

MediaWiki TEI extension

Thomas Pellissier Tanon - Télécom Paris



+



MediaWiki?

- An open source wiki software
- Used by >25k sites including Wikipedia
- >1k extensions
- A large set of features for collaboration and consumption

Good for journals, team work, students labs?



Already used for texts (with ProofreadPage)

Editing Page:Wind in the Willows (1913).djvu/196

Insert - - {} {}{} | [] []] "" = [[Category:]] #REDIRECT [[]] <nowiki></nowiki> Sign your posts: ~~~~

B *I* OCR [Advanced](#) [Special characters](#) [Help](#) [Proofread tools](#)

Header (noinclude):

Page body (to be transcluded):

it up late with his old comrade. It was still too hot to think of staying indoors, so he lay on some cool dock-leaves, and thought over the past day and its doings, and how very good they all had been.

The Rat's light footfall was presently heard approaching over the parched grass. "O, the blessed coolness!" he said, and sat down, gazing thoughtfully into the river, silent and preoccupied.

"You stayed to supper, of course?" said the Mole presently.

"Simply had to," said the Rat. "They wouldn't hear of my going before. You know how kind they always are. And they made things as jolly for me as ever they could, right up to the moment I left. But I felt a brute all the time, as it was clear to me they were very unhappy, though they tried to hide it. Mole, I'm afraid they're in trouble. Little Portly is missing again; and you know what a lot his father thinks of him, though he never says much about it."

"What, that child?" said the Mole lightly.

Footer (noinclude):

<references/>



THE WIND IN THE WILLOWS

it up late with his old comrade. It was still too hot to think of staying indoors, so he lay on some cool dock-leaves, and thought over the past day and its doings, and how very good they all had been.

The Rat's light footfall was presently heard approaching over the parched grass. "O, the blessed coolness!" he said, and sat down, gazing thoughtfully into the river, silent and preoccupied.

"You stayed to supper, of course?" said the Mole presently.

"Simply had to," said the Rat. "They wouldn't hear of my going before. You know how kind they always are. And they made things as jolly for me as ever they could, right up to the moment I left. But I felt a brute all

Wikitext

From Wikipedia, the free encyclopedia

The **Text Encoding Initiative (TEI)** is a **text-centric community of practice** in the academic field of digital humanities, operating continuously since the 1980s. The community currently runs a mailing list, meetings and conference series, and maintains an **eponymous technical standard**, a **journal**, a **wiki**, a **GitHub repository** and a **toolchain**.

Contents [[show](#)]

TEI guidelines [[edit](#) | [edit source](#)]

The *TEI Guidelines*, which collectively define an XML format, are the defining output of the community of practice. The format differs from other well-known [open formats](#) for text (such as [HTML](#) and [OpenDocument](#)) in that it's primarily *semantic* rather than presentational; the semantics and interpretation of every tag and attribute are specified. Some 360 different textual components and concepts (word^[1] sentence^[2] character^[3] glyph^[4] person^[5] etc.): each is grounded in one or more academic disciplines and examples are given.

Technical details [\[edit | edit source \]](#)

The standard is split into two parts, a discursive textual description with extended examples and discussion and set of tag-by-tag definitions. Schemata in most of the modern formats ([DTD](#), [RELAX NG](#) and [W3C Schema](#)) are generated automatically from the tag-by-tag definitions. A number of tools support the production of the guidelines and the application of the guidelines to specific projects.

A number of special tags are used to circumvent restrictions imposed by the underlying [Unicode](#): glyph to allow representation of characters that don't qualify for Unicode inclusion^[1] and choice to allow overcome the required strict linearity^[6]



```
{{short description|An academic community concerned with practices for semantic markup of texts}}
```

[[File:Text Encoding InitiativeTEI Logo.svg|thumb|

The "Text Encoding Initiative" ("TEI") is a [written language|text-centric] [community of practice] in the [List of academic disciplines|academic field] of [digital humanities], operating continuously since the 1980s. The community currently runs a mailing list, meetings and conference series, and maintains an [eponymous] [technical standard], a [scientific journal|journal], a [wiki], a [GitHub] repository and a [toolchain].

