

Home

Woodstox, the Fast XML-processor

Woodstox is a high-performance validating namespace-aware StAX-compliant (JSR-173) Open Source XML-processor written in Java. XML processor means that it handles both input (== parsing) and output (== writing, serialization)), as well as supporting tasks such as validation.

For the impatient, you can quickly proceed to [Download](#) page; or browse [Documentation](#).

Latest news

- 23-Apr-2012: Woodstox 4.1.3 patch version released (see [Download page](#))
- 26-Aug-2011: Woodstox 4.1.2 patch version released.
- 27-Jan-2011: Woodstox 4.1.1 (and 4.0.10) patch versions released.
- 13-Dec-2010: **Woodstox 4.1.0** released
- 05-May-2010: Maintenance release **4.0.8**: (from [Download page](#)): miscellaneous fixes; plus version 3.0.2 of Stax API jar (to resolve issues with Maven repos)
- 16-Dec-2009: Maintenance release **4.0.7**:: fixes one problem with accessing Base64-encoded binary content (in coalescing mode)
- 01-Oct-2009: Maintenance release **4.0.6**: fixes a nasty bug with long CDATA sections, `XMLStreamReader.getElementText()`.
- 09-Jun-2009: Maintenance release **4.0.5**, fixes minor W3C Schema validation issues
- 08-May-2009: Maintenance releases **4.0.4** ("Page Not Found" release?) and **3.2.9**; latter likely the last release from 3.2 branch.
- 12-Apr-2009: Added related API javadocs (Stax, SAX) for convenience.
- 04-Mar-2009: **Woodstox 4.0.3** released: just a single bug fix, but important one for W3C Schema validation. Also started 4.1 development branch (trunk is for 5.0)
- 25-Feb-2009: **Woodstox 4.0.2** released; 2 bug fixes, bit improved maven deployment (version range deps should finally work)
- 29-Jan-2009: **Woodstox 4.0.1** released to resolve Maven2 repo problems experienced with 4.0.0.
- 01-Jan-2009: **Woodstox 4.0.0**, "**Tequila**" released!
- 26-Dec-2008: **Third 4.0 release candidate** (version 3.9.9-3) released. Bug fixes to StreamSource handling, an API change regarding base64 handling.
- 26-Dec-2008: **3.2.8** maintenance release: work-around for QName problem with old app servers, other minor fixes.
- 17-Dec-2008: **Second 4.0 release candidate** (version 3.9.9-2) released! One critical bug fix; improved packaging, and OSGi service definition added.
- 21-Nov-2008: **First 4.0 release candidate** (version 3.9.9-1) released! Now with full Typed Access API, W3C Schema validation (using MSV), and OSGi compatibility.
- 05-Jun-2007: Released updated versions of [ValidateXML](#) and [DTDFlatten](#) tools: previous versions were based on ancient Woodstox version (2.0.3), new builds are based on 3.2.1.
- 28-Dec-2006: **3.2.0 released** (although it really should be numbered "3.4"... guess why?). The most significant new feature is **full SAX2 API implementation**. In addition, writer-side had bit of TLC given to it, resulting in 10-20% speed increase, as well as numerous fixes.
- 02-Nov-2006: **3.1 (final) released**: implements Xml:id, properly reports SPACE in non-validating mode, and tries to preserve prefix mappings and namespace declarations in repairing mode.
- 11-Aug-2006: Finally added a more recent version (0.9) of [StaxMate](#). This is a significant upgrade, and makes full use of Java 5 features (meaning it also requires JDK 1.5 or above) amongst other things. I will try to write a tutorial for it at [Let's Talk About Stax](#) .
- 07-Aug-2006: **3.0.0 (3.0 final) released**.
- 12-Jun-2006: I am starting to write a blog ([Let's talk about Stax](#)) about Stax, Woodstox, XML in general; this should become a good resource about Woodstox as well as about general Stax issues.

You can also check out full [News](#) for the full record of news events for Woodstox project.

Woodstox features

Woodstox implements StAX (STreaming Api for Xml processing) version 1.0. StAX specifies interface for standard J2ME "pull-parsers" (as opposed to "push parser" like SAX API ones): see [StAX specification](#) for details.

Features of the latest release (from 'current' branch) include:

- Full StAX 1.0 implementation, including all optional features.
- Full namespace (1.0, 1.1; latter with 3.0+) support.
- XML 1.0 and 1.1 compliant (see [XML compatibility page](#) for some discussion on implementation details)
- Support for validation:
 - Full native DTD support, including bi-directional (for both stream readers and writers) validation (writer-side validation with 3.0 and above) validation.
 - RelaxNG validation via Sun MSV (3.0 and above)
 - W3C Schema validation via Sun MSV (3.9 and above)
- Full SAX/SAX2 API implementation, usable directly or via JAXP (3.2 and above)

Features as well as lots of other related information about Woodstox is available from the [Documentation](#) page.

