DAW2015 Introduction to Digital Editions with TEI: Exercises

These exercises will allow you to experiment with applying some basic TEI markup to some excerpted literary texts from Edgar Allan Poe. You will be using a paper copy of the text as well as the oXygen XML editor to work on this exercise. All of the necessary electronic files for the exercise can be downloaded from the **Dropbox folder:** http://bit.ly/1SR4ioX.

Exercise 1: Structural I (analog)

Take a look at the paper copy of your text.

- Use **<front>**, **<body>**, and **<back>** tags to demarcate the front matter, body, and back matter.
- Within each of those regions, identify the main divisions and surround them with **<div>** tags, supplying an appropriate type for each one (e.g., <div type="contents">).
- If the division has a head, enclose it with **<head>** tags.
- What are some of the other structural divisions or elements that we might want to tag?

Exercise 2: Structural II (+ getting familiar with oXygen)

1. Launch <oXygen>, and open up the file called **poe.txt.xml** (**CTRL-O /⌘-O**, or select "**Open**" from the File menu). This is a plain text version of your text, with a single XML tag (**<xml>**) to serve as the root element.

2. Let's get some practice using the oXygen XML Editor by adding some basic structural tags to our poem, short story, and verse drama excerpts.

- Add the tags that you defined on the paper copy to your document: **<front>, <body>, <back>, <div>, <head>**
- Begin to encode **The Raven**, using **<lg>** to enclose each stanza and **<l>** to enclose each line. Note that if you try to type an opening tag, oXygen will immediately try to supply the closing tag. This is helpful in some contexts, but not for what we're doing here. Try these methods instead:

  - ✴To enclose the first stanza in an **<lg>** tag, highlight the whole stanza and press **CTRL-E** (Win) or ⌘-**E** (Mac) (or select **Document > XML Refactoring > Surround** with tags…) and type **lg** into the pop-up box. Use the same technique to surround each line with an **<l>** tag, one line at a time.

  - ✴Enclose the second stanza in an **<lg>** tag, as above. While the stanza is still highlighted, press **CTRL-E** or ⌘-**E** again to surround the stanza with an **<l>** tag. Then, place your cursor at the end of the first line and press **Alt-Shift-D** (Win) or **CTRL-Opt-D** (Mac) (or select **Document > XML Refactoring > Split Element**) to split the line at that point. Do the same at the end of each line of the stanza (except the last one!)

- When you get tired of tagging verse, go to "The Angel of the Odd" and add **<p>** tags around each paragraph.   (Note that the boundaries of the first paragraph are a bit ambiguous.)

- When you get tired of *that*, move on to "Politian". Add **<speaker>** tags around each speaker's name, **<sp>** tags around each speech, and **<stage>** tags around the stage directions (this would include the setting description as well as descriptors like "*sighing.*") Because this is a verse drama, you could also treat each speech as a stanza and mark it with **<lg>** and **<l>** tags.

- Insert a pair of **<figure>** tags in the appropriate place beneath the <head> on page 7. Within <figure>, include **<graphic url="figure1.jpg"/>** to indicate the location of the digitized image. Beneath <graphic>, include a **<figDesc>** that contains a brief description of the image.

- Here are more steps you can take:

  ✳ Add tags to mark page breaks **<pb/>**. You can add page numbers as attributes: <pb n="1"/>. You can also mark line breaks **<lb/>**.

  ✳ Use an attribute to define your line groups <lg> or other <div> elements. *Hint: consult the TEI Tags Cheat sheet.*

3. Encoding the Front matter

- Tag the title page with the **<titlePage>** element

  ✳ We cannot add text directly within the <titlePage> element, so we'll need additional tags for the title **<title>** and publication information **<docImprint>**.

  ✳ Here we have a main title and a subtitle, so we can use the **<titlePart>** element with the type attributes "main" and "sub".

  ✳ **<docImprint>** also requires that we have child elements to contain our data. Tag the publisher with the **<publisher>** element and the date with the **<docDate>** element.

- Use **<div type="contents">** to enclose the Table of Contents. Within this <div> we have a **<head>** and a **<list>** with **<item>**s.

4. Encoding the Back matter

- Tag the Notes section with the **<div type="notes">** element. There is also a heading and a paragraph here…

Exercise 3: teiHeader (and TEI schemas)

1. Now let's jump ahead a bit by creating a new TEI Lite document and inserting a version of the Poe document with basic structural tags.

