

NAME

sq, *unsq* – squeeze or unsqueeze a sorted word list

SYNOPSIS

sq < infile > outfile

unsq < infile > outfile

DESCRIPTION

sq compresses a sorted list of words (a dictionary). For example:

```
sort -u /usr/dict/words | sq | gzip -9 > words.sq.Z
```

will compress dict by about a factor of 5.

unsq uncompress the output of *sq*. For example:

```
gunzip < words.sq.gz | unsq | sort -f -o words
```

will uncompress a dictionary compressed with *sq*.

The squeezing is achieved by eliminating common prefixes, and replacing them with a single character which encodes the number of characters shared with the preceding word. The prefix size is encoded as a single printable character: 0-9 represent 0-9, A-Z represent 10-35, and a-z represent 36-61.

BUGS

sq and *unsq* can only handle words of up to 256 characters. The input must be sorted, and duplicates must be suppressed.

AUTHOR

Mike Wexler

SEE ALSO

compress(1), sort(1).