



**dblp**

computer science bibliography

# Praxis der Autoren-Disambiguierung

## Einleitung

Marcel R. Ackermann


4. Dezember 2018



# Authorities und Bibliographien

- Euler, Astrid von
- Euler, Brigitte
- Euler, Chad W.
- Euler, Christian
- Euler, Craig
- Euler, Dieter
- Euler, Ekkehard
- Euler, Ellen
- Euler, James A.
- Euler, Juliane
- Euler, L.
- Euler, Leonhard
- Euler, Lutz
- Euler, Matthew J.
- Euler, Peter
- Euler, R.
- Euler, Reinhardt
- Euler, Sebastian
- Euler, Stefan
- Euler, Stephan
- Euler, Thomas
- Euler, Thorsten
- Euler, Timm
- Euler, William B.
- Euler-Chelpin, Astrid von
- Euler-Rolle, Nikolaus
- Eulgem, Stefan
- Euliano, Neil R.
- Euliano, Tammy Y.
- Eulig, N.
- Eulino, Joao

## Euler, Leonhard

Author ID: [euler.leonhard](#) 

Published as: Euler, L.; Euler, Leonard; Euler, Leonh.; Euler, Leonhard; Eulerus, L.; Eulerus, Leonhardus

External Links: [MGP](#) · [Wikidata](#) · [dblp](#) · [GND](#) · [MacTutor](#)

## Found 109 Documents (Results 1–100)

Hits per Page  Display Format  Result Sorting

- Euler, Leonhard  
**Correspondence of Leonhard Euler with Johann Andreas von Segner and other scholars in Halle. Edited by Andreas Kleinert and Thomas Steiner (to appear). (Briefwechsel von Leonhard Euler mit Johann Andreas von Segner und anderen Gelehrten aus Halle.)** (English) [Zbi 06897956](#)  
Leonhardi Euleri Opera Omnia. Series Quarta A: Commercium Epistolicum 8. New York, NY: Birkhäuser/Springer (ISBN 978-3-319-75942-5/hbk). xiv, 713 p. (2018).  
MSC: [01A70](#) [01A50](#) [01A75](#)  
[PDF](#) [BibTeX](#) [XML](#) [Cite](#)
- Euler, Leonhard  
**Commentationes astronomicae ad theoriam perturbationum pertinentes. Part 2. Edited by Andreas Verdun (to appear).** (English) [Zbi 06739152](#)  
Leonhardi Euleri Opera Omnia, Series 2: Opera Mechanica et Astronomica 26. Basel: Birkhäuser/Springer (ISBN 978-3-7643-1472-9/hbk; 978-3-0346-0649-3/ebook). 800 p. (2018).  
MSC: [01A75](#) [01A50](#) [85-03](#) [70-03](#) [01A70](#)  
[PDF](#) [BibTeX](#) [XML](#) [Cite](#)
- Euler, Leonhard  
**Commentationes astronomicae ad theoriam perturbationum pertinentes. Part 3. Edited by Andreas Verdun (to appear).** (English) [Zbi 06739142](#)  
Leonhardi Euleri Opera Omnia, Series 2: Opera Mechanica et Astronomica 27. Basel: Birkhäuser/Springer (ISBN 978-3-7643-1473-6/hbk; 978-3-0346-0645-5/ebook). (2018).  
MSC: [01A75](#) [01A50](#) [85-03](#) [70-03](#) [01A70](#)  
[PDF](#) [BibTeX](#) [XML](#)
- Euler, Leonhard  
**Correspondence of Leonhard Euler with L. Bertrand, Ch. Bonnet, M. M. Bousquet, J. de Castillon, G. Cramer, Ph. Cramer, G. Cuenz, A. von Haller, G. L. Lesage, J. M. von Loen and J. C. Wettstein. Edited by Siegfried Bodenmann, Vanja Hug, Mirjana Ilić and Andreas Kleinert. (Correspondance de Leonhard Euler avec L. Bertrand, Ch. Bonnet, M. M. Bousquet, J. de Castillon, G. Cramer, Ph. Cramer, G. Cuenz, A. von Haller, G. L. Lesage, J. M. von Loen, J. C. Wettstein.)** (French) [Zbi 1382.01013](#)  
Leonhardi Euleri Opera Omnia. Series Quarta A: Commercium Epistolicum 7. Basel: Birkhäuser (ISBN 978-3-7643-8743-3/hbk). xii, 621 p. (2017).  
MSC: [01A75](#) [01A70](#) Reviewer: [Franz Lemmermeyer \(Jagstzell\)](#)  
[PDF](#) [BibTeX](#) [XML](#) [Cite](#)
- Euler, Leonhard  
**Correspondence of Leonhard Euler with Daniel, Johann II and Johann III Bernoulli. Correspondence of Johann Albrecht Euler with Daniel Bernoulli. Correspondence of Daniel Bernoulli with officials of the Petersburg Academy of Sciences (selection) and Nicolaus Fuss. Part II. Edited by Emil A. Fellmann and Gleb K. Mikhajlov. With the collaboration of Andreas Kleinert, Martin Mattmüller, Uta Monecke and Andreas**



# Problemfälle

## ▪ Homonyme

### Introduction of the Minitrack on Information Security and Privacy

Tung Bui  
University of Hawaii at Manoa

Eric Clemons  
University of Pennsylvania  
clemons@wharton.upenn.edu

David Wang  
University of Hawaii at Manoa  
twwang@hawaii.edu

### A Practical and Portable Solids-State Electronic Terahertz Imaging System

Ken Smart<sup>1</sup>✉, Jia Du<sup>1,\*</sup>✉, Li Li<sup>1</sup>✉, David Wang<sup>1</sup>✉, Keith Leslie<sup>1</sup>✉, Fan Ji<sup>2</sup>✉

### A hidden Markov model for detecting recombination al detection microarrays

Hilary Renshaw<sup>1,2</sup>, Scott Weaver<sup>3,4,5</sup>, Robert B. Tesh<sup>3</sup> and David Wang<sup>1</sup>

### Real-Time Final Product Quality Prediction in Batch Process Operation

David Wang, *Member, IEEE*

### Mechanisms, Overheads and Scaling

Brinda Ganesh<sup>†</sup>, Aamer Jaleel<sup>‡</sup>, David Wang<sup>†</sup>, and Bruce Jacob<sup>†</sup>



# Problemfälle

## ▪ Homonyme

**Introduction of the Minitrack on Information Security and Privacy**

Tung Bui  
University of Hawaii at Manoa

Eric Clemons  
University of Pennsylvania  
clemons@wharton.upenn.edu

David Wang  
University of Hawaii at Manoa  
twwang@hawaii.edu

**A Practical and Portable Solids-State Electronic  
Terahertz Imaging System**

Ken Smart<sup>1</sup>✉, Jia Du<sup>1,\*</sup>✉, Li Li<sup>1</sup>✉, David Wang<sup>1</sup>✉, Keith Leslie<sup>1</sup>✉, Fan Ji<sup>2</sup>✉

**a hidden Markov model for detecting recombination  
al detection microarrays**

Hilary Renshaw<sup>1,2</sup>, Scott Weaver<sup>3,4,5</sup>, Robert B. Tesh<sup>3</sup> and David Wang<sup>1</sup>

**Real-Time Final Product Quality Prediction in  
Batch Process Operation**

David Wang, *Member, IEEE*

**Mechanisms, Overheads and Scaling**

Brinda Ganesh<sup>†</sup>, Aamer Jaleel<sup>‡</sup>, David Wang<sup>†</sup>, and Bruce Jacob<sup>†</sup>

## ▪ Synonyme

**Surface Prediction for Spatial Augmented Reality**

Adam Gomes<sup>(✉)</sup>, Keegan Fernandes, and David Wang

**Degenerate Motions in Multicamera Cluster SLAM with Non-overlapping Fields**

Michael J. Tribou<sup>a,\*</sup>, David W. L. Wang<sup>b</sup>, Steven L. Waslander<sup>a</sup>

**Collision Detection Using a Flexible Link Manipulator**

S. J. Moorehead and D. Wang



# Disambiguierung in der Praxis

- Redaktionelle Kuration

- Selbstkuration durch die Forscher

The image shows three overlapping screenshots of web pages. The top-left screenshot is the dblp (computer science bibliography) homepage, featuring a search bar and navigation links. The top-right screenshot is the zbMATH website, which is described as 'the first resource for mathematics', with a search bar and various filters. The bottom screenshot is the Wikidata 'Welcome to Wikidata' page, highlighting it as 'the free knowledge base with 52,308,442 data items that anyone can edit'.

The image shows two overlapping screenshots of researcher profiles. The top screenshot is a ResearchGate profile for Sarah Hartmann, with the text 'Discover the world's scientific knowledge' and 'With 15+ million researchers and 118+ million publications, this is where everyone can access science'. The bottom screenshot is a Google Scholar profile for Sarah Hartmann, showing a list of publications. The first publication is 'Public libraries in the knowledge society: Core services of libraries in informational world cities' (2013), and the second is 'Looking for friends and followers: a global investigation of governmental social media use' (2015).



# Disambiguierung in der Praxis

- Redaktionelle Kuration

- Selbstkuration durch die Forscher

A collage of three web pages. The top left is the dbip (computer science bibliography) homepage, showing a search bar and navigation menu. The top right is the zbmATH (first resource for mathematics) homepage, featuring a search bar and navigation options. The bottom is the Wikidata homepage, displaying a network graph and a 'Welcome to Wikidata' message.

A collage of two web pages, ResearchGate and Google Scholar, with a large grey 'X' overlaid across them. The ResearchGate page shows a search bar and a list of publications. The Google Scholar page shows a search bar and a list of publications, including one by Sarah Hartmann.



# Über uns



- Projekt im Leibniz-Wettbewerb:  
*„Skalierbare Autoren-Disambiguierung in  
Literatur-Datenbanken“*



Heidelberg Institute for  
Theoretical Studies



- Marcel Ackermann
- Florian Reitz
- Nicolas Roy
- Adam Bannister
- Olaf Teschke
- Christoph Müller
- Michael Strube



# Agenda

- Praxis bei dblp
- Praxis bei zbMATH
- NLP-Methoden des HITS
- Diskussion: Praxis bei Ihnen?







# Autoren-Disambiguierung bei dblp

Marcel R. Ackermann

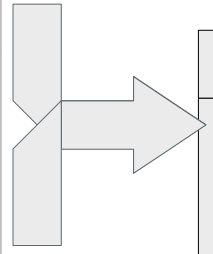
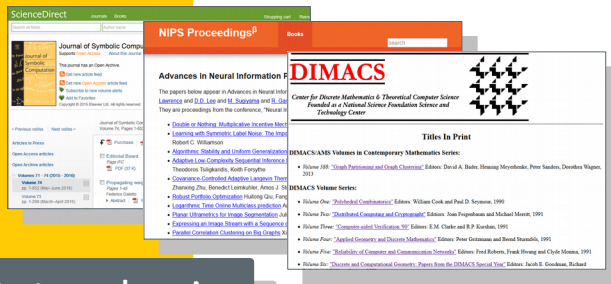
4. Dezember 2018



# Der dblp-Workflow

## 1. Datenakquise

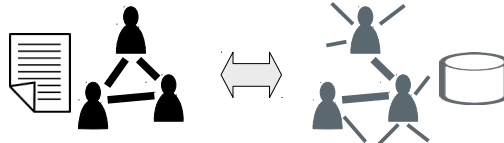
- Webcrawler oder Datenlieferungen
- Normalisierung



XML

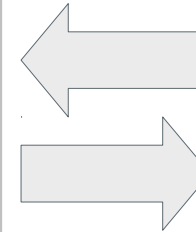
```
<article key="jou
<author>Gregor Ke
<title>Using exte
invariant theory.
<pages>161-181</p
<year>2016</year>
<volume>72</volum
<journal>J. Symb
<ee>http://dx.doi
<url>db/journals/
</article>
```

## 2. Initiale Datensäuberung

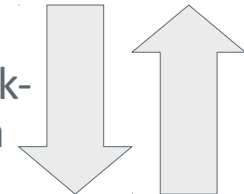


- Zuordnen der Autorenschaft
- Homonym/Synonym-Erkennung
- Korrektur der Neudaten
- Vervollständigungen (zB: abgek. Namen, fehlende Namensteile)
- Anreicherung (Homepages, Affiliations, PIDs, zB ORCIDs, ...)

