ENCODING TEXTS FOR VISUALIZATION AND ANALYSES USING THE TEI STANDARD

William Chong
D’Linkup Pte. Ltd.
Data, for everyone
http://dlinkup.com/workshops.html
big data

digitally ready

encoded

analyzable

meaningful
ATTACK PLAN

Overview
XML
TEI

Hands-on + Content Expertise = THE FUTURE
Encoding

Annotation

Markup

Meaning
WHAT CAN BE ENCODED?

Metadata

Content

Textual Structure and Layout
Many elements

Formal method of encoding

Language

eXtensible Markup Language
THE NEED FOR A COMMON LANGUAGE

<person>Tan Quee Lan</person> is a seamstress.

Lim Teck is <name>Tan Quee Lan</name>’s husband.

Text Encoding Initiative – 1987

P5 Standards
EXAMPLES OF TEI

http://tapasproject.org/all_projects

http://tapasproject.org/proceedingsofblackpeople/files/narrative-proceedings-black-people-during-late-awful-calamity

HOW TO APPLY TEI

http://www.tei-c.org/Guidelines/P5/
Shakespeare: the first folio (1623) in electronic form

Shakespeare, William (1564-1616)

Oxford Text Archive

Freely available on a non-commercial basis.

1968

The first folio of Shakespeare, prepared by Charlton Hinman (The Norton Facsimile, 1968)
VISION & MISSION

Vision

A premier knowledge hub promoting the University’s vision as a leading global university centred in Asia

NUS Libraries aspires to be the NUS communities’ first stop when seeking information
NUS Libraries provides for the information needs of diverse communities in NUS in support of learning and research
NUS Libraries equips the NUS communities with information seeking skills to enhance learning, research and scholarly communication

Mission

To actively engage and partner the NUS community in advancing scholarship and research through innovative library services

D’Linkup Pte. Ltd.
Data, for everyone
Recently, I visited the <org>Design Centric Programme</org> (DCP) in Engineering. One of the student projects I saw was called Snowstorm. As you can see, it is a personal flying machine designed for indoor spaces. Snowstorm received extensive media attention. Our student team, led by <persName>Shawn Sim</persName>, and their faculty mentors, were invited to the <event>Founders Forum</event> in the <country>UK</country>, where they had the chance to explain their flying machine to <persName>Prince William</persName>. 
Well-formed Valid
ATTACK PLAN

Overview

XML

TEI

Hands-on + Content Expertise = THE FUTURE
eXtensible Markup Language
RULE 1: DECLARATION

ID Card
Name: Doe
Surname: John

Optional
“We’ve fallen on our feet and no mistake,” said Peter. “This is going to be perfectly splendid. That old chap will let us do anything we like.”

“I think he’s an old dear,” said Susan.

“Oh, come off it!” said Edmund, who was tired and pretending not to be tired, which always made him bad-tempered. “Don’t go on talking like that.”
RULE 3: ROOT ELEMENT

```xml
<?xml version="1.0"?>
<TEI>
  <div>
    <p>
      <s>.....<name>...</name>...</s>
    </p>
  </div>
  <div>.....<date>...</date>...</div>
  <div>.....<place>...</place>...</div>
</TEI>
```
RULE 4:
NESTING

<px <name>...</name>... </px> ☺
1. XML is case-sensitive
2. Element names must start with an alphabet or the “_”
3. May contain only alphanumerics characters (letters and digits) and “_” “-” “.”
4. The colon “:” is reserved for namespaces
1. Open Notepad
2. Save file as “test.xml”

3. Open Google Chrome
4. Drag “test.xml” into Google Chrome
5. Input xml declaration: `<?xml version="1.0"?>`
6. On the next line, write a sentence:
   My name is _______(your name), and I was born on _______(a date) at _______(a place).
7. `<TEI>` tag
8. `<p>` tag
9. `<name>`, `<date>`, `<place>` tags
10. case-sensitivity test
11. wrong nesting test
Well-formed
HAVE A BREAK,
HAVE A KIT KAT
<teiHeader>Metadata</teiHeader>

Content

Textual Structure and Layout
FOUR MAIN COMPONENTS

<feedDesc>
<encodingDesc>
<profileDesc>
<revisionDesc>
USING TEI WEBSITE

http://www.tei-c.org/Guidelines/P5/
“Request a Trial License”

OR

New Document > Choose a file template > TEI P5 > All

Open
“One Hundred Years’ History of the Chinese in Singapore: An annotated edition”
Add:
1. Author
2. Publisher

Check:
1. Well-formedness
2. Valid

D’Linkup Pte. Ltd.
Data, for everyone
<div><p>Metadata</p></div>

<div><p>Content</p></div>

<div><p>Textual Structure and Layout</p></div>
VISION & MISSION

Vision

A premier knowledge hub promoting the University’s vision as a leading global university centred in Asia

NUS Libraries aspires to be the NUS communities’ first stop when seeking information

NUS Libraries provides for the information needs of diverse communities in NUS in support of learning and research

NUS Libraries equips the NUS communities with information seeking skills to enhance learning, research and scholarly communication

Mission

To actively engage and partner the NUS community in advancing scholarship and research through innovative library services
WORKING AT OXYGEN

Copy text from: _02 Text of Chapter 2

Add:
1. Front (dedication)
2. Back (index)
3. Page breaks (pb/)
4. Line breaks (lb/)
5. Quotations (quote) p50
6. Footnotes (note)
STRUCTURAL MARKING

Markup:
1. paragraphs(<p>)
2. Sentences(<s>)
Metadata

<name>Content</name>

Textual Structure and Layout
Textual Structure and Layout
PERSONS AND NAMES

<persName>Raffles</persName>

“Major Farquhar” or “Farquhar”? → “encodingDesc”

xml:id

listPerson
COLLABORATIVE WORK

https://docs.google.com/document/d/1iHKrixiXPUQNaqHPx-cUo56t9mQdBr-5PLo-68fOfhM/edit

Share work using “03” file (listPerson)
<country>Singapore</country>

<region>Malacca</region>

<district>...</district>

<settlement>...</settlement>
<date>1999</date>

<date when="1999">1999</date>

<date when-iso="1819-06-11">11th June 1819</date>
CHECKING FOR ERRORS

1. Copy from google docs into OxygenXML Editor
2. Do validation
3. Make corrections
4. See "04" file for completed markups
BASIC VISUALIZATION

See "05" file

D’Linkup Pte. Ltd.
Data, for everyone
QUESTIONS?
BRAINSTORMING

How can TEI/XML be applied to your field?

How would you execute it?
D’LINKUP MISSION
Thank you!
HAVE A MINUTE?
HELP US IMPROVE
j.mp/RUfeedback
LIKE US ON FACEBOOK

NUSLIBRARIES
PARSING

XML → Parser step 1 → well-formed → Parser step 2 → valid 😊

not well-formed → not valid

/DTD/Schema e.g. TEI
Document Model 文件模型:
DTD, Relax NG, XML Schema

<xsl-fo>

XSLT

XPath

XQuery

XML Document

ENCODING 標記

PDF

OUTPUT 輸出

JScript

CSS

HTML

ePub

any XML

TRANSFORM 轉換

ePub