

Skördning webbsidor till Primo

Marika Lundqvist
Umeå universitetsbibliotek

Tillvägagångssätt

- CMS Drupal
 - Modul för export av valfria innehållstyper och fält till XML
- Verktyg för transformation av rå XML-data till XML-schema – MODS
 - Oxygen

XML? XSLT? MODS?

- XML = EXtensible Markup Language
- XSLT = EXtensible Stylesheet Language
- XPath = sökvägar i XML
- MODS = Metadata Object Description Schema
- Tutorials, ex:
 - [w3schools.com XML Tutorial](http://w3schools.com/xml/)
 - [W3schools.com Xpath Tutorial](http://w3schools.com/xpath/)
 - [W3schools.com XSLT Tutorial](http://w3schools.com/xslt/)
 - [MODS \(Library of Congress\)](http://www.loc.gov/mods/)
 - [Outline of Elements and Attributes in MODS Version 3.6](http://www.loc.gov/mods/outline/)

Exempel: Databas

Bibliotek / Sök / Artiklar och databaser / ZETEO

Sök
Böcker
E-böcker
Artiklar och databaser
Nyheter om e-resurser
Testperioder
Tidskrifter
Mobilapplikationer
Avhandlingar och forskningspublikationer
Studentuppsatser
Specialsamlingar
Offentligt tryck
Kartor
Dagstidningar
Sökhjälp och guider
Åtkomst utanför Campus
Användarrättigheter

ZETEO

Databas med all gällande lagstiftning och rättsfall.

Zeteeo är den klart innehållsrikaste juridiska databasen, som innehåller flest propositioner, flest rättsfall med mera. De äldre rättsfallen kan vara en bra källa för lokalthistoriskt material. Samsökningen med ett antal myndigheters webbplatser är mycket användbar.

Innehåller bl.a:

- Hela Blå lagboken, med det alfabetiska registret.
- Många av Norstedts kommentarer och laghandböcker i fulltext, även Rättspraxis i litteraturen finns i fulltext.

Om man i Avancerad sökning klickar i Allt material (fler sökträffar) så hittar man även hänvisningar till Norstedts kommentarer och laghandböcker som vi inte abonnerar på i Zeteeo, och som vi normalt har i pappersform. Databas med Sveriges Rikes Lag på Internet (all gällande lagstiftning, rättsfall - NJA I, RR, RH, Kammarrätterna, AD, och förarbeten - propositioner, NJA II m.m.)

Djupa kommentarer till dessa lagar:

- Brottsbalken
- Aktiebolagslagen
- Inkomstskattelagen
- Rättegångsbalken
- Sekretesslagen

[ZETEO](#)

Rådata XML (från Drupal)

```
1 <?xml version="1.0" encoding="UTF-8" ?>
```

```
<node>
  <titel>ZETEO </titel>
  <ingress>
    Databas med all gällande lagstiftning och rättsfall.
  </ingress>
  <body>
    Zeteo är den klart innehållsrikaste juridiska databasen, som /.../
  </body>
  <utgivare>Norstedts juridik</utgivare>
  <path>http://www.ub.umu.se/sok/artiklar-databaser/zeteo</path>
  <URL>http://proxy.ub.umu.se/login?qurl=http%3A//zeteo.wolterskluwer.se/autologin.html</URL>
  <Nid>12714</Nid>
  <taxonomi>Alla databaser inom juridik, Juridik</taxonomi>
  <LibrisID> </LibrisID>
  <ISSN> </ISSN>
</node>
```

kan vara

böcker som
rättsfall -

```
27 <ISSN></ISSN>
```

```
28 </node>
```

```
--
```

Transformer till MODS

```
1 <xsl:transform version="2.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
2   <xsl:output method="xml" indent="yes" encoding="utf-8"/>
3   <xsl:template match="/">
4     <modsCollection>
5       <xsl:for-each select="nodes/node">
6         <mods>
7           <titleInfo>
8             <title>
9               <xsl:value-of select="normalize-space(.)"/>
10            </title>
11          </titleInfo>
12          <originInfo>
13            <publish
14              <xsl:for-each select="tokenize(taxonomi,',')">
15                <subject>
16                  <xsl:value-of select="normalize-space(.)"/>
17                </subject>
18              </xsl:for-each>
19            </publish
20            </originInfo>
21            <xsl:for-each select="publisher">
22              <subject>
23                <xsl:value-of select="utgivare"/>
24              </subject>
25            </xsl:for-each>
26          <abstract>
27            <xsl:if test="ingress">
28              <xsl:value-of select="ingress"/>
29            </xsl:if>
30            <xsl:if test="body != ''">
31              <xsl:value-of select="body"/>
32            </xsl:if>
33          </abstract>
34        </mods>
35      </xsl:for-each>
36    </modsCollection>
37  </template>
38 </transform>
```

