Grupo de Especificação e Processamento de Linguagens

XCSL: XML Constraint Specification Language

Marta Jacinto marta.jacinto@itij.mj.pt
Giovani Librelotto grl@di.uminho.pt
José Carlos Ramalho jcr@di.uminho.pt
Pedro Rangel Henriques prh@di.uminho.pt

Departamento de Informática, Univ. Minho, Braga, Portugal

Motivation (1)

- Kinds of Semantic Constraints -

We want to contrain XML documents in the following situations:

- Case 1: Domain Range checking
  The price of a CD is less than 32€
- Case 2: Dependencies between two elements/attributes
  When a noun is singular, the verb should be singular too
- Case 3: Matching against a Regular Expression
  A telephone number follows a certain format
- Case 4: Complex constraints
  A key should be unique in a certain table (for which it is the primary key), but may occur any number of times in every other table of the DB

Motivation (2)

- None of the above is specifiable with DTDs
- Sometimes we just want to constrain certain parts of a document

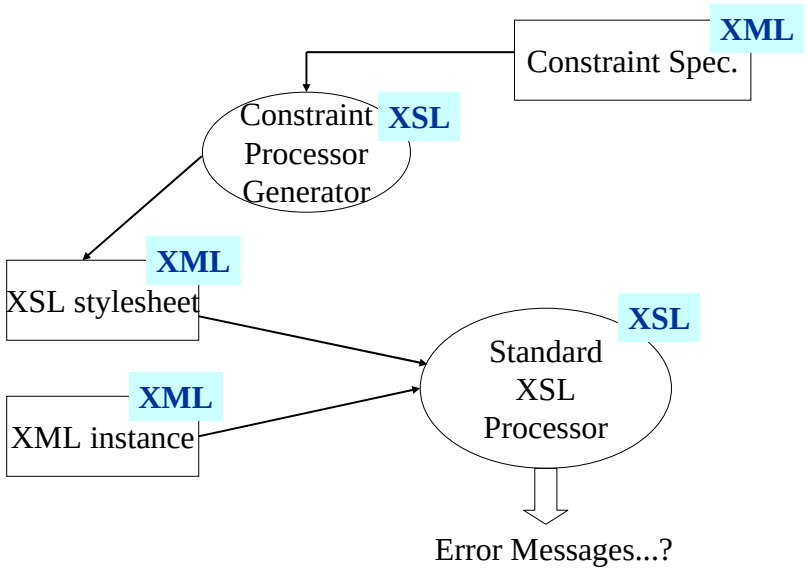
We propose

XCSL - Validation System
- Language

Outline

- Architecture
  - Validation System
  - Language
- Case Studies
  - Fiscal Certificate
  - 2nd Conference for a Divorce
- Skeletons (templates for XCSL constraints)
- Conclusion

Architecture (1)



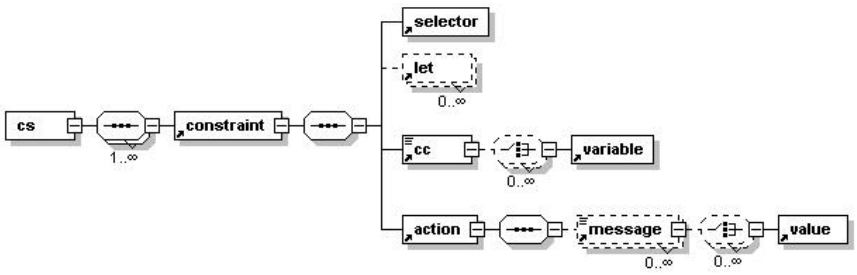
Architecture (2)

- Requirements -

- The solution should work with existing tools and standards
- The solution should work across every platform

XCSL

- Language (Constraint Specification) -



Case-Study 1 – Fiscal Certificate (1)

- What is it?
- Problems it raises:
  - Dates
  - Department
  - Cardinality/order of mixed content elements' sub-elements

