

Relaxed—on the Way Towards True Validation of Compound Documents

Petr Nálevka

University of Economics, Prague
*Dept. of Information and Knowledge
Engineering*

petr@nalevka.com

Jirka Kosek

University of Economics, Prague
*Dept. of Information and Knowledge
Engineering*

jirka@kosek.cz



Relaxed—on the Way Towards True Validation of Compound Documents

- Agenda
 - Benefits of validation
 - What is Relaxed
 - Limitations of current validation approaches
 - RELAX NG + Schematron
 - Comparison of Relaxed with W3C validator
 - Support for compound documents
 - NVDL, JNVDL and compound documents
 - Who is using Relaxed

Benefits of validation

- Ideal world
 - All browsers are implementing Web standards
 - All authors create pages according to Web standards
 - All pages work in all browsers, interoperability is reached
- How to reach ideal world
 - Web standards promotion
 - Conformance testing
 - Many aspects of standard compliance can be automatically tested – validated

What is Relaxed

- HTML and XHTML validation service
 - Web-based user interface for people
 - Web service interface for machines
- Set of XHTML schemas
 - Schemas can validate more than DTDs which are provided as a part of corresponding W3C recommendation
 - Powerful schema languages RELAX NG and Schematron are used to overcome DTD limitations

Weaknesses of DTD validation

- Weak data types support
 - Cannot express HTML datatypes
 - e. g. colors, lengths, multi-lengths, integers, date & time, URIs, ...
- No namespace support
 - Unable to validate compound documents
- Unable to express complex structural relationships
 - No rule-based validation
- W3C Markup Validation Service is DTD based and thus it suffers from all problems mentioned above

The power of RELAX NG and Schematron

- Advantages

- Ability to validate compound documents
- Optional restriction level thanks to straightforward modularity support
- Full expressive power of XPath and regular expressions
- Complex structural relationship constraints (Schematron rules)
- Standardized technology (ISO and OASIS standards)

- Disadvantages

- SGML/HTML 4.01 must be converted to well-formed XML before validation

```

<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head><title>W3C validator limitations demo</title></head>
  <body>
    <h1>Datatypes</h1>
    <table border="10%">
      <tbody>
        <tr>
          <td>A</td>
          <td><font color="Ivory">B</font></td>
        </tr>
      </tbody>
    </table>

    <h1>Nested forms</h1>
    <form name="form2" action="process.form">
      <div>
        <form action="process.subform">
          <p>Something is wrong</p>
        </form>
      </div>
    </form>

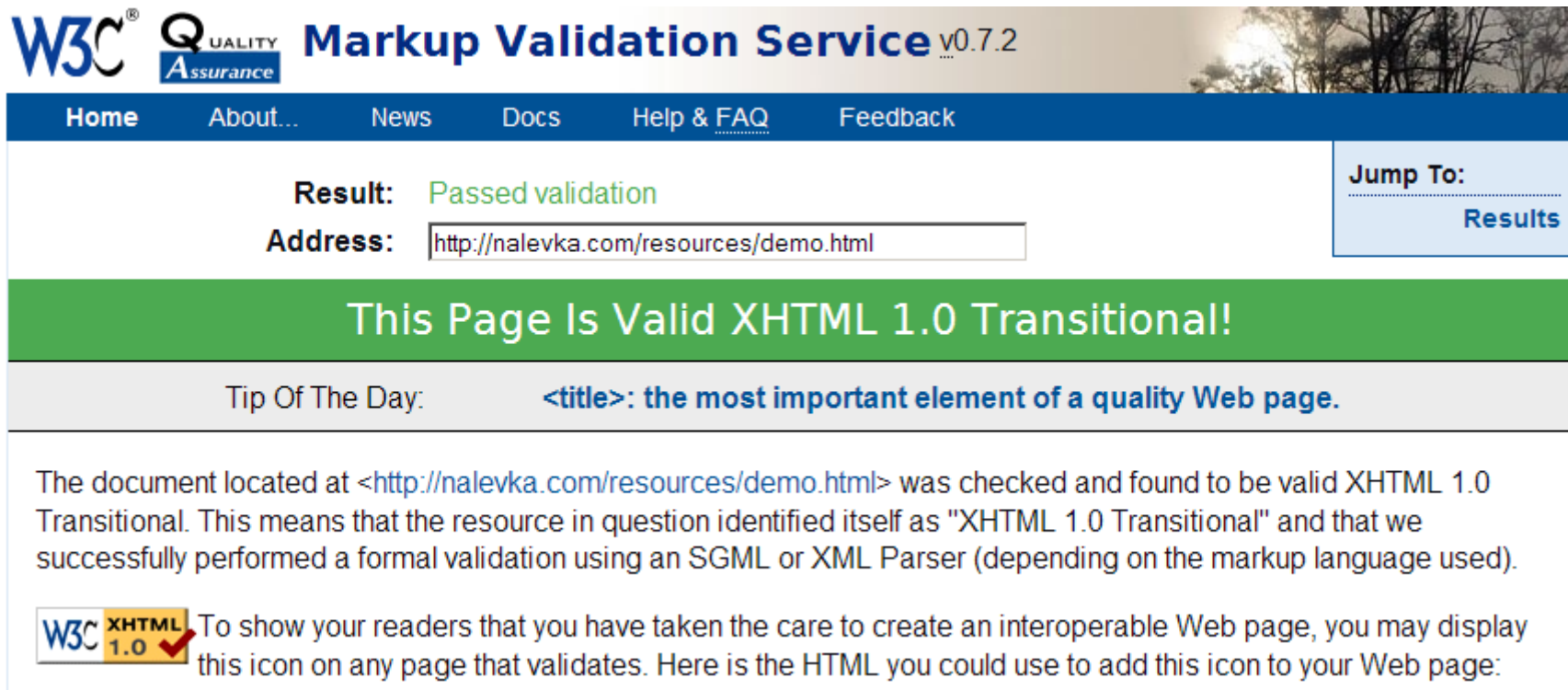
    <h1>NAME and ID inconsistency</h1>
    <form name="form1" id="form2" action="process.form">
      <p>Something is wrong</p>
    </form>
    <a name="form2">Something is wrong</a>
  </body>
</html>

