



GÖTEBORGS
UNIVERSITET

GÖTEBORGS UNIVERSITETSBIBLIOTEK

ELASTICSEARCH OCH KOHA

- En utvecklingsrapport

Vem är jag?

Stefan Berndtsson

IT-arkitekt och utvecklare

Jobbat på universitetsbiblioteket i evigheter

Sysslar med allt IT-relaterat och petar i allting

Vilka är vi?

Avdelningen Digitala tjänster är indelad i fem team. E-resurser, Digitalisering, Publicering och bibliometri, Webb samt Utveckling.

Vi ansvarar för drift, support och utveckling av vår digitala verksamhet, där bl a Koha ingår.

Ca 1,6 miljoner bibposter, 1,6 miljoner exemplar och 40000 aktiva låntagare.



Varför inte Zebra?

- Ger obegripliga resultat med obegriplig relevansrankning
- Långsam
- Fasetter fungerar inte
- Saknar gränssnitt för konfiguration
- Konfigurationen lagras i systemfiler istället för i databasen

Hur använder vi Elasticsearch?

- Sökning för personal
- Matchning vid Libris-import
- Presentation av lånestatus i Libris webbsök
- Blandade datauttag som är tunga att söka i databasen



Elasticsearch för vår personal

Ange nyckelord för sökning:

[Låna ut](#)
[Återlämna](#)
[Förnya](#)
[Sök katalogen](#)

Hem > Katalog > Sök efter 'grunder' med begränsning(ar): 'itype:2'

Förfin din sökning

Tillgänglighet
Visa bara tillgängliga exemplar

Författare
[Edberg, Anna-Karin, \(5\)](#)
[Castoriano, Martine, \(4\)](#)
[Bohlin, Alf, \(3\)](#)
[Davidson, Bo, \(2\)](#)
[Ehrenberg, Anna, \(2\)](#)
[Visa mer](#)

Format
[Book \(52\)](#)

Exemplartyper
[Kursbok \[x\]](#)
[Monografi \(44\)](#)
[REF \(41\)](#)

Placering
[Ekon bibl KURS \(20\)](#)
[Sambibl KURS \(15\)](#)
[Sambibl KURS LÄSESAL \(15\)](#)
[Biomed bibl KURS \(14\)](#)
[Ekon bibl KURS REF \(14\)](#)
[Visa mer](#)

52 resultat hittade för 'grunder' med begränsningar: 'itype:2'.

✓ Välj alla | ✗ Rensa alla | Ta bort markering |

1
2
3
Nästa >>
Sista

Resultat	Plats
<input type="checkbox"/> 1. Logistikens grunder / Kenth Lumsden av Lumsden, Kenth, 1944-. Materialtyp: Bok; Format: print ; Litterär form: Ej fiktion Förläggare: Lund : Studentlitteratur, 2006 Beskrivning: 767 s. : ill. ; 25 cm. ISBN: 9144028733; 9789144028736 (2. tr.). Upplaga: 2., [utök. och uppdaterade] uppl. <div style="text-align: right; font-size: small;"> Reservationer (0) Redigera post Redigera exemplar </div>	4 exemplar, 4 tillgänglig: <ul style="list-style-type: none"> Ekonomiska biblioteket Ekon bibl KURS [Kursbok] (4) Kursbok
<input type="checkbox"/> 2. Logistikens grunder / Kent Lumsden av Lumsden, Kenth, 1944-. Materialtyp: Bok; Format: print ; Litterär form: Ej fiktion Förläggare: Lund : Studentlitteratur, 2012 Beskrivning: 816 s. : ill. ; 25 cm. ISBN: 9789144081298. Upplaga: 3., [utök. och uppdaterade] uppl. <div style="text-align: right; font-size: small;"> Reservationer (0) Redigera post Redigera exemplar </div>	7 exemplar, 3 tillgänglig: <ul style="list-style-type: none"> Ekonomiska biblioteket Ekon bibl KURS [Kursbok] (2) Kursbok Ekonomiska biblioteket Ekon bibl KURS REF [Info & Lån] (1) Kursbok 4 utlånad: <ul style="list-style-type: none"> Samhällsvetenskapliga biblioteket Ekon bibl BALKONG 1 [658.5 12-19] (1) återlämningsdatum: 2018-10-22

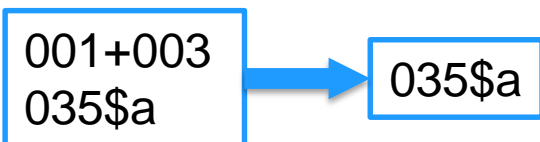
Ett gediget sökspråk

Elasticsearch har ett väl utbyggt och modernt sökspråk

- Phrase: “handbok för cirkelledare”
- Fuzzy: dsylexi~1
- Range: copydate:[1984 TO 1989]
- Boolean: (quick OR brown) AND fox

Librisimport

Matchning av poster
sker via Elasticsearch



Förbereda MARC-poster för import

- Välj en MARC-fil att förbereda i importreservoaren. Den kommer att läsas och varje giltig post kommer att förberedas för senare import till katalogen.
- Du kan ange ett namn för denna import. Det kan vara användbart att komma ihåg varifrån det föreslagna MARC-datat kommer när en post skapas!

