XML to visual tagging migration

Mario Essert, Mario Lončarić ¹
Marijana Tomić ², Boris Bosančić ³

Faculty of Mechanical eng. and Naval arch., University of Zagreb ¹
Department of Librarianship, University of Zadar ²
Faculty of Arts, University of Osijek³

September 2011
1. Introduction
   - Medieval manuscripts, XML
   - Tagging, OCR, no OCR
   - Modern way
   - Level, layer

2. XML, TEI
   - Notations and notions
   - XLink, XPath
   - XSLT

3. TEIMark
   - Good morning
   - Web-based editor

4. DocMark
   - Web-based system
   - Visual tags, measurements, analysis

5. HI - History integrator
   - Time + space + information = history
Presentation and workshop

The digital side of document research

- An emphasis on medieval manuscripts
- From already known (XML, TEI) to new jQuery/WebGL technology (TEIMark program)
- The new world of digital visualization - program DocMark
- WEB-integration of time, place and info - program History integrator
Presentation and workshop

The digital side of document research

- An emphasis on medieval manuscripts
- From already known (XML, TEI) to new jQuery/WebGL technology (TEIMark program)
- The new world of digital visualization - program DocMark
- WEB-integration of time, place and info - program History integrator
Presentation and workshop

The digital side of document research

- An emphasis on medieval manuscripts
- From already known (XML, TEI) to new jQuery/WebGL technology (TEIMark program)
- The new world of digital visualization - program DocMark
- WEB-integration of time, place and info - program History integrator
Presentation and workshop

The digital side of document research

- An emphasis on medieval manuscripts
- From already known (XML, TEI) to new jQuery/WebGL technology (TEIMark program)
- The new world of digital visualization - program DocMark
- WEB-integration of time, place and info - program History integrator
Presentation and workshop

The digital side of document research

- An emphasis on medieval manuscripts
- From already known (XML, TEI) to new jQuery/WebGL technology (TEIMark program)
- The new world of digital visualization - program DocMark
- WEB-integration of time, place and info - program History integrator
Problems

You and 'digital' technology

- Tirelessly tagging texts
- ... all tags are there, now what???
- For handwriting there are no OCR programs
- There are hundreds or thousands of manuscript pages
Problems

You and 'digital' technology

- Tirelessly tagging texts
- ... all tags are there, now what???
- For handwriting there are no OCR programs
- There are hundreds or thousands of manuscript pages
Problems

You and 'digital' technology

- Tirelessly tagging texts
- ... all tags are there, now what???
- For handwriting there are no OCR programs
- There are hundreds or thousands of manuscript pages
Problems

- Tirelessly tagging texts
- ... all tags are there, now what???
- For handwriting there are no OCR programs
- There are hundreds or thousands of manuscript pages

You and 'digital' technology
Problems

- Tirelessly tagging texts
- ... all tags are there, now what???
- For handwriting there are no OCR programs
- There are hundreds or thousands of manuscript pages
In old times...

Are you faced with the same thing now, only in a modern way?

- All right, we will not go this harder way, let's take the easier way!!
In old times...

Are you faced with the same thing now, only in a modern way?

- All right, we will not go this harder way, let’s take the easier way!!
Text and markup

Texts are more than sequences of encoded glyphs

- They have structure
- Content
- Multiple readings

The point of markup is to make explicit to a machine what is implicit to a person!
Texts are more than sequences of encoded glyphs

- They have structure
- Content
- Multiple readings

The point of markup is to make explicit to a machine what is implicit to a person!
Text and markup

Texts are more than sequences of encoded glyphs

- They have structure
- Content
- Multiple readings

The point of markup is to make explicit to a machine what is implicit to a person!
Text and markup

Texts are more than sequences of encoded glyphs
- They have structure
- Content
- Multiple readings

The point of markup is to make explicit to a machine what is implicit to a person!
Text and markup

Texts are more than sequences of encoded glyphs

- They have structure
- Content
- Multiple readings

The point of markup is to make explicit to a machine what is implicit to a person!
Concept and Technical process

The units of information in markup structures

- **Annotation level** – referring to the conceptual level of information represented in markup
- **Annotation layer** – referring to the technical realisation of markup
- In this lecture we will speak only at the 'layer' level, which is technology - tagging
- In the workshop we will also work with 'level' marking - annotations

---

Concept and Technical process

The units of information in markup structures

- **Annotation level** – referring to the conceptual level of information represented in markup
- **Annotation layer** – referring to the technical realisation of markup
- In this lecture we will speak only at the ‘layer’ level, which is technology - tagging
- In the workshop we will also work with ‘level’ marking - annotations
Concept and Technical process

