



# Relevance Ranking und Querytype Detection

VuFind Anwendertreffen

Hajo Seng

Hamburg / 16.9.2013

# ÜBERSICHT

- Wozu Relevance Ranking? Verschiedene Suchstrategien
- Technische Hintergründe
- Heuristisches Vorgehen: Metadaten clustern und Schwellwerte ermitteln
- Ergebnisse boosten
- Ausblick: Querytype Detection

# Verschiedene Suchstrategien

## EXAKTE SUCHE

ZIELGENAUE SUCHE MIT HILFE DEFINIERTER SUCHSCHLÜSSEL ABER OHNE MÖGLICHKEITEN NACHTRÄGLICHER FILTERUNG

HOCHGRADIG STRUKTURIERTE DATEN MIT BEKANNTER STRUKTUR

ZIELGENAUE SUCHEN MÖGLICH; GERINGE ABER „GUTE“ TREFFERMENGEN

SUCHSCHLÜSSEL UND DATENSTRUKTUR SOLLTEN BEKANNT SEIN

MAN WEISS, WAS MAN SUCHT, KENNT Z.B. TITEL ODER AUTOR DES GESUCHTEN MEDIUMS ODER HAT GENAUE VORSTELLUNGEN ÜBER DEN INHALT

## SUCHSTRATEGIE

### DATENBASIS

### VORTEILE

### NACHTEILE

### SINNVOLLE USECASES

## SUCHE ALS PROZESS

SUCHE MIT NACHTRÄGLICHER SUKZESSIVER FILTERUNG – AUSGEHEND VON EINER MUTMASSLICH SINNVOLLEN EINGANGSAUSWAHL

WENIG STRUKTURIERTE DATEN ODER DATEN MIT WENIG BEKANNTEN STRUKTUREN

KEINE KENNTNIS DER DATENSTRUKTUR ERFORDERLICH; „INSPIRIERENDES“ SUCHEN MÖGLICH

ZIELGENAUES SUCHEN SCHWIERIGER OFT GROSSE TREFFERMENGEN

MAN HAT NUR UNGEFÄHRE VORSTELLUNGEN UND WÜRDTE GERNE IN EINEM SUCHBEREICH „STÖBERN“, BZW. SICH EINEN ÜBERBLICK VERSCHAFFEN

# Verschiedene Suchstrategien

## „traditionelle“ bibliothekarische Suche:

(hochgradig) strukturierte Metadaten  
dazu passende Suchschlüssel

## Suche über einen solr-Index:

geeignetes Relevance Ranking  
dazu passende Facettierungen der Ergebnisse

## Konflikte:

qualifizierte Metadaten  $\leftrightarrow$  „flache“ Strukturen / möglichst Volltexte  
Intransparentes Relevance Ranking  $\leftrightarrow$  Suchschlüssel Know-how

# Technische Hintergründe: Bewertung

## Vektorraum-Modell (VSM)

Abfrage  $q$  und Dokumente  $d_j$  werden als Vektoren dargestellt:

$$q = (x_{1,q}, x_{2,q}, x_{3,q}, \dots)$$

$$d_j = (x_{1,q}, x_{2,q}, x_{3,q}, \dots)$$

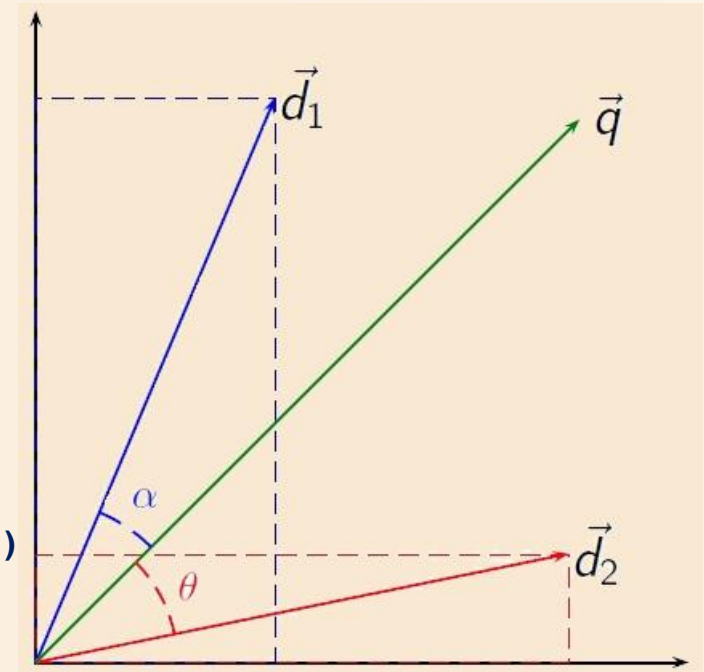
**Maß für die Ähnlichkeit:**  $\cos \alpha = d_1 \cdot q / |d_1| |q|$