<xsl:for-each select="nodes/node">

<xsl:for-each select="tokenize(taxonomi,',')">

<subject>

<xsl:value-of select="normalize-space(.)"/>

</subject>

</xsl:for-each>

</titleInfo>

<originInfo>

<publisher>

<xsl:value-of select="utgivare"/>

</publisher>

</originInfo>

Transformerat till MODS

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <modsCollection>
3 <mods>
4 <titleInfo>
5 <title>Z
6 </titleInfo>
7 <originInfo>
8 <publisher>
9 </originInfo>
10 <subject>Al
11 <subject>Ju
12 <abstract>
13 Databas med al
14 Zeteo är den k
15 en bra källa för
16 Innehåller bl.
17 Hela Blå la
18 Många av No
19 Om man i Avancera
20 vi inte abonnerar på i zeteo, och som vi normalt har i pappersform.Databas med Sveriges Rikes Lag på Internet (all gällande lagstiftning, rättsfall -
21 NJA I, RR, RH, Kammarrätterna, AD, och förarbeten - propositioner, NJA II m.m)
22 Djupa kommentarer till dessa lagar:
23 Brottsbalken
24 Aktiebolagslagen
25 Inkomstskattelagen
26 Rättegångsbalken
27 Sekretesslagen
28 </abstract>
29 <location>
30 <url>http://www.ub.umu.se/sok/arti
31 </location>
32 <recordInfo>
33 <recordIdentifier>12714</recordIdentifier>
34 </recordInfo>
35
```

```
<titleInfo>
  <title>ZETEO </title>
</titleInfo>
<originInfo>
  <publisher>Norstedts juridik</publisher>
</originInfo>
<subject>Alla databaser inom juridik</subject>
<subject>Juridik</subject>
```

rättsfall med mera. De äldre rättsfallen kan vara
et användbar.

ns i fulltext.

ill Norstedts kommentarer och laghandböcker som

Originaldata:

```
<taxonomi>Alla databaser inom juridik, Juridik</taxonomi>
```

Och nu över till Primo...

Back Office: File splitter

- Mapping tables > File splitters
 - Splitter class: `com.exlibris.primo.publish.platform.harvest.splitters.generic.DomXmlSplitter`

Mapping Table Rows

Enabled	Name*	Splitter Class	Description
<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input checked="" type="checkbox"/>	MARC Exchange splitter	<code>com.exlibris.primo.publish.platform.harvest.splitters</code>	MARC Exchange splitter
<input checked="" type="checkbox"/>	Extensions File Splitter	<code>com.exlibris.primo.publish.platform.harvest.splitters</code>	Extensions File Splitter
<input checked="" type="checkbox"/>	OAI splitter	<code>com.exlibris.primo.publish.platform.harvest.splitters</code>	OAI Splitter
<input checked="" type="checkbox"/>	Static OAI splitter	<code>com.exlibris.primo.publish.platform.harvest.splitters</code>	Static OAI Splitter
<input checked="" type="checkbox"/>	WARC splitter	<code>com.exlibris.primo.publish.platform.harvest.splitters</code>	WARC splitter
<input checked="" type="checkbox"/>	SFX XML splitter	<code>com.exlibris.primo.publish.platform.harvest.splitters</code>	SFX XML splitter
<input checked="" type="checkbox"/>	MODS splitter	<code>com.exlibris.primo.publish.platform.harvest.splitters</code>	MODS splitter

Back Office: File splitter

- Mapping tables > File splitters params
 - IdentifierXPath
 - RootXPath
 - FullRecordXPath