The units of information in markup structures

- **Annotation level** – referring to the conceptual level of information represented in markup
- **Annotation layer** – referring to the technical realisation of markup

- In this lecture we will speak only at the ‘layer’ level, which is technology - tagging
- In the workshop we will also work with ‘level’ marking – annotations

---

Introduction

Level, layer

Concept and Technical process

The units of information in markup structures

- **Annotation level** – referring to the conceptual level of information represented in markup
- **Annotation layer** – referring to the technical realisation of markup
- In this lecture we will speak only at the ‘layer’ level, which is technology - tagging
- In the workshop we will also work with ‘level’ marking - annotations

---

Concept and Technical process

The units of information in markup structures

- **Annotation level** – referring to the conceptual level of information represented in markup

- **Annotation layer** – referring to the technical realisation of markup

In this lecture we will speak only at the ‘layer’ level, which is technology - tagging

In the workshop we will also work with ‘level’ marking - annotations

---


Essert, Tomić, Bosančić, (Faculty of Mechanical eng. and Naval arch., University of Zagreb Department of Librarianship, University of Zadar Faculty of Arts, University of Osijek)
XML and TEI

TEI- Text Encoding Initiative = XML framework

- **XML notations** – elements, attributes, processing instruction etc.
- **XML notions** – well-formed and validated
- **TEI** - a framework for the definition of more than 500 textual distinctions
- **TEI schemes** (*RELAX NG*) make those distinctions
XML and TEI

TEI- Text Encoding Initiative = XML framework

- **XML notations** – elements, attributes, processing instruction etc.
- **XML notions** – well-formed and validated
- **TEI** - a framework for the definition of more than 500 textual distinctions
- **TEI schemes** (*RELAX NG*) make those distinctions
XML and TEI

TEI- Text Encoding Initiative = XML framework

- **XML notations** – elements, attributes, processing instruction etc.
- **XML notions** – well-formed and validated
- **TEI** - a framework for the definition of more than 500 textual distinctions
- **TEI schemes** (*RELAX NG*) make those distinctions
XML and TEI

- **XML notations** – elements, attributes, processing instruction etc.
- **XML notions** – well-formed and validated
- **TEI** - a framework for the definition of more than 500 textual distinctions
- **TEI schemes** (*RELAX NG*) make those distinctions
XML and TEI

TEI- Text Encoding Initiative = XML framework

- **XML notations** – elements, attributes, processing instruction etc.
- **XML notions** – well-formed and validated
- **TEI** - a framework for the definition of more than 500 textual distinctions
- **TEI schemes** (*RELAX NG*) make those distinctions
Extensible Stylesheet Language Transformations (XSLT) is a language that lets you convert XML documents into other XML documents, into HTML documents, or into almost anything you like.\(^a\)

Transformer

XLink, then XPath, and finally XSLT

Extensible Stylesheet Language Transformations (XSLT) is a language that lets you convert XML documents into other XML documents, into HTML documents, or into almost anything you like.\(^a\)

# A brief tutorial

1.1 What is XSLT (and XSL, and XPath)?  3  
1.2 A simple XSLT stylesheet  8  
1.3 More element and attribute manipulation  13  
1.4 Summing up the tutorial  16

## 1.1 What is XSLT (and XSL, and XPath)?

Extensible Stylesheet Language Transformations (XSLT) is a language that lets you convert XML documents into other XML documents, into HTML documents, or into almost anything you like. When you specify a series of XSLT instructions for converting a class of XML documents, you do so by creating a “stylesheet,” an XML document that uses specialized XML elements and attributes that describe the
"Good morning," said the little prince.

"Why are you selling those?" asked the little prince.

"And what do I do with those fifty-three minutes?"

"As for me," said the little prince to himself, "if I had fifty-three minutes to spend as I liked, I should walk at my leisure toward a spring of fresh water."

"(The Little Prince, chapter 23)"
"Good morning," said the little prince.

"Why are you selling those?" asked the little prince.

"And what do I do with those fifty-three minutes?"

"As for me," said the little prince to himself, "if I had fifty-three minutes to spend as I liked, I should walk at my leisure toward a spring of fresh water." a

---

a(The Little Prince, chapter 23)
"Good morning," said the little prince.

"Why are you selling those?" asked the little prince.

"And what do I do with those fifty-three minutes?"