tf – idf – Modell mit Ausdruck  $t$  in  $q$ :

term frequency (inkl. Feldnorm):  $tf_{t,d_i} \sim (|\{t_i \text{ in } d_j\}| / |\{t' \text{ in } d_j\}|)^{0,5}$

inverse document frequency:  $idf_{t,d_i} \sim \log(|\{\text{alle } d\}| / |\{d \text{ mit } t_i\}|)$

Vektorkoordinate:  $\omega_{t,d_i} = tf_{t,d_i} \cdot idf_{t,d_i}$



**bewerten → projizieren (Skalarprodukt) → normieren → Ähnlichkeitsmaß**

# Technische Hintergründe: Bewertung

## Beispiel „Alsterdampfer“, Rang 1, im GBV-Index der Auswahl für Beluga

**21.704868** = (MATCH) sum of:

**6.2555776** = (MATCH) max [plus 0 times others] of:

**0.62967247** = (MATCH) weight(title\_full\_unstemmed:alsterdampfer^300.0 in 1059051) [DefaultSimilarity], result of:

**0.2166948** = (MATCH) weight(allfields\_unstemmed:alsterdampfer^500.0 in 1059051) [DefaultSimilarity], result of:

**6.2555776** = (MATCH) weight(title\_short:alsterdampf^900.0 in 1059051) [DefaultSimilarity], result of:

**1.1032299** = (MATCH) weight(title:alsterdampf^450.0 in 1059051) [DefaultSimilarity], result of:

**0.08829611** = (MATCH) weight(allfields:alsterdampf^250.0 in 1059051) [DefaultSimilarity], result of:

**0.30645275** = (MATCH) weight(title\_full:alsterdampf^150.0 in 1059051) [DefaultSimilarity], result of:

**0.71473706** = (MATCH) weight(format:Book^5000.0 in 1059051) [DefaultSimilarity], result of:

**14.734553** = (MATCH) FunctionQuery(ord(publishDate)), product of:

**tie („tiebreaker“)**

**unstemmed**

**allfields**

**FunctionQuery**

# Technische Hintergründe: Bewertung

## Beispiel „Alsterdampfer“: title\_short

**6.2555776** =  $\text{score}(\text{doc}=1059051, \text{freq}=1.0 = \text{termFreq}=1.0$

), product of:

**0.36937854** = **queryWeight**, product of:

**900.0** = **boost**

**16.935411** = **idf**(**docFreq**=2, **maxDocs**=24990929)

**2.4234463E-5** = **queryNorm**

**16.935411** = **fieldWeight** in 1059051, product of:

**1.0** = **tf**(**freq**=1.0), with **freq** of:

1.0 = **termFreq**=1.0

**16.935411** = **idf**(**docFreq**=2, **maxDocs**=24990929)

**1.0** = **fieldNorm**(**doc**=1059051)

$$\text{queryWeight} = \text{boost} \cdot \text{idf} \cdot \text{queryNorm}$$

$$\text{fieldWeight} = \text{tf} \cdot \text{idf} \cdot \text{fieldNorm}$$

# Technische Hintergründe: Bewertung

## Beispiel „Alsterdampfer“: format

**0.71473706** =  $\text{score}(\text{doc}=1059051, \text{freq}=1.0 = \text{termFreq}=1.0$

), product of:

**0.29428956** = **queryWeight**, product of:

**5000.0** = **boost**

**2.4286864** = **idf**(**docFreq**=5988412, **maxDocs**=24990929)

**2.4234463E-5** = **queryNorm**

**2.4286864** = **fieldWeight** in 1059051, product of:

**1.0** = **tf**(**freq**=1.0), with **freq** of:

1.0 = **termFreq**=1.0

**2.4286864** = **idf**(**docFreq**=5988412, **maxDocs**=24990929)

**1.0** = **fieldNorm**(**doc**=1059051)



# Technische Hintergründe: Bewertung

## Beispiel „Alsterdampfer“: `ord(publishDate)`

**14.734553** = (MATCH) `FunctionQuery(ord(publishDate)`

), product of:

**608.0** = `ord(publishDate)=608`

**1000.0** = `boost`

**2.4234463E-5** = `queryNorm`

**21.704868** = (MATCH) **sum of:**

**6.2555776** = (MATCH) **max [plus 0 times others] of:**

**6.2555776** = (MATCH) `weight(title_short:alsterdampf^900.0 in 1059051)`

**0.71473706** = (MATCH) `weight(format:Book^5000.0 in 1059051)`

**14.734553** = (MATCH) `FunctionQuery(ord(publishDate))`

# Technische Hintergründe: VuFind

## Konfiguration in der searchspecs.yaml

### AllFields:

#### DismaxFields:

- title\_short^900
- title\_full\_unstemmed^300
- title\_full^150

....

#### QueryFields:

- 
- [or, 50]
- title\_short:
  - [onephrase, 900]
- title\_full\_unstemmed:
  - [onephrase, 300]
  - [and, 300]
- title\_full:
  - [onephrase, 150]

....

### Standard Query Parser:

**komplexe Abfragen (und/oder)  
präzise Abfragen**

### DisMax Query Parser:

**einfache Abfragen  
fehlertolerante Abfragen  
boosting**

### Extended DisMax Query Parser:

**umfasst und erweitert Standard und  
Dismax Query Parser**

# Technische Hintergründe: VuFind

## Konfiguration in der searchspecs.yaml

### AllFields:

....

#### DismaxParams:

```
-[bf, ord(publishDate)^1000]  
-[bq, format:Book^5000]
```

....

### QueryFields:

....

```
FilterQuery: "format:Book"
```

....

### Standard Query Parser:

```
q=((title_short:"Alsterdampfer")^900 OR  
  title_full_unstemmed:"Alsterdampfer")^300 OR ...)  
qf=((format:Book))
```

### (E-)DisMax Query Parser:

```
qt=dismax  
q=Alsterdampfer  
qf=title_short^900 title_full_unstemmed^300 ...  
bq=format:Book^5000  
bf=ord(publishDate)^1000
```

```
q=({_query_:""!dismax qf="title_short^900  
  title_full_unstemmed^300 ... bq=format:Book^5000  
  bf=ord(publishDate)^1000"}Alsterampfer")
```

# Heuristisches Vorgehen: Metadaten

## Felder-Analyse, Beispiel „Alsterdampfer : die weisse Flotte am Jungfernstieg“ → Titel und Autorendaten

### solr Index

title\_short: **Alsterdampfer :**  
title: **Alsterdampfer : die weisse Flotte am Jungfernstieg /**  
title\_full(\_unstemmed): **Alsterdampfer : die weisse Flotte am Jungfernstieg / Hans H. Müller**  
author: **Müller, Hans H.**  
abstract(\_unstemmed): **Frühere Ausg. u.d.T.: Müller, Hans H.: Chronik der weißen Flotte**

### marc 21

100|a: **Müller, Hans H.**  
245|a: **Alsterdampfer :**  
245|b: **die weisse Flotte am Jungfernstieg /**  
245|c: **Hans H. Müller**  
500|a: **Frühere Ausg. u.d.T.: Müller, Hans H.: Chronik der weißen Flotte**

### pica

4000: **@Alsterdampfer : die weisse Flotte am Jungfernstieg / Hans H. Müller**  
4201: **Frühere Ausg. u.d.T.: Müller, Hans H.: Chronik der weißen Flotte**

