

xml.dom — The Document Object Model API

Source code: [Lib/xml/dom/__init__.py](#)

The Document Object Model, or “DOM,” is a cross-language API from the World Wide Web Consortium (W3C) for accessing and modifying XML documents. A DOM implementation presents an XML document as a tree structure, or allows client code to build such a structure from scratch. It then gives access to the structure through a set of objects which provided well-known interfaces.

The DOM is extremely useful for random-access applications. SAX only allows you a view of one bit of the document at a time. If you are looking at one SAX element, you have no access to another. If you are looking at a text node, you have no access to a containing element. When you write a SAX application, you need to keep track of your program’s position in the document somewhere in your own code. SAX does not do it for you. Also, if you need to look ahead in the XML document, you are just out of luck.

Some applications are simply impossible in an event driven model with no access to a tree. Of course you could build some sort of tree yourself in SAX events, but the DOM allows you to avoid writing that code. The DOM is a standard tree representation for XML data.

The Document Object Model is being defined by the W3C in stages, or “levels” in their terminology. The Python mapping of the API is substantially based on the DOM Level 2 recommendation.

DOM applications typically start by parsing some XML into a DOM. How this is accomplished is not covered at all by DOM Level 1, and Level 2 provides only limited improvements: There is a `DOMImplementation` object class which provides access to `Document` creation methods, but no way to access an XML reader/parser/Document builder in an implementation-independent way. There is also no well-defined way to access these methods without an existing `Document` object. In Python, each DOM implementation will provide a function `getDOMImplementation()`. DOM Level 3 adds a Load/Store specification, which defines an interface to the reader, but this is not yet available in the Python standard library.

Once you have a DOM document object, you can access the parts of your XML document through its properties and methods. These properties are defined in the DOM

specification; this portion of the reference manual describes the interpretation of the specification in Python.

The specification provided by the W3C defines the DOM API for Java, ECMAScript, and OMG IDL. The Python mapping defined here is based in large part on the IDL version of the specification, but strict compliance is not required (though implementations are free to support the strict mapping from IDL). See section [Conformance](#) for a detailed discussion of mapping requirements.

See also:**Document Object Model (DOM) Level 2 Specification**

The W3C recommendation upon which the Python DOM API is based.

Document Object Model (DOM) Level 1 Specification

The W3C recommendation for the DOM supported by `xml.dom.minidom`.

Python Language Mapping Specification

This specifies the mapping from OMG IDL to Python.

Module Contents

The `xml.dom` contains the following functions:

`xml.dom.registerDOMImplementation(name, factory)`

Register the *factory* function with the name *name*. The factory function should return an object which implements the `DOMImplementation` interface. The factory function can return the same object every time, or a new one for each call, as appropriate for the specific implementation (e.g. if that implementation supports some customization).

`xml.dom.getDOMImplementation(name=None, features=())`

Return a suitable DOM implementation. The *name* is either well-known, the module name of a DOM implementation, or `None`. If it is not `None`, imports the corresponding module and returns a `DOMImplementation` object if the import succeeds. If no name is given, and if the environment variable `PYTHON_DOM` is set, this variable is used to find the implementation.

If *name* is not given, this examines the available implementations to find one with the required feature set. If no implementation can be found, raise an `ImportError`. The features list must be a sequence of (*feature*, *version*)

pairs which are passed to the `hasFeature()` method on available `DOMImplementation` objects.

Some convenience constants are also provided:

`xml.dom.EMPTY_NAMESPACE`

The value used to indicate that no namespace is associated with a node in the DOM. This is typically found as the `namespaceURI` of a node, or used as the `namespaceURI` parameter to a namespaces-specific method.

`xml.dom.XML_NAMESPACE`

The namespace URI associated with the reserved prefix `xml`, as defined by [Namespaces in XML](#) (section 4).

`xml.dom.XMLNS_NAMESPACE`

The namespace URI for namespace declarations, as defined by [Document Object Model \(DOM\) Level 2 Core Specification](#) (section 1.1.8).

`xml.dom.XHTML_NAMESPACE`

The URI of the XHTML namespace as defined by [XHTML 1.0: The Extensible HyperText Markup Language](#) (section 3.1.1).

