XML *Extensible Markup Language*

DR WIOLETA SZWOCH

DEPARTMENT OF INTELLIGENT INTERACTIVE SYSTEMS

Agenda

XML fundamentals

XML document

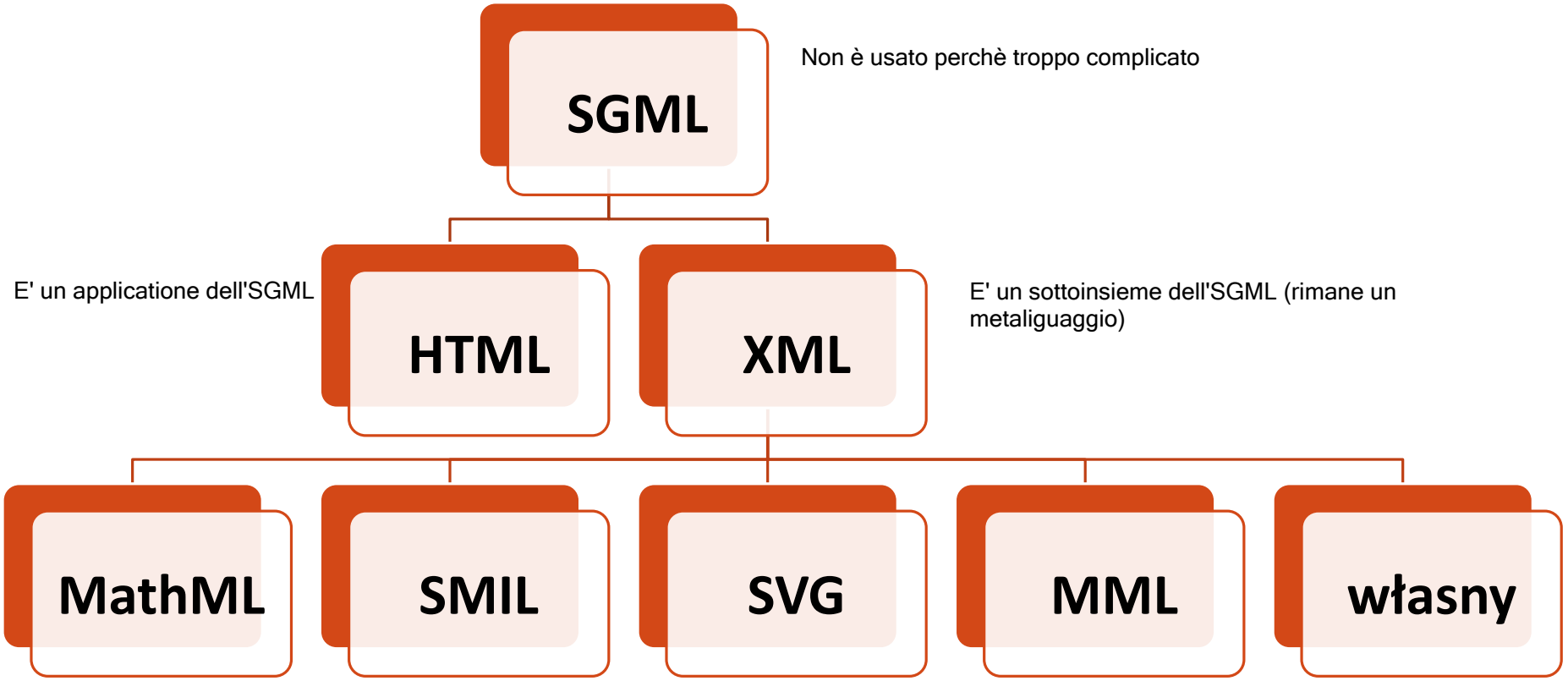
Markup Language

GML

- 1969 E.Mosher, R.Lorie, Ch. F. Goldfarb
- first markup language

Standard General Markup Language (SGML)

- ISO standard of exchange and storage of the data (1986)
- separating form (how it looks) from content (what it is)
- tags do not determine style of presentation
- set of tags is defined in separate declaration - DTD



Are languages to create XML documents and are instances of XML meta languages

HTML	vs	XML
web description language		meta-language for defining languages
the meaning of tags and attributes is predetermined		the meaning of tags and attributes is determined by the user or application
instance of SGML		subset of SGML

XML - eXtensible Markup Language

Extensible – tags using in documents are designed by the user (document schema designer)

Language – the document is built according to syntax rules, and based on a defined alphabet - tags

Markup – the document is built on tags

Non sappiamo con interprentar i numeri: l'HTML non interpreta i numeri

Why XML?

Problems with HTML

- the set of tags is strictly (rigorosamente) defined
- is used to describe the way of presentation not data
- what does the code mean?