```

Relaxed beats W3C validator



Relaxed beats W3C validator



W3C QUALITY Assurance **Markup Validation Service** v0.7.2

Home About... News Docs Help & FAQ Feedback

Result: Passed validation


Address:

Jump To:
[Results](#)

This Page Is Valid XHTML 1.0 Transitional!

Tip Of The Day: **<title>: the most important element of a quality Web page.**

The document located at <http://nalevka.com/resources/demo.html> was checked and found to be valid XHTML 1.0 Transitional. This means that the resource in question identified itself as "XHTML 1.0 Transitional" and that we successfully performed a formal validation using an SGML or XML Parser (depending on the markup language used).

 To show your readers that you have taken the care to create an interoperable Web page, you may display this icon on any page that validates. Here is the HTML you could use to add this icon to your Web page:

Petr Nálevka, Jirka Kosek – WWW2006, 25th May 2006, Edinburgh, Scotland
Relaxed – on the Way Towards True Validation of Compound Documents




```

<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head><title>W3C validator weaknesses demo</title></head>
  <body>
    <h1>Datatypes</h1>
    <table border="10%">
      <tbody>
        <tr>
          <td>A</td>
          <td><font color="Ivory">B</font></td>
        </tr>
      </tbody>
    </table>

    <h1>Nested forms</h1>
    <form action="process.form">
      <div>
        <form action="process.subform">
          <p>Somethinkg's wrong</p>
        </form>
      </div>
    </form>

    <h1>NAME and ID consistency</h1>
    <form name="form1" id="form2" action="process.form">
      <p>Somethinkg's wrong</p>
    </form>
    <a name="form2">Something is wrong</a>
  </body>
</html>

```

Specification violations

Relaxed beats
W3C validator



Relaxed
the HTML validator

www.relaxed.cz

RL
http://nalevka.com/resources/demo.html

Restriction

W3C XHTML1.0/HTML4.01

Doctype

Autodetect

view source brief output

dirty parser

Validate

Relaxed beats W3C validator

Your document is invalid.

