

Working with Rational ClearCase Unicode Type Manager

Skill Level: Intermediate

[Manivannan Veeramani \(manivannan.veeramani@in.ibm.com\)](mailto:manivannan.veeramani@in.ibm.com)
Software Engineer
IBM

[Gopikrishnan Sidhardhan \(gsidhard@in.ibm.com\)](mailto:gsidhard@in.ibm.com)
Advisory Software Engineer
IBM

27 Aug 2009

Get an overview of the IBM® Rational® ClearCase® Unicode Type Manager (UTM) introduced in Version 7.1, which helps you get started quickly by using the same version of Unicode files.

Unicode is the widely used industry standard for representing characters independent of the underlying program and language. With software being developed and deployed worldwide, Unicode is being used pervasively. Therefore, specifying versions of files with Unicode data is essential for software development.

From IBM® Rational® ClearCase®; Version 7.1, versioning Unicode data has become simple and efficient, with the inclusion of a new Unicode Type Manager. This article provides an overview of this new type manager and explains how you can use it to start versioning Unicode data. This article also explains the Unicode types supported, the differences between the new type manager and the existing text file manager, and how to convert existing text files to UTM-supported files.

Warning:

Do not implement this new element type in a replicated environment if not all sites are at least at ClearCase 7.1.x. As this feature is new in ClearCase 7.1 you are unable to import a packet which creates a UTF-8 file element on 7.0.1.x VOB servers since that feature is not supported in that version.

Unicode Type Manager overview

Elements in a ClearCase Versioned Object Base (VOB) have an element type associated with them. There are several predefined element types provided by default (the number depends on the Rational ClearCase release), such as `file`, `text_file`, and `compressed_file`. Each element type is bound to a type manager that assists ClearCase in handling different operations, such as creating a new version or reading a specific version. A type manager, in short, is a suite of programs that manipulates files with a particular data format. Different type managers process files with different formats.

The UTM assists ClearCase in manipulating files that are in Unicode format. It is based on the same algorithm as that of `text_file_delta` type manager. The usability remains the same. The Unicode Type Manager supports the following file types in addition to ASCII:

- UTF-8
- UTF-16LE
- UTF-16BE
- UTF-32LE
- UTF-32BE

Advantages of using the UTM

- Unicode is the pervasively used data format.
- UTM enables differentiating and merging two or more Unicode files.
- Storage of Unicode files is efficient, because the UTM stores only the differences between the versions of a file.
- It supports all languages supported by UTF-8, UTF 16-LE, UTF 16BE, UTF 32LE, and UTF32BE.
- None of the existing type managers support Unicode data.

How it works

A new line-oriented difference-finding tool that supports Unicode files evolved in Java™. The underlying algorithm of the new tool was similar to that of the existing `text_file_delta`. Other interfaces were developed around this differentiating tool to support all of the type manager operations, and they were packaged as an executable file. Five new map file entries for each type of the Unicode type managers were created (UTF-8, 16LE, 16BE, 32LE, and 32BE), along with the

associated operation and program.

The information for mapping the type manager name to operations and to the program is stored in a map file on a Microsoft® Windows® system; whereas, it is created as a symbolic link in a UNIX® or similar operating system. The entries for the Unicode file `delta` also can be found in the map file.

The map file in Windows is located in the `<clearcase-home-dir>\lib\mgrs` directory and has the information shown in Table 1 for Unicode files, where XX is either 8, 16le, 16be, 32le, or 32be.

Table 1. Mapping information for Unicode files

Type Manager	Operation	Program
utfXX_file_delta	construct_version	..\bin\utfXXfdmgr.exe
utfXX_file_delta	create_branch	..\bin\utfXXfdmgr.exe
utfXX_file_delta	create_element	..\bin\utfXXfdmgr.exe
utfXX_file_delta	create_version	..\bin\utfXXfdmgr.exe
utfXX_file_delta	delete_branches_versions	..\bin\utfXXfdmgr.exe
utfXX_file_delta	compare	..\bin\utfXXcleardiffmrg.exe
utfXX_file_delta	xcompare	..\bin\ucleardiffmrgutfXX.exe
utfXX_file_delta	merge	..\bin\utfXXcleardiffmrg.exe
utfXX_file_delta	xmerge	..\bin\ucleardiffmrgutfXX.exe
utfXX_file_delta	annotate	..\bin\utfXXfdmgr.exe
utfXX_file_delta	get_cont_info	..\bin\utfXXfdmgr.exe

On Linux® and UNIX systems, the `utf8_file_delta` directory is created under `/usr/atria/lib/mgrs`. For each of the operations previously mentioned, there is a file entry that points to the corresponding program.