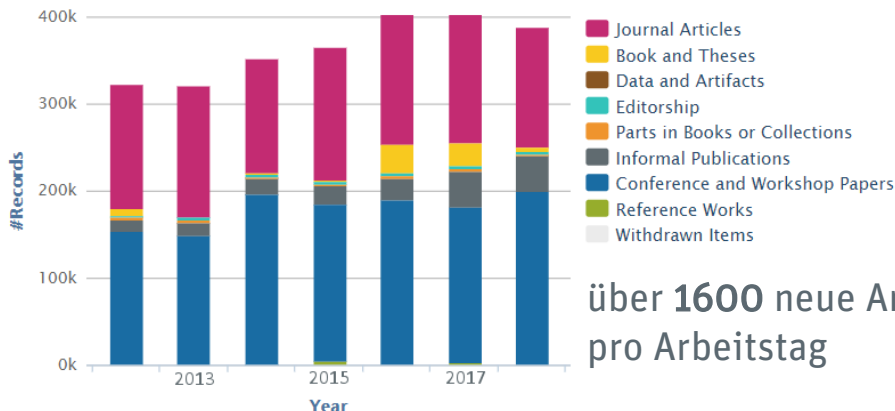
eigener Korpus als Authority



Netzwerk-  
analysen

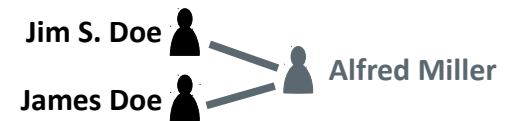


## Effizienz: Neuaufnahmen pro Jahr



über 1600 neue Artikel  
pro Arbeitstag

## 3. Kontinuierliche Qualitätskontrolle



- Homonym/Synonym-Erkennung
- Korrekturen am bestehenden Corpus
- neue Informationen durch neuste Ergänzungen



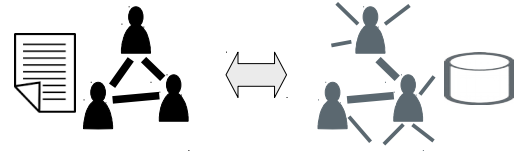
# Der dblp-Workflow

## 1. Datenakquise

- Webcrawler oder Datenlieferungen
- Normalisierung

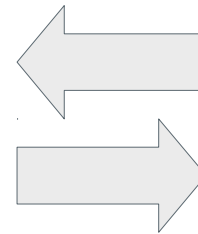


## 2. Initiale Datensäuberung



- Zuordnen der Autorenschaft
- Homonym/Synonym-Erkennung
- Korrektur der Neudaten
- Vervollständigungen (zB: abgek. Namen, fehlende Namensteile)
- Anreicherung (Homepages, Affiliations, PIDs, zB ORCIDs, ...)

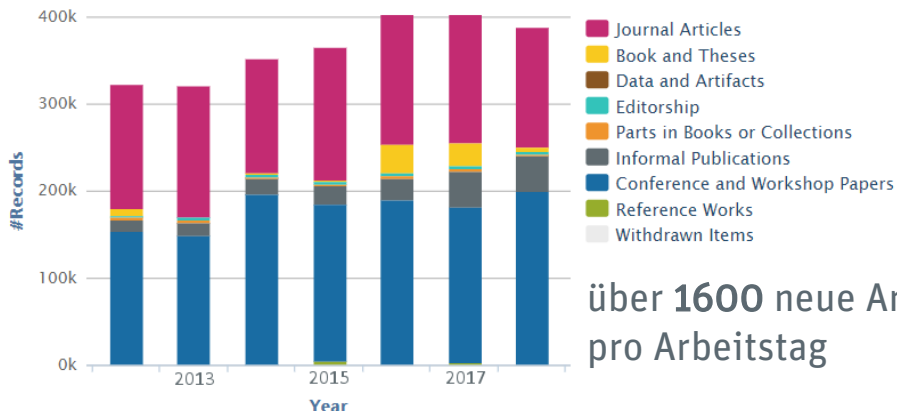
eigener Korpus als Authority



Netzwerk-  
analysen



## Effizienz: Neuaufnahmen pro Jahr



über 1600 neue Artikel  
pro Arbeitstag

## 3. Kontinuierliche Qualitätskontrolle

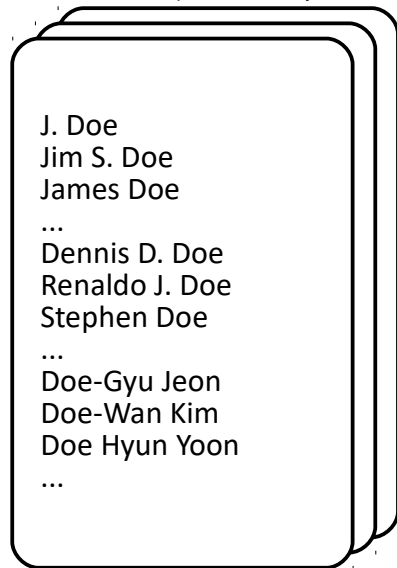


- Homonym/Synonym-Erkennung
- Korrekturen am bestehenden Corpus
- neue Informationen durch neuste Ergänzungen



# 1. Initiale Kandidatenmenge erzeugen

- Suche auf gesamtem Korpus (rein syntaktisch)
- benötigt sehr schnelle Ähnlichkeitsmaße und Datenstrukturen

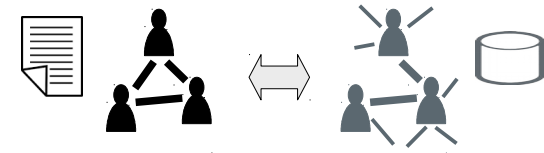


Simple Blocking-Strategien, z.B.:

- letzter Namensteil
- längster Namensteil
- seltenster Namensteil
- vorkommende Konsonanten

alle Vergleiche unter Unifikation  
von diakritischen Zeichen

## Algorithmus *dblpi*



# 1. Initiale Kandidatenmenge erzeugen

## 2. Zu große Kandidatenmengen adaptiv filtern

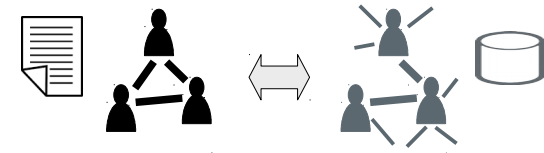
- zu groß: mehr als 10 Kandidaten (z.B. „W. Wang“  $\approx$  15000)
- Filter darf teurere Maße verwenden als initiale Erstellung



### Mögliche Filterregel:

- alle Namensteile als Präfix enthalten
- in eine oder andere Richtung
- ggf. permutierte Reihenfolge

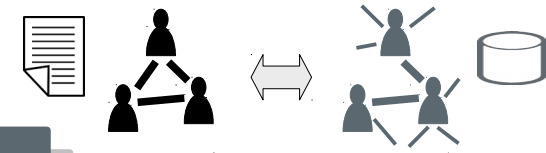
# Algorithmus *dblpi*



# 1. Initiale Kandidatenmenge erzeugen

# Algorithmus *dblpi*

## 2. Zu große Kandidatenmengen adaptiv filtern

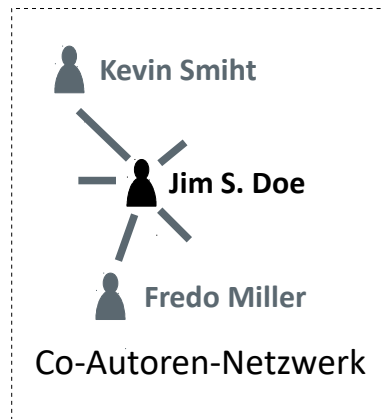


## 3. Weitere Kandidaten im Kontext finden

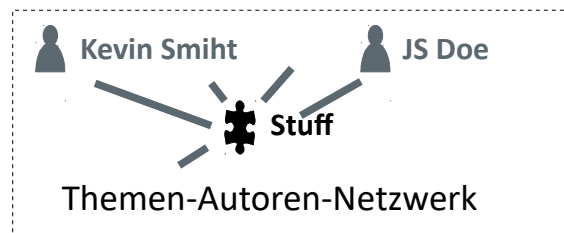
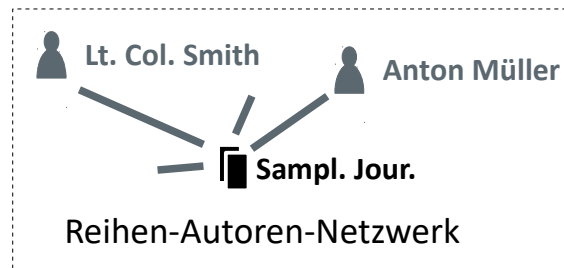
- Ausnutzen von semantischen Zusammenhängen
- darf teurere Ähnlichkeitsmaße verwenden

J. Doe, A. Miller, K. Smith: „Important Stuff.“ Sampl. Jour., 1975

Kontext der Kandidatenmenge:



weitere Kontexte, z.B.:



Mögliche Maße:

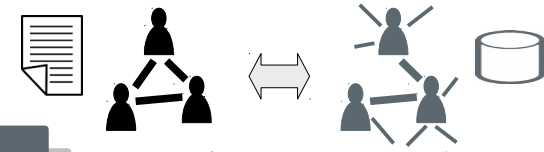
- Levenshtein
- Jaro-Winkler
- n-Gramme
- Soundex-Varianten mit Anpassungen an Personennamen



# 1. Initiale Kandidatenmenge erzeugen

# Algorithmus *dblpi*

## 2. Zu große Kandidatenmengen adaptiv filtern



## 3. Weitere Kandidaten im Kontext finden

## 4. Kandidaten bewerten

- Bewertung anhand mehrdimensionaler Features
- Werte je zwischen 0 und 1



Kontext  
Kandidat

- ähnlicher Name
- gemeinsame Publikation
- zeitnahe Publikationen
- gleiche Affiliation
- publiziert in gleicher Reihe
- publiziert in *ähnlicher* Reihe
- verwendet gleiche Themen
- ggf. weitere ...

- Ranking mittels gewichteter Summe

Co-A

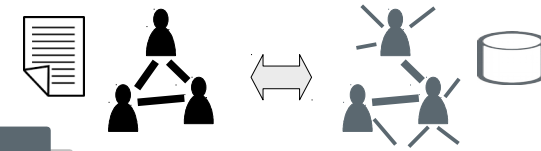
Themen-Autoren-Netzwerk



# 1. Initiale Kandidatenmenge erzeugen

# Algorithmus *dblp*

## 2. Zu große Kandidatenmengen adaptiv filtern



## 3. Weitere Kandidaten im Kontext finden

## 4. Kandidaten bewerten

- Bewert...
- Werte...

## 5. Autoren zuordnen

Kontext  
Kandidat

- ähnl...
- gem...
- zeitn...
- gleich...