<td> 3472096</td>

```
<p>Wioleta Szwoch
<br/> Department of Intelligent
Interactive Systems <br/> 3472096
```

HTML code

Wioleta Szwoch
Department of Inteligent Interactive Systems
3472096

processing by the browser

Is the algorithm able to interpret this code?

QUi è molto più semplice

```
< employee>
 <name> Wioleta Szwoch</name>
 <department>
        Department of Inteligent Interactive Systems
 </ department >
 <phone>3472096</phone>
</ employee>
```

code processing by the browser?

Wioleta Szwoch

Department of Inteligent Interactive Systems

phone: 347 20 96

Wioleta Szwoch
Department of Inteligent Interactive Systems
phone: 347-20-96

Wioleta Szwoch
Department of Inteligent Interactive Systems
3472096

XML is not …

A replacement for HTML

- but HTML can be generated from XML

A presentation format

- but XML can be converted into one

A programming language

- but it can be used with almost any language

A network transfer protocol

- but XML may be transferred over a network

A database

- but XML may be stored into a database

But then – what is it?

XML is a meta markup language for text documents / textual data

XML allows to define languages („applications“) to represent text documents / textual data

```
< employee>
    <name> Wioleta Szwoch</name>
    <department>
        Department of Inteligent Interactive Systems
    </ department >
    <phone> 3472096 </phone>
</ employee>
```

- Easy to understand for human users
- Very expressive (semantics along with the data)
- Well structured, easy to read and write from programs

```
< a1>
   <b20> Wioleta Szwoch</b20>
   <cX>
        Department of Inteligent Interactive Systems
   </ cX>
   <tX> 3472096 </tX>
</ a1>
```

- Hard to understand for human users
- Not expressive (semantics along with the data)
- Well structured, easy to read and write from programs

< **data**>
12n23ng3h4j5ats7899ducyytt6n33n3hdgsgc237675
</ **data**>

- Impossible to understand for human users
- Not expressive (no semantics along with the data)
- Unstructured, read and write only with special programs

XML Document

The XML document can optionally have an XML declaration.

- XML version
- the character encoding used in the document
  - UTF-8; UTF-16; ISO-8859-1; ISO-8859-2; windows-1250
- processing external DTD

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes"?>
```

```
<person>
   Some information about
   <FirstName>Anna</FirstName>
   <LastName>Kowalska</LastName>
</person>
```

XML documents

What‘s in an XML document?

elements

attributes

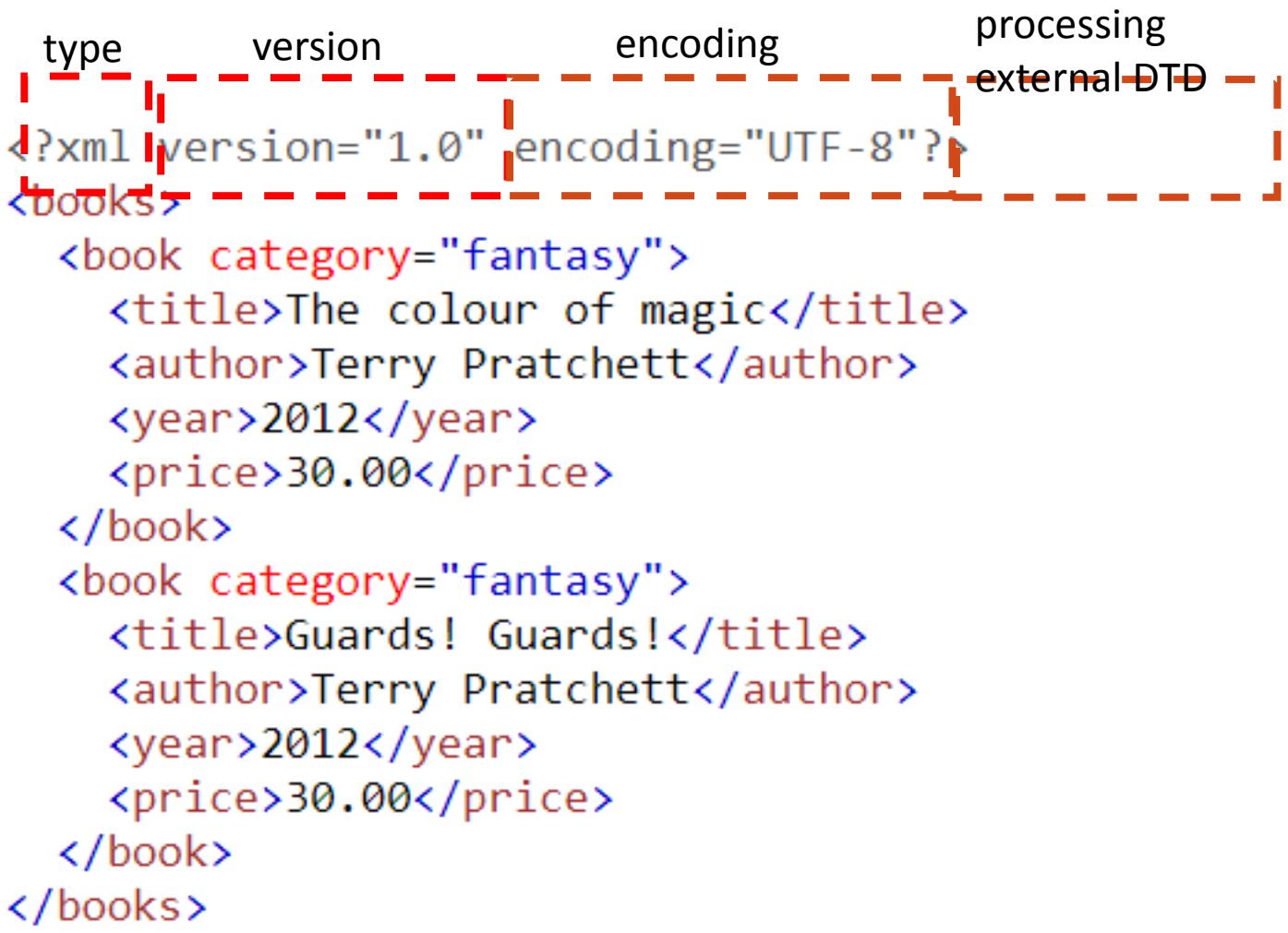
declaration

entity

processing instructions

XML declaration (FACOLTATIVA)

