NAME

regexp — interactive regular-expression tester

SYNOPSIS

regexp [*inputfile*]

DESCRIPTION

The **regexp** command invokes an interactive pattern-matcher that is useful for formulating and refining regular expressions. *Regular expressions* provide a generic means for defining patterns of characters (see tutorial in Section 6). Innumerable UNIX and Humdrum commands make use of regular expressions. The **regexp** command allows the user to test interactively various expressions using a sample text. If no sample text is supplied by the user *(inputfile)* then a short default text is used.

Once invoked, the user may interactively input a regular expression followed by a carriage return. The sample text is scanned for occurrences of the defined regular expression. Any text lines containing the matched expression are displayed on the screen; **regexp** differs from the UNIX **grep** command in that the precise locations of the matched pattern are explicitly marked. (See EXAMPLES below.) Note that only the first occurrence of a matching pattern is identified in each line of text. (This is how most software tools make use of regular expressions.)

The entire sample text file may be viewed by typing the regular expression `.*` or by simply typing a carriage return. Viewing the sample text is helpful in identifying character-strings that are not identified by a given regular expression.

The **regexp** command is terminated by typing an end-of-file marker (control-D on UNIX; control-Z on DOS or OS/2).

The **regexp** command implements the same regular expression features found in the UNIX **awk** command. This includes all so-called "extended" regular expression features with the exception of \> and \<.

OPTIONS

The **regexp** command provides only a help option:

-h displays a help screen summarizing the command syntax

Options are specified in the command line.

EXAMPLES

Imagine the case where the sample text file specified in the command line contains the following three records:

```
The quick brown fox jumped over the lazy dogs.
Once upon a time, long, long ago ...
It was the best of times, it was the worst of times.
```

The following regular expression defines any character string beginning with the lower-case letter 'b', followed by zero or one instance of any single character, followed by a lower-case vowel.

```
b.?[aeiou]
```

Given this regular expression, the corresponding output would appear as follows:

```
The quick brown fox jumped over the lazy dogs.
          ^ ^

It was the best of times, it was the worst of times.
          ^^
```

Notice that only those text lines matching the defined regular expression are displayed in the output.

WARNINGS

The regular-expression features provided by **regexp** depend on the local UNIX **awk** utility — as accessed via the `AWK_VER` shell variable. Available features may change depending on the version of **awk** used.

FILES

The default text file is `$HUMDRUM/regexp.txt`.

PORTABILITY

DOS 2.0 and up, with the MKS Toolkit. OS/2 with the MKS Toolkit. UNIX systems supporting the *Korn* shell or *Bourne* shell command interpreters, and revised *awk* (1985).

SEE ALSO

awk (UNIX), **expr** (UNIX), **grep** (UNIX), **gres** (UNIX), **patt** (4), **pattern** (4), **regexp** (6), **sed** (UNIX)