INFO Detected document type: -//W3C//DTD XHTML 1.0 Transitional//EN

ERROR Line number 9 Column 25 attribute "border" has a bad value: "10%" does not satisfy the "nonNegativeInteger" type

```
... <table border="10%">
```

ERROR Line number 13 Column 28 attribute "color" has a bad value: the value does not match the regular expression "[bB][lL][aA][cC][kK][gG][rR][eE][eE][nN][sS][iI][lL][vV][e ... A][aA][qQ][uU][aA]#[0-9A-Fa-f]{3}|#[0-9A-Fa-f]{6}".

```
<td><font color="Ivory">B</font></td>
```

ERROR Line number 28 Column 0 line 28 : Id and name attribute (if present both) needs to have the same value for element form (" form2 " != " form1 ").

```
<form name="form1" id="form2" action="process.form">
```

ERROR Line number 28 Column 0 line 28 : The id and name attributes share the same name space. This means that they cannot both define an anchor with the same name in the same document. The id attribute (with value " form2 ") of element form collides with the a name attribute of some following element.

```
<form name="form1" id="form2" action="process.form">
```

ERROR Line number 19 Column 0 line 19 : form element cannot have any nested form elements

```
<form action="process.form">
```

XML



Relaxed
the HTML validator

www.relaxed.cz

RELAX NG

Example of datatype modelling

```
<!-- Color: Black, Green, Silver, Lime, Gray, Olive, White, Yellow,
      Maroon, Navy, Red, Blue, Purple, Teal, Fuchsia, Aqua, #custom -->
<define name="Color.datatype">
  <data type="string">
    <param name="pattern">[bB][lL][aA][cC][kK]|[gG][rR][eE][eE][nN] | ...
      ... [aA][qQ][uU][aA] | #[0-9A-Fa-f]{3}| #[0-9A-Fa-f]{6}</param>
  </data>
</define>
```

```
<!-- Pixels: a pixel is restricted to a non-negative integer. -->
<define name="Pixels.datatype">
  <data type="nonNegativeInteger">
    <param name="pattern">[0-9]+</param>
  </data>
</define>
```



Modelling complex relationships using Schematron

```
<sch:rule context="html:*">  
  <sch:report test="string-length(@id) > 0 and  
    ((preceding::html*/@name = @id) or  
     (following::html*/@name = @id))">  
    The id and name attributes share the same namespace, they shall not  
    collide.  
  </sch:report>  
</sch:rule>
```

```
<sch:rule context="html:form">  
  <sch:report test="descendant::html:form">  
    Forms cannot have any nested forms.  
  </sch:report>  
</sch:rule>
```

What does Relaxed validate

- W3C Recommendations
 - HTML 4.01, XHTML 1.0 – Strict/Transitional/Frameset
 - Widely used in real world
 - WCAG 1.0 (partial)
- Compound documents
 - Arbitrary foreign elements and attributes are allowed/disallowed
 - XHTML1.0 + SVG1.1
 - XHTML1.0 + MathML2.0
 - XHTML1.0 + MathML2.0 + SVG1.1

W3C validator does not support compound documents

Validation results for XHTML page with embedded SVG

W3C QUALITY Assurance **Markup Validation Service** v0.7.2

Home About... News Docs Help & FAQ Feedback

Result: Failed validation, 25 errors

Address:

Encoding: utf-8

Doctype: XHTML 1.0 Transitional

Root Namespace: <http://www.w3.org/1999/xhtml>

Other Namespaces ♦ <http://www.w3.org/2000/svg>

Jump To:
[Results](#)

This page is **not Valid XHTML 1.0 Transitional!**

Below are the results of checking this document for XML well-formedness and validity.

- Error** Line 15 column 15: **there is no attribute "xmlns"**.

```
<svg xmlns="http://www.w3.org/2000/svg" width="300" height="200">
```

You have used the attribute named above in your document, but the document type you are using does not support that attribute for this element. This error is often caused by incorrect use of the "Strict" document type with a document that uses frames (e.g. you must use the "Transitional" document type to get the "target" attribute), or by using vendor proprietary extensions such as "marginheight" (this is usually fixed by using CSS to achieve the desired effect instead).
- Error** Line 15 column 68: **element "svg" undefined**.

```
...p://www.w3.org/2000/svg" width="300" height="200" >
```

Petr Nálevka, Jirka Kosek – WWW2006, 25th May 2006, Edinburgh, Scotland
Relaxed – on the Way Towards True Validation of Compound Documents