--TEI guidelines--

The "TEI Guidelines", which collectively define an XML format, are the defining output of the community of practice. The format differs from other well-known [Open formats] for text (such as [HTML] and [OpenDocument]) in that it's primarily semantic rather than presentational; the semantics and interpretation of every tag and attribute are specified. Some 500 different textual components and constructs

[illegible]

Text Encoding Initiative (Q780920)

an academic community concerned with practices for semantic markup of texts

TE

edit

[🔽 In more languages](#) [Configure](#)

Language	Label	Description	Also known as
English	Text Encoding Initiative	an academic community concerned with practices for semantic markup of texts	TEI
French	Text Encoding Initiative	<i>No description defined</i>	
Spanish	Text Encoding Initiative	consorcio y estándar para el mercado semántico de textos digitales	
German	Text Encoding Initiative	Organisation und Datenformat	TEI TEI-Konsortium

All entered languages

Wikipedia (11 entries) [edit](#)

da	Text Encoding Initiative
de	Text Encoding Initiative
en	Text Encoding Initiative
es	Text Encoding Initiative
fr	Text Encoding Initiative
it	Text Encoding Initiative
ja	Text Encoding Initiative
oc	TEI
pl	Text Encoding Initiative
ta	எழுத்தின் குறியேற்ற முறை
zh	文本編碼規範

Wikibooks (0 entries) [edit](#)

Wikinews (0 entries) [edit](#)

Wikiquote (0 entries) [edit](#)

Wikisource (0 titles) [edit](#)

Wikiversity (0 entries)  [edit](#)

Statements

instance of

community of practice

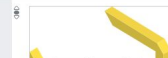
edit

0 references

+ add reference

+ add value

logo image



What about making MediaWiki support TEI?

- Powerful collaboration platform
- Proper markup for texts
- Open source code and specifications



<https://www.mediawiki.org/wiki/Extension:TEI>

XML editor and validation based on an ODD

Editing TestTEI

A ▾

✎ ▾

```
1 <text xmlns="http://www.tei-c.org/ns/1.0">
2   <body>
3     <figure><graphic url="Text_Encoding_InitiativeTEI_Logo.svg" width="221" height="244"/><figDesc>TEI logo</figDesc></figure>
4     <p xml:lang="en"><ref target="https://en.wikipedia.org/wiki/Text_Encoding_Initiative" rend="bold"><abbr>TEI</abbr></ref> is a nice <hi rend="italic">markup language</hi> <hi rend="sup">with</hi> root tag <hi rend="var">TEI</hi></p>
5     <p><formula notation="tex">\frac{2^3}{4\sqrt{5}}</formula></p>
6     <list type="unordered">
7       <item>1</item>
8       <item>2</item>
9       <item>3</item>
10    </list>
11    <list type="ordered">
12      <item>1</item>
13      <item>2</item>
14    </list>
15    <table>
16      <row>
17        <cell role="label">Col 1</cell>
18        <cell role="label">Col 2</cell>
19      </row>
20    </table>
21  </body>
22 </text>
```

Summary:

☐ This is a minor edit ☒ Watch this page

Please note that all contributions to TEI are considered to be released under the Creative Commons Attribution-ShareAlike (see [Project:Copyrights](#) for details). If you do not want your writing to be edited mercilessly and redistributed at will, then do not submit it here.

You are also promising us that you wrote this yourself, or copied it from a public domain or similar free resource. **Do not submit copyrighted work without permission!**

Save changes

Show preview

Show changes

Cancel

Editing help (opens in new window)

HTML rendering for (most of) simple print

TestTEI



TEI logo

TEI is a nice *markup language* with root tag *TEI*

$$\frac{2^3}{4\sqrt{5}}$$

- 1
- 2
- 3

1. 1

2. 2

Col 1 Col 2 Col 3

Row 1 2 3

Revision history of "TestTEI"

? Help

[Expand]

Compare selected revisions

- Compare selected revisions

Tpt (talk | contribs | block)

[← Older edit](#)

Tpt ([talk](#) | [contribs](#) | [block](#)) [[rollback 5 edits](#)]

Line 3:

$\frac{2^3}{4\sqrt{5}}$

```
<list type="unordered">
```


<item>1</item>

$\frac{2^3}{9\sqrt{5}}$

```
<list type="unordered">
```

<item>1</item>

VisualEditor



TEI logo

TEI is a nice markup language with root tag TEI

2^3
 $9\sqrt{5}$

- 1
- 2
- 3

1. 1

2. 2

▼

Col 1	Col 2	Col 3
Row 1	2	3

📄

Table

Delete

Properties

Still a work in progress

Future work

- Better documentation
- ODD configuration (currently hardcoded)
- Improve TEI <-> HTML conversion
- Make VisualEditor less limited
- Automated documentation from ODD
- Integration with ProofreadPage
- ...

Please join the effort if you are interested!

Live demo at the poster session



+



<https://www.mediawiki.org/wiki/Extension:TEI>

A special thanks to Gimena del Rio Riande for her help