Overview of executable files

Two new Java™ Archive (JAR) files and executable files are introduced as part of the new UTM. They can be found in the following paths, where XX is either 8, 16le, 16be, 32le, or 32be.

Windows

- `<clearcase-home-dir>\bin\ucleardiffmerge.jar`
- `<clearcase-home-dir>\bin\unicodetypemgr.jar`

- <clearcase-home-dir>\clearcase\bin\utfXXdeltamgr.exe
- <clearcase-home-dir>\clearcase\bin\utfXXcleardiffmerge.exe

Systems similar to the UNIX system

- /opt/rational/clearcase/bin/ucleardiffmerge.jar
- /opt/rational/clearcase/bin/unicodetypemgr.jar
- /opt/rational/clearcase/bin/utfXXdeltamgr.exe
- /opt/rational/clearcase/bin/utfXXcleardiffmerge.exe

Getting started

Follow these steps to get started with the new type manager:

1. Create a new element type (this is usually done by an administrator).
2. Create an element, and associate it with the new element type.
3. Run any ClearCase command.

Create an element type

For the 7.1 release, there is no default Unicode element type provided in Rational ClearCase. Users need to create this type and associate it with the Unicode Type Manager. The super type for this element will be the 'file' type.

For each Unicode encoding, you need an element type in Rational ClearCase, as Table 2 shows.

Table 2. Element types

Encoding	Element type
UTF-8	utf8_file
UTF-16 Big Endian	utf16be_file
UTF-16 Little Endian	utf16le_file
UTF-32 Big Endian	utf32be_file
UTF-32 Little Endian	utf32le_file

Creating an element type necessitates mentioning the name of the element type that has to be created, along with the type manager that it has to be associated with and the super type of the element type.

For example:

```
cleartool mkeltype -supertype file -manager utfXX_file_delta utfXX_file
```

Where:

XX can be any of 8, 16le, 16be, 32le, and 32be

utfXX_file is the name of the element type

utfXX_file_delta is the name of the type manager that is associated with it

file is the super type

Create an element

After you have created the element type, the next step is to create an element and associate it with the preferred element type.

For example:

```
cleartool mkelem -eltype utfXX-file -ncomments foo.c
```

Operations supported

Table 3 shows the operations that are supported.

Table 3. Description of operations supported

Operation	Description
construct_version	Used to construct a cleartext copy of a particular version
create_branch	Used to create a new branch in an element's version tree
create_element	Used to create an element's initial data container
create_version	Used to create a new version of an element
delete_branches_versions	Used to delete a branch/version of an element
Compare	Used when comparing from command line
Xcompare	Used when comparing from GUI
Merge	Used when merging from command line
Xmerge	Used when merging from GUI
Annotate	Used to produce an annotated listing of a

	version's contents
get_cont_info	Used to get information on the contents of a container

Using the following ClearCase commands on an element invokes the type manager:

- mkelem
- mkbranch
- checkin
- rmver
- rmbranch
- annotate
- diff
- merge

Differentiating and merging operations

You can perform the "diff" (or "differencing") and merge operations for Unicode files by using either `cleartool diff` or `utfXXcleardiffmerge` commands. `Cleartool diff` is supported only for versioned elements; however, you can use `utfXXcleardiffmerge` for versioned or nonversioned files. Both the diff and merge operations can support comparing two or more files, with a limit of 32.

Sample command

Diff

```
cleartool diff file1 file2
utfXXcleardiffmerge file1 file2
```

Merge

```
utfXXcleardiffmerge -out mergedfile file1 file2
cleartool merge -out mergedfile file1 file2
```

Where `XX` can be any of these: 8, 16le, 16be, 32le, and 32be.

End-of-line characters

Because UTM performs a line-oriented file comparison, segregating a line becomes important. The line segregation is based on recognizing an end-of-line (EOL) character. The following characters are treated as the end-of-line characters for Unicode:

- Carriage Return
- Line Feed
- Carriage Return + Line Feed*
- Form Feed
- Line Separator
- Paragraph Separator
- New Line

*CR+LF is higher-priority than CR.

