

IV1023 vt2022

Avancerad Datahantering med XML

XML-baserade språk, RDF/RDFS

nikos dimitrakas
nikosd@kth.se
08-161295
Rum 2423

Läsanvisningar
Kapitel 1, 4, 16, 18 i kursboken
Kapitel 31.4 i Database Systems (Connolly, Begg) upplaga 5

XML-baserade språk

- **Definition av struktur**
- **Definition av semantik**
- **XML-grundregler**
 - Alfabet, Vokabulär
- **XML Schema (eller DTD)**
 - Grammatik, Syntax
- **XML Schema-förklaringen (för människor)**
 - Semantik, Betydelse

XML-baserade språk - exempel

- **RDF och RDFS (Resource Description Framework)**
 - RDF XML (RDF är mer än bara ett XML-baserat språk)
- **DC (Dublin Core)**
- **MathML**
- **CML (Chemical Markup Language)**
- **RecipeML**
- **ODF (Open Document Format)**
- **OOXML (Open Office XML)**
- **SVG (Scalable Vector Graphics)**
- **WSDL (Web Services Description Language)**
- **RSS (Really Simple Syndication)**
- **och många många fler**

3

DC (Dublin Core)

- **Katalogisering av böcker, artiklar, etc.**
 - Nära relation till RDF
- **15 element**
 - Title, Subject, Description, Type, Source, Relation, Coverage, Creator, Publisher, Contributor, Rights, Date, Format, Identifier, Language
- **Många flera metadata-termerna**
 - omdefinierar även de 15 elementen

```
<bok xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>Det fina havet</dc:title>
  <dc:description>Fin bok om havet med bilder</dc:description>
  <dc:publisher>Fina Böcker Förlag</dc:publisher>
  <!-- etc -->
</bok>
```

4

MathML

- Representation av matematiska ekvationer
 - Visuell representation
 - Semantisk representation

$$x^3+5$$

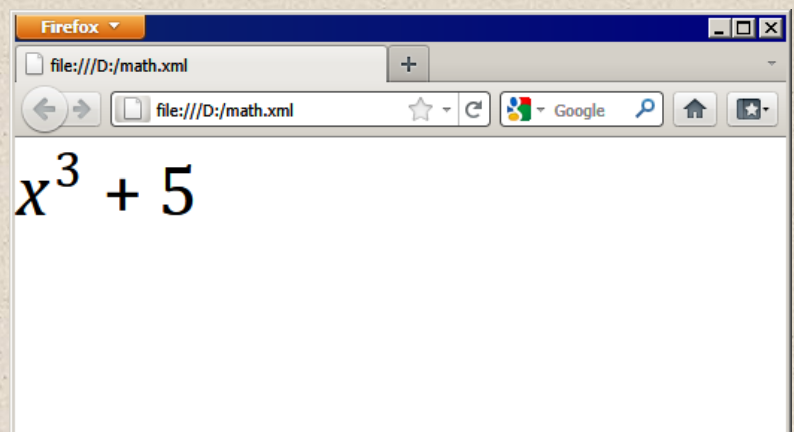
```
<math>
<mrow>
  <msup>
    <mi>x</mi>
    <mn>3</mn>
  </msup>
  <mo>+</mo>
  <mn>5</mn>
</mrow>
</math>
```

```
<math>
  <apply>
    <plus />
    <apply>
      <power />
      <ci>x</ci>
      <cn>3</cn>
    </apply>
    <cn>5</cn>
  </apply>
</math>
```

