

Proposal for C2X

WG14 N2549

Title: Allow for binary integer constants
Author: Jörg Wunsch <j.gnu@uriah.heep.sax.de>
Date: 2020-07-30
Proposal category: New feature
Reference: N2478 ISO/IEC 9899 working draft February 2020

Abstract:

Binary constants for integer numbers, consisting solely of the digits '0' and '1', are currently missing from the standard. By now, the standard specifies integer numbers for base 8, 10, and 16.

Internally, all current computer systems use binary numbers to represent data, even though their presentation at the user interface usually differs a lot (decimal integers, floating-point numbers, printable characters). However, developers working at the lowest levels of computer software are frequently confronted with the actual binary representation inside the machine, and have thus repeatedly expressed the wish to be able to specify that number format in C source code in the past.

A number of existing C compilers have thus added such an extension during the last two decades.

Background

With C being the predominant language these days in machine-level software (ranging from tiny microcontrollers with memory footprints in the kilobyte range to operating systems of versatile computer systems), developers are frequently confronted with data representations at the bit-level. While many C programmers can comprehend bits to more compact number representations like hexadecimal (where four bits form one printed character), developers operating at that level have repeatedly expressed the wish to be able to specify binary constants in C language in the past. This can sometimes lead to better comprehensible code (e.g. for bit patterns that eventually form graphic elements like glyphs of a font), or it can be better matched to bit-level documentation of the hardware registers, to name two examples.

State of the Art

Due to that, a number of compilers have meanwhile added binary integer constants as a language extension. Pioneering in that respect were compilers primarily targetted for microcontroller environments about 20 years ago. Later on, larger compilers like GCC, SDCC, and Clang joined in.

There appears to be consensus to specify binary integer constants similar to the existing way for hexadecimal constants, with the string `0b` as introductory sequence, and obviously the digits `0` and `1` as the only allowable digits inside the number. This can easily fit into the lexical scanner of existing compilers, and due to its similarity to hexadecimal constants, it is easy to cophrend to any C developer.

Consequently, developers started using that feature. As an example, searching on the common opensource site `github.com` for `0b00001111` (just one out of many possible binary constants) yields almost 6000 occurrences. Not necessarily all of them belong to C source code, but the vast majority does, and it outlines the feature is important even for other programming languages as well.

C++ standardized the feature in C++14.

Thus, it would make sense to standardize the existing practice.

Suggested change:

6.4.4.1 Integer constants

Syntax

Change

```
integer-constant:  
    decimal-constant integer-suffixopt  
    octal-constant integer-suffixopt  
    hexadecimal-constant integer-suffixopt
```

to

```
integer-constant:  
    decimal-constant integer-suffixopt  
    octal-constant integer-suffixopt  
    hexadecimal-constant integer-suffixopt  
    binary-constant integer-suffixopt
```

After the definition of *hexadecimal-constant*, insert

binary-constant:
 binary-prefix binary-digit
 binary-constant binary-digit

After the definition of *hexadecimal-prefix*, insert

binary-prefix: one of
 0b 0B

After the definition of *hexadecimal-digit*, insert

binary-digit: one of
 0 1

Description

To paragraph 3, append the sentence

A binary constant consists of the prefix 0b or 0B followed by a sequence of the digits 0 or 1.

Semantics

Change the first sentence of paragraph 4 from

The value of a decimal constant is computed base10; that of an octal constant, base8; that of a hexadecimal constant, base16.

to

The value of a decimal constant is computed base10; that of an octal constant, base8; that of a hexadecimal constant, base16; that of a binary constant, base2.

Acknowledgements

I would like to thank Manuel López-Ibáñez, Gerald Pfeifer, Dave Korn, Mark Mitchell, Joseph S. Myers for supporting me in implementing binary constants in GCC, many years ago.

Special thanks to Philipp Klaus Krause for helping me out with all the WG14 procedures.