Ranking

Co-A

Develop - dblpi interface Load new data 5 / 18

©2019 De Koster, Tom Van Cutsem, Theo D'Hondt  
Domains: safe-sharing, among, actors  
11-22  
https://doi.org/10.1145/2414639.2414644#view

authors: Ehsan khamespaneh **Ehsan Khamespanahi** \* 2a 3a 4a 5a  
Zeynab Sabahi-Kaveeni **Zeynab Sabahi-Kaviani** \* 1a 3a 4a 5a  
Ramtin Khorzavi **Ramtin Khorzavi** \* 1a 2a 4a 5a  
Marjan Sirjani **Marjan Sirjani** \* 1a 2a 3a 5a Behzod Sirjani \* Abolfazl Sirjani Arash Shahabi-Sirjani  
Mohammad-Javad Izadi **Mohammad-Javad Izadi** \* 1a 2a 3a 4a Mohammad Izadi \* M. Hamed Izadi \* Mohammad Hadi Izadi  
M. Izadi M. A. Izadi

Ask Yahoo: [Names] [Title] [Names & Title]  
Ask Google: [Names] [Title] [Names & Title] (e) | Google Scholar: [Names] [Title] [Names & Title] (e) | DeepDyve

title: Timed-rebecca schedulability and deadlock-freedom analysis using floating-time transition system.

pages: 23-34  
ee: <http://doi.acm.org/10.1145/2414639.2414645> visit

©Thierry Renaux, Ludic Hoque, Stefan Mann, Wolfgang De Meuter  
Digital domain recognition with soft real-time guarantees  
35-46  
https://doi.org/10.1145/214839.214841#view  
©Hajm W. Kuntz, Arnd Poetzsch-Heffter  
A regional trace logic for simple hierarchical actor-based component systems  
47-59  
https://doi.org/10.1145/2414639.2414647#view



- manuelle Zuordnung anhand der Kandidatenlisten
- im Zweifel: Recherche mit externen Quellen
- Fehler im Korpus werden durch die Vorschläge oft sichtbar!







the first resource for mathematics

# Autoren-Disambiguierung @ zbMATH

Nicolas D. Roy

Dec. 02, 2018



the first resource for mathematics

# zbMATH – die Herausforderungen

Umfang:

- 4 Mio. Dokumente → 7 Mio. Autherschaften
- 1 Mio. Autoren

# zbMATH – die Herausforderungen

Umfang:

- 4 Mio. Dokumente → 7 Mio. Autherschaften
- 1 Mio. Autoren

Herausforderungen:

- alte Daten (1826) → unvollständige Namen

# zbMATH – die Herausforderungen

**Müller**

**Complanation of the cones of second degree. (Complanation der Kegel II. Ordnung.) (German)** JFM 18.0405.01

Pr. Realg. Chemnitz (1886).

Das Integral

$$F = \frac{1}{2} \int_{\xi_0}^{\xi} \frac{\sqrt{h^2(f'^2\xi + f'^2\eta) + (\xi f'\xi + \eta f'\eta)^2}}{f'\eta} d\xi,$$

welches den Inhalt des von  $\xi_0$  bis  $\xi$  genommenen Oberflächenstückes eines Kegel mit der Leitlinie  $f(\xi, \eta) = 0$  ausdrückt, wird zunächst in den Fällen, wo die Leitlinie ein Kegelschnitt ist, auf elliptische Integrale zurückgeführt. Alsdann wird der Fall betrachtet, wo der Fusspunkt  $O$  der Höhe  $h = HO$  (wo  $O$  der Anfangspunkt,  $H$  die Spitze des Kegels) in eine Axe des Kegelschnittes fällt.

# zbMATH – die Herausforderungen

## Umfang:

- 4 Mio. Dokumente → 7 Mio. Autherschaften
- 1 Mio. Autoren

## Herausforderungen:

- alte Daten (1826) → unvollständige Namen
- Dok. aus nicht-romanischer Literatur → Transliterationsprobleme

# zbMATH – die Herausforderungen

Umfang:

- 4 Mio. Dokumente → 7 Mio. Authorschaften

## Chebyshev, Pafnutii L'vovich

**Author ID:** chebyshev.p-l   

**Published as:** Chebyshev, P. L.; Tchebichef, P.; Tchebycheff; Tchebycheff, P.; Tchébycheff; Tchébycheff, P.; Tschebychew, P.; Tschebyscheff, P.; Tschebyscheff, P. L.; Tschebyschew, P.; Tschebyschew, P. L.; Čebišev, P. L.

# zbMATH – die Herausforderungen

Umfang:

- 4 Mio. Dokumente → 7 Mio. Authorschaften
- 1 Mio. Autoren

Herausforderungen:

- alte Daten (1826) → unvollständige Namen
- Dok. aus nicht-romanischer Literatur → Transliterationsprobleme
- **Single-Author-Tradition** in Mathematik → Koautoren-basierte Algorithmen weniger effizient (*1,77 Koautoren im Durchschnitt*)

# Halbautom. Autoren-Disambiguierung

1. algorithmische Disambiguierung
2. editorielle Kuration

*3,5% der Autoren - 19% der Autorschaften*



# Halbautom. Autoren-Disambiguierung

1. algorithmische Disambiguierung
2. editorielle Kuration

*3,5% der Autoren - 19% der Autorschaften*

Datenstruktur erlaubt:

- "fuzzy/multi"-Zuordnungen
- negative Zuordnungen

# Halbautom. Autoren-Disambiguierung

1. algorithmische Disambiguierung

2. editorielle Kuratierung

*3,5% der Autoren - 19*



Datenstruktur erlaubt:

- "fuzzy/multi"-Zuordnungen
- **negative** Zuordnungen

# Halbautom. Autoren-Disambiguierung

1. algorithmische Disambiguierung

2. editorielle Kuratierung

3,5% der Autoren - 19



Datenstruktur erlaubt:




- "fuzzy/multi"-Zuordnungen
- **negative** Zuordnungen

Autorenprofile mit:

- **biographischen** Daten (geb., PhD...)
- **Verlinkung** mit anderen Diensten

# Halbautom. Autoren-Disambiglierung

## Dieudonné, Jean Alexandre

**Author ID:** dieudonne.jean-alexandre   

**Published as:** Dieudonné, J.; Dieudonné, J. A.; Dieudonné, Jean;  
Dieudonné, Jean A.; Dieudonné, Jean Alexandre

**External Links:** [MGP](#) · [Wikidata](#) · [dblp](#) · [GND](#) · [MacTutor](#)

**Member of Collective:** [Bourbaki, Nicolas](#)

- **negative Zuordnungen**

Autorenprofile mit:

- **biographischen** Daten (geb., PhD...)
- **Verlinkung** mit anderen Diensten

# Halbautom. Autoren-Disambiguierung

1. algorithmische Disambiguierung

2. editorielle Kuratierung

3,5% der Autoren - 19



Datenstruktur erlaubt:

- "fuzzy/multi"-Zuordnungen
- **negative** Zuordnungen

Autorenprofile mit:

- **biographischen** Daten (geb., PhD...)
- **Verlinkung** mit anderen Diensten

Dokumente mit:

- **semantischen Metadaten**

**Delzant, Thomas; Gromov, Misha**

**Cuts in Kähler groups. (English)** Zbl 1116.32016

**MSC:**

- 32Q15 Kähler manifolds
- 20F65 Geometric group theory
- 57M07 Topological methods in group theory

**Keywords:**

fundamental groups of Kähler manifolds; small cancelation groups

Dokumente mit:

- semantischen Metadaten

(KW, MSC)

# Algorithmische Disambiguierung

basiert auf:

- **Autorenname**  
(*mit Abkürzung, Transliterations-Varianten...*)

# Algorithmische Disambiguierung

basiert auf:

- **Autorenname**  
(mit Abkürzung, Transliterations-Varianten...)





# Algorithmische Disambiguierung




basiert auf:

- **Autorenname**  
(*mit Abkürzung, Transliterations-Varianten...*)
- **zeitliche Information**  
(*Publikationsjahr, biographische Daten, Publikationsprofile*)

A  
b

iguierung

**Müller, Jan Steffen**

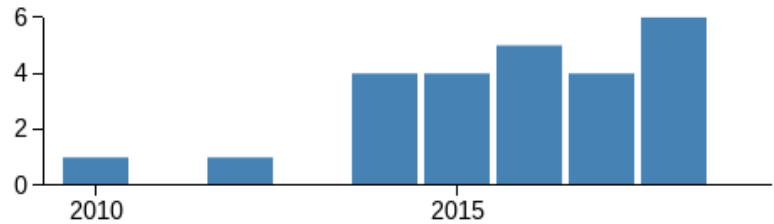
Author ID: [muller.jan-steffen](#)   

Published as: Mueller, J.; Müller, J.; Müller, J. Steffen; Müller, J.-S.; Müller, Jan; Müller, Jan Steffen

External Links: [MGP](#)

Documents Indexed: [25 Publications](#) since 2010

Publications by Year



Year	Publications
2010	1
2011	1
2012	4
2013	4
2014	5
2015	4
2016	6

**Müller, Johann**

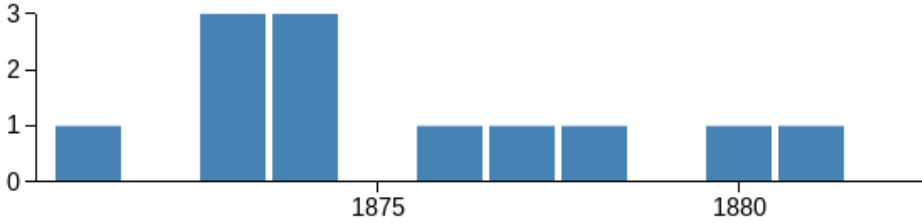
Author ID: [muller.johann](#)

Published as: Müller, Johann

External Links: [MGP](#)

Documents Indexed: [10 Publications](#) since 1870

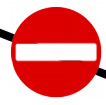
Publications by Year



Year	Publications
1870	1
1871	3
1872	3
1873	1
1874	1
1875	1
1876	1
1877	1
1878	1
1879	1
1880	1

ographische Daten,

[Müller, J.; Vlachou, V.; Yavrian, A.](#)  
 Overconvergent series of rational functions and universal Laurent series. (English) [Zbl 1156.30030](#)  
 J. Anal. Math. 104, 235-245 (2008).



# Algorithmische Disambiguierung

basiert auf:

- **Autorenname**  
*(mit Abkürzung, Transliterations-Varianten...)*
- **zeitliche Information**  
*(Publikationsjahr, biographische Daten, Publikationsprofile)*
- **Koautoren**

# Verlinkung mit anderen Diensten

- In Autorenprofilen werden **Identifiers** gespeichert:  
*Wikidata, GND, Math. Genealogy Project, Mathnet.ru, dblp, arXiv, MathOverflow, ResearchGate, ...*
- *Die Profile werden algorithmisch und manuell **gematcht***
- Aus den Verlinkungen werden **Metadaten** gesammelt:  
*Geburtsjahr, PhD-Jahr, andere Identifiers...*

Author: Thalheim, Bernhard 

code: thalheim.bernhard 





































ID: 416205

### Time events

birth start phd stop death

1952 YYYY 1979 YYYY YYYY

### External Ids

- Agnes Scott    
- MacTutor    
- Theses.fr    
- DBLP    
- arXiv    
- Zbl Rev.    
- MGP    
- Wikidata    
- Math-Net.Ru    
- GND    
- ORCID    
- Celebratio    
- ResearchGate    
- MathOverflow    
- Google Scholar    
- mathscinet    
- Homepage    
- Email    

Author: Thalheim, Bernhard 

code: thalheim.bernhard 

ID: 416205



















### Time events

birth start phd stop death

1952 YYYY 1979 YYYY YYYY



### External Ids

- Agnes Scott 
- MacTutor 
- Theses.fr 
- DBLP 
- arXiv 
- Zbl Rev. 
- MGP 
- Wikidata 
- Math-Net.Ru 
- GND 
- ORCID 
- Celebratio 
- ResearchGate 
- MathOverflow 
- Google Scholar 
- mathscinet 
- Homepage 
- Email 

Automatic matching

Author: Thalheim, Bernhard

code: thalheim.bernhard

ID: 416205

### Time events

birth start phd stop death

1952 YYYY 1979 YYYY YYYY

Source: wikidata

### External Ids

- Agnes Scott
- MacTutor
- Theses.fr
- DBLP  t/BernhardThalheim
- arXiv
- Zbl Rev. 6270
- MGP 139910
- Wikidata  Q91452
- Math-Net.Ru
- GND 111056888
- ORCID  0000-0002-7909-7786
- Celebratio
- ResearchGate
- MathOverflow
- Google Scholar
- mathscinet
- Homepage
- Email

Automatic matching

Author: Thalheim, Bernhard

code: thalheim.bernhard

ID: 416205

### Time events

birth start phd stop death

1952

YYYY

1979

YYYY

YYYY

Source: wikidata

### External Ids

Agnes Scott

MacTutor

Theses.fr

DBLP

arXiv

Zbl Rev.