**allfields > allfields\_unstemmed > title\_full > title\_full\_unstemmed > title > title\_short**

**allfields > allfields\_unstemmed > title\_full > title\_full\_unstemmed > author**

# Heuristisches Vorgehen: Metadaten

## Felder-Analyse, Beispiel „Fliegende Kühe über‘m Alsterdampfer“ → Titel und Autorendaten

### solr Index

title\_short: Fliegende Kühe über‘m  
Alsterdampfer :  
title\_sub: Heiteres und Ernstes von der  
HHA /  
title: Fliegende Kühe über‘m Alsterdampfer  
: Heiteres und Ernstes von der HHA /  
title\_full(\_unstemmed): Fliegende Kühe  
über‘m Alsterdampfer : Heiteres und  
Ernstes von der HHA / Wolfgang Plat. Mit  
Zeichnungen von Jürgen Pieplow  
author: Plat, Wolfgang  
author2: Pieplow, Jürgen

### marc 21

100|a: Plat, Wolfgang  
245|a: Fliegende Kühe über‘m  
Alsterdampfer :  
245|b: Heiteres und Ernstes von der HHA /  
245|c: Wolfgang Plat. Mit Zeichnungen von  
Jürgen Pieplow  
700|a: Pieplow, Jürgen

### pica

3000: Plat, Wolfgang![627788998](#)!Plat,  
Wolfgang \*1923-1995\* ; ID: gnd/141434376  
3010: Pieplow, Jürgen![636297608](#)!Pieplow,  
Jürgen \*1935-\* ; ID: gnd/14247746X  
4000: Fliegende Kühe über‘m  
Alsterdampfer : Heiteres und Ernstes von  
der HHA / Wolfgang Plat. Mit  
Zeichnungen von Jürgen Pieplow

[allfields](#) > [allfields\\_unstemmed](#) > [title\\_full](#) > [title\\_full\\_unstemmed](#) > [title](#) > [title\\_short](#) | [title\\_sub](#)

[allfields](#) > [allfields\\_unstemmed](#) > [title\\_full](#) > [title\\_full\\_unstemmed](#) > [author](#) | [author2](#)

# Heuristisches Vorgehen: Metadaten

## Felder-Analyse, Beispiel „Alsterdampfer : die weisse Flotte am Jungfernstieg“

### → Schlagwörter

#### solr Index

topic: Firmen-Festschrift / Hamburg /  
Schiffahrt / Alster: Schiffahrt / Alster /  
Fahrgastschiffahrt / Binnenschiffahrt /  
Alster / Fahrgastschiffahrt /  
Binnenschiffahrt  
topic\_unstemmed: Firmen-Festschrift /  
Hamburg / Schiffahrt / Alster: Schiffahrt /  
Alster / Fahrgastschiffahrt /  
Binnenschiffahrt / Alster /  
Fahrgastschiffahrt / Binnenschiffahrt  
topic\_title: ....

#### marc 21

650|a: Fahrgastschiffahrt  
650|a: Binnenschiffahrt  
651|a: Alster  
653|a: Alster River (Germany) ...  
...  
689|a: Alster  
689|a: Fahrgastschiffahrt  
689|a: Binnenschiffahrt  
982|a: Firmen-Festschrift  
982|a: Hamburg  
982|a: Schiffahrt  
982|a: Alster: Schiffahrt  
...

#### pica

5500: |a| |a|Inland water transportation  
|a|Germany |a|Alster River  
5500: |a| |a|Alster River (Germany)  
|a|Navigation  
5500: |a| |a|Hamburg (Germany)  
|a|Description and travel  
5550: [!106397370!](#)Alster ; ID: gnd/4001368-6  
5550: [!105211931!](#)Fahrgastschiffahrt ; ID:  
gnd/4195432-4  
5550: [!106376365!](#)Binnenschiffahrt ; ID:  
gnd/4006761-0  
6800: [!051612755!](#)Alster: Schiffahrt

# Heuristisches Vorgehen: Metadaten

## Felder-Analyse, Beispiel „Alsterdampfer : die weisse Flotte am Jungfernstieg“

### → Klassifikationen

#### solr Index

class: **Alsterdampfer / Bibliothek Gerhard Ahrens / Alster / Alsterschiffahrt / Binnenschiffahrt / Buch / Neuzeit**  
class\_local: **Alsterdampfer / Bibliothek Gerhard Ahrens / Alster / Alsterschiffahrt / Binnenschiffahrt / Buch / Neuzeit**

