

# HexaURL Character Code

Yota Inomoto

“HexaURL is a fixed-length, compressed, case-insensitive, URL-safe text format. It is designed to serve as a unique identifier for various applications, including short URLs, database record IDs, and file hash values, all while enabling rapid searches.”,

February 09, 2025

## Contents

1. Goals .....	3
2. Context .....	4
3. Overview .....	5

## 1. Goals

- Define a fixed-length, compressed, case-insensitive, URL-safe character encoding.
- Ensure that the encoding is compatible with search algorithms and more efficient than standard ASCII.
- Provide negligibly efficient methods for converting between HexaURL and ASCII.

## 2. Context

Although UTF-8 strings are ubiquitous in modern software and networks, practical URI safety mandates the use of only ASCII alphanumeric characters, hyphens, and underscores. Since many systems handle strings in a case-insensitive manner, it is beneficial to employ identifiers that ignore letter casing. In practical terms, if percent encoding is not applied, the necessary character set consists of 38 characters—26 letters (A–Z), 10 digits (0–9), and 2 delimitation marks (hyphen and underscore). Given that

$$5 < \log_2 38 \leq 6 \tag{1}$$

, each character can be efficiently encoded using 6 bits.

Historically, when “1 byte = 8 bits” was not the prevailing standard, systems such as those from Digital Equipment Corporation (DEC) utilized 6-bit character codes on machines with word lengths in multiples of 6. [DEC SIXBIT](#)<sup>°</sup> is a six-bit representation covering 64 characters, typically ranging from the space character up to the underscore in the [ASCII table](#)<sup>°</sup>, while excluding non-printable control characters. This encoding maps each character to the lower 6 bits of its ASCII code after subtracting 32, thereby providing the necessary components—digits, uppercase letters, hyphens, and underscores—for a case-insensitive and URL-safe set.

Building on these established principles, **HexaURL** is defined as a compressed character set that rigorously validates URL-safe characters based on the DEC SIXBIT scheme, while also supporting seamless conversion to and from standard ASCII. Most notably, when used as the key in collections, HexaURL offers enhanced search efficiency. Its fixed-length encoding and streamlined character set not only reduce computational overhead during comparisons but also enable rapid, scalable lookups in various data models.

### **3. Overview**