KDI : Knowledge and Data Integration

Resource Description Framework (RDF)

W8.L15.M5.T15.1.1and4

Contents

Contents

What is XML?

- XML stands for eXtensible Markup Language
- XML is a markup language much like HTML
- XML was designed to store and transport data
- XML is a W3C Recommendation
- All major browsers have a built-in XML parser to access and manipulate XML.

XML vs. HTML

The Adventures of Tom Sawyer

Front piece of *The Adventures of Tom Sawyer*

Author	Mark Twain
Cover artist	created by Mark Twain
Country	United States
Language	English, Limited Edition(Spanish)
Genre	Bildungsroman, picaresque, satire, folk, children's novel
Publisher	American Publishing Company
Publication date	1876[1]
OCLC	47052486
Dewey Decimal	Fic. 22
LC Class	PZ7.T88 Ad 2001
Followed by	*Adventures of Huckleberry Finn*

HTML: focus on presentation

```
<h2>The adventures of Tom Sawyer</h2>
…
<b>Author: </b> Mark Twain <br>
<b>Cover artist: </b> created by <a href="http://…">Mark Twain </a>
…
```

XML: focus on metadata

```
<book>
        <title> The adventures of Tom Sawyer </title>
        <author> Mark Twain </author>
        <genre> Bildungsroman </genre>
        <genre> picaresque </genre>
        …
        <publisher> American Publishing Company </publisher>
        <year>1876</year>
</book>
```

XML and HTML (Similarities)

- They both use tags and attributes
- Tags may be nested
- Human can read and interpret both XML and HTML quite easily
- Machines can read and interpret only to some extent
- Both XML and HTML were influenced and derived from SGML (Standard Generalized Markup Language)

XML vs. HTML (Differences)

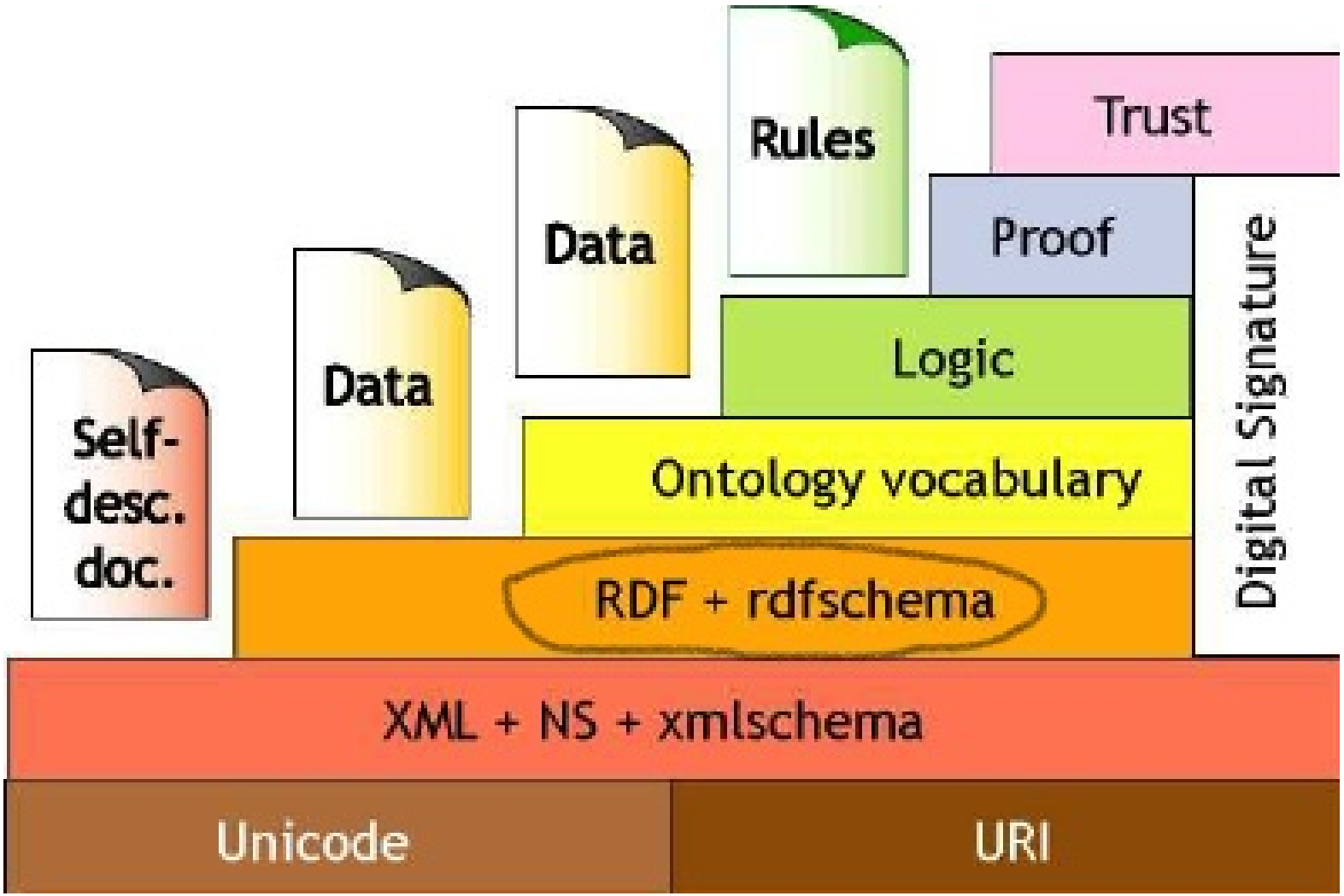
- HTML is to tell machines about how to interpret formatting for graphical presentation **(focus on how data looks)**
- XML is to tell machines about metadata content and relationships between different pieces of information **(focus on what data is)**
- XML allows the definition of constraints on values
- HTML tags are fixed, while XML tags are user defined
- HTML is not Case sensitive, whereas XML is Case sensitive.

