

TRIPxml v3.1-0

Change History

2018-08-07

Copyright (c) 2018 infinIT Services GmbH

Email: de.tripsupport@infinet-services.de

Web: <http://infinet-services.de>

End User License Agreement

All rights to this software, its documentation and logotypes of the TRIP product family and software (altogether "Software") supplied by infinIT Services GmbH (infinIT) are exclusively owned by infinIT.

The transfer of this Software, solutions or parts thereof requires the prior written agreement of infinIT. Furthermore, the customer has the right to use licensed Software and / or process solutions supplied by infinIT to the extent specified in his contract with infinIT.

The free-to-use non-commercial version doesn't require a prior written agreement with infinIT but such customers, organizations and/or third parties agree by using the software and / or solution of infinIT to be strongly obliged to keep all rights to this software, documentation and logotypes of the TRIP product family absolutely un infringed and protected.

Table of Contents

About This Document.....	5
Version 3.1-0.....	5
News in v3.1-0.....	5
XPath parser restructuring [XML-50].....	5
Corrected in v3.1-0.....	5
Problem installing TRIPxml on Linux [TRIP-145].....	5
XPath may crash [XML-51].....	5
Version 3.0-3.....	5
News in v3.0-3.....	5
Support for TRIPsystem 7 [XML-49].....	5
Corrected in v3.0-3.....	5
Version 3.0-2.....	5
News in v3.0-2.....	5
Corrected in v3.0-2.....	5
Memory leak when retrieving XML fragment sets [XML-48].....	6
XPath comparison predicates may result in zero hits [XML-45].....	6
Version 3.0-1.....	6
News in v3.0-1.....	6
Corrected in v3.0-1.....	6
Invalid XML fragment sets from documents with namespaces [XML-46].....	6
XPath comparison predicates may result in zero hits [XML-45].....	6
Version 3.0-0.....	6
News in v3.0-0.....	6
TRIPxml namespace renamed [XML-38].....	6
XPath as Query Language [XML-26].....	6
XML Fragment Retrieval [XML-1].....	7
Corrected in v3.0-0.....	7
Installation fails if %TEMP% and install dir is not on the C: drive [XML-44].....	7
Invalid character entity replacement for double quote [XML-37].....	7
Crash in validation against DTD stored in database [XML-36].....	7
Version 2.2-2.....	7
News in v2.2-2.....	7
Updated XML parser libraries [R1135].....	7
Unicode database support [R1137].....	7
Corrected in v2.2-2.....	8
Crash when updating an XML record [B1127].....	8
XML highlight markup invalid in attribute values [B1134].....	8
Character entities break words into multiple part records [B1142].....	8
Special characters not encoded during XML reconstruction [B1163].....	8
Version 2.2-1.....	8
News in v2.2-1.....	8
Corrected in v2.2-1.....	8
Memory deallocation error in import filter [B0752].....	8
Version 2.2-0.....	8

News in v2.2-0.....	8
Updated XML parser libraries [R0587].....	8
Support for HP-UX [R0698].....	9
Corrected in v2.2-0.....	9
Memory leaks in XML parsing routine [B0697].....	9
Multibyte characters removed by reconstructive export [B0700].....	9
Version 2.1-0.....	9
News in v2.1-0.....	9
More Encodings.....	9
Entity References.....	9
Refurbished Storage Algorithm.....	9
Stream Transfer.....	9
XML Schema Support.....	10
Improved Validation.....	10
New C APIs.....	10
Corrected in v2.1-0.....	10
Reconstruction With Original Encoding.....	10
Broken Reconstruction.....	10
Txlinker Only Worked With Default Credentials.....	10
Txlinker Overloaded TRIP Daemon Job Queue.....	10
Updlinks Only Worked With Default Credentials.....	10
Mklinkdb Missing Under Unix.....	11
Upgrade Tool XRS Broken Under Unix.....	11
Invalid Returns From Filter tripxmlgetlinks On Failure.....	11
Version 2.0-1.....	11
News in v2.0-1.....	11
Corrected in v2.0-1.....	11
Attribute Paths.....	11
URL Alias Generation.....	11
Known Remaining Issues.....	11
TXlinker link database update crashes on AIX [XML-43].....	11
Streamed I/O hangs on AIX [XML-42].....	11
Loss of detail in internal DTD [XML-23].....	11
Loss of DOCTYPE-declaration without PUBLIC/SYSTEM [XML-22].....	12
Internal Copy of External DTD on Reconstructive Export [XML-21].....	12
Date Extraction [XML-20].....	12
Line Break Replacement [XML-19].....	12
Block-oriented import and export [XML-18].....	12
Default values and entities turn explicit on reconstructive export [XML-17].....	12
Streamed I/O on multiprocessor machines [XML-16].....	12

