011)".

The main content area features a video player with a red progress bar. Above the player, a "Filtration" section shows a diagram of the process. The "TOKENIZATION" step shows the text "y2k and the world switched to 2000 few y2k bugs were reported in several labs [...]". The "FILTRATION" step shows the text "y2k world computers world switched 2000 y2k bugs reported labs [...]". A red arrow points from the tokenization step to the filtration step. Below the diagram, a list of bullet points explains the process:

- 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

Below the list, a video player shows a segment titled "be or not to be?". The player has a progress bar and a volume icon. Below the player, a "Citation of segment" field contains the URL <https://doi.org/10.5446/362#t=35:04:37:51>. An "Embed Code" field contains the following code:

```
<iframe width="560" height="315" scrolling="no" src="//av.tib.eu/player/362" frameborder="0" allowfullscreen>
</iframe>
```

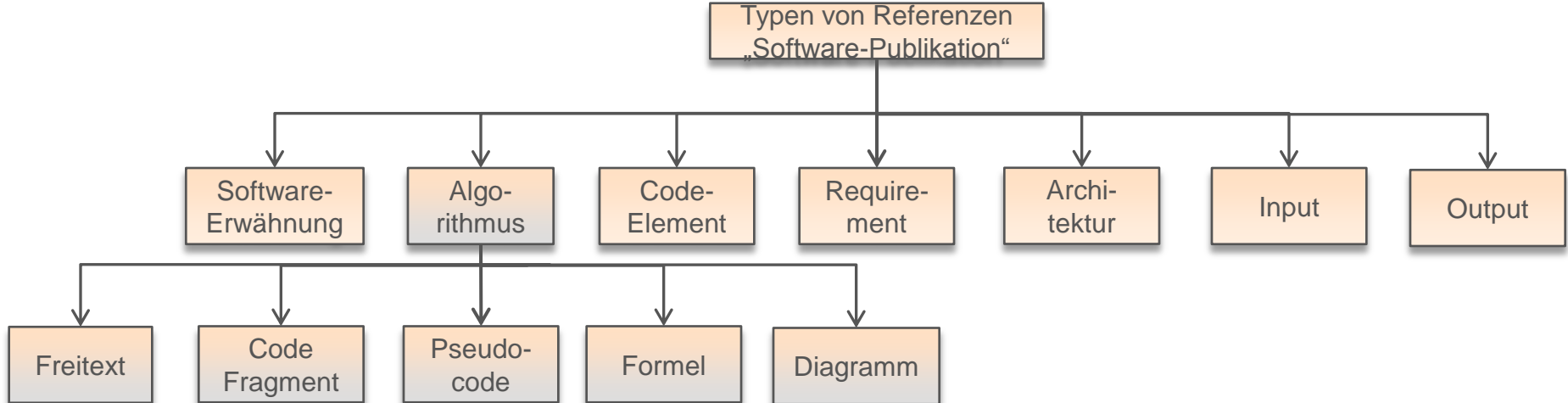
On the right side of the interface, there is a section titled "Automated Media Analysis" with a "BETA" label. It has two tabs: "Recognized Entities" (selected) and "Speech transcript". Below the tabs is a search bar. There are three buttons: "Speech" (orange), "Text in the video" (blue, checked), and "Image content" (green). Below these buttons are two tabs: "Social class" and "Rule of inference". The "Rule of inference" tab is selected. Below the tabs is a list of recognized entities, including "Query language", "Maize", "System programming", and "Computer". Below the list is a "Chi-squared distribution" button. Below the distribution button is a list of entities, including "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)", and "Formal language".

Exploration der Bezüge zwischen Publikationen und wissenschaftlicher Software

Ziel:

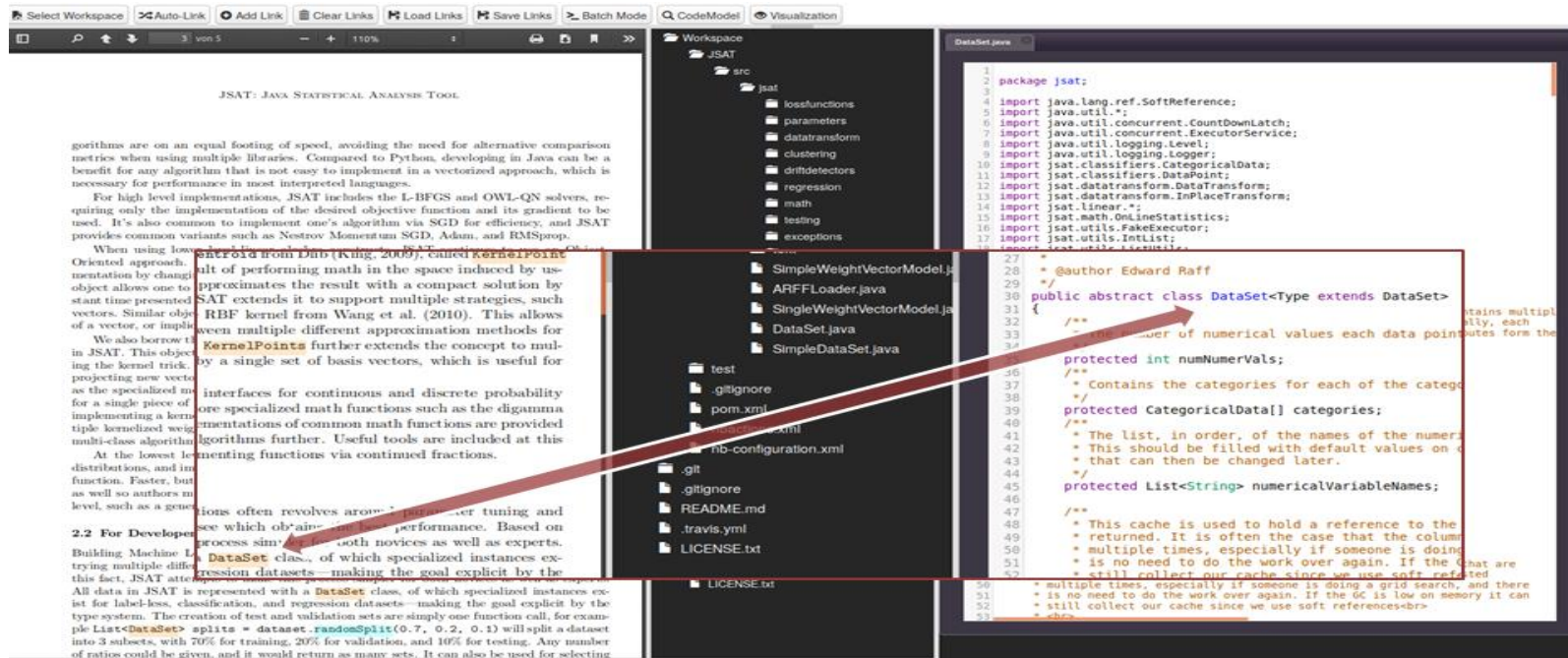
- Exploration von Softwarebezügen in Publikationen
- Für **Verständnis** von Algorithmen & **Reproduzierbarkeit** von Ergebnissen

Verknüpfe semi-automatisch Publikation mit Software-Implementierung



Exploration der Bezüge zwischen Publikationen und wissenschaftlicher Software

Ansatz: Ein semi-automatisches Analytics-Tool

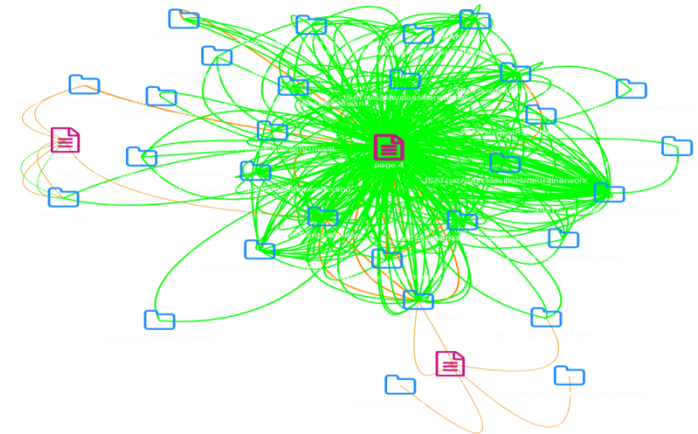
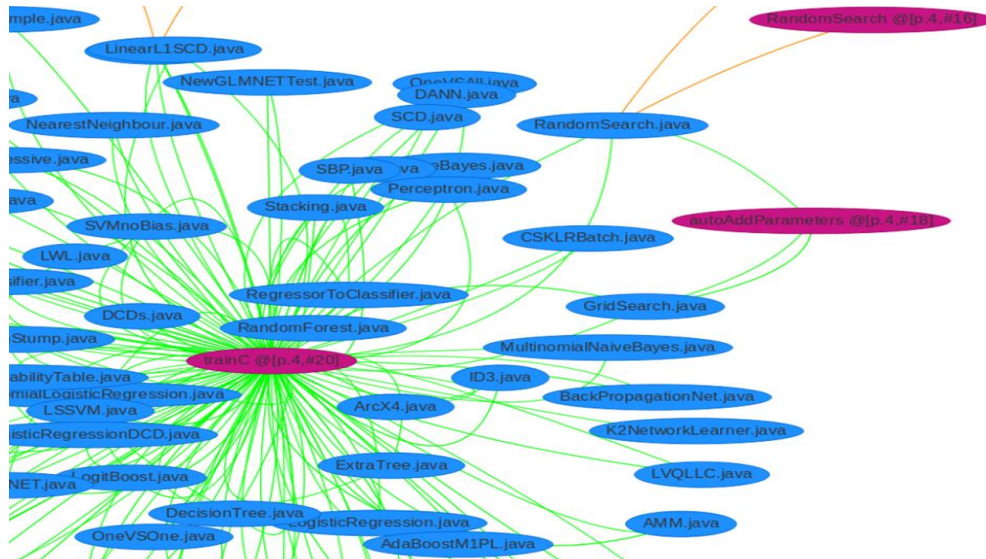


