TECHNISCHE UNIVERSITÄT DRESDEN

FOUNDATIONS OF SEMANTIC WEB TECHNOLOGIES

Overview and XML

Sebastian Rudolph

DRESDEN concept
Exzellenz aus Wissenschaft und Kultur

Agenda

- Introduction of Lecturer
- Organizational Matters
- What is the “Semantic Web”?

Agenda

- Introduction of Lecturer
- Organizational Matters
- What is the “Semantic Web”?

Introduction of Lecturer

Prof. Dr. Sebastian Rudolph

Since Apr 2013	Full Professor for Computational Logic at the Computer Science Department, TU Dresden
2006 – 2013	Research Assistant $\rightarrow$ Project Leader $\rightarrow$ Privatdozent at the Chair of Knowledge Management, Institute AIFB, University of Karlsruhe $\rightarrow$ Karlsruhe Institute of Technology
2003 – 2006	Research Assistant at the Chair of Psychology of Teaching and Learning, TU Dresden
2000 – 2003	PhD Scholarship Holder Graduate School, TU Dresden
1995 – 2000	Studies for high-school teaching (Math, Physics, CS), TU Dresden

Introduction of Lecturer

scientific interests

- logic-based knowledge representation and reasoning
- semantic technologies
- complexity and decidability problems
- computational linguistics
- theory of databases
- (and much more)

Agenda

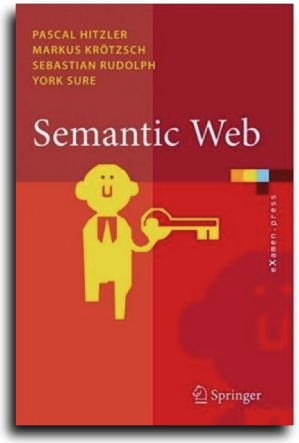
- Introduction of Lecturer
- Organizational Matters
- What is the “Semantic Web”?

Organizational Matters: Time and Place

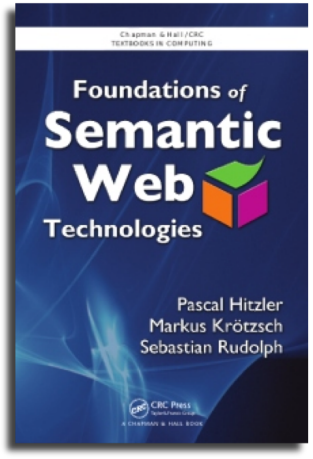
- Wednesdays, 09:20 – 10:50 (2.DS) and 11:10 – 12:40 (3.DS)
- Mondays, 09:20 – 10:50 (2.DS)
- exact schedule see webpage
- INF E005
- accompanying web page:
  `https://ddll.inf.tu-dresden.de/web/Foundations_of_Semantic_Web_Technologies_%28SS2017%29/en`

Literature

Hitzler, Krötzsch, Rudolph, Sure
“Sematic Web Grundlagen”
Springer-Verlag



Hitzler, Krötzsch, Rudolph, Sure “Foundations of Semantic Web Technologies”
CRC Press



Agenda

- Introduction of Lecturer
- Organizational Matters
- What is the “Semantic Web”?