About This Document

This document contains the change history for TRIPxml.

Version 3.1-0

News in v3.1-0

XPath parser restructuring [XML-50]

The XPath parser has been overhauled for maintainability purposes. No changes to functionality have been made.

Corrected in v3.1-0

Problem installing TRIPxml on Linux [TRIP-145]

The script for installing TRIP on Linux would sometimes fail to properly identify the TRIPsystem installation, causing the installation to fail. The error message would in this case also be unclear.

IXPath may crash [XML-51]

The IXPath command line tool may crash if the database being opened with it is not an XML database or does not exist.

Version 3.0-3

News in v3.0-3

Support for TRIPsystem 7 [XML-49]

TRIPxml can now be used with TRIPsystem version 7.0 and later.

Corrected in v3.0-3

There are no corrected issues in version 3.0-3.

Version 3.0-2

News in v3.0-2

There are no new features added in version 3.0-2.

Corrected in v3.0-2

Memory leak when retrieving XML fragment sets [XML-48]

A memory leak occurred in TRIPxml when retrieving XML fragment sets from XML documents with namespaces.

XPath comparison predicates may result in zero hits [XML-45]

The fix for this issue as released in version 3.0-1 caused several types of predicate comparisons to stop working. TRIPxml would return error messages like 'Location step expressions for axis "self" are not yet supported.'

Version 3.0-1

News in v3.0-1

There are no new features added in version 3.0-1.

Corrected in v3.0-1

Invalid XML fragment sets from documents with namespaces [XML-46]

When retrieving a fragment set documents with namespaces, namespace prefixes were inserted without their corresponding definitions. This resulted in an invalid returned fragment set document.

XPath comparison predicates may result in zero hits [XML-45]

An XPath expression with a predicate in the style `//MyElement[MyOtherElement='myvalue']` would result in zero hits while its equivalent `//MyElement/MyOtherElement[text()='myvalue']/parent::MyElement` worked fine.

Version 3.0-0

News in v3.0-0

TRIPxml namespace renamed [XML-38]

The namespace for the highlighting tags has been renamed to "http://www.tieto.com/trip/xml". This namespace also applies to the XPath fragment sets, a feature introduced in this version.

XPath as Query Language [XML-26]

The XPath language can now be used to query a TRIPxml database. APIs for access of this feature have been added to TRIPnxp and TRIPjxp version 2.1-0.

The use of this functionality requires a Unicode session and a TRIPsystem version 6.2-7 or later.

XML Fragment Retrieval [XML-1]

It is now possible to retrieve XML fragments from a search set containing XML records. The node set that form the roots of the fragments is specified by an XPath expression.

APIs for access of this feature have been added to TRIPnpx and TRIPjxp version 2.1-0.

The use of this functionality requires a Unicode session and a TRIPsystem version 6.2-7 or later.

Corrected in v3.0-0

Installation fails if %TEMP% and install dir is not on the C: drive [XML-44]

The installation program for Windows failed if the %TEMP% environment variable and the installation directory are not on the C: drive. The error message was "Unable to move the replacement file to the file to be replaced. The file to be replaced has retained its original name".

(changed in version 3.0-0:1)

Invalid character entity replacement for double quote [XML-37]

When reconstructing an XML document (e.g. for applying highlight markup), TRIPxml replaces various characters with their character entity equivalents. The entity replacement for the double quote was incorrectly spelled, which resulted in an unreadable result.

Crash in validation against DTD stored in database [XML-36]

The XML parser could crash when releasing memory allocated by TRIPxml. The memory was used for temporary storage of a DTD automatically exported from a TRIPxml database in the process of validating an XML document.