MathML

- Stödjs i vissa webbläsare:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<math xmlns="http://www.w3.org/1998/Math/MathML">
  <mrow>
    <msup>
      <mi>x</mi>
      <mn>3</mn>
    </msup>
    <mo>+</mo>
    <mn>5</mn>
  </mrow>
</math>
```



CML (Chemical Markup Language)

- **Representation av kemiska formler**
 - molekyler, reaktioner, etc
- **Element**
 - atom, atomParity, atomSet, atomType
 - electron, molecule, particle
 - action, reaction, reactant
 - angle, bond, crystal, dimension
 - formula, isotope, spectrum
 - etc.
- **Namespace**
 - <http://www.xml-cml.org/schema>

RecipeML

- **Beskriver recept**
 - ingredienser
 - process
- **Element**
 - recipe
 - ingredients och ing
 - directions och step

```
<recipeml>
  <recipe>
    <head>
      <title>Chocolate Milk</title>
    </head>
    <ingredients>
      <ing>
        <amt><qty>1</qty><unit>cup</unit></amt>
        <item>milk</item>
      </ing>
      <ing>
        <amt><qty>2</qty><unit>tsp.</unit></amt>
        <item>chocolate powder</item>
      </ing>
    </ingredients>
    <directions>
      <step>Pour the milk in a cup</step>
      <step>Add the chocolate powder</step>
      <step>Mix until powder dissolves</step>
    </directions>
  </recipe>
</recipeml>
```

ODF (Open Document Format)

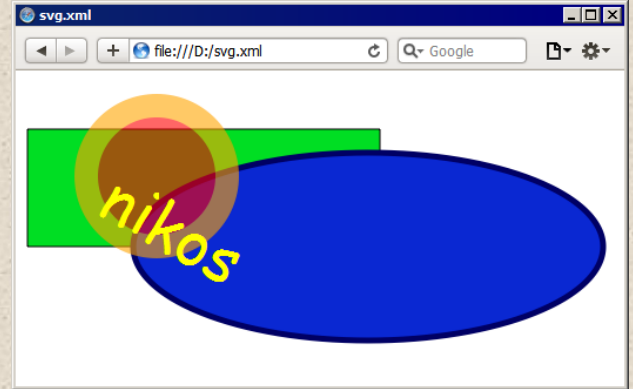
- **För representation av**
 - dokument
 - kalkylblad
 - presentationer
 - grafer, diagram, tabeller
 - databaser
- **OASIS-standard och ISO-standard**
- **Använder MathML**
- **Den senaste versionen har en formell relation till RDF**

OOXML (Open Office XML)

- **Microsofts motsvarighet till ODF**
 - Blev standard efter mycket lobbying
- **Flera delspråk**
 - WordprocessingML
 - SpreadsheetML
 - PresentationML
 - OfficeMathML
 - DrawingML

SVG (Scalable Vector Graphics)

- Representerar 2D-grafik
- Stödjer
 - vektorgrafik
 - bilder
 - text
 - interaktivitet
 - animation
- Stöds delvis i de flesta webbläsare
 - Vissa (t ex IE8) kräver plugin



```
<svg width="100%" height="100%" version="1.1" xmlns="http://www.w3.org/2000/svg">
  <rect x="10" y="50" width="300" height="100" style="fill:rgb(0,222,35); stroke-width:1; stroke:rgb(0,0,0)"/>
  <ellipse cx="300" cy="150" rx="200" ry="80" style="fill:rgb(10,40,210); stroke:rgb(0,0,100); stroke-width:5"/>
  <circle cx="120" cy="90" r="60" stroke="orange" stroke-width="20" fill="red" opacity="0.6"/>
  <text x="120" y="70" font-family="Comic Sans MS" font-size="54" fill="yellow" transform="rotate(30)">nikos</text>
</svg>
```

11

WSDL (Web Services Description Language)

- Används för att beskriva Web Services
 - definiera datatyper
 - definiera anrop (operationer med input och output)
 - definiera detaljer om protokoll
- Används tillsammans med SOAP