#### marc 21

983|a: **Alsterdampfer**  
983|a: **Bibliothek Gerhard Ahrens**  
983|b: **Alster**  
983|b: **Alsterschiffahrt**  
983|b: **Binnenschiffahrt**  
983|b: **Buch**  
983|b: **Neuzeit**

#### pica

6700: **Bibliothek Gerhard Ahrens**

# Heuristisches Vorgehen: Metadaten

## Felder-Analyse, Beispiel „Fliegende Kühe über‘m Alsterdampfer“

### → Schlagwörter und Klassifikationen

#### solr Index

topic(\_unstemmed): **Texte eines  
einzelnen Autors / Deutsche Literatur /  
Hamburger Hochbahn Aktiengesellschaft  
/ Erzählung / Hamburger Hochbahn-  
Aktiengesellschaft / Erzählung**  
class: **Texte eines einzelnen Autors /  
Deutsche Literatur / Buch**  
bklname: **Texte eines einzelnen Autors /  
Deutsche Literatur**

#### marc 21

084|9: **Texte eines einzelnen Autors**  
084|9: **Deutsche Literatur**  
653|a: **Anthologie**  
689|a: **Hamburger Hochbahn-  
Aktiengesellschaft**  
689|a: **Erzählung**  
689|a: **Anthologie**  
983|b: **Buch**

#### pica

5301: [!181571714!](#)17.97\$j**Texte eines  
einzelnen Autors**  
5302: [!106405047!](#)18.10\$j**Deutsche  
Literatur**  
5550: [!101781830!](#)**Hamburger Hochbahn-  
Aktiengesellschaft ; ID: gnd/2013043-0**  
5550: [!106336673!](#)**Erzählung ; ID:  
gnd/4015464-6**  
5550: |f| **Anthologie**



# Heuristisches Vorgehen: Clustern

## Gewichtung und Schwellwerte

### Metadatencluster

#### Titelcluster

title\_short, title\_sub < title < title\_full(\_unstemmed) (marc 245)

title\_alt (marc 130, 240, 246, 730, 740), title\_new (marc 247), title\_old (marc 780)

#### Schlagwortcluster

class (marc 084, 983), class\_local (marc 983) < topic(\_unstemmed) (marc 650, 653, 689, 982, 983)

bklname (BK) (marc 084), dewey-xxx (DDC) (marc 082, 083) < topic(\_unstemmed)

#### Autorencluster

author (marc 100, 110) < title\_full(\_unstemmed)

author2 (marc 700, 710)

#### Weitere (Serien, Bandtitel etc.)

journal (marc 773), series (marc 490, 830), series2 (marc 490), ...

# Heuristisches Vorgehen: Clustern

## Gewichtung und Schwellwerte

interne Gewichtung der Cluster / Beispiel „Alsterdampfer ...“

abhängige Felder:

**Feldinhalte:** title\_short < title < title\_full\_unstemmed < title\_full < all\_fields\_unstemmed < all\_fields

**Feldnorm (fn):** title\_short > title > title\_full\_unstemmed > title\_full > all\_fields\_unstemmed > all\_fields

1.0 > 0.375 > 0.3125 > 0.3125 > 0.047 > 0.039

**idf:** title\_short > title > title\_full\_unstemmed > title\_full > all\_fields\_unstemmed > all\_fields

16.935 > 16.425 > 16.648 > 16.425 > 16.425 > 16.242

**df:** 2 < 4 < 3 < 4 < 4 < 5

**Rohwert:** 286,8 < 101,2 < 86,6 < 84,3 < 12,7 < 10,2

(fn · idf<sup>2</sup>)

**Fragen:** Bedeutung der Felder in Hinblick auf die Relevanz des Gesamtdatensatzes (Z.B.: Hauptansetzung)  
Wie homogen sind die Feldinhalte (z.B. title\_alt)?  
Tiebreaker setzen (und damit gering bewertete Felder mitspielen lassen)?