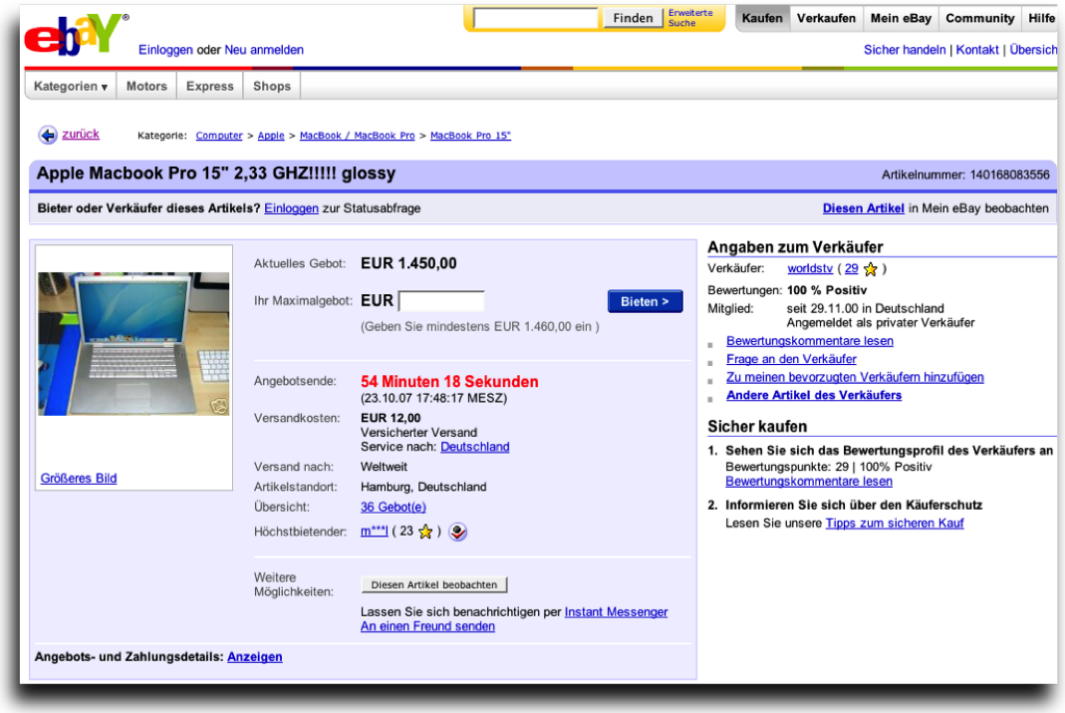
The Web

The Web is at the heart of the transition from industrial to information society, providing the infrastructure for a novel quality of handling information in terms of retrieval and provision

- high availability
- high up-to-date-ness
- low cost

The Web

Commercialization on all levels



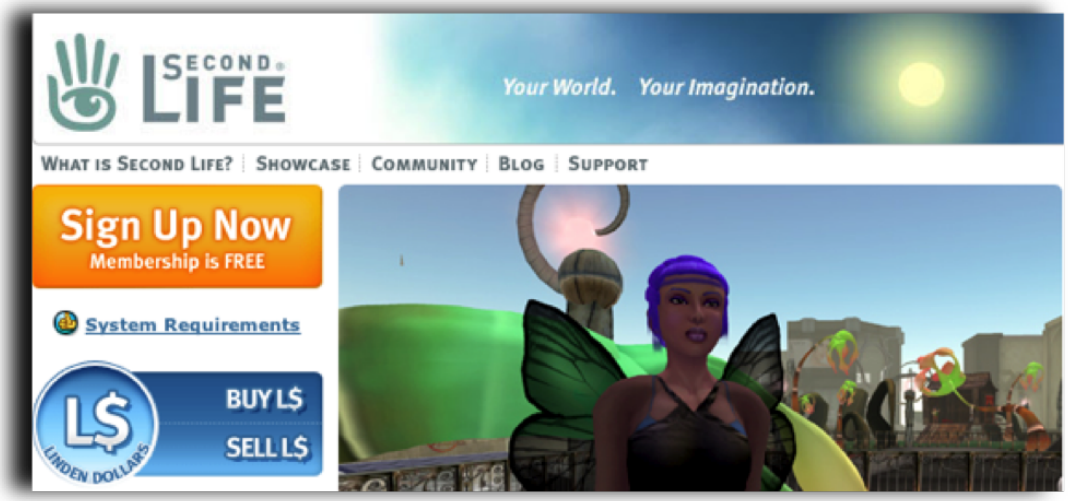
The Web

Commercialization on all levels

The Web

Further aspects of daily life are being “webized”:

- authorities, administration (eGovernment)
- education (eLearning, eEducation)
- social contacts (social networking platforms, dating sites)
- everyday life?