Version 2.2-2

News in v2.2-2

Updated XML parser libraries [R1135]

The third party libraries used for XML parsing have been updated to Xerces 3.0 and ICU 4.0.

Unicode database support [R1137]

TRIPxml databases can now be unicode-enabled. A template database design for a unicode-enabled XML database is provided (unixml.def for use via TRIP classic, and unixml_def.xml for use via TRIP manager).

IMPORTANT: When using unicode XML databases, the TRIP session must always be set up to use the UTF-8 character set (a.k.a. a unicode session). If this is not done, TRIPxml will not perform its character transcoding operations correctly, likely resulting in corrupted and unsearchable text being stored in the database.

Corrected in v2.2-2

Crash when updating an XML record [B1127]

The TRIPxml routine for updating a previously imported XML record caused the process to crash. This problem primarily affected the Windows version of TRIPxml.

XML highlight markup invalid in attribute values [B1134]

Search hit markup <TRIPHLBEGIN> and <TRIPHLEND> was incorrectly applied to attribute nodes; only text nodes are valid locations.

Character entities break words into multiple part records [B1142]

Storing XML data where the text contains character entities would cause TRIPxml to store the text into as many part records as there were character entities in the text. For example, the word "täysistunto" would be stored as a single "t" in one part record and the rest of the word in the next part record. In other words, it would be impossible to find the complete word in this record.

Special characters not encoded during XML reconstruction [B1163]

Special characters in text nodes, such as the less-than sign '<', were not encoded during XML reconstruction. This resulted in invalid XML documents.

Version 2.2-1

News in v2.2-1

There are no new features added in version 2.2-1.

Corrected in v2.2-1

Memory deallocation error in import filter [B0752]

A memory deallocation call in the TRIPxml import filter used the wrong heap and caused a crash. This problem only affected the Windows platform.

Version 2.2-0

News in v2.2-0

Updated XML parser libraries [R0587]

The third party libraries used for XML parsing have been updated to Xerces 2.6 and ICU 3.2.

Support for HP-UX [R0698]

TRIPxml release 2.2-0 adds support for the HP-UX 11.0 platform.

Corrected in v2.2-0

Memory leaks in XML parsing routine [B0697]

The XML parser deinitialization leaked 4096 bytes per indexed document.

Multibyte characters removed by reconstructive export [B0700]

Reconstructive export would incorrectly trim off some multibyte characters. This affected encodings such as gb2312 and others where valid bytes may exist outside the ASCII printable character range.

Version 2.1-0

News in v2.1-0

More Encodings

Support for more than 200 different character encodings has been added. Among others are Cyrillic (ISO-8859-5), Arabic (ISO-8859-6), Greek (ISO-8859-7), Hebrew (ISO-8859-8), Traditional Chinese (Big5), and Simplified Chinese (GBK).

See the standard documentation for a complete list.

Entity References

Entity references are now resolved. It is therefore now possible to search in the content of an entity.

If there is a section in an xml document like this:

```
<author>&myname;</author>
```

And the &myname; entity reference resolves to "Smith", then you can issue a CCL find order like "FIND author=Smith".

Refurbished Storage Algorithm

The algorithm used to parse and store documents has been improved with regard to speed. Large documents will now process a lot faster. In order to further maximize performance, it is strongly suggested that TRIPsystem v3.5 is used.

Stream Transfer

Stream import/export functionality is now provided. Suitable for very large documents, this takes a lot less of memory resources than doing a memory-based import/export (which is done via TRIP@client or TRIPjtk).

XML Schema Support

Support for XML schema based validation has been added. To enable schema based validation, specify the XML_SCHEMA setting in the server-side TRIPrcs file. See appendix C in the users guide for more details.

Improved Validation

The validation algorithm can now load DTD and XML schema documents from the web, via stream transfer, from local hard drive, or use an already imported version of the requested document.

New C APIs

A new set of C API functions have been created for use with TRIP toolkit (delivered with TRIPsystem and TRIP@client). There are both server and client versions of the new APIs, making it possible to use them with either TRIPsystem or TRIP@client.

