

Elasticsearch-Konkordanz von Katalogs-IDs zu RVK auf Grundlage der Culturegraph Aggregatdaten

Pascal Christoph (dr0i)

Offene Infrastruktur, Hochschulbibliothekszentrum NRW (hbz)



KIM-Workshop, WWW, 05.05.2020

Diese Präsentation:

<http://slides.lobid.org/kim-ws-2020-lightning-talk/>



Was ist RVK?

Was ist RVK?

Die Regensburger Verbundklassifikation (RVK, auch Regensburger Systematik) ist eine Klassifikation

Was ist RVK?

Die Regensburger Verbundklassifikation (RVK, auch Regensburger Systematik) ist eine Klassifikation

Gegliedert in Hauptgruppen und thematische Klassen, die durch eine Notation identifiziert sind (ähnlich Dewey Decimal Classification)

Was ist RVK?

Die Regensburger Verbundklassifikation (RVK, auch Regensburger Systematik) ist eine Klassifikation

Gegliedert in Hauptgruppen und thematische Klassen, die durch eine Notation identifiziert sind (ähnlich Dewey Decimal Classification)

Z.B.: "CI 5310" für "Abhandlungen im Kontext französischsprachiger Literatur der Philosophie"

Was ist RVK?

Die Regensburger Verbundklassifikation (RVK, auch Regensburger Systematik) ist eine Klassifikation

Gegliedert in Hauptgruppen und thematische Klassen, die durch eine Notation identifiziert sind (ähnlich Dewey Decimal Classification)

Z.B.: "CI 5310" für "Abhandlungen im Kontext französischsprachiger Literatur der Philosophie"

CC0 lizenziert

Wozu RVK Anreicherung ?

Inhalterschließung, z. B. virtuelle Aufstellungssystematik

The screenshot shows the website of the University of Dortmund library. At the top left is the logo 'tu technische universität dortmund'. To its right is the text 'Universitätsbibliothek'. Below this are two tabs: 'Search' and 'Classification', with 'Classification' being the active tab. A large green banner below the tabs contains the text 'Search in classification'. Below the banner, there are two sub-tabs: 'Classification' and 'Search in classification', with 'Classification' being the active tab. A help icon (?) is also present. The main content area is titled 'Classification' and lists three categories with their respective result counts:

- > Allgemeines (RVK: A): 50986 Results
- > Theologie und Religionswissenschaften (RVK: B): 50178 Results
- ▼ Philosophie (RVK: CA - CK): 37862 Results
 - Zeitschriften und Reihenwerke (RVK: CA 1000): 255 Results

Wie

Wie

Culturegraph stellt geclusterte Bestände der Verbünde (u.a. mit RVK) bereit

Wie

Culturegraph stellt geclusterte Bestände der Verbünde (u.a. mit RVK) bereit

Dieses MARC21-XML filtern mit dieser Metafactory-Morph:

Wie

Culturegraph stellt geclusterte Bestände der Verbünde (u.a. mit RVK) bereit

Dieses MARC21-XML filtern mit dieser Metafactory-Morph:

```
<combine name="@rvk" value="{a}" sameEntity="true" >
  <data source="084??.a" name="a"/>
  <data source="084??.2">
    <regexp match="rvk" />
  </data>
</combine>
<combine name="@id" value="{a}">
  <data source="035??.a" name="a">
    <regexp match="^(DE-605\)(.*)" format="{1}"/>
  </data>
</combine>
<combine name="rvk" value="{a}" >
  <data source="@id"/>
  <data source="@rvk" name="a"/>
</combine>
<combine name="id" value="{a}">
  <concat delimiter=", " name="a" >
    <data source="@id"/>
  </concat>
  <data source="@rvk" />
</combine>
```

Wie

Wie

=> der Output als Elasticsearch JSON bulk:

Wie

=> der Output als Elasticsearch JSON bulk:

```
{"index":{"_index":"cgrvk","_type":"rvk"}}  
{"rvk":["CI 1125","CI 5604","CI 5603","IH 34381","CI 1100","CI 5  
"EC 2430"],"id":["HT013166356, HT018625006, TT000577460"]}  
...
```

Wie

=> der Output als Elasticsearch JSON bulk:

```
{"index":{"_index":"cgrvk","_type":"rvk"}}  
{"rvk":["CI 1125","CI 5604","CI 5603","IH 34381","CI 1100","CI 5  
"EC 2430"],"id":["HT013166356, HT018625006, TT000577460"]}  
...
```

... das per curl nach Elasticsearch indexieren

Wie

=> der Output als Elasticsearch JSON bulk:

```
{"index":{"_index":"cgrvk","_type":"rvk"}}  
{"rvk":["CI 1125","CI 5604","CI 5603","IH 34381","CI 1100","CI 5  
"EC 2430"],"id":["HT013166356, HT018625006, TT000577460"]}  
...
```

... das per curl nach Elasticsearch indexieren

... Elasticsearch verwenden als Lookuptabelle

Links

- Das Tool auf github
- Aggregatdatei (MARC21-XML, 6.1 GB)
- Die RVK
- Hans-Georg Becker: Mit CultureGraph zu einer virtuellen Systematik