What means “Semantic”?

Syntax vs. Semantics

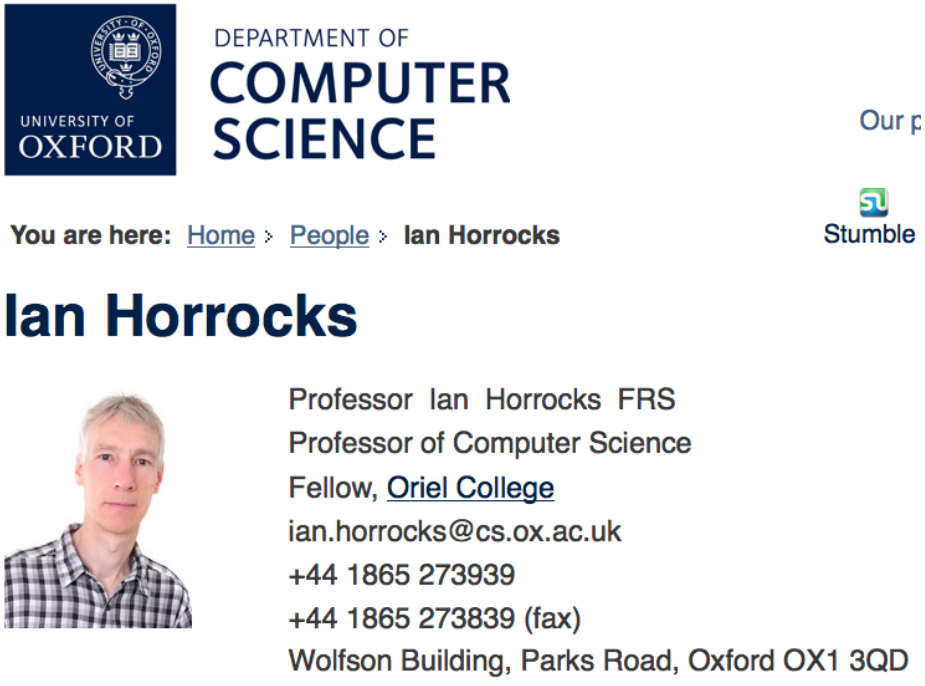
Syntax (from greek $\sigma\upsilon\nu\tau\alpha\xi\iota\varsigma$ composition, sentential structure) denotes the (normative) structure of data, i.e., it characterizes what makes data “well-formed”

Semantics (greek $\sigma\eta\mu\alpha\upsilon\tau\iota\kappa o\varsigma$ belonging to the sign) denotes the meaning of data, i.e., it characterizes what conclusions can be drawn from it.

$4+) = ($	$3 + 4 = 12$	$3 + 4 = 7$
syntactically wrong	syntactically correct	syntactically correct
–	semantically wrong	semantically correct

Problems of the Web

- plethora of information
- targeted at human end user



Problems of the Web

- plethora of information
- targeted at human end user

Prof. Dr. Carsten Lutz
Visitors: Room 3090, MZH, Bibliotheksstr. 1
Tel.: ++49 421 218-64431
Fax: ++49 421 218-9864431
EMail: clu@informatik.uni-bremen.de

Universität Bremen
Fachbereich 03
Postfach 330440
28334 Bremen

AG Theory of Artificial Intelligence.

