# DHW

# An online introductory toolset for XML encoding

# **Alejandro Bia**

Miguel Hernández University XML Prague 2017

# Contents

1. THE PROBLEM

2. THE SOLUTION

3. AUTOMATIC MARKUP

4. VISUALIZATION

5. DEMO

6. CONCLUSIONS



THE PROBLEM

# THE PROBLEM



THE PROBLEM Motivation	
The idea for this project emerged, while <b>teaching XML/XSLT hands-on</b> workshops abroad.	
One of the usual problems was <b>to have the necessary tools properly installed</b> in a computer lab before the course started.	
<ul> <li>Most of the times, the instructions for preparing the tools were sent by emai resulting in incomplete or faulty set-ups.</li> </ul>	I,
We asked ourselves: ¿why not use the ubiquity of the web, and have the tools installed on a server, properly tested in advance and available from everywhere?	
Alejandro Bia DHW	3

THE PROBLEM

# Main objective

To build an **online service** with **a collection of useful text processing tools as online services**.

- For instance:
  - Tools related to XML documents: TEI, DTDs, Schemas, XSLT processing, Xpath, HTML, etc.
  - Tools to graphically visualize and design markup vocabularies and XML document instances.
  - Tools for automatic TEI markup from a Markdown-style lightweight markup language.
  - Tools to validate, pretty-print, improve and transform XML documents.
  - Tools to render XML documents

Saving the time of software installation.

Alejandro Bia

DHW 4



# What for • Teaching XML / TEI / DH methods • Support begginer's projects • Batch production through customizable pipelines • Testbed platform for new [experimental] tools

Alejandro Bia

dhw 5

# THE PROBLEM $\Delta$ **Objectives of the project Functionality** Work area with file management functionality Editing and processing tools Colaborative work environment **Teaching facilities** • Interactive examples Hands-on tools • Help Maintenance • Easy to maintain and upgrade • Scalable (both tools and users)











**File manager**: similar to the online services of Google Drive or Dropbox, and to conventional file explorers.

Allows operations like: **file upload** and **download**, **copy**, **move**, **rename**, **delete** and **selection** of files **for editing or processing**.

**Editor functionality** partially resembling Google Docs, or simple XML/XSLT editors.

Up to three files in parallel folders: input, transformation and output files.

Partial **"responsive design**": since the target of the tools are not small screens.

#### **Multilingual support**

Alejandro Bia

► Command = 'java-jar', %toolsPath%\saxon6\saxon6.jar.''. @\$param.' " \$\_SESSION["input1]. " " \$\_SESSION["input1]." " \$\_SESSION["input1]." " \$\_SESSION["input1]." " \$\_SESSION["input1]."" \*

#### • The HTML form with parameters for each service

```
<div class="checkbox">
<label>
<input type="checkbox" name="param_3" id="param_3" value="-w0">
<?php echo $s->getStringInterface($_SESSION['currentLang'], 3); ?>
</label>
</div>
```

#### • The multilingual interface and help messages

```
Saxon-6_6 => array('es' => 'Tratar errores recuperables como fatales', 'en' => 'Treat recoverable errors as fatal'),
```

#### • Inlude flags

• Indicate when to include the new commands in the dropdown menu, in the toolbar icons, etc.

**Run update-modules.xsl** to generate the PHP software to link new modules to the interface

One module may provide several services

**One service** may be provided by **several modules** (as **alternatives**, or **secuential pipelining**)

Alejandro Bia

dhw 11

THE SOLUTION

## Three types of users

• The online application allows for three types of users:

**NOT REGISTERED** (can use the online tools but cannot store documents in the cloud),

**REGISTERED** users (have their own work area where documents can be left to be used in future work sessions),

**ADMINISTRATORS** (can perform application management tasks).

Alejandro Bia

DHW 12



Bootstrap for interface rendering

Alejandro Bia

dhw 13

 $\bigtriangleup$ 

THE SOLUTION

# **Implementation: initial list of tools**

• The following is a non-exhaustive list of the services and tools included in the





AUTOMATIC MAR	KUP				
	Μ	arkdown	to TEI		
• We mana	ged to generate	XML-TEI from e	extended multimation	arkdown.	
• We've buil	t a prototype pa	arser to test the i	dea		
dtei document	body	sciidoc-like parser	proto-tei (XML)	SLT arser TEI do	cument
Aleiandro Bia					DHW 16



```
16 :name: Isabel Santiago
17 :resp: Correctora
18 :item: Revisión del formato
19
```