Förbered poster i reservoaren

Välj fil att förbereda: 180529.mrc

Uppladdningsförlopp: 100%

Använd mall för MARC-modifiering:

Modifiera post med följande mall: ▾

[Visa avancerade inställningar](#)

Förbered för import

+ Förbered MARC-poster

☰ Hantera förberedda poster

Förbereda MARC-poster för import

MARC förbereda resultat :

- Bearbetar bibliografiska poster
- 1 poster i fil
- 0 poster inte förberedda på grund av MARC-fel
- 1 poster förberedda
- 1 poster med minst en träff i katalog per matchningsregel "KohaBiblio"
- 0 exemplarsposter hittade och förberedda

Hantera förberedda MARC-poster › Sats 14331

Filnamn: 180529.mrc

User:

Kommentarer: (ingen)

Typ: Bibliografiska poster

Förberedda: 2018-10-13 16:36:46

Status: Förberedd

Lägg till nya bibliografiska poster i detta ramverk: ▾

Visar 1 till 1 av 1 Visa ▾ -objekt

#	Citat	Status	Matchtyp	Matchningsdetaljer
25338	Svart på vitt :	Förberedda	Matchning hittad	Matches bibliographic record 2598983 (score=1000):Svart på vitt :

Visa ▾ -objekt

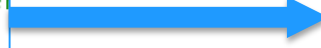


Lånestatus i Libris

```

{
  _index: "koha_koha_biblios",
  _type: "data",
  _id: "1771091",
  _version: 10,
  found: true,
  - _source: {
    - provider: [
      "Ungern",
      "Lund :",
      "Studentlitteratur,",
      "2011"
    ],
    - system-control-number__sort: [
      "(LIBRIS)12180090"
    ],
    - title: [
      "Forskningsmetodikens grunder :",
      "att planera, genomföra och rapportera",
      "Runa Patel, Bo Davidson."
    ],
    - author: [
      "Patel, Runa,",
      "1945-",
      "Runa Patel, Bo Davidson.",
      "Davidson, Bo,"
    ]
  }
}

```



forskningsmetodikens grunder [Mina bibliotek](#)

Sökning: [forskningsmetodikens grunder](#) > Forskningsmetodiken...

6 av 7 [Föregående post](#) | [Nästa post](#) | [Till träfflistan](#)

Översikt

Forskningsmetodikens grunder : att planera, genomföra och rapportera en undersökning / Runa Patel, Bo Davidson.

Patel, Runa, 1945- (författare)
Davidson, Bo, 1956- (författare)
 ISBN 9789144068688
 4., [uppdaterade] uppl.
 Publicerad : Lund : Studentlitteratur, 2011
 Tillverkad : Ungern
 Svenska 149 s.

Göteborgs universitetsbibliotek Humanistiska biblioteket (G) [ANGE SOM FAVORIT](#)

[Titeln i bibliotekets lokala katalog](#) [Find@GU](#)

Placering: Kursbok
 Placering: ÖM D 11/274

▼ **Utlånad?** ► **Öppettider, adress m.m.**

#	Placering	Hylla	Lånepolitik	Status	Datum
1.	Biomedicinska biblioteket	Kursbok	Ej fjärrlån	Utlånad	Åter:2018-10-23
2.	Biomedicinska biblioteket	Kursbok	Ej fjärrlån	Utlånad	Åter:2018-10-22
3.	Biomedicinska biblioteket	Kursbok	Ej fjärrlån	Utlånad	Åter:2018-10-17
4.	Biomedicinska biblioteket	Kursbok	Ej fjärrlån	Utlånad	Åter:2018-10-22
5.	Biomedicinska biblioteket	Kursbok	Ej fjärrlån	Tillgänglig	
6.	Biomedicinska biblioteket	Kursbok	Utlånas ej	Tillgänglig	
7.	Biomedicinska biblioteket	Kursbok	Utlånas ej	Tillgänglig	
8.	Ekonomiska biblioteket	Kursbok	Ej fjärrlån	Tillgänglig	
9.	Ekonomiska biblioteket	Kursbok	Ej fjärrlån	Utlånad	Åter:2018-10-18
10.	Ekonomiska biblioteket	Kursbok	Ej fjärrlån	Utlånad	Åter:2018-10-19
11.	Ekonomiska biblioteket	Kursbok	Ej fjärrlån	Tillgänglig	
12.	Ekonomiska biblioteket	Kursbok	Ej fjärrlån	Utlånad	Åter:2018-10-18
13.	Ekonomiska biblioteket	Kursbok	Ej fjärrlån	Utlånad	Åter:2018-10-25
14.	Ekonomiska biblioteket	Kursbok	Ej fjärrlån	Utlånad	Åter:2018-10-26
15.	Ekonomiska biblioteket	Kursbok	Ej fjärrlån	Tillgänglig	

