NAME

xmlsimple − add facilities for writing simple one-line scripts with the gawk-xml extension, and also simplify writing more complex scripts.

SYNOPSIS

```
@include "xmlsimple"

parentpath = XmlParent(path)
test = XmlMatch(path)
scopepath = XmlMatchScope(path)
ancestorpath = XmlMatchAttr(path, name, value, mode)

XmlGrep()
```

DESCRIPTION

The *xmlsimple* library facilitates writing simple one-line scripts based on the gawk-xml extension. Also provides higher-level functions that simplify writing more complex scripts. It is an alternative to the *xmllib* library. A key difference is that **$0** is not changed, so *xmlsimple* is compatible with awk code that relies on the *gawk-xml* core interface.

Short token variable names

To shorten simple scripts, *xmlsimple* provides two-letter named variables that duplicate predefined token-related core variables:

XD Equivalent to XMLDECLARATION.

SD Equivalent to XMLSTARTDOCT.

ED Equivalent to XMLENDDOCT.

PI Equivalent to XMLPROCINST.

SE Equivalent to XMLSTARTELEM.

EE Equivalent to XMLENDELEM.

TX Equivalent to XMLCHARDATA.

SC Equivalent to XMLSTARTCDATA.

EC Equivalent to XMLENDCDATA.

CM Equivalent to XMLCOMMENT.

UP Equivalent to XMLUNPARSED.

EOI Equivalent to XMLENDDOCUMENT.

Collecting character data

Character data items between element tags are automatically collected in a single **CHARDATA** variable. This feature simplifies processing text data interspersed with comments, processing instructions or CDATA markup.

CHARDATA

Available at every **XMLSTARTELEMENT** or **XMLENDELEMENT** token. Contains all the character data since the previous start- or end-element tag.

Whitespace handling

The **XMLTRIM** mode variable controls whether whitespace in the **CHARDATA** variable is automatically trimmed or not. Possible values are:

XMLTRIM = 0

Keep all whitespace

XMLTRIM = 1 (default)
Discard leading and trailing whitespace, and collapse contiguous whitespace characters into a single space char.

XMLTRIM = -1
Just collapse contiguous whitespace characters into a single space char. Keeps the collapsed leading or trailing whitespace.

Record ancestors information
The **ATTR** array variable automatically keeps the attributes of every ancestor of the current element, and of the element itself.

ATTR[path@attribute]
Contains the value of the specified *attribute* of the ancestor element at the given *path*.

Example

While processing a `/books/book/title` element, `ATTR["/books/book@on-loan"]` contains the name of the book loaner.

Path related functions
A fixed path is a slash delimited list of direct child elements (`/name/name/...`). A path expression accepts also an asterisk (`*`) to match any name, and a double slash (`//`) to represent a descendant at any level. An absolute path starts with a slash (path from the root element). A relative path without a leading slash can start at any level (path from some ancestor).

XmlParent(path)
Returns the path of the parent element. I.e., the *path* argument without the last `/name` part. The path argument is optional. If not given the **XMLPATH** is used.

XmlMatch(path)
Tests whether the current **XMLPATH** matches the *path* expression argument, anchored at the end.

XmlMatchScope(path)
Returns the **XMLPATH** prefix not matched by the matching *path* expression argument. Returns a null value if there is no match.

XmlMatchAttr(path, name, value, mode)
Returns the path of the innermost ancestor that matches the *path* argument and also has a *name* attribute with the given *value*. The mode argument is optional. If non-null then the value is handled as a regular expression instead of a fixed value.

Grep-like facilities
XmlGrep()
If invoked at the **XMLSTARTELEM** event, causes the whole element subtree to be copied to the output.

NOTES

The *xmlsimple* library includes both the *xmlbase* and *xmlcopy* libraries. Their functionality is implicitly available.

BUGS

The path related functions only operate on elements. Comments, processing instructions or CDATA sections are not taken into account.

XmlGrep() cannot be used to copy tokens outside the root element (XML prologue or epilogue).

SEE ALSO

XML Processing With gawk, *xmlbase*(3am), *xmlcopy*(3am), *xmltree*(3am), *xmlwrite*(3am).

AUTHOR

Manuel Collado, **m-collado@users.sourceforge.net**.

COPYING PERMISSIONS