In addition, `xml.dom` contains a base `Node` class and the DOM exception classes. The `Node` class provided by this module does not implement any of the methods or attributes defined by the DOM specification; concrete DOM implementations must provide those. The `Node` class provided as part of this module does provide the constants used for the `nodeType` attribute on concrete `Node` objects; they are located within the class rather than at the module level to conform with the DOM specifications.

Objects in the DOM

The definitive documentation for the DOM is the DOM specification from the W3C.

Note that DOM attributes may also be manipulated as nodes instead of as simple strings. It is fairly rare that you must do this, however, so this usage is not yet documented.

Interface	Section	Purpose
<code>DOMImplementation</code>	DOMImplementation Objects	Interface to the underlying implementation.

Interface	Section	Purpose
<code>Node</code>	Node Objects	Base interface for most objects in a document.
<code>NodeList</code>	NodeList Objects	Interface for a sequence of nodes.
<code>DocumentType</code>	DocumentType Objects	Information about the declarations needed to process a document.
<code>Document</code>	Document Objects	Object which represents an entire document.
<code>Element</code>	Element Objects	Element nodes in the document hierarchy.
<code>Attr</code>	Attr Objects	Attribute value nodes on element nodes.
<code>Comment</code>	Comment Objects	Representation of comments in the source document.
<code>Text</code>	Text and CDATASection Objects	Nodes containing textual content from the document.
<code>ProcessingInstruction</code>	ProcessingInstruction Objects	Processing instruction representation.

An additional section describes the exceptions defined for working with the DOM in Python.

DOMImplementation Objects

The `DOMImplementation` interface provides a way for applications to determine the availability of particular features in the DOM they are using. DOM Level 2 added the ability to create new `Document` and `DocumentType` objects using the `DOMImplementation` as well.

```
DOMImplementation.hasFeature(feature, version)
```

Return `True` if the feature identified by the pair of strings *feature* and *version* is implemented.

`DOMImplementation.createDocument(namespaceUri, qualifiedName, doctype)`

Return a new `Document` object (the root of the DOM), with a child `Element` object having the given *namespaceUri* and *qualifiedName*. The *doctype* must be a `DocumentType` object created by `createDocumentType()`, or `None`. In the Python DOM API, the first two arguments can also be `None` in order to indicate that no `Element` child is to be created.

`DOMImplementation.createDocumentType(qualifiedName, publicId, systemId)`

Return a new `DocumentType` object that encapsulates the given *qualifiedName*, *publicId*, and *systemId* strings, representing the information contained in an XML document type declaration.

Node Objects

All of the components of an XML document are subclasses of `Node`.

`Node.nodeType`

An integer representing the node type. Symbolic constants for the types are on the `Node` object: `ELEMENT_NODE`, `ATTRIBUTE_NODE`, `TEXT_NODE`, `CDATA_SECTION_NODE`, `ENTITY_NODE`, `PROCESSING_INSTRUCTION_NODE`, `COMMENT_NODE`, `DOCUMENT_NODE`, `DOCUMENT_TYPE_NODE`, `NOTATION_NODE`. This is a read-only attribute.

`Node.parentNode`

The parent of the current node, or `None` for the document node. The value is always a `Node` object or `None`. For `Element` nodes, this will be the parent element, except for the root element, in which case it will be the `Document` object. For `Attr` nodes, this is always `None`. This is a read-only attribute.

`Node.attributes`

A `NamedNodeMap` of attribute objects. Only elements have actual values for this; others provide `None` for this attribute. This is a read-only attribute.

`Node.previousSibling`

The node that immediately precedes this one with the same parent. For instance the element with an end-tag that comes just before the *self* element's start-tag. Of course, XML documents are made up of more than just elements so the previous sibling could be text, a comment, or something else. If this node is the first child of the parent, this attribute will be `None`. This is a read-only attribute.

Node.**nextSibling**

The node that immediately follows this one with the same parent. See also [previousSibling](#). If this is the last child of the parent, this attribute will be `None`. This is a read-only attribute.

Node.**childNodes**

A list of nodes contained within this node. This is a read-only attribute.

Node.**firstChild**

The first child of the node, if there are any, or `None`. This is a read-only attribute.

Node.**lastChild**

The last child of the node, if there are any, or `None`. This is a read-only attribute.