"As for me," said the little prince to himself, "if I had fifty-three minutes to spend as I liked, I should walk at my leisure toward a spring of fresh water."\(^{a}\)

\(^{a}\)(The Little Prince, chapter 23)
"Good morning," said the little prince.

"Why are you selling those?" asked the little prince.

"And what do I do with those fifty-three minutes?"

"As for me," said the little prince to himself, "if I had fifty-three minutes to spend as I liked, I should walk at my leisure toward a spring of fresh water."\(^a\)

\(^a\)(The Little Prince, chapter 23)
"Good morning," said the little prince.

"Why are you selling those?" asked the little prince.

"And what do I do with those fifty-three minutes?"

"As for me," said the little prince to himself, "if I had fifty-three minutes to spend as I liked, I should walk at my leisure toward a spring of fresh water."

\[a\]

\[a\] (The Little Prince, chapter 23)
This example shows how TinyMCE can be extended with TEI XML tags.

Select text and then select TEI tag from Styles menu.

- TinyMCE visual editor which is expanded to support TEI
- Tagging = selecting text + choosing TEI tag
- Supported nested tags
- Visual distinction between different tags
This example shows how TinyMCE can be extended with TEI XML tags. Select text and then select TEI tag from Styles menu.

- TinyMCE visual editor which is expanded to support TEI
- Tagging = selecting text + choosing TEI tag
- Supported nested tags
- Visual distinction between different tags
TEI(glagolitic)Mark: web-based editor

This example show how TinyMCE can be extended with TEI XML tags

Select text and then select TEI tag from Styles menu.

Editor

- TinyMCE visual editor which is expanded to support TEI
- Tagging = selecting text + choosing TEI tag
- Supported nested tags
- Visual distinction between different tags
TEI(glagolitic)Mark: web-based editor

This example shows how TinyMCE can be extended with TEI XML tags. Select text and then select TEI tag from Styles menu.

- TinyMCE visual editor which is expanded to support TEI
- Tagging = selecting text + choosing TEI tag
- Supported nested tags
- Visual distinction between different tags
**TEI (glagolitic)Mark: web-based editor**

This example shows how TinyMCE can be extended with TEI XML tags.

Select text and then select TEI tag from Styles menu.

---

**Editor**

- **TinyMCE** visual editor which is expanded to support TEI
- Tagging = selecting text + choosing TEI tag
- Supported nested tags
- Visual distinction between different tags
Markup the digitalized image
DocMark: WebGL annotation system

- Provides the optimal joint collaboration and research
- Ratio between independent level and layer is 1:1
- Inserting, deleting and change of tags are simple
- Measuring tools are added
- No extra informatics knowledge or programming skills are needed
- Data processing is not specific but general
DocMark: WebGL annotation system

- Provides the optimal joint collaboration and research
- Ratio between independent level and layer is 1:1
- Inserting, deleting and change of tags are simple
- Measuring tools are added
- No extra informatics knowledge or programming skills are needed
- Data processing is not specific but general
DocMark: WebGL annotation system

- Provides the optimal joint collaboration and research
- Ratio between independent level and layer is 1:1
- Inserting, deleting and change of tags are simple
- Measuring tools are added
- No extra informatics knowledge or programming skills are needed
- Data processing is not specific but general
DocMark: WebGL annotation system

- Provides the optimal joint collaboration and research
- Ratio between independent level and layer is 1:1
- Inserting, deleting and change of tags are simple
- Measuring tools are added
- No extra informatics knowledge or programming skills are needed
- Data processing is not specific but general
DocMark: WebGL annotation system

- Provides the optimal joint collaboration and research
- Ratio between independent level and layer is 1:1
- Inserting, deleting and change of tags are simple
- Measuring tools are added
- No extra informatics knowledge or programming skills are needed
- Data processing is not specific but general
DocMark: WebGL annotation system

- Provides the optimal joint collaboration and research
- Ratio between independent level and layer is 1:1
- Inserting, deleting and change of tags are simple
- Measuring tools are added
- No extra informatics knowledge or programming skills are needed
- Data processing is not specific but general
DocMark: WebGL annotation system

- Provides the optimal joint collaboration and research
- Ratio between independent level and layer is 1:1
- Inserting, deleting and change of tags are simple
- Measuring tools are added
- No extra informatics knowledge or programming skills are needed
- Data processing is not specific but general
History integrator - a new kind of visual encyclopaedia