Problems of the Web

- plethora of information
- targeted at human end user

漆桂林

教授，博导
东南大学计算机与工程学院

电话：025 52090910
传真：0 25 52090880
电子邮箱：gqi at seu dot edu dot cn

个人简介:

漆桂林，东南大学教授，博士生导师。中国计算机协会会员，ACM会员。1998年宜春学院数学专业毕业，2002年江西师范大学数学与信息系硕士研究生，2006年获英国贝尔法斯特女皇大学计算机博士学位。2006年8月至2009年8月在德国Karlsruhe大学AIFB研究所做博士后研究。

长期从事人工智能和语义网络的推理方面科研及教学工作。发表高质量学术论文60多篇。特别是在国际人工智能顶级会议IJCAI、AAAI、KR和UAI，以及国际语义网络顶级会议ISWC、ESWC发表多篇会议文章,在国际顶级杂志

Problems of the Web

- plethora of information
- targeted at human end user

Guilin Qi

Professor
School of Computer Science and Engineering
Southeast University
China

Tel : +86 (0) 25 52090910
Fax : +86 (0) 25 52090880
E-mail : gqi@seu.edu.cn

Curriculum Vitae (PDF)

Brief Introduction : Dr. Guilin Qi is a professor working at Southeast University in China. His research topics include knowledge representation and reasoning, semantic Web, uncertainty reasoning. His current research interests include the areas of

- **Knowledge representation**: belief merging, belief revision, inconsistency handling, nonmonotonic reasoning, information fusion, argumentation, paraconsistent logic

only computer-readable layout information

```
<h1>Ian Horrocks</h1> <table><tr>
  <td class="personImg">
    <img src="IH005-1.jpg" alt="Photo Ian Horrocks"/>
  </td>
  <td>
    <div class="personinfo">
      <div>Professor Ian Horrocks FRS</div>
      <div>Professor of Computer Science</div>
      <div>Fellow, <a href="http://www.oriel.ox.ac.uk">Oriel College</a></div>
      <div>ian.horrocks@cs.ox.ac.uk</div>
      <div>+44 1865 273939</div>
      <div>+44 1865 273839 (fax)</div>
    </div>
    <p>Wolfson Building, Parks Road, Oxford OX1 3QD</p>
  </td>
</tr></table>
```

Problems of the Web

- localizing information problematic
- today's search engines good but mostly keyword-based
- desirable: search for content $\rightarrow$ semantic search

Problems of the Web

- Heterogeneity of present information on diverse levels:
  - character encoding (e.g. ASCII vs. Unicode)
  - used natural languages
  - positioning of information on webpages
- desirable: cross-web information integration

Problems of the Web

- **implicit knowledge**, i.e. many pieces of information are not provided explicitly, but follow from the combination of the given data
- requires methods from formal logics
- automated deduction

Problems of the Web

Approaches toward a solution:

1. Ad hoc: Deployment of AI methods (most notably NLP techniques) to evaluate existing unstructured information on the Web
2. A priori: structure information on the Web at authoring time in a way facilitating later automated deployment

Problems of the Web

Approaches toward a solution:

1. Ad hoc: Deployment of AI methods (most notably NLP techniques) to evaluate existing unstructured information on the Web
2. A priori: structure information on the Web at authoring time in a way facilitating later automated deployment