The screenshot displays the JSAT (Java Statistical Analysis Tool) interface. It is divided into several panes:

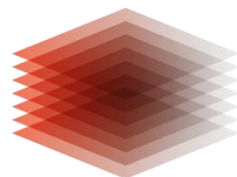
- Left Pane:** Contains documentation text. Key sections include:
 - A paragraph discussing performance in Java compared to Python.
 - A section titled "KernelPoints" which describes a method for approximating results in a space induced by multiple strategies, such as RBF kernel from Wang et al. (2010).
 - A section titled "Interfaces for continuous and discrete probability distributions" which lists specialized math functions like the digamma function.
 - A section titled "2.2 For Developers" which describes the `DataSet` class and its methods for splitting data into training, validation, and testing sets.
- Center Pane:** A file explorer showing the project structure:
 - src
 - jsat
 - lossfunctions
 - parameters
 - datatransform
 - clustering
 - drift detectors
 - regression
 - math
 - testing
 - exceptions
 - test
 - .gitignore
 - pom.xml
 - .git
 - .gitignore
 - README.md
 - .travis.yml
 - LICENSE.txt
- Right Pane:** Shows the source code for `DataSet.java`. Key elements include:
 - Package declaration: `package jsat;`
 - Imports for various Java utilities and JSAT classes.
 - Class declaration: `public abstract class DataSet<Type> extends DataSet<>`
 - Comments describing the class's purpose: "Abstract base class for numerical values each data point..."
 - Attributes: `protected int numNumerVals;` and `protected CategoricalData[] categories;`
 - Documentation comments for the `categories` attribute.
 - Cache-related comments: "This cache is used to hold a reference to the returned..."

Exploration der Bezüge zwischen Publikationen und wissenschaftlicher Software

...mit automatischer Bezüge-Erkennung, Möglichkeiten zur manuellen Annotation, Visualisierungen



LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY



TIB

Fazit

Member of

Leibniz
Leibniz Association

Vorteile wissensbasierter Informationsflüsse für Fachdomänen

Open Research Knowledge Graph:

Forschungsbeiträge können besser mit Forschungsdaten vernetzt werden

Vorteile

- Eindeutige **Identifikation** aller relevanten Artefakte, Konzepte, Attribute, Beziehungen
→ Terminologische und konzeptuelle **Präzision** und Schärfe, weniger Ambiguität
- **Maschinenlesbarkeit** des ORKG
→ Neue Such-, Retrieval-, Mining- und Assistenzanwendungen
- Bessere und explizite Vernetzung aller relevanten Artefakte und Informationsquellen
→ **Nachvollziehbarkeit**

Vorteile wissenschaftlicher Informationsflüsse

Vorteile

- Vermeidung von Medienbrüchen in den verschiedenen Phasen wissenschaftlicher Arbeit
→ Steigerung der **Effizienz**
- Nutzung von Konzepten und Beziehungen über Fachgrenzen hinweg
→ **Interdisziplinarität und Transdisziplinarität**
- **Eindämmung der Proliferation** wissenschaftlicher Publikationen
- **Erleichterung des Einstiegs** von jungen oder fachfremden WissenschaftlerInnen oder Laien (Open Science)

Positionspapier zum ORKG

Sören Auer et. Al: *Towards an Open Research Knowledge Graph*

<https://zenodo.org/record/1157185#.WwUuYEdApKk>

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY



Vielen Dank für Ihre Aufmerksamkeit!

Fragen?

Mail: ralph.ewerth@tib.eu

Member of
Leibniz
Leibniz Association