MGP

Wikidata

Math-Net.Ru

GND

ORCID

ResearchGate

MathOverflow

Google Scholar

mathscinet

Homepage

Email

Source: wikidata

Automatic matching

Source: wikidata



Author: Thalheim, Bernhard

code: thalheim.bernhard

ID: 416205

### Time events

birth	start	phd	stop	death
1952	YYYY	1979	YYYY	YYYY

Source: wikidata

Source: MGP

### External Ids

Agnes Scott		
MacTutor		
Theses.fr		
DBLP	t/BernhardThalheim	
arXiv		
Zbl Rev.	6270	
MGP	139910	
Wikidata	Q91452	
Math-Net.Ru		
GND	111056888	
ORCID	0000-0002-7909-7786	
ResearchGate		
MathOverflow		
Google Scholar		
mathscinet		
Homepage		
Email		

Source: wikidata

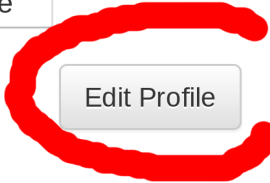
Automatic matching

Source: wikidata

# Editorielle Kuration der Daten

Autorschaften-Zuordnungen können korrigiert/berarbeitet werden von:

- dem editorielle Team
- jedem zbMATH-Benutzer über **User-Interface**



Edit Profile

# Müller, Jan Steffen

Author ID: muller.jan-steffen    

Published as: Mueller, J.; Müller, J.; Müller, J. Steffen; Müller, J.-S.; Müller, Jan; Müller, Jan Steffen

External Links: [MGP](#)

Compute Distance To:

Paul Erdős

Documents Indexed: [24 Publications](#) since 2010

## Co-Authors

all ▾

- 3 single-authored
- 3 Balakrishnan, Jennifer S.
- 2 Besser, Amnon
- 2 Kühn, Ulf
- 1 Busch, Vincenz

## Journals

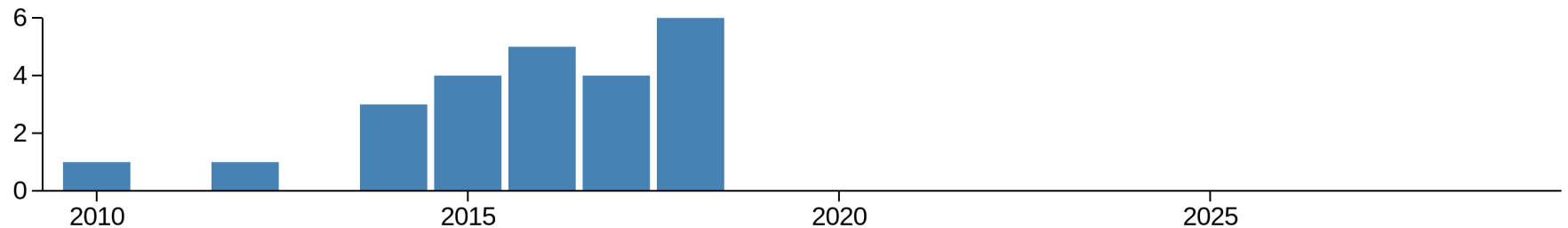
all ▾

- 3 Mathematics of Computation
- 1 Mathematical Proceedings of the Cambridge Philosophical Society
- 1 Functiones et Approximatio. Commentarii Mathematici

## Fields

- 9 Number theory (11-XX)
- 8 Algebraic geometry (14-XX)

## Publications by Year



## Author muller.jan-steffen

Relation to author:

[Other](#)[Change](#)Author ID: [muller.jan-steffen](#)

Name: Müller, Jan Steffen

Published as: Müller, Jan Steffen; Müller, J. Steffen; Müller, J.; Mueller, J.

[Confirm/Exclude](#)[Merge](#)[Add External Links](#)[Comment](#)

## Confirm or exclude documents currently assigned to muller.jan-steffen

[Cancel](#)[Done](#)

Documents without a confirm button have already been verified.

7/16 publications confirmed

Hofmann, F.; **Müller, J.**[Relativistic tests with lunar laser ranging](#) [Zbl 1382.83021](#)

Classical Quantum Gravity 35, No. 3, Article ID 035015, 26 p. (2018).

[Confirm](#)[Exclude](#)Balakrishnan, Jennifer S.; **Müller, J. Steffen**; Stein, William A.[A  \$p\$ -adic analogue of the conjecture of Birch and Swinnerton-Dyer for modular abelian varieties.](#)[Zbl 06552505](#)

Math. Comput. 85, No. 298

[Exclude](#)Kohler, D.; **Müller, J.**; Wever, U.[Cellular non-deterministic automata and partial differential equations](#) [Zbl 1364.37036](#)

Physica D 311-312, 1-16 (2015).

[Confirm](#)[Exclude](#)Kühn, Ulf; **Müller, Jan Steffen**[A height inequality for rational points on elliptic curves implied by the abc-conjecture](#) [Zbl 1395.11090](#)

Funct. Approximatio, Comment. Math. 52, No. 1, 127-132 (2015).

[Exclude](#)[Müller, Jan Steffen](#)[Exclude](#)

## Author muller.jan-steffen

Relation to author:  
[Other](#)  
[Change](#)

**Author ID:** [muller.jan-steffen](#)  
**Name:** Müller, Jan Steffen  
**Published as:** Müller, Jan Steffen; Müller, J. Steffen; Müller, J.; Mueller, J.

Personal Website

### Add Links or IDs from these Services

Service	Link	or	ID
MGP	<input type="text" value="http://genealogy.math.ndsu.nodak.edu/id.php?id=38586"/>		<input type="text" value="38586"/>
			We found this ID: <a href="#">150323</a> <input type="button" value="Copy ID"/>
Wikidata	<input type="text" value="http://www.wikidata.org/wiki/Q77141"/>		<input type="text" value="Q77141"/>
Math-Net.Ru	<input type="text" value="http://www.mathnet.ru/php/person.phtml?option_lang=eng&amp;personid=14007"/>		<input type="text" value="14007"/>
ORCID	<input type="text" value="http://orcid.org/0000-0003-1279-3709"/>		<input type="text" value="0000-0003-1279-3709"/>



# NLP-basierte Autoren-Disambiguierung @ HITS

Mark-Christoph Müller

4. Dezember 2018



# Autoren-Disambiguierung @ HITS

- *Ähnlichkeit* von Titeln, Abstracts, etc.

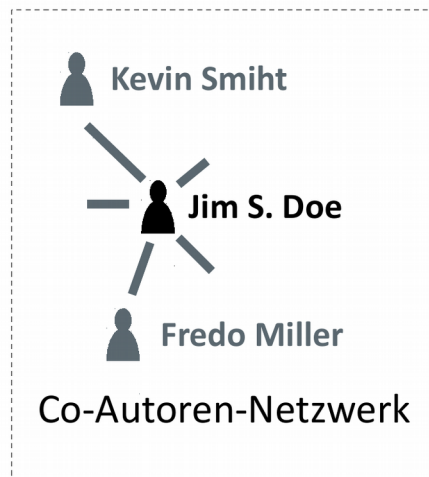
# Autoren-Disambiguierung @ HITS

## 3. Weitere Kandidaten im Kontext finden

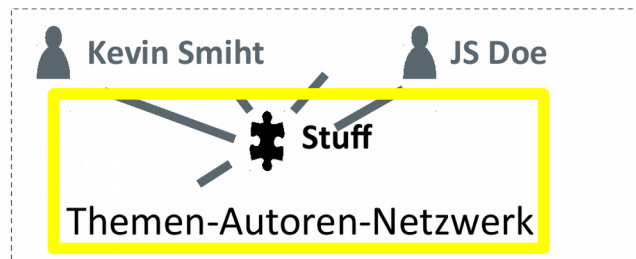
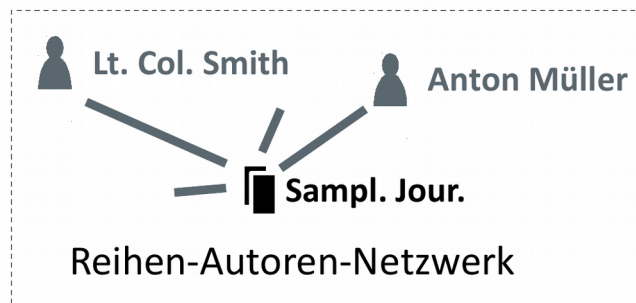
- Ausnutzen von **semantischen** Zusammenhängen
- darf **teurere** Ähnlichkeitsmaße verwenden

  J. Doe,  A. Miller,  K. Smith: „Important  Stuff.“  Sampl. Jour.,  1975

Kontext der  
Kandidatenmenge:



weitere Kontexte, z.B.:



Mögliche Maße:

- Levenshtein
- Jaro-Winkler
- n-Gramme
- Soundex-Varianten mit Anpassungen an Personennamen

Teil des  
*dblp*  
Workflow



# Autoren-Disambiguierung @ HITS

- *Ähnlichkeit* von Titeln, Abstracts, etc.
- Ziel
  - Anwendung aktueller NLP-Methoden
- Motivation
  - Verbesserung der Effektivität
  - Höhere Bedeutung bei schwachen Co-Autor-Netzwerken („*Single-author*“-Tradition)
- Weitere Anforderungen
  - Effizienz
  - Nachvollziehbarkeit

# Semantik in der A-D: Beispiel

2004 IEEE International Conference on Multimedia and Expo (ICME)

## ARTICULATORY CLASS BASED SPECTRAL ENVELOPE REPRESENTATION FOR VOICE FONTS

*Ashish Verma*

IBM India Research Lab  
Indian Institute of Technology  
New Delhi, India 110016  
vashish@in.ibm.com

*Arun Kumar*

Center for Applied Research in Electronics  
Indian Institute of Technology  
New Delhi, India 110016  
arunkm@care.iitd.ernet.in

IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 6, NOVEMBER 2006

## A Multiresolution Model of Auditory Excitation Pattern and Its Application to Objective Evaluation of Perceived Speech Quality

Abhijit Karmakar, *Student Member, IEEE*, *Arun Kumar*, *Member, IEEE*, and R. K. Patney, *Senior Member, IEEE*

# Semantik in der A-D: Beispiel

2004 IEEE International Conference on Multimedia and Expo (ICME)

## ARTICULATORY CLASS BASED SPECTRAL ENVELOPE REPRESENTATION FOR VOICE FONTS

*Ashish Verma*

*Arun Kumar*

IBM India Research Lab  
Indian Institute of Technology  
New Delhi, India 110016  
vashish@in.ibm.com

Center for Applied Research in Electronics  
Indian Institute of Technology  
New Delhi, India 110016  
arunkm@care.iitd.ernet.in

Keine  
String-  
ähnlichkeit

IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 6, NOVEMBER 2006

## A Multiresolution Model of Auditory Excitation Pattern and Its Application to Objective Evaluation of Perceived Speech Quality

Abhijit Karmakar, *Student Member, IEEE*, Arun Kumar, *Member, IEEE*, and R. K. Patney, *Senior Member, IEEE*