Case-Study 1 – Fiscal Certificate (2)

- DTD:

```
<!ELEMENT fcert (header, body, ending)>
<!ELEMENT header (#PCDATA | department)*>
<!ELEMENT department (#PCDATA)>
<!ATTLIST department
   place CDATA "0101">
<!ELEMENT body (requester, request)>
<!ELEMENT requester (#PCDATA | name | CF | address)*>
<!ELEMENT name (#PCDATA)>
<!ELEMENT CF (#PCDATA)>
<!ELEMENT address (#PCDATA)>
<!ELEMENT request (#PCDATA | affinity | name | date | village |
parish | municipality)*>
<!ELEMENT affinity (#PCDATA)>
<!ELEMENT date (#PCDATA)>
<!ATTLIST date
   value CDATA "19000101">
<!ELEMENT village (#PCDATA)>
<!ELEMENT parish (#PCDATA)>
<!ATTLIST parish
   place CDATA "010101">
<!ELEMENT municipality (#PCDATA)>
<!ATTLIST municipality
   place CDATA "0101">
<!ELEMENT ending (#PCDATA | place | date)*>
<!ELEMENT place (#PCDATA)>
```


Case-Study 1 – Fiscal Certificate (3)

- XML:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE fcert SYSTEM "fcert_cm.dtd">
<fcert>
   <header>
      Dear Sir, Chief of the Finance Department of
      <department place="110504">Lisbon's 4th Fiscal Parish</department>
   </header>
   <body>
      <requester>
         <name>Rita Santos </name>
         taxpayer Ner.
         <CF>31988455</CF>
         with the address
         <address>Pedras tortas Street, Ner 7 - 5423 Ranholas
       </address>
      </requester>
      <request>
         requests your Excellency to certify if, on behalf of the death of her
         ...
         <name>Francelestina Pereira e Santos</name>
         who died on the
         <date value="19990913">13th of September 1999</date>
         ...
```

Case-Study 1 – Fiscal Certificate (3a)

- XML:

```
            parish of
            <parish place="100611">Salir de Matos</parish>
            municipality of
            <municipality place="1006">Caldas da Rainha</municipality>
            and maried she was with
            ...
        </request>
    </body>
    <ending>
        Ask that her request be granted
        <place>Caldas da Rainha</place>
        <date value="19991020">20th of October 1999</date>
        The requester
    </ending>
</fcert>
```


Case-Study 1 – Fiscal Certificate (4)

- Problems it raises:
  - Dates
  - Department
  - Mixed Content

Case-Study 1 – Fiscal Certificate (4a)

- XML:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE fcert SYSTEM "fcert_cm.dtd">
<fcert>
   ...
   <body>
      ...
      <request>
         requests your Excellency to certify if, on behalf of the death of her
         ...
         <name>Francelestina Pereira e Santos</name>
         who died on the
         <date value="20010803">3rd of August 2001</date>
         ...
      </request>
   </body>
   <ending>
      Ask that her request be granted
      <place>Caldas da Rainha</place>
      <date value="20010607“>7th of June 2001</date>
      The requester
   </ending>
</fcert>
```


Case-Study 1 – Fiscal Certificate (4b)

- XCSL restriction:

```
<constraint>
   <selector selexp="//request/date"/>
   <cc>
      @value < /fcert/ending/date/@value
   </cc>
   <action>
      <message>
                  The date of the death pointed out:
         <value selexp="/fcert/body/request/date"/>, is
         posterior to the request date:
         <value selexp="/fcert/ending/date"/>
         </message>
   </action>
</constraint>
```

Case-Study 1 – Fiscal Certificate (4c)

First attribute value - 20010803
Second one - 20010607

- XCSL error output:

```
<err-message>
        The date of the death pointed out: 3rd of August 2001,
        is posterior to the request date:  7th of June 2001
</err-message>
```


Case-Study 1 – Fiscal Certificate (5)

