



WIKIPEDIA
The Free Encyclopedia

[Main page](#)
[Contents](#)
[Featured content](#)
[Current events](#)
[Random article](#)
[Donate to Wikipedia](#)
[Wikipedia store](#)

Interaction

[Help](#)
[About Wikipedia](#)
[Community portal](#)
[Recent changes](#)
[Contact page](#)

Tools

[What links here](#)
[Related changes](#)
[Upload file](#)
[Special pages](#)
[Permanent link](#)
[Page information](#)
[Wikidata item](#)
[Cite this page](#)

Print/export

[Create a book](#)
[Download as PDF](#)
[Printable version](#)

Languages

[Русский](#)

 [Edit links](#)

[Create account](#) [Log in](#)

Article

[Talk](#)

[Read](#)

[Edit](#)

[More](#) ▾



Incremental encoding

From Wikipedia, the free encyclopedia

Incremental encoding, also known as **front compression**, **back compression**, or **front coding**, is a type of [delta encoding compression algorithm](#) whereby common [prefixes](#) or [suffixes](#) and their lengths are recorded so that they need not be duplicated. This algorithm is particularly well-suited for compressing [sorted data](#), e.g., a list of [words](#) from a [dictionary](#).

For example:

Input	Common prefix	Compressed output
myxa	no preceding word	0 myxa
myxophyta	'myx'	3 ophyta
myxopod	'myxop'	5 od
nab	no common prefix	0 nab
nabbed	'nab'	3 bed
nabbing	'nabb'	4 ing
nabit	'nab'	3 it
nabk	'nab'	3 k
nabob	'nab'	3 ob
nacarat	'na'	2 carat
nacelle	'nac'	3 elle
64 bytes		46 bytes

The encoding used to store the common prefix length itself varies from application to application. Typical techniques are storing the value as a single byte; [delta encoding](#), which stores only the change in the common prefix length; and various [universal codes](#). It may be combined with other general [lossless data compression](#) techniques such as [entropy encoding](#) and [dictionary coders](#) to compress the remaining suffixes.

Applications [\[edit\]](#)

Incremental encoding is widely used in information retrieval to compress the lexicons used in [search indexes](#); these list all the words found in all the documents and a pointer for each one to a list of locations. Typically, it compresses these indexes by about 40%.^[1]

As one example, incremental encoding is used as a starting point by the [GNU locate](#) utility, in an index of filenames and directories. The [GNU locate](#) utility further uses [bigram](#) encoding to further shorten popular filepath prefixes.

References [\[edit\]](#)

- ↑ Ian H. Witten, Alistair Moffat, Timothy C. Bell. Managing Gigabytes. Second edition. Academic Press. ISBN 1-55860-570-3. Section 4.1: Accessing the lexicon, subsection Front coding, pp.159–161.



*This [computer storage](#)–related [software](#) article is a *stub*. You can help Wikipedia by [expanding it](#).*

Categories: [Lossless compression algorithms](#) | [Database index techniques](#) | [Storage software stubs](#)

This page was last modified on 16 April 2013, at 10:44.

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.

[Privacy policy](#) [About Wikipedia](#) [Disclaimers](#) [Contact Wikipedia](#) [Developers](#) [Mobile view](#)



a
WIKIMEDIA
project



Powered By
MediaWiki