# Semantik in der A-D: Beispiel

2004 IEEE International Conference on Multimedia and Expo (ICME)

## ARTICULATORY CLASS BASED SPECTRAL ENVELOPE REPRESENTATION FOR VOICE FONTS

*Ashish Verma*

*Arun Kumar*

IBM India Research Lab  
Indian Institute of Technology  
New Delhi, India 110016  
vashish@in.ibm.com

Center for Applied Research in Electronics  
Indian Institute of Technology  
New Delhi, India 110016  
arunkm@care.iitd.ernet.in

Semantische  
Nähe

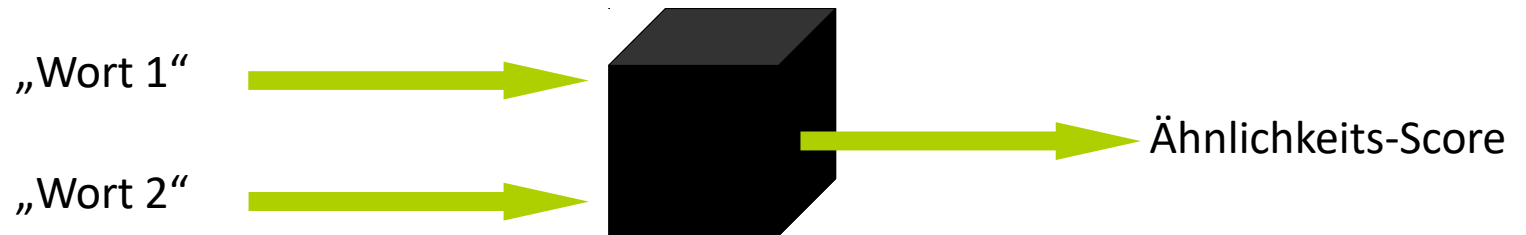
IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 6, NOVEMBER 2006

## A Multiresolution Model of Auditory Excitation Pattern and Its Application to Objective Evaluation of Perceived Speech Quality

Abhijit Karmakar, *Student Member, IEEE*, Arun Kumar, *Member, IEEE*, and R. K. Patney, *Senior Member, IEEE*

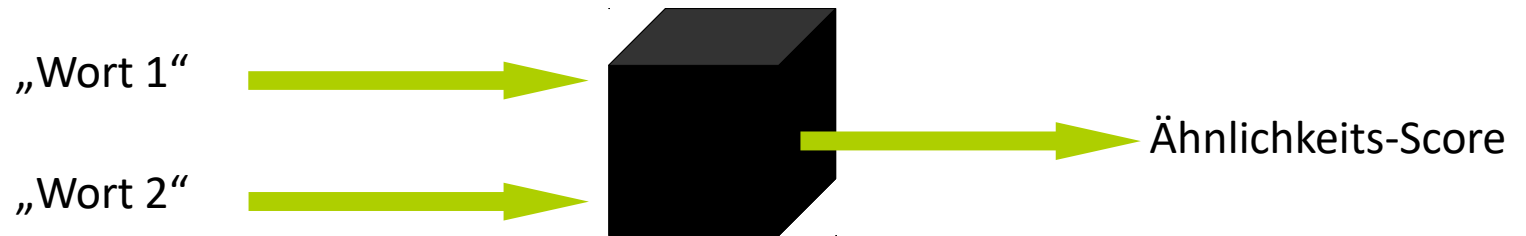
# NLP-Semantik

- Aktueller Standard: *Word Embeddings*
- Repräsentation von Wortbedeutung als Folge von (z.B. 200) Zahlen
- Wortvergleich durch einfache **Berechnung**



# NLP-Semantik

- Aktueller Standard: *Word Embeddings*
- Repräsentation von Wortbedeutung als Folge von (z.B. 200) Zahlen
- Wortvergleich durch einfache **Berechnung**



- Vollautomatisch generierbar aus großen Textmengen, z.B.
  - Wikipedia, Nachrichten: Allgemein

# NLP-Semantik

- Aktueller Standard: *Word Embeddings*

- Repräsentation von Wörtern als Folge von

- Wort

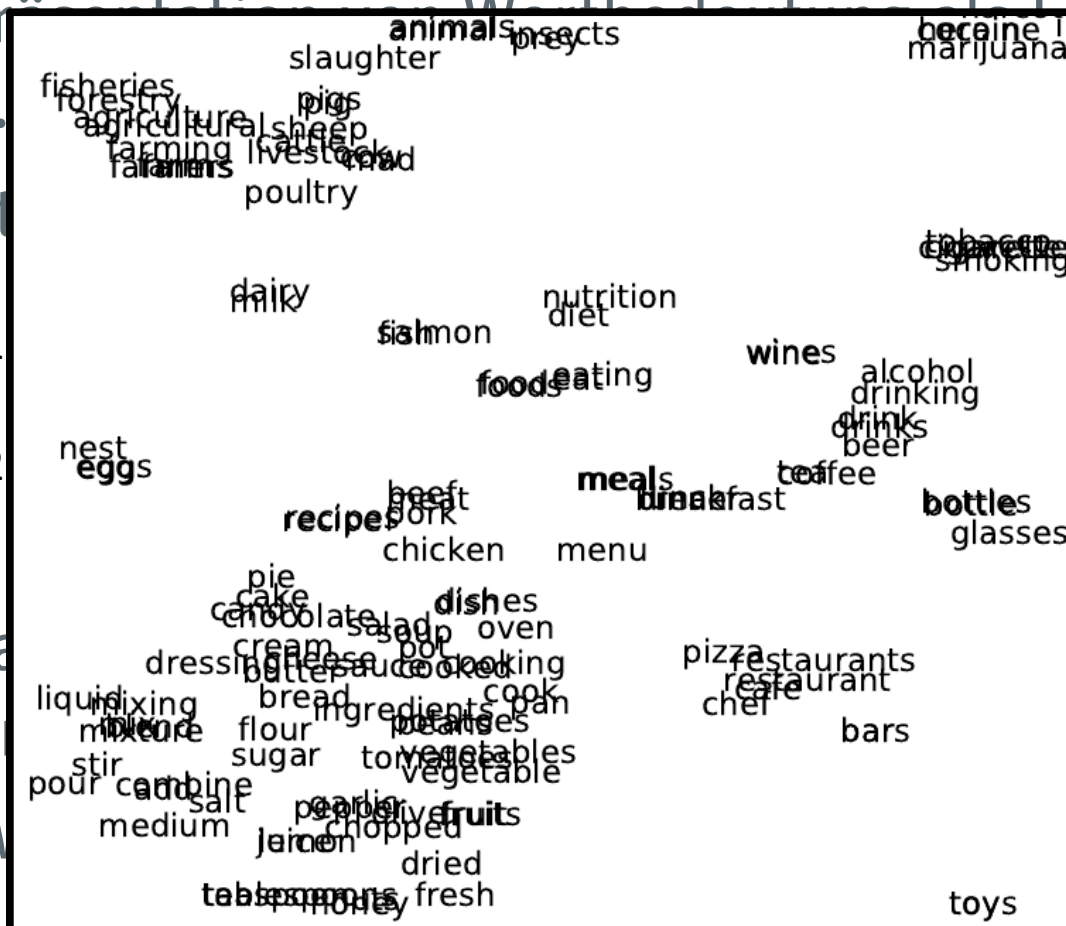
„Wort 1

„Wort 2

- Volla

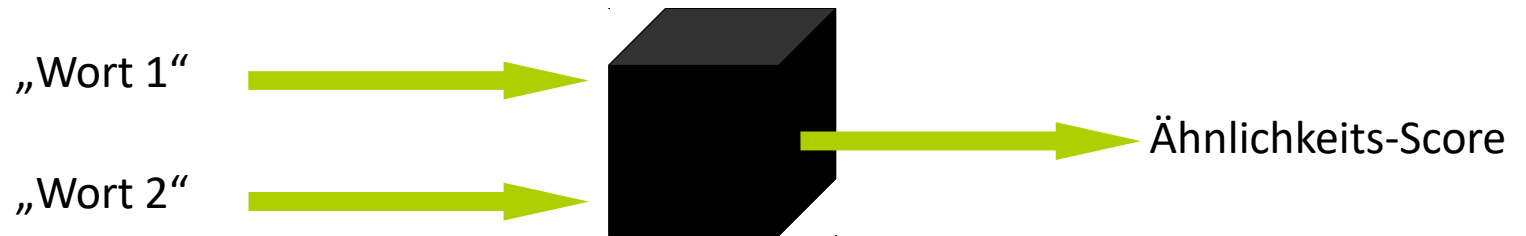
Text

- W



# NLP-Semantik

- Aktueller Standard: *Word Embeddings*
- Repräsentation von Wortbedeutung als Folge von (z.B. 200) Zahlen
- Wortvergleich durch einfache **Berechnung**



- Vollautomatisch generierbar aus großen Textmengen, z.B.
  - Wikipedia, Nachrichten: Allgemein
  - dblp-Titel: Domänenspezifisch



# NLP-Semantik

- Aktueller Standard: *Word Embeddings*

- Repr (z.B.)