# Heuristisches Vorgehen: Clustern

## Gewichtung und Schwellwerte

### externe Gewichtung der Cluster

#### Cluster:

Titelcluster

Schlagwortcluster

Autorencluster

Weitere Cluster

#### Kriterien: Datenqualität:

Sind die Felder immer gefüllt (z.B. DDC, BK)?

Wie gut sind die Feldinhalte qualifiziert (z.B. Schlagwörter)?

Wie homogen sind die Cluster (z.B. „Weitere Cluster“)?

#### Medientypen:

Unterschiedliche Darstellung unterschiedlicher Medien

(z.B. E-Zeitschriften ↔ Monographien; Artikel ↔ Zeitschriften)

# Heuristisches Vorgehen: weitere Aspekte

## Beispiel Schlagwörter

Schlagwort „Fahrgastschiffahrt“ (Auszug aus der WinIBW):

**035** gnd/4195432-4

**150** Fahrgastschiffahrt

**450** Passagierschiffahrt

**450** Personenschiffahrt

**450** Passagierschiffahrt

**450** Personenschiffahrt

**550** [!105807109!](#)Schiffahrt ; ID: gnd/4116383-7\$4obal

**550** [!106202308!](#)Personenverkehr ; ID: gnd/4045298-0\$4obal

**035** gnd/4116383-7

**150** Schiffahrt

**450** Schiffsverkehr

**450** Schiffahrt

# Heuristisches Vorgehen: weitere Aspekte

## Beispiel Zeitschriften

### Suche nach „Journal of coastal research“:

Journal of coastal research  
Zeitschrift, 1985

Rang 1

27.381084 = sum of:

7.3857694 = max of:

7.3857694 = weight(title\_short:  
"journal of coastal research"^900.0

3.0201528 = weight(format:Journal^4000.0

16.975163 = FunctionQuery(ord(publishDate)),  
product of:

617.0 = ord(publishDate)=617

1000.0 = boost

Journal of coastal research  
E-Zeitschrift, 1985

Rang 2

27.02681 = sum of:

7.400678 = max of:

7.400678 = weight(title\_short:  
"journal of coastal research"^900.0

2.9625764 = weight(format:eJournal^2000.0

16.663555 = FunctionQuery(ord(publishDate)),  
product of:

605.0 = ord(publishDate)=605

1000.0 = boost

Journal of coastal research  
Serial Volume 25.2009, 2009

Rang 4

24.259615 = sum of:

6.5303187 = max of:

6.5303187 = weight(series2:  
"journal of coastal research"^450.0

17.729296 = FunctionQuery(ord(publishDate)),  
product of:

646.0 = ord(publishDate)=646

1000.0 = boost

# Heuristisches Vorgehen: weitere Aspekte

## Beispiel Zeitschriften

Journal of coastal research; Zeitschrift

$7.3857694 = \text{weight}(\text{title\_short}:"\text{journal of coastal research}"^{\wedge}900.0$

$7.3857694 = \text{score}(\text{doc}=270083, \text{freq}=1.0 = \text{phraseFreq}=1.0$

), product of:

$0.6047815 = \text{queryWeight}$ , product of:

$900.0 = \text{boost}$

**24.424587** = idf(), sum of:

$7.0908265 = \text{idf}(\text{docFreq}=55935, \text{maxDocs}=24711766)$

$2.3762057 = \text{idf}(\text{docFreq}=6240583, \text{maxDocs}=24711766)$

$8.893877 = \text{idf}(\text{docFreq}=9217, \text{maxDocs}=24711766)$

$6.0636797 = \text{idf}(\text{docFreq}=156233, \text{maxDocs}=24711766)$

$2.7512418\text{E-}5 = \text{queryNorm}$

$12.212294 = \text{fieldWeight}$  in 270083, product of:

$1.0 = \text{tf}(\text{freq}=1.0)$ , with freq of:

$1.0 = \text{phraseFreq}=1.0$

**24.424587** = idf(), sum of:

$7.0908265 = \text{idf}(\text{docFreq}=55935, \text{maxDocs}=24711766)$

$2.3762057 = \text{idf}(\text{docFreq}=6240583, \text{maxDocs}=24711766)$

$8.893877 = \text{idf}(\text{docFreq}=9217, \text{maxDocs}=24711766)$

$6.0636797 = \text{idf}(\text{docFreq}=156233, \text{maxDocs}=24711766)$

$0.5 = \text{fieldNorm}(\text{doc}=270083)$

Journal of coastal research; Serial Volume 25.2009

$6.5303187 = \text{weight}(\text{series2}:"\text{journal of coastal research}"^{\wedge}450.0$