Two lines with identical data but with different EOL characters are still reported as a difference. To avoid the difference, use the `blanks_ignore` option. For example:
`utf8cleardiffmerge -blanks_ignore file2 file2`

Graphical diff and merge

The graphical utility for UTM is included in the Rational ClearCase 7.1 release. The executable file is named `ucleardiffmrgutfXX.exe`. You can invoke the utility by using either the `cleartool` command or by calling the stand-alone `ucleardiffmrgutfXX` executable file. However, the former will work only with ClearCase elements.

Sample command

Diff

```
cleartool diff -graphical file1 file2
ucleardiffmrgutfXX file1 file2
```

Merge

```
cleartool merge -out mergedfile file1 file2
ucleardiffmrgutfXX -out mergedfile file1 file2
```

Where `XX` can be any of these versions: 8, 16le, 16be, 32le, or 32be.

Comparison with `text_file_delta`

Usability

The use and options provided for UTM are similar to that of `text_file_delta`. Therefore, you can have a seamless transition from `text_file_delta` to UTM.

Support

UTM supports files with Unicode and ASCII data; however, `text_file_delta` supports only ASCII format.

Limitations

- UTM does not support files with extended ASCII data.
- Using `utfXXcleariffmerge` to perform a diff or merge operation on files of different Unicode types produces undesirable results. For example: `utf8cleariffmerge <utf8_file> <utf16le_file>`
- The maximum size of a line supported by UTM is 8000 bytes (see Note). If a line exceeds 8000 bytes, it fails.

Note:

The maximum of 8000 bytes of data is applicable across all of UTF-8, -16, and -32, where multibytes represent a single character. Therefore, the actual number of characters supported in a line might not be 8000 characters.

Resources

Learn

- Learn more about ClearCase:
 - For technical information and other resources for developers, see the [Rational ClearCase product page](#).
 - Visit the [Rational ClearCase area on developerWorks](#) for articles and tutorials.
 - Find answers to terminology questions in the [ClearCase Glossary](#).
- Explore the [Software Configuration Management Information Center](#) to learn how features such as local and remote access, a proven use model, a wide range of supported environments, transparent access to files, and parallel development support give your team members instant, controlled access to the project assets that they need to create, update, build, reuse, and maintain your software.
- Learn about other applications in the [IBM Rational Software Delivery Platform](#), including collaboration tools for parallel development and geographically dispersed teams, plus specialized software for architecture management, asset management, change and release management, integrated requirements management, process and portfolio management, and quality management. You can find product manuals, installation guides, and other documentation in the [IBM Rational Online Documentation Center](#).
- Visit the [Rational software area on developerWorks](#) for technical resources and best practices for Rational Software Delivery Platform products.
- Explore [Rational computer-based, Web-based, and instructor-led online courses](#). Hone your skills and learn more about Rational tools with these courses, which range from introductory to advanced. The courses on this catalog are available for purchase through computer-based training or Web-based training. Additionally, some "Getting Started" courses are available free of charge.
- Subscribe to the [Rational Edge newsletter](#) for articles on the concepts behind effective software development.
- Subscribe to the [IBM developerWorks newsletter](#), a weekly update on the best of developerWorks tutorials, articles, downloads, community activities, webcasts and events.
- Browse the [technology bookstore](#) for books on these and other technical topics.

Get products and technologies

- Download [trial versions of IBM Rational software](#).
- Download these [IBM product evaluation versions](#) and get your hands on application development tools and middleware products from DB2®, Lotus®, Tivoli®, and WebSphere®.

Discuss

- Check out [Rational ClearCase forum](#) to post questions, find answers, and share solutions about configuration management and Rational ClearCase.
- Check out [developerWorks blogs](#) and get involved in the [developerWorks community](#).

About the authors

Manivannan Veeramani

Manivannan Veeramani has worked with IBM Rational Change and Configuration Management team for the past two years. He was a developer on the team that developed the Unicode Type Manager.

Gopikrishnan Sidhardhan

Gopikrishnan Sidhardhan has more than 10 years of industry experience, and three years with IBM. He was a developer on the team that developed the Unicode Type Manager.

Trademarks

IBM, Rational, and the IBM logo are trademarks of International Business Machines Corporation in the United States, other countries or both.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.