- Wort

„Wort 1

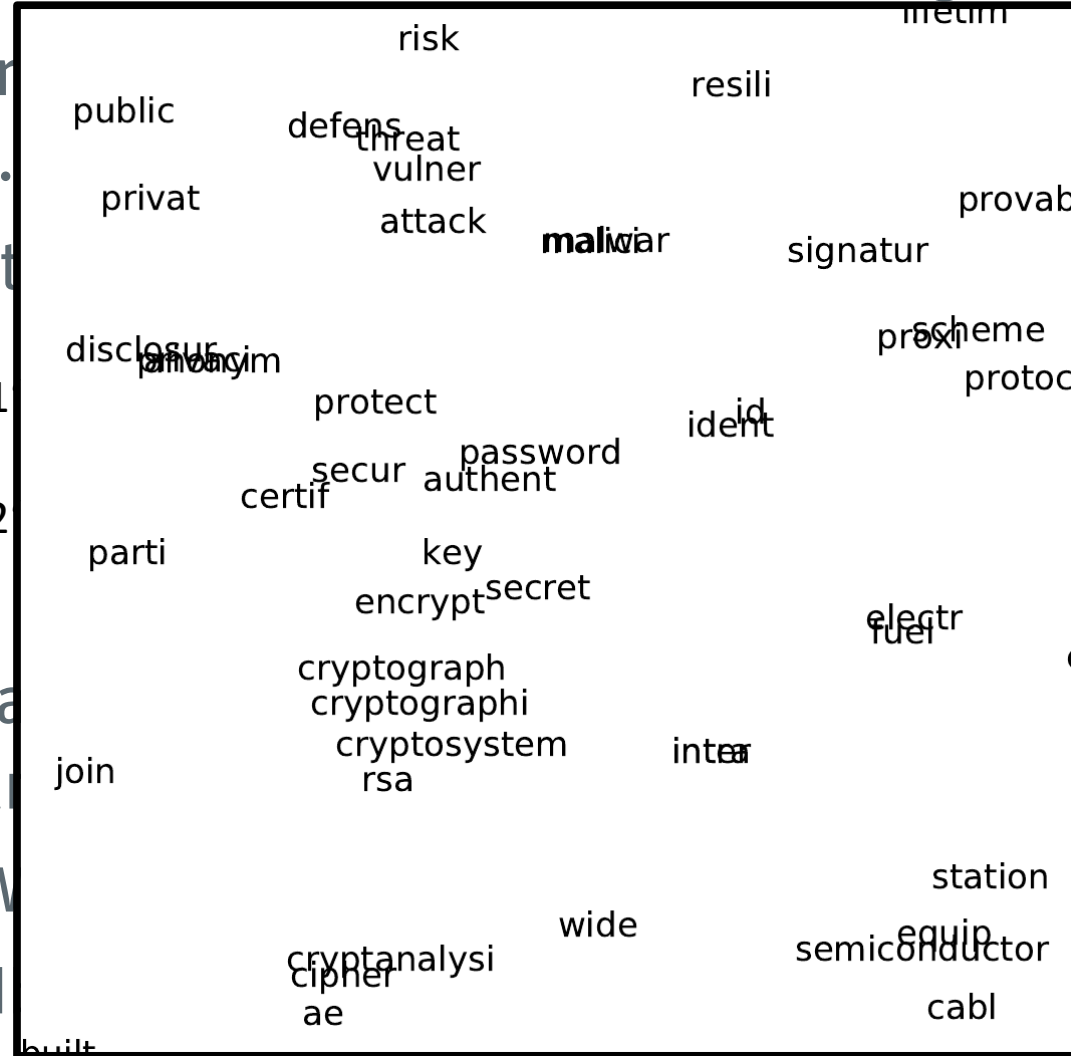
„Wort 2

- Volla

Text

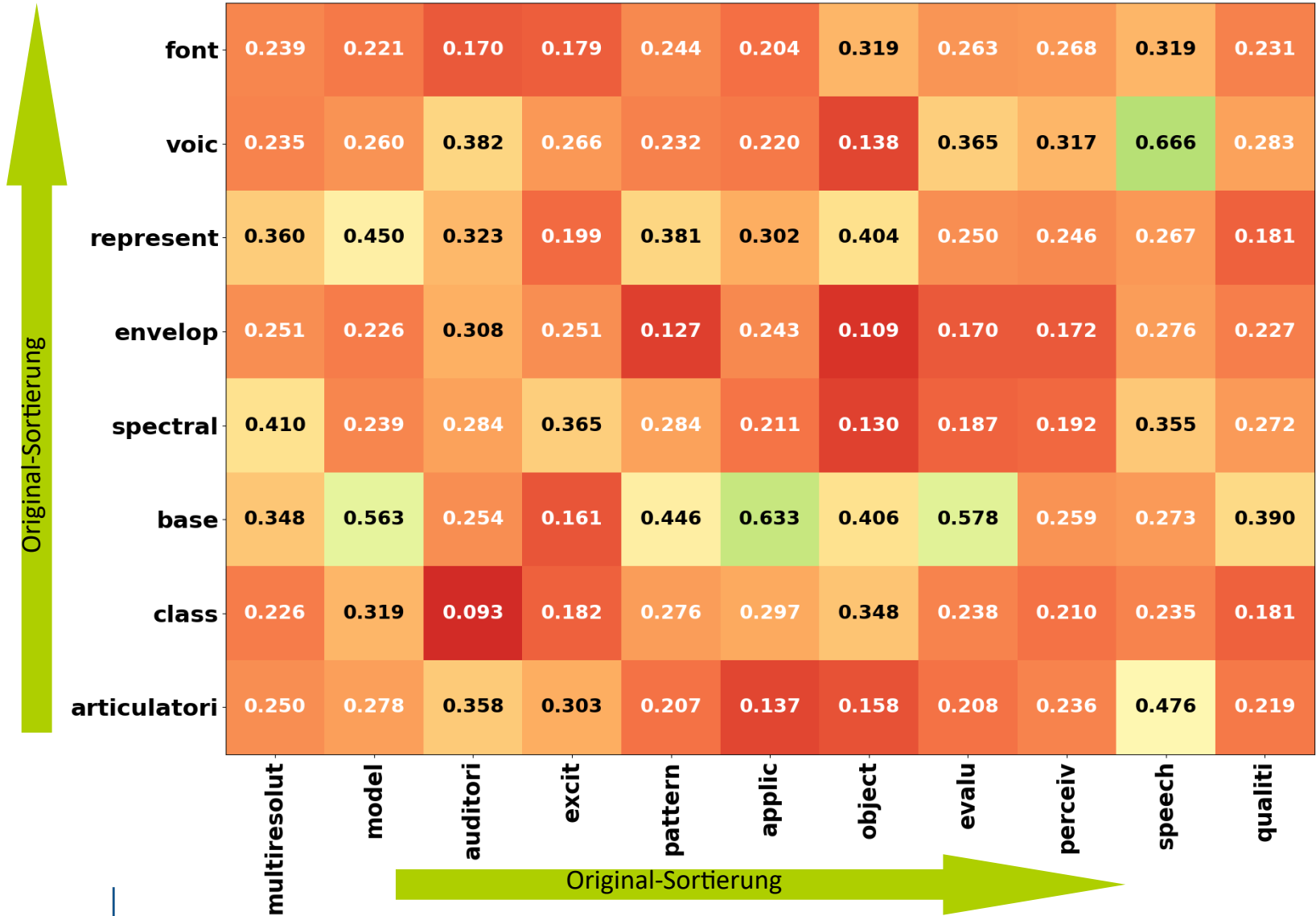
- W

- d



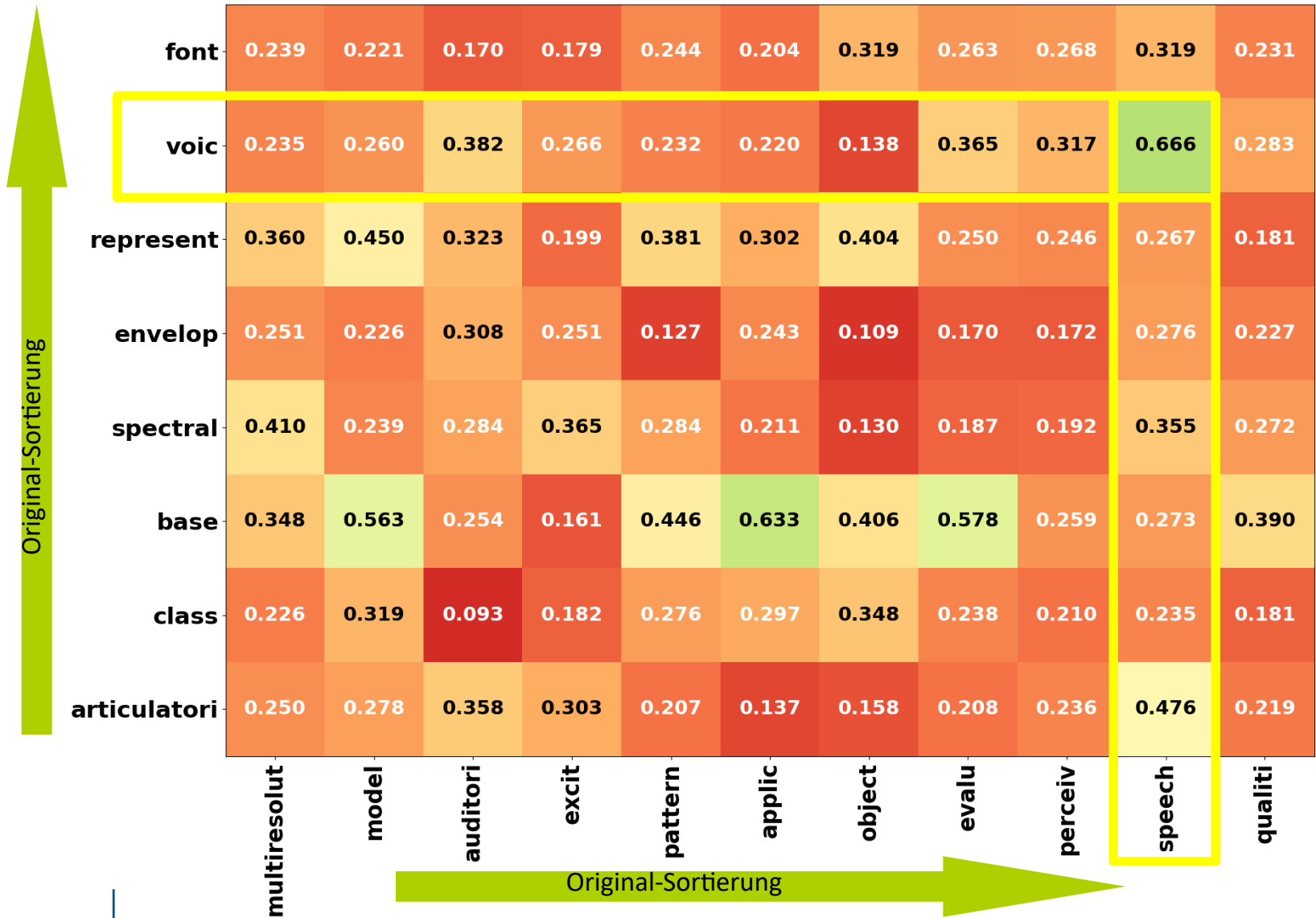
# Semant. Matching von Titeln

['algo:sg', 'dataset:dblp', 'dims:200', 'fold:1', 'iters:20', 'mc:5', 'neg:20', 'phrases:no', 'unit:stem', 'w:5']



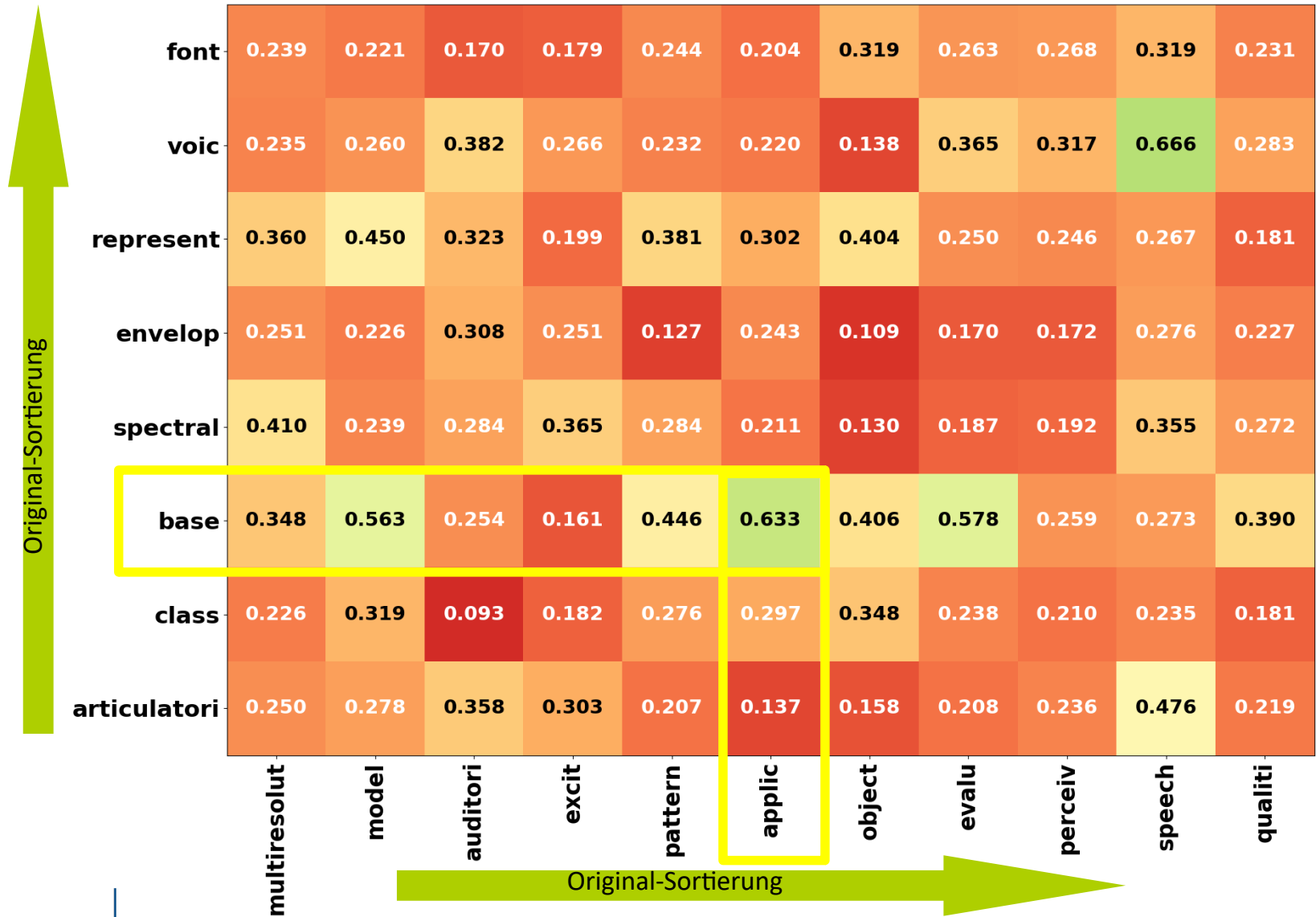
# Semant. Matching von Titeln

['algo:sg', 'dataset:dblp', 'dims:200', 'fold:1', 'iters:20', 'mc:5', 'neg:20', 'phrases:no', 'unit:stem', 'w:5']



