

Wiki · Import (version 1.2.804)

LERA natively supports plain text (TXT) and Extensible Markup Language (XML) files for the import of text documents (witnesses) according to a defined scheme described in this article.

In General

LERA supports Unicode according to UTF-8. Other encodings such as ISO-8859-1 or UTF-16le are automatically converted to UTF-8 when the files are uploaded. The focus of LERA is on comparing the content of text versions (witnesses), accordingly, only a few structural and special elements are processed and displayed. The native XML format follows the guidelines of the [Text Encoding Initiative \(TEI\)](#), whereby a certain part being processed and represented by LERA. These elements and attributes are described below. Additional elements and attributes in the XML files do not interfere, but are not treated either.

Meta Data

Some meta data for the text witness is recognized by LERA on upload when encoded in a specific way to a XML file (i.e. within <TEI> ... <teiHeader> ... <sourceDesc> ... <bibl> ...). You can specify the following data:

- <title> - used as name of the text witness
- <abbr> - used as an unique siglum for the witness
- <date> - the year of publication
- <note> - a short description of the text witness

Structural Elements

Line Breaks

- XML: <lb/>
- TXT: single line break – there may a carriage return (\r = U+000d) followed by a line feed (\n = U+000a) or line separator (U+2028) or `␣break␣line␣␣`

Column Breaks

- XML: <cb>
- TXT: `␣break␣column␣␣`

Paragraphs

- XML: <p>
- TXT: empty line – there may a carriage return (\r = U+000d) and than a paragraph separator (U+2029) or two line feeds (\n = U+000a) or two line separators (U+2028)

Page breaks

- XML: <pb>
- TXT: a form feed (\f = U+000c) or `␣break␣page␣␣`

Special Elements

Page Numbers

- XML: `<pb n="42"/>`
- TXT: `⌘pagenumber⌘42⌘`

Text Alterations

This includes various interventions by the author himself or by the editors, for example to correct errors or to resolve abbreviations. In general, a value `original` is used for representation, while in text comparison the value `modified` is used. Currently three types, errata (`sic`), abbreviations (`abbr`) and regularization (`orig`), are provided:

Errata

- XML: `<choice><sic>original</sic><corr>modified</corr></choice>`
- TXT: `⌘alteration⌘sic⌘original⌘modified⌘`

Abbreviations

- XML: `<choice><abbr>NY</abbr><expan>New York</expan></choice>`
- TXT: `⌘alteration⌘abbr⌘NY⌘New York⌘`

Regularization

- XML: `<choice><orig>Fluß</orig><reg>Fluss</reg></choice>`
- TXT: `⌘alteration⌘orig⌘Fluß⌘Fluss⌘`

Placeholder for figures such as sketches or diagrams

- XML: `<figure><figdesc>frame with picture</figdesc></figure>`
- TXT: `⌘figure⌘frame with picture⌘`

Notes and Marginals

- XML: `<note place="inline" n="†">additional details ... </note>`
- TXT: `⌘note⌘inline⌘†⌘additional details ... ⌘`

Proper Names and Referencing Strings

- XML: `<name type="person" ref="#raynal" key="Abbe Raynal">Guillaume Thomas François Raynal</name>`
- XML: `<rs type="person" ref="#raynal" key="Abbe Raynal">Guillaume Thomas François Raynal</rs>`
- TXT: `⌘name⌘person⌘#raynal⌘Guillaume Thomas François Raynal⌘Abbe Raynal⌘`

Headlines

- XML: `<head>A Headline</head>`
- TXT: `⌘head⌘A Headline⌘`

Editorial Markers

Other Special elements can be indicated with `<metamark>` to render them differently or ignore them on text comparison. Examples might be lacunas, line fillers, special symbols or text decoration.

- XML: `<metamark function="lacuna">[]</metamark>`
- TXT: `⌘metamark⌘lacuna⌘[]⌘`

Anchors

Anchors can be used to control the alignment while comparing text witnesses.

- XML: `<anchor>Chapter One</anchor>`
- TXT: `⌘anchor⌘Chapter One⌘`

To turn on prioritized alignment of segments with matching anchors, set the check box `prioritize anchors (<anchor>)`.

Gaps

- XML: `<gap quantity="5"/>`
- XML: `<gap extent="5 chars"/>`
- TXT: `⌘gap⌘5⌘`

Text Emphasis

Bold

- XML: `<hi rend="bold">...</hi>`
- TXT: `⟨...⟩` (U+2770 and U+2771)
- ... Normal Text ... **Emphasized Text**...

Italic

- XML: `<hi rend="italic">...</hi>`
- TXT: `{...}` (U+2774 and U+2775)
- ... Normal Text ... *Emphasized Text* ...

Larger Fontsize

- XML: `<hi rend="large">...</hi>`
- TXT: `⸀...⸁` (U+2e22 and U+2e23)
- ... Normal Text ... **Emphasized Text** ...

Smaller Fontsize

- XML: `<hi rend="small">...</hi>`
- TXT: `⸂...⸃` (U+2e24 and U+2e25)
- ... Normal Text ... **Emphasized Text** ...

Small Caps

(convert lowercase letters to uppercase displayed in a smaller font size)

- XML: `<hi rend="smallcaps">...</hi>`
- TXT: [...] (U+230a and U+230b)
- ... Normal Text ... EMPHASIZED TEXT ...

Spaced Letters

- XML: `<hi rend="spaced">...</hi>`
- TXT: ⟨...⟩ (U+27ea and U+27eb)
- ... Normal Text ... Emphasized Text ...

Strikethrough

- XML: `<hi rend="strikethrough">...</hi>` or `...`
- TXT: [...] (U+2045 and U+2046)
- ... Normal Text ... ~~Emphasized Text~~ ...

Subscript

- XML: `<hi rend="subscript">...</hi>` or `_{...}`
- TXT: [...] (U+298f and U+298e)
- ... Normal Text ... Emphasized Text ...

Superscript

- XML: `<hi rend="superscript">...</hi>` or `^{...}`
- TXT: [...] (U+298d and U+2990)
- ... Normal Text ... ^{Emphasized Text} ...

Unclear

- XML: `<hi rend="unclear">...</hi>` or `<unclear>...</unclear>`
- TXT: (...) (U+276a and U+276b)
- ... Normal Text ... Emphasized Text ...

Underlined

- XML: `<hi rend="underline">...</hi>`
- TXT: [...] (U+298b and U+298c)
- ... Normal Text ... Emphasized Text ...