NAME

ustar – Pure Tcl USTAR Archive Reader

SYNOPSIS

(If package installed globally)

```
package require ustar
```

(Otherwise)

```
source ustar.tcl
```

`::ustar::contents` *stream*

`::ustar::file_contents` *path*

`::ustar::gzip_contents` *path*

`::ustar::extract` *stream callback*

`::ustar::file_extract` *path callback*

`::ustar::gzip_extract` *path callback*

`::ustar::format_contents` *contents*

DESCRIPTION

This module provides a simple interface for reading USTAR archive files, to either enumerate their contents or extract them completely. The following procedures are provided.

`::ustar::contents` *stream*

Scan the archive by reading data from *stream* (which must be a stream channel open in binary mode). The stream is not closed at the end of the operation. This returns a list of metadata values, one per file, in the order in which they appear in the archive. Each is a list of 13 values:

size The length, in bytes, of the file. Things without lengths (such as directories) will have a zero in this field.

type The file type. This is a single character exactly as found in the USTAR header field. A null byte or ASCII `0` digit indicates a regular file. Other than null, the type characters will be ASCII alphanumeric characters. A `1` indicates the file is a hard link to another file already recorded in the archive. A `2` indicates that this entry describes a symbolic link. Types `3`, `4`, `5`, and `6` indicate character devices, block devices, directories, and FIFOs respectively. Other type values may be used to designate vendor-specific custom types of files.

name The full pathname of the file. This might contain arbitrary Unicode characters.

mode The binary file mode as described in `lstat`(2).

uid The numeric user ID of the file's owner.

uname If provided in the archive, the user name of the file's owner.

gid The numeric group ID of the file's owner.

gname If provided in the archive, the group name of the file's owner.

mtime The file's modification time as a UNIX `time_t` value. If the archive supports it, this may be a floating-point value giving fractional seconds.

link For links, this is the pathname of the file being linked to.

major For devices, this is the numeric major device number. Otherwise zero.

minor For devices, this is the numeric minor device number. Otherwise zero.

attrs If any extended attributes were given for the file, they are listed here. This is a list with an even number of elements, which are taken in pairs as the name of the attribute followed by its value. These are completely arbitrary. Any which map directly to the above metadata fields are automatically updated in the other

fields, but regardless of that, the entire extended attribute set appears here.

`::ustar::file_contents` *path*
: As an convenience function, this opens *path* as a binary file and then calls `contents` on it, returning the result. This does close the file when finished reading from it.

`::ustar::gzip_contents` *path*
: As with `file_contents` but decompresses the data as described for `gzip_extract` below.

`::ustar::extract` *stream callback*
: This reads the archive data from the open binary file stream channel *stream* like the `contents` procedure does, but it also extracts the actual file data in addition to the metadata for each file. As each file is read, the supplied *callback* procedure is invoked to handle the disposition of the file's data, so the caller can decide whether to process the data in-memory or save it somewhere, perhaps after other processing takes place. The callback is invoked with two arguments: the metadata as already described above for the `contents` procedure, and the binary data of the file itself.

`::ustar::file_extract` *path callback*
: As an convenience function, this opens *path* as a binary file and then calls `extract` on it, saving the caller from the trouble of opening and closing the archive.

`::ustar::gzip_extract` *path callback*
: As with `file_extract` but also decompresses the contents of *path* while reading from it. This uses a `gzip`(1)-compatible decompresion suitable for tar archives created with `tar`'s `-z` option.

`::ustar::format_contents` *contents*
: Given a list of metadata values exactly as returned by the `contents` procedure, this returns a multi-line string with that information formatted in a human-readable presentation. Not all extended attributes are included.

DIAGNOSTICS

An exception is thrown if the archive is unable to be processed for reasons such as unexpected end of file, unparseable header fields, checksum verification failures, etc.

SEE ALSO

`gzip`(1), `lstat`(2), `pax`(1), `tar`(1).

AUTHOR

Steve Willoughby / steve@madscience.zone.

HISTORY

Initial version created 16 July 2020.

BUGS

Does not support base-256-encoded numeric values in headers, nor global extended header blocks. Neither of those is commonly found enough to prioritize implementation but that may change in the future.

COPYRIGHT