#### • This is the output:

```
1
            <?xml version="1.0" encoding="UTF-8"?>
            <?xml-stylesheet href="teixlite2015.css" type="text/css"?>
        2
            <!DOCTYPE TEI SYSTEM "cervantes2015.dtd">
        3
        4 □<TEI>
        5
          <teiHeader>
        6 🛓
                <fileDesc>
          þ
        7
                 <titleStmt>
        8
                   <title type="main">Carta abierta</title>
        9
                   <author>Güiraldes, Ricardo </author>
       10
                  </titleStmt>
                  <publicationStmt>
       11
          ¢
       12
                   <publisher>Biblioteca Virtual Miguel de Cervantes Saavedra</publisher>
       13
                    <pubPlace>Universidad de Alicante</pubPlace>
       14
                   <idno>004229</idno>
       15
          Ē
                   <availability status="free">
       16
                   Copyright (c) Universidad de Alicante, Banco Santander Central Hispano. Accesible
                      desde http://cervantesvirtual.com
       17
                    </availability>
       18
                    <date day="21" month="11" year="2000"/>
       19
                  </publicationStmt>
          É
                  <sourceDesc>
       21
                    Martín Fierro, segunda época, año II, núm. 14 y 15, Buenos Aires, enero 24 de 1925.
       22
            Obra cedida por la Biblioteca de la Academia Argentina de las Letras. Digitalización realizada
            por Verónica Zumárraga.
       23
                  </sourceDesc>
                </fileDesc>
       24
       25 🛱
                <profileDesc>
          H
       26
                 <langUsage>
       27
                    <language ident="es">Español</language>
       28
                 </langUsage>
       29
                </profileDesc>
       30 白
                <revisionDesc>
       31 白
                 <change>
       32
                   <date day="01" month="12" year="2008"/>
       33 白
                   <respStmt>
       34
                     <name>Lorenzo García Pérez</name>
       35
                     <resp>Supervisor</resp>
       36
                    </respStmt>
       37
                    <item>Supervisión del texto </item>
       38
                  </change>
       39 白
                  <change>
       40
                   <date day="11" month="12" year="2008"/>
       41 白
                   <respStmt>
       42
                     <name>Isabel Santiago</name>
       43
                     <resp>Supervisora</resp>
       44
                    </respStmt>
       45
                    <item>Revisión del formato</item>
       46
                  </change>
       47
                  <change>
                    <date day="06" month="01" year="2009"/>
       48
       49
                    <respStmt>
       50
                  <name>José Lorenzo Pérez</name>
                                                                                                 T
       51
                  <resp>Corrector</resp>
       52
                    </respStmt>
                                                                                                         17
Alejandro Bia
                                                                                                  DHW
```

AUTOMATIC MARKUP

# dtei example: prose

### • This is the input text:

20	# Carta abierta #
21	docAuthor: Ricardo <u>Güiraldes</u>
22	
23	Amigos:
24	
25	He leído hoy el último número de MARTÍN FIERRO. Patria chica en el papel, grande en el anhelo. Yuestra juventud sube hacia mi rostro, como un aliento de pampa, cuando sobre la gramilla iluminada de rocío (emoción de la madrugada, que vuelve a encontrar su mundo) me aferro al optimismo ascendente de los nuevos crecimientos. El hombre se siente pequeño ante la infinita transmutación que anuncia lo porvenir, pero crece con sentirse capaz de comprenderla.
26	
27	En vuestra valentía se produce una vez más el eterno amanecer del espíritu. ¿No es ése el misterio de la anunciación?
28	
29	Vienen y vendrán los ataques. Inútil sorprenderse. Caminos sin pantanos no son caminos de hombres libres y los más duros son los que más espolean el deseo de llegar. Caer no es nada. Las osamentas sirven de mojón a los que después de uno sienten el vértigo del desierto. Así se conquistan horizontes. Así se regala el bien habido a los timoratos.
31	Uno de los nuestros ha pedido piedad. Y es que dio el rostro a lo que siempre debió volver la espalda. ¿Cómo la cobardía momentánea del fuerte puede pedir ayuda a la cobardía constante de los débiles? Dice un refrán gaucho: "No hay que mudar caballo en medio 'el río'".
33	En el camino de las ideas la duda equivale a mudar de caballo.
34	THE NEW WINNING OF AND IC MEMIE NIME A HEADER OF HEADER.
35	Además,;qué puede esperar el que cargó sobre sus hombros con la responsabilidad de partir, de aquellos que se aferraron a la inmovilidad? Solamente un reproche, una acusación de soberbia, de