```
<?xml version="1.0" encoding="UTF-8"?>
<wsdl:definitions xmlns:wsdl="http://schemas.xmlsoap.org/wsdl/" xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <wsdl:types>
    <complexType name="DTCustomer">
      <sequence><element name="customerName" nillable="true" type="xsd:string"/>
        <element name="userName" nillable="true" type="xsd:string"/></sequence>
    </complexType>
  </wsdl:types>
  <wsdl:message name="loginRequest">
    <wsdl:part name="in0" type="xsd:string"/>
    <wsdl:part name="in1" type="xsd:string"/>
  </wsdl:message>
  <wsdl:message name="loginResponse">
    <wsdl:part name="loginReturn" type="DTCustomer"/>
  </wsdl:message>
  <wsdl:portType name="B2BService">
    <wsdl:operation name="login" parameterOrder="in0 in1">
      <wsdl:input message="loginRequest" name="loginRequest"/>
      <wsdl:output message="loginResponse" name="loginResponse"/>
    </wsdl:operation>
  </wsdl:portType>
  <wsdl:service name="B2BServiceService">
    <wsdl:port name="B2BWebService">
      <wsdlsoap:address location="http://iv1023.kth.se/services/B2BWebService"/>
    </wsdl:port>
  </wsdl:service>
</wsdl:definitions>
```

12

RSS (Really Simple Syndication)

- Används ofta för nyhetsflöden, bloggar, annonser, etc.
- Tidigare "RDF Site Summary" och relaterat till RDF
- Kallat även "Rich Site Summary"
- Kan även kombineras med DC

```
<rss version="2.0">
<channel>
  <link>http://www.cbsnews.com/</link>
  <title>Breaking News: CBS News</title>
  <item>
    <title>Stocks turn lower as optimism about jobs fades</title>
    <link>http://feeds.cbsnews.com/~r/CBSNewsMain/~3/WdSyTMQjwds/main20088553.shtml</link>
    <description>Day after worst single-day drop since 2008, Dow jumps on strong hiring report - then quickly loses gains</description>
    <pubDate>Fri, 05 Aug 2011 07:47:45 EDT</pubDate>
  </item>
  <item>
    <title>Syrian troops fire on defiant protesters; 8 dead</title>
    <link>http://feeds.cbsnews.com/~r/CBSNewsMain/~3/5eg3GhnUoG0/main20088559.shtml</link>
    <description>Regime's security forces open fire on tens of thousands across nation calling for downfall of President Bashar Assad</description>
    <pubDate>Fri, 05 Aug 2011 07:58:53 EDT</pubDate>
  </item>
</channel>
</rss>
```

13

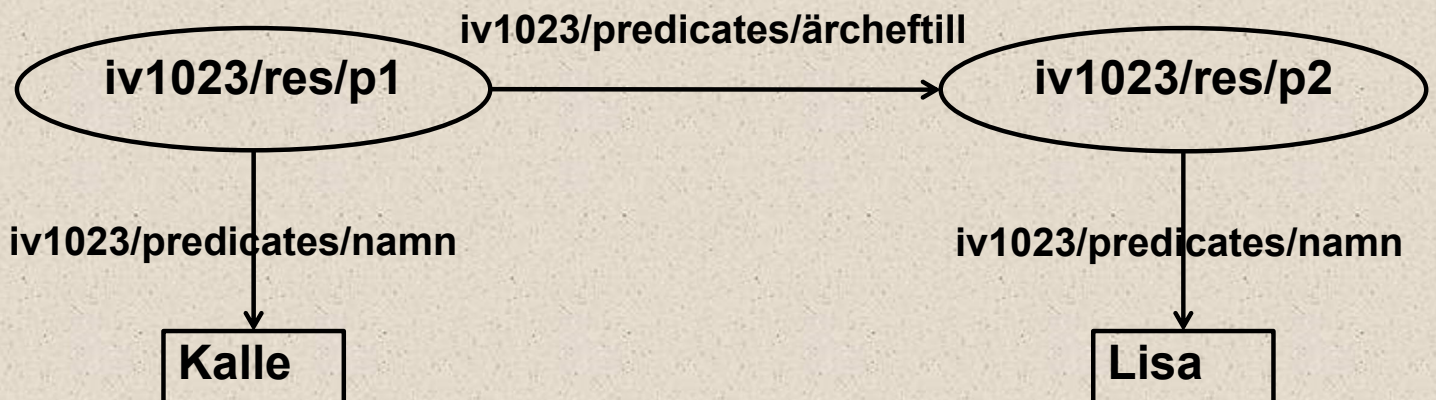
RDF (Resource Description Framework)

- Används för att beskriva resurser (allt i universum)
 - Identifiera en resurs unikt med en URI
 - Påstå något om den resursen
- RDF-påståenden
 - Ett subjekt (själva resursen)
 - Ett predikat (en property, motsvarar attribut i andra terminologier)
 - » Predikatet i sig är en resurs
 - Ett objekt (värdet, vilket kan vara en annan resurs)
- RDF-representationer
 - Graf
 - XML (så "RDF/XML" är ett XML-baserat språk)
- RDF 1.1
 - IRI istället för URI
 - Flera datatyper (från XSD)
 - Nya serialiseringsmöjligheter

14

RDF exempel

- Kalle är chef till Lisa
- Kalle är en resurs
 - Kalle är namnet
- Lisa är en resurs
 - Lisa är namnet



15

RDF exempel i XML