More About XML

- XML is considered a meta-markup language, i.e. a language for defining markup languages
- Query Languages for XML include:- XQuery, XML-QL, XQL, YATL etc.
- The XML DOM (Document Object Model) defines a standard way for accessing and manipulating XML documents. It presents an XML document as a tree-structure.
- XPath can be used to navigate through elements and attributes in an XML document.
- XSL (eXtensible Stylesheet Language) is a styling language for XML. XSLT stands for XSL Transformations.
- By using XML, RDF information can easily be exchanged between different types of computers (RDF/XML Serialization).

Contents

RDF and the Semantic Web



What is RDF?

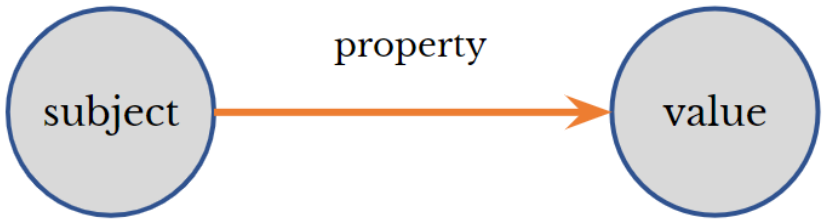
- A language for representing Web resources and information about them in the form of metadata [RDF Primer]
- A language to represent all kinds of things that can be identified on the Web [RDF Primer]
- A domain independent data model for representing information on the Web [G. Antoniou and F. van Harmelen, 2004]
- A language with an underlying model designed to publish data on the Semantic Web [F. Giunchiglia et al., 2010]

Contents

RDF language and data model

RDF language:

- A language for expressing simple statements of the form subject-property-value (binary predicates), with reasoning and inferencing capabilities
- The data model in RDF is a graph data model
- An edge with two connecting nodes forms a triple



RDF language

Formal syntax:

- RDF has been given a syntax in XML and inherits all its benefits
- Statements in RDF are machine comprehensible

Resources:

- An object, an entity or anything we want to talk about (e.g. authors, books, publishers, places, people, facilities)

RDF language (Contd.)