Node.**localName**

The part of the `tagName` following the colon if there is one, else the entire `tagName`. The value is a string.

Node.**prefix**

The part of the `tagName` preceding the colon if there is one, else the empty string. The value is a string, or `None`.

Node.**namespaceURI**

The namespace associated with the element name. This will be a string or `None`. This is a read-only attribute.

Node.**nodeName**

This has a different meaning for each node type; see the DOM specification for details. You can always get the information you would get here from another property such as the `tagName` property for elements or the `name` property for attributes. For all node types, the value of this attribute will be either a string or `None`. This is a read-only attribute.

Node.**nodeValue**

This has a different meaning for each node type; see the DOM specification for details. The situation is similar to that with `nodeName`. The value is a string or `None`.

`Node.hasAttributes()`

Return `True` if the node has any attributes.

`Node.hasChildNodes()`

Return `True` if the node has any child nodes.

`Node.isSameNode(other)`

Return `True` if *other* refers to the same node as this node. This is especially useful for DOM implementations which use any sort of proxy architecture (because more than one object can refer to the same node).

Note: This is based on a proposed DOM Level 3 API which is still in the “working draft” stage, but this particular interface appears uncontroversial. Changes from the W3C will not necessarily affect this method in the Python DOM interface (though any new W3C API for this would also be supported).

`Node.appendChild(newChild)`

Add a new child node to this node at the end of the list of children, returning *newChild*. If the node was already in the tree, it is removed first.

`Node.insertBefore(newChild, refChild)`

Insert a new child node before an existing child. It must be the case that *refChild* is a child of this node; if not, `ValueError` is raised. *newChild* is returned. If *refChild* is `None`, it inserts *newChild* at the end of the children’s list.

`Node.removeChild(oldChild)`

Remove a child node. *oldChild* must be a child of this node; if not, `ValueError` is raised. *oldChild* is returned on success. If *oldChild* will not be used further, its `unlink()` method should be called.

`Node.replaceChild(newChild, oldChild)`

Replace an existing node with a new node. It must be the case that *oldChild* is a child of this node; if not, `ValueError` is raised.

`Node.normalize()`

Join adjacent text nodes so that all stretches of text are stored as single `Text` instances. This simplifies processing text from a DOM tree for many applications.

`Node.cloneNode(deep)`

Clone this node. Setting *deep* means to clone all child nodes as well. This returns the clone.

NodeList Objects

A `NodeList` represents a sequence of nodes. These objects are used in two ways in the DOM Core recommendation: an `Element` object provides one as its list of child nodes, and the `getElementsByTagName()` and `getElementsByTagNameNS()` methods of `Node` return objects with this interface to represent query results.

The DOM Level 2 recommendation defines one method and one attribute for these objects:

`NodeList.item(i)`

Return the *i*th item from the sequence, if there is one, or `None`. The index *i* is not allowed to be less than zero or greater than or equal to the length of the sequence.

`NodeList.length`

The number of nodes in the sequence.

In addition, the Python DOM interface requires that some additional support is provided to allow `NodeList` objects to be used as Python sequences. All `NodeList` implementations must include support for `__len__()` and `__getitem__()`; this allows iteration over the `NodeList` in `for` statements and proper support for the `len()` built-in function.

If a DOM implementation supports modification of the document, the `NodeList` implementation must also support the `__setitem__()` and `__delitem__()` methods.

DocumentType Objects

Information about the notations and entities declared by a document (including the external subset if the parser uses it and can provide the information) is available from a `DocumentType` object. The `DocumentType` for a document is available from the `Document` object's `doctype` attribute; if there is no `DOCTYPE` declaration for the document, the document's `doctype` attribute will be set to `None` instead of an instance of this interface.

`DocumentType` is a specialization of `Node`, and adds the following attributes:

`DocumentType`. **publicId**

The public identifier for the external subset of the document type definition. This will be a string or `None`.

`DocumentType`. **systemId**

The system identifier for the external subset of the document type definition. This will be a URI as a string, or `None`.

`DocumentType`. **internalSubset**

A string giving the complete internal subset from the document. This does not include the brackets which enclose the subset. If the document has no internal subset, this should be `None`.