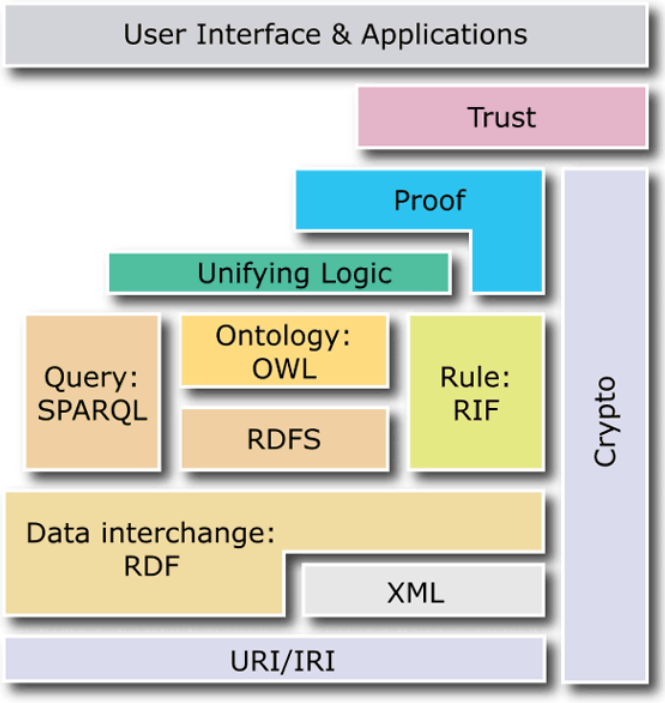
$\Rightarrow$ Semantic Web

Problems of the Web

two essential prerequisites for the implementation:

1. open standards for describing information
   - clearly defined
   - flexible
   - extendable
2. methods for eliciting information from such descriptions

Semantic Web – Standards



1994	First public presentation of the Semantic Web idea
1998	Start of standardization of data model (RDF) and a first ontology languages (RDFS) at W3C
2000	Start of large research projects about ontologies in the US and Europe (DAML & Ontoknowledge)
2002	Start of standardization of a new ontology language (OWL) based on research results
2004	Finalization of the standard for data (RDF) and ontology (OWL)
2008	Standardization of a query language (SPARQL)
2009	Extension of OWL to OWL 2.0
2010	Standard Rule Interchange Format (RIF)

Agenda

- XML – Motivation/Idea
- XML – Syntax
- IRIs
- Name Spaces

Agenda

- XML – Motivation/Idea
- XML – Syntax
- IRIs
- Name Spaces

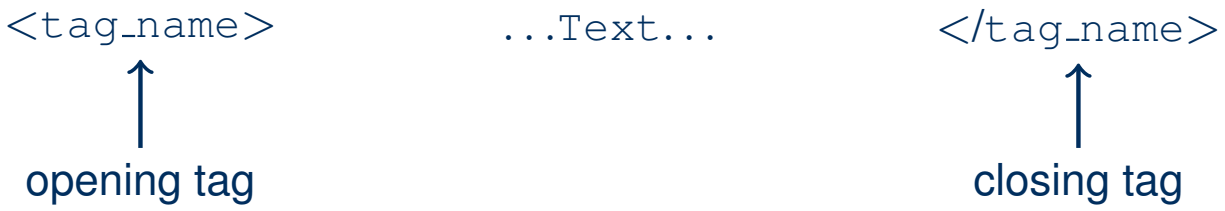
Annotation with Mark-up Languages

- basic idea of mark-up: endow (unstructured) text with additional information (or structure)
- synonym: annotate text

text	= data
additional information	= metadata

Annotation with Mark-up Languages

- common strategy: include to-be-annotated text in so-called tags:



- Additional information is read and interpreted by processing software

Annotation with Mark-up Languages

- most prominent example: HTML tags encode visual presentation information:
  `<i>This book</i> has the title <b>Foundations of Semantic Web Technologies</b>.`
- Output of web browser:
  This book has the title **Foundations of Semantic Web Technologies**.
- Strategy also suited for annotation of content, e.g.:
  `<firstname>Sebastian</firstname> <lastname>Rudolph</lastname> works in <city>Dresden</city>.`

Annotation with Markup-Languages

```
<lecture>
    <title>
      Deduction Systems
    </title>
    <lecturer>
      <title>
        Prof. Dr.
      </title>
      <firstname>
        Sebastian
      </firstname>
      <lastname>
        Rudolph
      </lastname>
    </lecturer>
</lecture>
```