$6.5303187 = \text{score}(\text{doc}=565096, \text{freq}=1.0 = \text{phraseFreq}=1.0$

), product of:

$0.4016224 = \text{queryWeight}$ , product of:

$450.0 = \text{boost}$

**32.51969** = idf(), sum of:

$7.7017 = \text{idf}(\text{docFreq}=30486, \text{maxDocs}=24809963)$

$5.8544884 = \text{idf}(\text{docFreq}=193351, \text{maxDocs}=24809963)$

$11.915289 = \text{idf}(\text{docFreq}=450, \text{maxDocs}=24809963)$

$7.048215 = \text{idf}(\text{docFreq}=58602, \text{maxDocs}=24809963)$

$2.7444732\text{E-}5 = \text{queryNorm}$

$16.259846 = \text{fieldWeight}$  in 565096, product of:

$1.0 = \text{tf}(\text{freq}=1.0)$ , with freq of:

$1.0 = \text{phraseFreq}=1.0$

**32.51969** = idf(), sum of:

$7.7017 = \text{idf}(\text{docFreq}=30486, \text{maxDocs}=24809963)$

$5.8544884 = \text{idf}(\text{docFreq}=193351, \text{maxDocs}=24809963)$

$11.915289 = \text{idf}(\text{docFreq}=450, \text{maxDocs}=24809963)$

$7.048215 = \text{idf}(\text{docFreq}=58602, \text{maxDocs}=24809963)$

$0.5 = \text{fieldNorm}(\text{doc}=565096)$

# Heuristisches Vorgehen: Schwellwerte

## Gewichtung und Schwellwerte

### Schwellwertermittlung / Beispiel „Alsterdampfer“ Plätze 2 und 3

#### Gewichtungs-Parameter: `title_short^300`

Die Hamburger Alsterdampfer : hrsg. zum 125jährigen  
Geburtstag der Hamburger Alsterflotte / Wolfgang Reimer

17.047792 = sum of:

1.0379449 = max of:

1.0379449 = `weight(title_short:alsterdampf^300.0)`

0.7577349 = `weight(title:alsterdampf^450.0)`

0.73751765 = `weight(format:Book^5000.0)`

15.272329 = `FunctionQuery(ord(publishDate))`

#### Rang 2

Alsterschiffahrt : die "Weiße Flotte" im Herzen Hamburgs /  
Christian Müller

16.951725 = sum of:

0.49269205 = max of:

0.49269205 = `weight(topic_unstemmed:alsterdampfer^700.0)`

0.73751765 = `weight(format:Book^5000.0)`

15.721516 = `FunctionQuery(ord(publishDate))`

#### Rang 3

# Heuristisches Vorgehen: Schwellwerte

## Gewichtung und Schwellwerte

### Schwellwertermittlung / Beispiel „Alsterdampfer“ Plätze 2 und 3

Gewichtungs-Parameter: `title_short^200`

Die Hamburger Alsterdampfer : hrsg. zum 125jährigen  
Geburtstag der Hamburger Alsterflotte / Wolfgang Reimer

16.767582 = sum of:

0.7577349 = max of:

0.69196326 = `weight(title_short:alsterdampf^200.0`

0.7577349 = `weight(title:alsterdampf^450.0`

0.73751765 = `weight(format:Book^5000.0`

15.272329 = `FunctionQuery(ord(publishDate))`

**Rang 8**

Alsterschiffahrt : die "Weiße Flotte" im Herzen Hamburgs /  
Christian Müller

16.951725 = sum of:

0.49269205 = max of:

0.49269205 = `weight(topic_unstemmed:alsterdampfer^700.0`

0.73751765 = `weight(format:Book^5000.0`

15.721516 = `FunctionQuery(ord(publishDate))`

**Rang 3**



# Heuristisches Vorgehen: Boosting

## Globale Steuerungen des Rankings (Auswahl):

**Boosten einzelner Feldinhalte via Boosting Queries**

**Beispiel: Angleichung unterschiedlicher Medientypen**

**Boosten einzelner Feldinhalte via Boosting Functions**

**Beispiel: Aufwertung aktuellerer Veröffentlichungen**

**Höhere Gewichtung des allfields Feldes**

**Beispiel: Ausgleich unregelmäßig besetzter Felder**

**Einsetzen eines Tiebreakers**

**Beispiel: Aufwertung ansonsten nicht beachteter Felder nach Suchfeldauswahl**

**etc.**

# Heuristisches Vorgehen: Boosting

## Globale Steuerungen des Rankings (Auswahl):

### Boosten einzelner Feldinhalte via Boosting Queries

**Beispiel: Angleichung unterschiedlicher Medientypen (format Feld)**  
(Query: Journal Crystal Growth)

Journal of crystal growth: Zeitschrift

format boosting: 4000

23.336092 = (MATCH) sum of:  
0.94628406 = (MATCH) sum of:  
0.33666772 = (MATCH) max of:  
0.33553445 = (MATCH) max of:  
0.27408192 = (MATCH) max of:  
3.481653 = (MATCH) weight(format:Journal^4000.0  
18.908155 = (MATCH) FunctionQuery(ord(publishDate))

Rang 2

Journal of crystal growth: E-Zeitschrift

format boosting: 2000

23.21532 = (MATCH) sum of:  
0.94628406 = (MATCH) sum of:  
0.33666772 = (MATCH) max of:  
0.33553445 = (MATCH) max of:  
0.27408192 = (MATCH) max of:  
3.3608801 = (MATCH) weight(format:eJournal^2000.0  
18.908155 = (MATCH) FunctionQuery(ord(publishDate))

Rang 3

# Heuristisches Vorgehen: Boosting

## Globale Steuerungen des Rankings (Auswahl):

Boosten einzelner Feldinhalte via Boosting Functions

Beispiel: Aufwertung aktueller Medien (format Feld)

(Query: Journal Crystal Growth)

Crystal growth & design: E-Zeitschrift (2001)

format boosting: 4000

23.440609 = (MATCH) sum of:

0.6126873 = (MATCH) sum of:

**0.0022138513** = (MATCH) max of: (journal)

0.3362606 = (MATCH) max of: (crystal)

0.27421284 = (MATCH) max of: (growth)

3.3815858 = (MATCH) weight(format:Journal^4000.0

**19.446335** = (MATCH) FunctionQuery(ord(publishDate))

Rang 1

Journal of crystal growth: E-Zeitschrift (1967)

format boosting: 2000

23.21532 = (MATCH) sum of:

0.94628406 = (MATCH) sum of:

**0.33666772** = (MATCH) max of: (journal)

0.33553445 = (MATCH) max of: (crystal)

0.27408192 = (MATCH) max of: (growth)

3.3608801 = (MATCH) weight(format:eJournal^2000.0

**18.908155** = (MATCH) FunctionQuery(ord(publishDate))

Rang 3

# Ausblick: Querytype Detection

Eine grundsätzliche Schwierigkeit ergibt sich dadurch, dass der Versuch unternommen werden soll, eine Ranking-Konfiguration für alle denkbaren Suchen / Suchstrategien zu finden. Besser wäre es, Suchstrategien zu erkennen und das Relevance-Ranking entsprechend anzupassen.

- known item search ↔ thematische Suche
- Suche in Feldern, die ansonsten eher unterbewertet sind (z.B. „weitere Autoren“)
- Suche nach übergeordneten Einheiten (Reihen, Zeitschriften)  
↔ Suche nach deren Inhalten (Bände, Artikel)
- Spezielle Suchen (ppn, Signatur)

# Vielen Dank

**Hajo Seng**

Von-Melle-Park 3  
20146 Hamburg

040 / 4 28 38-8336  
hajo.seng@sub.uni-hamburg.de

[www.sub.uni-hamburg.de](http://www.sub.uni-hamburg.de)

 [facebook.com/stabihh](https://facebook.com/stabihh)

 [twitter.com/stabihh](https://twitter.com/stabihh)