These scripts are written to explore the practice of processing XML documents using the xml.dom API (Application Programming Interface). The following scripts are explored;

Python XML DOM Minidom.

‘xml.dom.minidom’ is a minimal DOM implementation.

xml.dom.minidom is a minimal implementation of the Document Object Model interface, with an API similar to that in other languages.

It is intended to be simpler than the full DOM and also significantly smaller.

Users who are not already proficient with the DOM should consider using the xml.etree.ElementTree module for their XML processing instead.

Python XHTML Parser.

‘html.parser’ is a simple HTML and XHTML parser.

This module defines a class HTMLParser which serves as the basis for parsing text files formatted in HTML (HyperText Mark-up Language) and XHTML.

Python XML Parsers Expat.

‘xml.parsers.expat’ is a fast XML parsing using Expat.

Warning: The pyexpat module is not secure against maliciously constructed data.

If you need to parse untrusted or unauthenticated data see XML vulnerabilities.

The xml.parsers.expat module is a Python interface to the Expat non-validating XML parser.

The module provides a single extension type, xmlparser, that represents the current state of an XML parser.

After an xmlparser object has been created, various attributes of the object can be set to handler functions.

When an XML document is then fed to the parser, the handler functions are called for the character data and markup in the XML document.

Python XML Pulldom.

‘xml.dom.pulldom’ is a support for building partial DOM trees.

The xml.dom.pulldom module provides a “pull parser” which can also be asked to produce DOM-accessible fragments of the document where necessary.

The basic concept involves pulling “events” from a stream of incoming XML and processing them.

In contrast to SAX which also employs an event-driven processing model together with callbacks, the user of a pull parser is responsible for explicitly pulling events from the stream, looping over those events until either processing is finished or an error condition occurs.

Python XML Etree Element Tree.

‘xml.etree.ElementTree’ is the the ElementTree XML API.

The xml.etree.ElementTree module implements a simple and efficient API for parsing and creating XML data.