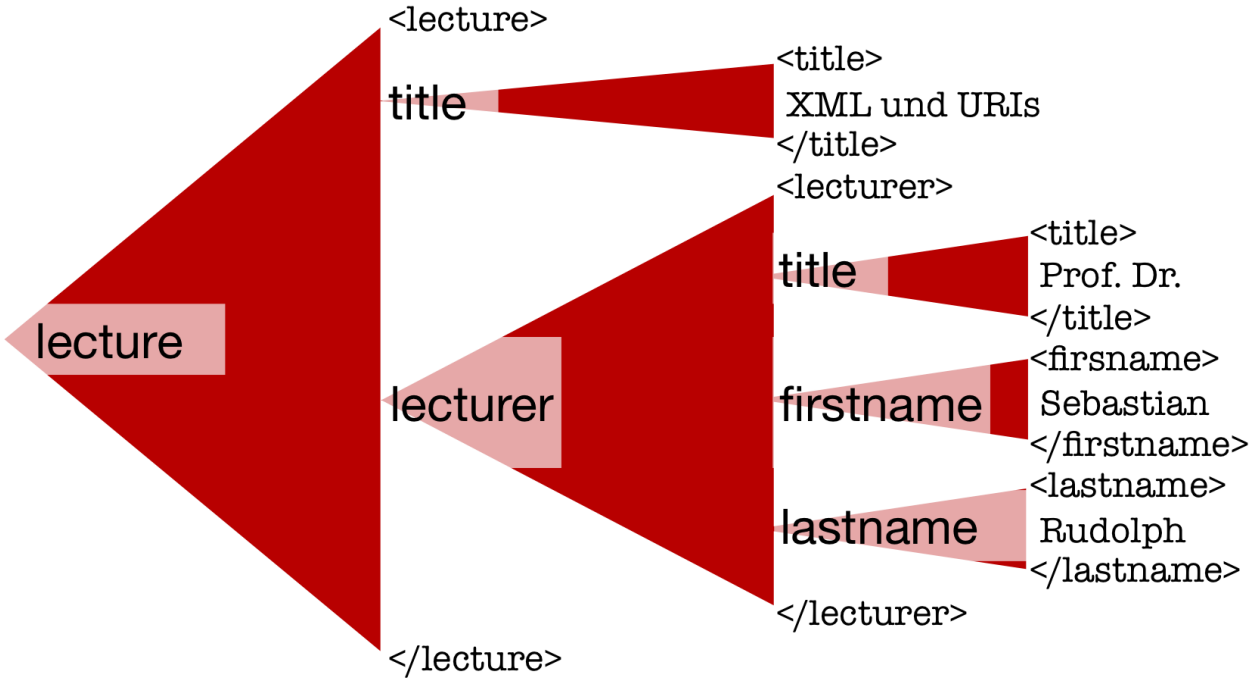
- nesting of tags is permitted

Annotation with Markup-Languages

```
<lecture>
    <title>
      Deduction Systems
    </title>
    <lecturer>
      <title>
        Prof. Dr.
      </title>
      <firstname>
        Sebastian
      </firstname>
      <lastname>
        Rudolph
      </lastname>
    </lecturer>
</lecture>
```

- nesting of tags is permitted
- multiple usage of tags is permitted

Annotation with Markup-Languages



- nesting of tags is permitted
- multiple usage of tags is permitted
- XML tags constitute a tree structure

Agenda

- XML – Motivation/Idea
- XML – Syntax
- IRIs
- Name Spaces

XML

- eXtensible Markup Language
- Origin: structured text (HTML4.0 $\in$ XML $\subset$ SGML)
- web standard (W3C) for data exchange:
  - input and output data can be described by means of XML
  - industry only has to agree on standardized tag names (the vocabulary)
- complementary language for HTML:
  - HTML describes presentation
  - XML describes content
- database perspective: XML as a data model for semi-structured data