```
<?xml version="1.0" encoding="UTF-8"?>
<books>
  <book category="fantasy">
    <title>The colour of magic</title>
    <author>Terry Pratchett</author>
    <year>2012</year>
    <price>30.00</price>
  </book>
  <book category="fantasy">
    <title>Guards! Guards!</title>
    <author>Terry Pratchett</author>
    <year>2012</year>
    <price>30.00</price>
  </book>
</books>
```


type
version
encoding
processing
external DTD
<?xml version="1.0" encoding="UTF-8"?>
<books>
<book category="fantasy">
<title>The colour of magic</title>
<author>Terry Pratchett</author>
<year>2012</year>
<price>30.00</price>
</book>
<book category="fantasy">
<title>Guards! Guards!</title>
<author>Terry Pratchett</author>
<year>2012</year>
<price>30.00</price>
</book>
</books>

root (parent of the other elements):
tutto è contenuto all'interno

element content

```
<?xml version="1.0" encoding="UTF-8"?>
<books>
  <book category="fantasy">
    <title>The colour of magic</title>
    <author>Terry Pratchett</author>
    <year>2012</year>
    <price>30.00</price>
  </book>
  <book category="fantasy">
    <title>Guards! Guards!</title>
    <author>Terry Pratchett</author>
    <year>2012</year>
    <price>30.00</price>
  </book>
</books>
```

element
(contenitore di informazioni)

attribute: part of element that contains more information about the element

attribute content

Qualsiasi nome va bene

XML - element

Basic concept - defines the meaning of the content of information

<name> Kowalski Jan </name>

Element names are case-sensitive

<name > Kowalski Jan </Name >

The tag name can contain:

- letters, numbers
- - . (not the beginning)
- _ :

<NAME></NAME>
<Name></NAme>
<name ></ name >

_name.i:.1-2

XML - elements

An element can contain:

- text
- attributes
- other elements
- or a mix of the above

Empty element

```
<person>
   Definition for a person
   <name>Anna</name>
   <surname >Kowalska</surname>
    with mixed value
</person>
```