```

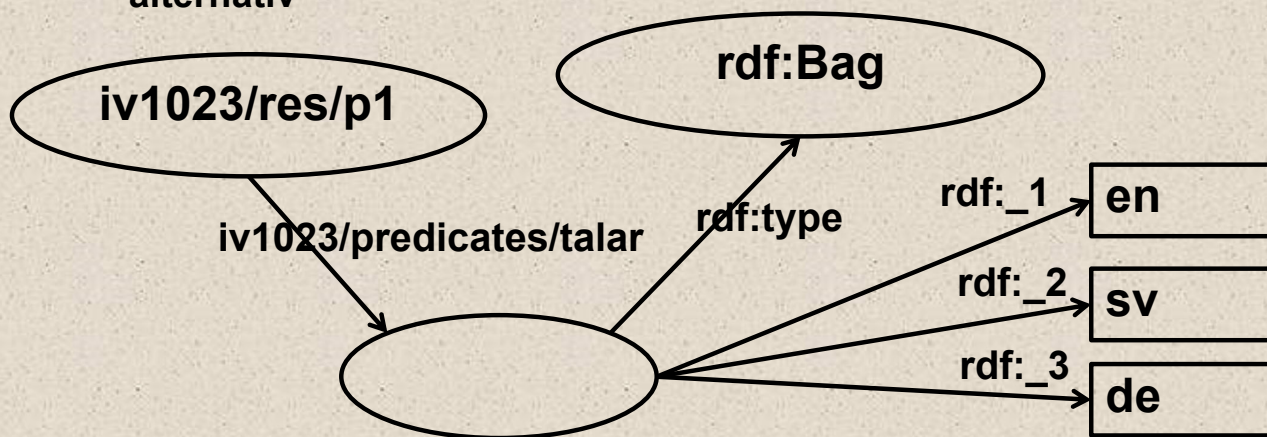
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:iv1023="iv1023/predicates/">
  <rdf:Description about="iv1023/res/p1">
    <iv1023:namn>Kalle</iv1023:namn>
    <iv1023:ärcheftill rdf:resource="iv1023/res/p2" />
  </rdf:Description>
  <rdf:Description about="iv1023/res/p2">
    <iv1023:namn>Lisa</iv1023:namn>
  </rdf:Description>
</rdf:RDF>
  
```

- about - global URI
- ID - lokal identifierare

16

RDF containers

- Flera resurser/literaler
- **rdf:Bag** (subklass av **rdfs:Container**)
 - ingen ordning, möjligt med dubletter
- **rdf:Seq** (subklass av **rdfs:Container**)
