


<oo> → <dh> Digital humanities

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

Last modified: 2018-03-02T00:17:53+0000

Test #4: XPath

The task

Using *Bad Hamlet*, provide an XPath expression that retrieves:

1. All speeches (<sp>) by Ophelia that contain Hamlet's name. Requires, at least in our solution, **contains()**. (There are two such speeches.)
2. A semicolon-separated list of all unique speakers (<speaker>) in Act IV, without duplicates. Requires, at least in our solution, **string-join()** and **distinct-values()**. Your list will include, among other, "Rosencrantz", "Guildenstern", and, because they sometimes speak together, "Rosencrantz and Guildenstern". For the purpose of this test, you don't have to get rid of the last of these (but see the bonus question, below).
3. The number of speeches (<sp>) in each act (**//body/div**). Our solution requires **count()**. (The number of speeches you should find are 251 for Act 1, 201 for Act 2, 249 for Act 3, 179 for Act 4, and 257 for Act 5.)
4. The speaker elements (<speaker>) for all speeches (<sp>) that are greater than 4000 characters long. Requires, at least in our solution, **string-length()**. Hint: to approach this in stages: 1) find all speeches; 2) filter them to keep just the ones that are more than 4000 characters in length; 3) find the speakers of those speeches. (There are two such speeches, one by "Hamlet" and one by "Ghost".)
5. This question has four parts, which build incrementally on one another:
 - a. The number of lines (<l> elements) in each speech (<sp> element). Here and in the other parts of this question, count the descendants of the speech, not just the children.
 - b. The number of lines in the longest speech (measured in <l> elements). The answer is 58.
 - c. The longest speech (<sp>) itself. It's the speech by Hamlet that begins with the line that has the **@xml:id** value "sha-ham202553".
 - d. The **<head>** of the scene that contains that speech. The answer is "Act 2, Scene 2".

Bonus

How can you use XPath to get the semi-spurious "Rosencrantz and Guildenstern" out of the answer to #2? Your answer should cater to the following possibilities:

- There could be other pairs of simultaneous speakers, that is, Horatio and Ophelia might, in principle also speak together.
- Although Rosencrantz and Guildenstern both happen to speak individually in this act, in principle there might be characters who speak together but not separately, and you need to list them individually in your list. For example, if Rosencrantz spoke separately and with Guildenstern, and Guildenstern spoke only together with Rosencrantz, your list of speakers should include both of them individually.

- It's possible in principle for there to be more than two simultaneous speakers, and in that case they might be joined with both commas and "and", along the lines of "King, Gertrude, Hamlet, and Servant". These persons may or may not also speak individually, so you need to be sure that they all make it into your list as separate entries.