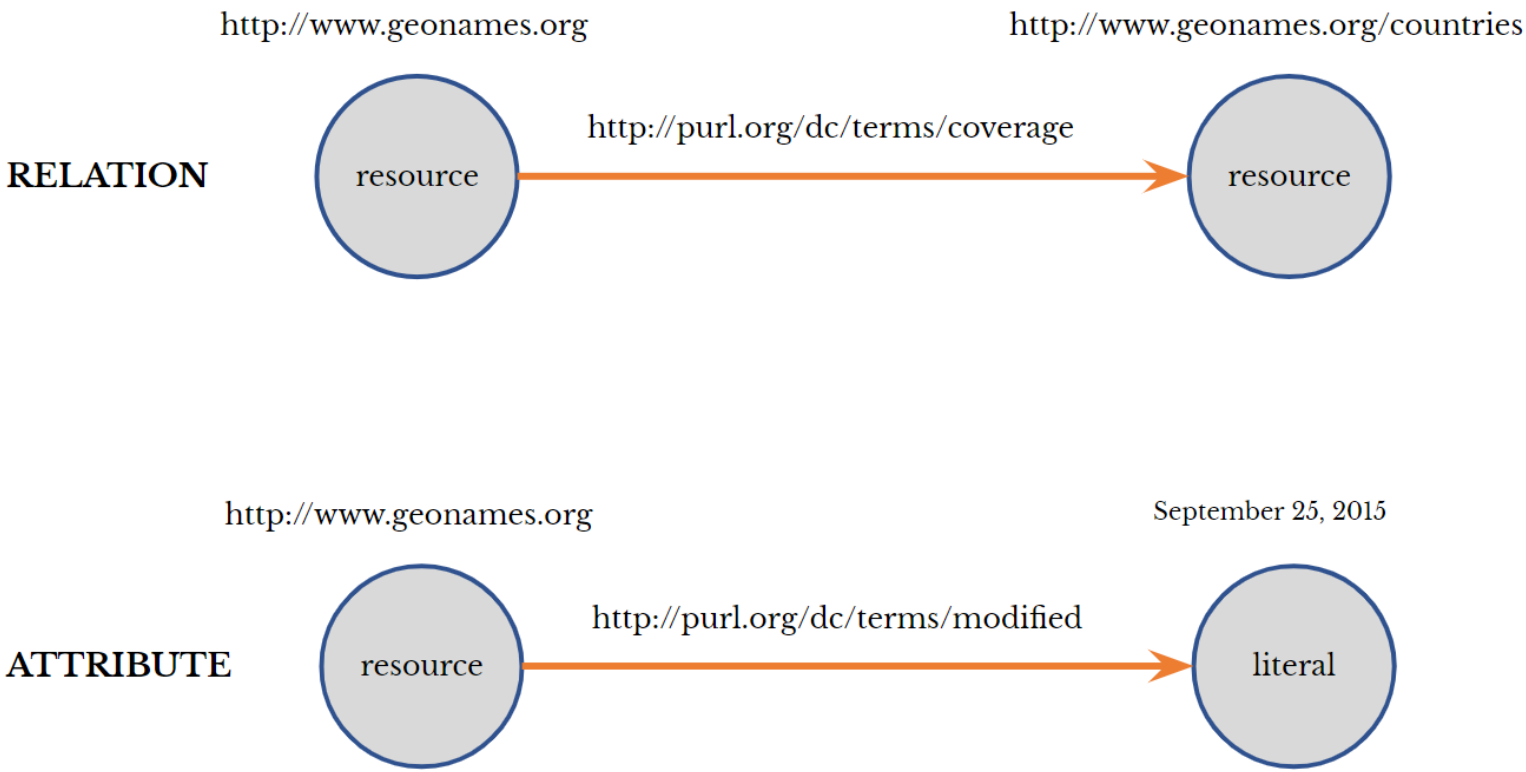
Properties:

- They codify **relations** (e.g. written-by, friend-of, located-in, . . .) and **attributes** (e.g. age, date of birth, length . . .)