```
<modsCollection>  
  <mods>  
    [post]  
  </mods>  
<mods>  
  [post]  
</mods>  
</modsCollection>
```

IdentifierXPath	//recordInfo/recordIdentifi	MODS splitter	XPath to identifier	20
RootXPath	modsCollection	MODS splitter	XPath to root	20
FullRecordXPath	modsCollection/mods	MODS splitter	XPath to single record	20

Back Office: Data Source

- Source Name
- Source Code
- Source System
- File Splitter
- Source Format

Skapa pipe

> Define Pipe

Owner: Umeå universitetsbibliotek

Pipe Name: UMUB_MODS_DATABASES_R

Pipe Type: Regular

Pipe Description:

Source

Normalization and Enrichment

Errors

Data Source: *

UmUB databaser MODS

Harvesting method: *

Copy

Source directory: *

/exlibris/primis/data_i

Delete after copy

Start harvesting files/records from: *

Apr 09, 2015

Start time: 13 : 53 : 56

Normalization Mapping Set: *

UMUB_MODS_DATABASES

Enrichment Set: *

No Enrichments - Template

Harvested File Format: *

*.tar

System Last Stage: *

PERSISTENCE

Priority: *

Medium

Maximum error threshold:

5 % (1-100)

Force DEDUP

Force FRBR

Back Office: Normaliseringsregler

- Xpath för att definiera reglerna
- Delivery and Scoping: Online Resource
- Resource type: Website
- Länkfältet: Add to the beginning of string \$\$U

Lägg Search Scope i vyn

- Lägg till Search Scope
 - a) Search Scope Name: hämtas från aktuell Data Source, fältet Source Code
 - b) Description: hämtas från aktuell Data Source, fältet Source Name

> Search Scope List

View : Umeå universitetsbiblic ▼

Search Scope List

Name	Description	Search Type		
UMUB_MODS_DATABASER	UmUB databaser MODS	Local	Edit	Delete
UMUB_FOARK_SAMLINGAR_MODS	UmUB foark samlingar	Local	Edit	Delete
UmUB_Aleph	Aleph	Local	Edit	Delete
default_scope_umub	Default scope för UmUB	Both	Edit	Delete
LIBRIS	Libris deep link	Both	Edit	Delete
UmUB_SFX	SFX	Local	Edit	Delete

Aktivera Search Scope till vald/valda scopes

General Attributes for Search Scopedefault_scope_umub

Scope Name :

default_scope_umub

Search Scope Description :

Display Text:

General:

Search Scope Values

	Type	Name
<input checked="" type="checkbox"/>	other	UmUB databaser MODS
<input type="checkbox"/>	other	DiVA MODS
<input checked="" type="checkbox"/>	other	Foark arkiv MODS
<input type="checkbox"/>	collection	Homepages
<input type="checkbox"/>	collection	Barn faktaböcker
<input type="checkbox"/>	collection	Kursböcker
<input type="checkbox"/>	collection	Läromedel
—		

Exempel databas

1 **ZETEO** Online

[Fulltext](#) [Detaljer](#) [Spara/dela](#)

Databas

Titel: ZETEO

Ämnesord: Alla databaser inom juridik; Juridik

Länkar

[Länk till resurs](#)

Utgivare: Norstedts juridik

Beskrivning: Databas med all gällande lagstiftning och rättsfall. Zeteeo är den klart innehållsrikaste juridiska databasen, som innehåller flest propositioner, flest rättsfall med mera. De äldre rättsfallen kan vara en bra källa för lokalhistoriskt material. Samsökningen med ett antal myndigheters webbplatser är mycket användbar. Innehåller bl.a: Hela Blå lagboken, med det alfabetiska registret. Många av Norstedts kommentarer och laghandböcker i fulltext, även Rättspraxis i litteraturen finns i fulltext. Om man i Avancerad sökning klickar i Allt material (fler sökträffar) så hittar man även hänvisningar till Norstedts kommentarer och laghandböcker som vi inte abonnerar på i Zeteeo, och som vi normalt har i pappersform. Databas med Sveriges Rikes Lag på Internet (all gällande lagstiftning, rättsfall - NJA I, RR, RH, Kammarrätterna, AD, och förarbeten - propositioner, NJA II m.m) Djupa kommentarer till dessa lagar: Brottsbalken Aktiebolagslagen Inkomstskattelagen Rättegångsbalken Sekretesslagen

Källa: XML

Böcker (30)
E-Böcker (5)
Databaser (3)
Avhandlingar (1)
Artiklar (1)

[Fler alternativ](#)

Utgivningsår

Från Till [Begränsa](#)

1982 2014

[2005 till 2010](#) (8)
[Efter 2010](#) (8)
[2000 till 2004](#) (6)
[Före 1990](#) (5)
[1990 till 1999](#) (5)

Språk



[Svenska](#) (30)
[Engelska](#) (1)


[Fler alternativ](#)

Ämne


[Sverige. Aktiebolagslagen](#) (12)
[Sverige](#) (10)

Länkat till datakällans index


Juridik  

Träffar 1 - 10 av 12 1 2 


sorterade på: Relevans Datum (nyaste först)

1  **JP Djurnet**
● Online
[Fulltext](#) [Detaljer](#) [Spara/dela](#)

Databas

2  **JP Miljönet**
● Online
[Fulltext](#) [Detaljer](#) [Spara/dela](#)

Databas

3  **Norstedts lagkommentarer**
● Online
[Fulltext](#) [Detaljer](#) [Spara/dela](#)

Databas