


<oo> → <dh> Digital humanities

Maintained by: David J. Birnbaum (djbitt@gmail.com) 

Last modified: 2015-02-08T19:27:00+0000

XPath assignment #1

You can find an XML (TEI) version of Shakespeare's *Hamlet* at <http://dh.obdurodon.org/bad-hamlet.xml>. We've deliberately damaged some of the markup in this edition to introduce some inconsistencies, but the file is well-formed XML, which means that you can use XPath to explore it. You should download this file to your computer (typically that means right-clicking on the link and selecting "save as") and open it in <oXygen/>.

After you've completed your homework, save your answers to a file and upload it to CourseWeb as an attachment. (Please use an attachment! If you paste your answer into the text box, CourseWeb may munch the angle brackets.) Some of these tasks are thought-provoking, and even difficult. If you get stuck, do the best you can, and if you can't get a working answer, give the answers you tried and explain where they failed to get the results you wanted. Sometimes doing that will help you figure out what's wrong, and even when it doesn't, it will help us identify the difficult moments. These tasks require the use of path expressions, predicates, and the functions `count()` and `not()`, but they should not require any other XPath functions. There may be more than one possible answer.

Using the *Bad Hamlet* document and the XPath browser window in <oXygen/>, construct XPath expressions that will do the following (give the full XPath expressions in your answers, and not just the results):

1. Hamlet, like a typical Shakespearean tragedy, contains five acts, each of which contains scenes. Both acts and scenes are encoded as division (<div>) elements.
 - a. How can XPath tell them apart?
 - b. What XPath would find just the acts?
 - c. What XPath would find just the scenes?
 - d. What XPath would find just the scenes in Act III?
2. Stage directions (<stage>) occur in a variety of contexts.
 - a. What XPath would find all of the stage directions that are inside a metrical line (<l>), that is, between the starting <l> and the ending </l>. How many are there?
 - b. What XPath would find all of the stage directions that are directly inside a speech (<sp>), that is, inside a speech but not inside a line within a speech?
 - c. What XPath would find all of the stage directions that are not directly inside a speech or a line. How many are there?
 - d. For the stage directions you identified in #2c, above, write an XPath expression that will return not the <stage> elements themselves, but their parent elements, whatever they might be. What are those parent elements? (You haven't yet learned the XPath to return just the names of the parent elements [rather than the elements themselves], but you can locate them, click on each one in the list <oXygen/> returns, and look at it directly.)