

Kompresjon

Algoritmer og strategier

Birk og Theodor

Oblig D-2: Alternativ kompresjonsalgoritme

Original:

Gjør kompresjon hver gang man appender en posting.

Posting komprimeres og legges så til.

Komprimeres med Variable Byte Encoding.

Ny:

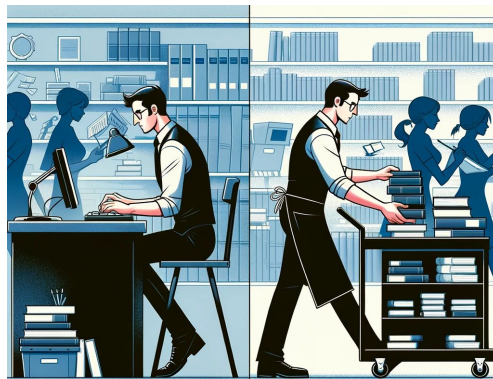
Gjør ingen komprimering før i finalize-steget.

Postings legges til som de er, og komprimeres til slutt når alt er lagt inn.

Komprimeres med DEFLATE-algoritmen gjennom Zlib-biblioteket.

Original: Den grundige bibliotekaren

Bibliotekar som sorterer og setter tilbake bøker med en gang de leveres.



Ny: Bulkhåndteringsbibliotekaren

Bibliotekar som samler opp en haug med bøker, og sorterer og setter tilbake på plass på slutten av dagen.



Fordelen med å gjøre alt i finalization-steget

Vi kan bruke ordbok-baserte kompresjonsalgoritmer!

Vi har brukt Zlib som bruker DEFLATE som bruker Huffman-algoritmen.

Demo av Huffman-komprimering

Testing time to **append** 100000 postings:

Original: 0.1708 seconds

New: 0.0619 seconds

Testing time to **finalize** 100000 postings:

Original: 0.0000 seconds

New: 0.0029 seconds

In total:

Appending and finalizing sum: New is 163.93% faster

Original: 800000 -> 383760 bytes

800000 -> 298228 bytes

Kompresjonsnivåer

Mengden tid brukt på å lage
ordboken / huffman-tree

Compression level 0 took 0.0005 seconds with a compression ratio of 1.00:1

Compression level 1 took 0.0100 seconds with a compression ratio of 2.50:1

Compression level 2 took 0.0134 seconds with a compression ratio of 2.55:1

Compression level 3 took 0.0165 seconds with a compression ratio of 2.72:1

Compression level 4 took 0.0179 seconds with a compression ratio of 2.68:1

Compression level 5 took 0.0238 seconds with a compression ratio of 2.79:1

Compression level 6 took 0.0606 seconds with a compression ratio of 2.86:1

Compression level 7 took 0.0859 seconds with a compression ratio of 2.87:1

Compression level 8 took 0.1934 seconds with a compression ratio of 2.91:1

Compression level 9 took 0.6409 seconds with a compression ratio of 2.93:1

Diskusjon

Hva er best?

Kommer an på kontekst: Hvor ofte legges ny data til?

Eks. masteroppgaver vs. chat-historikk?