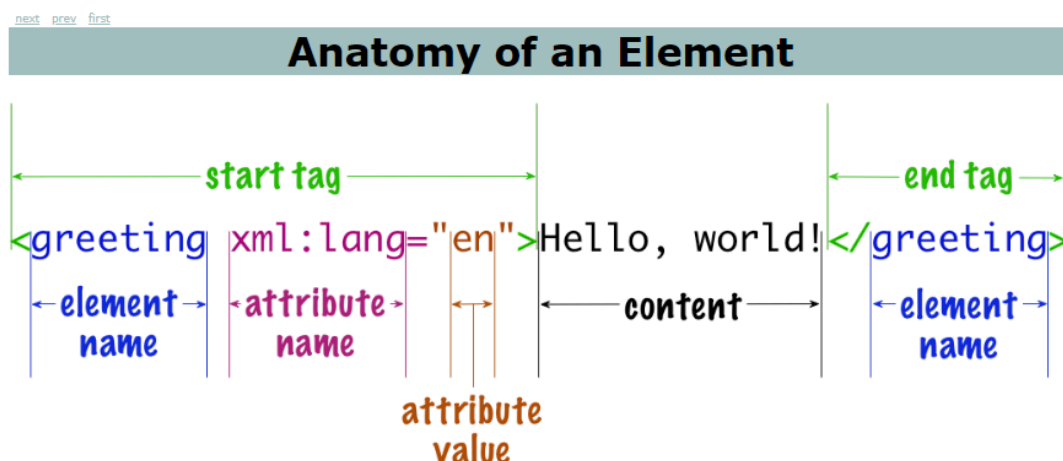


Troubleshooting Text Encoding

[Click to schedule a meeting with the DITI Team](#)

Text encoding, while it can be intimidating at first, is just like learning new rules of grammar. If you can approach text encoding like a grammar, then learning how to encode text can become a matter of just looking up the particular rules of the grammar when you don't know them. Even experienced text encoders spend a lot of time looking up the definitions of tags in manuals like the [TEI guidelines](#). If you are a new learner and find yourself overwhelmed by the newness of encoding, don't feel discouraged—it's just part of the learning process.

As a reminder, text encoding is a process which involves formatting text with a series of tags according to some set rules of guidelines, usually the TEI guidelines. Start tags should always be accompanied with an end tag which sandwiches the content you wish to describe using tags.



Well-formedness Errors

Many of the initial errors or troubleshooting issues you are likely to experience when you are starting off with encoding are likely to be related to missing end tags, improper formatting of tags, and otherwise missing elements provided in the above graphic. This type of error is called a **well-formedness error** or **syntax error**. Here is a checklist which you can look over to ensure that your tags are properly formatted:

- ☐ Start tags have matching end tags
- ☐ All tags are spelled correctly and are without typos

Developed by: Avery Blankenship

Questions? Contact us: nulab.info@gmail.com

- ☐ All attribute values are enclosed on quotation marks
- ☐ Tags do not include extra punctuation marks (<,>, /, “”, etc.)
- ☐ Start and end tag names match
- ☐ Empty tags are terminated with “/>”

Structural Errors

Another form of error that XML documents can commonly have is what is called a structural error. This type of error typically involves tags occurring in the incorrect order or being nested incorrectly. For example, paragraph tags (<p>) cannot occur outside of division tags (<div>) and similarly tags cannot be broken up by other tags (<p><div></p></div>). The following checklist can provide some guidance for catching structural errors:

- ☐ Your document includes the appropriate <div>, <header>, and <p> tags. The TEI guidelines provides a series of [sample documents](#) which you can reference to ensure that your document has the correct headers and formatting tags
- ☐ If you are creating a TEI compliant document, do you begin with the correct root tag (<TEI>)?
- ☐ Are all of the structural tags properly terminated (for example, the header tag closed with an ending tag before beginning the text tag)
- ☐ There are no rags or text outside of the root tag

Validation Errors

One of the more tricky errors to deal with when working with text encoding is a validation error. Validation errors are errors that are caused by invalid uses of tags or otherwise uses of XML that break the rules of the XML schema. Sometimes these errors can involve using a tag that isn't allowed by the schema, using an invalid data type for a specific tag, or using a tag in the wrong place. In addition to keeping an eye on the validation warnings that the platform you are using to encode might draw your attention to automatically (Oxygen, for example, might draw a red line under invalid tags), some additional items to pay attention to are:

- ☐ Tags do not include any duplicate attributes
- ☐ All attribute names are allowed in the tags according to the TEI guidelines
- ☐ All attribute values are allowed for the attribute name
- ☐ All tags are real tags in the TEI guidelines
- ☐ All attributes required by the schema are present
- ☐ Special characters are correctly encoded (for example, & should be & instead)

Developed by: Avery Blankenship

Questions? Contact us: nulab.info@gmail.com