- Problems it raises:
  - Dates
  - Department
  - Mixed Content

Case-Study 1 – Fiscal Certificate (5a)

- XML:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE fcert SYSTEM "fcert_cm.dtd">
<fcert>
  <header>
    Dear Sir, Chief of the Finance Department of
    <department place="110504">Lisbon's 4th Fiscal Parish</department>
  </header>
  <body>
    ...
    <request>
      ...
      parish of
      <parish place="100611">Salir de Matos</parish>
      municipality of
      <municipality place="1006">Caldas da Rainha</municipality>
      and maried she was with
      ...
    </request>
  </body>
  ...
</fcert>
```


Case-Study 1 – Fiscal Certificate (5b)

- XCSL restriction:

```
<constraint>
  <selector selexp="//fcert/body/request"/>
  <cc>
     parish/@place = /fcert/header/department/@place
    or
    municipality/@place = /fcert/header/department/@place
   </cc>
  <action>
    <message>
       The request for this certificate shall not be delivered in this department
       <value selexp="/fcert/header/department"/>, but in the
          department in charge of the <value selexp="parish"/>'s parish,
       <value selexp="municipality"/>'s municipality.
    </message>
  </action>
</constraint>
```

Case-Study 1 – Fiscal Certificate (5c)

department/@place=110504
parish/@place=100611
municipality/@place=1006

- XCSL error output:

```
<err-message>
    The request for this certificate shall not be delivered in this department
    Lisbon's 4th Fiscal Parish, but in the department in charge of the
    Salir de Matos's parish, Caldas da Rainha's municipality.
</err-message>
```


Case-Study 1 – Fiscal Certificate (6)

- Problems it raises:
  - Dates
  - Department
  - Mixed Content (requester element)

Case-Study 1 – Fiscal Certificate (6a)

- XML:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE fcert SYSTEM "fcert_cm.dtd">
<fcert>
   ...
   <body>
      <requester>
         <name>Rita Santos </name>
         taxpayer Ner.
         <CF>31988455</CF>
         with the address
         <address>Pedras tortas Street, Ner 7 - 5423 Ranholas
        </address>
      </requester>
      ...
   </body>
</fcert>
```


Case-Study 1 – Fiscal Certificate (6b)

- XCSL restriction:

```
<constraint>
   <selector selexp="//fcert/body/requester"/>
   <cc>
      (count(name) = 1) and
      (count(CF) = 1) and
      (count(address) = 1) and
      name(name[1]/following::*) ='CF' and
      name(CF[1]/following::*) ='address'
   </cc>
   <action>
      <message>
            Either -requester- sub-elements occur in a wrong order,
            either  they occur a wrong number of times.
      </message>
   </action>
</constraint>
```

Case-Study 1 – Fiscal Certificate (6c)

If the XML instance had two name elements

- XCSL error output:

```
<err-message>
        Either -requester- sub-elements occur in a wrong order,
    either  they occur a wrong number of times.
</err-message>
```


Case-Study 2 – 2nd Conference for a Divorce (1)

- What is it?
- Problems it raises:
  - Days since the first petition

Case-Study 2 – 2nd Conference for a Divorce (2)

- DTD:

```
<!ELEMENT div_2c (header, body, ending)>
<!ELEMENT header (sender, addressee)>
<!ELEMENT sender (#PCDATA | cdepart)*>
<!ELEMENT cdepart (#PCDATA)>
<!ELEMENT addressee (#PCDATA | court)*>
<!ELEMENT court (#PCDATA)>
<!ELEMENT body (requesters, request)>
<!ELEMENT requesters (#PCDATA | name)*>
<!ELEMENT name (#PCDATA)>
<!ELEMENT request (#PCDATA | date | article)*>
<!ELEMENT date (#PCDATA)>
<!ATTLIST date
   value CDATA "19000101"    >
<!ELEMENT article (#PCDATA)>
<!ELEMENT ending (text, place, date, signature, signature)>
<!ELEMENT place (#PCDATA)>
<!ELEMENT signature (#PCDATA)>
<!ELEMENT text (#PCDATA)>
```