 - ordning, möjligt med dubletter
- **rdf:Alt** (subklass av **rdfs:Container**)
 - alternativ



17

RDF containers, XML

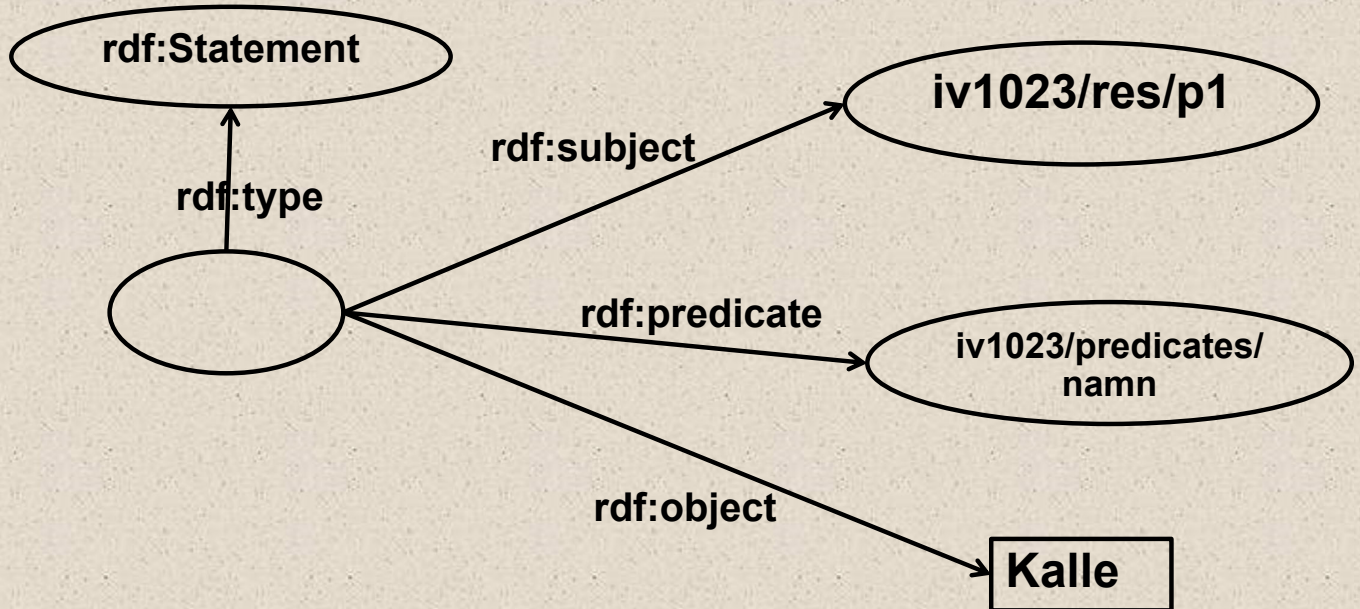
```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:iv1023="iv1023/predicates/">
  <rdf:Description about="iv1023/res/p1">
    <iv1023:talar>
      <rdf:Bag>
        <rdf:li>en</rdf:li>
        <rdf:li>sv</rdf:li>
        <rdf:li>de</rdf:li>
      </rdf:Bag>
    </iv1023:talar>
  </rdf:Description>
</rdf:RDF>
  
```

18

RDF Statement

- Själva påståendet är en resurs
- Kallas "Reifiering" (Reification)



19

RDF Statement