`DocumentType`. **name**

The name of the root element as given in the `DOCTYPE` declaration, if present.

`DocumentType`. **entities**

This is a `NamedNodeMap` giving the definitions of external entities. For entity names defined more than once, only the first definition is provided (others are ignored as required by the XML recommendation). This may be `None` if the information is not provided by the parser, or if no entities are defined.

`DocumentType`. **notations**

This is a `NamedNodeMap` giving the definitions of notations. For notation names defined more than once, only the first definition is provided (others are ignored as required by the XML recommendation). This may be `None` if the information is not provided by the parser, or if no notations are defined.

Document Objects

A `Document` represents an entire XML document, including its constituent elements, attributes, processing instructions, comments etc. Remember that it inherits properties from `Node`.

`Document`. **documentElement**

The one and only root element of the document.

`Document`. **createElement**(*tagName*)

Create and return a new element node. The element is not inserted into the document when it is created. You need to explicitly insert it with one of the other methods such as `insertBefore()` or `appendChild()`.

`Document.createElementNS(namespaceURI, tagName)`

Create and return a new element with a namespace. The *tagName* may have a prefix. The element is not inserted into the document when it is created. You need to explicitly insert it with one of the other methods such as `insertBefore()` or `appendChild()`.

`Document.createTextNode(data)`

Create and return a text node containing the data passed as a parameter. As with the other creation methods, this one does not insert the node into the tree.

`Document.createComment(data)`

Create and return a comment node containing the data passed as a parameter. As with the other creation methods, this one does not insert the node into the tree.

`Document.createProcessingInstruction(target, data)`

Create and return a processing instruction node containing the *target* and *data* passed as parameters. As with the other creation methods, this one does not insert the node into the tree.

`Document.createAttribute(name)`

Create and return an attribute node. This method does not associate the attribute node with any particular element. You must use `setAttributeNode()` on the appropriate `Element` object to use the newly created attribute instance.

`Document.createAttributeNS(namespaceURI, qualifiedName)`

Create and return an attribute node with a namespace. The *tagName* may have a prefix. This method does not associate the attribute node with any particular element. You must use `setAttributeNode()` on the appropriate `Element` object to use the newly created attribute instance.

`Document.getElementsByTagName(tagName)`

Search for all descendants (direct children, children's children, etc.) with a particular element type name.

`Document.getElementsByTagNameNS(namespaceURI, localName)`

Search for all descendants (direct children, children’s children, etc.) with a particular namespace URI and localname. The localname is the part of the namespace after the prefix.

Element Objects

`Element` is a subclass of `Node`, so inherits all the attributes of that class.

`Element.tagName`

The element type name. In a namespace-using document it may have colons in it. The value is a string.

`Element.getElementsByTagName(tagName)`

Same as equivalent method in the `Document` class.

`Element.getElementsByTagNameNS(namespaceURI, localName)`

Same as equivalent method in the `Document` class.

`Element.hasAttribute(name)`

Return `True` if the element has an attribute named by *name*.

`Element.hasAttributeNS(namespaceURI, localName)`

Return `True` if the element has an attribute named by *namespaceURI* and *localName*.

`Element.getAttribute(name)`

Return the value of the attribute named by *name* as a string. If no such attribute exists, an empty string is returned, as if the attribute had no value.

`Element.getAttributeNode(attrname)`

Return the `Attr` node for the attribute named by *attrname*.

`Element.getAttributeNS(namespaceURI, localName)`

Return the value of the attribute named by *namespaceURI* and *localName* as a string. If no such attribute exists, an empty string is returned, as if the attribute had no value.

`Element.getAttributeNodeNS(namespaceURI, localName)`

Return an attribute value as a node, given a *namespaceURI* and *localName*.

`Element.removeAttribute(name)`

Remove an attribute by name. If there is no matching attribute, a `NotFoundError` is raised.

`Element.removeAttributeNode(oldAttr)`

Remove and return *oldAttr* from the attribute list, if present. If *oldAttr* is not present, `NotFoundError` is raised.

`Element.removeAttributeNS(namespaceURI, localName)`

Remove an attribute by name. Note that it uses a *localName*, not a *qname*. No exception is raised if there is no matching attribute.

`Element.setAttribute(name, value)`

Set an attribute value from a string.

