<oo>→<dh> Digital humanities

Maintained by: David J. Birnbaum (djbpitt@gmail.com)
Last modified: 2015-08-24T19:05:34+0000

XPath assignment #2

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. Most (not all) speeches in Hamlet contain, as their immediate children, mostly metrical line (`<l>`) and "anonymous block" (`<ab>`) elements (an anonymous block is the TEI element that the tagger used to represent a non-metrical speech line). Speeches also typically contain, as immediate child elements, `<speaker>` elements, and may also contain stage directions (`<stage>`). We have deliberately left out at least one other type of subelement found as an immediate child of speech elements. Based on this understanding:
   - What XPath would find all of the speeches that do not contain any metrical lines as immediate children? How many are there?
   - What XPath would find all of the speeches that do not contain either any metrical lines (`<l>`) or any anonymous blocks (`<ab>`) as immediate children? How many are there? What do they contain instead? (As in #2d in XPath assignment #1, you haven't yet learned the XPath to return a list of the types of elements they do contain, but if you find them all, you can scan the brief list that <oXygen/> returns, click on each one to see it in context, and see what's going on.)
2. Explain why the following four XPath expressions return two different results, and describe in prose what each of them does return, and why:
   - //sp[@who="Hamlet"]/l[1]
   - /descendant::sp[@who="Hamlet"]/l[1]
   - (//sp[@who="Hamlet"]/l)[1]
   - (/descendant::sp[@who="Hamlet"]/l)[1]