

Compression and Git

a new object in the repository each time a changed file is committed.

et crowded as a result.

e, it *compresses* each object.

d then (such as when sending or receiving from another it packs objects together into a single file: a "packfile."

sticking the files together, uses a technique called *compression*.

Delta Compression (II)

o versions

V1

fully open to my awful

t once to Roderick and oration. I shall tell him

ed my forgotten moral

V2

My eyes are fully open to my awful situation.

I shall go at once to Roderick and make him an oration.

I shall tell him I've recovered my forgotten moral senses, and don't give twopence halfpenny for any consequences.

V1

lines from V2]

V2

My eyes are fully open to my awful situation.

I shall go at once to Roderick and make him an oration.

I shall tell him I've recovered my forgotten moral senses, and don't give twopence halfpenny for any consequences.

LZ77 and DEFLATE

uses a different scheme from LZW for compression: a of LZ77 and Huffman coding.

of like delta compression, but within the same text.

xt such as

ippi, two Mississippi

ng like

ippi, two <11,7>

1,7> is intended to mean "the next 11 characters come t that ends 7 characters before this point."

symbols to the alphabet to represent these (length, lusions.

Huffman encode the result.

Lecture #38: Compression

s Friday in class. *Extra points awarded to those who*

Delta Compression

ere will be many versions of a file in a Git repository: nd previous edits of it, each in different commits.

keep track explicitly of which file came from where, hard in general:

file is split into two, or two are spliced together?

ss that files with same name and (roughly) same size in are probably versions of the same file.

happens, store one of them as a pointer to the other, changes.

Compression Techniques

h Hug