```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Statement>
    <rdf:subject resource="iv1023/res/p1"/>
    <rdf:predicate resource="iv1023/predicates/namn"/>
    <rdf:object>Kalle</rdf:object>
  </rdf:Statement>
</rdf:RDF>
  
```

20

RDFS (RDF Schema)

- **Används för att modellera/definiera resurstyper**
 - Varje resurs är av en klass
 - Varje property är en resurs av klassen `rdf:Property` (har `rdf:type = rdf:Property`)
 - Varje klass är en resurs av klassen `rdfs:Class` (har `rdf:type = rdfs:Class`)
- **Klasser kan ärva av varandra**
- **Properties har**
 - `rdfs:domain` - vilka resurstyper denna property kan appliceras på (vilka resurser kan vara subject med denna property som predicate)
 - `rdfs:range` - vilka värden denna property kan ha (vilka resurser eller literaler kan vara object)
- **RDFS är property-centric**
 - UML är class-centric

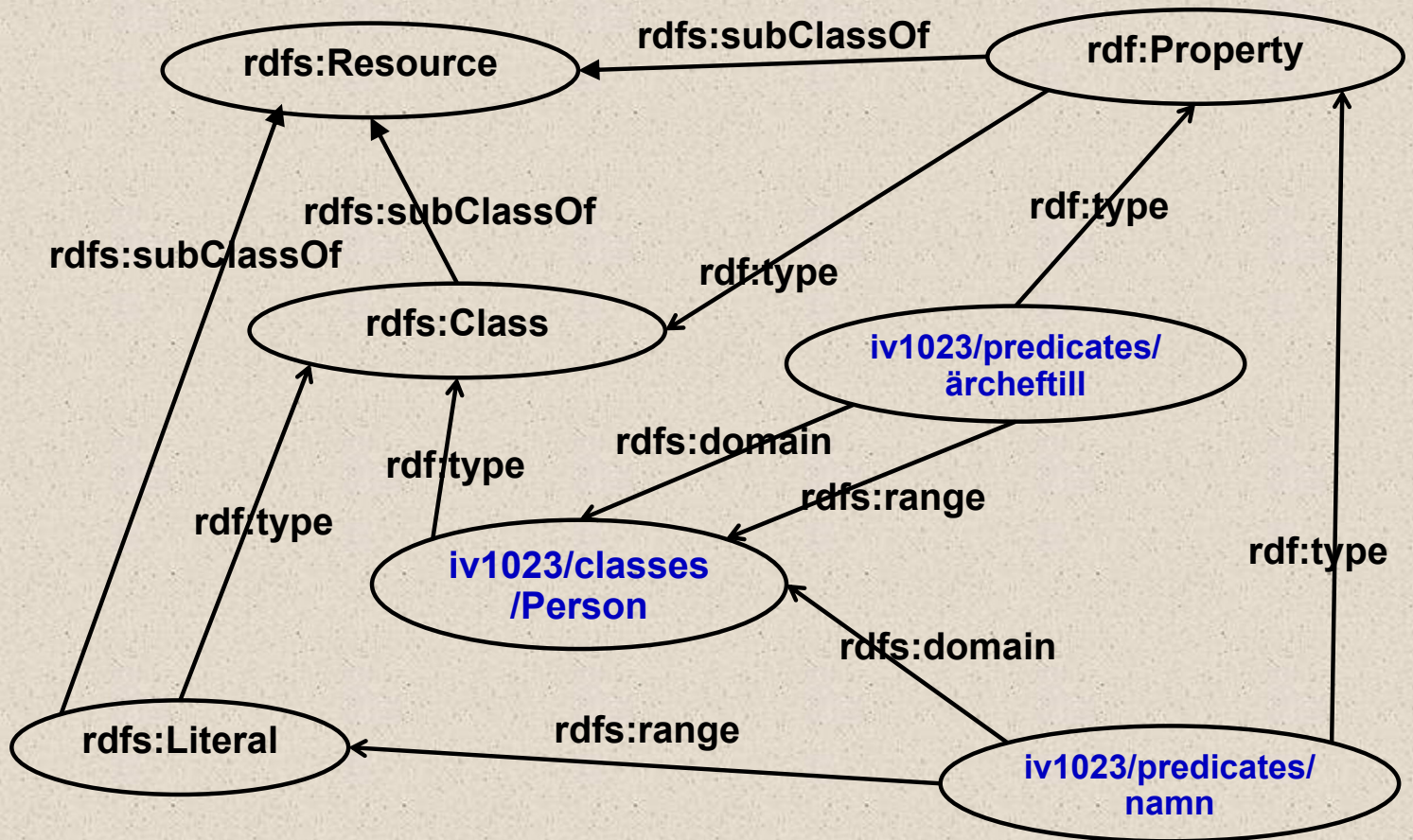
21

RDFS (RDF Schema)

- **`rdfs:Resource` - superklass till**
 - `rdfs:Class`
 - `rdfs:Literal`
 - `rdf:Property`
- **vilket innebär att alla instanser är resurser**
- **så literaler är en speciell typ av resurser**

22

RDFS - exempel



23

RDFS - exempel XML