`Element.setAttributeNode(newAttr)`

Add a new attribute node to the element, replacing an existing attribute if necessary if the *name* attribute matches. If a replacement occurs, the old attribute node will be returned. If *newAttr* is already in use, `InuseAttributeErr` will be raised.

`Element.setAttributeNodeNS(newAttr)`

Add a new attribute node to the element, replacing an existing attribute if necessary if the *namespaceURI* and *localName* attributes match. If a replacement occurs, the old attribute node will be returned. If *newAttr* is already in use, `InuseAttributeErr` will be raised.

`Element.setAttributeNS(namespaceURI, qname, value)`

Set an attribute value from a string, given a *namespaceURI* and a *qname*. Note that a *qname* is the whole attribute name. This is different than above.

Attr Objects

`Attr` inherits from `Node`, so inherits all its attributes.

`Attr.name`

The attribute name. In a namespace-using document it may include a colon.

`Attr.localName`

The part of the name following the colon if there is one, else the entire name. This is a read-only attribute.

`Attr.prefix`

The part of the name preceding the colon if there is one, else the empty string.

`Attr.value`

The text value of the attribute. This is a synonym for the `nodeValue` attribute.

NamedNodeMap Objects

`NamedNodeMap` does *not* inherit from `Node`.

`NamedNodeMap.length`

The length of the attribute list.

`NamedNodeMap.item(index)`

Return an attribute with a particular index. The order you get the attributes in is arbitrary but will be consistent for the life of a DOM. Each item is an attribute node. Get its value with the `value` attribute.

There are also experimental methods that give this class more mapping behavior. You can use them or you can use the standardized `getAttribute*()` family of methods on the `Element` objects.

Comment Objects

`Comment` represents a comment in the XML document. It is a subclass of `Node`, but cannot have child nodes.

`Comment.data`

The content of the comment as a string. The attribute contains all characters between the leading `<!--` and trailing `-->`, but does not include them.

Text and CDATASection Objects

The `Text` interface represents text in the XML document. If the parser and DOM implementation support the DOM's XML extension, portions of the text enclosed in CDATA marked sections are stored in `CDATASection` objects. These two interfaces are identical, but provide different values for the `nodeType` attribute.

These interfaces extend the `Node` interface. They cannot have child nodes.

`Text.data`

The content of the text node as a string.

Note: The use of a `CDATASection` node does not indicate that the node represents a complete CDATA marked section, only that the content of the node was part of a CDATA section. A single CDATA section may be represented by more than one node in the document tree. There is no way to determine whether two adjacent `CDATASection` nodes represent different CDATA marked sections.

ProcessingInstruction Objects

Represents a processing instruction in the XML document; this inherits from the `Node` interface and cannot have child nodes.

`ProcessingInstruction.target`

The content of the processing instruction up to the first whitespace character. This is a read-only attribute.

`ProcessingInstruction.data`

The content of the processing instruction following the first whitespace character.

Exceptions

The DOM Level 2 recommendation defines a single exception, `DOMException`, and a number of constants that allow applications to determine what sort of error occurred. `DOMException` instances carry a `code` attribute that provides the appropriate value for the specific exception.

The Python DOM interface provides the constants, but also expands the set of exceptions so that a specific exception exists for each of the exception codes defined by the DOM. The implementations must raise the appropriate specific exception, each of which carries the appropriate value for the `code` attribute.

exception `xml.dom.DOMException`

Base exception class used for all specific DOM exceptions. This exception class cannot be directly instantiated.

exception `xml.dom.DomstringSizeErr`

Raised when a specified range of text does not fit into a string. This is not known to be used in the Python DOM implementations, but may be received from DOM implementations not written in Python.

exception `xml.dom.HierarchyRequestErr`

Raised when an attempt is made to insert a node where the node type is not allowed.

exception `xml.dom.IndexSizeErr`

Raised when an index or size parameter to a method is negative or exceeds the allowed values.

exception `xml.dom.InuseAttributeErr`

Raised when an attempt is made to insert an `Attr` node that is already present elsewhere in the document.

exception `xml.dom.InvalidAccessErr`

Raised if a parameter or an operation is not supported on the underlying object.