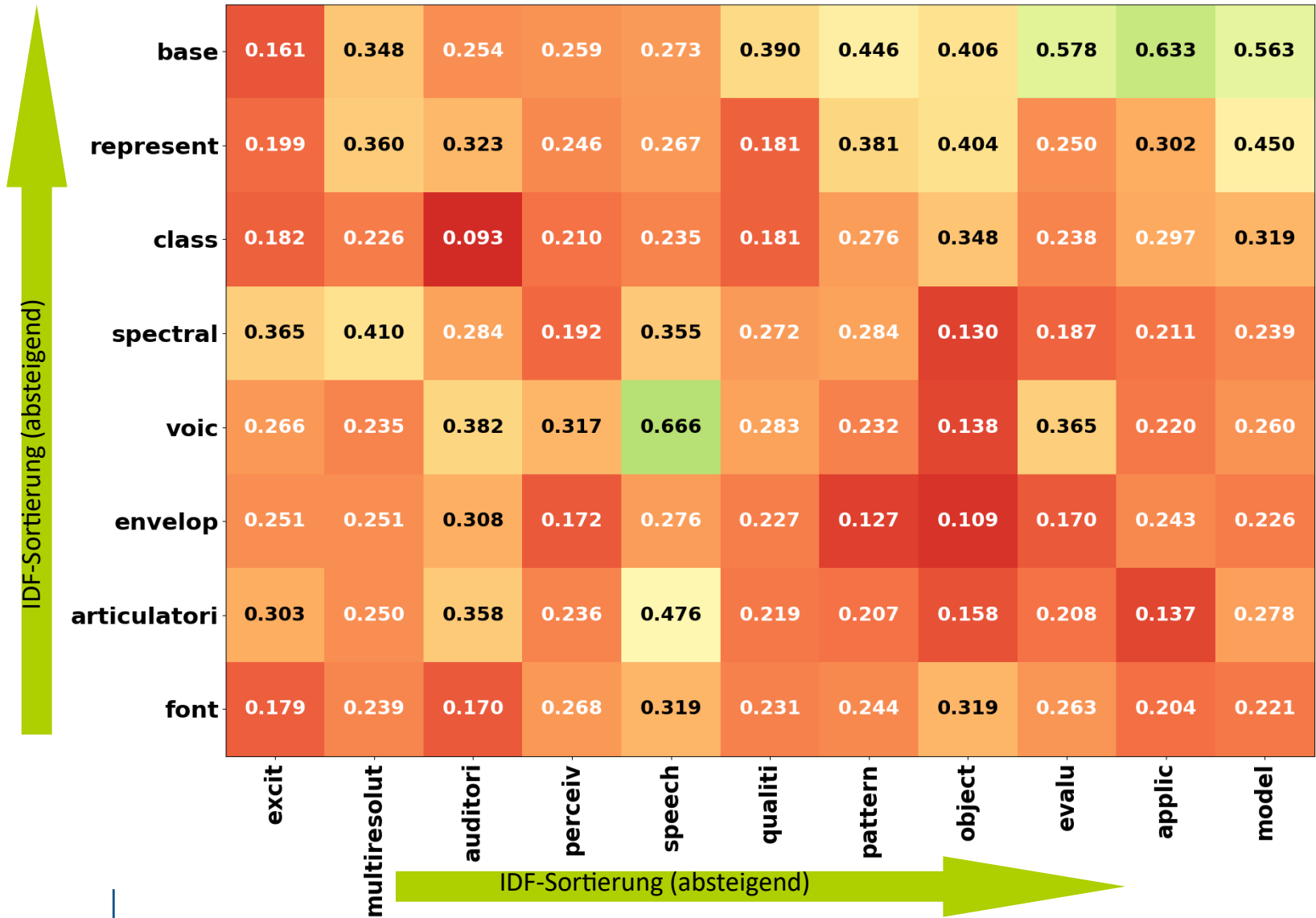
# Semant. Matching von Titeln

['algo:sg', 'dataset:dblp', 'dims:200', 'fold:1', 'iters:20', 'mc:5', 'neg:20', 'phrases:no', 'unit:stem', 'w:5']



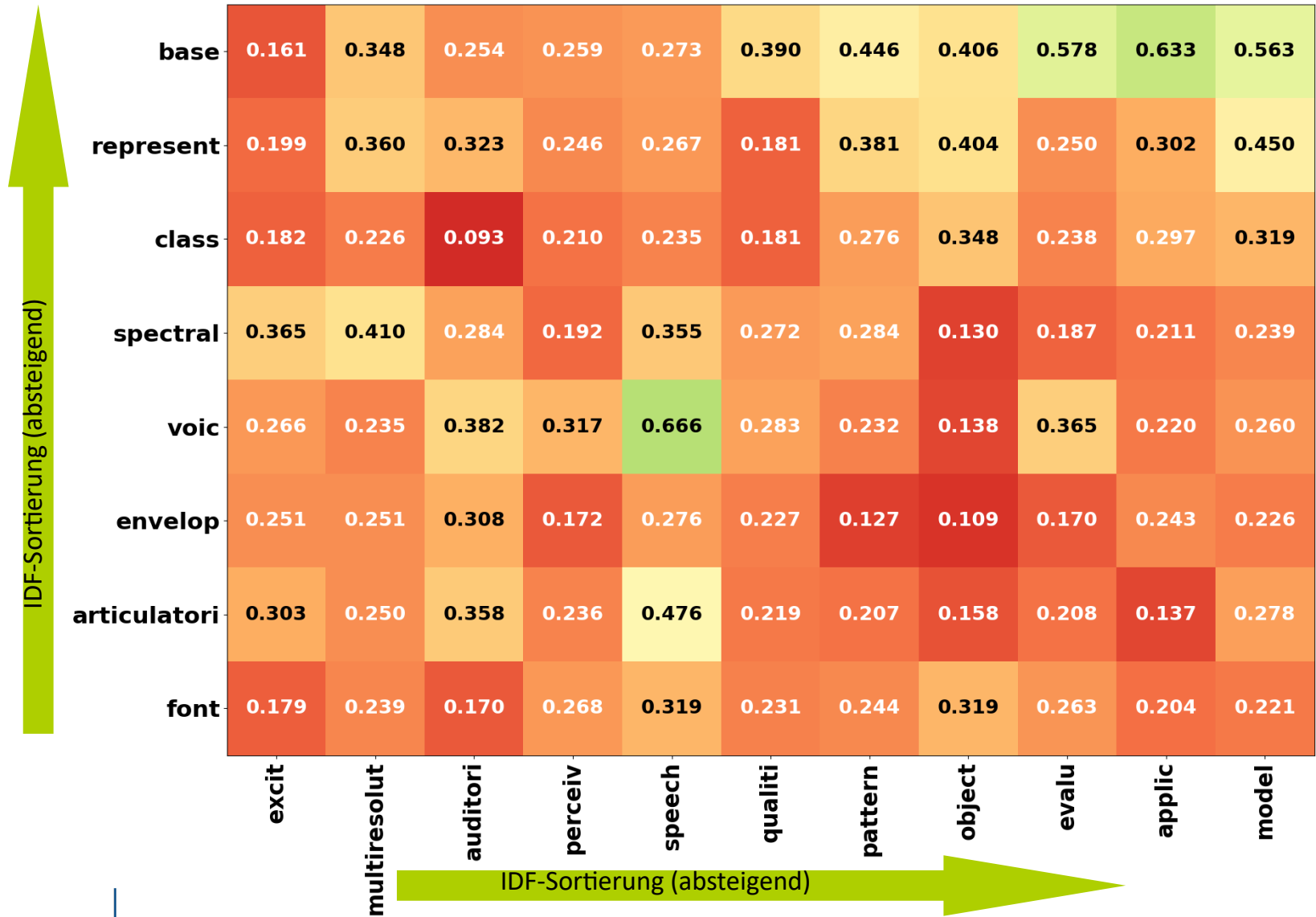
# Semant. Matching von Titeln

[ 'algo:sg', 'dataset:dblp', 'dims:200', 'fold:1', 'iters:20', 'mc:5', 'neg:20', 'phrases:no', 'unit:stem', 'w:5' ]



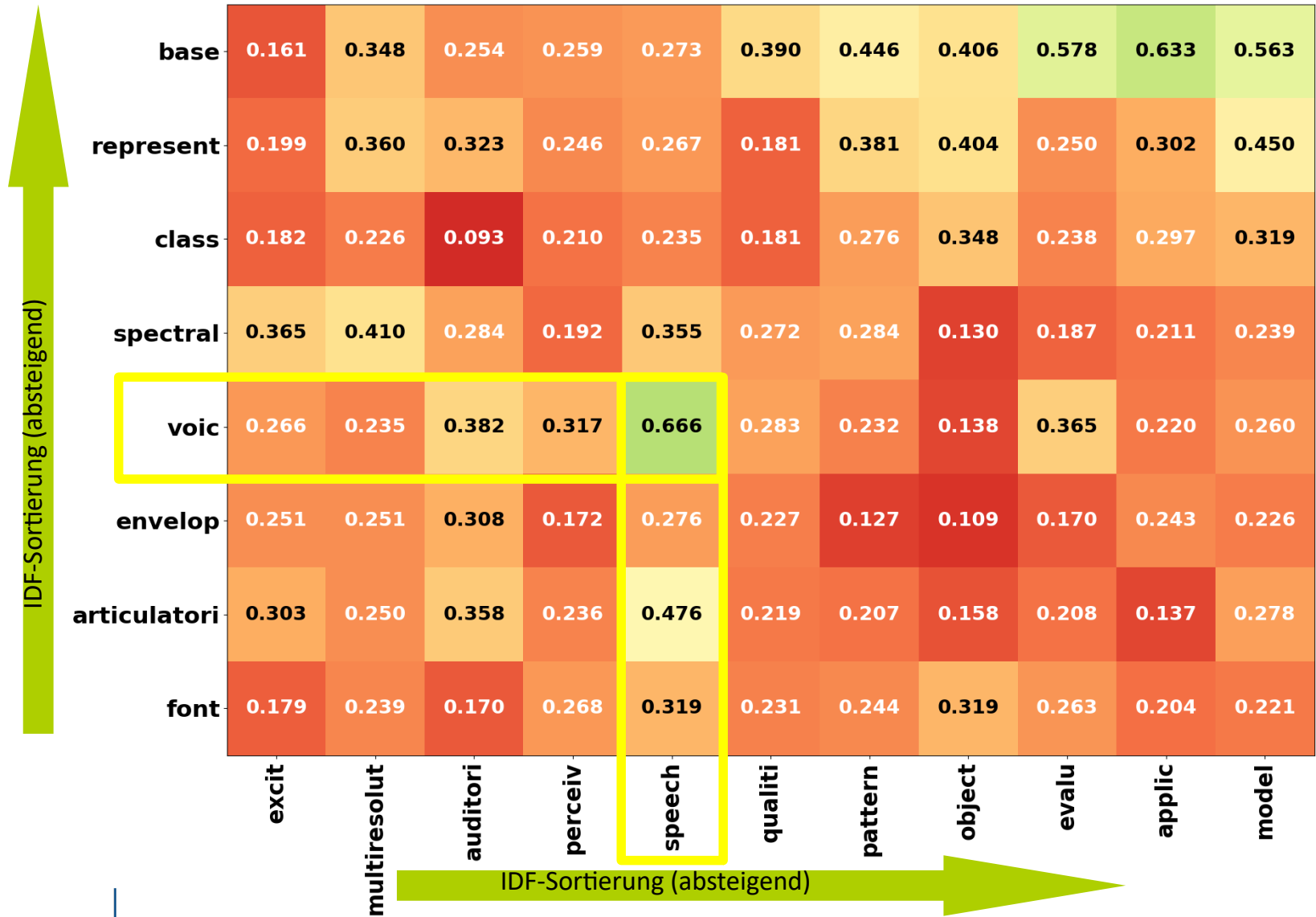
# Semant. Matching von Titeln

['algo:sg', 'dataset:dblp', 'dims:200', 'fold:1', 'iters:20', 'mc:5', 'neg:20', 'phrases:no', 'unit:stem', 'w:5']