XML-Syntax – Preamble

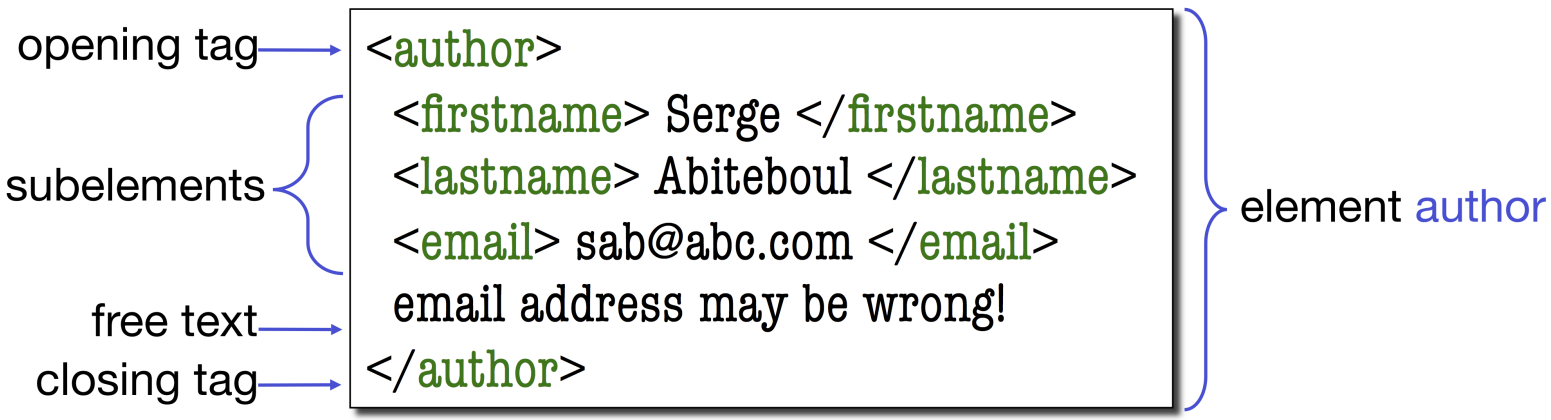
- XML document is a text document
- starts with declaration
  - contains version number of the standard used
  - optional: character encoding information

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

XML-Syntax – XML element

XML element:

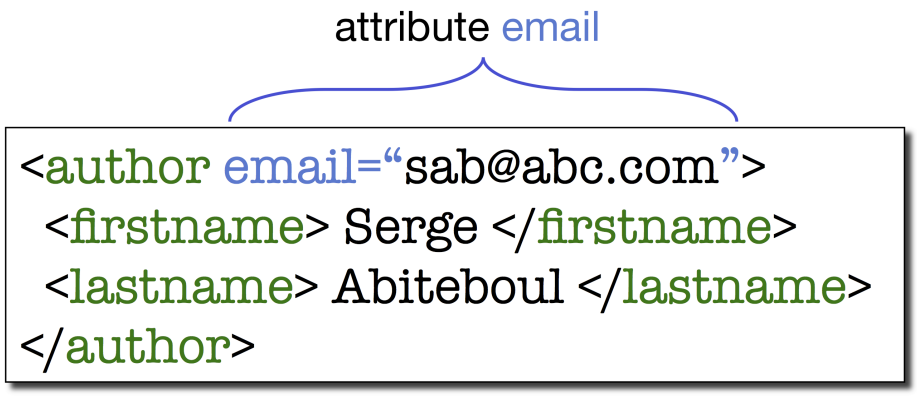
- description of an object enclosed by matching tags
- content of an elements: text and/or further elements (arbitrary nesting possible)
- empty elements: `<year></year>` short: `<year/>`
- “outermost” element is called root element (and there can be only one per document)



XML-Syntax – XML attributes

XML attribute:

- pair of name and string-value in start or self-closing tag
- associated with one XML element
- alternative option for describing data



Further possible description of the same data:

```
<author firstname="Serge" lastname="Abiteboul" email="sab@abc.com"/>
```

HTML vs. XML

- HTML: fixed vocabulary (set of tags) and semantics (visual presentation of text)
- XML: free choice of names for describing application-specific syntax and semantics
- XML $\subset$ SGML

```
<h1> Bib </h1>
<p>
   <i> Foundations of Databases </i>
   Serge Abiteboul
   </br> Addison Wesley, 1997
<p>
...
```