Case-Study 2 – 2nd Conference for a Divorce (3)

- XML:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE div_2c SYSTEM "div_2c02.dtd">
<div_2c>
 <header>
  ...
 </header>
 <body>
  ...
  <request>
   identified in the referred Action of Divorce official papers,
   having accomplished the first conference in the
   <date value="20010406">6th of April of 2001</date>
   and both maintaining their will to divorce, come,
   ...
  </request>
 </body>
 <ending>
    <date value="20010506">6th of May of 2001</date>
  ...
 </ending>
</div_2c>
```

Case-Study 2 – 2nd Conference for a Divorce (4)

- Problems it raises:
  - Days since the first petition

Case-Study 2 – 2nd Conference for a Divorce (4a)

- XML:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE div_2c SYSTEM "div_2c02.dtd">
<div_2c>
 <header>
  ...
 </header>
 <body>
  ...
  <request>
   identified in the referred Action of Divorce official papers,
   having accomplished the first conference in the
   <date value="20010406">6th of April of 2001</date>
   and both maintaining their will to divorce, come,
   ...
  </request>
 </body>
 <ending>
   <date value="20010506">6th of May of 2001</date>
  ...
 </ending>
</div_2c>
```

Case-Study 2 – 2nd Conference for a Divorce (4b)

XCSL restriction:

```
<constraint>
  <selector selexp="//div_2c"/>
  <LET NAME="a" value="(floor((14-substring(ending/date/@value,5,2)) div 12))"/>
  <LET NAME="y" value="(substring(ending/date/@value,1,4)  + 4800 - $a)"/>
  <LET NAME="m" value="(substring(ending/date/@value,5,2) + 12 * $a - 3)"/>
  <LET NAME="t" value="(substring(ending/date/@value,7,2)
                + floor((153 * $m + 2) div 5) +
                (365 * $y) + floor($y div 4) -
                floor($y div 100) +
                floor($y div 400) - 32045)"/>
   ...
  <cc>
    ($t - $t2) >= 90
  </cc>
   <action>
     <message lang="en">
       Only <value selexp="($t - $t2)"/> days undergone since the first conference...
       You will have to wait a little longer!!
     </message>
     <message lang="pt">
       Só passaram <value selexp="($t - $t2)"/> dias desde a primeira conferência...
       Têm que esperar mais algum tempo!!
     </message>
   </action>
</constraint>
```


Case-Study 2 – 2nd Conference for a Divorce (4c)

First attribute value - 20010406
Second one - 200105606

- XCSL error output (no information provided - default is lang="en"):

```
<err-message>
  Only 30 days undergone
  since the first conference...
  You will have to wait a little longer!!
</err-message>
```

- XCSL error output (specifying lang="pt"):

```
<err-message>
  Só passaram 30 dias desde a
  primeira conferência...
  Têm que esperar mais algum tempo!!
</err-message>
```

- XCSL error output (specifying lang="all"):

```
<err-message>
  Only 30 days undergone
  since the first conference...
  You will have to wait a little longer!!
</err-message>
<err-message>
  Só passaram 30 dias desde a
  primeira conferência...
  Têm que esperar mais algum tempo!!
</err-message>
```

Skeletons - Templates

By analysing the cases we dealt with

(the ones we presented here and a lot more),

we can present the following skeletons (templates)

one for each kind of semantic constraint

Skeletons – Templates (Case 1)

- Domain range checking

```
<constraint>
   <selector selexp="path to the element"/>
   <cc>
     . | @attname < value
    </cc>
   <action>
     <message>
         Message...
         <value selexp="path to any element/attribute |
         any expression applied to any element/attribute"/>
     </message>
   </action>
</constraint>
```

Skeletons – Templates (Case 2)