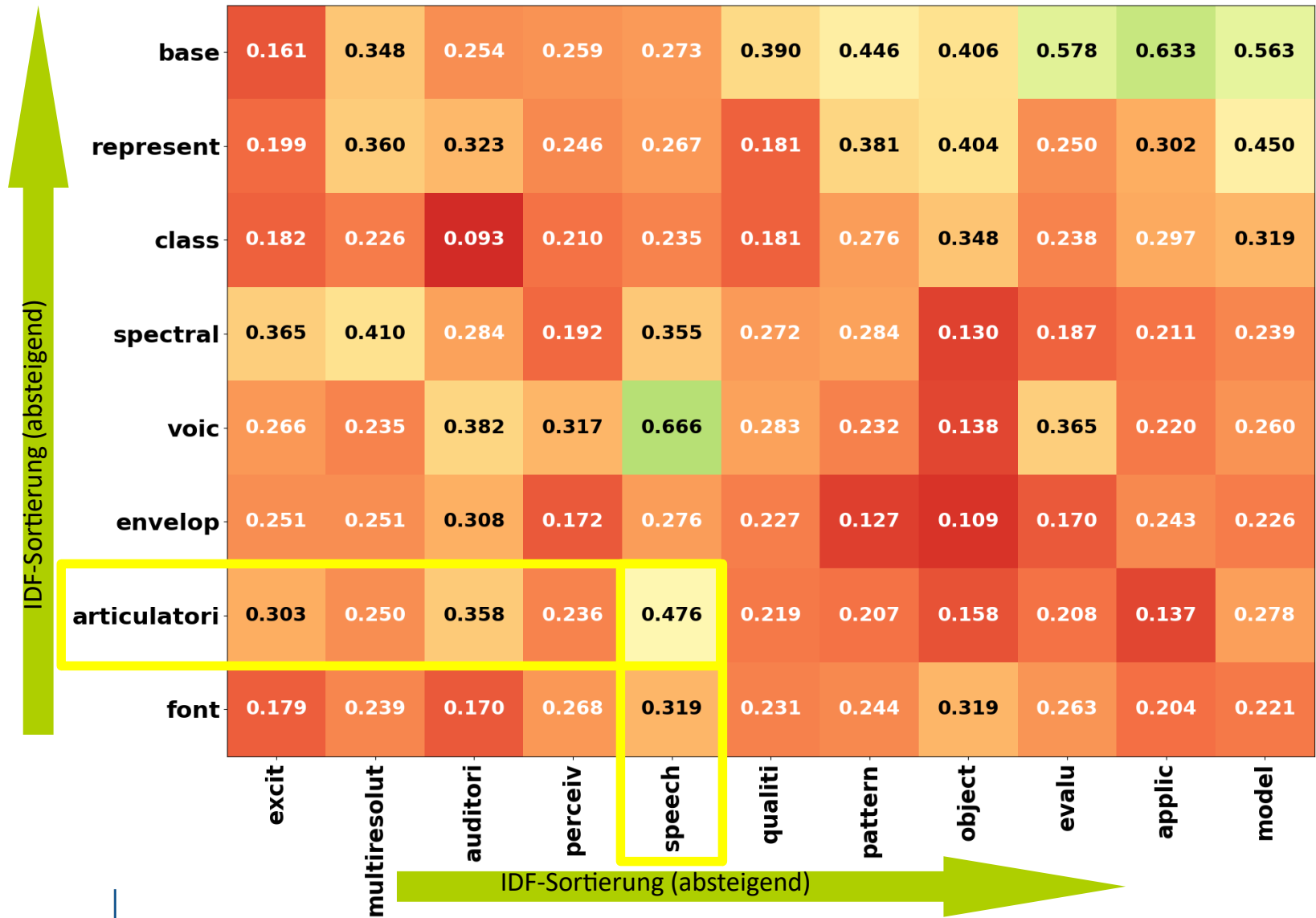
# Semant. Matching von Titeln

['algo:sg', 'dataset:dblp', 'dims:200', 'fold:1', 'iters:20', 'mc:5', 'neg:20', 'phrases:no', 'unit:stem', 'w:5']



# Semant. Matching von Titeln

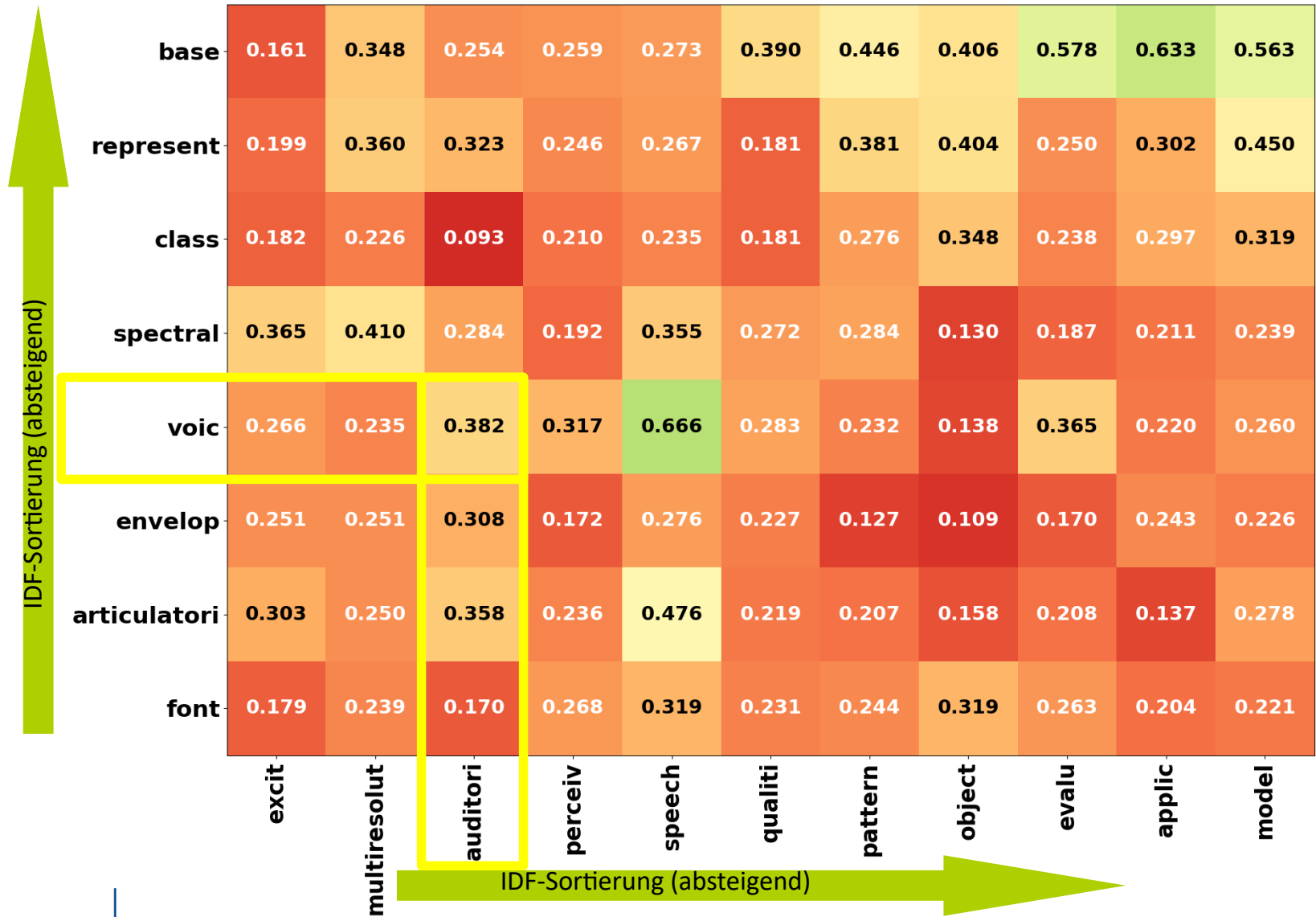
[ 'algo:sg', 'dataset:dblp', 'dims:200', 'fold:1', 'iters:20', 'mc:5', 'neg:20', 'phrases:no', 'unit:stem', 'w:5' ]





# Semant. Matching von Titeln

['algo:sg', 'dataset:dblp', 'dims:200', 'fold:1', 'iters:20', 'mc:5', 'neg:20', 'phrases:no', 'unit:stem', 'w:5']



# Integration in die Praxis

- Unterstützung existierender Prozesse, „*human in the loop*“
- Semantische Methoden
  - sind teu(r)er, aber
  - werden erst in späterer Disambiguierungs-Phase eingesetzt, um
    - fehlende bzw. schwache Co-Autor- und/oder String-Matching-Evidenz zu ergänzen

# Integration in die Praxis

- Unterstützung existierender Prozesse, „*human in the loop*“
- Semantische Methoden
  - sind teu(r)er, aber
  - werden erst in späterer Disambiguierungs-Phase eingesetzt, um
    - fehlende bzw. schwache Co-Autor- und/oder String-Matching-Evidenz zu ergänzen
- Weiteres numerisches Feature für Kandidaten-Ranking

# Integration in die Praxis

## 4. Kandidaten bewerten

- Bewertung anhand mehrdimensionaler Features
- Werte je zwischen 0 und 1



ähnlicher Name



publiziert in gleicher Reihe



gemeinsame Publikation



publiziert in *ähnlicher* Reihe



zeitnahe Publikationen



verwendet gleiche Themen



gleiche Affiliation

ggf. weitere ...

- Ranking mittels gewichteter Summe

# Integration in die Praxis

- Unterstützung existierender Prozesse, „*human in the loop*“
- Semantische Methoden
  - sind teu(r)er, aber
  - werden erst in späterer Disambiguierungs-Phase eingesetzt, um
    - fehlende bzw. schwache Co-Autor- und/oder String-Matching-Evidenz zu ergänzen
- Weiteres numerisches Feature für Kandidaten-Ranking
- Zustandekommen der Ähnlichkeits-Scores für menschl. Bearbeiter nachvollziehbar

# Vielen Dank!



<https://dblp.org>  
dblp@dagstuhl.org



<https://zbmath.org>  
authorid@zbmath.org

Heidelberg Institute for  
Theoretical Studies



<https://h-its.org/nlp>  
mark-christoph.mueller@h-its.org

# In Kontakt bleiben?

Mailingliste zum gemeinsamen Austausch  
für alle Interessierten:

**authid@portal.dagstuhl.de**

**Anmeldung:**

**<https://tinyurl.com/authid-mailinglist>**