```
<link href='http://www.wp.pl'> </link>
```

empty element tag

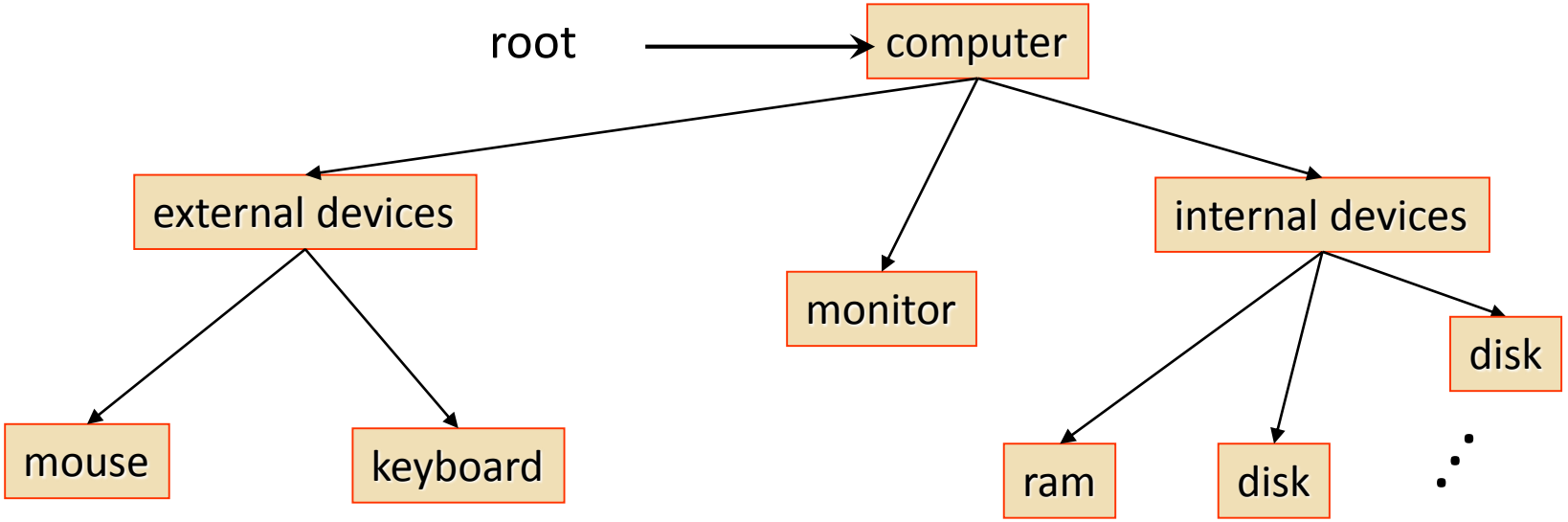
```
<link href='http://www.wp.pl' />
```

attribute

Sono corrette entrambe le versioni

Solo gerarchia, nulla da segnalare

XML – elements hierarchy




```
<computer>
  <monitor>
    LCD
  </computer>
</monitor>
```

```
<computer>
  <monitor>
    LCD
  </monitor>
</computer>
```

XML - attribute

any order of attributes

qualsiasi ordine di attributi

In one element **only different** attribute **names**

In un elemento solo nomi di attributi diversi

limitations on the name of the attribute as on the element

limitazioni sul nome dell'attributo come sull' elemento

attributes have no structure, **simply strings**
(while elements can have subelements)

gli attributi non hanno struttura, semplicemente stringhe (mentre gli elementi possono avere sottoelementi)

```
<surname maiden_name =”no”> Kowalski </surname>
```

~~<name which=”first” which=”second”> Jan </name>~~

XML - declarations

```
<!
<!
[
```

```
<message>
        Bad nesting <comp> <mon>  LCD  </comp> </mon>
</message>
```