exception `xml.dom.InvalidCharacterErr`

This exception is raised when a string parameter contains a character that is not permitted in the context it's being used in by the XML 1.0 recommendation. For example, attempting to create an `Element` node with a space in the element type name will cause this error to be raised.

exception `xml.dom.InvalidModificationErr`

Raised when an attempt is made to modify the type of a node.

exception `xml.dom.InvalidStateErr`

Raised when an attempt is made to use an object that is not defined or is no longer usable.

exception `xml.dom.NamespaceErr`

If an attempt is made to change any object in a way that is not permitted with regard to the [Namespaces in XML](#) recommendation, this exception is raised.

exception `xml.dom.NotFoundErr`

Exception when a node does not exist in the referenced context. For example, `NamedNodeMap.removeNamedItem()` will raise this if the node passed in does not exist in the map.

exception `xml.dom.NotSupportedErr`

Raised when the implementation does not support the requested type of object or operation.

exception `xml.dom.NoDataAllowedErr`

This is raised if data is specified for a node which does not support data.

exception `xml.dom.NoModificationAllowedErr`

Raised on attempts to modify an object where modifications are not allowed (such as for read-only nodes).

exception `xml.dom.SyntaxErr`

Raised when an invalid or illegal string is specified.

exception `xml.dom.WrongDocumentErr`

Raised when a node is inserted in a different document than it currently belongs to, and the implementation does not support migrating the node from one document to the other.

The exception codes defined in the DOM recommendation map to the exceptions described above according to this table:

Constant	Exception
DOMSTRING_SIZE_ERR	DomstringSizeErr
HIERARCHY_REQUEST_ERR	HierarchyRequestErr
INDEX_SIZE_ERR	IndexSizeErr
INUSE_ATTRIBUTE_ERR	InuseAttributeErr
INVALID_ACCESS_ERR	InvalidAccessErr
INVALID_CHARACTER_ERR	InvalidCharacterErr
INVALID_MODIFICATION_ERR	InvalidModificationErr
INVALID_STATE_ERR	InvalidStateErr
NAMESPACE_ERR	NamespaceErr
NOT_FOUND_ERR	NotFoundErr
NOT_SUPPORTED_ERR	NotSupportedErr
NO_DATA_ALLOWED_ERR	NoDataAllowedErr
NO_MODIFICATION_ALLOWED_ERR	NoModificationAllowedErr
SYNTAX_ERR	SyntaxErr
WRONG_DOCUMENT_ERR	WrongDocumentErr

Conformance

This section describes the conformance requirements and relationships between the Python DOM API, the W3C DOM recommendations, and the OMG IDL mapping for Python.

Type Mapping

The IDL types used in the DOM specification are mapped to Python types according to the following table.

IDL Type	Python Type
boolean	bool or int
int	int
long int	int
unsigned int	int
DOMString	str or bytes
null	None

Accessor Methods

The mapping from OMG IDL to Python defines accessor functions for IDL `attribute` declarations in much the way the Java mapping does. Mapping the IDL declarations

```
readonly attribute string someValue;  
attribute string anotherValue;
```

yields three accessor functions: a “get” method for `someValue` (`_get_someValue()`), and “get” and “set” methods for `anotherValue` (`_get_anotherValue()` and `_set_anotherValue()`). The mapping, in particular, does not require that the IDL attributes are accessible as normal Python attributes: `object.someValue` is *not* required to work, and may raise an `AttributeError`.

The Python DOM API, however, *does* require that normal attribute access work. This means that the typical surrogates generated by Python IDL compilers are not likely to work, and wrapper objects may be needed on the client if the DOM objects are ac-

cessed via CORBA. While this does require some additional consideration for CORBA DOM clients, the implementers with experience using DOM over CORBA from Python do not consider this a problem. Attributes that are declared `readonly` may not restrict write access in all DOM implementations.

In the Python DOM API, accessor functions are not required. If provided, they should take the form defined by the Python IDL mapping, but these methods are considered unnecessary since the attributes are accessible directly from Python. “Set” accessors should never be provided for `readonly` attributes.

The IDL definitions do not fully embody the requirements of the W3C DOM API, such as the notion of certain objects, such as the return value of `getElementsByTagName()`, being “live”. The Python DOM API does not require implementations to enforce such requirements.