### • This is the output:

65	¢	<text lang="es"></text>
66	¢	<front></front>
67	¢	<titlepage></titlepage>
68	¢	<doctitle></doctitle>
69	¢	<titlepart></titlepart>
70		<title type="main">Carta abierta</title>
71	-	
72	-	
73		<docauthor>Ricardo Güiraldes</docauthor>
74	-	
75	-	
76	自	<pre><body></body></pre>
77	þ	<div></div>
78		Amigos:
79	þ	He leído hoy el último
80		<u>número</u> de MARTÍN FIERRO. <u>Patria chica en el papel, grande</u> en el
81		anhelo. Vuestra juventud sube hacia mi rostro, como un aliento de pampa, cuando
82		sobre la gramilla iluminada de rocío (emoción de la madrugada,
83		que vuelve a encontrar su mundo) me aferro al optimismo ascendente de los
84		nuevos crecimientos. El hombre se siente pequeño ante la infinita
85		transmutación que anuncia lo porvenir, pero crece con sentirse capaz de
86	F	comprenderla.
87	自	En vuestra valentía se produce una
88		vez más el eterno amanecer del espíritu. ¿No es ése
89	F	el misterio de la anunciación?
90	Ê	Vienen y vendrán los ataques.
91		Inútil sorprenderse. Caminos sin pantanos no son caminos de hombres

 $\bigtriangleup$ 



Alejandro Bia

DHW 19







- The **DTD** or **Schema is presented as a mind map beginning in a central oval node**, and from it each element definition is a subtree (root in green).
- Each subtree describes a content model, and can be folded and unfolded by clicking on its green root node.
- Sequence or choice operators that conform to the rules of the content model are represented in the middle nodes of the subtree (in yellow).
- Leaf nodes (in red) are the elements referred by the content model, and point (both graphically with curved arrows and by hyperlink) to the subtree definition of the corresponding element (elsewhere in the diagram).
- This allows for easy navigation and comprehension of the global graph structure.

Alejandro Bia

DHW 21











• Another example:







CONCLUSIONS
THE DHW PLATFORM
The online DHW platform serves to <b>integrate</b> some interesting tools (plus other not covered in this presentation) and <b>make them available</b> on the Web.
• Future work
Beta website for testing the platform:
http://kaos.umh.es/dhw/
Once finished, the project will appear at:
http://dhw.umh.es/dhw/
(now there is old information there) ☞
<ul> <li>In a future stage:</li> </ul>
<ul> <li>PoC Prototypes =&gt; Stable production</li> </ul>
Move from <b>proof-of-concept prototypes</b> to a <b>stable production</b>
version (implies rewriting some modules).

20 of 23

### • Pipelining and Batch production

The project will include the implementation of some **power production workflows** to process **whole collections of documents at once** using **several tools altogether** following customizable **pipelines**.

We may use **Xproc** pipelines and containers...

#### Collaborative environment

• Features will be enhanced

#### **Optimize client-server interaction**

- Using technologies like React, Angular or Ionic
- Cosmetic improvements
  - e.g. graphic icons

#### Alejandro Bia

DHW 26

 $\sim$ 

# **AUTOMATIC MARKUP**

- Simplifies / speeds-up the creation of XML-TEI documents.
- Now we can use a simpler, Markdown based, lightweight markup notation to obtain valid XML-TEI documents.
- Furure work...
  - Include DocBook generation from Markdown or alike

```
Alejandro Bia
```

CONCLUSIONS	
VIS	UALIZATION
<ul> <li>Advanced visualizations are complex text and metadata s</li> </ul>	usefull to teach, learn, understand, explain structures.
• Furure work	
<ul> <li>Browser visualization of F folding/unfolding</li> </ul>	reeplane/Freemind like mindmaps with
<ul> <li>Existing Freemind web obsolete</li> </ul>	visualization flash or applet plugins are now
<ul> <li>Use other web based diag</li> </ul>	graming tool
Generate SVG	
<ul> <li>Generate static mindmap</li> </ul>	images (not the purpose)
Alejandro Bia	DHW 28
JUNULUSIUNS	END