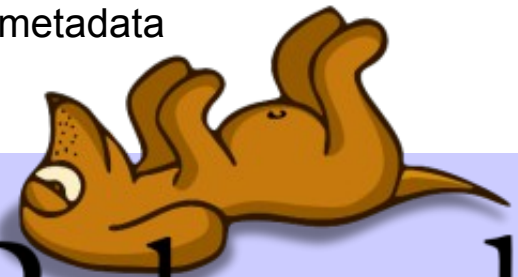
Relaxed
the HTML validator

www.relaxed.cz

Compound documents

- Documents combining more XML grammars together
 - There are many XML languages whose combination can bring a real value-added (rich-client, web-design, semantic queries...)
 - Already supported in some browsers
 - e. g. SVG+XHTML in Firefox and Opera

SVG	SMIL	MathML	EGIX	presentation
VoiceXML	XForms	XHTML	XLink	rich-client
RDF	RSS	XTM	vCard in XML	metadata



NVDL (ISO/IEC 19757-4)

- NVDL = Namespace-based Validation and Dispatching Language
- International standard for compound document validation
- Advantages
 - Validator transparent
 - NVDL engine distributes XML fragments from particular namespace to appropriate validators
 - Schema language neutral
 - Different schema languages can be combined (W3C XML Schema, RELAX NG, ...)
 - Real life schemas are written in many different languages
 - Standardized and flexible way for expressing which grammars may be used in particular context

Relaxed and compound documents

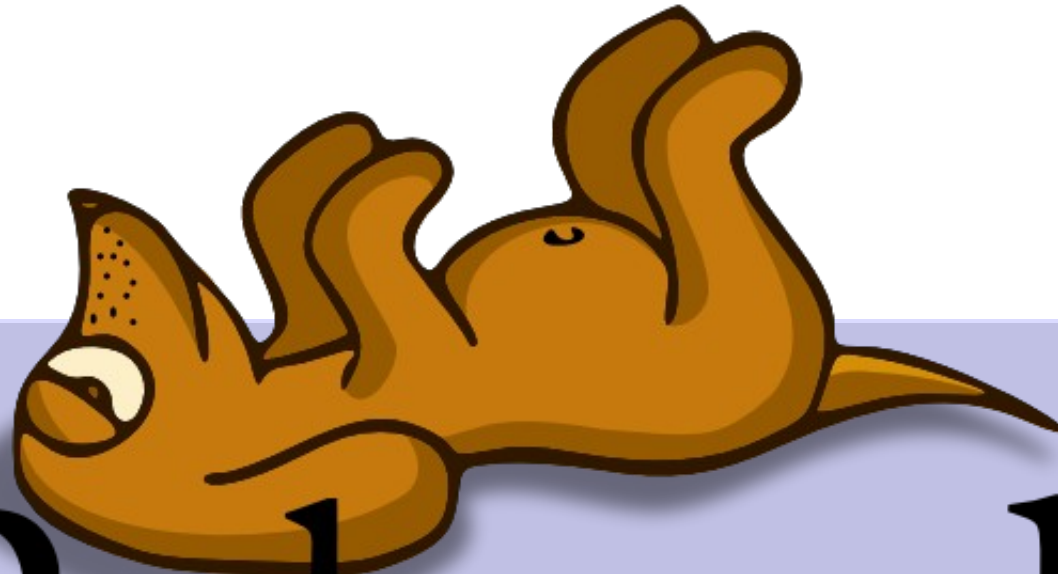
- Today
 - Compound document validation using RELAX NG
 - Special compound document schema must be created manually for each combination of mark-up vocabularies
 - All schemas must be converted to RELAX NG prior combining
- Future
 - JNVDL (<http://sourceforge.net/projects/jnvd1>)
 - Java-based NVDL implementation
 - Straightforward compound document validation

Relaxed in use

- On-line service for HTML document authors
 - Available at <http://relaxed.cz>
 - Better outputs, compound document support...
- European Internet Accessibility Observatory
 - Uses Relaxed validation engine
 - “The EIAO project will establish the technical basis for a European Internet Accessibility Observatory. Frequently updated assessment data will be available online from a data warehouse providing a basis for benchmarking, policymaking, research and actions to develop accessibility to Internet.”
- Henri Sivonen's validation service
 - Available at <http://hsivonen.iki.fi/validator/>
 - Uses Relaxed XHTML and WCAG schemas



Thank you for your attention



Relaxed
the HTML validator www.relaxed.cz

Try <http://relaxed.cz> yourself!