- Dependencies between two elements/attributes

```
<constraint>
   <selector selexp="path to the 1st element"/>
   <cc>
     . | @attname < path to the 2nd element/[ . | @attname]
    </cc>
   <action>
     <message>
          Message...
          <value selexp="path to any element/attribute |
          any expression applied to any element/attribute"/>
     </message>
   </action>
</constraint>
```


Skeletons – Templates (Case 3)

- Pattern Matching against a Regular Expression

```
<constraint>
   <selector selexp="path to the element"/>
   <cc>
      substring(.| @attname,i,n1)=literal_value and
      (string-length(number(substring(.| @attname,j,n2))) = value
   </cc>
   <action>
      <message>
         Message...
         <value selexp="path to any element/attribute |
         any expression applied to any element/attribute"/>
      </message>
   </action>
</constraint>
```

Values like:
Literal_value value_digits

Skeletons – Templates (Case 4)

- Complex constraints –mixed content

```
<constraint>
   <selector selexp="path to the parent element"/>
   <cc>
      (count(elt1)=c_elt1) and (count(elt2)=c_elt2) and ... (count(eltn)=c_eltn) and
      name(elo1[1]/following::*)='elo2' and
      name(elo2[1|2]/following::*)='elo3' and
         ...
   </cc>
   <action>
      <message>
         Message...
         <value selexp="path to any element/attribute |
         any expression applied to any element/attribute"/>
      </message>
   </action>
</constraint>
```

ATTENTION
elo2 may differ from *elt2* !!!

Skeletons – Templates (Case 4')

- Complex constraints – unicity problem

```
<constraint>
  <selector selexp="path to X branch"/>
  <LET NAME="nameKey1" value="elementX | @attributeX"/>
  <cc>
    (count(path to Y branch[elementY | @attributeY = $nameKey1]) = 1)
  </cc>
  <action>
   <message>
      Message...
      element | @attribute:
      <value selexp="$nameKey1"/>.
   </message>
  </action>
</constraint>
```

Every value of *element | @attribute* that appears in the X branch exists in the Y branch.

Conclusion (1)

- Comparison between similar approaches -

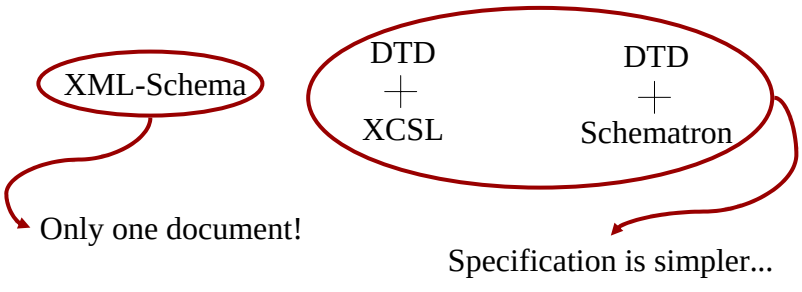
Constraint Language / Kind of constraint	XCSL	Schematron	XML-Schemas
1- Domain Range checking	X	X	X
2- Dependencies between two elements/attributes	X	X	
3- Pattern Matching against a Regular Expression	X	X	X
4- Complex Constraints (mixed content)	X	X	X
4'- Complex constraints (unicity problem)	X	X	

Easier

No personalized output

Conclusion (2)

- XCSL is simpler than Schematron
- XML-Schema does not substitute the other two approaches
- When a constraint is specifiable with XML-Schema, there are three options:



Conclusion (3)

- XML language
- Processed in XSL
- Allows the specification of any restriction

Do it simple and with existing technology!

Grupo de Especificação e Processamento de Linguagens

For more information:

http://www.di.uminho.pt/~jcr/PROJS/xcsl-www/

Marta Jacinto	marta.jacinto@itij.mj.pt
Giovani Librelotto	grl@di.uminho.pt
José Carlos Ramalho	jcr@di.uminho.pt
Pedro Rangel Henriques	prh@di.uminho.pt

Departamento de Informática, Univ. Minho, Braga, Portugal