- Integration of history (time, space and information) as a whole.
- Visual tags are set at time axis and geographical maps
- Information with multimedia (video, image, link,...) is written in the WORD like editor
- Controlled dictionary; glossary and thesaurus
- All saved details may be changed or deleted
- On-line Collaboration available
History integrator - a new kind of visual encyclopaedia

- Integration of history (time, space and information) as a whole.
- Visual tags are set at time axis and geographical maps
- Information with multimedia (video, image, link, ...) is written in the WORD like editor
- Controlled dictionary; glossary and thesaurus
- All saved details may be changed or deleted
- On-line Collaboration available
History integrator - a new kind of visual encyclopaedia

- Integration of history (time, space and information) as a whole.
- Visual tags are set at time axis and geographical maps
- Information with multimedia (video, image, link, . . . ) is written in the WORD like editor
- Controlled dictionary; glossary and thesaurus
- All saved details may be changed or deleted
- On-line Collaboration available
History integrator - a new kind of visual encyclopaedia

- Integration of history (time, space and information) as a whole.
- Visual tags are set at time axis and geographical maps.
- Information with multimedia (video, image, link,...) is written in the WORD like editor.
- Controlled dictionary; glossary and thesaurus.
- All saved details may be changed or deleted.
- On-line Collaboration available.

Essert, Tomić, Bosančić, (Faculty of Mechanical eng. and Naval arch., University of Zagreb Department of Librarianship, University of Zadar Faculty of Arts, University of Osijek)
History integrator - a new kind of visual encyclopaedia

- Integration of history (time, space and information) as a whole.
- Visual tags are set at time axis and geographical maps
- Information with multimedia (video, image, link, ...) is written in the WORD like editor
- Controlled dictionary; glossary and thesaurus
- All saved details may be changed or deleted
- On-line Collaboration available
History integrator - a new kind of visual encyclopaedia

- Integration of history (time, space and information) as a whole.
- Visual tags are set at time axis and geographical maps.
- Information with multimedia (video, image, link,...) is written in the WORD like editor.
- Controlled dictionary; glossary and thesaurus.
- All saved details may be changed or deleted.
- On-line Collaboration available.
History integrator - a new kind of visual encyclopaedia

- Integration of history (time, space and information) as a whole.
- Visual tags are set at time axis and geographical maps.
- Information with multimedia (video, image, link, ...) is written in the WORD like editor.
- Controlled dictionary; glossary and thesaurus.
- All saved details may be changed or deleted.
- On-line Collaboration available.

Essert, Tomič, Bosančić, (Faculty of Mechanical eng. and Naval arch., University of Zagreb Department of Librarianship, University of Zadar Faculty of Arts, University of Osijek)
"It is strange," I said to the little prince. "Everything is ready for use: the pulley, the bucket, the rope..."

"Do you hear?" said the little prince.

"We have wakened the well, and it is singing..."\(^a\)

\(^a\)(The Little Prince, chapter 25)
Wells of Sahara

"It is strange," I said to the little prince. "Everything is ready for use: the pulley, the bucket, the rope..."

"Do you hear?" said the little prince.

"We have wakened the well, and it is singing..."\(^a\)

\(^a\)(The Little Prince, chapter 25)
"It is strange," I said to the little prince. "Everything is ready for use: the pulley, the bucket, the rope..."

"Do you hear?" said the little prince.

"We have wakened the well, and it is singing..."\(^a\)

\(^a\)(The Little Prince, chapter 25)
"It is strange," I said to the little prince. "Everything is ready for use: the pulley, the bucket, the rope..."

"Do you hear?" said the little prince.

"We have wakened the well, and it is singing..."\(^a\)

\(^a\)(The Little Prince, chapter 25)
We have wakened the well, and it is singing

- It was "XML to visual tagging migration”
- :-) 
- Thank you very much for your attention!!

\(^a\) (I believe that for his escape he took advantage of the migration of a flock of wild birds.)
We have wakened the well, and it is singing

- It was "XML to visual tagging migration"\(^a\)
- :-)  
- Thank you very much for your attention!!

\(^a\)(I believe that for his escape he took advantage of the migration of a flock of wild birds.)
We have wakened the well, and it is singing

- It was ”XML to visual tagging migration”
- :-)
- Thank you very much for your attention!!

(a(I believe that for his escape he took advantage of the migration of a flock of wild birds.)
We have wakened the well, and it is singing

- It was "XML to visual tagging migration"\(^a\)
- :-)
- Thank you very much for your attention!!

\(^a\)(I believe that for his escape he took advantage of the migration of a flock of wild birds.)