

Schema-based transformation: architectural forms for the 21st Century

John Cowan

LexisNexis

Copyright 2013 by John Cowan

Licensed under the GNU GPL

(but the poems are fair use)

Why

- Traditionally we validate, transform, and validate again
 - Or skip either or both validations
- Simple transformations can and should be deduced from the schemas
- Architectural forms did this for SGML
- Nobody does this in XML

How

- Source document
- Source schema with decorations
- Target schema
- (Optional) Name of renaming attribute (aka form attribute)
- (Optional) Names of renaming attributes that we want to be ignored

Examplotron

- Schemas are Russian doll instances
- Child elements, attributes, and mixed content all model themselves (attributes are optional)
- Unbracketed text is a prototype (string, number, date); text in {braces} is an XSD type; [bracketed] text is a default value (new in Examplotron 0.8)
- Occurrence types: 0, 1, ?, *, +
- Content model definitions and references

Element renaming

Source limerick

<limerick>

<a>I sat next to the Duchess for tea:

<a>It was just as I feared it would be.

Her rumblings abdominal

Were simply phenomenal—

<a>And everyone thought it was me!

<limerick>

Limerick schema

<limerick>

<a>There once was a (person) from (place)

<a>Whose (body part) was (special case).

When (event) would occur

It would cause (him or her)

<a>To violate (law of time/space).

</limerick>

Quintet schema

```
<quintet>  
  <line>Line 1</line>  
  <line>Line 2</line>  
  <line>Line 3</line>  
  <line>Line 4</line>  
  <line>Line 5</line>  
</quintet>
```

Limerick schema: quintet

```
<limerick quintet="[quintet]">  
  <a quintet="[line]">There once ...</a>  
  <a quintet="[line]">Whose ...</a>  
  <b quintet="[line]">When ...</b>  
  <b quintet="[line]">It would ...</b>  
  <a quintet="[line]">To violate ...</a>  
</limerick>
```

Target quintet

<quintet>

<line>I sat next to the Duchess for tea:</line>

<line>It was just as I feared it would be.</line>

<line>Her rumblings abdominal</line>

<line>Were simply phenomenal—</line>

<line>And everyone thought it was me!</line>

</quintet>

Multiple targets

Limerick schema: bilingual

```
<limerick quintet="[quintet]" estrofa="[estrofa]" >  
  <a quintet="[line]" estrofa="[línea]" eg:define="line">  
    Plain text allowed here  
  </a>  
  <a eg:content="line"/>  
  <b eg:content="line"/>  
  <b eg:content="line"/>  
  <a eg:content="line"/>  
</limerick>
```

Bilingual target quintet?

<estrofa>

<línea>Había un hombre de Aztlán,</línea>

<línea>que vivía con mucho elán.</línea>

<línea>Comentó su amiga:</línea>

<línea>que Diós te bendiga</línea>

<línea>y te cuide, como dice el refrán.</línea>

</estrofa>

No such luck

<estrofa>

<línea>I sat next to the Duchess for tea:</línea>

<línea>It was just as I feared it would be.</línea>

<línea>Her rumblings abdominal</línea>

<línea>Were simply phenomenal—</línea>

<línea>And everyone thought it was me!</línea>

<estrofa>

Occurrences

Extended limerick

<limerick>

<a>There was a young man from Bengal

<a>Who had no phone manners at all.

He'd yak and he'd yak

And you'd have to lean back

<a>And position yourself by the wall,

<a>And wait for the end of the call.

</limerick>

Truncated quintet, unfortunately

<quintet>

<a>There was a young man from Bengal

<a>Who had no phone manners at all.

He'd yak and he'd yak

And you'd have to lean back

<a>And position yourself by the wall,

</quintet>

Truncated limerick

<limerick>

<a>There was an old man from Peru

<a>Whose limericks stopped at line two.

<limerick>

Revised quintet schema

<quintet>

<line eg:occurs="?">[MISSING!]</line>

<line eg:occurs="?">[MISSING!]</line>

<line eg:occurs="?">[MISSING!]</line>

<line eg:occurs="?">[MISSING!]</line>

<line eg:occurs="?">[MISSING!]</line>

</quintet>

The best we can do

<quintet>

<line>There was an old man from Peru</line>

<line>Whose limericks stopped at line two.</line>

<line>MISSING!</line>

<line>MISSING!</line>

<line>MISSING!</line>

<quintet>

Very truncated limericks

<limerick>

<a>There was a young man from Verdun,
</limerick>

<limerick>

<!-- This limerick is about the Emperor Nero,
but it's not possible to actually quote it. -->
</limerick>

Stanza schema

<stanza>

<line eg:occurs="+">Line</line>

<!-- eg:occurs allows one or more lines -->

</stanza>

Limerick schema: stanza

```
<limerick stanza="[stanza]" >  
  <a stanza="[line]">A</a>  
  <a stanza="[line]">A</a>  
  <b stanza="[line]">B</b>  
  <b stanza="[line]">B</b>  
  <a stanza="[line]">A</a>  
</limerick>
```

Target stanza

<stanza>

<line>I sat next to the Duchess for tea:</line>

<line>It was just as I feared it would be.</line>

<line>Her rumblings abdominal</line>

<line>Were simply phenomenal—</line>

<line>And everyone thought it was me!</line>

</stanza>

Reordering

Haiku target schema

```
<haiku>  
  <first>Just five syllables</first>  
  <middle>Plus seven can't say much, but</middle>  
  <last>That's haiku for you.</last>  
</haiku>
```

Reverse haiku source

<haiku>

<last>Abolishing traditions</last>

<middle>Create a new style</middle>

<first>Reflection of the old one</first>

</haiku>

Haiku target

<haiku>

<first>Reflection of the old one</first>

<middle>Create a new style</middle>

<last>Abolishing traditions</last>

</haiku>

Mixed content

Mixed content limerick

<limerick>

<a>Said a widow whose singular vice

<a>Was to keep her dead husband on ice:

<q>"It's been hard since I lost him,</q>

<q>I'll <i>never</i> defrost him.</q>

<a><q>Cold comfort, but cheap at the price."</q>

</limerick>

Mixed content limerick schema

```
<limerick>  
  <a eg:define="qandi">  
    <q>Quotes</q> and <i>italics</i>  
  </a>  
  <a eg:content="qandi"/>  
  <b eg:content="qandi"/>  
  <b eg:content="qandi"/>  
  <a eg:content="qandi"/>  
</limerick>
```

Stanza schema with emphasis

<stanza>

<line eg:occurs="+"><i>Italics</i> roman</line>

<!-- eg:occurs allows any number of lines -->

</stanza>

Mixed content stanza

<stanza>

<line>Said a widow whose singular vice</line>

<line>Was to keep her dead husband on ice:</line>

<line>"It's been hard since I lost him,</line>

<line>I'll <i>never</i> defrost him.</line>

<line>Cold comfort, but cheap at the price."</line>

</stanza>

What else?

- Unwanted character content and unwanted or unrenamable child elements are discarded
- Attribute names can be changed by placing pairs of tokens in the renaming attribute
- Character content can become an attribute or vice versa
- Tokens in attribute values can be changed
- Examplotron allows XPath assertions

XML content limerick

```
<limerick>  
  <a>XQuery ampersand DOM</a>  
  <a>Entity parser dot-com</a>  
  <b>Abstract schemata</b>  
  <b>XPointer errata</b>  
  <a>InfoSet Unicode BOM</a>  
</limerick>
```