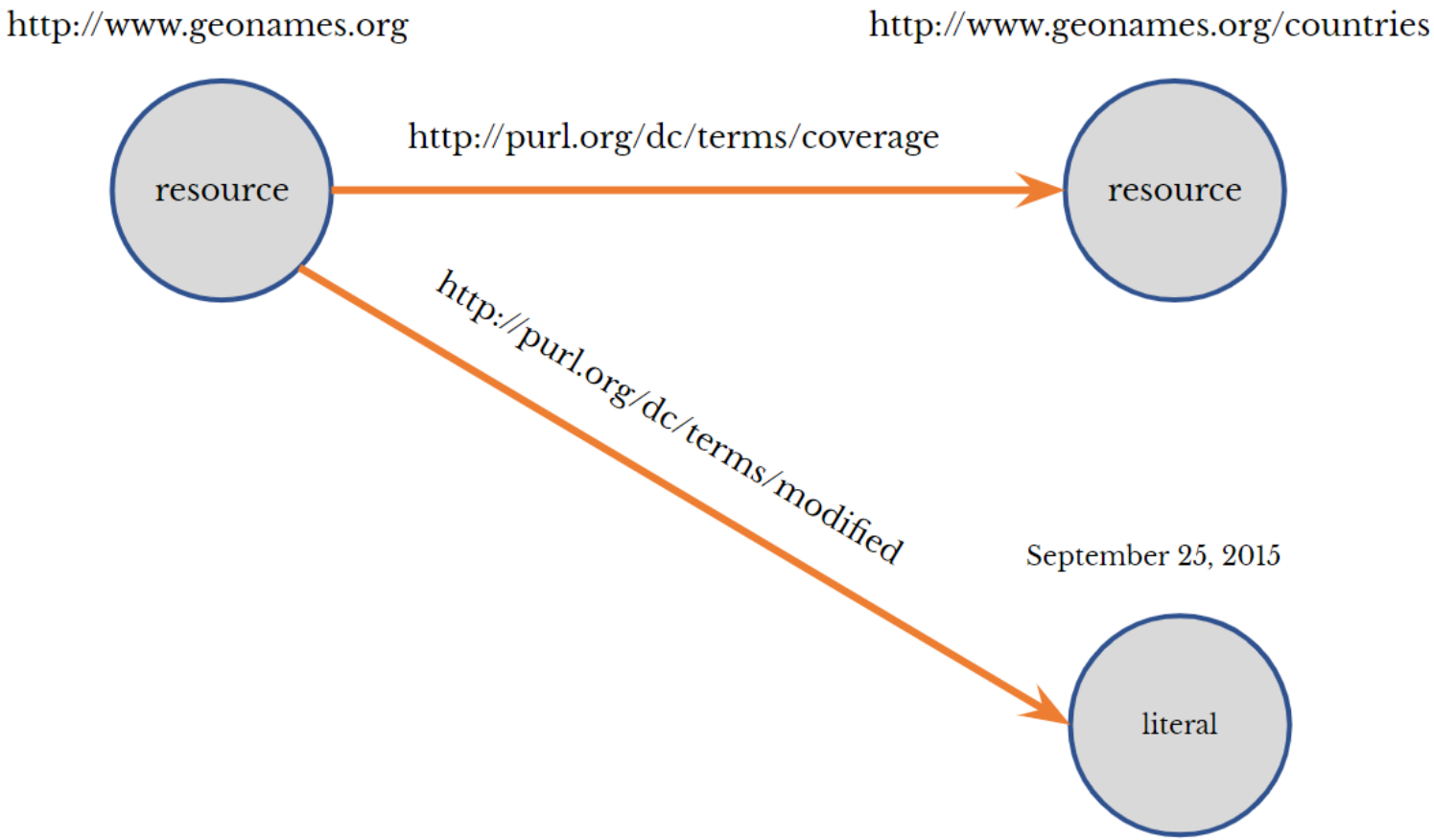
Statements:

- Statements assert the properties of resources in form of triples subject-property-value
- Every resource and property has a URI (an URL or any other identifier)
- Values can be resources (for relations) or literals (for attributes)

RDF Data Model



RDF Graph


http://www.geonames.org
http://www.geonames.org/countries
resource
http://purl.org/dc/terms/coverage
resource
http://purl.org/dc/terms/modified
September 25, 2015
literal

