NAME

detoxrc — configuration file for detox(1)

OVERVIEW

detox allows for configuration of its sequences through config files. This document describes how these files work.

IMPORTANT

When setting up a new set of rules, the *safe* and *wipeup* filters should always be run after a translating filter (or series thereof), such as the *utf_8* or the *uncgi* filters. Otherwise, the risk of introducing difficult characters into the filename is introduced.

SYNTAX

The format of this configuration file is C-like. It is based loosely off the configuration files used by **named**. Each statement is semicolon terminated, and modifiers on a particular statement are generally contained within braces.

sequence "*name*" { . . . };
Defines a sequence of filters to run a filename through. *name* specifies how the user will refer to the particular sequence during runtime. Quotes around the sequence name are generally optional, but should be used if the sequence name does not start with a letter.

There is a special sequence, named *default*, which is the default sequence used by **detox**. This can be overridden through the command line option **-s** or the environmental variable DETOX_SEQUENCE.

Sequence names are case sensitive and unique throughout all sequences; that is, if a system-wide file defines *normal_seq* and a user has a sequence with the same name in their .detoxrc, the users' *normal_seq* will replace the system-wide version.

iso8859_1;

iso8859_1 {**filename** "*/path/to/filename*";};
This transliterates ISO 8859-1 characters between 0xA0 and 0xFF into lower ASCII equivalents. The output is not necessarily safe, and should also be run through the *safe* filter. The default version of this filter also includes transliterations for *CP-1252* for characters between 0x80 and 0x9F.

Under normal circumstances, the filename syntax is not needed. **detox** looks in several locations for a file called iso8859_1.tbl, which is a set of rules defining how an ISO 8859-1 character should be translated. If **detox** can't find the translation table, it will fall back on its internal translation table, based on the stock one.

You can also download or create your own, and tell **detox** the location of it using the filename syntax shown above.

You can chain together multiple *iso8859_1* filters, as long as the default value of all but the last one it empty. This is explained in detox.tbl(5).

This filter is mutually exclusive with the *utf_8* filter.

utf_8;

utf_8 {**filename** "*/path/to/filename*";};
This transliterations Unicode characters, encoded using UTF-8, into lower ASCII equivalents.

This operates in a manner similar to *iso8859_1*, except it looks for a translation table called unicode.tbl.

Similar to the *iso8859_1* filter, an internal table exists, based on the stock translation table.

uncgi;
This translates CGI-escaped strings into their ASCII equivalents. The output of this is not necessarily safe, and should be run through the *safe* filter, at the least.

safe;

safe {**filename** "*/path/to/filename*";};
This could also be called "safe for UNIX-like operating systems". It translates characters that are difficult to work with in UNIX environments into characters that are not.

Similar to the *iso8859_1* and *utf_8* filters, this can be controlled using a translation table. This filter also has an internal version of the translation table based on the stock version.

See the **SAFE** section for more details on what this filter translates by default.

wipeup;

wipeup {**remove_trailing**;};
This condenses sequences of dashes and underscores into a single dash, and removes leading dashes, underscores, and octothorpes (A.K.A. pound signs or hash tags).

The *remove_trailing* option removes a dash or underscore followed immediately by a period.

See the **WIPEUP** section for more details on what this filter translates.

max_length {**length** *value*;};
This trims a filename down to the length specified (or less). It is conscious of extensions and attempts to preserve anything following the last period in a filename.

For instance, given a max length of 12, and a filename of `this_is_my_file.txt`, the filter would output `this_is_.txt`.

lower;
This translates uppercase characters into lowercase characters. It only works on ASCII characters.

 # comments
Any thing after a # on any line is ignored.

EXAMPLE

```
sequence default {
  uncgi;
  iso8859_1 {
    filename "iso8859_1.tbl";
  };
# utf_8 {
#   filename "unicode.tbl";
# };
  safe {
    filename "safe.tbl";
  };
  wipeup {
    remove_trailing;
  };
# max_length {
#   length 128;
# };
```

```
};
```

SAFE

The following characters are translated by the stock *safe* filter. They can be tuned by updating `safe.tbl` or creating a copy of `safe.tbl` and updating your `detoxrc` file.

Rules that apply anywhere in the filename:

Safe	Original
and	&
_	*space* ‘ ! @ $ * \ | : ; " ’ < > ? /
-	( ) [ ] { }

WIPEUP

The following characters are translated by the *wipeup* filter.

Rules that apply anywhere in the filename:

Wipeup	Original
-	-_
-	_-
-	--
_	__

Rules that apply only at the beginning of a filename:

Any leading dashes are stripped to prevent programs from interpreting these files as command line options.

Wipeup	Original
removed	- _ #

Rules that apply when remove trailing is enabled:

Wipeup	Original
.	.-
.	-.
.	._
.	_.

SEE ALSO

`detox`(1), `inline-detox`(1), `detox.tbl`(5), `ascii`(7), `cp1252`(7), `iso_8859-1`(7), `unicode`(7), `utf-8`(7).

AUTHORS

detox was written by Doug Harple.