Corrected in v2.1-0

Reconstruction With Original Encoding

Reconstructive export, i.e. when the document is put together from its parts, would previously only produce ISO-8859-1 encoded text, no matter what encoding the original document was encoded in.

Now the document will be exported in the original encoding. Note that in order for this to work, the document must have been imported using v2.1-0 of TRIPxml.

Broken Reconstruction

Reconstructive export now no longer produce corrupt result for documents containing empty elements with attributes or has extensive usage of namespaces.

Txlinker Only Worked With Default Credentials

The link support tool txlinker would not accept username and password arguments. These have now been added.

Txlinker Overloaded TRIP Daemon Job Queue

This occurred when txlinker removed several hundred dead link records, i.e. link records whose corresponding XML database records are removed.

Updlinks Only Worked With Default Credentials

The link support tool updlinks ignored the username and password arguments.

Mklinkdb Missing Under Unix

The link support tool mklinkdb was missing from non-Windows builds.

Upgrade Tool XRS Broken Under Unix

The database version upgrade utility program XRS had a flaw that caused it to crash.

Invalid Returns From Filter tripxmlgetlinks On Failure

The return buffer would in some cases be NULL, which caused the packet encoder routine in the tbserver to fail when the filter was called from a client tool (TRIP@client or TRIPjtk).

Version 2.0-1

News in v2.0-1

There are no new features added in version 2.2-1.

Corrected in v2.0-1

Attribute Paths

The field N_PATH that is used for element path searches were not provided with a value for attribute nodes, essentially making attributes unsearchable.

URL Alias Generation

The URL alias value is, among other things, used for link resolution. In some cases TRIPxml would produce a URL alias missing a directory delimiter, effectively disabling functionality depending on a proper URL alias.

Known Remaining Issues

TXlinker link database update crashes on AIX [XML-43]

Using the link database on an AIX does not work. The txlinker program which updates the TRIPxml link database crashes.

Streamed I/O hangs on AIX [XML-42]

Importing XML documents via a streamed I/O does not work when the server is an AIX. The host process (e.g. tbserver) will stop responding.

Loss of detail in internal DTD [XML-23]

When XML documents specifies an internal DTD, not everything in it may be parsed

by TRIPxml. This is especially true when there is also an external DTD.

The recommendation is to avoid internal DTDs when possible and only use external ones. If that is not an option, then there are two things to remember:

- Do not disable the storage of the original document in the D_XMLDOC field.
- Do not retrieve the document using reconstructive export. Only retrieve the contents of the D_XMLDOC field.

Loss of DOCTYPE-declaration without PUBLIC/SYSTEM [XML-22]

A doctype declaration without PUBLIC or SYSTEM (for example `<!DOCTYPE article>`) will not be parsed.

Avoid using such declarations if possible.

Internal Copy of External DTD on Reconstructive Export [XML-21]

This occurs if the document has been imported with validation turned on.

While this is mostly harmless, it may be an issue in cases where the DTD has changed significantly.

Date Extraction [XML-20]

The algorithm for date extraction may sometimes interpret a non-date number sequence as a date. This is harmless, but may result in storage of dates that are valid, but is essentially garbage. This will not affect reconstruction of the document.

Line Break Replacement [XML-19]

On Windows platforms, the filter routine makes sure that line breaks are represented by pairs of carriage return and line feed when the XML document is reconstructed. Sometimes, line breaks can be omitted from conversion, resulting in a document where line breaks follow two standards.

Block-oriented import and export [XML-18]

Block-oriented import and export may not always produce the desired results. If block-oriented I/O is not critical, it is recommended one of the other techniques for import and export be used (see section "Use the API calls" in the Users Guide for details).

Default values and entities turn explicit on reconstructive export [XML-17]

When doing a reconstructive export of a document with entities or whose schema specifies default values (for text nodes, attribute values, etc), these turn explicit when exporting the document.

Note: Reconstructive export is ALWAYS done if the document has been imported via stream import and when highlighting is requested.

Streamed I/O on multiprocessor machines [XML-16]

Importing XML documents with streamed I/O sporadically crash the import process on Solaris running on multiprocessor SPARC machines.