

236369 - Managing Data on the World-Wide Web

Merabi Shmulik Chicvashvili - 317655157

Eyal Tolchinsky - 311505564

12/28/14

Abstract

Ziping algorithms this days are basically based on searching patterns in the source text/file. When such patterns are discovered each ziping algorithm makes its own operations and outputs zipped text/file. Eventually there are more benefits than drawbacks, most of the files contain a lot of patterns and redundant information so ziping is recommended. It is estimated that the total amount of data that is stored on the world's storage devices could be further compressed with existing compression algorithms by a remaining average factor of 4.5:1. It is estimated that the combined technological capacity of the world to store information provides 1,300 exabytes of hardware digits in 2007, but when the corresponding content is optimally compressed, this only represents 295 exabytes of Shannon information.



Benefits

Probabilistic most of the files/texts contain patterns, and big ones! So its easy to understand that we reduce the traffic over the internet and the size of the files. When reducing the size of files there is one immediate benefit, we can easily send more files and resources over one connection. Furthermore, zipped files will reach to the server faster therefore browsing in a site which handles and accepts zipped files will work faster.

Drawbacks

File compression algorithms are intense mathematical operations, therefore compression and decompression of large files are heavy operations which will take time. It's better to use zipping for big amount of small files. Furthermore, in files which will not contain patterns and any redundant information that in that case we will add information more than we will reduce.

Good Scenario

Lets take for example the song 'Get Lucky' by Daft Punk:

Like the legend of the Phoenix All ends with beginnings What keeps
the planets spinning (uh) The force from the beginning

(Look)

We've come too far to give up who we are So let's raise the bar and
our cups to the stars

She's up all night to the sun I'm up all night to get some She's up
all night for good fun I'm up all night to get lucky

We're up all night to the sun

We're up all night to get some

We're up all night for good fun

We're up all night to get lucky

We're up all night to get lucky

We're up all night to get lucky

We're up all night to get lucky

We're up all night to get lucky

The present has no ribbon Your gift keeps on giving, What is this
I'm feeling? If you wanna leave I'm with it (ah)

We've come too far to give up who we are So let's raise the bar and
our cups to the stars

She's up all night to the sun I'm up all night to get some She's up
all night for good fun I'm up all night to get lucky

We're up all night to the sun

We're up all night to get some

We're up all night for good fun

We're up all night to get lucky

We're up all night to get lucky

We're up all night to get lucky

We're up all night to get lucky

We're up all night to get lucky

(We're up all night to get

We're up all night to get

We're up all night to get

We're up all night to get)

(We're up all night to get (together)

As we can see there are many patterns in this song, so it will make a good compressed file.

Bad Scenario

Lets take for example the next phrase:

The quick brown fox jumps over the lazy dog

The phrase contains 43 characters but in the end after zipping it we get 62 characters, this could be explained by that there is no patterns in that phrase.