XML syntax example

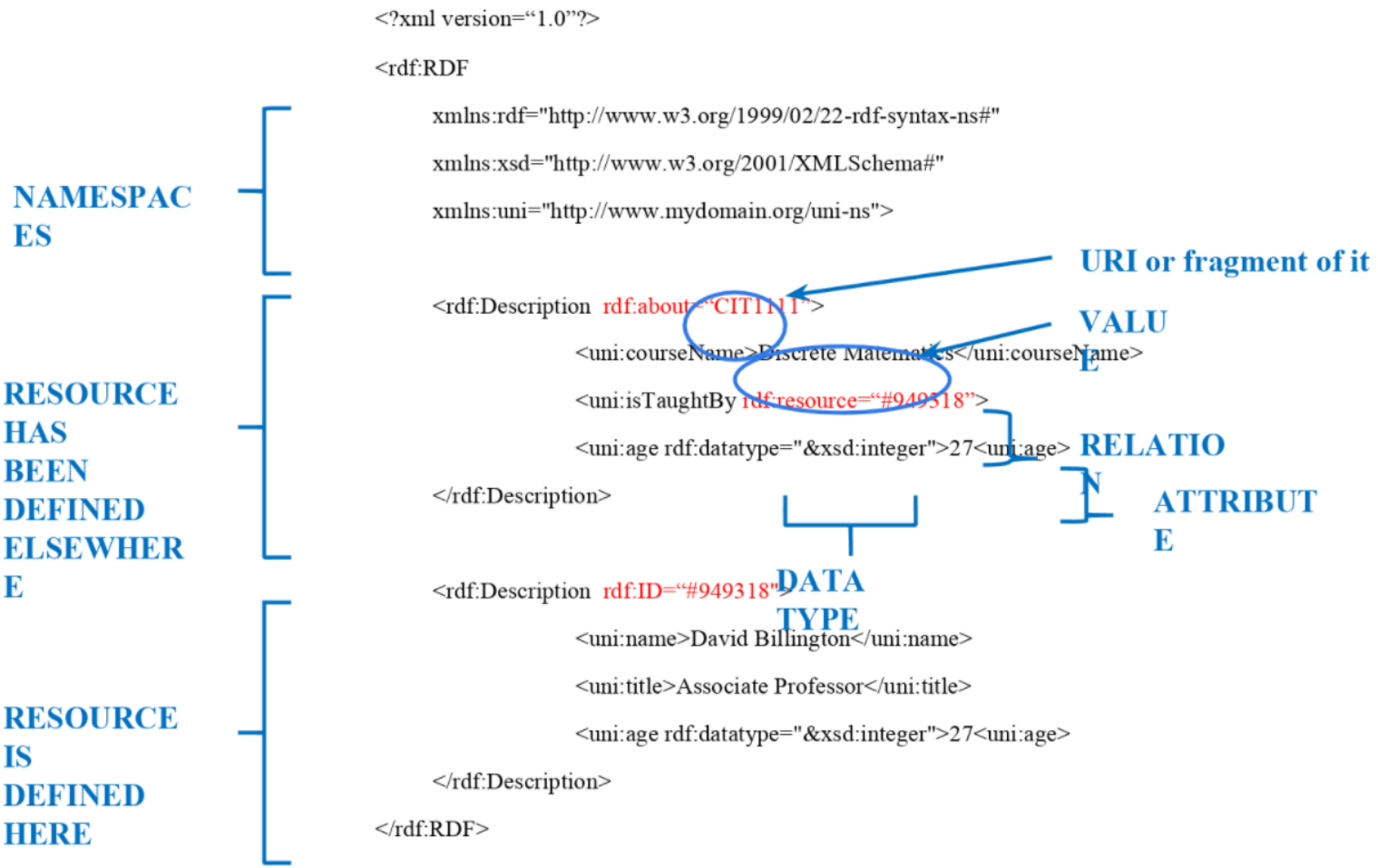
```
<?xml version="1.0"?>

 <rdf:RDF
          xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
          xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
          xmlns:dc="http://purl.org/dc/terms#">

<rdf:Description rdf:about="http://www.geonames.org">
     <rdfs:label>GeoNames</rdfs:label>
     <dc:coverage rdf:resource="http://www.geonames.org/countries"/>
     <dc:modified>September 25, 2015</dc:modified>
</rdf:Description>

</rdf:RDF>
```