Åtgärder för att få Elasticsearch att fungera

- Sökresultat och rankning
- Fasetter
- Paginering
- Zebra-fält och Elasticsearch-fält
- Indexeringstid
- Sammanslagning av fältdata vid uppgraderingar



Sökning, fasetter och paginering

- Sökning som innehåller “/” ger ett fel, rättas av [bugg 20334](#).
- Fasetter och paginering sker helt via URL-fältet, och det är därför viktigt att detta genereras rätt mellan sidladdningarna. [Bugg 20114](#) åtgärdar ett antal fel som annars orsakar sökfel.
- Valt sorteringsfält hanteras fel om fältnamnet innehåller “_”, vilket redan rättats via [bugg 19564](#). Denna är accepterad av Koha.



Zebra-fält och Elasticsearch-fält

Zebra-fältnamnen används lite här och var i Koha. Dessa namn återfinns inte alltid med samma namn i Elasticsearch.

Bugg 19575 och ett par till gör så att alla Zebra-fält återfinns som Elasticsearch-fält eller fält-alias och går att söka.

- Fixar...
 - ...avancerad sökning
 - ...länkning till sökning (t ex ämnesordslänkar och liknande)
 - ...vissa kombinationssökningar fungerar (**Bugg 19197**)
 - ...sökning på årtalsintervall med formen YYYY-YYYY (**Bugg 20492**)



Indexeringstid och Catmandu

Catmandu är en komponent som konverterar mellan MARC-data och Elasticsearch-indexformat (JSON).

Bugg 19893 är vår andra stora fix som tillsammans med **bugg 19884** snabbar upp indexeringstiden avsevärt.

- Minskar vår indexeringstid (1.6M poster) från ca 24tim till ca 4tim
- Tar bort 2-3s uppstartstid Catmandu har som skapar fördröjning vid utlån/återlämning
- Sparar MARC-posten i MARC21 eller MARCXML i indexet så att sökträfflistan inte behöver konvertera lika mycket och tar mindre plats på disk.



Sammanslagning av fält vid uppgraderingar

- Se till att nya fält från Koha läggs till, tas bort eller uppdateras i samband med uppgraderingar utan att tappa befintliga fält

Bugg 19707

Sökmotorinställningar

Varning: Eventuella ändringar i konfigurationen träder i kraft först efter en fullständig o sökningen inte fungerar korrekt

Sök fält **Bibliographic records** Auktoriteter

Namn	Etikett	Typ
abstract	abstract	▼
acqsource	acqsource	▼
ar	archival-control	▼
arl	arl	▼
arp	arp	▼
author	author	Sträng ▼
author-in-order	author-in-order	▼
author-name-corporate	author-name-corporate	▼
author-name-personal	author-name-personal	▼
author-personal-bibliogræ	author-personal-bibliogræ	▼
author-title	author-title	▼



Kvar att göra

- Sortering (annat än relevans)
- Auktoritetsposter
- Konfigurerbar viktning (**Bugg 20589** och/eller **bugg 18316**)
- Få in koden i Koha



När kan ni börja använda det då?!?

- **Bugg 19575** (Korrekta fältnamn)
- **Bugg 19893** (Plocka bort Catmandu)
- **Bugg 19884** (Snabbare exemplarhämtning)
- **Bugg 19707** (Sammanslagning av fält)
- **Bugg 20334** (“/” i sökningar)
- **Bugg 20114** (URL:er för fasetter och paginering)
- ~~**Bugg 19564**~~ (Sorteringsfält med “_”)
- **Bugg 20589 / 18316** (Konfigurerbar viktning)



GÖTEBORGS
UNIVERSITET

GÖTEBORGS UNIVERSITETS BIBLIOTEK

Frågor?