

XPath for processing XML and managing projects

DHSI 2019 Course 49 (Week 2, 10–14 June, 2019)

[Course description](#) [Syllabus](#) [Resources and references](#) [Course Pack](#) [View on GitHub](#)

Instructors: Elisa Beshero-Bondar and David J. Birnbaum

Description:

Learn XPath intensively and gain superpowers with XML processing! Whether you’ve recently learned XML and want to build something with it, or whether you’ve worked with XPath before but are rusty, new and experienced coders alike will benefit from our course. XPath is usually not the center of a DHSI class, and people often gain hasty “ad hoc” experience with it when learning it only along the way to doing something else. Concentrating intensively for a week on XPath will “power up” what you can do with XML, and will help you refine the way you code your documents. Our course will assist XML coders (whether beginners or experienced) with complex processing of information from markup and from plain text. Our goals are 1) to increase our participants’ confidence and fluency in reading and extracting information coded in XML archives and databases, and 2) to share strategies for systematically reviewing, designing, and building those archives and databases.

Because we can “dig” latent information out of the document “strata” of texts, we think of working with XPath as something like planning an archaeology project, turning an XML project into a carefully managed digital dig site for cultural data! In our course you’ll gain experience with writing precise and powerful XPath to illuminate information that isn’t obvious on a human reading. For example, we’ll write XPath to calculate how frequently you have marked a certain phenomenon, or locate which names of persons are mentioned together in the same chapter, paragraph, sentence, stanza, footnote, or other structural unit. We’ll apply XPath to check for accuracy of text encoding—to write schema rules to manage your coding (or your project team’s coding). You will learn how XPath can help you to pull data from your documents into lists, tables, and graphic visualizations.

XPath is the center of the course, but we will explore how it applies in multiple XML processing contexts so that you learn how these work similarly and how these are used, respectively, to validate documents and to transform them for publication and other reuse. Thus we devote serious, sustained attention to writing *and* applying XPath by surveying how it is expressed in a variety of frameworks (including XSLT, XQuery, and Schematron), with a variety of materials (including XML and plain-text documents), and involving a variety of task types (such as date arithmetic to calculate how much time elapsed between dates and string surgery to look for and manipulate patterns inside your coded elements). You’ll gain fluency with XPath expressions and

patterns, including predicates, operators, functions (from the core library and user-defined), regular expressions, and other features, and we'll practice these in different XML-related contexts, starting with XQuery, and moving to XSLT and Schematron). Whether you are an XML beginner or a more experienced coder, you'll find that XPath will help you with systematic encoding, document processing, and project management.

This is a hands-on course. Consider this offering in complement with, and / or to be built on by: Text Encoding Fundamentals and their Application, Out-of-the-Box Text Analysis for the Digital Humanities, Text Processing - Techniques & Traditions, XML Applications for Historical and Literary Research. No advanced knowledge of XML processing is necessary but those with interests in document processing who have taken Digital Documentation and Imaging for Humanists; Advanced TEI Concepts / TEI Customization; A Collaborative Approach to XSLT; or Geographical Information Systems in the Digital Humanities will certainly benefit.

[Link to register for DHSI courses](#)

UpTransformation is maintained by **ebeshero**.