```
<!– an example of comment and CDATA section -->
<message >
  <![CDATA[Bad nesting <comp> <mon>  LCD  </comp> </mon>]]>
</message >
```

basic declaration

- comments <!–- comment content -->
- CDATA <![CDATA[text]]>

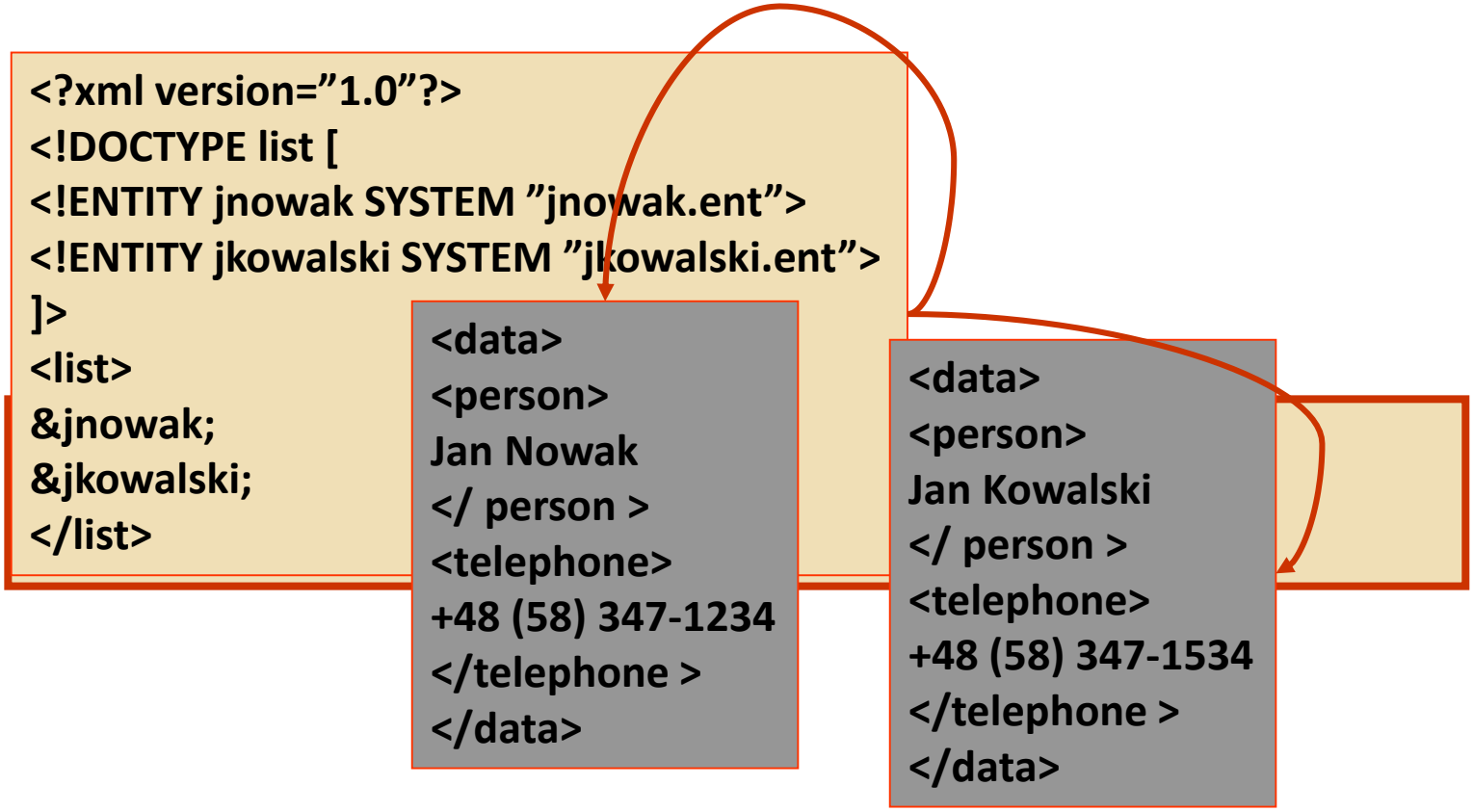
declaration used in DTD

- <!DOCTYPE ...> <!ATTLIST …> …

Caracter Data: you let to know to the parser that a particolar contains no markup and shoud be treated as a regular text

XML - entity

dividing the document into smaller parts



XML - processing instructions

allow documents to contain instructions for applications

<? target instructions?>

can be used to pass the information to applications (rarely used)

they are mostly used to link XML documents to a stylesheet so such process i will you on class and project.
Let's the parser run a piece of code

```
<?xml-stylesheet type="text/xsl" href="journey.xslt"?>
```

```
<?php
  mysql_connect(”base.pl”, ”kowalski”, ”password”);
  mysql_close()
?>
```


26

XML – a good style

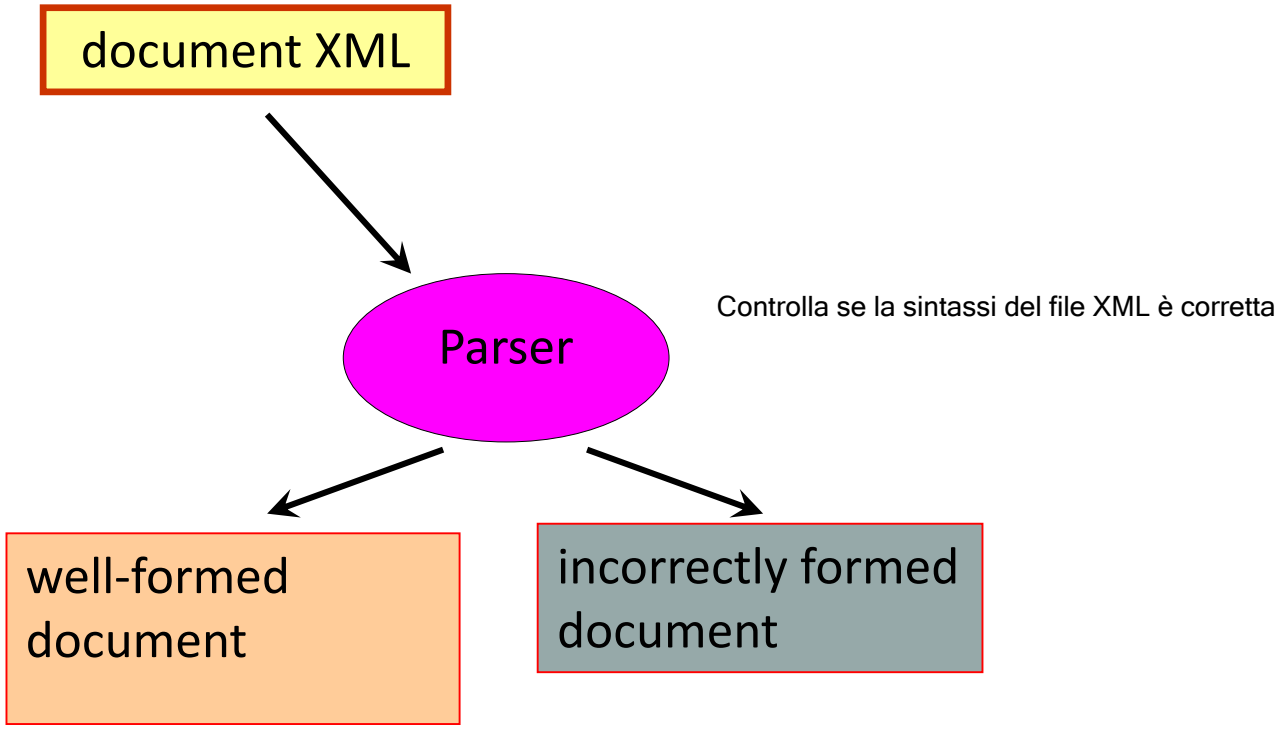
the appearance of the document