- Type **CTRL-N** or ⌘-**N** (or select *New* from the *File* menu) to launch a new document. In the New dialog box, scroll down to the *Framework Templates* folder and open the *TEI P5* folder. Select "**TEI Lite**" and click the *Create* button.

- Notice what oXygen provides automatically: an **XML declaration** (line 1), a **schema declaration** (line 2), the **root <TEI> element** including a **namespace declaration** (line 3), a minimal **TEI header**, and a minimal **text** with a **body** and a single **<p>**.

*** *Keep in mind, the header contains information about the text, not a transcription of the title page. This is our metadata.* ***

2. Complete the minimal header:

- The title is "Digital Edition of The Incomplete Works of Edgar Allan Poe".

- The publication statement could say something like, "Produced for 'Digital Antiquarian TEI Workshop, 2015".

- The source statement might say, "Excerpted from electronic texts at the University of Virginia Library."

- Delete the <text> tags (and the <body> and <p> tags within). Leave your cursor at that spot.

- Go to **Document > File > Insert File**… and select **poe.tagged.xml**.

  ✳Validate the document by clicking the red check mark in the toolbar at the top of the screen. You should see "Document is valid" at the bottom of your screen.

  ✳Save the new document as **poe.xml.**

Exercise 4: Presentational/semantic (if time allows)

Notice that certain formatting and typographical features (bold, italics, superscript, font variations, page numbers, centering, etc.) are absent from our markup. Your job as encoder is to decide which of these are significant and how best to represent them in your encoding. Some possibilities include:

- Add a **rend** attribute to the last line of the first stanza of "The Raven," to indicate that it should be indented. (You can repeat this with each stanza).

- Add a **rend** attribute to the <head>s that should be centered. (ex. <head rend="center">)

- Add a **<hi>** tag with a **rend** attribute to the first word of The Raven to indicate small caps.

- Use the print text to identify chunks of text that are italicized due to emphasis, and supply **<emph>** tags. (See the first and third full stanzas on p. 3, and the speech at the bottom of p. 7.)

- In the first paragraph on p. 4, mark the word "truffe" as **<foreign xml:lang="fr">** to indicate that it is French. Later in the paragraph, mark the titles of the literary works mentioned with **<title>** tags.

- In the same paragraph, encode the quote with **<quote>** tags. Within <quote>, encode the lines as verse using the appropriate tags.

Exercise 5: Contextual

1. People, places, dates

- Open the **Parker.xml** document

- Tag a few proper names. – Tag a few named places. – Are there any geographic names?:
  <persName>, **<placeName>**, **<geogName>**

- Tag a few dates: **<date when="yyyy-mm-dd">**, or *notBefore*, *notAfter*, *from*…

2. Referencing entities

- Add a ref attribute with a direct web reference to a tagged person or place: **<persName ref="url">**

- Add a key attribute to a tagged person or place: **<persName key="description here">**

- Create an "authority record" with an **xml:id** and use a ref attribute with a tagged person or place to refer back to your authority record: in back matter **<person xml:id="abc">**; in text **<persName ref="#abc">**

Exercise 6: Editorial

1. Additions, deletions, etc.

- Open the **ParkerJournal.jpg** file - this page image corresponds to the first couple of paragraphs in our Parker.xml file

- Scan for additions, deletions, and substitutions

- Find the corresponding text in the **Parker.xml** file and tag accordingly: **<add />**, **<del />**, **<subst />**

2. Editing and Notes

- Here you can work with either the **Poe text** (guided) or with the **Parker journal** (freestyle!)

- Normalize a word or phrase and encode both the original form and the normalized form

  - Poe: Note the nonstandard spelling "visiter" in the 1st and 3rd stanzas of "The Raven". For each occurrence, enclose the word in **<orig>** tags, then immediately after include the normalized spelling within **<reg>** tags. Surround both with a pair of **<choice>** tags.

  - Poe: Oops – "December" is misspelled in the 2nd stanza. Tag it with **<sic>** tags, then supply the corrected spelling within **<corr>** tags. Surround both with **<choice>** tags.

- Add a note to the back matter and point to the note from the point of attachment

  - Poe: The notes at the end of the text (in the <back> matter) are lumped together as a paragraph. Remove the <p> tag, and encode each note with a separate **<note>** tag. Be sure to use the **xml:id** attribute to give each note a unique identifier.

- Encode a note at the point of attachment (hint: remember to identify **type**, **place**, and **responsibility**)

  - Poe: Find the footnote markers within the text (around lines 57, 189, and 219 of the electronic version). Enclose each marker with **<ref target="#???">**, where ??? is the unique identifier of the corresponding note.