```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  >
  <rdfs:Class rdf:about="iv1023/classes/Person" />
  <rdf:Property about="iv1023/predicates/namn">
    <rdfs:domain rdf:resource="iv1023/classes/Person"/>
    <rdfs:range rdf:resource="rdfs:Literal"/>
  </rdf:Property>
  <rdf:Property about="iv1023/predicates/ärcheftill">
    <rdfs:domain rdf:resource="iv1023/classes/Person"/>
    <rdfs:range rdf:resource="iv1023/classes/Person"/>
  </rdf:Property>
</rdf:RDF>

```

24

Resurser baserade på RDFS

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:iv1023="iv1023/predicates/" xmlns="iv1023/classes/">
  <Person rdf:about="iv1023/res/p1">
    <iv1023:namn>Kalle</iv1023:namn>
    <iv1023:ärcheftill rdf:resource="iv1023/res/p2" />
  </Person>
  <Person rdf:about="iv1023/res/p2">
    <iv1023:namn>Lisa</iv1023:namn>
  </Person>
</rdf:RDF>
```

Det andra Person-elementet är likvärdigt detta:

```
<rdf:Description about="iv1023/res/p2" type="iv1023/classes/Person">
  <iv1023:namn>Lisa</iv1023:namn>
</rdf:Description>
```

RDF-förkortningar (RDF abbreviations)

Information i form av ett reifierat påstående

```
<rdf:Statement>
  <rdf:subject resource="iv1023/res/p1"/>
  <rdf:predicate resource="iv1023/predicates/namn"/>
  <rdf:object>Kalle</rdf:object>
</rdf:Statement>
```

Samma information direkt som Person-element

```
<Person rdf:about="iv1023/res/p1">
  <iv1023:namn>Kalle</iv1023:namn>
</Person>
```

Även med attribut för namnet

```
<Person rdf:about="iv1023/res/p1" namn="Kalle"/>
```

Eller som en rdf-resurs med typen specificerad

```
<rdf:Description about="iv1023/res/p1" type="iv1023/classes/Person">
  <iv1023:namn>Kalle</iv1023:namn>
</rdf:Description>
```

RDF-förkortningar (RDF abbreviations)

- **Resursens typ blir elementnamn**
 - istället för `rdf:Description` med `rdf:type`
- **En resurs predicates och objects blir nästlade element**
 - istället för att hanteras som separata påståenden
- **Element som har literal som värde kan hanteras som attribut**
 - istället för att vara subelement
- **En resurs kan vara helt nästlad i en annan**
 - antingen som en egen `rdf:Description` eller förkortad

RDF, DC och andra språk

- **Allt kan ses som resurser**
 - RSS channel, item
 - MathML
 - CML
 - RecipeML
- **Alla resurser kan ha DC-metadata**
 - `rdf:Statement` kan ha `dc:creator`, `dc:date`
 - `rss:channel` och `rss:item` kan ha `dc:date`, `dc:language`, `dc:title`
 - etc.

Fortsättning

- Quiz om RDF, RDFS, XML-baserade språk
- Lektionsuppgifter
- Inlupp 7 (Artiklar om frågespråk för SSD & XML, samt om XML-baserade språk)