```
<?xml version="1.0" encoding="UTF-8"?>
<books>
  <book category="fantasy">
    <title>The colour of magic</title>
    <author>Terry Pratchett</author>
    <year>2012</year>
    <price>30.00</price>
  </book>
  <book category="fantasy">
    <title>Guards! Guards!</title>
    <author>Terry Pratchett</author>
    <year>2012</year>
    <price>30.00</price>
  </book>
</books>
```


27

Correctness of the document

well-formed document



Correctness of the document

WELL-FORMED DOCUMENT

There must be exactly one root element

Every start-tag has a matching end-tag

Attribute values must be quoted

Elements may nest, but must not overlap

…

XML (*Extensible Markup Language*)

10 II 1998 r XML 1.0

meta language

- a set of rules on the basis of which other languages can be created

extensible, formalized

designed to store and transport data

XML (*Extensible Markup Language*)

open standard

not licensed

text language

flexible

- flexible document structure

self-describing

contextual

- the ability to store data with their meaning

XML (*Extensible Markup Language*)

platform independent

hierarchical

separate form (how it looks) from content (what it is)

the possibility of a varied presentation

- XSLT, FO

modular

Uses of XML

Web pages

- WML, XHTML
- greater formalization of information, its processing

XML can be used to exchange the information between organizations and systems

Description of resources

- information retrieval, resource descriptors
  - RDF (*Resource Description Framework*), OWL (*Web Ontology Language*)
  - WSDL (*Web Services Description Language*)

Representation of semi-structural information

Databases - intermediate layer

Guarda quanto è complesso scrivere una semplice formula con MathXML

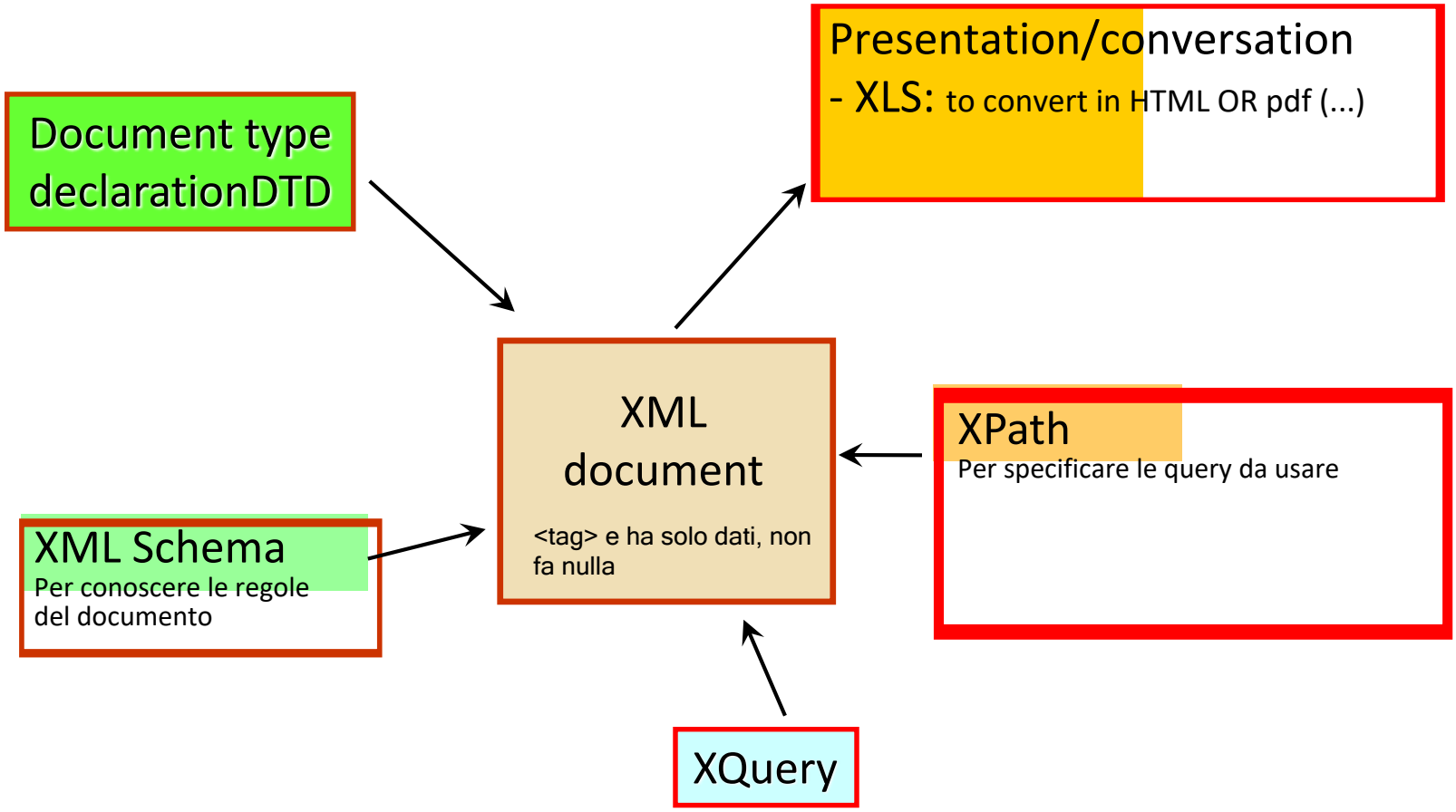
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