Why use StAX parsers?

StAX parsers are usually a good compromise between convenience offered by tree-based API (DOM, JDom, Dom4j) implementations, and efficiency offered by streaming API (like SAX) implementations.

"As fast as SAX, almost as convenient as DOM" is one way to summarize the benefits.

Why use Woodstox of all available StAX implementations?

Woodstox has following benefits:

- It has most complete and conformant StAX API support of existing implementations.
- It has most complete XML support (including full DTD support, entities, validation, notations) and conformance (which for 2.9 may be second best, after Xerces, of active Java-based xml parsers).
- It is the fastest implementation for most test cases, from small documents to very large documents (tested with 500 MB ones, should handle bigger ones as well).
- It aims to not only detect all XML problems, but to accurately report them (including full location information).
- Beyond plain StAX API, it has the most configurability; from performance settings to convenience ones (including some settings for relaxed verifications). There are even many things one can do to support "almost well-formed" documents (like legacy (X)HTML content), or to do alternate non-compliant processing.

Where can I find sources and binaries?

You can find binaries (jars) and sources (tar, zip) on the [Download](#) page.

Also, Woodstox sources are stored in Codehaus Subversion; you can access them using anonymous read-only access:

```
svn co https://svn.codehaus.org/woodstox/wstx/trunk
```

or, if you want the whole contents of the repository, not just trunk:

```
svn co https://svn.codehaus.org/woodstox
```

and registered developers can access it similarly, but adding "--username" (and "--password") switch to allow changes to be committed back in.

Help for using Woodstox

There are two kinds of support for Woodstox:

- Volunteer support:
 - Woodstox authors and users can be most easily reached via Woodstox [mailing lists](#).
 - Another useful mailing list is the "official" [StAX mailing list](#), which is used for more general discussion regarding Stax specification, and issues common to implementations.
- Professional support:
 - [FasterXML](#) offers professional support service for Woodstox as well as related services.

Companion Projects

Due to both versatility and focus of Woodstox codebase, there are projects that are not included in Woodstox core functionality or package, but that are built on top of it, as separate tools, libraries or applications.

These projects include:

- [StaxMate](#), "the perfect companion for StAX" is an extension that builds on top of raw StAX interface, and adds many convenience features with limited (or, in some cases, negligible) overhead. While it should work over any StAX implementation, it is especially well suited to be used with Woodstox.
- [DTDFlatten](#) is a simple utility that can be used for "flattening" (serializing, pre-processing) of DTDs that consist of multiple physical files. This is often useful for simplicity, performance or debugging reasons. For example, it may be beneficial to create a single physical DTD file for one's customize DocBook flavour, instead of a collection of dozens (or hundreds...) of smaller override files that is needed to cleanly override basic DocBook definitions.
- [ValidateXML](#) is a simple validator tool that uses Woodstox validation methods (currently just DTD) to validate one or more documents. Its main benefits are good error diagnostics, possibility to override document-specific schema settings (validate against different DTDs), and efficient batch validation features.
- [StaxMisc](#) is a loose collection of StAX-utilities, adapters for using StAX with other libraries and frameworks and such, that are not core components of Woodstox nor fall under any other category.

Interesting Related Things

- <https://stax-utils.dev.java.net/> has lots of nifty utilities for integrating StAX with SAX components, amongst other things.
- Sun folks wrote a [comparison of StAX parser performance](#) , including Woodstox as well as the StAX reference implementation. It is interesting reading, although does not include comparison to best SAX parsers.
- Michael Kay (of Saxon fame) [previewed Woodstox](#) along with other fast streaming parsers, and had interesting comments about the subject (and the whole [Saxonica Blog](#) is good reading too).
- Sing Li has written a nice introduction to using Stax2 extension API: "[StAX the odds with Woodstox](#)"

Things That Do Woodstox

- [NUX](#) has support for StAX parsers as xml content source, and has been extensively tested with Woodstox to verify stax builder functionality.
- [SemmlerCode tool](#) (free Code quality improvement tool) [uses](#) Woodstox for xml processing.
- [XFire](#) is based on StAX parsing, and Woodstox is one of tested and suggested implementations to use with it.

Future Plans

List of planned and wished for features can be found from the [Wishlist](#) page.