RDF/XML elements



RDF typing

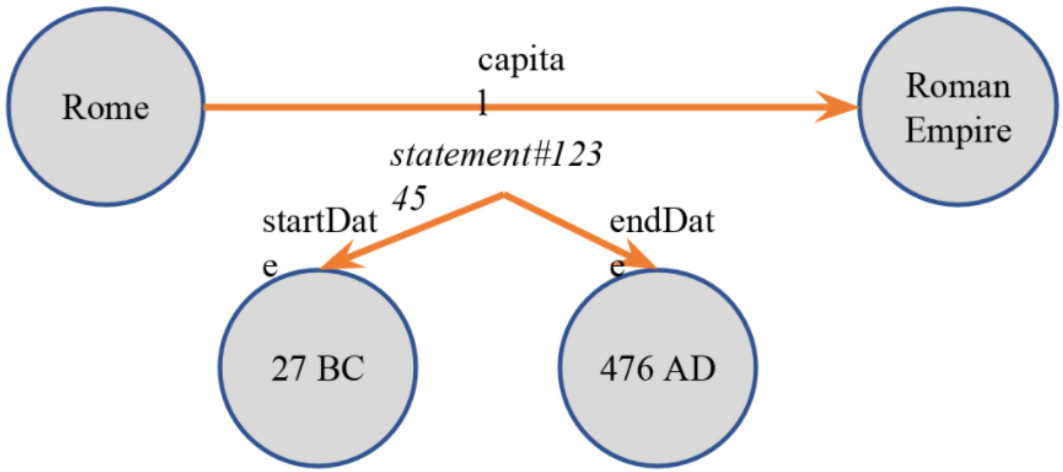
```
<rdf:Description rdf:ID="CIT1111">
        <rdf:type
    rdf:resource="http://www.mydomain.org/uni-ns#course"/>
        <uni:courseName>Discrete Maths</uni:courseName>
        <uni:isTaughtBy rdf:resource="#949318"/>
</rdf:Description>

<rdf:Description rdf:ID="949318">
        <rdf:type
    rdf:resource="http://www.mydomain.org/uni-ns#lecturer"/>
        <uni:name>David Billington</uni:name>
        <uni:title>Associate Professor</uni:title>
</rdf:Description>
```

RDF Reification

RDF Reification can be used to represent:

- Generic statements about statements
- Structured attributes (e.g. address)
- Units of measure
- Provenance information
- Time validity and other contextual information



Contents

RDF Schema (RDFS)

RDF:

- RDF is a universal language that lets users describe resources in their own vocabularies
- RDF by default does not assume, nor defines semantics of any particular application domain

RDF Schema (RDFS) [Contd.]

RDF Schema (RDFS): A language defined to provide mechanisms to add semantics to RDF resources, in terms of:

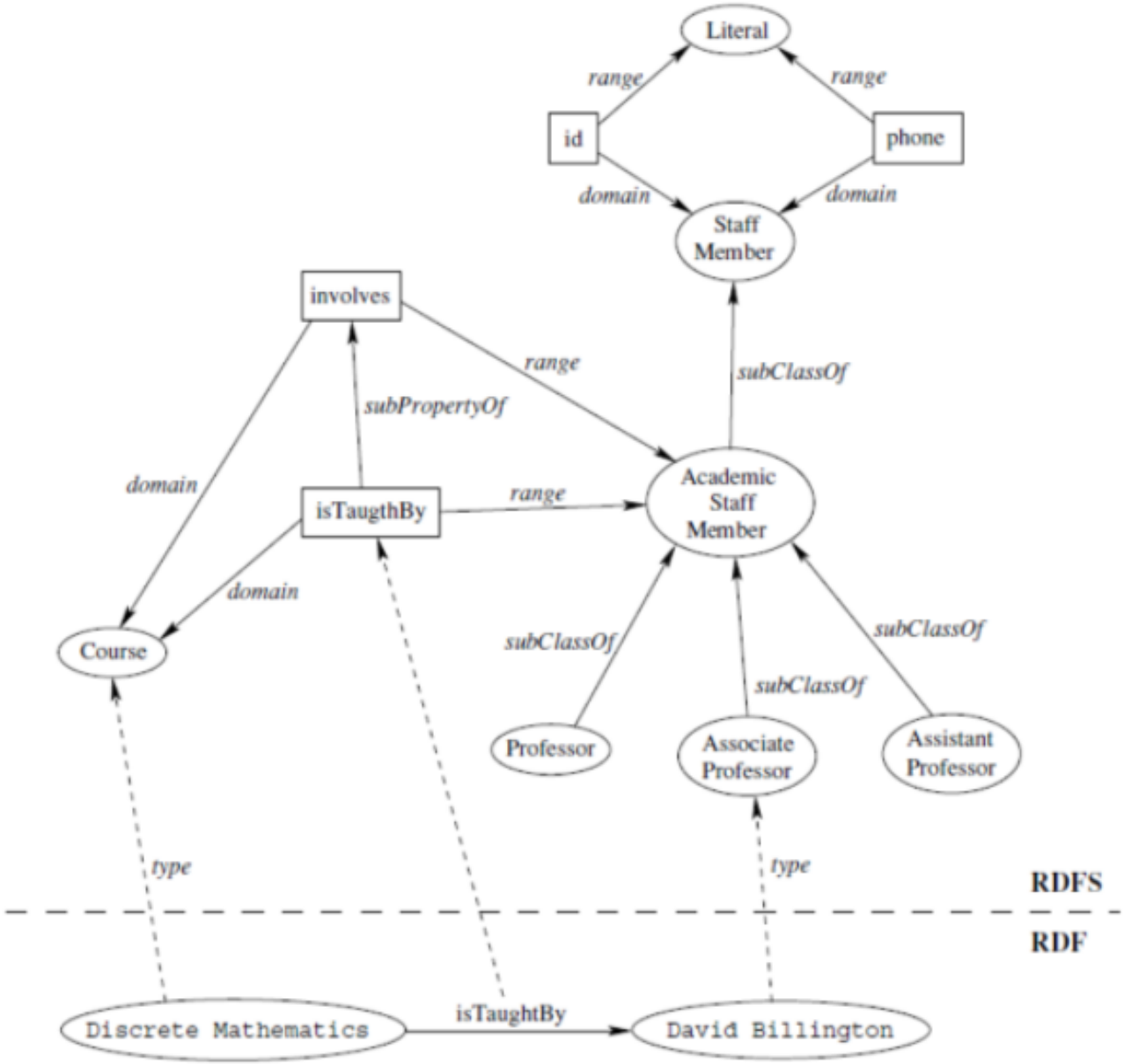
- Classes (**rdfs:Class**) and Properties (**rdfs:Property**)
- Class Hierarchies and Inheritance (**rdfs:subClassOf**)
- Property Hierarchies (**rdfs:subPropertyOf**)
- Domain (**rdfs:domain**) and range (**rdfs:range**) of properties

RDF Schema (RDFS) [Contd.]

- It is similar to the object-oriented programming (OOP) paradigm with the difference that in OOP the central notion is the class (and properties are defined for them), while in RDF the central notion is the property and classes are used to specify their domain/range.

- **Classes and instances**: Individual objects that belong to a class are referred to as instances of that class (**rdf:type**).