```
<math mode="display" xmlns="http://www.w3.org/1998/Math/MathML">
 <semantics>
  <mrow>
    <mi>x</mi>
    <mo>=</mo>
    <mfrac>
      <mrow>
        <mo form="prefix">&#x2212;<!-- − --></mo>
        <mi>b</mi>
        <mo>&#x00B1;<!-- &PlusMinus; --></mo>
        <msqrt>
          <msup>
            <mi>b</mi>
            <mn>2</mn>
          </msup>
          <mo>&#x2212;<!-- − --></mo>
          <mn>4</mn>
          <mo>&#x2062;<!-- &InvisibleTimes; --></mo>
          <mi>a</mi>
          <mo>&#x2062;<!-- &InvisibleTimes; --></mo>
          <mi>c</mi>
        </msqrt>
      </mrow>
      <mrow>
        <mn>2</mn>
        <mo>&#x2062;<!-- &InvisibleTimes; --></mo>
        <mi>a</mi>
      </mrow>
    </mfrac>
  </mrow>
  <annotation encoding="TeX">
     x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}
  </annotation>
  <annotation encoding="StarMath 5.0">
     x={-b plusminus sqrt {b^2 - 4 ac}} over {2 a}
  </annotation>
 </semantics>
</math>
```

Che sarebbe cortissima in altri linguaggi

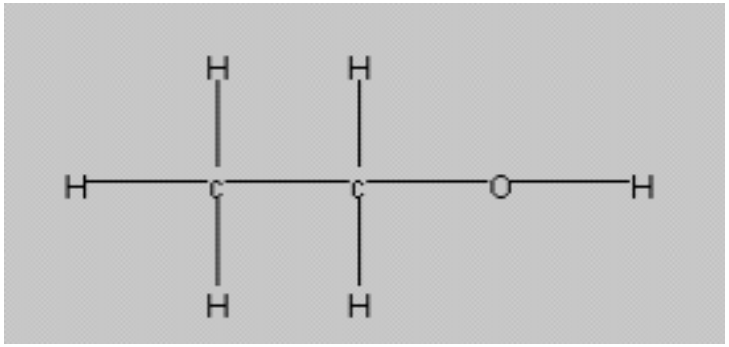
```
x=\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}
```

XML document



Esempio dello stardard di XML per descrivere le formule chimiche (nelle slides dopo ci sono altri esempi

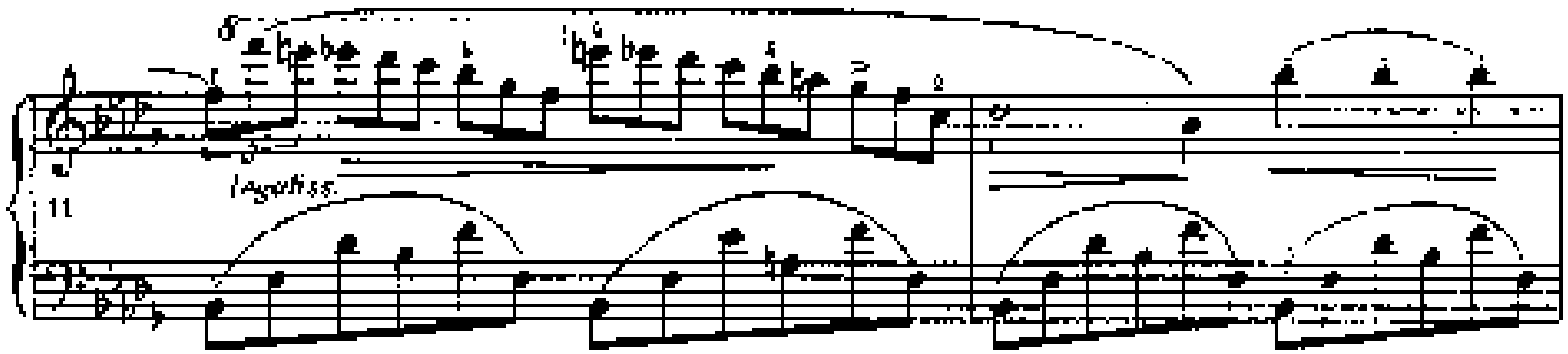
```
<CML:molecule id="test">
  <CML:atomArray builtin="elsym">
    c c O H H H H H H</CML:atomArray>
  <CML:atomArray builtin="x2" type="float">
    0 0 0 0 0 -2 -2 2 2</CML:atomArray>
  <CML:atomArray builtin="y2" type="float">
    0 2 4 -2 6 0 2 0 2</CML:atomArray>
  <CML:bondArray builtin="atid1">
    1 1 1 1 2 2 2 3</CML:bondArray>
  <CML:bondArray builtin="atid2">
    2 4 6 8 7 9 3 5</CML:bondArray>
  <CML:bondArray builtin="order" type="integer">
    1 1 1 1 1 1 1 1
  </CML:bondArray>
</CML:molecule>
```