HTML

```
<Bib id="o1">
   <paper id="o12">
      <title> Foundations of Databases </title>
      <author>
         <firstname> Serge </firstname>
         <lastname> Abiteboul </lastname>
      </author>
      <year> 1997 </year>
      <publisher> Addison Wesley </publisher>
   </paper>
   ...
</Bib>
```

XML

Agenda

- XML – Motivation/Idea
- XML – Syntax
- IRIs
- Name Spaces

IRIs – Idea

- IRI = Internationalized Resource Identifier
- serve for denoting resources in a world-wide unique way
- a resource can be any object that has (in the context of a given application) a clear identity (e.g. books, cities, persons, publishers, relations between those, abstract concepts etc.)
- in certain domains, something similar already exists: ISBN number for books

IRIs – Syntax

- extension of the notion of URLs; not every IRI relates to a Web document but mostly a Web document is referred to by using its URL as IRI
- starts with the so-called IRI schema, which is separated by a colon (:) from the subsequent part (e.g.: http, ftp, mailto)
- IRIs often hierarchically structured

IRIs – Self-defined IRIs

- necessary, if for a certain resource no IRI exists or is known (yet)
- strategy in order to avoid unintentional double use of an IRI for different things: use http-IRIs of a webpage that you control
- allows for providing a documentation describing the IRI under this address

The Describing vs. the described

- Separation of IRIs for (non-information) resources and their documentation (information resources) by IRI references (appended fragments starting with “#”) or content negotiation
- e.g.: as a IRI for Shakespeare’s “Othello”, http://de.wikipedia.org/wiki/Othello should not be used, but rather http://de.wikipedia.org/wiki/Othello#IRI

Agenda

- XML – Motivation/Idea
- XML – Syntax
- IRIs
- Name Spaces

XML Name Spaces: Motivation

- in XML documents, element and attribute names (“markup vocabulary”) have universal validity
- in an XML application, these names are interpreted uniformly
- if XML data from several sources is merged, name conflicts / clashes may occur
- name spaces help avoid such conflicts

XML Name Spaces

- XML name spaces are similar to the notion of modules in programming languages
- disambiguation of tag names through usage of different “prefixes”
- a prefix is separated from a local name by a colon (:), thereby `prefix:name` tags come into being
- name space bindings are ignored by some tools: socalled “shallow name spaces”

Name Space Bindings

- prefixes are associated with name space IRIs by inserting an attribute `xmlns:prefix` into the relevant element or some of its predecessor elements: `prefix:name`$_1$, ..., `prefix:name`$_n$
- the value of the attribute `xmlns:prefix` is an IRI, that may point to a description of the syntax of the name space
- an element can use bindings for several (different) name spaces by using separate attributes `xmlns:prefix`$_1$, ..., `xmlns:prefix`$_m$

Example: Without Name Spaces

```
<lecture>
    <title> Deduction Systems </title>
    <lecturer>
        <title> Prof. Dr. </title>
        <firstname> Sebastian </firstname>
        <lastname> Rudolph </lastname>
    </lecturer>
</lecture>
```

`title` is an ambiguous element name

Two Distinct Name Spaces

```
<lec:lecture  xmlns:lec = "http://www.example.org/lectures"
              xmlns:per = "http://www.example.org/person">
    <lec:title> Deduction Systems </lec:title>
    <lec:lecturer>
        <per:title> Prof. Dr. </per:title>
        <per:firstname> Sebastian </per:firstname>
        <per:lastname> Rudolph </per:lastname>
    </lec:lecturer>
</lec:lecture>
```

`title` has been disambiguated by using the prefixes `lec` and `per`

Agenda

- XML – Motivation/Idea
- XML – Syntax
- IRIs
- Name Spaces