Graphical Example



RDF Schema in XML

```
<rdfs:Class rdf:about="#lecturer">
        <rdfs:subClassOf rdf:resource="#staffMember"/>
</rdfs:Class>
<rdf:Property rdf:ID="phone">
        <rdfs:domain rdf:resource="#staffMember"/>
        <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
</rdf:Property>
```

Utility Properties:

- **rdfs:seeAlso** relates a resource to another resource that explains it
- **rdfs:isDefinedBy** is a subproperty of **rdfs:seeAlso** and relates a resource to the place where its definition, typically an RDF schema, is found
- **rdfs:comment** support comments that can be associated with a resource
- **rdfs:label** is a human-friendly name associated with a resource

```
<rdfs:Class rdf:ID="course">
        <rdfs:comment>The class of courses</rdfs:comment>
</rdfs:Class>

<rdf:Property rdf:ID="isTaughtBy">
        <rdfs:comment>
                Inherits its domain ("course") and range ("lecturer")
                from its superproperty "involves"
        </rdfs:comment>
        <rdfs:subPropertyOf rdf:resource="#involves"/>
</rdf:Property>
```

Contents

Inference System

Sound and complete set of inference rules:

- The RDF inference system consists of inference rules
- **Sound:** inference rules prove only formulas that are valid with respect to its semantics
- **Complete:** every formula having a certain property can be derived) inference systems

Examples of rules:

(transitivity)
IF E contains the triples (?u,rdfs:subClassOf,?v) and (?v,rdfs:subclassOf,?w)
THEN E also contains the triple (?u,rdfs:subClassOf,?w)

(inheritance)
IF E contains the triples (?x,rdf:type,?u) and (?u,rdfs:subClassOf,?v)
THEN E also contains the triple (?x,rdf:type,?v)

RDF Inferencing by example

Type (rdf:type) propagation through rdfs:subClassOf

:Fausto Giunchiglia	rdf:type	:Professor
:Professor	rdfs:subClassOf	:Faculty
:Fausto Giunchiglia	rdf:type	:Faculty (inferred)

Relationship propagation through rdfs:subPropertOf

:professorshipAt	rdfs:subProperytOf	:affiliationWith
:Fausto Giunchiglia	:professorshipAt	:UniTN
:Fausto Giunchiglia	:affiliationWith	:UniTN (inferred)

Type identification through rdfs:domain

:professorshipAt	rdfs:domain	:Person
:Fausto Giunchiglia	:professrshipAt	:UniTn
:Fausto Giunchiglia	rdf:type	:Person (inferred)

RDF Inferencing by example

Type identification through rdfs:range

:professorshipAt	rdfs:range	:Educational_Institution
:Fausto_Giunchiglia	:professrshipAt	:UniTn
:UniTn	rdf:type	:Educational_Institution (inferred)

Inferencing through rdfs:domain and rdfs:subClassOf

:Researcher	rdfs:subClassOf	:Scientist
:hIndex	rdfs:domain	:Researcher
:Fausto Giunchiglia	:hIndex	44
:Fausto Giunchiglia	rdf:type	:Researcher (inferred)
:Fausto Giunchiglia	rdf:type	:Scientist (inferred)

Inferencing through rdfs:range and rdfs:subClassOf

:Educational_Institution	rdfs:subClassOf	:Organization
:professorshipAt	rdfs:range	: Educational Institution
:Fausto Giunchiglia	:professorshipAt	:UniTn
:UniTn	rdf:type	:Educational Institution (inferred)
:UniTn	rdf:type	:Organization (inferred)

Intersection and union in RDF

- **Set Intersection** (if an entity e is in X, it is also in both Y and Z)

X	rdfs:subClassOf	Y
X	rdfs:subClassOf	Z
e	rdf:type	X
e	rdf:type	Y (inferred)
e	rdf:type	Z (inferred)

- **Set Union** (any entity e that belongs either to Y or Z also belongs to X)

Y	rdfs:subClassOf	X	
Z	rdfs:subClassOf	X	
e	rdf:type	Y	or
e	rdf:type	Z	
e	rdf:type	X (inferred)	

Contents

Summary

- RDF provides a domain-independent foundation for representing and processing metadata, using a graph-based data model
- RDF has a decentralized philosophy and allows incremental building of knowledge, and its sharing and reuse across the Web
- RDF has a (XML-based) syntax and a semantics (via RDF Schema)
- RDF Schema provides a mechanism for describing specific domains
- RDF Schema is a primitive ontology language, and it offers certain modelling primitives with fixed meaning.

KDI : Knowledge and Data Integration

W8.L15.M5.T15.1.1and4

Resource Description Framework (RDF)