```
<?xml version="1.0"?>
<math xmlns="http://www.w3.org/1998/Math/MathML"
      style="color: #000000; font-family: times">
  <mrow>
    <mi>z</mi>
    <mo>=</mo>
    <mfrac>
      <mrow>
        <mn>2</mn>
        <mo>&InvisibleTimes;</mo>
        <msub>
          <mi>M</mi>
          <mn>i</mn>
        </msub>
      </mrow>
      <msqrt>
        <mrow>
          <msup>
            <mi>x</mi>
            <mn>3</mn>
          </msup>
          <mo>+</mo>
          <mn>5</mn>
          <mo>-</mo>
          <mi>y</mi>
        </mrow>
      </msqrt>
    </mfrac>
  </mrow>
</math>
```

$$z = \frac{2M_i}{\sqrt{x^3 + 5 - y}}$$

```
<?xml version="1.0" standalone="no"?>
<svg width="100%" height="100%" xmlns='http://www.w3.org/2000/svg'>
  <desc></desc>
  <g id="polygonMask" style="stroke: none;">
    <polygon style="fill: #0192BF;" points="30,60 30,230 108,208"
             stroke="1"/>
    <polygon style="fill: #9EA4A6;" points="30,60  108,208  125,155"
             stroke="1"/>
    <polygon style="fill: #89A5C7;" points="30,60  125,155  144,161"
             stroke="1"/>
    <polygon style="fill: #488B9B;" points="30,60  144,161  138,115"
             stroke="1"/>
    <polygon style="fill: #9EA4A6;" points="30,60  138,115  180,114"
             stroke="1"/>
    <polygon style="fill: #EC1608;" points="30,60  180,114  160,60"
             stroke="1"/>
</g>
</svg>
```




```
<bar barid="treb-11"> F 5(3F En Eb D C Eb D C B) G
                                    5(En Eb D C B An G F C)
        <notation>
                <slurend end="slur11" beat="3" />
        </notation>
</bar>
<bar barid="treb-12"> D:2 D (5B:2)3
</bar>
```

Summary

XML

- **meta language = we can create another languages using XML**
- **separate form (appearance) from content = il modo di presentare è diviso dai dati. XML è solo dati**
- text
- expandable
- flexible
- self-describing
- contextual
- open standard